

ნ. ოთხოზორია, ლ. ტოკაძე, ვ. ოთხოზორია

სტატისტიკური ინფორმაციის
დამუშავების პროგრამული
უზრუნველყოფა

„ტექნიკური უნივერსიტეტი“

საქართველოს ტექნიკური უნივერსიტეტი

ნ. ოთხოზორია, ლ. ტოკაძე, ვ. ოთხოზორია

სტატისტიკური ინფორმაციის დამუშავების
პროგრამული უზრუნველყოფა



რეგისტრირებულია სტუ-ს
სარედაქციო-საგამომცემლო
საბჭოს მიერ

თბილისი
2009

დამხმარე სახელმძღვანელო მოიცავს მოკლე თეორიულ ცნობებს, სტატისტიკური ფუნქციების აღწერას და მათ პრაქტიკულ გამოყენებას. დაწვრილებით არის განხილული მონაცემთა ანალიზის მეთოდები და მათი გამოყენება Microsoft Excel-ისა და მონაცემთა ანალიზის პაკეტის Statistica-ს მიხედვით. ის დახმარებას გაუწევს როგორც სტუდენტებს სასწავლო გეგმით გათვალისწინებული პროგრამის შესრულებაში, გაანალიზებასა და მათ პრაქტიკულ გამოყენებაში, ასევე მომხმარებლებს მონაცემთა დამუშავებაში და ანალიზის სპეციფიკურ საკითხებში.

რეცენზენტი ტ.მ.კ. ო. ხუციშვილი

© საგამომცემლო სახლი „ტექნიკური უნივერსიტეტი“, 2009

ISBN 978-9941-14-606-0

<http://www.gtu.ge/publishinghouse/>



ყველა უფლება დაცულია. ამ წიგნის არც ერთი ნაწილი (იქნება ეს ტექსტი, ფოტო, ილუსტრაცია თუ სხვა) არანაირი ფორმით და საშუალებით (იქნება ეს ელექტრონული თუ მექანიკური), არ შეიძლება გამოყენებულ იქნას გამომცემლის წერილობითი ნებართვის გარეშე.

სავტორო უფლებების დარღვევა ისჯება კანონით.

შინაარსი

შესავალი	4
1. სტატისტიკური ანალიზი Microsft Excel-ის გამოყენებით	7
1.1. ძირითადი სტატისტიკური მახასიათებლები	9
1.1.1. სტატისტიკური მახასიათებლები და მათი შესაბამისი ფუნქციები Excel-ში	12
მდგომარეობის მაჩვენებლები	12
გაბნევის მაჩვენებლები	20
ასიმეტრიის მაჩვენებლები	25
განაწილების მაჩვენებლები	26
1.2. მონაცემთა ანალიზის სერვისის ჩართვა	27
1.2.1. აღწერითი სტატისტიკის რეჟიმი	28
1.2.2. "ჰისტოგრამის" რეჟიმში მუშაობის ტექნოლოგია	33
1.2.3. "ამონარჩევის" რეჟიმში მუშაობის ტექნოლოგია	37
1.2.4. რეგრესიული ანალიზი	41
1.2.5. ერთფაქტორიანი დისპერსიული ანალიზი	53
1.2.6. კორელაცია	55
1.2.7. კოვარიაცია	56
1.2.8. შემთხვევითი რიცხვების გენერირება	57
1.2.9. სტანდარტული ნორმალური განაწილება	58
2. სტატისტიკური ანალიზი მონაცემთა ანალიზის პაკეტის Statistica გამოყენებით	60
2.1. კვლევის მონაცემების პირველადი დამუშავება Basic Statistics / Tables მოდულის მხედვით	60

2.2. პროცედურა Descriptive statistics (აღწერითი სტატისტიკა)	63
2.3. რეგრესიული ანალიზი Multiple Regressions მოდულის გამოყენებით	70
2.4. პროცედურა Correlation matrices (კორელაციური მატრიცები)	81
გამოყენებული ლიტერატურა	85

შესავალი

საბაზრო რეჟიმის და მკაცრი კონკურენციის პირობებში კომერციული ფირმების, ბანკების, სადაზღვევო კომპანიების და სხვა ორგანიზაციების წარმატებული ფუნქციონირებისათვის საჭიროა პროდუქციის წარმოების, ექსპლუატაციის, ასევე კონკურენტების შესახებ ინფორმაციის ანალიზი და საიმედო და დასაბუთებული დასკვნების მიღება. ამიტომაც მონაცემთა სტატისტიკური ანალიზის საშუალებებზე მოთხოვნა საკმაოდ მაღალია. შესაბამისად სტატისტიკური პროგრამების ბაზარი მნიშვნელოვნად განვითარდა და დღეს ამ ბაზარზე წარმოდგენილია ათასზე მეტის სხვადასხვა პროგრამა, რომლებიც ერთმანეთისაგან განსხვავდებიან სამომხმარებლო ინტერფეისით, გამოყენების შესაძლებლობებით, მოწყობილობებისადმი მოთხოვნებით და ა.შ. სხვადასხვა კრიტერიუმებით. ძირითადი ნაწილი ამ პროგრამებისა შეიძლება დავყოთ ორ ჯგუფად: უნივერსალური ანუ ზოგადი დანიშნულების სტატისტიკური პაკეტები და სპეციალიზირებული პროგრამული პროდუქტები.

უნივერსალური პაკეტები არ არის ორიენტირებული სპეციფიკურ სფეროზე. ამ ტიპის პაკეტებიდან ფართოდ არის გავრცელებული: BAS, SPSS, Systat, Minilab, Statgraphics, STATISTICA.

სპეციალიზირებულ პაკეტებში, როგორც წესი ხდება კონკრეტული მეთოდების რეალიზება კონკრეტული სფეროსათვის. ყველაზე ხშირად ეს პაკეტები ორიენტირებულნი არიან დროითი

მწკრივების, კორელაციურ-რეგრესიულ, ფაქტორულ და კლასტერულ ანალიზზე. ამ პაკეტებიდან გავრცელებულია ODA, WinSTAT, Statit, STADIA.

სტატისტიკურ პაკეტებს გააჩნიათ შემდეგი შესაძლებლობები:

- მონაცემთა საკმაოდ დიდი რაოდენობის დამუშავება;
- ამოცანების აღწერა ჩაშენებული პროგრამული ენის საშუალებით;
- მათ საფუძველზე ინფორმაციის დამუშავების სისტემის შექმნა მთელი საწარმოსათვის;
- სპეციალიზირებული ანალიზის მეთოდების გამოყენება.

სტატისტიკური პაკეტის შერჩევა მონაცემთა ანალიზისათვის დამოკიდებულია გადასაწყვეტი ამოცანის ხასიათზე, დასამუშავებელი მონაცემების მოცულობასა და სპეციფიკაზე, მომხმარებლის კვალიფიკაციაზე, არსებულ აპარატურულ რესურსებზე და ა.შ.

ჩვენი მიზანია ჩავატაროთ გავრცელებული სტატისტიკური პაკეტების შესაძლებლობების ანალიზი, რათა მომხმარებელს საკუთარი კომპეტენციის ფარგლებში გაუადვილდეს საჭირო პროგრამული პაკეტის შერჩევა.

1. სტატისტიკური ანალიზი Microsoft Excel-ის გამოყენებით

ელექტრონულ ცხრილებთან დაკავშირებული პროგრამები წარმოადგენენ ყველაზე პოპულარულ და გავრცელებულ პროგრამებს ტექსტური რედაქტორების შემდეგ. ასეთი სახის პროგრამული უზრუნველყოფა საშუალებას აძლევს მომხმარებელს ერთ ანგარიშში ან ერთ სამუშაო წიგნში მოახდინოს მონაცემების, მათემატიკური ფორმულების, ტექსტების და გრაფიკების კომბინირება. ამიტომაც ელექტრონული ცხრილები გახდა არა მხოლოდ ბიზნესის წარმოების, არამედ სამეცნიერო გამოკვლევების ჩატარების შეუცვლელი ინსტრუმენტი. პროგრამა Excel, გამოყენების სიმარტივის და ფართო შესაძლებლობების გამო ერთ-ერთი პოპულარული გახდა ამ ტიპის პროგრამებს შორის.

Excel მოიცავს 81 სტატისტიკურ და 59 მათემატიკურ ფუნქციას. ამის გარდა, არსებობს ასევე მრავალი სპეციალური ფუნქცია, რომელიც ფართოდ გამოიყენება ინჟინრებისა და ბიზნესმენების მიერ. Excel-ის ბაზისური სტატისტიკური ფუნქციები განეკუთვნება აღწერით ფუნქციებს, მაგ. საშუალო, სტანდარტული გადახრა, რანგი და ა.შ. დისკრეტული და უწყვეტი განაწილების ფუნქციების აღსაწერად გათვალისწინებულია ინტეგრალური და ალბათობათა სიმკვრივის ფუნქციები.

პროგრამა Excel შეიცავს ასევე დაამტებით მოდულს Analysis ToolPak, რომელშიც შედის შემდეგი კომპონენტები:

- დისპერსიული ანალიზის ინსტრუმენტები, ერთფაქტორიანი, ორფაქტორიანი პარალელური დაკვირვებების გარეშე და პარალელური დაკვირვებების შემთხვევაში;
- კორელაციის და კოვარიაციის მატრიცის გამოსათვლელი ინსტრუმენტები;
- აღწერითი სტატისტიკის ცხრილები;
- ერთპარამეტრიანი ექსპონენციალური განაწილების ინსტრუმენტები;
- ჰისტოგრამები;
- შემთხვევითი რიცხვების გენერატორი სხვადასხვა განაწილებისათვის;
- მრავლობითი წრფივი რეგრესიის ინსტრუმენტები;
- მონაცემთა შემთხვევითი შერჩევის ინსტრუმენტები;
- t-ტესტები ერთნაირი და განსხვავებული დისპერსიების შემთხვევაში;
- z-ტესტები.

1. 1. ძირითადი სტატისტიკური მახასიათებლები

საშუალო მნიშვნელობა - შემთხვევითი სიდიდის საშუალო მნიშვნელობა წარმოადგენს მის ყველაზე ტიპიურ, ყველაზე ალბათურ მნიშვნელობას, თავისებურ ცენტრს, რომლის გარშემოც არის გაბნეული ნიშან-თვისებების ყველა მნიშვნელობა.

მედიანა - შემთხვევითი სიდიდის ის მნიშვნელობაა, რომელიც ამონარჩევის ყველა შემთხვევას ყოფს ორ ტოლ რიცხოვრივ ნაწილად;

სტანდარტული გადახრა - შემთხვევითი სიდიდის სტანდარტული გადახრა წარმოადგენს ვარიაციის ზომას. ის გვიჩვენებს, საშუალოდ რა ერთეულით ხდება გადახრა საშუალო მნიშვნელობიდან. განსაკუთრებულ მნიშვნელობას იძენს ნორმალური განაწილების კვლევის პროცესში. ნორმალური განაწილების შემთხვევაში მისი მნიშვნელობათა 68% მოთავსებულია \pm საშუალოდან ერთმაგი გადახრის ინტერვალში, 95% - \pm საშუალოდან ორმაგი გადახრის ინტერვალში, 99,7% კი სამმაგი გადახრის ინტერვალში.

დისპერსია - წარმოადგენს გაბნევის მახასიათებელს და გვიჩვენებს საშუალო მნიშვნელობიდან გადახრის კვადრატულ მნიშვნელობას. დისპერსია ეს არის მოდელის ან პროცესის ერთ-ერთი არსებითი მაჩვენებელი, ასევე ერთ-ერთი მნიშვნელოვანი კრიტერიუმი არსებითი მოდელის შესაქმნელად;

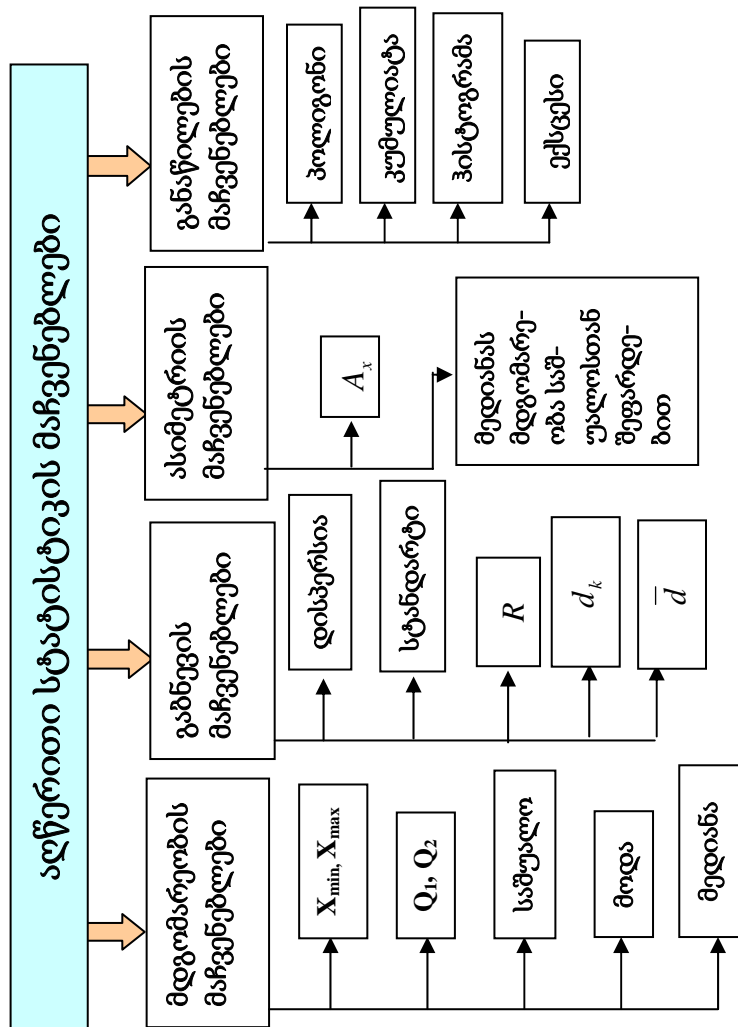
საშუალო კვადრატული გადახრა - ეს არის მნიშვნელობა რომლითაც განსხვავდება ამონარჩევის საშუალო მნიშვნელობა გენერალური ერთობლიობის

საშუალო მნიშვნელობისაგან, იმ პირობით, რომ განაწილების ტიპი უახლოვდება ნორმალურს.

გაქანება - მაქსიმალურ და მინიმალურ მნიშვნელობებს შორის სხვაობა;

ასიმეტრია - ახასიათებს ვარიაციული რიგის წანაცვლების ხარისხს საშუალო მნიშვნელობიდან გარკვეული სიდიდით და მიმართულებით. სიმეტრიული მრუდის შემთხვევაში ასიმეტრიის კოეფიციენტი ნულის ტოლია. 0,5-ზე ნაკლები მნიშვნელობა ითვლება მცირე ასიმეტრიად.

ექსცესი - ახასიათებს საშუალო მნიშვნელობის ირგვლივ მნიშვნელობათა კონცენტრაციის ხარისხს და წარმოადგენს მრუდის მახვილწვერიანობის ტავისებურ მახასიათებელს. ნორმალური განაწილების შემთხვევაში ექსცესის მნიშვნელობა ნულის ტოლია. თუ ექსცესის მნიშვნელობა დადებითია, მაშინ განაწილების წირს გააჩნია ნორმალურზე მაღალი, მახვილი წვერო, წინააღმდეგ შემთხვევაში, თუ უარყოფითია წვერო შედარებით დაბალია და ბლაგვი. ექსცესის უარყოფითი მნიშვნელობის ზღვარს წარმოადგენს -2 სიდიდე, ხოლო დადებითი მნიშვნელობის ზღვარი არ არსებობს.



სტატისტიკური მახასიათებლები და მათი შესაბამისი ფუნქციები Excel-ში

მდგომარეობის მაჩვენებლები

1. ფუნქცია *MIN*

MIN (მნიშვნელობა1, მნიშვნელობა2,...)

ფუნქცია MIN პოულობს მინიმალურ მნიშვნელობას მონაცემთა სიმრავლეში

SKEW					
=MIN(B3:B17)					
	A	B	C	D	E
1	ინდივიდუალური ბინათშენებლობის მოცულობა წარმოგენილი ფართობის ხაზით				
2	ქალაქი	ფართობი, მ ²			
3	სოხუმი	718			
4	ზუგდიდი	800			
5	ქუთაისი	1987			
6	ზესტაფონი	458			
7	მესტია	123			
8	გორი	654	მინიმალური მნიშვნელობა	45	=MIN(B3:B17)
9	ოზურგეთი	298			MIN(number1;
10	ფოთი	8123			
11	სამტრედია	345			
12	თბილისი	2098			
13	რუსთავი	1045			
14	სენაკი	321			
15	თელავი	100			
16	საგარეჯო	45			
17	სიღნაღი	590			

2. ფუნქცია მინიმალური MINA

MINA (მასივი;K)

ფუნქცია MINA მასივში პოულობს უმცირეს რიგით K მნიშვნელობას

ინდივიდუალური ბინათშენებლობების მოცულობა წარმოდგენილი ფართობის ხაზით						
ქალაქი	ფართობი, მ ²					
სოხუმი	23					
ზუგდიდი	45					
ქუთაისი	98					
ზესტაფონი	574					
მესტია	1					
გორი	23	უმცირესი მე-3 მნიშვნელობა		1	=MINA(H3:H17;3)	
ოზურგეთი	12					
ფოთი	32					
სამტრედია	54					
თბილისი	987					
რუსთავი	45					
სენაკი	32					
თელავი	45					
საგარეჯო	98					
სიღნაღი	12					

3. ფუნქცია MAX

MAX (მნიშვნელობა1, მნიშვნელობა2,...)

ფუნქცია MAX პოულობს მაქსიმალურ მნიშვნელობას მონაცემთა სიმრავლეში

ინდივიდუალური ბინათშენებლობების მოცულობა წარმოდგენილი ფართობის ხაზით					
ქალაქი	ფართობი, მ ²				
სოხუმი	718				
ზუგდიდი	800				
ქუთაისი	1987				
ზესტაფონი	458				
მესტია	123				
გორი	654	მაქსიმალური მნიშვნელობა		8123	=MAX(O3:O16)
ოზურგეთი	298				
ფოთი	8123				
სამტრედია	345				
თბილისი	2098				
რუსთავი	1045				
სენაკი	321				
თელავი	100				
საგარეჯო	45				

4. ფუნქცია მაქსიმალური MAXA
MAXA (მასივი;K)

ფუნქცია MAXA მასივში პოულობს უდიდეს რიგით K მნიშვნელობას

S		T	U	V	W	X
1	ინდივიდუალური ბინათშენებლობების მოცულობა წარმოდგენილი ფართობის ხაზით					
2	ქალაქი	ფართობი, მ ²				
3	სოხუმი	54				
4	ზუგდიდი	87				
5	ქუთაისი	98				
6	ზესტაფონი	58				
7	მესტია	12				
8	გორი	11	მაქსიმალური მე-5 მნიშვნელობა	201	=MAXA(T3:T16;5)	
9	ოზურგეთი	2				
10	ფოთი	3				
11	სამტრედიას	5				
12	თბილისი	4				
13	რუსთავი	7				
14	სენაკი	102				
15	თელავი	201				
16	საგარეჯო	3				

5. ფუნქცია კვარტილი QUARTILE
QUARTILE (მასივი;K).

ფუნქცია QUARTILE კვარტილი ითვლის დისკრეტული ვარიაციული რიგის კვარტილს.

ფუნქცია კვარტილი ამასთანავე ითვლის:

- მინიმალურ მნიშვნელობას, თუ K=0
- პირველი კვარტილს, თუ K=1
- მედიანას, თუ K=2
- მესამე კვარტილს, თუ K=3
- მაქსიმალურ მნიშვნელობას, თუ K=4

	Y	Z	AA	AB
1	მარტის თვეის განმავლობაში მოთხოვნა "Reebok"-ის სპორტულ ტანსაცმალზე			
2	ზომა	გაყიდვლის რაოდენობა	რიგით	დაგროვილი სისშირე
3	46	57	48	48
4	48	48	57	105
5	50	95	60	165
6	52	60	77	242
7	54	77	95	337
8				
9	ფორმულა	შედეგი		
10	=QUARTILE(Z3:Z7;0)	48		
11	=QUARTILE(Z3:Z7;1)	57		
12	=QUARTILE(Z3:Z7;2)	60		
13	=QUARTILE(Z3:Z7;3)	77		
14	=QUARTILE(Z3:Z7;4)	95		
15				

6. ფუნქცია *AVERAGE* - საშუალო არითმეტიკული

AVERAGE (მნიშვნელობა1, მნიშვნელობა2,...)

ფუნქცია ითვლის მეუწონელ საშუალო არითმეტიკულს მონაცემთა სიმრავლეში.

SKEW X ✓ ✘ =AVERAGE(B3:B17)					
	A	B	C	D	E
1	ინდივიდუალური ბინათშენებლობების მოცულობა წარმოდგენილი ფართობის სახით				
2	ქალაქი	ფართობი, მ ²			
3	სოხუმი	718			
4	ზუგდიდი	800			
5	ქუთაისი	1987			
6	ზესტაფონი	458			
7	მესტია	123			
8	გორი	654			
9	ოზურგეთი	298	საშუალო არითმეტიკული	1180	=AVERAGE(B3:B17)
10	ფოთი	8123			
11	სამტრედი	345			
12	თბილისი	2098			
13	რუსთავი	1045			
14	ხენაკი	321			
15	თელავი	100			
16	საგარეჯო	45			
17	სიღნაღი	590			
18					

7. ფუნქცია *AVERAGEA* საშუალო გეომეტრიული

AVERAGEA (მნიშვნელობა1, მნიშვნელობა2,...)

ფუნქცია ითვლის საშუალო გეომეტრიულს დადებით მონაცემთა მნიშვნელობებისათვის.

SKEW X ✓ ✘ =AVERAGEA(O4:O7)				
	M	N	O	P
1	რძის წარმოების დინამიკა ქართლში 2004-2008 წლებში			
2	წლები	რძის წარმოება, ათასი ტ.	ზრდის კოეფიციენტი	
3	2004	310,12		
4	2005	321,50	1,04	
5	2006	340,70	1,06	
6	2007	315,40	0,93	
7	2008	335,90	1,06	
8	საშუალო შეფარდებითი ზრდის ტემპი		1,02	=AVERAGEA(O4:O7)
9				

8. ფუნქცია *MODA*

MODA (მნიშვნელობა1, მნიშვნელობა2,...)

ფუნქცია გვხმარება მოვძებნოთ ის მნიშვნელობები, რომლებიც უფრო ხშირად გვხვდება მოცემულ ინტერვალში.

	T	U	V	W	X	Y	Z	AA
1	მთხუნა სპორტულ ტანსაცმელზე 2009 წლის იანვარში							
2	ზომა	გაყიდვის რაოდენობა						
3	46	57						
4	48	48	მოდა=	#N/A				
5	50	95	მოდა=	#N/A				
6	52	60	მოდა=	48	=MODE(T3:T7;U3:U7)			
7	54	77	მოდა=	50	=INDEX(T3:T7;MATCH(MAX(U3:U7);U3:U7))			
8								INDEX(array; row_num; [column_num])

გაზნევის მაჩვენებლები

1. ფუნქცია VAR - დისპერსია

ფუნქცია ითვლის დისპერსიას ამორჩევის ერთობლიობის მიხედვით.

ფუნქცია ითვლის დისპერსიას, იმ შემთხვევაში, როცა საწყისი მონაცემები ქმნიან ამორჩევით ერთობლიობას. თუ ერთობლიობა წარმოადგენს გენერალურს, მაშნ გამოიყენება ფუნქცია VARP.

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

	A	B	C	D	E
1	ინდივიდუალური ბინათშენებლობების მოცულობა წარმოადგენილი ფართობის ხაზით				
2	ქალაქი	ფართობი, მ ²			
3	სოხუმი	437			
4	ზუგდიდი	718			
5	ქუთაისი	730			
6	ზესტაფონი	919			
7	მესტია	1319			
8	გორი	1321			
9	ოზურგეთი	2121	დისპერსია	4795472	=VAR(B3:B19)
10	ფოთი	8872			
11	სამტრედიო	632			
12	თბილისი	218			
13	რუსთავი	206			
14	სენაკი	457			
15	თელავი	3011			
16	ზაგარეჯო	1363			
17	სიღნაღი	1389			
18	ქარელი	4728			
19	საჩხერე	3439			

9. ფუნქცია MEDIAN

MEDIAN(მნიშვნელობა1, მნიშვნელობა2,...)

ფუნქცია ითვლის მოცემული დისკრეტული ვარიაციული რიგის მედიანას.

	AA	AB	AC	AD	AE	AF	AG
1	ინდივიდუალური ბინათშენებლობების მოცულობა წარმოადგენილი ფართობის ხაზით						
2	ქალაქი	ფართობი, მ ²					
3	სოხუმი	718					
4	ზუგდიდი	800					
5	ქუთაისი	1987					
6	ზესტაფონი	458					
7	მესტია	123					
8	გორი	654	მედიანა	590,0	=MEDIAN(AB3:AB17)		
9	ოზურგეთი	298					
10	ფოთი	8123					
11	სამტრედიო	345					
12	თბილისი	2098					
13	რუსთავი	1045					
14	სენაკი	321					
15	თელავი	100					
16	ზაგარეჯო	45					
17	სიღნაღი	590					

2. ფუნქცია *VARP* დისპერსია

ფუნქცია ითვლის გენერალური ერთობლიობის აუწონელ დისპერსიას.

$$D_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

ზოგჯერ გენერალურ დისპერსიას აღნიშნავენ σ^2 .

FACT <input type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> =VARP(H3:H19)						
	G	H	I	J	K	L
1	ინდივიდუალური ბინათშენებლობების მოცულობა წარმოდგენილი ფართობის ხაზით					
2	ქალაქი	ფართობი, მ ²				
3	სოხუმი	437				
4	ზუგდიდი	718				
5	ქუთაისი	730				
6	ზესტაფონი	919				
7	მესტია	1319				
8	გორი	1321				
9	ოზურგეთი	2121				
10	ფოთი	8872				
11	სამშრედია	632	დისპერსია	4513386	=VARP(H3:H19)	
12	თბილისი	218				
13	რუსთავი	206				
14	სენაკი	457				
15	თელავი	3011				
16	საგარეჯო	1363				
17	სიღნაღი	1389				
18	ყვარელი	4728				
19	ბოლნისი	3439				

3. ფუნქცია *STDEV* - საშუალო კვადრატული გადახრა

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

FACT <input type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> =STDEV(N3:N10)							
	M	N	O	P	Q	R	S
1	ინდივიდუალური ბინათშენებლობების მოცულობა წარმოდგენილი ფართობის ხაზით						
2	ქალაქი	ფართობი, მ ²					
3	სოხუმი	437					
4	ზუგდიდი	718					
5	ქუთაისი	730					
6	ზესტაფონი	919	სტანდარტული გადახრა			2803	=STDEV(N3:N10)
7	მესტია	1319					
8	გორი	1321					
9	ოზურგეთი	2121					
10	ფოთი	8872					
11							

4. ფუნქცია *STDVP* - საშუალო კვადრატული გადახრა

ფუნქცია ითვლის საშუალო კვადრატულ გადახრას.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

FACT		=STDEVP(U3:U19)						
	T	U	V	W	X	Y	Z	
1	ინდივიდუალური ბინათშენებლობების მოცულობა წარმოდგენილი ფართობის სახით							
2	ქალაქი	ფართობი, მ ²						
3	სოხუმი	437						
4	ზეგდიდი	718						
5	ქუთაისი	730						
6	ზესტაფონი	919						
7	მესტია	1319						
8	გორი	1321	სტანდარტული გადახრა			2124	=STDEVP(U3:U19)	
9	ოზურგეთი	2121						
10	ფოთი	8872						
11	სამტრედია	632						
12	თბილისი	218						
13	რუსთავი	206						
14	სენაკი	457						
15	თელავი	3011						
16	საგარეჯო	1363						
17	სიღნაღი	1389						
18	ყვარელი	4728						
19	ბოლბუ	3439						

5. ფუნქცია *AVEDEV* - საშუალო წრფივი გადახრა

$$\bar{d} = \pm \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

FACT		=AVEDEV(AD3:AD10)					
	AC	AD	AE	AF	AG	AH	AI
1	ინდივიდუალური ბინათშენებლობების მოცულობა წარმოდგენილი ფართობის სახით						
2	ქალაქი	ფართობი, მ ²					
3	სოხუმი	437					
4	ზეგდიდი	718					
5	ქუთაისი	730					
6	ზესტაფონი	919	საშუალო წრფივი გადახრა			1721	=AVEDEV(AD3:AD10)
7	მესტია	1319					
8	გორი	1321					
9	ოზურგეთი	2121					
10	ფოთი	8872					

ასიმეტრიის მაჩვენებლები

1. ფუნქცია *SKEW* - ასიმეტრიულობა

SKEW (მნიშვნელობა₁, მნიშვნელობა₂,...)

ფუნქცია ითვლის ასიმეტრიულობის კოეფიციენტს ამონარჩევში.

$$A_x = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$$

თუ მონაცემები ქმნიან არა ამორჩევით, არამედ გენერალურ ერთობლიობას, მაშინ ასიმეტრიულობა უნდა გამოითვალოს სტანდარტული ფუნქციის მიხედვით.

$$A_x = \frac{\mu_3}{\sigma^3}$$

FACT X ✓ ✖ =SKEW(B3:B10)						
	A	B	C	D	E	F
1	ინდივიდუალური ბინამშენებლების მოცულობა წარმოდგენილი ფართობის სახით					
2	ქალაქი	ფართობი, მ ²				
3	სოხუმი	437				
4	ზუგდიდი	718				
5	ქუთაისი	730	ასიმეტრიულობა	2,64	=SKEW(B3:B10)	
6	ზესტაფონი	919				
7	მესტია	1319				
8	გორი	1321				
9	ოზურგეთი	2121				
10	ფოთი	8872				

განაწილების მაჩვენებლები

1. ფუნქცია *KURT* - ექსცესი

KURT (მნიშვნელობა₁, მნიშვნელობა₂,...)

ფუნქცია ითვლის ამონარჩევის ექსცესის კოეფიციენტს.

$$E_x = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \right]$$

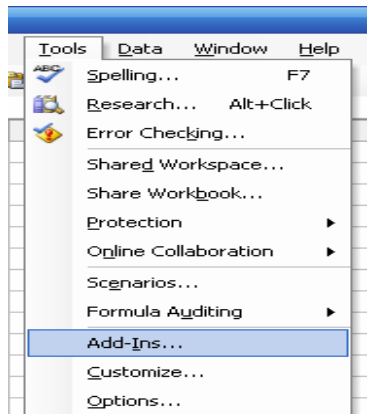
თუ მონაცემები ქმნიან არა ამორჩევით, არამედ გენერალურ ერთობლიობას, მაშინ ექსცესი უნდა გამოითვალოს სტანდარტული ფუნქციის მიხედვით.

$$E_x = \frac{\mu_4}{\sigma^4} - 3$$

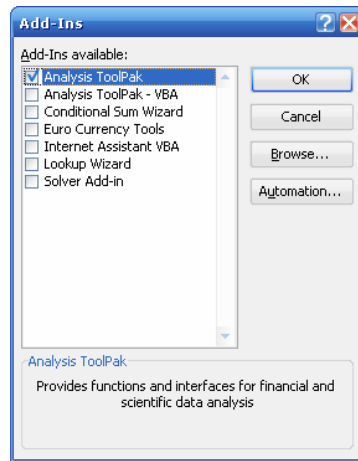
FACT X ✓ ✖ =KURT(I3:I10)						
	H	I	J	K	L	M
1	ინდივიდუალური ბინამშენებლების მოცულობა წარმოდგენილი ფართობის სახით					
2	ქალაქი	ფართობი, მ ²				
3	სოხუმი	437				
4	ზუგდიდი	718				
5	ქუთაისი	730				
6	ზესტაფონი	919	ექსცესი	7,18	=KURT(I3:I10)	
7	მესტია	1319				
8	გორი	1321				
9	ოზურგეთი	2121				
10	ფოთი	8872				

1.2. მონაცემთა ანალიზის სერვისის ჩართვა

მონაცემთა ანალიზის გასააქტიურებლად შესრულებულ შემდეგი ბრძანებების თანმიმდევრობა Tools – Add-Ins...(ნახ.2.1) , Add-Ins დასახელების ფანჯარაში მონიშნეთ Analysis ToolPak (ნახ.2.2.)

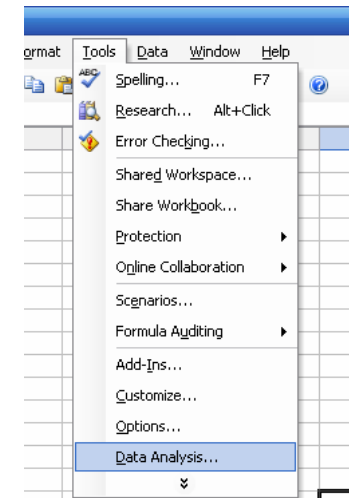


ნახ. 2.1



ნახ. 2.2

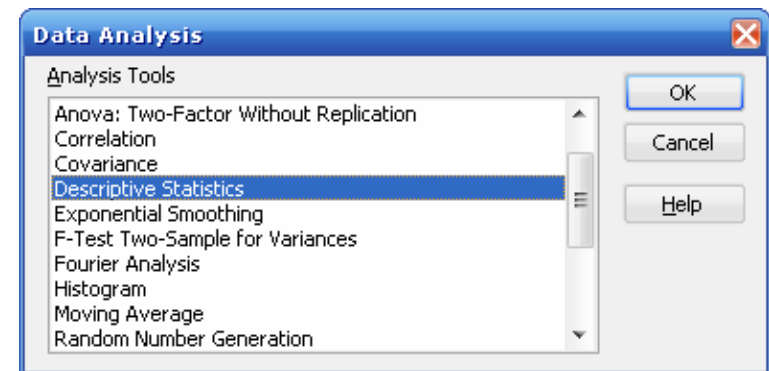
Tools მენიუს დაემატება ჩანართი Data Analysis (ნახ.2.3), რომლის გამოძახების შემდეგ ეკრანზე გამოდის მონაცემთა ანალიზის სხვადასხვა პაკეტები.



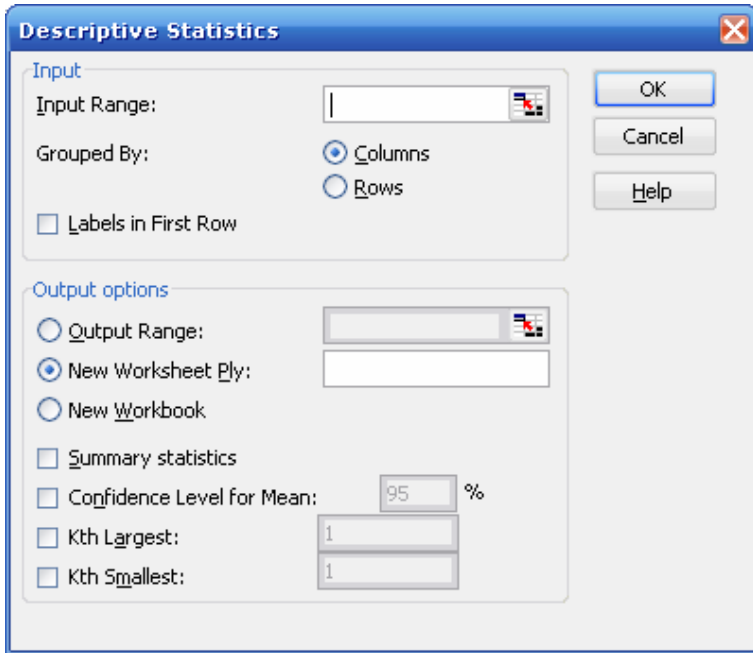
ნახ. 2.3

1.2.1. აღწერითი სტატისტიკა

მონაცემთა ანალიზის პაკეტიდან შეირჩიეთ Descriptive Statistics (აღწერითი სტატისტიკა) (ნახ.2.1.1).



ნახ. 2.1.1



ნახ. 2.1.2

ეკრანზე გამოჩნდება ფანჯარა (ნახ. 2.1.2) შემდეგი პარამეტრებით :

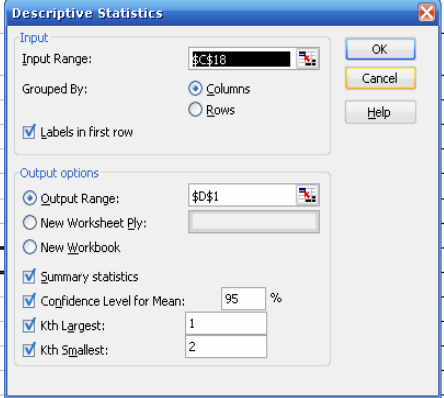
1. ველი Input Range – მონაცემების მისამართი, შესაძლებელია მონაცემები დაჯგუფებული იყოს რამდენიმე სვეტად ან რამდენიმე სტრიქონად, შესაბამისა აღწერითი სტატისტიკის მაჩვენებლები გამოთვლილი იქნება თითოეული მწკრივისათვის ერთ ცხრილში ცალკეული სვეტების სახით.
2. გადამრთველი Grouped by: Columns
Rows;
მოინიშნება იმის შესაბამისად, თუ რა სახით არის განლაგებული გასაანალიზებელი მონაცემები

სვეტების (Columns) თუ სტრიქონების სახით (Rows);

3. Labels in First Rows - მოინიშნება იმ შემთხვევაში, თუ მონაცემთა პირველი სტრიქონი შეიცავს სათაურს;
4. Output Options - გამოსასვლელი ინტერვალი (გამოსასვლელი სიდიდე) / New Worksheet Ply - ახალი სამუშაო ფურცელი/ New Workbook - ახალი სამუშაო წიგნი - მიუთითეთ მისამართი მონიშნულის გასწვრივ.
5. Summary Statistics – ამ ველის მონიშვნის შემთხვევაში გამოითვლება აღწერითი სტატისტიკის ყველა ძირითადი მაჩვენებელი;
6. Confidence Level for Mean - ნდობის ალბათობის ველი მოინიშნება იმ შემთხვევაში, თუ მისი მნიშვნელობა დასამუშავებელი მონაცემები–სათვის განსხვავდება ავტომატურად დაფიქსირებული (95%) მნიშვნელობისაგან;
7. Kth Largest, Kth Smallest - ეს ველი გააქტიურდება იმ შემთხვევაში თუ შემაჯამებელ ცხრილში აუცილებელია K-ური მინიმალური ან მაქსიმალური ელემენტის ჩვენება;

ველებში შესაბამისი მონაცემების დაფიქსირების შემდეგ (ნახ. 2.1.3) დააჭირეთ Ok ღილაკს, ღილაკზე დაჭერის შემდეგ სამუშაო ფურცელზე გამოჩნდება მითითებული მონაცემების მიხედვით სტატისტიკური ანალიზის შედეგები (ნახ. 2.1.4)

1	ინდივიდუალური ბინათმშენებლობების მოცულობა წარმოდგენილი ფართობის სახით	
2	ქალაქი	ფართობი, მ2
3	სოსუმი	718
4	ზუგდიდი	800
5	ქუთაისი	1987
6	ზესტაფონი	458
7	მესტია	123
8	გორი	654
9	ოზურგეთი	298
10	ფოთი	8123
11	სამტრედი	345
12	თბილისი	2098
13	რუსთავი	1045
14	სენაკი	321
15	თელავი	100
16	საგარეჯო	45
17	სიღნაღი	590
18		



ნახ. 2.1.3

	A	B	C	D	E
1	ინდივიდუალური ბინათმშენებლობების მოცულობა წარმოდგენილი ფართობის სახით			ფართობი, მ2	
2	ქალაქი	ფართობი, მ2			
3	სოსუმი	718		Mean	1213.357143
4	ზუგდიდი	800		Standard Error	558.5363192
5	ქუთაისი	1987		Median	524
6	ზესტაფონი	458		Standard Deviation	2089.851544
7	მესტია	123		Sample Variance	4367479.478
8	გორი	654		Kurtosis	10.83617863
9	ოზურგეთი	298		Skewness	3.782891297
10	ფოთი	8123		Range	8078
11	სამტრედი	345		Minimum	45
12	თბილისი	2098		Maximum	8123
13	რუსთავი	1045		Sum	16987
14	სენაკი	321		Count	14
15	თელავი	100		Largest(1)	8123
16	საგარეჯო	45		Smallest(2)	100
17	სიღნაღი	590		Confidence Level(95.0%)	1206.644355
18					

ნახ. 2.1.4

შემაჯამებელ ცხრილში მოყვანილია შემდეგი მაჩვენებლები :

- Mean - საშუალო არითმეტიკული
- Standard Error - სტანდარტული ცდომილება - საშუალო კვადრატული გადახრა μ
- Median - მედიანა
- Mode - მოდა
- Standard Deviation - სტანდარტული გადახრა σ_x
- Sample Variance - ამონარჩევის დისპერსია
- Kurtosis - ექსცესი
- Skewness - ასიმეტრია
- Range - ინტერვალი - ვარიაციის გაქანება - R
- Minimum - მინიმუმი
- Maximum - მაქსიმუმი
- Sum - ჯამი
- Count - რაოდენობა
- Largest(2) - სიდიდით მეორე მინიმალური ელემენტი
- Confidence Level(95.0%) - ნდობის ალბათობა

უნდა აღინიშნოს აღწერითი სტატისტიკის მაჩვენებლების გამოთვლის შემდეგი თავისებურებები:

1. სტანდარტული გადახრა, ამონარჩევის დისპერსია, ექსცესი, ასიმეტრიულობა - გამოითვლება გენერალური ერთობლიობისათვის და არა ამონარჩევისათვის;
2. არ არის საჭირო მონაცემების წინასწარი რანჟირება. გამოთვლების მიმდინარეობისა ეს პროცესი ავტომატურად ხორციელდება;
3. მოდის მაჩვენებლის გასწვრივ შეცდომის ინდიკატორის გამოჩენა მიუთითებს, რომ საწყის

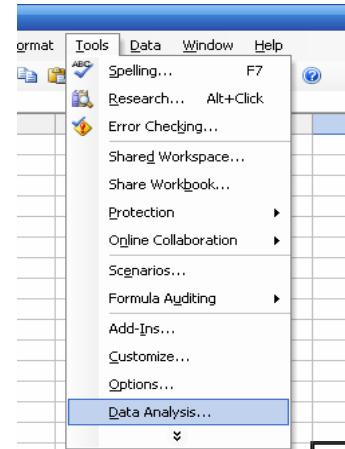
მონაცემებში არ გვხვდება ერთნაირი მნიშვნელობის ელემენტები. ამ შემთხვევაში მოდის მნიშვნელობად მიიღება ის მნიშვნელობა, რომელიც შეესაბამება მაქსიმალურ ორდინატას განაწილების მრუდზე;

4. ექსცესის ან ასიმეტრიის გასწვრივ შეცდომის ინდიკატორის გამოჩენა მიუთითებს, რომ შემაჯამებელ ცხრილში სტანდარტული გადახრის მნიშვნელობა 0-ის ტოლია ან შესასვლელი პარამეტრები მოიცავს 4-ზე ნაკლებ ელემენტს.

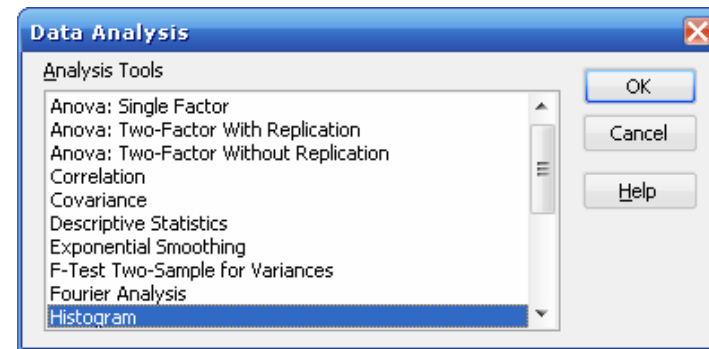
1.2.2. "ჰისტოგრამის" რეჟიმში მუშაობის ტექნოლოგია (საცნობარო ინფორმაცია)

"ჰისტოგრამის" რეჟიმის საშუალებით შეგვიძლია დავადგინოთ მონაცემების მოხვედრის სიხშირე მითითებულ ინტერვალში, ასევე შესაძლებელია განაწილების ინტერვალური ვარიაციული რიგის ჰისტოგრამის აგება.

Histogram ბრძანების არჩევის შემდეგ ეკრანზე ჩნდება დიალოგური ფანჯარა, რომელშიც უნდა მივუთითოთ დასამუშავებელი მონაცემების მისამართი, ასევე გამოსასვლელი პარამეტრების რეჟიმები:



ნახ. 2.1.5



ნახ. 2.1.6

დიალოგურ ფანჯარაში შეავსეთ აუცილებელი ველები:

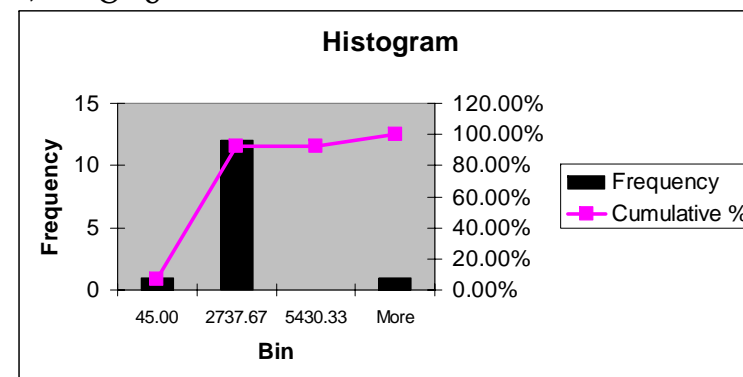
- Input Range – მონაცემების მისამართი;
- Bin Range - მონიშნეთ ის მისამართი, სადაც წინასწარ მითითებული გაქვთ ინტერვალის საზღვრები. თუ ველს არ შეავსებთ, ავტომატურად შეირჩევა ინტერვალის სიგრძე;

Labels - მონიშნება იმ შემთხვევაში, თუ მონაცემთა პირველი სტრიქონი შეიცავს სათაურს;

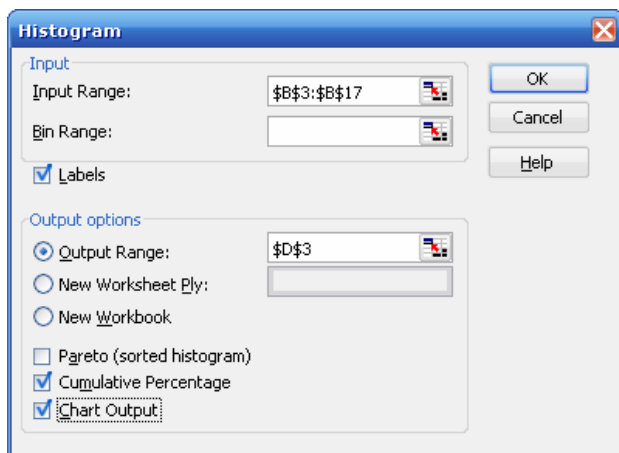
Output Options - გამოსასვლელი ინტერვალი (გამოსასვლელი სიდიდე) / New Worksheet Ply - ახალი სამუშაო ფურცელი/ New Workbook - ახალი სამუშაო წიგნი - მიუთითეთ მისამართი მონიშნულის გასწვრივ.

<i>Bin</i>	<i>Frequency</i>	<i>Cumulative %</i>
45.00	1	7.14%
2737.67	12	92.86%
5430.33	0	92.86%
More	1	100.00%

ბ) ჰისტოგრამა



ნახ. 2.1.8



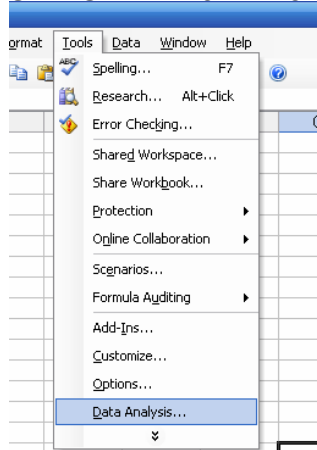
ნახ. 2.1.7

მიიღება შემდეგი მონაცემები:

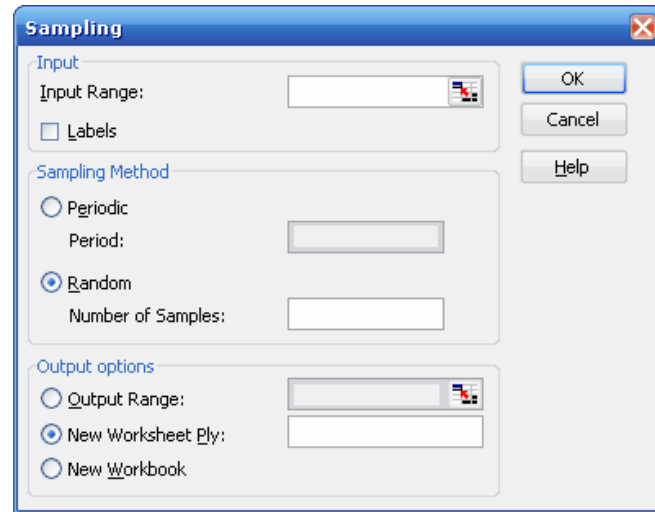
ა) ცხრილი, რომელშიც მოცემულია ინტერვალებში მონაცემების მოხვედრის სიხშირე (Frequency) და მონაცემების ინტერვალებში მოხვედრის სიხშირის ალბათობა (Cumulative %)

1.2.3. "ამონარჩევის" რეჟიმში მუშაობის ტექნოლოგია (საცნობარო ინფორმაცია)

რეჟიმი- "ამონარჩევი" გამოიყენება გენერალური ერთობლიობიდან შემთხვევითი შერჩევის გზით ამონარჩევის ჯგუფის ფორმირებასთვის.

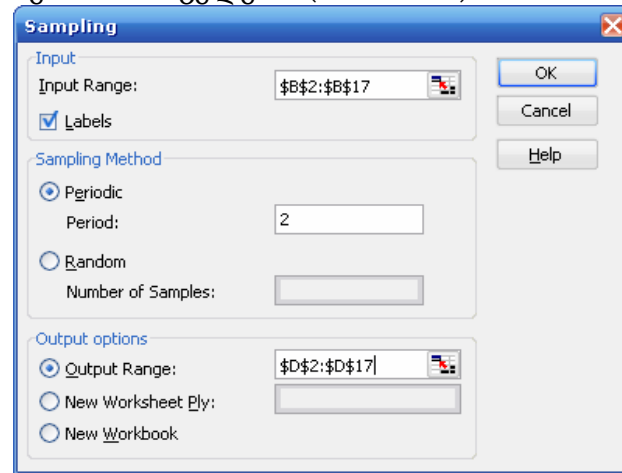


ნახ. 2.1.9

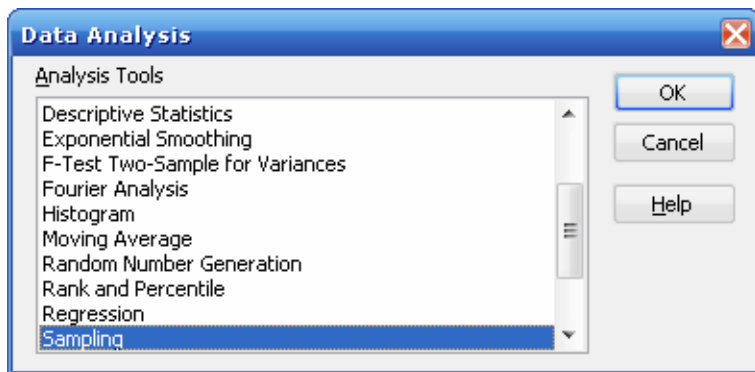


ნახ. 2.1.11

გამოსულ დიალოგურ ფანჯარაში შევავსოთ შესაბამისი ველები (ნახ. 2.1.11)



ნახ. 2.1.12



ნახ. 2.1.10

მიიღება შემდეგი მონაცემები:

	A	B	C	D
	ინდივიდუალური ბინათმშენებლობების მოცულობა წარმოადგენილი ფართობის სახით			
1				
2	ქალაქი	ფართობი, მ2		800
3	სოსუმი	718		458
4	ზუგდიდი	800		654
5	ქუთაისი	1987		8123
6	ზეესტაფონი	458		2098
7	მესტია	123		321
8	გორი	654		45
9	ოზურგეთი	298		
10	ფოთი	8123		
11	სამტრედიი	345		
12	თბილისი	2098		
13	რუსთავი	1045		
14	სენაკი	321		
15	თელავი	100		
16	საგარეჯო	45		
17	სიღნაღი	590		

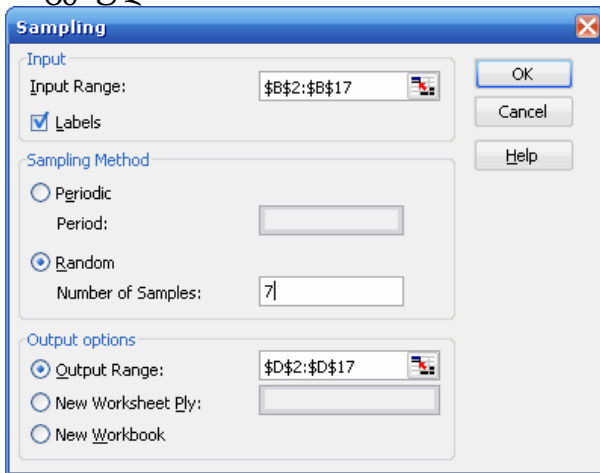
მიღებული შედეგები მოცემულია ნახ. 2.1.15-ზე.

	A	B	C	D	E
	ინდივიდუალური ბინათმშენებლობების მოცულობა წარმოადგენილი ფართობის სახით				
1					
2	ქალაქი	ფართობი, მ2			1987
3	სოსუმი	718			123
4	ზუგდიდი	800			123
5	ქუთაისი	1987			345
6	ზეესტაფონი	458			1987
7	მესტია	123			8123
8	გორი	654			654
9	ოზურგეთი	298			
10	ფოთი	8123			
11	სამტრედიი	345			
12	თბილისი	2098			
13	რუსთავი	1045			
14	სენაკი	321			
15	თელავი	100			
16	საგარეჯო	45			
17	სიღნაღი	590			

ნახ. 2.1.15

ნახ. 2.1.13

გამოვიძახოთ ისევ რეჟიმი "ამონარჩევი" (Sampling) და დიალოგური ფანჯრის ველები შევავსოთ ნახ. 2.1.14-ზე მოცემული სახით



ნახ. 2.1.14

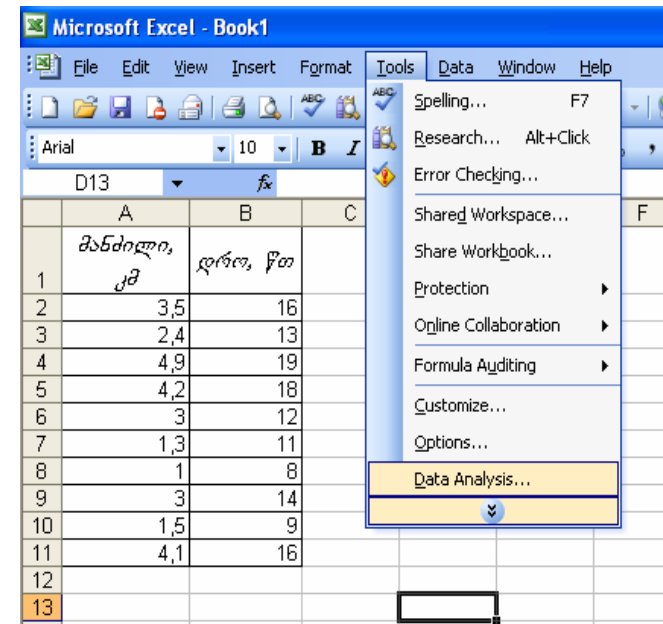
1.2.4. რეგრესიული ანალიზი მეთოდური რეკომენდაციები

რეგრესიული ანალიზის ჩასატარებლად და პროგნოზირებისათვის აუცილებელია:

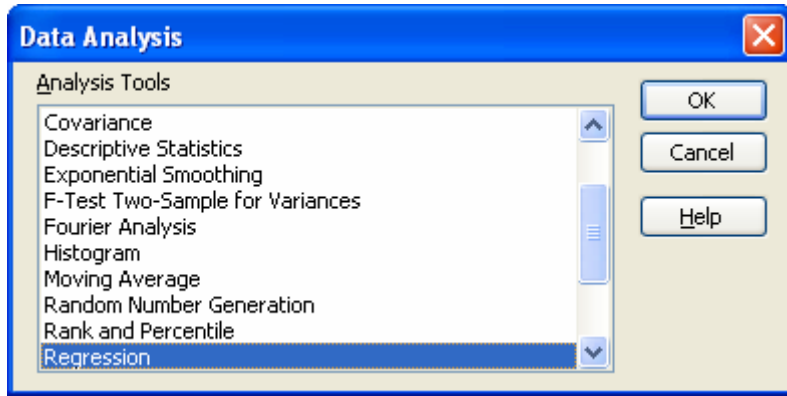
1. ავადოთ გრაფიკი საწყისი მონაცემების მიხედვით და დაახლოებით განვსაზღვროთ დამოკიდებულების ხასიათი;
2. შევირჩიოთ რეგრესიის ფუნქციის სახე, რომელიც შესაძლებელია აღწერდეს საწყის მონაცემებს შორის კავშირს;
3. განვსაზღვროთ რეგრესიის ფუნქციის რიცხვითი კოეფიციენტები უმცირეს კვადრატთა მეთოდით;
4. შეაფასეთ მოძებნილი რეგრესიული დამოკიდებულების კავშირის სიმლიერე დეტერმინაციის კოეფიციენტის საფუძველზე;
5. გააკეთეთ პროგნოზი (როცა $R^2 \geq 75\%$) ან გამოიტანეთ დასკვნა მიღებული რეგრესიული დამოკიდებულების მიხედვით პროგნოზირების შეუძლებლობის შესახებ

MS Excel-ის ანალიზთა პაკეტის მიხედვით რეგრესიული ანალიზის ჩატარების შესახებ საცნობარო ინფორმაცია

1. Tools მენიუში ვირჩევთ ბრძანებას Data Analysis - regression. (ნახ. 3.1 ნახ. 3.2)
2. გამოსულ დიალოგურ ფანჯარაში ვირჩევთ

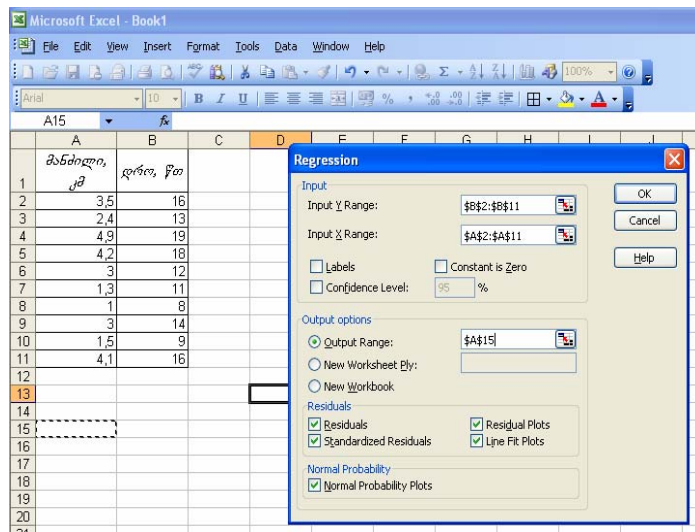


ნახ. 3.1



ნახ. 3.2

გამოსულ დიალოგურ ფანჯარაში (ნახ. 3.3) შევირჩიოთ შემდეგი პარამეტრები:



ნახ. 3.3

1. Input Y Range - შესავალი სიდიდის Y-ის მნიშვნელობები;
2. Input X Range - შესავალი სიდიდის X-ის მნიშვნელობები;
3. Labels - ჩანართის მონიშვნა საჭიროა იმ შემთხვევაში, თუ პირველ სტრიქონში მოცემულია სათაური;
4. Confidence Levels - ნდობის ალბათობის ველის მონიშვნა საჭიროა იმ შემთხვევაში, თუ აუცილებელია მითითებული ნდობის ალბათობის მნიშვნელობის შეცვლა;
5. Constant is Zero - მონიშვნა აუცილებელია თუ რეგრესიის ხაზმა უნდა გაიაროს კოორდინატთა სისტემის დასაწყისში ($a_0=0$).
6. Output Options - გამოსასვლელი ინტერვალი (გამოსასვლელი სიდიდე) / New Worksheet Ply - ახალი სამუშაო ფურცელი/ New Workbook - ახალი სამუშაო წიგნი - მიუთითეთ მისამართი მონიშნულის გასწვრივ.
7. ჯგუფში *Residuals* ალმის ჩართვის შემთხვევაში შედეგების ცხრილში გამოჩნდება სხვაობების შესაბამისი მნიშვნელობები და ასევე მივიღებთ შესაბამის გრაფიკულ გამოსახულებას;
8. Normal Probability - ნორმალური განაწილების ალბათური გრაფიკის ველის გააქტიურების შემთხვევაში Y სიდიდის ავტომატურად ფორმირებული ინტერვალების დამოკიდებულების წერტილოვან გრაფიკს.

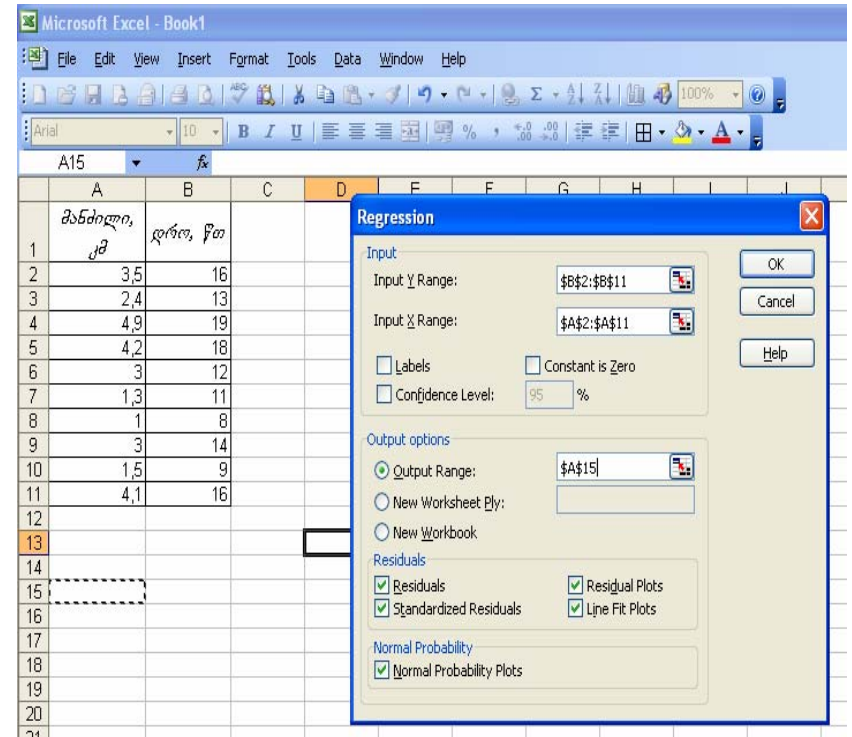
ამოცანა: ფირმა ახორციელებს სხვადასხვა ტიპის ტვირთის გადატანას ქალაქის შიგნით. შეაფასეთ ასეთი მომსახურების ღირებულება დახარჯული დროზე დამოკიდებულებით. მნიშვნელოვან ფაქტორად, რომელიც გავლენას ახდენს მიწოდების დროზე არის გავლილი მანძილი.

მოცემულია საწყისი მონაცემები:

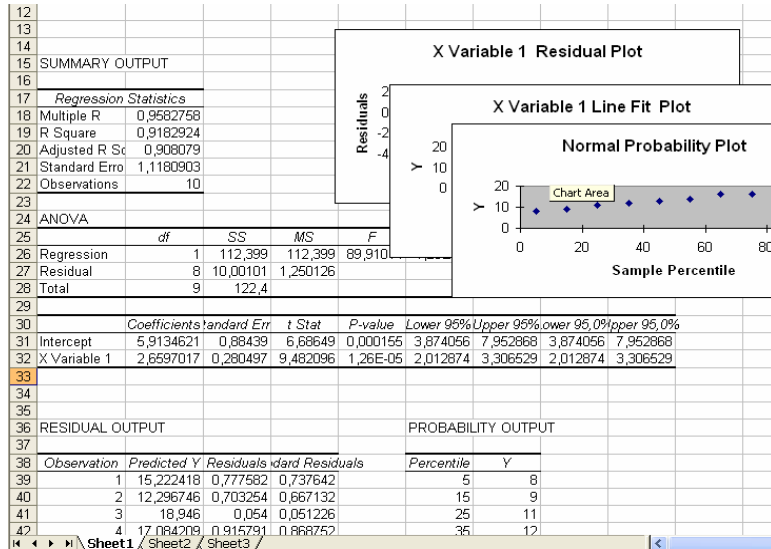
მანძილი, კმ	3,5	2,4	4,9	4,2	3,0	1,3	1,0	3,0	1,5	4,1
დრო, წთ	16	13	19	18	12	11	8	14	9	16

დაადგინეთ მანძილსა და დახარჯულ დროს შორის დამოკიდებულების ხასიათი, გაანალიზეთ უმცირეს კვადრატთა მეთოდის გამოყენების არსი, ააგეთ რეგრესიის განტოლება, გაანალიზეთ რეგრესიულიკავშირის სიძლიერე. გამოიანგარიშეთ რეგრესიული მოდელის ცდომილება და ააგეთ შესაბამისი გრაფიკი

ზემოთაღწერილი მიმდევრობის მიხედვით ჩავატაროთ რეგრესიული ანალიზი MS Excel-ის რეგრესიის რეჟიმის გამოყენებით (ნახ. 3.4, ნახ. 3.5).



ნახ. 3.4



ნახ. 3.5

გავაანალიზოთ მიღებული მნიშვნელობები:

ცხრილი 3.1

SUMMARY OUTPUT	
<i>Regression Statistics</i>	
Multiple R	0,958
R Square	0,918
Adjusted R Square	0,908
Standard Error	1,118
Observations	10

სიდიდე R^2 -კვადრატი (R Square), რომელსაც ასევე უწოდებენ განსაზღვრულობის ზომას, ახასიათებს მიღებული რეგრესიის წრფის ადეკვატურობას. ადეკვატურობა გამოისახება საწყისი მონაცემებისა და რეგრესიული მოდელის (განტოლებით ნაწინასწარ-მეტყველავი მნიშვნელობები) შესაბამისობის ხარისხით. განსაზღვრულობის ზომა გამოისახება მნიშვნელობებით ინტერვალში $[0;1]$. განხილულ ამოცანაში მისი მნიშვნელობა ტოლია 0,918-ის, რაც მიგვითითებს რეგრესიის წრფის საუკეთესო მიახლოებაზე საწყის მონაცემებთან. ე.ი. წრფივი მოდელით განისაზღვრა მიწოდების დროის ვარიაცია 91,8%, რაც ნიშნავს, რომ ფაქტორი (მანძილი) შერჩეულია სწორად. მოდელი ვერ ხსნის მოძრაობის დროის ვარიაციის $100\% - 91,8\% = 8,2\%$ -ს, რომელიც განპირობებულია სხვადასხვა ფაქტორებით, რომელიც გავლენას ახდენს მიწოდების დროზე, მაგრამ არ არის განხილული მოცემულ მოდელში.

მნიშვნელოვნობის საანგარიშო დონე $\alpha_p = 1,26E-05 < 0,05$ (F-კრიტერიუმის მნიშვნელობა) ამტკიცებს R^2 არსებობას.

მრავლობითი R (Multiple R) - მრავლობითი კორელაციის კოეფიციენტი R - განსაზღვრავს დამოუკიდებელი (X) ცვლადის და დამოკიდებული (Y) ცვლადის დამოკიდებულებას და ტოლია დეტერმინაციის კოეფიციენტიდან კვადრატული ფესვის მნიშვნელობის, ეს სიდიდე ღებულობს მნიშვნელობებს 0-დან ერთამდე. მარტივი წრფივი რეგრესიული ანალიზის დროს მისი მნიშვნელობა (R) ტოლია

პირსონის კორელაციის კოეფიციენტის მნიშვნელობის, რომელიც გამოითვლება ფორმულით:

$$r_{xy} = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] \cdot [n \sum y^2 - (\sum y)^2]}}$$

ცხრილში 3.2. მოცემულია რეგრესიის კოეფიციენტები a_1 (2,65970168) და ორდინატა ღერძის მიმართ წანაცვლება a_0 (5,913462144).

ცხრილი 3.2

	Coefficients კოეფიციენტები	Standard Error სტანდარტული ცდომილება	t Stat t-სტატისტიკა	P-value P- მნიშვნელობა
Y	5,913462144	0,884389599	6,686489927	0,00015485
X 1	2,65970168	0,280497238	9,482095791	1,26072E-05

გამოთვლებიდან გამომდინარე, შესაძლებელია განტოლების ჩაწერა შემდეგი სახით:

$$y^d = 5,91346 + 2,65970x.$$

შევამოწმოთ a_0 და a_1 რეგრესიის კოეფიციენტების არსებითობა. როგორც ცხრილიდან ჩანს, კოეფიციენტების აბსოლუტური მნიშვნელობა მეტია მათ სტანდარტულ ცდომილებაზე. ასევე ამ კოეფიციენტების

არსებითობაზე მიგვანიშნებს P-მნიშვნელობები, რომელიც მცირეა მნიშვნელოვნობის საწყის დონესთან ($\alpha=0,05$) შედარებით.

ცხრილი 3.3.

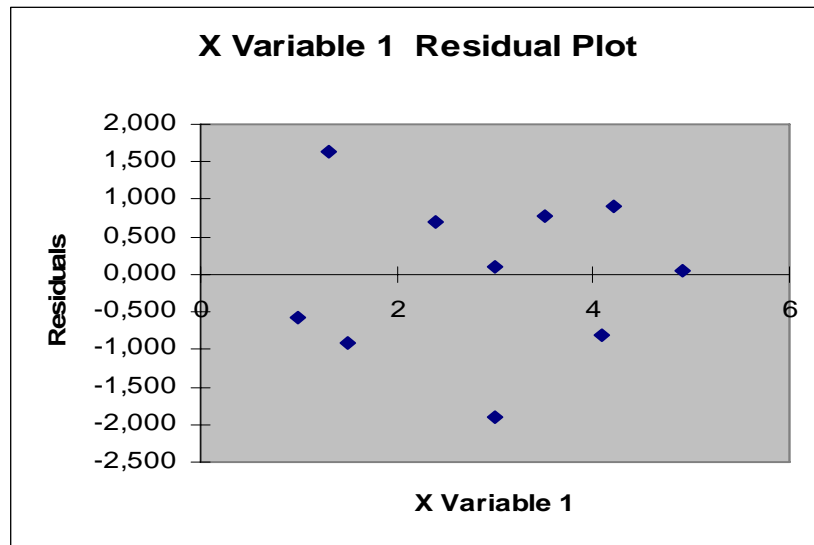
Observation დაკვირვება	Predicted Y წინასწარმეტყველი Y	Residuals სხვაობები	Standard Residuals სტანდარტული სხვაობები
1	15,22241803	0,777581975	0,737641894
2	12,29674618	0,703253823	0,667131568
3	18,94600038	0,053999622	0,051225961
4	17,0842092	0,915790799	0,868751695
5	13,89256718	-1,892567185	-1,795356486
6	9,371074328	1,628925672	1,545256778
7	8,573163824	-0,573163824	-0,543723571
8	13,89256718	0,107432815	0,101914586
9	9,903014664	-0,903014664	-0,8566318
10	16,81823903	-0,818239033	-0,776210624

ცხრილში 3.3 მოცემული მონაცემებიდან ჩანს, რომ სხვაობის უდიდესი აბსოლუტური მნიშვნელობა ჩვენს შემთხვევაში ტოლია - 1,89256, უმცირესი - 0,05399. ამ მონაცემების ინტერპრეტაციისათვის აგებულია რეგრესიის წრფე.

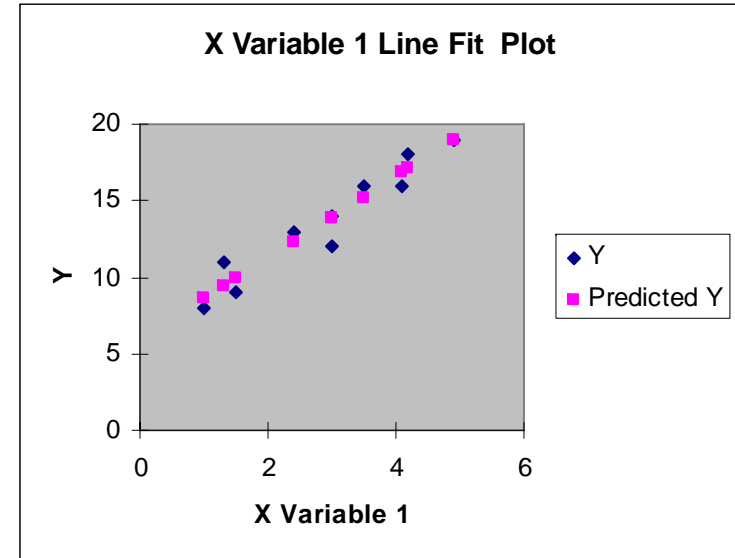
მიახლოებითი, მაგრამ საკმაოდ მარტივი და თვალსაჩინო საშუალებაა გადახრების გრაფიკული წარმოდგენა რეგრესიული მოდელის ადეკვატურობის შესამოწმებლად

Y ღერძზე გადავზომოთ $(y_i^p - y_i)$ გადახრები, ყოველი y_i მნიშვნელობისათვის (ნახ.449). თუ რეგრესიული მოდელის მიახლოებულა რეალურ დამოკიდებულებასთან, მაშინ გადახრას ექნება შემთხვევითი ხასითი და მათი ჯამი მიუახლოვდება 0-ს.

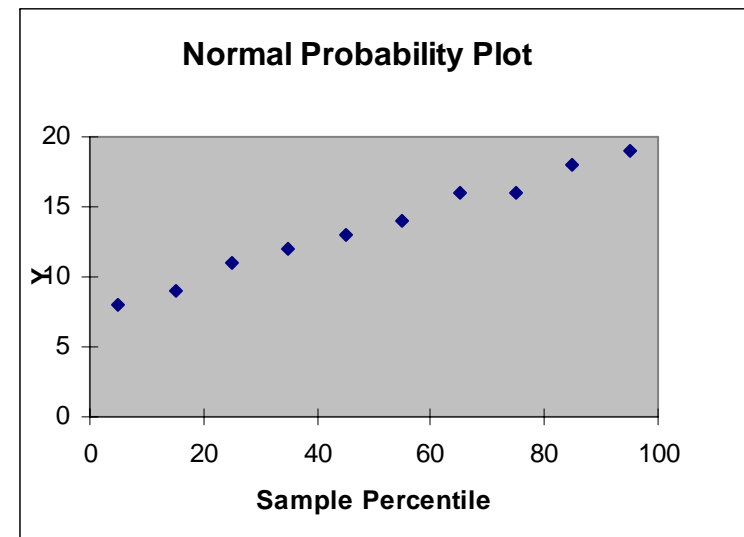
განხილულ მაგალითში $\sum_{i=1}^n (y_i^p - y_i) = 0,004$



ნახ. 3.6

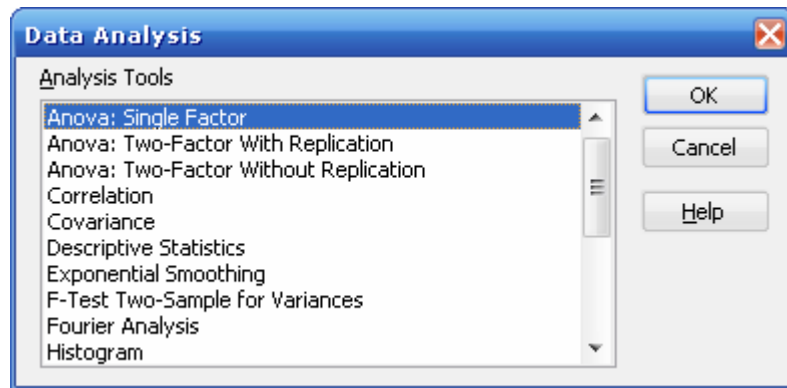


რეგრესიის ხაზი
ნახ. 3.5



ნახ. 3.6

1.2.5. ერთფაქტორიანი დისპერსიული ანალიზი



	A	B	C	D	E	F	G	H	I
1									
2									
3									
4									
5				a1	a2	a3	a4		
6		1	2	3	4	5			
7		2	6	7	5	9			
8		3	7	11	5	3			
9		4	5	6	7	8			
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									

Anova: Single Factor

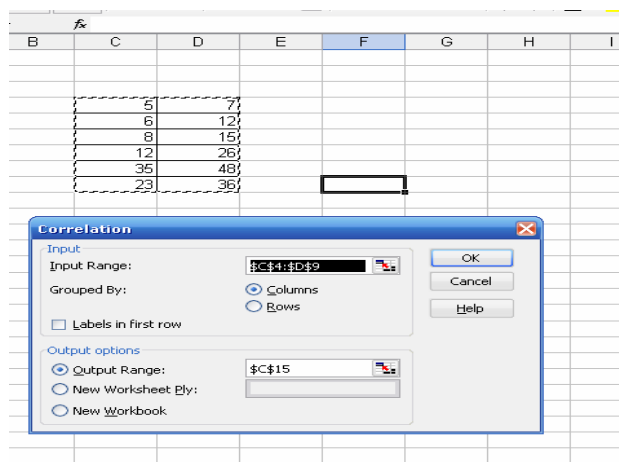
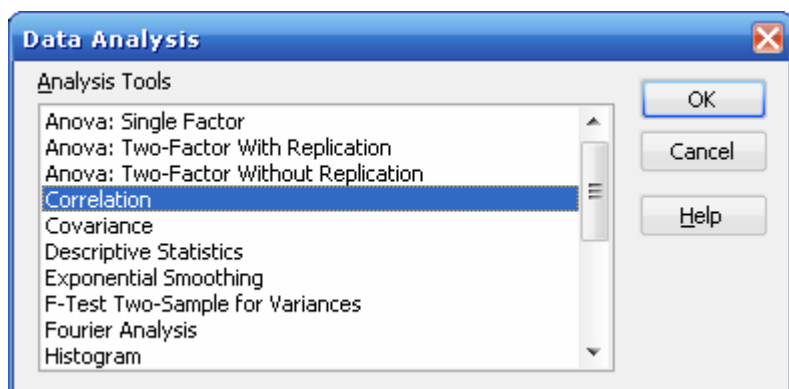
SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	4	20	5	4.666667
Column 2	4	27	6.75	10.91667
Column 3	4	21	5.25	1.583333
Column 4	4	25	6.25	7.583333

ANOVA

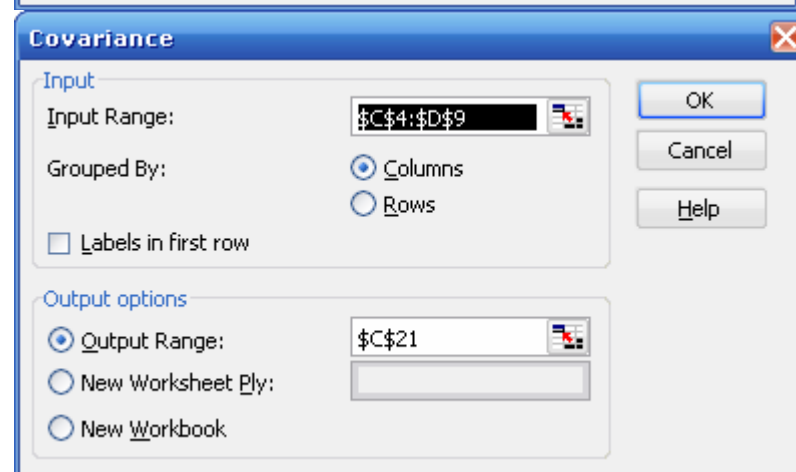
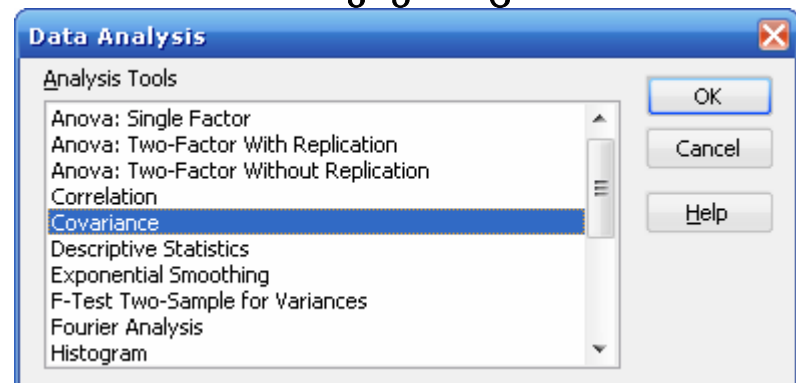
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	8.1875	3	2.729167	0.441077	0.727866	3.490295
Within Groups	74.25	12	6.1875			
Total	82.4375	15				

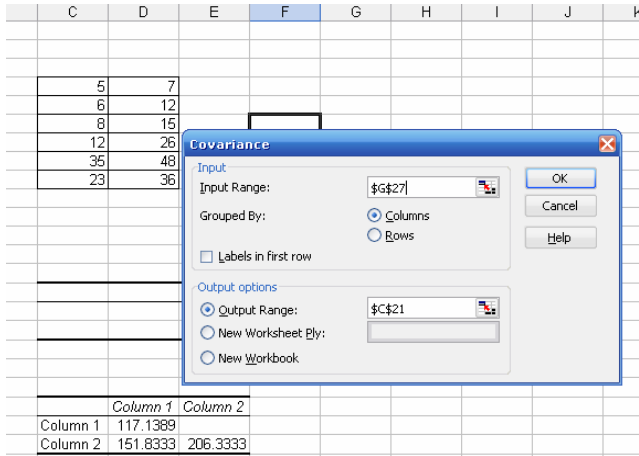
1.2.6. კორელაცია



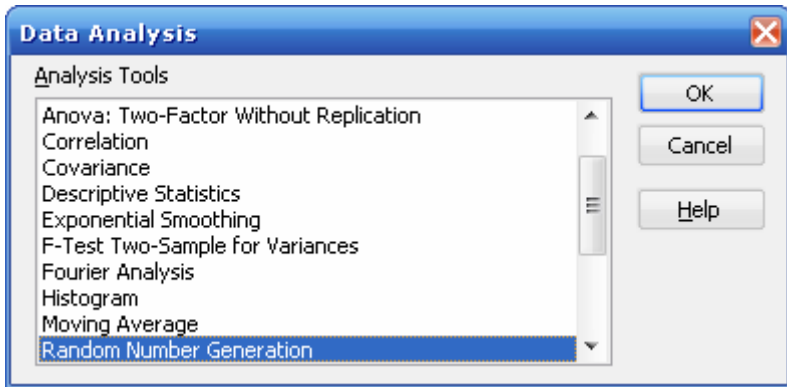
	Column 1	Column 2
Column 1	1	
Column 2	0.976634	1

1.2.7. კოვარიაცია



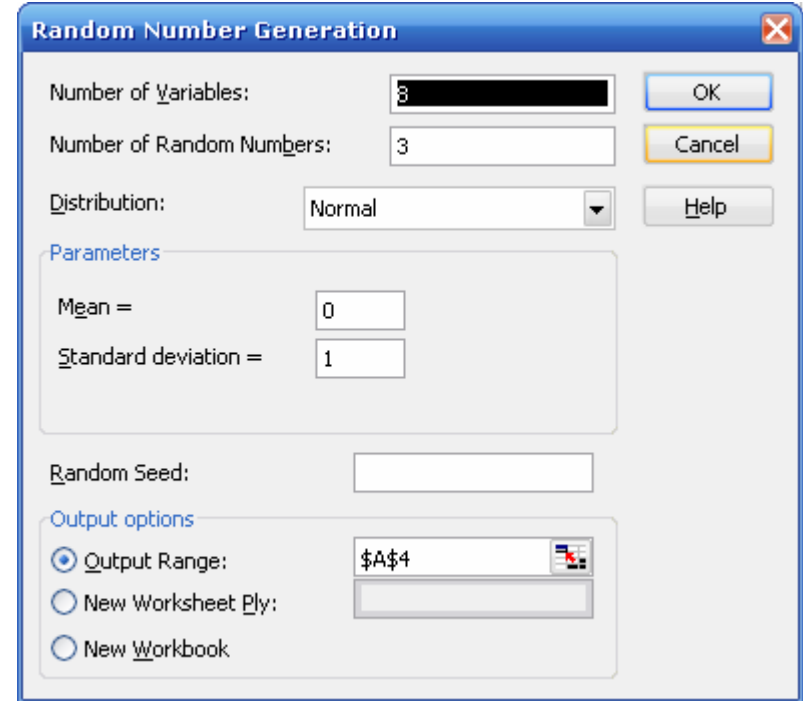


1.2.8. შემთხვევითი რიცხვების გენერირება



1.2.9. სტანდარტული ნორმალური განაწილება

მიუთითეთ რიცხვების რაოდენობა, Distribution - ველში აირჩიეთ Normal, ავტომატურად შეირჩევა სტანდარტული ნორმალური განაწილების პარამეტრები $\mu=0$; $\sigma=1$.



თუ საჭიროა ნორმალური განაწილების მქონე რიცხვები სხვადასხვა მათემატიკური ლოდინითა და საშუალო კვადრატული გადახრით, მაშინ შესაბამის

ველებში მიუთითეთ სასურველი მნიშვნელობები.

Parameters

Mean =

Standard deviation =

Microsoft Excel - Book2

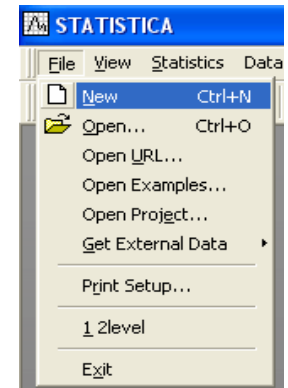
	B	C	D	E	F	G	H
1							
2							
3							
4	-0.79998	-0.87877	-1.19945	1.40044	0.815783	0.627564	-0.09281
5	-0.11874	0.55164	-0.92589	0.942816	1.378546	-0.15573	-0.5416
6	-0.87036	-1.20008	-1.53959	1.651802	1.8423	0.694675	0.547815
7							

2. სტატისტიკური დამუშავება მონაცემთა ანალიზის პაკეტის Statistica გამოყენებით

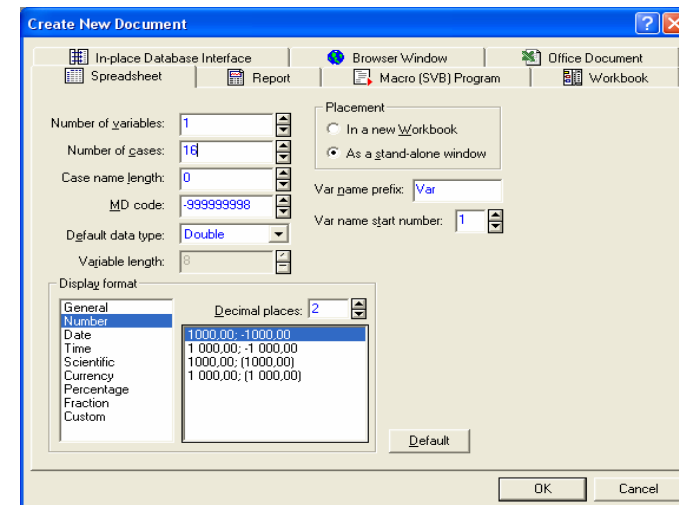
2.1. კვლევის მონაცემების პირველადი დამუშავება **Basic Statistics / Tables**

მოდულის მხედვით

1. ახალი დოკუმენტის შექმნა File-New (ნახ.2.1.1) . გაოსულ დიალოგურ ფანჯარაში (ნახ.2.1.2) შეავსეთ შესაბამისი ველები. მიუთითეთ ცვლადების რაოდენობა (Number of Variables), დაკვირვებათა რაოდენობა (Number of cases), ცვლადის ტიპი და ფორმატი (Display Format) და ა.შ.

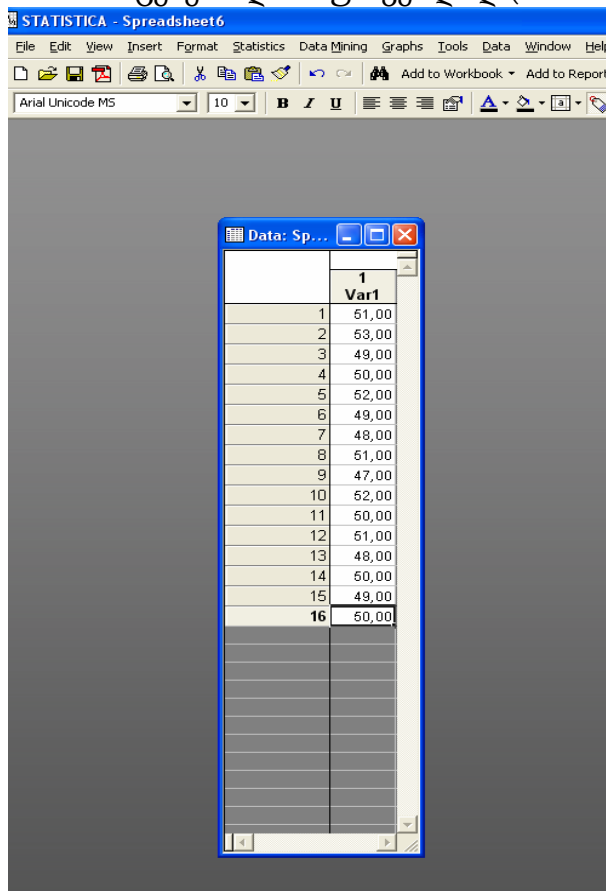


ნახ 2.11



ნახ 2.1.2

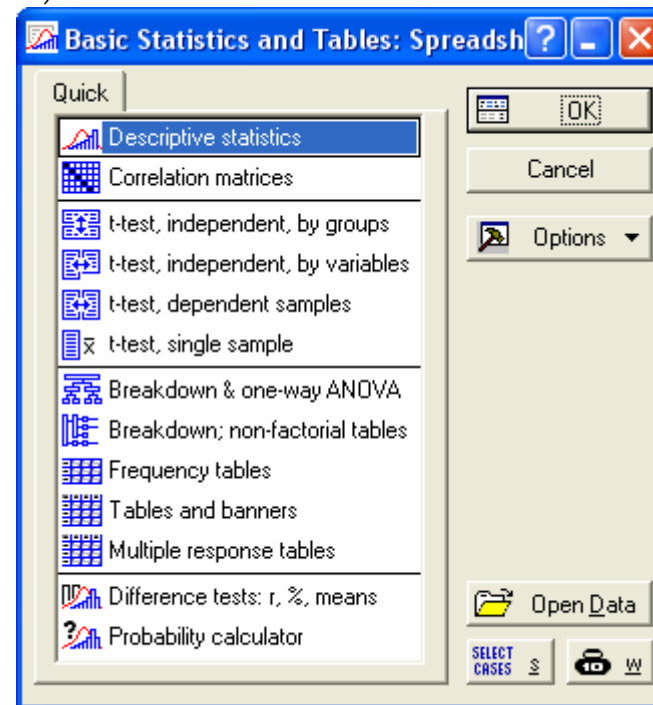
შევიტანოთ მონაცემები დასამუშავებლად (ნახ.2.1.3).



ნახ 2.1.3

მონაცემების დასამუშავებლად შევირჩიოთ დამუშავების სახე. მონაცემთა დამუშავება ხორციელდება Basic Statistics / Tables მოდულით. ამ მოდულში გაერთიანებულია მონაცემთა დამუშავების საწყის ეტაპზე ყველაზე ხშირად გამოყენებული

პროცედურები. საწყის პანელზე გამოდის ამ მოდულის სტატისტიკური პროცედურების ჩამონათვალი (ნახ.2.1.4).



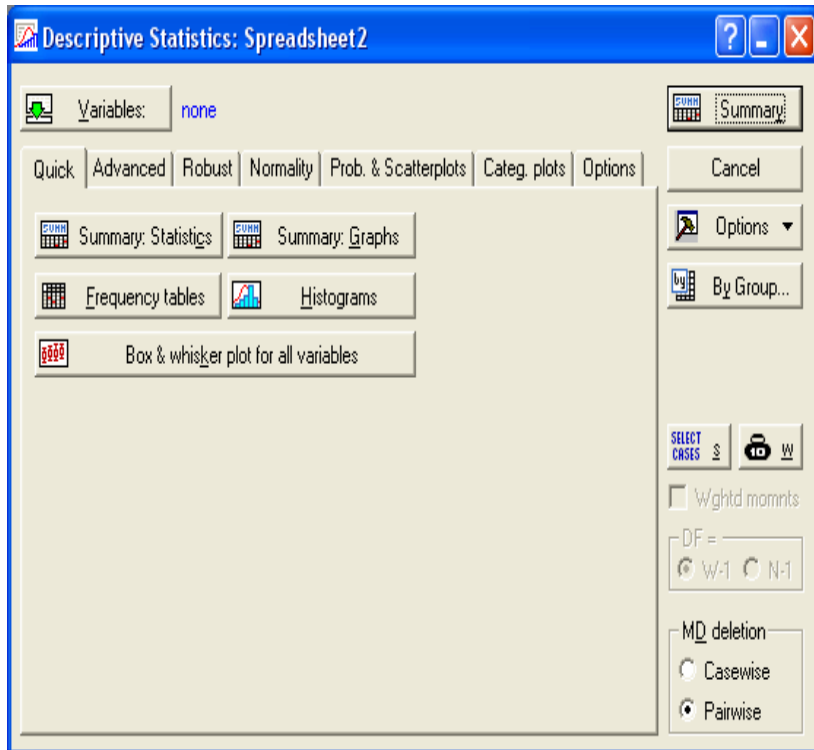
ნახ 2.1.4

Descriptive statistics - აღწერითი სტატისტიკა;
Correlation matrices- კორელაციური მატრიცა;
t-test for independent samples - t-ტესტი დამოუკიდებელი ამონარჩევისათვის
t-test for dependent samples - t-ტესტი პარალელური დაკვირვებების შემთხვევაში და ა.შ.

2.2. პროცედურა *Descriptive statistics* (აღწერითი სტატისტიკა)

განვიხილოთ ამ პროცედურის შესაძლებლობები მაგალითის საფუძველზე.

აღწერითი სტატისტიკის პროცედური შერჩევის შემდეგ ეკრანზე ჩნდება ანალოგური დასახელების დიალოგური ფანჯარა (ნახ.2.2.1)



ნახ.2.2.1

Variables- ანალიზისათვის ცვლადების შერჩევა (ნახ.2.2.2);

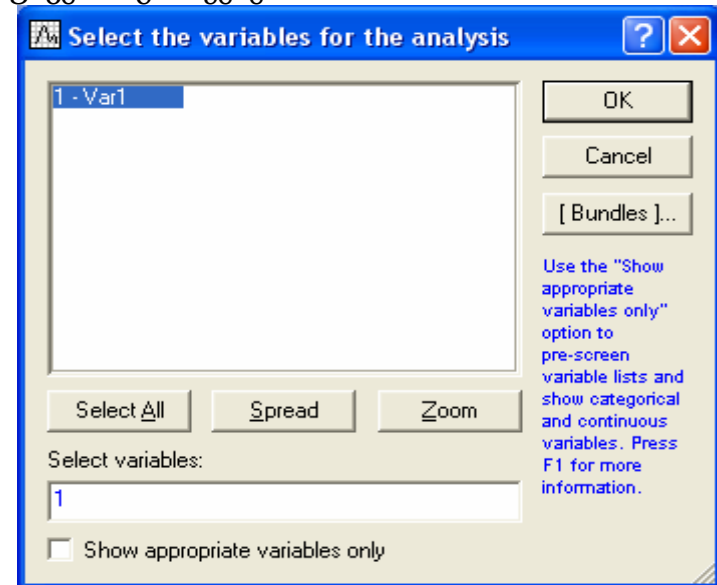
Summary Statistics – აღწერითი სტატისტიკის მაჩვენებლების გამოთვლა. გამოსაანგარიშებელი მაჩვენებლების განსაზღვრა ხდება დამატებით Advanced ჩანართის გააქტიურებით (ნახ.2.2.3)

Summary: Graphs – გრაფიკული ანალიზის სერვისის ჩართვა;

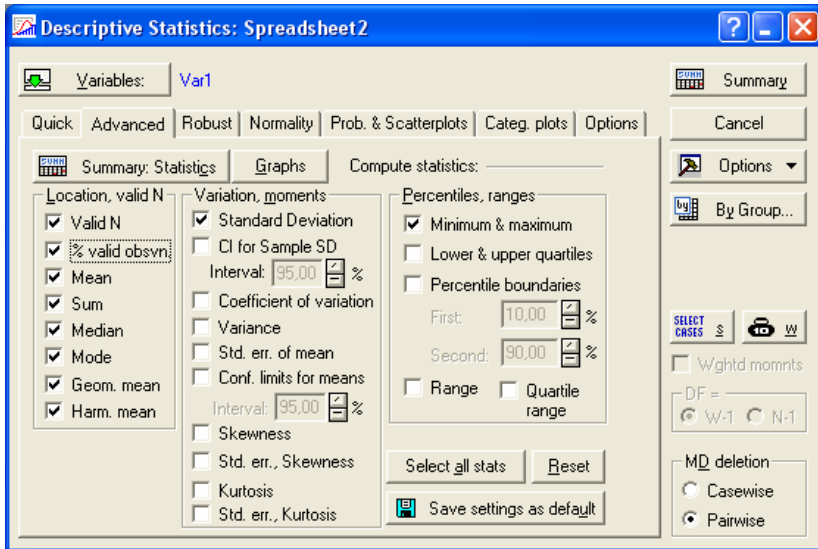
Frequency tables – სიხშირული ცხრილის აგება ინტერვალთა მეთოდით;

Histograms – ჰისტოგრამები;

Box @ Whisker plot for all variables - აღწერითი სტატისტიკის მაჩვენებლების ვიზუალიზაცია ჩარჩო-გრაფიკების გამოყენებით.



ნახ 2.2.2



ნახ 2.2.3

Valid N - ამონარჩევის მოცულობა;

Mean - საშუალო არითმეტიკული;

შემთხვევითი სიდიდის საშუალო მნიშვნელობა წარმოადგენს ყველაზე ტიპიურ, ალბათურ მნიშვნელობას, გარკვეულ თვისებურ ცენტრს, რომლის მიმართ გაბნეულია თვისების ყველა მნიშვნელობა.

Sum - ჯამი;

Median - მედიანა;

Standard Deviation - სტანდარტული გადახრა;

Variance - დისპერსია;

Standard error of mean - საშუალო მნიშვნელობის საშუალო ცდომილება;

95% confidence limits of mean - 95%-იანი ნდობითი ინტერვალი საშუალო მნიშვნელობისათვის;

Minimum, maximum - მინიმალური და მაქსიმალური მნიშვნელობა;

Lower, upper quartiles - ზედა და ქვედა კვარტილი;

Range - გაქანება;

Quartile range - ინტერკვარტილური გაქანება;

Skewness - ასიმეტრია;

Standard error of Skewness - ასიმეტრიის

სტანდარტული ცდომილება;

Kurtosis - ექსცესი;

Standard error of Kurtosis - ექსცესის სტანდარტული

ცდომილება;

Statistica პაკეტი არ იძლევა საშუალებას გამოვთვალოთ ყველაზე ხშირად გამოყენებული სტატისტიკური მნიშვნელობები ვარიაციის კოეფიციენტი და საშუალო მნიშვნელობის ფარდობითი ცდომილება (ცდის სიზუსტე).

$$\text{ვარიაციის კოეფიციენტი} = \frac{S \text{ standardDeviation}}{\text{Mean}} \cdot 100\%$$

$$\text{ცდის სიზუსტე} = \frac{S \text{ standardErrorOfMean}}{\text{Mean}} \cdot 100\%$$

ჰისტოგრამისა და სიხშირეთა ცხრილის ასაგებად გამოიყენება Descriptive statistics ფანჯრის **Distribution** ჯგუფის ღილაკები. დაჯგუფებული მონაცემების ინტერვალების (კლასების) რაოდენობა რეგულირდება **Number of intervals** გადამრთველის დახმარებით. **Distribution** ღილაკის მარჯვნივ მდებარეობს **Categorization** ჯგუფი, რომელიც საშუალებას იძლევა

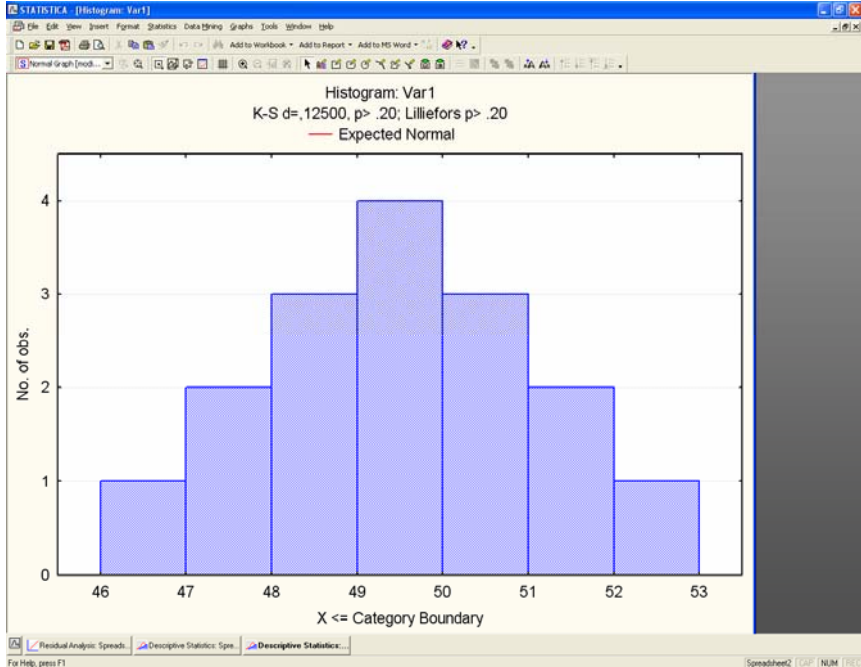
ინტერვალის სიგრძე იყოს მთელი რიცხვით გამოსახული **Integer intervals (categories)**.

დაჯგუფების შედეგად მიღებული ინტერვალები მოცემულია ცხრილში (ნახ.2.2.4) .

ჰისტოგრამის ასაგებად გამოვიყენოთ ღილაკი **Histograms**. ჰისტოგრამაზე აუცილებლობის შემთხვევაში შესაძლებელია დავამატოთ ნორმალური განაწილების სიმკვრივის გრაფიკი და შევამოწმოთ მისი მიახლოება განაწილების ნორმალურ სახესთან კოლმოგოროვ-სმირნოვის, ლილიეფორსის კრიტერიუმების გამოყენებით; ასევე შესაძლებელია შაპირო-უილკსის სტატისტიკის გამოანგარიშება. მიღებული მნიშვნელობები აისახება პირდაპირ ჰისტოგრამაზე (ნახ.2.2.5).

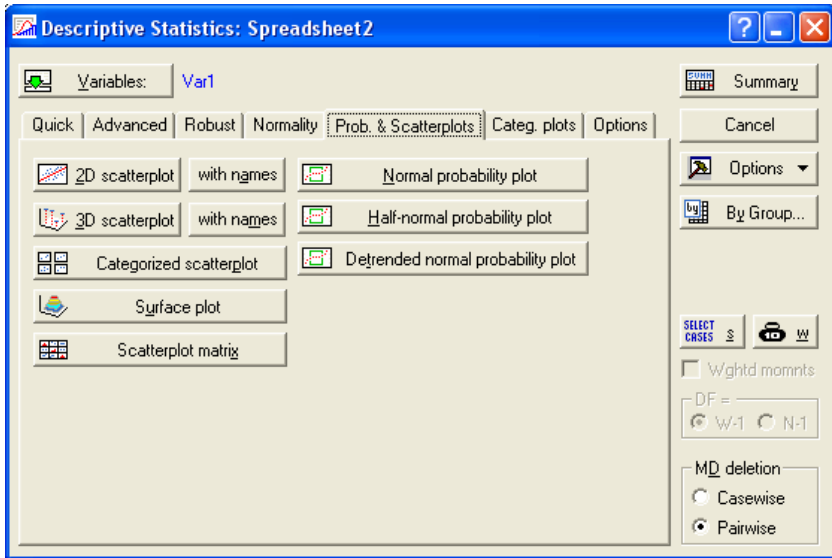
Category	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
46,00000 <x <= 47,00000	1	1	6,25000	6,25000	6,25000	6,25000
47,00000 <x <= 48,00000	2	3	12,50000	18,75000	12,50000	18,75000
48,00000 <x <= 49,00000	3	6	18,75000	37,50000	18,75000	37,50000
49,00000 <x <= 50,00000	4	10	25,00000	62,50000	25,00000	62,50000
50,00000 <x <= 51,00000	3	13	18,75000	81,25000	18,75000	81,25000
51,00000 <x <= 52,00000	2	15	12,50000	93,75000	12,50000	93,75000
52,00000 <x <= 53,00000	1	16	6,25000	100,00000	6,25000	100,00000
Missing	0	16	0,00000	0,00000	0,00000	100,00000

ნახ 2.2.4

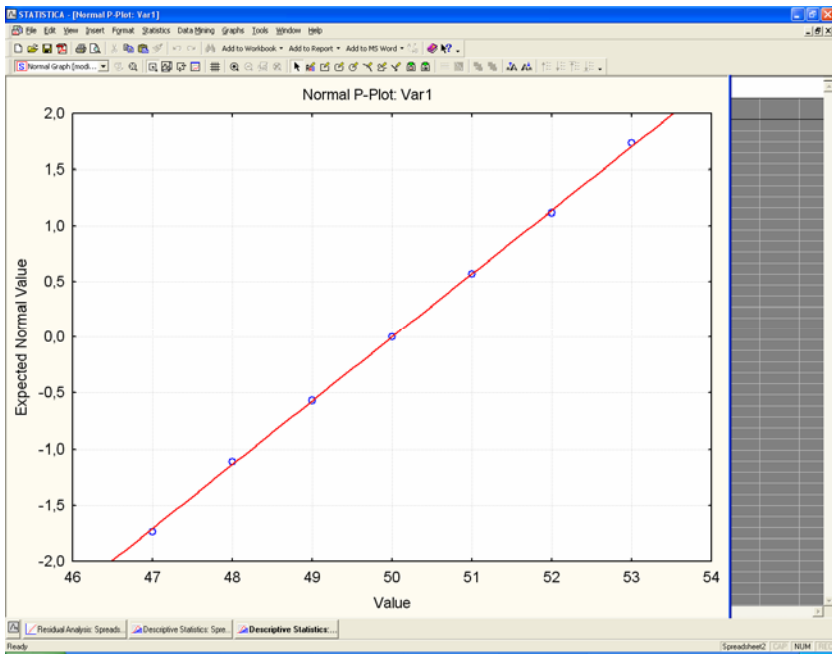


ნახ 2.2.5

ნორმალური განაწილების შესახებ შეიძლება ვიმსჯელოთ გრაფიკის მიხედვით. ამისათვის გამოვიყენოთ Normal probability plots ღილაკი (ნახ.2.2.6). რაც უფრო ახლოსაა განაწილება ნორმალური განაწილების სახესთან, მით უფრო კარგად განთავსდებიან მნიშვნელობები წრფეზე (ნახ.2.2.7). შეფასების ეს ხერხი ფაქტობრივად წარმოადგენს მხედველობით შეფასებას. საექვო შემთხვევებში საჭიროა მახლოების შემოწმება სხვადასხვა სტატისტიკური კრიტერიუმებით (კოლმოგოროვ-სმირნოვი, ომეგა-კვადრატი). ამისათვის საჭიროა მონაცემთა მნიშვნელოვანი რაოდენობა (ზოგიერთი ავტორის აზრით არანაკლებ 100-სა).



5sb 2.2.6



5sb 2.2.7

იბეჭდება ავტორთა მიერ წარმოდგენილი სახით

გადაეცა წარმოებას 28.05.2009. ხელმოწერილია დასაბეჭდად 12.06.2009. ქალაქის ზომა 60X84 1/16. პირობითი ნაბეჭდი თაბახი 6,5. ტირაჟი 100 ეგზ.

საგამომცემლო სახლი „ტექნიკური უნივერსიტეტი“, თბილისი,
კოსტავას 77



ი.მ. „გონა დალაქიშვილი“,
ქ. თბილისი, ვარკეთილი 3, კორპ. 333, ბინა 38