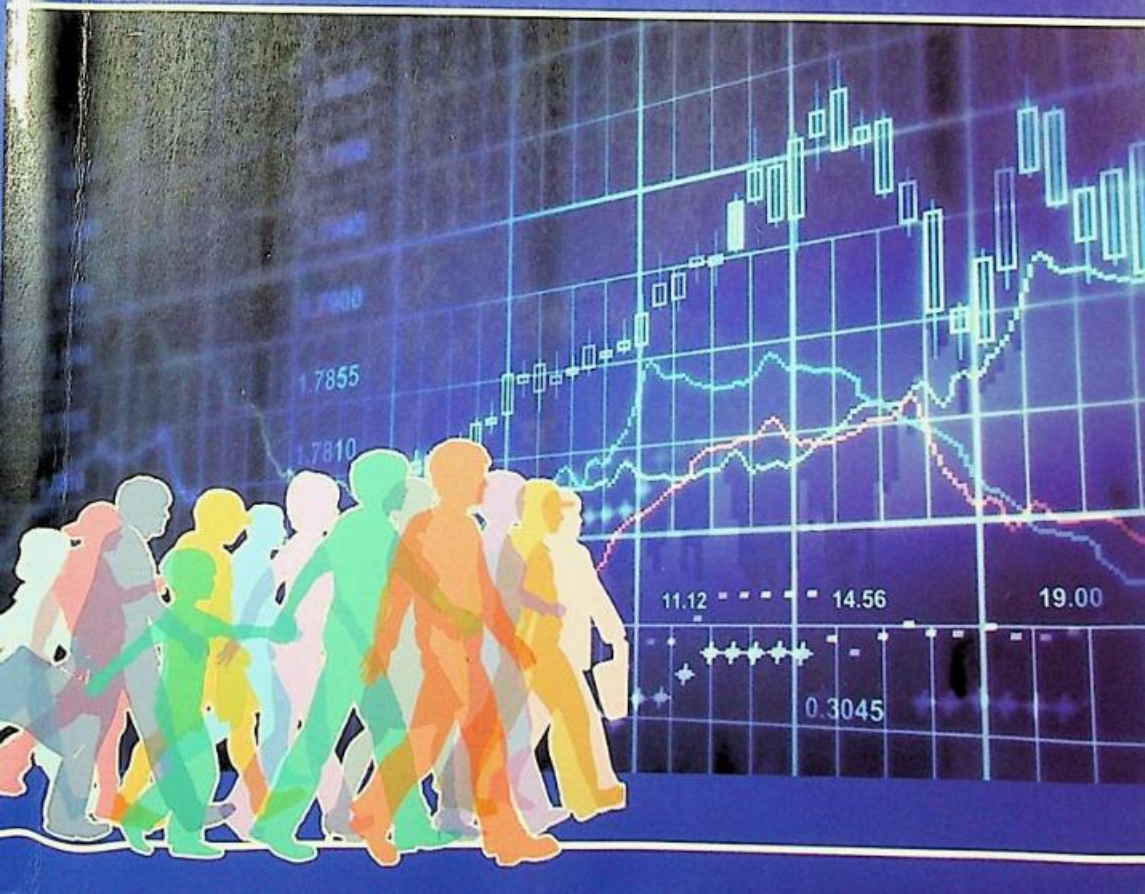
31:316
ნ-36

ჩაჯა ფრანკფორტ ნაჩმიასი ანა ლეონ-გერერო

სოციალური სტატისტიკა

მრავალფეროვანი საზოგადოებისათვის

მეექვსე გამოცემა



31:316
ნ-36

სოციალური სტატისტიკა

მრავალფეროვანი საზოგადოებისათვის

მეექვსე გამოცემა

სოციალური სტატისტიკა

მრავალფეროვანი საზოგადოებისათვის

მეექვსე გამოცემა

ჩავა ფრანკფორტ ნაჩმიასი
ვისკონსინ-მილუოკის უნივერსიტეტი

ანა ლეონ-გერერო
წყნარი ოკეანის ლუთერანული უნივერსიტეტი

საქართველოს უნივერსიტეტების ასოციაციის ბიბლიოთეკა
№ 41014

U.S. – GEORGIA
CELEBRATING TWENTY YEARS

დაფინანსებულია საქართველოში აშშ-ის საელჩოს წიგნების თარგმნის პროგრამის ფარგლებში

Funded through the Book Translation Program, U.S. Embassy in Georgia

თბილისის უნივერსიტეტის გამომცემლობა

ჩავა ფრანკფორტ ნაჩმიასისა და ანა ლეონ-გერეროს წიგნის „სოციალური სტატისტიკა მრავალფეროვანი საზოგადოებისათვის“, ინგლისური მე-6 გამოცემის ავტორიზებული თარგმანი. გამოცემის უფლების მფლობელი: „SAGE PUBLICATIONS INC“.

Chava Frankfort-Nachimas, Anna Leon-Guerrero “Social Statistics for a Diverse Society”, Original Publisher: SAGE PUBLICATIONS INC (Original Publisher in the United States, London and New Delhi). 2011 ISBN 978-1-4129-9253-4

სამეცნიერო რედაქტორი: სრული პროფესორი *სიმონ გელაშვილი*

რედაქტორი: სრული პროფესორი *ქეთევან მარშავა*

თარგმნელები: *მარინე მინდორაშვილი* (1-10 თავები)
maia.mindorashvili@tsu.ge
გიორგი კ. ყუფარაძე (11-14 თავები)
giorgi.kuparadze@tsu.ge

დაკაბადონება: *ზაზა გულაშვილი*
zaza.gulashvili@tsu.ge

წიგნის ქართული თარგმანი მომზადდა ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტში, ეკონომიკური და სოციალური სტატისტიკის კათედრაზე, საქართველოში ამერიკის შეერთებული შტატების საელჩოს ფინანსური მხარდაჭერით.

ყველა საავტორი უფლება დაცულია. SAGE PUBLICATIONS, INC.-ის თანხმობის გარეშე არ შეიძლება წიგნის რომელიმე ნაწილის ასლის გადაღება, გამრავლება ან გავრცელება რაიმე ფორმით, ელექტრონული ან მექანიკური საშუალებებით, ფოტოკოპირებით ან ინფორმაციული შენახვის რაიმე სხვა საშუალებით.

© თბილისის სახელმწიფო უნივერსიტეტი, 2012

ISBN 987-9941-0-4804-3

მოკლე სარჩევი

1. სტატისტიკა – რატომ და რისთვის?	1
2. ინფორმაციის თავმოყრა სიხშირული განაწილებები	32
3. გრაფიკული პრეზენტაცია	73
4. ცენტრალური ტენდენციის საზომები	105
5. ვარიაციის საზომები	145
6. ნორმალური განაწილება	185
7. შერჩევა და შერჩევითი განაწილებები	214
8. შეფასება	248
9. ჰიპოთეზების ტესტირება	279
10. ურთიერთკავშირი ორ ცვლადს შორის – კროს-ტაბულაცია	321
11. ხი-კვადრატ კრიტერიუმი	369
12. ნომინალურ და ორდინალურ ცვლადებს შორის კავშირის გაზომვა	402
13. რეგრესია და კორელაცია	442
14. დისპერსიული ანალიზი	503
დანართი ა	543
დანართი ბ	546
დანართი გ	551
დანართი დ	553

დანართი ე 554
დანართი ვ 556
ამოხსნები რუბრიკისათვის
„ცოდნის შემოწმება“ 563
კენტ ნომრიანი სავარჯიშოების
პასუხები 573
გლოსარიუმი 621
შენიშვნები 628

სარჩევი

1. სტატისტიკა – რატომ და რისთვის? 1
კვლევის პროცესი 2
კვლევის კითხვების დასმა 3
თეორიის როლი 5
ჰიპოთეზების ჩამოყალიბება 6
დამოუკიდებელი და დამოკიდებული ცვლადები: მიზეზობრიობა 9
დამოუკიდებელი და დამოკიდებული ცვლადები 11
მონაცემების შეგროვება 12
გაზომვის დონეები 13
გაზომვის ნომინალური დონე 14
გაზომვის ორდინალური დონე 14
ინტერვალურ-ფარდობითი გაზომვის დონე 15
გაზომვის დონეების კუმულაციური მახასიათებელი 15
დიქოტომიური ცვლადების გაზომვის დონეები 16
დისკრეტული და უწყვეტი ცვლადები 19
მონაცემთა ანალიზი და ჰიპოთეზების შეფასება 20
აღწერითი და დასკვნითი სტატისტიკა 21
ჰიპოთეზების შეფასება 22
სოციალური განსხვავებების გათვალისწინება 23
სტატისტიკა SPSS-ის გამოყენებით 27
ამოცანა SPSS-ისათვის 30
თავის საკარგიშოები 30

2. ინფორმაციის თავმოყრა სიხშირული განაწილებები 32
სიხშირული განაწილებები 33
კოეფიციენტები და პროცენტები 34
ხვედრითი წილების (პროცენტული) განაწილება 37
შედარებები 38

სტატისტიკა პრაქტიკაში: სამუშაო ძალის აქტივობა ლათინო-ამერიკულ მოსახლეობაში 39
სიხშირული განაწილებების აგება 41
სიხშირული განაწილებები ნომინალური ცვლადებისთვის 41
სიხშირული განაწილებები ორდინალური ცვლადებისთვის 44
სიხშირული განაწილებები ინტერვალურ-ფარდობითი ცვლადებისთვის 47
კუმულაციური განაწილებები 49
დონეები 54
სტატისტიკა პრაქტიკაში: ქორწინებათა და განქორწინებათა დროითი დინამიკა 55
კვლევითი ლიტერატურის მიმოხილვა: სტატისტიკური ცხრილები 56
საბაზისო პრინციპები 56
სხვადასხვა ფორმატის ცხრილები 59
დასკვნა 60
სტატისტიკა SPSS-ის გამოყენებით 62
ამოცანები SPSS-ისთვის 65
თავის სავარჯიშოები 65

3. გრაფიკული პრეზენტაცია 73
წრიული დიაგრამა: განდაზმულთა რასა და ეთნიკური ნიშანი 74
სვეტებიანი დიაგრამა: განდაზმულთა საცხოვრებელი პირობები და ეკონომიკური აქტივობა 76
სტატისტიკური რუკა: განდაზმულების გეოგრაფიული განაწილება 79
ჰისტოგრამა 80
სტატისტიკა პრაქტიკაში: ამერიკის „დაბერება“ 82
ხაზოვანი დიაგრამა 84
დროითი მწკრივების დიაგრამები 87
სტატისტიკა პრაქტიკაში: მრავალფეროვნება ერთი თვალის შევლებით 91
სტატისტიკა SPSS-ის გამოყენებით 95
ამოცანები SPSS-ისთვის 98
თავის სავარჯიშოები 99

4. ცენტრალური ტენდენციის საზომები 105
მოდა 106
მედიანა 109
მედიანის პოვნის ხერხი თანმიმდევრობით დალაგებული მონაცემებისათვის 110

დაკვირვებათა კენტი რიცხვი 110
დაკვირვებათა ლუწი რიცხვი 112
მედიანის პოვნის ხერხი სიხშირულ განაწილებებში 114
სტატისტიკა პრაქტიკაში: შემოსავლების გენდერული უთანაბრობა 115
პერცენტილების ადგილი სიხშირულ განაწილებაში 116
საშუალო არითმეტიკული 117
საშუალოს გაანგარიშება 119
არითმეტიკული საშუალოს რამდენიმე მნიშვნელოვანი თვისება 122
გაზომვის ინტერვალურ-ფარდობითი დონე 122
მიზიდულობის ცენტრი 123
მგრძნობიარობა ექსტრემალური (კიდურა) წევრების მიმართ 123
განაწილების ფორმა: ტელევიზია, განათლება და და-ძმები 126
სიმეტრიული განაწილება 127
პოზიტიურად გადახრილი განაწილება 128
ნეგატიურად გადახრილი განაწილება 130
მითითებები განაწილების ფორმის განსაზღვრისათვის 130
ცენტრალური ტენდენციის საზომის შერჩევა 131
გაზომვის დონე 132
გადახრილი განაწილება 132
სიმეტრიული განაწილება 133
სტატისტიკა SPSS-ის გამოყენებით 135
ამოცანები SPSS-ისთვის 138
თავის სავარჯიშოები 139

5. ვარიაციის საზომები 145
ვარიაციის გაზომვის მნიშვნელობა 146
თვისობრივი ვარიაციის ინდექსი: მოკლე შესავალი 148
IQV-ის გაანგარიშების საფეხურები 149
IQV-ს ასახვა პროცენტულად 151
სტატისტიკა პრაქტიკაში: მრავალფეროვნება აშშ-ის საზოგადოებაში 152
დიაპაზონი 153
ინტერკვარტილური დიაპაზონი: მატება უფროსი ასაკის მოსახლეობაში 155
მართკუთხა დიაგრამა 158
დისპერსია და სტანდარტული გადახრა: ცვლილებები ასაკოვან მოსახლეობაში 161
გადახრის გაანგარიშება საშუალოდან 162
დისპერსიისა და სტანდარტული გადახრის გაანგარიშება 165
ვარიაციის საზომის შეჩევის საკითხები 169
კვლევითი ლიტერატურის მიმოხილვა: განსხვავებები ლათინო-ამერიკელი მოზარდების

მისწრაფებებსა და მოლოდინებს შორის კოლეჯში შესვლამდე 170
სტატისტიკა SPSS-ის გამოყენებით 173
ამოცანები SPSS-ისთვის 177
თავის სავარჯიშოები 178

6. ნორმალური განაწილება 185

ნორმალური განაწილების დამახასიათებელი ნიშნები 186
ნორმალურ განაწილებასთან მიახლოებული ემპირიული განაწილებები 187
მაგალითი: ფინალური ქულები სტატისტიკაში 187
არეები ნორმალური მრუდის ქვეშ 189
სტანდარტული გადახრის ინტერპრეტაცია 190
სტანდარტული (Z) ქულები 190
ნედლი ქულის გადაყვანა Z ქულაში 191
Z ქულის გადაყვანა ნედლ ქულაში 191
სტანდარტული ნორმალური განაწილება 193
სტანდარტული ნორმალური ცხრილი 194
სტანდარტული ნორმალური ცხრილის სტრუქტურა 194
Z ქულის გამოსახვა კოეფიციენტითა და პროცენტით 196
არის პოვნა საშუალოსა და რაიმე დადებით Z ქულას შორის 196
არის პოვნა საშუალოსა და რაიმე უარყოფით Z ქულას შორის 197
არის პოვნა დადებითი Z ქულის ზემოთ და უარყოფითი Z ქულის ქვემოთ 198
კოეფიციენტების (ან პროცენტების) გარდაქმნა Z ქულებად 199
Z ქულის პოვნა ზედა მოსაზღვრე არიდან 200
Z ქულის პოვნა ქვედა მოსაზღვრე არიდან 200
პერცენტილების გამოყენება ნორმალურ განაწილებაში 202
საშუალოზე მაღალი პერცენტილური რანგის პოვნა 202
საშუალოზე დაბალი პერცენტილური რანგის პოვნა 203
50-ზე მაღალი პერცენტილური რანგის შესაბამისი ნედლი ქულის პოვნა 204
50-ზე დაბალი პერცენტილური რანგის შესაბამისი ნედლი ქულის პოვნა 205
ბოლო შენიშვნა 206
სტატისტიკა SPSS-ის გამოყენებით 207
ამოცანები SPSS-ისთვის 209
თავის სავარჯიშოები 210

7. შერჩევა და შერჩევითი განაწილებები 214

შერჩევის მიზნები 215
ალბათობის ძირითადი პრინციპები 217
ალბათობის განსაზღვრა 217
ფარდობითი სიხშირის მეთოდი 218
ნორმალური განაწილება და ალბათობები 219
ალბათური შერჩევა 219
მარტივი შემთხვევითი შერჩევა 220
სისტემატური შემთხვევითი შერჩევა 222
სტრატიფიცირებული შემთხვევითი შერჩევა 223
შერჩევითი განაწილების არსი 225
მთლიანი ერთობლიობა 226
შერჩევითი ერთობლიობა 227
დილემა 228
შერჩევითი განაწილება 228
საშუალო სიდიდის შერჩევითი განაწილება 229
თვალსაჩინოება 229
მიმოხილვა 232
შერჩევითი განაწილების საშუალო სიდიდე 232
საშუალოს სტანდარტული შეცდომა 233
ცენტრალური ზღვრის თეორემა 234
შერჩევის ზომა 237
შერჩევითი დაკვირვებისა და ცენტრალური ზღვრის თეორემის მნიშვნელობა 238
სტატისტიკა SPSS-ის გამოყენებით 241
ამოცანები SPSS-ისთვის 244
თავის სავარჯიშოები 244

8. შეფასება 248
შეფასების განსაზღვრა 249
შეფასების მიზეზები 250
წერტილოვანი და ინტერვალური შეფასება 250
ნდობის ინტერვალების შეფასების პროცედურები საშუალოებისთვის 252
ნდობის ინტერვალის განსაზღვრა 254
საშუალოს სტანდარტული შეცდომის გაანგარიშება 254
ნდობის ხარისხისა და მისი შესაბამისი Z სიდიდის პოვნა 255
ნდობის ინტერვალის გაანგარიშება 255
შედეგების ინტერპრეტაცია 255
რისკის შემცირება 256
სიგმას გაანგარიშება 258
საშუალოსათვის შეფასებული სტანდარტული შეცდომის გაანგარიშება 258

ნდობის ხარისხის შერჩევა და შესაბამისი Z სიდიდის პოვნა 258
ნდობის ინტერვალის გაანგარიშება 258
შედეგების ინტერპრეტაცია 259
ერთობლიობის ზომა და ნდობის ინტერვალები 259
სტატისტიკა პრაქტიკაში: ლათინო-ამერიკელების მიგრაცია და შემოსავლები 262
ნდობის ინტერვალები ხვედრითი წილებისთვის 264
პროპორციების შეფასების პროცედურები 266
პროპორციის სტანდარტული შეცდომის შეფასების გაანგარიშება 267
ნდობის ხარისხისა და მისი შესაბამისი Z სიდიდის პოვნა 267
ნდობის ინტერვალის გაანგარიშება 267
შედეგების ინტერპრეტაცია 267
სტატისტიკა პრაქტიკაში: ჯანმრთელობის დაცვის რეფორმა 268
პროპორციის შეფასებული სტანდარტული შეცდომის გაანგარიშება 269
ნდობის სასურველი ხარისხისა და მისი შესაბამისი Z სიდიდის განსაზღვრა 269
ნდობის ინტერვალის გაანგარიშება 269
შედეგების ინტერპრეტაცია 270
სტატისტიკა SPSS-ის გამოყენებით 272
ამოცანები SPSS-ისთვის 275
თავის სავარჯიშოები 275

9. ჰიპოთეზების ტესტირება 279
ჰიპოთეზების სტატისტიკური ტესტირების პირობები 280
კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება 281
კვლევითი ჰიპოთეზა (H_1) 281
ნულოვანი ჰიპოთეზა (H_0) 282
უფრო ვრცლად კვლევითი ჰიპოთეზის შესახებ: ცალმხრივი და ორმხრივი ტესტები 283
საკმარისი არადამაჯერებლობის განსაზღვრა: ალბათური სიდიდეები და ალფა 284
ხუთი საფეხური ჰიპოთეზის ტესტირებაში: შეჯამება 289
შეცდომები ჰიპოთეზის ტესტირებაში 290
t სტატისტიკა და სტანდარტული შეცდომა 291
t განაწილება და თავისუფლების ხარისხი 292
t და Z სტატისტიკების შედარება 293
სტატისტიკა პრაქტიკაში: თეთრკანიანი ქალების შემოსავლები 294
ორი შერჩევის ჰიპოთეზების ტესტირება 297
დამოუკიდებელი შერჩევის პირობა 298

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება 298
საშუალოებს შორის განსხვავების შერჩევითი განაწილება 299
სტანდარტული შეცდომის შეფასება 300
შეფასებული სტანდარტული შეცდომის გაანგარიშება 300
t სტატისტიკა 301
თავისუფლების ხარისხის გაანგარიშება საშუალოებს შორის განსხვავების ტესტისთვის 301
ჰიპოთეზების ტესტირება საშუალოებს შორის განსხვავების შესახებ: შეჯამება 302
სტატისტიკა პრაქტიკაში: ნარკოტიკების გამოყენების პირველი შემთხვევა სტუდენტებს შორის 304
ორი შერჩევის ხვედრით წილებს შორის განსხვავების მნიშვნელოვნობის ტესტირება 307
სტატისტიკა პრაქტიკაში: პარტიულობა და მთავრობისადმი ნდობა 307
სტატისტიკა პრაქტიკაში: განათლება და მთავრობისადმი ნდობა 309
კვლევითი ლიტერატურის მიმოხილვა: სტატისტიკური ჰიპოთეზის ტესტირების ანგარიში 311
სტატისტიკა SPSS-ის გამოყენებით 314
ამოცანები SPSS-ისთვის 317
თავის სავარჯიშოები 318

10. ურთიერთკავშირი ორ ცვლადს შორის – კროს-ტაბულაცია 321
დამოუკიდებელი და დამოკიდებული ცვლადები 322
როგორ შევქმნათ ბივარიაციული ცხრილი: რასა და ბინის მფლობელობა 323
როგორ გავიანგარიშოთ პროცენტები ბივარიაციულ ცხრილში 326
პროცენტების გაანგარიშება დამოუკიდებელი ცვლადის ყველა კატეგორიისათვის 326
დამოუკიდებელი ცვლადის სხვადასხვა კატეგორიების პროცენტების შედარება 327
რომელი მოვიძიოთ ცვლადებს შორის არაერთგვაროვანი ურთიერთკავშირის შემთხვევაში 329
კვლევითი ლიტერატურის მიმოხილვა: გარდაცვალების ადგილი ამერიკაში 331
ბივარიაციული ურთიერთკავშირის მახასიათებლები 334
ურთიერთკავშირის არსებობა 334
ურთიერთკავშირის სიმჭიდროვე 335
ურთიერთკავშირის მიმართულება 336
შემაჯამებელი 339

არამცდარობის ტესტირება: მეხანძრეები და ქონების დაზიანება 340
ჩარეული ურთიერთკავშირი: რელიგია და შეზღუდულება აბორტზე 344
პირობითი ურთიერთკავშირები: უფრო მეტი აბორტის შესახებ 349
შემუშავების საზღვრები 351
სტატისტიკა პრაქტიკაში: ოჯახის მხარდაჭერა საშუალო სკოლის დამთავრების შემდეგ 352
სტატისტიკა SPSS-ის გამოყენებით 356
ამოცანები SPSS-ისთვის 359
თავის საკვანძო სიტყვები 361

11. ხი-კვადრატ კრიტერიუმი 369
ხი-კვადრატის კონცეფცია, როგორც სტატისტიკური კრიტერიუმი 372
სტატისტიკური დამოუკიდებლობის კონცეფცია 373
ჰიპოთეზის შემოწმების სტრუქტურა ხი-კვადრატ კრიტერიუმის მიხედვით 374
დაშვებები 374
კვლევითი და ნულოვანი ჰიპოთეზების ჩამოყალიბება 374
მოსალოდნელი სიხშირეების კონცეფცია 375
მოსალოდნელი სიხშირეების გამოთვლა 375
თანხმობის ხი-კვადრატ კრიტერიუმის გამოთვლა 377
ხი-კვადრატის შერჩევითი განაწილება 379
თავისუფლების ხარისხის განსაზღვრა 380
საბოლოო გადაწყვეტილების მიღება 382
მიმოხილვა 382
სტატისტიკა პრაქტიკაში: განათლება და ჯანმრთელობის მდგომარეობის შეფასება 387
კვლევითი ლიტერატურის მიმოხილვა: და-ძმების მხარდაჭერა და აკადემიური მიღწევები 389
კვლევითი ლიტერატურის მიმოხილვა: ძალადობრივი ქცევა სქესის, ასაკისა და რასის მიხედვით 392
სტატისტიკა SPSS-ის გამოყენებით 394
ამოცანები SPSS-ისათვის 396
თავის საკვანძო სიტყვები 397

12. ნომინალურ და ორდინალურ ცვლადებს შორის კავშირის გაზომვა 402
შეცდომის პროპორციული შემცირება: მოკლე მიმოხილვა 402
ლამბდა: სტატისტიკური კავშირის გაზომვა ნომინალური ცვლადებისათვის 406
ლამბდას გამოთვლის მეთოდი 408
ლამბდას გამოთვლის ზოგიერთი ასპექტი 410
კრამერის V: ხი-კვადრატთან დაკავშირებული

სტატისტიკური კავშირის საზომი ნომინალური ცვლადებისათვის 411
გამა: კავშირის საზომი ორდინალური ცვლადებისათვის 412
ორდინალურ ცვლადებს შორის კავშირის ანალიზი: განათლება და ფინანსური მდგომარეობით კმაყოფილება 413
წყვილების შედარება 414
წყვილთა ტიპები 416
წყვილთა რაოდენობის დათვლა 417
მსგავსი წყობის წყვილები 417
საპირისპირო წყობის წყვილები 418
დაკავშირებული წყვილები 419
გამას გამოთვლა 420
დადებითი და უარყოფითი გამა 421
გამა, როგორც შეცდომის პროპორციული შემცირების საზომი 422
კენდალის ტაუ-ბ-ს გამოთვლა 423
სტატისტიკა პრაქტიკაში: განათლება და ჰომოსექსუალობის მიმართ დამოკიდებულება 424
ორდინალურ ცვლადებს შორის კავშირის საზომების გამოყენება დიქოტომიური ცვლადებისათვის 425
კვლევითი ლიტერატურის მიმოხილვა: მსოფმხედველობრივი შეხედულებები და აბორტის შესახებ აზრი 426
მონაცემთა განხილვა 428
მონაცემთა ინტერპრეტირება 429
სტატისტიკა SPSS-ის გამოყენებით 431
ამოცანები SPSS-სათვის 434
თავის საკვანძო სიტყვები 435

13. რეგრესია და კორელაცია 442
წერტილთა გაბნევის დიაგრამა 444
წრფივი კავშირი და პროგნოზირების წესები 446
წრფეების აგება 449
რეგრესიის წრფის პოვნა 451
დეტერმინისტული შეცდომა 452
სხვაობების კვადრატთა ჯამი ($\sum e2$) 453
უმცირეს კვადრატთა წრფე 453
მიმოხილვა 453
a და b პარამეტრების გამოთვლა რეგრესიის განტოლებისათვის 454
a და b კოეფიციენტების ინტერპრეტირება 457
სტატისტიკა პრაქტიკაში: საოჯახო მეურნეობების საშუალო შემოსავლები და პრიმინალური ძმედებები 459

რეგრესიის სიზუსტის შეფასების მეთოდები 463
პროგნოზირების შეცდომა 465
დეტერმინაციის r^2 კოეფიციენტი, როგორც შეცდომის პროპორციული შემცირების საზომი 467
r^2-ის გამოთვლა 470
პირსონის კორელაციის კოეფიციენტი (r) 471
პირსონის r-ის თვისებები 472
სტატისტიკა პრაქტიკაში: ფეხმძიმობა მოზარდებში და სოციალური უთანასწორობა 473
სტატისტიკა პრაქტიკაში: ქორწინება და შემოსავლები 477
მრავლობითი რეგრესია 479
სტატისტიკა SPSS-ის გამოყენებით 484
ამოცანები SPSS-ისთვის 491
თავის სავარჯიშოები 493

14. დისპერსიული ანალიზი 503
დისპერსიური ანალიზის შინაარსი 505
ჰიპოთეზის შემოწმების სტრუქტურა დისპერსიული ანალიზის მიხედვით 507
დაშვებები 507
კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება, ალფას მნიშვნელობის არჩევა 507
შიგაჯგუფური და ჯგუფთაშორისი დიპერსიის კონცეფცია 508
F სტატისტიკა 511
გადაწყვეტილების მიღება 513
შეჯამება: ჰიპოთეზის შემოწმების ხუთი ნაბიჯი 515
R^2 მნიშვნელოვნების შეფასება დისპერსიული ანალიზის გამოყენებით 518
დისპერიული ანალიზი მრავლობითი წრფივი რეგრესიისათვის 522
კვლევითი ლიტერატურის მიმოხილვა: პიროვნული თვითშეფასება და ეთნიკური იდენტობა 523
კვლევითი ლიტერატურის მიმოხილვა: სტრესი და დაძაბულობა 525
სტატისტიკა SPSS-ის გამოყენებით 529
ამოცანები SPSS-ისთვის 534
თავის სავარჯიშოები 535

დანართი ა 543
დანართი ბ 546
დანართი გ 551

დანართი დ 553
დანართი ე 554
დანართი ვ 556
ამოხსნები რუბრიკისათვის
„ცოდნის შემოწმება“ 563
კენტ ნომრიანი სავარჯიშოების
პასუხები 573
გლოსარიუმი 621
შენიშვნები 628

ავტორების შესახებ

ჩავა ფრანკფორტ-ნაჩმიასი ახლა ემერიტუს-პროფესორია ვისკონსინის უნივერსიტეტში, მილუოკიში, სადაც აქამდე ებრაული სწავლების ცენტრს ხელმძღვანელობდა. ის კითხულობდა ლექციებს კვლევის მეთოდებში, სტატისტიკასა და გენდერში. დევიდ ნაჩმიასთან ერთად გამოაქვეყნა „კვლევის მეთოდები სოციალურ მეცნიერებებში“. მისი თანარედაქტორობით გამოიცა „საფო წმინდა მიწაზე“. იგი აგრეთვე არის ავტორი მრავალრიცხოვანი პუბლიკაციებისა ეთნიკურ კუთვნილებასა და განვითარებაზე, ურბანული გარემოს გაჯანსაღებაზე, მეცნიერებასა და სქესზე, ქალების ცხოვრების შესახებ ისრაელში. ვისკონსინის უნივერსიტეტის სისტემური სწავლების გაუმჯობესებისთვის მან დაიმსახურა გრანტი: „რასის, ეთნიკურობისა და გენდერის ინტეგრაცია სოციალური სტატისტიკისა და კვლევის მეთოდების კურიკულუმებში“.

ანა ლეონ-გერერო სოციოლოგიის პროფესორია წყნარი ოკეანის ლუთერანულ უნივერსიტეტში, ვაშინგტონში. სოციოლოგიის დოქტორის აკადემიური ხარისხი მიენიჭა კალიფორნიის უნივერსიტეტში, ლოს-ანჯელესში. ლექციებს კითხულობს სტატისტიკაში, სოციალურ თეორიასა და სოციალურ პრობლემებში. მისი კვლევებისა და პუბლიკაციების სფეროებია: საოჯახო ბიზნესი, სოციალური კეთილდღეობის პოლიტიკა და სოციალური მომსახურების პროგრამის შეფასება. ის აგრეთვე არის ავტორი მონოგრაფიისა „სოციალური პრობლემები: საზოგადოება, პოლიტიკა და სოციალური მოქმედება“ და „თანამედროვე სოციალური პრობლემების“ თანარედაქტორი.

შესავალი

ამ შესავალს, ალბათ, სტატისტიკის შესწავლის პირველივე კვირაში წაიკითხავთ და კითხვებიც დაგებადებათ სტატისტიკის რაობის შესახებ. მაინც რა წარმოდგენა გაქვთ მასზე? მათემატიკა, ფორმულები, გამოთვლები? დიახ, რა თქმა უნდა, მაგრამ ეს მხოლოდ ნაწილია. გარდა ამისა, კიდევ ბევრი რამაა სასწავლი.

სტატისტიკა ყოველდღიურობისა და პროფესიული საქმიანობის თანამდევია, ის ჩვენი ცხოვრების ნაწილია, თუმცა მას ჯეროვან ყურადღებას არ ვაქცევთ. რა შეხედულება აქვთ ამერიკელებს პოლიტიკურ და სოციალურ საკითხებზე, როგორიცაა სკოლების უსაფრთხოება, გარემო, ეკონომიკა, აბორტები, ჯანდაცვის რეფორმა, ან თუნდაც პრეზიდენტი? — ყოველივე ეს დაკვირვებებითა და გამოკითხვებით იზომება და ყოველდღიურად ქვეყნდება მედიის ახალ ამბებში. აი, უკანასკნელმა სიახლემ ქალთა ჯანმრთელობის შესახებ დილის თოქ-შოუშიც გაიჟღერა. და ის ტანსაცმელიც, რომელიც ახლახანს შეიძინეთ, სხვებისთვის უკვე შეუმჩნეველი აღარ არის. სამომხმარებლო ტრენდების შესწავლა, რომელიც თინეიჯერებსა და მოზარდებზეა ორიენტირებული, ხელს უწყობს კომერციულ პროგრამირებას, პროდუქტის რეკლამირებას და საბოლოო ანგარიშით, სამომხმარებლო დანახარჯების ზრდას.

სტატისტიკა ჩვენს ცხოვრებაში მხოლოდ ახალი ამბებისა და ინფორმაციის სახით არ შემოდის, ის არც მხოლოდ ციფრებია. ამ წიგნით გასწავლით, თუ როგორ უნდა აიმაღლოთ ცნობიერება სტატისტიკის უბრალო მომხმარებელთან შედარებით და გამოიყენოთ ის რეალური სამყაროს შეცნობისათვის. ჩვენთვის, როგორც სოციალური სფეროს მეცნიერთათვის, კარგად არის ცნობილი სტატისტიკის მნიშვნელობა ამერიკულ საზოგადოებასა და მსოფლიოში არსებული განსხვავებების ანალიზისა და გაცნობიერებისათვის. ჩვენ სტატისტიკას ვიყენებთ დემოგრაფიული ტრენდების დადგენისა და საზოგადოების სხვადასხვა ჯგუფებს შორის არსებული განსხვავებების შეფასებისთვის სოციალურ პოლიტიკასა და სოციალურ ძვრებზე გავლენის მოსახდენად. სტატისტიკა გვეხმარება იმ საკითხების სიღრმისეულ შესწავლაში, რაც ჩვენ რეალურ ცხოვრებას განსაზღვრავს.

სწავლებისა და სწავლის მიზნები

სწავლებისა და სწავლის მიზნები ამ წიგნის სახელმძღვანელო პრინციპებად რჩება, რადგან ის უკვე ხუთჯერ გამოიცა.

პირველი მიზანია სოციალური სტატისტიკის გაცნობა და მისი მნიშვნელობის ჩვენება. იცოდეთ, თუ სტატისტიკას თქვენს სტუდენტურ კვლევებში არ გამოიყენებთ, სამუშაოზე მაინც მოგიწევთ მეცნიერთა და პროფესიონალთა პუბლიკაციებიდან ან მას-მედიიდან სტატისტიკური ინფორმაციის წაკითხვა და ინტერპრეტირება. ეს წიგნი ისე გასწავლით კონცეფციების გაგებას სტატისტიკის მიღმა, რომ თქვენ შეძლოთ შეაფასოთ, რომელ სიტუაციაში რა სტატისტიკა უნდა გამოიყენოთ.

მეორე მიზანია სოციალური კვლევის არსისა და სტატისტიკური მეთოდების რეალური კავშირის დემონსტრირება. ეს წიგნი იმითაა განსაკუთრებული, რომ აქ ინტეგრირებულია სტატისტიკური კვლევის არსენალი და სოციალური მეცნიერებების ამოსავალი დებულებები. თქვენი სასწავლო პროცესი სტატისტიკური გამოთვლებითა და ფორმულებით არ შემოიფარგლება. პირიქით, სტატისტიკური მეთოდების გამოყენებისას რეალურ ცხოვრებისეულ მაგალითებზე დაყრდნობით თქვენ უამრავ რამეს ისწავლით სოციალური განსხვავებისა და უთანაბრობის შესახებ. თანამედროვე მსოფლიო დიდი მრავალფეროვნებით ხასიათდება, სადაც სულ უფრო მეტად იგრძნობა რასობრივი, კლასობრივი, გენდერული და სხვა ნიშნების გავლენა, როგორც ცალკეული ადამიანზე, ისე მთლიანად საზოგადოებრივ ცხოვრებაზე. ეს წიგნი ისე გასწავლით საბაზისო სტატისტიკას, რომ ფეხი აუწყოთ სოციალური ცვლადების დინამიურ ურთიერთქმედებას და მათ ასახვას სოციალურ კვლევებში.

ბევრ თქვენგანს ალბათ აკლია მათემატიკაში გაცნობიერებულობა და გარკვეული თვალსაზრისით "მათემატიკის სინდრომით" არის შეპყრობილი. ეს ნერვიულობა ხშირად არახელსაყრელ გარემოს ქმნის სტატისტიკის შესწავლისათვის, რადგან იმის მაგივრად, რომ ძირითად არსს ჩაწვდნენ, სტუდენტები ზედმიწევნით იზეპირებენ სტატისტიკურ მეთოდებს. ამიტომ, ჩვენი მესამე მიზანია, დაივიწყოთ ნერვიულობა და ტექსტის რუტინული დამახსოვრებისა და ფორმულების გამოყვანის ნაცვლად სტატისტიკური კონცეფციების ასახსნელად მოუხმოთ ლოგიკას, ინტუიციასა და საღ აზროვნებას.

წიგნის გამორჩეული და განახლებული მახასიათებლები

ჩვენ მიერ დასახული მიზნები უამრავ კონკრეტულ და სპეციფიკურ პუნქტში აისახება.

მჭიდრო კავშირი პრაქტიკულ სტატისტიკას, მნიშვნელოვან სოციალურ საკითხებსა და რეალურ მაგალითებს შორის. წიგნი იმითაა განსაკუთრებული, რომ აქ

სტატისტიკური მეთოდოლოგია ინტეგრირებულია საზოგადოებისა და სოციალური მეცნიერების ინტერესის სფეროებთან. ხაზგასმულია სოციალური შეხედულებებისა და გამოკითხვის მეთოდების მუდმივი ურთიერთქმედების როლი სოციალური მეცნიერების განვითარებაში. გარდა ამისა, სტუდენტებისთვის ცხადი რომ გავხადოთ სტატისტიკური არგუმენტების მნიშვნელობა, წიგნში მოყვანილია მაგალითები ახალი ამბებიდან, სამთავრობო ანგარიშებიდან, მეცნიერთა ნაშრომებიდან და სხვადასხვა კვლევითი ცენტრების კვლევებიდან. ჩვენ გვსურს ნათლად დავანახოთ სტუდენტებს, თუ რაოდენ მნიშვნელოვანია სოციალური საკითხების შესწავლა რეალურ მონაცემებზე დაყრდნობით, სტატისტიკური დებულებებისა და მეთოდების გამოყენებით.

ფოკუსირება მრავალფეროვნებაზე: აშშ-ისა და მსოფლიოს მასშტაბით. წიგნში გაკეთებულია აქცენტები რასობრივი, კლასობრივი და გენდერული ნიშნების არსებით როლზე სოციალური ტრენდების ჩამოყალიბებაში. სტატისტიკური მეთოდები ფართოდაა გამოყენებული სოციალურ განსხვავებათა მრავალფეროვნების კვლევის პროცესებში, როგორც აშშ-ში, ისე საზღვარგარეთ. ისინი გვეხმარება ჩვენს სოციუმსა და სოციალურ ქცევაზე რასობრივი, კლასობრივი, გენდერული და სხვა ნიშნების გავლენის შეფასებაში.

კვლევითი ლიტერატურის გაცნობა. თქვენი სტუდენტური ცხოვრების განმავლობაში ან სამუშაო ადგილზე მოგიხდებათ სხვების მიერ წარმოდგენილი სტატისტიკური ინფორმაციის ინტერპრეტირება. ასეთი ტიპის სტატისტიკური ანალიზები თავისი ბუნებით ბევრად უფრო კომპლექსურია, ვიდრე სახელმძღვანელოში მოტანილი ან სააუდიტორიო პრეზენტაციები. იმისათვის, რომ გასწავლოთ კვლევითი ლიტერატურის წაკითხვა და გაგება, წიგნში თავების მიხედვით ჩართულია ამონარიდები ექსპერტების კვლევითი ხასიათის პუბლიკაციებიდან სტატისტიკის სადისკუსიო საკითხებზე.

სწავლების ეფექტური ხერხები. თითოეული თავი შეიცავს ძირითად განმარტებებსა და საკვანძო ტერმინებს. ტექსტურ ნაწილში აგრეთვე გამოკვეთილია ძირითადი განმარტებები, ცოდნის შესამოწმებელი კითხვები და დავალებები. ძირითადი განმარტებები აგრეთვე მოცემულია წიგნის ბოლოს დართულ გლოსარიუმში. აქვეა კენტ-ნომრიანი სავარჯიშოებისა და ცოდნის შესამოწმებელი კითხვების პასუხები. ისინი შეიძლება მოიძიოთ სასწავლო საიტზეც www.pineforge.com/ssds6e. ლექტორებისთვის ხელმისაწვდომია სავარჯიშოების ამოხსნის მითითებები.

კომპიუტერული გამოთვლები. წიგნში კომპიუტერული პროგრამა SPSS ყველგან არის გამოყენებული, თუმცა ტექსტის შესასწავლად კომპიუტერი საჭირო არ არის. რეალური მონაცემების გამოყენება აძლიერებს სტუდენტების მოტივაციას სტატისტიკის შესწავლისათვის. ეს მონაცემები მოტანილია ზოგადი სოციალური კვლევიდან (GSS), ჯანდაცვის ინფორმაციის ეროვნული ტრენდების გამოკვლევი-

დან (HINTS) და მომავლის მონიტორინგის კვლევიდან (MTF), რაც ხელმისაწვდომია საიტზე www.pineforge.com/ssds6e. თითოეულ თავის ბოლოს ერთვის მონაცემების დამუშავება SPSS-ის გამოყენებით და სავარჯიშოები SPSS-ისთვის.

მეექვსე გამოცემის ღირსებები

მეექვსე გამოცემაში შევიდა ცვლილებები, რომელთა საფუძვლად გამოვიყენეთ ლექტორებისა (და მათი სტუდენტების) და სხვა დაინტერესებული პირების ფასეული შენიშვნები მეხუთე გამოცემასთან დაკავშირებით.

• სავარჯიშოები და მაგალითები რეალური სამყაროდან: პირველი ხუთი გამოცემის სავიზიტო ბარათი იყო სახელმძღვანელოს თავების სავარჯიშოებსა და მაგალითებში რეალური სამყაროს მონაცემების გამოყენება უამრავი წყაროდან. მეექვსე გამოცემაში ჩვენ განვაახლეთ სავარჯიშოებისა და მაგალითების უმრავლესობა, რომლებიც ეყრდნობა SPSS-ის, MTF-ის, HINTS-ისა და აშშ-ის მოსახლეობის აღწერის მასალებს.

• SPSS-ის ვერსია 18.0: ტექსტთან ერთად, ფაკულტატურ საფუძველზე შეგიძლიათ გამოიყენოთ IBM ®SPSS 18.0-ის სტუდენტური ვერსია. განახლებული სავარჯიშოები და მაგალითები მე-18 ვერსიის ფორმატს ეყრდნობა. საიტზე www.pineforge.com/ssds6e შეგიძლიათ მიიღოთ ინფორმაცია SPSS-ის სტუდენტური ვერსიის პაკეტის შეკვეთასთან დაკავშირებით, ან დაეკონტაქტეთ გამომცემლობას ნომერზე: (805) 499-4224.

• GSS 2008, HINTS 2007 და MTF 2008: SPSS-ის ამოცანებისა და სავარჯიშოებისთვის ჩვენ შევქმენით მონაცემთა 5 ბაზა. SPSS-ის სტუდენტური ვერსიის ფარგლებში თქვენ იმუშავებთ 4 დამოუკიდებელ ფაილთან: GSS Module A, GSS Module B, HINTS 2007 და MTF2008. GSS2008.SAV მოიცავს ცვლადებისა და დაკვირვებების გავრცობილ ნაკრებს GSS 2008-იდან. HINTS2007.SAV მოიცავს 50 ცვლადს 2007 წლის ჯანდაცვის ინფორმაციის ეროვნული ტენდენციების დაკვირვებიდან (HINTS), რომელიც ტარდება კიბოს ეროვნული ცენტრის მიერ. მისი მიზანია ჯანდაცვის სწრაფად განვითარებადი საკომუნიკაციო სფეროების კვლევა როგორც ინგლისურენოვან, ისე ესპანურენოვან მოსახლეობაში. MTF2008.SAV მოიცავს მიჩიგანის უნივერსიტეტის კვლევას, რომელსაც ეწოდება მომავლის მონიტორინგი (MTF). ეს არის მე-12 კლასელთა კვლევა, რომელიც შეიცავს კითხვებს ნარკოტიკების მოხმარებასა და კანონსაწინააღმდეგო ქცევებზე. სავარჯიშოებში SPSS-ისათვის თითოეული თავის ბოლოს გამოყენებულია განსაზღვრული ცვლადები მონაცემთა ყველა მოდულიდან.

• დამატებითი საშუალებები და მნიშვნელოვანი თემები: მეექვსე გამოცემაში დასკვნით სტატისტიკაში კონცენტრაცია ხდება კვლავ Z, t და ხი-კვადრატზე. შერ-

ჩეულ თავებში დამატებულია შეცდომის წყაროები, სადაც გარჩეულია გავრცელებული შეცდომები და შეზღუდვები მონაცემთა შეგროვებასა და ანალიზში.

მადლობა[1]

ჩვენ ორივე მადლიერებას გამოვთქვამთ SAGE Publication-ის რედაქტორის, ჯერი ვესთბის მიმართ, რომელმაც უდიდესი წვლილი შეიტანა წიგნის გამოცემის პროცესში.

ექვსი გამოცემის მანძილზე თავდაუზოგავად იშრომეს კორექტორებმა და გამომცემლობის თანამშრომლებმა, რისთვისაც მათ უღრმეს მადლობას ვუხდით.

ჩვენ უდიდეს მადლიერებას გამოვთქვამთ ოზგარ ავსის მიმართ. მისი ყოვლისმომცველი თანადგომის გარეშე ეს წიგნი ვერ გამოიცემოდა.

ჩავა ფრანკფორტ ნაჩმიასი მადლობას უხდის მეგობრებსა და კოლეგებს მხარში დგომისა და ყოველმხრივი ხელშეწყობისათვის.

მე მადლიერი ვარ იმ სტუდენტებისა, ვინც რთული სტატისტიკური იდეების გამარტივება შთამაგონა. წიგნში წარმოდგენილია მრავალწლიანი სააუდიტორიო ტესტირების შედეგები. მადლობას ვუხდი სტუდენტებს მოთმინებისა და ამ საქმეში წვლილის შეტანისათვის.

და ბოლოს, მადლობას ვუძღვნი ჩემს პარტნიორს, მერლინ შტერნს და აგრეთვე ჩემს გოგონებს, ანეტსა და თალიას, სიყვარულის, მხარდაჭერისა და ნდობისთვის.

ანა ლეონ-გერერო მადლობას უხდის თავის ასისტენტებსა და სტუდენტებს. ჩემი მონდომება, რომ წლიდან-წლამდე უკეთ მესწავლებინა სტატისტიკა, ყოველ სემესტრულად და ყოველ ახალ ჯგუფში სულ უფრო იზრდებოდა. მე მადლიერი ვარ იმ გამოცდილებისათვის, რაც მივიღე სტატისტიკის სწავლებისა და სწავლის სფეროში.

მინდა მადლიერება გამოვთქვა ჩემი მეგობრებისა და კოლეგების მისამართით, ვინც მამხნევებდა და მეხმარებოდა. სიყვარულსა და მადლიერებას ვუძღვნი ჩემს მეუღლეს, ბრაიან სალივანს.

ჩავა ფრანკფორტ ნაჩმიასი
ვისკონსინ-მილუოკის უნივერსიტეტი

ანა ლეონ-გერერო
წყნარი ოკეანის ლუთერანული უნივერსიტეტი

1 იბეჭდება მცირედი შემოკლებით.

თავი 1

სტატისტიკა – რატომ და რისთვის?

თავის სწავლების მიზნები

- კვლევის პროცესის არსი
- დამოუკიდებელი და დამოკიდებული ცვლადების განსაზღვრა და განსხვავება
- გაზომვის სამი დონის განსაზღვრა და განსხვავება
- აღწერითი და დასკვნითი სტატისტიკური პროცედურების განსხვავება

რატომ აირჩიეთ სტატისტიკა? იმიტომ ხომ არა, რომ სასწავლო პროგრამაში ძირითადი საგნების ნუსხაში შეგხვდათ? თვლით, რომ არ არის საინტერესო, არა? თუ ასეა, ცოტა შეცბუნებულიც უნდა ბრძანდებოდეთ, რადგან, ალბათ, გსმენიათ, რომ სტატისტიკა ციფრებთან და მათემატიკასთანაა დაკავშირებული. თანაც, თქვენ შეშფოთებას მხოლოდ მათემატიკასთან უხერხული დამოკიდებულება არ იწვევს. ალბათ, დიდი ეჭვი გაქვთ, რომ ციფრები და მათემატიკა კაბინეტების კედლებს არასოდეს გასცდება, ან რა შეხება უნდა ჰქონდეთ მათ ადამიანებთან და თუნდაც თქვენს ცხოვრებასთან? ზოგადად შეიძლება დავსვათ კითხვა: რა კაეშირი აქვს სტატისტიკას ადამიანის შეცნობასთან, მის სოციალურ ქცევასთან ან საზოგადოებასთან?

ჩვენ ამ წიგნში გიჩვენებთ, რომ სტატისტიკა იმაზე ბევრად უფრო საინტერესო და ადვილად გასაგებია, ვიდრე ეს ოდესმე შეგეძლოთ წარმოგედგინათ. შევეცდებით დავეყრდნოთ თქვენს ცოდნასა და გამოცდილებას და დავაკავშიროთ ისინი საინტერესო და მნიშვნელოვან სოციალურ საკითხებთან. თქვენ თანდათან მიხვდებით, რომ ეს მხოლოდ მორიგი საგანი კი არ არის კურიკულუმში, არამედ სასარგებლო და საჭიროც კია.

ორი მიზეზია იმისათვის, რომ დავასაბუთოთ სტატისტიკის შესწავლის მნიშვნელობა. პირველი, ცხოვრებაში თქვენ ყოველ დღე აწყდებით სტატისტიკას. იქნება

ეს პოპულარულ ჟურნალ-გაზეთებსა და გამოცემებში გამოქვეყნებული მარკეტინგული და სოციალური კვლევის შედეგები, თუ საარჩევნო კამპანიები. სტატისტიკის შესწავლით თქვენ აღნიშნული ინფორმაციის უფრო გამჭრიახი მომხმარებელი გახდებით. მეორე, როგორც სოციალურ მეცნიერებათა სპეციალისტს, თქვენს საქმიანობაში ხშირად მოგიწევთ სტატისტიკური ინფორმაციის გაცნობა და ინტერპრეტაცია. მომავალი საქმიანობის სპეციფიკიდან გამომდინარე, შეიძლება უშუალოდ თქვენ არ მოგიწიოთ კვლევის ჩატარება და მოგიხდეთ სხვების მიერ ჩატარებული კვლევის მასალების გაცნობა და გაგება თუნდაც იმისთვის, რომ დაწეროთ სტატისტიკურ ანალიზზე დაფუძნებული ანგარიში.

მაგრამ, მაინც რა არის სტატისტიკა? შეიძლება ეს სიტყვა დააკავშიროთ რიცხვებთან, როგორიცაა: შობადობა, დანაშაულის რაოდენობა, შემოსავალი ერთ სულ მოსახლეზე, ან თუნდაც ქორწინებისა და განქორწინების დონეები და ა.შ. ამასთანავე სიტყვა **სტატისტიკას** შეხება აქვს მთელ რიგ მეთოდებთან, რომელსაც სოციალური დარგის მეცნიერები იყენებენ. ეს არის ინფორმაციის თავმოყრა, განზოგადება და გადმოცემა. სტატისტიკური ანალიზის საგანი შეიძლება მხოლოდ ციფრობრივი ინფორმაცია გახდეს, რომელსაც **მონაცემები** ეწოდება; მკვლევარები სტატისტიკურ პროცედურებს მონაცემთა ანალიზის მიზნით იყენებენ, რადგან პასუხები გასცენ კვლევის კითხვებს და შეამოწმონ თეორიები. აი, სწორედ ეს გახლავთ ამ სახელმძღვანელოს მიზანიც.

სტატისტიკა სოციალური დარგის მეცნიერთა მიერ გამოყენებული მეთოდების ერთობლიობა ინფორმაციის თავმოყრის, განზოგადებისა და გადმოცემისათვის.

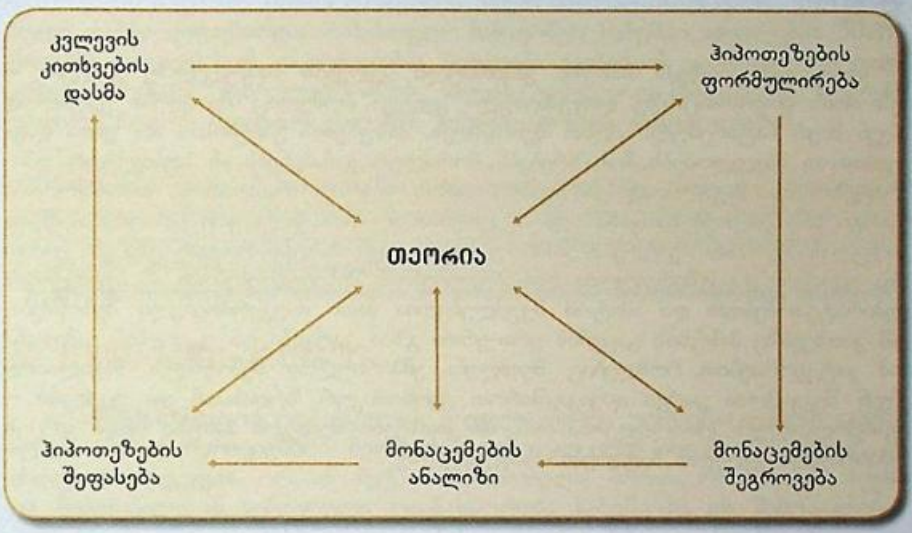
მონაცემები რიცხვებით გამოსახული ინფორმაცია, რომელიც შეიძლება იყოს სტატისტიკური ანალიზის ობიექტი.

კვლევის პროცესი

სოციალურ კვლევებში სტატისტიკის როლის შესახებ უფრო ნათელი წარმოდგენა რომ შეგექმნათ, დავაკვირდეთ **კვლევის პროცესს**. სოციალური სფეროს მეცნიერები ამ პროცესში სხვადასხვა სახის საქმიანობას ეწევიან იმისათვის, რომ პასუხები გასცენ კითხვებს, ან შეამოწმონ იდეებისა და თეორიების ჭეშმარიტება. როგორც ნაჩვენებია ნახაზზე 1.1, კვლევის პროცესი ხუთი საფეხურისგან შედგება:

1. კვლევის კითხვების დასმა
2. ჰიპოთეზების ფორმულირება
3. მონაცემების შეგროვება
4. მონაცემების ანალიზი
5. ჰიპოთეზების შეფასება

ნახაზი 1.1 კვლევის პროცესი



თითოეული საფეხური ზემოქმედებას ახდენს თეორიაზე, ხოლო ისინი თავის მხრივ განიცდიან თეორიის ზეგავლენას. როგორც მომდევნო თავებში დავინახავთ, მონაცემთა სტატისტიკური ანალიზი მკვლევარებს ეხმარება მათი ჰიპოთეზების საიმედოობისა და სიზუსტის შემოწმებაში.

კვლევის პროცესი საქმიანობათა ერთობლიობა, რომლის მიხედვითაც სოციალური სფეროს მეცნიერები პასუხობენ კითხვებს მოსაზრებათა და ჰიპოთეზების შემოწმების მიზნით.

კვლევის კითხვების დასმა

კვლევების უმრავლესობა კითხვების დასმით იწყება. განვიხილოთ სოციოლოგიური ჟურნალებიდან მოტანილი შემდეგი კითხვები:

ახდენს თუ არა რეგულირებადი ჯანდაცვა გავლენას მის ხარისხზე?

უფრო გახშირდა თუ არა სექსუალური ძალადობა უკანასკნელ ათწლეულში?

ახდენს თუ არა სოციალური კლასი გავლენას ამომრჩეველთა ქცევაზე?

გენდერული სეგრეგაციით აიხსნება – ფაქტია, რომ საქმიანობის ტიპისა და სახეობის მიხედვით ამერიკელი ქალებისა და მამაკაცების განაწილება განსხვავებულია. მაგალითად, 2008 წელს მოსახლეობის აღწერის კითხვარში ჩამოთვლილი 14 დარგიდან მხოლოდ ერთში იყო კონცენტრირებული დასაქმებული ქალების ერთი მესამედი (34%), ანუ დაახლოებით 69 მილიონი.[4]

რა არის ნიშანდობლივი სამუშაო ადგილზე გენდერული სეგრეგაციისათვის? საზოგადოებაში, ადამიანების საქმიანობა და დაკავებული თანამდებობა მჭიდროდაა დაკავშირებული მათ პრესტიჟთან, ღირსებასთან და შემოსავლებთან. სამუშაო, სადაც ქალები სეგრეგაციას განიცდიან, არა მარტო განსხვავებულია, არამედ უთანასწოროც. მიუხედავად იმისა, რომ დასაქმებულ ქალთა ხვედრითი წილი მნიშვნელოვნად გაიზარდა, ქალები ჯერ კიდევ კონცენტრირებულნი არიან დაბალპრესტიჟულ, დაბალანაზღაურებად და უპერსპექტივო სამუშაოებზე. ამიტომ, სამუშაო ადგილებზე გენდერული სეგრეგაცია ასოცირდება ანაზღაურების, ღირსებისა და სტატუსის უთანასწორობასთან. ქალთა სეგრეგაცია სამუშაო ადგილებზე ყველაზე ხშირად მამაკაცებთან შედარებით დაბალი ანაზღაურების მიზეზი ხდება. ერთი და იგივე სამუშაოს შესრულებაში იგივე განათლების, უნარ-ჩვევებისა და გამოცდილების პირობებში, ქალები გაცილებით ნაკლებ ანაზღაურებას იღებენ, ვიდრე მამაკაცები.

ჰიპოთეზების ჩამოყალიბება

ამგვარად, ჩვენ განვიხილეთ ქალებისა და მამაკაცების შემოსავლების უთანაბრობასთან დაკავშირებული რამდენიმე საკითხი. აგრეთვე შევეხეთ აღნიშნულის შესაძლო ახსნასაც – თეორიას, რომელიც გვეხმარება სინათლე მოვფინოთ ანაზღაურებაში გენდერული უთანაბრობის საკითხებს. მაგრამ არის კი ეს საკმარისი? ახლა რაღა უნდა ვქნათ?

შემდგომი საფეხურია გენდერული სეგრეგაციის თეორიის მიერ შემოთავაზებული რამდენიმე იდეის ტესტირება. მაგრამ ეს თეორია, რამდენადაც მისაღები და ლოგიკური არ უნდა იყოს ჩვენთვის, ძალიან ზოგადია და არ შეიცავს ტესტირებისათვის საკმარის სპეციფიკურ ინფორმაციას. სამაგიეროდ, თეორია გვთავაზობს სპეციფიკურ კონკრეტულ პროგნოზებს იმის შესახებ, თუ როგორ არის ერთმანეთთან დაკავშირებული ცალკეული ინდივიდებისა თუ ჯგუფების მახასიათებლები. ეს პროგნოზებია, რომელსაც **ჰიპოთეზებს** ვუწოდებთ. ჰიპოთეზები – სავარაუდო პასუხებია დასმულ კითხვებზე. ჰიპოთეზები დროებითია, რადგან მათი დადასტურება მხოლოდ ემპირიული ტესტირების შემდეგ არის შესაძლებელი.[5] მაგალითად, ერთ-ერთი ჰიპოთეზა, რომელიც გენდერული სეგრეგაციის თეორიიდან შეიძლება წამოვაყენოთ, შემდეგში მდგომარეობს: ანაზღაურება იმ საქმიანობებში, სადაც უმრავლესობა ქალია, უფრო დაბალია ანაზღაურებაზე იმ სფეროებში, სადაც უმრავლესობა მამაკაცია.

ჰიპოთეზა სავარაუდო პასუხი საკვლევ საკითხებზე.

ყველა ჰიპოთეზა არ გამომდინარეობს უშუალოდ თეორიებიდან. *ჰიპოთეზები* შეიძლება სხვადასხვა გზით ჩამოვაყალიბოთ. მაგალითად, თეორიებიდან, უშუალოდ კვლევებიდან, ან თუნდაც ინტუიციაზე დაყრდნობით. ჩვენი აზრით, ჰიპოთეზების ჩამოყალიბებისათვის ყველაზე მნიშვნელოვანი წყარო პროფესიული ლიტერატურის გაცნობაა. პროფესიული ლიტერატურის კრიტიკული ანალიზი სწავლისა და დაგროვილი გამოცდილების გაზიარების წინაპირობაა.

დავუბრუნდეთ ჩვენ მიერ ჩამოყალიბებულ ჰიპოთეზას:

ანაზღაურება დასაქმების იმ სფეროებში, სადაც უმრავლესობა ქალია, უფრო დაბალია ანაზღაურებაზე იმ დარგებში, სადაც უმრავლესობა მამაკაცია.

შენიშნავდით, რომ ჰიპოთეზაში მოცემულია ორ მახასიათებელს შორის ურთიერთკავშირი: *ანაზღაურებასა* და *დასაქმების სქესობრივ სტრუქტურას* შორის, რომლებიც ვარიაციას განიცდიან. ამ მახასიათებლებს ცვლადები ეწოდება. *ცვლადი* წარმოადგენს ადამიანების ან საგნების მახასიათებელს, რომელმაც შეიძლება ორი ან მეტი მნიშვნელობა მიიღოს. მაგალითად, ადამიანები შეიძლება დავაჯგუფოთ სოციალური კლასის მიხედვით: მაღალი, საშუალო და მუშათა. ასევე, ადამიანებს აქვთ განსხვავებული განათლება, ამიტომ *განათლების დონე* ცვლადია. *ოჯახის შემოსავალი* აგრეთვე ცვლადია; ის შეიძლება იცვლებოდეს ნულიდან ათასებამდე და ზევითაც. ასევე, ცვლადია *სქესიც*. ქალთა (მამაკაცთა) ხვედრითი წილი დასაქმებაში ნულიდან ასამდე იცვლება. (იხ. ნახ. 1.2, სადაც მოცემულია რამდენიმე ცვლადი და მათი შესაძლო მნიშვნელობები).

ცვლადი ადამიანების ან საგნების მახასიათებელი, რომელმაც შეიძლება ორი ან მეტი მნიშვნელობა მიიღოს.

თითოეული ცვლადი *ამომწურავ* და *სრულიად ექსკლუზიურ* კატეგორიებს უნდა შეიცავდეს. ამომწურავი ნიშნავს, რომ ცვლადს თითოეული დაკვირვების დაჯგუფებისათვის საკმაო რაოდენობის მახასიათებელი უნდა გააჩნდეს. მაგალითად, *ქორწინებაში სტატუსის* მახასიათებლები: „დაქორწინებული“, „დაუქორწინებელი“, „განქორწინებული“ და „ქვრივი“, ეწინააღმდეგება ამომწურავობის მოთხოვნას. ეს არ გვაძლევს საშუალებას, რომ დავაფიქსიროთ ზოგიერთი ერთსქესიანი ან ჰეტეროსექსუალური წყვილი, რომლებიც ოფიციალურ ქორწინებაში არ იმყოფებიან. (ჩვენ შეგვიძლია ნებისმიერი ცვლადი ამომწურავად გადავაქციოთ, თუ მახასიათებლების ჩამონათვალს დავამატებთ კატეგორიას „სხვა“. ამასთან გასათვალისწინებელია, რომ ეს პრაქტიკა მიუღებელია იმ შემთხვევებში, თუ მას თეორიული

რა ფაქტორები ახდენს გავლენას დასაქმებული ქალების ეკონომიკურ მობილობაზე?

ყველა ამ კითხვას პასუხი **ემპირიული კვლევის** საფუძველზე გაეცემა – ეს არის ინფორმაციაზე დაფუძნებული კვლევა, რომელიც შეიძლება უშუალოდ ჩვენ მიერ ჩატარებული ცდით შემოწმდეს. პასუხების გაცემისას არ უნდა დავეყრდნოთ მსჯელობებს, მოსაზრებებს, მორალურ განაჩენებს, ან სუბიექტურ უპირატესობებს. მაგალითად, კითხვებზე: „არის თუ არა რასობრივი თანასწორობა კარგი საზოგადოებისათვის?“ და „უკეთესია თუ არა ქალაქური ცხოვრების წესი სოფლისაზე?“, არ შეიძლება ემპირიულად გაეცეს პასუხი, რადგან ცნებები *კარგი* და *უკეთესი* დაკავშირებულია ღირებულებებთან, შეხედულებებთან ან სუბიექტურ უპირატესობებთან და ამიტომ შეუძლებელია მათი მიუკერძოებელი შემოწმება. ამ კითხვებზე პასუხის გაცემის ერთ-ერთი გზაა *კარგისა* და *უკეთესის* გამიჯვნა იმ კატეგორიებით, რომლებიც შეიძლება ემპირიულად შემოწმდეს. მაგალითად, ჩვენ შეგვიძლია *კარგი* დავაკავშიროთ ეკონომიკურ ზრდასთან და *უკეთესი* – ფსიქოლოგიურ კეთილდღეობასთან. ამ კითხვებს შემდგომ პასუხი უნდა გაეცეს ემპირიული კვლევის ჩატარების საფუძველზე.

ემპირიული კვლევა მონაცემებზე დაფუძნებული კვლევა, რომელიც შეიძლება შემოწმდეს ჩვენი უშუალო ცდით.

ალბათ გაინტერესებთ, თუ როგორ უნდა დასვათ კვლევის კითხვა. უპირველესად ის კითხვა უნდა შეარჩიოთ, რომელიც თქვენ გაინტერესებთ. თუ თავს არ ენდობით, ირგვლივ მიმოიხედეთ! იდეები საკვლევი პრობლემის გარშემო ყველგან მოიპოვება – იქნება ეს მედია-რესურსები თუ პირადი გამოცდილება, ან საკუთარი ინტუიცია. გაესაუბრეთ ადამიანებს, ჩაინიშნეთ თქვენი მოსაზრებები და შენიშვნები, ან ისწავლეთ იმ სოციალური სფეროს მეცნიერთაგან, რომლებსაც უკვე შესწავლილი აქვთ ეს საკითხი.

მაგალითისათვის განვიხილოთ სქესი და სამუშაო. ალბათ, კოლეჯში სწავლისას, სანამ დასაქმებული არ ხარ, გიფიქრია ქალებისა და მამაკაცების სამუშაო გამოცდილების მსგავსება-განსხვავებაზე და შენ სამომავლო შესაძლებლობებზე. მოვიყვანთ ფაქტებს, რომლებიც სხვადასხვა კვლევებსა და პირად გამოცდილებას ეყრდნობა: 2008 წელს სრულ შტატზე მომუშავე ქალების საშუალო კვირეული ანაზღაურება $682 იყო, ხოლო მამაკაცებისა – $876.[1] მამაკაცებისა და ქალების სამუშაო აგრეთვე ძალიან განსხვავებულია. ქალები ისევ უმცირესობაში არიან მაღალრეიტინგულ და მაღალანაზღაურებად პროფესიულ და მენეჯერულ პოზიციებზე. 2008 წელს ქალები შეადგენდნენ ინჟინრების 10.4%-ს, ექიმების 30.5%-ს, სტომატოლოგების 27.2%-ს და არქიტექტორების 24.7%-ს. ამის საპირისპიროდ, სკოლამდელ და საბავშვო ბაღის მასწავლებლებს შორის 97.6% ქალები

იყვნენ. ასევე, 2008 წელს მდივან-რეფერენტთა 96.1%-ს ქალები შეადგენდნენ.[2] 2010 წლის იანვარში აშშ-ის დასაქმების ისტორიაში აღსანიშნავი და განსხვავებული ფაქტი დაფიქსირდა: არასასოფლო-სამეურნეო სამუშაო ადგილების 50.3% ქალებმა დაიკავეს[3] ამ დაკვირვებებმა შეიძლება გვიბიძგოს კითხვისკენ: ერთიდაიმავე სამუშაოს შესრულებისათვის ქალებს საშუალოდ მამაკაცებზე ნაკლებს უხდიან? როგორ შეიცვალა ქალების სამუშაო დროთა განმავლობაში?

ჩამოაყალიბე ერთი ან ორი კითხვა სოციალურ სფეროში, რომელიც ემპირიულ კვლევას მოითხოვს. უმაღლესში სწავლისას, თქვენ აუცილებლად მოგთხოვენ კვლევითი პროექტის განხორციელებას. დაიჭირე თადარიგი და დაიწყე ფიქრი იმ კარგ კითხვაზე, რომელსაც შენ დასვამ მომავალში.

✓ *ცოდნ შეძოწმებ*

თეორიის როლი

ალბათ შეამჩნევდით, რომ კვლევის ყოველი კითხვა *ურთიერთკავშირს* გამოხატავს. ეს ურთიერთკავშირი შეიძლება იყოს ცალკეულ ინდივიდებს შორის, ან ინდივიდთა ჯგუფის ორ ან მეტ მახასიათებელს შორის, როგორიცაა სქესი და შემოსავალი, ან გენდერული უთანასწორობა სამუშაოზე და შემოსავლების უთანაბრობა. ინდივიდებისა და ჯგუფების ცალკეულ მახასიათებლებსა ან თვისებებს შორის ურთიერთკავშირი სოციალური მეცნიერების თეორიის ქვაკუთხედია.

ხშირად ტერმინ თეორიას ცხოვრებისეული მოვლენებისა და ცდების ასახსნელად ვიყენებთ. შეიძლება ჩვენი „თეორია“ გვქონდეს იმის შესახებ, თუ რატომ იყო ადრე ბოსი ჩვენდამი კეთილგანწყობილი, ან რატომ ვერ დავწერეთ კარგად ისტორიის ტესტი. სოციალური სფეროს მეცნიერებიც ამის მსგავსად ცდილობენ სოციალური რეალობის ახსნას. თუ სხვადასხვა მოვლენის შესახებ ჩვენი თეორიები ჩვენივე ცოდნისა და გამოცდილების შედეგია, სოციალური სფეროს მეცნიერთათვის თეორია მოვლენების უფრო ზუსტ შეფასებას წარმოადგენს, რაც ხშირად კვლევებით მოწმდება.

თეორია ეს არის ორი ან მეტი ინდივიდის, ან ჯგუფის მახასიათებლების ურთიერთკავშირის ახსნა. თეორია ცდილობს დაგვანახოს კავშირი მონაცემებსა და ჩვენ მიერ აღქმულ რეალობას შორის. მაგალითად, გვინდა ავხსნათ განსხვავების მიზეზები ქალისა და მამაკაცის ანაზღაურებაში; ვინტერესდებით: ერთი ტიპის სამუშაოებს ასრულებენ თუ არა ქალები და მამაკაცები? და ვფიქრობთ, რომ ორგანიზაციამ, სადაც ისინი მუშაობენ, რაღაც უნდა მოუხერხოს მათ ხელფასებს.

თეორია ცალკეული ინდივიდების, ან ჯგუფების ორი ან მეტი მახასიათებლის ურთიერთკავშირის მეცნიერული ახსნა.

გენდერული უთანასწორობა ანაზღაურებაში *სამუშაო ადგილზე არსებული*

გენდერული სეგრეგაციით აიხსნება – ფაქტია, რომ საქმიანობის ტიპისა და სახეობის მიხედვით ამერიკელი ქალებისა და მამაკაცების განაწილება განსხვავებულია. მაგალითად, 2008 წელს მოსახლეობის აღწერის კითხვარში ჩამოთვლილი 14 დარგიდან მხოლოდ ერთში იყო კონცენტრირებული დასაქმებული ქალების ერთი მესამედი (34%), ანუ დაახლოებით 69 მილიონი.[4]

რა არის ნიშანდობლივი სამუშაო ადგილზე გენდერული სეგრეგაციისათვის? საზოგადოებაში, ადამიანების საქმიანობა და დაკავებული თანამდებობა მჭიდროდაა დაკავშირებული მათ პრესტიჟთან, ღირსებასთან და შემოსავლებთან. სამუშაო, სადაც ქალები სეგრეგაციას განიცდიან, არა მარტო განსხვავებულია, არამედ უთანასწოროც. მიუხედავად იმისა, რომ დასაქმებულ ქალთა ხვედრითი წილი მნიშვნელოვნად გაიზარდა, ქალები ჯერ კიდევ კონცენტრირებულნი არიან დაბალპრესტიჟულ, დაბალანაზღაურებად და უპერსპექტივო სამუშაოებზე. ამიტომ, სამუშაო ადგილებზე გენდერული სეგრეგაცია ასოცირდება ანაზღაურების, ღირსებისა და სტატუსის უთანასწორობასთან. ქალთა სეგრეგაცია სამუშაო ადგილებზე ყველაზე ხშირად მამაკაცებთან შედარებით დაბალი ანაზღაურების მიზეზი ხდება. ერთი და იგივე სამუშაოს შესრულებაში იგივე განათლების, უნარ-ჩვევებისა და გამოცდილების პირობებში, ქალები გაცილებით ნაკლებ ანაზღაურებას იღებენ, ვიდრე მამაკაცები.

ჰიპოთეზების ჩამოყალიბება

ამგვარად, ჩვენ განვიხილეთ ქალებისა და მამაკაცების შემოსავლების უთანაბრობასთან დაკავშირებული რამდენიმე საკითხი. აგრეთვე შევეხეთ აღნიშნულის შესაძლო ახსნასაც – თეორიას, რომელიც გვეხმარება სინათლე მოვფინოთ ანაზღაურებაში გენდერული უთანაბრობის საკითხებს. მაგრამ არის კი ეს საკმარისი? ახლა რაღა უნდა ვქნათ?

შემდგომი საფეხურია გენდერული სეგრეგაციის თეორიის მიერ შემოთავაზებული რამდენიმე იდეის ტესტირება. მაგრამ ეს თეორია, რამდენადაც მისაღები და ლოგიკური არ უნდა იყოს ჩვენთვის, ძალიან ზოგადია და არ შეიცავს ტესტირებისათვის საკმარის სპეციფიკურ ინფორმაციას. სამაგიეროდ, თეორია გვთავაზობს სპეციფიკურ კონკრეტულ პროგნოზებს იმის შესახებ, თუ როგორ არის ერთმანეთთან დაკავშირებული ცალკეული ინდივიდებისა თუ ჯგუფების მახასიათებლები. ეს პროგნოზებია, რომელსაც **ჰიპოთეზებს** ვუწოდებთ. ჰიპოთეზები – სავარაუდო პასუხებია დასმულ კითხვებზე. ჰიპოთეზები დროებითია, რადგან მათი დადასტურება მხოლოდ ემპირიული ტესტირების შემდეგ არის შესაძლებელი.[5] მაგალითად, ერთ-ერთი ჰიპოთეზა, რომელიც გენდერული სეგრეგაციის თეორიიდან შეიძლება წამოვაყენოთ, შემდეგში მდგომარეობს: ანაზღაურება იმ საქმიანობებში, სადაც უმრავლესობა ქალია, უფრო დაბალია ანაზღაურებაზე იმ სფეროებში, სადაც უმრავლესობა მამაკაცია.

ჰიპოთეზა სავარაუდო პასუხი საკვლევ საკითხებზე.

ყველა ჰიპოთეზა არ გამომდინარეობს უშუალოდ თეორიებიდან. *ჰიპოთეზები* შეიძლება სხვადასხვა გზით ჩამოვაყალიბოთ. მაგალითად, თეორიებიდან, უშუალოდ კვლევებიდან, ან თუნდაც ინტუიციაზე დაყრდნობით. ჩვენი აზრით, ჰიპოთეზების ჩამოყალიბებისათვის ყველაზე მნიშვნელოვანი წყარო პროფესიული ლიტერატურის გაცნობაა. პროფესიული ლიტერატურის კრიტიკული ანალიზი სწავლისა და დაგროვილი გამოცდილების გაზიარების წინაპირობაა.

დავუბრუნდეთ ჩვენ მიერ ჩამოყალიბებულ ჰიპოთეზას:

ანაზღაურება დასაქმების იმ სფეროებში, სადაც უმრავლესობა ქალია, უფრო დაბალია ანაზღაურებაზე იმ დარგებში, სადაც უმრავლესობა მამაკაცია.

შენიშნავდით, რომ ჰიპოთეზაში მოცემულია ორ მახასიათებელს შორის ურთიერთკავშირი: *ანაზღაურებასა და დასაქმების სქესობრივ სტრუქტურას* შორის, რომლებიც ვარიაციას განიცდიან. ამ მახასიათებლებს ცვლადები ეწოდება. **ცვლადი** წარმოადგენს ადამიანების ან საგნების მახასიათებელს, რომელმაც შეიძლება ორი ან მეტი მნიშვნელობა მიიღოს. მაგალითად, ადამიანები შეიძლება დავაჯგუფოთ სოციალური კლასის მიხედვით: მაღალი, საშუალო და მუშათა. ასევე, ადამიანებს აქვთ განსხვავებული განათლება, ამიტომ *განათლების დონე* ცვლადია. *ოჯახის შემოსავალი* აგრეთვე ცვლადია; ის შეიძლება იცვლებოდეს ნულიდან ათასებამდე და ზევითაც. ასევე, ცვლადია *სქესიც*. ქალთა (მამაკაცთა) ხვედრითი წილი დასაქმებაში ნულიდან ასამდე იცვლება. (იხ. ნახ. 1.2, სადაც მოცემულია რამდენიმე ცვლადი და მათი შესაძლო მნიშვნელობები).

ცვლადი ადამიანების ან საგნების მახასიათებელი, რომელმაც შეიძლება ორი ან მეტი მნიშვნელობა მიიღოს.

თითოეული ცვლადი *ამომწურავ* და *სრულიად ექსკლუზიურ* კატეგორიებს უნდა შეიცავდეს. ამომწურავი ნიშნავს, რომ ცვლადს თითოეული დაკვირვების დაჯგუფებისათვის საკმაო რაოდენობის მახასიათებელი უნდა გააჩნდეს. მაგალითად, *ქორწინებაში სტატუსის* მახასიათებლები: „დაქორწინებული“, „დაუქორწინებელი“, „განქორწინებული“ და „ქვრივი“, ეწინააღმდეგება ამომწურავობის მოთხოვნას. ეს არ გვაძლევს საშუალებას, რომ დავაფიქსიროთ ზოგიერთი ერთსქესიანი ან ჰეტეროსექსუალური წყვილი, რომლებიც ოფიციალურ ქორწინებაში არ იმყოფებიან. (ჩვენ შეგვიძლია ნებისმიერი ცვლადი ამომწურავად გადავაქციოთ, თუ მახასიათებლების ჩამონათვალს დავამატებთ კატეგორიას „სხვა“. ამასთან გასათვალისწინებელია, რომ ეს პრაქტიკა მიუღებელია იმ შემთხვევებში, თუ მას თეორიული

მნიშვნელობის მქონე კატეგორიების ან დაკვირვებების არსებითი რაოდენობის გაუთვალისწინებლობამდე მივყავართ).

სრულიად ექსკლუზიური ნიშნავს, რომ თითოეული დაკვირვებისათვის მხოლოდ ერთი მახასიათებელია მისაღები. მაგალითად, *რელიგიის* კლასიფიცირება ისე უნდა მოვახერხოთ, რომ არავინ მოხვდეს ერთზე მეტ კატეგორიაში. კერძოდ,

ნახაზი 1.2 ცვლადები და სიდიდის კატეგორიები

ცვლადები	*კატეგორიები*
სოციალური კლასი	მაღალი საშუალო მუშათა
რელიგია	ქრისტიანი ებრაელი მუსლიმანი
თვიური შემოსავალი	$1,000 $2,500 $10,000 $15,000
სქესი	მამრობითი მდედრობითი

კატეგორიები: „პროტესტანტი“ და „მეთოდისტი“, არსებითად ექსკლუზიური არ არის, რადგან მეთოდისტებს შორისაც არიან პროტესტანტები, ამიტომ ისინი ორივე კატეგორიაში მოხვდებიან.

✓ *ცოდნის შემოწმება*

გადახედეთ „ამომწურავი“ და „სრულიად ექსკლუზიური“ ცნებების განმარტებებს. დააკვირდით ნახ. 1.2-ს. რომელი სხვა კატეგორიები შეიძლება დაემატოს რელიგიის ცვლადს, რომ ის ამომწურავი და სრულიად ექსკლუზიური გახდეს? რომელი კატეგორიები შეიძლება დაემატოს სოციალურ კლასს? შემოსავალს?

სოციალური სფეროს მეცნიერები თვითონ წყვეტენ სოციალური ცხოვრების რომელ დონეზე აპირებენ კვლევის ჩატარებას. მათი ყურადღება შეიძლება შეჩერდეს ცალკეულ ინდივიდებზე ან ადამიანთა ჯგუფებზე, როგორიცაა ოჯახი, ორგანიზაციები ან ეროვნებები. ეს განსხვავება შესასწავლ **საანალიზო ერთეულზე** დამოკიდებული. ცვლადი ნებისმიერი ერთეულის მახასიათებელია, რომლის შესწავლასაც ვაპირებთ. ცვლადები ახასიათებენ ცალკეულ ინდივიდებს, ჯგუფებს

(მაგ. ოჯახი ან სოციალური კლასი), ორგანიზაციებს (მაგ. საავადმყოფო ან უნივერსიტეტი), ან საზოგადოებებს (მაგ. ქვეყანა ან ერი). მაგალითად, თუ ვსწავლობთ ინდივიდის განათლების დონესა და მის ანაზღაურებას შორის ურთიერთკავშირს, ცვლადი *შემოსავალი* ახასიათებს ცალკეული ინდივიდის შემოსავალს. მეორე მხრივ, თუ ვსწავლობთ კორპორაციის მიერ მიღებული ამონაგების ურთიერთკავშირს მასში დასაქმებულთა დამატებით შეღავათებთან, აქ ცვლადი *ამონაგები* ორგანიზაციის (კორპორაციის) მახასიათებელია. ცვლადები – *ანაზღაურება* და *სქესობრივი სტრუქტურა* მოცემულ მაგალითში დასაქმების მახასიათებლებია. ნახაზზე 1.3 ასახულია სოციალური სფეროში ხშირად გამოყენებული საანალიზო ერთეულები.

საანალიზო ერთეული სოციალური დარგის მეცნიერთა ინტერესის სფეროში მოქცეული სოციალური ცხოვრების დონეები. ერთობლიობის ერთეულის განსხვავებული დონეებია ცალკეული ინდივიდები და ჯგუფები.

გახსოვს ის კითხვა, რომლის დასმაც გსურდა? შეგიძლია თუ არა ჩამოაყალიბო ჰიპოთეზა, სანამ მის ტესტირებას მოახდენ? დაიმახსოვრე, რომ ცვლადმა შეიძლება მიიღოს ორი ან მეტი მნიშვნელობა და აუცილებელია ერთობლიობის ერთეულის განსაზღვრა.

✓ ცოდნის შემოწმება

დამოუკიდებელი და დამოკიდებული ცვლადები: მიზეზობრიობა

ჩვეულებრივ, ჰიპოთეზების ჩამოყალიბება *დამოუკიდებელი* და *დამოკიდებული* ცვლადების ურთიერთკავშირის საფუძველზე ხდება. კვლევაში ძალიან მნიშვნელოვანია დამოუკიდებელი და დამოკიდებული ცვლადების გარჩევა. სოციალურ თეორიებში ხშირად კეთდება მინიშნება ცვლადების მიზეზობრივი კავშირების განსაზღვრის შესახებ. მაგალითად, გენდერული სეგრეგაციის თეორიის თანახმად, მამაკაცებისა და ქალების განსხვავებული ანაზღაურების მთავარი მიზეზი (რა თქმა უნდა, არა ერთადერთი) სამუშაო ადგილზე არსებული გენდერული სეგრეგაციაა. რატომ უხდიდან ქალებს მამაკაცებზე უფრო ნაკლებს იმ სამუშაოს შესრულებაში, სადაც ისინი უმრავლესობაში არიან? ერთ-ერთი ახსნა ისაა, რომ

> შესასრულებელი სამუშაოს მიუხედავად, საზოგადოებები სათანადოდ არ აფასებენ ქალის შრომას, რადგან მას ქალები ასრულებენ... მაგალითად, ჩვენი კულტურული შეხედულებებიდან გამომდინარე, აღმზრდელების შრომის სათანადო დაფასება იმის გამო არ ხდება, რომ ამას ქალები აკეთებენ. ამიტომ, აღმზრდელების (ძიძების) საქმიანობა დაბალანაზღაურებადია.[6]

ნახაზი 1.3 კვლევის პროცესი

ინდივიდი, როგორც საანალიზო ერთეული

რამდენი წლის ხართ?
რომელ პოლიტიკურ პარტიას ემხრობით?
რას საქმიანობთ?



ოჯახი, როგორც საანალიზო ერთეული

რამდენი ბავშვია ოჯახში?
ვინ ასრულებს სახლში სამუშაოს?
რამდენს აქვს ხელფასი?



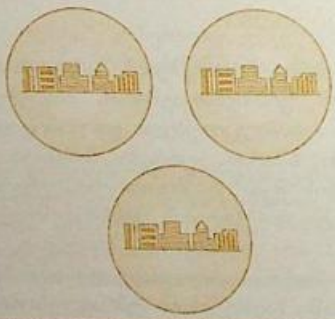
ორგანიზაცია, როგორც საანალიზო ერთეული

რამდენი დასაქმებულია?
როგორია გენდერული თანაფარდობა?
არის თუ არა თქვენი ოფისი მრავალფეროვანი?



ქალაქი, როგორც საანალიზო ერთეული

როგორი იყო დამნაშავეობა გასულ წელს?
როგორია მოსახლეობის სიმჭიდროვე?
რა ტიპის მთავრობა იღებს გადაწყვეტილებებს?



ცვლადს, რომლის ახსნასაც ცდილობს მკვლევარი, კვლევის ენაზე ეწოდება **დამოკიდებული ცვლადი**, ხოლო იმ ცვლადს, რომელიც იწვევს ამ დამოკიდებული ცვლადის ცვლილებას, *დამოუკიდებელი* ეწოდება. ამიტომ, ჩვენ მაგალითში *დასაქმების გენდერული სტრუქტურა* დამოუკიდებელი ცვლადია, ხოლო ანაზღაურება — დამოკიდებული.

დამოკიდებული ცვლადი ცვლადი, რომელიც უნდა ავხსნათ („შედეგი").

დამოუკიდებელი ცვლადი ცვლადი, რომელიც ახდენს ზეგავლენას (გამომწვევი „მიზეზი") დამოკიდებულ ცვლადზე.

სოციალურ მეცნიერებებში მიზეზ-შედეგობრივი დამოკიდებულებების შესახებ დასკვნების გაკეთება ადვილი არ არის. იმისათვის, რომ დადგინდეს მიზეზ-შედეგობრივი ურთიერთკავშირი ორ ცვლადს შორის, უნდა დაკმაყოფილდეს სამი პირობა: (1) მიზეზმა შედეგი უნდა გამოიწვიოს დროში, (2) მიზეზსა და შედეგს შორის უნდა იყოს ემპირიული ურთიერთკავშირი, და (3) ეს ურთიერთობა არ უნდა აიხსნას სხვა ფაქტორებით.

გავიხსენოთ წლების წინ დაწყებული დისკუსია ჩადენილ დანაშაულზე დაწესებული სასჯელის საწინააღმდეგოდ პრევენციული ღონისძიებების გატარების შესახებ. ზოგიერთი ფიქრობს, რომ დანაშაულის პირველი შემთხვევისას ჭკუის დარიგება და ახალგაზრდაზე განხორციელებული მკაცრი კონტროლი დამნაშავეობას შეამცირებს. ხოლო დანარჩენებს გამოსავლად პატიმრობის შეფარდება მიაჩნიათ. 90-იანი წლების დასაწყისში ვაშინგტონსა და კალიფორნიაში მიიღეს კანონი გრძელვადიანი პატიმრობის შეფარდებაზე მათთვის, ვინც სამჯერ იქნებოდა შემჩნეული სისხლის სამართლის დანაშაულში. 2004 წლისათვის დანაშაულის ჩადენაზე „ხისტი" პოლიტიკის გატარების მიზნით ანალოგიური კანონმდებლობა დანარჩენ 24 შტატშიც გავრცელდა. წარმოვიდგინოთ, რომ წლების შემდეგ მართლაც შემცირდა დამნაშავეობა ამ შტატებში; აღნიშნული ზომის გატარების ინიციატორებმა საკუთარ წარმატებად აღიქვეს დანაშაულის შემცირება. ნიშნავს თუ არა ეს, რომ სწორედ ამ ღონისძიებებმა გამოიწვია დანაშაულის შემცირება? სულაც არ არის აუცილებელი! იქნებ დანაშაულის შემცირება ეკონომიკის გაჯანსაღებამ გამოიწვია და არა ახალმა კანონმდებლობამ? იმისათვის, რომ მოვახდინოთ მიზეზ-შედეგობრივი კავშირის დემონსტრირება, უნდა ვაჩვენოთ სამი რამ: (1) კანონის შემოღება ემპირიულად დაკავშირებული იყო დამნაშავეობის შემცირებასთან, (2) დანაშაულის შემცირება ნამდვილად მოჰყვა ახალი კანონმდებლობის შემოღებას, და (3) დანაშაულის შემცირებისა და ახალი კანონმდებლობის ურთიერთკავშირზე გავლენა არ მოუხდენია სხვა ცვლადს (მაგ. ეკონომიკური პირობების გაუმჯობესებას).

დამოუკიდებელი და დამოკიდებული ცვლადები

რადგან სოციალურ მეცნიერებებში გარკეული შეზღუდვებია დაწესებული მიზეზ-შედეგობრივ ურთიერთკავშირებზე, ამიტომ სიფრთხილეა საჭირო ცვლადებს შორის ტერმინების *მიზეზისა* და *შედეგის* გამოყენებისას. ამასთან, უფრო მიზანშეწონილია *დამოუკიდებელი* და *დამოკიდებული* ცვლადების ტერმინების გამოყენება, სანამ რეალურად დადგინდება მიზეზ-შედეგობრივი დამოკიდებულება. გთავაზობთ რამდენიმე მითითებას დამოუკიდებელი და დამოკიდებული ცვლადების იდენტიფიკაციასთან დაკავშირებით:

1. დამოკიდებული ცვლადი ყოველთვის ის მახასიათებელია, რომელსაც ახსნა სჭირდება: ის ყოველთვის კვლევის ობიექტია.
2. დამოუკიდებელი ცვლადი დროში უფრო ადრე წარმოიქმნება, ვიდრე დამოკიდებული.
3. დამოუკიდებელი ცვლადი ზეგავლენას ახდენს დამოკიდებულზე პირდაპირი თუ არაპირდაპირი გზით.

კვლევის მიზანი უნდა დაგვეხმაროს იმის გარკვევაში, რომელია დამოუკიდებელი და რომელი — დამოკიდებული ცვლადი. რეალურ ცხოვრებაში, ცვლადები არც დამოკიდებულია და არც დამოუკიდებელი. ამ თვისებას ისინი იძენენ კონკრეტულ პრობლემასთან მიმართებაში. შეიძლება თუ არა, რომ ერთი კვლევის დამოკიდებული ცვლადი მეორე კვლევაში დამოუკიდებელი გახდეს?[7] მაგალითად, *განათლების დონე* შეიძლება დამოუკიდებელი ცვლადი იყოს კვლევაში, რომელიც ცდილობს დაადგინოს განათლების როლი პოლიტიკურ შეხედულებებზე. თუმცა კვლევაში, სადაც სწავლობენ სოციალური სტატუსის გავლენას პიროვნების განათლებაზე, *განათლების დონე* დამოკიდებული ცვლადია. ისეთი ცვლადები, როგორიცაა: რასა, ასაკი და ეთნიკური წარმოშობა თავიდანვე მოცემულობაა, რაც არ აიხსნება სოციალური სფეროს მეცნიერების მიერ. ამიტომ ისინი დამოკიდებული ცვლადები ვერ იქნება სოციალურ ანალიზში.

✓ *ცოდნის შემოწმება*

განსაზღვრეთ დამოუკიდებელი და დამოკიდებული ცვლადები შემდეგ ჰიპოთეზებში:

- *ჯანდაცვის რეფორმებს ახალგაზრდა ამერიკელები უფრო მეტად უჭერენ მხარს, ვიდრე ხანდაზმულები.*
- *ადამიანები, ვინც ხშირად დადიან ეკლესიაში, უფრო მეტად ეწინააღმდეგებიან აბორტს, ვიდრე ისინი, ვინც არ დადიან.*
- *ხანდაზმული ქალების უფრო დიდი ნაწილი ცხოვრობს მარტო, ვიდრე კაცების.*
- *უმაღლესდამთავრებულებს უფრო ცოტა შვილი ჰყავთ, ვიდრე ნაკლებად განათლებულებს.*

რომელია დამოუკიდებელი და დამოკიდებული ცვლადები თქვენს ჰიპოთეზებში?

მონაცემების შეგროვება

რადგან სასწავლო პროცესში ჩავრთეთ კვლევის კითხვები, ჰიპოთეზები და ცვლადები, დროა უკვე გადავიდეთ კვლევის ციკლის მეორე საფეხურზე. ეს საფეხური მოიცავს ჩვენი ცვლადების გაზომვას და მონაცემების შეგროვებას. როგორც მკვლევარებმა, თვითონ უნდა გადავწყვიტოთ, თუ როგორ გავზომოთ ჩვენთვის საინტერესო ცვლადები, როგორ შევარჩიოთ მოვლენები კვლევისათვის და მონაცემთა შეგროვების რომელი ხერხები გამოვიყენოთ. ჩვენთვის მონაცემთა შეგროვების მრავალი ხერხია ხელმისაწვდომი, უშუალო დაკვირვებიდან დაწყე-

ბული, კვლევის, ექსპერიმენტისა და მეორადი წყაროების ჩათვლით. ასევე ჩვენ შეგვიძლია გამოვიყენოთ გაზომვის უამრავი ხერხი. ეს ინსტრუმენტები შეიძლება ისევე მარტივი იყოს, როგორც მხოლოდ ერთკითხვიანი კითხვარი, ან ისეთი კომპლექსური, როგორიც ორი ან მეტი კითხვარის საფუძველზე მიღებული კრებსითი საზომი. მონაცემთა შეგროვების მეთოდისა და ხერხის შერჩევა ცვლადების გაზომვისათვის შესასწავლ ობიექტზეა დამოკიდებული.

ვთქვათ, გადავწყვიტეთ, შევისწავლოთ სოციალური ჯგუფების დამოკიდებულება აბორტების მიმართ. ეს პრობლემა უშუალო დაკვირვებით არ გადაწყდება. აღნიშნულთან დაკავშირებით მათი დამოკიდებულებისა და შეხედულებების გასარკვევად ადამიანთა სხვადასხვა ჯგუფები უნდა გამოვკითხოთ. ამ შემთხვევაში მონაცემთა შეგროვებისათვის ყველაზე მოსახერხებელი იქნება *გამოკითხვის* ჩატარება კითხვარების ან ინტერვიუების საშუალებით, რათა მივიღოთ ვერბალური პასუხები რესპონდენტებისაგან. კითხვარი უნდა მოიცავდეს შეკითხვების ჩამონათვალს, რაც საშუალებას მოგვცემს შევაფასოთ აბორტთან დაკავშირებით სოციალური ჯგუფების შეხედულებები და ასევე სხვა ცვლადებიც, რომლებიც ამ საკითხის კვლევას სჭირდება.

როგორ უნდა შევაგროვოთ მონაცემები, რომ შევაფასოთ დასაქმებულთა ანაზღაურებასა და სქესობრივ სტრუქტურას შორის დამოკიდებულების ჰიპოთეზა? გვჭირდება მონაცემების შეგროვება სხვადასხვა პროფესიის დასაქმებულებში ქალებისა და მამაკაცების პროპორციებისა და საშუალო ხელფასების შესახებ. როგორც წესი, ამ ტიპის ინფორმაციის მოზიდვა მთავრობის მიერ ხდება და ქვეყნდება აშშ-ის შრომის ბიუროს ბიულეტენებსა და ქვეყნის შემაჯამებელ სტატისტიკურ პუბლიკაციებში (*Statistical Abstract of United States*). სწორედ ამ წყაროებიდან მიღებული მონაცემები შეიძლება გამოვიყენოთ ანალიზსა და ჰიპოთეზების ტესტირებაში.

გაზომვის დონეები

მონაცემთა სტატისტიკური ანალიზი მრავალ მათემატიკურ ოპერაციას მოიცავს, მარტივი დათვლიდან დაწყებული, შეკრებისა და გამრავლების ჩათვლით. თუმცა ყველა ოპერაციის გამოყენება არ შეიძლება ყველა ცვლადთან. სტატისტიკური ოპერაციის ტიპი დამოკიდებულია გასაზომ ცვლადზე. მაგალითად, ცვლადი *სქესისათვის*, შეგვიძლია გამოვიყენოთ ციფრი 1 — ქალების აღსანიშნავად და ციფრი 2 — მამაკაცებისათვის. ასევე შეგვიძლია რაოდენობრივი კოდი 1-ის გამოყენება კატეგორიისათვის „ერთი ბავშვი“ ცვლად *ბავშვების რაოდენობაში*. გასაგებია, რომ პირველ შემთხვევაში ციფრი 1-ის გამოყენება ქალების აღსანიშნავად სიმბოლურ ხასიათს ატარებს, თუმცა მეორე შემთხვევაში, ის წარმოადგენს თვისების „ერთი ბავშვი“ პირდაპირ რაოდენობრივ გამოხატულებას. ჩვენ მიერ გამოყენებული სტატისტიკური ოპერაციების ტიპს განსაზღვრავს შესაბამისობა გასაზომ თვისებასა და ამ თვისების ამსახველ ციფრებს შორის. შესაბამისობის ხარისხს მივყავართ გაზომვის სხვადასხვა გზებთან, ანუ მკაფიოდ ჩამოყალიბებულ გაზომვის დონეებთან. სახელმძღვანელოს ამ ნაწილში, ვიმსჯელებთ გაზომვის *სამ დონეზე: ნომინალურ, ორდინალურ და ინტერვალურ-ფარდობითზე.*

გაზომვის ნომინალური დონე

გაზომვის ნომინალურ დონეზე ციფრები და სიმბოლოები გამოიყენება სახელდების, აღნიშვნის ან დაკვირვებების დაჯგუფების მიზნით. *სქესი* არის ნომინალური დონის ცვლადის მაგალითი. ციფრების 1 და 2 გამოყენებით ჩვენ შეგვიძლია დაკვირვებების დაჯგუფება კატეგორიებში „ქალი“ და „მამაკაცი“, სადაც 1-იანით აღნიშნულია ქალი, და 2-იანით კაცი. ნომინალური ცვლადის სხვადასხვა კატეგორიისათვის ჩვენ შეგვიძლია უამრავი სიმბოლოს გამოყენება, თუმცა ამ შემთხვევაში, განსხვავებული ციფრები სხვაობას არ გამოხატავენ არც ამ კატეგორიების სიდიდეში და არც რაოდენობაში. განსხვავებული კატეგორიები (მაგ. ქალები და მამაკაცები) ერთმანეთის მიმართ ცვალებადნი არიან *თვისებრივად* და არა რაოდენობრივად. ნომინალურ ცვლადებს ხშირად თვისებრივსაც უწოდებენ. ნომინალური ცვლადების სხვა მაგალითებია პოლიტიკური პარტიები, რელიგია და რასა.

ნომინალური გაზომვა სხვადასხვა კატეგორიისათვის ციფრებისა და სიმბოლოების მინიჭება სახელდების, აღნიშვნის ან დაკვირვებების დაჯგუფების მიზნით.

გაზომვის ორდინალური დონე

თუ ჩვენ კატეგორიებს მივანიჭებთ ციფრებს რანგული თანმიმდევრობით, დაბლიდან მაღლისკენ, მივიღებთ გაზომვის ორდინალურ დონეს. სოციალური ჯგუფი წარმოადგენს ორდინალურ ცვლადს. *სოციალური სტატუსიდან* გამომდინარე, შეგვიძლია ინდივიდები მივაკუთვნოთ „მაღალ“, „საშუალო“ ან „მუშათა“ კლასს. ჩვენ შეგვიძლია ვთქვათ, რომ „მაღალი“ სოციალური ფენის წარმომადგენელს უფრო მაღალი სტატუსი გააჩნია „საშუალოსთან“ შედარებით (ან „საშუალო“ ფენის პოზიცია უფრო მაღალია, ვიდრე „მუშათა“ კლასის), მაგრამ ჩვენ არ ვიცით მათ შორის სხვაობის სიდიდე – თუ რამდენად აღემატება „მაღლის“ სტატუსი „საშუალოს“.

ცხრილი 1.1 ორდინალური რანგების სკალა

რანგი	სიდიდე
1	სრულიად ვეთანხმები
2	ვეთანხმები
3	არც ვეთანხმები და არც უარვყოფ
4	არ ვეთანხმები
5	სრულიად არ ვეთანხმები

სოციალურ მეცნიერებებში გასაზომი მრავალი თვისებრიობა ორდინალურ ცვლადს წარმოადგენს. ავიღოთ მაგალითად, ერთსქესიანი ქორწინებებისადმი ადამიანების დამოკიდებულების შეფასებისათვის გამოყენებული დებულება: "ერთსქესიან პარტნიორებს უნდა ჰქონდეთ დაქორწინების უფლება". რესპონდენტებს სთხოვენ მონიშნონ თანხმობის ან უარყოფის ხარისხი აღნიშნულთან მიმართებით. ცხრილში 1.1 მოცემულია ერთ-ერთი ფორმა, რომელიც ციფრებით გამოსახავს პასუხებს. ამასთან, ციფრებს შორის განსხვავება აღნიშნავს ერთსქესიანი ქორწინებების მხარდაჭერის ან უარყოფის მაღალ და დაბალ დონეებს, ხოლო მათ შორის განსხვავების სიდიდეს არა აქვს განმსაზღვრელი მნიშვნელობა.

ორდინალური გაზომვა ციფრები გამოიყენება რანგული თანმიმდევრობის კატეგორიების აღსანიშნავად დაბლიდან მაღლისაკენ.

ინტერვალურ-ფარდობითი გაზომვის დონე

თუ ცვლადის კატეგორიებს (ან მნიშვნელობებს) რანგები გააჩნიათ და ყველა შემთხვევაში მათი გაზომვის შედეგები ერთიდაიმავე ერთეულებშია გამოსახული, აღმოვაჩენთ გაზომვის ინტერვალურ-ფარდობით დონეს. ინტერვალურ-ფარდობითი დონით იზომება ასაკი, შემოსავლები, შეფასების ქულები (SAT). ამ ცვლადებით შეგვიძლია შევადაროთ არა მხოლოდ მათი სიდიდეები, ანუ რომელი მათგანია დიდი თუ მცირე, არამედ ისიც, თუ რამდენად აღემატება ერთი სიდიდე მეორეს. გაზომვის დონეების შესახებ მსჯელობისას, თქვენ დაინახავთ განსხვავებას ინტერვალურ-ფარდობით სიდიდეებს შორის, რომელთაც გააჩნიათ პირობითი 0-ის და ათვლის წერტილის მნიშვნელობა. ცვლადებს, რომელთაც პირობითი 0-ის მნიშვნელობა გააჩნია (0 – ნიშნავს ამ თვისების არქონას), *ფარდობით სიდიდეებს* უწოდებენ. მაგალითად, წონას და სიმაღლეს პირობითი 0-ის მნიშვნელობა გააჩნია, ხოლო ტემპერატურას – ათვლის წერტილი. სტატისტიკურ პრაქტიკაში ფარდობითი სიდიდეები განიხილება, როგორც ინტერვალური და ამის გამო იგნორირებას უკეთებენ მათ ფარდობით თვისებებს. წიგნში არავითარი განსხვავება არ არის ამ ორი ტიპის სიდიდეს შორის.

ინტერვალურ-ფარდობითი გაზომვა ყველა დაკვირვება ერთსადაიმავე ერთეულში იზომება.

გაზომვის დონეების კუმულაციური მახასიათებელი

ცვლადები, რომლებიც ინტერვალური სკალით იზომება, ასევე შეიძლება გაიზომოს ორდინალური და ნომინალური სკალებითაც. როგორც წესი, თუ მახასი-

ცხრილი 1.2 ოთხი ძირითადი პროფესიის მომუშავეთა გენდერული განაწილება

საქმიანობის ჯგუფი	დასაქმებული ქალების ხვედრითი წილი(%)
მენეჯერები, სპეციალისტები და მომიჯნავე კატეგორიები	50.8
მომსახურების სფეროს მომუშავეები	57.2
წარმოება, ტრანსპორტირება და სასაწყობო მეურნეობა	22.4
ბუნებრივი რესურსები, მშენებლობა და რემონტი	4.2

წყარო: U.S. Bureau of the Census, *Statistical Abstract of the United States*, 2010, Table 603.

ათებლებლები იზომება უმაღლესი (ინტერვალური – უმაღლესია) სკალით, მაშინ ისინი შეიძლება გავზომოთ დაბალ დონეზეც, მაგრამ – არა პირიქით. ავიღოთ, მაგალითად, *დასაქმების გენდერული სტრუქტურა*, რომელიც ჩვენს მაგალითში დამოუკიდებელი ცვლადია. 2010 წლის სტატისტიკურ ბიულეტენში მოცემულია აშშ-ში ქალების ხვედრითი წილები ოთხი ძირითადი საქმიანობის მიხედვით.

გენდერული სტრუქტურა (რომელიც გაიზომება დასაქმებული ქალების ხვედრითი წილით) ინტერვალური ცვლადია და ამიტომ გააჩნია როგორც ნომინალური, ისე ორდინალური საზომის ნიშნები. მაგალითად, ჩვენ შეგვიძლია ვთქვათ, რომ მენეჯერთა ჯგუფი განსხვავდება ბუნებრივი რესურსების ჯგუფისაგან (ნომინალური შედარება), მომსახურების სფეროს მომუშავეებში სხვა ჯგუფებთან შედარებით სჭარბობენ ქალები (ორდინალური შედარება) და სერვისის სფეროში 34.8 პროცენტით მეტი ქალია დასაქმებული წარმოების სფეროსთან შედარებით (ინტერვალური შედარება). გაზომვის თითოეული დონისათვის შესაძლებელი შედარების ვარიანტები შეჯამებულია ცხრილში 1.3 და ნახაზზე 1.4. აღსანიშნავია, რომ განსხვავებები შესამჩნევია გაზომვის სამივე დონეზე, თუმცა ამ განსხვავების სიდიდე მხოლოდ ინტერვალურ დონეზე გაიზომება.

ცოდნის შემოწმება

დარწმუნდით, რომ კარგად გაიგეთ გაზომვის დონეების შინაარსი. კურსის შესწავლისას თქვენი ლექტორი დაგისვამთ კითხვას სხვადასხვა მონაცემების ანალიზისათვის საჭირო სტატისტიკური პროცედურების შესახებ. სწორი არჩევანის გასაკეთებლად უნდა იცოდეთ მონაცემთა გაზომვის დონეები.

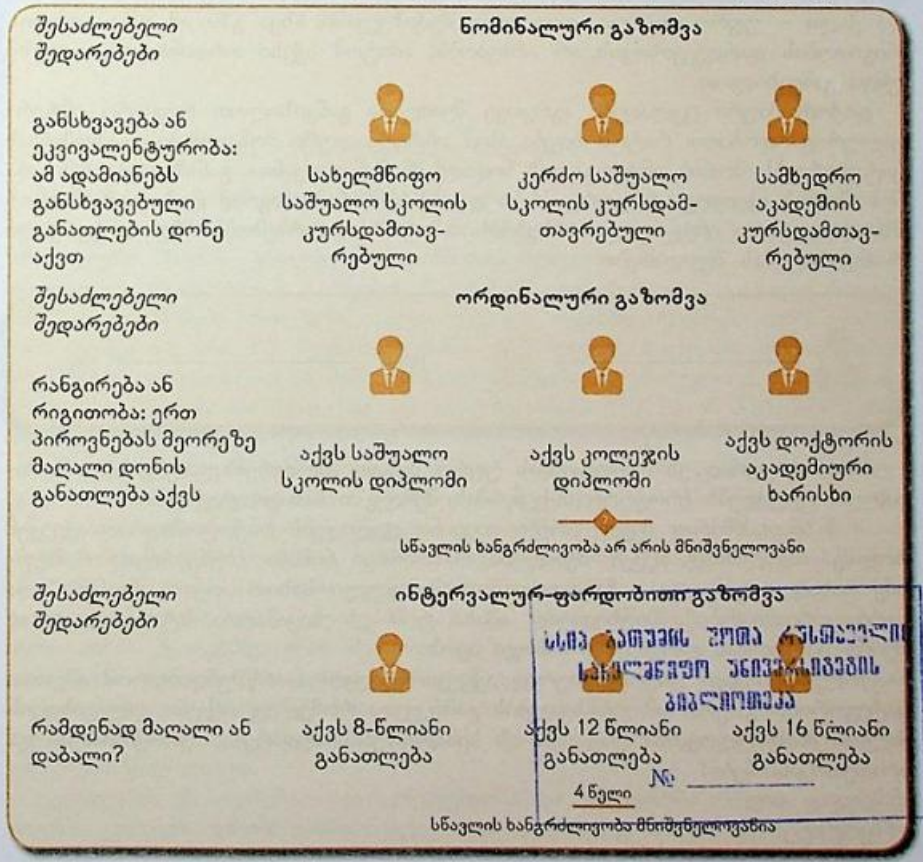
დიქოტომიური ცვლადების გაზომვის დონეები

ცვლადს, რომელსაც მხოლოდ ორი მნიშვნელობა აქვს, **დიქოტომიური ცვლადი** ეწოდება. სოციალური ფაქტორები, როგორიცაა: სქესი, სტატუსები დასაქმებასა და ქორწინებაში – დიქოტომიებია, რაც იმას ნიშნავს, რომ შეიძლება იყო მამაკაცი ან ქალი, დასაქმებული ან უმუშევარი, დაქორწინებული ან დაუქორწინებელი. როგორც ჩანს, ეს ცვლადები მხოლოდ ნომინალურ დონეზე გაიზომება, ისინი

ცხრილი 1.3 გაზომვის დონეები და შესაძლებელი შედარებები

დონე	განსხვავებულია ან ეკვივალენტურია	მაღალია ან დაბალია	რამდენად მაღალია
ნომინალური	დიახ	არა	არა
ორდინალური	დიახ	დიახ	არა
ინტერვალურ-ფარდობითი	დიახ	დიახ	დიახ

ნახაზი 1.4 გაზომვის დონეები და შესაძლებელი შედარებები: განათლების დონის გაზომვა ნომინალურ, ორდინალურ და ინტერვალურ-ფარდობით დონეებზე



-4014-

ს.ს.ი. ა.თ.უ.ი.ს. ჭ.ო.მ.ა. რ.ე.ს.უ.რ.ს.ე.ბ.ი.ს. ს.ა.ი.მ.ე.დ.ნ.ი.ე.რ.ო. უ.ნ.ი.ვ.ე.რ.ს.ი.ტ.ე.ტ.ი.ს. ბ.ი.ბ.ლ.ი.ო.თ.ე.კ.ა. №

შეიძლება მივაკუთვნოთ ან ერთ კატეგორიას, ან მეორეს. არც ერთი კატეგორია ბუნებრივად არ არის მეორეზე მეტი ან ნაკლები და შეუძლებელია მათი დალაგება რიგობითი თანმიმდევრობით.

დიქოტომიური ცვლადი ცვლადი, რომელსაც მხოლოდ ორი მნიშვნელობა აქვს.

ამასთან ერთად, მიუხედავად იმისა, რომ დიქოტომიურ ცვლადებს მხოლოდ ორი მნიშვნელობა გააჩნიათ, მათი გაზომვა ორდინალურ ან ინტერვალურ დონეებზეც შეიძლება. მაგალითად, ჩვენ შეგვიძლია რიგითობის საფუძვლად მივიჩნიოთ „ქალობა", ამის გამო „ქალი" უფრო მაღალი რანგის იქნება, ვიდრე „მამაკაცი". თუ რიგითობის საფუძვლად „მამაკაცობას" მივიჩნევთ, მამაკაცი იქნება უფრო მაღალი და ქალი – უფრო დაბალი. თითოეულ შემთხვევაში, სხვა გზა ამ ორი ჯგუფის რიგითობის დარღვევისთვის, არ არსებობს; ამიტომ სქესი ორდინალურ დონეზე უნდა განვიხილოთ.

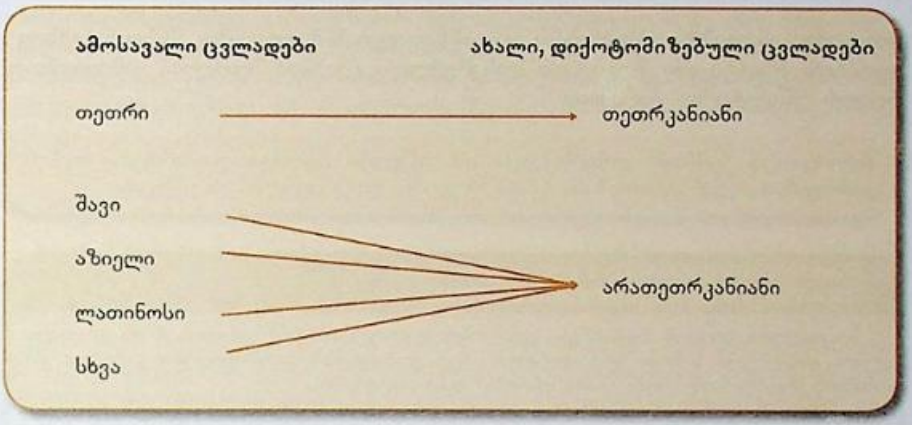
დიქოტომიური ცვლადები აგრეთვე შეიძლება განვიხილოთ როგორც ინტერვალურ-ფარდობითი. რატომ ხდება ასე? ინტერვალური მონაცემების გაზომვისას, კატეგორიებს შორის ინტერვალის სიდიდე მნიშვნელოვანია. განსხვავება 4-სა და 7-ს შორის ისეთივეა, როგორც 11-სა და 14-ს შორის. მაგრამ დიქოტომიისათვის, მხოლოდ ერთი ინტერვალი არსებობს. ამიტომ არ არსებობს სხვა ინტერვალი, რომელსაც მას შევადარებთ.



მათემატიკურად, ეს დიქოტომიას უფრო მეტად აძლიერებს, ვიდრე სხვა ნომინალურ ცვლადებს (როგორც თქვენ ამას შემდეგში ნახავთ ტექსტში).

ამ მიზნით, ხშირად მკვლევარები თავიანთ ცვლადებს დიქოტომიურად აქცევენ ხოლმე. მაგალითად, თქვენ შეიძლება რასობრივი ნიშანი (ბუნებრივად რამდენიმე რასად დაყოფილი) ნახოთ დიქოტომიზებული სახით „თეთრკანიანები" და „არათეთრკანიანები". მიუხედავად იმისა, რომ ეს საკამათოა, სტატისტიკურად შეიძლება ძალიან ლოგიკური ნაბიჯი იყოს.

როდესაც თქვენ დიქოტომიურად აქცევთ ცვლადს, დარწმუნდით, რომ ამ ორი კატეგორიით გსურთ იმ განსხვავების გამოკვეთა, რომელიც თქვენი კვლევისთვის ძალიან მნიშვნელოვანია (მაგ. აშშ-ის სენატორების შედარება: თეთრკანიანები და არათეთრკანიანები).



დისკრეტული და უწყვეტი ცვლადები

ჩვენ მიერ გამოყენებული სტატისტიკური ოპერაციები აგრეთვე განპირობებულია ცვლადების დისკრეტულობით ან უწყვეტობით. *დისკრეტულ* ცვლადებს მინიმალური ზომის ერთეულები გააჩნიათ, რომელთა დაყოფა შეუძლებელია. თითოეული ოჯახისათვის ბავშვების რაოდენობა დისკრეტული ცვლადია, რადგან მინიმალური საზომი ერთეული — ერთი ბავშვია. ოჯახში შეიძლება იყოს ორი ან სამი ბავშვი და არა 2.5. ჩვენ მაგალითში განხილული *ხელფასი*, აგრეთვე დისკრეტული ცვლადია, რადგან მისი მინიმალური საზომი ერთეული 1 ცენტია, რაც არ დაიყოფა. ადამიანს შეიძლება ჰქონდეს ხელფასი $101.21 ან $101.22, მაგრამ არა $101.21843. ხელფასებს შორის მინიმალური განსხვავების საზომია 1 ცენტი.

დისკრეტულისაგან განსხვავებით, *უწყვეტ* ცვლადებს არ გააჩნიათ მინიმალური საზომი ერთეული; მათი მნიშვნელობები შეიძლება უსასრულოდ მცირე ნაწილებად დაიყოს. *სიგრძე* უწყვეტი ცვლადების მაგალითია, რადგან მას არ გააჩნია მინიმალური სიგრძის საზომი. საგნის სიგრძე შეიძლება იყოს 12 ინჩი, 12.5 ინჩი, ან 12.532011 ინჩი. ყოველთვის შეუძლებელიც კია აბსოლუტური სიზუსტით ამ უსასრულო ცვლადების გაზომვა. *პრინციპში*, ჩვენ შეიძლება ვისაუბროთ ინჩის მეათედზე, ინჩის მეათიათასედზე, ან ინჩის მეტრლიონედ ნაწილზე. *დასაქმების გენდერული სტრუქტურა* უწყვეტი ცვლადის მაგალითია, რადგან ის გამოსახულია პროცენტობით და პროპორციას წარმოადგენს (მაგ. ქალების ხვედრითი წილი მედიცინაში), ამიტომ ის შეიძლება წარმოვადგინოთ მცირე ან უმცირესი წილადებით.

ცვლადების ეს თვისება — დისკრეტულობა და უწყვეტობა — დიდ გავლენას ახდენს კვლევის დროს გამოყენებულ ოპერაციებზე, გაზომვის პროცედურებზე,

მონაცემთა ანალიზზე, დასკვნებისა და განზოგადების მეთოდებზე. ამასთან, გახსოვდეთ, რომ პრაქტიკაში ზოგიერთი დისკრეტული ცვლადი შეიძლება გამოვიყენოთ როგორ უწყვეტი და პირიქით.

✓ *ცოდნის შემოწმება*

ჩამოთვალეთ სამ-სამი დისკრეტული და უწყვეტი ცვლადი. განმარტეთ, რატომ გამოიყენეთ თქვენ ჰიპოთეზაში ისინი როგორც დისკრეტული, ან უწყვეტი.

დეტალების განხილვა 1.1
შეცდომის წყარო: გაზომვის შეცდომა

სოციალური სფეროს მეცნიერები ცდილობენ დაგვაჯერონ, რომ კვლევის პროცესი უშეცდომოა, დაწყებული საზომი ერთეულებიდან. ჩვენ ყურადღებას გავამახვილებთ გაზომვის ორ მახასიათებელზე: საიმედოობა (სარწმუნოობა) და ობიექტურობა.

საიმედოობა ნიშნავს, რომ გაზომვა ყოველი გამოყენებისას სტაბილურ შედეგს გვაძლევს. მაგალითად, თუ შევეკითხებით ადამიანებს: „უჭერთ თუ არა მხარს პრეზიდენტ ობამას საქმიანობას?“ – ეს უფრო სარწმუნო იქნება, ვიდრე შეკითხვა: “რას ფიქრობთ პრეზიდენტ ობამას საქმიანობაზე?“. მიუხედავად იმისა, რომ მეორე კითხვაზე პასუხები მნიშვნელოვანია, ისინი მაინც გაუგებარი და გაფანტული იქნება და შეიძლება სხვადასხვა ინტერპრეტაციის საფუძველი გახდეს. მკვლევარები ყურადღებას ამახვილებენ დროში გაზომვების მდგრადობაზე სხვა დაკავშირებულ გაზომვებთან მიმართებაში, ან სხვა, ორი ან მეტი მკვლევარის მიერ მიღებულ შედეგებთან შედარებით. საიმედოობა ობიექტურობის წინაპირობაა: ჩვენ არ შეგვიძლია რაიმე ფენომენის გაზომვა, თუ ჩვენი გაზომვები არასტაბილურ შედეგს იძლევა.

ობიექტურობა ახასიათებს საზომი ერთეულის შესაბამისობას გასაზომთან. მიუხედავად იმისა, რომ *IQ* ტესტები სარწმუნოა, ჯერ კიდევ დაობენ იმის თაობაზე, შეიძლება თუ არა მათი საშუალებით ადამიანის შესაძლებლობების გაზომვა. გაზომვა შეიძლება არ იყოს ობიექტური ინდივიდუალური შეცდომების (ადამიანებს შეიძლება სურდეთ სოციალურად მათთვის მისაღები პასუხების დაფიქსირება) ან მეთოდოლოგიური შეცდომების (კითხვები არასწორად ან გაუგებრად არის ჩამოყალიბებული) დაშვების გამო.

კვლევის მეთოდების შესასწავლ კურსებზე განხილვის მუდმივი ობიექტია საიმედოობისა და ობიექტურობის დაცვისა და განმტკიცების სპეციალური ხერხები და მათი დანერგვა პრაქტიკაში.

მონაცემთა ანალიზი და ჰიპოთეზების შეფასება

მონაცემების შეგროვების შემდეგ მკვლევარები აანალიზებენ მათ და აფასებენ შესასწავლ ჰიპოთეზებს. დაკვირვებების აღნიშვნის მიზნით, მონაცემები წარმოდგენილია კოდებითა და ციფრებით. ჩვენ მაგალითში, საქმიანობის თითოეული ჯგუფი წარმოდგენილია ორი სიმრავლით: (1) ქალების ხვედრითი წილი და (2) საშუალო ხელფასი. ჩვენ რომ გვქონოდა მონაცემები 100 საქმიანობის მიხედვით, გვექნებოდა 200 სიმრავლე, თითოეულში კი ორ-ორი. ჩვეულებრივ, კვლევაში მრავალი ცვლადია ჩართული და ამიტომ მკვლევარი მონაცემთა დიდ მოცულობას აწყდება. ახლა უნდა ვიპოვოთ გზა, რომ მოვაწესრიგოთ მონაცემები, გავაანალ-

იზოთ და სხვადასხვა გარდაქმნების საფუძველზე დავადგინოთ, რას უნდა ნიშნავდეს ისინი. ეს ბოლო საფეხურები *სტატისტიკური ანალიზის* ეტაპებია, რაც ამ სახელმძღვანელოს ძირითადი თემაა. სტატისტიკური გარდაქმნები დაგვეხმარება კვლევითი ჰიპოთეზისა და იმ თეორიის *შეფასებაში*, საიდანაც ჰიპოთეზა ჩამოყალიბდა.

აღწერითი და დასკვნითი სტატისტიკა

სტატისტიკური მეთოდები ორ ჯგუფად იყოფა: *აღწერითი* და *დასკვნითი*. სანამ ავხსნიდეთ მათ შორის განსხვავებას, უნდა გავარკვიოთ, რა არის *მთლიანი* და *შერჩევითი ერთობლიობები*. **მთლიანი ერთობლიობა** არის ინდივიდების, საგნების ჯგუფების ან მოვლენების სრული ნაკრები. მაგალითად, თუ ჩვენ დავინტერესდებით საპრეზიდენტო არჩევნებზე ხმის მიცემის დროს ადამიანების ქცევით, ჩვენთვის ერთობლიობა იქნება ის მოქალაქეები, რომლებიც არჩევნებში მონაწილეობენ. თუ გვაინტერესებს ჩვენ შტატში ლათინური წარმოშობის მოსახლეობის დასაქმების საკითხები, ამ შემთხვევაში ერთობლიობა იქნება დასაქმებული ლათინო-ამერიკელი მოსახლეობა.

მთლიანი ერთობლიობა საკვლევი ინდივიდების, საგნების ჯგუფების ან მოვლენების სრული ნაკრები.

მიუხედავად იმისა, რომ ჩვენი ინტერესის საგანს ყოველთვის მთლიანი ერთობლიობა წარმოადგენს, ხშირად მისი შესწავლა შეზღუდული დროისა და რესურსების გამო შეუძლებელია. წარმოიდგინეთ უკანასკნელ არჩევნებზე მისული ყველა მოქალაქის გამოკითხვა, ან თუნდაც შტატის ლათინური წარმოშობის მოსახლეობის გამოკითხვა დასაქმებასთან დაკავშირებით. ეს არამარტო ძალიან ძვირი დაჯდებოდა და დიდ დროს დავკარგავდით, არამედ სხვა ბევრ პრობლემასაც გადავაწყდებოდით! საბედნიეროდ, ჩვენ შეგვიძლია დიდი ინფორმაციის მიღება მთლიანი ერთობლიობის შესახებ, თუ ყურადღებით შევარჩევთ და შევისწავლით მის ნაწილს. მთლიანი ერთობლიობიდან შერჩეულ ნაწილს **შერჩევითი ერთობლიობა** ეწოდება. ჩვეულებრივ, მკვლევარები მონაცემებს შერჩევითი ერთობლიობიდან იღებენ და შემდეგ ავრცელებენ მთლიან ერთობლიობაზე.

შერჩევითი ერთობლიობა მთლიანი ერთობლიობიდან შერჩეული შედარებით მცირე ნაწილი.

აღწერითი სტატისტიკა მოიცავს მთლიანი ან შერჩევითი ერთობლიობიდან მოპოვებული ინფორმაციის თავმოყრისა და აღწერის მეთოდებს. როგორც

წესი, მთლიანი ერთობლიობიდან მონაცემების შეგროვება მოსახლეობის აღწერების დროს ხდება. მეორე მხრივ, დასკვნითი სტატისტიკა კონცენტრირებულია მთლიანი ერთობლიობის შესახებ პროგნოზების ან დასკვნების გაკეთებაზე დაკვირვებებისა და შერჩევითი ერთობლიობების ანალიზის საფუძველზე. მაგალითად, ზოგადი სოციალური კვლევა (GSS), საიდანაც ამ წიგნში უამრავი მაგალითია მოტანილი, ეროვნული აზრის კვლევის ცენტრის (NORC) მიერ ყოველ ორ წელიწადში ერთხელ ტარდება. ის იკვლევს რეპრეზენტატულ შერჩევით ერთობლიობას რესპონდენტების რამდენიმეათასიანი შემადგენლობით (მაგ. 2006 წელს გამოიკითხა 4,510 და 2008-ში 2,023 ადამიანი). ეს კვლევა, რომელიც ასეულობით კითხვას მოიცავს, იმისთვისაა შემუშავებული, რომ სოციალური სფეროს მეცნიერებს მიაწოდოს გამზადებული მონაცემთა ბაზა სოციალურად რელევანტური დამოკიდებულებების, ქცევებისა და მახასიათებლების შესახებ მთელი აშშ-ის მოზრდილი მოსახლეობის მასშტაბით. NORC-ი ადასტურებს, რომ მიღებული შედეგები ძალიან უახლოვდება მოსახლეობის აღწერის შედეგებს. მაგრამ რადგან ეს კვლევა შერჩევით ერთობლიობაზე ტარდება, საშუალო მაჩვენებლები აღწერების საშუალოებს არ ემთხვევა. მაგალითად, GSS-მა 2008 წელს გამოკითხა მოსახლეობა განათლებაზე დახარჯული წლების რაოდენობის შესახებ. GSS-მა შედეგად მიიღო 13.34 წელი, რაც ცოტა მეტია საშუალო სკოლის დიპლომის მიღებისათვის საჭირო დროზე. ეს საშუალო ალბათ განსხვავდება მთლიანი ერთობლიობის მაჩვენებლისგან. დასკვნითი სტატისტიკის მეთოდები გვეხმარება მკვლევარების მიერ შერჩევითი ერთობლიობიდან მიღებული საშუალოს სიზუსტის დადგენაში.

აღწერითი სტატისტიკა მეთოდები, რომლებიც გვეხმარება მონაცემთა თავმოყრასა და აღწერაში შერჩევითი ან მთლიანი ერთობლიობიდან.

დასკვნითი სტატისტიკა მთლიანი ერთობლიობის შესახებ პროგნოზებისა და დასკვნების გაკეთების ლოგიკური თანმიმდევრობა და მეთოდები, რომელიც მიიღება შერჩევით ერთობლიობაზე დაკვირვებისა და მისი ანალიზის საფუძველზე.

ჰიპოთეზების შეფასება

განვიხილეთ რა აღწერითი და დასკვნითი სტატისტიკის მეთოდები, ჩვენ შეგვიძლია წინ წავიწიოთ კვლევის პროცესში: საანალიზო მონაცემების საფუძველზე შევიმუშაოთ და შევაფასოთ ჰიპოთეზები. ამ საფეხურზე შეიძლება ახალი საკითხები წამოიჭრას — მონაცემებში შემჩნეულ ახალ ტენდენციებთან და ამოსავალ ცვლადებთან ერთად სხვა ცვლადების შემოტანასთან დაკავშირებით. მაგალითად, დასაქმებაში ქალთა ხვედრითი წილისა და შემოსავლების კვლევის დროს შეიძლება აღმოვაჩინოთ, რომ ეს ცვლადები დასაქმების რამდენიმე ჯგუფთან

მიმართებაში უნდა განვიხილოთ და არა სხვა ჯგუფებში. ანალოგიურად, ამ ცვლადებს შორის ურთიერთკავშირი შეიძლება განვიხილოთ რამდენიმე რასობრივ/ეთნიკური ჯგუფის ჭრილში და არა დანარჩენებისათვის.

ეს ყოველივე გვიბიძგებს იმის დადგენისაკენ, თუ როგორ კავშირშია ჩვენ მიერ მიღებული მონაცემები თეორიულ დებულებებთან. ახალ გარემოებათა საფუძველზე შეიძლება მოგვიწიოს ჩვენი ჰიპოთეზებისა და თეორიების გადახედვა. ჩვენი წარმოდგენა სამუშაო ადგილებზე ქალთა სეგრეგაციის შესახებ ბოლო კვლევებმა შეცვალა. კერძოდ, ამ კვლევებმა აჩვენა, რომ აშშ-ში დასაქმების სტრუქტურას განსაზღვრავს როგორც რასობრივი, ისე გენდერული ნიშანი და ეს გვეხმარება შემოსავლების უთანაბრობის ახსნაში. თავის მხრივ, რეფორმირებული თეორია ახალი ჰიპოთეზების დაყენებასა და ახალი კვლევის ჩატარებას მოითხოვს, რაც სათავეს უდებს ახალ წრიულ პროცესს.

სტატისტიკა უზრუნველყოფს თეორიისა და კვლევის მნიშვნელოვან კავშირს. როგორც გენდერული სეგრეგაციის მაგალითი მოწმობს, სტატისტიკური მეთოდების გამოყენება კვლევითი პროცესის განუყოფელი ნაწილია. სტატისტიკური ანალიზის შედეგები გვეხმარება ჰიპოთეზებისა და თეორიების შეფასებაში, აღმოაჩენს ცვალებად მოვლენებსა და ტენდენციებს და ბიძგს აძლევს თეორიების ჩამოყალიბებასა და განვითარებას. ამასთან, სტატისტიკის აუცილებლობა არამც და არამც არ აკნინებს კვლევის შემდგომი ეტაპების მნიშვნელობას. ასევე არ ამცირებს სტატისტიკა მთლიან პროცესში მსჯელობის აუცილებლობას. სტატისტიკური ანალიზი კვლევის პროცესის შედარებით მცირე ნაწილია და ყველაზე სკურპულოზური სტატისტიკური მეთოდებიც კი თავისით ვერ ამეტყველდებიან. თუ კვლევის კითხვები ძნელად აღსაქმელია და მონაცემები ხარვეზებით სავსეა, შედეგი აბსოლუტურად გამოუსადეგარი იქნება.

სოციალური განსხვავებების გათვალისწინება

თუ არსებული ტენდენციები მიმდინარე საუკუნის შუა წლებისთვის შენარჩუნდება, აშშ აღარ იქნება დომინანტურად ევროპული ქვეყანა. ძლიერი იმიგრაცია და გაზრდილი ბუნებრივი მატება გამოიწვევს აშშ-ის ტრანსფორმაციას „გლობალურ საზოგადოებად“, რომელშიც მოსახლეობის დაახლოებით ნახევარი აფრიკული, აზიური, ლათინო-ამერიკული და ადგილობრივი ამერიკული წარმოშობის იქნება. ეხება თუ არა ამერიკული საზოგადოების ზრდადი მრავალფეროვნება სოციალური სფეროს მეცნიერებს? რა გავლენას იქონიებს ეს ფაქტორი ჩვენ მიერ გამოყენებულ კვლევის მეთოდოლოგიებზე?

ამ განსხვავებულ საზოგადოებაში – რასის, ეთნიკური ნიშნის, სოციალური ჯგუფებისა და სქესის მიხედვით, სოციალური ურთიერთობების ნაკლებად დეტალიზებული და გამრუდებული ახსნა სასურველ შედეგამდე ვერ მოგვიყვანს. რეალურ შედეგს მივიღებთ მხოლოდ იმ შემთხვევაში, თუ მკვლევარები, კვლევის მონაწილეები და თვით კვლევის პროცესი სათანადოდ ასახავს აღნიშნულ მრა-

ვალფეროვნებას. ეს მრავალფეროვნება განსაზღვრავს კვლევის კითხვებს, დაკვირვების ხერხებსა და დასკვნებს.

როგორ ხდება ამ სოციალური განსხვავებების აღქმა სტატისტიკის მიერ? როგორ ხდება რასობრივი, გენდერული და სხვა დემოგრაფიული ცვლილებების მოქცევა სტატისტიკის ობიექტივში? სტატისტიკა თავისი მეთოდებით სწავლობს სოციალური განსხვავებების, როგორიცაა რასობრივი, კლასობრივი, გენდერული და დემოგრაფიული — მათ შორის ასაკობრივი, რელიგიური, სქესობრივი — ზეგავლენას ჩვენ სოციუმსა და სოციალურ ქცევაზე. ამ შემთხვევაში, მრავალი სტატისტიკური მეთოდი შეიძლება გამოვიყენოთ, რასაც მომდევნო თავიდან შევუდგებით. ახლა კი მოკლედ მიმოვიხილოთ ის მეთოდები, რომლებიც სოციალური განსხვავებების ანალიზში გამოგვადგება.

მეორე თავში ჩვენ გავეცნობით ინფორმაციის თავმოყრის აღწერით ხერხებს, როგორიც არის სიხშირული განაწილება, პროცენტული განაწილება, შეფარდებითი სიდიდეები და კოეფიციენტები. ეს ხერხები გამოიყენება ავრეთვე ცალკეული ჯგუფების ისეთი მახასიათებლების გამოსავლენად, რაც ნაკლებად შესამჩნევი იყო სხვებთან შედარებით. სწორედ ეს აღწერითი მეთოდები იქნა გამოყენებული აშშ-ის აღწერების ბიუროს მიერ უკანასკნელ წლებში გამოქვეყნებულ სპეციალურ ანგარიშებში საზღვარგარეთ დაბადებულთა, ლათინო-ამერიკელებისა და ხანდაზმულ ამერიკელთა შესასწავლად.

მესამე თავში ნაჩვენებია გრაფიკული მეთოდის შესაძლებლობები განსხვავებების თვალსაჩინოებისათვის. გრაფიკები გვეხმარება ავსახოთ აშშ-ის საზოგადოებაში არსებული სოციალური ჯგუფების მსგავსებები და განსხვავებები და მივუთითოთ ამ სტრუქტურული ცვლილებების სისწრაფეზე. აშშ-ის აღწერების ბიუროს მონაცემების გამოყენებით ჩვენ ავაგებთ სხვადასხვა სახის გრაფიკებს და ვაჩვენებთ მსგავსებებსა და განსხვავებებს ხანდაზმული ამერიკელების დასაქმებაში. ასე მაგალითად, მარტივი სვეტებიანი დიაგრამების საშუალებით გაჩვენებთ, რომ ხანდაზმულ ამერიკელთა შორის თითოეულ ასაკობრივ კატეგორიაში ქალები უფრო ნაკლებად არიან დასაქმებული, ვიდრე მამაკაცები.

თუ მსგავსება და საერთო თვისებები საშუალოების (თავი 4) საფუძველზე გამოვლინდება, სოციალურ ჯგუფებში არსებული განსხვავებები და მრავალფეროვნება უკვე ვარიაციის მაჩვენებლებით აიხსნება. მაგალითად, თუ გვინდა შევისწავლოთ ასაკობრივი სტრუქტურის ცვლილება, ან შევადაროთ 50 შტატის მიხედვით რასობრივი, ეთნიკური ან რელიგიური მრავალფეროვნების ხარისხი, უნდა გამოვიყენოთ საშუალო კვადრატული გადახრა და თვისებრივი ვარიაციის ინდექსი (*IQV*). მაგალითად, ამ ინდექსის საფუძველზე ჩვენ შეიძლება ვთქვათ, რომ მეინის შტატი ყველაზე ნაკლებად მრავალფეროვანია, ხოლო ჰავაი — პირიქით (თავი 5).

მე-6-დან მე-14 თავის ჩათვლით განხილულია დასკვნითი სტატისტიკისა და ბივარიაციული ანალიზის საკითხები. თავიდან შევუდგებით დასკვნითი სტატისტიკის საფუძვლების შესწავლას — ნორმალური განაწილება, შერჩევა, ალბათობა და შეფასებები — თავები 6-8. მე-9-დან მე-14 თავების ჩათვლით ვიმუშავებთ შერჩევაზე, გავზომავთ ურთიერთკავშირს კლასებს, სქესს ან ეთნიკურობასა და

სხვა სოციალურ ქცევებსა და მახასიათებლებს შორის. დასკვნითი სტატისტიკიდან t ტესტი, ხი-კვადრატი და *F* სტატისტიკა დაგვეხმარება შერჩევის შეცდომის გამოვლენაში. ამასთან ერთად, ჩვენ შევეხებით ბივარიაციულ ანალიზსაც, რაც განსაკუთრებით სასარგებლოა ურთიერთკავშირების შესწავლისათვის ისეთ ცვლადებს შორის, როგორიცაა რასა, სქესი, კლასი, ეთნიკურობა და რელიგია. ანალიზის აღნიშნულ მეთოდებს ვიყენებთ არა მხოლოდ იმისთვის, რომ დავაკვირდეთ თითოეული ცვლადის დამოუკიდებელ ზეგავლენას ჩვენ ქცევაზე, არამედ იმისთვისაც, თუ როგორ ურთიერთქმედებენ ისინი საზოგადოებაში ჩვენი ინდივიდუალური გამოცდილების ჩამოსაყალიბებლად.

დეტალების განხილვა 1.2
მარტივი არითმეტიკა: კულტურის ზეგავლენა თვლაზე

მეორე კლასის მასწავლებელმა ბავშვებს წაუკითხა ამოცანა: „ხეზე იჯდა 4 შაშვი. ვთქვათ, შურდულით მიზანში ამოიღე ერთი მათგანი და მოკალი. რამდენი დარჩა?“

„სამი“, დარწმუნებით უპასუხა შვიდი წლის ევროპელმა. „ოთხს გამოვაკლოთ ერთი, უდრის სამს“.

„არც ერთი“, ასეთივე დარწმუნებით უპასუხა შვიდი წლის აფრიკელმა. „თუ ერთს მოკლავ, დანარჩენები გაფრინდებიან“*

---

* *Working Woman*, January 1991, p.45

სოციალური კვლევის რომელი მოდელიც არ უნდა აირჩიოთ, ტრადიციული თუ თვისებრივ ანალიზთან ინტეგრირებული, სოციალურ განსხვავებებზე თუ სოციალური ქცევის სხვა ასპექტებზე ორიენტირებული, დაიმახსოვრეთ, რომ სტატისტიკური ოპერაციების განხორციელება საბაზისო სტატისტიკური ცნებებისა და მეთოდების ცოდნას მოითხოვს. ეს შესავალი კურსი სწორედ იმისთვისაა, რომ უკეთ გაეცნოთ სოციალურ მეცნიერებებში გამოყენებული აღწერითი და დასკვნითი სტატისტიკის არსენალს. ამასთან, რა თქმა უნდა, ადამიანურ ფაქტორზე ბევრია დამოკიდებული სტატისტიკური კვლევებისას. მხოლოდ აღნიშნულის გაცნობიერებით შეიძლება გახდეს სტატისტიკა სასარგებლო იარაღი სოციალური მოვლენების შესასწავლად.

დეტალების განხილვა 1.3
სტატისტიკის გამო ნერვიულობთ?

ზოგიერთმა თქვენგანმა ალბათ გარკვეული ეჭვითა და ძალიან მცირე ენთუზიაზმით შეარჩიეთ ეს საგანი. სიტყვა სტატისტიკა ცოტა ნერვიულობასთან არის დაკავშირებული, რადგან მისი წარმოთქმისას თვალწინ წარმოგიდგებათ ციფრები, ფორმულები, აბსტრაქტული მსჯელობები და ფიქრობთ, რომ ეს ყველაფერი დაუძლეველი და ძალიან რთულია. როგორც ჩანს, სტატისტიკა ისე ახლოს არ არის თქვენთან, როგორც სხვა საგნები.

შესაძლებელია, რომ სტატისტიკა ყველაზე დიდი თავის ტკივილია სოციალურ მეცნიერებათა კურიკულუმში. ეს უსიამოვნო ფიქრები იქიდან მოდის, რომ სტატისტიკის სასწავლო გარეშო საშუალოზე დაბალია; ხშირად სტუდენტები კონცენტრირებულნი არიან სტატისტიკური პროცედურების ზედმიწევნით დამახსოვრებაზე და ნაკლებად ცდილობენ ჩაწვდნენ მათ აზრს.

ჩვენი მრავალწლიანი პრაქტიკის საფუძველზე მივედით იმ დასკვნამდე, რომ სტატისტიკა-სადმი სტუდენტების შიშის ძირითადი მიზეზი, მათი აზრით, უამრავი უაზროდ დასამახსოვრებელი ფორმულის არსებობაა.

არავინ უარყოფს, რომ სტატისტიკაში ბევრი უცნაური სიმბოლო და ტერმინია. ისიც მართალია, რომ ციფრა მათემატიკის ცოდნაც საჭიროა. თუმცა ყველაფერი, რაც უნდა იცოდეთ, ოთხი მათემატიკური ოპერაციაა: შეკრება, გამოკლება, გამრავლება და გაყოფა. სტატისტიკის ენაც შეიძლება რთულად მოგეჩვენოთ, რადგან ეს ოპერაციები (და მათი კომბინაცია) ისეა ჩაწერილი, რომ თქვენთვის გაუგებარია. აბსტრაქტული მსჯელობები კი სტატისტიკური ენის ნაწილია; როგორც უცხო ენის შესწავლისას ხდება ხოლმე, ჯერ უნდა სტატისტიკის ანბანი ისწავლოთ, რომ „ალაპარაკდეთ ამ ენაზე„. როდესაც თქვენ სტატისტიკურ ლექსიკონს დაეუფლებით და შეძლებთ სიმბოლოების და ჩანაწერების გაგებას, უფრო დამშვიდებით შეუდგებით სტატისტიკის გაცნობიერებას*.

---

* ეს დისკუსია ეფუძნება შეილა ტობიასის ნაშრომს. Sheila Tobias, *Overcoming Math Anxiety* (New York: Norton, 1995), chapters 2 and 8.

ძირითადი განმარტებები

• სტატისტიკა არის მეთოდების ერთობლიობა, რომელიც გამოიყენება სოციალური დარგის მეცნიერთა მიერ ინფორმაციის თავმოყრის, განზოგადებისა და გადმოცემისათვის.

• კვლევის პროცესი წარმოადგენს საქმიანობათა ერთობლიობას, რომელსაც ახორციელებენ სოციალური სფეროს მეცნიერები იმისათვის, რომ გასცენ პასუხები კითხვებს, შეამოწმონ მოსაზრებები და ჰიპოთეზები.

• თეორია არის ცალკეული ინდივიდებისა ან ჯგუფების ორი ან მეტი მახასიათებლის ურთიერთკავშირის მეცნიერული ახსნა.

• თეორია გვთავაზობს სპეციფიკურ კონკრეტულ პროგნოზებს იმის შესახებ, თუ როგორ არიან ერთმანეთთან დაკავშირებული ცალკეული ინდივიდებისა თუ ჯგუფების მახასიათებლები. ეს პროგნოზებია, რომელსაც ჰიპოთეზებს ვუწოდებთ.

• ცვლადები წარმოადგენს ადამიანის ან საგნის მახასიათებლებს, რომლებმაც შეიძლება მიიღოს ორი ან მეტი მნიშვნელობა. ცვლადს, რომლის ახსნასაც ცდილობს მკვლევარი („შედეგი“) ეწოდება დამოკიდებული ცვლადი, ხოლო იმ ცვლადს, რომელიც იწვევს ამ დამოკიდებული ცვლადის ცვლილებას („მიზეზი“), დამოუკიდებელი ეწოდება.

• იმისათვის, რომ დადგინდეს მიზეზ-შედეგობრივი ურთიერთკავშირი ორ ცვლადს შორის, უნდა დაკმაყოფილდეს სამი პირობა: (1) მიზეზმა შედეგი უნდა გამოიწვიოს დროში, (2) მიზეზსა და შედეგს შორის უნდა იყოს ემპირიული ურთიერთკავშირი, და (3), ეს ურთიერთობა არ უნდა აიხსნას სხვა ფაქტორებით.

• გაზომვის ნომინალურ დონეზე ციფრები და სიმბოლოები გამოიყენება სახელდების, აღნიშვნის ან დაკვირვებების დაჯგუფების მიზნით. გაზომვის ორდინალურ დონეზე კატეგორიებს ვანიჭებთ რანგებს დაბლიდან მაღლისკენ (ან პირიქით) თანმიმდევრობით. ინტერვალურ-ფარდობითი გაზომვის დროს გაზომვის ყველა მნიშვნელობა ერთიდაიმავე საზომი ერთეულებით გამოისახება.

• მთლიანი ერთობლიობა მკვლევარისათვის საინტერესო საკვლევი ინდივიდების, საგნების, ჯგუფების ან მოვლენების სრული ნაკრებია. შერჩევითი ერთობლიობა მთლიანი ერთობლიობიდან შერჩეული შედარებით მცირე ნაწილია.

• აღწერითი სტატისტიკა მოიცავს მეთოდებს, რომლებიც გვეხმარება მონაცემთა თავმოყრასა და აღწერაში შერჩევითი ან გენერალური ერთობლიობიდან. დასკვნითი სტატისტიკა მოიცავს გენერალური ერთობლიობის შესახებ პროგნოზებსა და დასკვნებს შერჩევით ერთობლიობაზე დაკვირვებისა და მისი ანალიზის საფუძველზე.

საკვანძო ტერმინები

მონაცემი
დამოკიდებული ცვლადი
აღწერითი სტატისტიკა
დიქოტომიური ცვლადი
ემპირიული კვლევა
ჰიპოთეზა
დამოუკიდებელი ცვლადი
დასკვნითი სტატისტიკა
ინტერვალურ-ფარდობითი გაზომვა
ნომინალური გაზომვა
ორდინალური გაზომვა
მთლიანი ერთობლიობა
კვლევის პროცესი
შერჩევითი ერთობლიობა
სტატისტიკა
თეორია
საანალიზო ერთეული
ცვლადი

დამატებითი რესურსები

ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

მონაცემების მწკრივებისა და ცვლადების გაცნობა

ყოველი თავის ბოლოს ჩვენ გამოვიყენებთ კომპიუტერულ მონაცემებსა და სავარჯიშოებს. კომპიუტერული სავარჯიშოები დაფუძნებულია პროგრამაზე IBM ® SPSS ვერსია 18.0 (ფორმალურად პროგრამას დაერქვა PASW ® Statistics). აქ მოცემულია პროგრამის ორი ვერსია: სტანდარტული ვერსია ცვლადებისა და დაკვირვებების შეუზღუდავი რაოდენობით და სტუდენტური

ვერსია ლიმიტირებული 1 500 დაკვირვებითა და 50 ცვლადით. შეეკითხეთ თქვენ ლექტორს, რომელი ვერსიაა მისაწვდომი თქვენს სასწავლებელში.

ამ სახელმძღვანელოს ფარგლებში თქვენ იმუშავებთ მონაცემთა ორ მწკრივზე. GSS2008.SAV მოიცავს ცვლადებისა და დაკვირვებების ნაწილს GSS 2008-იდან, რომელიც 1972 წლიდან მოყოლებული ორ წელიწადში ერთხელ ტარდება. ეს კვლევა, რომელიც NORC-ის მიერ ჩიკაგოს უნივერსიტეტში ტარდება, იმისთვისაა შემუშავებული, რომ სოციალური სფეროს მეცნიერებს მიაწოდოს გამზადებული მონაცემთა ბაზა სოციალურად რელევანტური დამოკიდებულებების, ქცევებისა და მახასიათებლების შესახებ მთელი აშშ-ის მოზრდილი მოსახლეობის მასშტაბით.

2008 წელს საერთო ჯამში 2,023 კვლევა ჩატარდა. მონაცემები, რომელიც შეირჩა მრავალსაფეხურიანი ალბათური შერჩევით, მოიცავდა 18 წლისა და ზევით ასაკის ამერიკელებს. ეს ნიშნავს, რომ GSS მონაცემები საშუალებას გვაძლევს შევაფასოთ ინგლისურენოვანი ამერიკელი ფიზიკური პირების მახასიათებლები, შეხედულებები და ქცევები.

MTF2008.SAV მოიცავს მიჩიგანის უნივერსიტეტის კვლევას, რომელსაც ეწოდება მომავლის მონიტორინგი (MTF). ეს არის მე-12 კლასელთა კვლევა და წარმოადგენს 1975 წელს დაწყებული კვლევების სერიის ნაწილს, რომელიც შეისწავლის ახალგაზრდა თაობის ფასეულობებს, ქცევებს, ცხოვრების სტილს. 2008 წლის კვლევა შეეცადა კითხვებს ნარკოტიკების მოხმარებასა და კანონსაწინააღმდეგო ქცევებზე.

HINTS2007.SAV მოიცავს მონაცემებს 2007 წლის ჯანდაცვის ინფორმაციის ეროვნული ტენდენციების დაკვირვებიდან (HINTS), რომელიც ტარდება კიბოს ეროვნული ცენტრის მიერ. მისი მიზანია ჯანდაცვის სწრაფადგანვითარებადი საკომუნიკაციო სფეროების კვლევა როგორც ინგლისურენოვან, ისე ესპანურენოვან მოსახლეობაში. ეს კვლევა მოიცავს 7,674 რესპონდენტის 472 განსხვავებულ ცვლადს.

ვინც აირჩევს სტუდენტურ ვერსიას, ის იმუშავებს GSS-ის ორ განსხვავებულ ფაილთან: GSS08PFP-A, რომელშიც მოცემულია სქესი და ასაკი და GSS08PFP-B, სადაც მოცემულია რასა და ინფორმაცია სამთავრობო პოლიტიკის შესახებ. თითოეული მათგანი მოიცავს 1,500 დაკვირვებას და 50 ცვლადს. თქვენ შეგიძლიათ გამოიყენოთ აგრეთვე MTF2008 SPSS-ის სტუდენტურ ვერსიასთან ერთად (უფრო ვრცელი მონაცემთა ბაზა). ამ სახელმძღვანელოს სასწავლო საიტზე მოცემულია SPSS-ის საბაზისო პროცედურები სტუდენტური ვერსიისთვის. გირჩევთ SPSS-ის სავარჯიშოების გაკეთებამდე გაეცნოთ ამ დანართს.

როდესაც თქვენ დაიწყებთ ამ მონაცემების დამუშავებას, მოგიწევთ გაიხსენოთ ზოგიერთი რამ ცვლადების შესახებ. რას ეწოდება ცვლადი? როგორ იზომება? რას ნიშნავს ეს? არსებობს რამდენიმე გზა ამის გასაკეთებლად.

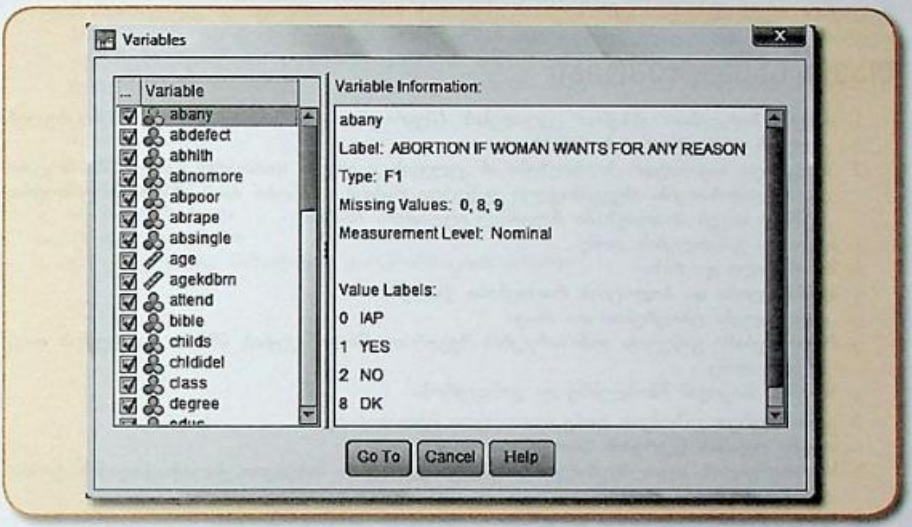
ამ მონაცემების სანახავად ჯერ უნდა გახსნათ მონაცემთა ფაილი. SPSS-ში ფაილების გახსნა ხდება File-ზე დაკლიკებით, შემდეგ Open და მერე Data. დირექტორიის შეჩევის შემდეგ აირჩიეთ მონაცემები და დაკლიკეთ Open. ეს ოპერაციები ყოველთვის უნდა გაიმეოროთ, როდესაც ხსნით ფაილს. SPSS ავტომატურად ხსნის მონაცემთა თითოეულ ფაილს ფანჯარაში Data Editor, რომლის სახელია Labeled Data View. თვალსაჩინოებისათვის ჩვენ გამოვიყენებთ GSS08PFP-A.SAV-ს.

ერთ-ერთი გზა მონაცემთა ჩამონათვალის დათვალიერებისათვის არის Utilities დაკლიკება ძირითადი მენიუდან, შემდეგ შეარჩიეთ ცვლადები Variables ქვემენიუდან. გაიხსნება დიალოგური ფანჯარა (როგორც ეს ნაჩვენებია ნახ. 1.5). ცვლადების სახელები, რომელთა სიგრძე შეზღუდულია 8 სიმბოლოთი, მოცემულია ჩამონათვალის ყუთში (მარცხენა სვეტი). როდესაც მოინიშნება ცვლადი, მისი აღწერითი სახელდება გამოჩნდება გამოტოვებულ სიდიდეების ჩვენებით და თუ ეს შესაძლებელია, თითოეული ცვლადის ვარიაციი მითითებული იქნება. (ხანდახან SPSS-ში სწორად არ არის მითითებული ცვლადის გაზომვის დონე. მაგალითად, ცვლადისთვის HEALTH, SPSS უთითებს ზომს „nominal", თუმცა HEALTH არის ორდინალური ცვლადი. ყოველთვის შეამოწმეთ გაზომვის დონის სისწორე). ცვლადები ალფაბეტური თანმიმდევრობით არის მოცემული.

მონაცემების დათვალიერების მეორე გზა არის პირდაპირ Variable View Window-ში შესვლა. აქ შეგხვდებათ ორი ეტიკეტი Data View და Variable View. დაკლიკეთ Data View და როგორც ნახ. 1.6-ზეა ნაჩვენები, თქვენ დაინახავთ ყველა ცვლადის მნიშვნელობას. თითოეული სვეტი სპეციფიკურ ინფორმაციას შეიცავს ცვლადების შესახებ. სვეტები „Label" და „Values" უზრუნველყოფენ

ცვლადის სახელდებას (მოკლე აღწერილობა იმისა, თუ რას ვზომავთ) და მნიშვნელობების სახელდებას (თითოეული ცვლადის კატეგორიისათვის).

ნახაზი 1.5 ცვლადების დიალოგური ფანჯარა – Utilities



ნახაზი 1.6 ფანჯარა Data View Window GSS02PFP-ისთვის

1 : abany | 0

	abany	abdefect	abhlth	abnomore	abpoor	abrape	absingle
1	0	0	0	0	0	0	0
2	1	1	1	2	2	1	2
3	1	1	1	1	2	1	2
4	0	0	0	0	0	0	0
5	2	2	2	2	2	2	2
6	0	0	0	0	0	0	0
7	2	2	2	2	2	2	2
8	0	0	0	0	0	0	0
9	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1

ამოცანა SPSS-ისათვის

Utilities Variables ოპციაზე დაყრდნობით, თქვენ შეგიძლიათ ცვლადების დათვალიერება GSS08P-FP-A-დან (ან GSS08PFP-B-დან). შეგიძლიათ თუ არა გამოავლინოთ სამ-სამი ნომინალური, ორდინალური და ერთი მაინც ინტერვალური ცვლადი? Variable View დიალოგური ფანჯრის საშუალებით თქვენ შეძლებთ განსაზღვროთ ცვლადის სახელი, იარლიყი და კატეგორიის სიდიდე.

თავის სავარჯიშოები

1. თქვენი სიტყვებით ახსენით ცვლადების (შეგროვილი და საანალიზო) დამოკიდებულება კვლევის პროცესთან (იხ. ნახ.1.1)
2. შეადგინეთ სავარაუდო ჰიპოთეზები ან კვლევის კითხვები თითოეულ მაგალითში მოყვანილი ცვლადისათვის. ასევე, მოკლედ დაწერეთ, რატომ ფიქრობთ რომ არსებობს ურთიერთკავშირი თქვენ ჰიპოთეზაში მოყვანილ ცვლადებს შორის.

ა. სქესი და განათლების დონე
ბ. შემოსავალი და რასა
გ. დაბნაშავეობა და პოლიციის რაოდენობა ქალაქში
დ. კმაყოფილება ცხოვრებით და ასაკი
ე. შეიარაღებაზე გაწეული დანახარჯების ხვედრითი წილი ქვეყნის მშპ-ში და ქვეყნის თავდაცვის დონე
ვ. ზრუნვა ასაკოვან მშობლებზე და ეთნიკურობა

3. განსაზღვრეთ გაზომვის დონე თითოეული ცვლადისათვის:

ა. თქვენი ოჯახის წევრების რაოდენობა
ბ. საცხოვრებლის ადგილმდებარეობა შემდეგი ჯგუფების მიხედვით: ქალაქი, ქალაქის ტიპის დასახლება, სოფელი
გ. სტუდენტების რაოდენობა სახელმწიფო უმაღლეს სასწავლებლებში
დ. სახელმძღვანელოს ხარისხის რეიტინგი სკალაზე „ძალიან კარგიდან„ დაწყებული, „ცუდის“ ჩათვლით
ე. სამსახურამდე სამგზავრო ტრანსპორტის ტიპი (მაგ. ავტობუსი, ფეხით, ავტომობილი)
ვ. თქვენი წლიური შემოსავალი
ზ. უმუშევრობის დონე აშშ-ში
თ. საპრეზიდენტო კანდიდატი, რომელსაც მიეცით ხმა 2008 წლის არჩევნებში

4. თითოეული ცვლადისათვის, რომელიც ინტერვალურია სავარჯიშო 3-ში, გაარკვიეთ დამოუკიდებელია ის თუ დამოკიდებული.
5. როგორ ფიქრობთ, ქალები და მამაკაცები საშუალოდ განსხვავებულ ანაზღაურებას იღებენ? გამოიყენეთ სამი დამოუკიდებელი ცვლადი თქვენ თეორიაში, სადაც შემოსავალი დამოკიდებული ცვლადი იქნება. გამოიყენეთ ჰიპოთეზები და დააკავშირეთ თითოეული დამოუკიდებელი ცვლადი დამოკიდებულთან.
6. თითოეული ქვემოთ ჩამოთვლილი მაგალითისთვის მიუთითეთ, რომელ სტატისტიკას გამოიყენებთ: აღწერითს თუ დასკვნითს. დაასაბუთეთ თქვენი არჩევანი.

ა. უმუშევარი ადამიანების რაოდენობა აშშ-ში
ბ. განსაზღვრეთ სტუდენტების აზრი კაფეტერიაში საკვების ხარისხის შესახებ 100 სტუდენტისაგან შემდგარი შერჩევითი ერთობლიობის საფუძველზე
გ. კობოთი დაავადებულთა რაოდენობა აზიელ ქალთა შორის ქვეყნის მასშტაბით
დ. ჩაატარეთ კვლევა ახალი ავტომობილის ხარისხის რეიტინგის გამოსავლენად 1,000 ახალი მფლობელის გამოკითხვით
ე. სხვადასხვა მიმართულების ბაკალავრების (მაგ. სოციოლოგია, ფსიქოლოგია, ინგლისური ენა) საშუალო GPA
ვ. აშშ-ში იმიგრირებული სამხრეთ აზიელების რიცხოვნობის ცვლილება 2005-2010 წლებში

7. ჩამოაყალიბეთ სამი სოციალური პრობლემა, რომლებიც სტატისტიკის გამოყენებით უნდა იქნეს შესწავლილი. (ამის ერთ-ერთი მაგალითია სიძულვილის ნიადაგზე ჩადენილი სასტიკი დანაშაულები). რომელი პრობლემის შესწავლაა უფრო ძნელი? რომელია უფრო ადვილი? რატომ?
8. მოიყვანეთ პოლიტიკური აქტივობის მაგალითები ნომინალურ, ორდინალურ და ინტერვალურ დონეებზე (შეგიძლიათ გამოიყენოთ ისეთი ქცევები, როგორიცაა ხმის მიცემის სიხშირე და პოლიტიკური პარტიის წევრობა). ჩამოთვალეთ თითოეულის ძლიერი და სუსტი მხარეები.
9. ცვლადები შეიძლება გავზომოთ ერთი ან რამდენიმე საზომით. ქვემოთ მოყვანილი ცვლადებისათვის დაასახელეთ ერთზე მეტი საზომი. არის თუ არა ერთი საზომი მეორეზე უკეთესი? ახსენით.

ა. ინდივიდის ასაკი
ბ. წლიური შემოსავალი
გ. რელიგიურობა
დ. სტუდენტის ქცევა
ე. სოციალური კლასი
ვ. დამოკიდებულება პოზიტიური დისკრიმინაციის მიმართ

თავი 2

ინფორმაციის თავმოყრა
სიხშირული განაწილებები

თავის სწავლების მიზნები

- როგორ გავიანგარიშოთ და გავაანალიზოთ სიხშირული, პროცენტული და კუმულაციური განაწილებები
- როგორ გავიანგარიშოთ კოეფიციენტები და პროცენტები
- განსხვავებები ნომინალური, ორდინალური და ინტერვალურ-ფარდობითი ცვლადების სიხშირულ განაწილებებში
- სტატისტიკური ცხრილების წაკითხვა კვლევითი ხასიათის ლიტერატურიდან

სოციალური სფეროს მეცნიერებს ხშირად უწევთ მონაცემთა დიდი მოცულობის დამუშავება. მაგალითად, ჩვეულებრივ, გამოკვლევის ბოლოს ათასობით ინდივიდუალური პასუხი გვიგროვდება, რომელიც ციფრების გროვის სახით არის წარმოდგენილი. იმისათვის, რომ რაიმე აზრი გამოვიტანოთ აქედან, საჭიროა მონაცემების თავმოყრა და განზოგადება, ანუ მათი სისტემატიზაცია. ამ შემთხვევაში მთავარი მეთოდი დაკვირვებების დაჯგუფებაა სიხშირული განაწილებებით. **სიხშირული განაწილება** ეს არის ცხრილი, რომელიც ასახავს საანალიზო ცვლადის თითოეულ კატეგორიაში მოხვედრილი დაკვირვებების რაოდენობას. სიხშირული განაწილების აგება მონაცემთა სტატისტიკური ანალიზის პირველი საფეხურია.

სიხშირული განაწილება ცხრილის სახით წარმოდგენილი ცვლადის თითოეულ კატეგორიაში მოხვედრილი დაკვირვებების რაოდენობა.

სიხშირული განაწილებები

2006 და 2007 წლებში აშშ-ის კონგრესში ბუშის ადმინისტრაციის მიერ საიმიგრაციო კანონმდებლობის ქარ-ცეცხლში გატარებამ იმიგრანტთა სტატუსის შესახებ, მთელი ქვეყნის მასშტაბით დებატები გამოიწვია. გლობალიზაციამ ბიძგი მისცა მიგრაციის ინტენსივობას, განსაკუთრებით 21-ე საუკუნის დასაწყისში. აზღვარგარეთიდან სამუშაო ძალის მიგრაციის გაფართოება სამუშაო ადგილისა და უკეთესი ცხოვრების პირობების ძებნამ განაპირობა. მონაცემები მოწმობენ, რომ აშშ მრავალი მიგრანტისათვის მუდმივი საცხოვრებელი ადგილი გახდა.[1] ტერმინს „საზღვარგარეთ დაბადებულები“ აშშ-ის ცენზების ბიურო მოსახლეობის იმ კატეგორიის აღსანიშნავად იყენებს, ვინც დაბადებისას აშშ-ის მოქალაქე არ ყოფილა. ცენზების ბიუროს გამოთვლებით, აშშ-ის მოსახლეობის დაახლოებით 12%, ანუ 37 მილიონი ადამიანი საზღვარგარეთაა დაბადებული.[2] დღეისათვის, აშშ-ის ყოველი რვა მოსახლიდან ერთი საზღვარგარეთ არის დაბადებული[3]. იმიგრანტები ერთ ჰომოგენურ ჯგუფს არ ქმნიან, აქ მრავალი განსხვავებული ჯგუფია. ცხრილში 2.1 ნაჩვენებია აშშ-ის მოსახლეობაში საზღვარგარეთ დაბადებულთა სიხშირული განაწილება მსოფლიოს მასშტაბით.

ცხრილი 2.1 მსოფლიოს რეგიონების მიხედვით აშშ-ში მცხოვრები საზღვარგარეთ დაბადებული მოსახლეობის სიხშირული განაწილება 2008 წელს

დაბადების ადგილი მსოფლიოს რეგიონების მიხედვით	სიხშირე (f)
ევროპა	4,644,000
აზია	9,978,000
ლათინური ამერიკა	20,034,000
სხვა დანარჩენი	2,034,000
სულ (N)	37,258,000

წყარო: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 43.

დააკვირდით, რომ სიხშირული განაწილება ცხრილის სახით არის წარმოდგენილი, ცხრილს აქვს ნომერი (2.1) და აღწერითი სათაური. სათაურში მოცემულია, თუ რა სახის მონაცემებია წარმოდგენილი ცხრილში: „მსოფლიოს რეგიონების მიხედვით აშშ-ში მცხოვრები საზღვარგარეთ დაბადებული მოსახლეობის სიხშირული განაწილება 2008 წელს“. ცხრილი შედგება ორი სვეტისაგან. პირველ სვეტში მოცემულია ცვლადი (დაბადების ადგილი მსოფლიოს რეგიონების მიხედვით) და მისი კატეგორიები. მეორე სვეტი, რომლის სათაურია „სიხშირე (f)“ გვიჩვენებს თითოეულ კატეგორიაში მოხვედრილი დაკვირვებების რაოდენობას ($N = 37,258,000$). დააკვირდით აგრეთვე ცხრილის ბოლოს მკაფიოდ მითითებულ საინფორმაციო წყაროს. ის გვიჩვენებს, რომ მონაცემები მოტანილია აშშ-ის

აღწერის ბიუროს 2010 წლის ანგარიშიდან. ზოგადად, საინფორმაციო წყაროს მითითება ყველგან საჭიროა, სადაც მონაცემებია მოტანილი.

რა ინფორმაცია მივიღეთ ცხრილი 2.1-დან? ცხრილიდან ჩანს, რომ 2008 წლისათვის დაახლოებით 37 მილიონი ადამიანი იყო კლასიფიცირებული საზღვარგარეთ დაბადებულთა ჯგუფში. ჯგუფში უმრავლესობას ლათინო-ამერიკელები (დაახლოებით 20 მლნ.) წარმოადგენდნენ, 10 მლნ-ს აზიელები და 4.6 მლნ-ს ევროპელები.

კოეფიციენტები და პროცენტები

სიხშირული განაწილება გვეხმარება ინფორმაციის წარმოდგენაში კომპაქტური ფორმით. როდესაც დაკვირვებათა რიცხვი ძალიან დიდია, სიხშირეების აღქმადობა ძნელდება. ამიტომ მათი სტანდარტიზაციის თვალსაზრისით, ნედლი სიხშირეები გადაგვყავს ფარდობით (რელატიურ) სიხშირეებში, ანუ კოეფიციენტებსა და პროცენტებში.

კოეფიციენტი არის ფარდობითი სიხშირე, რომელიც მიიღება თითოეული კატეგორიის სიხშირის ფარდობით დაკვირვებათა რიცხვზე. იმისათვის, რომ მივიღოთ კოეფიციენტი (p), თითოეული კატეგორიის სიხშირე (f) უნდა გავყოთ დაკვირვებათა რიცხვზე (N):

$$p = \frac{f}{N} \qquad (2.1)$$

სადაც

f = სიხშირე

N = დაკვირვებათა რიცხვი

აქედან, ლათინურ ამერიკაში დაბადებულთა კოეფიციენტი იქნება:

$$\frac{20,034,000}{37,258,000} = 0.54$$

აზიაში დაბადებულთა კოეფიციენტია:

$$\frac{9,978,000}{37,258,000} = 0.27$$

ევროპაში დაბადებულთა კოეფიციენტი:

$$\frac{4,644,000}{37,258,000} = 0.12$$

და ბოლოს, სხვა ადგილებში დაბადებულთა კოეფიციენტია:

$$\frac{2,602,000}{37,258,000} = 0.07$$

კოეფიციენტების ჯამი ყოველთვის 1.00-ის ტოლია (დასაშვებია დამრგვალების გამო მცირე უზუსტობა). ამიტომ, ჩვენ მაგალითში კოეფიციენტების ჯამი ტოლია:

$$0.54 + 0.27 + 0.12 + 0.07 = 1.00$$

კოეფიციენტიდან სიხშირის საპოვნელად საჭიროა კოეფიციენტის გადამრავლება დაკვირვებათა რიცხვზე:

$$f = p(N) \tag{2.2}$$

აქედან, აზიაში დაბადებულთა სიხშირე ტოლი იქნება:

$$0.27(37,258,000) = 10,059,660$$

აღსანიშნავია, რომ მიღებული სიხშირე ცოტა განსხვავდება რეალური სიხშირისაგან – 9,978,000. ეს განსხვავება გამოიწვია კოეფიციენტის დამრგვალებამ. თუ დამრგვალებულის ნაცვლად თავდაპირველ კოეფიციენტს გამოვიყენებთ, მივიღებთ სიხშირის სწორ მნიშვნელობას:

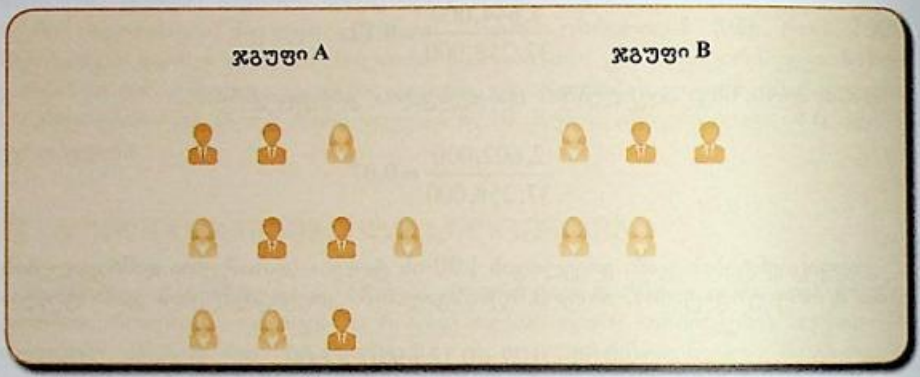
$$0.0267808255945032(37,258,000) = 9,978,000$$

კოეფიციენტი ფარდობითი სიხშირე, რომელიც მიიღება თითოეული კატეგორიის სიხშირის ფარდობით დაკვირვებათა რიცხვზე.

ნახ. 2.1 შეადარეთ ჯგუფი A ჯგუფ B-ს და გაეცით პასუხი შემდეგ კითხვებს: რომელ ჯგუფშია ქალების მეტი რაოდენობა? რომელ ჯგუფშია უფრო მაღალი ქალების კოეფიციენტი?

✓ ცოდნი
შემოწმება

ნახაზი 2.1 რიცხვები და კოეფიციენტები



კოეფიციენტების გამოსახვა შესაძლებელია აგრეთვე პროცენტების სახით. **პროცენტი** არის ფარდობითი სიხშირე, რომელიც მიიღება თითოეული კატეგორიის სიხშირის ფარდობით დაკვირვებათა რიცხვზე და 100-ზე გამრავლებით. სტატისტიკურ ანგარიშებში კოეფიციენტები ძირითადად პროცენტებშია წარმოდგენილი. პროცენტი გამოსახავს სიხშირის ზომას 100-თან მიმართებით.

პროცენტის გასაანგარიშებლად, კოეფიციენტი მრავლდება 100-ზე:

$$\text{პროცენტი } (\%) = \frac{f}{N}(100) \tag{2.3}$$

ან

$$\text{პროცენტი } (\%) = p\,(100) \tag{2.4}$$

ამგვარად, რესპონდენტთა ხვედრითი წილი (პროცენტი), რომლებიც წარმოშობით აზიიდან არიან, იქნება:

$$0.27\,(100) = 27\%$$

რესპონდენტთა ხვედრითი წილი, რომლებიც წარმოშობით ლათინური ამერიკიდან არიან:

$$0.54\,(100) = 54\%$$

პროცენტი ფარდობითი სიხშირე, რომელიც მიიღება თითოეული კატეგორიის სიხშირის ფარდობით დაკვირვებათა რიცხვზე და 100-ზე გადამრავლებით.

გაიანგარიშეთ ქალებისა და მამაკაცების კოეფიციენტები სტატისტიკის ჯგუფში. როგორი კოეფიციენტი აქვთ ქალებს?

✓ *ცოდნის შემოწმება*

ხვედრითი წილების (პროცენტული) განაწილება

ხვედრითი წილები პროცენტული განაწილებებია. **ხვედრითი წილების (პროცენტული)** განაწილება წარმოადგენს ცხრილს, რომელშიც ასახულია ცვლადის თითოეული კატეგორიაში მოხვედრილი დაკვირვებების ხვედრითი წილი. მაგალითად, ცხრილში 2.2 ნაჩვენებია საზღვარგარეთ დაბადებული მოსახლეობის სიხშირული განაწილება 2008 წელს (ცხრილი 2.1) შესაბამის პროცენტულ განაწილებებთან ერთად. ხვედრითი წილების (ან კოეფიციენტების) განაწილებაში ყოველთვის უნდა ჩანდეს შესადარი ბაზა (N), რომლის მიმართაც ვახდენთ გაანგარიშებებს. აქედან, ცხრილში 2.2 ბაზა, რომლის მიმართაც ვახდენთ გაანგარიშებებს ტოლია $N = 37,258,000$.

ცხრილი 2.2 საზღვარგარეთ დაბადებული მოსახლეობის სიხშირული და პროცენტული განაწილება 2008 წელს

დაბადების ადგილი მსოფლიოს რეგიონების მიხედვით	სიხშირე (f)	პროცენტი
ევროპა	4,644,000	12.5
აზია	9,978,000	26.8
ლათინური ამერიკა	20,034,000	53.8
სხვა დანარჩენი	2,034,000	7.0
სულ (*N*)	37,258,000	100.1

წყარო: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 43.

პროცენტული (ხვედრითი წილების) განაწილება ცხრილის სახით წარმოდგენილი ცვლადის თითოეულ კატეგორიაში მოხვედრილი დაკვირვებების ხვედრითი წილები.

შედარებები

ცხრილში 2.2, ვაჩვენეთ, რომ აშშ-ში ცხოვრობს საზღვარგარეთ დაბადებულთა ოთხი ძირითადი ჯგუფი. ჯგუფების გამოყოფა გვეხმარება თითოეული ჯგუფის სპეციფიკური მახასიათებლებისა და წარმოშობის დადგენაში. ჩვენ უნდა ვძლიოთ უცხოეთში დაბადებულების მხოლოდ ერთ ჯგუფში განაწილების ცდუნებას და დავსვათ კითხვები: რომელ ჯგუფში უფრო მაღალია განათლების დონე? ერთ ჯგუფში უფრო მაღალია ახალგაზრდების ხვედრითი წილი, თუ მეორეში?

ჯგუფების განხილვა შესაძლებელია სათითაოდ ან ამორჩევით, ეს ინფორმაციის მოცულობასა და კვლევის კითხვაზეა დამოკიდებული.

მაგალითად, ჩვენთვის ცნობილია, რომ 2008 წელს საზღვარგარეთ დაბადებული იმიგრანტების 16.5% სიღარიბეში ცხოვრობდა.[4] საზღვარგარეთ დაბადებულთა შორის სიღარიბის ყველაზე მაღალი დონე იყო ლათინო-ამერიკელებს შორის (21%), ხოლო ყველაზე დაბალი – ევროპელებს შორის (9%). რას გვეუბნება ეს ციფრები საზღვარგარეთ დაბადებულთა დემოგრაფიული მახასიათებლების შესახებ? ლათინო-ამერიკელების შესახებ? ევროპელების შესახებ? იმისათვის, რომ პასუხი გავცეთ ამ კითხვებს და გავარკვიოთ, აქვს თუ არა ამ ორ რეგიონს მნიშვნელოვნად განსხვავებული სოციალური მახასიათებლები, ისინი უნდა შევადაროთ.

სტუდენტები, სოციალური სფეროს მეცნიერები ან უბრალოდ მომხმარებლები, ხშირად აწყდებიან პრობლემებს, რომელთა გადაწყვეტა სანდო და დამაჯერებელ შედარებებს მოითხოვს. მაგალითად, 2008 წელს ბავშვების 22.8% ცხოვრობდა მხოლოდ დედასთან.5 დიდია თუ მცირე ეს ციფრი? 2008 წელს 18 და ზევით ასაკის მოსახლეობაში იმ პირთა ხვედრითი წილი, რომლებიც ქორწინებაში არასოდეს არ ყოფილან, 26%-ს შეადგენდა.6 ასახავს თუ არა ეს მაჩვენებლები ცვლილებებს ამერიკულ ოჯახებში? იმისათვის, რომ პასუხი გავცეთ კითხვებს და რაიმე დასკვნამდე მივიდეთ, თითოეულ ამ შემთხვევაში საჭიროა შესადარი ინფორმაცია.

განხილული მაგალითები გვიჩვენებს ასევე შედარების რამდენიმე გზას. ის, თუ რა უნდა შევადაროთ, დიდადაა დამოკიდებული კითხვაზე, რომელსაც ვსვამთ. მკაფიოდ ჩამოყალიბებული კითხვის გარეშე რთულია შედარების ტიპის შერჩევა. სოციალურ მეცნიერებებში რამდენიმე ტიპის შედარება გამოიყენება. ერთ-ერთია სხვადასხვა მახასიათებლებით შედგენილ ჯგუფების შედარება – მაგალითად, მოხუცებისა და ახალგაზრდების, თეთრკანიანებისა და წყნარი ოკეანეთის აზიელების, ან როგორც ჩვენ მაგალითშია, შედარებები საზღვარგარეთ დაბადებულთა სხვადასხვა კატეგორიებს შორის. ზოგჯერ ჩვენ გვაინტერესებს ჯგუფების შიგნით რეგიონული შედარებები, ან საზოგადოების სხვადასხვა სეგმენტის შედარება. თქვენ შეიძლება ახალ ამბებში წაგიკითხავთ ხმის მიცემის კონტრასტების შესახებ აშშ-ის ჩრდილოეთსა და სამხრეთში, ან ბინათმფლობელთა ხვედრითი წილების შესახებ ცენტრალურ ქალაქებსა და გარეუბნებში. ასევე საინტერესოა ერთიდაიმავე ჯგუფის შიგნით დროის რაიმე პერიოდში მომხდარი ცვლილებები, როგორიცაა უცხოეთში დაბადებულთა ხვედრითი წილების ცვლილება უკანასკნელ ათწლეულში, ან არის თუ არა ცვლილებები მოსახლეობის გადანაცვლებაში ქალაქებიდან გარეუბნებისკენ?

სტატისტიკა პრაქტიკაში: სამუშაო ძალის აქტივობა ლათინო-ამერიკულ მოსახლეობაში

ძალიან ხშირად, როდესაც სხვადასხვა ზომის ორ ჯგუფს ერთმანეთს ვადარებთ, პროცენტების გამოყენება განსაკუთრებით სასარგებლოა. მაგალითად, ჩვენ ვიცით, რომ მოსახლეობის სოციალ-ეკონომიკურ სტატუსში არსებული განსხვავებები მათ განსხვავებულ ეკონომიკურ შესაძლებლობებზე მიანიშნებს. შრომითი აქტივობა (დასაქმება ან სამუშაოს ძებნა) ეკონომიკური შესაძლებლობების მისაწვდომობის მნიშვნელოვანი ინდიკატორია და უშუალო კავშირშია სოციალ-ეკონომიკურ სტატუსთან. ცხრილში 2.3 მოცემულია *სამუშაო ძალის აქტივობის* ნედლი სიხშირული განაწილებები ლათინური (ესპანური წარმოშობის) მოსახლეობის სამივე კატეგორიის მიხედვით.

ცხრილი 2.3 მოსახლეობის სტატუსი დასაქმებაში, ესპანური წარმოშობის კატეგორიებისათვის, 2008 წელი

სტატუსი დასაქმებაში	მექსიკელები	პუერტო-რიკოელები	კუბელები
დასაქმებული	12,931,000	1,634,000	841,000
უმუშევარი	1,078,000	188,000	57,000
ეკონომიკურად არააქტიური	6,465,000	1,032,000	525,000
სულ (*N*)	20,474,000	2,854,000	1,423,000

წყარო: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 39.

რომელ ჯგუფშია თავმოყრილი სამუშაო ძალის გარეთ მყოფი მოსახლეობის დიდი რაოდენობა? რადგან სამივე ჯგუფში მოსახლეობის სხვადასხვა რაოდენობაა, ნედლი სიხშირეებით ძნელია ამ კითხვაზე პასუხის გაცემა. მეტი დამაჯერებლობისათვის უნდა შევადაროთ თითოეული ჯგუფის პროცენტული სიხშირეები. აღნიშნული წარმოდგენილია ცხრილში 2.4. დაიმახსოვრეთ, რომ პროცენტული განაწილებები უფრო აადვილებს ჯგუფებს შორის არსებული განსხვავებების გამოვლენას. დასაქმებულთა ხვედრითი წილი უფრო მაღალია მექსიკელთა შორის (63%) პუერტო-რიკოელებთან (57%) და კუბელებთან (59%) შედარებით. ამ ჯგუფებს შორის, კუბელებს შორის ყველაზე დაბალია უმუშევართა ხვედრითი წილი (4%).

ცხრილი 2.4 მოსახლეობის სტატუსი დასაქმებაში, ესპანური წარმოშობის კატეგორიებისათვის, 2008 წელი

სტატუსი დასაქმებაში	მექსიკელები (%)	პუერტო-რიკოელები (%)	კუბელები (%)
დასაქმებული	63.2	57.2	59.1
უმუშევარი	5.3	6.6	4.0
ეკონომიკურად არააქტიური	31.6	36.2	36.9
სულ	100.1	100.0	100.0
(N)	20,474,000	2,854,000	1,423,000

წყარო: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 39.

✓ ცოდნის შემოწმება

ცხრილი 2.4-ის საფუძველზე უპასუხეთ შემდეგ კითხვებს: მექსიკელების რა პროცენტია დასაქმებული? რა არის ამ ხვედრითი წილების გაანგარიშების ბაზა (N)? კუბელების რამდენი პროცენტია ეკონომიკურად არააქტიური? დაასახელეთ ამ უკანასკნელის გაანგარიშების ბაზა (N).

როდესაც ერთ ჯგუფს მეორესთან ვადარებთ, ყველაზე მნიშვნელოვან დასკვნებს ფარდობითი სიხშირული განაწილებების საფუძველზე ვიღებთ. ფაქტია, რომ ჩვენთვის ნაკლებადსაინტერესოა მხოლოდ ცალკეული განაწილება. ყველაზე საინტერესო კითხვები სოციალურ მეცნიერებებში ორ ან მეტ ჯგუფს შორის არსებული განსხვავებების შესახებ დაისმება.[7] სამუშაო ძალის აქტივობასთან დაკავშირებით ლათინურ ჯგუფში აღმოჩენილმა განსხვავებებმა სამართლიანი კითხვა გააჩინა: არის კი ლათინური ჯგუფი ერთადერთი, შედარებით ჰომოგენური ეთნიკური ჯგუფი? შემდგომმა ანალიზმა უნდა დაგვანახოს, თუ რატომ არის ლათინო-ამერიკელების განსხვავებული იდენტობა დაკავშირებული მათ განსხვავებულ ეკონომიკურ აქტივობასთან. ამ ტიპის კითხვებზე პასუხის გასაცემად უფრო რთული მულტივარიაციული მიდგომა უნდა გამოვიყენოთ, რომელიც ორზე მეტ ცვლადს მოიცავს. პროცენტული განაწილებები ამ რთული მეთოდების გამოყენების საფუძველს წარმოადგენს.

სანამ გავაგრძელებდეთ, გვინდა შეგახსენოთ, რომ კომპლექსური მონაცემების გამოყენებამდე, პირველ ხუთ თავში შემოთავაზებული მეთოდები მხოლოდ დროებითი დასკვნების ჩამოყალიბების საშუალებას იძლევა ჯგუფებს შორის არსებულ განსხვავებებზე. უფრო დამაჯერებელი, საფუძვლიანი შედარებებისათვის ჩვენ უნდა განვიხილოთ შერჩევის მეთოდები და დასკვნითი სტატისტიკა, რაც მოცემული იქნება თავებში 6-დან 14-ის ჩათვლით. თუ გულმოდგინედ შეისწავლით

სახელმძღვანელოს თავებს და კარგად აითვისებთ სტატისტიკურ კონცეფციებს საფუძვლიანი დასკვნების ჩამოსაყალიბებლად, სამომავლოდ მონაცემების უფრო კომპლექსურ ინტერპრეტირებასაც შეძლებთ.

სიხშირული განაწილებების აგება

ჩვენ უკვე შევისწავლეთ სიხშირული განაწილების ზოგადი კონცეფცია და ვნახეთ, რომ მონაცემები შეიძლება გამოვსახოთ ნედლი სიხშირეებით, კოეფიციენტებით, ან ხვედრითი წილებით. ჩვენ აგრეთვე ვნახეთ, თუ როგორ უნდა გამოვიყენოთ პროცენტები განაწილებების შესადარებლად სხვადასხვა ჯგუფებში.

ამ ნაწილში თქვენ ისწავლით სიხშირული განაწილებების აგებას. რა თქმა უნდა, ეს შეიძლება კომპიუტერით გაკეთდეს, თუმცა მნიშვნელოვანია თქვენ თვითონ გაიაროთ მთელი ეს პროცესი.

ნომინალური და ორდინალური ცვლადებისათვის სიხშირული განაწილების აგება საკმაოდ მარტივია. დათვალეთ დაკვირვებათა რიცხვი თითოეული კატეგორიისათვის და ჩაინიშნეთ დაკვირვებათა საერთო რაოდენობასთან (N) ერთად. თვალსაჩინოებისთვის მივმართოთ 40 რესპონდენტისაგან შემდგარ შერჩევით ერთობლიობას, რომელიც ამოღებულია GSS-ის კვლევიდან და ჩავწეროთ ქულები შემდეგი მაჩვენებლების მიხედვით: სქესი, ნომინალური ცვლადი; განათლების დონე, ორდინალური ცვლადი, და ასაკი და ბავშვების რაოდენობა, – ორივე ინტერვალური ცვლადები. რესპონდენტების სქესი წარმოდგენილია GSS-ის შესაბამისად: „მამაკაცი“ და „ქალი“.

რესპონდენტების სქესი კვლევის დაწყებისას იქნა ჩანიშნული. შემდეგ რესპონდენტებს ჰკითხეს მათ მიერ მიღებული განათლების უმაღლესი ხარისხის შესახებ: საშუალო სკოლაზე დაბალი, საშუალო სკოლა, კოლეჯი, ბაკალავრის ხარისხი და სამეცნიერო ხარისხი. ამათგან პირველი აღნიშნავს ყველაზე დაბალ საფეხურს. რესპონდენტის ასაკის გაანგარიშება მისი დაბადების წლიდან მოხდა. ბავშვების რაოდენობა დადგინდა შემდეგი კითხვის საფუძველზე: „რამდენი ბავშვი გყოლიათ საერთოდ?“. რესპონდენტების პასუხები ასახულია ცხრილში 2.5. თითოეული სტრიქონი ასახავს რესპონდენტს, ხოლო სვეტი – ცვლადს. ეს ფორმატი ძალიან მოსახერხებელია სოციალურ მეცნიერებებში.

როგორც მიხვდით, თვალის ერთი გადავლებით საკმაოდ ძნელია ცხრილიდან რაიმე დასკვნის გამოტანა. რამდენი მამაკაცია რესპონდენტებს შორის? რამდენმა თქვა, რომ სამეცნიერო ხარისხი აქვს? რამდენი ადამიანი იყო 50 წელზე ასაკოვანი? კითხვებზე პასუხის გასაცემად ოთხივე ცვლადისთვის უნდა ავაგოთ სიხშირული განაწილებები.

სიხშირული განაწილებები ნომინალური ცვლადებისთვის

დავიწყოთ ნომინალური ცვლადიდან, *სქესი*. ჯერ დავთვალოთ მამაკაცები, შემდეგ ქალები (ილუსტრაციის მიზნით დათვლა გაკეთებულია ხაზებით და მოტა-

ნილია ცხრილში 2.6). შემდეგ დათვლის შედეგები გამოყენებულია იმავე ცხრილში სიხშირული განაწილების წარმოსადგენად. ცხრილს აქვს სათაური, რომელიც აღწერს მის შინაარსს (სიხშირეების განაწილება ცვლადისათვის სქესი: GSS-ის ქვეშერჩევა). კატეგორიები (მამაკაცები და ქალები) და მათი შესაბამისი სიხშირეები მკაფიოდ არის ჩამოწერილი. ამასთან ერთად, მოყვანილია დაკვირვებათა საერთო რიცხვი (*N*). პროცენტების სვეტში მოტანილია ამ ცვლადის პროცენტული განაწილებები. სიხშირეების სვეტის ხვედრით წილებად გარდაქმნისათვის, უბრალოდ საჭიროა თითოეული სიხშირის გაყოფა დაკვირვებათა საერთო რაოდენობაზე და შემდეგ 100-ზე გამრავლება. პროცენტული განაწილებები ყოველთვის თან ახლავს სიხშირის ცხრილებს და ეს განსაკუთრებით გამოსადეგია, თუ ჯგუფების შედარებას აპირებთ. მაშინვე მიხვდებით, თუ როგორ ადვილდება ინფორმაციის წაკითხვა. შერჩევაში 25 ქალია და 15 მამაკაცი. სიხშირულ განაწილებაზე დაყრდნობით, ჩვენ შეგვიძლია ვთქვათ, რომ შერჩევაში უმრავლესობა ქალია.

ცხრილი 2.5 GSS-ის 40 რესპონდენტისაგან შემდგარი ქვეშერჩევა

რესპონდენტის სქესი	ხარისხი	ბავშვების რაოდენობა	ასაკი
M	ბაკალავრი	1	43
F	საშუალო სკოლა	2	71
F	საშუალო სკოლა	0	71
M	საშუალო სკოლა	0	37
M	საშუალო სკოლა	0	28
F	საშუალო სკოლა	6	34
F	საშუალო სკოლა	4	69
F	სამეცნიერო	0	51
F	ბაკალავრი	0	76
M	სამეცნიერო	2	48
M	სამეცნიერო	0	49
M	საშუალო სკოლაზე დაბალი	3	62
F	საშუალო სკოლა	8	71
F	საშუალო სკოლა	1	32
F	საშუალო სკოლა	1	59
F	საშუალო სკოლა	1	71
M	ბაკალავრი	0	34
M	ბაკალავრი	0	39
F	კოლეჯი	2	50

M	საშუალო სკოლა	3	82
F	საშუალო სკოლა	1	45
M	საშუალო სკოლა	0	22
M	საშუალო სკოლა	2	40
F	საშუალო სკოლა	2	46
M	საშუალო სკოლა	0	29
F	საშუალო სკოლა	1	75
F	ბაკალავრი	0	23
M	ბაკალავრი	2	35
M	საშუალო სკოლა	3	44
F	საშუალო სკოლა	3	47
M	სამეცნიერო	1	84
F	საშუალო სკოლაზე დაბალი	1	45
F	სამეცნიერო	3	24
F	საშუალო სკოლაზე დაბალი	0	47
F	სამეცნიერო	5	67
F	საშუალო სკოლა	1	21
F	საშუალო სკოლა	0	24
F	საშუალო სკოლა	3	49
F	საშუალო სკოლა	3	45
F	სამეცნიერო	3	37

შენიშვნა: M-ით აღნიშნულია მამაკაცები, F-ით ქალები.

ცხრილი 2.6 სიხშირეების განაწილება ცვლადისათვის სქესი: GSS-ის ქვეშერჩევა

სქესი	თვლის ერთეულები	სიხშირე (f)	პროცენტები
მამაკაცი	𝍸 𝍸 𝍸	15	37.5
ქალი	𝍸 𝍸 𝍸 𝍸 𝍸	25	62.5
სულ (N)		40	100.0

ააგეთ სიხშირული და პროცენტული განაწილებები ქალებისა და მამაკაცებისათვის სტატისტიკის სალექციო ჯგუფში.

✓ ცოდნი შემოწმება

სიხშირული განაწილებები ორდინალური ცვლადებისთვის

ორდინალური ცვლადებისათვისაც ისევე აიგება სიხშირული განაწილება, როგორც ნომინალურის შემთხვევაში. ცხრილში 2.7 მოტანილია სიხშირული განაწილება ცვლადისთვის ხარისხი. ცხრილიდან ჩანს, რომ შერჩევაში უმრავლესობამ (60%) მიღებული განათლების ყველაზე მაღალ საფეხურად დაასახელა საშუალო სკოლის დიპლომი.

ცხრილი 2.7 სიხშირეების განაწილება ცვლადისათვის ხარისხი: GSS-ის ქვეშერჩევა

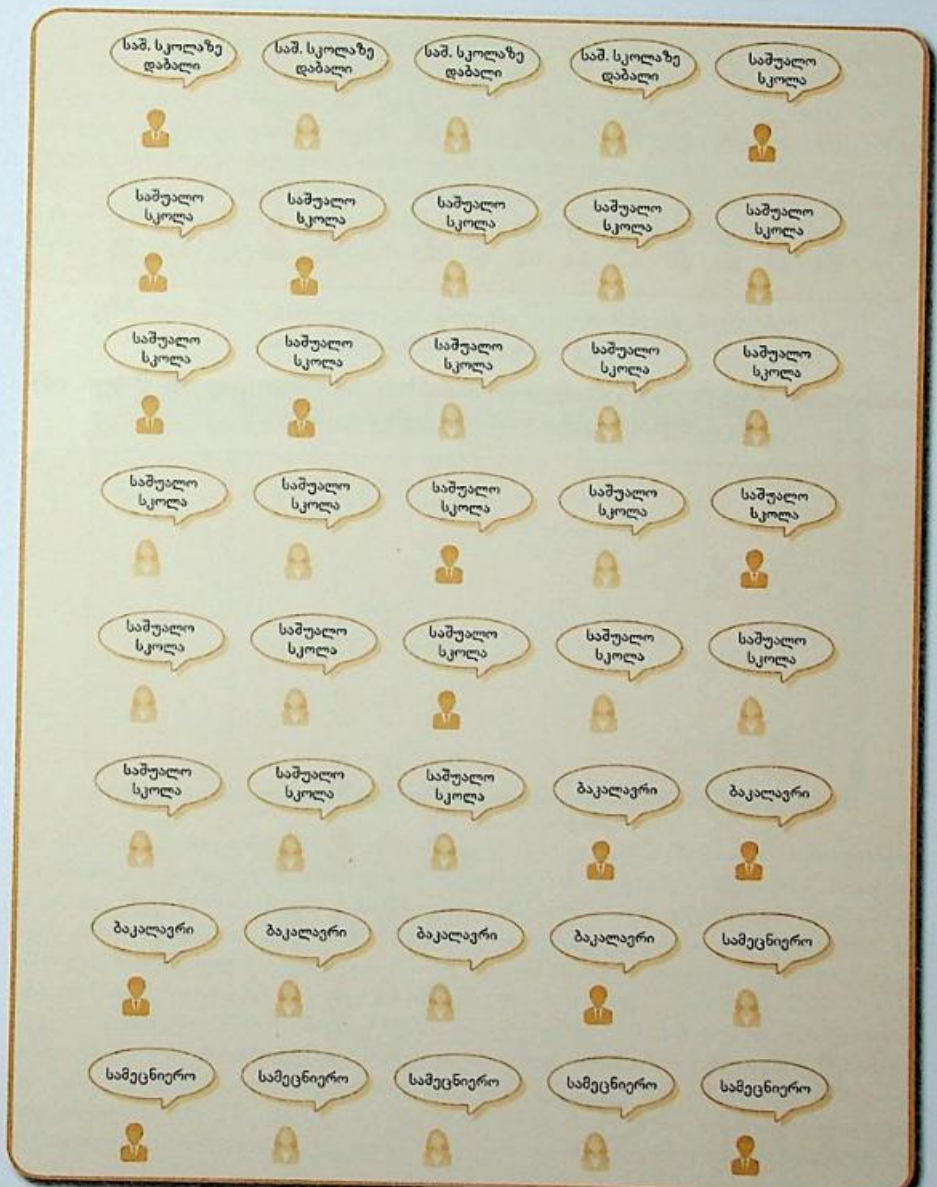
სქესი	თვლის ერთეულები	სიხშირე (*f*)	პროცენტები
საშუალო სკოლაზე დაბალი	𝍸	4	10.0
საშუალო სკოლა	𝍸 𝍸 𝍸 𝍸 𝍸	24	60.0
სამეცნიერო	𝍸	6	15.0
ბაკალავრი	𝍸	6	15.0
სულ (*N*)		40	100.0

ნომინალური და ორდინალური ცვლადების სიხშირული განაწილებების მთავარი განსხვავება ამ კატეგორიების ჩამონათვალშია, ანუ მათ თანმიმდევრობაში. ნომინალური ცვლადებისათვის ჩამონათვლის თანმიმდევრობას არავითარი მნიშვნელობა არ გააჩნია. მაგალითად, შეიძლებოდა თანმიმდევრობაში პირველად დაგვეწერა ქალები და მერე მამაკაცები და ეს არ შეცვლიდა განაწილების ბუნებას. ორდინალური ცვლადებისა და კატეგორიებისათვის კი, რომლებსაც მინიჭებული აქვთ რანგები, მნიშვნელოვანია მათი განლაგება თანმიმდევრობით, რანგების შესაბამისად – უდაბლესიდან უმაღლესისკენ, ან პირიქით. ამიტომ, მონაცემები ცხრილში 2.7 წარმოდგენილია ზრდადი თანმიმდევრობით „საშუალო სკოლაზე დაბლიდან“ – „სამეცნიერომდე“ (განათლების ყველაზე მაღალი კატეგორია).

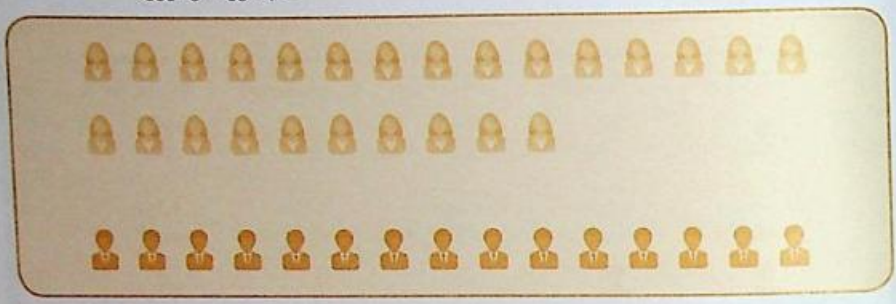
ცოდნის შოწმება

ნახაზები 2.2, 2.3 და 2.4 თვალსაჩინოდ ასახავს იმ საფეხურებს, რაც წარმოდგენილია ცხრილებში 2.5, 2.6 და 2.7. იმაში დასარწმუნებლად, თუ როგორ აიოლებს რესპონდენტების დათვლას დაჯგუფება სქესისა (ნახ.2.3) და ხარისხის (ნახ.2.4) მიხედვით, დაუბრუნდით კვლავ ნახაზს 2.2. და უპასუხეთ შემდეგ კითხვებს: რამდენი მამაკაცია ჯგუფში? რამდენი ქალია? რამდენმა გასცა პასუხი, რომ აქვს ბაკალავრის ხარისხი? შემდეგ მიუბრუნდით ნახაზს 2.3: რამდენი მამაკაცია ჯგუფში? რამდენი ქალია? და ბოლოს შეხედეთ ნახაზს 2.4: რამდენმა გასცა პასუხი, რომ აქვს ბაკალავრის ხარისხი?

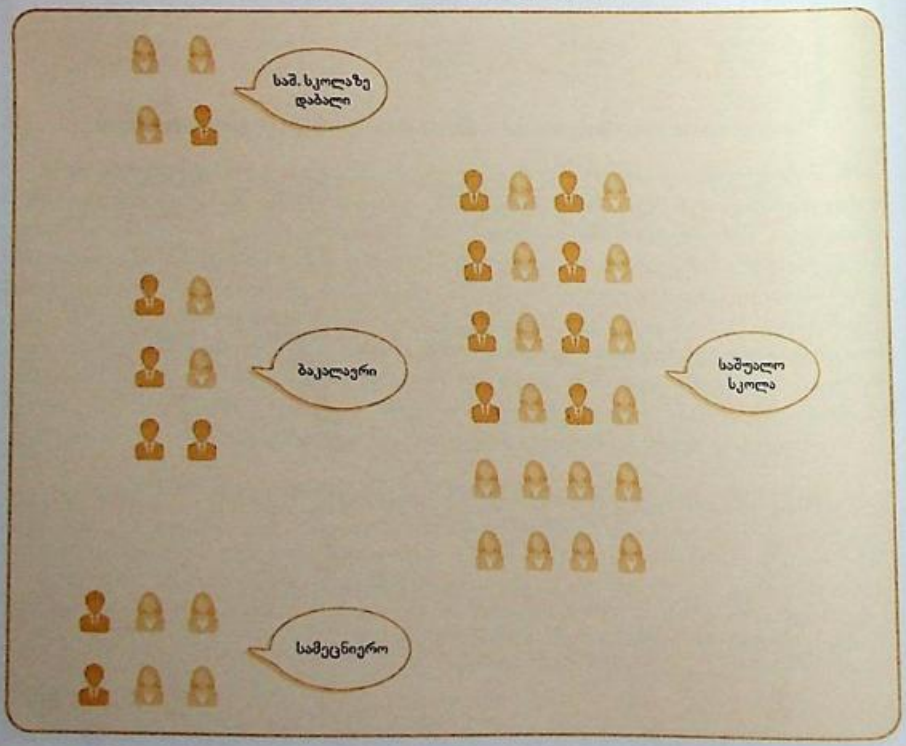
ნახაზი 2.2 ორმოცი რესპონდენტი GSS-ის ქვეშერჩევიდან, მათი სქესი და განათლების ხარისხი (იხ. ცხრილი 2.5)



ნახაზი 2.3 სქესის მიხედვით დაჯგუფებული ორმოცი რესპონდენტი GSS-ის ქვეშერჩევიდან



ნახაზი 2.4 სქესისა და განათლების ხარისხის მიხედვით დაჯგუფებული ორმოცი რესპონდენტი GSS-ის ქვეშერჩევიდან



სიხშირული განაწილებები ინტერვალურ-ფარდობითი ცვლადებისთვის

ვფიქრობთ, დაგვეთანხმებით, რომ სიხშირული განაწილებების აგება ნომინალური და ორდინალური ცვლადებისთვის ძალიან ადვილია. საჭიროა მხოლოდ კატეგორიების ჩამოთვლა და შემდეგ თითოეული კატეგორიის ჯგუფში მოხვედრილი დაკვირვებების დათვლა. ასევე ადვილია შედარებით მცირე რაოდენობის ცვლადების ინტერვალური სიხშირული განაწილების აგებაც. მაგალითად, ოჯახში ბავშვების რაოდენობის სიხშირული განაწილების აგებისათვის საკმარისია ჩამოვწეროთ ბავშვების რაოდენობები და გვერდზე მივუწეროთ მათი სიხშირეები, როგორც ეს ნაჩვენებია ცხრილში 2.8.

ცხრილი 2.8 სიხშირული განაწილება ცვლადისთვის ბავშვების რაოდენობა: GSS-ის ქვეშერჩევა

ბავშვების რაოდენობა	სიხშირე (f)	ხვედრითი წილი (%)
0	13	32.5
1	9	22.5
2	6	15.0
3	8	20.0
4	1	2.5
5	1	2.5
6	1	2.5
7+	1	2.5
სულ (N)	40	100.0

ძალიან ხშირად ინტერვალურ ცვლადებს ფართო დიაპაზონის მნიშვნელობები გააჩნიათ და მარტივი სიხშირული განაწილებების წაკითხვა ძნელდება. მაგალითად, თუ შევხედავთ ასაკის განაწილებას ცხრილში 2.9, დავინახავთ, რომ ასაკს აქვს მნიშვნელობები 21-დან 84 წლამდე. მოხერხებულობისთვის მონაცემთა ეს გრძელი ჩამონათვალი ჯგუფების მცირე რაოდენობამდე უნდა შევამციროთ, სადაც ასაკის მნიშვნელობები ჯგუფებში იქნება თავმოყრილი. ცხრილში 2.10 მოცემულია ცხრილი 2.9-ის მონაცემების დაჯგუფებული სიხშირული განაწილება. თითოეული ჯგუფი, რომელსაც *ინტერვალს* უწოდებენ, 1-ის ნაცვლად 10 შესაძლო მნიშვნელობას შეიცავს. მაგალითად, ასაკები 21, 22, 23, 24, 28 და 29 მოხვდება ინტერვალში 20–29. მეორე სვეტი (სიხშირე) მიანიშნებს იმაზე, თუ რამდენი რესპონდენტი მოხვდა ამ ჯგუფში. მაგალითად, ინტერვალურ ჯგუფში 20–29

მოხვდა 7 რესპონდენტი. როდესაც დავაჯგუფებთ ასაკებს, ვნახავთ, რომ ყველაზე დიდი ინტერვალური ჯგუფია 40-49 (აქ მოხვდა 40-დან 12 ასაკი, ანუ შერჩევის 30%). ინტერვალური ჯგუფების ხვედრითი წილები, რომლებიც დავამატეთ ცხრილს 2.10, უფრო უკეთ ასახავს ასაკობრივ განაწილებას.

რა თქმა უნდა, გადაწყვეტილებას ჯგუფების რაოდენობისა და ინტერვალების სიდიდის შესახებ მკვლევარი იღებს და ეს დამოკიდებულია მისი კვლევის მიზანზე. თუმცა გასათვალისწინებელია, რომ ინტერვალი საკმარისად, მაგრამ არც ძალიან დიდი უნდა იყოს: ჩვენ უნდა შევძლოთ, რომ თავი ავარიდოთ კატეგორიების მრავალ მნიშვნელობას, მაგრამ არ უნდა გამოგვეპაროს დაკვირვებებს შორის მნიშვნელოვანი განსხვავებები.[§] ცხადია, ინტერვალების რაოდენობა თითოეული ინტერვალის სიდიდეზეა დამოკიდებული. მაგალითად, თუ თქვენ მუშაობთ ცვლადზე, რომლის მნიშვნელობები იცვლება 10-დან 60-მდე და ინტერვალის სიდიდედ მიიჩნევთ 10-ს, გექნებათ 5 ინტერვალი.

ცხრილი 2.9 სიხშირული განაწილება ცვლადისთვის ასაკი: GSS-ის ქვეშერჩევა

რესპონდენტების ასაკი	სიხშირე (f)	რესპონდენტების ასაკი	სიხშირე (f)
21	1	59	1
22	1	62	1
23	1	67	1
24	2	69	1
28	1	71	4
29	1	75	1
32	1	76	1
34	2	82	1
35	1	84	1
37	2		
39	1		
40	1		
43	1		
44	1		
45	3		
46	1		
47	2		
48	1		
49	2		
50	1		
51	1		

შეგიძლიათ თუ არა დაამტკიცოთ, რომ ცხრილი 2.10 სწორად არის აგებული? გამოიყენეთ ცხრილი 2.9, რათა განსაზღვროთ დაკვირვებების სიხშირეები, რომლებიც მოხვდნენ ცხრილი 2.10-ის ინტერვალებში.

✓ *ცოდნის შემოწმება*

ცხრილი 2.10 დაჯგუფებული სიხშირული განაწილება ცვლადისთვის ასაკი: GSS-ის ქვეშერჩევა

რესპონდენტების ასაკი	სიხშირე (f)	ჯგუფური წილები (%)
20-29	7	17.5
30-39	7	17.5
40-49	12	30.0
50-59	3	7.5
60-69	3	7.5
70-79	6	15.0
80-89	2	5.0
სულ (N)	40	100.0

თუ თქვენ გარკვეული სიძნელები გექმნებათ ნომინალური, ორდინალური და ინტერვალური ცვლადების გარჩევაში, გირჩევთ დაუბრუნდეთ პირველ თავს და თავიდან განიხილოთ ისინი. გაზომვის ამ დონეების გარჩევა ძალიან მნიშვნელოვანია.

✓ *ცოდნის შემოწმება*

კუმულაციური განაწილებები

ხანდახან, ჩვენ შეიძლება დაგვაინტერესოს მთლიან განაწილებაში რომელიმე ქულის ფარდობითმა მდგომარეობამ. მაგალითად, გვინდა გავარკვიოთ, ერთობლიობიდან რა რაოდენობა ან პროცენტია 40 წელზე ახალგაზრდა, ან 60 წელზე ხნიერი. ამისათვის, განაწილების სიხშირეები შესაძლებელია კუმულაციური სახით წარმოვადგინოთ. **კუმულაციური განაწილება** საშუალებას გვაძლევს წარმოვადგინოთ სიხშირეები ცვლადის თითოეული კატეგორიისა და მისი წინა სიხშირეების საშუალებით.

კუმულაციური სიხშირული განაწილება სიხშირეებს გვიჩვენებს ცვლადის თითოეული კატეგორიისა (ინტერვალის ან ქულის) და მისი წინა სიხშირეების საშუალებით.

კუმულაციური სიხშირეები გამოიყენება მხოლოდ ორდინალურ ცვლადებსა და გაზომვის უფრო მაღალი დონეზე. ისინი მიიღება თითოეული კატეგორიის სიხშირისათვის მისი წინა სიხშირეების დამატებით.

დააკვირდით ცხრილს 2.11, სადაც ნაჩვენებია ცხრილი 2.10-ის ბაზაზე მიღებული კუმულაციური სიხშირეები. კუმულაციური სიხშირეების სვეტი, რომელიც აღნიშნულია *Cf*-ით, გვიჩვენებს ადამიანების რაოდენობას თითოეული ინტერვალში და მის წინ. ჩვენ ვხედავთ, რომ 40 რეპონდენტიდან 14 იყო 39 წლისა ან უფრო ახალგაზრდა, ხოლო 29 რესპონდენტი – 59 წლისა და ან უფრო ახალგაზრდა.

კუმულაციური სიხშირული განაწილების ასაგებად უნდა დავიწყოთ ყველაზე დაბალი ინტერვალის (ან ყველაზე დაბალი ქულიდან, თუ მონაცემები არ არის დაჯგუფებული) სიხშირიდან და დავუმატოთ მას მომდევნო, უფრო მაღალი ინტერვალის სიხშირეები და ასე გავაგრძელოთ, სანამ არ მივალთ ბოლო ინტერვალამდე. ბოლო ინტერვალის სიხშირე დაკვირვებათა მთლიანი რაოდენობის (N) ტოლია. ცხრილში 2.11, პირველი ინტერვალის (20-29) სიხშირეა 7. მისი კუმულაციური სიხშირეც 7-ის ტოლია, რადგან მის წინ სხვა ინტერვალი არ არის. მეორე ინტერვალის სიხშირეც 7-ია. კუმულაციური სიხშირე მოცემული ინტერვალისთვის არის 7 + 7 = 14. იმისათვის, რომ მივიღოთ მესამე ინტერვალის კუმულაციური სიხშირე (26), მის სიხშირეს ვუმატებთ მეორე ინტერვალის

ცხრილი 2.11 დაჯგუფებული სიხშირული განაწილება და კუმულაციური სიხშირები ცვლადისთვის ასაკი: GSS-ის ქვეშერჩევა

რესპონდენტების ასაკი	სიხშირე (f)	*Cf*
20-29	7	7
30-39	7	14
40-49	12	26
50-59	3	29
60-69	3	32
70-79	6	38
80-89	2	40
სულ (N)	40	

კუმულაციურ სიხშირეს (14). ამ პროცესს ვაგრძელებთ იქამდე, სანამ არ მივალთ ბოლო ინტერვალამდე. ამიტომ, ბოლო ინტერვალის კუმულაციური სიხშირე 40-ის, ანუ დაკვირვებათა საერთო რაოდენობის ტოლია.

ჩვენ აგრეთვე შეგვიძლია პროცენტული კუმულაციური განაწილების (*C %*) აგებაც, რომელსაც უფრო ფართო გამოყენება აქვს კუმულაციურ სიხშირულ

განაწილებებთან (*Cf*) შედარებით. პროცენტული კუმულაციური განაწილება გვიჩვენებს ხვედრით წილებს ცვლადის თითოეული კატეგორიისა (ინტერვალის ან ქულის) და მისი წინა ხვედრითი წილების საშუალებით. პროცენტული კუმულაციური განაწილების ასაგებად იგივე პროცედურებია საჭირო, რაც სიხშირული კუმულაციური განაწილებისათვის, მხოლოდ იმ განსხვავებით, რომ სიხშირეების ნაცვლად ხვედრით წილებს ვიყენებთ.

დეტალების განხილვა 2.1
რეალური ზღვრები, დაწესებული ზღვრები და ინტერვალის შუა წერტილები

ინტერვალები, რომლებიც ჩვენ გამოვიყენეთ ცხრილში 2.10, გარკვეულ ჩარჩოში ათავსებს რესპონდენტების კვლევის ცვლადს ასაკი. პირველ თავში ხაზი გაეუსვით იმ პირობას, რომ ჩვენი ცვლადები ამომწურავი და სრულიად ექსკლუზიური უნდა იყოს. ეს პირობები ვრცელდება აგრეთვე ინტერვალებზეც, რაც ნიშნავს, რომ 40-დან ყოველი რესპონდენტი ერთ და მხოლოდ ერთ ინტერვალში უნდა მოვათავსოთ. ამასთან, უნდა მოვახერხოთ ცვლადის ყველა არსებული მნიშვნელობის დაჯგუფება.

ჩვენ მაგალითში ეს პირობები დაცულია: თითოეული დაკვირვების მნიშვნელობა მხოლოდ ერთ ინტერვალშია მოათავსებული და ინტერვალებშიც ცხრილი 2.10-ის ყველა მნიშვნელობაა მოთავსებული. ამასთან, თუ კარგად დააკვირდებით ცხრილს 2.10, ნახავთ, რომ მოცემულ ინტერვალებს შორის დატოვებულია 1 წლის ტოლი განსხვავება, რაც პრობლემებს შეგვიქმნის ათწილადურ სიდიდეებთან. მოცემულობის თანახმად, ასაკები დამრგვალებულია, მაგრამ წარმოვიდგინოთ, რომ რესპონდენტების ასაკი უფრო ზუსტი სიდიდეებით არის მოცემული. რომელ ინტერვალში უნდა დავაჯგუფოთ რესპონდენტი, თუ მისი ასაკი 49.25 წელია? როგორც ვხედავთ, მისი ასაკი ორ ინტერვალს 40-49 და 50-59-ს შორისაა. ამ პრობლემის თავიდან ასაცილებლად უნდა გამოვიყენოთ არა ცხრილში 2.10 დაწესებული ზღვრები, არამედ რეალური ზღვრები.

რეალური ზღვრები ინტერვალის ქვედა და ზედა ზღვრებს 0.5-ით აფართოებს. მაშინ რეალური ზღვრები ინტერვალისათვის 40-49 იქნება 39.5-49.5; ინტერვალისათვის 50-59 იქნება 49.5-59.5 და ა.შ. (ქულები, რომლებიც ზუსტად დაემთხვევა ინტერვალის ქვედა და ზედა ზღვრებს, მაგ. 59,5 ან 49.5, უნდა დავამრგვალოთ უახლოეს ლუწ რიცხვამდე. ე.ი. 59.5 უნდა დამრგვალდეს 60-მდე და ამის გამო, ის მოხვდება ინტერვალში 59.5-69.5). ქვემოთ მოტანილ ცხრილში ჩვენ ვხედავთ როგორც დაწესებულ, ისე რეალურ ზღვრებს ასაკობრივი განაწილებისათვის. აბა, სად უნდა მოხვდეს რესპონდენტის ასაკი, თუ მან განაცხადა, რომ 49.25 წლისაა? (პასუხი: ინტერვალში 39.5-49.5). რას იტყვით 19.9-ზე? (ინტერვალში 19.5-29.5).

შუა წერტილი ერთადერთი რიცხვია, რომლითაც შეგვიძლია მთელი ინტერვალის წარმოდგენა. ის გაიანგარიშება ინტერვალის ქვედა და ზედა რეალური ზღვრების შეკრებით და 2-ზე გაყოფით. მაგალითად, ინტერვალის 19.5-29.5 შუა წერტილია (19.5 + 29.5) : 2 = 24.5. ცხრილის მესამე სვეტში მოცემულია ყველა ინტერვალის შუა წერტილები.

მიუხედავად იმისა, რომ დაჯგუფებული სიხშირული განაწილებები ძალიან მნიშვნელოვანია ინფორმაციის განზოგადებისთვის, ისინი მხოლოდ დაჯამებისათვის გამოიყენება და არ იძლევა დეტალებზე ყურადღების გამახვილების საშუალებას. მას შემდეგ, რაც მკვლევარებისა და სტუდენტებისათვის ხელმისაწვდომია კომპიუტერები, დაჯგუფებული სიხშირეები მხოლოდ მაშინ გამოიყენება, როდესაც ნედლი მონაცემები არ გვაქვს. შემდეგ თავებში აღწერილ სტატისტიკურ მეთოდებში ნედლ მონაცემებს გამოვიყენებთ.

რესპონდენტების ასაკი			
დაწესებული ზღვრები	რეალური ზღვრები	შუა წერტილი	სიხშირე (f)
20-29	19.5-29.5	25.4	7
30-39	29.5-39.5	34.5	7
40-49	39.5-49.5	44.5	12
50-59	49.5-59.5	54.5	3
60-69	59.5-69.5	54.5	3
70-79	69.5-79.5	74.5	6
80-89	79.5-89.5	84.5	2
სულ (N)		40	

პროცენტული კუმულაციური განაწილება გვიჩვენებს ხვედრით წილებს ცვლადის თითოეული კატეგორიისა (ინტერვალის ან ქულის) და მისი წინა ხვედრითი წილების საშუალებით.

ცხრილში 2.12 სიხშირულ და პროცენტულ განაწილებებს (ცხრილი 2.10) ჩვენ დავამატეთ კუმულაციური განაწილებაც. კუმულაციური პროცენტული განაწილება გვიჩვენებს, მაგალითად, რომ შერჩევის 35 პროცენტი 40 წელზე ახალგაზრდაა, ანუ 39 წლისა ან უფრო ახალგაზრდაა.

ცხრილი 2.12 დაჯგუფებული სიხშირული განაწილება და კუმულაციური ხვედრითი წილები ცვლადისთვის ასაკი: GSS-ის ქვეშერჩევა

რესპონდენტების ასაკი	სიხშირე (f)	ხვედრითი წილები (%)	C %
20-29	7	7	17.5
30-39	7	14	35.0
40-49	12	26	65.0
50-59	3	29	72.5
60-69	3	32	80.0
70-79	6	38	95.0
80-89	2	40	100.0
სულ (N)	40		

ისევე როგორც ზემოთ იყო აღწერილი, პროცენტულ განაწილებებთან ერთად, კუმულაციური პროცენტული განაწილებებიც ფართოდ გამოიყენება ჯგუფებს შორის განსხვავებების შესასწავლად. აღნიშნულის თვალსაჩინოებისათვის, ჩვენ გამოვიყენეთ GSS-ის 2008 წლის მონაცემები, რათა ხაზი გაგვესვა კონტრასტებზე ერთი მხრივ, თეთრკანიანი მამაკაცებისა და ქალების, და მეორე მხრივ, სხვა ამერიკული ოჯახების ცხოვრების დონეებს შორის. რესპონდენტებს დაუსვეს შემდეგი კითხვა: „მოცემული პირობებიდან გამომდინარე, მე და ჩემ ოჯახს ცხოვრების დონის ამაღლების კარგი შესაძლებლობა გვაქვს – ეთანხმებით, თუ არა?“

ცხრილში 2.13 ნაჩვენებია პროცენტული და კუმულაციური პროცენტული განაწილებები თეთრკანიანი ქალებისა და მამაკაცებისათვის. კუმულაციური პროცენტული განაწილება გვიჩვენებს, რომ მამაკაცების უფრო მეტი რაოდენობა ეთანხმება იმ მოსაზრებას, რომ მათი ცხოვრების დონე გაუმჯობესდება. ორი სქესობრივი ჯგუფი ერთმანეთისგან 6.6%-ით განსხვავდება – ქალების 54,3%-მა უჩვენა, რომ სრულიად ეთანხმება ან ეთანხმება ამ მოსაზრებას და მამაკაცების 60.9%-მა. (აღსანიშნავია, რომ ქალების უფრო მეტი პროცენტი არ ეთანხმება ამ მოსაზრებას, ვიდრე მამაკაცების). რით შეიძლება ამ განსხვავებების ახსნა? თეთრკანიანი მამაკაცებისა და ქალების განსხვავებული მოსაზრებები შეიძლება გენდერული და სახელფასო დისკრიმინაციით აიხსნას. ეს მონაცემები ბევრ კითხვას ბადებს გენდერული (რასობრივი) ფაქტორის როლის შესახებ შემოსავლების უთანაბრობაში. საინტერესოა, არის თუ არა ანალოგიური განსხვავებები სხვა რასობრივ ჯგუფებში? როგორი იქნება განსხვავებები, თუ შევადარებთ თეთრკანიან და შავკანიან მამაკაცებს? თეთრკანიან და ლათინური წარმოშობის ქალებს?

ცხრილი 2.13 გაუმჯობესდება თუ არა ცალკეული პიროვნების ცხოვრების დონე: თეთრკანიანი მამაკაცები თეთრკანიანი ქალების წინააღმდეგ

	თეთრკანიანი მამაკაცები		თეთრკანიანი ქალები	
	%	C %	%	C %
სრულიად ვეთანხმები	18.6	18.6	14.4	14.4
ვეთანხმები	42.3	60.9	39.9	54.3
არც ერთი	13.0	73.9	16.5	70.8
არ ვეთანხმები	24.2	98.1	24.8	95.6
სრულიად არ ვეთანხმები	2.0	100.1	4.4	100.0
სულ	100.1		100.0	
(N)	355		411	

წყარო: General Social Survey, 2008.

დონეები

ტერმინები, როგორიცაა *შობადობა, უმუშევრობის დონე, ქორწინების კოეფიციენტი*, ხშირად გამოიყენება სოციალური სფეროს მეცნიერებისა და დემოგრაფების მიერ. ასევე ხშირად ხდება მათი ციტირება მედიაში მოსახლეობის ტრენდების დასახასიათებლად. რა არის დონე და როგორ ხდება მისი გამოთვლა? **დონე** წარმოადგენს რიცხვს, რომელიც მიიღება მოცემულ პერიოდში რეალური ხდომილობების რაოდენობის ფარდობით შესაძლო ხდომილობების საერთო რაოდენობასთან. მაგალითად, 2008 წელს არსებული სიღარიბის დონის გასაანგარიშებლად, აშშ-ის ცენზების ბიურო ღარიბი ქალებისა და მამაკაცების რაოდენობას (რეალური ხდომილობა) აფარდებს მოსახლეობის მთლიან რიცხოვნობასთან. ეს დონე 2008 წლისათვის შემდეგნაირად გამოისახება:

$$\text{სიღარიბის დონე 2008 წელს} = \frac{\text{ღარიბების რაოდენობა 2008 წელს}}{\text{მთელი მოსახლეობა 2008 წელს}}$$

2008 წელს 40,136,000 ადამიანი ღარიბი იყო, ხოლო მოსახლეობის რიცხოვნობამ 304,060,000 შეადგინა. შესაბამისად, 2008 წლისთვის სიღარიბის დონე შემდეგნაირად გამოითვლება:

$$\text{სიღარიბის დონე 2008 წელს} = \frac{40,136,000}{304,060,000} = 0.132$$

აშშ-ის აღწერის ბიუროს მიერ გამოცხადებული უმუშევრობის დონე 2008 წელს 13.25% (*0.132 × 100*) იყო. დონეებს ხშირად ათასზე ან ასი ათასზე ამრავლებენ, რათა მოვახდინოთ ათწილადების ელიმინირება და რიცხვი ადვილად წასაკითხი გახდეს. მაგალითად, ათასზე გადაანგარიშებით, ცხოვრების დონე ტოლი იქნება 132-ის (*0.132 × 1,000 = 132*). ეს ნიშნავს, რომ ყოველი ათასი კაციდან, 132 ღარიბი იყო.

ზემოთ მოყვანილი სიღარიბის მაჩვენებელი შეიძლება ზოგად *(უხეშ)* დონედ მივიჩნიოთ, რადგან ბაზად გამოყენებულია მთელი მოსახლეობა. დონე შეიძლება გავიანგარიშოთ როგორც მთელ მოსახლეობასთან, ისე უფრო ვიწრო, შერჩეულ ჯგუფებთან მიმართებითაც. მაგალითად, შეგვიძლია გავიანგარიშოთ 18 წლის ქვემოთ ასაკის მოსახლეობის სიღარიბის დონე:

$$\text{სიღარიბის დონე (18 წლის ქვემოთ, 2008)} = \frac{13,454,000}{73,922,000} = 0.182 \times 1,000 = 182$$

აგრეთვე შეგვიძლია გავიანგარიშოთ ხანდაზმული ამერიკელების სიღარიბის დონეც:

$$\text{სიღარიბის დონე (65 წელზე ზევით, 2008)} = \frac{3,842,000}{38,812,00} = 0.099 \times 1,000 = 99$$

დონე რიცხვი, რომელიც მიიღება მოცემულ პერიოდში რეალური ხდომილობების რაოდენობის ფარდობით შესაძლო ხდომილობების საერთო რაოდენობასთან.

სამართალდამცავი ორგანოები მუდმივად აქვეყნებენ მონაცემებს დამნაშავეობის დონეების (დანაშაულები მთელ მოსახლეობაზე გაანგარიშებით), დაპატიმრებათა დონეების (დაპატიმრებების რაოდენობა ჩადენილ დანაშაულთან მიმართებით) და სასჯელის მისჯის დონეების (სასჯელის მისჯის რაოდენობა დანაშაულებრივ ქმედებების მიმართ) შესახებ. მოიფიქრეთ რამდენიმე მაგალითი სხვა ცვლადებზე, რომლებიც შეიძლება ავსახოთ დონეებით.

✓ *ცოდნის შემოწმება*

სტატისტიკა პრაქტიკაში: ქორწინებათა და განქორწინებათა დროითი დინამიკა

როგორ უნდა შევამოწმოთ ცვლილებები ქორწინებისა და განქორწინების დონეებში? შემცირდა თუ არა ქორწინების დონეები? გაიზარდა თუ არა განქორწინების დონეები? პროცენტების მსგავსად, დონეებიც გვეხმარება განსხვავების დადგენაში სხვადასხვა ჯგუფებს შორის და დროში. მხოლოდ 1990 წლის ზოგადი დონე არაფერს გვეუბნება ქორწინებების ზრდაზე ან კლებაზე, თუმცა თუ

ცხრილი 2.14 ქორწინებისა და განქორწინების დონეები ყოველ 1,000 კაცზე

წელი	ქორწინების დონე	განქორწინების დონე
1970	10.6	3.7
1975	10.0	4.8
1980	10.6	5.2
1985	10.1	5.0
1990	9.8	4.7
1995	8.9	4.4
2000	8.5	4.2
2005	7.5	3.6
2007	7.3	3.6

წყარო: U.S. Census Bureau, *Statistical Abstract of the United States*, 2003, Table 83; *Statistical Abstract of the United States*, 2008, Table 121; *Statistical Abstract of the United States*, 2010, Table 126.

შევხედავთ ცხრილს 2.14, ვნახავთ, რომ უკანასკნელი ოცწლეულის განმავლობაში ქორწინებები შემცირდა, განქორწინებები 80-იან წლებში გაიზარდა, თუმცა 2005 წლისათვის ისევ შემცირდა და დაუბრუნდა 1970 წლის დონეს.

ოდნის
რწმება

დარწმუნდით იმაში, რომ თქვენ შეგიძლიათ ცხრილის წაკითხვა. ახსენით, როგორ მივედით ზემოთ მოტანილ დასკვნამდე ცხრილი 2.14-ის გამოყენებით?

კვლევითი ლიტერატურის მიმოხილვა: სტატისტიკური ცხრილები[3]

სიხშირული განაწილებისა და სხვა სახის სტატისტიკური ინფორმაციის შემცველი ცხრილები ფაქტობრივად თითქმის ყველა წიგნში, სტატიასა თუ გაზეთში გვხვდება. თუმცა სტატისტიკური ცხრილების ჩართვა ანგარიშებსა და სტატიებში არ ნიშნავს, რომ კვლევა აუცილებლად სამეცნიერო ხასიათისაა ან დამაჯერებელია. თქვენ ყოველთვის უნდა დასვათ კითხვა იმის თაობაზე, რას გვეუბნება ცხრილი, არის თუ არა საიმედო და ობიექტური ინფორმაცია და ანალიზი? სოციალურ ლიტერატურაში წარმოდგენილი ცხრილები გაცილებით უფრო რთულია იმასთან შედარებით, რასაც ჩვენ ამ თავში ვეხებით. ერთიდაიგივე ინფორმაცია შეიძლება სხვადასხვა სახით წარმოვადგინოთ. სივრცის შეზღუდულობის გამო მკვლევარმა შეიძლება ინფორმაციის დეტალები მინიმუმამდე დაიყვანოს.

ამ ნაწილში, მოგაწვდით მითითებებს იმის შესახებ, თუ როგორ უნდა წაიკითხოთ ცხრილები და როგორი ინტერპრეტაცია უნდა გაუკეთოთ სიხშირული განაწილების ცხრილებს. მიზანი იმაში მდგომარეობს, რომ გაჩვენოთ მოცემულ თავში აღწერილი მეთოდების სწორი გამოყენების მაგალითები. დაიმახსოვრეთ, რომ ცხრილების წაკითხვის უნარ-ჩვევების გამომუშავება გარკვეულ დროსა და პრაქტიკას მოითხოვს. ზოგჯერ, ცხრილების ინტერპრეტაციის დროს გამოცდილი მკვლევარებიც უშვებენ შეცდომებს. ამიტომ დახარჯეთ დრო იმისათვის, რომ შეისწავლოთ აქ მოტანილი ცხრილები, გააკეთეთ თავის სავარჯიშოები და თქვენ იგრძნობთ, რომ დროთა განმავლობაში ბევრად გაგიადვილდებათ ცხრილების წაკითხვა, ინტერპრეტირება და გაგება.

საბაზისო პრინციპები

ნებისმიერი სტატისტიკური ცხრილის წასაკითხად, უპირველეს ყოვლისა უნდა გავიგოთ, თუ რის თქმას ცდილობს მკვლევარი. ამისათვის მკვლევარს სათანადო მიზანი და უნარი უნდა გააჩნდეს. ცხრილის დათვალიერება სათაურიდან იწყება. როგორც წესი, აქ გადმოცემული უნდა იყოს ცხრილის მთავარი შინაარსი. წაიკითხეთ შენიშვნები, რომელიც ახლავს ცხრილს. აქ აღწერილია ინფორმაციის, ან ცხრილის წყაროები, ან ნებისმიერი სხვა ინფორმაცია, რომელსაც მკვლევარი საჭიროდ ჩათვლის. შემდეგ ნახეთ სვეტებისა და სტრიქონების სათაურები და ქვე-

სათაურები. აქ ხდება ცვლადების, კატეგორიებისა და მოტანილი სტატისტიკური ინფორმაციის იდენტიფიკაცია, როგორიცაა ნედლი სიხშირეები თუ პროცენტები. სტატისტიკური ცხრილის თითოეული შემადგენელი ცვლადი, ან ჯგუფი შესაბამის ასახვას მოითხოვს (სიხშირეები, ხვედრითი წილები, დონეები და ა.შ.) ანუ ისე, როგორც ეს განსაზღვრულია სათაურებსა და ქვესათაურებში. ცხრილი 2.15 მოტანილია პროფესორების ერიკ ფონგისა და კუმიკო შიბუიას სტატიიდან, სადაც ისინი იხილავენ სხვადასხვა რასობრივი და ეთნიკური ჯგუფების საცხოვრისების საკუთრებრივ საკითხებს. მკვლევარები შეეცადნენ ეჩვენებინათ განსხვავებები ბინათმფლობელობის საკითხში თეთრკანიანებს, შავკანიანებს, ხუთ მთავარ აზიურ ჯგუფსა და სამ ძირითად ესპანურ ჯგუფს შორის. მონაცემები აღებულია 1990 წლის 1%-იანი საჯარო გამოყენების მიკრომონაცემებიდან (PUMS) და მოიცავს 15 სტატისტიკურ ზონას.

ცხრილი 2.15 ბინათმოსარგებლეთა პროცენტული განაწილება ძირითადი რასობრივი და ეთნიკური ჯგუფების სტატუსის მიხედვით, ადგილმდებარეობის შესაბამისად

	მესაკუთრე გარეუბანში	მესაკუთრე ცენტრალურ ქალაქში	დამქირავებელი გარეუბანში	დამქირავებელი ცენტრალურ ქალაქში	სულ
თეთრკანიანები	50.4	16.8	17.1	15.7	100.0
შავკანიანები	14.6	27.2	13.9	44.3	100.0
აზიელები	34.2	20.6	14.7	30.5	100.0
ჩინელები	33.7	26.1	9.9	30.3	100.0
იაპონელები	39.2	20.8	17.8	22.2	100.0
კორეელები	29.8	13.8	17.1	39.3	100.0
ვიეტნამელები	31.0	11.8	24.3	32.9	100.0
ფილიპინელები	38.1	21.5	15.8	24.6	100.0
ესპანურენოვანი	22.9	15.3	20.8	40,9	99.9
კუბელები	36.3	16.6	16.2	30.9	100.0
მექსიკელები	26.0	15.5	26.8	31,8	100.1
პუერტო-რიკოელები	7.1	14.2	8.4	70.3	100.0
სულ	41.7	18.5	16.8	23.0	100.0

წყარო: ადაპტირებულია 1990 U.S. 1% PUMS Data. Eric Fong and Kumiko Shibuya, „Suburbanization and Home Ownership: The Spatial Assimilation Process in U.S. Metropolitan Cities"-დან, *Sociological Perspectives* 43, no. 1(2000):143.

ცხრილში 2.15 მკვლევარებმა წარმოადგინეს ძირითადი რასობრივი და ეთნიკური ჯგუფების ბინათმოსარგებლეთა პროცენტული განაწილება სტატუსის მიხედვით ადგილმდებარეობის შესაბამისად 1990 წლისათვის. აღსანიშნავია, რომ

სიხშირეები (f) თითოეული კატეგორიისათვის არ არის წარმოდგენილი. მიუხედავად იმისა, რომ ცხრილი საკმაოდ მარტივია, ინფორმაციის სწორად აღქმის მიზნით აუცილებელია მისი დაკვირვებით შესწავლა სათაურების ჩათვლით.

ოდნის
ნწმება

შეამოწმეთ ცხრილი 2.15 და პასუხი გაეცით შემდეგ კითხვებს:

- *რა არის ამ ცხრილის წყარო?*
- *რამდენი ცვლადია წარმოდგენილი?*
- *რა არის ციფრებით წარმოდგენილი პირველ სვეტში? მესამე სვეტში?*

რის თქმა სურთ ავტორებს ამ ცხრილით?

ცხრილი 2.15 [ცხრილი 1], რომელიც გვიჩვენებს ბინის მესაკუთრეთა ჯგუფებს შორის დაჯგუფებას ადგილდებარეობასთან, ოთხი მკვეთრად გამოხატული ასპექტისგან შედგება. პირველ სვეტში, რომელიც ასახავს გარეუბნებში ბინების მესაკუთრეთა პროპორციებს, ჯგუფებს შორის მნიშვნელოვანი ვარიაციები შეინიშნება. მიუხედავად იმისა, რომ 15 მეტროპოლიის მცხოვრებთა 42% მესაკუთრე იყო, ციფრები მოწმობენ, რომ ეს ატიპურია რამდენიმე ჯგუფისათვის. თეთრკანიანების ხვედრითი წილი ყველაზე მაღალია, მათი თითქმის 50% – ბინათმფლობელია. მეორე მხრივ, შავკანიანები და პუერტო-რიკოელები აქ უკიდურეს უმცირესობაში არიან. აზიელების ხვედრითი წილები ბინათმფლობელობაში (43%) სჭარბობს ესპანური მოსახლეობის მაჩვენებელს (23%). თუმცა ამ ორი ჯგუფის შიგნით დიდი განსხვავებები შეინიშნება – 39% იაპონელებს აქვთ და მხოლოდ 7% პუერტო-რიკოელებს.

ცენტრალურ ქალაქებში ბინათმფლობელობის სტატუსი (სვეტი 2) განსხვავებულ სურათს იძლევა, აქ შავკანიანები ლიდერობენ. თითქმის მათი 27% ფლობს ბინას ცენტრალურ ქალაქებში. ფაქტობრივად მათი ორი მესამედი ამ არეებშია განთავსებული. ეს ფაქტი იმით აიხსნება, რომ ბინების ფასები გარეუბნებში უფრო მაღალია, ვიდრე ქალაქებში. შესაძლოა შავკანიანების გამოძევება გარეუბნებიდან საცხოვრისების ბაზარზე არსებული დისკრიმინაციითაც აიხსნებოდეს. აზიური ეთნიკური ჯგუფებიც მნიშვნელოვან ვარიაციას განიცდიან ცენტრალურ ქალაქებში. ხვედრითი წილები იცვლება ვიეტნამელების 12%-იდან ჩინელების 26%-მდე. დიდ ქალაქებში ჩინელების მჭიდრო დისლოკაცია შესაძლოა იქ მათი ტრადიციული „ჩაინათაუნების“ არსებობას უკავშირდება. დიდ ქალაქებში ესპანურ ეთნიკურ ჯგუფებს მსგავსი მაჩვენებლები აქვთ, ისინი იცვლება კუბელების 14 პროცენტიდან მექსიკელების 17 პროცენტამდე.

მესამე სვეტში მოცემულია ბინების დამქირავებელთა მაჩვენებლები გარეუბნებში, რაც ფრიად განსხვავებულია გარეუბნებში მესაკუთრეთა სტატისტიკისაგან. შედეგები მიგვანიშნებენ თეთრკანიანებისა და ესპანური მოსახლეობის მაღალ ხვედრით წილებზე გარეუბნებში ბინების დაქირავების მიხედვით. ასევე დაბალია საშუალოდ აზიელი დამქირავებლების ხვედრითი წილები, თუმცა ცალკეულ ეთნიკურ ჯგუფებში, როგორიც არიან ვიეტნამელები, უფრო მაღა-

ლია დამქირავებელთა ხვედრითი წილი, ვიდრე თეთრკანიანებს შორის. ვიეტნამელების მაღალი მაჩვენებელი შეიძლება უკავშირდებოდეს მათ განთავსებას ლტოლვილთა პროგრამების მიხედვით. აგრეთვე, აზიური ეთნიკური კომუნების არარსებობას ცენტრალურ ქალაქებში...

მეოთხე სვეტში მოცემულია დამქირავებელთა განაწილება ცენტრალურ ქალაქებში. სხვა ჯგუფებთან შედარებით, ამ კატეგორიაში უმაღლესი მაჩვენებლები შავკანიანებს და პუერტო-რიკოელებს აქვთ. პუერტო-რიკოელების თითქმის 70% და შავკანიანების 44% ბინას ქირაობს ცენტრალურ ქალაქებში. აზიელებს აქ მოკრძალებული მაჩვენებლები აქვთ, ყველაზე დაბალი — იაპონელებს (22%) და ყველაზე მაღალი — კორეელებს (39%). თეთრკანიანები ყველაზე ნაკლებად ქირაობენ ბინებს ცენტრალურ ქალაქებში (16%-მდე).

საერთო ჯამში, კვლევამ აჩვენა საცხოვრისებთან დაკავშირებული საკითხების რთული ბუნება რასობრივ/ეთნიკურ ჭრილში. გარეუბნებსა და ცენტრალურ ქალაქებში ბინათმფლობელობის სტატუსის კვლევის შედეგი ოთხ ძირითად ჭრილშია წარმოდგენილი, რაც შეიძლება არც გამოვლენილიყო, თუ ცალ-ცალკე განვიხილავდით ურბანულ და საკუთრებრივ საკითხებს. აღნიშნული კვლევით შესაძლებელია მარტივად ავსახოთ ის განსხვავებები, რაც არსებობს სხვადასხვა ჯგუფების სოციალ-ეკონომიკურ რესურსებსა და კულტურულ დონეებს შორის ორივე ადგილმდებარეობის მიხედვით.[10]

ამ ცვლადების ურთიერთკავშირების უფრო საფუძვლიანი შესწავლისათვის საჭიროა უფრო რთული მეთოდები ბივარიაციული ანალიზიდან და დასკვნითი სტატისტიკიდან, რასაც მე-9 თავიდან შევეხებით.

სხვადასხვა ფორმატის ცხრილები

ზოგჯერ ცხრილებში მონაცემები წარმოდგენილია მხოლოდ შერჩევითი ერთობლიობის ნაწილიდან. მაგალითად, ცხრილში 2.16, რომელიც ეფუძნება 2008 წელს ჩატარებული მექსიკელი ამერიკელების აღწერის მასალებს, მხოლოდ შერჩეული ცვლადების ხვედრითი წილებია. ცხრილში თითოეული ცვლადისათვის ნაწილობრივი ინფორმაციაა წარმოდგენილი, ამიტომ ამ პროცენტების ჯამი 100%-ის ტოლი ვერ იქნება. მაგალითად, მექსიკელების 55.2%-ს საშუალო სკოლისა და

ცხრილი 2.16 შერჩეული ეკონომიკური და სოციალური მაჩვენებლები მექსიკელებისთვის, 2008

მაჩვენებლები	ხვედრითი წილები %
საშუალო სკოლა ან უფრო მაღალი დონის განათლება	55.2
სიღარიბის ზღვარს ქვემოთ მცხოვრები მოსახლეობა	22.9
მესაკუთრის მიერ დაკავებული საცხოვრისი	51.4
65 წლის და ზევით	4.4

წყარო: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 39.

უფრო მაღალი დონის განათლება აქვს. დანარჩენი 44.8%, ვისაც უფრო დაბალი განათლება აქვს, ცხრილში არ არის წარმოდგენილი და გამოტოვებულია.

მიუხედავად ამისა, რომ ცხრილი 2.16 სასარგებლო ინფორმაციას შეიცავს, ჩვენ ხშირად ისეთი ინფორმაცია გვაინტერესებს, რაც სცილდება ცვლადების განაწილების მარტივ აღწერას. მკვლევართა უმრავლესობა ძირითადად სხვადასხვა ჯგუფების შედარებას იყენებს, ან აკვირდება ერთი ჯგუფის შიგნით მიმდინარე ცვლილებას დროში. მაგალითად, თუ გვსურს, რომ ცხრილში 2.16 მოტანილი ინფორმაცია უფრო მნიშვნელოვანი გავხადოთ, შეიძლება ის შევადაროთ სხვა ლათინურ ჯგუფებს. ეს საშუალებას მოგვცემს პასუხი გავცეთ კითხვებს, თუ რამდენად „მაღალია“ საშუალო სკოლისა და უფრო მაღალი დონის განათლების მქონეთა პროცენტი, ან „მაღალია“ თუ „დაბალია“ 65 წლის და ზევით ასაკის მოსახლეობის ხვედრითი წილი.

შეხედეთ ცხრილს 2.17. ის შეიცავს ინფორმაციას ცხრილიდან 2.16 და ასევე დამატებულია შესაბამისი ინფორმაცია კუბელ ამერიკელებზე. თქვენ შენიშნავთ განსხვავებებს ცხრილში მოყვანილ ყველა მაჩვენებელს შორის; მაგალითად, კუბელი ამერიკელების უფრო მეტ პროცენტს აქვს საშუალო სკოლისა და უფრო მაღალი დონის განათლება (80%), ვიდრე მექსიკელ ამერიკელებს (55,2%).

ცხრილი 2.17 შერჩეული ეკონომიკური და სოციალური მაჩვენებლები მექსიკელებისა და კუბელებისათვის, 2008

მაჩვენებლები	მექსიკელების ხვედრითი წილები %	კუბელების ხვედრითი წილები %
საშუალო სკოლა ან უფრო მაღალი დონის განათლება	55.2	80.0
სიღარიბის ზღვარს ქვემოთ მცხოვრები მოსახლეობა	22.9	14.6
მესაკუთრის მიერ დაკავებული საცხოვრისი	51.4	62.0
65 წლის და ზევით	4.4	17.9

წყარო: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 39.

დასკვნა

ამ თავის შესავალ ნაწილში ჩვენ აღვნიშნეთ, რომ სიხშირული განაწილების აგება, როგორც წესი, სტატისტიკური ანალიზის პირველი საფეხურია; ვიმეორებთ, ახლა დაგვეთანხმებით, რომ საბაზისო სიხშირული და პროცენტული განაწილებების აგება ძალიან მარტივია. როგორც თქვენ დარწმუნდით ამ თავში მოტანილი მაგალითებიდან, განაწილების ცხრილები ეხმარება მკვლევარებს მონაცემების თავმოყრაში, დაჯამებაში, გამოსახვასა და აღწერაში. ჯგუფებს შორის

მსგავსებისა და განსხვავების ტენდენციები შეიძლება გამოვავლინოთ მარტივი განაწილების ცხრილებით.

მომდევნო თავებში განაწილების ცხრილებს ბაზად გამოვიყენებთ მონაცემების გრაფიკული ასახვისა და ცენტრალური ტენდენციისა და ვარიაციის გაზომვისათვის. სხვა სიტყვებით რომ ვთქვათ, თქვენ კიდევ და კიდევ შეხვდებით სიხშირეებსა და პროცენტულ განაწილებებს, ამიტომ ახლავე დარწმუნდით, რომ შეძლებთ ააგოთ და წაიკითხოთ განაწილების ცხრილები, სანამ ახალ თავზე გადახვიდოდეთ.

ძირითადი განმარტებები

• მონაცემების თავმოყრის ძირითადი მეთოდია დაკვირვებების დაჯგუფება სიხშირულ განაწილებებში – ცხრილში, რომელიც ასახავს საანალიზო ცვლადის თითოეულ კატეგორიაში მოხვედრილი დაკვირვებების რაოდენობას.

• მონაცემთა სტატისტიკურ ანალიზში სიხშირული განაწილების აგება – პირველი საფეხურია.

• ნომინალური და ორდინალური ცვლადებისათვის სიხშირული განაწილების მისაღებად, დათვალეთ და ჩაწერეთ ცვლადის თითოეული კატეგორიისათვის დაკვირვებათა რაოდენობა დაკვირვებათა მთლიან რაოდენობასთან (N) ერთად.

• პროცენტული (ხვედრითი წილების) სიხშირე ცხრილის სახით წარმოდგენილი ცვლადის თითოეულ კატეგორიაში მოხვედრილი დაკვირვებების ხვედრითი წილებია. პროცენტული სიხშირე ჩვეულებრივ ემატება სიხშირულ ცხრილებს განსაკუთრებით მაშინ, თუ საჭიროა ჯგუფებს შორის შედარებები.

• კუმულაციური სიხშირული განაწილება საშუალებას გვაძლევს განვსაზღვროთ მოცემული სიდიდის ფარდობითი ადგილი განაწილებაში. ის აიგება თითოეული კატეგორიის სიხშირისათვის მისი წინა სიხშირეების დამატებით.

• სიხშირული განაწილების მისაღებად ინტერვალური ცვლადებისათვის, რომელთა მნიშვნელობების ვარიაციის დიაპაზონი საკმაოდ მაღალია, თავიდან ისინი გააერთიანეთ შედარებით მცირე ჯგუფებში, ანუ ინტერვალებში.

• კოეფიციენტები და ხვედრითი წილები შეფარდებითი სიდიდეებია. კოეფიციენტი მიიღება თითოეული კატეგორიის სიხშირის (f) ფარდობით დაკვირვებათა რიცხვზე (*N*). იმისათვის, რომ მივიღოთ ხვედრითი წილი, თითოეული კატეგორიის სიხშირე (f) უნდა გავყოთ დაკვირვებათა რიცხვზე (*N*) და გავამრავლოთ 100-ზე.

• კუმულაციურ პროცენტულ განაწილებებს უფრო ფართო გამოყენება აქვს. კუმულაციურ სიხშირულ განაწილებებთან შედარებით ისინი აიგება თითოეული კატეგორიის პროცენტული სიხშირისათვის მისი წინა სიხშირეების დამატებით.

• ნედლი სიხშირეების შეფარდებით

სიდიდეებში გამოსახვის ერთ-ერთი მეთოდია დონეების გამოყენება. ის წარმოადგენს რიცხვს, რომელიც მიიღება მოცემულ პერიოდში რეალური ხდომილობების რაოდენობის ფარდობით შესაძლო ხდომილობების საერთო რაოდენობასთან. დონე ხშირად მრავლდება 10-ის რომელიმე ხარისხზე ათწილადების ელიმინირებისა და ციფრების ინტერპრეტირების გაადვილების მიზნით.

საკვანძო ტერმინები

კუმულაციური სიხშირული განაწილება
კუმულაციური პროცენტული განაწილება
სიხშირული განაწილება
ხვედრითი წილი
პროცენტული განაწილება
კოეფიციენტი
დონე

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

[MTF 2008]

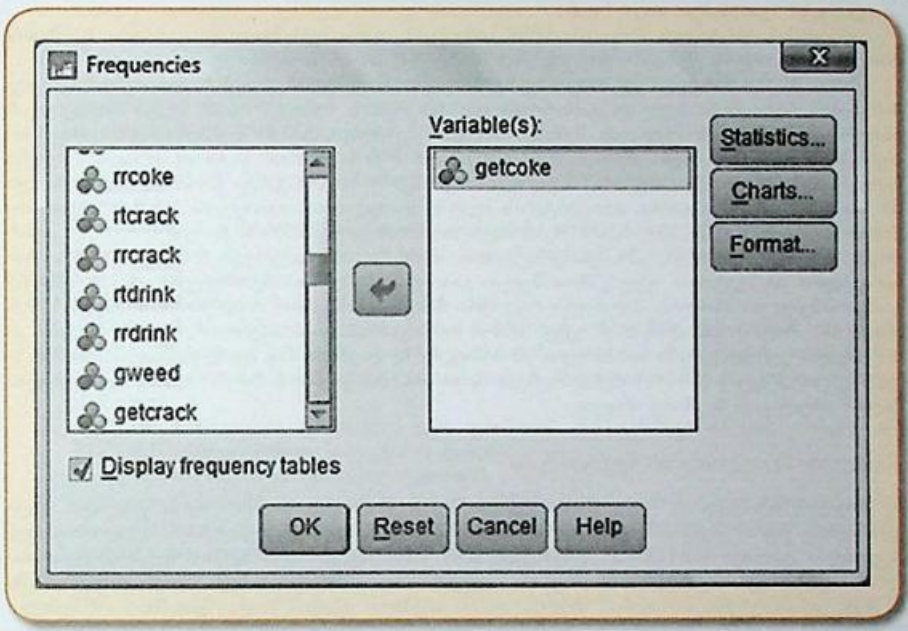
მაგალითი 1: სიხშირული განაწილებების აგება

SPSS-ში თქვენ შეგიძლიათ შექმნათ სიხშირული განაწილებები როგორც ერთი, ისე ერთდროულად რამდენიმე ცვლადისათვის. სიხშირესთან დაკავშირებული პროცედურები მოთავსებულია მენიუში Descriptive Statistics, Analyze-ის ქვემოთ. ამ თავისთვის ჩვენ გამოვიყენებთ მონაცემებს Monitoring the Future 2008-იდან.

სიხშირეების დიალოგურ ფანჯარაში (ნახ. 2.5), დაკლიკეთ ცვლადის სახელზე მარცხენა სვეტში და გადაიტანეთ ის ცვლადების ყუთში, მარჯვენივ (ნახაზზე 2.5 ცვლადები განლაგებულია რიგითობის მიხედვით, თუმცა ისინი შეიძლება ალფაბეტურადაც იყოს განლაგებული). დაიმახსოვრეთ, რომ ერთბაშად რამდენიმე ცვლადის სიხშირის გაანგარიშებაა შესაძლებელი.

ამ მაგალითისათვის შევარჩიოთ ცვლადი GETCOKE („ადვილია თუ არა კოკაინის შოვნა?“). სიხშირეების ასაგები პროცესის დასაწყებად, დაკლიკეთ OK-ზე. რესპონდენტებს უნდა გაეცათ კოდირებული პასუხები შემდეგი სახით: 1. ალბათ შეუძლებელია; 2. ძალიან ძნელია; 3. საკმაოდ ძნელია; 4. საკმაოდ ადვილია და 5. ძალიან ადვილია.

ნახაზი 2.5 სიხშირეების დიალოგური ფანჯარა



ნახაზი 2.6 სიხშირული ცხრილი ცვლადისთვის GETCOKE

How Easy to Get Cocaine

		Frequency	*Percent*	*Valid Percent*	*Cumulative Percent*
Valid	PROB IMP: (1)	216	14.4	14.9	14.9
	VRY DIFFF: (2)	266	17.7	18.4	33.3
	FRLY DIF: (3)	375	25.0	25.9	59.1
	FRLY EAS: (4)	360	24.0	24.8	84.0
	VRY EASY: (5)	232	15.5	16.0	100.0
	Total	1449	96.6	100.0	
Missing	MISSING: (-9)	51	3.4		
Total		1500	100.0		

შედეგების ფანჯარაში SPSS შექმნის ორ ცხრილს – სტატისტიკურს და სიხშირეების. თქვენ გჭირდებათ სიხშირეების ცხრილი, ამიტომ ცოტა ქვევით ჩამოინაცვლეთ. გაზომვის რა დონე აქვს ამ ცვლადს? (გაიხსენეთ თავი 1-დან).

პირველ, სტატისტიკურ ცხრილში SPSS გამოიტანს ყველა შევსებული და ცარიელი პასუხების რაოდენობას. პასუხს ეწოდება missing, თუ მასზე პასუხი არ მოიძებნა (გამოუპასუხებელია).

ცხრილში 2.6 პირველ სვეტში კითხვაზე პასუხები მოცემულია კოდების მითითებით, აგრეთვე, პასუხების რაოდენობა სულ და გამოტოვებული პასუხებიც. შემდეგი ოთხი სვეტი მნიშველოვან ინფორმაციას შეიცავს ცვლადის შესახებ. როგორც ვხედავთ, 232-მა რესპონდენტმა გასცა პასუხი, რომ კოკაინის შოვნა ძალიან ადვილია, ხოლო 360-მა – რომ საკმაოდ ადვილია. თქვენც იგივე პასუხებს ელოდით, თუ არა? სულ ჯამში კითხვაზე პასუხი გასცა 1,449 ადამიანმა, ხოლო 51 პასუხი ცარიელი დარჩა. პროცენტების სვეტში დათვლილია თითოეული პასუხის ხვედრითი წილი მთელ შერჩევაში, ანუ 1,500 რესპონდენტიდან. ამიტომ, 15.5%-მა ჩათვალა, რომ კოკაინის შოვნა ადვილია. სინამდვილეში, მესამე სვეტიდან Valid Percent მიღებული მონაცემები უფრო სასარგებლოა. ამ სვეტიდან უკვე გამოთიშულია გამოუპასუხებლობის შემთხვევები და პროცენტები გაანგარიშებულია მხოლოდ მიღებული რეალური პასუხების საერთო რაოდენობიდან, რაც 1,449-ს შეადგენს. ამიტომ გამოდის, რომ უკვე 16%-მა უპასუხა „ძალიან ადვილია". ბოლო სვეტში კი მოცემულია კუმულაციური სიხშირეები. ამ მონაცემებზე დაყრდნობით შეიძლება ვთქვათ, რომ ნახევარზე ცოტა მეტმა (59.1%) რესპონდენტმა მიუთითა, რომ კოკაინის შოვნა ალბათ შეუძლებელია, ძალიან ძნელია ან საკმაოდ ძნელია.

მაგალითი 2: ცვლადების რეკოდირება

ზოგჯერ მონაცემების უკეთ წარმოდგენისა და აღქმისათვის საჭირო ხდება ცვლადის კატეგორიების უფრო შემჭიდროვება ან კომპაქტური გამოსახვა ინტერვალებში. მაგალითად, ჩვენ შეგვიძლია ცვლადი GETCOKE წარმოვადგინოთ სამი კატეგორიით: ალბათ შეუძლებელია; ძალიან ძნელია/საკმაოდ ძნელია; და საკმაოდ ადვილია/ძალიან ადვილია. ჩვენ გავაერთიანეთ კოდში 2 შემდეგი პასუხები: 2. ძალიან ძნელია და 3. საკმაოდ ძნელია, ხოლო კოდში 3 – შემდეგი პასუხები: 4. საკმაოდ ადვილია და 5. ძალიან ადვილია. პირველი კატეგორია, 1. ალბათ შეუძლებელია, უცვლელი დავტოვეთ. ამ ოპერაციის შესასრულებლად, ჩვენ უნდა გამოვიყენოთ SPSS-ის მენიუდან ბრძანება Transform - Recode Into Different Variables.

რეკოდირების შემდეგ, ჩვენ მივიღეთ ახალი ცვლადი RGETCOKE. ამ ცვლადის სიხშირეები მოცემულია ნახაზზე 2.7.

ნახაზი 2.7 სიხშირული ცხრილი ცვლადისთვის GETCOKE

rgetcoke Recoded GETCOKE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 probably impossible	216	14.4	14.9	14.9
	2 very difficulty/fairly difficult	641	42.7	44.2	59.1
	3 fairly easy/very easy	592	39.5	40.9	100.0
	Total	1449	96.6	100.0	
Missing	System	51	3.4		
Total		1500	100.0		

ამოცანები SPSS-ისათვის

[MTF 2008]

1. გამოიყენეთ SPSS-ის ბრძანება Frequencies და შექმენით სიხშირული ცხრილი ცვლადისთვის GREWUP. აგვიხსენით, სად გაიზარდა შერჩევაში მონაწილე სტუდენტების უმეტესობა?
ა. სტუდენტების რა პროცენტია პატარა ქალაქებიდან?
ბ. სტუდენტების რა პროცენტია დიდი ქალაქებიდან?
გ. სტუდენტების რა პროცენტია სასოფლო ფერმებიდან?
დ. სტუდენტების რა პროცენტს მიიჩნევთ გარეუბნებიდან? (ჩართეთ ყველა შესაბამისი კატეგორია).
2. MTF 2008-ის გამოკვლევაში შევიდა კითხვების სერია ნარკოტიკებთან სტუდენტების მისაწვდომობის შესახებ. ჩვენ უკვე გავაკეთეთ მაგალითი კოკაინზე. დანარჩენი ცვლადები მოიცავს მარიჯუანას და კოკაინ კრეკს.
ა. შექმენით სიხშირეები სამივე ნარკოტიკისთვის.
ბ. მოამზადეთ შემაჯამებელი ანგარიში სამივე სიხშირული ანალიზის ცხრილიდან. თითოეულ ცვლადს მიუთითეთ გაზომვის დონე. როგორ დაახასიათებთ ნარკოტიკების მისაწვდომობას? იდენტურია თუ განსხვავდება ერთმანეთისგან სიტუაცია სამივე ნარკოტიკთან დაკავშირებით?
3. შექმენით სიხშირული ცხრილები ცვლადებისთვის RTCOKE (გასინჯვის რისკი) და RRCOKE (რეგულარული გამოყენების რისკი).
ა. გაზომვის რა დონე აქვს თითოეულ ცვლადს?
ბ. მოამზადეთ შემაჯამებელი ანგარიში სამივე სიხშირული ანალიზის ცხრილიდან. არის თუ არა განსხვავება ამ რისკებს შორის? რომელი სახის რისკი უფრო მაღალია? თვალსაჩინოებისათვის დაეყრდენით ცხრილების მონაცემებს.
4. MTF 2008-ში შესულია კითხვა: "რომელ კლასში მოწიეთ პირველად სიგარეტი?" (GRSMOKE). ააგეთ სიხშირული ცხრილი ამ ცვლადისათვის. გადაიყვანეთ ეს ინტერვალური ცვლადი ორდინალურში (გამოტოვეთ ისინი, ვისაც საერთოდ არ მოუწევია). რამდენი კატეგორია მიიღეთ? მოამზადეთ კუმულაციური სიხშირული ცხრილი ახალი რეკოდირებული ცვლადისათვის GRSMOKE.

თავის სავარჯიშოები

1. წარმოიდგინეთ, რომ თქვენ ჩაატარეთ 30-კაციანი კვლევა და დაუსვით კითხვები რეპონდენტებს რასობრივი კუთვნილების, გადატანილი ტრავმებისა და სოციალური სტატუსის შესახებ. თქვენ მიერ შეგროვილი ნედლი მონაცემები წარმოდგენილია ქვემოთ ცხრილში:
ა. გაზომვის რომელი დონეები აქვთ ცვლადებს: რასა და სოციალური ჯგუფი?
ბ. ააგეთ ნედლი სიხშირეების ცხრილი ცვლადებისთვის: რასა და სოციალური ჯგუფი.
გ. როგორია არათეთრკანიანი მოსახლეობის ხვედრითი წილი? თეთრკანიანი მოსახლეობის?
დ. რამდენმა პროცენტმა მიაკუთვნა თავი საშუალო ფენას?
2. გამოიყენეთ მონაცემები და ნედლი სიხშირის ცხრილები სავარჯიშო 1-დან, ააგეთ სიხშირული განაწილება სოციალური სტატუსისთვის.

რასა	ჯგუფი	ტრავმა	რასა	ჯგუფი	ტრავმა
W	L	1	W	W	0
W	M	0	W	M	2
W	M	1	W	W	1
N	M	1	W	W	1
N	L	2	N	W	0

W	W	0	N	M	2
N	W	0	W	M	1
W	M	0	W	M	0
W	M	1	N	W	1
N	W	1	W	W	0
N	W	2	W	W	0
N	M	0	N	M	0
N	L	0	N	W	0
W	U	0	N	W	1
W	W	1	W	W	0

წყარო: Data based on General Social Survey Files for 1987 to 1991.

შენიშვნა: რასა: W, თეთრკანიანი; N, არათეთრკანიანი; სოციალური ჯგუფი: L, დაბალი; M, საშუალო; U, მაღალი; W, მუშები.

ა. რომელია ყველაზე მცირე ჯგუფი?
ბ. რომელ ორ ჯგუფშია პროცენტულად ყველაზე მეტი ადამიანი?
3. გამოიყენეთ მონაცემები სავარჯიშო 1-დან, ააგეთ სიხშირული განაწილება ტრავმისთვის.
ა. გაზომვის რომელი დონეა გამოყენებული ცვლადისთვის ტრავმა?
ბ. რა უფრო დამახასიათებელია ადამიანებისთვის უკანასკნელი წლის განმავლობაში, ერთი ტრავმა თუ არც ერთი?
გ. რესპონდენტების რამდენმა პროცენტმა მიიღო ერთი და მეტი ტრავმა უკანასკნელი წლის განმავლობაში?
4. დაუშვათ, თქვენ ატარებთ გამოკვლევას განათლებაზე GSS 2008-ის შერჩევითი ერთობლიობიდან. კვლევა შეიცავს კითხვას განათლების ხანგრძლივობის შესახებ წლებში. 1,497 რესპონდენტის სიხშირული განაწილება წარმოდგენილია ცხრილში:
ა. გაზომვის რა დონე აქვს ცვლადს განათლების ხანგრძლივობა?
ბ. ამ ცვლადისთვის ააგეთ სიხშირული ცხრილი კუმულაციური ხვედრითი წილებით.

განათლების ხანგრძლივობა წლებში	სიხშირე
0	5
2	2
3	4
4	8
5	4
6	21
7	9
8	28
9	35
10	41
11	90
12	432

13	115
14	200
15	69
16	235
17	42
18	87
19	28
20	42

წყარო: General Social Survey, 2008.

გ. რამდენ რესპონდენტს აქვს მითითებული ხანგრძლივობა 8 წელი ან მასზე ნაკლები? შერჩევის რამდენ პროცენტს შეადგენს ეს რაოდენობა?

დ. დაუშვათ, რომ უფრო სიღრმისეულად გინდათ განათლების შესწავლა და გსურთ ოთხ ჯგუფში გააერთიანოთ კატეგორიები, რათა უკეთესად წარმოადგინოთ თქვენი ინტერესის საგანი. შევთანხმდეთ, რომ ყველა, ვისაც 12-წლიანი განათლება აქვს, საშუალო სკოლადამთავრებულია, ხოლო ვინც 16 წელი ისწავლა, კოლეჯი აქვს დამთავრებული. ააგეთ კუმულაციური პროცენტული განაწილების ცხრილი ამ ოთხი ჯგუფისათვის. შერჩევის რა პროცენტმა დაამთავრა კოლეჯი? შერჩევის რა პროცენტს არ დაუმთავრებია სკოლა?

5. გელაპის ორგანიზაციამ 2007 წლის ივნისში ჩაატარა კვლევა, რომელშიც შეეკითხა ამერიკელებს, ეთანხმებოდნენ თუ არა ისინი აშშ-ის მთავრობის გადაწყვეტილებებს არალეგალურ იმიგრანტებთან დაკავშირებით. ქვემოთ მოყვანილ ცხრილში ინფორმაცია წარმოდგენილია რასობრივ ჭრილში პროცენტობით (ყველა პასუხი აქ არ არის დაფიქსირებული, ამიტომ შეიძლება პროცენტების ჯამი არ უდრიდეს 100-ს). ადასტურებს თუ არა ეს მონაცემები მოსაზრებას, რომ არალეგალური იმიგრაციის შესახებ ადამიანების შეხედულებები რასობრივი ნიშნით არის განპირობებული? პასუხი ორივე შემთხვევაში დაასაბუთეთ.

	არა-ესპანელი თეთრკანიანები	*შავკანიანები*	*ესპანელები*
ვეთანხმები	32	29	37
არ ვეთანხმები	60	60	58

6. აშშ-ში გრძელდება იმიგრაციის პროცესი, ხშირად იმიგრანტები უპირატესობას ანიჭებენ მშობლიურ ენაზე ლაპარაკს, ყოველ შემთხვევაში, სახლში მაინც. ქვინსი ნიუ-იორკში ყველაზე მრავალეროვანი ადგილია. სიხშირული განაწილების ცხრილში მოცემულია 2000 წლის მოსახლეობის აღწერის შედეგები ძირითადი იმიგრანტების სალაპარაკო ენის შესახებ.

არაბები	12,504
ჩინელები	12,650
გერმანელები	12,054
ბერძნები	39,418

უნგრელები	4,753
ინდიელების ენებზე მოლაპარაკე	95,484
იტალიელები	44,411
იაპონელები	5,300
კორეელები	57,447
პოლონელები	20,883
პორტუგალიელები	9,479
რუსები	36,517
ესპანელები	48,504
ტაგალოგები	25,436
ვიეტნამელები	3,361
იდიშზე მოლაპარაკე	5,472

ა. ააგეთ სიხშირული და პროცენტული განაწილების ცხრილი.
ბ. რატომ იქნება კუმულაციური სიხშირეების გამოთვლა უსარგებლო ამ შემთხვევაში?
გ. რომელია 4 ყველაზე გავრცელებული ენა ქვინსში?
დ. რომელი ორი ენა არის ყველაზე ნაკლებად გავრცელებული ქვინსში?

7. ქვემოთ მოტანილია ცხრილები განათლების დონის შესახებ ასაკობრივ-გენდერულ ჭრილში. დააკვირდით და პასუხი გაეცით კითხვებს.

განათლება	სქესი	
	მამაკაცი (f)	ქალი (f)
საშუალო სკოლაზე დაბალი	100	116
საშუალო სკოლა	327	407
დაუმთავრებელი კოლეჯი	68	76
კოლეჯი	190	216

წყარო: General Social Survey, 2008.

განათლება	რასა/ეთნიკურობა		
	თეთრკანიანი (f)	შავკანიანი (f)	ესპანელი (f)
საშუალო სკოლაზე დაბალი	131	50	21
საშუალო სკოლა	567	109	25
დაუმთავრებელი კოლეჯი	113	19	5
კოლეჯი	335	28	5

წყარო: General Social Survey, 2008.

ა. ააგეთ განათლების პროცენტული და კუმულაციური სიხშირული ცხრილები სქესობრივ და რასობრივ ჭრილში.

ბ. მამაკაცების რამდენმა პროცენტმა გააგრძელა სწავლა საშუალო სკოლის შემდეგ? ქალების რამდენმა პროცენტმა გააგრძელა სწავლა საშუალო სკოლის შემდეგ?

გ. თეთრკანიანების რამდენმა პროცენტმა მიიღო საშუალო სკოლის ან უფრო დაბალი დონის განათლება? როგორია შავკანიანების ხვედრითი წილი? ესპანურენოვანი მოსახლეობის?

დ. არის თუ არა მსგავსი კუმულაციური სიხშირეები მამაკაცებისა და ქალებისათვის ან რასობრივი და ეთნიკური ჯგუფებისთვის? (სხვა სიტყვებით, სად შეინიშნება ყველაზე მეტი უთანაბრობა?) ახსენით.

ე. ამ თავის წინა ნაწილში ცხრილში 2.13 მოცემული იყო თეთრკანიანი მოსახლეობის აზრი იმის შესახებ, გაუმჯობესდებოდა თუ არა მათი ცხოვრების დონე. შემდეგი ცხრილები, რომელიც GSS 2008-ის საფუძველზეა აგებული, წარმოგვიდგენს ფარდობით სიხშირულ განაწილებას უფრო მცირე შერჩევითი ერთობლიობებისათვის სქესობრივ ჭრილში (საშუალო სკოლაზე უფრო დაბალი განათლებით და ბაკალავრებს).

	რესპონდენტების სქესი			
	მამაკაცი		ქალი	
	f	%	f	%
სრულიად ვეთანხმები	38	18.8	34	14.5
ვეთანხმები	88	43.6	103	43.8
არც ერთი	26	12.9	36	15.3
არ ვეთანხმები	45	22.3	51	21.7
სრულიად არ ვეთანხმები	5	2.5	11	4.7
სულ (*N*)	202	100.1	235	100.0

წყარო: General Social Survey, 2008.

	რესპონდენტების განათლების დონის მაჩვენებლები			
	საშუალო სკოლაზე დაბალი		ბაკალავრი	
	f	%	f	%
სრულიად ვეთანხმები	8	11.6	16	20.0
ვეთანხმები	36	52.2	40	50.0
არც ერთი	5	7.2	6	7.5
არ ვეთანხმები	15	21.7	16	20.0
სრულიად არ ვეთანხმები	5	7.2	2	2.5
სულ (*N*)	69	99.9	80	100.0

წყარო: General Social Survey, 2008.

ა. როგორია კუმულაციური ხვედრითი წილი ქალებისათვის, რომლებიც ეთანხმებიან ან სრულიად ეთანხმებიან მოსაზრებას, რომ მათი ცხოვრების დონე ამაღლდება? როგორია მამაკაცების ხვედრითი წილი, რომლებიც არ ეთანხმებიან, ან სრულიად არ ეთანხმებიან ამ აზრს?

ბ. როგორ თანაფარდობაშია სკოლადამთავრებულებზე უფრო დაბალი განათლების მქონეთა და ბაკალავრების კუმულაციური პროცენტები, რომლებსაც სჯერათ?

გ. როგორია კუმულაციური პროცენტები ქალებისა და მამაკაცებისთვის, რომელთაც სჯერათ, რომ მათი ცხოვრების დონე გაუმჯობესდება? (გააერთიანეთ პასუხები „ვეთანხმები“ და „სრულიად ვეთანხმები“).

9. MTF 2008-ის საფუძველზე წარმოვადგენთ უკანასკნელი 12 თვის მანძილზე მოძრაობის წესების დარღვევების გამო ქალებისა და მამაკაცების მიერ მიღებული საჯარიმო ქვითრების მონაცემებს.

	მამაკაცები (%)	ქალები (%)
არც ერთი	71.3	80.4
1	17.1	13.0
2	7.4	5.0
3	2.1	0.7
4	2.1	0.9
სულ	100.0	100.0

წყარო: Monitoring the Future, 2008.

ა. რა არის ცვლადის გაზომვის დონე?

ბ. ქალებისა და მამაკაცების რამდენ პროცენტს აქვს მიღებული 3 ან მეტი საჯარიმო ქვითარი? პროცენტები გაიანგარიშეთ ცალ-ცალკე ქალებისთვის და მამაკაცებისთვის.

გ. გადაივანეთ პროცენტები სახშირეებში. მამაკაცების რაოდენობაა 624 და ქალების საერთო რაოდენობაა 699.

დ. რა შეიძლება ითქვას მიღებული ქვითრების რაოდენობებს შორის განსხვავებაზე ქალებსა და მამაკაცებს შორის?

10. MTF 2008-ის რესპონდენტებს დაუსვეს კითხვა იმის შესახებ, თუ რამდენი დღე ჰქონდათ გაცდენილი სკოლაში უკანასკნელი 4 კვირის განმავლობაში. მონაცემები წარმოდგენილია ცხრილში.

	მამაკაცები	ქალები
0	403	371
1	98	138
2	55	71
3	46	53
4–5	26	46
6–10	9	18
11+	5	10
სულ (N)	642	707

წყარო: Monitoring the Future, 2008.

ა. გაიანგარიშეთ კუმულაციური სიხშირული განაწილება მამაკაცებისთვის.

ბ. გაიანგარიშეთ კუმულაციური სიხშირული განაწილება ქალებისთვის.

11. გელაპმა 2004 წლის ივნისში ჩაატარა კვლევა იმის შესახებ, ეხმარება თუ არა აშშ-ის მთავრობა უმცირესობებს თავიანთი სოციალ-ეკონომიკური მდგომარეობის გაუმჯობესებაში

(ყველა პასუხი არ არის მოტანილი). როგორ აფასებთ თეთრკანიანებისა და უმცირესობების განსხვავებულ შესაძლებლობებს?

რა როლი ენიჭება მთავრობას უმცირესობების სოციალ-ეკონომიკური მდგომარეობის გაუმჯობესებაში?	არა-ესპანელი თეთრკანიანები (%)	შავკანიანები (%)	ესპანელები (%)
მთავარი როლი	32	69	67
უმნიშვნელო როლი	51	22	21
არანაირი	16	9	8

წყარო: Jeffrey Jones, „Blacks More Pessimistic Than Whites About Economic Opportunities", Gallup Pool News Service, The Gellup Organization. July 9, 2004. გამოყენებულია ნებართვით.

12. ამ სავარჯიშოში ჩვენ ვამოწმებთ მოძრაობის საჯარიმო ქცევების ურთიერთდამოკიდებულებას რასობრივ კუთვნილებასთან.

	შავკანიანი (%)	თეთრკანიანი (%)	ლათინო-ამერიკელი (%)
არც ერთი	138	597	153
1	17	149	17
2	8	64	8
3	1	13	4
4	2	14	2
სულ	166	837	184

წყარო: Monitoring the Future, 2008.

ა. გაიანგარიშეთ კუმულაციური სიხშირეები და ხვედრითი წილები თითოეული ჯგუფისათვის.
ბ. რამდენ პროცენტს არ ჰქონდა დარღვევები უკანასკნელი 12 თვის მანძილზე თითოეული რასობრივი ჯგუფიდან? რამდენ პროცენტს ჰქონდა 2 და მეტი დარღვევა?

13. 2007 წელს გელაპმა ჩაატარა გამოკვლევა საზოგადოების ტოლერანტულობის შესახებ. განსხვავებულ დემოგრაფიულ ჯგუფებს დაუსვეს კითხვა იმის შესახებ, მისაღები იყო თუ არა მათთვის ჰომოსექსუალიზმი, როგორც განსხვავებული ცხოვრების წესი. განიხილეთ თითოეული დემოგრაფიული ჯგუფის მოსაზრება და შეაჯამეთ განსახილველი საკითხი.

	დიახ (%)	არა (%)
სქესი		
კაცი	53	44
ქალი	61	35
ასაკი		
18-34	75	23
35-54	58	39
55+	45	51
პოლიტიკური შეხედულება		
რესპუბლიკელი	36	58
დემოკრატიკელი	60	36

დემოკრატი	72	27
რელიგიურ მომსახურებაზე დასწრება		
ყოველკვირეულად	33	64
დაახლოებით ყოველკვირეულად/ყოველთვიურად	57	40
იშვიათად/არასოდეს	74	22

წყარო: Lydia Saad, „Tolerance for Gay Rights at High-Water Mark", 2007.

14. ამ ცხრილში მოტანილია არჩევნებში მონაწილე ამერიკელების დემოგრაფიული მონაცემები 2008 წლის არჩევნებამდე რამდენიმე დღით ადრე. (დამრგვალების ან გამოუპასუხებლობის გამო ცვლადების ყველა კატეგორიების ჯამები შეიძლება არ უდრიდეს 100-ს).

	ბარაკ ობამა (%)	ჯონ მაკკეინი (%)
სქესი		
კაცი	49	51
ქალი	56	44
ასაკი		
30 წლის ქვემოთ	61	39
30–49	53	47
50–64	54	46
65-ზე ზევით	46	54
რასა		
თეთრკანიანი (ლათინო-ამერიკელების ჩათვლით)	45	55
არათეთრი	90	10
არა-ესპანელი თეთრი	44	56
არათეთრი (ლათინო-ამერიკელების ჩათვლით)	85	15
შავკანიანი	99	1
განათლება		
დაწყებითი სკოლა	67	33
საშუალო სკოლა	46	54
დაუმთავრებელი კოლეჯი	51	49
კოლეჯი	50	50
კოლეჯის შემდგომი	64	36
პროფკავშირები		
დარგობრივი პროფკავშირები	63	37
სამხედრო სამსახური		
ვეტერანი	41	59
სამხედრო შინამეურნეობა	47	53
იარაღის ფლობა		
მფლობელი	36	64
არამფლობელი	63	37

წყარო: Gallup Organizacion, „Blacks, Postgrads, Young Adults Help Obama Prevail", 2008.

ა. თითოეული ცვლადისათვის გაარკვიეთ გაზომვის დონე.
ბ. როგორ დაახასიათებთ ობამას მხარდამჭერებს? მაკკეინის მხარდამჭერებს?

თავი 3

გრაფიკული პრეზენტაცია

თავის სწავლების მიზნები

- წრიული, სვეტებიანი, ხაზოვანი, დროითი მწკრივების დიაგრამისა და ჰისტოგრამის აგება და ინტერპრეტაცია
- დიაგრამებისა და სქემების ანალიზი და ინტერპრეტაცია

ალბათ, გსმენიათ გამოთქმა: „ათჯერ გაგონილს ერთხელ ნანახი სჯობიაო“. იგივე შეიძლება ითქვას სტატისტიკურ გრაფიკებზეც, რადგან მათი საშუალებით ასობით და ათასობით რიცხვების განზოგადებაა შესაძლებელი. ბევრი ადამიანი გაურბის სიხშირული განაწილების ცხრილებისა და სხვა ტაბულარული ფორმით წარმოდგენილი ინფორმაციის გამოყენებას, მაგრამ მათთვის მისაღებია გრაფიკები, რადგან ისინი უფრო ადვილად აღსაქმელი და გასაგებია. გრაფიკები შინაარსს უფრო „ნახატებით“ გადმოგვცემენ, ვიდრე სიტყვებით, ან რიცხვებით. მათი საშუალებით ჩვენი ყურადღების გამახვილება უფრო მეტად საკითხის არსზე ხდება, ვიდრე პრეზენტაციის ტექნიკურ დეტალებზე.

ამ თავში თქვენ შეისწავლით რამდენიმე ყველაზე გავრცელებული გრაფიკის აგების მეთოდებს. ნაკლებ ყურადღებას დავუთმობთ თვით აგების ტექნიკურ მხარეს და უფრო მეტს – თუ როგორ შევარჩიოთ ისინი კონკრეტული ინფორმაციის წარმოსადგენად. აგრეთვე თქვენ ყურადღებას შევაჩერებთ გრაფიკულად წარმოდგენილი ინფორმაციის ინტერპრეტაციაზე.

გრაფიკების აგების მეთოდებთან ერთად, ჩვენ გასწავლით მათ გამოყენებას ინფორმაციის გადმოსაცემად, ანუ „ამბის“ მოსაყოლად. ერთ-ერთი ასეთი ამბავი იქნება აშშ-ის ხანდაზმული მოსახლეობის შესახებ. სხვადასხვა გრაფიკებით შევეცდებით მოსახლეობის დაბერების ასპექტების გადმოცემას უკანასკნელი ორმოცი წლის მანძილზე. ზოგადად, მოხუცებზე ვსაუბრობთ როგორც ჰომოგენურ ჯგუფზე, მაგრამ გრაფიკული მეთოდებით ასახული ინფორმაცია საშუალებას გვაძლევს

ავსახოთ ვარიაციები, რომელიც გვხვდება 65 და ზევით ასაკის მოსახლეობის ეკონომიკურ მახასიათებლებში, საცხოვრებელ პირობებში და ოჯახთან დაკავშირებულ სტატუსებში. ამ თავში წარმოდგენილი სტატისტიკური ინფორმაცია ძირითადად აშშ-ის აღწერის ბიუროსა და სხვა სამთავრობო სააგენტოების მიერ შეგროვილ მასალებს ეყრდნობა როგორც აშშ-ის, ისე სხვა ქვეყნების ხანდაზმულებზე.

გრაფიკული გამოსახვის უამრავი ხერხი არსებობს, თუმცა ჩვენ შევჩერდებით სოციალურ მეცნიერებებში ყველაზე მეტად გავრცელებულ მეთოდებზე. მაგალითად, სვეტებიანი და წრიული დიაგრამები გამოიყენება ნომინალურ და ორდინალურ ცვლადებთან, ჰისტოგრამა და ხაზოვანი დიაგრამები კი – ინტერვალურ-ფარდობით ცვლადებთან. ჩვენ აგრეთვე შევეხებით დროითი მწკრივების სტატისტიკურ სქემებს და რუკებს. სტატისტიკური რუკები ყველაზე ხშირად ინტერვალურ ცვლადებთან გამოიყენება, ხოლო დროითი მწკრივების სტატისტიკური სქემები – დროში ცვალებადი მოვლენების ასახვისათვის.

▣ ᲬᲠᲘᲣᲚᲘ ᲓᲘᲐᲒᲠᲐᲛᲐ: ᲮᲐᲜᲓᲐᲖᲛᲣᲚᲗᲐ ᲠᲐᲡᲐ ᲓᲐ ᲔᲗᲜᲘᲙᲣᲠᲘ ᲜᲘᲨᲐᲜᲘ

აშშ-ის ხანდაზმული მოსახლეობა რასობრივად ჰეტეროგენულია. როგორც ცხრილში 3.1-ია ნაჩვენები, 2008 წლისათვის ხანდაზმულთა (იგულისხმება 65 წლის და ზევით მოსახლეობა) შორის 33.7 მლნ თეთრკანიანი[1] იყო, 3.3 მლნ შავკანიანი, 1.3 მლნ აზიელი ამერიკელი, 36,000 ადგილობრივი ჰავაელი ან წყნარი ოკეანის კუნძულელი და 261,000 – ორი და მეტი რასის ნარევი.

ცხრილი 3.1 აშშ-ის ხანდაზმული მოსახლეობა რასობრივ ჭრილში, 2008

რასა	სიხშირე (f)	პროცენტი (%)
თეთრკანიანები	33,737,000	86.8
შავკანიანები	3,314,000	8.5
ამერიკელი ინდიელები	225,000	0.6
აზიელები	1,294,000	3.3
ჰავაელები და კუნძულელები	36,000	0.1
ორი ან მეტი რასის კომბინაცია	261,000	0.7

წყარო: U.S. Census Bureau, Statistical Abstract of the United States, 2010, Table 10.

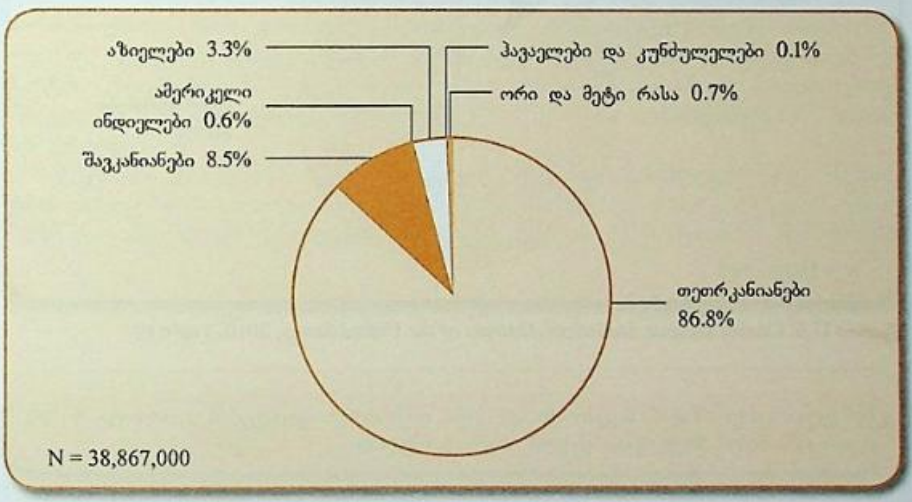
წრიული დიაგრამა გვიჩვენებს განსხვავებებს ნომინალური ან ორდინალური ცვლადების სიხშირეებს ან ხვედრით წილებს შორის. მასში კატეგორიები სეგმენტების სახითაა წარმოდგენილი, რომელთა ჯამი 100%-ის ტოლია. ნახაზზე 3.1 იგივე მონაცემებია წარმოდგენილი, რაც ცხრილში 3.1. მიუხედავად იმისა, რომ ამ

მონაცემების ნახვა შესაძლებელია ცხრილში 3.1, წრიული დიაგრამიდან უფრო ადვილია ინფორმაციის ინტერპრეტაცია. მასში ნაჩვენებია, რომ ასაკოვან მოსახლეობაში თეთრკანიანებს დომინანტური ადგილი უკავიათ (86.6%) შავკანიან მოსახლეობასთან შედარებით (8.5%).

წრიული დიაგრამა ყველა იმ ინფორმაციას შეიცავს, რასაც სიხშირული განაწილება. მას აქვს საიდენტიფიკაციო ნომერი, შინაარსის ამსახველი სათაური და მითითებულია საინფორმაციო წყარო. სიხშირე ან ხვედრითი წილები წარმოდგენილია როგორც ვიზუალურად, ისე ციფრებით.

✓ *ცოდნის შემოწმება*

ნახაზი 3.1 65 წლის და ზევით აშშ-ის მოსახლეობის წლიური მაჩვენებლები რასობრივ ჭრილში, 2008



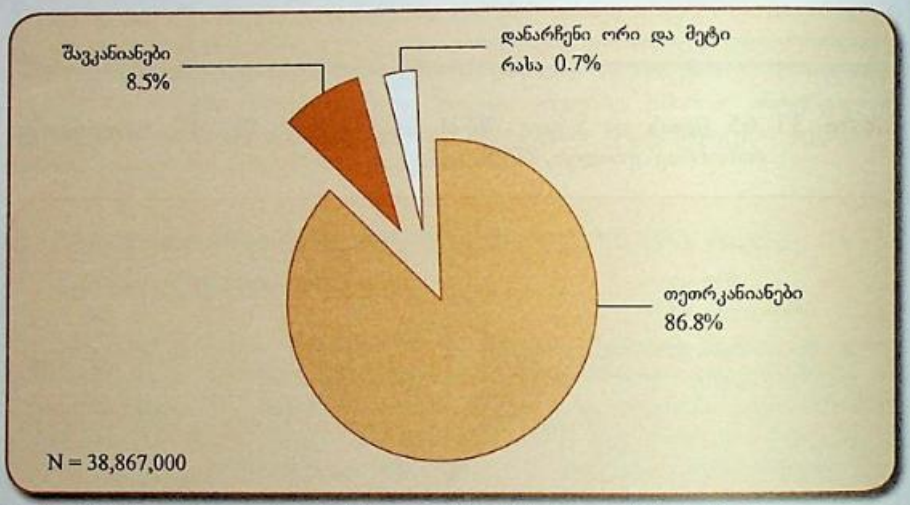
წყარო: U.S. Census Bureau, *The 2010 Statistical Abstract*, Table 10.

წრიული დიაგრამა გვიჩვენებს განსხვავებებს ნომინალური ან ორდინალური ცვლადების სიხშირეებს ან ხვედრით წილებს შორის. მასში კატეგორიები სეგმენტების სახითაა წარმოდგენილი, რომელთა ჯამი 100%-ის ტოლია.

როგორც ხედავთ, რამდენიმე რასობრივი ჯგუფის პროცენტების ჯამი დაახლოებით 3%-ია. უკეთესი იქნება, თუ ამერიკელ ინდიელებს, აზიელებსა და ჰავაელებს ერთ ჯგუფში გავაერთიანებთ — „სხვა რასები“. დაგვრჩება მხოლოდ სამი ძირითადი კატეგორია: თეთრკანიანები, შავკანიანები და სხვა რასები. განახლე-

ბული წრიული დიაგრამა 3.2 აგრეთვე მიღებულია ცხრილი 3.1-ის საფუძველზე. თუ სეგმენტებს წრიდან ცოტა გამოვწევთ, ყურადღება მეტად გამახვილდება ასაკოვანი მოსახლეობის მრავალფეროვნებაზე.

ნახაზი 3.2 აშშ-ის 65 წლის და ზევით ასაკის მოსახლეობის წლიური მაჩვენებლები რასობრივ ჭრილში, 2008



წყარო: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 10.

ცოდნის შემოწმება

წრიული დიაგრამიდან შავკანიანი და სხვა რასობრივი ჯგუფების გამოყოფა მოხდა თეთრკანიანების ხვედრითი წილის ხაზგასასმელად.

სვეტებიანი დიაგრამა: ხანდაზმულთა საცხოვრებელი პირობები და ეკონომიკური აქტივობა

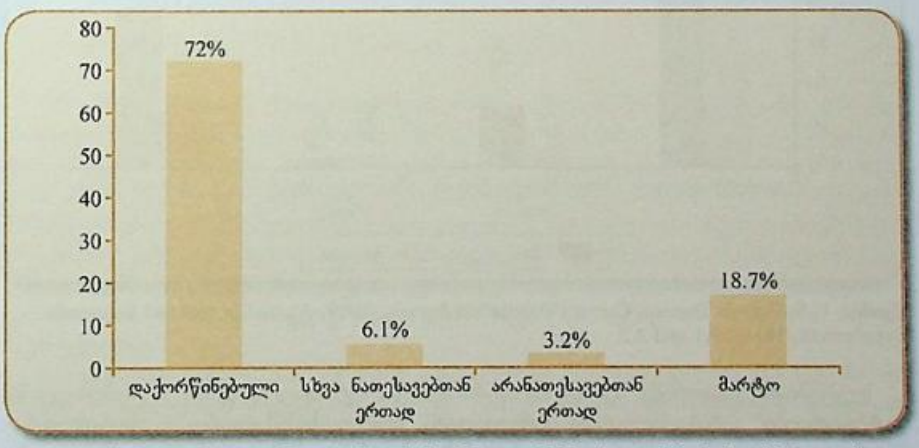
სვეტებიანი დიაგრამა ნომინალური და ორდინალური ცვლადების წარმოდგენის ალტერნატული საშუალებაა. ის გვიჩვენებს განსხვავებებს ნომინალური და ორდინალური ცვლადების სიხშირეებსა ან ხვედრით წილებს შორის. კატეგორიები წარმოდგენილია თანაბარი სიგანის მართკუთხედების სახით, რომელთა სიმაღლე სიხშირის, ან ხვედრითი წილის პროპორციულია.

სვეტებიანი დიაგრამა გვიჩვენებს განსხვავებებს ნომინალური და ორდინალური ცვლადების სიხშირეებსა, ან ხვედრით წილებს შორის. კატეგორიები წარმოდგენილია თანაბარი სიგანის მართკუთხედების სახით, რომელთა სიმაღლე სიხშირის, ან ხვედრითი წილის პროპორციულია.

მოვახდინოთ სვეტებიანი დიაგრამის საშუალებით ხანდაზმულ ადამიანთა საცხოვრებელი პირობების თვალსაჩინო ასახვა, რომელიც ასაკის მატებასთან ერთად მნიშვნელოვნად იცვლება – ხანდაზმულთა სულ უფრო მეტი რაოდენობა ცხოვრობს მარტო, ან ნათესავებთან ერთად. ნახაზზე 3.3 წარმოდგენილია ხანდაზმული მამაკაცების პროცენტული განაწილება საცხოვრებელი პირობების მიხედვით 2009 წელს. ამ დიაგრამის ინტერპრეტირებაც წრიული დიაგრამის მსგავსად ხდება, თუმცა კატეგორიები აბსცისთა ღერძზეა განლაგებული ჰორიზონტალურად (მათი განლაგება შესაძლებელია ვერტიკალურადაც, ორდინატთა ღერძზე). დიაგრამა გვიჩვენებს, რომ 2009 წელს ხანდაზმული მამაკაცების 18.7% მარტო ცხოვრობდა, 72% დაქორწინებული იყო და მეუღლესთან ერთად ცხოვრობდა, 6.1% ნათესავებთან ერთად, ხოლო 3,2% – არანათესავებთან ერთად ცხოვრობდა.

სვეტებიანი დიაგრამის აგებისას, უპირველეს ყოვლისა, სახელები უნდა მივანიჭოთ კატეგორიებს ჰორიზონტალური ღერძის გასწვრივ. თითოეულ კატეგორიას თანაბარი სიგანის და განსხვავებული სიგრძის ერთი მართკუთხედი შეესაბამება.

ნახაზი 3.3 მამაკაცების (65 წლის და ზევით) საცხოვრებელი პირობები აშშ-ში, 2009



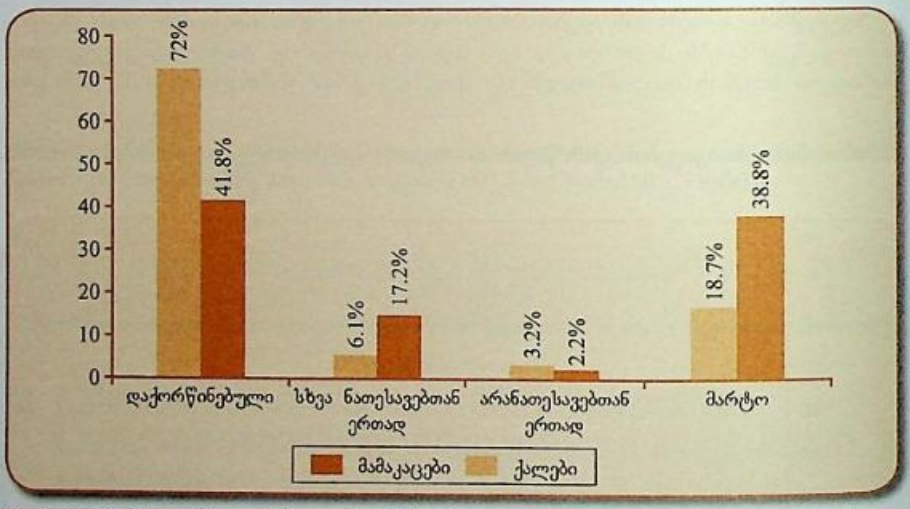
წყარო: U.S. Census Bureau, Current Population Survey, 2009, Annual Social and Economic Sypplement, Tables A1 and A2.

მისი სიგრძე სიხშირის ან ხვედრითი წილის პროპორციულია. აღსანიშნავია, რომ სვეტები ერთმანეთისგან დაშორებულია, რაც იმაზე მიანიშნებს, რომ კატეგორიები ნომინალურია.

სვეტებიანი დიაგრამები ხშირად გამოიყენება ერთი ან რამდენიმე კატეგორიის შედარებისათვის სხვადასხვა ჯგუფებში. მაგალითად, დიდია ალბათობა, რომ ხანდაზმულობაში ქალები მარტო იცხოვრებენ. საცხოვრებელი პირობებისათვის გენდერული განსხვავების მთავარი ფაქტორი ქალთა სიცოცხლის ხანგრძლივობაა.[2] ამასთან, ხანდაზმულ მამაკაცებს განმეორებითი ქორწინების უფრო დიდი შანსი აქვთ, ვიდრე ქალებს. გარდა ამისა, ხანდაზმული ქალების ახალ თაობას უფრო განვითარებული სოციალური დაცვის მექანიზმები გააჩნია.[3]

ვთქვათ, გვსურს ვაჩვენოთ, თუ როგორ განსხვავდება ერთმანეთისგან მამაკაცებისა და ქალების საცხოვრებელი პირობები. ნახაზზე 3.4 მოცემულია 2009 წლისათვის 65 წლისა და ზევით ასაკის მარტო ან სხვებთან ერთად მცხოვრები ქალებისა და მამაკაცების ხვედრითი წილების შედარება. აშკარაა, რომ ქალებისათვის მარტო ცხოვრების ალბათობა უფრო მაღალია.

ნახაზი 3.4 ხანდაზმულთა საცხოვრებელი პირობები აშშ-ში გენდერულ ჭრილში, 2009

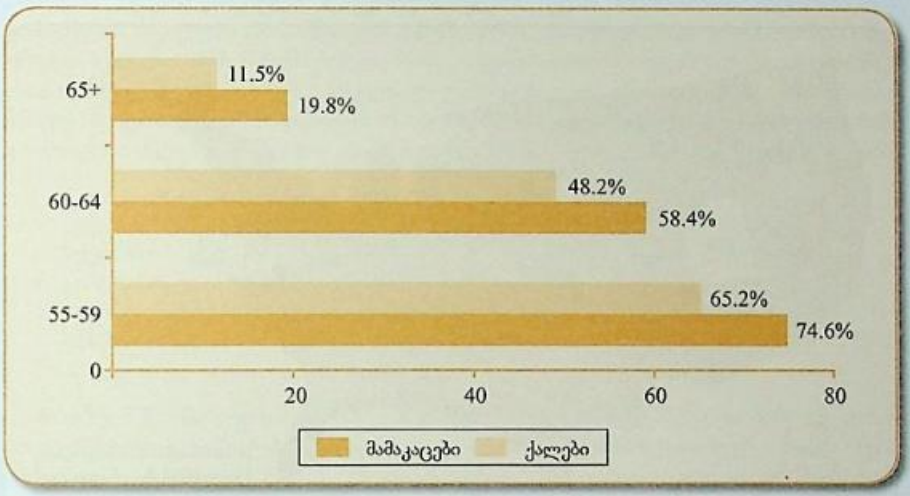


წყარო: U.S. Census Bureau, Current Population Survey, 2009, Annual Social and Economic Supplement, Tables A1 and A2.

სვეტებიანი დიაგრამა შეიძლება ჰორიზონტალურადაც ავაგოთ. ამ შემთხვევაში კატეგორიები ვერტიკალურად განლაგდება, ხოლო სიხშირეები — ჰორიზონტალურად. ეს ფორმატი ნაჩვენებია ნახაზზე 3.5, სადაც შედარებულია 55 წლისა და ზევით ასაკის ქალებისა და მამაკაცების შრომითი აქტივობა 2008 წელს.

აშკარაა, რომ ყველა ასაკობრივ კატეგორიაში მამაკაცები მეტად არიან დასაქმებული, ვიდრე ქალები.

ნახაზი 3.4 55 წლის და ზევით დასაქმებული მამაკაცები და ქალები, 2008



წყარო: U.S. Census Bureau, American Community Survey 2006-2008, Table B23001.

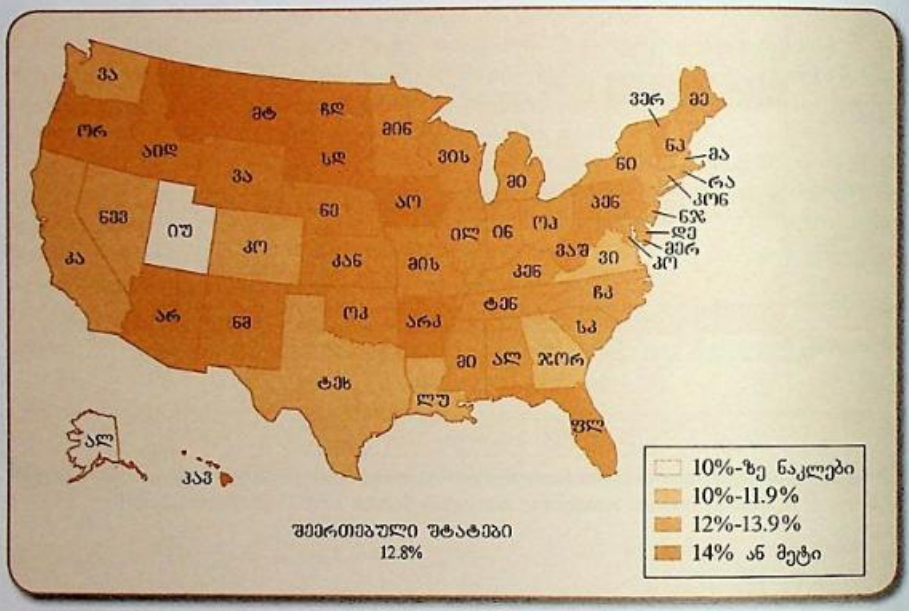
სტატისტიკური რუკა: ხანდაზმულების გეოგრაფიული განაწილება

60-იანი წლებიდან ხანდაზმულები ძირითადად აშშ-ის სამხრეთსა და აღმოსავლეთში არიან თავმოყრილნი. პროგნოზის თანახმად, 2020 წლისათვის აქ მათი ხვედრითი წილი 80%-მდე გაიზრდება (უკანასკნელი აღწერის მასალებით, 2008 წელს დაწყებულმა რეცესიამ შეანელა ეს პროცესი). ამერიკული საზოგადოების ამ მნიშვნელოვან გეოგრაფიულ ცვლილებებს ჩვენ გამოვსახავთ რუკებით (კარტოგრამებით). კარტოგრამები განსაკუთრებით კარგად გეოგრაფიულ ჭრილში ასახავს ისეთი ცვლადების ვარიაციას, როგორიცაა მოსახლეობის განაწილება, საარჩევნო გეოგრაფია, დამნაშავეობის დონე ან სამუშაო ძალის სტრუქტურა.

დააკვირდით ნახაზს 3.6. აქ ოთხ კატეგორიად წარმოდგენილია 65 წლისა და ზევით მოსახლეობის ხვედრითი წილები აშშ-ის თითოეული შტატის მიხედვით 2008 წლისათვის: 10%-ზე ნაკლები, 10%-იდან 11.9%-მდე, 12%-იდან 13.9%-მდე და 14% და მეტი. თითოეული კატეგორია განსხვავებული ფერით არის წარ-

მოდგენილი და შტატებიც შესაბამისად არის შეფერადებული. იმისათვის, რომ კარტოგრამა ადვილი წასაკითხი და აღსაქმელი იყოს, ნუ გაზრდით ძალიან კატეგორიების რაოდენობას, ხუთი კატეგორია სავსებით საკმარისია.

ნახაზი 3.6 ხანდაზმული მოსახლეობის ხვედრითი წილი, 2008



წყარო: U.S. Census Bureau, The 2010 Statistical Abstract, Table 16.

კარტოგრამების ობიექტი შეიძლება გახდეს ქალაქები, ქვეყნები, ქალაქის ცალკეული რაიონები, სააღწერო უბნები და სხვა ობიექტები. გეოგრაფიული არეალის მასშტაბი კვლევის საკითხზეა დამოკიდებული.

ცოდნის შემოწმება

მოიფიქრეთ, კიდევ რა მონაცემების აღწერაა შესაძლებელი სტატისტიკური კარტოგრამების გამოყენებით. მოიყვანეთ მაგალითები. ქვეყნის მასშტაბით, რა მონაცემებს აღწერდით კარტოგრამებით?

ჰისტოგრამა

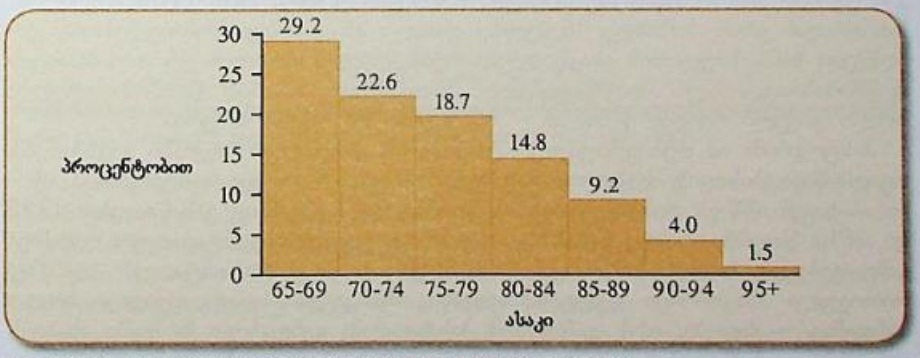
ჰისტოგრამა გამოიყენება ინტერვალური ცვლადების სიხშირეებისა და ხვედრითი წილების განსხვავებათა გამოსახვის მიზნით. კატეგორიები გამოსახულია ერთმანეთზე მიწყობილი სვეტების სახით, რომელთა სიგანე კატეგორიის მნიშვნელობას

შეესაბამება, ხოლო სიმაღლე – სიხშირეს. ჰისტოგრამა ძალიან წააგავს სვეტებიან დიაგრამას, იმ განსხვავებით, რომ სვეტები ერთმანეთს ეხება და მათი სიგანეც განსხვავებული შეიძლება იყოს. სვეტებიან დიაგრამაში დაცილებული სვეტები იმის მანიშნებელია, რომ ეს განსხვავებული კატეგორიებია. მაგალითად, *სტატუსი ქორწინებაში* (დაქორწინებული, დაუქორწინებელი), *სქესი* (მამრობითი, მდედრობითი), *დასაქმების სტატუსი* (დასაქმებული, უმუშევარი). ჰისტოგრამაში ერთმანეთზე მიწყობილი სვეტები იმის მანიშნებელია, რომ კატეგორიები ან ინტერვალები დალაგებულია დაბლიდან მაღლისაკენ გარკვეული რიგითობით და მნიშვნელობით. მაგალითად, *სწავლაზე დახარჯული საათების რაოდენობა, ასაკი, მიღებული განათლების ხანგრძლივობა წლებში* – ყველა ეს მომიჯნავე, რიგობით დალაგებული ინტერვალებია.

ჰისტოგრამა გამოიყენება ინტერვალური ცვლადების სიხშირეებისა და ხვედრითი წილების განსხვავებათა გამოსახვის მიზნით. კატეგორიები გამოსახულია ერთმანეთზე მიწყობილი სვეტების სახით, რომელთა სიგანე კატეგორიის მნიშვნელობას შეესაბამება, ხოლო სიმაღლე – სიხშირეს.

ნახაზი 3.7 ჰისტოგრამაა, რომელზეც გამოსახულია ხანდაზმული მოსახლეობის პროცენტული განაწილება ასაკის მიხედვით. ჰისტოგრამაზე ასახული მონაცემები მოტანილია ცხრილიდან 3.2. ამ ჰისტოგრამის ასაგებად ასაკობრივი ინტერვალები დაიტანეთ აბსცისთა ღერძზე, ხოლო ხვედრითი წილები – ვერტიკალურზე. თითოეული სვეტის სიგანე ასაკობრივი ინტერვალის სიდიდეს შეესაბამება, ხოლო სიმაღლე იმაზე მოგვანიშნებს, თუ მოსახლეობის რა პროცენტი მოხვდა აღნიშნულ ასაკობრივ ინტერვალში. ჰისტოგრამის თითოეული სვეტი ერთმანეთთან მიჯრითაა განლაგებული, რაც იმაზე მიანიშნებს, რომ ისინი მომიჯნავე კატეგორიებს ასახავენ.

ნახაზი 3.7 აშშ-ის ხანდაზმული მოსახლეობის ასაკობრივი განაწილება, 2008



წყარო: U.S. Census Bureau, The 2010 Statistical Abstract, Table 10.

ცხრილი 3.2 აშშ-ის ხანდაზმული მოსახლეობის პროცენტული განაწილება, 2008

ასაკი (წლები)	ხვედრითი წილები (%)
60-69	29.2
70-74	22.6
75-79	18.7
80-84	14.8
85-89	9.2
90-94	4.0
95+	1.5
სულ (N)	100.0 (38,870,000)

ცოდნის მოწმება

როდესაც სვეტებიანი დიაგრამა და ჰისტოგრამა ერთი ცვლადის კატეგორიების სიხშირეების ასახვისთვის გამოიყენება, კატეგორიები აბსცისთა ღერძზე გამოისახება, ხოლო სიხშირეები — ორდინატთა ღერძზე. ჰორიზონტალური სვეტებიანი დიაგრამის ან ჰისტოგრამის დროს ეს პირიქით ხდება.

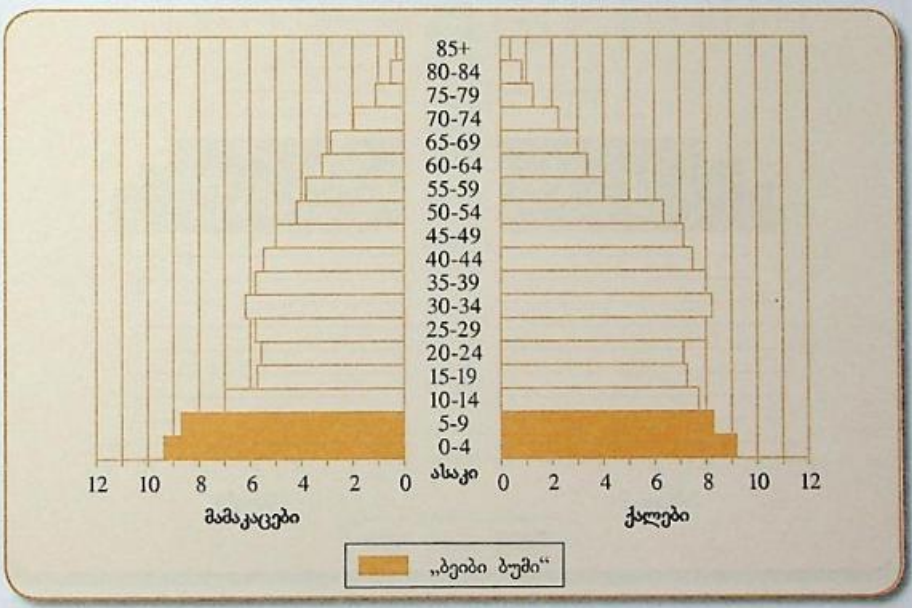
სტატისტიკა პრაქტიკაში: ამერიკის „დაბერება“

ჰისტოგრამის გამოყენება უფრო კომპლექსური ტრენდების ასახვისთვისაა მიზანშეწონილი, როგორიცაა მაგალითად, ამერიკის „დაბერება“. ერთი წუთით წარმოიდგინეთ შემდეგი ტრენდები: ხანდაზმული მოსახლეობა 10-ჯერ უფრო მეტია, ვიდრე 1900 წელს და ის ორჯერ მეტი გახდება 2030 წლისათვის. როგორც ერთმა ჟურნალისტმა აღნიშნა, ავტომობილი რომ კოლონიალიზმის ეპოქაში გამოეგონათ, ახალი მიწის რეზიდენტების ნახევარზე მეტი ამით ვერ ისარგებლებდა: ორიდან ერთი მოსახლე 16 წელზე დაბალი ასაკისა იყო. უმრავლესობა ვერ აღწევდა ხანდაზმულობის ასაკს. დღეისათვის ყოველი ოთხიდან ერთი — მართვის მოწმობისათვის ჯერ მოუმწიფებელი ახალგაზრდა თაობის წარმომადგენელია, ხოლო ყოველი რვიდან ერთი კი — 65 წლისა და უფრო ხნიერია.[4]

ჰისტოგრამა ამ დემოგრაფიული ტრენდების ვიზუალურ აღქმაში გვეხმარება. თვალსაჩინოებისთვის შევხედოთ ნახაზებს 3.8 და 3.9. ორივე ჰისტოგრამაა. ისინი ასახავენ აშშ-ის მოსახლეობის ასაკობრივ და სქესობრივ განაწილებას 1955 და 2010 წლებში. ორივე ნახაზზე ასაკობრივი ჯგუფები ვერტიკალურ ღერძზეა განლაგებული, ხოლო სიხშირეები (მლნ კაცი) — ჰორიზონტალურ ღერძზე. თითოეული ასაკობრივი ჯგუფის მარცხნივ მამაკაცებია დაჯგუფებული, ხოლო მარჯვნივ — ქალები. იმის გამო, რომ ჰისტოგრამა გენდერულ ჭრილში ასახავს ასაკობრივ განაწილებას, მას ასაკობრივ-სქესობრივ პირამიდასაც უწოდებენ.

ვიზუალურად შევადაროთ ერთმანეთს ნახაზებზე წარმოდგენილი მონაცემების სხვადასხვა ნაწილები. ნახაზზე ასაკობრივი ჯგუფების კონცენტრაცია დროთა განმავლობაში ასაკობრივ განაწილებაში არსებულ მთავარ მოძენტებზე მეტყველებს. ორივე ნახაზზე შენიშნავდით განსხვავებულ ფორმებსაც. თუ 1955 წელს ყველაზე დიდი ასაკობრივი ჯგუფი იყო 0-დან 9 წლამდე, 2010-ში უდიდესი ჯგუფი უკვე 45-დან 54 წლამდეა. ეს დრამატული ცვლილებები გამოიწვია „ბეიბი ბუმ"-ის (დემოგრაფიული აფეთქება) თაობის (1946-1965) „დაბერებამ". ამ წლებში ამერიკაში 84 მილიონი ბავშვი დაიბადა, რაც 60%-ით სჭარბობს შემდეგ 20-წლეულში დაბადებულთა რაოდენობას. 2010 წლისათვის, როდესაც ეს თაობა 45-64 წლის გახდა, საშუალო და ხანდაზმული მოსახლეობის რიცხოვნობამ საგრძნობლად მოიმატა.

ნახაზი 3.8 აშშ-ის მოსახლეობა სქესისა და ასაკის მიხედვით, 1955 (მლნ)



წყარო: U.S. Bureau of Census, *Current Population Reports*, p 23-178, 1992.

დააკვირდით აგრეთვე მამაკაცებისა და ქალების რიცხოვნობას ასაკის მატებასთან ერთად. განსხვავებები განსაკუთრებით მკვეთრია ნახაზზე 3.9. მაგალითად, 70-დან 74 წლამდე ასაკებში ქალები სჭარბობენ მამაკაცებს პროპორციით 5:4.2; 85 წლისა და ზევით ასაკებში კი – 2:1. ეს სხვაობა ნიშნავს, რომ ყოველ ასაკში მამაკაცთა მოკვდაობა უფრო მაღალია, ვიდრე ქალებისა.

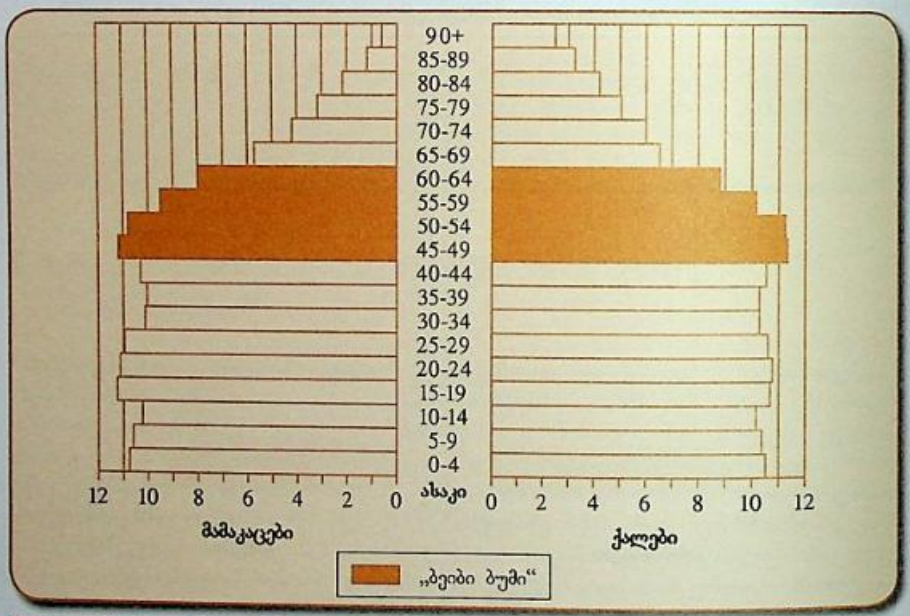
მიუხედავად იმისა, რომ დღეს სახეზეა ამერიკის „დაბერება", ახლო მომავლის პროგნოზებით, ე.წ. Y თაობა (1980-1995 წლებში დაბადებულები) გადააჭარბებს

„ბეები ბუმ“-ის თაობას 2015 წლისათვის. 2015 წლის პროგნოზები წარმოდგენილია ნახაზზე 3.10. ამ ფენომენის უკან ემიგრანტთა დიდი ტალღა დგას – ლეგალური და არალეგალური მიგრანტები, რომლებიც აშშ-ში 1975 წლიდან შემოვიდნენ. მათი შვილები შეუერთდნენ *Y* თაობას და თავიანთი წვლილი შეიტანეს დაბერებული ბუმერების გამოწვევაში.[5]

✓ ცოდნის შემოწმება

დაიმახსოვრეთ, როდესაც ჰისტოგრამას ვიყენებთ ჯგუფების შედარებისთვის, ჰისტოგრამა უნდა ვაჩვენოთ თითოეული ჯგუფისათვის (იხ. ნახაზები 3.8, 3.9, და 3.10). თუ ჩვენ ვადარებთ ჯგუფებს სვეტებიან დიაგრამაზე, აქ შესაძლებელია ორი ან მეტი ჯგუფის შედარება (იხ. ნახაზი 3.4).

ნახაზი 3.9 აშშ-ის მოსახლეობა სქესისა და ასაკის მიხედვით, 2010 (მლნ)



წყარო: U.S. Census Bureau, *The 2010 Statistical Abstract, Resident Population Projections by Sex and Age: 2010 to 2050*, Table 8.

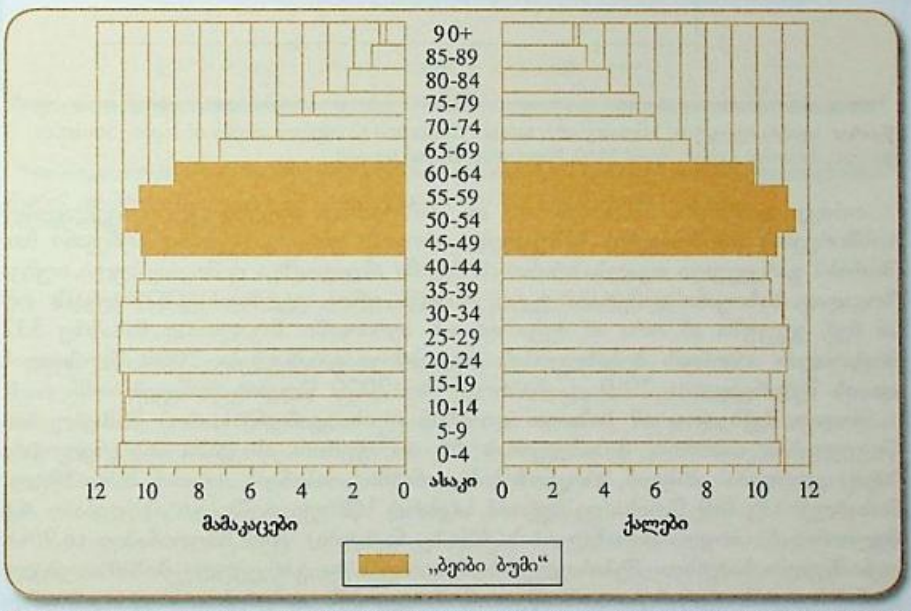
▣ საზოვანი დიაგრამა

ხანდაზმული ადამიანების რიცხოვნობის ზრდა მთელ მსოფლიოში შეინიშნება, როგორც განვითარებულ, ისე განვითარებად ქვეყნებში. 1994 წლისათვის მსოფლიოს 30 ქვეყანაში ხანდაზმულთა რიცხოვნობამ 2 მილიონს გადააჭარბა, 2020 წლის პროგნოზებით ასეთი ერების რიცხვი 55-მდე გაიზრდება. იაპონია

მძაფრად განიცდის ხანდაზმულთა მატებას. ნახაზზე 3.11 ხაზოვანი დიაგრამით მოცემულია იაპონიის მოსახლეობის განაწილება ასაკის მიხედვით.

ხაზოვანი დიაგრამა ინტერვალურ-ფარდობითი ცვლადების წარმოდგენის კიდევ ერთი ხერხია. წერტილები, რომლებიც თითოეული კატეგორიის სიხშირეს შეესაბამება, ამ კატეგორიის შუა წერტილის თავზე თავსდება და შემდეგ სწორი ხაზით ერთდება. დააკვირდით, ნახაზზე 3.11 ასაკობრივი ინტერვალები ჰორიზონტალურ ღერძზეა მოთავსებული, ხოლო სიხშირეები – ვერტიკალურზე. სვეტების ნაცვლად, აქ სიხშირეების წარმოდგენისათვის გამოყენებულია წერტილები, რომლებიც ინტერვალების შუა წერტილის გასწვრივ და ზემოთ თავსდება. შემდეგ ეს წერტილები ხაზებით ერთდება.

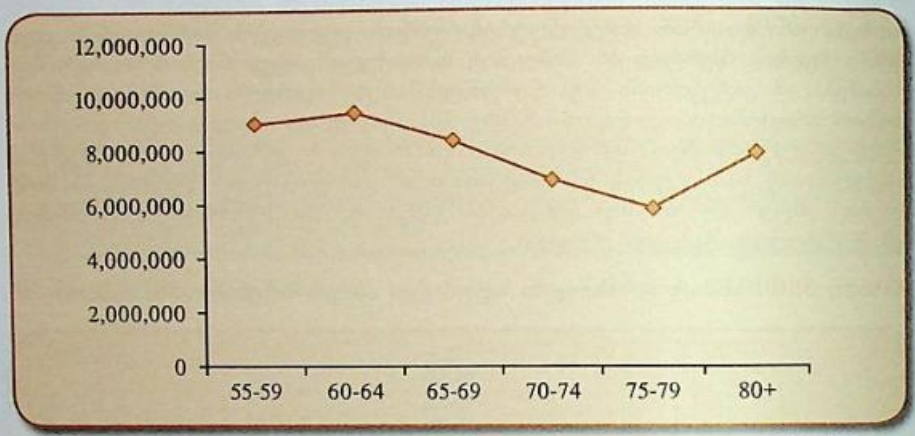
ნახაზი 3.10 აშშ-ის მოსახლეობა სქესისა და ასაკის მიხედვით, 2015 (მლნ)



წყარო: U.S. Census Bureau, *The 2010 Statistical Abstract, Resident Population Projections by Sex and Age: 2010 to 2050*, Table 8.

ხაზოვანი დიაგრამა გვიჩვენებს განსხვავებებს ინტერვალურ-ფარდობითი ცვლადების სიხშირეებსა და ხვედრით წილებში. წერტილები, რომლებიც თითოეული კატეგორიის სიხშირეს შეესაბამება, ამ კატეგორიის შუა წერტილის თავზე თავსდება და შემდეგ სწორი ხაზით ერთდება.

ნახაზი 3.11 იაპონიის მოსახლეობა 55 წლის ასაკში და ზევით, 2009



წყარო: ადაპტირებულია Ministry of Internal Affairs and Communications of Japan, Statistics Bureau, *Monthly Report April 2010*, Population Estimates-დან.

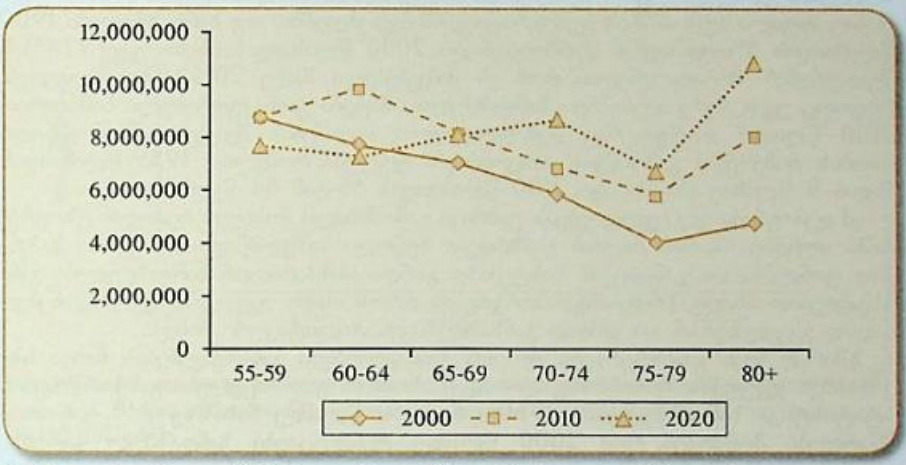
ორივე დიაგრამა, ჰისტოგრამაც და ხაზოვანიც, ინტერვალური ცვლადების სიხშირეების წარმოდგენის საშუალებაა. როგორ უნდა გავაკეთოთ არჩევანი მათ შორის? გარკვეული თვალსაზრისით, არჩევანი ინდივიდზეა დამოკიდებული, თუმცა ზოგადად, ხაზოვანი დიაგრამა უკეთ ასახავს ერთი ცვლადის განაწილებას ორ ან მეტ ჯგუფში ან ორი ან მეტი დროის პერიოდში. მაგალითად, ნახაზზე 3.12 მოტანილია იაპონიის მოსახლეობის ასაკობრივი განაწილება 2000 წლისათვის და ის შედარებულია 2010 და საპროგნოზო 2020 წლების მონაცემებთან.

მოდით, შევხედოთ ამ ხაზოვან დიაგრამას. ის გვიჩვენებს, რომ მომავალ ათწლეულებში იაპონიის მოსახლეობას 55 და უფროს ასაკებში მნიშვნელოვანი ზრდა ელოდება. ამასთან, პროგნოზების თანახმად, იაპონიის ძალიან ხანდაზმული მოსახლეობაც (80 წლისა და ზევით), საკმაოდ სწრაფი ტემპებით იზრდება – 4.8 მილიონიდან (მთელი მოსახლეობის 4%-ზე ნაკლები) 10.8 მილიონამდე (8.9%). აღნიშნულმა საპენსიო შემოსავლების შემცირება და გარკვეული მოსამზადებელი ღონისძიებების გატარება გამოიწვია მოსახლეობის დაბერების სოციალური და ეკონომიკური ეფექტების თავიდან ასაცილებლად.[6]

ცოდნის შემოწმება

დააკვირდით ხაზოვან დიაგრამას ნახაზზე 3.12, სადაც მოტანილია 2010 და 2020 წლების შედარებები. როგორ დაახასიათებთ იაპონიის ხანდაზმული მოსახლეობის მატებას?

ნახაზი 3.12 იაპონიის მოსახლეობა 55 წლის ასაკში და ზევით, 2000, 2010 და 2020 წლებში



წყარო: ადაპტირებულია U.S. Census Bureau, Center for International Research, International Database, 2007-დან.

დროითი მწკრივების დიაგრამები

ჩვენ ხშირად გვაინტერესებს მოვლენების ცვლილება დროში. მაგალითად, შეიძლება დაგვაინტერესოს უკანასკნელი ათწლეულის განმავლობაში ლათინოსი მოსახლეობის შრომითმა აქტივობამ, ან როგორ იცვლება საზოგადოების დამოკიდებულება აბორტებთან მიმართებაში, ან თუნდაც ქორწინებებისა და განქორწინების დონეები. დროითი მწკრივების დიაგრამა ცვლადის განსხვავებებს დროის სხვადასხვა მომენტებში ასახავს. ის შეიცავს ორ ცვლადს: (1) დროს, რომელიც ჰორიზონტალურ ღერძზე დაიტანება და (2) ჩვენთვის საინტერესო ცვლადს (სიხშირეებს, ხვედრით წილებს ან დონეებს), რომელიც ვერტიკალურ ღერძზე გადაიზომება. დროითი მწკრივების დიაგრამის ასაგებად დროის თითოეული პერიოდისათვის ცვლადების მნიშვნელობების აღსანიშნავად წერტილების სერია გამოიყენეთ და ისინი სწორი ხაზებით შეაერთეთ.

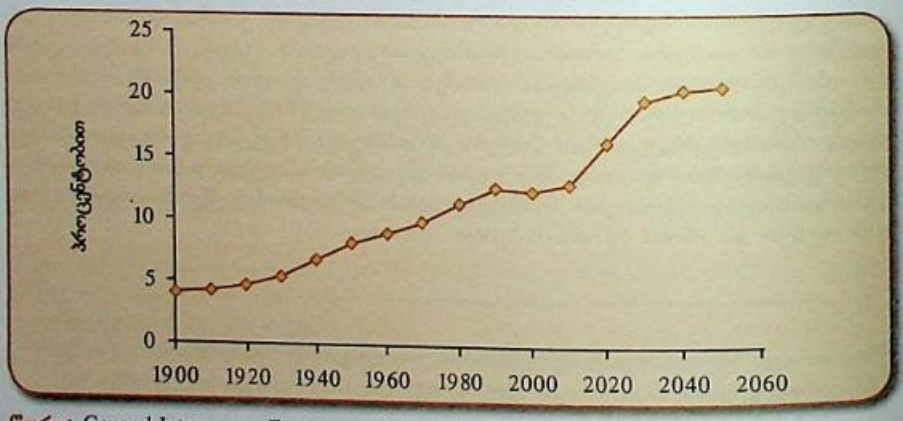
დროითი მწკრივების დიაგრამა ასახავს ცვლადის განსხვავებებს დროის სხვადასხვა მომენტებში. ის გვიჩვენებს დროს (იზომება დროის ერთეულებში როგორიცაა წელი, თვე და ა.შ.) აბსცისთა ღერძზე და სიხშირეს (ხვედრით წილებს, დონეებს) ორდინატთა ღერძზე.

ნახაზზე 3.13 ნაჩვენებია აშშ-ის ხანდაზმული მოსახლეობის ხვედრითი წილების დროითი მწკრივები 1900 წლიდან 2050 წლამდე (პროგნოზი). ეს მწკრივები საშუალებას გვაძლევს უფრო თვალსაჩინოდ დავინახოთ ის დრამატული ცვლილებები, რასაც ადგილი აქვს აშშ-ის დემოგრაფიულ მაჩვენებლებში. თუ ხანდაზმულები 1900 წლისათვის 5%-ით იყვნენ წარმოდგენილი, 2000 წლისათვის, ისინი უკვე 12.4%-ს შეადგენდნენ. მოსალოდნელია, რომ ეს მაჩვენებელი მალე 20%-ს მიუახლოვდეს. როგორც უკვე ადრე აღვნიშნეთ, ხანდაზმულთა მატება განსაკუთრებული სიმძაფრით 2010 წლიდან დაიწყო, რაც დემოგრაფიული აფეთქების შედეგად წარმოქმნილი თაობის დაბერებამ გამოიწვია. ასაკობრივმა ჯგუფმა, რომელიც 1955 წლისათვის 0-დან 9 წლამდე ასაკში იყო, 2010 წლისათვის 55-დან 64 წელს მიაღწია.

ამ დემოგრაფიულ ცვლილებებს უამრავი გამოძახილი მოსდევს რეალურ ცხოვრებაში. თუნდაც ის, რომ ძალიან გაიზრდება ზეწოლა სახელმწიფო და კერძო საპენსიო ფონდებსა და ჯანდაცვის სისტემაზე. გარდა ამისა, რადგან ხანდაზმულები განსხვავდებიან ახალგაზრდებისგან, არჩევნების დროს ისინი უკვე გასათვალისწინებელ ძალად გვევლინებიან და დიდად განსაზღვრავენ პოლიტიკურ ამინდს.

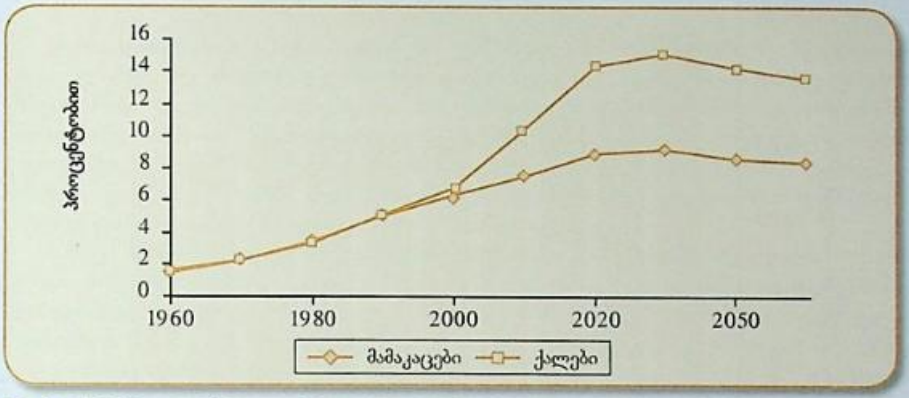
ხშირად ჩვენ გვაინტერესებს დროთა განმავლობაში განსხვავებების ნახვა სხვადასხვა ჯგუფებს შორის. მაგალითად, ვნახოთ როგორ იცვლება განქორწინების სტატისტიკა ხანდაზმულებში 1960-დან 2050 წლამდე ნახაზზე 3.14. დროითი მწკრივები მოწმობენ, რომ 2000 წლამდე ტენდენციები ხანდაზმულ ქალებსა და მამაკაცებს შორის მსგავსი იყო. ორივე ჯგუფის ხვედრითი წილი თითქმის 2%-იდან (1960 წელი) 5%-მდე გაიზარდა 1990 წლისათვის.[7] პროგნოზების თანახმად, მნიშვნელოვანი ზრდაა მოსალოდნელი როგორც მამაკაცების, ისე განსაკუთრებით ქალების განქორწინების სტატისტიკაში: კერძოდ, 5%-იდან 8.4%-მდე მამაკაცებში და 13.6%-მდე ქალებში. ეს მკვეთრი გენდერული ცვლილება გარკვევითაა ნაჩვენები ნახაზზე 3.14.

ნახაზი 3.13 აშშ-ის მოსახლეობის პროცენტული განაწილება 65 და ზევით ასაკებში, 1990–2050



წყარო: General Interagency Forum on Aging Related Statistics, *Older Americans 2004: Key Indicators of Well Being*, 2004.

ნახაზი 3.14 აშშ-ის განქორწინებული მოსახლეობის პროცენტული განაწილება 65 და ზევით ასაკებში გენდერულ ჭრილში, 1960–2040



წყარო: U.S. Bureau of Census, *65+ in America, Current Population Reports, Special Studies*, P23-190, Table 6-1.

რით განსხვავდება დროითი მწკრივების დიაგრამა ხაზოვანი დიაგრამისაგან? განსხვავება იმაშია, რომ თუ ხაზოვან დიაგრამაზე ასახულია მხოლოდ ერთი ცვლადის სიხშირული განაწილების ცვლილებები, დროითი მწკრივების დიაგრამაზე — ორი ცვლადია ასახული. ამასთან, დრო ყოველთვის ერთ-ერთი ცვლადია დროითი მწკრივების დიაგრამაზე.

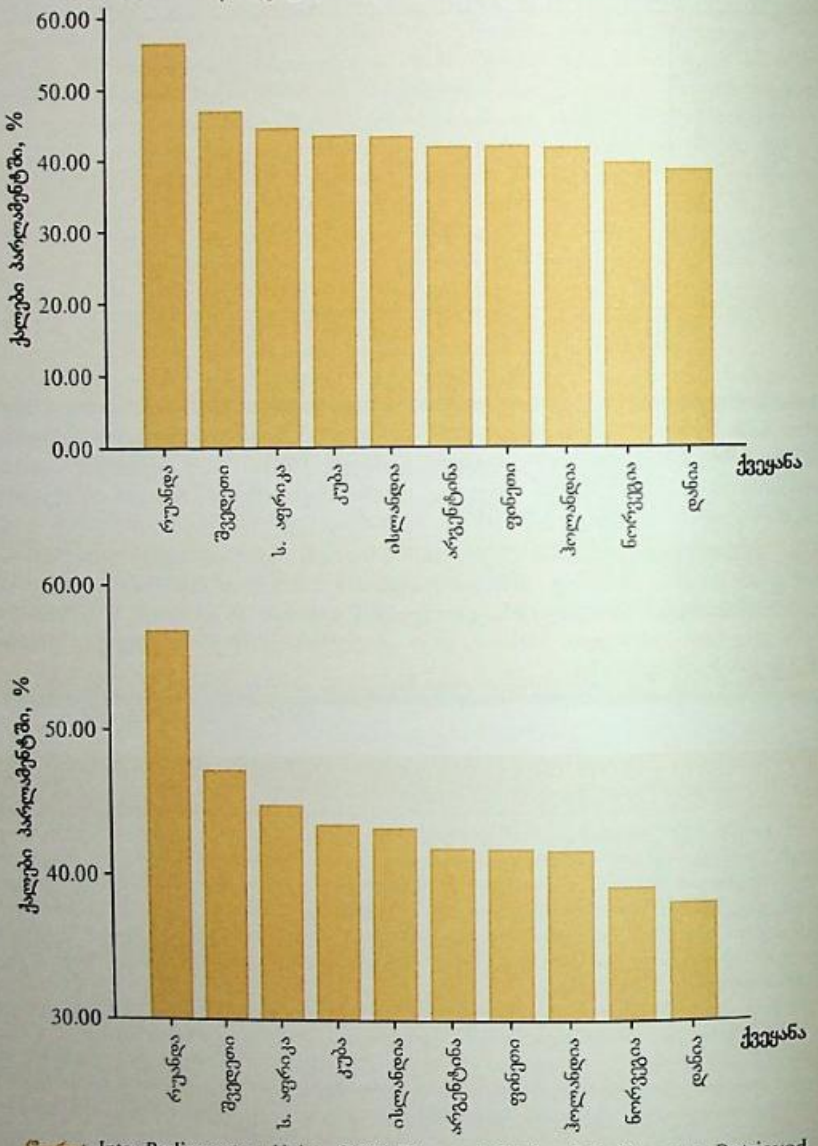
✓ *ცოდნის შემოწმება*

▣ დეტალების განხილვა 3.1
შეცდომის წყარო: გრაფიკების უზუსტობები

ამ თავში ჩვენ დავინახეთ, რომ გრაფიკები მოვლენის სწრაფად აღქმის საშუალებას გვაძლევს. თუმცა სწრაფ ინფორმაციასთან ერთად, მათ შეიძლება შეცდომაშიც შეგვიყვანონ. რადგან ჩვენთვის ზოგადი აღქმა უფრო მნიშვნელოვანია, ვინემ დეტალებში ჩაღრმავება, ადვილი შესაძლებელია, რომ არასწორად აგებული გრაფიკების ტყვეობაში მოვექცეთ. 1983 წელს გამოქვეყნებულ ნაშრომში „რაოდენობრივი ინფორმაციის ვიზუალური ჩვენება", ედვარდ ტაფტი არა მხოლოდ გრაფიკების დადებით ეფექტებზე საუბრობს, არამე გვიჩვენებს მათი გამოყენებისა და ინტერპრეტაციის ხარვეზებსაც.

ალბათ ყველაზე უფრო გავრცელებული დარღვევები მაშინ ჩნდება, როდესაც დისტანცია აბსცისასა და ორდინატას შორის შეცვლილია (ათვლის წერტილი შეიძლება 0 იყოს, — ან არა, როგორც ეს ნაჩვენებია ნახაზებზე 3.15ა, ან 3.15ბ). ორდინატთა ღერძები შეიძლება დავაგრძელოთ ან დავამოკლოთ დასახული მიზნის მისაღწევად. ნახაზებზე 3.15ა და 3.15ბ წარმოდგენილია ქვეყნების პარლამენტებში ქალთა წარმომადგენლობის საერთაშორისო მაჩვენებლები 2009 წელს. ნახეთ, როგორ მოხდა მონაცემების ყოველგვარი შეცვლის გარეშე სურათის უფრო მკვეთრი წარმოდგენა, როდესაც 0-ის ნაცვლად ბაზად აღებულია 30% (იხ. ნახ. 3.15ბ.).

ნახაზი 3.15 ქალთა წარმომადგენლობა ეროვნულ პარლამენტებში, 2009: (ა) ბაზად აღებულია 0 და (ბ) ბაზად გამოყენებულია 30



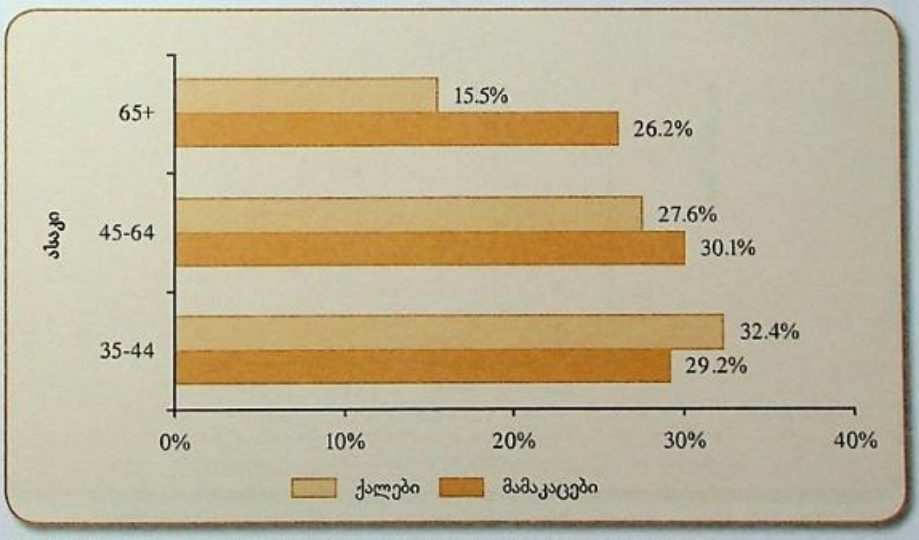
წყარო: Inter-Parliamentary Union. 2009. Women in National Parliaments. Retrieved December 22, 2009, www.ipu.org/wmn-e/classif.htm

სტატისტიკა პრაქტიკაში: მრავალფეროვნება ერთი თვალის შევლებით

ახლა, დამატებით გზებს გიჩვენებთ განსხვავების ვიზუალური ხაზგასმისათვის. კერძოდ, თუ როგორ გვეხმარება დიაგრამები: (ა) ამერიკული საზოგადოების სხვადასხვა ჯგუფების მსგავსებისა და განსხვავების წარმოჩენასა და (ბ) მასში მიმდინარე სწრაფი სტრუქტურული ცვლილებების ასახვაში. როდესაც თქვენ შეხედავთ მონაცემებს, ამერიკული საზოგადოების ჰეტეროგენულობის გამო, დაისმება კითხვა: „რასთან შედარებით?“ ეს კითხვა არა მარტო რაოდენობრივი აზროვნების,[8] არამედ, ზოგადად, აზროვნების საფუძველია.

სამივე სახის გრაფიკი — სვეტებიანი, ხაზოვანი და დროითი მწკრივების — ძალიან მოსახერხებელია ჯგუფებს შორის შედარებისათვის. დავიწყოთ სვეტებიანი დიაგრამით, რომელიც გამოსახულია ნახაზზე 3.16. მასზე შედარებულია კოლეჯის ხარისხის მქონე 35 წლისა და ზევით ასაკის მოსახლეობა გენდერულ ჭრილში. ის გვიჩვენებს, რომ ბაკალავრის ხარისხი უფრო მეტ მამაკაცს აქვს 45 და ზევით ასაკებში, თუმცა ამ მაჩვენებლით ქალები მამაკაცებს სჭარბობენ უფრო ახალგაზრდა ჯგუფებში (35-44). ამ ასაკობრივ ჯგუფში მცირე განსხვავება აღინიშნება — მამაკაცების ხვედრითი წილი, ვისაც ბაკალავრის ან უფრო მაღალი ხარისხი აქვს, 29.2%-ს შეადგენს, ხოლო ქალებისა — 32.4%-ს.

ნახაზი 3.16 კოლეჯდამთავრებულთა ხვედრითი წილები 35 წლისა და უფრო მაღალ ასაკობრივ ჯგუფებში გენდერულ ჭრილში, 2008



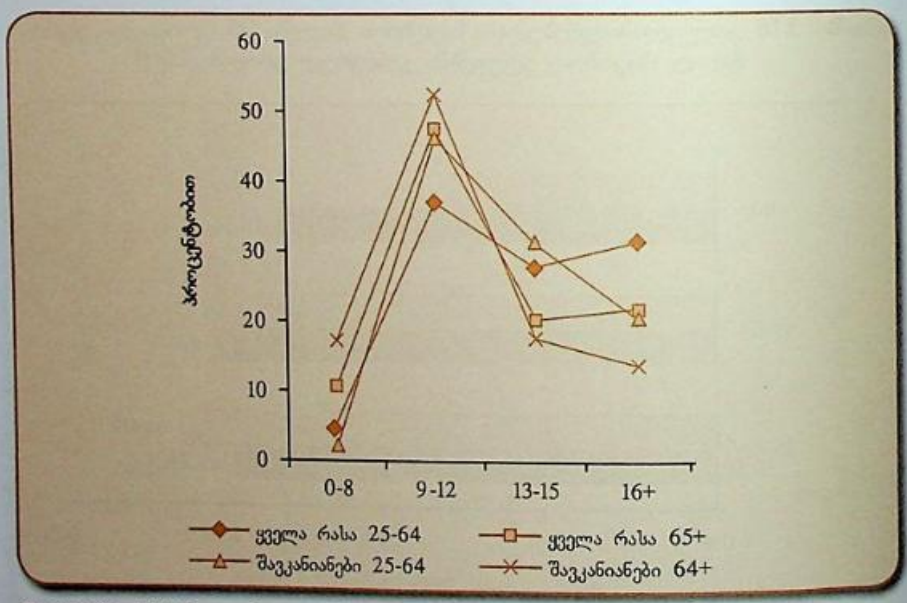
წყარო: U.S. Census Bureau, *American Community Survey*, 2008, Table B15001.

ხაზოვანი დიაგრამა ერთმანეთს ადარებს განსხვავებებს სქესში, რასობრივ/ეთნიკურ ნიშნებში, ან სხვა განსხვავებებს, როგორიცაა სოციალური ჯგუფი, ასაკი ან სექსუალური ორიენტაცია. მაგალითად, ნახაზზე 3.17 გამოსახულია შავკანიანების სწავლის ხანგრძლივობა 25 და ზევით ასაკობრივ ჯგუფებში, ამერიკელი მოსახლეობის ამავე ასაკობრივ ჯგუფებთან მიმართებით.

როგორც ვხედავთ, ამერიკელების ხვედრითი წილი, ვისაც 8-წლიანი განათლება აქვს, 65 წლის და ზევით ასაკის მოსახლეობაში 10.7%-იდან 4.2%-მდე შემცირდა 25-64 წლიანთა ჯგუფში. შავკანიანი ამერიკელებისათვის ეს კლება უფრო გამოკვეთილია, 17%-იდან 2.2%-მდე. ნახაზზე 3.17 აგრეთვე ნაჩვენებია როგორც შავკანიანთათვის, ისე ყველა ამერიკელისათვის სწავლის გაზრდილი ხანგრძლივობა 13-დან 15 წლამდე, ან 16-იდან და ზევით. მაგალითად, 65 და ზევით ასაკებში შავკანიანთა 17.3%-მა ისწავლა 13-დან 15 წლამდე, 31.3%-თან შედარებით 25-64 წლიანებში.

ნახაზზე 3.17 ასახულია მასობრივი განათლების განვითარება აშშ-ში უკანასკნელი 50 წლის განმავლობაში. ამერიკელების ხვედრითი წილი, ვისაც 4-წლიანი საშუალო განათლება აქვს, 40%-იდან (1940 წ.) 86.6%-მდე გაიზარდა 2008 წლისათვის. ანალოგიურად, თუ 1940 წლისათვის ამერიკელთა მხოლოდ 5%-ს ჰქონდა კოლეჯის 4 კლასის ან მეტი განათლება, 2008 წლისათვის მათი ხვედრითი წილი უკვე 29.4%-ია.

ნახაზი 3.17 აშშ-ში მიღებული განათლება რასისა და ასაკის მიხედვით, 2008



წყარო: The U.S. Census Bureau, *Educational Attainment in the United States: 2009*, Detailed Tables, Table 1.

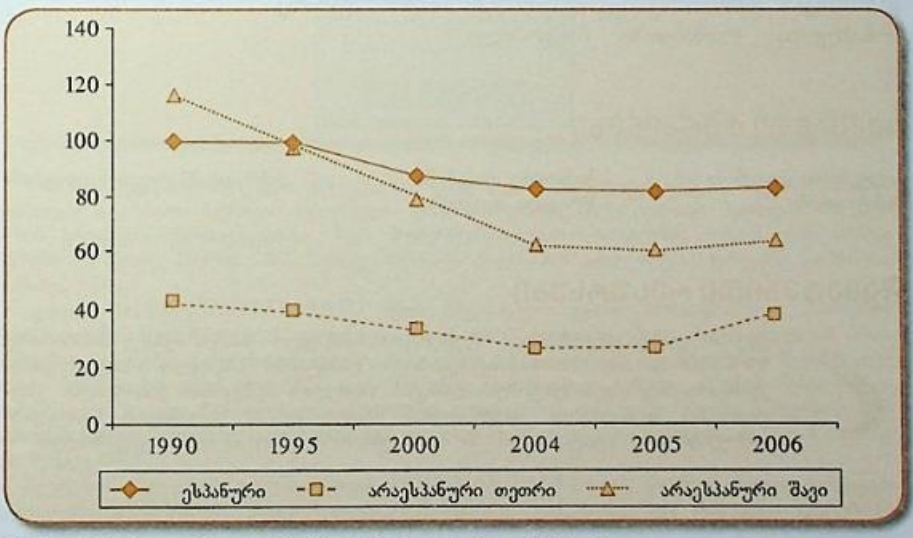
დიაგრამა 3.17 გვიჩვენებს, რომ ახალგაზრდა ამერიკელები (25-64) უფრო განათლებულები არიან, ვიდრე ხანდაზმულები. მიუხედავად ამისა, არსებობს განსხვავებები ამ მაჩვენებლებში, როდესაც ვადარებთ "შავკანიანებსა" და "ყველა რასას". აღნიშნული განიხილეთ ნახაზზე და იპოვეთ განსხვავებები. რას გვეუბნებიან ციფრები განათლების ხანგრძლივობის შესახებ?

✓ *ცოდნის შემოწმება*

და ბოლოს, ნახაზზე 3.18 მოცემულია ცვლილებები თინეიჯერების შობადობის კოეფიციენტებში (დაბადებულთა რაოდენობა ყოველ 1,000 ქალზე) დროითი მწკრივების დიაგრამის საშუალებით თეთრკანიან, შავკანიან და ესპანურ მოსახლეობაში. ის გვიჩვენებს, რომ 1990-დან 2006 წლამდე თინეიჯერების შობადობის კოეფიციენტი შემცირდა; შობადობის კოეფიციენტი ყველაზე მაღალია ესპანურ და არაესპანურ შავკანიან მოსახლეობაში, ყველაზე დაბალი კი — თეთრკანიან არაესპანურ ქალთა ჯგუფში.

საერთო ჯამში, ამ სამი ტიპის გრაფიკის განხილვამ დაგვანახა, თუ როგორ გვეხმარება გრაფიკული ასახვის მეთოდი სოციალური საკითხების კომპლექსურ წარმოჩენაში ისეთი მახასიათებლების მიხედვით, როგორიცაა, ასაკი, სქესი და ეთნიკური ნიშანი. მსგავსებებისა და განსხვავებების გრაფიკული გამოსახვით ისინი გვეხმარებიან სოციალური სამყაროს მრავალფეროვნებისა და კომპლექსურობის დანახვაში.

ნახაზი 3.18 აშშ-ის ესპანური წარმოშობის თინეიჯერების შობადობის კოეფიციენტები 1990, 1995, 2000, 2004, 2005, 2006



წყარო: U.S. Census Bureau, *The 2010 Statistical abstrct*, Table 84.
შენიშვნა: ყოველ 1000 ქალზე თითოეულ ჯგუფში.

ძირითადი განმარტებები

• წრიული დიაგრამა გვიჩვენებს განსხვავებებს ნომინალური ან ორდინალური ცვლადების სიხშირეებს ან ხვედრით წილებს შორის. მასში კატეგორიები სეგმენტების სახითაა წარმოდგენილი, რომელთა ჯამი 100%-ის ტოლია.

• სვეტებიანი დიაგრამა გვიჩვენებს განსხვავებებს ნომინალური და ორდინალური ცვლადების სიხშირეებს ან ხვედრით წილებს შორის. კატეგორიები წარმოდგენილია თანაბარი სიგანის მართკუთხედების სახით, რომელთა სიმაღლე სიხშირის, ან ხვედრითი წილის პროპორციულია.

• ჰისტოგრამა გამოიყენება ინტერვალური ცვლადების სიხშირეებისა და ხვედრითი წილების განსხვავებათა გამოსახვის მიზნით. კატეგორიები გამოსახულია ერთმანეთზე მიწყობილი სვეტების სახით, რომელთა სიგანე კატეგორიის მნიშვნელობას შეესაბამება, ხოლო სიმაღლე — სიხშირეს.

• ხაზოვანი დიაგრამა გვიჩვენებს განსხვავებებს ინტერვალურ-ფარდობითი ცვლადების სიხშირეებსა და ხვედრით წილებში. წერტილები, რომლებიც თითოეული კატეგორიის სიხშირეს შეესაბამება, თავსდება ამ კატეგორიის შუა წერტილის თავზე და შემდეგ სწორი ხაზით ერთდება.

• დროითი მწკრივების დიაგრამა ასახავს ცვლადის განსხვავებებს დროის სხვადასხვა მონაკვეთებში. ის გვიჩვენებს დროს (იზომება დროის ერთეულებში როგორიცაა წელი, თვე და ა.შ.) აბსცისთა ღერძზე და სიხშირეს (ხვედრით წილებს, დონეებს) ორდინატთა ღერძზე.

საკვანძო ტერმინები

სვეტებიანი დიაგრამა
ჰისტოგრამა
ხაზოვანი დიაგრამა
წრიული დიაგრამა
დროითი მწკრივების დიაგრამა

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

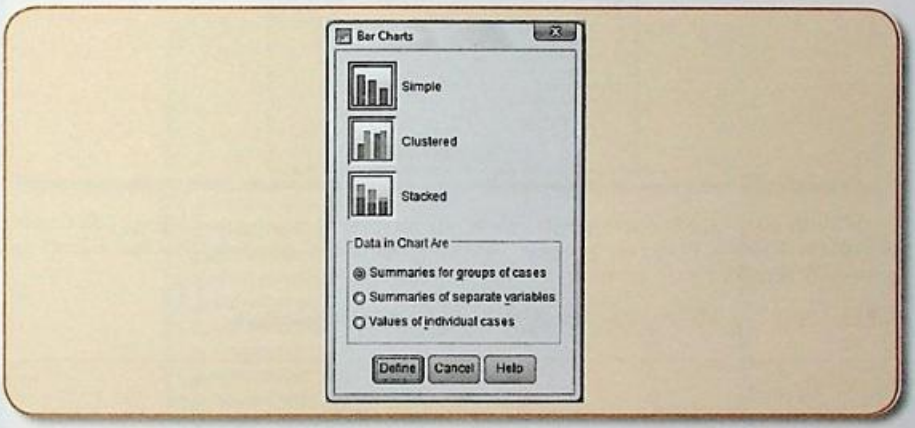
[HINTS 2007]

მაგალითი 1: სვეტებიანი დიაგრამის შექმნა

SPSS მნიშვნელოვნად ამარტივებს გრაფიკების შექმნას. პროგრამა გვთავაზობს ცალკეული არჩევანის საშუალებას მთავარი მენიუდან Graphs, რომელშიც წარმოდგენილია ათეულობით გრაფიკები. ამ მაგალითში ჩვენ გამოვიყენებთ HINTS2007-ს.

Graph მენიუს ქვეშოთ შეარჩიეთ Legacy Dialogs, შემდეგ Bar, რომელიც შემოგვთავაზებს სვეტებიანი დიაგრამის მრავალ ტიპს. სვეტებიან დიაგრამას გამოვიყენებთ ნომინალური ცვლადისათვის HDO6MARITALSTATU (რესპონდენტის სტატუსი ქორწინებაში). დაკლიკეთ Bar-ზე და იხილავთ საწყის დიალოგურ ფანჯარას, რომელიც ნაჩვენებია ნახაზზე 3.19.

ნახაზი 3.19 სვეტებიანი დიაგრამის დიალოგური ფანჯარა

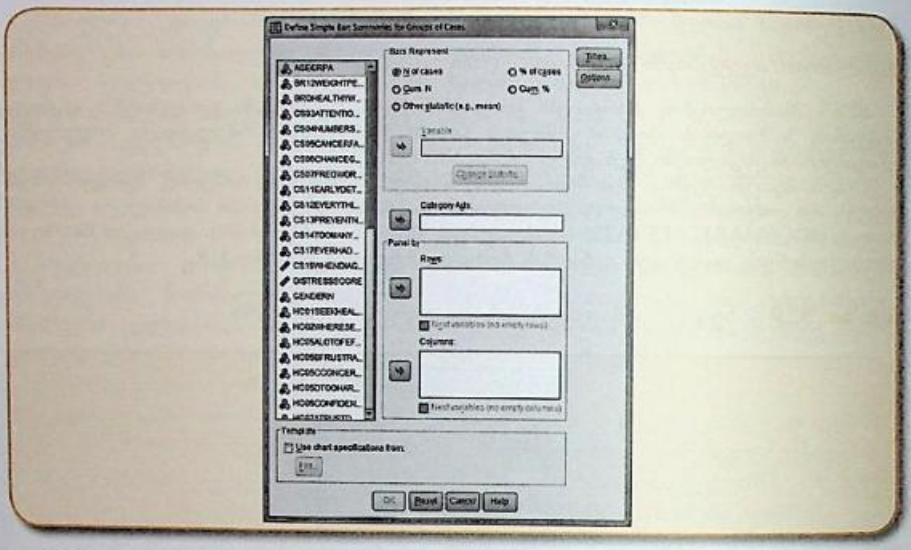


SPSS-ში ყველა გრაფიკული პროცედურა დიალოგური ფანჯრიდან იწყება, რაც საშუალებას გვაძლევს შევარჩიოთ ჩვენთვის სასურველი დიაგრამა. ერთზე მეტი ცვლადი შეიძლება მრავალი სახის გრაფიკით წარმოვადგინოთ. ჩვენ შევეჩერდებით უფრო მარტივზე, ამიტომ დააწკაპუნეთ Simple-ზე, შემდეგ Define. ამის შემდეგ, მარტივი დიაგრამის დიალოგური ფანჯარა გაიხსნება. (ნახაზი 3.20).

ცვლადი HDO6MSARITALSTATU უნდა მოვათავსოთ ყუთში სახელწოდებით „Category Axis". ყუთში „Bars Represent" დაკლიკეთ „% of cases" ღილაკზე. ეს რიცხვებს შეცვლის პროცენტებით, რაც შედარებებისთვის უფრო მისაღებია. დაკლიკეთ OK-ზე ბრძანების შესასრულებლად. (უნდა გახსოვდეთ, რომ SPSS ავტომატურად გამორიცხავს გამოტოვებულ მნიშვნელობებს. აღნიშნულის შეცვლა შესაძლებელია Option-ზე დაკლიკებით. დაკლიკეთ ყუთზე სახელწოდებით „Display groups defined by missing values", თუ გსურთ ამ არჩევანზე შეჩერება. შემდეგ Continue და შემდეგ OK).

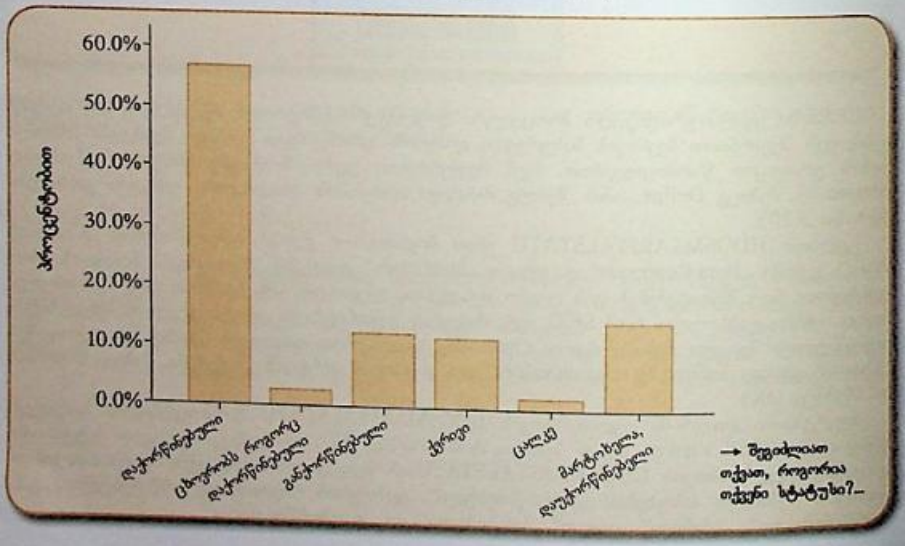
სვეტებიანი დიაგრამა ცვლადისათვის HDO6MARITALSTATU წარმოდგენილია საშედეგო ფანჯარაში SPSS Viewer. თქვენ ხედავთ, რომ ნახაზზე 3.21, რომელზეც ასახულია სვეტებიანი დიაგრამა ცვლადისთვის HDO6MARITALSTATU, ის შეიცავს მხოლოდ ექვს სვეტს, რადგან ამ კითხვის შესაძლო პასუხებია: „დაქორწინებული", „ცხოვრობ როგორც დაქორწინებული", „განქორწინებული", „ქვრივი", „ცალკე", „მარტოხელა" და „დაუქორწინებელი".

ნახაზი 3.20 მარტივი სვეტებიანი დიაგრამის დიალოგური ფანჯარა



SPSS-ის დიაგრამების რედაქტირება შესაძლებელია Edit-ზე დაკლიკებით, შემდეგ Edit Content, In Separate Window, რომელიც გრაფიკს მიმართავს საკუთარი ფანჯრისკენ (Chart Editor) და გამოიტანს რედაქტირების ხერხებსა და ფართო არჩევანს.

ნახაზი 3.21 ქორწინებაში სტატუსის სვეტებიანი დიაგრამა

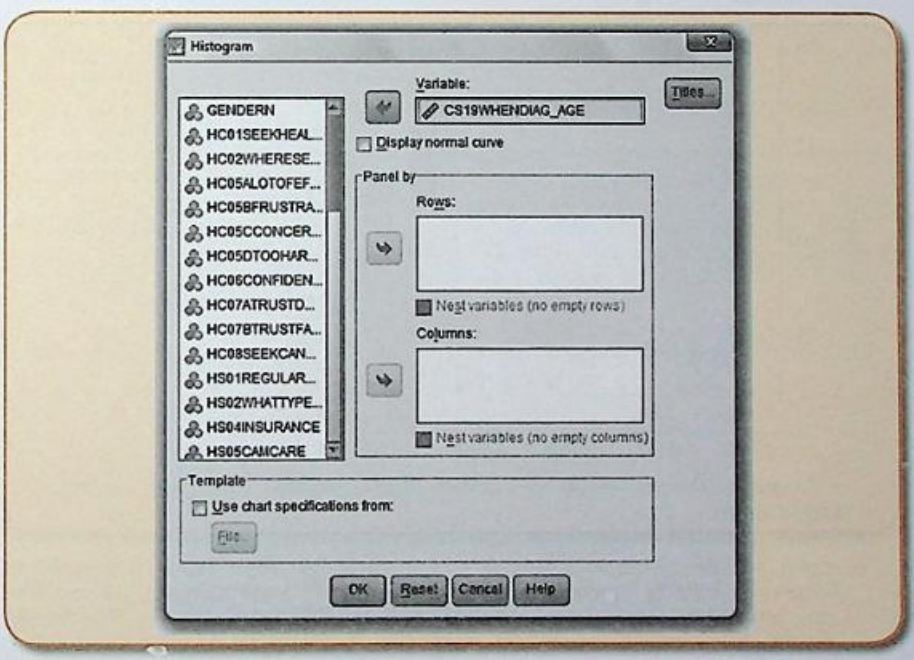


მაგალითი 2: ჰისტოგრამის შექმნა

ჰისტოგრამა გამოიყენება ინტერვალური ცვლადების ასახვისათვის. ჩვენ გამოვიყენებთ ცვლადს CS19WHENDIAG_AGE (კიბოს დადგენის ასაკი) HINTS2007-ის ფაილიდან. Graphs-Legacy Dialogs მენიუს ქვემოთ აირჩიეთ Histogram. დაკლიკეთ [illegible] არჩევანზე და დაინახავთ დიალოგურ ფანჯარას ნახაზზე 3.22.

ჰისტოგრამა ერთ ჯერზე ერთი ცვლადისთვის კეთდება (ამიტომ არ გაიხსნა დიალოგური ფანჯარა, როგორც სვეტებიანი დიაგრამის დროს). პირველ ცარიელ ჯუთში მარტივად გადაათრიეთ ცვლადი, რომელიც თქვენ გსურთ. არ იფიქროთ გამოტოვებულ მნიშვნელობებზე, ჰისტოგრამა მათ ავტომატურად გამორიცხავს (სვეტებიანი დიაგრამისგან განსხვავებით). დაიმახსოვრეთ, თითოეული ცვლადის გაზომვის დონის მისათითებლად SPSS ხატებს შეიცავს. ინტერვალურ-ფარდობითი ცვლადები მონიშნულია სახაზავიანი ხატით. ბრძანების შესასრულებლად დაკლიკეთ OK ღილაკზე (მარცხენა ქვედა კუთხეში). საშედეგო ჰისტოგრამა მოცემულია ნახაზზე 3.23.

ნახაზი 3.22 ჰისტოგრამის დიალოგური ფანჯარა



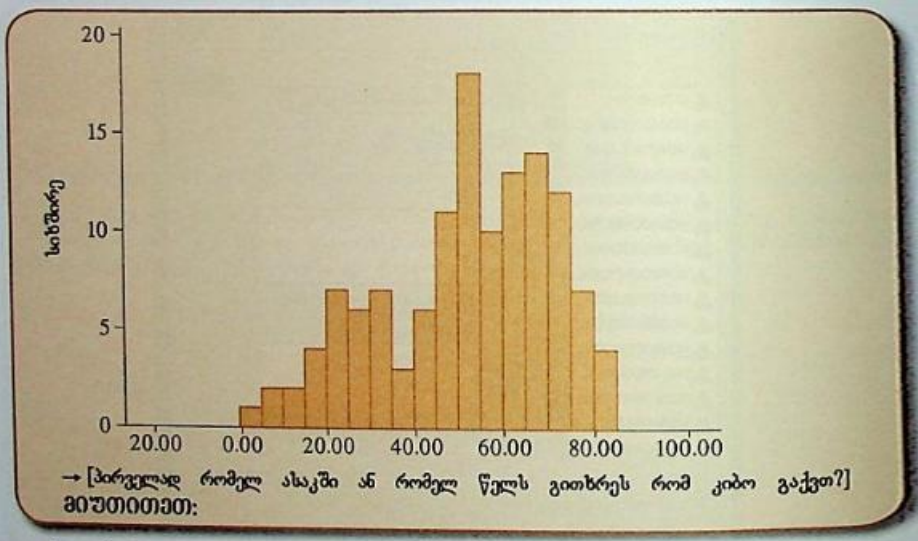
SPSS ავტომატურად იღებს გადაწყვეტილებას თითოეული ინტერვალის სიგანის შესარჩევად, ეს დამოკიდებულია ცვლადის დიაპაზონზე და ეკრანზე გამოსატანი სვეტების ოპტიმალურ რაოდენობაზე. თითოეული სვეტის ქვემოთ მოთავსებული რიცხვი შეესაბამება ინტერვალის შუა წერტილს. მაგალითად, 20 წლის ასაკის შესაბამის სვეტი მოიცავს ყველას 20-დან 25 წლამდე, ვისაც პირველად დაუდგინდა კიბო. ჰისტოგრამა აგრეთვე მოიცავს საშუალო სიდიდისა და საშუალო კვადრატული გადახრის გაანგარიშებასაც, რასაც შევეხებით თავებში 4 და 5.

ამოცანები SPSS-ისთვის

[HINTS 2007]

1. თქვენ გადაწყვიტეთ, რომ ნახოთ განსხვავებები მამაკაცებისა და ქალების ასაკებში მონაცემთა ბაზიდან HINTS2007. ამ მიზნით შეამოწმეთ ცვლადი AGEGRPA (რესპონდენტის ასაკობრივი ჯგუფი).
ა. ააგეთ სვეტებიანი დიაგრამა ცვლადისათვის AGEGRPA (Graph მენიუს ქვემოთ შეარჩიეთ Legacy Dialogs, შემდეგ Bar).
ბ. ააგეთ სვეტებიანი დიაგრამები ცალ-ცალკე ქალებისა და მამაკაცებისთვის (ჩასვით ცვლადი AGEGRPA Panel by/Columns Box-ში)
გ. მოკლედ აღწერეთ ასაკობრივი განაწილება და განსხვავებები ქალებისა და მამაკაცების ასაკობრივ ჯგუფებს შორის.

ნახაზი 3.23 ჰისტოგრამა ცვლადისთვის CS19WHENDIAG_AGE (კიბოს დადგენის ასაკი)



2. იღებენ თუ არა უმაღლესი განათლების მქონე ადამიანები ინფორმაციას ჯანდაცვასა და სამედიცინო თემებზე სავარაუდოდ ნებისმიერი წყაროდან? ააგეთ გამოცალკევებული წრიული დიაგრამები ცვლადისათვის HC01SEEKHEALTHIN (ეძებს თუ არა რესპონდენტი ინფორმაციას სამედიცინო თემებზე?). თქვენ უნდა შეარჩიოთ Pie, მენიუდან Graph, შემდეგ Legacy Dialogs. პირველ დიალოგურ ყუთში შეარჩიეთ „Summaries for groups of cases". შემდეგ დარწმუნდით, რომ შეარჩიეთ „% of cases" Slices Represent-ის ქვეშ. ყუთში Define slices by, შეიტანეთ ცვლადი HC01SEEKHEALTHIN და Panel by/Columns Box-ში ჩასვით HD07EDUCATION (ხარისხი). შეადარეთ წრიული დიაგრამები. რა განსხვავებებია ინფორმაციის მოძიებაში სხვადასხვა განათლების მქონე ადამიანებს შორის?
3. შეამოწმეთ, არის თუ არა განსხვავება მამაკაცებისა და ქალების პასუხებს შორის შემდეგი ცვლადების მიხედვით: მათ მიერ მიღებული სტრესის დონე (DISTRESSSCORE), დასაქმების სტატუსი (HD05OCCUPATIONS) და დაუსვეს თუ არა ოდესმე კიბოს დი-

აგნოზი (CS17EVERHADCANCE). შეარჩიეთ დიაგრამა თითოეული ცვლადის გაზომვის დონიდან გამომდინარე. ცალკე ააგეთ გრაფიკები მამაკაცებისა და ქალებისთვის. არის თუ არა თვალსაჩინო განსხვავებები მონაცემებში?

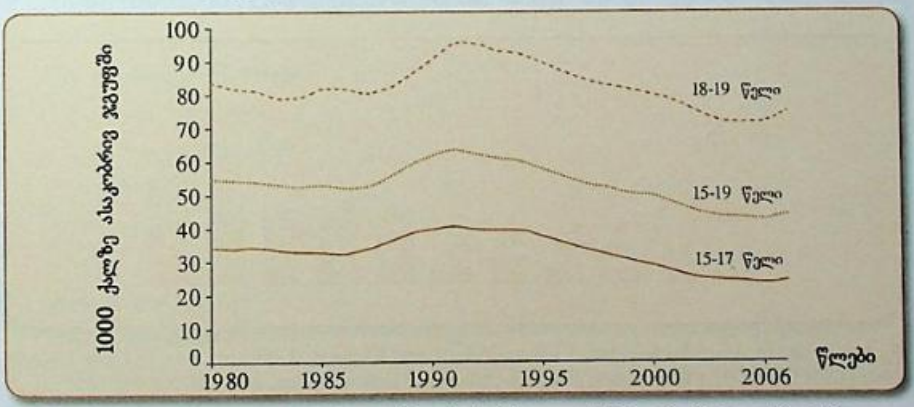
4. განსაზღვრეთ შემდეგი ცვლადების გრაფიკული ასახვის საუკეთესო გზები:
   HG16RENTOROWN – რესპონდენტი ბინას ქირაობს თუ მისი მესაკუთრეა?
   HD09RACE1 – რესპონდენტის რასა
   HD15INCOME – რესპონდენტის ოჯახის წლიური შემოსავალი
   HS08QUALITYCARE – როგორ აფასებს რესპონდენტი მიღებული ჯანდაცვის მომსახურების ხარისხს უკანასკნელი 12 თვის მანძილზე

შენიშვნა: თუ თქვენ აპირებთ დიაგრამის აგებამდე შეხედოთ ცვლადის სიხშირეს, გახსენებთ, რომ SPSS-ში ზოგიერთ ცვლადს არ აქვს მითითებული გაზომვის დონე, ამიტომ უპირველეს ყოვლისა, ჯერ დადასტურეთ მისი გაზომვის დონე და მერე შეუდექით შემდეგი ბრძანებების შესრულებას.

თავის სავარჯიშოები

1. ნახაზზე 3.24 მოცემულია თინეიჯერების შობადობის კოეფიციენტები დროითი მწკრივების დიაგრამით 1980 წლიდან 2006-მდე.
   ა. დაწერეთ 60 სიტყვიანი ანგარიში და აღწერეთ თინეიჯერების შობადობის ვარიაციები სამი ასაკობრივი კატეგორიისათვის.
   ბ. თქვენი აზრით, რა ახდენს გავლენას სხვადასხვა ასაკობრივი ჯგუფის შობადობის განსხვავებებზე? არის თუ არა წლები ან ათწლეულები, როდესაც ცვლილებები დრამატული იყო?
2. ჩვენ შევარჩიეთ ადამიანები სოციალური კვლევის საერთაშორისო პროგრამიდან (ISSP) 2000. ნედლ მასალაში წარმოდგენილია მათი სქესი (SEX), სოციალური ჯგუფი (CLASS) და შინამეურნეობის წევრთა რაოდენობა (HOMPOP). სოციალური ჯგუფის შემთხვევაში რესპონდენტები უთითებენ შემდეგ პასუხებს: L – დაბალი, W – მუშათა, M – საშუალო და U – მაღალი.

ნახაზი 3.24 თინეიჯერების შობადობის დონე ასაკების მიხედვით, 1980-2006



წყარო: Brady Hamilton, Jouce Martin, and Stephanie Ventura, "Births: Preliminary Data for 2006", *National Vital Haelth Statistics* 56, no. 7 (2007) (Hyattsville, MD: National Center for Health Statistics).

სქესი	Hompop	კლასი	სქესი	Hompop	კლასი
F	1	L	F	2	U
F	3	W	M	2	M
F	1	M	F	4	W
M	2	M	M	2	U
F	3	M	M	4	M
M	2	U	M	4	W
F	7	L	M	2	M
M	3	M	F	1	W
M	4	M	M	2	U
F	1	M	M	4	W
F	2	M	F	7	M
M	5	L	M	3	M
M	4	M	M	4	M
M	4	L	F	4	M
M	3	W	F	5	L

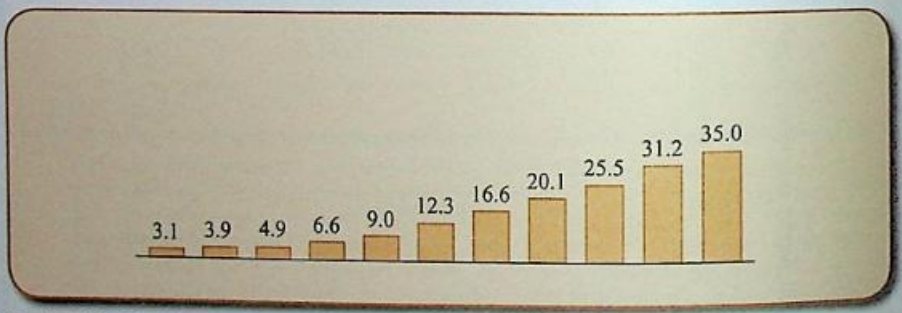
ა. ააგეთ წრიული დიაგრამა სქესის განაწილების ასახვის მიზნით (არ დაგავიწყდეთ სათაური, პროცენტები და შესაბამისი ჭდეები).

ბ. ააგეთ წრიული დიაგრამა სოციალური ჯგუფების პროცენტული განაწილების მიხედვით.

გ. ააგეთ ორი წრიული დიაგრამა და ერთმანეთს შეადარეთ სოციალური ჯგუფების პროცენტული განაწილების სქესობრივი მაჩვენებლები.

3. ვაგრძელებთ აშშ-ის ხანდაზმული მოსახლეობის ანალიზს. ამჯერად ვაკვირდებით ხანდაზმულების რიცხოვნობას 1900-დან 2000 წლამდე. სვეტებიანი დიაგრამა წარმოდგენილია ნახაზზე 3.25. წერილობით მოკლედ აღწერეთ მოტანილი მონაცემები.

ნახაზი 3.25 65 წლისა და ზევით ასაკის მოსახლეობა, 1900-2000 (მლნ)



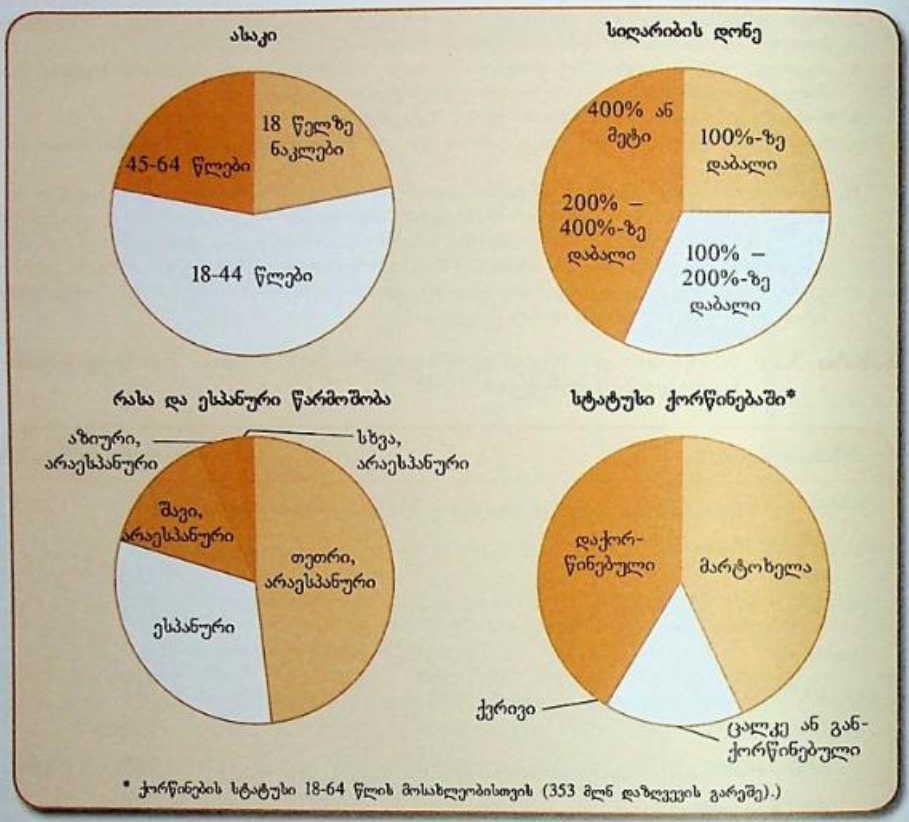
წყარო: Wan He, Manisha Segupta, Victoria Velkof, and Kimberly DeBarros. *65+ in United States: 2005* (CPS P23-209), (Washington, DC: Government Printing Office, 2005).

4. გამოიყენეთ მონაცემები მეორე სავარჯიშოდან და ააგეთ სვეტებიანი დიაგრამები სქესისა და სოციალური ჯგუფის პროცენტული განაწილების მიხედვით. (არ დაგავიწყდეთ სათაური, პროცენტები და შესაბამისი ჭდეები).

5. ვთქვათ, გსურთ შეადაროთ შინამეურნეობების წევრთა რაოდენობა ქალებისა და მამაკაცებისათვის (მეორე სავარჯიშოს მონაცემების მიხედვით).
   ა. ააგეთ ჯგუფური სვეტებიანი დიაგრამა (ნახაზი 3.4-ის მსგავსად) სქესის მიხედვით შინამეურნეობების წევრთა რაოდენობის პროცენტული განაწილებით.
   ბ. რომელ ჯგუფშია შინამეურნეობის უფრო მეტი წევრი?
   გ. რატომ გამოიყენეთ პროცენტული განაწილება და არა სიხშირეები სვეტებიანი დიაგრამის ასაგებად?
6. 2008 წლის საპრეზიდენტო არჩევნების დროს ამომრჩევლებმა მნიშვნელოვანი ადგილი მიანიჭეს ჯანდაცვასა და ჯანდაცვის დაზღვევას. პოლიტიკის ანალიტიკოსებმა დააფიქსირეს დაზღვეულთა რაოდენობის ზრდა აშშ-ში. ნახაზზე 3.26 მოტანილია ჯანდაცვის ეროვნული ცენტრის მონაცემები. ვის არ ჰქონდა ჯანდაცვის დაზღვევა 2005 წელს? როგორ განსხვავდება დაუზღვეველთა ხვედრითი წილები ეთნიკური ნიშნის, რასის, ასაკის, ქორწინების სტატუსისა და სიღარიბის დონის მიხედვით?

ნახაზი 3.26 65 წლისა და ზევით დაუზღვეველი მოსახლეობა შერჩეული მახასიათებლების მიხედვით, 2005

მახასიათებლები	პროცენტები
ასაკი	
18 წელზე ქვემოთ	16.2
18-44	61.6
45-64	22.2
სიღარიბე	
100%-ის ქვემოთ	25.1
100%-200%	31.6
200%-400%	28.7
400% და მეტი	14.6
რასა და ესპანური წარმოშობა	
ესპანური წარმოშობის	31.3
მხოლოდ თეთრკანიანები	48.0
შავკანიანები	13.9
აზიელები	4.3
სხვა	2.4
სტატუსი ქორწინებაში	
დაუქორწინებელი	40.2
დაქორწინებული	41.8
განქორწინებული/ცალკე მცხოვრები	16.3
ქვრივი	1.7



7. MTF 2008-ის ფარგლებში სტუდენტები გამოკითხეს მარიჰუანის მოხმარებასთან დაკავშირებით. ცხრილში მოტანილია შავკანიანი, თეთრკანიანი და ესპანური წარმოშობის სტუდენტების პასუხები. ააგეთ გრაფიკი ან სქემა, რომელიც უკეთესად ასახავს ამ ინფორმაციას (რასობრივი/ეთნიკური ჯგუფები ცალკე უნდა იყოს გამოყოფილი).

სტუდენტების დამოკიდებულება მარიჰუანის მოხმარებისადმი	შავკანიანი	თეთრკანიანი	ესპანური წარმოშობის
არ უარყოფ	73	434	69
უარყოფ	44	190	54
სრულიად უარყოფ	68	242	76
სულ (N)	185	866	199

8. დაუშვათ, თქვენ ატარებთ კვლევას უფროსი ასაკის მოსახლეობაზე, რომელიც შვილიშვილებთან ერთად ცხოვრობს. აღწერის მიხედვით, 2000 წელს აშშ-ში 2.4 მლნ. ადამიანი უწევდა მეთვალყურეობას თავის შვილიშვილს, რაც იმას ნიშნავს, რომ მათ ეკისრებოდათ შვილიშვილებზე პასუხისმგებლობა 18 წლამდე. აშშ-ის აღწერის ანგარიშიდან „Grandpar-

ents Living With Grandchildren: 2000" (C2KBR-31, October 2003) თქვენ ნახეთ, რომ ბაბუებთან და ბებიებთან 6 თვეზე ნაკლებ პერიოდში ცხოვრობდა შვილიშვილების 12%, 11% ცხოვრობდა 6-11 თვე, 23% – 1-დან 2-წლამდე, 15% 3-დან 4 წლამდე და 39% 5 და მეტი წელი.

ა. ააგეთ გრაფიკი ან სქემა, რომელიც უკეთესად ასახავს ამ ინფორმაციას.

ბ. დაასაბუთეთ თქვენი არჩევანი.

9. გამოიყენეთ მონაცემები განათლების დონეზე (თავი 2, სავარჯიშო 4) და უპასუხეთ კითხვებს:

ა. რა დონეზე იზომება „განათლების ხანგრძლივობა"? რატომ შეგიძლიათ წრიულ დიაგრამასთან ერთად ჰისტოგრამის გამოყენება ამ განაწილების გამოსახისათვის?

ბ. ააგეთ ჰისტოგრამა 4-წლიანი თანაბარი ინტერვალებით. ამ გრაფიკზე არ გამოიყენოთ ხვედრითი წილები.

10. GSS 2008-ის მონაცემები წარმოდგენილია რასობრივ ჭრილში:

ა. ააგეთ განათლების დონის შესახებ ორი ჰისტოგრამა, ერთი თეთრკანიანებისთვის (N=636) და მეორე შავკანიანებისთვის (N=136).

ბ. გამოიყენეთ გრაფიკები და ახსენით რასობრივი განსხვავებები განათლებაში.

განათლების წლები	თეთრკანიანები	შავკანიანები
0	2	0
1	0	0
2	1	1
3	2	0
4	0	0
5	3	1
6	5	0
7	2	0
8	13	1
9	8	6
10	17	8
11	27	6
12	180	36
13	56	18
14	101	33
15	27	5
16	103	13
17	32	1
18	32	2
19	12	1
20	13	4

11. გამოიყენეთ მეორე თავში სავარჯიშო 11-ის მონაცემები უმცირესობების უფლებებზე. რომელი გრაფიკული პრეზენტაცია გამოგადგებათ ამ შემთხვევაში? პასუხი დაასაბუთეთ.

12. დააკვირდით ნახაზს 3.27, სხვა ენებზე მოსაუბრე ამერიკელების შესახებ.

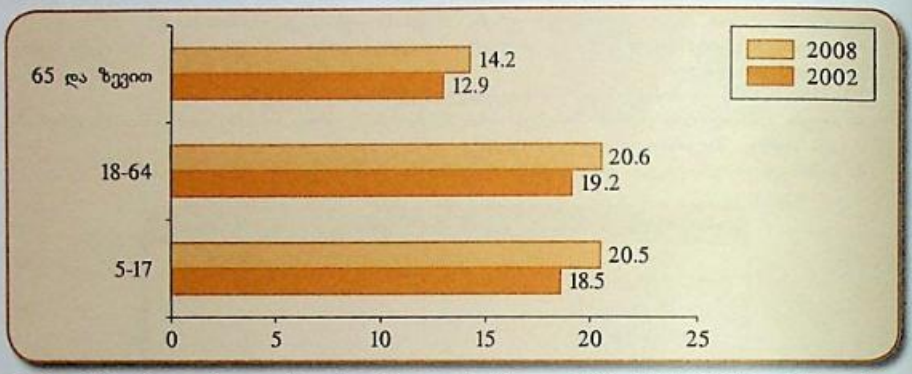
ა. საერთო ჯამში, რომელ ასაკობრივ ჯგუფში საუბრობს უფრო მეტი ადამიანი ინგლისურისგან განსხვავებულ ენაზე? უფრო ნაკლები?

ბ. აღწერეთ განსხვავებები 2002 და 2008 წლების პროცენტებში.

13. ნახეთ თავი 2, სავარჯიშო 7 განათლების მონაცემებზე. რომელი გრაფიკი გამოგადგებათ აღნიშნული ინფორმაციის ადვილად გასაგები ფორმატით გადმოსაცემად?

ა. რომელს აირჩევთ, სვეტებიან თუ წრიულ დიაგრამას? ახსენით.
ბ. ააგეთ წრიული ან სვეტებიანი დიაგრამა (თქვენი პასუხიდან გამომდინარე). არ დაგავიწყდეთ სათაური, პროცენტები და სახელდებები.

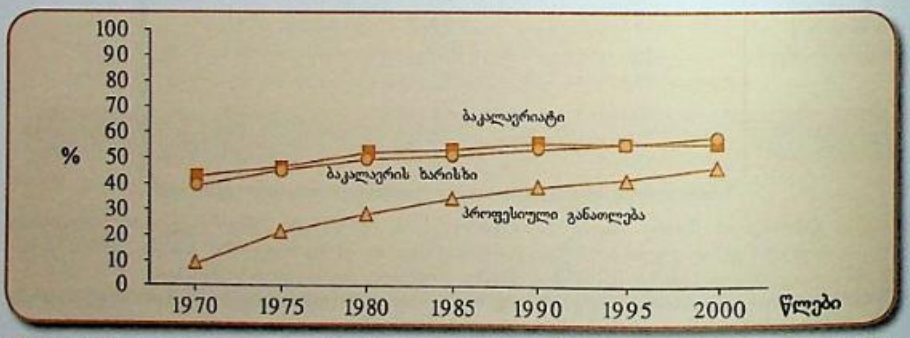
ნახაზი 3.27 ინგლისურისგან განსხვავებულ ენებზე მოსაუბრე 5 წლისა და ზევით მოსახლეობის ხვედრითი წილი ასაკის მიხედვით 2002 და 2008 წლებში



წყარო: U.S. Census Bureau, American Community Survey, 2002, Table PO35, and American Community Survey, 2008, Table C16007.

14. როგორც კეტრინ ფრიმენი (2004) აღნიშნავდა, ქალები უფრო წარმატებულნი იყვნენ საშუალო სკოლის შემდგომ განათლების მიღებაში, ვიდრე მამაკაცები. ისინი უფრო სწრაფად აგრძელებენ სწავლას კოლეჯებში და მაღალი რეიტინგები აქვთ. ნახაზზე 3.28 წარმოდგენილია ბაკალავრიატში სწავლის პროცესში და აგრეთვე ბაკალავრებისა და პროფესიული განათლების მქონე ქალების ხვედრითი წილები.

ნახაზი 3.28 ქალების ხვედრითი წილი უმაღლესი სასწავლებლების სხვადასხვა საფეხურებზე 1970-2000 წლებში



წყარო: Catherine Freeman, *Trends in Educational Equity for Girls and Women: 2004* (NCES 2005–016). U.S. Department of Education, National Center for Education Statistics (Washington, DC: Government Printing Office, 2004).

თავი 4

ცენტრალური ტენდენციის საზომები

თავის სწავლების მიზნები

- ცენტრალური ტენდენციის საზომების განსაზღვრებები, განსხვავებები მათ შორის, დადებითი და უარყოფითი მხარეები
- მოდის განსაზღვრა მოცემულ განაწილებაში
- მედიანისა და პერცენტილების პოვნა, ან გაანგარიშება
- საშუალო სიდიდის გაანგარიშება
- განაწილების ფორმის განსაზღვრა

მე-2 და მე-3 თავებში ჩვენ გავეცანით სიხშირულ განაწილებებს და მონაცემების გრაფიკული გამოსახვის მეთოდებს, რაც მათი აღწერისთვის გამოგვადგება. ამ მეთოდების საშუალებით შესაძლებელია მონაცემების დაჯგუფება და მათი გამოსახვა ადვილად აღსაქმელი ფორმით. თუმცა, ხშირად საჭირო ხდება მონაცემთა დიდი ბაზებისა და უამრავი ცვლადის დამუშავება. ამ დროს ცხრილებისა და გრაფიკების გამოყენება ნაკლებად ეფექტურია. მაგალითად, ჩვენ გვჭირდება სქესობრივ ჭრილში მოსახლეობის მონაცემების დამუშავება შემოსავლის, განათლების, ან პოლიტიკური შეხედულებების მიხედვით. რაც უფრო იზრდება ცვლადების რაოდენობა, მით უფრო რთულია მათი პრეზენტაცია.

განაწილების ასახვა შესაძლებელია აგრეთვე რაიმე ერთი რიცხვის შერჩევით, რომელიც მოკლედ შეაჯამებს და აღწერს მას. ასეთი რიცხვებით გამოისახება განაწილებისთვის დამახასიათებელი ტიპური ნიშნები. მაგალითად, კოლეჯის დონის განათლების მქონე ლათინო-ამერიკელების საშუალო შემოსავალი, ან ყველაზე სასურველი პოლიტიკური პარტია სოფლად მცხოვრები ღარიბებისათვის. განაწილებისათვის დამახასიათებელი საშუალო ან ტიპური ნიშნების ამსახველ რიცხვებს **ცენტრალური ტენდენციის საზომები** ეწოდება.

ცენტრალური ტენდენციის საზომები კატეგორიები ან მნიშვნელობები, რომლებიც ახასიათებს განაწილების საშუალო ან ტიპურ ნიშნებს.

ამ თავში ჩვენ შევეხებით ცენტრალური ტენდენციის სამ საზომს: მოდას, მედიანასა და საშუალოს. ალბათ, თქვენ გარკვეული წარმოდგენა გაქვთ ამ საზომების შესახებ. მედია-საშუალებებში ხშირად გამოიყენება ტერმინები საშუალო შემოსავალი და მედიანური შემოსავალი. თითოეული მათგანი განაწილების ყველაზე ტიპურ, დამახასიათებელ და რეპრეზენტატულ ნიშანს გამოსახავს. ამ თავში შევეხებით აგრეთვე მათ შორის არსებულ განსხვავებებს. ჩვენ ვნახავთ, რომ ცენტრალური ტენდენციის საზომებს შორის არჩევანი სამი ფაქტორითაა განპირობებული: (1) როგორ იზომება ცვლადი (მათი გაზომვის დონე), (2) განაწილების ფორმა და (3) კვლევის მიზანი.

მოდა

მოდა ეს ის კატეგორია, ან მნიშვნელობაა, რომელსაც განაწილებაში უდიდესი სიხშირე ან ხვედრითი წილი გააჩნია. საშუალო სიდიდეებს შორის მოდის გამოცნობა ყველაზე იოლია. უნდა ვიპოვოთ კატეგორია, რომელსაც განაწილებაში ყველაზე მაღალი სიხშირე აქვს.

მოდა კატეგორია ან მნიშვნელობა, რომელსაც განაწილებაში ყველაზე მაღალი სიხშირე (ან ხვედრითი წილი) აქვს.

მოდის საშუალებით შეიძლება დავახასიათოთ დღეისათვის აშშ-ში ყველაზე გავრცელებული უცხო ენა. რა თქმა უნდა, აშშ-ში საზოგადოებრივი კომუნიკაციის ენა ინგლისურია, მაგრამ აშშ-ის აღწერის ბიუროს მონაცემებიდან ჩანს, რომ ყოველი ხუთი ამერიკელიდან ერთი, სახლში ინგლისურისგან განსხვავებულ 155 ენაზე საუბრობს. აღნიშნულს ხელი შეუწყო 1980 წლიდან იმიგრანტთა მნიშვნელოვანმა ზრდამ.[1]

რომელი უცხო ენაა ყველაზე გავრცელებული დღეს აშშ-ში? ამის გასარკვევად დააკვირდით ცხრილს 4.1, რომელშიც ჩამოთვლილია 10 ყველაზე გავრცელებული ენა აშშ-ში და ამ ენებზე მოსაუბრე ადამიანების რიცხოვნობა. ცხრილიდან ჩანს, რომ ყველაზე მეტად გავრცელებულია ესპანური. ამ ენაზე 30 მილიონი ადამიანი საუბრობს. ამ მაგალითში მოდა არის „ესპანური ენა“ — კატეგორია, რომელსაც მოცემულ განაწილებაში ყველაზე დიდი სიხშირე გააჩნია.

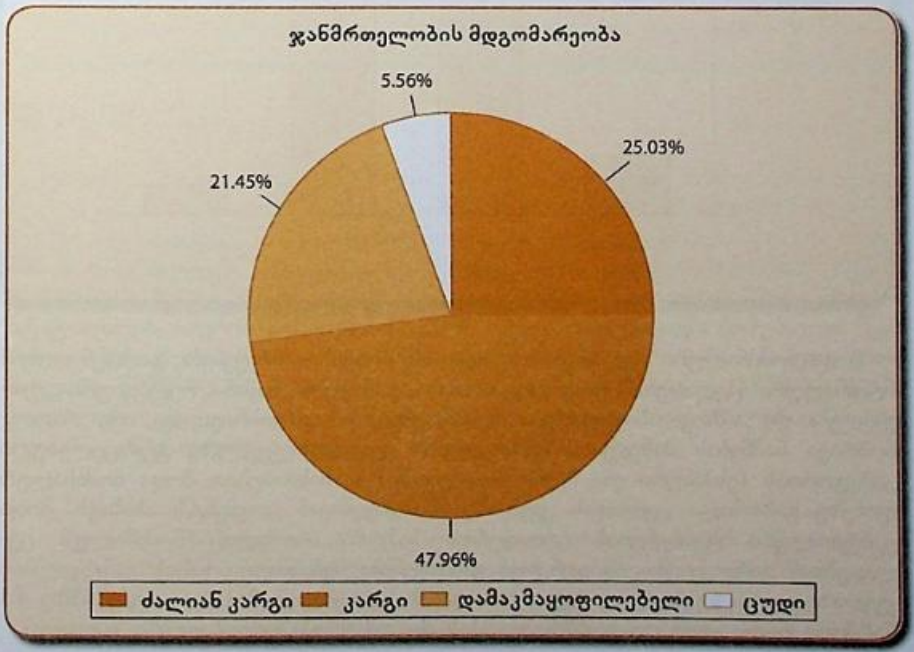
ცხრილი 4.1 ათი ყველაზე გავრცელებული ენა, რომელზეც საუბრობენ ამერიკელები, 2007

საუბარი ენა	მოსაუბრეთა რიცხვობა
ესპანური	34,547,000
ჩინური	2,465,000
ფრანგული	1,985,000
ტაგალოგი (ფილიპინური)	1,480,000
გერმანული	1,104,000
ვიეტნამური	1,207,000
კორეული	1,062,000.
იტალიური	799,000
რუსული	851,000
არაბული	767,000

წყარო: U.S. Census Bureau, The 2010 Statistical abstract, Table 53.

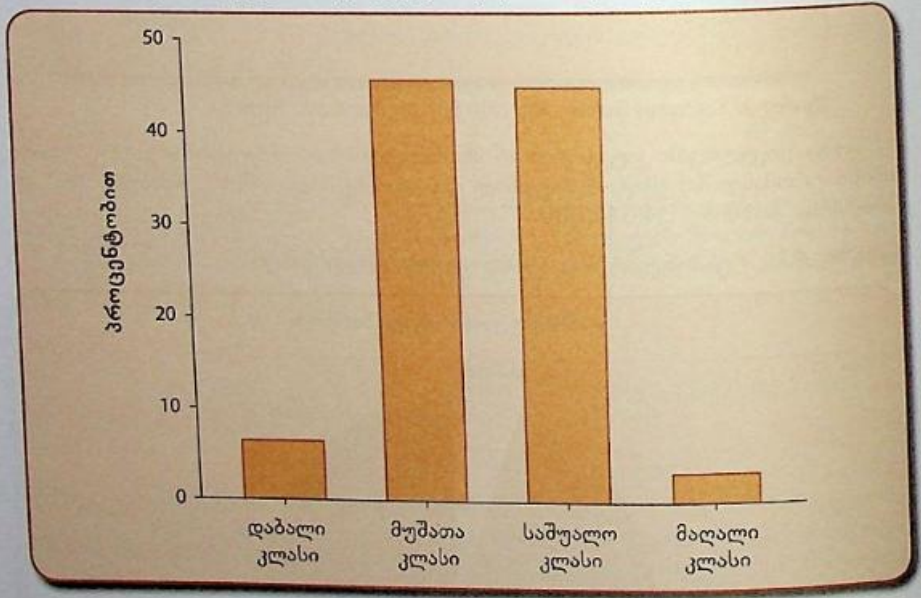
მოდა ყოველთვის კატეგორია ან მნიშვნელობაა და არა სიხშირე. არ აურიოთ ისინი ერთმანეთში! ამიტომ, მოყვანილ მაგალითში მოდა არის „ესპანური ენა“ და არა მისი სიხშირე 34,547,000.

ნახაზი 4.1 რესპონდენტების ჯანმრთელობის მდგომარეობა



აუცილებელი არ არის, რომ მოდის კატეგორიაში დაკვირვებათა უმრავლესობა შედიოდეს (ანუ, 50%-ზე მეტი), როგორც ეს ცხრილში 4.1-ია. მარტივად, ეს ის კატეგორიაა, რომელზეც დაკვირვებათა ყველაზე დიდი რაოდენობა (ან ხვედრითი წილი) მოდის. მაგალითად, ნახაზზე 4.1 წარმოდგენილია GSS-ის 2008 წლის კვლევის შედეგები რესპონდენტებისათვის დასმულ კითხვაზე: „როგორია თქვენი ჯანმრთელობის მდგომარეობა: ძალიან კარგი, კარგი, დამაკმაყოფილებელი თუ ცუდი?“. როგორც ხედავთ, ყველაზე მაღალი პროცენტი (47.96%) აქვთ რესპონდენტებს, რომელთაც პასუხში მიუთითეს „კარგი“. ამიტომ, პასუხი „კარგი“ ამ შემთხვევაში მოდას წარმოადგენს.

ნახაზი 4.2 რესპონდენტების სუბიექტური კლასობრივი იდენტიფიკაცია



მოდა ნომინალური ცვლადების დახასიათებისთვის გამოიყენება. გაიხსენეთ, რომ ნომინალური ცვლადების, როგორიცაა: რასა, ეთნიკური ნიშანი, რელიგიური კუთვნილება და ა.შ., კლასიფიცირება შესაძლებელია თვისობრივი და არა რაოდენობრივი ნიშნების მიხედვით. ნომინალური ცვლადის ყველაზე გავრცელებული კატეგორიის (ესპანური ენა, ჩვენს მაგალითში) დახასიათებით მოდა ნომინალურ დონეზე გაზომილი ცვლადის ყველაზე მნიშვნელოვან ელემენტს ასახავს. მოდა ცენტრალური ტენდენციის ერთადერთი საზომია, რომელიც ნომინალურ ცვლადებთან გამოიყენება. ის აგრეთვე გამოიყენება ნებისმიერი სახის განაწილების ყველაზე გავრცელებული კატეგორიის დასახასიათებლად. მაგალითად, ნახაზზე 4.1 წარმოდგენილი ცვლადი *ჯანმრთელობის მდგომარეობა*, ორდინალური ცვლადია.

ზოგ განაწილებაში შეიძლება შეგვხვდეს არა ერთი, არამედ ორი კატეგორია ან მნიშვნელობა, რომელსაც უდიდესი სიხშირე აქვს. ასეთ განაწილებას ორი მოდა გააჩნია და მას *ბიმოდალური* ეწოდება. მაგალითად, ნახაზზე 4.2 წარმოდგენილია რესპონდენტების პასუხები GSS-ის 2008 წლის გამოკვლევიდან. მათ დაუსვეს შემდეგი კითხვა: „თუ თქვენ ოთხიდან ერთი პასუხის მითითების საშუალება გექნებათ, რომელ კლასს მიაკუთვნებთ საკუთარ თავს: დაბალს, მუშათა კლასს, საშუალოს თუ მაღალს?“ რესპონდენტების ერთი და იგივე რაოდენობამ (დაახლოებით 44%-მა) მიუთითა „მუშათა კლასი“ და „საშუალო კლასი“. პასუხის ამ ორივე კატეგორიას უმაღლესი სიხშირე აქვს და ამის გამო, ორივე მოდაა. ასეთ განაწილებას ჩვენ ვახასიათებთ, როგორც ბიმოდალურს. როდესაც ორი მნიშვნელობის ან ორი კატეგორიის სიხშირეები ძალიან ახლოსაა (მაგრამ არა ერთმანეთის ტოლი), განაწილება „არსებითად“ მაინც ბიმოდალურია. ამ შემთხვევაში თქვენ შეგიძლიათ მიუთითოთ მოდის არა ზუსტი მნიშვნელობა, არამედ ორი უმაღლესი სიხშირის მქონე კატეგორია.

ქვემოთ მოცემულია 15 პიროვნების განაწილება პოლიტიკურ პარტიებში. იპოვეთ მოდა.

✓ *ცოდნი შემოწმება*

დემოკრატი	რესპუბლიკელი	დემოკრატი	რესპუბლიკელი	რესპუბლიკელი
დამოუკიდებელი	დემოკრატი	დემოკრატი	დემოკრატი	რესპუბლიკელი
დამოუკიდებელი	დემოკრატი	დამოუკიდებელი	რესპუბლიკელი	დემოკრატი

მედიანა

მედიანა ცენტრალური ტენდენციის საზომია იმ ცვლადებისთვის, რომლებიც სულ მცირე, ორდინალურ დონეზე გაიზომებიან. მედიანა განაწილების ზუსტად შუაში მდებარეობს; ეს ის მნიშვნელობაა, რომელიც განაწილებას ორ თანაბარ ნაწილად ყოფს. ამიტომ დაკვირვებათა ერთი ნაწილი მედიანის ქვემოთ მდებარეობს, ხოლო მეორე ნაწილი — მის ზემოთ. მაგალითად, აშშ-ის შრომის სტატისტიკის ბიუროს მონაცემებით, 2009 წელს მთელდღიურ მომუშავეთა კვირეული მედიანური ანაზღაურება $739 იყო.[2] ეს ნიშნავს, რომ მომუშავეთა ერთი ნახევარი $739-ზე ნაკლებს იღებდა, ხოლო მეორე ნახევარი — $739-ზე მეტს. სოციალურ კვლევებში გამოყენებული მრავალი ცვლადი ორდინალურია, ამიტომ მედიანა ცენტრალური ტენდენციის მნიშვნელოვანი საზომია.

მედიანა მნიშვნელობა, რომელიც განაწილებას ორ თანაბარ ნაწილად ყოფს ისე, რომ ერთი ნახევარი მის ქვემოთ მდებარეობს, ხოლო მეორე ნახევარი — მის ზემოთ.

მაგალითად, ჩვენ გვაინტერესებს, რა შეხედულება აქვთ ამერიკელებს დასახადებზე? როგორ შეიძლება აშშ-ში გადასახადების დონის რეიტინგები აღწერა? ეს კითხვა დაუსვეს რესპონდენტებს GSS 2008-ის კვლევაში. პასუხები შემდეგი სახის იყო: „ძალიან მაღალია“, „მაღალია“, „რეალურთან მიახლოებულია“, „დაბალია“ და „ძალიან დაბალია“. გადასახადების დონის რეიტინგები ორდინალური ცვლადებია. ამიტომ, საშუალო რეიტინგის გამოსავლენად ორდინალური ცვლადებისათვის შესაბამისი ცენტრალური ტენდენციის საზომი უნდა გამოვიყენოთ, როგორიცაა მედიანა. მედიანის გამოყენება ისეთი ცვლადებისთვისაა მისაღები, რომელთა კატეგორიები ან მნიშვნელობები სიდიდის მიხედვით არის დალაგებული, უმცირესიდან — უდიდესისკენ. სწორედ ამიტომ გამოიყენება მედიანა ორდინალური და ინტერვალურ-ფარდობითი და არა ნომინალური ცვლადებისათვის, რომელთა მნიშვნელობების რანჟირება შეუძლებელია.

მედიანის პოვნის ხერხი თანმიმდევრობით დალაგებული მონაცემებისათვის

მედიანის მოძებნა ძალიან ადვილია. ძირითადად, ეს დალაგებული მწკრივის დათვალიერებით ხდება. მედიანის განსხვავებული ადგილმდებარეობა დაკვირვებათა რიცხვზეა დამოკიდებული. კერძოდ, კენტია ის, თუ ლუწი. პირველად განვიხილოთ კენტი რაოდენობის ორი შემთხვევა.

დაკვირვებათა კენტი რიცხვი

ვთქვათ, ხუთ რესპონდენტს დავუსვით შემდეგი კითხვა: „როგორ შეაფასებდით დღეს ამ ქვეყნის ეკონომიკურ მდგომარეობას?“ ქვემოთ მოყვანილია პირობითი რესპონდენტების პასუხები:

ცუდად
კარგად
დამაკმაყოფილებლად
ცუდად
ძალიან კარგად
სულ (N) = 5

მედიანის ადგილმდებარეობის გასარკვევად პასუხები დაალაგეთ უმცირესიდან უდიდესისკენ (ან პირიქით):

ცუდად
ცუდად

დამაკმაყოფილებლად
კარგად
ძალიან კარგად
სულ (N) = 5

მედიანა ასოცირდება შუაში მდებარე დაკვირვებასთან. როდესაც დაკვირვებათა რიცხვი (*N*) კენტია, მაშინ შუა წევრის საპოვნელად, მას უნდა დავუმატოთ 1 და შემდეგ გავყოთ 2-ზე: *(N+1)/2*. რადგან N არის 5-ის ტოლი, გვექნება: *(5+1)/2=3*. როგორც ვნახეთ, შუაში მდებარეობს მესამე დაკვირვება და ამიტომ მედიანა არის კატეგორია „დამაკმაყოფილებლად". როგორც ვიცით, მედიანა განაწილებას ორ ტოლ ნაწილად ყოფს, ამიტომ მწკრივში ორი რესპონდენტი უფრო მეტადაა კმაყოფილი და დანარჩენი ორი კი — ნაკლებად.

ახლა ვნახოთ მეორე მაგალითი. ქვემოთ მოტანილია აშშ-ის 9 უდიდეს შტატში 2007 წელს სიძულვილის ნიადაგზე ჩადენილ დანაშაულთა სტატისტიკა.[3]

რიცხვი
1,400
166
323
748
493
312
83
242
75
სულ (*N*) = 9

მედიანის ადგილმდებარეობის გასარკვევად პასუხები დაალაგეთ უმცირესიდან უდიდესისკენ: შუაში მდებარე დაკვირვებაა *(9+1)/2=5*, მეხუთე შტატი, ოჰაიო. მედიანა 312-ის ტოლია, ეს რიცხვი კი ოჰაიოს შეესაბამება. ის ორ ტოლ ნაწილად ყოფს განაწილებას, ისე, რომ ოთხ შტატში ნაკლები დანაშაულებია, ხოლო დანარჩენ ოთხში — უფრო მეტი (აღნიშნული ნაჩვენებია ნახაზზე 4.3ა).

რიცხვი
1,400
166
323
748

493
312
83
242
75

სულ (N) = 9

აქ შტატები ჩამოთვლილია შემდეგი თანმიმდევრობით: ჩრდილოეთ კაროლინა, პენსილვანია, ფლორიდა, ტეხასი, ოჰაიო, ვირჯინია, ნიუ-იორკი, ნიუ-ჯერსი და კალიფორნია.

დაკვირვებათა ლუწი რიცხვი

შტატების რაოდენობიდან ამოვიღოთ ბოლო შტატი და მწკრივი ლუწად გადავაქციოთ (ნახაზი 4.3ბ). მნიშვნელობები ისევ ზრდადი თანმიმდევრობით არის დალაგებული.

თავიდან ვიპოვოთ მედიანის ადგილმდებარეობა ზრდად მწკრივში:

ნახაზი 4.3 სიძულვილის ნიადაგზე ჩადენილი დანაშაულების მედიანური რიცხვის პოვნა: ა) ცხრა შტატისთვის და ბ) ათი შტატისთვის

ა) დაკვირვებათა კენტი რიცხვი

1. დაალაგეთ დაკვირვებები უმცირესიდან უდიდესისკენ



75 83 166 242 312 323 493 748 1,400

2. ამ შემთხვევაში ჩვენ გვჭირდება მე-5 დაკვირვება: (9+1) + 2 = 5



75 83 166 242 312 323 493 748 1,400

4 დაკვირვება მედიანის ქვემოთ — 4 დაკვირვება მედიანის ზემოთ

ბ) დაკვირვებათა ლუწი რიცხვი

1. დაალაგეთ დაკვირვებები უმცირესიდან უდიდესისკენ



75 83 166 242 312 323 493 748

2. ამ შემთხვევაში ჩვენ გვჭირდება ორი შუათანა დაკვირვება: (8+1) + 2 = 4.5



75 83 166 | 242 | 312 | 323 493 748

3 დაკვირვება მედიანის ქვემოთ — 3 დაკვირვება მედიანის ზემოთ

რიცხვი
75
83
166
242
312
323
493
748
სულ (*N*) = 8

როდესაც შტატების რაოდენობა (*N*) ლუწია, შუაში აღარ გვაქვს მხოლოდ ერთი დაკვირვება. მედიანა ახლა ორ შუა წევრს შორის მდებარეობს. ვიპოვოთ ორი შუა დაკვირვება შემდეგი სახით: *(N + 1)/2* ან *(8+1)/2=4,5*. ეს ნიშნავს, რომ მედიანა მდებარეობს ორ შტატს, ტეხასსა და ოჰაიოს შორის. ამ შტატებს შეესაბამება დანაშაულების რაოდენობები 242 და 312. ინტერვალურ-ფარდობითი ცვლადის მედიანის საპოვნელად, მარტივად გაიანგარიშეთ ამ ორი რიცხვის საშუალო:

$$\text{მედიანა} = \frac{242 + 312}{2} = 277$$

ამრიგად, მედიანა 277-ის ტოლია.

დაიმახსოვრეთ, როდესაც მონაცემები ორდინალურია, შუა ორი წევრის საშუალოს გაანგარიშება არ ხდება. ამ შემთხვევაში მედიანა მოთავსებულია შუა ორ წევრს შორის.

✓ *ცოდნის შემოწმება*

იპოვეთ ინტერვალურ-ფარდობითი ცვლადის განაწილების მედიანა: 22, 15, 18, 33, 17, 5, 11, 28, 40, 19, 8, 20.

მედიანის პოვნის ხერხი სიხშირულ განაწილებებში

ხშირად ჩვენი მონაცემები სიხშირული განაწილებების სახით არის წარმოდგენილი. ავიღოთ, მაგალითად, სიხშირული განაწილება, რომელიც მოცემულია ცხრილში 4.2. მასში ასახულია GSS-ის რესპონდენტების პოლიტიკური შეხედულებები 2008 წელს.

ცხრილი 4.2 GSS-ის რესპონდენტების პოლიტიკური შეხედულებები, 2008

პოლიტიკური შეხედულებები	სიხშირე (f)	*Cf*	ხვედრითი წილი, %	*C %*
უკიდურესად ლიბერალური	52	52	3.6	3.6
ლიბერალური	175	227	12.2	15.8
სუსტად ლიბერალური	172	399	12.0	27.8
ზომიერი	545	944	38.0	65.8
სუსტად კონსერვატული	203	1,147	14.1	79.9
კონსერვატული	241	1,388	16.8	96.7
უკიდურესად კონსერვატული	48	1,436	3.3	100.0
სულ (N)	1,436			

მედიანის გასაანგარიშებლად საჭიროა ვიპოვოთ განაწილების შუაში მოთავსებული დაკვირვების კატეგორია. ჯერ გავიგოთ, რას უდრის N, რესპონდენტთა რიცხვნობა. ამ შემთხვევაში, N = 1,436-ს. შემდეგ გამოვიყენოთ ფორმულა *(N+1)/2*, ან *(1,436+1)/2=718,5*. მედიანა მოთავსებულია 718-ე და 719-ე დაკვირვებებს შორის. 718-ის და 719-ის შესაბამისი კუმულაციური სიხშირე (*Cf*) ხვდება კატეგორიაში „ზომიერი". ამიტომ მედიანა არის „ზომიერი". შეიძლება უცნაურად მოგეჩვენოთ, მაგრამ მედიანა ყოველთვის პასუხის კატეგორიის მნიშვნელობაა და არა სიხშირე.

სიხშირულ განაწილებაში მედიანის პოვნის მეორე გზაა კუმულაციური სიხშირეების გამოყენება, რაც ცხრილი 4.2-ის ბოლო სვეტშია მოთავსებული. ამ მაგალითში სიხშირეების დაგროვება (კუმულაცია) „უკიდურესად ლიბერალურიდან" „უკიდურესად კონსერვატულისკენ" ხდება. შეგვიძლია პირიქითაც მოვახდინოთ კუმულაცია „უკიდურესად კონსერვატულიდან" „უკიდურესად ლიბერალურამდე". მედიანის ადგილის საპოვნელად უნდა მოვძებნოთ პასუხის ის კატეგორია, რომლის პროცენტული კუმულაციური სიხშირე 50%-ის ტოლია. მედიანა სწორედ ამ დაკვირვების შესაბამისი კატეგორიის სიდიდე იქნება.[4] თუ შევხედავთ ცხრილს, ვნახავთ, რომ 50%-ის ტოლი კუმულაციური სიხშირე ხვდება კატეგორიაში „ზომიერი". ამიტომ ამ განაწილების მედიანა არის „ზომიერი". თუ არ ხართ

დარწმუნებული, რომ სწორად გაიგეთ, კიდევ ერთხელ შეხედეთ კუმულაციური პროცენტების სვეტს (*C* %). ნახავთ, რომ დაკვირვებათა 27.8% თავმოყრილია კატეგორია „ზომიერი“-ს ქვემოთ, ხოლო 65.8%-ში შედის როგორც „ზომიერი“, ისე მის წინ მდებარე დაკვირვებები. აქედან ჩვენ უკვე ვხვდებით, რომ 50% მოთავსებულია სადღაც კატეგორია „ზომიერი“-ს ფარგლებში.

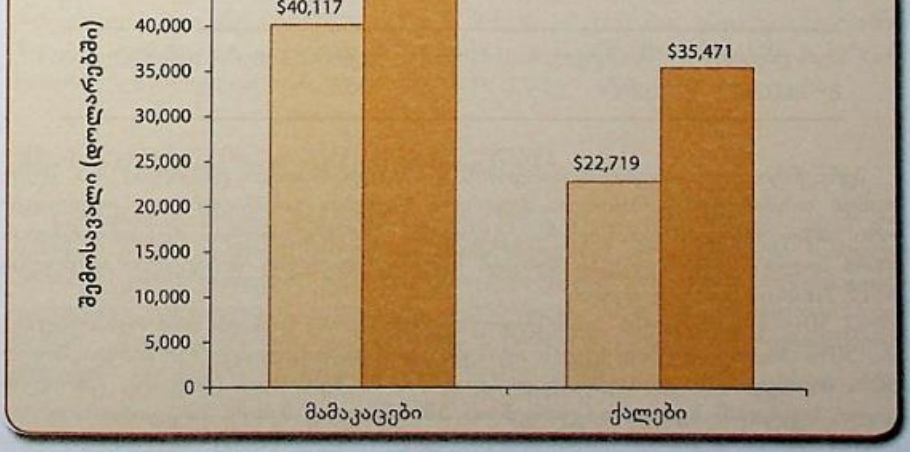
იხილეთ მეორე თავი კუმულაციური სიხშირეების გამორების მიზნით.

✓ *ცოდნა შემოწმება*

სტატისტიკა პრაქტიკაში: შემოსავლების გენდერული უთანაბრობა

მედიანის გამოყენება შესაძლებელია ჯგუფების შედარებისთვისაც. უკანასკნელი წლების მანძილზე მნიშვნელოვანი ძვრები მოხდა აშშ-ში ქალებისა და მამაკაცების შემოსავლების დონეებში. შემოსავლის დონე ძირფესვიანად ცვლის ჩვენ სოციალურ და ეკონომიკურ მდგომარეობას; მაღალი შემოსავალი განათლებისა და სამუშაო გამოცდილების მაღალ დონესთან ასოცირდება როგორც მამაკაცების, ისე ქალებისთვის.

ნახაზი 4.4 მამაკაცებისა და ქალების მედიანური შემოსავლები 1973 და 2008 წლებში



წყარო: 1973 წლის მონაცემები აღებულია U.S. Census Bureau, Current Population Reports P60-226, *Poverty and Health Insurance Coverage in United States: 2003*-დან. 2008 წლის მონაცემები აღებულია U.S. Census Bureau, American Community Survey, 2008-იდან.

ნახაზზე 4.4 შედარებულია მამაკაცებისა და ქალების მედიანური შემოსავლები 1973 და 2008 წლებში. მედიანის საშუალებით შესაძლებელია მოსახლეობის სხვადასხვა ქვეჯგუფში არსებული განსხვავებების ან დროის გარკვეულ პერიოდებში მომხდარი ცვლილებების გამოსახვა. ამ მაგალითიდან ნათლად ჩანს, რომ 1973-დან 2008 წლამდე ქალთა შემოსავლები მნიშვნელოვნად გაიზარდა, თუმცა ის ჯერაც საშუალოდ $10,000-ით ნაკლებია კაცებისაზე.

ცოდნის შემოწმება

ნახაზზე 4.4 დაკვირდით განსხვავებებს მამაკაცებისა და ქალების შემოსავლებში უკანასკნელი ათწლეულის მანძილზე. რა დასკვნა შეიძლება გამოვიტანოთ აქედან?

პერცენტილების ადგილი სიხშირულ განაწილებაში

მედიანა ცენტრალური ტენდენციის საზომების კერძო შემთხვევაა. უფრო ზოგად საზომს **პერცენტილი** ეწოდება. პერცენტილი არის მნიშვნელობა, რომლის გასწვრივ ან ქვემოთ ხვდება განაწილების გარკვეული პროცენტი. *n*-ური პერცენტილი არის მნიშვნელობა, რომლის ქვემოთ მოხვდება განაწილების *n* %. მაგალითად, 75-ე პერცენტილი არის მნიშვნელობა, რომელიც განაწილებას ისე ყოფს, რომ დაკვირვებათა 75% მის ქვემოთ ხვდება. მედიანა 50-ე პერცენტილია. ის განაწილებას ისე ყოფს, რომ დაკვირვებათა 50% მის ქვემოთ ხვდება. მედიანის მსგავსად, პერცენტილებიც გამოიყენება ორდინალურ და ინტერვალურ-ფარდობით ცვლადებთან. პერცენტილების განსაზღვრა უფრო იოლია, თუ მონაცემები სიხშირული განაწილების სახით არის წარმოდგენილი.

პერცენტილი მნიშვნელობა, რომლის ქვემოთ ხვდება განაწილების გარკვეული პროცენტი.

პერცენტილების ადგილმდებარეობის განსაზღვრისათვის, ცხრილში 4.3 შევიტანეთ სიხშირული განაწილება, ხვედრითი წილების განაწილება და კუმულაციური პროცენტული განაწილება. ცხრილში წარმოდგენილია რესპონდენტების აზრი პოლიციელების ქცევის შესახებ სამუშაო ადგილზე. მონაცემები აღებულია MTF 2008-ის გამოკვლევიდან.

აქ 50-ე პერცენტილია „დამაკმაყოფილებელი", რაც ნიშნავს, რომ რესპონდენტთა 50% პოლიციელების ქცევას აფასებს „დამაკმაყოფილებელზე" მაღლა, ხოლო 50% მიიჩნევს მას „დამაკმაყოფილებლად", ან მასზე დაბლა (როგორც ცხრილის ბოლო სვეტიდან ხედავთ, დაკვირვებათა 50% სადღაც მესამე კატეგორიის ფარგლებში ხვდება, რომელიც „დამაკმაყოფილებელს" შეესაბამება). ასევე, მე-20 პერცენტილი იქნება „ცუდი", რადგან რესპონდენტთა 20% პოლიციელების ქცევას აფასებს „ცუდად" ან „ცუდზე" უფრო უარესად.

პერცენტილები ფართოდ გამოიყენება ფარდობითი მაჩვენებლების შესაფასებლად სტანდარტიზებულ ტესტებში, როგორიცაა SAT და ACT. დაუშვათ, თქვენ ACT-ში 29 ქულა დააგროვეთ. კოლეჯში მისაღებად თქვენ ქულას გადაიყვანენ პერცენტილურ რანგებში და შემდეგ მოხდება მათი შედარება. თუ თქვენთან შედარებით დანარჩენი სტუდენტების 90%-მა დაბალი ქულა (დანარჩენ 10%-ს უფრო მაღალი ქულა აქვს) მიიღო, მაშინ თქვენი პერცენტილური რანგი 90-ის ტოლი იქნება. თუ 15%-მა თქვენზე მაღალი ქულა მიიღო, მაშინ თქვენი პერცენტილური რანგი მხოლოდ 85 იქნება.

ცხრილი 4.3 პოლიციელების ქცევის სიხშირული განაწილება: MTF 2008-ის რესპონდენტები

პოლიციელების ქცევა	სიხშირე (f)	ხვედრითი წილები, %	C%
ძალიან ცუდი	125	9.9	9.9
ცუდი	201	15.9	25.8
დამაკმაყოფილებელი	452	35.7	61.5
კარგი	380	30.0	91.5
ძალიან კარგი	108	8.5	100.0
სულ (N)	1,266	100.0	

შენიშვნა: MTF, Monitoring the Future.

ლოკაციის შემდეგი გავრცელებული საზომია *კვარტილი*. ქვედა კვარტილი 25-ე პერცენტილის ტოლია, ხოლო ზედა კვარტილი – 75-ე პერცენტილის. (მოძებნეთ ზედა კვარტილი ცხრილში 4.3). კოლეჯში მისაღები კომისიის წარმომადგენელი აპლიკანტების ზედა 25%-ს არჩევს, ამიტომ მან უნდა გაიანგარიშოს ზედა კვარტილი და კოლეჯში ჩარიცხოს ყველა, ვინც მოაგროვა 75-ე პერცენტილის ეკვივალენტური ან უფრო მაღალი ქულა.

საშუალო არითმეტიკული

საშუალო არითმეტიკული ცენტრალური ტენდენციის ყველაზე გავრცელებული საზომია. ის ძირითადად „საშუალოს“ სახელწოდებით არის ცნობილი. საშუალო არითმეტიკული გამოყენება ცენტრალური ტენდენციის გასაზომად ინტერვალურ-ფარდობით ცვლადებთან, როგორიცაა შემოსავალი, ასაკი და განათლება. თქვენ ალბათ გეცოდინებათ მისი გაანგარიშება. უბრალოდ ყველა მნიშვნელობა უნდა დაჯამდეს და გაიყოს მნიშვნელობების რაოდენობაზე.

საშუალო ცენტრალური ტენდენციის საზომი, რომელიც მიიღება ყველა მნიშვნელობის შეკრებით და მიღებული ჯამის გაყოფით მნიშვნელობების საერთო რაოდენობაზე. ეს არის არითმეტიკული საშუალო.

საშუალო არითმეტიკულს ხშირად იყენებენ დანაშაულის სტატისტიკის ანალიზში. ყოველწლიურად, აშშ-ის შინამეურნეობათა 25% სხვადასხვა დანაშაულის მსხვერპლი ხდება. აშშ-ში საკმაოდ გავრცელებულია ძალადობრივ ნიადაგზე ჩადენილი დანაშაულები, თუმცა სხვა განვითარებული ქვეყნების მაჩვენებლებს შორის ყველაზე მაღალი არ არის. მაგალითად, მკვლელობები აშშ-ში 5-ჯერ აჭარბებს ამ მაჩვენებელს ევროპაში.

ცხრილში 4.4 ნაჩვენებია მკვლელობის დონეები 2007 წელს (ყოველ 100,000 კაცზე გაანგარიშებით) აშშ-ის 15 უდიდესი ქალაქის მიხედვით. საჭიროა ცხრილში მოცემული ინფორმაციის განზოგადება ცენტრალური ტენდენციის რომელიმე საზომის გამოყენებით. რადგან ცვლადი „მკვლელობის დონე“ ინტერვალურ-ფარდობითია, უმჯობესია ამ მიზნით არითმეტიკული საშუალო გამოვიყენოთ.

ცხრილში 4.4 მონაცემების საშუალო არითმეტიკულის გასაანგარიშებლად უნდა შევკრიბოთ მკვლელობის დონეები თხუთმეტივე ქალაქის მიხედვით და გავყოთ ქალაქების რაოდენობაზე:

$$\text{საშუალო} = \frac{(6.0 + 10.2 + 15.7 + 16.2 + 27.3 + 13.8 + 4.7 + 8.9 + 9.3 + 16.1 + 45.8 + 3.5 + 2.1 + 14.3 + 15.4)}{15} = \frac{209.3}{15} = 13.95$$

ამრიგად, აშშ-ში 15 ქალაქის მიხედვით მკვლელობის საშუალო დონე 13.95-ია.[5]

ცხრილი 4.4 მკვლელობის დონეები (100,000 კაცზე) აშშ-ის 15 უდიდესი ქალაქის მიხედვით 2007 წელს

ქალაქი	მკვლელობის დონე 100,000 კაცზე
ნიუ-იორკი	6.0
ლოს-ანჯელესი	10.0
ჩიკაგო	15.7
ჰიუსტონი	16.2
ფილადელფია	27.3
ფენიქსი	13.8
სან-დიეგო	4.7
ლას-ვეგასი	8.9
სან-ანტონიო	9.3
დალასი	16.1
დეტროიტი	45.8
სან-ხოსე	3.5
ჰონოლულუ	2.1
ინდიანაპოლისი	14.3
ჯექსონვილი	15.4

შენიშვნა: U.S. Census Bureau, The 2010 statistical Abstract, Table 298.

საშუალოს გაანგარიშება

არსებობს არითმეტიკული საშუალოს გაანგარიშების მეორე გზა, რაც ფორმულების გამოყენებაში მდგომარეობს. სახელმძღვანელოს ამ ნაწილიდან ჩვენ ვიწყებთ ფორმულების გამოყენებას სხვადასხვა სტატისტიკური მაჩვენებლის გასაანგარიშებლად. ფორმულა შემოკლებული სახით გვიჩვენებს სხვადასხვა ოპერაციის თანმიმდევრობას განსაზღვრული შედეგის მისაღებად. მაგალითად, იმის მაგივრად, რომ ვთქვათ: „საშუალოს მისაღებად ყველა მნიშვნელობა უნდა შევკრებოთ და მიღებული ჯამი გავყოთ მნიშვნელობების საერთო რაოდენობაზე", შეგვიძლია უბრალოდ დავწეროთ ფორმულა:

$$\overline{Y} = \frac{\sum Y}{N} \qquad (4.1)$$

ცოტა ხანს შევჩერდეთ ახალი სიმბოლოების განმარტებებზე. Y-ით აღნიშნულია ცვლადის ნედლი მნიშვნელობები განაწილებაში; $\overline{Y}$ - გამოითქმის როგორც Y – ხაზიანი და აღნიშნავს ცვლადის საშუალოს. ბერძნული ასოთი გამოხატული სიმბოლო Σ გამოითქმის როგორც „სიგმა", ის ჯამის ნიშანია (+ ნიშნის მსგავსად), მიანიშნებს ამ სიმბოლოს შემდგომ ჩაწერილი ცვლადების შეჯამებაზე. ამიტომ $\sum Y$ ნიშნავს შემდეგს: „შეაჯამეთ Y-ის ყველა ნედლი მნიშვნელობა". და ბოლოს, N-ით აღნიშნულია დაკვირვებების მთლიანი რაოდენობა განაწილებაში.

მოდით, შევაჯამოთ სიმბოლოების მნიშვნელობები:

Y = ცვლადი Y-ის ნედლი მნიშვნელობები
$\overline{Y}$ = Y-ის საშუალო
$\sum Y$ = Y-ის მნიშვნელობების ჯამი
N = დაკვირვებების რაოდენობა

ახლა უკვე ცნობილია სიმბოლოების მნიშვნელობები. გავარჩიოთ შემდეგი მაგალითი. სტატისტიკის ჯგუფში ათი სტუდენტია. მათი ასაკებია:

21, 32, 23, 41, 20, 30, 36, 22, 25, 27

რას უდრის სტუდენტების საშუალო ასაკი?

ამ მონაცემებისათვის Y-ით აღნიშნულია სტუდენტების ასაკი, ხოლო N-ით მათი რაოდენობა, რომელიც 10-ის ტოლია. სტუდენტების ასაკების ჯამი $\sum Y$ ტოლია:

$$\sum Y = 21+32+23+41+20+30+36+22+25+27=277$$

აქედან, საშუალო ასაკი ტოლი იქნება:

$$\bar{Y} = \frac{\sum Y}{N} = \frac{277}{10} = 27.7$$

საშუალოს გაანგარიშება აგრეთვე შესაძლებელია სიხშირულ განაწილებებშიც. ქვემოთ, დეტალების განხილვაში 4.1 მოტანილია მაგალითი სიხშირული განაწილებით.

დეტალების განხილვა 4.1
საშუალოს პოვნა სიხშირულ განაწილებაში

როდესაც მონაცემები სიხშირული განაწილების სახით არის წარმოდგენილი, თითოეული მნიშვნელობა უნდა შევწონოთ, რაც ნიშნავს მის გადამრავლებას შესაბამის სიხშირეზე:

$$\bar{Y} = \frac{\sum fY}{N}$$

სადაც

Y = ცვლადი Y-ის ნედლი მნიშვნელობები

$\bar{Y}$ = Y-ის საშუალო

$\sum fY$ = fY-ის ყველა მნიშვნელობის ჯამი

N = დაკვირვებების რაოდენობა

ზემოთ მოყვანილი ფორმულის საშუალებით გავიანგარიშოთ საშუალო არითმეტიკული. GSS 2008-ის გამოკვლევაში რესპონდენტებს დაუსვეს კითხვა ოჯახში ბავშვების იდეალური რაოდენობის შესახებ. პასუხები წარმოდგენილია შემდეგი ცხრილის სახით:

ბავშვების იდეალური რაოდენობა ოჯახში: GSS 2008

ბავშვების რაოდენობა (Y)	სიხშირე (f)	სიხშირე $\times Y$ (fY)
0	14	0
1	24	24
2	475	950
3	251	753
4	73	292
5	11	55
6	2	12
7	2	14
სულ	N = 852	$\sum fY$ = 2,100

აღსანიშნავია, რომ $\sum fY$-ის (სვეტი 3) გამოსათვლელად თითოეული მნიშვნელობა (სვეტი 1) უნდა გავამრავლოთ მის სიხშირეზე (სვეტი 2) და მიღებული ნამრავლები შევკრიბოთ. როდესაც ჩავსვამთ ფორმულაში, მივიღებთ:

$$\bar{Y} = \frac{\sum fY}{N} = \frac{2,100}{852} = 2.46$$

ამრიგად, ბავშვების იდეალური საშუალო რაოდენობა 2.46-ია.

თუ სიძნელეები შეგექმნათ სიხშირულ განაწილებაში საშუალოს გაანგარიშებასთან დაკავშირებით, განიხილეთ ქვემოთ მოყვანილი ცხრილი. აქ ახსნილია გაანგარიშების მთელი პროცესი ყოველგვარი ანიშნების გარეშე.

✓ ცოდნის შემოწმება

საშუალოს პოვნა სიხშირულ განაწილებაში

	ადამიანების რაოდენობა ერთ სახლში	სახლების რაოდენობა	ადამიანების რაოდენობა ამ სახლებში
	1	3	3
	2	5	10
	3	1	3
	4	1	4

ადამიანების საერთო რაოდენობა: 20
სახლების საერთო რაოდენობა: 10
ადამიანების საშუალო რაოდენობა ერთ სახლში: 20/10 = 2

შემდეგი მაგალითი. ქვემოთ მოყვანილ ცხრილებში წარმოდგენილია ამერიკელი ინდიელების, ან ადგილობრივი ალასკელებისა და ესპანური წარმოშობის ამერიკელების განათლების ხანგრძლივობა GSS 2008-იდან. თითოეული ჯგუფისთვის გამოიყენეთ საშუალო არითმეტიკულის ფორმულა სიხშირული განაწილებისთვის.

ამერიკელი ინდიელების, ან ადგილობრივი ალასკელების განათლების ხანგრძლივობა, GSS 2008

განათლება (*Y*)	სიხშირე (f)	სიხშირე x Y (fY)
5	1	5
6	1	6
8	1	8
9	1	9
11	1	11
12	5	60
13	5	65
14	5	70
15	1	15
16	1	16
18	1	18
სულ	$N = 23$	$\sum fY = 283$
	$\overline{Y} = 12.30$	

ესპანური წარმოშობის ამერიკელების განათლების ხანგრძლივობა, GSS 2008

განათლება (Y)	სიხშირე (f)	სიხშირე × Y (fY)
0	2	0
3	1	3
4	1	4
6	4	24
8	3	24
9	4	36
10	1	10
11	7	77
12	15	180
13	3	39
14	6	84
15	2	30
16	6	96
18	1	18
სულ	$N = 56$	$\sum fY = 625$
	$\overline{Y} = 11.16$	

დააკვირდით ცხრილებს და აღმოაჩინეთ ამ ორ ჯგუფს შორის მსგავსება და განსხვავება. განათლება სოციალური კლასის მნიშვნელოვანი მახასიათებელია. ამ თავის ბოლოს SPSS-ის მესამე მაგალითში მითითებულია, თუ როგორ უნდა დავადგინოთ ურთიერთკავშირი სოციალურ კლასსა და წყვილებისათვის სასურველი ბავშვების რაოდენობას შორის.

✓ **ცოდნის შემოწმება** *მოცემულია იგივე განაწილება, რაც მედიანის ცოდნის შემოწმების დროს: 21, 32, 23, 41, 20, 30, 36, 22, 25, 27. გაიანგარიშეთ საშუალო. ემთხვევა თუ არა ერთმანეთს მედიანა და საშუალო?*

არითმეტიკული საშუალოს რამდენიმე მნიშვნელოვანი თვისება

ჩვენ განვიხილავთ საშუალო არითმეტიკულს სამ მათემატიკურ მახასიათებელს, რაც მას ცენტრალური ტენდენციის უმნიშვნელოვანეს საზომად აქცევს. ფაქტობრივად, ეს კონცეფცია ამოსავალია უფრო რთული სტატისტიკური მეთოდების გამოყენებისათვის.

გაზომვის ინტერვალურ-ფარდობითი დონე

საშუალოს გასაანგარიშებლად საჭიროა შეკრებისა და გაყოფის ოპერაციები. ამის გამო, მისი გაანგარიშება მხოლოდ ინტერვალურ-ფარდობითი ცვლადები-

სთვისაა შესაძლებელი. ეს ცვლადების გაზომვის ერთადერთი დონეა, რომელიც შეკრებასა და გაყოფას მოითხოვს.

მიზიდულობის ცენტრი

რადგან საშუალო (მოდისა და მედიანისგან განსხვავებით) განაწილების ყველა მნიშვნელობას აერთიანებს, მას შეიძლება მიზიდულობის (სიმძიმის) ცენტრი ვუწოდოთ. ეს ნიშნავს, რომ საშუალო ის წერტილია, რომელიც ძალიან კარგად აბალანსებს განაწილების მნიშვნელობებს. თუ თითოეულ მნიშვნელობას გამოვაკლებთ საშუალოს და სხვაობებს შევკრებთ, ჯამი ყოველთვის 0-ის ტოლი იქნება.

✓ *ცოდნის შემოწმება*

რატომ მიაჩნიათ საშუალო სიდიდე განაწილების მიზიდულობის ცენტრად? დაუშვათ, პარკში საქანელაზე (აიწონა-დაიწონა) ქანაობთ მეგობართან ერთად (შეიძლება ეს დიდი ხნის წინ იყო), რომელიც თქვენზე მძიმეა. თქვენ მარცხნივ ხართ გადაქანებული იქამდე, სანამ ის არ მიუახლოვდება საქანელის ცენტრს. მოკლედ, საქანელის დაბალანსება შეუძლია მსუბუქ ადამიანს, რომელიც ცენტრიდან (საშუალოდან) მოშორებულია, თუ მძიმეწონიანი ადამიანი ცენტრს (საშუალოს) უფრო მიუახლოვდება. შეგიძლიათ თუ არა ამ პრინციპის დემონსტრირება შემოსავლების მარტივი განაწილების მაგალითზე?

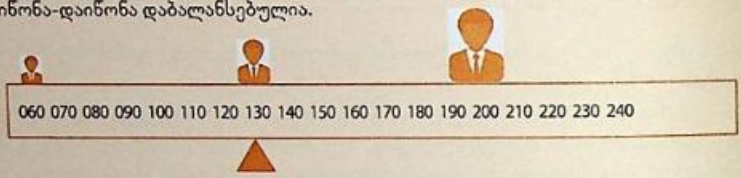
მგრძნობიარობა ექსტრემალური (კიდურა) წევრების მიმართ

ზემოთ მოყვანილი მაგალითებიდან დავრწმუნდით, რომ მოდისა და მედიანისაგან განსხვავებით, საშუალოს გაანგარიშებაში ცვლადის ყველა მნიშვნელობა მონაწილეობს. ეს კი განაპირობებს მის დამოკიდებულებას კიდურა მნიშვნელობებზე – საშუალო შეიძლება გადაიხაროს ძალიან დიდი ან ძალიან მცირე მნიშვნელობებისკენ. ამაში ადვილად დარწმუნდებით, თუ შეხედავთ ნახაზს 4.5.

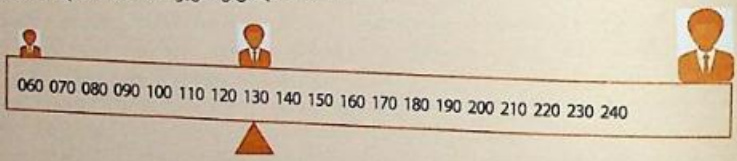
ნახაზზე 4.5ა და 4.5ბ გამოსახულია 10 ადამიანის შემოსავალი. ნახაზზე 4.5ბ ერთი ადამიანის შემოსავალი 5,000 აშშ დოლარიდან 35,000 აშშ დოლარამდე შეიცვალა. აღსანიშნავია, რომ ამან ზეგავლენა მოახდინა საშუალოზე. მისი მნიშვნელობა 3,000-დან 6,000-მდე გაიზარდა! საშუალომ არაპროპორციული ზეგავლენა განიცადა შედარებით მაღალი შემოსავლის მხრიდან და შეცდომაში შეგვიყვანა, რადგან უკვე ვეღარ ასახა განაწილების ცენტრალური ტენდენცია. შენიშნეთ, რომ მედიანა საშუალოზე უკეთ ასახავს ინფორმაციას ამ ჯგუფის ტიპური შემოსავლის შესახებ, ის არ განიცდის ექსტრემალური მნიშვნელობების ზეგავლენას და ისევ 3,000-ის ტოლია. შემდეგ ვნახავთ, რომ თუ განაწილების ერთ მხარეს რამდენიმე ექსტრემალური მნიშვნელობაა, საშუალო ვეღარ ასახავს ცენტრალურ ტენდენციას (რამდენიმე ექსტრემალური სიდიდე პრობლემას არ შექმნის, თუ ის განაწილების მხოლოდ ერთ მხარეს არ არის განლაგებული).

აიწონა-დაიწონას პრინციპის ილუსტრაცია

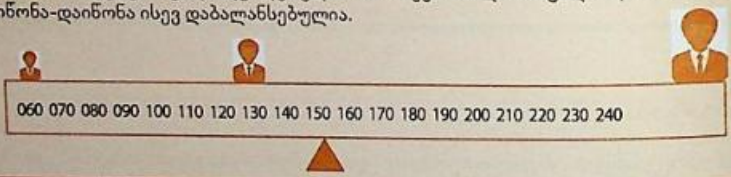
ა. სამი ადამიანი, რომლებიც იწონიან 60, 120 და 180 ფუნტს, დგას აიწონა-დაიწონაზე. სიმძიმის ცენტრი გადანაცვლებულია 120-სკენ. საშუალო ტოლია (60+120+180)/3. აიწონა-დაიწონა დაბალანსებულია.



ბ. 180-ფუნტიანი შეცვალა 240 ფუნტიანმა. სიმძიმის ცენტრი არ გადაგვინაცვლებია. აიწონა-დაიწონა მარჯვნივ გადაიხარა.



გ. სიმძიმის ცენტრი გადაადგილებულია 140-ისკენ. ახალი საშუალოა (60+120+240)/3. აიწონა-დაიწონა ისევ დაბალანსებულია.



✓ *ცოდნის შემოწმება*

როდესაც განაწილებისთვის შესაბამისი ცენტრალური ტენდენციის საზომის შერჩევა გჭირდებათ, საკმარისი არ არის მხოლოდ ცვლადის გაზომვის დონის ცოდნა. თუ ცვლადი ინტერვალურია, საშუალო ყველაზე კარგი არჩევანია, თუმცა თუ ექსტრემალური მნიშვნელობები განაწილების ერთ მხარესაა მოქცეული, მაშინ საშუალო არარეპრეზენტატული ხდება და უმჯობესია მედიანისა და მოდის გამოყენება.

ნახაზი 4.5 საშუალო განიცდის ექსტრემალური წევრების ზემოქმედებას: (ა) ექსტრემალური მნიშვნელობების გარეშე და (ბ) ერთი ექსტრემალური მნიშვნელობა

ა) ექსტრემალური მნიშვნელობების გარეშე: საშუალო არის $3,000

შემოსავალი (Y)	სიხშირე (f)	*fY*
1,000	1	1,000
2,000	2	4,000
3,000	4	12,000
4,000	2	8,000
5,000	1	5,000
	N = 10	Σ fY = 30,000

$$\text{საშუალო} = \frac{\Sigma fY}{N} = \frac{30,000}{10} = \$3,000$$

მედიანა = $3,000

ბ) ერთი ექსტრემალური მნიშვნელობა: საშუალო არის $6,000

შემოსავალი (Y)	სიხშირე (f)	*fY*
1,000	1	1,000
2,000	2	4,000
3,000	4	12,000
4,000	2	8,000
35,000	1	35,000
	N = 10	Σ fY = 60,000

$$\text{საშუალო} = \frac{\Sigma fY}{N} = \frac{60,000}{10} = \$6,000$$

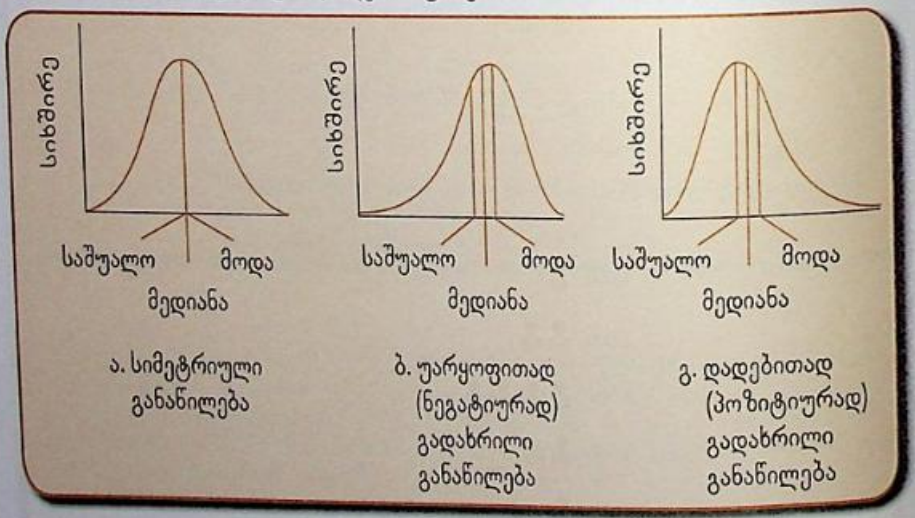
მედიანა = $3,000

განაწილების ფორმა: ტელევიზია, განათლება და და-ძმები

ამ თავში ჩვენ გავარჩიეთ, თუ როგორ ასახავს მოდა, მედიანა და საშუალო განაწილების ცენტრალურ ტენდენციას. განაწილების (აქ შევეხებით მხოლოდ ინტერვალური ცვლადების განაწილებას) აღწერა შესაძლებელია აგრეთვე მათი ზოგადი ფორმის მიხედვით, რაც ვიზუალურად ადვილად გამოისახება. განაწილება შეიძლება იყოს სიმეტრიული, ან გადახრილი, რაც დამოკიდებულია განაწილების ერთ მხარეს არსებულ ექსტრემალურ მნიშვნელობებზე.

განაწილება **სიმეტრიულია** (ნახაზი 4.6ა), თუ სიხშირეები განაწილების ორივე მხარეს იდენტურია; ანუ, თუ მას ორ ნაწილად გავყოფთ, ერთი ნაწილი მეორის სარკისებრ ანარეკლს წარმოადგენს. უნიმოდალურ, სიმეტრიულ განაწილებაში საშუალო, მოდა და მედიანა ერთმანეთის იდენტურია.

ნახაზი 4.6 სიხშირული განაწილების ტიპები



გადახრილი განაწილების ერთ მხარეზე რამდენიმე ექსტრემალური სიდიდე მდებარეობს. თუ განაწილების ექსტრემალური მნიშვნელობები დაბალია, მაშინ ის ნეგატიურად (უარყოფითად) არის გადახრილი (ნახაზი 4.6ბ). ამ დროს საშუალო სიდიდე ყოველთვის ნაკლებია მედიანაზე. სხვა სიტყვებით რომ ვთქვათ, საშუალო დაბალი მნიშვნელობებისკენ იხრება. საპირისპიროდ, თუ განაწილების ექსტრემალური მნიშვნელობები მაღალია, მაშინ ის პოზიტიურად (დადებითად) გადახრილია (ნახაზი 4.6გ). ამ შემთხვევაში საშუალო ყოველთვის მედიანაზე მეტია, ის მაღალი მნიშვნელობებისკენ მიისწრაფვის.

სიმეტრიული განაწილება სიხშირეები განაწილების მარჯვენა და მარცხენა მხარეს იდენტურია. ერთი ნაწილი მეორის სარკისებრ ანარეკლს წარმოადგენს.

გადახრილი განაწილება განაწილება ერთ მხარეზე წარმოდგენილი ექსტრემალური მნიშვნელობებით.

ნეგატიურად გადახრილი განაწილება განაწილება რამდენიმე ექსტრემალურად დაბალი მნიშვნელობით.

პოზიტიურად გადახრილი განაწილება განაწილება რამდენიმე ექსტრემალურად მაღალი მნიშვნელობით.

მოვახდინოთ ამ სამი განსხვავებული განაწილების დემონსტრირება GSS 2008[6]-ის ცვლადების მაგალითზე. ამ ცვლადების სიხშირული განაწილება მოცემულია ცხრილებში 4.5, 4.6 და 4.7, ხოლო მათი შესაბამისი გრაფიკები წარმოდგენილია ნახაზებზე 4.7, 4.8 და 4.9.

ცხრილი 4.5 ყოველდღიურად ტელევიზორთან გატარებული საათების რაოდენობა

ყოველდღიურად ტელევიზორთან გატარებული საათების რაოდენობა	სიხშირე (f)	fY	პროცენტი	C%
1	201	201	33.0	33.0
2	261	522	42.9	75.9
3	147	441	24.1	100.
სულ	609	1,164	100.0	

$$\bar{Y} = \frac{\sum fY}{N} = \frac{1,164}{609} = 1.91$$

მედიანა = 2.0
მოდა = 2.0

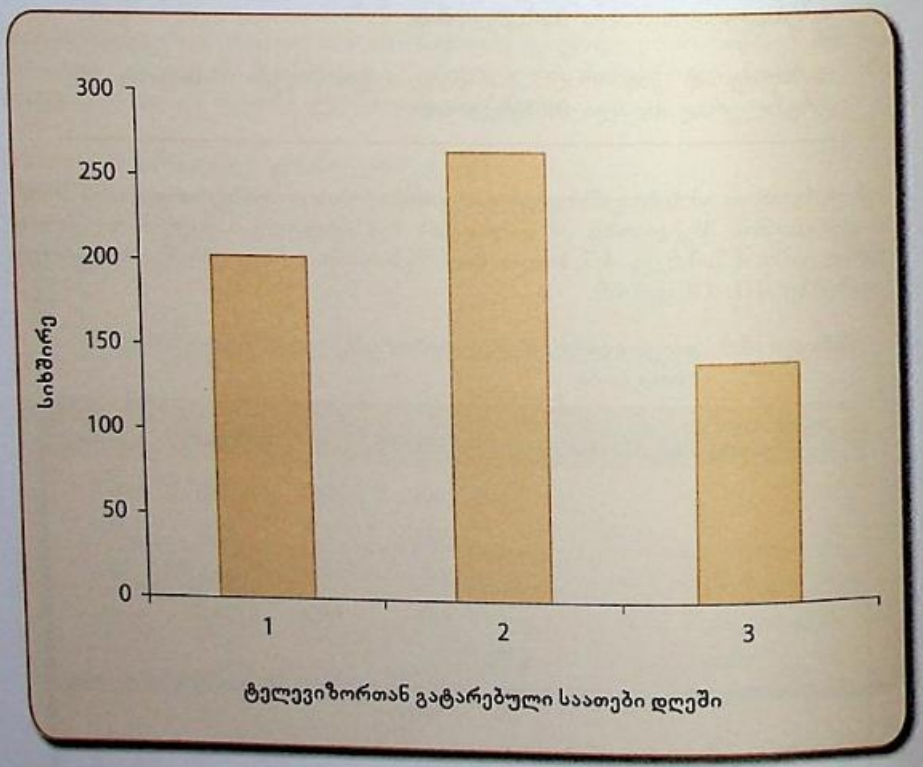
სიმეტრიული განაწილება

უპირველესად, განვიხილოთ ცხრილი 4.5 და ნახაზი 4.7. აღსანიშნავია, რომ ადამიანების ყველაზე დიდი რაოდენობა (261) დღეში 2 საათს უყურებს

ტელევიზორს (მოდა = 2.0), აგრეთვე დაახლოებით იგივე რაოდენობამ (201 და 147) მიუთითა 1 ან 3 საათი. როგორც ჩანს ნახაზზე 4.7, მოდა, მედიანა და საშუალო იდენტური სიდიდეებია და ისინი ერთმანეთს ემთხვევა განაწილების შუაში.

ეს განაწილება დაახლოებით სიმეტრიულია, რადგან მოდა, მედიანა და საშუალო ერთმანეთის იდენტურია.

ნახაზი 4.7 ყოველდღიურად ტელევიზორთან გატარებული საათების რაოდენობა



პოზიტიურად გადახრილი განაწილება

ახლა შევხედოთ ცხრილს 4.6 და ნახაზს 4.8, სადაც ასახულია რესპონდენტების და-ძმების რაოდენობა.

ცხრილი 4.6 და-ძმების რაოდენობა

და-ძმების რაოდენობა	სიხშირე (f)	fY	პროცენტი	C%
1	68	0	4.8	4.8
2	286	286	20.2	25.0
3	301	602	21.3	46.3
4	239	717	16.9	63.2
5	189	756	13.4	76.6
6	117	585	8.3	84.9
7	87	522	6.2	91.1
8	70	490	4.9	96.0
სულ	57	456	4.0	100.0

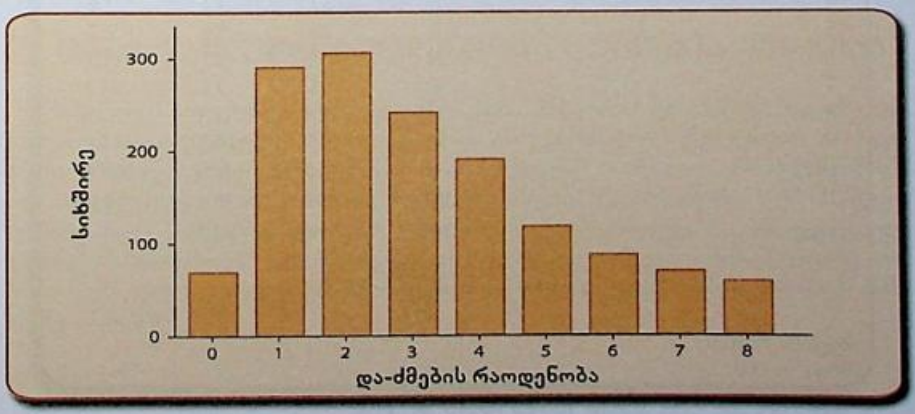
$$\bar{Y} = \frac{\sum fY}{N} = \frac{4,414}{1,414} = 3.12$$

მედიანა = 3.0
მოდა = 2.0

აღსანიშნავია, რომ რესპონდენტების უდიდეს რაოდენობას (301) ორი ძმა და/ან და ყავს. ამასთან, ექვსი და მეტი დედმამიშვილი მათ მცირე რაოდენობას ყავს. გარდა ამისა, საშუალო, მოდა და მედიანა განსხვავებულია, რომელთაგან მოდას ყველაზე დაბალი მნიშვნელობა აქვს (მოდა = 2.0), შემდეგ მოდის მედიანა (მედიანა = 3.0) და უდიდესია საშუალო არითმეტიკული (საშუალო = 3.12).

ეს განაწილება ასახულია ცხრილში 4.6 და ნახაზზე 4.8, როგორც პოზიტიურად გადახრილი. როგორც წესი, გადახრილ განაწილებებში საშუალო, მოდა და მედიანა ერთმანეთს არ ემთხვევა. საშუალო სიდიდე იქით იხრება, სადაც განაწილების მნიშვნელობები უმცირესია.

ნახაზი 4.8 და-ძმების რაოდენობა



ნეგატიურად გადახრილი განაწილება

ახლა შეეხედოთ ცხრილს 4.7 და ნახაზს 4.9, რომელიც ასახავს სკოლაში გატარებული წლების რაოდენობას იმ რესპონდენტებისათვის, ვისაც საშუალო სკოლის დიპლომი არ აუღია. აქ საპირისპირო სურათს ნახავთ. ეს განაწილება ნეგატიურად გადახრილია. უპირველესად აღსანიშნავია, რომ სკოლაში გატარებული წლების უდიდესი რაოდენობა კონცენტრირებულია მაღალი მნიშვნელობებისკენ (11 წელი) და დაბალი მნიშვნელობების მხარეს სულ ცოტა რესპონდენტია. როგორც წინა მაგალითში, აქაც საშუალო, მოდა და მედიანა ერთმანეთისგან განსხვავებულია. ამგვარად, აქ მოდას უდიდესი მნიშვნელობა აქვს (მოდა = 11.0), შემდეგ მოდის მედიანა (მედიანა = 10.0) და საშუალო უმცირესია (საშუალო = 8.89).

მითითებები განაწილების ფორმის განსაზღვრისათვის

მოვიყვანთ საჭირო მითითებებს განაწილების ფორმის განსაზღვრისათვის.

1. უნიმოდალურ განაწილებაში, სადაც მოდა, მედიანა და საშუალო ერთმანეთს ემთხვევა ან ერთმანეთის იდენტურია, განაწილება სიმეტრიულია.
2. როდესაც საშუალო მედიანაზე მეტია (ან მდებარეობს მედიანის მარჯვნივ), განაწილება პოზიტიურად გადახრილია.
3. როდესაც საშუალო მედიანაზე ნაკლებია (ან მდებარეობს მედიანის მარცხნივ), განაწილება ნეგატიურად გადახრილია.

ცხრილი 4.7 უდიპლომო რესპონდენტების მიერ სკოლაში გატარებული წლების რაოდენობა

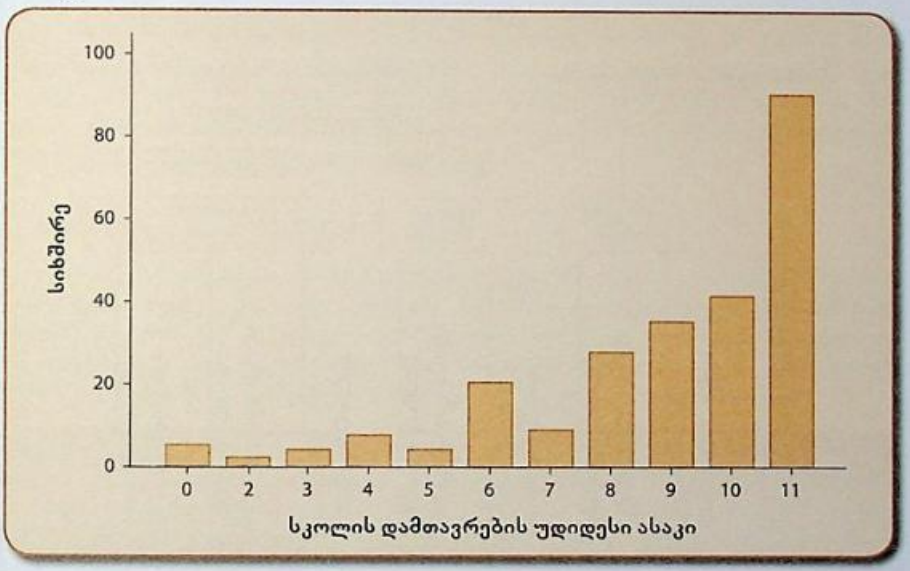
წლების რაოდენობა	სიხშირე (f)	fY	პროცენტი	C%
1	5	0	2.0	2.0
2	2	4	0.8	2.8
3	4	12	1.6	4.4
4	8	32	3.2	7.6
5	4	20	1.6	9.2
6	21	126	8.5	17.7
7	9	63	3.6	21.3
8	28	224	11.3	32.6
9	35	315	14.2	46.8
10	41	410	16.6	63.4
11	90	990	36.4	99.8[ა]
სულ	247	2,196	99.8	

$$\bar{Y} = \frac{\sum fY}{N} = \frac{2,196}{247} = 8.89$$

მედიანა = 10.0
მოდა = 11.0

ა. დამრგვალების გამო პასუხი 100-ის ტოლი არ არის.

ნახაზი 4.9 უდიპლომო რესპონდენტების მიერ სკოლაში გატარებული წლების რაოდენობა



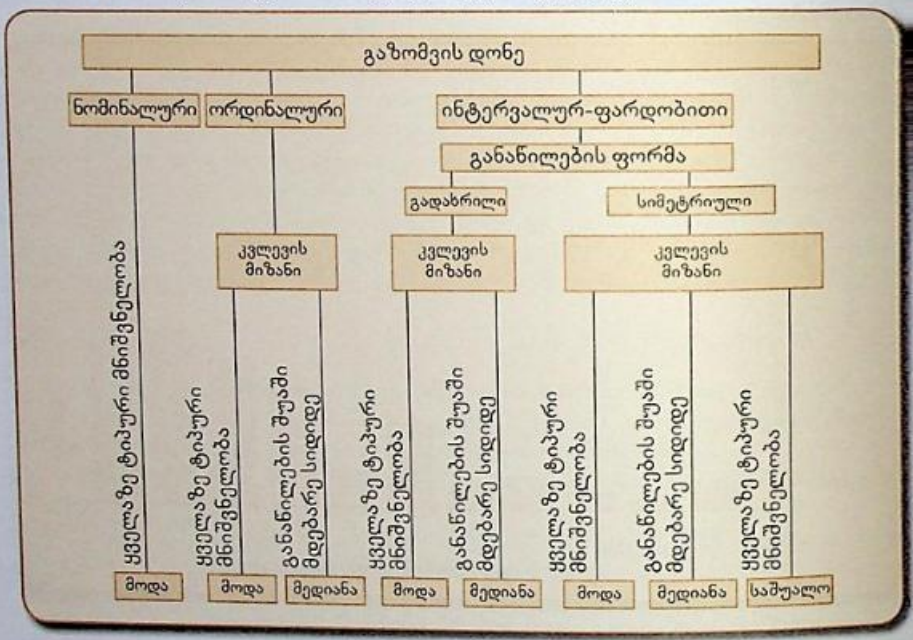
პოზიტიურად და ნეგატიურად გადახრილი განაწილებების განსაზღვრისათვის დახედეთ განაწილების ბოლოს. თუ ის მარჯვნივ მდებარეობს (აბსცისთა ღერძის დადებითი ბოლო), მაშინ განაწილება პოზიტიურად გადახრილია, ხოლო თუ ბოლო მარცხნივაა გადახრილი, (უარყოფითი ან აბსცისთა ღერძის პოტენციურად უარყოფითი ბოლო), მაშინ ის ნეგატიურად გადახრილია.

✓ *ცოდნის შემოწმება*

ცენტრალური ტენდენციის საზომის შერჩევა

აქამდე ჩვენ განვიხილეთ საშუალოების სამი ძირითადი ტიპი: მოდა, მედიანა და საშუალო არითმეტიკული. თითოეული მათგანი ცენტრალურ ტენდენციას ასახავს, მაგრამ რომელი უნდა გამოვიყენოთ? მოდა? მედიანა? საშუალო არითმეტიკული? თუ სამივე ერთად? ამ კითხვაზე მარტივი პასუხი არ არსებობს. თუმცა ჩვეულებრივ, ჩვენი არჩევანი ერთ-ერთ მათგანზე ჩერდება ხოლმე და ის გარკვეულ განხილვას მოითხოვს. აღნიშნულთან დაკავშირებული მსჯელობის პროცესი და მიღებული გადაწყვეტილებები წარმოდგენილია გადაწყვეტილების მიღების ხის სახით ნახაზზე 4.10.

ნახაზი 4.10 როგორ შევარჩიოთ ცენტრალური ტენდენციის საზომი



გაზომვის დონე

ცენტრალური ტენდენციის საზომის არჩევის დროს დიდი ყურადღება უნდა დაეთმოს ცვლადის გაზომვის დონეს. საშუალოების სამივე ტიპისათვის გამოიყენება მათთვის შესაბამისი ან უფრო მაღალ დონეზე გაზომილი ცვლადები. ამიტომ, როგორც ნაჩვენებია ნახაზზე 4.10, ნომინალური ცვლადების შემთხვევაში ჩვენი არჩევანი მხოლოდ მოდის გამოყენებით შემოიფარგლება.

ორდინალურ მონაცემებთან ორი არჩევანი გვაქვს: მოდა ან მედიანა (ზოგჯერ ორივე). არჩევანი დამოკიდებულია იმაზე, თუ რის გაგება გვინდა განაწილებიდან. თუ გვაინტერესებს განაწილების ტიპური სიდიდის ჩვენება, ვიყენებთ მოდას, ხოლო თუ გვაინტერესებს, რომელი სიდიდეა მოთავსებული განაწილების შუაში, მაშინ ვიყენებთ მედიანას.

როდესაც მონაცემები ინტერვალურ-ფარდობით დონეზეა გაზომილი, მაშინ არჩევანი ცოტა რთულდება და შეზღუდულია განაწილების ფორმის გათვალისწინებით.

გადახრილი განაწილება

გადახრილი განაწილების დროს, რადგან ის ექსტრემალური მნიშვნელობების გავლენას განიცდის, საშუალო სიდიდემ შეიძლება არასწორი ინფორმაცია მოგვა-

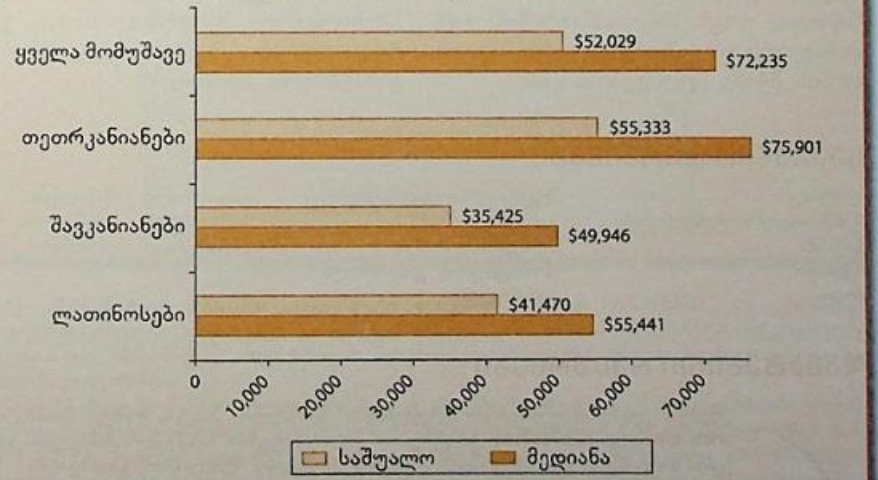
წოდოს ცენტრალური ტენდენციის შესახებ. ასეთ შემთხვევებში უმჯობესია მოდისა და მედიანის გამოყენება (იხ. დეტალების განხილვა 4.2), რადგან ისინი ექსტრემალური მნიშვნელობების გავლენას არ განიცდიან. მაგალითად, როდესაც ვიხილავდით უდიპლომო რესპონდენტების სკოლაში გატარებული წლების ხანგრძლივობას (ცხრილი 4.7, ნახაზი 4.9), საშუალომ რეპრეზენტატულად ვერ დაახასიათა წლების „ტიპური" რაოდენობა, ამიტომ გამოვიყენეთ მოდა და მედიანა. ამრიგად, ამ სიდიდეთაგან ერთ-ერთი მათგანის გამოყენება „საშუალოს" სახით, კვლევის მიზანზეა დამოკიდებული.

სიმეტრიული განაწილება

როდესაც განაწილება სიმეტრიულია, შეგვიძლია სამივე ტიპის საშუალოს გამოყენება. თუმცა, ეს ისევ კვლევის მიზანზეა დამოკიდებული და აგრეთვე იმაზეც, თუ რის გაგება გვსურს განაწილების შესახებ. მთლიანობაში, საშუალო არითმეტიკული საუკეთესო არჩევანია, რადგან ინფორმაციის დიდ მოცულობას აერთიანებს და მისი გამოყენება კომპლექსურ ანალიზში უფრო ადვილია.

დეტალების განხილვა 4.2
შეცდომის ზჭარე: შემოსავლის ასახვა

ინდივიდუალური შემოსავალი ხშირად პოზიტიურად არის გადახრილი, რადგან მხოლოდ მცირე რაოდენობის ადამიანებს აქვთ მაღალი შემოსავალი. ამის გამო, საშუალო არითმეტიკული რეპრეზენტატულად ვერ ახასიათებს "საშუალო" შემოსავალს. მაგალითად, 2008 წელს ჩატარებულმა ამერიკული საზოგადოების კვლევამ გამოაქვეყნა მონაცემები თეთრკანიანების, შავკანიანების და ლათინო-ამერიკელების შინამეურნეობებში საშუალოწლიური და მედიანური შემოსავლების შესახებ. ეს მაჩვენებლები შედარებულია ქვემოთ მოტანილ ნახაზზე.



წყარო: U.S. Census Bureau, *American Community Survey*, 2008.

როგორც ხედავთ, თითოეულ ჯგუფში საშუალო მეტია მედიანაზე. ეს იმაზე მიუთითებს, რომ შინამეურნეობების შემოსავლები აშშ-ში ძალიან გადახრილია, ამიტომ საშუალო რეალურ სურათს ვერ ასახავს. მოსახლეობის სხვადასხვა ჯგუფებში საშუალო შემოსავლების ასახვისათვის უფრო მისაღებია მედიანის გამოყენება (როგორც ეს ნაჩვენებია).

აშკარაა, რომ თეთრკანიანების შემოსავლები ყველაზე მაღალია, ამიტომ საინტერესოა თითოეული ჯგუფის მედიანური შემოსავლის შედარება აღნიშნულ ჯგუფთან. მაგალითად, თეთრკანიანებისათვის გადახდილ ყოველ $1-ზე შავკანიანებს უხდიდნენ 64 ცენტს ($35,425/$55,333), ხოლო ლათინო-ამერიკელებს – 75 ცენტს ($41,470/$55,333).

ძირითადი განმარტებები:

• მოდა, მედიანა და საშუალო ცენტრალური ტენდენციის საზომებია – განაწილებისთვის ტიპური ან საშუალო მახასიათებლები.

• მოდა არის კატეგორია ან მნიშვნელობა, რომელსაც ყველაზე მაღალი სიხშირე (ან ხვედრითი წილი) აქვს განაწილებაში. ის ხშირად გამოიყენება ყველაზე მეტად განმეორებადი ნომინალური კატეგორიის დასახასიათებლად.

• მედიანა არის მნიშვნელობა, რომელიც განაწილებას ორ თანაბარ ნაწილად ყოფს. ის გაიანგარიშება ორდინალური და გაზომვის უფრო მაღალი დონის მქონე ცვლადებისათვის.

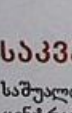
• საშუალო სიდიდე ძირითადად გამოიყენება ინტერვალურ-ფარდობითი ცვლადების ცენტრალური ტენდენციის დასახასიათებლად, როგორიცაა შემოსავალი, ასაკი, განათლება. საშუალო არითმეტიკული მიიღება ყველა მნიშვნელობის შეკრებით და მიღებული ჯამის შეფარდებით მნიშვნელობათა რიცხვზე (N).

• სიმეტრიულ განაწილებაში სიხშირეები განაწილების მარჯვენა და მარცხენა ბოლოებთან იდენტურია. გადახრილ განაწილებებში გვაქვს როგორც ექსტრემალურად მაღალი (პოზიტიურად გადახრილი), ისე ექსტრემალურად დაბალი (ნეგატიურად გადახრილი) სიდიდეები.

საკვანძო ტერმინები

საშუალო
ცენტრალური ტენდენციის საზომი
მედიანა
მოდა
ნეგატიურად გადახრილი განაწილება
პერცენტილი
პოზიტიურად გადახრილი განაწილება
გადახრილი განაწილება
სიმეტრიული განაწილება
დამატებითი რესურსები

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

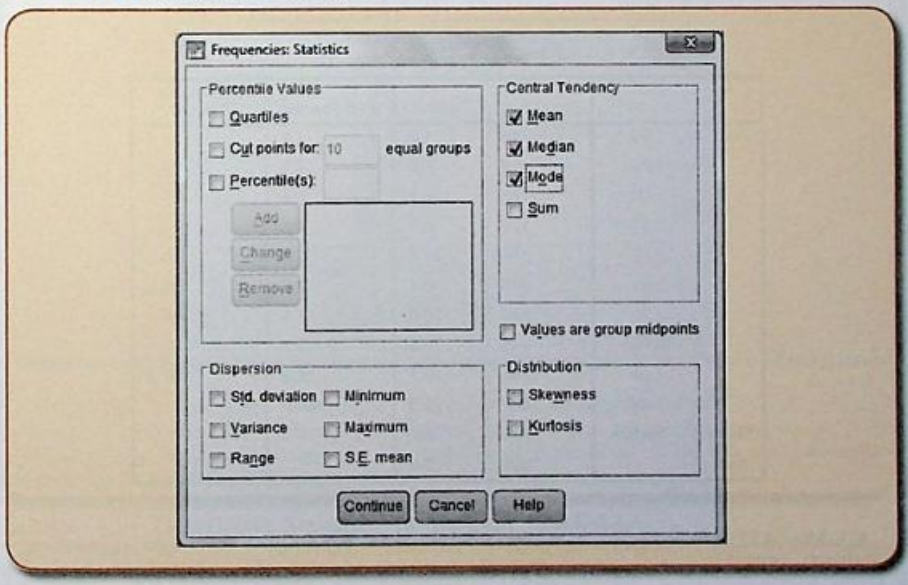
[MTF2008]

მაგალითი 1: ცენტრალური ტენდენციის საზომების გაანგარიშება ბრძანებით Frequencies

მე-2 თავში განხილული სიხშირეების (Frequences) ბრძანების დახმარებით შესაძლებელია აგრეთვე ამ თავში განხილული ცენტრალური ტენდენციის სამივე საზომის გაანგარიშება. ამ თავში ჩვენ გამოვიყენებთ Frequences ბრძანებას ცვლადების GPA და ATWEED (დამოკიდებულება მარიჰუანის მოხმარებისადმი) ცენტრალური ტენდენციის საზომების გაანგარიშებისთვის.

დაკლიკეთ Analyse, Descriptive Statistics და შემდეგ Frequences. მოათავსეთ GPA და ATWEED ცვლადების ჯგუფში, შემდეგ დაკლიკეთ ღილაკი Statistics. თქვენ დაინახავთ ცენტრალური ტენდენციის (Central Tendency) ფანჯრის (ნახაზი 4.11) ამოსარჩევ ჩამონათვალს, საიდანაც მხოლოდ პირველი სამი გვჭირდება (Mean, Median, Mode). შემდეგ დაკლიკეთ Continue და Ok.

ნახაზი 4.11 დიალოგური ფანჯარა Statistics



აქ გამოსახულია Statistics დიალოგური ფანჯარა GPA და ATWEED ცვლადებისათვის, აგრეთვე მოცემულია სიხშირული განაწილება GPA-სთვის (ნახაზი 4.12). GPA ინტერვალური ცვლადია, რაც იმას ნიშნავს, რომ მისთვის შესაძლებელია როგორც მოდის, ისე მედიანას და საშუალოს გაანგარიშება.

GPA-ის მედიანა 3.30-ის ტოლია და ძალიან ახლოსაა საშუალოს მნიშვნელობასთან. გავიანგარიშეთ რა, მედიანის მნიშვნელობა, შეგვიძლია მიახლოებით ვთქვათ, რომ რესპონდენტების ნახევარს 3.30-ზე ნაკლები რეიტინგი აქვს, ხოლო მეორე ნახევარს – 3.30-ზე მეტი. ამასთან, რადგან საშუალო მედიანაზე ნაკლებია (3.13<3.30), GPA-ის განაწილება ნეგატიურად გადახრილია. და ბოლოს, GPA ბიმოდალური განაწილებაა, რადგან აქ ორი მოდაა, ანუ ორი კატეგორია გამოიყოფა

თითქმის ერთიდაიგივე უმაღლესი სიხშირით. ასეთ შემთხვევაში SPSS მოდალურად აცხადებს მოდის უფრო მცირე მნიშვნელობას, თუმცა სქოლიოში უჩვენებს, რომ განაწილებაში ორი მოდაა. თუ სიხშირულ განაწილებას დავაკვირდებით ნახაზზე 4.12, ვნახავთ, რომ 3.70 GPA ჰქონდა 248 რესპონდენტს. ასევე, რესპონდენტების დაახლოებით იგივე რაოდენობამ (246) მიუთითა, რომ მათი GPA 3.00-ია. ამიტომ, GPA-ის განაწილება არის ბიმოდალური და მისი მოდებია 3.00 და 3.70.

ნახაზი 4.12 შერჩეული შედეგები GPA და ATWEED-ისათვის

Statistics

		Grade Point Average	Attitude Toward Trying Marijuana
N	Valid	1400	1454
	Missing	100	46
Mean		3.1281	1.86
Median		3.3000	2.00
Mode		3.70	1

Grade Point Average

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	19	1.3	1.4	1.4
	1.70	38	2.5	2.7	4.1
	2.00	78	5.2	5.6	9.6
	2.30	121	8.1	8.6	18.3
	2.70	174	11.6	12.4	30.7
	3.00	246	16.4	17.6	48.3
	3.30	240	16.0	17.1	65.4
	3.70	248	16.5	17.7	83.1
	4.00	236	15.7	16.9	100.0
	Total	1400	93.3	100.0	
Missing	System	100	6.7		
Total		1500	100.0		

ნახაზზე 4.12 გამოსახულია აგრეთვე ცენტრალური ტენდენციის საზომები ცვლადისთვის ATWEED. SPSS-ისთვის უცნობია, რომ ეს ცვლადი ორდინალურია. ამიტომ ის შედეგების ფანჯარაში იგივე ცენტრალური ტენდენციის საზომებს გამოიტანს. ცენტრალური ტენდენციის სწორი საზომის შერჩევა თქვენზეა დამოკიდებული.

რადგან ATWEED ორდინალური ცვლადია, ცენტრალური ტენდენციის საზომებად შეიძლება გამოვიყენოთ მოდა და მედიანა. მედიანა 2-ის ტოლია და შესაბამება პასუხს "არ ვადასტურებ" (disapprove). კუმულაციური პროცენტული განაწილებიდან (აქ არ არის ნაჩვენები) შეგვიძლია ვნახოთ, რომ რესპონდენტების 68.4% ან „არ ადასტურებს" ან „არ უარყოფს" (don't disapprove) მარიხუანის მოხმარებას. დანარჩენი 31.6%, რომელიც აქ არ არის ნაჩვენები, „სრულიად უარყოფს" მარიჰუანის მოხმარებას. მოდა არის „არ უარვყოფ", რაც იმას ნიშნავს, რომ ეს პასუხი ყველაზე ხშირად განმეორდა.

[GSS08PFP-B]

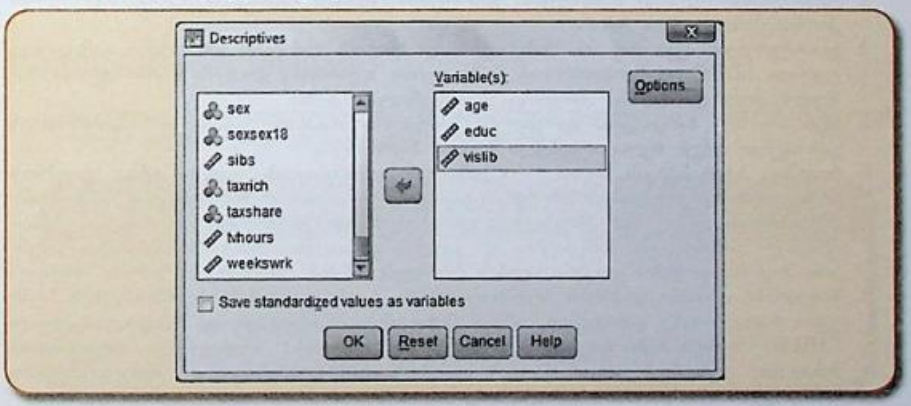
მაგალითი 2: ცენტრალური ტენდენციის საზომების მიღება ბრძანებით Descriptives

თუ შედეგების მიღება სქესობრივ ჭრილში გსურთ, უნდა შეასრულოთ შემდეგი ოპერაციები: შეარჩიეთ Data, Split File, Organize Output by Groups. შეიტანეთ ცვლადი sex ყუთში სახელწოდებით „Groups Based on". დაკლიკეთ Ok. ახლა SPSS გაფილტრავს შედეგებს სქესის მიხედვით.

თუ გსურთ ინტერვალური ცვლადის საშუალოს გაანგარიშება, მაგრამ არ გაინტერესებთ სიხშირული ცხრილის ნახვა, მაშინ უნდა აირჩიოთ Descriptives. მისი მოძებნის თანმიმდევრობაა: Analyze, Descriptive Statistics და ბოლოს Descriptives.

Descriptives დიალოგური ფანჯარა (ნახაზი 4.13) რთული არ არის და მოითხოვს მხოლოდ ჩვენთვის საინტერესო ცვლადის (AGE, EDUC, VISLIB) მითითებას ცვლადების ყუთში (Variables box). დათქმის თანახმად, Descriptives გაიანგარიშებს საშუალოს, სტანდარტულ გადახრას (მე-5 თავში იქნება განხილული), მინიმუმს, მაქსიმუმს და ვალიდური პასუხების რაოდენობას.

ნახაზი 4.13 დიალოგური ფანჯარა Descriptives



აღწერითი სტატისტიკის მონაცემები ქალებისათვის შედეგების ფანჯრის ქვედა ნაწილში განლაგდება. ნახაზზე 4.13 ნაჩვენებია აღწერითი სტატისტიკა ცვლადებისათვის ასაკი (Age), უკანასკნელ წელს საჯარო ბიბლიოთეკაში ვიზიტების რაოდენობა (VISLIB) და განათლების ხანგრძლივობა წლებში (EDUC) ქალებისათვის 2008 წელს.

ნახაზი 4.14 Descriptive Statistics (აღწერითი სტატისტიკა)

FEMALE	AGE OF RESPONDENT	803	18	89	48.16	17.981
	HIGHEST YEAR OF SCHOOL COMPLETED	805	0	20	13.33	3.036
	HOW OFTEN R VISISTED PUBLIC LIBRARY LAST YEAR	574	0	250	13.29	26.599
	Valid N (listwise)	567				

ა. რესპონდენტების სქესი = ქალები

მივიღეთ შემდეგი შედეგები: ქალების საშუალო ასაკი 48.16 წელი, სწავლის ხანგრძლივობა დაახლოებით 13 წელი, წინა წლის განმავლობაში საჯარო ბიბლიოთეკაში საშუალოდ 13 ვიზიტი.

ამოცანები SPSS-ისთვის

[GSS08PFP-A]

1. ცვლადისათვის HOMOSEX შექმენით სიხშირული განაწილება ცენტრალური ტენდენციის შესაბამისი საზომებით.
ა. ცენტრალური ტენდენციის რომელი საზომებია უფრო მისაღებია ამ შემთხვევაში, საშუალო თუ მედიანა? ახსენით რატომ.
ბ. ვთქვათ, გვაინტერესებს ქალებისა და მამაკაცების დამოკიდებულება ჰომოსექსუალური ურთიერთობების მიმართ. შექმენით სიხშირული განაწილება ცვლადისათვის HOMOSEX. ქალებისა და მამაკაცებისათვის ცალ-ცალკე (შეარჩიეთ Data, Split File, Organize Output by Groups. შეიტანეთ ცვლადი SEX ჯგუფში სახელწოდებით „Groups Based on"). არის თუ არა განსხვავება მათი ცენტრალური ტენდენციის საზომებში? ახსენით (დაიმახსოვრეთ, სავარჯიშოს დასრულების შემდეგ გამორთეთ Split File ბრძანება შემდეგი თანმიმდევრობით: Data, Split File, Analyze All Cases, Do Not Create Groups).
გ. დაუბრუნდით (ბ)-ს, ამ შემთხვევაში გამოიყენეთ ცვლადი PREMARSX (დამოკიდებულება ქორწინებამდე სექსის მიმართ).
2. გვაინტერესებს, აქვთ თუ არა მამაკაცებსა და ქალებს თანაბარი განათლება. გამოიყენეთ ცვლადი EDUC და Frequances-ის საშუალებით სქესობრივ ჭრილში წარმოადგინეთ საშუალო, მოდა და მედიანა (როგორც ეს აღწერილია 1.ბ-ში).
ა. აქვთ თუ არა მამაკაცებსა და ქალებს საშუალოდ თანაბარი განათლება? პასუხისთვის გამოიყენეთ თქვენ ხელთ არსებული მთელი ინფორმაცია.
ბ. როდესაც სტატისტიკას სოციალური სამყაროს აღწერისათვის ვიყენებთ, უნდა ვიფიქროთ იმაზე, გვეხმარება თუ არა ეს მონაცემები განსხვავებული სოციალური ჯგუფების შესწავლაში, ანუ, გაანიავთ თუ არა მათ პრაქტიკული ღირებულება. ფიქრობთ თუ არა, რომ ჩვენ მიერ აღმოჩენილი განსხვავებები მამაკაცებისა და ქალების განათლების დონეებში საკმარისია სამუშაოს შოვნისა და შემოსავლების მიღების შესაძლებლობების ახსნისათვის? პასუხი ახსენით.
3. ზოგიერთი ადამიანი ფიქრობს, რომ სოციალური კლასი გავლენას ახდენს ოჯახში ბავშვების რაოდენობაზე. გამოიყენეთ აშშ-ის მონაცემები ამ საკითხის შესასწავლად. (ცვლადი CHILDS ზომავს რესპონდენტების ბავშვების რაოდენობას). აუცილებელი ინფორმაციის მისაღებად, ჯერ შეასრულეთ ფაილის დაყოფის (Split) ოპერაცია და შემდეგ გადადით სიხშირეების (Frequances) გამოყენებაზე.
ა. ცენტრალური ტენდენციის რომელი საზომია საუკეთესო ბავშვების რაოდენობისათვის? რატომ?
ბ. რომელ სოციალურ კლასს ჰყავს უფრო მეტი ბავშვი?
გ. გააკეთეთ ანალიზი ცვლადისათვის CHILDIDEL (ბავშვების იდეალური რაოდენობა). არის თუ არა განსხვავება სოციალურ ჯგუფებს შორის? ახსენით.
4. განაწილებისათვის შესაბამისი ცენტრალური ტენდენციის საზომის შერჩევა უმნიშვნელოვანესი უნარ-ჩვევაა. GSS08PFP-B-ზე დაყრდნობით შეარჩიეთ შესაბამისი საზომი(ები) შემდეგი ცვლადებისათვის:
ა. რა სიხშირით უყურებენ რესპონდენტები ტელევიზორს? [TVHOURS]
ბ. რესპონდენტების პოლიტიკური შეხედულებები [POLVIEWS]
გ. უკანასკნელ წელს ნამუშევარი კვირების რაოდენობა [WEEKSWRK]
დ. აქვთ თუ არა სახლში იარაღი [OWNGUN]
ე. შერჩევაში მოხვედრილთა და-ძმები [SIBS]
ვ. უჭერენ თუ არა მხარს მარიხუანის ლეგალიზაციას [GRASS]
5. დეტალების განხილვაში 4.1 შედარებულია ამერიკელი ინდიელების, აღგილობრივი ალიასკელებისა და ლათინო-ამერიკელების განათლების დონეები. გამოიყენეთ SPSS სხვა ცვლადების ანალოგიური შედარებებისათვის. ავიღოთ ცვლადი „ქორწინებით კმაყოფილების დონე" (HAPMAR), ამისათვის შეგადაროთ ქალებისა და მამაკაცებისათვის ყოველდღიუ-

რად ტელევიზორის ყურების ხანგრძლივობა. ამ ოპერაციის შესასრულებლად რამდენიმე გზა არსებობს, მათგან ყველაზე ადვილია ბრძანება Split File (როგორც აღწერილია I.ბ-ში). გამოიყენეთ GSS08PFP-A მონაცემთა ბაზა. დაკლიკეთ Data, Split File, Organize Output by Groups, მოათავსეთ ცვლადები HAPMAR da SEX (ამ თანმიმდევრობით) ყუთში „Groups Based on" და დაკლიკეთ Ok. HAPMAR-ს აქვს სამი პასუხი: 1 = ძალიან ბედნიერი, 2 = საკმაოდ ბედნიერი, 3 = არც ისე ბედნიერი. SEX-ს 2 ვალიდური პასუხი აქვს: 1 = მამაკაცი, 2 = ქალი. ამჯერად თქვენ მიეცით დავალება SPSS-ს თითოეული ჯგუფისათვის ცალ-ცალკე გამოიტანოს შედეგები ცვლადების HAPMAR და SEX კომბინაციის შედეგად.

ა. შექმენით სიხშირული განაწილება. დაკლიკეთ Analyze, Descriptive Statistics, Frequances. შეარჩიეთ ცვლადი TVHOURS და აირჩიეთ სასურველი ცენტრალური ტენდენციის საზომი ოპციიდან Statistics. დაკლიკეთ Ok. SPSS უამრავ შედეგს გამოიტანს. თქვენ უნდა შეარჩიოთ შესაბამისი სიხშირული განაწილება და ცენტრალური ტენდენციის საზომები. მაგალითად, მამაკაცებისათვის ქორწინებით კმაყოფილების დასადგენად, მამაკაცების სიხშირულ ცხრილში მოძებნეთ = „ძალიან ბედნიერი ვარ" და რესპონდენტის სქესი = მამაკაცი.

ბ. რა განსხვავებები შეამჩნიეთ ქალებსა და მამაკაცებს შორის ტელევიზორის ყურებასა და ქორწინებით კმაყოფილებაში?

თავის სავარჯიშოები

1. მოცემული სიხშირული განაწილება მოიცავს შემდეგ ინფორმაციას: ბავშვების დამოკიდებულება დღეში 1 კოლოფი სიგარეტის მოწევისადმი.

დამოკიდებულება დღეში 1 კოლოფი სიგარეტის მოწევისადმი	სიხშირე	ხვედრითი წილი, პროცენტი	კუმულაციური ხვედრითი წილი
1. არ უარვყოფ	288	19.8	19.8
2. არ ვეთანხმები	418	28.7	48.5
3. სრულიად არ ვეთანხმები	749	51.5	100.0
სულ	1,455	100.0	

წყარო: Monitoring The Future, 2008.

ა. იპოვეთ მოდა
ბ. იპოვეთ მედიანა
გ. გაუკეთეთ ინტერპრეტაცია მოდასა და მედიანას
დ. რატომ არ ანგარიშობთ საშუალოს ამ ცვლადისათვის?

2. ზოგიერთი სექსუალური კავშირი პოლიტიკური განსჯის საგნად იქცა. GSS 2008-ის რესპონდენტებს დაუსვეს კითხვა ჰომოსექსუალური ურთიერთობების შესახებ. პასუხები ოთხ კატეგორიად ჩამოყალიბდა. იხილეთ პასუხების სიხშირული განაწილება:

ჰომოსექსუალური ურთიერთობები	სიხშირე	ხვედრითი წილი, პროცენტი	კუმულაციური ხვედრითი წილი
1. ყოველთვის არასწორია	468	49.3	49.3
2. თითქმის ყოველთვის არასწორია	35	3.7	53.0
3. ხანდახან არასწორია	73	7.7	60.7
4. არ არის სწორი	374	39.3	100.0
სულ	950	100.0	

ა. რომელ დონეზე იზომება ეს ცვლადი? რა არის ამ ცვლადის შიდა?
ბ. გაიანგარიშეთ მედიანა ამ ცვლადისთვის. როგორ დაახასიათებთ ზოგადად საზოგადოების დამოკიდებულებას ამ ურთიერთობების მიმართ?
3. ქვემოთ მოყვანილი სიხშირული განაწილება აღწერს 65 მოზრდილი ლათინო-ამერიკელის მიერ ნამუშევარი კვირების რაოდენობას წინა წლის განმავლობაში.

წინა წელს ნამუშევარი კვირების რაოდენობა	სიხშირე	ხვედრითი წილი, პროცენტი	კუმულაციური ხვედრითი წილი
0	10	15.4	15.4
8	1	1.5	16.9
12	2	3.1	20.0
15	1	1.5	21.5
24	1	1.5	23.0
25	1	1.5	24.5
26	1	1.5	26.0
30	1	1.5	27.5
32	1	1.5	29.0
36	2	3.1	32.1
38	1	1.5	33.6
39	1	1.5	35.1
40	5	7.7	42.8
44	2	3.1	45.9
45	1	1.5	47.4
47	2	3.1	50.5
48	1	1.5	52.0
49	1	1.5	53.5
50	2	3.1	56.6
51	1	1.5	58.1
52	27	41.5	99.6
სულ	65	99.6	

წყარო: General Social Survey, 2008.

ა. რომელ დონეზე იზომება „წინა წელს ნამუშევარი კვირების რაოდენობა"? რა არის ამ ცვლადის შიდა? მედიანა?
ბ. წინა წელს ნამუშევრები კვირების რაოდენობისთვის ააგეთ კვარტილები. რას უდრის 25-ე პერცენტილი? 50-ე პერცენტილი? 75-ე პერცენტილი? 50-ე პერცენტილის გაანგარიშება აღარ არის საჭირო, რატომ?
4. MTF 2008-დან მოტანილია მონაცემები რესპონდენტების მეგობრების მიერ ალკოჰოლური სასმელის გამოყენების შესახებ:

მეგობრების რაოდენობა, ვინც სვამს ალკოჰოლურ სასმელს	სიხშირე
არავინ	125
ძალიან ცოტა	177
ცოტა	262
უმრავლესობა	521
ყველა	216

ა. გაიანგარიშეთ მედიანა.
ბ. რომელი კატეგორია შედის მე-20 და მე-80 პერცენტილებში?

5. იცვლება თუ არა ჯანმრთელობის მდგომარეობა ასაკების მიხედვით? GSS 2008-იდან მოტანილია მონაცემები ჯანმრთელობის მდგომარეობის შესახებ სხვადასხვა ასაკობრივი ჯგუფის მიხედვით (ყველა ჯგუფი არ არის განხილული).

ჯანმრთელობის მდგომარეობა	ასაკობრივი ჯგუფი			
	18-29	30-39	40-49	50-59
ძალიან კარგი	64	57	46	37
კარგი	86	87	94	83
დამაკმაყოფილებელი	36	27	37	47
ცუდი	2	5	8	15

ა. გაიანგარიშეთ მოდა და მედიანა თითოეული ჯგუფისათვის.

ბ. გამოიყენეთ ეს ინფორმაცია და აღწერეთ, როგორ იცვლება ჯანმრთელობის მდგომარეობა ასაკობრივ ჯგუფებში. რომელი უფრო უკეთესად აღწერს მონაცემებს, მოდა თუ მედიანა? სტატისტიკურ მონაცემებზე დაყრდნობით დაასაბუთეთ, რომ ზოგიერთი ასაკობრივი ჯგუფი უფრო ჯანმრთელია, ვიდრე დანარჩენი.

6. როგორც მოსალოდნელი იყო, ამერიკელების მოთხოვნილება სამედიცინო მომსახურებაზე იზრდება „ბებიბუმერების" თაობის ასაკის მომატებასთან ერთად. ქვემოთ მოტანილია ეს ინფორმაცია 2000 და 2005 წლებისთვის. შენიშვნა: ციფრები ასახულია ათასებში.

შტატი	2000	2005
ალაბამა	685	740
დელავერი	112	125
ფლორიდა	2,804	3,008
ილინოისი	1,635	1,674
მინესოტა	654	691
ნიუ-ჰემპშირი	170	185
ნიუ-იორკი	2,715	2,758
ვაშინგტონი	736	807

წყარო: U.S. Census Bureau, The 2008 *Statistical Abstract*, Table 135.

ა. გაიანგარიშეთ ამერიკელების საშუალო მიმართვიანობა ჯანდაცვის მომსახურებაზე რვავე შტატის მიხედვით 2000 და 2005 წლებში. როგორ დაახასიათებდით აღნიშნულ ცვლილებებს? ახასიათებს თუ არა ადეკვატურად მიმართვების საშუალო სიდიდე ამ განაწილების ცენტრალურ ტენდენციას? რატომ დიახ, ან არა?

ბ. ამოიღეთ განაწილებიდან ფლორიდა, ილინოისი და ნიუ-იორკი. ახასიათებს თუ არა დარჩენილი 5 შტატის საშუალო სიდიდე უკეთესად ცენტრალურ ტენდენციას ორივე წლისათვის? ახსენით.

7. გასული წლების განმავლობაში აშშ-ის შინამეურნეობების ზომა სულ უფრო შემცირდა. GSS 2008-ის კვლევა შეიცავს მონაცემებს რესპონდენტების შინამეურნეობებში მცხოვრები 18 წლისა და ზევით მოსახლეობის რიცხოვნობის შესახებ. გაიანგარიშეთ 2008 წლისთვის აშშ-ის შინამეურნეობებში მცხოვრებთა საშუალო რაოდენობა.

შინამეურნეობის ზომა	სიხშირე
1	384
2	541
3	228
4	195
5	98
6	36
7	12
8	4
9	1
11	1
სულ	1,500

8. მე-6 სავარჯიშოში უკვე გაიანგარიშეთ ამერიკელების საშუალო მოთხოვნილება სამედიცინო დახმარებაზე. ახლა ვნახოთ, ეს განაწილებები სიმეტრიულია, თუ გადახრილი?
ა. რვა შტატის მაგალითზე გაიანგარიშეთ მოდა და მედიანა. ამ მონაცემებზე დაყრდნობით, როგორ დახასიათებდით ამერიკელების განაწილებას ორივე წლის მიხედვით?
ბ. რომელი უფრო უკეთესად ახასიათებს განაწილებას, მოდა თუ მედიანა? რატომ?
გ. თუ აღმოაჩენთ, რომ განაწილება გადახრილია, რა სახის სტატისტიკურ განმარტებას დაურთავთ აღნიშნულ მოსაზრებას?
9. გამოიყენეთ მე-7 სავარჯიშოს მონაცემები და ააგეთ შინამეურნეობების ზომის ამსახველი ჰისტოგრამა.
ა. რა შეგიძლიათ თქვათ ჰისტოგრამაზე დაკვირვების შედეგად — როგორ არის გადახრილი განაწილება: დადებითად თუ უარყოფითად? რატომ?
ბ. ახლა გაიანგარიშეთ ამ განაწილების მედიანა და შეადარეთ ის მე-7 სავარჯიშოში გაანგარიშებულ საშუალოს. დაგეხმარებათ თუ არა ეს ციფრები თქვენი მოსაზრების დასაბუთებაში, თუ როგორ არის განაწილება გადახრილი? რატომ ფიქრობთ, რომ შინამეურნეობების ზომის განაწილება ასიმეტრიულია?
10. მე-3 სავარჯიშოს მონაცემების გამოყენებით გაიანგარიშეთ:
ა. წლიურად ნამუშევარი კვირების საშუალო რაოდენობა.
ბ. შეადარეთ საშუალო და თქვენ მიერ უკვე გაანგარიშებული მედიანა. ჰისტოგრამის აგების გარეშე აღწერეთ, რომელ მხარეს და როგორ არის გადახრილი განაწილება?
11. ვთქვათ, თქვენ უსმენთ პოლიტიკოსების დებატებს აშშ-ის ეკონომიკურ მდგომარეობაზე. ერთი პოლიტიკოსი ამბობს, რომ ამერიკელი მომუშავეების საშუალო წლიური შემოსავალი $72,235-ია, ხოლო მეორე ამტკიცებს, რომ მდგომარეობა არც თუ სახარბიელოა, რადგან ამერიკელი მომუშავეების საშუალო წლიური შემოსავალი $52,029-ია. შეიძლება თუ არა, რომ ორივე პოლიტიკოსი მართალი იყოს? თუ ასეა, ახსენით.
12. იმსჯელეთ ცენტრალური ტენდენციის სამივე საზომის დადებითი და უარყოფითი მხარეების შესახებ. დარწმუნებული ხართ თუ არა იმაში, რომ ერთ-ერთი მათგანი ყველაზე უკეთ ასახავს ცენტრალურ ტენდენციას? თუ ასეა, ახსენით რატომ?
13. იცვლება თუ არა მკვლელობის დონე ქალაქების ზომასთან ერთად? შეისწავლეთ ქვემოთ მოტანილი მონაცემები და ქალაქების ორივე ჯგუფისათვის გაიანგარიშეთ საშუალო და მედიანა. სად არის მკვლელობის დონე უფრო მაღალი? განსხვავდება თუ არა საშუალო და მედიანა ქალაქების ამ ორ ჯგუფში?

მკვლელობების დონე 100,000 კაცზე 2005 წელს			
ტოპ ათეული მოსახლეობის რაოდენობით	მკვლელობების დონე	ქვედა ათეული მოსახლეობის რაოდენობით	მკვლელობების დონე
ნიუ-იორკი	6.6	კორპუს-კრისტი	2.8
ლოს ანჟელესი	12.6	ბაკერსფილდი	11.2
ჩიკაგო	15.6	ბუფალო	19.8
ჰიუსტონი	16.3	სტოკტონი	14.6
ფილადელფია	25.6	ნიუარკი	34.5
ფენიქსი	15.0	სენტ-პაული	8.6
ლას-ვეგასი	11.3	ენქორიჯი	5.8
სან-დიეგო	4.0	ლექსინგტონი	5.6
სან-ანტონიო	6.8	სენტ-პიტერსბერგი	11.8
დალასი	16.4	მობაილი	14.0

წყარო: U.S. Census Bureau, The 2008 *Statistical Abstract*, Table 302.

14. პოლიტიკოსებს აინტერესებთ მოზარდების აზრი ნარკოტიკების მოხმარებაზე. გამოიყენეთ ქვემოთ მოტანილი ინფორმაცია სქესობრივ ჭრილში:

დამოკიდებულება მარიჰუანის მოწევისადმი	მამაკაცები	ქალები
არ უარვყოფ	325	302
არ ვეთანხმები	160	157
სრულიად არ ვეთანხმები	185	260
სულ	670	719

წყარო: Monitoring The Future, 2008.

გაიანგარიშეთ ცენტრალური ტენდენციის შესაბამისი მაჩვენებლები მამაკაცებისა და ქალებისათვის. როგორ აღწერდით მათ დამოკიდებულებას ნარკოტიკების მოხმარებისადმი? თქვენც იგივე აზრის ხართ თუ არა? რატომ? თქვენი მოსაზრების დასაბუთებისთვის დაეყრდენით თქვენ მიერ გაანგარიშებულ ცენტრალური ტენდენციის საზომს.

15. პირადი მონაცემების ქურდობა ზრდადი ტენდენციით ხასიათდება. ქვემოთ მოყვანილია აშშ-ის ფედერალური სავაჭრო ორგანიზაციის 2006 წლის მონაცემები 11 შტატის მიხედვით.

შტატი	პირადი მონაცემების ქურდობის შემთხვევები
არკანზასი	384
ფლორიდა	17,780
ილინოისი	10,080
მიჩიგანი	6,784
ნიუ-ჯერსი	6,394
ნიუ-იორკი	16,452

ჩრდილოეთ დაკოტა	189
ორეგონი	2,815
როდ-აილენდი	615
ტეხასი	20,006
ვისკონსინი	2,536

წყარო: Monitoring The Future, 2008.

ა. გაიანგარიშეთ საშუალო და მედიანა 11 შტატის მიხედვით.
ბ. არის თუ არა განაწილება გადახრილი? თუ ასეა, საიდან მიხვდით?
გ. ორი გზით ახსენით, რატომ აქვს ზოგ შტატს ქურდობის უფრო მაღალი მაჩვენებელი, ვიდრე სხვებს?

თავი 5

ვარიაციის საზომები

თავის სწავლების მიზნები

- ვარიაციის გაზომვის მნიშვნელობა
- თვისობრივი ვარიაციის ინდექსის (IQV), დიაპაზონის, ინტერკვარტილური დიაპაზონის, დისპერსიისა და სტანდარტული გადახრის (საშუალო კვადრატული გადახრის) გაანგარიშება და ინტერპრეტაცია
- ვარიაციის საზომის შერჩევის კრიტერიუმი

წინა თავში ჩვენ განვიხილეთ ცენტრალური ტენდენციის საზომები: საშუალო არითმეტიკული, მედიანა და მოდა. ამ საზომების საფუძველზე მხოლოდ ერთი რიცხვით შეიძლება განაწილების საშუალო ან ტიპური ტენდენციის დახასიათება. საშუალოების გაანგარიშებას დიდი მნიშვნელობა აქვს კვლევის პროცესში, თუმცა მხოლოდ მათი გამოყენება არასაკმარისია, რადგან შეიძლება შეცდომაში შეგვიყვანოს. სტატისტიკური კვლევის შემდეგი ეტაპია განაწილების ვარიაციისა და მრავალფეროვნების აღწერა რიცხვების საშუალებით. ასეთ რიცხვებს ვარიაციის საზომები ეწოდება. მონაცემების აღწერისათვის მკვლევარები ხშირად ერთად იყენებენ ცენტრალური ტენდენციისა და **ვარიაციის საზომებს**.

ვარიაციის საზომები რიცხვები, რომლებიც გამოიყენება ცვლადის განაწილების მრავალფეროვნებისა და ვარიაციის დახასიათებისათვის.

ამ თავში ჩვენ განვიხილავთ ვარიაციის ხუთ საზომს: თვისობრივი ვარიაციის ინდექსს (IQV), დიაპაზონს, ინტერკვარტილურ დიაპაზონს, დისპერსიასა და სტანდარტულ გადახრას. ჯერ განვიხილოთ მათი მნიშვნელობა.

ვარიაციის გაზომვის მნიშვნელობა

განვიხილოთ ვარიაციის შესწავლის მნიშვნელობა აშშ-ის ქალების განსხვავებული გამოცდილების მაგალითზე. ხდება თუ არა ქალების გაერთიანება მსგავსების საფუძველზე ან დაყოფა განსხვავებების მიხედვით? სწორი პასუხია ორივე. სწორია. მსგავსების განხილვა განსხვავებების შესწავლის გარეშე „შეცდომაში შეგვიყვანს იმ საკითხებზე მსჯელობისას, რაც აშორებს და აახლოებს ქალებს ერთმანეთთან".[1] მაშინაც კი, როდესაც ქალების ერთ კონკრეტულ ჯგუფზე ვახდენთ კონცენტრაციას, მნიშვნელოვანია როგორც მსგავსებების, ისე განსხვავებების განხილვა. ავიღოთ, მაგალითად აზიური წარმოშობის ამერიკელი ქალები. როგორც ჯგუფს, მათ საერთო ნიშნები ახასიათებთ.

სხვა ეთნიკურ ჯგუფებთან შედარებით ისინი უფრო აქტიურად ეწევიან საქმიანობას სამუშაო ადგილზე. ბევრი მათგანი მხარში უდგას დანარჩენებს, ხშირად თავიანთ ცხოვრებას დიდი ოჯახების მოთხოვნილებებს უქვემდებარებენ. თუმცა, გარემოებათა გამო ეს გამოცდილება არასაკმარისია მათი მდგომარეობის აღსაწერად. აზიელ-ამერიკელ ქალთა შორის ბევრია აშშ-ში დაბადებული და ისინიც... ვინც... შემდეგ ჩამოვიდნენ. მათი წარმოშობის ქვეყნებიც განსხვავებულია: ჩინეთი, იაპონა, ფილიპინები, კორეა... და ინდოეთი. ისინი ძირითადად „მუშათა კლასს" მიეკუთვნებიან, მათ შორის ბევრია „ღარიბი" და „საშუალო კლასის" წარმომადგენელი „წარმატებულის" სტერეოტიპის ნაცვლად.[2]

როგორც ეს მაგალითი გვიჩვენებს, სტერეოტიპების ჩამოყალიბების საფუძვლად გვევლინება რაიმე ერთი ჯგუფის დახასიათება მისი ცენტრალური სიდიდით, ჯგუფის შიგნით არსებული განსხვავებების იგნორირების ხარჯზე. სოციოლოგები ხშირად ხელს უწყობენ ასეთი სტერეოტიპების ჩამოყალიბებას, როდესაც მათი ემპირიული განზოგადებები მხოლოდ საშუალოებს შორის განსხვავებების მარტივი ახსნით შემოიფარგლება. ყოველივე ეს საშუალოებთან ერთად ვარიაციის გამოყენების აუცილებლობაზე მიგვითითებს, როდესაც ვახასიათებთ ან ვადარებთ ჯგუფებს. თუ აზიელი-ამერიკელი ქალების მსგავსებები ცენტრალური ტენდენციით ხასიათდება, მათ შორის განსხვავებების ასახვა მხოლოდ ვარიაციის მაჩვენებლებით არის შესაძლებელი.

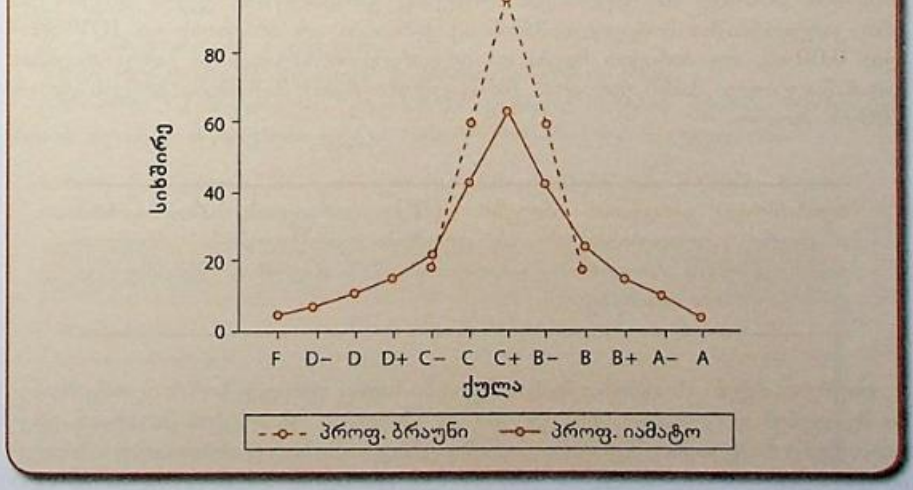
ვარიაციის კონცეფცია არა მხოლოდ აზიელი ქალების შემთხვევაში გამოგადგებათ, არამედ ყოველდღიურ ცხოვრებაშიც. აკადემიური პერსონალის მუდმივ საზრუნავს წარმოადგენს კურიკულუმების მორგება სტუდენტთა მოთხოვნებზე. განვიხილოთ კოლეჯში სტატისტიკის ლექტორის მაგალითი.

სტატისტიკა სოციალურ მეცნიერებათა კურიკულუმში ალბათ ყველაზე „საინერვიულო" საგანია. შეიძლება ითქვას, რომ ის თითქმის ყოველთვის „უკანასკნელი წინაღობაა" კურსის დამთავრების გზაზე. ერთ-ერთ ხელისშემშლელ ფაქტორად „მათემატიკისადმი შიში" გვევლინება, რომელიც არახელსაყრელ გარემოს ქმნის სტატისტიკის შესწავლისათვის; ეს იმაში გამოიხატება, რომ სტუდენტები არსის გაგების ნაცვლად მეთოდების დაზეპირებას ცდილობენ.

წარმოვიდგინოთ საუნივერსიტეტო კომიტეტის სხდომა, რომლის დღის წესრიგშიც აღნიშნული საკითხების მოგვარებაა. კომიტეტის წევრები ცდილობენ შეფასების სისტემების შედარებას სხვადასხვა ფაკულტეტებზე წაკითხული სტატისტიკის კურსების შესაფასებლად. სოციოლოგიის ფაკულტეტზე დანერგილ სისტემას იყენებს პროფესორი ბრაუნი, ხოლო სოციალური სამსახურის სკოლაში გამოყენებულ სისტემას – პროფესორი იამატო. კომიტეტმა აღმოაჩინა, რომ წლების მანძილზე პროფესორი ბრაუნის კლასებში საშუალო შეფასება C+ იყო, ასევე იყო პროფესორ იამატოსთანაც. თუ ვიფიქრებთ, რომ ორივე პროფესორს ერთნაირი შეფასების სისტემა ჰქონდა, შეცდომას დავუშვებთ. ჩვენ დეტალურად უნდა განვიხილოთ ქულების განაწილება თითოეულ კლასში. ეს განსხვავებები ასახულია ნახაზზე 5.1, რომელიც წარმოადგენს ორივე კლასის სიხშირული განაწილების პოლიგონს.

შეადარეთ ერთმანეთს ამ განაწილებების ფორმები. პროფესორ იამატოს კლასში ქულები A-დან F-მდეა გაბნეული, ხოლო ბრაუნის კლასში საშუალოს გარშემოა განლაგებული და B-დან C-მდე იცვლება. მიუხედავად იმისა, რომ ორივე პროფესორის ქულების საშუალოები ერთმანეთის ტოლია, იამატოს ქულები უფრო განსხვავებულია, ვიდრე ბრაუნის. ამ ორი კლასის შედარება არც ისე ადვილი ყოფილა, როგორც ჩვენ გვეგონა!

ნახაზი 5.1 ქულების განაწილება პროფესორების ბრაუნისა და იამატოს სტატისტიკის კლასებში



როგორც ამ მაგალითიდან ჩანს, ინფორმაცია საშუალოს ირგვლივ ქულების გაბნევის შესახებ ისეთივე მნიშვნელოვანია, როგორც განაწილების ცენტრალური ტენდენციის შესახებ. ასეთი ტიპის ინფორმაცია ვარიაციის საზომებიდან მიიღება.

ოდნის
ოწება

კარგად დააკვირდით ნახაზს 5.1. თქვენ რომელ კლასს აირჩევდით? თუ ჩაჭრაზე ნერვიულობთ, ბრაუნის კლასს აირჩევდით, რადგან აქ ჩაჭრილები არ არიან. თუ მუშაობა და GPA-ის ამაღლება გსურთ, მაშინ იამატოს კლასი სჯობს. თუ არჩევანს მხოლოდ საშუალოების საფუძველზე გააკეთებთ, თქვენი არჩევანი ცუდად ინფორმირებული იქნება.

თვისობრივი ვარიაციის ინდექსი: მოკლე შესავალი

მიმდინარე ეტაპზე აშშ დემოგრაფიული ცვლილებების წინაშეა: მიმდინარეობს დომონანტურად ევროპული მოსახლეობის ჩანაცვლების პროცესი რასობრივი, ეთნიკური და კულტურული მრავალფეროვნებით. ეს გარკვეული გამოწვევაა აშშ-ის ევროპული მოსახლეობის შესახებ დაგროვილი გამოცდილების გადასასინჯად. მრავალი კითხვა ისმის რასობრივად და ეთნიკურად განსხვავებული ჯგუფების შესახებ. მაგალითად, გვჭირდება შედარებები რასობრივი/ეთნიკური ნიშნის მიხედვით სხვადასხვა ქალაქებში, რეგიონებში ან შტატებში, ან გვაინტერესებს დროთა განმავლობაში ჯგუფების შიგნით მიმდინარე განსხვავებების მრავალფეროვნების საკითხი.

თვისობრივი ვარიაციის ინდექსი (IQV) ვარიაციის საზომია ნომინალური ცვლადებისათვის, როგორიცაა რასა და ეთნიკურობა. ინდექსმა შეიძლება მიიღოს 0.00-დან 1.00-მდე მნიშვნელობები. როდესაც განაწილებაში ყველა დაკვირვება ერთ კატეგორიაშია მოქცეული, მაშინ აქ ვარიაცია არ არსებობს და IQV ტოლია 0.00-ის. თუ პირიქით ხდება და დაკვირვებები სხვადასხვა კატეგორიებშია გადანაწილებული, მაშინ ვარიაცია (მრავალფეროვნება) მაქსიმუმს აღწევს და ის 1.00-ის ტოლია.

თვისობრივი ვარიაციის ინდექსი (IQV) ვარიაციის საზომია ნომინალური ცვლადებისათვის. ის ეფუძნება განაწილებაში არსებული განსხვავებების რაოდენობის ფარდობას მაქსიმალურად შესაძლებელ განსხვავებათა რიცხვზე.

დაუშვათ, თქვენ ცხოვრობთ მეინის შტატში, სადაც უმრავლესობას თეთრკანიანები შეადგენენ და ლათინო-ამერიკელები და აზიელები უმცირესობაში არიან. ვთქვათ, თქვენი საუკეთესო მეგობარი ჰავაიზე ცხოვრობს, სადაც ძირითადად აზიელები და ადგილობრივი ჰავაელები ცხოვრობენ. ცხრილში 5.1 მოცემულია ამ ორი შტატის მოსახლეობის განაწილებები. სად უფრო მეტია განსხვავება? რა თქმა უნდა, ჰავაიზე, რადგან აქ მოსახლეობის ნახევარს აზიელები ან ადგილობრივი ჰავაელები შეადგენენ, ხოლო მეინის შტატში აზიელები უმცირესობაში არიან.

ცხრილი 5.1 ხუთი ძირითადი რასობრივი/ეთნიკური ჯგუფის ხვედრითი წილი ორი შტატის მიხედვით, 2008

რასობრივი/ეთნიკური ჯგუფი	მეინი	ჰავაი
თეთრკანიანი	96.4	29.7
ლათინო-ამერიკელი	1.0	8.7
აზიელი	1.0	39.3
ადგილობრივი ჰავაელი ან წყნარ-ოკეანელი	–	9.0
სხვა*	1.6	13.3
სულ	100.0	100.0

* კატეგორია „სხვა“ განიხილება როგორც რასობრივი/ეთნიკური ჯგუფი.

წყარო: U.S. Census Bureau, The 2010 *Statistical Abstract*, Table 19

IQV-ის გაანგარიშების საფეხურები

IQV-ის გაანგარიშებისათვის ვიყენებთ ფორმულას:

$$IQV = \frac{K(100^2 - \sum Pct^2)}{100^2(K-1)} \qquad (5.1)$$

სადაც

K = კატეგორიების რაოდენობა

$\sum Pct^2$ = კვადრატში აყვანილი ხვედრითი წილების ჯამი

ცხრილში 5.2 მოტანილია კვადრატში აყვანილი მატების ტემპები მეინისა და ჰავაის შტატში მცხოვრები ყველა რასობრივი/ეთნიკური ჯგუფისთვის.

ნაბაზი 5.2 კვადრატში აყვანილი მატების ტემპები ორ შტატში ხუთი რასობრივი/ეთნიკური ჯგუფისთვის

რასობრივი / ეთნიკური ჯგუფი	მეინი		ჰავაი	
	%	$(\%)^2$	%	$(\%)^2$
თეთრკანიანი	96.4	9,292.96	29.7	882.09
ლათინო-ამერიკელი	1.0	1.0	8.7	75.69
აზიელი	1.0	1.0	39.3	1,544.49
ადგილობრივი ჰავაელი ან წყნარ-ოკეანელი	–	–	9.0	81.0
სხვა*	1.6	2.56	13.3	176.89
სულ	100.0	9,297.52	100.0	2,760.16

* კატეგორია „სხვა“ განიხილება როგორც რასობრივი/ეთნიკური ჯგუფი.

მეინის შტატისთვის IQV ტოლია:

$$IQV = \frac{K(100^2 - \sum Pct^2)}{100^2(K-1)} = \frac{4(100^2 - 9,297.52)}{100^2(5-1)} = \frac{2,809.92}{30,000} = 0.09$$

ჰავაის შტატისთვის IQV ტოლია:

$$IQV = \frac{K(100^2 - \sum Pct^2)}{100^2(K-1)} = \frac{5(100^2 - 2,760.16)}{100^2(5-1)} = \frac{36,199.20}{40,000} = 0.90$$

შენიშნავდით, რომ გაანგარიშებებმა დაადასტურა ჩვენი წინა მსჯელობები ჰავაიში, სადაც IQV ტოლია 0.90-ის, უფრო მეტი რასობრივი/ეთნიკური ვარიაციაა, ვიდრე მეინის შტატში, სადაც IQV ტოლია 0.09-ის.

ხაზგასასმელია, რომ IQV ნაწილობრივ კატეგორიების რაოდენობის ფუნქციაა. ამ მაგალითში კატეგორიების რაოდენობა არის 4 მეინში და 5 ჰავაიზე. რაც უფრო იზრდება კატეგორიების რიცხვი, მით უფრო მაღალია IQV.

შევაჯამოთ IQV-ის გაანგარიშების საფეხურები:

1. ავაგოთ პროცენტული განაწილება
2. ავიყვანოთ კვადრატში თითოეული კატეგორიის ხვედრითი წილი
3. შევაჯამოთ კვადრატში აყვანილი სიდიდეები
4. გავიანგარიშოთ IQV შემდეგი ფორმულის გამოყენებით:

$$IQV = \frac{K(100^2 - \sum Pct^2)}{100^2(K-1)}$$

დეტალების განხილვა 5.1
სტატისტიკა პრაქტიკაში: მრავალფეროვნება ბერკლის უნივერსიტეტში წლების განმავლობაში[1]

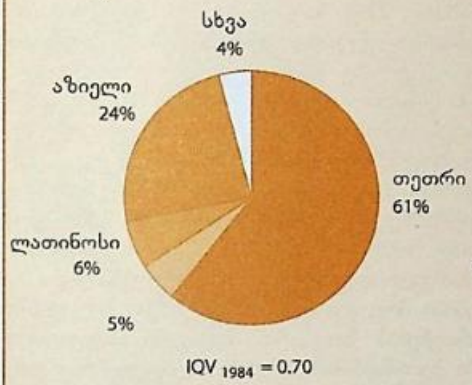
„ბერკლი, კალიფორნია, ლიდერების შეკრება სპროულ ჰოლში – 1960-იანი წლების დასაწყისის ფოტოზე ადამიანები ბერმუდის შორტებსა და სიმბოლოიან სვიტერებში უძველესი ცივილიზაციის არტეფაქტებს მოგვაგონებენ. ისინი სულ ახალგაზრდები არიან... წარმოუდგენელია, მაგრამ ყველა თეთრკანიანებია“.[2]

დებატები კამპუსებში პოზიტიური დისკრიმინაციის შესახებ დიდი ხანია მოძველდა. ბერკლის კამპუსი ერთ-ერთი პირველთაგანი იყო, რომელიც მიესალმა პოზიტიურ დისკრიმინაციას ჩარიცხვების დროს და დღეს მას აშშ-ში ერთ-ერთი ყველაზე მრავალფეროვანი საუნივერსიტეტო ქალაქი აქვს.

ქვემოთ მოტანილ დიაგრამებზე ასახულია ბერკლის სტუდენტების რასობრივი და ეთნიკური ხვედრითი წილები 1984 და 2009 წლებში. ამ პერიოდში ბერკლის კამპუსი საგრძნობლად გამრავალფეროვანდა.

ბერკლის უნივერსიტეტის სტუდენტების რასობრივი/ეთნიკური შემადგენლობა

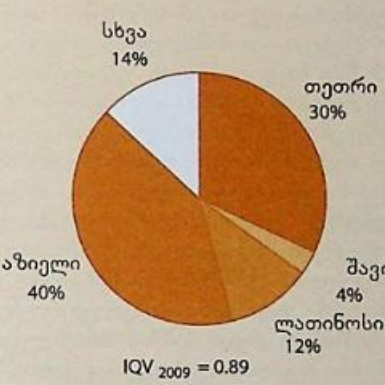
ა)



IQV $_{1984}$ = 0.70

1984: ძირითადად თეთრკანიანები

ბ)



IQV $_{2009}$ = 0.89

2009: მრავალფეროვანი შემადგენლობა

წყარო: * The New York Times, June 4. © 1995 The New York Times.
† 2009 მონაცემები მოტანილია Office of Planning and Analysis, University of California, Berkeley, UC BerkeleyEnrollment Data-დან.
‡The Office of Planning and Analysis, University of California, Berkeley. Cal Stats 2007.

IQV-ს ასახვა პროცენტულად

IQV კოეფიციენტების გარდა შეიძლება პროცენტებშიც ავსახოთ. ამისათვის კოეფიციენტი 100-ზე უნდა გავამრავლოთ. პროცენტებში ასახული IQV გვიჩვენებს რასობრივი/ეთნიკური მრავალფეროვნების ხვედრით წილს მაქსიმალურად შესაძლებელ განსხვავებებში. ამიტომ, როდესაც IQV 0.09-ის ტოლია, ეს ნიშნავს, რომ მეინში რასობრივი განსხვავებები მაქსიმალურად შესაძლებელი განსხვავებების 9.0%-ია (0.09×100). ჰავაისათვის IQV როდესაც 0.90-ის ტოლია, ეს ნიშნავს, რომ აქ რასობრივი/ეთნიკური განსახვავებების რიცხვი მაქსიმალურად შესაძლებლის 90%-ს შეადგენს.

დაკვირვებით შეისწავლეთ დეტალების განხილვა 5.1 და განიხილეთ კატეგორიების რაოდენობის გავლენა IQV-ზე. რა მოხდება, თუ ბერკლის მაგალითში აზიელებს ორ კატეგორიად დავყოფთ: 19% ჩინელი ამერიკელები და 21% – დანარჩენი აზიელები. (ამ კითხვაზე პასუხის გასაცემად თავიდან უნდა გაკეთოთ გაანგარიშებები ამ ახალი მონაცემებით).

✓ *ცოდნა შემოწმება*

სტატისტიკა პრაქტიკაში: მრავალფეროვნება აშშ-ის საზოგადოებაში

დემოგრაფიული პროგნოზების თანახმად, 21-ე საუკუნის შუა წლებში აშშ აღარ იქნება დომინანტურად თეთრკანიანთა ქვეყანა. პროგნოზების მიხედვით 2015 წლისათვის აშშ-ის უმცირესობათა ყველაზე დიდი ჯგუფები – აფრიკელი ამერიკელები, აზიელი ამერიკელები და ლათინო-ამერიკელები – 116 მილიონს მიაღწევენ.[3] 1990-იანი წლებიდან უმცირესობების გადაადგილება აშშ-ის კონკრეტულ რეგიონებსა და მეტროპოლიებში ხდება.[4] დემოგრაფები ამას ჯაჭვურ მიგრაციას უწოდებენ: მიგრანტები არსებითად სოციალურ კაპიტალს იყენებენ – კონკრეტულ ცოდნას მიგრაციულ პროცესებზე (ანუ ერთი ადგილიდან წამოსვლა და მეორეში გადასახლება).[5] მაგალითად, აშშ-ში 2008 წლისათვის ლათინო-ამერიკელების განთავსების მიხედვით პირველ ადგილზე ლოს-ანჯელესი იყო.

როგორ ვადარებთ მრავალფეროვნების მოცულობას სხვადასხვა ქალაქებში შტატებსა და რეგიონებში? მრავალფეროვნებას ზოგიერთი ჩვენგანი ინტუიციით გრძნობს. მაგალითად, როდესაც ქალაქში სეირნობისას ან მეზობლად უამრავ განსხვავებულ ადამიანს ვხვდებით სხვადასხვა ჯგუფიდან.[6]

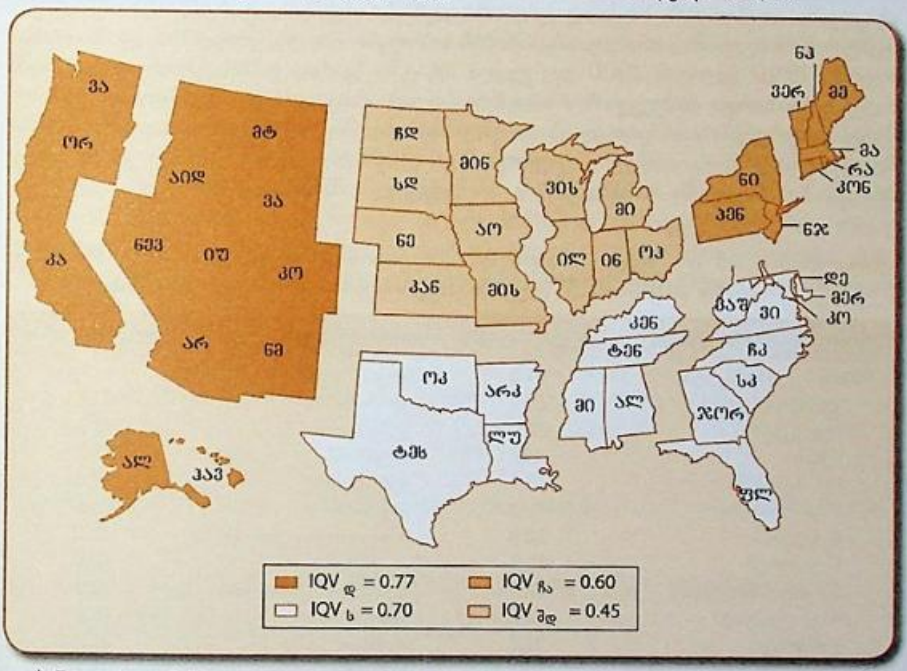
IQV-ს გაანგარიშება შეგვიძლია აგრეთვე სხვადასხვა რეგიონისათვისაც. ცხრილში 5.3 ასახულია მოსახლეობის რასობრივი განაწილება აშშ-ის 4 რეგიონის მიხედვით. ამ მონაცემებზე დაყრდნობით და 5.1 ფორმულის საშუალებით გავიანგარიშეთ IQV თითოეული რეგიონისათვის. ერთი რიცხვით მრავალფეროვნების გამოსახვის დადებითი მხარე ნაჩვენებია ნახაზზე 5.2, რომელიც რეგიონულ განსხვავებებს ასახავს. აღსანიშნავია, რომ დასავლეთი რეგიონი, სადაც IQV ტოლია 0.77-ის, ყველაზე მრავალფეროვანია. ამის საპირისპიროდ, შუა დასავლეთი ყველაზე ჰომოგენური რეგიონია, აქ ძირითადად თეთრკანიანები ცხოვრობენ, ამიტომ IQV 0.45-ის ტოლია.

ცხრილი 5.3 რეგიონებში მოსახლეობის რასობრივი განაწილების ხვედრითი წილები 2008 წელს

რეგიონი	თეთრკანიანი	შავკანიანი	ლათინოსი	აზიელი	სხვა	IQV
ჩრდილო-აღმოსავლეთი	70.4	10.9	11.7	5.1	1.9	0.60
შუა დასავლეთი	79.1	9.9	6.4	2.3	2.3	0.45
სამხრეთი	61.8	18.7	14.8	2.5	2.2	0.70
დასავლეთი	54.5	4.5	27.9	8.8	4.3	0.77

წყარო: U.S. Census Bureau, *American Community Survey*, 2008.

ნახაზი 5.2 რასობრივი მრავალფეროვნება აშშ-ში* 2008 წელს (IQV)



* შტატები აღნიშნულია საფოსტო კოდების მიხედვით

დიაპაზონი

დიაპაზონი ვარიაციის უმარტივესი და პირდაპირი მაჩვენებელია, რომელიც ინტერვალურ-ფარდობითი ცვლადების ვარიაციას ზომავს. ეს არის სხვაობა განაწილების უდიდეს (მაქსიმალურ) და უმცირეს (მინიმალურ) მნიშვნელობებს შორის:

დიაპაზონი = უდიდესი მნიშვნელობა − უმცირესი მნიშვნელობა

აშშ-ის 2008 წლის კვლევაში უხუცესი მონაწილე 89 წლის იყო, ყველაზე ახალგაზრდა კი – 18 წლის. აქედან, დიაპაზონი ტოლია 89 − 18 = 71 წელი.

დიაპაზონი ინტერვალურ-ფარდობითი ცვლადების ვარიაციის საზომი. ეს არის სხვაობა განაწილების უდიდეს (მაქსიმალურ) და უმცირეს (მინიმალურ) მნიშვნელობებს შორის.

შეიძლება გავიანგარიშოთ პროცენტების დიაპაზონიც. მაგალითად, 1980 წლიდან ასაკოვანი ადამიანების საკმაოდ დიდი რაოდენობა ისვენებს არა საპენსიო მექანიზმებით დებულ სან ბელთში[7], არამედ არკანზასის ოზარკსა და კოლორადოსა და მონტანის მთებში. 1990 წლიდან 2000-მდე ყველა შტატში (გარდა ვაშინგტონისა, კოლუმბიის ოლქი) გაიზარდა მოხუცების რაოდენობა, თუმცა სხვადასხვა ოდენობით. ბეიბი ბუმერების დაბერებასთან ერთად ეს ტრენდი მომავალშიც გაგრძელდება. ცხრილში 5.4 მოცემულია აშშ-ის აღწერის ბიუროს მიერ 2008-დან 2015 წლამდე გაკეთებული პროგნოზი ასაკოვანი მოსახლეობის პროცენტული მატების შესახებ.[8]

ცხრილი 5.4 65 და ზევით ასაკის მოსახლეობის პროცენტული მატების პროგნოზი 2008-2015 წლებში რეგიონებისა და შტატების მიხედვით

რეგიონი, შტატი, ოლქი	მატების ტემპი	რეგიონი, შტატი, ოლქი	მატების ტემპი
ჩრდილო-აღმოსავლეთი	16.0	სამხრეთი (გაგრძ.)	
კონექტიკუტი	20.9	ჯორჯია	21.1
დელავერი	22.3	კენტუკი	12.5
მეინი	25.6	ლუიზიანა	23.0
მასაჩუსეტსი	17.7	მერილენდი	23.1
ნიუ-ჰემპშირი	28.4	მისისიპი	16.4
ნიუ-ჯერსი	20.4	ჩრდილოეთი კაროლინა	20.7
ნიუ-იორკი	12.8	ოკლაჰომა	12.8
პენსილვანია	12.5	სამხრეთი კაროლინა	22.1
როდ-აილენდი	18.2	ტენესი	18.2
ვერმონტი	31.4	ტეხასი	25.9
		ვირჯინია	26.8
შუა დასავლეთი	14.0	ვაშინგტონი, კ.ო.	–14.1
ინდიანა	11.3	დასავლეთი ვირჯინია	15.4
ილინოისი	12.9		
აიოვა	11.2	დასავლეთი	27.0
კანზასი	14.2	ალიასკა	50.0
მიჩიგანი	15.5	არიზონა	36.8
მინესოტა	19.2	კალიფორნია	27.1
მისური	14.5	კოლორადო	22.7
ნებრასკა	12.4	ჰავაი	18.8
ჩრდილოეთი დაკოტა	12.6	აიდაჰო	20.2
ოჰაიო	12.4	მონტანა	27.0
სამხრეთი დაკოტა	9.4	ნევადა	42.1
ვისკონსინი	17.6	ნიუ-მექსიკო	31.9
		ორეგონი	17.1
სამხრეთი	22.8	იუტა	13.8
ალაბამა	15.1	ვაშინგტონი	23.2
არკანზასი	14.7	ვაიომინგი	36.9
ფლორიდა	29.7		

წყარო: U.S. Census Bureau, The 2010 *Statistical Abstract*, Tables 16 and 18.

რას უდრის ასაკოვანი მოსახლეობის მატების დიაპაზონი? მოვძებნოთ ყველაზე მაღალი მაჩვენებელი, ეს არის ალასკა 50%-ით და ყველაზე დაბალი მაჩვენებელი, ვაშინგტონი, კოლუმბიის ოლქი (-14.1%). ამიტომ დიაპაზონი 64.1% იქნება, ანუ 50%-იდან -14.1%-მდე.

მიუხედავად იმისა, რომ დიაპაზონი მარტივი და სწრაფი გასაანგარიშებელია, ის მაინც ვარიაციის უხეში საზომია, რადგან მხოლოდ უდიდეს და უმცირეს მნიშვნელობებს ეყრდნობა. ეს ორი მნიშვნელობა შეიძლება ძალიან ექსტრემალური და ატიპური იყოს და შეცდომაში შეგვიყვანოს განაწილების ვარიაციასთან დაკავშირებით. მაგალითად, პროცენტული კლება მხოლოდ ვაშინგტონში, კოლ. ოლქი, შეინიშნება, ხოლო ალასკის მსგავსი მატება მხოლოდ ნევადაშია ნაჩვენები. დიაპაზონმა გვიჩვენა განსხვავება მხოლოდ ვაშინგტონსა (კოლ. ოლქი) და ალასკას, და არა სხვა შტატებს შორის.

რატომ არ ვიყენებთ დიაპაზონს ვარიაციის დასახასიათებლად ნომინალურ ცვლადებთან? დიაპაზონი შეიძლება გამოვიყენოთ ავრეთვე ორდინალურ ცვლადებთან (მაგ. შეგვიძლია ვთქვათ, რომ პასუხების დიაპაზონი იცვლება „ნაწილობრივ კმაყოფილიდან“ „სრულიად უკმაყოფილომდე“), მაგრამ მას რაოდენობრივი მნიშვნელობა არ ექნება. რატომ?

✓ *ცოდნის შემოწმება*

ინტერკვარტილური დიაპაზონი: მატება უფროსი ასაკის მოსახლეობაში

დიაპაზონის შეზღუდულობის დასაძლევად ჩვენ შეგვიძლია ალტერნატული, ინტერკვარტილური დიაპაზონის გამოყენება. **ინტერკვარტილური დიაპაზონი** (IQR) წარმოადგენს ინტერვალური ცვლადების ვარიაციის საზომს და ასახავს შუა 50%-ის საზღვრებს (სიგანეს). ის გაიანგარიშება როგორც ქვედა და ზედა კვარტილების სხვაობა (Q_1 და Q_3).

$$IQR=Q_3-Q_1$$

გაიხსენეთ, პირველი კვარტილი (Q_1) არის 25-ე პერცენტილი. წერტილი, რომლის ქვევით ხვდება დაკვირვებების 25%, ხოლო ზევით — 75%. მესამე კვარტილი (Q_3) არის 75-ე პერცენტილი, წერტილი, რომლის ქვევით ხვდება დაკვირვებების 75%, ხოლო ზევით — 25%. ამიტომ IQR გვიჩვენებს შუაში მდებარე დაკვირვებების 50%-ის ვარიაციას.

ინტერკვარტილური დიაპაზონი (IQR) გვიჩვენებს განაწილების შუა 50%-ის საზღვრებს (სიგანეს). ის გაიანგარიშება როგორც ქვედა და ზედა კვარტილების სხვაობა (Q_1 და Q_3).

დიაპაზონის მსგავსად, IQR მხოლოდ ორ დაკვირვებას ეყრდნობა, მაგრამ რადგან შუალედურ დაკვირვებებს ითვალისწინებს და არა ექსტრემალურს, გარკვეულად დაზღვეულია არასტაბილურობისგან, რაც დიაპაზონს ახასიათებს.

IQR-ის გაანგარიშების საფეხურებია:

1. Q_1-ისა და Q_3-ის საპოვნელად დაალაგეთ დაკვირვებები უმაღლესიდან უმცირესი მნიშვნელობებისკენ, ან პირიქით. ცხრილში 5.5 დალაგებულია ცხრილი 5.4-ის მონაცემები ალასკიდან (50.0%) ვაშინგტონამდე (კოლუმბიის ოლქი) (–14.1%).

2. შემდეგ უნდა განვსაზღვროთ პირველი კვარტილი Q_1, ანუ 25-ე პერცენტილი. ასაკოვანთა პროცენტული მატება ისე უნდა განვსაზღვროთ აშშ-ის შტატების მიხედვით, რომ შტატების 25% მოხვდეს ქვემოთ, ხოლო 75% – ზემოთ. ამისათვის, N-ს ვამრავლებთ 0.25-ზე:

$$(N)(0.25)=(51)(0.25)=12.75$$

პირველი კვარტილი მოხვდა მე-12 და მე-13 შტატებს შორის. ქვემოდან თუ გადავთვლით, მე-12 შტატი ილინოისია, რომელსაც უკავშირდება პროცენტული მატება 12.9. მეცამეტე შტატია იუტა, რომლის მატება 13.8-ია. პირველი კვარტილის განსაზღვრისათვის ვანგარიშობთ 12.9-ისა და 13.8-ის საშუალოს. ამიტომ, (12.9+13.8)/2=13.35 არის პირველი კვარტილი Q_1.

3. Q_3-ის განსაზღვრის მიზნით, ვეძებთ შტატს, რომელიც განაწილებას ისე გაყოფს, რომ მის ქვევით მოხვდეს დაკვირვებების 75%, ხოლო ზევით – 25%. ამ შემთხვევაში N-ს ვამრავლებთ 0.75-ზე:

$$(N)(0.75)=(51)(0.75)=38.25$$

მესამე კვარტილი ხვდება 38-ე და 39-ე შტატებს შორის. ქვემოდან თუ გადავთვლით, 38-ე შტატი არის ვაშინგტონი და მისი მატების პროცენტია 23.2. 39-ე შტატია მეინი, მატებით – 25.6. მესამე კვარტილის საპოვნელად, უნდა გავიანგარიშოთ 23.2-ისა და 25.6-ის საშუალო. ამიტომ, (23.2+25.6)/2=24.4 არის მესამე კვარტილი Q_3.

4. ახლა უკვე შეგვიძლია IQR-ის პოვნა:

$$IQR=Q_3-Q_1=24.4-13.35=11.05$$

ასაკოვანი მოსახლეობის IQR არის 11.05 პროცენტული პუნქტი.

აღსანიშნავია, რომ IQR უკეთეს ინფორმაციას იძლევა, ვინემ დიაპაზონი. ის გვაძლევს 64.1 პუნქტიან გაქანებას, 50%-იდან –14.1%-მდე. IQR გვიჩვენებს, რომ შტატების ნახევარი შეკუმშულია 24.4-სა და 13.35-ს შორის და ეს გაქანება

(დიაპაზონი) გაცილებით ვიწროა. ექსტრემალური კიდურა სიდიდეები გავლენას ვერ ახდენენ IQR-ზე. განსხვავება დიაპაზონსა და ინტერკვარტალურ დიაპაზონს შორის ნაჩვენებია ნახაზზე 5.3, როდესაც 10 ბავშვის შემთხვევაში ექსტრემალური სიდიდე გავლენას ახდენს დიაპაზონზე, მაგრამ არა IQR-ზე.

რატომ არის IQR ვარიაციის უკეთესი საზომი, ვიდრე დიაპაზონი, განსაკუთრებით მაშინ, როდესაც განაწილებაში ექსტრემალური მნიშვნელობებია? ამ კითხვაზე პასუხის გასაცემად იხილეთ ნახაზი 5.3.

✓ ცოდნი შემოწმება

ნახაზი 5.3 დიაპაზონი ინტერკვარტილური დიაპაზონის წინააღმდეგ: ბავშვების რაოდენობა ქალების ორ ჯგუფში

	ჯგუფი 1	ჯგუფი 2	
ბავშვების რაოდენობა	ნაკლებად ცვალებადი	მეტად ცვალებადი	
0	1	1	
Q_1 1	3	3	Q_1
2	3	2	
Q_3 3	3	2	
4			
5			
6		1	Q_3
7			
8		2	
9			
10	2	1	
	დიაპაზონი = 10	დიაპაზონი = 10	
	ინტერკვარტილური დიაპაზონი = 2	ინტერკვარტილური დიაპაზონი = 5	

მართკუთხა დიაგრამა

მართკუთხა დიაგრამა ვიზუალურად ასახავს დიაპაზონს, ინტერკვარტილურ დიაპაზონს, მედიანას, უმცირეს (მინიმალურ) და უდიდეს (მაქსიმალურ) მნიშვნელობებს. მისი საშუალებით შეგვიძლია ვნახოთ ცენტრი, ვარიაცია და ინტერვალურ-ფარდობითი განაწილების ფორმა.

ნახაზზე 5.4 წარმოდგენილია ასაკოვანი მოსახლეობის მატების ტემპის პროგნოზი 2008-2015 წლებისთვის (ცხრილიდან 5.5). მართკუთხა დიაგრამის ასაგებად გამოვიყენეთ განაწილების უმცირესი და უდიდესი მნიშვნელობები, ზედა და ქვედა კვარტილები და მედიანა. თუ მიყვებით ამ ინსტრუქციებს, ადვილად შეძლებთ მის ხელით დახაზვასაც:

ცხრილი 5.5 ასაკოვანი მოსახლეობის მატების ტემპის პროგნოზი 2008-2015 წლებისათვის შტატების მიხედვით, დალაგებულია კლებადი თანმიმდევრობით

შტატი	მატების ტემპი	შტატი	მატების ტემპი	შტატი	მატების ტემპი
ალასკა	50.0	დელავერი	22.3	ალაბამა	15.1
ნევადა	42.1	სამ. კაროლინა	22.1	არკანზასი	14.7
ვაიომინგი	36.9	ჯორჯია	21.1	მისური	14.5
არიზონა	36.8	კონექტიკუტი	20.9	კანზასი	14.2
ნიუ-მექსიკო	31.9	ჩრ. კაროლინა	20.7	იუტა	13.8
ვერმონტი	31.4	ნიუ-ჯერსი	20.4	ილინოისი	12.9
ფლორიდა	29.7	აიდაჰო	20.2	ნიუ-იორკი	12.8
ნიუ ჰემპშირი	28.4	მინესოტა	19.2	ოკლაჰომა	12.8
კალიფორნია	27.1	როდ აილენდი	18.2	კენტუკი	12.5
მონტანა	27.0	ჰავაი	18.8	ჩრდ. დაკოტა	12.6
ვირჯინია	26.8	ტენესი	18.2	პენსილვანია	12.5
ტეხასი	25.9	მასაჩუსეტსი	17.7	ნებრასკა	12.4
მეინი	25.6	ვისკონსინი	17.6	ოჰაიო	12.4
ვაშინგტონი	23.2	ორეგონი	17.1	ინდიანა	11.3
მერილენდი	23.1	მისისიპი	16.4	აიოვა	11.2
ლუიზიანა	23.0	მიჩიგანი	15.5	სამხ. დაკოტა	9.4
კოლორადო	22.7	დას. ვირჯინია	15.4	ვაშინგტ., კ.	–14.1

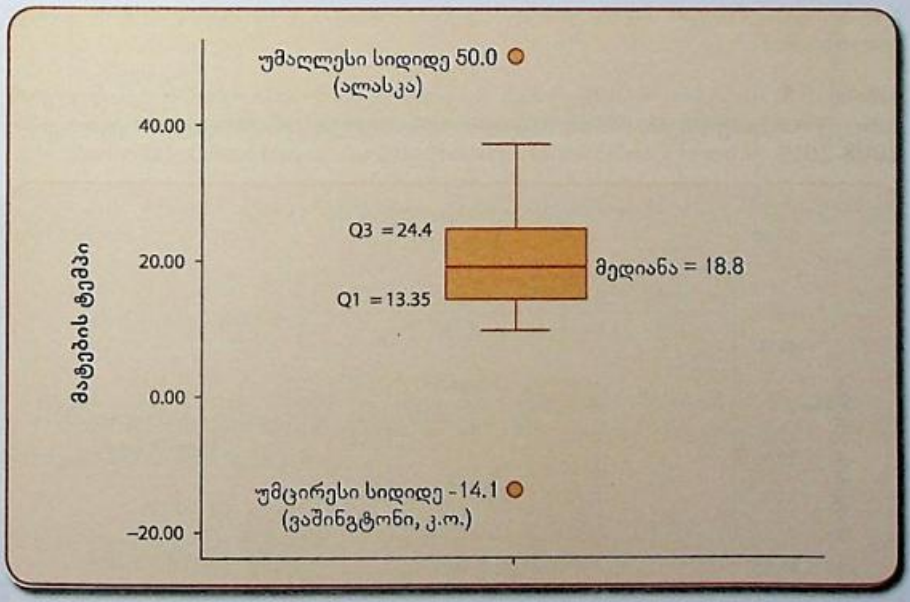
წყარო: U.S. Census Bureau, The 2010 *Statistical Abstract*, Tables 16 and 18.

1. დახაზეთ მართკუთხედი უმცირეს და უდიდეს კვარტილებს შორის.
2. მართკუთხედის შიგნით ჩახაზეთ უწყვეტი ხაზი მედიანის აღსანიშნავად.
3. მართკუთხედის გარეთ გაავლეთ ვერტიკალური ხაზები უმცირეს და უდიდეს მნიშვნელობებამდე.

რის დანახვა შეიძლება ამ დიაგრამიდან? პირველი, მთლიანი ხაზი ნათლად გვიჩვენებს მედიანას. მეორე, რადგან მართკუთხედი მოთავსებულია უმცირეს და

უდიდეს მნიშვნელობებს შორის, IQR ასახულია მართკუთხედის სიგრძეში. ასევე, მართკუთხედის ორივე მხრიდან გამოსული ვერტიკალური ხაზები წარმოადგენენ განაწილების დიაპაზონს.[9] როგორც დიაპაზონი, ისე ინტერკვარტილური დიაპაზონიც გვაძლევს განაწილების გაქანების ვიზუალური აღქმის საშუალებას. და ბოლოს, მართკუთხედის შეფარდებითი მდებარეობა და მასში მედიანის მდებარეობა საშუალებას იძლევა დავინახოთ განაწილების სიმეტრიულობა ან გადახრილობა. სიმეტრიული განაწილების დროს მართკუთხედი დიაპაზონის შუაში იქნება მოთავსებული, ხოლო მედიანა — მართკუთხედის შუაში. როდესაც ირღვევა განაწილების სიმეტრია, მაშინ მართკუთხედი, ისევე როგორც მედიანა, ცენტრში აღარ იქნება მოთავსებული; თუ განაწილებაში დაბალი მნიშვნელობები სჭარბობს, მაშინ მართკუთხედი ქვედა კვარტილთან იქნება ახლოს, ხოლო, თუ მაღალი მნიშვნელობები სჭარბობს — ზედა კვარტილთან.

ნახაზი 5.4 ასაკოვანი მოსახლეობის მატების ტემპის საპროგნოზო განაწილების მართკუთხა დიაგრამა, 2008-2015



ნახაზზე 5.4 მართკუთხა დიაგრამაზე გამოსახული განაწილება სიმეტრიულია, თუ გადახრილი?

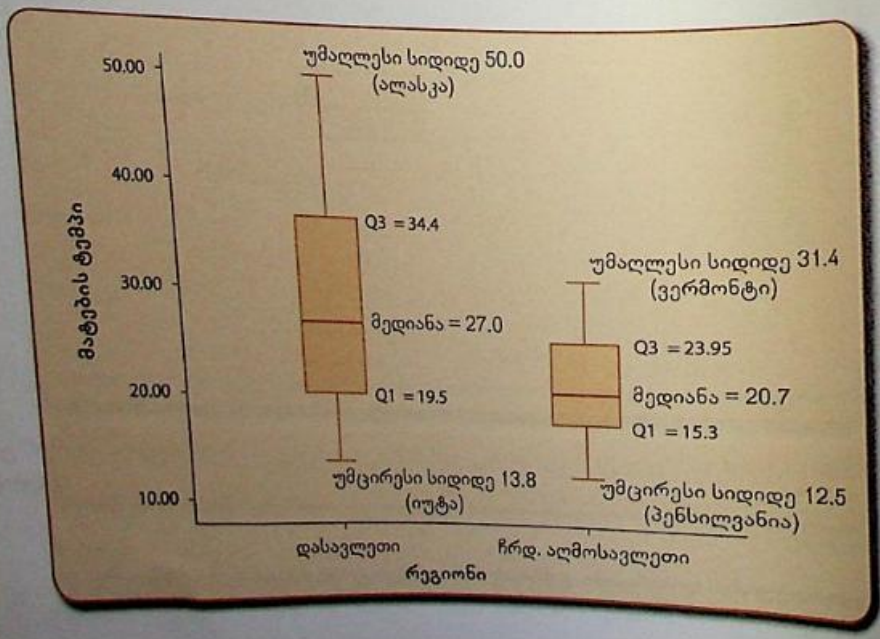
✓ *ცოდნ*
შემოწმება

მართკუთხა დიაგრამები განსაკუთრებით გამოგვადგება განაწილებების შედარების დროს. აღნიშნულის თვალსაჩინოებისთვის განვიხილოთ ნახაზზე 5.5 წარ-

მოდგენილი ცხრილი 5.5-ის მონაცემები ასაკოვანი მოსახლეობის მატების ტემპების შესახებ აშშ-ის ჩრდილო-აღმოსავლეთ და დასავლეთ რეგიონებში.

როგორც ხედავთ, მართკუთხა დიაგრამები ერთმანეთისგან მნიშვნელოვნად განსხვავდება. რა დასკვნები შეიძლება გამოვიტანოთ აქედან? პირველი, მედიანების პოზიციები ხაზს უსვამს ასაკოვანი მოსახლეობის მნიშვნელოვან მატებას დასავლეთ ნაწილში. თუ ჩრდილოეთ ნაწილში (მედიანა = 20.7%) პროგნოზით მოსალოდნელია ასაკოვანი მოსახლეობის ზრდის ტენდენცია, დასავლეთში ეს ტენდენცია უფრო მყარი ჩანს (მედიანა = 27%). მეორე, დიაპაზონიც (მართკუთხედიდან გამოსული ხაზები) და ინტერკვარტილური დიაპაზონიც (მართკუთხედის სიგრძე) უფრო განიერია დასავლეთისკენ (დიაპაზონი = 36.2%; IQR = 13.97%) ვიდრე ჩრდილოეთისკენ (დიაპაზონი = 18.9%; IQR = 8.8%), რაც გვიჩვენებს, რომ დასავლეთით უფრო დიდია ვარიაცია, ვიდრე ჩრდილოეთით. და ბოლოს, მართკუთხედების შეფარდებითი პოზიცია გვიჩვენებს განსხვავებას ამ განაწილებების ფორმებს შორის. რადგან მართკუთხედი თითქმის დიაპაზონის შუაში მდებარეობს, ჩრდილოეთის განაწილება სიმეტრიულია; ხოლო დასავლეთის — პოზიტიურად გადახრილია, რადგან მართკუთხედი უფრო ახლოს არის განაწილების დაბალ ბოლოსთან.

ნახაზი 5.5 ასაკოვანი მოსახლეობის მატების ტემპის საპროგნოზო განაწილების მართკუთხა დიაგრამები, აშშ-ის ჩრდილო-აღმოსავლეთ და დასავლეთ რეგიონებში, 2008-2015



დისპერსია და სტანდარტული გადახრა: ცვლილებები ასაკოვან მოსახლეობაში

1900 წელთან შედარებით აშშ-ში ასაკოვანი მოსახლეობა 10-ჯერ გაიზარდა და პროგნოზების თანახმად, კიდევ გაიზრდება. ამ დემოგრაფიული ცვლილებების ტემპები და მიმართულება სოციალური, ეკონომიკური და ეთიკური არჩევანის წინაშე დააყენებს ცალკეულ ადამიანებს, ოჯახებსა და მთავრობას.

ცხრილში 5.6 წარმოდგენილია ასაკოვანი მოსახლეობის მატების საპროგნოზო ტემპები აშშ-ის ყველა რეგიონის მიხედვით 2008-2015 წლებში.

ნახაზი 5.6 ასაკოვანი მოსახლეობის მატების საპროგნოზო ტემპები აშშ-ის ყველა რეგიონის მიხედვით, 2008-2015

რეგიონი	მატების ტემპი
ჩრდილო-აღმოსავლეთი	16.0
სამხრეთი	22.8
შუა დასავლეთი	14.0
დასავლეთი	27.0
საშუალო ($\bar{Y}$)	19.95

წყარო: U.S. Census Bureau, The 2010 *Statistical Abstract*, Tables 16 and 18

მატების ტემპები გაანგარიშებულია აშშ-ის აღწერის ბიუროს მიერ შემდეგი ფორმულით:

$$\text{მატების ტემპი} = \left[\frac{(\text{2015 მოსახლეობა} - \text{2008 მოსახლეობა})}{\text{2008 მოსახლეობა}}\right] \times 100$$

მაგალითად, 1990 წელს დასავლეთ რეგიონში ასაკოვანი მოსახლეობა 5,773,363 იყო, ხოლო 2000 წელს 6,922,129. ამიტომ მატების ტემპი 1990 წლიდან 2000-მდე შეადგენს:

$$\text{მატების ტემპი} = \left[\frac{(6,922,129 - 5,773,363)}{5,773,363}\right] \times 100 = 19.9$$

ცხრილიდან 5.6 ჩანს, რომ პროგნოზის თანახმად, ასაკოვანი მოსახლეობა 2008-2015 წლებში საშუალოდ 19.95%-ით გაიზრდება. მაგრამ ეს საშუალო სიდიდე არაფერს გვეუბნება რეგიონული ვარიაციის შესახებ. მაგალითად, იქნება თუ არა ჩრდილო-აღმოსავლეთ შტატებში საშუალოზე უფრო ნაკლები მატება იმის გამო, რომ ასაკოვანი მოსახლეობა თბილ-კლიმატიან შტატებს არჩევს? იქნება თუ არა სამხრეთში პროგნოზთან შედარებით ასაკოვანთა მაღალი მატების ტემპი იმიგრაციასთან დაკავშირებით?

მიუხედავად იმისა, რომ მნიშვნელოვანია ქვეყნის მასშტაბით მატების
როგორც ტემპის ცოდნა, ასევე საინტერესოა, განსხვავდება თუ არა რეგიონ
მატებები ეროვნული საშუალოსაგან. თუ რეგიონული საშუალოები ეროვ
საშუალოს უახლოვდება, მაშინ ციფრები ეროვნული საშუალოს გარშემო
ჯგუფდება, ხოლო თუ მათ შორის დიდი განსხვავებაა, მაშინ რეგიონული
დიდები ეროვნული საშუალოს გარშემო ფართოდ გაიბნევა.

ცხრილიდან 5.6 ჩანს, რომ რეგიონებში მნიშვნელოვანი ვარიაცია შეიმჩნ
პროცენტები იცვლება 27%-დან (დასავლეთი), 14.0%-მდე (შუა დასავლეთი),
ტომ დიაპაზონი 13.0%-ია (27.0% − 14.0% = 13.0%). უფრო მეტიც, შტატე
უმრავლესობაში პროგნოზით გათვალისწინებულია მნიშვნელოვანი გადახრ
ეროვნული საშუალოდან. რამდენად დიდია საშუალოდ ეს გადახრები? ჩ
გვჭირდება ისეთი საზომი, რომელიც ყველა შტატისთვის ერთობლივად გაზომ
გადახრებს საშუალოდან. დიაპაზონისა და IQR-ის შემთხვევაში მხოლოდ
მნიშველობას ვეყრდნობოდით.

ასეთი ტიპის საზომმა საშუალოდ უნდა ასახოს თითოეული დაკვირვებ
გადახრა ისეთი ცენტრალური წერტილიდან, როგორიცაა საშუალო არითმეტიკ
ლი. ამ შემთხვევაში ჩვენ საშუალო არითმეტიკულს ვეყრდნობით და არა მოდა
და მედიანას, რადგან საშუალო არითმეტიკულში ყველა დაკვირვებაა გათვა
ისწინებული. დამოკიდებულება ექსტრემალურ სიდიდეებზე ვრცელდება აგრეთ
საშუალო გადახრებზეც, რადგან მისი გაანგარიშება საშუალოთი ხდება. ათვლ
წერტილად არითმეტიკული საშუალოს გამოყენება განპირობებულია მისი ალგ
ბრული მახასიათებლების შესაბამისობით ვარიაციის უფრო რთულ საზომებთან
დისპერსია და **სტანდარტული გადახრა** ერთმანეთთან მჭიდროდ დაკავშირ
ბული ვარიაციის საზომებია, რომელთა სიდიდე დამოკიდებულია საშუალოს ირ
გვლივ მნიშვნელობების დაჯგუფების სიახლოვეზე. **დისპერსია** წარმოადგენ
განაწილების ცენტრიდან (საშუალოდან) გადახრების კვადრატების საშუალო
ხოლო **სტანდარტული გადახრა** მისგან ამოღებული კვადრატული ფესვია. ორივ
ზომავს ინტერვალურ-ფარდობითი ცვლადების ვარიაციას.

დისპერსია ვარიაციის საზომი ინტერვალურ-ფარდობითი ცვლადებისთვის; ის საშუალოდან გადახრების კვადრატების საშუალოს წარმოადგენს.

სტანდარტული გადახრა ვარიაციის საზომი ინტერვალურ-ფარდობითი ცვლადებისთვის; ის უდრის კვადრატულ ფესვს დისპერსიიდან.

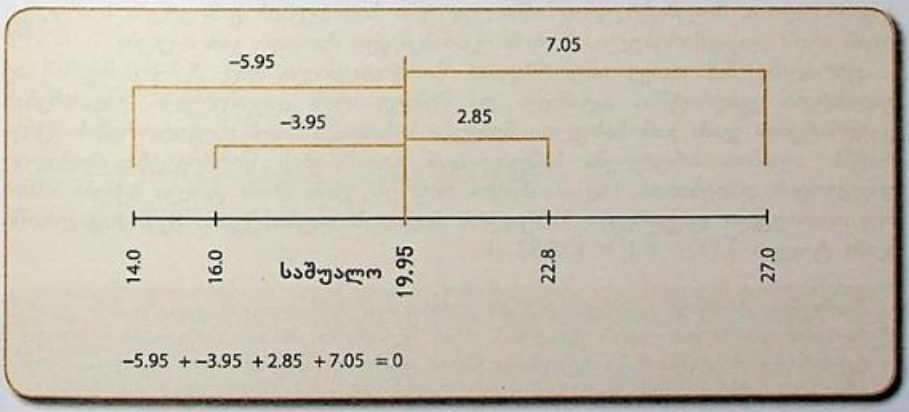
გადახრის გაანგარიშება საშუალოდან

კვლავ განვიხილოთ ასაკოვანი ადამიანების მატების ტემპების განაწილება აშშ-ის ოთხ შტატში. რადგან ჩვენ გვინდა ეროვნული საშუალოდან თითოეული რეგიონის გადახრის საშუალოს გაანგარიშება, ჯერ შევხედოთ განსხვავებებს რეგიონებისა და ეროვნულ საშუალოს შორის. ეს სხვაობა, რომელსაც საშუა-

ლოდან გადახრას უწოდებენ, აღინიშნება შემდეგნაირად: $(Y - \overline{Y})$, ხოლო მათი ჯამი $\sum(Y - \overline{Y})$.

გაანგარიშებები თითოეული რეგიონისთვის მოცემულია ცხრილში 5.7 და ნახაზზე 5.6. ჩვენ ავრეთვე შევაჯამეთ ეს გადახრები. აღსანიშნავია, რომ თითოეული რეგიონისთვის გადახრა შეიძლება იყოს ან დადებითი, ან უარყოფითი. გადახრა დადებითია, როდესაც მატების ტემპი საშუალოზე მაღალია, ხოლო ნეგატიურია, როდესაც მატების ტემპი საშუალოზე დაბალია. ამიტომ, მაგალითად ჩრდილო-აღმოსავლეთის გადახრა ტოლია −3.95-ის, ნიშნავს, რომ ასაკოვანი მოსახლეობის მატების ტემპი 3.95 პუნქტით დაბალია საშუალოზე.

ნახაზი 5.1 საშუალოდან გადახრების ილუსტრაცია



ცხრილი 5.7 რეგიონების მიხედვით ასაკოვანთა მატების საპროგნოზო ტემპები და საშუალოდან გადახრები, 2008-2015

რეგიონი	მატების ტემპი	$(Y - \overline{Y})$
ჩრდილო-აღმოსავლეთი	16.0	16.0 − 19.95 = −3.95
სამხრეთი	22.8	22.8 − 19.95 = 2.85
შუა დასავლეთი	14.0	14.0 − 19.95 = −5.95
დასავლეთი	27.0	27.0 − 19.95 = 7.05
	$\sum(Y) = 79.8$	$\sum(Y - \overline{Y}) = 0$

$$\overline{Y} = \frac{\sum Y}{N} = \frac{79.8}{4} = 19.95$$

ალბათ გაინტერესებთ, შეგვიძლია თუ არა ამ გადახრების საშუალოს გაანგარიშება უბრალოდ შევკამებითა და გაყოფით. სამწუხაროდ, ეს შეუძლებელია,

რადგან საშუალოდან გადახრების ჯამი ყოველთვის ნულის ტოლია, ალგებრულად $\Sigma(Y-\bar{Y})=0$, სხვა სიტყვებით რომ ვთქვათ, თუ თითოეულ მნიშვნელობას გამოვაკლებთ საშუალოს და შემდეგ ამ გადახრებს შევაჯამებთ, როგორც ეს გავაკეთეთ ცხრილში 5.7, ჯამი 0-ის ტოლი იქნება, რაც თავის მხრივ იქამდე მიგვიყვანს, რომ საშუალო გადახრა (ანუ საშუალო განსხვავება) 0-ის ტოლი იქნება. ეს ყოველთვის ასე მოხდება, რადგან საშუალო განაწილების ცენტრის ზიზიდულობის ცენტრს წარმოადგენს.

მათემატიკურად ამ პრობლემის დაძლევა შესაძლებელია, თუ ყურადღებას არ მივაქცევთ პლუს და მინუს ნიშნებს და გამოვიყენებთ გადახრის აბსოლუტურ სიდიდეებს, ან კვადრატში ავიყვანთ გადახრებს – ეს ნიშნავს, რომ უარყოფითი გადახრების თავიდან ასაცილებლად, თითოეული გადახრა თავის თავზე უნდა გავამრავლოთ. მატემატიკურად, აბსოლუტური სიდიდეების გამოყენება სიძნელეებთან არის დაკავშირებული, ამიტომ უკანასკნელი მეთოდი გამოიყენება.

ცხრილში 5.8 იგივე ინფორმაციაა წარმოდგენილი, რაც 5.7-ში, მაგრამ აქ გადახრები კვადრატშია აყვანილი და შემდეგ არის დაჯამებული. გადახრების კვადრატების ჯამი გამოსახულია შემდეგი სახით $\Sigma(Y-\bar{Y})^2$. გადახრების კვადრატში აყვანით სრულდება საშუალოდან გადახრების წარმოდგენა, რომელიც ყოველთვის დადებითია. (აღსანიშნავია, რომ ეს ჯამი 0-ის ტოლი იქნება მაშინ, თუ თითოეული დაკვირვება საშუალოს სიდიდეს დაემთხვევა). ჩვენ მაგალითში ჯამი ტოლია $\Sigma(Y-\bar{Y})^2 = 108.82$-ის.

ცოდნის შემოწმება

შეამოწმეთ თავიდან ცხრილი 5.8 და განიხილეთ საშუალოდან გადახრების კვადრატების ჯამში დასავლეთ რეგიონის მიერ შეტანილი არაპროპორციული წვლილი (კვადრატების ჯამში მისი ხვედრითი წილი დაახლოებით 45%-ია). შეგიძლიათ თუ არა ამის ახსნა? (მინიშნება: ამას გარკვეული კავშირი აქვს საშუალოს მგრძნობიარობასთან ექსტრემალური სიდიდეების მიმართ).

ცხრილი 5.8 რეგიონების მიხედვით ასაკოვანთა მატების საპროგნოზო ტემპები 2008-2015 წლებში, საშუალოდან გადახრები და საშუალოდან გადახრების კვადრატები

რეგიონი	მატების ტემპი	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$
ჩრდილო-აღმოსავლეთი	16.0	$16.0 - 19.95 = -3.95$	15.60
სამხრეთი	22.8	$22.8 - 19.95 = 2.85$	8.12
შუა დასავლეთი	14.0	$14.0 - 19.95 = -5.95$	35.40
დასავლეთი	27.0	$27.0 - 19.95 = 7.05$	49.70
	$\Sigma(Y) = 79.8$	$\Sigma(Y-\bar{Y}) = 0$	$\Sigma(Y-\bar{Y})^2 = 108.82$

საშუალო $= \bar{Y} = \frac{\sum Y}{N} = \frac{79.8}{4} = 19.95$

დისპერსიისა და სტანდარტული გადახრის გაანგარიშება

კვადრატული გადახრების ჯამს **დისპერსია** ეწოდება. დისპერსია აღინიშნება შემდეგი სახით S_Y^2. დაიმახსოვრეთ, რომ ჩვენი ინტერესის საგანია საშუალოდან გადახრების კვადრატების საშუალოს გაანგარიშება. ამიტომ საშუალოდან გადახრების კვადრატების ჯამი უნდა გავყოთ დაკვირვებათა რაოდენობაზე (N). თუმცა, საშუალო არითმეტიკულისაგან განსხვავებით, N-ის ნაცვლად ჩვენ გამოვიყენებთ $N - 1$-ს.[10] დისპერსიის ფორმულა ასე გამოისახება:

$$S_Y^2 = \frac{\sum (Y - \overline{Y})^2}{N - 1} \qquad (5.2)$$

სადაც

S_Y^2 = დისპერსია

$(Y - \overline{Y})$ = საშუალოდან გადახრა

$\sum (Y - \overline{Y})^2$ = გადახრების კვადრატების ჯამი

N = დაკვირვებათა რაოდენობა

აღსანიშნავია, რომ ამ ფორმულაში მონაწილეობს უკვე ყველა გამოყენებული სიმბოლო. ფორმულა ასე წაიკითხება: დისპერსია საშუალოდან გადახრების კვადრატების საშუალოს ტოლია.

დისპერსიის გაანგარიშება რამდენიმე საფეხურს აერთიანებს:

1. გაიანგარიშეთ საშუალო, $\overline{Y} = \sum(Y)/N$.
2. გადახრის საპოვნელად თითოეულ დაკვირვების მნიშვნელობას გამოაკელით საშუალო, $Y - \overline{Y}$.
3. თითოეული გადახრა აიყვანეთ კვადრატში, $(Y - \overline{Y})^2$.
4. შეაჯამეთ გადახრების კვადრატები, $\sum (Y - \overline{Y})^2$.
5. მიღებული ჯამი გაყავით $N - 1$-ზე, $\sum (Y - \overline{Y})^2/(N-1)$.
6. გაიანგარიშეთ დისპერსიის სიდიდე.

ცოდნის შემოწმების მიზნით დაუბრუნდით ცხრილს 5.8 და ჩამოთვლილი საფეხურების მიხედვით გაიანგარიშეთ დისპერსია. ახლა ფორმულაში 5.2 ჩავსვათ მნიშვნელობები და გავიანგარიშოთ დისპერსია:

$$S_Y^2 = \frac{\sum (Y - \overline{Y})^2}{N - 1} = \frac{108.82}{3} = 36.27$$

დისპერსიის პრობლემა იმაში მდგომარეობს, რომ გადახრები კვადრატშია აყვანილი და ამის გამო მნიშვნელობები თავდაპირველ საზომ ერთეულებში აღარ არის გამოსახული. მაგალითად, რთულია დისპერსიის მნიშვნელობის 36.27 ინტერპრეტაცია, რომელიც ასაკოვანი მოსახლეობის მატების ტემპს ასახავს, რადგან ის კვადრატულ პროცენტებშია გამოსახული. ამიტომ ფესვი ამოგვაქვს დისპერსიიდან და შემდეგ ვახდენთ ამ სიდიდის ინტერპრეტირებას. ყოველივე აღნიშნულის შედეგად ვიღებთ *სტანდარტულ გადახრას*, S_Y^2-ს.

სტანდარტული გადახრა წარმოადგენს კვადრატულ ფესვს დისპერსიიდან:

$$S_Y = \sqrt{S_Y^2}$$

სტანდარტული გადახრა ჩვენ მაგალითში არის:

$$S_Y = \sqrt{S_Y^2} = \sqrt{36.27} = 6.02$$

სტანდარტული გადახრის ფორმულაში იგივე სიმბოლოები გამოიყენება, რაც დისპერსიაში:

$$S_Y = \sqrt{\frac{\sum (Y - \bar{Y})^2}{N-1}} \qquad (5.3)$$

თუ მოვახდენთ ფორმულის ინტერპრეტაციას, ჩვენ ვიტყვით, რომ სტანდარტული გადახრა ტოლია საშუალოდან გადახრების კვადრატების საშუალოს კვადრატული ფესვისა.

სტანდარტული გადახრის დადებითი მხარეა ის, რომ ამოსავალ მონაცემებს იმავე საზომ ერთეულებში გამოსახავს, რაც მათ თავიდან აქვთ. კერძოდ, ჩვენი მაგალითის მიხედვით, სტანდარტული გადახრა 6.02-ს უდრის. რადგან ამოსავალი მონაცემები პროცენტებში იყო გამოსახული, ეს რიცხვიც პროცენტი იქნება. სხვა სიტყვებით რომ ვთქვათ, „სტანდარტული გადახრა 6.02%-ის ტოლია". მაგრამ რას ნიშნავს ეს? თავისთავად ეს რიცხვი ბევრს არაფერს გვეუბნება, თუმცა საშუალოს ირგვლივ დისპერსიის შეფასების საშუალებას გვაძლევს.

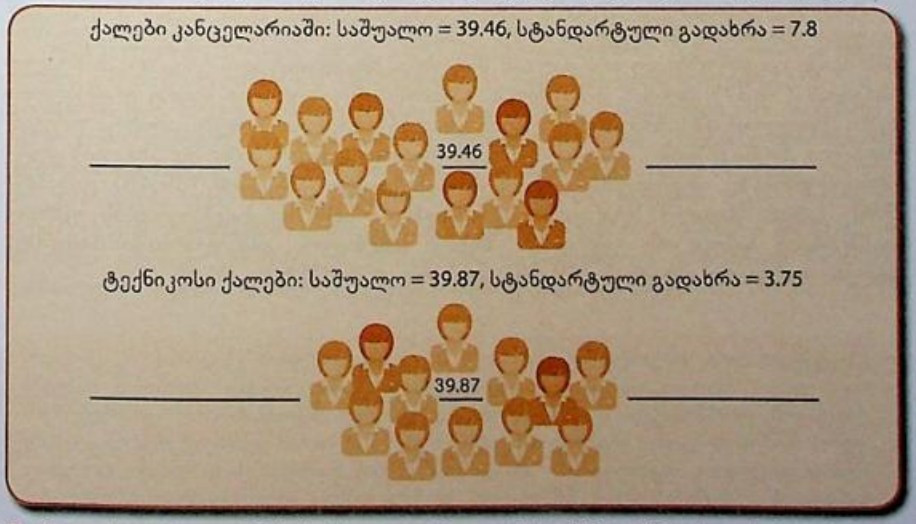
განაწილებაში, სადაც ყველა მნიშვნელობა იდენტურია, სტანდარტული გადახრა ნულის (0) ტოლია. ნული სტანდარტული გადახრის უმცირესი შესაძლებელი მნიშვნელობაა: იდენტურ განაწილებაში, ყველა დაკვირვება ერთმანეთის ტოლია, ტოლია აგრეთვე საშუალო არითმეტიკული, მოდა და მედიანა. ამ დაკვირვებებს არც ვარიაცია გააჩნიათ და არც დისპერსია.

რაც უფრო სცილდება გადახრა ნულს, მით მეტი ვარიაციაა განაწილებაში. სტანდარტული გადახრის სიდიდეს ზედა ზღვარი არ გააჩნია. ჩვენ მაგალითში შეიძლება დავასკვნათ, რომ 6.02%-ის ტოლი სტანდარტული გადახრა ნიშნავს

შემდეგს: აშშ-ის ოთხ რეგიონში ასაკოვანი მოსახლეობის საპროგნოზო პროცენტული ცვლილებები ფართოდ არის გაბნეული საშუალოს, 19.95%-ის ირგვლივ.

სტანდარტული გადახრა შეიძლება მივიჩნიოთ სტანდარტად, რომელსაც შეიძლება შევადაროთ მნიშვნელობების რელატიური განთავსება საშუალოსთან ან ერთმანეთთან. როგორც მე-6 თავში ვნახავთ, უმრავლეს განაწილებებში, თუ ისინი ძალიან არ არის გადახრილი, მნიშვნელობების 34% ხვდება საშუალოსა და საშუალოზე 1 სტანდარტული გადახრით ზედა არეში. დანარჩენი 34% საშუალოსა და საშუალოზე 1 სტანდარტული გადახრით ქვედა არეში. ამიტომ ჩვენ მოველით, რომ დაკვირვებების უმეტესობა (68%) საშუალოდან ერთი სტანდარტული გადახრის არეში მოხვდება. მაგალითად მოვიყვანოთ MTF 2008-ის დაკვირვებიდან რესპონდენტების ქულების შეფასების საშუალო რეიტინგის GPA-ს განაწილება. GPA-ის მნიშვნელობა 4.0-იან სკალზე 3.13-ის ტოლია, ხოლო სტანდარტული გადახრა – 0.68-ის. ჩვენი მოლოდინით, რესპონდენტების GPA მოხვდება საზღვრებში დიაპაზონით 2.45-დან (*3.13 - 0.68*) 3.81-მდე (*3.13 + 0.68*). ამიტომ, საშუალოსა და სტანდარტულ გადახრაზე დაყრდნობით, ჩვენ შეგვიძლია საკმაოდ კარგად წარმოვიდგინოთ, თუ რის ტოლი იქნება ტიპური GPA სტუდენტების უმრავლესობისთვის. მაგალითად, განვიხილოთ სტუდენტი, რომელმაც მიიღო 3.9-ის ტოლი GPA. სხვა სტუდენტებთან შედარებით, ის გამონაკლისი იქნება. ამ სტუდენტისგან განსხვავებით, სტუდენტების ორ მესამედზე მეტი საშუალოს მიდამოებში მოხვდება.

ნახაზი 5.7 საშუალო ასაკისა და სტანდარტული გადახრების ილუსტრაცია



წყარო: ადაპტირებულია: Marjorie Armstrong-Stassen, "The Effect of Gender and Organizational Level on How Survivors Appraiseand Cope With Organizational Downsizing," *Journal of Applied Behavioral Science* 34, no. 2 (June 1998): 125–142. იბეჭდება ნებართვით.

სტანდარტული გადახრის ინტერპრეტირებისათვის მეორე გზას წარმოადგენს მისი შედარება სხვა განაწილებასთან. მაგალითად, ცხრილში 5.9 მოტანილია ჟურნალ ფორჩუნიდან (Fortune) 100 კორპორაციის საფუძველზე შედგენილი მომუშავეთა ორი ერთობლიობის ასაკების სტანდარტული გადახრები. ერთობლიობები დაყოფილია კანცელარიაში და ტექნიკურ სამუშაოებზე დასაქმებულ ქალებად. აღსანიშნავია, რომ ორივე ერთობლიობაში ქალების საშუალო ასაკი დაახლოებით 39 წელია. სტანდარტული გადახრებით ვხვდებით, რომ ასაკობრივი განაწილება ამ ჯგუფებში განსხვავებულია. ეს არაერთგვაროვნება კარგად არის ასახული ნახაზზე 5.7.

ცხრილი 5.9 კანცელარიაში და ტექნიკური საქმიანობით დაკავებული ქალების ასაკი

მახასიათებლები	კანცელარიაში N=22	ტექნიკოსი ქალები N=22
საშუალო ასაკი	39.46	39.87
სტანდარტული გადახრა	7.8	3.75

წყარო: ადაპტირებულია: Marjorie Armstrong-Stassen, "The Effect of Gender and Organizational Level on How Survivors Appraiseand Cope With Organizational Downsizing," *Journal of Applied Behavioral Science* 34, no. 2 (June 1998): 125–142. იხეჭდება ნებართვით.

ტექნიკოსი ქალების ჯგუფში მცირე სტანდარტული გადახრა მიანიშნებს ამ ჯგუფის ასაკობრივ ჰომოგენურობაზე. შეიძლება ითქვას, რომ ქალების უმრავლესობის ასაკი იდენტური თუ არა, ძალიან მსგავსია. სტანდარტული გადახრა საშუალო ასაკიდან (39.87) ტოლია 3.75 წლის. საპირისპირო სურათია მეორე ერთობლიობაში. კანცელარიაში მომუშავე ქალების ასაკების სტანდარტული გადახრა ორჯერ მაღალია ტექნიკოსებზე. ეს მიგვანიშნებს ფართო დისპერსიასა და ასაკების ჰეტეროგენურობაზე. შეგვიძლია ვთქვათ, რომ საშუალო ასაკიდან (39.46) გადახრა 7.80 წელია. მაღალი სტანდარტული გადახრა ნიშნავს დაკვირვებების ფართო დისპერსიას საშუალოს ქვევით და ზევით. საშუალოდ, კანცელარიაში მომუშავე ქალები საშუალო ასაკიდან (39.46) საკმაოდ დაშორებული არიან.

✓ ცოდნის შემოწმება

გარკვეული დრო დაუთმეთ დისპერსიისა და სტანდარტული გადახრის შესწავლას, რაც უფრო რთული მეთოდების გამოყენებისას გამოგვადგება. შეიძლება თქვენს ლექტორმა ფორმულების დამახსოვრება მოგთხოვოთ, მაგრამ მათი ინტერპრეტირება უფრო მნიშვნელოვანია. შეგიძლიათ გამოთვლები ხელით ან კომპიუტერული პროგრამებით გააკეთოთ, მაგრამ ისინი არ აგიხსნიან რას ნიშნავს ეს საზომები. როდესაც ვარიაციის საზომების არსს ჩაწვდებით, ფორმულებსაც ადვილად დაიმახსოვრებთ.

ვარიაციის საზომის შეჩევის საკითხები

ჩვენ უკვე განვიხილეთ ვარიაციის ზუთი საზომი: (1) IQV, (2) დიაპაზონი, (3) IQR, (4) დისპერსია და (5) სტანდარტული გადახრა. თითოეული მათგანი ვარიაციის ხარისხს ასახავს განაწილებაში, მაგრამ რომელი მათგანი უნდა გამოვიყენოთ? ამ კითხვაზე მარტივი პასუხი არ არსებობს. თუმცა, როგორც წესი, ჩვენ ერთ-ერთ მათგანს ვირჩევთ და ამ არჩევანს შესაბამისი განსჯის საფუძველზე ვაკეთებთ. ეს საკითხები და მათი გავლენა ჩვენ არჩევანზე ასახულია გადაწყვეტილებათა ხის სახით ნახაზზე 5.8.

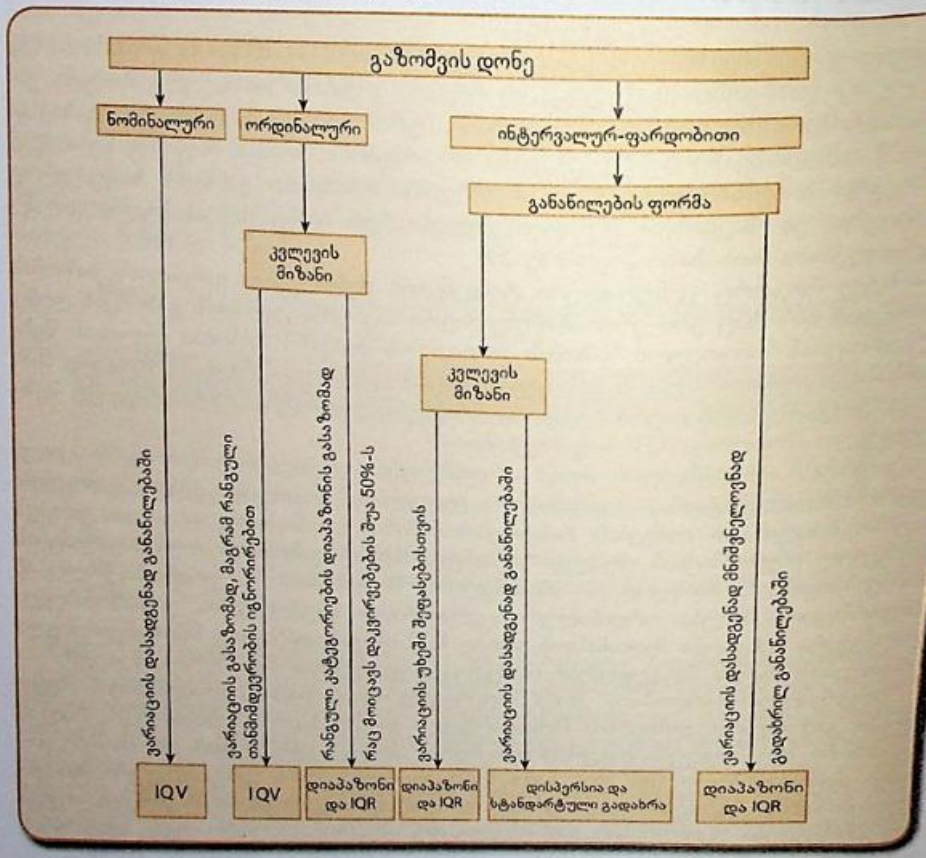
ისევე როგორც ცენტრალური ტენდენციის შემთხვევაში, ვარიაციის საზომის შერჩევის დროსაც ერთ-ერთი მნიშვნელოვანი საკითხია ცვლადის გაზომვის დონე. Y ვარიაციის თითოეული საზომის გამოყენების მთავარი პირობაა ცვლადის შესაბამისი ან უფრო მაღალი გაზომვის დონე, როგორც ეს ნაჩვენებია ნახაზზე 5.8.

გაზომვის ნომინალური დონე: ნომინალური ცვლადებისთვის არჩევანი შეზღუდულია მხოლოდ IQV-ს გამოყენებით.

გაზომვის ორდინალური დონე: ორდინალური ცვლადებისთვის არჩევანი კიდევ უფრო პრობლემატურია. ორდინალურ ცვლადებთან შეიძლება IQV-ს გამოყენება, მაგრამ რადგან ის რანგების რანჟირების მიმართ მგრძნობიარე არ არის, კარგავს გარკვეულ ინფორმაციას. შემდეგი ვარიანტია IQR-ის გამოყენება, თუმცა ვარიაციის ასახვისათვის ის მხოლოდ ორ მნიშვნელობას შორის მანძილს ეყრდნობა, ხოლო ამ ინფორმაციის მიღება ორდინალური ცვლადებიდან შეუძლებელია. კომპრომისული გადაწყვეტა იქნებოდა მედიანასთან ერთად IQR-ის გამოყენება, თუ მას რანჟირებული სიდიდეების რანგს მივაკუთვნებთ, რომელიც დაკვირვებათა შუა 50%-ს მოიცავს.[11]

ინტერვალურ-ფარდობითი დონე: ინტერვალურ-ფარდობითი ცვლადებისთვის შეგიძლიათ გამოიყენოთ დისპერსია (სტანდარტული გადახრა), დიაპაზონი ან IQR. რადგან დიაპაზონი და ნაკლები ინტენსივობით, მაგრამ IQR-იც განაწილების ორ მნიშვნელობაზეა დამოკიდებული (და ამის გამო მგრძნობიარეები არიან ექსტრემალური მნიშვნელობებისადმი), ჩვეულებრივ დისპერსიას და/ან სტანდარტულ გადახრას ირჩევენ. თუმცა თუ განაწილება უკიდურესად გადახრილია, ანუ ისე, რომ საშუალო უკვე აღარ არის ცენტრალური ტენდენციის ამსახველი, შეიძლება გამოიყენოთ დიაპაზონი და IQR. ეს საზომები აგრეთვე გამოგადგებათ ცხრილების წაკითხვასა და დისპერსიის ხარისხზე წარმოდგენის შექმნის მიზნით მონაცემების სწრაფ გადათვალიერებაში.

ნახაზი 5.8 როგორ შევარჩიოთ ვარიაციის საზომი



კვლევითი ლიტერატურის მიმოხილვა: განსხვავებები ლათინო-ამერიკელი მოზარდების მისწრაფებებსა და მოლოდინებს შორის კოლეჯში შესვლამდე

მე-2 თავში აღვნიშნეთ, რომ განაწილებები ხშირად გვხვდება პროფესიულ ლიტერატურაში. ხაზი გავუსვით იმასაც, რომ სოციოლოგიურ ლიტერატურაში გამოყენებული სტატისტიკური ცხრილები სახელმძღვანელოში განხილულზე ბევრად რთულია. იგივე შეიძლება ითქვას ცენტრალური ტენდენციისა და ვარიაციის საზომებზე. სტატიების უმრავლესობაში ცენტრალური ტენდენციისა და

ვარიაციის საზომებს ისე იყენებენ, რომ ის სცილდება ერთი ცვლადის ცენტრალური ტენდენციისა და ვარიაციის აღწერის საზღვრებს. ამ ნაწილში ჩვენ გამოვიყენებთ როგორც საშუალოს, ისე სტანდარტულ გადახრას, რადგან გამოკვლევების უმრავლესობაში სტანდარტული გადახრა საშუალოსთან ერთად არის მოცემული.

ცხრილში 5.10 მონაცემები აღებულია ჟურნალ „Social Problems"[12]-ში გამოქვეყნებული სამეცნიერო სტატიიდან. აქ წარმოდგენილია საშუალოსა და სტანდარტული გადახრის ზოგადი გამოყენება. ავტორები იკვლევენ ეთნიკური ნიშნის ზეგავლენას კოლეჯში შესვლის წინ ახალგაზრდების მისწრაფებებსა და მოლოდინებზე. მათი კვლევა ფოკუსირებულია სამ ყველაზე დიდ ლათინო-ამერიკულ ჯგუფზე „აღნიშნულ საკითხში მათ შორის არსებულ განსხვავებებსა და ამ განსხვავებების გამომწვევ მიზეზებზე".[13] საშუალოსა და სტანდარტული გადახრის გამარტივებული მაგალითის წარმოდგენის მიზნით ჩვენ მხოლოდ კუბელებისა და მექსიკელების მონაცემებზე შევჩერდებით. ეთნიკურობისა და ახალგაზრდების მისწრაფებებისა და მოლოდინების ურთიერთკავშირის შესწავლა ძალიან მნიშვნელოვანია, რადგან ლათინო-ამერიკელების რიცხოვნობა სხვა უმცირესობებთან შედარებით უფრო სწრაფად იზრდება, ხოლო განსხვავებული კანის ფერის მოსახლეობაში ისინი ყველაზე ნაკლებად განათლებულებად რჩებიან.

კვლევის მონაცემები აღებულია მოზარდთა ჯანმრთელობის გრძელვადიანი ეროვნული კვლევიდან, რაც მე-7-დან მე-12 კლასის ჩათვლით ამერიკელი მოზარდების რეპრეზენტატულ კვლევას წარმოადგენს. გათვალისწინებულია ისეთი ფაქტორები, როგორიცაა გეოგრაფიული მდებარეობა, სკოლის ტიპი, რასობრივი შემადგენლობა და ა.შ.

ცხრილი 5.10 ეთნიკურობა და მისწრაფებები და მოლოდინები კოლეჯში შესვლის წინ

	კუბელები		მექსიკელები	
	საშუალო	სტანდარტული გადახრა	საშუალო	სტანდარტული გადახრა
I. რამდენად უნდა რესპონდენტს კოლეჯში შესვლა	4.50	1.80	4.20	1.30
II. როგორია ალბათობა, რომ რესპონდენტი კოლეჯში შევა	4.30	2.00	3.70	1.40

წყარო: ადაპტირებულია Stephanie A. Bohon, Monica Kirkpatrick Johnson, and Bridget K. Gorman, "College Aspirations and Expectations Among Latino Adolescents in the United States,"-დან, *Social Problems* 53, no. 2 (2006): 207–225. Published by theUniversity of California Press.

შენიშვნა: სტატიაში ავტორები ვარიაციას სხვა ეთნიკურ ჯგუფებშიც იხილავენ. ამ მაგალითის გამარტივების მიზნით ჩვენ ყურადღებას გავამახვილებთ კუბელებისა და მექსიკელების აღწერითი სტატისტიკის მონაცემებზე.

რესპონდენტებს უამრავი კითხვა დაუსვეს, თუმცა ავტორებმა ყურადღება გაამახვილეს კონკრეტულად მისწრაფებებსა და მოლოდინებზე. ანალიზის ლტოლვა კოლეჯისაკენ და მოლოდინები შეფასდა 1-დან 5-მდე სკალაზე. კითხვაზე „რამდენად გინდათ სწავლის გაგრძელება კოლეჯში?“, 1-იანი აღნიშნავდა მისწრაფების დაბალ დონეს, ხოლო 5-იანი – მაღალს. მოლოდინებიც მსგავსი სკალით იზომებოდა 1-დან 5-მდე დასმული კითხვის საფუძველზე: “როგორია იმის ალბათობა, რომ თქვენ სწავლას კოლეჯში გააგრძელებთ?“

რა დასკვნა შეიძლება გამოვიტანოთ ამ ცვლადების საშუალოებისა და სტანდარტული გადახრების დათვალიერებით? უპირველესად, შევხედოთ საშუალოებს. მსგავსია ისინი, თუ განსხვავებული? ცხრილიდან ჩანს, რომ მექსიკელებს საშუალოდ უფრო ნაკლები მისწრაფებები (4.20) და მოლოდინები (3.70) აქვთ, ვიდრე კუბელებს (4.50 და 4.30, შესაბამისად). სტანდარტული გადახრა კუბელებთან უფრო მაღალია, ვიდრე მექსიკელებთან.

მკვლევარებმა ცხრილში 5.10 მოტანილი მონაცემები ასე აღწერეს:

ხუთიდან ჯგუფში გამოვლინდა მაღალი მისწრაფება და მოლოდინები კოლეჯში შესვლასთან დაკავშირებით. თუმცა ჯგუფებს შორის მნიშვნელოვანი განსხვავებები შეიმჩნევა. როგორც მოსალოდნელი იყო, მექსიკელების მისწრაფებები და მოლოდინები საშუალოზე უფრო დაბალია... ხოლო კუბელების – უფრო მაღალი.[14]

რატომ მოხდა ასე? მსჯელობის დასაბუთებისთვის ავტორებს შემდეგი არგუმენტები მოყავთ:

განსხვავებული მისწრაფებები და მოლოდინები შეიძლება ავხსნათ განსხვავებებით ოჯახებსა და შინამეურნეობებს შორის, მშობლების მოლოდინებით შვილების განათლებიდან მიღებულ სიკეთეზე, აკადემიური უნარ-ჩვევებითა და გამოხატვის უნარით.[15]

ძირითადი განმარტებები:

• ვარიაციის საზომები – რიცხვები, რომლებიც გამოიყენება ცვლადის განაწილების მრავალფეროვნებისა და ვარიაციის დახასიათებისათვის.

• თვისობრივი ვარიაციის ინდექსი (IQV) – ვარიაციის საზომია ნომინალური ცვლადებისათვის. ის ეფუძნება განაწილებაში არსებული განსხვავებების რაოდენობის ფარდობას მაქსიმალურად შესაძლებელ განსხვავებათა

ინდექსმა შეიძლება მიიღოს 0.00-დან 1.00-მდე მნიშვნელობები.

• დიაპაზონი – ინტერვალურ-ფარდობითი ცვლადების ვარიაციის საზომი. ეს არის სხვაობა განაწილების უდიდეს (მაქსიმალურ) და უმცირეს (მინიმალურ) მნიშვნელობებს შორის. ორდინალური ცვლადების

• ინტერკვარტილური დიაპაზონი (IQR) – ზომავს განაწილების შუა 50%-ის საზღვრებს (სიგანეს). ის გაიანგარიშება როგორც ქვედა და ზედა კვარტილების სხვაობა (Q_1 და Q_3). ორდინალური ცვლადებისთვის მხოლოდ მიუთითეთ ზედა და ქვედა კვარტილები, გამოკლების გარეშე.

• მართკუთხა დიაგრამა – ვიზუალურად ასახავს დიაპაზონს, ინტერკვარტილურ დიაპაზონს, მედიანას, უმცირეს (მინიმალურ) და უდიდეს (მაქსიმალურ) მნიშვნელობებს. მისი საშუალებით შეგვიძლია ვნახოთ ცენტრი, ვარიაცია და ინტერვალურ-ფარდობითი განაწილების ფორმა.

• დისპერსია და სტანდარტული გადახრა – ერთმანეთთან მჭიდროდ დაკავშირებული ვარიაციის საზომებია, რომელთა სიდიდე დამოკიდებულია საშუალოს ირგვლივ მნიშვნელობების დაჯგუფების სიახლოვეზე. დისპერსია წარმოადგენს განაწილების ცენტრიდან (საშუალოდან) გადახრების კვადრატების საშუალოს, ხოლო სტანდარტული გადახრა მისგან ამოღებული კვადრატული ფესვია.

საკვანძო ტერმინები

ინტერკვარტილური დიაპაზონი (IQR)
ინტერკვარტილური დიაპაზონი (IQR)
ვარიაციის საზომები
დიაპაზონი
დისპერსია
სტანდარტული გადახრა

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

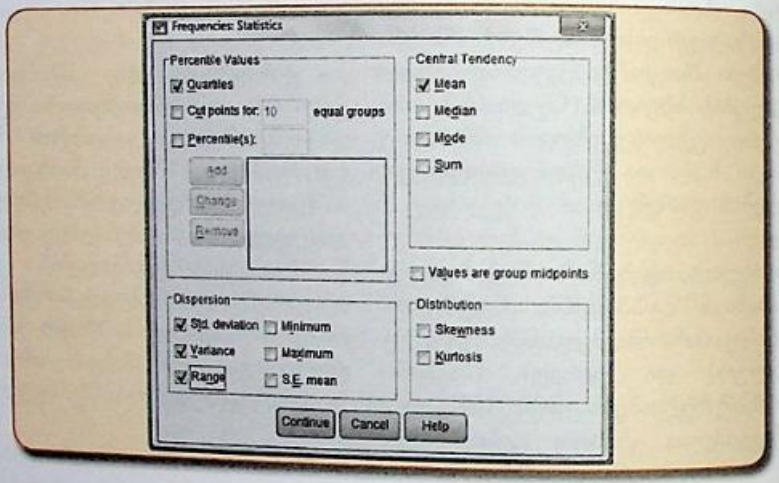
[GSS08PFP-A]

მაგალითი 1: ვარიაციის საზომების გაანგარიშება ბრძანებით Frequencies

IQV-ის გარდა SPSS-ის Frequencies-ის ბრძანებით შესაძლებელია ამ თავში განხილული ვარიაციის ყველა საზომის გაანგარიშება. (IQV-ის გასასაანგარიშებლად შესაძლებელია SPSS-ის პროგრამირება, თუმცა ეს ოპერაციები სცილდება სახელმძღვანელოს ფარგლებს).

ჩვენ დავიწყებთ ბრძანებით Frequencies და გავიანგარიშებთ უამრავ მაჩვენებელს ცვლადისთვის AGE. დაკლიკეთ Analyze, Descriptive Statistics, Frequencies, შემდეგ Statistics ღილაკი და შეარჩიეთ ვარიაციის შესაბამისი საზომი.

ნახაზი 5.9 დიალოგური ფანჯარა Statistics

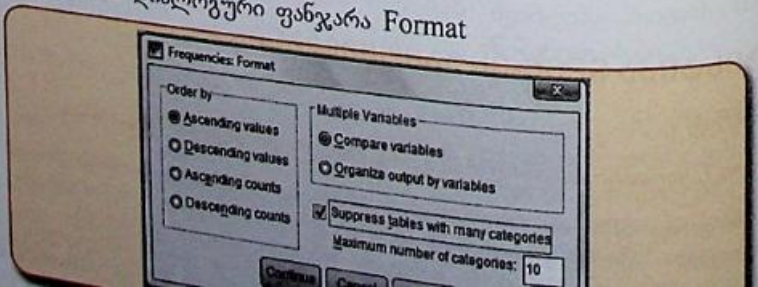


ვარიაციის საზომების შერჩევა შესაძლებელია ყუთიდან Dispersion, რომელიც დიალოგური ფანჯრის ბოლოს მდებარეობს (იხ. ნახაზი 5.9). ჩვენ შევარჩიეთ სტანდარტული გადახრა standard deviation, დისპერსია variance, დიაპაზონი range და საშუალო არითმეტიკული mean (ცენტრალური ტენდენციის ყუთიდან). პერცენტილური სიდიდეების ყუთიდან Percentile Values box შევარჩიეთ კვარტილები Quartiles, რათა SPSS-ს მივცეთ ბრძანება გამოითვალოს 25-ე, 50-ე და 75-ე პერცენტილები. SPSS საშუალებას გვაძლევს განვსაზღვროთ კონკრეტული პერცენტილები, თუ მათ პერცენტილების ყუთში ჩავწერთ (მაგ. 34-ე პერცენტილი) და დავამატებთ პერცენტილების ყუთს დდ ღილაკით.

ჩვენ უკვე ვნახეთ სიხშირული ცხრილები ცვლადისთვის AGE. ამიტომ მას შემდეგ, რაც დაკლიკეთ Continue, კელავ დაკლიკეთ Format, რათა გამოვრთოთ ცხრილის გამოტანის ფუნქცია. ეს შეიძლება გაკეთდეს „Suppress tables with many categories"-ზე დაკლიკებით (იხ. ნახაზი 5.10) აქ არის ფორმატირების ოფციები, რომლებიც შემდგომში შეგიძლიათ გამოიყენოთ.

დაკლიკეთ Continue, შემდეგ OK. SPSS გამოიტანს ყველა მოთხოვნილ საზომს (ნახაზი 5.11). ასაკის დიაპაზონია 71 წელი (18-დან 89-მდე), სტანდარტული გადახრა 17.381, რაც გვიჩვენებს, რომ ასაკებში ზომიერი დისპერსიაა (ამის დანახვა ავრეთვე შესაძლებელია მესამე თავში განხილული ასაკის ჰისტოგრამიდან), დისპერსია 302.102-ს ტოლია, რაც სტანდარტული გადახრის კვადრატს წარმოადგენს.

ნახაზი 5.10 დიალოგური ფანჯარა Format



ნახაზი 5.11 აღწერითი სტატისტიკა ცვლადისთვის AGE

Statistics

AGE OF RESPONDENT

N	Valid	1494
	Missing	6
Mean		47.69
Std. Deviation		17.381
Variance		302.102
Range		71
Percentiles	25	34.00
	50	47.00
	75	60.00

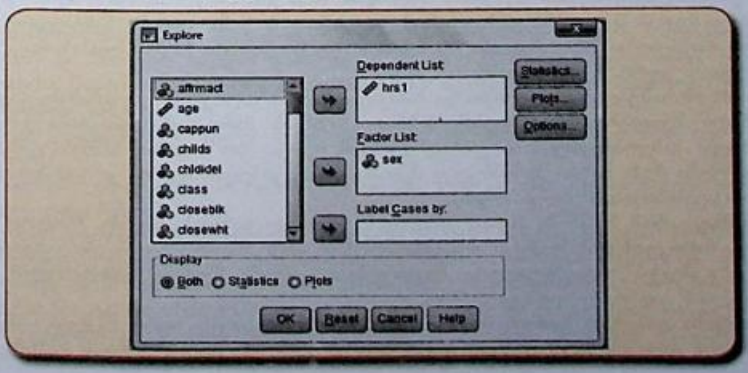
25-ე პერცენტილი არის 34, 50-ე პერცენტილი (რომელიც ამავე დროს მედიანაა) არის 47 და 75-ე პერცენტილის სიდიდეა 60. მიუხედავად იმისა, რომ SPSS არ ანგარიშობს IQV-ის, ჩვენ შეგვიძლია თავად შევასრულოთ ეს მოქმედება და 75-ე პერცენტილს დავაკლოთ 25-ე პერცენტილი. პასუხი იქნება 26 წელი. შეადარეთ ეს სიდიდე სტანდარტულ გადახრას.

[GSS08PFP-B]

მაგალითი 2: ვარიაციის საზომებისა და მართკუთხა დიაგრამების შექმნა ბრძანებით Explore

ვარიაციის საზომების წარმოება შეუძლია აგრეთვე Explore-ს, რომელიც მოთავსებულია Analyze მენიუს Descriptive Statistics განყოფილებაში. Explore-ის საშუალებით შეგვიძლია აგრეთვე მართკუთხა დიაგრამის აგება. მთავარ დიალოგურ ფანჯარაში (ნახაზი 5.12) ცვლადი უნდა მოვათავსოთ Dependent List box -ში. შეგიძლიათ ერთი ან რამდენიმე ნომინალური ცვლადის მოთავსება Factor List Box-ში. Explore ცალკე გამოიტანს სტატისტიკას ნომინალური ცვლადის ყველა კატეგორიისათვის.

ნახაზი 5.12 დიალოგური ფანჯარა Explore



მოათავსეთ ცვლადი HRS1 (უკანასკნელ კვირაში ნამუშევარი საათების რაოდენობა) Dependent box-ში, ხოლო SEX ცვლადი Factor box-ში განცალკევებული პასუხების მისაღებად სქესობრივ ჭრილში. დაკლიკეთ OK. დათქმის თანახმად, Explore თვითონ გამოიტანს ეკრანზე მთელ სტატისტიკას და დიაგრამებს სხვა ბრძანების მიცემის გარეშე. ამ ბრძანების საფუძველზე Explore არ გამოიტანს ჰისტოგრამასა და პერცენტილებს, თუმცა მას ეს ოფციები გააჩნია და ორივეს გაკეთება შეუძლია.

ნახაზი 5.13 მამაკაცების მიერ წინა კვირაში ნამუშევარი საათების რაოდენობა, ამორჩეული შედეგები

RESPONDENTS SEX				Statistic	Std. Error
NUMBER OF HOURS WORKED LAST WEEK	MALE	Mean		45.49	.698
		95% Confidence Interval for mean	Lower Bound	44.12	
			Upper Bound	46.86	
		5% Trimmed Mean		45.34	
		Median		43.50	
		Variance		231.236	
		Std. Deviation		15.206	
		Minimum		1	
		Maximum		89	
		Range		88	
		Interquartile Range		12.50	
		Skewness		.126	.112
		Kurtosis		.927	.224

ნახაზზე 5.13 მოცემულია ამორჩეული შედეგები მამაკაცების მიერ ნამუშევარი სააათების მიხედვით წინა კვირაში. Case Processing Summary Table-დან ჩანს, რომ კითხვაზე პასუხი გასცა 474 მამაკაცმა. ქალთაგან პასუხი გასცა 420-მა. ცხრილიდან ჩანს, რომ მამაკაცების მიერ წინა კვირაში ნამუშევარი საათების საშუალო რაოდენობაა 45.49, მედიანა 43.50, სტანდარტული გადახრა 15.206, დიაპაზონი 88, ინტერკვარტილური დიაპაზონი 12.5[16], რაც ძალიან ვიწროა დიაპაზონთან და სტანდარტულ გადახრასთან შედარებით. (დათქმის თანახმად შედეგების ფანჯარაში გამოტანილი იქნება აგრეთვე ფოთლებიანი ღეროების დიაგრამა, მაგრამ ამ სახელმძღვანელოში მას არ განვიხილავთ, ოფციის შეცვლით შეგვიძლია მისი გამორთვა). აგრეთვე აქ არ არის წარმოდგენილი, მაგრამ შედეგებში ვნახავთ, რომ ქალებისათვის წინა კვირაში ნამუშევარი საათების საშუალო რაოდენობაა 38.62, მედიანა 40, სტანდარტული გადახრა 12.346, დიაპაზონი 79. ვარიაცია ქალებთან უფრო მცირეა, ვიდრე მამაკაცებთან.

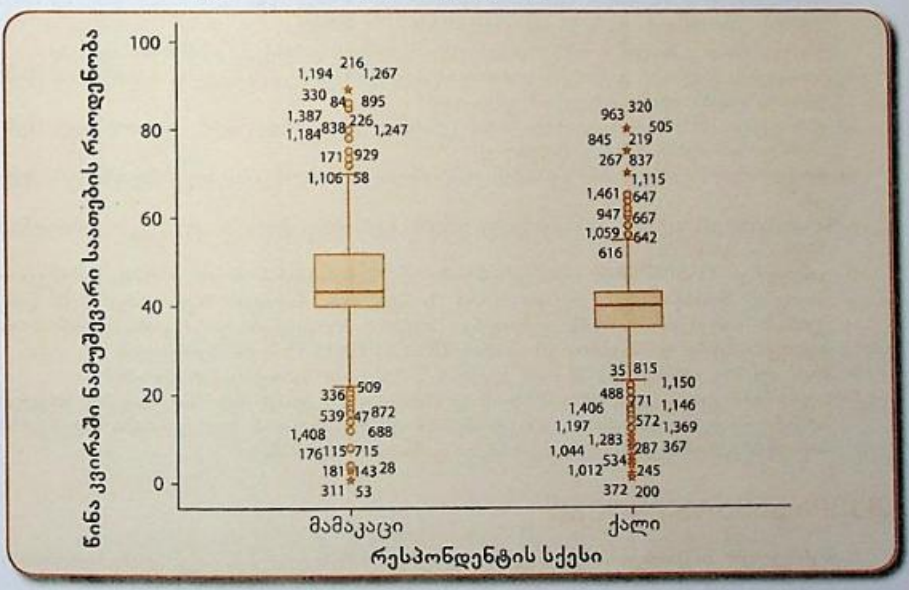
Explore-ის საშუალებით ქალებისა და მამაკაცებისათვის მართკუთხა დიაგრამები ცალ-ცალკეა გამოტანილი შედარების გაადვილების მიზნით. SPSS-ის მართკუთხა დიაგრამა გარკვეულად განსხვავდება სახელმძღვანელოში განხილულისაგან, თუმცა მათ საერთო ნიშნებიც ახასიათებთ. მართკუთხედში მუქი ხაზით აღნიშნულია მედიანა, მართკუთხედის სიგრძე კი გამოსახავს IQR-ს (12.5 მამაკაცებისთვის, 8 ქალებისთვის).

შევნიშნავთ, რომ SPSS-ში მართკუთხედიდან გამოსულ ხაზების სიგრძე მხოლოდ 1½-ჯერ აღემატება მართკუთხედის სიგრძეს. თუ განაწილებაში ამაზე მაღალი სიდიდეები გვხვდება, SPSS-ს ცალკე გამოაქვს ინდივიდუალური დაკვირვებები: კიდურა წევრები გამოსახულია წრეებით, ხოლო ძალიან კიდურა – ვარსკვლავებით.

მართკუთხა დიაგრამა გვიჩვენებს, რომ ვარიაცია ქალებისთვის უფრო ნაკლებია მამაკაცებისაზე. ამასთან ქალების მართკუთხედი უფრო მცირეა. IQR მამაკაცებში 40-დან 52.5-მდეა, ხოლო

ქალებისთვის –35-დან 43-მდე. ორივე სქესის დიაგრამაზე შეინიშნება კიდურა სიდიდეების არსებობა.

ნახაზი 5.14 მამაკაცებისა და ქალების მიერ წინა კვირაში ნამუშევარი საათების რაოდენობის მართკუთხა დიაგრამები



ამოცანები SPSS-ისთვის

[GSS08PFP-A]

1. გამოიყენეთ Frequencies პროცედურები და შეისწავლეთ რესპონდენტის მიმდინარე ასაკისა (AGE) და პირველი ბავშვის დაბადებისას ასაკის (AGKDBRN) ცვლადები. დაკლიკეთ Analyze, Descriptive Statistics, Frequencies, და შემდეგ Statistics. შეარჩიეთ ვარიაციის შესაბამისი საზომები.
   ა. რომელ ცვლადს აქვს უფრო მეტი ვარიაცია? პასუხისთვის გამოიყენეთ ერთზე მეტი საზომი.
   ბ. სოციალური თვალსაზრისით, რატომ ხასიათდება ერთი ცვლადი უფრო მაღალი ვარიაციით, ვიდრე დანარჩენები?
2. Explore პროცედურის გამოყენებით, ცვლადში AGKDBRN დააცალკევეთ სტატისტიკური მაჩვენებლები ქალებისა და მამაკაცებისთვის, SEX-ის შერჩევით Explore-ის ფანჯარაში, როგორც factor variable. დაკლიკეთ Analyze, Descriptive Statistics, Explore, შემდეგ AGKDBRN ჩასვით Dependent List-ში და SEX Factor List-ში. რა განსხვავებაა ქალებისა და მამაკაცების ასაკებში მათი პირველი შვილის დაბადებისას? შეაფასეთ ეს განსხვავებები ცენტრალური ტენდენციისა და ვარიაციის საზომებით.

3. გაიმეორეთ მე-2 სავარჯიშოს მოქმედებები და შეისწავლეთ დისპერსია ცვლადებში EDUC (education) და PRESTG80 (occupational prestige score). შეარჩიეთ თქვენი ფაქტორული (ნომინალური) ცვლადი, რომ შეადაროთ ფაქტორებს, როგორიცაა CLASS, RACECEN1, ან სხვა რომელიმე. დაკლიკეთ Analyze, Descriptive Statistics, Explore, შემდეგ ჩასვით EDUC და PRESTG80 Dependent List-ში და თქვენ მიერ შერჩეული ფაქტორული ცვლადი Factor List-ში. შეარჩიეთ ვარიაციის შესაბამისი საზომები და შეაფასეთ ვარიაცია.
4. MTF2008-ის გამოყენებით შეისწავლეთ რესპონდენტების აზრი პოლიციის ქცევისა (POLICE) და სამხედროების ქცევის (MILITARY) შესახებ.

ა. უპირველესად, გამოიყენეთ SPSS თითოეული ცვლადის გაზომვის დონის დასადგენად.

ბ. თითოეული ცვლადის გაზომვის დონის გათვალისწინებით, ცენტრალური ტენდენციის რომელი საზომები იქნება მისაღები? ვარიაციის?

გ. გამოიყენეთ SPSS და კალკულატორი ცენტრალური ტენდენციისა და ვარიაციის შესაბამისი საზომების გასაანგარიშებლად.

დ. რომელს უფრო დადებითად აფასებენ რესპონდენტები, პოლიციისა თუ სამხედროების ქცევას?

ე. ნახეთ, იცვლება თუ არა (დ) კითხვაზე პასუხი სქესობრივ ჭრილში. შეხსენება: გამოიყენეთ Data და Split File.

5. გამოიყენეთ GSS08PFP-B თეთრკანიანებისა და შავკანიანების მიერ კვირაში ნამუშევარი საათების შესასწავლად. ცვლადი HRS1 ზომავს რესპონდენტის მიერ ინტერვიუს წინა კვირაში ნამუშევარ საათებს. გამოიყენეთ Explore პროცედურა ვარიაციის შესასწავლად თეთრკანიანებსა და შავკანიანებს შორის (RACECEN1) GSS-ის შერჩევიდან.

ა. არის თუ არა განსხვავება ამ ორი ჯგუფის ნამუშევარი საათების ვარიაციაში?

ბ. აჯეთ მართკუთხა დიაგრამა SPSS-ის დახმარებით და დიაგრამის საფუძველზე მოკლედ აღწერეთ თქვენი დასკვნები თეთრკანიანებისა და შავკანიანების მიერ კვირაში ნამუშევარი საათების ცენტრალური ტენდენციისა და ვარიაციის შესახებ.

თავის სავარჯიშოები

1. ამერიკელებს საკმაოდ განსხვავებული პოლიტიკური შეხედულებები აქვთ, ლიბერალურიდან დაწყებული, კონსერვატულის ჩათვლით. გამოიყენეთ GSS 2008-იდან მოტანილი მონაცემები და შეისწავლეთ პოლიტიკური შეხედულებების განსხვავებულობა. ხვედრით წილებში ასახულია რესპონდენტების პოლიტიკური შეხედულებების თვითშეფასება. (გამოუპასუხებელი კითხვები ამოღებულია ამ მაგალითიდან).

პოლიტიკური შეხედულებები	ხვედრითი წილი, %
უკიდურესად ლიბერალური	3.6
ლიბერალური	12.2
სუსტად ლიბერალური	12.0
ზომიერი	38.0
სუსტად კონსერვატული	14.1
კონსერვატული	16.8
უკიდურესად კონსერვატული	3.3
სულ	100.0

ა. რამდენი კატეგორიაა (K) ამ მაგალითში?

ბ. გაიანგარიშეთ ხვედრითი წილების კვადრატების ჯამი, ანუ $\sum Pct^2$.

გ. რას უდრის IQV ამ ცვლადებისთვის? თქვენს მოლოდინთან შედარებით, ეს საზომი მაღა-

ლია (1-თან ახლოსაა) თუ დაბალი (0-თან ახლოსაა)? გამოგადგათ თუ არა IQV თქვენი მოლოდინების განმტკიცებაში და რა დასკვნის გამოტანა შეიძლება ცხრილიდან?

2. ქვემოთ მოტანილი GSS 2008-ის მონაცემების საფუძველზე უპასუხეთ კითხვებს განათლების დონესთან დაკავშირებით:

უმაღლესი საგანმანათლებლო ხარისხი	მამაკაცები	ქალები
საშუალო სკოლაზე დაბალი	14.6	14.2
საშუალო სკოლა	47.7	49.9
კოლეჯი	9.9	9.3
ბაკალავრი	18.1	17.2
მაგისტრი	9.6	9.3
სულ[ა]	99.9	99.9

ა. ჯამი 100%-ის ტოლი არ არის დამრგვალების გამო.

ა. რას უდრის K-ს სიდიდე?

ბ. გაიანგარიშეთ ხვედრითი წილების კვადრატების ჯამი, ანუ $\sum Pct^2$ ქალებისა და მამაკაცებისათვის ცალ-ცალკე.

გ. გამოიყენეთ (ა) და (ბ)-ს გაანგარიშებები და გამოთვალეთ IQV მამაკაცებისა და ქალებისათვის. არის თუ არა დიდი განსხვავება განათლების დონეებში სქესობრივ ჭრილში?

3. კორუფცია სამოქალაქო სექტორში მუდმივი განხილვის საგანია. ნახეთ აშშ-ის სამართლის დეპარტამენტის მონაცემები 1990 და 2007 წლებში.

სამოქალაქო სექტორში გამოვლენილი კორუფციის შემთხვევები წლების მიხედვით

1990		2007	
მთავრობის დონე	დანაშაულის რაოდენობა	მთავრობის დონე	დანაშაულის რაოდენობა
ფედერალური	583	ფედერალური	405
შტატი	79	შტატი	85
ადგილობრივი	225	ადგილობრივი	275

წყარო: U.S. Census Bureau, The 2010 *Statistical Abstract*, Table 327.

ა. რას უდრის დიაპაზონი 1990 წელს? 2007 წელს? რომელი უფრო მაღალია?

ბ. რას უდრის დანაშაულის საშუალო რაოდენობა 1990 და 2007 წლებში?

გ. გაიანგარიშეთ სტანდარტული გადახრა 1990 და 2007 წლებში.

დ. სტანდარტული გადახრის მიხედვით, რომელ წელს უფრო მაღალი იყო ვარიაცია? შეესაბამება თუ არა ეს შედეგები დიაპაზონით მიღებულ ინფორმაციას?

4. დაწერეთ ანგარიში სამართალდარღვევებზე აშშ-ის სამართლის დეპარტამენტის მონაცემების საფუძველზე. მასში უნდა გამოიყენოთ ცენტრალური ტენდენციისა და ვარიაციის საზომები და 2005 და 1995 წლების შედარებები. დაურთეთ განმარტებები აღნიშნულ შედარებებთან დაკავშირებით.

1990		2005	
დანაშაულის ტიპი	ეჭვმიტანილთა რაოდენობა	დანაშაულის ტიპი	ეჭვმიტანილთა რაოდენობა
ძალადობა	5,720	ძალადობა	5,485
ქონება	31,759	ქონება	25,570
ნარკოტიკები	31,686	ნარკოტიკები	40,038
საზოგადოებრივი წესრიგი	19,036	საზოგადოებრივი წესრიგი	21,583
იარაღი	5,376	იარაღი	13,689
იმიგრაცია	7,256	იმიგრაცია	36,559

წყარო: U.S. Census Bureau, The 2010 *Statistical Abstract*, Table 317.

5. ამ სავარჯიშოში გამოიყენეთ ცხრილი 5.4-ის მონაცემები და გააგრძელეთ შედარებები რეგიონების მიხედვით. გამოიყენეთ მხოლოდ დასავლეთ და შუა-დასავლეთ შტატების მონაცემები.
   ა. შეადარეთ დასავლეთის და შუა-დასავლეთის შტატების დიაპაზონები. რომელი რეგიონის დიაპაზონია უფრო მაღალი ასაკოვანი მოსახლეობის საპროგნოზო მატების ტემპის მიხედვით?
   ბ. გაიანგარიშეთ თითოეული რეგიონის IRQ. რომელი უფრო მაღალია?
   გ. გამოიყენეთ ვარიაციის საზომები და დაახასიათეთ განსხვავებები ამ ორი რეგიონის მატების ტემპებში. არის თუ არა ერთ შტატში ვარიაცია მეორეზე მაღალი? თუ ასეა, დაასაბუთეთ.
6. საქმიანობის პრესტიჟი წარმოადგენს სოციოლოგების კვლევის საგანს და ზომავს საქმიანობის სტატუსს. ის აგრეთვე პიროვნების სოციალ-ეკონომიკური სტატუსის შემადგენელი ნაწილია. საშუალოდ, მაღალი განათლების ხარისხის მქონე ადამიანებს უფრო მაღალი საქმიანობის პრესტიჟი აქვთ. ჩვენ შევისწავლეთ აღნიშნული GSS 2008-ის მონაცემების საფუძველზე ცვლადის PRESTG80-ის და EXPLORE პროცედურის დახმარებით SPSS-ში, რაც ნაჩვენებია ნახაზზე 5.15.
   ა. მოდისა და მედიანის მონაცემებზე დაკვირვების შედეგად, არის თუ არა საშუალო სკოლის დიპლომის მქონეთა საქმიანობის პრესტიჟის განაწილება გადახრილი? ბაკალავრებისთვის? რატომ დიახ, ან არა?
   ბ. ახსენით, რატომ ფიქრობთ, რომ ვარიაცია ორივე ჯგუფში ერთნაირია, ან განსხვავებულია?

ნახაზი 5.15 საქმიანობის პრესტიჟის აღწერითი სტატისტიკა მიღებული განათლების უმაღლესი ხარისხის მიხედვით

PRESTG80			Statistic
RS OCCUPATIONAL PRESTIGE SCORE (1980)	High School Diploma	Mean	40.41
		Median	39.00
		Std. Deviation	11.756
		Minimum	17
		Maximum	75
		Range	58
		Interquartile Range	18
	Bachelor's Degree	Mean	51.93
		Median	51.00
		Std. Deviation	12.934
		Minimum	17
		Maximum	86
		Range	69
		Interquartile Range	21

7. აშშ-ის აღწერის ბიურო ინფორმაციას აგროვებს განქორწინების დონის შესახებ. ქვემოთ მოტანილია შემაჯამებელი მონაცემები 2007 წლისთვის. გამოიყენეთ ცხრილი და უპასუხეთ კითხვებს:

შტატი	განქორწინების დონე 1,000 კაცზე
ალასკა	4.3
ფლორიდა	4.7
აიდაჰო	4.9
მეინი	4.5
მერილენდი	3.1
ნევადა	6.5
ნიუ-ჯერსი	3.0
ტეხასი	3.3
ვერმონტი	3.8
ვისკონსინი	2.9

წყარო: U.S. Census Bureau, The 2010 *Statistical Abstract*, Table 126.

ა. გაიანგარიშეთ დიაპაზონი და IRQ. რომელი უფრო უკეთ ზომავს ვარიაციას? რატომ?
ბ. გაიანგარიშეთ და მოახდინეთ საშუალოსა და სტანდარტული გადახრის ინტერპრეტაცია.
გ. გამოავლინეთ შტატებს შორის განქორწინების დონეებში არსებული განსხვავების ორი ახსნა.

8. HINTS 2007-ის მონაწილეებმა თავიანთი ფსიქოლოგიური მდგომარეობა განსაზღვრეს 0-დან 24-მდე სკალაზე. ცხრილში მოტანილია განსხვავებული მონაცემები ორი ჯგუფის მიხედვით: რესპონდენტებისთვის, რომლებსაც დაუდგინდათ კიბოს დიაგნოზი, და რომლებსაც არ დაუდგინდათ.

	ფსიქოლოგიური სტრესის ქულა	
	დიაგნოზირებული	არადიაგნოზირებული
$\overline{Y}$	3.9	4,87
$\sum Y$	729	5,849
$\sum (Y - \overline{Y})^2$	3,059.14	25,180.20
N	187	1,200

ა. ორივე ჯგუფისთვის გაიანგარიშეთ დისპერსია და სტანდარტული გადახრა.
ბ. რა შეგიძლიათ თქვათ ამ ჯგუფებში ვარიაციის შესახებ? რატომ შეიძლება, რომ ისინი განსხვავებული იყოს?
გ. როგორ უნდა გამოვიყენოთ საშუალო სიდიდე დისპერსიისა და სტანდარტული გადახრის გასაანგარიშებლად?

9. თქვენ გაინტერესებთ ჩადენილი დანაშაულების ვარიაცია და პოლიციის დანახარჯები აშშ-ის აღმოსავლეთ და შუა-დასავლეთ შტატებში. ქვემოთ მოტანილია აშშ-ის აღწერის ბიუროს მონაცემები 2006 წლისთვის 21 შტატში.

შტატი	დანაშაულის რაოდენობა 100,000 კაცზე	პოლიციის დანახარჯები (მლნ. დოლარი)
მეინი	2,635	221
ნიუ-ჰემპშირი	2,013	274
ვერმონტი	2,442	136
მასაჩუსეტსი	2,838	1,673
როდ-აილენდი	2,815	286
კონექტიკუტი	2,785	905
ნიუ-იორკი	2,488	7,585
ნიუ-ჯერსი	2,644	3,010
პენსილვანია	2,883	2,535
ოჰაიო	4,029	2,689
ინდიანა	3,817	1,039
ილინოისი	3,562	3,761
მიჩიგანი	3,775	2,333
ვისკონსინი	3,102	1,434
მინესოტა	3,398	1,294
აიოვა	3,087	570
მისური	4,373	1,179
ჩრდ. დაკოტა	2,128	106
სამხრ. დაკოტა	1,791	132
ნებრასკა	3,623	334
კანზასი	4,175	635

წყარო: U.S. Census Bureau, The 2010 *Statistical Abstract*, Tables 297 and 431

ა. გაიანგარიშეთ საშუალო არითმეტიკული თითოეული ცვლადისთვის.
ბ. გაიანგარიშეთ სტანდარტული გადახრა თითოეული ცვლადისთვის.
გ. თითოეული ცვლადისთვის შეადარეთ საშუალო არითმეტიკული და სტანდარტულ გადახრა. სად უფრო მაღალია ვარიაცია, დანაშაულის რაოდენობაში, თუ პოლიციის დანახარჯებში? რომელ შტატებშია მაღალი ვარიაცია?

10. მე-9 სავარჯიშოზე დაყრდნობით ააგეთ მართკუთხა დიაგრამა. იმსჯელეთ დიაგრამის როლზე თქვენი დასკვნების გასაძლიერებლად.
11. ამ სავარჯიშოში გამოიყენეთ ცხრილი 5.5-ის მონაცემები.
ა. გაიანგარიშეთ სტანდარტული გადახრა.
ბ. მიღებული სტატისტიკა შეადარეთ IQR-სა და ნახაზზე 5.4 გამოსახულ მართკუთხა დიაგრამას. რომელი უფრო მაღალია: IRQ თუ სტანდარტული გადახრა?
გ. IRQ-სა და მართკუთხა დიაგრამიდან გამომდინარე, იგივე დასკვნებამდე მივყავართ თუ არა სტანდარტულ გადახრას?
12. თქვენ გადაწყვიტეთ გამოიყენოთ MTF 2008-ის მონაცემები ახალგაზრდა ამერიკელების მიერ ალკოჰოლის გამოყენებასა (ATDRINK) და სიგარეტის მოწევასთან (ATSMOKE) დაკავშირებით.

მოზარდების დამოკიდებულება ალკოჰოლისა და სიგარეტისადმი	ალკოჰოლის მოხმარება	ერთი კოლოფი სიგარეტი დღეში
არ უარეყოფ	72.1%	19.8%
არ ვეთანხმები	14.4%	28.7%
სრულიად უარეყოფ	13.5%	51.5%
სულ	100%	100%
	N = 1,473	N = 1,455

ა. ვარიაციის რომელი საზომები იქნება მისაღები ამ ცვლადებისთვის? რატომ?
ბ. გაიანგარიშეთ ვარიაციის მისაღები საზომები ამ ცვლადებისთვის.
გ. რომელ ცვლადს აქვს უფრო მაღალი ვარიაცია? ახსენით.

13. 2006 წლისთვის მოცემულია 10 ქვეყნის ქალების სიცოცხლის ხანგრძლივობა. გაიანგარიშეთ ცენტრალური ტენდენციისა და ვარიაციის მაჩვენებლები როგორც ევროპული, ისე არაევროპული ქვეყნებისთვის. რომელ ჯგუფშია უფრო მაღალი ვარიაციის დონე? პასუხი ახსენით.

ქვეყანა	სიცოცხლის ხანგრძლივობა დაბადებისას
ევროპული ქვეყნები	
საფრანგეთი	84.4
გერმანია	82.4
ჰოლანდია	81.9
ესპანეთი	84.4
თურქეთი	74.0
არაევროპული ქვეყნები	
იაპონია	85.8
ავსტრალია	82.7
მექსიკა	78.1
ისლანდია	83.0
ახალი ზელანდია	82.2

წყარო: U.S. Census Bureau, The 2010 *Statistical Abstract*, Table 1304.

14. თქვენ გევალებათ მოკლე მიმოხილვის მომზადება ამერიკის ახალგაზრდა მძღოლების შესახებ. MTF 2008-ის ბაზაზე მიღებული სტატისტიკის საფუძველზე დაწერეთ 1-2 აბზაცი ამერიკის ახალგაზრდა მძღოლებზე. ახსენით თქვენი მიგნებები.

დარღვევების რაოდენობა ავტომობილის მართვის დროს უკანასკნელ 12 თვეში	N	საშუალო	სტანდარტული გადახრა
მოძრაობის წესების დარღვევა	1,357	0.38	0.804
საავტომობილო შემთხვევები	1,348	0.28	0.610

წყარო: Monitoring the Future, 2008.

15. მოქალაქეების მონაწილეობა არჩევნებში დემოკრატიის მნიშვნელოვანი შემადგენელია. შეამოწმეთ ამომრჩეველთა აქტივობა განათლების დონის მიხედვით საპრეზიდენტო არჩევნებზე 2008 წელს:

განათლების დონე	ხმა მისცა (ათასი კაცი)
საშუალოზე დაბალი	9,046
საშუალო	35,866
კოლეჯი	41,477
ბაკალავრი	29,330
მაგისტრი	15,425

წყარო: U.S. Census Bureau, Current *Population Survey*, 2008.

ა. იპოვეთ ამომრჩეველთა რაოდენობაში მე-20 და 80-ე პერცენტილები.
ბ. რამდენიმე წინადადებით ახსენით, რატომ გვმართებს დაკვირვება (ა) პასუხის ინტერპრეტაციის დროს. (გაიხსენეთ განსხვავება ნედლ და შეფარდებით სიხშირეებს შორის).

16. ცხრილში შეკამბულია რასობრივი განსხვავება განათლებისა და ბავშვების იდეალური რაოდენობის მიხედვით ჩინელ და ფილიპინელ ამერიკელებს შორის. საშუალოსა და სტანდარტულ გადახრაზე (ფრჩხილებში) დაყრდნობით, როგორ განსხვავდება ჩინელი და ფილიპინელი ამერიკელების მოსაზრება ბავშვების იდეალურ რაოდენობასთან დაკავშირებით?

	ჩინელი ამერიკელები	ფილიპინელი ამერიკელები
განათლება (წლები)	15.83 (2.691)	13.87 (3.335)
ბავშვების იდეალური რაოდენობა	2.20 (0.422)	4.00 (2.769)

წყარო: General Social Survey, 2008.

თავი 6

ნორმალური განაწილება

თავის სწავლების მიზნები

- ნორმალური განაწილების მნიშვნელობა და მისი გამოყენება სტატისტიკაში
- ნორმალური განაწილების დამახასიათებელი ნიშნები
- ნედლი ქულის გარდაქმნა სტანდარტულ (Z) ქულად და პირიქით
- სტანდარტული ნორმალური ცხრილის გამოყენება
- Z ქულის გამოსახვა კოეფიციენტით (ან პროცენტით) და პირიქით
- ქულის პერცენტილური რანგის პოვნა და მისი განმარტება

წინა თავებში რამდენიმე მნიშვნელოვანი საკითხი ვისწავლეთ განაწილებების შესახებ: როგორ გარდავქმნათ ისინი სიხშირულ განაწილებად, როგორ გამოვსახოთ გრაფიკულად და როგორ აღვწეროთ ცენტრალური ტენდენცია და ვარიაცია ისეთი საზომებით, როგორიცაა საშუალო და სტანდარული გადახრა. ჩვენ აგრეთვე ვისწავლეთ, რომ განაწილებებს განსხვავებული ფორმა აქვთ. ზოგიერთი განაწილება სიმეტრიულია, ხოლო ზოგიერთი — ნეგატიურად ან პოზიტიურად გადახრილი. ჩვენ მიერ უკვე აღწერილი განაწილებები *ემპირიულია*, რაც იმას ნიშნავს, რომ ისინი რეალურ მონაცემებზეა დაფუძნებული.

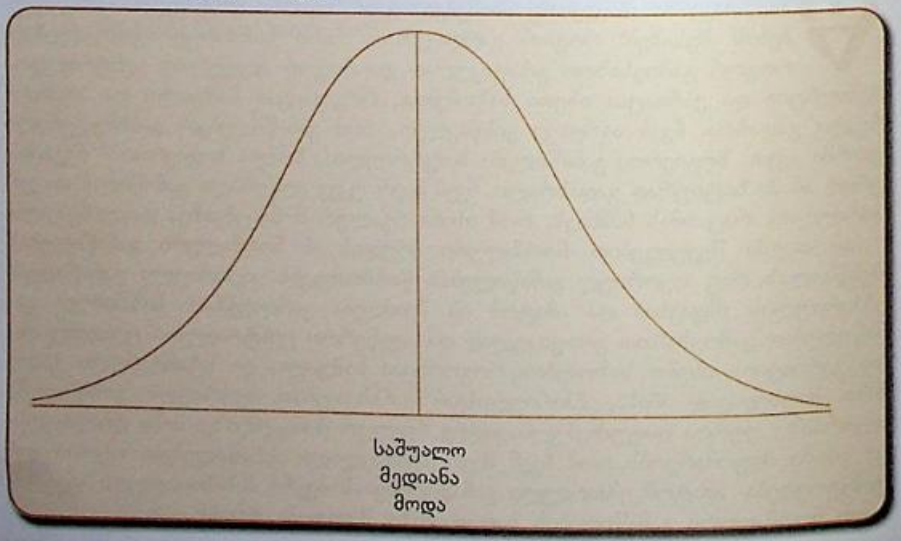
ამ თავში შევუდგებით *ნორმალური მრუდის* ან **ნორმალური განაწილების** შესწავლას, რაც თეორიულ განაწილებას წარმოადგენს. *თეორიული განაწილება* ემპირიულის მსგავსია და ამიტომ ის შეიძლება გარდავქმნათ სიხშირულ განაწილებად, გამოვსახოთ გრაფიკულად და აღვწეროთ ცენტრალური ტენდენციისა და ვარიაციის ისეთი საზომებით, როგორიცაა საშუალო და სტანდარული გადახრა. მიუხედავად ამისა, ემპირიულისგან განსხვავებით, თეორიული განაწილება თეორიაზე უფროა დაფუძნებული, ვიდრე რეალურ მონაცემებზე. მისი ღირებულება იმაში მდგომარეობს, რომ ჩვენ მიერ შესწავლილი განაწილებები თეორიულს უახლოვდება. ამიტომ ემპირიული განაწილებების ბევრი მახასიათებელი თეორიული ნორმალური განაწილების საფუძველზე შეიძლება შევისწავლოთ.

ნორმალური განაწილება არის ზარის ფორმის სიმეტრიული თეორიული განაწილება, რომლის საშუალო, მედიანა და მოდა მის უმაღლეს წერტილში ერთმანეთს ემთხვევა, ხოლო სიხშირეები თანდათან მცირდება მრუდის ბოლოებისკენ.

ნორმალური განაწილების დამახასიათებელი ნიშნები

ნორმალურ მრუდს (ნახაზი 6.1) ზარის ფორმის სიხშირული პოლიგონის სახე აქვს. ამის გამო მას ხშირად *ზარის ფორმის მრუდს* უწოდებენ. სიმეტრიულობა ნორმალური განაწილების ყველაზე გამოკვეთილი დამახასიათებელი ნიშანია. თუ ნახაზს 6.1 შუაზე გაკეცავთ, მიიღებთ ორ თანაბარ ნაწილს, ერთი მეორის სარკისებრ ანარეკლს წარმოადგენს. ეს ნიშნავს, რომ განაწილების ორივე მხარეს დაკვირვებების თანაბარი რაოდენობა მოხვდება. ამასთან, ნორმალური განაწილების შუა წერტილს უმაღლესი სიხშირე აქვს. ეს სწორედ ის წერტილია, სადაც სამი საზომი ერთმანეთს ემთხვევა: მოდა (წერტილი, რომელსაც უმაღლესი სიხშირე აქვს), მედიანა (წერტილი, რომელიც განაწილებას ორ ტოლ ნაწილად ყოფს) და საშუალო (ყველა დაკვირვების საშუალო სიდიდე). აღსანიშნავია, რომ დაკვირვებათა უმრავლესობა შუაშია თავმოყრილი, რომელთა სიხშირეები თანდათან მცირდება განაწილების ბოლოებისკენ.

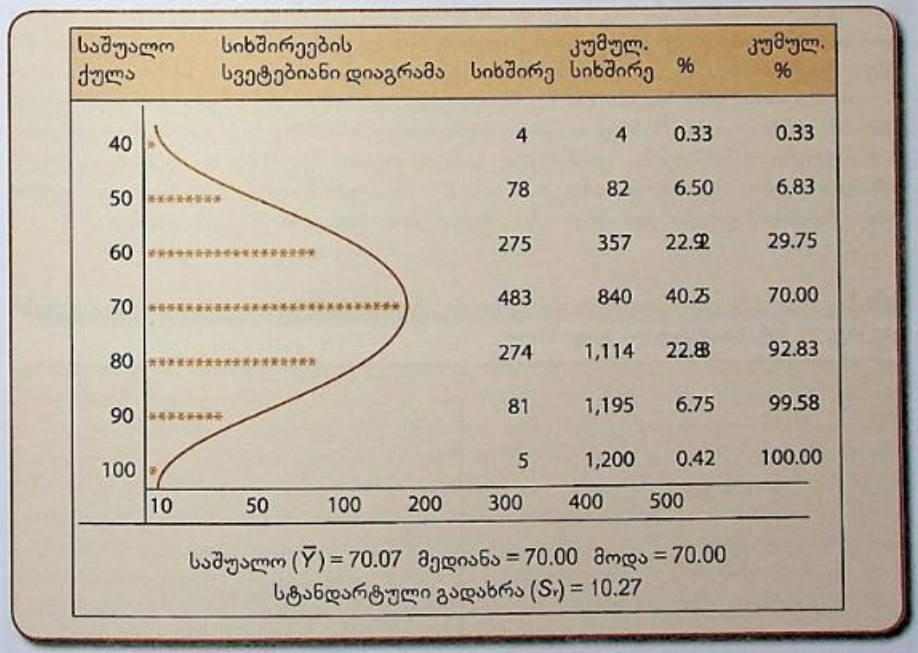
ნახაზი 6.1 ნორმალური მრუდი



ნორმალურ განაწილებასთან მიახლოებული ემპირიული განაწილებები

ნორმალური მრუდი თეორიული იდეალია, რომელსაც რეალური განაწილებები ზუსტად არასოდეს არ ემთხვევა. მიუხედავად ამისა, მკვლევარები მრავალ ცვლადს სწავლობენ (სტანდარტიზებული ტესტები, როგორიცაა SAT, ACT, GRE; სიმაღლე; ფიზიკური შესაძლებლობა: მრავალი სოციალური და პოლიტიკური შეხედულება), რომლებიც ძალიან ჰგავს თეორიულ მოდელს. როდესაც ვამბობთ, რომ ცვლადს აქვს „ნორმალური განაწილება", ვგულისხმობთ დაახლოებით ზარის ფორმის სიმეტრიულ განაწილებას, რომლის იდეალური მოდელი გამოსახულია ნახაზზე 6.1. ეს გვაძლევს იმის საშუალებას, რომ ემპირიული განაწილებები ნორმალური მრუდის ცოდნის საფუძველზე აღვწეროთ.

ცხრილი 6.1 1,200 სტუდენტის ფინალური ქულები სტატისტიკაში (1983-1993): ნორმალურ განაწილებასთან მიახლოებული



საშუალო ქულა	სიხშირეების სვეტებიანი დიაგრამა	სიხშირე	კუმულ. სიხშირე	%	კუმულ. %
40	*	4	4	0.33	0.33
50	********	78	82	6.50	6.83
60	**********	275	357	22.92	29.75
70	**********	483	840	40.25	70.00
80	**********	274	1,114	22.83	92.83
90	********	81	1,195	6.75	99.58
100	*	5	1,200	0.42	100.00
	10 50 100 200 300 400 500				

საშუალო ($\bar{Y}$) = 70.07 მედიანა = 70.00 მოდა = 70.00
სტანდარტული გადახრა (S_Y) = 10.27

მაგალითი: ფინალური ქულები სტატისტიკაში

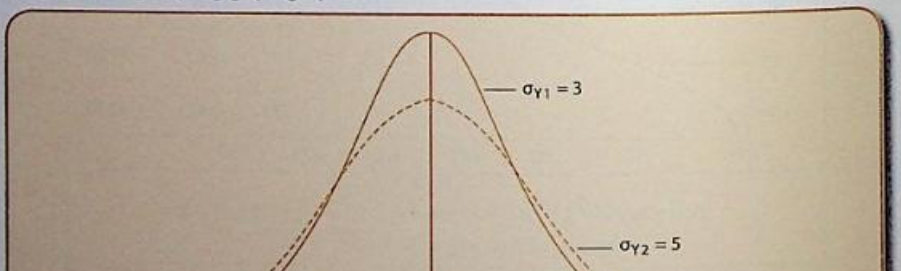
ნორმალური განაწილების თვისებების გაგება გაგვიადვილდება, თუ მას ნორმალურთან მიახლოებული რეალური განაწილების პირობებით განვიხილავთ. გან-

ვიხილოთ ცხრილში 6.1 წარმოდგენილი სიხშირეები და სვეტებიანი დიაგრამა. ეს არის პროფესორ ფრანკფორტ-ნაჩმიასის 1,200 სტუდენტის ფინალური შეფასებები სოციალურ სტატისტიკაში ვისკონსინ-მილუოკის უნივერსიტეტში 1983-1993 წლებში. იმაში დასარწმუნებლად, რომ *სტატისტიკის ფინალური ქულები* ნორმალურად არის განაწილებული, ჩვენ გავაკეთ ნორმალური მრუდი ცხრილში 6.1 ასახულ განაწილებაზე. შენიშნავდით, როგორ მიუახლოვდა ჩვენი განაწილება ნორმალურ მრუდს!

სტუდენტების მიერ ყველაზე ხშირად მიღებული ქულა 70-ის ტოლია, ამიტომ ეს არის განაწილების მოდა. რადგან სტუდენტების თითქმის ნახევარი ამ ქულის მაღლა (49.99%) ან დაბლა (50.01%)იმყოფება, საშუალოც (70.07) და მედიანაც (70) მიახლოებით 70-ის ტოლია. ცხრილში აგრეთვე ნაჩვენებია სტუდენტების რაოდენობის თანდათანობით კლება, ვინც მიიღო 70-ზე მეტი, ან ნაკლები ქულა. ძალიან ცოტა სტუდენტმა მიიღო 90-ზე მეტი ან 50-ზე დაბალი ქულა.

როდესაც ჩვენ ვიყენებთ ტერმინს *ნორმალური მრუდი*, ამით არ იგულისხმება, რომ ყველა განაწილება იდენტურია. ნორმალური განაწილების ფორმა ცვალებადია და ის დამოკიდებულია ცალკეული განაწილების საშუალოსა და სტანდარტულ გადახრაზე (μ_y-თი გამოისახება მთლიანი ერთობლიობის საშუალო, ხოლო σ_y-ით მთლიანი ერთობლიობის სტანდარტული გადახრა. ჩვენ კიდევ მივუბრუნდებით ამ სიმბოლოებს შემდეგ თავში). მაგალითად, ნახაზზე 6.2 გამოსახულია ორი ნორმალური განაწილება იდენტური საშუალოებით (μ_y=12) და განსხვავებული სტანდარტული გადახრებით(σ_{y1}=3, და σ_{y2}=5). განაწილება, რომელსაც უფრო დიდი სტანდარტული გადახრა აქვს, შედარებით განიერი და გაშლილია.

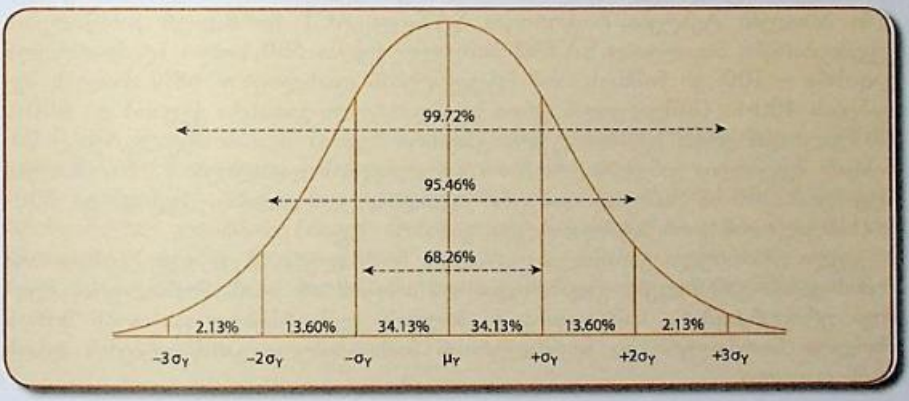
ნახაზი 6.2 ორი ნორმალური განაწილება ტოლი საშუალო სიდიდით და განსხვავებული სტანდარტული გადახრით



არეები ნორმალური მრუდის ქვეშ

ნორმალური მრუდის ფორმიდან გამომდინარე, ნორმალურ და მასთან მიახლოებულ მრუდებში საშუალოსა და მოცემულ წერტილს შორის არსებული არე (ფართობი), როდესაც მას სტანდარტული გადახრის ერთეულებში ვზომავთ, მუდმივი თანაფარდობით ხასიათდება. კონცეპტუალურად ეს ფართობი შერჩევაში მონაწილე დაკვირვებების ხვედრით წილს ან თანაფარდობას წარმოადგენს. ამიტომ მრუდის ქვეშ არსებული არე 1.00-ის ან დაკვირვებათა 100%-ის (1.00 × 100) ტოლია. რადგან ნორმალური მრუდი სიმეტრიულია, დაკვირვებათა ზუსტად 0.50 ან 50% ცენტრიდან მარჯვნივ ან ზემოთ განლაგდება, ხოლო დანარჩენი ნახევარი – ცენტრიდან მარცხნივ, ან ქვემოთ.

ნახაზი 6.3 ხვედრითი წილები ნორმალური მრუდის ქვეშ



ნახაზზე 6.3 გამოსახულია ხვედრითი წილები, რომლითაც დაკვირვებები საშუალოდან დაცილებულია 1,2, და 3 სტანდარული გადახრით ზევით და ქვევით. საშუალო სიდიდე დაკვირვებებს ზუსტად ორ ნაწილად ყოფს; მათი 34.13% საშუალოსა და 1 სტანდარტულ გადახრას შორის, მარჯვნივ მდებარეობს, ხოლო ამდენივე პროცენტი – საშუალოს მარცხნივ. პლუს ნიშანი გვეუბნება, რომ სტანდარტული გადახრა საშუალოს ზევით არის, ხოლო მინუსი – პირიქით. აქედან გამომდინარე, საშუალოსა და ±1 სტანდარტულ გადახრას შორის დაკვირვებათა 68.26% მოექცევა; საშუალოსა და ±2 სტანდარტულ გადახრას შორის დაკვირვებათა 95.46%; ხოლო საშუალოსა და ±3 სტანდარტულ გადახრას შორის დაკვირვებათა 99.72% მოექცევა.

მიმოიხილეთ ნორმალური მრუდის დამახასიათებელი ნიშნები. რას უდრის ფართობი მრუდის ქვეშ ? განაწილების რამდენი პროცენტი მოხვდება 1 სტანდარტული გადახრის ფარგლებში? 2 და 3 სტანდარტული გადახრის ფარგლებში? აღნიშნული დაადასტურეთ ნახაზზე 6.3 მოცემული პროცენტების დაჯამებით.

✓ *ცოდნის შემოწმება*

სტანდარტული გადახრის ინტერპრეტაცია

საშუალოსა და სტანდარტულ გადახრას შორის მდებარე ფართობის პროპორციები მუდმივია და მას დიდი პრაქტიკული მნიშვნელობა გააჩნია. რადგან განაწილება ნორმალურია და ჩვენთვის ცნობილია საშუალოსა და სტანდარტული გადახრის სიდიდეები, შეგვიძლია განვსაზღვროთ საშუალოსა და ნებისმიერ დაკვირვებას შორის არსებული შეფარდებითი (რელატიური) სიხშირე (პროპორცია, თანაფარდობა ან ხვედრითი წილი).

ეს თვისება ძალიან მნიშვნელოვანია ნორმალურთან მიახლოებული ემპირიული განაწილების სტანდარტული გადახრის ინტერპრეტაციისთვის. ასეთი განაწილებებისთვის, სადაც ცნობილია საშუალო და სტანდარტული გადახრა, შეგვიძლია საშუალოდან ნებისმიერი მანძილით დაცილებული დაკვირვების ხვედრითი წილის განსაზღვრა სტანდარული გადახრის საზომ ერთეულებში. ცნობილია, რომ კოლეჯში მისაღები ტესტები, როგორიცაა SAT და ACT ნორმალურ განაწილებას ექვემდებარება. მაგალითად, SAT-ის საშუალო ქულაა 500, ხოლო სტანდარტული გადახრა – 100. ეს ნიშნავს, რომ სტუდენტების დაახლოებით 68% მიიღებს შეფასებას 400-სა (საშუალოდან ერთი სტანდარტული გადახრა ქვევით) და 600-ს (საშუალოდან ერთი სტანდარტული გადახრა ზევით) შორის. ჩვენ აგრეთვე წინასწარ შეგვიძლია განვჭვრიტოთ, რომ სტუდენტების დაახლოებით 95% მიიღებს შეფასებას 300-სა (საშუალოდან ორი სტანდარტული გადახრა ქვევით) და 700-ს (საშუალოდან ორი სტანდარტული გადახრა ზევით) შორის.

ყველა ემპირიული განაწილება არ არის ნორმალური. ჩვენ უკვე ვნახეთ, რომ შემოსავლის განაწილება გადახრილია და ამიტომ არ არის ნორმალური. მუდმივი ურთიერკავშირი საშუალოსა და მრუდის ქვეშ არსებულ ფართობს შორის მხოლოდ ნორმალური და ნორმალურთან მიახლოებული განაწილებების დროს არის დაცული.

▣ სტანდარტული (Z) ქულები

განაწილების ნებისმიერ მნიშვნელობასა და საშუალოს შორის განსხვავება შეიძლება *სტანდარტული ქულებით*, ანუ *Z ქულებით* გამოისახოს. **Z ქულა** გვიჩვენებს, რამდენი სტანდარტული გადახრით არის ნედლი ქულა (ან დაკვირვების ქულა) საშუალოზე მაღალი ან დაბალი. ნედლი ქულა შეიძლება Z ქულაში გადავიყვანოთ, რათა ვნახოთ, რამდენი სტანდარტული გადახრით არის ის მეტი ან ნაკლები საშუალოზე.

სტანდარტული (Z) ქულა სტანდარტული გადახრების რაოდენობა, რომელიც გვიჩვენებს მოცემული ნედლი ქულის საშუალოდან დაშორების ზომას ზემოთ ან ქვემოთ.

ნედლი ქულის გადაყვანა Z ქულაში

ნედლის ქულის გადასაყვანად Z ქულაში ნედლ ქულასა და საშუალოს შორის სხვაობა უნდა გავყოთ სტანდარტულ გადახრაზე. მაგალითად, სტატისტიკის ფინალური ქულის გადასაყვანად, ფინალურ ქულას ვაკლებთ საშუალოს 70.07 და ვყოფთ სტანდარტულ გადახრაზე 10.27. ამიტომ, 80-ის Z ქულა იქნება:

$$\frac{80-70.07}{10.27}=0.97$$

ან 0.98 სტანდარტული გადახრა საშუალოს ზემოთ. ამის მსგავსად, 60-ის Z ქულა იქნება:

$$\frac{60-70.07}{10.27}=-0.98$$

ან 0.98 სტანდარტული გადახრა საშუალოს ქვემოთ; უარყოფითი ნიშანი გვიჩვენებს, რომ ეს ქულა საშუალოს ქვემოთ მდებარეობს.

ეს გაანგარიშება, რომელშიც ნედლ ქულასა და საშუალოს შორის განსხვავება სტანდარტულ გადახრაზე იყოფა, წარმოადგენს სტანდარტიზაციის მეთოდს და ცნობილია როგორც *ნედლი ქულის გადაყვანა Z ქულაში* (ცნობილია აგრეთვე, როგორც სტანდარტული ქულა). Z ქულის გასაანგარიშებელი ფორმულაა:

$$Z=\frac{Y-\bar{Y}}{S_Y} \qquad (6.1)$$

Z ქულა საშუალებას გვაძლევს გამოვსახოთ ნედლი ქულა მისი დამოკიდებულებით საშუალოსა და სტანდარტულ გადახრასთან. ის გვიჩვენებს, თუ რამდენად არის დაშორებული მოცემული ქულა საშუალოსგან სტანდარტული გადახრის ერთეულებში. დადებითი Z ქულა გვიჩვენებს, რომ ნედლი ქულა საშუალოზე მაღალია, ხოლო უარყოფითი Z ქულა გვიჩვენებს, რომ ნედლი ქულა საშუალოზე დაბალია. რაც უფრო მაღალია Z ქულა, მით უფრო დიდია სხვაობა ნედლ ქულასა და საშუალოს შორის (ცხრილი 6.2).

Z ქულის გადაყვანა ნედლ ქულაში

ნორმალური მრუდის გამოყენების დროს ზოგჯერ შებრუნებული პროცესი გვჭირდება, ანუ Z ქულების გარდაქმნა ნედლ ქულებად. Z ქულების გარდაქმნა ნედლ ქულებად ხდება საშუალოდან გარკვეულ მანძილზე მდგომი ნედლი ქულის საპოვნელად, როდესაც მანძილი სტანდარტული გადახრის ერთეულებშია

გამოსახული. ვთქვათ, გვინდა სტატისტიკის კლასში ფინალური ქულის პოვნა, რომელიც 1 სტანდარტული გადახრით ზემოთ არის საშუალოდან. ამ პრობლემის გადასაწყვეტად დავწეროთ Z ქულის გასაანგარიშებელი ფორმულა:

$$Z = \frac{Y - \overline{Y}}{S_Y}$$

ამ შემთხვევაში ჩვენთვის ცნობილია Z ($Z = 1$), საშუალო ($\overline{Y} = 70.07$) და სტანდარტული გადახრა $S_y = 10.27$. ჩვენ გვჭირდება Y-ის მნიშვნელობის გაგება:

$$1 = \frac{Y - 70.07}{10.27}$$

ცხრილი 6.2 Z ქულებად გარდაქმნილი სოციალური სტატისტიკის ფინალური ქულები

ფინალური ქულები	*Z ქულები*
40	$Z = \frac{40 - 70.07}{10.27} = \frac{-30.07}{10.27} = -2.93$
50	$Z = \frac{50 - 70.07}{10.27} = \frac{-20.07}{10.27} = -1.95$
60	$Z = \frac{60 - 70.07}{10.27} = \frac{-10.07}{10.27} = -0.98$
70	$Z = \frac{70 - 70.07}{10.27} = \frac{-0.07}{10.27} = -0.01$
80	$Z = \frac{80 - 70.07}{10.27} = \frac{9.93}{10.27} = 0.97$
90	$Z = \frac{90 - 70.07}{10.27} = \frac{19.93}{10.27} = 1.94$
100	$Z = \frac{100 - 70.07}{10.27} = \frac{29.93}{10.27} = 2.91$
$\overline{Y} = 70.07$	$S_y = 10.27$

მარტივი ალგებრული გარდაქმნებით, Y ტოლია:

$$Y = 70.07 + 1(10.27) = 80.34$$

80.34 საშუალოზე 1 სტანდარტული გადახრით (ან 1 Z ქულით) ზემოთ მდებარეობს.

Z ქულის ნედლ ქულაში გადასაყვანი ზოგადი ფორმულაა

$$Y = \overline{Y} + Z(S_y) \tag{6.2}$$

ამრიგად, Z ქულის ნედლ ქულაში გადასაყვანად Z ქულა უნდა გავამრავლოთ სტანდარტულ გადახრაზე და დავუმატოთ საშუალოს.

ახლა განვსაზღვროთ, რომელი ქულა მდებარეობს საშუალოზე 1.5 სტანდარტული გადახრით ქვემოთ? რადგან ქულა საშუალოს ქვემოთ მდებარეობს, ე.ი. Z ქულა ნეგატიურია. ამიტომ,

$$Y=70.07+(-1.5)(10.27)=70.07-15.41=54.66$$

54.66 ქულა საშუალოდან 1.5 სტანდარტული გადახრით ქვემოთ მდებარეობს.

ცხრილში 6.2 Z ქულები უკან გადაიყვანეთ ნედლ ქულებში. პასუხები უნდა დაემთხვეს ცხრილში მოცემულ ნედლ ქულებს.

✓ *ცოდნის შემოწმება*

▣ სტანდარტული ნორმალური განაწილება

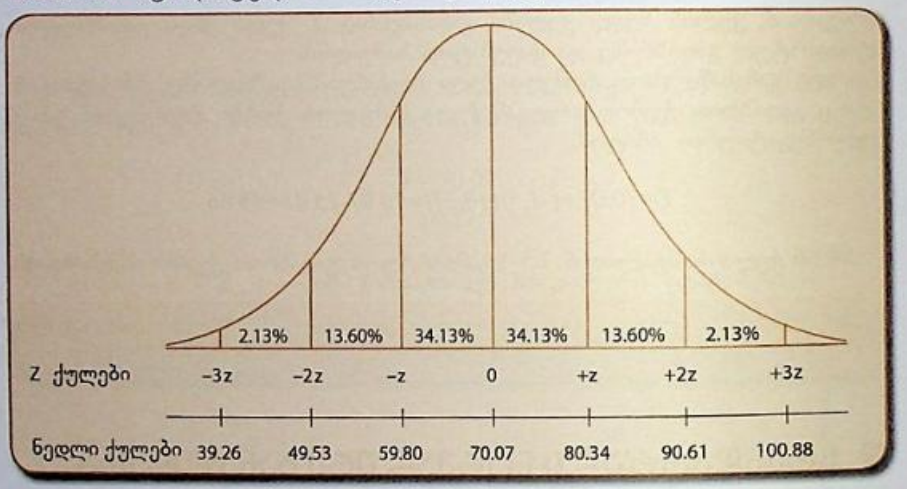
როდესაც ნორმალური განაწილება სტანდარტულ ქულებშია გამოსახული, მას **სტანდარტული ნორმალური განაწილება** ეწოდება. სტანდარტული, ანუ Z ქულები ეს არის რიცხვები, რომელიც სტანდარტული გადახრის ერთეულებით გამოსახავს მანძილს ფაქტიურ ქულებსა და საშუალოს შორის. სტანდარტულ ნორმალურ განაწილებაში საშუალო 0.00-ის ტოლია, ხოლო სტანდარტული გადახრა – 1.0-ის.

სტანდარტული ნორმალური განაწილება სტანდარტული ქულებით (Z) წარმოდგენილი ნორმალური განაწილება, სადაც საშუალო = 0-ს, ხოლო სტანდარტული გადახრა = 1-ს.

ნახაზზე 6.4 გამოსახულია ნორმალური სტანდარტული განაწილება მრუდის ქვეშ მდებარე არეებით, რაც ასოცირდება საშუალოს ზემოთ და ქვემოთ 1, 2 და 3 სტანდარტულ გადახრასთან. ნედლი და სტანდარტული ქულების დამოკიდებულების უკეთ გაგებისათვის, ნახაზზე ნაჩვენებია სტატისტიკის კლასის როგორც ნედლი, ისე სტანდარტული ქულები. მაგალითად, როდესაც სტატისტიკის კლასში ქულების განაწილების საშუალო 70.07-ის ტოლია, მისი შესაბამისი Z ქულა – სტანდარტული ნორმალური განაწილების საშუალო – 0-ია. 83.34 ქულა საშუა-

ლოდან ერთი სტანდარტული გადახრით ზემოთ მდებარეობს (70.07 + 10.27 = 80.34), ამიტომ მისი შესაბამისი Z ქულა არის +1; შესაბამისად, 59.80 ქულა საშუალოდან 1 სტანდარტული გადახრით ქვემოთ არის (70.07 − 10.27 = 59.80) და მისი შესაბამისი Z ქულა არის −1.

ნახაზი 6.4 სტანდარტული ნორმალური განაწილება



სტანდარტული ნორმალური ცხრილი

Z ქულების დახმარებით შეგვიძლია გავარკვიოთ დაკვირვებების რა პროცენტია მოთავსებული საშუალოსა და Z ქულებს შორის ნორმალურ განაწილებაში. სტანდარტული ნორმალური მრუდის ქვეშ არეები ან პროპორციები, რომლებიც Z ქულებს ან მის ნაწილებს შეესაბამება, თავმოყრილია სპეციალურ ცხრილში, რომელსაც **სტანდარტული ნორმალური ცხრილი** ეწოდება. ეს ცხრილი მოცემულია დანართში (ბ). ამ ნაწილში ჩვენ აგიხსნით ცხრილის გამოყენების წესს.

სტანდარტული ნორმალური ცხრილი არე სტანდარტული ნორმალური მრუდის ქვეშ (პროპორცია, რაც შეიძლება პროცენტულად გამოისახოს), რომელსაც შეესაბამება Z ქულა ან მისი ნაწილი.

სტანდარტული ნორმალური ცხრილის სტრუქტურა

ცხრილში 6.3 წარმოდგენილია სტანდარტული ნორმალური ცხრილის მცირე ნაწილი. აღსანიშნავია, რომ ცხრილი სამი სვეტისგან შედგება (ერთი გრძელი

ცხრილის ნაცვლად, აქ წარმოდგენილია ცხრილის ნაწილი – სამი სვეტის მაგივრად, ექვსსვეტიან ფორმატში).

ცხრილი 6.3 სტანდარტული ნორმალური ცხრილი

(ა)	(ბ)	(გ)	(ა)	(ბ)	(გ)
Z	არე საშუალოსა და Z-ს შორის	არე Z-ის გარეთ	Z	არე საშუალოსა და Z-ს შორის	არე Z-ის გარეთ
0.00	0.0000	0.5000	0.21	0.0832	0.4168
0.01	0.0040	0.4960	0.22	0.0871	0.4129
0.02	0.0080	0.4920	0.23	0.0910	0.4090
0.03	0.0120	0.4880	0.24	0.0948	0.4052
0.04	0.0160	0.4840	0.25	0.0987	0.4013
0.05	0.0199	0.4801	0.26	0.1026	0.3974
0.06	0.0239	0.4761	0.27	0.1064	0.3936
0.07	0.0279	0.4721	0.28	0.1103	0.3897
0.08	0.0319	0.4681	0.29	0.1141	0.3859
0.09	0.0359	0.4641	0.30	0.1179	0.3821
0.10	0.0398	0.4602	0.31	0.1217	0.3783
0.11	0.0439	0.4562	0.32	0.1255	0.3745
0.12	0.0478	0.4522	0.33	0.1293	0.3707
0.13	0.0517	0.4483	0.34	0,1331	0.3669
0.14	0.0557	0.4443	0.35	0.1368	0.3632
0.15	0.0596	0.4404	0.36	0.1406	0.3594
0.16	0.0636	0.4364	0.37	0.1443	0.3557
0.17	0.0675	0.4325	0.38	0.1480	0.3520
0.18	0.0714	0.4286	0.39	0.1517	0.3483
0.19	0.0753	0.4247	0.40	0.1554	0.3446
0.20	0.0793	0.4207			

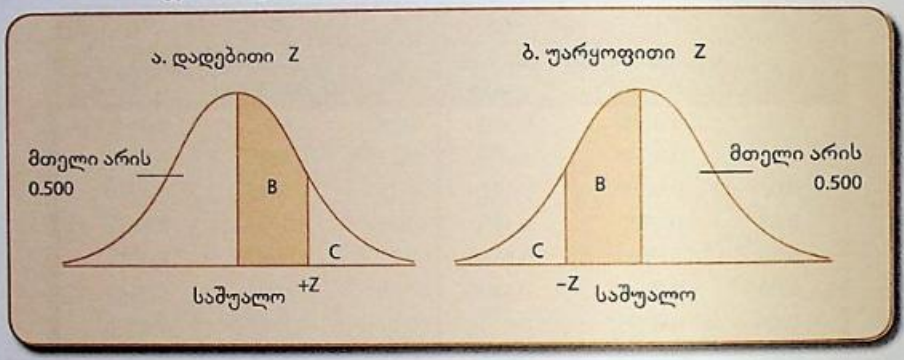
(ა) სვეტში მოცემულია დადებითი Z ქულები. რადგან ნორმალური მრუდი სიმეტრიულია, პროპორციები როგორც დადებითი, ისე უარყოფითი Z ქულებისთვის, იდენტურია.

სვეტში (ბ) მოცემულია საშუალოსა და (ა) სვეტში მოცემულ Z ქულებს შორის არეები. აღსანიშნავია, რომ როდესაც Z დადებითია, არე მოთავსებულია საშუალოს მარჯვნივ (იხ. ნახაზი 6.5ა), უარყოფითი Z ქულებისთვის კი ეს არე საშუალოს მარცხნივ იქნება (ნახაზი 6.5ბ).

სვეტში (გ) მოცემულია (ა) სვეტში ჩამოთვლილი Z ქულების გარეთ არსებული არეების პროპორციები. დადებითი Z ქულების შესაბამისი არეები მრუდის

მარჯვენა მხარეს (იხ. ნახაზი 6.5ა) არის, ხოლო უარყოფითის შესაბამისი – მარცხნივ (ნახაზი 6.5ბ).

ნახაზი 6.5 არეები საშუალოსა და Z-ს შორის (B) და Z-ს მიღმა (C)



Z ქულის გამოსახვა კოეფიციენტითა და პროცენტით

გიჩვენებთ, თუ როგორ უნდა გამოიყენოთ დანართი ბ სტატისტიკაში სტუდენტების ფინალური ქულების მაგალითზე (იხ. ცხრილი 6.1). ამ ნაწილში განხილული მაგალითები მოითხოვს Z ქულების გადაყვანას ათწილადებში (ან პროცენტებში).

არის პოვნა საშუალოსა და რაიმე დადებით Z ქულას შორის

საშუალოსა და განსაზღვრულ დადებით Z ქულას შორის არის საპოვნელად გამოიყენეთ სტანდარტული ნორმალური ცხრილი. საშუალოსა (70.07) და 85 ქულას შორის სტუდენტების პროცენტის საპოვნელად მიჰყევით შემდეგ საფეხურებს:

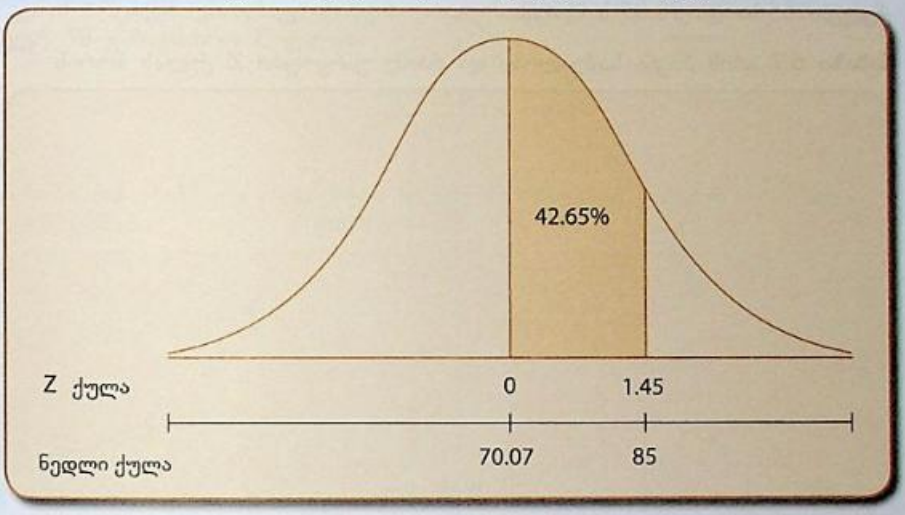
1. გარდაქმენით რიცხვი 85 Z ქულად:

$$Z = \frac{85 - 70.07}{10.27} = 1.45$$

2. მოძებნეთ 1.45 დანართი (ბ)-ს (ა) სვეტში და მისი შესაბამისი არე (ბ) სვეტში, 0.4265. ეს ათწილადი გადაიყვანეთ პროცენტებში (0.4265×100=42.65%) და მიიღებთ არის პროცენტს, რომელიც მოთავსებულია საშუალოსა და Z ქულას, 1.45-ს შორის.

3. აქედან, სტუდენტების 42.65%-მა მიიღო შეფასება 70.07-სა და 85-ს შორის.

ნახაზი 6.6 არის პოვნა საშუალოსა და რაიმე დადებით Z ქულას შორის



სტუდენტების ფაქტობრივი რიცხვის საპოვნელად, ვინც მიიღო ქულები 70.07-სა და 85-ს შორის, სტუდენტების საერთო რაოდენობა გაამრავლეთ 0.4265-ზე. ეს დაახლოებით 512 სტუდენტია (0.4265×1,200=512).

არის პოვნა საშუალოსა და რაიმე უარყოფით Z ქულას შორის

სტუდენტების რამდენმა პროცენტმა მიიღო ქულები 65-სა და 70.07-ს შორის? გამოიყენეთ სტანდარტული ნორმალური ცხრილი და შეასრულეთ შემდეგი გაანგარიშებები:

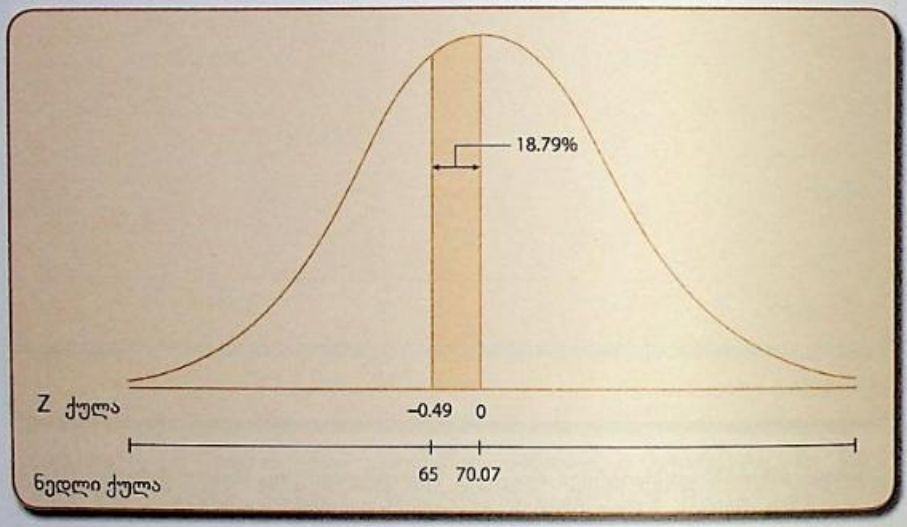
1. გადაიყვანეთ რიცხვი 65 Z ქულაში:

$$Z = \frac{65 - 70.07}{10.27} = 0.49$$

2. რადგან დადებით და უარყოფით Z ქულებს იდენტური არეები შეესაბამება, Z-ის უარყოფით ნიშანს არ ვაქცევთ ყურადღებას და სვეტში (ა) ვეძებთ 0.49-ს. Z ქულა 0.49-ს შეესაბამება 0.1879. ეს ნიშნავს, რომ არის 0.1879-ედი მოთავსებულია საშუალოსა და Z ქულა −0.49-ს შორის (ნახაზი 6.7). შემდეგ გადაგვყავს ეს ათწილადი პროცენტში 18.79% (0.1879×100=18.79%).

3. ამგვარად, დაახლოებით 225-მა (0.1879×1,200=225) სტუდენტმა მიიღო ქულები 65-სა და 70.07-ს შორის.

ნახაზი 6.7 არის პოვნა საშუალოსა და რაიმე უარყოფით Z ქულას შორის



არის პოვნა დადებითი Z ქულის ზემოთ და უარყოფითი Z ქულის ქვემოთ

ჩვენ შეგვიძლია შევადაროთ ერთმანეთს ძალიან კარგი და ძალიან ცუდი ქულების მქონე სტუდენტები სხვა სტუდენტებს.

ძალიან კარგი ქულების ათვლა დავიწყეთ 85 ქულიდან. 85-ზე მაღალი ქულის მქონე სტუდენტების რაოდენობის განსაზღვრისათვის, ჯერ 85 გადავიყვანოთ Z ქულაში:

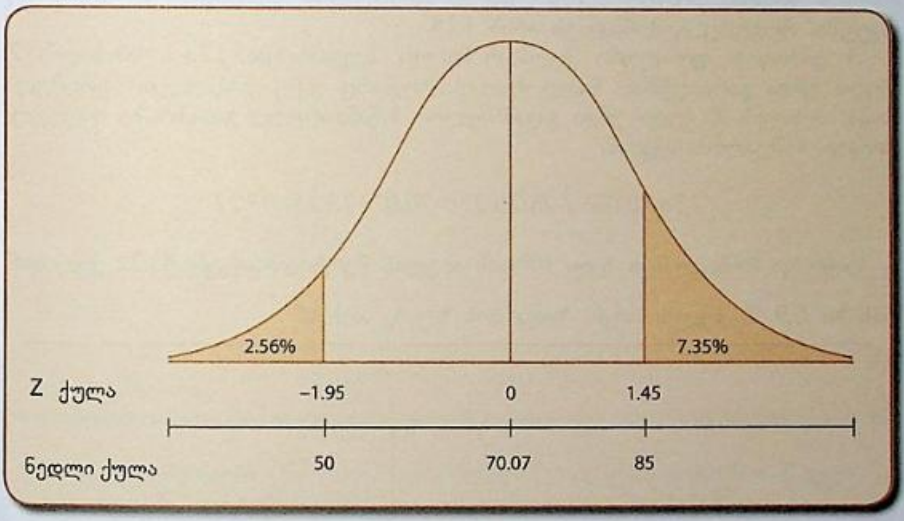
$$Z = \frac{85 - 70.07}{10.27} = 1.45$$

1.45-ის მიღმა არეში მოხვდნენ ის სტუდენტები, რომლებმაც 85 ქულაზე მაღალი შეფასება დაიმსახურეს. ეს არე ნაჩვენებია ნახაზზე 6.8. მათი პროპორციის საპოვნელად Z ქულა 1.45-ის გასწვრივ (გ) სვეტში ვიპოვოთ 0.0735. ეს ნიშნავს, რომ სტუდენტთა 7.35%-მა (0.0735×100=7.35%) 85 ქულაზე მაღალი შეფასება მიიღო. სტუდენტების ფაქტობრივი რიცხვის საპოვნელად, 0.0735 სტუდენტთა საერთო რიცხვზე უნდა გავამრავლოთ 1,200×0.0735=88. ამრიგად, 10 წლის განმავლობაში 88 სტუდენტმა დაიმსახურა 85 ქულაზე მაღალი შეფასება.

ასეთივე მოქმედებები უნდა შევასრულოთ დაბალი შეფასების მქონე სტუდენტების რიცხვის განსაზღვრისათვის. ათვლის წერტილად მივიჩნიოთ 50 ქულა. ჯერ 50 გამოვსახოთ Z ქულით:

$$Z = \frac{50 - 70.07}{10.27} = -1.95$$

ნახაზი 6.8 არის პოვნა დადებითი Z ქულის ზემოთ და უარყოფითი Z ქულის ქვემოთ



Z ქულის –1.95-ის ქვემოთ ის სტუდენტები მოხვდნენ, რომლებმაც 50-ზე დაბალი შეფასება მიიღეს. ეს არე აგრეთვე ნაჩვენებია ნახაზზე 6.8. Z ქულის განსაზღვრის შემდეგ მოვძებნოთ 1.95-ის შესაბამისი არე (გ) სვეტში (გაიხსენეთ, რომ დადებითი და უარყოფითი Z ქულების შესაბამისი პროპორციები იდენტურია). ეს ათწილადია 0.0256. ამიტომ ჯგუფის 2.56%-მა (0.0256×100=2.56%) ან დაახლოებით 31-მა (0.0256×1,200) სტუდენტმა მიიღო დაბალი ქულები სტატისტიკაში.

კოეფიციენტების (ან პროცენტების) გარდაქმნა Z ქულებად

ამ ნაწილში მაგალითები მოითხოვს კოეფიციენტების (ან პროცენტების) გარდაქმნას Z ქულებად.

Z ქულის პოვნა ზედა მოსაზღვრე არიდან

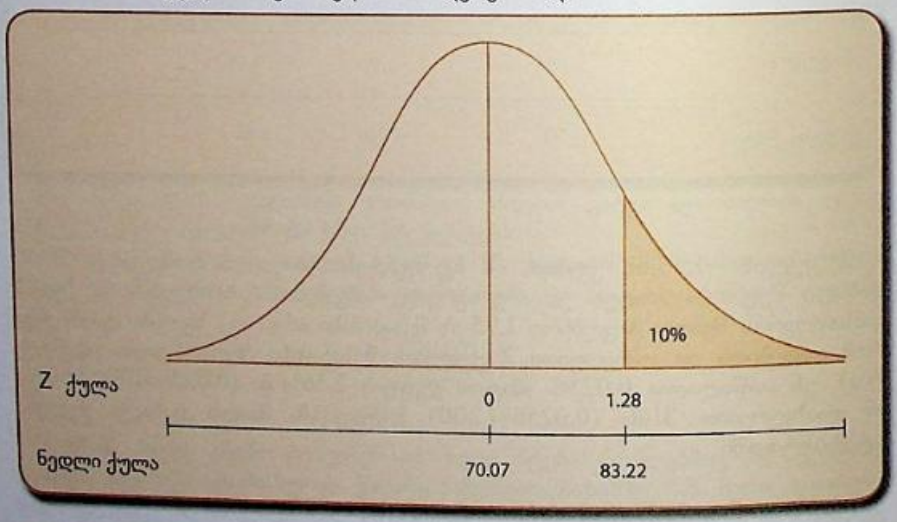
თუ დაუშვებთ, რომ A შეფასება სტუდენტების 10%-მა მიიღო, რამდენ ქულიანი სტუდენტები შევიდოდნენ ამ ჯგუფში? ამ კითხვაზე პასუხის გასაცემად გაანგარიშების 2 ეტაპია გასავლელი:

1. განვსაზღვროთ Z ქულა, რომელსაც ესაზღვრება სტატისტიკის სტუდენტების ზედა 10% ან 0.1000 (0.1000×100=10%) (ნახაზი 6.9).
2. შემდეგ დანართში ბ მოვნახოთ არეების (გ) სვეტში 0.1000-თან ყველაზე ახლოს მდგომი სიდიდე 0.1003 და ამ ათწილადის შესაბამისი Z ქულა (ა) სვეტში. როგორც აღმოჩნდა, ეს არის 1.28.
3. ვიპოვოთ ფინალური ქულა, რომელიც შეეესაბამება 1.28-ს. ამისთვის Z ქულა უნდა გარდავქმნათ ნედლ ქულად. როგორც უკვე ვისწავლეთ (ფორმულა 6.2), ამისთვის Z ქულა უნდა გავამრავლოთ სტანდარტულ გადახრაზე და დავუმატოთ საშუალოს. აქედან,

$$Y=70.07+1.28(10.27)=70.07+13.15=83.22$$

როგორც ჩანს, კლასის ზედა 10%-ის ათვლის წერტილი იწყება 83.22 ქულიდან.

ნახაზი 6.9 Z ქულის პოვნა ზედა მოსაზღვრე არიდან



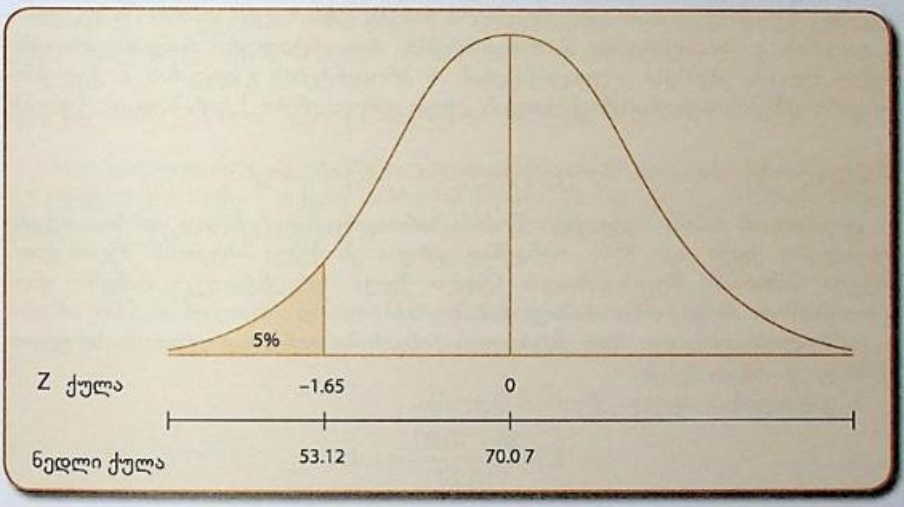
Z ქულის პოვნა ქვედა მოსაზღვრე არიდან

ახლა დაუშვათ, რომ *F* შეფასება კლასის 5%-მა მიიღო. რა იქნება ათვლის

წერტილი დაბალი ქულებისთვის? ამ კითხვაზე პასუხის გასაცემადაც გაანგარიშების 2 ეტაპია გასავლელი:

1. განვსაზღვროთ Z ქულა, რომელსაც ესაზღვრება სტატისტიკის სტუდენტების ქვედა 5% ან 0.0500 (0.0500×100=5%) (ნახაზი 6.10).

ნახაზი 6.10 Z ქულის პოვნა ქვედა მოსაზღვრე არიდან



2. შემდეგ დანართში (B) მოვნახოთ არეების (გ) სვეტში 0.0500-თან ყველაზე ახლოს მდგომი სიდიდე 0.0495 და ამ ათწილადის შესაბამისი Z ქულა (ა) სვეტში. ეს არის 1.65. რადგან ჩვენ ვეძებთ საშუალოს მარცხნივ მდებარე არის პროპორციას, მისი შესაბამისი Z ქულა უარყოფითი იქნება. ამიტომ, Z ქულა 0.0500-ის ან 0.0495-ის შესაბამისი იქნება –1.65.

3. ახლა ვიპოვოთ ფინალური ქულა, რომელიც შეესაბამება –1.65-ს. ამისთვის Z ქულა უნდა გარდავქმნათ ნედლ ქულად:

$$Y=70.07+(-1.65)(10.27)=70.07-16.95=53.12$$

დაბალი ქულები იწყება 53.12 ქულიდან.

იპოვეთ სტუდენტების რაოდენობა, რომლებმაც სტატისტიკის გამოცდაზე სულ მცირე 90 ქულა მიიღეს. რამდენმა სტუდენტმა მიიღო 60-ზე დაბალი ქულა?

✓ ცოდნის შემოწმება

პერცენტილების გამოყენება ნორმალურ განაწილებაში

მეოთხე თავში პერცენტილები განვსაზღვრეთ, როგორც მნიშვნელობები, რომლის ქვემოთ განაწილების გარკვეული პროცენტი ხვდებოდა. მაგალითად, 95-ე პერცენტილი არის მნიშვნელობა, რომელიც განაწილებას ისე ყოფს, რომ დაკვირვებათა 95% მის ქვემოთ ხვდება, ხოლო 5% – მის ზემოთ. როგორ ხდება პერცენტილური რანგის განსაზღვრა? როგორ უნდა გარდავქმნათ პერცენტილური რანგი ნედლ ქულად? ნედლი ქულიდან პერცენტილური რანგის განსაზღვრა მოითხოვს Z ქულის გადაყვანას კოეფიციენტებში ან პროცენტებში. პერცენტილური რანგის გარდაქმნა ნედლ ქულაში ემყარება კოეფიციენტების ან პროცენტების გადაყვანას Z ქულებში. ამ გარდაქმნების თვალსაჩინოებისათვის კვლავ გამოვიყენებთ სტატისტიკის ქულებს.

საშუალოზე მაღალი პერცენტილური რანგის პოვნა

დაუშვათ, ამ 1,200 სტუდენტს შორის ხართ, ვინც სტატისტიკა აირჩია. თქვენი ფინალური ქულა იყო 85%. რამდენად კარგია ეს ქულა სხვებთან შედარებით? თქვენი ნაშრომის შეფასებისთვის ნედლი ქულა პერცენტილურ რანგში უნდა გადაიყვანოთ. ამ საკითხს ასახავს ნახაზი 6.11.

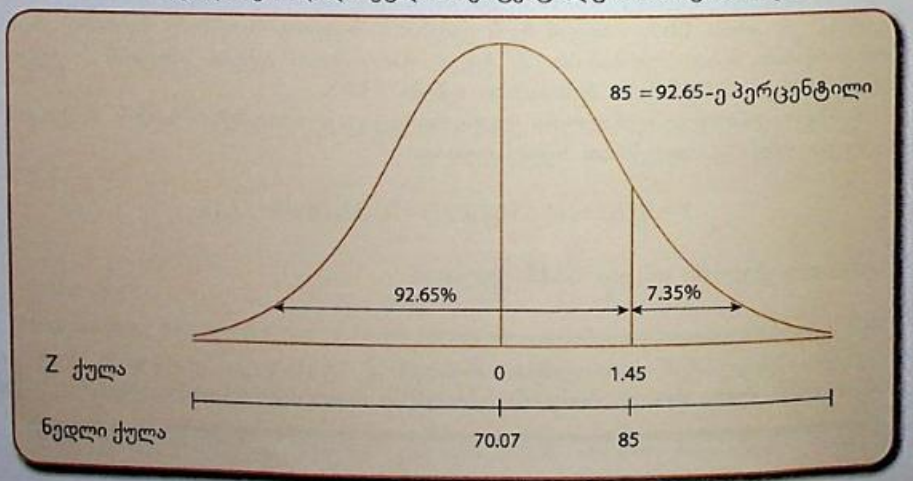
საშუალოზე მაღალი პერცენტილური რანგის საპოვნელად უნდა შეასრულოთ შემდეგი გაანგარიშებები:

1. გადაიყვანეთ ნედლი ქულა Z ქულაში:

$$Z = \frac{85 - 70.07}{10.27} = 1.45$$

ნედლ ქულას 85 შეესაბამება Z ქულა 1.45.

ნახაზი 6.11 საშუალოზე მაღალი ქულის პერცენტილური რანგის პოვნა



2. იპოვეთ Z-ის ზედა არე დანართში ბ, სვეტში (გ). 1.45-ის შესაბამისი არეა 0.0735.

3. პერცენტილური რანგის მისაღებად, გამოაკელით 1-ს არე 0.0735 და გაამრავლეთ 100-ზე:

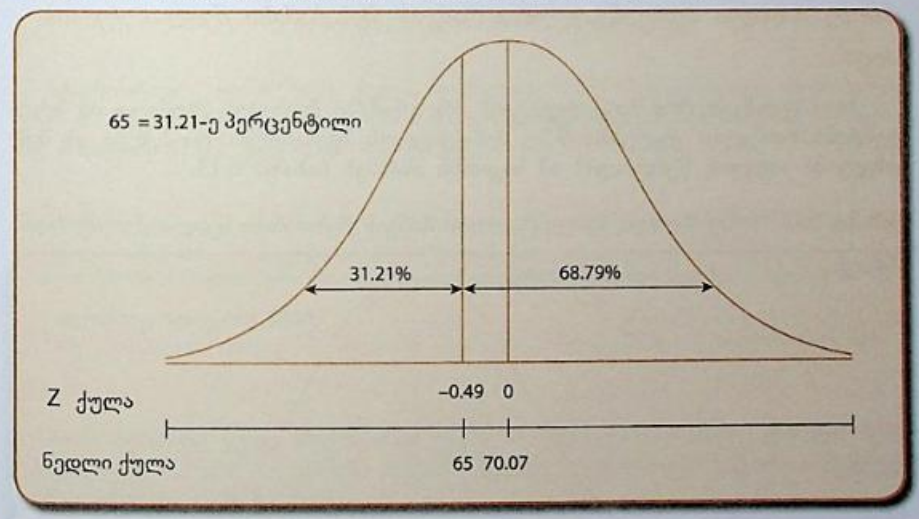
პერცენტილური რანგი = (1.0000-0.0735 = 0.9265) × 100 = 92.65%

92.65-ე პერცენტილი ნიშნავს, რომ სტუდენტების 92.65%-მა მიიღო 85 ქულაზე დაბალი შეფასება, ხოლო 7.35%-მა მასზე მაღალი.

საშუალოზე დაბალი პერცენტილური რანგის პოვნა

ახლა დაუშვათ, რომ არ გაგიმართლათ და 65 ქულა მიიღეთ. რას უდრის თქვენი პერცენტილური რანგი? თქვენი ნაშრომის შესაფასებლად კვლავ მოგიწევთ ნედლი ქულების გადაყვანა პერცენტილურ რანგებში. ამ საკითხს ასახავს ნახაზი 6.12.

ნახაზი 6.12 საშუალოზე დაბალი პერცენტილური რანგის პოვნა



საშუალოზე დაბალი პერცენტილური რანგის საპოვნელად უნდა შეასრულოთ შემდეგი გაანგარიშებები:

1. გადაიყვანეთ ნედლი ქულა Z ქულაში:

$$Z = \frac{65 - 70.07}{10.27} = 0.49$$

ნედლ ქულას 65 შეესაბამება Z ქულა −0.49.

2. იპოვეთ Z-ის ქვედა არე დანართში ბ, სვეტში (გ). −0.49-ის შესაბამისი არეა 0.3121.

3. პერცენტილური რანგის მისაღებად, ეს არე 0.3121 გაამრავლეთ 100-ზე:

პერცენტილური რანგი = 0.3121 × 100 = 31.21%

31.21-ე პერცენტილური რანგი ნიშნავს, რომ სტუდენტების 31.21%-მა თქვენზე უარესი ქულები მიიღო, ხოლო 68.79%-მა − თქვენზე კარგი.

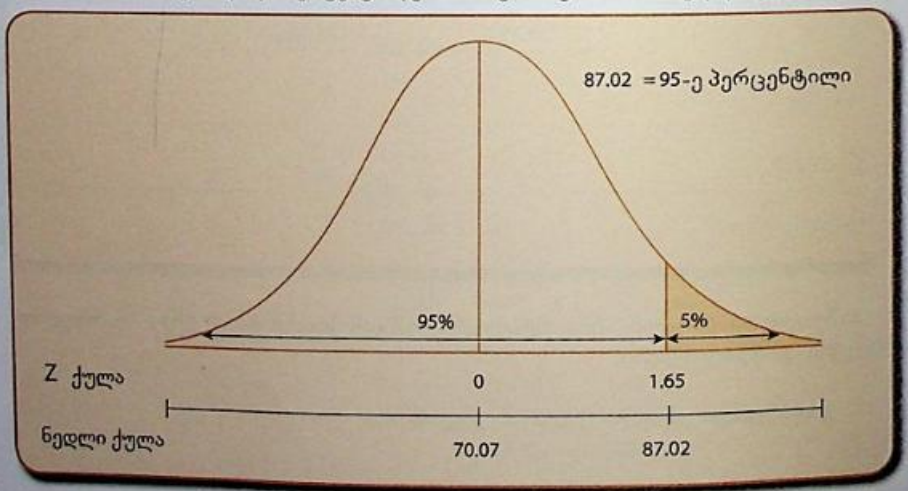
✓ **ცოდნის შემოწმება**

თავში 4 ჩვენ ვისწავლეთ პერცენტილების განსაზღვრა განაწილების კუმულაციური ხვედრითი წილებით. ნახეთ ცხრილი 6.1 და იპოვეთ 92-ე პერცენტილი. განსხვავდება თუ არა თქვენი პასუხი ადრინდელისგან (საშუალოზე მაღალი პერცენტილური რანგის პოვნა)? თუ თანახმა ხართ, ახსენით რატომ.

50-ზე მაღალი პერცენტილური რანგის შესაბამისი ნედლი ქულის პოვნა

ეხლა დაუშვათ, რომ სოციოლოგიის პროგრამაზე მიიღებენ მხოლოდ იმ სტუდენტებს, რომელთა ქულებიც 95-ე პერცენტილს შეესაბამება. რა იქნება ამ შემთხვევაში ათვლის წერტილი? ამ საკითხს ასახავს ნახაზი 6.13.

ნახაზი 6.13 50-ზე მაღალი პერცენტილური რანგის შესაბამისი ნედლი ქულის პოვნა



50-ზე მაღალი პერცენტილური რანგის შესაბამისი ნედლი ქულის პოვნისათვის საჭიროა შემდეგი გაანგარიშებები:

1. პერცენტილური რანგის ქვედა არის საპოვნელად გაყავით პერცენტილური რანგი 100-ზე:

$$\frac{95}{100} = 0.95$$

2. პერცენტილური რანგის ზედა არის საპოვნელად 1.00-ს გამოაკელით პერცენტილური რანგის ქვედა არე:

$$1.00-0.95=0.05$$

3. იპოვეთ პერცენტილური რანგის ზედა არის შესაბამისი Z ქულა.
დანართში B, სვეტში (გ) მოძებნეთ 0.0500 (ან მასთან ყველაზე ახლოს მდგომი რიცხვი). ასეთია 0.0495. მოძებნეთ მისი შესაბამის Z ქულა (ა) სვეტში. ეს არის 1.65.
4. გადაიყვანეთ Z ქულა ნედლ ქულაში:

$$Y=70.07+1.65(10.27)=70.07+16.95=87.02$$

სტატისტიკის ფინალური ქულა, რომელიც 95-ე პერცენტილს შეესაბამება, არის 87.02. ეს ნიშნავს, რომ სოციოლოგიის პროგრამაზე შესვლა ნებადართულია, თუ მიიღებთ 87.02 ქულას ან მეტს.

რამდენი სტანდარტული გადახრა შეესაბამება 95-ე პერცენტილს ნორმალურ განაწილებაში? თუ არ გაქვთ პასუხი, თავიდან გადაათვალიერეთ ეს ნაწილი.

✓ *ცოდნი*
შემოწმება

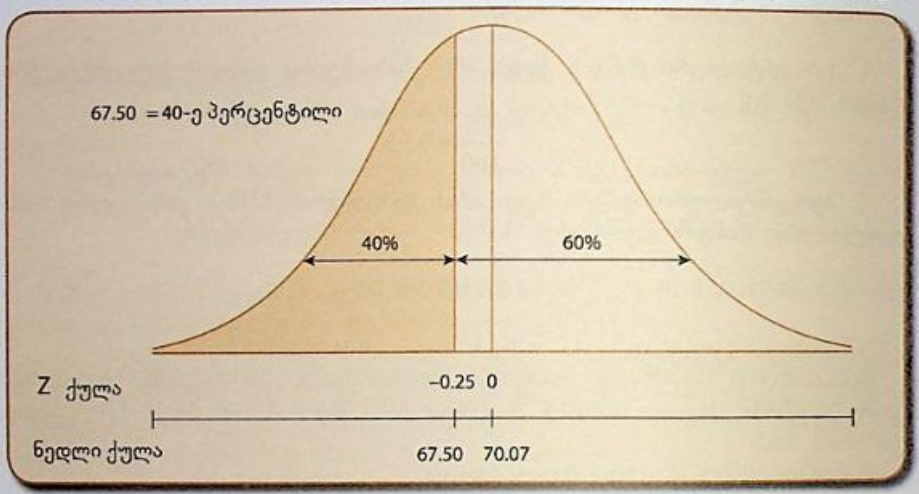
50-ზე დაბალი პერცენტილური რანგის შესაბამისი ნედლი ქულის პოვნა

და ბოლოს, რა ქულა შეესაბამება მე-40 პერცენტილს? 50-ზე დაბალი პერცენტილური რანგის შესაბამისი ნედლი ქულის პოვნისათვის საჭიროა შემდეგი გაანგარიშებები (ნახაზი 6.14):

1. პერცენტილური რანგის ქვედა არის საპოვნელად პერცენტილური რანგი გაყავით 100-ზე:

$$\frac{40}{100} = 0.40$$

ნახაზი 6.14 50-ზე დაბალი პერცენტილური რანგის შესაბამისი ნედლი ქულის პოვნა



2. იპოვეთ ამ არის შესაბამისი Z ქულა.

დანართში B, სვეტში (გ) მოძებნეთ 0.4000 (ან მასთან ყველაზე ახლოს მდგომი რიცხვი). ასეთია 0.4013. მოძებნეთ მისი შესაბამისი Z ქულა (ა) სვეტში. ეს არის –0.25.

3. გადაიყვანეთ Z ქულა ნედლ ქულაში:

$$Y=70.07+(-0.25)(10.27)=70.07-2.568=67.50$$

მე-40 პერცენტილს შეესაბამება ფინალური ქულა 67.50. ეს ნიშნავს, რომ სტუდენტების 40%-მა მიიღო 67.50-ზე დაბალი ქულა, ხოლო 60%-მა – მაღალი.

✓ *ცოდნის შემოწმება*

რომელი ნედლი ქულა შეესაბამება 50-ე პერცენტილს?

ბოლო შენიშვნა

ამ თავში განვიხილეთ თეორიული ნორმალური განაწილების მახასიათებლები, რომლებიც გვეხმარება ნორმალურთან მიახლოებული ემპირიული განაწილების აღწერაში. ნორმალურ მრუდს სხვა პრაქტიკული გამოყენებაც აქვს. კერძოდ, შემდეგ თავებში დავინახავთ, რომ ნორმალური მრუდის დახმარებით აღვწერთ დასკვნითი სტატისტიკის უმნიშვნელოვანეს თეორიულ განაწილებას: შერჩევით განაწილებას. ამ თავში ნასწავლი მეთოდები – ქულების გარდაქმნა და ნორმალური მრუდის ქვეშ არეების პოვნა – შემდგომ თავებში გამოგვადგება, ამიტომ დარწმუნდით, რომ კარგად გაითავისეთ ეს მასალა.

ძირითადი განმარტებები:

• ნორმალური განაწილება ამოსავალია დასკვნითი სტატისტიკის თეორიაში. ის მოდელად გამოიყენება ნორმალურთან მიახლოებული ემპირიული განაწილებებისთვის.

• ნორმალურ ან ნორმალურთან მიახლოებულ განაწილებებში მუდმივი პროპორცია არსებობს მრუდის ქვეშ საშუალოსა და საშუალოდან რაიმე მანძილით დაშორებულ არეებს შორის, თუ ეს მანძილი სტანდარტული გადახრის ერთეულებით იზომება.

• სტანდარტული ნორმალური განაწილება ეს არის სტანდარტული ქულებით, ანუ Z ქულებით წარმოდგენილი განაწილება, სადაც საშუალო 0-ის ტოლია, ხოლო სტანდარტული გადახრა – 1-ის. Z ქულა გამოსახავს სტანდარტული გადახრის რიცხვს, რომელიც გვიჩვენებს, თუ რამდენად ზემოთ ან ქვემოთ მდებარეობს ეს ქულა საშუალოდან. Z ქულის, ან მისი წილადების შესაბამისი პროპორციები თავმოყრილია სპეციალურ ცხრილში, რომელსაც სტანდარტული ნორმალური ცხრილი ეწოდება.

საკვანძო ტერმინები

ნორმალური განაწილება
სტანდარტული ნორმალური განაწილება
სტანდარტული ნორმალური ცხრილი
სტანდარტული (Z) ქულა

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

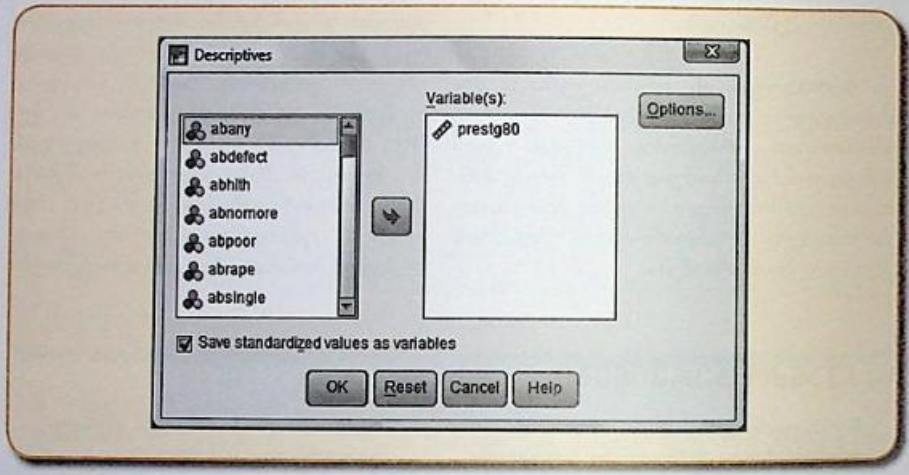
[GSS08PFP-A]

Z ქულების შექმნა SPSS-ის საშუალებით

ამ თავში ჩვენ განვიხილეთ თეორიული ნორმალური მრუდი, Z ქულები და დამოკიდებულება ნედლ და Z ქულებს შორის. SPSS-ის Descriptives პროცედურებით შესაძლებელია ნებისმიერი განაწილების Z ქულების პოვნა. მონაცემებს ავიღებთ GSS 2008-ის ფაილიდან საქმიანობის პრესტიჟულობასთან დაკავშირებით. დაკლიკეთ Analyze მენიუ, ქვემოთ Descriptive Statistics, შემდეგ Descriptives. ჩვენ შეგვიძლია ერთი ან მეტი ცვლადის მოთავსება Variable(s) box-ში; ამჯერად ამ ყუთში მოვათავსოთ ცვლადი PRESTG80 (occupational prestige score) (ნახაზი 6.15).

ჰქვ, რომელიც ყუთის მარცხენა ქვედა ნაწილშია მოთავსებული, SPSS-ს აძლევს ბრძანებას შექმნას ახალი სტანდარტიზებული ცვლადები, ანუ Z ქულები. ნებისმიერი ახალი ცვლადი თავსდება მონაცემთა ფანჯრის Data View ახალ სვეტში და შესაძლებელია მისი შემდგომი ანალიზი. დაკლიკეთ OK, ბრძანების შესასრულებლად.

ნახაზი 6.15 დიალოგური ფანჯარა Descriptives



შედეგების ფანჯარაში გამოტანილია Descriptives (ნახაზი 6.16) შემოკლებული სახით. მოცემულია საშუალო, სტანდარტული გადახრა, მინიმუმი, მაქსიმუმი და ვალიდური დაკვირვებების რაოდენობა ცვლადისთვის PRESTG80.

ნახაზი 6.15 საქმიანობის პრესტიჟულობის აღწერითი მონაცემები Descriptives

	N	Minimum	Maximum	Mean	Std. Deviation
RS OCCUPATIONAL PRESTIGE SCORE (1980)	1425	17	86	43.98	14.052
Valid N (listwise)	1425				

მიუხედავად იმისა, რომ შედეგების ფანჯარაში Output window არ არის ნაჩვენები, SPSS-მა შექმნა ახალი Z ქულიანი ცვლადი PRESTG80-ისთვის. ამ ცვლადის სანახავად გადადით ეკრანზე Data View. შემდეგ გასრიალდით მარჯვნივ, სულ ბოლო სვეტამდე (ნახაზი 6.17). დათქმის მიხედვით, SPSS ცვლადის სახელს ამატებს Z-ს და ამიტომ ახალ ცვლადს ჰქვია ZPRESTG80. პირველი დაკვირვების (case) Z ქულა არის –1.5642, რაც იმას ნიშნავს, რომ მოცემული პიროვნების პრესტიჟის ქულა საშუალოზე (43.98) დაბალია. თუ ჩვენ თანმიმდევრობით განვათავსებთ რესპონდენტის PRESTG80 ქულას, ვნახავთ, რომ ამ პიროვნების ქულა 22 (არ არის ასახული ნახაზზე) ქულით დაბლაა საშუალოზე, როგორც მოველოდით.

ნახაზი 6.17 Z ქულების შექმნა ცვლადისთვის PRESTG80

	speduc	sppres80	tvhours	inequal5	Zprestg80
1	97	0	8	0	-1.56420
2	19	43	1	0	1.49593
3	97	0	-1	4	-.56788
4	16	64	2	0	1.92293
5	97	0	-1	8	-1.06604
6	97	0	2	0	1.49593
7	97	0	-1	8	.49961
8	16	58	3	0	1.78059
9	97	0	-1	5	1.06894
10	99	0	-1	5	1.56710

თუ მონაცემთა ფაილს შეინახავთ, Z ქულაც შეინახება ამოსავალ მონაცემებთან ერთად და შემდგომშიც შეგიძლიათ მისი გამოყენება ანალიზში. ამასთან, თუ SPSS-ს მივცემთ ბრძანებას და გავაანგარიშებინებთ ZPRESTG80 ცვლადისთვის საშუალოს და სტანდარტულ გადახრას, ვნახავთ რომ ისინი შესაბამისად 0 და 1.00-ის ტოლი იქნება.

ამოცანები SPSS-ისთვის

[GSS08PFP-A]

1. სოციალურ მეცნიერებებში ცვლადების უმრავლესობა ნორმალურად არ არის განაწილებული. როდესაც კვლევის მიზანი საშუალოსა და სტანდარტული გადახრის გაანგარიშებაა, ან თუ შერჩევის ზომა 50-ს სცილდება (აღნიშნული განხილული იქნება შემდეგ თავებში), ეს პრობლემას არ ქმნის. მიუხედავად ამისა, თუ ვიკვლევთ ერთ შერჩევას ან მთლიან ერთობლიობას, არანორმალურმა განაწილებამ შეიძლება პრობლემები შეგვიქმნას. აღნიშნულის ჩვენება შესაძლებელია GSS-ის მონაცემთა ბაზიდან.

ა. შექმენით ჰისტოგრამა ცვლადისთვის AGE (დაკლიკეთ Graphs, Legacy Dialogs, Histogram; შეიტანეთ ცვლადი age), მონიშნეთ ნორმალური მრუდი (დაკლიკეთ ოფცია Display Normal Curve). როგორ არის ცვლადის AGE მრუდი გადახრილი ნორმალური მრუდიდან?

ბ. გაიანგარიშეთ AGE-ისთვის საშუალო და სტანდარტული გადახრა ამ შერჩევით ერთობლიობაში, გამოიყენეთ Frequencies ან Descriptives პროცედურები.

გ. AGE -ის განაწილება მიიჩნიეთ ნორმალურად და გაიანგარიშეთ 25 წლისა და უფრო დაბალი ასაკის მოსახლეობის რაოდენობა.

დ. გამოიყენეთ Frequencies პროცედურა და ააგეთ დაკვირვებათა ხვედრითი წილის ცხრილი ასაკის თითოეულ სიდიდეში. შეადარეთ (გ) კითხვის თეორიული გაანგარიშებები ასაკის ფაქტობრივ განაწილებას. რამდენი პროცენტია 25 წელზე მაღალი ან დაბალი ასაკის? ახლოსაა თუ არა ეს სიდიდე თქვენ მიერ გაანგარიშებულთან? რატომ შეიძლება იყოს აქ განსხვავება?

2. SPSS-ს შეუძლია სტანდარტული ქულების გაანგარიშება ნებისმიერი განაწილებისთვის. შეამოწმეთ ცვლადის EDUC (years of school completed) განაწილება.
ა. SPSS-ის საშუალებით გაიანგარიშეთ Z ქულები EDUC-ისთვის.
ბ. რომელია ეკვივალენტური Z ქულა მათთვის, ვინც მიიღო 18-წლიანი განათლება?
გ. გამოიყენეთ Frequencies პროცედურა 18-ის პერცენტილური რანგის საპოვნელად.
დ. შეესაბამება თუ არა თქვენ მიერ გაანგარიშებული პერცენტილური რანგი Z ქულას სიდიდე 18-ისთვის? სხვა სიტყვებით, არის თუ არა წლების განაწილება ნორმალური? თუ ასეა, მაშინ Z ქულა, რომელსაც SPSS გაიანგარიშებს, ძალიან ახლოს უნდა იყოს ამავე ქულის პერცენტილურ რანგთან შესაბამის არეში გადაყვანის შემდეგ.
ე. შექმენით ჰისტოგრამები ცვლადებისთვის EDUC და ZEDUC. ახსენით, რატომ აქვთ მათ ერთნაირი ფორმა.
3. გაიმეორეთ 2-ე სავარჯიშოს პროცედურები. ამ შემთხვევაში ცვლადის EDUC ბაზაზე დამოუკიდებელი ანალიზი გააკეთეთ ქალებისა და მამაკაცებისთვის (SEX) და თეთრკანიანებისა და შავკანიანებისთვის (RACECEN1). განცალკევებული ანალიზისთვის უნდა გამოიყენოთ ბრძანება Data Split File. დაკლიკეთ Data, Split File, Organize Output by Groups და შეარჩიეთ SEX ან RACECEN1. არის თუ არა განსხვავება ცვლადში EDUC ქალებსა და მამაკაცებს, ან თეთრკანიანებსა და შავკანიანებს შორის? როგორ დაახასიათებთ ცვლადის EDUC განაწილებას ოთხი ჯგუფისთვის?

თავის სავარჯიშოები

1. ჩვენ აღმოვაჩინეთ, რომ 2008 წელს GSS-ის 978 რესპონდენტი ყოველდღიურად ტელევიზორს უყურებდა საშუალოდ 2.95 საათის განმავლობაში, 2.58 საათი სტანდარტული გადახრით. გაეცით პასუხი შემდეგ კითხვებს, თუ მიიჩნევთ, რომ საათების განაწილება ნორმალურია:
ა. რომელი Z ქულა შეესაბამება პირს, რომელიც ტელევიზორს უყურებს 8 საათზე მეტ ხანს?
ბ. ადამიანების რა ნაწილი უყურებს ტელევიზორს 5 საათზე ნაკლებს ყოველდღიურად? რამდენი ადამიანია ასეთი შერჩევაში?
გ. საათების რა ოდენობა შეესაბამება Z ქულას +1?
დ. ადამიანების რამდენი პროცენტი უყურებს ტელევიზორს ყოველდღე 1-დან 6 საათამდე?
2. დაუშვათ, განაწილებაში, რომელსაც თქვენ იკვლევთ და ის არ არის ნორმალურად განაწილებული, სიძნელეები წარმოიშვა განაწილების ქვეშ არეების განსაზღვრასა და ნედლი ქულების გადაყვანისას Z ქულებში. არის თუ არა ეს დებულება სწორი? რატომ დიახ, ან არა?
3. დაუშვათ, რომ განათლება ნორმალურად არის განაწილებული. GSS-ის მონაცემებიდან ვნახეთ, რომ განათლების საშუალო ხანგრძლივობა 13.44 წელია, 3.1 სტანდარტული გადახრით. გამოკვლევაში ჩართული იყო 1,497 ადამიანი. გამოიყენეთ ეს რიცხვები და უპასუხეთ შემდეგ კითხვებს:
ა. რას უდრის თქვენი Z ქულა, თუ განათლების საშუალო ხანგრძლივობა 13.44 წელია?
ბ. საშუალოდ რამდენი წელი სწავლობდა თქვენი მეგობარი, თუ ის 95-ე პერცენტილში მოხვდა?
გ. რამდენი ადამიანი მოხვდება თქვენი და თქვენი მეგობრის განათლების საშუალო ხანგრძლივობებს შორის?
4. კრიმინოლოგმა რეციდივიზმის დონის გასაზომად ჩაატარა ტესტი, სადაც დაბალი ქულები მიანიშნებდა არაკანონიერი საქციელის განმეორების მცირე ალბათობაზე. ტესტის საშუალო ქულაა 140 და სტანდარტული გადახრა 40.
ა. რას უდრის 172 ქულის პერცენტილური რანგი?
ბ. რას უდრის 200 ქულის შესაბამისი Z ქულა?
გ. ქულების რამდენი პროცენტი მოსვდება 100-სა და 160-ს შორის?
დ. რესპონდენტების რა ნაწილი მიიღებს 190-ზე მაღალ ქულებს?
ე. დაუშვათ, პიროვნებას ამ ტესტში 67-ე პერცენტილი უკავია. რა არის მისი შესაბამისი რეციდივიზმის ქულა?

5. GSS ატარებს კვლევას დაბალი, მუშათა, საშუალო და მაღალი კლასების განათლების საშუალო დონის შესახებ სტანდარტულ გადახრასთან ერთად.

	საშუალო	სტანდარტული გადახრა	N
დაბალი კლასი	11.36	2.96	121
მუშათა კლასი	12.73	2.79	676
საშუალო კლასი	14.40	3.04	636
მაღალი კლასი	15.49	2.95	53

ა. თუ დაუშვებთ, რომ განაწილება ნორმალურია, მუშათა კლასის რა ნაწილს აქვს განათლება 12-დან 16 წლამდე? მაღალი კლასის რა ნაწილს აქვს განათლება 12-დან 16 წლამდე?
ბ. როგორია იმის ალბათობა, რომ მუშათა კლასიდან შემთხვევით შერჩეული პირის განათლება 16 წელზე მეტი იქნება? საშუალო ფენიდან?
გ. როგორია იმის ალბათობა, რომ დაბალი და მაღალი კლასიდან შემთხვევით შერჩეული პირის განათლება 16 წელზე მეტი იქნება?
დ. იპოვეთ ზედა და ქვედა ზღვრები, რომელშიც მუშათა კლასიდან რესპონდენტების 50% გაერთიანდება.
ე. როგორ შეიცვლება დანარჩენი პასუხები, თუ განათლება დადებითად არის გადახრილი?

6. ქვემოთ ცხრილში მოტანილია 2009 წელს უმუშევრების განაწილება აშშ-ში ყველა შტატისა და კოლუმბიის ოლქის მიხედვით. უმუშევართა რიცხოვნობა გამოსახულია ათას კაცებში. როდესაც ეს ციფრები გამოქვეყნდა, ალაბამაში უმუშევართა რიცხოვნობამ 212,000-ს მიაღწია.

უმუშევრობა (ათასი კაცი) აშშ-ში: 2009					
ალაბამა	212	კენტუკი	218	ჩრდ. დაკოტა	16
ალასკა	29	ლუიზიანა	141	ოჰაიო	611
არიზონა	284	მეინი	57	ოკლაჰომა	114
არკანზასი	100	მერილენდი	209	ორეგონი	217
კალიფორნია	2,086	მასაჩუსეტსი	293	პენსილვანია	519
კოლორადო	208	მიჩიგანი	665	როდ-აილენდი	64
კონექტიკუტი	156	მინესოტა	236	სამხ. კაროლინა	255
დელავერი	35	მისისიპი	123	სამხ. დაკოტა	21
კოლუმბიის ოლქი	34	მისური	283	ტენესი	317
ფლორიდა	966	მონტანა	31	ტეხასი	911
ჯორჯია	457	ნებრასკა	45	იუტა	90
ჰავაი	43	ნევადა	161	ვერმონტი	25
აიდაჰო	60	ნიუ-ჰემპშირი	47	ვირჯინია	278
ილინოისი	665	ნიუ-ჯერსი	418	ვაშინგტონი	314
ინდიანა	320	ნიუ-მექსიკო	69	დას. ვირჯინია	63
აიოვა	100	ნიუ-იორკი	813	ვისკონსინი	262
კანზასი	102	ჩრდ. კაროლინა	484	ვაიომინგი	19

წყარო: U.S. Bureau of Labor Statistics, New Release, March 3, 2010, USDL-10-0231, Table 1.

ა. რას უდრის საშუალო სიდიდე და სტანდარტული გადახრა ყველა შტატისთვის?
ბ. (ა) კითხვის პასუხებზე დაყრდნობით, რამდენი შტატი მოხვდა საშუალოდან 1 სტანდარტული გადახრით ზემოთ? ეს როგორი რიცხვია თეორიული ნორმალური მრუდის განაწილებიდან მოსალოდნელთან შედარებით? რის შეთავაზება შეგიძლიათ იმისათვის, რომ ამ შტატებს სიძნელეები არ შეექმნას გაბრდილ უმუშევრობასთან დაკავშირებით?
გ. ააგეთ ჰისტოგრამა 50 შტატისა და კოლუმბიის ოლქის უმუშევართა რაოდენობების მიხედვით. ნორმალურია თუ არა ეს განაწილება? პასუხი დაასაბუთეთ.
7. ცხრილში მოტანილია ინფორმაცია თეთრკანიანებისა და შავკანიანების საქმიანობის პრესტიჟულობის შესახებ.

	საშუალო	სტანდარტული გადახრა	N
შავკანიანები	40.75	13.63	190
თეთრკანიანები	44.70	14.03	1094

წყარო: General Social Survey, 2008.

ა. თეთრკანიანების რამდენ პროცენტს ექნება პრესტიჟულობის ქულა 60-ზე მაღალი? რამდენ თეთრკანიანს ექნება პრესტიჟულობის ქულა 60-ზე მაღალი?
ბ. შავკანიანების რამდენ პროცენტს ექნება პრესტიჟულობის ქულა 60-ზე მაღალი? რამდენ შავკანიანს ექნება პრესტიჟულობის ქულა 60-ზე მაღალი?
გ. თეთრკნიანების რა ნაწილს ექნება პრესტიჟულობის ქულა 30-დან 70-მდე? რამდენ თეთრკანიანს ექნება პრესტიჟულობის ქულა 30-დან 70-მდე?
დ. რამდენ შავკანიანს ექნება პრესტიჟულობის ქულა 30-დან 70-მდე?
8. SAT-ის ქულები ისეა ნორმირებული, რომ მათემატიკის ან ვერბალური ტესტების საშუალო 500 უნდა იყოს, სტანდარტული გადახრა კი – 100. დაუშვათ, რომ ეს მართალია (სინამდვილეში ეს მხოლოდ მიახლოებით არის ნამდვილი, რადგან ქულები SAT-ის შემოღების შემდეგ შემცირდა.) გაეცით პასუხი შემდეგ კითხვებს:
ა. სტუდენტების რამდენი პროცენტი იღებს ყოველწლიურად მათემატიკაში 625 ქულაზე მაღალს?
ბ. სტუდენტების რამდენი პროცენტი იღებს 400-დან 600 ქულამდე ვერბალურ ნაწილში?
გ. კოლეჯში მისაღები პოლიტიკის ლიბერალიზაციის მიზნით პირველ ეტაპზე გადაწყდა გამოეთიშათ ის სტუდენტები, რომლებიც ვერბალურ ნაწილში 20-ე პერცენტილის დაბლა მოხვდნენ. გადაიყვანეთ პერცენტილი Z ქულაში. შემდეგ გაიანგარიშეთ მისი ეკვივალენტური SAT-ის ვერბალური ტესტის ქულა.
9. 2007 წელს მონაცემები სისხლის სამართლის დანაშაულზე შეგროვდა 300 შემთხვევაზე დაყრდნობით. თითოეული დანაშაულის მიხედვით მსხვერპლის საშუალო სიდიდემ 1.28 შეადგინა, ხოლო სტანდარტულმა გადახრამ – 0.82. დაუშვათ, რომ მსხვერპლის განაწილება ნორმალურია, გაეცით პასუხი შემდეგ კითხვებს:
ა. დანაშაულების რა ნაწილს მოჰყვა 2-ზე მეტი მსხვერპლი?
ბ. როგორია იმის ალბათობა, რომ შემთხვევას 1-ზე მეტი მსხვერპლი მოჰყვება?
გ. დანაშაულების რა ნაწილს მოჰყვა 4-ზე ნაკლები მსხვერპლი?
10. 894 რესპონდენტის კვლევის შედეგად ნახეს, რომ წინა კვირაში ნამუშევარი საათების საშუალო რაოდენობა 42.26 იყო, ხოლო სტანდარტული გადახრა 14.34 საათი.
ა. დაუშვათ, განაწილება ნორმალურია. როგორი ალბათობაა, რომ შერჩევიდან ვინმე 60 საათზე მეტს იმუშავებდა? შერჩევიდან რამდენ ადამიანს შეეძლო ემუშავა 60 საათი და მეტი?
ბ. როგორია იმის ალბათობა, რომ შერჩევიდან ვინმე 30 საათზე ნაკლებს იმუშავებდა? შერჩევიდან რამდენ ადამიანს შეეძლო ემუშავა 30 საათზე ნაკლები?
გ. შეესაბამება თუ არა კვირაში ნამუშევარი საათების რაოდენობა 60-ე პერცენტილს?

11. 2007-2008 წლებში აკადემიური პროგრესის საშუალო დონემ 1,000-დან 945.81 ქულა შეადგინა, სტანდარტული გადახრით 50.35. შერჩევაში კალათბურთის 341 გუნდი მონაწილეობდა. დაუშვათ, რომ განაწილება ნორმალური იყო.
ა. მოხვდება თუ არა 975 ქულით გუნდი განაწილების ზედა კვარტილში (ზედა 25%)?
ბ. რა უნდა იყოს გუნდის ქულა, რომ მან აჯობოს გუნდების 75%-ს?
გ. რას უდრის ამ ქულის Z ქულა?
12. კალათბურთის ორი გუნდის მონაცემების – არჩევის უფლებისა და დამაგრების საფუძველზე განსაზღვრეთ:

	საშუალო	სტანდარტული გადახრა	A	B
არჩევის უფლება	957.19	54.32	917	962
დამაგრება	925.86	59.44	913	962

ა. რომელი კრიტერიუმით (არჩევის უფლება, დამაგრება) სჯობს A გუნდი დანარჩენებს? გაიანგარიშეთ შესაბამისი მაჩვენებლები.
ბ. რომელი კრიტერიუმით (არჩევის უფლება, დამაგრება) სჯობს B გუნდი დანარჩენებს? გაიანგარიშეთ შესაბამისი მაჩვენებლები.
გ. გუნდების რა ნაწილი ჩამორჩება B გუნდს დამაგრებით?
დ. რომელი პერცენტილური რანგი აქვს გუნდი A-ს არჩევის უფლებას?
13. რას უდრის ნორმალური სტანდარტული განაწილების საშუალო? რას უდრის ნორმალური სტანდარტული განაწილების სტანდარტული გადახრა? ახსენით.
14. ყოველდღიურად თავშესაფარში საშუალოდ ღამეს ათევს 250 ქალი, სტანდარტული გადახრით 75. ქალების განაწილება ნორმალურია.
ა. დაკმაყოფილდება თუ არა მოთხოვნა 95%-ით, თუ თავშესაფრების ტევადობა 350 ადგილით განისაზღვრება? თუ არა, რამდენი ადგილი უნდა დაემატოს?
ბ. თავშესაფრების ნაწილი დაიხურა და ახლა საწოლების რაოდენობა 220-ია. ქალების რამდენი პროცენტი დარჩება ადგილების გარეშე?
15. ამ თავში განხილულიდან გამომდინარე:
რა ნიშნები ახასიათებს ნორმალურ განაწილებას? რატომ უწოდებენ მას "ნორმალურს"? რას ნიშნავს "დადებითი" Z ქულა? "უარყოფითი" Z ქულა?

თავი 7

შერჩევა და შერჩევითი განაწილებები

თავის სწავლების მიზნები

- შერჩევის მიზნები და ალბათობის ძირითადი პრინციპები
- შერჩევის სახეები
- შერჩევითი განაწილების არსი და გამოყენება
- ცენტრალური ზღვრის თეორემის ბუნება და არსი

ჯამდე არ განგვიხილავს საკითხი იმის შესახებ, თუ რას ან ვის უნდა დავაკვირდეთ მონაცემთა შეგროვების დროს, ან როგორ უნდა გავავრცელოთ მთლიან ერთობლიობაზე შერჩევითი დაკვირვებიდან მიღებული შედეგები. საქმე იმაშია, რომ ჩვენ იშვიათად გვაქვს საშუალება შევისწავლოთ და დავაკვირდეთ ყველას ან ყველაფერს. მიუხედავად იმისა, რომ ანალიზის მრავალი მეთოდი შევისწავლეთ, უნდა გვახსოვდეს, რომ ეს მხოლოდ მცირე ნაწილია იმისა, რაც ჩვენ გვინდოდა გაგვეკეთებინა. მოვიყვანთ შემდეგ მაგალითებს.

მაგალითი 1. უნივერსიტეტის მუსულმან სტუდენტთა ასოციაციას სურს ჩაატაროს კამპუსში გამოკვლევა მრავალფეროვნების შესახებ. მას გააჩნია მხოლოდ 300 სტუდენტის გამოკვლევის საშუალება, თუმცა კამპუსში 20,000 სტუდენტია. თქვენ რას გააკეთებდით?

მაგალითი 2. გარემოს დაცვის ადგილობრივ აქტივისტებს უსარგებლო ნარჩენების შემცირების მიზნით სურთ დანერგონ ნარჩენების გადამუშავების პრაქტიკა. უნივერსიტეტი ემსახურება 30,000 სტუდენტს, ფაკულტეტის ადმინისტრაციას და აკადემიურ პერსონალს. რა უნდა მოიმოქმედონ აქტივისტებმა?

მაგალითი 3. სტუდენტების კავშირი ცდილობს დაეხმაროს ქალაქგარეთ მცხოვრებ სტუდენტებს და მათ მოთხოვნებთან დაკავშირებით დაგეგვალათ კვლე-

ვის ჩატარება. თქვენ მხოლოდ 500 სტუდენტის გამოკითხვის სახსრები გაქვთ. ამ სამუშაოს შესრულება შეუძლებელია?

რა საერთო აქვს ამ მაგალითებს? ყველა შემთხვევაში მთავარი პრობლემაა დიდი მოცულობის ინფორმაცია და რესურსების ნაკლებობა მათი შეგროვებისა და ანალიზისთვის.

შერჩევის მიზნები[1]

სოციალური სფეროს მკვლევარებს ასევე არ აქვთ საკმარისი დრო და სახსრები მთლიანი ერთობლიობიდან მათთვის საინტერესო ინფორმაციის შესაგროვებლად. **მთლიან ერთობლიობაში** შედის ყველა დაკვირვება (ქეისი - ინდივიდები, საგნები ან ჯგუფები), რისი შესწავლაც მკვლევარს აინტერესებს. პირველ მაგალითში მთლიანი ერთობლიობა არის 20,000 სტუდენტი, მეორეში – 30,000 სტუდენტი, ადმინისტრაცია და ლექტორები, მესამეში – 15,000 ქალაქგარეთ მცხოვრები სტუდენტი.

მთლიანი ერთობლიობა ჯგუფი, რომელიც მოიცავს დაკვირვებების მთლიან სიმრავლეს (ინდივიდებს, საგნებს ან ჯგუფებს), რაც მკვლევარს აინტერესებს.

თუმცა, ჩვენ შეგვიძლია ბევრი რამ გავიგოთ მთლიან ერთობლიობაზე, თუ დაკვირვებით შევარჩევთ ქვესიმრავლეს. ამ ქვესიმრავლეს **შერჩევითი ერთობლიობა** ეწოდება. *შერჩევითი ერთობლიობის* საფუძველზე ჩვენ ვცდილობთ მცირე ჯგუფიდან ან შერჩევიდან მიღებული შედეგები გავავრცელოთ მთლიან ერთობლიობაზე. ეს არის *დასკვნითი სტატისტიკის საფუძველი* – პროგნოზების ან დასკვნების ჩამოყალიბება მთლიანი ერთობლიობის შესახებ შერჩევაზე განხორციელებული დაკვირვების შედეგად.

შერჩევითი ერთობლიობა მთლიანი ერთობლიობიდან შერჩეული დაკვირვებების ქვესიმრავლე.

ტერმინი **პარამეტრიც** მთლიან ერთობლიობას უკავშირდება. ის მთლიანი ერთობლიობის განაწილების საზომს წარმოადგენს. მაგალითად, სტუდენტების გადაადგილების საშუალო დრო კამპუსში 15,000-ივე სტუდენტისათვის პარამეტრია, რადგან სტუდენტთა მთლიან ერთობლიობას ეხება. წინა თავებში განვიხილეთ განაწილების პროპორციები, საშუალო სიდიდეები და სტანდარტული გადახრა. თუ მათი საშუალებით მთლიანი ერთობლიობის განაწილებას აღვწერთ, მაშინ ყველა ის პარამეტრია.

პარამეტრი საზომი (მაგ. საშუალო ან სტანდარტული გადახრა), რომელიც მთლიან ერთობლიობას ახასიათებს.

ტერმინ **სტატისტიკას** ანალოგიური მახასიათებლებისთვის ვიყენებთ შერჩევაში. მაგალითად, სტუდენტების გადაადგილების საშუალო დრო სტუდენტების შერჩევითი ერთობლიობისთვის სტატისტიკაა. ანალოგიურად, — შერჩევის საშუალო, პროპორცია, ან სტანდარული გადახრა — სტატისტიკაა.

სტატისტიკა საზომი (მაგ. საშუალო ან სტანდარტული გადახრა), რომელიც შერჩევით ერთობლიობას ახასიათებს.

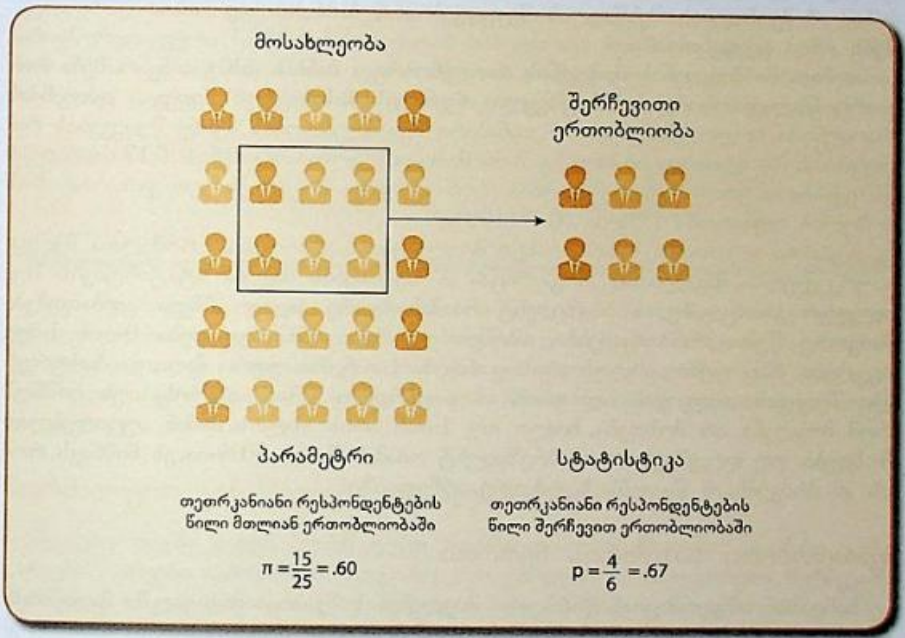
ამ თავში, ისევე როგორც მე-8 და მე-9 თავებში, ვიმსჯელებთ შერჩევითი ერთობლიობიდან მიღებული შედეგების გავრცელების შესახებ მთლიან ერთობლიობაზე და სხვადასხვა აღნიშვნებს გამოვიყენებთ პარამეტრებისა და სტატისტიკისთვის. ცხრილში 7.1 მოცემულია შერჩევის აღნიშვნები და მისი შესაბამისი მთლიანი ერთობლიობის აღნიშვნები.

განსხვავება შერჩევასა და მთლიან ერთობლიობას შორის და სტატისტიკასა და პარამეტრს შორის თვალსაჩინოდ ასახულია ნახაზზე 7.1. მთლიანი ერთობლიობისთვის აქ შევიტანეთ პარამეტრი 0.60 — თეთრკანიანი რესპონდენტების ხვედრითი წილი მთლიან ერთობლიობაში. რადგან მთლიანი ერთობლიობიდან ინფორმაციის შესაგროვებლად არასოდეს არ გვყოფნის რესურსები, ჩვენ ძალიან იშვიათად ვიცით პარამეტრის სიდიდე. ამიტომ, როგორც წესი, მთლიანი ერთობლიობის პარამეტრის მისაღებად მკვლევარები მიმართავენ შერჩევითი ერთობლიობის შესწავლას. ამის გამო, შერჩევის თეორიისა და დასკვნითი სტატისტიკის მთავარი მიზანია გაანგარიშებების საფუძველზე შერჩევითი სტატისტიკიდან ადვილად მივიღოთ მთლიანი ერთობლიობის უცნობი პარამეტრები.[2]

ცხრილი 7.1 შერჩევითი და მთლიანი ერთობლიობების აღნიშვნები

საზომის დასახელება	შერჩევითი ერთობლიობა	მთლიანი ერთობლიობა
საშუალო	$\bar{Y}$	μ
პროპორცია	P	π
სტანდარტული გადახრა	S_y	σ_y
დისპერსია	S_y^2	σ_y^2

ნახაზი 7.1 თეთრკანიანი რესპონდენტების ხვედრითი წილები მთლიან და შერჩევით ერთობლიობებში



ტერმინების გაგება, როგორიცაა მთლიანი ერთობლიობა, შერჩევითი ერთობლიობა, პარამეტრი, სტატისტიკა, ძალიან მნიშვნელოვანია. შეეცადეთ თქვენი სიტყვებით ახსნათ ისინი. თუ ვერ ახერხებთ, კვლავ დაუბრუნდით ამ მასალას. მთლიანი და შერჩევითი ერთობლიობების აღნიშვნებს შემდგომში ხშირად შეხვდებით. მათი დამახსოვრება გაგიადვილებთ დასკვნითი სტატისტიკის ფორმულების გაგებას.

✓ *ცოდნის შემოწმება*

ალბათობის ძირითადი პრინციპები

ალბათობის განსაზღვრა

ალბათობის კონცეფციას ხშირად ვიყენებთ ყოველდღიურ საუბარში. ჩვენ შეიძლება ვიკითხოთ, „როგორი ალბათობაა, რომ ხვალ იწვიმებს?“ ან „როგორია ალბათობა, რომ ტესტს კარგად დავწერთ?“ ყოველდღიურ საუბარში ამ კითხვებზე პასუხები იშვიათად არის სისტემატური, თუმცა სტატისტიკაში ალბათობას ბევრად ზუსტი მნიშვნელობა გააჩნია.

შემდეგ ნაწილებში ვიმსჯელებთ სოციალურ სტატისტიკაში მიღებულ შერჩევის მეთოდებზე. ყველა ეს მეთოდი *ალბათური შერჩევის* მიდგომას ექვემდებარება. სანამ ამ მეთოდების შესწავლას შევუდგებით, ალბათობის თეორიასა და პრინციპებს უნდა გავეცნოთ.

ალბათობა მოვლენის მოხდენის რაოდენობრივი შანსია. ის გაიანგარიშება როგორც წილადი, რომლის მრიცხველია ჩვენთვის სასურველი შედეგის დადგომის რაოდენობა, ხოლო მნიშვნელი – თანაბარი შესაძლებლობის მქონე შედეგების რაოდენობა. მაგალითად, კამათელზე 3-ის მოსვლის ალბათობა 1/6 ან 0.17-ია, რადგან ეს ექვსიდან ერთი შესაძლებლობაა. ჩვენ აგრეთვე შეგვიძლია ვთქვათ, რომ 3-ის მოსვლის ალბათობა 17%-ია (0.17×100).

ტერმინი ალბათობა ხშირად ისეთ მოვლენებთან გამოიყენება, რომელთა შედეგი გაურკვეველია, მაგალითად, როგორიცაა A შეფასების მიღება სტატისტიკის შუალედურ გამოცდაზე, ან პრეზიდენტ ობამას მეორე ვადით არჩევა. ალბათობები როგორც წესი პროპორციებით იზომება და მათი მნიშვნელობები 0-დან 1-მდე იცვლება. რაც უფრო ახლოს არის ფარდობა 1-თან, მით უფრო მაღალია სასურველი შედეგის დადგომის ალბათობა. თუ ფარდობა 0-თან ახლოსაა, ეს ნიშნავს, რომ მოვლენა არ მოხდება, ხოლო თუ 1-თან არის ახლოს, მაშინ აუცილებლად მოხდება. თუ დავუშვებთ, რომ პრეზიდენტ ობამას შანსი 50%-ია, ეს ნიშნავს, რომ ის ან მოიგებს, ან წააგებს ხელმეორე არჩევნებზე.

ფარდობითი სიხშირის მეთოდი

ხანდახან ინფორმაციას წარსული მოვლენების შესახებ მომავალში მათი მოხდენის ალბათობის განსაზღვრისთვის ვიყენებთ. ასეთ მეთოდს ფარდობითი (რელატიური) სიხშირის მეთოდი ეწოდება. მაგალითად, MTF 2008-ის 1,266 რესპონდენტს დაუსვეს კითხვა სამუშაოზე პოლიციელების ქცევის შესახებ; მათი პასუხები შეჯამებულია ცხრილში 7.2.

ცხრილი 7.2 შეხედულებები პოლიციელების ქცევაზე

	სიხშირე	ხვედრითი წილი (პროპორცია)
ძალიან ცუდი	125	0.10
ცუდი	201	0.16
დამაკმაყოფილებელი	452	0.36
კარგი	380	0.30
ძალიან კარგი	108	0.09
სულ	1,266	1.01

რესპონდენტების თანაფარდობა ყველა რესპონდენტთან, ვინც „კარგად“ შეაფასა პოლიციელების ქცევა, არის 380:1,266, ან მიახლოებით 1:3. თანაფარდობის

პროპორციაში გადასაყვანად, მრიცხველს (380) ვყოფთ მნიშვნელზე (1,266), როგორც ეს ნაჩვენებია ცხრილში 7.2. ამიტომ, ალბათობა 380:1,266, ეკვივალენტურია 0.3-ის (380:1,266=0.3). ეხლა წარმოვიდგინოთ, რომ ფურცელზე ჩამოვწერეთ ყველა 1,266 რესპონდენტის სახელი და ქუდში ჩავყარეთ. იმის გამო, რომ პროპორცია (0.3) 0-თან უფრო ახლოს არის, ვიდრე 1-თან, იმის ალბათობა, რომ ვინც „კარგი" შეფასება მისცა პოლიციელებს, იმას ამოვიღებთ – მცირეა.

რა არის იმის ალბათობა, რომ კარტის კოლოდიდან ტუზი ამოვა? ეს 1/52 არ არის, რადგან კოლოდაში 4 ტუზია, ამიტომ ტუზის ამოსვლის ალბათობა იქნება 4/52 ან 1/13. პროპორცია იქნება 0.08. ყვავის ტუზის ამოსვლის ალბათობა იქნება 1/52 ან პროპორცია 0.02.

✓ ცოდნის შემოწმება

ჩვენ მიერ განხილული ფარდობითი სიხშირეები მხოლოდ მიახლოებულია რეალურ ალბათობასთან. ნამდვილი ალბათობის გამოვლენა მხოლოდ მაშინ არის შესაძლებელი, თუ ერთიდაიმავე პირობებში მრავალჯერ გავიმეორებთ ცდას. ამ შემთხვევაში გრძელვადიანი ფარდობითი სიხშირე ნამდვილ ალბათობას მიუახლოვდება.

ნორმალური განაწილება და ალბათობები

მეექვსე თავში თქვენ შეისწავლეთ ნორმალური განაწილება და მისი გარე „არეები". არეები გამოვსახეთ როგორც თანაფარდობა ან პროპორცია შერჩევაში მონაწილე დაკვირვებების რიცხვთან, ხოლო მათი საშუალოდან დაშორების ასახვა სტანდარტული გადახრის საზომ ერთეულებში მოვახდინეთ. ეს პროპორციები შესაძლებელს ხდის გავზომოთ ამ დაკვირვებების მოხდენის ალბათობა. მაგალითად, 200 თინეიჯერი გოგონას გამოკითხვის შედეგად დადგინდა, რომ დღეში ერთი გოგონა საშუალოდ 70 მესიჯს წერს, სტანდარტული გადახრით – 10. ჩვენ შეგვიძლია გამოვთვალოთ ამ ერთობლიობიდან გოგონას შერჩევის ალბათობა, რომელიც დღეში 70-დან 80 მესიჯამდე გზავნის. ეს არის 0.3413 (გადაათვალიერეთ თავი 6). თქვენ აგრეთვე შეგიძლიათ თქვათ: იმის შანსი, რომ შემთხვევით შერჩეული გოგონა 70-დან 80-მდე მესიჯს აგზავნის დღეში, 43.13%-ია.

ალბათური შერჩევა

ჩვენ მიერ ინფორმაციის ყოველდღიური შეგროვებისგან განსხვავებით, სოციალური სფეროს მეცნიერები უფრო თანმიმდევრულები არიან. ისინი სხვადასხვა მიდგომას იყენებენ მთლიანი ერთობლიობიდან შერჩევითი ერთობლიობის ამოსარჩევად. არსებობს მხოლოდ ერთი ძირითადი მიდგომა, *ალბათური შერჩევა*, რომელიც მკვლევარებს საშუალებას აძლევს გამოიყენონ დასკვნითი სტატისტიკის პრინციპები და გაავრცელონ მიღებული შედეგები მთელ ერთობლიობაზე.

ალბათური შერჩევა არის მეთოდი, რომელიც საშუალებას გვაძლევს თითოეული დაკვირვებისთვის (ქეისი) განვსაზღვროთ შერჩევაში მოხვედრის ალბათობა. ალბათური შერჩევის მიზანი იმაში მდგომარეობს, რომ შერჩევა მთლიანი ერთობლიობის რეპრეზენტატული იყოს.[3] შერჩევითი ერთობლიობის ამორჩევა ისე ხდება, რომ ალბათობის პრინციპებზე დაყრდნობით მოვახდინოთ შერჩევიდან მიღებული შედეგების გენერალიზაცია ანუ გავრცელება მთლიან ერთობლიობაზე. ალბათური შერჩევა საშუალებას გვაძლევს შევაფასოთ, თუ რამდენად განსხვავებული შეიძლება იყოს შერჩევითი გამოკვლევის შედეგები მთლიანი ერთობლიობისგან.

ალბათური შერჩევა მეთოდი, რომელიც საშუალებას გვაძლევს განვსაზღვროთ შერჩევაში მოხვედრის ალბათობა თითოეული დაკვირვებისთვის.

მიუხედავად იმისა, რომ შერჩევის შეცდომების ზუსტი გაზომვა მხოლოდ ალბათური შერჩევებიდან არის შესაძლებელი, სოციალური სფეროს მეცნიერები ხშირად მიმართავენ არაალბათურ შერჩევებს, რადგან მათი გამოყენება უფრო მოხერხებული და იაფია. ასეთი შერჩევების მთავარი შეზღუდვა მდგომარეობს დასკვნითი სტატისტიკის მეთოდის გამოყენების აკრძალვაში ერთობლიობაზე შერჩევის შედეგების გავრცელების დროს. ამ და მე-8 თავებში მხოლოდ დასკვნით სტატისტიკას შევეხებით, ამიტომ არაალბათურ განაწილებას აქ არ განვიხილავთ. შემდგომ ნაწილებში შევეხებით ალბათური შერჩევის სამ სახეს: (1) მარტივ შემთხვევით შერჩევას, (2) სისტემატურ (მექანიკურ), და (3) სტრატიფიცირებულ შერჩევას.[4]

მარტივი შემთხვევითი შერჩევა

მარტივი შემთხვევითი შერჩევა შერჩევის საბაზისო მოდელია, ის აგრეთვე ჩართულია შერჩევის უფრო რთულ სახეებშიც. **მარტივ შემთხვევით შერჩევაში:** (1) მთლიანი ერთობლიობის ყველა წევრს შერჩევაში მოხვედრის თანაბარი შანსი აქვს და (2) N წევრების თითოეულ კომბინაციას შერჩევაში მოხვედრის თანაბარი შანსი აქვს.

მარტივი შემთხვევითი შერჩევა შერჩევის მოდელი, რომელიც ისეა აგებული, რომ (1) მთლიანი ერთობლიობის ყველა წევრს შერჩევაში მოხვედრის თანაბარი შანსი აქვს, და (2) N წევრების თითოეულ კომბინაციას შერჩევაში მოხვედრის თანაბარი შანსი აქვს.

თვალსაჩინოებისთვის განვიხილოთ ძალიან მარტივი მაგალითი. ვთქვათ, დანახარჯების შემცირებასთან დაკავშირებით გამოკვლევას ვატარებთ ჩვენ რეგიონში

10 საავადმყოფოზე და გვინდა ინტენსიურად დავაკვირდეთ 2 მათგანს. ქაღალდის პატარა ნაჭრებზე დავიტანოთ ათივე საავადმყოფოს ნომრები, ჩავყაროთ ქუდში და ავურიოთ. ამოვიღოთ ერთი ნაჭერი, შემდეგ კლავ გავიმეოროთ ცდა და ამოვიღოთ მეორე ნაჭერი. ჩვენ უკვე შეექმნით შერჩევა, რომელშიც 2 საავადმყოფოა: (1) იმის გამო, რომ ნაჭრები კარგად ავურიეთ, ჩვენთვის უცნობი იყო, რომელი საავადმყოფოები იქნებოდა შერჩეული. ეს შერჩევა მარტივი შემთხვევითია, რადგან ყველა საავადმყოფოს შერჩევაში მოხვედრის თანაბარი შანსი ჰქონდა და (2) საავადმყოფოების ყოველ კომბინაციას (N = 2), შერჩევაში მოხვედრის თანაბარი შანსი ჰქონდა.

ჩვეულებრივ, შემთხვევითი შერჩევისთვის მკვლევარები კომპიუტერულ პროგრამებს ან შემთხვევითი რიცხვების ცხრილებს იყენებენ. დანართში 1 შემოკლებული სახით წარმოდგენილია შემთხვევითი რიცხვების ცხრილი. ცხრილის გამოყენებისთვის მთლიანი ერთობლიობის თითოეულ წევრს მიანიჭეთ ნომერი. ცხრილში ნებისმიერი მიმართულებით წაიკითხეთ ციფრი: ზევით, ქვევით, მარჯვნივ, მარცხნივ, მიმართულებას არა აქვს არანაირი მნიშვნელობა, და თუ ერთობლიობის წევრის ნომერს დაემთხვევა, მაშინ ის შეარჩიეთ. ეს პროცესი გაიმეორეთ შერჩევის სასურველი რაოდენობის შევსებამდე.

ახლა წარმოიდგინეთ, რომ ასეთივე გამოკვლევა გინდათ ჩაატაროთ პაციენტების ანკეტებზე. 300 ჩანაწერიდან უნდა შეარჩიოთ 5. თქვენ უნდა გაიაროთ შემდეგი საფეხურები:

1. პაციენტების ანკეტები დანომრეთ 001-დან 300-ის ჩათვლით.
2. გამოიყენეთ დანართი ა შემთხვევითი რიცხვების მოსაძებნად (თქვენ შეგიძლიათ უბრალოდ თვალები დახუჭოთ და ფანქარი დაადოთ ციფრს). თვალსაჩინოებისთვის დავიწყოთ რიცხვების პირველი მწკრივიდან. რადგან პაციენტების ერთობლიობა 3 ციფრისგან შედგება, ბოლო სვეტში 5 – თანრიგიანი რიცხვებია, ბოლო ორ თანრიგს ყურადღება არ მიაქციოთ (ასევე შეგიძლიათ რიცხვების ბოლო 3 თანრიგი გამოიყენოთ).
3. ბოლოდან ორი ციფრის ჩამოცილების შემდეგ, მიიღებთ შემდეგ სამ-თანრიგიან რიცხვებს:

104*	375	963	289*
223*	779	895	635
241*	995	854	094*
421			

ამ რიცხვებს შორის 5 მათგანი პაციენტების ანკეტების ნომრებს შეესაბამება. ისინი ვარსკვლავით არის აღნიშნული. ბოლო რიცხვი მე-13 სვეტიდანაა. მეტი ძებნა აღარ გჭირდებათ, რადგან 5 ანკეტა უკვე შერჩეულია. ეს რიცხვებია: 104, 223, 241, 289 და 094.

სისტემატური შემთხვევითი შერჩევა

ახლა განვიხილოთ შერჩევის სახე, რომლის გამოყენება მარტივ შემთხვევით შერჩევასთან შედარებით უფრო ადვილია. *სისტემატური შემთხვევითი შერჩევა* ასევე არ არის ნამდვილი ალბათური შერჩევა, თუმცა მიღებული შედეგები მარტივი შემთხვევითი შერჩევის მსგავსია. ის იყენებს კოეფიციენტს, რომელიც მიიღება მთლიანი ერთობლობის ზომის ფარდობით შერჩევის ზომასთან:

$$K = \frac{\text{მთლიანი ერთობლიობის ზომა}}{\text{შერჩევითი ერთობლიობის ზომა}}$$

სისტემატური შემთხვევითი შერჩევა არის მეთოდი, რომლითაც მთლიანი ერთობლიობიდან ნებისმიერი K-ური წევრის (K კოეფიციენტია, რომელიც მიიღება მთლიანი ერთობლობის ზომის ფარდობით შერჩევის ზომასთან) შერჩევა მოხდება მხოლოდ მაშინ, როდესაც შემთხვევით მოხდება შერჩევის პირველი წევრის ამორჩევა საწყისი K წევრებიდან.

სისტემატური შემთხვევითი შერჩევა არის მეთოდი, რომლითაც მთლიანი ერთობლიობიდან ნებისმიერი K-ური წევრის (K კოეფიციენტია, რომელიც მიიღება მთლიანი ერთობლობის ზომის ფარდობით შერჩევის ზომასთან) შერჩევა მოხდება მხოლოდ მაშინ, როდესაც შემთხვევით მოხდება შერჩევის პირველი წევრის ამორჩევა საწყისი K წევრებიდან.

გავიხსენოთ ჩვენი მაგალითი გარეუბნებში მცხოვრები სტუდენტების შესახებ. 15,000-დან უნდა გამოგვეკვლია 500. ამ მაგალითიდან,

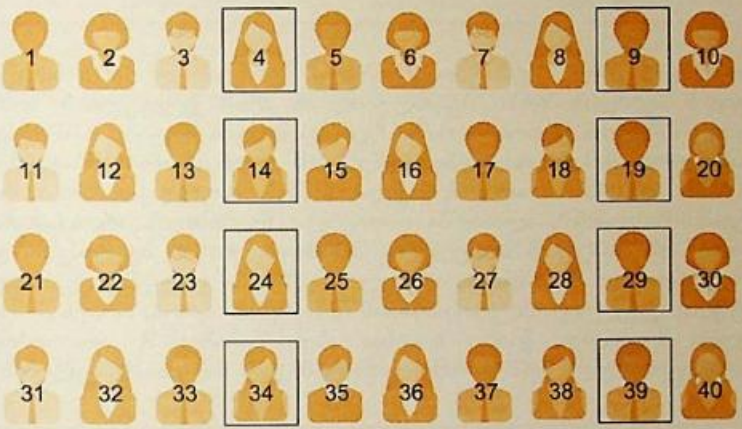
$$K = \frac{15,000}{500} = 30$$

სისტემატური შემთხვევითი შერჩევის მეთოდის გამოყენებით სტუდენტების სიის საწყისი 30 წევრიდან შემთხვევით შეარჩიეთ პირველი სტუდენტი და შემდეგ მის ნომერს ყოველ ჯერზე 30 მიუმატეთ, სანამ არ მიაღწევთ 500-მდე. დაუშვათ, პირველი სტუდენტის ნომერია 8, მაშინ მეორე სტუდენტი იქნება 38-ე (8+30=38), მესამე იქნება 38+30=68, მეოთხე 68+30=98 და ა.შ. სისტემატური შემთხვევითი შერჩევის მაგალითი განხილულია ნახაზზე 7.2.

✓ *ცოდნის შემოწმება* *განსხვავდება თუ არა სისტემატური შემთხვევითი შერჩევა მარტივი შემთხვევითი შერჩევისგან?*

ნახაზი 7.2 სისტემატური შემთხვევითი შერჩევა

40 სტუდენტის ერთობლიობიდან სისტემატური შემთხვევითი შერჩევით ამოვარჩიოთ 8 სტუდენტი. K-ს სიდიდე იქნება 5 (40 / 8 = 5). შემთხვევითი რიცხვების ცხრილიდან ჯერ ვირჩევთ ციფრებს 1-დან 5-მდე. ვთქვათ, ავირჩიეთ 4. შემდეგ ვინყებთ მე-4 სტუდენტიდან და ვირჩევთ ყოველ მეხუთეს:



შემთხვევითი რიცხვების ცხრილიდან ყველა სტუდენტს აქვს შერჩევაში მოხვედრის შანსი

სტრატიფიცირებული შემთხვევითი შერჩევა

ალბათური შერჩევის მესამე სახეა *სტრატიფიცირებული შემთხვევითი შერჩევა*. **სტრატიფიცირებული შემთხვევითი შერჩევა** მიიღება: (1) ჩვენთვის მნიშვნელოვანი 1 ან 2 ცვლადის ბაზაზე ერთობლიობის ქვეჯგუფებად დაყოფით და (2) თითოეულიდან მარტივი შემთხვევითი შერჩევის განხორციელებით. შეიძლება სტრატიფიკაცია მოვახდინოთ რასის, ეთნიკურობის მიხედვით; შევქმნათ რასობრივი ნიშნით ქვეჯგუფები და შემდეგ იქიდან ამოვირჩიოთ. ვთქვათ, გვაინტერესებს აბორტებზე ლათინო-ამერიკელი რეპონდენტების მოსაზრების შედარება თეთრკანიანებთან და შავკანიანებთან. მთლიანი ერთობლიობა შედგება 1,000 ინდივიდისაგან, რომელშიც 700 (ან 70%) თეთრკანიანია, 200 (ან 20%) შავკანიანი და 100 (ან 10%) ლათინო-ამერიკელი. ასეთ **პროპორციულად სტრატიფიცირებულ შერჩევაში** თითოეული ქვეჯგუფიდან ამორჩეული შერჩევა მთლიან ერთობლიობაში ქვეჯგუფის ზომის პროპორციულია.

სტრატიფიცირებული შემთხვევითი შერჩევა მიიღება: (1) ჩვენთვის მნიშვნელოვანი 1 ან 2 ცვლადის ბაზაზე ერთობლიობის ქვეჯგუფებად

დაყოფით და (2) თითოეულიდან მარტივი შემთხვევითი შერჩევის განხორციელებით.

პროპორციული სტრატიფიცირებული შერჩევა თითოეული ქვეჯგუფიდან ამორჩეული შერჩევა მთლიან ერთობლიობაში ქვეჯგუფის ზომის პროპორციულია.

არაპროპორციული სტრატიფიცირებული შერჩევის დროს თითოეული ქვეჯგუფიდან შერჩევის ზომა სპეციალურად არ არის ქვეჯგუფების ზომის პროპორციული მთლიან ერთობლიობაში. ჩვენ მაგალითში შეგვიძლია შერჩევის ზომა განვსაზღვროთ 180-ით, რომელშიც იქნება 90 თეთრკანიანი (50%), 45 შავკანიანი (25%) და 45 ლათინო-ამერიკელი (25%). მიუხედავად იმისა, რომ ასეთ შერჩევაში მოხვედრის ალბათობა თითოეული წევრისთვის არათანაბარია, ისინი ცნობილია და დასკვნით პროცესებში შესაძლებელია ამ შეცდომების გაზომვა.[5] არაპროპორციული სტრატიფიცირებული შერჩევის გამოყენება განსაკუთრებით ეფექტურია ქვეჯგუფების ერთმანეთთან შედარების დროს, როდესაც რომელიმე ქვეჯგუფი შედარებით მცირე ზომისაა. ასეთ შემთხვევაში არაპროპორციული სტრატიფიცირებული შერჩევა სანდო ინფორმაციის მიღების საშუალებას იძლევა.

არაპროპორციული სტრატიფიცირებული შერჩევა შერჩევის ზომა თითოეული ქვეჯგუფიდან სპეციალურად არ არის პროპორციული ქვეჯგუფების ზომისა მთლიან ერთობლიობაში.

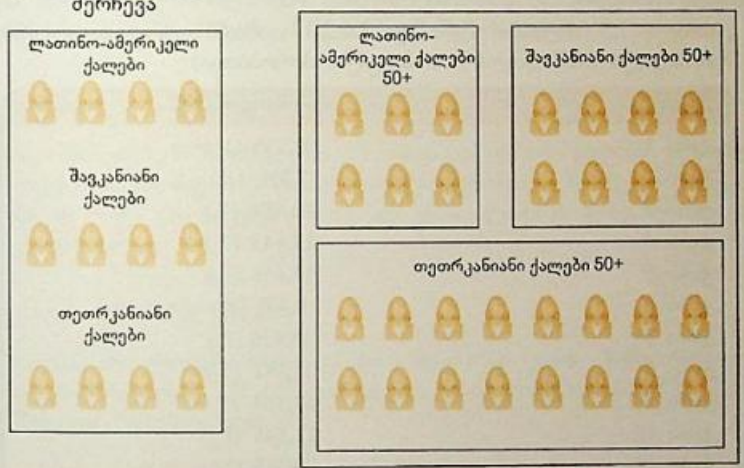
დეტალების განხილვა 7.3
არაპროპორციული სტრატიფიცირებული შერჩევა და მრავალფეროვნება

არაპროპორციული სტრატიფიცირებული შერჩევა განსაკუთრებით მნიშვნელოვანია ამერიკული საზოგადოების ზრდადი მრავალფეროვნების ფონზე. მრავალფეროვან საზოგადოებაში ისეთი ფაქტორები, როგორიცაა რასა, ეთნიკურობა, კლასი და სქესი ისევე მნიშვნელოვანია, როგორც ასაკი, რელიგია და სექსუალური ორიენტაცია. ეს ფაქტორები განსაზღვრავენ საზოგადოების სოციალურ სტრუქტურას და ერთობლივ ზემოქმედებას ახდენენ ჩვენზე.* მაგალითად, თეთრკანიან მამაკაცს შეუძლია გაიზიაროს მისი ასაკის ქალის ზოგადი გამოცდილება, მაგრამ გარკვეული სიძნელეები შეექმნება, თუ მათ შორის რასობრივი განსხვავებაა. მეტიც, რასობრივ-გენდერული გამოცდილების სისტემა განპირობებულია სოციალური კლასითაც. მაგალითად, სქესობრივთან ერთად მამაკაცი განიცდის რასობრივი, კლასობრივი და სექსუალური ორიენტაციის გავლენას. ჰომოსექსუალ შავკანიან მამაკაცს აბსოლუტურად განსხვავებული გამოცდილება გააჩნია თეთრკანიანი ჰეტეროსექსუალი კაცისგან განსხვავებით.

რა მიდგომები უნდა გამოვიყენოთ კვლევაში, რომ ხაზი გავუსვათ სოციალურ განსხვავებებს? რა თქმა უნდა, გამოვყოთ ქალები და მამაკაცები სქესის შესწავლის დროს, თეთრკანიანები და ფერადკანიანები რასის შესწავლის დროს და ა.შ. ასეთი კომპლექსური მიდგომა საშუალებას იძლევა უფრო მეტი ყურადღება გავამახვილოთ სხვადასხვა ჯგუფის გამოცდილებაზე და უფრო გამოვკვეთოთ მათი თავისებურებები.

შერჩევის რა მეთოდები უნდა გამოვიყენოთ ამ შემთხვევებში? დაუშვათ, გსურთ ლათინო-ამერიკელი და შავკანიანი ქალების სამუშაო გამოცდილების შესწავლა 50-ზე ზემოთ ასაკებში და გინდათ მათი შედარება ამავე ასაკის თეთრკანიან ქალებთან. ლათინო-ამერიკელი და შავკანიანი ქალები მოსახლეობაში მცირე ხვედრითი წილით არიან წარმოდგენილი. პროპორციულ შერჩევაში შეიძლება არ შევიდეს იმდენი ლათინო-ამერიკელი და შავკანიანი ქალი, რომ შედარების ბაზა ადეკვატური იყოს. ამის გამო, არაპროპორციული სტრატიფიცირებული შერჩევა დაგვჭირდება, რომელიც ლათინო-ამერიკელ და შავკანიან ქალებს ისე წარმოადგენს, რომ მათი შერჩევის ზომა შესაბამისი იქნება.

ნახაზი 7.3 სტრატიფიცირებული შერჩევა რასა/ერთობლიობა



* Margaret L. Andersen and Patricia Hill Collins, Race, Class and Gender (Belmont, CA: Wadsworth, 2009).

მოიფიქრეთ საკითხები ისეთი კვლევებისთვის, სადაც გამოიყენებთ არაპროპორციულ სტრატიფიცირებულ შერჩევას. რა შემთხვევაში გამოიყენება პროპორციული სტრატიფიცირებული შერჩევა?

✓ *ცოდნი შემოწმება*

შერჩევითი განაწილების არსი

ამ თავის დასაწყისში რამდენიმე მაგალითი მოვიყვანეთ, სადაც თვალსაჩინოდ იყო ახსნილი, თუ რატომ იყენებენ სოციალური სფეროს მეცნიერები მთლიანი ერთობლიობის ნაცვლად შერჩევით ერთობლიობებს. ისინი შერჩევის სტატისტიკურ

მახასიათებლებს მთლიანი ერთობლიობის ანუ პარამეტრების გასაზომად იყენებენ. ამ ნაწილში გაეცნობით დასკვნითი სტატისტიკის ერთ-ერთ უმნიშვნელოვანეს — *შერჩევითი განაწილების* კონცეფციას. მისი დახმარებით ვზომავთ შერჩევითი სტატისტიკის ალბათობას, რაც მისი გავრცელების საშუალებას იძლევა მთლიან ერთობლიობაზე.

მთლიანი ერთობლიობა

განვიხილოთ 20 ინდივიდისაგან შემდგარი ერთობლიობა, რომელიც წარმოდგენილია ცხრილში 7.3.[6] ჩვენი ცვლადია შემოსავალი Y, რომელიც დოლარებშია გაზომილი, ამ შემთხვევაში უნდა განვიხილოთ პარამეტრი საშუალო შემოსავალი.

ცხრილი 7.3 მთლიანი ერთობლიობა: 20 ადამიანის ინდივიდუალური შემოსავალი (აშშ დოლარი, მონაცემები პირობითია)

ინდივიდი	შემოსავალი (Y)
ქეისი 1	11,350 (Y_1)
ქეისი 2	7,859 (Y_2)
ქეისი 3	41,654 (Y_3)
ქეისი 4	13,445 (Y_4)
ქეისი 5	17,458 (Y_5)
ქეისი 6	8,451 (Y_6)
ქეისი 7	15,436 (Y_7)
ქეისი 8	18,342 (Y_8)
ქეისი 9	19,354 (Y_9)
ქეისი 10	22,545 (Y_{10})
ქეისი 11	25,345 (Y_{11})
ქეისი 12	68,100 (Y_{12})
ქეისი 13	9,368 (Y_{13})
ქეისი 14	47,567 (Y_{14})
ქეისი 15	18,923 (Y_{15})
ქეისი 16	16,456 (Y_{16})
ქეისი 17	27,654 (Y_{17})
ქეისი 18	16,452 (Y_{18})
ქეისი 19	23,890 (Y_{19})
ქეისი 20	25,671 (Y_{20})
საშუალო σ_Y = 22,766	სტანდარტული გადახრა (σ_Y) = 14,687

მთლიანი ერთობლიობის აღსანიშნავად ვიყენებთ სიმბოლოს μ_Y, აქ μ აღნიშნავს საშუალო სიდიდეს, ხოლო Y შემოსავალს. ფორმულა 4.1-ის საშუალებით შეგვი-

ძლია გავიანგარიშოთ მთლიანი ერთობლიობის საშუალო:

$$\mu_Y = \frac{\sum Y}{Y} = \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + \cdots + Y_{20}}{20}$$
$$= \frac{11,350 + 7,859 + 41,654 + 13,445 + 17,458 + \cdots + 25,671}{20}$$
$$= 22,766$$

5.3 ფორმულით ჩვენ შეგვიძლია აგრეთვე სტანდარტული გადახრის გაანგარიშება მთლიანი ერთობლიობის განაწილებისთვის. σ (სიგმით) აღნიშნულია სტანდარტული გადახრა, ხოლო Y აღნიშნავს შემოსავალს:

$$\sigma_Y = 14,687$$

ცხადია, რომ უმეტეს შემთხვევაში ვერ დავაკვირდებით მთლიან ერთობლიობას, ამიტომ ვიკვლევთ შერჩევით ერთობლიობას, ვანგარიშობთ შერჩევის საშუალოს – სტატისტიკას – და მას ვიყენებთ მთლიანი ერთობლიობის საშუალოს – პარამეტრის შესაფასებლად.

შერჩევითი ერთობლიობა

დაუშვათ, μ_Y არ არის ცნობილი და მისი შეფასება გვინდა სამი ინდივიდისაგან შემდგარი ერთობლიობის ($N = 3$ საშუალოს გაანგარიშების ბაზაზე. შერჩევაში შევიდა შემდეგი შემოსავლები:

ქეისი 8	18,342
ქეისი 16	16,456
ქეისი 17	27,654

ახლა გავიანგარიშოთ ამ შერჩევის საშუალო:

$$\overline{Y} = \frac{18,342 + 16,456 + 27,654}{3} = 20,817$$

როგორც ხედავთ, შერჩევის საშუალო ($\overline{Y}$)= $20,817-ს, ხოლო ერთობლიობის პარამეტრია $22,766. ეს განსხვავება შერჩევის შეცდომითაა გამოწვეული. **შერჩევის შეცდომა** არის განსხვავება შერჩევის საშუალებით შეფასებული მთლიანი ერთობლიობის პარამეტრსა და მთლიანი ერთობლიობის რეალურ პარამეტრს

შორის. ჩვენ შემთხვევაში შერჩევის შეცდომა არის 1,949 (22,766 – 20,817 = 1,949).

შერჩევის შეცდომა არის განსხვავება შერჩევის საშუალებით შეფასებული მთლიანი ერთობლიობის პარამეტრსა და მთლიანი ერთობლიობის რეალურ პარამეტრს შორის.

ახლა, შერჩევისათვის სხვა ინდივიდები ამოვარჩიოთ. შერჩეული შემოსავლები იქნება:

ქეისი 15	18,923
ქეისი 5	17,458
ქეისი 17	27,654

ამ შერჩევის საშუალო იქნება:

$$\overline{Y} = \frac{18,923 + 17,458 + 27,654}{3} = 21,345$$

შერჩევის შეცდომა ამ შემთხვევაში იქნება 1,421 (22,766 – 21,345 = 1,421), ცოტა ნაკლები, ვიდრე პირველ შემთხვევაში.

დილემა

ჩვენი გაანგარიშებების საიმედოობის შესამოწმებლად კარგი იქნებოდა, თუ ერთმანეთს შევადარებდით მთლიანი და შერჩევითი ერთობლიობების საშუალოებს. მაგრამ რეალურად ნამდვილი ერთობლიობის შესახებ ინფორმაცია ძალიან იშვიათად გვაქვს. რომ გვქონოდა, ამ გაანგარიშებებსაც არ ჩავატარებდით! ამასთან, შერჩევის გაზომვები იშვიათად (თუ ეს საერთოდ ხდება) ემთხვევა მთლიანი ერთობლიობის გაზომვებს. სწორედ აქ დგება დილემა: თუ შერჩევის შედეგები განსხვავებულია და მათ უმრავლესობაში შერჩევის შეცდომებია, რამდენად უნდა ვენდოთ მათ? რა საფუძველზე უნდა გავაკეთოთ შერჩევიდან დასკვნები მთლიანი ერთობლიობის შესახებ?

შერჩევითი განაწილება

ამ დილემაზე პასუხის გასაცემად არსებობს საშუალება, რასაც შერჩევითი განაწილება ჰქვია. **შერჩევითი განაწილება** ჩვენთვის საინტერესო სტატისტიკი-

სთვის საჭირო ყველა შესაძლებელი შერჩევითი სიდიდის თეორიული ალბათური განაწილებაა. თუ ჩვენ შევძლებთ მთლიანი ერთობლიობიდან ერთი და იმავე ზომის ყველა შესაძლებელი შემთხვევითი შერჩევის ამორჩევას, გამოვითვლით შერჩევისთვის ყველა სტატისტიკას, ავსახავთ მათ სიხშირულ განაწილებას, მივიღებთ მიახლოებით ალბათურ განაწილებას. თითოეულ სტატისტიკას — მაგალითად, პროპორციას, საშუალოს, ვარიაციას, — შერჩევითი განაწილება გააჩნია. რადგან მასში შედის შერჩევის ყველა შესაძლო სიდიდე, შერჩევითი განაწილება იმის საშუალებას იძლევა, რომ ერთმანეთთან შევადაროთ სხვა შერჩევის სიდიდეები და განვსაზღვროთ ამ შედეგის ალბათობა.[7]

შერჩევითი განაწილება ჩვენთვის საინტერესო სტატისტიკისთვის საჭირო ყველა შესაძლებელი შერჩევითი სიდიდის თეორიული ალბათური განაწილება.

საშუალო სიდიდის შერჩევითი განაწილება

შერჩევითი განაწილებები თეორიული განაწილებებია, რაც ნიშნავს, რომ მათზე დაკვირვება შეუძლებელია. რეალურად შერჩევითი განაწილებების აგებისათვის საჭიროა მთლიანი ერთობლიობიდან ყველა შესაძლებელი ფიქსირებული ზომის შემთხვევითი შერჩევის მოცვა. ეს ძალიან მოსაწყენი და დამღლელი პროცედურა იქნება და შერჩევების უამრავ რაოდენობას მოიცავს. მიუხედავად ამისა, ვნახოთ, როგორ ხდება ერთ-ერთი მათგანის აგება შერჩევების ზღვრული რაოდენობიდან.

თვალსაჩინოება

თვალსაჩინოებისთვის გამოვიყენოთ შერჩევითი განაწილების ერთ-ერთი ყველაზე გავრცელებული სახე — საშუალო სიდიდის შერჩევითი განაწილება. საშუალო სიდიდის შერჩევითი განაწილება შერჩევითი საშუალოების თეორიული განაწილებაა, რომელიც შეიძლება მივიღოთ მთლიანი ერთობლიობიდან ერთი ზომის ყველა შესაძლებელი შერჩევის განხილვით.

საშუალო სიდიდის შერჩევითი განაწილება შერჩევითი საშუალოების თეორიული ალბათური განაწილებაა, რომელიც შეიძლება მივიღოთ მთლიანი ერთობლიობიდან ერთი ზომის ყველა შესაძლებელი შერჩევის განხილვით.

დავუბრუნდეთ ჩვენ მაგალითს, სადაც ვიხილავდით 20 ადამიანის საშუალო

შემოსავალს. ცხრილში 7.3 მოცემული ერთობლიობიდან ახლა შემთხვევითი შერჩევის გზით ამოვარჩიოთ 50 სამწევრიანი შერჩევა და მათი შენაცვლებით მივიღოთ საშუალოები.

პირველ შერჩევაში შევიდა შემდეგი შემოსავლები: $8,451, $41,654 და $18,923. ამ შერჩევის საშუალო ტოლია:

$$\overline{Y} = \frac{8,451 + 41,654 + 18,923}{3} = 23,009$$

ახლა ვცადოთ შემდეგი სამი წევრისგან შემდგარი შერჩევისთვის საშუალოს გაანგარიშება:

$$\overline{Y} = \frac{15,436 + 25,345 + 16,456}{3} = 19,079$$

ანალოგიური ოპერაცია კიდევ 48-ჯერ გავიმეოროთ. ცხრილში 7.4 მოტანილია ყველა ეს შემთხვევა ($\sum\overline{Y}$ ყველა გაანგარიშებული საშუალოს ჯამია, ხოლო M-ით შერჩევების რაოდენობაა აღნიშნული, ანუ 50).

ცხრილში 7.5 მოცემულია 50 შერჩევის საშუალოების დაჯგუფებული სიხშირული განაწილებები და ნახაზზე 7.4 – ამ განაწილების ჰისტოგრამა.

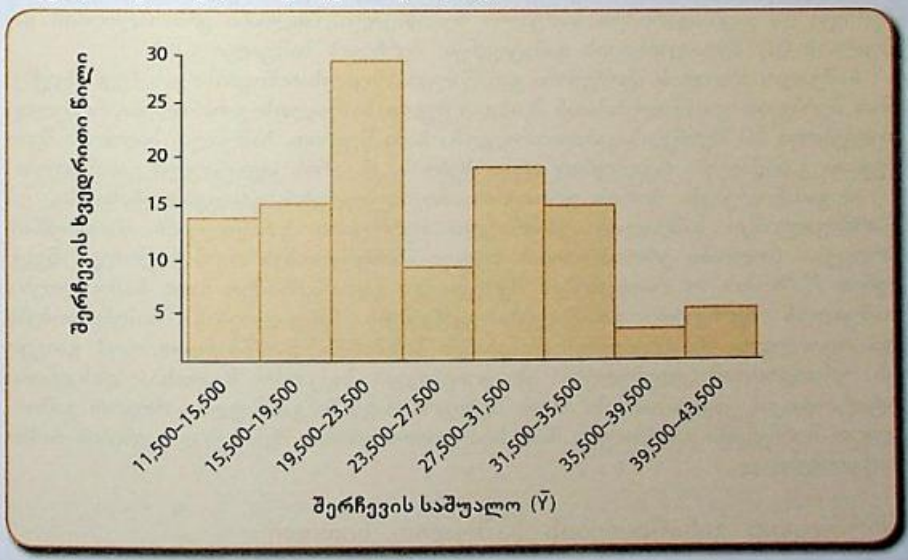
ცხრილი 7.4 საშუალო შემოსავალი სამწევრიანი 50 შერჩევის მიხედვით

შერჩევა	საშუალო
პირველი	23,009
მეორე	19,079
მესამე	18,873
მეოთხე	26,885
მეხუთე	21,847
⋮	⋮
ორმოცდამეათე	26,645
სულ (M) = 50	$\sum\overline{Y}$ = 1,237,482

ცხრილი 7.5 20 ინდივიდის ერთობლიობიდან 50 სამწევრიანი შერჩევის მიხედვით საშუალო შემოსავლის შერჩევითი განაწილება

შერჩევის საშუალოს ინტერვალები	სიხშირე	ხვედრითი წილი
11,500-15,500	6	12
15,500-19,500	7	14
19,500-23,500	14	28
23,500-27,500	4	8
27,500-31,500	9	18
31,500-35,500	7	14
35,500-39,500	1	2
39,500-43,500	2	4
სულ (M)	50	100

ნახაზი 7.4 20 ინდივიდის ერთობლიობიდან 50 სამწევრიანი შერჩევის მიხედვით საშუალო შემოსავლის შერჩევითი განაწილება



აღსანიშნავია, რომ შერჩევითი განაწილება თავისი სტრუქტურით ნედლი ქულების სიხშირულ განაწილებას ჰგავს, თუმცა აქ თითოეულ ქულას შერჩევის საშუალო შეესაბამება, ხოლო სიხშირეს — შერჩევის რაოდენობა შესაბამისი საშუალო მნიშვნელობით. მაგალითად, ცხრილში 7.5 მესამე ინტერვალის სიხშირეა

14, ანუ 28%. ეს ნიშნავს, რომ ჩვენ განაწილებაში 14 შერჩევა შეგვხვდა საშუალოს დიაპაზონით $19,500-დან $23,500-მდე.

დაიმახსოვრეთ, რომ ცხრილში 7.5 და ნახაზზე 7.4 ემპირიული განაწილებაა მოცემული, ხოლო შერჩევითი განაწილება თეორიული განაწილებაა. რეალობაში ჩვენ არ შეგვიძლია შერჩევითი განაწილების აგება. მიუხედავად ამისა, ეს ემპირიული განაწილება გვეხმარება შერჩევითი განაწილების დამახასიათებელი ნიშნების თვალსაჩინოებაში.

მიმოხილვა

სანამ გავაგრძელებდეთ, მოკლედ მიმოვიხილოთ განაწილების სამი ტიპი.

მთლიანი ერთობლიობა: ჩვენ დავიწყეთ 20 ინდივიდისგან შემდგარი მთლიანი ერთობლიობის განაწილებიდან. ეს განაწილება ნამდვილია და ემპირიული განაწილებაა, რომელიც ჩვეულებრივ ჩვენთვის უცნობია. გვაინტერესებს ამ მთლიანი ერთობლიობის საშუალო შემოსავლის შეფასება.

შერჩევითი ერთობლიობა: ჩვენ მოვახდინეთ შერჩევა. შერჩევითი განაწილება ემპირიული განაწილებაა, რომელიც ჩვენთვის ნაცნობია და გვეხმარება მთლიანი ერთობლიობის საშუალოს გაანგარიშებაში. ჩვენ შევარჩიეთ 50 სამწევრიანი შერჩევა და გავიანგარიშეთ საშუალო შემოსავალი. მთლიანი ერთობლიობის საშუალოს (μ_Y შეფასებისთვის გამოვიყენეთ შერჩევის საშუალო ($\overline{Y}$).

საშუალო სიდიდის შერჩევითი განაწილება: თვალსაჩინოებისთვის ჩვენ შევქმენით შერჩევით განაწილებასთან მიახლოებული საშუალოს განაწილება, რომელიც შედგებოდა 50 შერჩევისაგან, თითოეულში სამი წევრით. საშუალო სიდიდის შერჩევითი განაწილება რეალურად არ არსებობს. ეს არის თეორიული განაწილება.

ამ განაწილებებს შორის ურთიერთდამოკიდებულების გასაგებად ნახაზზე 7.5 წარმოვადგინეთ საშუალოს ემპირიული შერჩევითი განაწილების წარმოქმნის პროცესი. მთლიანი ერთობლიობის ნედლი მნიშვნელობებისაგან (Y_i) ჩვენ შევქმენით N ზომის M რაოდენობის შერჩევა და გავიანგარიშეთ მათი საშუალოები. საშუალოს შერჩევითი განაწილება გვიჩვენებს საშუალოების მნიშვნელობებს და თითოეული მნიშვნელობის შესაბამის სიხშირეს. დარწმუნდით, რომ გაიგეთ ეს ურთიერთდამოკიდებულება. ეს კონცეფცია საკვანძო საკითხია დასკვნითი სტატისტიკის გაგებისათვის. ამ და შემდგომ თავებში გასწავლით როგორ გამოიყენოთ შერჩევითი განაწილება მთლიანი ერთობლიობის შესახებ დასკვნების ჩამოსაყალიბებლად.

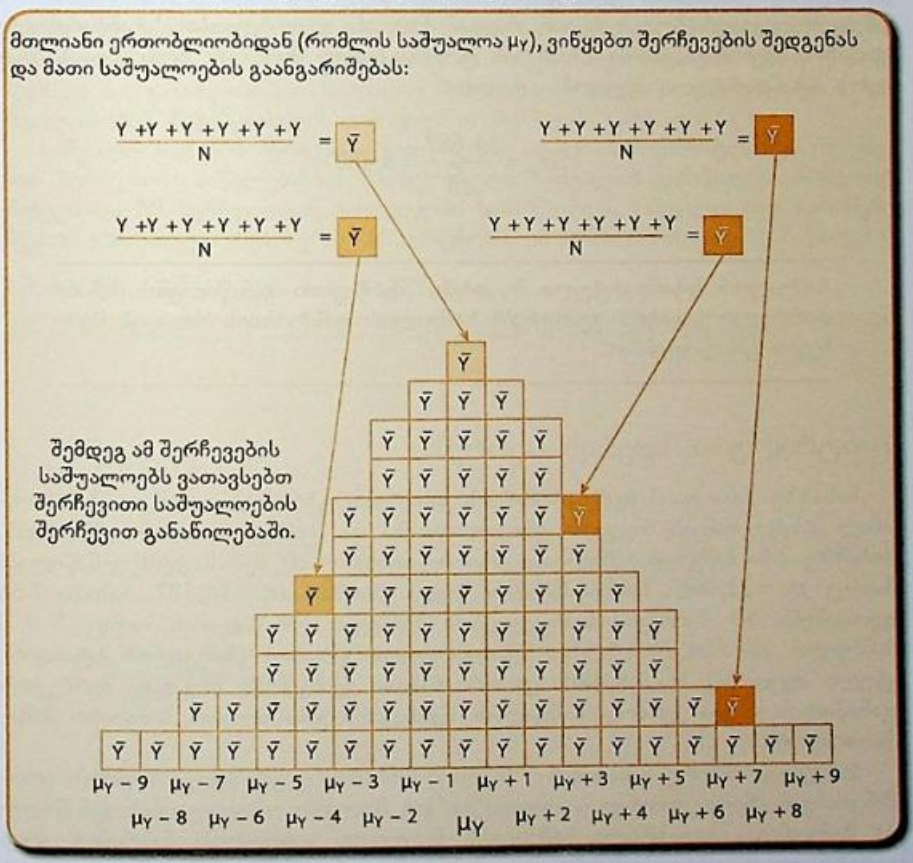
შერჩევითი განაწილების საშუალო სიდიდე

შერჩევისა და მთლიანი ერთობლიობის მსგავსად, შერჩევითი განაწილებისთვისაც შეიძლება საშუალოსა და სტანდარტული გადახრის გაანგარიშება. სიმბოლოს $\mu_{\overline{Y}}$ ვიყენებთ შერჩევითი განაწილების საშუალოს აღნიშვნისთვის. $\overline{Y}$ აქ მიანიშნებს, რომ განაწილების ცვლადი საშუალოა. შერჩევითი განაწილების საშუალო სიდიდის გაანგარიშებისთვის ინდივიდუალური შერჩევების საშუალოები

უნდა შევკრიბოთ $\sum \bar{Y} = 1,237.482$ და გავყოთ შერჩევების რაოდენობაზე ($M = 50$). ამიტომ, შერჩევითი განაწილების საშუალო სიდიდე რეალურად, საშუალოების საშუალოა:

$$\mu_{\bar{Y}} = \frac{\sum \bar{Y}}{M} = \frac{1,237,482}{50} = 24,750$$

ნახაზი 7.5 საშუალოს შერჩევითი განაწილების შექმნა



საშუალოს სტანდარტული შეცდომა

შერჩევითი განაწილების სტანდარტულ გადახრას *საშუალოს სტანდარტულ შეცდომასაც* უწოდებენ. საშუალოს სტანდარტული შეცდომა აღინიშნება სიმბო-

ლოთი $\sigma_{\bar{y}}$ და გვიჩვენებს დისპერსიის სიდიდეს შერჩევით განაწილებაში, ანუ თუ როგორია საშუალოს ვარიაცია შერჩევიდან შერჩევამდე:

$$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{N}}$$

ამ ფორმულიდან, საშუალოს სტანდარტული შეცდომა ტოლია მთლიანი ერთობლიობის სტანდარტული გადახრის σ_y ფარდობისა შერჩევის ზომიდან (N) ამოღებულ კვადრატულ ფესვთან. რადგან მთლიანი ერთობლიობის სტანდარტული გადახრა ჩვენ მაგალითში 14,687-ის ტოლია, ხოლო შერჩევის ზომა 3-ია, საშუალოს სტანდარტული შეცდომა ტოლია:

$$\sigma_{\bar{y}} = \frac{14,687}{\sqrt{3}} = 8,480$$

საშუალოს სტანდარტული შეცდომა შერჩევითი განაწილების სტანდარტული გადახრა. გვიჩვენებს საშუალოს დისპერსიის სიდიდეს შერჩევით განაწილებაში.

ცენტრალური ზღვრის თეორემა

ნახაზზე 7.6ა და ბ, ჩვენ ერთმანეთს ვადარებთ ცხრილებში 7.2 და 7.4 მოტანილ ჰისტოგრამებს მთლიანი ერთობლიობისა და შერჩევითი განაწილებისთვის. ნახაზზე 7.6ა ნაჩვენებია მთლიანი ერთობლიობის – 20 შემოსავლის განაწილება, სადაც $\mu_{\bar{y}} = 22,766$, ხოლო სტანდარტული გადახრა $\sigma_{\bar{y}} = 14,687$. ნახაზი 7.6ბ გვიჩვენებს 50 შერჩევის საშუალოების შერჩევით განაწილებას, სადაც $N =$ საშუალო $\mu_{\bar{y}} = 24,749$, ხოლო სტანდარტული გადახრა (საშუალოს სტანდარტული შეცდომა) $\sigma_{\bar{y}} = 8,480$. ეს ორი ნახაზი გვიჩვენებს ზოგადად შერჩევითი განაწილებისა და კერძოდ საშუალოს შერჩევითი განაწილების საბაზისო მახასიათებლებს.

პირველი, როგორც ნახაზიდან 7.6ა და ბ ჩანს, ორივე განაწილების ფორმა ერთმანეთისგან მნიშვნელოვნად განსხვავდება. თუ მთლიანი ერთობლიობის განაწილება მარჯვნივაა გადახრილი, საშუალოს შერჩევითი განაწილება ნაკლებად არის გადახრილი – უფრო სიმეტრიულია და ნორმალურ განაწილებას უახლოვდება.

მეორე, მაშინ, როდესაც მხოლოდ რამდენიმე შერჩევის საშუალო ემთხვევა ზუსტად მთლიანი ერთობლიობის საშუალოს ($22,766), შერჩევითი განაწილების ცენტრი ამ სიდიდეზე გადის. შერჩევითი განაწილების საშუალო საკმაოდ უახლოვდება მთლიანი ერთობლიობის საშუალოს.

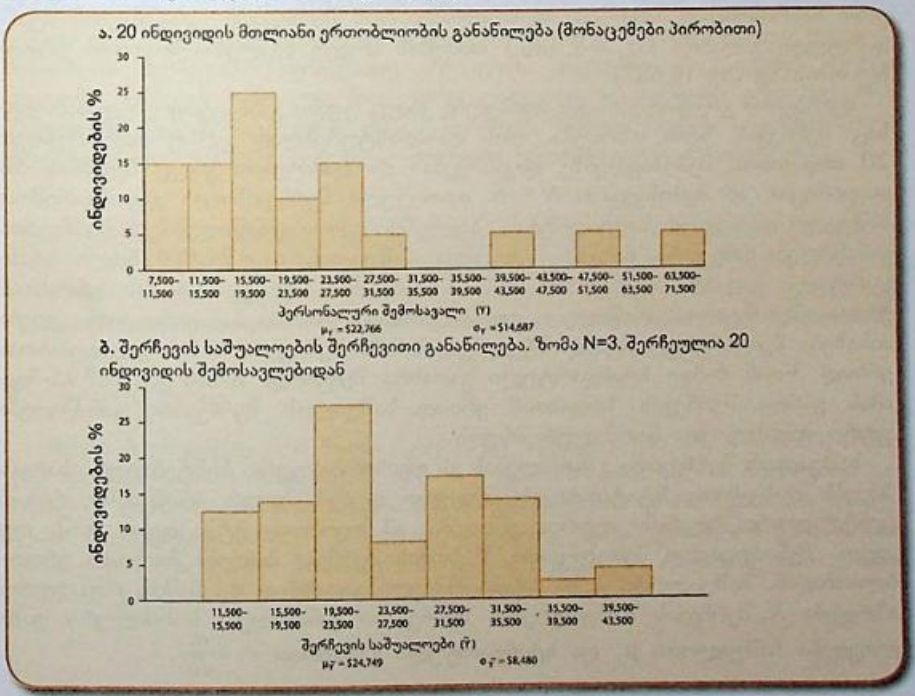
ზემოთ მოყვანილი განხილვის დროს ჩვენ ხშირად ვახსენებდით სამივე განაწილების საშუალოსა და სტანდარტულ გადახრას. მათი განსხვავების მიზნით სხვადასხვა სიმბოლოები გამოვიყენეთ ინდექსებად. შენიშნავდით, რომ როგორც მთლიანი, ისე შერჩევითი განაწილებისთვის, ვიყენებთ ბერძნულ სიმბოლოებს.

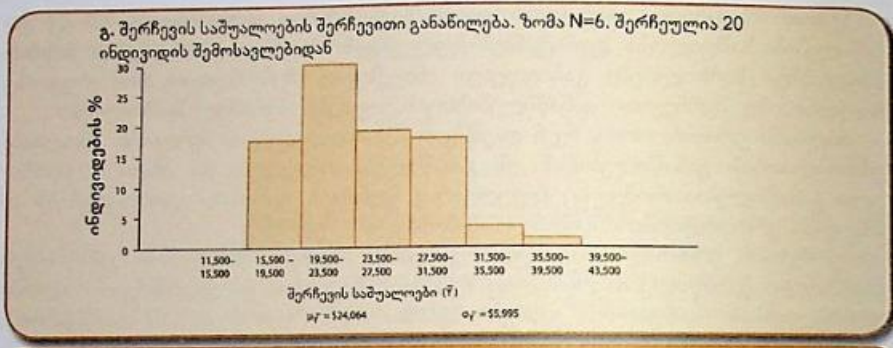
მთლიანი ერთობლიობა: ჩვენ დავიწყეთ 20 ინდივიდისგან შემდგარი მთლიანი ერთობლიობის განაწილებიდან. ეს განაწილება ნამდვილია და ამასთან ემპირიული განაწილებაა, რომელიც ჩვეულებრივ ჩვენთვის უცნობია. გვაინტერესებს ამ მთლიანი ერთობლიობის საშუალო შემოსავლის შეფასება.

შერჩევითი ერთობლიობა: ჩვენ მოვახდინეთ შერჩევა. შერჩევითი განაწილება ემპირიული განაწილებაა, რომელიც ჩვენთვის ნაცნობია და გვეხმარება მთლიანი ერთობლიობის საშუალოს გაანგარიშებაში. ჩვენ შევარჩიეთ 50 სამწევრიანი შერჩევა და გავიანგარიშეთ საშუალო შემოსავალი. მთლიანი ერთობლიობის საშუალოს (μ_Y) შეფასებისთვის გამოვიყენეთ შერჩევის საშუალო ($\overline{Y}$).

საშუალო სიდიდის შერჩევითი განაწილება: თვალსაჩინოებისთვის ჩვენ შევქმენით შერჩევით განაწილებასთან მიახლოებული საშუალოს განაწილება, რომელიც შედგებოდა 50 შერჩევისაგან, თითოეულში სამი წევრით. საშუალო სიდიდის შერჩევითი განაწილება რეალურად არ არსებობს. ეს არის თეორიული განაწილება.

ნახაზი 7.6 შემოსავლის სამი განაწილება





	საშუალო	სტანდარტული გადახრა
შერჩევითი განაწილება	$\bar{Y}$	S_Y
მთლიანი ერთობლიობის განაწილება	μ_Y	σ_Y
$\bar{Y}$-ის შერჩევითი განაწილება	$\mu_{\bar{Y}}$	$\sigma_{\bar{Y}}$

მესამე, შერჩევითი განაწილების ვარიაცია მთლიან ერთობლიობასთან შედარებით საგრძნობლად ნაკლებია. აღსანიშნავია, რომ შერჩევითი განაწილების სტანდარტული გადახრა ($\sigma_{\bar{Y}} = 8,480$) თითქმის 2-ჯერ ნაკლებია მთლიანი ერთობლიობისაზე (σ_Y=14,687).

შერჩევითი განაწილების ეს თვისებები კიდევ უფრო განსაცვიფრებელია, როდესაც შერჩევის ზომა იზრდება. ამის დასადასტურებლად კვლავ დავუბრუნდეთ 20 ინდივიდის შემოსავლებზე დაკვირვებას და დამატებით კიდევ შევქმნათ 50 დაკვირვება. ამ შემთხვევაში $N = 6$. თითოეული შერჩევისთვის გავიანგარიშოთ საშუალო და ავაგოთ საშუალოების ახალი შერჩევითი განაწილება. ეს შერჩევითი განაწილება ნაჩვენებია ნახაზზე 7.გ. მისი საშუალოა $\mu_{\bar{Y}} = 24,064$, ხოლო სტანდარტული გადახრა $\sigma_{\bar{Y}} = 5,995$, აღსანიშნავია, რომ შერჩევის ზომის ზრდასთან ერთად შერჩევითი განაწილება უფრო კომპაქტური ხდება. ვარიაციის კლება აისახება შემცირებულ სტანდარტულ გადახრაში: შერჩევის ზომის გადიდებასთან ერთად 3-დან 6-მდე სტანდარტული გადახრა შემცირდა 8,480-დან 5,995-მდე. ამის გარდა, შერჩევის ზრდასთან ერთად, საშუალოს შერჩევითი განაწილება უფრო მიუახლოვდა ნორმალურ მრუდს.

საშუალოს შერჩევითი განაწილების ეს თვისებები უფრო სისტემატურ ასახვას პოვებს დასკვნითი სტატისტიკის ერთ-ერთ ფუძემდებლურ პრინციპში, რასაც **ცენტრალური ზღვრის თეორია** ეწოდება. ამ თეორიით მტკიცდება, რომ, ოდესაც შესაძლებელი შემთხვევითი N ზომის შერჩევა მიიღება მთლიანი ერთობლიობიდან, საშუალოთი μ_Y და სტანდარტული გადახრით σ_Y, მაშინ, რაც უფრო იზრდება N, შერჩევის საშუალოების შერჩევითი განაწილება ნორმალურს უახლოვდება საშუალოთი $\mu_{\bar{Y}}$ და სტანდარტული გადახრით $\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{N}}$.

ცენტრალური ზღვრის თეორია თუ ყველა შესაძლებელი შემთხვევითი N ზომის შერჩევა მიიღება მთლიანი ერთობლიობიდან, საშუალოთი და სტანდარტული გადახრით μ_Y, მაშინ, რაც უფრო იზრდება σ_Y, შერჩევის საშუალოების შერჩევითი განაწილება ნორმალურს უახლოვდება საშუალოთი $\mu_{\bar{Y}}$ და სტანდარტული გადახრით

$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{N}}$$

ცენტრალური ზღვრის თეორიის არსი შემდეგში მდგომარეობს: მთლიანი განაწილების ფორმისგან დამოუკიდებლად, *შერჩევის საკმარისი ზომის* შემთხვევაში, საშუალოების შერჩევითი განაწილება ნორმალური იქნება.

ამიტომ, მაშინაც კი, როდესაც მთლიანი ერთობლიობის განაწილება გადახრილია, შეგვიძლია ვთქვათ, რომ, თუ შერჩევის ზომა საკმარისია, საშუალოს შერჩევითი განაწილება ნორმალურია. უფრო მეტიც, ცენტრალური ზღვრის თეორიიდან გამომდინარეობს, რომ (1) რაც უფრო იზრდება შერჩევის ზომა, შერჩევითი განაწილების საშუალო მთლიანი ერთობლიობის საშუალოს ტოლი ხდება და (2) რაც უფრო იზრდება შერჩევის ზომა, საშუალოს სტანდარტული შეცდომა (საშუალოს შერჩევითი განაწილების სტანდარტული გადახრა) მცირდება. სტანდარტული შეცდომა მიანიშნებს შერჩევიდან შერჩევამდე ვარიაციის სიდიდეზე. რაც უფრო მცირეა შეცდომა, შერჩევის საშუალო მით უფრო უახლოვდება მთლიანი ერთობლიობის საშუალოს. აქედან, რაც უფრო დიდია შერჩევის ზომა, შერჩევის სტატისტიკური კლასტერი მით უფრო უახლოვდება მთლიანი ერთობლიობის პარამეტრებს.

დარწმუნდით, რომ სწორად გაიგეთ განსხვავება შერჩევების რაოდენობასა და შერჩევის ზომას შორის. თუ შერჩევების რაოდენობა თეორიულად უსასრულოა, შერჩევის ზომა მკვლევარზეა დამოკიდებული.

✓ ცოდნი შემოწმება

შერჩევის ზომა

იმისდა მიუხედავად, რომ რაიმე მკაცრი და სწრაფი წესი არ არსებობს, ზოგადად მიღებულია, რომ როდესაც $N = 50$ ან მეტია მასზე, საშუალოს შერჩევითი განაწილება დაახლოებით ნორმალური იქნება, განაწილების ფორმისგან დამოუკიდებლად. მიუხედავად ამისა, ჩვენ შეგვიძლია აგრეთვე დავასკვნათ, რომ განაწილება ნორმალური იქნება $N = 30$-ის შემთხვევაშიც, თუ ვიცით, რომ მთლიანი ერთობლიობის განაწილება ნორმალურს უახლოვდება.

ცოდნის შემოწმება

რა არის მთლიანი ერთობლიობის ნორმალური განაწილება? თუ პასუხი არ გაქვთ, დაუბრუნდით მე-6 თავს. თქვენ უნდა კარგად გაიგოთ ნორმალური განაწილების კონცეფცია დასკვნითი სტატისტიკის მეთოდების შესწავლისათვის.

შერჩევითი დაკვირვებისა და ცენტრალური ზღვრის თეორემის მნიშვნელობა

წინა განყოფილებებში ბევრ აბსტრაქტულ მასალას შეეხეთ. თქვენ ალბათ კითხვები გაგიჩნდათ ამასთან დაკავშირებით. რა მნიშვნელობა აქვს ცენტრალური ზღვრის თეორემას? ამ კითხვაზე პასუხის გასაცემად მივუბრუნდეთ და მიმოვიხილოთ 20 შემოსავლის მაგალითი.

მთლიანი ერთობლიობის ანუ 20 ინდივიდის საშუალო შემოსავლის გაზომვის მიზნით ჩვენ შევადგინეთ სამწევრიანი შერჩევა და გამოვითვალეთ შერჩევის საშუალო. შერჩევის საშუალო $\overline{Y} = 20,817$, განსხვავდება ნამდვილი ერთობლიობის პარამეტრისგან $\mu_Y = 22,766$. როდესაც განსხვავებულ შერჩევებს ვაკვირდებოდით, ვნახეთ, რომ ისინი განსხვავდებოდნენ მთლიანი ერთობლიობის საშუალოსაგან. ეს შერჩევის შეცდომის გამო ხდებოდა. რამდენი ახალი შერჩევაც არ უნდა დავათვალიეროთ, ისინი იშვიათი გამონაკლისის გარდა, მაინც მთლიანი ერთობლიობისაგან განსხვავებული იქნება საშუალოს სიდიდის მიხედვით. ჩვეულებრივ, მხოლოდ ერთი შერჩევის სტატისტიკას ვიყენებთ მთლიანი ერთობლიობის პარამეტრის შესაფასებლად.

ისევ ხელახლა მივუბრუნდეთ დილემას: თუ შერჩევის შედეგები განსხვავებულია და მათ უმრავლესობაში შერჩევის შეცდომებია, რამდენად უნდა ვენდოთ მათ? რა საფუძველზე უნდა გავაკეთოთ შერჩევიდან დასკვნები მთლიანი ერთობლიობის შესახებ?

საკითხის გადაწყვეტის გასაღები შერჩევით განაწილებასა და მის მახასიათებლებშია. რადგან შერჩევითი განაწილება თეორიული განაწილებაა, რომელიც ყველა შესაძლებელი შერჩევის შედეგს მოიცავს, ჩვენი შერჩევის შედეგი შეგვიძლია მას შევადაროთ და შევაფასოთ მისი მოხდენის ალბათობა.

თუ შერჩევითი განაწილება თეორიულია, საიდან უნდა ვიცოდეთ მისი ფორმა და მახასიათებლები ამ შედარებებისთვის? ჩვენი ცოდნა ემყარება ცენტრალური ზღვრის თეორემას, საიდანაც ვგებულობთ, რომ თუ თქვენი შერჩევის ზომა (არანაკლებ 50 დაკვირვება) საკმარისად დიდია, შერჩევითი საშუალოების უმრავლესობა მთლიანი ერთობლიობის საშუალოს მიუახლოვდება. ამ მოვლენის მოხდენის ალბათობა საკმაოდ დიდია.

მეექვსე თავში ვნახეთ, რომ ნორმალურ მრუდში საშუალოსა და მის გარკვეებს შორის მუდმივი პროპორცია არსებობს, როდესაც ეს დაცილება სტანდარტული გადახრის საზომებით იზომება, მათ Z ქულებს უწოდებენ. ეს პროპორციები შეიძლება მოვძებნოთ ნორმალური განაწილების ცხრილში (დანართი ბ).

როდესაც ვიცით, რომ საშუალოების შერჩევითი განაწილება დაახლოებით

ნორმალურია, საშუალოთი $\mu_{\bar{y}}$ და სტანდარტული გადახრით $\frac{\sigma_y}{\sqrt{N}}$ (საშუალოს სტანდარტული შეცდომა), დანართი ბ-ს საშუალებით შეგვიძლია განვსაზღვროთ ალბათობა, რომლითაც შერჩევის საშუალო გარკვეულ დისტანციაზე მოხვდება, რაც $\mu_{\bar{y}}$-ის ან μ_y-ის Z ქულებში გამოისახება. მაგალითად, სავარაუდოდ, შერჩევითი საშუალოების 68% (ანუ ალბათობა 0.68-ის ტოლია) $\mu_{\bar{y}}$-დან ან μ_y-დან ±1 სტანდარტული შეცდომის ფარგლებში მოხვდება. ანალოგიურად, თუ ალბათობა 0.95-ია, შერჩევითი საშუალოების 95% $\mu_{\bar{y}}$-დან ან μ_y-დან ± 2 სტანდარტული შეცდომის ფარგლებში მოხვდება. შემდეგ თავში დაწვრილებით გვექნება აღნიშნულის განხილვის საშუალება.

✓ *ცოდნის შემოწმება*

დაუშვათ, მთლიანი ერთობლიობის საშუალო $\mu_y = 150$, ხოლო სტანდარტული გადახრა – $\sigma_y = 30$. თქვენ შეადგინეთ შემთხვევითი შერჩევა $N = 100$ დაკვირვებით (ქეისით). რა იქნება იმის ალბათობა, რომ საშუალო 147-სა და 153-ს შორის იქნება? რა იქნება იმის ალბათობა, რომ საშუალო 153-ს გადასცდება? გაგიკვირდებათ, თუ საშუალო 159-ის ტოლი იქნება? რატომ? (შეხსენება: ამ კითხვებზე პასუხის გასაცემად იხ. თავი 6, Z ქულები და და არეები ნორმალური მრუდის ქვეშ, დანართი ბ). ნელი მნიშვნელობის Z ქულაში გადასაყვანად ვიყენებთ ფორმულას:

$$Z = \frac{Y - \overline{Y}}{S_Y}$$

რადგან აქ შერჩევით განაწილებას ვიხილავთ, Y შეანაცვლეთ $\overline{Y}$-ით, $\overline{Y}$ შერჩევითი განაწილების საშუალოთი $\mu_{\bar{y}}$ და σ_y საშუალოს სტანდარტული შეცდომით $\frac{\sigma_y}{\sqrt{N}}$

$$Z = \frac{\overline{Y} - \mu_{\bar{y}}}{\sigma_y / \sqrt{N}}$$

ძირითადი განმარტებები:

• შერჩევის პროცესში მკვლევარები ცდილობენ ქვესიმრავლიდან (შერჩევითი ერთობლიობიდან) მიღებული მახასიათებლები გაავრცელონ მთლიან ერთობლიობაზე. ტერმინი პარამეტრი უკავშირდება მთლიან ერთობლიობაში ჩვენთვის საინტერესო ინფორმაციას, ხოლო სტატისტიკა – შერჩევითი დაკვირვების გაანგარიშებებს.

• ალბათური შერჩევა გვაძლევს იმის საშუალებას, რომ შევაფასოთ, როგორი იქნება შერჩევითი დაკვირვების შედეგების განსხვავების ალბათობა მთლიანისაგან.

• მარტივი შემთხვევითი შერჩევა ისეა აგებული, რომ მთლიანი ერთობლიობის ყველა წევრს და წევრების

თითოეულ კომბინაციას შერჩევაში მოხვედრის თანაბარი შანსი აქვს.

• სისტემატურ შემთხვევით შერჩევაში მთლიანი ერთობლიობიდან ნებისმიერი K-ური წევრის შერჩევა მხოლოდ მაშინ მოხდება, როდესაც შემთხვევით მოხდება შერჩევის პირველი წევრის ამორჩევა საწყისი K წევრებიდან.

• სტრატიფიცირებული შემთხვევითი შერჩევა მიიღება: (1) ჩვენთვის მნიშვნელოვანი 1 ან 2 ცვლადის ბაზაზე ერთობლიობის ქვეჯგუფებად დაყოფით და (2) თითოეულიდან მარტივი შემთხვევითი შერჩევის განხორციელებით.

• შერჩევითი განაწილება ჩვენთვის საინტერესო სტატისტიკისთვის საჭირო ყველა შესაძლებელი შერჩევითი სიდიდის თეორიული ალბათური განაწილებაა. საშუალო სიდიდის შერჩევითი განაწილება შერჩევითი საშუალოების თეორიული განაწილებაა, რომელიც შეიძლება მივიღოთ მთლიანი ერთობლიობიდან ერთი ზომის ყველა შესაძლებელი შერჩევის განხილვით.

• ცენტრალური ზღვრის თეორიით მტკიცდება, რომ, თუ ყველა შესაძლებელი შემთხვევითი N ზომის შერჩევა მიიღება მთლიანი ერთობლიობიდან, რომლის საშუალოა μ_y და სტანდარტული გადახრა σ_y, მაშინ, რაც უფრო იზრდება N, შერჩევის საშუალოების შერჩევითი განაწილება ნორმალურს უახლოვდება საშუალოთი $\mu_{\bar{y}}$ და სტანდარტული გადახრით $\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{N}}$.

• ცენტრალური ზღვრის თეორიიდან გამომდინარეობს, რომ შერჩევის საკმარისი ზომის შემთხვევაში, როდესაც მთლიანი ერთობლიობის განაწილება გადახრილია, შეგვიძლია ვთქვათ, საშუალოს შერჩევითი განაწილება ნორმალურია.

საკვანძო ტერმინები

ცენტრალური ზღვრის თეორემა
არაპროპორციული სტრატიფიცირებული შერჩევა
პარამეტრი
მთლიანი ერთობლიობა
ალბათური შერჩევა
პროპორციული სტრატიფიცირებული შერჩევა
შერჩევითი ერთობლიობა
საშუალოს შერჩევითი განაწილება
შერჩევის შეცდომა
მარტივი შემთხვევითი შერჩევა
საშუალოს სტანდარტული შეცდომა
სტატისტიკა
სტრატიფიცირებული შემთხვევითი შერჩევა
სისტემატური შემთხვევითი შერჩევა

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

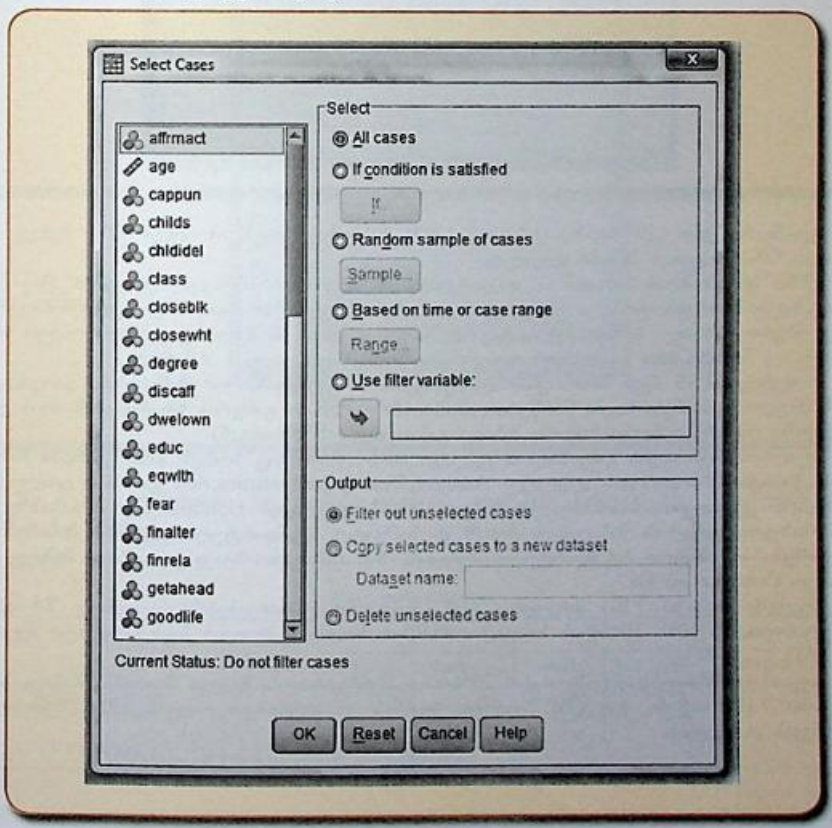
[GSS08PFP-A]

შემთხვევითი შერჩევის ამორჩევა

ამ თავში სხვადასხვა სახის შერჩევა განვიხილეთ და განვსაზღვრეთ საშუალოს სტანდარტული შეცდომა. ჩვეულებრივ, SPSS-ში შეყვანილი მონაცემები უკვე მთლიანი ერთობლიობიდან არის შერჩეული. SPSS-ს გააჩნია შემთხვევითი შერჩევის მექანიზმი, მისი საშუალებით სისტემატური და სტარტიფიცირებული შერჩევის განხორციელებაც შესაძლებელია, თუმცა ამ შემთხვევაში პროგრამირების ცოდნაა საჭირო.

რა შემთხვევაში გამოგვადგება SPSS-ის შერჩევის პროცედურები? ერთ-ერთია – ძალიან დიდი მონაცემთა ბაზის წინასწარი ანალიზი. მაგალითად, თუ ადგილობრივ საავადმყოფოში მუშაობთ და გაქვთ ავადმყოფების სრული ბაზა, პირველადი ანალიზისათვის ყველა მონაცემი არ არის საჭირო. თქვენ შეგიძლიათ ინდივიდების შერჩევა და ამ ქვესიმრავლის წინასწარი ანალიზი. მოგვიანებით მთლიანი ბაზის ანალიზიც დაგჭირდებათ ანალიზის სისრულისათვის.

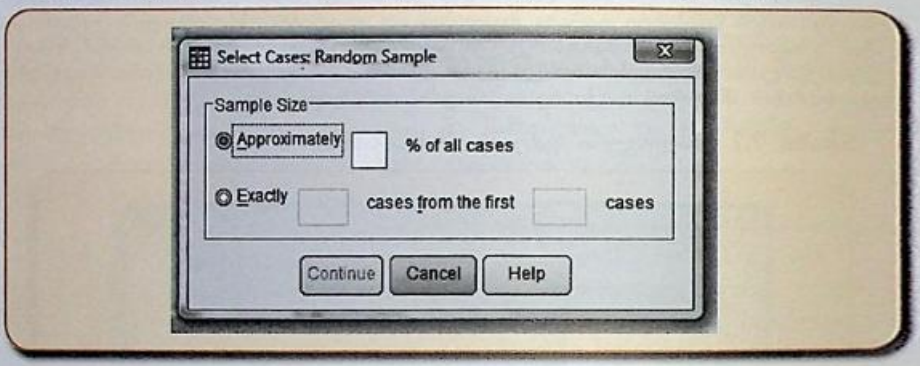
ნახაზი 7.7 დიალოგური ფანჯარა Select Cases



შერჩევის პროცედურების გამოსაყენებლად დაკლიკეთ ძირითადი მენიუდან Data, შემდეგ Select Cases. გაიხსნება დიალოგური ფანჯარა (ნახაზი 7.7) ქვესიმრავლის 5 არჩევანით. დათქმის მიხედვით აირჩიეთ All Cases. შემდეგ დაკლიკეთ Random sample of cases ღილაკზე და Sample ღილაკზე SPSS-ისთვის ჩვენი პირობების მისაცემად.

შემდეგ დიალოგური ფანჯარას (ნახაზი 7.8) აქვს ორი ოპცია. ყველაზე მოსახერხებელია პირველი, რომლითაც SPSS-ს ვეუბნებით, თუ დაკვირვებების რა პროცენტი ამოარჩიოს მონაცემთა დიდი ბაზიდან. ალტერნატულად, SPSS-ს შეიძლება მივუთითოთ ამოარჩიოს დაკვირვებების ზუსტი რაოდენობა. მეორე ვარიანტი მისაღებია, რადგან SPSS პირველ ოფციაში ამოარჩევს მხოლოდ მიახლოებულ პროცენტს.

ნახაზი 7.8 შერჩევის ზომის განსაზღვრა ბრძანებაში Selecting Cases



ყუთში ჩავწერთ „10"-ს, იმისთვის, რომ 1,500 რესპონდენტიდან ამოირჩეს 10%. შემდეგ დაკლიკეთ OK, პროცესის შესასრულებლად.

SPSS არ ამოშლის ბაზიდან იმ დაკვირვებებს, რომლებიც არ მოხვდა შერჩევაში. ამის ნაცვლად, ხდება მათი ფილტრაცია (მონაცემთა ფანჯარაში შეგიძლიათ მათი ნახვა ნელდ მონაცემებზე გავლებული დახრილი ხაზით). ეს ნიშნავს, რომ ჩვენ ყოველთვის შეგვიძლია დავუბრუნდეთ სრულ ბაზას თუ Select Cases დიალოგურ ფანჯარაში მოვნიშნავთ ღილაკს All cases.

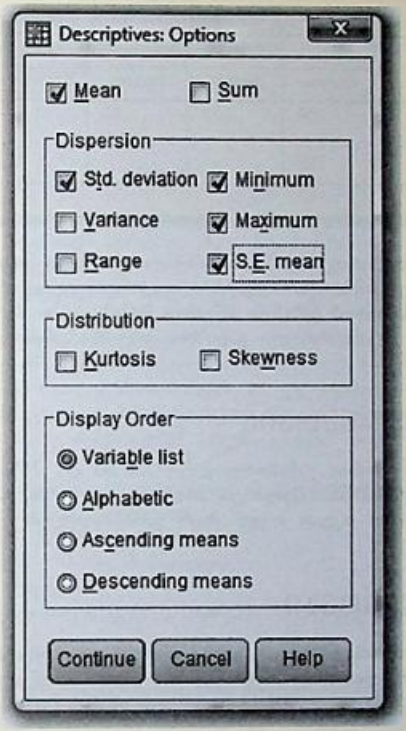
როდესაც SPSS ჩვენ მოთხოვნას დაამუშავებს, ის გვეტყვის, რომ მონაცემები გაიფილტრა. აღნიშნულის მინიშნება ხდება წარწერით „Filter On" SPSS-ის ფანჯრის სტატუსების არის ქვედა ნაწილში (ამ არეში ბევრი საჭირო მესიჯი განთავსდება SPSS-ისგან).

თვალსაჩინოებისთვის განვიხილოთ ცვლადი HRS1, რომელიც ზომავს წინა კვირაში ნამუშევარი საათების რაოდენობას. დაკლიკეთ Analyze, Descriptive Statistics, რომ გაიხსნას დიალოგური ფანჯარა. ცვლადების ჩამონათვალში ჩასვით HRS1. დაკლიკეთ Options ღილაკი ნახაზზე 7.9 გამოსახული ფანჯრის მისაღებად. მონიშნეთ საშუალო, სტანდარტული გადახრა, მინიმუმი და მაქსიმუმი. დავამატოთ სტანდარტული შეცდომა S.E. mean box-ზე დაკლიკებით. შემდეგ დაკლიკეთ Continue და Ok.

შედეგში (ნახაზი 7.10) ნაჩვენებია, რომ ვალიდური დაკვირვებების რაოდენობა 84-ია, ანუ მთელი ფაილის 10%. HRS1-ის საშუალო 41.95-ია, ხოლო საშუალოს სტანდარტული შეცდომა – 1.613.

თუ ამავე პროცესს გავიმეორებთ 25%-იანი შერჩევისთვის, შედეგი მიიღებს შემდეგ სახეს (ნახაზი 7.11). თქვენი შედეგები შეიძლება ჩვენსას არ დაემთხვეს, რადგან SPSS შემთხვევით შერჩევას ასრულებს.

ნახაზი 7.9 დიალოგური ფანჯარა Descriptive Statistics



ნახაზი 7.10 აღწერითი სტატისტიკა წინა კვირაში ნამუშევარი საათებისთვის, 10%-იანი შერჩევა

Descriptive Statistics

	N	Minimum	Maximum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
NUMBER OF HOURS WORKED LAST WEEK	84	4	89	41.95	1.613	14.784
Valid N (listwise)	84					

ნახაზი 7.11 აღწერითი სტატისტიკა წინა კვირაში ნამუშევარი საათებისთვის, 25%-იანი შერჩევა

Descriptive Statistics

	N	Minimum	Maximum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
NUMBER OF HOURS WORKED LAST WEEK	206	1	80	41.43	.972	13.953
Valid N (listwise)	206					

რამდენად მიუახლოვდა ორივე შერჩევის საშუალო მთლიანი ერთობლიობის საშუალოს? 894 რესპონდენტის (დანარჩენ 606 რესპონდენტს არ ჰქონდა ვალიდური პასუხი) საშუალო 42.26 წელია; ამიტომ სარწმუნოა, რომ SPSS-მა სწორედ შემთხვევითი შერჩევა განახორციელა. ორივე შერჩევის საშუალო და სტანდარტული გადახრა მთლიანი ერთობლიობის პარამეტრების საზღვრებში მოხვდა.

ამოცანები SPSS-ისთვის

GSS08PFP-A-ს გამოყენებით განახორციელეთ 25, 50 და 75%-იანი შერჩევები ცვლადებისთვის MAEDUC და PAEDUC. შეადარეთ მთლიანი და შერჩევითი სტატისტიკის აღწერითი მონაცემები. რა შეგიძლიათ თქვათ თქვენ მიერ განხორციელებული შემთხვევითი შერჩევების სიზუსტეზე?

თავის სავარჯიშოები

1. ახსენით ჩამოთვლილთაგან რომელია სტატისტიკა და რომელი პარამეტრი.
ა. ამერიკელების საშუალო ასაკი 2010 წლის ათწლიანი აღწერიდან.
ბ. მთავრობის მიერ დიდი ბაზიდან განსაზღვრული აშშ-ის მოზრდილი მოსახლეობის უმუშევრობის დონე.
გ. 1,000 რეზიდენტისაგან შემდგარი შერჩევიდან ჯანდაცვის რეფორმის მოწინააღმდეგე ტეხასელების ხვედრითი წილი.
დ. დაქირავებულთა საშუალო ხელფასი თქვენ სკოლაში (მაგ. ადმინისტრაცია, ფაკულტეტი, ტექნიკური მუშაკები).
ე. ფინანსური დახმარების მიმღები სტუდენტების ხვედრითი წილი.
2. თქვენი ქალაქის მერი გადასახადების მომატებაზე ალაპარაკდა. საქალაქო გაზეთი საზოგადებრივი აზრის დასაფიქსირებლად იყენებს რედაქციაში მოქალაქეებისგან მისულ წერილებს.
ა. როგორ ფიქრობთ, ეს წერილები შემთხვევითი შერჩევის მაგალითია? რატომ დიახ, ან არა?
ბ. შერჩევის რომელ ალტერნატულ მეთოდს შესთავაზებდით მერს?
3. ქვემოთ მოტანილი 4 სიტუაცია ახასიათებს შერჩევის ორგანიზებას და შერჩევისა და მთლიანი ერთობლიობის ურთიერთდამოკიდებულებას.
ა. თქვენი მეგობარი საავჭრო ცენტრის მთავარ გასასვლელთან იღებს ინტერვიუს ყოველი მე-10 გამოსული მყიდველისაგან. რომელი სახის შერჩევას ატარებს ის? როგორ შეიძლება მთლიანი ერთობლიობის განსაზღვრა, რომლიდანაც ის შერჩევას ახორციელებს?
ბ. პოლიტიკური არჩევნების შემსწავლელი ფირმა თითოეული შტატიდან არჩევს არჩევნებში

დარეგისტრირებულ 50 რესპონდენტს მომავალი არჩევნების შესახებ აზრის შესასწავლად. რა სახისაა ეს შერჩევა? საკმარისია თუ არა ინფორმაცია აღნიშნულის განსაზღვრისათვის?

ბ. პოლიტიკური არჩევნების შემსწავლელი სხვა ფირმა თითოეული შტატიდან განსხვავებული წესით არჩევს რესპონდენტებს. პირველად ის ალფაბეტური თანმიმდევრობით ალაგებს მონაწილეთა გვარებს, შემდეგ ა-ზე დაწყებულ პირველ 20 გვარს ამოიღებს, შემდეგ ბ-დან ასევე 20-ს და ა.შ., ჰ-ს ჩათვლით (ამიტომ შერჩევის ზომა იქნება 20×33, ან 660). არის თუ არა ეს ალბათური განაწილება?

დ. სოციოლოგი დაკვირვებით არჩევს 20 ადამიანისგან შემდგარ შერჩევას ნიუ-იორკის მოსახლეობის ფართო ფენების კვლევის მიზნით. ჩაღრმავებული ინტერვიუს საშუალებით მას სურს გაარკვიოს ქალაქში ჩამოყალიბებული რასობრივი ურთიერთობები. ეს შერჩევა ალბათურია? რა სახის შერჩევას იყენებს ის?

4. ქალაქის უნივერსიტეტის ზედა დონის სოციოლოგიის კლასში 120 სტუდენტია, მათ შორის 34 მეოთხეკურსელი, 57 მესამეკურსელი, 22 მეორეკურსელი და 7 პირველკურსელი.

ა. ვთქვათ ამ კლასიდან შეარჩიეთ მხოლოდ ერთი სტუდენტი. რა ალბათობაა, რომ ეს სტუდენტი მესამეკურსელი იქნება?

ბ. რა ალბათობაა, რომ ეს სტუდენტი პირველკურსელი იქნება?

გ. თუ სტარტიფიცირებული 30 კაციანი შერჩევის ჩატარებას აპირებთ, თითოეული კურსიდან რამდენი სტუდენტი მოხვდება შერჩევაში?

დ. თუ საჭიროა 20 კაციანი არაპროპორციული შერჩევის ჩატარება თითოეული კურსიდან სტუდენტების თანაბარი რაოდენობით, რამდენი პირველკურსელი მოხვდება შერჩევაში?

5. შეიძლება თუ არა, რომ ცვლადის სტანდარტული შეცდომა იმავე ცვლადის სტანდარტული გადახრის ტოლი ან მასზე მეტი იყოს? პასუხი გაეცით როგორც ფორმულის, ისე მსჯელობის სახით.

6. როგორ შეიცვლება საშუალოს სტანდარტული შეცდომა, თუ დიდი ერთობლიობიდან შემთხვევით შერჩევას განახორციელებთ და

ა. შერჩევის ზომა 100-დან 1,500-მდე გაიზრდება?

ბ. შერჩევის ზომა 300-დან 150-მდე შემცირდება?

გ. შერჩევის ზომა გაორმაგდება?

7. ახალი ამბების სატელევიზიო სააგენტო ატარებს სწრაფ გამოკითხვებს 800 სატელეფონო ნომერზე დარეკვით.

ა. არის თუ არა ეს ალბათური შერჩევა? რატომ დიახ, ან არა?

ბ. დაახასიათეთ მთლიანი ერთობლიობა, საიდანაც მიიღება შერჩევა.

8. ცხრილში მოტანილია შეიარაღებულ ძალებში დასაქმებულთა რაოდენობა შტატების მიხედვით (კოლუმბიის ოლქის ჩათვლით) 2006 წელს.

ალაბამა	9,742	მონტანა	3,589
ალასკა	20,363	ნებრასკა	6,784
არიზონა	21,997	ნევადა	9,127
არკანზასი	4,905	ნიუ-ჰემპშირი	912
კალიფორნია	149,481	ნიუ-ჯერსი	6,293
კოლორადო	29,932	ნიუ-მექსიკო	10,834
კონექტიკუტი	6,594	ნიუ-იორკი	26,240
დელავერი	3,297	ჩრდ. კაროლინა	102,845
კოლუმბიის ოლქი	12,351	ჩრდ. დაკოტა	7,013
ფლორიდა	58,100	ოჰაიო	6,745
ჯორჯია	68,928	ოკლაჰომა	25,064
ჰავაი	45,366	ორეგონი	558
აიდაჰო	4,042	პენსილვანია	2,979
ილინოისი	24,536	როდ-აილენდი	2,403

ინდიანა	805	სამხ. კაროლინა	38,090
აიოვა	390	სამხ. დაკოტა	3,150
კანზასი	17,645	ტენესი	2,441
კენტუკი	38,799	ტეხასი	119,176
ლუიზიანა	15,069	იუტა	5,188
მეინი	2,096	ვერმონტი	70
მერილენდი	29,626	ვირჯინია	128,515
მასაჩუსეტსი	2,175	ვაშინგტონი	49,887
მიჩიგანი	1,073	დას. ვირჯინია	373
მინესოტა	729	ვისკონსინი	449
მისისიპი	10,158	ვაიომინგი	3,043
მისური	16,241		

წყარო: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 496.

ა. გაიანგარიშეთ მთლიანი ერთობლიობისთვის საშუალო და სტანდარტული გადახრა.

ბ. გააკეთეთ 5 კაციანი 10 შერჩევა. გამოიყენეთ შემთხვევითი ან სისტემატური შერჩევის მეთოდი. ისარგებლეთ შემთხვევითი რიცხვების ცხრილით დანართში A. თითოეული შერჩევისთვის გაიანგარიშეთ საშუალო.

გ. რადგან უკვე გაიანგარიშეთ თითოეული შერჩევისთვის საშუალო, ახლა გაიანგარიშეთ საშუალოების საშუალო, ანუ შეკრიბეთ საშუალოები და გაყავით 10-ზე. შეადარეთ მისი მნიშვნელობა ყველა შტატის მიხედვით გაანგარიშებულ საშუალოს.

დ. შეადარეთ 8ა-ში გაანგარიშებული სტანდარტული გადახრის მაჩვენებელი სტანდარტული შეცდომის სიდიდეს (შერჩევითი განაწილების სტანდარტულ გადახრას).

ე. ააგეთ ორი ჰისტოგრამა. ერთი მთლიანი ერთობლიობის სიდიდეების განაწილებისთვის და მეორე 8ბ-დან აღებული სხვადასხვა შერჩევების საშუალოების განაწილებისთვის. აღწერეთ და ახსენით განსხვავებები ამ ორ განაწილებას შორის.

ვ. ძალიან მნიშვნელოვანია იცოდეთ მთლიანი ერთობლიობის განმარტება ამ სავარჯიშოზე მუშაობისას. რა არის მთლიანი ერთობლიობა?

9. თქვენ დაგავალეს შეისწავლოთ სტუდენტების ხვედრითი წილი, ვინც მხარს უჭერს მამაკაცებს შორის ერთსქესიან ქორწინებას. თქვენ გინდათ თანაკურსელი სტუდენტების შემთხვევითი შერჩევის გაკეთება. ახსენით, ჩამოთვლილთაგან რომელია შემთხვევითი შერჩევა:

ა. თქვენ გამოკითხავთ კაფეტერიაში 12:30-ზე ლანჩზე შემოსულ სტუდენტებს ლანჩზე.

ბ. თქვენ გამოკითხავთ ყოველ მე-10 სტუდენტს საერთო სიიდან.

გ. თქვენ გამოკითხავთ ყოველ მე-10 სტუდენტს სტუდენტთა კავშირის სიიდან.

დ. შერჩევის რომელ მეთოდს გაუწევთ რეკომენდაციას სრული კვლევის ჩასატარებლად?

10. 2008 წლის მონაცემებით ვისკონსინში ოჯახების საშუალოწლიური შემოსავალი $64,750-ია, ხოლო სტანდარტული გადახრა – $59,750.

ა. ვთქვათ, შერჩევა გააკეთეთ შტატის 200 რესპონდენტზე. როგორის ალბათობა, რომ შერჩევის საშუალო $61,000-სა და $64,700-ს შორის იქნება?

ბ. რესპონდენტების იმავე რაოდენობისთვის, როგორია ალბათობა, რომ შერჩევის საშუალო 75,000 დოლარს გადააჭარბებს?

11. მცირე მთლიანი ერთობლიობის $N = 10$ მნიშვნელობებია 4, 7, 2, 11, 5, 3, 4, 6, 10 და 1.

ა. გაიანგარიშეთ საშუალო და სტანდარტული გადახრა.

ბ. აარჩიეთ სამწევრიანი 10 შერჩევა და თითოეულისთვის გაიანგარიშეთ საშუალო.

გ. გაიანგარიშეთ ყველა ამ შერჩევის საშუალო და საშუალოს სტანდარტული გადახრა. რამდენად უახლოვდება ეს საშუალო მთლიანი ერთობლიობის საშუალოს? სტანდარტული გადახრა?

12. ვთქვათ, მუშაობთ მთლიან ერთობლიობაზე, რომელიც 21,473 რესპონდენტს მოიცავს.

თითოეულს მიანიჭეთ ნომერი 0-დან 21,473-მდე და შერჩევის გაკეთებას ცდილობთ შემთხვევითი რიცხვების ცხრილიდან, დანართი ა. დაიწყეთ მე-7 სვეტიდან, პირველი ხაზი. დაასახელეთ პირველი 5 რესპონდენტი, რომელიც შევა ამ შერჩევაში.

13. ცხრილში მოტანილია საგარეო ვალის სიდიდე 2007 წელს შვიდი ქვეყნისთვის (მლნ. აშშ დოლ.)

ქვეყანა	ვალი (მლნ. აშშ დოლარი)
ბრაზილია	237,500
ჩილე	58,600
კოლუმბია	45,000
ინდოეთი	221,000
მექსიკა	178,100
სამხრეთ აფრიკა	43,400
თურქეთი	251,500

წყარო: U.S. Census Bureau, *Statistical Abstract of the United States*, 2010, Table 1367.

ა. დაუშვათ, რომ $\sigma_y = 93,500$. გაიანგარიშეთ სტანდარტული შეცდომა და გაუკეთეთ ინტერპრეტაცია. შეხსენება: ნახეთ სტანდარტული შეცდომის ფორმულა. რადგან უკვე იცით მთლიანი ერთობლიობის სტანდარტული გადახრა, სტანდარტული შეცდომის გაანგარიშება მინიმალურ გამოთვლებს მოითხოვს.

ბ. დაწერეთ ანგარიში, სადაც განიხილავთ: სტანდარტული შეცდომის შედარებას მთლიანი ერთობლიობის სტანდარტულ გადახრასთან, შერჩევითი განაწილების ფორმას და შეთავაზებებს სტანდარტული შეცდომის შემცირებისათვის.

თავი 8

შეფასება

თავის სწავლების მიზნები

- შეფასების არსი და მიზეზები
- ნდობის ინტერვალების შეფასება საშუალოებისთვის
- რისკის არსი და მისი შემცირება
- ნდობის ინტერვალების შეფასება ხვედრითი წილებისთვის

ამ თავში განვიხილავთ მთლიანი ერთობლიობის საშუალოებისა და ხვედრითი წილების (პროპორციების) შეფასების მეთოდებს. აღნიშნული მეთოდები ემყარება მე-7 თავში განხილულ შერჩევითი დაკვირვებისა და დასკვნითი სტატისტიკის პრინციპებს. შერჩევითი განაწილების ცოდნა საშუალებას გვაძლევს შერჩევიდან მიღებული შედეგებით შევაფასოთ მთლიანი ერთობლიობის საშუალოები და ხვედრითი წილები და გავზომოთ მათი საიმედოობა.

მაგალითი 1. საპრეზიდენტო არჩევნების პრაიმერებზე 862 დემოკრატიული ორიენტაციის ამომრჩევლის გამოკითხვებზე დაყრდნობით პოლიტიკური აზრის შემსწავლელი სოციოლოგები ვარაუდობდნენ, რომ სენატორი ბარაქ ობამა ნიუჰემპშირის შტატში 13 პროცენტით მოუგებდა სენატორ ჰილარი კლინტონს. თუმცა მეორე დღეს, როდესაც შედეგები გამოქვეყნდა, ყველამ დაინახა, რომ კლინტონმა 3 პროცენტით მეტი ხმა მიიღო. ეს წინასწარმეტყველება გამოკითხვების ისტორიაში ერთ-ერთ ყველაზე დიდ მარცხად არის მიჩნეული.

მაგალითი 2. ყოველთვიურად, აშშ-ის შრომის სტატისტიკის ბიურო ატარებს 50,000 მოზრდილი ამერიკელის გამოკითხვას სამუშაოსთან დაკავშირებულ საკითხებზე. ამ მონაცემებზე დაყრდნობით ხდება უმუშევრობის დონის (უმუშევართა ხვედრითი წილი), საშუალო შემოსავლების, არასრული სამუშაო დღით დასაქმებულების პროცენტისა და უმუშევრობის შემწეობის მქონეთა რაოდენობის

გაანგარიშება. ეს შეფასებები ძალიან მნიშვნელოვანია, რადგან მათ შეუძლიათ ზემოქმედება მოახდინონ საფონდო ბაზრებსა და ფედერალური მთავრობის პოლიტიკაზე.

მაგალითი 3. მას შემდეგ, რა არიზონას შტატში წინააღმდეგობრივი საიმიგრაციო კანონი მიიღეს, ამ საკითხზე საზოგადოებრივი აზრის შესასწავლად გელაპის კვლევითმა ორგანიზაციამ სატელეფონო გამოკითხვა ჩაატარა. აღმოჩნდა, რომ 1,013 რესპონდენტიდან 51% მხარს უჭერდა კანონს, ხოლო 39% – ეწინააღმდეგებოდა (წყარო: „More Americans Favor Than Oppose Arizona Immigration Law," Gallup Poll", April 27–28, 2010).

მაგალითი 4. საზოგადოებრივი აზრის კვლევის ეროვნული ცენტრი (NORC) ორ წელიწადში ერთხელ ატარებს 1,500 რესპონდენტის გამოკითხვას (GSS). ეს გამოკითხვა სოციალურ სფეროში მომუშავე მკვლევარებს აწვდის ინფორმაციას აშშ-ის მოზრდილი მოსახლეობის სოციალურად მნიშვნელოვან შეხედულებებზე, ქცევებსა და მახასიათებლებზე. მაგალითად, GSS 2008-ის კვლევებიდან აღმოჩნდა, რომ რესპონდენტების სწავლის საშუალო ხანგრძლივობა 13.43 წელია. ეს საშუალო ალბათ მთლიანი ერთობლიობის საშუალო მახასიათებლისგან განსხვავებულია. მიუხედავად ამისა, ჩვენ შეიძლება დავადგინოთ, რომ უმეტეს შემთხვევებში შერჩევისა და მთლიანი ერთობლიობის საშუალო სიდიდეები საკმაოდ ერთნაირია.

ზემოაღნიშნულის წაკითხვის შემდეგ შეიძლება დაიბადოს კითხვა ზოგიერთი რიცხვის სანდოობის შესახებ. დაუშვეს თუ არა შეცდომები მკვლევარებმა, როდესაც არასწორი პროგნოზი გააკეთეს წინასაარჩევნო პერიოდში? როგორია უმუშევრობის ნამდვილი დონე აშშ-ში? რა შეხედულება აქვს აშშ-ის მოსახლეობას არიზონას საიმიგრაციო კანონზე? როგორია სინამდვილეში აშშ-ში განათლების დონე?

შეფასების განსაზღვრა

ნიუ-ჰემპშირის პრაიმერის ოფიციალური შედეგები, უმუშევარ ამერიკელთა ხვედრითი წილი, ამერიკელების შეხედულება არიზონას ახალ კანონზე და ამერიკელთა განათლების დონე, ყველა ეს მაჩვენებელი მთლიანი ერთობლიობის პარამეტრია. წინასაარჩევნო პროგნოზები, შრომის სტატისტიკის ბიუროს მიერ გაანგარიშებული უმუშევრობის დონე, გელაპის კვლევის შედეგები საიმიგრაციო კანონზე და GSS-ის მიერ გაანგარიშებული განათლების საშუალო დონე, ყველა ეს მაჩვენებელი მთლიანი ერთობლიობის პარამეტრების შერჩევითი შეფასებებია. შერჩევის გაზომვები მთლიანი ერთობლიობის პარამეტრების გასაანგარიშებლად გამოიყენება; GSS-ის კვლევიდან მიღებული შედეგები განათლებაზე დახარჯული წლების საშუალო რაოდენობის შესახებ შეიძლება გამოვიყენოთ მთლიანად აშშ-ის მოზრდილი მოსახლეობის განათლების დონის შესაფასებლად. ასევე, მოზრდილი ამერიკელების ეროვნული შერჩევის ბაზაზე შეიძლება შევაფასოთ ამერიკელების დამოკიდებულება წინააღმდეგობრივი საიმიგრაციო კანონის შესახებ.

ყველა ზემოთ მოყვანილი მაგალითი შეფასებას წარმოადგენს. **შეფასება** ეს არის პროცესი, რომლის საშუალებით ვაკეთებთ მთლიანი ერთობლიობიდან შემ-

თხვევით შერჩევას და შემდეგ შერჩევის სტატისტიკას ვიყენებთ მთლიანი ერთობლიობის პარამეტრის შესაფასებლად. ჩვენ შეგვიძლია მთლიანი ერთობლიობის ხვედრითი წილების შესაფასებლად შერჩევის ხვედრითი წილების გამოყენება. შერჩევის საშუალოების გამოყენება მთლიანი ერთობლიობის საშუალოების შესაფასებლად და შერჩევის ვარიაციის საზომებისა — მთლიანი ერთობლიობის ვარიაციის შესაფასებლად.

შეფასება ეს არის პროცესი, რომლის საშუალებით ვაკეთებთ მთლიანი ერთობლიობიდან შემთხვევით შერჩევას და შემდეგ შერჩევის სტატისტიკას ვიყენებთ მთლიანი ერთობლიობის პარამეტრის შესაფასებლად.

შეფასების მიზეზები

რატომ უნდა შევაფასოთ? კვლევების უმრავლესობის ძირითადი მიზანია მთლიანი ერთობლიობის პარამეტრის პოვნა. ჯერ-ჯერობით ჩვენ თითქმის არასოდეს არ გაგვაჩნია საკმაო რესურსები მთლიანი ერთობლიობის შესახებ მონაცემების შეგროვებისთვის. ასევე იშვიათად ვიცით მთლიანი ერთობლიობის პარამეტრი. მეორე მხრივ, ჩვენ ბევრი რამის გაგება შეგვიძლია მთლიან ერთობლიობაზე, თუ შემთხვევით შევარჩევთ შერჩევით ერთობლიობას და მივიღებთ მთლიანი ერთობლიობის პარამეტრის შეფასებას. შერჩევის თეორიისა და სტატისტიკური დასკვნების მთავარი მიზანია შერჩევის სტატისტიკიდან მთლიანი ერთობლიობის პარამეტრის შეფასების გაკეთება.

წერტილოვანი და ინტერვალური შეფასება

მთლიანი ერთობლიობის მახასიათებლების შეფასება ორად იყოფა: წერტილოვანი შეფასებები და ინტერვალური შეფასებები. **წერტილოვანი შეფასებები** შერჩევის სტატისტიკაა, რომელიც გამოიყენება მთლიანი ერთობლიობის პარამეტრის ზუსტი სიდიდის შესაფასებლად. როდესაც გელაპის ორგანიზაცია აქვეყნებს ანგარიშს იმის შესახებ, რომ ამერიკელების 51% მხარს უჭერს არიზონას საიმიგრაციო კანონს, ის წერტილოვან შეფასებას იყენებს. ასევე, თუ ვიტყვით, რომ მოზრდილი ამერიკელების სწავლის საშუალო ხანგრძლივობა 13.43 წელია, ჩვენ გამოვიყენებთ წერტილოვან შეფასებას.

წერტილოვანი შეფასება შერჩევის სტატისტიკაა, რომელიც გამოიყენება მთლიანი ერთობლიობის პარამეტრის ზუსტი სიდიდის შესაფასებლად.

წერტილოვანი შეფასებების პრობლემა იმაშია, რომ შერჩევითი შეფასებები ჩვეულებრივ ცვლებადია, რაც შერჩევის სხვადასხვა სახის შეცდომაში აისახება. შედეგად, როდესაც ერთობლიობის პარამეტრის დასახასიათებლად შერჩევის სტატისტიკას ვიყენებთ, რეალურად არასოდეს არ ვიცით, თუ რამდენად ზუსტია ის.

წერტილოვანთან შედარებით, სიზუსტის გაზრდის თვალსაზრისით, უკეთესია ინტერვალური შეფასების გამოყენება. ამ შემთხვევაში ვსაზღვრავთ სიდიდეების დიაპაზონს, რომელშიც მთლიანი ერთობლიობის პარამეტრი მოხვდება. სიდიდეების ამ დიაპაზონს **ნდობის ინტერვალი** ეწოდება (CI). ერთი სიდიდით (13.43 წელი) განათლების დონის დახასიათების ნაცვლად შეგვიძლია ვთქვათ, რომ მთლიანი ერთობლიობის საშუალო სადღაც 12-სა და 14 წელს შორისაა.

ნდობის ინტერვალი (CI) ნდობის ხარისხით განსაზღვრული სიდიდეების დიაპაზონი, რომელშიც მოხვდება მთლიანი ერთობლიობის პარამეტრი. ზოგჯერ ნდობის ინტერვალს ცდომილების ზღვარს უწოდებენ.

როდესაც ნდობის ინტერვალს მთლიანი ერთობლიობის პარამეტრის შესაფასებლად ვიყენებთ, როგორიცაა მაგალითად განათლების საშუალო დონე, ჩვენ აგრეთვე შეგვიძლია გამოვთვალოთ ამ შეფასების სიზუსტეც, თუ გავიანგარიშებთ ალბათობას, რომ საშუალო სიდიდე რომელიმე ინტერვალში მოხვდება. ამ ალბათობას, რომელიც ან პროცენტობით არის გამოსახული, ან ალბათობით, **ნდობის ხარისხი** (სანდოობა) ეწოდება. ნდობის ინტერვალები ნდობის ხარისხით გამოისახება. ამიტომ, თუ ნდობის ხარისხად შევარჩევთ 95%-ს, შეგვიძლია ვთქვათ, რომ 0.95 ალბათობით – ან 100-დან 95 შემთხვევაში – მთლიანი ერთობლიობის საშუალო სიდიდე განსაზღვრულ ინტერვალში მოხვდება. ნდობის ინტერვალები ნებისმიერ დონეზე შეიძლება ავაგოთ, თუმცა ყველაზე გავრცელებულია 90%, 95% და 99%-იანი დონეები. ზოგჯერ ნდობის ინტერვალებს ცდომილების ზღვრითაც გამოსახავენ. ცდომილების ზღვარი მარტივად ნდობის ინტერვალის რადიუსს წარმოადგენს. თუ ჩვენ შევარჩევთ 95%-იან ნდობის ინტერვალს, ცდომილება 5 პროცენტი იქნება.

ნდობის ხარისხი ალბათობა, რომელიც გამოსახულია პროცენტობით, ან ალბათობით, რომ მთლიანი ერთობლიობის პარამეტრი განსაზღვრულ ინტერვალში მოხვდება.
ცდომილების ზღვარი ნდობის ინტერვალის რადიუსი.

ნდობის ინტერვალების აგება შესაძლებელია მრავალი განსხვავებული პარამეტრისთვის მათი შესაბამისი შერჩევის სტატისტიკიდან. ამ თავში ჩვენ განვიხილავთ მიზეზებსა და პროცედურებს საშუალოებისა და პროპორციების ნდობის ინტერვალების ასაგებად.

ცოდნის შემოწმება

რა განსხვავებაა წერტილოვან და ინტერვალურ შეფასებებს შორის?

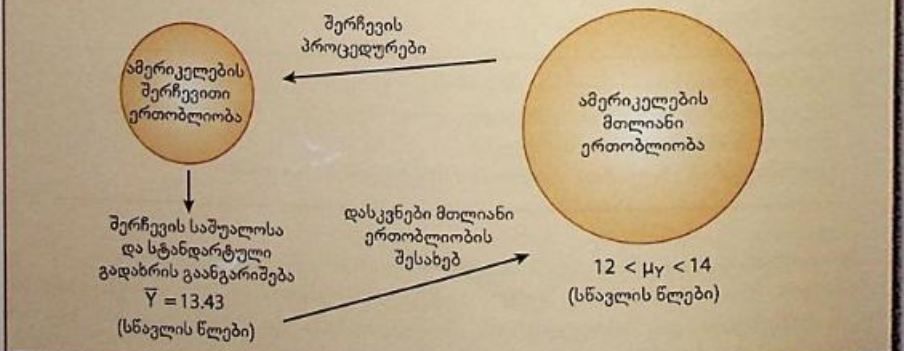
ნდობის ინტერვალების შეფასების პროცედურები საშუალოებისთვის

ნდობის ინტერვალების შეფასების პროცედურების თვალსაჩინოებისთვის გავიხსენოთ საუნივერსიტეტო ქალაქის გარეთ მცხოვრები სტუდენტების მაგალითი 7-ე თავიდან.

როგორც გახსოვთ, ჩვენ მხოლოდ 500 სტუდენტის გამოკვლევის საშუალება გვქონდა. ჩვენი მიზანია შევაფასოთ 15,000 სტუდენტის მგზავრობისთვის საჭირო საშუალო დრო – მთლიანი ერთობლიობის პარამეტრი. ამ შეფასების მისაღებად ვანგარიშობთ შერჩევის საშუალოს. დაუშვათ, შერჩევის საშუალო არის $\bar{Y} = 7.5$ საათი კვირის განმავლობაში და გვინდა ის გამოვიყენოთ სტუდენტების მთლიანი ერთობლიობის მგზავრობის რეალური საშუალო ხანგრძლივობის შესაფასებლად. რადგან ეს მონაცემი შერჩევიდან მიიღება, ის შერჩევის შეცდომის ობიექტია. თუმცა ცენტრალური ზღვრის თეორემიდან ჩვენთვის ცნობილია, რომ შერჩევის საკმაოდ დიდი ზომის შემთხვევაში შერჩევის საშუალო მთლიანი ერთობლიობის საშუალოს უახლოვდება. ამიტომ არ არის მოსალოდნელი, რომ შერჩევის საშუალო $\bar{Y} = 7.5$ საათი კვირის განმავლობაში, ნამდვილი ერთობლიობის საშუალოდან ძალიან გადაიხრება.

დეტალების განხილვა 8.1
შეფასება, როგორც დასკვნის სახეობა

დასკვნითი სტატისტიკის ძირითადი მიზანია შერჩევითი ერთობლიობის განხილვის საფუძველზე რაიმე მნიშვნელოვანის თქმა მთლიან ერთობლიობის შესახებ. ნდობის ინტერვალი სწორედ ამისთვისაა: შერჩევის საშუალოს, შერჩევის ზომისა და სტანდარტული გადახრის საფუძველზე შეფასების გასაკეთებლად იმ მთლიანი ერთობლიობის შესახებ, საიდანაც მოხდა შერჩევა.



თუ ჩვენთვის ცნობილია შერჩევის საშუალოს სიდიდე სტანდარტულ გადახრასა და შერჩევის ზომასთან ერთად, მათი კომბინაციის საფუძველზე შეიძლება განისაზღვროს ის დიაპაზონი, რომელშიც უეჭველად მოხვდება მთლიანი ერთობლიობის საშუალო.

ჩვენთვის ცნობილია, რომ საშუალოს შერჩევითი განაწილება დაახლოებით ნორმალურია, როდესაც მთლიანი ერთობლიობის საშუალო μ_Y შერჩევის საშუალო სიდიდის $\mu_{\bar{Y}}$-ს ტოლია, ხოლო სტანდარტული შეცდომა $\sigma_{\bar{Y}}$ (შერჩევითი განაწილების სტანდარტული გადახრა) ტოლია:

$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{N}} \tag{8.1}$$

ეს ინფორმაცია საშუალებას გვაძლევს გამოვიყენოთ ნორმალური განაწილება და განვსაზღვროთ ალბათობა, რომ შერჩევის საშუალო $\mu_{\bar{Y}}$-დან ან μ_Y-დან გარკვეული დაშორებით – რომელიც სტანდარტული გადახრის (სტანდარტული შეცდომის) ერთეულებში ან Z ქულებით იზომება – მოხვდება. აღნიშნულის საფუძველზე ჩვენ შეიძლება დავასკვნათ, რომ:

- შერჩევის საშუალოების 68% მოხვდება მთლიანი ერთობლიობის საშუალოდან 1 სტანდარტული გადახრის ფარგლებში.
- შერჩევის საშუალოების 95% მოხვდება მთლიანი ერთობლიობის საშუალოდან 2 სტანდარტული გადახრის ფარგლებში.
- შერჩევის საშუალოების 99% მოხვდება მთლიანი ერთობლიობის საშუალოდან 3 სტანდარტული გადახრის ფარგლებში.

ამ დასკვნებისა და სტანდარტული შეცდომის საფუძველზე შეგვიძლია შევქმნათ სიდიდეების დიაპაზონი – ნდობის ინტერვალი – რომელშიც სავარაუდოდ მოხვდება ნამდვილი ერთობლიობის საშუალო. აგრეთვე შესაძლებელია ამ შეფასების სიზუსტის განსაზღვრა, თუ გავიანგარიშებთ იმის ალბათობას, რომ ნამდვილი ერთობლიობის საშუალო ამ დიაპაზონში მოხვდება.

ნდობის ნებისმიერი ხარისხისთვის ნდობის ინტერვალის (*CI*) მთავარი გასაანგარიშებელი ფორმულაა:

$$CI = \bar{Y} \pm Z\,(\sigma_{\bar{Y}}) \tag{8.2}$$

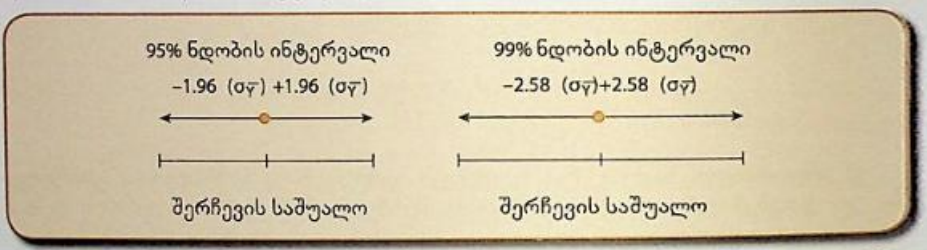
ნდობის ინტერვალის გასაანგარიშებლად შერჩევით საშუალოს ვუმატებთ ან ვაკლებთ Z ქულისა და სტანდარტული შეცდომის ნამრავლს. Z ქულის შერჩევა სასურველ ნდობის ხარისხთან არის დაკავშირებული. მაგალითად, 95%-იანი ნდობის ინტერვალისთვის Z-ის მნიშვნელობად ავირჩევთ 1.96-ს, რადგან ჩვენთვის ცნობილია (დანართი ბ), რომ მრუდის ქვეშ ფართობის 95% ±1.96-ს შორისაა. ამის მსგავსად, 99%-იანი ნდობის ინტერვალისთვის Z-ის მნიშვნელობად

ავირჩევთ 2.58-ს. ეს ურთიერთობა ნდობის ხარისხსა და Z-ს შორის ასახულია ნახაზზე 8.1, 95% და 99% ნდობის ალბათობისთვის.

✓ ცოდნის შემოწმება

ნდობის ხარისხსა და Z-ს შორის ურთიერთობის გასარკვევად დაუბრუნდით მე-6 თავს. რა იქნება 98%-ის შესაბამისი Z ქულა?

ნახაზი 8.1 ურთიერთდამოკიდებულება ნდობის ხარისხსა და Z-ს შორის 95% და 99%-იანი ნდობის ინტერვალებისთვის



წყარო: David Freedman, Robert Pisani, Roger Purves, and Ani Akhikari, Statistics, 2nd ed. (New York: W. W. Norton, 1991)- dan. Copyright ©1991 by W. W. Norton & Company, Inc. გამოყენებულია W. W. Norton and Company, Inc-ის ნებართვით.

ნდობის ინტერვალის განსაზღვრა

ნდობის ინტერვალის განსაზღვრისათვის შეასრულეთ შემდეგი მოქმედებები:

1. გაიანგარიშეთ საშუალოს სტანდარტული შეცდომა.
2. განსაზღვრეთ ნდობის ხარისხი და იპოვნეთ მისი შესაბამისი Z სიდიდე.
3. გაიანგარიშეთ ნდობის ინტერვალი.
4. მოახდინეთ შედეგების ინტერპრეტაცია.

ახლა ისევ დაუბრუნდით ქალაქგარეთ მცხოვრები სტუდენტების საკითხს და განსაზღვრეთ მათი მგზავრობის საშუალო ხანგრძლივობა. როგორ უნდა იპოვნოთ 95%-იანი ნდობის ინტერვალი?

საშუალოს სტანდარტული შეცდომის გაანგარიშება

დაუშვათ, გარეუბნებში მცხოვრები სტუდენტების სტანდარტული გადახრა σ_Y=1.5. გავიანგარიშოთ სტანდარტული შეცდომა საშუალოს შერჩევითი განაწილებისთვის:

$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{N}} = \frac{1.5}{\sqrt{500}} = 0.07$$

ნდობის ხარისხისა და მისი შესაბამისი Z სიდიდის პოვნა

ჩვენ გადავწყვიტეთ ნდობის ხარისხად მივიჩნიოთ 95%. მას შეესაბამება Z სიდიდე 1.96.

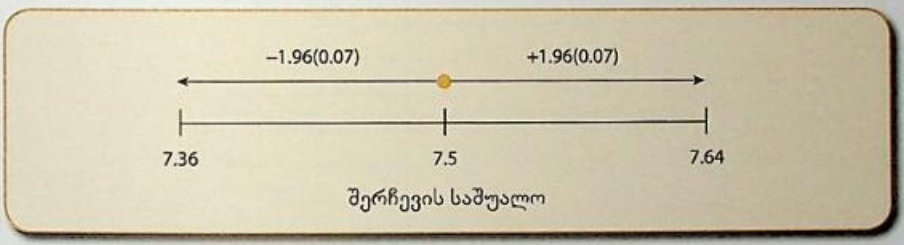
ნდობის ინტერვალის გაანგარიშება

ნდობის ინტერვალის გასაანგარიშებლად შერჩევით საშუალოს ვუმატებთ ან ვაკლებთ Z ქულისა და სტანდარტული შეცდომის ნამრავლს.

$$95\%\ CI = 7.5 \pm 1.96(0.07)$$
$$= 7.5 \pm 0.14$$
$$= 7.36\text{-დან } 7.64\text{-მდე}$$

95%-იანი ნდობის ინტერვალი საშუალოსათვის ნაჩვენებია ნახაზზე 8.2.

ნახაზი 8.2 95%-იანი ნდობის ინტერვალი მგზავრობის საშუალო ხანგრძლივობისათვის ($N = 500$)



შედეგების ინტერპრეტაცია

95%-ით დარწმუნებულები უნდა ვიყოთ, რომ მგზავრობის ნამდვილი საშუალო ხანგრძლივობა 7.36-დან 7.64 საათამდე იქნება. სხვა სიტყვებით, თუ შერჩევის საკმაოდ დიდ რიცხვს ($N = 500$) დავაკვირდებით, 100-დან 95 შემთხვევაში ნამდვილი საშუალო ნდობის ინტერვალს არ გასცდება. 95%-იანი სანდოობის პირობით, მხოლოდ 5%-იანი რისკი არსებობს, რომ ეს ასე არ მოხდება, ანუ რომ 100-დან 5 შემთხვევაში მთლიანი ერთობლიობის საშუალო ამ ინტერვალში არ მოხვდება.

დაიმახსოვრეთ, არ შეიძლება დარწმუნებული ვიყოთ, რომ მთლიანი ერთობლიობის საშუალო უეჭველად ნდობის ინტერვალში მოხვდება. როდესაც შერჩევა გაკეთებულია და ნდობის ინტერვალი ნაპოვნია, არ შეიძლება დანამდვილებით ითქვას, რომ მთლიანი ერთობლიობის საშუალო აუცილებლად მის საზღვრებშია მოთავსებული.

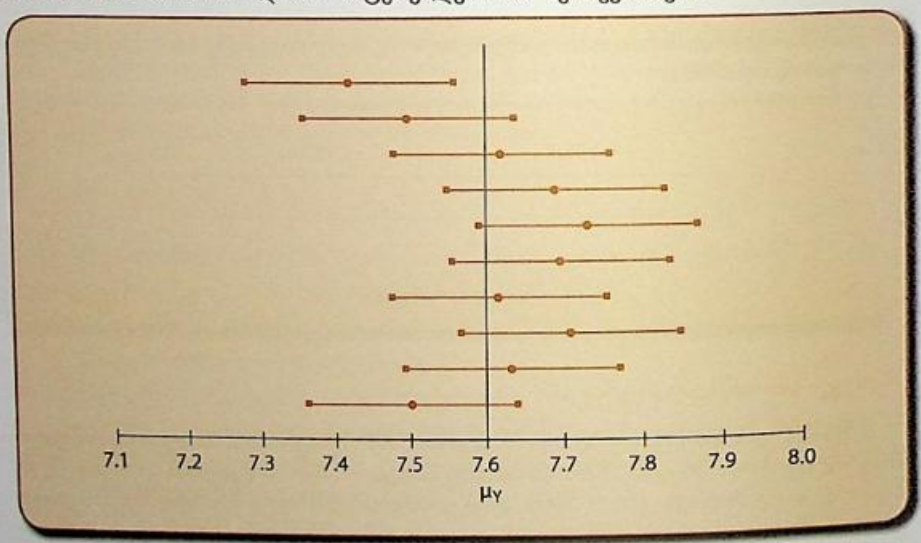
რა იქნება 90%-იანი ნდობის ინტერვალი მგზავრობის საშუალო ხანგრძლივობისათვის? (შეხსენება: პირველად მოძებნეთ ნდობის ხარისხის შესაბამისი Z ქულა)

✓ ცოდნი შემოწმება

ვაგრძელებთ ნდობის ინტერვალის ილუსტრაციას. დაუშვათ, სტუდენტების ერთობლიობიდან ათი სხვადასხვა შერჩევა შევქმენით ($N = 500$) და თითოეული შერჩევის საშუალოსათვის 10 ნდობის ინტერვალი გავიანგარიშეთ. ეს ინტერვალები ასახულია ნახაზზე 8.3. თითოეული ჰორიზონტალური ხაზი ასახავს 95%-იან ნდობის ინტერვალს შერჩევის საშუალოს (აღნიშნულია წრით) ირგვლივ.

ვერტიკალური წრფით გამოსახულია მთლიანი ერთობლიობის საშუალო. ჰორიზონტალური და ვერტიკალური ხაზების გადაკვეთაზე ხვდება ინტერვალები, რომლებიც შეიცავენ ნამდვილი ერთობლიობის საშუალოს. 10-დან მხოლოდ 1 ინტერვალი არ გადაკვეთს ვერტიკალურ ხაზს, რაც იმას ნიშნავს, რომ ის არ შეიცავს ერთობლიობის საშუალოს. რა მოხდება, თუ ჩვენ გავაგრძელებთ ასეთივე ზომის შერჩევების შექმნას და თითოეული შერჩევის საშუალოსათვის ავაგებთ 95%-იან ინტერვალს? შერჩევების 95%-ში იქნება მთლიანი ერთობლიობის საშუალო, ხოლო 5%-ში — არა.

ნახაზი 8.3 95%-იანი ნდობის ინტერვალები 10 შერჩევისთვის



რისკის შემცირება

რისკის შემცირების ერთ-ერთი გზა ნდობის ხარისხის გაზრდაში მდგომარეობს. მაგალითად, ჩვენი ნდობის ინტერვალის სანდოობა შეგვიძლია 95-იდან 99%-მდე გავზარდოთ:

$$99\%\ CI = 7.5 \pm 2.58(0.07)$$
$$= 7.5 \pm 0.18$$
$$= 7.32\text{-დან } 7.68\text{-მდე}$$

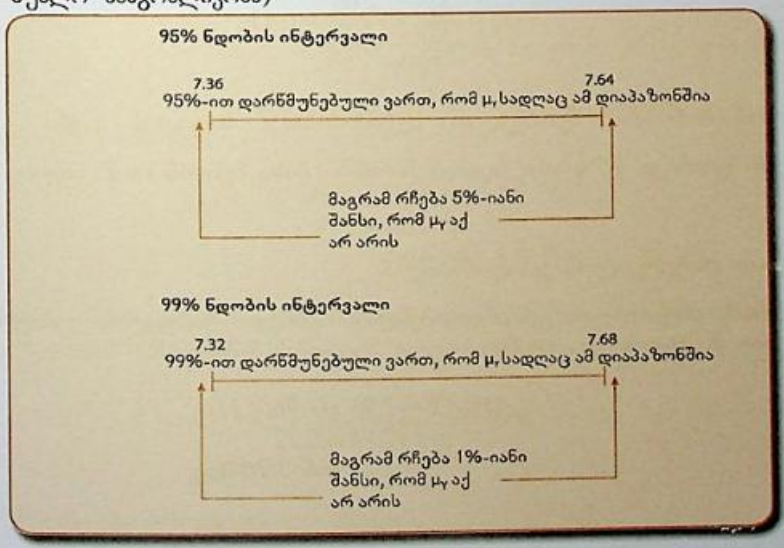
როდესაც 99%-იან ნდობის ინტერვალს ვიყენებთ, რისკი იმისა, რომ ერთობლიობის საშუალო ამ ინტერვალში არ მოხვდება, მხოლოდ 1%-ია. ჩვენ აგრეთვე დარწმუნებული უნდა ვიყოთ, რომ საშუალო 7.32-დან 7.68-მდე ინტერვალში მოხვდება. აღსანიშნავია, რომ ნდობის ხარისხთან ერთად ჩვენ გავზარდეთ აგრეთვე ნდობის ინტერვალის დიაპაზონი 0.28-დან (7.36-7.64) 0.36-მდე (7.32-7.68-მდე), რამაც ჩვენი გაანგარიშება ნაკლებად ზუსტი გახადა.

თქვენ ხედავთ, რომ ჩვენ კომპრომისის ძიებაში ვართ სანდოობასა და სიზუსტეს შორის. რაც უფრო ვზრდით ნდობის ხარისხს (მაგ. 99%), და შესაბამისად ჩვენ მოლოდინს, რომ მთლიანი ერთობლიობის საშუალო ნდობის ინტერვალში მოხვდება, შეფასება ნაკლებად ზუსტი ხდება და ინტერვალის სიგანე იზრდება. ამასთან, 95%-იანი ნდობის შემთხვევაში ნდობის ინტერვალი 7.36-7.64 უფრო ზუსტია, ვიდრე 99%-ის შემთხვევაში – 7.32-7.68. ეს დამოკიდებულება ნდობის ინტერვალებსა და მათ სიზუსტეს შორის ნაჩვენებია ნახაზზე 8.4. ცხრილში 8.1 წარმოდგენილია სამი გავრცელებული ნდობის ხარისხი მათი შესაბამისი Z სიდიდეებით.

ცხრილი 8.1 ნდობის ხარისხები მათი შესაბამისი Z სიდიდეები

ნდობის ინტერვალი	Z სიდიდე
90%	1.65
95%	1.96
99%	2.58

ნახაზი 8.4 ნდობის ინტერვალები, 95% თუ 99% (მგზავრობის საშუალო ხანგრძლივობა)



სიგმას გაანგარიშება

ნდობის ინტერვალების გაანგარიშებისთვის საჭიროა შერჩევითი განაწილების სტანდარტული შეცდომის $\sigma_{\bar{Y}}$-ის პოვნა. როგორც წესი, მთლიანი ერთობლიობისთვის არც საშუალო μ_Y ვიცით და არც სტანდარტული გადახრა σ_Y. როდესაც $N \geq 50$, შერჩევითი ერთობლიობის S_Y კარგი საშუალებაა $\sigma_{\bar{Y}}$-ის შეფასებისთვის. სტანდარტული შეცდომა მაშინ ასე გაიანგარიშება:

$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}}$$

თვალსაჩინოებისთვის განვიხილოთ საშუალოდ ყოველდღიურად ამერიკელების მიერ ტელევიზორთან გატარებული საათების რაოდენობა GSS კვლევის მონაცემებიდან. შერჩევის $N = 652$, ხოლო $\bar{Y} = 2.98$ საათია, სტანდარტული გადახრა S_Y=2.66 საათს. ამ მონაცემებისთვის ნდობის ინტერვალი განვსაზღვროთ 95%-ით.

საშუალოსათვის შეფასებული სტანდარტული შეცდომის გაანგარიშება

შეფასებული სტანდარტული შეცდომა შერჩევითი განაწილების საშუალოსთვის ტოლია:

$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}} = \frac{2.66}{\sqrt{562}} = 0.11$$

ნდობის ხარისხის შერჩევა და შესაბამისი Z სიდიდის პოვნა

ჩვენ ავირჩიეთ 95%-იანი ნდობის ხარისხი. მისი შესაბამისი Z სიდიდე 1.96-ია.

ნდობის ინტერვალის გაანგარიშება

ნდობის ინტერვალის გასაანგარიშებლად შერჩევით საშუალოს ვუმატებთ ან ვაკლებთ Z ქულისა და სტანდარტული შეცდომის ნამრავლს:

$$\begin{aligned} 95\%\ CI &= 2.98 \pm 1.96(0.11) \\ &= 2.98 \pm 0.22 \\ &= 2.76\text{-დან } 3.20\text{-მდე} \end{aligned}$$

შედეგების ინტერპრეტაცია

შეგვიძლია 95%-ით დარწმუნებულები ვიყოთ, რომ GSS-ის კვლევის მონაცემებიდან ყოველდღიურად ამერიკელების მიერ ტელევიზორთან გატარებული საათების საშუალო რაოდენობა არ იქნება 2.76 საათზე ნაკლები და 3.20 საათზე მეტი. სხვა სიტყვებით, თუ მთლიანი ერთობლიობიდან რესპონდენტების საკმაოდ დიდ რაოდენობას შევარჩევთ ($N = 562$), მაშინ 100-დან 95 შემთხვევაში ნამდვილი ერთობლიობის საშუალო ჩვენ მიერ გაანგარიშებულ ინტერვალში მოხვდება.

ერთობლიობის ზომა და ნდობის ინტერვალები

შეფასებების სიზუსტის გაზრდა მკვლევარებს შერჩევის ზომის გადიდებით შეუძლიათ. მე-7 თავში ვისწავლეთ, რომ დიდი ზომის შერჩევებში მცირე სტანდარტული შეცდომებია და ამიტომ შერჩევითი განაწილებები მთლიანი ერთობლიობის საშუალოს ირგვლივ განლაგდებიან (ნახაზი 7.6). შერჩევითი განაწილების ასეთი მჭიდრო კლასტერი იმაზე მიანიშნებს, რომ ნდობის ინტერვალი უფრო შევიწროებულია და უფრო ზუსტია. შერჩევის ზომას, სტანდარტულ შეცდომასა და ნდობის ინტერვალს შორის დამოკიდებულების თვალსაჩინოების მიზნით გავიანგარიშოთ 95%-იანი ნდობის ინტერვალი ჩვენი მონაცემებისთვის GSS-იდან როდესაც: (1) შერჩევის $N = 195$ და (2) შერჩევის $N = 1,987$.

როდესაც შერჩევის $N = 195$, შერჩევითი განაწილების შეფასებული სტანდარტული გადახრა არის:

$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}} = \frac{2.66}{\sqrt{195}} = 0.19$$

ხოლო 95%-იანი ნდობის ინტერვალი ტოლია:

$$\begin{aligned} 95\%\ CI &= 2.98 \pm 1.96(0.19) \\ &= 2.98 \pm 0.37 \\ &= 2.61\text{-დან } 3.35\text{-მდე} \end{aligned}$$

როდესაც შერჩევის $N = 1,987$, შერჩევითი განაწილების შეფასებული სტანდარტული გადახრა არის:

$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}} = \frac{2.66}{\sqrt{1,987}} = 0.06$$

და 95%-იანი ნდობის ინტერვალი ტოლია

$$95\%\ CI = 2.98 \pm 1.96(0.06)$$
$$= 2.98 \pm 0.12$$
$$= 2.86\text{-დან } 3.10\text{-მდე}$$

ცხრილში 8.2 შეჯამებულია 95%-იანი ნდობის ინტერვალები ტელევიზორის ყურების საშუალო ხანგრძლივობისათვის სამი სხვადასხვა ზომის შერჩევაში: $N = 195$, $N = 562$ და $N = 1{,}987$.

ცხრილი 8.2 95%-იანი ნდობის ინტერვალები და მათი სიდიდე სამი სხვადასხვა ზომის შერჩევაში ტელევიზორის ყურების საშუალო ხანგრძლივობისათვის

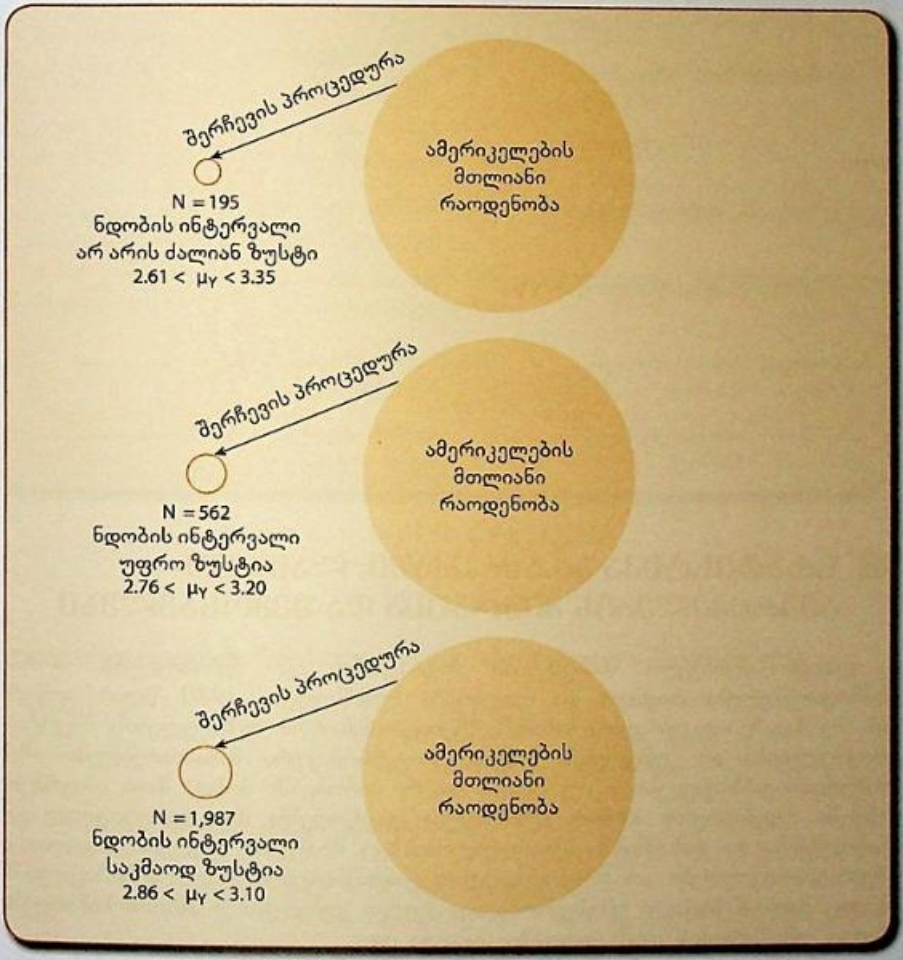
შერჩევის ზომა	ნდობის ინტერვალი	ინტერვალის სიდიდე	S_x	$S_{\bar{x}}$
195	2.61-3.35	0.74	2.66	0.19
562	2.76-3.20	0.44	2.66	0.11
1,987	2.86-3.10	0.24	2.66	0.06

შერჩევის ზომასა და ნდობის ინტერვალის სიდიდეს შორის უკუკავშირი შეიმჩნევა. შერჩევის გაზრდილი ზომა პირდაპირ უკავშირდება ნდობის ინტერვალის გაზრდილ სიზუსტეს. GSS-ის 195 რესპონდენტის შერჩევაში 95%-იანი სიზუსტით ნდობის ინტერვალი 0.74-ის ტოლია, ხოლო ინტერვალის სიდიდე 0.44-დან 0.24-მდე მცირდება, როდესაც შერჩევის ზომა $N = 562$-დან $N = 1{,}987$-მდე იზრდება. ამ შემთხვევაში თითქმის გავაორმაგეთ შერჩევის ზომა (562-დან 1,987-მდე), რომ გაგვენახევრებინა ნდობის ინტერვალი (0.44-დან 0.24-მდე). მიხედავად იმისა, რომ შერჩევის ზომის ზრდასთან ერთად შეფასების სიდიდე განუზრელად იზრდება, თუ ის 1,987-ს მიაღწევს, შედეგები შეიძლება ძალიან მოკრძალებული აღმოჩნდეს. აქ უმნიშვნელოვანესი მომენტია შერჩევის ზომის ზრდასთან ერთად კვლევის დანახარჯების გადიდება, ამიტომ მკვლევარმა უნდა გაითვალისწინოს რა ზომით უნდა გაზარდოს სიზუსტე, რომ ამან დიდი დანახარჯები არ გამოიწვიოს.

✓ **ცოდნის შემოწმება**

რატომ იწვევს შერჩევის ზომის ზრდა ნდობის ინტერვალების გადიდებას? (იხ. ნახაზი 8.5) (შეხსენება: შეადარეთ საშუალოს სტანდარტული შეცდომა სამი სხვადასხვა ზომის შერჩევისთვის).

ნახაზი 8.5 ურთიერთდამოკიდებულება შერჩევის ზომასა და ინტერვალის სიდიდეს შორის



დეტალების განხილვა 8.2
რა ახდენს გავლენას ინტერვალის სიდიდეზე? შეჯამება

„თუ სხვა ფაქტორები უცვლელია...“

თუ შერჩევის ზომა იზრდება	↑	ნდობის ინტერვალი უფრო ზუსტია	→ ←
თუ შერჩევის ზომა მცირდება	↓	ნდობის ინტერვალი ნაკლებად ზუსტია	← →
თუ შერჩევის სტანდარტული გადახრა იზრდება	↑	ნდობის ინტერვალი ნაკლებად ზუსტია	← →
თუ შერჩევის სტანდარტული გადახრა მცირდება	↓	ნდობის ინტერვალი უფრო ზუსტია	→ ←
თუ ნდობის ხარისხი იზრდება (95%-დან 99%-მდე)	↑	ნდობის ინტერვალი ნაკლებად ზუსტია	← →
თუ ნდობის ხარისხი მცირდება (99%-დან 95%-მდე)	↓	ნდობის ინტერვალი უფრო ზუსტია	→ ←

სტატისტიკა პრაქტიკაში: ლათინო-ამერიკელების მიგრაცია და შემოსავლები

ლათინო-ამერიკელი მამაკაცების მიგრაციასა და შემოსავლებს შორის დამოკიდებულება თიენდამ და უილსონმა შეისწავლეს.[1] 1980 წლის აღწერის შერჩევაზე დაკვირვებით ისინი ამტკიცებდნენ, რომ მექსიკელების, პუერტო-რიკოელებისა და კუბელების სოციალურ-ეკონომიკური მახასიათებლები იმის მიხედვით განსხვავდებოდა, თუ როდის და რა პირობებში მოხდა მათი იმიგრაცია აშშ-ში. მეცნიერების აზრით, აღნიშნულმა ფაქტორებმა მათ გეოგრაფიულ განაწილებასა და დასაქმების შესაძლებლობებზეც მოახდინა გავლენა. მაგალითად, პუერტო-რიკოელები არაპროპორციულად განთავსდნენ ჩრდილო-აღმოსავლეთში, სადაც ძალიან მაღალი უმუშევრობა იყო, ხოლო კუბელები — აშშ-ის სამხრეთში, სადაც უმუშევრობის დონე ყველაზე დაბალი იყო.

თიენდამ და უილსონმა შეისწავლეს აგრეთვე ამ ჯგუფების განათლების დონე. კუბელებთან შედარებით პუერტო-რიკოელები და მექსიკელები დაბალი დონით გამოირჩეოდნენ. მაგალითად, საშუალო სკოლა არ ჰქონდა დამთავრებული პუერტო-რიკოელებისა და მექსიკელების 60%-ს, ხოლო კუბელების 42%-ს. ასევე კუბელების 17%-ს კოლეჯი ჰქონდა დამთავრებული, ხოლო პუერტო-რიკოელებისა და მექსიკელების მხოლოდ 4%-ს.

მოსალოდნელი იყო, რომ ეს განსხვავებები აისახებოდა ამ ჯგუფების შემოსავლებზე. თიენდა და უილსონი მოელოდნენ, რომ კუბელების შემოსავლები და-

ნარჩენ ჯგუფებზე მეტი იქნებოდა. ჩვენ მოვახდინეთ ამ მოსაზრების ტესტირება და 2000 წლის მოსახლეობის აღწერიდან შევარჩიეთ 29,233 კუბელი, 34,620 მექსიკელი და 66,933 პუერტო-რიკოელი. როგორც მოსალოდნელი იყო, შემოსავლების მიხედვით სათავეში მოექცნენ კუბელები, საშუალო შემოსავლით $24,018 ($S_{\bar{y}}$ = $36,298), შუაში პუერტო-რიკოელები, საშუალო შემოსავლით $18,748 ($S_{\bar{y}}$ = $25,694), ხოლო ბოლოში მექსიკელი მამაკაცები $16,537 საშუალო შემოსავლით ($S_{\bar{y}}$ = $23,502).

თიენდასა და უილსონს არ გაუანგარიშებიათ ნდობის ინტერვალები თავიანთი შეფასებებისთვის, ამიტომ ჩვენ გამოვიყენებთ შემოსავლების განახლებულ მაჩვენებლებს 2000 წლის აღწერიდან[2] სამივე ჯგუფის მამაკაცებისთვის გავიანგარიშებთ 95%-იან ნდობის ინტერვალებს.

95%-იან ნდობის ინტერვალის მოსაძებნად კუბელების შემოსავლებისთვის, პირველად ვაფასებთ სტანდარტულ შეცდომას:

$$S_{\bar{\bar{y}}} = \frac{S_{\bar{y}}}{\sqrt{N}} = \frac{36,298}{\sqrt{29,233}} = 212.29$$

და 95%-იანი ნდობის ინტერვალი ტოლია:

$$\begin{aligned} 95\%\ CI &= 24,018 \pm 1.96(212,29) \\ &= 24,018 \pm 416 \\ &= 23,602\text{-დან } 24,434\text{-მდე} \end{aligned}$$

პუერტო-რიკოელების შემოსავლებისთვის სტანდარტული შეცდომა არის

$$S_{\bar{\bar{y}}} = \frac{S_{\bar{y}}}{\sqrt{N}} = \frac{25,694}{\sqrt{66,933}} = 99.32$$

და 95%-იანი ნდობის ინტერვალი ტოლია:

$$\begin{aligned} 95\%\ CI &= 18,748 \pm 1.96(99.32) \\ &= 18,748 \pm 196 \\ &= 18,553\text{-დან } 18,943\text{-მდე} \end{aligned}$$

და ბოლოს მექსიკელების შემოსავლებისთვის სტანდარტული შეცდომა არის

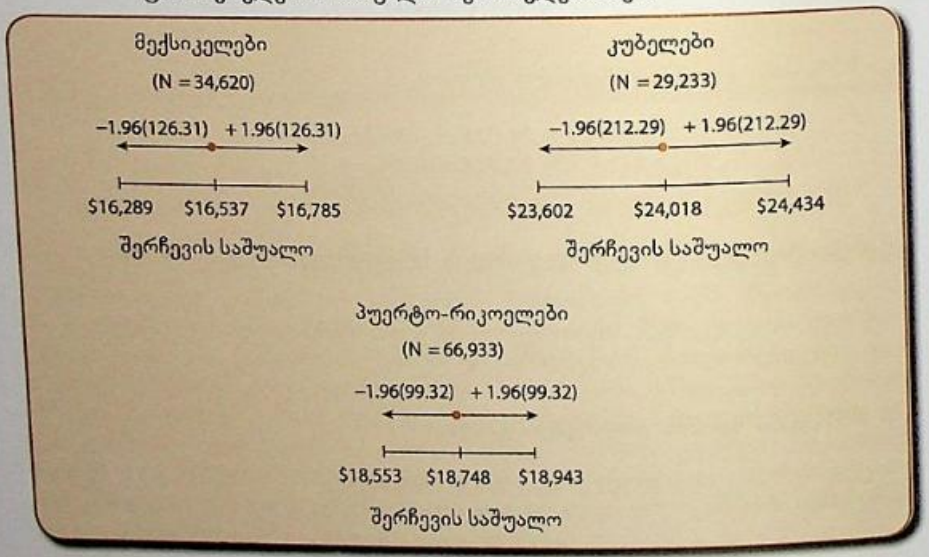
$$S_{\bar{\bar{y}}} = \frac{S_{\bar{y}}}{\sqrt{N}} = \frac{23,502}{\sqrt{34,620}} = 126.31$$

და 95%-იანი ნდობის ინტერვალი ტოლია:

$$95\%\ CI = 16{,}537 \pm 1.96(126.31)$$
$$= 16{,}537 \pm 196$$
$$= 16{,}289\text{-დან } 16{,}785\text{-მდე.}$$

მექსიკელების, პუერტო-რიკოელებისა და კუბელი იმიგრანტების საშუალო შემოსავლების ნდობის ინტერვალები ასახულია ნახაზზე 8.6. 95%-იანი სანდოობით შეიძლება ვთქვათ, რომ მათი რეალური შემოსავლების საშუალო სიდიდე შესაბამის ნდობის ინტერვალში მოხვდება. მიაქციეთ ყურადღება, რომ ნდობის ინტერვალები ერთმანეთს არ კვეთს, რაც შემოსავლებს შორის დიდ განსხვავებაზე მეტყველებს. ინტერვალის ყველაზე მაღალი შეფასება აქვთ კუბელებს, შემდეგ მექსიკელებსა და პუერტო-რიკოელებს.

ნახაზი 8.6 95%-იანი ნდობის ინტერვალები კუბელების, მექსიკელებისა და პუერტო-რიკოელების საშუალო შემოსავლებისთვის



ნდობის ინტერვალები სვედრიოთი ფილებისთვის

მთლიანი ერთობლიობის პროპორციების ან პროცენტების შეფასების მიზნით აგრეთვე შესაძლებელია შერჩევის პროპორციების ან პროცენტების ნდობის ინ-

ტერვალების გაანგარიშება. გაანგარიშების მიმდინარეობა ჩვენ მიერ უკვე განხილულის იდენტურია. ნდობის ინტერვალებს ვეძებ კოეფიციენტებში გავანგარიშებთ და შემდეგ 100-ზე გადავამრავლებთ.

თეორიული დებულებებიც, რაც შერჩევისა და დასკვნითი სტატისტიკის საფუძველს წარმოადგენს, ასევე ამოსავალია მთლიანი ერთობლიობის ხვედრითი წილების შეფასებისთვის.

ადრე ჩვენ ვნახეთ, რომ შერჩევის საშუალოებიდან მთლიანი ერთობლიობის საშუალოების შეფასებას საფუძვლად უდევს საშუალოების შერჩევითი განაწილება. ანალოგიურად, *ხვედრითი წილების შერჩევითი განაწილება* საფუძვლად უდევს მთლიანი ერთობლიობის ხვედრითი წილების შეფასებას შერჩევითი ხვედრითი წილებიდან. ცენტრალური ზღვრის თეორიიდან ცნობილია, რომ ხვედრითი წილების შერჩევითი განაწილება დაახლოებით ნორმალურია, როდესაც საშუალო – μ_p მთლიანი ერთობლიობის პროპორციის π-ის ტოლია, ხოლო სტანდარტული შეცდომა (ხვედრითი წილების შერჩევითი განაწილების სტანდარტული შეცდომა) ტოლია:

$$\sigma_p = \sqrt{\frac{(\pi)(1-\pi)}{N}} \tag{8.4}$$

სადაც

σ_p = პროპორციების სტანდარტული შეცდომა

π = მთლიანი ერთობლიობის პროპორცია

N = მთლიანი ერთობლიობის ზომა

რადგან ჩვენთვის მთლიანი ერთობლიობის პროპორცია π უცნობია (სწორედ მის შეფასებას ვცდილობთ), მისი შეფასებისთვის შეგვიძლია გამოვიყენოთ p, შერჩევის პროპორცია. მაშინ შეფასებული სტანდარტული შეცდომა იქნება:

$$S_p = \sqrt{\frac{(p)(1-p)}{N}} \tag{8.5}$$

სადაც

S_p = პროპორციების შეფასებული სტანდარტული შეცდომა

p = შერჩევის პროპორცია

N = შერჩევის ზომა

განვიხილოთ, მაგალითად, გელაპის მიერ ჩატარებული კვლევა 1,013 მოზრდილზე არიზონას საიმიგრაციო კანონთან დაკავშირებით. ამ კანონის მხარდამჭერთა ხვედრითი წილი 51%-ით შეფასდა. ფორმულა 8.5-ის საშუალებით, როდესაც $p = 0.51$, $1 - p = (1-0.51) = 0.49$ და $N = 1,013$, სტანდარტული შეცდომა იქნება

$S_p = \sqrt{(0.51)(1-0.51)/1,013} = 0.016$ ნორმალური განაწილების პირობასთან შესაბამისობისთვის ორი ფაქტორი უნდა გავითვალისწინოთ: (1) შერჩევის ზომა N და (2) შერჩევის პროპორციები p და $1-p$. როდესაც p და $1-p$ დაახლოებით 0,50-ია, შერჩევის ზომა 50-ზე ნაკლები არ უნდა იყოს. მაგრამ როდესაც $N > 0.50$ (ან $1-p < 0.50$), ნორმალური განაწილების პირობასთან შესაბამისობისთვის შერჩევის ზომა უფრო დიდი უნდა იყოს. ჩვეულებრივ, მთლიანი ერთობლიობის პროპორციის შეფასებისთვის შერჩევის ზომა 100 ან მეტი უნდა იყოს.

პროპორციების შეფასების პროცედურები

რადგან პროპორციების შერჩევითი განაწილება დაახლოებით ნორმალურია, პროპორციებისთვის იმავე წესით შეგვიძლია ნორმალური განაწილების ნდობის ინტერვალების გამოყენება, როგორც საშუალოებისთვის.

ნდობის ნებისმიერი ხარისხისთვის ნდობის ინტერვალების მთავარი ფორმულა პროპორციებისთვის არის:

$$CI = p \pm Z(S_p) \qquad (8.6)$$

სადაც

CI = ნდობის ინტერვალი

p = შერჩევითი დაკვირვების პროპორცია

Z = ნდობის ინტერვალის შესაბამისი Z

S_p = პროპორციების სტანდარტული შეცდომის შეფასება

უფრო დეტალურად განვიხილოთ ფორმულა. რაიმე ხარისხის ნდობის ინტერვალის მისაღებად შერჩევის პროპორციას ვუმატებთ ან ვაკლებთ Z-ის სიდიდისა და სტანდარტული შეცდომის ნამრავლს. Z-ის სიდიდის შერჩევა ნდობის სასურველ ხარისხზეა დამოკიდებული. ფართობი საშუალოსა და შერჩეულ ±Z-ს შორის ნდობის ხარისხის შესაბამისი უნდა იყოს.

მაგალითად, 95%-იანი ნდობის ინტერვალის მისაღებად, ჩვენ უნდა მოვძებნოთ 1.96-ის შესაბამისი Z-ის სიდიდე (დანართი ბ-დან), რადგან ვიცით, რომ მრუდის ქვეშ ფართობის 95% მოთავსებულია ±1.96-ის ფარგლებში. ანალოგიურად, 99%-იანი ნდობის ინტერვალის მისაღებად, უნდა მოვძებნოთ 2.58-ის შესაბამისი Z-ის სიდიდე (დამოკიდებულება ნდობის ინტერვალსა და Z სიდიდეებს შორის მოცემულია ნახაზზე 8.1).

პროპორციისთვის ნდობის ინტერვალის გაანგარიშება საშუალოსათვის ნდობის ინტერვალის გაანგარიშების ეტაპებს იმეორებს:

1. გაიანგარიშეთ პროპორციის შეფასებული სტანდარტული შეცდომა.
2. განსაზღვრეთ ნდობის ხარისხი და იპოვნეთ მისი შესაბამისი Z სიდიდე.

3. გაიანგარიშეთ ნდობის ინტერვალი.

4. მოახდინეთ შედეგების ინტერპრეტაცია.

ამ საფეხურების თვალსაჩინოებისთვის ისევ მივმართოთ გელაპის 2010 წლის გამოკვლევას არიზონას საიმიგრაციო კანონის მხარდამჭერთა ხვედრით წილთან დაკავშირებით.

პროპორციის სტანდარტული შეცდომის შეფასების გაანგარიშება

სტანდარტული შეცდომა 0.51 ან 51% პროპორციისათვის, როდესაც შერჩევის ზომა $N = 1,013$-ია არის 1.96.

ნდობის ხარისხისა და მისი შესაბამისი Z სიდიდის პოვნა

ჩვენ გადავწყვიტეთ ნდობის ხარისხად მივიჩნიოთ 95%. მას შეესაბამება 1.96-ის ტოლი Z სიდიდე.

ნდობის ინტერვალის გაანგარიშება

ნდობის ინტერვალის გასაანგარიშებლად შერჩევის პროპორციას ვუმატებთ ან ვაკლებთ Z ქულისა და სტანდარტული შეცდომის ნამრავლს.

$$95\%\ CI = 0.51 \pm 1.96(0.016)$$
$$= 0.51 \pm 0.03$$
$$= 0.48\text{-დან } 0.54\text{-მდე}$$

შედეგების ინტერპრეტაცია

95%-ით დარწმუნებულები ვართ, რომ ნამდვილი ერთობლიობის პროპორცია 0.48-სა 0.54-ს შორის იქნება. სხვა სიტყვებით, თუ შერჩეული მოზრდილი მოსახლეობის საკმაოდ დიდ რიცხვს დავაკვირდებით, 100-დან 95 შემთხვევაში ნამდვილი ერთობლიობის პროპორცია ნდობის ინტერვალს არ გასცდება. შეგვიძლია ეს შედეგი პროცენტებშიც გამოვსახოთ და ვთქვათ, რომ 95%-ით დარწმუნებული ვართ იმაში რომ, საიმიგრაციო კანონის მომხრეთა ნამდვილი ხვედრითი წილი მოხვდება 48%-იდან 54%-მდე ინტერვალში.

გაიანგარიშეთ გელაპის მონაცემებისთვის ნდობის ინტერვალი არა პროპორციების, არამედ პროცენტების გამოყენებით. თქვენ ნახავთ, რომ შედეგები იდენტური იქნება, მხოლოდ გამოსახული იქნება პროცენტებში.

✓ *ცოდნის შემოწმება*

აღსანიშნავია, რომ 95%-იანი სანდოობის პირობებში შეცდომის 5%-იანი რისკი არსებობს. თუ ჩვენ გავაგრძელებთ შერჩევებს, 100-დან 5 შერჩევაში მთლიანი ერთობლიობის პროპორცია ამ ინტერვალში არ მოხვდება.

შეგვიძლია რისკის შემცირება, თუ ნდობის ხარისხს 95%-იდან 99%-მდე გავზრდით.

$$99\%\ CI = 0.51 \pm 2.58(0.016)$$
$$= 0.51 \pm 0.04$$
$$= 0.47\text{-დან}\ \ 0.55\text{-მდე}$$

99%-იანი ინტერვალის გამოყენების შემთხვევაში შეგვიძლია დარწმუნებული ვიყოთ (99 შემთხვევაში 100-დან), რომ ნამდვილი ერთობლიობის პროპორცია მოხვდება 0.47-დან (47%) 0.55-მდე (55%) ინტერვალში. თუმცა, როგორც ადრე ვნახეთ, საჭიროა კომპრომისის მოძებნა ნდობის გაზრდილ ხარისხსა და შეფასების სიზუსტეს შორის. ნდობის გაზრდა 99%-მდე (ანუ რისკის შემცირება 5%-იდან 1%-მდე) შეფასების სიზუსტეზე ახდენს გავლენას და იწვევს აგრეთვე ნდობის ინტერვალის გადიდებას.[3,4]

სტატისტიკა პრაქტიკაში: ჯანმრთელობის დაცვის რეფორმა

შერჩევითი კვლევა შეიძლება მხოლოდ ერთი პარამეტრის შეფასებით შემოიფარგლოს. მაგალითად, კვლევითმა ორგანიზაციამ შეიძლება შეაფასოს იმ ამერიკელთა ხვედრითი წილი, ვინც მხარს უჭერს ჯანდაცვის რეფორმას. ზოგადად, კვლევების უმეტესობა მხოლოდ ერთი შეფასებით არ შემოიფარგლება. ხშირად დამოუკიდებელი შეფასებები კეთდება მთლიანი ერთობლიობის ქვეჯგუფებისთვის. 2010 წლის 8-11 აპრილს გამოქვეყნებულ ანგარიშში გელაპმა ამერიკელების რეაქცია ჯანდაცვის რეფორმასთან დაკავშირებით პოლიტიკური შეხედულებების მიხედვით შეისწავლა.[5]

როდესაც შეფასებები ქვეჯგუფებში კეთდება, მოსალოდნელია, რომ ჯგუფებში ნდობის ინტერვალები განსხვავებული იქნება. თითოეული ნდობის ინტერვალი ეყრდნობა ნდობის ხარისხს, პროპორციის სტანდარტულ შეცდომას (რომელიც უნდა შეფასდეს p-დან) და შერჩევის ზომას. მთლიანი ერთობლიობისათვის ნდობის ინტერვალის განსაზღვრის შემდეგ, ადვილად შევძლებთ ქვეჯგუფებისათვის დამოუკიდებელი ნდობის ინტერვალების გაანგარიშებას, თუ ჩავრთავთ ნდობის ხარისხს და თითოეული ქვეჯგუფის ზომას.

აღნიშნულის თვალსაჩინოებისთვის გავიანგარიშოთ 95%-იანი ნდობის ინტერვალები დემოკრატებისა და რესპუბლიკელების პროპორციებისთვის, ვინც მხარი დაუჭირა ჯანდაცვის რეფორმას. გამოკითხული 470 დემოკრატიდან რეფორმას მხარი დაუჭირა 0.81 (ან 81%-მა). ამის საპირისპიროდ, რესპუბლიკელთა მხოლოდ 0.10-მა (10%) დააფიქსირა მხარდაჭერა აღნიშნულისადმი.

პროპორციის შეფასებული სტანდარტული შეცდომის გაანგარიშება

პროპორციის შეფასებული სტანდარტული შეცდომა დემოკრატებისთვის იქნება:

$$S_p = \sqrt{\frac{(0.81)(1-0.81)}{470}} = 0.02$$

პროპორციის შეფასებული სტანდარტული შეცდომა დემოკრატებისთვის არის:

$$S_p = \sqrt{\frac{(0.10)(1-0.10)}{459}} = 0.01$$

ნდობის სასურველი ხარისხისა და მისი შესაბამისი Z სიდიდის განსაზღვრა

ჩვენ გადავწყვიტეთ ნდობის ხარისხად მივიჩნიოთ 95%. მას შეესაბამება 1.96-ის ტოლი Z სიდიდე.

ნდობის ინტერვალის გაანგარიშება

დემოკრატებისთვის,

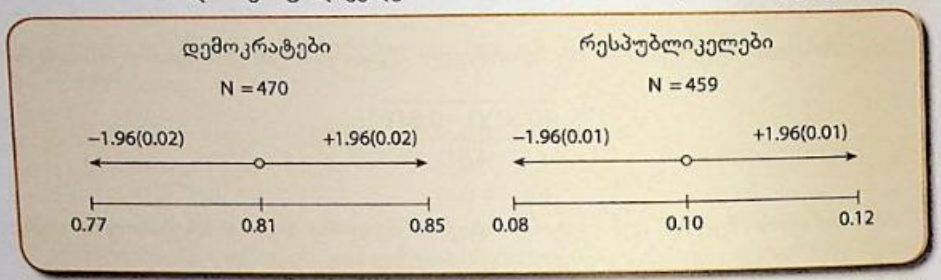
$$95\%\ CI = 0.81 \pm 1.96(0.02)$$
$$= 0.81 \pm 0.04$$
$$= 0.77\text{-დან}\ \ 0.85\text{-მდე}$$

რესპუბლიკელებისთვის,

$$95\%\ CI = 0.10 \pm 1.96(0.01)$$
$$= 0.10 \pm 0.02$$
$$= 0.08\text{-დან}\ \ 0.12\text{-მდე}$$

ნახაზზე 8.7 95%-იანი ნდობის ინტერვალით ნაჩვენებია ჯანდაცვის რეფორმის მხარდამჭერთა ხვედრითი წილები დემოკრატებსა და რესპუბლიკელებს შორის

ნახაზი 8.7 ჯანდაცვის რეფორმის მხარდამჭერთა ხვედრითი წილები დემოკრატებსა და რესპუბლიკელებს შორის 95%-იანი ნდობის ინტერვალით



წყარო: მონაცემები მოტანილია Frank Newport, "Americans' Views on Healthcare Law Remain Stable," *Gallup*, April 15, 2010-დან.

შედეგების ინტერპრეტაცია

95%-იანი სანდოობით შეიძლება ითქვას, რომ რეფორმის მხარდამჭერი დემოკრატების პროპორცია 0.77-სა და 0.85-ს შორის (77%-სა და 85%-ს შორის) იქნება მოთავსებული, ხოლო რესპუბლიკელებისთვის – 0.08-სა და 0.12-ს შორის (8%-სა და 12%-ს შორის).

დეტალების განხილვა 8.3
შეცდომის ზღვარი: ცდომილების ზღვარი

შეფასების გამოყენების ყველაზე გავრცელებული სფეროა არჩევნებსა და საზოგადოებრივი აზრის კვლევებში ნდობის ინტერვალების (ავრეთვე უწოდებენ ცდომილების ზღვარს) განსაზღვრა. მაგალითად, 2009 წელს ნიუ-იორკ თაიმსის გამოკითხვის შედეგად 6 708 გამოკითხული უმუშევრიდან 69%-მა დააფიქსირა, რომ ჩვეულებრივზე მეტად სტრესულ მდგომარეობაში იმყოფებოდა. ცდომილების ზღვრად მითითებული იყო ±4%. ცდომილების ზღვარი გვიჩვენებს, რამდენად კარგად ასახავს შემთხვევით შერჩეული ერთობლიობა მთლიან ერთობლიობას. ნიუ-იორკ თაიმსის[6] გამოკითხვის შედეგი გვეუბნება, რომ სტრესულ მდგომარეობაში მყოფი უმუშევრების ხვედრითი წილი სადღაც 65%-სა (69% – 4%) და 73%-ს (69% + 4%) შორისაა.

ცდომილების 4%-იანი ზღვარი 708 გამოკითხულისთვის ნიშნავს შემდეგს: 95%-იანი სანდოობით შეიძლება დარწმუნებული ვიყოთ, რომ ნამდვილი შედეგი 69% ±4%-ის ფარგლებს არ გასცდება.

ხაზგასასმელია, რომ ცდომილების ზღვარი მხოლოდ შემთხვევით შერჩეული ერთობლიობის შეცდომებს ზომავს. სამწუხაროდ, საარჩევნო და მოსახლეობის აზრის გამოკითხვებში შეცდომის სხვა წყაროებიც არსებობს, როგორიცაა არასწორად დასმული კითხვა და ინტერვიუერის შეცდომა. ეს შეცდომები თავს მაშინაც იჩენს, თუ კვლევა სწორ მეთოდოლოგიურ პრინციპებს ემყარება. ასეთმა სისტემატურმა შეცდომებმა შეიძლება დიდი გავლენა მოახდინოს შერჩევის სიზუსტეზე. ამასთან, მათი რაოდენობრივი დახასიათება თითქმის შეუძლებელია.

სისტემატური და არადათვლადი შეცდომების კლასიკური მაგალითია ნიუ-ჰემპშირის საპრეზიდენტო პრაიმერი, როდესაც სენატორ ობამას უწინასწარმეტყველებდნენ 10-15%-იან გამარჯვებას. შედეგების შეჯამების შემდეგ აღმოჩნდა, რომ კლინტონმა ობამას 39.4%-იდან 36.8%-მდე მოუგო.

შერჩევების ისტორიაში ეს ერთ-ერთი მნიშნელოვანი მარცხი იყო, ამიტომ ექსპერტებმა შეისწავლეს ეს საკითხი.

შერჩევების ექსპერტების[7] აზრით, ერთ-ერთი პრობლემა იყო ამომრჩეველთა დაყოფა სოციალ-ეკონომიკური ჯგუფების მიხედვით. კლინტონმა ობამას მოუგო 12 პუნქტით (47% და 35%) იმ ამომრჩევლებს შორის, რომელთა შემოსავალი $50,000-ზე ნაკლები იყო. ობამამ თავის მხრივ 5 პუნქტით გადაასწრო კლინტონს იმ ჯგუფში, სადაც შემოსავლები $50,000-ზე მეტი იყო.

აქ თავისი როლი ითამაშა აგრეთვე განსხვავებებმა განათლების დონეებში. კლინტონმა გაიმარჯვა მათ შორის, ვისაც არასოდეს უსწავლია კოლეჯში (43% და 35% შესაბამისად), ხოლო ობამას კოლეჯდამთავრებულთა კატეგორიამ (39% და 34% შესაბამისად) დაუჭირა მხარი. ერთ-ერთ ფაქტორად განიხილებოდა აგრეთვე წინასაარჩევნო პერიოდში თეთრკანიანი ღარიბი მოსახლეობის მიერ შავკანიანი კანდიდატის მხარდაჭერის გადაჭარბებული მოლოდინი.[8]

ძირითადი განმარტებები:

• მრავალი გამოკვლევის მიზანია მთლიანი ერთობლიობის პარამეტრების პოვნა. შერჩევის თეორიისა და სტატისტიკური დასკვნების მთავარი მიზანია შერჩევის სტატისტიკიდან მთლიანი ერთობლიობის პარამეტრების შეფასების გაკეთება.

• მკვლევარები აკეთებენ წერტილოვან და ინტერვალურ შეფასებებს. წერტილოვანი შეფასებები შერჩევის სტატისტიკაა, რომელიც გამოიყენება მთლიანი ერთობლიობის პარამეტრის ზუსტი სიდიდის შესაფასებლად. ინტერვალური შეფასების დროს ვსაზღვრავთ სიდიდეების დიაპაზონს, რომელშიც მთლიანი ერთობლიობის პარამეტრი მოხვდება.

• ნდობის ინტერვალები შეიძლება გამოვიყენოთ მთლიანი ერთობლიობის ისეთი პარამეტრების შესაფასებლად, როგორიცაა საშუალო ან პროპორცია. მათი სიზუსტე დამოკიდებულია ნდობის ხარისხზე. ყველაზე გავრცელებულია 95% და 99%-იანი ნდობის ხარისხები.

• საშუალოს ან პროპორციის ნდობის ინტერვალის გასაანგარიშებლად საშუალოს ან პროპორციას უნდა დავამატოთ ან დავაკლოთ სტანდარტული შეცდომისა და ნდობის ხარისხის შესაბამისი Z სიდიდის ნამრავლი.

საკვანძო ტერმინები

ნდობის ინტერვალი (ინტერვალური შეფასება)
ნდობის ხარისხი
შეფასება
ცდომილების ზღვარი
წერტილოვანი შეფასება

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

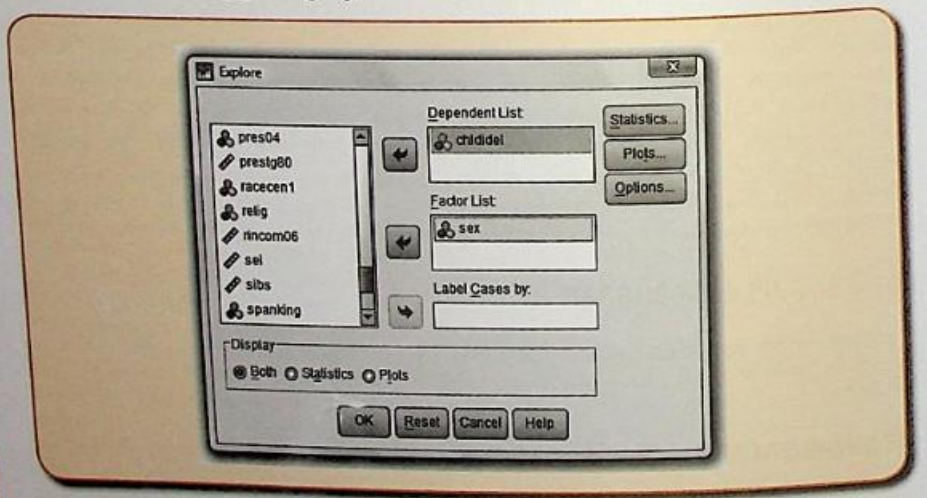
სტატისტიკა SPSS-ის გამოყენებით

[GSS08PFP-A]

SPSS შერჩევის საშუალოსა და პროპორციების ირგვლივ ნდობის ინტერვალებს ანგარიშობს Explore პროცედურების საშუალებით. გამოვიკვლიოთ ბავშვების იდეალური რაოდენობა მამაკაცებისა და ქალებისთვის.

გააქტიურეთ Explore პროცედურები შემდეგი თანმიმდევრობით: Analyze menu, Descriptive Statistics, შემდეგ Explore. დიალოგურ ფანჯარას, რომელიც გაიხსნება, გააჩნია ადგილი როგორც დამოკიდებული (dependent), ისე დამოუკიდებელი (independent) ცვლადებისთვის. Explore პროცედურების დაწყებამდე, ცვლადის CHILDIDEL კატეგორიას „as many as want“ (კოდი 98) გამოვაცხადებთ როგორც missing. ამისათვის შედით Data მენიუში, შემდეგ Define Variable Properties, ი პოვეთ ცვლადი CHILDIDEL ცვლადების ჩამონათვალში და მოათავსეთ ყუთში "Variables to Scan," შემდეგ დაკლიკეთ Continue; ახალ დიალოგურ ფანჯარაში, რომელიც გაიხსნება, შეარჩიეთ პატარა კვადრატი სვეტში Missing რომელიც შეესაბამება კატეგორიას „as many as want,“ მისი კოდია 98 (ეს პროცედურა აქ არ არის ასახული თვალსაჩინოდ). მოათავსეთ CHLDIDEL ყუთში Dependent List box, ხოლო SEX მოათავსეთ Factor List box-ში, როგორც ეს ნაჩვენებია ნახაზზე 8.8. (ამ ანალიზისთვის ჩვენ მოვახდინეთ შემდეგი კატეგორიების ელიმინირება: კატეგორიის 8 [as many as I want] და 9 [don't know or no answer].)

ნახაზი 8.8 დიალოგური ფანჯარა Explore

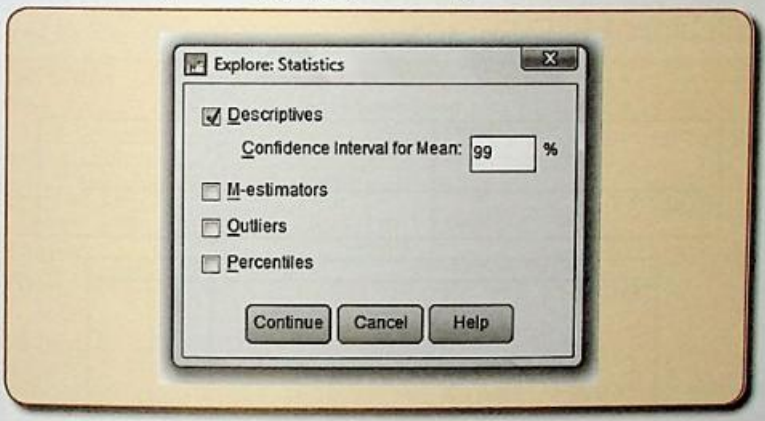


დაკლიკეთ ღილაკი Statistics. შენიშნავთ, რომ არჩევანი Descriptives აგრეთვე შეიცავს ნდობის ინტერვალის არჩევანს 95%. შევცვალოთ ეს არჩევანი 99%-ით (წაშალეთ 95 და ჩაწერეთ 99) (ნახაზი 8.9).

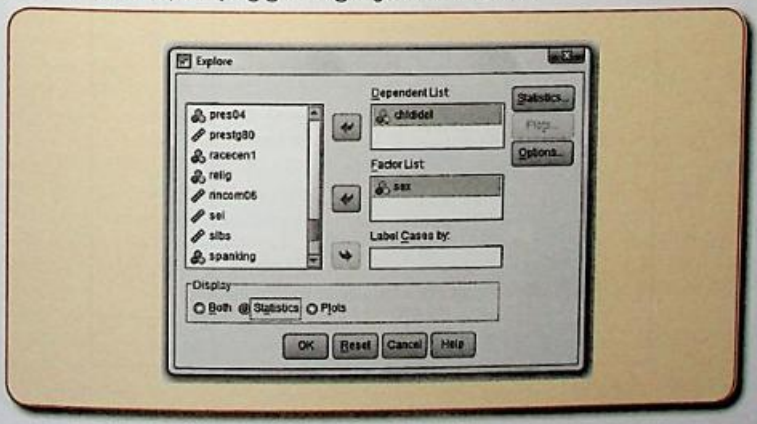
დაკლიკეთ Continue მთავარ დიალოგურ ფანჯარაზე გადასასვლელად. დათქმის თანახმად, Explore აწარმოებს სხვადასხვა სტატისტიკასა და გრაფიკებს. ამ მაგალითში ჩვენ ისინი არ გვჭირდება, ამიტომ დაკლიკეთ Statistics ღილაკი Display განყოფილებაში. თქვენი ეკრანი ნახაზი 8.10-ის მსგავსად უნდა გამოიყურებოდეს. დაკლიკეთ OK პროცედურების შესასრულებლად.

Explore პროცედურების შედეგების ფანჯარა (ნახაზი 8.11) დაყოფილია ორ ნაწილად, ერთი მამაკაცებისთვის და მეორე ქალებისთვის. მამაკაცებისთვის ბავშვების იდეალური საშუალო რაოდენობაა 2.40; ხოლო ქალებისთვის — 2.52. როგორც მონაცემებიდან ჩანს, ბავშვების იდეალური რაოდენობა ქალებისთვის მცირედ განსხვავდება მამაკაცებისაგან.

ნახაზი 8.9 ნდობის ინტერვალის მითითება



ნახაზი 8.10 დიალოგური ფანჯარა Explore



99%-იანი ნდობის ინტერვალი მამაკაცებისთვის მოიცავს 2.29-დან 2.51 ბავშვს. მონაცემების ინტერპრეტაცია შემდეგი სახით ხდება: 100 შერჩევის ბაზაზე, შერჩევის ზომით თითოეულში 398 მამაკაცი აშშ-ის მოზრდილი მოსახლეობიდან, შეიძლება 100-დან 99 შემთხვევაში განვსაზღვროთ საშუალოდ ბავშვების იდეალური რაოდენობა, რომელიც ნდობის ინტერვალში მოხვდება. როგორც უკვე ადრე ავხსენით, ჩვენ ვერ ვიქნებით დარწმუნებული, რომ ამ კერძო შემთხვევაში მთლიანი ერთობლიობის საშუალო (რომელიც ჩვენთვის უცნობია) ნდობის ინტერვალში მოხვდება. ჯერ-ჯერობით საუკეთესო ვარიანტია ნდობის დიაპაზონი 0.22. 99%-იანი სანდოობით ქალებისთვის ნდობის ინტერვალი თითქმის იგივეა და იცვლება 2.42-დან 2.63 ბავშვამდე, ან მხოლოდ 0.21.

ნახაზი 8.11 SPSS-ის შედეგების ფანჯარა ცვლადისთვის CHILDIDEL სქესობრივ ჭრილში

Case Processing Summary

	RESPONDENTS SEX	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
IDEAL NUMBER OF CHILDREN	MALE	398	58.1%	287	41.9%	685	100.0%
	FEMALE	454	55.7%	361	44.3%	815	100.0%

Descriptives

RESPONDENTS SEX				Statistic	Std. Error
IDEAL NUMBER OF CHILDREN	MALE	Mean		2.40	.042
		99% Confidence Interval for Mean	Lower Bound	2.29	
			Upper Bound	2.51	
		5% Trimmed Mean		2.38	
		Median		2.00	
		Variance		.719	
		Std. Deviation		.848	
		Minimum		0	
		Maximum		7	
		Range		7	
		Interquartile Range		1	
		Skewness		.887	.122
		Kurtosis		3.720	.244
	FEMALE	Mean		2.52	.040
		99% Confidence Interval for Mean	Lower Bound	2.42	
			Upper Bound	2.63	
		5% Trimmed Mean		2.50	
		Median		2.00	
		Variance		.731	
		Std. Deviation		.855	
		Minimum		0	
		Maximum		7	
		Range		7	
		Interquartile Range		1	
		Skewness		.744	.115
		Kurtosis		2.502	.229

ამოცანები SPSS-ისთვის

[GSS08PFP-A]

1. GSS-ის შერჩევა მოიცავს 18-დან 89 წლამდე მამაკაცებსა და ქალებს. აქვს თუ არა მნიშვნელობა ასაკს? შეიცვლება თუ არა პასუხები, თუ უფრო ახალგაზრდა ასაკის მოსახლეობას გამოვკითხავთ?

ა. ამ ამოცანის გადასაწყვეტად გამოიყენეთ Select Cases პროცედურა რესპონდენტების შესარჩევად ცვლადით AGE, ვინც 35 წლისაა ან უფრო ახალგაზრდა. ამისთვის შეარჩეთ Data, შემდეგ Select Cases. შემდეგ შეარჩიეთ If Condition is satisfied და დაკლიკეთ If. იპოვეთ და ჩასვით ცვლადი AGE ქვემოთა მარცხენა ყუთში. დაკლიკეთ ისარი ამ ყუთთან და ცვლადი AGE უკვე მარჯვენა ყუთში აღმოჩნდება. ახლა შეგიძლიათ დაავალოთ SPSS-ს ამოარჩიოს 35 წლის ან უფრო ახალგაზრდა რესპონდენტები. ყუთი ამ დავალებას წაიკითხავს ასე: AGE ≤ 35. დაკლიკეთ Continue და შემდეგ OK.

ბ. ახალგაზრდა რესპონდენტების შერჩევისათვის გაიმეორეთ Explore პროცედურები. რა განსხვავება არსებობს ქალებსა და მამაკაცებს შორის ამ ასაკებში? განსხვავდება თუ არა ეს შედეგები მთლიანი შერჩევის შედეგებისაგან?

2. გაიანგარიშეთ 90%-იანი ნდობის ინტერვალი შემდეგი ცვლადებისთვის და შეადარეთ დაბალი, მუშათა, საშუალო და მაღალი კლასების მონაცემები (CLASS) GSS-ის შერჩევაში. პირველად დაკლიკეთ Data, Select Cases, შემდეგ All Cases და OK. შემდეგ გამოიყენეთ Explore პროცედურები და ცვლადი CLASS როგორც ფაქტორული ცვლადი (Analyze, Descriptive Statistics, Explore). შეაჯამეთ მიღებული შედეგები.

ა. CHILDS (ბავშვების რაოდენობა შინამეურნეობაში)
ბ. EDUC (სწავლის ხანგრძლივობა)
გ. PAEDUC (მამის სწავლის ხანგრძლივობა)
დ. PRESTG80 (რესპონდენტის საქმიანობის პრესტიჟი)
ე. MAEDUC (დედის სწავლის ხანგრძლივობა)

თავის სავარჯიშოები

1. დანაშაულების კვლევის დროს დაზვერვის ფედერალურმა ბიურომ (FBI) აღმოაჩინა, რომ ამერიკელთა 13.2% დანაშაულის მსხვერპლი გახდა უკანასკნელი 1 წლის მანძილზე. შედეგი გამომდინარეობდა 1,105 მოზრდილის გამოკვლევიდან.

ა. შეაფასეთ მსხვერპლის ხვედრითი წილი 90%-იანი სანდოობით. სიტყვიერად ჩამოაყალიბეთ შედეგი.

ბ. შეაფასეთ მსხვერპლის ხვედრითი წილი 99%-იანი სანდოობით.

გ. დაუშვათ, ბიურომ გააორმაგა დაკვირვების ზომა და ახალ შერჩევაში აღმოაჩინა, რომ მსხვერპლის ხვედრითი წილი ისევ 13.2%-ია. რამდენით შემცირდება 90%-იანი ნდობის ინტერვალი? რამდენით შემცირდება 99%-იანი ნდობის ინტერვალი?

დ. ა, ბ და გ კითხვებზე გაცემული პასუხების მიხედვით, შეგიძლიათ მიხვდეთ თუ არა, რატომ იყენებენ 1,000-დან 1,500-მდე შერჩევის ზომას ეროვნულ კვლევის ორგანიზატორები Gallup, Roper, და *The New York Times*?

2. გამოიყენეთ მონაცემები განათლების შესახებ მეექვსე თავიდან, სავარჯიშო 5.

ა. გაიანგარიშეთ 95%-იანი ნდობის ინტერვალი განათლების საშუალო ხანგრძლივობისათვის დაბალ და საშუალო კლასებში.

ბ. გაიანგარიშეთ 99%-იანი ნდობის ინტერვალი განათლების საშუალო ხანგრძლივობისათვის დაბალ და საშუალო კლასებში.

გ. როგორ შეიცვლება ნდობის ინტერვალის სიდიდე, თუ ჩვენი სანდოობა გაიზრდება? ახსენით, რატომ ხდება ასე.

3. უკანასკნელ პერიოდში ბევრს კამათობენ გლობალური დათბობის შესახებ. 2006 წელს[9] Pew

Research Center-მა აღნიშნულ საკითხთან დაკავშირებით გამოიკვლია 1,501 ამერიკელი. აქედან 615 ადამიანმა დააფიქსირა, რომ გლობალური დათბობა სერიოზული პრობლემაა.

ა. შეაფასეთ იმ ადამიანების პროპორცია, რომლებიც ფიქრობენ, რომ გლობალური დათბობა სერიოზული პრობლემაა 95%-იანი ნდობის ინტერვალით.

ბ. შეაფასეთ იმ ადამიანების პროპორცია, რომლებიც ფიქრობენ, რომ გლობალური დათბობა სერიოზული პრობლემაა 99%-იანი ნდობის ინტერვალით.

გ. თუ ამ საკითხთან დაკავშირებით დააპირებთ ანგარიშის დაწერას, რომელ ნდობის ინტერვალს ამჯობინებთ, 95%-ს, თუ 99%-ს? პასუხი ახსენით.

4. გამოიყენეთ მონაცემები განათლების შესახებ მეხუთე თავიდან, სავარჯიშო 6, საქმიანობის პრესტიჟულობისა და განათლების შესახებ.

PRESTG80			Statistic
RS OCCUPATIONAL PRESTIGE SCORE (1980)	High school diploma	Mean	40.41
		Median	39.00
		Standard deviation	11.756
		Minimum	17
		Maximum	75
		Range	58
		Interquartile range	18
	Bachelor's degree	Mean	51.93
		Median	51.00
		Standard deviation	12.934
		Minimum	17
		Maximum	86
		Range	69
		Interquartile range	21

ა. გაიანგარიშეთ 90%-იანი ნდობის ინტერვალი საქმიანობის პრესტიჟულობისათვის მხოლოდ საშუალო სკოლის დიპლომის მქონე რესპონდენტებისთვის ($N = 706$).

ბ. გაიანგარიშეთ 90%-იანი ნდობის ინტერვალი საქმიანობის პრესტიჟულობისათვის ბაკალავრის დიპლომის მქონე რესპონდენტებისთვის ($N = 254$).

გ. ეს სტატისტიკა გამოიყენეთ საქმიანობის პრესტიჟულობის განსხვავებების ასახსნელად განათლების დონის მიხედვით.

5. 2010 წლის 1-დან 25 აპრილამდე გელაპი ატარებდა ამომრჩეველთა წინასაარჩევნო კვლევას.[10] 5,490 რეგისტრირებული ამომრჩევლიდან მამაკაცთა 51%-მა მხარი დაუჭირა რესპუბლიკელ კანდიდატს. გელაპი გისვამთ კითხვას: დღეს რომ არჩევნები იყოს, მოსალოდნელია თუ არა რესპუბლიკელების გაიმარჯვება? პასუხი ახსენით.

6. HINTS 2007-ის მონაცემებზე დაყრდნობით გამოიკვლიეთ რამდენად იმყოფებიან სტრესის ქვეშ მოზრდილი ამერიკელები. მონაცემების შეფასება ხდება 0-24 ქულით.

ა. HINTS 2007-ის მონაცემებიდან 1,390 ადამიანის სტრესის საშუალო ქულაა 4.75, ხოლო სტანდარტული გადახრა 4.53. ააგეთ 95%-იანი ნდობის ინტერვალი სტრესის ნამდვილი საშუალო ქულისათვის.

ბ. დაუშვათ, თქვენი თანაკურსელი, რომელიც ამბობს, რომ სტატისტიკაში ძლიერია, ამტკიცებს, რომ სტრესის ქულები არ არის ნორმალურად განაწილებული და ამიტომ ნდობის ინ-

ტერვალის გაანგარიშებას აზრი არ აქვს. დაუშვათ, განაწილების თაობაზე ის მართალია. გულისხმობს თუ არა ეს, რომ ნდობის ინტერვალის გაანგარიშება შეუსაბამოა? რატომ დიახ, ან არა?

7. GSS 2008-ის ქვეშერჩევიდან ვგებულობთ, რომ რესპონდენტთა 78.6%-ს სჯერა სიცოცხლისა სიკვდილის შემდეგ ($N = 1,303$).
   ა. როგორი იქნება 95%-იანი ნდობის ინტერვალი აშშ-ის მოსახლეობის იმ ნაწილისთვის, ვისაც სჯერა სიცოცხლისა სიკვდილის შემდეგ?
   ბ. გაანგარიშებების გარეშე მიხვდით, როგორი იქნება 90%-იანი და 99%-იანი ინტერვალების ზედა და ქვედა ზღვრები.
8. სოციალური სააგენტო აპირებს კვლევის ჩატარებას კლიენტების საშუალო შემოსავლის შეფასების მიზნით. დირექტორს სურს, რომ საშუალოს გაანგარიშება ძალიან ზუსტად მოხდეს ±$500. 2 წლის წინ ჩატარებული კვლევის მიხედვით ამ ერთობლიობისთვის სტანდარტული გადახრა დაახლოებით $5,000 იყო. თქვენი მიზანია შერჩევის საჭირო ზომის დადგენა შერჩევის შეცდომის შესამცირებლად ±$500-მდე.
   ა. საჭიროა თუ არა მიმდინარე საშუალო შემოსავლების შეფასების ცოდნა ამ კითხვაზე პასუხის გასაცემად? რატომ დიახ, ან არა?
   ბ. რა უნდა იყოს შერჩევის ზომა დირექტორის მოთხოვნების შესასრულებლად 95%-იანი სანდოობით? (შეხსენება: შერჩევის ზომის განსაზღვრისათვის გამოიყენეთ ნდობის ინტერვალის ფორმულა).
   გ. რა უნდა იყოს შერჩევის ზომა დირექტორის მოთხოვნების შესასრულებლად 99%-იანი სანდოობით?
9. GSS 2008-ის კვლევიდან ბავშვების საშუალო რიცხვი ერთ რესპონდენტზე 1.94-ია, ხოლო სტანდარტული გადახრა – 1.70. ამ კითხვაზე პასუხი გასცა 2,020 ადამიანმა. შეაფასეთ ერთ სულზე ბავშვების საშუალო რაოდენობა 90%-იანი სანდოობით.
10. 2008 წელს ჩატარებული MTF-ის კვლევამ აჩვენა, რომ მოზარდები მიდრეკილებას ამჟღავნებენ სოციალური თემებისადმი. ფაქტობრივად 1,488 მოზარდიდან 79.9%-მა გასცა პასუხი, რომ ხანდახან, ხშირად/ყოველთვის ფიქრობს სოციალურ საკითხებზე. შეაფასეთ 95% და 99% სანდოობით მოზარდების ხვედრითი წილი, რომლებიც ხანდახან ან ხშირად ფიქრობენ სოციალურ საკითხებზე.
11. Pew Research Center-ის მიერ 2008 წელს ჩატარებული კვლევის[11] თანახმად, 18-დან 29 წლამდე ამერიკელი ახალგაზრდების 27%-მა 2008 წლის საპრეზიდენტო კანდიდატების შესახებ ინფორმაცია სოციალური ქსელების საშუალებით მიიღო. განსაზღვრეთ მოსახლეობის ამ კატეგორიისთვის 95%-იანი ნდობის ინტერვალი.
12. 2010 წლის თებერვალში Pew Research Center-ის გამოქვეყნებული კვლევის[12] შედეგების მიხედვით, მილენიუმის თაობის ახალგაზრდების 61% ფიქრობს, რომ მათ უნიკალური და გამორჩეული იდენტობა აქვთ ($N = 527$).
   ა. განსაზღვრეთ ამ თაობის ხვედრითი წილის შეფასების ნდობის ინტერვალი 95%-იანი სანდოობით.
   ბ. განსაზღვრეთ ამ თაობის ხვედრითი წილის შეფასების ნდობის ინტერვალი 99%-იანი სანდოობით.
   გ. არის თუ არა ეს შედეგები იმ მოსაზრების შესაბამისი, რომ მილენიუმის თაობის უმეტესობა ასე ფიქრობს?
13. ზოგიერთი ფიქრობს, რომ ჰომოსექსუალიზმი ბიოლოგიური ნიშანია, ხოლო დანარჩენს ის გარემოს ან სოციალიზაციის შედეგად მიაჩნია. ამიტომ არ არის გასაკვირი, რომ რელიგიურმა ჯგუფებმა, რომლებიც ეწინააღმდეგებიან ჰომოსექსუალიზმს, გაბედეს, რომ მათ შეუძლიათ ამ სენისგან განკურნება. ამის შემდეგ მათ დაასკვნეს, რომ ჰომოსექსუალიზმი არ არის ბიოლოგიური ნიშანი, რომლითაც ადამიანი იბადება. 2008 წელს GSS-ის 1,269 რესპონდენტს ჰკითხეს, თუ რას ფიქრობდნენ ისინი ჰომოსექსუალიზმის შესახებ. აღნიშნულისადმი უარყოფითი დამოკიდებულება დააფიქსირა 51.5%-მა, ხოლო დადებითი – 38.1%-მა.

ა. თითოეული დასახელებული ხვედრითი წილისთვის გაიანგარიშეთ ნდობის 95%-იანი ინტერვალი.

ბ. რესპონდენტთა 10% სადღაც შუაში აღმოჩნდა, რომლებიც აფიქსირებენ, რომ აღნიშნული ურთიერთობები ცუდია, ან ხანდახან ცუდია. გაიანგარიშეთ 95%-იანი ნდობის ინტერვალი.

გ. რა დასკვნები შეიძლება გამოიტანოთ ამ საკითხზე ხალხის აზრის შეფასების შემდეგ?

14. 2008 წელს MTF-ის კვლევის შედეგად გამოვლინდა, რომ 184 შავკანიანი ახალგაზრდიდან 44%-მა დადებითად უპასუხა სხვა რასის წარმომადგენლებთან მეგობრობის შესაძლებლობას, ხოლო 867 თეთრკანიანიდან ამ კითხვას დადებითად უპასუხა მხოლოდ 39%-მა.

ა. თითოეული დასახელებული ხვედრითი წილისთვის გაიანგარიშეთ ნდობის 95%-იანი ინტერვალი.

ბ. აღნიშნული განსხვავების მინიმუმ 2 მიზეზის განხილვის საფუძველზე დაწერეთ შემაჯამებელი მიმოხილვა.

15. მრავალ ქალს უხდება ბავშვების გაზრდასთან ერთად სახლის გარეთ მუშაობა. მეცნიერებმა შეისწავლეს აღნიშნულის გავლენა ჰეტეროსექსუალურ ოჯახებზე.

ა. რას ფიქრობენ ამერიკელები ამ ტრენდის შესახებ? არის თუ არა განსხვავება მოსახლეობის სხვადასხვა ქვეჯგუფში? თუ დიახ, რომელი ქვეჯგუფებისთვის უფრო მისაღებია ეს ტრენდი? თუ ფიქრობთ, რომ განსხვავება არ არის, ახსენით რატომ?

ბ. GSS 2008-ის კვლევის თანახმად, 623 გამოკითხული მამაკაციდან 17% სრულიად ეთანხმება აზრს, რომ ქალის მუშაობა ზიანს არ აყენებს ბავშვს. ამ სტატისტიკისთვის ააგეთ 90%-იანი ნდობის ინტერვალი.

გ. 697 გამოკითხული ქალიდან 35.6% სრულიად ეთანხმება აზრს, რომ ქალის მუშაობა ზიანს არ აყენებს ბავშვს. ამ სტატისტიკისთვის ააგეთ 90%-იანი ნდობის ინტერვალი.

დ. ახსენით მამაკაცებისა და ქალების პასუხებში არსებული განსხვავების მიზეზები.

თავი 9

ჰიპოთეზების ტესტირება

თავის სწავლების მიზნები

- ჰიპოთეზის ტესტირების განსაზღვრა
- ჰიპოთეზის ტესტირების კომპონენტების განსაზღვრა და გამოყენება: კვლევითი და ნულოვანი ჰიპოთეზები, შერჩევითი განაწილება და კრიტერიუმის სტატისტიკა
- რას ნიშნავს ნულოვანი ჰიპოთეზის უარყოფა ან მიღება
- ორი შერჩევის საშუალოების ან ხვედრითი წილების ჰიპოთეზების ტესტირება

ფასების ზრდა ბენზინზე წარსულში შეიძლება მნიშვნელოვან ეროვნულ ან გლობალურ მოვლენასთან დაგვეკავშირებინა, როგორიც იყო ისრაელ-ლიბანის ომი ან ქარიშხალი კატრინა, თუმცა 2005 წელს 1 გალონი ჩვეულებრივი ბენზინის ფასმა 3 დოლარს მიაღწია და დიდხანს აღარ შეცვლილა. არაპროგნოზირებადი ფასები საწვავზე დღემდე შენარჩუნებულია მთელ ქვეყანაში, მაგრამ ეს ტვირთი უფრო საგრძნობია ცალკე აღებულ სოციალ-ეკონომიკურ ჯგუფებსა და გეოგრაფიულ რაიონებში.

დაბალი შემოსავლის მქონე ამერიკელები 8-ჯერ უფრო მეტს ხარჯავენ ბენზინზე, ვიდრე მდიდრები.[1] უილკოკსში, ალაბამას შტატი, 1 სატრანსპორტო საშუალების ბენზინზე ადამიანები შემოსავლის 12.72%-ს ხარჯავენ, ხოლო ჰანთერდონში, ნიუ-ჯერსის შტატი, მხოლოდ 1.52%-ს. ეროვნულ დონეზე 1 სატრანსპორტო საშუალების ბენზინით გამართვისათვის ამერიკელები თავიანთი შემოსავლის 3.8%-ს ხარჯავენ.

3 დოლარი გალონზე პირველად 2005 წელს დაფიქსირდა კალიფორნიაში. განსაკუთრებით დიდი დანაკარგები განიცადეს გაზის მომხმარებლებმა: როგორც ავტომობილების მფლობელებმა, ისე საზოგადოებრივი ტრანსპორტით მოსარგებლეებმა. ანალიტიკოსები ქვეყნის მასშტაბით ისევ ფასების ზრდას პროგნოზირებდნენ. შემცირდა სამომხმარებლო დანახარჯები; ამ ფაქტორს უკავშირებდნენ აგრეთვე შემცირებულ ნდობას ეკონომიკისადმი.

2010 წლისთვის ბენზინის ფასები კვლავ მაღალი იყო დასავლეთ სანაპიროს შტატებში, კერძოდ, ალასკასა და კალიფორნიაში. დაუშვათ, კალიფორნიის ავტოგასამართი სადგურებიდან ჩვენ შევარჩიეთ 100 მათგანი და გავიანგარიშეთ 1 გალონი ჩვეულებრივი გაზის საშუალო ფასი აპრილის ერთი კვირის განმავლობაში. მომხმარებლების მიერ მოწოდებული ინფორმაციიდან[2] ჩვენთვის ცნობილია, რომ ქვეყნის მასშტაბით საშუალო ფასი 2.86 დოლარია, ხოლო სტანდარტული გადახრა — 0.17. მაშასადამე, ჩვენ შეგვიძლია კალიფორნიაში გაზის საშუალო ფასი შევადაროთ ქვეყნის საშუალო მაჩვენელს აპრილის ერთი კვირის განმავლობაში. ამ შედარების დროს ჩვენ ვსვამთ კითხვას იმის თაობაზე, არის თუ არა კალიფორნიაში შერჩეული გასამართი სადგურები მთლიანად აშშ-ის ავტოგასამართი სადგურების რეპრეზენტატული. ფაქტობრივად, ჩვენ მოველით, რომ კალიფორნიის მაგალითი არარეპრეზენტატულია, რადგან აქ ფასები ქვეყნის საშუალო მაჩვენებელზე მაღალია.

ჩვენ შერჩევაში გაზის ფასი $3.11-ია, რაც ქვეყნის საშუალოზე $2.86 მაღალია. მაგრამ არის კი ფასებს შორის 25 ცენტიანი განსხვავება იმის გარანტია, რომ ჩვენი შერჩევა არარეპრეზენტატულია?

ამ კითხვაზე მარტივი პასუხი არ არსებობს. შერჩევის საშუალო 3.11 დოლარი ქვეყნის მასშტაბით გაანგარიშებულ საშუალოზე მაღალია, მაგრამ ჩვენი შეფასება მხოლოდ ერთ შერჩევას ემყარება. აქედან გამომდინარე, ეს უნდა ნიშნავდეს შემდეგს: (1) გაზის საშუალო ფასი კალიფორნიაში მართლა მაღალია ქვეყნის საშუალო მაჩვენებელზე, ან (2) გაზის საშუალო ფასი კალიფორნიაში დაახლოებით ისეთივეა, როგორც ქვეყანაში, მაგრამ კონკრეტულად ამ შერჩევაში ის უფრო მაღალი აღმოჩნდა.

როგორ უნდა განვსაზღვროთ, რომელი ახსნა უფრო მართებულია? ერთეული შერჩევა ან სხვადასხვა შერჩევები განსხვავებულ შედეგებს იძლევა, ამიტომ მთლიანი ერთობლიობის შესახებ შეფასებებს ვერ გავაკეთებთ. ჩვენ ისეთი მეთოდი გვჭირდება, რომელიც შერჩევის სტატისტიკაზე დაყრდნობით საშუალებას მოგვცემს შევაფასოთ ჰიპოთეზები მთლიანი ერთობლიობის პარამეტრების შესახებ. ამ მეთოდს **ჰიპოთეზის ტესტირება** ეწოდება.

ჰიპოთეზის ტესტირება მეთოდი, რომელიც შერჩევის სტატისტიკაზე დაყრდნობით საშუალებას გვაძლევს შევაფასოთ ჰიპოთეზები მთლიანი ერთობლიობის პარამეტრების შესახებ.

ჰიპოთეზების სტატისტიკური ტესტირების პირობები

ჰიპოთეზების სტატისტიკური ტესტირება რამდენიმე პირობის დაკმაყოფილებას მოითხოვს. ეს პირობებია: ცვლადის გაზომვის დონე, შერჩევის მეთოდი, მთლიანი ერთობლიობის განაწილების ფორმა და შერჩევის ზომა. კონკრეტული

დასკვნები და დაშვებები შეიძლება განსხვავებული იყოს ტესტის ან ტესტირების პირობებიდან გამომდინარე. მიუხედავად ამისა, *ყველა* ტესტი, გამონაკლისის გარეშე, გულისხმობს შემთხვევით შერჩევას. საშუალოს შესახებ ჰიპოთეზების ტესტირება აგრეთვე გულისხმობს გაზომვის ინტერვალურ-ფარდობით დონეს და ერთობლიობის ნორმალურ განაწილებას, ან 50-ზე მეტი შერჩევის ზომას.

ჩვენ მონაცემებზე დაყრდნობით შეგვიძლია მოვახდინოთ ჰიპოთეზის ტესტირება, რომ გაზის საშუალო ფასი კალიფორნიაში უფრო მაღალია, ვიდრე მთელი ქვეყნის მასშტაბით. ტესტი, რომელსაც ჩვენ ვიხილავთ, აკმაყოფილებს შემდეგ პირობებს:

1. კალიფორნიის გაზის გასამართი სადგურების შერჩევა მოხდა შემთხვევით.
2. *ერთი გალონი გაზის ფასი* ინტერვალურ-ფარდობით დონეზეა გაზომილი.
3. ჩვენ არ შეგვიძლია დავუშვათ, რომ მთლიანი ერთობლიობა ნორმალურ განაწილებას ექვემდებარება, მაგრამ, რადგან შერჩევის ზომა ($N = 50$) საკმაოდ დიდია, ცენტრალური ზღვრის თეორემიდან გამომდინარე ვიცით, რომ საშუალოს განაწილება დაახლოებით ნორმალურია.

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება

ჰიპოთეზების ჩამოყალიბება ხდება ცვლადებს შორის ურთიერთკავშირის საფუძველზე და ემყარება ძირითად თეორიებს. ადრე *ჰიპოთეზები* განვსაზღვრეთ როგორც სავარაუდო პასუხები საკვლევ საკითხებზე. მათ წინასწარი, სავარაუდო ხასიათი აქვთ, რადგან მტკიცებულებას მხოლოდ ემპირიული შემოწმების შედეგად ვიღებთ. ჰიპოთეზების ტესტირება უმნიშვნელოვანესი ეტაპია მტკიცებულებების შეგროვების პროცესში.

კვლევითი ჰიპოთეზა (H_1)

პირველი ნაბიჯი მდგომარეობს ჰიპოთეზის გამოსახვაში ისეთი ფორმით, რომ ის დაექვემდებაროს სტატისტიკურ ტესტირებას. ძირითად ჰიპოთეზას **კვლევით ჰიპოთეზას** უწოდებენ და ის აღინიშნება H_1-ით. ის ყოველთვის მთლიანი ერთობლიობის პარამეტრებით გამოისახება, რადგან ჩვენი მიზანია შერჩევის სტატისტიკაზე დაყრდნობით დასკვნების ჩამოყალიბება მთლიანი ერთობლიობის შესახებ.

კვლევითი ჰიპოთეზა (H_1) ასახავს ძირითად ჰიპოთეზას. ის ყოველთვის მთლიანი ერთობლიობის პარამეტრებით გამოისახება, მაგრამ მისი კონკრეტული ფორმა განსხვავებულია.

ჩვენ კვლევით ჰიპოთეზაში (H_1), ვაყენებთ დებულებას, რომ გაზის საშუალო ფასი კალიფორნიაში უფრო მაღალია, ვიდრე მთელი ქვეყნის მასშტაბით. თუ μ_Y-ით აღვნიშნავთ მთლიანი ერთობლიობის საშუალოს, ჰიპოთეზა შემდეგი სახით გამოისახება:

$$H_1: \mu_Y > \$2.86$$

ზოგადად, კვლევითი ჰიპოთეზა (H_1) განმარტავს, რომ მთლიანი ერთობლიობის პარამეტრი ჩამოთვლილთაგან ერთ-ერთია:

1. არ არის რაიმე კონკრეტული სიდიდის ტოლი: $\mu_Y \neq$ რაიმე კონკრეტული სიდიდის
2. მეტია რაიმე კონკრეტულ სიდიდეზე: $\mu_Y >$ რაიმე კონკრეტული სიდიდეზე
3. ნაკლებია რაიმე კონკრეტულ სიდიდეზე: $\mu_Y <$ რაიმე კონკრეტული სიდიდეზე

ნულოვანი ჰიპოთეზა (H_0)

შესაძლებელია თუ არა, რომ კალიფორნიასა და მთლიანად აშშ-ში გაზის საშუალო ფასებს შორის განსხვავება არ არსებობდეს და 0.25 ცენტის სხვაობა მხოლოდ შემთხვევითობა იყოს? სტატისტიკური დასკვნა ალბათობის თეორიას ეყრდნობა, ამიტომ კვლევითი ჰიპოთეზის მტკიცება ან უარყოფა პირდაპირ შეუძლებელია. უკეთეს შემთხვევაში ჩვენ მხოლოდ შეგვიძლია შევაფასოთ ალბათობა, რომ ეს ჭეშმარიტებაა ან მცდარია.

ამ ალბათობის შესაფასებლად სტატისტიკოსები ქმნიან ისეთ ჰიპოთეზას, რომელიც ეწინააღმდეგება კვლევით ჰიპოთეზას. **ნულოვანი ჰიპოთეზა**, რომელიც H_0-ით აღინიშნება, წინააღმდეგობაშია კვლევით ჰიპოთეზასთან და აყენებს დებულებას, რომ მთლიანი და შერჩევითი ერთობლიობების საშუალოებს შორის განსხვავება არ არის. ამ ჰიპოთეზას „არანაირი განსხვავების“ ჰიპოთეზაც შეიძლება დავარქვათ. ჩვენი ნულოვანი ჰიპოთეზა შეიძლება გამოისახოს შემდეგი სახით:

$$H_0: \mu_Y = \$2.86$$

ძირითადი ჰიპოთეზის (H_1) ტესტირების ნაცვლად, რომ კალიფორნიასა და ქვეყნის მასშტაბით ფასები განსხვავებულია, ჩვენ ვაკეთებთ ნულოვანი ჰიპოთეზის (H_0) ტესტირებას, რომ ფასებს შორის განსხვავება არ არსებობს. ნულოვანი ჰიპოთეზის უარყოფით ვაძლიერებთ კვლევით ჰიპოთეზას. ამით ვაძლიერებთ ჩვენ რწმენას იმ თეორიის სარგებლიანობისა და მნიშვნელობის შესახებ, საიდანაც ეს კვლევითი ჰიპოთეზა მივიღეთ.

ნულოვანი ჰიპოთეზა (H_0) „არანაირი განსხვავების“ პირობა, რომელიც ეწინააღმდეგება კვლევით ჰიპოთეზას და ყოველთვის გამოისახება მთლიანი ერთობლიობის პარამეტრებით.

უფრო ვრცლად კვლევითი ჰიპოთეზის შესახებ: ცალმხრივი და ორმხრივი ტესტები

ცალმხრივ ტესტში კვლევითი ჰიპოთეზა მიმართულების აღმნიშვნელია: ეს ნიშნავს, რომ ის აკონკრეტებს მთლიანი ერთობლიობის საშუალო ნაკლებია (<), თუ მეტია (>) რაიმე კონკრეტულ სიდიდეზე. ჩვენი კვლევითი ჰიპოთეზა შეიძლება გამოვსახოთ ამ სახით:

$$H_1: \mu_Y < \text{რაიმე კონკრეტულ სიდიდეზე}$$

ან

$$H_1: \mu_Y > \text{რაიმე კონკრეტულ სიდიდეზე}$$

ჩვენ მიერ დასმული ჰიპოთეზა კალიფორნიაში ერთი გალონი გაზის საშუალო ფასის შესახებ ცალმხრივი ჰიპოთეზაა.

როდესაც ცალმხრივ ჰიპოთეზაში დასმულია მოსაზრება, რომ მთლიანი ერთობლიობის საშუალო *მეტია* რაიმე კონკრეტულ სიდიდეზე, ასეთ ტესტს **მარჯვენამხრივი ტესტი** ეწოდება, რადგან შედეგი შერჩევითი განაწილების მარჯვენა მხარეს უნდა შევაფასოთ. თუ ცალმხრივ ჰიპოთეზაში დასმულია მოსაზრება, რომ მთლიანი ერთობლიობის საშუალო *ნაკლებია* რაიმე კონკრეტულ სიდიდეზე, ასეთ ტესტს **მარცხენამხრივი ტესტი** ეწოდება, რადგან შედეგი შერჩევითი განაწილების მარცხენა მხარეს უნდა შევაფასოთ. ჩვენ მაგალითში კრიტერიუმი მარჯვენამხრივია, რადგან კვლევით ჰიპოთეზაში ჩამოყალიბებულია, რომ კალიფორნიაში ერთი გალონი გაზის საშუალო ფასი მეტია 2.86 დოლარზე. (იხ. ნახაზი 9.1)

ხანდახან, რაღაც თეორიული საფუძველი გაგვაჩნია იმის სარწმუნოდ, რომ ჯგუფებს შორის განსხვავება არსებობს, თუმცა ამ განსხვავების მიმართულებას ვერ ვარკვევთ. მაგალითად, შეიძლება გვქონდეს საფუძველი, რომ დავიჯეროთ კალიფორნიაში *განსხვავებული* ფასის არსებობა, თუმცა გვაკლდეს კვლევის დამაჯერებლობა და სიღრმე იმის განსაზღვრისათვის, *მეტია* ის თუ *ნაკლები* მთლიანი ერთობლიობის საშუალო ფასზე. თუ არ გაგვაჩნია თეორიული საფუძველი კვლევითი ჰიპოთეზის მიმართულების გასარკვევად, მივმართავთ **ორმხრივ ტესტს.** კვლევითი ჰიპოთეზა გვეუბნება, რომ მთლიანი ერთობლიობის საშუალო არ უდრის რაიმე კონკრეტულ სიდიდეს. მაგალითად, გაზის საშუალო ფასის ირგვლივ ჰიპოთეზა შეიძლება ასე ჩავწეროთ:

$$H_1: \mu_Y \neq \$2.86$$

როგორც ცალმხრივ, ისე ორმხრივ ჰიპოთეზებთან ერთად ნულოვანი ჰიპოთეზა უცვლელი რჩება და გამოისახება:

H_1: μ_Y = რაიმე კონკრეტული სიდიდე

ცალმხრივი კრიტერიუმი ჰიპოთეზის ტესტირების სახე, რომელიც მოიცავს კვლევითი ჰიპოთეზის მიმართულებას. ის აკონკრეტებს, მთლიანი ერთობლიობის საშუალო ნაკლებია თუ მეტია რაიმე კონკრეტულ სიდიდეზე.

მარჯვენამხრივი კრიტერიუმი ცალმხრივი კრიტერიუმი, რომელშიც შერჩევის შედეგი შერჩევითი განაწილების მარჯვენა მხარეს უნდა შევაფასოთ.

მარცხენამხრივი კრიტერიუმი ცალმხრივი კრიტერიუმი, რომელშიც შერჩევის შედეგი შერჩევითი განაწილების მარცხენა მხარეს უნდა შევაფასოთ.

ორმხრივი კრიტერიუმი ჰიპოთეზის ტესტირების სახე, რომელიც მოიცავს კვლევით ჰიპოთეზას მიმართულების გარეშე. ჩვენ ერთნაირად გვაინტერესებს რამდენად მეტია ან ნაკლებია ერთი სიდიდე მეორეზე. შერჩევის შედეგი შეიძლება მოთავსებული იყოს შერჩევითი განაწილების როგორც მაღალ, ისე დაბალ დაბოლოებებზე.

საკმარისი არადამაჯერებლობის განსაზღვრა: ალბათური სიდიდეები და ალფა

ახლა ერთად მოვუყაროთ თავი მთელ ინფორმაციას. ჩვენი ვარაუდით, ნულოვანი ჰიპოთეზა სწორია და გვინდა შევამოწმოთ, საეჭვოა თუ არა ჩვენი მოსაზრება იმის შესახებ, რომ კალიფორნიაში ერთი გალონი გაზის საშუალო ფასი 2.86 დოლარზე მეტია, ანუ $\mu_Y > \$2.86$. როგორია იმის შანსი, რომ ჩვენ მიერ კალიფორნიაში შერჩეულ გასამართ სადგურებში გაზის საშუალო ფასი ქვეყნის საშუალო მაჩვენებელზე მაღალი აღმოჩნდება? შანსი, ანუ ალბათობა შეიძლება განვსაზღვროთ შერჩევითი განაწილებისა და სვედრითი წილების ცოდნის საფუძველზე. ცენტრალური ზღვრის თეორიაზე დაყრდნობით, შეგვიძლია ვთქვათ, რომ თუ შერჩევის ზომა 50-ზე მეტია, საშუალოს შერჩევითი განაწილება დაახლოებით ნორმალურია, $\mu_{\bar{Y}} = \mu_Y$ და სტანდარტული გადახრა (სტანდარტული შეცდომა) ტოლია:

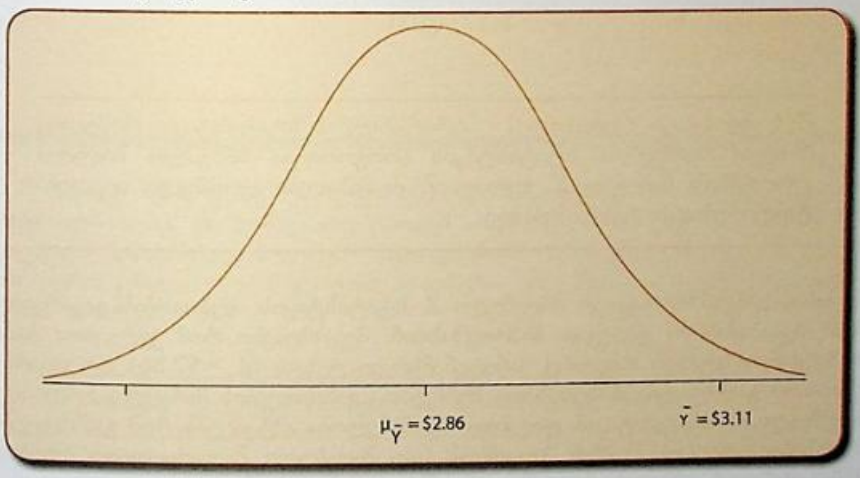
$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{N}}$$

ჩვენ ვაპირებთ დავასკვნათ, რომ ნულოვანი ჰიპოთეზა სწორია და შემდეგ ვნახოთ, შერჩევის შედეგი საეჭვოა თუ არა. მთლიანი ერთობლიობის საშუალო $\mu_Y = \$2.86$, სტანდარტული გადახრა $\sigma_Y = 0.17$. შერჩევის ზომა $N = 100$, ხოლო შერჩევის საშუალო $3.11-ია. ჩვენ შეგვიძლია დავასკვნათ, რომ $N = 100$ ზომის ყველა შესაძლო შერჩევისათვის საშუალოს განაწილება ნორმალურთან მიახლოებული იქნება საშუალოთი $2.86 და სტანდარტული განაწილებით:

$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{N}} = \frac{0.17}{\sqrt{100}} 0.017$$

ეს შერჩევითი განაწილება ნაჩვენებია ნახაზზე 9.1. აქ აგრეთვე ნაჩვენებია საშუალო ფასი, რომელსაც ჩვენ დავაკვირდით კალიფორნიის შერჩეულ სადგურებზე.

ნახაზი 9.1 შერჩევის საშუალოს შერჩევითი განაწილება, როდესაც H_0 სწორია შერჩევისთვის $N = 100$



რადგან შერჩევითი განაწილება ნორმალურია, შეგვიძლია გამოვიყენოთ დანართი ბ იმისთვის, რომ გავარკვიოთ იმის ალბათობა, შერჩევის საშუალო $3.11 ტოლია თუ მეტია მთლიანი ერთობლიობის საშუალოზე. მთლიანი ერთობლიობის შერჩევის საშუალოსთან შეფარდებითი მდებარეობის გასარკვევად, შერჩევის საშუალო გადავიყვანოთ Z ქულებში. მე-6 თავში შევისწავლეთ, თუ როგორ გადავიყვანოთ ნედლი ქულა Z ქულაში ფორმულა 6.1-ის დახმარებით:

$$Z = \frac{Y - \bar{Y}}{S_Y}$$

რადგან ჩვენ ვიხილავთ შერჩევით განაწილებას, სადაც ნედლი ქულა არის $\bar{Y}$ – საშუალო, ხოლო სტანდარტული გადახრა (სტანდარტული შეცდომა) არის $\sigma_Y / \sqrt{N}$, ფორმულა უნდა გარდავქმნათ შემდეგი სახით:

$$Z = \frac{\bar{Y} - \mu_Y}{\sigma_Y / \sqrt{N}} \qquad (9.1)$$

შერჩევის საშუალოს გარდაქმნას Z ქულებად ეწოდება *კრიტერიუმის სტატისტიკა*. Z სიდიდეს, რომელსაც მივიღებთ, **Z სტატისტიკა (მიღებული)** ეწოდება. მიღებული Z გვაძლევს სტანდარტული გადახრების (შეცდომის) რიცხვს, რომლითაც ჩვენი შერჩევა დაშორებულია საჰიპოთეზო სიდიდისაგან (μ_Y ან $\mu_{\bar{Y}}$), იმ დაშვებით, რომ ნულოვანი ჰიპოთეზა სწორია. ჩვენ მაგალითში მიღებული Z ტოლია:

$$Z = \frac{3.11 - 2.86}{0.17 / \sqrt{100}} = 14.70$$

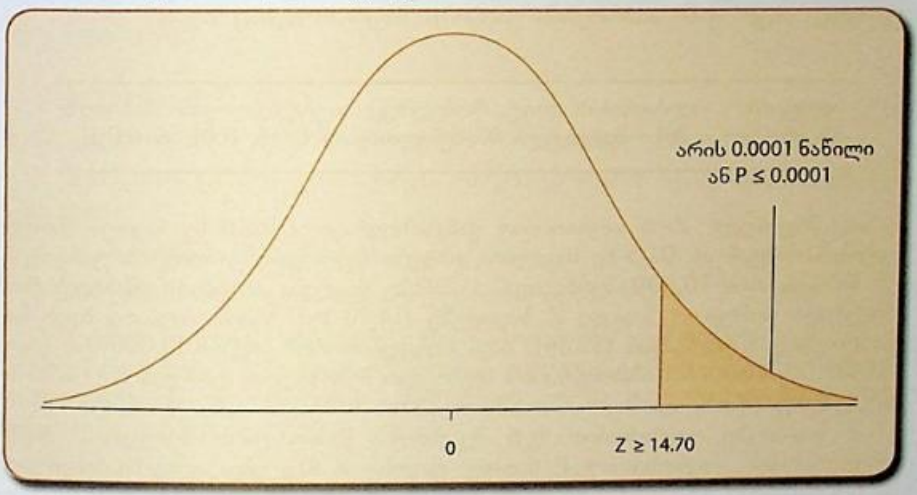
Z სტატისტიკა (მიღებული) კრიტერიუმის სტატისტიკა, რომელიც მიიღება შერჩევითი სტატისტიკის (როგორიცაა საშუალო სიდიდე) გარდაქმნის შედეგად Z ქულად. Z-ის მისაღები ფორმულა იცვლება ტესტირებიდან ტესტირებამდე.

სანამ განვსაზღვრავდეთ მიღებული Z სტატისტიკის ალბათობას, გავარკვიოთ მისი შესაბამისობა კვლევით ჰიპოთეზასთან. შეგახსენებთ, რომ კვლევითი ჰიპოთეზა განვსაზღვრეთ როგორც მარჯვენამხრივი ტესტი ($\mu_Y > \$2.86$) და ვივარაუდეთ, რომ განსხვავება მოხვდებოდა შერჩევითი განაწილების მარჯვენა ბოლოსკენ. მიღებული Z სტატისტიკის დადებითი მნიშვნელობა ამტკიცებს, რომ განსხვავების შეფასება მარჯვენა მხარეს მოგვიწევს (თუ მიღებული Z უარყოფითი იქნებოდა, მაშინ მარცხენა მხარეს მოგვიწევდა განსხვავების გაზომვა).

Z სიდიდის – 14.70, ალბათობის საპოვნელად, როდესაც ნულოვანი ჰიპოთეზა სწორია, დანართში B მოძებნეთ სიდიდე, რომელიც 14.70-ის მარჯვნივ (ზემოთ) მდებარეობს. ჩვენ მიერ გაანგარიშებული Z სიდიდე (14.70) დანართში B არ იძებნება, ამიტომ უნდა დავუკვირდეთ ბოლო Z სიდიდეს, რომელიც დანართში გვხვდება – 4.00. გაიხსენეთ ზე-ნ თავიდან, სადაც ვითვლიდით Z სიდიდეებს და მათ ალბათობებს, რომ Z სიდიდეები მოცემულია სვეტში A. P სიდიდე, ანუ ალბათობა, მოთავსებულია Z-ის მარჯვნივ, სვეტში C „არე Z სიდიდის მიღმა". ეს არე მოიცავს ყველა შერჩევით საშუალოს ზუსტად წილს, რომელიც $3.11-ის ტოლი ან მასზე მეტია. ზუსტადი წილი 0.0001-ზე ნაკლებია (ნახაზი 9.2).

ეს სიდიდე არის შედეგის მიღების ალბათობა, როდესაც ნულოვანი ჰიპოთეზა სწორია. ის აღინიშნება P სიმბოლოთი. ამიტომ ჩვენ მაგალითში $P \leq 0.0001$.

ნახაზი 9.2 ალბათობა (P), როდესაც $Z \geq 14.70$



P სიდიდე შეიძლება განვსაზღვროთ როგორც მიღებული Z სიდიდის ფაქტობრივი ალბათობა. ის ზომავს, თუ რაოდენ იშვიათად შეესაბამება, ან შეუსაბამოა მიღებული სტატისტიკა ნულოვან ჰიპოთეზასთან. რაც უფრო მცირეა *P* სიდიდე, მით უფრო ცხადია, რომ ნულოვანი ჰიპოთეზა უნდა უარვყოთ კვლევითი ჰიპოთეზის სასარგებლოდ.

P სიდიდე მიღებული Z სიდიდის შესაბამისი ალბათობა.

მეცნიერები ჩვეულებრივ წინასწარ საზღვრავენ, რომელია საკმარისად არადამაჯერებელი Z-ის მნიშვნელობა. ამისათვის ისინი უთითებენ ათვლის წერტილს, რომლის ქვემოთ P სიდიდის მოხვედრა გვაძლევს ნულოვან ჰიპოთეზაზე უარის თქმის საფუძველს. ამ წერტილს **ალფას** ეძახიან და ბერძნული ასოთი α აღინიშნება. მისი მნიშვნელობებია 0.05, 0.01 ან 0.001. დაუშვათ, ჩვენ გადავწყვიტეთ ნულოვანი ჰიპოთეზის უარყოფა, თუ $P \leq 0.05$. 0.05 არის α. ის განსაზღვრავს რომელი მნიშვნელობა არის საკმარისად წარმოუდგენელი იმისთვის, რომ გავრისკოთ და უარვყოთ ნულოვანი ჰიპოთეზა. $\alpha = 0.05$ ნიშნავს, რომ მაშინაც კი, თუ მიღებული Z სტატისტიკა შერჩევის შეცდომის საფუძველზეა გაანგარიშებული, როდესაც ნულოვანი ჰიპოთეზა სწორია, ჩვენ შეგვიძლია 5%-იანი რისკის ქვეშ

უარვყოთ ის. ალფას მნიშვნელობები 0.01 და 0.001 უფრო მეტი წინდახედულობით უნდა გამოვიყენოთ. განსხვავება P-სა და ალფას შორის იმაშია, რომ P წარმოადგენს მიღებული Z სიდიდის *ფაქტობრივ ალბათობას*, ხოლო α *ალბათობის წინასწარ განსაზღვრული* დონეა, რომელზეც ხდება ნულოვანი ჰიპოთეზის უარყოფა. ნულოვანი ჰიპოთეზის უარყოფა ხდება, როდესაც $P \leq \alpha$.

ალფა (α) ალბათობის დონე, რომელზეც ხდება ნულოვანი ჰიპოთეზის უარყოფა. მისი მიღებული მნიშვნელობებია 0.05, 0.01 ან 0.001.

თუ მიღებული Z-ის ალბათობად განვსაზღვრავთ 0.0001-ზე ნაკლებ მნიშვნელობას, რადგან ის 0.05-ზე ნაკლებია, ვახდენთ ნულოვანი ჰიპოთეზის უარყოფას. ეს ნიშნავს, რომ 10,000 შერჩევიდან 1 შანსზე ნაკლები არსებობს იმისთვის, რომ მიღებული საშუალო სიდიდე Z სიდიდეზე (14.70-ზე) ზემოთ მდებარეობდეს ჰიპოთეზო საშუალოსგან ($2.86). სხვა სიტყვებით რომ ვთქვათ, 10,000-დან ერთი შანსია (ან 0.0001%) იმისთვის, რომ შევარჩევთ შემთხვევით შერჩევას $Z \geq 14.70$-ით, თუ კალიფორნიასა და მთლიანად აშშ-ში გაზის საშუალო ფასი ერთნაირი იქნება.

P სიდიდეზე დაყრდნობით ჩვენ შეგვიძლია წამოვაყენოთ შედეგების „მნიშვნელოვნობის“ საკითხი. თუ P სიდიდე ტოლია ან ნაკლებია ალფაზე, მიღებული Z სტატისტიკა *სტატისტიკურად მნიშვნელოვანია*, ანუ ხდომილობის შემთხვევითი შესაძლებლობა, ან შერჩევის შეცდომა, ძალიან ნაკლებად შესაძლებელია. ჩვენ შეგვიძლია ვთქვათ, რომ კალიფორნიასა და მთლიანად აშშ-ში გაზის საშუალო ფასებს შორის განსხვავება მნიშვნელოვანია 0.05 დონეზე, ან რომ მნიშვნელოვნობის დონე 0.0001-ზე ნაკლებია.

გაიხსენეთ, რომ ჩვენი ჰიპოთეზა ცალმხრივი იყო ($\mu_Y > \$2.86$). ორმხრივ ჰიპოთეზაში შერჩევის შედეგები შერჩევითი განაწილების ორივე ბოლოსკენ განთავსდება. ამგვარად, ნულოვანი ჰიპოთეზა გაბათილდება, თუ შერჩევის შედეგი შერჩევითი განაწილების მარჯვენა ან მარცხენა ბოლოში მოხვდება. მაგალითად, 0.05 ალფა ან P სიდიდე ნიშნავს, რომ შერჩევის შედეგი შერჩევითი განაწილების ზედა ან ქვედა 5%-ში მოხვდა.

დაუშვათ, გაზის საშუალო ფასის შესახებ ჩვენი კვლევითი ჰიპოთეზა ასე ჩამოყალიბდა:

$$H_1: \mu_1 \neq \$2.86$$

ნულოვანი ჰიპოთეზა კვლავ არის $H_0: \mu_Y = \$2.86$ და ჩვენი მიღებული Z გაანგარიშებულია ფორმულა (9.1)-ის ბაზაზე, როგორც ცალმხრივი ტესტის დროს. ორმხრივი ტესტის დროს P-ს საპოვნელად მოვძებნოთ დანართი ბ-ს გ სვეტი და Z-ის შესაბამისი რიცხვი გავამრავლოთ 2-ზე, რომ მივიღოთ ორმხრივი ალბათობა. აქედან, ორმხრივი P სიდიდე Z-ისთვის (14.70) არის $0.0001 \times 2 = 0.0002$. ეს ალბათობა 0.05 ალფაზე ნაკლებია, ამიტომ ჩვენ უარვყოფთ ნულოვან ჰიპოთეზას.

ნუთი საფეხური ჰიპოთეზის ტესტირებაში: შეჯამება

კონკრეტული პრობლემის განხილვის დროს ჰიპოთეზის ტესტირება მოიცავს ხუთ საფეხურს:

1. პირობების განსაზღვრა
2. კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას შერჩევა
3. შერჩევითი განაწილების შერჩევა და კრიტერიუმის სტატისტიკის განსაზღვრა
4. კრიტერიუმის სტატისტიკის გაანგარიშება
5. დასკვნის ჩამოყალიბება და შედეგების ინტერპრეტაცია.

პირობების დაშვება. სტატისტიკური ჰიპოთეზების ტესტირება გარკვეული პირობების დაშვებას მოითხოვს, როგორიცაა ცვლადის გაზომვის დონე, შერჩევის მეთოდი, მთლიანი ერთობლიობის განაწილების ფორმა და შერჩევის ზომა. ჩვენ მაგალითში დავაყენეთ შემდეგი პირობები:

1. გამოვიყენეთ შემთხვევითი შერჩევა.
2. ცვლადი *ერთი გალონის ფასი* გავზომეთ ინტერვალურ-ფარდობით დონეზე.
3. რადგან $N > 50$, მთლიანი ერთობლიობის ნორმალურობის დაშვება აღარ არის საჭირო.

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას შერჩევა. *ძირითად ჰიპოთეზას* უწოდებენ კვლევით ჰიპოთეზას და აღნიშნავენ H_1-ით. ის ყოველთვის მთლიანი ერთობლიობის პარამეტრებით გამოისახება, რადგან ჩვენი მიზანია შერჩევის სტატისტიკაზე დაყრდნობით დასკვნების ჩამოყალიბება მთლიანი ერთობლიობის შესახებ. ჩვენი კვლევითი ჰიპოთეზა იყო:

$$H_1: \mu_Y > \$2.86$$

ნულოვანი ჰიპოთეზა ეწინააღმდეგება კვლევით ჰიპოთეზას დებულებით, რომ მთლიანი ერთობლიობის საშუალოსა და საჰიპოთეზო სიდიდეს შორის არავითარი განსხვავება არ არის. ჩვენ მაგალითში ნულოვანი ჰიპოთეზა ასე გამოისახება:

$$H_0: \mu_Y > \$2.86$$

ალფას მნიშვნელობად მივიღეთ 0.05, რაც იმას ნიშნავს, რომ ჩვენ უარვყოფთ ნულოვან ჰიპოთეზას, თუ მიღებული Z-ის მნიშვნელობის ალბათობა ნაკლებია ან ტოლია 0.05-ის.

შერჩევითი განაწილების შერჩევა და კრიტერიუმის სტატისტიკის განსაზღვრა. ნორმალური განაწილება და Z სტატისტიკა გამოიყენება ნულოვანი ჰიპოთეზის ტესტირებისთვის.

კრიტერიუმის სტატისტიკის გაანგარიშება. ფორმულა 9.1-ის საშუალებით გაანგარიშებული Z სტატისტიკა არის 14.70.

დასკვნის ჩამოყალიბება და შედეგების ინტერპრეტაცია. ჩვენ ვასაბუთებთ, რომ მიღებული Z სიდიდე მდებარეობს განაწილების მარჯვენა მხარეს და შეესაბამება კვლევით ჰიპოთეზას. მიღებულ Z სტატისტიკაზე (14.70) დაყრდნობით ვსაზღვრავთ P სიდიდეს, რომელიც 0.0001-ის ტოლია და ის 0.05 ალფაზე ნაკლებია. ჩვენ გვაქვს იმის უფლება, რომ უარვყოთ ნულოვანი ჰიპოთეზა კალიფორნიასა და მთლიანად აშშ-ში გაზის საშუალო ფასებში „არანაირი განსხვავების“ შესახებ. ამიტომ ვასკვნით, რომ კალიფორნიაში გაზის საშუალო ფასი მნიშვნელოვნად მაღალია საერთო ეროვნულ საშუალო ფასზე.

შეცდომები ჰიპოთეზის ტესტირებაში

ხაზგასასმელია, რომ რადგან ჩვენი დასკვნები შერჩევის მონაცემებს ეყრდნობა, რეალურად არასოდეს გვეცოდინება, სწორია თუ არა ნულოვანი ჰიპოთეზა. ფაქტობრივად, როგორც ვნახეთ, არსებობს 0.01% შესაძლებლობა, რომ ნულოვანი ჰიპოთეზა სწორია და მისი უარყოფით ჩვენ შეცდომას ვუშვებთ.

ნულოვანი ჰიპოთეზა შეიძლება იყოს სწორი ან არასწორი და ის შეიძლება უარვყოთ ან არ უარვყოთ. თუ ნულოვანი ჰიპოთეზა სწორია და მას უარვყოფთ, არასწორ გადაწყვეტილებას ვიღებთ. ასეთი ტიპის შეცდომას ეწოდება **I რიგის შეცდომა**. ამის საწინააღმდეგოდ, თუ ნულოვანი ჰიპოთეზა არასწორია და არ უარვყოფთ, მაშინ ვუშვებთ **II რიგის შეცდომას**.

I რიგის შეცდომა ალბათობა, რომელიც ასოცირდება ნულოვანი ჰიპოთეზის უარყოფასთან, როდესაც ის სწორია.

II რიგის შეცდომა ალბათობა, რომელიც ასოცირდება ნულოვანი ჰიპოთეზის არ უარყოფასთან, როდესაც ის არასწორია.

ცხრილში 9.1 ნაჩვენებია დამოკიდებულება ნულოვანი ჰიპოთეზის საფუძველზე მიღებულ გადაწყვეტილებებსა და ორივე ტიპის შეცდომას შორის. I რიგის შეცდომის ალბათობა – სწორი ჰიპოთეზის უარყოფა – ალფას შერჩეული დონის ტოლია. მაგალითად, როდესაც ჩვენ ვარჩევთ ალფას 0.05 დონეზე, ვიცით, რომ ნულოვანი ჰიპოთეზის სისწორის ალბათობა 0.05 ანუ 5%-ია.

ცხრილი 9.1 I და II რიგის შეცდომები

მიღებული გადაწყვეტილება	ნდობის ინტერვალი	
	H_0 სწორია	H_0 არასწორია
H_0-ის უარყოფა	I ტიპის შეცდომა (α)	სწორი გადაწყვეტილება
H_0-ის არ უარყოფა	სწორი გადაწყვეტილება	II ტიპის შეცდომა

ჩვენ შეგვიძლია სწორი ჰიპოთეზის უარყოფის რისკის შემცირება ალფას მანიპულირებით. მაგალითად, თუ ალფას მნიშვნელობად ავირჩევთ 0.01-ს, მაშინ I რიგის შეცდომის დაშვების რისკს 1%-ით შევამცირებთ. სამწუხაროდ, I და II რიგის შეცდომები უკუდამოკიდებულებაში არიან ერთმანეთთან, რაც იმას ნიშნავს, რომ ალფას შემცირებით და I რიგის შეცდომის დაშვების რისკის შემცირებით ჩვენ ვზრდით II რიგის შეცდომის დაშვების რისკს. (ცხრილი 9.1).

რადგან ჩვენი დასკვნები შერჩევით სტატისტიკას და არა მთლიანი ერთობლიობის პარამეტრებს ეფუძნება, განუსაზღვრელობის ხარისხი უნდა მივიღოთ როგორც დასკვნითი სტატისტიკის პროცესის შემადგენელი ნაწილი.

კვლევის შედეგების დანერგვა პრაქტიკაში ერთნაირად არ ხდება. მაგალითად, მკვლევარებმა შეიძლება წამოაყენონ ჰიპოთეზა, რომ ისპანახის გამოყენება ზრდის ძალოსანთა შესაძლებლობებს. შეიძლება ზიანის გარეშე არ ჩაიაროს ნულოვანი ჰიპოთეზის დაყენებამ, რომ ისპანახის გამოყენება არავითარ ეფექტს არ იძლევა და მისმა უარყოფამ. მეცნიერების მხრიდან I რიგის შეცდომის რისკის გაწევის გამო, ყველა ათლეტმა შეიძლება ისპანახის გამოყენება დაიწყოს. თუმცა, როდესაც კვლევის შედეგები ძალიან მნიშვნელოვანია, ბალანსის დაცვას I და II რიგის შეცდომებს შორის დიდი როლი ენიჭება. შეგიძლიათ თუ არა დაასახელოთ შემთხვევები, როდესაც მეცნიერები ცდილობდნენ I რიგის შეცდომების მინიმუმამდე დაყვანას? II რიგის შეცდომების მინიმუმამდე დაყვანას?

✓ ცოდნა შემოწმება

t სტატისტიკა და სტანდარტული შეცდომა

Z სტატისტიკა გავიანგარიშეთ (ფორმულა 9.1) ჰიპოთეზის ტესტირებისათვის კალიფორნიის გაზის გასამართი სადგურების შერჩევისათვის, რომელიც ეყრდნობოდა დაშვებას, რომ მთლიანი ერთობლიობის სტანდარტული გადახრა σ_Y ცნობილია. σ_Y გვჭირდება სტანდარტული შეცდომის გასაანგარიშებლად

$$\frac{\sigma_Y}{\sqrt{N}}$$

უმეტესად, σ_Y არ არის ცნობილი და მოგვიწევს მისი გაანგარიშება შერჩევის სტანდარტული გადახრის S_Y-ის საშუალებით. ასეთ შემთხვევებში Z სტატისტი-

კის ნაცვლად ნულოვანი ჰიპოთეზის ტესტირებაში ვიყენებთ t სტატისტიკას. t სტატისტიკის გასაანგარიშებელი ფორმულაა:

$$t = \frac{\overline{Y} - \mu_Y}{S_Y / \sqrt{N}} \qquad (9.2)$$

t სიდიდეს, რომელსაც ჩვენ გავიანგარიშებთ, **t სტატისტიკა (მიღებული)** ეწოდება. t სტატისტიკა წარმოადგენს სტანდარტული გადახრის (ან სტანდარტული შეცდომის) ერთეულებს, რომლითაც ჩვენი შერჩევის საშუალო დაცილებულია საჰიპოთეზო μ_Y სიდიდისაგან, როდესაც კეთდება დაშვება, რომ ნულოვანი ჰიპოთეზა სწორია.

---

t სტატისტიკა (მიღებული) კრიტერიუმის სტატისტიკა მთლიანი ერთობლიობის საშუალოს ნულოვანი ჰიპოთეზის ტესტირებისათვის, როდესაც მთლიანი ერთობლიობის სტანდარტული გადახრა უცნობია და მისი გაანგარიშება ხდება შერჩევის სტანდარტული გადახრით.

---

t განაწილება და თავისუფლების ხარისხი

t სტატისტიკის გასაგებად საჭიროა ჯერ გავეცნოთ მის განაწილებას. **t განაწილება** ფაქტობრივად წარმოადგენს მრუდების ოჯახს, რომელთაგან თითოეული განსაზღვრულია *თავისუფლების ხარისხით*. თავისუფლების ხარისხის კონცეფცია რამდენიმე სტატისტიკის გაანგარიშებაში გამოიყენება, t სტატისტიკის ჩათვლით. თავისუფლების ხარისხი (*df*) წარმოადგენს ქულების რიცხვს, რომლებიც თითოეული სტატისტიკის შემთხვევაში თავისუფლად იცვლება.

---

t განაწილება მრუდების ოჯახი, რომელთაგან თითოეული განსაზღვრულია თავისუფლების ხარისხით (*df*). ის მაშინ გამოიყენება, როდესაც მთლიანი ერთობლიობის სტანდარტული გადახრა უცნობია და სტანდარტული შეცდომის გაანგარიშება ხდება შერჩევის სტანდარტული გადახრით.

თავისუფლების ხარისხი (df) წარმოადგენს ქულების რიცხვს, რომლებიც თითოეული სტატისტიკის შემთხვევაში თავისუფლად იცვლება.

თავისუფლების ხარისხის განსაზღვრისთვის საჭიროა შერჩევის ზომის ცოდნა და კონკრეტული სტატისტიკის გაანგარიშებისთვის არსებული შეზღუდვები. თავისუფლების ხარისხის განსაზღვრისთვის შერჩევის ზომიდან ხდება შეზღუდვების რაოდენობის გამოკლება. როდესაც t სტატისტიკას ვანგარიშობთ ცალმხრივი კრიტერიუმისთვის, ვიწყებთ შერჩევის ზომიდან N და მას ვაკლებთ თავისუფლების 1 ხარისხს იმ მთლიანი ერთობლიობის სტანდარტული გადახრისთვის, რომელსაც ვაფასებთ.[3] აღსანიშნავია, რომ შერჩევის ზომის ზრდასთან ერთად იზრდება თავისუფლების ხარისხიც. მხოლოდ ერთი შერჩევის საშუალოსთვის df გაიანგარიშება შემდეგი სახით:

$$df = N - 1 \qquad (9.3)$$

t და Z სტატისტიკების შედარება

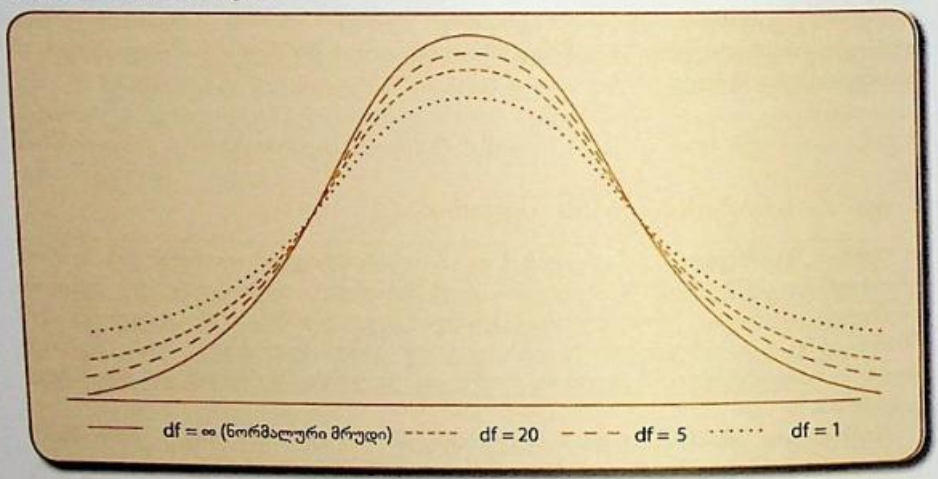
ალბათ შეამჩნევდით მსგავსებებს t და Z სტატისტიკების ფორმულებს შორის. ერთადერთი განსხვავება მნიშვნელშია. Z-ის მნიშვნელია სტანდარტული შეცდომა, რომელიც მთლიანი ერთობლიობის სტანდარტულ გადახრას σ_Y ეყრდნობა. t-ს მნიშვნელში გამოსახულებას $\sigma_Y / \sqrt{N}$ ვცვლით გამოსახულებით $S_Y / \sqrt{N}$, შეფასებული სტანდარტული შეცდომით, რომელიც ეფუძნება შერჩევის სტანდარტულ გადახრას.

მიუხედავად ამისა, Z-სა და t სტატისტიკებს შორის არის კიდევ ერთი მნიშვნელოვანი განსხვავება: რადგან t სტატისტიკის მნიშვნელი შერჩევის მონაცემებიდან გაიანგარიშება, ის შერჩევის შეცდომის ობიექტს წარმოადგენს. t სტატისტიკის შერჩევითი განაწილება არ არის ნორმალური და ამიტომ აქ სტანდარტული ნორმალური განაწილების გამოყენება შეუძლებელია მასთან დაკავშირებული ალბათობების განსაზღვრისათვის.

ნახაზზე 9.3 წარმოდგენილია t განაწილება რამდენიმე *df*-ისათვის. სტანდარტული ნორმალური განაწილების მსგავსად, t განაწილებასაც ზარის ფორმა აქვს. Z სტატისტიკის მსგავსად, t სტატისტიკასაც შეიძლება ჰქონდეს დადებითი და უარყოფითი სიდიდეები. დადებითი t სტატისტიკა შეესაბამება განაწილების მარჯვენა მხარეს, ხოლო უარყოფითი – მარცხენას. აღსანიშნავია, რომ როდესაც *df* მცირეა, t განაწილება უფრო გაბრტყელებულია, ვიდრე ნორმალური მრუდი. მაგრამ რაც უფრო იზრდება თავისუფლების ხარისხი, t განაწილების ფორმა ნორმალურს უახლოვდება და ისინი აბსოლუტურად იდენტური ხდება, როდესაც *df* 120-ზე მეტია. დანართში B შეჯამებულია t განაწილება. ცხრილში 9.2 მოტანილია ამ დანართის მხოლოდ მცირე ნაწილი. დააკვირდით და ნახავთ, რომ t ცხრილი Z ცხრილისგან რამდენადმე განსხვავებულია. პირველი, ცხრილის მარცხენა სვეტში მოცემულია თავისუფლების ხარისხები. t სტატისტიკა იცვლება თავისუფლების ხარისხის მიხედვით, რომელიც ჯერ უნდა გავიანგარიშოთ ($df = N - 1$). მეორე, ალბათობები, ანუ ალფა, რომელიც გამოიხატება მნიშვნელოვნობის დონეებით, მოთავსებულია ცხრილის თავში ორ ადგილზე. პირველი განკუთვნილია ცალმხრივი, ხოლო მეორე – ორმხრივი ტესტისთვის. და ბოლოს, t-ს

სიდიდეები წარმოადგენს შემდეგი ჩამონათვალის ფუნქციას: (1) თავისუფლების ხარისხი, (2) მნიშვნელოვნობის დონე (ან ალბათობა) და (3) ტესტი ცალმხრივია თუ ორმხრივი.

ნახაზი 9.3 ნორმალური განაწილება და t განაწილებები 1, 5, 20 და ∞ თავისუფლების ხარისხისთვის



ამ ცხრილის გამოყენების თვალსაჩინოებისთვის განვსაზღვროთ 2.021 t სიდიდის ნახვის ალბათობა თავისუფლების ხარისხით 40 და ორმხრივი ტესტისთვის. სტრიქონის $df = 40$ და სვეტის (ორმხრივი კრიტერიუმი) გადაკვეთაზე ვპოულობთ t სტატისტიკას 2.021, რომელიც შეესაბამება მნიშვნელოვნობის .05 დონეს. გადახედვის შემდეგ ჩვენ შეგვიძლია ვთქვათ, რომ t სტატისტიკის 2.021 მიღების ალბათობა .05-ია, ან 100-დან 5-ზე ნაკლები შანსია, რომ შემთხვევითი შერჩევიდან მივიღებთ 2,021-ის ტოლ t სტატისტიკას, როდესაც ნულოვანი ჰიპოთეზა სწორია.

სტატისტიკა პრაქტიკაში: თეთრკანიანი ქალების შემოსავლები

t სტატისტიკის გამოყენების თვალსაჩინოებისთვის მოვახდინოთ მთლიანი ერთობლიობის საშუალოს μ_Y ორმხრივი ჰიპოთეზის ტესტირება. GSS 2002-დან ვაკეთებთ თეთრკანიანი ქალების შერჩევას ($N = 371$), რომლებიც სრული დღის განმავლობაში მუშაობენ. ვპოულობთ საშუალო შემოსავლის სიდიდეს $28,889 და სტანდარტულ გადახრას $S_Y = \$21,071$. მოსახლეობის მიმდინარე დაკვირვებიდან[4] ვიცით, რომ 2002 წელს აშშ-ში ქალების საშუალო შემოსავალი იყო $\mu_Y = \$24,146$. თუმცა ჩვენ არ ვიცით მთლიანი ერთობლიობის სტანდარტული

გადახრა. ჩვენ გვინდა გავარკვიოთ, იყო თუ არა შერჩევა 2002 წელს მთელდღიურად მომუშავე ქალების მთლიანი ერთობლიობის რეპრეზენტატული. აგრეთვე, ჩვენ გარკვეული ეჭვი გვაქვს, რომ თეთრკანიანი ქალების საშუალო შემოსავალი მაღალია ქალების მთლიან ერთობლიობასთან შედარებით. ამის გამო, სტატისტიკური ტესტი ორმხრივია.

გამოვიყენოთ ხუთ-საფეხურიანი მოდელი ჰიპოთეზის ტესტირებისთვის, რომ თეთრკანიანი ქალების საშუალო შემოსავალი განსხვავდება მთელდღიურად მომუშავე ქალების საშუალო შემოსავლისგან 2002 წელს.

პირობების დაშვება. ჩვენი დაშვებები შემდეგია:

1. გამოვიყენეთ შემთხვევითი შერჩევა.
2. რადგან $N > 50$, მთლიანი ერთობლიობის ნორმალურობის დაშვება აღარ არის საჭირო.
3. ცვლად *შემოსავლის* გაზომვის დონე ინტერვალურია.

ცხრილი 9.2 t განაწილების სიდიდეები

	მნიშვნელოვნობის დონე ცალმხრივი ტესტისთვის					
	.10	*.05*	*.025*	*.01*	*.005*	*.0005*
	მნიშვნელოვნობის დონე ორმხრივი ტესტისთვის					
df	*.20*	*.10*	*.05*	*.02*	*.01*	*.001*
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
10	1.372	1.812	2.228	2.764	3.169	4.587
15	1.341	1.753	2.131	2.602	2.947	4.073
20	1.325	1.725	2.086	2.528	2.845	3.850
25	1.316	1.708	2.060	2.485	2.787	3.725
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
80	1.289	1.658	1.980	2.358	2.617	3.737
∞	1.282	1.645	1.960	2.326	2.576	3.291

წყარო: შემოკლებულია: R. A. Fisher and F. Yates, Statistical Tables for Biological, Agricultural and Medical Research, Table III. Copyright © □R. A. Fisher and F. Yates, 1963. Published by Pearson Education Limited-დან.

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას შერჩევა. ჩვენი კვლევითი ჰიპოთეზაა:

$$H_1: \mu_Y \neq \$24,146$$

ჩვენს მაგალითში ნულოვანი ჰიპოთეზა ასე გამოისახება:

$$H_0: \mu_Y = \$24,146$$

ალფას მნიშვნელობად მივიღეთ .05, რაც იმას ნიშნავს, რომ ჩვენ უარვყოფთ ნულოვან ჰიპოთეზას, თუ მიღებული სტატისტიკის ალბათობა ნაკლებია ან ტოლია .05-ის.

შერჩევითი განაწილების შერჩევა და კრიტერიუმის სტატისტიკის განსაზღვრა. ნულოვანი ჰიპოთეზის ტესტირებისთვის ვიყენებთ t სტატისტიკას და t განაწილებას.

კრიტერიუმის სტატისტიკის გაანგარიშება. პირველად ვანგარიშობთ ჩვენს კრიტერიუმთან დაკავშირებულ *df*-ს:

$$df = N - 1 = (371 - 1) = 370$$

იმისათვის, რომ შევაფასოთ შერჩევითი საშუალოს $28,889 მიღების ალბათობა, როდესაც ვუშვებთ, რომ თეთრკანიანი ქალების საშუალო შემოსავალი ეროვნული საშუალოს – $24,146-ის ტოლია, საჭიროა მიღებული t სტატისტიკის გაანგარიშება ფორმულით 9.2:

$$t = \frac{\overline{Y} - \mu_Y}{S_Y / \sqrt{N}} = \frac{29,889 - 24,146}{21,071 / \sqrt{371}} = 4.33$$

დასკვნის ჩამოყალიბება და შედეგების ინტერპრეტაცია. კვლევითი ჰიპოთეზის თანახმად, ჩვენ ჩავატარებთ ორმხრივ ტესტირებას. t სიდიდის (4.33) 370 თავისუფლების ხარისხით განსაზღვრისათვის, ჩავხედოთ ცხრილს 9.2. პირველი სვეტის ჩამონათვალში არ გვხვდება თავისუფლების ხარისხი 370, ამიტომ გამოვიყენოთ ბოლო სტრიქონი, $df = \infty$, რომ მოვათავსოთ მიღებული t სტატისტიკა.

მიუხედავად იმისა, რომ მიღებული t სტატისტიკა (4.33) არ გვხვდება t სტატისტიკის სტრიქონში, ფაქტობრივად ის ჩამოთვლილთაგან უკანასკნელ სიდიდეზე 3.291 უფრო მეტია. 3.291-ის t სტატისტიკა შეესაბამება .001 მნიშვნელოვნობის დონეს ორმხრივი ტესტისთვის. თუ კიდევ ერთხელ გადავხედავთ, შეგვიძლია ვთქვათ, რომ ჩვენ მიერ მიღებული t სტატისტიკა 4.33 მეტია 3.291-ზე, ან t სტატისტიკის (4.33) მიღების ალბათობა .001-ზე ნაკლებია ($P < .001$). ეს P სიდიდე ჩვენს .05 ალფას დონეზე დაბალია. განსხვავების $4,743 ($28,889 – $24,146) მიღების ალბათობა, თეთრკანიანი ქალების საშუალო შემოსავალსა და ეროვნულ საშუალოს შორის,

როდესაც ნულოვანი საშუალო სწორია, ძალიან მცირეა. ჩვენ საკმაო მტკიცებულებები გავაჩნია იმისთვის, რომ უარვყოთ ნულოვანი ჰიპოთეზა და დავასკვნათ, რომ 2002 წელს თეთრკანიანი ქალების საშუალო შემოსავალი განსხვავდებოდა ქალების ეროვნული საშუალო მაჩვენებლისგან. განსხვავება მნიშვნელოვანია .05 დონეზე. ჩვენ აგრეთვე შეგვიძლია ვთქვათ, რომ მნიშვნელოვნობის დონე .001-ზე ნაკლებია.

ორი შერჩევის ჰიპოთეზების ტესტირება

ორივე მაგალითი, რომელიც ამ თავში განვიხილეთ, მოიცავდა თითო შერჩევის მონაცემების შედარებას მთლიან ერთობლიობასთან. პრაქტიკაში მეცნიერებს ორი პარამეტრი (შერჩევა) უფრო აინტერესებთ, ვიდრე ერთი. მაგალითად, ჩვენ შეიძლება გვაინტერესებდეს ერთი რასობრივი/ეთნიკური ჯგუფის განათლების საშუალო ხანგრძლივობის შედარება სხვა ჯგუფებთან.

დაკვირვებების მონაცემებით, სხვა ჯგუფებთან შედარებით, აზიელების და წყნარი ოკეანის კუნძულელების განათლების ხანგრძლივობა უფრო მეტია. ის მოიცავს საშუალო სკოლისა და უფრო მაღალი, ან კოლეჯისა და უფრო მაღალი განათლების მქონეთა ხვედრით წილებს. მიუხედავად იმისა, რომ 1990 წლიდან შავკანიანებისა და ესპანური წარმოშობის ჯგუფების განათლების დონე გაიზარდა, ის მაინც ჩამორჩება აზიელებისა და კუნძულელების დონეს.

GSS 2008-დან ჩვენ შევამოწმეთ განსხვავება შავკანიანებისა და ესპანური წარმოშობის ჯგუფებს შორის განათლების ხანგრძლივობაში. შავკანიანების საშუალო მაჩვენებელი განისაზღვრა 12.80 წლით, ხოლო ესპანური წარმოშობის ჯგუფის – 10.63 წლით, როგორც ნაჩვენებია ცხრილში 9.3. ამ შერჩევების საშუალოები შეიძლება ნიშნავდეს შემდეგს: (1) შავკანიანების საშუალო მაჩვენებელი მაღალია ესპანური წარმოშობის ჯგუფის საშუალოზე, ან (2) შავკანიანების საშუალო მაჩვენებელი თითქმის იგივეა, რაც ესპანური წარმოშობის ჯგუფის საშუალო, მაგრამ ჩვენმა შერჩევამ უფრო მაღალი მაჩვენებელი აჩვენა. რასაც ჩვენ ეხლა აქ ვიყენებთ, ბივარიაციული ანალიზია (უფრო მეტი ინფორმაციისთვის იხ. თავი 10), მეთოდი, რომელიც საშუალებას გვაძლევს აღმოვაჩინოთ და აღვწეროთ ურთიერთდამოკიდებულება ორ ცვლადს შორის – რასა/ეთნიკურობა და განათლების დონე.

ცხრილი 9.3 განათლების წლები შავკანიანი და ესპანური წარმოშობის მამაკაცებისა და ქალებისთვის, GSS 2008

	შავკანიანები (შერჩევა 1)	ესპანური წარმოშობის (შერჩევა 2)
საშუალო	12.80	10.63
სტანდარტული გადახრა	2.55	3.76
დისპერსია	6.50	14.14
N	279	82

სტატისტიკური პროცედურები, რომლებიც ჩვენ უკვე განვიხილეთ, საშუალებას გვაძლევს ჩავატაროთ ტესტირება იმის თაობაზე, არის თუ არა ორ შერჩევაში აღმოჩენილი განსხვავებები იმის საფუძველი, რომ მთლიანი ერთობლიობები, საიდანაც მოხდა მათი შერჩევა, ასევე განსხვავებული იყოს.

ჰიპოთეზების ტესტირება ორი შერჩევისთვის იმავე საფეხურებს გაივლის, რასაც მხოლოდ ერთ-შერჩევიანი ტესტირება: პირობების დაშვება, კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას შერჩევა, შერჩევითი განაწილების შერჩევა და კრიტერიუმის სტატისტიკის განსაზღვრა, კრიტერიუმის სტატისტიკის გაანგარიშება და დასკვნის ჩამოყალიბება და გადაწყვეტილების მიღება — უარვყოთ თუ არა ნულოვანი ჰიპოთეზა.

დამოუკიდებელი შერჩევის პირობა

ერთ- და ორ-შერჩევიანი ჰიპოთეზების ტესტირებას შორის მნიშვნელოვანი განსხვავება შერჩევის პროცედურებს უკავშირდება. ორ-შერჩევიანი შემთხვევისთვის ჩვენ ვაკეთებთ დაშვებას, რომ ეს შერჩევები ერთმანეთისგან დამოუკიდებელია. ერთ შერჩევაში წევრების მოხვედრა გავლენას არ ახდენს მეორე შერჩევის წევრების ამორჩევაზე. ჩვენ მაგალითში ვაკეთებთ დაშვებას, რომ შავკანიანების შერჩევა დამოუკიდებელია ესპანური ჯგუფის წევრების შერჩევისგან. (დამოუკიდებლობის მოთხოვნა მაშინაც დაკმაყოფილებულია, როდესაც ერთი შერჩევა შემთხვევით ხდება და შემდეგ ის დაიყოფა შესაბამის ქვეჯგუფებად. მაგალითად, შეგვიძლია შემთხვევით მოვახდინოთ შერჩევა და შემდეგ დავყოთ ჯგუფებად სქესის, რელიგიის, შემოსავლის, ან სხვა ნებისმიერი ნიშნის მიხედვით, რაც ჩვენ გვაინტერესებს).

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება

შემდეგი განსხვავება ერთ- და ორ-შერჩევიან ტესტებში მდგომარეობს ფორმის შერჩევაში კვლევითი და ნულოვანი ჰიპოთეზისთვის. ერთ-შერჩევიან ტესტში როგორც ნულოვანი, ისე კვლევითი ჰიპოთეზა ერთი პარამეტრის — მთლიანი ერთობლიობის საშუალოს μ_Y ირგვლივ კეთდება. ამის საპირისპიროდ, ორ-შერჩევიან ტესტში ჩვენ ერთმანეთს ვადარებთ ორი ერთობლიობის პარამეტრებს.

ჩვენი კვლევითი ჰიპოთეზაა (H_1), რომ შავკანიანების განათლების საშუალო ხანგრძლივობა ესპანური წარმოშობის ჯგუფზე მაღალია. ჩვენ ვაყენებთ ჰიპოთეზას რასობრივ/ეთნიკურ ნიშანსა და განათლებას შორის ურთიერთკავშირის შესახებ მთლიან ერთობლიობაში შავკანიანების განათლების საშუალო ხანგრძლივობის შედარებით ესპანური წარმოშობის ჯგუფის მაჩვენებელთან. სიმბოლოს μ ვიყენებთ მთლიანი ერთობლიობის საშუალოს გამოსახვისთვის: ინდექსებს — 1 და 2 — პირველი (შავკანიანი) და მეორე შერჩევის (ესპანური წარმოშობის ჯგუფი) აღსანიშნავად. მაშინ ჩვენი კვლევითი ჰიპოთეზა იქნება:

$$H_1: \mu_1 \neq \mu_2$$

რადგან H_1 განსაზღვრავს, რომ შავკანიანებისა და ესპანური ჯგუფების განათლების დონეები არ არის ერთმანეთის მსგავსი, ამიტომ ეს ჰიპოთეზა მიმართულების გარეშეა. ე.ი. ჩვენი ტესტი ორმხრივია. ალტერნატულად, თუ საკმარისი საფუძველი გვექნებოდა, რომ გადაგვეწყვიტა, რომელი ერთობლიობის საშუალო იქნებოდა მეტი, ჩვენი ტესტი ერთმხრივი იქნებოდა:

$$H_1: \mu_1 < \mu_2 \text{ ან } H_1: \mu_1 > \mu_2$$

ნებისმიერ შემთხვევაში, ნულოვანი ჰიპოთეზა ამბობს, რომ ორი ერთობლიობის საშუალოებს შორის განსხვავება არ არის:

$$H_0: \mu_1 = \mu_2$$

ჩვენ დაინტერესებული ვართ ნულოვანი ჰიპოთეზის უარყოფაში, რათა საკმაოდ გავაძლიეროთ კვლევითი ჰიპოთეზა.

ჩამოაყალიბეთ კვლევითი და ნულოვანი ჰიპოთეზები შემდეგი სიტუაციებისთვის:

- *სოციოლოგებისა და ბიზნესის სტუდენტების მიერ მიღებულ სტატისტიკის ქულებს შორის არის განსხვავება.*
- *ორ-შვობლიან ოჯახებში ბავშვების რაოდენობა შავკანიანებსა და თეთრკანიანებს შორის განსხვავებულია.*
- *სპორტით დაკავებულ გოგონებში საშუალო აკადემიური მოსწრება უფრო მაღალია, ვიდრე მათ შორის, ვინც არ არის სპორტით დაკავებული.*

✓ *ცოდნა შემოწმება*

საშუალოებს შორის განსხვავების შერჩევითი განაწილება

შერჩევითი განაწილება საშუალებას გვაძლევს შევადაროთ შერჩევის ყველა შესაძლო შედეგი და შევაფასოთ მათი მოხდენის ალბათობა. ორი შერჩევის საშუალოებს შორის განსხვავების ტესტი ემყარება საშუალოებს შორის განსხვავების შერჩევით განაწილებას. **ორი შერჩევის საშუალოებს შორის განსხვავების შერჩევითი განაწილება** არის თეორიული ალბათური განაწილება, რომელიც მიიღება საშუალოებს შორის ყველა განსხვავების $(\bar{Y}_1 - \bar{Y}_2)$ გაანგარიშების საფუძველზე ორი ერთობლიობიდან N_1 და N_2 ზომის ყველა შესაძლო შერჩევის განხორციელების გზით.

ორი შერჩევის საშუალოებს შორის განსხვავების შერჩევითი განაწილება თეორიული ალბათური განაწილება, რომელიც მიიღება საშუალოებს შორის ყველა განსხვავების $(\bar{Y}_1 - \bar{Y}_2)$ გაანგარიშების საფუძველზე ორი ერთობლიობიდან N_1 და N_2 ზომის ყველა შესაძლო შერჩევის განხორციელების გზით, სადაც N_1 და N_2 50-ზე მეტია.

ორი შერჩევის საშუალოებს შორის განსხვავების შერჩევითი განაწილება ემყარება ცენტრალური ზღვრის თეორემას. თეორემიდან გამომდინარეობს, რომ ჩვენი შერჩევები გაკეთებულია დამოუკიდებლად ნორმალური ერთობლიობიდან, მაგრამ იმ შემთხვევაში, თუ შერჩევის ზომა დამაკმაყოფილებელია ($N_1 > 50$, $N_2 > 50$) ორი შერჩევის საშუალოებს შორის განსხვავების შერჩევითი განაწილება მიახლოებით ნორმალური იქნება, თუნდაც მთლიანი ერთობლიობები არ იყოს ნორმალური. შერჩევითი განაწილების საშუალოები იქნება $\mu_{\bar{Y}_1}$ და $\mu_{\bar{Y}_2}$, ხოლო სტანდარტული გადახრა (სტანდარტული შეცდომა) იქნება:

$$\sigma_{\bar{Y}_1-\bar{Y}_2} = \sqrt{\frac{\sigma_{Y_1}^2}{N_1} + \frac{\sigma_{Y_2}^2}{N_2}} \qquad (9.4)$$

რომელიც ემყარება თითოეული ერთობლიობის დისპერსიას ($\sigma_{Y_1}^2$ და $\sigma_{Y_2}^2$).

სტანდარტული შეცდომის შეფასება

ფორმულა 9.4-დან გამომდინარეობს, რომ მთლიანი ერთობლიობის დისპერსიები ჩვენთვის ცნობილია და აქედან შეგვიძლია სტანდარტული შეცდომის $\sigma_{\bar{Y}_1-\bar{Y}_2}$ (შერჩევითი განაწილების სტანდარტული გადახრა) გაანგარიშება. თუმცა, უმრავლეს შემთხვევაში, ჩვენთვის რეალურად მხოლოდ შერჩევის მონაცემებია ცნობილი და მთლიანი ერთობლიობების ნამდვილი დისპერსიები $\sigma_{Y_1}^2$ და $\sigma_{Y_2}^2$ არ ვიცით. ამიტომ ჩვენ უნდა შევაფასოთ საშუალოებს შორის განსხვავების სტანდარტული შეცდომა, რომელიც გამოისახება როგორც $S_{\bar{Y}_1-\bar{Y}_2}$, ნაცვლად $\sigma_{\bar{Y}_1-\bar{Y}_2}$-სა.

შეფასებული სტანდარტული შეცდომის გაანგარიშება

როდესაც ჩვენ დავუშვებთ, რომ ორი ერთობლიობის დისპერსია ერთმანეთის ტოლია, შევაერთებთ ორი შერჩევის დისპერსიების მონაცემებს და გავიანგარიშებთ შეფასებულ სტანდარტულ შეცდომას.

$$S_{\bar{Y}_1-\bar{Y}_2} = \sqrt{\frac{(N_1-1)S_{Y_1}^2 + (N_2-1)S_{Y_2}^2}{(N_1+N_2)-2}} \sqrt{\frac{N_1+N_2}{N_1N_2}} \qquad (9.5)$$

როგორც წესი, თუ ერთი შერჩევის დისპერსია მეორეზე ორჯერ მეტია, ჩვენ ვეღარ დავუშვებთ, რომ ორი მთლიანი ერთობლიობის დისპერსია ერთმანეთის ტოლია და ამ შემთხვევაში უნდა გამოვიყენოთ ფორმულა 9.8 დეტალების განხილვიდან 9.1.

t სტატისტიკა

ისევე როგორც ერთი შერჩევის შემთხვევაში, ჩვენ ვიყენებთ t განაწილებასა და t სტატისტიკას, როდესაც ვაფასებთ სტანდარტულ შეცდომას საშუალოებს შორის განსხვავების ტესტში. t სიდიდე, რომელსაც ვითვლით, მიღებული t სიდიდეა. ის წარმოადგენს სტანდარტული გადახრების რიცხვს (ან სტანდარტული შეცდომების ერთეულებს), რომლითაც ჩვენი საშუალოების განსხვავება $(\overline{Y}_1 - \overline{Y}_2)$ დაცილებულია საჰიპოთეზო სიდიდისაგან $\mu_1 - \mu_2$ იმის დაშვებით, რომ ნულოვანი ჰიპოთეზა სწორია.

t სტატისტიკის გასაანგარიშებელი ფორმულა საშუალოებს შორის განსხვავების ტესტისთვის არის

$$t = \frac{\overline{Y}_1 - \overline{Y}_2}{S_{\overline{Y}_1 - \overline{Y}_2}} \qquad (9.6)$$

სადაც $S_{\overline{Y}_1 - \overline{Y}_2}$ არის შეფასებული სტანდარტული შეცდომა.

თავისუფლების ხარისხის გაანგარიშება საშუალოებს შორის განსხვავების ტესტისთვის

t განაწილების გამოსაყენებლად ორი შერჩევის საშუალოებს შორის განსხვავების ტესტირებისთვის, გვჭირდება თავისუფლების ხარისხის გაანგარიშება. როგორც ჩვენთვის ცნობილია, თავისუფლების ხარისხი (*df*) წარმოადგენს ქულების რაოდენობას, რომლებიც თავისუფლად იცვლება თითოეული სტატისტიკის გაანგარიშებისას. როდესაც ვანგარიშობთ t სტატისტიკას ორ-შერჩევიანი ტესტისთვის, ჩვენ ვკარგავთ თავისუფლების 2 ხარისხს, თითოს, თითოეული მთლიანი ერთობლიობის დისპერსიის შეფასებისთვის. როდესაც დავუშვებთ, რომ ერთობლიობების დისპერსიები ტოლია, ან თითოეული შერჩევის ზომა 50-ზე მეტია, *df* შემდეგი სახით გაიანგარიშება:

$$df = (N_1 + N_2) - 2 \qquad (9.7)$$

როდესაც ჩვენ ვერ ვუშვებთ, რომ ერთობლიობის დისპერსიები ტოლია და ერთ-ერთი ან ორივე შერჩევის ზომა ტოლია ან ნაკლებია 50-ზე, ჩვენ ვიყენებთ ფორმულას 9.9, დეტალების განხილვიდან 9.1 თავისუფლების ხარისხის გასაანგარიშებლად.

დეტალების განხილვა 9.1
შეფასებული სტანდარტული შეცდომისა და თავისუფლების ხარისხის (*df*) გაანგარიშება, როდესაც მიგვაჩნია, რომ მთლიანი ერთობლიობების დისპერსიები ერთმანეთის ტოლი არ არის

თუ ერთი შერჩევის დისპერსია ($S^2_{Y_1}$ და $S^2_{Y_2}$) მეორეზე ორჯერ მეტია, მაშინ შეფასებული სტანდარტული შეცდომა გაიანგარიშება:

$$S_{\bar{Y}_1-\bar{Y}_2} = \sqrt{\frac{S^2_{Y_1}}{N_1} + \frac{S^2_{Y_2}}{N_2}} \tag{9.8}$$

როდესაც ერთობლიობის დისპერსიები არ არის ტოლი და ერთ-ერთი ან ორივე შერჩევის ზომა ტოლია ან ნაკლებია 50-ზე, ჩვენ ვიყენებთ t სტატისტიკასთან დაკავშირებული თავისუფლების ხარისხის ფორმულას[5]:

$$df = \frac{(S^2_{Y_1}/N_1 + S^2_{Y_2}/N_2)^2}{(S^2_{Y_1}/N_1)^2/(N_1-1) + (S^2_{Y_2}/N_2)^2/(N_2-1)} \tag{9.9}$$

ჰიპოთეზების ტესტირება საშუალოებს შორის განსხვავების შესახებ: შეჯამება

როგორც ერთი შერჩევის, ისე ორი შერჩევის საშუალოს შემთხვევაში, ჰიპოთეზების ტესტირება 5 საფეხურს მოიცავს:

1. პირობების განსაზღვრა
2. კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას შერჩევა
3. შერჩევითი განაწილების ამორჩევა და კრიტერიუმის სტატისტიკის განსაზღვრა
4. კრიტერიუმის სტატისტიკის გაანგარიშება
5. დასკვნის ჩამოყალიბება და შედეგების ინტერპრეტაცია.

პირობების განსაზღვრა. ჩვენ მაგალითში გვაქვს შემდეგი პირობები:

1. გამოყენებულია დამოუკიდებელი შემთხვევითი შერჩევა.
2. რადგან $N_1 > 50$, $N_2 > 50$ მთლიანი ერთობლიობის ნორმალურობის დაშვება არ არის საჭირო.
3. მთლიანი ერთობლიობების დისპერსიები განიხილება როგორც ერთმანეთის ტოლი.
4. მთლიანი ერთობლიობის დისპერსიები ერთმანეთის ტოლად არის მიჩნეული.

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას შერჩევა. ჩვენი კვლევითი ჰიპოთეზა იმაში მდგომარეობს, რომ შავკანიანების განათლების საშუალო ხანგრძლივობა ესპანური წარმოშობის ჯგუფისაგან განსხვავებულია, რაც იმაზე მიანიშნებს, რომ ჩვენი ტესტი ორმხრივია. კვლევითი ჰიპოთეზა ასე გამოისახება:

$$H_1: \mu_1 \neq \mu_2$$

სადაც μ_1-ით აღნიშნულია შავკანიანთა განათლების საშუალო დონე, ხოლო μ_2-ით ესპანური წარმოშობის ჯგუფის საშუალო.

ნულოვანი ჰიპოთეზა აყენებს დებულებას, რომ განათლების საშუალო სიდიდეებს შორის არავითარი განსხვავება არ არის, ანუ

$$H_0: \mu_1 = \mu_2$$

ჩვენ უნდა ვიპოვოთ საკმაო მტკიცებულება იმისთვის, რომ უარვყოთ ნულოვანი ჰიპოთეზა და გავაძლიეროთ კვლევითი ჰიპოთეზა. ნულოვანი ჰიპოთეზის უარყოფისთვის, მიღებული t ტოლი უნდა იყოს ან ნაკლები .05-ზე (ან ალფაზე).

შერჩევითი განაწილების ამორჩევა და კრიტერიუმის სტატისტიკის განსაზღვრა. t განაწილება და t სტატისტიკა გამოიყენება ორი ერთობლიობის საშუალოებს შორის განსხვავების მნიშვნელოვნობის ტესტირებისთვის.

კრიტერიუმის სტატისტიკის გაანგარიშება. ნულოვანი ჰიპოთეზის ტესტირებისთვის საჭიროა დაკვირვების შედეგად მიღებული განსხვავებები უნდა გადავიყვანოთ სტანდარტულ შეცდომაში t სტატისტიკით (ცხრილში 9.3 მოტანილი მონაცემების საფუძველზე). მიღებული t სტატისტიკა გაიანგარიშება ფორმულით 9.6:

$$t = \frac{\overline{Y}_1 - \overline{Y}_2}{S_{\overline{Y}_1 - \overline{Y}_2}}$$

სადაც $S_{\overline{Y}_1 - \overline{Y}_2}$ შერჩევითი განაწილების შეფასებული სტანდარტული შეცდომაა. რადგან ერთობლიობების დისპერსიები ერთმანეთის ტოლია $df = (N_1 + N_2) - 2 = (279 + 82) - 2 = 359$, ახლა შეგვიძლია შევაერთოთ ორივე შერჩევიდან მიღებული ინფორმაცია დისპერსიებზე და გავიანგარიშოთ სტანდარტული შეცდომა (ფორმულა 9.5):

$$S_{\overline{Y}_1 - \overline{Y}_2} = \sqrt{\frac{(279-1)2.55^2 + (82-1)3.76^2}{(279+82)-2}} \sqrt{\frac{279+82}{279(82)}} = (2.87)(0.13) = 0.37$$

ჩავსვათ ეს მნიშვნელობა t სტატისტიკის ფორმულის 9.6 მნიშვნელში, მივიღებთ:

$$t = \frac{12.80 - 10.63}{0.37} = 5.86$$

დასკვნის ჩამოყალიბება და შედეგების ინტერპრეტაცია. ჩვენ ვადასტურებთ, რომ მიღებული t განაწილების მარჯვენა მხარეს მდებარეობს. რადგან მიღებული t სტატისტიკა (5.86) მეტია სიდიდეზე $t = 3.291$ ($df = \infty$, ორმხრივი, იხ. დანართი B), შეგვიძლია ვთქვათ, რომ მისი ალბათობა .001-ზე ნაკლებია. ეს ჩვენი ალფას დონეზე (.05) დაბალია, ამიტომ შეგვიძლია უარვყოთ არანაირი განსხვავების ნულოვანი ჰიპოთეზა. ჩვენ შეგვიძლია დავასკვნათ, რომ შავკანიანებს საშუალოდ განათლების მნიშვნელოვნად მაღალი დონე აქვთ, ვიდრე ესპანური წარმოშობის ჯგუფს.

სტატისტიკა პრაქტიკაში: ნარკოტიკების გამოყენების პირველი შემთხვევა სტუდენტებს შორის

მომავლის მონიტორინგის კვლევა (MTF) სწავლობს მე-8, მე-10 და მე-12 კლასელების შეხედულებებს ნარკოტიკული ნივთიერებების გამოყენებაზე, როგორიცაა სიგარეტი, ალკოჰოლი, მარიჰუანა, კოკაინი და მეტამფეტაინი.[6]

ვნახოთ MTF 2008-ის მონაცემები გოგონებსა და ვაჟებში სიგარეტის, ალკოჰოლისა და მარიჰუანის პირველად მოხმარების შესახებ.

ცხრილიდან 9.4 ვიგებთ, რომ ვაჟები უფრო ადრე იწყებენ ალკოჰოლისა და მარიჰუანის მოხმარებას, ვიდრე გოგონები. მაგალითად, მარიჰუანის პირველი მოხმარების საშუალო კლასია 4.61 გოგონებისთვის და 4.39 ვაჟებისთვის. თუმცა გოგონები უფრო ადრე იწყებენ სიგარეტის მოხმარებას, ვიდრე ვაჟები.

მეტყველებს თუ არა განსხვავება 0.22 (4.61-4.39) იმაზე, რომ ვაჟები უფრო ადრე იწყებენ მარიჰუანის მოხმარებას? ჩვენ შეგვიძლია გამოვიყენოთ მონაცემები ცხრილიდან 9.4 და მოვახდინოთ ჰიპოთეზის ტესტირება საშუალოებს შორის განსხვავების გარშემო, რათა პასუხი გავცეთ ამ კითხვას. ჩვენ ვანგარიშობთ t სტატისტიკას, რათა მოვახდინოთ ტესტირება, არის თუ არა ჯგუფებს შორის დაკვირვების შედეგად მიღებული განსხვავება საკმაოდ დიდი იმისათვის, რომ საფუძველი გვქონდეს ამ შერჩევებს შორის მნიშვნელოვანი განსხვავების მტკიცებისათვის.

პირობების განსაზღვრა. ჩვენ მაგალითში გვაქვს შემდეგი პირობები:

1. გამოყენებულია დამოუკიდებელი შემთხვევითი შერჩევა.
2. ცვლადი *მარიჰუანის პირველად მოხმარების კლასი* არის ინტერვალურ-ფარდობითი.
3. რადგან $N_1 > 50$. $N_2 > 50$ მთლიანი ერთობლიობის ნორმალურობის დაშვება არ არის საჭირო.
4. მთლიანი ერთობლიობების დისპერსიები განიხილება როგორც ერთმანეთის ტოლი.

ცხრილი 9.4 ნარკოტიკების პირველად მოხმარების საშუალო კლასი

	მამაკაცები	ქალები
სიგარეტი	საშუალო = 4.90 S_y = 1.74 N = 150	საშუალო = 4.76 S_y = 1.73 N = 169
ალკოჰოლი	საშუალო = 4.22 S_y = 1.70 N = 617	საშუალო = 4.29 S_y = 1.64 N = 677
მარიჰუანა	საშუალო = 4.39 S_y = 1.70 N = 380	საშუალო = 4.61 S_y = 1.62 N = 394

წყარო: Monitoring the Future, 2008.

შენიშვნა: გამოყენებულია მხოლოდ ვალიდური პასუხები. ვინც უპასუხა, რომ არ აქვს გასინჯული ნარკოტიკი, არ არის შეტანილი მონაცემებში.

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას შერჩევა. კვლევითი ჰიპოთეზაა იმის ტესტირება, რომ ვაჟებისათვის მარიჰუანის პირველად მოხმარების საშუალო კლასი გოგონებზე უფრო დაბალია. კვლევითი ჰიპოთეზა ასე გამოისახება:

$$H_1: \mu_1 < \mu_2$$

სადაც μ_1-ით აღნიშნულია ვაჟების საშუალო კლასი, ხოლო μ_2-ით გოგონების მიერ მარიჰუანის პირველად მოხმარების საშუალო კლასი. აღსანიშნავია, რომ, რადგან ვაჟებისათვის მარიჰუანის პირველად მოხმარების საშუალო კლასი გოგონებზე უფრო დაბალია, ჩვენი ტესტი მარცხენამხრივი იქნება.

ნულოვანი ჰიპოთეზა აყენებს დებულებას, რომ მარიჰუანის პირველად მოხმარების საშუალო კლასის სიდიდეებს შორის არავითარი განსხვავება არ არის, ანუ

$$H_0: \mu_1 = \mu_2$$

ჩვენ უნდა ვიპოვოთ საკმაო მტკიცებულება იმისთვის, რომ უარვყოთ ნულოვანი ჰიპოთეზა და გავაძლიეროთ კვლევითი ჰიპოთეზა. ნულოვანი ჰიპოთეზის უარყოფისთვის მიღებული t ტოლი უნდა იყოს ან ნაკლები .01-ზე.

შერჩევითი განაწილების ამორჩევა და კრიტერიუმის სტატისტიკის განსაზღვრა. t განაწილება და t სტატისტიკა გამოიყენება ორი ერთობლიობის საშუალოებს შორის განსხვავების მნიშვნელოვნობის ტესტირებისთვის.

კრიტერიუმის სტატისტიკის გაანგარიშება. ნულოვანი ჰიპოთეზის ტესტირებისთვის მარიჰუანის პირველად მოხმარების საშუალო კლასის განსხვავების შესახებ 12-ე კლასელ ვაჟებსა და გოგონებს შორის, დაკვირვების შედეგად მიღებული განსხვავებები უნდა გადავიყვანოთ სტანდარტულ შეცდომაში t სტატისტიკით. მიღებული t სტატისტიკა გაიანგარიშება ფორმულით 9.6:

$$t = \frac{\overline{Y}_1 - \overline{Y}_2}{S_{\overline{Y}_1 - \overline{Y}_2}}$$

სადაც $S_{\overline{Y}_1 - \overline{Y}_2}$ შერჩევითი განაწილების შეფასებული სტანდარტული შეცდომაა. რადგან ერთობლიობების დისპერსიები ერთმანეთის ტოლია, $df = (N_1 + N_2) - 2 = (380 + 394) - 2 = 772$.

ახლა შეგვიძლია შევაერთოთ ორივე შერჩევიდან მიღებული ინფორმაცია დისპერსიებზე და გავიანგარიშოთ სტანდარტული შეცდომა (ფორმულა 9.5):

$$S_{\overline{Y}_1 - \overline{Y}_2} = \sqrt{\frac{(N_1 - 1)S_{Y_1}^2 + (N_2 - 1)S_{Y_2}^2}{(N_1 + N_2) - 2}} \sqrt{\frac{N_1 + N_2}{N_1 N_2}}$$

ჩვენი მაგალითისთვის, $S_{\overline{Y}_1 - \overline{Y}_2}$ ტოლია:

$$S_{\overline{Y}_1 - \overline{Y}_2} = \sqrt{\frac{(380-1)1.70^2 + (394-1)1.62^2}{(380+394)-2}} \sqrt{\frac{380+394}{380(394)}} = (1.66)(0.07) = 0.12$$

ჩავსვათ ეს მნიშვნელობა t სტატისტიკის ფორმულის (9.6) მნიშვნელში, მივიღებთ:

$$t = \frac{4.39 - 4.61}{0.12} = -1.83$$

დასკვნის ჩამოყალიბება და შედეგების ინტერპრეტაცია. ჩვენ ვადასტურებთ, რომ მიღებული t განაწილების მარცხენა მხარე მდებარეობს. მისი ალბათობის განსაზღვრისათვის ჩვენ ვიყენებთ t-ს აბსოლუტურ მნიშვნელობას. დანართი B-ს გამოყენებით ვიცით, რომ $t = -1.83$-ს შეესაბამება P სიდიდე, რომელიც .01-ზე მეტია. რადგან ალფას განსაზღვრეული დონე .01-ია და ალბათობა .01-ს აღემატება, ამიტომ არ შეგვიძლია უარვყოთ „არანაირი განსხვავების“ ნულოვანი ჰიპოთეზა. ჩვენ შეგვიძლია დავასკვნათ, რომ დაკვირვების შედეგად მიღებული განსხვავება გოგონებსა და ვაჟებს შორის არ ასახავს განსხვავებას, რომელსაც ადგილი უნდა ჰქონოდა მთლიან ერთობლიობაში.

შეცვლიდით თუ არა თქვენ გადაწყვეტილებას წინა მაგალითში, თუ ალფას მნიშვნელობად შევარჩევდით .05-ს?
ცხრილში 9.4 მოტანილი მონაცემების საფუძველზე მოახდინეთ ნულოვანი ჰიპოთეზის ტესტირება, რომ ალკოჰოლის პირველად მოხმარების საშუალო კლასი ვაჟებში უფრო დაბალია, ვიდრე გოგონებში. ალფა მიიჩნიეთ .05-ის ტოლად. რა დასკვნები გამოიტანეთ ტესტირებიდან?

✓ ცოდ
შემოწმე

ორი შერჩევის ხვედრით წილებს შორის განსხვავების მნიშვნელოვნობის ტესტირება

აქამდე ჩვენ ვიხილავდით მთლიანი ერთობლიობების საშუალოებს შორის განსხვავების მნიშვნელოვნობის საკითხს, როდესაც ცვლადი ინტერვალურ-ფარდობით დონეზე იზომებოდა. სოციალურ მეცნიერებებში უამრავი ცვლადი იზომება ნომინალურ და ორდინალურ დონეებზე. ასეთი ცვლადები ხვედრითი წილებით გამოისახება. მაგალითად, ჩვენ შეიძლება გვაინტერესებდეს მათი ხვედრითი წილების შედარება, ვინც მხარი დაუჭირა საიმიგრაციო კანონს ესპანური და არაესპანური წარმოშობის რესპონდენტებს შორის, ან დემოკრატიული პარტიის კანდიდატის მხარდამჭერი ქალებისა და მამაკაცების ხვედრითი წილების შედარება. ამ ნაწილში წარმოვადგენთ დასკვნითი სტატისტიკის მეთოდებს, თუ როგორ მოვახდინოთ შერჩევის ხვედრით წილებს შორის განსხვავებების მნიშვნელოვნობის ტესტირება.

ორი შერჩევის ხვედრითი წილების ჰიპოთეზის ტესტირება სტრუქტურულად იგივე საფეხურებს იმეორებს, რასაც ადრე წარმოდგენილი ტესტები: პირობების განსაზღვრა, კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას შერჩევა, შერჩევითი განაწილების ამორჩევა და კრიტერიუმის სტატისტიკის განსაზღვრა, კრიტერიუმის სტატისტიკის გაანგარიშება, დასკვნის ჩამოყალიბება და შედეგების ინტერპრეტაცია.

სტატისტიკა პრაქტიკაში: პარტიულობა და მთავრობისადმი ნდობა

საზოგადოებრივი აზრის შეფასების მიზნით სოციოლოგები ხშირად აგროვებენ მონაცემებს მიმდინარე სოციალ-პოლიტიკურ საკითხებზე. განსხვავებების კლასიფიცირება ხშირად ხდება შემოსავლების დონის, რასობრივი/ეთნიკური ნიშნის, სქესის, ან რეგიონისა და პოლიტიკური შეხედულების მიხედვით. მაგალითად, ერთნაირად ენდობიან თუ არა მთავრობას რესპუბლიკელები და დემოკრატები? ჩვენ შეგვიძლია GSS 2008-ის მონაცემებზე დაყრდნობით მოვახდინოთ ნულოვანი ჰიპოთეზის ტესტირება, რომ რესპუბლიკელები და დემოკრატები ერთნაირად ენდობიან მთავრობას. რესპუბლიკელების ხვედრითმა წილმა, რომლებიც დადებითად

ეხმაურებიან ამ კითხვას, 0.21 (p_1) შეადგინა, ხოლო დემოკრატთაგან – 0.06 (p_2). ამ კითხვას პასუხი გასცა 141-მა რესპუბლიკელმა (N_1) და 267 დემოკრატმა (N_2).

პირობების განსაზღვრა. ჩვენი პირობებია:

1. გამოყენებულია შემთხვევითი შერჩევები ზომით $N_1 > 50$ და $N_2 > 50$.
2. ცვლადის გაზომვის დონე ნომინალურია.

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას შერჩევა. ჩვენ ვაყენებთ ორმხრივ ტესტს, რომ რესპუბლიკელების და დემოკრატების ხვედრითი წილები განსხვავებულია.

$$H_1: \pi_1 \neq \pi_2$$

$$H_0: \pi_1 = \pi_2$$

ალფა განისაზღვრა .05-ის დონეზე.

შერჩევითი განაწილების ამორჩევა და კრიტერიუმის სტატისტიკის განსაზღვრა. მთლიან ერთობლიობაში დიქოტომიების განაწილება არ არის ნორმალური. მიუხედავად ამისა, ცენტრალური ზღვრის თეორემაზე დაყრდნობით ჩვენ ვიცით, რომ შერჩევის ხვედრით წილებს შორის განსხვავების შერჩევითი განაწილება ნორმალურია, თუ შერჩევის ზომა საკმაოდ დიდია ($N_1 > 50$ და $N_2 > 50$), საშუალო სიდიდით $\mu_{p_1-p_2}$ და სტანდარტული გადახრით $S_{p_1-p_2}$. ამიტომ ჩვენ შეგვიძლია შერჩევითი განაწილება ნორმალურად მივიჩნიოთ და კრიტერიუმის სტატისტიკად გამოვიყენოთ Z მნიშვნელობა.[7]

Z სტატისტიკის ფორმულას ხვედრით წილებს შორის განსხვავების გაანგარიშებისთვის შემდეგი სახე აქვს:

$$Z = \frac{p_1 - p_2}{S_{p_1-p_2}} \qquad (9.10)$$

სადაც p_1 და p_1 რესპუბლიკელების და დემოკრატების ხვედრითი წილებია, ხოლო $S_{p_1-p_2}$ შერჩევის ხვედრით წილებს შორის განსხვავების შერჩევითი განაწილების შეფასებული სტანდარტული შეცდომაა.

შეფასებული სტანდარტული შეცდომა გაიანგარიშება ფორმულით:

$$S_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{N_1} + \frac{p_2(1-p_2)}{N_2}} \qquad (9.11)$$

კრიტერიუმის სტატისტიკის გაანგარიშება. 9.11 ფორმულით ვანგარიშობთ სტან-

დარტულ შეცდომას:

$$S_{p_1-p_2} = \sqrt{\frac{0.21(1-0.21)}{141} + \frac{0.06(1-0.06)}{267}} = .037 = 0.04$$

ჩავსვათ ეს მნიშვნელობა ფორმულა 9.10-ის მნიშვნელში, მივიღებთ:

$$Z = \frac{0.21-0.06}{0.04} = 3.75$$

დასკვნის ჩამოყალიბება და შედეგების ინტერპრეტაცია. Z სიდიდე 3.75-ის ტოლია, რაც იმაზე მეტყველებს, რომ განსხვავება ორ ხვედრით წილს შორის Z განაწილების მარჯვენა ბოლოსკენ შეფასდება. Z სიდიდე 3.75-ის ალბათობის განსაზღვრისათვის, როდესაც ნულოვანი ჰიპოთეზა სწორია, უნდა მოვნახოთ დანართში ბ (სვეტი გ) არე, მიღებული Z სიდიდის მარჯვნივ (ზემოთ).

Z სიდიდეს (3.75) შეესაბამება 0.001 P სიდიდე. ორმხრივი კრიტერიუმისთვის P სიდიდე უნდა გავამრავლოთ 2-ზე (*.0001×2 = .0002*). 3.75-ის ალბათობა ორმხრივი ტესტისთვის ალფას დონეზე (.05) ნაკლებია (*.0002 < .05*).

ამის გამო, ჩვენ უარვყოფთ ნულოვან ჰიპოთეზას და ვასკვნით, რომ რესპუბლიკელებსა და დემოკრატებს შორის მნიშვნელოვანი განსხვავებაა მთავრობისადმი ნდობის საკითხში. რესპუბლიკელები უფრო მეტად ენდობიან ალმასრულებელ ხელისუფლებას.

შეცვლიდით თუ არა თქვენ დასკვნას ალკოჰოლის პირველად მოხმარების მაგალითში, თუ ალფა .01 იქნება?

✓ *ცოდნი შემოწმება*

სტატისტიკა პრაქტიკაში: განათლება და მთავრობისადმი ნდობა

იქნება თუ არა ჩვენი შედეგები განსხვავებული, თუ განათლების დონეებს შევადარებთ? GSS 2008-დან ჩვენ შემთხვევით შევარჩიეთ რესპონდენტები საშუალო სკოლისა და კოლეჯის დონის განათლებით და გვინდა შევისწავლოთ მათი განსხვავებული შეხედულება მთავრობის შესახებ.

საშუალო სკოლა და უფრო დაბალი	კოლეჯი ან ბაკალავრის ხარისხი
$p_1 = 0.11$	$p_2 = 0.09$
$N_1 = 837$	$N_2 = 494$

პირობების განსაზღვრა. ჩვენი პირობებია:

1. გამოყენებულია შემთხვევითი შერჩევები ზომით $N_1 > 50$ და $N_2 > 50$.
2. ცვლადის გაზომვის დონე ნომინალურია.

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას შერჩევა. ჩვენ ვაყენებთ ორმხრივ ტესტს, რომ რესპუბლიკელების და დემოკრატების ხვედრითი წილები განსხვავებულია.

$$H_1: \pi_1 \neq \pi_2$$

$$H_0: \pi_1 = \pi_2$$

ალფა განისაზღვრა .05-ის დონეზე.

შერჩევითი განაწილების ამორჩევა და კრიტერიუმის სტატისტიკის განსაზღვრა. შერჩევითი განაწილება ნორმალურია და კრიტერიუმის სტატისტიკად გამოვიყენოთ Z მნიშვნელობა.

კრიტერიუმის სტატისტიკის გაანგარიშება. სტანდარტული შეცდომა გაიანგარიშება 9.11 ფორმულით:

$$S_{p_1-p_2} = \sqrt{\frac{0.11(1-0.11)}{837} + \frac{0.09(1-0.09)}{494}} = 0.02$$

ჩავსვათ ეს მნიშვნელობა ფორმულა 9.10-ის მნიშვნელში, მივიღებთ:

$$Z = \frac{0.11-0.09}{0.02} = 1$$

დასკვნის ჩამოყალიბება და შედეგების ინტერპრეტაცია. მიღებული Z სიდიდე 1-ის ტოლია, რაც იმაზე მეტყველებს, რომ განსხვავება ორ ხვედრით წილს შორის განაწილების მარჯვენა ბოლოსკენ შეფასდება. 1-ის ტოლ Z სიდიდეს შეესაბამება .1587-ის ტოლი P სიდიდე. ორმხრივი ტესტისთვის P სიდიდე უნდა გავამრავლოთ 2-ზე (*.1587 × 2 = .3174*). ეს ალბათობა ორმხრივი ტესტისთვის ალფაზე (.05) საკმაოდ დიდია. ამიტომ ჩვენ არ შეგვიძლია ნულოვანი ჰიპოთეზის უარყოფა და გამოვაქვს დასკვნა, რომ განსხვავება, რომელიც დაფიქსირდა შერჩევების დონეზე განათლების დონის მიხედვით მთავრობის შეფასებასთან დაკავშირებით, სავარაუდოდ არ არის დამახასიათებელი მთლიანი ერთობლიობებისთვის.

დეტალების განხილვა 9.2
შეცდომის ზღვარი: არის თუ არა მნიშვნელოვანი განსხვავება?

2010 წელს მასმედიის საშუალებებმა გააკეთეს კომენტარი დაავადებათა კონტროლის ცენტრის (CDC) საშიანობაზე, რომელმაც 2005 წლისათვის გამოაქვეყნა ინფორმაცია წყვილების ქორწინების ხანგრძლივობის შესახებ ორი ერთობლიობისათვის: ვინც ქორწინებამდე ერთად

ცხოვრობდა და ვინც არ ცხოვრობდა. ახალი ამბების ზოგიერთმა სააგენტომ გააკეთა შენიშვნა, რომ წყვილების ერთად ცხოვრებას გარკვეული სიძნელეები შეექმნა. ქორწინების შეთე წლისთავზე ქორწინებამდე ადრე ერთად მცხოვრებთა კავშირი უფრო ნაკლებად ინარჩუნებდა სიმტკიცეს, ვიდრე მათი ქორწინებები, ვინც ადრე ერთად არ ცხოვრობდა. დეტალების განხილვამ მნიშვნელოვანი მომენტები გამოიტანა სააშკარაოზე.

CDC-ს წარმომადგენლები გუდვინი, ჩანდრა და მოშერი (2010) წერდნენ, რომ მამაკაცებისთვის ქორწინებამდე თანაცხოვრების გამოცდილება მნიშვნელოვნად იყო დამოკიდებული ქორწინების გადარჩენის ალბათობაზე. მეორე მხრივ, ალბათობა, რომ ქორწინება მინიმუმ 10 წელი მაინც გაგრძელდებოდა, უფრო დაბალი იყო იმ ქალთა შორის, რომლებსაც ქორწინებამდე ჰქონდათ მამაკაცთან თანაცხოვრების გამოცდილება (60%) იმ ქალებთან შედარებით, ვისაც არ გააჩნდა ეს გამოცდილება (66%). მკვლევარები წერდნენ, რომ 2002 წლის მონაცემებში ეს განსხვავება მნიშვნელოვანი არ იყო 5%-ის დონეზე.[8]

ამ თავში უკვე გვქონდა საუბარი იმაზე, რომ ორ საშუალოსა და ხვედრით წილში განსხვავება შეიძლება გამოწვეული იყოს ან მთლიან ერთობლიობაზე რეალური ზემოქმედებით, ან შერჩევის შეცდომით. მნიშვნელოვანი განსხვავება იმაზე მიანიშნებს, რომ დამოუკიდებელი ცვლადის, როგორიცაა თანაცხოვრება ქორწინებამდე, გავლენა მართლა რეალურია. როგორც გამოკვლევიდან ჩანს, მამაკაცებისთვის ამ ფაქტორს დიდი მნიშვნელობა აქვს, ხოლო ქალებისთვის, — არა.

ამიტომ სიფრთხილით მოეკიდეთ განცხადებების გაკეთებას, სადაც აქცენტი არ კეთდება მნიშვნელოვნობის ფაქტორზე. შეიძლება განსხვავება არსებობდეს, მაგრამ უნდა დასვათ კითხვა: რამდენად მნიშვნელოვანია ეს განსხვავება?

კვლევითი ლიტერატურის მიმოხილვა: სტატისტიკური ჰიპოთეზის ტესტირების ანგარიში

ახლა ვნახოთ, როგორ ხდება სოციალურ კვლევით ლიტერატურაში სტატისტიკური ჰიპოთეზის ტესტირების წარმოდგენა. მხედველობაში იქონიეთ ის ფაქტი, რომ კვლევით ლიტერატურაში ასე დეტალურად არ ხდება ამ საკითხის გაშუქება. მაგალითად, სტატიების უმეტესობაში გამოტოვებულია ფორმალური მსჯელობები ნულოვან ჰიპოთეზაზე ან შერჩევით განაწილებაზე. ანალიზის და დეტალიზაციის ფორმატი საჟურნალო პოლიტიკასა და დისციპლინის ფორმატის სტანდარტზეა დამოკიდებული.

ერთი სტატიისთვის სრულიად ბუნებრივია 10-დან 20-მდე ტესტის შედეგის განხილვა. შედეგები წარმოდგენილია მოკლედ და შეჯამებული სახით. დასკვნები ცხრილის სახითაა წარმოდგენილი (შერჩევის სტატისტიკა, მიღებული კრიტერიუმის სტატისტიკა, მნიშვნელოვნობის დონე).[9]

მეცნიერებმა რობერტ ემეტ ჯონსმა და შირლი ა. რეინიმ (2006) შეისწავლეს რასობრივი ნიშნის, გარემო პირობებისა და სამართლის ურთიერთობები.[10] მათ დოკუმენტურად აღწერეს, როგორ უხდებათ ფერადკანიან და ღარიბ ადამიანებს ჯანმრთელობისთვის საშიშ გარემოში ცხოვრება. მათ 2001-2003 წლებში შეისწავლეს ტენესის შტატში წითელი მდინარის რეგიონის თეთრკანიანი და ფერადკანიანი რესპონდენტები.

მეცნიერებმა შექმნეს სკალური ინდექსების სერია, რომელშიც აისახა მოსახლეობის შეხედულებები გარემო პირობებზე. გარემოს დაცვის ინდექსი (EC) მოიცავდა ისეთი საკითხების შეფასებას, როგორიცაა: სასმელი წყალი, ნაგავსაყრელები, ტყვიისა და სხვა საწამლავის შემცველობა საღებავებში, მწვანე ზოლების მდგომარეობა, წყაროები და მდინარეები. EC II ზომავს გარემოს ხარისხს სამეზობლოში, EC III ზომავს გარემოს პრობლემების სერიოზულობას. ინდექსების მაღალი დონე მიანიშნებს პრობლემების სერიოზულობაზე. EH ინდექსი სწავლობს გარემოს ზემოქმედებას ჯანმრთელობაზე: თავის ტკივილი, ნერვული აშლილობა, წონის მნიშვნელოვანი მატება ან კლება, გამონაყარი კანზე ან სუნთქვის პრობლემები. ეს ინდექსი ზომავს ალბათობას, რომ ადამიანებს შეიძლება ჯანმრთელობის პრობლემები შეექმნათ გარემოს დაბინძურების გამო. და ბოლოს, ეკოლოგიური სამართლიანობის ინდექსი EJ, რომელიც ზომავს ადამიანების მოლოდინებს სამთავრობო წრეებიდან ეკოლოგიურ პრობლემებზე რეაგირების ხარისხთან დაკავშირებით. მაღალი ქულა მიანიშნებს, რომ მთავრობა თავს ვერართმევს ამ სახის პრობლემების გადაწყვეტას. ეს ინფორმაცია რასობრივ ჭრილში მოცემულია ცხრილში 9.5.

ცხრილი 9.5 ზრუნვა გარემოზე (EC), გარემოს ჰიგიენა (EH) და ეკოლოგიური სამართლიანობა (EJ)

ინდიკატორი	ჯგუფი	საშუალო	სტანდარტული გადახრა	t	მნიშვნელოვნობა (ცალმხრივი)
EC ინდექსი	შავკანიანები	56.2	13.7	6.2	<0.001
	თეთრკანიანები	42.6	15.5		
EC-II	შავკანიანები	4.4	1.0	5.6	<0.001
	თეთრკანიანები	3.5	1.3		
EC-III	შავკანიანები	3.4	1.1	6.7	<0.001
	თეთრკანიანები	2.3	1.0		
EH ინდექსი	შავკანიანები	23.0	10.5	5.1	<0.001
	თეთრკანიანები	16.0	7.3		
EJ ინდექსი	შავკანიანები	31.0	7.3	3.8	<0.001

წყარო: Robert E. Jones and Shirley A. Rainey, "Examining Linkages Between Race, Environmental Concern, Health and Justice in a Highly Polluted Community of Color," Journal of Black Studies 36, no. 4 (2006): 473–496.

შენიშვნა: $N = 78$ შავკანიანი, 113 თეთრკანიანი.

ახლა უფრო დეტალურად განვიხილოთ ეს ცხრილი. თითოეული სტრიქონი აღწერს მხოლოდ ერთ ინდექსს. აქ მოტანილია საშუალოსა და სტანდარტული გადახრის მონაცემები რასობრივ ჭრილში. ბოლო და ბოლოს წინა სვეტებში მოტანილია მიღებული კრიტერიუმის სტატისტიკა და ალბათობა t კრიტერიუმისთვის ($P < .001$). ყველა ინდექსის შედარება მნიშვნელოვანია .001 დონეზე.

ჯონსი და რეინი კონკრეტულად არ იხილავენ თეთრკანიანებსა და შავკანიანებს

შორის t კრიტერიუმის განსხვავებებს, განზოგადებული სახით მათი დასკვნა შემდეგში მდგომარეობს:

შავკანიანები უფრო განიცდიან ეკოლოგიურ პრობლემებს (EC)... შედეგები... აგრეთვე იმაზე მეტყველებენ, რომ ცუდი გარემო უფრო პრობლემურია შავკანიანთა ჯანმრთელობის მდგომარეობისთვის (EH)... შავკანიანებს უფრო მეტად მიაჩნიათ, რომ მთავრობა და პასუხისმგებელი პირები ვერ წყვეტენ ეკოლოგიურ პრობლემებს ეფექტურად და სამართლიანად (EJ). (გვ. 485).

ძირითადი განმარტებები:

• სტატისტიკური ჰიპოთეზის ტესტირება არის გადაწყვეტილების მიღების პროცესი, რომელიც გვაძლევს საშუალებას განვსაზღვროთ, მოხდება თუ არა კონკრეტული შერჩევის შედეგი შესაძლებლობის მისაღები დიაპაზონის ფარგლებში. ჰიპოთეზის ტესტირების პროცესი ხუთ საფეხურს მოიცავს: (1) პირობების განსაზღვრა, (2) კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას შერჩევა, (3) შერჩევითი განაწილების შერჩევა და კრიტერიუმის სტატისტიკის განსაზღვრა, (4) კრიტერიუმის სტატისტიკის გაანგარიშება, (5) დასკვნის ჩამოყალიბება და შედეგების ინტერპრეტაცია.

• სტატისტიკური ჰიპოთეზის ტესტირება შეიძლება მოიცავდეს მთლიანი ერთობლიობისა და შერჩევის საშუალოების შედარებას, ან ორი შერჩევის საშუალოების შედარებას. თუ საშუალოების შედარებისას ჩვენთვის ცნობილია მთლიანი ერთობლიობის დისპერსი(ები)ა, მაშინ შეგვიძლია Z სტატისტიკის და ნორმალური განაწილების გამოყენება. თუმცა პრაქტიკაში ეს ნაკლებად არის შესაძლებელი.

• თუ ჩვენთვის უცნობია მთლიანი ერთობლიობის დისპერსი(ები)ა, მაშინ საშუალოების შედარებისას შეგვიძლია t სტატისტიკის და t განაწილების გამოყენება.

• ტესტები, რომლებიც მოიცავს ხვედრით წილებს შორის განსხვავებას, იმავე მიმდევრობით შემოწმდება, როგორც განსხვავება საშუალოებს შორის, როდესაც მთლიანი ერთობლიობის დისპერსი(ები)ა ცნობილია. კრიტერიუმის სტატისტიკა ამ შემთხვევაში არის Z, ხოლო შერჩევითი განაწილება მიახლოებულია ნორმალურთან.

საკვანძო ტერმინები

ალფა (α)
თავისუფლების ხარისხები (*df*)
მარცხენამხრივი ტესტი
ნულოვანი ჰიპოთეზა (H_0)
ცალმხრივი ტესტი
P სიდიდე
კვლევითი ჰიპოთეზა (H_1)
მარჯვენამხრივი ტესტი
საშუალოებს შორის განსხვავების შერჩევითი განაწილება
სტატისტიკური ჰიპოთეზის ტესტირება
t განაწილება
t სტატისტიკა (მიღებული)
ორმხრივი ტესტი
I რიგის შეცდომა
II რიგის შეცდომა
Z სტატისტიკა (მიღებული)

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

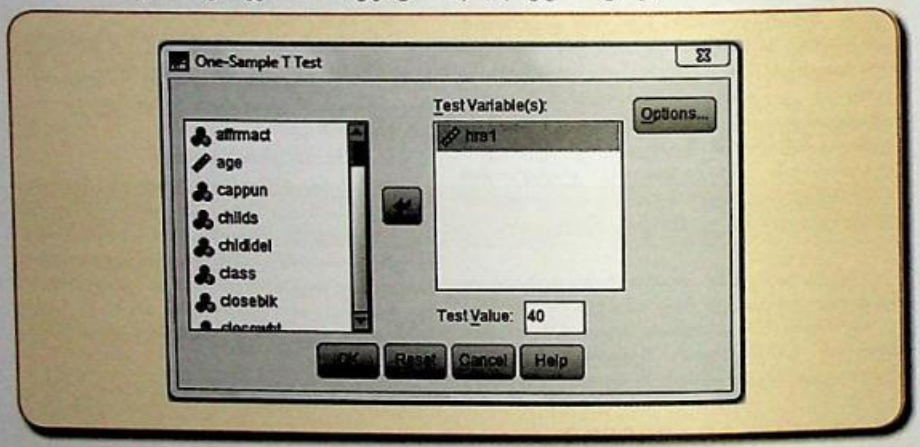
სტატისტიკა SPSS-ის გამოყენებით

[GSS08PFP-A]

მაგალითი 1: ერთ-შერჩევიანი (One-Sample T Test) ტესტის შექმნა

ამ თავში ჩვენ განვიხილეთ მთლიანი და შერჩევითი ერთობლიობების საშუალოებს შორის განსხვავებების ტესტირების მეთოდები. ამ პროცედურის შესასრულებლად SPSS შეიცავს ერთ-შერჩევიან T ტესტს. ამ მიზნისთვის SPSS იყენებს არა Z სტატისტიკას, არამედ – t სტატისტიკას. ერთ-შერჩევიანი T ტესტის პროცედურას იპოვით Analyze მენიუს არჩევანში, შემდეგ კი Compare Means, სადაც ნახავთ ჭდეს One-Sample T Test. დიალოგური ფანჯარა, რომელიც გაიხსნება (ნახ. 9.4) მოგთხოვთ სულ მცირე ერთი ცვლადის მოთავსებას Test Variable(s) box-ში. შემდეგ უნდა დაკონკრეტდეს შესამოწმებელი სიდიდე.

ნახაზი 9.4 ერთ-შერჩევიანი T ტესტის დიალოგური ფანჯარა



თვალსაჩინოებისთვის გამოვიყენებთ GSS 2008-ის მონაცემებს. სტანდარტული სამუშაო კვირა 40 საათია. შევამოწმოთ, მუშაობენ თუ არა ამერიკელები ყოველ კვირა ამდენ საათს. ამისათვის ცვლადი HRS1 მოვათავსოთ ყუთში Test Variable(s) box, ხოლო „40“ – Test Value box-ში. შემდეგ დაკლიკეთ OK.

One-Sample T Test-ის საშედეგო ფანჯარა არ არის მოცულობითი (იხ. ნახაზი 9.5). ყოველკვირეულად ნამუშევარი საათების შესახებ კითხვაზე პასუხი გასცა 894 ადამიანმა. საშუალოდ ნამუშევარი საათების რაოდენობაა 42.26, სტანდარტული გადახრით 14.344. ამის ქვემოთ SPSS-ის დიალოგურ ფანჯარაში მოცემულია სატესტო სიდიდე, 40. ის მოიცავს ორმხრივ მნიშვნელოვნობას, ანუ ალბათობას ერთ-შერჩევიანი ტესტისთვის. ეს სიდიდეა .000, მოცემულია გაანგარიშებული t სტატისტიკა 4.717 – 893 თავისუფლების ხარისხით. ამიტომ, .01 მნიშვნელოვნობის დონეზე, ჩვენ უნდა უკუვაგდოთ ნულოვანი ჰიპოთეზა და დავასკვნათ, რომ მოზრდილი ამერიკელები კვირაში 40 საათზე მეტს მუშაობენ.

SPSS აგრეთვე გვთავაზობს 95% ნდობის ინტერვალს სატესტო და შერჩევითი საშუალოების განსხვავებისათვის. აქ ნდობის ინტერვალი 1.32-დან 3.20-მდეა, რომელიც აფასებს, თუ 40 საათზე რამდენით მეტს მუშაობენ ამერიკელები ყოველკვირეულად.

მაგალითი 2: საშუალოებს შორის განსხვავების ტესტის შექმნა

ამ თავში ჩვენ ავრეთვე შემოგთავაზეთ მეთოდები ორი შერჩევას (ჯგუფს) შორის საშუალოების ან ხვედრითი წილების განსხვავების ტესტირებისთვის. ეს პროცედურა Two-Sample T Test უნდა იპოვოთ Analyze მენიუს არჩევანში. შემდეგ ქვემოთ Compare Means, სადაც ნახავთ ჭდეს Independent-Samples T Test.

ნახაზი 9.5 One-Sample T Test-ის საშედეგო ფანჯარა

One-Sample Statistics

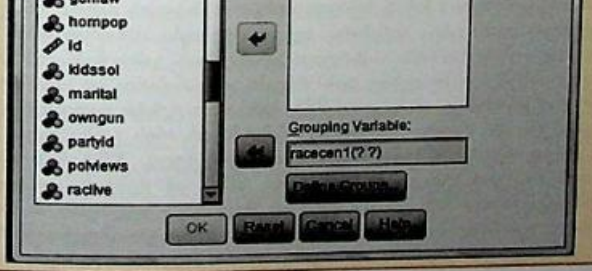
	N	Mean	Std. Deviation	Std. Error Mean
NUMBER OF HOURS WORKED LAST WEEK	894	42.26	14.344	.480

One-Sample Test

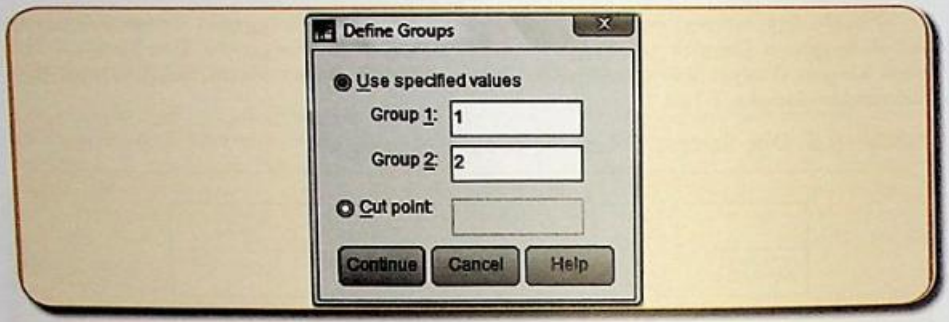
	Test Value = 40					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
NUMBER OF HOURS WORKED LAST WEEK	4.717	893	.000	2.263	1.32	3.20

გახსნილი დიალოგური ყუთი მოგთხოვთ სატესტო ცვლადებს მითითებას: დამოკიდებულს (dependent variable) და ერთ დამოუკიდებელ (independent) ან დაჯგუფების ცვლადს (grouping variable) (ნახ. 9.6). ჩვენ უნდა მოვახდინოთ ნულოვანი ჰიპოთეზის ტესტირება ცვლადით HRS1, რომ თეთრკანიანები და შავკანიანები თანაბარი რაოდენობის საათების განმაელობაში მუშობენ. მოათავსეთ ეს ცვლადი ყუთში Test Variable(s) box და RACECEN1 ყუთში Grouping Variable box. როდესაც ამას გააკეთებთ, RACECEN1-თან გამოვა კითხვა, თქვენ უნდა მიუთითოთ ორი სიდიდე ორი ჯგუფის შესაქმნელად (independent samples), დაკლიკეთ Define Groups. შემდეგ ჩასვით „1" პირველ ყუთში, „2" კი მეორე ყუთში (1 = თეთრკანიანები და 2 = შავკანიანები), როგორც ნაჩვენებია ნახაზზე 9.7. შემდეგ დაკლიკეთ Continue და OK პროცედურის დასამთავრებლად.

ნახაზი 9.6 Independent-Samples T Test-ის დიალოგური ფანჯარა



ნახაზი 9.7 Define Groups-ის დიალოგური ფანჯარა



Independent Samples T test-ის საშედეგო (ნახაზი 9.8) ცხრილი უფრო დეტალურია და მეტ ინფორმაციას მოიცავს. ცხრილის პირველ ნაწილში მოცემულია შავკანიანებისა და თეთრკანიანების მიერ ნამუშევარი საათების საშუალო რაოდენობა, რესპონდენტების რაოდენობა თითოეულ ჯგუფში, სტანდარტული გადახრა და საშუალოს სტანდარტული შეცდომა. როგორც ვხედავთ, ყოველკვირეულად თეთრკანიანები 1.35 საათით უფრო მეტს მუშაობენ, ვიდრე თეთრკანიანები (42.48-41.13=1.35).

ნახაზი 9.8 Independent-Samples T Test-ის საშედეგო ფანჯარა

Group Statistics

	WHAT IS RS RACE 1ST MENTION	N	Mean	Std. Deviation	Std. Error Mean
NUMBER OF HOURS WORKED LAST WEEK	WHITE	658	42.48	14.090	.549
	BLACK OR AFRICAN AMERICAN	136	41.13	15.221	1.305

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NUMBER OF HOURS WORKED LAST WEEK	Equal variances assumed	.029	.865	1.003	792	.316	1.349	1.346	-1.293	3.991
	Equal variances not assumed			.953	185.855	.342	1.349	1.416	-1.444	4.143

გაიხსენეთ ამ თავიდან, რომ t სტატისტიკის გაანგარიშებისთვის ძალიან მნიშვნელოვანია, არის თუ არა ორი ჯგუფის დისპერსია თანაბარი. თუ დისპერსიები თანაბრად არის მიჩნეული, მაშინ შეგიძლიათ თავისუფლების ხარისხის მარტივი ფორმულის გამოყენება. SPSS-ს შეუძლია იმ შემთხვევის გათვალისწინებაც, როდესაც დისპერსიები არათანაბარია და აღნიშნულის შესაბამისი t სტატისტიკის გაანგარიშებაც. ამისთვის SPSS პირდაპირ ტესტს იყენებს, რომ გამოააშკარაოს არის თუ არა დისპერსიები ტოლი (ეს ტესტი ტექსტში არ არის განხილული). ეს არის ლევინის ტესტი, რომელიც მეორე საშედეგო ცხრილშია მოცემული. ტესტის ნულოვანი ჰიპოთეზა მდგომარეობს დისპერსიების ტოლობის დაშვებაში. SPSS ამ ტესტის ალბათობად საზღვრავს .865-ს. ასე რომ, მაგალითად .05 დონეზე ჩვენ არ შეგვიძლია ნულოვანი ჰიპოთეზის უარყოფა. ანუ ჩვენ ვასკვნით, რომ ნულოვანი ჰიპოთეზა სწორია. ჩვენი გადაწყვეტილების თანახმად, უკვე ვიცით, რომელი t ტესტი გამოვიყენოთ. ამისთვის ვიყენებთ ლევინის t ტესტის პირველ სტრიქონს.

ეს სიდიდეა 1.003, რომლის თავისუფლების ხარისხია 792. ორმხრივი მნიშვნელოვნობაა (Sig. [2-tailed]) 0.316. SPSS ანგარიშობს ზუსტ მნიშვნელოვნობას, ამიტომ დანართში ჩახედვა აღარ არის საჭირო. ეს ალბათობა ბევრად აღემატება .05 ან .01 დონეს, აქედან ჩვენ ვასკვნით, რომ მნიშვნელოვანი განსხვავება ნამუშევარი საათების რაოდენობას შორის არ არის. ჩვენ არ შეგვიძლია ნულოვანი ჰიპოთეზის უარყოფა.

რა მოხდება, თუ ჩვენ ამის ნაცვლად ცალმხრივ ტესტს გამოვიყენებთ? SPSS ამას პირდაპირ არ აკეთებს, მაგრამ გაანგარიშება არ არის ძნელი. თუ კვლევით ჰიპოთეზას განუსაზღვრავთ მიმართულებას – მაგალითად ისე, რომ მამაკაცები უფრო მეტს მუშაობენ, ვიდრე ქალები, – ჩვენ უნდა ავირჩიოთ SPSS-ის მიერ განსაზღვრული ალბათობა და ორზე გავყოთ ცალმხრივი ჰიპოთეზისათვის. რადგან ალბათობა მაღალია, ცალმხრივ ტესტს გავაკეთებთ, თუ ორმხრივს, ორივე შემთხვევაში დასკვნა იგივე იქნება.

ცხრილის ბოლოს თითოეულ სტრიქონზე მოცემულია 95% ნდობის ინტერვალი ჯგუფური საშუალოების განსხვავებისთვის. აღნიშნული საკითხი შეგიძლიათ გადაამოწმოთ 8-ე თავში. ეს ძალიან მნიშვნელოვანია, რადგან რეალური განსხვავება საშუალოებს შორის შერჩევის საშუალოთი იზომება (აქ 1.349 სთ.), რომელიც ერთი შერჩევიდან მეორემდე იცვლება. ეს ინტერვალი კი გვაძლევს დიაპაზონს, რომელშიც მოხვდება საშუალოს ვარიაცია.

ამოცანები SPSS-ისთვის

1. გამოიყენეთ 2008 GSS (GSS08PFP-B) ფაილი იმის გამოსაკვლევად, ჰყავს თუ არა ერთ ამერიკელს საშუალოდ ორი ბავშვი. გამოიყენეთ One Sample T Test პროცედურა ამ ტესტისთვის, ცვლადით CHILDS. მნიშვნელოვნობის დონეა .01. რა აღმოაჩინეთ? ნამდვილად ორი შვილი ჰყავს ერთ ამერიკელს, თუ მეტი ან ნაკლები?
2. გამოიკვლიეთ განსხვავება ინდივიდებს შორის, რომლებიც მხარს უჭერენ ან არ უჭერენ მარიჰუანის ლეგალიზაციას GSS08PFP-B ბაზიდან. ცვლადი GRASS მიიჩნიეთ დამოკიდებულ ან ჯგუფის ცვლადად (1 = ლეგალური და 2 = არალეგალური). გამოავლინეთ, არის თუ არა ამ ჯგუფებს შორის მნიშვნელოვანი განსხვავება ასაკის (AGE), ბავშვების რაოდენობის (CHILDS) და განათლების დონის (EDUC) მიხედვით. ალფა არის .05 ორმხრივი ტესტისთვის. დაწერეთ შემაჯამებელი ანალიზი.
3. განავრცეთ 2-ე სავარჯიშოს ანალიზი ამჯერად ინდივიდების შედარებით, ვინც მხარს უჭერს/არ უჭერს სიკვდილით დასჯას, (CAPPUN). გამოიყენეთ იგივე დამოკიდებული ცვლადები, AGE, CHILDS, and EDUC, t ტესტის შეფასებისთვის. ალფა არის .05 ორმხრივი ტესტისთვის. დაწერეთ შემაჯამებელი ანალიზი.
4. GSS08PFP-B მოიცავს მიღებული განათლების უმაღლეს დონეს (DEGREE). არის თუ არა მნიშვნელოვანი განსხვავება საშუალო სკოლადამთავრებულებს (კოდი 0) და ბაკალავრებს (კოდი 3) შორის ტელევიზორის ყურებისა და (TVHOURS) საჯარო ბიბლიოთეკებში ვიზიტების რაოდენობის მიხედვით (VISLIB). ალფა არის .05 ორმხრივი ტესტისთვის. დაწერეთ შემაჯამებელი ანალიზი.
5. ამ სავარჯიშოსთვის გამოიყენებთ MTF 2008 მონაცემებს და შეადარებთ GPA-ს სტუდენტების რასობრივი/ეთნიკური ჯგუფების მიხედვით. პირველად უნდა ნახოთ ცვლადის RACE სიხშირეები. გაითვალისწინეთ, რომ თითოეული რასობრივი/ეთნიკური ჯგუფი კოდირებულია (მაგ. შავკანიანი სტუდენტების კოდია „1“). MTF 2008 მოიცავს სამ კატეგორიას ცვლადისთვის RACE; თქვენ ორი მათგანი უნდა შეარჩიოთ როგორც დამოუკიდებელი (ფაქტორული) ცვლადი. GPA იქნება დამოკიდებული ცვლადი. გაიანგარიშეთ t ტესტი GPA-ს ორი სახის შედარებისთვის: მაგ. შავკანიანები და თეთრკანიანები და შავკანიანები და ესპანური წარმოშობის სტუდენტები. ალფა არის .05 ორმხრივი ტესტისთვის.

თავის სავარჯიშოები

1. ცნობილია, რომ ჯანდაცვის ორგანიზაციაში მომუშავე ექიმს უნდა ჰქონდეს საშუალოდ 13.5 წლიანი გამოცდილება, სტანდარტული გადახრით 7.6 წელი. ჯანდაცვის ორგანიზაციის დასავლეთ შტატის აღმასრულებელ დირექტორს აინტერესებს, აქვთ თუ არა მის ექიმებს საშუალო ეროვნულზე ნაკლები გამოცდილება. 150 კაციანი შერჩევის შედეგად აღმოჩნდა, რომ მათი საშუალო გამოცდილება მხოლოდ 10.9 წელია.
ა. ჩამოაყალიბეთ კვლევითი და ნულოვანი ჰიპოთეზები ამ მოსაზრების ტესტირებისთვის.
ბ. მოახდინეთ ტესტირება, თუ ამ ტესტში ალფა .01-ის ტოლია.
2. სამხრეთ-დასავლეთის სამხედრო ბაზაზე უსაფრთხოების სამსახურის დანიშნულებაა შემოჭრილი ჯაშუშების გამოვლენა. თუმცა, ხანდახან, სიგნალიზაცია არ მუშაობს, ან ოფიცრები ვერ ამჩნევენ მას. ნებისმიერ მოცემულ დროს უსაფრთხოების პერსონალმა უნდა ამოირჩიოს ორი ალტერნატივიდან ერთი:

H_0: ყველაფერი კარგადაა; არავინ ცდილობს შემოჭრას.

H_1: არსებობს პრობლემები; ვიღაც ცდილობს შემოჭრას.

ამ ინფორმაციიდან გამომდინარე, შეავსეთ გამოტოვებული ადგილები დებულებებში:
ა. „შეუმჩნეველი განგაში“ არის —— გვარის შეცდომა და მისი მოხდენის ალბათობა აღინიშნება როგორც ——.
ბ. „ცრუ განგაში“ არის —— შეცდომა.
3. თითოეული სიტუაციისთვის განსაზღვრეთ რომელი ტესტია შესაბამისი: ცალმხრივი, თუ ორმხრივი. ჩამოაყალიბეთ კვლევითი და ნულოვანი ჰიპოთეზები.
ა. თქვენ გაინტერესებთ, განსხვავდება თუ არა თქვენი შტატის რეზიდენტების საშუალო შემოსავალი ეროვნული საშუალოსაგან. 2010 წლის აღწერის მიხედვით, ერთი შინამეურნეობის საშუალო შემოსავალმა შეადგინა $50,303.
ბ. თქვენ მიგაჩნიათ, რომ ლიბერალური ხელოვნების კოლეჯების სტუდენტები ყოველთვიურად უფრო მეტ წვეულებას ესწრებიან, ვიდრე ზოგადად სტუდენტები ქვეყნის მასშტაბით. სტუდენტები ქვეყნის მასშტაბით თვეში საშუალოდ 3.2 წვეულებას მართავენ. ლიბერალური ხელოვნების კოლეჯის სტუდენტების შერჩევა მოხდა შემთხვევითი წესით ამ ტიპის კოლეჯებიდან.
გ. სოციოლოგს მიაჩნია, რომ ხანდაზმული ქალების შემოსავალი ხანდაზმულ მამაკაცებზე ნაკლებია.
დ. არის თუ არა განსხვავება კამპუსში და მის გარეთ მცხოვრები სტუდენტების სასწავლო საათებს შორის საშუალო კვირის განმავლობაში? თქვენ გსურთ მოამზადოთ შერჩევა თითოეული ჯგუფის სტუდენტების მიერ დახარჯული საშუალო დროის შესახებ.
ე. მოსალოდნელია, რომ გაძლიერებული სწავლების მესამე კლასელთა ქულები კითხვაში უფრო მაღალი იქნება ჩვეულებრივ მესამე კლასელებთან შედარებით.
ვ. პროგნოზის თანახმად, სტრესის დონე (იზომება ორდინალურ დონეზე. უფრო დაბალია იმ მოზრდილებში, ვისაც შინაური ცხოველი ჰყავს სახლში.
4.
ა. სავარჯიშო 3-ის თითოეული სიტუაციისთვის აღწერეთ I და II გვარის შეცდომები.
ბ. დაახასიათეთ I და II გვარის შეცდომები.
გ. რა შემთხვევებში არის საჭირო I ან II გვარის შეცდომების მინიმიზაცია?
5. შერჩევის რეპრეზენტატულობის შესამოწმებლად ერთმანეთს ადარებენ მთლიანი ერთობლიობისა და შერჩევის მახასიათებლებს. ამ მიზნისთვის ტიპური ცვლადია ასაკი. GSS 2008-ის შერჩევის საფუძველზე გამოვლინდა მოზრდილი ამერიკელების საშუალო ასაკი, რომელიც 47.71 წლის ტოლია, სტანდარტული გადახრით 17.35 წელი. შერჩევაში მონაწილეობდა 2,013 მოზრდილი. ამასთან, მოსახლეობის აღწერიდან მივიღეთ, რომ ამერიკელი მოზრდილების საშუალო ასაკი 37.7 წელია. გამოიყენეთ ეს ინფორმაცია და უპასუხეთ შემდეგ კითხვებს.

ა. ჩამოაყალიბეთ კვლევითი და ნულოვანი ჰიპოთეზები ორმხრივი ტესტისთვის.
ბ. გაიანგარიშეთ t სტატისტიკა და ჩამოაყალიბეთ ნულოვანი ჰიპოთეზა .001 მნიშვნელოვნობის დონეზე. რა აღმოაჩინეთ?
გ. რა გადაწყვეტილება მიიღეთ ნულოვან ჰიპოთეზასთან დაკავშირებით? არის თუ არა შერჩევა რეპრეზენტატული?

6. სიგარეტის პირველად მოხმარების საშუალო კლასის გამოსავლენად გამოიყენეთ ცხრილში 9.4 მოტანილი ინფორმაცია. გამოიყენეთ t სტატისტიკა და ჩაატარეთ ნულოვანი ჰიპოთეზის ცალმხრივი ტესტირება, რომ მამაკაცებისთვის საშუალო კლასი უფრო მაღალია, ვიდრე ქალებისთვის. ალფა მიიჩნიეთ .05-ის ტოლად. რა დასკვნა გამოიტანეთ? შეიცვლება თუ არა თქვენი დასკვნა, თუ ორმხრივ ტესტს გამოიყენებთ?
7. ამ სავარჯიშოში ჩვენ შევამოწმებთ ლიბერალებისა და კონსერვატორების შეხედულებას სამუშაო ადგილზე პოზიტიური დისკრიმინაციის შესახებ. GSS 2008-ის მონაცემებიდან ჩანს, რომ ლიბერალთა 10% ($N = 424$). და კონსერვატორთა 28% ($N = 336$). „ძალიან უჭერს მხარს“ და „უჭერს მხარს“ აფრიკელი ამერიკელების პოზიტიურ დისკრიმინაციას სამუშაო ადგილზე.
ა. რომელია შესაბამისი კრიტერიუმის სტატისტიკა? რატომ?
ბ. მოახდინეთ ნულოვანი ჰიპოთეზის ტესტირება ცალმხრივი ტესტით (კონსერვატორები უფრო ნაკლებად უჭერენ მხარს ამ მოსაზრებას); ალფა .05-ია. რა დასკვნა გამოიტანეთ ლიბერალებისა და კონსერვატორების შეხედულებებს შორის განსხვავების შესახებ?
გ. თუ ორმხრივ ტესტს ჩაატარებთ, სადაც ალფა .05-ია, შეიცვლება თუ არა თქვენი გადაწყვეტილება?

8. განვაგრძოთ ლიბერალებისა და კონსერვატორების შეხედულებების შესწავლას ამჯერად მათი განათლების ხანგრძლივობის გათვალისწინებით. GSS 2008-ის გამოკვლევიდან ჩანს, რომ ლიბერალებისთვის სწავლის საშუალო ხანგრძლივობა 13.90 წელია ($S_y = 3.27$), ხოლო კონსერვატორებისთვის – 13.55 წელი ($S_y = 2.82$). მონაცემები აღებულია 187 ლიბერალისა და 227 კონსერვატორის გამოკითხვიდან.
ა. მოახდინეთ კვლევითი ჰიპოთეზის ტესტირება, რომ ლიბერალებისა და კონსერვატორების განათლების ხანგრძლივობა განსხვავებულია, ალფა უდრის .01-ს.
ბ. შეიცვლება თუ არა თქვენი გადაწყვეტილება, თუ ალფა .05-ია?

9. 2008 წლის საპრეზიდენტო არჩევნების დროს თვლიდნენ, რომ ამომრჩეველთა სქესი მნიშვნელოვანი იყო ჰილარი კლინტონისთვის. სოციოლოგები მუდმივად აქცენტებდნენ ანგარიშებს იმის თაობაზე, რომ ასაკოვანი ქალები უფრო მეტად უჭერდნენ მას მხარს, ვიდრე კაცები. კვლევითმა კომპანიამ Pew Research Centre გამოაქვეყნა ინფორმაცია, რომ 240 გამოკითხული მამაკაციდან 41%-მა პირველ არჩევნად მიუთითა კლინტონი, ხოლო 381 გამოკითხული ქალიდან 49%-მა მიუთითა იგივე. არის თუ არა გენდერული განსხვავება კლინტონის მხარდამჭერებს შორის?
ა. თქვენი კვლევითი ჰიპოთეზა იმაში მდგომარეობს, რომ ქალების უფრო მაღალი ხვედრითი წილი უჭერს მხარს კლინტონს, ვიდრე მამაკაცების. რომელ ტესტს აირჩევთ, ცალმხრივს თუ ორმხრივს?
ბ. მოახდინეთ ნულოვანი ჰიპოთეზის ტესტირება .05 ალფას დონეზე.
გ. შეიცვლება თუ არა თქვენი გადაწყვეტილება, თუ ალფა .01-ია?

10. MTF 2008-ის გამოკვლევის თანახმად ალკოჰოლის მოხმარება დაადასტურა ქალების 75.7%-მა (651-დან 493). და მამაკაცების 70.4%-მა (712-დან 501). თქვენ გაინტერესებთ, არის თუ არა განსხვავება ქალებისა და მამაკაცების ერთობლიობებში ალკოჰოლის მოხმარებაში (ეს ცვლადი არ გულისხმობს რეგულარულ მოხმარებას, მხოლოდ პირველ ცდას). გამოიყენეთ პროპორციებს შორის განსხვავების ტესტი აღნიშნულის შესასწავლად.
ა. რა არის აქ კვლევითი ჰიპოთეზა? რომელ ტესტს ჩაატარებთ, ცალმხრივს თუ ორმხრივს? რატომ?
ბ. მოახდინეთ ჰიპოთეზის ტესტირება .05 დონეზე. რა დასკვნა გამოიტანეთ?

11. მეცნიერთა ჯგუფმა შეისწავლა ინტერნეტის მოხმარება ამერიკელ და ინდოელ სტუდენტებს შორის.[11] ინტერნეტის გამოყენება ორ კატეგორიად დაიყო: პერსონალური და საკურსოზე.

ქვემოთ ცხრილში მოცემულია ეს განსხვავებები.

ინტერნეტის საშუალო მოხმარება და სტანდარტული გადახრა ამერიკელ და ინდოელ სტუდენტებს შორის

	საკურსოზე დაბარებული საათები	პერსონალური საათები
ამერიკელი სტუდენტები	$\overline{Y} = 1.76$	$\overline{Y} = 2.08$
$N = 149$	$S_Y = 1.52$	$S_Y = 1.91$
ინდოელი სტუდენტები	$\overline{Y} = 0.73$	$\overline{Y} = 0.87$
$N = 306$	$S_Y = 0.79$	$S_Y = 0.78$

წყარო: Marcelline Fusilier, Subhash Durlabhji, Alain Cucchi, and Michael Collins, "A four country investigation of factors facilitating student internet use," CyberPsychology and Behavior 8(5): 454–464. Copyright © Mary Ann Liebert, Inc. Publishers.

ა. განსაზღვრეთ, მოიხმარენ თუ არა ამერიკელები ინტერნეტს უფრო მეტად საკურსო სამუშაოს შესასრულებლად ინდოელებთან შედარებით. ალფა უდრის .05-ს.

ბ. არის თუ არა მნიშვნელოვანი განსხვავება ამერიკელებისა და ინდოელების მიერ ინტერნეტის პერსონალურ მოხმარებაში? ალფა უდრის .01-ს.

12. აქვთ თუ არა ქალებსა და მამაკაცებს განსხვავებული მოსაზრება ოჯახში ბავშვების იდეალური რაოდენობის შესახებ? გამოიყენეთ GSS 2008-ის მონაცემები და მიღებული t სტატისტიკა, რა დასკვნა გამოიტანეთ? (შეხსენება: გამოიყენეთ ორმხრივი ტესტი, ალფა=.05).

	მამაკაცები	ქალები
ბავშვების იდეალური საშუალო რაოდენობა	3.06	3.22
სტანდარტული გადახრა	1.92	1.99
N	604	678
მიღებული t სტატისტიკა	−1.450	

13. განსხვავდება თუ არა კოკაინთან მისაწვდომობა სქესის მიხედვით? MTF 2008-ის გამოკვლევიდან ჩანს, რომ მამაკაცთა 18%-მა ($N = 670$) გასცა პასუხი, რომ „ძალიან ადვილია", ხოლო ქალების 14%-მა ($N = 723$). არის თუ არა მნიშვნელოვანი განსხვავება ამ პროპორციებს შორის? ალფა უდრის .01-ს. რა დასკვნა გამოიტანეთ?

14. ბავშვების იდეალური რაოდენობა შეადარეთ ქალების ჯგუფებს შორის, ვინც „ძალიან ბედნიერია" და „არც ისე ბედნიერია". შედეგები ასახულია ქვემოთ ცხრილში.

	მამაკაცები	ქალები
ბავშვების იდეალური საშუალო რაოდენობა	3.49	2.73
სტანდარტული გადახრა	2.24	1.35
N	228	90
მიღებული t სტატისტიკა	2.98	

რა დასკვნას გამოიტანთ ორმხრივი ტესტის გამოყენებით, როდესაც ალფა=.05-ს?

თავი 10

ურთიერთკავშირი ორ ცვლადს შორის

კროს-ტაბულაცია

თავის სწავლების მიზნები

- ბივარიაციული ცხრილის შედგენა
- არაერთგვაროვანი ურთიერთკავშირები ცვლადებს შორის
- ბივარიაციული ურთიერთკავშირის მახასიათებლების განსაზღვრა: არსებობა, სიძლიერე და მიმართულება
- ურთიერთკავშირის შემუშავება ცვლადებს შორის: არამცდარობა, ჩარევა, პირობითი ურთიერთკავშირები

სოციალური მეცნიერების ერთ-ერთი უმნიშვნელოვანესი მიზანია მოვლენებს შორის ურთიერთკავშირის ახსნის საფუძველზე ჩაწვდეს ჰუმანური და სოციალური გამოცდილების არსს. სწორედ ამიტომ, ურთიერთკავშირის ენა სოციალური მეცნიერების საფუძველთა საფუძველია. განვიხილოთ რამდენიმე მაგალითი სტატიებიდან და სამეცნიერო ანგარიშებიდან:

მაგალითი 1. 45 წელზე ზევით ასაკის მოსახლეობა უფრო მეტად უჭერს მხარს 2010 წლის არიზონის საიმიგრაციო კანონს, ვიდრე ახალგაზრდობა. ახალი კანონმდებლობა პოლიციას ნებას რთავს დააკავოს ქვეყანაში არალეგალურად შემოჭრილი პირები[1] (ეს მაგალითი გვიჩვენებს ურთიერთკავშირს ასაკსა და საიმიგრაციო რეფორმას შორის).

მაგალითი 2. სტერეოტიპის საწინააღმდეგოდ, თეთრკანიანები უფრო ხშირად იყენებენ სოციალური დაცვის პროგრამებს, ვიდრე შავკანიანები და ლათინო-ამერიკელები, ამიტომ ეროვნულ უმცირესობებთან შედარებით, მათ სიღარიბისგან თავის დაღწევის უფრო მეტი შანსი აქვთ.[2] (ეს მაგალითი გვიჩვენებს ურთიერთკავშირს რასასა და სახელმწიფო დახმარების მიღებას შორის).

მაგალითი 3. საშუალოდ, შავკანიანებსა და ლათინო-ამერიკელებს ჯანდაცვის მომსახურების მიღების უფრო ნაკლები შანსი აქვთ, ვიდრე თეთრკანიანებს.[3] (ეს მაგალითი გვიჩვენებს ურთიერთკავშირს რასასა და ჯანდაცვის მომსახურების მიღებას შორის).

თითოეულ შემთხვევაში ურთიერთკავშირი იმის მანიშნებელია, რომ ერთი ცვლადის ზოგიერთი მნიშვნელობა მეორე ცვლადის ზოგიერთ მნიშვნელობასთან არის დაკავშირებული. მაგალითად, პირველ შემთხვევაში საიმიგრაციო კანონი დაწყვილებულია ასაკთან, მეორე შემთხვევაში, რასობრივი ფაქტორი დაწყვილებულია სოციალური დახმარების მიღების სიხშირესთან, მესამე შემთხვევაშიც რასობრივი ფაქტორია დაწყვილებული ჯანდაცვის მომსახურების მიღებასთან, თუმცა თეთრკანიანები უფრო მაღალი ალბათობით არიან დაწყვილებული ამ მომსახურებასთან, ხოლო ლათინო-ამერიკელები უფრო ნაკლები ალბათობით.

ამ თავში თქვენ გაეცნობით ორ ცვლადს შორის ურთიერთკავშირის ანალიზში გამოყენებულ ერთ-ერთ ყველაზე გავრცელებულ მეთოდს, რომელსაც *კროს-ტაბულაცია* (ჯვარედინი შეჯამება) ეწოდება. *კროს-ტაბულაცია* ცხრილის სახით წარმოდგენილ ორ ცვლადს შორის ურთიერთკავშირის ანალიზის მეთოდია. კროს-ტაბულაცია ბივარიაციული ანალიზის სახეობაა, რომელიც გამოიყენება ორი ნომინალური ან ორდინალური ცვლადის ურთიერთკავშირის აღმოჩენისა და აღწერისათვის. ჩვენ არამარტო აღმოვაჩენთ, არამედ განვსაზღვრავთ კიდეც ამ ურთიერთკავშირის სიძლიერეს და სადაც შესაძლებელია, მის მიმართულებასაც. აგრეთვე ვნახავთ, თუ როგორ გამოიყენება ეს მეთოდები რეალური კვლევის დროს. ბივარიაციულ, ანუ ორგანზომილებიან ანალიზს ჩვენ ერთხელ უკვე შევეხეთ მე-9 თავში ორ საშუალოს ან ხვედრით წილს შორის განსხვავების განსაზღვრისათვის t და Z სიდიდეების დახმარებით.[4]

კროს-ტაბულაცია ცხრილის სახით წარმოდგენილ ორ ნომინალურ ან ორდინალურ ცვლადს შორის ურთიერთკავშირის ანალიზის მეთოდია.

ბივარიაციული ანალიზი მეთოდი, რომელიც გამოიყენება ორი ნომინალური ან ორდინალური ცვლადის ურთიერთკავშირის აღმოჩენისა და აღწერისათვის.

დამოუკიდებელი და დამოკიდებული ცვლადები

სოციალურ მეცნიერებებში კვლევის დიზაინისა და სტატისტიკის მნიშვნელოვანი ასპექტია *დამოუკიდებელი* და *დამოკიდებული ცვლადების* განსხვავება. ეს ტერმინები პირველივე თავში გაგაცანით, ისინი ამ და მომდევნო თავებშიც ხშირად გამოიყენება, ამიტომ მნიშვნელოვანია მათ შორის განსხვავების გაგება. განვიხილოთ, მაგალითად, გარემო პირობების გავლენა ადამიანების ცხოვრების დონის უარყოფით ცვლილებებზე. მიუხედავად იმისა, რომ გარემო პირობების აღქმა პირდაპირ ზეგავლენას არ ახდენს პიროვნების გადაწყვეტილებაზე მიიღოს

გარემო პირობების გამო გარკვეული უარყოფითი ცვლილებები, ის პიროვნების გადაწყვეტილების მიღებაზე ზემოქმედებს მისი გამოცდილების – განათლების, დასაქმებისა და სოციალიზაციის საკითხებში – საფუძველზე, რაც მისი შეხედულებებისა და აზრების ჩამოყალიბებაზე ახდენს გავლენას. თუ ჩავთვლით, რომ ადამიანის მიერ გარემო პირობების გამო უარყოფითი ცვლილებების მიღება (ეს არის ცვლადი, რომელიც უნდა ახსნას მკვლევარმა) იცვლება გამომდინარე მისი დამოკიდებულებიდან გარემოსადმი (რომელიც სავარაუდოდ მონაწილეობს გადაწყვეტილების მიღების პროცესში), მაშინ გარემოს მისეული აღქმა დამოუკიდებელი ცვლადია, ხოლო უარყოფითი ცვლილებების მიღება – დამოკიდებული ცვლადი.

თითოეულ მაგალითში, რომელიც დასაწყისში მოვიტანეთ, ორი ცვლადია: დამოუკიდებელი და დამოკიდებული. პირველ მაგალითში მკვლევარის მიზანია საიმიგრაციო კანონმდებლობის რეფორმის მხარდაჭერის ახსნა. სავარაუდოდ, ამ მხარდაჭერასთან დაკავშირებული ცვლადია ასაკი. ამიტომ, საიმიგრაციო კანონმდებლობის რეფორმის მხარდაჭერა დამოკიდებული ცვლადია, ხოლო ასაკი – დამოუკიდებელი. მეორე მაგალითში მკვლევარი სწავლობს რასობრივი ნიშნის გავლენას სამთავრობო დახმარების გამოყენების სტერეოტიპზე. ამიტომ სამთავრობო დახმარების გამოყენება დამოკიდებული ცვლადია, ხოლო რასა – დამოუკიდებელი. აღნიშნულის მსგავსად, ჯანდაცვის მომსახურების მისაწვდომობა დამოკიდებული ცვლადია, რადგან ეს არის საკვლევი ცვლადი, ხოლო რასა, რომელიც გვეხმარება მის ახსნაში, დამოუკიდებელია.

ამ თავში, ისევე როგორც 12-ე და 13-ე თავებში, განხილულია ის მეთოდები, რომლებიც საშუალებას მოგვცემს შევაფასოთ დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის ურთიერთკავშირის სიმჭიდროვე.

ზოგიერთი ცვლადის როლი – დამოუკიდებელია ის თუ დამოკიდებული – კვლევის კითხვიდან გამომდინარეობს. ამიტომ, თუ ჯერ კიდევ ვერ ასხვავებთ ერთმანეთისგან დამოუკიდებელ და დამოკიდებულ ცვლადებს, გადახედეთ პირველ თავს.

✓ *ცოდნის შემოწმება*

▣ როგორ შევქმნათ ბივარიაციული ცხრილი: რასა და ბინის მფლობელობა

ბივარიაციულ ცხრილში ერთი ცვლადი მეორე ცვლადის კატეგორიების გასწვრივ არის განლაგებული. ის მიიღება დაკვირვებების კლასიფიკაციით ორი ნომინალური ან ორდინალური ცვლადის ქულების შერწყმის საფუძველზე. შეიძლება ითქვას, რომ ეს არის ერთ ცხრილში გაერთიანებული სიხშირული განაწილების სერიები. ცხრილში 10.1 მოტანილია GSS-ის კვლევის შედეგები, სადაც რესპონდენტები განაწილებული არიან რასისა და ბინის ფლობის სტატუსის მიხედვით (ამ შემთხვევაში ორივე ცვლადი ნომინალურ დონეზე იზომება).

თითოეულ შემთხვევაში ურთიერთკავშირი იმის მანიშნებელია, რომ ერთი ცვლადის ზოგიერთი მნიშვნელობა მეორე ცვლადის ზოგიერთ მნიშვნელობასთან არის დაკავშირებული. მაგალითად, პირველ შემთხვევაში საიმიგრაციო კანონი დაწყვილებულია ასაკთან, მეორე შემთხვევაში, რასობრივი ფაქტორი დაწყვილებულია სოციალური დახმარების მიღების სიხშირესთან, მესამე შემთხვევაშიც რასობრივი ფაქტორია დაწყვილებული ჯანდაცვის მომსახურების მიღებასთან, თუმცა თეთრკანიანები უფრო მაღალი ალბათობით არიან დაწყვილებული ამ მომსახურებასთან, ხოლო ლათინო-ამერიკელები უფრო ნაკლები ალბათობით.

ამ თავში თქვენ გაეცნობით ორ ცვლადს შორის ურთიერთკავშირის ანალიზში გამოყენებულ ერთ-ერთ ყველაზე გავრცელებულ მეთოდს, რომელსაც *კროს-ტაბულაცია* (ჯვარედინი შეჯამება) ეწოდება. *კროს-ტაბულაცია* ცხრილის სახით წარმოდგენილ ორ ცვლადს შორის ურთიერთკავშირის ანალიზის მეთოდია. კროს-ტაბულაცია ბივარიაციული ანალიზის სახეობაა, რომელიც გამოიყენება ორი ნომინალური ან ორდინალური ცვლადის ურთიერთკავშირის აღმოჩენისა და აღწერისათვის. ჩვენ არამარტო აღმოვაჩენთ, არამედ განვსაზღვრავთ კიდეც ამ ურთიერთკავშირის სიძლიერეს და სადაც შესაძლებელია, მის მიმართულებასაც. აგრეთვე ვნახავთ, თუ როგორ გამოიყენება ეს მეთოდები რეალური კვლევის დროს. ბივარიაციულ, ანუ ორგანზომილებიან ანალიზს ჩვენ ერთხელ უკვე შევეხეთ მე-9 თავში ორ საშუალოს ან ხვედრით წილს შორის განსხვავების განსაზღვრისათვის t და Z სიდიდეების დახმარებით.[4]

კროს-ტაბულაცია ცხრილის სახით წარმოდგენილ ორ ნომინალურ ან ორდინალურ ცვლადს შორის ურთიერთკავშირის ანალიზის მეთოდია.

ბივარიაციული ანალიზი მეთოდი, რომელიც გამოიყენება ორი ნომინალური ან ორდინალური ცვლადის ურთიერთკავშირის აღმოჩენისა და აღწერისათვის.

დამოუკიდებელი და დამოკიდებული ცვლადები

სოციალურ მეცნიერებებში კვლევის დიზაინისა და სტატისტიკის მნიშვნელოვანი ასპექტია *დამოუკიდებელი* და *დამოკიდებული ცვლადების* განსხვავება. ეს ტერმინები პირველივე თავში გაგაცანით, ისინი ამ და მომდევნო თავებშიც ხშირად გამოიყენება, ამიტომ მნიშვნელოვანია მათ შორის განსხვავების გაგება. *განვიხილოთ*, მაგალითად, გარემო პირობების გავლენა ადამიანების ცხოვრების ხანის უარყოფით ცვლილებებზე. მიუხედავად იმისა, რომ გარემო პირობების ცვლა პირდაპირ ზეგავლენას არ ახდენს პიროვნების გადაწყვეტილებაზე მიიღოს

გარემო პირობების გამო გარკვეული უარყოფითი ცვლილებები, ის პიროვნების გადაწყვეტილების მიღებაზე ზემოქმედებს მისი გამოცდილების — განათლების, დასაქმებისა და სოციალიზაციის საკითხებში — საფუძველზე, რაც მისი შეხედულებებისა და აზრების ჩამოყალიბებაზე ახდენს გავლენას. თუ ჩავთვლით, რომ ადამიანის მიერ გარემო პირობების გამო უარყოფითი ცვლილებების მიღება (ეს არის ცვლადი, რომელიც უნდა ახსნას მკვლევარმა) იცვლება გამომდინარე მისი დამოკიდებულებიდან გარემოსადმი (რომელიც სავარაუდოდ მონაწილეობს გადაწყვეტილების მიღების პროცესში), მაშინ გარემოს მისეული აღქმა დამოუკიდებელი ცვლადია, ხოლო უარყოფითი ცვლილებების მიღება — დამოკიდებული ცვლადი.

თითოეულ მაგალითში, რომელიც დასაწყისში მოვიტანეთ, ორი ცვლადია: დამოუკიდებელი და დამოკიდებული. პირველ მაგალითში მკვლევარის მიზანია საიმიგრაციო კანონმდებლობის რეფორმის მხარდაჭერის ახსნა. სავარაუდოდ, ამ მხარდაჭერასთან დაკავშირებული ცვლადია ასაკი. ამიტომ, საიმიგრაციო კანონმდებლობის რეფორმის მხარდაჭერა დამოკიდებული ცვლადია, ხოლო ასაკი — დამოუკიდებელი. მეორე მაგალითში მკვლევარი სწავლობს რასობრივი ნიშნის გავლენას სამთავრობო დახმარების გამოყენების სტერეოტიპზე. ამიტომ სამთავრობო დახმარების გამოყენება დამოკიდებული ცვლადია, ხოლო რასა — დამოუკიდებელი. აღნიშნულის მსგავსად, ჯანდაცვის მომსახურების მისაწვდომობა დამოკიდებული ცვლადია, რადგან ეს არის საკვლევი ცვლადი, ხოლო რასა, რომელიც გვეხმარება მის ახსნაში, დამოუკიდებელია.

ამ თავში, ისევე როგორც 12-ე და 13-ე თავებში, განხილულია ის მეთოდები, რომლებიც საშუალებას მოგვცემს შევაფასოთ დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის ურთიერთკავშირის სიმჭიდროვე.

ზოგიერთი ცვლადის როლი — დამოუკიდებელია ის თუ დამოკიდებული — კვლევის კითხვიდან გამომდინარეობს. ამიტომ, თუ ჯერ კიდევ ვერ ასხვავებთ ერთმანეთისგან დამოუკიდებელ და დამოკიდებულ ცვლადებს, გადახედეთ პირველ თავს.

✓ *ცოდნის შემოწმება*

▣ როგორ შევქმნათ ბივარიაციული ცხრილი: რასა და ბინის მფლობელობა

ბივარიაციულ ცხრილში ერთი ცვლადი მეორე ცვლადის კატეგორიების გასწვრივ არის განლაგებული. ის მიიღება დაკვირვებების კლასიფიკაციით ორი ნომინალური ან ორდინალური ცვლადის ქულების შერწყმის საფუძველზე. შეიძლება ითქვას, რომ ეს არის ერთ ცხრილში გაერთიანებული სიხშირული განაწილების სერიები. ცხრილში 10.1 მოტანილია GSS-ის კვლევის შედეგები, სადაც რესპონდენტები განაწილებული არიან რასისა და ბინის ფლობის სტატუსის მიხედვით (ამ შემთხვევაში ორივე ცვლადი ნომინალურ დონეზე იზომება).

თითოეულ შემთხვევაში ურთიერთკავშირი იმის მანიშნებელია, რომ ერთი ცვლადის ზოგიერთი მნიშვნელობა მეორე ცვლადის ზოგიერთ მნიშვნელობასთან არის დაკავშირებული. მაგალითად, პირველ შემთხვევაში საიმიგრაციო კანონი დაწყვილებულია ასაკთან, მეორე შემთხვევაში, რასობრივი ფაქტორი დაწყვილებულია სოციალური დახმარების მიღების სიხშირესთან, მესამე შემთხვევაშიც რასობრივი ფაქტორია დაწყვილებული ჯანდაცვის მომსახურების მიღებასთან, თუმცა თეთრკანიანები უფრო მაღალი ალბათობით არიან დაწყვილებული ამ მომსახურებასთან, ხოლო ლათინო-ამერიკელები უფრო ნაკლები ალბათობით.

ამ თავში თქვენ გაეცნობით ორ ცვლადს შორის ურთიერთკავშირის ანალიზში გამოყენებულ ერთ-ერთ ყველაზე გავრცელებულ მეთოდს, რომელსაც *კროს-ტაბულაცია* (ჯვარედინი შეჯამება) ეწოდება. *კროს-ტაბულაცია* ცხრილის სახით წარმოდგენილ ორ ცვლადს შორის ურთიერთკავშირის ანალიზის მეთოდია. კროს-ტაბულაცია ბივარიაციული ანალიზის სახეობაა, რომელიც გამოიყენება ორი ნომინალური ან ორდინალური ცვლადის ურთიერთკავშირის აღმოჩენისა და აღწერისათვის. ჩვენ არამარტო აღმოვაჩენთ, არამედ განვსაზღვრავთ კიდეც ამ ურთიერთკავშირის სიძლიერეს და სადაც შესაძლებელია, მის მიმართულებასაც. აგრეთვე ვნახავთ, თუ როგორ გამოიყენება ეს მეთოდები რეალური კვლევის დროს. ბივარიაციულ, ანუ ორგანზომილებიან ანალიზს ჩვენ ერთხელ უკვე შევეხეთ მე-9 თავში ორ საშუალოს ან ხვედრით წილს შორის განსხვავების განსაზღვრისათვის t და Z სიდიდეების დახმარებით.[4]

კროს-ტაბულაცია ცხრილის სახით წარმოდგენილ ორ ნომინალურ ან ორდინალურ ცვლადს შორის ურთიერთკავშირის ანალიზის მეთოდია.

ბივარიაციული ანალიზი მეთოდი, რომელიც გამოიყენება ორი ნომინალური ან ორდინალური ცვლადის ურთიერთკავშირის აღმოჩენისა და აღწერისათვის.

დამოუკიდებელი და დამოკიდებული ცვლადები

სოციალურ მეცნიერებებში კვლევის დიზაინისა და სტატისტიკის მნიშვნელოვანი ასპექტია *დამოუკიდებელი* და *დამოკიდებული ცვლადების* განსხვავება. ეს ტერმინები პირველივე თავში გაგაცანით, ისინი ამ და მომდევნო თავებშიც ხშირად გამოიყენება, ამიტომ მნიშვნელოვანია მათ შორის განსხვავების გაგება. განვიხილოთ, მაგალითად, გარემო პირობების გავლენა ადამიანების ცხოვრების დონის უარყოფით ცვლილებებზე. მიუხედავად იმისა, რომ გარემო პირობების აღქმა პირდაპირ ზეგავლენას არ ახდენს პიროვნების გადაწყვეტილებაზე მიიღოს

გარემო პირობების გამო გარკვეული უარყოფითი ცვლილებები, ის პიროვნების გადაწყვეტილების მიღებაზე ზემოქმედებს მისი გამოცდილების — განათლების, დასაქმებისა და სოციალიზაციის საკითხებში — საფუძველზე, რაც მისი შეხედულებებისა და აზრების ჩამოყალიბებაზე ახდენს გავლენას. თუ ჩავთვლით, რომ ადამიანის მიერ გარემო პირობების გამო უარყოფითი ცვლილებების მიღება (ეს არის ცვლადი, რომელიც უნდა ახსნას მკვლევარმა) იცვლება გამომდინარე მისი დამოკიდებულებიდან გარემოსადმი (რომელიც სავარაუდოდ მონაწილეობს გადაწყვეტილების მიღების პროცესში), მაშინ გარემოს მისეული აღქმა დამოუკიდებელი ცვლადია, ხოლო უარყოფითი ცვლილებების მიღება — დამოკიდებული ცვლადი.

თითოეულ მაგალითში, რომელიც დასაწყისში მოვიტანეთ, ორი ცვლადია: დამოუკიდებელი და დამოკიდებული. პირველ მაგალითში მკვლევარის მიზანია საიმიგრაციო კანონმდებლობის რეფორმის მხარდაჭერის ახსნა. სავარაუდოდ, ამ მხარდაჭერასთან დაკავშირებული ცვლადია ასაკი. ამიტომ, საიმიგრაციო კანონმდებლობის რეფორმის მხარდაჭერა დამოკიდებული ცვლადია, ხოლო ასაკი — დამოუკიდებელი. მეორე მაგალითში მკვლევარი სწავლობს რასობრივი ნიშნის გავლენას სამთავრობო დახმარების გამოყენების სტერეოტიპზე. ამიტომ სამთავრობო დახმარების გამოყენება დამოკიდებული ცვლადია, ხოლო რასა — დამოუკიდებელი. აღნიშნულის მსგავსად, ჯანდაცვის მომსახურების მისაწვდომობა დამოკიდებული ცვლადია, რადგან ეს არის საკვლევი ცვლადი, ხოლო რასა, რომელიც გვეხმარება მის ახსნაში, დამოუკიდებელია.

ამ თავში, ისევე როგორც 12-ე და 13-ე თავებში, განხილულია ის მეთოდები, რომლებიც საშუალებას მოგვცემს შევაფასოთ დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის ურთიერთკავშირის სიმჭიდროვე.

ზოგიერთი ცვლადის როლი — დამოუკიდებელია ის თუ დამოკიდებული — კვლევის კითხვიდან გამომდინარეობს. ამიტომ, თუ ჯერ კიდევ ვერ ასხვავებთ ერთმანეთისგან დამოუკიდებელ და დამოკიდებულ ცვლადებს, გადახედეთ პირველ თავს.

✓ *ცოდნი შემოწმება*

▣ როგორ შევქმნათ ბივარიაციული ცხრილი: რასა და ბინის მფლობელობა

ბივარიაციულ ცხრილში ერთი ცვლადი მეორე ცვლადის კატეგორიების გასწვრივ არის განლაგებული. ის მიიღება დაკვირვებების კლასიფიკაციით ორი ნომინალური ან ორდინალური ცვლადის ქულების შერწყმის საფუძველზე. შეიძლება ითქვას, რომ ეს არის ერთ ცხრილში გაერთიანებული სიხშირული განაწილების სერიები. ცხრილში 10.1 მოტანილია GSS-ის კვლევის შედეგები, სადაც რესპონდენტები განაწილებული არიან რასისა და ბინის ფლობის სტატუსის მიხედვით (ამ შემთხვევაში ორივე ცვლადი ნომინალურ დონეზე იზომება).

ცხრილი 10.1 რასა და ბინის მფლობელობა GSS-ის 17 რესპონდენტის მიხედვით

რესპონდენტი	რასა	ბინის მფლობელობა
1	შავკანიანი	ფლობს
2	შავკანიანი	ფლობს
3	თეთრკანიანი	ქირაობს
4	თეთრკანიანი	ქირაობს
5	თეთრკანიანი	ფლობს
6	თეთრკანიანი	ფლობს
7	თეთრკანიანი	ფლობს
8	შავკანიანი	ქირაობს
9	შავკანიანი	ქირაობს
10	შავკანიანი	ქირაობს
11	თეთრკანიანი	ფლობს
12	თეთრკანიანი	ფლობს
13	თეთრკანიანი	ქირაობს
14	თეთრკანიანი	ფლობს
15	შავკანიანი	ქირაობს
16	თეთრკანიანი	ფლობს
17	შავკანიანი	ქირაობს

ამ მონაცემებში გასარკვევად უნდა ავაგოთ ცხრილი, რომელშიც კლასიფიცირებული იქნება მოცემული ინდივიდუალური მნიშვნელობები. ცხრილში 10.2 17 რესპონდენტი კლასიფიცირებულია რასისა და ბინის მფლობელობის ნაერთი ქულების მიხედვით.

ცხრილი 10.2 ბინის მფლობელობა რასის მიხედვით (აბსოლუტური სიხშირეები), GSS 2006

ბინის მფლობელობა	რასა: შავი	რასა: თეთრი		
ფლობს	2	7	9	სტრიქონების მარგინალები (სტრიქონების ჯამი)
ქირაობს	5	3	8	
	7	10	17	ქეისები სულ (N)
	სვეტების მარგინალები (სვეტების ჯამი)			

ამ ცხრილს აქვს ტიპური ბივარიაციული ცხრილისათვის დამახასიათებელი თვისებები:

1. ცხრილის სათაური აღწერითია და მის შინაარსს განსაზღვრავს ორი ცვლადის საშუალებით.
2. მას ორი განზომილება აქვს: ერთი რასობრივი ნიშნისთვის და მეორე ბინის მფლობელობისთვის. *საბლის მფლობელობა* წარმოდგენილია ცხრილის სტრიქონებში: ერთი სტრიქონი ეთმობა მფლობელებს და მეორე დამქირავებლებს. ცხრილის სვეტები ფორმირებულია *რასის* მიხედვით. ცხრილს შეიძლება უფრო მეტი სვეტი და სტრიქონი ჰქონდეს, რაც იმაზეა დამოკიდებული, თუ რამდენი კატეგორია აქვს ცვლადს. მაგალითად, თუ ლათინო-ამერიკელებს ჩავამატებდით რასაში, გვექნებოდა სამი სვეტი (სტრიქონების ჯამის სვეტის გარდა). უფრო ხშირად, დამოუკიდებელი ცვლადი **სვეტის ცვლადია**, ხოლო დამოკიდებული — **სტრიქონის ცვლადი**.
3. სვეტებისა და სტრიქონების გადაკვეთას **უჯრედს** უწოდებენ. მაგალითად, ცხრილის მარცხენა ზედა უჯრედში წარმოდგენილია 2 შავკანიანი პირი, რომელიც ფლობს ბინას.
4. სვეტებისა და სტრიქონების ჯამები თითოეული ცვლადის სიხშირულ განაწილებას წარმოადგენს. სვეტების ჯამი *რასის* სიხშირული განაწილებაა, ხოლო სტრიქონების ჯამი — *ბინის მფლობელობის*. ზოგჯერ სვეტებისა და სტრიქონების ჯამებს **მარგინალებს,** ან შეჯამებულ სიხშირეებს უწოდებენ. (ეს ელემენტები ცხრილში დაჭდეებულია).
5. ეს ცხრილი არის 2×2 ცხრილი, რადგან მას ორი სვეტი და 2 სტრიქონი აქვს. ჩვეულებრივ ამ ცხრილს r×c ცხრილს ეძახიან, სადაც r სტრიქონებია, ხოლო c — სვეტები. ცხრილში, სადაც სტრიქონის ცვლადს სამი კატეგორია აქვს, ხოლო სვეტის ცვლადს — 2, ეწოდება 3×2 ცხრილი.
6. მონაცემთა წყარო აგრეთვე გარკვევით უნდა იყოს მითითებული. ეს შეესაბამება მონაცემთა თავმოყრის პირობებს და ფორმულირებულია 2-ე თავში.

ბივარიაციული ცხრილი ცხრილი, რომელშიც ერთი ცვლადის განაწილება სხვა ცვლადის კატეგორიების გასწვრივ არის განლაგებული.
სვეტის ცვლადი ცვლადი, რომლის კატეგორიები ბივარიაციული ცხრილის სვეტებშია მოთავსებული.
სტრიქონის ცვლადი ცვლადი, რომლის კატეგორიები ბივარიაციული ცხრილის სტრიქონებშია მოთავსებული.
უჯრედი სვეტისა და სტრიქონის ურთიერთგადაკვეთა ბივარიაციულ ცხრილში.
მარგინალები სვეტებისა და სტრიქონის ჯამები ბივარიაციულ ცხრილში.

✓ **ცოდნის შემოწმება**

კარგად შეისწავლეთ ცხრილი 10.2 და დარწმუნდით, რომ ყველაფერი სწორად გაიგეთ. შეგიძლიათ თუ არა ამ ცხრილში დამოუკიდებელი და დამოკიდებული ცვლადების განსაზღვრა? აღნიშნულის ცოდნა მნიშვნელოვანია სიხშირეების გარდაქმნისათვის პროცენტებად.

როგორ გავიანგარიშოთ პროცენტები ბივარიაციულ ცხრილში

ბინათმფლობელობის შედარებისათვის შავკანიანებსა და თეთრკანიანებს შორის საჭიროა სტრიქონების ნედლი სიხშირეების გარდაქმნა პროცენტებად, რადგან, სვეტების ჯამები არათანაბარია. 2-ე თავიდან გაიხსენეთ, რომ პროცენტები ხშირად გამოიყენება სხვადასხვა ზომის ორი ან მეტის ჯგუფის შედარებისთვის. არსებობს ბივარიაციულ ცხრილში პროცენტების გაანგარიშებისა და ანალიზის ორი ძირითადი წესი:

1. პროცენტების გაანგარიშება დამოუკიდებელი ცვლადის ყველა კატეგორიისათვის.
2. ცხრილის ინტერპრეტაცია უნდა მოხდეს დამოუკიდებელი ცვლადის ყველა კატეგორიის პროცენტული განსხვავებების საფუძველზე.

პროცენტების გაანგარიშება დამოუკიდებელი ცვლადის ყველა კატეგორიისათვის

პირველი წესი იმაში მდგომარეობს, რომ პროცენტების გაანგარიშება უნდა მოხდეს ცვლადის ყველა კატეგორიისათვის, რომელსაც მკვლევარი დამოუკიდებელ ცვლადად მიიჩნევს. როდესაც დამოუკიდებელი *ცვლადი სვეტებში* მოთავსდება, თითოეული სვეტისთვის ცალკე უნდა გავიანგარიშოთ პროცენტები. თითოეული უჯრედისა და სტრიქონის მარგინალების სიხშირეები იმ სვეტების ჯამით არის გაყოფილი, რომლებშიც ისინი არიან მოთავსებული, ხოლო სვეტების ჯამები 100% უნდა იყოს. როდესაც დამოუკიდებელი ცვლადი თავსდება სტრიქონებში, თითოეული სტრიქონისთვის ცალკე უნდა გავიანგარიშოთ პროცენტები. თითოეული უჯრედისა და სვეტის მარგინალების სიხშირეები იმ სტრიქონების ჯამით არის გაყოფილი, რომლებშიც ისინი არიან მოთავსებული, ხოლო სტრიქონების ჯამები 100% უნდა იყოს.

ჩვენ მაგალითში გვაინტერესებს *რასის*, როგორც დამოუკიდებელი ცვლადის ურთიერთკავშირი *ბინის მფლობელობასთან*. ამიტომ უნდა გამოვთვალოთ თითოეული რასობრივი ჯგუფის პროცენტები სვეტების ჯამების საფუძველზე. მაგალითად, შავკანიანი ბინათმფლობელი რესპონდენტების პროცენტის მისაღებად, შავკანიანი ბინათმფლობელი რესპონდენტების რაოდენობა უნდა გავყოთ შავკანიანთა მთლიან რაოდენობაზე შერჩევაში:

$$(100)\frac{2}{7} = 28.6\%$$

ცხრილი 10.3 ბინის მფლობელობა რასობრივ ჭრილში (პროცენტობით)

რესპონდენტი	რასა		სულ
	შავი	თეთრი	
ფლობს	28.6%	70.0%	52.9%
ქირაობს	71.4%	30.0%	47.1%
სულ	100%	100%	100%
(*N*)	(7)	(10)	(17)

ცხრილში 10.3 გაანგარიშებული პროცენტები ეფუძნება ცხრილი 10.2-ის მონაცემებს. ყოველთვის უჩვენეთ *N*-ის მნიშვნელობები, რადგან აქედან ხდება პროცენტების გაანგარიშება — ამ შემთხვევაში სვეტების ჯამების.

დამოუკიდებელი ცვლადის სხვადასხვა კატეგორიების პროცენტების შედარება

მეორე წესის შესაბამისად, უნდა შევადაროთ ბინათმფლობელობის სტატუსი თეთრკანიანებსა და შავკანიანებს შორის. აქ უნდა მოვახდინოთ განსხვავებების შედარება დამოუკიდებელი ცვლადის სხვადასხვა კატეგორიებს შორის. ზოგიერთი მკვლევარი შედარების ზღვრად მინიმუმ 10%-იან განსხვავებას აწესებს. ჩვენ მაგალითში ვხედავთ, რომ თეთრკანიან (70%) და შავკანიან (28.6%) ბინის მესაკუთრეებს შორის განსხვავება 41.4 პროცენტია. სხვა სიტყვებით, უფრო დიდი ალბათობაა, რომ ბინის მესაკუთრეები თეთრკანიანები იქნებიან, ვიდრე შავკანიანები.[5] ამიტომ, შეიძლება დავასკვნათ, რომ რასობრივი ნიშანი დაკავშირებულია ბინის ფლობის ალბათობასთან.

ასეთივე დასკვნამდე მივალთ, თუ რასობრივ ჭრილში შევადარებთ ბინის ქირაობის მაჩვენებლებს. რადგან თითოეულ რასობრივ ჯგუფში ბინის მფლობელებისა და დამქირავებლების პროცენტების ჯამი 100%-ის ტოლია, ჩვენ მხოლოდ ერთი შედარება გვჭირდება. ფაქტობრივად, 2×2 ცხრილში მხოლოდ ერთი შედარებაა საჭირო მისი ინტერპრეტაციისთვის. უფრო დიდ ცხრილში ერთზე მეტი შედარება იქნება საჭირო.

იუარჯიშეთ ბივარიაციული ცხრილის აგებაში. გამოიყენეთ ცხრილი 10.1 და შექმენით ბივარიაციული პროცენტული ცხრილი. თქვენი ცხრილი შეადარეთ ცხრილს 10.3. გახსოვთ, ყველა ნაწილი? სწორია, თქვენი დასკვნები? თუ არა, ხელახლა დაუბრუნდით ტექსტს. ამაში დაგეხმარებათ დეტალების განხილვა 10.1. ის ასახავს ბივარიაციული ცხრილების აგებისა და პროცენტების პოვნის პროცესს. დაიმახსოვრეთ, თქვენ უნდა სწორად განსაზღვროთ დამოუკიდებელი ცვლადი და მიხვდეთ, მისი შესაბამისი პროცენტის ადგილი სტრიქონებშია, თუ სვეტებში.

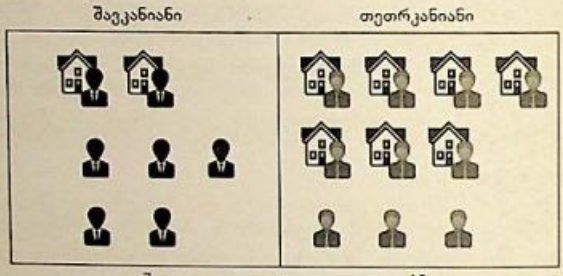
✓ *ცოდნის შემოწმება*

დეტალების განხილვა 10.1
ბივარიაციული ცხრილის პროცენტების პოვნა

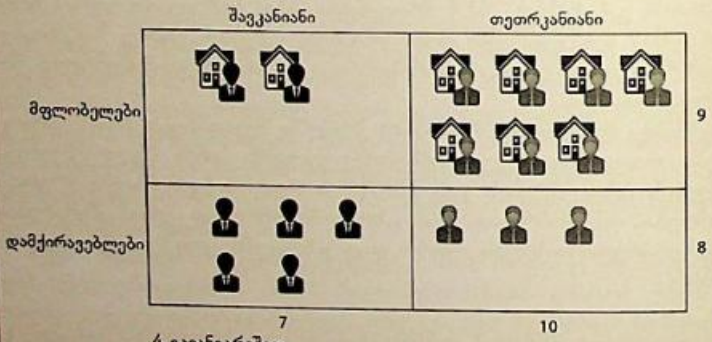
1. თეთრკანიანი და შავკანიანი ბინის მფლობელები და დამქირავებლები



2. დაყავით რესპონდენტები რასის მიხედვით (დამოუკიდებელი ცვლადი); დათვალეთ თითოეულ ჯგუფში მათი რაოდენობა სვეტების ჯამების მისაღებად.



3. დაყავით თითოეული ჯგუფი ბინის მფლობელებისა და დამქირავებლების მიხედვით (დამოკიდებული ცვლადი); დათვალეთ თითოეულ ჯგუფში მათი რაოდენობა სტრიქონების ჯამების მისაღებად.



4. გაიანგარიშეთ თითოეული უჯრედი

	შავი	თეთრი	
მფლობელები	2	7	9
დამქირავებლები	5	3	8
	7	10	

5. შავკანიანი მფლობელების პროცენტი: (100)2/7 = 29%
თეთრკანიანი მფლობელების პროცენტი: (100)7/10 = 70%
შავკანიანი დამქირავებლების პროცენტი: (100)5/7 = 71%
თეთრკანიანი დამქირავებლების პროცენტი: (100)3/10 = 30%

6. შეადარეთ პროცენტები: 29% და 70%
71% და 30%

როგორ მოვიქცეთ ცვლადებს შორის არაერთგვაროვანი ურთიერთკავშირის შემთხვევაში

ზოგჯერ გაურკვეველია, ცვლადი დამოუკიდებელია, თუ დამოკიდებული. ამიტომ სხვა გზა უნდა მოვნახოთ აღნიშნულის დასაზუსტებლად. კერძოდ, შეგიძლიათ გაიანგარიშოთ სტრიქონებისა და სვეტების პროცენტები. მაგალითად, ცხრილში 10.4 წარმოდგენილია ცვლადებისთვის SPANKING და FEFAM (GSS 2008, 127 რესპონდენტი) მონაცემების სამი სვეტი: (ა) აბსოლუტური სიხშირეები, (ბ) სვეტის პროცენტები, და (გ) სტრიქონის პროცენტები. ცვლადის SPANKING კვლევის კითხვაა: „მიმართავთ თუ არა ბავშვის ფიზიკურ დასჯას აღზრდის პროცესში?“. ცვლადი FEFAM ზომავს ეთანხმება თუ არა რესპონდენტი შემდეგ მოსაზრებას: „მამაკაცმა უნდა იმუშაოს და ქალი სახლში უნდა იჯდეს“. ცხრილიდან 10.4ბ ჩანს, რომ რესპონდენტები, რომლებიც სრულიად არ ეთანხმებიან ბავშვის ფიზიკურ დასჯას, უფრო ნაკლებად იზიარებენ მეორე დებულებას (10% და 50%). ცხრილიდან 10.4გ ავრეთვე ჩანს, რომ პირები, რომლებიც სრულიად ეთანხმებიან მოსაზრებას, რომ მამაკაცმა უნდა იმუშაოს და ქალი სახლში უნდა იჯდეს, უფრო მეტად ეთანხმებიან ბავშვის ფიზიკურ დასჯას, ვიდრე ისინი, ვინც არ იზიარებს მამაკაცისა და ქალის როლებს მოცემულ დებულებაში (94% და 63% შესაბამისად).

ცხრილი 10.4 პროცენტების გაანგარიშების სხვადასხვა გზები: ცვლადი SPANKING ცვლადი FEFAM-ის მიხედვით

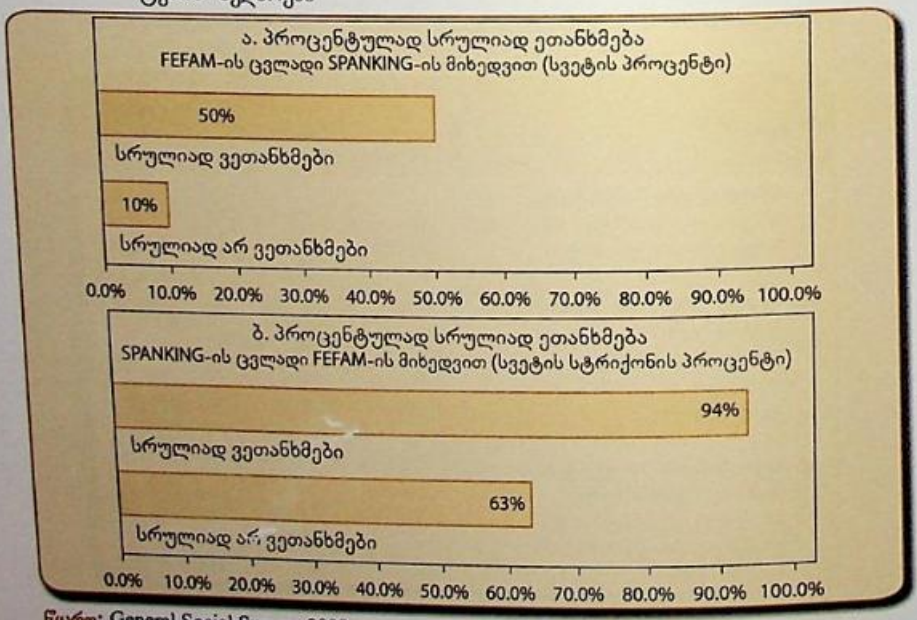
FEFAM	SPANKING		სტრიქონების ჯამი
	სრულიად ვეთანხმები	*სრულიად არ ვეთანხმები*	
ა. აბსოლუტური სიხშირეები			
სრულიად ვეთანხმები	48	3	51
სრულიად არ ვეთანხმები	48	28	76
სვეტის ჯამი	96	31	127
ბ. სვეტის პროცენტი (ბაზა — სვეტის ჯამი)			
სრულიად ვეთანხმები	50%	10%	40%
სრულიად არ ვეთანხმები	50%	90%	60%
სვეტის ჯამი	100%	100%	100%
	(96)	(31)	(127)
გ. სტრიქონების პროცენტები (ბაზა — სვეტის ჯამი)			
სრულიად ვეთანხმები	94%	6%	100% (51)
სრულიად არ ვეთანხმები	63%	37%	100% (76)
სვეტის ჯამი	76%	24%	100% (127)

წყარო: General Social Survey, 2008.

რადგან ჩვენ ვსაზღვრავთ თითოეული *სვეტის* პროცენტს (ცხრილი 10.4ბ), ეს გვეხმარება იმ ჰიპოთეზის შემოწმებაში, რომ დასჯა (დამოუკიდებელი ცვლადი) დაკავშირებულია FEFAM-ის დებულებასთან (დამოკიდებული ცვლადი). როდესაც ჩვენ ვსაზღვრავთ თითოეული *სტრიქონის* პროცენტს (ცხრილი 10.4გ), ვამოწმებთ იმ ჰიპოთეზას, რომ FEFAM-ის დებულებასთან დათანხმება ან არდათანხმება (დამოუკიდებელი ცვლადი) შეიძლება დაკავშირებული იყოს SPANKING-თან (დამოკიდებული ცვლადი).[6] ნახაზი 10.1ა და ბ წარმოადგენს ცხრილში 10.4ბ და გ წარმოდგენილი პროცენტების გაანგარიშებისა და შედარების ორი მეთოდის გრაფიკულ ილუსტრაციას.

და ბოლოს, მნიშვნელოვანია იმის გაგება, თუ რით ხელმძღვანელობს მკვლევარი ბივარიაციული ცხრილის აგებისა და ინტერპრეტაციის დროს მის წინაშე მდგომი კითხვის ასახსნელად. მიუხედავად იმისა, რომ ეს კონკრეტული მაგალითი გვეხმარება სვეტებისა და სტრიქონების პროცენტების ანალიზში, ყველა მონაცემის ამგვარი ინტერპრეტაცია შეუძლებელია. მაგალითად, მამაკაცებისა და ქალების შეხედულებები სამუშაო ადგილზე სექსუალური ძალადობის შესახებ მხოლოდ ერთი მიმართულებით შეიძლება აიხსნას. სქესმა შეიძლება გავლენა მოახდინოს პიროვნების შეხედულებაზე სექსუალური ძალადობის შესახებ, თუმცა სექსუალური ძალადობა სქესზე ვერ მოახდენს ზეგავლენას. ამგვარად, ცვლადების განთავსების მიხედვით სვეტებსა და სტრიქონებში, უნდა გამოვიყენოთ ან სვეტების პროცენტები, ან სტრიქონის პროცენტები და არა ორივე.

ნახაზი 10.1 სტრიქონებისა და სვეტების მიხედვით გაანგარიშებული პროცენტების შედარება



წყარო: General Social Survey, 2008.

კვლევითი ლიტერატურის მიმოხილვა: გარდაცვალების ადგილი ამერიკაში

ის მითითებები, რომლებზეც ამ თავში გავამახვილეთ ყურადღება, ყოველთვის მკაცრად არ არის დაცული ბივარიაციული ცხრილების აგებისა და ანალიზის დროს. პროფესიულ ლიტერატურაში გამოყენებული ბივარიაციული ცხრილები აქ განხილულზე ბევრად უფრო რთულია. განვიხილოთ ზოგიერთი მათგანი. ეს მაგალითი მოტანილია ავტორთა ჯგუფის მიერ 2007 წელს ჩატარებული კვლევის მასალების მიხედვით, რომელიც შეეხება გარდაცვალების ადგილების ვარიაციას ამერიკაში. მათი კვლევის შედეგად დაფიქსირდა ამერიკელების შეხედულების ამაღლება გარდაცვალებამდე მოვლის ხარისხზე. გაირკვა, რომ რესპონდენტების უმრავლესობა გარდაცვალების ადგილად სახლს ირჩევს, თუმცა უმეტესობა სიცოცხლეს მზრუნველობის სხვადასხვა დაწესებულებებში ამთავრებს. ამ მხივ, ქრონიკულად დაავადებულთათვის, გადაუდებელი სამედიცინო დახმარების საავადმყოფოები პირველ ადგილზეა. მეცნიერები ამ საკითხს ასე ხსნიან:

> ინდივიდების გადაწყვეტილებაზე, თუ სად გაატარონ სიცოცხლის უკანასკნელი დღეები, ისევე ახდენს გავლენას მათი დემოგრაფიული და სამედიცინო მახასიათებლები, როგორც მათი დაინტერესების ხარისხი აღნიშნული საკითხით და კეთილდღეობის დონე (გვ. 359).

მოცემული კვლევის საფუძველზე მეცნიერებმა შეისწავლეს გარდაცვალების განსხვავებული ადგილები სქესის, ასაკის, ქორწინებითი მდგომარეობის, რასის/ეთნიკურობის, განათლების დონისა და გარდაცვალების მიზეზის მიხედვით. მათ გამოიყენეს ბუნებრივი მოძრაობის სტატისტიკის ეროვნული სისტემის (NVSS) 1997 წლის მონაცემები და შეისწავლეს გარდაცვალების 1,402,167 შემთხვევა გარდაცვალების სხვადასხვა ადგილის მიხედვით.

შედეგები მოტანილია ცხრილში 10.5.

განვიხილოთ ცხრილი შემდეგი საფეხურების მიხედვით:

1. განსაზღვრეთ დამოკიდებული ცვლადი და დაკვირვების ობიექტი მისი დახასიათებისთვის (ინდივიდი, ქალაქი ან ბავშვი). აქ დამოკიდებული ცვლადია *გარდაცვალების ადგილი 1997 წელს*. დაკვირვების ობიექტია ინდივიდი.

2. განსაზღვრეთ ცხრილში წარმოდგენილი დამოუკიდებელი ცვლადები და მათი კატეგორიები. აქ ხუთი დამოუკიდებელი ცვლადია: სქესი, ასაკი, ქორწინებითი მდგომარეობა, რასა/ეთნიკურობა და განათლება. პირველი ცვლადი — სქესი, ორ კატეგორიას მოიცავს: "მამაკაცი" და „ქალი". დააკვირდით დანარჩენი ცვლადების კატეგორიებს.

ცხრილი 10.5 გარდაცვალების ადგილი 1997 წელს სქესის, ასაკის, ქორწინებითი მდგომარეობის, რასის/ეთნიკურობის და განათლების მიხედვით

	გარდაცვალების ადგილი, 1997		
	საავადმყოფო (N=740,405)	სახლი (N=330,447)	მზრუნველობის სახლი (N=331,315)
სქესი			
მამაკაცი	57.3	25.5	17.2
ქალი	48.7	21.8	29.5
ასაკი			
<65	64.4	29.2	6.4
65-74	59.3	28.1	12.6
75-84	52.7	22.7	24.6
85-94	40.8	16.9	42.3
95+	28.0	14.5	57.6
ქორწინებითი მდგომარეობა			
დაუქორწინებელი	55.9	21.5	22.7
დაქორწინებული	59.0	26.9	14.1
ქვრივი	45.2	19.9	34.9
განქორწინებული	54.6	26.1	19.2
უცნობია	57.1	24.9	18.0
რასა/ეთნიკურობა			
თეთრკანიანი	49.7	24.2	26.1
შავკანიანი	66.4	20.2	13.5
ესპანური	65.2	22.7	12.1
სხვა/უცნობია	63.4	21.7	14.9
განათლება (წლები)			
<9	50.8	20.3	29.0
9-11	54.7	23.0	22.3
12	53.5	23.8	22.7
13-15	51.9	26.4	21.7
16+	50.8	27.7	21.6
უცნობია	56.8	18.8	24.4

წყარო: Andrea Gruneir, Vincent Mor, Sherry Weitzen, Rachael Truchil, Joan Teno, and Jason Roy, "Where People Die: A Multilevel Approach to Understanding Influences on Site of Death in America," Medical Care Research Review 64 (2007): 351–378.

3. შეისწავლეთ ცხრილის სტრუქტურა. როგორც ხედავთ, დამოუკიდებელი ცვლადები სტრიქონებშია წარმოდგენილი, ხოლო დამოკიდებული, *გარდაცვალების ადგილი*, სვეტებში. ცხრილი დაყოფილია ხუთ ნაწილად დამოუკიდებელი ცვლადების მიხედვით. რადგან დამოუკიდებელი ცვლადები სტრიქონებშია მო-

თავსებული, თითოეული სტრიქონისთვის ცალკეა გაანგარიშებული ხვედრითი წილები, ხოლო ბაზად აღებულია სტრიქონის ჯამური პროცენტები (არ არის ცხრილში ნაჩვენები). როგორც ვხედავთ, გარდაცვალების ყველაზე დიდი რიცხვი ფიქსირდება საავადმყოფოებში (740,405), სახლებთან (330,447) და მზრუნველობის სახლებთან (331,315) შედარებით. თუ გავაერთიანებთ გარდაცვალების შემთხვევებს საავადმყოფოებსა და მზრუნველობის სახლებში, მივიღებთ, რომ ამერიკაში უფრო მეტი ადამიანი იღუპება ინსტიტუციურ დაწესებულებებში, ვიდრე სახლებში. აქ არ არის გამოთვლილი, მაგრამ სტრიქონების პროცენტების ჯამმა 100% უნდა შეადგინოს.

4. ცხრილის დამოუკიდებელი მონაცემების მიხედვით შეგვიძლია ერთმანეთს შევადაროთ გარდაცვალების ადგილები მაგალითად, სქესის მიხედვით. ორივე სქესის მიხედვით გარდაცვალების დიდი რაოდენობა მოდის საავადმყოფოებზე. პროცენტები თითქმის ერთნაირია ორივე სქესისთვის სახლში გარადაცვალების მიხედვით (25.5% მამაკაცები, 21.8 ქალები), თუმცა სქესთა შორის დიდია განსხვავება მზრუნველობის სახლებში (17.2% მამაკაცები, 29.5% ქალები) და საავადმყოფოებში გარდაცვალების მიხედვით (57.3 მამაკაცები, 48.7% ქალები). საავადმყოფოებში უფრო მეტი მამაკაცი გარდაიცვალა, ხოლო მზრუნველობის სახლებში — უფრო მეტი ქალი.

თქვენ შეგიძლიათ მსგავსი შედარებები გააკეთოთ დანარჩენი ცვლადების მიხედვით.

5. ამ ანალიზის შედეგად, რა დასკვნები შეიძლება გამოიტანოთ გარდაცვალების ადგილების ვარიაციის შესახებ? მეცნიერების აზრით,

მზრუნველობის სახლებში გარდაცვალების სიხშირე იზრდება ასაკის მატებასთან ერთად. ქალებისა და მამაკაცების ყველაზე მეტი ნაწილი (29.5% და 17.2%) იღუპება მზრუნველობის სახლებში, თუმცა სხვა ადგილებში საწინააღმდეგო სურათია. დაქორწინებულების და განქორწინებულების უდიდესი ნაწილი სახლშია გარდაცვლილი (26.9% და 26.1%), მაშინ როდესაც ქვრივების გარდაცვალების სიხშირე უფრო მაღალია მზრუნველობის სახლებში (34.9%). მოხუცებული თეთრკანიანების თითქმის ნახევარია გარდაცვლილი საავადმყოფოებში, ხოლო სხვა რასობრივ/ეთნიკური ჯგუფების მოხუცების — 60%. საავადმყოფოებს გარეთ გარდაცვლილი თეთრკანიანი მოხუცები თანაბრად არიან განაწილებული სახლებსა და მზრუნველობის სახლებში, ხოლო დანარჩენი ჯგუფები უფრო ხშირად სახლებში კვდებიან, ვიდრე მზრუნველობის სახლებში (გვ. 363).

დაასაბუთეთ ცხრილიდან 10.5 მეცნიერების მიერ ჩამოყალიბებული შემდეგი დასკვნები: (1) მზრუნველობის სახლები მოხუცთა შორის გარდაცვალების ყველაზე გავრცელებული ადგილია. (2) მზრუნველობის სახლებში უფრო მეტი ქალი გარდაიცვალა. (3) ქორწინებითი მდგომარეობა სახლში გარდაცვალებას უკავშირდება. (4) მოხუცებული თეთრკანიანების თითქმის ნახევარი საავადმყოფოებშია გარდაცვლილი, ხოლო სხვა რასობრივ/ეთნიკური ჯგუფების მოხუცების — 60%. შეგიძლიათ თუ არა აღნიშნულის ახსნა? რომელი კითხვები დაგეჭმარება ამისთვის?

✓ *ცოდნი შემოწმება*

ბივარიაციული ურთიერთკავშირის მახასიათებლები

ჩვენ უკვე განვიხილეთ ბივარიაციული ურთიერთკავშირის ძირითადი პრინციპები. ახლა განვიხილოთ უფრო სიღრმისეული საკითხები. ამ ნაწილში უფრო დეტალურად განვიხილავთ ბივარიაციული დამოკიდებულებების „მახასიათებლებს“. ეს მახასიათებლები სამი კითხვით შეიძლება წარმოვადგინოთ:[7]

1. აქვს თუ არა ადგილი ურთიერთკავშირს?
2. როგორია მისი სიძლიერე?
3. რა მიმართულება აქვს ამ ურთიერთკავშირს?

ურთიერთკავშირის არსებობა

როგორც ადრე ვნახეთ, ბივარიაციული ცხრილის ანალიზში აუცილებელია ხვედრითი წილების გაანგარიშება და მათი შედარება. ცხრილი 10.6-ის მიხედვით შევისწავლოთ ეკლესიაში სიარულის სიხშირის დამოკიდებულება აბორტის მხარდაჭერასთან. კვლევაში აბორტის მხარდაჭერა გაიზომა კითხვით: “მისაღებია თუ არა თქვენთვის ორსულის მიერ ლეგალური აბორტის გაკეთება, თუ ამის მიზეზი არსებობს?“. ეკლესიაში სიარულის სიხშირე გაიზომა კითხვით: “რამდენად ხშირად ესწრებით რელიგიურ მომსახურებას?“[8]

ცხრილი 10.6 აბორტის მხარდაჭერა ეკლესიაში სიარულის სიხშირის მიხედვით

	ეკლესიაში სიარული			
	არასოდეს	ზოგჯერ	ხშირად	სულ
დიახ	55%	50%	26%	43%
არა	45%	50%	74%	57%
სულ	100%	100%	100%	100%
(*N*)	(111)	(212)	(157)	(480)

წყარო: General Social Survey, 2008.

დაუშვათ, რომ სიცოცხლის შენარჩუნების მომხრე ადამიანები უფრო ხშირად დადიან ეკლესიაში. ჩვენ არ ვამბობთ, რომ ეკლესიაში სიარული აუცილებლად „განაპირობებს“ სიცოცხლის მხარდაჭერას, თუმცა სავარაუდოდ ამ ორს შორის არაპირდაპირი კავშირია. მაგალითად, შეიძლება ვივარაუდოთ, რომ ეკლესიაში მოსიარულეები სხეულის ფორმების შენარჩუნების გამო არ იღებენ გადაწყვეტილებებს აბორტის გასაკეთებლად. (არაპირდაპირი კავშირების წარმოშობაზე მოგვიანებით გვექნება საუბარი ამ თავში).

აღნიშნული ფორმულირებით, ეკლესიაში სიარული „გავლენას ახდენს“ აბორტებზე, ამიტომ ის დამოუკიდებელი ცვლადია, ამიტომ ხვედრითი წილები ეკლესიაში სიარულის სიხშირის ყველა კატეგორიისათვის უნდა გავიანგარიშოთ (ეკლესიაში სიარული სვეტის ცვლადია). ახლა უნდა ვნახოთ, არსებობს თუ არა კავშირი ამ ორ ცვლადს შორის.

ბივარიაციული ცხრილის ორ ცვლადს შორის ურთიერთკავშირი არსებობს იმ შემთხვევაში, თუ დამოუკიდებელი ცვლადის, ამ შემთხვევაში, ეკლესიაში სიარულის სიხშირე, ხვედრითი წილები სხვადასხვა კატეგორიებში ვარიაციას განიცდიან. ჩვენ ადვილად შევამჩნევთ ცვლელებებს აბორტის მხარდაჭერაში ეკლესიაში სიარულის სხვადასხვა დონეების მიხედვით. მათ შორის, ვინც არ დადის ეკლესიაში, 55% არჩევანის მომხრეა; ვინც იშვიათად დადის – 50%, ხოლო ვინც ხშირად დადის – 26%.

ცხრილი გვიჩვენებს, რომ ჩვენი პროგნოზი გამართლდა.

ურთიერთდამოკიდებულების არარსებობის შემთხვევაში ეკლესიაში სიარულის სიხშირის მიუხედავად, ყველა დონეზე ერთნაირი ხვედრითი წილები უნდა მიგვეღო – როგორც არჩევანის მხარდასაჭერად, ისე მის საწინააღმდეგოდ. ცხრილში 10.7 წარმოდგენილია ურთიერთკავშირის არარსებობის ჰიპოთეტური შემთხვევა, როდესაც აბორტებზე შეხედულებასა და ეკლესიაში სიარულს შორის არავითარი კავშირი არ არსებობს. როგორც ვხედავთ, ეკლესიაში სიარულის სიხშირის მიუხედავად, არჩევანის მომხრეთა ხვადრითი წილი უცვლელია და 43%-ს შეადგენს.

ცხრილი 10.7 აბორტის მხარდაჭერა ეკლესიაში სიარულის მიხედვით (ურთიერთკავშირის არარსებობის ჰიპოთეტური ილუსტრაცია)

	ეკლესიაში სიარული			
	არასოდეს	იშვიათად	ხშირად	სულ
დიახ	43%	43%	43%	43%
არა	57%	57%	57%	57%
სულ	100%	100%	100%	100%
(*N*)	(111)	(212)	(157)	(480)

ურთიერთკავშირის სიმჭიდროვე

როდესაც დავადგენთ, რომ ცვლადებს შორის ურთიერთკავშირი არსებობს, დღის წესრიგში დგება ამ ურთიერთკავშირის სიმჭიდროვის განსაზღვრის საკითხი. ყველაზე სწრაფი მეთოდია დამოუკიდებელი ცვლადის სხვადასხვა კატეგორიების ხვედრითი წილების განხილვა. რაც უფრო დიდია განსხვავების ხარისხი კატეგორიებს შორის, მით უფრო მჭიდროა ურთიერთკავშირი. წინა მაგალითში აბორტის მხარდაჭერასთან დაკავშირებით ეკლესიაში სიარულის მიხედვით (ცხრილი 10.7), სვეტებს შორის განსხვავება 0%-ია. მეორე უკიდურესობა იქნებოდა, თუ ყველა, ვინც არ დადის ეკლესიაში, მხარს არჩევანს დაუჭერდა და ვინც ეკლესიაში

დადის, მის წინააღმდეგ იქნებოდა. ამ შემთხვევაში სრულყოფილი ურთიერთკავშირი დაფიქსირდებოდა 100%-იანი განსხვავების დონეზე. თუმცა ურთიერთკავშირების უმეტესობა ამ ორ უკიდურესობას შორის იმყოფება. ფაქტობრივად, იშვიათად ვაკვირდებით 0%-იან ან 100%-იან განსხვავებას. ყველაზე დიდი განსხვავება ცხრილში 10.6 დაფიქსირდა მათ შორის, ვინც არასოდეს დადის ეკლესიაში და ვინც ხშირად დადის (55%-26%=29%). განსხვავება მათ შორის, ვინც იშვიათად დადის და ვინც ხშირად დადის (50%-26%=24%), იმდენად დიდი არ არის, ნაკლებად არსებითია და მიანიშნებს ეკლესიაში სიარულის სიხშირესა და აბორტის მხარდაჭერას შორის ზომიერ კავშირზე.

პროცენტული განსხვავება ორ ცვლადს შორის — ურთიერთკავშირის უხეში საზომია. მოგვიანებით ამავე თავში კავშირის სიმჭიდროვის უფრო სტანდარტიზებულ საზომებზე გვექნება საუბარი.

ურთიერთკავშირის მიმართულება

თუ დამოუკიდებელი და დამოკიდებული ცვლადები ბივარიაციულ ცხრილში ორდინალურ ან ინტერვალურ-ფარდობით დონეზე იზომება, მაშინ მათ შორის ურთიერთკავშირი ან დადებითი იქნება, ან უარყოფითი. **დადებითი** ბივარიაციული ურთიერთკავშირი არსებობს, თუ ცვლადები ერთი მიმართულებით იცვლებიან. ერთი ცვლადის მაღალი მნიშვნელობა „დაწყვილებულია“ მეორე ცვლადის მაღალ მნიშვნელობასთან. **უარყოფითი** ურთიერთკავშირის დროს ცვლადები საპირისპირო მიმართულებით იცვლება: ერთი ცვლადის მაღალი მნიშვნელობა მეორის დაბალ მნიშვნელობასთან არის „დაწყვილებული“ და პირიქით.

დადებითი ურთიერთკავშირი ორდინალურ ან უფრო მაღალ დონეზე გაზომილ ორ ცვლადს შორის ბივარიაციული ურთიერთკავშირი, თუ ცვლადები ერთი მიმართულებით იცვლება.

უარყოფითი ურთიერთკავშირი ორდინალურ ან უფრო მაღალ დონეზე გაზომილ ორ ცვლადს შორის ბივარიაციული ურთიერთკავშირი, თუ ცვლადები საპირისპირო მიმართულებით იცვლება.

ცხრილში 10.8 მოტანილია სოციალური კვლევის საერთაშორისო პროგრამის 2000 წლის კვლევის შედეგები და ავლენს დადებით ურთიერთკავშირს მზადყოფნაში მაღალი გადასახადებისა და მაღალი ფასების გადახდას შორის. დაკვირვებით შეამოწმეთ თითოეული კატეგორია. ნახავთ, რომ იმ რესპონდენტებისთვის, ვინც არ ემხრობა მაღალ ფასებს, უარყოფითი განწყობა მაღალი გადასახადების მიმართ უფრო დამახასიათებელია (91.5%), ვინც ფასების მიმართ ინდიფერენტულია, ასევე ინდიფერენტულია მაღალი გადასახადების მიმართაც (55.1%), ხოლო ვინც ემხრობა მაღალ ფასებს, ის მაღალ გადასახადებსაც ემხრობა (57.8%). ეს

დადებითი ურთიერთკავშირის მაგალითია, როდესაც მაღალი ფასების გადახდა დაკავშირებულია მაღალი გადასახადების გადახდასთან და მაღალი ფასების გადახდისადმი არმიმხრობა დაკავშირებულია მაღალი გადასახადების არგადახდასთან.

ცხრილი 10.8 მაღალი გადასახადების გადახდის მხარდაჭერა მაღალი ფასების გადახდის მიხედვით: დადებითი ურთიერთკავშირი

განწყობა მაღალი გადასახადების მიმართ	განწყობა მაღალი ფასების მიმართ		
	არ ემხრობა	ინდიფერენტულია	ემხრობა
არ ემხრობა	91.5%	36.4%	23.6%
ინდიფერენტულია	5.1%	55.1%	18.6%
ემხრობა	3.4%	8.5%	57.8%
სულ	100%	100%	100%
(*N*)	(529)	(352)	(532)

წყარო: International Social Survey Programme, 2000.

ცხრილში 10.9 ნაჩვენებია უარყოფითი ურთიერთკავშირი განათლების დონესა და რელიგიურ მომსახურებაზე დასწრებას შორის.[9] განათლების არმქონე პირებისთვის დამახასიათებელია რელიგიურ მომსახურებაზე დასწრება თვეში 2-3-ჯერ, ან მეტჯერ (66.2%). საშუალო განათლების მქონეები იშვითად ესწრებიან რელიგიურ მომსახურებას (35%), ხოლო უნივერსიტეტდამთავრებულებისთვის ტიპური პასუხია „არა“ (37.3%). კავშირი აქ უარყოფითია, რადგან განათლების დონის ზრდასთან ერთად მცირდება რელიგიურ მომსახურებაზე დასწრების სიხშირე.

ცხრილი 10.9 რელიგიურ მომსახურებაზე დასწრება განათლების დონის მიხედვით: უარყოფითი ურთიერთდამოკიდებულება

დასწრება რელიგიურ მომსახურებაზე	განათლების დონე		
	არასაკმარი	საშუალო	საუნივერსიტეტო
არასოდეს	5.2%	32.5%	37.3%
იშვიათად	28.6%	35.0%	34.9%
2-3-ჯერ თვეში ან მეტი	66.2%	32.5%	27.8%
სულ	100%	100%	100%
(*N*)	(77)	(237)	(126)

წყარო: International Social Survey Programme, 2000.

ჯანდაცვის ინფორმაციის ეროვნული ტრენდების კვლევამ (HINTS) 2007 წელს გამოკითხა რესპონდენტები კიბოს შესახებ ინფორმაციის მოძიების თაობაზე. ცხრილში 10.10 მოტანილია პასუხები ასაკობრივი ჯგუფების მიხედვით.

ყველა ასაკობრივ ჯგუფში პასუხების უმრავლესობა იყო „არა“, თუმცა, თუ უფრო დეტალურად განვიხილავთ დამოუკიდებელი ცვლადის, ასაკის კატეგორიებს, ვნახავთ, რომ დამოკიდებული ცვლადის, კიბოს შესახებ ინფორმაციის მოძიება, უარის ყველაზე მაღალი დონე დაფიქსირდა დაბალ ასაკობრივ ჯგუფში (62.9%) დანარჩენ ჯგუფებთან შედარებით (59.9% და 51.9 უფროს ასაკებში). ასევე, ინფორმაციის მოძიების ცდაში ყველაზე მაღალი ხვედრითი წილი აქვთ მაღალი ასაკის ადამიანებს (48.1%) ახალგაზრდა ასაკებთან შედარებით (40.7% და 37.1% შესაბამისად). სხვა სიტყვებით, თითოეული ცვლადის დაბალი კატეგორიები დაკავშირებულია ერთმანეთთან და იგივე შეიძლება ითქვას მაღალ კატეგორიებზეც. აქედან გამომდინარე, შეიძლება ვთქვათ, რომ ამ ორ ცვლადს შორის ურთიერთკავშირი დადებითია.

ცხრილი 10.10 ინფორმაციის მოძიება კიბოს შესახებ ასაკის მიხედვით

მოიძია თუ არა ოდესმე ინფორმაცია კიბოს შესახებ	ასაკობრივი ჯგუფი		
	18-34	35-44	45+
არა	62.9%	59.3%	51.9%
დიახ	37.1%	40.7%	48.1%
სულ	100%	100%	100%
(N)	(224)	(194)	(1,068)

წყარო: HINTS, 2007.

ცხრილში 10.11 მონაცემები აგრეთვე მოტანილია HINTS 2007-დან, რომელიც ზომავს ურთიერთკავშირს სტრესის (დეპრესიის) ხარისხსა (0-დან 24 ქულამდე) და ასაკს შორის. ცვლადი DISTRESSSCORE, 5 შემადგენლით არის წარმოდგენილი რესპონდენტის მიმდინარე ფსიქოლოგიური მდგომარეობის დასახასიათებლად. სტრესის ქულები ორ ჯგუფად წარმოვადგინეთ, პირველ ჯგუფში გაერთიანდნენ რესპონდენტები სტრესის ქულებით 1-დან 3-ის ჩათვლით, ხოლო მეორეში, სტრესის მაღალი ქულით – 4. სტრესის მედიანური ქულაა 3, ამიტომ რესპონდენტების ერთი ნახევარი მოხვდა პირველ ჯგუფში, ხოლო მეორე ნახევარი – უმაღლეს ჯგუფში.

ცხრილი 10.11 დეპრესიის ქულა ასაკების მიხედვით

დეპრესიის ქულა	ასაკობრივი ჯგუფი		
	18-34	35-44	45+
1-3	32.0%	35.5%	45.8%
4+	68.0%	64.5%	54.2%
სულ	100%	100%	100%
(N)	(197)	(166)	(825)

წყარო: HINTS, 2007.

ამ მაგალითში დეპრესიის ქულასა და ასაკს შორის უარყოფითი ურთიერთდამოკიდებულებაა. მაღალი ასაკობრივი ჯგუფები დეპრესიის უფრო დაბალ ქულებთან ასოცირდება ახალგაზრდა ასაკებთან შედარებით (45% და 35.5% და 32% შესაბამისად). ასევე, დეპრესიის ხარისხი უფრო მაღალია დაბალ ასაკებში. სხვა სიტყვებით, რაც უფრო იზრდება ასაკი, მით უფრო მცირდება დეპრესიის ქულა, ამიტომ ეს ცვლადები უარყოფით დამოკიდებულებაში არიან.

შემდეგ თავში ვნახავთ ორდინალური და ინტერვალურ-ფარდობითი ცვლადების ურთიერთკავშირის საზომებს და მათ დადებით ან უარყოფით მნიშვნელობებს მათი ურთიერთკავშირის მიმართულების შესაბამისად.

10.9 ცხრილის საფუძველზე რელიგიურ მომსახურებაზე დასწრება გაანაწილეთ 2 კატეგორიაში: არასდროს და ხანდახან. გაიანგარიშეთ ბივარიაციული ცხრილი პროცენტებით. თქვენი შედეგები შეადარეთ ცხრილი 10.9-ის შედეგებს. რა შეგიძლიათ თქვათ განათლების დონესა და რელიგიურ მომსახურებაზე დასწრებას შორის ურთიერთკავშირის ცვლილებაზე?

✓ *ცოდნ შემოწმება*

შემუშავება

წინა განყოფილებებში ჩვენ განვიხილეთ ურთიერთკავშირი ორ ცვლადს (დამოუკიდებელ და დამოკიდებული) შორის, თუმცა ეს მონაცემთა ანალიზის მხოლოდ პირველი ეტაპია. მას შემდეგ, რაც ბივარიაციული ანალიზის საფუძველზე ცვლადებს შორის ურთიერთკავშირს დავადგენთ, ვიწყებთ ამ კავშირების ბუნებისა და ინტერპრეტაციის სხვა შესაძლებლობების ძიებას. ამ ნაწილში ჩვენ განვიხილავთ მეთოდს, რომელსაც *შემუშავება* ეწოდება. **შემუშავება** ეს არის პროცესი, რომელიც ბივარიაციული ანალიზის შემდგომი გაღრმავებისთვის არის შექმნილი და ითვალისწინებს ახალი ცვლადების შემოტანას, რომლებსაც **საკონტროლო ცვლადებს** უწოდებენ. საკონტროლო ცვლადების შემოტანით ჩვენ ვცდილობთ, შევამოწმოთ ამ ცვლადის ეფექტი ბივარიაციულ ურთიერთკავშირზე. თითოეული პოტენციური საკონტროლო ცვლადი განსახილველი ბივარიაციული კავშირის ალტერნატული ახსნის საშუალებაა.

შემუშავება პროცესი, რომელიც ბივარიაციული ანალიზის შემდგომი გაღრმავებისთვის არის შექმნილი და ითვალისწინებს საკონტროლო ცვლადების შემოღებას.

საკონტროლო ცვლადი ბივარიაციულ ანლიზში შემოტანილი დამატებითი ცვლადი. ცვლადის მიერ კონტროლი ხორციელდება მისი ეფექტის გათვალისწინებით ბივარიაციულ ანალიზში.

ახალი ცვლადის შემოტანას ბივარიაციულ ანალიზში სამი ძირითადი მიზანი აქვს:

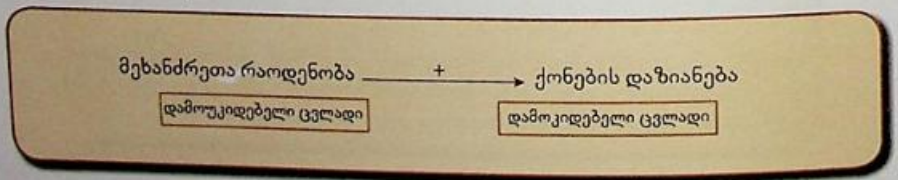
- შემუშავება გვეხმარება არამცდარობის ტესტირებაში. მიზეზ-შედეგობრივი კავშირის შემუშავება მოითხოვს არა მარტო დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის კავშირის ჩვენებას, არამედ მათ შორის დროითი წესრიგის გათვალისწინებასაც, რაც თეორიულად და ემპირიულად დაასაბუთებს, რომ თუ ურთიერთობა არამცდარია, მისი ახსნა სხვა ცვლადებით შეუძლებელია.
- შემუშავება ხსნის ბივარიაციული ურთიერთკავშირის მიზეზ-შედეგობრივ თანმიმდევრობას დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის ჰიპოთეზო ცვლადის შემოტანის გზით.
- შემუშავება განსაზღვრავს სხვადასხვა პირობებს, რომელთა გავლენით შეიძლება ადგილი ჰქონდეს საწყის ურთიერთკავშირს.

წინა განყოფილებებში ჩვენ შევისწავლეთ, თუ როგორ დავადგინოთ ორ ცვლადს შორის კავშირის არსებობა. ახლა გავეცნობით ბივარიაციული ურთიერთკავშირის შემუშავების თეორიულ და სტატისტიკურ მოსაზრებებს. ამისთვის ჩვენ გამოვიყენებთ სამ მაგალითს. პირველი შეეხება არამცდარობის ტესტირებას, მეორე – მიზეზობრივ თანმიმდევრობას, რომელშიც დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის მესამე ცვლადის ჩარევა ხდება და მესამე, რომელიც გვიჩვენებს, თუ როგორ შეუძლია შემუშავებას პირობითი ურთიერთობების ახსნა.

არამცდარობის ტესტირება: მეხანძრეები და ქონების დაზიანება

დავიწყოთ მცდარი ურთიერთობების მაგალითით. მეცნიერებმა დაამტკიცეს ძლიერი ბივარიაციული ურთიერთკავშირი *მეხანძრეების რაოდენობასა* (დამოუკიდებელი ცვლადი) და *დაზიანებული ქონების ოდენობას* შორის (დამოკიდებული ცვლადი). რაც უფრო მეტია მეხანძრეების რაოდენობა, მით უფრო დიდია ზარალი. ამ დამოკიდებულებას შეიძლება მოჰყვეს თქვენი უხერხული დასკვნა (ასახულია ნახაზზე 10.2) იმის თაობაზე, რომ მეხანძრეებმა გამოიწვიეს ზარალი.

ნახაზი 10.2 ბივარიაციული ურთიერთკავშირი მეხანძრეების რაოდენობასა და მიყენებულ ზარალს შორის

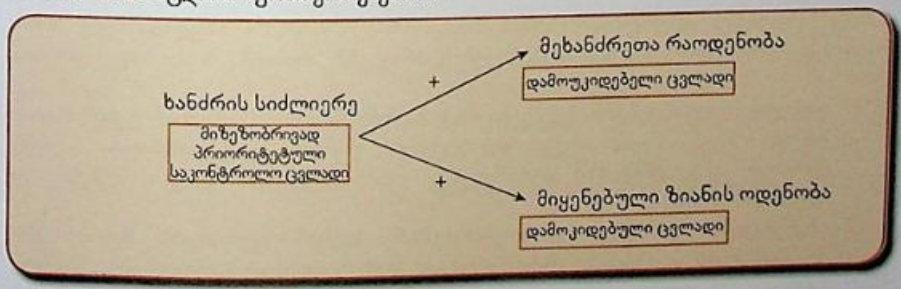


ნახაზზე 10.2 ასახულია, თუ როგორი შეიძლება იყოს *პირდაპირი მიზეზ-შედეგობრივი კავშირი* მეხანძრეების რაოდენობასა და მიყენებულ ზარალს შორის. **პირდაპირი მიზეზობრივი ურთიერთკავშირის** დროს ორ ცვლადს შორის ურთიერთობა არ შეიძლება აიხსნას სხვა თეორიულად მისაღები ცვლადების საშუალებით. აქ კი ცხადია, რომ კავშირი მეხანძრეების რაოდენობასა და მიყენებულ ზარალს შორის უნდა აიხსნას მესამე, მიზეზობრივად პრიორიტეტული ცვლადით – ხანძრის სიძლიერე. როდესაც ძლიერი ხანძარია, მის ჩასაქრობად მეტი მეხანძრეები მუშაობენ და ზარალის ოდენობაც უფრო მეტია. ასევე, როდესაც ხანძარი არ არის ძლიერი, მის ჩასაქრობად ნაკლები რაოდენობის მეხანძრეები მუშაობენ და ზარალიც სავარაუდოდ უფრო ნაკლებია.

პირდაპირი მიზეზობრივი ურთიერთკავშირი ბივარიაციული ურთიერთკავშირი, რომელიც არ შეიძლება აიხსნას სხვა თეორიულად მისაღები ცვლადების საშუალებით.

ალტერნატული ახსნა ნაჩვენებია ნახაზზე 10.3. აღსანიშნავია, რომ ნახაზზე შემოთავაზებული საჰიპოთეზო მიზეზობრივი თანმიმდევრობის მიხედვით, როგორც მეხანძრეთა რაოდენობა, ისე ზარალის ოდენობა, ორივე დაკავშირებულია ცვლადთან – *ხანძრის სიძლიერე*, მაგრამ არა ერთმანეთთან. ხანძრის სიძლიერე აქ *საკონტროლო ცვლადია*, ხოლო ნახაზზე 10.2 ასახული ურთიერთკავშირი მეხანძრეთა რაოდენობასა და მიყენებულ ზარალს შორის *მცდარი* ურთიერთიერთკავშირია. **მცდარი ურთიერთიერთკავშირი** არის ორ ცვლადს შორის ურთიერთიერთკავშირი, როდესაც როგორ დამოუკიდებელი, ისე დამოკიდებული ცვლადები განიცდიან მიზეზობრივად პრიორიტეტული საკონტროლო ცვლადის გავლენას და მათ შორის მიზეზობრივი კავშირი არ არსებობს. ამ შემთხვევაში ბივარიაციული კავშირი დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის შეიძლება აიხსნას მხოლოდ საკონტროლო ცვლადის შემოტანის საშუალებით.

ნახაზი 10.3 მცდარი ურთიერთკავშირი



მცდარი ურთიერთიერთკავშირი ურთიერთიერთკავშირი, როდესაც დამოუკიდებელი და დამოკიდებული ცვლადები განიცდიან მიზეზობრივად პრიორიტეტული საკონტროლო ცვლადის გავლენას და მათ შორის მიზეზობრივი კავშირი არ არსებობს. ამ შემთხვევაში ბივარიაციული კავშირი დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის შეიძლება აიხსნას მხოლოდ საკონტროლო ცვლადის შემოტანის საშუალებით.

მეცნიერებმა შეიმუშავეს წესი, რომელიც დაგვეხმარება ცვლადებს შორის ურთიერთკავშირის გარკვევაში: პირდაპირია (მიზეზობრივი) ეს ურთიერთობა, თუ მცდარი. თუ ორი ან მეტი მიზეზობრივად პრიორიტეტული და თეორიულად მისაღები საკონტროლო ცვლადის შემოტანის შედეგად ბივარიაციული ურთიერთკავშირი ორ ცვლადს შორის დაახლოებით იგივე რჩება, მაშინ თავდაპირველი ბივარიაციული ურთიერთკავშირი პირდაპირია (მიზეზობრივი ურთიერთკავშირი). მეორე მხრივ, თუ თავდაპირველი ბივარიაციული ურთიერთკავშირი მნიშვნელოვნად მცირდება (ან ქრება), მაშინ ის მცდარია.

ვნახოთ ამ წესის გამოყენება მეხანძრეების მაგალითზე. ამისათვის ერთ-ერთი გზაა ხანძრების დაყოფა ძლიერ და მცირე ხანძრებად და ცალ-ცალკე თითოეულ ჯგუფში ბივარიაციული ურთიერთკავშირის ტესტირება. თუ საწყისი ურთიერთკავშირი გაქრება ან მნიშვნელოვნად შემცირდება, მაშინ ადგილი ექნება ნახაზზე 10.3 შემოთავაზებულ ახსნას, ხოლო თუ ურთიერთკავშირი შენარჩუნდება, მაშინ უნდა შევჩერდეთ ნახაზზე 10.2 გამოსახულ ახსნაზე ან ახალი ალტერნატული გზები მოვძებნოთ ამისთვის.

ნახაზზე 10.4 გამოსახულია ბივარიაციული ურთიერთკავშირი მეხანძრეების რაოდენობასა და მიყენებულ ზარალს შორის (10.4ა) და კონტროლის პროცესი *ხანძრის სიძლიერით* (10.4ბ). აღსანიშნავია, რომ კონტროლმა გამოიწვია მეხანძრეების რაოდენობასა და მიყენებულ ზარალს შორის კავშირის ზომის მნიშვნელოვანი შემცირება (40%-იდან 12%-მდე). ეს შედეგი ამტკიცებს, რომ ხანძრის სიძლიერე განსაზღვრავს მეხანძრეების რაოდენობასა და მიყენებულ ზარალს, ამიტომ მეხანძრეების რაოდენობასა და მიყენებულ ზარალს შორის ურთიერთკავშირი მცდარია.

საკონტროლო ცვლადის, *ხანძრის სიძლიერის* შემოტანა თავდაპირველ ბივარიაციულ ურთიერთკავშირში შემუშავების პროცესს ასახავს. მისი საფეხურებია:

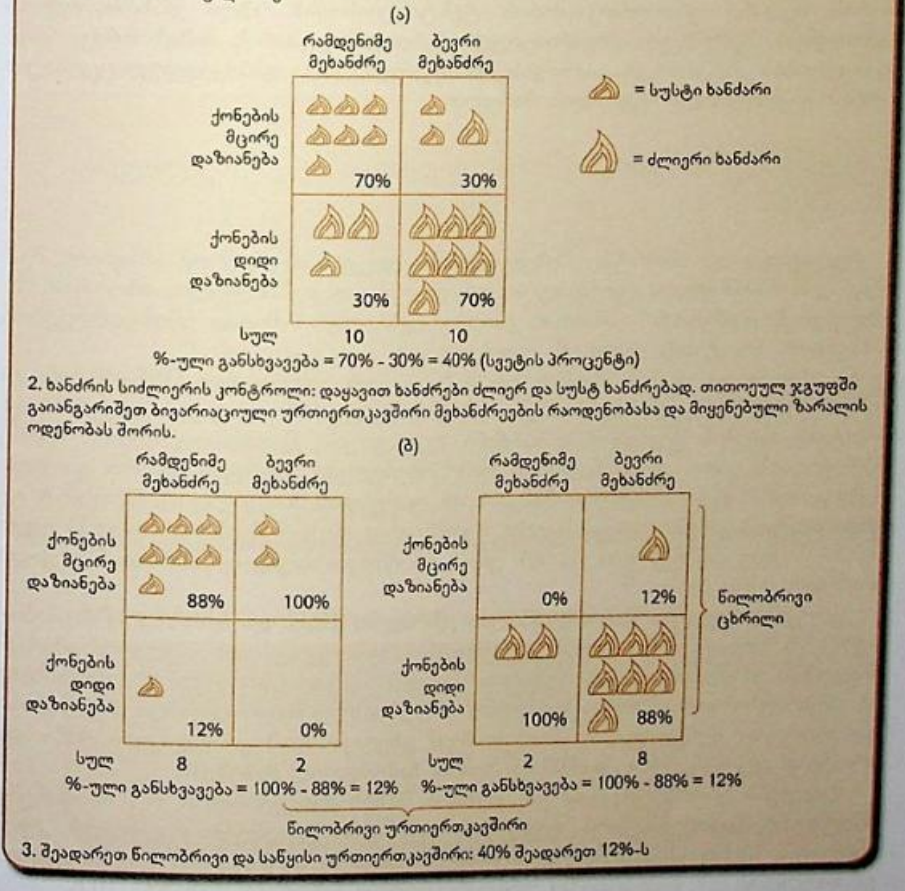
1. დაყავით დაკვირვებები ქვეჯგუფებად საკონტროლო ცვლადით ტესტირებისათვის. ჩვენ იმდენი ქვეჯგუფი გვექნება, რამდენი კატეგორიაცაა საკონტროლო ცვლადში (ამ შემთხვევაში ორი ქვეჯგუფი გვაქვს: სუსტი და ძლიერი ხანძარი).
2. ხელახლა შეამოწმეთ ურთიერთკავშირი საწყის ცვლადებს შორის საკონტროლო ცვლადის ქვეჯგუფებში. დამოუკიდებელ ცხრილებს **წილობრივი ცხრილები** ეწოდება. ისინი ასახავენ **წილობრივ ურთიერთკავშირს**

დამოუკიდებელ (მეხანძრეთა რაოდენობა) და დამოკიდებულ (მიყენებული ზარალი) ცვლადებს შორის საკონტროლო ცვლადის ცალკეულ კატეგორიებში (ძლიერი და სუსტი ხანძარი).

3. შეადარეთ წილობრივი ურთიერთკავშირები და თავდაპირველი ურთიერთკავშირი მთლიან ჯგუფში. პირდაპირი კავშირის დროს წილობრივი და თავდაპირველი ურთიერთკავშირები ერთმანეთთან ძალიან ახლოს იქნება, ხოლო მცდარი კავშირის დროს წილობრივი ურთიერთკავშირი ბევრად სუსტი იქნება თავდაპირველ ურთიერთკავშირთან შედარებით.

ნახაზი 10.4 ბივარიაციული ურთიერთკავშირის შემუშავება

1. ბივარიაციული ურთიერთკავშირი მეხანძრეების რაოდენობასა და მიყენებული ზარალის ოდენობას შორის 20 ხანძრის მაგალითზე



წილობრივი ცხრილები ასახავს წილობრივ ურთიერთკავშირს დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის საკონტროლო ცვლადის საშუალებით.

წილობრივი ურთიერთკავშირი ურთიერთკავშირი დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის, რომელიც ნაჩვენებია წილობრივ ცხრილში.

შემუშავების პროცედურა ჩვენ გამოვიყენეთ მეხანძრეებსა და მიყენებულ ზიანს შორის მცდარი ურთიერთკავშირის ტესტირებისთვის. ახლა ვნახოთ, როგორ გამოიყენება შემუშავება ბივარიაციული ურთიერთკავშირის მიზეზობრივი თანმიმდევრობის ინტერპრეტაციისთვის დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის საკონტროლო ცვლადის ჩარევით.

ჩარეული ურთიერთკავშირი: რელიგია და შეხედულება აბორტზე

რელიგიისა და აბორტზე შეხედულების კვლევამ აჩვენა, რომ არსებობს შემდეგი კანონზომიერება: რელიგიური კუთვნილება დაკავშირებულია აბორტის მხარდაჭერის დონესთან.[10] ამასთან, კათოლიკეები უფრო მეტად ეწინააღმდეგებიან მას, ვიდრე პროტესტანტები ან იუდეველები.[11]

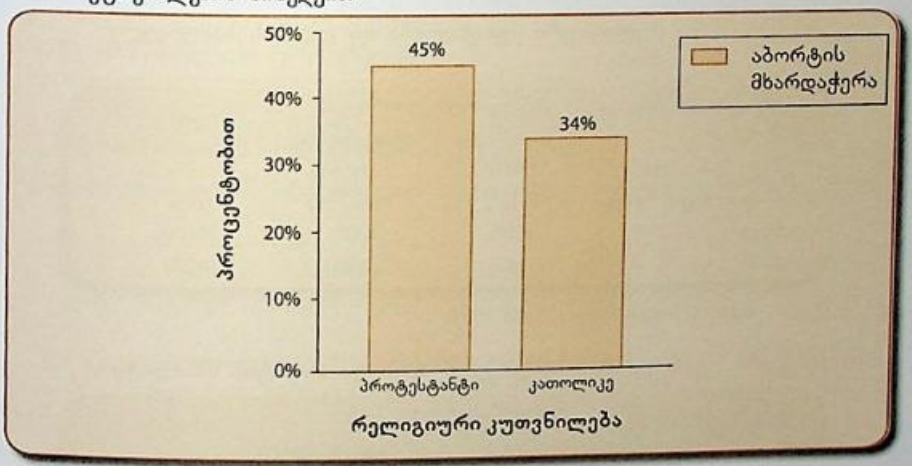
ამ ურთიერთკავშირის შესწავლის მიზნით ჩვენ გამოვიყენეთ GSS-ის მონაცემები 1988-1991 წლებისთვის. ჩვენ შემოვიფარგლეთ მხოლოდ კათოლიკებისა და პროტესტანტების კვლევით დანარჩენი რელიგიური წარმომადგენლების მცირე რაოდენობის გამო. რესპონდენტის დამოკიდებულება აბორტის მიმართ გაიზომა შემდეგ სამ სიტუაციაზე დათანხმების ან უარყოფის საფუძველზე: (1) ქალს არ უნდა ბავშვის გაჩენა, რადგან ოჯახი ხელმოკლეა და ვერ შეინახავს მას, (2) ქალი არ არის დაქორწინებული და არ უნდა ქორწინება ბავშვის მამასთან, (3) ქალს არ სურს მეტი ბავშვი.[12]

კვლევის შედეგები წარმოდგენილია ცხრილში 10.12 და ნახაზზე 10.5. ჰიპოთეზის მიხედვით, რელიგიური კუთვნილება დამოუკიდებელი ცვლადია, ამიტომ ანალიზში ვიყენებთ სვეტის სვედრით წილებს. შედეგში გარკვეულწილად აისახა იმ ჰიპოთეზისადმი მხარდაჭერა, რომ არსებობს ურთიერთკავშირი რელიგიასა და აბორტზე შეხედულებას შორის. ჩვენ ვხედავთ, რომ კათოლიკეთა 45% და პროტესტანტების 34% ეთანხმება ზემოთ ჩამოთვლილ მიზეზებს.

შეიძლება ვიფიქროთ, რომ ეს კავშირი მიზეზ-შედეგობრივია. თუ ამ ინტერპრეტაციას მივიღებთ, სხვა ფაქტორებისგან დამოუკიდებლად, რელიგიურ კუთვნილებას სხვადასხვა შეხედულებებამდე მივყავართ აბორტებთან დაკავშირებით. გრაფიკულად ეს საჰიპოთეზო ურთიერთკავშირი ნაჩვენებია ნახაზზე 10.6.

სხვადასხვა კვლევები ეჭვის ქვეშ აყენებს მოსაზრებას, რომ რელიგიურ კუთვნილებასა და აბორტის მხარდაჭერას შორის მიზეზ-შედეგობრივი კავშირი არსებობს (როგორც შემოთავაზებულია ნახაზზე 10.6). ამ ტიპის კვლევით ლიტერატურაში არსებობს აზრი, რომ კათოლიკებსა და პროტესტანტებს შორის გარკვეული განსხვავებები შეიძლება აიხსნას *ოჯახის სასურველი ზომით.*[13] დამტკიცებულია, რომ რელიგია მუდმივად დაკავშირებულია ოჯახის სასურველ ზომასთან: კათოლიკები უფრო დიდი ზომის ოჯახებს ამჯობინებენ არა-კათოლიკებთან შედარებით. ანალოგიურად, თუ აბორტს განვიხილავთ ოჯახის ზომის გაკონტროლების ალტერნატულ საშუალებად, მაშინ მისი მხარდაჭერა ასევე შეიძლება განვიხილოთ ოჯახის სასურველ ზომასთან კავშირში. ამგვარად, ოჯახის სასურველი ზომა გვევლინება ჩარევის საშუალებად ურთიერთკავშირში რელიგიასა და აბორტებზე შეხედულებებს შორის.

ნახაზი 10.5 აბორტის მხარდამჭერთა ხვედრითი წილები რელიგიური კუთვნილების მიხედვით



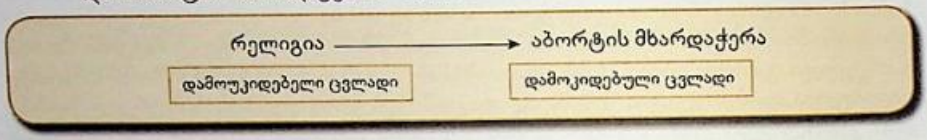
წყარო: General Social Survey, 1988-1991.

ცხრილი 10.12 რელიგიური კუთვნილება და აბორტის მხარდაჭერა

მხარდაჭერა	რელიგიური კუთვნილება		სულ
	კათოლიკე	პროტესტანტი	
დიახ	34%(56)	45%(109)	41%
არა	66%(107)	55%(131)	59%
სულ	100%	100%	100%
(*N*)	(163)	(240)	(403)

წყარო: General Social Survey, 1988-1991.

ნახაზი 10.6 ბივარიაციული ურთიერთკავშირი რელიგიურ კუთვნილებასა და აბორტის მხარდაჭერას შორის



ამ მოსაზრებების ტესტირების მიზნით ჩვენ გავაანალიზეთ ურთიერთკავშირი ოჯახის სასურველ ზომასა და რელიგიას შორის (ცხრილი 10.13) და ოჯახის სასურველ ზომასა და აბორტის მხარდაჭერას შორის (ცხრილი 10.14).[14] აღსანიშნავია, რომ, რადგან თეორიულად ოჯახის სასურველი ზომა გვევლინება ჩარევის მექანიზმად რელიგიასა და აბორტებზე შეხედულებებს შორის ურთიერთკავშირში, ის ცხრილში 10.13 გაანალიზდება როგორც დამოკიდებული ცვლადი, ხოლო ცხრილში 10.14 – როგორც დამოუკიდებელი ცვლადი.

ცხრილი 10.13 რელიგიური კუთვნილება და ოჯახის სასურველი ზომა

ოჯახის სასურველი ზომა	რელიგიური კუთვნილება		სულ
	კათოლიკე	პროტესტანტი	
2-ზე მეტი შვილი	52%(85)	27%(65)	37%
2-ზე ნაკლები შვილი	48%(78)	73%(175)	63%
სულ	100%	100%	100%
(*N*)	(163)	(240)	(403)

წყარო: General Social Survey, 1988-1991.

ცხრილი 10.14 ოჯახის სასურველი ზომა და აბორტის მხარდაჭერა

მხარდაჭერა	ოჯახის სასურველი ზომა		სულ
	2-ზე მეტი შვილი	2 ან ნაკლები შვილი	
დიახ	25%(38)	50%(127)	41%
არა	75%(112)	50%(126)	59%
სულ	100%	100%	100%
(*N*)	(150)	(253)	(403)

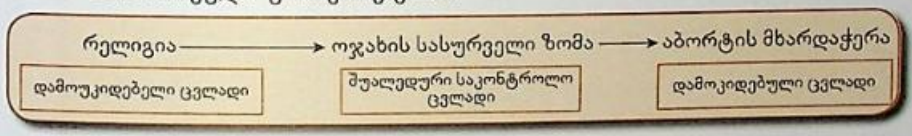
წყარო: General Social Survey, 1988-1991.

ცხრილებში 10.13 და 10.14 მოტანილი მონაცემები ადასტურებს კავშირებს ერთი მხრივ, ოჯახის სასურველ ზომასა და რელიგიას და მეორე მხრივ, ოჯახის სასურველ ზომასა და აბორტის მხარდაჭერას შორის: პირველი, უფრო მეტი კათოლიკე (52%), ვიდრე პროტესტანტი (27%) უჭერს მხარს დიდ ოჯახს (ცხრილი 10.13). მეორე, უფრო მეტი რესპონდენტი, ვინც მცირე ოჯახს ემხრობა,

მხარს უჭერს ქალების უფლებას აბორტზე (50%), იმათთან შედარებით, ვინც დიდ ოჯახს ამჯობინებს (25%) (ცხრილი 10.14). ამ ინტერპრეტაციის მიხედვით ოჯახის სასურველი ზომა არამარტო დაკავშირებულია რელიგიასა და აბორტის მხარდაჭერასთან, არამედ მათ შორის შუალედური რგოლის ფუნქციასაც ასრულებს. აქედან შეიძლება ვივარაუდოთ, რომ რელიგიასა და აბორტზე შეხედულებებს შორის ურთიერთკავშირი *არაპირდაპირია* და *გაშუალებულია* საკონტროლო ცვლადით – ოჯახის სასურველი ზომა.

საკონტროლო ცვლადს ამ ურთიერთკავშირში *ჩარევის ცვლადი* ეწოდება. **ჩარევის ცვლადი** ეწოდება საკონტროლო ცვლადს, რომელიც მიზეზობრივ თანმიმდევრობაში უკან მოსდევს დამოუკიდებელ ცვლადს, თუმცა წინ უსწრებს დამოკიდებულ ცვლადს. რადგან ოჯახის სასურველი ზომა უკან მოსდევს დამოუკიდებელ ცვლადს, რელიგია, მაგრამ დამოკიდებული ცვლადის, აბორტებზე შეხედულების წინაა, ის მიიჩნევა ჩარეულ ცვლადად. ნახაზზე 10.7 გამოსახული ურთიერთკავშირი რელიგიასა და აბორტის მხარდაჭერას შორის **ჩარეული ურთიერთკავშირია**. ამ კავშირის დროს ორ ცვლადს შორის ჩარეულია საკონტროლო ცვლადი.

ნახაზი 10.7 ჩარეული ურთიერთკავშირი



ჩარევის ცვლადი საკონტროლო ცვლადი, რომელიც უკან მოსდევს დამოუკიდებელ ცვლადს, თუმცა წინ უსწრებს დამოკიდებულ ცვლადს.

ჩარეული ურთიერთკავშირი ურთიერთკავშირი, რომლის დროს დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის საკონტროლო ცვლადია ჩარეული.

ჩვენ შეგვიძლია ნახაზზე 10.7 გამოსახული მოდელის ტესტირება, თუ დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის შევიტანთ საკონტროლო ცვლადს, ოჯახის სასურველი ზომა. ამისთვის უნდა დავაცალკეოთ ის რესპონდენტები, რომლებიც ამჯობინებენ დიდ ოჯახებს იმათგან, ვინც ამჯობინებს მცირე ოჯახებს. თუ ჰიპოტეზა, რომელიც დაყენებულია ნახაზზე 10.7 სწორია, მაშინ ურთიერთკავშირი რელიგიასა და აბორტის მხარდაჭერას შორის ან უნდა გაქრეს, ან მნიშვნელოვნად შემცირდეს.

ამ მოსაზრების თვალსაჩინოებისთვის წარმოდგენილია ცხრილი 10.15 და ნახაზი 10.8. წილობრივ ცხრილებში ურთიერთკავშირი რელიგიასა და აბორტის მხარდაჭერას შორის შემცირებულია თავდაპირველ ბივარიაციულ ცხრილთან შედარებით (ცხრილი 10.12). დიდი ოჯახების მხარდამჭერ რესპონდენტებს შორის

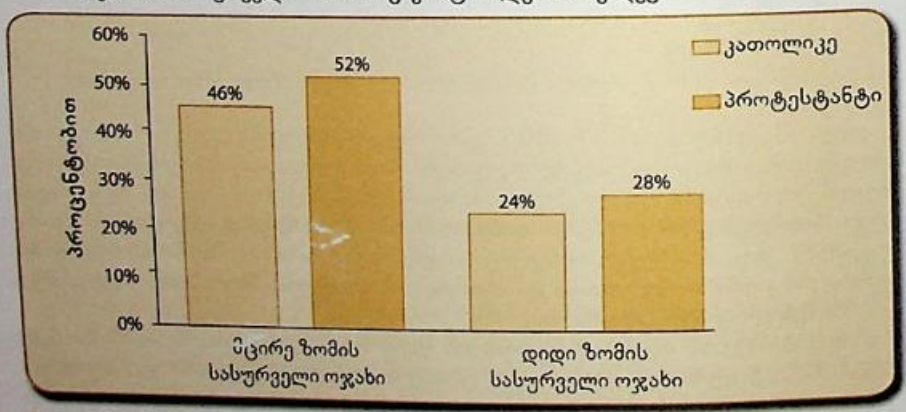
კათოლიკებსა და პროტესტანტებს შორის მცირე სხვაობაა აბორტთან დაკავშირებით. პროტესტანტების 52% და კათოლიკების 46% მხარს უჭერს ლეგალურ აბორტს. აქედან ჩვენ შეიძლება დავასკვნათ, რომ კათოლიკები უფრო ნაკლებად უჭერენ მხარს აბორტს (ვიდრე პროტესტანტები), რადგან ისინი დიდ ოჯახებს ამჯობინებენ. ეს მიგნებები აფართოებს ჩვენ ცოდნას რელიგიურ კუთვნილებასა და აბორტის მხარდაჭერას შორის არსებული თავდაპირველი ბივარიაციულ ურთიერთკავშირის შესახებ.

ცხრილი 10.15 რელიგიური კუთვნილება და აბორტის მხარდაჭერა ოჯახის სასურველი ზომით გაკონტროლების შემდეგ

მხარდაჭერა	რელიგიური კუთვნილება		სულ
	კათოლიკე	პროტესტანტი	
ოჯახის სასურველი ზომა: 2 ან ნაკლები შვილი			
დიახ	46%(36)	52%(91)	50%
არა	54%(42)	48%(84)	50%
სულ	100%	100%	100%
(*N*)	(78)	(175)	(253)
ოჯახის სასურველი ზომა: 2-ზე მეტი შვილი			
დიახ	24%(20)	28%(18)	25%
არა	76%(65)	72%(47)	75%
სულ	100%	100%	100%
(*N*)	(85)	(65)	(150)

წყარო: General Social Survey, 1988-1991.

ნახაზი 10.8 რელიგიის მიხედვით აბორტის მხარდაჭერის ხვედრითი წილები ოჯახის სასურველი ზომით გაკონტროლების შემდეგ



წყარო: General Social Survey, 1988-1991.

როგორც თქვენ უკვე ნახეთ, მცდარი ურთიერთკავშირისა და შუალედური ურთიერთკავშირის ტესტები იდენტურია: ორივე შემთხვევაში უნდა დაკმაყოფილდეს პირობა იმის შესახებ, რომ წილობრივი ურთიერთკავშირები უნდა გაქრეს, ან მნიშვნელოვნად შემცირდეს! როგორ განასხვავოთ ისინი ერთმანეთისგან? განსხვავება კეთდება უფრო თეორიულ საფუძველზე, ვიდრე ემპირიულად. როდესაც ურთიერთკავშირი მცდარია, დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის მიზეზობრივი კავშირი არ არსებობს: ორივე მათგანი განიცდის მიზეზობრივად პრიორიტეტული საკონტროლო ცვლადის გავლენას. შუალედური ურთიერთკავშირის დროს დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის არაპირდაპირი მიზეზობრივი კავშირია; საკონტროლო ცვლადი უკან მოსდევს დამოუკიდებელ ცვლადს და წინ უსწრებს დამოკიდებულ ცვლადს.

✓ ცოდნ
შემოწმე

პირობითი ურთიერთკავშირები: უფრო მეტი აბორტის შესახებ

კიდევ რომელი სხვა ცვლადით შეიძლება რელიგიასა და აბორტს შორის ურთიერთკავშირის ახსნა? ერთ-ერთი შესაძლებელი ვარიანტია *რელიგიური მონაწილეობა*. აბორტთან დაკავშირებულ კვლევაში მეცნიერებმა[15], რომლებმაც საკონტროლო მიზნით გამოიყენეს ცვლადი — რელიგიური მონაწილეობა, აღმოაჩინეს, რომ დიდი განსხვავება არ არის კათოლიკებისა და პროტესტანტების შეხედულებებს შორის აბორტზე, თუ ისინი ეკლესიაში იშვიათად (თვეში ერთხელაც არა) დადიან. თუმცა მათ შორის, რომლებიც ერთზე მეტჯერ დადიან ეკლესიაში, კათოლიკები უფრო ნაკლებად უჭერენ მხარს აბორტს, ვიდრე პროტესტანტები.[16]

რაზე მეტყველებს საზოგადოდ ეს მაგალითები? ისინი განსაზღვრავენ განსხვავებულ პირობებს რელიგიასა და აბორტებს შორის ურთიერთკავშირის განხორციელებისთვის. მაგალითად, მეცნიერთა ერთი ჯგუფი თვლის, რომ კათოლიკებსა და პროტესტანტებს შორის განსხვავებას იწვევს საკონტროლო ცვლადის — *რელიგიური მონაწილეობა* — მხოლოდ ერთი (ეკლესიაში სიარული თვეში ერთზე მეტჯერ) და არა სხვა (ეკლესიაში სიარული თვეში ერთზე ნაკლებად) პირობა. ასევე, შეიძლება განსხვავებული იყოს ურთიერთკავშირი მამაკაცებისა და ქალებისთვის, მოხუცებისა და ახალგაზრდებისთვის. როდესაც ბივარიაციული ურთიერთკავშირი საკონტროლო ცვლადის სხვადასხვა პირობისთვის განსხვავებულია, მიიჩნევენ, რომ ეს არის **პირობითი ურთიერთკავშირი**. პირობითი ურთიერთკავშირის აღწერის მეორე გზაა იმის თქმა, რომ საკონტროლო ცვლადსა და დამოუკიდებელ ცვლადს შორის *სტატისტიკური ურთიერთქმედებაა*.

პირობითი ურთიერთკავშირი ურთიერთკავშირი, რომლის დროს დამოკიდებულ ცვლადზე საკონტროლო ცვლადის გავლენა პირობითია მისი ურთიერთქმედების გამო დამოუკიდებელ ცვლადთან. ურთიერთკავშირი დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის შეიძლება შეიცვალოს საკონტროლო ცვლადის სხვადასხვა პირობების შესაბამისად.

პირობითი ურთიერთკავშირები სოციოლოგიაში ძალიან გავრცელებულია. ერთ-ერთი ასეთი მაგალითია პოზიცია ლეგალური აბორტების მიმართ და აბორტის მორალურობა. მიუხედავად იმისა, რომ თითქმის ყველა რეპონდენტი აბორტს ზნეობრივად არასწორად მიიჩნევს, არჩევანის მომხრე ყველა რესპონდენტს ეს არ მიაჩნია სწორად. ამასთან, არჩევანის მრავალი მომხრე ლეგალურ აბორტს ირჩევს მორალური მოსაზრებების საწინააღმდეგოდ.[17] ბივარიაციული ურთიერთკავშირი აბორტის მორალურობასა და ლეგალური აბორტის მიმართ პოზიციას შორის ასახულია ცხრილში 10.16.

ცხრილი 10.16 აბორტის მორალურობა და პოზიცია ლეგალური აბორტის მიმართ

აბორტის მორალურობა	პოზიცია ლეგალური აბორტების მიმართ		სულ
	არჩევანის მომხრე	სიცოცხლის მომხრე	
ყოველთვის არასწორია	37%	98%	57%
არ არის არასწორი	63%	2%	43%
სულ	100%	100%	100%
(*N*)	(337)	(162)	(499)

წყარო: ადაპტირებულია Jacqueline Scott, "Conflicting Belief About Abortion: Legal Approval and Moral Doubts," Social Psychology Quarterly 52, no. 4 (1989): 319–326-იდან.

რადგან პოზიცია ლეგალური აბორტის მიმართ დამოუკიდებელი ცვლადია, ხვედრითი წილები სვეტებშია გაანგარიშებული. აბორტის მოწინააღმდეგეთა შორის 98% ფიქრობს, რომ აბორტი არ არის მორალური. მათ შორის კი, ვინც ლეგალური აბორტის მომხრეა, შეუსაბამობის დონე შედარებით მაღალია: მთლიანობაში 37% ემხრობა აბორტს, თუმცა თვლის, რომ ეს მორალურად არ არის სწორი, ან არ აქვს ჩამოყალიბებული პოზიცია.[18]

თუმცა ქალებისა და მამაკაცების შეხედულებებში ლეგალური აბორტების მიმართ მცირე განსხვავებაა, ზოგიერთი მეცნიერი თვლის, რომ ქალები უფრო მეტად მიიჩნევენ მას არასწორად. მაგალითად, კეროლ გილიგანი[19] ამტკიცებს, რომ იმ დროს, როდესაც მამაკაცები უფლებებსა და კანონებზე არიან კონცენტრირებული, ქალებს უფრო ურთიერთობების პრობლემები აწუხებთ. ამიტომ, აბორტი ქალისთვის უფრო დიდი მორალური დილემაა, ვიდრე მამაკაცისთვის. იმ ჰიპოთეზის ტესტირებისთვის, რომ მორალური პრინციპების მიუხედავად, ქალები უფრო მეტად ემხრობიან აბორტებს, ვიდრე მამაკაცები, სკოტმა საკონტროლო ცვლადად ჩართო სქესი და თავდაპირველი ურთიერთობა ლეგალური აბორტების მიმართ პოზიციასა და აბორტის ლეგალურობას შორის შეამოწმა ამ ცვლადის საშუალებით, რაც ასახულია ცხრილში 10.17. ცხრილში ნათლად ჩანს გენდერული განსხვავება ამ ურთიერთკავშირში. ამასთან, ცხრილი 10.17-ის საფუძველზე ჩვენ შეიძლება დავასკვნათ, რომ მიუხედავად იმისა, რომ აბორტის მორალურობასა და ლეგალურ აბორტებზე პოზიციას შორის ურთიერთკავშირი არსებობს, ის უფრო

ძლიერია მამაკაცებისთვის (პროცენტული განსხვავება 96%-29%=67%), ვიდრე ქალებისთვის (პროცენტული განსხვავება 100%-46%=54%).

რადგან დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის ურთიერთკავშირი განსხვავებულია თითოეულ წილობრივ ცხრილში, ამიტომ ეს პირობითი ურთიერთკავშირია, რაც იმას ნიშნავს, რომ თავდაპირველი ბივარიაციული ურთიერთკავშირი საკონტროლო ცვლადზეა დამოკიდებული. ჩვენ მაგალითში აბორტის მორალურობასა და ლეგალური აბორტების მიმართ პოზიციას შორის, ურთიერთკავშირის სიძლიერე განპირობებულია სქესით. ეს ურთიერთკავშირი ასახულია ნახაზზე 10.9.

შემუშავების საზღვრები

შემუშვების მაგალითები სოციალური სამყაროს სირთულეში გვარწმუნებს. ეს თავი იმით დავიწყეთ, რომ ჩვენ გარშემო ყველაფერი ურთიერთკავშირშია. თუ უფრო დავაზუსტებთ, შეიძლება ითქვას, რომ ყველაფერი „ჩახლართულია“ და სოციალური მეცნიერების ერთ-ერთი მიზანია ამ „გორგალის გახსნა“. შემუშავება ის პროცედურაა, რომელიც გვეხმარება ბივარიაციული ურთიერთკავშირების „გახსნაში“.

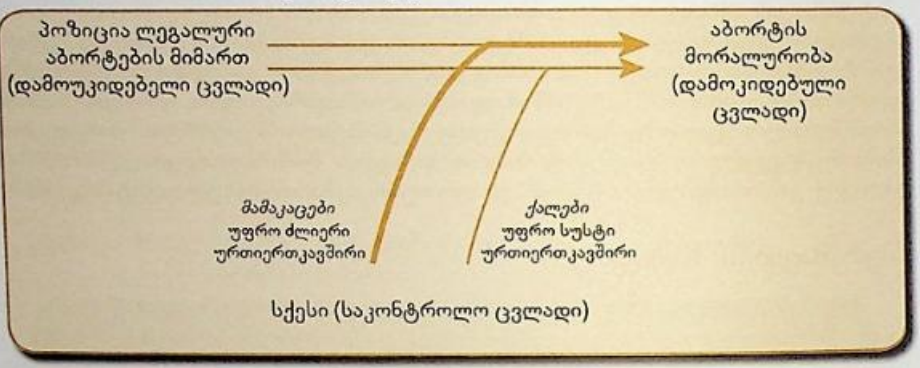
ცხრილი 10.17 ლეგალური აბორტების მიმართ პოზიციასა და აბორტის ლეგალურობას შორის ურთიერთკავშირის კონტროლი სქესის საშუალებით

მამაკაცები				ქალები			
	პოზიცია ლეგალური აბორტების მიმართ		სულ		პოზიცია ლეგალური აბორტების მიმართ		სულ
აბორტის მორალურობა	არჩევანის მომხრე	სიცოცხლის მომხრე		აბორტის მორალურობა	არჩევანის მომხრე	სიცოცხლის მომხრე	
ყოველთვის არასწორია, ან გააჩნია	29%	96%	50%	ყოველთვის არასწორია, ან გააჩნია	46%	100%	64%
არ არის არასწორი	71%	4%	50%	არ არის არასწორი	54%	0%	36%
სულ	100%	100%	100%	სულ	100%	100%	100%
(*N*)	(172)	(78)	(250)	(*N*)	(165)	(84)	(249)

წყარო: ადაპტირებულია Jacqueline Scott, "Conflicting Belief About Abortion: Legal Approval and Moral Doubts," Social Psychology Quarterly 52, no. 4 (1989): 319–326-დან.

ამ ნაწილში წარმოდგენილი ნახაზების საშუალებით ჩვენ ვნახეთ ბივარიაციული ურთიერთკავშირები, რომელთა დაზუსტება და ხელახლა გააზრება საკონტროლო ცვლადის შემოტანის საშუალებით მოხდა. როგორ უნდა გავიგოთ, რომელი ცვლადი უნდა გამოვიყენოთ კონტროლოს მიზნით? ფაქტობრივად, თეორია მნიშვნელოვან მითითებებს გვთავაზობს უთიერთკავშირებზე, რომელიც გვაინტერესებს და იმ ცვლადების სახეობებზე, რომელიც კონტროლისთვის გამოგვადგება. თეორიის გარეშე ჩვენი ცდები სიბნელეში უმისამართო გასროლებს დაემსგავსება,

ნახაზი 10.9 პირობითი ურთიერთკავშირი



და არა ანალიზს. თეორიის გამოყენების შემთხვევაშიც სტატისტიკური ანალიზი ბევრად უფრო რთულია, ვიდრე ამ წიგნშია წარმოდგენილი. როდესაც საკონტროლო ცვლადი შემოვიტანეთ, ურთიერთკავშირის „ნამდვილი“ ბუნება სწრაფად აღვიქვით, თუმცა ეს ყოველთვის ასე არ ხდება. სინამდვილეში, უფრო ხშირად, თეორიასა და ანალიზს შორის საკმაო უფსკრულია. ეს რა თქმა უნდა, იმას არ ნიშნავს, რომ უარი უნდა თქვათ აბურდული გორგალის გახსნაზე. უბრალოდ უნდა გახსოვდეთ თეორიის მნიშვნელობა და სტატისტიკური ანალიზის საზღვრები.

✓ ცოდნის შემოწმება

ამ თავში თქვენ შეგხვდათ ახალი, მნიშვნელოვანი ტერმინები. მონახეთ და ამოიწერეთ მათი განმარტებები: შემუშავება, საკონტროლო ცვლადი, ჩარევის ცვლადი, მიზეზობრივად პრიორიტეტული ცვლადი, მცდარი ურთიერთკავშირი, წილობრივი ურთიერთკავშირი, წილობრივი ცხრილი და პირობითი ურთიერთკავშირი.

სტატისტიკა პრაქტიკაში: ოჯახის მხარდაჭერა საშუალო სკოლის დამთავრების შემდეგ

წინა თავებში ვნახეთ, რომ სტატისტიკა გვეხმარება რასის, კლასის, ასაკის, სქესის გავლენის შესწავლაში ჩვენთვის საინტერესო მოვლენებზე. თუმცა, ჩვენი ყურადღება კონცენტრირებული იყო ამ კატეგორიების დამოუკიდებელ ზეგავლენაზე (ჩვენ ვადარებდით მამაკაცებსა და ქალებს, ახალგაზრდებსა და მოხუცებს, მუშათა კლასს და საშუალო კლასს და ა.შ.). ახლა გვჭირდება ამ კატეგორიების გაერთიანებული ზემოქმედების გააზრება ჩვენი გამოცდილების ფორმირების პროცესზე.[20] თითოეულ ჩვენთაგანს რასის, სოციალური მდგომარეობის, ასაკისა და სქესის საკუთარი კომბინაცია გააჩნია და ამ ლინზის ქვეშ ალიქვამს სამყაროს. ამ ურთიერთგადამკვეთი ფაქტორების ანალიზის

საფუძველზე ჩვენ შეგვიძლია იმის გაგება, თუ როგორ აღიქვამენ რეალობას დანარჩენები.

როდესაც რასას, ასაკს, სქესს, ურთიერთგადამკვეთ სისტემებად მოვიაზრებთ, ვხვდებით, რომ თეთრკანიან და ფერადკანიან ქალებს სქესობრივი თვალსაზრისით საერთო გამოცდილება აქვთ, თუმცა მათი რასობრივი გამოცდილება განსხვავებულია. ასევე, რასობრივი და კლასობრივი განსხვავების საფუძველზე მამაკაცებს განსხვავებული სქესობრივი გამოცდილება გააჩნიათ. მაგალითად, სამრეწველო საწარმოების დახურვამ ბევრი შავკანიანი და ესპანური წარმოშობის მამაკაცი უმუშევრად დატოვა, ხოლო ფერადკანიანმა ქალებმა გააფართოვეს თავიანთი სამუშაო შესაძლებლობები და მომსახურებასა და მაღალ ტექნოლოგიურ სამუშაოებზე დასაქმდნენ.[21]

ბივარიაციული ანალიზის, და კერძოდ, შემუშავების სტატისტიკური მეთოდები განსაკუთრებით მნიშვნელოვანია რასის, კლასის, სქესის გავლენის შესწავლაში ადამიანის სოციალურ ქცევაზე. ცხრილში 10.18 მოტანილია მონაცემები იმის შესაფასებლად, თუ როგორ ეხმარებოდნენ ოჯახები საშუალო სკოლადამთავრებულ ქალებს (რომლებმაც შემდეგ მენეჯერული და პროფესიული პოზიციები დაიკავეს) კოლეჯისკენ გარდამავალ ეტაპზე. ამ მაგალითში ერთდროულად ხდება რასისა და კლასის ზეგავლენის შესწავლა შემუშავების დახმარებით. მოცემული მაგალითი აგრეთვე საშუალებას გვაძლევს ვნახოთ, თუ როგორ რადიკალურად განსხვავებულ შედეგს მივიღებდით ორივე ან ერთ-ერთი ფაქტორის იგნორირების ხარჯზე.[22]

ცხრილი 10.18 მოიცავს 5 ტიპის ოჯახურ მხარდაჭერას, რომლებიც ხუთ დამოკიდებულ ცვლადს წარმოადგენენ: (1) ინფორმაციულს – მისაღები გამოცდებისთვის, (2) ინფორმაციულს – კანდიდატების მიმართ წაყენებული მოთხოვნებისთვის, (3) ფინანსურს – სწავლის და სხვა გადასახადებზე, (4) ემოციურს და (5) კარიერულს. თითოეული დამოკიდებული ცვლადისთვის მხოლოდ ერთი კატეგორიაა მოცემული და ეს კატეგორია არის ქალების ხვედრითი წილი, რომლებმაც ჩამონათვალში არსებული კონკრეტული ტიპის დახმარება მიიღეს. მაგალითად, საშუალო კლასის შავკანიანი ქალების 42%-მა უჩვენა, რომ ოჯახები დაეხმარნენ კოლეჯში მისაღები გამოცდების შესახებ ინფორმაციის მოძიებაში. დანარჩენ 58%-ს (აქ არ არის ნაჩვენები) ასეთი დახმარება არ მიუღია.

რასა და კლასობრივი წარმოშობა ორი დამოუკიდებელი ცვლადია ამ ანალიზში. კლასობრივი ფაქტორის ზეგავლენის შესასწავლად ჩვენ შევადარეთ საშუალო და მუშათა კლასების ოჯახებიდან გამოსული თეთრკანიანი და შავკანიანი ქალები. მონაცემები მეტყველებენ, რომ კლასობრივი განსხვავება საკმაო გავლენას ახდენს ოჯახური დახმარების სახეობაზე. მაგალითად, თუ სწავლის საფასური საშუალო კლასის შავკანიანი ოჯახების 90%-მა დააფინანსა, მუშათა კლასის მხოლოდ 56%-მა გააკეთა ეს. ასეთივე მდგომარეობას ვაკვირდებით თეთრკანიან ოჯახებში. ასევე, თავიანთ გოგონებს ემოციური მხარდაჭერა გაუწიეს როგორც შავკანიანი (86%), ისე თეთრკანიანი (70%) საშუალო კლასის ოჯახებმა.

რასობრივი განსხვავების საფუძველზე შედარებების დროს, განსხვავებები დიდი არ იყო ინფორმაციული და ფინანსური მხარდაჭერის მიხედვით, თუმცა მნიშვნელოვანი იყო ემოციურ მხარდაჭერაში. მაგალითად, საშუალო კლასის

შავკანიანებმა უფრო მეტი ემოციური მხარდაჭერა (86%) გამოიჩინეს შვილებისადმი თეთრკანიანებთან შედარებით (70%), ასევე მოხდა მუშათა კლასის შემთხვევაშიც (64% და 56%).

საშუალო კლასიდან ყველაზე განსხვავებულია შავკანიანთა ჯგუფი, რომელმაც ყველაზე დიდი ემოციური და კარიერული მხარდაჭერა მიიღო მშობლებისგან.

ცხრილი 10.18 პროცენტული განსხვავებები ოჯახიდან მიღებული მხარდაჭერის მიხედვით სკოლიდან კოლეჯისაკენ გარდამავალ ეტაპზე, რასობრივი და კლასობრივი განსხვავებების ფონზე

ოჯახის მხარდაჭერის ტიპი	შავკანიანი		თეთრკანიანი	
	მუშათა კლასი (N = 50)	საშუალო კლასი (N = 50)	მუშათა კლასი (N = 50)	საშუალო კლასი (N = 50)
ინფორმაციული				
მისაღები გამოცდები	22%	42%	24%	40%
მოთხოვნები კანდიდატებისადმი	20%	34%	20%	30%
ფინანსური				
სწავლის გადასახადი	56%	90%	62%	88%
ემოციური				
ემოციური მხარდაჭერა	64%	86%	56%	70%
კარიერული მხარდაჭერა	56%	60%	40%	52%

წყარო: ადაპტირებულია: Lynn Weber, Elizabeth Higginbotham, and Marianne L. A. Leung, "Race and Class Bias in Research on Women: A Methodological Note," Research paper 5, presented at the Center for Research on Women,University of Memphis, 198-დან.

საერთო ჯამში, გამოვლინდა ძლიერი ურთიერთკავშირი რასობრივ წარმოშობასა და ფინანსურ და საინფორმაციო მხარდაჭერას შორის. ამასთან, მონაცემებმა აჩვენა რასობრივი და კლასობრივი წარმოშობის კავშირი ემოციურ მხარდაჭერასთან. ამ კვლევიდან რასობრივი და კლასობრივი წარმოშობა რომ გამოგვეთიშა, სრულიად განსხვავებული წარმოდგენა შეგვექმნებოდა პროფესიონალი და მენეჯერი ქალების ოჯახური მხარდაჭერის შესახებ სკოლიდან კოლეჯში გარდამავალ ეტაპზე.[23]

ეს შეშუშავების გამოყენების კიდევ ერთი მაგალითია. ამ შემთხვევაში კლასობრივი წარმოშობა რასასა და ოჯახურ მხარდაჭერას შორის ურთიერთკავშირის შეშუშავებისთვის საკონტროლო ცვლადის ფუნქციას ასრულებს. ამერიკულ საზოგადოებაში რასა კლასობრივ წარმოშობასთან არის დაკავშირებული (შავკანიანები უმეტესად მუშათა კლასის ოჯახებში იზრდებიან), ხოლო კლასი დაკავშირებულია ოჯახურ მხარდაჭერასთან (მუშათა კლასი ნაკლებად უზრუნველყოფს დახმარებას). ეს ფაქტორები რომ არ მიგველო მხედველობაში, დავასკვნიდით, რომ შავკანიანი ქალები თეთრკანიანებთან შედარებით ყველა პუნქტის მიხედვით ნაკლებ მხარდაჭერას იღებენ ოჯახებიდან, რაც ჩამოაყალიბებდა სტერეოტიპს, რომ შავკანიანები ნაკლებ მზრუნველობას იჩენენ შვილებისადმი,

ვიდრე თეთრკანიანები. ეს დაამახინჯებდა რეალურ სურათს, რადგან მუშათა კლასიდან გამოსული გოგონები, როგორც თეთრკანიანი, ისე შავკანიანები, ნაკლებ დახმარებას იღებენ ოჯახებიდან.[24]

დასასრულ, ეს მაგალითი თვალსაჩინოდ ასახავს რასობრივი და კლასობრივი ნიშნების ერთობლივი ზემოქმედების ეფექტის მნიშვნელობას ქალთა ცხოვრებაში. სტატისტიკურ ანალიზში ეს რასობრივი, კლასობრივი და სქესობრივი შედარებების მხოლოდ ერთი მაგალითია. ასეთი ტიპის კვლევებისათვის საზღვრების დაწესება შეუძლებელია. ამ ცვლადების შემოტანა ართულებს ანალიზს, თუმცა ახალ შესაძლებლობებსაც გვაძლევს ჩვენი გამოცდილების გასამდიდრებლად.

ძირითადი განმარტებები:

• ბივარიაციული ანალიზი სტატისტიკური მეთოდია ორ ცვლადს შორის ურთიერთკავშირის გამოვლენისა და აღწერისათვის. ურთიერთკავშირი სახეზეა, როდესაც ერთი ცვლადის ცალკეული სიდიდეები მეორე ცვლადის სიდიდეებთან არის დაწყვილებული.

• ბივარიაციულ ცხრილში ერთი ცვლადი მეორე ცვლადის კატეგორიების გასწვრივ არის განლაგებული. ის მიიღება დაკვირვებების კლასიფიკაციით ორი ნომინალური ან ორდინალური ცვლადის ქულების შერწყმის საფუძველზე.

• ბივარიაციულ ცხრილებში პროცენტების გაანგარიშება ორ ცვლადს შორის ურთიერთკავშირის შემოწმების მიზნით ხდება. პროცენტები ყოველთვის გაიანგარიშება დამოუკიდებელი ცვლადის თითოეული კატეგორიისათვის.

• ბივარიაციულ ცხრილებში პროცენტების შედარება ხდება დამოუკიდებელი ცვლადის სხვადასხვა კატეგორიის გასწვრივ. ურთიერთკავშირი სახეზეა, თუ დამოუკიდებელი ცვლადის სხვადასხვა კატეგორიის პროცენტული განაწილება ვარიაციას განიცდის.

• ორდინალურ ან ინტერვალურ-ფარდობით დონეზე გაზომილი ცვლადები დადებითად ან უარყოფითად არიან დაკავშირებული. დადებითი კავშირის დროს ერთი ცვლადის მაღალ მნიშვნელობას მეორე ცვლადის მაღალი მნიშვნელობა შეესაბამება. უარყოფითის დროს კი, ცვლადის მაღალ მნიშვნელობას მეორე ცვლადის დაბალი მნიშვნელობა შეესაბამება.

• შემუშავება ეს არის პროცესი, რომელიც ბივარიაციული ანალიზის შემდგომი გაღრმავებისთვის არის შექმნილი და ითვალისწინებს საკონტროლო ცვლადების შემოტანას დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის კავშირის ინტერპრეტაციის მიზნით.

• მცდარ ურთიერთკავშირებში დამოუკიდებელი და დამოკიდებული ცვლადები განიცდიან მიზეზობრივად პრიორიტეტული ცვლადის გავლენას და მათ შორის მიზეზ-შედეგობრივი კავშირი არ არსებობს.

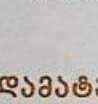
• ჩარეულ ურთიერთკავშირებში საკონტროლო ცვლადი უკან მოსდევს დამოუკიდებელ ცვლადს და წინ უსწრებს დამოკიდებულ ცვლადს.

• პირობითი ურთიერთკავშირის დროს ბივარიაციული ურთიერთკავშირი დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის განსხვავებულია თითოეულ წილობრივ ცხრილში.

საკვანძო ტერმინები

ბივარიაციული ანალიზი
ბივარიაციული ცხრილი
უჯრედი
სვეტის ცვლადი
პირობითი ურთიერთკავშირი
საკონტროლო ცვლადი
კროსტაბულაცია
პირდაპირი მიზეზობრივი კავშირი
შემუშავება
ჩარეული ურთიერთკავშირი
ჩარევის ცვლადი
მარგინალები
უარყოფითი ურთიერთკავშირი
წილობრივი ურთიერთკავშირი
წილობრივი ცხრილები
დადებითი ურთიერთკავშირი
სტრიქონის ცვლადი
მცდარი ურთიერთკავშირი

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

[GSS08PFP-A]

მაგალითი 1: ბივარიაციული ცხრილების შექმნა

SPSS-ს გააჩნია დამოუკიდებელი პროცედურა სპეციალურად კროს-ტაბულაციის ცხრილების საწარმოებლად. მას კროს-ტაბულაციის პროცედურა ეწოდება და შეგიძლიათ ნახოთ Descriptive Statistics ქვემოთ Analyze მენიუში. Crosstabs დიალოგური ყუთი (ნახაზი 10.10) მოგვთხოვს სვეტისა და სტრიქონის ცვლადის განსაზღვრას ცხრილისათვის. შევამოწმოთ ურთიერთკავშირი დაქორწინებული ქალების მიერ, რომელთაც აღარ უნდათ მეტი შვილი, ლეგალური აბორტის მხარდაჭერასა (ABNOMORE) და რელიგიურ კუთვნილებას (RELIG) შორის. ამ მაგალითში RELIG კატეგორიები 5-დან 13-მდე რეკოდირებულია 5 – სხვა სახით.

დათქმის თანახმად, SPSS დათვლის ყველა უჯრედის მონაცემს. უნდა დაკლიკოთ ღილაკი Cells პროცენტების გაანგარიშებისთვის (ნახაზი 10.11). როგორც წესი, ჩვენ ვითხოვთ პროცენტებს იმ ცხრილისთვის, რომელიც განმსაზღვრელი ცვლადისთვის შედგება, ამ შემთხვევაში რელიგიური კუთვნილებისთვის. დამოუკიდებელი ცვლადი მოთავსდება სვეტებში, ხოლო დამოკიდე-

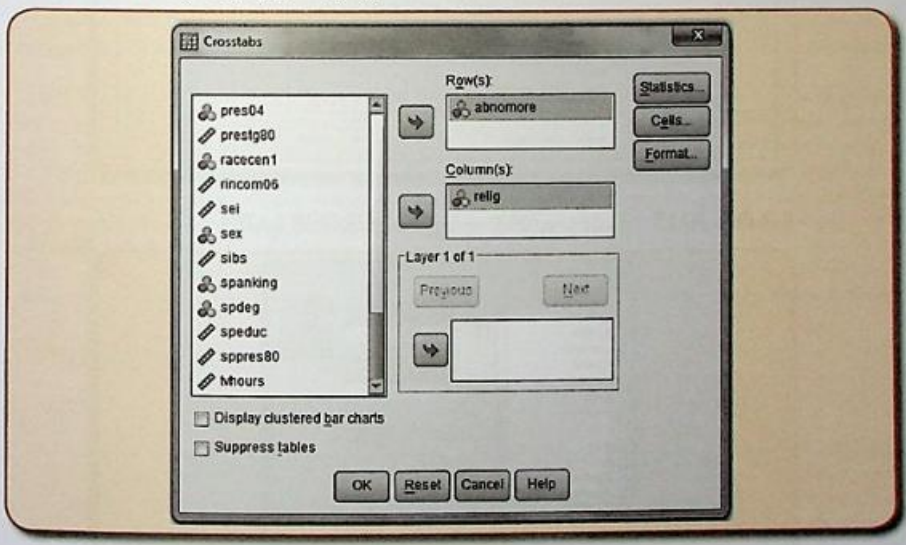
ბული – სტრიქონებში. დაკლიკეთ ჩამრთველი „Column“ სვეტებში რელიგიური პროცენტების გასაანგარიშებლად. დათქმის თანახმად, უკვე არჩეულია „Observed“ სექციაში Counts. (ავრეთვე დათქმის თანახმად შერჩეულია მთელი პროცენტების გაანგარიშების ბრძანება სექციაში Noninteger Weights section, ანუ „Round cell counts“ უკვე შერჩეულია.)

დაკლიკეთ Continue, შემდეგ OK, ცხრილის მისაღებად, რომელიც გამოსახულია ნახაზზე 10.12. SPSS გამოიტანს თითოეული უჯრედის დათვლილ და სვეტის პროცენტულ მნიშვნელობებს. ცხრილის ზედა მარცხენა კუთხეში მითითებულია ჭდეები „Count“ და „% within RELIG“, რის საშუალებითაც SPSS გვახსენებს, თუ რა გაიანგარიშა. ცხრილში ავტომატურად გაანგარიშებულია სვეტებისა და სტრიქონების მიხედვით ჯამები. ცხრილმა დიდი განსხვავებები აჩვენა რელიგიური კუთვნილების მიხედვით ქალების შეხედულებაში აბორტზე. ამ შეხვედრაში კათოლიკებმა და პროტესტანტებმა ძირითადად უარი გამოხატეს, ხოლო სხვა რელიგიების წარმომადგენლებმა მხარი დაუჭირეს აბორტს. ეს შედეგები თქვენთვისაც გასაკვირი იყო, თუ შეესაბამება თქვენს წარმოდგენებს აღნიშნულ საკითხზე?

მაგალითი 2. ცხრილების შექმნა საკონტროლო ცვლადით

საკონტროლო ცვლადის შემოტანა ხდება კროს-ტაბულაციის დიალოგური ყუთის ქვედა ნაწილში (ნახაზი 10.13). (მისი სახელწოდებაა „Layer 1 of 1“, რადგან შესაძლებელია სხვა საკონტროლო ცვლადების შემოტანაც ღილაკზე დაკლიკებით Next button.) ჩვენ ვტოვებთ უცვლელად დამოკიდებულ ცვლადს ABNOMORE, სვეტის ცვლადია RACECEN1, ხოლო საკონტროლო ცვლადი – SEX. უჯრედის ყუთში ალარაფერს არ ვცვლით. ნახაზზე 10.14 მოცემულია აბორტის მხარდაჭერის ბივარიაციული ცხრილები სქესის მიხედვით.

ნახაზი 10.10 დიალოგური ფანჯარა CROSSTABS



ნახაზი 10.11 კროსტაბულაცია: დიალოგური ფანჯარა Cell Display

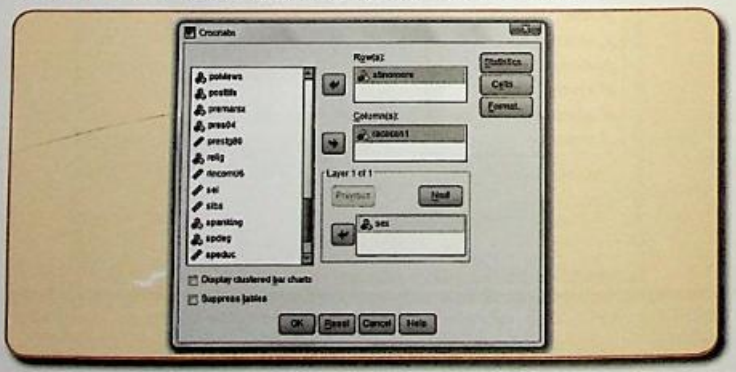


ნახაზი 10.12 საშედეგო ცხრილი Cross-Tabulation Output: ABNOMORE ცვლადის RELIG მიხედვით

MARRIED--WANTS NO MORE CHILDREN * RS RELIGIOUS PREFERENCE Crosstabulation

			RS RELIGIOUS PREFERENCE					
			PROTESTANT	CATHOLIC	JEWISH	NONE	OTHER (SPECIFY)	Total
MARRIED--WANTS NO MORE CHILDREN	YES	Count	191	81	14	109	34	429
		% within RS RELIGIOUS PREFERENCE	38.0%	38.6%	77.8%	67.7%	50.0%	44.7%
	NO	Count	312	129	4	52	34	531
		% within RS RELIGIOUS PREFERENCE	62.0%	61.4%	22.2%	32.3%	50.0%	55.3%
Total		Count	503	210	18	161	68	960
		% within RS RELIGIOUS PREFERENCE	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

ნახაზი 10.13 დიალოგური ფანჯარა CROSSTABS



ნახაზი 10.14 საშედეგო ცხრილი Cross-Tabulation Output: ABNOMORE ცვლადის RACECEN1 მიხედვით საკონტროლო ცვლადისთვის SEX

MARRIED--WANTS NO MORE CHILDREN * WHAT IS RS RACE 1ST MENTION * RESPONDENTS SEX Crosstabulation

RESPONDENTS SEX				WHAT IS RS RACE 1ST MENTION		
				WHITE	BLACK OR AFRICAN AMERICAN	Total
MALE	MARRIED--WANTS NO MORE CHILDREN	YES	Count	171	23	194
			% within WHAT IS RS RACE 1ST MENTION	50.0%	48.9%	49.9%
		NO	Count	171	24	195
			% within WHAT IS RS RACE 1ST MENTION	50.0%	51.1%	50.1%
	Total		Count	342	47	389
			% within WHAT IS RS RACE 1ST MENTION	100.0%	100.0%	100.0%
FEMALE	MARRIED--WANTS NO MORE CHILDREN	YES	Count	166	31	197
			% within WHAT IS RS RACE 1ST MENTION	42.9%	36.0%	41.6%
		NO	Count	221	55	276
			% within WHAT IS RS RACE 1ST MENTION	57.1%	64.0%	58.4%
	Total		Count	387	86	473
			% within WHAT IS RS RACE 1ST MENTION	100.0%	100.0%	100.0%

SPSS ცხრილში გამოიტანს სიხშირეებს ყველა შესასწავლი რასის მიხედვით, თუმცა ჩვენ მხოლოდ ორ მათგანს განვიხილავთ: თეთრკანიანები და შავკანიანები ან აფრო-ამერიკელები. ორივე რასის მამაკაცებში აბორტზე თანხმობა და უარი თითქმის ერთნაირი პროცენტებით ხასიათდება. უარის ყველაზე დიდი პროცენტი შავკანიან მამაკაცებშია (51.1%), თეთრკანიანებთან შედარებით (50%). ქალებში ორივე რასის მიხედვით უარყოფითი დამოკიდებულებაა, თუმცა შავკანიანებში უარის უფრო მაღალი პროცენტია დაფიქსირებული (64% და 57.1%). სწორედ ამ განსხვავებების აღმოჩენას ემსახურება საკონტროლო ცვლადის შემოტანა.

მომდევნო თავებში Statistics ღილაკს Crosstabs dialog box-ში გამოვიყენებთ დამატებითი ცხრილების შესაქმნელად ბივარიაციული ურთიერთკავშირის შემდგომი ინტერპრეტაციის მიზნით.

ᲐᲛᲝᲪᲐᲜᲔᲒᲘ SPSS-ᲘᲡᲗᲕᲘᲡ

[GSS08PFP-A]

1. GSS-ის მონაცემები შეიცავს პასუხებს კითხვაზე რესპონდენტის ბედნიერების შესახებ ზოგადად (HAPPY) და მის კლასობრივ კუთვნილებაზე (CLASS). ააგეთ კროს-ტაბულაციის პროცენტული ცხრილი. (დაკლიკეთ Analyze, Descriptive Statistics, Crosstabs).

ა. მუშათა კლასის რესპონდენტებიდან, რამდენმა პროცენტმა უწვენა, რომ „ძალიან ბედნიერია“?

ბ. დაბალი კლასის რესპონდენტებიდან, რამდენმა პროცენტმა უწვენა, რომ „ძალიან ბედნიერია“?

გ. საშუალო და მაღალი კლასის რესპონდენტებიდან რამდენმა პროცენტმა უწვენა, რომ „საკმაოდ ბედნიერია“?

დ. უმეტესად რომელი ორი კლასიდან იყვნენ რესპონდენტები, ვინც უჩვენა, რომ „ძალიან ბედნიერია“?

ე. არის თუ არა ურთიერთკავშირი კლასსა და ბედნიერებას შორის? თუ არის – აღწერეთ ის. ძლიერია ის თუ სუსტი? (კლასი დამოუკიდებელი ცვლადია).

ვ. დაუბრუნდით ანალიზს და საკონტროლო ცვლადად ჩართეთ RACECEN1. არის თუ არა განსხვავება ურთიერთკავშირში კლასსა და ბედნიერებას შორის რასის მიხედვით?

2. გაანალიზეთ ურთიერთკავშირი ჯანმრთელობის მდგომარეობასა და ბედნიერებას შორის.

ა. გამოიყენეთ SPSS ცხრილის შესაქმნელად ცვლადებისთვის (HEALTH) და (HAPPY). HEALTH დამოუკიდებელი ცვლადია. შემდეგ ეს ურთიერთკავშირი გააკონტროლეთ ცვლადით SEX.

ბ. მთლიანობაში, ქალები უფრო ბედნიერები არიან, თუ მამაკაცები?

გ. ბედნიერების მაღალი დონის ქალებს შორის რამდენი პროცენტი აფიქსირებს ძალიან კარგი ან ცუდი ჯანმრთელობის დონეს? არის თუ არა ურთიერთკავშირი ჯანმრთელობასა და ბედნიერებას შორის? განამტკიცეთ თქვენი დასკვნები ცხრილის მონაცემებით.

3. ადგილობრივ გაზეთში დაიბეჭდა ინფორმაცია იმის თაობაზე, რომ დაბალი განათლების მქონე ადამიანები უფრო მეტად უჭერენ მხარს ბავშვის ფიზიკურ დასჯას. ამ მოსაზრების შესამოწმებლად გამოიყენეთ GSS 2008-ის მონაცემები SPANKING და DEGREE.

ა. SPPS-ის დახმარებით ააგეთ კროსტაბულაციის ცხრილი.

ბ. ააგეთ სვეტებიანი დიაგრამა, რომელიც დაჯგუფებული იქნება DEGREE-ის მიხედვით იგივე ცხრილის გამოსახვისთვის. დაკლიკეთ Graphs, Legacy Dialogs, და Bar. შეარჩიეთ Clustered და შემდეგ Define. ჯუთში სახელწოდებით „Category Axis,“ შეიტანეთ ცვლადი SPANKING, ჯუთში სახელწოდებით Define Clusters By, შეიტანეთ ცვლადი DEGREE. საბოლოოდ, დიალოგური ჯუთის სექციაში „Bars Represent“-ის ქვეშოთ აირჩიეთ % of cases.

გ. არის თუ არა ურთიერთკავშირი ამ ცვლადებს შორის? როგორია მისი მიმართულება?

დ. გაამყარა თუ არა ამ მონაცემებმა გაზეთის შეხედულება?

ე. გამოიყენეთ კროს-ტაბულაციის პროცედურა საკონტროლო ცვლადით SEX და შექმენით დამოუკიდებელი ცხრილები ქალებისთვის და მამაკაცებისთვის.

ვ. არის თუ არა განსხვავება დასჯის მხარდაჭერაში მამაკაცებს შორის განსხვავებული განათლების დონით?

ზ. არის თუ არა განსხვავება დასჯის მხარდაჭერაში ქალებს შორის განსხვავებული განათლების დონით?

4. შეამოწმეთ SEX-ის ურთიერთკავშირი ცვლადებთან ABPOOR (უნდა გაიკეთოს თუ არა აბორტი ქალმა, რომელსაც არ სურს მეტი შვილი?) და ABHLTH (უნდა გაიკეთოს თუ არა ქალმა აბორტი, თუ მის ჯანმრთელობას სერიოზული საფრთხე ემუქრება?) რა დასკვნა გამოიტანეთ?

5. GSS-ის 2008 წლის კვლევაში რესპონდენტებს ჰკითხეს, თუ ვის მისცეს ხმა 2004 წლის საპრეზიდენტო არჩევნებზე (PRES04) და მათ მიერ ბიბლიის აღქმის შესახებ (BIBLE). არის თუ არა ამ ორ ცვლადს შორის ურთიერთკავშირი? მისცემდა თუ არა ხმას ჯონ კერის ის ადამიანი, რომელიც ფიქრობს, რომ ბიბლია „იგავების წიგნია“?

ა. რომელი ცვლადი უნდა მივიჩნიოთ დამოკიდებულად? ახსენით, რატომ.

ბ. SPSS-ში ამ ცვლადების დახმარებით შექმენით კროს-ტაბულაციის ცხრილი. ახსენით ურთიერთდამოკიდებულება ამ ცვლადებს შორის. (დაიმახსოვრეთ, რომ თქვენი დასკვნებიდან უნდა გამოთიშოთ ის ადამიანები, ვისაც ხმა არ მიუცია).

გ. თქვენ თვითონ შეარჩიეთ საკონტროლო ცვლადი და შეამოწმეთ ამ ცვლადებს შორის კავშირი.

თავის სავარჯიშოები

1. ამ სავარჯიშოს ამოსახსნელად გამოიყენეთ ცხრილში მოტანილი მონაცემები რასის, შიშის და ბინის მფლობელობის შესახებ. ცვლადები ზომავს რესპონდენტის რასას, ეშინია თუ არა რესპონდენტს ღამე მარტო სიარულის და მის მიერ სახლის ფლობის სტატუსს.

რესპონდენტი	რასა	ეშინია ღამით მარტო სიარულის	ქირაობს/ფლობს
1	W	N	R
2	B	N	R
3	W	Y	R
4	B	N	R
5	W	N	R
6	B	Y	O
7	W	Y	R
8	W	Y	R
9	W	N	O
10	W	N	O
11	W	Y	R
12	W	N	R
13	B	Y	O
14	W	N	R
15	B	N	O
16	B	N	R
17	W	N	O
18	W	N	O
19	B	N	R
20	W	N	O
21	B	Y	R

წყარო: General Social Survey, 2002

შეიშვნა: რასა: B = შავკანიანი, W = თეთრკანიანი; ეშინია: Y = დიახ, N = არა; ბინის ფლობა: R = ქირაობს, O = ფლობს

ა. ააგეთ ბივარიაციული ცხრილი რასისა და ღამე მარტო სიარულის შიშის მიხედვით. რომელი ცვლადია დამოუკიდებელი?

ბ. დამოუკიდებელ ცვლადზე დაფუძნებული ცხრილისთვის გაიანგარიშეთ პროცენტები. ცხრილის გამოყენებით აღწერეთ ურთიერთკავშირი რასასა და ღამე მარტო სიარულის შიშს შორის. შერჩევის რომელი პუნქტებია აქ გამოყენებული?

გ. გამოიყენეთ მონაცემები ბივარიაციული ცხრილის ასაგებად ბინის მფლობელებსა და მოიჯარეებს შორის ღამე სიარულის შიშის შესადარებლად. ამ მიზნისთვის გამოიყენეთ პროცენტები.

2. განსხვავდება თუ არა მამაკაცებისა და ქალების შეხედულებები პოზიტიურ დისკრიმინაციასთან დაკავშირებით? ცხრილში 10.15 მოტანილია GSS 2008-ის კვლევიდან მონაცემები რესპონდენტების სქესზე (SEX) და შეხედულებებზე საზოგადოებრივად პოზიტიურ დისკრიმინაციასთან დაკავშირებით (DISCAFF: მიადგათ თუ არა ზიანი თეთრკანიანებს პოზიტიური დისკრიმინაციის შედეგად?).

ნახაზი 10.15 კროს-ტაბულაცია: ზიანი პოზიტიური დისკრიმინაციის შედეგად (DISCAFF) სქესის (SEX) მიხედვით

WHITES HURT BY AFF. ACTION * RESPONDENTS SEX Crosstabulation

Count

		RESPONDENTS SEX		
		MALE	FEMALE	Total
WHITES HURT BY AFF. ACTION	VERY LIKELY	66	93	159
	SOMEWHAT LIKELY	193	260	453
	NOT VERY LIKELY	172	179	351
Total		431	532	963

ა. რომელია დამოუკიდებელი ცვლადი?

ბ. რა განსხვავებაა ქალებისა და მამაკაცების შეხედულებებში?

გ. სხვა რომელმა მიზეზებმა შეიძლება იქონიოს გავლენა შეხედულებაზე სოციალური კეთილდღეობის დანახარჯების გაწევის შესახებ? დაასახელეთ მინიმუმ 2 მიზეზი.

3. ჰომოსექსუალების უფლებების დამცველები ამტკიცებენ, რომ ჰომოსექსუალობა არჩევანი კი არა ორიენტაციაა, რის შეცვლაც შეუძლებელია. თქვენი აგვუფელი ამტკიცებს, რომ შეხედულება ჰომოსექსუალიზმზე გავლენას ახდენს პოლიტიკურ შეხედულებაზე. ის, ვინც ფიქრობს, რომ ჰომოსექსუალიზმი არასწორია, უფრო კონსერვატული შეხედულებით გამოირჩევა იმათთან შედარებით, ვინც ასე არ ფიქრობს. გამოიყენეთ ცხრილი და პასუხი გაეცით კითხვებს.

პოლიტიკური შეხედულებები	ჰომოსექსუალური ურთიერთობები		სულ
	ყოველთვის არასწორია	ყოველთვის არასწორი არ არის	
ლიბერალური	61	167	228
	14%	45.8%	28.5%
ზომიერი	164	135	299
	37.7%	37%	37.4%
კონსერვატული	210	63	273
	48.3%	17.3%	34.1%
სულ	435	365	800
	100.0%	100.0%	100.0%

წყარო: General Social Survey, 2008.

ა. რომელია დამოუკიდებელი ცვლადი? რომელია დამოკიდებული ცვლადი?

ბ. რამდენმა პროცენტმა მიუთითა, რომ ჰომოსექსუალური ურთიერთობები ყოველთვის არასწორია?

გ. გამოიყენეთ ცხრილის პროცენტები და აღწერეთ ურთიერთკავშირი ჰომოსექსუალურ ურთიერთობებსა და პოლიტიკურ ორიენტაციას შორის. რამდენად ძლიერია ეს ურთიერთკავშირი?

4. ვაგრძელებთ შეხედულებების განხილვას ჰომოსექსუალურ ურთიერთობებთან დაკავშირებით. დაუშვათ, თქვენი ჯგუფელი ამბობს, რომ ჰომოსექსუალური ურთიერთობები შეიძლება აიხსნას ეკლესიაში სიარულის სიხშირით. ის იყენებს GSS 2008-ის შერჩევას.

ჰომოსექსუალური ურთიერთობები	ეკლესიაში სიარული			სულ
	არასოდეს	წელიწადში რამდენჯერმე	ყოველ კვირა	
ყოველთვის არასწორია	62	40	109	211
ყოველთვის არასწორი არ არის	114	50	35	199
სულ	176	90	144	410

წყარო: General Social Survey, 2008.

ა. რომელია დამოუკიდებელი ცვლადი? რომელია დამოკიდებული ცვლადი?
ბ. გაიანგარიშეთ პროცენტები თითოეული სტრიქონისთვის დამოუკიდებელი ცვლადის ეკლესიაში სიარულის გამოყენებით. არის თუ არა ურთიერთკავშირი ჰომოსექსუალობაზე შეხედულებასა და ეკლესიაში სიარულს შორის? თუ ასეა, რამდენად ძლიერია ის?
გ. დაუშვათ, თქვენ არ ეთანხმებით თქვენი ჯგუფელის პოზიციას, რომ ეკლესიაში სიარული გავლენას ახდენს ხალხის შეხედულებაზე ჰომოსექსუალიზმთან დაკავშირებით და ამტკიცებთ, რომ ეს ერთეულების მოსაზრებაა. რა პრობლემა აქვს თქვენ არგუმენტს?

5. მომავლის მონიტორინგის 2008 წლის კვლევიდან ჩვენ შევამოწმეთ, ჰქონდათ თუ არა თინეიჯერებს გასინჯული ალკოჰოლი. ქვემოთ მოტანილია მონაცემები 585 ბიჭის შესახებ რასის მიხედვით.

ალკოჰოლის მოხმარება	რასა			სულ
	თეთრკანიანი	შავკანიანი	ესპანური წარმოშობის	
არა	90	24	24	138
დიახ	326	53	68	447
სულ	416	77	92	585

წყარო: MTF, 2008.

ა. განსაზღვრეთ თეთრკანიანი, შავკანიანი და ესპანური წარმოშობის თინეიჯერების ხვედრითი წილები, ვისაც გასინჯული აქვს ალკოჰოლი.
ბ. არის თუ არა განსხვავება ამ რასობრივი ჯგუფების მიხედვით?

6. მე-20 საუკუნის განმავლობაში ამერიკელების განათლების დონე მნიშვნელოვნად ამაღლდა. აშშ-ის აღწერის ბიუროს მონაცემები ახასიათებს 25 წლის და უფროსი ასაკის მოსახლეობის განათლების დონეს წლების მიხედვით.

ალკოჰოლის მოხმარება	განათლების დონე	
	საშუალო სკოლა ან მეტი, %	კოლეჯი ან მეტი, %
1980	66.5	16.2
1990	77.6	21.3
1995	81.7	23.0
2000	84.1	25.6
2006	85.5	28.0

წყარო: U.S. Census Bureau, Statistical Abstract of the United States, 2008, Table 217.

ა. რა მიმართულება აქვს წლების ურთიერთკავშირს განათლების დონესთან?

ბ. გამოიყენეთ პროცენტული განსხვავებები ურთიერთკავშირის დასახასიათებლად. რატომ უდრის წლების მიხედვით სტრიქონების პროცენტების ჯამი 100%-ს? არის თუ არა რაიმე პრობლემა ამ ცხრილის ანალიზში?

3. შეიძლება თუ არა ამ ცხრილის გამოყენება იმის დასამტკიცებლად, რომ 2006 წელს განათლების დონე აჭარბებდა 1980 ან 1990 წლების დონეს?

7. აქვთ თუ არა საშუალო სკოლელ უფროსკლასელ გოგონებს და ვაჟებს ერთნაირი მიზნები კოლეჯში შესასვლელად? უფროსკლასელები გამოკითხეს კოლეჯთან დაკავშირებული გეგმებისა და შემდგომი 4-წლიანი განათლების მიღების თაობაზე. მონაცემები მოტანილია სამი წლის მიხედვით:

	1980		1990		2001	
	ქალები (%)	მამაკაცები (%)	ქალები (%)	მამაკაცები (%)	ქალები (%)	მამაკაცები (%)
უფროსკლასელთა გეგმები						
კოლეჯი	33.6	35.6	50.8	45.8	62.4	51.1
აუცილებლად კოლეჯის დამთავრება (4-წლიანი პროგრამა)						
ალბათ დიახ	21.3	23.5	20.5	24.0	20.2	24.8
აუცილებლად/ ალბათ არა	45.0	41.0	28.8	30.2	17.5	24.1

წყარო: U.S. Department of Education, Trends in Educational Equity of Girls and Women, NCES 2005–016. Washington, DC: 2005.

ა. როგორ შეიცვალა 1980 წლიდან გოგონების გეგმები კოლეჯთან დაკავშირებით?

ბ. აქვთ თუ არა ვაჟებს გოგონების მსგავსი გეგმები? რატომ დიახ, ან არა?

8. პირველი სავარჯიშოდან ვნახეთ, რომ თეთრკანიანებს უფრო მეტად ეშინიათ ღამე მარტო სიარულის, ვიდრე შავკანიანებს. თქვენ გაინტერესებთ, თუ რატომ ეშინიათ თეთრკანიანებს, მათ ხომ უსაფრთხო და კეთილმოწყობილ რაიონებში აქვთ საკუთარი სახლები? სხვა სიტყვებით, თქვენ გინდათ სცადოთ შემუშავების გამოყენება.

ა. გამოიყენეთ პირველი სავარჯიშოს მონაცემები რასასა და შიშს შორის ურთიერთკავშირის გასაკონტროლებლად სახლის ფლობის ან ქირაობის ცვლადით.

ბ. დაგეხმარათ, საკონტროლო ცვლადი ურთიერთკავშირის გარკვევაში?

გ. გვიჩვენა თუ არა მფლობელობის ჩართვამ, რომ ურთიერთკავშირი მცდარია, ან რომ სახლის მფლობელობა ჩარევის ცვლადია? ახსენით.

9. სეგრეგაცია სკოლებში იზრდება კლასობრივ (რასობრივ) ნიადაგზე საჯარო და კერძო სკოლებში. ცხრილში მოტანილია თეთრკანიანი და ფერადკანიანი მოსწავლეების განაწილება საჯარო და კერძო დაწყებით სკოლებში სამი წლის მიხედვით. აღწერეთ მოსწავლეების რასობრივი შემადგენლობა და ინსტიტუციის ტიპი. შეიცვალა თუ არა წლების განმავლობაში ეს ურთიერთკავშირი?

ტიპური შვკანიანის კლასელები არიან	დაწყებითი სკოლა (საშუალო)			დაწყებითი სკოლა (ცვრძო)		
	1980	1990	2005	1980	1990	2005
თეთრკანიანი მოსწავლეები	80.9	78.9	75.2	90.7	88.2	83.0
ფერადკანიანი მოსწავლეები	19.1	21.1	24.8	9.3	11.8	17.0

წყარო: U.S. Census Bureau, Statistical Abstract of the United States, 2008, Table 213.

10. მეხუთე სავარჯიშოში დავაკვირდით ურთიერთკავშირს თინეიჯერი ბიჭების რასასა და ალკოჰოლის პირველად გამოყენებას შორის. ქვემოთ მოტანილია ინფორმაცია გოგონების შესახებ.

ალკოჰოლის მოხმარება	რასა				სულ
	თეთრკანიანი	შავკანიანი	ესპანური წარმოშობის		
არა	121	37	22		180
დიახ	315	57	69		441
სულ	436	94	91		621

წყარო: MTF 2008.

11. ორგანიზაცია ლობირებს თქვენ შტატში პორნოგრაფიის აკრძალვას, რადგან მიაჩნია, რომ ეს ზნეობას დაბლა სწევს. ცხრილი შეიცავს მონაცემებს პორნოგრაფიის გამოყენებასა და შეხედულებაზე ქორწინებამდე სექსის შესახებ.

ქორწინებამდე სექსი	X რეიტინგული ფილმის ნახვა გასულ წელს		სულ
	დიახ	არა	
ყოველთვის არასწორია	228	852	1,080
თითქმის ყოველთვის არასწორია	59	111	170
ზანდახან არასწორია	31	38	69
არ არის არასწორი	12	17	29
სულ	330	1,018	1,348

წყარო: General Social Survey, 2008.

ა. ამყარებს თუ არა ორგანიზაციის მოსაზრებებს ქვემოთ მოტანილი ინფორმაცია? რატომ? (ურთიერთკავშირის გასარკვევად გაიანგარიშეთ პროცენტები).

ბ. თქვენი მეგობარი ამტკიცებს, რომ პორნოგრაფიის ეფექტი ქორწინებამდელ სექსზე სქესის მიხედვით განსხვავებულია. გვაძლევს თუ არა აღნიშნულის მტკიცების საფუძველს ქვემოთ მოტანილი ცხრილი? რატომ დიახ, ან არა?

ქორწინებამდე სექსი	X რეიტინგული ფილმის ნახვა გასულ წელს		სულ
	დიახ	არა	
მამაკაცები			
ყოველთვის არასწორია	143	315	458
თითქმის ყოველთვის არასწორია	44	52	96
ხანდახან არასწორია	18	20	38
არ არის არასწორი	7	10	17
სულ	212	397	609
ქალები			
ყოველთვის არასწორია	85	537	622
თითქმის ყოველთვის არასწორია	15	59	74
ხანდახან არასწორია	13	18	31
არ არის არასწორი	5	7	12
სულ	118	621	739

წყარო: General Social Survey, 2008.

3. რა დასკვნა გამოიტანეთ ურთიერთკავშირის შესახებ, საკონტროლო ცვლადის — სქესის შემოტანის შემდეგ?

12. წინა სავარჯიშოში შეისწავლეთ სქესობრივი განსხვავებები შეხედულებებზე პორნოგრაფიასა და ქორწინებამდე სექსს შორის. ახლა ვნახოთ რასობრივი განსხვავებები ამ ურთიერთკავშირში.

ქორწინებამდე სექსი	X რეიტინგული ფილმის ნახვა გასულ წელს		სულ
	დიახ	არა	
თეთრკანიანები			
ყოველთვის არასწორია	158	674	832
თითქმის ყოველთვის არასწორია	41	94	135
ხანდახან არასწორია	21	25	46
არ არის არასწორი	10	14	24
სულ	230	807	1,037
შავკანიანები			
ყოველთვის არასწორია	43	100	143
თითქმის ყოველთვის არასწორია	12	8	20
ხანდახან არასწორია	5	8	13
არ არის არასწორი	2	1	3
სულ	62	117	179

წყარო: General Social Survey, 2008.

ა. აღწერეთ ურთიერთკავშირი პორნოგრაფიულ მასალებსა და შეხედულებას შორის ქორწინებამდე სექსზე თეთრკანიანებისა და შავკანიანებისთვის ცალ-ცალკე.

ბ. როგორი ურთიერთკავშირი აჩვენა საკონტროლო ცვლადის — რასის შემოტანამ: მცდარი, თუ პირობითი? ახსენით.

13. კროს-ტაბულაციის გამოყენების დროს ძალიან მნიშვნელოვანია იმის განსაზღვრა, ცვლადი დამოუკიდებელია, თუ დამოკიდებული. როდესაც ორივე ცვლადი შეხედულებას ან აზრს წარმოადგენს, მნიშვნელოვანია იმ მიზეზის ცოდნა, თუ რატომ თვლით, რომ ცვლადი დამოუკიდებელი ან დამოკიდებულია.

ა. ქვემოთ მოცემულია ორი ცვლადი თავისი კატეგორიებით. იმსჯელეთ იმის თაობაზე, თუ რომელი ცვლადი უნდა იყოს დამოუკიდებელი და რომელი – დამოკიდებული.

ცვლადი 1	*ცვლადი 2*
მეცნიერებას მეტი ზიანი მოაქვს, ვიდრე სიკეთე	*მეცნიერება წყვეტს გარემოს პრობლემებს*
სრულიად ვეთანხმები	სრულიად ვეთანხმები
ვეთანხმები	ვეთანხმები
არც ვეთანხმები, არც ვუარყოფ	არც ვეთანხმები, არც ვუარყოფ
არ ვეთანხმები	არ ვეთანხმები
სრულიად არ ვეთანხმები	სრულიად არ ვეთანხმები

წყარო: International Survey Programme, 2000.

ბ. 13ა-ში მიღებული ცვლადების კლასიფიკაციის შესაბამისად გაიანგარიშეთ ქვემოთ მოტანილ ცხრილში პროცენტები.

გ. რა დასკვნა გამოიტანეთ ამ ცვლადებს შორის ურთიერთკავშირიდან?

მეცნიერება წყვეტს გარემოს პრობლემებს	*მეცნიერებას მეტი ზიანი მოაქვს, ვიდრე სიკეთე*					
	სრულიად ვეთანხმები	*ვეთანხმები*	*არც ვეთანხმები, არც ვუარყოფ*	*არ ვეთანხმები*	*სრულიად არ ვეთანხმები*	*სულ*
სრულიად ვეთანხმები	316	217	162	404	291	1,390
ვეთანხმები	289	1,728	1,088	980	2,585	6,670
არც ვეთანხმები, არც ვუარყოფ	159	877	2,253	718	2,120	6,127
არ ვეთანხმები	230	1,640	1,826	1,306	4,228	9,320
სრულიად არ ვეთანხმები	308	484	668	998	1,160	3,618
სულ	1,392	4,946	5,997	4,406	10,384	27,125

წყარო: International Survey Programme, 2000.

14. 2000 წელს რამდენიმე ევროპული ქვეყნის რესპონდენტს ჰკითხეს, აარიდებდნენ თუ არა თავს ავტომობილის გამოყენებას გარემოს დაცვის მიზნით. პასუხები მოტანილია ქვემოთ ცხრილში.

თავის არიდება ავტომობილისგან გარემოს დაცვის მიზნით	ქვეყანა					
	დიდი ბრიტანეთი	ესპანეთი	ავსტრია	ირლანდია	ჰოლანდია	ნორვეგია
ყოველთვის	33	15	40	19	47	53
ხშირად	96	51	165	71	287	206
ხანდახან	260	122	297	225	541	507
არასოდეს	343	413	268	655	411	515
სულ	732	601	770	970	1,286	1,281

წყარო: International Survey Programme, 2000.

ა. არის თუ არა ურთიერთკავშირი რესპონდენტის ქვეყანასა და თავის არიდებას შორის ავტომობილისგან გარემოს დაცვის მიზნით?

ბ. დაახლოებით ავსტრიელების 27%-მა და ჰოლანდიელების 26%-მა დაფიქსირა, რომ ყოველთვის, ან ხშირად თავს არიდებენ ავტომობილს გარემოს დაცვის მიზნით, მაშინ, როდესაც ირლადიელებისთვის ეს ციფრი 9%-ის ტოლია. სოციოლოგიური თვალსაზრისით, რატომ არის ამ ციფრებს შორის ასეთი დიდი განსხვავება?

თავი 11

ხი-კვადრატ კრიტერიუმი

თავის სწავლების მიზნები

- ჰიპოთეზების შემოწმება ხი-კვადრატ კრიტერიუმის გამოყენებით
- სტატისტიკური დამოუკიდებლობის განმარტება
- ხი-კვადრატ კრიტერიუმის შეზღუდვები-შერჩევის ზომა და სტატისტიკური მნიშვნელოვნება

ამერიკის შეერთებული შტატების განათლების დეპარტამენტის მონაცემების მიხედვით შეერთებულ შტატებში განათლების საერთო დონე იზრდება. როგორც თავდაპირველად მე-3 თავში აღინიშნა, იმ ამერიკელთა პროცენტული წილი, რომლებიც ამთავრებდნენ საშუალო განათლების სწავლების ბოლო ოთხ ან მეტ წელს, 1940 წლის 40%-დან გაიზარდა 86.6%-მდე 2008 წლისათვის. 1940 წელს იმ ამერიკელთა წილი, რომლებიც კოლეჯის განათლებს საფეხურის ოთხ ან მეტ წელს ამთავრებდნენ დაახლოებით 5%-ს შეადგენდნენ 2008 წლის 29.4%-თან შედარებით[1].

მიუხედავად იმ საერთო მიღწევებისა, რაც ზემოთა აბზაცშია მოყვანილი განათლების დონე კვლავ განსხვავდება ისეთი დემოგრაფიული ფაქტორების მიხედვით, როგორიცაა: რასა/ეთნიკურობა, სოციალური ჯგუფი და სქესი. როგორც მე-10 თავის ბოლოს ვნახეთ განათლების არათანაბარი დონე ნაწსენები დემოგრაფიული ფაქტორების მიხედვით განხილულია ლინ ვებერისა და მისი კოლეგების კვლევაში (1987). მკვლევარებმა შეისწავლეს მდედრობითი სქესის სტუდენტებისადმი გამოვლენილი ოჯახური მხარდაჭერის სახეები (ინფორმაციული, ფინანსური და ემოციური) საშუალო განათლებიდან უმაღლეს განათლებაზე გადასვლის პერიოდში. შეკრებილმა მონაცემებმა აჩვენა, რომ არსებობდა მჭიდრო კავშირი სოციალურ ჯგუფებსა, ინფორმაციის ხელმისაწვდომობასა და ფინანსურ

მხარდაჭერას შორის; ასევე გამოვლინდა კავშირი რასას, სოციალურ ჯგუფსა და ემოციურ მხარდაჭერას შორის.

ჩვენი მაგალითი განათლების კვლევაზე განვიხილოთ მხოლოდ იმ სტუდენტებზე კონცენტრირებით, რომლებიც წარმოადგენენ კოლეჯის განათლების პირველ თაობას ე.ი. ის სტუდენტები, რომლთა მშობლებს არასოდეს მიუღიათ კოლეჯის განათლება. ბოლო წლებში, ამერიკის შეერთებულ შტატებში კოლეჯის საწყისი კურსების სტუდენტების საერთო რაოდენობაში პირველი თაობის კოლეჯის სტუდენტების პროპორციული წილი შემცირდა, რაც თავის მხრივ მეტყველებს აშშ-ს განათლების სისტემის მიღწევებზე მთელი მოსახლეობის მასშტაბით. მაგრამ მიუხედავად ამისა, რასობრივი/ეთნიკური სხვაობები პირველი თაობის კოლეჯის სტუდენტებს შორის კვლავ რჩება. მაგალითად, ესპანური წარმოშობის ეთნიკური ჯგუფები სხვა ეთნიკურ ჯგუფებთან შედარებით იკავებენ მაღალ პროცენტულ წილს პირველი თაობის კოლეჯის სტუდენტებს შორის (38,2%) კოლეჯის ოთხწლიან განათლების საფეხურზე.

პირველი თაობის სტუდენტების უმრავლესობა სწავლას იწყებენ ორწლიან ან საზოგადოებრივ კოლეჯებში. ელიოტ ვ. ინმენისა და ლარი მეიესის (1999) მონაცემების მიხედვით საზოგადოებრივი კოლეჯების სტუდენტების უმრავლესობას პირველი თაობის სტუდენტები შეადგენენ, რომელთაც განსხვავებული მიზნების წყება და შეზღუდვები აქვთ. მათი დემოგრაფიული მახასიათებლების შესწავლით შესაძლებელია გამოვლინდეს მათი ჩაბარების, სწავლისა და დამთავრების მიზნობრივი მოტივები. ინმენმა და მეიესმა შეისწავლეს პირველი თაობის სტუდენტების განსხვავებული ცხოვრებისეული გამოცდილება. მკვლევარებმა უპირველესად კოლეჯის სტუდენტებს შორის ცადეს განესაზღვრათ ის ჯგუფები, რომლებიც ყველაზე დიდი ალბათობით იქნებოდნენ პირველი თაობის სტუდენტები.

ინმენისა და მეიესის კვლევის მონაცემები ნაჩვენებია ცხრილ 11.1-ში. ორგანზომილებიანი ცხრილი მოიცავს პირველი თაობის სტუდენტების დაყოფას სქესის მიხედვით. ცხრილის მონაცემების მიხედვით ჩანს, რომ ქალთა პროცენტული წილი, რომლებიც წარმოადგენენ პირველი თაობის კოლეჯის სტუდენტებს აღემატება იგივე მაჩვენებელს მამაკაცი სტუდენტების შემთხვევაში, 46,6 და 35,4 პროცენტები შესაბამისად.

ცხრილი 11.1 ქალი და მამაკაცი სტუდენტების პროცენტული წილები პირველი თაობის კოლეჯის სტუდენტებს შორის

პირველი თაობა	მამაკაცი	ქალი	ჯამი
პირველი თაობის სტუდენტი	35.4% (691)	46.6% (1,245)	41,9% (1,936)
სხვა თაობის სტუდენტი	64.6% (1,259)	53.4% (1,425)	58.1% (2,684)
ჯამი (*N*)	100.0% (1,950)	100.0% (2,670)	100.0% (4,620)

წყარო: ადაპტირებულია-W. Elliot Inman and Larry Mayes, "The Importance of being First: Unique Characsteritics of First Generation Community College Students," Community College Review 26, no. 3 (1999): 8. Copyright © North Carolina State University. Published by Sage Publications

კოლეჯის სტუდენტებს შორის (ცხრილი 11.1) სქესის მიხედვით განსხვავებული მაჩვენებლები მეტყველებს გარკვეული ურთიერთკავშირის არსებობაზე.

ჩვეულებრივ, დასკვნით სტატისტიკაში გამოტანილი დასკვნები გენერალური ერთობლიობის(პოპულაციის) შესახებ ეყრდნობა იმ დასკვნებს, რაც მიღებულ იქნა შერჩევითი ერთობლიობის ანალიზის შედეგად. ჩვენს მაგალითში პრობლემატურია საკითხი იმის შესახებ თუ რამდენად სრულად ასახავს პოპულაციის ტენდენციას მოცემული შერჩევითი ერთობლიობა და მისი სქესის მიხედვით განსხვავებები? არის თუ არა შერჩევითი ერთობლიობის ეს მახასიათებლები შემთხვევით მიღებული? თუ დამატებით შერჩევით ერთობლიობას ავიღებთ და მასზე გამოთვლით იგივე მაჩვენებლებს, ამით არსებული მაჩვენებლების განსხვავებები პოპულაციისა და შერჩევით ერთობლიობას შორის გაიზრდება თუ შემცირდება?

დავუშვათ, რომ პოპულაციიდან აღებულ შერჩევით ერთობლიობაში არსებობს თანაბარი მოსალოდნელობა იმისა, რომ პირველი თაობის სტუდენტი იქნება ქალი ან მამაკაცი. რა იქნება მოსალოდნელი პროცენტული წილები პირველი თაობის ქალი და მამაკაცი სტუდენტებისათვის?

თუ სქესი და პირველი თაობის სტუდენტის სტატუსი ერთმანეთს არ უკავშირდება, მაშინ შესაბამისი პროცენტული წილები ერთმანეთს უნდა უტოლდებოდეს. ასევე, იგივე მიდგომის ფარგლებში პროცენტული წილები ქალი და მამაკაცი სტუდენტების მიხედვით ერთმანეთის ტოლი უნდა იყოს სხვა თაობის სტუდენტებისათვისაც. პროცენტული მაჩვენებლები ერთმანეთის ტოლი უნდა იყოს პირველი და სხვა თაობის რესპონდენტებისათვის, როგორც შერჩევით ერთობლიობაში, ასევე პოპულაციაში(აღნიშნული კატეგორიები გამოიყენეს ინმენმა და მეიესმა თავიანთ კვლევაში). ცხრილ 11.1-ის ბოლო სვეტში ნაჩვენებია ეს პროცენტული წილები: მთელი რესპონდენტების 41.9% იყო პირველი თაობის სტუდენტი, ხოლო 58.1% სხვა თაობის სტუდენტი. აქედან გამომდინარე, თუ არ იარსებებდა რაიმე კავშირი სქესსა და პირველი თაობის კოლეჯის სტუდენტის სტატუსს შორის, ჩვენი მოლოდინი უნდა ყოფილიყო თანაბარი მაჩვენებლების მიღება 41.9% მამაკაცი სტუდენტებისათვის და მისი ტოლი 41.9% ქალი სტუდენტებისათვის. მსგავსი მაჩვენებლის მიღება 58.1% პროცენტის შემთხვევაში შეუძლებელი იქნებოდა.

ცხრილ 11.2-ში ნაჩვენებია ზემოთ მოყვანილი პირობითი პროცენტული მაჩვენებლები. რადგან აქ მოყვანილი პროცენტული წილები ცვლად „პირველი თაობის სტუდენტის სტატუსთან“ მიმართებაში თანაბარია, როგორც ქალი, ასევე მამაკაცი სტუდენტების შემთხვევაში, შეგვიძლია ვთქვათ, რომ ცხრილ 11.2 –ში მოტანილი მაჩვენებლები მეტყველებს სრულყოფილ მოდელზე, რომელშიც არ არსებობს კავშირი ცვლად „სქესსა“ და „პირველი თაობის სტუდენტის სტატუსს“ შორის.

ცხრილი 11.2 პირველი თაობის კოლეჯის სტუდენტები სქესის მიხედვით, პირობითი მონაცემები, ცვლადებს შორის კაშირის არარსებობა

პირველი თაობა	მამაკაცი	ქალი	ჯამი
პირველი თაობის სტუდენტი	41,9%	41,9%	41,9% (1,936)
სხვა თაობის სტუდენტი	58,1%	58.1%	58.1% (2,684)
ჯამი (*N*)	100.0% (1,950)	100.0% (2,670)	100.0% (4,620)

თუ ორ ცვლადს – „სქესსა“ და „პირველის თაობის სტუდენტის სტატუსს“ შორის არსებობს კავშირი, მაშინ დაკვირვების მონაცემები, რომელიც მოყვანილია ცხრილ 11.1-ში უნდა განსხვავდებოდეს ცხრილ 11.2-ში მოყვანილი პირობითი მონაცემებისაგან. მოერეს მხრივ თუ სქესი და პირველი თაობის სტუდენტის სტატუსი ერთმანეთს არ უკავშირდება, დაკვირვების მონაცემები უნდა უახლოვდებოდეს ცხრილ 11.2-ში მოტანილ მოსალოდნელ მნიშვნელობებს. თუ შევადარებთ ცხრილ 11.1-ს და ცხრილ 11.2-ის უჯრებში მოტანილ მნიშვნელობებს დავინახავთ, რომ არსებობს საკმაო სხვაობა დაკვირვებისა და მოსალოდნელ მონაცემებს შორის. მაგალითად, ცხრილ 11.1-ის 1:1 უჯრის მონაცემი (პირველი თაობის მამაკაცი სტუდენტი) განსხვავდება ანალოგიური ცვლადის მნიშვნელობისაგან, რომელიც 11.2 ცხრილშია მოყვანილი. ასევე იგივე განსხვავებებია სხვა დანარჩენი სამი უჯრების მნიშვნელობებში, ცხრილ 11.1-ისა და ცხრილ 11.2 ის მიხედვით.

არის თუ არა დაკვირვებისა და მოსალოდნელ მნიშვნელობებს შორის ზემოთ აღწერილი განსხვავებები საკმარისი საფუძველი იმისათვის, რომ ვამტკიცოთ იგივე მთლიანად პოპულაციისათვის? ამ კითხვაზე პასუხის გასაცემად უნდა გამოვიყენოთ ხი-კვადრატ კრიტერიუმი. იგი მიიღება ორი რიგის მაჩვენებლების შედარებით-დაკვირვების მონაცემების შედარებით, იმ მოსალოდნელ მნიშვნელობებთან, რომლის პირობებშიც ხდება დაშვება, რომ ორ ცვლადს შორის არ არსებობს კავშირი. თუ დაკვირვებისა და მოსალოდნელი მნიშვნელობები ერთმანეთთან ახლოს დგას, მაშინ ხი-კვარდატ კრიტერიუმი იქნება მცირე. თუ ამ უკანასკნელ ორ შორის განსხვავება დიდია, მაშინ ხი-კვადრატ კრიტერიუმის მნიშვნელობა იქნება შესაბამისად მაღალი. ქვემოთ განვიხილავთ, თუ როგორ უნდა გამოთვალოთ ხი-კვადრატ კრიტერიუმი რათა განვსაზღვროთ, არის თუ არა შემთხვევითი ის განსხვავებები, რაც მიღებულ იქნა მამაკაცი და ქალი სტუდენტების შერჩევით ერთობლიობაზე დაკვირვებით.

ხი-კვადრატის კონცეფცია, როგორც სტატისტიკური კრიტერიუმი

ხი-კვადრატ კრიტერიუმი (იწერება როგორც χ^2) არის დასკვნითი სტატისტიკის მეთოდი, რომელიც გამოიყენება ორ ცვლადს შორის არსებული კავშირის მნიშვნელოვნების შესაფასებლად, ორგანზომილებიანი ცხრილის გამოყენებით. ამ კრიტერიუმს სოციალურ კვლევებში აქვს საკმაოდ ფართო გამოყენება. ხი-კვადრატ კრიტერიუმის გამოყენება არ საჭიროებს წინასწარ დაშვებას იმ პოპულაციის მნიშვნელობების განაწილების ფორმის შესახებ, საიდანაც ხდება შერჩევით ერთობლიობაზე დაკვირება. ხი-კვადრატ კრიტერიუმი გამოიყენება, როგორც რაოდენობრივი, ასევე რიგობითი მონაცემების ანალიზის დროს.

ხი-კვადრატ კრიტერიუმი დასკვნითი სტატისტიკის მეთოდი, რომელიც გამოიყენება ორ რაოდენობრივ ან რიგობით ცვლადებს შორის არსებული კავშირის მნიშვნელოვნების შესაფასებლად, ორგანზომილებიანი ცხრილის გამოყენებით.

ხი-კვადრატ კრიტერიუმი ასევე შესაძლებელია გამოყენებულ იქნას ერთი ცვლადის მნიშვნელობების განაწილების მიმართ. ხი-კვადრატ კრიტერიუმი ასევე მოიხსენიება, როგორც თანხმობის ხი კვარდატ კრიტერიუმი. როგორც აღნიშნა ეს კრიტერიუმი გამოიყენება ცვლადის მოსალოდნელი მნიშვნელობების შესადარებლად ცვლადის ფაქტიურ მნიშვნელობებთან.

სტატისტიკური დამოუკიდებლობის კონცეფცია

ორი ცვლადის ერთმანეთისაგან დამოუკიდებლობის შემთხვევაში, როგორც ეს ცხრილ 11.2-ის პირობით მონაცემებშია, მოიხსენიება როგორც სტატისტიკურად დამოუკიდებელი ცვლადები. სხვაგვარად ეს ნიშნავს, რომ პირის მიმართ მიყენებული ერთი კრიტერიუმის მნიშვნელობა დამოუკიდებელია მის მიმართ მიყენებული სხვა კრიტერიუმის მნიშვნელობისაგან. სტატისტიკური დამოუკიდებლობა ორგანზომილებიან ცხრილში განისაზღვრება დამოკიდებული ცვლადის მნიშვნელობის განაწილების შედარებით დამოუკიდებელი ცვლადის მნიშვნელობების განაწილებასთან. როცა ორი ცვლადი სტატისტიკურად დამოუკიდებელია, დამოკიდებული ცვლადის პროცენტული განაწილება თითოეული კატეგორიისათვის დამოუკიდებელი ცვლადის პროცენტული მნიშვნელობების განაწილების მსგავსია. ამ უკანასკნელის ილუსტრირება ცხრილ 11.2-ში მოყვანილი პირობითი მონაცემებია. ცხრილ 11.2-ზე დაყრდნობით შეგვიძლია ვთქვათ, რომ პირველი თაობის კოლეჯის სტუდენტის სტატუსი დამოუკიდებელია პირის სქესისაგან3.

სტატისტიკური დამოუკიდებლობა – კავშირის არარსებობა, ორ ურთიერთთანამკვეთ ცვლადს შორის. დამოკიდებული ცვლადის მნიშვნელობების პროცენტული განაწილება დამოუკიდებელი ცვლადის მნიშვნელობების პროცენტული განაწილების მსგავსია.

ხი კვადრატ კრიტერიუმის გამოსათვლელად გამოყენებული მონაცემები, აღებულია კლაუ ინმენისა და მეიესის კვლევიდან. განვიხილოთ დამოკიდებულება ასაკსა (დამოუკიდებელი ცვლადი) და პირველი თაობის კოლეჯის სტუდენტის სტატუსს(დამოკიდებული ცვლადი) შორის, ორგანზომილებიანი ცხრილის გამოყენებით.

✓ ცოდნის შემოწმება

ასაკი და პირველი თაობის კოლეჯის სტუდენტის სტატუსი

პირველი თაობის სტატუსი	ასაკი წლების მიხედვით		ჯამი
	19-ის ან 19 წლამდე	*20 წლის და 20 წლის ზემოთ*	
პირველი თაობის	*916 (33.7%)*	*1,018 (53.6%)*	*1,934 (41.9%)*
სხვა თაობის	*1,802 (66.3%)*	*881 (46.4%)*	*2,683 (58.1%)*
ჯამი (*N*)	*2,718 (100.0%)*	*1,899 (100.0%)*	*4,617 (100.0%)*

წყარო: ადაპტირებულია W. Elliot Inman and Larry Mayes, "The Importance of being First: Unique Characteristics of First Generation Community College Students," Coomunity College Review 26, no. 3 (1999): 8."

✓ ცოდნის შემოწმება

ამგვარ ორგანზომილებიანი ცხრილი, რომელიც აჩვენებს დამოკიდებულების არარსებობას ასაკისა და პირველი თაობის კოლეჯის სტუდენტის სტატუსს შორის.

ჰიპოთეზის შემოწმების სტრუქტურა ხი-კვადრატ კრიტერიუმის მიხედვით

ხი-კვადრატ კრიტერიუმის პრაქტიკული გამოყენება შედგება იგივე ხუთი საბაზისო საფეხურისაგან, როგორც ეს ამ წიგნის მე-9 თავში სხვა სტატისტიკური კრიტერიუმების შემთხვევებში იყო: 1. დაშვებების გაკეთება; 2. კვლევითი და ნულოვანი ჰიპოთეზების ჩამოყალიბება და ალფას მნიშვნელობის არჩევა; 3. შერჩევის განაწილების არჩევა და სტატისტიკური კრიტერიუმის დაზუსტება; 4. სტატისტიკური კრიტერიუმის გამოთვლა; 5. გადაწყვეტილების მიღება და შედეგების ინტერპრეტირება. სანამ უშუალოდ კონკრეტულ მაგალითზე დაყრდნობით განვიხილავთ ამ ხუთი ეტაპის პრაქტიკულ გამოყენებას, მანამ გამოვყოთ ხი-კვადრატ კრიტერიუმის ცალკეული მახასიათებლები.

დაშვებები

ხი-კვადრატ კრიტერიუმი არ საჭიროებს დაშვებებს იმ პოპულაციის მნიშვნელობების განაწილების ფორმის შესახებ, საიდანაც ხდება შერჩევითი ერთობლიობის მიღება. მაგრამ ამასთანავე, ყველა სხვა დასკვნითი მეთოდების მსგავსად ის საჭიროებს შემთხვევით შერჩევას. ხი-კვადრატ კრიტერიუმი გამოიყენება როგორც რაოდენობრივი, ასევე რიგობითი მონაცემების მიმართ.

კვლევითი და ნულოვანი ჰიპოთეზების ჩამოყალიბება

კვლევითი ჰიპოთეზა (H_1) უშვებს, რომ პოპულაციის ორი ცვლადი ერთმანეთს უკავშირდება.

H_1: პოპულაციის ორი ცვლადი სიდიდე ერთმანეთს უკავშირდება(სქესი და პირველი თაობის კოლეჯის სტუდენტის სტატუსი სტატისტიკურად დამოკიდებული ცვლადებია)

ყველა სხვა სტატისტიკური კრიტერიუმის მსგავსად, ხი-კვადრატ კრიტერიუმი ნულოვანი ჰიპოთეზის შემოწმების კრიტერიუმია. ნულოვანი ჰიპოთეზა (H_0) გულისხმობს ცვლადების სტატისტიკურ დამოუკიდებლობას.

H_0: პოპულაციის ორ ცვლადს შორის არ არსებობს კავშირი. (სქესი და პირველი თაობის კოლეჯის სტუდენტის სტატუსი სტატისტიკურად დამოუკიდებელია).

გამოიყენეთ წინა დავალების მონაცემები და განსაზღვრეთ არის თუ არა ორი ცვლადი "ასაკი" და "პირველი თაობის კოლეჯის სტუდენტის სტატუსი" სტატისტიკურად დამოუკიდებელი ? ჩამოაყალიბეთ კვლევითი და ნულოვანი ჰიპოთეზა თქვენი საკუთარი პრაქტიკული მონაცემებისათვის.

✓ ცოდნის შემოწმება

მოსალოდნელი სიხშირეების კონცეფცია

ნულოვანი ჰიპოთეზის მართებულობის დაშვებით გამოვთვალოთ თითოეული შემთხვევისათვის ცვლადის ის სიხშირეები, რაც უნდა მიგველო ცვლადების სტატისტიკურად დამოუკიდებლობის პირობებში. ეს სიხშირეები იწოდება მოსალოდნელ სიხშირეებად (აღინიშნება, როგორც f_e). ხი-კვადრატ ტესტი ეფუძნება ცხრილის თითოეული უჯრის მიხედვით მოსალოდნელ სიხშირეებსა (f_e) და დაკვირვების ფაქტიურ სიხშირეებს შორის (დაკვირვების სიხშირეები აღნიშნება 0) შედარებებს.

მოსალოდნელი სიხშირეები (f_e) ორგანზომილებიანი ცხრილის უჯრების სიხშირეთა ის მნიშვნელობები, რაც მიიღება იმ დაშვების საფუძველზე, რომ ცხრილში წარმოდგენილი ორი ცვლადი სიდიდე სტატისტიკურად დამოუკიდებელია.

დაკვირვების სიხშირეები (f_0) ორგანზომილებიანი ცხრილის უჯრების ფაქტობრივი დაკვირვების მნიშვნელობები.

მოსალოდნელი სიხშირეების გამოთვლა

f_0-სა f_e-ს შორის განსხვავება განსაზღვრავს ნულოვანი ჰიპოთეზის მიღებას ან უარყოფას. ნულოვანი ჰიპოთეზის სწორად მიჩნევა ნიშნავს, რომ ცვლადი სიდიდეები ერთმანეთისაგან სტატისტიკურად დამოუკიდებელია. თუ სხვაობა f_0-სა f_e-ს შორის დიდია, ეს ნიშნავს, რომ ნაკლებად მოსალოდნელია ცვლადების დამოუკიდებლობა და სავარაუდოდ ნულოვანი ჰიპოთეზა უარყოფილ უნდა იქნას. მეორეს მხრივ, თუ სხვაობა f_0-სა f_e-ს შორის მცირეა, ეს ნიშნავს რომ ცვლადი სიდიდეები ერთმანეთისაგან სტატისტიკურად დამოუკიდებელია და ნულოვანი ჰიპოთეზა სწორია(შესაბამისად არ მოხდება ნულოვანი ჰიპოთეზის უარყოფა).

ხი-კვადრატ კრიტერიუმის გამოყენებისას მნიშვნელოვანი ელემენტია მოსალოდნელი სიხშირეების განსაზღვრა. როგორც ზემოთ აღვნიშნეთ ეს სტატისტი-

კური კრიტერიუმი გამოიყენება შესაბამის ცხრილში ჯვარედინად განლაგებული ცვლადი სიდიდეების მიმართ. რადგან ხი-კვადრატი გამოითვლება ფაქტობრივ სიხშირეებზე დაყრდნობით, ნაცვლად პროცენტული მაჩვენებლებისა უპირველესად საჭიროა ნულოვან ჰიპოთეზაზე დაყრდნობით გამოვთვალოთ მოსალოდნელი სიხშირეები.

პრაქტიკაში, მოსალოდნელი სიხშირეების გამოთვლა უფრო მარტივია ცხრილის სვეტებისა და სტრიქონების მონაცემების მიხედვით, ვიდრე პროცენტული მაჩვენებლების მიხედვით. მოსალოდნელი სიხშირეების გამოთვლა შესაძლებელია შემდეგი ფორმულის გამოყენებით:

$$f_e = \frac{(\textit{სვეტების უჯრათა ჯამი})(\textit{სტრიქონების უჯრათა ჯამი})}{N} \qquad (11.1)$$

იმ ცხრილის ნებისმიერი უჯრის მოსალოდნელი სიხშირეების გამოსათვლელად, რომლისთვისაც გაკეთებულია დაშვება ცვლადი სიდიდეების დამოუკიდებლობის შესახებ, მოცემული უჯრის შემცველი სვეტის ჯამი უნდა გამრავლდეს სტრიქონის ჯამზე და გაიყოს ცხრილის ყველა უჯრის საერთო ჯამზე.

გამოვიყენოთ 11.1 ფორმულა მოსალოდნელი სიხშირეების გამოთვლისათვის ცხრილ 11.1 მონაცემებზე დაყრდნობით. განვიხილოთ პირველი თაობის კოლეჯის სტუდენტის სტატუსის მამაკაცი სტუდენტები(ზედა მარცხენა უჯრა). ამ უჯრისათვის მოსალოდნელი სიხშირე იქნება სვეტებისა (1,950) და სტრიქონების ჯამის (4,620) ნამრავლი გაყოფილი ყველა უჯრის საერთო ჯამზე (4,620):

$$f_e = \frac{1,936 \times 1,930}{4,620} = 817.14$$

შესაბამისი მოსალოდნელი სიხშირე სხვა თაობის კოლეჯის სტატუსის მქონე მამაკაცი სტუდენტებისათვის (ქვედა მარცხენა უჯრა) იქნება:

$$f_e = \frac{2,684 \times 1,950}{4,620} = 1,132.86$$

ასევე, გამოვთვალოთ მოსალოდნელი სიხშირე პირველის თაობის კოლეჯის სტუდენტის სტატუსის მქონე ქალი სტუდენტებისათვის (ზედა მარჯვენა უჯრა):

$$f_e = \frac{1,936 \times 2,670}{4,620} = 1,118.86$$

საბოლოოდ, გამოვთვალოთ მოსალოდნელი სიხშირე სხვა თაობის კოლეჯის სტატუსის მქონე ქალი სტუდენტებისათვის:

მოსალოდნელი სიხშირეები ნაჩვენებია 11.3 ცხრილში.

ცხრილი 11.3 მოსალოდნელი სიხშირეები პირველი თაობის კოლეჯის სტატუსის ქალი და მამაკაცი სტუდენტებისათვის

პირველი თაობის სტატუსის მქონე სტუდენტები	ქალი	მამაკაცი	ჯამი
პირველი თაობა	817.14	1,118.86	1,936
სხვა თაობა	1,132.86	1,551.14	2,684
ჯამი (*N*)	1,950	2,670	4,620

შევნიშნოთ, რომ მოსალოდნელი სიხშირეების ცხრილის კიდურა უჯრების მნიშვნელობები თავდაპირველი, 11.1 ცხრილის კიდურა უჯრების მნიშვნელობების ტოლია. მიუხედავად იმისა, რომ მოსალოდნელი სიხშირეები ჩვეულებრივ განსხვავდება დაკვირვების ფაქტობრივი მონაცემებისაგან (რაც თავის მხრივ დამოკიდებულია ცვლადებს შორის დამოკიდებულების სახეზე), ცხრილის სვეტების და სტრიქონების კიდურა უჯრების მნიშვნელობები საწყისი ცხრილის კიდურა უჯრების მნიშვნელობების ტოლია.

გამოიყენეთ 341-ე გვერდზე მოცემული მონაცემები. გამოთვალეთ მოსალოდნელი სიხშირეები ასაკისა და პირველი თაობის სტატუსის მქონე სტუდენტებისათვის და ააგეთ ორგანზომილებიანი ცხრილი. არის თუ არა სვეტისა და სტრიქონის კიდურა უჯრების მნიშვნელობები საწყისი ცხრილის კიდურა უჯრების მნიშვნელობების ტოლი?

✓ *ცოდნის შემოწმება*

თანხმობის ხი-კვადრატ კრიტერიუმის გამოთვლა

შემდეგი ნაბიჯი ხი-კვადრატის გამოსათვლელად არის ცხრილის თითოეული უჯრის მიხედვით მოსალოდნელ და დაკვირვების სიხშირეებს შორის სხვაობების შედარება. შევნიშნოთ, რომ ჩვენს მაგალითში მოსალოდნელ და ფაქტობრივ სიხშირეებს შორის სხვაობა თითოეული უჯრისათვის საკმაოდ მაღლია. მაგრამ რამდენად საკმარისია ამგვარი სხვაობის არსებობა, რომ იგი მნიშვნელოვნად მივიჩნიოთ? ამ კითხვაზე პასუხის გაცემისათვის უნდა გამოვთვალოთ თანხმობის ხი-კვადრატ კრიტერიუმი:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

სადაც,

f_0 = არის დაკვირვების მონაცემების სიხშირეები

f_e = არის მოსალოდნელი სიხშირეები

თანხმობის ხი-კვადრატ კრიტერიუმი სტატისტიკური კრიტერიუმი, რომელიც აჯამებს სხვაობას დაკვირვებისა (f_0) და მოსალოდნელ (f_e) სიხშირეებს შორის.

ცხრილი 11.4 დაკვირვებისა და მოსალოდნელი სიხშირეები პირველი თაობის სტატუსის ქალი და მამაკაცი სტუდენტებისათვის

პირველი თაობის სტუდენტები	მამაკაცები		ქალები		ჯამი
	f_0	f_e	f_0	f_e	
პირველი თაობა	691	817.14	1,245	1,118.86	1,936
სხვა თაობა	1,259	1,132.86	1,425	1,551.14	2,684
ჯამი (N)	1,950	2,670	4,620		

11.2 ფორმულის მიხედვით, ცხრილის თითოეული უჯრისათვის გამოვაკლოთ მოლსადონელი სიხშირეების მაჩვენებლებს დაკვირვების ფაქტობრივი მნიშვნელობები, ავიყვანოთ კვადრატში და გავყოთ მოსალოდნელ სიხშირეზე. თითოეული უჯრისათვის ამ მოქმედებების ჩატარების შემდეგ დავაჯამოთ შედეგები ხი-კვადრატ სტატისტიკის მისაღებად.

მივყვეთ ზემოთ აღწერილ მოქმედებებს ასევე ცხრილ 11.4-ისათვის, გამოთვლის შედეგები ნაჩვენებია ცხრილ 11.5-ში. თანხმობის ხი-კვადრატ კრიტერიუმი არის 57.99. აღნიშნული მაჩვენებელი აჯამებს სხვაობას მოსალოდნელ და ფაქტობრივი დაკვირვების სიხშირეებს შორის და ასევე აჩვენებს სხვაობას იმ შემთხვევისათვის, როცა ნულოვანი ჰიპოთეზა ჭეშმარიტია ე.ი. სქესი და პირველი თაობის კოლეჯის სტუდენტის სტატუსი არ არის დაკავშირებული. შემდეგ ეტაპზე საჭიროა მიღებული თანხმობის ხი-კვადრატ კრიტერიუმის ინტერპრეტირება და შეფასება იმისა, თუ რამდენად დიდია მისი მნიშვნელობა ნულოვანი ჰიპოთეზის უარსაყოფად.

ცხრილი 11.5 ხი-კვადრატი კრიტერიუმის გამოთვლა

პირველი თაობის სტუდენტები	f_0	f_e	$f_0 - f_e$	$(f_0 - f_e)^2$	$\frac{f_0 - f_e}{f_e}$
მამაკაცები/1-ლი თაობა	691	817.14	-126.14	15,911.2996	19.47
მამაკაცები/სხვა თაობა	1,259	1,132.86	126.14	15,911.2996	14.04
ქალები/1-ლი თაობა	1,245	1,118.86	126.14	15,911.2996	14.22
ქალები/სხვა თაობა	1,425	1,551.14	-126.14	15,911.2996	10.26

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 57.99$$

11.5 ცხრილის მსგავსი ფორმის გამოყენებით ააგეთ ცხრილი და გამოთვალეთ ხი-კვადრატ კრიტერიუმი შემდეგი ორი ცვლადისათვის: ასაკი და განათლების მიღწევები.

✓ *ცოდნის შემოწმება*

ხი-კვადრატის შერჩევითი განაწილება

როგორც მე-9 თავში ვისწავლეთ ისეთი კრიტერიუმის სტატისტიკებს, როგორიცაა Z და t სტატისტიკა, აქვთ თავიანთი მახასიათებელი შერჩევითი განაწილება, რომლებზე დაყრდნობითაც შესაძლებელია განვსაზღვროთ ნულოვანი პიპოთეზის მცდარობა ან ჭეშმარიტობა. მათ მსგავსად, ხი-კვადრატის შერჩევითი განაწილება გვაძლევს ხი-კვადრატ კრიტერიუმის მნიშვნელობების ცალეკულ ალბათობებს, იმ დაშვებით რომ შესასწავლ ცვლადებს შორის არ არსებობს კავშირი პოპულაციის დონეზე.

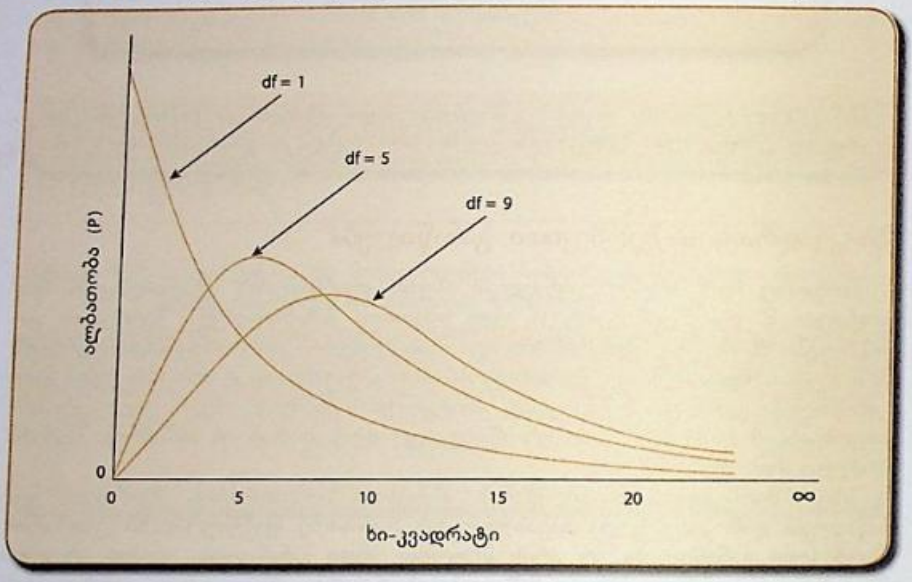
სხვა შერჩევითი განაწილებების მსგავსად, ხი-კვადრატის შერჩევითი განაწილება დამოკიდებულია თავისუფლების ხარისხზე. ფაქტიურად, ხი-კვადრატის შერჩევითი განაწილება არ არის მხოლოდ ერთი განაწილება, არამედ ეს არის, მსგავსად t განაწილებისა, განაწილებების ნაკრები. ხი-კვადრატ განაწილების კონკრეტული ფორმა დამოკიდებულია თავისუფლების ხარისხის მნიშვნელობაზე. ამ უკანასკნელის ილუსტრირება მოცემულია 11.1 ნახაზზე, რომელზეც ნაჩვენებია ხი-კვადრატ განაწილება თავისუფლების 1, 5 და 9 ხარისხებისათვის. ნახაზზე დაკვირვებით შესაძლებელია შევნიშნოთ ხი-კვადრატ განაწილების რამდენიმე მნიშვნელოვანი თვისება:

- განაწილების მნიშვნელობები ასიმპტოტურად უახლოვდება აბცისათა ღერძს დადებითი მიმართულებით; ხი-კვადრატისათვის კვლევის ჰიპოთეზა ყოველთვის არის ცალმხრივი კრიტერიუმი;

- ხი-კვადრატის მნიშვნელობები ყოველთვის არის დადებითი. მინიმალური მნიშვნელობაა 0, ხოლო მაქსიმალური დადებითი ზღვარი არ არსებობს. ხი-კვადრატის ნულთან ტოლობა ნიშნავს, რომ ცვლადები ერთმანეთისაგან სრულიად დამოუკიდებელია და დაკვირვების მნიშვნელობები ცხრილის თითოეული უჯრისათვის შეესაბამება მათსავე მოსალოდნელ სიხშირეების მნიშვნელობებს;

- თავისუფლების ხარისხის ზრდასთან ერთად ხი-კვადარტ განაწილების ფორმა ხდება უფრო სიმეტრიული, 30-ზე მეტი თავისუფლების ხარისხის შემთხვევაში ამ განაწილების ფორმა ნორმალურ განაწილების ფორმას უახლოვდება.

ნახაზი 11.1 ხი-კვარდატ განაწილება თავისუფლების 1, 5 და 9 ხარისხებით.



თავისუფლების ხარისხის განსაზღვრა

როგორც მე-9 თავში განსაზღვრეთ, თავისუფლების ხარისხი არის იმ ცვლადების მნიშვნელობების რაოდენობა, რომელთა მნიშვნელობები თავისუფლად ცვალებადობს. ჯვარედინად განლაგებულ მონაცემებში თავისუფლების ხარისხს შემდეგი ფორმულის გამოყენებით ვპოულობთ:

$$df = (r - 1)(c - 1)$$

სადაც,

r = სტრიქონების რაოდენობაა
c = სვეტების რაოდენობა

შესაბამისად, ცხრილ 11.1-ის (2 სვეტი და 2 სტრიქონი) თავისუფლების ხარისხი იქნება 1 (2-1)(2-1). თუ ცხრილს 3 სტრიქონი და 2 სვეტი ექნებოდა შესაბამისი თავისფლების ხარისხი იქნებოდა 2 (3-1)(2-1).

წიგნის ბოლოში არსებულ „დანართ D"-ში მოცემულია ხი-კვადრატ განაწილება სხვადასხვა თავისუფლების ხარისხებისათვის. დააკვირდით თუ როგორაა განაწილებული თავისუფლების ხარისხის მაჩვენებლები პირველი სვეტში,

დაღმავალი რიგით, ხოლო მნიშვნელოვნების დონეები (ან P მნიშვნელობები) ჰორიზონტალურად, პირველი სტრიქონის გასწვრივ. მაგალითად, თავისუფლების ხარისხის 5 მნიშვნელობის შემთხვევაში, ხი-კვადრატ კრიტერიუმის 15.086 მნიშვნელობის მიღების ალბათობა 0.01-ია. მიღებული ხი-კვადრატ კრიტერიუმი 15.086-ისათვის მხოლოდ ერთხელ დაფიქსირდება 100 შერჩევის ერთეულში.

ორგანზომილებიანი ცხრილისათვის თავისუფლების ხარისხი შესაძლოა განიმარტოს, როგორც იმ უჯრების რაოდენობა, რომლებისთვისაც შესაძლებელია მოსალოდნელი სიხშირეების ცვალებადობა იმ პირობით, რომ კიდურა უჯრების ჯამები უკვე განსაზღვრულია. 11.3 ცხრილის მონაცემებზე დაყრდნობით ჯერ გამოვთვალოთ მოსალოდნელი სიხშირეები პირველი თაობის კოლეჯის სტატუსის მამაკაცი სტუდენტებისათვის ($f_e = 817.14$). რადგან პირველი სვეტის მოსალოდნელი სიხშირეების ჯამი აღებულია 1,950-ის ტოლად, მოსალოდნელი სიხშირე სხვა თაობის კოლეჯის სტატუსის მქონე მამაკაცი სტუდენტებისათვის იქნება 1,132.86(1,950-817.14). ამასთან, სხვა უჯრების ზღვრული ჯამები წინასწარ განისაზღვრება და მათი მნიშვნელობები ცვალებადობას არ ექვემდებარება. შესაბამისად, ამ ცხრილის მონაცემების თავისუფლების ხარისხი 1-ის ტოლია.

ორგანზომილებიან ცხრილში მონაცემები შესაძლოა არასწორად იქნას შეტანილი შემთხვევითი დაკვირვების დროს, თუ მოცემულ ცვლადზე დაკვირვება სხვასთან შედარებით მეტად ან ნაკლებად ხდება, რაც თავის მხრივ გავლენას მოახდენს ხი-კვადრატის მნიშვნელობაზე. ამ უკანასკნელის კომპენსირება შესაძლებელია თავისუფლების ხარისხის მაჩვენებლის გამოყენებით, მაგრამ 2x2-ზე ზომის ცხრილის შემთხვევაში თავისუფლების მხოლოდ ერთი ხარისხი გვაქვს. ხი-კვადრატის მნიშვნელობები უნდა მოსწორდეს იეტის უწყვეტობის კორექციის კრიტერიუმის გამოყენებით. ფორმულა ამცირებს თითოეული სხვაობის აბსოლიტურ მნიშვნელობას ($f_0 - f_e$) ხუთი ერთეულით, შემდეგ სხვაობა ადის კვადრატში და იყოფა თითოეული უჯრის მოსალოდნელ სიხშირეებზე. იეტის უწვეტობის კორექციის კრიტერიუმის ფორმულა შემდეგია:

$$\chi_c^2 = \sum \frac{(|f_0 - f_e| - 0.5)^2}{f_e} \qquad (11.4)$$

✓ *ცოდნის შემოწმება*

„დანართ დ“-ს მიხედვით, განსაზღვრეთ ქვემოთ მოყვანილი თითოეული ხი-კვადრატ კრიტერიუმის მნიშვნელობების შესაბამისი ალბათობები (თავისუფლების ხარისხები მოცემულია ფრჩხილებში):

- *12.307 (15)*
- *20.337 (21)*
- *54.052 (24)*

საბოლოო გადაწყვეტილების მიღება

იეტსის უწყვეტობის კორექციის კოეფიციენტის გამოყენებით მიღებული ხი-კვადრატის მნიშვნელობა 57.54-ია. გამოთვლები ნაჩვენებია ცხრილ 11.6-ში.

ცხრილი 11.6 იეტის უწყვეტობის კორექციის კოეფიციენტის გამოთვლა

სქესი და პირველი თაობის სტუდენტის სტატუსი	$\lvert f_o - f_e \rvert$	$(\lvert f_o - f_e \rvert - 0.50)^2$	f_e	$\frac{(\lvert f_o - f_e \rvert - 0.5)^2}{f_e}$
პირველი თაობა (მამაკაცები)	-126.14	(125.64)2=15,785.41	817.14	19.32
სხვა თაობა (მამაკაცები)	126.14	(125.64)2=15,785.41	1,132.86	13.93
პირველი თაობა (ქალები)	126.14	(125.64)2=15,785.41	1,118.86	14.11
სხვა თაობა (ქალები)	-126.14	(125.64)2=15,785.41	1,551.14	10.18
ჯამი				57.54

ცხრილიდან შეგვიძლია შევნიშნოთ, რომ სიდიდე 57.54 არ ფიქსირდება ცხრილის პირველ რიგში(თავისუფლების ხარისხი 1-ის ტოლია). ფაქტობრივად იგი აჭარბებს ხი-კვადრატის მაქსიმალურ მნიშვნელობას 10.827, როცა (P=0.001). აქედან გამომდინარე, 0.001-ზე ნაკლებია ალბათობა ხი-კვადრატის იმ მნიშვნელობის მიღებისა, რომელიც ნაკლებია 57.54-ზე; იმ დაშვებით, რომ რომ ნულოვანი ჰიპოთეზა ჭეშმარიტია. თუ ალფას მნიშვნელობა იქნებოდა 0.05, მაშინ 10.827-ის მნიშვნელობის მიღების ალბათობა გაცილებით ნაკლები იქნებოდა. შესაბამისად, ჩვენ შეგვიძლია უარვყოთ ნულოვანი ჰიპოთეზა იმის თაობაზე, რომ სქესი და პირველი თაობის კოლეჯის სტუდენტის სტატუსი არ არის ურთიერთდაკავშირებული მთლიანად პოპულაციაში. დავიმახსოვროთ, რომ რაც უფრო მეტია ხი-კვადრატ კრიტერიუმის მნიშვნელობა და რაც უფრო მცირეა შესაბამისი P-ს მნიშვნელობა შეგვიძლია მეტი დარწმუნებით უარვყოთ ნულოვანი ჰიპოთეზა. ჩვენ შეგვიძლია კიდევ უფრო დარწმუნებულნი ვიყოთ ჩვენს დასკვნაში ნულოვანი ჰიპოთეზის უარყოფის თაობაზე, განსაკუთრებით იმ ფაქტზე დაყრდნობით, რომ ამ შემთხვევაში შერჩევის შეცდომა 0.001-ია.

მიმოხილვა

შევაჯამოთ ზემოთ განხილული ჰიპოთეზის შემოწმების ზუთსაფეხურიანი პროცესი:

დაშვებების გაკეთება. ჩვენს მიერ გაკეთებული დაშვებები შემდეგია:

1. შემთხვევითი შერჩევის ზომა: =4,620
2. გაზომვის სკალა ცვლად „სქესისათვის“: ნომინალური
3. გაზომვის სკალა ცვლად „პირველი თაობის კოლეჯის სტუდენტის სტატუსი“: ნომინალური

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება და ალფას მნიშვნელობის არჩევა. კვლევის ჰიპოთეზა H_1 შემდეგში მდგომარეობს: არსებობს კავშირი სქესსა და პირველი თაობის კოლეჯის სტუდენტის სტატუსს შორის (ე.ი. სქესი და პირველი თაობის კოლეჯის სტუდენტის სტატუსი სტატისტიკურად დამოკიდებულია). ნულოვანი ჰიპოთეზა H_0: არ არსებობს კავშირი სქესსა და პირველი თაობის კოლეჯის სტუდენტის სტატუსს შორის(ე.ი. სქესი და პირველი თაობის კოლეჯის სტუდენტის სტატუსი სტატისტიკურად დამოუკიდებელია). ალფას მნიშვნელობა 0.05.

შერჩევითი ერთობლიობის განაწილების არჩევა და სტატისტიკური კრიტერიუმის დაზუსტება. შერჩევითი ერთობლიობის განაწილების ფორმა და სტატისტიკური კრიტერიუმი ორივე შემთხვევაში ხი-კვადრატია.

სტატისტიკური კრიტერიუმის მნიშვნელობის გამოთვლა.
პირველ რიგში უნდა განვსაზღვროთ თავისუფლების ხარისხი, მოცემულ კრიტერიუმთან მიმართებაში:

$$df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = (1)(1) = 1$$

შემდეგი ნაბიჯია ხი-კვადრატის მნიშვნელობის გამოთვლა, მოსალოდნელ სიხშირეებს ვითვლით სტატისტიკური დამოუკიდებლობის დაშვებით. მოსალოდნელი სიხშირეების გამოთვლისათვის სვეტებისა და სტრიქონების კიდურა უჯრებში მოცემული ჯამები გავამრავლოთ და გავყოთ N-ზე. მოსალოდნელი სიხშირეები ნაჩვენებია ცხრილ 11.3-ში.

არის თუ არა ცხრილ 11.3-ში მოყვანიალი მოსალოდნელი სიხშირეები საკმარისად განსხვავებული დაკვირვების ფაქტობრივი მონაცემებისაგან რათა ნულოვანი ჰიპოთეზის უარყოფა მოხდეს? აღნიშნულ კითხვაზე პასუხის გაცემისათვის გამოვთვალოთ ხი-კვადრატი(იეტსის კორექციით) 57.54-ისათვის. გამოთვლის შედეგები ნაჩვენებია ცხრილ 11.6-ში.

გადაწყვეტილების მიღება და შედეგების ინტერპრეტირება. ხი-კვადრატის (57.54) შესაბამისი ალბათობის მნიშვნელობის პოვნისათვის მივმართოთ წიგნის „დ დანართს". თავისუფლების ერთი ხარისხით, ალბათობა იმისა რომ მივიღოთ 57.54-ის ტოლი მნიშვნელობა ნაკლებია 0.001-ზე (ალფას 0.05 მნიშვნელობაზე ნაკლები). აქედან გამომდინარე უარვყოფთ ნულოვან ჰიპოთეზას, რაც ნიშნავს რომ სტატისტიკურად არსებობს კავშირი სქესსა და პირველი თაობის კოლეჯის სტუდენტის სტატუსს შორის. შესაბამისად, მოცემული შედეგი შესაძლებელია განივრცოს იმ პოპულაციაზე, საიდანაც მიღებულია ჩვენი შერჩევითი ერთობლიობა. ჩვენს შერჩევით ერთობლიობაზე დაყრდნობით უკვე ვიცით, რომ ქალი სტუდენტების შემთხვევაში მეტი მოლოდინი არსებობს, რომ ისინი პირველი თაობის სტუდენტები არიან ვიდრე ეს მამაკაცი სტუდენტების შემთხვევაშია.

✓ *ცოდნის შემოწმება*

რა გადაწყვეტილებას მიიღებთ ორი ცვლადის "ასაკისა" და "პირველი თაობის კოლეჯის სტუდენტის სტატუსს" შორის სტატისტიკური კავშირის თაობაზე? უარყოფთ თუ არა ნულოვან ჰიპოთეზას ალფას 0.05 ან 0.01 დონისათვის?

დეტალების განხილვა 11.1
შეცდომის წყარო: შერჩევის ზომა და სტატისტიკური მნიშვნელობა ხი-კვადრატისათვის

როგორც წინა პარაგრაფებში ვნახეთ კავშირი ცვლად სიდიდეებს სქესსა და პირველი თაობის კოლეჯის სტუდენტის სტატუსს შორის სტატისტიკურად მნიშვნელოვნია, მაგრამ თავად კავშირის არსებობა არ გვაძლევს ინფორმაციას იმის შესახებ თუ რამდენად ძლიერია ეს კავშირი ან არსებითად მნიშვნელოვანია თუ არა მთელი პოპულაციისათვის. სტატისტიკური მნიშვნელოვნება გვეხმარება მხოლოდ იმის შეფასებაში არის თუ არა არსებული კავშირი შემთხვევითი ან მიზეზ-შედეგობრივი; მაგრამ ამასთანავე იგი არაფერს ამბობს იმის შესახებ არის თუ არა ეს კავშირი თეორიულად მნიშვნელოვანი შემდგომი შესწავლისათვის.

განსხვავება სტატისტიკურ და არსებით მნიშვნელოვნებას შორის მნიშვნელოვანია იმ ნებისმიერი სტატისტიკური კრიტერიუმის გამოყენებისას, რაც მე-9 თავში განვიხილეთ. განსაკუთრებით ეს ეხება ხი-კვადრატსაც შერჩევის ზომის მნიშვნელობიდან გამომდინარე. ხი-კვადრატის გამოთვლილი მნიშვნელობა პროპორციულად უკავშირდება შერჩევის ზომას, ცვლადებს შორის არსებული კავშირის სიძლიერის მიუხედავად.

მაგალითისათვის, ცხრილ 11.1-ში მოცემული დაკვირვების შედეგები გავყოთ ზუსტად ორზე, რაც ფაქტობრივად ნიშნავს შერჩევის ზომის განახევრებას. თავის მხრივ ეს ცვლილება გაელენას არ მოახდენს სქესის მიხედვით პირველი თაობის კოლეჯის სტატუსის პროცენტულ განაწილებაზე. აქედან გამომდინარე, პროცენტული სხვაობის ზომა და კავშირის სიძლიერე სქესსა და პირველი თაობის სტატუსს შორის დარჩება იგივე და არ შეიცვლება. ამასთან, დაკვირვების სიხშირეების რიცხვის განახევრება გამოთვლილი ხი-კვადრატსაც ასევე გაანახევრებს (თავად გადაამოწმეთ ეს ცვლილება!) 57.54-დან 28.77 მდე. თუ იგივე მოქმედებებს საწინააღმდეგოდ გავაკეთებდით – გავაორმაგებდით დაკვირვების სიხშირეებს შესაბამისად გაორმაგდებოდა ხი-კვადრატის მნიშვნელობა და ნულოვანი ჰიპოთეზის უარყოფაც გამარტივდებოდა.

ხი-კვადრატის მნიშვნელობის ცვლილება შერჩევის ზომიდან გამომდინარე, ნიშნავს რომ ცვლადების შორის ძლიერი კავშირის არსებობის შემთხვევაში კავშირის სიძლიერე არ გამოვლინდება თუ შერჩევის ზომა მცირეა. და პირიქით ცვლადებს შორის სუსტი კავშირის შემთხვევაში შერჩევის ზომის სიდიდე საბოლოოდ კავშირის სტატისტიკური მნიშვნელობის სიძლიერეში აისახება. აქედან გამომდინარე, თუ გამოთვლილი ხი-კვადრატის მნიშვნელობა დიდია და შესაბამისად ნულოვანი ჰიპოთეზის უარყოფა დიდი სხვაობით ხდება ეს საბოლოოდ არ ნიშნავს, რომ ცვლადებს შორის კავშირი ძლიერი და არსებითად მნიშვნელოვანია.

ხი-კვადრატის მეორე შეზღუდვა არის მისი დამოკიდებულება ცხრილის თითოულ უჯრაში მოსალოდნელი სიხშირეების მცირე მნიშვნელობებზე. ჩვეულებრივ, როცა ცხრილის ერთ ან მეტ უჯრაში მოსალოდნელი სიხშირეების მნიშვნელობები 5-ზე ნაკლებია, ხი-კვადრატის მნიშვნელობები ცვლებადია და მისმა გამოყენებამ შესაძლოა მნიშვნელოვანი შეცდომები მოგვტანოს. ამასთან, გასათვალისწინებელია ისიც, რომ მკაცრად განსაზღვრული წესი მოსალოდნელი სიხშირეების ზომასთან დაკავშირებით არ არსებობს. უმეტესობა მკვლევარებისა ზღუდავს ხი-კვადრატის გამოყენებას ან იყენებენ შემდეგ ორ მიდგომას: მოსალოდნელი სიხშირეების (f_e) მნიშვნელობას არ უშვებენ, როგორც 5-ზე ნაკლებს ან მთლიანად ცვლადების არაუმეტეს 20 %-ის მოსალოდნელი სიხშირეების მნიშვნელობა არის 5-ზე ნაკლები.

სტატისტიკური კრიტერიუმის გამოყენება ორ ცვლადის შემთხვევაში არის მხოლოდ მცირე, მაგრამ მნიშვნელოვანი ნაბიჯი ცვლადებს შორის კავშირის შესაფასებლად. ხი-კვადრატის მნიშვნელოვნება მეტყველებს იმაზე, რომ ცვლადებს შორის არსებობს კავშირი (სუსტი ან ძლი-

ერი), იგივე კავშირი მოსალოდნელია რომ არსებობს მთლიანად პოპულაციაში და ეს კავშირი არ არის მხოლოდ შერჩევის ცვლადებადობით განპირობებული. კავშირის სიძლიერის შესაფასებლად უნდა გამოვიყენოთ ისეთი მაჩვენებლები, როგორებიცაა: გამა, ლამბდა ან პირსონის r (თავი 12 და 13) . სტატისტიკური კრიტერიუმებისა და მაჩვენებლების ერთდროული გამოყენება გვეხმარება განვსაზღვროთ არსებობს თუ არა კავშირი ცვლადებს შორის და რამდენად მნიშვნელოვანია ეს კავშირი.

დემონსტრირების განხილვა 11.2
გადაწყვეტილების მიღება: ნულოვანი ჰიპოთეზის შარჭოფის არასაკმარისი პირობები

მე-9 თავში ვისწავლეთ პროპორციებს შორის განსხვავებების შემოწმება. შემოწმება 2x2-ზე ზომის ცხრილების შემთხვევაში შესაძლებელია განხორციელდეს ხი-კვადრატ კრიტერიუმის გამოყენებით. განვიხილოთ მაგალითი „ალკოჰოლის მოხმარების თაობაზე ქალებისა და მამაკაცების შეხედულებები“, 2x2-ზე ცხრილის გამოყენებით.

მოხმარების თაობაზე დადებითი აზრის მქონე რესპონდენტები დავყოთ ორ კატეგორიად: ეთანხმება და მკაცრად არ ეთანხმება. მონაცემები ნაჩვენებია ქვემოთა ცხრილში:

სქესი და ალკოჰოლის მოხმარების შესახებ აზრი

ალკოჰოლის მოხმარების შესახებ აზრი	*მამაკაცი*	*ქალი*	*ჯამი*
ეთანხმება	85.2% (508)	83.6% (514)	84.4% (1,022)
მკაცრად არ ეთანხმება	14.8% (88)	16.4% (101)	15.6% (189)
ჯამი (N)	100% (596)	100% (615)	100% (1,211)

წყარო: Monitoring the Future, 2008

კვლევის ჰიპოთეზა:

H_1: არსებობს კავშირი სქესსა და ალკოჰოლის მოხმარების შესახებ აზრს შორის(სქესი და ალკოჰოლის მოხმარების შესახებ აზრი პოპულაციაში სტატისტიკურად დამოკიდებულია)

ნულოვანი ჰიპოთეზა:

H_0: არ არსებობს კავშირი სქესსა და ალკოჰოლის მოხმარების შესახებ აზრს შორის. (პოპულაციაში სქესი და შეხედულება ალკოჰოლის მოხმარების შესახებ სტატისტიკურად დამოუკიდებელია).

ალფას მნიშვნელობა ავირჩიოთ 0.05-ის დონეზე. შერჩევის განაწილება ხი-კვადრატია, სტატისტიკური კრიტერიუმიც ასევე ხი-კვადრატია. თავისუფლების ხარისხი შემდეგია:

$$df = (r - 1)(c - 1)$$
$$= (2 - 1)(2 - 1)$$
$$= (1)(1) = 1$$

ხი-კვადრატის გამოთვლისათვის უპირველესად განვსაზღვროთ მოსალოდნელი სიხშირეები სტატისტიკური დამოკიდებულების არსებობის დაშვების ქვეშ. მოსალოდნელი სიხშირეები ნაჩვენებია ქვემოთა ცხრილში.

მოსალოდნელი სიხშირეები რესპონდენტის სქესისა და ალკოჰოლის შესახებ აზრს შორის

ალკოჰოლის მოხმარების შესახებ აზრი	მამაკაცები (f_e)	ქალები (f_e)	ჯამი (f_e)
ეთანხმება	508	503	5.00
სრულიად არ ეთანხმება	88	93	-5.00
სულ (N)	514	519	-5.00

ხი-კვადრატის გამოთვლა, ალკოჰოლის მოხმარების შესახებ აზრი სქესის მიხედვით

სქესი და ალკოჰოლის მოხმარების მიმართ დამოკიდებულება	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
მამაკაცი/დადებითი	508	503	5.00	25.00	0.05
მამაკაცი/მკვეთრად უარყოფითი	88	93	-5.00	25.00	0.27
ქალი/დადებითი	514	519	-5.00	25.00	0.05
ქალი/მკვეთრად უარყოფითი	101	96	5.00	25.00	0.26

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 0.63$$

იეტის უწყვეტობის კორექცის კოეფიციენტით, მოსწორებული ხი-კვადრატის მნიშვნელობა იქნება:

$$\chi_c^2 = \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e} = 0.51$$

იმისათვის, რომ განვსაზღვროთ არის თუ არა მნიშვნელოვანი ის განსხვავება, რაც დაფიქსირდა დაკვირვების სიხშირეებსა და მოსალოდნელ სიხშირეებს შორის, განვსაზღვროთ ხი-კვადრატის 0.455-ის შესაბამისი P მნიშვნელობა 0.50 ალბათობით (დანართი D). რადგან ალბათობის მნიშვნელობა შეტია ალფას 0.05-ის მნიშვნელობაზე, ნულოვანი ჰიპოთეზის უარყოფა ვერ მოხდება, რაც ნიშნავს რომ სქესის მიხედვით არ არის განსხვავება ალკოჰოლის მოხმარების შესახებ შეხედულებებში. არ გაქვს შესაბამისი საფუძველი, რომ ვამტკიცოთ სქესსა და ალკოჰოლის მოხმარების შესახებ კავშირის არსებობის შესახებ.

სტატისტიკა პრაქტიკაში: განათლება და ჯანმრთელობის მდგომარეობის შეფასება

ზოგადი სოციოლოგიური კვლევის კითხვარში (2006),რესპონდენტებს თხოვეს დაეფიქსირებინათ თავიანთი ბოლო, ყველაზე მაღალი საფეხურის განათლება და ჯანმრთელობის დონე (ცუდი, საშუალო, კარგი და შესანიშნავი). ეს მონაცემები მოტანილია ცხრილ 11.7-ში. ორგანზომილებიან ცხრილში ნაჩვენებია მკეთრად გამოხატული დადებითი მიმართულების კავშირი განათლებასა(დამოუკიდებელი ცვლადი) და ჯანმრთლების მდგომაროების შეფასებაში(დამოკიდებული ცვლადი). მაგალითად, კოლეჯის განათლების მქონე მოქალაქეების 35%-მა ჯანმრთელობის მდგომარეობა შეაფასა, როგორც „შესანიშნავი“, 12.5%-მა, რომელთაც ქონდათ კოლეჯის უმაღლეს განათლებაზე დაბალი განათლება ჯანმრთელობის მდგომარეობაზე იგივე შეფასება გააკეთა. ასევე, კოლეჯის განათლების მქონე პირების მხოლოდ 2%-მა შეაფასა საკუთარი ჯანმრთელობის მდგომარეობა როგორც „ცუდი“, იგივე შეფასება გააკეთა კოლეჯის უმაღლეს განათლებაზე დაბალი განათლების მქონე პირების 12.5%.

ცხრილი 11.7 ჯანმრთელობის მდგომარეობა განათლების დონის მიხედვით: საერთო სოციოლოგიური კვლევა 2006

	განათლების დონე			
ჯანმრთელობის დონე	საშუალო განათლება	უმაღლესი განათლება	კოლეჯის ან სხვა შესაბამისი განათლება	ჯამი
ცუდი	16 (12.5%)	26 (6.0%)	6 (2.1%)	48 (5.7%)
საშუალო	44 (34.4%)	79 (18.1%)	39 (13.9%)	162 (19.2 %)
კარგი	52 (40.6%)	213 (48.9%)	137 (48.9%)	402 (47.6%)
შესანაშნავი	16 (12.5%)	118 (27.1%)	98 (35.0%)	232 (27.5 %)
ჯამი	128 (100 %)	436 (100%)	280 (99.9%)	844 (100%)

როგორც მოყვანილი მონაცემებიდან ჩანს, ჯანმრთელობის დონეებსა და განათლების სამი ჯგუფის მიხედვით დაყოფილ რესპონდენტებს შორის არსებობს გარკვეული შესაბამისობა. მაგრამ ჩვენ არ ვიცით აღნიშნული მხოლოდ შემთხვევით დაფიქსირებული შედეგია თუ იგი რეალურად, პოპულაციაში არსებულ ტენდენციას აღწერს. ამ კითხვებზე პასუხის გასაცემად უნდა გამოვიყენოთ ხი-კვადრატ კრიტერიუმი, ზემოთა პარაგრაფებში აღწერილი 5 ეტაპის მიხედვით.

დაშვებების გაკეთება. ჩვენი დაშვებები შემდეგია:

1. შერჩევის ზომა $N = 844$
2. ცვლადი სიდიდის „განათლების“ გაზომვის მნიშვნელობები ორდინალურია
3. ცვლადი სიდიდის „ჯანმრთელობის დონის “ მნიშვნელობები ორდინალურია

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება. ჩვენი ჰიპოთეზები შემდეგია:

H_1: პოპულაციაში არსებობს კავშირი განათლების დონესა და ჯანმრთელობის დონის შეფასებას შორის (განათლება და ჯანმრთელობა სტატისტიკურად ურთიერთდაკავშირებულია)

H_0: პოპულაციაში არ არსებობს კავშირი განათლების დონესა და ჯანმრთელობის მდგომარეობას შორის (განათლება და ჯანმრთელობის მდგომარეობა სტატისტიკურად დამოუკიდებელია).

ალფას მნიშვნელობა 0.01.

შერჩევის განაწილების ფორმის არჩევა და შესაბამისი სტატისტიკური კრიტერიუმის დაზუსტება. შერჩევის განაწილება ხი-კვადრატი, ასევე სტატისტიკურ კრიტერიუმად ხი-კვადრატს გამოვიყენებთ.

სტატისტიკური კრიტერიუმის გამოთვლა: თავისუფლების ხარისხი ცხრილ 11.7-ისათვის:

$$df = (r - 1)(c - 1) = (4 - 1)(3 - 1) = (3)(2) = 6$$

ხი-კვადრატის გამოთვლისათვის პირველ რიგში გამოვთვალოთ მოსალოდნელი სიხშირეები სტატისტიკური დამოკიდებულების არსებობის დაშვების პირობებში. ცხრილის თითოეული უჯრისათვის მოსალოდნელი სიხშირეების მისაღებად, სვეტისა და სტრიქონების ჯამის ნამრავლი გავყოთ N-ზე. უკვე გამოვთვალოთ უშუალოდ ხი-კვადრატი, რათა განვსაზღვროთ არის თუ არა დაკვირვებისა და მოსალოდნელ სიხშირეებს შორის არსებული სხვაობა საკმარისი ნულოვანი ჰიპოთეზის უარსაყოფად. (რადგან ცხრილი განზომილებით 2x2-ზე მეტია, არ ვითვლით იეტის კორექციის კოეფიციენტს). გამოთვლების შედეგები ნაჩვენებია ცხრილ 11.8-ში.

გადაწყვეტილების მიღება და შედეგების ინტერპრეტირება. იმის განსასაზღვრად არის თუ არა მნიშვნელოვანი არსებული განსხვავება დაკვირვებისა და მოსალოდნელ სიხშირეებს შორის, შევადაროთ გამოთვლილი ხი-კვადრატის მნიშვნელობები ჩვენი წიგნის დამატებაში მოცემულ მნიშვნელობას. თავისუფლების ხარისხის 6-თან ტოლობის შემთხვევაში, ჩვენი მაგალითის ხი-კვადრატი – 53.96 აჭარბებს ხი-კვადრატის მოცემულ უდიდეს მნიშვნელობას 22.457-ს ($P = 0.001$). აქედან გამომდინარე, ალბათობა იმისა, რომ დაკვირვებისას შესაძლებელია მივიღოთ

ხი-კვადრატის 53.96-ის ტოლი მნიშვნელობა 0.001-ზე ნაკლებია. შესაბამისად, შეგვიძლია უარვყოთ ნულოვანი ჰიპოთეზა-კავშირის არარსებობის შესახებ განათლებასა და ჯანმრთელობის მდგომარეობას შორის. ვასკვნით, რომ იმ პოპულაციაში საიდანაც მიღებულია ჩვენი მაგალითის შერჩევითი ერთობლიობა ჯანმრთელობის მდგომარეობის შეფასება ცვალებადობს განათლების დონის მიხედვით.

ნახაზი 11.8 ხი-კვადრატის გამოთვლა „განათლების“ და „ჯანმრთელობის“ ცვლადი სიდიდეებისათვის

განათლება/ ჯანმრთელობა	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
საშუალო განათლება/ცუდი	16	7.3	8.7	75.9	10.37
საშუალო განათლება/საშუალო	44	24.6	19.4	376.36	15.30
საშუალო განათლება/კარგი	52	61.0	-9.0	81.0	1.33
საშუალო განათლება/შესანიშნავი	16	35.2	-19.2	368.64	10.47
უმაღლესი განათლება/ცუდი	26	24.8	1.2	1.44	0.06
უმაღლესი განათლება/საშუალო	79	83.7	-4.7	22.09	0.26
უმაღლესი განათლება/კარგი	213	207.7	5.3	28.09	0.13
უმაღლესი განათლება/შესანიშნავი	118	119.8	-1.8	3.24	0.03
არასრული კოლეჯის ან შემდგომი განათლება/ცუდი	6	15.9	-9.9	98.01	6.16
არასრული კოლეჯის ან შემდგომი განათლება/საშუალო	39	53.7	-14.7	216.09	4.02
არასრული კოლეჯის ან შემდგომი განათლება/კარგი	137	133.4	3.6	12.96	0.10
არასრული კოლეჯის ან შემდგომი განათლება/შესანიშნავი	98	77.0	21	441	5.73

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 53.96$$

კვლევითი ლიტერატურის მიმოხილვა: დამების მხარდაჭერა და აკადემიური მიღწევები

წინა თავებში ჩვენ განვიხილეთ რამდენიმე მაგალითი, რომლებშიც დავინახეთ თუ როგორაა წარმოდგენილი სტატისტიკური ანალიზი სხვადასხვა პროფესიულ ლიტერატურაში. ის სრული სტატისტიკური ანალიზის დეტალები, რაც ამ თავში გავიარეთ იშვიათად არის წარმოდგენილი კვლევით სტატიებში. უმრავლეს შემთხვევებში, კვლევით სტატიებში უკვე გამოთვილი ხი-კვადრატი და ბივარიანტული ანალიზის შედეგია ნაჩვენები.

ამგავრი მიდგომის მაგალითი მოყვანილია ცხრილ 11.9-ში, რომელიც აღებულია კარლ ბენკსტონის სტატიიდან (1984). სტატია ეხება კავშირს აკადემიურ მიღწევებსა და ოჯახურ მხარდაჭერას შორის ვიეტნამური წარმოშობის ამერიკულ ოჯახებში. კერძოდ, რა გავლენა აქვს და-ძმისა და დედმამიშვილისაგან მიღებულ დახმარებას სწავლის შედეგებზე? ბენკსტონმა დაუშვა, რომ ვიეტნამური წარმოშობის ოჯახებში არსებობს დამკვიდრებული აღქმა ოჯახის, როგორც ერთი მთლიანობის, როგორც – „თანამშრომლობითი ერთეულის“. ბენკსტონმა გამოიყენა გამოკითხვების 1994 წლის ბაზა, რომ შეემოწმებინა კავშირი აკადემიურ მიღწევებსა და და-ძმისაგან მიღებულ მხარდაჭერას შორის.

მონაცემები შეგროვებულ იქნა ორი ჩვეულებრივი საზოგადოებრივი უმაღლესი სკოლიდან და ერთი ნიჭიერთა საზოგადოებრივი უმაღლესი სკოლიდან ნიუ ორლეანში, ვიეტნამური დასახლებების ახლოს. კვლევის კითხვარები დარიგებულ იქნა 402 სტუდენტისათვის.

ბენკსტონი კვლევისას ორ ძირითად მაჩვენებელს დაეყრდნო. ანკეტაში ასახული სტუდენტთა ქულები ეფუძნებოდა მათ მიერვე შევსებულ მონაცემებს აკადემიური მოსწრების შემდეგი შკალის გამოყენებით: A, B, C ან ნაკლები. ხოლო და-ძმებისაგან მიღებული ან და-ძმისათვის გაცემული დახმარების აღსაწერად გამოყენებულ იქნა შემდეგი კატეგორიები: „არასოდეს“, „ხანდახან“ ან „ყოველთვის“.

ცხრილი 11.9-ში ნაჩვენებია დამოკიდებული ცვლადის საშუალო ქულის ჯვარედინი ტაბულაცია, შემდეგი ორი დამოუკიდებელი ცვლადის მიხედვით: 1. იღებს დახმარებას და-ძმებისაგან 2. ეხმარება თავის და-ძმას. ცხრილი უჯრებში თითოეული შემთხვევის პროცენტული განაწილებებია (დაკვირვების სიხშირეები) მოცემული.

ცხრილი 11.9 საშუალო ქულა სასკოლო სწავლებისას, და-ძმების დახმარების შემთხვევაში, პროცენტული განაწილება

	არასოდეს ეხმარება	*ხანდახან*	*ხშირად ან ყოველთვის*	*სტრიქონის ჯამი (N)*
ა) და-ძმისაგან მიღებული დახმარება				
C ან ნაკლები	35.0	13.7	6.8	19.6 (63)
B	50.5	58.3	52.3	55.0 (177)
A	14.6	28.0	40.9	25.5 (82)
სვეტის ჯამი (*N*)	32.0 (103)	54.3 (175)	13.7 (44)	100 (322)
ბ) და-ძმისათვის გაცემული დახმარება				
C ან ნაკლები	54.3	18.3	8.7	19.9 (67)
B	32.6	57.7	60.0	55.1 (185)
A	13.0	24.0	31.3	25.0 (84)
სვეტის ჯამი (*N*)	13.7 (46)	52.1 (175)	34.2 (115)	100 (336)

წყარო: ადაპტირებულია – Carl L. Bankston III, "Sibling Cooparation and Scholastic Perfomance Among Vietnamese American Secondary Students: An Ethnic Social Relations Theory", Sociological Perspectives 41, no.1(1998): 174,175. University of California Press.

შენიშვნები: ა $\chi^2 = 29.33$; $P < 0.001$.
ბ $\chi^2 = 44.32$; $P < 0.001$.

ბენკსტონის კვლევაში შემდგომ მოცემულია ცხრილები, რომლებშიც მოცემულია გამოთვილი ხი-კვადრატის მნიშვნელობები და მიღებული შედეგების სტატისტიკური მნიშვნელობები (P-ს მნიშვნელობები). აღნიშნული მონაცემები მეტყველებს, რომ სწავლის ქულებსა და და-ძმების მხარდაჭერაში არსებობს სტატისტიკურად მნიშვნელოვანი კავშირი. მნიშვნელოვნების დონის 0.001-ზე ნაკლები მნიშვნელობა ნიშნავს, რომ ხი-კვადრატი 29.33 ან 44.32 მნიშვნელობით 1000 დაკვირვების ერთეულში ერთზე ნაკლებად დაფიქსირდება, თუ ორი ცვლადი სიდიდე ურთიერთდაკავშირებული არ იქნებოდა. სხვა სიტყვებით, რომ ვთქვათ ცვლადებს შორის იმ კავშირის დაფიქსირების ალბათობა, რომელიც შერჩევის ცვალებადობის შედეგია, ერთზე ნაკლებია შერჩევის 1000 დაკვირვების ერთეულში.

ამ მონაცემებზე დაყრდნობით ბენკსტონი შემდეგ ანალიზს აკეთებს:

გამოკითხვაში მონაწილე სტუდენტების უმრავლესობა ავლენს კარგ აკადემიურ მოსწრებას მიუხედავად და-ძმებისაგან მიღებული დახმარებისა. კვლევაში მონაწილე სტუდენტების 55%, რომელთაც ყავთ და-ძმა საშუალო ქულას „B“ ასახელებენ; იმ სტუდენტების 51% რომელთაც არასოდეს მიუღიათ დახმარება და-ძმებისაგან იღებენ საშუალო „B“ ქულას; იმ სტუდენტების 58%, რომელთაც და-ძმანი დროდადრო ეხმარებოდნენ სწავლაში იღებენ „B“ შეფასებას; ის სტუდენტების 52% კი, რომლებიც და-ძმებისგან ხშირად ან ყოველთვის იღებენ დახმარებას აღწევენ „B“ შეფასებას სწავლაში.

მიუხედავად „B“ შეფასებათან უმრავლესობა კატეგორიის სტუდენტების თანაკვეთისა, ის სტუდენტები რომლებიც აფიქსირებდნენ ხშირი ან მუდმივი დახმარების არსებობას და-ძმებისაგან უფრო მეტად იღებენ A“A“ შეფასებას, ხოლო ის სტუდენტები, რომლებიც არ იღებდნენ საერთოდ დახმარებას და-ძმებისაგან უფრო მეტი ალბათობით ხვდებოდნენ C და უფრო დაბალი ქულების მიმღებთა რიცხვში. 35% მათგან ვინც დახმარებას არ იღებდა და-ძმისაგან ქონდა საშუალოდ C და უფრო დაბალი შეფასება, მხოლოდ 15% მათგან ვინც არასოდეს არ იღებდა დახმარებას და-ძმებისაგან იღებდა საშუალო „A“ შეფასებას. ასევე, საპირისპიროდ მხოლოდ 7% იმ სტუდენტებისა, რომლებიც ხშირად ან ყოველთვის იღებდნენ დახმარებას საშუალო შეფასების „C“ კატეგორიაში მოხვდნენ და მათგან 41% შეფასების „A“ კატეგორიაში მოხვდნენ.

ცხრილ 11.9(ბ)-ში წარმოდგენილია საშუალო ქულების ჯვარედინი ტაბულაცია და-ძმისაგან მიღებული დახმარების სიხშირის მიხედვით. შეადარეთ ერთმანეთს ცხრილი 11.9(ა) და 11.9(ბ), როგორც ჩანს, ასევე მხედველობაში მისაღებია ისიც, რომ არსებობს გარკვეული ტენდენცია-რესპონდენტები ნაკლებად გადმოსცემენ მათთვის გაწეულ დახმარებას, ხოლო გაცემულ დახმარებას უფრო მეტად აფიქსირებენ. მიუხედავად, ამ მიკერძოების ფაქტორის არსებობისა, დახმარების არსებობის ან არარსებობის სისტემატური დაკავშირება სწავლებასთან მეტყველებს იმაზე, რომ ეს მონაცემები რეალობას ასახავს გარკვეული ხარისხით. ის სტუდენტები, რომლებიც იღებენ მეტ დახმარებას

და-ძმებისაგან აღწევენ უფრო კარგ აკადემიურ შედეგებს ვიდრე ისინი ვინც საერთოდ არ იღებენ ასეთ დახმარებას.

რესპონდენტების მხრიდან მეტი დახმარების გაცემა და-ძმების მისამართით ასევე უკავშირდება მათ ზოგად მაღალ აკადემიურ მოსწრებას. ნახევარი მათგან ვინც მათივე პასუხებით არასოდეს დახმარებია თავიანთ და-ძმას იღებენ C ან ნაკლებ შეფასებებს; 10 დან ერთი ვინც ხშირად ან ყოველთვის ეხმარება თავიანთ და-ძმას სწავლაში იღებს საშუალოდ C ან ნაკლებ შეფასებას5.

კვლევითი ლიტერატურის მიმოხილვა: ძალადობრივი ქცევა სქესის, ასაკისა და რასის მიხედვით

მეორე მაგალითში, პაულ პიქუერომ და რობერტ ბრეიმმა[6] გამოიკვლიეს თუ რა განსხვავებული პირობები ქონდათ იმ პირებს, რომლებიც ავლენდნენ ძალმომრეობით ქცევას სხვების მიმართ, განსხვავებით მათგან ვინც არაძალმომრეობით აყენებდა სხვას შეურაწყოფას. ცხრილ 11.10-ში მოყვანილია დამოკიდებულება ძალმომრეობით და არაძალმომრეობით ქცევას შორის სქესის, რასის და ასაკის მიხედვით, ხი-კვადრატ ანალიზის გამოყენებით. მონაცემები შეგროვებულია 1,503 არასრულწოვანების დანაშაულების აღრიცხვით ქუისლენდში, ავსტრალიაში.

შევნიშნოთ, რომ მიღებული ხი-კვადრატის მნიშვნელობა და შესაბამისი ალბათობა მოყვანილია თითოეული ცვლადის წყვილებისათვის. შესაბამისი ორი მნიშვნელოვანი კატეგორია წყვილებისათვის არის: ძალმომრეობითი ქცევა სქესისა და ასაკის მიხედვით. ცხრილის თითოეულ უჯრაში მოტანილია შესაბამისი პროცენტები.

მკლევართა ანგარიშის მიხედვით, ბიჭები აბსოლიტური უმრავლესობით ჭარბობენ ძალმომრეობითი ქცევის ჩამდენებს შორის. კვლევის მიხედვით გოგონების 18.32% არის ადრესატი ძალმომრეობითი ქცევის, იგივე მაჩვენებელი ბიჭებისათვის 11.71%-ს შეადგენს.

კვლევის შედეგების მიხედვით მოულოდნელი აღმოჩნდა ის ფაქტი, რომ არ დაფიქსირდა განსხვავება ძალადობრივი ქცევის სიჭარბის კუთხით ადგილობრივ (11.83%) და ჩამოსულ ჯგუფებს (12.83%) შორის. ამასთან, პროპორციული გამოსახულებით ძალმომრეობითი ქცევის სიჭარბე უფრო მაღალ ასაკობრივ ჯგუფებში დაქსირდა(14.74%), დაბალ ასაკობრივ ჯგუფებთან შედარებით (9.76%), ეს განსხავება მნიშვნელოვანი იყო $P < 0.01$ დონისათვის[7].

ცხრილი 10.11 ძალმომრეობითი ქცევა ასაკის, რასის და ასაკის მიხედვით

	არაძალმომრეობითი ქცევა (%)	ძალმომრეობითი ქცევა (%)	ჯამი (%)
სქესი	1,146(88.29)	152(11.71)	1,298(100)
მდედრობითი	156(81.68)	35(18.32)	191(100)
მამრობითი	1,302	187	1,489
ჯამი			
			$\chi^2 = 6.331^{**}$
ადგილობრივის სტატუსი	815(87.17)	129(12.83)	935(100)
ადგილობრივი	477(88.17)	64(11.83)	541(100)
ჩამოსული			
			$\chi^2 = 0.317^{**}$
პირველი დანაშაულის	579(90.33)	62(9.67)	641(100)
ასაკი	723(85.26)	125(14.74)	848(100)
14 წლის ნაკლები			
14 წლის და მეტი			
			$\chi^2 = 8.539^{**}$

შენიშვნა: $^{**}P < 0.01$

✓ *ცოდნის შემოწმება*

ორგანზომილებიანი ცხრილისათვის: ასაკი და პირველი თაობის სტუდენტის სტატუსი მიღებული ხი-კვადრატის მნიშვნელობაა 181.15, თავისუფლების ხარისხი 1. წიგნის ბოლოს მოცემული დამატება D-ს გამოყენებით ვპოულობთ, რომ შესაბამისი ალბათობა ნაკლებია 0.001-ზე. ეს ალბათობა ნაკლებია ალფას მნიშვნელობაზე 0.05-ზე. შედეგად, ჩვენ უარვყოფთ ნულოვან ჰიპოთეზას კავშირის არარსებობის შესახებ ასაკსა და პირველი თაობის კოლეჯის სტუდენტის სტატუსს შორის. თუ შერჩევის ზომას გავანახევრებთ, ხი-კვდრატის მნიშვნელობა იქნება 90.85. განსაზღვრეთ P-ს მნიშვნელობა 90.58-ისათვის. რა გადაწყვეტილებას მიიღებთ ნულოვან ჰიპოთეზასთან დაკავშირებით?

ძირითადი განმარტებები

• ხი-კვადრატ კრიტერიუმი დასკვნითი სტატისტიკის მეთოდია, რომელიც ამოწმებს ორგანზომილებიანი ცხრილის სახით მოცემულ, ნომინალურ და ორდინალურ მონაცემებს შორის არსებული კავშირის მნიშვნელოვნებას. კრიტერიუმი გულისხმობს პოპულაციაში კავშირის არარსებობის ე.ი. ცვლადების სტატისტიკური დამოუკიდებლობის შესახებ ნულოვანი ჰიპოთეზის შემოწმებას.

• თანხმობის ხი-კვადრატ კრიტერიუმი (χ^2) შემავჯამებელი მაჩვენებელია

დაკვირვების შედეგად მიღებულ სიხშირეებსა (f_o) და მოსალოდნელ სიხშირეებს (f_e) შორის არსებული სხვაობებისათვის-იმ დაშვებით თუ ნულოვანი ჰიპოთეზა ჭეშმარიტი იქნებოდა და ცვლადებს შორის კავშირი არ იარსებებდა. 2x2 ზომის ყველა ცხრილებისათვის იეტის უწყვეტობის კორექციის კოეფიციენტი გამოიყენება.

- ხი-კვადრატის შერჩევითი განაწილების მიხედვით განისაზღვრება ხიკვადრატის გარკვეული მნიშვნელობის მიღების ალბათობა, იმ დაშვებაზე დაყრდნობით, რომ პოპულაციის დონეზე ცვლადებს შორის სტატისტიკური კავშირი არ არსებობს. კონკრეტული ხი-კვადრატის შერჩევითი განაწილება დამოკიდებულია შესაბამის თავისუფლების ხარისხზე.

საკვანძო ტერმინები

ხი-კვადრატ განაწილება
ხი-კვადრატ კრიტერიუმი
მოსალოდნელი სიხშირეები (f_e)
დაკვირვების სიხშირეები (f_o)
სტატისტიკური დამოუკიდებლობა

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

[GSS08PFP-A]

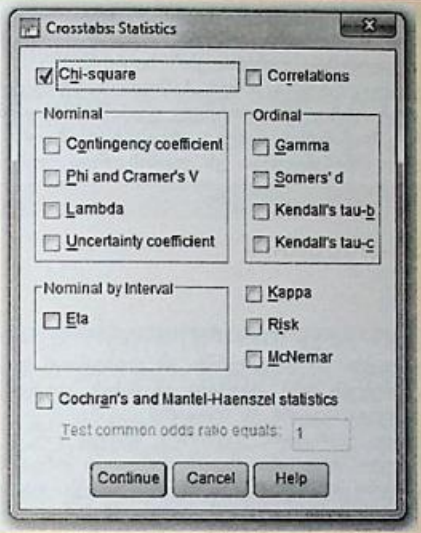
ხი-კვადრატ კრიტერიუმის მიღება ჯვარედინი ცხრილის მონაცემებისათვის

SPSS-ის მეშვეობით ჯვარედინი გამოთვლების (Cross-tabs) ჩატარება უკვე განვიხილეთ მე-10 თავში. იგივე ბრძანებები ასევე შეიძლება გამოყენებულ იქნას ხი-კვადრატ კრიტერიუმის გამოთვლისათვის ორგანოზომილებიანი ცხრილის ფორმით მოცემული მონაცემების შემთხვევაში.

SPSS-ის პროგრამის ჩატვირთვის და შესაბამისი მონაცემთა ბაზის გახსნის შემდეგ დაკლიკეთ შემდეგ ბრძანებათა თანმიმდევრობაზე, ბრძანებების პანელზე: Anayze (ანალიზი) - Descriprive Statictics (აღწერითი სტატისტიკა) – Crosstabs (კროსტაბულაცია) – Statistics (სტატისტიკა), დაკლიკეთ Chi-square ოპციაზე ზედა მარცხენა კუთხეში (ნახ.11.2).

გამოტანილ დიალოგურ ფანჯარაში(ნახ. 11.2) დაკლიკეთ გაგრძელების ბრძანებაზე (Continue). ამ მაგალითში განვიხილავთ შემდეგ ორ ცვლადს შორის კავშირს: განათლების ხარისხი (DEGREE) და პოლიტიკური შეხედულებები (POLVIEWS). გადაიტანეთ ცვლადი POLVIEWS, როგორც მონაცემთა ცხრილის სტრიქონი და ცვლადი DEGREE, როგორც მონაცემთა ცხრილის სვეტი. ბრძანების დასრულებისათვის დაკლიკეთ OK (თანხმობა).

ნახაზი 11.2 ჯვარედინი ტაბულაციის (Crosstabs) დიალოგური ფანჯარა



ბრძანებების მიცემის შემდეგ მივიღებთ შემდეგ შედეგს, რომელიც ნაჩვენებია ნახაზ 11.3-ზე. შედეგი შეიცავს ხი-კვადრატ კრიტერიუმის ანალიზის გამოყენების შედეგებს. ყურადღება გავამახვილოთ მოტანილი ცხრილის პირველ სტრიქონზე, პირსონის ხი-კვადრატზე.

ნახაზი 11.3 ხი-კვადრატ კრიტერიუმის გამოყენების შედეგი SPSS-ში ცვლად: POLVIEWS და DEGREE-ისთვის

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	98.686[a]	24	.000
Likelihood Ratio	100.589	24	.000
Linear-by-Linear Association	12.484	1	.000
N of Valid Cases	1436		

a. 2 cells (5.7%) have expected count less than 5. The minimum expected count is 4.68.

შენიშვნა: ორი უჯრისათვის (5.7%) მოსალოდნელი სიხშირე 5-ზე ნაკლებია.მინიმალური მოსალოდნელი სიხშირე 4.68.

პირსონის ხი-კვადრატის მნიშვნელობა 98.686-ია, თავისუფლების 24 ხარისხით. SPSS-ის გამოთვლების მიხედვით ხი-კვადრატის მნიშვნელობა 0.00-ია. ამ შედეგის ინტერპრეტირება შემდეგია: განათლება და პოლიტიკური შეხედულებები ურთიერთდაკავშირებულია. კერძოდ, განათლების ხარისხის ზრდასთან ერთად მოსალოდნელია რომ როგორც ქალი, ასევე მამაკაცი მოქალაქეები შეხედულებებით უფრო „უკიდურესად ლიბერალები“ ხდებიან.

SPSS-ის მიერ მოცემული შედეგის ბოლო ნაწილი საშუალებას გვაძლევს, შევამოწმოთ დაშვება იმის შესახებ რომ ცხრილის თითოეულ უჯრაში მონაცემი 5-ის ტოლი ან მასზე მეტია. პროგრამის მიერ გამოტანილი შედეგი აჩვენებს, რომ ორი უჯრის ან მონაცემების 5.7%-ს აქვს 5-ზე ნაკლები მნიშვნელობა. ეს უკანასკნელი უფრო ნაკლებია ვიდრე ჩვენი ზღვრული 20%.

ამოცანები SPSS-ისათვის

[GSS08PFP-A და MTF2008]

1. ზოგადი სოციოლოგიური კვლევა (GSS 2008) მოიცავს რიგ კითხვებს, რომელიც უკავშირდება ქალის როლს ოჯახსა და სამსახურში. ამ კითხვებთან დაკავშირებით თავდაპირველი მოსალოდნელობა არსებობს, რომ პასუხები ამ სახის კითხვებზე განსხვავებული იქნება სქესის მიხედვით, სწორია თუ არა ეს მიდგომა?

ა. SPSS-ის გამოიყენებით გაარკვიეთ კავშირის შემდეგ ორ ცვლადს შორის: SEX და FECHLD (დასაქმებული შვილიანი ქალი). გააკეთეთ ორგანზომილებიანი ცხრილი და გამოთვალეთ შესაბამისი მოსალოდნელი მნიშვნელობები და პროცენტები. აქვს თუ არა ამ უკანასკნელ ცხრილს ბევრი უჯრა, რომელთა მნიშვნელობები ნაკლებია 5-ზე? არის თუ არა მონაცემთა ბაზაში რაიმე განსხვავებული შედეგები წინასწარ მოლოდინებთან მიმართებაში?

ბ. უკვე მიღებული ცხრილისათვის გამოთვალეთ ხი-კვადრატის მნიშვნელობა SPSS-ის გამოყენებით.

გ. ჩამოაყალიბეთ ნულოვანი ჰიპოთეზა მნიშვნელოვნების 0.05 დონისათვის. რა დასკვნას გამოიტანთ?

დ. აირჩიეთ სხვა, დემოგრაფიასთან დაკავშირებული ცვლადი (DEGREES ან CLASS) და დაადგინეთ კავშირი შემდეგ ცვლადთან FECHLD.

2. სოციოლოგიური კვლევისას დასმულ იქნა შემდეგი კითხვა, როგორც ქალების ასევე მამაკაცების მისამართით: უმჯობესია კაცისათვის/ქალისათვის იყოს სახლის დიასახლისი ან იმუშაოს? დაადგინეთ კავშირი საოჯახო მდგომარეობას (დაქორწინებული, მარტოხელა-MARITAL) და ნახსენებ კითხვაზე პასუხს შორის (FEFAM). ალფას მნიშვნელობა აირჩიეთ 0.05-ის ტოლად. ორი ცვლადების ხი-კვადრატის მნიშვნელობების მიღების შემდეგ SPSS-ში რა დასკვნას გამოიტანთ?

3. დაადგინეთ კავშირი ცვლად SEX და FEPRESCH-ს შორის. იწვევს თუ არა დედის მუშაობა ბაღის ასაკამდელი ბავშვებისათვის ფსიქოლოგიურ პრობლემებს?

ა. გამოითვალეთ SPSS-ით ცვლადების ხი-კვადრატის მნიშვნელობები. ქალთა რა პროცენტი „არ ეთანხმება“ ან „მკვეთრად არ ეთანხმება“ ზემოთ მოყვანილ შეხედულებას. მამაკაცების რა პროცენტი ფიქრობს იგივეს?

ბ. შეამოწმეთ ნულოვანი ჰიპოთეზა 0.05 მნიშვნელოვნების დონისათვის რა დასკვნას გამოიტანთ?

4. მონაცემები კვლევიდან Monitoring the Future (2008) აღწერს არასრულწლოვანების დამოკიდებულებას სხვადასხვა ტიპის ნარკოტიკის გამოყენების მიმართ: ალკოჰოლი (ATDRINK), ჰეროინი (ATHERION), მარიხუანა (ATWEED) და კოკაინი (ATCOKE). შეადგინეთ ორგანზომილებიანი ცხრილები დასახელებული ცვლადების გამოყენებით და ასევე შემდეგ დემოგრაფიულ ინდიკატორების მიხედვით: სქესი, რასა ან ასაკი. SPSS-ით შესაბამისი პროცენტებისა და ხი-კვადრატის მნიშვნელობებზე დაყრდნობით (ალფა აიღეთ

0.05-ის ტოლად) რა კავშირი არსებობს არჩეულ დემოგრაფიულ მაჩვენებლებსა და ნახსენები ნარკოტიკების გამოყენების მიმართ არსებულ შეხედულებებს შორის?

თავის სავარჯიშოები

1. განხილულ იქნა კავშირი „სქესსა“ და სახლის მიმდებარე ტერიტორიებზე „ღამით სეირნობის შიშს“ შორის (FEAR), 2006 წლის სოციოლოგიური კვლევის მონაცემებზე დაყრდნობით (GSS 2006).

ეშინია ღამით სეირნობა	მამაკაცები	ქალები	ჯამი
დიახ	94	242	336
არა	343	295	638
ჯამი	437	537	974

ა. მოცემული ცხრილისათვის გამოთვალეთ თავისუფლების ხარისხი
ბ. შეამოწმეთ ნულოვანი ჰიპოთეზა იმის თაობაზე, რომ ღამით სეირნობის შიშსა და სქესს შორის არ არსებობს კავშირი ($\alpha = 0.05$). რა დასკვნას გამოიტანთ?
გ. თუ ალფას 0.01-ის ტოლად ავირჩევდით, როგორ შეიცვლებოდა უკვე მიღებული დასკვნა? პასუხი განმარტეთ.

2. საკითხი იმის შესახებ თუ რამდენი ოდენობის თანხა უნდა დაიხარჯოს ამერიკის შეერთებულ შტატებში არსებული სოციალური პრობლემების გადასაჭრელად ყოველთვის დებატების და უთანხმოების საკითხია. რასასა და სოციალურ კლასებს აქვს გავლენა იმ ხარჯების ხალხისათვის გადაცემაზე, რასაც მთავრობა გასწევს სოციალური ხარჯების სახით. ამ საკითხებთან დაკავშირებით სოციოლოგიური კვლევა მოიცავს რამდენიმე შეკითხვას (GSS 2006). ამ დავალებაში გამოიყენებთ ცვლადებს რასას და ცვლად NATFARE, ეს უკანასკნელი გულისხმობდა კითხვას იმის შესახებ თუ როგორ აფასებდა მოქალაქე გაწეულ სოციალურ ხარჯებს: „ზედმეტად ბევრი“, „ზედმეტად მცირე“, „შესაბამისი“.

	რესპონდენტის რასა			
სოციალური დანახარჯები	თეთრკანიანი	შავკანიანი	ესპანური წარმოშობის	ჯამი
ზედმეტად მცირე	128	41	21	190
შესაბამისი	197	32	21	250
ზედმეტად ბევრი	185	31	19	235
ჯამი	510	104	61	675

ა. დაადგინეთ ცხრილის თავისუფლების ხარისხი
ბ. შეამოწმეთ არის თუ არა ორი ცვლადი რასა და NATFAE დამოუკიდებელი ($\alpha = 0.05$). გააკეთეთ შესაბამისი დასკვნა.

3. ჩვენი ანალიზი პირველი სავარჯიშოდან განვაგრძოთ, რასის როგორც საკონტროლო ცვლადის დამატებით. ქვემოთ მოტანილია ორგანზომილებიანი ცხრილები თეთრკანიანი და შავკანიანი რესპონდენტებისათვის.

თეთრკანიანი

ეშინია დამთა სეირნობა	მამაკაცები	ქალები	ჯამი
დიახ	59	160	219
არა	253	224	477
ჯამი	312	384	696

შავკანიანი

ეშინია დამთა სეირნობა	მამაკაცები	ქალები	ჯამი
დიახ	7	52	59
არა	41	38	79
ჯამი	48	90	138

ა. რომელ რასობრივ ჯგუფს აქვს უფრო მაღალი პროცენტული წილი „ღამის სეირნობის შიშის“ მიხედვით?

ბ. რასის მიუხედავად, არის თუ არა მეტი ალბათობა იმისა, რომ გამოკითხული ქალი რესპონდენტის უმრავლესობა დადებით პასუხს გასცემს მამკაცებთან შედარებით?

გ. თითოეული ცხრილისათვის, შეამოწმეთ ჰიპოთეზა იმის თაობაზე, რომ სქესი და ღამით სეირნობის შიში ერთმანეთისაგან დამოუკიდებელია ($\alpha = 0.01$). რა დასკვნას გამოიტანთ?

4. გავაგრძელოთ ჩვენი ანალიზი მე-2 სავარჯიშოდან, ამჯერად დავაგდინოთ კავშირი სოციალურ კლასსა (CLASS) და სოციალური კეთილდღეობის დანახარჯს შორის შეხედულებაზე.

ა. რისი ტოლია ამ ცხრილისათვის თავისუფლების ხარისხი?

ბ. გამოთვალეთ ხი-კვადრატი. ალფას 0.01-თან ტოლობის პირობებში უარყოფთ თუ არა ნულოვან ჰიპოთეზას? პასუხები დაასაბუთეთ.

	სოციალური კლასი				
სოციალურ დანახარჯებზე აზრი	დაბალი კლასი	მუშების კლასი	საშუალო კლასი	მაღალი კლასი	ჯამი
ზედმეტად მცირე	23	92	76	8	199
შესაბამისი	12	113	133	10	268
ზედმეტად ბევრი	12	127	99	7	245
ჯამი	47	332	308	25	712

5. იცვლება თუ არა მარიჰუანასადმი ხელმისაწვდომობა არასრულწლოვნებში იმ ქალაქების (დასახლებების) ზომიდან გამომდინარე, სადაც ისინი ცხოვრობენ? კვლევაში (Monitoring the Future 2008) დასმულ იქნა შეკითხვა თუ რამდენად მარტივი იყო არასრულწოვანთათვის მარიჰუანის შოვნა? (GWEED). მათი პასუხები დასახლებელის ზომის მიხედვითაა დალაგებული-პატარა ქალაქი, საშუალო ზომის ქალაქი, დიდი ქალაქი.

ა. გამოთვალეთ ქვემოთ მოყვანილი ცხრილისათვის ხი-კვადრატის მნიშვნელობა და თავისუფლების ხარისხი

ბ. ცხრილის რამდენ უჯრას აქვს მოსალდნელი სიხშირის მნიშვნელობა 5-ზე ნაკლები?
გ. თუ ალფას 0.01-ის ტოლად ავირჩევთ, რა დასკვნას გამოვიტანთ? როგორ ცვალებადობს მარიჰუანასადმი ხელმისაწვდომობა დასახლებების ზომის მიხედვით?

	დასახლების ზომა			
GWEED	პატარა ქალაქი	საშუალო ზომის ქალაქი	დიდი ქალაქი	ჯამი
ალბათ შეუძლებელი	24	7	8	39
ძალიან რთული	13	7	7	27
საკმაოდ რთული	31	11	11	53
საკმაოდ ადვილი	78	45	32	155
ძალიან ადვილი	196	98	104	398
ჯამი	342	168	162	672

6. არასრულწლოვანების სოციოლოგიური კვლევისას (MTF 2008) დასმულ იქნა შეკითხვები ზოგადი სოციალური საკითხების შესახებ მათ აზრზე. ქვემოთ მოყვანილ ორგანზომილებიან ცხრილში მოყვანილია არასრულწლოვანთა შეხედულებები რასის მიხედვით.

	რესპონდენტის რასა			
რამდენად ხშირად ფიქრობთ სოციალურ საკითხებზე	შავკანიანი	თეთრკანიანი	ესპანური წარმოშობის	ჯამი
არასოდეს	9	39	5	53
იშვიათად	38	147	22	207
ხანდახან	96	426	107	629
ხშირად	39	222	57	318
საკმაოდ ხშირად	5	48	12	65
ჯამი	187	882	203	1,272

შეამოწმეთ ნულოვანი ჰიპოთეზის მართებულობა რასისა და სოციალურ საკითხებზე ფიქრის სიხშირეს შორის სტატისტიკური დამოკიდებულების არარსებობის შესახებ. ($\alpha = 0.05$)
7. რა დამოკიდებულება არსებობს ეკლესიურ ცხოვრებასა და ჰომოსექსუალობის შესახებ აზრს შორის? მონაცემები GSS 2006-დან მოყვანილია ქვემოთ ცხრილებში, შემდეგი ცვალდების მიხედვით ATTEN (რამდენად ხშირად ესწრებით საეკლესიო მსახურებებს), HOMOSEX (შეხედულებები ჰომოსექსუალობის შესახებ).
ა. შეამოწმეთ ნულოვანი ჰიპოთეზა ჰომოსექსუალობის შესახებ შეხედულებებსა და საკლესიო მსახურებებზე დასწრების სიხშირეს შორის კავშირის არარსებობის შესახებ. ($\alpha = 0.05$)
ბ. როგორ აღწერთ კავშირს ჰომოსექსუალობის შესახებ შეხედულებებსა და საეკლესიო მსახურებებზე დასწრების სიხშირეს შორის?

	საეკლესიო მსახურებებზე დასწრება			
ჰომოსექსუალობის შესახებ შეხედულებები	არასოდეს	წელიწადში რამდენჯერმე	ყოველ კვირას	ჯამი
ყოველთვის მიუღებელია	163	101	266	530
საერთოდ მიუღებელი არ არის	215	65	57	337
ჯამი	378	166	323	876

8. რა დამოკიდებულება არსებობს რასასა და ცეცხლსასროლი იარაღის ფლობას შორის? GGSS 2006-ის მონაცემებზე დაყრდნობით, შეეამოწმოთ არის თუ არა დამოუკიდებელი რასა და ცეცხლსასროლი იარაღის ფლობა ($\alpha = 0.01$)

	რესპონდენტის რასა			
აქვს იარაღი სახლში	თეთრკანიანი	შავკანიანი	ესპანური წარმოშობის	ჯამი
დიახ	571	53	25	649
არა	861	214	151	1,226
ჯამი	1,432	267	176	1,875

9. ინმენმა და მეიესმა (1999) ასევე შეისწავლეს დამოკიდებულება სტუდენტის რასასა და პირველი თაობის კოლეჯის სტუდენტის სტატუსს შორის. კვლევის მონაცემებზე დაყრდნობით განსაზღვროთ არის თუ არა დაკავშირებული რასა და პირველი თაობის სტუდენტის სტატუსი ($\alpha = 0.01$)

	სტუდენტის რასა				
პირველი თაობის სტუდენტის სტატუსი	თეთრკანიანი	შავკანიანი	ესპანური წარმოშობის	აზიური წარმოშობის	ჯამი
პირველი თაობის	1,742	102	19	6	1,910
სხვა თაობის	2,392	119	25	22	2,603
ჯამი	4,134	221	44	28	4,513

10. არის თუ არა არასრულწლოვნების დამოკიდებულება ალკოჰოლის მოხმარებისადმი კავშირში მათ საშუალო აკადემიურ მოსწრებასთან (GPA). MTF 2008-ის მონაცემებზე დაყრდნობით შეეამოწმოთ არის თუ არა ალკოჰოლის მოხმარებისადმი დამოკიდებულება კავშირში საშუალო აკადემიურ მოსწრებასთან ($\alpha = 0.05$). როგორც აღწერთ კავშირს ამ ორ უკანასკნელ ცვლადს შორის?

	GPA-საშუალო აკადემიური მოსწრება			
დამოკიდებულება ალკოჰოლის მოხმარებისადმი	$GPA \leq 2.0$	$2.0 < GPA \leq 3.0$	$3.0 < GPA \leq 4.0$	ჯამი
დადებითი	98	396	517	1,910
უარყოფითი	17	69	107	193
მკვეთრად უარყოფითი	16	73	100	189
ჯამი	131	538	724	1,393

11. MTF 2008-ის მონაცემებზე დაყრდნობით გავაფართოვოთ ჩვენი ანალიზი შემდეგი ცვლადის დამატებით: არსებობს თუ არა დამოკიდებულება მარიჯუანის მოხმარებისადმი შეხედულებებსა და საშუალო აკადემიურ მოსწრებას (GPA) შორის ($\alpha = 0.05$). როგორ აღწერთ ამ ორ უკანასკნელ ცვლადს შორის სტატისტიკურ კავშირს? რით განსხვავდება ეს კავშირი წინა სავარჯიშოს შემთხვევაში მოყვანილი მაგალითისაგან?

დამოკიდებულება მარიჰუანის მოხმარებისადმი	GPA-საშუალო აკადემიური მოსწრება			
	GPA≤2.0	2.0<GPA ≤ 3.0	3.0<GPA≤4.0	ჯამი
დადებითი	76	261	293	630
უარყოფითი	29	105	174	308
მკვეთრად უარყოფითი	27	166	245	438
ჯამი	132	532	712	1,376

12. არის თუ არა დამოკიდებული არასრულწოვანთა შეხედულებები კოკაინის მოხმარებაზე მათსავე საშუალო აკადემიურ მოსწრებაზე? MTF 2008-ის მონაცემების გამოყენებით შეამოწმეთ ჰიპოთეზა იმის, თაობაზე რომ კოკაინის მოხარებისადმი დამოკიდებულება არ არის კავშირში საშუალო აკადემიურ მოსწრებასთან ($\alpha = 0.01$). როგორ აღწერთ დამოკიდებულებას ამ ორ ცვლადს შორის? რით განსხვავდება ეს კავშირი წინა მე-10 და მე-11 სავარჯიშოების შემთხვევაში მოყვანილი მაგალითებისაგან?

დამოკიდებულება კოკაინის მოხმარებისადმი	GPA-საშუალო აკადემიური მოსწრება			
	GPA≤2.0	2.0<GPA ≤ 3.0	3.0<GPA≤4.0	ჯამი
დადებითი	28	63	59	150
უარყოფითი	27	89	100	216
მკვეთრად უარყოფითი	76	388	564	1,028
ჯამი	131	540	723	1,394

თავი 12

ნომინალურ და ორდინალურ ცვლადებს შორის კავშირის გაზომვა

თავის სწავლების მიზნები

- შეცდომის პროპორციული შემცირების (PRE) კონცეფცია
- ლამბდას მნიშვნელობის გამოთვლა და ინტერპრეტირება: ნომინალურ ცვალდებს შორის კავშირის შეფასება
- გამის და კენდალის ტაუ ხ-ს გამოთვლა და ინტერპრეტირება: ორდინალურ ცვლადებს შორის კავშირის შეფასება
- კრამერის V-ს ინტერპრეტირება: ხი-კვადრატთან დაკავშირებული სტატისტიკური კავშირის შეფასების კრიტერიუმი

შეცდომის პროპორციული შემცირება: მოკლე მიმოხილვა

მე-10 თავში გავეცანით ჯვარედინ გამოთვლებს, სადაც ცვლადებს შორის სტატისტიკური კავშირის შეფასება ეფუძნებოდა პროცენტული განაწილებების შედარებებს. მე-11 თავში განვიხილეთ დასკვნითი სტატისტიკის მეთოდი: ხი-კვდარატ კრიტერიუმის გამოყენება ცვლადებს შორის სტატისტიკური კავშირის შეფასებისათვის. 12-ე და 13-ე თავებში განვიხილავთ ცვლად სიდიდეებს შორის სტატისტიკური კავშირის შეფასების მაჩვენებლებს, როგორც ნომინალური ასვე რიგობითი მონაცემებისათვის. ეს მაჩვენებლები საშუალებას გვაძლევს ცვლად სიდიდეებს შორის არსებული სტატისტიკური დამოკიდებულების სახე შევაფასოთ ერთიანი მაჩვენებლით. ხი-კვადრატისაგან განსხვავებით ეს მაჩვენებლები ასახავს ცვლად სიდიდეებს შორის კავშირის სიძლიერეს და ცალკეულ შემთხვევებში ამ კავშირის მიმართულებას (დადებითს ან უარყოფითს). ასევე, მათზე დაყრდნობით

შესაძლებელია შეფასდეს თუ რამდენად რეალისტურია დამოკიდებული ცვლადის მნიშვნელობების პროგნოზირება დამოუკიდებელი ცვლადების მიხედვით.

ამ თავში განვიხილავთ კავშირის ოთხ მაჩვენებელს: ლამბდა (ნომინალური ცვლადებისათვის), გამა და კენდალის ტაუ b (ორდინალური ცვლადებისათვის) და კრამერის V (ხი-კვადრატთან დაკავშირებული მაჩვენებელი). ასევე, მე-13 თავში გაეცნობით კორელაციის პირსონის კოეფიციენტს, რომელიც გამოიყენება ინტერვალურ მონაცემებს შორის ბივარიანტული კავშირის შეფასებისათვის.

კავშირის საზომი – ერთიანი შემაჯამებელი მაჩვენებელი, რომელიც ასახავს ცვლადებს შორის კავშირის სიძლიერეს, აჩვენებს დამოკიდებული ცვლადების პროგნოზირების შესაძლებლობებს დამოუკიდებელი ცვლადების მიხედვით და ხშირ შემთხვევაში იძლევა კავშირის მიმართულების შეფასების საშუალებას.

ყველა სტატისტიკური კავშირის შეფასების საზომი, რომელიც განხილულია მე-12 და მე-13 თავებში ეფუძნება შეცდომის პროპორციული შემცირების კონცეფციას, შემოკლებითი აღვნიშნით PRE. ამ კონცეფციის მიხედვით, ორი ცვლადი სიდიდე ერთმანეთთან სტატისტიკურად დაკავშირებულია თუ ერთი ცვლადის მნიშვნელობები შესაძლებელია გამოყენებულ იქნას მეორე ცვლადის მნიშვნელობების პროგნოზირებაში.

შეცდომის პროპორციული შემცირების (PRE) – კონცეფცია ეფუძნება სტატისტიკური კავშირის შეფასების რამდენიმე საზომის განმარტებას და ინტერპრეტაციას. PRE-ის მაჩვენებლები მიიღება იმ შეცდომების შედარებით, რაც ერთის მხრივ მიიღება დამოკიდებული ცვლადების მნიშვნელობების პროგნოზირებისას დამოუკიდებელი ცვლადებისაგან განცალკევებით, და მეორეს მხრივ იმ შეცდომებთან, რაც მიიღება დამოკიდებული ცვლადების მნიშვნელობების პროგნოზირებით დამოუკიდებელი ცვლადების მხედველობაში მიღებით. ე.ი. ეს კონცეფცია დამოკიდებული ცვლადების პროგნოზირების შედეგებს ეფუძნება.

ცხრილ 12.1-ში მოცემული მონაცემები დაგვეხმარება შეცდომის პროპორციული შემცირების (PRE) კონცეფციის უკეთ გაგებაში. ზოგადი სოციოლოგიური კვლევის შედეგების გამოყენებით (GSS 2008) ცხრილ 12.1-ში ნაჩვენებია საშუალო სიძლიერის კავშირი დამოუკიდებელ ცვლად „განათლება"-სა და დამოკიდებულ ცვლად „აბორტის მიმართ თანხმობა, არაფერტილური ასაკის ღარიბი ქალბატონების აზრს" შორის. ცხრილის მონაცემების მიხედვით რესპონდენტების 64.3%, რომელთაც აქვთ საშუალო ან წინა საფეხურის განათლება არიან აბორ-

ტის წინააღმდეგი, იგივე კატეგორიაში მოხვდნენ უმაღლესი განათლების მქონეთა – 44.4%.

შეცდომის პროპორციული შემცირების (PRE) კონცეფციის ძირითადი ფორმულა შემდეგია1:

$$PRE = \frac{E_1 - E_2}{E_1} \qquad (12.1)$$

სადაც,

E_1 დამოკიდებული ცვლადის პროგნოზირების შეცდომაა, დამოუკიდებელი ცვლადის გამოტოვებით (პროგნოზირება 1).

E_2 დამოკიდებული ცვლადის პროგნიზირების შეცდომაა, დამოუკიდებელი ცვლადის მიხედვით (პროგნოზირება 2).

ცხრილი 12.1 აბორტის მხარდაჭერა განათლების დონის მიხედვით

	განათლების დონე		
აბორტის მხარდაჭერა	საშუალო ან უფრო დაბალი განათლება	უმაღლესი ან უმაღლესის შემდგომი განათლება	ჯამი
დიახ	399 64%	156 44.4%	555 57.1%
არა	222 35.7%	195 55.6%	417 42.9%
სულ	621 100%	351 100%	972 100%

წყარო: General Social Survey, 2008.

შეცდომის პროპორციული შემცირების ყველა მაჩვენებელი ეფუძნება პროგნოზირების იმ შეცდომების შედარებას, რაც მიიღება პროგნოზირების ორი მეთოდის გამოყენების შედეგად. დავუშვათ, გვსურს ვივარაუდოთ რესპონდენტების მოსალოდნელი პასუხები აბორტის შესახებ, მაგრამ არ გვაქვს მონაცემები მათი განათლების დონის შესახებ. ცხრილ 12.1-ის სტრიქონების ჯამის მიხედვით შეგვიძლია ვივარაუდოთ, რომ შერჩევის ყველა რესპონდენტი აბორტის წინააღმდეგია. ამ დაშვების გაკეთებით ჩვენ მივიღებთ 417 შეცდომას, რადგან ფაქტობრივად შერჩევაში 555 რესპონდენტი აბორტის წინააღმდეგია, ხოლო 417 აბორტს ემხრობა. (იხილეთ სრიქონების ჯამი 12.1 ცხრილში). შესაბამისად:

$$E_1 = 972 - 555 = 417$$

როგორ შეგვიძლია გავზარდოთ ჩვენი პროგნოზის სიზუსტე თუ გვეკნება ინფორმაცია ჩვენი რესპონდენტების განათლების დონის შესახებ? ჩვენი ახალი პროგნოზისათვის გამოვიყენოთ შემდეგი მიდგომა: თუ რესპონდენტს აქვს საშუალო ან უფრო დაბალი განათლება ის სავარაუდოდ აბორტის წინააღმდეგი იქნება, თუ რესპონდენტს საშუალოზე მაღალი განათლება აქვს ის აბორტის მომხრე იქნება; ეს მიდგომა რაციონალურია, რადგან უკვე ვიცით ცხრილ. 12.1 დან, რომ რესპონდენტები დაბალი განათლების დონით უფრო მეტად არიან აბორტის წინააღმდეგნი, ვიდრე უფრო მეტი განათლების მქონე რესპონდენტები. ამ მიდგომის მიხედვით პროგნოზირებისას დავუშვებთ 378 შეცდომას, ნაცვლად 417-ისა, რადგან 156 რესპონდენტი, რომელთაც აქვთ უმაღლესი განათლება არიან აბორტის წინააღმდეგნი, ხოლო უმაღლესი განათლების მქონე 222 რესპონდენტი არის აბორტის მომხრე (156+222=378), შესაბამისად:

$$E_1 = 156 + 222 = 378$$

ჩვენმა პირველმა პროგნოზირების მეთოდმა, დამოუკიდებელი ცვლადის (განათლების) იგნორირებით, მოგვიტანა 417 შეცდომა. პროგნოზირების მეორე მეთოდმა, დამოუკიდებელი ცვლადის შესახებ ინფორმაციის მხედველობაში მიღებით, მოგვიტანა 378 შეცდომა. თუ ცვლადებს შორის კავშირი არსებობს პროგნოზირების მეორე მეთოდი ნაკლებ შეცდომებს მოიტანს ვიდრე პირველი. რაც უფრო ძლიერია ცვლადებს შორის კავშირი მით უფრო მეტი იქნება პროგნოზირების შეცდომის შემცირების რიცხვი.

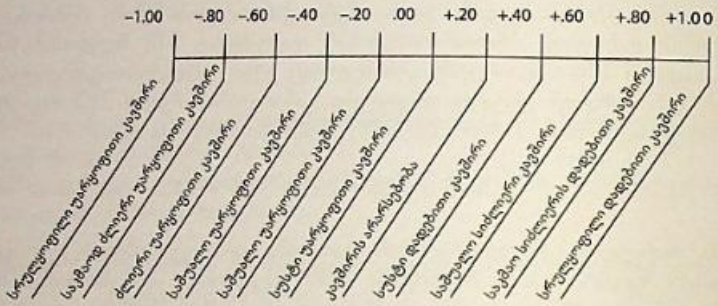
ცხრილ 12.1-ისათვის გამოთვალოთ შეცდომის პროპორციული შეცდომა 12.1 ფორმულის გამოყენებით. შეცდომის პროპორციული შემცირება დამოუკიდებელი ცვლადის „განთლების დონის“ გამოყენებით შემდეგია:

$$PRE = \frac{417 - 378}{417} = 0.09$$

კავშირის PRE მაჩვენებელი იღებს მნიშვნელობას 0.0-დან +/- 1.0-მდე. შეცდომის პროპორციული შეცდომის ნულთან ტოლობა ნიშნავს, რომ ცვალდებს შორის კავშირი არ არსებობს; დამოუკიდებელი ცვლადის შესახებ მონაცემები არ შეცვლის დამოკიდებული ცვლადის მნიშვნელობის პროგნოზირების შედეგებს. PRE-ის +/- 1.0 ტოლობა ნიშნავს ცვლადებს შორის სრულყოფილ დადებით ან უარყოფით კავშირს, შესაძლებელია დამოკიდებული ცვლადის მნიშვნელობის უშეცდომოდ პროგნოზირება დამოუკიდებელ ცვლადის მნიშვნელობებზე დაყრდნობით. PRE-ის შუალედური მაჩვენებლები ასახავენ ცვლადებს შორის კავშირის სიძლიერეს, რაც ამასთანავე გვაძლევს ინფორმაციას თითოეული ცვლადის გამოყენების სარგებლიანობაზე მეორე ცვლადის მნიშვნელობის პროგნოზირებისას. რაც უფრო მეტად სცილდება კავშირის საზომის მნიშვნელობა 0.00-ს ორივე მიმართულებით მით უფრო ძლიერია ცვლადებს შორის კავშირი. PRE მაჩვენებელი შესაძლებელია გამრავლდეს 100-ზე პროცენტული მაჩვენებლების მისაღებად.

დეტალების განხილვა 12.1
რას ნიშნავს ძლიერი, სუსტი სტატისტიკური კავშირი? ინტერპრეტირების წესები

სტატისტიკის შესწავლისას, რაც უფრო მეტ სტატისტიკური კავშირის მაჩვენებელს გაეცნობით თქვენი ცოდნა სტატისტიკური კავშირის შესახებ უფრო დაიხვეწება. ცოდნის მისაღებ დონეზე დაზუსტებამდე არსებობს რამდენიმე სახელმძღვანელო პრინციპი, რომელიც დაგეხმარებათ უკეთ აღიქვათ რა არის სუსტი და ძლიერი სტატისტიკური კავშირი.



დაიმახსოვრეთ, რომ ზემოთ მოყვანილი მაჩვენებლები შედარებით არაზუსტი დაჯგუფებებია. მაჩვენებლების კონკრეტული ინტერპრეტირება დამოკიდებულია თავად კვლევის კონტექსტზე. მაგალითად, +0.30 ერთი კვლევის შემთხვევაში შეიძლება ნიშნავდეს მცირედ განსხვავებას +0.30-თან მეორე კვლევის შემთხვევაში. მაგრამ ნული ყოველთვის ცვალდებს შორის კავშირის არარსებობას ნიშნავს.

PRE-ის 0.09-სთან ტოლობა ნიშნავს, რომ არსებობს სუსტი კავშირი რესპონდენტის განათლებასა და მის აბორტის შესახებ შეხედულებას შორის. (მიმართეთ ზემოთ 12.1 დეტალების განხილვას). შეცდომის პროპორციული შემცირების მაჩვენებლის 0.09-თან ტოლობა ნიშნავს, რომ ჩვენ პროგნოზირების შედეგები გავაუმჯობესეთ მხოლოდ 9%-ით (0.09x100=9%) ცვლად „განათლების დონის“ გამოყენებით.

ლამბდა: სტატისტიკური კავშირის გაზომვა ნომინალური ცვლადებისათვის

2008 წლის წინასაარჩევნო კამპანია დემოკრატების საპრეზიდენტო კანდიდატის დასახლებელისათვის ისტორიული მომენტი იყო. პირველად ამერიკის შეერთებული შტატების ისტორიაში ერთმანეთს ეჯიბრებოდა შავკანიანი სენატორი ბარაკ ობამა და თეთრკანიანი სენატორი ჰილარი კლინტონი. მათ ქონდათ რეალური შანსები წარედგინათ დემოკრატიული პარტია საპრეზიდენტო არჩევნებზე 2008 წელს. რადგან სენატორებს ქონდათ ბევრი მსგავსი

შეხედულება ეკონომიკურ, გლობალურ და სოციალურ საკითხებზე, ბევრი ანალიტიკოსი დაინტერესდა შეესწავლათ არსებობდა თუ არა განსხვავება კანდიდატების მიმართ შეხედულებებში რასისა და სქესის მიხედვით. ბევრს მიაჩნდა, რომ უფრო მეტად მოსალოდნელი იყო, რომ ქალების უმრავლესობა რასის მიუხედავად ხმას მისცემდა სენატორ კლინტონს გენდერული ფაქტორის გამო, ხოლო სხვებს მიაჩნდა რომ შავკანიანი მოსახლეობა სქესის მიუხედავად უფრო მეტად სენატორ ობამას მისცემდა ხმას, საერთო რასობრივი ფაქტორის გამო.

ამ ისტორიული მოვლენისათვის განვიხილოთ კავშირი ამომრჩევლის რასასა (დამოუკიდებელი ცვლადი) და დემოკრატი კანდიდატის მხარდაჭერას შორის (დამოკიდებული ცვლადი). ცხრილ 12.2-ში ნაჩვენებია ქვეყნის მასშტაბით დემოკრატი ამომრჩევლების გამოკითხვის შედეგები (Gallup, Inc. 2008 1 თებერვლის კვირა). 2,932 დემოკრატი ამომრჩეველი კლასიფიცირებულ იქნა მათი რასისა და კონკრეტული საპრეზიდენტო კანდიდატის მხარდაჭერის მიხედვით. რესპონდენტის რასა დამოუკიდებელი ცვლადია, ხოლო კონკრეტული დემოკრატი კანდიდატის მხარდაჭერა დამოკიდებული ცვლადი.

ცხრილი 12.2 დემოკრატი კანდიდატის მხარდაჭერა ამომრჩევლის რასის მიხედვით, 2008 წელი

2008 წლის დემოკრატი კანდიდატის მხარდაჭერა	რასა		
	შავკანიანები	თეთრკანიანები	სულის ჯამი
ბარაკ ობამა	330	966	1,296
ჰილარი კლინტონი	117	1,519	1,636
სულ	447	2,485	2,932

წყარო: Gallup Poll Daily, Febraury 1-9, 2008.

რადგან რასა და ამომრჩევლის ხმა ნომინალური ცვლადია, სტატისტიკური კავშირის შეფასების შესაბამისი მაჩვენებელი უნდა გამოვიყენოთ. ეს უკანასკნელი უნდა დაგვეხმაროს იმის განსაზღვრაში თუ რამდენად ძლიერია კავშირის ცვლადებს „რასასა“ და „არჩეულ საპრეზიდენტო კანდიდატს“ შორის. სწორედ ასეთი მაჩვენებელია ლამბდა.

ლამბდა კავშირის ასიმეტრიული საზომი; ლამბდა გამოიყენება ნომინალური ცვალდებისათვის და იღებს მნიშვნელობებს 0.0-დან 1.0-მდე. იგი გვიჩვენებს დამოკიდებულ და დამოუკიდებელ ცვალდებს შორის კავშირის სიძლიერეს.

ლამბდას გამოთვლის მეთოდი

დააკვირდით ცხრილ 12.2-ის სტრიქონების ჯამს, რომელიც გვიჩვენებს ცვლად „2008 წლის დემოკრატი საპრეზიდენტო კანდიდატის არჩევის“ განაწილებას. თუ მოვახდენდით საპრეზიდენტო კანდიდატის გამარჯვების პროგნოზირებას, ყველაზე შესაბამისი შედეგი სენატორი ჰილარი კლინტონის უპირატესობის გამოკვეთა იქნებოდა. ეს შედეგი გარკვეულ შეცდომას მოგვიტანდა. ამ მიდგომით დაშვებული შეცდომის რიცხვი 1,296-ია, რადგან 2,392-დან მხოლოდ 1,636-მა მისცა ხმა კლინტონს (2,392 – 1,636 = 1,296).

ამჯერად, ცხრილ 12.2-ის მონაცემებზე დაყრდნობით განვსაზღვროთ 2008 წლის საპრეზიდენტო დემოკრატი კანდიდატის არჩევნების შედეგები რასის მიხედვით. ამ შემთხვევაშიც კვლავ იგივე მიდგომა გამოვიყენოთ განცალკევებულად, თეთრკანიანი და შავკანიანი კანდიდატის მისამართით. მოდა შავკანიანი დემოკრატი ამომრჩევლისათვის არის 330 ხმა, შესაბამისად შეგვიძლია ვივარაუდოთ რომ ყველა შავკანიანი დემოკრატი ობამას მხარდამჭერია. ამ პროგნოზირების მეთოდით ვუშვებთ 117 შეცდომას, რადგან 117 შავკანაინი დემოკრატი 447-დან მხარს უჭერს კლინტონს (447–330=17). შემდეგ განვიხილოთ თეთრკანიანი დემოკრატების ჯგუფი. ამ ჯგუფისათვის მოდა არის „კლინტონი“. ეს უკანასკნელი იქნება ამ ჯგუფისათვის ჩვენი პროგნოზი. პროგნოზირების მეთოდი მოგვიტანს 966 შეცდომას (2,485–1,519=966). შესაბამისად, შეცდომების მთლიანი ჯამი არის 117+966 ან 1,083.

ეს შედეგები გავაერთიანოთ და ჩამოვაყალიბოთ ლამბდას გამოთვლის უფრო ზოგადი მეთოდი.

1. ვიპოვოთ E_1 – პროგნოზის შეცდომა, როცა მხედველობაში არ მიიღება დამოუკიდებელი ცვლადი. E_1-ის გამოთვლისათვის, ვიპოვოთ დამოკიდებული ცვლადის მოდა და გამოვაკლოთ მისი სიხშირე საერთო სიხშირეს N-ს. 12.2 ცხრილის მიხედვით:

$$E_1 = N - \text{მოდალური სიხშირე}$$

$$E_1 = 2,932 - 1,636 = 1,296$$

2. ვიპოვოთ E_2-პროგნოზის შეცდომა დამოუკიდებელ ცვლადზე დაყრდნობით. E_2-ის გამოთვლისათვის ვიპოვოთ მოდალური სიხშირე თითოეული კატეგორიისათვის, გამოვაკლოთ ეს სიდიდე თითოელი კატეგორიის სიხშირეთა ჯამს და საბოლოოდ შევკრიბოთ ყველა შეცდომის რიცხვი. ცხრილ 12.2-ის მიხედვით:

$$\text{შავკანიანი დემოკრატები} = 447 - 330 = 117$$

$$\text{თეთრკანიანი დემკარტები} = 2,485 - 1,519 = 966$$

$$E_2 = 117 + 966 = 1,083$$

3. გამოვთვალოთ ლამბდა 12.1 ფორმულის მიხედვით:

$$\text{ლამბდა} = \frac{E_1 - E_2}{E_1} = \frac{1,296 - 1,083}{1,296} = 0.16$$

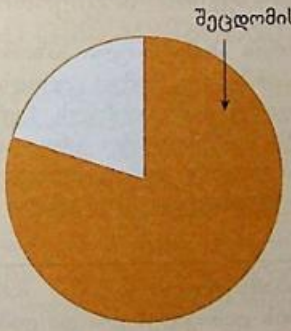
ლამბდას მნიშვნელობები შესაძლებელია ცვალებადობდეს 0.0-დან 1.0-მდე. ლამბდას ნოლთან ტოლობა ნიშნავს, რომ დამოუკიდებელი ცვლადის მნიშვნელობების გათვალისწინება ვერ შეცვლის დამოკიდებული ცვლადის პროგნოზირების შედეგებს. ლამბდას 1-თან ტოლობთან ნიშნავს, რომ დამოუკიდებელი ცვლადის გამოყენება შესაძლებელია დამოკიდებული ცვლადის მნიშვნელობის შეცდომის გარეშე პროგნოზირებისათვის. ჩვენს მაგალითში ლამბდას 0.16-იანი მნიშვნელობა ნაკლებია 1-სა და 0-ს შორის არსებულ ერთ მეოთხედზე, რაც ნიშნავს, რომ ამ რესპონდენტებისათვის რასა და საპრეზიდენტო კანდიდატის მხრდაჭერა მხოლოდ მცირედით უკავშირდება ერთმანეთს.

შეცდომის პროპორციული შემცირების მნიშვნელობის 100-ზე გამრავლებით მივიღებთ ამ მაჩვენებლის პროცენტულ ფორმას. ამ უკანასკნელის ინტერპრეტირება შემდეგია: რესპონდენტის რასის შესახებ ინფორმაციის გამოყენება 2008 წლის დემოკრატი საპრეზიდენტო კანდიდატის გამარჯვების პროგნოზირებისათვის ამცირებს პროგნოზის შეცდომას 16%-ით (0.16x100=16%). სხვა სიტყვებით, რომ ვთქვათ თუ პროგნოზისას გამოვიყენებთ რესპონდენტის რასის შესახებ მონაცემებს ჩვენ პროგნოზის შეცდომას ვამცირებთ 100-დან 16 ერთეულით ან 16 პროცენტით. (0.16x100).

ახსენით თუ რატომ არ იღებს ლამბდა უარყოფით მნიშვნელობებს? შეცდომის პროპორციული შემცირების კონცეფციის გაგება მნიშვნელოვანია სტატისტიკაში (მას ჩვენ კვლავ დაუბრუნდებით მე-13 თავში). ამ კონცეფციის არსის უკეთ გაგებისათვის დაკვირდით შემდეგ მონაცემებს და მათ გრაფიკულ ილუსტრაციას დიაგრამების მეშვეობით.

ძლიერი კავშირის არსებობა პირობითი მონაცემების მიხედვით:
სტრესის დონე და ძაღლის ყოლა, ლამბდა — 0.80

სტრესი \ ძაღლის ყოლა	დიახ	არა	
დიახ	5	45	50
არა	45	5	50
	50	50	100



ძლიერი კავშირის არსებობა პირობითი მონაცემების მიხედვით:
სტრესის დონე და ძაღლის ყოლა, ლამბდა — 0.40

ძაღლის ყოლა

სტრესი	დიახ	არა	
დიახ	15	35	50
არა	35	15	50
	50	50	100



ძლიერი კავშირის არსებობა პირობითი მონაცემების მიხედვით:
სტრესის დონე და ძაღლის ყოლა, ლამბდა — 0.12

ძაღლის ყოლა

სტრესი	დიახ	არა	
დიახ	22	28	50
არა	28	22	50
	50	50	100



ლამბდას გამოთვლის ზოგიერთი ასპექტი

ლამბდა კავშირის ასიმეტრიული საზომია. ეს ნიშნავს, რომ მისი მნიშვნელობა დამოკიდებულია იმაზე, თუ რომელ ცვლადს შევარჩევთ როგორც დამოუკიდებელს და რომელს როგორც დამოკიდებულს. ჩვენს მაგალითში, 2008 წლის გამარჯვებული დემოკრატი საპრეზიდენტო კანდიდატი ავირჩიეთ, როგორც დამოკიდებული ცვლადი და რასა როგორც დამოუკიდებელი ცვლადი. თუ ამ უკანასკნელ არჩევანს საწინააღმდეგოდ მოვახდენდით, მივიღებდით ლამბდას განსხვავებულ მნიშვნელობას.

კავშირის ასიმეტრიული საზომი მაჩვენებელი, რომლის მნიშვნელობა ცვალებადობს დამოკიდებული და დამოუკიდებელი ცვლადის შერჩევის მიხედვით.

ლამბდას გამოთვლის მეთოდი იგივე რჩება ცვლადების როლების შეცვლის შედეგადაც. მაგრამ გარკვეული შედომის წყარო შეიძლება იყოს ლამბდას გამოთვლა, როცა დამოუკიდებელი ცვლადი ცხრილის სტრიქონების ნაცვლად მოცემულია სვეტებში. შეცდომის თავიდან აცილებისათვის, ეცადეთ დამოუკიდებელი ცვლადი ცხრილის სვეტებში შეიტანოთ. ამასთან, დაიმახსოვრეთ, რომ მიუხედავად ლამბდას გამოთვლისათვის გამოყენებული მეთოდისა, საბოლოოდ ცვლადის დამოუკიდებლად ან დამოკიდებულად ჩათვლა თავად მკვლევარის მიერ დასმული საკვლევი საკითხიდან გამომდინარეობს.

ლამბდა ყოველთვის ნოლის ტოლია, როცა ყოველი კატეგორიის მიხედვით თითოეული დამოუკიდებელი ცვლადის მოდა ტოლია დამოკიდებული ცვლადის მოდის. ლამბდას ინტერპრეტირების საკითხი პრობლემატურია, როცა მისი მნიშვნელობა ნოლს უტოლდება, მაგრამ ამასთანავე კავშირის სხვა საზომები მეტყველებს რომ ცვლადებს შორის კავშირი არსებობს. ამ პოტენციური პრობლემის თავიდან აცილებისათვის უნდა განვიხილოთ ცხრილის მონაცემების მიხედვით არსებული პროცენტული სხვაობები. თუ პროცენტული სხვაობები მცირეა (ჩვეულებრივ 5% ან ნაკლები), ლამბდა არის კავშირის ადეკვატური საზომი არსებული მონაცემთა ცხრილისათვის. მაგრამ თუ პროცენტული სხვაობები დიდია, რაც ნიშნავს რომ ცვლადები ერთმანეთთან დაკავშირებულია, ლამბდა კავშირის ცუდი საზომია. ასეთ შემთხვევებში კავშირის გაზომვის ალტერნატიული მაჩვენებელი უნდა ავირჩიოთ.

ცხრილ 12.2-ისათვის ხელახლა გამოთვალეთ ლამბდას მნიშვნელობა, ამჯერად 2008 წლის დემოკრატი საპრეზიდენტო კანდიდატი აირჩიეთ, როგორც დამოუკიდებელი ცვლადი, ხოლო რასა, როგორც დამოკიდებული ცვლადი. როგორ შეიცვალა ლამბდას მნიშვნელობა?

✓ *ცოდნის შემოწმება*

კრამერის V: ხი-კვადრატთან დაკავშირებული სტატისტიკური კავშირის საზომი ნომინალური ცვლადებისათვის

კრამერის V სტატისტიკური კავშირის ალტერნატიული საზომია, რომელიც შესაძლებელია გამოყენებულ იქნას ნომინალური ცვლადებისათვის. იგი ეფუძნება ხი-კვადრატის (განხილულია მე-11 თავში) მნიშვნელობას და ცვალებადობს 0-დან 1-მდე. ნოლი აღნიშნავს კავშირის არარსებობას, ხოლო 1 სრულყოფილ კავშირს. რადგან ეს მაჩვენებელი არ იღებს უარყოფით მნიშვნელობას, იგი არ აჩვენებს კავშირის მიმართულებას. აქედან გამომდინარე კრამერის V გარკვეულწილად შეზღუდული მაჩვენებელია, მისი მნიშვნელობის ინტერპრეტირება არ ხდება შეცდომის შემცირების პროპორციული კონცეფციის მიხედვით. კრამერის V გამოითვლება შემდეგი ფორმულით:

$$\text{კრამერის } V = \sqrt{\frac{\chi^2}{N \times m}} \qquad (12.2)$$

სადაც, m = უმცირესია $(r - 1)$ ან $(c - 1)$-დან.

მე-11 თავში ჩვენ შევამოწმეთ ჰიპოთეზა პოპულაციაში განათლებასა და ჯანმრთელობის მდგომარეობის შეფასებას შორის კავშირის შესახებ (ცხრილები 11.7 და 11.8). ანალიზის შედეგად მივიღეთ ხი-კვადრატის 53.96-იანი მნიშვნელობა, რის საფუძვლებზეც უარვყავით ნულოვანი ჰიპოთეზა. საბოლოოდ დავასკვენით, რომ პოპულაციაში ჯანმრთელობის დონის შეფასება ცვალებადობს განათლების დონის მიხედვით.

შეგვიძლია გამოვიყენოთ კრამერის V იმისათვის, რომ შევაფასოთ ზემოთა აბზაცში ნახსენები კავშირის სიძლიერე (ფორმულა 12.2).

$$\text{კრამერის } V = \sqrt{\frac{\chi^2}{N \times m}} = \sqrt{\frac{53.96}{844 \times 2}} = \sqrt{0.032} = 0.18$$

კრამერის V-ის 0.18 მნიშვნელობა მეტყველებს, რომ განათლების დონესა და ჯანმრთელობის შეფასებას შორის კავშირი სუსტია.

გამა: კავშირის საზომი ორდინალური ცვლადებისათვის

ამ ქვეთავში განვიხილეთ ორ ორდინალურ ცვლადს შორის კავშირის გაზომვასა და ინტერპრეტირებას. თუ ორ ცვლადს შორის არსებობს კავშირი, ერთი ცვლადის მნიშვნელობების ცოდნა გვეხმარება მეორე ცვლადის მნიშვნელობების უკეთ განსაზღვრაში.

სანამ უშუალოდ ორდინალურ ცვლადებს შორის კავშირის გაზომვის მეთოდის განხილვაზე გადავიდოდეთ თავად ორდინალური ცვლადის განსაზღვრება განვიხილოთ. ორდინალური ცვლადი გულისხმობს გარკვეული კატეგორიების მიხედვით დაკვირვების მონაცემისათვის არჩეული რიცხვითი მნიშვნელობების მინიჭებას. ორდინალური ცვლადის ერთი მაგალითი სოციალური კლასია. ინდივიდების კლასიფიკაცის შესაძლებელია მოვახდინოთ შემდეგი კატეგორიების მიხედვით: „მაღალი კლასი“, „საშუალო“ ან „მუშათა კლასი“. პიროვნების მიკუთვნება კონკრეტულ კლასზე დამოკიდებულია ჩვენს მიერ არჩეულ ლოგიკურ კრიტერიუმებზე.

ორდინალური ცვლადები საკმაოდ ხშირად გამოიყენება სოციალურ მეცნიერებებში. ზოგადი სოციოლოგიური კვლევა (GSS) შეიცავს ბევრ კითხვას, რომლებზეც პასუხი ორდინალური სკალით ჩაიწერება, მაგალითად: „ძალიან ხშირად“, „საკმაოდ ხშირად“, „იშვიათად“ ან „თითქმის არასდროს“.

ორდინალურ ცვლადებს შორის კავშირის ანალიზი: განათლება და ფინანსური მდგომარეობით კმაყოფილება

განვიხილოთ კვლევის მაგალითი, რომელშიც გვხვდება ორ ორდინალურ ცვლადს შორის კავშირის ანალიზი. კვლევის ჰიპოთეზა შემდეგია: რაც უფრო მაღალია პირის განათლების დონე, მით უფრო მეტია მისი კმაყოფილება თავისი ფინანსური მდგომარეობით.

ცხრილი 12.3 ფინანსური მდგომარეობით კმაყოფილება განათლების დონის მიხედვით

კმაყოფილება ფინანსური მდგომარეობით	განათლება (X)		
	საშუალო ან უმაღლესი განათლება	საშუალო განათლებაზე დაბალი განათლება	ჯამი
	a	*b*	
კმაყოფილი	71.5% (1,194)	56.8% (193)	69% (1,387)
	c	*d*	
უკმაყოფილო	28.5% (477)	43.2% (147)	31% (624)
სულ (*N*)	100% (1,671)	100% (340)	100% (2,011)

წყარო: General Social Survey, 2008.

შენიშვნა: ცხრილის უჯრები დანომრილია ა-დან ბ-მდე.

კვლევის ჰიპოთეზის შემოწმებისათვის ავირჩიოთ ორი ცვლადი (GSS 2008). ცვლადი განათლება (EDUC) კოდირებულ იქნა, როგორც დიქოტომიური ცვლადი იმ რესპონდენტების პასუხების მიხედვით, რომლებმაც აღნიშნეს, რომ ჭონდათ 11 ან ნაკლები წლების განათლება ერთ კატეგორიაში, ხოლო 12 ან მეტი წლის განათლების მქონენი სხვა კატეგორიაში. მეორე ცვლადი „კმაყოფილება ფინანსური მდგომაროებით“ ასევე ჩაწერილ იქნა, როგორც დიქოტომიური ცვლადი შემდეგი პასუხების არჩევანით: „კმაყოფილი“ და „მეტ-ნაკლებად კმაყოფილი“ მიკუთვნებულ იქნა კატეგორიას „კმაყოფილი“, ხოლო პასუხი „საერთოდ უკამყოფილო“ მიკუთვნებულ იქნა კატეგორიას „უკმაყოფილო“.

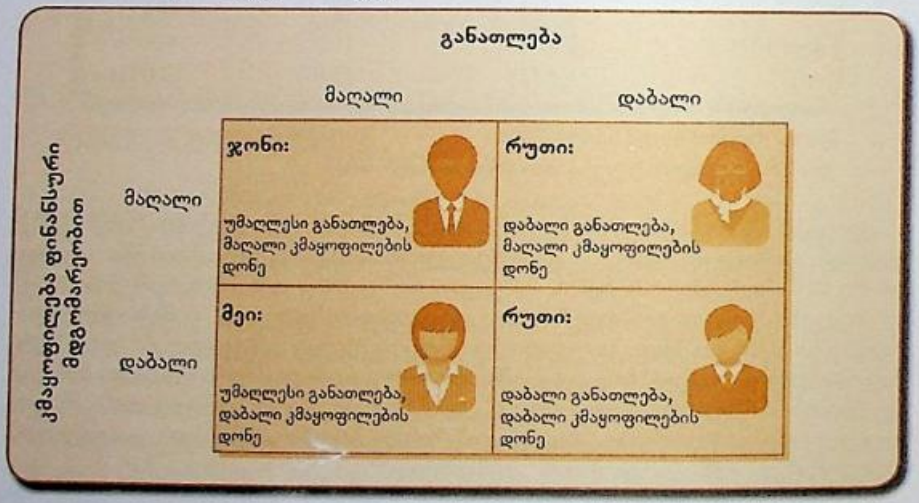
ცხრილ 12.3-ში ნაჩვენებია ორი ცვლადის ჯვარედინი მონაცემები, განათლება როგორც დამოუკიდებელი ცვლადი და კმაყოფილება, როგორც დამოკიდებული ცვლადი. მონაცემების მიხედვით საშუალო და უმაღლესი განათლების არმქონე რესპონდენტების 43.2% უკმაყოფილოები არიან საკუთარი ფინანსური მდგომაროებით, უმაღლესი განათლების მქონე 28.5%-თან შედარებით. პროცენტული სხვაობა (43.2%-28.5%=14.7%) მეტყველებს, რომ ცვლადებს შორის კავშირი არსებობს. თუ პროცენტულ სხვაობას

განვიხილავთ ფინანსური მდგომარეობით კმაყოფილ რესპონდენტებს შორის (71.5%-56.8%=14.7%), სხვაობის არსებობა დამატებითი არგუმენტია ცვლადებს შორის არსებული კავშირის შესახებ აზრის განსამტკიცებლად.

წყვილების შედარება

განვიხილოთ უფრო დეტალურად ცვლადებს შორის ზემოთ ნახსენები კავშირი. ორდინალურ ცვლადებს შორის კავშირის მაჩვენებლების გამოთვლის ლოგიკის უკეთ გაგებისათვის უნდა განვიხილოთ დაკვირვების მონაცემები წყვილების სახით. საკითხის დასაზუსტებლად კატეგორია „საშუალო განათლებაზე დაბალი განათლება“ ჩაითვლება, როგორც დაბალი განათლება, ხოლო სულ მცირე უმაღლესი განათლების ქონა ჩაითვლება უმაღლეს განათლებად. იგივე მიდგომა გამოიყენება ფინანსური მდგომარეობით კმაყოფილების შემთხვევაშიც. რესპონდენტები, რომლებიც უკმაყოფილონი არიან თავიანთი ფინანსური მდგომაროებით მიეკუთვნება კატეგორიას- „დაბალი კმაყოფილების დონე“, ხოლო ის რესპონდენტები, რომლებიც კმაყოფილნი არიან თავიანთი ფინანსური მდგომარეობით მიეკუთვნება კატეგორიას-„მაღალი კმაყოფილების დონე“.

ნახაზი 12.1 განათლების დონე და კმაყოფილება ფინანსური მდგომაროებით, ოთხი რესპონდენტის მიხედვით



ამ მაგალითის მიზნებისათვის რესპონდენტებს დავარქვათ საკუთარი სახელები. ჩვენი პირობითი მონაცემები ნაჩვენებია ნახაზ 12.1-ზე. გვყავს ოთხი რესპონდენტი და მათგან შემდგარი წყვილები განათლების დონისა და ფინანსური მდგომარეობით კმაყოფილების მიხედვით.

დაკვირვების წყვილობრივი მონაცემები დაკვირვების მონაცემები, რომლებიც მოცემულია მათი შესაბამისი პოზიციით დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის.

განვიხილოთ პირველი წყვილი ჯონი და არტურო. ჯონი, რომელსაც უმაღლესი განათლება აქვს და ასევე კმაყოფილია საკუთარი ფინანსური მდგომაროებით, ხოლო არტურო რომლესაც არ აქვს უმაღლესი განათლება და ასევე უკმაყოფილოა თავისი ფინანსური მდგომარეობით. ამ მონაცემის წყვილის განხილვით შეგვიძლია დავასკვნათ, რომ მაღალი განათლება უკავშირდება ფინანსური მდგომარეობით მაღალი კმაყოფილების დონეს ან განათლების დონის ზრდასთან ერთად იზრდება ფინანსური მდგომარეობით კმაყოფილების დონე.

შემდეგ განვიხილოთ რუთი და მეი. რუთი, რომლესაც არ აქვს უმაღლესი განათლება კმაყოფილია თავისი ფინანსური მდგომარეობით. მეორეს მხრივ კი მეი, რომელსაც აქვს უმაღლესი განათლება და ნაკლებადაა კმაყოფილი თავისი ფინანსური მდგომარეობით. ამ დაკვირვების მონაცემებზე დაყრდნობით შეგვიძლია დავასკვნათ, რომ რაც უფრო დაბალია რესპონდენტის განათლების დონეზე მით უფრო მაღალია მისი კმაყოფილება თავისი ფინანსური მდგომარეობით, ან სხვაგვარად რომ ვთქვათ კმაყოფილება ფინანსური მდგომარეობით მცირდება განათლების ზრდასთან ერთად.

ნახაზი 12.2 ოთხი რესპონდენტის მწკრივი განათლებისა და ფინანსური მდგომარეობით კმაყოფილების დონის მიხედვით





წყვილთა ტიპები

ორდინალურ ცვლადებს შორის კავშირს ყოველთვის აქვს მიმართულება,რადგან თავად ორდინლურ ცვლადებს აქვთ აღმავალი ან დაღმავალი რიგის მიხედვით დალაგებული მნიშვნელობები. ცვლადთა შორის კავშირის მიმართულება შეიძლება იყოს დადებითი ან უარყოფითი. დადებითი კავშირის შემთხვევაში თუ რესპონდენტი მაღალ პოზიციას იკავებს ერთი კატეგორიის მიხედვით სხვასთან შედარებით და ასევე მეორე კატეგორიის მიხედვით სხვასთან შედარეით უფრო მაღალ პოზიციას იკავებს, ამ სახის შეფარდებითი დამწკრივება იწოდება მსგავსი წყობის წყვილად. ამ ტიპის წყვილების რაოდენობის აღსანიშნავად გამოიყენება *N* სიმბოლო. მსგავსი წყობის წყვილებს შორის არსებობს დადებითი კავშირი. ჯონი და არტურო არიან მსგავსი წყობის წყვლები,დადებითი სტატისტიკური კავშირით. ჯონი განათლების დონით მაღლა დგას არტუროსთან შედარებით, იგივე ასევე მაღლა დგას არტუროსთან შედარებით თავის ფინანსური მდგომარეობის კმაყოფილების დონით. ჯონისა და არტუროსათვის ფინანსური მდგომარეობით კამყოფილება იზრდება განათლების დონის ზრდასთან ერთად.

მსგავსი წყობის წყვილები (Ns) დაკვირვების დაწყვილებული მონაცემები დადებითი კავშირით. რესპონდენტის მაღალი რიგობითი პოზიცია ერთი კატეგორიის მიხედვით დამოუკიდებელი ცვლადისათვის ეთანადება მის ასევე მაღალ რიგობით პოზიციას მეორე კატეგორიის მიხედვით დამოკიდებული ცვლადისათვის.

შევნიშნოთ, რომ ნახაზ 12.2-ზე ასევე არის ოთხი სხვა წყვილიც. ამ წყვილებს აქვს ერთიდაიგივე პოზიცია განათლების ან კმაყოფილების დონის მიხედვით. მაგალითად, რუთს და არტუროს სხვადასხვა კმაყოფილების დონე აქვს, მაგრამ ორივე მათგანი დაბალი განათლების მქონეა. წყვილებს, რომელთაც აქვთ ტოლი მნიშვნელობები ცალკეული ცვლადებისათვის დაკავშირებულ წყვილებად იწოდება. დაკავშირებული წყვილები გამოიყენება კავშირის საზომის — კენდალის ტაუ-b-ს გამოთვლისათვის, რომელსაც მოგვიანებით განვიხილავთ. გამა, რომელიც ყველაზე ხშირად გამოიყენება ორდინალურ ცვლადებს შორის კავშირის გაზომვისათვის რელევენატურია მხოლოდ არადაკავშირებული მონაცემთა წყვილებისათვის ე.ი. მსგავსი წყობისა და საპირისპირო წყობის მონაცემთა წყვილებისათვის.

წყვილთა რაოდენობის დათვლა

გამას გამოთვლისათვის ჯერ განვსაზღვროთ მსგავსი წყობის წყვილების (Ns) და საპირისპირო წყობის მონაცემთა წყვილების (Nd) რიცხვი, რაც შეიძლება მიღებულ იქნას ორგანზომილებიანი ცხრილიდან. მას შემდეგ, რაც ვიპოვით Ns-ს და Nd-ს, გამმას გამოთვლა საკმაოდ მარტივია. 2x2-ზე ზომის ცხრილებისათვის შესაბამისი წყვილების რაოდენობა მცირეა და შესაბამისად მარტივია მათი რაოდენობის დათვლა (როგორც ეს ნახაზ 12.2-ზეა).

მონაცემთა წყვილების ტიპების განსაზღვრის და მათი რაოდენობის დათვლის ილუსტრირებისათვის განვიხილოთ 2x2-ზე ზომის[2] ცხრილი 12.3. შევნიშნოთ, რომ ცხრილის ზედა მარცხენა კუთხეში მდებარე უჯრაში შეტანილია ცვლადის კატეგორიის უმაღლესი მნიშვნელობები (უმაღლესი განათლება და მაღალი კმაყოფილება ფინანსური მდგომარეობით), თუ X მცირდება მარცხნიდან მარჯვნისაკენ, ასევე Y მცირდება ზემოდან ქვემოთ. ამ სახის ცხრილის ანალიზისას, Ns-ის და Nd-ს პოვნისათვის თავდაპირველად დარწმუნდით, რომ ცხრილში მონაცემები სწორედ ნახსენები სახით არის განლაგებული. თუ მონაცემები არ არის ამგვარად განლაგებული, უკვე არსებულ ცხრილში ყოველთვის შეგიძლიათ შეცვალოთ მასში მონაცემთა წარმოდგენის ფორმა.

მსგავსი წყობის წყვილები

მსგავსი წყობის წყვილების (Ns) რაოდენობის პოვნისათვის 2x2-ზე ცხრილისათვის გაამრავლეთ იმ უჯრების სიხშირე, რომელთაც აქვთ მაღალი მნიშვნელობები, როგორც დამოკიდებული ასევე დამოუკიდებელი ცვლადებისათვის იმ უჯრების სიხშირეებზე, რომელთაც აქვთ დაბალი მნიშვნელობები როგორც დამოკიდებული ასევე დამოუკიდებელი ცვლადების შემთხვევაში. სწორედ ამ სიხშირეთა ნამრავლი არის Ns.

ცხრილ 12.3-ში ავიღოთ ზედა მარცხენა უჯრის სიხშირე 1,194 და გავამრავლოთ d უჯრის სიხშირეზე (147). ცხრილის a უჯრას აქვს მაღალი მნიშვნელობა როგორც დამოკიდებული, ასევე დამოუკიდებელი ცვლადისათვის და ასევე d უჯრას აქვს დაბალი მნიშვნელობა ორივე დამოკიდებული და დამოუკიდებელი ცვლადისათვის. მათი ნამრავლი შესაბამისად არის 175,518.

გამოთვლები ნაჩვენებია ნახაზ 12.3-ზე, რომელშიც a უჯრა (უმაღლესი განათლება, მაღალი კმაყოფილება ფინანსური მდგომარეობით) და d უჯრა (დაბალი განათლება, დაბალი კმაყოფილების დონე) გამუქებულია. მსგავსი წყობის წყვილების საერთო რაოდენობაა:

$$Ns = 1,194 \times 147 = 175,518$$

ნახაზი 12.3 მსგავსი წყობის წყვილების რაოდენობის გამოთვლა (Ns) 12.3 ცხრილის მიხედვით:

განათლება

კმაყოფილება ფინანსური მდგომარეობით		მაღალი	დაბალი
	მაღალი	a 1,194	b 193
	დაბალი	c 477	d 147

საპირისპირო წყობის წყვილები

საპირისპირო წყობის წყვილების რაოდენობის გამოთვლისათვის (Nd) მივყვეთ იგივე გამოთვლის გზას, რაც მსგავსი წყობის წყვილების შემთხვევაში გამოვიყენეთ, მაგრამ ერთი გამოკლისით. რადგან ჩვენ გვაინტერესებს წყვილები, რომელთა შემთხვევაში ერთი ცვლადის მიხედვით მნიშვნელობა მეორეს საპირისპიროა, ვამრავლებთ იმ უჯრების მნიშვნელობებს, რომელთაც აქვთ ერთი ცვლადის მიხედვით მაღალი მნიშვნელობა, ხოლო მეორე ცვლადის მიხედვით აქვს დაბალი მნიშვნელობა. ანუ b უჯრის (დაბალი განათლება, მაღალი კმაყოფილების დონე) მნიშვნელობას c უჯრის მნიშვნელობაზე (მაღალი განათლება, დაბალი კმაყოფილების დონე). გამოთვლები ნაჩვენებია ნახაზ 12.4-ზე. საპირისპირო წყვილების რაოდენობა Nd:

$$Nd = 193 \times 477 = 92,01$$

ნახაზი 12.4 საპირისპირო წყვილების რაოდენობის გამოთვლა (Nd) ცხრილ 12.3-ის მონაცემებზე დაყრდნობით

განათლება

კმაყოფილება ფინანსური მდგომარეობით		მაღალი	დაბალი
	მაღალი	a 1,194	b 193
	დაბალი	c 477	d 147

სანამ უშუალოდ გამას ინტერპრეტირებაზე გადავიდოდეთ, კვლავ განვიხილოთ საპირისპირო და მსგავსი წყობის წყვილების შინაარსი. მსგავსი წყობის წყვილები არის წყვილები ცხრილ 12.3-დან, რომელთაც აქვთ დადებითი კავშირი შემდეგი კატეგორიების მიხედვით: რაც უფრო მაღალია განათლების დონე მით უფრო მაღალია ფინანსური მდგომარეობით კმაყოფილების დონე. საპირისპირო წყობის წყვილები, არის დაკვირვების ის წყვილები, რომელთაც აქვთ უარყოფითი კავშირი შემდეგი კატეგორიების მიხედვით: რაც უფრო მაღალია განათლების დონე მით უფრო ნაკლებია ფინანსური მდგომარეობით კმაყოფილების დონე. ჩვენი ანალიზის მიზანია განვსაზღვროთ არის თუ არა კავშირი განათლების დონესა და ფინანსური მდგომარეობით კმაყოფილებას შორის. თუ ამ ცვლადებს შორის კავშირი არსებობს, ჩვენი მიზანია დავადგინოთ თუ როგორია ამ კავშირის მიმართულება: დადებითი ან უარყოფითი?. ამ კითხვებზე პასუხის გაცემისათვის გამა მაჩვენებელი დაგვეხმარება.

დაკავშირებული წყვილები

გამას გამოთვლისათვის გამოიყენება მსგავსი (Ns) და საპირისპირო (Nd) წყობის წყვილები, ხოლო დაკავშირებული წყვილები მხედველობაში არ მიიღება ე.ი. იმ წყვილების რაოდენობა, რომლებიც უკავშირდება დამოკიდებულ ცვლადს (დაკავშირებული დამოკიდებული წყვილები (Ty)) და ის წყვილები რომლებიც უკავშირდება დამოუკიდებელ ცვლადს (დაკავშირებული დამოუკიდებელი წყვილები [Tx]). რადგან კენდალის ტაუ *b*-ს გამოთვლისათვის საჭიროა დაკავშირებული წყვილების რაოდენობის გამოთვლა, ჩვენც ეს უკანასკნელი განვიხილოთ.

დაკავშირებული დამოკიდებული წყვილები (Ty). დამოკიდებული ცვლადის მიხედვით დაკავშირებული წყვილების რაოდენობის პოვნისათვის განვსაზღვროთ წყვილების რაოდენობა თითოეული სტრიქონის მიხედვით. გავამრავლოთ ცხრილის თითოეული უჯრის სიხშირე მის გვერდითა უჯრის სიხშირეზე იგივე სტრიქონში. თავდაპირველად დავიწყოთ მაღალი ფინანსური კმაყოფილების სტრიქონით. ცხრილის მონაცემების მიხედვით მაღალი განათლებისა და მაღალი ფინანსური კმაყოფილების მქონე რესპონდენტების რიცხვი 1,194-ია. მის მარჯვნივ მდებარე უჯრის შესაბამისი სიხშირე 193-ია (დაბალი განათლება, მაღალი კმაყოფილების დონე). შესაბამისად, 1,194 მრავლდება 193-ზე, ნამრავლის შედეგი 230,442 ნიშნავს დაკავშირებული წყვილების რაოდენობის რიცხვს ამ სტრიქონის დამოკიდებული ცვლადის მიხედვით. ამჯერად გადავიდეთ ცხრილის შემდეგ სტრიქონზე, რომელიც ასახავს იმ რესპონდენტების მონაცემებს, რომლებსაც აქვთ დაბალი კმაყოფილების დონე თავიანთი ფინანსური მდგომარეობით. ამ სტრიქონის პირველი უჯრის მნიშვნელობა 477-ია, მის გვერდით მდებარე უჯრის კი 147, მათი ნამრავლი გვაძლევს 70,119 მნიშვნელობას. უკვე შეკრიბოთ დაკავშირებული წყვილების რიცხვი დამოკიდებული ცვლადისათვის (Ty):

$$Ty = 230,442 + 70,119 = 300,561$$

დაკავშირებული წყვილები დამოუკიდებელი ცვლადის მიხედვით (Tx). წყვილები ასევე შესაძლებელია დაკავშირებულ იქნან დამოუკიდებელი ცვლადის შემთხვევაშიც. ამ სახის წყვილების საერთო რაოდენობის გამოთვლისათვის გამოვიყენოთ თითოეული სვეტის წყვილების რაოდენობა. ამ შემთხვევაშიც ზემოთა პარაგრაფში აღწერილ გამოთვლის გზას უნდა მივყვეთ, ოღონდ იმ გამონაკლისით რომ ამ შემთხვევაში სვეტებს ვიყენებთ, ერთი უჯრის სიხშირეს ვამრავლებთ მის ქვემოთა უჯრის სიხშირეზე. დაკავშირებული წყვილების რაოდენობის გამოთვლა დავიწყოთ იმ წყვილებით, რომელთაც მაღალი განათლების დონე აქვთ. ცხრილ 12.3-ის მიხედვით პირველი, უჯრის მნიშვნელობა 1,194-ია, მის ქვემოთ მდებარე უჯრის მნიშვნელობა კი 477-ია. შესაბამისად მათი ნამრავლი 569,538-ია. ამ სვეტისათვის გამოთვლების დასრულების შემდეგ გადავიდეთ მეორე სვეტზე, რომელიც მოიცავს დაბალი განათლების მქონე რესპონდენტების მონაცემებს. პირველ უჯრის მნიშვნელობა 193-ია (დაბალი განათლება, მაღალი კმაყოფილების დონე), მის ქვემოთ მდებარე უჯრის სიხშირე 147-ია. ნამრავლი გვაძლევს 28,371-ს (193x147=28,371). დაკავშირებული წყვილების რაოდენობა დამოუკიდებელი ცვლადისათვის იქნება:

$$Tx = 569{,}538 + 28{,}371 = 597{,}909$$

დაკავშირებული წყვილები დამოკიდებული ცვლადისათვის (Ty) დაკვირვების დაკავშირებული წყვილები დამოკიდებული ცვლადის მიხედვით.

დაკავშირებული წყვილები დამოუკიდებელი ცვლადისათვის (Tx) დაკვირვების დაკავშირებული წყვილები დამოუკიდებული ცვლადის მიხედვით. მიუხედავად იმისა, რომ გამას გამოთვლისათვის არ არის საჭირო წყვილთა რაოდენობის ცოდნა, იგი საჭიროა კენდალის ტაუ-b-ს გამოთვლისათვის.

გამას გამოთვლა

გამა კავშირის სიმეტრიული მაჩვენებელია, რომლის გამოყენება შესაძლებელია როგორც ორდინალური ასევე დიქოტომიური ნომინალური ცვლადის შემთხვევაშიც. მისი მნიშვნელობები ცვალებადობს 0.0-დან +1.0-მდე და ასახავს ცვლადებს შორის კავშირის სიძლიერესა და მიმართულებას. როდესაც გვაქვს მსგავსი წყობის წყვილების მეტი ოდენობა ვიდრე საპირისპირო წყობის წყვილების (Ns მეტია ვიდრე Nd), გამა იქნება დადებითი. გამა გამოითვლება შემდეგი ფორმულით:

$$\text{გამა} = \frac{Ns - Nd}{Ns + Nd} \qquad (12.3)$$

12.3 ფორმულისა და ზემოთა პარაგრაფების გამოთვლების მიხედვით ვიპოვოთ გამას მნიშვნელობა განათლებასა და კმაყოფილებას შორის კავშირისათვის:

$$\text{გამა} = \frac{175,518 - 92,061}{175,518 + 92,061} = 0.31$$

გამას 0.13-თან ტოლობა ნიშნავს, რომ ცვლად სიდიდეებს ,,განათლებას" და ,,კმაყოფილება ფინანსური მდგომარეობით" შორის არსებობს საშუალო სიძლიერის დადებითი კავშირი. აქედან გამომდინარე, რესპონდენტების განათლების შესახებ ინფორმაციის გამოყენება გააუმჯობესებს ფინანსური მდომარეობით კმაყოფილების დონის განსაზღვრის პროგნოზს 31%-ით.

გამა კავშირის ანალიზის სიმეტრიული მაჩვენებელია, რაც ნიშნავს რომ გამას მნიშვნელობა არ შეიცვლება მიუხედავად იმისა, თუ რომელი ცვლადი ჩაითვლება დამოკიდებულად ან დამოუკიდებლად. ე.ი. თუ პირიქით შევეცდებოდით განათლების დონის პროგნოზირებას ფინანსური კმაყოფილების დონის მიხედვით, ამით გამის მნიშვნელობა არ შეიცვლებოდა.

გამა-ცვლადებს შორის კავშირის სიმეტრიული საზომი ორდინალური ან დიქოტომიური ნომინალური ცვლადებისათვის. მნიშვნელობა ცვალებადობს 0.0დან +/-1.0 მდე და ასახავს ცვალადებს შორის კავშირის სიძლიერეს და მიმართულებას. როცა მონაცემებში მრავლადაა Ns წყვილები, გამას მნიშვნელობა დადებითია; ხოლო როცა მონაცემებში Nd წყვილების სიმრავლეა გამა უარყოფით მნიშვნელობას იღებს.

კავშირის სიმეტრიული საზომი საზომი რომლის მნიშვნელობა დამოუკიდებელია იმის მიუხედავად თუ რომელ ცვლადს ჩავთვლით დამოკიდებულად ან დამოუკიდებლად.

დადებითი და უარყოფითი გამა

შევნიშნოთ, რომ 12.3 ფორმულის მიხედვით გამას მნიშვნელობა და დადებითობა ან უარყოფითობა დამოკიდებულია მსგავსი წყობისა და საპირისპირო წყობის წყვილების ოდენობებზე. მსგავსი წყობის მონაცემთა წყვილების სიჭარბე გამას მნიშვნელობას დადებითს ხდის და პირიქით. რაც უფრო მეტია სხვაობა Ns და Nd წყვილების რაოდენობას შორის კოეფიციენტის მნიშვნელობა მით უფრო დიდია გამა (ნიშნის მიუხედავად). მაგალითად, როცა ყველა წყვილი მსგავსი წყობისაა Ns (Nd = 0), გამას მნიშვნელობა 1.00 ტოლია:

$$გამა = \frac{Ns - 0}{Ns + 0} = 1.0$$

გამას მნიშვნელობის 1-თან ტოლობა ნიშნავს, რომ ცვლადებს შორის არსებობს დადებითი კავშირი და დამოკიდებული ცვლადის მნიშვნელობის პროგნოზირება შესაძლებელია დამოუკიდებელი ცვლადის მიხედვით შეცდომის გარეშე. როდესაც Ns რაოდენობა ნოლია, მაშინ გამას მნიშვნელობა -1.0-ის ტოლი იქნება, რაც ნიშნავს, რომ ცვლადებს შორის არსებობს პირდაპირი უარყოფითი კავშირი:

$$გამა = \frac{0 - Ns}{0 + Nd} = -1.0$$

როცა Ns = Nd, გამას მნიშვნელობა ნოლია:

$$გამა = \frac{Ns - Nd}{Nd + Ns} = \frac{0}{Ns + Nd} = 0.0$$

გამას ნოლთან ტოლობა ნიშნავს, რომ ცვლადებს შორის კავშირი არ არის; შესაბამისად, ერთი ცვლადის მნიშვნელობების პროგნოზირებისათვის მეორე ცვლადის გამოყენება დამატებით არაფერს მოგვცემს და პირიქით.

გამა, როგორც შეცდომის პროპორციული შემცირების საზომი

შეცდომის პროპორციული შეცდომის სხვა საზომების მსგავსად გამაც ეფუძნება პროგნოზირების ორ მეთოდს. პროგნოზირების პირველი მეთოდისას არ გამოიყენება წყვილთა შეფარდებითი წყობა დამოუკიდებელი ცვლადის მიხედვით (მხოლოდ გაერთიანებული ცვლადთა წყვილებია ჩართული გამას გამოთვლაში), ხოლო პროგნოზირების მეორე მეთოდისას ეს უკანასკნელი მხედველობაში მიიღება. დავუშვათ, რომ გვსურს თითოეული წყვილის რიგობითი რანგის პროგნოზირება 267,579 (მსგავსი და საპირისპირო წყობის წყვილების ჯამი: $Ns + Nd = 175,518 + 92,061 = 167,579$) განათლების მიხედვით მათი შეფარდებითი რანგის შესახებ მონაცემების იგნორირებით. თუ თითოეული ამ წყვილის პროგნოზირებისათვის გამოვიყენებთ შემთხვევით შერჩევის მეთოდს, არსებობს 50%-იანი შანსი იმისა, რომ ჩვენ შედეგები იქნება სწორი. შესაბამისად, ჩვენ დავუშვებთ ნახევარ ე.ი. $(Ns + Nd) / 2 = (175,518 + 92,061) / 2 = 133,790$ შეცდომას.

ამჯერად შევეცადოთ გავაუმჯობესოთ ჩვენი პროგნოზი განათლების შესახებ მონაცემების გამოყენებით. პროგნოზის გაუმჯობესების ალბათობა დამოკიდებულია Ns-ის რაოდენობაზე Nd-ის რაოდენობასთან შედარებით. როცა Ns-ის რიცხვი მეტია Nd-ის რაოდენობაზე, როგორც ეს ჩვენს მაგალითშია, რაციონალურია დაშ-

ვება იმის შესახებ, რომ თუ პირს აქვს უმაღლესი განათლება, მაშინ მას მაღალი კმაყოფილებაც ექნება თავისი ფინანსური მდგომარეობით (ეს არის სწორედ Ns წყვილების მიერ ნაჩვენები წყობა). ამ დაშვებაზე დაფუძნებული პროგნოზით ჩვენ დავუშვებთ 92,061 შეცდომას (Nd წყვილთა რაოდენობა, რომელთათვის არ არის სწორი ნახსენები დაშვება), ან 41,729 შეცდომას (133,790 - 92,061 = 41729). გავყოთ აღნიშნული შედეგი თავდაპირველი შეცდომის ოდენობაზე:

$$\frac{41,729}{133,790} = 0.31$$

შედეგად მივიღებთ შეცდომის პროპორციული შემცირების მნიშვნელობას 0.310-ს, რაც გამა კოეფიციენტის მნიშვნელობის ტოლია, რომელიც 12.3 ფორმულის გამოყენებით მივიღეთ.

კენდალის ტაუ-b-ს გამოთვლა

კენდალის ტაუ-b კავშირის სიმეტრიული საზომია ორდინალური ცვალდებისათვის. განსხვავებით გამასაგან კენდალის ტაუ-b-ს გამოთვლაში მონაწილეობს როგორც დამოკიდებული, ასევე დამოუკიდებელი ცვლადების პარალელური მონაცემები. იგი ასახავს ცვლადებს შორის კავშირის სიძლიერეს და მიმართულებას; მნიშვნელობები ცვალებადობს 0.0-დან +/-1.0 მდე, ხოლო ცხრილის სახით მოცემული მონაცემების შემთხვევაში, რომელშიც სვეტების რიცხი სტრიქონების რიცხვის ტოლია კენდალის ტაუ-ბ-ს მნიშვნელობა +/-1-ია.

$$\text{კენდალის ტაუ-}b = \frac{Ns - Nd}{\sqrt{(Ns + Nd + Ty)(Ns + Nd + Tx)}}$$

კენდალის ტაუ-b ორდინალურ ცვლადებს შორის კავშირის სიმეტრიული საზომი. გამასაგან განსხვავებით ამ მაჩვენებლის გამოთვლისას მხედველობაში მიიღება, როგორც დამოკიდებული, ასევე დამოუკიდებელი ცვლადების მიხედვით დალაგებული მონაცემთა წყვილები. მნიშვნელობათა სიმრავლე 0.0-დან +/-1.0 შუალედია და ასახავს ცვლადთა შორის კავშირის სიძლიერესა და მიმართულებას.

ცხრილ 12.3-ის მონაცამების მიხედვით გამოვთვალოთ კენდალის ტაუ-*b*. რადგან უკვე ვიცით საპირისპირო (Nd) და მსგავსი წყობის (Nd) მონაცემთა წყვილების რაოდენობა, ასევე დაკავშირებული დამოკიდებული წყვილების (Ty) და დაკავშირებული დამოუკიდებელი წყვილების რაოდენობები (Tx), 12.4 ფორმულის მიხედვით:

$$\text{კენდალის ტაუ-}b = \frac{Ns - Nd}{\sqrt{(Ns + Nd + Ty)(Ns + Nd + Tx)}}$$

კენდალის ტაუ-b-ს 0.12-თან ტოლობა აღნიშნავს სუსტ დადებით კავშირს ცვლადებს-განათლებასა და ფინანსური მდგომარეობით კმაყოფილებას შორის. რესპონდენტის განათლების შესახებ ინფორმაციის გამოყენებით ფინანსური კმაყოფილების დონის პროგნოზისას დაშვებული შეცდომა შემცირდა 12%-ით. თუ გავიხსენებთ იგივე მონაცემებისათვის გამას მნიშვნელობა იყო 0.31. აქვე შევნიშნოთ, რომ კენდალის ტაუ-b-ის მნიშვნელობა ყოველთვის მცირე იქნება გამასთან შედარებით, რადგან კენდალის ტაუ-b-ს გამოთვლისას გამოიყენება როგორც დამოკიდებული, ასევე დამოკიდებული ცვლადების მნიშვნელობები.

სტატისტიკა პრაქტიკაში: განათლება და ჰომოსექსუალობის მიმართ დამოკიდებულება

არსებობს თუ არა სტატისტიკური კავშირი განათლების დონესა და ჰომოსექსუალობის მიმართ დამოკიდებულებაში? თუ ეს ასეა, როგორია ამ დამოკიდებულების მიმართულება? რამდენად მნიშვნელოვანია ეს დამოკიდებულება? ამ კითხვებზე პასუხის გაცემისათვის გამოვიყენოთ ზოგადი სოციოლოგიური კვლევის მონაცემები (GSS 2008). სიმარტივისათვის, განათლების დონე და ჰომოსექსუალობის მიმართ დამოკიდებულება ჩაწერილ იქნა როგორც დიქოტომიური ცვლადები. განათლების დონე მივიჩნიოთ, როგორც დამოუკიდებელი ცვლადი, ხოლო ჰომოსექსუალობის მიმართ დამოკიდებულება როგორც დამოკიდებული ცვლადი (ცხრილი 12.4). ცხრილის მონაცემების მიხედვით ვიზუალურად ჩანს, რომ განათლების დონესა და ჰომოსექსუალობის მიმართ დადებით დამოკიდებულებას შორის არსებობს სტატისტიკური კავშირი, ჰომოსექსუალობის მიმართ დადებითი შეხედულება გამოხატა უმაღლესი განთლების მქონე რესპონდენტების 41%, ხოლო უმაღლესი განათლების არმოქნე რესპონდენტებიდან მხოლოდ 24.5%-მა. გამას და კენდალის ტაუ-b-ს მნიშვნელობებიც სწორედ ამაზე მეტყველებს. SPSS-ის გამოყენებით გამას მნიშვნელობა 0.36-ის ტოლია, რაც ნიშნავს საშუალო სიძლიერის კავშირის არსებობას განათლების დონესა და ჰომოსექსუალობის მიმართ დამოკიდებულებას შორის. ასევე, გამას ეს რიცხვითი შედეგი ნიშნავს, რომ პროგნოზირების შეცდომა ჰომოსექსუალობის მიმართ შემცირდა 36% განათლების შესახებ მონაცემების გამოყენებით. აღნიშნულს დამატებით ამყარებს კენდალის ტაუ-*b*-ს მნიშვნელობაც. SPSS-ის გამოყენებით კენდალის ტაუ-b-ს მნიშვნელობა 0.13-ია, რაც ნიშნავს ცვალდებს შორის სუსტ დადებით კავშირს. დააკვირდით, რომ ორივე მაჩვენებლის მიხედვით ცვლადებს შორის კავშირი დადებითია, მიუხედავად მათ მიერ ნაჩვენები

სტატისტიკური კავშირის განსხვავებული სიძლიერისა. ეს უკანასკნელი შედეგი განპირობებულია იმით, რომ კენდალის ტაუ-b-ს გამოთვლისას გაითვალისწინება მონაცემთა წყვილები, როგორც დამოკიდებული, ასევე დამოუკიდებელი ცვლადებისათვის.

ცხრილი 12.4 ჰომოსექსუალობის მიმართ დამოკიდებულება განათლების დონის მიხედვით

ჰომოსექსუალობის მიმართ დამოკიდებულება (Y)	განათლება (X)		
	უმაღლესი ან მისი შემდგომი განათლ.	საშუალო განათლება	ჯამი
დადებითი	41.0% (431)	24.5% (52)	38.2% (483)
უარყოფითი	59.0% (621)	75.5% (160)	61.8% (781)
სულ (N)	100% (1,052)	100% (212)	100% (1,264)

წყარო: General Social Survey, 2008.
შენიშვნა: გამა = 0.36, კენდალის ტაუ-b = 0.13

ორდინალურ ცვლადებს შორის კავშირის საზომების გამოყენება დიქოტომიური ცვლადებისათვის

ორდინლურ ცვლადებს შორის კავშირის საზომები არ იცვლებიან მოდალური კატეგორიების მიხედვით, მაგალითად ესეთი საზომია გამა. შედეგად, კავშირის ეს სახეობები შესაძლებელია გამოყენებულ იქნას იქ სადაც არ ჩანს კავშირი გამა მაჩვენებლის მიხედვით. ეს მაჩვენებლები შესაძლებელია გამოყენებულ იქნას ორდინალურ შკალაზე ჩაწერილი მონაცემებისათვის. დიქოტომიური (ე.ი. ორი კატეგორიის მქონე) ცვალდები შესაძლებელია ჩაითვალოს ორდინალურ ცვლადებად ცალკეული კვლევის მიზნებისათვის. გავიხსენოთ, რომ ჩვენ ამ თავში ლაპარაკობდა გამოვთვალეთ აბორტის მიმართ დამოკიდებულებასა და განათლების დონეს შორის კავშირის შესაფასებლად (ცხრილი 12.1). მიუხედავად იმისა, რომ ორივე ეს ცვლადი სიდიდე შესაძლებელია მიჩნეულ იქნას ნომინალურ ცვლადად, რადგან ორივე დიქოტომიური ცვლადებია (დიახ/არა, უმაღლესი განათლების ქონა/არქონა), ისინი ასევე შესაძლებელია მივიჩნიოთ როგორც ორდინალური ცვლადები. შესაბამისად, ამ შემთხვევაში ცვლადებს შორის კავშირის შეფასებისათვის შეგვიძლია გამოვიყენოთ გამა-ორდინალურ ცვლადებს შორის კავშირის საზომი3.

გამოვთვალოთ გამას მნიშვნელობა 12.1 ცხრილისათვის. ამ ცხრილისათვის მსგავსი წყობის მონაცემთა წყვილების რაოდენობაა:

$$Ns = 399\ (195) = 77{,}805$$

საპირისპირო წყობის მონაცემთა წყვილების რაოდენობა ცხრილ 12.1-ისათვის:

$$Nd = 156\ \ (222) = 34{,}632$$

გამას გამოთვლა:

$$\frac{77{,}805 - 34{,}632}{77{,}805 + 34{,}632} = 0.38$$

გამას 0.38-იანი მნიშვნელობა ამყარებს ჩვენს ადრინდელ დასკვნას პოზიტიური კავშირის არსებობის შესახებ განათლების დონესა და აბორტის მხარდაჭერას შორის. გამოკითხული რესპონდენტები უმაღლესი განათლების დონით უფრო მეტი ალბათობით დაუჭერენ მხარს აბორტის შესაძლებლობას.

კვლევითი ლიტერატურის მიმოხილვა: მსოფმხედველობრივი შეხედულებები და აბორტის შესახებ აზრი

მოდით განვიხილოთ კვლევის მაგალითი, რომელშიც ვნახავთ თუ როგორ გამოიყენება გამა მაჩვენებელი სოციოლოგიური კვლევების ლიტერატურაში. ეს მაგალითი აღებულია კვლევიდან, რომელშიც განხილულია ჰიპოთეზა იმის შესახებ, რომ აბორტის შესახებ აზრზე გავლენა აქვს პიროვნების ზოგად მსოფმხედველობას, მორალურ შეხედულებებს. კრისტინ ლუგარის მიხედვით: „აბორტის შესახებ მოდებატეების ორივე მხარე ფარულად და მათგან დამოუკიდებლად იზიარებენ მსოფმხედველობრივად გაბატონებულ შეხედულებებს, რომლებიც ნათლად არ არის გამოკვეთილი და ფორმულირებული“[4]. ზოგადად, თუ პირი რელიგიურია და ეთანხმება ქალის ცამხრივ როლს ოჯახის მოვლასა და ბავშვების აღზრდაში ცალსახად არ ეთანხმება აბორტის იდეას. საპირისპიროდ, პირი რომელიც არ მისდევს რელიგიურ ცხოვრებას, ეთანხმება ქალის როლის ეგალიტარულ ხედვას, აქვს ლიბერალური სექსისტური შეხედულებები, და საკუთარ თავს მემარცხენედ მიიჩნევს მეტი ალბათობით დაეთანხმება ქალის უფლების არსებობას თავად აირჩიოს ან უარი თქვას აბორტზე[5]. კვლევის მთავარი ჰიპოთეზაა: „რაც უფრო კონსერვატიულია პიროვნების მსოფლმხედველობრივი ხედვა, მით უფრო მძაფრია მის მიერ აბორტის უარყოფა“[6].

აღნიშნული კვლევის ჰიპოთეზის შემოწმებისათვის დანიელ სპაისერმა გამოიყენა საერთო სოციოლოგიური გამოკითხვის მონაცემები (GSS, 1990), რომლებიც რეპრეზენტატიული იყო აშშ-ს მოსახლეობის მოწიფული ასაკის ნაწილისათვის. პიროვნების ზოგადი მსოფმხედველობის შეფასებისათვის სპაისერმა აირჩია ექვსი

ცვლადი: დამოკიდებულება ქორწილამდელი სექსისადმი, გენდერული როლები, რელიგიური ცხოვრების ინსტენსივობა, პოლიტიკური შეხედულებები, ფუნდამენტალიზმი და ბიბლიის მიმართ მიდგომა. ჩვენი მაგალითის მიზნებისათვის ნახსენები ცვლადებიდან მხოლოდ პირველ სამ ცვლადს განვიხილავთ.

ამ ცვლადების შესახებ მონაცემები მიიღებოდა შემდეგ კითხვებზე პასუხებით:

- გენდერული როლების შესახებ აზრი: „ეთანხმებით თუ არა შემდეგ აზრს?: ქალები უპირატესად უნდა ზრუნავდნენ ოჯახზე და მამაკაცებთან შედარებით ნაკლებად უნდა იყვნენ ჩართულნი ქვეყნის მართვაში.“

- დამოკიდებულება ქორწილამდელი სექსისადმი: „არსებობს აზრთა ბევრი წინააღმდეგობა მისაღებ მორალურ საკითხებზე. თუ ქალსა და მამაკაცს ქორწინებამდე ექნებათ სექსუალური ურთიერთობა, როგორ ფიქრობთ ეს ყოველთვის: არასწორია, ჩვეულებრივ არასწორია, გარკვეულწილად არასწორია ან მისაღებია?“
- რელიგიური ცხოვრების ინტენსივობა: „მიგაჩნიათ თუ არა თავი: ძლიერ, საშუალო, ან სუსტ მორწმუნედ თქვენი რელიგიის მიხედვით?“.

დამოკიდებული ცვლადის მონაცემები „აბორტის შესახებ აზრი“ მიღებულ იქნა შემდეგ ოთხ კითხვაზე პასუხით. ამასთან მხედველობაში იქნა მიღებული სხვადასხვა გარემოებები, რომელთა პირობებშიც დათანხმდებოდა პირი აბორტის არსებობას ან არ დათანხდებოდა. მიჩნეულ იქნა, რომ პირი პირი ეთანხმება აბორტს თუ ის დადებით პასუხს სცემს შემდეგი ოთხი შეკითხვიდან ყველას, აბორტი მისაღებია:

1. თუ ქალი დაქორწინებულია და აღარ სურს მეტი ბავშვი
2. თუ ქალის ოჯახს აქვს დაბალი შემოსავლები და მეტი ბავშვის მატერიალური რჩენა არ შეუძლია
3. თუ ქალი დაქორწინებული არ არის და არ სურს მოცემულ მამაკაცზე დაქორწინება
4. ქალი აბორტს მიმართავს სხვა ნებისმიერი მიზეზით

აღნიშნულ ოთხ კითხვაზე პასუხების მიხედვით, რესპონდენტების კლასიფიცირება მოხდა შემდეგ ორ ჯგუფად: „ვეთანხმები“, „არ ვეთანხმები“ აბორტის არჩევანს. ორკატეგორიანი ცვლადის პროცენტული განაწილება და გამას მნიშვნელობა თითოეული ცვლადისათვის მოცემულია ცხრილ 12.5-ში.

მონაცემთა განხილვა

განვიხილოთ ცხრილ 12.5-ში მოცემული მონაცემები. ცხრილში მონაცემები თითოეული დამოუკიდებელი ცვლადის მიხედვით სამ ნაწილადაა დაყოფილი. თითოეული ნაწილი შესაძლებელია განხილულ იქნას, როგორც დამოუკიდებელი პროცენტული განაწილება. დამოუკიდებელი ცვლადები ცხრილის სტრიქონებშია მოცემული, ხოლო დამოკიდებული ცვლადები სვეტებშია მოცემული. დამოუკიდებელი ცვლადის თითოეული კატეგორიასათვის მოცემულია რესპონდენტების რაოდენობა, რომელთაც კითხვას პასუხი გასცეს და ასევე თანხმობის შესაბამისი პროცენტები. მაგალითად, ცვლად გენდერულ როლების შესახებ აზრს აქვს ორი კატეგორია: „ლიბერალური" და „ტრადიციული". მაგალითად, 58 ტრადიციული შეხედულების მქონე რესპონდენტიდან 21%-მა მხარი დაუჭირა აბორტს, ხოლო 292 ლიბერალიდან 53% მხარი დაუჭირა აბორტის შესაძლებლობის არსებობას.

ცხრილი 12.5 აბორტის მიმართ მხარდაჭერა, პროცენტული მაჩვენებლები დამოუკიდებელი ცვლადების მიხედვით (GSS, 1990)

ცვლადი	კატეგორია	N	აბორტის მხარდაჭერა	გამა
ქორწინებამდელი სექსის მიმართ დამოკიდებულება	ყოველთვის არასწორია	92	15%	–.644
	ჩვეულებრივ არასწორია	42	36%	
	გარკვეულწილად არასწორია	75	51%	
	მისაღებია	41	71%	
გენდერული როლების შესახებ აზრი	ტრადიციული	58	21%	–.625
	ლიბერალური	292	53%	
რელიგიური ცხოვრების ინტენსივობა	ძლიერი	261	28%	–.433
	საშუალო	97	55%	
	სუსტი	312	59%	

წყარო: ადაპტირებულია Daniel N. Spicer, "World View and Abortion Beliefs: A Replication of Luker's Implicit Hypothesis", Sociological Inquiry 64, no.1 (1994), 115-116,120-121, Copyright (C) 1994 By the University of Texas Press. Published by Blackwell Publishing.

✓ *ცოდნის შემოწმება*

იმ რესპონდენტთა პროცენტული წილები, რომლებმაც უარყვეს აბორტის არჩევანის შესაძლებლობა არ არის მოცემული ცხრილში. მაგრამ ამ მონაცემების მიღება საკმაოდ მარტივია: თითოეული კატეგორიისათვის შესაბამისი თანხმობის პროცენტი გამოაკლეთ 100%-ს. მაგალითად, 58 ტრადიციული შეხედულების რესპონდენტიდან 79% (100%-21%) უარყო აბორტის არჩევანის შესაძლებლობა. აღნიშნულის მსგავსად შეავსეთ ცხრილის სხვა უჯრები, აბორტის უარყოფის შესაბამისი პროცენტული მაჩვენებლებით.

მონაცემთა ინტერპრეტირება

ცხრილის 12.5 მონაცემების ინტერპრეტირებისას დავაკვირდეთ როგორც პროცენტულ მაჩვენებლებს თითოეული კატეგორიის მიხედვით, ასევე გამას შესაბამის მნიშვნელობებს. დავიწყოთ პროცენტების შედარებით. როგორც მე-10 თავში ვისწავლეთ, როცა პროცენტები სტრიქონების მიხედვითაა მოყვანილი, როგორც ეს 12.5 ცხრილშია შედარებები უნდა მოხდეს სვეტების დაღმავალი რიგით. მაგალითად, „რელიგიური ცხოვრების ინტენსივობასა“ და „აბორტის შესახებ აზრს“ შორის დამოკიდებულების განხილვისათვის შევადაროთ 28%, 55% და 59%ები. თითოეული დამოუკიდებელი ცვლადის მიხედვით პროცენტული მაჩვენებლების შედარების შემდეგ მკვლევარი გვთავაზობს შემდეგ ინტერპრეტირებას:

„როგორც მონაცემების ანალიზიდან ჩანს, ქორწილამდელი სექსის მიმართ თანხმობის ზრდასთან ერთად, იზრდება აბორტის არჩევანის შესაძლებლობის მიმართ დადებითი აზრი. რესპონდენტები გენდერული როლების შესახებ ტრადიციული შეხედულებებით უფრო მეტი ალბათობით უარყოფენ აბორტის არჩევანის შესაძლებლობას, ხოლო ლიბერალური შეხედულებების მქონე რესპონდენტები ეთანხმებიან აბორტის არჩევანის შესაძლებლობას.. რესპონდენტები ხშირი რელიგიური ცხოვრებით უფრო მეტი სიხშირით უარყოფენ აბორტის შესაძლებლობას, ხოლო ნაკლებად რელიგიური რესპონდენტები მხარს უჭერენ აბორტის შესაძლებლობას“[7].

ნახსენები ინტერპრეტირება ცვალდებს შორის კავშირის შესახებ მყარდება ცხრილშივე მოცემული გამას მნიშვნელობებით. გამას მნიშვნელობები — 0.433 რელიგიური ცხოვრების ინტენსივობისათვის, — 0.644-მდე ქორწინებამდე სექსისადმი დამოკიდებულებისათვის ცვალებადობს. გამას ეს მნიშვნელობები აღნიშნავს საშუალოდან ძლიერ კავშირს პირის ზოგად მსოფმხედველობრივ შეხედულებებსა და აბორტის შესახებ მოსაზრებებს შორის.

შევნიშნოთ, რომ გამას ყველა მნიშვნელობები უარყოფითია. გამას უარყოფითი მნიშვნელობა ასახავს, რომ ერთი ცვლადის დაბალ მნიშვნელობებს მოსდევს მეორე ცვლადის მაღალი მნიშვნელობები და პირიქით.

ორდინალურ ცვლადებს შორის კავშირის მიმართულების ინერპრეტირებისას სიფრთხილეა საჭირო, რადგან ეს უკანასკნელი უშუალოდ მკვლევარის მიერ არჩეულ ცვლადთა კოდირებაზეა დამოკიდებული. მკლევარმა თავად უნდა მოიყვანოს კვლევის ანგარიშში არჩეული კოდირების შესახებ ინფორმაცია. სპაისერი ცვლადთა შორის კავშირის ინტერპრეტირებას შემდეგნაირად იძლევა: „მაღალი ქულები.. დამოუკიდებელი ცვლადებისათვის (მაღალი ქულები აღნიშნავს კონსერვატიულ სოციალურ შეხედულებებს) უკავშირდება დაბალ ქულებს (უარყოფითი პასუხი) აბროტის შესახებ დამოკიდებულებებს „გამას მნიშვნელობები უარყოფითია“[8].

სხვა სიტყვებით რომ ვთქვათ, გამას უარყოფითი მნიშვნელობა ქორწინებამდე სექსა, გენდერული როლების შესახებ აზრსა და რელიგიური ცხოვრების ინტენსივობას შორის ნიშნავს, რომ რეპოდენტები, რომელთაც კონსერვატიული სოციალური შეხედულებები აქვთ უფრო მეტად უარყოფენ აბორტის არჩევანის შესაძლებლობას. და პირიქით რესპონდენტები, რომელთაც შედარებით ლიბერალური სოციალური შეხედულებები აქვთ უფრო მეტი სიხშირით ეთანხმებიან აბორტის არჩევანის შესაძლებლობის არსებობას.

ძირითადი განმარტებები:

• კავშირის საზომი- ერთიანი შემაჯამებელი მაჩვენებელი, რომელიც ასახავს ცვლადებს შორის კავშირის სიძლიერეს, აჩვენებს დამოკიდებული ცვლადების პროგნოზირების შესაძლებლობებს დამოუკიდებელი ცვლადების მიხედვით და ხშირ შემთხვევაში იძლევა კავშირის მიმართულების შეფასების საშუალებას.

• შეცდომის პროპორციული შემცირების (PRE) – კონცეფცია ეფუძნება სტატისტიკური კავშირის შეფასების რამდენიმე საზომის განმარტებას და ინტერპრეტაციას. PRE-ის მაჩვენებლები მიიღება იმ შეცდომების შედარებით, რაც ერთის მხრივ მიიღება დამოკიდებული ცვლადების მნიშვნელობების პროგნოზირებისას დამოუკიდებელი ცვლადებისაგან განცალკევებით, და მეორეს მხრივ იმ შეცდომებთან, რაც მიიღება დამოკიდებული ცვლადების მნიშვნელობების პროგნოზირებით დამოუკიდებელი ცვლადების მხედველობაში მიღებით. ე.ი. ეს კონცეფცია დამოკიდებული ცვლადების პროგნოზირების შედეგებს ეფუძნება.

• კავშირის PRE მაჩვენებელი იღებს მნიშვნელობას 0.0-დან +/- 1.0-მდე. შეცდომის პროპორციული შეცდომის ნულთან ტოლობა ნიშნავს, რომ ცვალდებს შორის კავშირი არ არსებობს; დამოუკიდებელი ცვლადის შესახებ მონაცემები არ შეცვლის დამოკიდებული ცვლადის მნიშვნელობის პროგნოზირების შედეგებს. PRE-ის +/- 1.0 ტოლობა ნიშნავს ცვლადებს შორის სრულყოფილ დადებით ან უარყოფით კავშირს, შესაძლებელია დამოკიდებული ცვლადის მნიშვნელობის უშეცდომოდ პროგნოზირება დამოუკიდებელ ცვლადის მნიშვნელობებზე დაყრდნობით.

• კავშირის საზომი შესაძლებელია იყოს სიმეტრიული ან ასიმეტრიული. საზომის სიმეტრიულობა ნიშნავს მისი რიცხვითი მნიშვნელობის უცვლელობას დამოკიდებული და დამოუკიდებელი ცვლადების ცვლილების მიუხედავად. კავშირის ასიმეტრიული მაჩვენებელი ნიშნავს მისი რიცხვითი მნიშვნელობის ცვლილებას დამოკიდებული და დამოკიდებული ცვლადების ცვლილების შესაბამისად.

• ლამბდა – კავშირის ასიმტერიული საზომი; ლამბდა გამოიყენება ნომინალური ცვალდებისათვის და იღებს მნიშვნელობებს 0.0-დან 1.0-მდე. იგი გვიჩვენებს დამოკიდებულ და დამოუკიდებელ ცვალდებს შორის კავშირის სიძლიერეს.

• გამა-ცვლადებს შორის კავშირის სიმეტრიული საზომი ორდინალური ან დიქოტომიური ნომინალური ცვლადებისათვის. მნიშვნელობა ცვალებადობს 0.0დან +/-1.0 მდე და ასახავს ცავლადებს შორის კავშირის სიძლიერეს და მიმართულებას.

• კენდალის ტაუ-b ორდინალური ცვლადებს შორის კავშირის სიმეტრიული საზომი. გამასაგან განსხვავებით ამ მაჩვენებლის გამოთვლისას მხედველობაში მიიღება, როგორც დამოკიდებული, ასევე დამოუკიდებელი ცვლადების მიხედვით დალაგებული მონაცემთა წყვილები. მნიშველობათა სიმრავლე 0.0-დან +/-1.0 შუალედია და ასახავს

ცვლადთა შორის კავშირის სიძლიერესა და მიმართულებას

• კრამერის V ცვლადებს შორის კავშირის საზომი ნომინალური ცვლადებისათვის, ეფუძნება ხი-კვადრატის მნიშვნელობას და მნიშვნელობათა შუალედია 0.0-1.0. რადგან ეს მაჩვენებელი არ იღებს უარყოფით მნიშვნელობებს, იგი არ აჩვენებს ცვლადებს შორის კავშირის მიმართულებას.

საკვანძო ტერმინები

კავშირის ასიმეტრიული საზომი
გამა
საპირისპირო წყობის მონაცემთა წყვილები (Nd)
კენდალის ტაუ b
ლამბდა
კავშირის საზომი
დაკვირვების მონაცემებთა წყვილები
შეცდომის პროპორციული შემცირება
მსგავსი წყობის წყვილები (Ns)
კავშირის სიმეტრიული საზომი
დაკავშირებული დამოკიდებული წყვილები (Ty)
დაკავშირებული დამოუკიდებული წყვილები (Tx)

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

[GSS08PFP-A]

მაგალითი 1: ნომინალურ ცვლადთა შორის კავშირის საზომის გამოთვლა ბივარიანტული ცხრილებისათვის

როგორც გახსოვთ, ბივარიანტული ცხრილების მიღებისათვის SPSS-ში გამოვიყენეთ კროსტაბულაცია (Crosstabs). (მე-10 თავი). იგივე ბრძანებათა თანმიმდევრობა უნდა გამოვიყენოთ SPSS-ში, რომ მივიღოთ კავშირის საზომები. ჩვენი მაგალითი SPSS-ში დავიწყოთ კავშირის გამოთვლით ბიბლიისადმი დამოკიდებულებასა (BIBLE) და ქალებისათვის აბორტის კანონიერი უფლების მინიჭებისადმი მხარდაჭერას შორის (ABANY).

SPSS-ის ბრძანებათა პანელზე დაკლიკეთ Analyze (ანალიზი), Descriptive Statistics (აღწერითი სტატისტიკა), Crosstabs (ჯვარედინი გამოთვლები), მივიღებთ კროსტაბულაციის დიალოგურ ფანჯარას, სტრიქონის ოპციაში გადაიტანეთ ცვლადი ABANY, ხოლო სვეტში BIBLE. დაკლიკეთ Statistics ოპცია, დიალოგურ ფანჯარაში (ნახაზი 12.5) მოცემულია სტატისტიკური მაჩვენებლების ოთხი სხვადასხვა წყება, რომლებიც დაყოფილია ნომინალური და ორდინალური ცვლადებისათვის. ლამბდა მოცემულია ნომინალური ცვლადების კატეგორიაში, ხოლო გამა და

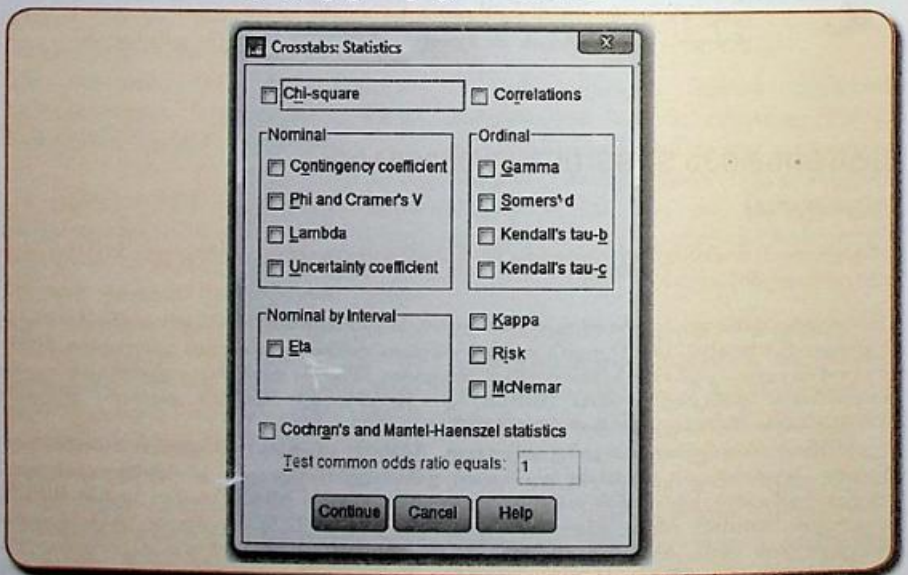
კენდალის ტაუ-*b* ორდინალური ცვლადების კატეგორიაში. რაც შეეხება კრამერის V-ს მიღება შესაძლებელია კრამერის V-ს დიალოგურ ფანჯარაზე დაკლიკვით. აქვეა სტატისტიკური კავშირის სხვა საზომები, როგორიცაა სომერის d და f, მათ ამ წიგნში არ განვიხილავთ. ხი-კვადრატ სტატისტიკა კი უკვე განვიხილეთ მე-11 თავში.

რადგან ჩვენს მიერ არჩეული ორივე ცვლადი ნომინალურია, დიალოგურ ფანჯარაში ავირჩიოთ ლამბდა. მნიშვნელოვანია გვახსოვდეს, რომ ჩვენ როგორც სტატისტიკური პროგრამის მომხმარებელმა უნდა ვიცოდეთ თუ რა შესაბამისი სტატისტიკის არჩევა გვჭირდება ჩვენი კვლევის მიზნებიდან გამომდინარე. SPSS ვერ დაგვეხმარება იმის განსაზღვრაში, თუ რომელი სტატისტიკური მაჩვენებლებია უფრო შესაბამისი ჩვენი კვლევისათვის. გამოტანილ დიალოგურ ფანჯარაში დაკლიკეთ Continue (გაგრძელება) და შემდეგ OK (თანხმობა).

SPSS-ს მიერ გამოტანილი პირველი საშედეგო ფანჯარა აჩვენებს მონაცემების შეჯამებას, სრული და ნაკლულ მონაცემების რიცხვს. მეორე საშედეგო ცხრილია — კავშირის მიმართულების ინდიკატორები, ცხრილი 12.6. ცხრილის პირველ სვეტში მოცემულია ლამბდას მნიშვნელობები. როგორც უკვე ვიცით ლამბდას მნიშვნელობები დამოკიდებულია იმაზე, თუ რომელ ცვლადს ვირჩევთ, როგორც დამოკიდებულს ან დამოუკიდებელს. ჩვენს მაგალითში, აბორტის მიმართ აზრი არის დამოკიდებული ცვლადი, შესაბამისი ლამბდა 0.182-ია. ეს უკანასკნელი ნიშნავს ცვლადებს შორის სუსტ კავშირს. შეგვიძლია დავასკვნათ, რომ პირის ბიბლიის მიმართ დამოკიდებულების შესახებ ცოდნა ზრდის ალბათობას მოვახდინოთ მისი აბორტის მიმართ დამოკიდებულების პროგნოზირება 18.2% იანი გაუმჯობესებული პროგნოზით.

SPSS ასევე ითვლის სიმეტრიულ ლამბდას იმ მონაცემთა ცხრილებისათვის, სადაც არ არის დამოკიდებული და დამოუკიდებელი ცვლადები. აღნიშნული ცდება ამ წიგნის მოცულობას. (ჩვეულებრივ, მარტივი არაა ლამბდს ორი მნიშვნელობის გასაშუალოება). დამატებით SPSS გვაძლევს გუდმენისა და კრუსტალის ტაუ-ს სტატისტიკას, ორდანალურ ცვლადებს შორის კავშირის საზომს, მიუხედავად იმისა რომ ეს მაჩვენებლები ჩვენ SPSS-სგან არ მოგვითხოვია. ეს მაჩვენებლები ლამბდას გამოთვლისას ყოველთვის ავტომატურად გამოითვლება SPSS-ში.

ნახაზი 12.5 SPSS-ის დიალოგიური ფანჯარა-Statistics



ნახაზი 12.6 SPSS-ის საშედეგო ფანჯარა, ცვლადებს BIBLE და ABANY-ს შორის კავშირის გამოთვლა

Directional Measures

			Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	.093	.026	3.327	.001
		ABORTION IF WOMAN WANTS FOR ANY REASON Dependent	.182	.031	5.315	.000
		FEELINGS ABOUT THE BIBLE Dependent	.021	.042	.507	.612
	Goodman and Kruskal tau	ABORTION IF WOMAN WANTS FOR ANY REASON Dependent	.136	.020		.000[c]
		FEELINGS ABOUT THE BIBLE Dependent	.058	.009		.000[c]

თუ SPSS-ის დიალოგურ ფანჯარაში Statistics ავირჩევდით კრამერის V-ს, მიეიღებდით შემდეგ საშედეგო ფანჯარას. ნახ 12.7

ნახაზი 12.7 SPSS-ის საშედეგო ფანჯარა, ცვლადებს BIBLE და ABANY-ს შორის კავშირის გამოთვლა კრამერის V-ს მიხედვით

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.368	.000
	Cramer's V	.368	.000
N of Valid Cases		957	

კრამერის V 0.368-ის ტოლია, რაც ცვლადებს-ბიბლიის რწმენა და მხარდჭერა ლეგალური აბორტისადმი, შორის საშუალო სიძლიერის კავშირს ნიშნავს.

მაგალითი 2.

ჩვენ შეგვიძლია ასევე გამოვიყენოთ იგივე გამოთვლები გამას მნიშვნელობის მისაღებად. ამ მაგალითში განვიხილავთ კავშირს შემდეგ ცვლადებს შორის „განათლების დონე" (DEGREE) და „დამოკიდებულება ერთსქესიანი ურთიერთობების მიმართ" (HOMOSEX). რესპონდენტებისადმი დასმულ იქნა შეკითხვა მათი მოსაზრების თაობაზე ერთსქესიანი ურთიერთობების მიმართ, შემდეგი პასუხების არჩევანით: ყოველთვის მიუღებელია, თითქმის ყოველთვის მიუღებელია, ხშირად მიუღებელია, სულაც არ არის მიუღებელი.

SPSS-ის ბრძანებათა პანელზე დაკლიკეთ ბრძანებათა შემდეგ მიმდევრობას: Analyze – Descriptive Statistics – Crosstabs. კროსტაბულაციის დიალოგური ფანჯრის გამოტანის შემდეგ, სტრიქონში გადაიტანეთ ცვლადი DEGREE, ხოლო სვეტში HOMOSEX. სტატისტიკის ოპციის დაკლიკვის შემდეგ გამოტანილ დიალოგურ ფანჯარაში აირჩიეთ კენდალის ტაუ-b. შედეგად

მივიღებთ ორ საშედეგო ცხრილს, პირველი მათგანი ბივარიანტული ცხრილია HOMOSEX da DEGREE ცვლადებისათვის, ხოლო მეორე კავშირის სიმეტრიული საზომებია-გამა, კენდალის ტაუ-ხ (ნახ.12.8)

ნახაზი 12.8 SPSS-ის საშედეგო ფანჯარა, კავშირი HOMOSEX-სა და DEGREE-ს შორის

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	.226	.027	8.382	.000
	Gamma	.347	.040	8.382	.000
N of Valid Cases		950			

ცხრილის პირველ სტრიქონში კენდალის ტაუ-ხ-ს მნიშვნელობას მოსდევს გამას მნიშვნელობა. მოცემული მონაცემებისათვის კენდალის ტაუ-ხ-ს 0.226 და გამას 0.347 მნიშვნელობები ასახავს დადებით, საშუალო სიძლიერის კავშირს განათლების დონესა და ერთსქესიანი ურთიერთობების მიმართ დამოკიდებულებას შორის. როგორც ამ თავში უკვე განვიხილეთ კენდალის ტაუ-ხ-ს გამოთვლისას გამოიყენება როგორც დამოკიდებული, ასევე დამოუკიდებელი ცვლადის მიხედვით დამოკიდებული მონაცემთა წყვილები, შესაბამისად ამ მაჩვენებლის რიცხვითი მნიშვნელობა განსხვავდება გამას შესაბამისი მნიშვნელობისაგან. კენდალის ტაუ-ხ-ს მიხედვით, რესპონდენტის ერთსქესიანი ურთიერთობების მიმართ დამოკიდებულების პროგნოზირების შეცდომის შემცირებას 23%-ის შეეძლებით თუ გამოვიყენებთ რესპონდენტის განათლების შესახებ ინფორმაციას. იგივე პირობებშიც, გამას მიხედვით შეცდომის შემცირება 35%-ითაა შესაძლებელი. შევნიშნოთ, რომ ცვლად HOMOSEX-ის ჩაწერის ფორმიდან გამომდინარე, დადებითი გამა ნიშნავს, რომ განათლების ზრდასთან ერთად რესპონდენტების მიერ ერთსქესიანი ურთიერთობების მიმართ მხარდაჭერის ალბათობა იზრდება.

ამოცანები SPSS-სათვის

[GSS08PFP-A]

1. გამოიკვლიეთ კავშირი რესპონდენტის ჯანმრთელობის მდგომარეობასა (HEALTH) და სოციალურ კლასს (CLASS) შორის. სოციალური კლასი მიიჩნიეთ დამოუკიდებელ ცვლადად.
   ა. პროგრამაში გამოთვალეთ კავშირის შესაბამისი საზომი, ცვლადთა შორის კავშირის აღწერისათვის
   ბ. დაამატეთ რესპონდენტის სქესი, როგორც საკონტროლო ცვლადი და გამოთვალეთ შესაბამისი კავშირის საზომები. არის თუ არა ნახსენებ ცვლადთა შორის კავშირი ძლიერი, სუსტი ქალებისათვის? მამაკაცებისათვის?
   გ. სხვა რომელი საკონტროლო ცვლადის გამოყენება შეიძლება? გააგრძელეთ ჯანმრთელობის მდგომარეობასა და სოციალურ კლასს შორის კავშირის შესწვალა სხვა საკონტროლო ცვლადების დამატებით.
2. გამოკითხვებისას რესპონდენტებს შეეკითხნენ თუ რამდენად ეთანხმებოდნენ ისინი შეხედულებას იმის თაობაზე, რომ ზოგადად ქალები არ არიან პოლიტიკისათვის შესაბამისნი. (FEPOL). გაარკვიეთ თუ როგორ უკავშირდება ეს ცვლადი სქესს (SEX) და რესპონდენტის განათლებას (DEGREE). SPSS-ში გამოთვალეთ შესაბამისი მაჩვენებლები. იქნება

თუ არა გამასა და კენდალის ტაუ-b-ს გამოყენება გამართლებული FEPOL და DEGREE ცვლადებისათვის? პასუხი ახსენით.

3. შეისწავლეთ აბორტის შესახებ დამოკიდებულებასა და სხვა დემოგრაფიულ ცვლადებს (სქესი, ასაკი, რასა და ა.შ.) შორის კავშირი. გამოთვალეთ კავშირის შესაბამისი საზომები. მაგალითად, გაარკვიეთ არის თუ არა აბორტის მიმართ დამოკიდებულება განსხვავებული სქესის მიხედვით? შეისწავლეთ ამ ცვალდებს შორის კავშირი სხვა დამატებითი საკონტროლო ცვლადების ჩართვით. ამ მიზნებისათვის დაგჭირდებათ გამოიტანოთ ცალკე ცხრილები სქესისა და რასის მიხედვით. როდესაც ანალიზს დაამთავრებთ დაწერეთ მიღებული შედეგების შემაჯამებელი ანაგარიში. ჩამოაყალიბეთ შესაძლო მიზეზობრივი ფაქტორები არსებული სტატისტიკური კავშირების ასახსნელად.
4. რესპონდენტებს დაუსვეს კითხვა, თუ რას ფიქრობდენ ბავშვის მიმართ მკაცრი მეთოდების, როგორც აღზრდის საშუალების გამოყენებაზე. (SPARKING). დაადგინეთ თუ ნახსენებ ცვლადთან როგორ მიმართებაშია რესპონდენტთა პასუხები სქესის (SEX), სოციალური კლასის (CLASS) და ქორწინების სტატუსის (MARITAL) მიხედვით. თითოეული წყვილისათვის გამოითვალეთ შესაბამის მაჩვენებლები.
5. გამოიყენეთ MTF 2008-ის მონაცემები რათა დაწეროთ მოკლე მიმოხილვა სოციალურ საკითხებში გათვითცნობიერებულობასა (SOCCISSUE) და სხვა რასის მეგობრის ყოლაზე აზრს (ATFRIENDS) შორის კავშირზე. ცვლადი SOCCISSUE გამოიყენეთ, როგორც დამოუკიდებელი ცვლადი, ხოლო ATFRIENDS როგორც დამოკიდებული ცვლადი. გამოიყენეთ კავშირის ის ყველა საზომი, რომელიც ამ თავში ისწავლეთ. ასევე ახსენით ამ მაჩვენებლების მნიშვნელობებს შორის განსხვავების მიზეზები.

თავის სავარჯიშოები

1. ამ თავში ბევრი საკითხი განვიხილეთ. შესწავლილი მასალის მიხედვით უპასუხეთ შემდეგ კითხვებს:
ა. რაში გამოიყენება კავშირების საზომები?
ბ. კავშირის შეფასებისას არჩეული საზომების შემთხვევაში რა ფაქტორებს უნდა მიექცეს ყურადღება? რა განსხვავებაა გამა და კენტალის ტაუ-ბ მაჩვენებლებს შორის?
გ. ჩამოაყალიბეთ ორი პირობითი კვლევის ჰიპოთეზა, რომელთა მიმართ შესაძლებელია კავშირის საზომი მაჩვენებლების გამოყენება.
2. არის თუ არა სტატისტიკური კავშირი ძალადობასა და რასას, ასევე ძალადობის მსხვერპლსა და რასას შორის? აშშ-ს იუსტიციის დეპარტამენტის მონაცემები მოცემულია ქვემოთ (Expanded Homicide Data Table 5, 2007).

ძალადობის მსხვერპლი	ძალადობის ჩამდენი		
	თეთრკანიანი	შავკანიანი	გამა
დადებითი	2,918	566	51
უარყოფითი	245	2,905	11
სულ (*N*)	50	25	95

ა. ცვლადი „ძალადობის ჩამდენი" მიიჩნიეთ, როგორც დამოუკიდებელი ცვლადი, ხოლო ძალადობის მსხვერპლის რასა როგორც დამოკიდებული ცვლადი; ეცადეთ მოახდინოთ ძალადობის მსხვერპლის რასის პროგნოზირება, რამდენ პროგნოზირების შეცდომას მივიღებთ?
ბ. თუ მხედველობაში მივიღებთ დამოუკიდებელ ცვლადს, რამდენი იქნება ძალადობის მსხვერპლი თეთრკანიანი, შავკანიანი, სხვა რასის პირების პროგნოზირების შეცდომა?
გ. გამოიყენეთ ა და ბ დავალებებში მიღებული შედეგები და გამოთვალეთ შეცდომის შემცი-

რების პროპორციული მაჩვენებელი დამოუკიდებელ ცვლადზე დაყრდნობით. რა დასკვნას გამოიტანთ ცვლადებს შორის არსებული კავშირის შესახებ?

3. გავაგრძელოთ ძალადობის ჩამდენთა და ძალადობის მსხვერპლთა შესახებ მონაცემების ანალიზი. ქვემოთა ცხრილში მოყვანილია ძალადობის ჩამდენთა და ძალადობის მსხვერპლთა სქესის შესახებ ინფორმაცია:

ძალადობის მსხვერპლთა სქესი	ძალადობის ჩამდენთა სქესი	
	მამაკაცები	ქალები
მამაკაცები	4,451	497
ქალები	1,747	171

ა. ძალადობის ჩამდენთა სქესი მიიჩნიეთ დამოუკიდებელ ცვლადად, პროგნოზირების რამდენ შეცდომას მიუშვებდით თუ დამოუკიდებელი ცვლადის მნიშვნელობებს მხედველობაში არ მიიღებთ?

ბ. შეცდომის შემცირების რა მასშტაბი გვექნება დამოუკიდებელი ცვლადის მონაცემების გამოყენებით?

გ. გამოიყენეთ ა და ბ დავალებებში მიღებული შედეგები ლამბდას გამოთვლისათვის. განიხილეთ ამ ორ ცვლადს შორის კავშირი.

დ. რომელი ცვლადის ლამბდა მიანიშნებს უფრო ძლიერი კავშირის არსებობას: ძალადობის ჩამდენთა და ძალადობის მსხვერპლთა სქესს შორის, ძალადობის ჩამდენთა და ძალადობის მსხვერპლთა რასას შორის.

4. ზოგადი სოციოლოგიური კვლევისას (GSS 2008) რესპონდენტებს შეეკითხნენ აზრს სამუშაო ადგილზე ე.წ. აფირმატიული პოლიტიკის შესახებ (ქალთა, ეთნიკური უმცირესობების და ა.შ. წილის შეგნებული ზრდა) აზრი (AFFRMACT). შედეგები მოცემულია რესპონდენტების განათლების მიხედვით.

აფირმატული ქმედება	ბაკალავრის ხარისხზე ნაკლები განათლება	სულ ბაკალავრის ხარისხი	ჯამი
არ ემხრობა	761	274	1,035
ემხრობა	166	54	220
ჯამი	927	328	1,255

წყარო: General Social Survey, 2008

ა. ბაკალავრის ხარისხზე ნაკლები განათლების მქონე რესპონდენტების რა პროცენტული წილი უჭერს მხარს აფირმატიულ ღონისძიებებს? სულ მცირე ბაკალავრის ხარისხის მქონე რესპონდენტების რა პროცენტული წილი უჭერს მხარს აფირმატიული ღონისძიებებს?

ბ. გამოთვალეთ მსგავსი (Ns) და საპირისპირო წყობის (Nd) წყვილების რაოდენობები.

გ. მსგავსი და საპირისპირო წყობის წყვილების რაოდენობების გამოყენებით გამოთვალეთ გამას მნიშვნელობა. დადებითია თუ უარყოფითი გამას მნიშვნელობა? გამას მიხედვით მოახდინეთ აღნიშნულ ცვლადებს შორის კავშირის მნიშვნელოვნების ინტერპრეტირება.

5. ქვემოთ მოყვანილი ცხრილი ეფუძნება ზოგადი სოციოლოგიური კვლევის მონაცემებს (GSS 2008). განიხილეთ სოციალურ კლასსა და ჯანმრთელობის მდგომარეობის შორის კავშირი. რესპონდენტებს თხოვეს შეეფასებინათ თავიანთი ჯანმრთელობის მდგომარეობა.

ჯანმრთელობის მდგომარეობა	დაბალი და მშრომელი კლასის წარმომადგენლები	საშუალო და მაღალი კლასის წარმომადგენლები	ჯამი
ცუდი/საშუალო	218	134	352
შესანიშნავი/კარგი	470	519	989
ჯამი	688	653	1,341

წყარო: General Social Survey, 2008

ა. გამოთვალეთ Ns, Nd, Ty და Tx წყვილების რიცხვი.
ბ. მოცემული ცხრილისათვის ა დავალებაში მიღებული შედეგების გამოყენებით გამოთვალეთ კენდალის ტაუ-ხ. დადებითია თუ უარყოფითი კენდალის ტაუ-ხ-ს მნიშვნელობა? დარწმუნდით, რომ იცით თუ როგორაა ცვლადი კოდირებული. რა კავშირია სოციალურ კლასსა და ჯანმრთელობის მდგომარეობას შორის?

6. არის თუ არა კავშირი ქორწინების სტატუსსა და პიროვნების ზოგად ბედნიერებას შორის? ზოგადი სოციოლოგიური კვლევისას (GSS 2008) რესპონდენტებს თხოვეს შეეფასებინათ თავიანთი ბედნიერების ხარისხი 3 ქულიანი შკალის მიხედვით. ჩვენი მაგალითის გამარტივებისათვის ავირჩიოთ ორი კატეგორია: „ბედნიერი“, „ნაკლებად ბედნიერი“. მონაცემები ნაჩვენებია ქვემოთა ცხრილში.

ბედნიერების დონე	დაქორწინებული	ქორწინებაში არ მყოფი	ჯამი
ბედნიერი	888	405	1,293
ნაკლებად ბედნიერი	81	123	204
ჯამი	969	528	1,497

წყარო: General Social Survey, 2008

ა. არის თუ არა ამ მონაცემებისათვის ლამბდა ცვლადთა შორის კავშირის ადეკვატური მაჩვენებელი? გამოთვლების გარეშე განსაზღვრეთ ლამბდას შესაძლო მნიშვნელობა.
ბ. გამოთვლეთ გამას მნიშვნელობა და მოახდინეთ შედეგის ინტერპრეტირება.

7. „გეშინიათ თუ არა ღამით ქუჩაში სეირნობა თქვენი სახლის სიახლოვეს?“ ეს კითხვა დასმულ იქნა ზოგადი სოციოლოგიური კვლევისას (GSS 2006). შეისწავლეთ ამ ცვლადის განაწილება სქესისა და რასის მიხედვით. არის თუ არა ამ კითხვაზე პასუხები განსხვავებული სქესის და რასის მიხედვით?

ღამით სეირნობის შიში	სქესი		ჯამი
	მამაკაცები	ქალები	
(ა) თეთრკანიანი და შავკანიანი რესპონდენტები			
დიახ	196	523	719
არა	667	607	1,274
ჯამი	863	1,130	1,993

(ბ) თეთრკანიანი			
დიახ	122	360	482
არა	504	466	970
ჯამი	626	826	1,452
(გ) შავკანიანი			
დიახ	24	96	120
არა	75	74	149
ჯამი	99	170	269

წყარო: General Social Survey, 2008

ა. გამოიყენეთ შეცდომის პროპორციული შემცირების შესაბამისი მაჩვენებლები ცვალდებს შორის კავშირის შეფასებისათვის, მთლიანად და რესპონდენტთა რასის მიხედვით. შეეცადეთ ახსნათ არსებული განსხვავებები.

8. როცა გამოთვლები ეფუძნება დიქოტომიურ ცვლადებს ლამბდას გამოთვლილი მნიშვნელობა ხშირად უტოლდება ნულს. ხოლო ორდინალური ცვლადების კავშირის საზომი ინდიკატორების მნიშვნელობებზე არ აქვს გავლენა მოდალურ კატეგორიებს, როგორც ეს ლამბდას შემთხვევაშია. შესაბამისად, ორდინალური ცვალდების კავშირის საზომი უნდა გამოვიყენოთ იმ სახის მონაცემებისათვის, როგორც ეს მე-7 ამოცანაშია.

ა. გამოთვალეთ გამა და მიღებული მნიშვნელობის მიხედვით მოახდინეთ სქესსა და ღამით სეირნობის შიშს შორის კავშირის ინტერპრეტირება.

ბ. გამოიყენეთ ეს მაჩვენებელი იგივე ცვლადებს შორის კავშირის სიძლიერის შესაფასებლად.

9. ამ დავალებაში განვიხილავთ იმას თუ როგორ განსაზღვრავს ჩვენი განათლება ჩვენსავე სოციალურ დამოკიდებულებებს. კონკრეტულად კი განვიხილავთ განათლებასა და პოლიტიკურ სიმპატიებს შორის კავშირს. გამოთვალეთ გამას მნიშვნელობა ქვემოთ მოყვანილი ცხრილისათვის.

სიმპათია პოლიტიკური პარტიისადმი	ბაკალავრის ხარისხზე ნაკლები განათლება	ბაკალავრის და უფრო მაღალი ხარისხის	ჯამი
დემოკრატები	723	260	983
რესპუბლიკელები	453	213	666
ჯამი	1,176	473	1,649

წყარო: General Social Survey, 2008

10. გააგრძელეთ თქვენი ანალიზი განათლებასა და სხვა სოციალურ დამოკიდებულებებს შორის კავშირის შესახებ. ამჯერად, განათლებასა და სამხედრო ძალების მიმართ ნდობას შორის. (GSS 2008). კენდალის ტაუ-ბ გამოთვლის შემდეგ პასუხი გაეცით შემდეგ შეკითხვებს: არის თუ არა ძლიერი კავშირი განათლებასა და სამხედრო ძალების მიმართ ნდობას შორის? როგორია ამ კავშირის მიმართულება?

ნდობა სამხედრო ძალებისადმი	განათლება: ბაკალავრის ხარისხზე ნაკლები	განათლება: ბაკალავრის და უფრო მაღალი ხარისხის
მეტ-ნაკლები ნდობა	863	329
უმცირესი ნდობა	103	50
ჯამი	966	379

წყარო: General Social Survey, 2008

11. სოციოლოგიური კვლევისას (MTF 2008) მოზარდებს შეეკითხნენ მათი დამოკიდებულების შესახებ სხვადასხვა ნარკოტიკული საშუალებების მოხმარების მიმართ. ქვემოთ მოყვანილ ცხრილში ნაჩვენებია კოკაინისა და მარიჰუანას მოხარებისადმი მოზარდების აზრი. მაგალითის გამარტივებისათვის პასუხები: „მიუღებელია", „აბსოლიტურად მიუღებელია" ჩვენს შემთხვევაში ჩაწერილ იქნა როგორც „მიუღებელია".

	სქესი: მამაკაცები	სქესი: ქალები	ჯამი
(ა) მარიჰუანის მოხმარებისადმი დამოკიდებულება			
მისაღებია	325	302	627
მიუღებელია	345	417	762
ჯამი	670	719	1,389
(ბ) კოკაინის მოხმარებისადმი დამოკიდებულება			
მისაღებია	98	61	159
მიუღებელი	579	671	1,250
ჯამი	677	732	1,409

ა. კავშირის შეფასების რომელი საზომია ამ ცხრილისათვის შესაბამისი?
ბ. გამოთვალეთ კავშირის შესაბამისი საზომები. გამოთვლების შემდეგ განსაზღვრეთ არის თუ არა სქესსა და მარიჰუანას მოხმარებისადმი აზრს შორის კავშირი უფრო ძლიერი ვიდრე სქესსა და კოკაინის მოხმარებისადმი აზრს შორის?

12. ამ სავარჯიშოში განვიხილავთ კავშირს ბიბლიის რწმენასა და დამოკიდებულებას: ქორწინებამდელ სექს (PREMARSX), ჰომოსექსუალობის შესახებ აზრს (HOMOSEX) და პორნოგრაფიის ლეგალიზების შესახებ აზრს შორის (PORNLAW). როგორც ქვემოთ მოტანილ მონაცემებში ხედავთ რესპონდენტების პასუხები თითოეულ კითხვაზე ჩაწერილ იქნა დიქოტომიური ცვლადის სახით. გამოთვალეთ თითოეულ მონაცემთა ცხრილისათვის ლამბდა. რა დასკვნას გამოიტანთ? არის თუ არა ბიბლიისადმი რწმენა კავშირში სხვა ცვლადებთან?

ა) ბიბლიისადმი რწმენა და *PREMAX*

დამოკიდებულება ქორწინებამდელი სექსისადმი	ბიბლიის შესახებ აზრი: სულიერების საშყარო	ამბების თბრობა	ჯამი
მისაღებია	935	82	1,017
მიუღებელია	590	252	842
ჯამი	1,525	334	1,859

ბ) ბიბლიისადმი რწმენა და *HOMOSEX*

დამოკიდებულება ქორწინებამდელი სექსისადმი	ბიბლიის შესახებ აზრი: სულიერების საშყარო	ამბების თბრობა	ჯამი
მისაღებია	1,116	127	1,243
მიუღებელი	396	199	595
ჯამი	1,512	326	1,838

გ) ბიბლიისადმი რწმენა და *PORNLAW*

დამოკიდებულება ქორწინებამდელი სექსისადმი	ბიბლიის შესახებ აზრი: სულიერების საშყარო	ამბების თბრობა	ჯამი
მისაღებია	1,529	302	1,831
მიუღებელია	48	18	66
ჯამი	1,577	320	1,897

წყარო: General Social Survey, 2008

13. კანადაში, ონტარიოში მოზარდების სექსუალური ქცევის შესწავლისას რესპონდენტებს შეეკითხნენ იყენებდნენ თუ არა კონდომებს და ჩასახვის საწინააღმდეგო აბებს. ქვემოთ მოტანილია მონაცემები ბიჭებისა და გოგონებისათვის, ორი ასაკობრივი ჯგუფის მიხედვით.

ბიჭები (N=2,918)

კონდომის გამოყენების სიხშირე	მოზარდის ასაკი: 12-15	16-17	ჯამი
ყოველთვის პერიოდულობით	928	745	1,673
არასოდეს	912	333	1,245
ჯამი	1,840	1,078	2,918

გოგონები (N=2,824)

დაცვის საშუალების გამოყენების სიხშირე	მოზარდის ასაკი 12-15	16-17	ჯამი
ყოველთვის ან პერიოდულობით	856	1,010	1,866
არასოდეს	710	248	958
ჯამი	1,566	1,258	2,824

წყარო: ადაპტირებულია: B. Helen Thomas, Alba DiCenso and Lauren Griffith, "Adolescent Sexual Behavoir: Results from an Ontario Sample: Part II. Adolescent use of protection ," Canadian Journal of Public Health 89, no.2 (1998):94-97, Table 1, p.95

მოცემული ცხრილებისათვის გამოთვალეთ გამა და მოახდინეთ მიღებული შედეგის ინტერპრეტაცია. არის თუ არა გოგოებისა და ბიჭების მიხედვით განსხვავებული სურათი?

14. განვიხილოთ კავშირი აბორტის შესახებ აზრსა და ბავშვების რაოდენობას შორის (თუ ქალი ღარიბია და მატერიალურად არ შეუძლია დამატებითი ბავშვის რჩენა). GSS 2008.

აბორტის მხარდაჭერა	მოზარდის ასაკი არ ყავს ან ყავს ერთი ბავშვი	ორი ან მეტი ბავშვი	ჯამი
არა	279 49.8%	467 63.5%	746 57.6%
დიახ	281 50.2%	269 36.5%	550 42.4%
ჯამი	560 100%	736 100%	1,296 100%

გამოთვალეთ კავშირის შესაბამისი საზომი ამ ცხრილისათვის. არსებობს თუ არა კავშირი ბავშვების ყოლასა და აბორტის მხადაჭერას შორის? პასუხები დაასაბუთეთ.

თავი 13

რეგრესია და კორელაცია

თავის სწავლების მიზნები

- წრფივი კავშირი და პროგნოზირების წესები
- წრფივი გრაფიკების აგება და ინტერპრეტირება, რეგრესიის წრფის პოვნა
- a და b პარამეტრების გამოთვლა და ინტერპრეტირება
- პროგნოზირების შეცდომა და მისი შინაარსი
- დეტერმინაციის (r^2) და პირსონის (r) კოეფიციენტების გამოთვლა და ინტერპრეტირება
- მრავლობითი რეგრესია

ბევრი კვლევა ეფუძნება ინტერვალურ-ფარდობით მონაცემებს შორის კავშირის ანალიზს. სოციოლოგიაში მეცნიერები ხშირად იკვლევენ ისეთ ცვლადებს, როგორიცაა: განათლება, ოჯახების ზომა, შემოსავლები და ა.შ. დავუშვათ დაინტერესებულნი ვართ განათლების დონესა და ოჯახის საშუალო შემოსავლებს შორის კავშირის გარკვევით. ბივარიანტული რეგრესიული ანალიზი საშუალებას გვაძლევს გამოვიკვლიოთ ორ ინტერვალურ-ფარდობით მონაცემს შორის კავშირი საკმაო სიზუსტით[1].

ამერიკის შეერთებული შტატები უკვე საკმაოდ დიდი ხნის მანძილზე მიიჩნევა, როგორც ადამიანების შესაძლებლობების რეალიზების ქვეყანა, სადაც ყველას შეუძლია წარმატების მიღწევა. მაგრამ მიუხედავად ამ რწმენისა ბევრ სოციოლოგს მიაჩნია, რომ წარმატების მასშტაბები შეზღუდულია მხოლოდ იმ სოციალური შესაძლებლობებით, რასაც იძლევა სოციალური კლასი, რომელსაც პიროვნება დაბადებიდან მიეკუთვნება. პიროვნება, რომელიც დაბადებიდანვე მაღალ სოციალურ კლასს მიეკუთვნება აქვს უფრო მეტი წარმატების მიღწევის შესაძლებლობები, ვიდრე დაბალი სოციალური კლასის წარმომადგენელს. ამ დაშვებით შევამოწმოთ არის თუ არა აღნიშნული მოსაზრება რეალობა ამერიკული საზოგადოებისათვის.

აშშ-ს მოსახლეობის აღწერის ბიურო აგროვებს მონაცემებს ამერიკის შეერთებული შტატებისა და მისი მოსახლეობის შესახებ. საოჯახო მეურნეობების შემოსავლებისა და განათლების შესახებ მონაცემები აშშ-ს აღწერის ბიუროს რეგულარული დაკვირვების ობიექტია. ცხრილ 13.1-ში ნაჩვენებია შტატების მიხედვით ბაკალავრის განათლების ქონის პროცენტული მაჩვენებელები, 2006 წლისათვის 10 ყველაზე მეტად დასახლებული შტატის მიხედვით. ასევე ცხრილშივე ნაჩვენებია საშუალო, ვარიაციისა და დიაპაზონის მნიშვნელობები ამ მონაცემებისათვის.

ცხრილი 13.1 ბაკალავრის ხარისხის მქონე მოქალაქეთა პროცენტული წილი, შტატების მიხედვით

შტატი	ბაკალავრის ხარისხის მქონე მოქალაქეთა წილი
კალიფორნია	29.80
ტეხასი	25.50
ნიუ-იორკი	32.20
ფლორიდა	27.20
ილინოისი	31.20
პენსილვანია	26.60
ოჰაიო	23.30
მიჩიგანი	26.10
ჯორჯია	28.10
ნიუ-ჯერსი	35.60

საშუალო $\overline{Y} = \frac{\sum Y}{N} = \frac{258.6}{10} = 28.56$

ვარიაცია $Y = S_Y^2 = \frac{\sum (Y - \overline{Y})^2}{N-1} = \frac{120.3}{9} = 13.7$

დიაპაზონი $Y = 35.60 - 23.30 = 12.3$

წყარო: U.S. Census Bureau, Statistical Abstract of the United States, 2008, Table 221.

ცხრილ 13.1 მონაცემებისა და აღწერითი სტატისტიკის მაჩვენებლების მიხედვით შენიშნეთ თუ როგორი ცვალებადობაა განათლების დონესთან დაკავშირებით შტატების მიხედვით. კონკრეტული შტატის მოსახლეობის პროცენტული წილები ცვალებადობს მინიმალურ 23.30%-დან ოჰაიოსათვის, მაქსიმალურ ნიუ-ჯერსის 35.60%-მდე.

განათლებაში ამ დიფერენცირების ერთ-ერთი ამხსნელი ფაქტორი შესაძლებელია იყოს განსხვავებული ეკონომიკური განვითარების დონე. სოციოლოგები დიდი ხანია დავობენ იმის შესახებ, რომ ეკონომიკური სიკეთეების შეზღუდულობას გავლენა აქვს განათლების დონეზე. რადგან ეკონომიკური კეთილდღეობის ერთ-ერთი ძირითადი მაჩვენებელი საოჯახო მეურნეობების საშუალო შემოსავ-

ლებია, უნდა ვივარაუდოთ, რომ შტატებს მაღალი საშუალო შემოსავლებით, ასევე აქვთ განათლების კუთხით მაღალი მაჩვენებლები. ცხრილ 13.2-ში ნაჩვენებია საშუალო შემოსავლები 2006 წლისათვის, ყველაზე მეტად დასახლებული 10 შტატისათვის. შევნიშნოთ, რომ საშუალო შემოსავლები მერყეობს მინიმალური ოჰაიოს 44,532$-დან ნიუ-ჯერსის მაქსიმალურ 64,470$-მდე.

▣ ფერტილთა გაბნევის დიაგრამა

მოდით განვიხილოთ განათლებასა და საშუალო შემოსავლებს შორის შესაძლო კავშირის არსებობა აშშ-ს შტატების მაგალითზე. ყველაზე სწრაფი მეთოდი ცვლადებს შორის პოტენციურად არსებული კავშირის ვიზუალიზაციისათვის არის წერტილთა გაბნევის დიაგრამა. წერტილთა გაბნევის დიაგრამის გამოყენება პირველი ნაბიჯია რეგრესიულ ანალიზში. მონაცემების ვიზუალიზაცია საშუალებას იძლევა მივიღოთ პირველადი წარმოდგენა ცვლადებს შორის შესაძლო კავშირის შესახებ.

ცხრილი 13.2 საოჯახო მეურნეობის საშუალო შემოსავალი ($)

შტატი	საოჯახო მეურნეობის საშუალო შემოსავალი
კალიფორნია	56,645
ტეხასი	44,922
ნიუ-იორკი	51,384
ფლორიდა	45,495
ილინოისი	52,006
პენსილვანია	46,259
ოჰაიო	44,532
მიჩიგანი	47,182
ჯორჯია	46,832
ნიუ-ჯერსი	64,470

საშუალო $\bar{Y} = \frac{\sum Y}{N} = \frac{499,727}{10} = \$49,972.70$

ვარიაცია $Y = S_Y^2 = \frac{\sum (Y - \bar{Y})^2}{N-1} = \frac{367,421,523}{9} = \$40,824,614$

დიაპაზონი $Y = \$64,470 - \$44,532 = \$19,938$

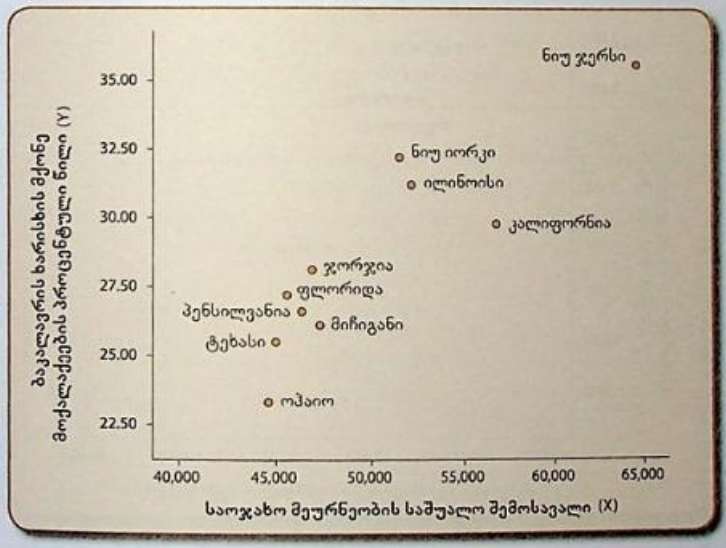
წყარო: US Census Bureau, American Community Survey, 2006.

წერტილთა გაბნევის დიაგრამა ინტერვალურ-ფარდობით მონაცემებს შორის არსებული კავშირის ვიზუალიზაციის მეთოდი.

13.1 ნახაზზე მოტანილ წერტილთა გაბნევის დიაგრამაზე ნაჩვენებია კავშირი განათლების დონესა და საოჯახო მეურნეობების საშუალო შემოსავლებს შორის, აშშ-ს ათი შტატისათვის. დიაგრამის X და Y ღერძებზე დატანილი ცვლადების მნიშვნელობები ადგენენ წერტილთა გაბნევის დიაგრამას. რადგან ჩვენივე დაშვებით, განათლების განსხვავებულ დონეებს სავარუდოდ განაპირობებს ეკონომიკური მდგომარეობის უთანაბრო პირობები, ამიტომ საშუალო შემოსავლები მივიჩნიოთ დამოუკიდებელ ცვლადად და დავიტანოთ ჰორიზონტალურ ღერძზე. დამოკიდებული ცვლადი- განათლება დატანილია ვერტიკალურ ღერძზე. ნახ. 13.1-ზე თითოეული წერტილი კონკრეტულ შტატს აღნიშნავს და ორივე ცვლადის თანაკვეთის ადგილზეა განთავსებული.

შევნიშნოთ, რომ ვიზუალურად აშკარად ჩანს გამოკვეთილი ტენდენცია: დაბალი შემოსავლების მქონე შტატებს (მაგ. ოჰაიო და ტეხასი) აქვთ ასევე დაბალი პროცენტული წილი ბაკალავრის განათლების მქონეთა რიცხვის მიხედვითაც. სხვაგვარად რომ ვთქვათ, ოჯახების საშუალო შემოსავლებსა და განათლებას შორის დადებითი კავშირია.

ნახაზი 13.1 წერტილთა გაბნევის დიაგრამა: განათლების დონე და ოჯახების საშუალო შემოსავლები



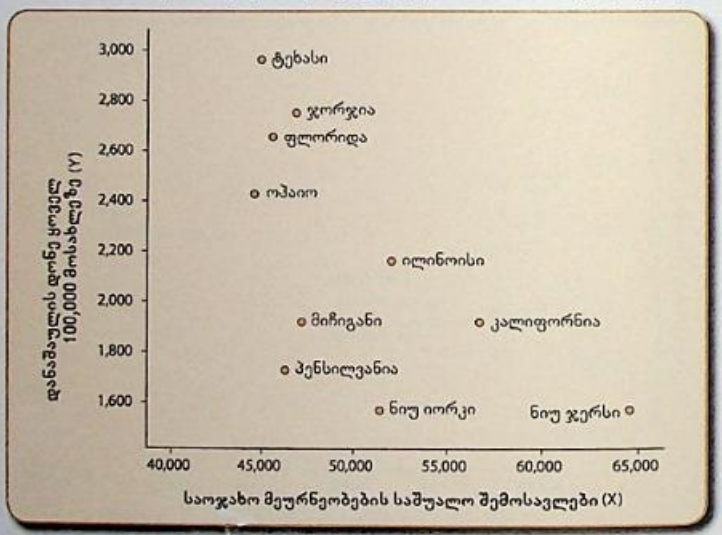
წერტილთა გაბნევის დიაგრამაზე ასევე შესაძლებელია ნაჩვენებ იქნას ცვლადებს შორის უარყოფითი კავშირიც. მაგალითად, ნახაზ 13.2-ზე ნაჩვენებია კავშირი საშუალო შემოსავლებსა და დანაშაულის დონეს შორის 10 ყველაზე მეტად დასახლებული შტატებისათვის. (მონაცემები აღებულია აშშ-ს სტატისტიკური წელიწდეულიდან). 13.2 ნახაზის მიხედვით დაბალი საშუალო შემოსავლები

უკავშირდება მაღალ დანაშაულის დონეს. საპირისპიროდ, მაღალი საშუალო შემოსავლები დაბალ დანაშაულის დონეს უკავშირდება. (მონაცემებისათვის იხილეთ ცხრილი 13.5).

ფრფივი კაშირი და პროგნოზირების წესები

წერტილთა გაბნევის დიაგრამები გვთავაზობენ ანალიზის სასარგებლო, მაგრამ საწყის ეტაპს. უფრო სიღრმისეული ანალიზისათვის გვჭირდება სისტემური მეთოდი ცვლადთა შორის კავშირის გამოთვლისათვის. მოდით ხელახლა მივუბრუნდეთ ნახ. 13.1 და 13.2-ს. ამ დიაგრამების მეშვეობით გვექმნება საწყისი წარმოდგენა განათლებასა და საშუალო შემოსავლებს, საშუალო შემოსავლებსა და დანაშაულის დონეს შორის კავშირის შესახებ. პირველ შემთხვევაში (ნახ.13.1) შემოსავლების ზრდა განათლების ზრდას უკავშირდება, ხოლო მეორე შემთხვევაში (ნახ.13.2) კი შემოსავლების ზრდას დანაშაულის შემცირება ახლავს.

ნახაზი 13.2 საოჯახო მეურნეობის საშუალო შემოსავლები და დანაშაულის დონე ყოველ 100,000 მოქალაქეზე



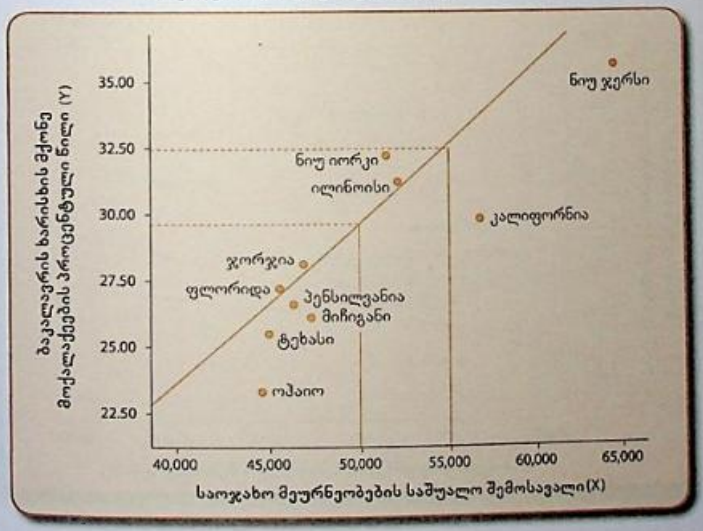
ერთი გზა ამ ცვალდებს შორის კავშირის შეფასებისათვის არის მათი გამოსახვა წრფივი კავშირის სახით. წრფივი კავშირი საშუალებას გვაძლევს წერტილოვან დიაგრამაზე გამოისახოს მონაცემებს შორის კავშირის ამსახველი შესაბამისი წრფე. სრულყოფილი წრფივი კავშირის შემთხვევაში კავშირის ამსახველი წრფე ემთხვევა წერტილთა გაბნევის დიაგრამის თანაკვეთის ყველა წერტილს

და ხშირად დეტერმინისტულ კავშირად იწოდება. ეს უკანასკნელი იძლევა Y-ის პროგნოზულ მნიშვნელობებს მოცემული X-ისათვის. მაგალითად, ნახაზ 13.3-ზე წერტილთა გაბნევის დიაგრამაზე მორგებულია წრფივი ხაზი, თავად წერტილთა გაბნევის დიაგრამა აღებულია ნახაზ 13.1-დან. ამ წრფის გამოყენებით ჩვენ შეგვიძლია მივიღოთ შტატების მოსახლეობის განათლების დონის (ბაკალავრის ხარისხის ფლობა) საპროგნოზო მაჩვენებლები მოცემული შემოსავლების დონისათვის. მაგალითად, ბაკალავრის ხარისხის მქონეთა პროცენტული წილი 50,000$-ის საშუალო შემოსავლების შემთხვევაში იქნება 30%-ზე ნაკლები. შესაბამისად, შტატი სადაც საშუალო შემოსავალი 55,000$-ია, ბაკალავრის ხარისხის მქონეთა წილი 32.50% იქნება.

წრფივი კავშირი ორ ინტერვალურ-ფარდობით ცვლადს შორის კავშირი, რომელიც შეიძლება გამოისახოს წერტილთა გაბნევის დიაგრამით და აღიწეროს წრფით.

დეტერმინისტული (სრულყოფილი) წრფივი კავშირი ორ ინტერვალურ-ფარდობით ცვლადს შორის კავშირი, რომლის მნიშვნელობები სრულად ემთხვევა წრფეს. კავშირის ამსახველი წრფის მიხედვით შესაძლებელია Y (ვერტიკალური ღერძი)-ის მნიშვნელობების მიღება ყოველი X (ჰორიზონტალური ღერძი)-ისათვის.

ნახაზი 13.3 განათლების დონე და ოჯახების საშუალო შემოსავლები, წრფივი გამოსახვა



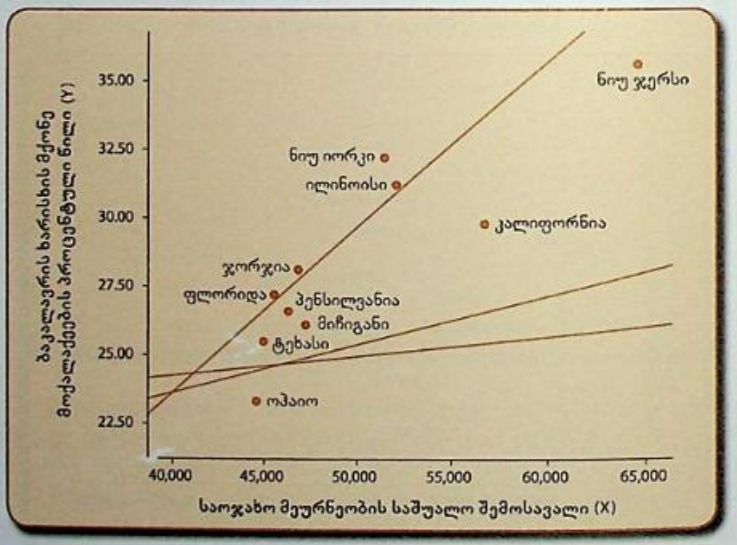
როგორც 13.3 ნახაზზეა ნაჩვენები, ათი შტატისათვის კავშირი შემოსავლების საშუალო დონესა და განათლებას შორის სრულყოფილად წრფივი ხასიათის არ არის. მიუხედავად იმისა, რომ ზოგი შტატის მაჩვენებელი ახლოსაა წრფესთან, არცერთი მათგანი ზუსტად წრფეზე არ მდებარეობს, ხოლო ნაწილი მათგანი საკმაოდაა წრფიდან დაშორებული. აღნიშნული შედეგის დანახვისას დაგვებადება კითხვა: არსებობს თუ არა სხვა უფრო შესაბამისი წრფე, რომელიც უფრო მეტი სიზუსტით აღწერს შემოსავლების საშუალო დონესა და განათლებას შორის კავშირს?

13.4-ზე ნაჩვენებია კიდევ ორი დამატებითი წრფე, რომლებიც წერტილთა გაბნევის დიაგრამის მიერ მოცემულ კავშირის სახეს შეესაბამებიან. შევნიშნოთ, მიუხედავად იმისა რომ, ცალკეული მაჩვენებლები ახლოსაა წრფეებთან ცალკეული შტატის მაჩვენებელები საკმაოდ დაშორებულია მოცემული წრფეებიდან. რომელია ამ წრფეებიდან ყველაზე მეტად შესაბამისი? როგორი უნდა იყოს ასეთი წრფის მახასიათებლები? სანამ უშუალოდ ჩამოვაყალიბებდეთ ყველაზე მეტად შესაბამისი წრფის აგების მეთოდს, მანამ განვიხილოთ წრფეების აგების რამდენიმე საბაზისო წესი.

✓ **ცოდნის შემოწმება**

გამოიყენეთ 13.3 ნახაზი, განსაზღვრეთ ბაკალავრის ხარისხის მქონე პირების პროცენტული წილი თუ საოჯახო მეურნეობის საშუალო შემოსავალი ა) 47,500$ და ბ) 50,000$ იქნება.

ნახაზი 13.4 განათლების დონე და ოჯახების საშუალო შემოსავლები, წრფივი კავშირის ამსახველი წრფეების სხვადასხვა ვერსიები



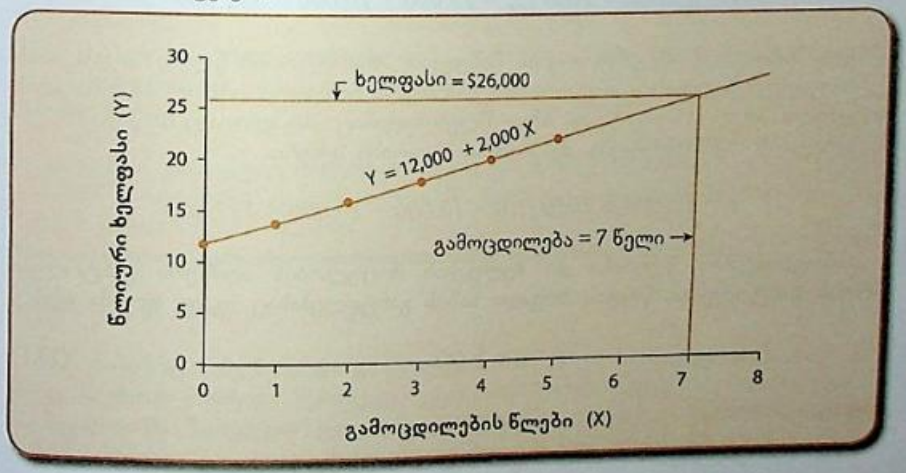
წრფეების აგება

წრფეების აგების საბაზისო საკითხების განხილვის მიზნით მოვიყვანოთ მარტივი მაგალითი. დავუშვათ ადგილობრივ სკოლაში მასწავლებლის ხელფასი მასწავლებლის გამოცდილების წლების მიხედვით განისაზღვრება. ახალი მასწავლებლის წლიური ხელფასი 12,000$-დან იწყება, თითოეული გამოცდილების წლის მიხედვით ხელფას 2000$ ემატება. მონაცემები ქვემოთა ცხრილშია მოყვანილი (ცხრილი 13.3).

ცხრილი 13.3 გამოცდილების წლები და ექვსი მასწავლებლის ხელფასი (პირობითი მონაცემები)

გამოცდილების წლები (X)	ხელფასი (დოლარებში) (Y)
0	12,000
1	14,000
2	16,000
3	18,000
4	20,000
5	22,000

ნახაზი 13.5 სრულყოფილი წრფივი კავშირი გამოცდილების წლების რაოდენობასა და წლიურ ხელფასს (ათას დოლარებში) შორის ექვსი მასწავლებლისათვის, (პირობითი მონაცემები)



ცხრილის მონაცემები გამოვსახოთ გრაფიკის სახით (ნაზ. 13.5). რადგან მუშაობის წლები განაპირობებს ხელფასის მოცულობას, ამიტომ ის დამოუკიდებელ ცვლადად ჩავთვალოთ და დავიტანოთ ჰორიზონტალურ (X) ღერძზე. ხელფასის მოცულობა დავიტანოთ ვერტიკალურ ღერძზე (Y). ექვსი მონაცემის თანაკვეთის წერტილების შეერთება გვაძლევს წრფეს (ნახ 13.5). ამ წრფის მიხედვით შესაძლებელია მარტივად განვსაზღვროთ ხელფასის მოცულობა სწავლების წლების მიხედვით, X ღერძის შესაბამისი წლების მნიშვნელობიდან მართობის აღმართით წრფის თანაკვეთამდე და შემდეგ ვერტიკალურ ღერძზე მართობის დაშვების გადაკვეთის წერტილიდან. მაგალითად, წრფეზე ნაჩვენებია 7 წლის გამოცდილების შესაბამისი ხელფასი. როგორც ნახაზზე ჩანს 7 წლის გამოცდილების შესაბამისი წლიური ხელფასი 26,000$-ია.

დამოკიდებულება ხელფასის ოდენობასა და სამუშაო გამოცდილების წლებს შორის, რომელიც ნაჩვენებია ცხრილი 13.3-სა და ნახ. 13.5, ასევე შესაძლებელია ჩაიწეროს განტოლების სახით:

$$Y = 12,000 + 2000\ (X)$$

სადაც,

X – სამუშაო გამოცდილების წლებია

Y – ხელფასის წლიური მოცულობა (დოლარებში)

ეს განტოლება საშუალებას გვაძლევს ზუსტად განვსაზღვროთ ხელფასის მოცულობა გამოცდილების წლების მიხედვით – წლების მაჩვენებლის განტოლებაში ჩასმით. მაგალითად, 5 წლიანი გამოცდილების მქონე მასწავლებლის ხელფასი იქნება:

$$Y = 12,000 + 2000\ (5) = 12,000 + 10,000 = \$22,000$$

შევნიშნოთ, რომ ამ განტოლების მიხედვით არამარტო იმ წლებისათვის არის შესაძლებელი ხელფასის მოცულობის გამოთვლა, რომელიც ცხრილ 13.3-ში არის მოცემული, არამედ ნებისმიერი სხვა წლებისთვისაც. მაგალითად, 10-წლიანი გამოცდილების მასწავლებლის შესაბამისი ხელფასი იქნება:

$$Y = 12,000 + 2000\ (10) = 12,000 + 20,000 = \$32,000$$

გამოცდილების წლებისა და ხელფასის მოცულობის აღმწერი განტოლება წრფის განტოლებაა. წრფის ზოგადი სახის განტოლებასაც იგივე ფორმა აქვს:

$$Y = a + b\ (X) \qquad (13.1)$$

სადაც,

Y = დამოკიდებული ცვლადის გამოთვლილი მნიშვნელობაა

X = დამოუკიდებელი ცვლადის მნიშვნელობაა

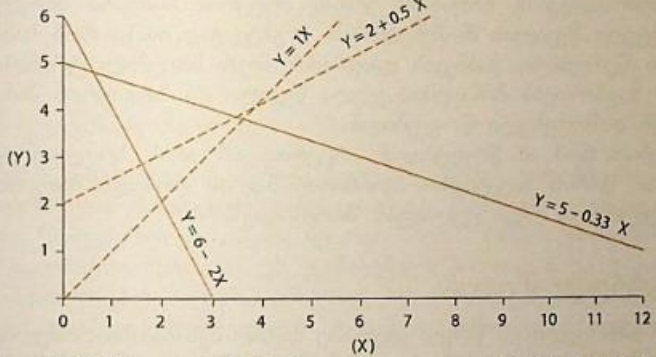
a = ღერძთან თანაკვეთაა; ე. ი. a არის Y-ის მნიშვნელობა, როცა X = 0.

b = რეგრესიის წრფის დახრილობაა, ან სხვაგვარად Y-ის ცვლილება, როცა X ერთი ერთეულით იცვლება. ჩვენს მაგალითში, a = 12,000 და b = 2,000. ეს ნიშნავს, რომ სამუშაო გამოცდილების არქმონე მასწავლებელი მიიღებს 12,000$ ხელფასს და შემდეგ ყოველ წელზე დაემატება 2000$.

ქვემოთ მოყვანილი ოთხი წრფის მიხედვით განსაზღვრეთ Y-ის მიმართულება, როცა X იზრდება ერთი ერთეულით? ამ კითხვაზე პასუხი დაასაბუთეთ, როგორც წრფის გრაფიკული გამოსახულებით, ასევე წრფის განტოლებითაც.

✓ *ცოდნის შემოწმება*

ოთხი წრფე: დახრილობისა და Y თანაკვეთის ილუსტრირება



დახრილობა (b) Y ცვლადის (დამოკიდებული ცვლადის) მნიშვნელობის ცვლილება X-ის (დამოუკიდებელი ცვლადის) ერთი ერთეულით ცვლილებისას

Y თანაკვეთა (a) წრფის Y ღერძთან თანაკვეთის წერტილი, Y-ის მნიშვნელობა, როცა X = 0.

გამოიყენეთ განტოლება, რომელიც აღწერს კავშირს გამოცდილების წლებსა და ანაზღაურებას შორის და გამოთვალეთ 12 წლის გამოცდილების მქონე მასწავლებლის მოსალოდნელი ხელფასი.

✓ *ცოდნის შემოწმება*

რეგრესიის წრფის პოვნა

13.5 ნახაზზე ნაჩვენები და განტოლებით ჩაწერილი (*Y = 12,000 + 2000 (X)*) წრფე საკმაოდ მარტივი შემთხვევაა ცვლადთა შორის კავშირის აღწერისათვის. ამ მაგალითში საშედეგო ცვლადის მნიშვნელობა სრულად განისაზღვრებოდა დამოუკიდებელი ცვლადით. მონაცემთა ყველა წყვილი უშუალოდ წრფეზე მდე-

ბარეობს, ანუ ცვლადებს შორის დეტერმინისტული, სრულყოფილი წრფივი კავშირია.

ცვლადებს შორის კავშირების უმრავლესობა, რომლებიც სოციალურ მეცნიერებებში შეისწავლება არ არის სრულყოფილი წრფივი კავშირის ტიპის. უმრავლეს შემთხვევებში ჩვენ შევხვდებით მაგალითებს, როცა ცვალდებს შორის კავშირი უახლოვდება წრფივ კავშირს, მაგრამ სრულყოფილი წრფივი კავშირი არ არის. მაგალითად, რეალობაში მასწავლების ხელაფასი მხოლოდ ნამუშევარი წლებით არ განისაზღვრება, შესაბამისად მარტო ნამუშევარი წლების მიხედვით ვერ განვსაზღვრავთ მასწავლებლის ხელფას რეალობაში.

როცა დამოკიდებული ცვლადის მნიშვნელობა (Y) სრულად არ განისაზღვრება მარტო დამოუკიდებელი ცვლადით (X), ასეთ შემთხვევაში მონაცემთა ყველა წყვილი კავშირის ამსახველ წრფეზე არ მდებაროებს. დააკვირდით კვლავ ნახაზ 13.4-ს განათლებასა და საშუალო შემოსავლებს შორის კავშირის შესახებ. მიუხედავად იმისა, რომ თითოეული წრფე აჩვენებს განათლების დონის ზრდას საშუალო შემოსავლების ზრდასთან ერთად, არცერთი მათგანი სრულყოფილად არ ასახვს ყველა წყვილის მნიშვნელობას. მიუხედავად, იმისა რომ სამივე წრფე უახლოვდება წერტილთა გაბნევის დიაგრამის მიერ მოცემულ ტრენდს, არცერთ წრფეზე არ მდებაროებს მონაცემთა ყველა წყვილი და ზოგიერთი მათგანი საკმაოდაც არის დაშორებული წრფეებიდან.

იმ დაშვებით, რომ ამ წრფეებიდან არცერთი არ არის სრულყოფილი, ჩვენი ამოცანაა მათ შორის საუკეთესო ავირჩიოთ. მაგრამ რომელი მათგანი აღწერს სხვასთან შედარებით უკეთ ცვლადებს შორის კავშირს?

დეტერმინისტული შეცდომა

კავშირის ამსახველი ის წრფეა ყველაზე მეტად შესაბამისი, რომელიც იძლევა შეცდომების მინიმალურ რიცხვს, ეს უკანასკნელი იწოდება ასევე ნაშთად. მოდით დავფიქრდეთ, რას ნიშნავს ნაშთი. დააკვირდით კვლავ ნახაზ 13.3-ს. ნახაზზე დაყრდნობით შემოსავლის თითოეული დონისათვის (X) შესაძლებელია მივიღოთ განათლების შესაბამისი დონე (Y). მაგალითად, ტეხასის საშუალო შემოსავალს – 44,922$, განათლების 26% შეესაბამება. მაგრამ მისი რეალური მნიშვნელობა 25.5%-ია (იხილე ცხრილი 13.1). შესაბამისად, გვაქვს Y-ის ორი მნიშვნელობა: 1. გამოთვლილი Y, რომელსაც ავღნიშნავთ როგორც $\hat{Y}$ და რომელიც მიიღება განტოლების (ასევე იწოდება წრფივ რეგრესიად) $Y = a + b(X)$ მეშვეობით; 2. დაკვირვების მნიშვნელობა, Y. ე. ი. ტეხასისათვის $\hat{Y}$ = 26% და Y=25.5%.

ნაშთი არის დაკვირვებისა (Y) და გამოთვლის შედეგად ($\hat{Y}$) მიღებულ მნიშვნელობებს შორის სხვაობა. თუ ნაშთს e სიმბოლოთი აღვნიშნავთ, მაშინ ნაშთი შემდეგნაირად გამოითვლება:

$$e = Y - \hat{Y}$$

ტეხასისათვის ნაშთი იქნება: 25.5%-26%=-0.5.

სხვაობების კვადრატთა ჯამი ($\sum e^2$)

რეგრესიის სავარაუდო განტოლებისა და რეგრესიის წრფის არჩევის კრიტერიუმს წარმოადგენს ნაშთების (e) მინიმალური მნიშვნელობა. რეგრესიის ცალკეული წრფეები ამცირებენ ნაშთის მნიშვნელობას დაკვირვების მონაცემების ერთი ჯგუფისათვის, ხოლო სხვებისათვის ეს მნიშვნელობები იზრდება. ჩვენთვის მისაღებია რეგრესიის ისეთი გატოლების პოვნა, რომელიც სხვაობის ნაშთებს შეამცირებს დაკვირვების ყველა მონაცემებისათვის.

ნაშთის მნიშვნელობის გამოთვლისათვის არსებობს რამდენიმე მათემატიკური მეთოდი. მაგალითად, შეგვიძლია გამოვიყენოთ: ნაშთების ალგებრული ჯამი $\sum(Y-\hat{Y})$, ნაშთების აბსოლუტური მნიშვნელობების ჯამი $\sum(|Y-\hat{Y}|)$ ან ნაშთების კვადრატების ჯამი $\sum(Y-\hat{Y})^2$. მათემატიკური მოსაზრებების გამო სტატისტიკოსები ამჯობინებენ გამოიყენონ მესამე მეთოდი: ნაშთების კვადრატების ჯამი. ნაშთების კვადრატების ჯამი აღინიშნება, როგორც $\sum e^2$ და გამოითვლება შემდეგი განტოლებით:

$$\sum e^2 = \sum (Y - \hat{Y})^2$$

უმცირეს კვადრატთა წრფე

რეგრესიის ამსახველი ის წრფეა ყველაზე მეტად მისაღები, რომელიც ნაშთების კვადრატების მინიმალურ ჯამს გვაძლევს. ასეთი წრფე იწოდება უმცირეს კვადრატთა წრფედ, ხოლო თავად მეთოდი ასეთი წრფის მისაღებად იწოდება უმცირეს კვადრატთა მეთოდად. ეს უკანასკნელი მოიცავს a და b პარამეტრების ისეთი მნიშვნელობების პოვნას, რომლებიც $Y = a + bx$ განტოლებისათვის $\sum e^2$-ის მინიმალურ შესაძლო მნიშვნელობას გვაძლევს. სანამ გავაგრძელებდეთ, მოდით განვიხილოთ უკვე რა ვისწავლეთ.

უმცირეს კვადრატთა წრფე წრფე, რომელიც ნაშთების კვადრატების უმცირეს ჯამს იძლევა, ე. ი. $\sum e^2$-ის მნიშვნელობა უმცირესია.

უმცირეს კვადრატთა მეთოდი უმცირეს კვადრატთა წრფის მიღების მეთოდი.

მიმოხილვა

1. ათი შტატის მონაცემების მიხედვით განვიხილეთ კავშირი საოჯახო მეურნეობის საშუალო შემოსავლებსა და ბაკალავრის ხარისხის მქონეთა პროცენტულ წილს შორის. ამ ცვლადებს შორის კავშირის ვიზუალურად გამოსახვისათვის წერტილთა გაბნევის დიაგრამა გამოვიყენეთ.

2. წერტილთა გაბნევის დიაგრამის მიხედვით ვივარაუდეთ, რომ ცვალდებს შორის კავშირი შესაძლებელია წრფივი ხასიათის ყოფილიყო. საშუალო შემოსავლების ზრდასთან ერთად იზრდება ბაკალავრის ხარისხის მქონეთა რიცხვიც.
3. უფრო სისტემური მეთოდი ცვლადებს შორის კავშირის შეფასებისათვის არის წრფივი განტოლებისა და კავშირის ამსახველი წრფივი გრაფიკის გამოყენება. როგორც ვნახეთ მონაცემთა გაბნეული წყვილებისათვის რამდენიმე წრფე შეიძლება მოიძებნოს.
4. კავშირის ამსახველი წრფეებიდან ყველაზე მეტად მისაღები ის წრფეა, რომლის შემთხვევაშიც ნაშთის $\sum e^2$ მნიშვნელობა მინიმალურია. ასეთი წრფე უმცირეს კვადრატთა წრფედ იწოდება და მოიცავს a და b პარამეტრების ისეთი მნიშვნელობების პოვნას, რომლებიც $\hat{Y} = a + bx$ განტოლებისათვის $\sum e^2$-ის მინიმალურ შესაძლო მნიშვნელობას გვაძლევს.

a და b პარამეტრების გამოთვლა რეგრესიის განტოლებისათვის

კალკულუსის გამოყენებით შეგვიძლია ვაჩვენოთ, რომ a და b პარამეტრების ისეთი მნიშველობების პოვნისათვის, რომლთათვის $\sum e^2$-ის მნიშვნელობა უმცირესია, დაგვჭირდება შემდეგი ფორმულების გამოყენება:

$$a = \overline{Y} - b(\overline{X}) \qquad (13.3)$$

სადაც,

S_{yx} = X და Y შორის კოვარიაციის მნიშვნელობა

S_x^2 = X-ის ვარიაციის მნიშვნელობა

$\overline{Y}$ = Y-ის საშუალო მნიშვნელობა

$\overline{X}$ = X-ის საშუალო მნიშვნელობა

a – ღერძთან თანკვეთის მნიშვნელობა

b – წრფის დახრილობა

ამ ფორმულების გამოყენებისას გაკეთებულია დაშვება, რომ X დამოუკიდებელი ცვლადია, ხოლო Y დამოკიდებული ცვლადია.

სანამ a და b-ს უშუალო გამოთვლამდე გადავიდოდეთ განვიხილოთ ზემოთ მოტანილი ფორმულები. b-ს ფორმულის მნიშვნელი X-ის ვარიაციაა, რომელიც თავის მხრივ შემდეგი ფორმულით გამოითვლება:

$$\text{ვარიაცია } (X) = S_X^2 = \frac{\sum (X - \overline{X})^2}{N-1}$$

ეს ფორმულა თქვენთვის უკვე ნაცნობია მე-5 თავიდან. მისი მრიცხველი კი ჩვენთვის სიახლეა. ეს უკანასკნელი წარმოადგენს X და Y-ს შორის კოვარიაციას და განისაზღვრება შემდეგნაირად:

$$\text{კოვარიაცია } (X,Y) = S_{YX} = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{N-1} \qquad (13.4)$$

კოვარიცია ასახავს თუ როგორ იცვლიან X და Y მნიშვნელობებს ერთმანეთთან მიმართებაში. კოვარიაციის საბაზისო შინაარსი შემდეგია: იგი ასახავს თუ რა ხარისხით იღებს ერთი ცვლადის მაღალი მნიშვენლობების შესაბამისად მეორე ცვლადი მაღალ (ამ შემთხვევაში გვაქვს დადებითი კოვარიაცია) ან დაბალ მნიშვნელობებს (უარყოფითი კოვარიაცია). განვიხილოთ ეს ფორმულა, მასში აღებულია სხვაობა X და Y-ის საშუალო და დაკვირვების ფაქტობრივ მონაცემებს შორის, შემდეგ მათი ჯამის ნამრავლი გაყოფილია N - 1-ზე.

ცხრილ 13.4-ში ნაჩვენებია a და b პარამეტრების გამოთვლები 10 შტატის მონაცემებისათვის. შემოსავლებისა და ბაკალავრის ხარისხის მქონეთა საშუალო რიცხვი მიიღება 1 და 2 სვეტების დაჯამებით და N-ზე გაყოფით. კოვარიაციის გამოთვლისათვის, უნდა ვიპოვოთ და X-სა, და Y-ს შორის სხვაობა (სვეტები: 3 და 5); მიღებული სხვაობები გავამრავლოთ (სვეტი 7). მაგალითად, დაკვირვების პირველი მონაცემი, კალიფორნიისათვის შემოსავლების საშუალო გადახრა არის 6,672.3 (56,645 - 49,972.7 = 6,672.3); ხოლო ბაკალავრის ხარისხის მქონეთა რიცხვისათვის იგივე მაჩვენებელი 1.24-ია (6,672x1.24=8,273.65). ამ მნიშვნელობების ჯამი მე-7 სვეტის ბოლოშია ნაჩვენები – 186,591.26-ია. ამ უკანასკნელის 9 (N - 1)-ზე გაყოფით მივიღებთ კოვარიაციის მნიშვნელობას – 20,732.36.

ცხრილი 13.4 რეგრესიის განტოლების a და b პარამეტრების გამოთვლა

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
შტატი	საოჯახო შეურნეობის საშუალო შემოსავალი X	ბაკალავრის ხარისხის მქონეთა პროცენტული წილი Y	$(X-\overline{X})$	$(X-\overline{X})^2$	$(Y-\overline{Y})$	$(Y-\overline{Y})^2$	$(X-\overline{X})(Y-\overline{Y})$
კალიფორნია	56,645	29.8	6,672.3	44,519,587	1.24	1.54	8,273.65
ტეხასი	44,922	25.5	-5,050.7	25,509,570	-3.06	9.36	15,455.14
ნიუ-იორკი	51,384	32.2	1,411.3	1,991,767	3.64	13.25	5,137.13
ფლორიდა	45,495	27.2	-4,477.7	20,049,797	-1.36	1.85	6,089.67

ილინოისი	52,006	31.2	2,033.3	4,134,309	2.64	6.87	5,367.91
პენსილვანია	46,29	26.6	-3,713.7	13,791,568	-1.96	3.84	7,278.85
ოჰაიო	44,532	23.3	-5,440.7	29,601,216	-5.26	27.67	28,618.08
მიჩიგანი	47,182	26.1	-2,790.7	7,788,006	-2.46	6.05	6,865.12
ჯორჯია	46,832	28.1	-3,140.7	9,863,996	-0.46	0.21	1,444.72
ნიუ-ჯერსი	64,470	35.6	14,497.3	210,171,707	7.04	49.56	102,060.99
	$\sum X = 499,727$	$\sum Y = 285.6$	0.00	367,421,526	0.001	120.3	186,591.26

საშუალო $X = \overline{X} = \frac{\sum X}{N} = \frac{499,727}{10} = 49,972.7$

საშუალო $Y = \overline{Y} = \frac{\sum Y}{N} = \frac{285.6}{10} = 28.56$

ვარიაცია $(Y) = S_Y^2 = \frac{\sum (Y - \overline{Y})^2}{N-1} = \frac{120.3}{9} = 13.37$

სტანდარტული გადახრა $(Y) = S_Y = \sqrt{13.37} = 3.66$

ვარიაცია $(X) = S_Y^2 = \frac{\sum (X - \overline{X})^2}{N-1} = \frac{367,421,523}{9} = 40,824,614$

სტანდარტული გადახრა $(X) = S_X = \sqrt{40,824,614} = 6,389.41$

კოვარიაცია $(X,Y) = S_{YX} = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{N-1} = \frac{186,591.26}{9} = 20,732.36$

ა. პასუხები შეიძლება დამრგვალების გამო განსხვავდებოდეს ცალკეულ შემთხვევებში, ამ სვეტების ზუსტი ჯამი ყოველთვის ნოლის ტოლი იქნება

კოვარიაცია ცვლადებს შორის წრფივი კავშირის საზომია და გვიჩვენებს როგორც ცვალდებს შორის კავშირის სიძლიერეს, ასევე მის მიმართულებას. კოვარიაციის მნიშვნელობა ნულთან ახლოსაა, როცა ცვლადებს შორის კავშირი არ არსებობს, და პირიქით თუ ცვლადებს შორის დადებითი კავშირია კოვარიაციის მნიშვნელობა ნოლზე მეტია და ნულზე ნაკლებია, როცა კავშირი უარყოფითია.

გამოვიყენოთ კოვარიაციისა და ვარიაციის მნიშვნელობები 13.4 ცხრილიდან b პარამეტრის გამოთვლისათვის:

$$b = \frac{S_{YX}}{S_X} = \frac{20,732.36}{40,824,614} = 0.0005$$

მას შემდეგ, რაც გამოვთვალეთ b-ს მნიშვნელობა, a თანაკვეთის პოვნა შედარებით ადვილია:

$$a = \overline{Y} - b\,(\overline{X}) = 28.56 - 0.0005\,(49,972.7) = 3.57$$

რეგრესიის განტოლებაა:

$$\hat{Y} = 3.57 + 0.0005\,(X)$$

ეს განტოლება შესაძლებელია გამოვიყენოთ მოცემული შემოსავლების შესაბამისი განათლების დონის გამოთვლისათვის. მაგალითად, შტატს, სადაც საშუალო შემოსავალი 48,000$ იქნება, ბაკალავრის ხარისხის მქონეთა შესაბამისი რიცხვი იქნება:

$$\hat{Y} = 3.57 + 0.0005\,(48,000) = 27.57$$

ასევე, შტატს 56,000$-ის საშუალო შემოსავლებით ბაკალავრის ხარისხის მქონეთა შემდეგი რიცხვი ექნება:

$$\hat{Y} = 3.57 + 0.0005\,(56,000) = 31.57$$

ახლა უკვე შეგვიძლია დავხაზოთ რეგრესიის განტოლების შესაბამისი წრფე. წრფის დახაზვისათვის გვჭირდება ვიპოვოთ ორი შესაბამისი წერტილთა კოორდინატები. ამ მიზნით ზემოთ ნაპოვნი ორი მნიშვნელობა შეგვიძლია გამოვიყენოთ: 1. $X = \$48,000, Y = \%27.57$ და 2. $X = \$56.000, Y = \%31.5$. 13.6 ნახაზზე ნაჩვენებია აგებული რეგრესიის წრფე, რომელიც დატანილია ჩვენს მიერ თავდაპირველად წერტილთა გაბნევის დიაგრამაზე (ნახ.13.1).

გამოიყენეთ რეგრესიის განტოლება ნიუ-იორკის, ჯორჯიის და ოჰაიოს Y მნიშვნელობების საპოვნელად. შეამოწმეთ გადის თუ არა ამ წერტილებზე ნახაზ 13.6-ზე მოცემული წრფე.

✓ ცოდნ
შემოწმება

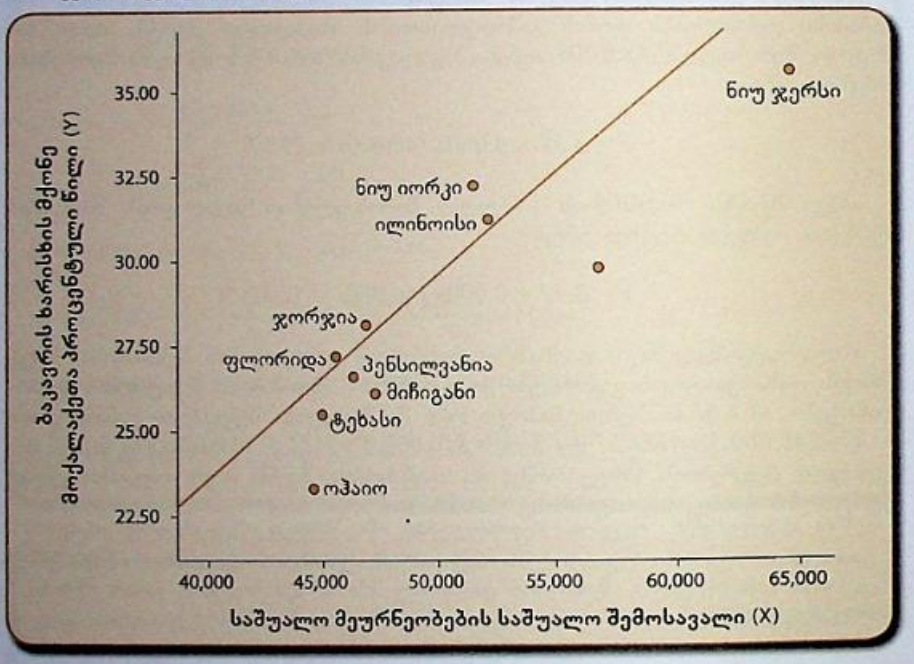
a და b კოეფიციენტების ინტერპრეტირება

რეგრესიის განტოლების დადგენის შემდეგ მოვახდინოთ a და b კოეფიციენტების ინტერპრეტირება. b კოეფიციენტის მნიშვნელობა 0.0005%-ია, რაც ნიშნავს რომ შტატის საშუალო შემოსავლების 1$-ით ზრდა ბაკალავრის განათლების მქონეთა რიცხვს 0.0005%-ით გაზრდის. ასევე, შტატის საშუალო შემოსავლების 10,000$-ით ზრდა, 5%-ით გაზრდის ბაკალავრის ხარისხის მქონეთა რიცხვს.

აქვე შევნიშნოთ, რომ ცვლადებს შორის კავშირი სოციალურ მეცნიერებებში არ არის აბსოლიტური სიზუსტის და ყოველ კონკრეტულ შემთხვევაში მიღებული შედეგი სიზუსტის გარკვეულ დონეს ასახავს. რეგრესიის განტოლებით მოცემული დამოკიდებულება შესაძლოა არ ასახავდეს ყველა კერძო შემთხვევას,

მაგრამ ის გვაძლევს წარმოდგენას იმის შესახებ თუ საშუალოდ როგორ შეიცვლება ბაკალავრის განათლების მქონეთა რიცხვი საშუალო შემოსავლების ზრდის შედეგად. შეგვიძლია ვთქვათ, რომ წრფის დახრილობა 0.0005% არის ცვლადებს შორის კავშირის შეფასება.

ნახაზი 13.6 საშუალო შემოსავლებსა და განათლების დონეს შორის კავშირის ამსახველი რეგრესიის წრფე



ღერძთან თანაკვეთა a არის Y-ის მნიშვნელობა როცა X=0, ე.ი. ეს არის რეგრესიის წრფის Y ღერძთან თანაკვეთა. თუ a = 3.57 ე.ი მოქალაქეთა მხოლოდ 3.57% ექნება ბაკალავრის ხარისხი, როცა საშუალო შემოსავალი 0-ის ტოლია. მაგრამ რეალობაში არცერთ შტატში არ არის საშუალო შემოსავალი ნულის ტოლი. ზოგადად, სიფრთხილით მოეკიდეთ Y-ის იმ მნიშვნელობების გამოთვლას, რომელთათვის X გადის მნიშვნელობათა დიაპაზონიდან. ე.ი. თუ X-ის მნიშვნელობა ნოლთან ახლოსაა, თანაკვეთას რაციონალური ინტერპრეტირება ვერ მოეძებნება.

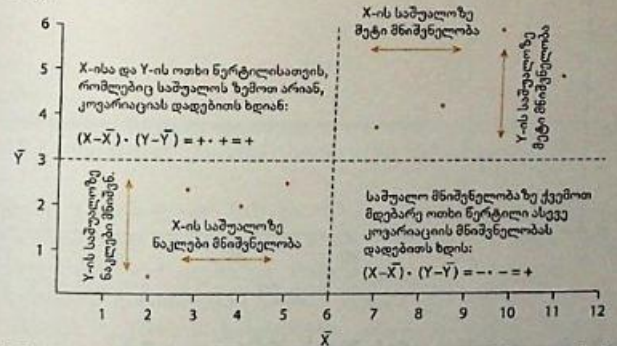
Y ღერძთან თანაკვეთა (a) წერტილი, სადაც რეგრესიის წრფე კვეთს Y ღერძს და X=0.

სტატისტიკა პრაქტიკაში: საოჯახო მეურნეობების საშუალო შემოსავლები და პრიმინალური ქმედებები

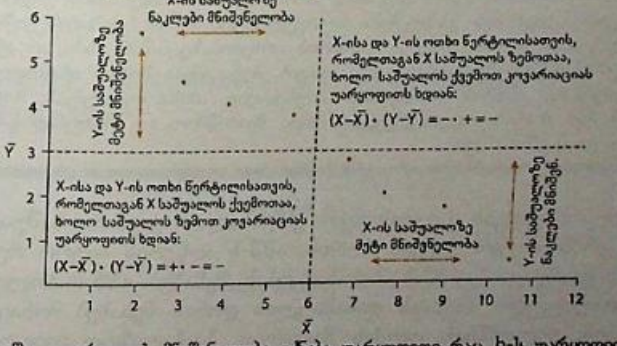
ჩვენს წინა მაგალითში განვიხილეთ კავშირი საშუალო შემოსავლებსა და განთლების დონეს შორის. რეგრესიის განტოლებამ, რომელიც ათი შტატის მონაცემების მიხედვით შევადგინეთ, გვაჩვენა რომ საშუალო შემოსავლების ზრდასთან ერთად იზრდება ბაკალავრის ხარისხის მქონეთა რიცხვიც. ეს დასკვნა ამტკიცებს აზრს განათლების დონესა და საშუალო შემოსავლებს შორის კავშირის არსებობის თაობაზე.

დეტალების განხილვა 13.1
კოვარიაციის შინაარსი

დავუშვათ, გვაქვს მონაცემთა რვა წყვილი X, რომლის საშუალო 6-ია, ხოლო Y-ის საშუალო 3-ია.



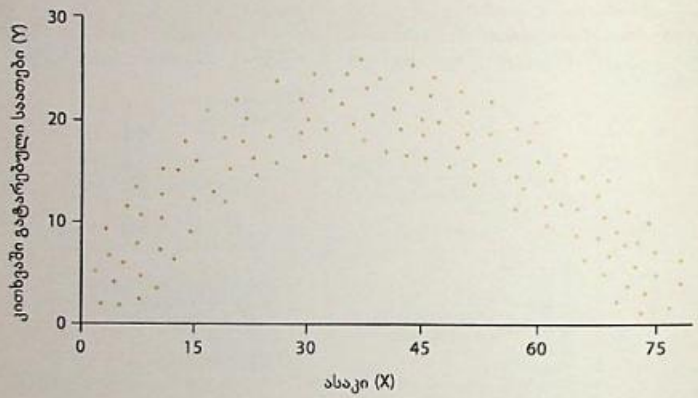
ასე, რომ ამ შემთხვევაში კოვარიაციის მნიშვნელობა დადებითი იქნება, a და b-ს დადებითი მნიშვნელობები დადებით r-ს გვაძლევს.
ახლა დავუშვათ, რომ გვაქვს მონაცემთა შემდეგი განლაგება:



ამ შემთხვევაში კოვარიაციის მნიშვნელობა იქნება უარყოფითი, რაც b-ს უარყოფით მნიშვნელობას გვაძლევს.

დეტალების განხილვა 13.2
რამდენიმე შენიშვნა არაწრფივ კავშირზე

საშუალო შემოსავლებსა და ბაკალავრის განათლების მქონეთა რიცხვს შორის კავშირის განხილვისას ჩვენ დავუშვით, რომ მათ შორის კავშირი წრფივი ხასიათის იყო. სოციალურ მეცნიერებებში ცვლადთა შორის კავშირების უმრავლესობა შესაძლებელია დაყვანილ იქნას წრფივი კავშირის სახემდე ან მოხდეს წრფივ კავშირთან მიახლოება. მაგრამ ხშირ შემთხვევაში ცვლადთა შორის კავშირი ნაცვლად წრფისა უფრო შესაბამისად სხვა ფორმის წირით გამოისახება. მაგალითად მოყვანილი წერტილთა გაბნევის დიაგრამა ასახავს არაწრფივ კავშირს ასაკსა და კითხვაში გატარებულ საათების რაოდენობას შორის(პირობითი მონაცემები). კითხვაში გატარებული საათები თავდაპირველად იზრდება ასაკის მატებასთან ერთად ოც წლამდე, რჩება სტაბილური სიდიდის ორმოც წლამდე და შემდეგ ასაკის მატებასთან ერთად კვლავ იწყებს კლებას.



ასაკსა და კითხვაში გატარებულ დროს შორის არაწრფივი კავშირი.

ყველაზე სწრაფი მეთოდი გაარკვიოთ არის თუ არა თქვენს ცვლადებს შორის წრფივი ან არაწრფივი კავშირი, ან ნახოთ საერთოდ არის თუ არა რაიმე სახის კავშირი არის წერტილთა გაბნევის დიაგრამის გამოყენება. თუ ცვლადებს შორის კავშირის ვიზუალური სახე არ შეტყველებს მათ შორის წრფივი კავშირის არსებობაზე, მაშინ აზრი არ აქვს რომელიმე წრფის გამოყენებას მოცემული მონაცემებისათვის. მონაცემებს შორის არაწრფივი კავშირების განხილვა ცდება ამ წიგნის საზღვრებს. მიუხედავად ამისა, მონაცემებს შორის კავშირის შესწავლისას მათ შორის თავდაპირველად უნდა შეამოწმოთ თუ რამდენად განსხვავდება ეს კავშირი წრფივი კავშირისაგან.

კრიმინალური ქცევის გაგებისათვის განვიხილოთ კავშირი საშუალო შემოსავლებსა და კრიმინალურ ქცევას შორის. აშშ-ს აღწერის ბიურო რეგულარულად აგროვებს ინფორმაციას აშშ-ს მოსახლეობის შესახებ. აშშ-ს ბოლო სტატისტიკური წელიწდეული ასახავს დანაშაულის დონის შესახებ მონაცემებს 50-ივე შტატისათვის, კოლუმბიის ოლქის ჩათვლით. საანალიზოდ ავიღოთ ქურდობის დონე ყოველ 100,000 მოსახლეზე. ცხრილ 13.5-ში ნაჩვენებია ქურდობის დონე

და საშუალო შემოსავლები აშშ-ს ათი ყველაზე დასახლებული შტატისათვის. ამ მონაცემებისათვის წერტილთა გაბნევის დიაგრამა ნაჩვენებია ნახაზ 13.2-ზე.

ცხრილი 13.5 საშუალო შემოსავლები და დანაშაულის დონე ყოველ 100,000 მოქალაქეზე, ათი შტატის მონაცემები

შტატი	ოჯახების საშუალო შემოსავალი X	დანაშაულის დონე X
კალიფორნია	56,645	1,917
ტეხასი	44,922	2,962
ნიუ-იორკი	51,384	1,570
ფლორიდა	45,495	2,658
ილინოისი	52,006	2,165
პენსილვანია	46,259	1,729
ოჰაიო	44,532	2,429
მიჩიგანი	47,182	1,918
ჯორჯია	46,832	2,751
ნიუ-ჯერსი	64,470	1,568

წყარო: U.S. Census Bureau, Statistical Abstract of the United States, 2008, Table 301.

მოდით კვლავ განვიხილოთ ნახაზი 13.2. წერტილთა გაბნევის დიაგრამის მიხედვით ჩანს, რომ ცვლადთა შორის არის წრფივი კავშირი, ასევე დიაგრამის მიხედვით ვგებულობთ, რომ კავშირის მიმართულება უარყოფითია ე.ი. საშუალო შემოსავლების ზრდას დანაშაულის დონის შემცირება მოსდევს.

ცვლადთა შორის კავშირის სისტემური ანალიზისათვის საჭიროა უმცირეს კვადრატთა მეთოდის გამოყენებით შევადგინოთ რეგრესიის განტოლება. რადგან ჩვენი მიზანი დანაშაულის დონის განსაზღვრაა, ამიტომ ეს უკანასკნელი დამოუკიდებელ ცვლადად მივიჩნიოთ (Y).

a და b კოეფიციენტების მიღებისათვის საჭირო გამოთვლები ნაჩვენებია ცხრილ 13.6-ში.

b კოეფიციენტის გამოთვლისათვის ჩავსვათ ფორმულაში კოვარიაციისა და ვარიაციის მაჩვენებელები:

$$b = \frac{S_{YX}}{S_X^2} = \frac{-2,038,953}{40,824,614} = -0.05$$

b-ს მნიშვნელობის მიღების შემდეგ ვიპოვოთ a, ღერძთან თანაკვეთა:

$$a = \overline{Y} - b(\overline{X}) = 2,166.7 - (-0.05)(49,972.7) = 4,665$$

რეგრესის განტოლება შესაბამისად იქნება:

$$\hat{Y} = 4,665 + (-0.05)\,X$$

ამ განტოლების გამოყენებით შესაძლებელია მივღოთ დანაშაულის დონე თითოეული შტატისათვის მოცემული საშუალო შემოსავლების შემთხვევაში.

ცხრილი 13.6 დანაშაულის დონე და საშუალო შემოსავლები, ათი შტატი

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
შტატი	საშუალო შეურნობის საშუალო შემოსავალი X	დანაშაულის დონე Y	$(X-\overline{X})$	$(X-\overline{X})^2$	$(Y-\overline{Y})$	$(Y-\overline{Y})^2$	$(X-\overline{X})(Y-\overline{Y})$
კალიფორნია	56,645	1,917	6,672.3	44,519,587	-249.7	62,350	-1,666,073
ტეხასი	44,922	2,962	-5,050.7	25,509,570	795.3	632,502	-4,016,822
ნიუ-იორკი	51,384	1,570	1,411.3	1,991,767	-596.7	356,051	-842,123
ფლორიდა	45,495	2,658	-4,477.7	20,049,797	491.3	241,376	-2,199,984
ილინოისი	52,006	2,165	2,033.3	4,134,309	-1.7	2.89	-3,457
პენსილვანია	46,29	1,729	-3,713.7	13,791,568	-437.7	-191,581	1,625,486
ოჰაიო	44,532	2,429	-5,440.7	29,601,216	262.3	68,801	-1,427,096
მიჩიგანი	47,182	1,918	-2,790.7	7,788,006	-248.7	-61,852	694,047
ჯორჯია	46,832	2,751	-3,140.7	9,863,996	584.3	341,406	-1,835,111
ნიუ-ჯერსი	64,470	1,568	14,497.3	210,171,707	-598.7	358,442	-8,679,534
	$\sum X$=499,727	$\sum Y$=285.6	0.00	367,421,526	0.002	2,314,364	-18,350,577

საშუალო $X = \overline{X} = \dfrac{\sum X}{N} = \dfrac{499,727}{10} = 49,972.7$

საშუალო $Y = \overline{Y} = \dfrac{\sum Y}{N} = \dfrac{21.667}{10} = 2,166.7$

ვარიაცია $(Y) = S_Y^2 = \dfrac{\sum (Y-\overline{Y})^2}{N-1} = \dfrac{2,314,364}{9} = 257,152$

სტანდარტული გადახრა $(Y) = S_Y = \sqrt{257,152} = 507.10$

ვარიაცია $(X) = S_X^2 = \dfrac{\sum (X-\overline{X})^2}{N-1} = \dfrac{367,421,523}{9} = 40,824,614$

სტანდარტული გადახრა $(X) = S_X = \sqrt{40,824,614} = 6,389.41$

$$\text{კოვარიაცია } (X,Y) = S_{YX} = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{N-1} = \frac{-18,350,577}{9} = -2,038,953$$

ა. პასუხები შეიძლება დამრგვალების გამო განსვავდებოდეს ცალკეულ შემთხვევებში, ამ სვეტების ზუსტი ჯამი ყოველთვის ნოლის ტოლი იქნება

განვიხილოთ a და b კოეფიციენტების ინტერპრეტირება. b კოეფიციენტი უდრის -0.05 რაც ნიშნავს, რომ ქურდობის დონე შემცირდება 0.05-ით ყოველ დამატებით 1$-ის შემოსავალზე. ასევე, საშუალო შემოსავლების 10,000$-ით ზრდა დანაშაულის დონეს 500 ერთეულით შეამცირებს (-0.05x10,000).

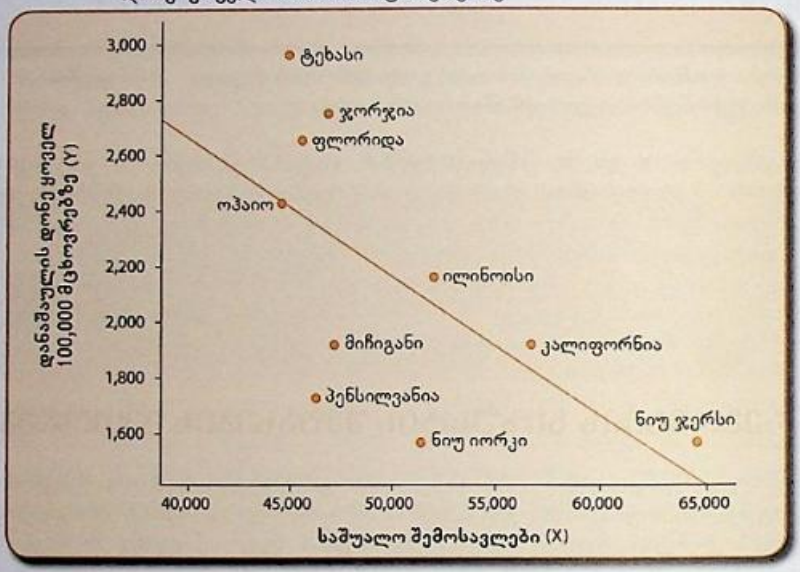
a კოეფიციენტი Y-ის მნიშვნელობაა, როცა $X = 0$ ე.ი. რეგრესიის წრფის Y ღერძთან თანაკვეთის წერტილი. $a = 4,665$ ნიშნავს, რომ შტატს სადაც საშუალო შემოსავლების ნულის ტოლია დანაშაულის დონე 4,665 იქნება.

რეგრესიის სიზუსტის შეფასების მეთოდები

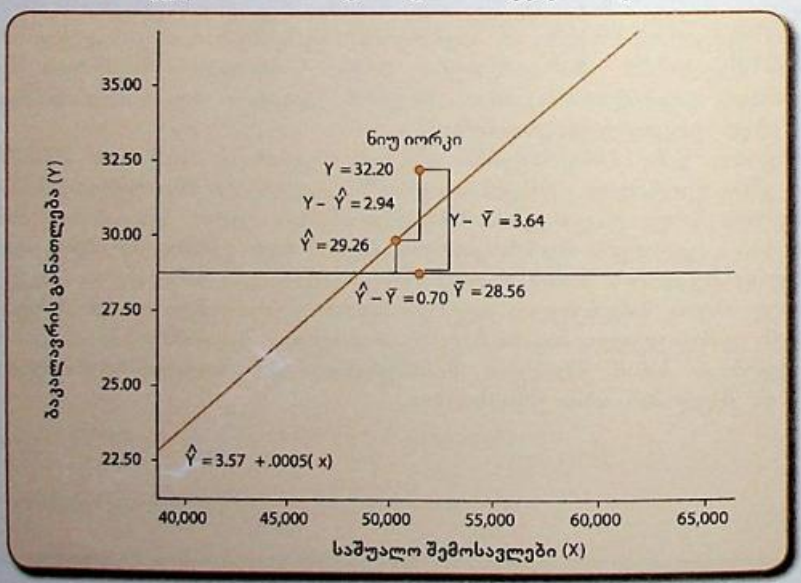
ჩვენ უკვე განვიხილეთ ორი რეგრესიის განტოლება, რომლის მეშვეობითაც შესაძლებელია მოვახდინოთ განათლების ან დანაშაულის დონის პროგნოზირება შტატების დონეზე. მაგრამ ორივე შემთხვევაში ჩვენი ანალიზი შორსაა სრულყოფილებისაგან. თუ განვიხილვათ ნახაზ 13.6-ს და 13.7-ს შევნიშნავთ, რომ პროგნოზი ზუსტი არცერთი შემთხვევისთვის არ არის. მიუხედევად იმისა, რომ ზოგიერთი შტატის მონაცემი საკმაოდ ახლოსაა რეგრესიის წრფესთან, არცერთი მათგანი უშუალოს წრფეზე არ მდებარეობს, რაც ნიშნავს, რომ გარკვეული მოცულობის შეცდომები იქნა დაშვებული. ბოლო წინადადების წაკითხვის შემდეგ, ბუნებრივია გაგიჩნდათ შეკითხვა: „როგორ შევაფასოთ თუ რამდენად ზუსტია ჩვენს მიერ გაკეთებული პროგნოზი?“.

როგორც წინა პარაგრაფებში ვახსენეთ რეგრესიის სისწორის შემოწმების ერთი გზაა წერტილთა გაბნევის დიაგარამზე დაკვირვება. რაც უფრო „ახლოსაა“ რეგრესიის წრფე დაკვირვების მონაცემებთან, მით უფრო შესაბამისად ასახავს რეგრესია ცვლადებს შორის კავშირს. მაგრამ ჩვენ გვჭირდება ისეთი მეთოდი, რომელიც შეცდომის ზომის რაოდენობრივი გამოსახვის საშუალებას მოგვცემდა. სწორედ ასეთი მაჩვენებელია დეტერმინაციის კოეფიციენტი r^2. ეს მაჩვენებელი ასახავს გამოთვლილი ბივარიანტული რეგრესიის შესაბამისობას დაკვირვების მონაცემებთან. სანამ უშუალოდ ამ მაჩვენებელს განვიხილავთ, მანამ თავად რეგრესიის შეცდომის არსი განვიხილოთ.

ნახაზი 13.7 რეგრესიის წრფე: საშუალო შემოსავლები და დანაშაულის დონე ყოველ 100,000 მცხოვრებზე



ნახაზი 13.8 შეცდომის ზომა ნიუ-იორკის მონაცემებისათვის



პროგნოზირების შეცდომა

განვიხილოთ ნახაზ 13.8. მასზე ნაჩვენებია რეგრესიის წრფე საშუალო შემოსავალსა (X) და ბაკალავრის ხარისხის მქონეთა რიცხვს (Y) შორის. თავდაპირველად ეს რეგრესიის წრფე და მისი შესაბამისი წერტილთა გაბნევის დიაგრამა ნაჩვენები იყო ნახაზ. 13.6-ზე.

ნახაზ 13.8-ზე განხილულია მხოლოდ ერთი შტატის, ნიუ-იორკის მონაცემები, Y ცვლადის მიღებული მნიშვნელობები. დავუშვათ, ჩვენთვის ცნობილი არ არის Y-ის რეალური მნიშვნელობები. რადგან საშუალო მაჩვენებელი შეცდომის კვადრატების მინიმალურ მნიშვნელობას იძლევა, Y-ის ყველაზე მეტად მისაღები მნიშვნელობა მისი საშუალოა ან 28.56%. ნახაზ 13.8-ზე ნაჩვენები ჰორიზონტალური წრფე ასახავს ამ საშაულო მაჩვენებელს. შევადაროთ, ნიუ-იორკის ცვლადის დაკვირვების მონაცემი 32.20%, რეგრესიის შედეგად მიღებულ მაჩვენებელს:

$$Y - \overline{Y} = 32.20 - 28.56 = 3.64$$

3.64 შეცდომის დონით, ნიუ-იორკისათვის მიღებული მონაცემი არც თუ ისე ზუსტია.

პროგნოზის გაუმჯობესების მიზნით გავითვალისწინოთ X (საშუალო შემოსავლების)-ის მნიშვნელობები. თუ რეგრესიის განტოლებაში ჩავსვამთ საშუალო შემოსავლების მნიშვნელობას $51,384, მივიღებთ:

$$\hat{Y} = 3.57 + 0.0005\ (X)$$

$$\hat{Y} = 3.57 + 0.0005\ (1,384) = 29.26$$

მივიღებთ $\hat{Y}$-ის გამოთვლილ მნიშვნელობას 29.26.

ამ მნიშვნელობისათვის ხელახლა გამოვთვალოთ შეცდომის მნიშვნელობა Y-სათვის, რეგრესიის განტოლებით მიღებულ $\hat{Y}$ და დაკვირვების მონაცემებს შორის:

$$Y - \hat{Y} = 32.20 - 29.26 = 2.94$$

მიუხედევად იმისა, რომ ეს გამოთვლა რეგრესიის მიხედვით არ არის სრულყოფილი წინა გამოთვლასთან შედარებით ის უფრო მეტად გაუჯობესებულია 0.70 ერთეულით (3.64-2.94=0.70).

ეს გაუმჯობესებული გამოთვლა ნაჩვენებია ნახაზ 13.8-ზე. შევნიშნოთ, რომ გამოთვლის ეს გაუმჯობესება უდრის $Y - \hat{Y} = 29.26 - 28.56 = 0.70$. ეს მნიშვნელობა ასახავს რეგრესიის შეცდომის შემცირებას წრფივი რეგრესიის გამოყენებისას.

მოდით შევაჯამოთ, რაც უკვე აქამდე განვიხილეთ. გვაქვს პროგნოზირების, რეგრესიის გამოთვლის ორი მეთოდი და ასევე შეცდომის ორი საზომი. პირველი პროგნოზირების მეთოდი გულისხმობს $\overline{Y}$-ის გამოთვლას X-ის მონაცემების გარეშე; შეცდომა გამოითვლება $Y - \overline{Y}$. მეორე მეთოდი გულისხმობს X-ის მნიშვნ-

ელობების გამოყენებას $\hat{Y}$-ის მნიშველობების მისაღებად. პროგნოზის შეცდომა გამოითვლება $Y - Y$.

შერჩევის ყველა მონაცემებისათვის შეცდომის გამოსათვლელად, სხვაობები ავიყვანოთ კვადრატში და გამოვთვალოთ მათი ჯამი. Y-ის თითოეული მნიშვნელობისათვის გვექნება:

$$\sum(Y - \overline{Y})^2$$

სხვაობების კვადრატების ჯამი კვდრატების ჯამად იწოდება ან შემოკლებული აღნიშვნით SST:

$$SST = \sum(Y - \overline{Y})^2$$

რეგრესიის წრფესთან სხვაობისათვის $\hat{Y}$ გვაქვს:

$$\sum(Y - \overline{Y})^2$$

რეგრესიის წრფიდან ვარიაციის კვადრატების ჯამი იწოდება ნაშთის კვადრატების ჯამად ან შემოკლებული აღნიშვნით SSE:

$$SSE = \sum(Y - \overline{Y})^2$$

(შეცდომისა და ნაშთის მნიშვნელობები უკვე განვიხილეთ ამ თავში).

პირველი წრფივი რეგრესიის შესაბამისობა შესაძლებელია შეფასდეს ნაშთის კვადრატების ჯამის შედარებით (SSE) სხვაობების კვადრატების ჯამთან (SST). ნაშთის კვადრატების ჯამის სხვაობების კვადრატების ჯამთან იწოდება რეგრესიის კვადრატულ ჯამად ან შემოკლებული აღნიშვნით – SSR და გამოითვლება:

$$SSR = SST - SSE$$

გამოვთვალოთ აღნიშნული მაჩვენებელი ჩვენი მონაცემებისათვის-საშუალო შემოსავლებსა (X) და განათლების დონეს შორის (Y). ცხრილ 13.4-დან უკვე გვაქვს რეგრესიის კვადრატების ჯამის მნიშვნელობა:

$$SST = \sum(Y - \overline{Y})^2 = 120.3$$

ნაშთის კვადრატების ჯამის გამოთვლისათვის, გამოვთვალოთ $\hat{Y}$-ის მნიშვნელობები თითოეული შტატისათვის, გამოვაკლოთ ეს მაჩვენებელი დაკვირვების მონაცემებს Y, ავიყვანოთ კვადრატში და შევკრიბოთ მიღებული მნიშვნელობები. გამოთვლები ნაჩვენებია ცხრილ 13.7-ში.

ნაშთის კვადრატების ჯამი შესაბამისად იქნება:

$$SSE = \sum(Y - \bar{Y})^2 = 25.57$$

ხოლო რეგრესიის კვადრატების ჯამი იქნება:

$$SSR = SST - SSE = 120.3 - 25.57 = 94.73$$

ცხრილი 13.7 შეცდომის კვადრატების ჯამის გამოთვლა

	(1)	(2)	(3)	(4)	(5)
შტატი	საუკეთესო მეურნეობის საშუალო შემოსავალი X	დანაშაულის დონე Y	$\hat{Y}$	$(Y-\hat{Y})$	$(Y-\hat{Y})^2$
კალიფორნია	56,645	1,917	31.89	-2.09	4.37
ტეხასი	44,922	2,962	26.03	-0.53	0.28
ნიუ-იორკი	51,384	1,570	29.26	2.94	8.64
ფლორიდა	45,495	2,658	26.32	0.88	0.77
ილინოისი	52,006	2,165	29.57	1.63	2.66
პენსილვანია	46,29	1,729	26.70	-0.10	0.01
ოჰაიო	44,532	2,429	25.84	-2.54	6.45
მიჩიგანი	47,182	1,918	27.16	-1.06	1.12
ჯორჯია	46,832	2,751	26.99	1.11	1.23
ნიუ-ჯერსი	64,470	1,568	35.81	-0.21	0.04
	$\sum X$=499,727	$\sum Y$=285.6			$\sum(Y-\hat{Y})^2$=25.57

დეტერმინაციის r^2 კოეფიციენტი, როგორც შეცდომის პროპორციული შემცირების საზომი

დეტერმინაციის r^2 კოეფიციენტი ასახავს რეგრესიის განტოლებით მიღებული შედეგების გაუმჯობესების ზომას. დეტერმინაციის r^2 კოეფიციენტი არის ცვლადთა შორის კავშირის შეცდომის პროპორციული შემცირების მაჩვენებელი. როგორც მე-12 თავში განვიხილეთ შეცდომის პროპორციული მაჩვენებელი PRE გამოითვლება შემდეგი ფორმულით:

$$PRE = \frac{E_1 - E_2}{E_1}$$

სადაც,

E_1 დამოკიდებული ცვლადის პროგნოზირების შეცდომაა, დამოუკიდებელი ცვ-

ლადის გამოტოვებით

E_2 დამოკიდებული ცვლადის პროგნოზირების შეცდომაა, დამოუკიდებელი ცვლადის მიხედვით

უკვე გვაქვს ყველა ელემენტი შეცდომის პროპორციული მაჩვენებლის გამოთვლისათვის. რადგან სხვაობების კვადრატების ჯამი (SST), ასახავს პროგნოზის შეცდომას დამოუკიდებელი ცვლადის გამოტოვებით, შეგვიძლია მივიჩნიოთ:

$$E_1 = SST$$

ასევე, რადგან ნაშთების კვადრატების ჯამი (SSE) ასახავს პროგნოზის ცდომილებას დამოუკიდებელი ცვლადის გამოყენებით, ამიტომ შეგვიძლია მივიჩნიოთ:

$$E_2 = \text{SSE}$$

უკვე შეგვიძლია განვსაზღვროთ დეტერმინაციის r^2 კოეფიციენტი. ეს მაჩვენებელი ასახავს შეცდომის პროპორციულ შემცირებას წრფივი რეგრესიის განტოლების გამოყენებისას Y-ის მნიშველობების გამოთვლისათვის:

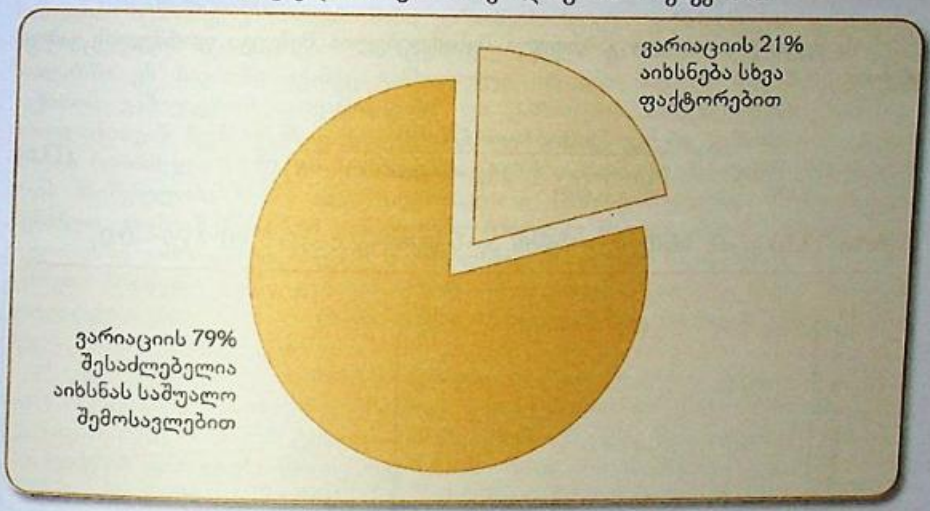
$$PRE = \frac{E_1 - E_2}{E_1} = \frac{\sum (Y - \bar{Y})^2 - \sum (Y - \hat{Y})^2}{\sum (Y - \bar{Y})^2}$$

ჩვენი მაგალითის მიხედვით:

$$r^2 = \frac{120.3 - 25.57}{120.3} = \frac{94.73}{120.3} = 0.79$$

დეტერმინაციის (r^2) კოეფიციენტი ასახავს დამოკიდებული ცვლადის (Y) მთლიან პროპორციულ ვარიაციას დამოუკიდებელი ცვლადის მიხედვით. r^2-ის 0.79-თან ტოლობა ნიშნავს: საშუალო შემოსავლების მიხედვით ბაკალავრის განათლების მქონეთა რიცხვის პროგნოზირებისას შეცდომის ზომა 79%-ით შემცირდა (0.79x100). სხვაგვარად რომ განვმარტოთ, დამოუკიდებელი ცვლადი (საშუალო შემოსავალი) დამოკიდებული ცვლადის ვარიაციის 79%-ს ხსნის (ნაჩვენებია ნახაზ 13.9-ზე).

ნახაზი 13.9 r^2-ის მნიშვნელობის გამოსახვა დიაგრამის მეშვეობით



დეტერმინაციის კოეფიციენტი (r^2) შეცდომის პროპორციული შემცირების მაჩვენებელი წრფივი რეგრესიისათვის. ასახავს დამოკიდებული ცვლადის მთლიან ვარიაციას (Y), რომელიც შესაძლებელია აიხსნას დამოუკიდებელი ცვლადით (X).

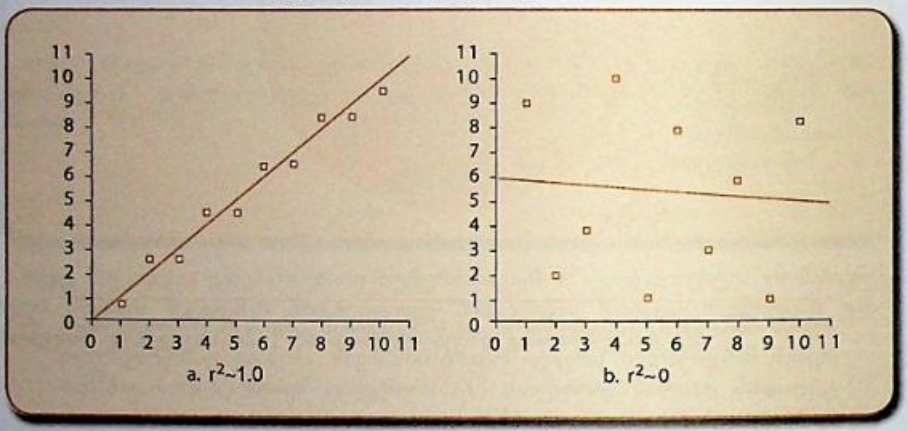
დეტერმინაციის კოეფიციენტი იღებს მნიშვნელობებს 0.0-დან 1.0-მდე. r^2-ის ერთთან ტოლობა ნიშნავს, რომ წრფივი რეგრესიის გამოყენებით, პროგნოზირების ცდომილება 100%-ით შემცირდა. ასევე, ეს ნიშნავს რომ დამოუკიდებელი ცვლადით შესაძლებელია აიხსნას დამოკიდებული ცვლადის ვარიაციის 100%. ამავე მნიშვნელობით დაკვირვების ყველა მონაცემი ემთხვევა რეგრესიის წრფეს და შეცდომის ზომა ნოლის ტოლია. r^2-ის ნოლთან ტოლობა ნიშნავს, რომ წრფივი რეგრესიის განტოლების გამოყენება არ აუმჯობესებს პროგნოზის შედეგებს. ნახაზ 13.10-ზე ნაჩვენებია r^2-ის 0-თან და 1-თან ტოლობის ილუსტრირება. a შემთხვევაში ნახაზზე ნაჩვენებია $r^2 = 1.0$ შემთხვევა, რომელიც კავშირის გამთთველის სრულყოფილი მაგალითია. ხოლო b ვარიანტში საპირისპირო შემთხვევაა ნაჩვენები $r^2 = 0.0$, რაც ნიშნავს რეგრესიის წრფის ცუდ შესაბამისობას ან b კოეფიციენტის ნოლთან ტოლობას.

r^2-ის გამოთვლა

r^2-ის შედარებით იოლი გამოთვლა შესაძლებელია შემდეგი ფორმულის გამოყენებით:

$$r^2 = \frac{[\text{კოვარიაცია } (X,Y)]^2}{[\text{ვარიაცია } (X)][\text{ვარიაცია } (Y)]} = \frac{S_{YX}^2}{S_X^2 S_Y^2} \qquad (13.6)$$

ნახაზი 13.10 r^2-ის მნიშვნელობების ილსუტრირება: ა) r^2~1.0 ბ) r^2~0.0



ფორმულის მიხედვით X და Y-ს შორის კოვარიაცია უნდა გავყოთ თითოეული ცვლადის ვარიაციის ნამრავლზე.

ჩვენს მაგალითში r^2-ის გამოთვლისათვის დავუბრუნდეთ კვლავ ცხრილ 13.4-ის მონაცემებს, სადაც უკვე გვაქვს კოვარიაციისა და ვარიაციის მაჩვენებლები გამოთვლილი ორივე ცვლადისათვის:

$$S_{YX} = 20,732.36$$

$$S_X^2 = 40,824,614$$

$$S_X^2 = 13.37$$

შესაბამისად,

$$r^2 = \frac{(20,732.36)^2}{(40,824,614)(13.37)} = \frac{429,830,751}{545,825,089} = 0.79$$

რადგან ჩვენ ჩვენი მაგალითის პირობებში განვიხილავთ რეალურ მონაცემებს საშუალო შემოსავლებისათვის მისი გაზომვის სკალა განსხვავდება დამოკიდებული ცვლადის-ბაკალავრების პროცენტული წილი გაზომვის სკალისაგან. ჩვენს ანალიზში ეს საკითხი აქამდე განხილული არ იყო, მაგრამ მას შემდეგ, რაც საჭიროა გამოვიყენოთ კოვარიაციისა და ვარიაციის მაჩვენებლები r^2-ის გამოთვლისათვის ჩვენ უნდა გავითვალისწინოთ განსხვავებები გაზომვის სკალებს შორის (ფორმულა 13.6). ამ პრობლემის გადაჭრა საკმაოდ მარტივია. მიღებული r^2-ის მნიშვნელობა 0.79 უნდა გავამრავლოთ 100-ზე, მივიღებთ 79-ს. თქვენს შესაძლო კითხვაზე: „რატომ ვამრავლებთ მიღებულ შედეგს ასზე?“ პასუხი შემდეგია: ასზე ვამრავლებთ, რადგან დამოკიდებული ცვლადი-ბაკალავრის ხარისხის მქონეთა პროცენტული წილი გაზომილია პროცენტულ სკალაზე 1-დან 100-მდე. დააკვირდით, ახლა სწორედ ის მნიშვნელობა მივიღეთ რაც თავდაპირველად ფორმულა 13.5-ის გამოყენებისას.

r^2 უნდა გამრავლდეს 100-ზე, რომ მივიღოთ დამოკიდებული ცვლადის ვარიაციის პროცენტული გამოსახულება, რაც შესაძლებელია აიხსნას დამოუკიდებელი ცვლადით. r^2-ის 0.79-თან ტოლობა ნიშნავს, რომ საშუალო შემოსავლების შესახებ მონაცემების გამოყენება წრფივ რეგრესიაში Y-ის პროგნოზირების ცდომილებას 79%-ით ამცირებს (0.79x100). ასევე ეს ნიშნავს, რომ დამოუკიდებელი ცვლადით (საშუალო შემოსავლებით) აიხსნება დამოკიდებული ცვლადის ვარიაციის 79%, როგორც ეს 13.9 ნახაზზეა ნაჩვენები.

პირსონის კორელაციის კოეფიციენტი (r)

სოციალურ მეცნიერებებში ფესვი r^2-დან r ცნობილია როგორც პირსონის კორელაციის კოეფიციენტი. ეს მაჩვენებელი საკმაოდ ხშირად გამოიყენება ინტერვალურ-ფარდობით მონაცემებს შორის კავშირის შეფასებისათვის:

$$r = \sqrt{r^2}$$

პირსონის კორელაციის კოეფიციენტი (r) კვადრატული ფესვი r^2-დან. ინტერვალურ-ფარდობით მონაცემებს შორის კავშირის საზომი, ასახავს ცვლადებს შორის წფრივი კავშირის სიძლიერეს. იღებს დადებით და უარყოფით მნიშვნელობებს.

პერსონის r ჩვეულებრივ მისი განსაზღვრების მიხედვით შედგენილი ფორმულით გამოითვლება:

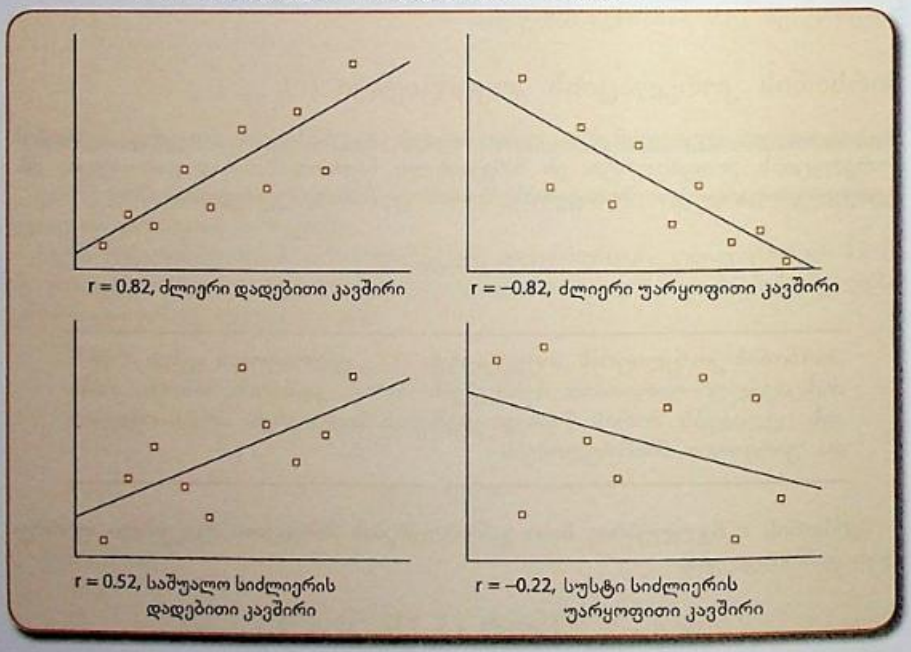
$$r = \frac{[\text{კოვარ. } (X,Y)]}{[\text{სტ.გად. } (X)][\text{სტ.გად. } (Y)]} = \frac{S_{XY}}{S_X S_Y}$$

ე.ი. r განისაზღვრება, როგორც X და Y-ს შორის კოვარიაციის შეფარდება მათივე სტანდარტული გადახრების ნამრავლთან.

პირსონის r-ის თვისებები

პირსონის r არის ინტერვალურ-ფარდობით მონაცემებს შორის კავშირის საზომი. გამას მსგავსად (განვიხილეთ მე-12 თავში) ეს მაჩვენებელი იღებს მნიშვნელობას 0.0-დან +/-1.0-მდე. r-ის +1.0-ის მნიშვნელობა აღნიშნავს ცვლადებს შორის სრულყოფილ დადებით კავშირს, ხოლო -1.0 კი სრულყოფილ უარყოფით კავშირს ასახავს. r-ის აბსოლიტური მნიშვნელობა ასახავს ცვლადებს შორის წრფივი კავშირის სიძლიერეს. (დეტალური განხილვისათვის მიმართეთ 12.1 ქვეპარაგრაფს). შესაბამისად, -0.75-ის ტოლი კორელაცია ასახავს უფრო ძლიერ წრფივ კავშირს ვიდრე 0.50-ის ტოლი კორელაცია. ნახაზ 13.11-ზე ნაჩვენებია ძლიერი დადებითი კავშირი, საშუალო სიძლიერის დადებითი და სუსტი უარყოფითი კავშირი.

ნახაზი 13.11 წერტილთა გაბნევის დიაგრამა ძლიერი დადებითი კავშირის, საშუალო სიძლიერის დადებითი და სუსტი უარყოფითი კავშირისათვის, r-ის აბსოლიტური მნიშვნელობების მიხედვით



b კოეფიციენტისაგან განსხვავებით r არის კავშირის სიმეტრიული საზომი. ე.ი. X და Y შორის კორელაცია Y და X-ს შორის კორელაციის იდენტურია. საპირისპიროდ, ცვლადების მიმდევრობის შეცვლისას b კოეფიციენტის მნიშვნელობაც იცვლება. მაგალითად, თუ შევცვლით დამოკიდებული და დამოუკიდებელი ცვლადების როლს.

ჩვენს მაგალითში r-ის გამოთვლისათვის გამოვიყენოთ 13.4 ცხრილის მონაცემები, რომელშიც უკვე გამთვლილია კოვარიაცა და სტანდარტული გადახრები X და Y-ისათვის:

$$r = \frac{S_{YX}}{S_X S_Y} = \frac{20732.36}{(6389.41)(3.66)} = \frac{20732.36}{23385.24} = 0.89$$

0.89 კორელაციის კოეფიციენტი ასახავს ძლიერი დადებითი კავშირის არსებობას საშუალო შემოსავლებსა და ბაკალავრის განათლების მქონეთა პროცენტულ წილს შორის.

შევნიშნოთ რომ r-ის გამოთვლისათვის მხოლოდ კვადრატული ფესვი გამოვთვალეთ r^2-დან, რადგან $r = \sqrt{r^2}$ ან $\sqrt{0.89} = 0.89$. და პირიქით თუ თავდაპირველად გამოვთვლით r-ს r^2-ის მისაღებად უბრალოდ კვადრატში აგვყავს r. (ამ შემთხვევაში ყურადღება მიაქციეთ დადებითობის ან უარყოფითობის ნიშანს).

სტატისტიკა პრაქტიკაში: ფეხმძიმობა მოზარდებში და სოციალური უთანასწორობა

ამერიკის შეერთებულ შტატებს ამჟამად ინდუსტრიულ ქვეყნებს შორის აქვს მოზარდების ფეხმძიმობის ყველაზე მაღალი მაჩვენებელი. 15-დან 19-წლამდე ასაკის ახალგაზრდების ფეხმძიმობის წილი 1990 წლისათვის იყო 95.9-% ყოველ 1000 ქალზე. მიუხედავად იმისა, რომ 2005 წლისათვის მოზარდთა ფეხმძიმობის წილმა მიაღწია ყველაზე დაბალ ნიშნულს ბოლო 30 წლის მანძილზე (69.5% ყოველ 1000 ქალზე)2, ეს მაჩვენებელი ჯერ კიდევ თითქმის ორჯერ მაღალია სხვა ინდუსტრიული ქვეყნების შესაბამის მაჩვენებელზე. ასეთი მაღალი მაჩვენებელი ამერიკის შეერთებულ შტატებში სხვა ფაქტორებთან ერთად აიხსნება სიღარიბის მაღალი დონითა და უთანასწორობით.

კავშირის არსებობა მოზარდების მაღალი ფეხმძიმობის დონესა, სიღარიბესა და სოციალურ უთანასწორობას შორის დადასტურებულ იქნა, როგორც ნაციონალურ, ასევე საერთაშორისო დონეზე. მოზარდთა ფეხძიმობის წილი მაღალია ღარიბ მოსახლეობაში, ხოლო იმ ინდუსტრიულმა ქვეყნებმა, რომლებმაც მიაღწიეს წარმატებას სოციალური უთანასწორობის შემცირებაში აღნიშნული

მაჩვენებლის ყველაზე დაბალი დონე აქვთ.[3] ცნობილი სოციოლოგი ვილიამ ვილსონი ამტკიცებდა, რომ ბოლო 25 წლის მანძილზე დაბალი კვალიფიკაციის სამუშაო ადგილების შემცირებამ გამოიწვია უმუშევრობის ზრდა, განსაკუთრებით მცირე ზომის შიდა ქალაქებში, რამაც გაზარდა ფეხმძიმობისა და სოციალური დამოკიდებულების დონე[4]. ახალგაზრდები, რომლებიც ცხოვრობენ მაღალი უმუშევრობის რეგიონებში სიღარიბე და შესაძლებლობების დაბალი დონე ექვსჯერ და შვიდჯერ ზრდის დაუქორწინებელ მშობლებად გახდომის ალბათობას[5].

ფეხმძიმობასა და ეკონომიკურ ფაქტორებს შორის კავშირის ანალიზისათვის განვიხილოთ შტატების მიხედვით უმუშევრობის მონაცემები 2004 წლისათვის და ახალგაზრდებში ფეხმძიმობის დონის შესახებ მონაცემები 2005 წლისათვის. უმუშევრობისა და ფეხმძიმობის ინტერვალურ-ფარდობითი მონაცემების გამოყენებით შეგვიძლია შევამოწმოთ ჰიპოთეზა, იმის თაობაზე რომ შტატები სადაც უმუშევრობის მაღალი დონეა ასევე ახალგაზრდების ფეხმძიმობის მაღალი მაჩვენებელი ფიქსირდება.

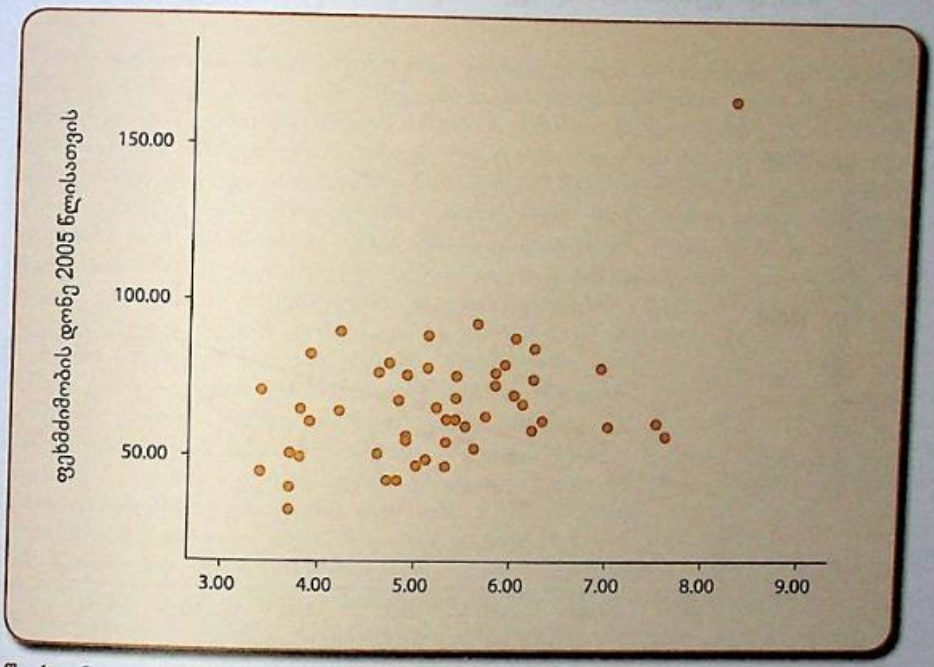
ნახაზ 13.12-ზე ნაჩვენებია წერტილთა გაბნევის დიაგრამა უმუშევრობისა და ფეხმძიმობის დონეებისათვის. რადგან ჩვენი დაშვებით, შეგვიძლია 2004 წლის უმუშევრობის შესახებ მონაცემების გამოყენებით განვსაზღვროთ ფეხმძიმობის დონე 2005 წლისათვის, უმუშევრობის დონე უნდა მივიჩნიოთ დამოუკიდებელ ცვლადად X; ხოლო ფეხმძიმობის დონე, როგორც დამოკიდებული ცვლადი . წერტილთა გაბნევის დიაგრამის მიხედვით შეგვიძლია დავასკვნათ, რომ ამ ორ ცვლადს შორის არსებობს წრფივი კავშირი, ასევე ამ დიაგრამის მიხედვით ვასკვნით, რომ მათ შორის დადებითი წრფივი კავშირია ე.ი. თუ შტატში უმუშევრობის მაჩვენებლები იზრდება ასევე გაიზრდება ახალგაზრდებში ფეხმძიმობის დონე.

ჩვენი ბივარიანტული რეგრესიის განტოლება 49 შტატისათვის[6] არის:

$$\hat{Y}=23.925+8.127\ (X)$$

ამ განტოლებაში წრფის დახრილობა (b) და თანაკვეთა (a) არის 23.925. დახრილობის დადებითი მნიშვნელობა 8.127 ამტკიცებს ჩვენს მიერ ადრე გაკეთებულ დაშვებას უმუშევრობასა და ფეხმძიმობის დონეს შორის დადებითი კავშირის შესახებ. სხვაგავარად რომ ვთქვათ, რაც უფრო მაღალია უმუშევრობის დონე, მით უფრო მაღალია ფეხმძიმობის დონე. b-ს 8.127-თან ტოლობა ნიშნავს, რომ ყოველი 1 პროცენტის ზრდა უმუშევრობის დონეში ფეხმძიმობის დონეს ზრდის 8.127-ით ყოველი 1000 ქალისათვის. თანაკვეთა a, მნიშვნელობით 23.925 ასახავს სრული დასაქმების (უმუშევრობის ნოლი დონით) შემთხვევაში ფეხმძიმობის დონეს ყოველ 1000 ქალზე. ამ კავშირის ამსახველი შესაბამისი რეგრესიის წრფე ნაჩვენებია ნახაზ 13.13-ზე.

ნახაზი 13.12 წერტილთა გაბნევის დიაგრამა, უმუშევრობისა და ფეხმძიმობის დონეები



წყარო: Guttmacher Institute, U.S. Teenage Pregnancies, Births and Abortions: National and State Trends and Trends by Race and Ethinicity, January 2012, and Statistical Abstract of the United States, 2007.

დახრილობა (b) დამოკიდებული ცვლადის მნიშვნელობის ცვლილება დამოუკიდებელი ცვლადის ერთი ერთეულის ცვლილების შედეგად.

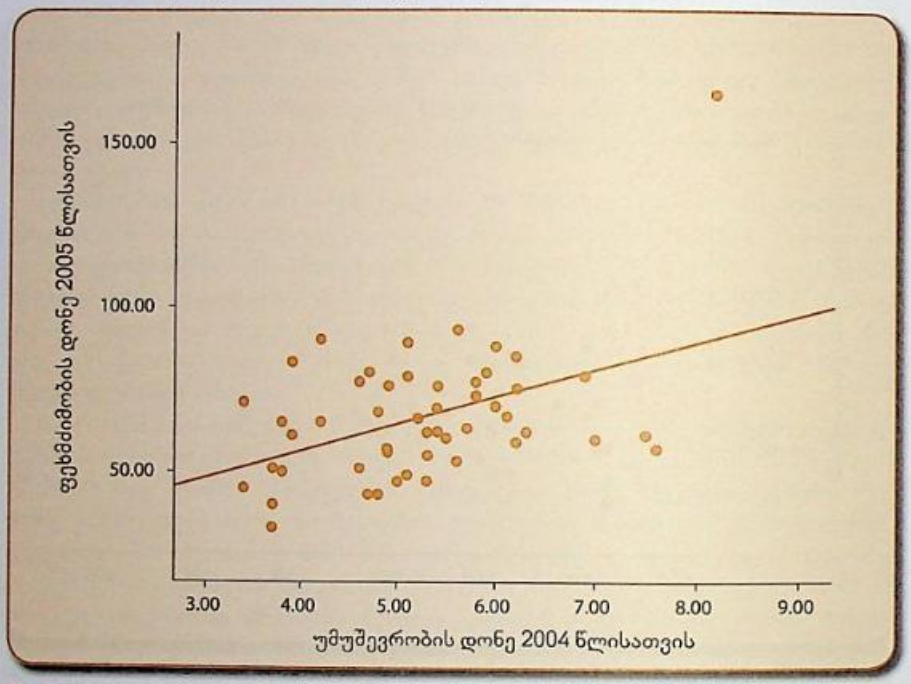
რეგრესიის განტოლების გამოყენებით შეგვიძლია განვსაზღვროთ ფეხმძიმობის დონე 2005 წლისათვის 2004 წლის უმუშევრობის მონაცემებზე დაყრდნობით. მაგალითად, 2004 წლის 5.8%-ის უმუშევრობის დონის მიხედვით ალაბამის შტატის ფეხმძიმობის დონე იქნება:

$$\hat{Y} = 23.925 + 8.127\,(5.8) = 71.06$$

უფრო მაღალი უმუშევრობის 7.5%-იანი დონისათვის, 2005 წლის საპროგნოზო მაჩვენებელი ფეხმძიმობის დონე იქნება:

$$\hat{Y} = 23.925 + 8.127\,(7.5) = 84.88$$

ნახაზი 13.13 წერტილთა გაბნევის დიაგრამა და რეგრესიის წრფე უმუშევრობისა და ფეხმძიმობის დონეებისათვის



წყარო: Guttmacher Institute, U.S. Teenage Pregnancies, Births and Abortions: National and State Trends and Trends by Race and Ethinicity, January 2012, and Statistical Abstract of the United States, 2007.

ასევე ამ მონაცემებისათვის გამოვთავალოთ r^2 და r მაჩვენებლები. გამოთვლების შედეგად მივიღებთ, $r = 044$ და $r^2 = 0.44^2 = 0.19$ r-ის 0.44 მნიშვნელობა ასახავს ცვლადებს შორის საშუალო სიძლიერის დადებით კავშირს.

დეტერმინაციის r^2 კოეფიციენტი ასახავს რეგრესიის შეცდომის პროპორციულ შემცირებას ფეხმძიმობის დონის გამოთვლისას. r^2-ის 0.19 მნიშვნელობა ასახავს რეგრესიის განტოლების გამოყენებით შედეგის 19%-იან გაუმჯობესებას. (0.19x100). ასევე, დამოუკიდებელი ცვლადი (უმუშევრობის დონე) დამოკიდებული ცვლადის ვარიაციის დაახლოებით 19%-ს ხსნის.

ეს ანალიზი ეხება მხოლოდ ერთ ფაქტორს ფეხმძიმობის დონეს მოზარდებში. სხვა სოციალურ-ეკონომიკური ფაქტორი, რომელსაც გავლენა აქვს ფეხმძიმობის დონეზე არის: სიღარიბის დონე, სახელმწიფოს კეთილდღეობის პოლიტიკა და განათლებაზე დანახარჯები. ყველა ეს ფაქტორი განხილულ უნდა იქნას საკითხის უფრო სრულყოფილად შესწავლისათვის.

სტატისტიკა პრაქტიკაში: ქორწინება და შემოსავლები

იმ ფაქტორებს შორის, რომლებიც განსაზღვრავენ პიროვნების შემოსავლებს არის ადამიანური კაპიტალისა (მაგ. ასაკი, განათლება, სამუშაო გამოცდილება და ჯანმრთელობის დონე) და შრომის ბაზრების მახასიათებლები (უმუშევრობის დონე და დასაქმების სტრუქტურა). პიროვნული ფაქტორები, როგორიცაა სქესი, რასა და ეთნიკურობა, ასევე განაპირობებს განსხვავებებს შემოსავლებში. ამ ფაქტორებთან ერთად განსხვავებებს შემოსავლებში განაპირობებს ქორწინებაში ყოფნა. მიუხედავად იმისა, რომ ქორწინებაში მყოფი მამაკაცებისათვის ეს შედარებით მაღალ შემოსავლებთან არის დაკავშირებული, ქალებისათვის ქორწინება გარკვეული სასჯელის როლს ატარებს. დაქორწინებული ქალები გაცილებით ნაკლებს გამოიმუშავებენ მიუხედავად განათლების დონისა, ვიდრე ქორწინებაში არმყოფი ქალები.

დაქორწინებული ქალების დაბალი შემოსავლები აიხსნება განსხვავებებით სამუშაო გამოცდილების ქონაში. დაქორწინება და შვილების ყოლა ზღუდავს ქალის არჩევანს ქონდეს მაღალ ანაზღაურებადი სამუშაო და კარიერული წინსვლის საშუალება, რადგან მოქნილი სამუშაო გრაფიკის მქონე სამუშაო ადგილები ჩვეულებრივ ნაკლებად ხელმისწვდომი და დაბალი ანაზღაურების მომტანია. ამასთან, დაქორწინებული ქალი უფრო მეტ ხანს რჩება შრომითი ბაზრის მიღმა, ვიდრე ქორწინებაში არმყოფი ქალი. როდესაც დაქორწინებული ქალი უბრუნდება შრომის ბაზარს, მას უწვეს კონკურენცია გაუწიოს უკვე სამუშაო გამოცდილებისა და პრაქტიკული ცოდნის მქონე თანამშრომლებს. (ქალებს, რომლებსაც განქორწინების ან დაქვრივების შემდეგ სურთ იყვნენ ფინანსურად დამოუკიდებელნი თითქმის იგივე ვალდებულებებს ინაწილებენ, რასაც აქტიურ ქორწინებაში მყოფი ქალები).

ამ ფაქტორებზე დაყრდნობით შეგვიძლია ვივარაუდოთ, რომ დაქორწინებული ქალებისათვის ფორმალური განათლებიდან შემოსავლები დაბალია დაუქორწინებელ ქალებთან შედარებით. შესაბამისად, დაუქორწინებელი ქალების წლიური შემოსავლები მაღალი იქნება დაქორწინებულ ქალებთან შედარებით. ეს საკითხი განვიხილოთ განათლების დონესა და ქალების შემოსავლებს შორის (ქორწინების სტატუსის მიხედვით) კავშირის შესწავლით. მონაცემები აღებულია ზოგადი სოციოლოგიური კვლევიდან 2008, 673 რესპონდენტის პასუხების მიხედვით.[7] ჩვენი დაშვებით, განათლების დონის (წლების რაოდენობა) მიხედვით შესაძლებელია განვსაზღვროთ პირადი შემოსავლები ე.ი. განათლება მივიჩნიოთ დამოუკიდებელ ცვლადად, X, ხოლო შემოსავლები დამოკიდებულ ცვლადად Y. რადგან ორივე ეს ცვლადი არის ფარდობით-ინტერვალური ცვლადთა შორის კავშირის შესწავლისათვის შეგვიძლია გამოვიყენოთ ბივარიანტული რეგრესიული ანალიზი.

რეგრესიის განტოლება დაუქორწინებელი ქალების (სრულგანაკვეთიანი დასაქმებით) შემოსავლებისათვის არის:

$$\hat{Y} = -47,254.46 + 6,541.19\ (X)$$

მოყვანილი განტოლების მიხედვით განათლების დონის ზრდას ერთეული წლების მიხედვით მოყვება $6,541.19-ის ზრდა წლიურ შემოსავლებში.

რეგრესიის განტოლება დაქორწინებელი ქალების (სრულგანაკვეთიანი დასაქმებით) შემოსავლებისათვის არის:

$$\hat{Y} = -9,536.24 + 4,014.63\ (X)$$

ამ განტოლების მიხედვით განათლების დონის ყოველ ერთეულობით ზრდას დაქორწინებული ქალებისათვის მოყვება საშუალო შემოსავლების $4,014.63-ით ზრდა. განათლების ყოველი დამატებითი წლისათვის, დაუქორწინებელი ქალის შემოსვალი იზრდება 2,526.56$-ით ($6,541.19-$4,014.63) მეტად ვიდრე ეს დაქორწინებული ქალის შემთხვევაშია.

გამოვიყენოთ რეგრესიის ეს განტოლებები რათა განვსაზღვროთ დაქორწინებული და დაუქორწინებელი ქალების საშუალო შემოსავლებს შორის განსხვავებები, განათლებისა და სამუშაო გამოცდილების 16 წლის შემთხვევაში:

$$\hat{Y} = -47,254.46 + 6,541.19\ (16) = 57,404.58$$

$$\hat{Y} = -9,536.24 + 4,014.63\ (16) = 54,697.84$$

წლიურ შემოსავლებს შორის განსხვავება სრული სამუშაო განაკვეთით მომუშავე დაქორწინებული და დაუქორწინებელი ქალებისათვის, კოლეჯის განათლებით, არის $2,706.74 ($57,404.58-$54,697.84).

ასევე გამოვთალოთ r და r^2, დაუქორწინებული ქალის შემთხვევაში $r = 0.48$, ხოლო დაქორწინებულისათვის $r = 0.23$. ამ კოეფიციენტების მიხედვით ორივე ჯგუფის შემთხვევაში არსებობს სუსტი-საშუალო (კავშირი უფრო ძლიერია დაუქორწინებელი ქალის შემთხვევაში) კავშირი განათლებასა და შემოსავლებს შორის.

იმის განსასაზღვრად თუ შემოსავლების ვარიაციის რა ნაწილი აიხსნება განათლების დონით, საჭიროა გამოვთვალოთ r^2:

$$r^2 \text{ (დაუქორწინებელი)} = 0.482 = 0.23$$

$$r^2 \text{ (დაქორწინებული)} = 0.232 = 0.05$$

რეგრესიის გამოყენებით ჩვენი განათლების მიხედვით შემოსავლების გამოთვლა დაუქორწინებელი ქალისათვის გაუმჯობესდა 23%-ით (0.23x100), ვიდრე ეს განათლების მონაცემების გამოყენების გარეშე იქნებოდა. იგივე მაჩვენებელი უფრო ნაკლებია დაქორწინებული ქალის შემთხვევაში – 5% (0.05x100).

ჩვენი ანალიზი ითვალისწინებს შემოსავლების დონეზე მოქმედ მხოლოდ ერთ ფაქტორს-განათლების დონეს. სრული ანალიზისათვის საჭიროა სხვა ფაქტორების გათვალისწინებაც: პროფესია, ასაკი, რასა/ეთნიკურობა და გენდერი.

მრავლობითი რეგრესია

ჩვენს მიერ აქამდე განხილულ მაგალითებში მხოლოდ ორი ცვლადი გვხვდებოდა: დამოკიდებული და დამოუკიდებელი. მაგალითად, წინა მაგალითში ფეხმძიმობის განსხვავებული დონეების ასახსნელად უმუშევრობის დონე გამოვიყენეთ. წრფივი ბივარიანტული რეგრესიის პირობებში უმუშევრობა დამოუკიდებელ, ხოლო ფეხმძიმობის დონე დამოკიდებულ ცვლადად მივიჩნიეთ. მიუხედავად იმისა, რომ ამ შემთხვევაში არსებული მონაცემების მიხედვით შედგენილმა რეგრესიის განტოლებამ მოგვცა მეტ-ნაკლებად ზუსტი შედეგი, გარდა დასაქმების დონისა საშედეგო ცვლადზე გავლენა ასევე ექნება სხვა ფაქტორებსაც.

მაგალითად, როგორც თეორიული მოსაზრებებიდან, ასევე ემპირიული ფაქტებიდან გამომდინარე ვიცით, რომ განათლების მიღების განსხვავებულ შესაძლებლობებს ასევე აქვს გავლენა ფეხმძიმობის დონეზე ახალგაზრდებში. ჩვენი ანალიზი უფრო სრულყოფილი გახდება თუ უმუშევრობის გარდა დამოუკიდებელ ფაქტორთა რიგში ასევე ჩავრთავთ განათლების მიღების შესაძლებლობებს.

მრავლობითი რეგრესია ბივარიანტული რეგრესიის გაფართოებაა. იგი საშუალებას იძლევა განხილულ იქნას ორ ან მეტი დამოუკიდებელი ფაქტორის გავლენა საშედეგო ცვლადის მნიშვნელობებზე. მრავლობითი რეგრესიის გამოთვლები ხელით საკმაოდ მონოტონურია, ხოლო SPSS-ის ან სხვა სტატისტიკური პროგრამის გამოყენებით მისი მნიშვენლობების მიღება საკმაოდ მარტივია.

მრავლობითი რეგრესია ბივარიანტული რეგრესიის გაფართოება, რომელიც იხილავს ორ ან მეტი დამოუკიდებელი ცვლადის გავლენას დამოკიდებულ ცვლადზე. მრაელობითი რეგრესიის განტოლების ზოგადი სახე ორი დამოუკიდებელი ცვლადით შემდეგია: $Y = a + b_1(X_1) + b_2(X_2)$

მრავლობითი რეგრესიის განტოლების ზოგადი სახე ორი დამოუკიდებელი ცვლადით:

$$\hat{Y} = a + b_1(X_1) + b_2(X_2)$$

სადაც,

$\hat{Y}$ = დამოკიდებული ცვლადის მნიშვნელობა

X_1 = პირველი დამოუკიდებელი ცვლადის მნიშვნელობა

X_2 = მეორე დამოუკიდებელი ცვლადის მნიშვნელობა

a = Y ღერძთან თანაკვეთა, Y-ის მნიშვენლობა, როცა ორივე დამოუკიდებელი ცვლადის მნიშვნელობა ნოლს უტოლდება

b_1 = Y-ის ცვლილება X_1-ის ერთეულობითი ცვლილების შედეგად.

b_2 = Y-ის ცვლილება X_2-ის ერთეულობითი ცვლილების შედეგად

მრავლობითი რეგრესიის ილუსტრირებისათვის, ჩვენს მაგალითში ფეხმძიმობის დონის შესახებ დავამატოთ ახალი დამოუკიდებელი ცვლადი სახელმწიფოს დანახარჯები თითოეულ მოსწავლეზე დაწყებით და საშუალო სკოლებში 2004 წლისათვის. გავაკეთოთ დაშვება, რომ რაც უფრო მაღალია სახელმწიფოს დანახარჯები თითოეულ მოსწავლეზე მით უფრო დაბალია ფეხმძიმობის დონე მოზარდებში, ასევე როგორც უკვე დავუშვით რაც უფრო მაღალია უმუშევრობის დონე, მით უფრო მაღალია ფეხმძიმობის დონე ახალგაზრდებში. რეგრესიის განტოლება განათლებაზე დანახარჯებისა და უმუშევრობის დონისათვის არის შემდეგი:

$$\hat{Y} = 70.783 + 3.364(X_1) - 0.003(X_2)$$

სადაც,

$\hat{Y}$ = ფეხმძიმობის დონე, 2005

X_1 = უმუშევრობის დონე, 2004

X_1 = შტატის დანახარჯი თითოეულ სკოლის მოსწავლეზე, 2004

ამ განტოლების მიხედვით, შტატში ახალგაზრდობაში ფეხმძიმობის დონე იზრდება 3.364-ით ყოველ 1%-იან ზრდაზე უმუშევრობის დონეში (X_1), მოსწავლის მიხედვით დანახარჯის მუდმივად შენარჩუნებით (X_2). ასევე, შტატში ფეხმძიმობის დონე იკლებს 0.003-ით მოსწავლეზე დანახარჯის ყოველ ერთ დოლარიანი ზრდის შედეგად, უმუშევრობის დონის უცვლელობის პირობებში (X_1).

ერთი ცვლადის მუდმივად შენარჩუნებით, შესაძლებელია გამოვაცალკევოთ მეორე დამოუკიდებელი ცვლადის ეფექტი. მაგალითად, თუ გვაქვს ორი შტატი უმუშევრობის ერთნაირი დონით, 1$-ის დამატებითი ხარჯი სკოლის მოსწავლეების განათლებაზე ამცირებს ფეხმძიმობის დონეს 0.003-ით. ასევე, თუ გვაქვს ორი შტატი ერთნაირი დანახარჯების მოცულობით სკოლის განათლებაზე, უმუშევრობის დონის 1%-იანი ზრდა მოიტანს 3.364-ის ზრდას ფეხმძიმობის დონეში ყოველ 1000 ქალზე. a კოეფიციენტის მნიშვნელობა 70.783 ასახავს ფეხმძიმობის დონეს, როცა უმუშევრობა და სკოლის მოსწავლეებზე დანახარჯები ნოლს უტოლდება. მიუხედავად იმისა, რომ რეალობაში ნაკლებადაა მოსალოდნელი, რომ რომელიმე შტატში უმუშევრობა ან სკოლაზე დანახარჯები ნოლს უტოლდებოდეს, a-ს მნიშვნელობას მაინც არსებითია ფეხმძიმობის დონის სწორად განსაზღვრისათვის.

რეგრესიის განტოლების გამოყენებით განვსაზღვროთ ფეხმძიმობის დონე მოზარდებში, უმუშევრობისა და სკოლის მოსწავლეებზე დანახარჯების მოცემული მნიშვნელობებისათვის. მაგალითად, 5.10%-ის უმუშევრობის დონითა და სკოლის მოსწავლეებზე $5,595-ის დანახარჯების პირობებში ფეხმძიმობის მოსალოდნელი დონე იქნება:

$$\hat{Y} = 70.783 + 3.364(5.10) - 0.003(5,595) = 71.15$$

ბივარიანტული რეგრესიის მსგავსად მრავლობითი რეგრესიის შემთხვევაშიც გვაქვს დეტერმინაციის კოეფიციენტი, რომელიც აღნიშნება r^2-ით (ბივარიანტულ რეგრესიაში – r^2). r^2 ასახავს შეცდომის პროპორციულ შემცირებას წრფივ რეგრესიაში, დამოკიდებულ ცვლადის მთლიან ვარიაციას, რომელიც აიხსნება ორი ან მეტი დამოუკიდებელი ცვლადით ერთობლივად. ჩვენს მაგალითში r_2 უდრის 0.20. ეს ნიშნავს, რომ სკოლის განათლებაზე დანახარჯებისა და უმუშევრობის დონის შესახებ მონაცემების გამოყენებით ფეხმძიმობის დონის შესახებ პროგნოზში გვაქვს შეცდომის 20%-იანი შემცირება (0.20*100). ასევე, დამოუკიდებელი ცვლადებით-უმუშევრობის დონითა და სკოლის მოსწავლეზე დანახარჯით შესაძლებელია აიხსნას დამოკიდებული ცვლადის ვარიაციის 20%.

დეტერმინაციის მრავლობითი კოეფიციენტი (r^2) მაჩვენებელი, რომელიც ასახავს დამოკიდებული ცვლადის მთლიან ვარიაციას, რაც შესაძლებელია აიხსნას ერთობლივად ორი ან მეტი დამოუკიდებელი ცვლადით.

სკოლის მოსწავლეებზე დანახარჯების შესახებ მონაცემების ჩართვამ დიდად არ გააუმჯობესა ჩვენი პროგნოზი ფეხმძიმობის დონის შესახებ. როგორც წინა პარაგრაფებში ვნახეთ უმუშევრობის დონე დამოკიდებული ცვლადის მთლიანი ვარიაციის 19%-ს ხსნიდა. ეს მაჩვენებელი მხოლოდ 1%-ით გაუმჯობესდა განათლებაზე დანახარჯების მონაცემების რეგრესიაში ჩართვით (20%-19%=1%).

როგორც ბივარიანტული რეგრესიის შემთხვევაში, ასევე მრავლობითი რეგრესიის შემთხვევაშიც კვადრატული ფესვი r^2-დან იწოდება პირსონის მრავლობითი კორელაციის კოეფიციენტად r. ჩვენს მაგალითში, $r = 0.45$. ეს უკანასკნელი აღნიშნავს, რომ ცვლადებს ფეხმძიმობის დონესა და ერთობლივად დამოუკიდებელ ცვლადებს-უმუშევრობის დონესა და სკოლის მოსწავლეებზე დანახარჯებს შორის არსებობს საშუალო სიძლიერის კავშირი.

პირსონის მრავლობითი კორელაციის კოეფიციენტი (r) წრფივი კავშირის მაჩვენებელი დამოუკიდებელ ცვლადსა და ორ ან მეტ დამოუკიდებელ ცვლადს შორის.

✓ *ცოდნის შემოწმება*

გამოყენეთ რეგრესიის განტოლება ფეხმძიმობის დონეს, უმუშევრობასა და სკოლის მოსწავლეების შორის დამოკიდებულების შესახებ და განსაზღვრეთ 2005 წლის ფეხმძიმობის დონე მოზარდებში იმ შტატისათვის, სადაც უმუშევრობის დონე 3%, ხოლო ერთ სკოლის მოსწავლზე დანახარჯი 6,000$-ია.

დეტალების განხილვა 13.3
შეცდომის ზჟარი: მოჩვენებითი კორელაცია და შემთხვევითი ეფექტები

ამ თავში განვიხილეთ კორელაციის r კოეფიციენტი, რომელიც ხშირად გამოიყენება ცვლადებს შორის კავშირის ანალიზისათვის. როგორც წინა მაგალითის განხილვისას ვნახეთ უმუშევრობასა და ფეხმძიმობის დონეს შორის კორელაცია 0.44-ია. კორელაციის ეს მნიშვნელობა განვმარტეთ, როგორც საშუალო სიძლიერის კავშირის მანიშნებელი აღნიშნულ ცვლადებს შორის.

მნიშვნელოვანია აღვნიშნოთ, რომ კორელაცის კოეფიციენტი მხოლოდ ასახავს ცვლადთა შორის კავშირს და არა მათ შორის მიზეზ-შედეგობრივ დამოკიდებულებას. გავრცელებული გამოთქმა „კოლერაცია არ არის მიზეზ-შედეგობრიობა“ სწორედ ნახსენებ თეზისს ასახავს. ასეთი კავშირი შესაძლოა იწოდებოდეს, როგორც „შემთხვევითი“, რადგან ორივე ცვლადზე გავლენა აქვს სხვა საკონტროლო, მესამე ცვლადს, ხოლო მათ შორის კავშირი კი მიზეზ-შედეგობრივი არ არის.

მე-10 თავში უკვე განვიხილეთ მაგალითი, სადაც ნაჩვენები იყო ცვლადებს შორის არასრულყოფილი, მოჩვენებითი კავშირი: მსროლელთა რიცხვი (დამოუკიდებელი ცვლადი) სროლის არეში და შენობების დაზიანება (დამოკიდებული ცვლადი). რაც უფრო მეტია მსროლელთა რიცხვი სროლის არეში, მით უფრო მეტია დაზიანებების რიცხვი. ამ კავშირმა შესაძლოა მცდარ დასკვნამდე მიგვიყვანოს-მსროლელთა რაოდენობა განსაზღვრავს დაზიანებების რიცხვს. მაგრამ ასეთი მიზეზობრივი დამოკიდებულება შესაძლებელია განპირობებულ იყოს, სხვა მესამე მიზეზობრივი ცვლადით- სროლის მასშტაბურობით. როდესაც მხარეში სროლა ძლიერდება მეტი ჯარისკაცი იგზავნება აღნიშნულ მხარეში,რაც თავის მხრივ იწვევს შენობათა დაზიანების მეტ რიცხვს. შესაბამისად, როცა სროლის დონე დაბალია ჯარისკაცთა შედარებით მცირე ოდენობა იგზავნება და დაზიანებათა რიცხვიც ნაკლებია.

ცვლადები მსროლელთა რიცხვი და დაზიანებათა ოდენობა, ორივე უკავშირდება მიზეზობრივ ცვლადს სროლის ინტენსივობას, მაგრამ თავად ამ ორ ცვალდს შორის პირდაპირი კავშირი არ არის. ცვლადი-სროლის ინტენსივობა იწოდება საკონტროლო ცვლადად, ხოლო მსროლელთა რიცხვსა და დაზიანებას შორის კავშირი მოჩვენებითია. შესაბამისად, დამოკიდებულ და დამოუკიდებელ ცვლადს შორის კავშირი შესაძლებელია ახსნილ იქნას მესამე, საკონტროლო ცვლადის შემოტანით.

კვლევით ლიტერატურაში ცვლადებს შორის მოჩვენებითი კავშირის უამრავი მაგალითი მოიპოვება. მაგალითად, 2004 წელს გამოქვეყნებულ სტატიაში მაიკლ ბენსონმა და მისმა კოლეგებმა[8] განიხილეს კავშირი ოჯახურ ძალადობასა და რასობრივ კუთვნილებას შორის. კვლევის მიხედვით ოჯახური ძალადობის დონე მაღალია იმ თემებში, სადაც აფრო-ამერიკული მოსახლეობა ჭარბობს. ნიშნავს თუ არა ამ კავშირის არსებობა ამ ცვლადებს შორის მიზეზ-შედეგობრივი კავშირის არსებობას? იმ აზრის განვითარება, რომ აფრო-ამერიკული მოსახლეობა უფრო მეტადაა მიდრეკილი ძალადობისაკენ, სულ მცირე რასობრივ-დისკრიმინაციული ხედვაა. ბენსონისა და მისი კოლეგების მიხედვით ძალადობასა და რასობრიობას შორის კავშირი შესაძლებელია აიხსნას ეკონომიკური კეთილდღეობის უთანაბრობით. ეკონომიკურად ნაკლებად განვითარებული რეგიონები უფრო მეტად აფრიკული წარმოშობის მოსახლეობით არის დასახლებული. ასეთ რეგიონებში კი ჩვეულებრივ ოჯახური ძალადობის დონე მაღალია. შეგვიძლია ვთქვათ, რომ რასასსა და ძალადობის დონეს შორის კავშირი აიხსნება მესამე ცვლადით-ეკონომიკური კეთილდღეობის უთანაბრობით.

როგორ დავდგინოთ, რომ ცვლადებს შორის კავშირის კორელაცია მოჩვენებითია? როგორც მე-10 თავში ვნახეთ, ასეთ შემთხვევაში უნდა ვაკონტროლოთ მესამე, თეორიულად მნიშვნელოვანი ცვლადი. ცვლადებს მსროლელთა რიცხვსა და დაზიანების ოდენობას შორის კავშირის ხელახალი განხილვისას უნდა ვაკონტროლოთ მესამე ცვლადი-სროლის ინტენსივობა. თუ ამ შემთხვევაში თავდაპირველი კავშირი შესუსტდება, ეს ნიშნავს რომ მათ შორის კავშირი მოჩვენებითი იყო.

ასევე, ეკონომიკური კეთილდღეობის მისაწვდომობასა, რასასა და ოჯახურ ძალადობას შო-

რის კავშირის განხილვისათვის ბენსონმა და ფოქსმა[9] ოჯახური ძალადობის დონეები შეისწავლეს აფრიკელ ამერიკელებსა და თეთრკანიანებს შორის, ასევე ეკონომიკური კეთილდღეობის დონეების მიხედვით. მათი კვლევის მიხედვით ოჯახური ძალადობის დონეებს შორის განსხვავებები მნიშვნელოვანი აღარაა, როცა ეკონომიკური განვითარების დონე მუდმივია. ოჯახურ ძალადობის დონეებს შორის განსხვავებები დაბალი ეკონომიკური კეთილდღეობის პირობებში თითქმის 50%-მდე მცირდება აფრიკელი ამერიკელებსა და თეთრკანიან მოსახლეობას შორის. ხოლო მაღალი შემოსავლების პირობებში ოჯახური ძალადობის დონე თითქმის ერთმანეთის ტოლია აფრიკელი ამერიკელებისა და თეთრკანიანი მოსახლეობისათვის. ამ შედეგების მიხედვით ოჯახური ძალადობის დონე რასის მიხედვით აიხსნება ეკონომიკური კეთილდღეობის ხელმისაწვდომობის განსხვავებებით.

ცვლადებს შორის მოჩვენებითი კორელაციისა და შემთხვევითი ეფექტების ანალიზისათვის შესაძლებელია ასევე მრავლობითი რეგრესიის გამოყენება. მრავლობითი რეგრესია საშუალებას გვაძლევს შევისწავლოთ დამოკიდებულ და დამოუკიდებელ ცვლადს შორის კავშირი, ერთი ან რამდენიმე საკონტროლო ცვლადის მუდმივად შენარჩუნებით. მაგალითად, წინა მაგალითში ჩვენ განვიხილეთ უმუშევრობის დონის კავშირი ფეხმძიმობის დონესთან, ერთ შემთხვევაში საკონტროლო ცვლადის- სკოლის მოსწავლეზე დანახარჯის მუდმივად შენარჩუნებით, ხოლო მეორე შემთხვევაში უმუშევრობის დონის მუდმივად შენარჩუნებით.

ძირითადი განმარტებები

• წერტილთა გაბნევის დიაგრამა: ინტერვალურ-ფარდობით მონაცემებს შორის არსებული კავშირის ვიზუალიზაციის მეთოდი. წერტილთა გაბნევის დიაგრამის შედგენა პირველი ეტაპია რეგრესიულ ანალიზში, იგი გვიჩვენებს ცვლადებს შორის პოტენციურად არსებულ კავშირს.

• წრფივი რეგრესიის განტოლებებს აქვს შემდეგი ზოგადი სახე:

$$\hat{Y} = a + b(X)$$

სადაც,

$\hat{Y}$ =დამოკიდებული ცვლადის გამოთვლილი მნიშვნელობაა

X = დამოუკიდებელი ცვლადის მნიშვნელობაა

$a = Y$ ღერძთან თანაკვეთა; ე. ი. a არის Y-ის მნიშვნელობა, როცა $X = 0$.

b = რეგრესიის წრფის დახრილობა, ან სხვაგვარად Y-ის ცვლილება, როცა X ერთი ერთეულით იცვლება.

• საუკეთესო შესაბამისობის რეგრესიის წრფე რეგრესიის წრფე, რომლისათვისაც ნაშთების კვადრატების ჯამი $\sum e^2$ მინიმალურია. ასეთი წრფე იწოდება უმცირეს კვადრატთა წრფედ, ხოლო თავად წრფის მიღების მეთოდი უმცირეს კვადრატთა მეთოდია.

• დეტერმინაციის r^2 კოეფიციენტი და პირსონის კორელაციის კოეფიციენტი r ასახავს რეგრესიის მოდელის მონაცემებთან შესაბამისობას. პირსონის r ასევე ზომავს ცვლადებს შორის კავშირის სიძლიერეს. დეტერმინაციის r^2 კოეფიციენტი რეგრესიის მოდელის შეცდომის პროპორციული შემცირების საზომია.

• მრავლობითი რეგრესიის განტოლების ზოგადი ფორმა, ორი დამოუკიდებელი ცვლადით შემდეგია:

$$\hat{Y} = a + b_1 (X_1) + b_2 (X_2)$$

სადაც,

$\hat{Y}$ = დამოკიდებული ცვლადის მნიშვნელობა

X_1 პირველი დამოუკიდებელი ცვლადის მნიშვნელობა

X_2 მეორე დამოუკიდებელი ცვლადის მნიშვნელობა

a = Y ღერძთან თანაკვეთა, Y-ის მნიშვნელობა, როცა ორივე დამოუკიდებელი ცვლადის მნიშვნელობა ნოლს უტოლდება

b_1 = Y-ის ცვლილება X_1-ის ერთეულობითი ცვლილების შედეგად.

b_2 = Y-ის ცვლილება X_2-ის ერთეულობითი ცვლილების შედეგად

- მრავლობითი დეტერმინაციის კოეფიციენტი (r^2) და პირსონის მრავლობითი კორელაციის კოეფიციენტი (r) ასახავს მიღებული რეგრესიის მოდელის შესაბამისობას დაკვირვების მონაცემებთან. პირსონის მრავლობითი კორელაციის კოეფიციენტი (r) ასევე დამოკიდებულ და დამოუკიდებელ ცვლადებს შორის კავშირის სიძლიერეს ასახავს. დეტერმინაციის კოეფიციენტი (r^2) ასევე შესაძლებელია განხილულ იქნას, როგორც რეგრესიის პროპორციული შემცირების საზომი.

საკვანძო ტერმინები

დეტერმინაციის კოეფიციენტი (r^2)
დეტერმისტული (სრულყოფილი) წრფივი კავშირი
უმცირეს კვადრატთა წრფე (საუკეთესო შესაბამისობის წრფე)
უმცირეს კვადრატთა მეთოდი
წრფივი კავშირი
მრავლობითი დეტერმინაციის კოეფიციენტი (r^2)
პირსონის კორელაციის კოეფიციენტი (r)
პირსონის მრავლობითი კორელაციის კოეფიციენტი (r)
წერტილთა გაბნევის დიაგრამა
დახრილობა (b)
Y– ღერძთან თანაკვეთა (a)

დამატებითი რესურსები



ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

[GSS08PFP-B]

მაგალითი 1: წერტილთა გაბნევის დიაგრამის მიღება

აქვს თუ არა უფრო მაღალი განათლების მქონე პიროვნებებს შედარებით მეტი სამუშაო კვირები? ამ კითხვაზე პასუხის გასაცემად გამოვიყენოთ ის მეთოდები, რაც ამ თავში განვიხილეთ SPSS-ზე დაყრდნობით. მონაცემთა ფაილში GSS08PFP-B გამოვიყენოთ ცვლადი წინა წელს ნამუშევარი კვირები (WEEKSWRK) და განათლების წლების რაოდენობა (EDUC). ორივე ეს ცვლადი კოდირებულია, როგორც ინტერვალურ-ფარდობითი ცვლადი.

თავდაპირველად ამ ორი ცვლადისათვის შევადგინოთ წერტილთა გაბნევის დიაგრამა. პროგრამის შესაბამისი ბრძანება არის მენიუს ბრძანებათა ხაზის Graphs ოპციაში. Graphs ოპციაზე დაწკაპუნების შემდეგ დაკლიკეთ Legacy Dialogs-Scatter/Dot (რაც ნიშნავს, რომ ცვლადებისათვის გვინდა შევადგინოთ წერტილთა გაბნევის დიაგრამა ორი ცვლადისათვის), ავირჩიოთ Simple Scatter (მარტივი დიაგრამა) – Define.

წერტილთა გაბნევის დიაგრამისათვის (ნახ. 13.14) საჭიროა ცვლადების შერჩევა როგორც ვერტიკალური (Y ღერძი) ასევე ჰორიზონტალური ღერძისათვის (X ღერძი). Y ღერძზე დავიტანოთ ცვლადი EDUC, როგორც დამოუკიდებელი ცვლადი, ხოლო X ღერძზე სამუშაო კვირები. დაკლიკეთ OK.

ნახაზი 13.14 წერტილთა გაბნევის დიაგრამის დიალოგური ფანჯარა

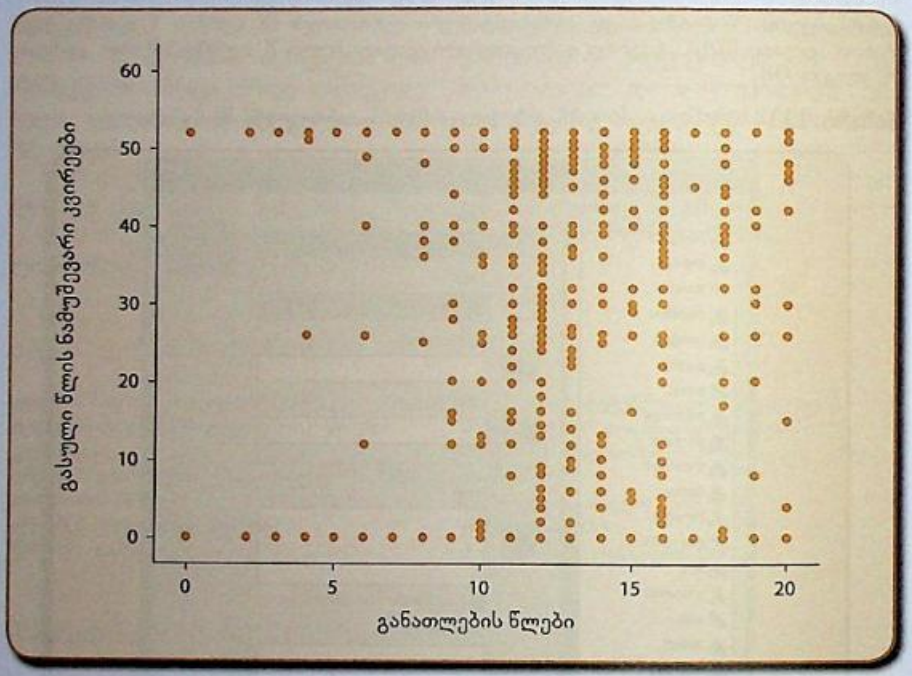

SPSS გამოგვიტანს მოთხოვნილ გრაფიკს (ნახ 13.15). გრაფიკის გარეგნული ფორმის შეცვლა შეგიძლიათ შედეგების ფანჯარაში გრაფიკზე ორჯერ დაწკაპუნებით.

მხოლოდ წერტილთა გაბნევის დიაგრამაზე დაკვირვებით დანამდვილებით ვერ ვიტყვით ცვლადებს შორის კავშირის არსებობის შესახებ. ამისათვის საჭიროა რეგრესიის წრფის დამატება, ამისათვის ორჯერ დაკლიკეთ დიაგრამაზე და გამოტანილ დიალოგურ ფანჯარაში აირჩიეთ ოპცია

Elements –Fit Line at Total, მეთოდებს შორის (Fit Method) აირჩიეთ Liner. შემდეგ: Apply-Close. საბოლოოდ, დიაგრამის რედაქტირების ფანჯარაში დაკლიკეთ File –Close. შედეგები ნაჩვენებია ნახაზ 13.16-ზე.

ნახაზი 13.15 წერტილთა გაბნევის დიაგრამა, განათლება და ნამუშევარი კვირეები



რადგან რეგრესიის მიხედვით განათლების წლების ზრდასთან ერთად იზრდება გასული წლის ნამუშევარი კვირეების რაოდენობა, ეს ნიშნავს რომ ამ ორ ცვლადს შორის დადებითი კავშირია. 20 წლის განათლების მქონე პიროვნების წლის მანძილზე საშუალოდ ნამუშევარი კვირეები 40, ხოლო 10 წლის განათლების მქონეთათვის კი 10 კვირაა. მაგრამ რეგრესიის მოდელის სიზუსტე სუსტია, რადგან რეგრესიის წრფის ირგვლივ დაკვირვების მონაცემები საკმაოდ გაბნეულია.

მაგალითი 2: კორელაციის კოეფიციენტების მიღება

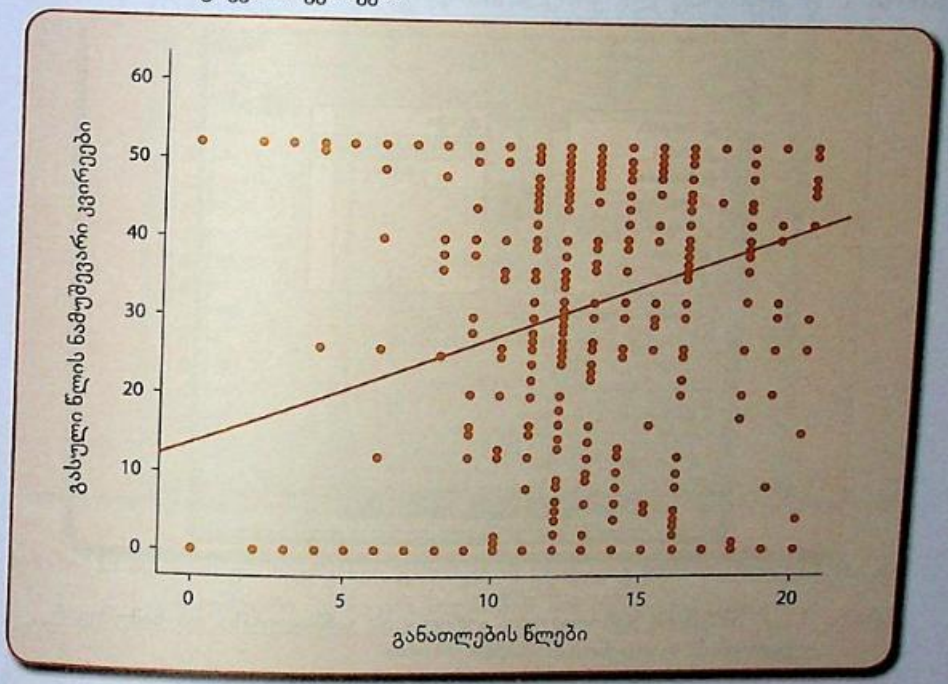
განათლების წლებსა და ნამუშევარ კვირეებს შორის კავშირის შესწავლისათვის გამოთვალოთ კორელაციის კოეფიციენტი. ამისათვის პროგრამის მენიუს ბრძანებათა პანელზე დაკლიკეთ Analyze, Corralate-Bivariate (ნახ.13.17). გადავიტანოთ შესაბამისი ცვლადები EDUC და WEEKSWRK და საბოლოოდ დაკლიკეთ OK.

SPSS გამოიტანს კორელაციების მატრიცას, რომელიც ნაჩვენებია ნახაზ 13.18-ზე. ჩვენ გვაინტერესებს მნიშვნელობა, რომელიც მოცემულია ზედა მარცხენა უჯრაში 0.189. კორელაციის ეს მნიშვნელობა უფრო 0-თან ახლოს, ვიდრე 1-თან, რაც ნიშნავს რომ განათლების წლების რაოდენობა ზუსტად ვერ განსაზღვრავს ნამუშევარი კვირეების რაოდენობას. კორელაციის კოეფიციენტის მნიშვნელობის ქვემოთ მოცემული რიცხვი 1,483 ვალიდური დაკვირვებების რიცხვია (N)-ე.ი.

იმ რესპონდენტთა რიცხვი, რომლებმაც პასუხი გასცეს ორივე შეკითხვას. მათი საერთო რიცხვი შემცირებულია, რადგან ყველა რესპონდენტის პასუხი არ არის ვალიდური.

კორელაცია მნიშვნელოვანია 0.01 დონეზე (ორმხრივი კრიტერიუმით).

ნახაზი 13.16 წერტილთა გაბნევის დიაგრამა და რეგრესიის წრფე, განათლება და ნამუშევარი კვირეები



მაგალითი 3: რეგრესიის წრფის მიღება

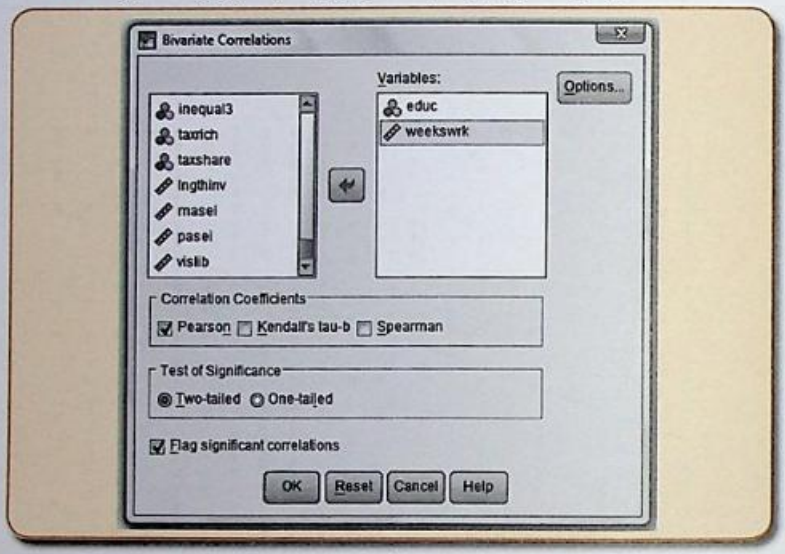
გამოვიყენოთ SPSS დეტრემინაციის კოეფიციენტისა და რეგრესიის წრფის მიღებისათვის. მიმართეთ ბრძანებათა მიმდევრობას პანელზე Analyze-Regresssion-Liner. მიღებულ ბრძანებათა დიალოგურ ფანჯარაში (ნახ. 13.19) გადაიტანეთ დამოკიდებული (ნამუშევარი კვირეები) და დამოუკიდებელი ცვლადები (განათლების წლები) შესაბამის ღერძებზე და დაკლიკეთ OK. წრფივი რეგრესიის ბრძანებათა ფანჯარა სხვა ოპციებსაც გვთავაზობს, მაგრამ ავტომატურად მოცემული ოპციები ჩვენს მიზნებს ამ შემთხვევაში აკმაყოფილებს.

სხვა მოწინავე სტატისტიკური პროგრამების მსგავსად SPSS გამოგვიტანს გამოთვლის საკმაოდ ტევად შედეგებს,. ამჯერად ჩვენ საშედეგო გამოთვლების მხოლოდ ორ ნაწილს გამოვიყენებთ. გამოთვლის შედეგებში დეტერმინაციის კოეფიციენტის დასახლებაა „R Square“. მისი მნიშვნელობა 0.036-ია, რომელიც სუსტ კავშირზე მიუთითებს. განათლების დონე ნამუშევარი კვირეების რაოდენობის ვარიაციის მხოლოდ მცირე ნაწილს ხსნის, დაახლოებით 4%-ს. ეს შედეგი არ არის გასაკვირი თეორიულადაც; მაგალითად, მცირე ბიზნესის მფლობელი, რომელსაც არ აქვს მაღალი განათლება თითქმის თავის სრულ დროს უთმობს საკუთარ ბიზნესს.

რეგრესიის განტოლების შედეგები ნაჩვენებია კოეფიციენტების ნაწილში, ნაჩვენებია b სტატ-

ში. b კოეფიციენტი ცვლად EDUC-ისათვის არის 1.384; თანაკვეთის მნიშვნელობა a 13.874-ია. შესაბამისად, შეგვიძლია ვთქვათ, რომ განათლების ყოველი დამატებითი წლისათვის იზრდება წლის მანძილზე ნამუშევარი კვირეების რაოდენობა, დაახლოებით 11/2 კვირით; ან მაღალი განათლების მქონენი საშუალოდ მუშაობენ 13.874+ (1.384) (12) კვირას ან დაახლოებით 30 კვირას.

ნახაზი 13.17 ბივარიანტული კორელაციის დიალოგური ფანჯარა



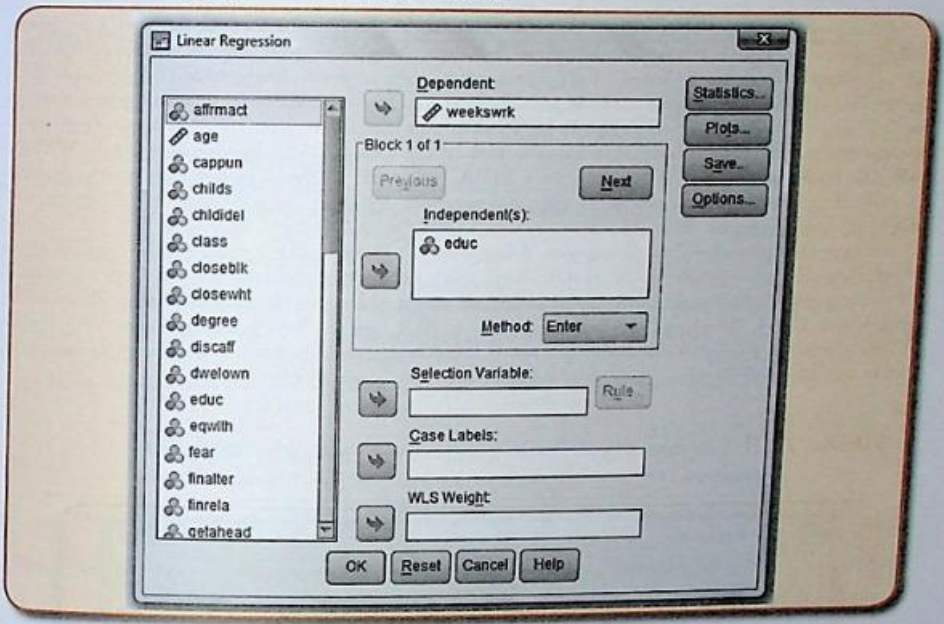
ნახაზი 13.18 SPSS-ის გამოთვლები: კორელაცია განათლებასა და ნამუშევარ კვირეების რაოდენობას შორის

Correlations

		HIGHEST YEAR OF SCHOOL COMPLETED	WEEKS R. WORKED LAST YEAR
HIGHEST YEAR OF SCHOOL COMPLETED	Pearson Correlation	1	.189**
	Sig. (2-tailed)		.000
	N	1496	1483
WEEKS R. WORKED LAST YEAR	Pearson Correlation	.189**	1
	Sig. (2-tailed)	.000	
	N	1483	1487

**. Correlation is significant at the 0.01 level (two-tailed).

ნახაზი 13.19 რეგრესიის დიალოგური ფანჯარა



ნახაზი 13.20 წრფივი რეგრესიის გამოთვლის შედეგები, განათლების დონესა და ნამუშევარი კვირეების ოდენობას შორის კავშირისათვის

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.189[a]	.036	.035	22.367

a. Predictors: (Constant), HIGHEST YEAR OF SCHOOL COMPLETED

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.874	2.572		5.394	.000
	HIGHEST YEAR OF SCHOOL COMPLETED	1.384	.187	.189	7.408	.000

a. Dependent Variable: WEEKS R. WORKED LAST YEAR

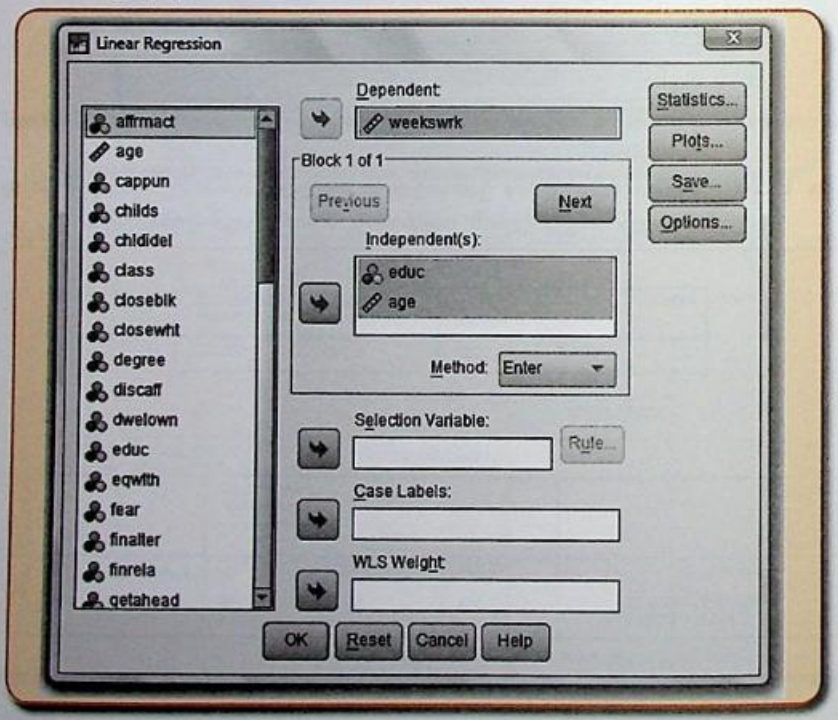
მაგალითი 4: მრავლობითი რეგრესიის განტოლების მიღება

განათლების გარდა სხვა რა ფაქტორები განსაზღვრავს წლის მანიძლზე სრული განაკვეთით ნამუშევარი კვირეების რაოდენობას? ერთი შესაძლო პასუხი ამ კითხვაზე არის „ასაკი“ (AGE). ამ ცვლადებს შორის კავშირის შესწვალისათვის გამოვიყენებთ მრავლობითი რეგრესიის განტოლებას SPSS-ის მიხედვით. აღნიშნული მეთოდის გამოყენება SPSS-ში ბივარიანტული რეგრესიის მსგავსია შემდეგი ბრძანებების მიმდევრობით: Analyze-Regression-Liner. წრფივი რეგრესიის დიალოგური ფანჯრის შესაბამის ოპციებში (ნახ.13.21) გადავიტანოთ შემდეგი ცვლადები WEEKSWRK (დამოკიდებული ცვლადი), ხოლო EDUC და AGE როგორც დამოუკიდებელი ცვლადი.

SPSS გამოგვიტანს საკმაოდ ტევად გამოთვლის შედეგებს, მათგან ჩვენ აქ მხოლოდ ორ ნაწილს გამოვიყენებთ. (ნახ. 13.22).

მოდელის გამოთვლების მიხედვით, მრავლობითი კორელაციის კოეფიციენტი აღნიშნულია როგორც „R“ და უდრის 0.431, რაც ნიშნავს რომ განათლებასა, ასაკსა და ნამუშევარ კვირეებს შორის კავშირი საშუალო სიძლიერისაა. SPSS ასევე გამოგვიტანს მოსწორებული R-ის მნიშვნელობას 0.184. ზოგადად, მოსწორებული r^2-ის მნიშვნელობა მით უფრო დაბალია, რაც მეტია საპროგნოზო ცვლადების რაოდენობა. r^2-ის 0.186-თან ტოლობა ნიშნავს, რომ განათლება და ასაკი ერთობლივად ნამუშევარი კვირეების რაოდენობის ვარიაციის 18.6%-ს ხსნის.

ნახაზი 13.21 წრფივი რეგრესიის დაილოგური ფანჯარა რამდენიმე ცვლადით



ნახაზი 13.22 მრავლობითი რეგრესიის გამოთვლის შედეგები: განათლების დონე, ასაკი და გასული წლის ნამუშევარი კვირეები

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.431[a]	.186	.184	20.540

a. Predictors: (Constant), AGE OF RESPONDENT, HIGHEST YEAR OF SCHOOL COMPLETED

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	41.187	2.898		14.214	.000
	HIGHEST YEAR OF SCHOOL COMPLETED	1.189	.172	.163	6.902	.000
	AGE OF RESPONDENT	-.510	.031	-.387	-16.396	.000

a. Dependent Variable: WEEKS R. WORKED LAST YEAR

რეგრესიის განტოლების გამოთვლების შედეგები ნაჩვენებია კოეფიციენტების ცხრილში, სვეტში. განათლებისათვის კოეფიციენტის მნიშვნელობა 1.189, ხოლო ასაკისათვის -0.510-ია. თანაკვეთა, a, რომელიც აღნიშნულია როგორც მუდმივა (Constant) და უდრის 41.187. აქედან გამომდინარე, შეგვიძლია ვთქვათ რომ ასაკის მუდმივად შენარჩუნებით, განათლების ყოველი დამატებითი წელი ნამუშევარი კვირეების რაოდენობას საშუალოდ 1 კვირით ზრდის. ასევე, განათლების მუდმივად შენარჩუნებით ყოველი დამატებითი წელი ასაკისათვის ამცირებს ნამუშევარი კვირეების რაოდენობას დაახლოებით ნახევარი კვირით. აქედან გამომდინარე შეგვიძლია გავაკეთოთ პროგნოზი-უმაღლესი განათლების მქონე და 20 წლის ასაკის პიროვნება საშუალოდ 41.187+ (1.189) (12)+ (-0.510) (20), 45 კვირას იმუშავებს წლის განმავლობაში სრული სამუშაო დატვირთვით.

ᲐᲛᲝᲪᲐᲜᲔᲑᲘ SPSS-ᲘᲡᲗᲕᲘᲡ

[GSSS08PFP-A]

1. გამოიყენეთ მონაცემთა ფაილი GSSS08PFP-A შემდეგ ცვლადებს შორის კავშირის შესწავლისათვის: და-ძმების რაოდენობა, რომელიც პირს ყავს (SIBS) და ამ პირის ბავშვების რაოდენობა (CHILDS).
ა. ნახსენები ორი ცვლადისათვის შეადგინეთ წერტილთა გაბნევის დიაგრამა და მასზე დატანეთ რეგრესიის წრფე SPSS-ში. აღწერეთ ცვლადებს შორის კავშირი.
ბ. გამოთვალეთ რეგრესიის განტოლების კოეფიციენტები, რა არის დახრილობისა და თანაკვეთის მნიშვნელობა? რას უდრის დეტერმინაციისა და კორელაციის კოეფიციენტები?
გ. მიღებული რეგრესიის მიხედვით რამდენი შვილი ეყოლება პირს, რომელსაც 3 და-ძმა ყავს?
დ. რამდენი ბავშვი ეყოლება პირს, რომელსაც და ან ძმა არ ყავს?
ე. იპოვეთ გზა, რომლითაც გამოთვლით პროგნოზის შეცდომასა და გამოთვილ მნიშვნელობას თითოეული ცვლადისათვის SPSS-ში, ეს გამოთვლები შეინახეთ ახალი ცვლადის სახით.

2. გამოიყენეთ 1 დავალებაში მოცემული ცვლადები, მაგრამ ამჯერად ანალიზი გააკეთეთ განცალკევებულად ქალისა და მამაკაცისათვის. ე.ი. დაამატეთ ცვლადი სქესი (SEX). პროგრამის ბრძანებათა პანელზე დაკლიკეთ Data-Split File-Organize Output by Groups. გადაიტანეთ ცვლადი SEX გამოტანილ ოპციაში, დაკლიკეთ OK. ამჯერად, SPSS გამოთვლის შედეგებს გამოიტანს სქესის მიხედვით.

ა. SPSS-გამოიტანს გამოთვლის შედეგებს ქალებისა და მამაკაცების მიხედვით. შეადარეთ რამდენად მსგავსია გამოთვლის ეს შედეგები?

ბ. რამდენი ბავშვი ეყოლება მამაკაცს, რომელსაც ორი და-ძმა, ექვსი და-ძმა ყავს? როგორია იგივე მონაცემი ქალების შემთხვევაში?

3. გამოიყენეთ 1 დავალების ცვლადები, ამ შემთხვევაში შედეგები გამოიტანეთ თეთრკანიანი და შავკანიანი რესპონდენტების მიხედვით. დამატებით აირჩიეთ ცვლადი RACEN1. დაკლიკეთ ბრძანეთა შემდეგ მიმდევრობას Data-Split File-Organize Output by Groups. გადაიტანეთ ცვლადი RACEN1 დიალოგური ფანჯრის ცვლადის ოპციაში და დაკლიკეთ OK. (ცვლადი SEX დააბრუნეთ უკან). ამჯერად SPSS გამოთელის შედეგებს გამოიტანს ცვლად RACEN1-ის მიხედვით:

ა. არის თუ არა განსვავებული რეგრესიის გამოთვლები შავკანიანი და თეთრკანიანი რეპოდენტებისათვის?

ბ. რა არის ბავშვთა მოსალოდნელი რაოდენობა თეთრკანიანი და შავკანიანი რესპონდენტებისათვის, რომელთაც ყავთ: ერთი, ოთხი და შვიდი და-ძმა?

4. ისევ გამოიყენეთ 1 დავალების ცვლადები, ამჯერად შედეგები დაყავით დაქორწინებული და განქორწინებული რესპონდენტების მიხედვით. ამისათვის გამოიყენეთ ცვლადი MARITAL. დაკლიკეთ ბრძანებათა შემდეგ მიმდევრობას: Data-Split File-Organize Output by Groups. გადიტანეთ ცვლადი MARITAL ცვლადის ოპციაში (არ დაგავიწყდეთ ცვლად RACEN1-ისა და ცვლად SEX-ის უკან დაბრუნება). SPSS გამოიტანს გამოთვლის შედეგებს ქორწინების სტატუსის მიხედვით:

ა. არის თუ არა განსხვავებული რეგრესიის გამოთვლის შედეგები დაქორწინებული და განქორწინებული რესპონდენტების მიხედვით?

ბ. რამდენია ბავშვთა რაოდენობა იმ დაქორწინებული და განქორწინებული რესპონდენტებისათვის, რომელთაც ყავთ ერთი, ოთხი და შვიდი და-ძმა?

გ. რა განსხვავებებს პოულობთ შედეგების მიხედვით, ასეთის არსებობის შემთხვევაში? არის თუ არა და-ძმების რაოდენობა უფრო კარგი მსაზღვრელი ფაქტორი ბავშვების რაოდენობის განსაზღვრისათვის ქალებისათვის ან დაქორწინებული წყვილებისათვის?

5. გამოიყენეთ ფაილი 2008 GSS[GSS08PFP-A], რათა შეისწავლოთ კავშირი რესპონდენტის განათლებასა (EDUC) და მისი მშობლების მამა და დედა განათლებას შორის (PAEDUC და MAEDUC შესაბამისად).

ა. SPSS-ის გამოყენებით გამოთვალეთ კორელაციის კოეფიციენტი, დეტერმინაციის კოეფიციენტი და რეგრესიის განტოლება რესპონდენტის განათლების განსასაზღვრად მათი მამის განათლების დონის მიხედვით. მოახდინეთ შედეგების ინტერპრეტირება.

ბ. SPSS-ის გამოყენებით გამოთვალეთ: მრავლობითი კორელაციის კოეფიციენტი, მრავლობითი დეტერმინაციის კოეფიციენტი და რეგრესიის განტოლება რესპონდენტის განათლებისათვის შემდეგი დამოუკიდებელი ცვლადების მიხედვით: მამისა და დედის განათლება. მოახდინეთ შედეგების ინტერპრეტირება.

გ. გააუმჯობესა თუ არა ცვლად „დედის განათლების“ ჩართვამ რეგრესიის გამოთვლები? იმსჯელეთ ამ საკითხზე ამავე დავალების წინა პუნქტში მიღებული შედეგების მიხედვით.

დ. გამოიყენეთ მიღებული რეგრესიის განტოლება და გამოთვალეთ იმ პირის მოსალოდნელი განათლება, რომელსაც ყავს მამა 12 წლის განათლებით. შემდეგ, კვლავ გაიმეორეთ ეს პროცედურა დედის 12 წლიანი განათლების დამატებით.

თავის სავარჯიშოები

1. სხვადასხვა ფაქტორების გავლენის გამო დღესდღეისობით მოსახლეობის სულ უფრო მეტი ოდენობაა დაინტერესებული გარემოს მდგომარეობით, ვიდრე ეს წარსულში იყო. შეიქმნა საკმაოდ ბევრი გარემოს დაცვითი ჯგუფები, რომელებიც ცდილობენ გარემოსდაცვითი პოლიტიკის შეცვლას ეროვნულ და გლობალურ დონეზე. რვა ქვეყნის მონაცემებზე დაყრდნობით განვსაზღვროთ კავშირი გარემოსდაცვითი საკითხებით დაინტერესებასა და გარემოს დაცვითი ჯგუფებისათვის ფულის შემოწირვისათვის მზაობას შორის.

ქვეყანა	დაინტერესებულთა პროცენტული წილი	თანხის შემოწირველთა პროცენტული წილი
ავსტრია	35.5	27.8
დანია	27.2	22.3
ნიდერლანდები	30.1	44.8
ფილიპინები	50.1	6.8
რუსეთი	29.0	1.6
სლოვენია	50.3	10.7
ესპანეთი	35.9	7.4
ამერიკის შეერთებული შტატები	33.8	22.8

წყარო: International Survey Programme, 2000.

ა. ააგეთ წერტილთა გაბნევის დიაგრამა ორი ცვლადისათვის, X-ღერძზე განათავსეთ ცვლადი „გარემოს საკითხებით დაინტერესება“, ხოლო Y-ღერძზე „თანხის შემოწირულობა“.

ბ. არის თუ არა ვიზუალურად ამ ორ ცვლადს შორის კავშირი წრფივი? აღწერეთ ეს წრფივი კავშირი.

გ. იპოვეთ პირსონის კორელაციის კოეფიციენტის მნიშვნელობა და მოახდინეთ მიღებული შედეგის ინტერპრეტირება.

2. მიჩნეულია, რომ არსებობს კავშირი პირის განათლების დონესა და მის ბავშვების რაოდენობას შორის. კვლევის ჰიპოთეზაა, რომ რაც უფრო მაღალია პირის განათლების დონე მით უფრო ნაკლები ბავშვები ყავს ამ პირს. შეისწავლეთ ეს კავშირი 25 რესპონდენტის მონაცემებზე დაყრდნობით, რომელიც ზოგადი სოციოლოგიური კვლევის 2006 წლის მონაცემებიდანაა აღებული. ქვემოთ მოტანილ ცხრილში ნაჩვენებია თითოეული რესპონდენტის ბავშვების რაოდენობა განათლებაზე დახარჯული წლების მიხედვით:

განათლების წლები	ბავშვების რაოდენობა	განათლების წლები	ბავშვების რაოდენობა
16	0	12	2
12	1	12	3
12	3	11	1
6	6	12	2
14	2	11	2
14	2	12	0
16	2	12	2

12	2	12	3
17	2	12	4
12	3	12	1
14	4	14	0
13	0	12	3
12	1		

ა. ამ მონაცემებისათვის გამოთვალეთ პირსონის კორელაციის კოეფიციენტი. ამართლებს თუ არა პირსონის კორელაციის მნიშვნელობა ჩვენს მიერ ჩამოყალიბებულ კვლევის ჰიპოთეზას?

ბ. გამოთვალეთ უმცირეს კვადრატთა მეთოდით რეგრესიის განტოლება, განათლება გამოიყენეთ, როგორც დამოუკიდებელი ცვლადი. რა უდრის წრფის დახრილობა და b კოეფიციენტის, a თანაკვეთის მნიშვნელობა?

გ. ჰყავს თუ არა რეგრესიით მიღებული ბავშვების რაოდენობა რომელიმე რესპონდენტს? თუ კი, რა არის მისი განათლების დონე? თუ პირველ კითხვაზე პასუხი უარყოფითია, ნიშნავს თუ არა ეს რომ რეგრესიის განტოლება არ შეესაბამება არსებულ მონაცემებს? პასუხი ახსენით.

3. ჯანმრთელობისა და გარემოს მდგომარეობა თანამედროვეობის აქტუალური პრობლემაა. განვიხილოთ კავშირი ქვეყნის მთლიან ეროვნულ შემოსავალსა (GNP) და იმ ადამიანების პროცენტულ წილს შორის, რომელთაც აქვთ სურვილი გადაიხადონ უფრო მაღალი ფასები გარემოს დაცვისათვის. ქვემოთა ცხრილში მოტანილია შერჩევითი მონაცემები ხუთი ქვეყნის მიხედვით.

ა. გამოთვალეთ კორელაციის კოეფიციენტის მნიშვნელობა ერთობლივ ეროვნულ შემოსავალსა (GNP) და იმ ადამიანების პროცენტულ წილს შორის, რომლებიც მზად არიან გადაიხადონ უფრო მაღალი ფასები გარემოს დაცვისათვის.

ბ. მოახდინეთ კოეფიციენტების მნიშვნელობების ინტერპრეტირება.

ქვეყანა	ერთობლივი ეროვნული პროდუქტი ერთ მოსახლეზე	გადახდის სურვილის მქონეთა პროცენტული წილი
აშშ	29.24	44.9
ირლანდია	18.71	53.3
ნიდერლანდები	24.78	61.2
ნორვეგია	34.31	40.7
შვედეთი	25.28	32.6

წყარო: International Survey Programme, 2000.

4. მე-5 თავის მე-9 სავარჯიშოში განვიხილეთ დანაშაულის დონესა და პოლიციაზე მთავრობის დანახარჯები აშშ-ს ალმოსავლეთ და ცენტრალურ შტატებში. ამჯერად შევამოწმოთ ჰიპოთეზა იმის თაობაზე, რომ რაც უფრო მაღალია შტატში დანაშაულის დონე, მით უფრო მაღალია პოლიციაზე დანახარჯი ამავე შტატში, რადგან დანაშაულის ზრდისას მთავრობა ზრდის პოლიციის შემადგენლობას და იზრდება პოლიციაზე დანახარჯებიც.

ა. შეადგინეთ წერტილთა გაბნევის დიაგრამა დანაშაულის დონისა და პოლიციაზე დანახარჯების მიხედვით, დანაშაულის დონე მიიჩნიეთ დამოუკიდებელ ცვლადად. რა შეგიძლიათ

თქვათ დიაგრამის მიხედვით ამ ორი ცვლადის კავშირის შესახებ?

ბ. შეადგინეთ უმცირეს კვადრატთა მეთოდის გამოყენებით რეგრესიის განტოლება. რას უდრის დახრილობა? თანაკვეთა?

გ. გამოთვალეთ დეტერმინაციის კოეფიციენტი (r^2) და მოახდინეთ მისი მნიშვნელობის ინტერპრეტირება.

დ. თუ შტატში დანაშაულის დონე 2,000-ით გაიზრდება პოლიციაზე დანახარჯების ზრდას რამდენი ათასით უნდა ველოდოთ?

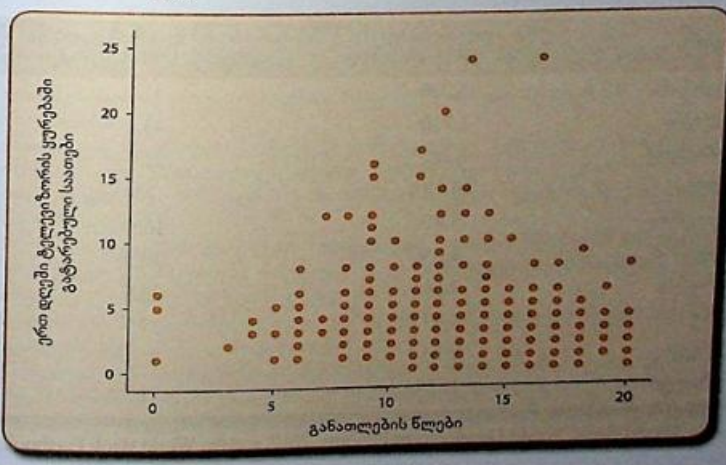
ე. აქვს თუ არა აზრი პოლიციაზე დანახარჯების პროგნოზირებას, როცა დანაშაულის დონე ნოლს უტოლდება? პასუხი ახსენით.

შტატი	დანაშაულის რიცხვი (ყველ 100,000 მოსახლეზე)	პოლიციაზე დანახარჯები (მილიონი დოლარი)
მეინი	2,635	221
ნიუ ჰემპშირი	2,013	274
ვერმონტი	2,442	136
მასაჩუსეტი	2,838	1,673
როდ აილენდი	2,815	286

წყარო: U.S. Census Bureau, The 2012 Statistical Abstarct, Tables 297,431.

5. კორელაციის კოეფიციენტისა და რეგრესიის განტოლების გამოთვლამდე ყოველთვის მნიშვნელოვანია ცვლადებს შორის კავშირი თავდაპირველად წერტილთა გაბნევის დიაგრამის მეშვეობით შევამოწმოთ და ვნახოთ თუ რამდენად კარგად შეესაბამება რეგრესიის წრფე მონაცემებს. თუ რეგრესიის წრფის ირგვლივ მონაცემები ძალიან არის გაფანტული, მაშინ სხვა სახის მრუდი უნდა გამოვიყენოთ ცვალდებს შორის კავშირის აღწერისათვის.

ნახაზი 13.23 წერტილთა გაბნევის დიაგრამა განათლებისა და ტელევიზორის ყურების საათების რაოდენობას შორის



SPSS-ში დამუშავებული წერტილთა გაბნევის დიაგრამა ნახაზ 13.23-ზე და გამოთვლის შედეგები ნახაზ 13.24-ზე ასახავს კავშირს განათლებასა (წლების მიხედვით) და ტელევიზორის ყურებას შორის (საათების მიხედვით) საერთო სოციოლოგიური კვლევის 2008 წლის მონაცემების მიხედვით (GSS2008). დავუშვათ ჰიპოთეზა, რომ განთლების ზრდასთან ერთად კლებულობს ტელევიზორის ყურებაში გატარებული საათები.

ნახაზი 13.24 წრფივი რეგრესიის გამოთვლა, კავშირი განათლებასა და ტელევიზორის ყურების საათების რაოდენობას შორის

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	5.575	.367		15.175	.000
	HIGHEST YEAR OF SCHOOL COMPLETED	-.196	.027	-.229	-7.335	.000

a. Dependent Variable: HOURS PER DAY WATCHING TV

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.229[a]	.052	.051	2.516

a. Predictors: (Constant), HIGHEST YEAR OF SCHOOL COMPLETED

ა. შეაფასეთ ორ ცვლადს შორის კავშირი წერტილთა გაბნევის დიაგრამისა და b კოეფიციენტის მიხედვით. არის თუ არა კავშირი ამ ორ ცვლადს შორის იმ დაშვების მსგავს, რაც გავაკეთეთ? დადებითია თუ უარყოფითი არსებული კავშირი?

ბ. აღწერეთ ამ ორ ცვლადს შორის კავშირი განათლების წლებისა და ტელევიზორის ყურების საათების შესახებ მონაცემების რეპრეზენტატიული რაოდენობის გამოყენებით.

გ. აღწერს თუ არა ამ ცვლადებს შორის (ვიზუალურად) კავშირს წრფე? პასუხი გამართეთ.

დ. სამხერთ ამერიკის ქვეყნების მონაცემების გამოყენებით შეისწავლოთ კავშირი ერთ სულ მოსახლეზე მთლიანი შიდა პროდუქტის რაოდენობასა (GDP) და ბავშვთა სიკვდილიანობის დონეს შორის (IMR).

ქვეყანა	ერთ სულ მოსახლეზე, 2008 აშშ დოლარებში	ბავშვთა სიკვდილიანობის დონე, 2012 წ. (ყოველ 1000 დაბადებულზე)
არგენტინა	8,236	11.1
ბოლივია	1,720	43.4
ბრაზილია	8,205	21.9
ჩილე	10,084	7.5
კოლუმბია	5,416	16.9
ეკვადორი	4,056	20.3
პარაგვაი	2,561	23.8
პერუ	4,477	27.7
ურუგვაი	9,654	11.0
ვენესუელა	11,246	21.2

წყარო: World Bank, World Development Indicatrs, 2012, and the World Bank Factbook, 2010.

ა. ააგეთ ამ მონაცემებისათვის წერტილთა გაბნევის დიაგრამა, მიიჩნიეთ GDP როგორც დამოუკიდებელი, ხოლო IMR, როგორც დამოკიდებული ცვლადი. რა კავშირია ამ ორ ცვლად შორის სამხრეთ ამერიკის 10 ქვეყნის მონაცემების მიხედვით?

ბ. ერგება თუ არა ვიზუალურად ამ მონაცემებს წრფე? პასუხი ახსენით.

გ. გამოთვალეთ კორელაციისა და დეტერმინაციის კოეფიციენტები. ამ კოფიციენტების მიხედვით დააზუსტეთ თქვენი პასუხი ბ კითხვაზე.

7. მეცნიერები დიდი ხანია დაინტერესებულნი არიან ამერიკის შეერთებულ შტატებში მისწრაფებებისა და მიღწევების შესწავლით. კვლევებმა სოციალურ მობილურობაზე, საზოგადოების წევრების სოციალურ სტატუსებზე და განათლების დონეზე მეცნიერებს მისცა საკმაო მასალები ბავშვებისა და მათი მშობლების სოციალურ-ეკონომიკურ მიღწევებს შორის კავშირის შესწავლისათვის. ზოგადი სოციოლოგიური კვლევის მონაცემებში (GSS 2006) შეტანილია მონაცემები რესპონდენტებისა და მათი დედების განათლების დონის შესახებ. ამ მონაცემების გამოყენებით შევეცადოთ გავარკვიოთ: აქვს თუ არა იმ რესპონდენტებს უფრო მაღალი განათლების დონე, რომელთა დედებსაც შედარებით მაღალი განათლების დონე აქვთ.

დედის განათლების წლები	რესპონდენტის განათლების წლები
0	12
15	13
6	9
9	12
16	16
12	12
6	16
18	14
12	13
14	12
14	18
7	12

ა. ააგეთ წერტილთა გაბნევის დიაგრამა, რესპონდენტისა და მისი დედის განათლების დონის მონაცემებისათვის.

ბ. გამოთვალეთ რეგრესიის განტოლება, დამოუკიდებელ ცვლადად გამოიყენეთ დედის განათლება, ხოლო დამოკიდებულ ცვლადად რესპონდენტის განათლება. დაიტანეთ რეგრესიის წრფე დიაგრამაზე. რას უდრის დახრილობა, თანაკვეთა? როგორ შეესაბამება რეგრესიის წრფე მონაცემებს?

გ. რას უდრის შეცდომის დონე მეორე რესპონდენტებისათვის (13 წლის საკუთარი, ხოლო 15 წლის დედის განათლებით)? რა უდრის იგივე მაჩვენებელი რესპონდენტისათვის 14 წლის საკუთარი და დედის 18 წლის განათლებით?

დ. რა არის იმ პირის მოსალოდნელი განათლების დონე, რომლის დედას აქვს 4 წლიანი განათლება, აქვს 12 წლიანი განათლება?

ე. გამოთვალეთ რესპონდენტებისა და რესპონდენტების დედების განათლების საშუალო მაჩვენებლები? აღნიშნეთ ეს მაჩვენებლები წერტილთა გაბნევის დიაგრამაზე. სად ხვდებიან ეს საშუალო მაჩვენებლები? შეგიძლიათ დაასაბუთოთ მათი ასეთი განლაგების შესაძლო მიზეზები?

8. მე-6 დავალებაში ჩვენ შევისწავლეთ კავშირი ბავშვების მოკვდავობის დონესა და მთლიან ეროვნულ შემოსავალს შორის სამხრეთ ამერიკის ქვეყნებში. იგივე მონაცემები 2010 წლისათვის არის შემდეგი:

ქვეყანა	დაბადებულთა ოდენობა 2010 წელი (პროგნოზი)
არგენტინა	18
ბოლივია	25
ბრაზილია	18
ჩილე	15
კოლუმბია	18
ეკვადორი	20
პარაგვაი	18
პერუ	19
ურუგვაი	14
ვენესუელა	13

წყარო: The World Factbook, 2010.

ა. ააგეთ წერტილთა გაბნევის ორი დიაგრამა, ერთი მათგანი დაბადების დონისა და მთლიანი ეროვნული პროდუქტისათვის, ხოლო მეორე ბავშვთა დაბადებისა და მოკვდავობის დონისათვის. შესაძლებელია თუ არა ამ ცვლადებს შორის კავშირი დახასიათდეს როგორც წრფივი?

ბ. გამოთვალეთ დეტერმინაციისა და კორელაციის კოეფიციენტები ამ მონაცემებისათვის.

გ. გამოიყენეთ წინა გამოთვლის შედეგები მონაცემთა შორის კავშირის აღწერისათვის.

9. 2010 წელს აშშ-ს მოსახლეობის აღწერის ბიურომ გამოაქვეყნა მონაცემები, რომლის მიხედვით აშშ-ს მოსახლეობის 13% 2007 წლისათვის ცხოვრობდა სიღარიბის ზღვარს ქვემოთ. ეს მაჩვენებელი აღემატება 200 წლის შესაბამის მაჩვენებელს 12.2%-ით, რაც ნიშნავს რომ აშშ-ს მოსახლეობის დაახლოებით 4.75 მილიონის ცხოვრების პირობები გაუარესდა 2007 წლისათვის. ამ ტიპის ოჯახები და პირები დგანან მრავალი პრობლემის წინაშე მაგალითად, მათთვის პრობლემაა ჯანდაცვის მომსახურების ხელმისაწვდომობა, რადგან მათ ჩვეულებრივ არ აქვთ სამედიცინო დაზღვევა. ქვემოთ მოყვანილი აშშ-ს აღწერის ბიუროს მონაცემებზე დაყრდნობით შეაფასეთ კავშირის სიღარიბის ზღვარს ქვემოთ ცხოვრებასა და ჯანდაცვის მომსახურების ხელმისაწვდომობას შორის.

შტატი	სიღარიბის ზღვარს ქვემოთ მცხოვრებთა პროცენტული წილი (2007)	სამედიცინო დაზღვევის არმქონეთა პროცენტული წილი (2007)
ალაბამა	16.9	12.0
კალიფორნია	12.4	18.2
აიდაჰო	12.1	13.9
ლუიზიანა	18.6	18.5
ნიუ-ჯერსი	8.6	15.8
ნიუ-იორკი	13.7	13.2

პენსილვანია	11.6	9.5
როდ-აილენდი	12.0	10.8
სამხრეთ კაროლინა	15.0	16.4
ტეხასი	16.3	25.2
ვაშინგტონი	11.4	11.3
ვისკონსინი	10.8	8.2

წყარო: U.S. Cenusu Bureau, The 2010 Statistical Abstract, 2010, Tables 693,150

ა. ააგეთ წერტილთა გაბნევის დიაგრამა დაზღვევის არმქონეთა და სიღარიბის ზღვარს ქვემოთ მცხოვრებთა მონაცემებისათვის.

ბ. გამოთვალეთ რეგრესიის განტოლება, დამოკიდებელ ცვლადად მიიჩნიეთ სამედიცინო დაზღვევის არმქონე მოსახლეობა, ააგეთ დიაგრამაზე რეგრესიის წრფე. რა უდრის წრფის დახრილობა? თანაკვეთის მნიშვნელობა? ეთანადება თუ არა რეგრესიის წრფე მონაცემებს? რომელი შტატის მონაცემები არ ძევს რეგრესიის წრფეზე?

გ. მოსახლეობის რა პროცენტი უნდა ცხოვრობდეს სიღარიბის ზღვარს ქვემოთ, რომ მათ 5%-ს არ ქონდეს სამედიცინო დაზღვევა?

დ. იმ მონაცემის დახასიათება, რომელიც გადის ორივე მონაცემთა ცნობილი რიგიდან ყოველთვის პრობლემატური და სტატისტიკურად ნაკლებად სარწმუნოა. აღნიშნული ეხება თქვენს პასუხს გ დავალებაში. მაგრამ სტატისტიკური მნიშვნელობის გარდა, ხშირად ამ სახის მონაცემის განხილვას აქვს გარკვეული აზრი. თქვენი აზრით ეხება თუ არა აღნიშნული გ დავალებაში მოცემულ პირობას?

10. განვიხილოთ კავშირი ერთ სულ მოსახლეზე ერთობლივ ეროვნულ პროდუქტსა და იმ მოსახლეობის პროცენტული წილს შორის, რომლებიც მზად არიან გადაიხადონ უფრო მაღალი გადასახადები.

ა. ამ თავის 13.4 ცხრილში ნაჩვენებია რეგრესიის წრფისა და ღერძთან თანაკვეთის მნიშვნელობის გამოთვლის გზა. ცხრილ 13.4-ის მიხედვით შეადგინეთ ცხრილი მსგავსი გამოთვლებით ნახსენები ცვლადებისათვის.

ბ. პირველ დავალებაში შესრულებული ცხრილის მიხედვით გამოთვალეთ a და b კოეფიციენტების მნიშვნელობა და შეადგინეთ რეგრესიის განტოლება.

გ. გამოთვალეთ შეცდომის ზომა 2 და მოახდინეთ მისი მნიშვნელობის ინტერპრეტირება.

დ. გ დავალებაში მიღებული შედეგების მიხედვით გამოთვალეთ PRE და r^2. მოახდინეთ შედეგების ინტერპრეტირება.

ე. მოსახლეობის რა პროცენტულ ნაწილს აქვს სურვილი გადაიხადოს მეტი გადასახადები თუ ერთ სულზე მთლიანი შიდა პროდუქტი 3.0 (ე. ი. 3,000$)-ს უტოლდება? იგივე შედეგები გამოთვალეთ GNP-ის 30.0 ($30,000)-თან ტოლობის შემთხვევაში?

ქვეყანა	ერთ სულ მოსახლეზე GNP-ის მოცულობა	მაღალი გადასახადების გადახდის მსურველთა პროცენტული წილი
კანადა	19.71	24.0
ჩილე	4.99	29.1
ფინეთი	24.28	12.0
ირლანდია	18.71	24.3
იაპონია	32.35	27.2
ლატვია	2.42	17.3

მექსიკა	3.84	34.7
ნიდერლანდები	24.78	51.9
ახალი ზელანდია	14.60	31.1
ნორვეგია	34.31	22.8
პორტუგალია	10.67	17.1
რუსეთი	2.66	29.9
ესპანეთი	14.10	22.2
შვედეთი	25.58	19.5
შვეიცარია	39.98	33.5
ამერიკის შეერთებული შტატები	29.24	31.5

წყარო: The World Bank Group, Development Education Program Learning Module: Economics, GNP per Capia, 2004. International Social Survey Programme, 2000.

11. მე-5 სავარჯიშოში განვიხილეთ კავშირი განათლების დონესა და ტელევიზორის ყურებაში გატარებულ დროს შორის და ვნახეთ რომ განათლების დონის ზრდასთან ერთად კლებულობს ტელევიზორის ყურებაში გატარებული საათები. ასევე, ტელევიზორის ყურების ხანგრძლივობაზე გავლენას ახდენს ოჯახში ბავშვების რაოდენობა, ამ უკანასკნელს შორის კი პირდაპირპროპორციული დამოკიდებულებაა. ქვემოთ მოყვანილ 13.25 ნახაზზე ნაჩვენებია SPSS -ის გამოთვლის შედეგები (GSS 2008) ტელევიზორის ყურებაში გატარებულ დროსა (საათების მიხედვით), განათლებასა (წლების მიხედვით) და ბავშვების რაოდენობას შორის. დავუშვათ ჰიპოთეზა, რომ განათლების დონის ზრდა იწვევს ტელევიზორის ყურების საათების კლებას, ხოლო ბავშვების ყოლას საპირისპირო შედეგი მოაქვს.

ნახაზი 13.25 მრავლობითი რეგრესია განათლება, ტელევიზორის ყურება და ბავშვების ყოლას შორის კავშირი.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	5.209	.399		13.067	.000
	HIGHEST YEAR OF SCHOOL COMPLETED	-.185	.027	-.216	-6.853	.000
	NUMBER OF CHILDREN	.117	.050	.074	2.348	.019

a. Dependent Variable: HOURS PER DAY WATCHING TV

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.240[a]	.058	.056	2.512

a. Predictors: (Constant), NUMBER OF CHILDREN, HIGHEST YEAR OF SCHOOL COMPLETED

ა. რას უდრის b კოეფიციენტის მნიშვნელობა ცვლად განათლებისათვის, ცვლად ბავშვების რაოდენობისათვის? გააკეთეთ შედეგების ინტერპრეტირება. შეესაბამება თუ არა გამოთვლის შედეგები ჩვენს მიერ გამოთქმულ ჰიპოთეზას.

ბ. რეგრესიის განტოლების გამოყენებით, რომელშიც ბავშვების რაოდენობა და განათლება დამოუკიდებელი ცვლადებია, გამოთვალეთ იმ პირის მიერ ტელევიზორის ყურებაში გატარებული დრო, რომელსაც აქვს 16 წლის განათლება და ყავს ორი ბავშვი? შეადარეთ ეს შედეგი მე-5 სავარჯიშოში მიღებულ შედეგებს.

გ. შეადარეთ მიღებული r^2-ის მნიშვნელობა მე-5 სავარჯიშოში მიღებულ r^2-ის მნიშვნელობას. ამცირებს თუ არა პროგნოზირების შეცდომას ცვლად განათლებისა და ბავშვების რაოდენობის ერთობლივად ჩართვა მოდელში?

12. 2004 წელს აშშ-ს მოსახლეობის აღწერის ბიურომ გამოაქვეყნა მონაცემები, რომლის მიხედვით ფედერალური სიღარიბის ზღვარს ქვემოთ მცხოვრები ადამიანების რიცხვი ყოველთვის მაღალი იყო. ჩვენი მიზანია გავარკვიოთ შესაძლებელია თუ არა შტატების მიხედვით იმ მოქალაქეთა პროცენტული წილის პროგნოზირება, რომლებიც ფედერალური სიღარიბის ზღვარს ქვემოთ იმყოფებიან მათ განათლებასა და თავად შტატის მოსახლეობის რასობრივი შემადგენლობის სტრუქტურის მიხედვით. ნაზად 13.26-ზე ნაჩვენებია მრავლობითი რეგრესიული ანალიზის შედეგები ($N = 50$ შტატს), მოცემული რეგრესიულ ანალიზი განსაზღვრავს სიღარიბის ზღვარს ქვემოთ მცხოვრები მოსახლეობის წილს 2002 და 2003 წლებში თითოეული შტატის ფერადკანიანი მოსახლეობისა და განათლების მონაცემებზე დაყრდნობით. გამოიყენეთ ეს მონაცემები ქვემოთ მოცემულ კითხვებზე საპასუხოდ.

ნახაზი 13.26 მრავლობითი რეგრესია, სიღარიბის ზღვარს ქვემოთ მცხოვრებთა წილის განსაზღვრა რასობრივი შემადგენლობისა და განათლების დონის მიხედვით

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	66.477	7.242		9.180	.000
	% BLACK RESIDENTS	–.045	.034	–.144	–1.304	.199
	% W/ HS DIPLOMA	–.635	.082	–.851	–7.718	.000

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.779*	.607	.591	1.91239

ა. რას უდრის b კოეფიციენტი შავკანიანი რეზიდენტებისათვის თითოეულ შტატში? უმაღლესი სკოლის დიპლომის მქონე რეზიდენტებისათვის? მოახდინეთ თითოეული კოეფიციენტის ინტერპრეტირება. ამყარებს თუ არა ეს შედეგები ჩვენს დაშვებას სიღარიბის დონის ნაწილობრივ განათლებისა და რასობრივი შემადგენლობით ახსნის შესახებ? პასუხი დაასაბუთეთ.

ბ. გამოიყენეთ რეგრესიის განტოლება სიღარიბის ზღვარს ქვემოთ მოსახლეობის რაოდენობის განსაზღვრისათვის შემდეგი მონაცემების მიხედვით: შავკანიანი მოსახლეობა თითოეულ

შტატში 2002 წლისათვის 10.2%, უმაღლესი განათლების დიპლომის მქონეთა წილი 85.6%. არის თუ არა მიღებული შედეგი დაბალი ან მაღალი საშუალო მაჩვენებელზე 11.7%-ზე?

გ. რა უდრის დეტერმინაციის კოეფიციენტი? რამდენად გაუმჯობესდა ჩვენი პროგნოზი სიღარიბის ზღვარს ქვემოთ მცხოვრები მოსახლეობის რაოდენობის შესახებ მრავლობითი რეგრესიის განტოლების გამოყენებით?

13. ამ თავის კარგად შესწვლის შემდეგ პასუხი გაეცით შემდეგ კითხვებს:

ა. სწორია თუ მცდარი: შესაძლებელია და ეს ხშირად დაც ხდება, როცა b კოეფიციენტის მნიშვნელობა დადებითია, ხოლო კორელაციის კოეფიციენტი r უარყოფითია.

ბ. რომელი ცვლადის დამოკიდებულის თუ დამოუკიდებელის მნიშვნელობის ცვლილებას ასახავს a და b კოეფიციენტები?

გ. დეტერმინაციის კოფიციენტი r^2 არის შეცდომის პროპორციული შემცირების საზომი. განმარტეთ ეს დებულება.

დ. სწორია თუ მცდარი: ყველა რეგრესიის განტოლება ასახავს მიზეზობრივ დამოკიდებულებას წრფივი ფორმით.

თავი 14

დისპერსიული ანალიზი

თავის სწავლების მიზნები

- დისპერსიული ანალიზის (ANOVA) არსი და გამოყენება
- სტატისტიკის მნიშვნელობა და ინტერპრეტაცია
- r^2-ის მნიშვნელოვნების შეფასება დისპერსიული ანალიზის გამოყენებით
- დისპერსიული ანალიზის გამოყენება მრავლობით წრფივ რეგრესიაში

ბევრი საკვლევი საკითხის შესწავლა მოითხოვს სულ მცირე ორი შერჩევითი ერთობლიობის ანალიზს. შესასწავლ საკითხთა რიგში შესაძლოა იყოს ეთნიკური იდენტობის (თეთრკანიანი, აფრო-ამერიკელი, აზიელი წარმოშობის ამერიკელი, ლათინო ამერიკელი) გავლენა რელიგიურ მსახურებებზე დასწრების სიხშირეზე, სოციალური ფენების (დაბალი, მუშათა, საშუალო და მაღალი) გავლენა პრეზიდენტ ობამას დასაქმების პოლიტიკაზე ან განათლების მიღწევების (საშუალო განათლება, უმაღლესი განათლება, კოლეჯის განათლება) გავლენა საშუალო შემოსავლებზე. ყველა ნახსენები საკითხის შესწავლა საჭიროებს მონაცემთა შედარებებს სხვადასხვაგვარი დემოგრაფიული ფაქტორების მიხედვით, როგორც ეს მე-9 თავში განვიხილეთ. ორი სოციალური ფაქტორის ან ჯგუფის მიხედვით ანალიზი შედარებით მარტივია, მაგრამ ჩვენი სოციალური სამყარო გაცილებით კომპლექსური და მრავალფეროვანია.

დავუშვათ გვინდა შევისწავლოთ ამერიკაში განათლების საშუალო მაჩვენებლები-ამერიკელი საშუალოდ განათლებაზე რამდენ წელს ხარჯავს? 2007 წლის აშშ-ს აღწერის ბიუროს მონაცემებით 25 წელს ზემოთ მოსახლეობის 84.5%-ს ქონდა სულ მცირე საშუალო განათლება, ხოლო მათაგან 27%-ზე მეტს ქონდა ბაკალავრის ხარისხი (Crissey, 2009).[1] პრეზიდენტი ობამა თავის საპრეზიდენტო ვადის დასაწყისში ამტკიცებდა, რომ 2020 წლისათვის შეერთებულ შტატე-

ბი მსოფლიოში მოწინავე ქვეყანა იქნებოდა კოლეჯის კურსდამთავრებულთა ოდენობით. ამ პერიოდისათვის განსაკუთრებული ყურადღება დაეთმო ლათინო-ამერიკული წარმოშობის მოქალაქეების განათლებას. აშშ-ს აღწერის ბიუროსა და განათლების დეპარტამენტის მონაცემებით ლათინო-ამერიკული წარმოშობის სტუდენტებს სხვა რასობრივ თუ ეთნიკურ ჯგუფებთან შედარებით ქონდათ ყველაზე დაბალი განათლების დონე.

მე-9 თავში- ჰიპოთეზების შემოწმება, განვიხილეთ სტატისტიკის მეთოდი, რომლის მეშვეობითაც შესაძლებელია შევაფასოთ ორ შერჩევის საშუალოს შორის განსხვავება. ჩვენს მაგალითში (ცხრილი 9.3.) ერთმანეთს შევადარეთ განათლების საშუალო მაჩვენებლები შავკანიანი და ესპანური წარმოშობის ამერიკელი მოსახლეობისათვის. მაგრამ შესაძლებელია ასევე დავინტერესდეთ არის თუ არა განათლების დონეში მნიშვნელოვანი განსხვავებები ქალისა და მამაკაცის მიხედვით? (შავკანიანი მამაკაცი, ესპანური წარმოშობის ქალი, შავკანიანი ქალი, ესპანური წარმოშობის მამაკაცი).

ცხრილი 14.1 განათლება (წლების მიხედვით) ოთხი ჯგუფისათვის, GSS 2006

შავკანიანი მამაკაცები n1=6	ესპანური წარმოშობის მამაკაცები n2=4	შავკანიანი ქალები n3=6	ესპანური წარმოშობის ქალები n4=5
16	14	16	14
12	12	18	12
14	11	16	12
12	11	14	13
12		16	14
12		12	

საანალიზოდ ავიღოთ 21 ქალი და მამაკაცი საერთო სოციოლოგიური კვლევის მონაცემებიდან ოთხი დემოგრაფიული კატეგორიის მიხედვით. (ცხრილი 14.1) t სტატისტიკის გამოყენებით შესაძლებელია მხოლოდ ორი შერჩევითი ერთობლიობის მონაცემთა ანალიზი (ეს საკითხი მე-9 თავში განვიხილეთ). ჩვენს შემთხვევაში კი უნდა განვიხილოთ განათლების მონაცემები ერთის მხრივ შავკანიანი და ესპანური წარმოშობის ქალებისათვის, მეორეს მხრივ შაკკანიანი და ესპანური წარმოშობის მამაკაცებისათვის და ა.შ. (მთლიანობაში 6 განსხვავებული მონაცემთა წყვილის განხილვა მოგვიწევს). ბოლოსათვის ჩვენ გვექნება t სტატისტიკის გამოთვლების მნიშვნელობები, მაგრამ არ გვექნება პასუხი ჩვენს თავდაპირველ კითხვაზე: არის თუ არა მნიშვნელოვანი განსხვავებები განათლების დონის მიხედვით ოთხ დემოგრაფიულ ჯგუფში?

სტატისტიკაში არსებობს მეთოდი, რომლის საშუალებითაც შესაძლებელია ოთხივე ჯგუფის ან შერჩევის ერთობლივი ანალიზი. ეს მეთოდი იწოდება დისპერსიულ ანალიზად (ANOVA). დისპერსიული ანალიზი მოიცავს 5 ეტაპს, როგორც ეს იყო Z, t და ხი-კვადრატ კრიტერიუმების შემთხვევებში (თავი 9, 11). ამ თავში განვიხილავთ დისპერიული ანალიზის გამოთვლებს და ასევე იმას თუ

როგორაა შესაძლებელი დეტერმინაციის r^2 კოეფიციენტისა და დეტერმინაციის მრავლობითი კოეფიციენტის r^2-ის მნიშვნელოვნების შემოწმება დისპერსიული ანალიზის მეშვეობით. ასევე განვიხილავთ დისპერსიული ანალიზის გამოყენების ორ პრაქტიკულ მაგალითს კვლევით ლიტერატურაზე დაყრდნობით.

დისპერიული ანალიზი (ANOVA) დასკვნითი სტატისტიკის მეთოდი, რომელიც გამოიყენება ორ ცვლადს, ცვლადთა ჯგუფსა ან ორ შერჩევით მონაცემებს შორის კავშირის მნიშვნელოვნების შეფასებისათვის.

დისპერსიული ანალიზის შინაარსი

გავიხსენოთ, რომ t კრიტერიუმი ასახავს ორ საშუალო მნიშვნელობას შორის სხვაობას, იმ პირობებშიც როცა ნულოვანი ჰიპოთეზის მიხედვით დაშვებულ იყო მათ შორის სხვაობის არარსებობა: $\overline{Y}_1 - \overline{Y}_2$. ნულოვანი ჰიპოთეზის უარყოფა ნიშნავს, რომ ორ საშუალო მნიშვნელობას შორის იყო განსხვავება. მე-9 თავში ჩვენ განვიხილეთ განსხვავება განათლების საშუალო მაჩვენებლების მიხედვით შავკანიან და ესპანური წარმოშობის მამაკაცებს შორის. t კრიტერიუმის მიხედვით ჩვენ უარვყავით ნულოვანი ჰიპოთეზა, რაც ნიშნავს რომ შავკანიან მამაკაცებს საშუალოდ აქვთ უფრო მეტი განათლება წლების მიხედვით ვიდრე ესპანური წარმოშობის ამერიკელ მამაკაცებს.

დისპერსიული ანალიზის ANOVA ლოგიკა t ტესტის მსგავსია, მაგრამ გამოიყენება ორი ან მეტი შერჩევითი ერთობლიობისათვის. ცხრილ 14.1-ში მოცემული მონაცემებისათვის დისპერსიული ანალიზის გამოყენება შესაძლებელია ოთხი საშუალო მნიშვნელობისათვის $(\overline{Y}_1, \overline{Y}_2, \overline{Y}_3, \overline{Y}_4)$, ხოლო ნულოვანი ჰიპოთეზა შესაძლებელია ჩამოყალიბდეს შემდეგნაირად: $\mu_1 = \mu_2 = \mu_3 = \mu_4$. ნულოვანი ჰიპოთეზის უარყოფა დისპერსიული ანალიზისათვის ნიშნავს, რომ არსებობს მნიშვნელოვანი ვარიაცია ოთხ შერჩევით ერთობლიობას შორის (ან იმ ოთხი პოპულაციისათვის, საიდანაც მიღებულ იქნა შერჩევითი ერთობლიობა) და სულ მცირე ერთი შერჩევითი ერთობლიობის საშუალო მაჩვენებელი განსხვავდება მეორესაგან. ჩვენს მაგალითში ეს უკანასკნელი ნიშნავს, რომ განათლების წლები (დამოუკიდებელი ცვლადი) განსხვავდება ჯგუფების მიხედვით (დამოკიდებული ცვლადი). როდესაც დისპერსიული ანალიზი გამოიყენება ერთი დამოკიდებული და ერთი დამოუკიდებელი ცვლადისათვის მას ცალმხრივი დისპერსიული ანალიზი ეწოდება.

ცალმხრივი დისპერსიული ანალიზი დისპერსიული ანალიზის გამოყენება ერთი დამოკიდებული და ერთი დამოუკიდებელი ცვლადის მიმართ.

საშუალო მაჩვენებლები, სტანდარტული გადახრა და შერჩევითი ერთობლიობის ვარიაციის გამოთვლის შედეგები ნაჩვენებია ცხრილ 14.2-ზე. შევნიშნოთ, რომ განათლების წლებისათვის ოთხი საშუალო მაჩვენებელი ერთმანეთის ტოლის არაა, მათ შორის ყველაზე მაღალი საშუალო მაჩვენებელი შავკანიან ქალებს აქვთ. ასევე, სტანდარტული გადახრის მნიშვნელობების მიხედვით შეგვიძლია ვთქვათ, რომ შერჩევითი ერთობლიობები შედარებით ჰომოგენურია სტანდარტული გადახრის 1.00-დან 2.07-მდე მნიშვნელობებით. მაგრამ ჩვენ ამავე დროს ვიცით, რომ შერჩევით ერთობლიობებს შორის არსებობს განსხვავება, აქ რჩება შეკითხვა: არის თუ არა ეს განსხვავება მნიშვნელოვანი? ასახავს თუ არა შერჩევითი ერთობლიობის მიხედვით დაფიქსირებული კავშირი განათლებასა და დემოგრაფიულ ფაქტორებს შორის პოპულაციაში არსებულ ტენდენციას?

✓ *ცოდნის შემოწმება*

ჩვენ გამოვთვალეთ საშუალო და სტანდარტული გადახრა თითოეული შერჩევითი ერთობლიობისათვის ცხრილ 14.2-ში. გამოთვალეთ თითოეულ მონაცემთა ჯგუფისათვის საშუალო(თავი 4) და სტანდარტული გადახრა(თავი 5) და შეამოწმეთ გამოთვლების სისწორე.

ცხრილი 14.2 საშუალო, ვარიაცია და სტანდარტული გადახრა ოთხი ჯგუფისათვის, GSS 2006

შავკანიანი მამაკაცები $n_1=6$	ესპანური წარმოშობის მამაკაცები $n_2=4$	შავკანიანი ქალები $n_3=6$	ესპანური წარმოშობის ქალები $n_4=5$
16	14	16	14
12	12	18	12
14	11	16	12
12	11	14	13
12		16	14
12		12	
$\overline{Y}_1 = 13.00$	$\overline{Y}_2 = 12.00$	$\overline{Y}_3 = 15.33$	$\overline{Y}_4 = 13.00$
$S_1 = 1.67$	$S_2 = 1.41$	$S_3 = 2.07$	$S_4 = 1.00$
$S_1^2 = 2.79$	$S_2^2 = 1.99$	$S_3^2 = 4.28$	$S_4^2 = 1.00$
	$\overline{Y} = 13.48$		

განსხვავებების მნიშვნელოვნების შეფასებისათვის დისპერსიულ ანალიზში გამოიყენება როგორც შერჩევითი მაჩვენებლები, ასევე სხვადასხვა ჯგუფების მიხედვით გამოთვლილი მაჩვენებლები და მათ შორის განსხვავებები. ეს განსხვავებები ასევე შეიძლება იწოდოს როგორც გადახრები, ან ვარიაცია და დისპერსია, სწორედ ამიტომ იწოდება ეს მეთოდი ANOVA დისპერიულ ანალიზად. რა განსხვავებაა ერთ შერჩევითი ერთობლიობის საშუალოსა და პოპულაციის საშუალოს

შორის? რას უდრის დაკვირვების ცალკეული მონაცემების ვარიაცია ერთი შერჩევის ფარგლებში? დიდია თუ არა ვარიაციის ეს მნიშვნელობები? დისპერსიული ანალიზი საშუალებას გვაძლევს შევაფასოთ რამდენად დიდია განსხვავება შერჩევითი ერთობლიობის ვარიაციასა და ცალკეული მონაცემების ვარიაციას შორის. თუ განსხვავება შერჩევითი ერთობლიობის ვარიაციასა და ცალკეული მონაცემების ვარიაციას შორის დიდია ეს ნიშნავს, რომ განათლების დონე მნიშვნელოვნად განსხვავდება შერჩევის მიხედვით. ეს უკანასკნელი კი მეტყველებს იმაზე, რომ ცალკეულ ჯგუფზე მიკუთვნება ხსნის განათლების განსხვავებულ დონეებს.

ჰიპოთეზის შემოწმების სტრუქტურა დისპერსიული ანალიზის მიხედვით

დაშვებები

დისპერსიული ანალიზი ითვალისწინებს რამდენიმე დაშვებას შერჩევის მეთოდთან, გაზომვის სკალასა და პოპულაციის განაწილების ფორმასა და ვარიაციის ჰომოგენურობასთან მიმართებაში:

1. საანალიზოდ დამოუკიდებელი შემთხვევითი შერჩევა გამოიყენება. ერთი შემთხვევითი შერჩევის გაკეთებას პოპულაციიდან არ აქვს გავლენა მოერე, მესამე და ა.შ. შერჩევაზე. მაგალითად, ესპანური წარმოშობის მამაკაცის შერჩევას არ აქვს გავლენა სხვა შერჩევებზე.
2. დამოკიდებული ცვლადი-განათლების წლები არის ინტერვალურ-ფარდობითი ცვლადი. ცალკეულ შემთხვევებში ANOVA შესაძლებელია გამოყენებულ იქნას ორდინალური ცვლადებისთვისაც.
3. პოპულაცია ნორმალური განაწილებისაა. მიუხედავად იმისა, რომ არ შეგვიძლია დავამტკიცოთ პოპულაციის ნორმალური განაწილება, შერჩევის ზომის სიმცირიდან გამომდინარე (N) შეგვიძლია მივიჩნიოთ, რომ ჩვენი ანალიზის პირობებში პოპულაციის მონაცემები ნორმალურადაა განაწილებული.
4. პოპულაციის ვარიაციები ტოლია. ჩვენი გამოთვლებით ცხრილ 14.2-ში, მიუხედევად იმისა, რომ შერჩევის ვარიაციები ერთმანეთის ტოლი არაა, ისინი საკმაოდ ჰომოგენურია.

კვლევითი და ნულოვანი ჰიპოთეზის ჩამოყალიბება, ალფას მნიშვნელობის არჩევა

კვლევის ჰიპოთეზის მიხედვით (H_1) სულ მცირე ერთი საშუალო მაჩვენებელი მაინც არის განსხვავებული. ჩვენ არ ვაკონკრეტებთ თუ რომელი საშუალო იქნება განსხვავებული, მაგრამ ეს დაშვება სულ მცირე ერთი ნებისმიერი საშუალოს განსხვავების არსებობას გულისხმობს.

H_1: სულ მცირე ერთი საშუალო მაინც არის განსხვავებული.

დისპერსიული ანალიზი არის ნულოვანი ჰიპოთეზის ტესტი. რადგან ჩვენ საქმე გვაქვს ოთხ შერჩევით ერთობლიობასთან, უნდა ჩავრთოთ ოთხი ჩვენს შერჩევითი ერთობლიობის ანალიზში.

$H_0: = \mu_1 = \mu_2 = \mu_3 = \mu_4$

როგორც სხვა ჰიპოთეზის შემოწმებისას, ამჯერადაც უნდა ავირჩიოთ ალფას მნიშვნელობა. ალფა ეს არის ალბათობის ის დონე, რომლისთვისაც ხდება ნულოვანი ჰიპოთეზის უარყოფა. ამ შემთხვევაში ავირჩიოთ ალფას 0.05 მნიშვნელობა.

შიგაჯგუფური და ჯგუფთაშორისი დისპერსიის კონცეფცია

სანამ უშუალოდ ანალიზზე გადავიდოდეთ, გავაკეთოთ ერთი მნიშვნელოვანი შენიშვნა: რადგან ვმუშაობთ ოთხ განსხვავებულ შერჩევით ერთობლიობაზე 21 რესპონდენტით, მოგვიწევს ბევრი გამოთვლების ჩატარება. იყავით თანმიმდევრულნი და სწორხაზოვნად მიყევით ყველა გამოთვლას.

ჩვენი თავდაპირველი გამოთვლები უკავშირდება ორი ტიპის დისპერსიას: შიგაჯგუფურ და ჯგუფთაშორის. თითოეული ტიპის დისპერსიის გამოთვლას აქვს ორი ნაწილი-კვადრატების ჯამი და თავისუფლების ხარისხი (*df*).

ჯგუფთაშორისი დისპერსია ან შემოკლებული აღნიშვნით SSB ასახავს განსხვავებას განათლების საშუალო მაჩვენებლის მიხედვით ოთხ ჯგუფს შორის. კვადრატების ჯამის გაფართოებული შინაარსია-გადახრების კვადრატების ჯამი. SSB ზომავს გადახრების კვადრატების ჯამს თითოეულ მონაცემის საშუალოსა და იმავე მონაცემთა ჯგუფის საშუალო მაჩვენებელს შორის. SSB-ის გამოთვლის ფორმულა შესაძლებლია შემდეგნაირად ჩაიწეროს:

$$SSB = \sum n_k (\overline{Y}_k - \overline{Y})^2 \qquad (14.1)$$

სადაც,

n_r = შერჩევის დაკვირვებათა რიცხვი (კ-შერჩევის რაოდენობა)

$\overline{Y}_k$ = შერჩევის მონაცემთა საშუალო მნიშვნელობა

$\overline{Y}$ = საერთო საშუალოს მნიშვნელობა

SSB ასევე შესაძლებელია განიმარტოს როგორც დამოკიდებული ცვლადის დისპერსია (განათლების წლები), რაც შესაძლებელია აიხსნას დამოუკიდებელი ცვლადით (ოთხი დემოგრაფიული ჯგუფი).

ჯგუფთაშორისი კვადრატების ჯამი, SSB დისპერსიის კვადრატების ჯამი შერჩევის თითოეული საშუალოსა და საერთო საშუალოს მიმართ.

შიგაჯგუფური კვადრატების ჯამი, SSW ასახავს ცვლადის მნიშვნელობათა ვარიაციას ერთი შერჩევის შიგნით, ჩვენი მაგალითის შემთხვევაში განათლების წლების მიხედვით ერთ შერჩევაში. SSW ასევე მოიხსენიება როგორც აუხსნელი დისპერსია, რადგან ეს სწორედ დისპერსიის ის ნაწილია, რაც რჩება დამოუკიდებელი ცვლადის გავლენის შესწავლის შემდეგ. ფორუმულის მიხედვით SSW ასახავს დისპერსიის კვადრატების ჯამს თითოეულ ჯგუფში, ცალეკული მონაცემისა და ჯგუფის საშუალოს მიხედვით:

$$SSW = \sum (Y_i - \overline{Y}_k)^2 \qquad (14.2)$$

სადაც,

Y_i = დაკვირვების ერთეულთა ცალკეული მნიშვნელობები

$\overline{Y}_k$ = შერჩევის საშუალო

მიუხედავად ჩვენი შერჩევის შედარებით მცირე ზომისა, 14.2 ფორმულის გამოყენების შემთხვევაში მოგვიწევდა საკმაოდ ბევრი გამოთვლების ჩატარება, ნაცვლად ამისა მოსახერხებელია შიგაჯგუფური დისპერსიის გამოთვლისათვის გამოვიყენოთ შემდეგი ფორმულა:

$$SSW = \sum Y_i^2 - \sum \frac{(\sum Y_k)^2}{n_k} \qquad (14.3)$$

სადაც,

Y_i^2 = თითოეული შერჩევითი ერთობლიობის დაკვირვების მნიშვნელობათა კვადრატებია

$\sum Y_k$ = შერჩევითი დაკვირების მნიშველობათა ჯამია

n_k = დაკვირვებათა რიცხვია თითოეული შერჩევითი ერთობლიობისათვის

შიგაჯგუფური კვადრატების ჯამი, SSW თითოეული ჯგუფის დისპერსიის კვადრატების ჯამი, ჯგუფისა და დაკვირვების მონაცემების საშუალოების მიხედვით.

ჯგუფშორისი (*SSB*) და შიგაჯგუფური დისპერსიების (*SSW*) დისპერსიების დაჯამებით მივიღებთ ჯამობრივ დისპერსიას. დისპერსიების მთლიანი ჯამი შშთ შესაძლებელია ჩაიწეროს შემდეგი ფორმით:

$$SST = \sum (Y_i - \overline{Y})^2 = SSB + SSW \qquad (14.4)$$

სადაც,

Y_i = დაკვირვების მნიშვნელობები

$\overline{Y}$ = საშუალოს მნიშვნელობა

კვადრატების მთლიანი ჯამი მთლიანი დისპერსია, რომელიც მიიღება *SSB*-ისა და *SSW*-ს დაჯამებით.

ჯგუფშორისი და შიგაჯგუფური დისპერსიების გამოთვლის მოერე ეტაპია თავისუფლების ხარისხის გამოთვლა. ეს საკითხი ასევე განხილულია მე-9 და მე-11 თავებში. დისპერსიული ანალიზისათვის დავჭირდება თავისუფლების ორი ხარისხის გამოთვლა. *SSB*-სათვის თავისუფლების ხარისხი განისაზღვრება:

$$df_b = k - 1 \tag{14.5}$$

სადაც k შერჩევის რიცხვია.

SSW-სათვის თავისუფლების ხარისხია:

$$df_w = N - k \tag{14.6}$$

სადაც

N = დაკვირვების მონაცემთა რაოდენობაა

k = შერჩევის რიცხვი

საბოლოოდ, ჩვენ შეგვიძლია გამოვთვალოთ ჯგუფშორის დისპერსია კვადრატების საშუალოს გამოთვლით. მარტივად, რომ ვთქვათ კვადრატების საშუალო გამოითვლება თითოეული კვადრატების ჯამის შეფარდებით თავისუფლების შესაბამის ხარისხთან.

$$\textit{მონაცემთა შორის კვადრატების საშუალო} = SSB / df_b \tag{14.7}$$

$$\textit{მონაცემებშიგა კვადრატების საშუალო} = SSW / df_w \tag{14.8}$$

მონაცემთა შორის კვადრატების საშუალო კვადრატების ჯამის შეფარდება შესაბამის თავისუფლების ხარისხთან.

მონაცემშიშიგა კვადრატების საშუალო კვადრატების ჯამის შეფარდება შესაბამის საშუალოსთან.

დეტალების განხილვა 14.1
SST-ს დეკომპოზიცია

14.4 ფორმულის მიხედვით, კვადრატების ჯამი (*SST*) უდრის:

$$SST = \sum(Y_i - \overline{Y})^2 = SSB + SSW$$

როგორც ფორმულის მიხედვით ვხედავთ მონაცემთა შორის კვადრატების საშუალო (ახსნადი დისპერსია) და მონაცემებშიგა კვადრატების საშუალო (აუხსნელი დისპერსია) განსაზღვრავს მთლიან ვარიაციას კონკრეტული დამოკიდებული ცვლადისათვის. როგორ ესადაგება აღნიშნული ჩვენს მაგალითს? ავიღოთ ცხრილ 14.1-დან შავკანიანი ქალი 16-წლის განათლებით.

ამ ცვლადის მთლიანი დისპერსია გამოითვლება მისი მნიშვნელობის სხვაობით საშუალო მნიშვნელობასთან (ფორმულა 14.4). ამ კონკრეტულ შემთხვევაში ცვლადის მნიშვნელობა მეტია განათლების საშუალო მნიშვნელობასთან შედარებით 13.84. სხვაობა 2.52-ია (16-13.48). ჯგუფშორისი დისპერსია შესაძლებელია გამოითვალოს ჯგუფის საშუალოსა და საერთო საშუალოს შორის სხვაობის პოვნით. (ფორმულა 14.4) ჩვენ უკვე ვიპოვეთ შავკანიანი ქალების განათლების დონის საშუალო მაჩვენებელი 15.33. ამ ჯგუფისათვის კი სხვაობა ჯგუფის საშუალო მნიშვნელობასთან იქნება 1.85 წელი (15.33-13.48). ბოლოს, შიგაჯგუფური კვადრატების ჯამი (ფორმულა 14.2) გამოვთვალოთ პირველი შავკანიანი ქალის განათლების წლების სხვაობით საშუალო მაჩვენებლთან: 0.67 (16-15.33). ასე,რომ ჩვენს მაგალითში პირველი შავკანიანი ქალისათვის მთლიანი დისპერსია უდრის $SSB + SSW = SST\ 1.85 + 0.67 = 2.52$. სრულ დისპერსიულ ანალიზში ვითვლით დისპერსიის ორივე მაჩვენებელს ჯამობრივი დისპერსიის მიღებისათვის (ფორმულა 14.4) შერჩევის თითოეული მნიშვნელობისათვის.

F სტატისტიკა

მონაცემთა შორის კვადრატების საშუალოსა (ფორმულა 14.7) და მონაცემებშიგა კვადრატების საშუალოს (ფორმულა 14.8) კომბინირებით მიიღება *F* ფარდობა და *F* სტატისტიკა. რ. ა. ფიშერის მიერ შემუშავებული სტატისტიკა არის ფარდობა მონაცემთა შორის კვადრატების საშუალოსა და მონაცემებშიგა კვადრატების საშუალოსი:

$$F = \frac{\text{მონაცემთა შორის კვადრატების საშუალო}}{\text{მონაცემებშიგა კვადრატების საშუალო}} = \frac{SSB / df_b}{SSW / df_w} \qquad (14.9)$$

F სტატისტიკის მაღალი რიცხვითი მნიშვნელობა მეტყველებს, რომ ჯგუფს შორის კვადრატების საშუალო აღემატება შიგაჯგუფურ კვადრატების საშუალოს, რაც თავის მხრივ ზრდის ნულოვანი ჰიპოთეზის უარყოფის ალბათობას. ცხრილ

14.3-ში წარმოდგენილია F სტატისტიკის გამოთვლის ეტაპები.

ჯგუფთაშორისი კვადრატების საშუალო და თავისუფლების ხარისხი გამოვთვალოთ 14.1 და 14.5 ფორმულების გამოყენებით. ჯგუფსშორის კვადრატების ჯამი უდრის:

$$\sum nk\,(\bar{Y}k - \bar{Y})2 = 6(13.00 - 13.48)^2 + 4(12.00 - 13.48)^2 + 6(15.33 - 13.48)^2 + 5(13.00 - 13.48)^2 = 31.83$$

ცხრილი 14.3 ANOVA-ს გამოთვლა

შვკანიანი მამაკაცები $n_1=6$	ესპანური წარმოშობის მამაკაცები $n_2=4$	შვკანიანი ქალები $n_3=6$	ესპანური წარმოშობის ქალები $n_4=5$
16	14	16	14
12	12	18	12
14	11	16	12
12	11	14	13
12		16	14
12		12	
$\bar{Y}_1 = 13.00$	$\bar{Y}_2 = 12.00$	$\bar{Y}_3 = 15.33$	$\bar{Y}_4 = 13.00$
$S_1 = 1.67$	$S_2 = 1.41$	$S_3 = 2.07$	$S_4 = 1.00$
$S_1^2 = 2.79$	$S_2^2 = 1.99$	$S_3^2 = 4.28$	$S_4^2 = 1.00$
$\sum Y_1 = 78$	$\sum Y_2 = 48$	$\sum Y_3 = 92$	$\sum Y_4 = 65$
$\sum Y_1^2 = 1028$	$\sum Y_2^2 = 582$	$\sum Y_3^2 = 1432$	$\sum Y_4^2 = 849$
$\bar{Y} = 13.48$			

თავისუფლების ხარისხი SSB-ისთვის არის k_1 ან $4 - 1 = 3$. ფორმულა 14.7-ის მიხედვით მონაცემთა შორის კვადრატების საშუალო იქნება:

$$\text{მონაცემთა შორის კვადრატების საშუალო} = \frac{31.83}{3} = 10.61$$

შიგაჯგუფური კვადრატების საშუალო და თავისუფლების 14.3 და 14.6 ფორმულების მიხედვით იქნება:

$$SSW = \sum Y_1^2 - \sum \frac{\left(\sum Y_k\right)^2}{n_k} = (1,028 + 582 + 1,432 + 849) - \left(\frac{78^2}{6} + \frac{48^2}{4} + \frac{96^2}{6} + \frac{65^2}{5}\right)$$

$$= 3,891 - 3,845.67$$
$$= 45.33$$

თავისუფლების ხარისხი *SSW*-ისათვის უდრის $N - k = 21 - 4 = 17$. 14.8 ფორმულის მიხედვით მონაცემებსშიგა კვადრატების საშუალო იქნება:

$$\text{მონაცემთაშიგა კვადრატების საშუალო} = \frac{45.33}{17} = 2.67$$

საბოლოოდ, 14.9 ფორმულის მიხედვით გამოვთვალოთ F სტატისტიკა:

$$F = \frac{10.61}{2.67} = 3.97$$

F სტატისტიკა ან F ფარდობა ტესტ სტატისტიკა დისპერსიული ანალიზისათვის, რომელიც გამოითვლება მონაცემთა შორის კვადრატების საშუალოსა და მონაცემებსშიგა კვადრატების საშუალოს ფარდობით.

გადაწყვეტილების მიღება

F სტატისტიკის 3.97 მნიშვნელობის ალბათობის პოვნისათვის მივმართოთ დამატებას – F განაწილების ცხრილი. წიგნის ბოლოში მოცემულ E დამატებაში არის F განაწილების მნიშვნელობები თავისუფლების სხვადასხვა ხარისხებისათვის და მნიშვნელოვნების ორი 0.05 და 0.01 დონეებისათვის. ხოლო ცხრილ 14.4-ში მოცემულია განაწილება 0.05 დონისათვის.

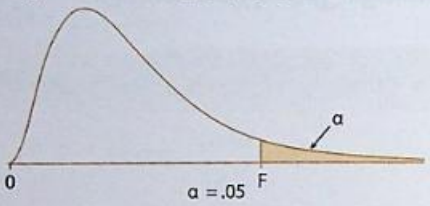
რადგან ალფას მნიშვნელობა არჩეულია 0.05-ის ტოლად, ცხრილში ვეძებთ „$P = 0.05$“ ტოლ შესაბამის მნიშვნელობას. შევნიშნოთ, წიგნის დამატებაში მოცემული F განაწილება თავისუფლების მხოლოდ ორი ხარისხისათვის არის მოცემული, ეს ჩვენი მაგალითის თავისუფლების ხარისხებსაც შეეხება: $df_1 = df_b$ და $df_2 = df_w$.

რადგან ჩვენი მაგალითის პირობებში გვაქვს თავისუფლების ორი ხარისხი F განაწილებისათვის ალბათობის განსაზღვრა t ტესტისა და ხი-კვადრატ კრიტერიუმისაგან განსხვავებულად მოგვიწევს. დისპერსიული ანალიზის ამ მაგალითისათვის, ჯერ უნდა განვსაზღვროთ შესაბამისი F, რომელიც ასევე F-ის კრიტიკულ მნიშვნელობად იწოდება, $df_b = 3$ და $df_w = 17$, $\alpha = .05$.

E დამატების მიხედვით F-ის კრიტიკული მნიშვნელობა 3.20-ის ტოლია, ხოლო F-ის გამოთვლილი მნიშვნელობა არის 3.97. ჩვენი გამოთვლილი F-ის მნიშვნელობა მეტია მის კრიტიკულ მნიშვნელობაზე F (3.97>3.20), ხოლო მისი ალბათობა <0.05-ზე, რაც ნახაზზე ნაჩვენებია გამუქებული ფართობით. (თუ ჩვენს

მიერ მიღებული F იქნებოდა <3.20-ზე, მაშინ მივიჩნევდით რომ ალბათობა მეტი იქნებოდა ვიდრე [α]-ს მნიშვნელობა, ანუ ნახაზის გაუმუქებელი ნაწილი, ნახ. 14.4). აქედან გამომდინარე შეგვიძლია უარვყოთ ნულოვანი ჰიპოთეზა განსხვავებების არარსებობის შესახებ და ვასკვნით, რომ ჯგუფების მიხედვით არსებობს მნიშვნელოვანი განსხვავება განათლების დონეში.

ცხრილი 14.4 განაწილება 0.05 მნიშვნელოვნების დონისათვის

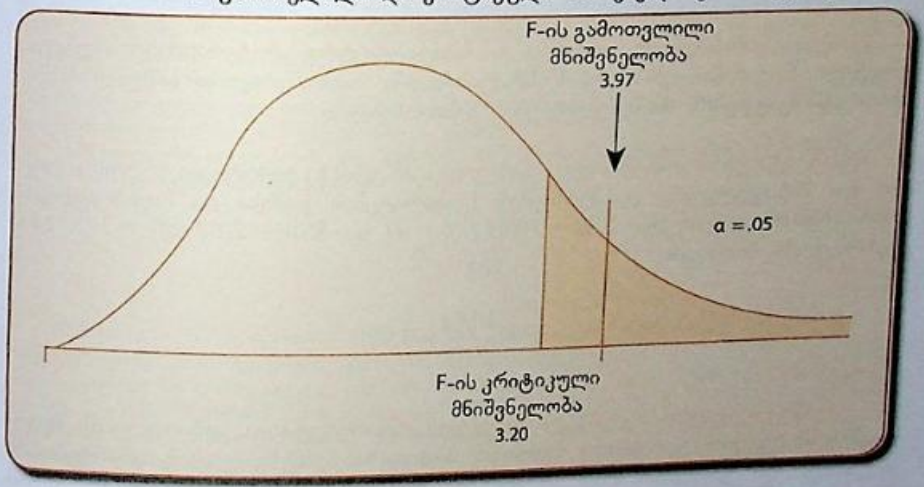


	df_1									
df_2	1	2	3	4	5	6	8	12	24	∞
1	161.40	199.50	215.70	224.60	230.20	234.00	238.90	243.90	249.00	254.30
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73

25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

წყარო: R.A. Fisher and F. Yates, Statistical Tables for Biological, Agricultural and Medical Research, 6th ed. CopyRight (C)R.A. Fisher and F. Yates 1963. Pearson Education Limited.

ნახაზი 14.1 F-ის გამოთვლილი და კრიტიკული მნიშვნელობების შედარება



შეჯამება: ჰიპოთეზის შემოწმების ცუთი ნაბიჯი

ჩვენ განვიხილეთ დისპერსიული ანალიზის მაგალითი განათლების საშუალო წლების მიმართ ოთხი სხვადასხვა დემოგრაფიული ჯგუფის მიმართ. შევაჯამოთ ყველა ეტაპი, რომელიც ამ ანალიზისას გამოვიყენეთ:

დაშვებების გაკეთება:

1. ანალიზი გამოიყენება დამოუკიდებელი შემთხვევითი შერჩევების მიმართ.
2. დამოკიდებული ცვლადი, განათლების წლები არის ინტერვალურ-ფარდობითი ტიპის ცვლადი.
3. პოპულაციები ნორმალურადაა განაწილებული.
4. პოპულაციის ვარიაციები იდენტურია.

კვლევითი და ნულოვანი ჰიპოთეზების ჩამოყალიბება, ალფას მნიშვნელობის არჩევა:

H_1: სულ მცირე ერთი საშუალო მნიშვნელობა მაინც განსხვავდება სხვებისაგან

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

$$\alpha = .05$$

შერჩევის განაწილების ფორმისა და სტატისტიკური კრიტერიუმის დაზუსტება. ვიყენებთ F განაწილებასა და F სტატისტიკას, ოთხ შერჩევით საშუალოს შორის განსხვავებების მნიშვნელოვნების შესაფასებლად.

სტატისტიკური კრიტერიუმის გამოთვლა. ამ ეტაპზე გვჭირდება ჯგუფთაშორისი და შიგაჯგუფური დისპერსიების (კვადრატების ჯამისა და თავისუფლების ხარისხების) გამოთვლა. $SSB = 10.61$ ($df_b = 3$) და $SSW = 2.67$ ($df_w = 17$). 14.9 ფორმულის მიხედვით:

$$F = \frac{10.61}{2.67} = 3.97$$

გადაწყვეტილების მიღება და შედეგების ინტერპრეტაცია. გამოთვლების შედეგების მიხედვით უარვყოფთ ნულოვან ჰიპოთეზას განსხვავებების არარსებობის შესახებ და ვასკვნით, რომ სხვადასხვა დემოგრაფიული ჯგუფების მიხედვით არსებობს განსხვავება განათლების დონეში. F-ის მიღებული მნიშვნელობაა 3.97, რაც მეტია F-ის კრიტიკულ მნიშვნელობაზე 3.20-ზე. 3.97-ის ტოლი მნიშვნელობის მიღების ალბათობა <0.05-ზე. F სტატისტიკა არ გვაძლევს ინფორმაციას თუ რომელი ჯგუფია განსხვავებული, მხოლოდ განათლების დონეა მნიშვნელოვნად განხვავებული დემოგრაფიული ჯგუფების მიხედვით. ამასთან, შერჩევის მონაცემების მიხედვით ვიცით, რომ განათლების საშუალო მაჩვენებელი 15.33 წელი აქვს მხოლოდ შავკანიან ქალებს. თუ დავაჯგუფებთ სხვა დემოგრაფიულ ჯგუფებსაც განათლების საშუალო მაჩვენებლების მიხედვით მეორე ადგილს დაიკავებს შავკანიანი მამაკაცები (13.00), შემდეგი ჯგუფი იქნება ესპანური წარმოშობის

ქალები (13.00) და ბოლოს ადგილზე იქნება ესპანური წარმოშობის მამაკაცები (12.00). ყველა ჯგუფის განათლების საშუალო მაჩვენებელი სულ მცირე 2 წლით მცირეა შავკანიანი ქალების განათლების საშუალო მაჩვენებელთან შედარებით. განათლების დონე მნიშვნელოვნად განსხვავდება რასისა და სქესის მიხედვით.

F-ის კრიტიკული მნიშვნელობა სტატისტიკური კრიტერიუმი, რომელიც შეესაბამება ალფას მნიშვნელობის დონეს, df_w, და df_b (წიგნის E დამატება).

F-ის მიღებული მნიშვნელობა გამოთვლების შედეგად მიღებული სტატისტიკური კრიტერიუმის მნიშვნელობა.

დეტალების განხილვა 14.2
ცვლადებს შორის კავშირის შეფასება

ჩვენს მაგალითში, F განაწილებისა და 5 ეტაპიანი გამოთვლების გამოყენებით, დავადგინეთ რომ არსებობდა მნიშვნელოვანი განსხვავება დემოგრაფიულ ჯგუფებში განათლების დონის მიხედვით. შესაბამისად, უარვყავით ნულოვანი ჰიპოთეზა და დავასკვენით, რომ განათლების წლების (დამოკიდებული ცვლადი) მიხედვით არსებობს სხვაობა დემოგრაფიულ ჯგუფებში (დამოუკიდებელი ცვლადი). მაგრამ შესაძლებელია თუ არა განვსაზღვროთ აღნიშნულზე დაყრდნობით, თუ როგორი სიძლიერისაა ეს კავშირი?

კორელაციური ფარდობა ან ეტა კვადრატი (η^2) ასახავს ცვლადთა შორის კავშირის სიძლიერეს. ეტა კვადრატი შემდეგი ფორმულით გამოითვლება:

$$\eta^2 = \frac{SSB}{SST}$$

SSB-სა და *SST*-ს ფარდობა (*SSB* + *SSW*) წარმოადგენს დისპერსიის იმ წილს, რომელიც აიხსნება დამოუკიდებელი ცვლადით (დემოგრაფიული ჯგუფი). ეტა კვადრატი ასახავს დამოკიდებულ და დამოუკიდებელ ცვლადს შორის კავშირს და იღებს მნიშვნელობებს 0-დან 1-მდე. ეტას მნიშვნელობის 0-თან მიახლოება ნიშნავს შედარებით სუსტ კავშირს, ხოლო მნიშვნელობის 1-თან მიახლოება ასახავს შედარებით ძლიერ კავშირს.

ჩვენი დისპერსიული ანალიზის მიხედვით:

$$\eta^2 = \frac{10.61}{10.61 + 2.67} = \frac{10.61}{13.28} = 0.80$$

მიღებულ შედეგზე დაყრდნობით, განათლების ვარიაციის 80% აიხსნება დემოგრაფიული ფაქტორებით ან სხვაგვარად დამოკიდებული ცვლადის ვარიაციის 80% აიხსნება დამოუკიდებელი ცვლადით. რამდენად ძლიერია ეს კავშირი? ამ საკითხის გადაწყვეტისათვის შეგვიძლია მივმართოთ მე-12 თავს (12.2), გამას მნიშვნელოვნების შეფასება. 12.1 ქვეპარაგრაფში მოცემული მიდგომის მიხედვით შეგვიძლია დავასკვნათ, რომ არსებობს საკმაოდ ძლიერი კავშირი განათლების დონესა და დემოგრაფიულ ჯგუფს მიკუთვნებას შორის.

R^2 მნიშვნელოვნების შეფასება დისპერსიული ანალიზის გამოყენებით

დისპერსიული ანალიზი შესაძლებელია გამოყენებულ იქნას რეგრესიული მოდელის მნიშვნელოვნების, რომელიც გამოხატულია r^2-ით, შესაფასებლად (იხილე მე-13 თავის ქვეთავი „დეტერმინაციის კოეფიციენტი, როგორც შეცდომის პროპორციული შემცირების საზომი“). ფაქტიურად, თუ დავაკვირდებით დისპერსიული და რეგრესიული ანალიზი საკმაოდ გავს ერთმანეთს. ორივე მეთოდში, დამოკიდებული ცვლადის ვარიაციის ახსნას დამოუკიდებელი ცვლადის მიხედვით ვახდენთ, ასევე დისპერსიულ ანალიზში დამოუკიდებელი ცვლადი კატეგორიული ტიპის ცვლადია (ნომინალური ან ორდინალური, მაგ, სქესი ან სოციალური კლასი), ხოლო რეგრესიულ ანალიზში ინტერვალურ-ფარდობითია (მაგ. შემოსავალი დოლარებში).

მე-13 თავში განვიხილეთ რეგრესის მოდელი, რომელშიც განვსაზღვრეთ „ბაკალავრის განათლების მქონეთა რაოდენობა“ დამოუკიდებელი ცვლადის – „შტატის საოჯახო მეურნეობების საშუალო შემოსავლების“ მიხედვით. როგორც იქვე აღვნიშნეთ დეტერმინაციის r^2 კოეფიციენტი ასახავს დამოკიდებული ცვლადის მთლიან ვარიაციას, რომელიც შესაძლებელია აიხსნას დამოუკიდებელი ცვლადით. r^2 კოეფიციენტის 0.79 მნიშვნელობა ნიშნავს, რომ დამოუკიდებელი ცვლადი (საშუალო შემოსავლები) განაპირობებს დამოკიდებული ცვლადის (ბაკალავარის ხარისხის მქონეთა რიცხვი) ვარიაციის 79%-ს.

დეტალების განხილვა 14.3

სტუდენტის საშუალო აკადემიური მოსწრება მოსწრება (GPA) და რასობრივი/ეთნიკური კუთვნილება

კვლევის დროს (Monitoring the Future 2006) შერჩეულ იქნა 17 მამაკაცი სტუდენტი, რათა შეესწავლათ კავშირი GPA-სა და რასობრივ კუთვნილებას შორის.

თეთრკანიანი მამაკაცები ($n_1 = 6$)	შავკანიანი მამაკაცები ($n_2 = 6$)	ესპანური წარმოშობის მამაკაცები ($n_3 = 6$)
3.4	3.2	3.6
3.5	2.7	2.1
3.4	2.8	1.9
3.4	2.1	2.0
3.0	2.9	2.0
3.5		2.2
$\bar{Y}_1 = 3.37$	$\bar{Y}_2 = 2.74$	$\bar{Y}_3 = 2.30$
$S_1 = 0.19$	$S_2 = 0.40$	$S_3 = 0.64$
$S_1^2 = 0.04$	$S_2^2 = 0.16$	$S_3^2 = 0.41$

$\sum Y_1 = 20.2$ $\quad \sum Y_2 = 13.7$ $\quad \sum Y_3 = 13.8$

$\sum Y_1^2 = 68.18$ $\quad \sum Y_2^2 = 38.19$ $\quad \sum Y_3^2 = 33.82$

$\bar{Y}_1 = 2.81$

ჩვენი ნულოვანი ჰიპოთეზაა, რომ GPA-ის მაჩვენებლები ტოლია. თუ არ მოხდება ნულოვანი ჰიპოთეზის უარყოფა, ეს ნიშნავს რომ რასობრივი კუთვნილების მიხედვით GPA-ის მაჩვენებლებში განსხვავება არ არის. გამოვთვალოთ *SSB*-ისა და *SSW*-ს მაჩვენებლები:

$$SSB = 6\,(3.37 - 2.81)^2 + 5\,(2.74 - 2.81)^2 + 6\,(2.30 - 2.81)^2$$
$$= 1.88 + 0.02 + 1.56$$
$$= 3.46$$

მონაცემთა შორის კვადრატების საშუალო უდრის: $3.46/2 = 1.73$

$$SSW = (68.18 + 38.19 + 33.82) - ()$$
$$= 140.19 - (68.01 + 37.54 + 31.74)$$
$$= 2.90$$

მონაცემებსშიგა კვადრატების საშუალო უდრის $2.90/14 = 0.21$

$$F = 1.73/0.21 = 8.24$$

გადაწყვეტილება: თუ ალფას 0.05-ის მნიშვნელობით ავირჩევთ F კრიტიკული იქნება 3.74. მიღებული F-ის მნიშვნელობის მიხედვით (8.24) უარვყოთ ნულოვანი ჰიპოთეზა და დასავკვნათ, რომ სულ მცირე GPA-ის ერთი მაჩვენებელი მაინც განსხვავდება მნიშვნელოვნად დანარჩენებისაგან. თეთრკანიან სტუდენტებს აქვს ყველაზე მაღალი GPA-ის მაჩვენებელი (3.37) შავკანიან და ესპანური წარმოშობის სტუდენტებთან შედარებით.

სხვა სტატისტიკური კრიტერიუმების მსგავსად r^2-იც ეფუძნება შერჩევის მონაცემებს. r^2-ის გამოთვლის შემდეგ უნდა შევაფასოთ იმის ალბათობა, რომ წრფივი კავშირი საშუალო შემოსავლებსა და ბაკალვრის განათლების ქონას შორის, რომელიც გამოხატულია r^2-ით ნოლის ტოლი იქნება პოპულაციაში. სხვა სიტყვებით, რომ ვთქვათ ჩვენ უნდა შევამოწმოთ r^2-ის სტატისტიკური მნიშვნელოვნება.

დისპერსიული ანალიზისას ჩვენ დამოკიდებული ცვლადის მთლიანი დისპერსიის დეკომპოზიცია მოვახდინეთ დამოუკიდებელი ცვლადის მიხედვით ახსნილ (*SSB*) და აუხსნელ (*SSW*) ნაწილებად. შემდეგ გამოვთვალეთ მონაცემთაშორის კვადრატების საშუალო (SSB/df_b) და მონაცემსშიგა კვადრატების საშუალო (SSW/df_w). სტატისტიკური F კრიტერიუმი არის მონაცემთაშორის კვადრატების საშუალოსა და მოანცემსშიგა კვადრატების საშუალოს ფარდობა 14.9 ფორმულის მიხედვით:

$$F = \frac{\text{მონაცემთა შორის კვადრატების საშუალო}}{\text{მონაცემებშიგა კვადრატების საშუალო}} = \frac{SSB / df_b}{SSB / df_w}$$

რეგრესიული ანალიზისას დამოკიდებული ცვლადის მთლიანი დისპერსიის დეკომპოზიცია მოვახდინეთ ახსნილ (*SSR*, საშუალო კვადარტული რეგრესია) და აუხსნელ (*SSE*, საშუალო კვადრატული ნაშთი) ნაწილებად. დისპერსიული ანალიზის მსგავსად, საშუალო კვადრატული ნაშთი და საშუალო კვადარტული რეგრესია გამოითვლება კვადრატების ჯამის შეფარდებით შესაბამის თავისუფლების ხარისხთან (*df*). *SSR*-თან დაკავშირებული თავისუფლების ხარისხი უდრის *K*-ს, რაც ასახავს დამოუკიდებელი ცვლადების რიცხვს რეგრესიის განტოლებაში.

$$\text{საშუალო კვადრატული რეგრესია} = \frac{SSR}{df_r} = \frac{SSR}{K} \qquad (14.11)$$

$$\text{საშუალო კვადრატული ნაშთი} = \frac{SSE}{df_e} = \frac{SSE}{[N-(K+1)]} \qquad (14.12)$$

რეგრესიის კვადრატების ჯამი (SSR) ასახავს პროგნოზის შეცდომის შემცირების ზომას წრფივი პროგნოზირების განტოლების გამოყენების შედეგად.

ნაშთის კვადრატების ჯამი (SSE) დაკვირვებისა და განტოლების საფუძველზე მიღებული დამოკიდებული ცვლადის (Y) მნიშვნელობებს შორის სხვაობის კვადრატების ჯამი.

ცხრილ 14.5-ში ნაჩვენებია დისპერსიული ანალიზის შემაჯამებელი მაჩვენებლები საოჯახო მეურნეობების საშუალო შემოსავლებსა და ბაკალავრის განათლების მქონეთა პროცენტული წილისათვის.

14.5 ცხრილში, სვეტში სათაურით „დისპერსიის წყარო“ რეგრესიის, ნაშთისა და მთლიანი კვადრატების ჯამებია მოყვანილი, ხოლო სვეტში სათაურით *df* მოყვანილია თავისუფლების ხარისხის შესაბამისი მნიშვნელობები რეგრესიისა და ნაშთების კვადრატების ჯამების მიხედვით. ბივარიანტული ანალიზის შემთხვევაში *SSR*-ს აქვს თავისუფლების ხარისხი 1; ხოლო *SSE*-სთან დაკავშირებული თავისუფლების ხარისხი არის *[N - (K + 1)]*, სადაც ასახავს დამოუკიდებელი ცვლადების რიცხვს რეგრესიის განტოლებაში. ბივარიანტულ ანალიზში, ერთი დამოუკიდებელი ცვლადით-საოჯახო მეურნეობის საშუალო შემოსავალი *SSE* აქვს თავისუფლების *N - 2* ხარისხი *[N - (1 + 1)]*.

ცხრილი 14.5 დისპერიული ანალიზი, შემაჯამებელი ცხრილი საშუალო შემოსავლები და ბაკალავრის განათლების ქონა

დისპერსიის წყარო	კვადრატების ჯამი	*df*	კვადრატული საშუალო	F
რეგრესია	94.758	1	94.758	29.675
ნაშთი	25.546	8	3.193	
ჯამი	120.304	9		

და ბოლოს, საშუალო კვადრატული რეგრესია (*MRS*) და საშუალო კვადრატული ნაშთი (*MSE*) გამოითვლება კვადრატების ჯამის თითოეული მნიშვნელობის შეფარდებით შესაბამის თავისუფლების ხარისხებთან. ჩვენი მაგალითის მიხედვით:

$$\text{საშუალო კვადრატული რეგრესია} = \frac{SSR}{1} = \frac{94.758}{1} = 94.758$$

$$\text{საშუალო კვადრატული ნაშთი} = \frac{SSE}{8} = \frac{25.546}{8} = 3.193$$

საშუალო კვადრატული რეგრესია საშუალო მაჩვენებელი, რომელიც გამოითვლება რეგრესიის კვადრატების ჯამის (*SSR*) შეფარდებით შესაბამის თავისუფლების ხარისხთან.

საშუალო კვადრატული ნაშთი საშუალო მაჩვენებელი, რომელიც გამოითვლება ნაშთის კვადრატების ჯამის (*SSE*) შეფარდებით შესაბამის თავისუფლების ხარისხთან.

F სტატისტიკა საშუალო კვადრატულ რეგრესიასა და საშუალო კვადრატულ ნაშთან ერთად ადგენენ *F* ფარდობასა და *F* სტატისტიკას. *F* სტატისტიკა არის საშუალო კვადრატული რეგრესიისა და საშუალო კვადრატული ნაშთის ფარდობა:

$$F = \frac{\text{საშუალო კვადრატული რეგრესია}}{\text{საშუალო კვადრატული ნაშთი}} = \frac{SSR / df_r}{SSE / df_e}$$

შესაბამისად, ფარდობა ასახავს საშუალი კვადრატული რეგრესიის ფარდობით ზომას საშუალო კვადრატულ ნაშთან; რაც უფრო დიდია ფარდობა, უფრო მეტია იმის ალბათობა, რომ r^2-ის მნიშვნელობა პოპულაციაში მნიშვნელოვნად იქნება ნოლისაგან განსხვავებული. ამ შემთხვევაში ჩვენ ვამოწმებთ ნულოვან

ჰიპოთეზას, იმის შესახებ რომ r^2 ნოლის ტოლია პოპულაციაში. მოდით გამოთვალოთ ფარდობა ჩვენი მაგალითის მიხედვით:

$$F = \frac{\text{საშუალო კვადრატული რეგრესია}}{\text{საშუალო კვადრატული ნაშთი}} = \frac{94.758}{3.193} = 29.68$$

გადაწყვეტილების მიღება. სტატისტიკის 29.68-ის მნიშვნელობის შესაბამისი ალბათობის მისაღებად გამოვიყენოთ წიგნის ე დანართი, რომელშიც მოცემულია განაწილების მნიშვნელობები თავისუფლების სხვადასხვა ხარისხებისა და მნიშვნელოვნების ორი 0.05 და 0.01 დონისათვის. ალფას მნიშვნელობა ავირჩიოთ 0.05-ის ტოლად და მივმართოთ განაწილების ცხრილის სვეტს „$P=0.05$“ სათაურით. მრიცხველი df_1 უკავშირდება df_r-ს საშუალო კვადრატული რეგრესიის, ხოლო მნიშვნელი df_2 უკავშირდება საშუალო კვადრატული ნაშთის df_e-ს. ჩვენს მაგალითში გამოთვლილი F სტატისტიკის მნიშვნელობა (29.68) შევადაროთ F-ის კრიტიკულ მნიშვნელობას. თუ თავისუფლების ხარისხი უდრის 1-ს (მრიცხველი), და 8-ს (მნიშვნელი), ხოლო $\alpha = 0.05$-ს შესაბამისი F-ის კრიტიკული მნიშვნელობაა 5.32. რადგან F-ის გამოთვლილი მნიშვნელობა მეტია F-ის კრიტიკულ მნიშვნელობაზე (29.68>5.32) უნდა უარვყოთ ნულოვანი ჰიპოთეზა პოპულაციაში r^2-ის ნოლთან ტოლობის შესახებ. ვასკვნით, რომ პოპულაციაში საშუალო შემოსავლებსა და ბაკალავრის განათლების მქონეთა პროცენტულ წილს შორს r^2-ით გამოსახული წრფივი კავშირი მეტია ნოლზე (ჩვენს შერჩევითი ერთობლიობის კოეფიციენტის მიხედვით).

✓ *ცოდნის შემოწმება*

შეამოწმეთ შემდეგი ნულოვანი ჰიპოთეზის მართებულობა: არსებობს წრფივი კავშირი ოჯახების საშუალო შემოსავლებსა და დანაშაულის დონეს შორის. საშუალო კვადრატული რეგრესია ურის 916,504.82-ს, თავისუფლების ხარისხი ერთის ტოლია, ხოლო საშუალო კვადრატული ნაშთი 174,732.41-ს ურდის თავისუფლების 8 ხარისხით. გამოთვალეთ F სტატისტიკა და შეაფასეთ მისი მნიშვნელოვნება.

დისპერიული ანალიზი მრავლობითი წრფივი რეგრესიისათვის

დისპერსიული ანალიზი მრავლობითი რეგრესიის შემთხვევაში თითქმის მარტივი წრფივი რეგრესიის მსგავსია, იმ განსხვავებით, რომ მრავლობითი რეგრესიის შემთხვევაში თავისუფლების ხარისხები შეესაბამება დამოუკიდებელ ცვლადთა რიცხვს მოდელში.

ჩვენ გამოვიყენეთ დისპერსიული ანალიზი, რომ შეგვეფასებინა წრფივი კავშირი მოზარდთა ფეხმძიმობასა და ერთ მოსწავლეზე დანარჯსა და უმუშევრობას შორის (იხილე მე-13 თავი), კავშირის გამოსახეთ r^2-ის მეშვეობით, ამ უკანასკნე-

ლის მნიშვნელობა კი ნოლის ტოლია. ამ ანალიზის შედეგები წარმოდგენილია ცხრილ 14.6-ში. თავისუფლების 2 და 47 ხარისხებით, F-ის 5.18 მნიშვნელობით უარვყოფთ ნულოვან ჰიპოთეზას, რომ $r^2 = 0$-ს, 0.01 დონეზე. რადგან F-ის გამოთვლილი მნიშვნელობა აღემატება კრიტიკულ მნიშვნელობას ($5.21 > 5.18$) უარვყოფთ ნულოვან ჰიპოთეზას $P < 0.01$ ალბათობით.

ცხრილი 14.6 შემაჯამებელი ცხრილი დისპერსიული ანალიზისათვის, მოზარდთა ფეხმძიმობის დონე, ერთ მოსწავლეზე დანახარჯი და უმუშევრობის დონე

დისპერსიის წყარო	კვადრატების ჯამი	*df*	საშუალო კვადრატი	F
რეგრესია	1,816.35	2	908.18	5.21
შეცდომის დონე	8,189.65	47	174.25	
ჯამი	10,006.00	49		

კვლევითი ლიტერატურის მიმოხილვა: პიროვნული თვითშეფასება და ეთნიკური იდენტობა

ბივარიანტული t ტესტისა და ხი-კვადრატ ანალიზის მსგავსად დისპერსიული ანალიზი საშუალებას გვაძლევს გავიგოთ თუ როგორ განსაზღვრავს ჩვენს სოციალურ ცხოვრებას ისეთი კატეგორიები როგორიცაა: რასა, სქესი, კლასი და ა.შ. პირველ მაგალითში ჩვენ განვიხილეთ განათლების მიღწევები სქესისა და რასის მიხედვით, დისპერსიული ანალიზი ასევე საშუალებას გვაძლევს გამოვიკვლიოთ სხვა სოციალური კატეგორიების გავლენა და მათ შორის განსხვავებები. დისპერსიული ანალიზის შესახებ ჩვენს მსჯელობას დავამთავრებთ ორი პრაქტიკული მაგალითის განხილვით, რომლებიც მოყვანილია სოციოლოგიური კვლევითი ლიტერატურიდან. დისპერსიული ანალიზი, როგორც სტატისტიკური კვლევის მეთოდი ფართოდ გამოიყენება სოციალურ ფსიქოლოგიაში. ორივე მაგალითი აღებულია სოციალური ფსიქოლოგიის ჟურნალებში გამოქვეყნებული კვლევებიდან.

ჩვენი პირველი მაგალითი ეფუძნება ელირეა ბონმანის 1999 წლის კვლევას ეთნიკურ იდენტობაზე სამხრეთ აფრიკაში[2]. პოლიტიკა, ენა, ეთნიკურობა, რასა და კულტურა თამაშობს მნიშვნელოვან როლს თვითშეფასებასა და ეთნიკურ იდენტობის თავისებურებებში სამხრეთ აფრიკაში. მაგრამ ბონმანის მიხედვით უფრო მნიშვნელოვანია, რომ ეთნიკურ იდენტობასა და თვითშეფასებას შორის კავშირის არსებობა განსაკუთრებით რელევანტურია ისეთ საზოგადოებებში, რომელიც ხასიათდება „რასობრივ და ეთნიკურ განსხვავებებზე დაფუძნებული კომპლექსური პლურალიზმით“[3]. ბონმანის კვლევითი ჰიპოთეზის მიხედვით – „თითოეული ჯგუფის დამოკიდებულება თავისი წევრების მიმართ, ასევე სოციალურ-ეკონომი-

კური და პოლიტიკური ცვლილებები, რასაც ადგილი ქონდა 1990-იანი წლებიდან სამხრეთ აფრიკაში გავლენა აქვს ეთნიკურ იდენტობასა და თვითშეფასებას შორის კავშირზე"[4].

ბონმანმა შეისწავლა განსხვავებები თვითშეფასებასა და ეთნიკურ იდენტობას შორის აფრიკულ ენაზე მოლაპარაკე თეთრკანიან, შავკანიან და ინგლისურად მოლაპარაკე თეთრკანიან რესპონდენტებს შორის. ფსიქოლოგიურ მეცნიერებათა კვლევების საბჭოს გამოკითხვების მონაცემებზე დაყრდნობით ბონმანმა გაანალიზა სამხრეთ აფრიკელი რესპონდენტების პასუხები თვითშეფასებისა და ეთნიკურ იდენტობისაზე არჩეულ სკალის მიხედვით. როზენბერგრის თვითშეფასების სკალა (1965) ზომავს, როგორც დადებით ასევე უარყოფით პასუხებს. ნეგატიურ სკალაზე მაღალი ქულა ასახავს უფრო ნეგატიურ თვითშეფასებას, ხოლო დადებით სკალაზე მაღალი ქულა ასახავს უფრო დადებით თვითშეფასებას[5]. ბონმანმა შექმნა ეთნიკური იდენტობის შეფასების სამი საზომი: 1. ეთნიკური იდენტობა (მაღალი ქულა ასახავს ძლიერ იდენტობას, ლოიალობას, პატივისცემასა და სიამაყეს საკუთარი ეთნიკური ჯგუფით); 2. ჩართულობა, იდენტობის ხარისხი (მაღლია ქულა ასახავს მაღალი ხარისხის ჩართულობას, აქტიურ მონაწილეობას კულტურულ და სხვა სახის აქტივობებში); 3. ამბივალენტურობა შენარჩუნების საპირწონედ. (მაღალი ქულა ასახავს იდენტობის დაცვისა და შენარჩუნების მაღალ სურვილს და ნაკლებ გაურკვეველობას ჯგუფის წევრობისადმი).[6] ცხრილ 147-ში ნაჩვენებია ბონმანის გამოთვლები-საშუალო მაჩვენებლები და სტანდარტული გადახრები სამი სამხრეთ აფრიკული ჯგუფისათვის. სტატისტიკის შესაბამისი თავისუფლების ხარისხები მოცემულია ცხრილის ბოლო ორ სვეტში. მნიშვნელოვნების დონე ამ ცხრილის მონაცემებისათვის არის 0.01.

მონაცემთა ანალიზის საფუძველზე ბონმანი წერს:

საშუალოდ, აფრიკულად მოსაუბრე თეთრკანიან მოსახლეობას აქვს ყველაზე დიდი დადებითი თვითშეფასება. მათი უარყოფითი თვითშეფასების საშუალო მნიშვნელობა ყველაზ დაბალი იყო, დადებითი შეფასების ყველაზე მაღალ მნიშვნელობასთან ერთად. საპირისპიროდ, შავკანიან რესპონდენტებს აქვთ ყველაზე მაღალი ნეგატიური თვითშეფასების ქულები ყველაზე დაბალ დადებით ქულებთან ერთად, რაც ნიშნავს, რომ მათ თეთრკანიან რესპონდენტებთან შედარებით აქვთ დაბალი თვითშეფასება. დისპერსიულმა ანალიზმა აჩვენა მნიშვნელოვანი სტატისტიკური განსხვავება სხვადასხვა ჯგუფების საშუალო მაჩვენებლებს შორის; ნეგატიური თვითშეფასებისას, $F(2,91)=99.47$, $p=0.0001$; დადებითი თვითშეფასებისას $F(2,921)=32.47$, $p=0.0001$.. აფრიკულად მოლაპარაკე თეთრკანიან რესპონდენტებს ასევე ქონდათ მაღალი ქულები ეთნიკური იდენტიფიკაციის მხრივაც, ხოლო ინგლისურად მოლაპარაკე თეთრკანიანი რესპონდენტებისათვის იგივე მაჩვენებელი იყო ყველაზე დაბალი. დისპერსიულმა ანალიზმა აჩვენა მნიშვნელოვანი სტატისტიკური განსხვავების არსებობა, $F(2,922)=8.52$, $p=0.0002$.. აფრიკულად მოლაპარაკე თეთრკანიან რესპონდენტებს ქონდათ ასევე უმაღლესი საშუალო ქულები ეთნიკურ იდენტობასთან დაკავშირებული ორი სხვა ცვლადის მიხედვითაც (მაგ: ჩართულობა,

იდენტობის ხარისხი). მაშინ როცა შავკანიან რესპონდენტებს ქონდათ დაბალი საშუალო მაჩვენებლები იგივე კატეგორიების მიხედვით. დისპერსიული ანალიზის შედეგები: $F(2,291)=16.79$, $p=0.0001$ ჩართულობისა და იდენტობის ხარისხისათვის; $F(2,922)=61.71$, $p=0.0001$ ამბივალენტურობისა და შენარჩუნების ფაქტორებისათვის7.

ცხრილი 14.7 საშუალო მნიშვნელობები და დისპერსიული ანალიზის შედეგები

	აფრიკულად მოლაპარაკე თეორკანიანი	შავკანიანი	ინგლისურად მოლაპარაკე თეორკანიანი	*df*	F
თვითშეფასება					
ნეგატიური	6.66 (2.39)	9.21 (2.89)	7.35 (2.89)	2, 921	99.47**
დადებითი	17.49 (2.08)	15.97 (3.31)	17.30 (2.29)	2, 921	32.47**
ეთნიკური იდენტურობის ფაქტორები					
ეთნიკური იდენტობა	37.92 (6.45)	36.71 (6.97)	35.01 (6.67)	2, 922	8.52**
ჩართულობა, იდენტობა და მონაწილეობა	22.84 (4.76)	20.73 (5.64)	21.13 (4.43)	2, 921	16.79**
ამბივალენტურობა და შენარჩუნება	15.36 (3.33)	12.64 (3.65)	13.42 (3.14)	2, 922	61.71**

შენიშვნები: სტანდარტული გადახრები მოცემულია ფრჩხილებში
** $P < 0.01$

შენიშნეთ, რომ ბონმანი F სტატისტიკის მნიშვნელობებს თავისუფლებს ხარისხთან და შესაბამის ალბათობებთან ერთად განიხილავს. მკვლევარი ცალკე არ აღნიშნავს საშუალო მნიშვნელობებს, იგი ჯგუფების მიხედვით განიხილავს უდიდეს და უმცირეს მნიშვნელობებს.

კვლევითი ლიტერატურის მიმოხილვა: სტრესი და დაძაბულობა

მუსილმა და მისმა კოლეგებმა (2009) შეისწავლეს ოჯახური დაძაბულობა და სტრესი იმ ბებიებსა და ბაბუებში, რომლებიც ჩართულნი არიან შვილიშვილების აღზრდაში[8]. წინა კვლევებმა აჩვენა, რომ ბებიები, რომლებიც ჩართულნი იყვნენ შვილიშვილების აღზრდაში უფრო მეტი სტრესის ნიშნები ქონდათ მათთან შედარებით ვინც არ იყო დაკავებული შვილიშვილების აღზრდით. შერჩევა მოიცავს

ბებიებს, რომლებიც დაყოფილნი არიან სამ კატეგორიად: მხოლოდ პირდაპირი შვილიშვილების აღზრდაში ჩაბმულნი, რადენიმე თაობის აღზრდაში ჩაბმულნი და შვილიშვილების აღზრდაში ჩაუბმელნი. ჯგუფი შემდეგნაირად იქნა განსაზღვრული: მხოლოდ შვილიშვილების პირდაპირ აღზრდაში ჩაბმულ ბებიებს აქვთ ვალდებულება აღზარდონ შვილიშვილები თავიანთ სახლებში ბავშვების მშობლების გარეშე; რამდენიმე თაობის აღზრდაში ჩაბმული ბებიები ზრდიან ერთ ან რადენიმე ბავშვს ბავშვის მშობლებთან ერთად; შვილიშვილების აღზრდაში ჩაუბმელნი არიან ისეთი ბებიები, რომლებიც არ არიან დაკავებულნი შვილიშვილების აღზრდით, მაგრამ რომელთაც 1 საათი მაინც უცხოვრიათ თავიანთ შვილიშვილებთან და აქვთ მათთან ურთიერთობა.[9]

მკვლევარებმა სტრესისა და დაძაბულობის ანალიზი მოახდინეს უკვე ცნობილი სკალის გამოყენებით:

- სტრესისა და დაძაბულობის დონე (კონფლიქტი, სირთულე) გაზომეს: ზოგადი ოჯახური სტრესის (ბავშვებს შორის კონფლიქტი, ბავშვების მართვის სირთულე); ოჯახური ცხოვრების სტრესი, ფინანსური (ფინანსური ვალდებულებების ზრდა), ტრანზიცია (ოჯახის წევრის დაკარგვა ან სამსახურიდან წამოსვლა, ახალ საცხოვრებელში გადასვლა), ოჯახური ძალადობა (ფიზიკური დაზიანების ინციდენტი ან აგრესია), ოჯახის წევრის დანაკარგი (ბავშვის გარდაცვალება), ჯანმრთელობის პრობლემა (ბავშვი გახდა სერიოზულად ავად ან დაშავდა) და ფეხმძიმობა (მოზარდის დაფეხმძიმება).
- სოციალური მხარდაჭერა შეფასდა დუკის სოციალური მხარდაჭერის ინდექსის მიეხდვით, რომელიც მოიცავს მხარდაჭერის სუბიექტურ და ინსტრუმენტალურ ასპექტებს. ინსტრსტრუმენტალური მხარდაჭერა გულისხმობს მხარდაჭერის იმ ოდენობას, რაც შესაძლებელია მიღებულ იქნას მეგობრებისაგან და ოჯახისაგან სპეციფიკურ სიტუაციებში. მაღალი ქულა მაღალ ინსტრუმენტალურ მხარდაჭერას ასახავდა. სუბიექტური მხარდაჭერა ასახავს დახმარების აღქმას და მეგობრებსა და ოჯახთან სიახლოვეს. მაღალი ქულა მაღალ სუბიექტურ მხადაჭერას ასახავს.
- რესურსებზე ხელმისაწვდომობა გაზომილ იქნა თვითშეფასების სკალით. მაღალი ქულა რესურსებზე ფართო ხელმისაწვდომობას ასახავდა.
- დეპრესიის სიმპტომები შეფასებულ იქნა 20 ბალიანი, ეპიდომოლოგიური კვლევების ცენტრის – დეპრესიის სკალის გამოყენებით. მაღალი ქულა ასახავს გაზრდილ კლინიკურ დეპრესიას.

მუსილისა და მისი კოლეგების კვლევის შედეგების ნაწილი წარმოდგენილია ცხრილ 14.8-ში. ცხრილში წარმოდგენილია სტრესისა და დაძაბულობის, სოციალური მხარდაჭერის, რესურსებზე ხელმისაწვდომობისა და დეპრესიის სიმპტომების საშუალო მაჩვენებლები და სტანდარტული გადახრები დისპერსიულ ანალიზზე დაყრდნობით. ქვემოთა პარაგრაფში მოტანილია მკვლევართა მიერ აღწერილი სხვაობები:

ჯგუფების მიხედვით დაფიქსირდა მნიშვნელოვანი სხვაობები შიგაოჯახურ სტრესის დონეებში: ყველაზე მეტი სტრესი ქონდათ შვილიშვილების პირდაპირ აღზრდაში ჩაბმულ ბებიებს (ცხრილი 1, ამ თავისათვის ცხრილი 14.8), სხვაობები არ იყო მნიშვენლოვანი ოჯახური ცხოვრების სტრესის მიხედვით. მაგრამ ამასთან იყო მნიშვნელოვანი სხვაობები რადენიმე თაობის აღზრდაში ჩაბმულ ბებიებში ცალკეული ოჯახური ცხოვრების სტრესის მიხედვით.. შემდგომმა ანალიზმა აჩვენა ნაკლები ფინანსური შეზღუდვების ქონა შვილიშვილების აღზრდაში ჩაუბმელი ბებიებისათვის სხვა კატეგორიებთან შედარებით. მნიშვნელოვანი სხვაობები არ დაფიქსირდა ოჯახურ ზრუნვასთან, ფეხმძიმობასა და დანაკარგებთან დაკავშირებული სტრესის კატეგორიებში. მნიშვნელოვანი სხვაობები იყო ჯგუფების მიხედვით მხარდაჭერასა და რესურსებზე ხელმისაწვდომობას შორის. პირდაპირი შვილიშვილების აღზრდაში ჩაბმულ ბებიებს ზოგადად ქონდათ სტრესის უფრო მაღალი დონე სხვა კატეგორიებთან შედარებით.[10]

შევნიშნოთ, რომ F ტესტის შედეგები არ არის წარმოდგენილი შემაჯამებელ ცხრილში. 14.8 ცხრილიდან ვიცით ცალკეული მოდელის მნიშვნელოვნება. მაგალითად, დისპერსიული ანალიზის მოდელის მიხედვით, შიგაოჯახური სტრესისათვის F-ის მიღებული მნიშვნელობაა 184 (0.01 დონისათვის). სტრესის საშუალო მაჩვენებლების განხილვაზე დაყრდნობით ყველაზე მაღალი დონის სტრესი აღმოაჩნდა შვილიშვილების პირდაპირ აღზრდაში ჩაბმულ პირებს (საშუალო მაჩვენებელი 4.4), შემდეგი კატეგორია სტრესის მიხედვით იყო რადენიმე თაობის აღზრდაში ჩაბმული პირები (საშუალო მაჩვენებელი 3.9). მუსილმა და მისმა კოლეგებმა დაასკვნეს, რომ ეს შედეგები ასახავს უფრო კომპლექსურ ოჯახურ ცხოვრებას ასეთი ტიპის ოჯახებში. გარდა უშუალოდ ბავშვების აღზრდისა ამგვარი ბებიები განიცდიან დამატებით სტრესს პირადი ცხოვრების ნაკლებობის, ნაკლები თავისუფალი დროისა და ოჯახური კონფლიქტის გამო.

ცხრილი 14.8 საშუალო მაჩვენებელი, სტანდარტული გადახრა და დისპერსიული ანალიზი ბავშვების აღზრდაში ჩაბმული ჯგუფებისათვის

	შვილიშვილების პირდაპირ აღზრდაში ჩაბმული ბებიები (n=183)		რამდენიმე თაობის აღზრდაში ჩაბმული ბებიები (n=136)		შვილიშვილების აღზრდაში ჩაუბმელი (n=167)		
ცვლადები	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F ტესტი*
შიგაოჯახური სტრესი	4.4	2.8	3.9	2.9	2.7	2.3	18.4***
ოჯახური ცხოვრების სტრესი (ზოგადი)	5.2	3.2	5.4	3.1	4.6	2.9	4.1
ფინანსური	1.2	0.8	1.1	0.8	0.9	0.8	5.8*
ტრანზიცია	1.7	1.5	2.3	1.7	1.9	1.5	4.7**
ოჯახური ძალადობა	0.7	0.9	0.4	0.7	0.4	0.7	9.9***

შიგაოჯახური სტრესი	4.4	2.8	3.9	2.9	2.7	2.3	18.4***
ოჯახური ცხოვრების სტრესი (ზოგადი)	5.2	3.2	5.4	3.1	4.6	2.9	4.1
ფინანსური	1.2	0.8	1.1	0.8	0.9	0.8	5.8*
ტრანზიცია	1.7	1.5	2.3	1.7	1.9	1.5	4.7**
ოჯახური ძალადობა	0.7	0.9	0.4	0.7	0.4	0.7	9.9***

შენიშვნა: * $P <= 0.05$, ** $P < 0.01$, *** $P < 0.001$.

ძირითადი განმარტებები

• დისპერსიული ანალიზი (ANOVA) საშუალებას გვაძლევს შევისწავლოთ ორზე მეტი შერჩევითი ერთობლიობების საშუალო მაჩვენებლების ვარიაცია. საშუალო მაჩვენებლებს შორის განსხვავების მნიშვნელოვნების დონის შეფასებისათვის დისპერსიული ანალიზის მეთოდი ითვალისწინებს, როგორც შერჩევით ერთობლიობებს შორის განსხვავებების, ასევე თითოეულ შერჩევით ერთობლიობაში განსხვავებების ანალიზს.

• ცალმხრივი დისპერსიული ანალიზი გამოყენება ერთი დამოკიდებული და ერთი დამოუკიდებელი ცვლადის მიმართ. ეს მეთოდი ჰიპოთეზის ტესტირებისას ხუთ ეტაპიანი პროცედურის გავლას გულისხმობს.

• დისპერსიული ანალიზისათვის სტატისტიკურ კრიტერიუმს წარმოადგენს. სტატისტიკა ჯგუფშორისი და შიგაჯგუფური დისპერსიების ფარდობას წარმოადგენს.

• დისპერსიული ანალიზი ასევე შესაძლებელია გამოყენებულ იქნას რეგრესიის მოდელის დეტერმინაციის კოეფიციენტისა (r^2) და მრავლობითი დეტერმინაციის კოეფიციენტის (R^2) მნიშვნელოვნების შესაფასებლად.

საკვანძო ტერმინები

დისპერსიული ანალიზი
ჯგუფშორისი კვადრატების ჯამი (*SSB*)
F კრიტიკული
F-ის გამოთვლილი მნიშვნელობა
F ფარდობა ან *F* სტატისტიკა
მონაცემთა შორის კვადრატების საშუალო
უმცირეს კვადრატთა რეგრესია
საშუალო კვადრატული ნაშთი
მონაცემთა შიგა კვადრატების ჯამი
ცალმხრივი დისპერსიული ანალიზი
რეგრესიის კვადრატების ჯამი (*SSR*)
ნაშთის კვადრატების ჯამი (*SSE*)
კვადრატების მთლიანი ჯამი (*SST*)
შიგაჯგუფური კვადრატების ჯამი (*SSW*)

დამატებითი რესურსები

ეწვიეთ ვებზე დაფუძნებულ სტუდენტთა სასწავლო საიტს, შემდეგ მისამართზე: **www.pineforge.com/ssds6e**. საიტზე იპოვით დამატებით სასწავლო მასალებს: კითხვებს, ამოცანებს, ვებ-რესურსებს, ლინკებს რომელთა მეშვეობით გადახვალთ სოციალური მეცნიერების ჟურნალების სტატიებზე. ამ სამეცნიერო სტატიებში გამოყენებულია სტატისტიკის ის მეთოდები, რომლებიც ამ თავში განვიხილეთ.

სტატისტიკა SPSS-ის გამოყენებით

[GSS08PFP-A]

მაგალითი 1: დისპერსიული ანალიზის მოდელის გამოთვლა

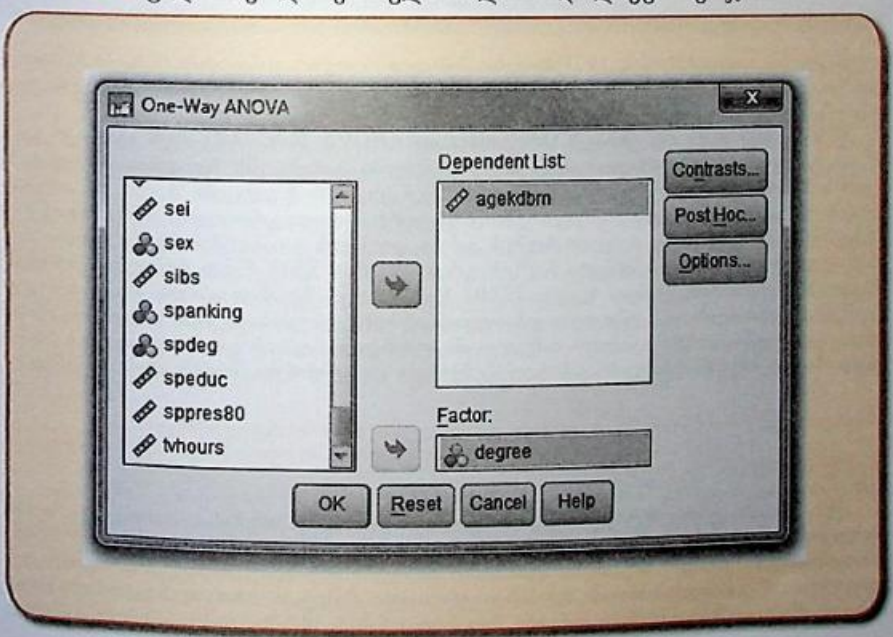
სოციოლოგებმა შეისწავლეს კავშირი ქალების მიერ ბავშვების ყოლის შესახებ გადაწყვეტილებასა და ხელფასის ოდენობას, ასევე დასაქმების სტატუსსა და განათლებას და სხვა სოციალურ-ეკონომიკურ და დემოგრაფიულ მაჩვენებელს შორის. კვლევამ აჩვენა, რომ განსხვავებულ სოციალურ ჯგუფებს აქვს განსხვავებული შეხედულებები ფერტილობასთან დაკავშირებით[11].

ჩვენს პირველ მაგალითში განვიხილავთ ქალის განათლების დონესა და მის მიერ პირველი ბავშვის ყოლის ასაკს შორის კავშირს. განათლება მივიჩნიოთ, როგორც დამოუკიდებელი ცვლადი, ხოლო პირველი ბავშვის დაბადების ასაკი, როგორც დამოკიდებული ცვლადი. შევაფასოთ ამ ორ ცვლადს შორის კავშირი.

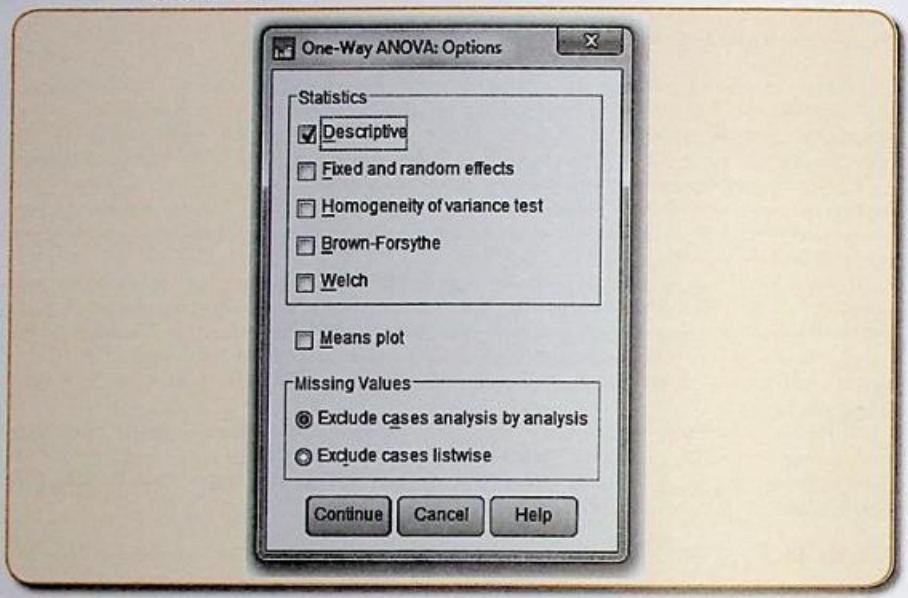
ჩვენი ანალიზისათვის გამოვიყენებთ ორ ცვლადს: DEGREE (განათლების დონის ხუთი კატეგორია) და AGEKDBRN (პირველი ბავშვის ყოლის ასაკი). თავდაპირველად, ჩვენი ანალიზი შემოიფარგლება მხოლოდ იმ ქალებით, რომელთა პასუხები მოტანილია ზოგადი სოციოლოგიური კვლევის მონაცემებში (GSS, მონაცემების გამოტანა: აირჩიეთ ბრძანება – Select Cases-if the condition is satisfied, ოპციის ფანჯარაში მიუთითეთ SEX=2 (ბაზიდან მხოლოდ ქალი რესპონდენტების შედეგებს მივიღებთ)).

SPPS-ში დისპერსიული ანალიზის შედეგების გამოთვლა ბრძანებათა შემდეგი თანმიმდევრობით ხდება: Analyze - Compare Means-One Way Anova. გამოტანილ დიალოგურ ფანჯარაში დამოკიდებული ცვლადის ოპციაში გადაიტანეთ ცვლადი AGEKBRN, ხოლო ოპციაში სახელწოდებით „Factor" ჩავსვით DEGREE. (იხ. ნახ. 14.2)

ნახაზი 14.2 ცალმხრივი დისპერსიული ანალიზის დიალოგური ფანჯარა SPSS-ში



ნახაზი 14.3 ცალმხრივი დისპერსიული ანალიზის დიალოგური ფანჯარის ოპციები SPSS-ში



დაკლიკეთ „ოპციებს“ (ნახ. 14.2), მიიღებთ შემდეგ დიალოგურ ფანჯარას (ნახ. 14.3) სადაც აირჩეთ Descriptive, შედეგად მიიღებთ საშუალოს, სტანდარტული გადახრისა და დისპერსიული ანალიზის გამოთვლის შედეგებს (ნახ.14.3). დაკლიკეთ Continue-OK.

SPSS გამოგვიტანს ორ ცხრილს, Descriptives და ANOVA (ნახ. 14.4). ჩვენ გვაინტერესებს F სტატისტიკა და დისპერსიული ანალიზის მნიშვნელობა. გამოთვლის შედეგების მიხედვით F ტოლია 53.742-ს 0.000 მნიშვნელოვნების დონით. ჩვენ ვიცით, რომ ჯგუფებს შორის განათლების დონის მიხედვით განსხვავება მნიშვნელოვანია, რაც უფრო მაღალი განათლების დონე, მით უფრო მოგვიანებით ასაკში ხდება პირველი ბავშვის გაჩენა. კოლეჯის განათლების მქონე ქალებს შორის ყველაზე დიდი ასაკი პირველი ბავშვის გაჩენისას არის 28.17 წელი, რომელსაც მოსდევს ბაკალავრის ხარისხის მქონეთა ჯგუფი (27.03 წელი). ყველაზე ახალგაზრდა ასაკი პირველი ბავშვის გაჩენისათვის აქვს უმაღლესი განათლების არმოქნე ქალების ჯგუფს. საშუალოდ, უმაღლესი განათლების არმქონე ქალების საშუალო ასაკი პირველი ბავშვის გაჩენისათვის არის 20.05 წელი, ხოლო სხვაობა უმაღლესი განათლების მქონეთა და არმქონეთა შორის არის 8.12 წელი.

მაგალითი 2: დისპერსიული ანალიზის გამოთვლები, ბივარიანტული რეგრესია

[GSS08PFP-B]

აქვს თუ არა უფრო მეტი განათლების მქონე პირებს მეტი სამუშაო დავირთვა კვირეების მიხედვით? ეს საკითხი ჩვენ უკვე განვიხილეთ მე-13 თავში SPPS-ის გამოყენებით. ჩვენ SPPS-ის გამოყენებით ვიპოვეთ საუკეთესო შესაბამისობის რეგრესიის წრფე და დეტერმინაციის კოეფიციენტი. ამ გამოთვლებისათვის შესაბამისი ბრძანებები შემდეგია: Analyze, Regression, Liner. წრფივი რეგრესიის დიალოგური ფანჯარა მოცემულია ნახაზ 14.5-ზე. დამოუკიდებელ ცვლადად

ჰადვიტანთ WEEKSWRK, ხოლო დამოუკიდებელ ცვლადად EDUC. წრფივი რეგრესიის დიალოგური ფანჯარა გვაძლევს ასევე სხვა არჩევანის შესაძლებლობებსაც, მაგრამ ამ შემთხვევაში აუტომატურად მოცემული ოფციები ჩვენთვის მისაღებია.

ნახაზი 14.4 დისპერსიული ანალიზის გამოთვლის შედეგები-პირველი ბავშვის დაბადების ასაკი და განთლება ქალებისათვის

Descriptives

R'S AGE WHEN 1ST CHILD BORN

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
LT HIGH SCHOOL	101	20.05	4.357	.434	19.19	20.91	13	36
HIGH SCHOOL	328	21.95	4.053	.224	21.51	22.39	13	39
JUNIOR COLLEGE	62	22.97	4.929	.626	21.72	24.22	15	40
BACHELOR	91	27.03	4.608	.483	26.07	27.99	18	44
GRADUATE	58	28.17	5.846	.768	26.64	29.71	16	39
Total	640	23.04	5.140	.203	22.64	23.44	13	44

ANOVA

R'S AGE WHEN 1ST CHILD BORN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4268.921	4	1067.230	53.742	.000
Within Groups	12610.179	635	19.859		
Total	16879.100	639			

SPSS გამოგვიტანს გამოთვლების საკმაო ბევრ შედეგებს. ჩვენ აქ მხოლოდ მათ ორ ნაწილს განვიხილავთ (ნახ. 14.6). მოდელის შეჯამებაში დეტერმინაციის კოეფიციენტი აღნიშნულია, როგორც „R Square", მისი მნიშვნელობა 0.036-ია, რომელიც სუსტ კავშირს ასახავს. განათლების დონე სამუშაო დატვირთვის ვარიაციის მხოლოდ მცირე ნაწილს ხსნის-3.6 %-ს. ეს შედეგი ალბათ მოულოდნელიც არა, მაგალითად, პირებს, რომლებსაც აქვთ საკუთარი მცირე ბიზნესი დიდ დროს უთმობენ თავიანთ სამუშაოს უმაღლესი განათლების არქონის პირობებში.

ANOVA-ს ცხრილი გვაძლევს დისპერსიული ანალიზის გამოთვლის შედეგებს. ცხრილის ბოლო მოიცავს ნაშთის კვადრატების ჯამს და კვადრატების საშუალოსაც. ნულოვანი ჰიპოთეზის r^2-ის ნოლთან ტოლობის შესახებ შემოწმებისათვის გავჭირდება ცხრილის სვეტი სახელწოდებით „Sig". ეს არის P-ს ის მნიშვნელობა, რომელიც უკავშირდება F ფარდობას, ცხრილის F სვეტში. F სტატისტიკის მნიშვნელობაა 54,879, ხოლო შესაბამისი P უდრის 0.000. ეს ნიშნავს, რომ არსებობს ძალიან მცირე ალბათობა (0.000) იმისა, რომ r^2 უდრის ნოლს პოპულაციაში, იმ პირობებში როცა დაკვირვების r^2 0.36-ია. შესაბამისად, შეგვიძლია უარვყოთ ნულოვანი ჰიპოთეზა, 0,05 დონეზე.

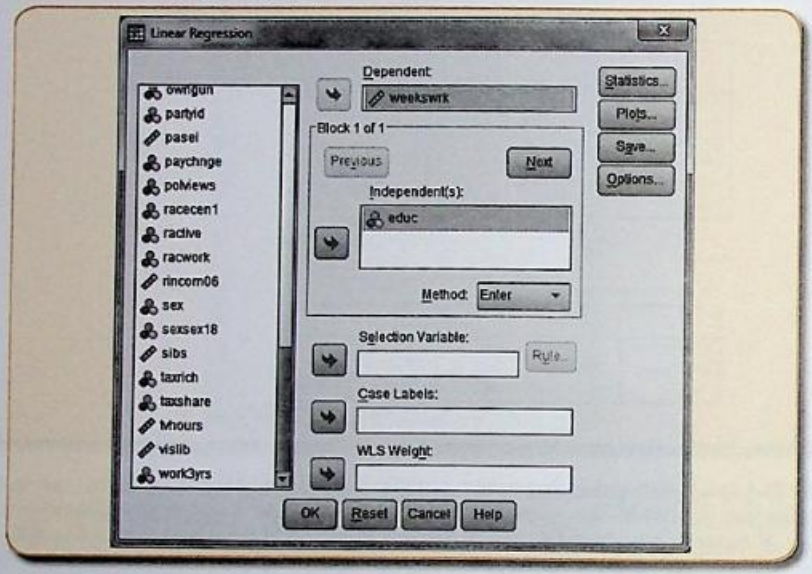
მაგალითი 3: დისპერსიული ანალიზის გამოთვლები, მრავლობითი რეგრესია

განათლების გარდა სხვა რა ფაქტორებს აქვს გავლენა სამუშაო კვირებით დატვირთვაზე? ამ კითხვაზე ერთი პასუხი ასაკია (AGE). ეს საკითხი ასევე უკვე განვიხილეთ მე-13 თავში. ამ კითხვაზე პასუხის გაცემისათვის გამოვიყენებთ SPSS-ს მრავლობითი რეგრესიის განტოლებისა და მრავლობითი დეტერმინაციის კოეფიციენტის გამოთვლისათვის. ამ შემთხვევაში პროგრამაში პრძანებათა თანმიმდევრობა ბივარიანტული რეგრესიის განტოლების მიღებისას გამოყენებულ პრძანებათა თანმიმდევრობის მსგავსია. პროგრამაში დაკლიკეთ შემდეგ ბრძანებათა შემდეგი მიმ-

დევრობა Analyze-Regression-Liner. წრფივი რეგრესიის დიალოგურ ფანჯარაში (ნახ. 14.7) დამოკიდებულ ცვლადად გადავიტანოთ WEEKSWRK, ხოლო დამოუკიდებელ ცვლადებად EDUC (განათლების წლები) და AGE (ასაკი წლებში).

SPSS გამოგვიტანს გამოთვლების საკმაო ბევრ შედეგებს. განვიხილოთ გამოთვლების ორი ნაწილი. (ნახ. 14.8).

ნახაზი 14.5 წრფივი რეგრესიის დიალოგური ფანჯარა ერთი დამოუკიდებელი ცვლადით



ნახაზი 14.6 გამოთვლის შედეგების ნაწილი განათლებისა და სამუშაო დატვირთვას შორის კავშირისათვის

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.189[a]	.036	.035	22.367

a. Predictors: (Constant), HIGHEST YEAR OF SCHOOL COMPLETED

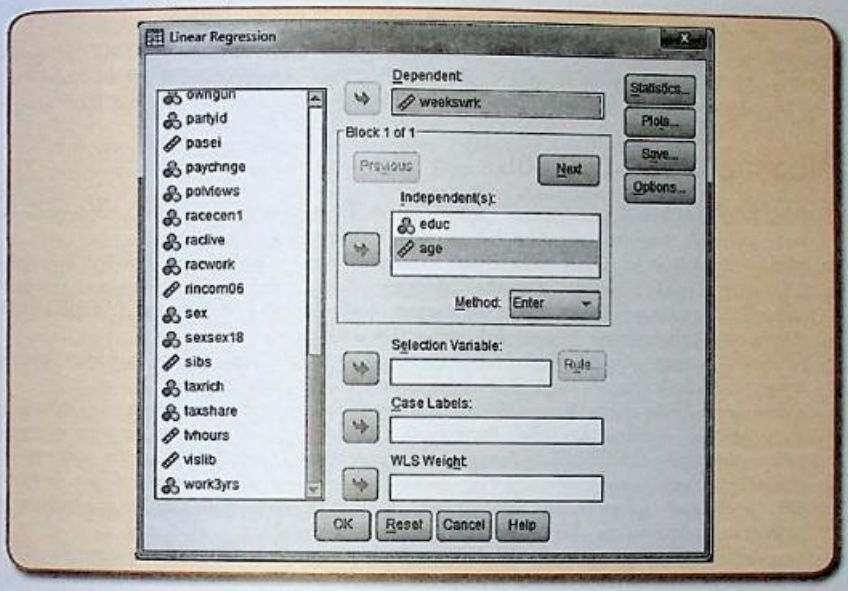
ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	27454.059	1	27454.059	54.879	.000[a]
	Residual	740888.921	1481	500.263		
	Total	768342.980	1482			

a. Predictors: (Constant), HIGHEST YEAR OF SCHOOL COMPLETED

b. Dependent Variable: WEEKS R. WORKED LAST YEAR

ნახაზი 14.7 წრფივი რეგრესიის დიალოგური ფანჯარა, ორი დამოუკიდებელი ცვლადით



ნახაზი 14.8 გამოთვლების შეჯამების ნაწილი, განათლებას, ასაკსა და სამუშაო დატვირთვას შორის კავშირისათვის

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.431[a]	.186	.184	20.540

a. Predictors: (Constant), AGE OF RESPONDENT, HIGHEST YEAR OF SCHOOL COMPLETED

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	141529.127	2	70764.564	167.732	.000[a]
	Residual	621021.391	1472	421.890		
	Total	762550.518	1474			

a. Predictors: (Constant), AGE OF RESPONDENT, HIGHEST YEAR OF SCHOOL COMPLETED

b. Dependent Variable: WEEKS R. WORKED LAST YEAR

მოდელის შემაჯამებელ ცხრილში მრავლობითი კორელაციის კოეფიციენტი აღნიშნულია, როგორც „R“ და უდრის 0.431. ამ მნიშვნელობის მიხედვით განათლებასა, ასაკსა და სამუშაო დატვირთვას შორის კავშირი საშუალო სიძლიერისაა. დეტერმინაციის კოეფიციენტი აღნიშნულია როგორც „R Square“ და მისი მნიშვნელობა 0.184-ის ტოლია. მოსწორებული R კვადრატი ით-

ვალისწინებს მოდელში დამოუკიდებელი ცვლადების რაოდენობას. ზოგადად, R^2-ის მნიშვნელობა მით უფრო ნაკლები იქნება R-თან შედარებით, რაც უფრო მეტია მოდელში დამოუკიდებელი ცვლადები. R^2-ის 0.186-თან ტოლობა ნიშნავს, რომ განათლება და ასაკი ერთობლივად ხსნიან სამუშაო დატვირთვის ვარიაციის დაახლოებით 18.6%-ს.

დისპერსიული ანალიზის ცხრილი ასახავს F სტატისტიკას (167.732) და P (0.000), რაც ნიშნავს რომ გვაქვს სტატისტიკურად მნიშვნელოვანი მოდელი. სხვა სიტყვებით, რომ თქვათ რადგან $P < 0.05$-ზე შეგვიძლია უარვყოთ ნულოვანი ჰიპოთეზა R^2-ის ნოლთან ტოლობის შესახებ 0.05 დონეზე.

ამოცანები SPSS-ისთვის

1. GSSS08PFP-A ბაზის გამოყენებით გავაგრძელოთ ბავშვების ყოლასა და განათლებას შორის კავშირის ანალიზი. ამჯერად მამაკაცების შემთხვევაა განვიხილოთ.
ა. ბაზიდან შესაბამისი ბრძანების გამოყენებით აირჩიეთ მხოლოდ მამაკაცების შესახებ მონაცემები
ბ. გამოთვალეთ დისპერსიული ანალიზის მოდელი მამაკაცებისათვის შემდეგი ცვლადების გამოყენებით: დამოკიდებული ცვლადი-პირველი ბავშვის დაბადების ასაკი (AGEKDBRN), დამოუკიდებელი ცვლადი განათლების დონე (DEGREE).
გ. SPSS-ის გამოთვლებზე დაყრდნობით შეგვიძლია თუ არა დავასკვნათ განათლებასა და პირველი ბავშვის ყოლის ასაკს შორის კავშირის არსებობის შესახებ?
დ. გამოთვალეთ მეორე დისპერსიული ანალიზის მოდელი მამაკაცებისათვის, რომელშიც დამოკიდებულ ცვლადად მიიჩნიეთ ბავშვების რაოდენობა (CHILDS), ხოლო დამოუკიდებელ ცვლადად განათლების დონე (DEGREE). გამოთვლის შედეგების მიხედვით რა დასკვნებს გამოიტანთ?
2. გაიმეორეთ 1.ბ-ის გამოთვლები, ამჯერად დამოუკიდებლ ცვლადად რესპონდენტის სოციალური კლასი გამოიყენეთ (CLASS) ქალებისა და მამაკაცებისათვის განცალკევებულად. რა კავშირია CLASS-სა და AGEKDBRN ცვლადებს შორის?
3. გავაგრძელოთ ბავშვის ყოლის შესახებ გადაწყვეტილებების ანალიზი. ანალიზში ჩავრთოთ პასუხები კითხვაზე: თქვენი აზრით რა არის შვილების ყოლის იდეალური რაოდენობა? (ცვლადი CHILDIDEL). GSSS08PFP-A ბაზის გამოყენებით, CHILDIDEL მიიჩნიეთ დამოკიდებულ, ხოლო DEGREE როგორც დამოუკიდებელ ცვლადად. არსებობს თუ არა განსხვავება ბავშვების იდეალური რაოდენობის შესახებ აზრსა და განათლებას შორის? (შეგიძლიათ ჩაატაროთ ანალიზი: მთლიანად ყველა რესპონდენტის მონაცემების, ქალი და მამაკაცი რესპონდენტების მონაცემების მიხედვით).
4. მე-13 თავში განხილული მაგალითის მსგავსად, გამოიყენეთ GSSS08PFP-A ბაზა რათა გაარკვიოთ კავშირი რესპონდენტის განათლებასა (EDUC) და მისი მშობლების (დედისა და მამის) განათლებას შორის (ცვლადის დასახელებები: PAEDUC, MAEDUC).
ა. რას უდრის F სტატისტიკა და რა არის მისი მნიშვნელობა? შესაძლებელია თუ არა ნულოვანი ჰიპოთეზის უარყოფა R^2-ის ნოლთან ტოლობის შესახებ?
ბ. აუხსენით თუ არა მოდელის გამოთვლის შედეგებს რესპონდენტის დედის განათლების შესახებ მონაცემების გამოყენება? შეადარეთ თქვენი გამოთვლის შედეგები 4.ა-ში მიღებულ შედეგებს, როცა მხოლოდ მამის განათლების დონე გაითვალისწინეს, როგორც დამოუკიდებელი ცვლადი.
5. GSSS08PFP-B მონაცემთა ბაზის გამოყენებით, შეაფასეთ რესპონდენტთა დამოკიდებულება აფირმაციულ ქცევასთან მიმართებაში შემდეგი ორი ცვლადის გამოყენებით: AFFRMACT და DISCAFF. AFFRMACT ცვლადი ასახავს რესპონდენტის მხარდაჭერას შავკანიანი პირების პრეფერენციული დასაქმებისა და დაწინაურების მიმართ (მაღალი ქულა ნიშნავს მკვეთრ უარყოფით დამოკიდებულებას). DISCAFF ცვლადი ასახავს რესპონდენტების დამოკიდებულებას იმ ფაქტის მიმართ, -როცა ერთდაიგივე კვალიფიკაციის თეთრკანიანი და შავკანიანი პირებიდან თეთრკანიანი ვერ იღებს დასაქმების შეთავაზებას, ხოლო შავკანიანი კი იღებს. (მაღლია ქულა ნიშნავს-ძალიან ვერ მოვიწონებ). AFFRMACT და DISCAFF როგორც დამოუკიდებელი ცვლადების გამოყენებით, დაადგინეთ არის თუ არა მნიშვნელოვანი განსხვავება აფირმატიული ქცევის მიმართ დამოკიდებულებაში სოციალური კლასების მიხედვით (CLASS)? (უნდა განიხილოთ დისპერსიული ანალიზის ორი მოდელი, თითოეული მათგანი CLASS დამოუკიდებელი ცვლადით).

6. გამოიყენეთ GSSS08PFP-B მონაცემთა ბაზა რათა შეისწავლოთ კავშირის ოჯახის წევრების რაოდენობასა (HOMPOP) და კვირის მანძილზე ნამუშევარ საათებს შორის (HRS1). დამოუკიდებელ ცვლადად მიიჩნიეთ HOMPOP, ხოლო დამოკიდებულ ცვლადად HRS1. გამოიყენეთ SPSS დისპერსიული ანალიზის გამოთვლებისათვის.
რას უდრის F სტატისტიკისა და P-ს მნიშვნელობა? არის თუ არა რესპონდენტის ოჯახის წევრების რაოდენობასა და სამუშაო საათებს შორის კავშირი მნიშვნელოვანი? მოახდინეთ მიღებული შედეგების ინტერპრეტირება.
7. მე-13 თავის SPSS 1 ამოცანის მსგავსად გამოიყენეთ იგივე ცვლადები, მაგრამ ამჯერად ანალიზი ჩაატარეთ ქალი და მამაკაცი რესპონდენტების მიხედვით. ამისათვის გამოიყენეთ ბრძანებათა შემდეგი მიმდევრობა Data-Split File-Organize Output by Groups (Sex). შემდგომ ამისა SPSS გამოთვლის შედეგებს გამოიტანს სქესის მიხედვით. მოახდინეთ მიღებული F სტატისტიკის მნიშვნელობების ინტერპრეტირება. არის თუ ეს მოდელი სტატისტიკურად მნიშვნელოვანი?
8. მე-13 თავის SPSS 1 ამოცანის მსგავსად გამოიყენეთ იგივე ცვლადები, მაგრამ ამჯერად ანალიზი მოახდინეთ თეთრკანიანი და შავკანიანი რესპონდენტებისათვის განცალკევებულად. გამოიყენეთ ცვლადი RACECEN1. პროგრამაში გამოიყენეთ ბრძანებათა შემდეგი მიმდევრობა: Data-Split File-Organize Output by Groups (RACECEN1). SPSS ამჯერად შედეგებს გამოიტანს რასის მიხედვით განცალკევებულს. არის თუ არა განსხვავება თეთკანიანი და შავკანიანი რესპონდენტებისათვის შედგენილ რეგრესიის განტოლებებში? შეადარეთ F სტატისტიკის მნიშვნელობები და განიხილეთ მოდელის სტატისტიკური მნიშვნელოვნება.
9. მე-13 თავის SPSS 1 ამოცანის მსგავსად გამოიყენეთ იგივე ცვლადები, მაგრამ ამჯერად ანალიზი მოახდინეთ დაქორწინებული და განქორწინებული რესპონდენტებისათვის განცალკევებულად. გამოიყენეთ ცვლადი MARITAL. პროგრამაში გამოიყენეთ ბრძანებათა შემდეგი მიმდევრობა: Data-Split File-Organize Output by Groups (MARITAL). SPSS ამჯერად შედეგებს გამოიტანს ქორწინების სტატუსის მიხედვით განცალკევებულს.
იპოვეთ F ფარდობები და მოდელის მნიშვნელოვნების დონე დაქორწინებული და განქორწინებული რესპონდენტების მიხედვით. ამ მნიშვნელობებზე დაყრდნობით, რომელი მოდელია უფრო შესაბამისი რესპონდენტების ბავშვების რაოდენობის განსაზღვრისათვის?

თავის სავარჯიშოები

1. მე-13 თავში განვიხილეთ კავშირი გარემოს დაცვის საკითხებით დაინტერესებას შორისა და გარემოს დაცვის ჯგუფებისათვის თანხის გადარიცხვას შორის კავშირის შესახებ, მონაცემები მოიცავდა მსოფლიოს რვა ქვეყანას.

ქვეყანა	დაინტერესება (%)	თანხის ლოტარეა (%)
ამერიკის შეერთებული შტატები	33.8	22.8
ავსტრია	35.5	27.8
ნიდერლანდები	30.1	44.8
სლოვენია	50.3	10.7
რუსეთი	29.0	1.6
ფილიპინები	50.1	6.8
ესპანეთი	35.9	7.4
დანია	27.2	22.3

წყარო: International Social Survey Programme, 2000.

ამ თავში განხილული ფორმულებისა და მე-13 თავში განხილული მეთოდების გამოყენებით გამოთვალეთ F სტატისტიკა, რას უდრის შესაბამისი P-ს მნიშვნელობა? არის თუ არა გარემოს დაცვის საკითხებით დაინტერესებასა და გარემოს დაცვით ჯგუფებისათვის თანხის დოტაციას შორის კავშირი სტატისტიკურად მნიშვნელოვანი? მოახდინეთ გამოთვლის შედეგების ინტერპრეტირება.

2. მე-13 თავში განვიხილეთ კავშირი რესპონდენტის განათლებასა და მისი შვილების რაოდენობას შორის. ჩვენი კვლევის ჰიპოთეზა იყო-განათლების ზრდასთან ერთად იკლებს ბავშვების ყოლის რაოდენობა. ამჯერად გავარკვიოთ თუ რამდენად მნიშვნელოვანია ეს კავშირი. შეისწავლეთ ეს კავშირი 25 შემთხვევის გამოყენებით, GSS 2006-ის მონაცემთა ბაზის მიხედვით. ქვემოთა ცხრილში მოყვანილია მონაცემები რესპონდენტის განათლების დონისა (წლებში) და შვილების რაოდენობის შესახებ.

განათლების წლები	ბავშვების რაოდენობა	განათლების წლები	ბავშვების რაოდენობა
16	0	12	2
12	1	12	3
12	3	11	1
6	6	12	2
14	2	11	2
14	2	12	0
16	2	12	2
12	2	12	3
17	2	12	4
12	3	12	1
14	4	14	0
13	0	12	3
12	1	12	2

გამოთვალეთ F სტატისტიკა და მოახდინეთ შედეგის ინტერპრეტირება. $\alpha = 0.05$-ს მნიშვნელობა მიიჩნიეთ 0.05-ის ტოლად და შეამოწმეთ შემდეგი ნულოვანი ჰიპოთეზის მართებულობა $r^2 = 0$.

3. ამ დავალებაში შეისწავლეთ კავშირი რესპონდენტის სქესსა და მისი პირველი ბავშვის დაბადების ასაკს შორის, GSS 2006 მონაცემთა ბაზის გამოყენებით. დისპერსიული ანალიზის გამოთვლები SPSS-ში მოცემულია ნახაზ 14.9-ზე.

ნახაზი 14.9 დისპერსიული ანალიზის გამოთვლები SPSS-ში.

Descriptives

RS AGE WHEN 1ST CHILD BORN

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Male	306	25.59	5.654	.323	24.95	26.22	13	50
Female	428	22.57	5.194	.251	22.08	23.06	14	51
Total	734	23.83	5.589	.206	23.42	24.23	13	51

ANOVA

RS AGE WHEN 1ST CHILD BORN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1625.356	1	1625.356	55.933	.000
Within Groups	21271.015	732	29.059		
Total	22896.371	733			

ა. საშუალოდ პირველი ბავშვის დაბადების ასაკის მიხედვით ქალების თუ მამაკაცების მაჩვენებელებია უფრო მეტი?

ბ. გამოთვლების მიხედვით რა დასკვნის გამოიტანთ სქესსა პირველი ბავშვის დაბადების ასაკს შორის კავშირის შესახებ? $\alpha = .05$

4. ნახაზ 14.10-ზე ნაჩვენები გამოთვლის შედეგები SPSS-ში ასახავს განათლებასა (წლების მიხედვით) და დღის მანძილზე ტელევიზორის ყურების ხანგრძლივობას შორის (საათებში) კავშირს, GSS 2006. შეგვიძლია დავუშვათ ჰიპოთეზა, რომ განათლების ზრდასთან ერთად კლებულობს ტელევიზორის ყურების დრო ე.ი. ცვლადებს შორის უარყოფითი კავშირია. შეაფასეთ მოდელის მნიშვნელოვნების დონე F სტატისტიკისა და P-ს მნიშვნელობების მიხედვით. არის თუ არა ეს კავშირის სტატისტიკურად მნიშვნელოვანი?

ნახაზი 14.10 დისპერსიული ანალიზის გამოთვლები, განათლება და ტელევიზორის ყურება

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.195[a]	.038	.037	2.337

a. Predictors: (Constant), HIGHEST YEAR OF SCHOOL COMPLETED

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	140.431	1	140.431	25.710	.000[a]
	Residual	3539.415	648	5.462		
	Total	3679.846	649			

a. Predictors: (Constant), HIGHEST YEAR OF SCHOOL COMPLETED
b. Dependent Variable: HOURS PER DAY WATCHING TV

5. მე-10 თავში განვიხილეთ რამდენიმე ბივარიანტული ცხრილი, რომლებშიც წარმოდგენილი იყო მონაცემები რესპონდენტის მიერ აბორტის არჩევანის შესაძლებლობის მხარადჭერასა და ოჯახის ზომის შესახებ რესპონდენტის აზრის შესახებ. ვნახეთ, რომ დიდი ოჯახის მხარდაჭერის შესახებ აზრს აბორტის არჩევანის შესაძლებლობის უარყოფითი შეფასება

უკავშირდებოდა. ეს ანალიზი განვავრცოთ დისპერსიული ანალიზის გამოყენებით GSS 2008 მონაცემთა ბაზის გამოყენებით, აბორტისადმი მხარადაჭერა მივიჩნიოთ დამოუკიდებელ ცვლადად (1 თანხმობა, 2 უარყოფა), ხოლო დამოკიდებულ ცვლადად მივიჩნიოთ ბავშვების ყოლის რაოდენობა (CHILD). SPSS-ში გამოთვლების შედეგები ნაჩვენებია ნახაზ 14.11-ზე.

ნახაზი 14.11 CHILD ბავშვების რაოდენობა

Descriptives

NUMBER OF CHILDREN

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
YES	417	1.52	1.550	.076	1.37	1.67	0	8
NO	553	2.17	1.828	.078	2.02	2.33	0	8
Total	970	1.89	1.744	.056	1.78	2.00	0	8

ANOVA

NUMBER OF CHILDREN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	101.438	1	101.438	34.509	.000
Within Groups	2845.411	968	2.939		
Total	2946.849	969			

ა. რომელ ჯგუფს აქვს ბავშვების ყოლის ყველაზე მაღალი მაჩვენებელი?

ბ. არის თუ არა მნიშვნელოვანი განსხვავება ჯგუფებს შორის ბავშვების ყოლის რაოდენობის მიხედვით? (დაუშვით, რომ $\alpha = 0.05$) პასუხი დაასაბუთეთ.

გ. გამოთვალეთ η^2.

6. მე-10 თავში გავაანალიზეთ კავშირი სოციალურ კლასსა და ჯანმრთელობის მდგომარეობის შეფასებას შორის. ამ დავალებაში კვლავ გავიმეოროთ ანალიზი შერჩევითი ერთბოლოიობის 32 შემთხვევის ანალიზით, GSS 2006. ჯანმრთელობის მდგომარეობის შეფასება ხორციელდება შემდეგი შკალის გამოყენებით: 1 = შესანიშნავი, 2 = კარგი, 3 = საშუალო, 2 = ცუდი. სოციალური კლასის მიხედვით დაყოფა: დაბალი, მუშათა, საშუალო და მაღალი. გამოიყენეთ დისპერსიული ანალიზის 5 ეტაპის გამოთვლები ($\alpha = 0.05$).

დაბალი კლასი	მუშათა კლასი	საშუალო საშუალო	მაღალი კლასი
3	2	2	2
2	1	3	1
2	3	1	1
2	2	1	2
3	2	2	1
3	2	3	1
4	3	3	1
4	3	1	2

7. მე-5 დავალებაში მოცემული ანალიზი გავაფართოვოთ ახალი დამოუკიდებელი ცვლადის CHLDIDEL-ის ჩართვით. გამოთვლის შედეგები SPSS-ში მოყვანილი ნახაზ 14.12-ზე.
   ა. რომელ ჯგუფს აქვს ყველაზე მაღალი მაჩვენებელი ბავშვების იდეალური რაოდენობის მიხედვით?
   ბ. არის თუ არა მნიშვნელოვანი განსხვავება ჯგუფებს შორის ბავშვების იდეალური რაოდენობის მიხედვით? ($\alpha = .05$). პასუხი დაასაბუთეთ.
   გ. შეიცვლება თუ არა თქვენი პასუხი ბ დავალებაზე, თუ დაუშვებთ რომ $\alpha = .05$-ს?

ნახაზი 14.12 ბავშვების ყოლის იდეალური რაოდენობა

Descriptives

IDEAL NUMBER OF CHILDREN

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
YES	185	2.84	1.755	.129	2.59	3.10	0	8
NO	273	3.11	1.866	.113	2.89	3.34	0	8
Total	458	3.00	1.825	.085	2.84	3.17	0	8

ANOVA

IDEAL NUMBER OF CHILDREN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.057	1	8.057	2.427	.120
Within Groups	1513.934	456	3.320		
Total	1521.991	457			

8. ამ დავალებაში განვიხილოთ კავშირი განათებასა და რელიგიურ მსახურებებზე დასწრების სიხშირეს შორის. ამ ანალიზისათვის ავირჩიეთ 30 შერჩევითი მონაცემი საერთაშორისო სოციოლოგიური კვლევის მონაცემებიდან-განათლების დონე (ფორმალური განათლების გარეშე, საშუალო განათლება, უმაღლესი განათლება) და რელიგიურ მსახურებებზე დასწრების სიხშირე (0 – არასოდეს, 1 – იშვიათად, 2 – თვეში ორჯერ, სამჯერ ან მეტად).

გამოიყენეთ დისპერსიული ანალიზის ხუთეტაპიანი გამოთვლები,

ფორმალური განათლების გარეშე	საშუალო განათლება	უმაღლესი განათლება
2	2	0
1	2	0
1	2	0
2	1	0
2	1	1
2	0	1
0	2	0
2	1	1
2	1	2

9. კვლევის მონაცემების (Monitoring the Future) 21 შემთხვევაზე დაყრდნობით გავაანალიზოთ კავშირი რესპონდენტის რასობრივი/ეთნიკური კუთვნილებასა და ბოლო 4 კვირის მანძილზე გაცდენილი სკოლის დღეების რაოდენობას შორის.
   ა. გამოიყენეთ დისპერსიული ანალიზის ხუთეტაპიანი გამოთვლები,
   ბ. შეიცვლება თუ არა თქვენი პასუხი ბ დავალებაზე, თუ დაუშვებთ რომ $\alpha = 0.01$-ს? პასუხი განმარტეთ.

თეთრკანიანი	შავკანიანი	ესპანური წარმოშობის
4	1	4
5	2	3
3	2	5
4	1	1
4	3	5
4	4	2

10. როგორც მე-13 თავში აღინიშნა მეცნიერები დიდი ხანია დაინტერესებულნი არიან ამერიკის შეერთებულ შტატებში მისწრაფებებისა და მიღწევების შესწავლით. GSS 2006 მონაცემთა ბაზაში წარმოდგენილია მონაცემები რესპონდენტებისა და მათი დედების განათლების დონის შესახებ. გამოიყენეთ ეს მონაცემები და დაადგინეთ იმ პირისათვის, რომლის დედასაც აქვს მაღლი განათლების დონე არის თუ არა უფრო მეტი ალბათობა, რომ მასაც მაღალი განათლების დონე ექნება.
გამოთვალეთ F სტატისტიკა. რას უდრის P-ს მნიშვნელობა? არის თუ არა კავშირი რესპონდენტისა და მისი დედის განათლებას შორის? მოახდინეთ გამოთვლის შედეგების ინტერპრეტირება.

რესპონდენტის დედის განათლება	რესპონდენტის განათლება
0	12
15	13
6	9
9	12
16	16
12	12
6	16
18	14
12	13
14	12
14	18

11. მე-13 თავის მე-6 სავარჯიშოში განვილიხილეთ კავშირი ბავშვთა სიკვდილიანობის დონესა და მთლიან შიდა პროდუქტის მოცულობას შორის სამხრეთ ამერიკის ქვეყნებში. შობადობის დონეები (დაბადებულთა რაოდენობა ყოველ 1000 მცხოვრებზე) ნაჩვენებია ქვემოთა ცხრილში.

ქვეყანა	შობადობის დონე 2010
არგენტინა	18
ბოლივია	25
ბრაზილია	18
ჩილე	15
კოლუმბია	18
ეკვადორი	20
პარაგვაი	18
პერუ	19
ურიგვაი	14
ვენესუელა	13

წყარო: The World Factbook, 2010

გამოთვალეთ F სტატისტიკა, დამოუკიდებულ ცვლადად მიიჩნიეთ მშპ ერთ სულზე, ხოლო დამოკიდებულ ცვლადად შობადობის დონე. რას უდრის P-ს მნიშვნელობა? არის თუ არა ბავშვთა სიკვდილიანობის დონესა და ერთ სულზე მთლიან შიდა პროდუქტის მოცულობას შორის კავშირი მნიშვნელოვანი? მოახდინეთ გამოთვლის შედეგების ინტერპრეტირება.

12. როგორც მე-13 თავში განვიხილეთ, სიღარიბის ზღვარს ქვემოთ მცხოვრები პირები და ოჯახები დგანან მრავალი სირთულის წინაშე, რომელთაგან ერთ-ერთია ჯანდაცვის მომსახურების ხელმისაწვდომობა. ხშირ შემთხვევაში სიღარიბის ზღვარს ქვემოთ მცხოვრებ ადამიანებს არ აქვთ სამედიცინო დაზღვევა. აშშ-ს აღწერის ბიუროს მონაცემების გამოყენებით გაანალიზეთ კავშირი სიღარიბის ზღვარს ქვემოთ ცხოვრებასა და ჯანდაცვის მომსახურების ხელმისაწვდომობას შორის.

დამოუკიდებელ ცვლადად მიიჩნიეთ სიღარიბის ზღვას ქვემოთ მცხოვრები ადამიანების პროცენტული წილი. შესაძლებელია თუ არა შემდეგი ნულოვანი ჰიპოთეზის უარყოფა: $r^2 = 0$, დონისათვის?

შტატი	სიღარიბის ზღვარს ქვემოთ მცხოვრებთა პროცენტული წილი (2007)	დაზღვევის არმქონეთა პროცენტული წილი (2007)
ალაბამა	16.9	12.0
კალიფორნია	12.4	18.2
აიდაჰო	12.1	13.9
ლუიზიანა	18.6	18.5
ნიუ-ჯერსი	8.6	15.8
ნიუ-იორკი	13.7	13.2
პენსილვანია	11.6	9.5
როდ-აილენდი	12.0	10.8
სამხერთ კაროლინა	15.0	16.4
ტეხასი	16.3	25.2
ვაშინგტონი	11.4	11.3
ვისკონსინი	10.8	8.2

წყარო: U.S. Census Bureau, The 2010 Statistical Abstract, Table 693 and Table 150; U.S. Bureau of the Census, Current Population Reports P60–226, Income, Poverty and Health Insurance Coverage in the United States: 2003.

13. კვლევის მონაცემებიდან შერჩეულ იქნა 14 რესპონდენტი (MTF 2006). ქვემოთ წარმოდგენილია საგზაო მოძრაობის წესების დარღვევები ბოლო 12 თვის მანძილზე საცხოვრებელი ადგილის მიხედვით (საცხოვრებელი ადგილი დამოუკიდებელი ცვლადია). ამ მონაცემებისათვის გამოიყენეთ დისპერსიული ანალიზის ზუთეტაპიანი გამოთვლები, $\alpha = .05$.

პატარა ქალაქი	საშუალო ზომის ქალაქი	დიდი ქალაქი
0	2	3
0	3	4
1	1	4
2	1	3
1	2	2

დანართი ა
შემთხვევითი რიცხვების ცხრილი

14,000 *შემთხვევითი ერთეულის ცხრილი*

სტრ/სვეტი	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	10480	15011	01536	02011	81647	91646	69179	14194	62590	36207	20969	99570	91291	90700
2	22368	46573	25595	85393	30995	89198	27982	53402	93965	34095	52666	19174	39615	99505
3	24130	48360	22527	97265	76393	64809	15179	24830	49340	32081	30680	19655	63348	58629
4	42167	93093	06243	61680	07856	16376	39440	53537	71341	57004	00849	74917	97758	16379
5	37570	39975	81837	16656	06121	91782	60468	81305	49684	60672	14110	06927	01263	54613
6	77921	06907	11008	42751	27756	53498	18602	70659	90655	15053	21916	81825	44394	42880
7	99562	72905	56420	69994	98872	31016	71194	18738	44013	48840	63213	21069	10634	12952
8	96301	91977	05463	07972	18876	20922	94595	56869	69014	60045	18425	84903	42508	32307
9	89579	14342	63661	10281	17453	18103	57740	84378	25331	12566	58678	44947	05585	56941
10	85475	36857	43342	53988	53060	59533	38867	62300	08158	17983	16439	11458	18593	64952
11	28918	69578	88231	33276	70997	79936	56865	05859	90106	31595	01547	85590	91610	78188
12	63553	40961	48235	03427	49626	69445	18663	72695	52180	20847	12234	90511	33703	90322
13	09429	93969	52636	92737	88974	33488	36320	17617	30015	08272	84115	27156	30613	74952
14	10365	61129	87529	85689	48237	52267	67689	93394	01511	26358	85104	20285	29975	89868
15	07119	97336	71048	08178	77233	13916	47564	81056	97735	85977	29372	74461	28551	90707
16	51085	12765	51821	51259	77452	16308	60756	92144	49442	53900	70960	63990	75601	40719
17	02368	21382	52404	60268	89368	19885	55322	44819	01188	65255	64835	44919	05944	55157
18	01011	54092	33362	94904	31273	04146	18594	29852	71585	85030	51132	01915	92747	64951
19	52162	53916	46369	58586	23216	14513	83149	98736	23495	64350	94738	17752	35156	35749
20	07056	97628	33787	09998	42698	06691	76988	13602	51851	46104	88916	19509	25625	58104
21	48663	91245	85828	14346	09172	30168	90229	04734	59193	22178	30421	61666	99904	32812
22	54164	58492	22421	74103	47070	25306	76468	26384	58151	06646	21524	15227	96909	44592
23	32639	32363	05597	24200	13363	38005	94342	28728	35806	06912	17012	64161	18296	22851
24	29334	27001	87637	87308	58731	00256	45834	15398	46557	41135	10367	07684	36188	18510
25	02488	33062	28834	07351	19731	92420	60952	61280	50001	67658	32586	86679	50720	94953
26	81525	72295	04839	96423	24878	82651	66566	14778	76797	14780	13300	87074	79666	95725
27	29676	20591	68086	26432	46901	20849	89768	81536	86645	12659	92259	57102	80428	25280

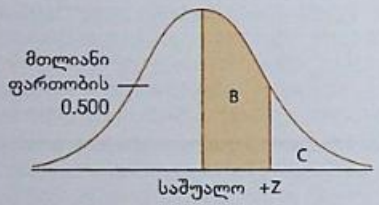
სტრ./ სვეტი	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
28	00742	57392	39064	66432	84673	40027	32832	61362	98947	96067	64760	64584	96096	98253
29	05366	04213	25669	26422	44407	44048	37937	63904	45766	66134	75470	66520	34693	90449
30	91921	26418	64117	94305	26766	25940	39972	22209	71500	64568	91402	42416	07844	69618
31	00582	04711	87917	77341	42206	35126	74087	99547	81817	42607	43808	76655	62028	76630
32	00725	69884	62797	56170	86324	88072	76222	36086	84637	93161	76038	65855	77919	88006
33	69011	65797	95876	55293	18988	27354	26575	08625	40801	59920	29841	80150	12777	48501
34	25976	57948	29888	88604	67917	48708	18912	82271	65424	69774	33611	54262	85963	03547
35	09763	83473	73577	12908	30883	18317	28290	35797	05998	41688	34952	37888	38917	88050
36	91567	42595	27958	30134	04024	86385	29880	99730	55536	84855	29080	09250	79656	73211
37	17955	56349	90999	49127	20044	59931	06115	20542	18059	02008	73708	83317	36103	42791
38	46503	18584	18845	49618	02304	51038	20655	58727	28168	15475	56942	53389	20562	87338
39	92157	89634	94824	78171	84610	82834	09922	25417	44137	48413	25555	21246	35509	20468
40	14577	62765	35605	81263	39667	47358	56873	56307	61607	49518	89656	20103	77490	18062
41	98427	07523	33362	64270	01638	92477	66969	98420	04880	45585	46565	04102	46880	45709
42	34914	63976	88720	82765	34476	17032	87589	40836	32427	70002	70663	88863	77775	69348
43	70060	28277	39475	46473	23219	53416	94970	25832	69975	94884	19661	72828	00102	66794
44	53976	54914	06990	67245	68350	82948	11398	42878	80287	88267	47363	46634	06541	97809
45	76072	29515	40980	07391	58745	25774	22987	80059	39911	96189	41151	14222	60697	59583
46	90725	52210	83974	29992	65831	38857	50490	83765	55657	14361	31720	57375	56228	41546
47	64364	67412	33339	31926	14883	24413	59744	92351	97473	89286	35931	04110	23726	51900
48	08962	00358	31662	25388	61642	34072	81249	35648	56891	69352	48373	45578	78547	81788
49	95012	68379	93526	70765	10593	04542	76463	54328	02349	17247	28865	14777	62730	92277
50	15664	10493	20492	38391	91132	21999	59516	81652	27195	48223	46751	22923	32261	85653
51	16408	81899	04153	53381	79401	21438	83035	92350	36693	31238	59649	91754	72772	02338
52	18629	81953	05520	91962	04739	13092	97662	24822	94730	06496	35090	04822	86772	98289
53	73115	35101	47498	87637	99016	71060	88824	71013	18735	20286	23153	72924	35165	43040
54	57491	16703	23167	49323	45021	33132	12544	41035	80780	45393	44812	12515	98931	91202
55	30405	83946	23792	14422	15059	45799	22716	19792	09983	74353	68668	30429	70735	25499
56	16631	35006	85900	98275	32388	52390	16815	69298	82732	38480	73817	32523	41961	44437
57	96773	20206	42559	78985	05300	22164	24369	54224	35083	19687	11052	91491	60383	19746
58	38935	64202	14349	82674	66523	44133	00697	35552	35970	19124	63318	29686	03387	59846
59	31624	76384	17403	53363	44167	64486	64758	75366	76554	31601	12614	33072	60332	92325
60	78919	19474	23632	27889	47914	02584	37680	20801	72152	39339	34806	08930	85001	87820
61	03931	33309	57047	74211	63445	17361	62825	39908	05607	91284	68833	25570	38818	46920
62	74426	33278	43972	10119	89917	15665	52872	73823	73144	88662	88970	74492	51805	99378
63	09066	00903	20795	95452	92648	45454	09552	88815	16553	51125	79375	97596	16296	66092
64	42238	12426	87025	14267	20979	04508	64535	31355	86064	29472	47689	05974	52468	16834
65	16153	08002	26504	41744	81959	65642	74240	56302	00033	67107	77510	70625	28725	34191
66	21457	40742	29820	96783	29400	21840	15035	34537	33310	06116	95240	15957	16572	06004
67	21581	57802	02050	89728	17937	37621	47075	42080	97403	48626	68995	43805	33386	21597
68	55612	78095	83197	33732	05810	24813	86902	60397	16489	03264	88525	42786	05269	92532
69	44657	66999	99324	51281	84463	60563	79312	93454	68876	25471	93911	25650	12682	73572
70	91340	84979	46949	81973	37949	61023	43997	15263	80644	43942	89203	71795	99533	50501
71	91227	21199	31935	27022	84067	05462	35216	14486	29891	68607	41867	14951	91696	85065
72	50001	38140	66321	19924	72163	09538	12151	06878	91903	18749	34405	56087	82790	70925
73	65390	05224	72958	28609	81406	39147	25549	48542	42627	45233	57202	94617	23772	07896

სტრ./სვეტი	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
74	27504	96131	83944	41575	10573	08619	64482	73923	36152	05184	94142	25299	84387	34925
75	37169	94851	39117	89632	00959	16487	65536	49071	39782	17095	02330	74301	00275	48280
76	11508	70225	51111	38351	19444	66499	71945	05422	13442	78675	84081	66938	93654	59894
77	37449	30362	06694	54690	04052	53115	62757	95348	78662	11163	81651	50245	34971	52924
78	46515	70331	85922	38329	57015	15765	97161	17869	45349	61796	66345	81073	49106	79860
79	30986	81223	42416	58353	21532	30502	32305	86482	05174	07901	54339	58861	74818	46942
80	63798	64995	46583	09765	44160	78128	83991	42865	92520	83531	80377	35909	81250	54238
81	82486	84846	99254	67632	43218	50076	21361	64816	51202	88124	41870	52689	51275	83556
82	21885	32906	92431	09060	64297	51674	64126	62570	26123	05155	59194	52799	28225	85762
83	60336	98782	07408	53458	13564	59089	26445	29789	85205	41001	12535	12133	14645	23541
84	43937	46891	24010	25560	86355	33941	25786	54990	71899	15475	95434	98227	21824	19585
85	97656	63175	89303	16275	07100	92063	21942	18611	47348	20203	18534	03862	78095	50136
86	03299	01221	05418	38982	55758	92237	26759	86367	21216	98442	08303	56613	91511	75928
87	79626	06486	03574	17668	07785	76020	79924	25651	83325	88428	85076	72811	22717	50585
88	85636	68335	47539	03129	65651	11977	02510	26113	99447	68645	34327	15152	55230	93448
89	18039	14367	61337	06177	12143	46609	32989	74014	64708	00533	35398	58408	13261	47908
90	08362	15656	60627	36478	65648	16764	53412	09013	07832	41574	17639	82163	60859	75567
91	79556	29068	04142	16268	15387	12856	66227	38358	22478	73373	88732	09443	82558	05250
92	92608	82674	27072	32534	17075	27698	98204	63863	11951	34648	88022	56148	34925	57031
93	23982	25835	40055	67006	12293	02753	14827	22235	35071	99704	37543	11601	35503	85171
94	09915	96306	05908	97901	28395	14186	00821	80703	70426	75647	76310	88717	37890	40129
95	50937	33300	26695	62247	69927	76123	50842	43834	86654	70959	79725	93872	28117	19233
96	42488	78077	69882	61657	34136	79180	97526	43092	04098	73571	80799	76536	71255	64239
97	46764	86273	63003	93017	31204	36692	40202	35275	57306	55543	53203	18098	47625	88684
98	03237	45430	55417	63282	90816	17349	88298	90183	36600	78406	06216	95787	42579	90730
99	86591	81482	52667	61583	14972	90053	89534	76036	49199	43716	97548	04379	46370	28672
100	38534	01715	94964	87288	65680	43772	39560	12918	86537	62738	19636	51132	25739	56947

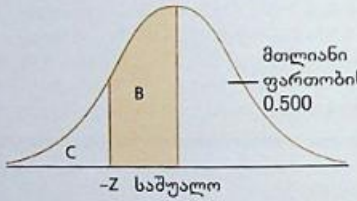
წყარო: William H. Beyer, ed., Handbook for Probability and Statistics, 2nd ed. Copyright © 1966 CRC Press, Boca Raton, Florida.

დანართი ბ
სტანდარტული ნორმალური განაწილების ცხრილი

A სვეტში მოცემულია Z მნიშვნელობები. B სვეტში მოცემულია Z-ის კონკრეტულ მნიშვნელობასა და საშუალო მნიშვნელობას შორის არსებული პროპორციული არე. C სვეტში მოცემულია Z-ის კონკრეტულ მნიშვნელობამდე მოქცეული არე. ცხრილში Z-ის მხოლოდ დადებითი მნიშვნელობებია მოცემული. რადგან ნორმალური განაწილების მრუდი სიმეტრიულია, Z-ის უარყოფითი მნიშვნელობების შესაბამისი არე იგივე იქნება როგორც ეს დადებითი Z-ის მნიშვნელობების შემთხვევაშია.



Z-სა და საშუალო მნიშვნელობას და Z-ის მნიშვნელობამდე მოქცეული არეები (დადებითი Z-ისათვის)



Z-სა და საშუალო მნიშვნელობას და Z-ის მნიშვნელობამდე მოქცეული არეები (უარყოფითი Z-ისათვის)

Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობამდე მოქცეული არე	Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობამდე მოქცეული არე	Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობამდე მოქცეული არე
0.00	0.0000	0.5000	0.11	0.0438	0.4562	0.21	0.0832	0.4168
0.01	0.0040	0.4960	0.12	0.0478	0.4522	0.22	0.0871	0.4129
0.02	0.0080	0.4920	0.13	0.0517	0.4483	0.23	0.0910	0.4090
0.03	0.0120	0.4880	0.14	0.0557	0.4443	0.24	0.0948	0.4052

Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობამდე მოქცეული არე	Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობამდე მოქცეული არე	Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობამდე მოქცეული არე
0.04	0.0160	0.4840	0.15	0.0596	0.4404	0.25	0.0987	0.4013
0.05	0.0199	0.4801	0.16	0.0636	0.4364	0.26	0.1026	0.3974
0.06	0.0239	0.4761	0.17	0.0675	0.4325	0.27	0.1064	0.3936
0.07	0.0279	0.4721	0.18	0.0714	0.4286	0.28	0.1103	0.3897
0.08	0.0319	0.4681	0.19	0.0753	0.4247	0.29	0.1141	0.3859
0.09	0.0359	0.4641	0.20	0.0793	0.4207	0.30	0.1179	0.3821
0.10	0.0398	0.4602	0.11	0.0438	0.4562	0.21	0.0832	0.4168
0.31	0.1217	0.3783	0.71	0.2611	0.2389	1.11	0.3665	0.1335
0.32	0.1255	0.3745	0.72	0.2642	0.2358	1.12	0.3686	0.1314
0.33	0.1293	0.3707	0.73	0.2673	0.2327	1.13	0.3708	0.1292
0.34	0.1331	0.3669	0.74	0.2703	0.2297	1.14	0.3729	0.1271
0.35	0.1368	0.3632	0.75	0.2734	0.2266	1.15	0.3749	0.1251
0.36	0.1406	0.3594	0.76	0.2764	0.2236	1.16	0.3770	0.1230
0.37	0.1443	0.3557	0.77	0.2794	0.2206	1.17	0.3790	0.1210
0.38	0.1480	0.3520	0.78	0.2823	0.2177	1.18	0.3810	0.1190
0.39	0.1517	0.3483	0.79	0.2852	0.2148	1.19	0.3830	0.1170
0.40	0.1554	0.3446	0.80	0.2881	0.2119	1.20	0.3849	0.1151
0.41	0.1591	0.3409	0.81	0.2910	0.2090	1.21	0.3869	0.1131
0.42	0.1628	0.3372	0.82	0.2939	0.2061	1.22	0.3888	0.1112
0.43	0.1664	0.3336	0.83	0.2967	0.2033	1.23	0.3907	0.1093
0.44	0.1700	0.3300	0.84	0.2995	0.2005	1.24	0.3925	0.1075
0.45	0.1736	0.3264	0.85	0.3023	0.1977	1.25	0.3944	0.1056
0.46	0.1772	0.3228	0.86	0.3051	0.1949	1.26	0.3962	0.1038
0.47	0.1808	0.3192	0.87	0.3078	0.1992	1.27	0.3980	0.1020
0.48	0.1844	0.3156	0.88	0.3106	0.1894	1.28	0.3997	0.1003
0.49	0.1879	0.3121	0.89	0.3133	0.1867	1.29	0.4015	0.0985
0.50	0.1915	0.3085	0.90	0.3159	0.1841	1.30	0.4032	0.0968
0.51	0.1950	0.3050	0.91	0.3186	0.1814	1.31	0.4049	0.0951
0.52	0.1985	0.3015	0.92	0.3212	0.1788	1.32	0.4066	0.0934
0.53	0.2019	0.2981	0.93	0.3238	0.1762	1.33	0.4082	0.0918
0.54	0.2054	0.2946	0.94	0.3264	0.1736	1.34	0.4099	0.0901
0.55	0.2088	0.2912	0.95	0.3289	0.1711	1.35	0.4115	0.0885
0.56	0.2123	0.2877	0.96	0.3315	0.1685	1.36	0.4131	0.0869
0.57	0.2157	0.2843	0.97	0.3340	0.1660	1.37	0.4147	0.0853
0.58	0.2190	0.2810	0.98	0.3365	0.1635	1.38	0.4612	0.0838

Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობაზე მოქცეული არე	Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობაზე მოქცეული არე	Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობაზე მოქცეული არე
0.59	0.2224	0.2776	0.99	0.3389	0.1611	1.39	0.4177	0.0823
0.60	0.2257	0.2743	1.00	0.3413	0.1587	1.40	0.4192	0.0808
0.61	0.2291	0.2709	1.01	0.3438	0.1562	1.41	0.4207	0.0793
0.62	0.2324	0.2676	1.02	0.3461	0.1539	1.42	0.4222	0.0778
0.63	0.2357	0.2643	1.03	0.3485	0.1515	1.43	0.4236	0.0764
0.64	0.2389	0.2611	1.04	0.3508	0.1492	1.44	0.4251	0.0749
0.65	0.2422	0.2578	1.05	0.3531	0.1469	1.45	0.4265	0.0735
0.66	0.2454	0.2546	1.06	0.3554	0.1446	1.46	0.4279	0.0721
0.67	0.2486	0.2514	1.07	0.3577	0.1423	1.47	0.4292	0.0708
0.68	0.2517	0.2483	1.08	0.3599	0.1401	1.48	0.4306	0.0694
0.69	0.2549	0.2451	1.09	0.3621	0.1379	1.49	0.4319	0.0681
0.70	0.2580	0.2420	1.10	0.3643	0.1357	1.50	0.4332	0.0668
1.51	0.4345	0.0655	1.91	0.4719	0.0281	2.31	0.4896	0.0104
1.52	0.4357	0.0643	1.92	0.4726	0.0274	2.32	0.4898	0.0102
1.53	0.4370	0.0630	1.93	0.4732	0.0268	2.33	0.4901	0.0099
1.54	0.4382	0.0618	1.94	0.4738	0.0262	2.34	0.4904	0.0096
1.55	0.4394	0.0606	1.95	0.4744	0.0256	2.35	0.4906	0.0094
1.56	0.4406	0.0594	1.96	0.4750	0.0250	2.36	0.4909	0.0091
1.57	0.4418	0.0582	1.97	0.4756	0.0244	2.37	0.4911	0.0089
1.58	0.4429	0.0571	1.98	0.4761	0.0239	2.38	0.4913	0.0087
1.59	0.4441	0.0559	1.99	0.4767	0.0233	2.39	0.4916	0.0084
1.60	0.4452	0.0548	2.00	0.4772	0.0228	2.40	0.4918	0.0082
1.61	0.4463	0.0537	2.01	0.4778	0.0222	2.41	0.4920	0.0080
1.62	0.4474	0.0526	2.02	0.4783	0.0217	2.42	0.4922	0.0078
1.63	0.4484	0.0516	2.03	0.4788	0.0212	2.43	0.4925	0.0075
1.64	0.4495	0.0505	2.04	0.4793	0.0207	2.44	0.4927	0.0073
1.65	0.4505	0.0495	2.05	0.4798	0.0202	2.45	0.4929	0.0071
1.66	0.4515	0.0485	2.06	0.4803	0.0197	2.46	0.4931	0.0069
1.67	0.4525	0.0475	2.07	0.4808	0.0192	2.47	0.4932	0.0068
1.68	0.4535	0.0465	2.08	0.4812	0.0188	2.48	0.4934	0.0066
1.69	0.4545	0.0455	2.09	0.4817	0.0183	2.49	0.4936	0.0064
1.70	0.4554	0.0466	2.10	0.4821	0.0179	2.50	0.4938	0.0062
1.71	0.4564	0.0436	2.11	0.4826	0.0174	2.51	0.4940	0.0060
1.72	0.4573	0.0427	2.12	0.4830	0.0170	2.52	0.4941	0.0059
1.73	0.4582	0.0418	2.13	0.4834	0.0166	2.53	0.4943	0.0057

Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობაზე მოქცეული არე	Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობაზე მოქცეული არე	Z	Z-სა და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობაზე მოქცეული არე
1.74	0.4591	0.0409	2.14	0.4838	0.0162	2.54	0.4945	0.0055
1.75	0.4599	0.0401	2.15	0.4842	0.0158	2.55	0.4946	0.0054
1.76	0.4608	0.0392	2.16	0.4846	0.0154	2.56	0.4948	0.0052
1.77	0.4616	0.0384	2.17	0.4850	0.0150	2.57	0.4949	0.0051
1.78	0.4625	0.0375	2.18	0.4854	0.0146	2.58	0.4951	0.0049
1.79	0.4633	0.0367	2.19	0.4857	0.0143	2.59	0.4952	0.0048
1.80	0.4641	0.0359	2.20	0.4861	0.0139	2.60	0.4953	0.0047
1.81	0.4649	0.0351	2.21	0.4864	0.0136	2.61	0.4955	0.0045
1.82	0.4656	0.0344	2.22	0.4868	0.0132	2.62	0.4956	0.0044
1.83	0.4664	0.0336	2.23	0.4871	0.0129	2.63	0.4957	0.0043
1.84	0.4671	0.0329	2.24	0.4875	0.0125	2.64	0.4959	0.0041
1.85	0.4678	0.0322	2.25	0.4878	0.0122	2.65	0.4960	0.0040
1.86	0.4686	0.0314	2.26	0.4881	0.0119	2.66	0.4961	0.0039
1.87	0.4693	0.0307	2.27	0.4884	0.0116	2.67	0.4962	0.0038
1.88	0.4699	0.0301	2.28	0.4887	0.0113	2.68	0.4963	0.0037
1.89	0.4706	0.0294	2.29	0.4890	0.0110	2.69	0.4964	0.0036
1.90	0.4713	0.0287	2.30	0.4893	0.0107	2.70	0.4965	0.0035
2.71	0.4966	0.0034	3.01	0.4987	0.0013	3.31	0.4995	0.0005
2.72	0.4967	0.0033	3.02	0.4987	0.0013	3.32	0.4995	0.0005
2.73	0.4968	0.0032	3.03	0.4988	0.0012	3.33	0.4996	0.0004
2.74	0.4969	0.0031	3.04	0.4988	0.0012	3.34	0.4996	0.0004
2.75	0.4970	0.0030	3.05	0.4989	0.0011	3.35	0.4996	0.0004
2.76	0.4971	0.0029	3.06	0.4989	0.0011	3.36	0.4996	0.0004
2.77	0.4972	0.0028	3.07	0.4989	0.0011	3.37	0.4996	0.0004
2.78	0.4973	0.0027	3.08	0.4990	0.0010	3.38	0.4996	0.0004
2.79	0.4974	0.0026	3.09	0.4990	0.0010	3.39	0.4997	0.0003
2.80	0.4974	0.0026	3.10	0.4990	0.0010	3.40	0.4997	0.0003
2.81	0.4975	0.0025	3.11	0.4991	0.0009	3.41	0.4997	0.0003
2.82	0.4976	0.0024	3.12	0.4991	0.0009	3.42	0.4997	0.0003
2.83	0.4977	0.0023	3.13	0.4991	0.0009	3.43	0.4997	0.0003
2.84	0.4977	0.0023	3.14	0.4992	0.0008	3.44	0.4997	0.0003
2.85	0.4978	0.0022	3.15	0.4992	0.0008	3.45	0.4997	0.0003
2.86	0.4979	0.0021	3.16	0.4992	0.0008	3.46	0.4997	0.0003
2.87	0.4979	0.0021	3.17	0.4992	0.0008	3.47	0.4997	0.0003
2.88	0.4980	0.0020	3.18	0.4993	0.0007	3.48	0.4997	0.0003

Z	Z-ს და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობაზე მოქცეული არე	Z	Z-ს და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობაზე მოქცეული არე	Z	Z-ს და საშუალო მნიშვნელობას შორის არე	C Z-ის მნიშვნელობაზე მოქცეული არე
2.89	0.4981	0.0019	3.19	0.4993	0.0007	3.49	0.4998	0.0002
2.90	0.4981	0.0019	3.20	0.4993	0.0007	3.50	0.4998	0.0002
2.91	0.4982	0.0018	3.21	0.4993	0.0007	3.60	0.4998	0.0002
2.92	0.4982	0.0018	3.22	0.4994	0.0006	3.70	0.4999	0.0001
2.93	0.4983	0.0017	3.23	0.4994	0.0006	3.80	0.4999	0.0001
2.94	0.4984	0.0016	3.24	0.4994	0.0006	3.90	0.4999	<0.0001
2.95	0.4984	0.0016	3.25	0.4994	0.0006	4.00	0.4999	<0.0001
2.96	0.4985	0.0015	3.26	0.4994	0.0006			
2.97	0.4985	0.0015	3.27	0.4995	0.0005			
2.98	0.4986	0.0014	3.28	0.4995	0.0005			
2.99	0.4986	0.0014	3.29	0.4995	0.0005			
3.00	0.4986	0.0014	3.30	0.4995	0.0005			

დანართი გ
t განაწილება

	მნიშვნელოვნების დონე ცალმხრივი კრიტერიუმისათვის					
	0.10	0.05	0.025	0.01	0.005	0.0005
	მნიშვნელოვნების დონე ორმხრივი კრიტერიუმისათვის					
df	0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850

	მნიშვნელოვნების დონე ცალმხრივი კრიტერიუმისათვის					
	0.10	0.05	0.025	0.01	0.005	0.0005
	მნიშვნელოვნების დონე ორმხრივი კრიტერიუმისათვის					
df	0.20	0.10	0.05	0.02	0.01	0.001
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

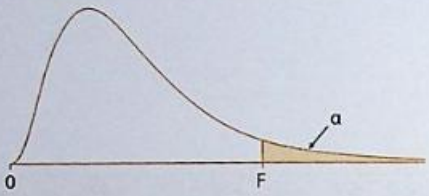
წყარო: Abridged from R. A. Fisher and F. Yates, Statistical Tables for Biological, Agricultural and Medical Research, 6th ed. Copyright © R. A. Fisher and F. Yates 1963.

დანართი დ
ხი-კვადრატ განაწილება

df	0.99	0.98	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01	0.001
1	0.03157	0.03628	0.00393	0.0158	0.064	0.14	0.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	0.0201	0.0404	0.103	0.211	0.466	0.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	0.115	0.185	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341	16.268
4	0.297	0.429	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.465
5	0.554	0.752	1.145	2.204	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.517
6	0.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.457
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	5.812	7.962	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	8.672	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	7.015	9.390	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	7.633	10.117	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	8.260	10.851	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	8.897	11.591	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	32.671	38.932	46.797
22	9.542	12.338	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	33.924	40.289	48.268
23	10.196	13.091	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	35.172	41.638	49.728
24	10.856	13.848	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	36.415	42.980	51.620
25	11.524	14.611	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	37.652	44.314	52.620
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	12.879	14.125	16.151	18.144	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.302
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703

წყარო: R. A. Fisher & F. Yates, Statistical Tables for Biological, Agricultural and Medical Research, 6th ed. Copyright © R. A. Fisher and F. Yates 1963.

დანართი ე
F განაწილება



$\alpha = 0.05$

df^2	df^1 1	2	3	4	5	6	8	12	24	∞
1	161.40	199.50	215.70	224.60	230.20	234.00	238.90	243.90	249.00	254.30
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64

$\alpha = 0.05$

df^2 \ df^1	1	2	3	4	5	6	8	12	24	∞
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

$\alpha = 0.01$

df^2 \ df^1	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98.49	99.01	99.17	99.25	99.30	99.33	99.36	99.42	99.46	99.50
3	34.12	30.81	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.27	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	3.91
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.08	2.65
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.23	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
∞	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

დანართი 3
საბაზისო მათემატიკის მიმოხილვა

ალბათ ყველას გსმენიათ, რომ სტატისტიკა ბევრი მათემატიკის ცოდნას მოითხოვს, ამ მიზეზის გამო სტატისტიკის კურს ბევრი შიშით უყურებს. მიუხედავად აღნიშნულის სისწორისა სტატისტიკის კურსის თავისებურება ის არის, რომ იგი უფრო მეტად ძირითადად მათემატიკური ცოდნის ინტერპრეტაციას გულისხმობს ვიდრე რთული მათემატიკური ამოცანების ამოხსნას. მიუხედავად ამისა, სტატიტიკის კურსის შესწავლისათვის უნდა იცოდეთ საბაზისო მათემატიკური ოპერაციები და გესმოდეთ ცალკეული სიმობოლების შინაარსი. ქვემოთ განვიხილავთ მათემატიკის იმ საბაზისო საკითხებს, რაც ამ კურსის წარმატებით სწავლაში დაგეხმარებათ.

სტატისტიკაში გამოყენებული სიმბოლოები და გამოსახულებები

სტატისტიკა გვაძლევს მონაცემთა აღწერისა და ანალიზის მეთოდებს. ცვლადი არის სიდიდე, რომლის მნიშვნელობა ცვალებადობს გარკვეული გზით. მაგალითად ცვლადი სიდიდეა პირის ასაკი, რადგან მისი მნიშვნელობები შესაძლებელია შეიცვალოს დაბადების ასაკიდან ას წლამდე. ასევე ცვლადებია „რასა“ და „სქესი“, მიუხედავად იმისა, რომ ეს ცვლადები უფრო ნაკლებად მრავალფეროვან მნიშვნელობებს იღებენ ასაკთან შედარებით. სტატისტიკაში ამ თუ იმ ცვლადთა წყების აღსანიშნავად იყენებენ სიმბოლოებს. მაგალითად, თუ გვინდა შევისწავლოთ სტუდენტების ასაკის რაიმე ასპექტი ჩვენს სტატისტიკის კლასში, ჩვენ გამოვიყენებთ Y-ს ცვლად „ასაკის“ აღსანიშნავად. სიმარტივისათვის დავუშვათ, რომ მხოლოდ პირველ რიგში მჯდარ სტუდენტებს დავუსვით ასაკის შესახებ კითხვა და მივიღეთ შემდეგი პასუხები – 19,21, 23 და 22. ეს ოთხი სიდიდე იქნება Y ცვლადის მნიშვნელობები.

სხვა სიმბოლო, რომელიც ყველაზე ხშირად გვხვდება სტატისტიკაში არის $\sum$ სიგმა. სიგმა არის ბერძნული ასო და სტატისტიკაში ჯამს აღნიშნავს. ანუ თუ რაიმე გამოსახულების წინ დგას $\sum$ სიმბოლო ეს ნიშნავს, რომ უნდა დაჯამოთ სიმბოლოს ქვეშ მოქცეული გამოსახულების მნიშვნელობები. მაგალითად, სტუდენტების ჩვენი შერჩევისათვის, რომელიც აღნიშნული გვაქვს Y-ით, სიგმას სიმბოლოს გამოყენება $\sum Y$ ნიშნავს Y-ის ყველა მნიშვნელობის შეჯამებას. ჩვენი მაგალითის მიხედვით ამ ჯამს ვიპოვით ცნობილი მნიშვნელობების შეკრებით:

$$19 + 21 + 23 + 32 = 95$$

ასე, რომ ცვლად „ასაკისათვის“ $\sum Y=95$

სიგმა ასევე გამოიყენება კვადრატში აყვანის სიმბოლოსთან ერთად, მაგალითად როგორც ეს ამ გამოსახულებაშია $\sum Y^2$ (იკითხება როგორც: მნიშვნელობების კვადრატების ჯამი). ეს ნიშნავს რომ ჯერ უნდა ავიყვანოთ კვადრატში Y-ის თითოეული მნიშვნელობა და შემდეგ ვიპოვოთ მათი ჯამი:

$$192 + 212 + 232 + 322 = 361 + 441 + 529 + 1.024 = 2.355$$

ასე, რომ ცვლად „ასაკისათვის“, $\sum Y^2 = 2.355$.

მსგავსი, მაგრამ მცირედით განსხვავებული გამოსახულებაა $(\sum Y)^2$, იკითხება როგორც: „მნიშვნელობათა ჯამის კვადრატი“. ამ გამოსახულებაში ფრჩხილების შემოტანა ნიშნავს, რომ ჯერ უნდა შეიკრიბოს მნიშვნელობები და ამის შემდეგ ავიდეს კვადრატში. ფრჩხილები საკმაოდ ხშირად გამოიყენება სტატისტიკაში და ნიშნავს, რომ ჯერ ფრჩხილებში მოთავსებული მოქმედება უნდა შესრულდეს და შემდეგ მის გარეთ დაწერილი მოქმედება. მოყვანილი გამოსახულების გამოსათვლელად პირველად შევაჯამოთ ცვლადის მნიშვნელობები. უკვე ნაპოვნი გვაქვს $\sum Y = 95$, ასე რომ $(\sum Y)^2$ საპოვნელად ავიყვანოთ კვადრატში 95:

$$95^2 = 9.025$$

ასე, რომ ცვლად „ასაკისათვის“ $(\sum Y)^2 = 9.025$.

თქვენ ასევე უნდა იცნობდეთ სხვადასხვა სიმბოლოებს, რომლებიც გამრავლებასა და გაყოფას აღნიშნავენ. გამრავლების სიმბოლო (×) ასევე შესაძლებელია სხვაგვადარაც ჩაიწეროს:

$$3(4) \quad (5)6 \quad (4)(2) \quad 7 \cdot 8 \quad 9 * 6$$

ამ ჩანაწერთაგან ამ წიგნში გამრავლების აღსანიშნავად ყველაზე ხშირად გამოიყენება პირველი სამი ნიშანი. ასევე არსებობს გაყოფის აღმნიშვნელი რამდენიმე სიმბოლო. მაგალითად, ყველაზე გავრცელებულია (÷), გაყოფა ასევე შესაძლებელია ჩაიწეროს:

$$4/6\frac{6}{3}$$

ამ ტექსტში უფრო ხშირად ბოლო ორი სიმბოლო გამოიყენება.

სტატისტიკაში ასევე შეგხვდებათ მეტობა-ნაკლებობის ნიშანი ($>, <$), მეტია ან ტოლი და ნაკლებია ან ტოლობის ნიშნები ($\geq, \leq$) და უტოლობის ნიშანი ($\neq$). მნიშვნელოვანია რომ ყველა ამ სიმბოლოს მნიშვნელობა გესმოდეთ და ისინი ერთმანეთში არ აურიოთ. გამოიყენეთ ქვემოთ მოტანილი გამოსახულებები თქვენი ცოდნის განსამტკიცებლად:

$4 > 2$ ნიშნავს, რომ 4 მეტია ორზე

$H_1 > 10$ ნიშნავს, რომ H_1 მეტია 10-ზე

$7 < 9$ ნიშავს, რომ 7 ნაკლებია 9-ზე

$a < b$ ნიშნავს, რომ a ნაკლებია b-ზე

$Y \geq 10$ ნიშნავს, რომ Y მეტია ან ტოლია 10-ზე

$a \leq b$ ნიშნავს, რომ a ნაკლებია ან ტოლია b-ზე.

$8 \neq 10$ ნიშნავს, რომ 8 არ უდრის 10

$H_1 \neq H_2$ ნიშნავს, რომ H_1 არ უდრის H_2-ს.

პროპორციები და პროცენტები

პროპორციები და პროცენტები ხშირად გამოიყენება ცალკეული მნიშვნელობების სიხშირეების გამოხატვისათვის. სტუდენტმა უნდა იცოდეს როგორ იპოვოს პროპორციები და პროცენტები.

ფარდობა ჩაიწერება P-თი და მის საპოვნელად გამოიყენება შემდეგი ფორმულა:

$$P = \frac{f}{N}$$

სადაც f აღნიშნავს ცალკეული შემთხვევების რიცხვს მთლიანობაში, ხოლო N აღნიშნავს შემთხვევათა მთლიან რიცხვს. მაგალითად ჩვენს მაგალითში თუ გვსურს გავიგოთ წინა რიგში მამაკაცი სტუდენტების წილი, გვექნება ორი კატეგორია ქალი და კაცი. რადგან შერჩევაში გვაქვს 3 ქალი და 1 კაცი, $N = 4$, და შემთხვევის რაოდენობა მამაკაცებისათვის არის 1. ფარდობითი წილი იქნება 1-ის 4-თან ფარდობა:

$$P = \frac{f}{N} \quad P = \frac{1}{4} = .25$$

ასე, რომ მამაკაცების ფარდობითი წილი წინა რიგში არის 0.25. ამ მაჩვენებლის პროცენტულ ფორმაში გადაყვანისათვის ეს მნიშვნელობა გავამრავლოთ **100**-ზე ან გამოიყენეთ პროცენტების ქვემოთ მოტანილი ფორმულა:

$$\% = \frac{1}{N} \times 100 \quad \% = \frac{1}{N} \times 100 = 25\%$$

უარყოფით რიცხვებთან მუშაობა

გამოკლება და მიმატება, გამრავლება, გაყოფა და რიცხვის კვადრატში აყვანა უმრავლესობათვის არ წარმოადგენს სირთულეს, მაგრამ მნიშვნელოვანია გახსოვდეთ ამ არითმეტიკული ოპერაციების ჩატარებისას უარყოფითი რიცხვების თავისებურებები:

1. უარყოფითი რიცხვის დამატება იგივე რიცხვის გამოკლების ტოლფასია:

$$5 + (-2) = 5 - 2 = 3$$

2. უარყოფითი რიცხვის გამოკლებისას იცვლება ნიშანი

$$8 - (-4) = 8 + 4 = 12$$

3. დადებითი რიცხვის უარყოფით რიცხვზე გამრავლების ან გაყოფისას შედეგი ყოველთვის უარყოფითია:

$$6\times - 4 = -24, -10 \div 5 = -2$$

4. ორი უარყოფითი რიცხვის გამრავლების ან გაყოფისას განაყოფი ან ნამრავლი დადებითია:

$$-3\times - 7 = 21, -12 \div -4 = 3$$

5. უარყოფითი რიცხვის კვადრატში აყვანა დადებით შედეგს იძლევა, რადგან კვადრატში აყვანა ამ შემთხვევაში ორი უარყოფითი რიცხვის ერთმანეთზე გამრავლებას ნიშნავს:

$-52 = 25$ მსგავსია, რაც $-5\times -5 = 25$

გამოთვლების მიმდევრობა და კომპლექსური გამოსახულებები

სტატისტიკაში შეხვდებით საკმაოდ გრძელ გამოსახულებსაც, რომელიც გამოთვლების რამდენიმე ეტაპს მოითხოვენ. იმის ცოდნისათვის თუ ჯერ რომელი მოქმედება უნდა შეასრულოთ დაიცავით შემდეგი ორი ძირითადი წესი: პირველი წესი იწოდება თანმიმდევრობის წესად, რომლის მიხედვითაც ჯერ უნდა მოხდეს კვადრატში აყვანა და ფესვის ამოღება, შემდეგ გამრავლება და გაყოფა და საბოლოდ მიმატება და გამოკლება მარცხნიდან მარჯვნისაკენ მიმართულებით. თუ გამოსახულებაში არის ფრჩხილები, მაშინ პირველად ფრჩხილებში მოცემული გამოთვლები უნდა შესრულდეს და შემდეგ დანარჩენი მოქმედებები. ეს ნიშნავს, რომ ფრჩხილების გამოყენებას შეუძლია შეცვალოს გამოთვლათა ტრადიციული მიმდევრობა. ეს წესები თავდაპირველად შესაძლოა აბსტრაქტულად გეჩვენოთ, მაგრამ მაგალითის განხილვა მას უფრო ცხადს გახდის:

$$4 + 6 \times 8 = 4 + 48 = 52$$

პირველად შეასრულეთ გამრავლების და შემდეგ შეკრების ოპერაცია. თუ არ გავითვალისწინებდით გამოთვლების თანმიმდევრობის წესს, მაშინ განსხვავებულ შედეგს მივიღებდით:

$$4 + 6 \times 8 = 10 \times 8 = 80$$

რაც არასწორია.

შემდეგი გამოსახულების მნიშვნელობის პოვნისათვის:

$$6 - 4(6)/3^2$$

პირველად ვიპოვოთ 3-ის კვადრატი:

$$6 - 4(6)/9$$

შემდეგ შევასრულოთ გამრავლებისა და გაყოფის ოპერაციები მარცხნიდან მარჯვნივ:

$$6 - \frac{24}{9} = 6 - 2.67$$

საბოლოოდ, შეასრულეთ გამოკლების ოპერაცია:

$$6 - 2.67 = 3.33$$

შემდეგი გამოსახულების მნიშვნელობის პოვნისათვის, პირველად შეასრულეთ ფრჩხილებში მოცემული მოქმედება:

$$(4 + 3) - 6(2)/(3 - 1)^2$$

ფრჩხილებში მოცემული მიმატებისა და გამოკლების ოპერაციების შემდეგ მივიღებთ:

$$(7) - 6(2)/(2)^2$$

ფრჩხილებში მოცემული გამოთვლების დასრულების შემდეგ მიყევით გამოთვლათა სტანდარტულ თანმიმდევრობას, გამოთვალეთ კვადრატის მნიშვნელობა:

$$(7) - 6(2)/4$$

შემდეგ შეასრულეთ გამრავლებისა და გაყოფის ოპერაციები:

$$(7) - \frac{12}{4} = (7) - 3$$

საბოლოოდ შეასრულეთ გამოკლების ოპერაცია:

$$7 - 3 = 4$$

ქვემოთ მოყვანილი გამოსახულება შედარებით რთული ჩანს, მაგრამ თუ მივყვებით გამოთვლის წესებს გამოთვლის შედეგებსაც მარტივად მივიღებთ:

$$\sqrt{(8(4 - 2)^2) / (12 / 4)^2}$$

პირველად შეასრულეთ ფრჩხილებში მოცემული გამოსახულება. ასევე მიაქციეთ ყურადღება, რომ თავად ფრჩილებშია ასევე მოცემულია სხვა ფრჩხილები. ამ შემთხვევაში თავდაპირველად შიგა ფრჩხილებში მოცემული გამოსახულების მნიშვნელობა იპოვეთ:

$$\sqrt{(8(2)^2) / 3^2}$$

შემდეგ შეასრულეთ გარეთა ფრჩხილებში მოცემული მოქმედება, თავად ფრჩხილის შიგნით ჯერ კვადრატში აიყვანეთ და შემდეგ გაამრავლეთ:

$$\sqrt{\frac{32}{3^2}}$$

აიყვანეთ მნიშვნელი კვადრატში და შეასრულეთ გაყოფის ოპერაცია:

$$\sqrt{\frac{32}{9}} = \sqrt{3.55}$$

გამოთვალეთ კვადრატული ფესვი:

1.88

ამოხსნები რუბრიკისათვის „ცოდნის შემოწმება“

თავი 4

(გვ. 109)

ცოდნის შემოწმება. *ქვემოთ მოცემულია 15 პიროვნების განაწილება პოლიტიკურ პარტიებში. იპოვეთ მოდა.*

დემოკრატი	რესპუბლიკელი	დემოკრატი	რესპუბლიკელი	რესპუბლიკელი
დამოუკიდებელი	დემოკრატი	დემოკრატი	დემოკრატი	რესპუბლიკელი
დამოუკიდებელი	დემოკრატი	დამოუკიდებელი	რესპუბლიკელი	დემოკრატი

პასუხი:

მოდა არის „დემოკრატი“, რადგან ამ კატეგორიას აქვს ყველაზე მაღალი სიხშირე- 7.

(გვ. 114)

ცოდნის შემოწმება. *იპოვეთ ინტერვალურ-შეფარდებითი ცვლადის განაწილების მედიანა: 22, 15, 18, 33, 17, 5, 11, 28, 40, 19, 8, 20.*

პასუხი:

თავდაპირველად ჩვენ დავალაგოთ ნომრები შემდეგნაირად: 5, 8, 11, 15, 17, 18, 19, 20, 22, 28, 33, 40. $(N + 1) / 2 = (12 + 1) / 2 = 6.5$. ამ შემთხვევაში მედიანა არის მეექვსე (18) და მეშვიდე (19) ნიშნების საშუალო არითმეტიკული.

$$\text{საშუალო} = \frac{18+19}{2} = 18.5$$

ე.ი. მედიანა უდრის 18.5-ს.

(გვ. 116)

ცოდნის შემოწმება. *ნახაზზე 4.4 დააკვირდით განსხვავებებს მამაკაცებისა და ქალების შემოსავლებში. უკანასკნელი სამი ათწლეულის მანძილზე. რა დასკვნა შეიძლება გამოვიტანოთ აქედან?*

პასუხი:

1970-იანი წლებიდან მოყოლებული ქალთა საშუალო შემოსავლები იზრდებოდა უფრო სწრაფად ვიდრე მამაკაცების. მაშინ როდესაც მამაკაცების საშუალო შემოსავალი გაიზარდა 13.5%-ით 2008 წელს 1873 წელთან შედარებით, იგივე წლებისათვის ქალების საშუალო შემოსავლები გაიზარდა 40%-ზე მეტად. მიუხედავად ამისა ამერიკის შეერთებულ შტატებში საშუალოდ მამაკაცების შემოსავლები აღემატება ქალების შემოსავლებს, რაც ბადებს აზრს რომ სქესსა და საშუალო შემოსავლებს შორის არსებობს კავშირი.

(გვ. 122)

ცოდნის შემოწმება. მოცემულია იგივე განაწილება, რაც მედიანის ცოდნის შემოწმების დროს 21, 32, 23, 41, 20, 30, 36, 22, 25, 27. გაიანგარიშეთ საშუალო. ემთხვევა თუ არა ერთმანეთს მედიანა და საშუალო?

პასუხი:

$$\text{საშუალო} = \frac{22+15+18+33+17+5+11+28+40+19+8+20}{12} = 19.67$$

ე.ი. საშუალო მნიშვნელობა 19.67 აღემატება მედიანის მნიშვნელობას 18.5-ს.

თავი 5

(გვ. 151)

ცოდნის შემოწმება. დაკვირვებით შეისწავლეთ „დეტალების განხილვა 5.1" და განიხილეთ კატეგორიების რაოდენობის გავლენა IQV-ზე. რა მოხდება, თუ ბერკლის მაგალითში აზიელებს ორ კატეგორიად დავყოფთ: 19% ჩინელი ამერიკელები და 21% – დანარჩენი აზიელები. (ამ კითხვაზე პასუხის გასაცემად თავიდან უნდა გააკეთოთ გაანგარიშებები ამ ახალი მონაცემებით).

პასუხი:

თუ აზიელებს გავყოფდით ორ კატეგორიად, რაც შეადგენდა ბერკლის სტუდენტების მთლიანი რაოდენობის 19%-სა და 21%-ს 2009 წლისათვის, მივიღებდით 6 რასობრივ/ეთნიკურ ჯგუფს, ნაცვლად ხუთისა. შესაბამისად IQV შემდეგნაირად გამოითვლება:

$$IQV_{2009} = \frac{K(100^2 - \sum Pct^2)}{100^2(K-1)} = \frac{6(100^2 - 2{,}058)}{100^2(6-1)} = \frac{47{,}652}{50{,}000} = 0.95$$

კატეგორიების რაოდენობის ზრდასთან ერთად გაიზარდა IQV-ს მნიშვნელობაც 0.89-დან (5 კატეგორიით) 0.95-მდე. ასე რომ სხვა თანაბარ პირობებში, რაც უფრო მეტია კატეგორიების რაოდენობა მით უფრო მაღალია ვარიაციის მნიშვნელობა.

(გვ. 157)

ცოდნის შემოწმება. რატომ არის IQR ვარიაციის უკეთესი საზომი, ვიდრე დიაპაზონი, განსაკუთრებით მაშინ, როდესაც განაწილებაში ექსტრემალური მნიშვნელობებია? ამ კითხვაზე პასუხის გასაცემად იხილეთ ნახაზი 5.4.

პასუხი:

მონაცემთა ექსტრემალური მნიშვნელობები გავლენას ახდენენ დიაპაზონზე ანუ უდიდეს და უმცი-

რეს მნიშვნელობებს შორის სხვაობაზე. შესაბამისად, თუ მონაცემებში მოიძებნება ექსტრემალური მნიშვნელობები (ძალიან დაბალი ან ძალიან მაღალი) დიაპაზონი მონაცემთა განაწილების ზუსტი საზომი არარაა. ხოლო IQR-ზე გავლენა არ აქვს მონაცემთა ექსტრემალურ მნიშვნელობებს. ამიტომაც ის დიაპაზონთან შედარებით ვარიაციის უკეთესი საზომია, როცა მონაცემებში მოიპოვება ექსტრემალური მნიშვნელობები.

(გვ. 165)

ცოდნის შემოწმება. ხელახლა განიხილეთ ცხრილი 5.8 და განიხილეთ საშუალოდან გადახრების კვადრატების ჯამში დასავლეთ რეგიონის მიერ შეტანილი არაპროპორციული წილი (კვადრატების ჯამში მისი ხვედრითი წილი დაახლოებით 45%-ია). შეგიძლიათ თუ არა ამის ახსნა? (მინიშნება: ამას გარკვეული კავშირი აქვს საშუალოს მგრძნობიარობასთან ექსტრემალური სიდიდეების მიმართ).

პასუხი:

დასავლეთის რეგიონებში ფიქსირდება ასაკოვანი მოსახლეობის მაღალი პროცენტული ცვლილება 2008 და 2015 წლების პროგნოზის მიხედვით, რაც 27%-ს შეადგენს. შესაბამისად ამ რეგიონის მაჩვენებლის გადახრა საშუალო მნიშვნელობიდან ყველაზე მაღალია. რაც უფრო მაღალია ცვლადის მნიშვნელობის გადახრა საშუალო მაჩვენებლიდან მით უფრო მაღალია კვადრატული გადახრაც და შესაბამისად ამგვარ ცვლადს უფრო დიდი წილი შეაქვს საშუალოდან გადახრების კვადრატში.

თავი 6

(გვ. 193)

ცოდნის შემოწმება. ცხრილში 6.2 Z ქულები უკან გადაიყვანეთ ნედლ ქულებში. პასუხები უნდა დაემთხვეს ცხრილში მოცემულ ნედლ ქულებს.

Z ქულები	ნედლი ქულები
−2.93	$Y = 70.07 - 2.93(10.27) = 40$
−1.95	$Y = 70.07 - 1.95(10.27) = 50$
−0.98	$Y = 70.07 - 0.98(10.27) = 60$
−0.01	$Y = 70.07 - 0.01(10.27) = 70$
0.97	$Y = 70.07 - 0.97(10.27) = 80$
1.94	$Y = 70.07 - 1.94(10.27) = 90$
2.91	$Y = 70.07 - 2.91(10.27) = 100$

პასუხი:

(გვ. 204)

ცოდნის შემოწმება. თავში 4 ჩვენ ვისწავლეთ პერცენტილების განაზღვრა განაწილების კუმულაციური ხვედრითი წილებით. ნახეთ ცხრილი 6.1 და იპოვეთ 92-ე პერცენტილი. განსხვავდება თუ არა თქვენი პასუხი ადრინდელისგან (საშუალოზე მაღალი პერცენტილური რანგის პოვნა)? პასუხი დაასაბუთეთ.

პასუხი:

ცხრილი 6.1-ის მიხედვით 92-ე პერცენტილი არის 80 რადგან სტუდენტების 92% ხვდება ამ კატეგორიაში (იხილეთ 80-ის კუმულაციური პერცენტილები). მაგრამ თუ გამოვთვლით 80-ის პერცენტილურ რანგს ჩვენ 92-ის ნაცვლად მივიღებთ 83.4-ს:

$$Z = \frac{(80 - 70.07)}{10.27} = 0.97$$

Z-ის 0.97 მნიშვნელობის ქვემოთ მდებარე არე არის 0.1660. ასე რომ პერცენტილური რანგი = (1 - 0.1660) 100 = 83.4, რაც ნიშნავს, რომ სოციალური სტატისტიკის კურსის სტუდენტების 83.4%-მა მიიღო 80 ქულაზე ნაკლები. ცხრილ 6.1-ის მიხედვით გამოთვლილი 80-ის პერცენტილური რანგი განსხვავდება ზემოთ გამოთვლილისაგან რადგან პირველი მეთოდი ეფუძნება კუმულაციურ პროცენტებს, ხოლო მეორე ეფუძნება საშუალოსა და სტანდარტულ გადახრას პერცენტილების მდებარეობის დასადგენად.

(გვ. 205)

ცოდნის შემოწმება. რამდენი სტანდარტული გადახრა შეესაბამება 95-ე პერცენტილს ნორმალურ განაწილებაში? თუ არ გაქვთ პასუხი, თავიდან გადაათვალიერეთ თავის ეს ნაწილი.

პასუხი:

საშუალო მნიშვნელობიდან სტანდარტული გადახრების რიცხვი არის რასაც ჩვენ ვეძახით Z მნიშვნელობებს. Z მნიშვნელობება 95-ე პერცენტილისათვის არის 1.65. ე. ი. 95 პერცენტილისათვის საშუალოდან სტანდარტული გადახრის მნიშვნელობა 1.65-ია.

(გვ. 206)

ცოდნის შემოწმება. რომელი ნედლი ქულა შეესაბამება 50-ე პერცენტილს?

პასუხი:

50-ე პერცენტილისათვის ნედლის ქულა არის მონაცემთა მედიანა.

თავი 7

(გვ. 239)

ცოდნის შემოწმება. დაუშვათ, მთლიანი ერთობლიობის საშუალო $\mu_Y = 150$, ხოლო სტანდარტული გადახრა – $\sigma_Y = 30$. თქვენ შეადგინეთ შემთხვევითი შერჩევა $N = 100$ დაკვირვებით. რა იქნება იმის ალბათობა, რომ საშუალო 147-სა და 153-ს შორის იქნება? რა იქნება იმის ალბათობა, რომ საშუალო 153-ს გადასცდება? გაგიკვირდებათ, თუ საშუალო 159-ის ტოლი იქნება? რატომ? (შეხსენება: ამ კითხვებზე პასუხის გასაცემად იხ. თავი 6, Z ქულები და არეები ნორმალური მრუდის ქვეშ, დანართი ბ). ნედლი მნიშვნელობის Z ქულაში გადასაყვანად ვიყენებთ ფორმულას:

$$Z = \frac{Y - \overline{Y}}{S_Y}$$

რადგან აქ შერჩევით განაწილებას ვიხილავთ, Y შეანაცვლეთ $\overline{Y}$-ით, $\overline{Y}$ შერჩევითი განაწილების საშუალოთი და საშუალოს სტანდარტული შეცდომით $\sigma_Y / \sqrt{N}$

პასუხი:

147-ის შესაბამისი Z ქულა არის:

$$Z = \frac{\overline{Y} - \mu_{\overline{Y}}}{\sigma_Y / \sqrt{N}} = \frac{147 - 150}{30 / \sqrt{N}} = \frac{-3}{3} = -1$$

153-ის შესაბამისი Z ქულა არის:

$$Z = \frac{\overline{Y} - \mu_{\overline{Y}}}{\sigma_Y / \sqrt{N}} = \frac{153 - 150}{30 / \sqrt{100}} = \frac{3}{3} = 1$$

სტანდარტული განაწილების ცხრილის გამოყენებით (დანართი ბ) ვპოულობთ, რომ ალბათობა საშუალოსა და 1-ის ტოლი სტანდარტული გადახრის ზემოთ ან ქვემოთ მდებარე არესათვის არის 0.3413. ასე, რომ ალბათობა იმისა, რომ საშუალო მნიშვნელობა არის 147-სა და 153 შორის, 1-ის ტოლი სტანდარტული გადახრით, არის 0.6826 (0.3413 + 0.3413), ან 68.26%.

ალბათობა საშუალოსა და ერთის ტოლის სტანდარტული გადახრის მქონე არესათვის არის 0.1587. შესაბამისად, ალბათობა იმისა, რომ საშუალო გადააჭარბებს 153-ს არის 0.1587 ან 15.87%.

159-ის შესაბამისი Z ქულა არის:

$$Z = \frac{\overline{Y} - \mu_{\overline{Y}}}{\sigma_Y / \sqrt{N}} = \frac{159 - 150}{30 / \sqrt{100}} = \frac{9}{3} = 3$$

ალბათობა 3-ის ტოლი სტანდარტული გადახრის არესათვის, ნორმალური განაწილების ცხრილის მიხედვით არის 0.0014. შესაბამისად, შესაძლოა მოულოდნელი იყოს, რომ 159-ის შესაბამისი საშუალო მნიშვნელობის მიღების ალბათობა ძალიან დაბალია (0.14%).

თავი 8

(გვ. 252)

ცოდნის შემოწმება. რა განსხვავებაა წერტილოვან და ინტერვალურ შეფასებებს შორის?

პასუხი:

როდესაც პოპულაციის პარამეტრის შეფასება ერთი მაჩვენებელია, იგი წერტილოვნად იწოდება. როდესაც შეფასება მნიშვნელობების დიაპაზონია, იგი ინტერვალურ შეფასებად იწოდება. ინტერვალური შეფასებებისათვის გამოიყენება ნდობის ინტერვალები.

(გვ. 254)

ცოდნის შემოწმება. ნდობის ხარისხსა და Z-ს შორის ურთიერთობის გასარკვევად დაუბრუნდით მე-6 თავს. რა იქნება 98%-ის შესაბამისი Z ქულა?

პასუხი:

98 %-იანი ნდობის ინტერვალის შესაბამისი Z მნიშვნელობა არის 2.33.

(გვ. 255)

ცოდნის შემოწმება. რა იქნება 90%-იანი ნდობის ინტერვალი მგზავრობის საშუალო ხანგრძლივობისთვის? (შეხსენება: პირველად მოძებნეთ ნდობის ხარისხის შესაბამისი Z ქულა)

პასუხი:

$$\begin{aligned} 90\%(\text{ნდობის ინტერვალი}) &= 7.5 \pm 1.65(0.07) \\ &= 7.5 \pm -0.12 \\ &= 7.38 - 7.62 \end{aligned}$$

(გვ. 260)

ცოდნის შემოწმება. რატომ იწვევს შერჩევის ზომის ზრდა ნდობის ინტერვალების გადიდებას? (იხ. ნახაზი 8.5) (შეხსენება: შეადარეთ საშუალოს სტანდარტული შეცდომა სამი სხვადასხვა ზომის შერჩევისთვის).

პასუხი:

შერჩევის ზომის შემცირებასთან ერთად, საშუალოს სტანდარტული შეცდომა იზრდება, რაც თავის მხრივ აისახება უფრო დიდ ნდობის ინტერვალში.

(გვ. 267)

ცოდნის შემოწმება. გაიხსენეთ გელაპის მონაცემებისთვის (Gallup Survey) ნდობის ინტერვალი არა პროპორციების, არამედ პროცენტების გამოყენებით. თქვენ ნახავთ, რომ შედეგები იდენტური იქნება, მხოლოდ გამოსახული იქნება პროცენტებში.

პასუხი:

$$\text{სტანდარტული შეცდომა} = \sqrt{(51)(100-51)/1,013} = 1.57$$

$$\begin{aligned} 95\%(\text{ნდობის ინტერვალი}) &= 51 \pm 1.96(1.57) \\ &= 51 \pm 3.08 \\ &= 47.92 - 54.08 \end{aligned}$$

ნდობის ინტერვალის გამოთვლა პროცენტების მიხედვით თითქმის მსგავსია პროპორციების მიხედვით ნდობის ინტერვალის გამოთვლის, გარდა იმ მცირე განსხვავებისა, რაც რიცხვების დამრგვალებითაა გამოწვეული.

თავი 9

(გვ. 307)

ცოდნის შემოწმება. შეცვლით თუ არა თქვენს გადაწყვეტილებას წინა დავალებაში თუ ალფას მნიშვნელობას 0.05-ის ტოლად აირჩევდით?

ცხრილი 9.4-ში მოცემული მონაცემების გამოყენებით, შეამოწმეთ შემდეგი ნულოვანი ჰიპოთეზის მართებულობა: მამაკაცი სტუდენტის მიერ ალკოჰოლის პირველადი მოხმარება უფრო დაბალ აკადემიურ საშუალო შეფასებაში აისახება, ვიდრე ეს ქალი სტუდენტების შემთხვევაშია. ალფა აირჩიეთ 0.05-ის ტოლად. რა დასკვნას გამოიტანთ?

პასუხი:

წიგნის C დამატების მიხედვით, t-ს გამოთვლილი მნიშვნელობა -1.83-ისათვის აღემატება ვიდრე ეს t = -1.83-ის შემთხვევაშია (α = 0.05, ცალმხრივი ტესტი). შეგვიძლია ვთქვათ, რომ -1.83-ის ალბათობა 0.05-ზე ნაკლებია, მოხდება ნულოვანი ჰიპოთეზის უარყოფა.

ალკოჰოლის პირველადი მოხმარებისათვის გამოთვლილი t-ს მნიშვნელობა -0.70-ია. t-ს გამოთვლი-

ლი მნიშვნელობა არის -0.70. გამოთვლილი t-ს მნიშვნელობა ნაკლებია -1.645-ზე. შესაბამისად არ მოხდება ნულოვანი ჰიპოთეზის უარყოფა.

(გვ. 309)

ცოდნის შემოწმება. შეიცვლებოდა თუ არა თქვენი გადაწყვეტილება t-ს გამოთვლისას თუ ალფას მნიშვნელობას 0.01-ის ტოლად აირჩევდით? პასუხი დაასაბუთეთ.

პასუხი:

თუ ალფას მნიშვნელობა 0.01-ის ტოლი იქნებოდა, ეს არ შეცვლიდა ჩვენს გადაწყვეტილებას, არ მოხდებოდა ნულოვანი ჰიპოთეზის უარყოფა.

თავი 10

(გვ. 325)

ცოდნის შემოწმება. განიხილეთ ცხრილი 10.2, დარწუნდით რომ შეგიძლიათ ყველა იმ წყვილის ტიპის განსაზღვრა, რომელიც ამ თავში განვიხილეთ და ასევე იცით თუ როგორ იქნა მიღებული ცალკეული მაჩვენებლები. განსაზღვრეთ ცხრილში დამოკიდებული და დამოუკიდებელი ცვლადები. ამისათვის დაგჭირდებათ სიხშირეების პროცენტებად გარდაქმნა.

პასუხი:

დამოუკიდებელი ცვლადია „რასა“, ხოლო დამოკიდებული ცვლადია კერძო სახლის ფლობა.

თავი 11

(გვ. 373)

ცოდნის შემოწმება. ხი-კვადრატ კრიტერიუმის გამოსათვლელად გამოყენებული მონაცემები, აღებულია კლავ ინმენისა და მეიესის კვლევიდან. განვიხილოთ დამოკიდებულება ასაკსა (დამოუკიდებელი ცვლადი) და პირველი თაობის კოლეჯის სტუდენტის სტატუსს (დამოკიდებული ცვლადი) შორის, ორვანზომილებიანი ცხრილის გამოყენებით.

პირველი თაობის სტატუსი	ასაკი წლების მიხედვით		ჯამი
	19-ის ან 19 წლამდე	20 წლის და 20 წლის ზემოთ	
პირველი თაობის	916 (33.7%)	1,018 (53.6%)	1,934 (41.9%)
სხვა თაობის	1,802 (66.3%)	881 (46.4%)	2,683 (58.1%)
ჯამი (N)	2,718 (100.0%)	1,899 (100.0%)	4,617 (100.0%)

წყარო: ადაპტირებულია W. Elliot Inman and Larry Mayes, "The Importance of being First: Unique Characterisitics of First Generation Community College Students," Coomunity College Review 26, no. 3 (1999): 8.

გვ. 374 – ცოდნის შემოწმება. ააგეთ ორვანზომილებიანი ცხრილი, რომელიც აჩვენებს დამოკიდებულების არარსებობას ასაკისა და პირველი თაობის კოლეჯის სტუდენტის სტატუსს შორის.

პასუხი:

ასაკი და პირველი თაობის კოლეჯის სტუდენტის სტატუსი

	19-ის ან 19 წლამდე	*20 წლის და 20 წლის ზემოთ*	*ჯამი*
პირველი თაობის	41.9%	41.9%	41.9% (1,934)
სხვა თაობის	58.1%	58.1%	58.1% (2,683)
	100.0%	100.0%	4617

ნულოვანი ჰიპოთეზა: პირველი თაობის კოლეჯის სტუდენტის სტატუსსა და ასაკს შორის კავშირი არ არის.

კვლევის ჰიპოთეზა: ასაკი და პირველი თაობის კოლეჯის სტუდენტის სტატუსი სტატისტიკურად დაკავშირებულია.

(გვ. 377)

ცოდნის შემოწმება. გამოთვალეთ მოსალოდნელი სიხშირეები ასაკისა და პირველი თაობის სტატუსის მქონე სტუდენტებისათვის და ააგეთ ორვანზომილებიანი ცხრილი. არის თუ არა სვეტისა და სტრიქონის კიდურა უჯრების მნიშვნელობები საწყისი ცხრილის კიდურა უჯრების მნიშვნელობების ტოლი?

პასუხი:

	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
19/პირველი თაობის	916	1138.53	222.53	49519.60	43.49
19/სხვა თაობის	1802	1579.47	222.53	49519.60	31.35
20/პირველი თაობის	1018	795.47	222.53	49519.60	62.25
20/სხვა თაობის	-881	1103.53	222.53	49519.60	44.87

ხი-კვადრატი = 181.96, იეტსის კორექციის კოეფიციენტი = 181.15

(გვ. 384)

ცოდნის შემოწმება. რა გადაწყვეტილებას მიიღებთ ორი ცვლადის „ასაკისა" და „პირველი თაობის კოლეჯის სტუდენტის სტატუსს" შორის სტატისტიკური კავშირის თაობაზე? უარყოფთ თუ არა ნულოვან ჰიპოთეზას ალფას 0.05 ან 0.01 დონისათვის?

პასუხი:

უნდა უარვყოთ ნულოვანი ჰიპოთეზა განსხვავების არარსებობის შესახებ. გამოთვლილი ხი-კვადრატის მნიშვნელობა მნიშვნელოვანია 0.05 და 0.01 დონეებისათვის. გვაქვს საკმარისი საფუძველი მტკიცებისათვის, რომ ასაკი და პირველი თაობის კოლეჯის სტუდენტის სტატუსი ურთიერთდაკავშირებულია- 53.6% 20 წლისა და მეტი ასაკის სტუდენტები არიან პირველი თაობის კოლეჯის სტუდენტები 19 წლის ან უფრო ახალგაზრდა სტუდენტების 33.7%-თან შედარებით.

(გვ. 393)

ცოდნის შემოწმება. ორვანოზმილებიანი ცხრილისათვის: ასაკი და პირველი თაობის სტუდენტის სტატუსი მიღებული ხი-კვადრატის მნიშვნელობაა 181.15, თავისუფლების ხარისხი 1. წივნის ბოლოს მოცემული დამატება Დ-ს გამოყენებით ვპოულობთ, რომ შესაბამისი ალბათობა ნაკლებია 0.001-ზე. ეს ალბათობა ნაკლებია ალფას მნიშვნელობაზე 0.05-ზე. შედეგად, ჩვენ უარვყოფთ ნულოვან ჰიპოთეზას კავშირის არარსეობის შესახებ ასაკსა და პირველი თაობის კოლეჯის სტუდენტის სტატუსს შორის. თუ შერჩევის ზომას გავანახევრებთ, ხი-კვადრატის მნიშვნელობა იქნება 90.85. განსაზღვრეთ P-ს მნიშვნელობა 90.58-ისათვის. რა გადაწყვეტილებას მიიღებთ ნულოვან ჰიპოთეზასთან დაკავშირებით?

პასუხი:

მოუხედავად ხი-კვადრატის განახევრებისა, ნულოვანი ჰიპოთეზის უარყოფა მაინც მოხდება.

თავი 13

(გვ. 448)

ცოდნის შემოწმება. გამოიყენეთ 13.3 ნახაზი, განსაზღვრეთ ბაკალავრის ხარისხის მქონე პირების პროცენტული წილი თუ საოჯახო მეურნეობის საშუალო შემოსავალი ა) 47,500$ და ბ) 50,000$ იქნება.

პასუხი:

ბაკალავრის ხარისხის მქონეთა პროცენტული წილი, იმ შტატში სადაც საშუალო შემოსავალი 47,500$-ია იქნება დაახლოებით 28%. შესადარი პროცენტული მაჩვენებელი იმ შტატში, სადაც საშუალო შემოსავალი 50,000$-ია იქნება დაახლოებით 29.5%.

(გვ. 451)

ცოდნის შემოწმება. ქვემოთ მოყვანილი ოთხი წრფის მიხედვით განსაზღვრეთ Y-ის მიმართულება, როცა X იზრდება ერთი ერთეულით? ამ კითხვაზე პასუხი დაასაბუთეთ, როგორც წრფის გრაფიკული გამოსახულებით, ასევე წრფის განტოლებითაც.

პასუხი:

წრფისათვის Y = 1X, X-ის 1 ერთეულით ზრდისას Y-იც ასევე 1 ერთეულით იზრდება. მეორე წრფის შემთხვევაში, Y = 2 + 0.5X, Y-ის 0.5 ერთეულით ზრდისას X 1 ერთეულით იზრდება. Y = 6 - 2X წრფის შემთხვევაში X-ის 1 ერთეულით ზრდის შემთხვევაში Y მცირდება 2 ერთეულით. და ბოლოს, მეოთხე წრფის შემთხვევაში X-ის 1 ერთეულით ზრდა მოყვება Y-ის შემცირება 0.33 ერთეულით.

(გვ. 451)

ცოდნის შემოწმება. გამოიყენეთ წრფივი განტოლება, რომელიც აღწერს კავშირს საშუალო გამოცდილების წლებსა და მასწავლებლების ხელფასის ოდენობას შორის რათა გამოთვალოთ იმ მასწავლებლის ხელფასი რომელსაც 12 წლიანი სამუშაო გამოცდილება აქვს.

პასუხი:

იმ მასწავლებლის მოსალოდნელი ხელფასი, რომელსაც 12 წლიანი გამოცდილება აქვს იქნება 36.000$ (Y = 12,000 + 2,000(12)).

(გვ. 457)

ცოდნის შემოწმება. *გამოიყენეთ რეგრესიის განტოლება ნიუ-იორკის, ვერმონტის და ოჰაიოს Y მნიშვნელობების საპოვნელად. შეამოწმეთ გადის თუ არა ამ წერტილებზე ნახაზ 13.6-ზე მოცემული წრფე.*

პასუხი:

$$\text{ნიუ-იორკი: } \hat{Y} = 3.57 + 0.0005(51,384) = 29.3\%$$

$$\text{ვერმონტი: } \hat{Y} = 3.57 + 0.0005(46,832) = 27.0\%$$

$$\text{ოჰაიო: } \hat{Y} = 3.57 + 0.0005(44,532) = 25.8\%$$

(გვ. 481)

ცოდნის შემოწმება. *გამოიყენეთ რეგრესიის განტოლება ფეხბურთობის დონესა, უმუშევრობასა და სკოლის მოსწავლეების შორის დამოკიდებულების შესახებ და განსაზღვრეთ 2005 წლის ფეხბურთობის დონე მოზარდებში იმ შტატისათვის, სადაც უმუშევრობის დონე 3%, ხოლო ერთ სკოლის მოსწავლეზე დანახარჯი 6,000$-ია.*

პასუხი:

$$\hat{Y} = 70.783 + 3.364(3) - 0.003(6,000) = 62.88$$

კენტ ნომრიანი სავარჯიშოების პასუხები

თავი 1. სტატისტიკა – რატომ და რისთვის?

1. მას შემდეგ, რაც ჩვენი კვლევის ობიექტი და ჰიპოთეზა, ასევე შესასწავლი ცვლადები არჩეულია გადავდივართ კვლევის შემდეგ ეტაპზე-მონაცემთა შეგროვება. მონაცემთა შეკრების კონკრეტულ გზასა და მეთოდს უშუალოდ კვლევის მიზნები განსაზღვრავს. მონაცემთა შეკრების შემდეგ უნდა ავირჩიოთ მონაცემთა დალაგებისა და ანალიზის სისტემური მეთოდი და ანალიზის პროცედურა.

3. ა. ინტერვალურ-ფარდობითი
ბ. ნომინალური
გ. ინტერვალურ-ფარდობითი
დ. ორდინალური
ე. ნომინალური
ვ. ინტერვალურ-ფარდობითი
ზ. ინტერვალურ-ფარდობითი
თ. ნომინალური

5. ასარჩევი ცვლადები საკმოდ მრავალფეროვანია. სტუდენტები ყველაზე ხშირად ირჩევენ შემდეგ ცვლადებს: დასაქმების სახე ან დარგი, სამუშაო გამოცდილება ან ტრენინგები და კვალიფიკაცია. სტუდენტებმა პირველად უნდა განიხილონ კავშირი ამ ცვლადებსა და გენდერს შორის: ,,კაცებს აქვს უფრო მეტი სამუშაო გამოცდილება ვიდრე ქალებს ცალკეულ პროფესიებში". სტუდენტებმა ასევე შესაძლებელია განიხილონ დისკრიმინაცია და სტრუქტურული თავისებურებები.

7. ზოგადად, კრიმინალური ქმედებების(სიძულვილის გამომხატველი ქმედებების ჩათვლით) შესწავლის სირთულე არის ის, რომ ვერ თავად კრიმინალური აქტის რეგისტრაცია უნდა მოხდეს. მაგრამ კრიმინალური მოქმედებების უმრავლესობის რეგისტრირება არ ხდება. კრიმინალური ქმედებების შესახებ ინფორმაციის სისტემური შეკრება ხორციელდება გამოძიების ფედერალური ბიუროსა და იუსტიციის ბიუროს მიერ.

9.

ინდივიდის ასაკი: ეს ცვლადი შესაძლებელია გაიზომოს როგორც ინტერვალურ-ფარდობითი ცვლადი. როგორც ამ თავში განიხილეთ ინტერვალურ-ფარდობითი ცვლადები არის გაზომვის უმაღლესი დონე და შესაძლებელია გაიზომოს, როგორც ორდინალურ ასევე ნომინალურ შკალაზე.

წლიური შემოსავალი: ეს ცვლადი შესაძლებელია გაიზომოს როგორც ინტერვალურ-ფარდობითი ცვლადი. პერიოდების მიხედვით მიღებულ შემოსავლების მიმართ.
რელიგიურობა: ეს ცვლადი შესაძლებელია გაიზომოს რამდენიმე გზით. მაგალითად, რელიგიურ მსახურებებზე დასწრება შეიძლება იყოს ორდინალური (თვის მანძილზე რელიგიურ მსახურებებზე დასწრების რიცხვი: ყოველ კვირას, სულ მცირე თვეში ორჯერ, ორჯერ ნაკლებად თვეში, საერთოდ არა).
სტუდენტის აკადემიური მოსწრება: შესაძლებელია გაიზომოს როგორც ინტერვალურ-ფარდობითი ცვლადი *GPA*-ის ან ტესტების შედეგების მიხედვით.
სოციალური კლასი: ეს ორდინალური ცვლადია შემდეგი კატეგორიებით: დაბალი, მუშათა, საშუალო და მაღალი.
დამოკიდებულება მტკიცებითი ქცევის მიმართ: ეს ორდინალური ცვლადია, შემდეგი კატეგორიებით: სრულიად არ ვეთანხმები, არ ვეთანხმები, არც ვეთანხმები და არც უარვყოფ, ვეთანხმები, სრულიად ვეთანხმები.

თავი 2. ინფორმაციის თავმოყრა: სიხშირეთა განაწილებები

1. ა. რასა ნომინალური ცვლადია. კლასი ორდინალური ცვლადია, რადგან კატეგორიების დალაგება შესაძლებელია დაბალიდან მაღალისაკენ.
ბ.

სიხშირეთა განაწილების ცხრილი ცვლადი-რასისათვის:

რასა	სიხშირე
თეთრკანიანი	17
სხვა	13

სიხშირეთა განაწილების ცხრილი ცვლად – კლასისათვის:

კლასი	სიხშირე
დაბალი	3
მუშათა	15
საშუალო	11
მაღალი	1

გ. არათეთრკანიანი (სხვა) რესპონდენტების წილია: *13/30=0.43*. თეთრკანიანი რესპონდენტების წილია *(17/30)×100=56.7%*
დ. საშუალო კლასის წილია – *11/30=0.37*.
3. ა.

სიხშირეთა ცხრილი ცვლად „ავადმყოფობა ტრამვებისათვის“

ტრამვების რიცხვი	სიხშირე (*f*)
0	15
1	11
2	4
ჯამი (*N*) = 30	30

ტრავმა ინტერვალურ-ფარდობითი ცვლადია, რადგან მას აქვს ნულის ათვლის წერტილი და გაზომვის შინაარსობრივი შკალაც.

ბ. კვლევის რესპონდენტებს უფრო მეტად ან გადაუტანიათ ტრამვა გასულ წელს (ჯგუფის 50%).

გ. იმ რესპონდენტთა წილის გამოთვლა, რომლებმაც გადაიტანეს ერთი ან მეტი ტრამვა გამოითვლება ჯერ 36.7%-ისა და 13.3%-ის შეკრებით =50%. შემდეგ 100-ზე გაყოფით.

5. ამ შემთხვევაში,დამოუკიდებელი ცვლადი რასობრივ-ეთნიკური კუთვნილებაა,ხოლო დამოკიდებული ცვლადი შეზღუდულებები არალეგალურ იმიგრაციის პოლიტიკასთან დაკავშირებით. თითოეული რასობრივ-ეთნიკური ჯგუფის უმრავლესობა (50%-ზე მეტი) არ დაეთანხმა მთავრობის მიმდინარე ღონისძიებებს არალეგალურ იმიგრაციასთან მიმართებაში. კავშირი ეთნიკურ კუთვნილებასა და არალეგალურ იმიგრაციის პოლიტიკაზე შეზღუდულებას შორის არ არის რადგან, ორივე ჯგუფის პროცენტული მაჩვენებლები თითქმის იდენტურია.

7. ა.

	f	Cf	%	C%
მამაკაცი				
არასრული	100	100	15	15
უმაღლესი განათლება	327	427	48	63
უმაღლესი განათლება	68	495	10	73
არასრული კოლეჯის განათლება	190	685	28	101
ჯამი	685		101	
ქალი				
არასრული	116	116	14	14
უმაღლესი განათლება	407	523	50	64
უმაღლესი განათლება	76	599	9	73
არასრული კოლეჯის განათლება	216	815	27	100
ჯამი	815		100	
თეთრკანიანი				
არასრული	131	131	11	11
უმაღლესი განათლება	567	698	49	60
უმაღლესი განათლება	113	811	10	70
არასრული კოლეჯის განათლება	335	1146	29	99
ჯამი	1146		99	
შავკანიანი				
არასრული	50	50	24	24
უმაღლესი განათლება	109	159	53	77
უმაღლესი განათლება	19	178	9	86
არასრული კოლეჯის განათლება	28	206	14	100
ჯამი	206		100	
ესპანური წარმოშობის				
არასრული	21	21	38	36
უმაღლესი განათლება	25	46	45	83
უმაღლესი განათლება	5	51	9	92
არასრული კოლეჯის განათლება	5	56	9	101
ჯამი	56		101	

ბ. მამაკაცებისათვის: *10 + 28 = 38%*, ქალებისათვის: *9 + 27 = 36%*

გ. თეთრკანიანებისათვის: 60%, შავკანიანებისათვის 77% და ესპანური წარმოშობის რესპონდენტებისათვის – 83%

დ. კუმულაციური პროცენტები ქალებისა და მამაკაცებისათვის უფრო მეტად მსგავსია. ამ მხრივ უთანასწორობა უფრო ეთნიკური ჯგუფების მიხედვით შეინიშნება. რესპონდენტთა წილი უმაღლესი ან უფრო დაბალი განათლებით უფრო მეტად მაღალია შავკანიან და ესპანური წარმოშობის რესპონდენტებში, ვიდრე თეთრკანიან რესპონდენტებში.

9. ა. ინტერვალურ-ფარდობითი

ბ. 4.2% მამაკაცებისათვის და 1.6% ქალებისათვის.

გ.

	ქალები (f)	მამაკაცები (f)
არცერთი	445	562
1	107	91
2	46	35
3	13	5
4	13	6
ჯამი	624	699

დ. ქალებისა და მამაკაცების უმრავლესობას არ აქვს მიღებული მოძრაობის სავალდებულო ქვითარი (71% მამაკაცებისათვის და 80% ქალებისათვის). მე-2, 3 და 4 კატეგორიებში მცირედით მაღალი პროცენტია მამაკაცების შემთხვევაში ქალებთან შედარებით.

11. ეთნიკური უმცირესობის წარმომადგენელი რესპონდენტების უმრავლესობამ აღნიშნა, რომ მთავრობამ უნდა ითამაშოს „მთავარი როლი“ ეთნიკური უმცირესობების სოციალურ-ეკონომიკური მდგომარეობის გაუმჯობესებაში. 69%-მა შავკანიანი რესპონდენტების და ესპანური წარმოშობის რესპონდენტების 67%-მა აღნიშნა ნახსენები თეზა თავის პასუხებში. მეორეს მხრივ ამ პოზიციას დაეთანხმა თეთრკანიანი რესპონდენტების 32% (ეს რიცხვი დაახლოებით 35-37%-ით ნაკლებია იმ მაჩვენებლებზე, რაც შავკანიანი და ესპანური წარმოშობის რესპონდენტების შემთხვევაში დაფიქსირდა). თეთრკანიანი რესპონდენტების უმრავლესობამ აღნიშნა(51%), რომ მთავრობამ უნდა ითამაშოს „მინიმალური როლი“.

13. რესპონდენტთა პასუხი „დიახ“, ასახავს ჰომოსექსუალობის მიმართ დადებით აზრს. სქესი: უმრავლესობა ქალია. ასაკი: ახალგაზრდა რესპონდენტები უფრო მეტად პასუხობენ დადებითად(ყველაზე მაღალი პროცენტი არის 18-დან 34 წლებს შორის, 75%); პოლიტიკური კუთვნილება: უფრო მეტმა დემოკრატმა (72%) და პარტიისაგან დამოუკიდებელმა რესპონდენტმა (60%) გასცა დადებითი პასუხი ვიდრე რესპუბლიკელებმა. რელიგიურ მსახურებებზე დასწრება: დადებითი პასუხი აღნიშნეს მათმა, ვინც არ ესწრება ყოველკვირეულად რელიგიურ მსახურებებს(33% 57%-ის სანაცვლოდ მათთვის ვინც ესწრება ყოველკვირეულ რელიგიურ მსახურებებს და 74% მათთვის ვინც საერთოდ არ ესწრება რელიგიურ მსახურებებს).

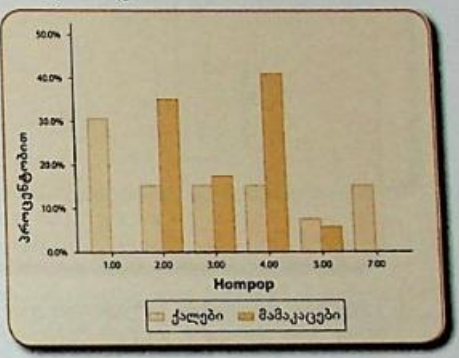
თავი 3. გრაფიკული პრეზენტაცია

1. ა. გრაფიკის მიხედვით ჩანს, რომ თითოული ჯგუფის მიხედვით გვაქვს შობადობის ზრდა 1988 წლიდან 1990 წლამდე, ხოლო კლება ფიქსირდება 1991 წელს. ზრდის დონე ყველაზე მაღალია 18 დან 19 წლამდე მოზარდებისათვის, ხოლო ყველაზე დაბალი ეს მაჩვენებელი არის 15-დან 17 წლამდე ასაკის რესპონდენტებისათვის. 2000 წლის ყველაზე დაბალი მაჩვენებლის შემდეგ გვაქვს მცირედი ზრდა 2005 და 2006 წლებში.

ბ. ზრდა დაფიქსირდა 1980-იან წლებში, რომელმაც პიკს 1991 წელს მიაღწია, განსაკუთრებით 18-19 წლის მოზარდების ჯგუფში. შობადობის დონეებში ცვლილება წლების მიხედვით შესაძლოა განპირობებულ იყოს ბევრი ფაქტორით მაგალითად როგორიცაა: აბორტების რაოდენობა, ეკონომიკის განვითარება, ფეხმძიმობის პრევენციის პროგრამების დაფინანსების ოდენობა.

3. ამ გრაფიკის მიხედვით, ხანდაზმული მოსახლეობის რაოდენობა გაიზარდა 3.1 მილიონიდან 1990 წელს 35.0 მილიონამდე 2000 წელს. ხანდაზმული მოსახლეობის რაოდენობა შედარებითი სტაბილურობით გამოირჩევა მე-20 საუკუნის საწყის წლებში – მათი რაოდენობა 10 მილიონზე ნაკლები იყო 1900-1940 წლებში. 1970 წლიდან მოყოლებული ხანდაზმული მოსახლეობის წილი სტაბილურად იზრდება, თითქმის გაორმაგდა 1970 წლიდან (20.1 მილიონი) 2000 წლის 35.0 მილიონამდე.

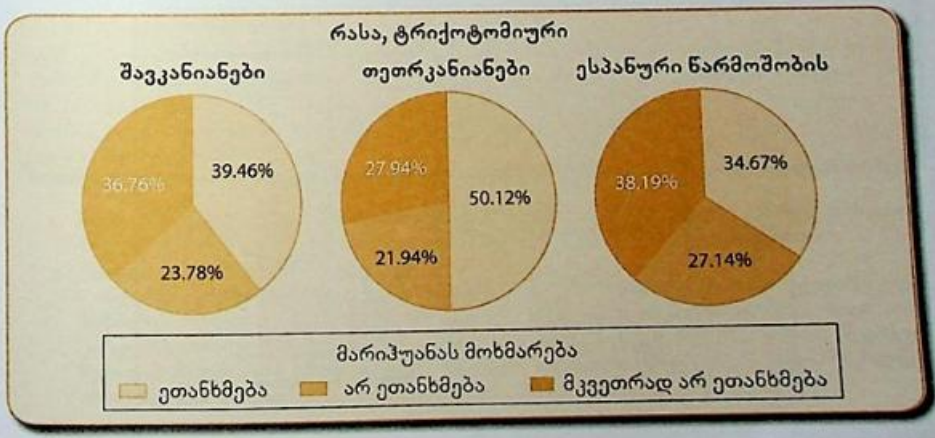
5. ა. საოჯახო მეურნეობების რაოდენობა სქესის მიხედვით (ISSP 2000)



ბ. 10%-ზე მეტმა ქალმა რესპონდენტებმა თავიანთი ოჯახის ზომა 7-ის ტოლად დააფიქსირეს. ხოლო კაცი რესპონდენტების ოჯახის მაქსიმალური ზომა 5-ს უტოლდება.

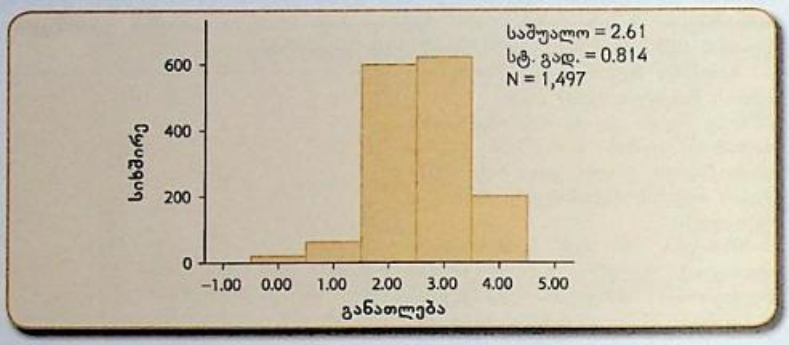
გ. სიხშირეები არ არის მოსწორებული თითოეულ ჯგუფში რესპონდენტების რაოდენობის მიხედვით. მამაკაცების რაოდენობა (17) აღემატება ქალი რესპონდენტების რაოდენობას (13), ასე რომ შესადარისობის უზრუნველსაყოფად გამოყენებულ უნდა იქნას პროცენტული მაჩვენებლები.

7. შევნიშნოთ, რომ თითოეული ჯგუფისათვის დაგჭირდებათ პროცენტული მაჩვენებლების გამოთვლა. ჩვენ მონაცემებს სხვადასხვა წრიული დიაგრამების სახით წარმოგიდგენთ. ესპანური წამოშობის სტუდენტები უფრო მეტად გამოხატავენ უარყოფით დამოკიდებულებას მარიჰუანას მოხმარებისადმი, ვიდრე თეთრკანიანი და შავკანიანი სტუდენტები. ამ კატეგორიის პროცენტული მაჩვენებელი ესპანური წარმოშობის სტუდენტებისათვის არის 38.19%. ყველაზე მაღალი მაჩვენებელი იმ სტუდენტებისა, რომლებიც არ დაეთანხმნენ მარიჰუანას მოხმარებას დაფიქსირდა თეთრკანიან სტუდენტებს შორის (50.12%).



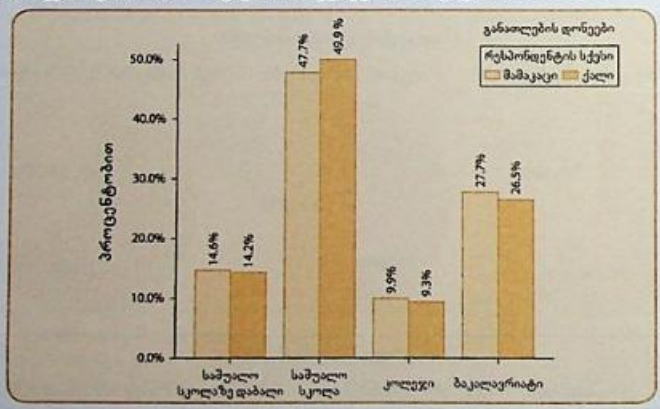
9. ა. განათლების წლები არის ინტერვალურ-ფარდობითი ცვლადი. შესაბამისად, ამ ცვლადისათვის შესაძლებელია გამოყენებულ იქნას ჰისტოგრამა.
ბ. ამ დიაგრამისათვის მონაცემები შემდეგ კატეგორიებად დაიყო: 0-4, 5-8, 9-12, 13-16, 17-20 წლები (0 შედის პირველ კატეგორიაში). პირველი კატეგორია უნდა იყოს 19, მეორე კატეგორია უნდა იყოს 62, მესამე კატეგორია უნდა იყოს 598, მეოთხე კატეგორია უნდა 619, მეხუთე კატეგორია უნდა იყოს 199.

განათლების დონე, N = 1,497, GSS 2008.

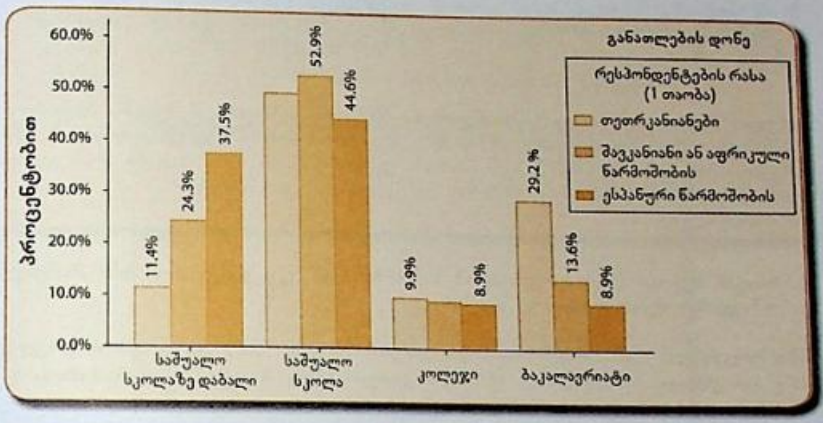


11. სვეტოვანი დიაგრამები გამოყენებულ უნდა იქნას პროცენტული მაჩვენებლების შესადარებლად სამი ეთნიკური ჯგუფის რესპონდენტების პასუხებისათვის. მიმართეთ ნახაზ 3.5-ს მაგალითისათვის. წრიული დიაგრამები შესაძლებელია გამოყენებულ იქნას ნომინალური და ორდინალური ცვლადებისათვის.
13. ა. თუ გვსურს შევადაროთ ქალებისა და მამაკაცების პასუხების მიხედვით პროცენტული მაჩვენებლები სვეტოვანი დიაგრამები იქნება ყველაზე მეტად შესაბამისი. ერთ გრაფიკზე შესაძლებელია ორი მწკრივის მონაცემების შედარება.
ბ. შენიშვნა: გამოყენებულ უნდა იქნას პროცენტული გამოთვლები.

განათლების დონეები, ქალი და მამაკაცი, პროცენტების მიხედვით, GSS 2008



განათლების დონე, პროცენტული მაჩვენებლები თეთრკანიანი, შავკანიანი და ესპანური წარმოშობის რესპონდენტებისათვის, GSS 2008



თავი 4 ცენტრალური ტენდენციის საზომები

1. ა. მოდა შესაძლებელია ვიპოვოთ ორი გზით: ან ყველაზე დიდი სიხშირის მიხედვით (749) ან ყველაზე მაღალი პროცენტული წილით(51.5%). მოდა ამ შემთხვევაში არის კატეგორია „კატეგორიულად არ ვეთანხმები“.

ბ. მედიანა შესაძლებელია ვიპოვოთ ორი გზით: სიხშირეთა სვეტის ან კუმულაციური პროცენტების გამოყენებით.

სიხშირეების გამოყენება	*კუმულაციური პროცენტების გამოყენება*
$\frac{N+1}{2} = \frac{1,455+1}{2} = 728$ - ე შემთხვევა	შევნიშნოთ, რომ დაკვირვების მონაცემების 48.5% ხვდება „არ ვეთანხმები“ კატეგორიის კუმულაციურ პროცენტში.
პირველი სვეტის სიხშირით(288) შეკრიბეთ სიხშირეები 728-ე შემთხვევამდე. 728-ე დაკვირვების მონაცემი შეესაბამება კატეგორიას „კატეგორიულად არ ვეთანხმები“, რაც თავის მხრივ მედიანაა.	50%-იანი საზღვარი ან მედიანა მდებარეობს „კატეგორიულად არ ვეთანხმები“ კატეგორიასთან. ასე რომ მედიანა არის „კატეგორიულად არ ვეთანხმები“.

გ. მოდა არის ყველაზე მეტი სიხშირის კატეგორია განაწილებაში. მედიანა განაწილებას ყოფს ორ თანაბარ ნაწილად, ასე რომ მედიანა დაკვირვების მონაცემთა შუა წერტილია.

დ. რადგან ეს ორდინალური ცვლადია.

3. ა. ინტერვალურ ფარდობითი. მოდა შესაძლებელია ვიპოვოთ ორი გზით: ან ყველაზე დიდი სიხშირის მიხედვით (27) ან ყველაზე მაღალი პროცენტული წილით(41.5%). მოდა არის კატეგორია, რომელსაც შეესაბამება „გასულ წელს 52 კვირის მუშაობა“. მოდა შესაძლებელია ვიპოვოთ ორი გზით: ან ყველაზე დიდი სიხშირის მიხედვით ან ყველაზე მაღალი პროცენტული წილით.

სიხშირეების გამოყენება	*კუმულაციური პროცენტების გამოყენება*
$\frac{N+1}{2} = \frac{65+1}{2} = 33$-ე შემთხვევა	შევნიშნოთ, რომ დაკვირვების მონაცემების 47.5% ხვდება „გასულ წელს 52 კვირის მუშაობა“ კატეგორიაში.
პირველი სვეტის სიხშირით (10) შეკრიბეთ სიხშირეები 33-ე შემთხვევამდე. 33-ე დაკვირვების მონაცემი შეესაბამება კატეგორიას „გასულ წელს 47 კვირის მუშაობა“, რაც თავის მხრივ მედიანაა.	50%-იანი საზღვარი ან მედიანა მდებარეობს „გასულ წელს 47 კვირის მუშაობა“ კატეგორიასთან. ასე რომ მედიანა არის „გასულ წელს 47 კვირის მუშაობა“.

ბ. რადგან მედიანა 50-ე პერცენტილის სინონიმია, ჩვენ უკვე ვიცით რომ მისი მნიშვნელობა 47 კვირის მუშაობაა გასულ წელს.

25-ე პერცენტილი = *(65×0.25) = 16.25*-ე დაკვირების მონაცემი = *25.5* კვირის მუშაობა გასულ წელს
75-ე პერცენტილი = *(65×0.75) = 48.75*-ე დაკვირების მონაცემი = „*52* კვირის მუშაობა გასულ წელს“.

5. მოდა შესაძლებელია ვიპოვოთ სვეტებში მოცემული სიხშირეების მიხედვით თითოეული ჯგუფისათვის, ჯგუფები მოცემულია ქვემოთ:
18-29: კარგი
30-39: კარგი
40-49: კარგი
50-59: კარგი
მედიანა შესაძლებელია ვიპოვოთ ორი გზით: სიხშირეთა სვეტის ან კუმულაციური პროცენტების გამოყენებით. რადგან ამ სავარჯიშოში მხოლოდ სიხშირეებია მოცემული ჩვენ მხოლოდ სიხშირეებს გამოვიყენებთ მედიანის საპოვნელად.

ასაკობრივი ჯგუფი

18-29	*30-39*	*40-49*	*50-59*
$\frac{N+1}{2} = \frac{188+1}{2} = 94.5$-ე შემთხვევა	$\frac{N+1}{2} = \frac{176+1}{2} = 88.5$-ე შემთხვევა	$\frac{N+1}{2} = \frac{185+1}{2} = 93$-ე შემთხვევა	$\frac{N+1}{2} = \frac{182+1}{2} = 91.5$-ე შემთხვევა
პირველი კატეგორიის სიხშირიდან (64) დაწყებული შეკრიბეთ სიხშირეები 94-ე და 95-ე შემთხვევამდე. ორივე დაკვირვების მონაცემი შეესაბამება კატეგორიას „კარგი“, რაც თავის მხრივ მედიანაა.	პირველი კატეგორიის სიხშირიდან (57) დაწყებული შეკრიბეთ სიხშირეები 88-ე და 89-ე შემთხვევამდე. ორივე დაკვირვების მონაცემი შეესაბამება კატეგორიას „კარგი“, რაც თავის მხრივ მედიანაა.	პირველი კატეგორიის სიხშირიდან (46) დაწყებული შეკრიბეთ სიხშირეები 93-ე შემთხვევამდე. ორივე დაკვირვების მონაცემი შეესაბამება კატეგორიას „კარგი“, რაც თავის მხრივ მედიანაა.	პირველი კატეგორიის სიხშირიდან (37) დაწყებული შეკრიბეთ სიხშირეები 91-ე და 92-ე შემთხვევამდე. ორივე დაკვირვების მონაცემი შეესაბამება კატეგორიას „კარგი“, რაც თავის მხრივ მედიანაა.

ბ. მოდა ყველა ასაკობრივი ჯგუფისათვის შესაძლებელია განპირობებულ იყოს იმით, თუ როგორ მოახდინეს რესპონდენტებმა შეკითხვის ინტერპრეტირება. მაგალითად, შესაძლებელია რომ პირის ჯანმრთელობის მდგომარეობა შეფასდა მისი ასაკიდან გამომდინარე. მოდა და მედიანა ერთანმეთთან შედარებით არ გვაძლევს უკეთეს ინფორმაციას, ვინაიდან ისინი ერთნაირ ინფორმაციას გვაწვდიან.

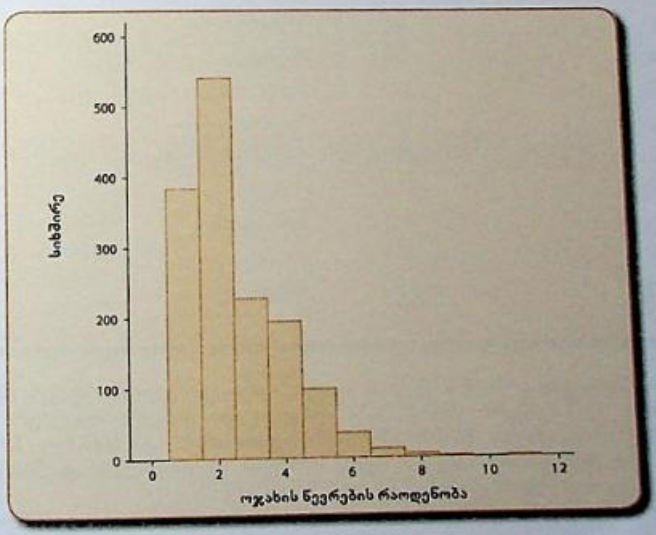
7. დავიწყოთ თითოეული ოჯახის ზომის გამრავლებით მის სიხშირეზე.

ოჯახის ზომა	სიხშირე	fY
1	384	384
2	541	1,082
3	228	684
4	195	780
5	98	490
6	36	216
7	12	84
8	4	32
9	1	9
11	1	11
ჯამი	1,500	3,772

$$\overline{Y} = \frac{\sum fY}{N} = \frac{3,772}{1,500} = 2.51$$

ასე, რომ ოჯახის წევრების საშუალო რიცხვი აშშ-ში არის 2.51.

9. ა. მონაცემებიდან ჩანს რომ ექსტრემალური მნიშვნელობის მქონე მონაცემების რიცხვი მცირეა, ასე რომ შეგვიძლია დავასკვნათ, რომ განაწილება დადებითი მიმართულებით არის გაშლილი.



ბ. მედიანა შესაძლებელია ვიპოვოთ ორი გზით: სიხშირეთა სვეტის ან კუმულაციური პროცენტების გამოყენებით. გვაქვს სიხშირეთა მონაცემები, შესაბამისად გამოვიყენებთ სიხშირეებს მედიანის საპოვნელად. რადგან მედიანა (2) ნაკლებია საშუალო მნიშვნელობაზე (2.48), შეგვიძლია დავასკვნათ, რომ განაწილება დადებითი მიმართულებით არის გადახრილი. ეს შედეგი კიდევ უფრო ამტკიცებს ჩვენს პასუხს 9ა დავალებაში.

სიხშირეთა გამოყენება

$$\frac{N+1}{2} = \frac{1,500+1}{2} = 750.5\text{-ე შემთხვევა}$$

პირველი კატეგორიის სიხშირედან დაწყებული (384) დაუკავშირეთ სიხშირეები 750-ე და 751-ე დაკვირვების მონაცემებამდე. ორივე მათგანი შეესაბამება კატეგორია „2“-ს, რაც თავის მხრივ მედიანაა.

11. დიახ, ორივე ეს პოლიტიკოსი შეიძლება იყოს სწორი, სულ მცირე ტექნიკური კუთხითაც მაინც. ერთი პოლიტიკოსი შესაძლებელია ეყრდნობოდეს საშუალო მნიშვნელობას, ხოლო მეორე მედიანას. უჩვეულო იქნებოდა ორივე ამ სტატისტიკური მაჩვენებლის ერთმანეთთან ტოლობა. ამერიკელთა საშუალო შემოსავალი მეტი იქნება ვიდრე მედიანა თუ შემოსავლების შესახებ მონაცემები დადებითი მიმართულებითაა გადახრილი.
13. პირველ რიგში დავალაგოთ მნიშვნელობები უმცირესიდან უდიდესის მიმართულებით.

მკვლელობების რაოდენობა ყოველ 100,000 მცხოვრებზე 2005 წელს

პირველი 10 მოსახლეობის რაოდენობის მიხედვით	მკვლელობების დონე	ბოლო 10 მოსახლეობის რაოდენობის მიხედვით	მკვლელობების დონე
სან-დიეგო	4.0	კორპუს კრისტი	2.8
ნიუ-იორკი	6.6	ლესინგტონი	5.6
სან ანტონიო	6.8	ანკორაჟი	5.8
ლას ვეგასი	11.3	სეინტ პაული	8.6
ლოს ანჟელესი	12.6	ბაკერსფილდი	11.2
ფოენიქსი	15.0	სანტ პეტერბურგი	11.8
ჩიკაგო	15.6	მობილი	14.0
ჰუსტონი	16.3	სტოკტონი	14.6
დალასი	16.4	ბუფალო	19.8
ფილადელფია	25.6	ნიუვარკი	34.5
	$\sum = 130.2$		$\sum = 128.7$

მედიანა მარტივად არის შუა წევრი. რადგან თითოეულ ჯგუფში გვაქვს 10 ქალაქი, ჩვენ გვჭირდება მე-5.5 წევრი *([10 + 1]/2)*. პირველი ათი ქალაქის შუა წევრი ხვდება ლოს ანჟელესსა და ფოენიქსს შორის. ასე რომ მედიანა არის *13.8 ([12.6 + 15.0] / 2)*. ხოლო მეორე ათი ქალაქისათვის, მე-5.5 წევრი ემთხვევა ბაკერსფილდსა და სანტ-პეტერბურგს შორის. ასე, რომ მედიანა არის *11.5 ([11.2 + 11.8] / 2)*.

საშუალო მაჩვენებელი თითოეული ჯგუფისათვის გამოითვლება შემდეგი ფორმულის გამოყენებით:

$$\overline{Y} = \frac{\sum Y}{N}$$

პირველი ათი ქალაქის მკვლელობის დონის საშუალო არის 13.02(130.2/10). ხოლო მეორე ათი ქალაქის ჯგუფისათვის იგივე მაჩვენებელი 12.87-ია(128.7/10). ნივარკსა და ნიუ-ჯერსს აქვს მკვლელობის ყველაზე მაღალი დონე (34.5). საშუალოდ ქალაქებს დიდი მოსახლეობით აქვთ მკვლელობების მაღალი დონე. განაწილება ამ მეორე ათი ქალაქის მკვლელობის დონის მონაცემების არის გადახრილი დადებითი მიმართულებით, რადგან მხოლოდ რამდენიმე ქალაქს აქვს ძალიან მაღალი მაჩვენებელი. მაგრამ განსხვავებულია პირველი ათის ქალაქის მონაცემები, რომელიც უარყოფითი მიმართულებით არის გადახრილი.

15. ა. დავიწყოთ შტატების დალაგება დაბალიდან მაღალისაკენ დაფიქსირებული ქურდობების რაოდენობის მიხედვით.

შტატი	რეგისტრირებული ქურდობის რიცხვი
ჩრდილოეთ დაკოტა	189
არკანზასი	384
როდ აილენდი	615
ვისკონსინი	2,536
ორეგონი	2,815
ნიუ-ჯერსი	6,394
მიჩიგანი	6,784
ილინოისი	10.080
ნიუ-იურკი	16,452
ფლორიდა	17,780
ტეხასი	26,006

რადგან ჩვენი დაკვირვებების რიცხვი კენტია(11), მედიანა არის მე-6 წევრი ანუ ნიუ-ჯერსი. შესაბამისად მედიანის მნიშვნელობაა 6,394. საშუალო მაჩვენებლის გამოთვლისათვის დავაჯამოთ ყველა შტატის მაჩვენებელი.

შტატი	რეგისტრირებული ქურდობის რიცხვი
ჩრდილოეთ დაკოტა	189
არკანზასი	384
როდ აილენდი	615
ვისკონსინი	2,536
ორეგონი	2,815
ნიუ-ჯერსი	6,394
მიჩიგანი	6,784
ილინოისი	10.080
ნიუ-იურკი	16,452
ფლორიდა	17,780
ტეხასი	26,006
	$\sum = 90,035$

$$\bar{Y} = \frac{\sum fY}{N} = \frac{90,035}{11} = 8.185$$

ბ. რადგან საშუალოს მნიშვნელობა (8,185) აღემატება მედიანის მნიშვნელობას (6,394), ქურდობის მონაცემების განაწილება დადებითი მიმართულებითაა გადახრილი.

გ. ყველაზე მეტად ლოგიკური პასუხი იქნებოდა თითოეულ შტატში მოსახლეობის რიცხვი. ასევე ჩვენი ახსნა მოიცავს კანონების თავისებურებებს, საერთო დანაშაულის დონეს და ა.შ.

თავი 5. ვარიაციის საზომები

1. ა. ცხრილში ნაჩვენებია ოთხი კატეგორია პოლიტიკურ შეხედულებებზე პასუხების მიხედვით

ბ. კვადრატული პროცენტების ჯამი, $\sum Pct^2$ უდრის 2,241.74

პოლიტიკური შეხედულებები	პროცენტული წილი (%)	პროცენტული წილი^2 (%)^2
უკიდურესად ლიბერალი	3.6	12.96
ლიბერალი	12.2	148.84
მცირედით ლიბერალი	12.0	144
ზომიერი	38.0	1,444
მცირედით კონსერვატორი	14.1	198.81
კონსერვატორი	16.8	282.24
უკიდურესად კონსერვატორი	3.3	10.89
ჯამი	100.0	2,241.74

გ. ფორმულის გამოყენებით *IQV*-ს შემდეგნაირად ვითვლით:

$$IQV = \frac{K(100^2 - \sum Pct^2)}{100^2(K-1)} = \frac{7(100^2 - 2,241.74)}{100^2(7-1)} = \frac{54,307.82}{60.000} = 0.91$$

0.91 მნიშვნელობა *IQV*-ათვის ნიშნავს, რომ ამერიკელები საკმაოდ მრავალფეროვანნი არიან თავიანთი პოლიტიკური შეხედულებებით, შესაბამისად ეს შედეგი ამყარებს ჩვენს დაკვირვებებს ცხრილის მონაცემებზე.

3. ა. კორუფციის დონის მონაცემების დიაპაზონი 1990 წელს არის (583-79)=504. იგივე მონაცემის დიაპაზონი 2007 წლისათვის არის (405-85)=320. დიაპაზონი უფრო დიდია 1990 წლისათვის 2007 წელთან შედარებით.

ბ. კორუფციის მონაცემთა საშუალო არის 295.67 1990 წლისათვის, ხოლო 2007 წლისათვის 255.

გ. სტანდარტული გადახრა არის 359.32 1990 წლისათვის, ხოლო 2007 წლისათვის 160.93.

1990

სამთავრობო დონე	დანაშაულის შემთხვევათა რაოდენობა	$Y - \bar{Y}$	$(Y - \bar{Y})^2$
ფედერალური	583	287.33	82,558.53
შტატი	79	-216.67	46,945.89
ადგილობრივი	225	-70.67	4,994.25
ჯამი	887	0.01	134,498.67

$$\bar{Y} = 295.67$$

$$S_Y = \sqrt{S_Y^2} = \sqrt{\frac{\sum (Y - \bar{Y})}{N-1}} = \sqrt{\frac{134,498.67}{2}} = 259.32$$

2007

სამთავრობო დონე	დანაშაულის შემთხვევათა რაოდენობა	$Y - \bar{Y}$	$(Y - \bar{Y})^2$
ფედერალური	403	150	22,500
შტატი	85	-170	28,900
ადგილობრივი	275	20	400
ჯამი	765	0.00	51,800

$$\bar{Y} = 255$$

$$S_Y = \sqrt{S_Y^2} = \sqrt{\frac{\sum (Y - \bar{Y})}{N-1}} = \sqrt{\frac{51,800}{2}} = 160.93$$

დ. სტანდარტული გადახრა 1990 წლისათვის უფრო მაღალია 2007 წელთან შედარებით, რაც ასახავს უფრო მაღალ ვარიაციას 1990 წლისათვის 2007 წელთან შედარებით. ეს უკანასკნელი კიდევ უფრო ამყარებს 3ა დავალებაში მიღებულ პასუხს.

5. ა. დასავლეთის შტატებისათვის ხანდაზმული მოსახლეობის გამოთვლილი პროცენტული ზრდა არის 36.2%(50.0%-13.8%). პროცენტული ზრდის დიაპაზონი დასავლეთის შუა შტატებისათვის არის 9.8%(19.2%-9.4%). დასავლეთის შტატებს აქვთ უფრო დიდი დიაპაზონი.

ბ. *IQR* დასავალეთის შტატებისათვის არის *14.85% (34.35% - 19.5%)*, ხოლო *IQR* დასავლეთის შუა შტატებისათვის არის *3.2% (14.5% - 11.3%)*. კვლავ დასავლეთის შტატების მაჩვენებელი უფრო მაღალია.

გ. დასავლეთის შტატების ხანდაზმული მოსახლეობის გამოთვლილ პროცენტულ ზრდაში საკმოდ მაღალია ვარიაციის მაჩვენებელი, რაც ძირითადად განპირობებულია დიდი ზრდით ნევადაში, არიზონასა, ვაიომინგსა და ალასკაში.

7. ა. დიაპაზონი ტოლია *3.6 (6.5-2.9)*. 25-ე პერცენტილი 3.05 ნიშნავს, რომ დაკვირვების მონაცემთა 25% ხვდება 3.05-ს ქვემოთ განქორწინების რიცხვისათვის ყოველ 1000 მცხოვრებზე. მსგავსად, 75-ე პეცენტილი ნიშნავს, რომ 75% დაკვირვების მონაცემების ხვდება 4.6-ს ქვემოთ განქორწინების რიცხვისათვის ყოველ 1000 მცხოვრებზე.

25-ე პერცენტილი	10(0.25)=2.5-ე დაკვირვება	ასე, რომ (3.0+3.1)/2=3.05
75-ე პერცენტილი	10(0.75)=7.5-ე დაკვირვება	ასე,რომ (4.5+4.7)/2=4.6

ბ.

შტატი	განქორწინების რიცხვი ყოველ 1000 მცხოვრებზე	$Y - \overline{Y}$	$(Y - \overline{Y})^2$
ალასკა	4.3	0.2	0.04
ფლორიდა	4.7	0.6	0.36
აიდაჰო	4.9	0.8	0.64
მაინი	4.5	0.4	0.16
მერილენდი	3.1	-1	1
ნევადა	6.5	2.4	5.76
ნიუ-ჯერსი	3.0	-1.1	1.21
ტეხასი	3.3	-0.8	0.64
ვერმონტი	3.8	-0.3	0.09
ვისკონსინი	2.9	-1.2	1.44
ჯამი	41	0.00	11.34

$$\overline{Y} = \frac{\sum Y}{N} = \frac{41}{10} = 4.1$$

$$S_y = \sqrt{S_Y^2} = \sqrt{\frac{\sum (Y - \overline{Y})}{N-1}} = \sqrt{\frac{11.34}{9}} = 1.12$$

განქორწინებათა საშუალო რიცხვი 4.1-ია. საშუალოდ შტატების მაჩვენებლები გადახრის საშუალო მნიშვნელობიდან 1.12 პროცენტით ყოველ 1000 მოსახლეზე.

8. განქორწინებათა რიცხვი შესაძლებელია განსხვავდებოდეს ისეთი ფაქტორების გავლენის გამო, როგორიცაა: განსხვავებები რელიგიურობის დონეებში, სახელმწიფო პოლიტიკის თავისებურებები ან დასაქმების შესაძლებლობები.

9. ა. დანაშაული საშუალო დონეა 3,066.8 და პოლიციაზე გაწეული ხარჯების საშუალო მაჩვენებელი მილიონ დოლარებში არის 1,530.05.

დანაშაულის შემთხვევების ოდენობა	პოლიციაზე გაწეული დანახარჯები
$\overline{Y} = \frac{\sum Y}{N} = \frac{64,403}{21} = 3,066.8$	$\overline{Y} = \frac{\sum Y}{N} = \frac{32,131}{21} = 1,530.05$

ბ. დანაშაულის რიცხვის სტანდარტული გადახრა არის 725.58, ხოლო დანახარჯების სტანდარტული გადახრა არის $1,754.80.

დანაშაულის რიცხვი	$Y - \overline{Y}$	$(Y - \overline{Y})^2$	პოლიციაზე დანახარჯი	$Y - \overline{Y}$	$(Y - \overline{Y})^2$
2,635	431.8	186,451.24	221	1,309.05	1,713,611.90
2,013	1,053.8	1,110,494.44	274	1,256.05	1,577,661.60
2,442	624.8	390,375.04	136	1,394.05	1,943,375.40
2,838	228.8	52,349.44	1,673	142.95	20,434.70
2,815	251.8	63,403.24	286	1,244.05	1,547,660.40
2,785	281.8	79,411.24	905	625.05	390,687.50
2,488	578.8	335,009.44	7,585	6,054.95	36,662,419.5
2,644	422.8	178,759.84	3,010	1,479.95	2,190,252.00
2,883	183.8	33,782.44	2,535	1,004.95	1,009,924.50
4,029	962.2	925,828.84	2,689	1,158.95	1,343,165.10
3,817	750.2	562,800.04	1,039	491.05	241,130.10
3,562	495.2	245,223.04	3,761	2,230.95	4,977,137.90
3,775	708.2	501,547.24	2,333	802.95	644,728.70
3,102	35.2	1,239.04	1,434	96.05	9,225.60
3,398	331.2	109,693.44	1,294	236.05	55,719.60
3,087	20.2	408.04	570	960.05	921,696.00
4,373	1,306.2	1,706,158.44	1,179	351.05	123,236.10
2,128	938.8	881,345.44	106	1,424.05	2,027,918.40
1,791	1,275.8	1,627,665.64	132	1,398.05	1,954,543.80
3,623	556.2	309,358.44	334	1,196.05	1,430,535.60
4,175	1,108.2	1,228,107.24	635	895.05	801,114.50

გ. რადგან დანაშაულის რიცხვი და პოლიციაზე დანახარჯები არ არის აღრიცხული ერთნაირი საზომი ერთეულით, ამიტომ სწორი არ იქნება მათი საშუალო და სტანდარტული გადახრები პირდაპირ შევადაროთ ერთმანეთს. უფრო სწორია ეს მაჩვენებლები ცალ-ცალკე განვიხილოთ. დანაშაულის შესახებ მონაცემების განხილვით ვიცით, რომ მისი სტანდარტული გადახრა დიდია 725.58, რაც საშუალოდან მონაცემთა ფართო გაფანტვაზე მიუთითებს. სტანდარტული გადახრის ამ დიდ მნიშვნელობაში წვლილი შეაქვს მისურისა და დაკოტას შტატებს- მათი მნიშვნელობები დაშორებული საერთო საშუალოს მნიშვნელობისაგან. ასევე პოლიციაზე დანახარჯების მონაცემების სტანდარტული გადახრაც დიდია- 1,530.05, ხოლო სტანდარტული გადახრა იგივე მონაცემებისათვის არის 1,754.80. ამ სტანდარტული გადახრის სიდიდეში წვლილი შეაქვს ჩრდილეთ დაკოტას და ნიუ-იორკს, რადგან მათი შესაბამისი მნიშვნელობები განსხვავდება საერთო საშუალოსგან.

დ. სხვა ფაქტორებთან ერთად მხედველობაში უნდა მივიღოთ თითოეული შტატის ეკონომიკური მდგომარეობა. ეკონომიკური განვითარების დაბალმა დონემ შესაძლებელია შეიტანოს წვლილი დანაშაულის დონეს და ერთ სულ მოსახლეზე პოლიციის დანახარჯების ოდენობებში.

11. ა. რადგან ცხრილი 5.5 თავისთავად დიდია, ამიტომ მხოლო ერთი სტანდარტული გადახრის ფორმულაა მოყვანილი. სტანდარტული გადახრა ასაკოვანი მოსახლეობის ზრდის მაჩვენებლისათვის 2008 წლიდან 2015 წლამდე არის 9.76%.

$$S_Y = \sqrt{S_Y^2} = \sqrt{\frac{\sum(Y-\bar{Y})}{N-1}} = \sqrt{\frac{4,758.28}{50}} = 9.76$$

ბ. *IQR* უდრის 11.05-ს.

გ. დიახ, სტანდარტული გადახრის მაჩვენებელიც იგივე დასკვნამდე მიგვიყვანს, როგორც *IQR* და მარკუთხა დიაგრამა. მართკუთხა დიაგრამა აჩვენებს მნიშვნელობების ფაქტობრივ დიაპაზონებს, რაც სტანდარტული გადახრების შემთხვევაში არ ჩანს.

13. რადგან ცვლადი ინტერვალურ-ფარდობითია უნდა გამოვიყენოთ ვარიაცია (ან სტანდარტული გადახრა), დიაპაზონი ან *IQR*. ამ სამ მაჩვენებელს შორის სტანდარტული გადახრა უფრო მისაღებია. როგორც ეს მე-4 თავში განვიხილეთ ცენტრალური ტენდენციის გასაზომად თუ ვეძებთ საშუალო სიცოცხლის ხანგრძლივობას, მაშინ უნდა გამოვიყენოთ საშუალო არითმეტიკული.

საშუალოდ, არაევროპულ ქვეყნებს აქვთ სიცოცხლის მცირედით მეტი ხანგრძლივობა ადრეული ასაკისათვის. როგორც მედიანა ასევე საშუალო მაღალია არაევროპული ქვეყნებისათვის ევროპულ ქვეყნებთან შედარებით. ასევე, ევროპის ქვეყნების მონაცემებში უფრო მეტი ვარიაცია შეიმჩნევა. ევროპული ქვეყნებისათვის სტანდარტული გადახრა არის 4.30 წელი, ხოლო არაევროპული ქვეყნებისათვის იგივე მაჩვენებელი 2.76 წელს შეადგენს. ასევე მაღალია ევროპული ქვეყნებისათვის *IQR* მაჩვენებელი(5.45), არაევროპული ქვეყნებისათვის *IQR = 2.70*.

ეს განსხვავებები შესაძლებელია აიხსნას ჯანდაცვის მომსახურების ხელმისწვდომობით ან შესაძლოა განსხვავებები შემთხვევითი ფაქტორით იყოს გამოწვეული, ამ მაგალითის შერჩევის მცირე ზომის გამო. შესაძლებელია სხვა შედეგი მივიღოთ თუ სხვა ქვეყნებსაც ჩავრთავთ ანალიზში.

შედეგების ცხრილი ნაჩვენებია ქვემოთ:

სიცოცხლის ხანგრძლივობა		სტატისტიკა
ევროპული ქვეყნები	საშუალო	81.42
	მედიანა	82.40
	ვარიაცია	18.50
	სტანდარტული გადახრა	4.30
	მინიმუმი	74.00
	მაქსიმუმი	84.40
	დიაპაზონი	10.40
	ინტერკვარტალური დიაპაზონი	5.45
არაევროპული ქვეყნები	საშუალო	82.36
	მედიანა	82.70
	ვარიაცია	7.63
	სტანდარტული გადახრა	2.76
	მინიმუმი	78.10
	მაქსიმუმი	85.80
	დიაპაზონი	7.70
	ინტერკვარტალური დიაპაზონი	2.70

15. ა. მე-20 პერცენტილი არის „საშუალო განათლების მქონე“ *(0.20 × 131,144 = 26,228.8)*. მე-80 პერცენტილი არის „ბაკალავრის ხარისხი“ *(0.80 × 131,144 = 104,915.2)*.
ბ. იმ ფაქტის მხედველობაში მიღებით, რომ ამერიკელების მცირე პროცენტს აქვს კოლეჯის განათლება და მათგან კიდევ უფრო მცირეს აქვს უმაღლესი განათლება, აქედან გამომდინარე კავშირის გაკვლება არჩევნებში მონაწილეობასა და განათლებას შორის პრობლემატური იქნებოდა. წარმოდგენილი მონაცემები ნედლი სიხშირეებია.

თავი 6. ნორმალური განაწილება

1. ა. *Z* მნიშვნელობა პირისათვის, რომელიც დღეში 8 საათის განმავლობაში უყურებს ტელევიზორს არის:

$$Z = \frac{8 - 2.95}{2.58} = 1.96$$

ბ. პირველად გამოთვალოთ იმ პირისათვის *Z* მნიშვნელობა, რომელიც დღეში 5 საათს უყურებს ტელევიზორს:

$$Z = \frac{5 - 2.95}{2.58} = 0.79$$

Z-ისა და საშუალო მნიშვნელობას შორის მდებარე არე არის 0.2852. შემდეგ საჭიროა შევაჯამოთ 0.50 და 0.2852 იმისათვის, რომ ვიპოვოთ იმ რესპონდენტების პროპორცია რომლებიც დღის მანძილზე 5 საათზე ნაკლებ დროს უთმობენ ტელევიზორის ყურებას. ე.ი. შეგვიძლია დავასკვნათ, რომ იმ რესპონდენტების წილი, რომლებიც დღეში 5 საათზე ნაკლებ დროს უყურებენ ტელევიზორს არის 0.7852. ეს შეესაბამება 767.93 რესპონდენტს *(0.7852 × 978)*.
გ. 5.38 საათი ტელევიზორის ყურება შეესაბამება *Z* მნიშვნელობას +1-ს.

$$Y = Y + Z\,(Sy) = 2.95 + 1\,(2.58) = 5.33 \text{ საათი/დღეში}$$

დ. *Z* მნიშვნელობა იმ პირისათვის, რომელიც დღეში 1 საათს უყურებს ტელევიზორს არის:

$$Z = \frac{1 - 2.95}{2.58} = -0.76$$

Z-ისა და საშუალო მნიშვნელობას შორის მდებარე არე არის 0.2764.
Z მნიშვნელობა იმ პირისათვის, რომელიც დღეში 6 საათს უყურებს ტელევიზორს არის:

$$Z = \frac{6 - 2.95}{2.58} = 1.18$$

Z-ისა და საშუალო მნიშვნელობას შორის მდებარე არე არის 0.3810.
შესაბამისად, იმ რესპონდენტების პროცენტული წილი, რომლებიც უყურებენ *D* დღეში 1-დან 6 საათამდე ტელევიზორს არის *65.74% (0.2764 + 0.3810 = 0.6574 × 100)*.

3. ა. 13.44 წლიანი განათლების მქონე რესპონდენტისათვის Z მნიშვნელობა იქნება:

$$Z = \frac{13.44 - 13.44}{3.1} = 0$$

ბ. რადგან ჩვენი მეგობრის განათლების წლების რაოდენობა თავსდება მე-60 პერცენტილში, ჩვენ გვჭირდება Y-ის მიმართ ამოხსნა. თავდაპირველად უნდა გამოვიყენოთ ნორმალური განაწილების ლოგიკა Z-ის პოვნისათვის. ნებისმიერი ნორმალური განაწილებისათვის, დაკვირვების მონაცემების 50% ხვდება საშუალო მნიშვნელობის ზემოთ. რადგან ჩვენი მეგობრის განათლების წლები მე-60 პერცენტილშია, ვიცით რომ არე საშუალო მნიშვნელობასა და ამ კონკრეტულ მნიშვნელობას შორის 0.10-ია. ასევე, ამ მნიშვნელობამდე არე 0.10-ია. შეგვიძლია გამოვიყენოთ ბ დამატების ბ სვეტი 0.10-ის მნიშვნელობისათვის ბ ან გ სვეტში 0.40-ისათვის. ცხრილში ვპოულობთ, რომ ამ მნიშვნელობების შესაბამისი Z უდრის 0.25, მაშინ Y-ისათვის მივიღებთ:

$$Y = \overline{Y} + Z\,(Sy) = 13.44 + 0.25\,(3.1) = 14.22$$

გ. რადგან ვიცით რომ პროპორცია განათლების წლებსა(13.44) და ჩვენი მეგობრის განათლების წლებს შორის (14.22) 0.10-ია, შეგვიძლია გავამრავალოთ ეს პროპორცია $N(1,497)$-ზე. ე.ი. 149.7 ადამიანს აქვს 13.44-სა და 14.22 წლებს შორის განათლება.

5. ა. მუშათა კლასის წევრი რესპონდენტებისათვის:
12-იანი მნიშვნელობისათვის შესაბამისი Z იქნება:

$$Z = \frac{12 - 12.73}{2.79} = -0.26$$

16-იანი მნიშვნელობისათვის შესაბამისი Z იქნება:

$$Z = \frac{13 - 12.73}{2.79} = 1.17$$

არე $Z = -026$-სა და საშუალოს შორის არის 0.1026. ხოლო არე $Z = 1.17$-სა და საშუალოს შორის არის 0.3790, ასე რომ:

$$\text{არე} = 0.1026 + 0.3790 = 0.4816$$

ასე, რომ იმ რესპონდენტთა პროპორციული წილი, რომელთაც აქვთ 12-დან 16 წლამდე განათლება უდრის 0.4816-ს.
მაღლი კლასის რესპონდენტებისათვის:
12-იანი მნიშვნელობისათვის შესაბამისი Z იქნება:

$$Z = \frac{12 - 15.49}{2.95} = -1.18$$

16-იანი მნიშვნელობისათვის შესაბამისი Z იქნება:

$$Z = \frac{16 - 15.49}{2.95} = 0.17$$

არე $Z = -1.18$-სა და საშუალოს შორის არის 0.3810. არე $Z = 0.17$-სა და საშუალოს შორის არის 0.0675. ასე რომ მთლიანი არე იქნება:

$$\text{არე} = 0.3810 + 0.0675 = 0.4485$$

ასე,რომ მაღლი კლასის წარმომადგენელ იმ რესპონდენტთა პროპორციული წილი,რომელთაც აქვთ 12-დან 16 წლამდე განათლება უდრის 0.4485-ს.

ბ. მუშათა კლასის წევრი რესპონდენტებისათვის:
16-იანი მნიშვნელობისათვის შესაბამისი Z იქნება:

$$Z = \frac{16 - 12.73}{2.79} = 1.17$$

არე $Z = 1.17$-სა და განაწილების კუდს შორის (გ სვეტი) არის 0.1210. ასე რომ,ალბათობა იმისა, რომ მუშათა კლასის წარმომადგენელ რესპონდენტს ექნება 16 წელზე მეტი ხნის განათლება არის 0.1210.

საშუალო კლასის წარმომადგენელი რესპონდენტებისათვის:
16-იანი მნიშვნელობისათვის შესაბამისი Z იქნება:

$$Z = \frac{16 - 14.40}{3.04} = 0.53$$

არე $Z = 0.53$-სა და განაწილების კუდს შორის (გ სვეტი) არის 0.2981. ასე რომ,ალბათობა იმისა, რომ საშუალო კლასის წარმომადგენელ რესპონდენტს ექნება 16 წელზე მეტი ხნის განათლება არის 0.2981.

გ. დაბალი კლასის წარმომადგენელი რესპონდენტებისათვის:
12-იანი მნიშვნელობისათვის შესაბამისი Z იქნება:

$$Z = \frac{12 - 11.36}{2.96} = 0.22$$

არე $Z = 0.22$-სა და საშუალო მნიშვნელობას შორის არის 0.9871. ამ მნიშვნელობას უნდა დაეუმატოთ 0.50(განაწილების მეორე ნახევარი) და 0.0596. ასე რომ, ალბათობა იმისა, რომ დაბალი კლასის წარმომადგენელ რესპონდენტს ექნება 12 წელზე ნაკლები ხნის განათლება არის 0.1190.

დ. თავდაპირველად ვიპოვოთ საშუალომდე მდებარე ის Z,რომელსაც აქვს 25%-იანი 0.25-ის ტოლი მნიშვნელობა. ეს არის $Z = 0.68$-ს. ქვედა ზღვარი არის:

$$Y = \overline{Y} + Z\,(Sy) = 12.73 + (-0.68)\,(2.79) = 10.83$$

ზედა ზღვარი არის:

$$Y = \overline{Y} + Z\,(Sy) = 12.73 + 0.68\,(2.79) = 14.63$$

ასე რომ, მუშათა კლასის წარმომადგენელთა 50% ხედება 10.83-14.63 შუალედში.

ე. თუ განათლების წლების შესახებ მონაცემები იქნება დადებითი მიმართულებით განლაგებული მაღალი განათლების წლები იქნება ნაკლები ვიდრე ეს ნორმალური განაწილებისას

იქნებოდა. ეს ნიშნავს, მაგალითად, რომ მაღალი განათლების ალბათობების მნიშვნელობები დაბალი იქნებოდა.

7. ა. სამსახურის პრესტიჟულობის 60-იანი მნიშვნელობას შეესაბამება შემდეგი *Z*:

$$Z = \frac{60 - 44.70}{14.03} = 1.09$$

არე $Z = 1.09$-სა და განაწილების კუდს შორის არის 0.1379. ასე რომ, თეთრკანიანი რესპონდენტების დაახლოებით 14%-ს აქვს სამსახურის პრესტიჟულობის ქულის 60-ზე ნაკლები ქულა. რადენობრივად ეს შეესაბამება 151 თეთრკანიან რესპონდენტს (0.1379*1094) ჩვენს შერჩევაში.

ბ. სამსახურის პრესტიჟულობის 60 ქულას შეესაბამება შემდეგი *Z* მნიშვნელობა:

$$Z = \frac{60 - 40.75}{13.63} = 1.41$$

არე $Z = 1.41$-სა და განაწილების კუდს შორის არის 0.0793. ასე რომ, შავკანიანი რესპონდენტების დაახლოებით 8%-ს აქვს სამსახურის პრესტიჟულობის 60 ქულა. რაოდენობრივად ეს დაახლოებით 15 რესპონდენტია (0.0793×190).

გ. სამსახურის პრესტიჟულობის 30 ქულას შეესაბამება შემდეგი *Z* მნიშვნელობა:

$$Z = \frac{30 - 44.70}{14.03} = -1.05$$

არე $Z = -1.05$-სა და საშუალო მნიშვნელობას შორის არის 0.3531.
სამსახურის პრესტიჟულობის 70 ქულას შეესაბამება შემდეგი *Z* მნიშვნლობა:

$$Z = \frac{70 - 44.70}{14.03} = 1.80$$

არე $Z = 1.80$-სა და საშუალო მნიშვნელობას შორის არის 0.4641. ასე რომ, იმ თეთრკანიანი რესპონდენტების წილი, რომლებსაც აქვთ სამსახურის პრესტიჟულობის ქულა 30-სა და 70-ს შორის არის $0.8172\ (0.3531 + 0.4641)$. ე.ი. დაახლოებით 894 თეთრკანიან რესპონდენტს ჩვენს შერჩევაში აქვს 30-დან 70-მდე სამსახურის პრესტიჟულობის ქულა.

დ. სამსახურის პრესტიჟულობის 30 ქულას შეესაბამება შემდეგი *Z* მნიშვნელობა:

$$Z = \frac{70 - 40.75}{13.63} = -0.79$$

არე $Z = -0.79$-სა და საშუალო მნიშვნელობას შორის არის 0.2852.
სამსახურის პრესტიჟულობის 60 ქულას შეესაბამება შემდეგი *Z* მნიშვნლობა:

$$Z = \frac{60 - 40.75}{13.63} = 1.41$$

არე $Z = 1.41$-სა და საშუალო მნიშვნელობას შორის არის 0.4207. ასე რომ იმ შავკანიანი რესპონდენტების წილი, რომელთაც სამსახურის პრესტიჟულობის 30-დან 60-მდე ქულა აქვთ არის 0.7059, ე.ი. დახლოებით 124 შავკანიან რესპონდენტს აქვს სამსახურის პრესტიჟულობის ქულა 30-დან 60-მდე საზღვრებში.

9. ა.

$$Z = \frac{2 - 1.28}{0.82} = 0.88$$

განაწილების დახლოებით 0.1894 ალემატება Z მნიშვნელობას, ივიეე ეხება დანაშაულების რაოდენობას 2-ზე მეტი მსხვერპლით.

ბ.

$$Z = \frac{1 - 1.28}{0.82} = -0.34$$

საშუალოსა და Z-ს შორის მდებარე არე არის დახლოებით 0.1331, ასე რომ მთლიანი არე ერთი მსხვერპლის შემთხვევებისათვის არის 0.50+0.1331=0.6331 ან 63.31%.

გ.

$$Z = \frac{4 - 1.28}{0.82} = 3.32$$

საშუალოსა და Z-ს შორის მდებარე არე არის დახლოებით 0.4995, ასე რომ მთლიანი არე ერთი მსხვერპლის შემთხვევებისათვის არის $0.50 + 0.4995 = 0.4995$.

11. ა. იმ გუნდისათვის, რომლის *APR* ქულა 975-ის ტოლია არის:

$$Z = \frac{975 - 945.81}{50.35} = 0.58$$

წიგნის *B* დამატების მიხედვით 0.58-მდე არე 0.2810-ის ტოლია.

ბ. Z-ის მნიშვნელობა, რომელიც შეესაბამება გუნდის იმ ქულას, რომლის შესაბამისი არე განაწილების კუდის 0.25-ს არის 0.67. აღნიშნულის გარდაქმნა გუნდის ქულად შემდეგნაირად მოხდება:

$$\frac{\text{კიდურა წერის მნიშვნელობა} - 945.81}{50.35} = 0.67$$

$$\text{კიდურა წერის მნიშვნელობა} \cong 980$$

შემდეგ უნდა დავამრგვალოთ აღნიშნული მაჩვენებელი, რომ მაღალი მნიშვნელობა, რომელიც ახლოს დგას Z-ის 0.67 მნიშვნელობასთან.

გ. Z-ის მნიშვნელობა უდრის 0.67-ს.

13. ნებისმიერი Z განაწილებისათვის საშუალო ნოლის ტოლია. Z განაწილების სტანდარტუ

ლი გადახრა 1-ის ტოლია. Z განაწილებები ცენტრირებულია მისი საშუალო მნიშვნელობის მიმართ და მისი განსაზღვრების მიხედვით მისი საშუალო ნოლის ტოლია. Z-ის 1-თან ან -1-თან ტოლობა ნიშნავს რომ ეს მნიშვნელობები მის თავდაპირველ მნიშვნელობებს შეესაბამება საშუალოდან 1 სტანდარტული გადახრის მანძილზე.

15. ა. ნორმალური განაწილების მრუდი ზარის ფორმის მრუდია. ეს ნიშნავს, რომ განაწილების ნახევარი ხვდება ცენტრის ერთ მხარეს, ხოლო მეორე ნახევარი მეორე მხარეს. ცენტრი არის წერტილი სადაც თავს იყრის მოდა, მედიანა და საშუალო მნიშვნელობა. განაწილება იწოდება ნორმალურად, რადგან ბევრი ემპირიულად ნანახი განაწილებები უახლოვდება მას.

ბ. ზოგადად, Z-ის მნიშვნელობების სტანდარტიზებული მნიშვნელობები გვაძლევს საშუალებას გამოვსახოთ მანძილი საშუალოსა და კრძო დაკვირვების წერტილს შორის. Z მნიშვნელობა არის სტანდარტული გადახრის ერთეულების რიცხვი მისი საშუალო მნიშვნელობიდან. უარყოფითი Z აღნიშნავს მისი ეს კონკრეტული მნიშვნელობა დაბალია მის საშუალო მნიშვნელობაზე, ხოლო დადებითობა კი ნიშნავს, რომ იგი აღემატება მის საშუალო მნიშვნელობას.

თავი 7. შერჩევა და შერჩევითი განაწილებები

1. ა. მიუხედავად იმისა, რომ მთელი მოსახლეობის შესახებ მონაცემების შეკრება რთულია, აღწერის მონაცემები მიჩნეულია სრულად, ასე რომ საშუალო ასაკი იქნება პარამეტრი.

ბ. სტატისტიკა, რადგან გამოთვლილია შერჩევითი ერთობლიობიდან.

გ. სტატისტიკა, რადგან გამოთვლილია შერჩევითი ერთობლიობიდან.

დ. პარამეტრი, რადგან სკოლას აქვს მონაცემები მის ყველა დასაქმებულზე.

ე. პარამეტრი, რადგან სკოლას აქვს მის ყველა მოსწავლეზე ინფორმაცია.

3. ა. იმ დაშვებით, რომ პოპულაციას წარმოადგენს იმ ხალხის ერთობლიობას, რომლებიც მოცეშულ დღეს ვაჭრობენ სავაჭრო ცენტრში, იგი ახდენს სიმეტრიული შერჩევითი ერთობლიობის შერჩევას.

ბ. ეს არის სტრატიფიცირებული შერჩევა, რადგან ხმები თავდაპირველად დაჯგუფებულ იქნა ქვეყნების მიხედვით. ჩვენ შეგვიძლია დავუშვათ, რომ ეს იყო ალბათური შერჩევა, მაგრამ ჩვენ არ ვიცით თუ როგორ იქნა სიიდან 50 ამომრჩეველი შერჩეული. იმ დაშვებით, რომ პოპულაციას მთელი ამერიკის მოსახლეობა წარმოადგენს ამ სახის შერჩევა წარმოადგენს არაალბათურ შერჩევას.

გ. ეს არ არის არც მარტივი შემთხვევითი შერჩევა არც სიმეტრიული შემთხვევითი შერჩევა. შესაძლოა ვიფიქროთ, რომ შერჩევა სტრატიფიცირებულია გვარის მიხედვით, მაგრამ შემდეგ პირველი 20 სახელის არჩევა ალარ არის შემთხვევითი შერჩევის პროცესი.

დ. ეს არ არის ალბათული შერჩევა. ნაცვლად ამისა, ეს არის გამიზნული შერჩევა რომლის მიზანიც ნიუ-იორკის მოსახლეობის შერჩევაა.

5. კავშირი სტანდარტულ გადახრასა და სტანდარტულ შეცდომას შორის შემდეგია $\sigma_{\bar{y}} = \sigma_y / \sqrt{N}$, სადაც $\sigma_{\bar{y}}$ სტანდარტული შეცდომაა, ხოლო σ_y სტანდარტული გადახრაა. რადგან $\sigma_{\bar{y}}$ იყოფა $\sqrt{N}$-ზე $\sigma_{\bar{y}}$ ყოველთვის მცირე იქნება σ_y-სთან შედარებით, გარდა იმ მარტივი შემთხვევისა, როცა $N = 1$.

თეორიულად საშუალოს დისპერსია უნდა იყოს ნაკლები, ვიდრე ნებისმიერი ცალკეული მნიშვნელობების დისპერსია. ეს გულისხმობს, რომ საშუალოს სტანდარტული შეცდომა ნაკლებია სტანდარტულ გადახრაზე.

7. ა. გამოკითხვები არ არის ალბათური შერჩევები. ტელევიზიის მიერ არ ხდება შერჩევა ნომერ 800 დამრეკავ აბონენტებს შორის.

ბ. პოპულაცია არის ყველა ის, ვინც უყურა ტელევიზიას და ნახა 800 ნომრის რეკლამა.

9. ა. ეს არ არის შემთხვევითი შერჩევა. სტუდენტები, რომლებიც მიირთმევენ ლანჩს სამშაბათობით, არ წარმოადგენენ აუცილებლად სკოლის ყველა სტუდენტს და არ გვაქვს გზა გამოთვალოთ ნებისმიერი სტუდენტის ჩართვის ალბათობა. ზოგი სტუდენტი შესაძლოა საერთოდ არ ჭამდეს ლანჩს კაფეტერიაში და შესაბამისად ასეთ სტუდენტს არ აქვს შანსი მოხვდეს თქვენს შერჩევაში.

ბ. ეს არის სიმეტრიული შემთხვევითი შერჩევა, რადგან სახელების შერჩევა სისტემატურად ხდება ჩარიცხული სტუდენტების სიიდან.

გ. თავდაპირველად შესაძლოა მოგვეჩვენოს, რომ ეს არის სიმეტრიული შერჩევა, როგორც ეს ბ-შია, მაგრამ აქაც იგივე პრობლემაა რაც კომპიუტერიის მაგალითის შემთხვევაში იყო. სტუდენტურ კავშირთან ჩაცლა გარკვეულ სტუდენტებს ამ შერჩევაში მოხვედრის შანს არ აძლევს.

დ. მეორე პროცედურა (ყოველი მე-10 სტუდენტის არჩევა ჩარიცხული სტუდენტების სიიდან) ყველაზე კარგი არჩევანია, რადგან იგი უზრუნველყოფს შემთხვევით შერჩევას.

11. ა. *საშუალო = 5.3; სტანდარტული გადახრა = 3.27.*

ბ. არის 10 საშუალო შერჩევითი ერთობლიობიდან 3: 6.33, 5.67, 3.33, 5.00, 7.33, 2.33, 6.00, 6.33, 7.00, 3.00.

გ. ათი შერჩევის საშუალო 5.23-ია, ხოლო სტანდარტული გადახრა 1.76. შერჩევის საშუალო ახლოს დგას პოპულაციის საშუალოსთან. შერჩევის სტანდარტული გადახრა ნაკლებია პოპულაციის სტანდარტულ გადახრასთან შედარებით. შერჩევის საშუალოს სტანდარტული გადახრა სტანდარტული გადახრის შეფასებაა შერჩევის 3 ზომისათვის.

13. ა. სტანდარტული გადახრა შემდეგნაირად გამოითვლება:

$$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{N}} = \frac{93,500}{\sqrt{7}} = 35,283.02$$

ეს მნიშვნელობა ასახავს საშუალო გადახრას შერჩევების საშუალოების საშუალოდან. მას ასევე შერჩევის განაწილების სტანდარტულ გადახრასაც უწოდებენ.

ბ. იმ შემთხვევის გამორიცხვით, როცა $N = 1$, სტანდარტული გადახრა ყოველთვის ნაკლები იქნება პოპულაციის სტანდარტულ გადახრასთან შედარებით. ეს უკანასკნელი შემდეგნაირად გამოისახება:

$$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{N}}$$

განაწილების ფორმა ნორმალურია; ე.ი იმ შემთხვევაშიც კი როცა საქმე გვაქვს განაწილების გადახრილ ფორმასთან შერჩევის განაწილება ნორმალურია. შერჩევის შეცდომის შემცირებისათვის საჭიროა შერჩევის ზომის გაზრდა.

თავი 8. შეფასება

1. ა. საუკეთესო შეფასება 90%-იანი ნდობის ინტერვალის ფარგლებში ხვდება 11.52 %-14.88%-ს შორის. ეს ნიშნავს რომ არსებობს 100-დან 90 შანსი იმისა, რომ ჩვენ მიერ გამოთვლილ ნდობის ინტერვალში მოხვდება დანაშუალის მსხვერპლი. ამ მონაცემებისათვის გამოთვლილი ინტერვალი შეიცავს ან არ შეიცავს პოპულაციის საშუალოს.

$$\text{სტანდარტული შეცდომა} = \sqrt{\frac{(13.2)(86.8)}{1,105}} = 1,018$$

$$\begin{aligned}\text{ნდობის ინტერვალი} &= 13.2 \pm 1.65\ (1.018) \\ &= 13.2 \pm 1.68 \\ &= 11.52 - 14.88\end{aligned}$$

ბ.

$$\begin{aligned}\text{ნდობის ინტერვალი} &= 13.2 \pm 2.58\ (1.018)\\ &= 13.2 \pm 2.63\\ &= 10.57 - 15.83\end{aligned}$$

გ. თუ შერჩევის ზომას გავაორმაგებთ, ხოლო პროცენტული წილი 13.2-ის ტოლად დარჩება, ნდობის ინტერვალი შემცირდება $1/\sqrt{2}$ ან დაახლოებით 30%-ით.

დ. შერჩევის ზომა 1000 და 1500 რიგით არის სიზუსტესა და ხარჯებს შორის ალტერნატივა. ამ ზომის შერჩევა გვაძლევს ნდობის ინტერვალს დაახლოებით შეცდომის ±3%-ის ოდენობით 95%-იანი ნდობის ინტერვალისათვის, რომელიც საკმარისია უმრავლესი კვლევების შემთხვევაში.

3. ა.

$$S_p = \sqrt{\frac{(0.41)(1-0.41)}{1,501}} = 0.013$$

$$\begin{aligned}\text{ნდობის ინტერვალი} &= 0.41 \pm 1.96\ (0.013)\\ &= 0.41 \pm 0.025\\ &= 0.385 - 0.35\end{aligned}$$

ბ.

$$\begin{aligned}\text{ნდობის ინტერვალი} &= 0.41 \pm 2.58\ (0.013)\\ &= 0.41 \pm 0.034\\ &= 0.376\text{-}0.444\end{aligned}$$

გ. ამ შემთხვევაში ნდობის 95%-იან და 99%-იან ინტერვალებს შორის სხვაობა დიდი არაა შერჩევის დიდი ზომიდან გამომდინარე. პირველი ინტერვალი მხოლოდ 1/2-ით არის ნაკლები, ხოლო მეორე 2/3-ით. ის ორგანიზაციები, რომლებიც ყველაზე ფართო ნდობის ინტერვალს იყენებენ, ჩვეულებრივ 95% იღებენ. ჩვენი დასკვნა ამერიკელების აზრის შესახებ გლობალურ დათბობაზე დარჩება იგივე ნებისმიერ შემთხვევაში.

5.

$$\text{სტანდარტული შეცდომა} = \sqrt{\frac{(51)(49)}{5,490}} = 0,67\%$$

$$\begin{aligned}\text{ნდობის ინტერვალი} &= 51 \pm 1.96\ (0.67)\\ &= 49.69\% - 52.31\%\end{aligned}$$

ნდობის ინტერვალი 95%-ის ტოლად ავირჩიოთ, მაგრამ არ აქვს მნიშვნელობა ავირჩევთ 90%, 95%-იან ან 99% იან ნდობის ინერვალს, გამოთვლილ ინტერვალში მაინც მოხვდება რესპუბლიკელი კანდიდატის მონაცემების 50%-მდე. შესაბამისად, თქვენ არ შეგიძლიათ თქვათ, რომ სავარაუდოდ რესპუბლიკელი კანდიდატი მოიგებს მამაკაცი ამომრჩევლების ხმების მიხედვით.

7. ა.

$$S_p = \sqrt{\frac{(0.786)(1-0.786)}{1,303}} = 0.011$$

ნდობის ინტერვალი=0.786 1.96(0.011)=0.786 0.022=0.764-0.808 ან 76.4%-80.8%

ბ. ზედა სავარჯიშოს პასუხის მიხედვით ჩვენ ვიცით, რომ 90%-იანი ნდობის ინტერვალი იქნება უფრო ზუსტი ვიდრე ნდობის ინტერვალი, რომლის ქვედა ზღვარი 76.4%, ხოლო ზედა ინტერვალი 80.8%-ია. შესაბამისად 90%-იან ნდობის ინტერვალს აქვს უფრო დაბალი ქვედა ზღვარი, რომელიც მეტია 76.4%-ზე, ხოლო ზედა ზღვარი ნაკლებია 80.8%-ზე. დამატებით, ასევე ვიცით, რომ 99%-იანი ნდობის ინტერვალი უფრო ნაკლები სიზუსტის იქნება ვიდრე ა დავალებაში გამოთვლილი ნდობის ინტერვალი. შესაბამისად, 99%-იანი ნდობის ინტერვალის ქვედა ზღვარი 76.4%-ზე ნაკლები იქნება, ხოლო ზედა ზღვარი 80.8%-ზე მეტი.

9.

$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}} = \frac{1.70}{\sqrt{2,020}} = 0.038$$

$$\begin{aligned} \text{ნდობის ინეტრვალი} &= 1.96 \pm 1.65\ (0.038) \\ &= 1.94 \pm 0.063 \\ &= 1.877 - 2.003 \end{aligned}$$

11.

$$\text{სტანდარტული შეცდომა} = \sqrt{\frac{27(100-27)}{225}} = 2.96$$

$$\begin{aligned} \text{ნდობის ინტერვალი} &= 27 \pm 1.65\ (2.96) \\ &= 27 \pm 5.80 \\ &= 21.2\% - 32.8\% \end{aligned}$$

13. ა. იმ რესპონდენტების შემთხვევა, რომლებიც ფიქრობენ, რომ ჰომოსექსუალური ურთიერთობები ყოველთვის არასწორია:

$$S_p = \sqrt{\frac{(51.5)(100-51.5)}{1,269}} = 1.403$$

$$\begin{aligned} \text{ნდობის ინტერვალი} &= 51.5 \pm 1.96\ (1.403) \\ &= 51.5 \pm 2.75 \\ &= 48.75\% - 54.25\% \end{aligned}$$

იმ რესპონდენტების შემთხვევა, რომლებიც ფიქრობენ, რომ ჰომოსექსუალური ურთიერთობები ყოველთვის არასწორია:

$$S_p = \sqrt{\frac{(38.1)(100-38.1)}{1,269}} = 1.363$$

$$\begin{aligned} \text{ნდობის ინეტრვალი} &= 38.1 \pm 1.96\ (1.363) \\ &= 38.1 \pm 2.67 \\ &= 35.43\% - 40.77\% \end{aligned}$$

ბ.

$$S_p = \sqrt{\frac{(10)(100-10)}{1,269}} = 0.842$$

ნდობის ინტერვალი $= 10 \pm 1.96\ (0.842)$
$= 10 \pm 1.65$
$= 8.35\% - 11.65\%$

გ. რადგან 90%-იანი ნდობის ინტერვალი იმ რესპონდენტებისათვის, რომლებიც ფიქრობენ, რომ ჰომოსექსუალური ურთიერთობები ყოველთვის არასწორია არ მოიცავს 50%-ზე ნაკლებ მნიშვნელობებს, ჩვენ გადაჭრით არ შეგვიძლია ვთქვათ, რომ ამერიკელთა უმრავლესობა მიიჩნევს, რომ ჰომოსექსუალური ურთიერთობები ყოველთვის არასწორია.

15. ა. ლექტორმა ამ კითხვაზე პასუხისათვის უნდა წაახალისოს სტუდენტები, რომ მათ თავად გამოიცნონ ამ კითხვაზე პასუხი.

ბ. მამაკაცი რესპონდენტებისათვის:

$$S_p = \sqrt{\frac{(17)(100-17)}{623}} = 1.50$$

ნდობის ინეტრვალი $= 17 \pm 1.65\ (1.50)$
$= 17 \pm 2.48$
$= 14.52\% - 19.48\%$

გ. ქალი რესპონდენტებისათვის:

$$S_p = \sqrt{\frac{(35.6)(100-35.6)}{697}} = 1.81$$

ნდობის ინტერვალი $= 35.6 \pm 1.65\ (1.81)$
$= 35.6 \pm 2.99$
$= 32.61\% - 38.59\%$

დ. ლექტორმა ამ კითხვაზე პასუხისათვის უნდა წაახალისოს სტუდენტები, რომ მათ თავად გამოიცნონ ამ კითხვაზე პასუხი.

თავი 9. ჰიპოთეზების ტესტირება

გთხოვთ გაითვალისწინოთ, რომ ამ თავში ძალიან მცირე რიცხვების კვადრატში აყვანა და გაყოფა მოგვიწევს. ამან შეიძლება გამოთვლებში გარკვეული უზუსტობა გამოიწვიოს და სტუდენტებმა ქვემოთ მოყვანილი გამოთვლის შედეგებისაგან განსხვავებული შედეგები მიიღონ.

1. ა. $H_0: \mu_Y = 13.5$ წელი; $H_1: \mu_Y < 13.5$ წელი.

ბ. გამოთვლილი Z მნიშვნელობა არის -4.19. P-ს შესაბამისი მნიშვნელობა ნაკლებია 0.001-ზე ცალმხრივი ტესტისათვის. იგი ასევე ნაკლებია $\alpha = 0.01$ დონისათვის, ასე რომ ვახდენთ ნულოვანი ჰიპოთეზის უარყოფას და ვასკვნით, რომ ამერიკის ჯანდაცვის ორგანიზაციაში ექიმებს აქვთ ნაკლები გამოცდილება ვიდრე ექიმებს მთლიანად მთელს ჯანდაცვის ორგანიზაციებში.

3. ა. ორმხრივი ტესტი, $\mu_1 \neq \$50,303$; ნულოვანი ჰიპოთეზა, $\mu_1 = \$50,303$

ბ. ერთმხრივი ტესტი, 1>3.2; ნულოვანი ჰიპოთეზა, 1=3.2
გ. ერთმხრივი ტესტი, 1> 2; ნულოვანი ჰიპოთეზა, 1= 2
დ. ორმხრივი ტესტი, 1 §50,303; ნულოვანი ჰიპოთეზა, 1=§50,303
ე. ერთმხრივი ტესტი, 1> 2; ნულოვანი ჰიპოთეზა, 1= 2
ვ. ერთმხრივი ტესტი, 1< 2; ნულოვანი ჰიპოთეზა, 1= 2

5. ა. $H_0: \mu_1 = 37.7; H_1: \mu_1 \neq 37.7$

ბ. ორმხრივი ტესტის მიხედვით 0.05 მნიშვნელოვნების დონისათვის, t არის 25.88, ხოლო მისი P მნიშვნელობა ნაკლებია <0.001-ზე.

$$t = \frac{47.71 - 37.7}{\frac{17.35}{\sqrt{2,013}}} = \frac{10.01}{.39} = 25.67$$

გ. შეგვიძლია უარვყოთ ნულოვანი ჰიპოთეზა. არსებობს განსხვავება GSS-ის შერჩევის მონაცემების საშუალოსა და მთლიანად ამერიკელი მოზარდების საშუალო ასაკს შორის. შეგვიძლია დავასკვნათ, რომ ასაკთან მიმართებაში GSS-ის შერჩევა არ არის მთელი ამერიკელი მოზარდების რეპრეზენტატული.

7. ა. როდესაც საქმე გვაქვს პროცენტულ წილებთან შესაბამისი ტესტია Z.

ბ. გამოთვლილი Z არის -6, <0.0001. რადგან P *(0.0001)* < α *(0.05)*, უარვყოფთ ნულოვან ჰიპოთეზას. ეს უკანასკნელი ასახავს მტკიცებითი ქცევის მიმართ კონსერვატორებსა და ლიბერალების შეხედულებებს შორის სტატისტიკურ განსხვავებას. ლიბერალები უფრო მეტი ალბათობით დაუჭერენ მხარს სამუშაო ადგილზე მტკიცებით ქცევას ვიდრე კონსერვატორები:

$$S_{p_1 - p_2} = \sqrt{\frac{0.10(1-0.10)}{424} + \frac{0.28(1-0.28)}{336}} = 0.03$$

$$Z = \frac{0.10 - 0.28}{0.03} = -6$$

გ. ორმხრივი ტესტისათვის, P უნდა გამრავლდეს 2-ზე, $0.0001 \times 2 = 0.0002$. ჩვენი გადაწყვეტილება ნულოვანი ჰიპოთეზის უარყოფის შესახებ იგივე რჩება.

9. ა. ეს იქნება ცალმხრივი ტესტი. $H_0: \pi_1 = \pi_2; H_1: \pi_1 = \pi_2;$

ბ. რადგან გამოთვლილი Z-ის ალბათობა არის 0.0228 (შესამე სვეტი $Z = 2.000$), უარვყოფთ ნულოვან ჰიპოთეზას. შეგვიძლია დავასკვნათ, რომ არსებობს კავშირი რესპონდენტის სქესსა და ჰილარი კლინტონის მხარდაჭერას შორის. ქალების პროცენტული წილი, რომლებმაც მხარი დაუჭირეს კლინტონის კანდიდატურას გაცილებით მეტია კლინტონის მამაკაც მხარდაჭერებთან შედარებით.

$$S_{p_1 - p_2} = \sqrt{\frac{0.41(1-0.41)}{240} + \frac{0.49(1-0.49)}{381}} = 0.04$$

$$Z = \frac{0.41 - 0.49}{0.04} = -2.0$$

გ. თუ ალფას მნიშვნელობას შევცვლით 0.01-ის ტოლად, ნულოვანი ჰიპოთეზის უარყოფა ვერ მოხერხდება. მიღებული *Z*-ის *P* მნიშვნელობა აღემატება 0.01.-ს.

11. ა. ცალმხრივი ტესტი, $\alpha = 0.05$.

$$df = 149 + 306 - 2 = 453$$

$$S_{\bar{Y}_1-\bar{Y}_2} = \sqrt{\frac{(149-1)1.52^2+(306-1)0.79^2}{(149+306)-2}}\sqrt{\frac{149+306}{149(306)}} = 0.11$$

$$t = \frac{1.76-0.73}{0.11} = 9.36$$

t-ს მიღებული მნიშვნელობის 9.36-ის მიხედვით უარვყოფთ ნულოვან ჰიპოთეზას. აშშ-ს სტუდენტები საკურსო სამუშაოებისას იყენებენ უფრო დიდი ხნით ინტერნეტს ვიდრე ინდოელი სტუდენტები.

ბ. უნდა შევნიშნოთ, რომ ვარიაციების მნიშვნელობები ერთმანეთისაგან განსხვავდება. სტანდარტული შეცდომა და თავისუფლების ხარისხი უნდა გამოთვალოთ ფორმულა 9.8 და 9.9-ის გამოყენებით.

$$S_{\bar{Y}_1-\bar{Y}_2} = \sqrt{\frac{1.91^2}{149}+\frac{0.78^2}{306}}$$
$$= \sqrt{.024483892+.001988235}$$
$$= \sqrt{.026472127} = .16$$

$$df = \frac{(1.91^2/149+0.78^2/306)^2}{(1.91^2/149)/148+(0.78^2/306)/305} = \frac{0.00070077356}{0.000171951} = 4.075$$

$$t = \frac{(2.08-0.87)}{0.16} = 7.56$$

t-ს მიღებული მნიშვნელობის 7.56-ის მიხედვით, უარვყოფთ ნულოვან ჰიპოთეზას.

13.

$$S_{p_1-p_2} = \sqrt{\frac{0.18(1-0.18)}{670}+\frac{0.14(1-0.14)}{723}} = 0.02$$

$$Z = \frac{0.18-0.14}{0.02} = 2$$

Z(2)-ის მიღების ალბათობაა $0.0228 \times 2 = 0.0456$. *P*-ს მნიშვნელობა უფრო მეტია ვიდრე ალფა, ასე, რომ უარვყოფთ ნულოვან ჰიპოთეზას.

თავი 10. ორ ცვლადს შორის კავშირი: ჯვარედინი გამოთვლები

1. ა. დამოუკიდებელი ცვლადი რასაა; დამოკიდებული ცვლადი ღამით სიარნობის შიში

ღამის სიარნობის შიში	რასა	
	შვკანიანი	თეთრკანიანი
დიახ	3	4
არა	5	9

ბ. თეთრკანიანი რესპონდენტების დაახლოებით 69%(69.2%)-ს არ ეშინია ღამით სახლის საიხლოვეს სეირნობა, ხოლო იგივე მაჩვენებელი შავკანიანი რესპონდენტებისათვის 63%(62.5%)-ია. ამ მაჩვენებლებს შორის მხოლოდ მცირედი 7%-ით განსხვავება ცვლადებს შორის სუსტი სიმჭიდროვის კავშირზე მიუთითებს.

ღამის სიარნობის შიში	რასა	
	შავკანიანი	თეთრკანიანი
დიახ	37.5%	30.8%
არა	62.5%	69.2%

გ. ღამით სახლის სიახლოვეს სეირნობის შიშის მიმართ არის განსხვავება იმ რესპონდენტებს შორის, რომელთაც აქვთ საკუთარი სახლი და იმ რესპონდენტებისათვის, რომელთაც ქირით აქვთ საცხოვრებელი სახლი. მთლიანობაში სახლის მფლობელთა 25.0%-ს, ხოლო ქირით მცხოვრებთა 38.5%-ს ეშინია ღამით სეირნობა. ამ ორ ჯგუფს შორის განსხვავება მხოლოდ 13.5%-ია. შესაბამისად, კავშირის სიმჭირდოვე სახლის მფლობელობასა და ქირით ცხოვრებას შორის ღამით სეირნობის შიშის მისამართით სუსტიდან საშუალომდე სიძლიერისაა.

ღამის სიარნობის შიში	სახლის ფლობა	
	დიახ	არა
დიახ	2	5
	25.0%	38.5%
არა	6	8
	75.0%	61.5%

3. ა. სტუდენტის არგუმენტის მიხედვით, დამოკიდებული ცვლადი პოლიტიკური შეხედულებებია, ხოლო დამოუკიდებელი ცვლადი კი ჰომოსექსუალობის მიმართ დამოკიდებულება.

ბ. გამოთვლისათვის გვჭირდება სვეტის ბოლო უჯრაში მოცემული მნიშვნელობა(435) გავყოთ ჯამზე(800) და შემდეგ მიღებული განაყოფი გავამრავლოთ 100-ზე. მივიღებთ 54.4%-ს.

გ. იმ კონსერვატორთაგან, რომლებსაც მიაჩნიათ, რომ რომ ჰომოსექსუალური ურთიერთობები ყოველთვის არასწორია 48.3% შეადგენს, ხოლო მათ 17.3%-ს კი მიაჩნია, რომ ჰომოსექსუალური ურთიერთობები შესაძლებელია. 31%-იანი სხვაობა მიანიშნებს, რომ არსებობს საშუალო სიძლიერის კავშირი. ასევე, იგივე შედარება შეგვიძლია გავაკეთოთ ლიბერალებს შემთხვევაშიც და მივიღებთ 31.8%-ის სხვაობას.

5. ა. იმ თეთრკანიანი რესპონდენტების წილი, რომელთაც მოიხმარეს ალკოჰოლი 0.78. ეს შედეგი მიღებულ იქნა იმ რესპონდენტების რიცხვზე გაყოფით, რომლებმაც უპასუხეს „დიახ" 326, სვეტის ჯამზე 416-ზე გაყოფით. შავკანიანი რესპონდენტების წილი არის 0.69, ხოლო ესპანური წარმოშობის მამაკაცი რესპონდენტებისათვის ეს მაჩვენებელი 0.74-ია.

ბ. ზოგადად, თითოეულ რასობრივ ჯგუფში რესპონდენთა უმრავლესობამ მათივე პასუხების მიხედვით მოიხმარა ალკოჰოლი. შევნიშნოთ, რომ ეს მაჩვენებელი არ ასახავს ალკოჰოლის ახლანდელი მოხმარების დონეს. მამაკაცებში ალკოჰოლის მოხმარების ყველაზე მაღალი დონე თეთრკანიან რესპონდენტებს აქვთ.

7. ა. ქალების შემთხვევაში მათი წილი ვინც აღნიშნა, რომ ნამდვილად ჩაბარებდნენ კოლეჯში გაიზარდა 33.6-დან 62.4-მდე. ეს არის 28.8%-ის ზრდა 21 წელიწადში. მეორე დრამატული ცვლილება არის იმ რესპონდენტებისათვის, რომლებმაც დააფიქსირეს, რომ ალბათ ჩააბარებენ კოლეჯში 1980 წლის 45%-დან 2001 წლის 17.5%-მდე. გვაქვს 27.5%-იანი კლება.

ბ. ეს ტენდენცია ასევე სწორია მამაკაცების შემთხვევაშიც. ყველაზე დიდი ცვლილებები არის კატეგორიაში „ნამდვილად ჩავაბარებ" და „ალბათ არ ჩავაბარებ". ბიჭების შემთხვევაში ცვლილება ნაკლებია ვიდრე ეს გოგონების შემთხვევაშია, მაგალითად იმ ბიჭების შემთხვევაში, რომლებმაც დაფიქსირეს, რომ „ნამდვილად ჩავაბარებ" და „ალბათ ჩავაბარებ" 1990 წლიდან 2001 წელთან შედარებით არის 30.2-24.1=6.1%. იგივე პერიოდისათვის გოგონებისათვის სხვაობა 28.8-17.5=11.3%-ია.

9. თეთრკანიანი სტუდენტების რიცხვი კერძო სკოლებში რჩება 83%-ის ფარგლებში, რაც საჯარო სკოლის მაჩვენებელს აღემატება. საჯარო დაწყებითი სკოლების შემთხვევაში იყო 5.7%-იანი კლება 80.9%-იდან 75.2%-მდე. ასევე ფიქსირდება კლება კერძო დაწყებითი სკოლების შემთხვევაშიც 90.7%-დან 83.0%-მდე. სტუდენტების რასობრივი შემადგენლობა იზრდება წლების მანძილზე და ეს ზრდა უფრო მეტია კერძო სკოლების შემთხვევაში.

11. ა. იმ რესპონდენტების 69.1% რომლებმაც ნახეს *X* რეიტინგის კინო გასულ წლებში, დააფიქსირეს რომ ექსტრამარიტალური სექსი არასწორია, ხოლო იმ რესპონდენტების 83.7%-მა რომლებსაც არ უნახიათ *X* რეიტინგის კინო გასულ წლებში, დააფიქსირეს რომ ექსტრამარიტალური სექსი არასწორია. ასე რომ გვაქვს 14.6%-იანი სხვაობა, შესაბამისად ეს მეტყველებს მხოლოდ სუსტ მხარდაჭერაზე ორგანიზაციული კუთვნილების მხრივ.

		ნახა X რეიტინგის ფილმი გასულ წელს		
		დიახ	არა	სულ
ექსტრამარიტალური სექსი	ყოველთვის არასწორია	69.1%	83.7%	80.1%
	თითქმის ყოველთვის არასწორია	17.9%	10.9%	12.6%
	ხანდახან არასწორია	9.4%	3.7%	5.1%
	სერთოდ არ არის არასწორი	3.6%	1.7%	2.2%
ჯამი		100.00%	100.00%	100.00%

ბ. იმ მამაკაცი რესპონდენტების 67.5%-მა, რომლებმაც ნახეს X რეიტინგის კინო გასულ წლებში, დააფიქსირეს რომ ექსტრამარიტალური სექსი არასწორია, ხოლო იმ რესპონდენტების 70.3%-მა რომლებსაც არ უნახიათ X რეიტინგის კინო გასულ წლებში, დააფიქსირეს რომ ექსტრამარიტალური სექსი არასწორია. ასე რომ გვაქვს 11.8%-იანი სხვაობა, რაც სუსტ კავშირს ასახავს.

გ. ეს არ არის პირობითი კავშირის მაგალითი, რადგან სქესის მიუხედავად პორნოგრაფიული

მასალების გამოყენებასა და ექსტრამარიტალური სექსის შესახებ აზრს შორის ურთიერთ სუსტია. შეგვიძლია დავასკვნათ, რომ მიუხედავად არსებული ინფორმაციისა, ჩვენ ეს კავშირი არ არის სქესის მიხედვით განსხვავებული.

			ნაბა X რეიტინგის ფილმი გასულ წელს		
			დიახ	არა	სულ
მამაკაცი	ექსტრამარიტალური სექსი	ყოველთვის არასწორია	67.5%	79.3%	75.2%
		თითქმის ყოველთვის არასწორია	20.5%	13.1%	15.8%
		ხანდახან არასწორია	8.5%	5.0%	6.2%
		სერთოდ არ არის არასწორი	3.3%	2.5%	2.8%
ჯამი			100.1%	99.9%	100.0%
ქალი	ექსტრამარიტალური სექსი	ყოველთვის არასწორია	72.0%	86.5%	84.2%
		თითქმის ყოველთვის არასწორია	12.7%	9.5%	10.0%
		ხანდახან არასწორია	11.0%	2.9%	4.2%
		სერთოდ არ არის არასწორი	4.2%	1.1%	1.6%
ჯამი			99.9%	100.0%	100.0%

13. ა. ეს საკითხი სხვადასხვაგვარად შეიძლება განვიხილოთ. აზრი იმის შესახებ, რომ მეცნიერებას მოაქვს მეტი ზიანი ვიდრე სარგებელი გარემოსათვის ძალიან ზოგადია. თუ დედუქციურად ვიფიქრებთ აზრი იმის შესახებ, რომ მეცნიერება ჭრის გარემოს დაცვის საკითხებს გამომდინარეობს იმ დაშვებიდან, რომ მეცნიერების განვითარებას უფრო მეტი ზიანი მოაქვს გარემოსათვის ვიდრე სიკეთე.

ბ. არსებობს ორი სცენარი ცვლადების მიმართების მიხედვით.

მეცნიერება წყვეტს გარემოს დაცვის პრობლემებს	მეცნიერებას უფრო მეტი ზიანი მოაქვს ვიდრე სარგებელი					
	ბოლომდე ეთანხმები	ეთანხმები	არც ეთანხმები არც უარყოფ	არ ეთანხმები	კატეგორიულად არ ეთანხმები	ჯამი
ბოლომდე ეთანხმები	22.7%	4.4%	2.7%	2.8%	9.2%	5.3%
ეთანხმები	20.8%	34.9%	18.1%	24.9%	22.2%	24.6%
არც ეთანხმები არც უარყოფ	11.4%	17.7%	27.6%	20.4%	16.3%	22.6%
არ ეთანხმები	23.0%	33.2%	30.4%	40.7%	29.6%	34.4%
კატეგორიულად არ ეთანხმები	22.1%	9.8%	11.1%	11.2%	22.7%	13.3%
ჯამი	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

მეცნიერება წვლილს გარემოს დაცვის პრობლემებს	მეცნიერებას უფრო მეტი ზიანი მოაქვს ვიდრე სარგებელი					
	ბოლომდე ეთანხმები	ეთანხმები	არც ეთანხმები არც უარყოფ	არ ეთანხმები	კატეგორიულად არ ეთანხმები	ჯამი
ბოლომდე ეთანხმები	22.7%	4.3%	2.6%	3.4%	8.5%	5.1%
ეთანხმები	15.6%	25.9%	14.3%	17.6%	1.4%	18.2%
არც ეთანხმები არც უარყოფ	11.7%	16.3%	36.8%	19.6%	18.5%	22.1%
არ ეთანხმები	20.9%	38.8%	34.6%	45.4%	32.1%	38.3%
კატეგორიულად არ ეთანხმები	29.1%	14.7%	11.7%	14.0%	27.6%	16.2%
ჯამი	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

გ. თუ ცვლადს „მეცნიერებას მოაქვს გარემოსათვის მეტი ზიანი" მივიჩნევთ დამოუკიდებელ ცვლადად, იმ რესპოდენტების 22.7%, სავსებით ეთანხმებიან აზრს, რომ მეცნიერებას მოაქვს მეტი ზიანი გარემოსათვის ამავდროულად ეთანხმებიან რომ მეცნიერებას შეუძლია გარემოს დაცვის პრობლემების გადაჭრა, ხოლო იმ რესპონდენტების მხოლოდ 2.7%, რომლებიც ეთანხმებიან აზრს იმის შესახებ, რომ მეცნიერებას შეუძლია გარემოს დაცვის საკითხების გადაჭრა ამავდროულად არიან ინდიფერენტულნი იმის შესახებ მოაქვს თუ არა მეცნიერების განვითარებას დადებითი ან უარყოფითი შედეგები გარემოს პრობლემებისადმი.

თავი 11. ხი-კვადრატ კრიტერიუმი

1. ა. თავისუფლების ხარისხი $= (2 - 1)(2 - 1) = 1$

ბ. ხი-კვადრატი $= 59.25$ (იეტსის კორექციით 58.22). ამ შემთხვევაში გამოთვლილი ხი-კვადრატის ალბათობა ნაკლებია ალფაზე (0.001). შეგვიძლია უარვყოთ ნულოვანი ჰიპოთეზა და დავასკვნათ, რომ ღამით სეირნობის შიში და სქესი დამოკიდებული ცვლადებია. უფრო მაღალმა პროცენტმა ქალების (45.1%) დააფიქსირა, რომ ეშინია ღამით სეირნობის ვიდრე იგივე პასუხზე მამაკაცებმა (21.5%).

შიში და სქესი	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
მამაკაცი/დიახ	94	150.8	56.8	3,226.24	21.39
მამაკაცი/არა	343	286.2	56.8	3,226.24	11.27
ქალი/დიახ	242	185.2	56.8	3,226.24	17.42
ქალი/არა	295	351.8	56.8	3,226.24	9.17

$\chi^2 = 59.25$

იეტსის კორექციის კოეფიციენტით:

შიში და სქესი	$\lvert f_0 - f_e \rvert$	$(\lvert f_0 - f_e \rvert - .50)^2$	f_e	$(f_0 - f_e)^2$	$\frac{(\lvert f_0 - f_e \rvert - .5)^2}{f_e}$
მამაკაცი/დიახ	94	150.8	56.8	3,226.24	21.39
მამაკაცი/არა	343	286.2 56.8	56.8	3,226.24	11.27
ქალი/დიახ	242	185.2 56.8	56.8	3,226.24	17.42
ქალი/არა	295	351.8	56.8	3,226.24	9.17
ჯამი					58.22

გ. თუ ალფას მნიშვნელობას შეცვლით 0.1-ით, მაინც შეგეძლებთ ნულოვანი ჰიპოთეზის უარყოფას. გამოთვლილი ხი-კვადრატისათვის შესაბამისი ალბათობა ალფაზე ნაკლებია.

3. ა. შავკანიანი რესპონდენტების მაღალმა პროცენტმა 42.8%(59/138) დააფიქსირა ღამით სეირნობის შიში. იგივე მაჩვენებელი თეთრკანიანი რესპონდენტების შემთხვევაში 31.5%(219/696).

ბ. რასის მიუხედავად ქალების მიერ ღამით სეირნობის შიში უფრო მაღალია კაცებთან შედარებით.

გ.

თეთრკანიანები, $\chi^2 = 41.40$

შავკანიანები, $\chi^2 = 23.80$

ორივე შემთხვევაში უარვყოფთ ნულოვან ჰიპოთეზას და ვასკვნით, რომ სქესი და ღამით სეირნობის შიში დამოკიდებული ცვლადებია.

თეთრკანიანი რესპონდენტებისათვის

შიში და სქესი	f_0	f_e	$f_0 - f_e$	$(f_0 - f_e)^2$	$\frac{(f_0 - f_e)^2}{f_e}$
მამაკაცი/დიახ	59	98.2	39.2	1,536.64	15.65
მამაკაცი/არა	253	213.8	39.2	1,536.64	7.19
ქალი/დიახ	160	120.8	39.2	1,536.64	12.72
ქალი/არა	224	263.2	39.2	1,536.64	5.84

$\chi^2 = 41.40$

იეტსის კორექციის კოეფიციენტით ხი-კვადრატი იქნება 40.35, როგორც ეს ქვემოთაა ნაჩვენები:

შიში და სქესი	$\lvert f_0 - f_e \rvert$	$(\lvert f_0 - f_e \rvert - .50)^2$	f_e	$\frac{(\lvert f_0 - f_e \rvert - .5)^2}{f_e}$
მამაკაცი/დიახ	39.2	1,497.69	98.2	15.25
მამაკაცი/არა	39.2	1,497.69	213.8	7.01
ქალი/დიახ	39.2	1,497.69	120.8	12.40
ქალი/არა	39.2	1,497.69	263.2	5.69

$\chi_c^2 = 40.35$

შავკანიანი რესპონდენტებისათვის

შიში და სქესი	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
მამაკაცი/დიახ	7	20.5	13.5	182.25	8.89
მამაკაცი/არა	41	27.5	13.5	182.25	6.63
ქალი/დიახ	52	38.5	13.5	182.25	4.74
ქალი/არა	38	51.5	13.5	182.25	3.54
			$\chi^2 = 23.80$		

იეტსის კორექციის კოეფიციენტით ხი-კვადრატი იქნება 22.06, როგორც ეს ქვემოთაა ნაჩვენები:

შიში და სქესი	$\lvert f_o - f_e \rvert$	$(\lvert f_o - f_e \rvert - .50)^2$	f_e	$\frac{(\lvert f_o - f_e \rvert - .5)^2}{f_e}$
მამაკაცი/დიახ	13.5	169	20.5	8.24
მამაკაცი/არა	13.5	169	27.5	6.15
ქალი/დიახ	13.5	169	38.5	4.39
ქალი/არა	13.5	169	51.5	3.28

5. ა. ხი-კვადრატი = *5.93*; *df* = *8*.

დასახლება/GWEED	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
მცირე/ალბათ შეუძლებელი	24	19.8	4.2	17.64	0.89
მცირე/ძალიან რთული	13	13.7	0.7	0.49	0.04
მცირე/საკმაოდ რთული	31	27.0	4.0	16.00	0.59
მცირე/საკმაოდ ადვილი	78	78.9	0.9	0.81	0.01
მცირე/ძალიან ადვილი	196	202.6	6.6	43.56	0.22
საშუალო/ალბათ შეუძლებელი	7	9.8	2.8	7.84	0.80
საშუალო/ძალიან რთული	7	6.8	0.2	0.04	0.01
საშუალო/საკმაოდ რთული	11	13.3	2.3	5.29	0.40
საშუალო/საკმაოდ ადვილი	45	38.8	6.2	38.44	0.99
საშუალო/ძალიან ადვილი	98	99.5	1.5	2.25	0.02
დიდი/ალბათ შეუძლებელი	8	9.4	1.4	1.96	0.21
დიდი/ძალიან რთული	7	6.5	0.5	0.25	0.04

დიდი/საკმაოდ რთული	11	12.8	1.8	3.24	0.25
დიდი/საკმაოდ ადვილი	32	37.4	5.4	29.16	0.78
დიდი/ძალიან ადვილი	104	95.9	8.1	65.61	0.68
	$\chi^2 = 5.93$				

ბ. ცხრილის არცერთ უჯრაში მოცემული სიხშირე არ არის 5-ზე ნაკლები.

გ. ჩვენი გამოთვლილი ხი-კვადრატის მნიშვნელობა ეკუთვნის 0.50-0.70 შუალედს. ჩვენ არ შეგვიძლია ნულოვანი ჰიპოთეზის უარყოფა და შესაბამისად ვასკვნით, რომ საცხოვრებელი ადგილი და მარიჰუანასადმი ხელმისაწვდომობა დაკავშირებული არ არის. რესპონდენტების უმრავლესობისათვის (59% და მეტი), მიუხედავად საცხოვრებელი ქალაქის ზომისა, მარიჰუანის შოვნა „საკმაოდ მარტივი" იყო. ეს მაჩვენებელი განსკუთრებით მაღალი დიდ ქალაქებში მაცხოვრებლებისათვის არის(64.2%).

7. ა.

საკლესიო მსახურებებზე დასწრება/ჰომოსექსუალობის შესახებ შეხედულებები	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
არასოდეს/ მიუღებელია	163	231.1	68.1	4,637.61	20.07
არასოდეს/საერთოდ მიუღებელი არ არის	215	146.9	68.1	4,637.61	31.57
რამდენჯერმე/მიუღებელია	101	101.5	0.5	0.25	0.00
რამდენჯერმე/საერთოდ მიუღებელი არ არის	65	64.5	0.5	0.25	0.00
ყოველ კვირას/მიუღებელია	266	197.5	68.5	4,692.25	23.76
ყოველ კვირას/საერთოდ მიუღებელი არ არის	57	125.5	68.5	4,692.25	37.39
	$\chi^2 = 112.79$				

ხი-კვადრატი= 112.79, თავისუფლების ორი ხარისხით, ჩვენ მიერ გამოთვლილი ხი-კვადრატის მნიშვნელობა ნაკლებია 0.001-ზე, უარვყოფთ ნულოვან ჰიპოთეზას.

ბ. რაც უფრო იზრდება საეკლესიო მსახურებებზე დასწრების სიხშირე, მით უფრო იზრდება იმის შანსი რომ რესპონდენტი უარყოფითად შეაფასებს ჰომოსექსუალურ ურთიერთობებს. იმ რესპონდენტების 80%-ზე მეტად, რომლებიც ყოველკვირეულად ესწრებიან საკლესიო მსახურებებს ჰომოსექსუალურ ურთიერთობებს აფასებენ, როგორც „ყოველთვის არასწორს" იმ 43%-თან შედარებით, რომლებიც არასოდეს არ ესწრებიან საკლესიო მსახურებებს.

9.

რასა/პირველი თაობის სტუდენტის სტატუსი	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
თეთრკანიანი /პირველი თაობის	1,742	1,749.6	7.6	57.76	0.03
თეთრკანიანი /სხვა თაობის	2,392	2,384.4	7.6	57.76	0.02
შავკანიანი/ პირველი თაობის	102	93.5	8.5	72.25	0.77
შავკანიანი/ სხვა თაობის	119	127.5	8.5	72.25	0.57

ადგილობრივი ამერიკელი/ პირველი თაობის	41	36.4	4.6	21.16	0.58
ადგილობრივი ამერიკელი/სხვა თაობის	45	49.6	4.6	21.16	0.43
ესპანური წარმოშობის/პირველი თაობის	19	18.6	0.4	0.16	0.01
ესპანური წარმოშობის/ სხვა თაობის	25	25.4	0.4	0.16	0.01
აზიური წარმოშობის/ პირველი თაობის	6	11.9	5.9	34.81	2.93
აზიური წარმოშობის/ სხვა თაობის	22	16.1	5.9	34.81	2.16
	$\chi^2 = 7.51$				

ხი-კვადრატი=7.51. თავისუფლების ხარისხი=(2-1)(5-1)=4.
ვერ ვუარყოფთ ნულოვან ჰიპოთეზას, ჩვენ მიერ გამოთვლილი ხი-კვადარატის მნიშვნელობის ალბათობა ეკუთვნის 0.20-0.10 შუალედს.

11.

GPA/მარიჯუანას მოხმარება	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
დადებითი/GPA ≤ 2.0	76	60.4	15.6	243.36	4.03
უარყოფითი/GPA ≤ 2.0	29	29.5	0.5	0.25	0.01
მკვეთრად უარყოფითი/GPA ≤ 2.0	27	42.0	15	225.00	5.36
დადებითი/2.0 < GPA ≤ 3.0	261	243.6	17.4	302.76	1.24
უარყოფითი/2.0 < GPA ≤ 3.0	105	119.1	14.1	198.81	1.67
მკვეთრად უარყოფითი/2.0 < GPA ≤ 3.0	166	169.3	3.3	10.89	0.06
დადებითი/3.0 < GPA ≤ 4.0	293	326.0	33	1089.00	3.34
უარყოფითი/3.0 < GPA ≤ 4.0	174	159.4	14.6	213.16	1.34
მკვეთრად უარყოფითი/3.0 < GPA ≤ 4.0	245	226.6	18.4	338.56	1.49
	$\chi^2 = 18.54$				

ხი-კვადრატი = *18.54*. თავისუფლების 4 ხარისხით. ჩვენ მიერ გამოთვლილი ხი-კვადრატის მნიშვნელობა ნაკლებია 0.001-ზე, უარვყოფთ ნულოვან ჰიპოთეზას და ვასკვნით, რომ არსებობს კავშირი საშუალო აკადემიურ მოსწრებასა და მარიჯუანას მოხმარების მიმართ აზრს შორის. თუ ამ შედეგებს შევადარებთ მე-10 დავალებაში მიღებულ შედეგებს ჩანს რომ არსებობს ძლიერი სიმჭიდროვის კავშირი საშუალო აკადემიურ მოსწრებასა და მარიჯუანას მოხმარებისადმი აზრს შორის, ხოლო მე-10 დავალების შემთხვევაში GPA-სა და ალკოჰოლის მოხმარებისადმი აზრს შორის არ იყო კავშირი.

თავი 12. ნომინალურ და ორდინალურ ცვლადებს შორის კავშირის გაზომვა

1. ა. ცვლადებს შორის კავშირის საზომები გამოიყენება ცვლადებს შორის სტატისტიკური კავშირის შეფასებისათვის. ეს მაჩვენებლები ასევე საშუალებას გვაძლევს განვსაზღვროთ შეცდომის შემცირების ზომა დამოუკიდებელი ცვლადის გამოყენების შესახებ.

ბ. როდესაც ვირჩევთ ცვლადებს შორის კავშირის საზომთაგან ერთ-ერთს, მხედველობაში უნდა მივიღოთ როგორც დამოკიდებული, ასევე დამოუკიდებელი ცვლადის საზომი შკალა. გამა განსხვავდება კენდალის ტაუ-*b*-საგან, პირველი მათგანი ითვალისწინებს დამოუკიდებული და დამოუკიდებელი ცვლადების მიხედვით დაკავშირებულ წყვილებს.

გ. პასუხები შესაძლოა სხვადასხვა იყოს.

3. ა. დანაშაულის ჩადენის სქესის შესახებ მონაცემების იგნორირებით დავუშვებთ 1,918 შეცდომას, 6,866-4,948

ბ. ცვლად „დანაშაულის ჩამდენი რესპონდენტის სქესის" მონაცემთა გამოყენებით, დანაშაულის მსხვერპლის სქესის პროგნოზირების მიზნით ქალი რესპონდენტების შემთხვევაში დავუშვებთ 1,918 შეცდომას. ხოლო მამაკაცი დანაშაულის ჩამდენი რესპონდენტების შემთხვევაში დავუშვებთ 1,747 შეცდომას, იგივე მაჩვენებელი ქალი რესპონდენტების შემთხვევაში 171 იქნება.

გ. ლამბდა = *(1,918 - 1,918) / 1,918 = 0.* რესპოდენტის სქესის შესახებ მონაცემების გამოყენებით ვერ ვიღებთ შეცდომის ზომის შემცირებას.

დ. მე-2 დავალებაში მიღებული ლამბდას მნიშვნელობა უფრო მეტი სიძლიერის კავშირს ასახავს, ვიდრე ეს მე-3 დავალებისას მიღებული ლამბდას მნიშვნელობის მიმართაა. დანაშაულის ჩამდენის რასა უკავშირდება დანაშაულის მსხვერპლის რასას.

5. ა. *Ns = 113,142; Nd = 62,980; Ty = 273,142; Tx = 172,006.*

ბ. კენდალის ტაუ-*b* = *0.13*

$$\frac{113{,}142 - 62{,}980}{\sqrt{(113{,}142 + 62{,}980 + 273{,}142)(113{,}142 + 62{,}980 + 172{,}006)}} = 0.13$$

კენდალის ტაუ-*b* ასახავს სუსტი სიმჭიდროვის დადებით კავშირს სოციალურ კლასსა და ჯანმრთელობის მდგომარეობას შორის.

7. რადგან FEAR და SEX ნომინალური ცვლადებია, ამ შემთხვევაში ლამბდა შეცდომის პროპორციული შემცირების შესაბამისი საზომია. პირველი ცხრილისათვის ლამბდას მნიშვნელობა 0-ის ტოლია, რაც ნიშნავს რომ ცვლადებს შორის კავშირი არ არსებობს. შავკანიანი რესპონდენტებისათვის ლამბდას მნიშვნელობა 0.1833-ის ტოლია. თუ თეთრკანიანი რესპონდენტების შემთხვევაში ცვლადებს შორის კავშირი არ ფიქსირდება, ხოლო შავკანიანი რესპონდენტებისათვის ცვლადებს შორის არსებობს სუსტი კავშირი.

9. გამა ამ ცხრილისათვის 0.13-ს უდრის.

$$Ns = 723\ (213) = 153{,}999$$
$$Nd = 260\ (453) = 117{,}780$$

განათლების დონის შესახებ მონაცემების გამოყენებით პოლიტიკური პარტიის მხარდაჭერის პროგნოზირებისათვის პროგნოზირების შეცდომას 13%-ით შევამცირებთ, რაც ასახავს ცვლადებს შორის სუსტ კავშირს.

11. ა. ამ ცხრილში მოცემული ცვლადებისათვის კენდალის ტაუ-*b* ან გამა არის კავშირის შესაბამისი საზომი. შესაძლებელია გამოვთვალოთ ლამბდას მნიშვნელობა, მაგრამ ის შესაძლოა რეალურად არ იყოს შესაბამისი, რადგან თითოეული კატეგორიის დამოკიდებული და დამოუკიდებელი ცვლადის მიხედვით მოდა ხვდება ერთნაირ კატეგორიაში.

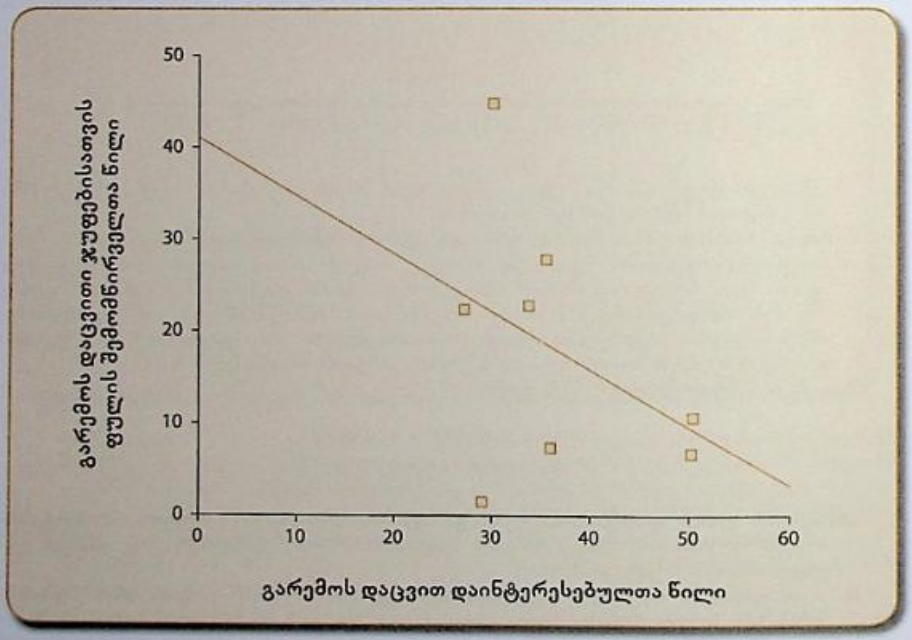
ბ. მიუხედავად იმის, რომ აქ მხოლოდ გამას გამოთვლების შედეგებია წარმოდგენილი, ასევე მისაღებია კენდალის ტაუ-*b*-ს გამოთვლა. ამ შემთხვევაში გამას მნიშვნელობა სქესსა და მარიჰუანას მოხმარებას შორის 0.13-ია, რაც ასახავს სუსტ დადებით კავშირს და 13%-ის ტოლ შეცდომის პროპორციულ შემცირებას. გამას მნიშვნელობა სქესსა და კოკაინის მოხმარებას შორის უდრის 03.0-ს, რაც ასახავს საშუალო სიძლიერის კავშირს და 30%-ის

ტოლ შეცდომის შემცირებას. შეგვიძლია დავასკვნათ, რომ არსებობს შედარებით ძლიერი სიმჭიდროვის კავშირი რესპონდენტის სქესსა და კოკაინის მოხმარების შესახებ მონაცემებს შორის, ვიდრე ეს სქესსა და მარიჰუანას მოხმარების შესახებ მონაცემებს შორისაა.

13. გამას მნიშვნელობა მამაკაცი რესპონდენტების შემთხვევაში −0.37-ის ტოლია, ხოლო იგივე მაჩვენებელი ქალი რესპონდენტების შემთხვევაში −0.54-ის ტოლია. შეცდომის პროპორციული შემცირების ზომა უფრო მაღალი ქალი რესპონდენტების შემთხვევაშია(54%), ვიდრე ეს მამაკაცი რესპონდენტების შემთხვევაშია(37%). ორივე ცხრილის მონაცემების მიხედვით არსებობს საშუალო სიმჭიდროვის უარყოფითი კავშირი ორ ცვლადს შორის- 12-დან 15 წლამდე ასაკის რესპონდენტები უფრო მეტი ალბათობით არ გამოიყენებენ დამცავ საშუალებებს, ხოლო 16-17 წლის ასაკის რესპონდენტები უფრო მეტი ალბათობით ყოველთვის გამოიყენებენ დაცვის საშუალებებს.

თავი 13. რეგრესია და კორელაცია

1. ა. ქვემოთ მოყვანილ წერტილთა გაბნევის დიაგრამაზე, დახაზულია რეგრესის წრფე რაც უფრო ცხადს ხდის ცვლადებს შორის კავშირის მიმართულებას.



ბ. წერტილთა გაბნევის დიაგრამა აჩვენებს ცვლადთა შორის კავშირს ზოგადი სახით. ჩვენი მაგალითის შემთხვევაში წერტილთა გაბნევა დიდი არ არის, მათი განლაგების მიხედვით შეგვიძლია შევნიშნოთ, რომ გარემოს დაცვის საკითხებით დაინტერესების ზრდასთან ერთად იზრდება შემომწირველთა რიცხვიც.

გ. ორ ცვლადს შორის კავშირისათვის პირსონის კორელაციის კოეფიციენტი -0.40-ის ტოლია. ეს მნიშვნელობა ლოგიკურიცაა, უარყოფითი კავშირი ასევე ჩანს წერტილთა გაბნევის დიაგრამაზე.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	დაინტერესებულთა წილი	შეშფოთებულთა წილი					
ქვეყანა	X	Y	$(X-\bar{X})$	$(X-\bar{X})^2$	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
აშშ	33.8	22.8	−2.69	7.22	4.77	22.80	−12.83
ავსტრია	35.5	27.8	−0.99	0.98	9.77	95.55	−9.67
ნიდერლანდები	30.1	44.8	−6.39	40.80	26.77	716.90	−171.06
სლოვენია	50.3	10.7	13.81	190.79	−7.33	53.66	−101.23
რუსეთი	29.0	1.6	−7.49	56.06	−16.43	269.78	123.06
ფილიპინები	50.1	6.8	13.61	185.30	−11.23	126.00	−152.84
ესპანეთი	35.9	7.4	−0.59	0.35	−10.63	112.89	6.27
დანია	27.2	22.3	−9.29	86.26	4.27	18.28	−39.67
	$\sum X = 291.9$	$\sum Y = 144.2$	−0.02	567.76	0.04[a]	1,415.85	−357.97

$$\text{საშუალო } X = \bar{X} = \frac{\sum X}{N} = \frac{291.9}{8} = 36.49$$

$$\text{საშუალო } Y = \bar{Y} = \frac{\sum Y}{N} = \frac{144.2}{8} = 18.03$$

$$\text{ცვლადი } (Y) = S_Y^2 = \frac{\sum (Y-\bar{Y})^2}{N-1} = \frac{1,415.9}{7} = 202.3$$

$$\text{სტანდარტული გადახრა } (Y) = S_Y = \sqrt{202.3} = 14.22$$

$$\text{ცვლადი } (X) = S_X^2 = \frac{\sum (X-\bar{X})^2}{N-1} = \frac{567.8}{7} = 81.11$$

$$\text{სტანდარტული გადახრა } (X) = S_X = \sqrt{81.11} = 9.01$$

$$\text{კოვარიაცია } (X,Y) = S_{YX} = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{N-1} = \frac{-357.97}{7} = -51.14$$

$$r = \frac{S_{YX}}{S_X S_Y} = \frac{-51.14}{(9.01)(14.22)} = -40^a$$

3. ა. კორელაციის კოეფიციენტი უდრის -0.45-ს.

ერთ სულზე მთლიანი ეროვნული პროდუქტი	გადახდის მსურველთა რიცხვი	$(X-\bar{X})$	$(X-\bar{X})^2$	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
29.24	44.9	2.72	7.40	−1.64	2.69	−4.46
18.71	53.3	−7.81	61.00	6.76	45.70	−52.80

24.78	61.2	−1.74	3.03	14.66	214.92	−25.51
34.31	40.7	7.79	60.68	−5.84	34.11	−45.49
25.58	32.6	−0.94	0.88	−13.94	194.32	13.10
$\sum X = 132.62$	$\sum X = 232.7$	0.02[a]	132.99	0.0	491.74	−151.16

$$\text{საშუალო } X = \overline{X} = \frac{\sum X}{N} = \frac{132.62}{5} = 26.52$$

$$\text{საშუალო } Y = \overline{Y} = \frac{\sum Y}{N} = \frac{232.7}{5} = 46.54$$

$$\text{ცვლადი } (Y) = S_Y^2 = \frac{\sum (Y-\overline{Y})^2}{N-1} = \frac{491.74}{4} = 122.94$$

$$\text{სტანდარტული გადახრა } (Y) = S_Y = \sqrt{122.94} = 11.09$$

$$\text{ცვლადი } (X) = S_X^2 = \frac{\sum (X-\overline{X})^2}{N-1} = \frac{132.99}{4} = 33.25$$

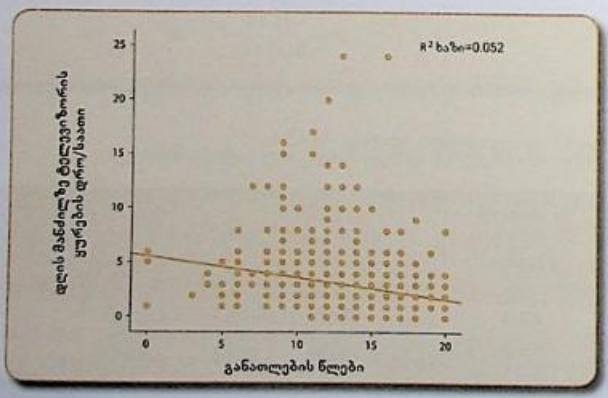
$$\text{სტანდარტული გადახრა } (X) = S_X = \sqrt{33.25} = 5.77$$

$$\text{კოვარიაცია } (X,Y) = S_{YX} = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{N-1} = \frac{-115.16}{4} = -28.79$$

$$r = \frac{S_{YX}}{S_X S_Y} = \frac{-28.79}{(11.09)(5.77)} = -0.45^{a}$$

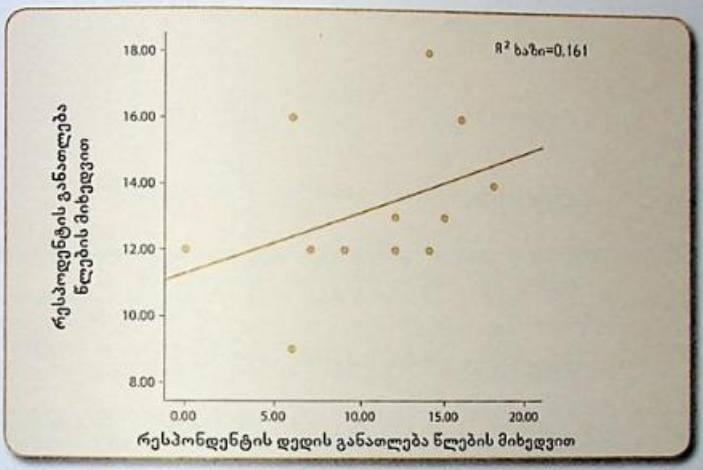
5. ა. დიახ, როგორც ამას b-ს უარყოფითი მნიშვნელობა ასახავს -0.196, ცვლადებს შორის კავშირი უარყოფით ხასიათს ატარებს.

ბ. საშედეგო კოეფიციენტის მიხედვით შეგვიძლია განვსაზღვროთ, რომ 16 წლის განათლების მქონე პირი უყურებს ტელევიზორს 2.439 საათს კვირაში. საპირისპიროდ, 12 წლის განათლების მქონე პირი სავარაუდოდ უყურებს ტელევიზორს 3.223 საათს კვირაში. ამ გამოთვლებისათვის გამოვიყენოთ შემდეგი რეგრესიის განტოლება $\hat{Y} = 5.575 + (-0.196)\,X$.



გ. წერტილთა გაბნევის დიაგრამაზე დაკვირვებით ჩანს, რომ რეგრესიის წრფე უახლოვდება წერტილთა გაბნევის კონცენტრირებულ არეებს, თუმცა წერტილთა საკმაო ოდენობა რჩება წრფის მიღმა.

7. ა. წერტილთა გაბნევის დიაგრამა ასახავს: სუსტიდან საშუალო სიძლიერემდე სიმჭიდროვის კავშირს ცვლადებს შორის, დადებით კავშირს რესპონდენტის დედის განათლების დონესა და რესპონდენტის განათლებას შორის.



ბ. დაკვირდით ზემოთ მოცემულ რეგრესიის წრფეს წერტილთა გაბნევის დიაგრამაზე. რეგრესიის წრფის დახრილობა 0.186-ის ტოლია, ხოლო თანაკვეთა 11.25-ია. წრფე მონაცემთა გაბნევას საკმაოდ კარგად ერგება. ზოგადად, რაც უფრო მაღალი რესპონდენტის დედის განათლება მით უფრო მაღალია რესპონდენტის განათლება.

MEADUC	EDUC	$(X-\bar{X})$	$(X-\bar{X})^2$	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
0	12	10.75	115.56	1.25	1.56	13.44
15	13	4.25	18.06	0.25	0.06	1.06
6	9	4.75	22.56	4.25	18.06	20.19
9	12	1.75	3.06	1.25	1.56	2.19
16	16	5.25	27.56	2.75	7.56	14.44
12	12	1.25	1.56	1.25	1.56	1.56
6	16	4.75	22.56	2.75	7.56	13.06
18	14	7.25	52.56	0.75	0.56	5.44
12	13	1.25	1.56	0.25	0.06	0.31
14	12	3.25	10.56	1.25	1.56	4.06
14	18	3.25	10.56	4.75	22.56	15.44
7	12	3.75	14.06	1.25	1.56	4.69

$\sum X = 129$	$\sum Y = 159$	0.0	300.22	0.0	64.22	55.78

$$\text{საშუალო } X = \overline{X} = \frac{\sum X}{N} = \frac{129}{12} = 10.75$$

$$\text{საშუალო } Y = \overline{Y} = \frac{\sum Y}{N} = \frac{159}{12} = 13.25$$

$$\text{ცვლადი } (Y) = S_Y^2 = \frac{\sum (Y-\overline{Y})^2}{N-1} = \frac{64.22}{11} = 5.84$$

$$\text{სტანდარტული გადახრა } (Y) = S_Y = \sqrt{5.84} = 2.42$$

$$\text{ცვლადი } (X) = S_X^2 = \frac{\sum (X-\overline{X})^2}{N-1} = \frac{300.22}{11} = 27.29$$

$$\text{სტანდარტული გადახრა } (X) = S_X = \sqrt{27.29} = 5.22$$

$$\text{კოვარიაცია } (X,Y) = S_{YX} = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{N-1} = \frac{55.78}{11} = 5.07$$

$$b = \frac{S_{YX}}{S_X^2} = \frac{5.07}{27.29} = 0.186^{\text{a}}$$

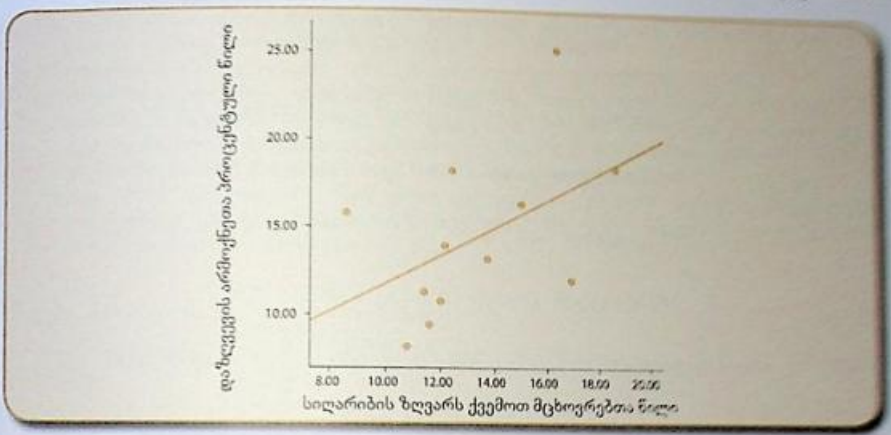
$$a = \overline{Y} - b\overline{X} = 13.25 - (0.186)(10.75) = 11.25^{\text{a}}$$

ბ. მეორე შემთხვევისათვის პროგნოზირების შეცდომა 1.04-ია. ე.ი. ჩვენი გამოთვლების მიხედვით პირს ექნებოდა 14.04 წლის განათლება $\hat{Y} = 11.25 + 0.186(15)$, მაგრამ ამ პირს რეალურად აქვს 13 წლის განათლება. შეცდომის ზომა იმ რესპონდენტისათვის, რომლის დედასაც აქვს 18 წლის განათლება, ხოლო თავად კი 14 წლის განათლება, იქნება −.0598. $\hat{Y} = 11.25 + 0.186(12)$, ე.ი ჩვენი პროგნოზით პირს უნდა ჰქონოდა 14.598 წლის განათლება, მას კი რეალურად უფრო ნაკლები წლების განათლება ჰქონდა.

დ. იმ პირისათვის, რომლის დედასაც აქვს 4 წლის განათლება, ჩვენი რეგრესიის განტოლების მიხედვით ამ პირს ექნება 11.99 წლის განათლება $\hat{Y} = 11.25 + 0.186(4)$. იმ პირისათვის, რომლის დედასაც აქვს 12 წლის განათლება რეგრესიის განტოლების მიხედვით ექნება 13.48 წლის განათლება $\hat{Y} = 11.25 + 0.186(12)$.

ე. უმცირეს კვადრატთა წრფეზე. რადგან უმცირეს კვადრატთა წრფის მიღებისათვის გამოიყენება ორივე საშუალო მნიშვნელობა, ეს ნიშნავს, რომ დაკვირვების ეს წერტილი უნდა იყოს თავად წრფეზე.

9. ა. წრფე არ ეთანადება დაკვირვების წერტილთა გაბნევას წერტილთა გაბნევის დიაგრამაზე.



ბ. განტოლების მიხედვით, $\hat{Y} = 3.76 + 0.8X$ რეგრესიის წრფე საუკეთესოდ უთანხმებს ზონა ცემთა გაფანტვას. b კოეფიციენტი უდრის 0.8-ს, რაც ნიშნავს რომ 1%-ით ზრდა ამ მოსახლეობის რაოდენობის, რომლებიც სიღარიბის ზღვას ქვემოთ ცხოვრობენ პროცენტს 0.8%-იანი ზრდა ჯანმრთელობის დაზღვევის არმქონეთა რიცხვში.

სიღარიბის ზღვარს ქვემოთ მცხოვრებთა წილი	საშედიცინო დაზღვევის არმქონეთა წილი	$(X-\bar{X})$	$(X-\bar{X})^2$	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
16.9	12	3.62	13.10	−2.42	5.86	−8.76
12.4	18.2	−0.88	0.77	3.78	14.29	−3.33
12.1	13.9	−1.18	1.39	−0.52	0.27	0.62
18.6	18.5	5.32	28.30	4.08	16.65	21.71
8.6	15.8	−4.68	21.90	1.38	1.90	−6.46
13.7	13.2	0.42	0.18	−1.22	1.49	−0.51
11.6	9.5	−1.68	2.82	−4.92	24.21	8.27
12	10.8	−1.28	1.64	−3.62	13.10	4.64
15	16.4	1.72	2.96	1.98	3.92	3.41
16.3	25.2	3.02	9.12	10.78	116.21	32.56
11.4	11.3	−1.88	3.53	−3.12	9.73	5.87
10.8	8.2	−2.48	6.15	−6.22	38.69	15.43
$\sum X = 159.4$	$\sum Y = 173$	0.0	91.86	0.0	246.32	73.45

საშუალო $X = \bar{X} = \frac{\sum X}{N} = \frac{159.4}{12} = 13.28$

$$\text{საშუალო } Y = \bar{Y} = \frac{\sum Y}{N} = \frac{173}{12} = 14.42$$

$$\text{ცვლადი } (Y) = S_Y^2 = \frac{\sum (Y-\bar{Y})^2}{N-1} = \frac{246.32}{11} = 22.39$$

$$\text{სტანდარტული გადახრა } (Y) = S_Y = \sqrt{22.39} = 4.73$$

$$\text{ცვლადი } (X) = S_X^2 = \frac{\sum (X-\bar{X})^2}{N-1} = \frac{91.86}{11} = 8.35$$

$$\text{სტანდარტული გადახრა } (X) = S_X = \sqrt{8.35} = 2.89$$

$$\text{კოვარიაცია } (X,Y) = S_{YX} = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{N-1} = \frac{73.45}{11} = 6.68$$

$$b = \frac{S_{YX}}{S_X^2} = \frac{6.68}{8.35} = 0.8$$

$$a = \bar{Y} - b\bar{X} = 14.42 - 0.8(13.28) = 3.796$$

$$\hat{Y} = 3.796 + 0.8X$$

გ. 5=2.796+0.8X
X=1.51%. ასე, რომ იმისათვის, რომ მოსახლეობაში იყოს მხოლოდ 5% იმ ადამიანების, რომელთაც არ აქვთ ჯანმრთელობის დაზღვევა სიღარიბის ზღვარს ქვემოთ მცხოვრებთა წილი უნდა იყოს 1.51%.

დ. ამ შემთხვევაში დაგჭირდებათ მოტანილი მონაცემების გარეთ მსჯელობა, რათა დამატებით შეაფასოთ კავშირი სიღარიბის ზღვარს ქვემოთ მცხოვრებთა რაოდენობასა და ჯანმრთელობის დაზღვევის არმქონეთა შორის. მიუხედავად იმისა, რომ ჩვენ არ შეგვიძლია სტატისტიკური განზოგადება, ეს მაგალითი გვეხმარება იმის შეფასებაში თუ როგორაა ეს ორი ცვლადი ერთმანეთთან დაკავშირებული.

11. ა. განათლებისათვის b კოეფიციენტის მნიშვნელობა −0.185-ის ტოლია. განათლების წლების 1 წლით ზრდა ამცირებს ტელევიზორის ყურების დროს 0.185 საათით. ბავშვების რაოდენობისათვის b კოეფიციენტის მნიშვნელობა 0.117-ია. თუ ოჯახში ბავშვების რაოდენობა ერთით იზრდება, პირის მიერ ტელევიზორის ყურების დრო 0.117-ით იზრდება. ამ ცვლადებს შორის კავშირის არსებობის შესახებ დაშვება ჰიპოთეზის სახითაა ჩამოყალიბებული.

ბ. იმ პირისათვის, რომელსაც აქვს 16 წლის განათლება და ყავს 2 ბავშვი, ტელევიზორის ყურების პროგნოზირებული დრო კვირის მანძილზე 2.483-ია. მე-5 დავალების მიხედვით 16 წლის განათლების მქონე პირისათვის მოსალოდნელი ტელევიზორის ყურების დრო კვირის მანძილზე დაახლოებით 2.439-ია. ბავშვების რაოდენობის დამატებით რეგრესიის განტოლებაში, ტელევიზორის ყურების დროს ზრდის 0.044 საათით:

$$\hat{Y} = 5.029 - 0.185(16) + 0.117(2) = 2.483$$

გ. მე-5 სავარჯიშოში r^2-ის მნიშვნელობა მხოლოდ განათლების ფაქტორის გამოყენების შემთხვევაში არის 0.052. შეცდომა მცირდება 5.2%-ით, როდესაც განათლების მონაცემების გამოყენება ხდება ტელევიზორის ყურების ხანგრძლივობის გამოსათვლელად. როდესაც ვამატებთ მეორე ფაქტორს-ოჯახში ბავშვების რაოდენობას R^2 უტოლდება 0.058-ს. შეცდომა მცირდება 5.8%-ით როცა ხდება ორივე ფაქტორის გამოყენება.

13. ა. მცდარია, როგორც b-ს ასევე r-ს ყოველთვის ეწერება მსგავსი ნიშანი, რაც ცვლადებს შორის კავშირის მიმართულებას ასახავს.
ბ. ორივე a და b კოეფიციენტი ასახავს დამოკიდებულ ცვლადში ცვლილებას.
გ. დეტერმინაციის კოეფიციენტი r^2 არის შეცდომის პროპორციული შემცირების საზომი. PRE აღნიშნავს შეცდომის პროპორციულ შემცირებას ასახავს თუ არა რეგრესიის განტოლება ცვლადებს შორის მიზეზობრივ კავშირს უნდა შეფასდეს მიზეზობრივი კავშირების შესახებ იმ მოთხოვნების გამოყენებით, რაც 1 თავში იყო განხილული.
რეგრესიის განტოლებები ჩვეულებრივ გამოიყენება ცვლადებს შორის კავშირების რაოდენობრივი გამოსახვისათვის, რომლის დროსაც მკვლევარი სწავლობს ერთი ცვლადის მიხედვით მეორის ცვლილებას.

თავი 14. დისპერსიული ანალიზი

1. ჩვენ SST, SSE-სა და SSR-ის გამოთვლას დავიწყებთ ქვემოთ მოცემული ცხრილის გამოყენებით:

$$SST = \sum(Y-)2 = 1,415.858$$

$$SSE = \sum(Y - \hat{Y})^2 = 1,190.158$$

$$SSR = SST - SSE = 1,415.858 - 1,190.158 = 115.700$$

Y	$(Y - \bar{Y})$	$(Y - \bar{Y})^2$	$\hat{Y}$	$(Y - \hat{Y})$	$(Y - \hat{Y})^2$
22.8	4.775	22.801	19.70	3.098	9.596
27.8	9.775	95.551	18.63	9.171	84.098
44.8	26.775	716.901	22.04	22.763	518.159
10.7	−7.325	53.656	9.29	1.409	1.986
1.6	−16.425	269.781	22.73	−21.131	446.519
6.8	−11.225	126.001	9.42	−2.617	6.848
7.4	−10.625	112.891	18.38	−10.977	120.497
22.3	4.275	18.276	23.87	−1.567	2.455
$\bar{Y} = 18.025$		$\sum(Y - \bar{Y})^2 = 1,415.858$		$\sum(Y - \hat{Y})^2 = 1,190.158$	

ჩავსვათ მოცემული მნიშვნელობები შემდეგ ფორმულებში:

$$MSR = \frac{SSR}{df_r} = \frac{SSR}{K} = \frac{225.700}{1} = 225.700$$

$$MSR = \frac{SSR}{df_e} = \frac{SSE}{[N-(K+1)]} = \frac{1,190.158}{[8-(1+1)]} = 198.359$$

საბოლოოდ, გამოვთვალოთ F ფარდობა:

$$F = \frac{MSR}{MSE} = \frac{225.700}{198.359} = 1.138$$

განაწილების მიხედვით $\alpha = 0.05$, $df_1 = 1$ და $df_1 = 6$, F ფარდობა 1.138 ნაკლებია ვიდრე F-ის კრიტიკული მნიშვნელობა 5.99. F-ის კრიტიკული მნიშვნელობა გამოიყენება კავშირის მნიშვნელოვნების შეფასებისათვის. ე. ი. P-ს მნიშვნელობა აღემატება 0.05-ს, რაც ნიშნავს რომ კავშირი სტატისტიკურად მნიშვნელოვანი არაა.

3. ა. მამაკაცები=25.59, ქალები= 22.57. საშუალოდ მამაკაცების ასაკი(3.02-ით) მცირედით აღემატება ქალების ასაკს პირველი ბავშვის დაბადებისას.

ბ. F-ის გამოთვლილი მნიშვნელობის მიხედვით 55.933, ვახდენთ ნულოვანი ჰიპოთეზის უარყოფას. SPSS-ის გამოთვლის მიხედვით F-ის გამოთვლილი მნიშვნელობის ალბათობა 0.000, რაც ნაკლებია ალფას 0.05 მნიშვნელობასთან შედარებით. ეს შედეგი ასახავს, რომ მამაკაცები საშუალოდ უფრო მეტი ასაკის არიან პირველი ბავშვის დაბადებისას ვიდრე ეს ქალების შემთხვევაშია.

5. ა. ჯგუფი ყველაზე მაღალი ბავშვების რაოდენობით არის ის ჯგუფი, რომელმაც განაცხადა „არა“ აბორტის შესაძლებლობის შესახებ, 2.17 ბავშვი.

ბ. SPSS-ის გამოთვლის მიხედვით არსებობს კავშირი ორ ჯგუფს შორის. F-ის გამოთვლილი მნიშვნელობა არის 34.509, რომელიც 0.000 დონეზეა მნიშვნელოვანი (რაც ალფას 0.05-ზე ნაკლებია).

გ. $101.438 / 2946.849 = 0.034 = 0.03$. ვარიაციის მხოლოდ 3% არის შესაძლებელი იხსნას ცვლადით „აბორტის შესაძლებლობის მხარდაჭერა“.

7. ა. ყველაზე მაღალი მაჩვენებლის მქონე ჯგუფი არის იმ რესპონდენტთა ჯგუფი, რომელთაც უპასუხეს „არა“, 3.11 ბავშვით.

ბ. ჯგუფებს შორის იდეალური ბავშვების რაოდენობის მიხედვით განსხვავება მნიშვნელოვანი არ არის. F-ის მიღებული მნიშვნელობა 2.427; 0.120 მნიშვნელოვნების დონით (რაც ალფას 0.05 მნიშვნელობას აღემატება). შესაბამისად, უარვყოფთ ნულოვან ჰიპოთეზას.

გ. თუ ალფას მნიშვნელობას 0.01-ის ტოლად ავირჩევდით, მაინც მოხდებოდა ნულოვანი ჰიპოთეზის უარყოფა. მიღებული F-ის მნიშვნელობის მიღების ალბათობა 0.120-ია, რაც აღემატება 0.01-ს.

9. ა.

$\bar{Y}_1 = 4.29$	$\bar{Y}_1 = 2.29$	$\bar{Y}_1 = 3.14$
$\sum Y_1 = 30$	$\sum Y_2 = 16$	$\sum Y_3 = 22$
$\sum Y_1^2 = 134$	$\sum Y_2^2 = 44$	$\sum Y_3^2 = 84$
$n_1 = 7$	$n_2 = 7$	$n_3 = 7$
	$\bar{Y} = 3.24$	
	$N = 21$	

$$SSB = 7(4.29 - 3.24)^2 + 7(2.29-3.24)^2 + 7(3.14-3.24)^2$$
$$= 7(1.10) + 7(0.01)$$
$$= 7.70 + 6.30 + 0.07$$
$$SSB = 14.07$$
$$df_b = 3-1$$
$$df_b = 2$$

მონაცემთა შორის კვადრატების საშუალო $= 14.07/2 = 7.035$

$$SSW = (134 + 44 + 84) - [(302/7) + (162)/7 + (222/7)]$$
$$= 262 - (128.57 + 36.57 + 69.14)$$

$$= 262 - 234.28$$
$$SSW = 27.72$$
$$df_w = 21-3$$
$$df_w = 18$$

მონაცემთა შიგა კვადრატების ჯამი $= 17.72/18 = 1.54$

$$F = 7.035/1.54$$
$$F = 4.57$$

გადაწყვეტილება: თუ ალფას 0.05-ის ტოლად ავირჩევთ, F-ის კრიტიკული მნიშვნელობა იქნება 3.55 ($df_1 = 2$ და $df_2 = 18$). F-ის გამოთვლილი მნიშვნელობის მიხედვით 4.57 უარვყოფთ ნულოვან ჰიპოთეზას და ვასკვნით, რომ სულ მცირე ერთი საშუალო მაჩვენებელი განსხვავდება სხვებისაგან. საშუალოდ, თეორკანიან რესპონდენტებს აქვთ სკოლის დღეების გაცდენის ყველაზე მაღალი მაჩვენებელი ბოლო 4 კვირის მანძილზე (4.29), რასაც მოსდევს ესპანური წარმოშობის რესპონდენტები (3.14), ხოლო შემდეგ შავკანიანი რესპონდენტები მოდიან (2.29).

ბ. თუ ალფას მნიშვნელობას 0.01-ით შევცვლით, F კრიტიკულის მნიშვნელობა იქნება 6.01, ამ შემთხვევაში ვერ მოხდება ნულოვანი ჰიპოთეზის უარყოფა.

11. *SST*, *SSE* და *SSR* მაჩვენებლების გამოთვლა დავიწყოთ ქვემოთ მოყვანილი ცხრილის გამოყენებით:

$$SST=\sum(Y-)2=103.6$$

$$SSE=\sum(Y-\hat{Y})2=37.38$$

$$SSR=SST-SSE=103.6-37.28=66.22$$

Y	$(Y - \bar{Y})$	$(Y - \bar{Y})^2$	$\hat{Y}$	$(Y - \hat{Y})$	$(Y - \hat{Y})^2$
18	0.2	0.04	15.25	2.75	7.56
25	7.2	51.84	21.77	3.23	10.43
18	0.2	0.04	15.29	2.71	7.34
15	−2.8	7.84	13.41	1.59	2.53
18	0.2	0.04	18.07	−0.07	0.01
20	2.2	4.84	19.43	0.57	0.32
18	0.2	0.04	20.93	−2.93	8.58
19	1.2	1.44	19.01	−0.01	0.00
14	−3.8	14.44	13.84	0.16	0.03
13	−4.8	23.04	12.24	0.76	0.58
$\bar{Y} = 17.8$		$\sum(Y - \bar{Y})^2 = 103.6$		$\sum(Y - \hat{Y})^2 = 37.38$	

ცხრილში მოყვანილი მონაცემები ჩავსვათ შემდეგ ფორმულებში:

$$MSR = \frac{SSR}{df_r} = \frac{SSR}{K} = \frac{66.22}{1} = 66.22$$

$$MSR = \frac{SSR}{df_e} = \frac{SSE}{[N-(K+1)]} = \frac{37.38}{[10-(1+1)]} = 4.67$$

საბოლოოდ, გამოვთვალოთ F ფარდობა:

$$F = \frac{MSR}{MSE} = \frac{66.22}{4.67} = 14.18$$

F განაწილების მიხედვით $\alpha = 0.05$, თავისუფლების ერთი და 8 ხარისხებით, ცხადია რომ F ფარდობის მნიშვნელობა აღემატება F-ის კრიტიკულ მნიშვნელობას 5.32-ს, შესაბამისად ვასკვნით რომ GDP და შობადობის დონე სტატისტიკურად დაკავშირებულია.

13.

$\overline{Y}_1 = 0.28$	$\overline{Y}_1 = 1.75$	$\overline{Y}_1 = 3.20$
$\sum Y_1 = 4$	$\sum Y_2 = 7$	$\sum Y_3 = 16$
$\sum Y_i^2 = 6$	$\sum Y_3^2 = 15$	$\sum Y_3^2 = 54$
$n_1 = 5$	$n_2 = 4$	$n_3 = 5$
	$\overline{Y} = 1.93$	
	$N = 14$	

$$SSB = 5(0.8 - 1.93)^2 + 4(1.75 - 1.93)^2 + 5(3.20 - 1.93)^2$$
$$= 5(1.2769) + 4(0.0324) + 5(1.6129)$$
$$= 6.3845 + 0.1296 + 8.0645$$
$$SSB = 14.58$$
$$df_b = 3\text{-}1$$
$$df_b = 2$$

მონაცემთა შორის კვადრატების საშუალო $= 14.58/2 = 7.29$

$$SSW = (6 + 15 + 54) - [(42/5) + (72)/4 + (162/5)]$$
$$= 75 - (3.2 + 12.25 + 51.2)$$
$$= 75 - 66.65$$
$$SSW = 8.35$$
$$df_w = 14 - 3$$
$$df_w = 11$$

მონაცემთა შიგა კვადრატების ჯამი=8.35/11=0.76

$$F = 7.29/0.76$$
$$F = 9.59$$

გადაწყვეტილების მიღება: თუ ალფას 0.05-ის ტოლად ავირჩევთ, F კრიტიკული იქნება 3.98 ($df_1 = 2$ და $df_2 = 11$). F-ის გამოთვლილი მნიშვნელობის მიხედვით 9.59 უარვყოფთ ნულოვან ჰიპოთეზას, სულ მცირე ერთი საშუალო მაინცაა სხვებისაგან განსხვავებული. საგზაო მოძრაობის დარღვევათა საშუალო რიცხვი ყველაზე მაღალი დიდ ქალაქებშია(3.2), შემდეგ ადგილზეა საშუალო ზომის ქალაქები (1.75), და ბოლოს, პატარა ზომის ქალაქები (0.8).

გლოსარიუმი

ალბათური შერჩევა მეთოდი, რომელიც საშუალებას გვაძლევს განვსაზღვროთ შერჩევაში მოხვედრის ალბათობა თითოეული დაკვირვებისთვის.

ალფა (α) ალბათობის ის დონე რომლის შემთხვევაში ხდება ნულოვანი ჰიპოთეზის უარყოფა. ალფას მნიშვნელობა აირჩევა: 0.05, 0.01-ის ან 0.001-ის ტოლად.

არაპროპორციული სტრატიფიცირებული შემთხვევითი შერჩევა შერჩევა, რომლის დროსაც თითოეული ქვეჯგუფიდან არჩეული შერჩევის ზომა არაპროპორციულია თავად ამ ქვეჯგუფის ზომასთან მიმართებაში, მთლიან გენერალურ ერთობლიობაში.

არაპროპორციული სტრატიფიცირებული შერჩევა შერჩევის ზომა თითოეული ქვეჯგუფიდან სპეციალურად არ არის პროპორციული ქვეჯგუფების ზომისა მთლიან ერთობლიობაში.

აღწერილობითი სტატისტიკა მოიცავს მეთოდებს, რომლებიც გვეხმარება მონაცემთა თავმოყრასა და აღწერაში შერჩევითი ან გენერალური ერთობლიობიდან.

ბივარიანტული ანალიზი სტატისტიკის მეთოდი, რომელიც გამოიყენება ორ ცვლადს შორის კავშირის აღწერისა და შესწავლისათვის.

ბივარიანტული ცხრილი ცხრილი, რომელიც შეიცავს ერთი ცვლადის განაწილებას მეორე ცვლადის კატეგორიების მიხედვით.

ცალმხრივი დისპერსიული ანალიზი დისპერსიული ანალიზის გამოყენება ერთი დამოკიდებული და ერთი დამოუკიდებელი ცვლადის მიმართ.

ცალმხრივი დისპერსიული ანალიზი ვარიაციის ანალიზი ერთი დამოკიდებული და ერთი დამოუკიდებელი ცვლადის მიმართ.

ცალმხრივი კრიტერიუმი ჰიპოთეზის ტესტირების სახე, რომელიც მოიცავს კვლევითი ჰიპოთეზის მიმართულებას. ის აკონკრეტებს, მთლიანი ერთობლიობის საშუალო ნაკლებია თუ მეტია რაიმე კონკრეტულ სიდიდეზე.

ცალმხრივი ტესტი ჰიპოთეზის კრიტერიუმის სახე, რომელიც მოიცავს ცვლადის მნიშვნელობების გარკვეული მიმართულების შესახებ დაშვებას. ამ სახის ჰიპოთეზის მიხედვით ერთი ჯგუფის მნიშვნელობები არის მეტი ან ნაკლები პოპულაციის ცალკეულ მნიშვნელობებთან შედარებით.

ცენტრალური ტენდენციის საზომები რიცხვები, რომლებიც აღწერენ განაწილების საშუალო ან ტიპიურ მნიშვნელობებს

ცენტრალური ზღვარითი თეორემა თუ პოპულაციიდან მიღებულია ყველა შესაძლო შერჩევითი ერთობლიობა N ზომით, საშუალო $\mu_{\bar{x}}$ და სტანდარტული გადახრის $\sigma_{\bar{x}}$ მნიშვნელობით, N-ის ზომის ზრდასთან ერთად შერჩევის მნიშვნელობების განაწილება უახლოვდება ნორმალურ განაწილებას, $\mu_{\bar{x}}$ საშუალოთი და $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ სტანდარტული გადახრით.

ცვლადი წარმოადგენს ადამიანების ან საგნების მახასიათებელს, რომელმაც შეიძლება ორი ან მეტი მნიშვნელობა მიიღოს.

დაკავშირებული წყვილები დამოკიდებული ცვ-

ლადისათვის (T_y) დაკვირვების დაკავშირებული წყვილები დამოკიდებული ცვლადის მიხედვით.

დაკავშირებული წყვილები დამოუკიდებელი ცვლადისათვის (T_x) დაკვირვების დაკავშირებული წყვილები დამოუკიდებული ცვლადის მიხედვით. მიუხედავად იმისა, რომ გამას გამოთვლისათვის არ არის საჭირო წყვილთა რაოდენობის ცოდნა, იგი საჭიროა კენდალის ტაუ-b-ს გამოთვლისათვის.

დაკვირვების სიხშირეები (f_o) ორგანზომილებიანი ცხრილის უჯრების ფაქტობრივი დაკვირვების მნიშვნელობები.

დაკვირვების წყვილობრივი მონაცემები დაკვირვების მონაცემები, რომლებიც მოცემულია მათი შესაბამისი პოზიციით დამოუკიდებელ და დამოკიდებულ ცვლადებს შორის.

დამოკიდებული ცვლადი ცვლადი, რომლის მნიშვნელობათა ცვლილება სხვა ცვლადის მეშვეობით აიხსნება.

დამოუკიდებელი ცვლადი ცვლადი, რომელიც განაპირობებს დამოკიდებული ცვლადის მნიშვნელობათა ცვლილებას.

დასკვნითი სტატისტიკა ლოგიკური მსჯელობა და მეთოდები, რომლებიც მოიცავს გენერალური ერთობლიობის შესახებ პროგნოზებსა და დასკვნებს შერჩევით ერთობლიობაზე დაკვირვებისა და მისი ანალიზის საფუძველზე.

დახრილობა (b) დამოკიდებული ცვლადის მნიშვნელობის ცვლილება დამოუკიდებელი ცვლადის ერთი ერთეულის ცვლილების შედეგად.

დეტერმინაციის კოეფიციენტი r^2 შეცდომის პროპორციული შემცირების საზომია, რომელიც ასახავს შეცდომის პროპორციული შემცირებას წრფივი რეგრესიის მოდელის გამოყენების შედეგად. ის ასახავს მთლიანი ვარიაციის წილს დამოკიდებულ ცვლადში, რომელიც დამოუკიდებელი ცვლადით აიხსნება.

დეტერმინისტული (სრულყოფილი) წრფივი კავშირი კავშირის სახე ორ იტერვალურ-ფარდობით ცვლადს შორის, რომლის დროსაც დაკვირვების ყველა მონაცემი ემთხვევა კავშირის გამომსახველ წრფეს. წრფე იძლევა საპროგნოზო ცვლადის (Y) ყველა მნიშვნელობას დამოკიდებული ცვლადის (X) ნებისმიერი მნიშვნელობისათვის.

დიაპაზონი ინტერვალურ-შეფარდებითი ცვლადების ვარიაციის საზომი. ეს არის სხვაობა განაწილების უდიდეს (მაქსიმალურ) და უმცირეს (მინიმალურ) მნიშვნელობებს შორის.

დიქოტომიური ცვლადი ცვლადი, რომელსაც მხოლოდ ორი მნიშვნელობა აქვს.

დისპერსიული ანალიზი (ANOVA) დასკვნითი სტატისტიკის მეთოდი, რომელიც გამოიყენება ორ ცვლადს, ცვლადთა ჯგუფსა ან ორ შერჩევით მონაცემებს შორის კავშირის მნიშვნელოვნების შეფასებისათვის.

დროითი მწკრივების დიაგრამა ასახავს ცვლადის განსხვავებებს დროის სხვადასხვა მომენტებში. ის გვიჩვენებს დროს (იზომება დროის ერთეულებში როგორიცაა წელი, თვე და ა.შ.) აბცისათა ღერძზე და სიხშირეს (ხვედრით წილებს, დონეებს) ორდინატთა ღერძზე.

დროითი მწკრივების დიაგრამა დიაგრამა, რომელიც ასახავს ცვლადის განსხვავებებს დროის სხვადასხვა მონაკვეთებში. ის გვიჩვენებს დროს (იზომება დროის ერთეულებში როგორიცაა წელი, თვე და ა.შ.) აბსცისათა ღერძზე და სიხშირეს (ხვედრით წილებს, დონეებს) ორდინატთა ღერძზე.

ემპირიული კვლევა მტკიცებულებებზე დაყრდნობით კვლევა, რომელიც შესაძლებელია გადამოწმდეს ჩვენი პირდაპირი გამოცდილების მიხედვით.

ერთობლიობის ერთეული სოციოლოგების ინტერესის სფეროში მოქცეული სოციალური ცხოვრების დონეები. ერთობლიობის ერთეულის განსხვავებული დონეებია ცალკეული ინდივიდები და ჯგუფები.

F ფარდობა ან F სტატისტიკა სტატისტიკური კრიტერიუმი დისპერსიული ანალიზისათვის, რომელიც მონაცემთა შიგა საშუალო კვადრატებისა და მონაცემთა შორის საშუალო კვადრატების შეფარდებით გამოითვლება.

F კრიტიკული სტატისტიკური კრიტერიუმი, რომელიც შეესაბამება ალფას მნიშვნელობის დონეს, df_w, და df_b.

F-ის მიღებული მნიშვნელობა გამოთვლების შედეგად მიღებული F სტატისტიკური კრიტერიუმის მნიშვნელობა.

გადახრილი განაწილება განაწილება ერთ მხარეზე წარმოდგენილი ექსტრემალური მნიშვნელობებით.

გამა ცვლადებს შორის კავშირის სიმეტრიული საზომი ორდინალური ან დიქოტომიური ნომინალური ცვლადებისათვის. მნიშვნელობა ცვალებადობს 0.0დან +/-1.0 მდე და ასახავს ცვლადებს შორის კავშირის სიძლიერეს და მიმართულებას.

გამოკვლევის პროცესი საქმიანობათა ერთობლიობაა, რომელსაც ახორციელებენ სოციოლოგები იმისათვის, რომ გასცენ პასუხები კითხვებს, შეამოწმონ მოსაზრებები და ჰიპოთეზები.

გაზომვის ნომინალურ დონეზე ციფრები და სიმბოლოები გამოიყენება სახელდების, აღნიშვნის ან დაკვირვებების დაჯგუფების მიზნით. გაზომვის ორდინალურ დონეზე კატეგორიებს ვანიჭებთ, რანგებს თანმიმდევრობით დაბალიდან მაღლისკენ (ან პირიქით). ინტერვალურ-ფარდობითი გაზომვის დროს ყველა დაკვირვება ერთიდაიმავე სიდიდეებით გამოისახება.

გენერალური ერთობლიობა მკვლევარისათვის საინტერესო საკვლევი ინდივიდების, საგნების, ჯგუფების ან მოვლენების სრული ნაკრებია. შერჩევითი ერთობლიობა გენერალური ერთობლიობიდან შერჩეული შედარებით მცირე ნაწილია.

ჰიპოთეზა კვლევის პრობლემის პირობითი პასუხი.

ჰიპოთეზის ტესტირება მეთოდი, რომელიც შერჩევის სტატისტიკაზე დაყრდნობით საშუალებას გვაძლევს შევაფასოთ ჰიპოთეზები მთლიანი ერთობლიობის პარამეტრების შესახებ.

ჰისტოგრამა გრაფიკი, რომელიც გვიჩვენებს ინტერვალური ცვლადების სიხშირეებისა და ხვედრითი წილებს შორის განსხვავებებს. კატეგორიები გამოსახულია ერთმანეთზე მიწყობილი სვეტების სახით, რომელთა სიგანე კატეგორიის მნიშვნელობას შეესაბამება, ხოლო სიმაღლე — სიხშირეს.

I ტიპის შეცდომა ალბათობა, რომელიც ასოცირდება ნულოვანი ჰიპოთეზის უარყოფასთან, როდესაც ის სწორია.

II ტიპის შეცდომა ალბათობა, რომელიც ასოცირდება ნულოვანი ჰიპოთეზის არ უარყოფასთან, როდესაც ის არასწორია.

ინტერკვარტილური დიაპაზონი (IQR) ზომავს განაწილების შუა 50%-ის საზღვრებს (სიგანეს). ის გაიანგარიშება როგორც ჭედა და ზედა კვარტილების სხვაობა (Q_1 და Q_2).

ინტერვალურ-ფარდობითი გაზომვა ცვლადის მნიშვნელობათა გაზომვა საერთო ერთეულით გამოსახვის გზით.

ჩარეული კავშირი კავშირის სახე, რომლის დროსაც საკონტროლო ცვლადი ახდენს შუალედურ გავლენას დამოკიდებულ და დამოუკიდებელ ცვლადს შორის კავშირზე.

ჩარეული ცვლადი საკონტროლო ცვლადი, რომელიც უკავშირდება დამოუკიდებელ ცვლადს და მას აქვს შუალედური ეფექტი დამოკიდებულ და დამოუკიდებელ ცვლადებს შორის.

ჯგუფშიდა კვადრატების ჯამი, SSW თითოეული ჯგუფის დისპერსიის კვადრატების ჯამი, ჯგუფისა და დაკვირვების მონაცემების საშუალოების მიხედვით.

ჯგუფთა შორის კვადრატების ჯამი, SSB დისპერსიის კვადრატების ჯამი შერჩევის თითოეული საშუალოსა და საერთო საშუალოს მიმართ.

ჯგუფთაშორისი კვადრატების ჯამი (SSB) დისპერსიის კვადრატების ჯამი შერჩევის თითოეული საშუალოსა და საერთო საშუალოს შორის.

ჯვარედინი ტაბულაცია ორ ცვლადს შორის კავშირის ანალიზის ტექნიკა, რომლის დროსაც ორი ცვლადის მნიშვნელობები ორგანიზებულია ორგანზომილებიანი ცხრილის სახით.

კავშირის ასიმეტრიული საზომი საზომი, რომლის მნიშვნელობა ცვალებადობს იმაზე დამოკიდებულებით თუ რომელ ცვლადს აირჩევთ დამოკიდებულ ან დამოუკიდებელ ცვლადად.

კავშირის საზომი ერთიანი შემაჯამებელი მაჩვენებელი, რომელიც ასახავს ცვლადებს შორის კავშირის სიძლიერეს, აჩვენებს დამოკიდებული ცვლადების პროგნოზირების შესაძლებლობებს დამოუკიდებელი ცვლადების მიხედვით და ხშირ შემთხვევაში იძლევა კავშირის მიმართულების შეფასების საშუალებას.

კავშირის სიმეტრიული საზომი საზომი რომლის მნიშვნელობა დამოუკიდებელია იმის მიუხედავად თუ რომელ ცვლადს ჩავთვლით დამოკიდებულად ან დამოუკიდებლად.

კენდალის ტაუ-b ორდინალური ცვლადებს შორის კავშირის სიმეტრიული საზომი. გამოსაგან

განსხვავებით ამ მაჩვენებლის გამოთვლისას მხედველობაში მიიღება, როგორც დამოკიდებული, ასევე დამოუკიდებელი ცვლადების მიხედვით დალაგებული მონაცემთა წყვილები. მნიშვნელობათა სიმრავლე 0.0-დან +/-1.0 შუალედია და ასახავს ცვლადთა შორის კავშირის სიძლიერესა და მიმართულებას.

კლასეთი ჰიპოთეზა (H_1) ასახავს ძირითად ჰიპოთეზას. ის ყოველთვის მთლიანი ერთობლიობის პარამეტრებით გამოისახება, მაგრამ მისი კონკრეტული ფორმა განსხვავებულია.

ლამბდა კავშირის ასიმეტრიული საზომი; ლამბდა გამოიყენება ნომინალური ცვლადებისათვის და იღებს მნიშვნელობებს 0.0-დან 1.0-მდე. იგი გვიჩვენებს დამოკიდებულ და დამოუკიდებელ ცვალდებს შორის კავშირის სიძლიერეს.

მარცხენა განაწილების ტესტი ცალმხრივი ტესტი, რომლის დროსაც გაკეთებულია დაშვება რომ შემთხვევითი შერჩევის შედეგი მოხვდება განაწილების ბოლო მარცხენა არეში.

მარცხენამხრივი კრიტერიუმი ცალმხრივი კრიტერიუმი, რომელშიც შერჩევის შედეგი შერჩევითი განაწილების მარცხენა მხარეს უნდა შეესაბამოთ.

მარჯვენამხრივი კრიტერიუმი ცალმხრივი კრიტერიუმი, რომელშიც შერჩევის შედეგი შერჩევითი განაწილების მარჯვენა მხარეს უნდა შეესაბამოთ.

მარტივი შემთხვევითი შერჩევა შერჩევის მოდელი, რომელიც ისეა აგებული, რომ (1) მთლიანი ერთობლიობის ყველა წევრს შერჩევაში მოხვედრის თანაბარი შანსი აქვს და (2) N წევრების თითოეულ კომბინაციას შერჩევაში მოხვედრის თანაბარი შანსი აქვს.

მედიანა მნიშვნელობა, რომელიც განაწილებას ორ თანაბარ ნაწილად ყოფს. ის გაიანგარიშება ორდინალური და გაზომვის უფრო მაღალი დონის მქონე ცვლადებისათვის.

მოდა კატეგორია ან მნიშვნელობა, რომელსაც ყველაზე მაღალი სიხშირე (ან ხვედრითი წილი) აქვს განაწილებაში. ის ხშირად გამოიყენება ნომინალური ყველაზე მეტად განმეორებადი კატეგორიის დასახასიათებლად.

მონაცემები რიცხვების სახით წარმოდგენილი ინფორმაცია, რაც ექვემდებარება სტატისტიკურ ანალიზს.

მონაცემთა შორის კვადრატების საშუალო მნიშვნელობათა კვადრატების ფარდობა თავისუფლების შესაბამის ხარისხთან.

მოსალოდნელი სიხშირეები (f_e) ორგანზომილებიანი ცხრილის უჯრების სიხშირეთა ის მნიშვნელობები, რაც მიიღება იმ დაშვების საფუძველზე, რომ ცხრილში წარმოდგენილი ორი ცვლადი სიდიდე სტატისტიკურად დამოუკიდებელია.

მრავლობითი დეტერმინაციის კოეფიციენტი (R^2) საზომი, რომელიც ასახავს მთლიანი ვარიაციის წილს დამოკიდებულ ცვლადში, რაც შესაძლებელია აიხსნას ორ ან მეტი დამოუკიდებელი ცვლადით.

მრავლობითი რეგრესია ბივარიანტული რეგრესიის გაფართოება, რომელიც მოიცავს ორ ან მეტ დამოუკიდებელ ცვლადს. მისი ზოგადი ფორმა ორი დამოუკიდებელი ცვლადით შემდეგია: $\hat{Y} = a + b_1(X_1) + b_2(X_2)$

მსგავსი წყობის წყვილები (Ns) დაკვირვების დაწყვილებული მონაცემები დადებითი კავშირით. რესპონდენტის მაღალი რიგობითი პოზიცია ერთი კატეგორიის მიხედვით დამოუკიდებელი ცვლადისათვის ეთანადება მის ასევე მაღალ რიგობით პოზიციას მეორე კატეგორიის მიხედვით დამოკიდებული ცვლადისათვის.

ნაშთის კვადრატების ჯამი (SSE) დაკვირვებისა და განტოლების საფუძველზე მიღებული დამოკიდებული ცვლადის (Y) მნიშვნელობებს შორის სხვაობის კვადრატების ჯამი.

ნდობის ინტერვალი მნიშვნელობათა დიაპაზონი, რაც ნდობის ხარისხითაა განსაზღვრული.

ნდობის ხარისხი მოსალოდნელობა, რომელიც გამოსახულია პროცენტული ან ალბათობის სახით, რაც შეესაბამება იმ ინტერვალს რომელშიც ხვდება პოპულაციის პარამეტრი.

ნეგატიურად გადახრილი განაწილება განაწილება რამდენიმე ექსტრემალურად დაბალი მნიშვნელობით.

ნომინალური გაზომვა სხვადასხვა კატეგორიისათვის ციფრებისა და სიმბოლოების მინიჭება სახელდების, აღნიშვნის ან დაკვირვებების დაჯგუფების მიზნით.

ნორმალური განაწილება არის ზარის ფორმის

სიმეტრიული თეორიული განაწილება, რომლის საშუალო, მედიანა და მოდა მის უმაღლეს წერტილში ერთმანეთს ემთხვევა, ხოლო სიხშირეები თანდათან მცირდება მრუდის ბოლოებისკენ.

ნულოვანი ჰიპოთეზა (H_0) „არანაირი განსხვავების“ პირობა, რომელიც ეწინააღმდეგება კვლევით ჰიპოთეზას და ყოველთვის გამოისახება მთლიანი ერთობლიობის პარამეტრებით.

ორ შერჩევის საშუალოებს შორის განსხვავების შერჩევითი განაწილება თეორიული ალბათური განაწილება, რომელიც მიიღება საშუალოებს შორის ყველა განსხვავების ($\bar{Y}_1$-$\bar{Y}_2$) გაანგარიშების საფუძველზე ორი ერთობლიობიდან N_1 და N_2 ყველა შესაძლო შერჩევის განხორციელების გზით, სადაც N_1 და N_2 50-ზე მეტია.

ორმხრივი კრიტერიუმი ჰიპოთეზის ტესტირების სახე, რომელიც მოიცავს კვლევით ჰიპოთეზას მიმართულების გარეშე. ჩვენ ერთნაირად გვაინტერესებს თუ რამდენად მეტია ან ნაკლებია ერთი სიდიდე მეორეზე. შერჩევის შედეგი შეიძლება მოთავსებული იყოს შერჩევითი განაწილების როგორც მარცხენა, ისე დაბალ დაბოლოებებზე.

P სიდიდე მიღებული Z სიდიდის შესაბამისი ალბათობა.

პარამეტრი მაჩვენებელი (მაგ. საშუალო ან სტანდარტული გადახრა), რომელიც გამოიყენება პოპულაციის განაწილების აღწერისათვის.

პერცენტილი მნიშვნელობა, რომლის ქვემოთ განაწილების გარკვეული პროცენტი ხვდება.

პირდაპირი მიზეზ-შედეგობრივი კავშირი ბივარიანტული კავშირი, რომელიც სხვა დამატებითი, თეორიულად შესაბამისი ცვლადით არ აიხსნება.

პირობითი კავშირი კავშირი რომლის დროსაც საკონტროლო ცვლადის ეფექტი დამოკიდებულ ცვლადზე პირობითია დამოუკიდებულ ცვლადთან მიმართებიდან გამომდინარე. დამოკიდებულ და დამოუკიდებელ ცვლადს შორის კავშირი იცვლება საკონტროლო ცვლადის განსხვავებული პირობების მიხედვით.

პირსონის კორელაციის კოეფიციენტი (r) კვადრატული ფესვი r^2-დან. ინტერვალურ-ფარდობით მონაცემებს შორის კავშირის საზომი, ასახავს ცვლადებს შორის წრფივი კავშირის სიძლიერეს. იღებს დადებით და უარყოფით მნიშვნელობებს.

პირსონის მრავალობითი დეტერმინაციის კოეფიციენტი (R) წრფივი კავშირის საზომი ერთ დამოკიდებულ და ორ ან მეტ დამოუკიდებელ ცვლადებს შორის.

პოპულაცია ყველა, რომელიც შეიძლება შეიცავდეს ბების მთლიან სიმრავლეს (ინდივიდებს, საგნებს ან ჯგუფებს), რაც მკვლევარს აინტერესებს.

პოზიტიურად გადახრილი განაწილება განაწილება რამდენიმე ექსტრემალურად მაღალი მნიშვნელობით.

პროცენტული (ხვედრითი წილების) განაწილება ცხრილის სახით წარმოდგენილი ცვლადის თითოეულ კატეგორიაში მოხვედრილი შემთხვევების ხვედრითი წილები.

პროცენტული წილი ფარდობითი სიხშირე, რომელიც მიიღება თითოეული კატეგორიის სიხშირის გაყოფით საერთო ჯამურ სიხშირეზე და 100-ზე გამრავლებით.

პროპორციული სტრატიფიცირებული შერჩევა თითოეული ქვეჯგუფიდან ამორჩეული შერჩევა მთლიან ერთობლიობაში ქვეჯგუფის ზომის პროპორციულია.

რეგრესიის კვადრატების ჯამი (SSR) ასახავს პროგნოზის შეცდომის შემცირების ზომის წრფივი პროგნოზირების განტოლების გამოყენების შედეგად.

საკონტროლო ცვლადი დამატებითი ცვლადი ბივარიანტული კავშირისას. ცვლადი წარმოადგენს საკონტროლო ცვლადს, როდესაც შედეგობაში მიიღება მისი ეფექტი სხვა ცვლადებზე ბივარიანტული კავშირისას.

საპირისპირო წყობის მონაცემთა წყვილი (Nd) დაკვირვების მონაცემთა წყვილები, რომელთა შორის არსებობს უარყოფითი კავშირი. წყვილის ერთი წევრი იღებს მაღალ მნიშვნელობას დამოუკიდებელი ცვლადის მიხედვით, ხოლო დაბალ მნიშვნელობას დამოკიდებული ცვლადის მიხედვით.

საშუალო არითმეტიკული საშუალო, რომელიც მიიღება ცვლადის მნიშვნელობების ჯამის გაყოფით მათსავე რაოდენობაზე.

საშუალო კვადრატული ნაშთი საშუალო მაჩვენებელი, რომელიც გამოითვლება ნაშთების კვადრატების ჯამის (*SSE*) გაყოფით თავისუფლების შესაბამის ხარისხთან.

საშუალო კვადრატული რეგრესია (SSR) საშუალო მაჩვენებელი, რომელიც გამოითვლება რეგრესიის კვადრატების ჯამის გაყოფით თავისუფლების შესაბამის ხარისხზე.

საშუალო სიდიდის შერჩევითი განაწილება შერჩევითი საშუალოების თეორიული ალბათური განაწილებაა, რომელიც შეიძლება მივიღოთ მთლიანი ერთობლიობიდან ერთი ზომის ყველა შესაძლებელი შერჩევის განხილვით.

საშუალოს სტანდარტული შეცდომა შერჩევითი განაწილების სტანდარტული გადახრა. გვიჩვენებს საშუალოს დისპერსიის სიდიდეს შერჩევით განაწილებაში.

საუკეთესო შესაბამისობის რეგრესიის წრფე რეგრესიის წრფე, რომლისათვისაც ნაშთების კვადრატების ჯამი $\sum e^2$ მინიმალურია.

საზღვრითი მნიშვნელობები სვეტებისა და სტრიქონების ჯამი ორგანზომილებიან ცხრილში.

შეცდომის პროპორციული შემცირების (PRE) კონცეფცია, რომელიც ეფუძნება სტატისტიკური კავშირის შეფასების რამდენიმე საზომის განმარტებას და ინტერპრეტაციას. PRE-ის მაჩვენებლები მიიღება იმ შეცდომების შედარებით, რაც ერთის მხრივ მიიღება დამოკიდებული ცვლადების მნიშვნელობების პროგნოზირებისას დამოუკიდებელი ცვლადებისაგან განცალკევებით, და მოერეს მხრივ იმ შეცდომებთან, რაც მიიღება დამოკიდებული ცვლადების მნიშვნელობების პროგნოზირებით დამოუკიდებელი ცვლადების მხედველობაში მიღებით.

შეცდომის საზღვრები ნდობის ინტერვალის რადიუსი.

შეფასება პროცედურა, რომლის დროსაც ვირჩევთ შერჩევით ერთობლიობას და მასზე მიღებულ გამოთვლის შედეგებს ვიყენებთ პოპულაციის პარამეტრის მისაღებად.

შერჩევის შეცდომა არის განსხვავება შერჩევის საშუალებით შეფასებული მთლიანი ერთობლიობის პარამეტრსა და მთლიანი ერთობლიობის რეალურ პარამეტრს შორის.

შერჩევითი ერთობლიობა მთლიანი ერთობლიობიდან შერჩეული დაკვირვებების ქვესიმრავლე.

შერჩევითი განაწილება ჩვენთვის საინტერესო სტატისტიკისთვის საჭირო ყველა შესაძლებელი შერჩევითი სიდიდის თეორიული ალბათური განაწილება.

სიმეტრიული განაწილება სიხშირეები განაწილების მარჯვენა და მარცხენა მხარეს იდენტურია, ერთი ნაწილი მეორის სარკისებრ ანარეკლს წარმოადგენს.

სისტემატური შემთხვევითი შერჩევა არის მეთოდი, რომელშიც მთლიანი ერთობლიობიდან ნებისმიერი K-ური წევრის (K კოეფიციენტია, რომელიც მიიღება მთლიანი ერთობლობის ზომის ფარდობით შერჩევის ზომასთან) შერჩევა მოხდება მხოლოდ მაშინ, როდესაც შემთხვევით მოხდება შერჩევის პირველი წევრის ამორჩევა საწყისი K წევრებიდან.

სიხშირის კუმულაციური განაწილება განაწილება, რომელიც გვიჩვენებს თითოეული კატეგორიისათვის ცვლადის პროცენტულ წილებს.

სტანდარტული ნორმალური განაწილება სტანდარტული ქულებით (Z) წარმოდგენილი ნორმალური განაწილება, სადაც საშუალო = 0-ს, ხოლო სტანდარტული გადახრა = 1-ს.

სტანდარტული Z ქულა სტანდარტული გადახრების რაოდენობა, რომელიც გვიჩვენებს მოცემული ნედლი ქულის საშუალოდან დაშორების ზომას ზემოთ ან ქვემოთ.

სტატისტიკა პროცედურების ერთობლიობაა, რომელიც გამოიყენება სოციალური დარგის მეცნიერების მიერ ინფორმაციის თავმოყრის, განზოგადებისა და გადმოცემისათვის.

სტრატიფიცირებული შემთხვევითი შერჩევა მიიღება: (1) ჩვენთვის მნიშვნელოვანი 1 ან 2 ცვლადის ბაზაზე ერთობლიობის ქვეჯგუფებად დაყოფით და (2) თითოეულიდან მარტივი შემთხვევითი შერჩევის განხორციელებით.

სვეტებიანი დიაგრამა დიაგრამა, რომელიც გვიჩვენებს განსხვავებებს ნომინალური და ორდინალური ცვლადების სიხშირეებს ან ხვედრით წილებს შორის. კატეგორიები წარმოდგენილია თანაბარი სიგანის მართკუთხედების სახით, რომელთა სიმაღლე სიხშირის, ან ხვედრითი წილის პროპორციულია.

სვეტის ცვლადი ცვლადი, რომლის კატეგორიები წარმოდგენილია ცხრილის სვეტებში.

t განაწილება მრუდების ოჯახი, რომელთაგან თითოეული განსაზღვრულია თავისუფლების

ხარისხით (*df*). ის მაშინ გამოიყენება, როდესაც მთლიანი ერთობლიობის სტანდარტული გადახრა უცნობია და სტანდარტული შეცდომის გაანგარიშება ხდება შერჩევის სტანდარტული გადახრით.

t სტატისტიკა (მიღებული) კრიტერიუმის სტატისტიკა მთლიანი ერთობლიობის საშუალოს ნულოვანი ჰიპოთეზის ტესტირებისათვის, როდესაც მთლიანი ერთობლიობის სტანდარტული გადახრა უცნობია და მისი გაანგარიშება ხდება შერჩევის სტანდარტული გადახრით.

თანხმობის ხი-კვადრატ კრიტერიუმი შემოკმებელი მაჩვენებელია დაკვირვების შედეგად მიღებულ სიხშირეებსა (f_0) და მოსალოდნელ სიხშირეებს (f_e) შორის არსებული სხვაობებისათვის.

თავისუფლების ხარისხი (df) მნიშვნელობათა რაოდენობა, რომლებიც თავისუფლად ცვალებადობს სტატისტიკური მაჩვენებლების გამოთვლისას.

თეორია ცალკეული ინდივიდებისა ან ჯგუფების ორი ან მეტი მახასიათებლის ურთიერთდამოკიდებულების მეცნიერული ახსნაა.

თვისობრივი ვარიაციის ინდექსი (IQV) ვარიაციის საზომია ნომინალური ცვლადებისათვის. ის ეფუძნება განაწილებაში არსებული განსხვავებების რაოდენობის ფარდობას მაქსიმალურად შესაძლებელ განსხვავებათა რიცხვზე.

უარყოფითი კავშირი ბივარიანტული კავშირი ორ ცვლადს შორის, რომლის დროსაც ცვლადები მნიშვნელობებს საწინააღმდეგო მიმართულებით იღებს.

უჯრა ორგანზომილებიანი ცხრილის სვეტისა და სტრიქონის თანაკვეთა.

უმცირეს კვადრატთა მეთოდი უმცირეს კვადრათა წრფის მიღების ტექნიკა.

ვარიაციის საზომები რიცხვები, რომლებიც გამოიყენება ცვლადის განაწილების მრავალფეროვნებისა და ვარიაციის დახასიათებისათვის.

წერტილოვანი შეფასება შერჩევის სტატისტიკა, რომელიც გამოიყენება მთლიანი ერთობლიობის პარამეტრის ზუსტი სიდიდის შესაფასებლად.

წერტილთა გაბნევის დიაგრამა ინტერვალურ-ფარდობით მონაცემებს შორის არსებული კავშირის ვიზუალიზაციის მეთოდი.

წრფივი გრაფიკი გრაფიკი, რომელიც გვიჩვენებს ორ ინტერვალურ-ფარდობითი ცვლადების სიხშირეებსა და ხვედრითი წილებს შორის განსხვავებებს. ცვლადის თითოეული მნიშვნელობები გამოსახულია წერტილებით და შეერთებულია სწორი ხაზით.

წრფივი კავშირი კავშირის სახე ორ ინტერვალურ-ფარდობით ცვლადს შორის, რომლის დროსაც ცვლადებს შორის კავშირი შესაძლებელია გამოისახოს წრფის სახით.

წრიული დიაგრამა დიაგრამა, რომელიც გვიჩვენებს განსხვავებებს ნომინალური ან ორდინალური ცვლადების სიხშირეებს ან ხვედრით წილებს შორის. მასში კატეგორიები სეგმენტების სახითაა წარმოდგენილი, რომელთა ჯამი 100%-ის ტოლია.

ხაზოვანი დიაგრამა გვიჩვენებს განსხვავებებს ინტერვალურ-შეფარდებითი ცვლადების სიხშირეებსა და ხვედრით წილებში. წერტილები, რომლებიც თითოეული კატეგორიის სიხშირეს შეესაბამება, თავსდება ამ კატეგორიის შუა წერტილის თავზე და შემდეგ სწორი ხაზით ერთდება.

ხი-კვადრატ კრიტერიუმი დასკვნითი სტატისტიკის მეთოდია, რომელიც ამოწმებს ორგანზომილებიანი ცხრილის სახით მოცემულ, ნომინალურ და ორდინალურ მონაცემებს შორის არსებული კავშირის მნიშვნელოვნებას.

Z სტატისტიკა (მიღებული) კრიტერიუმის სტატისტიკა, რომელიც მიიღება შერჩევითი სტატისტიკის (როგორიცაა საშუალო სიდიდე) გარდაქმნის შედეგად Z ქულად. Z-ის მისაღები ფორმულა იცვლება ტესტირებიდან ტესტირებამდე.

შენიშვნები

თავი 1

1. U.S. Census Bureau, American Community Survey, 2008
2. U.S. Census Bureau, Statistical Abstract of the United States, 2010, Table 603.
3. Rampell, Catherine. "Women Now a Majority in American Workplaces," The New York Times, February 5, 2010.Retrieved February 5, 2010, from *www.nytimes.com/2010/02/06/business/economy/06women.html.*
4. U.S. Census Bureau, Statistical Abstract of the United States, 2010, Table 607
5. Chava Frankfort-Nachmias and David Nachmias, Research Methods in the Social Sciences (New York: Worth Publishers, 2000), p. 56.
6. Barbara Reskin and Irene Padavic, Women and Men at Work (Thousand Oaks, CA: Pine Forge Press, 2002), p.65., 2002, p. 144
7. Frankfort-Nachmias and Nachmias, 2000, p. 50.
8. იხ. ზემოთ გვ. 52
9. Patricia Hill Collins, "Toward a New Vision: Race,Class and Gender as Categories of Analysis and Connection"(Keynote address at Integrating Race and Gender Into the College Curriculum, a workshop sponsored by the Center for Research on Women, Memphis State University, Memphis, TN, 1989).

თავი 2

1. Gary Hytrek and Kristine Zentgraf, America Transformed: Globalization, Inequality and Power (New York: Oxford University Press, 2007).
2. U.S. Census Bureau, American Community Survey, 2008
3. იხ. ზემოთ, 2008
4. U.S. Census Bureau, Statistical Abstract for the United States, 2010, Table 43.
5. იხ. ზემოთ, 2010 ცხრილი 69
6. იხ. ზემოთ, 2010 ცხრილი 56
7. David Knoke and George W. Bohrnstedt, Basic Social Statistics (New York: Peacock, 1991), p. 25.
8. იხ. ზემოთ, გვ 41
9. კვლევითი ლიტერატურის მიმოხილვის სექციის იდეა, რომელიც ამ წიგნის უმრავლეს თავებშია მოცემული ალებულია შემდეგი წიგნიდან: Joseph F. Healey, Statistics: A Tool for Social Research, 5th ed. (Belmont, CA:Wadsworth, 1999).
10. Eric Fong and Kumiko Shibuya, "Urbanization and Home Ownership: The Spatial Assimilation Process in U.S. Metropolitan Areas," Sociological Perspectives 43, no. 1 (2000): 137–157.

თავი 3

1. The U.S. Census Bureau notes that persons of Hispanic origin may be of any race.
2. U.S. Census Bureau, Marital Status and Living
3. Arrangements: March 1996, Current Population Reports,P20–496, 1998, p. 5.
4. U.S. Census Bureau, 65+ in America, Current Population Reports, Special Studies, P23–190, 1996, pp. 6-1, 6-8.
5. USA Today, November 10, 1992.
6. Thomas G. Donlan, A World of Wealth: How Capitalism Turns Profits Into Progress (Upper Saddle River, NJ: FT Press,2008), p. 166
7. U.S. Census Bureau, 65+ in America, Current Population Reports, Special Studies, P23–190, 1996, pp. 2–3.
8. იხ. ზემოთ გვ. 6-2

9. Edward R. Tufte, The Visual Display of QuantitativeInformation (Cheshire, CI: Graphics Press, 1983), p. 53.
10. U.S. Census Bureau, Statistical Abstract for the UnitedStates, 2010, Table 224.
11. Catherine Freeman, Trend in Educational Equity of Girls and Women: 2004 (Washington, DC: U.S. Departmentof Education, National Center for Education Statistics, 2004), pp. 9-10.

თავი 4

1. Douglas S. Massey. "The Social and Economic Origins of Immigration," Annals, AAPSS (July 1990): 510.
2. Bureau of Labor Statistics, Current Population Survey 2009, Household Data Annual Averages, Table 37.
3. U.S. Census Bureau, Statistical Abstract for the United States, Table 313: 2007 Hate Crimes Reported by State.
4. ეს წესი ადაპტირებულია: David Knoke and GeorgeW. Bohrnstedt, Basic Statistics (New York: Peacock Publishers,1991), pp. 56–57.
5. ცხრილ 4.4-ში წარმოდგენილი განაკვეთები გამოთვლილია სხვადასხვა ქალაქების ზომების მიხედვით. საშუალო 13.95 იწოდება შეუწონელ საშუალოდ. ეს არ არის იგივე, რაც პოპულაციაში დანაშაულის დონე ყველა ქალაქისათვის.
6. სამი ცვლადი TVHOURS, SIBS, და EDUC აღებულია ზოგადი სოციოლოგიური კვლევის მონაცემებიდან. ცვლადი EDUC ჩაწერილ იქნა სხვა ცვლადის სახით, რომელიც მხოლოდ იმ რესპონდენტებს მოიცავს, რომელთაც არ აქვთ უმაღლესი განათლება.

თავი 5

1. Johnneta B. Cole, "Commonalities and Differences," in Race, Class, and Gender, ed. Margaret L. Andersen and Patricia Hill Collins (Belmont, CA: Wadsworth, 1998), pp. 128–129.
2. იხ. ზემოთ, გვ. 12-13
3. U.S. Census Bureau, Statistical Abstract for the United States, 2010, Table 11.
4. William H. Frey, "The Diversity Myth," American Demographics 20, no. 6 (1998): 39-43.
5. Douglas S. Massey and Kristin Espinosa, "What's Driving Mexico–U.S. Migration? A Theoretical, Empirical, and Policy Analysis," American Journal of Sociology 102 (1997): 939–999.
6. Michael White, "Segregation and Diversity Measures in Population Distribution," Population Index 52, no. 2 (1986):198–221.
7. ბოლო აღწერის მონაცემებმა აჩვენა, რომ რეცესიამ მიგრაციის ეს ტენდენცია შეანელა 2008-2009 წლებში.
8. 65 წელს ზემოთ მოსახლეობის პროცენტული ზრდა თითოეული შტატისა და რეგიონისათვის მიღებულ იქნა შემდეგი ფორმულის გამოყენებით:
9. პროცენტული ზრდა = [(2015 პოპულაცია − 2008 პოპულაცია)/2008 პოპულაცია] ×100
10. ექსტრემალური მნიშვნელობები იწოდება გარე წევრებად. SPSS-ის გამოთვლები მოიცავს გარე წევრებს მართკუთხა დიაგრამების შედგენისა და IQR მაჩვენებლის გამოთვლისას. თუ მაჩვენებლებში არის IQR-ზე 1.5-ჯერ ნაკლები მონაცემები, SPSS გამოიტანს ინდივიდუალურ შემთხვევებს. ეს უკანასკნელი ყოველთვის უნდა გავითვალისწინოთ, როცა განაწილების ფორმაზე ვმსჯელობთ.
11. ვარიაციის გამოსათვლელ ფორმულაში გამოიყენება $N - 1$, რადგან იგი გამოითვლება შერჩევითი ერთობლიობიდან პოპულაციის ტენდენციის შესწავლის მიზნით. $N - 1$ იძლევა უკეთეს მიახლოებას და იგი ასევე გამოიყენება SPSS-ის გამოთვლებისას.
12. Stephanie A. Bohon, Monica Kirkpatrick Johnson,and Bridget K. Gorman, "College Aspirations and Expectations among Latino Adolescents in the United States," Social Problems53, no. 2 (2006): 207–225.
13. იხ. ზემოთ გვ. 210
14. იხ. ზემოთ გვ.213
15. იხ. ზემოთ
16. პროგრამის არჩეული პარამეტრებიდან გამომდინარე SPSS ან დაამრგვალებს ან არა მოცემულ რიცხვებს. ჩვენს მაგალითში თქვენ შესაძლოა მიიღოთ IQR-ის მნიშვნელობა 13-ის ტოლი ნაცვლად 12.50-ისა. თუ ეს ასე მიიღეთ, გააქტიურეთ აღწერითი

სტატიტიკის ცხრილი შედეგების ფანჯარაში მასზე ორჯერ დაწკაპუნებით. აირჩიეთ ცხრილის ის უჯრა, რომლის შეცვლაც თქვენ გინდათ (ჩვენს მაგალითში უჯრა, რომელშიც შეიცავს რიცხვ „13“-ს). შედეგ დაკლიკეთ „უჯრათა ოფციაზე“. ბრძანების „მნიშვნელობის ფორმატირება“ გექნებათ არჩევანი შეცვალოთ რიცხვების ჩაწერის ფორმა თქვენთვის მისაღები ფორმატით. არჩევის შემდეგ დაკლიკეთ „Apply“ ან „OK“.

თავი 7

1. Richard Maisel and Caroline Hodges Persell, How Sampling Works (Thousand Oaks, CA: Pine Forge Press, 1996).
2. C. Stephen Layman's, ThePower of Logic (Mountain View, CA: Mayfield, 2004).
3. აღნიშნული შემოთავაზებულ იქნა: David C. Howell in Statistical Methods for Psychology (Duxbury, MA: Wadsworth, 2006)
4. იხილეთ: Chava Frankfort-Nachmias and David Nachmias, Research Methods in the Social Sciences (New York: Worth Publishers, 2007), pp. 167–177.
5. შერჩევის შეცდომას უფრო დაწვრილებით შემდეგ ქვეთავებში განვიხილავთ.
6. 20 ინდივიდის ზომის პოპულაცია ცხრილ 7.3-ში არის სასრული პოპულაცია. სასრული პოპულაცია არის წევრთა სასრული წევრებისაგან შემდგარი პოპულაცია. ჩვენი სასრული პოპულაციის მაგალითი მოიცავს ყველა ქალს 2008 წელს სამუშაო ძალაში და ყველა საზოგადოებრივ საავადმყოფოს ნიუ-იორკში. პოპულაცია მიჩნეულია უსასრულოდ თუ მის წევრების რაოდენობა არათვლადია. უსასრულო პოპულაციის მაგალითია ყველა ქალის რაოდენობა სამუშაო ძალაში როგორც წარსულში, ასევე მომავალში. იმ შერჩევების უმრავლესობა, რომლებსაც სოციალური დარგის მეცნიერებები სწავლობენ სასრული პოპულაციებიდან არის აღებული.
7. აქ ჩვენ ვიყენებთ იდეალიზებულ მაგალითს, რომელშიც შერჩევის განაწილების რეალური გამოთვლა ხდება. მაგრამ გახსოვდეთ, რომ რეალობაში არ ხდება შერჩევის განაწილების გამოთვლა, რადგან ის ჩვეულებრივ უსასრულოა.

თავი 8

1. ადაპტირებულია: Marta Tienda and Franklin D. Wilson, "Migration and the Earnings of Hispanic Men," American Sociological Review 57 (1992): 661–678.
2. The U.S. Census 2000, IPUMS (Integrated Public Use Micro data Series).
3. გამოთვლებისას კავშირს შერჩევის ზომასა და ინტერვალის საზღვრებს შორის კავშირს ასევე აქვს მნიშვნელობა შერჩევის პროპორციებში.
4. "More Americans Favor than Oppose Arizona Immigration Law," Gallup Poll, April 27–28, 2010
5. Frank Newport, "Americans' Views on Healthcare Law Remain Stable," Gallup, April 15, 2010.
6. Michael Luo and Megan Thee Brenan,"Poll reveals Trauma of Joblessness in U.S." The New York Times, December 15, 2009
7. Andrew Kohut, "Getting It Wrong," The New York Times, January 10, 2008.
8. Data sources: *http://stats.org/faq_margin.htm; www.charneyresearch.com; www.pollingreport.com/sampling.htm;\ www.robertniles.com/stats/margin.shtml; http://us-elections.suite101.com/article.cfm/new_hampshire_primary_poll_wrong*
9. Data from "Little Consensus on Global Warming," Pew Research Center, July 12, 2006.
10. Data from "Enthusiastic Voters Prefer GOP by 20 Points in 2010 Vote," Gallup, April 27, 2010.
11. Data from "Internet's Broader Role in Campaign 2008," Pew Research Center, January 11, 2008.
12. Data from "The Millenials: Confident. Connected. Open to Change," Pew Research Center, February 24, 2010.

თავი 9

1. Steve Hargreaves, Gas prices hit working class, 2007. Retrieved February 11, 2008, from *http://money.cnn.com/2007/11/13/news/economy/gas_burden/index.htm*

2. American Automobile Association, 2010. Daily Fuel Gauge Report, April 27. Retrieved April 27, 2010, from *www.fuelgaugereport.com/*
3. შერჩევის ვარიაციის გამოთვლისათვის თავდაპირველად უნდა გამოვთვალოთ საშუალო მნიშვნელობა. რადგან საშუალოდან გადახრების ჯამი უნდა უტოლდებოდეს 0-ს, მხოლოდ *N - 1* ოდენობის მნიშვნელობები თავისუფლად ცვალებადობენ.
4. Current Population Survey, PINC-02. People 18 YearsOld and Over, by Total Money Income in 2008.
5. თავისუფლების ხარისხის ფორმულა ეფუძნება: Dennis Hinkle,William Wiersma, and Stephen Jurs, Applied Statistics for the Behavioral Sciences (Boston: Houghton Mifflin, 1998), p. 268.
6. Lloyd D. Johnson, Patrick M. O'Malley, Jerald G. Bachman, and John E. Schulenberg, Monitoring the Future National Results on Adolescent Drug Use: Overview of Key Findings, 2008 (Bethesda, MD: National Institute on Drug Abuse, 2009).
7. შერჩევის პროპორციები არის მიუკერძოებელი შეფასებები შესაბამისი პოპულაციის პროპორციების. შესაბამისად, შეგვეძლია გამოვიყენოთ Z სტატისტიკა, მიუხედავად იმისა რომ ამ შემთხვევაში სტანდარტული შეცდომა შერჩევითი ერთობლიობიდან არის მიღებული.
8. Paula Y. Goodwin, William D. Mosher, and Anjani Chandra, "Marriage and cohabitation in the United States: A statistical portrait based on Cycle 6 (2002) of the National Surveyof Family Growth," Vital Health Statistics 23, no. 28 (2010): 1–45.
9. მსგავსი განხილვა წარმოდგენილია: Joseph F. Healey, Statistics: A Tool for Social Research, 3rd ed. (Belmont, CA:Wadsworth, 1999), pp. 216–217.
10. Robert E. Jones and Shirley A. Rainey, "Examining Linkages Between Race, Environmental Concern, Health and Justice in a Highly Polluted Community of Color," Journal of Black Studies 36, no. 4 (2006): 473–496.
11. Marcelline Fusilier, Subhash Durlabhji, Alain Cucchi, and Michael Collins, "A Four Country Investigation of Factors Facilitating Student Internet Use," CyberPsychology and Behavior 8, no. 5 (2005): 454–464.

თავი 10

1. Damien Cave, "A Generation Gap Over Immigration," The New York Times, May 17, 2010. Retrieved May 2010, from *www.nytimes.com/2010/05/18/us/18divide.html*
2. USA Today, October 9, 1992.
3. National Center for Health Statistics, Summary Health Statistics for U.S. Population, U.S. National Health Interview Survey, 2002.
4. კავშირის არსებობის სრული შესწავლისათვის საჭიროა დასკვნითი სტატისტიკის მეთოდების გამოყენება. ეს საკითხები განხილულია მე-9, მე-11 და მომდევნო თავებში.
5. შევნიშნოთ, რომ ეს არის მონაცემთა მცირე ჯგუფი, რომელიც აღებულია ზოგადი სოციოლოგიური კვლევის მონაცემებიდან (GSS). აქ დაფიქსირებული კავშირი კერძლო სახლის ფლობასა და რასას შორის არ ნიშნავს ამ კავშირის აუცილებელ არსებობას უფრო დიდი შერჩევის შემთხვევაშიც.
6. პროცენტული გამოსახვის სხვა გზაა შემთხვევათა მთლიანი რაოდენობის მიხედვით გამოსახვა. მაგალითად, 7 რესპონდენტი (7.9%) 89 რესპონდენტიდან მხარს უჭერს აბორტის შესაძლებლობასა და უსაფრთხო სამუშაო გარემოს არსებობას. საერთო პროცენტულ გამოსახვებს არ აქვთ ისეთი ფართო გამოყენება კვლევებში, როგორც ეს სვეტებისა და სტრიქონების მიხედვით გამოსახული პროცენტული მაჩვენებლების შემთხვევაშია.
7. იგივე სამი თვისება განხილულია შემდეგ ნაშრომში: Joseph P. Healey in Statistics: A Tool for Social Research, 5th ed. (Belmont,CA: Wadsworth, 1999), pp. 314–320.
8. რელიგიურ მსახურებებზე დასწრება ჩაწერილ იქნა სამი კატეგორიის მიხედვით
9. ილუსტრირების მიზნებისათვის, განათლების დონისა და რელიგიურ მსახურებებზე დასწრების სიხშირის მხოლოდ ცალკეული შემთხვევებია აღებული.
10. მაგალითად, იხილეთ: Harris Mills, "Religion, Values, and Attitudes Toward Abortion," Journal for the Scientific Study of Religion 24, no. 2 (1985): 119–236.
11. Mario Renzi, "Ideal Family Size as an Inter-

vening Variable Between Religion and Attitudes Toward Abortion," Journal for the Scientific Study of Religion 14 (1975): 23–27.

12. იხ. ზემოთ
13. იხ. ზემოთ
14. ოჯახის იდეალური ზომა რესპონდენტის პასუხების მიხედვით ბავშვების ფოლის იდეალური რაოდენობის შესახებ კითხვაზე. მათ ვინც გასცეს პასუხი 2 ან ნაკლები ბავშვი ჩაწერილ იქნა როგორც მცირე ზომის ოჯახის მოსურნე, ხოლო ისინი რომელთაც გასცეს პასუხი სამი ან მეტი ბავშვი, ჩაწერილ იქნა როგორც დიდი ოჯახის მოსურნე.
15. William R. Arney and William H. Trescher, "Trends in Attitudes Toward Abortion, 1972–1975," Family Planning Perspective 8 (1976): 117–124.
16. William V. D'Antonio and Steven Stack, "Religion, Ideal Family Size, and Abortion: Extending Renzi's Hypothesis," Journal for the Scientific Study of Religion 19 (1980): 397–408.
17. Jacqueline Scott, "Conflicting Belief About Abortion: Legal Approval and Moral Doubts," Social Psychology Quarterly52, no. 4 (1989): 319–326.
18. იხ. ზემოთ, გვ. 322
19. Carol Gilligan, In a Different Voice (Cambridge, MA: Harvard University Press, 1982).
20. Patricia Hill Collins, "Toward a New Vision: Race, Class, and Gender as Categories of Analysis and Connection," key-note address at Integrating Race and Gender Into the College Curriculum, a workshop sponsored by the Center for Research on Women, Memphis State University, Memphis, TN, 1989.
21. Stanley Eitzen and Maxine Baca Zinn, "Structural Transformation and Systems of Inequality," in Race, Class, and Gender, ed. Margaret L. Andersen and Patricia Hill Collins (Belmont, CA: Wadsworth, 1998), pp. 233–237.
22. Lynn Weber Cannon, Elizabeth Higginbotham, and Marianne L. A. Leung, "Race and Class Bias in Research on Women: A Methodological Note," Research paper 5, presented at the Center for Research on Women, Memphis, TN, 1987.
23. იხ. ზემოთ
24. იხ. ზემოთ
25. ილუსტრირების მიზნებისათვის, მე-5 კატეგორია "სხვა", მოიცავს შემდეგ რელიგიურ ჯგუფებს: ბუდისტი, ჰინდუსი, სხვა ევროპული რელიგიები, ისლამი, მართლმადიდებელი ქრისტიანი, ქრისტიანი, ადგილობრივი ამერიკელი და შინაგანად არადენომინირებულები. თქვენს შედეგებში ეს ჯგუფები განცალკევებულად იქნება გამოტანილი.
26. ნაჩვენებია მხოლოდ თეთრკანიანები და შავკანიანები.

თავი 11

1. U.S. Census Bureau, Statistical Abstract to the United States, 2010, Table 224
2. Victor Saenz et al., First in My Family: A Profile of First-Generation College Students at Four Year Institutions Since 1971 (Los Angeles, CA: Higher Education Research Institute, 2007).
3. რადგან სტატისტიკური დამოკიდებულება სიმეტრიული თვისებაა, დამოკიდებული ცვლადის განაწილება თითოეული კატეგორიის მიხედვით იქნება ასევე იდენტური. ე.ი თუ სქესი და პირველი თაობის კოლეჯის სტუდენტის სტატუსი იქნება სტატისტიკურად დამოუკიდებელი, უნდა ველოდოთ სქესის იდენტურ განაწილებას თითოეული კატეგორიის მიხედვით.
4. Carl L. Bankston III,"Sibling Cooperation and Scholastic Performance Among Vietnamese American Secondary Students: An Ethnic Social Relations Theory," Sociological Perspectives 41, no. 1 (1998): 167–184.
5. იხ. ზემოთ გვ. 175
6. Paul Mazerolle, Alex Piquero, and Robert Brame, "Violent Onset Offenders: Do Initial Experiences Shape Criminal Career Dimensions?" International Criminal Justice Review 20, no. 2 (2010): 132–146.
7. იხ. ზემოთ გვ. 136

თავი 12

1. მიუხედავად იმისა, რომ ეს ფორმულა გვაძლევს საერთო წარმოდგენას შეცდომის პროპორციული შემცირების ყველა მაჩვენებელის გამოთვლისათვის, აქ მხოლოდ ლამბ-

დასთან ტოლობაა მოცემული. გამა, რომელიც ქვემოთა პარაგრაფებშია განხილული გამოითვლება სხვა ფორმულით.

2. წყვილების გამოთვლის ამგვარი გზა ასევე განხილულია წიგნში: Chava Frankfort-Nachmias and David Nachmias, Research Methods in the Social Sciences (New York: Worth, 2007), pp. 368–372.
3. 2×2 ზომის ცხრილისათვის გამას მსგავსი საზომია იულეს Q-რომელიც პირველად წარმოდგენილ იქნა სტატისტიკოს უდნი იულეს მიერ. მაგრამ მაშინ როდესაც გამა შესაძლოა გამოყენებულ იქნას ნებისმიერი ზომის ცხრილისათვის, იულეს Q მხოლოდ 2×2-ზე ზომის ცხრილისათვის გამოიყენება.
4. Kristin Luker, Abortion and the Politics of Motherhood(Berkeley: University of California Press, 1984), p. 59.
5. Daniel N. Spicer, "World View and bortion Beliefs: A Replication of Luker's Implicit Hypothesis," Sociological Inquiry 64, no. 1 (1994): 115–116 and 120–121. Copyright © 1994 by the University of Texas Press.
6. იხ. ზემოთ გვ. 115-116
7. იხ. ზემოთ გვ. 120-121
8. იხ. ზემოთ გვ. 121

თავი 13

1. მრავლობითი რეგრესიის უფრო დაწვრილებითი განხილვისათვის იხილეთ: Paul Allison's Multiple Regression: A Primer (Thousand Oaks, CA: Pine Forge Press, 1999).
2. თუ r-ს გამოთვლით r^2-დან ფესვის ამოღებით ყურადღება მიაქციეთ ნიშანს. (r^2 ყოველთვის დადებითი იქნება, მაგრამ r შეიძლება იყოს უარყოფითი).
3. Guttmacher Institute, U.S. Teenage Pregnancies, Births and Abortions: National and State Trends and Trends by Race andEthnicity, January 2010.
4. J. M. Greene and C. L. Ringwalt, "Pregnancy Among Three National Samples of Runaway Homeless Youth," Journal of Adolescent Health 23 (1998): 370–377.
5. William J. Wilson, The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy (Chicago: University of Chicago Press, 1987).
6. Stephanie Coontz, "The Welfare Discussion We ReallyNeed," Christian Science Monitor (December 29, 1994): 19.
7. ანალიზიდან ამოღებულ იქნა კოლუმბიის ოლქი, ამ შტატის მაღალი მოზარდთა ფეხმძიმობის დონის გამო.
8. ანალიზი მხოლოდ 40 და მეტი წლის ასაკის ქალებს მოიცავს.
9. Michael L. Benson, John Wooldredge, Amy B. Thistlethwaite, and Greer Litton Fox, "The Correlation Between Race and Domestic Violence Is Confounded With Community Context," Social Problems 51 no. 3 (2004): 326–342.

თავი 14

1. Sarah Crissey, Educational Attainment in the United States: 2007, Current Population Report, P20-560 (Washington, DC: U.S. Census Bureau, 2009).
2. რადგან ჩვენ მაგალითში შერჩევა მცირეა ($N = 21$), გვჭირდება დაშვებები ვარიაციის ჰომოგენურობისა და განაწილების ნორმალური ხასიათის შესახებ. ჩვენ ავირჩიეთ მცირე, რომ მოგვეხდინა F სტატისტიკის გამოთვლების ილუსტრირება, ნახსენები დაშვებებით. თუ მკლევარი არ აკეთებს ნახსენებ დაშვებებს მცირე შერჩევის შემთხვევაში, მას შეუძლია გაზარდოს შერჩევის ზომა.
3. Elirea Bornman, "Self Image and Ethnic Identification in South Africa," Journal of Social Psychology 139, no. 4 (1999): 411–425.
4. იხ. ზემოთ გვ. 412
5. იხ. ზემოთ გვ. 414
6. M. Rosenberg, Society and the Adolescent Self-Image (Princeton, NJ: Princeton University Press, 1965).
7. Bornman, pp. 415–416 Phinney's (1992) Multigroup Measure of Ethnic Identity and Bornman's own 1988 ethnic identity scale.
8. იხ. ზემოთ გვ. 417-418
9. Carol Musil, Camille Warner, Jaclene Zauszniewski, May Wykle, and Theresa Standing, "Grandmother Caregiving, Family Stress and Strain, and Depressive Symptoms," Western Journal of Nursing Research 31, no. 3 (2009): 389–408.
10. იხ. ზემოთ გვ. 395
11. იხ. ზემოთ. გვ. 391

