შოთა რუსთაველის სახელმწიფო უნივერსიტეტი

თენგიზ ხინიკაძე

STATISTICA

პროგრამული პაკეტის გამოყენების საფუძვლები

დამტკიცებულია სახელმძღვანელოდ
რსუ აკადემიური საბჭოს მიერ
21.07.2011 დადგენილება №93

გამომცემლობა
„შოთა რუსთაველის სახელმწიფო უნივერსიტეტი"
ბათუმი – 2011

სახელმძღვანელოში განხილულია პროგრამული პაკეტის STATISTICA გამოყენების შესაძლებლობები მონაცემთა თანამედროვე ანალიზის სფეროში. კორელაციური, რეგრესიული, დისპერსიული, დისკრიმინანტული, ფაქტორული, კლასტერული ანალიზების ჩატარების ხერხები და მეთოდები; ასევე კლასიფიკაციის ხეებისა და გრაფიკული ინსტრუმენტების მართვის მრავალრიცხოვანი საშუალებების გამოყენება. მოყვანილია ზოგადი ცნობები ნეირონული ქსელების შესახებ; აღწერილია ტექნიკური ნეირონის მოდელი. ნეირონული ქსელების საშუალებით გადაწყვეტილია აპროქსიმაციის, კლასიფიკაციისა და პროგნოზირების ამოცანები. იგი განკუთვნილია სტუდენტების, მაგისტრანტების, დოქტორანტებისა და ინსტრუმენტალური საშუალებების გამოყენებით მონაცემთა ანალიზის შესრულების ყველა მსურველისთვის.

რეცენზენტები:

გ. გოგიჩაიშვილი - ტექნიკურ მეცნიერებათა დოქტორი, საქართველოს პოლიტექნიკური უნივერსიტეტის ორგანიზაციული მართვის დეპარტამენტის სრული პროფესორი

ი. დიდმანიძე - ფიზიკა-მათემატიკის მეცნიერებათა დოქტორი, შოთა რუსთაველის სახელმწიფო უნივერსიტეტის კომპიუტერულ მეცნიერებათა დეპარტამენტის სრული პროფესორი

წინასიტყვაობა

მონაცემთა ანალიზის ადექვატური ტექნოლოგიების გარეშე ინფორმაციით გადატვირთულ გარემოში ადამიანი უსუსურია და არა აქვს რაციონალური გადაწყვეტილების მიღების საშუალება. მონაცემთა ანალიზის თანამედროვე სტატისტიკური საშუალებები საშუალებას იძლევიან კომპაქტურად აღიწეროს მონაცემები, გასაგები გახდეს მათი სტრუქტურა, ჩატარდეს კლასიფიკაცია, შემთხვევითი მოვლენების ქაოსში გამოიკვეთოს კანონზომიერებანი. ყოველივე ეს კი ზრდის სწორად მიღებულ გადაწყვეტილებათა წილს.

დღეისათვის მონაცემთა ანალიზისა და დამუშავების კომპიუტერული სტატისტიკური სისტემები ფართოდ გამოიყენებიან ეკონომიკის, წარმოების, მედიცინის და სხვა დარგებში. მათ შეუძლიათ გადაჭრან მონაცემთა „დაზვერვითი“ ანალიზის ფართო სპექტრი, ჩაატარონ პარამეტრებს შორის დამოკიდებულებათა სტატისტიკური კვლევა, ექსპერიმენტების დაგეგმვა, დროითი მწკრივების ანალიზი და პროგნოზირება, არარიცხვითი ბუნების მონაცემთა ანალიზი და სხვა.

მრავალ სტატისტიკურ პაკეტს შორის გამორჩეული ადგილი უჭირავს პაკეტს STATISTICA. ის ყველაზე დინამიურად განვითარებად პაკეტს წარმოადგეს და მრავალრიცხოვანი რეიტინგებით მსოფლიო ლიდერია სტატისტიკური პროგრამული უზრუნველყოფის ბაზარზე, ამიტომ მისი შესწავლა კომპიუტერული ტექნოლოგიების სპეციალისტისათვის მნიშვნელოვნად მიგვაჩნია.

STATISTICA-ს დამმუშავებელია ფირმა StatSoft, Inc.,(აშშ). ამ პროგრამის პირველი ვერსია DOS-ისათვის გამოვიდა 1991 წელს, ხოლო 1994 წელს-ვერსია Windows-ისათვის.

STATISTICA-ს პროგრამული პროდუქტების ხაზი დაფუძნებულია ყველაზე თანამედროვე ტექნოლოგიებზე, სრულად შეესაბა-

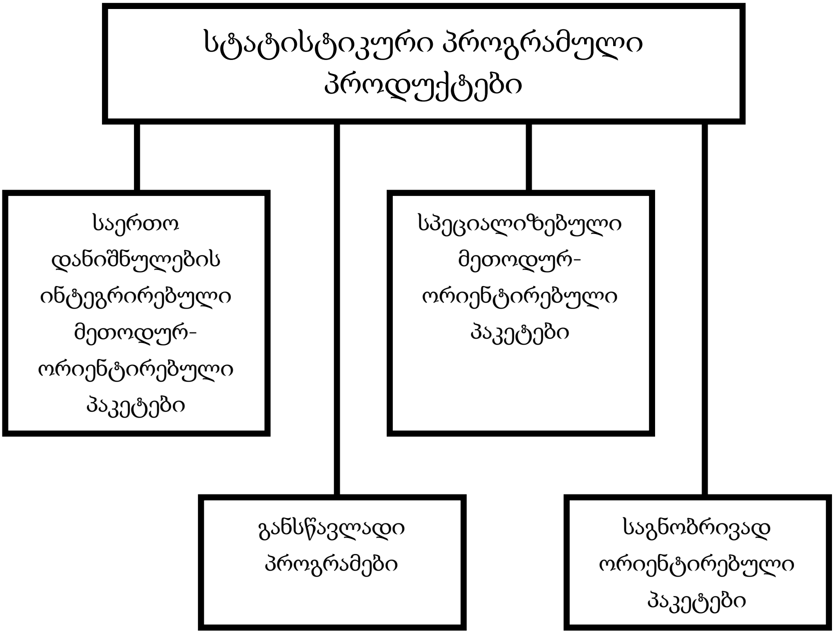
მება ინფორმაციული ტექნოლოგიების სფეროში უკანასკნელ მიღწევებს, საშუალებას იძლევა მონაცემთა ანალიზისა და დამუშავების სფეროში გადაწყვიტოს ყველაზე რთული ამოცანები, იდეალურად ესადაგება ნებისმიერ სფეროში გამოყენებას: მარკეტინგში, ფინანსებში, დაზღვევაში, ეკონომიკაში, ბიზნესში, წარმოებაში, მედიცინაში და სხვა.

STATISTICA-ის გარემოში არსებობს იმის შესაძლებლობა, რომ ჩატარდეს როგორც კლასიკური ისე უახლესი მეთოდებით მონაცემთა ანალიზი: კლასტერული, ფაქტორული, კორელაციური, დისპერსიული, წრფივი და არაწრფივი რეგრესიები, ნეირონული ქსელები და სხვა. საწყისი, საშუალედო და გამოსასვლელი მონაცემების ვიზუალიზაცია შეიძლება განხორციელდეს სხვადასხვა გრაფიკების, პიქტოგრამებისა და დიაგრამების დიდი რიცხვიდან ამორჩევით.

რასაკვირველია, ანალიზური პაკეტი ეს მხოლოდ ინსტრუმენტია. თავის-თავად ის არ იძლევა სწორი შედეგის მიღების გარანტიას. პაკეტის შესაძლებლობების სრულად გამოყენებისათვის საჭიროა მისით მართებულად სარგებლობა. ნებისმიერი ანალიზური ინსტრუმენტი მოითხოვს განსაზღვრულ ცოდნას ალბათობის თეორიასა და მათემატიკურ სტატისტიკაში.

1. სტატისტიკური პაკეტების მოკლე მიმოხილვა

სტატისტიკური პროგრამული პროდუქტების (სპპ) ბაზარი წარმოუდგენლად მრავალფეროვანია. არსებობს მსოფლიო ბაზარზე გავრცელებული ათასამდე პაკეტი, რომელიც წყვეტს მონაცემთა სტატისტიკური ანალიზის ამოცანებს. შეიძლება გამოვყოთ სტატისტიკური პაკეტების ოთხი ძირითადი ჯგუფი (ნახ. 1.1.).



ნახ. 1.1.

შევჩერდეთ დაწვრილებით მეთოდურ-ორიენტირებულ პაკეტებზე (ცხრ. 1).

ცხრილი 1

მეთოდურ-ორიენტირებული სტატისტიკური პროგრამების კლასიფიკაცია

სტატისტიკური პროგრამული პროდუქტების კლასი	სტატისტიკური პროგრამული პროდუქტების დასახელება
საერთო დანიშნულების უნივერსალური (ინტეგრირებული) სტატისტიკური პაკეტები	SAS, SPSS/W, SYSTAT, MINITAB, Statgraphics, BMDP Dynamic, STATISTICA/W, Stat View და Super ANONA
მკვლევარის ინსტრუმენტა-ლური საშუალება, რომელიც შეიცავს მძლავრ სტატისტი-კურ კომპონენტას	IMSL, S-Plus
განზომილების შემამცირებე-ლი და კლასიფიკაციის სპე-ციალიზირებული პაკეტები	КЛАСС-МАСТЕР, Stat-Media, PALMODA (ЛОРЕГ), STARC, КВА-ЗАР, PolyAnalyst, MVSP, CART
ზოგიერთი სხვა სპეციალიზი-რებული და უნივერსალური სტატისტიკური პროგრამული პროდუქტი	MESOSAUR, SANI, Stat View for Wi dows, STADIA, ОЛИМП, РОСТАН, NCSS Statistical Software, ODA, SOL STATlab Pro, UNISTAT, STATIT, WinSTAT, Multivariance 7, JMP, BM STAT, DATA DESK, SAM-86, STATMOST, POWERSTAT
პაკეტები და პროგრამები, რომლებიც წყვეტენ კლასიფი-კაციის მომიჯნავე ამოცანებს	«Статистик-Консультант», BMDP д Windows, TURBO Spring-Stat-Win, STATISTIX, SigmaStat, StatXact-3, M Excel-7.0

სტატისტიკური ექსპერტული სისტემები	СТАТЭКС, Statistical Navigator Pro, STAREX

უნივერსალურ პაკეტებში, რომლებიც გვთავაზობენ სტატისტიკური მეთოდების ფართო დიაპაზონს, არ არსებობს ორიენტაცია კონკრეტულ საგნობრივ არეზე. ამ პაკეტებიდან ყველაზე ფართო გავრცელება მოიპოვეს კომპიუტერულმა სისტემებმა SAS, SPSS, SYSTAT, Minitab, Statgraphics, STATISTICA.

სპეციალიზებულ პაკეტები, როგორც წესი, შეიცავენ რამდენიმე სტატისტიკურ მეთოდს ან მეთოდებს, რომლებიც გამოიყენებიან კონკრეტულ საგნობრივ არეში. ყველაზე ხშირად ეს სისტემები ორიენტირებულებია დროითი მწკრივების ანალიზზე, კორელაციურ-რეგრესიულ, ფაქტორულ ან კლასტერულ ანალიზებზე. **„ნახევრად-სპეციალიზებულ“ და „ნახევრად-უნივერსალურებად“** შეიძლება ჩავთვალოთ პაკეტები STADIA, ОЛИМП, РОСТАН. ამავე კლასს უნდა მივაკუთვნოთ პაკეტები ODA, WinSTAT, Statit, UNISTAT, Multivariance 7, JMP, SOLO, STATlab.

კლასიფიკაციისა და განზომილების შემცირების სპეციალიზებულ პაკეტებს შეიძლება მივაკუთვნოთ პაკეტები КЛАСС-МАСТЕР, КВАЗАР, PALMODA, Stat-Media, STARC, MVSP.

ფართოდაა ცნობილი ისეთი პაკეტები, რომლებიც წყვეტენ კლასიფიკაციასთან მომიჯნავე ამოცანებს: BMDP/W, SigmaStat, Statistix, TURBO Spring-Stat-Win, „Статистик-Консультант для Windows“. გარდა ამისა, არსებობენ **სტატისტიკური ექსპერტული სისტემები,** მაგალითად СТАТЭКС, Statistical Navigator Pro. არა სტატისტიკური პაკეტებიდან, რომლებიც წყვეტენ კლასიფიკაციის ამოცანებს შეიძლება გამოვყოთ პაკეტები PolyAnalyst, ДА-система, АРГОНАВТ, ЛОРЕГ, პაკეტი ОТЭКС და სხვადასხვა ნეირონულქსე-ლური პაკეტები.

მეთოდურ-ორიენტირებულ სპპ-ში შეიძლება შედიოდნენ შემდეგი ფუნქციონალური ბლოკები.

I. აღწერითი სტატისტიკისა და საწყისი მონაცემების დაზვერვითი ანალიზის ბლოკი ითვალისწინებს:
   - მრავალგანზომილებიანი ნიშან-თვისების შერეული ბუნე-ბის ანალიზი და საწყისი მონაცემების ჩაწერის უნიფიკაცია;
   - მკვეთრად გამორჩეული დაკვირვებების ანალიზი;
   - გამოტოვებული დაკვირვებების აღდგენა;
   - დაკვირვებათა სტატისტიკური დამოუკიდებლობის შემოწმება;
   - ძირითადი რიცხვითი მახასიათებლების განსაზღვრა და საწყისი მონაცემების სიხშირული დამუშავება (ჰისტოგრამე-ბის, სიხშირული პოლიგონების აგება, შერჩევითი საშუალო-ების, დისპერსიების გამოთვლა);
   - პარამეტრების სტატისტიკური შეფასება;
   - ალბათობების განაწილების მოდელური კანონების (ნორმა-ლური, ბინომიალური, პუასონის, Chi-კვადრატ და სხვა) გა-მოთვლა;
   - გასაანალიზებელი მრავალგანზომილებიანი სტატისტიკური მონაცემების ვიზუალიზაცია და სხვა.

II. დამოკიდებულებების სტატისტიკური კვლევის ბლოკი ითვა-ლისწინებს:
   - კორელაციურ და რეგრესიულ ანალიზებს;
   - დისპერსიულ და კოვარიაციულ ანალიზებს;
   - რეგრესიული ექსპერიმენტების დაგეგმვას და შერჩევით გამოკვლევებს;

- დროითი მწკრივების ანალიზს (დროითი მწკრივების წინასწარი ანალიზი; დროითი მწკრივების ტრენდის გამოვლენა; ფარული პერიოდულობის გამოვლენა, დროითი მწკრივის სპექტრალური ანალიზი; დროითი მწკრივის შემთხვევითი ნაშთების ანალიზი; სტატისტიკური ჰიპოთეზების შემოწმება რიგის სტაციონალურობის, მისი წევრების დამოუკიდებლობის, „ასაგები“ მოდელის ადექვატურობის შესახებ) და სხვა.

III. კლასიფიკაციისა და განზომილების შემცირების ბლოკი ითვალისწინებს:
   - დისკრიმინანტულ ანალიზს;
   - განაწილებათა ნარევების სტატისტიკურ ანალიზს;
   - კლასტერ-ანალიზს;
   - განზომილების შემცირების გარე ინფორმაციულობისა და ავტოინფორმაციულობის კრიტერიუმთა შესაბამისად და ზოგიერთი სხვა.

IV. არა რიცხვითი ინფორმაციის სტატისტიკური ანალიზის მეთოდებისა და ექსპერტული შეფასებების ბლოკი.

ამ ბლოკში გამოყენებულ მათემატიკურ-სტატისტიკურ ინსტრუმენტებს შორის შეიძლება ავღნიშნოთ: შეუღლებული ცხრილების ანალიზი, ლოგ-წრფივი მოდელები, სუბიექტური ალბათობები, ლოგოტ- და პრობიტ-ანალიზი, რანგული მეთოდები და ა.შ.

V. ექსპერიმენტების დაგეგმვისა და შერჩევითი გამოკვლევების ბლოკი.

VI. დამხმარე პროგრამების ბლოკი ითვალისწინებს ეგმ-ზე სტატისტიკურ მოდელირებას, ერთ და მრავალგანზომილებიან დაკვირვებათა გენერირების გათვალისწინებით.

უნივერსალურ მეთოდურ-ორიენტირებულ სტატისტიკურ პაკეტებს შორის ერთ-ერთ ყველაზე დინამიურად განვითარებადს წარმოადგენს ამერიკული ფირმა StatSoft-ის (http://www.statsoft. com) პროდუქტი Statistica for Windows (შემდგომში STATISTICA), რომელიც მრავალრიცხოვანი რეიტინგებით მსოფლიო ლიდერია.

STATISTICA-ს ღირსება იმაში მდგომარეობს, რომ ის თავიდანვე დაფუძნებული იყო Windows-ტექნოლოგიებზე: არ მომხდარა მისი გადმოტანა სხვა პლატფორმიდან. შეიძლება ითქვას, რომ STATISTICA Windows-გარემოა, რომლის დანიშნულებაა მონაცემების ყოველმხრივი სტატისტიკური ანალიზის ჩატარება. არსებობს 6 მიზეზი, რომლის გამოც სხვა სისტემებს შორის უპირატესობა STATISTICA-ს უნდა მივანიჭოთ: უპირველესად, სრული შეთანხმებულობა Windows-ის სტანდარტებთან, ადვილი ათვისება (ზევრად დაკავშირებული Windows-ტექნოლოგიების რეალიზაციაზე, რაც ასახავს სტატისტიკოსების ინტუიციურ წარმოდგენას მონაცემთა ანალიზის გარემოს შესახებ), მინიმალური მოთხოვნები კომპიუტერის რესურსების მიმართ, უნიკალური სამეცნიერო და პრეზენტაციული გრაფიკა, სისტემაში წარმოდგენილი კლასიკური და თანამედროვე სტატისტიკური მეთოდების ამომწურავი ნაკრები, ფასი, რომელიც მისაწვდომია მომხმარებელთა ფართო წრისათვის.

STATISTICA იყენებს ელექტრონული ცხრილების სტანდარტულ ინტერფეისს. მონაცემთა მიმდინარე ფაილი ყოველთვის აისახება ელექტრონული ცხრილის სახით. მონაცემები ორგანიზებულნი არიან დაკვირვებებისა და ცვლადების სახით. დაკვირვებები შეიძლება განვიხილოთ როგორც ელექტრონული ცხრილების სვეტების ექვივალენტები. თითოეული დაკვირვება შედგება ცვლადების მნიშვნელობათა ნაკრებისაგან.

მონაცემთა ანალიზისათვის სისტემაში რეალიზებულია ე. წ. გრაფიკულ-ორიენტირებული მიდგომა, რომლის არსიც იმაში მდგომარეობს, რომ მივიღოთ მონაცემების შესახებ ყოველმხრივი ვიზუალური წარმოდგენა სტატისტიკური დამუშავების ყველა ეტაპზე და მის საფუძველზე ავირჩიოთ ანალიზის შემდეგი ბიჯი.

სისტემა აგებულია მოდულური პრინციპით და შეიცავს სტატისტიკური ანალიზის ყველა ცნობილ მეთოდს. თითოეული მოდული შეიცავს პროცედურათა განსაზღვრულ კლასს. თითქმის ყველა პროცედურა ინტერაქტიულია, ანუ დამუშავების გასაშვებად საჭიროა მენიუდან ავირჩიოთ ცვლადები და ვუპასუხოთ სისტემის რიგ შეკითხვებს. ეს ფრიად მოსახერხებელია დამწყები მომხმარებლისათვის, თუმცა ანელებს გამოცდილის საქმიანობას და არ იძლევა საშუალებას ეფექტურად გაიმეოროს ერთი და იგივე პროცედურა რამდენიმეჯერ. მონაცემები ადვილად შეიძლება იქნან ექსპორტირებული პოპულარულ მონაცემთა ბაზებში და მათგან იმპორტირებულნი მაუსის ღილაკის რამდენიმე დაწკაპუნებით.

STATISTICA-ში არის შესანიშნავი საშუალება-ალბათური კალკულატორი, რომლითაც სარგებლობა ისევე მარტივია, როგორც ჩვეულებრივი კალკულატორით, და რომელიც საშუალებას გვაძლევს სწრაფად გამოვითვალოთ სხვადასხვა განაწილებების პროცენტული წერტილები, ასევე გამოვითვალოთ სხვადასხვა სტატისტიკების კრიტიკული მნიშვნელობები.

STATISTICA-ში გრაფიკთა ასობით სხვადასხვა ტიპია, რომელთა დანიშნულებაა საწყისი მონაცემების ვიზუალიზაცია, „დაზვერვითი“ ანალიზის ჩატარება, შედეგების გრაფიკული გამოტანა და ანალიზის მომდევნო მიმართულებების არჩევა. ისეთი უნიკალური გრაფიკები, როგორებიცაა ჩერნოვის სახეები, ვორონოვის დიაგრამები, მატრიცულები, საშუალებას იძლევიან, მაგალითად, მოვახ-

დინოთ კორელაციური მატრიცის „ვიზუალიზაცია“, კატეგორიზებული გრაფიკები, ტრასირების გრაფიკები და სხვა. აგრეთვე დიდი არჩევანი ორ და სამგანზომილებიანი სამეცნიერო და საქმიანი გრაფიკებისა და დიაგრამების ადვილად მისაწვდომი ხდება მომხმარებლისათვის მაუსის რამდენიმე დაწკაპუნებით.

გარდა გრაფიკთა სტანდარტული ტიპებისა, STATISTICA-ში არის სპეციალიზებული სტატისტიკური გრაფიკების დიდი რაოდენობა. გრაფიკები ავტომატურად განახლდებიან მათთან დაკავშირებული მონაცემთა ფაილების ცვლილებასთან ერთად. მონაცემების ვიზუალური ანალიზისათვის განსაკუთრებით მნიშვნელოვანია საშუალება „ფუნჯი“.

სისტემაში ჩართული პროგრამირების ენა STATISTICA BASIC საშუალებას იძლევა გაფართოვდეს სისტემის შესაძლებლობები, მოხდეს საკუთარი ორიგინალური მეთოდების დაპროგრამება.

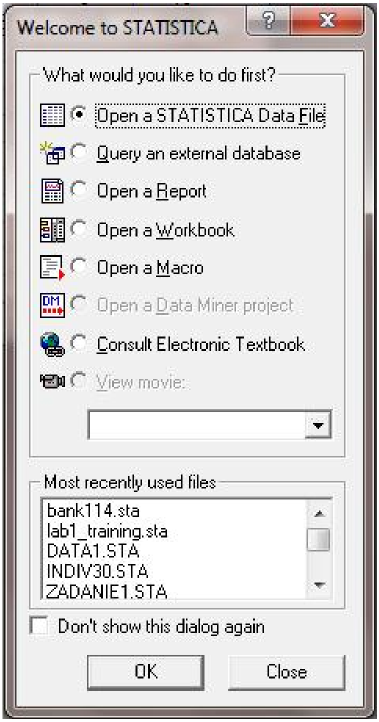
2. პაკეტ STATISTICA-თან მუშაობის საწყისები

2.1. სტატისტიკური პაკეტის ინსტალაცია და გაშვება

სისტემის ინსტალაცია ძალიან წააგავს ნებისმიერი სხვა Windows-დამატების ინსტალაციას. სისტემა მთლიანად ინტეგრირდება Windows-თან და ზოგიერთ სხვა დამატებასთან (მაგალითად, Microsoft Word-თან, Excel-თან), საშუალებას იძლევა გამოვიძახოთ, შევქმნათ და დავუმატოთ STATISTICA-ს დოკუმენტები ამ დამატებებს და პირიქით. მოცემული პროგრამული პროდუქტი ურთიერ-

თმოქმედებს ძირითად გრაფიკულ პაკეტებთან (მაგალითად, Adobe Photoshop-თან, Illustrator-თან).

პაკეტის თავდაპირველი გაშვებისას ჩნდება ნახ. 2.1-ზე წარმოდგენილი დიალოგური ფანჯარა, რომლის მეშვეობითაც სისტემა გვთავაზობს შევასრულოთ შემდეგი საწყისი მოქმედებები:

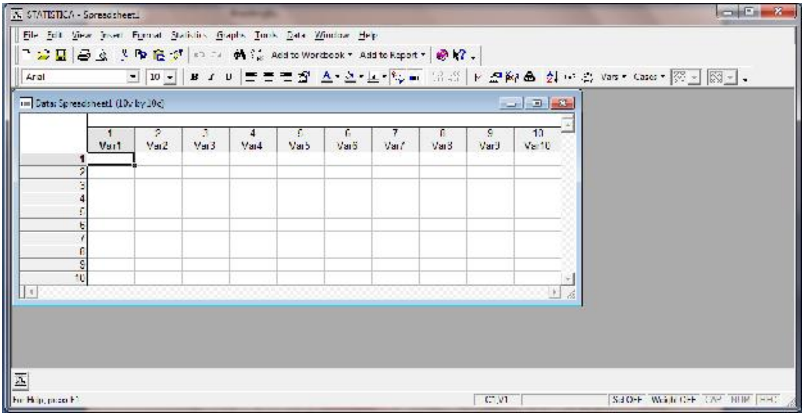


ნახ. 2.1.

1. Open a STATISTICA Data File – STATISTICA-ს მუშა ფურცლის გახსნა.
2. Query an external database – გარე მონაცემთა ბაზაზე მოთხოვნის შესრულება.
3. Open a Report – STATISTICA-ს ანგარიშის გახსნა.
4. Open a Workbook – STATISTICA-ს მუშა დავთრის გახსნა.

5. Open a Macro – STATISTICA-ს მაკროსის გახსნა.
6. Open a Data Miner project – „მონაცემების მომპოვებლის“ პროექტის გახსნა.
7. Consult Electronic Textbook – მომხმარებლის ელექტრონული სახელმძღვანელოს დახმარების მოთხოვნა.
8. View movie – ვიდეოს დათვალიერება.
9. Most recently used filed – ფანჯარაში წარმოდგენილია ყველაზე ხშირად გამოყენებულ ფაილთა სია.
10. Don’t show this dialog agein - ამ ადგილზე ალამის არსებობა ნიშნავს, რომ მომავალში სისტემის გაშვებისას ეს დიალოგური ფანჯარა აღარ გამოვა.

ამ მენიუთი სარგებლობა მოხერხებულია სისტემის მომდევნო გაშვებისას და იმ პირობებში, როცა პერსონალური კომპიუტერით სარგებლობს მომხმარებელთა მცირე რიცხვი. ჩვენს შემთხვევაში უმჯობესია დავაყენოთ ალამი ველში Don’t show this dialog agein და დავაწვეთ OK-ის. წინა პლანზე გამოდის გაშვებისას გახსნილი მუშა ფურცელი, რომელიც წააგავს Excel-ის დოკუმენტს, ზომებით 10*10 უჯრედი (ნახ. 2.2.). თუ მონაცემთა არსებული მასივი არ შეესაბამება ფურცლის შემოთავაზებულ ზომებს, იგი შეიძლება დახუროთ და გადახვიდეთ თქვენი ფაილის შექმნაზე (იხ. ქვემოთ). გასათვალისწინებელია, რომ ყოველი მომდევნო ჩართვისას სისტემა ავტომატურად ტვირთავს ბოლო გამოყენებულ ფაილს, ხოლო თუ მისი აღმოჩენა ვერ ხერხდება, მაშინ სისტემა ისევ ავტომატურად ტვირთავს ცარიელ მუშა ფურცელს.



ნახ. 2.2.

3. მონაცემების შეტანა პროგრამაში STATISTICA

3.1. მუშა დავთრის შექმნა

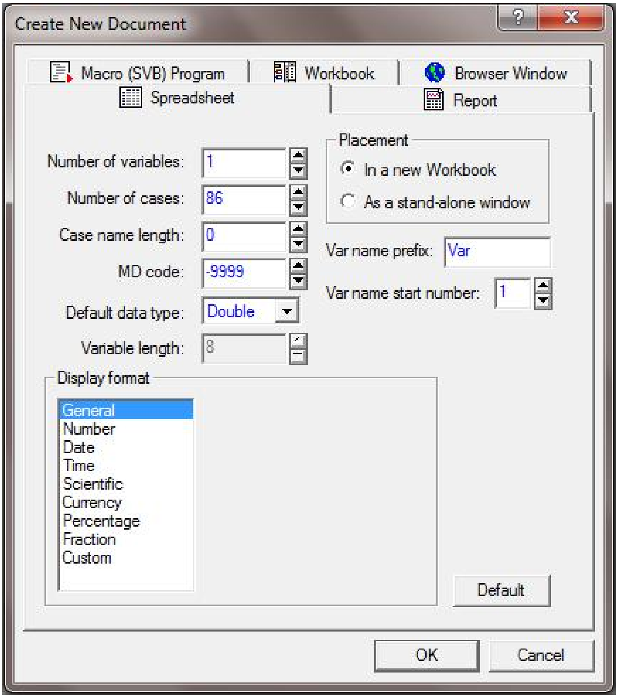
გამოყენებითი პროგრამული პაკეტებით სტატისტიკური ანალიზის ნებისმიერი ამოცანის ამოხსნა იწყება მონაცემების შეტანით. პაკეტ STATISTICA-ს არეში მონაცემთა შეტანის პროცედურა შემდეგნაირად ხორციელდება. უპირველესად საჭიროა ფაილის შექმნა, რისთვისაც სისტემის მთავარ მენიუში ირჩევთ პუნქტს File, ხოლო შემდეგ ოპციას New. პროგრამა გთავაზობთ რამდენიმე ტიპის ფაილის შექმნის შესაძლებლობას:

1. Spreadsheet – მუშა ფურცელი (ელექტრონული ცხრილი).
2. Report – ანგარიში.
3. Maqro (SVB) Program – მაკროსი.

4. Workbook – მუშა დავთარი.
5. Browser Window – ქსელში მითითება.

მუშა დავთარი (Workbook) ყველაზე ზოგადი ცნებაა. მუშა დავთარში შეიძლება ჩართული იყოს STATISTICA-სთვის მისაწვდომი ნებისმიერი ელემენტი, მაგალითად მუშა ფურცლები, ანგარიშები, MicrosofT office-ს დოკუმენტები.

მონაცემების შესატანად დაგჭირდებათ ელექტრონული ცხრილი. ეკრანზე მისი გამოტანისათვის საჭიროა აირჩიოთ ჩანართი Spreadsheet (ნახ. 3.1).



ნახ. 3.1.

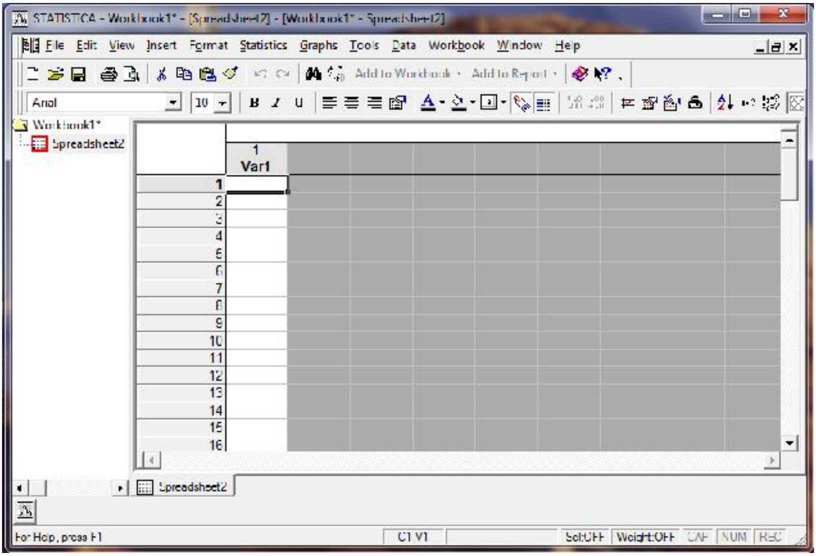
მომხმარებელს საშუალება ეძლევა განსაზღვროს ცხრილის ზომები, ანუ მიუთითოს ცვლადების რიცხვი (number of variables) და დაკვირვებათა რიცხვი (number of cases). ცვლადების რიცხვი - ეს

ვარირებულ ნიშან-თვისებათა რიცხვია, რომელთა მნიშვნელობები დარეგისტრირებულებია შესასწავლი სტატისტიკური ერთობლიობის თითოეულ ერთეულში. დაკვირვებათა რიცხვი - ერთობლიობის ერთეულთა რიცხვია (ერთობლიობის მოცულობა). შემდეგ ხდება იმ ადგილის შერჩევა, სადაც შენახულ უნდა იქნეს შესაქმნელი ელექტრონული ცხრილი (Placement). მიზანშეწონილია ავირჩიოთ "In a new Workbook", რაც ახალი მუშა ფურცლის (ელექტრონული ცხრილის) შექმნას ნიშნავს ახალი მუშა დავთრის შიგნით. პროცედურა "As a stand-alone window" ნიშნავს ცალკეული (დამოუკიდებელი) მუშა ფურცლის შექმნას.

შემდეგ შეიძლება ავირჩიოთ მონაცემების გამოტანის სხვადასხვა ვარიანტები. ველში Case name length ფიქსირდება სტრიქონის სახელის სიგრძე, ველში MD Code - გამოტოვებული უჯრედების გამოტანის კოდი (მითითების გარეშე მათი გათვალისწინება არ ხდება მაჩვენებლების გამოთვლისას), ველით Default data type ხდება მონაცემთა ფორმატის შერჩევა (რიცხვითი, ტექსტური, ბინარული; ჩუმათობის პრინციპით არჩეულია ტიპი Double, რაც ნიშნავს, რომ სისტემა აღითქვამს როგორც ტექსტურ, ისე რიცხვით მონაცემებს), ველის Variable length დანიშნულებაა მონაცემების სიგრძის დაფიქსირება მონაცემების ტექსტური ფორმატის დაფიქსირების შემთხვევაში. ველი Display format აფიქსირებს გამოყვანის პარამეტრებს შესატანი ან გამოთვლილი მონაცემებისათვის (ჩვეულებრივი, რიცხვითი, პროცენტული და ა.შ.), ველი Var name prefix გვთავაზობს შესაქმნელი ცვლადების დასახელების მითითებას, ხოლო ველი Var name start number უთითებს პირველი ცვლადის სასტარტო რიგით ნომერს. ღილაკი Default საშუალებას იძლევა დადგეს ყველა გაწყო-

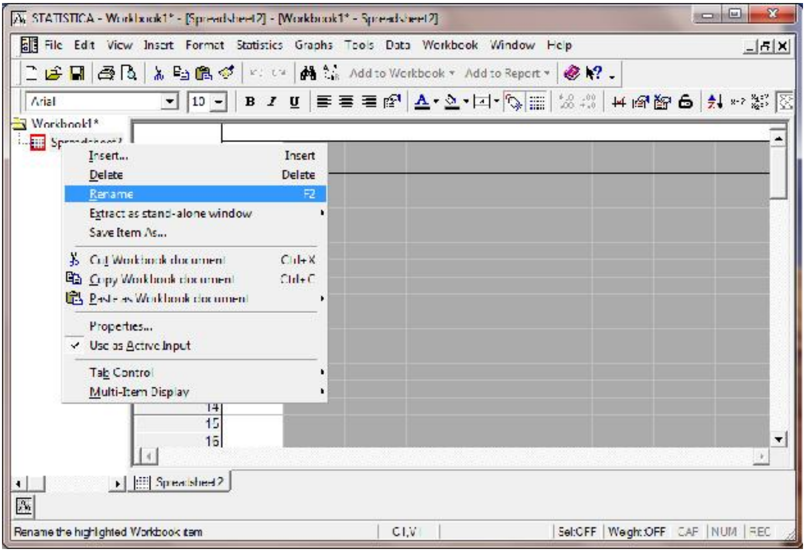
ზები მითითების გარეშე. ყველა აუცილებელი პირობის არჩევის შემდეგ ვაწვებით OK ღილაკს.

გამონათდება მუშა ფურცელი, რომელიც შეესაბამება მოცე-მულ პირობებს, ხოლო მისგან მარცხნივ წარმოდგენილია მუშა დავ-თარის ხე (ნახ. 3.2.). ჩუმათობის პრინციპით ახალ მუშა დავთარს Workbook1 ჰქვია. ამ მუშა დავთრის შიგნით მოთავსებული



ნახ. 3.2.

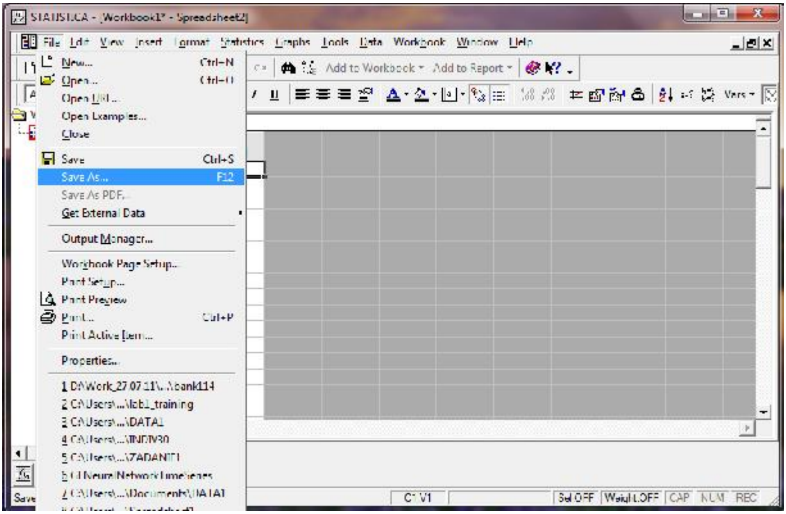
იქნება ყველა ის ელემენტი, რომლებიც შეიქმნა საწყისი მონაცემების საფუძველზე. მათზე შეიძლება ყველა ელემენტალური გარდაქმნის (დამახსოვრება, წაშლა, სახელის გადარქმევა) შესრულება. ამისათვის საკმარისია მუშა დავთრის ხეში მაუსის მარჯვენა ღილა-კის მათზე დაწკაპუნება და კონტექსტურ მენიუში შესაბამისი პუნქ-ტის ამორჩევა (ნახ. 3.3.).



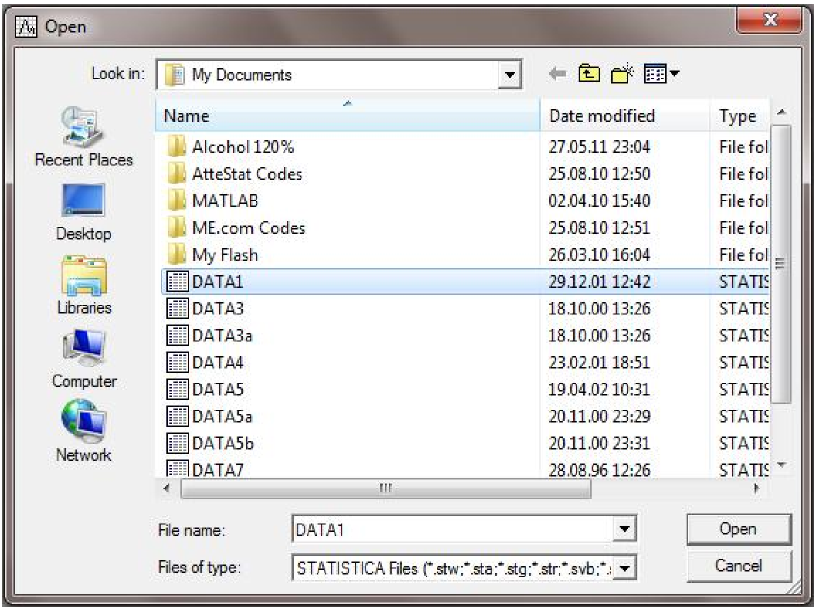
ნახ. 3.3.

შექმნილ მუშა დავთრებს უნდა გააჩნდეთ სტანდარტული და კონკრეტულ მომხმარებელთან ადვილად იდენტიფიცირებული სახელები. მუშა დავთრზე სახელის მისაკუთვნებლად ვირჩევთ მენიუს პუნქტს File და მასში ოპციას Save as (ნახ. 3.4.).

ყურადღებას ვამახვილებთ იმაზე, რომ მუშა დავთრისათვის დამახასიათებელია გაფართოება *.stw. STATISTICA-ს გაშვებისას იხსნება ფაილი, რომელიც გამოიყენებოდა ბოლო სეანსის დროს. სასურველი ფაილის გასახსნელად საჭიროა ვისარგებლოთ File/Open მენიუთი, ავირჩიოთ ჩვენთვის სასარგებლო საქაღალდე და ფაილი (ნახ. 3.4, 3.5.). ამასთან სისტემა გვთავაზობს ყველა შესაძლო მხარდაჭერილ ფორმატს (*.stw,*.sta,*.stg).


STATISTICA - Workbook1* - Spreadsheet2
File Edit View Insert Format Statistics Graphs Tools Data Workbook Window Help
New... Ctrl+N
Open... Ctrl+O
Open URL...
Open Examples...
Close
Save Ctrl+S
Save As... F12
Save As PDF...
Get External Data
Output Manager...
Workbook Page Setup...
Print Setup...
Print Preview
Print... Ctrl+P
Print Active Item...
Properties...
Add to Workbook
Add to Report
Sel OFF
Weight OFF
CAP
NUM
REC

ნახ. 3.4.


Open
Look in:
My Documents
Recent Places
Desktop
Libraries
Computer
Network
Name
Date modified
Type
Alcohol 120% 27.05.11 23:04 File fol
AtteStat Codes 25.08.10 12:50 File fol
MATLAB 02.04.10 15:40 File fol
ME.com Codes 25.08.10 12:51 File fol
My Flash 26.03.10 16:04 File fol
DATA1 29.12.01 12:42 STATIS
DATA3 18.10.00 13:26 STATIS
DATA3a 18.10.00 13:26 STATIS
DATA4 23.02.01 18:51 STATIS
DATA5 19.04.02 10:31 STATIS
DATA5a 20.11.00 23:29 STATIS
DATA5b 20.11.00 23:31 STATIS
DATA7 28.08.96 12:26 STATIS
File name:
DATA1
Files of type:
STATISTICA Files (*.stw;*.sta;*.stg;*.str;*.svb;*.
Open
Cancel

ნახ. 3.5.

3.2. მონაცემთა იმპორტი STATISTICA-ს გარემოში

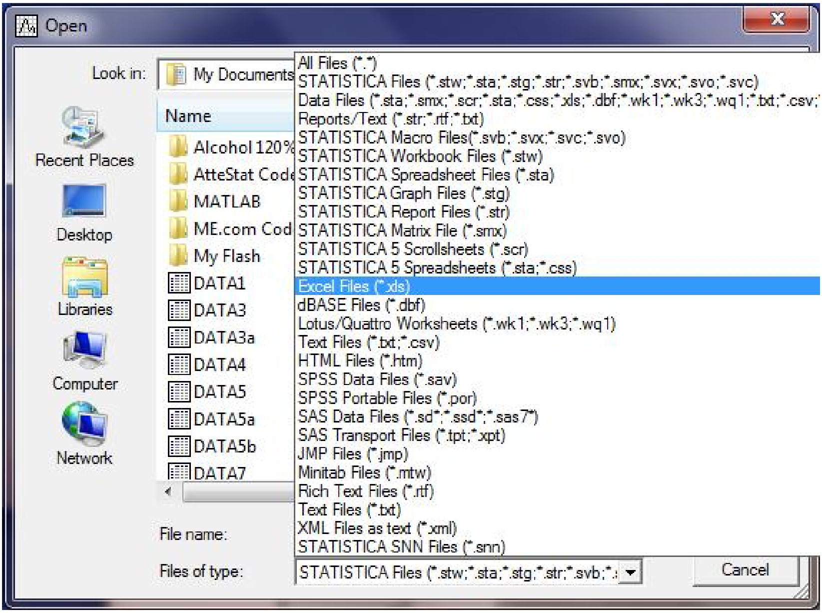
როგორც უკვე ავღნიშნეთ, STATISTICA ურთიერთმოქმედებს სხვა Windows-დამატებებთან. ეს საშუალებას იძლევა წარმატებით მოხდეს მონაცემების იმპორტირება და ექსპორტირება. არსებობს STATISTICA-ში სხვა დამატებებიდან მონაცემების შეტანის რამდენიმე ხერხი:

1) მონაცემთა კოპირების ოპერაციების შესრულება Windows-ის გაცვლის ბუფერით;
2) DDE დინამიური კავშირის მექანიზმის გამოყენება სისტემა STATISTICA-ს მონაცემებსა და სხვა დამატებების მონაცემებს შორის;
3) მონაცემების იმპორტი ყველაზე პოპულარული დამატებებიდან.

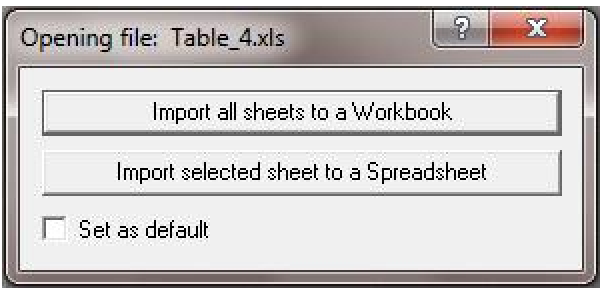
განვიხილოთ მონაცემების იმპორტის მაგალითი MS Excel-იდან. ამისათვის ავირჩიოთ მენიუს პუნქტი File და მასში ბრძანება Open. ავტომატურად სისტემა გვთავაზობს ავირჩიოთ STATISTICA-ს ფაილები (*.sta, *.stw, და ა.შ.). იმისათვის, რომ მოვახდინოთ მონაცემების იმპორტირება Excel-იდან, საჭიროა ველში ფაილების ტიპი ავირჩიოთ Excel Files (*.xls), სადაც xls - ფაილის გაფართოებაა (ნახ. 3.6).

საჭირო ფაილის შერჩევის შემდეგ ჩნდება დიალოგური ფანჯარა, რომელიც გვთავაზობს ან ყველა არსებული მონაცემის იმპორ-

ტირებას Excel-ის ფაილიდან მუშა დავთარში (Import all sheets to a Workbook), ან ირჩევენ მკლევარისათვის საჭირო მუშა ფურცლებს STATISTICA-ს ელექტრონულ ცხრილში (Import selected sheet to a Spreadsheet) (ნახ. 3.7.).

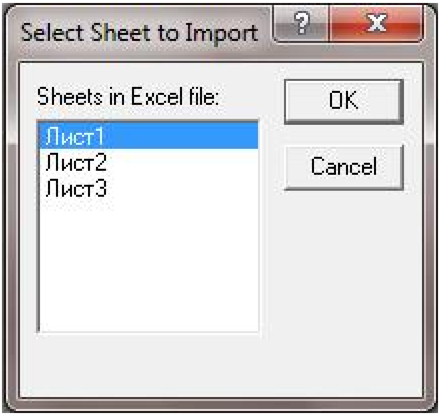


ნახ. 3.6.



ნახ. 3.7.

მაგალითისათვის ავირჩიოთ მეორე ვარიანტი, ამის შემდეგ ჩნდება დიალოგური ფანჯარა, რომელშიც შეიძლება მონაცემების იმპორტისათვის ამოვირჩიოთ Excel-ის ფაილში არსებული მუშა ფურცლები (ნახ. 3.8.).

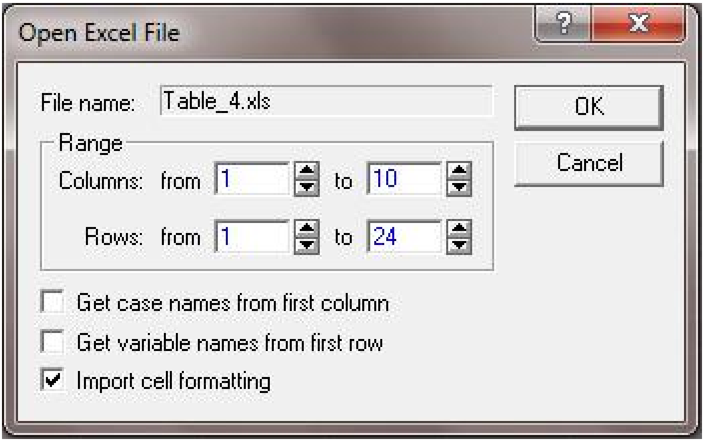


ნახ. 3.8.

საჭირო მუშა ფურცლის არჩევის შემდეგ , სისტემა მომხმარებელს სთავაზობს მონაცემების იმპორტის პარამეტრების დაზუსტებას, კერძოდ:

- იმპორტისათვის სვეტებისა და სტრიქონების ამორჩევა მათი ნომრების მითითებით (Columns: from… to; Rous: from … to);
- ალმების არსებობა Get variable names from first row და Get case names from first column ველებში ნიშნავს, რომ ცვლადებსა და დაკვირვებებს STATISTICA-ს დოკუმენტში მიენიჭებათ სახელები შესაბამისად Excel-ის დოკუმენტის პირველი სტრიქონიდან და პირველი სვეტიდან.

- დაიმახსოვრეთ Excel-ში არსებული უჯრედთა ფორმატი (Import cell formatting). პარამეტრების არჩევის შემდეგ დააწქით ღილაკს OK (ნახ. 3.9.).



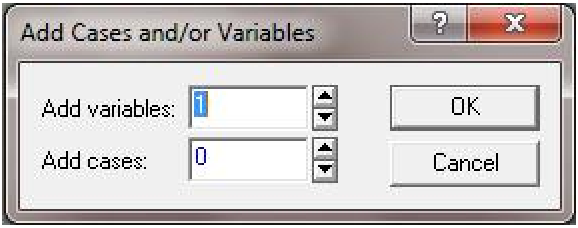
ნახ. 3.9.

მონაცემების შეტანა ხდება სტანდარტულად და დამატებით კომენტარებს არ საჭიროებს (რიცხვის ჩაწერის შემდეგ აწვებით Enter კლავიშას, ათობითი წილადები შეგაქვთ წერტილით).

3.3. ძირითადი ოპერაციები სვეტებთან და სტრიქონებთან

ისევე როგორც მუშა დავთრის ელემენტები, სტრიქონები და სვეტები სისტემის მუშა ფურცელზე შეიძლება დავამატოთ, გადავაადგილოთ, წავშალოთ, გადავაკოპიროთ.

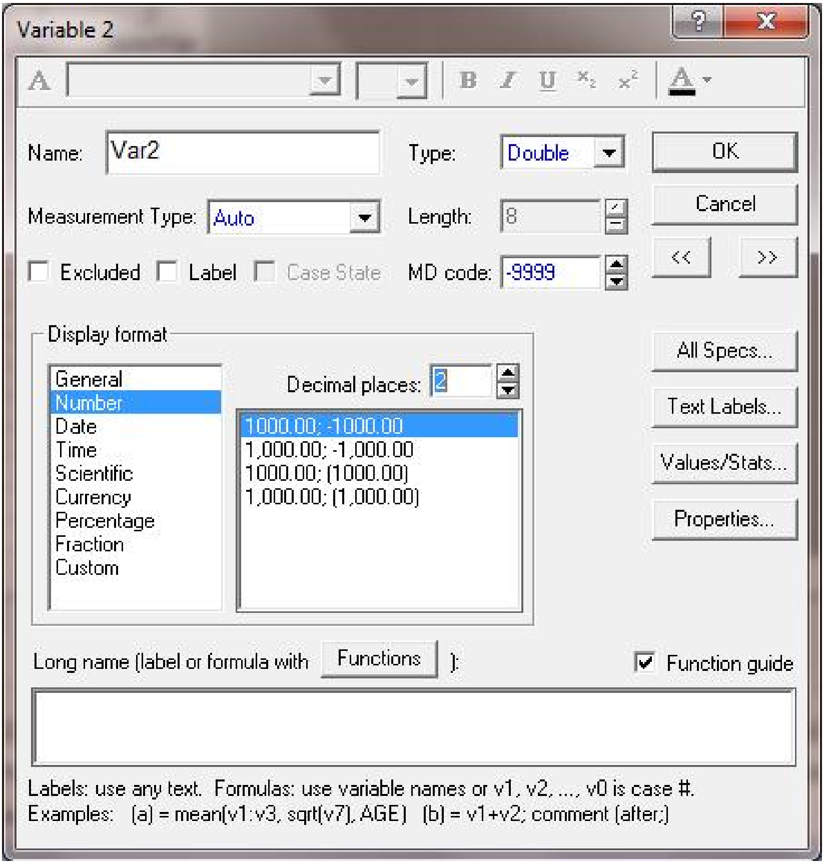
ახალი ცვლადის შექმნის სტანდარტული პროცედურა შემდეგში მდგომარეობს: საჭიროა ორჯერ დავაწკაპუნოთ მაუსის მარცხენა ღილაკი არსებული ცვლადის სათაურიდან მარჯვნივ მდებარე რუხ ველში. წარმოჩნდება დიალოგური ფანჯარა (ნახ. 3.10.), რომელიც გვთავაზობს ცვლადების (Add Variables) ან დაკვირვებების (Add Cases) დამატებას.



ნახ. 3.10.

მაგალითად, დავუმატოთ ერთი ცვლადი, ამისთვის Add Variables ველში ჩავწეროთ ციფრი 1. დავაწკაპუნოთ OK-ზე. ეკრანზე გამონათდება პანელი ახალი ცვლადის პარამეტრების ამოსარჩევად (ნახ. 3.11.).

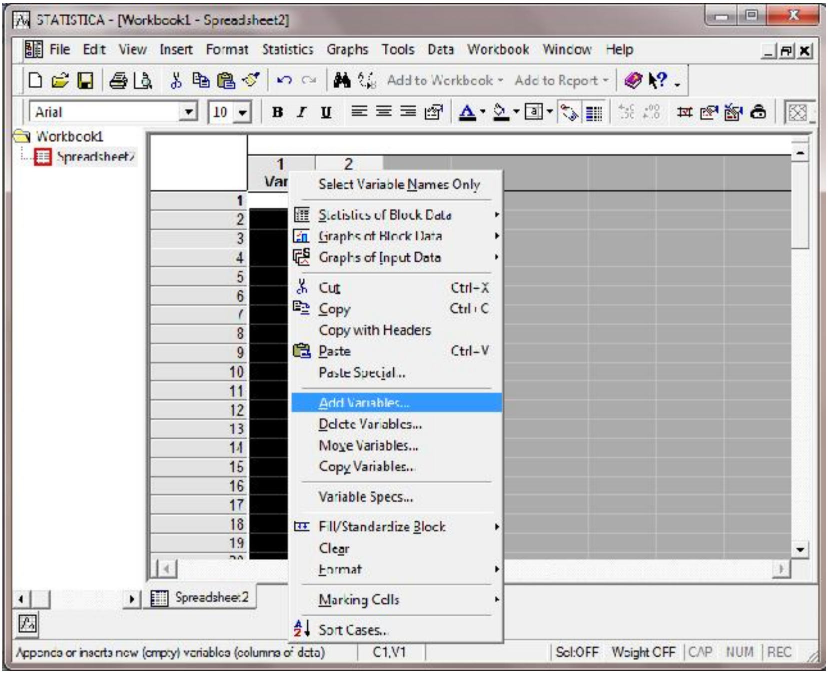
გამონათებულ ფანჯარაში შეგიძლიათ მიუთითოთ ახალი ცვლადის სახელი, მაგალითად Var2. ფანჯრის ზედა მარჯვენა კუთხეში Cansel ღილაკის ქვემოთ მოთავსებული ღილაკები „>>“ და „<<“ საშუალებას გვაძლევენ გადავადგილდეთ ცვლადებში და მოვმართოთ მათი ფუნქციები. მომავალში, მოცემული ფანჯრის გამოძახების აუცილებლობისას, საჭიროა ორჯერ დააწკაპუნოთ მაუსის მარცხენა ღილაკი ნებისმიერი არსებული ცვლადის სათაურზე.



ნახ. 3.11.

ველში Display Format აგრეთვე შეგვიძლია მივუთითოთ ცვლადის ფორმატი, ჩუმათობის პრინციპით ის ჩვეულებრივია (General), მაგრამ ჩვენ გვაინტერესებს რიცხვითი ფორმატი (Number). ამ ფორმატის გამოყენება საშუალებას გვაძლევს დავაყენოთ გამოთვლების სასურველი სიზუსტე, ანუ მძიმის შემდეგ ათობით ნიშანთა რაოდენობა (ველი Decimal places) (იხ. ნახ. 3.11.).

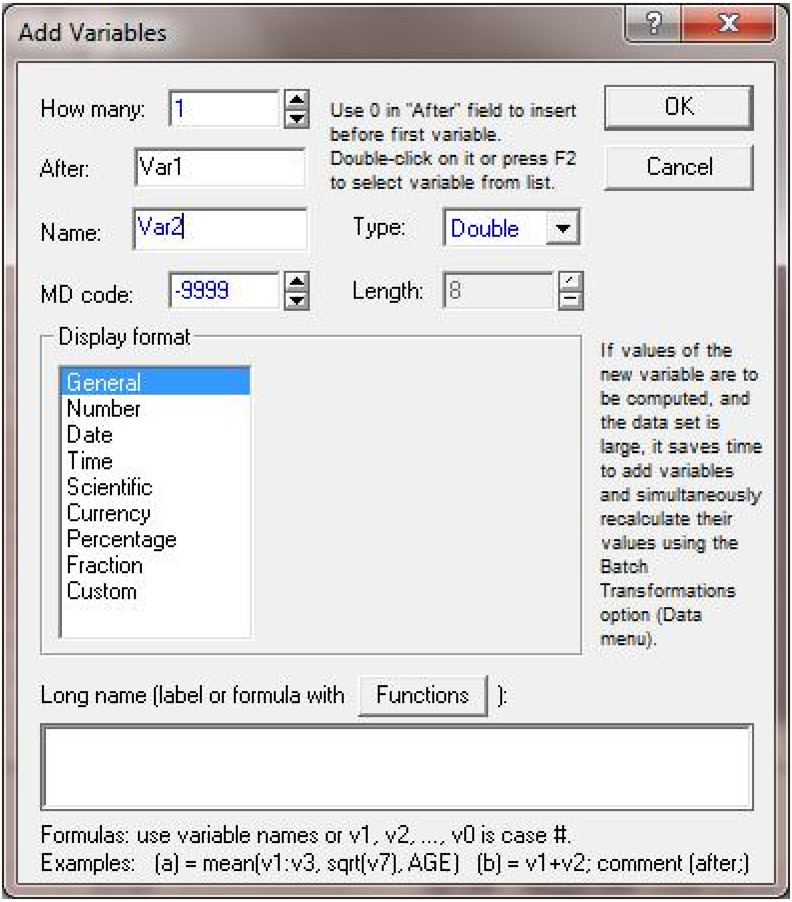
როცა მუშა ფურცელზე რამდენიმე ცვლადი იმყოფება, მაშინ შეგვიძლია დავუმატოთ ახლები მათი დამატების ადგილის მითითებით. ამისათვის აუცილებელია დავაწკაპუნოთ მაუსის მარჯვენა ღილაკი ნებისმიერი სვეტის დასათაურებაზე და ავირჩიოთ ფუნქცია Add Variables (ნახ. 3.12.) (იგივე მენიუში შეიძლება ავირჩიოთ ცვლადების გადაადგილების, წაშლისა და კოპირების ფუნქციები, რომლებიც ანალოგიურად სრულდებიან).



ნახ. 3.12.

გამონათებულ ფანჯარაში გარდა ჩვენთვის უკვე ცნობილი პროცედურებისა არსებობს ველი After, რომელშიც ეთითება იმ

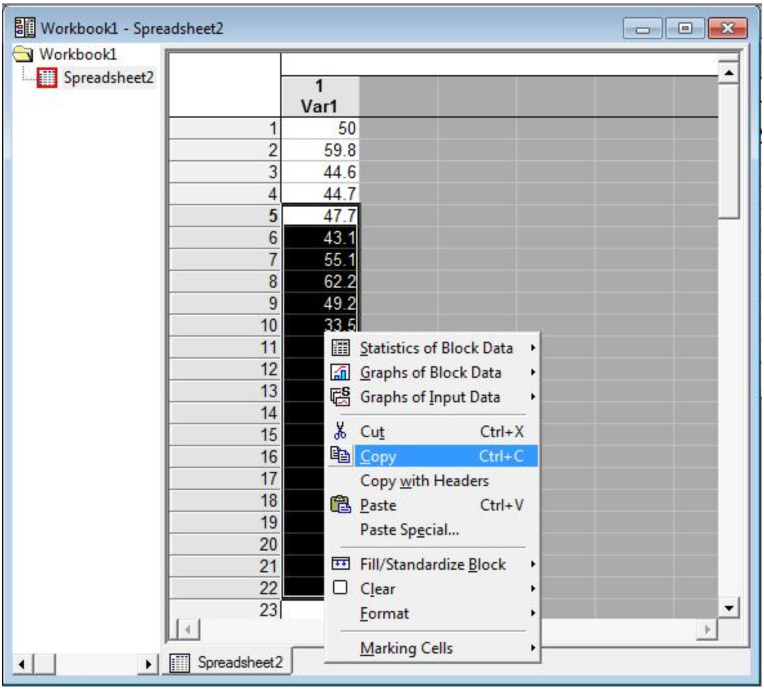
ცვლადის სახელი, რომლის შემდეგაც ჩვენ გვსურს მოვათავსოთ ახლად შექმნილი ცვლადი (ნახ. 3.13.).



ნახ. 3.13.

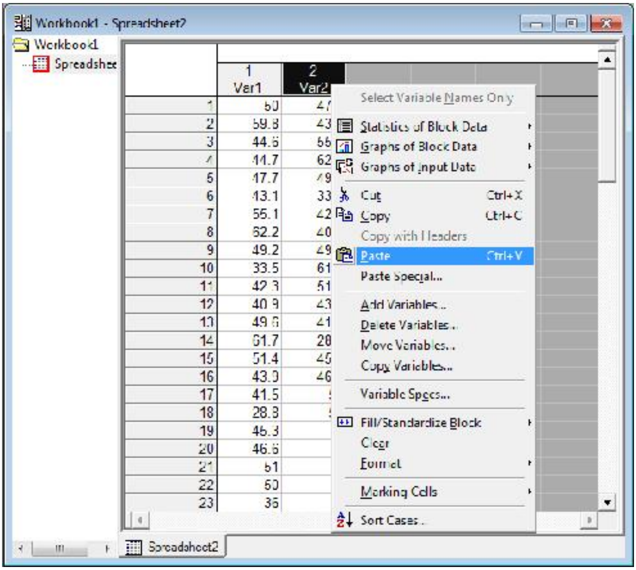
სხვა ფუნქციების არჩევისას ამ მენიუში აგრეთვე მოგეთხოვებათ ცვლადების სახელების მითითება, მაგალითად რომელი ცვლადიდან რომელამდე გჭირდებათ სვეტების წაშლა.

მიმდინარე მუშა ფურცლის, სხვა მუშა ფურცლების, Excel-ის დოკუმენტების სვეტებიდან მონაცემები შეიძლება ერთნაირად გადავაკოპიროთ. ამისათვის გამოვყოფთ ჩვენთვის საინტერესო ზონას, შემდეგ მაუსის მარჯვენა ღილაკზე დაწკაპუნების შემდეგ გამონათებულ ქვემენიუში ვირჩევთ ფუნქციას Copy (ნახ. 3.14.).



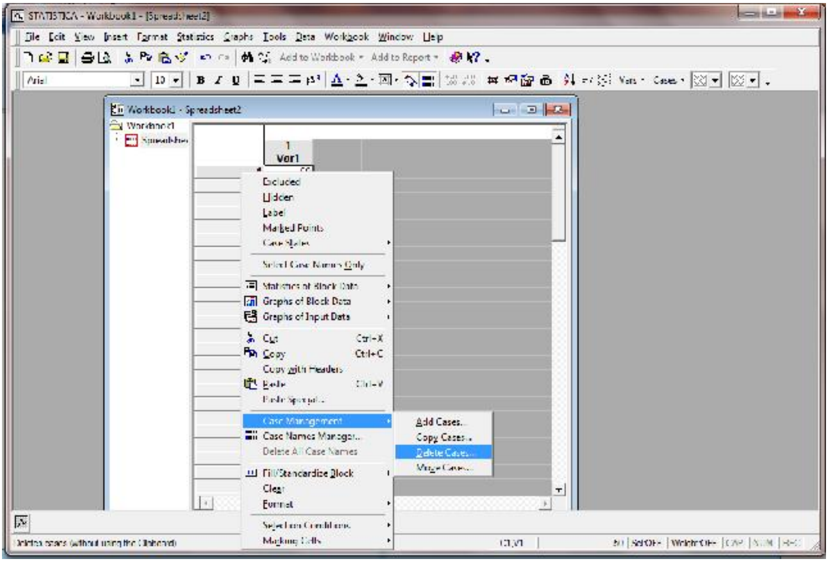
ნახ. 3.14.

შემდეგ ვაწკაპუნებთ მაუსის მარჯვენა ღილაკს ახალი ცვლადის დასათაურებაზე და გამონათებულ მენიუში ვირჩევთ ფუნქციას Paste (ჩასმა) (ნახ. 3.15.).

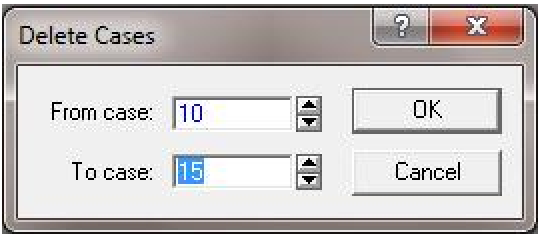


ნახ. 3.15.

ყველა ეს ფუნქცია სამართლიანია სტრიქონებისთვისაც. მაგალითად, თუ გვსურს წავშალოთ სტრიქონები 10-დან 15-მდე საჭიროა გავაწკაპუნოთ მაუსის მარჯვენა ღილაკი სტრიქონის ნებისმიერ ნომერზე და ავირჩიოთ ფუნქცია Case Management/ Delete cases (ნახ. 3.16.). შემდეგ გამონათებულ ფანჯარაში ველში From case ვსვამთ იმ სტრიქონის ნომერს, რომლიდანაც დაწყებული გვსურს მონაცემების წაშლა; ველში To case ვსვამთ იმ სტრიქონის ნომერს, რომლის ჩათლითაც საჭიროა მონაცემების წაშლა (ნახ. 3.17.).

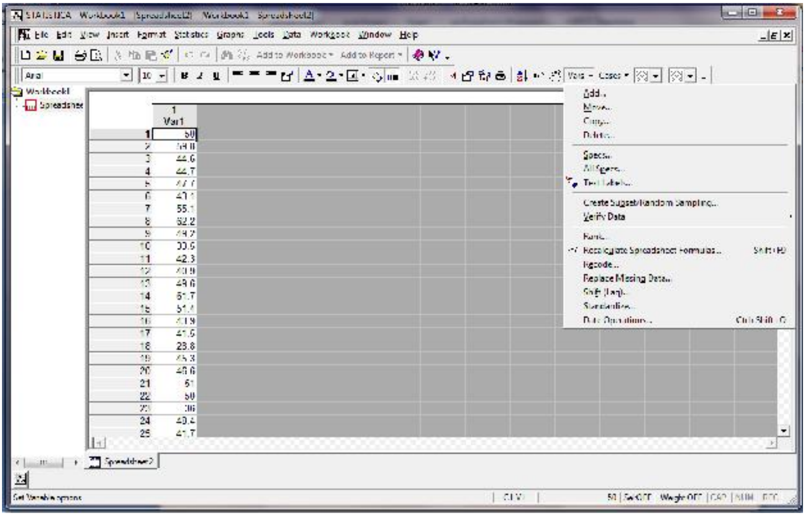


ნახ. 3.16.



ნახ. 3.17.

ყველა ფუნქცია ცვლადებითა და დაკვირვებებით შეიძლება გავუშვათ ღილაკების დახმარებით: Vars და Cases შესაბამისად ინსტ-რუმენტთა პანელიდან (ნახ. 3.18.).



ნახ. 3.18.

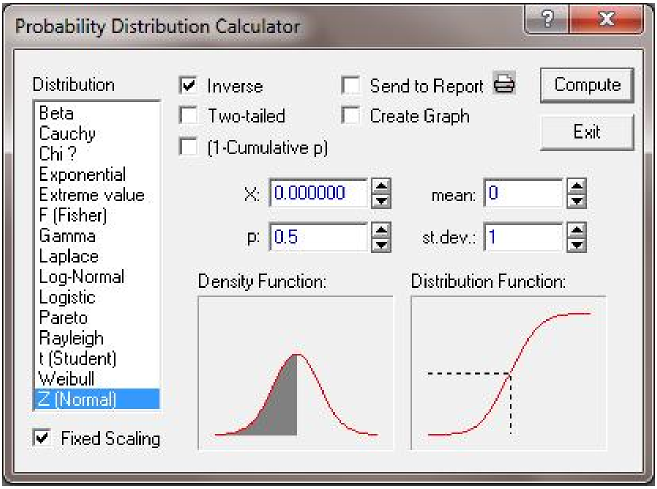
4. ალბათური კალკულატორი
4.1.მოდელური განაწილებების გეომეტრიული არსის კვლევა და ცხრილების აგება

ალბათური კალკულატორის (Probabiliti Calculator) გაშვება ხორციელდება Basic Statistics and Tables (ძირითადი სტატისტიკები და ცხრილები) მოდულის სასტარტო პანელიდან.

მაგალითი 4.1. ნორმალური განაწილების N(a; σ) გეომეტრიული არსის გამოკვლევა.

დავუშვათ a=0, σ=1. ფანჯარაში Probabiliti Calculator ველში Distribution მაუსის დახმარებით აღნიშნეთ სტრიქონი Z(Normal), შე-ავსეთ ველები: mean:0, sd.dev.:1, p:0,5. დააყენეთ ალამი Fixed Scaling, შემ-დეგ დააწქით ღილაკს Compute.

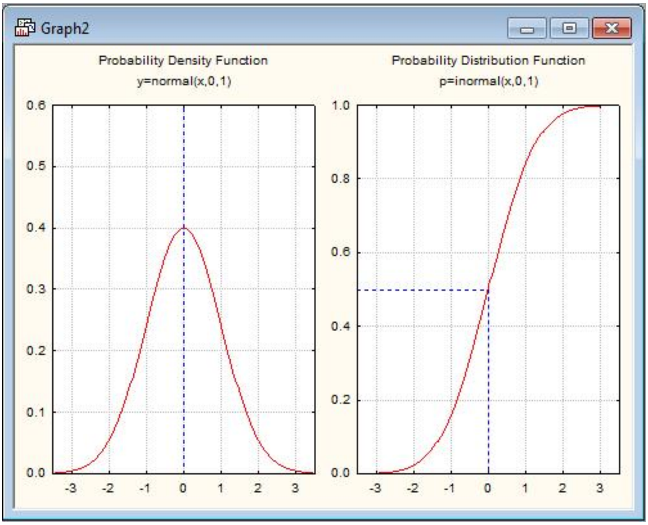
გახსნილი ფანჯრის X ველში წარმოჩნდება მნიშვნელობა 0.00000. ეს ნორმალური განაწილების 0.5 ქვანტილია, ანუ F(Z)=0.5 გან-ტოლების ფესვი. ველში Density Function გამოისახება განაწილების მრუდი დაშტრიხული არით (ნახ. 4.1.).



ნახ. 4.1.

მონიშნული არის ფართობი ტოლია მითითებული p=0,5 მნიშვ-ნელობისა. შემდეგ აღნიშნეთ ველი Create Graph და დააწქით ღილაკს Compute. ეკრანზე გამონათდება სიმკვრივის გრაფიკი ქვანტილის აღნიშნული წყვეტილი ხაზით (ნახ. 4.2.). გრაფიკიდან ჩანს, რომ 0.5- ქვანტილი წარმოადგენს ნორმალური განაწილების მოდასა და მედიანას. ბრძანებების მოყვანილი თანმიმდევრობების გამე-ორებით mean-ის სხვადასხვა მნიშვნელობებისათვის (a=1; 2; -2; ...) დარწმუნდით, რომ მნიშვნელობა ნორმალური განაწილების

სიმკვრივის ფუნქციის მაქსიმუმის წერტილია. (ნორმალური განაწილების სიმკვრივის გრაფიკი ინაცვლებს ორდინატთა ღერძზე საშუალოს ცვლილების დროს. საშუალოს ზრდისას გრაფიკები მარჯვნივ ინაცვლებენ). ნორმალური განაწილების სიმკვრივის პიკი მოთავსებულია წერტილში ორდინატით, რომელიც საშუალო მნიშვნელობის ტოლია. ეს მნიშვნელობა ეთითება ველში mean (საშუალო).



ნახ. 4.2.

sd.dev.(σ) ველის მნიშვნელობის ცლილებით a-ს და p-ს მუდმივი მნიშვნელობების დროს დარწმუნდით, რომ σ-ს ზრდისას ნორმალური განაწილების სიმკვრივე ნაწილდება (ბლაგვდება) a-ს მიმართ, ხოლო f_{max} მცირდება. σ-ს შემცირებისას სიმკვრივე იკუმშება, კონცენტრირდება მაქსიმუმის წერტილთან, f_{max} იზრდება.

მაგალითი 4.2. ნორმალური განაწილების მქონე ς შემთხვევითი სიდიდის, რომლის პარამეტრებია a=176,6, σ=7,63 ალბათობის P(175<ς<185) გამოთვლა.

ალბათური კალკულატორის (Probability Distribution Calculator) ფანჯარაში შეავსეთ ველები: Distribution: Z(Normal), : mean:176,6; sd.dev.:7,63; X: 185 . შემდეგ დააჭირეთ ღილაკს Compute. ველში P გამოვა მნიშვნელობა: 0.891022 . დაიმახსოვრეთ ის.

შეცვალეთ X-ს მნიშვნელობა 175-ზე, დააჭირეთ ღილაკს Compute. დაიმახსოვრეთ ველის ახალი მნიშვნელობა p:0.468661. გამოითვალეთ $P(175<\varsigma<185)= 0.891022-0.468661=0.422361\approx 0.4$.

2 და 3 „სიგმას" წესები.

ვთქვათ, გვაქვს ნორმალური განაწილების მქონე შემთხვევითი სიდიდე ξ a–ს ტოლი მათემატიკური ლოდინით და σ^2 დისპერსიით. განვსაზღვროთ მისი მოხვედრა $(a - 3\sigma;\ a + 3\sigma)$ ინტერვალში, ანუ იმის ალბათობა, რომ ξ ღებულობს მნიშვნელობებს, რომლებიც მათემატიკური ლოდინიდან განსხვავდებიან არა უმეტეს, ვიდრე სამი საშუალო კვადრატული გადახრა.

$$P(a - 3\sigma< \xi < a + 3\sigma)=\Phi (3) - \Phi (-3)=2\Phi (3)$$

ცხრილში ვპოულობთ $\Phi (3)=0,49865$, საიდანაც გამოდის, რომ $2\Phi (3)$ პრაქტიკულად ერთის ტოლია. ამრიგად, შეგვიძლია გავაკეთოთ მნიშვნელოვანი დასკვნა: ნორმალური შემთხვევითი სიდიდე ღებულობს მნიშვნელობებს, რომლებიც იხრებიან მისი მათემატიკური ლოდინიდან არა უმეტეს 3σ–ისა.

აქ რიცხვი 3 არჩევა პირობითია. შეგვეძლო აგვეღო 2,8, 2,9 ან 3,2 და მიგვეღო იგივე ალბათური შედეგი. იმის გათვალისწინებით, რომ $\Phi (2)=0,477$ შეგვეძლო გვესაუბრა 2 „სიგმას" წესზე.

თუ ნორმალური განაწილების სიმკვრივის საშუალოს წერტილიდან ან მაქსიმუმის წერტილიდან მარცხნივ და მარჯვნივ შესაბამისად გადავდებთ ორი ან სამი სტანდარტული გადახრის (2 და 3 „სიგმა") ტოლ მონაკვეთებს, მაშინ ნორმალური სიმკვრივის გრაფიკის ქვემოთ მოთავსებული ფართობი ტოლი იქნება 95,45% და 99,73%-ისა გრაფიკის ქვეშ მოქცეული მთელი ფართობისა. (ანუ ყველა დამოუკიდებელი დაკვირვებების 95,45% და 99,73% მოთავსებულებია

საშუალო მნიშვნელობიდან 2 და 3-ის ტოლი სტანდარტული გადახრების ტოლ რადიუსში).

მაგალითი 4.3. 2 და 3 „სიგმას“ წესების შემოწმება. შევამოწმოთ, რომ თუ $X\sim N(a;\sigma)$, მაშინ $P(|X-a|<2\sigma)=0.9545$, $P(|X-a|<3\sigma)=0.9973$ დამოუკიდებელია a და σ -ს მნიშვნელობებიდან.

ალბათური კალკულატორის ფანჯარაში (Probability Distribution Calculator) ველში Distribution: გამოყავით Z(Normal).

მონიშნეთ ოპცია Two-tailed (ორმხრივი), ვინაიდან უტოლობა მოდულით ორმხრივია. მიუთითეთ mean:0, sd.dev.:1. ვინაიდან $2\sigma=2$, ამიტომ ველში X ჩასვით 2, დააწქეთ ღილაკს Compute.

სტრიქონში p გამონათდება რიცხვი 0.954500. ველში Density Function (სიმკვრივის ფუნქცია) სიმკვრივის გრაფიკის ქვემოთ მოთავსებული დაშტრიხული ფართობი შეადგენს მთელი ფართობის 95,45%-ს. შეასრულეთ იგივე 3σ-სთვის. დარწმუნდით, რომ დაშტრიხული ფართობი მიაღწევს 99,73%-ს.

a და σ-ს სხვადასხვა მნიშვნელობების ჩასმით დარწმუნდით, რომ 2 და 3 „სიგმას“ წესებს ადგილი აქვს ნორმალური განაწილების ნებისმიერი მნიშვნელობებისათვის.

მაგალითი 4.4. გამოვითვალოთ χ^2 განაწილების 0.95 და 0.99 ქვანტილები თავისუფლების 7 ხარისხით. გამოვიკვლიოთ თავისუფლების ხარისხის გავლენა განაწილების მრუდის ფორმასა და განლაგებაზე.

ალბათური კალკულატორის ფანჯარაში (Probability Distribution Calculator) ველში Distribution: გამოყავით Chi I სტრიქონი. შეავსეთ ველები df:7, p:0,95 და დააწქეთ ღილაკს Compute. ველში Chi I გამოვა რიცხვი: 14.068419. ეს 95%-იანი წერტილია (.95 - ქვანტილი), ანუ F(I)=0.95 განტოლების ფესვია. ეს ნიშნავს, რომ $P(\chi^2\leq 14{,}068419)=0.95$. საწინააღმდეგო უტოლობის ალბათობის გამოსათვლელად დააყენეთ ალამი ველში (1-Cumulative p) და დააწქეთ ღილაკს Compute.

შეცვალეთ p ველის მნიშვნელობა 0.99-ზე. დააწექით ღილაკს Compute. ველში Chi I გამონათდება რიცხვი 18,477779. ეს 99%-იანი წერტილია (0.99 - ქვანტილია). აირჩიეთ ოპცია Create Graph. დააწექით ღილაკს Compute. ააგეთ სიმკვრივის გრაფიკი და Chi-კვადრატ ფუნქციის განაწილება თავისუფლების 7 ხარისხით. k პარამეტრის სხვადასხვა მნიშვნელობების (2; 5; 12; ...) df ველში ცვლილებით დარწმუნდით, რომ k-ს ზრდისას განაწილების სიმკვრივის პიკი ქვემოთ და მარჯვნივ ეშვება. სიმკვრივის გრაფიკი ხდება უფრო სიმეტრიული, ფორმით ემსგავსება გაუსის მრუდს.

მაგალითი 4.5. თავისუფლების ხარისხის ზემოქმედების გავლენა სტიუდენტის განაწილების მრუდის ფორმასა და განლაგებაზე.

ველში Distribution: გამოყავით სტრიქონი t (Student). შეავსეთ ველები: df: 5, p: 0.5. t ველს სისტემა შეავსებს რიცხვით 0. აღნიშნეთ ოპცია Create Graph, შემდეგ დააწექით ღილაკს Compute. დაათვალიერეთ გრაფიკი და გაიმეორეთ ალგორითმი df=10, 35, 50, 100-ისათვის. დარწმუნდით იმაში, რომ t-განაწილების სიმკვრივის გრაფიკი სიმეტრიულია Oy ღერძის მიმართ და გვაგონებს გაუსის მრუდს. თავისუფლების k ხარისხის ზრდასთან ერთად სიმკვრივის მაქსიმალური მნიშვნელობა იზრდება, ფერდები უფრო ციცაბოდ მცირდებიან 0-საკენ.

ველში p 0,5; 0,7; 0,95; 0,99 მნიშვნელობების შეტანით შეადგინეთ t-განაწილების ფუნქციის მნიშვნელობათა ცხრილი თავისუფლების ხარისხით 10 (ქვანტილთა ცხრილი).

t	0	0.54	1.812460	2.763770
F(t)	0.5	0.7	0.95	0.99

პირიქით, შეიტანეთ t ველში მნიშვნელობა 1. სისტემა გამოითვლის p: 0.829553. გამომდინარე, P(t <1) =0.829553. დააყენეთ ალამი (1-Cumulative p). p ველის მნიშვნელობა შეიცვლება 0.170447-ზე. კალკულატორმა გამოითვალა საწინააღმდეგო შემთხვევის ალბათობა: P(t ≥1) =0.170447.

მაგალითი 4.6. ფიშერის განაწილება. ალბათური კალკულატორის დახმარებით დარწმუნდით, რომ F-განაწილება თავმოყრილია დადებით ნახევარღერძზე. განსაზღვრეთ $F_{10,10}$ -განაწილების 0.5 და 0.75 ქვანტილები. გამოითვალეთ $P(F_{10,10} \leq 1)$ და $P(F_{10,10} \leq 2)$ ალბათობები.

ველში Distribution: გამოყავით სტრიქონი F. შეავსეთ ველები: p: 0.5; df1: 10; df2: 10, შემდეგ დააწექით ღილაკს Compute. კალკულატორი გამოითვლის ველის მნიშვნელობას F: 1. შეცვალეთ p: ველის მნიშვნელობა 0.75-ზე. F: ველის მნიშვნელობა შეიცვლება 1,551256-ზე. შეცვალეთ p: ველის მნიშვნელობა 2-ზე, შემდეგ 1-ზე. კალკულატორი გამოითვლის ალბათობებს $P(F_{10,10} \leq 2)=0,144846$ და $P(F_{10,10} \leq 1)=0,5$.

df1 და df2-ზე სხვადასხვა მნიშვნელობების მინიჭებისას დააკვირდით გრაფიკებს. ყურადღება მიაქციეთ იმას, რომ ნორმალურისაგან განსხვავებით F-განაწილების მრუდი არასიმეტრიულია თავისუფლების ხარისხის მცირე მნიშვნელობების დროს (n და k<30). n და k-ს ზრდისას F-განაწილება ნელა უახლოვდება ნორმალურ მრუდს.

ს ა ვ ა რ ჯ ი შ ო ე ბ ი

ააგეთ სტიუდენტის განაწილების სიმკვრივის გრაფიკი თავისუფლების 5 ხარისხის დროს. იპოვეთ t-მნიშვნელობა p: 0.95 დონის დროს. ააგეთ სტიუდენტის განაწილების გრაფიკი თავისუფლების 25 ხარისხის დროს. გრაფიკულად შეადარეთ სტიუდენტის განაწილების სიმკვრივე სტანდარტული ნორმალური განაწილების სიმკვრივეს.

დავალება 4.1. სამუშაოსათვის.

ალბათური კალკულატორის დახმარებით გადაწყვიტეთ შემდეგი ამოცანები.

1. ამოცანა გულივერებისა და ლილიპუტების შესახებ.

წარმოიდგინეთ, რომ მოხვდით ქვეყანაში, სადაც მოზრდილი მამაკაცების სიმაღლეს მიახლოებით აქვს ნორმალური განაწილება 176,6 სმ საშუალოთი და სტანდარტული გადახრით 7,63 სმ. რაა იმის ალბათობა, რომ შემთხვევით არჩეული მამაკაცის სიმაღლე 195 სმ-ზე მეტი იქნება, ანუ ის გულივერი იქნება?

რაა იმის ალბათობა, რომ შემთხვევით არჩეული მამაკაცის სიმაღლე 155 სმ-ზე ნაკლები იქნება, ანუ ის ლილიპუტი იქნება?

2. ნორმალური განაწილებისათვის არჩეული პარამეტრებით გამოითვალეთ იმ ინტერვალში მოხვედრის ალბათობა, რომელიც შეიცავს mean-ს და არ შეიცავს mean-ს.
3. შეადგინეთ ნორმალური,Chi-კვადრატ, სტიუდენტის და ფიშერის განაწილებათა ცხრილები (10-10 მნიშვნელობით). გამოითვალეთ მოდელური განაწილებების 0,95 და 0,99 -ქვანტილები პარამეტრების სხვადასხვა მნიშვნელობებისათვის.
4. გააანალიზეთ განაწილების პარამეტრების ზემოქმედება სიმკრივის მრუდების ფორმაზე შემდეგი უწყვეტი განაწილებებისათვის: ექსპონენციალური, ნორმალური ფიშერის, სტიუდენტის, χ^2.

4.2.ბინომიალური განაწილება და თამაშთა ამოცანები

ბინომიალური განაწილების პარამეტრებს წარმოადგენს წარმატების ალბათობა p (q=1-p) და ცდების n რიცხვი. m-წარმატების ალბათობა n-ცდაში გამოითვლება შემდეგი ფორმულით:

$$p(m;n)=B(m;n)*p^m(1-p)^{n-m},\ m=0,1,\dots,n,\ B(m;n)=n!/((n-m)!*m!)$$

შექმენით ცარიელი ელექტრონული ცხრილი 1v*10c, ფაილს დაარქვით testsm.sta სახელი. VAR1 ცვლადს დაართქვით „albatoba“,

ქვედა ველში Long Name შეიტანეთ ცვლადის განმსაზღვრავი გამოსახულება: =Binom(v0, 0.3, 10), დააწექით OK -ის.

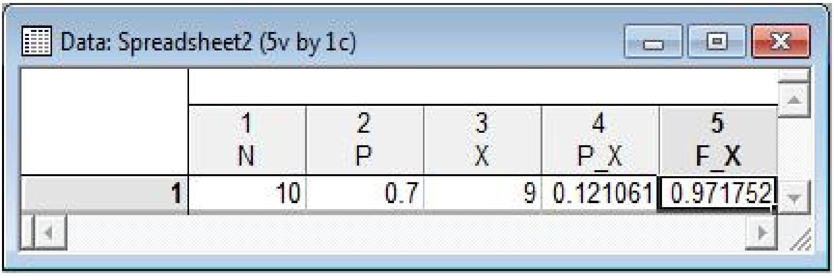
პროგრამა გამოითვლის წარმატების ალბათობას და შეიტანს მას პირველი ცვლადის მნიშვნელობად. მოცემულ ცხრილში გერბის მოსვლის ალბათობა 0,3-ის ტოლია. ცხრილიდან ჩანს, რომ 10 აგდებიდან ერთი გერბის მოსვლის ალბათობა 0.12106-ია, ხოლო 10 აგდებაში 2 გერბის მოსვლის ალბათობა 0.2334-ის ტოლია და ა. შ.

წარმატების ალბათობის შეცვლა ადვილია, მაგალითად თუ მას გავხდით 0,5-ის ტოლად. ეს ნიშნავს, რომ ვაგდებთ სიმეტრიულ მონეტას და წარმატების ალბათობა უდრის წარუმატებლობის ალბათობას. ველში Long Name საკმარისია შევცვალოთ ფორმულა - 0.3-ის ნაცვლად ჩავწეროთ 0.5.

თუ დაგავიწყდათ ბინომიალური ალბათობების გამოსათვლელი ფუნქცია შეგიძლიათ ისარგებლოთ საშუალებით Function Wizard. ცვლადის სპეციფიკაციის ფანჯარაში ღილაკ Functions-ზე დაწოლით გამონათდება დიალოგის ფანჯარა Function Wizard, რომელშიც ფანჯარაში Category აირჩიეთ Distributions, ფანჯარაში Name აირჩიეთ Binom. დააწექით Insert-ს. ბინომიალური განაწილების ფუნქცია გამონათდება ცვლადის სპეციფიკაციის ფანჯარაში ველში Long Name. დაგრჩათ მხოლოდ საჭირო პარამეტრების მითითება და გამოთვლაზე გაშვება. შემდგომში ჩვენ გამოვითვლით არა მხოლოდ ბინომიალურ ალბათობებს, არამედ ბინომიალურ კოეფიციენტებს B(m;n). ეს ადვილი შესასრულებელია წარმატების ალბათობის ბინომიალური ალბათობების p=1/2 გადამრავლებით 2 ხარისხში n-ზე.

შევასრულოთ ახლა გამოთვლები ბინომიალური განაწილებისათვის პარამეტრებით n=10 და p=0.7 წერტილში x=9. შევიტანოთ ცხრილში მოცემული სიდიდეები: N=10, P=0.7, X=9. მეოთხე სვეტის სპეციფიკაციის ფანჯარაში (სვეტი P_X), ველში Long Name შევიტანოთ ფორმულა ბინომიალური განაწილებისათვის =Binom(9; 0,7; 10), დააწექით OK-ის. ანალოგიურად მეხუთე სვეტის (F_X) სპეციფიკაცი

ის ფანჯარაში შევიტანოთ ფორმულა =IBINOM(9; 0,7; 10) სახის ბინომიალური განაწილების ფუნქციისათვის, დააწექით OK-ის.


Data: Spreadsheet2 (5v by 1c)

	1 N	2 P	3 X	4 P_X	5 F_X
1	10	0.7	9	0.121061	0.971752

ნახ. 4.3.

შედეგად მივიღებთ შემდეგ პასუხებს: P{X=9}=0.121; F(9)=0.972.

შევალიე დე მერეს ამოცანა

ერთხელ აზარტულმა მოთამაშემ იკითხა, უღირს თუ არა მას ჩამოვიდეს ფსონი ორი ექვსიანის ერთდროულად მოსვლაზე ორი კამათლის 24-ჯერ აგდების დროს?

შექმენით მუშა ფაილი play.sta. ორჯერ დააწკაპუნეთ მაუსის მარცხენა ღილაკი ცვლადის სახელზე და გახსენით var1 ცვლადის სპეციფიკაციის ფანჯარა. ველში Long Name ჩაწერეთ ფორმულა =Binom(v0,1/36,24), შემდეგ დააწექით OK-ის. პროგრამა გამოითვლის ბინომიალურ ალბათობებს. ამ ცხრილის პირველ სვეტში თანმიმდევრობითაა მოყვანილი ორი ექვსიანის ერთდროულად მოსვლის ალბათობები ერთხელ, ორჯერ, სამჯერ და ა.შ. ჩვენ გვჭირდება, უმცირესი, ერთი წყვილი ექვსიანების მოსვლის ალბათობის გამოთვლა. გამომდინარე, ყველა ეს ალბათობა უნდა შევკრიბოთ. ამრიგად, უმცირეს, ერთი წყვილი ექვსიანების მოსვლის ალბათობა კამათლების 24-ჯერ გაგორებისას 0.49140-ის ტოლია. თამაშის გრძელ სერიაში, რომელიც კამათლების 24 გაგორებიდან

შედგება, მოთამაშე, რომელიც ჩამოდის ფსონს ორი ექვსიანის ერთდროულად მოსვლაზე, საშუალოდ მდგრადად აგებს.

კითხვა: როგორ უნდა შევცვალოთ თამაშის პირობები, რომ მოგებული დავრჩეთ?

შევალიე დე მერეს შეცვლილი ამოცანა

დავუშვათ, რომ შევალიე დე მერემ დაიწყო ფსონის ჩამოსვლა წყვილი ექვსიანების მოსვლაზე 25 გაგორებიდან.

გაიმეორეთ წინა ამოცანის ყველა მოქმედება ცვლადისათვის var2. ველში Long Name ჩაწერეთ ფორმულა =Binom(v0,1/36,25), შემდეგ გააწკაპუნეთ OK-ზე. მეორე სვეტში მონაცემების შეკრებით, ადვილად შეგვიძლია ვიპოვოთ, რომ ორი ექვსიანის მოსვლის ალბათობა 25 აგდებიდან უმცირეს 0,5-ზე მეტია.

მოთამაშის კიდევ ერთი ამოცანა

ერთხელ ერთმა ინგლისელმა (ს. პეპაისმა) მიწერა ნიუტონს წერილი, რომელშიც რჩევას თხოვდა თუ რაზეა უკეთესი ფსონის ჩამოსვლა:

- ერთი ექვსიანის მოსვლაზე კამათლის 6-ჯერ აგდებისას?
- ორი ექვსიანის მოსვლაზე კამათლის 12-ჯერ აგდებისას?
- სამი ექვსიანის მოსვლაზე კამათლის 18-ჯერ აგდებისას?
- ოთხი ექვსიანის მოსვლაზე კამათლის 24-ჯერ აგდებისას?

ისევ გამოვიყენოთ ფაილი play.sta. გავზარდოთ მისი ზომები 14 შემთხვევის დამატებით. (Cases – Add – 14. After case: 10) – OK. დავიწყოთ პირველი კითხვით. ჩავწეროთ ბინომიალური ალბათობა პირველი კითხვისათვის var1 ცვლადის შემთხვევაში. ველში Long

Name ჩაწერეთ ფორმულა =Binom(v0,1/6,6), შემდეგ დააწექით OK-ის. შემდეგ იგივე შეასრულეთ var2, var3, var4 ცვლადებისათვის, შესაბამისი ალბათობების ჩასმით მეორე, მესამე და მეოთხე შეკითხვებისთვის.

მოცემულ ფაილში i ნომრის მქონე სტრიქონში მოცემულია i-ური ექვსიანის ამოსვლის ალბათობა პირველ, მეორე, მესამე და მეოთხე შეკითხვებში. სვეტებში ალბათობათა მნიშვნელობების აჯამვით მიიღებთ:

- 0.665 პირველი შემთხვევისათვის;
- 0.619 მეორე შემთხვევისათვის;
- 0.597 მესამე შემთხვევისათვის;
- 0.584 მეოთხე შემთხვევისათვის.

5. კორელაციური ანალიზი

5.1.ზოგადი ცნობები

კორელაციური მახასიათებლების შეფასების მეთოდების ერთობლიობასა და შერჩევითი მონაცემებით მათ შესახებ სტატისტიკური ჰიპოთეზების შემოწმებას კ ო რ ე ლ ა ც ი უ რ ა ნ ა ლ ი ზ ს უწოდებენ. კორელაციური ანალიზის მთავარი ამოცანაა ცვლად სიდიდეთა შორის ურთიერთკავშირთა შეფასება შერჩევითი მონაცემების საფუძველზე.

კორელაციურ ანალიზში იყენებენ შემდეგ ძირითადი ხერხებს:

1) კორელაციური ველის აგება (გაფანტვის დიაგრამის) ორი ეკონომიკური მაჩვენებლისათვის ან ორგანზომილებიანი კვეთისათვის, თუ საქმე ეხება მათ დიდ რაოდენობას;
2) კორელაციის შერჩევითი კოეფიციენტების განსაზღვრა ან კორელაციურ მატრიცათა შექმნა;
3) კავშირის პარამეტრების ინტერვალური შეფასება;
4) სტატისტიკური ჰიპოთეზების შემოწმება მაჩვენებლებს შორის კავშირების მნიშვნელოვნების შესახებ.

5.2. საკონტროლო მაგალითის შედეგების შემოწმება BASIC STATISTICA-ში

მაგალითი 5.1. განვიხილოთ კორელაციური მატრიცის აგება მანქანადმშენებლობის საწარმოს სამეურნეო საქმიანობის მაჩვენებელთა ანალიზის მაგალითზე. საწყის მონაცემებად გამოვიყენოთ მონაცემები data1.sta ფაილიდან.

გვაქვს $Y_1,\dots,Y_3,X_4,\dots,X_{17}$ ცვლადთა სისტემა. განვიხილოთ შემდეგი მახასიათებლები:

Y_1 - შრომისნაყოფიერება;

Y_2 - პროდუქციის თვითღირებულების შემცირების ინდექსი;

Y_3 - რენტაბელობა;

X_4 - პროდუქციის ერთეულის შრომატევადობა;

X_5 - მუშათა ხვედრითი წილი სამეურნეო-საწარმოო პერსონალის შემადგენლობაში;

X_6 - შეძენილი ნაკეთობის ხვედრითი წილი;

X_7 - დანადგარების ცვლის კოეფიციენტი;

X_8 - პრემიები და საჩუქრები ერთ მუშაკზე;

X_9 - წუნიდან მიღებული დანაკარგების ხვედრითი წილი;

X_{10}- ფონდუკუგება;

X_{11}- სამეურნეო-საწარმოო პერსონალის საშუალო წლიური რაოდენობა;

X_{12}- ძირითადი საწარმოო ფონდების საშუალო წლიური ღირებულება;

X_{13}- ხელფასის საშუალო წლიური ფონდი;

X_{14}- შრომის ფონდაღჭურვილობა;

X_{15}- ნორმირებული საბრუნავი საშუალებების ბრუნვადობა;

X_{16}- არანორმირებული საბრუნავი საშუალებების ბრუნვადობა;

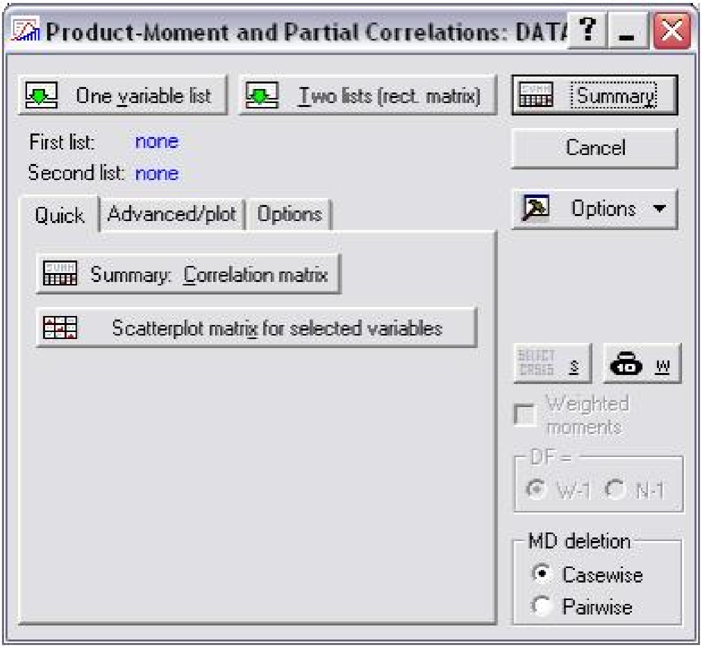
X_{17}- არასაწარმოო ხარჯები.

მოდულში Basic Statistics and Tables (ძირითადი სტატისტიკები) ადვილად შეიძლება გავითვალოთ და გავაანალიზოთ ჩვენს მიერ არჩეული ცვლადების კორელაციური მატრიცა. დასაწყისისათვის ჩავატაროთ Y_1, X_4, X_5 და X_6 ცვლადების კორელაციური ანალიზი:

-გაუშვით პროგრამა STATISTICA ბრძანებით Start/Programs /STATISTICA 7.0/ STATISTICA.

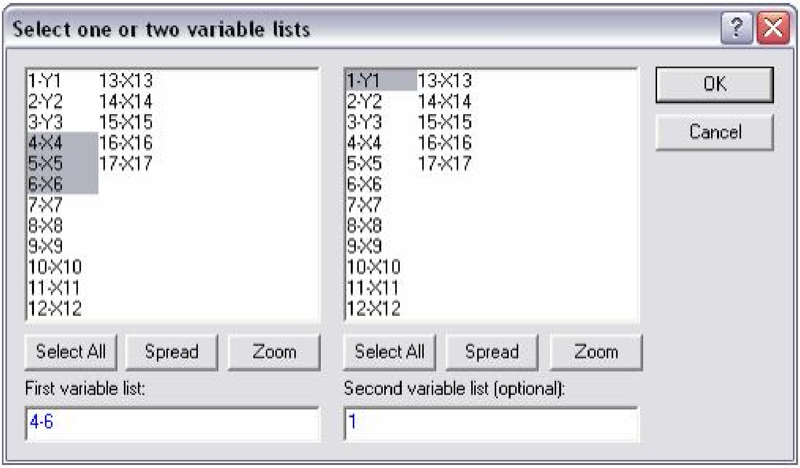
-აირჩიეთ ბრძანება File/Open (გახსნას მონაცემთა ფაილი) და გახსენით ფაილი data.sta. გადაერთეთ მოდულში Basic Statistics and Tables ბრძანებით Statistics/ Basic Statistics.

-Basic Statistics and Tables (ძირითადი სტატისტიკები) მოდულის სასტარტო პანელში აირჩიეთ პუნქტი Correlation matrices (კორელაციური მატრიცები). ორჯერ დააწკაპუნეთ მაუსის მარცხენა ღილაკი მასზე, ან გამოაჩინეთ და დააწექით OK ღილაკს. ეკრანზე გამონათდება Product-Moment Correlation and Partial Correlations (პირსონის კორელაცია) ფანჯარა (ნახ. 5.1).



ნახ. 5.1.

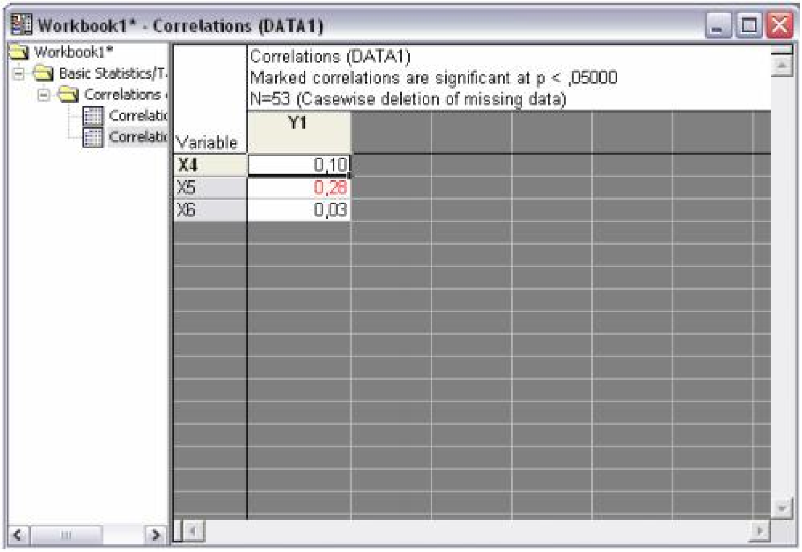
-დააწექით ღილაკს Two list (ორი სია). რის შემდეგაც გამონათდება ცვლადების შერჩევის ფანჯარა. აირჩიეთ ცვლადები ისე, როგორც ეს ნაჩვენებია ნახ. 5.2-ზე. ამრიგად, ჩვენ განვსაზღვრეთ ცვლადების ორი სია X_4 – X_6 – First variables list (ცვლადების პირველი სია) და Y_1 – Second variables list (ცვლადების მეორე სია). ჩვენ გვსურს გამოვითვალოთ კორელაციები Y_1 ცვლადსა და X_4 – X_6 ცვლადებს შორის. დააწკაპუნეთ OK ღილაკზე თქვენი არჩევანის დასადასტურებლად და Product-Moment Correlation and Partial Correlations ფანჯარაში დასაბრუნებლად.



ნახ. 5.2.

-ფანჯარაში Product-Moment Correlation and Partial Correlations დააწექით ღილაკს Summary. ეკრანზე დაინახავთ კორელაციურ მატრიცას (ნახ. 5.3).

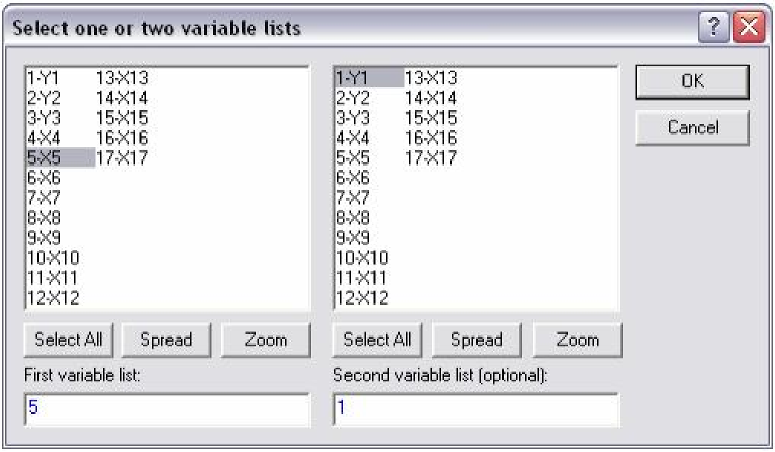
ამ მატრიცაში მხოლოდ ერთი სვეტია, ვინაიდან მეორე სიაში ჩვენ ავირჩიეთ მხოლოდ ერთი ცვლადი. სვეტში მოყვანილია კორელაციის კოეფიციენტები Y_1 და X_4 – X_6 ცვლადებს შორის. ჩვენს კორელაციურ მატრიცაში წითელი ფერით ავტომატურადაა გამოყოფილი კოეფიციენტები, რომლებიც მნიშვნელოვნებია P < 0,05 დონისთვის. სწორედ ამ კოეფიციენტებზეა საჭირო დიდი ყურადღების გამახვილება. უხეშად რომ ვთქვათ, დამოკიდებულე-



ნახ. 5.3.

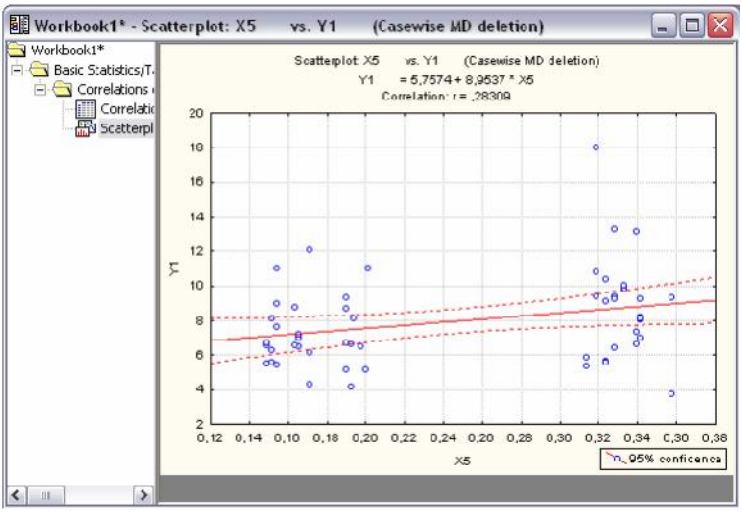
ბა იმ ცვლადებს შორის, რომელთა კორელაციის კოეფიციენტები გამოყოფილია წითელი ფერით, მნიშვნელოვანია. ჩვენს შემთხვევაში Y_1 ცვლადი ყველაზე მეტადაა დამოკიდებული X_5 ცვლადისაგან. ამ ცვლადებს შორის კორელაციის კოეფიციენტი ტოლია 0,28-ს. ვინაიდან $0{,}28 > 0$, ამიტომ ჩვენ შეგვიძლია ჩავთვალოთ, რომ X_5 ცვლადის ზრდისას Y_1 ცვლადიც აგრეთვე იზრდება. განვიხილოთ ეს ცვლადები უფრო ყურადღებით. სასარგებლოა ვნახოთ Y_1 და X_5 ცვლადებს შორის დამოკიდებულება გრაფიკულად.

-დაბრუნდით ფანჯარაში Product-Moment Correlation and Partial Correlations. დააწექით ღილაკს Two list (ორი სია). გამონათდება ცვლადების არჩევის ფანჯარა (მოცემულ შემთხვევაში გრაფიკისთვის). აირჩიეთ X_5 და Y_1 ცვლადები (ნახ. 5.4) და დააწექით OK ღილაკს.



ნახ. 5.4.

-ფანჯარაში Product-Moment Correlation and Partial Correlations ჩანართში Advansed/Plot დააწექით ღილაკს 2D scatterplots (გაფანტვის დიაგრამა 2D). ამის შემდეგ გამოჩნდება გაფანტვის დიაგრამა (ნახ. 5.5) შერჩეული ცვლადებისათვის.



ნახ. 5.5.

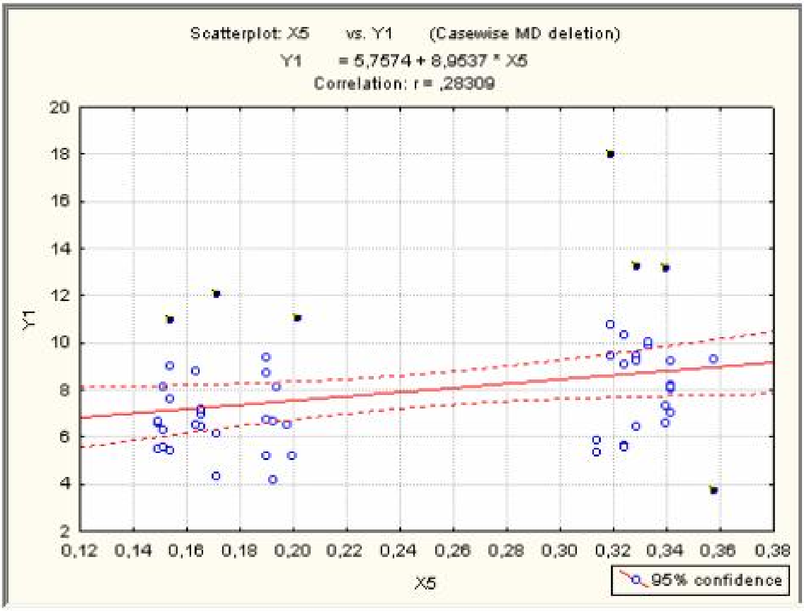
-გრაფიკიდან ცხადად ჩანს, რომ ცვლადებს შორის დამოკიდებულება არ არის წრფივი - წრფე ძალიან ცუდად „დევს“ მონაცემებზე. გრაფიკზე წარმოდგენილია საუკეთესო წრფე. როგორც არ უნდა ვცვალოთ დახრის კოეფიციენტი, მორგება უფრო ცუდი იქნება. STATISTICA გვთავაზობს შესაძლებლობებს, რომლებიც საშუალებას მოგვცემენ ჩავატაროთ მონაცემების უფრო სიღრმისეული განხილვა.

ნახ. 5.6.

-აირჩიეთ საშუალება „კისტი“ ზემოთ მოთავსებულ ინსტრუმენტთა პანელზე ღილაკზე  მაუსის მარცხენა ღილაკის დაწ-

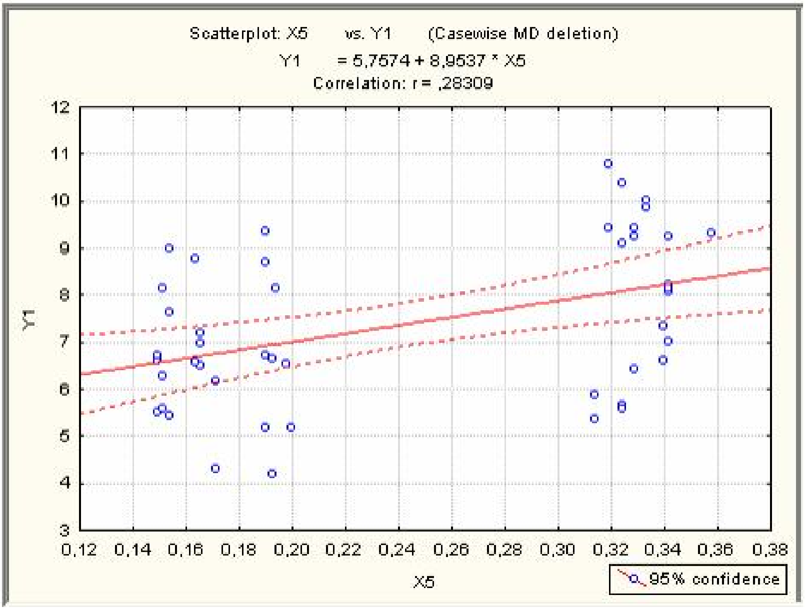
კაპუნებით. თქვენს წინ მარჯვნივ გამოჩნდება პანელი Brushing 2D (კისტი). პანელზე Brushing შეასრულეთ მომართვები, როგორც ეს ნაჩვენებია ნახ. 4.6-ზე და დააწექით ღილაკს Update.

-შედით გრაფიკში (უბრალოდ დააწკაპუნეთ მისი სივრცის ნებისმიერ წერტილზე, რითაც გრაფიკი გააქტიურდება) და აღნიშნეთ ლასსოთი ის წერტილები, რომლებიც, თქვენი აზრით, ყველაზე ძლიერ არიან გრაფიკზე წარმოდგენილი წრფიდან გადახრილები. ჩვენ ლასსოს დახმარებით ვახდენთ წერტილების გამოყოფას (შემოვუვლით რა მათ ფანქრით, თითქოს ვახდენთ ლასსოთი დაჭერას). ადრე აღნიშნული იყო ოპცია Lasso (ლასსო) ინსტრუმენტთა პანელზე Brushing (კისტი). მაგალითად, თუ ავირჩევთ ოფციას Simple (მარტივი), მაშინ ჩვენ წავშლიდით წერტილებს თანმიმდევრობით ერთმანეთის მიყოლებით. აირჩიეთ, მაგალითად, წერტილები წრფის ზემოთ, როგორც ეს ნაჩვენებია ნახ. 5.7-ზე.



ნახ. 5.7.

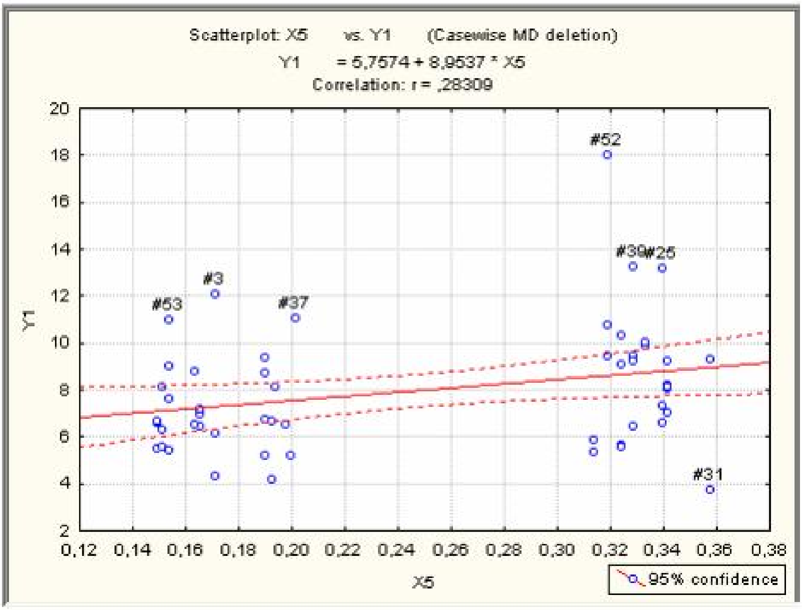
-დააწკაპუნეთ შემდეგ პანელ Brushing-ის ღილაკზე Apply (განახლება), გამონათდება შემდეგი გრაფიკი (ნახ. 5.8).



ნახ. 5.8.

ეხლა მონაცემები უკეთ „ჯდებიან“ წრფეზე. შეიძლება კვლევის გაგრძელება. ფრიად შესაძლებელია, რომ არსებობს რაღაც კანონზომიერება. საჭიროა ამ კანონზომიერებათა დამატებითი კვლევა. თქვენ ადვილად შეგიძლიათ განსაზღვროთ, თუ რომელი შემთხვევები წაშალეთ. ამისათვის შეგიძლიათ ისარგებლოთ ღილაკით Label (ნიშანი).

-დააწკაპუნეთ ღილაკზე Reset All (ყველა ამორჩევის გაუქმება) Brush პანელის ზემოთ. აღნიშნეთ Brush პანელზე Label ოპცია და ისევ მიიტაცეთ ლასსოთი საჭირო წერტილები. შემდეგ დააწექით ღილაკს Apply და ეკრანზე დაინახავთ გრაფიკს, რომელზეც გამოყოფილი წერტილების გვერდით გამოჩდნენ იმ შემთხვევათა სახელები, რომლებსაც ისინი მიეკუთვნებიან (ნახ. 5.9).

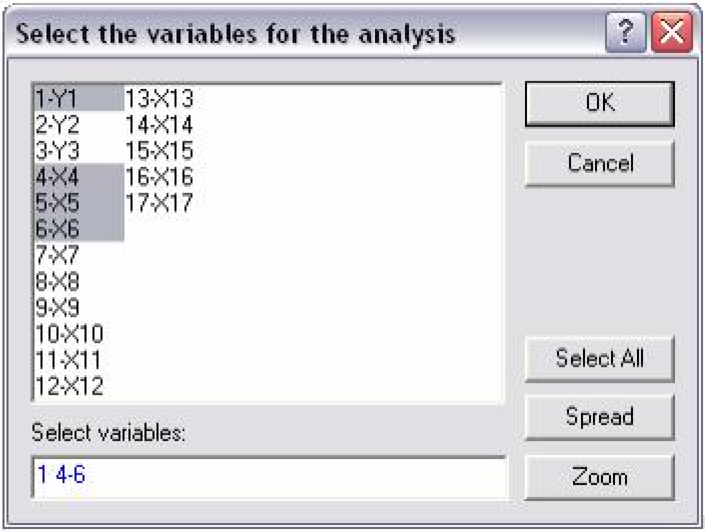


ნახ. 5.9.

სწორედ ეს შემთხვევები მოითხოვენ დამატებით კვლევას. მაგალითად, მათმა განხილვიდან ამოღებამ შეიძლება მიგვიყვანოს საკვლევი კორელაციის კოეფიციენტის მნიშვნელოვან ცვლილებასთან. იმ შემთხვევაში, თუ კორელაციურ მატრიცაში არის რამდენიმე გამოშუქებული კორელაციის კოეფიციენტი, მაშინ შემდეგ თქვენ დაგჭირდებათ განიხილოთ მონაცემები სხვა გამოშუქებული კორელაციის კოეფიციენტებით, ააგოთ დამოკიდებულებების გრაფიკები, იმუშაოთ ინსტრუმენტთან Brushing (კისტი). კორელაციის კოეფიციენტები კარგად ესადაგებიან წრფივი კავშირების აღწერას და ცუდად, თუ ცვლადებს შორის დამოკიდებულება არაა წრფივი. თქვენ შეგიძლიათ „გრაფიკულად“ დაათვალიეროთ კორელაციური მატრიცა ფანჯარაში Product-Moment Correlation and Partial Correlations ღილაკით Matrix (მატრიცული გრაფიკი).

-დაბრუნდით ფანჯარაში Product-Moment Correlation and Partial Correlations.

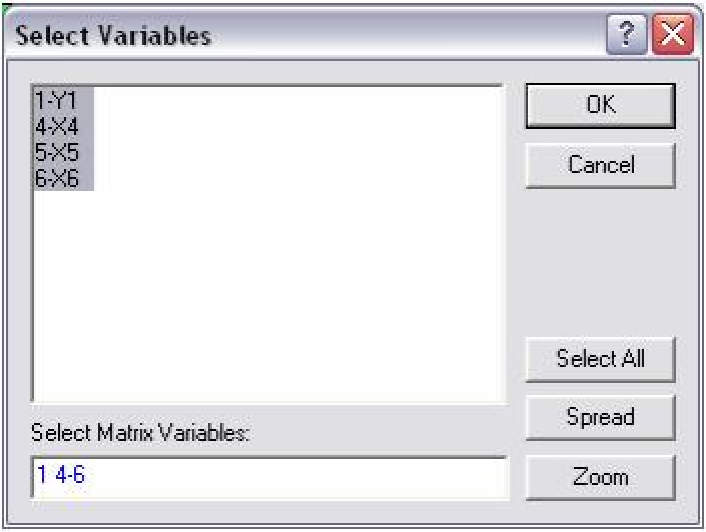
-დააწკაპუნეთ ღილაკზე One variable list (ცვლადების ერთი სია). აირჩიეთ ცვლადები Y_1, $X_4 – X_6$ როგორც ეს ნაჩვენებია ნახ. 5.10-ზე, ხოლო შემდეგ დააწკაპუნეთ ღილაკზე OK.



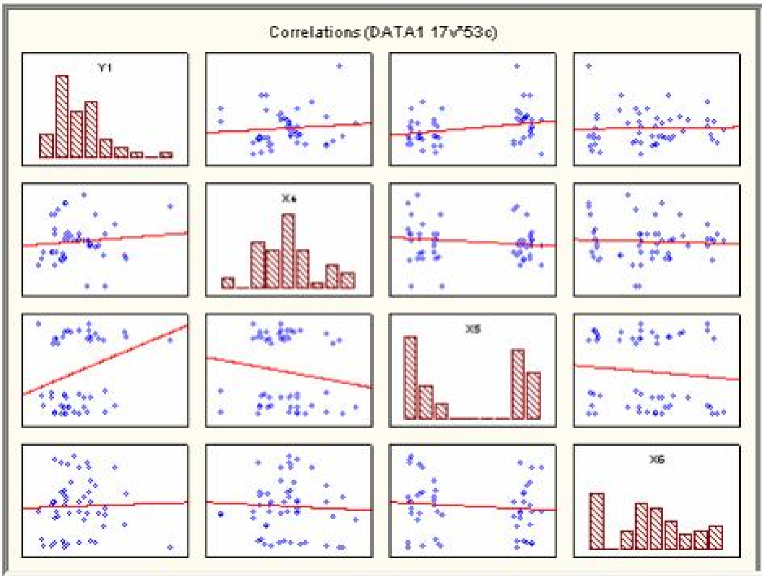
ნახ. 5.10.

-შემდეგ ფანჯარაში Product-Moment Correlation and Partial Correlations დააწექით ჩანართში Advansed/Plot ღილაკს Scatterplot Matrix. ამის შემდეგ გაიღება ცვლადების შერჩევის ფანჯარა გრაფიკების ასაგებად (ნახ. 5.11).

-ცვლადების შერჩევის ფანჯარაში აირჩიეთ ყველა ცვლადი ღილაკზე Select All (ყველას არჩევა) დაწოლით. დაადასტურეთ თქვენი არჩევანი OK ღილაკზე დაწოლით. ეკრანზე გამონათდება კორელაციური მატრიცა გრაფიკული სახით, რომელიც საშუალებას იძლევა შევაფასოთ წრფივი კავშირები ვიზუალურად (ნახ. 5.12).



ნახ. 5.11.



ნახ. 5.12.

5.3. ანგარიშის შექმნა

STATISTICA-ში ანალიზის შედეგები წარმოდგებიან სამი სახით: ფანჯარაში, მუშა დავთარში (Workbook) და ანგარიშში (Report).

STATISTICA-ში თითოეული ანგარიში გამოდის თავის საკუთარ ფანჯარაში სისტემის მუშა არეში. როგორც კი ეს ფანჯარა აქტიურდება, მაშინვე იცვლებიან ინსტრუმენტთა პანელები და მენიუ. მასში ჩნდებიან ბრძანებები და ინსტრუმენტები, მისაწვდომნი ამ ტიპის დოკუმენტებისათვის.

მუშა დავთარი წარმოადგენს შუალედურ ვარიანტს შედეგების შესანახად. არსებობს ამ ფაილების ერთადერთი ფორმატი *.stw.

ანგარიში - სისტემა STATISTICA-ს დოკუმენტია (მითითების გარეშე ფაილის ფორმატია *.str), რომელშიც შეიძლება შეინახოთ ნებისმიერი ტექსტური, რიცხვითი ან გრაფიკული ინფორმაცია. ანგარიში აუცილებელია ანალიზის შედეგების საბოლოო წარმოდგენისა და მათი ბეჭდვაზე გამოყვანისათვის. გარდა ამისა, ანგარიშები შეიძლება ინახებოდნენ შემდეგ ფორმატებში:

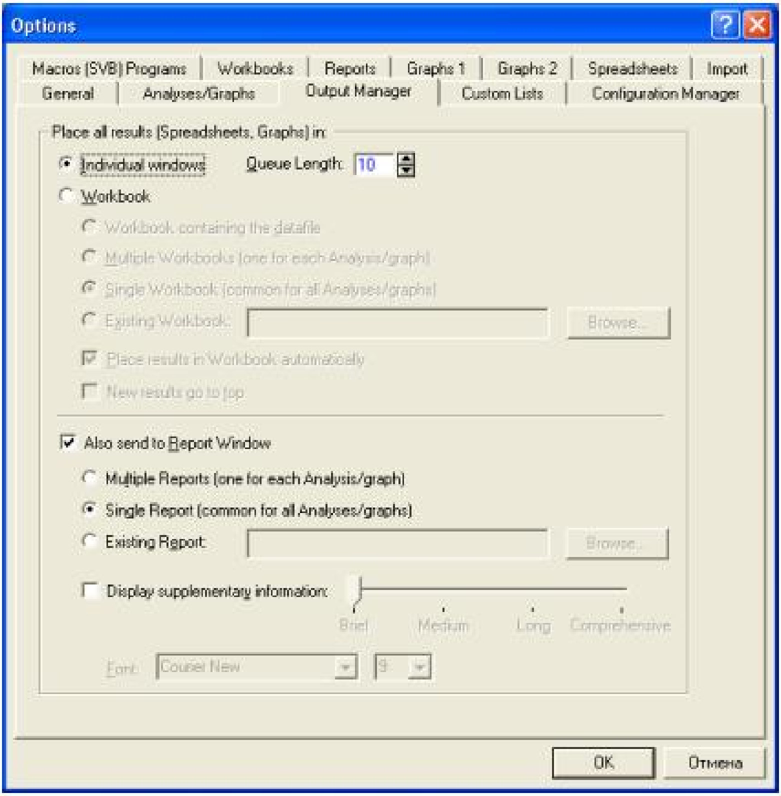
*.rtf - ტექსტური დოკუმენტების შენახვის ერთ-ერთი ღია ფორმატი. მოსახერხებელია შესრულებული სამუშაოს შესახებ ანგარიშის მომზადებისათვის. საშუალებას იძლევა მასში გამოყვანილ იქნას როგორც ტექსტური, რიცხვითი, ასევე გრაფიკული ინფორმაციაც. თქვენ შეგიძლიათ ცვლილებები შეიტანოთ ამ ანგარიშში ან STATISTICA-ის ფარგლებში, ან ჩატვირთოთ ის რომელიმე ტექსტურ რედაქტორში, მაგალითად MS Word-ში;

*.txt - ამ ოფციის არჩევისას ელექტრონული ცხრილის შემცველობა გამოდის ტექსტურ ფაილში, გრაფიკული ინფორმაციის დაკარგვით;

*.htm - შედეგების გამოტანა ხდება Web-დოკუმენტებში.

მუშა დავთრებისა და ანგარიშების მომართვებში ჩაღრმავების გარეშე, ავღნიშნოთ, რომ ამ და მომდევნო ინდივიდუალური

დავალებების შესრულებისათვის, მიზანშეწონილია გამოვიყენოთ მხოლოდ ერთადერთი ანგარიში ანალიზის ყველა შედეგის ავტომატური გამოყვანისათვის. შესაბამისი მომართვები შეგვიძლია გავაკეთოთ ნახ. 5.13-ზე წარმოდგენილი ფანჯრის მიხედვით, რომლის გამოძახებაც ხდება ბრძანებით Tools/Options ჩანართში Output Manager.



ნახ. 5.13.

შ ე ნ ი შ ვ ნ ა: ფაილი საწყისი მონაცემებით მოცემულ ანგარიშში ავტომატურად არ თავსდება.

6. რეგრესიული ანალიზი

6.1. ზოგადი ცნობები. მარტივი წრფივი რეგრესია

6.1.1. რეგრესიის ფუნქცია

რეგრესიისა და კორელაციის ცნებები მჭიდროდ არიან ერთმანეთთან დაკავშირებულნი, მაგრამ ამასთან ერთად არსებობს მათ შორის განსხვავება. კორელაციურ ანალიზში ფასდება სტოქასტიური კავშირის სიმჭიდროვე, ხოლო რეგრესიულ ანალიზში ხდება ფორმების კვლევა.

რეგრესიულ ანალიზში შეისწავლება კავშირი და განისაზღვრება რაოდენობრივი დამოკიდებულება დამოკიდებელ ცვლადსა და ერთ ან რამდენიმე დამოუკიდებელ ცვლადებს შორის, რომლებიც განიხილებიან როგორც არაშემთხვევითი სიდიდეები. ვთქვათ, Y ცვლადი დამოკიდებულია ერთ x ცვლადზე. ამასთან იგულისხმება, რომ x ცვლადი ღებულობს მოცემულ ფიქსირებულ მნიშვნელობებს, ხოლო დამოკიდებულ Y ცვლადს გააჩნია გაზომვის შეცდომების შემთხვევითი გაბნევა, სხვადასხვა გაუთვალისწინებელი ფაქტორების და ა. შ. გამო. თითოეულ x მნიშვნელობას შეესაბამება შემთხვევითი Y სიდიდის ალბათური განაწილების რაღაც კანონი. დავუშვათ, რომ Y „საშუალოდ" წრფივადაა დამოკიდებული x ცვლადის მნიშვნელობებზე. ეს ნიშნავს, რომ Y შემთხვევითი სიდიდის პირობით მათემატიკურ ლოდინს x-ის მოცემული მნიშვნელობის დროს აქვს შემდეგი სახე:

$$M(Y/x) = a_0 + a_1 \cdot x$$

მოცემულ ფუნქციას ეწოდება Y-ის x-ზე რეგრესიის წრფივი თეორიული ფუნქცია, ხოლო a_0 და a_1 პარამეტრებს - წრფივი რეგრესიის პარამეტრები (რეგრესიის კოეფიციენტები). პრაქტიკაში რეგრესიის პარამეტრები განისაზღვრებიან Y და x ცვლადებზე დაკვირვე-

ბის შედეგების მიხედვით, მათ შორის კავშირი კი შეიძლება შემდეგნაირად ჩავწეროთ:

$$Y = a_0 + a_1 \cdot x + \varepsilon$$

სადაც ε - დაკვირვების შემთხვევითი შეცდომაა.

6.2.რეგრესიული ანალიზის თანმიმდევრობა

- ამოცანის ფორმულირება.
- ცვლადების იდენტიფიკაცია (შესასვლელი და გამოსასვლელი ცვლადების განსაზღვრა).
- სტატისტიკური მონაცემების შეკრება.
- რეგრესიის ფუნქციის სპეციფიკაცია (მოდელის სახის განსაზღვრა).
- რეგრესიის ფუნქციის პარამეტრების შეფასება.
- რეგრესიული ანალიზის სიზუსტის შეფასება:
  1) მთელი მოდელის ადექვატურობის შემოწმება, ანუ თანხმობაშია თუ არა გამოსასვლელი სიდიდის პროგნოზირებული მნიშვნელობები დაკვირვების მონაცემებთან;
  2) მოდელის პარამეტრების მნიშვნელოვნების შეფასება, ანუ მნიშვნელოვნად განსხვავდებიან თუ არა ისინი ნულისგან.
- შედეგების ინტერპოლირება, ანალიზი, ოპტიმიზაცია და პროგნოზირება.

რეგრესიული ანალიზის ჩატარების წინაპირობები

- დაკვირვებათა შემთხვევით სიდიდეებს აქვთ განაწილების ნორმალური კანონი

$$\varepsilon \to N(0,\sigma), \qquad M(\varepsilon) = 0, \qquad D(\varepsilon) = \sigma^2 = const$$

- ავტოკორელაციის არ არსებობა დაკვირვებათა შეცდომებს შორის, ანუ ε_i - ის მიმდევრობითი მნიშვნელობები არ არიან დამოკიდებულნი ერთმანეთზე.

უმცირეს კვადრატთა მეთოდი

დაკვირვებათა შედეგების მიხედვით მოდელის პარამეტრების შეფასებების საპოვნელად გამოიყენება უმცირეს კვადრატთა მეთოდი. ვთქვათ, ჩატარებულია შემთხვევითი Y სიდიდის n დამოუკიდებელი დაკვირვება x - ის შესაბამისი მნიშვნელობების დროს, რომელთა ერთობლივი განაწილების კანონი ცნობილი არაა. გამომდინარე, რეგრესიის თეორიული ფუნქციის პოვნას ჩვენ ვერ შევძლებთ. ჩვენი ამოცანაა შევაფასოთ რეგრესიის ემპირული ფუნქცია

$$\tilde{y} = \widetilde{a_0} + \widetilde{a_1} \cdot x$$

უმცირეს კვადრატთა მეთოდის თანახმად, პარამეტრები ისეთი სახით შეირჩევიან, რომ მოხდეს დაკვირვებად მონაცემებსა და მოდელით გამოთვლილ მონაცემებს შორის სხვაობათა კვადრატების ჯამის მინიმიზირება

$$F = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \widetilde{y_i}) = \sum_{i=1}^{n} (y_i - \widetilde{a_0} - \widetilde{a_1} \cdot x)^2 \to min$$

სადაც y_i- გამოსასვლელი ცვლადის დაკვირვებადი მნიშვნელობებია; $\widetilde{y_i}$- მოდელით გამოთვლილი გამოსასვლელი ცვლადის მნიშვნელობები.

მინიმუმის აუცილებელი პირობიდან

$$\begin{cases} \dfrac{\partial F}{\partial \widetilde{a_0}} = -2 \cdot \sum_{i=1}^{n} (y_i - \widetilde{a_0} - \widetilde{a_1} \cdot x_i) = 0 \\ \dfrac{\partial F}{\partial \widetilde{a_1}} = -2 \cdot \sum_{i=1}^{n} (y_i - \widetilde{a_0} - \widetilde{a_1} \cdot x_i) = 0 \end{cases}$$

ვპოულობთ a_0 და a_1 პარამეტრების შეფასებებს (აქ და შემდგომ, თუ ეს აღთქმას არ უშლის ხელს, $\sim$ ნიშანი პარამეტრების ზემოთ გამოტოვებული იქნება). მათი განსაზღვრა მოხდება ორი წრფივი განტოლებისგან შემდგარი სისტემის ამოხსნით:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum x_i = \sum y_i \\ a_0 \cdot \sum x_i + a_1 \cdot \sum x_i^2 = \sum y_i \cdot x_i \end{cases}$$

აქ და შემდგომ, თუ ეს განსაკუთრებით არაა აღნიშნული, აჯამვა ხორციელდება i=1-დან n-მდე. დაკვირვებების შემთხვევითი შეცდომების მიმართ წინაპირობათა შესრულების შემთხვევაში უმცირეს კვადრატთა მეთოდით მიღებული პარამეტრების შეფასებებს ექნებათ შემდეგი თვისებები:

- ჩაუნაცვლებლობა;
- ძალმოსილება;
- ეფექტურობა.

მოდელის ადექვატურობის შემოწმება

მოდელის ადექვატურობის ჰიპოთეზის შესამოწმებლად აუცილებელია კვადრატთა ორი ჯამის შედარება:

1) კვადრატთა ნარჩენი ჯამი, რომელიც ახასიათებს რეგრესიიდან გადახრას

$$Q_e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \widetilde{y}_i)^2$$

2) რეგრესიით განპირობებული კვადრატთა ჯამი

$$Q_R = \sum_{i=1}^{n} (\tilde{y}_i - \bar{y})^2$$

სადაც

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

მაშინ F-ის შერჩეული მნიშვნელობა, რომელსაც გააჩნია ფიშერის განაწილება

$$F = \frac{Q_R/k}{Q_e/(n-k-1)}$$

შეიძლება გამოყენებულ იქნას ადექვატურობის შესამოწმებლად მნიშვნელოვნების მოცემული λ დონისთვის (ჩვეულებრივ უმრავლესი ამოცანებისათვის λ=0,05) და თავისუფლების $f_1=k$; $f_2=n-k-1$ ხარისხებისთვის, სადაც k- შესაფასებელი პარამეტრების რიცხვია, თავისუფალი კოეფიციენტის გამოკლებით.

თუ $F \geq F_{\lambda;f1;f2}$- მაშინ მოდელი ადექვატურია (დამატება 1). შეცდომის ნარჩენი დისპერსია

$$S^2 = Q_e/(n-k-1)$$

შეიძლება გამოვიყენოთ შემთხვევითი სიდიდის σ² დისპერსიის შესაფასებლად. ადექვატურობის შემოწმების შედეგები მოსახერხებელია წარმოვიდგინოთ შემდეგი სახით (ცხრილი 6.1).

ცხრილი 6.1.

ცვლილების წყარო	კვადრატების ჯამი	თავისუფლების ხარისხთა ჯამი	დისპერსიის შეფასება
მოდელი	$Q_R = \sum_{i=1}^{n} (\tilde{y}_i - \bar{y})^2$	k	$S_R^2 = Q_R/k$

შეცდომა	$Qe = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \widetilde{y}_i)^2$	*n-k-1*	$s^2 = Qe/(n-k-1)$
ჯამი	$Q^e + Q^R = \sum_{i=1}^{n}(y_i - \bar{y})^2$	*n-1*	

წრფივი რეგრესიის სასარგებლო მახასიათებელს წარმოადგენს დეტერმინაციის კოეფიციენტი, რომელიც გამოითვლება ფორმულით

$$R^2 = \frac{Q_R}{Q_R + Q_e} = 1 - \frac{Q_e}{Q_R + Q_e}$$

დეტერმინაციის კოეფიციენტი ტოლია დაკვირვებათა შედეგების იმ წილის ჰორიზონტალური $y = \bar{y}$ წრფის მიმართ, რომელიც აიხსნება რეგრესიის განტოლებით. სიდიდე $R = +\sqrt{R^2}$ წარმოადგენს დაკვირვებათა შედეგებსა და გამოთვლილ $\widetilde{y}_i$ მნიშვნელობებს შორის მრავლობითი კორელაციის კოეფიციენტის შეფასებას. თუ R^2=0,75, მაშინ ეს ნიშნავს, რომ მოდელი მუშაობს 75%-ზე, ხოლო 25% მოდის შეცდომაზე ან მოდელში გაუთვალისწინებელ ფაქტორებზე (პრაქტიკული მიზნებისათვის მიზანშეწონილია, რომ $R^2 \geq 0,75$). დაკვირვებათა მცირე მნიშვნელობისათვის ($n<30$) აუცილებელია ვისარგებლოთ დეტერმინაციის გაკორექტირებული კოეფიციენტით

$$R^{*2} = 1 - \frac{n-1}{n-k-1}(1 - R^2)$$

მოდელის პარამეტრების მნიშვნელოვნების შემოწმება

შემოწმების შედეგად დგინდება რეგრესიის პარამეტრების შე-ფასების ნულისგან განსხვავების სტატისტიკური მნიშვნელობა ან არამნიშვნელოვნება. ეს შემოწმება ხორციელდება მოდელის თი-თოეული პარამეტრისათვის ცალ-ცალკე. რეგრესიის კოეფიციენტე-ბის მნიშვნელოვნების შესაფასებლად შეიძლება ვისარგებლოთ შემდეგი წესით: თუ რეგრესიის კოეფიციენტის აბსოლუტური სი-დიდე მეტია სარწმუნოვნების ინტერვალზე, მაშინ ჰიპოთეზა კოე-ფიციენტის არამნიშვნელოვნების შესახებ უკუიგდება:

$$\tilde{a}_i - t_{f,\lambda/2} \cdot S_{ai} < a_i < \tilde{a}_i + t_{f,\lambda/2} \cdot S_{ai}, |\tilde{a}_i| \geq t_{f,\lambda/2} \cdot S_{ai},$$

სადაც $t_{f,\lambda/2}$ - სტიუდენტის კრიტერიუმის მნიშვნელობაა, განსა-ზღვრული თავისუფლების ხარისხთა რიცხვით f=n-k-1 და λ=0,05 -ით (დამატება 2); S_{ai} - რეგრესიის კოეფიციენტების შეცდომების საშუალო კვადრატული გადახრებია. მარტივი წრფივი რეგრესიისა-თვის $y = a_0 + a_1 \cdot x$ ისინი შეიძლება შესაბამისად შემდეგნაირად გამოვითვალოთ

$$S_{a0} = \sqrt{\frac{S^2 \sum x^2}{n \sum x^2 - (\sum x)^2}}, \qquad S_{a1} = \sqrt{\frac{n \cdot S^2}{n \sum x^2 - (\sum x)^2}}$$

კოეფიციენტების მნიშვნელოვნება შეიძლება შევამოწმოთ t - კრიტერიუმით. ვისარგებლოთ ფორმულით

$$t = \frac{|\tilde{a}_i|}{S_{ai}}$$

ხდება გამოთვლილი მნიშვნელობის შედარება ცხრილურ-თან და თუ $t \geq t_{f,\lambda/2}$, მაშინ კოეფიციენტი მნიშვნელოვანია. წინა-აღმდეგ შემთხვევაში შესაბამისი ცვლადი შეიძლება გამოვრიცხოთ მოდელიდან და ყველა გამოთვლა თავიდან შევასრულოთ.

6.3. მრავლობითი წრფივი რეგრესია

მრავლობითი წრფივი რეგრესია წარმოადგენს გამოსახულე-ბას

$$\tilde{y} = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \cdots + a_k \cdot x_k$$

k=2 შემთხვევისათვის დაკვირვებათა შედეგების მიხედვით უმცირეს კვადრატთა მეთოდით მიღებულ ნორმალურ განტოლებათა სისტემას ექნება შემდეგი სახე

$$\begin{cases} a_0 \cdot n + a_1 \cdot \sum x1_i + a_2 \cdot \sum x2_i = \sum y_i \\ a_0 \cdot \sum x1_i + a_1 \cdot \sum x1_i^2 + a_2 \cdot \sum x1_i \cdot x2_i = \sum y_i \cdot x1_i \\ a_0 \cdot \sum x2_i + a_2 \cdot \sum x1_i \cdot x2_i + a_2 \cdot \sum x2_i^2 = \sum y_i \cdot x2_i \end{cases}$$

შემდგომში მსჯელობისათვის მოხერხებულია გამოვიყენოთ შემდეგი მატრიცული აღნიშვნები

$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}$- დაკვირვებათა ვექტორი, $a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}$ - პარამეტრების ვექტორი,

$A = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & & x_{k2} \\ & \vdots & & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}$-რეგრესიული მატრიცა (nk+1).

განტოლებათა ნორმალურ სისტემას აქვს შემდეგი სახე: $(A^T A) \cdot a = A^T Y$. პირობით, თუ $A^T A$ - არაა გადაგვარებული მატრი-ცა, სისტემის ამოხსნა შეგვიძლია ჩავწეროთ შემდეგნაირად

$$a = (A^T A)^{-1} A^T Y$$

რეგრესიული მოდელის პარამეტრების შეფასების კოვარი-აციული მატრიცა იქნება

$$K_{a_i a_j} = S^2 (A^T A)^{-1}.$$

მოდელის პარამეტრების დისპერსია განისაზღვრება გამოსახულებით:

$$S_{a_i}^2 = K_{a_i a_j}.$$

მრავლობით წრფივ რეგრესიაში რეგრესიული ანალიზის წინაპირობები და მისი წარმართვა მთლიანად ემთხვევა მარტივ წრფივ რეგრესიას. მრავლობითი რეგრესიის თავისებურებას წარმოადგენს დამოუკიდებელი ცვლადების კორელაცია. სასურველია მოდელში არ ჩავრთოთ წრფივად დამოკიდებული ცვლადები.

6.4. არაწრფივი მოდელები, რომლებიც წრფივებზე დაიყვანებიან

არსებობს რეგრესიული მოდელების ორი სახის არა წრფივობა:

არაწრფივები დამოუკიდებელი ცვლადების მიმართ.

მაგალითად,

$$\tilde{y} = a_0 + a_1 x_1 + a_2 x_2^2 + a_3 x_1 x_2$$

ამ შემთხვევაში უბრალოდ საჭიროა ცვლადების შეცვლა:

$$x_2^2 = z_1, \quad x_1 x_2 = z_2,$$

მივიღებთ: $\tilde{y} = a_0 + a_1 x_1 + a_2 z_1 + a_3 z_2$

არაწრფივები რეგრესიის პარამეტრების მიმართ.

მაგალითად,

$$\tilde{y} = \frac{1}{a_0 + a_1 x}$$

შევასრულოთ ფუნქციონალური გარდაქმნა:

ვთქვათ $z = \frac{1}{\tilde{y}}$ - მაშინ, $z = a_0 + a_1 x$.

სამწუხაროდ, ყოველთვის არ შეიძლება ფუნქციონალური გარდაქმნებით არაწრფივი მოდელებიდან გადავიდეთ წრფივებზე.

გარდა ამისა, მხედველობაში უნდა ვიქონიოთ, რომ უმცირეს კვადრატთა მეთოდით პარამეტრების გამოთვლის დროს ხდება გარდაქმნილი გადახრების კვადრატების ჯამის მინიმიზირება და არა საწყისი მონაცემების.

6.5. რეგრესიული ანალიზის წინაპირობების შემოწმება

6.5.1. შეცდომების განაწილების კანონის ნორმალურობის შემოწმება

შეცდომების ანალიზი ხორციელდება შემდეგი სქემის მიხედვით. დავუშვათ, რომ $\varepsilon_i \sim N(0, \sigma)$, მაშინ $\varepsilon_i \sim N(0,1)$. თუ მოდელი სწორია, მაშინ ნარჩენების დისპერსია, რომელიც ახასიათებს დაკვირვებათა შედეგების აპროქსიმაციის ხარისხს

$$S^2 = \frac{Q_e}{n-k-1} = \frac{\sum_{i=1}^{n}(e_i - \bar{e})^2}{n-k-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1}$$

ემსახურება დაკვირვების შეცდომების დისპერსიის - σ^2 სიდიდის შეფასებას, სადაც $\bar{e}$- გადახრების საშუალო მნიშვნელობაა. $e_i/_S$ შემთხვევითი სიდიდე წარმოადგენს ერთეულოვან ნორმალურ გადახრებს. თუ ეს გადახრები იქმნებიან [-2 ; 2] ინტერვალში, მაშინ, გამომდინარე, ჩვენი წარმოდგენა იმის შესახებ, რომ $\varepsilon_i \sim N(0, \sigma)$, არაა შეცდომითი.

6.5.2. შემთხვევითი შეცდომების ერთგვაროვნებაზე შემოწმება

$D(\varepsilon)=const.$ დისპერსიის ერთგვაროვნებაზე შემოწმებისთვის მიზანშეწონილია ვისარგებლოთ გოლდფერდის მეთოდით. Y შემთხვევითი სიდიდის მნიშვნელობათა თანმიმდევრობა იყოფა ორ n_1 და n_2 მოცულობების თანმიმდევრობებად, სადაც $(n_1+n_2=n)$. თითოეული თანმიმდევრობისათვის გამოითვლებიან აღდგენის S_1^2 და S_2^2 დისპერსიები. მაშინ ფარდობას

$$F = \frac{S_2^2}{S_1^2},$$

$S_1^2 < S_2^2$ დროს ექნება ფიშერის განაწილება თავისუფლების ხარისხებით $f_1=n_1-k-1$, $f_2=n_2-k-1$. თუ F-ის მნიშვნელობა აღემატება ცხრილურს, მაშინ ჰიპოთეზა დისპერსიის ერთგვაროვნების შესახებ უკუიგდება. კრიტერიუმის მგრძნობიარობა იზრდება, თუ გამოვრიცხავთ საშუალო დაკვირვებებს.

იმ შემთხვევაში თუ დისპერსიები აღმოჩდნენ არაერთგვაროვნები, ხშირად სასარგებლოა გამოსასვლელი ცვლადის მასშტაბის შეცვლა. შემოაქვთ გამოსასვლელი ცვლადის რაღაც ფუნქცია, მაგალითად $\ln y$ ან $\sqrt{y}$.

6.5.3. შემთხვევითი შეცდომების ავტოკორელაციაზე შემოწმება

შეცდომების ავტოკორელაციის არსებობა შეიძლება შევამოწმოთ დარბინ - უოტსონის კრიტერიუმის დახმარებით:

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

დარბინ - უოტსონის კრიტერიუმი იცვლება $0 \le DW \le 4$ დიაპაზონში. ავტოკორელაციის არ არსებობისას $DW=2$. დამატება 3 - ში

მოყვანილია კრიტერიუმის ქვედა და ზედა საზღვრები d_1 და d_2 თავისუფლების $f_1=n; f_2=k$ ხარისხებისათვის.

თუ $0 \le DW \le d_{\text{ქ}}$, არსებობს დადებითი ავტოკორელაცია,

$4-d_{\text{ქ}} \le DW \le 4$, არსებობს უარყოფითი ავტოკორელაცია,

$d_{\text{ზ}} \le DW \le 4-d_{\text{ზ}}$, ავტოკორელაცია არ არსებობს,

$d_{\text{ქ}} \le DW \le d_{\text{ზ}}$ ან $4-d_{\text{ზ}} \le DW \le 4-d_{\text{ქ}}$, საჭიროა დამატებითი კვლევების ჩატარება.

პ რ ა ქ ტ ი კ უ ლ ი რ ჩ ე ვ ა: სასურველია განუსაზღვრელობის ზონა მივუერთოთ ავტოკორელაციის არ არსებობის ჰიპოთეზის უკუგდების ზონას.

6.6. ტიპიური მაგალითის განხილვა

ცხრილში 6.2. მოყვანილია y_i ცვლადის მნიშვნელობები შესასვლელი x_i ცვლადის მოცემული მნიშვნელობების დროს. მოდელს აქვს შემდეგი სახე

$$\tilde{y} = a_0 + a_1 x.$$

ცხრილი 6.2.

X	1	2	3	4
Y	2	4	5	7

1. გამოვითვალოთ მათემატიკური ლოდინის შეფასება, დისპერსიებისა და საშუალო კვადრატული გადახრების ჩანაცვლებული და ჩაუნაცვლებელი შეფასებები:

$$\bar{x} = \frac{1}{n}\sum x_i = 2{,}5\,, \quad \bar{y} = \frac{1}{n}\sum y_i = 4{,}5\,,$$

$$S_x^2 = \frac{1}{n}\sum (x_i - \bar{x})^2 = 1{,}250\,, \quad S_y^2 = \frac{1}{n}\sum (y_i - \bar{y})^2 = 3{,}251\,,$$

$$S_{XH}^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = 1{,}666\,, \quad S_{YH}^2 = \frac{1}{n-1}\sum (y_i - \bar{y})^2 = 4{,}333\,,$$

$$S_x = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2} = 1{,}118\,,\ S_y = \sqrt{\frac{1}{n}\sum(y_i - \bar{y})^2} = 1{,}803,$$

$$S_{XH} = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2} = 1{,}291\,,\quad S_{YH} = \sqrt{\frac{1}{n-1}\sum(y_i - \bar{y})^2}$$
$$= 2{,}082$$

2.განვსაზღვროთ მოდული. გამოვითვალოთ შესაბამისი ჯამები და შევადგინოთ ორი წრფივი განტოლების სისტემა:

$$\sum x_i = 10, \sum y_i = 18, \sum x_i^2 = 30, \quad \sum x_i y_i = 53, \quad \sum y_i^2 = 94$$

რომლის ამოხსნით ვღებულობთ: a_0=0,5, a_1=1,6. მოდელს აქვს შემდეგი სახე:

$$\tilde{y} = 0{,}5 + 1{,}6x.$$

3.გამოვითვალოთ მოდელური მნიშვნელობები განტოლებაში შესასვლელი ცვლადის მნიშვნელობების ჩასმით (ცხრ. 6.3).

ცხრილი 6.3.

[illegible]	2,1	3,7	5,3	6,9

4.განვსაზღვროთ მოდელის ადექვატურობა. ამისათვის გამოვითვალოთ კვადრატების საერთო ჯამი, რეგრესიასთან დაკავშირებული კვადრატების ჯამი და ნარჩენების კვადრატების ჯამი:

$$Q = \sum(y_i - \bar{y})^2 = \sum y^2 - \frac{(\sum y)^2}{n} = 94 - \frac{(18)^2}{4} = 13$$

$$Q_R = \sum(\tilde{y}_i - \bar{y})^2 = \sum \tilde{y}^2 - \frac{(\sum y)^2}{n} = 93{,}8 - \frac{(18)^2}{4} = 12{,}8$$

$$Q_e = \sum e_i^2 = (2 - 2{,}1)^2 + (4 - 3{,}7)^2 + (5 - 5{,}3)^2 + (7 - 6{,}9)^2 = 0{,}2$$

მიღებული გამოთვლების შესამოწმებლად საჭიროა, რომ $Q = Q_R + Q_e$. ჩვენს შემთხვევაში ტოლობა სრულდება.

ფიშერის კრიტერიუმის გამოთვლილი მნიშვნელობა ტოლია

$$F = \frac{Q_R/k}{Q_e/(n-k-1)} = \frac{S_R^2}{S^2} = \frac{12{,}8/1}{0{,}2/(4-1-1)} = 128$$

ფიშერის კრიტერიუმის გამოთვლილი მნიშვნელობა მეტია ვიდრე ცხრილური მნიშვნელობა $F > F_{ცხრ.} = F_{0,05;\,1;\,2} = 18{,}512$.

დ ა ს კ ვ ნ ა 1: მოდელი ადექვატურია, საწყისი მონაცემები კარგად ესადაგებიან მოდელურ მონაცემებს.

5.ოპერატიულად მოდელის ადექვატურობა შეიძლება შემოწმდეს დეტერმინაციისა და კორელაციის კოეფიციენტებით. დეტერმინაციის კოეფიციენტი, კორექტირებული დეტერმინაციის კოეფიციენტი და კორელაციის კოეფიციენტი შემდეგნაირად განისაზღრებიან:

$$R^2 = \frac{Q_R}{Q_R + Q_e} = \frac{Q_R}{Q} = \frac{12{,}8}{13} = 0{,}9846$$

$$R^{*2} = 1 - \frac{n-1}{n-k-1}(1 - R^2) = 1 - \frac{4-1}{4-1-1}(1 - 0{,}9846) = 0{,}977$$

$$R = \sqrt{R^2} = 0{,}9922$$

დეტერმინაციის კოეფიციენტის მიღებული მნიშვნელობა მეტია 0,75-ზე, გამომდინარე მოდელი შეგვიძლია ადექვატურად ჩავთვალოთ.

6.განვსაზღვროთ მოდელის პარამეტრების მნიშვნელობები. გამოვითვალოთ რეგრესიის კოეფიციენტების შეცდომების საშუალო კვადრატული გადახრები:

$$S_{a_0}^2 = S^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}\right)^2 = \frac{S^2\sum x_i^2}{n\sum x_i^2 - (\sum x_i)^2} = \frac{0{,}1\cdot 30}{4\cdot 30 - (10)^2} = 0{,}15$$

$$S_{a_1}^2 = \frac{S^2}{\sum(x_i - \bar{x})^2} = \frac{nS^2}{n\sum x_i^2 - (\sum x_i)^2} = \frac{4\cdot 0{,}1}{4\cdot 30 - (10)^2} = 0{,}02$$

$$S_{a_0} = \sqrt{0{,}15} = 0{,}387;\ S_{a_1} = \sqrt{0{,}02} = 0{,}141$$

სტიუდენტის კრიტერიუმის ცხრილური მნიშვნელობა $t_{2;0,025}$=4,303. ვღებულობთ სტიუდენტის გამოთვლილ მნიშვნელობებს:

$$t_{a_0} = 0{,}5/0{,}387 = 1{,}291;$$
$$t_{a_1} = 1{,}6/0{,}141 = 11{,}31.$$

დ ა ს კ ვ ნ ა 2: პარამეტრი a_1 მნიშვნელოვანია, ვინაიდან სტიუდენტის კრიტერიუმის გამოთვლილი მნიშვნელობა ცხრილურზე მეტია, ხოლო a_0 პარამეტრი არა მნიშვნელოვანია.

ანალოგიური დასკვნა შეგვიძლია სხვაგვარადაც გავაკეთოთ, თუ გამოვითვლით სარწმუნოვნების ინტერვალებს მოდელის პარამეტრებისათვის:

$$a_0 = 0{,}5 \pm 4{,}303 \cdot 0{,}387 = 0{,}5 \pm 1{,}665;$$
$$a_1 = 1{,}6 \pm 4{,}303 \cdot 0{,}141 = 1{,}6 \pm 0{,}607.$$

ვინაიდან a_1 პარამეტრის აბსოლუტური მნიშვნელობა იქნება მეტი ვიდრე მისი სარწმუნოვნების ინტერვალი, ამიტომ პარამეტრი მნიშვნელოვანია. a_0 პარამეტრი არა მნიშვნელოვანია.

7.შეცდომის საშუალო კვადრატული გადახრის შეფასება S=$\sqrt{\bar{S}^2}$=$\sqrt{0{,}1}$=0,316. ყველა გადახრის გაყოფით ამ სიდიდეზე მივიღებთ ნორმირებულ გადახრას, რომლებიც ყველა [-2 ; +2] ინტერვალში იმყოფება.

დ ა ს კ ვ ნ ა 3: პირველი წინაპირობა სრულდება. შემთხვევით შეცდომას გააჩნია განაწილების ნორმალური კანონი ნულოვანი მათემატიკური ლოდინით.

8.დარბინ - უოტსონის კრიტერიუმი შემდეგნაირად გამოითვლება:

DW=((0,3+0,1)²+(-0,3-03)²+(0,1+0,3)²)/0,2=3,4

1,5≤DW≤4-1,5 (d_B დაახლოებით 1,5-ის ტოლია (დამატება 3)).

დ ა ს კ ვ ნ ა 4: მეორე წინაპირობა არ სრულდება. შემთხვევითი სიდიდის მიმდინარე მნიშვნელობებს შორის არსებობს უარყოფითი ავტოკორელაცია.

სამუშაოს დასასრულს მოვიყვანოთ ნარჩენების და ნორმირებული გადახრების ჯამური ცხრილი (ცხრ. 6.4).

ცხრილი 6.4.

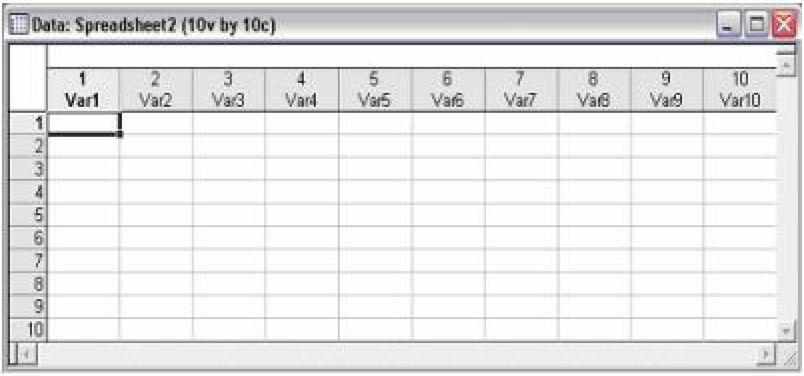
N/N	Y-ის მნიშვნელობა	Y-ის შეფასება	ნარჩენები	ნორმირებული გადახრები
1	2	2,1	-0,1	-0,316
2	4	3,7	0,3	0,948
3	5	5,3	-0,3	-0,948
4	7	6,9	0,1	0,316

6.7. საკონტროლო მაგალითის შედეგების შემოწმება MULTIPLE REGRESSION-ში

წინასწარი დამუშავების მოდულთან გამოთვლების ჩასატარებლად საჭიროა შემდეგი მოქმედებების შესრულება:

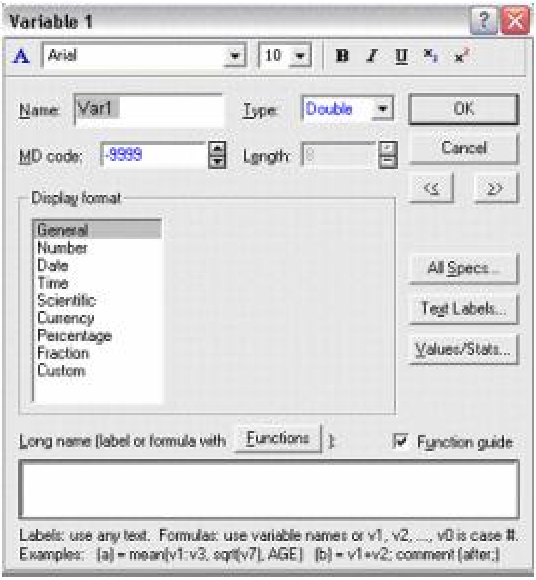
- გაუშვით პროგრამა STATISTICA ბრძანებით Start/All Programs/Statistica 7.0/ Statistics.
- გამონათებულ ფანჯარაში დახურეთ დოკუმენტების ყველა ფანჯარა და შეასრულეთ ბრძანება File/New Data. ფანჯარაში Create New Document ეთითება შესაქმნელი სტრიქონებისა და სვეტების რიცხვი, ჩუმათობის პრინციპით იქმნება მონაცემების შესატანად ცხრილი ზომებით (10 * 10) (ნახ. 6.1).

- შეიტანეთ საწყისი მონაცემები X და Y ცვლადებისათვის VAR 1 და VAR 2 სვეტებში ცხრილიდან 6.2.
- გამოყავით VAR 3 – VAR 10 სვეტების ბლოკი და ინსტრუმენტთა პანელზე დააწექით ღილაკს Vars. გამონათებულ მენიუში აირჩიეთ ბრძანება Delete, ხოლო შემდეგ დიალოგის ფანჯარაში დააწექით ღილაკს OK.
- გამოყავით სტრიქონთა 5 – 10 ბლოკი და დააწექით ინსტრუმენტთა პანელზე ღილაკს Cases. გამონათებულ მენიუში აირჩიეთ ბრძანება Delete, ხოლო შემდეგ დიალოგის ფანჯარაში დააწექით ღილაკს OK.

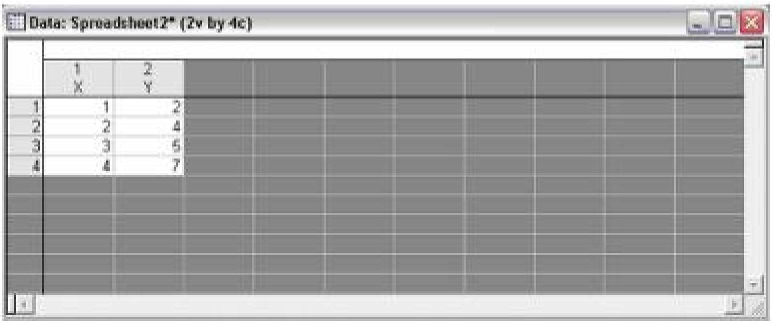


ნახ. 6.1.

- დააწკაპუნეთ მაუსის მარჯვენა ღილაკი VAR 1 სვეტზე და აირჩიეთ ბრძანება Variable Specs … . გამონათდება დიალოგის ფანჯარა (ნახ. 6.2).
- შეიტანეთ X ცვლადის სახელი და დააწექით OK-ის.
- ანალოგიური პროცედურა ჩაატარეთ VAR 2 სვეტისათვის.
- შედეგად თქვენ უნდა მიიღოთ შემდეგი სახის მონაცემების ცხრილი (ნახ. 6.3).

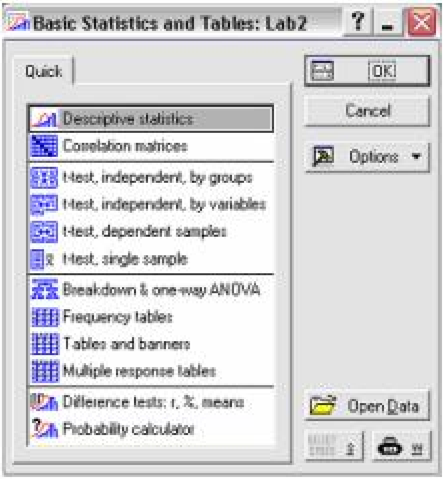


ნახ. 6.2.



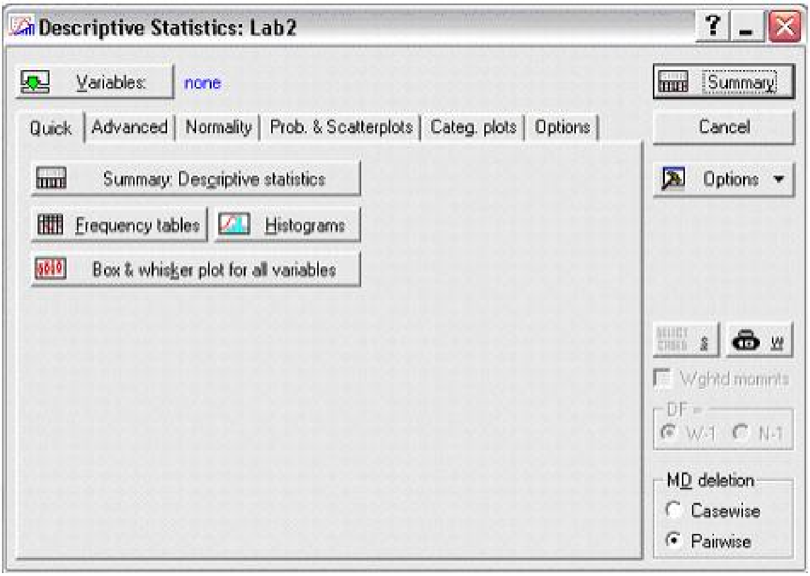
ნახ. 6.3.

- დაიმახსოვრეთ მონაცემების მიღებული ფაილი ბრძანებით File/Save As.
- გაააქტიურეთ ფანჯარა მონაცემების ცხრილით, ხოლო შემდეგ შეასრულეთ ბრძანება Statistics/Basic statistics. გამონათდება ფანჯარა (ნახ. 6.4).



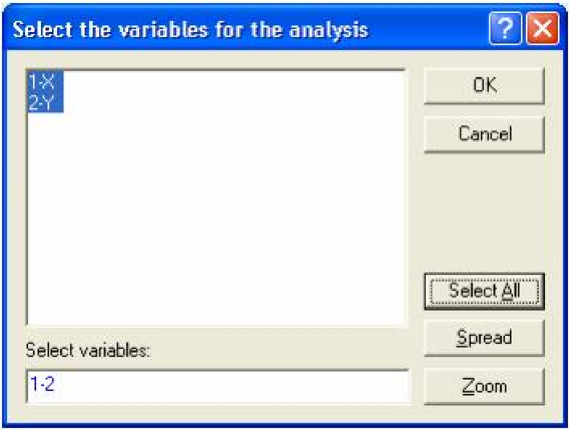
ნახ. 6.4.

- აირჩიეთ პუნქტი Descriptive statistics და დააწექით OK-ის. გამონათდება ფანჯარა (ნახ. 6.5).



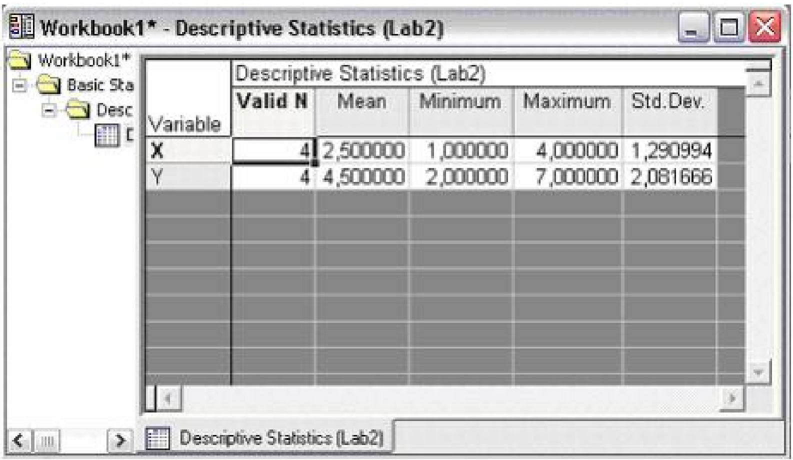
ნახ. 6.5.

- დააწექით ღილაკს Variables. გამონათებულ დიალოგის ფანჯარაში (ნახ. 6.6) დააწექით ღილაკს Select All და OK.



ნახ. 6.6.

- პროგრამა შეასრულებს წინა ფანჯარაში გადასვლას, რომელშიც საჭიროა დააწვეთ ღილაკს Summary: Descriptive statistics. გამონათდება ცხრილი გამოთვლების შედეგებით (ნახ. 6.7).

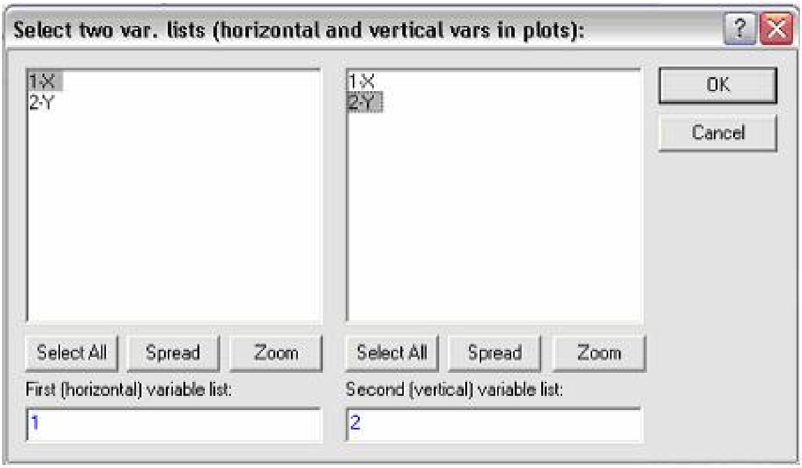


ნახ. 6.7.

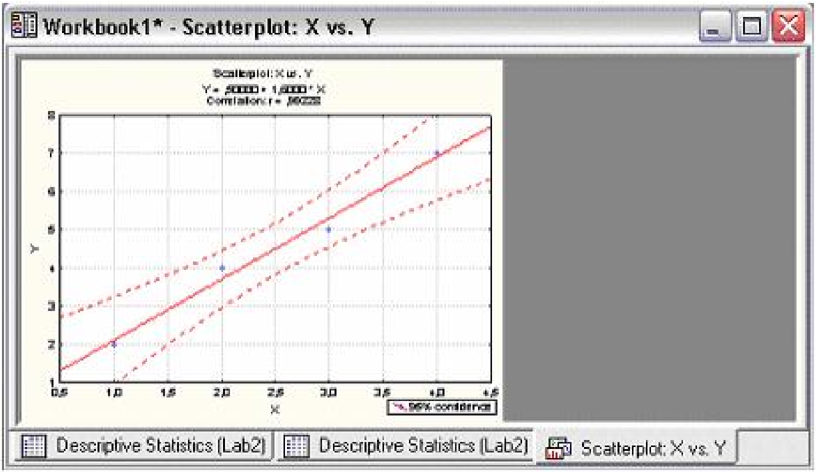
- დააწექით ღილაკს Descriptive statistics ან Statistics/ Resume (ნახ. 6.8). სრულდება გადასვლა წინა ფანჯარაში.

ნახ. 6.8.

- აირჩიეთ ჩანართი Prob.&Scatterplots, დააწექით ღილაკს 2D scatterplot. გამონათდება ფანჯარა ცვლადების შესარჩევად (ნახ. 6.9).
- შეასრულეთ საჭირო ცვლილებები და დააწექით OK ღილაკს.
- გამონათდება ფანჯარა (ნახ. 6.10) წრფივი რეგრესიული მოდელის გრაფიკით, მოდელის კოეფიციენტებისა და კორელაციის კოეფიციენტის გამოთვლილი მნიშვნელობებით. სამუშაოს პირველი ნაწილი შესრულებულია.



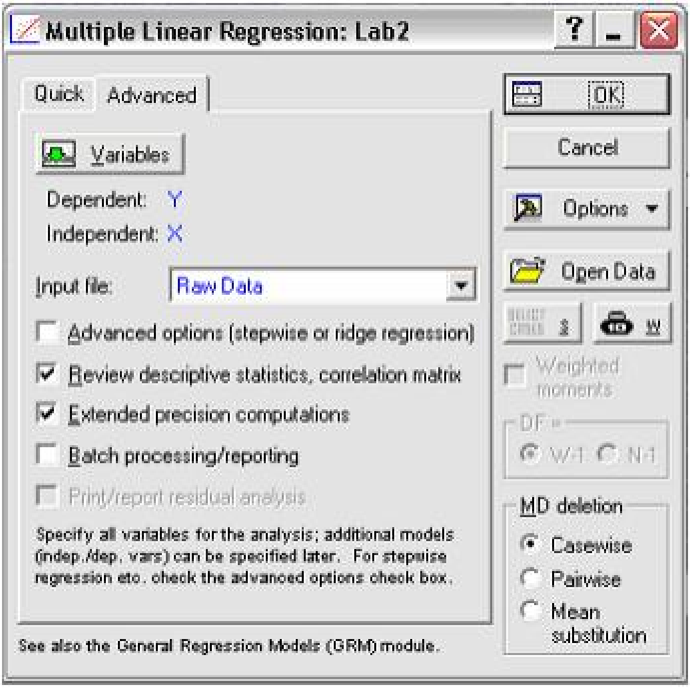
ნახ. 6.9.



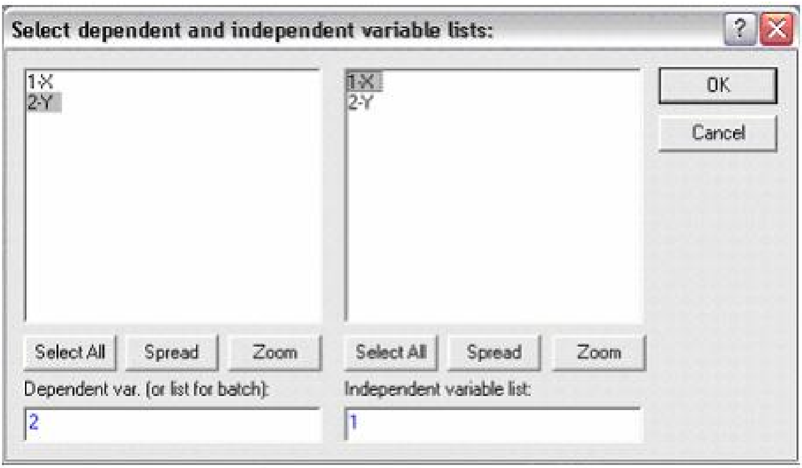
ნახ. 6.10.

შემდგომი გამოთვლები სრულდებიან Multiple Regression მოდულის გამოყენებით, რომელში გადასასვლელადაც შეასრულეთ ბრძანება Statistics/Multiple Regression. გამონათდება სასტარტო პანელი, რომელშიც უნდა შეასრულოთ შემდეგი მომართვები (ნახ. 6.11) და განსაზღვროთ დამოკიდებული (Dependent) და დამოუკიდებელი (Independent) ცვლადები (ნახ. 6.12), წინასწარ Variables ღილაკზე დაწოლით.

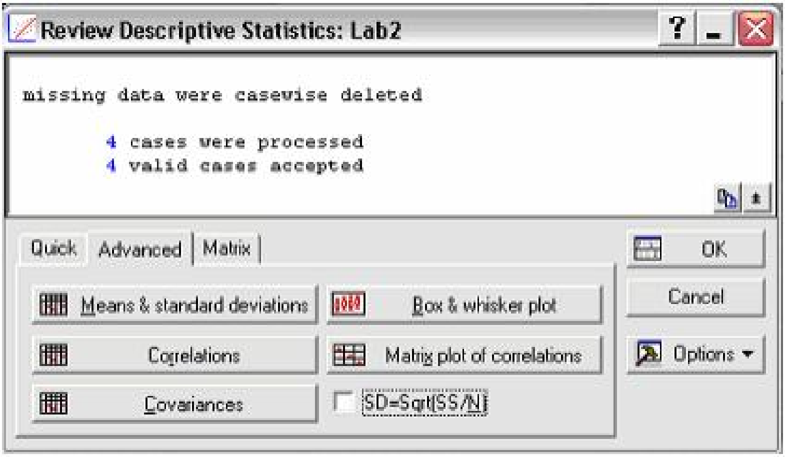
- ამის შემდეგ ფანჯარაში Multiple Linear Regression დააწექით OK-ის.
- გამონათდება ფანჯარა Review Descriptive Statistics (ნახ.6.13).
- განვსაზღვროთ ჩანაცვლებული საშუალო კვადრატული გადახრები. დააყენეთ ალამი ველში SD=Sums of Squares/N და დააწექით Means&Standard Deviations (ნახ. 6.14). ანალოგიურად შეიძლება გამოვითვალოთ ჩაუნაცვლებელი საშუალო კვადრატული გადახრები. ამ დროს ალამი არ უნდა იდგეს (ნახ. 6.15).



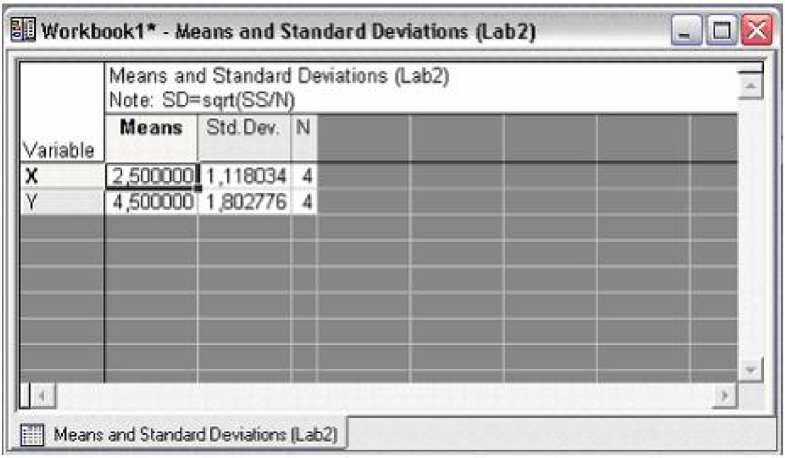
ნახ. 6.11.



ნახ. 6.12.

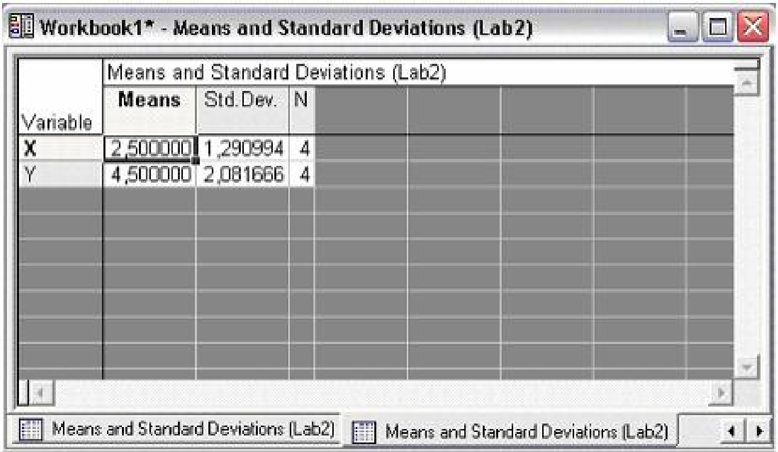


ნახ. 6.13.



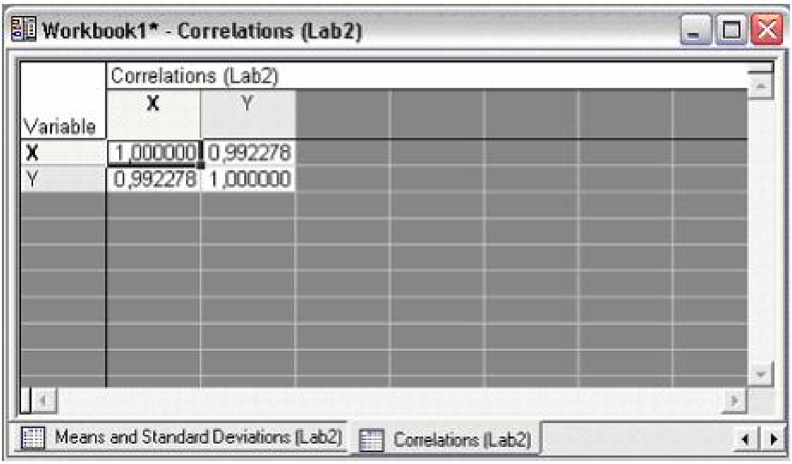
ნახ. 6.14.

- ფანჯარაში Review Descriptive Statistics კორელაციის კოეფიციენტის გამოსათვლელად დააწექით ღილაკს Correlations (ნახ. 6.16).

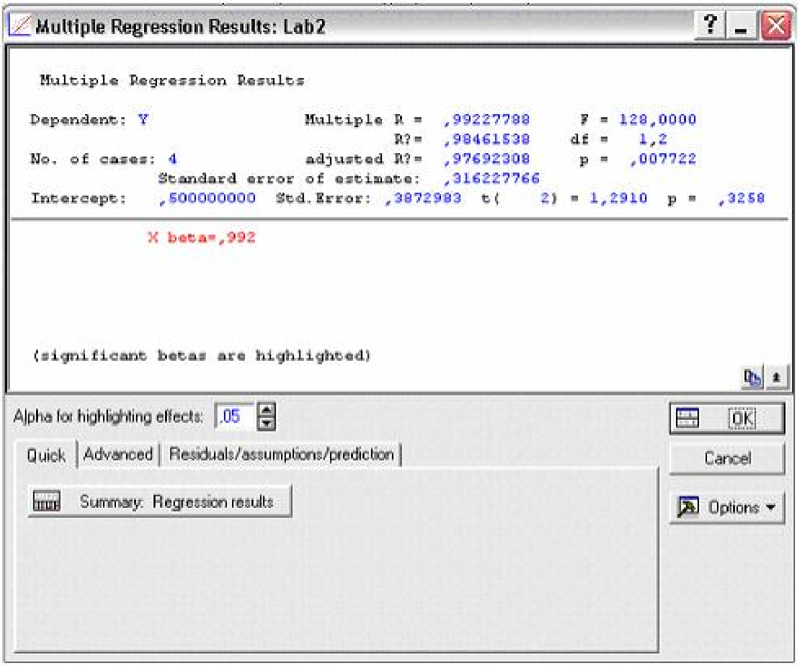


ნახ. 6.15.

- შემდეგ ისევ ფანჯარაში Review Descriptive Statistics დააწქით OK-ის. გამონათდება ფანჯარა (ნახ. 6.17).

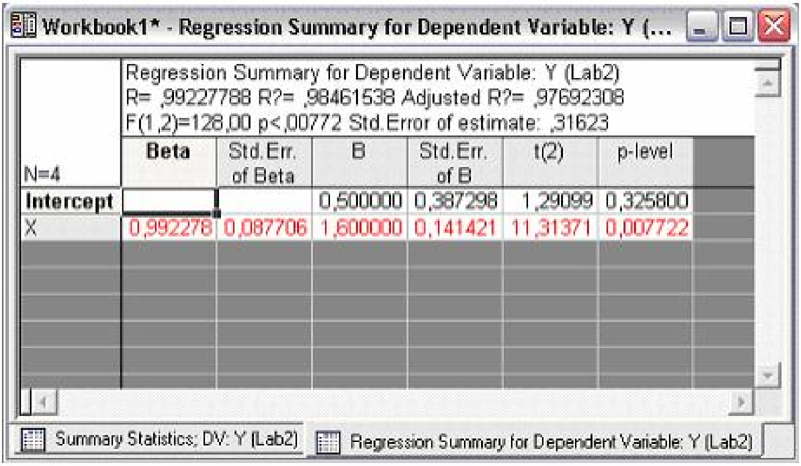


ნახ. 6.16.



ნახ. 6.17.

- აირჩიეთ ჩანართი Advansed და დააწექით ღილაკს Summary: Regression results მოდელის პარამეტრებისა და სტიუდენტის კოეფიციენტების გამოსათვლელად (ნახ. 6.18).
- შემდეგ ფანჯარაში Multiple Regression Results დააწექით ღილაკს ANOVA (Overall goodness of fit) ადექვატურობის ცხრილისა და შესაბამისი Q_R, Q_E, Q , k, n-k-1, F, S (ნახ. 6.19) გამოსათვლელად.
- ფანჯარაში Multiple Regression Results დააწექით OK-ის. გამონათდება ფანჯარა Residual Analysis (ნახ. 6.20).
- ჩანართში Advansed დააწექით Summary: Residuals&predicted ნარჩენების ანალიზისათვის (ნახ. 6.21).



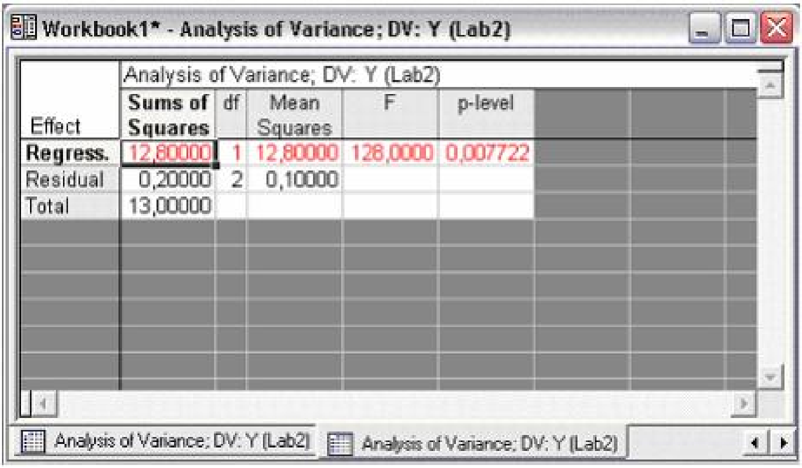
Workbook1* - Regression Summary for Dependent Variable: Y (...

Regression Summary for Dependent Variable: Y (Lab2)
R= ,99227788 R?= ,98461538 Adjusted R?= ,97692308
F(1,2)=128,00 p<,00772 Std.Error of estimate: ,31623

N=4	Beta	Std.Err. of Beta	B	Std.Err. of B	t(2)	p-level
Intercept			0,500000	0,387298	1,29099	0,325800
X	0,992278	0,087706	1,600000	0,141421	11,31371	0,007722

Summary Statistics; DV: Y (Lab2) | Regression Summary for Dependent Variable: Y (Lab2)

ნახ. 6.18.



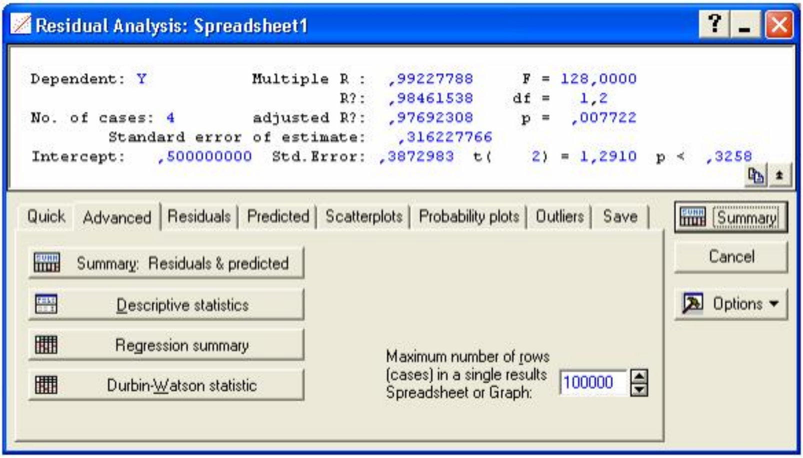
Workbook1* - Analysis of Variance; DV: Y (Lab2)

Analysis of Variance; DV: Y (Lab2)

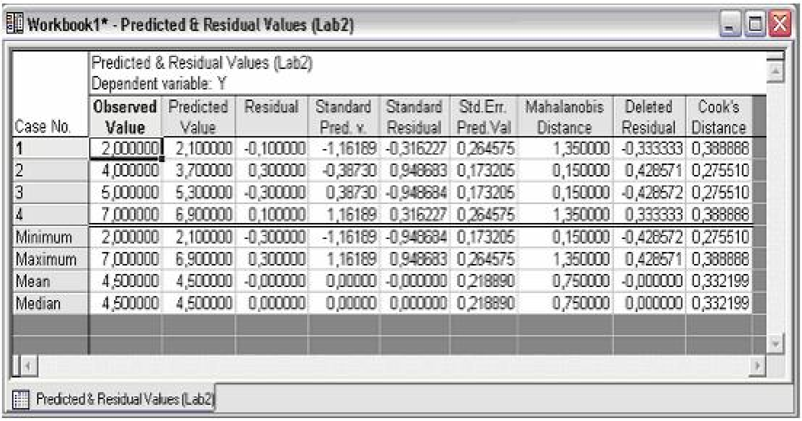
Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	12,80000	1	12,80000	128,0000	0,007722
Residual	0,20000	2	0,10000		
Total	13,00000				

Analysis of Variance; DV: Y (Lab2) | Analysis of Variance; DV: Y (Lab2)

ნახ. 6.19.



ნახ. 6.20.


Workbook1* - Predicted & Residual Values (Lab2)

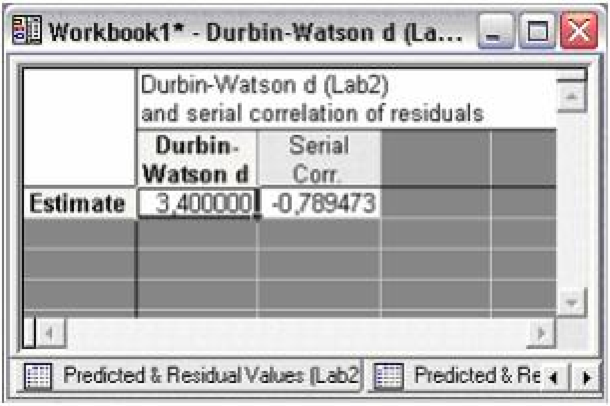
Predicted & Residual Values (Lab2)
Dependent variable: Y

Case No.	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual	Cook's Distance
1	2,000000	2,100000	-0,100000	-1,16189	-0,316227	0,264575	1,350000	-0,333333	0,388888
2	4,000000	3,700000	0,300000	-0,38730	0,948683	0,173205	0,150000	0,428571	0,275510
3	5,000000	5,300000	-0,300000	0,38730	-0,948684	0,173205	0,150000	-0,428572	0,275510
4	7,000000	6,900000	0,100000	1,16189	0,316227	0,264575	1,350000	0,333333	0,388888
Minimum	2,000000	2,100000	-0,300000	-1,16189	-0,948684	0,173205	0,150000	-0,428572	0,275510
Maximum	7,000000	6,900000	0,300000	1,16189	0,948683	0,264575	1,350000	0,428571	0,388888
Mean	4,500000	4,500000	-0,000000	0,00000	-0,000000	0,218890	0,750000	-0,000000	0,332199
Median	4,500000	4,500000	0,000000	0,00000	0,000000	0,218890	0,750000	0,000000	0,332199

Predicted & Residual Values (Lab2)

ნახ. 6.21.

- შემდეგ დააწექით ღილაკს Durbin – Watson statistic დარბინ-უოტსონის კრიტერიუმის გამოსათვლელად (ნახ. 6.22).
- დახურეთ პროგრამა STATISTICA-ს ფანჯარა.



ნახ. 6.22.

7. დისპერსიული ანალიზი

7.1. ერთფაქტორიანი დისპერსიული ანალიზი

დისპერსიული ანალიზის ამოცანას წარმოადგენს ერთი ან რამდენიმე ფაქტორის განსახილველ ნიშან-თვისებაზე ზემოქმედების შესწავლა.

ერთფაქტორიანი დისპერსიული ანალიზი გამოიყენება იმ შემთხვევებში, როცა ჩვენს განკარგულებაშია სამი ან მეტი დამოუკიდებელი შერჩევა, მიღებულნი ერთი გენერალური თანმიმდევრობიდან რომელიმე დამოუკიდებელი ფაქტორის ცვლილების გზით, რომლისთვისაც რაღაც მიზეზების გამო არ არსებობს რაოდენობრივი გაზომვები.

ამ შემთხვევებისათვის იგულისხმება, რომ მათ გააჩნიათ განსხვავებული შერჩეული საშუალოები და ერთნაირი შერჩეული დისპერსიები. ამიტომ საჭიროა ვუპასუხოთ კითხვას, მოახდინა თუ არა ამ ფაქტორმა მნიშვნელოვანი ზემოქმედება შერჩეულ საშუალოთა გაბნევაზე თუ გაბნევა წარმოადგენს იმ შემთხვევათა შედეგს, რომლებიც გამოწვეულია შერჩევათა მცირე მოცულობებით. სხვა სიტყვებით, თუ შერჩევები მიეკუთვნებიან ერთ და იგივე გენერალურ თანმიმდევრობას, მაშინ მონაცემთა გაბნევა შერჩევებს შორის (ჯგუფებს შორის) უნდა იყოს არა უმეტეს, ვიდრე მონაცემთა გაბნევა ამ შემთხვევათა შიგნით (ჯგუფებს შიგნით).

ვთქვათ x_{ik} –k–შერჩევის (k=$\overline{1,n}$) i-ური ელემენტია (i=$\overline{1,n_k}$), სადაც m –შერჩევათა რიცხვია, n_k - მონაცემთა რიცხვია k-ურ შერჩევაში. მაშინ $\bar{x}_k$ – k-ური საშუალოს შერჩეული საშუალო განისაზღვრება ფორმულით

$$\bar{x}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ik}.$$

საერთო საშუალო გამოითვლება ფორმულით

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{m}\sum_{i=1}^{n_k} x_{ik} \quad ,\text{სადაც } n = \sum_{k=1}^{m} n_k$$

დისპერსიული ანალიზის ძირითად იგივეობას აქვს შემდეგი სახე:

$$Q = Q_1 + Q_2,$$

სადაც Q_1 -შერჩეულ საშუალოთა $\bar{x}_k$ გადახრების კვადრატთა ჯამია $\bar{x}$ საერთო საშუალოდან (ჯგუფებს შორის გადახრების კვადრატთა ჯამი); Q_2 - დაკვირვებით მიღებულ მნიშვნელობათა x_{ik} გადახრების კვადრატების ჯამია $\bar{x}_k$ შერჩეული საშუალოდან (ჯგუფის შიგნით გადახრების კვადრატების ჯამია); Q - დაკვირვებით მიღე-

ბულ მნიშვნელობათა x_{ik} - საერთო საშუალოდან $\bar{x}$ გადახრების კვადრატების საერთო ჯამია.

გადახრების კვადრატთა ამ ჯამის გამოთვლა ხორციელდება შემდეგი ფორმულებით:

$$Q = \sum_{k=1}^{m}\sum_{i=1}^{n_k}(x_{ik} - \bar{x})^2 = \sum_{k=1}^{m}\sum_{i=1}^{n_k} x_{ik}^2 - n\cdot\bar{x}^2,$$

$$Q_1 = \sum_{k=1}^{m} n_k(\bar{x}_k - \bar{x})^2 = \sum_{k=1}^{m} n_k\bar{x}_k^2 - n\bar{x}^2,$$

$$Q_2 = \sum_{k=1}^{m} n_k(\bar{x}_k - \bar{x})^2 = \sum_{k=1}^{m} n_k\bar{x}_k^2 - n\bar{x}^2$$

კრიტერიუმად საჭიროა გამოვიყენოთ ფიშერის კრიტერიუმი:

$$F = \frac{Q_1/(m-1)}{Q_2/(n-m)}.$$

თუ ფიშერის კრიტერიუმის გამოთვლილი მნიშვნელობა ნაკლებია ვიდრე ცხრილური მნიშვნელობა $F_{\lambda\,;\,m-1;\,n-m}$, მაშინ არ არსებობს იმის საფუძველი, რომ ჩავთვალოთ რომ დამოუკიდებელი ფაქტორი ახდენს საშუალო მნიშვნელობის გაბნევაზე ზემოქმედებას, წინააღმდეგ შემთხვევაში დამოუკიდებელი ფაქტორი მნიშვნელოვან ზეგავლენას ახდენს საშუალო მნიშვნელობების გაბნევაზე (λ-მნიშვნელოვნების დონეა, რისკის დონეა, ჩვეულებრივ ეკონომიკური ამოცანებისათვის λ=0,05).

ერთფაქტორიანი ანალიზის ნაკლოვანება: იმ შემთხვევათა გამოყოფის შეუძლებლობა, რომლებიც განსხვავდებიან სხვებისაგან. ამ მიზნისათვის საჭიროა გამოვიყენოთ შეფფეს მეთოდი ან მოვახდინოთ შერჩევათა წყვილ-წყვილად შედარება.

მაგალითი 7.1. გამყიდველთა სამი ჯგუფი ყიდდა ცალობით საქონელს, დაფასოებულს სხვადასხვა საფუთავში. გაყიდვის ვადის გასვლის შემდეგ ჩაატარეს ტესტური კონტროლი ყოველი ჯგუფიდან შემთხვევითად ამორჩეულ გამყიდველებს შორის (ცხრ. 7.1).

ცხრილი 7.1.

ჯგუფის ნომერი	გამყიდველების მიერ ჩატარებული გაყიდვების რიცხვი, x_{ik}	გაყიდვათა საერთო რაოდენობა	გამყიდველთა რაოდენობა, n_k
1 2 3	1 3 2 1 0 2 1 2 3 2 1 4 - - 4 5 3 - - - -	10 12 12	7 5 3

თუ შემთხვევათა რიცხვი m=3, ხოლო გაყიდვათა რიცხვი ყველა შერჩევაში n=15, მაშინ $\bar{x}_1$=10/7=1,428; $\bar{x}_2$=12/5=2,4; $\bar{x}_3$=12/3=4; $\bar{x}$=(10+12+12)/15=2,226.

თუ

$$\sum_{k=1}^{m}\sum_{i=1}^{n_k} x_{ik}^2 = 1+9+4+1+4+1+4$$
$$+9+4+1+16+16+25+9=104$$
$$\sum_{k=1}^{m} n_k\,\bar{x}_k^2 = 7\cdot 1{,}428^2 + 5\cdot 2{,}4^2 + 3\cdot 4^2 = 91{,}074,$$

მაშინ

$$Q = 104 - 15\cdot 2{,}226^2 = 26{,}93;$$
$$Q_1 = 91{,}074 - 15\cdot 2{,}226^2 = 14{,}01;$$
$$Q_2 = Q - Q_1 = 26{,}93 - 14{,}01 = 12{,}92.$$

გამოვითვალოთ ფიშერის კრიტერიუმი

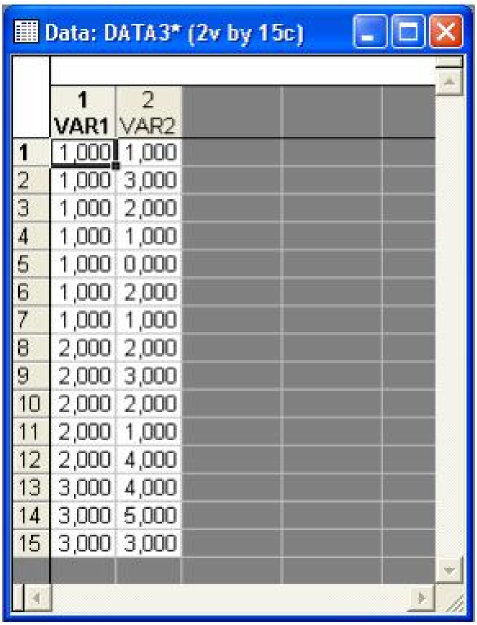
$$F = \frac{Q_1/(m-1)}{Q_2/(n-m)} = \frac{14{,}01/2}{12{,}92/12} = 6{,}52$$

ამ მნიშვნელობის ცხრილურთან შედარებით $F > F_{0{,}05\,;\,2\,;\,12}$ = 3,885 (იხ. დამატება 1), ვაკეთებთ დასკვნას, რომ შეფუთვა (განსაკუთრებით შელამაზებული!) გავლენას ახდენს გაყიდვების რაოდენობაზე.

7.2. საკონტროლო მაგალითის შედეგების შემოწმება ANOVA-ში

განვიხილოთ განხილული ამოცანის ამოხსნის პროცედურა დისპერსიული ანალიზის მეთოდით სისტემაში STATISTICA.

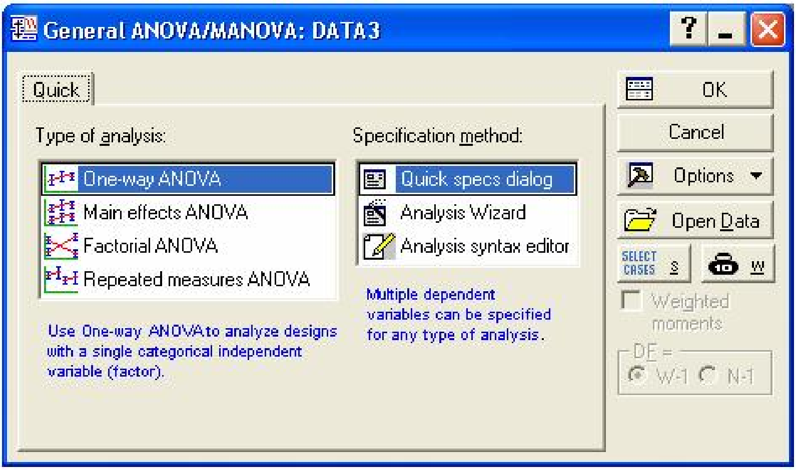
- გაუშვით პროგრამა STATISTICA ბრძანებით Start/All Programs/Statistica 7.0/ Statistics.
- გახსნილ ფანჯარაში დახურეთ დოკუმენტების ყველა ფანჯარა და შეასრულეთ ბრძანება File/New Data. ფანჯარაში Create New Document ეთითება შესაქმნელი სტრიქონებისა და სვეტების რიცხვი, მითითების გარეშე იქმნება მონაცემების შესატანად ცხრილი ზომებით (10 * 10).
- ცვლადებისათვის VAR 1 და VAR 2 სვეტებში შეიტანეთ საწყისი მონაცემები შემდეგი სახით (მოგვიხდება 5 cases დამატება) (ნახ. 7.1).



	1 VAR1	2 VAR2
1	1,000	1,000
2	1,000	3,000
3	1,000	2,000
4	1,000	1,000
5	1,000	0,000
6	1,000	2,000
7	1,000	1,000
8	2,000	2,000
9	2,000	3,000
10	2,000	2,000
11	2,000	1,000
12	2,000	4,000
13	3,000	4,000
14	3,000	5,000
15	3,000	3,000

ნახ. 7.1.

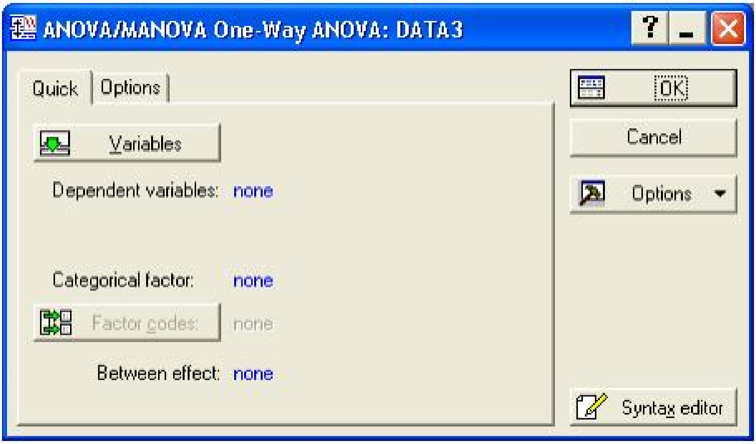
- მარჯვენა კლავიშის VAR 1 და VAR 2 სვეტებზე დაწკაპუნებით აირჩიეთ კონტექსტური მენიუ, გამოყავით პუნქტი Variable Spess ... და შეცვალეთ ცვლადების სახელები თუ არის ამის აუცილებლობა.
- File/Save As ბრძანებით დაიმახსოვრეთ მიღებული ფაილი.
- მენიუ Statistics-ში აირჩიეთ მოდული ANOVA. გამონათდება დიალოგური ფანჯარა General ANOVA/MANOVA. აირჩიეთ ანალიზის ტიპი One-wey ANOVA (ნახ. 7.2) და დააწექით OK-ის. პანელზე ANOVA/MANOVA (ნახ. 7.3). დააწექით ღილაკს Variables და განსაზღვრეთ კატეგორიალური ფაქტორი (VAR 1) და დამოკიდებული ცვლადი (VAR 2) (ნახ. 7.4). შემდეგ დააწექით OK-ის. გამონათებულ პანელზე დააწექით ღილაკს Factor codes და აირჩიეთ ყველა კოდი, მიაღწიეთ შედეგს (ნახ. 7.5). მოცემულ დიალოგურ ფანჯარაში დააწექით OK-ის.



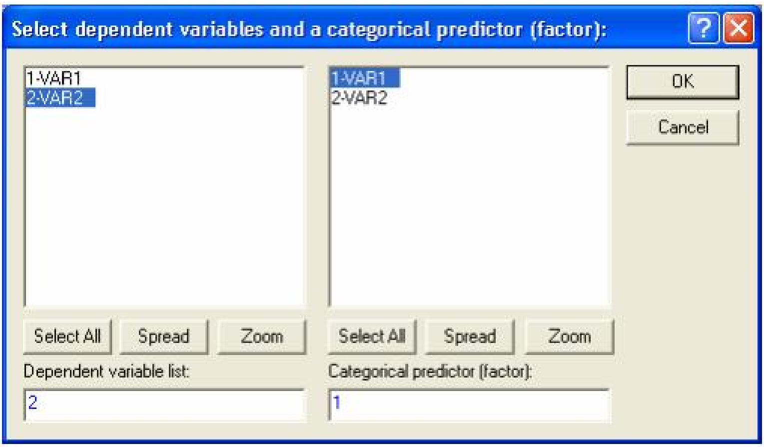
ნახ. 7.2.

- მოცემული ამოცანის ამოსახსნელად საკმარისია დააწვეთ ღილაკს All effects (ნახ. 7.6) და ეკრანზე გამონათდება საერთო დისპერსიული ანალიზის შედეგები (ნახ. 7.7). თუ ეს

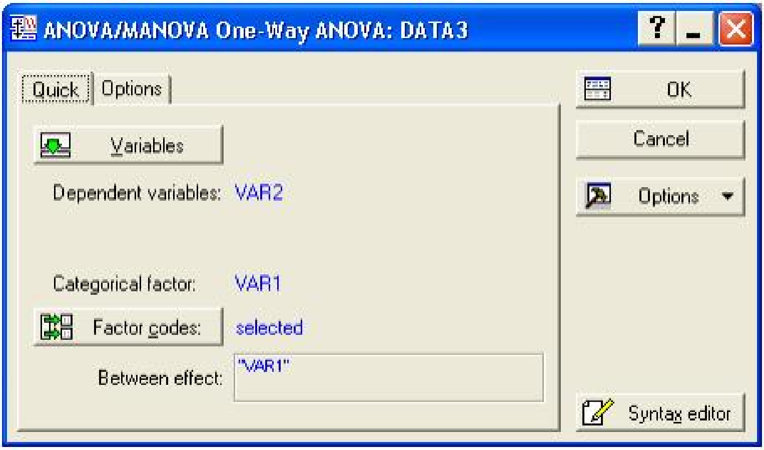
შედეგები გამოყოფილია წითელი ფერით, მაშინ ფაქტორი მნიშვნელოვან ზეგავლენას ახდენს, რასაც ჩვენ ვაკვირდე-ბით ეკრანზე. უფრო ზუსტი დასკვნის გაკეთება შეიძლება ფიშერის კრიტერიუმის გამოყენებით.



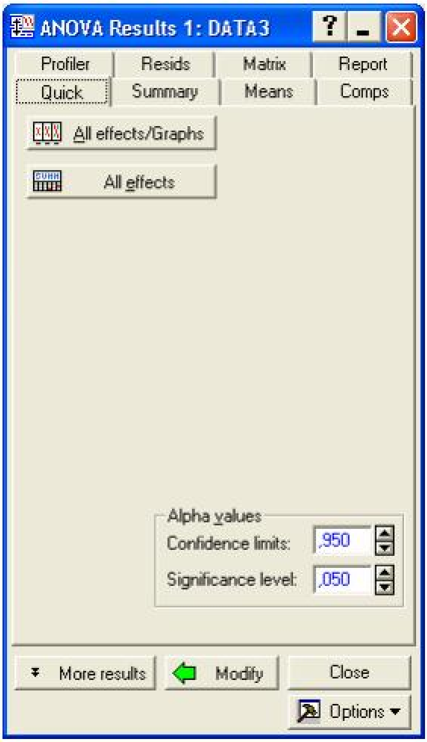
ნახ. 7.3.



ნახ. 7.4.


ANOVA/MANOVA One-Way ANOVA: DATA3
Quick
Options
OK
Cancel
Options
Variables
Dependent variables: VAR2
Categorical factor: VAR1
Factor codes: selected
Between effect: "VAR1"
Syntax editor

ნახ. 7.5.


ANOVA Results 1: DATA3
Profiler
Resids
Matrix
Report
Quick
Summary
Means
Comps
All effects/Graphs
All effects
Alpha values
Confidence limits: ,950
Significance level: ,050
More results
Modify
Close
Options

ნახ. 7.6.

Effect	Univariate Tests of Significance for Var2 (data1) Sigma-restricted parameterization Effective hypothesis decomposition SS	Degr. of Freedom	MS	F	p
Intercept	90,63501	1	90,63501	84,21837	0,000001
"Var1"	14,01905	2	7,00952	6,51327	0,012153
Error	12,91429	12	1,07619		

ნახ. 7.7.

7.3. მრავალფაქტორიანი დისპერსიული ანალიზი

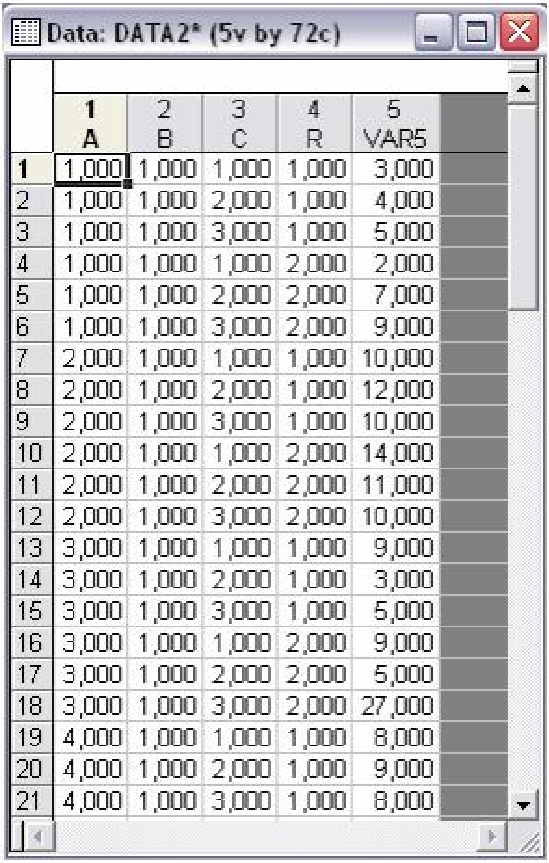
მაგალითი 7.2. გამყიდველთა ოთხი ჯგუფი ყიდდა ცალობით საქონელს. თითოეული ჯგუფი მომზადებული იყო განსაზღვრული მეთოდიკით (ფაქტორი A=1,2,3,4). საქონელი რეკლამირდებოდა ტელევიზიით, გაზეთში და რადიოთი (ფაქტორი B=1,2,3). გარდა ამისა, ის დაფასოებული იყო სხვადასხვა საფუთავში (ფაქტორი C=1,2,3). ექსპერიმენტი ორჯერ ჩატარდა (R=1,2 თამაშობს შემთხვევითი ფაქტორის როლს). გაყიდვათა ორი ვადის ამოწურვის შემდეგ მიღებულ იქნა გაყიდვების რაოდენობათა შესახებ შემდეგი შედეგები: (ცხრ. 7.2).

ცხრილი 7.2.

გამეო-რება		B1				B2				B3			
		A1	A2	A3	A4	A1	A2	A3	A4	A1	A2	A3	A4
R1	C1	3	10	9	8	24	8	9	3	2	8	9	8
R1	C2	4	12	3	9	22	7	16	2	2	2	7	2
R	C	5	1	5	8	2	9	1	3	2	8	6	3

1	3		0			3		7					
R2	C1	2	14	9	13	29	16	11	3	2	7	5	3
R2	C2	7	11	5	8	28	18	10	6	6	6	5	9
R2	C3	9	10	27	8	28	16	11	7	8	9	8	15

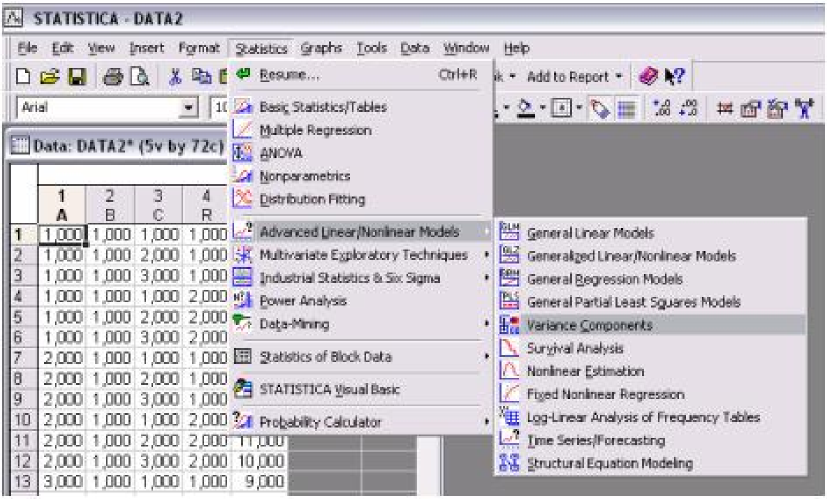
ცხრილის ფრაგმენტი საწყისი მონაცემებით მოყვანილია ნახ. 7.8-ზე.


Data: DATA2* (5v by 72c)

	1 A	2 B	3 C	4 R	5 VAR5
1	1,000	1,000	1,000	1,000	3,000
2	1,000	1,000	2,000	1,000	4,000
3	1,000	1,000	3,000	1,000	5,000
4	1,000	1,000	1,000	2,000	2,000
5	1,000	1,000	2,000	2,000	7,000
6	1,000	1,000	3,000	2,000	9,000
7	2,000	1,000	1,000	1,000	10,000
8	2,000	1,000	2,000	1,000	12,000
9	2,000	1,000	3,000	1,000	10,000
10	2,000	1,000	1,000	2,000	14,000
11	2,000	1,000	2,000	2,000	11,000
12	2,000	1,000	3,000	2,000	10,000
13	3,000	1,000	1,000	1,000	9,000
14	3,000	1,000	2,000	1,000	3,000
15	3,000	1,000	3,000	1,000	5,000
16	3,000	1,000	1,000	2,000	9,000
17	3,000	1,000	2,000	2,000	5,000
18	3,000	1,000	3,000	2,000	27,000
19	4,000	1,000	1,000	1,000	8,000
20	4,000	1,000	2,000	1,000	9,000
21	4,000	1,000	3,000	1,000	8,000

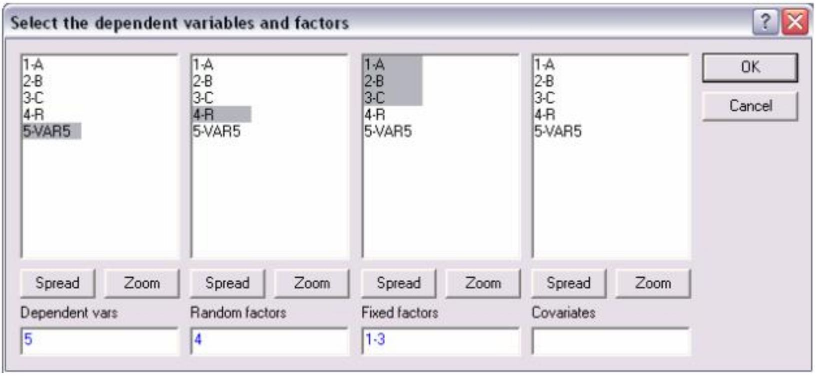
ნახ. 7.8.

- მენიუ Statistics-ში აირჩიეთ მოდული Advanced Linear/Nonlinear Models, მიუთითეთ ანალიზის ტიპი Variance Components (ნახ. 7.9).

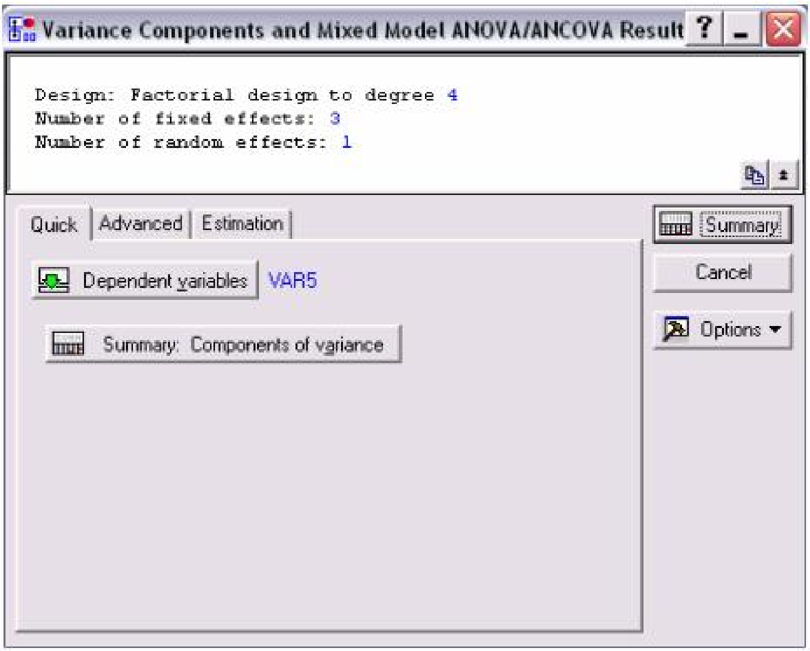


ნახ. 7.9.

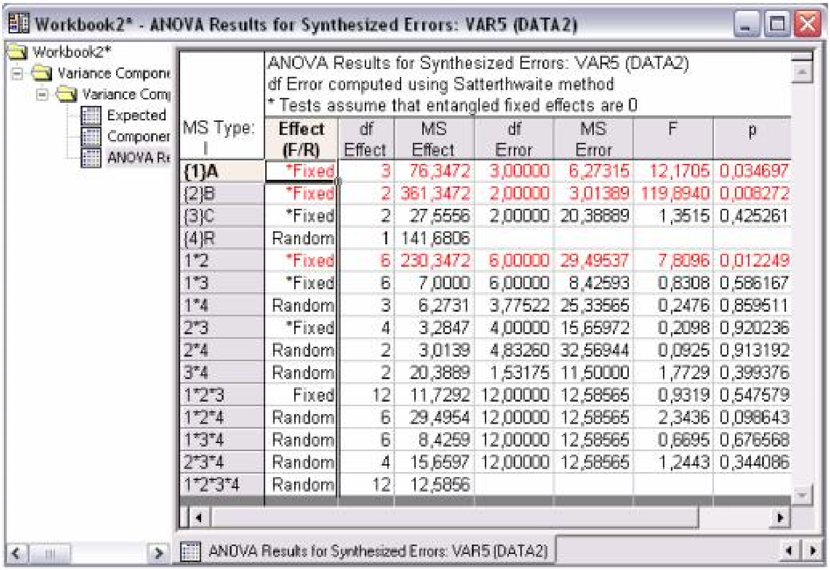
- დააწექით ღილაკს Variables და განსაზღვრეთ ცვლადები, როგორც ეს ნაჩვენებია ნახ. 7.10-ზე, შემდეგ დააწექით OK-ის.
- დააწექით OK-ის. გამონათდება პანელი Variance Components and Mixed Model ANOVA/MANOVA Resuts (ნახ. 7.11). მოცემული ამოცანის ამოსახსნელად საკმარისია დააწვეთ ღილაკს Summary, ეკრანზე გამონათდება საერთო დისპერსიული ანალიზის შედეგები (ნახ. 7.12). თუ ეს შედეგები გამოყოფილია წითელი ფერით, მაშინ ფაქტორი მნიშვნელოვან ზეგავლენას ახდენს. უფრო მკაცრად, ფიშერის კრიტერიუმის ან p-level მნიშვნელობის გამოყენებით, შეგვიძლია ვთქვათ, რომ გაყიდვათა რაოდენობაზე მნიშვნელოვან ზეგავლენას ახდენს ფაქტორი A (სწავლების მეთოდიკა), B (რეკლამა) და AB (მათი ერთობლივი მოქმედება).



ნახ. 7.10.



ნახ. 7.11.


Workbook2* - ANOVA Results for Synthesized Errors: VAR5 (DATA2)

ANOVA Results for Synthesized Errors: VAR5 (DATA2)
df Error computed using Satterthwaite method
* Tests assume that entangled fixed effects are 0

MS Type: I	Effect (F/R)	df Effect	MS Effect	df Error	MS Error	F	p
{1}A	*Fixed	3	76,3472	3,00000	6,27315	12,1705	0,034697
{2}B	*Fixed	2	361,3472	2,00000	3,01389	119,8940	0,008272
{3}C	*Fixed	2	27,5556	2,00000	20,38889	1,3515	0,425261
{4}R	Random	1	141,6806				
1*2	*Fixed	6	230,3472	6,00000	29,49537	7,8096	0,012249
1*3	*Fixed	6	7,0000	6,00000	8,42593	0,8308	0,586167
1*4	Random	3	6,2731	3,77522	25,33565	0,2476	0,859511
2*3	*Fixed	4	3,2847	4,00000	15,65972	0,2098	0,920236
2*4	Random	2	3,0139	4,83260	32,56944	0,0925	0,913192
3*4	Random	2	20,3889	1,53175	11,50000	1,7729	0,399376
1*2*3	Fixed	12	11,7292	12,00000	12,58565	0,9319	0,547579
1*2*4	Random	6	29,4954	12,00000	12,58565	2,3436	0,098643
1*3*4	Random	6	8,4259	12,00000	12,58565	0,6695	0,676568
2*3*4	Random	4	15,6597	12,00000	12,58565	1,2443	0,344086
1*2*3*4	Random	12	12,5856				

ნახ. 7.12.

8. კლასტერული ანალიზი

8.1.ზოგადი ცნობები

კლასტერული ანალიზი მრავლობითი ანალიზის ერთ-ერთი მეთოდია, რომლის დანიშნულებაა ელემენტების ერთობლიობის დაჯგუფება (კლასტერიზაცია), რომლებიც ხასიათდებიან მრავალი ფაქტორით, და ერთგვაროვანი ჯგუფების (კლასტერების) მიღება. კლასტერებად დაყოფა ხდება რაღაც მეტრიკის დახმარებით, მაგალითად ევკლიდეს ზომით. კლასტერული ანალიზის ამოცანა მდგომარეობს ელემენტების შესახებ საწყისი ინფორმაციის შეკუმშული სახით წარმოდგენაში მისი მნიშვნელოვანი დანაკარგების გარეშე.

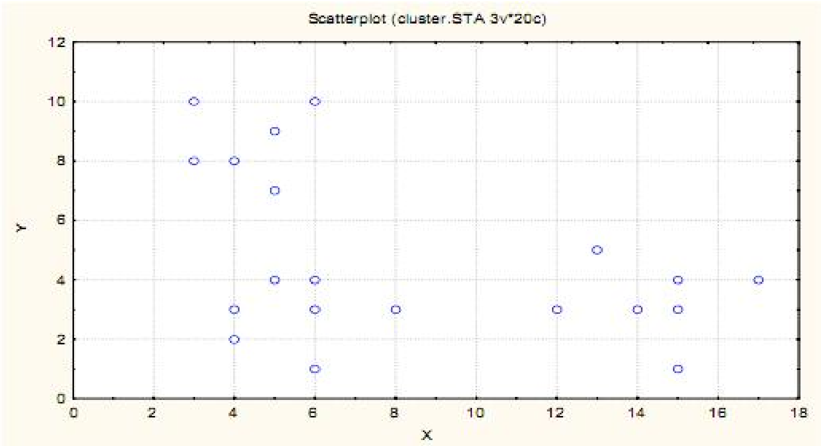
მაგალითი 8.1. განვიხილოთ აბსტრაქტული მაგალითი. ოცმა ბანკმა, რომელთა აქციებიც კოტირდებიან ბაზარზე, წარმოადგინეს შემდეგი ინფორმაცია (ცხრ. 8.1), სადაც X - გასულ პერიოდში გაწეული დანახარჯებია, Y- გასულ პერიოდში მიღებული მოგებაა. საჭიროა გამოვარკვიოთ თუ რომელი ბანკის აქციების შეძენას აქვს აზრი (Buy), რომლები შევინახოთ (Hold), ხოლო რომლები მოვიცილოთ (Sell).

საწყისი მონაცემების გრაფიკული წარმოდგენა მოყვანილია ნახ. 8.1-ზე. ჩვენ წინასწარ გავიადვილეთ სამუშაო ჩვენთვის და პროგრამული პაკეტი STATISTICA-სთვის ამ მონაცემების კლასტერებად დაყოფით. სულაც არაა აუცილებელი გაგვაჩნდეს ცოდნა კლასტერული ანალიზის სფეროში, რომ გადავჭრათ ეს ამოცანა. ბანკების პირველი (1,20,7,16,8,17,11) და მეორე ჯგუფების (2,15,6,3,12,19) შედარებით შეგვიძლია ვთქვათ, რომ მეორე ჯგუფი უმჯობესია, ვინაიდან ერთი და იგივე დანახარჯების დროს მეორე ჯგუფი ღებულობს უფრო მეტ მოგებას. ბანკების პირველი ჯგუფის მესამესთან (4,13,5,9,10,14,18) შედარებით ვასკვნით რომ უმჯობესია პირველი

ჯგუფი, ვინაიდან ერთი და იგივე მოგების შემთხვევაში მისი დანახარჯები უფრო მცირეა.

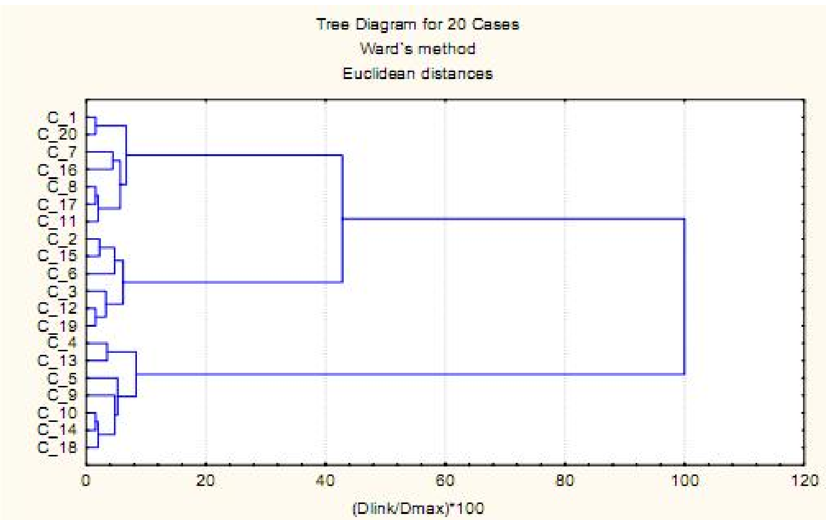
ცხრილი 8.1

ბანკის ნომერი	დანახარჯები X	მოგება Y	რეკომენდაცია
1	4	2	Hold
2	6	10	Buy
3	5	7	Buy
4	12	3	Sell
5	17	4	Sell
6	3	10	Buy
7	6	1	Hold
8	6	3	Hold
9	15	1	Sell
10	15	4	Sell
11	5	4	Hold
12	3	8	Buy
13	13	5	Sell
14	15	3	Sell
15	5	9	Buy
16	8	3	Hold
17	6	4	Hold
18	14	3	Sell
19	4	8	Buy
20	4	3	Hold



ნახ. 8.1.

ანალოგიური შედეგები მივიღეთ პაკეტ STATISTICA-სა და მისი მოდულის Cluster Analysis გამოყენებით (ნახ. 8.2).



ნახ. 8.2.

ოღონდ პრაქტიკაში ეს ასე მარტივი არაა, როგორც ერთი შეხედვით ჩანს. კლასტერული ანალიზის ამოცანების ამოხსნის დროს გვიხდება რიგ პრობლემებთან შეჯახება:

- ელემენტები (ჩვენს შემთხვევაში ბანკები) ხასიათდებიან ფაქტორების დიდი რაოდენობით, რომლებსაც გააჩნიათ გაზომვის განსხვავებული ერთეულები და განსხვავებული აბსოლუტური სიდიდეები, სრულიად შეუდარებლებია ერთმანეთთან და მოაქვთ ინფორმაციის განსხვავებული მოცულობა;
- თავდაპირველად უცნობია კლასტერთა ის რიცხვი, რომელზეც საჭიროა დავყოთ ელემენტთა საწყისი ერთობლიობა, და მრავალგანზომილებიან შემთხვევაში ვიზუალურ დაკვირვებებს უბრალოდ არ მივყავართ წარმატებამდე;
- რომელი მეტრიკები გამოვიყენოთ ელემენტებს შორის მანძილის (სიახლოვის ზომის) ზომად;
- რომელი მიზნობრივი ფუნქცია ან მეთოდი გამოვიყენოთ ელემენტების კლასტერებში გასაერთიანებლად.

8.2.მონაცემების სტანდარტიზაცია

კლასტერულ ანალიზში კლასტერებად დაყოფა არსებითადაა დამოკიდებული საწყისი მონაცემების აბსოლუტური მნიშვნელობებისაგან. ამ პრობლემას წყვეტენ სტანდარტიზაციის (ნორმირების) დახმარებით. ამისათვის თითოეული ფაქტორის ყველა მნიშვნელობიდან აკლებენ ამ ფაქტორის შერჩეულ საშუალოს და მიღებულ შედეგებს ყოფენ საშუალო კვადრატულ გადახრაზე:

$$x^* = \frac{x - \bar{x}}{S_H},$$

სადაც x - საწყისი მონაცემებია; $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ - შერჩეული საშუალოა; $S_H = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ - საშუალო კვადრატული გადახრაა.

ამასთან სტანდარტიზებულ მნიშვნელობებს ექნებათ ნულის ტოლი შერჩეული საშუალო, ხოლო შერჩეული დისპერსიები ერთის ტოლია. სხვა სიტყვებით, ჩვენ ყველა ფაქტორი დავიყვანეთ ერთ წონით კატეგორიაში, როგორც მოჭიდავეები შეჯიბრების წინ. ამ ოპერაციის განსახორციელებლად პაკეტ STATISTICA-ში საჭიროა გამოვიძახოთ მოდული Data Management.

8.3. კლასტერული ანალიზის მეთოდები

იგულისხმება, საწყის მონაცემთა მატრიცას აქვს სახე X[n,k], სადაც n- რაოდენობაა, k- ფაქტორთა რაოდენობა. ამიტომ ელემენტთა კლასტერიზაციის დროს პაკეტ STATISTICA-ში საჭიროა ავირჩიოთ რეჟიმი: cases (rows) - სტრიქონები, ხოლო ფაქტორების კლასტერიზაციისას: variables (columns) - სვეტები.

ანალიზის ძირითად მეთოდად პაკეტი STATISTICA გვთავაზობს Joining (tree clustering-ს - იერარქიული მეთოდების ჯგუფს (7 სახე), რომლებიც იმ შემთხვევაში გამოიყენებიან, თუ კლასტერების რიცხვი წინასწარ უცნობია, და K-Means Clastering (K-საშუალოთა მეთოდი), რომელშიც მომხმარებელი წინასწარ განსაზღვრავს კლასტერების რაოდენობას.

იერარქიული მეთოდების ჯგუფში შედის Ward's method- უორდის მეთოდი, რომელიც კარგად მუშაობს ელემენტთა მცირე რაოდენობასთან და გამიზნულია ისეთი კლასტერების შერჩევაზე, რომლებსაც წევრების დაახლოებით ერთნაირი რაოდენობა აქვთ.

დაცილების მეტრიკებად პაკეტი გვთავაზობს სხვადასხვა ზომებს, მაგრამ ყველაზე მეტად გამოყენებადებს წარმოადგენენ Euclidean distance (ევკლიდეს ზომა):

$$\rho(x_i, x_z) = \sqrt{\sum_{m=1}^{k} (x_{im} - x_{lm})^2},$$

სადაც i,z=1,2,3,…,n;

ან Squared Euclidean distance (კვადრატული ევკლიდური ზომა):

$$\rho(x_i, x_z) = \sum_{m=1}^{k} (x_{im} - x_{lm})^2$$

რაც შეეხება მიზნის ფუნქციას, ყველაზე უფრო გავრცელებულ მიზნის ფუნქციას წარმოადგენს შიგა ჯგუფური კვადრატების ჯამი. ასეთი მიზნის ფუნქციის გამოყენებისას კლასტერული ანალიზის ალგორითმი შეიძლება შემდეგზე დავიდეს: თუ გვაქვს n ელემენტი და მათ შორის მანძილთა მატრიცა, ჯერ ითვლება, რომ თითოეული ელემენტი წარმოადგენს ცალკეულ კლასტერს. შემდეგ თითოეულ ბიჯზე ერთიანდება ისეთი ორი კლასტერი, რომლებსაც მივყავართ მიზნის ფუნქციის მინიმალურ ზრდასთან.

კლასტერული ანალიზის ალგორითმების მრავალფეროვნება ხშირად ახდენს მომხმარებლის დეზორიენტაციას, ამიტომ მან შეიძლება გამოიყენოს რამდენიმე ალგორითმი და უპირატესობა მიანიჭოს რომელიმე დასკვნას სამუშაოს შედეგების ერთობლიობის კომპლექსური შეფასების საფუძველზე.

8.4. საკონტროლო მაგალითის შედეგების შემოწმება Claster Analysis-ში.

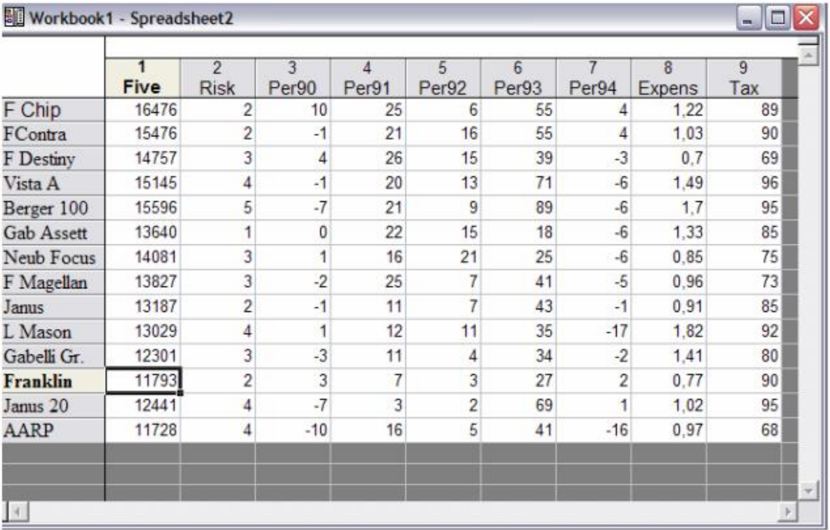
მაგალითი 8.2. ამ მაგალითში გამოკვლეული იქნება 16 ცნობილი ინვესტიციური ფონდი მათი მდგომარეობის შეფასების მიზნით. ცვლადებად გამოიყენებიან შემდეგი მახასიათებლები: შემოსავლიანობა ხუთწლიანი პერიოდის განმავლობაში, რისკი, შემოსავლის ყოველწლიური პროცენტი ყოველი წლის მიხედვით, გასავლის ნაწილი და საგადასახადო რეიტინგები. საწყისი მონაცემები ინვესტიციური ფონდების შესახებ მოყვანილია ცხრ. 8.2-ში.

ცხრილი 8.2.

Fund	Five	Risk	Per90	Per91	Per92	Per93	Per94	Expens	Tax	Recom
Chip	16476	2	10	25	6	55	4	1,22	89	Buy
FContra	15476	2	-1	21	16	55	4	1,03	90	Buy
F Destiny	14757	3	4	26	15	39	-3	0,7	69	Buy
Vista A	15145	4	-1	20	13	71	-6	1,49	96	Hold
Berger 100	15596	5	-7	21	9	89	-6	1,7	95	Hold
Gab Assett	13640	1	0	22	15	18	-6	1,33	85	Buy
Neub Focus	14081	3	1	16	21	25	-6	0,85	75	Buy
F Magellan	13827	3	-2	25	7	41	-5	0,96	73	Buy
Janus	13187	2	-1	11	7	43	-1	0,91	85	Sell
L Mason	13029	4	1	12	11	35	-17	1,82	92	Hold
Gabelli Gr.	12301	3	-3	11	4	34	-2	1,41	80	Sell
Franklin	11793	2	3	7	3	27	2	0,77	90	Sell
Janus 20	12441	4	-7	3	2	69	1	1,02	95	Sell
AARP	11728	4	-10	16	5	41	-16	0,97	68	Sell
Kemper	11386	4	-6	2	-2	67	4	1,09	86	Sell
20th Cent Gr	11258	4	-8	15	-4	32	0	1	60	Sell

- გაუშვით პაკეტი STATISTICA.
- გახსენით ახალი პროექტი. შეასრულეთ ბრძანება File/New. ცხრილის ზომებად მიუთითეთ 9 სვეტი (Vars-ცვლადი) და 14 სტრიქონი (Cases-შემთხვევები).
  - გამონათდება ელექტრონული ცხრილი Workbook: Spreadsheet საწყისი მონაცემების შესატანად და მათ გარდასაქმნელად.

- შეიტანეთ საწყისი მონაცემები ცვლადებისათვის სვეტებში VAR 1 და VAR 9 შემდეგი სახით (საჭიროა 6 Cases დამატება) (ნახ. 8.3).

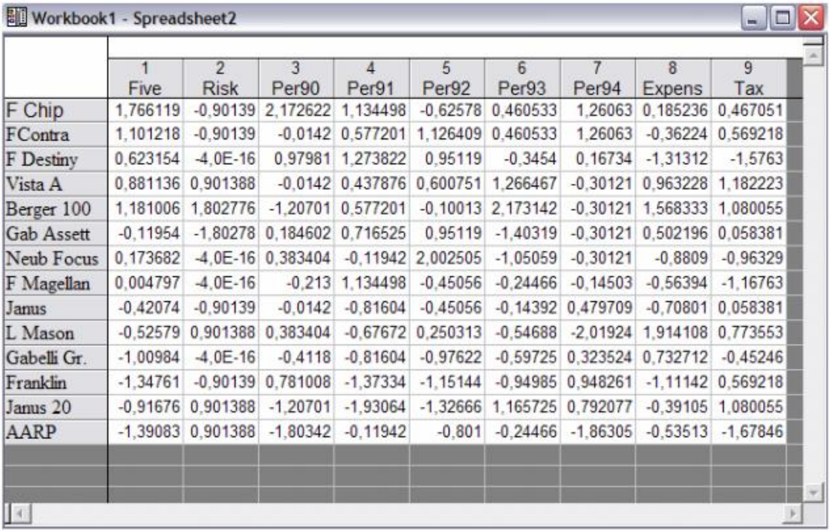


Workbook1 - Spreadsheet2

	1 Five	2 Risk	3 Per90	4 Per91	5 Per92	6 Per93	7 Per94	8 Expens	9 Tax
F Chip	16476	2	10	25	6	55	4	1,22	89
FContra	15476	2	-1	21	16	55	4	1,03	90
F Destiny	14757	3	4	26	15	39	-3	0,7	69
Vista A	15145	4	-1	20	13	71	-6	1,49	96
Berger 100	15596	5	-7	21	9	89	-6	1,7	95
Gab Assett	13640	1	0	22	15	18	-6	1,33	85
Neub Focus	14081	3	1	16	21	25	-6	0,85	75
F Magellan	13827	3	-2	25	7	41	-5	0,96	73
Janus	13187	2	-1	11	7	43	-1	0,91	85
L Mason	13029	4	1	12	11	35	-17	1,82	92
Gabelli Gr.	12301	3	-3	11	4	34	-2	1,41	80
Franklin	11793	2	3	7	3	27	2	0,77	90
Janus 20	12441	4	-7	3	2	69	1	1,02	95
AARP	11728	4	-10	16	5	41	-16	0,97	68

ნახ. 8.3.

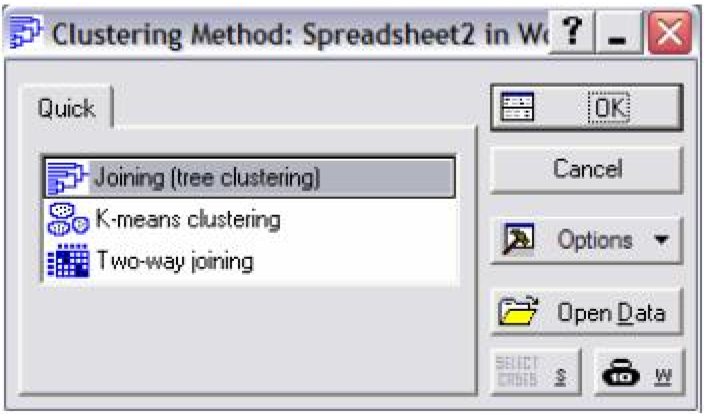
- შემდეგ მიუთითეთ ბრძანება Data/Standartize … . აირჩიეთ ყველა ცვლადი და ყველა სტრიქონი. მივიღებთ საწყისი მონაცემების სტანდარტიზებულ მნიშვნელობებს (ნახ. 8.4).
- შეასრულეთ ბრძანება Statistics/Myltivariate Exploratory Techniques/Cluster Analysis. გამონათებულ Clustering Method ფანჯარაში აირჩიეთ Joining (tree clustering). დააწექით OK-ის (ნახ. 8.5).



Workbook1 - Spreadsheet2

	1 Five	2 Risk	3 Per90	4 Per91	5 Per92	6 Per93	7 Per94	8 Expens	9 Tax
F Chip	1,766119	-0,90139	2,172622	1,134498	-0,62578	0,460533	1,26063	0,185236	0,467051
FContra	1,101218	-0,90139	-0,0142	0,577201	1,126409	0,460533	1,26063	-0,36224	0,569218
F Destiny	0,623154	-4,0E-16	0,97981	1,273822	0,95119	-0,3454	0,16734	-1,31312	-1,5763
Vista A	0,881136	0,901388	-0,0142	0,437876	0,600751	1,266467	-0,30121	0,963228	1,182223
Berger 100	1,181006	1,802776	-1,20701	0,577201	-0,10013	2,173142	-0,30121	1,568333	1,080055
Gab Assett	-0,11954	-1,80278	0,184602	0,716525	0,95119	-1,40319	-0,30121	0,502196	0,058381
Neub Focus	0,173682	-4,0E-16	0,383404	-0,11942	2,002505	-1,05059	-0,30121	-0,8809	-0,96329
F Magellan	0,004797	-4,0E-16	-0,213	1,134498	-0,45056	-0,24466	-0,14503	-0,56394	-1,16763
Janus	-0,42074	-0,90139	-0,0142	-0,81604	-0,45056	-0,14392	0,479709	-0,70801	0,058381
L Mason	-0,52579	0,901388	0,383404	-0,67672	0,250313	-0,54688	-2,01924	1,914108	0,773553
Gabelli Gr.	-1,00984	-4,0E-16	-0,4118	-0,81604	-0,97622	-0,59725	0,323524	0,732712	-0,45246
Franklin	-1,34761	-0,90139	0,781008	-1,37334	-1,15144	-0,94985	0,948261	-1,11142	0,569218
Janus 20	-0,91676	0,901388	-1,20701	-1,93064	-1,32666	1,165725	0,792077	-0,39105	1,080055
AARP	-1,39083	0,901388	-1,80342	-0,11942	-0,801	-0,24466	-1,86305	-0,53513	-1,67846

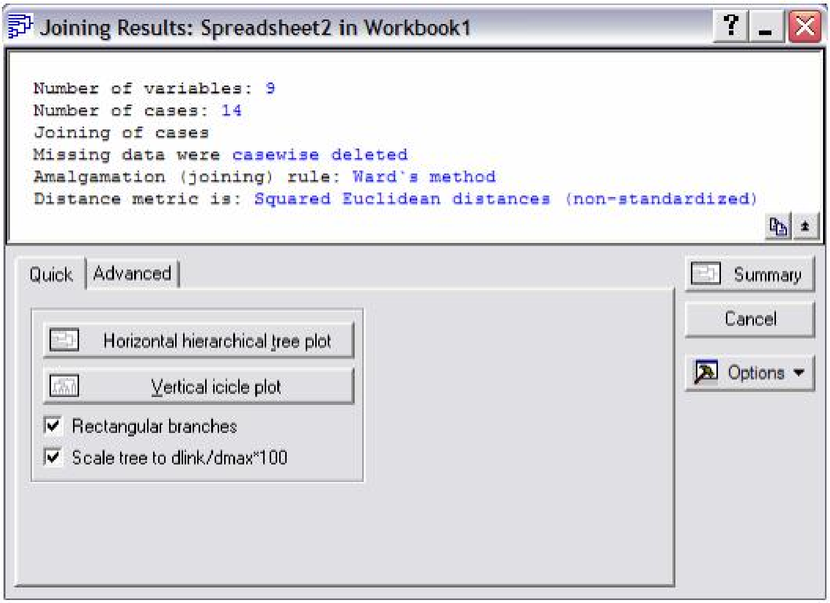
ნახ. 8.4.



ნახ. 8.5.

- განსაზღვრეთ ყველა ცვლადი, მეთოდი და მანძილის ზომა (ნახ. 8.6). დააწექით OK-ის.
- გამონათდება შესაბამისი დიალოგის ფანჯარა, რომელშიც საჭიროა განვსაზღვროთ გრაფიკის განლაგება (ვერტიკალური ან ჰორიზონტალური) (ნახ. 8.7). დააწექით Summary-ს.

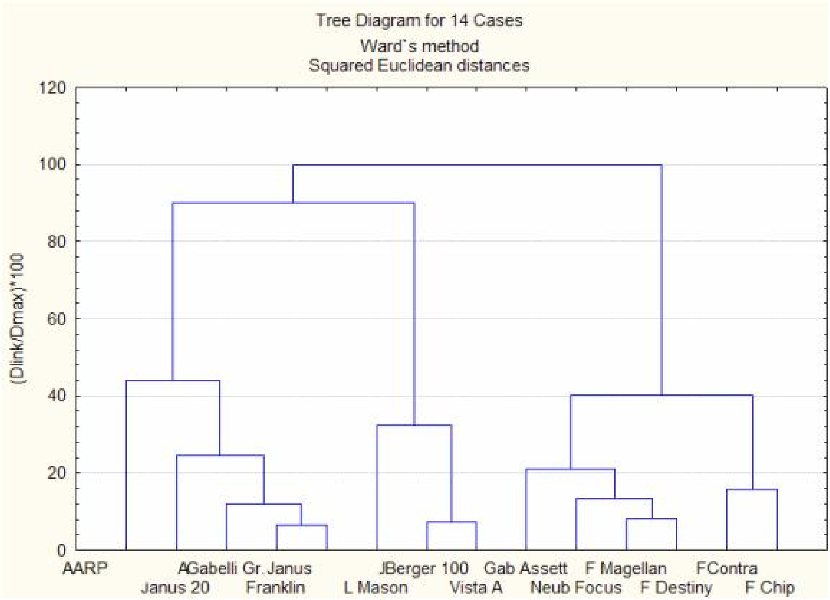
ნახ. 8.6.



ნახ. 8.7.

გრაფიკზე (ნახ. 8.8) ხშირად ჩნდება კლასტერთა სამი ჯგუფი. ამ კლასტერთა ანალიზი (რომელი ინვესტიციური ფონდები მივაკუ-

თვნოთ პერსპექტიულებს, რომლები - არაპერსპექტიულებს) მოით-ხოვს სპეციალურ ცოდნას ამ სფეროში. შეიძლება ვისარგებლოთ შემ-დეგი მოსაზრებებით:



ნახ. 8.8.

პირველ კლასტერში (მარჯვენაში) ჩანს, რომ დანახარჯები გო-ნივრული იყო: 1990 წელს მცირე შემოსავლებისას მომდევნო წლებ-ში ამ კლასტერის ფონდების მდგომარეობა მნიშვნელოვნად გაუმ-ჯობესდა. რისკის არამაღალი რეიტინგისას საგადასახადო მოსაკ-რებლები აგრეთვე საკმაოდ დაბალი იყო, მიზანშეწონილია ამ ფონ-დების აქციათა შესყიდვა;

მეორე კლასტერში (საშუალოში) ყველაზე დიდი დანახარჯე-ბი იყო, თუმცა ხუთდღიან პერიოდში შემოსავლები მაღალი იყო. რისკის შეფასება და საგადასახადო მოსაკრებლები აღმოჩნდნენ მაქ-სიმალურები ყველა კლასტერს შორის, ე.ი. ამ ფონდების აქციები სა-ჭიროა შევინახოთ;

მესამე კლასტერის შესახებ შეგვიძლია ვთქვათ, რომ მას უჭირავს მეორე ადგილი დანახარჯებით შემოსავლების მიმართ ხუთდღიანი პერიოდის განმავლობაში. რისკის შეფასება ყველაზე მაღალია, მაგრამ საგადასახადო მოსაკრებლები მნიშვნელოვნად დაბალია, ვიდრე აქვს პირველ კლასტერს, ამიტომ მიზანშეწონილია ამ ფონდების აქციათა გაყიდვა.

9. დისკრიმინანტული ანალიზი

9.1. ზოგადი ცნობები

რაში მდგომარეობს კლასტერულ და დისკრიმინანტულ ანალიზებს შორის განსხვავება?

კლასტერული ანალიზის დანიშნულება ისაა, რომ დააჯგუფოს ელემენტები ერთგვაროვან ჯგუფებში (კლასტერებში). ეს ერთგვაროვნება განისაზღვრება ნიშან-თვისებათა (ფაქტორების) საფუძველზე, რომლებიც ერთვებიან პარამეტრებად კლასტერულ ანალიზში. წინასწარ ცნობილი არაა ჯგუფების რაოდენობა. არაა შემაჯამებელი ნიშან-თვისება ან დამოკიდებული ცვლადი. კლასტერული ანალიზი ხშირად გამოიყენება ბაზრის აპოსტერიალური სეგმენტაციისთვის.

დისკრიმინანტული ანალიზი ცოტა სხვაგვარად მოქმედებს. განიხილება რაღაც „დამოკიდებული“ ცვლადი, რომელიც განსაზღვრავს ჩვენს აზრს (ექსპერტის აზრს) მოსალოდნელი დაჯგუფების მიმართ. შემდეგ განისაზღვრებიან წრფივი კლასიფიკაციური მოდელები, რომლებიც საშუალებას იძლევიან „ვიწინასწარმეტყველოთ“ ახალი ელემენტის ქცევა დამოკიდებული ცვლადის ტერმინებში რიგი იმ დამოუკიდებელი ცვლადების (ფაქტორების, მაჩვენებლების) გაზომვის საფუძველზე, რომლებითაც ისინი ხასიათდებიან.

მაგალითად, არის მომხმარებლის ლოიალურობის სამი დონე საქონლის განსაზღვრული მარკის მიმართ და გვაქვს მისი ცხოვრების სტილის რიგი მაჩვენებლების გაზომვის შედეგები. ვაგებთ წრფივ მოდელებს, რომლებშიც მნიშვნელობების ჩასმამ სტილური ცვლადებიდან შეიძლება გაგვცეს პასუხი მომხმარებლის ლოიალურობის შესახებ მოცემული საქონლის მიმართ. ეს მოდელი უფრო ინფორმაციულია, ვინაიდან იძლევა „ზემოქმედების სიძლიერის“ შესახებ წარმოდგენას. დისკრიმინანტული ანალიზი გამოიყენება ბაზრის აპრიორული სეგმენტაციისათვის.

9.2. საკონტროლო მაგალითის შედეგების შემოწმება DISCRIMINANT ANALYSIS-ში

განვიხილოთ პრაქტიკული ამოცანის ამოხსნის პროცედურა დისკრიმინანტული ანალიზის მეთოდით სისტემაში STATISTICA. გავარჩიოთ დისკრიმინანტული ანალიზის ჩატარების პრინციპი (უფრო სწორად, შემსწავლელი შერჩევების ფორმირება) ცხრ. 9.1-ში წარმოდგენილი მონაცემების საფუძველზე.

ცხრილი 9.1.

	X1	X2	X3	X4	X5	X6	X7	CLASS1
1	-107.000	5868.000	531.000	450.000	63.000	22.300	1608.000	1.000
2	-903.000	6330.000	636.000	401.000	69.000	17.600	1768.000	1.000
3	-18.000	6793.000	620.000	487.000	104.000	19.400	1775.000	2.000
4	1.300	4731.000	447.000	405.000	64.000	10.400	979.000	2.000
5	403.100	2969.000	382.000	274.000	29.000	5.700	728.000	3.000
6	-205.000	4924.000	284.000	292.000	35.000	17.500	1010.000	3.000
7	-256.000	7622.000	342.000	223.000	26.000	14.100	634.000	3.000
8	-2142.00	4318.000	257.000	151.000	33.000	16.500	985.000	4.000
9	-1394.00	3140.000	218.000	241.000	47.000	8.500	592.000	4.000
10	-1571.00	4617.000	171.000	137.000	13.000	13.100	484.000	4.000
11	-728.300	5448.000	348.000	215.000	28.000	5.700	367.000	4.000
12	-1796.00	2902.000	161.000	182.000	22.000	11.400	631.000	4.000
13	-1955.20	3634.000	334.000	361.000	59.000	10.100	925.000	4.000
14	-1294.00	3499.000	204.000	129.000	27.000	6.800	398.000	4.000
15	-1500.00	6368.000	288.000	169.000	27.000	13.300	601.000	4.000
16	-1879.00	3058.000	169.000	86.000	23.000	5.600	307.000	5.000
17	-197.000	5110.000	82.000	57.000	11.000	1.100	174.000	5.000
18	-2310.70	4166.000	207.000	183.000	32.000	9.800	487.000	5.000
19	-1437.00	5168.000	151.000	96.000	8.000	10.700	359.000	5.000
20	-482.000	2061.000	78.000	47.000	4.000	2.900	110.300	5.000

ცხრილში მოყვანილია მონაცემები სოფლის მეურნეობის 20 საწარმოს შესახებ, რომლებიც ამორჩეული და მიკუთვნებული იქნენ შესაბამის ჯგუფებზე ექსპერტული ხერხით.

კლასიფიკაციაში მონაწილე მაჩვენებლები შემდეგია:

X1 - მოგება (ათასი ლარი);

X2 - მთლიანი პროდუქცია 1 მუშაკზე (ათასი ლარი);

X3 - მთლიანი პროდუქცია 1 ჰა სავარგულზე (ათასი ლარი);

X4 - რძის წარმოება 1 ჰა სავარგულზე (კგ);

X5 - ხორცის წარმოება 1 ჰა სავარგულზე (კგ);

X6 - პროდუქციიდან ამონაგები 1 მუშაკზე (ათასი ლარი);

X7 - ამონაგები 1 ჰა სავარგულზე (ათასი ლარი).

-გაუშვით პაკეტი STATISTICA.

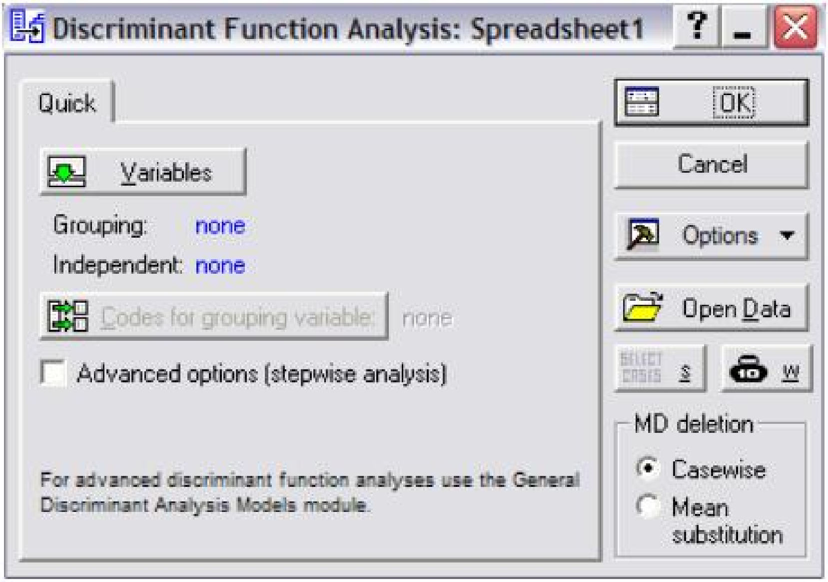
-გახსენით ახალი პროექტი. შეასრულეთ ბრძანება File/New. ცხრილის ზომად მიუთითეთ 8 სვეტი (Vars-ცვლადი) და 20 სტრიქონი (Cases-შემთხვევები).

-გამონათდება ელექტრონული ცხრილი Workbook: Spreadsheet საწყისი მონაცემების შესატანად და მათ გარდასაქმნე-

ლად. შეიტანეთ საწყისი მონაცემები ცვლადებისათვის სვეტებში VAR 1 და VAR 8. მიზანშეწონილია შევცვალოთ ცვლადების სახელები.

-შეასრულეთ ბრძანება Statistics/Myltivariate Exploratory Techniques/Discriminant Analysis.

-ეკრანზე გამონათდება Discriminant Function Analysis (დისკრიმინანტული ფუნქციების ანალიზი) მოდულის სასტარტო პანელი (ნახ. 9.1), რომელშიც ღილაკი Variables საშუალებას გვაძლევს ავირჩიოთ Grouping (დამაჯგუფებელი ცვლადი) და Independent (დამოუკიდებელი ცვლადები). Codes for grouping variable (კოდები ცვლადების ჯგუფებისათვის)

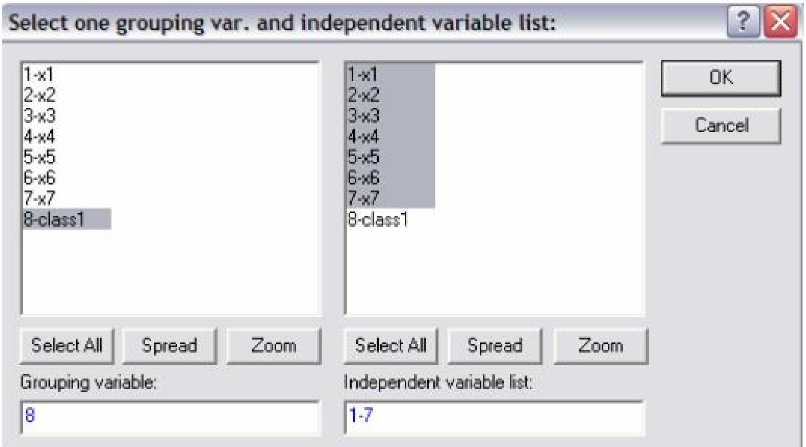


ნახ. 9.1.

უთითებენ გასაანალიზირებელი ობიექტების ჯგუფების რაოდენობას. MD (Missing data-გამოტოვებული ცვლადები) საშუალებას იძლევა სიიდან წაიშალოს სტრიქონული ცვლადები, ან შეცვალონ ისინი საშუალო მნიშვნელობებზე. File /Open ხნის ფაილს

მონაცემებით. შეგვიძლია მივუთითოთ მონაცემთა ბაზებიდან დაკვირვებების არჩევის პირობები ღილაკით Select Cases და ცვლადების წონები სიიდან ღილაკ W-ს ამორჩევით.

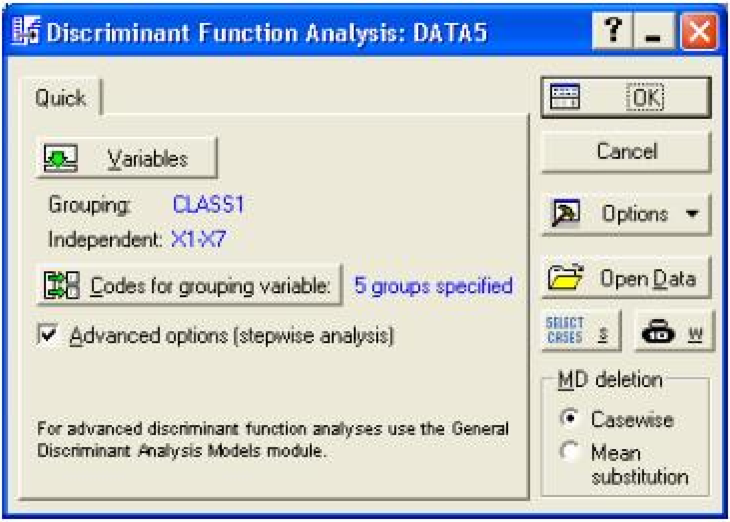
-ღილაკ Variables-ზე მაუსის დაწკაპუნებით იხსნება ცვლადების არჩევის დიალოგის ფანჯარა (ნახ. 9.2).



ნახ. 9.2.

მარცხენა ნაწილში ირჩევა დამაჯგუფებელი ცვლადები, მარჯვენაში კი დასაჯგუფებელი ცვლადები. ცვლადების სახელები მარცხენა და მარჯვენა ნაწილებში არ უნდა იკვეთებოდნენ. მოცემულ მაგალითში დამაჯგუფებელ ცვლადად არჩეულია ცვლადი CLASS 1, ხოლო დასაჯგუფებელ ცვლადებად X1 – X7. Select All (გამოიყოს ყველა) გამოყოფს ყველა ცვლადს, Spread (წვრილმანები) - გრძელი სახელის დათვალიერებისთვისაა, Zoom (ინფორმაცია ცვლადის შესახებ) საშუალებას იძლევა დავათვალიეროთ ინფორმაცია ცვლადის შესახებ: მისი სახელი, რიცხვითი მნიშვნელობის ფორმატი, აღწერითი სტატისტიკები (ნომერი ჯგუფში, საშუალო მნიშვნელობა, სტატისტიკური გადახრა). ღილაკ Variables-ზე მაუსის დაწკაპუნებით ვირჩევთ დამაჯგუფებელ (Grouping) ცვლადად CLASS 1, ხოლო დამოუკიდებელ (Independent) ცვლადებად- X1 – X7. დააწკაპუნეთ

მაუსის მაჩვენებელი ღილაკს Codes for grouping variables და ALL ღილაკზე დაწკაპუნებით აირჩიეთ ყველა კოდი დამაჯგუფებელი ცვლადისათვის. შესაბამისი არჩევებისა და OK-ზე დაწკაპუნების შემდეგ Discriminant Function Analysis ფანჯარა უნდა იყოს ისე წარმოდგენილი, როგორც ეს ნაჩვენებია ნახ. 9.3-ზე.



ნახ. 9.3.

OK ღილაკზე მაუსის დაწკაპუნების შემდეგ გაიღება დიალოგის ფანჯარა Model Definition (მოდელის განსაზღვრა) (ნახ. 9.4).

დიალოგის ფანჯარაში Model Definition შემოთავაზებულია მნიშვნელოვან ცვლადთა ამორჩევის მეთოდის არჩევა. Method-ად შეიძლება შემოთავაზებულ იქნეს Standart (სტანდარტული), Forward stepwise (ბიჯური ჩართვით) და Backward stepwise (ბიჯური გამორიცხვით). ჩანართი Descriptive (აღწერითი სტატისტიკები) საშუალებას გვაძლევენ მივიღოთ:

Pooled within-groups covariances&correlations (გაერთიანებული შიგაჯგუფური კოვარიაციები და კორელაციები);

Means&number of cases (საშუალო მნიშვნელობა თითოეული ცვლადისთვის);

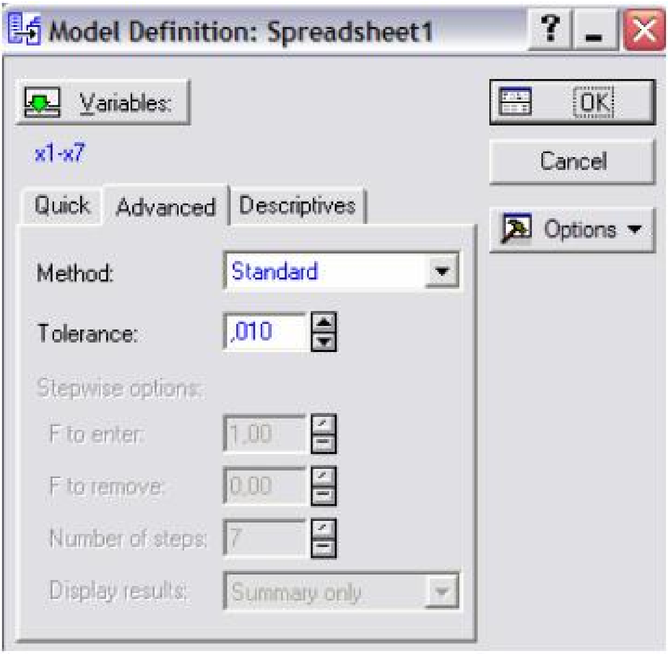
Within-groups standart deviations (თითოეულ ჯგუფში ცვლადების სტანდარტული გადახრები);

Categjrized histogram by group (კატეგორიზებული ჰისტოგრამები ჯგუფების მიხედვით თითოეული ცვლადისთვის);

Box plot of means by group (გაბნევის დიაგრამები ცალკე აღებული ჯგუფებისთვის);

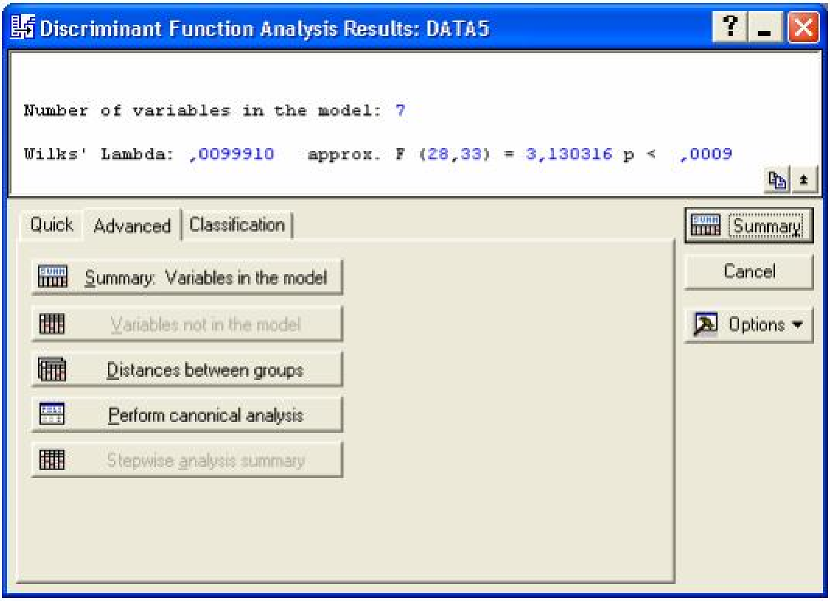
Categorized scatterplot by group (ნებისმიერი ორი ცვლადისთვის);

Categorized normal probability plot by group (კატეგორიზებული ნორმალური გრაფიკი ნებისმიერი ცვლადისთვის ჯგუპების მიხედვით).



ნახ. 9.4.

მეთოდად (Method) ავირჩიოთ Standard და დავაწვეთ OK-ის. გამოთვლების მსვლელობაში სისტემით მიღებული შედეგები წარმოდგენილებია ფანჯარაში Discriminant Function Analisis Results (დისკრიმინანტული ფუნქციების ანალიზი) (ნახ. 9.5).



ნახ. 9.5.

9.3. შედეგების გამოტანა და მისი ანალიზი

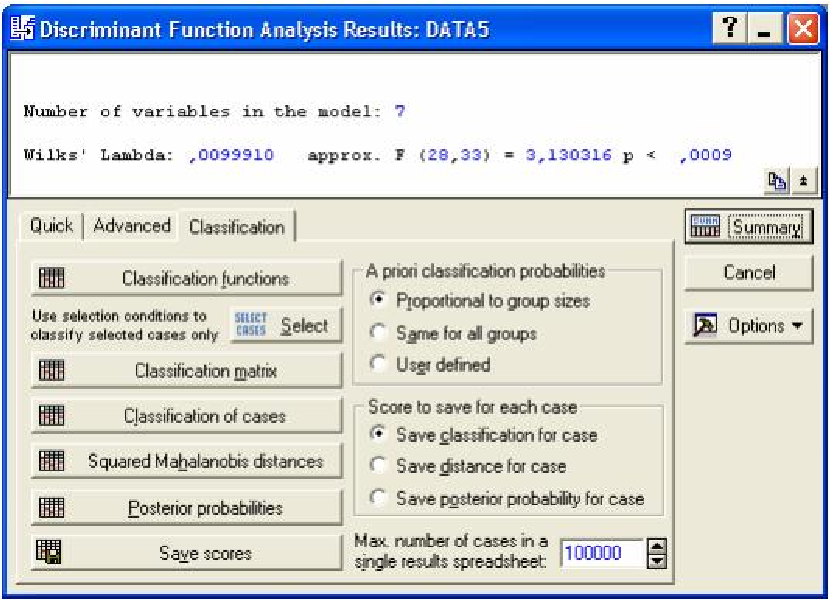
დიალოგური ფანჯრის Discriminant Function Analisis Results (დისკრიმინანტული ფუნქციების ანალიზი) ინფორმაციული ნაწილი გვატყობინებს, რომ:

- Number of variables in the model (მოდელში ცვლადების რაოდენობა) =7;

- Wilks lambda (უილკსის ლამბდის მნიშვნელობა) =0,0099910;
- Approx. F(28, 33) (F-სტატისტიკის მიახლოებითი მნიშვნელობა, დაკავშირებული უილკსის ლამბდასთან) =3,130316;
- Fკრიტერიუმის P-მნიშვნელოვნების დონე 3,130316 მნიშვნელობისათვის.

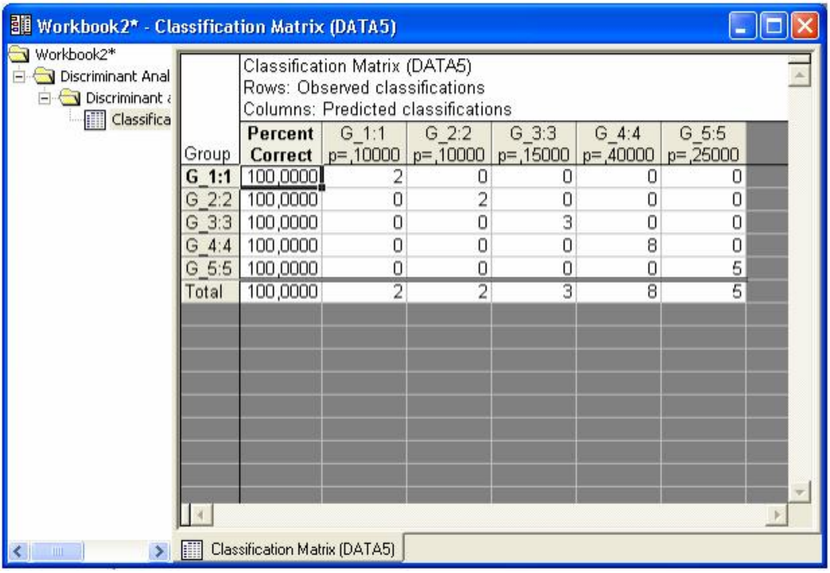
უილკსის სტატისტიკის მნიშვნელობა მოთავსებულია [0 , 1] ინტერვალში. 0-თან ახლოს მდგომი უილკსის სტატისტიკის მნიშვნელობები მეტყველებენ კარგ დისკრიმინაციაზე, ხოლო 1-თან ახლოს მყოფები - ცუდ დისკრიმინაციაზე.

0,00999-ს ტოლი Wilks lambda-ს (უილკსის ლამბდის მნიშვნელობის) და 3,1303-ის ტოლი F-კრიტერიუმის მნიშვნელობების მიხედვით შეიძლება დავასკვნათ, რომ მოცემული კლასიფიკაცია პრაქტიკულად კორექტულია. ავირჩიოთ ჩანართი Classification (ნახ. 9.6).



ნახ. 9.6.

შემსწავლელი შერჩევების კორექტულობის შესასწავლად დავათვალიეროთ კლასიფიკაციური მატრიცის შედეგები ჩანართ Classification-ში ღილაკზე Classification matrix დაწკაპუნებით (ნახ. 9.7). მანამდე წინასწარ უნდა ავირჩიოთ Propotional to group sizes ფანჯარა Discriminant Function Analysis Results-ის მარჯვენა ნაწილში.



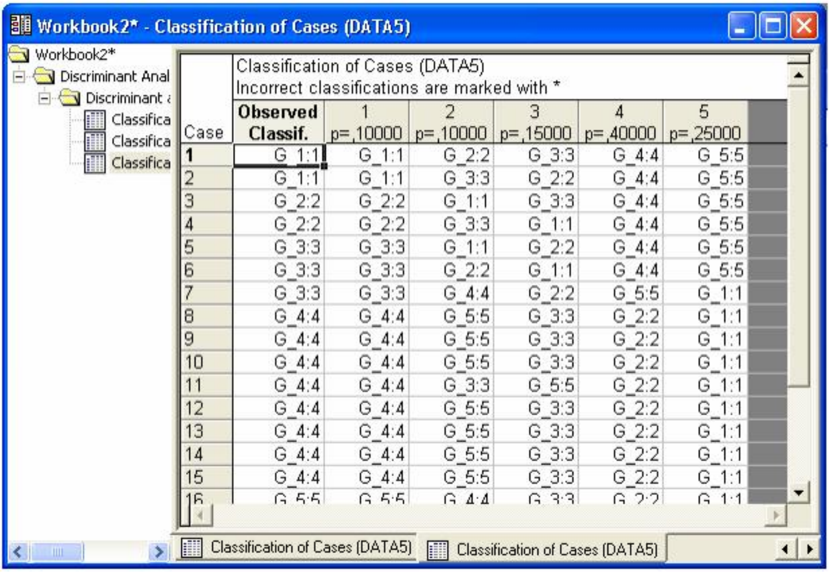
Group	Percent Correct	G_1:1 p=,10000	G_2:2 p=,10000	G_3:3 p=,15000	G_4:4 p=,40000	G_5:5 p=,25000
G_1:1	100,0000	2	0	0	0	0
G_2:2	100,0000	0	2	0	0	0
G_3:3	100,0000	0	0	3	0	0
G_4:4	100,0000	0	0	0	8	0
G_5:5	100,0000	0	0	0	0	5
Total	100,0000	2	2	3	8	5

ნახ. 9.7.

კლასიფიკაციური მატრიციდან შეიძლება დავასკვნათ, რომ ობიექტები ექსპერტული წესით სწორად იყვნენ მიკუთვნებულნი გამოყოფილ ჯგუფებზე. მაგრამ თუ არის შესაბამის ჯგუფებზე არასწორად მიკუთვნებული წარმოებები, მაშინ შეიძლება დავათვალიეროთ Classification of cases (შემთხვევათა კლასიფიკაცია) (ნახ. 9.8).

შერჩევათა კლასიფიკაციის ცხრილში არაკორექტულად მიკუთვნებული წარმოებები აღინიშნებიან ვარსკვლავით (*). ამრიგად, კორექტული შემსწავლელი შერჩევების მიღების ამოცანა იმაში მდგომარეობს, რომ შემსწავლელი შერჩევებიდან გამოვრიცხოთ ის

ობიექტები, რომლებიც თავისი მახასიათებლებით არ შეესაბამებიან ერთგვაროვანი ჯგუფის შემქმნელი წარმოებების უმრავლესობას.



Classification of Cases (DATA5)
Incorrect classifications are marked with *

Case	Observed Classif.	1 p=,10000	2 p=,10000	3 p=,15000	4 p=,40000	5 p=,25000
1	G_1:1	G_1:1	G_2:2	G_3:3	G_4:4	G_5:5
2	G_1:1	G_1:1	G_3:3	G_2:2	G_4:4	G_5:5
3	G_2:2	G_2:2	G_1:1	G_3:3	G_4:4	G_5:5
4	G_2:2	G_2:2	G_3:3	G_1:1	G_4:4	G_5:5
5	G_3:3	G_3:3	G_1:1	G_2:2	G_4:4	G_5:5
6	G_3:3	G_3:3	G_2:2	G_1:1	G_4:4	G_5:5
7	G_3:3	G_3:3	G_4:4	G_2:2	G_5:5	G_1:1
8	G_4:4	G_4:4	G_5:5	G_3:3	G_2:2	G_1:1
9	G_4:4	G_4:4	G_5:5	G_3:3	G_2:2	G_1:1
10	G_4:4	G_4:4	G_5:5	G_3:3	G_2:2	G_1:1
11	G_4:4	G_4:4	G_3:3	G_5:5	G_2:2	G_1:1
12	G_4:4	G_4:4	G_5:5	G_3:3	G_2:2	G_1:1
13	G_4:4	G_4:4	G_5:5	G_3:3	G_2:2	G_1:1
14	G_4:4	G_4:4	G_5:5	G_3:3	G_2:2	G_1:1
15	G_4:4	G_4:4	G_5:5	G_3:3	G_2:2	G_1:1
16	G_5:5	G_5:5	G_4:4	G_3:3	G_2:2	G_1:1

ნახ. 9.8.

ამისათვის მაქალანობისის მეტრიკის დახმარებით განისაზღვრება მანძილი ყველა n ობიექტიდან თითოეული ჯგუფის სიმძიმის ცენტრამდე (საშუალოთა ვექტორი), რომლებიც განსაზღვრულებია შემსწავლელი შერჩევით. ექსპერტის მიერ i-ური ობიექტის j-ურ ჯგუფზე მიკუთვნება ითვლება მკაცრად, თუ მაქალანობისის მანძილი ობიექტიდან მისი ჯგუფის ცენტრამდე მნიშვნელოვნად მეტია, ვიდრე მისგან სხვა ჯგუფების ცენტრამდე, ხოლო თავის ჯგუფში მოხვედრის აპოსტერიორული ალბათობა ნაკლებია კრიტიკულ მნიშვნელობაზე. ამ შემთხვევაში ობიექტი ითვლება არაკორექტულად მიკუთვნებულად და გამორიცხული უნდა იქნეს შერჩევიდან.

შემსწავლელი შერჩევიდან ობიექტის გამორიცხვის პროცედურა იმაში მდგომარეობს, რომ საწყის მონაცემთა ცხრილში ობიექტს, რომელიც უნდა გამოირიცხოს შერჩევიდან (ის აღნიშნულია „ *

“-ით), ცილდება ამ ჯგუფზე მიკუთვნების ნომერი, რის შემდეგაც ტესტირების პროცესი მეორდება. სავარაუდოდ, ჯერ ირიცხება ის ობიექტი, რომელიც ყველაზე მეტად არ ეკუთვნის განსაზღვრულ ჯგუფს, ანუ რომელსაც ყველაზე დიდი მაქალანობისის მანძილი და უმცირესი აპოსტერიორული ალბათობა აქვს.

ჯგუფიდან მორიგი ობიექტის გამორიცხვისას უნდა გვახსოვდეს, რომ ამ დროს იძვრის ჯგუფის სიმძიმის ცენტრი (საშუალოთა ვექტორი), ვინაიდან ის გამოითვლება დარჩენილი დაკვირვებების მიხედვით. შემსწავლელი შერჩევის სიიდან მორიგი წარმოების წაშლის შემდეგ გამორიცხული არაა, რომ გაჩნდნენ არაკორექტულად მიკუთვნებული წარმოებები, რომლებიც წაშლამდე გათვალისწინებულები იყვნენ, როგორც სწორად მიკუთვნებულები. ამიტომ მოცემული პროცედურა უნდა ჩავატაროთ თითოეულ ბიჯზე მხოლოდ ერთი ობიექტის გამორიცხვით და მისი დაბრუნებით ისევ შემსწავლელ შერჩევებში. თუ ამ ობიექტის გამორიცხვისას მოხდა მეტად ძლიერი ცვლილებები, მაშინ უმრავლესი წარმოებები, რომლებიც მიკუთვნებულები იყვნენ კორექტულებს, მოინიშნებიან როგორც არაკორექტულად მიკუთვნებული წარმოებები.

დაკვირვებების გამორიცხვის პროცედურა გრძელდება მანამ, სანამ კლასიფიკაციის მატრიცაში კორექტურობის საერთო კოეფიციენტი მიაღწევს 100%-ს, ანუ შემსწავლელი შერჩევების ყველა დაკვირვება სწორად იქნება მიკუთვნებული შესაბამის ჯგუფებს.

შემსწავლელი შერჩევების მიღებული შედეგები, წარმოდგენილია ფანჯარაში Discriminant Function Analysis Results (დისკრიმინანტული ფუნქციების ანალიზი). ჩატარებული ანალიზის შედეგად შემსწავლელი შერჩევის კორექტულობის საერთო კოეფიციენტი უნდა იყოს 100%-ის ტოლი.

ობიექტების კლასიფიკაცია. მიღებული შემსწავლელი შერჩევების საფუძველზე შეიძლება იმ ობიექტების ხელმეორე კლასიფიკაციის ჩატარება, რომლებიც ვერ მოხვდნენ შემსწავლელ შერჩევებში,

და ნებისმიერი სხვა ობიექტის, რომლებიც დაჯგუფებას ექვემდებარებიან. მოცემული ამოცანის ამოსახსნელად არსებობს ორი ვარიანტი: პირველი - კლასიფიკაციის ჩატარება დიალოგის ფანჯრის Discriminant Function Analysis Results სტატისტიკური კრიტერიუმების საფუძველზე, მეორე - კლასიფიკაციის ფუნქციების საფუძველზე.

პირველ შემთხვევაში საჭიროა Discriminant Function Analysis Results დიალოგის ფანჯარაში დიალოგის დაუხურავად საწყისი გაკორექტირებული მონაცემების ცხრილში დავამატოთ ახალი შემთხვევები. იმისათვის, რომ გავიგოთ თუ რომელ კლასს მიეკუთვნება ეს ობიექტი, დააწექით ღილაკს Posterior probabilities (აპოსტერიორული ალბათობები). ამის შემდეგ თქვენ იხილავთ ცხრილს აპოსტერიორული ალბათობებით. იმ ჯგუფებს (კლასებს), რომლებსაც მაქსიმალური ალბათობები ექნებათ, შეიძლება მივაკუთვნოთ ახალი შემთხვევები.

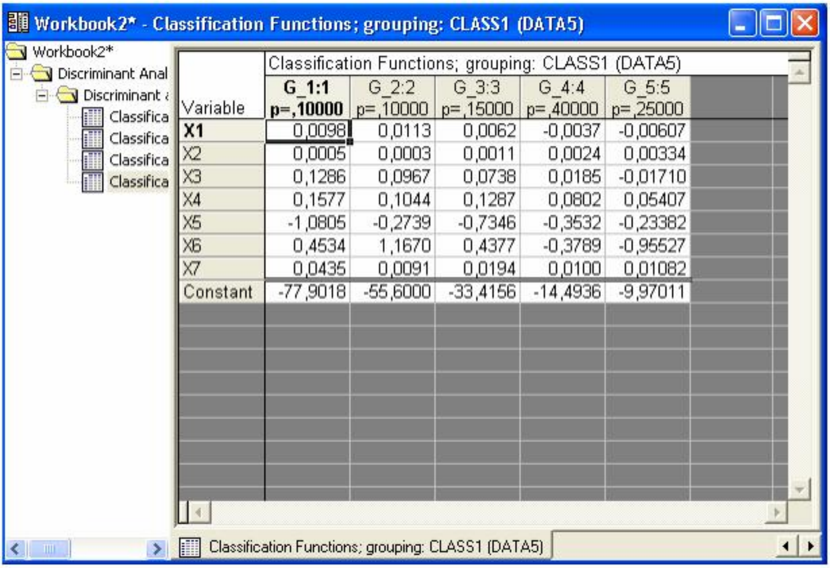
მეორე ვარიანტში საჭიროა დიალოგის ფანჯრის Discriminant Function Analysis Results ფანჯარაში დააწკაპუნოთ ღილაკზე Classification functions. გამონათდება ფანჯარა, რომლიდანაც შეიძლება ამოვიწეროთ თითოეული კლასისათვის კლასიფიკაციური ფუნქციები (ნახ. 9.9). მაგალითად, პირველი ორი კლასისათვის ფუნქციას ექნება სახე:

$$Y_1 = -77{,}9018 + 0{,}0098X1 + 0{,}0005X2 + 0{,}1286X3 + 0{,}1577X4 - 0{,}1577X5 + 0{,}4534X6 + 0{,}0435X7$$

$$Y_2 = -55{,}6000 + 0{,}0113X1 + 0{,}0003X2 + 0{,}0967X3 + 0{,}1044X4 - 0{,}2739X5 + 1{,}1670X6 + 0{,}0091X7$$

ამ ფუნქციების დახმარებით მომავალში შეიძლება ახალი ობიექტების კლასიფიცირება. ახალი ობიექტების მიკუთვნება მოხდება იმ კლასზე, რისთვისაც კლასიფიკაციური მნიშვნელობა მაქსიმალური იქნება. საბოლოო კლასიფიკაციის მეთოდის არჩევა დამოკიდებულია ახალი ობიექტების რაოდენობაზე. თუ ისინი ბევრი არაა, შეიძლება გამოვიყენოთ სტატისტიკურ კრიტერიუმებზე დაფუძნებული მეთოდი. თუ ახალი ობიექტების რაოდენობა დიდია, მაშინ

უფრო რაციონალურია შემსწავლელი შერჩევების მიხედვით მივიღოთ კლასიფიკაციური ფუნქციები, გავმართოთ ფორმულები და ჩავატაროთ საბოლოო კლასიფიკაცია.



Variable	G_1:1 p=,10000	G_2:2 p=,10000	G_3:3 p=,15000	G_4:4 p=,40000	G_5:5 p=,25000
X1	0,0098	0,0113	0,0062	-0,0037	-0,00607
X2	0,0005	0,0003	0,0011	0,0024	0,00334
X3	0,1286	0,0967	0,0738	0,0185	-0,01710
X4	0,1577	0,1044	0,1287	0,0802	0,05407
X5	-1,0805	-0,2739	-0,7346	-0,3532	-0,23382
X6	0,4534	1,1670	0,4377	-0,3789	-0,95527
X7	0,0435	0,0091	0,0194	0,0100	0,01082
Constant	-77,9018	-55,6000	-33,4156	-14,4936	-9,97011

ნახ. 9.9.

10. ფაქტორული ანალიზი

10.1. ზოგადი ცნობები

ფაქტორული ანალიზის ამოცანას წარმოადგენს მაჩვენებლების, ნიშან-თვისებების, რომლებიც ახასიათებენ ეკონომიკურ პროცესს ან ობიექტს, დიდი რაოდენობის გაერთიანება მათ საფუძველზე ხელოვნურად აგებული ფაქტორების უფრო მცირე რაოდენობაში, რომ შედეგად მიღებული ფაქტორთა სისტემა (ისევე კარგად აღმწერი შერჩეული მონაცემების, როგორც საწყისი) იყოს საკმაოდ მოხერხებული შინაარსობრივი ინტერპრეტაციის თვალსაზრისით. ფაქტორებს თქვენ არ აკვირდებით, ისინი ჰიპოთეზურად არსებობენ.

10.2. საკონტროლო მაგალითის შედეგების შემოწმება FACTOR ANALYSIS-ში

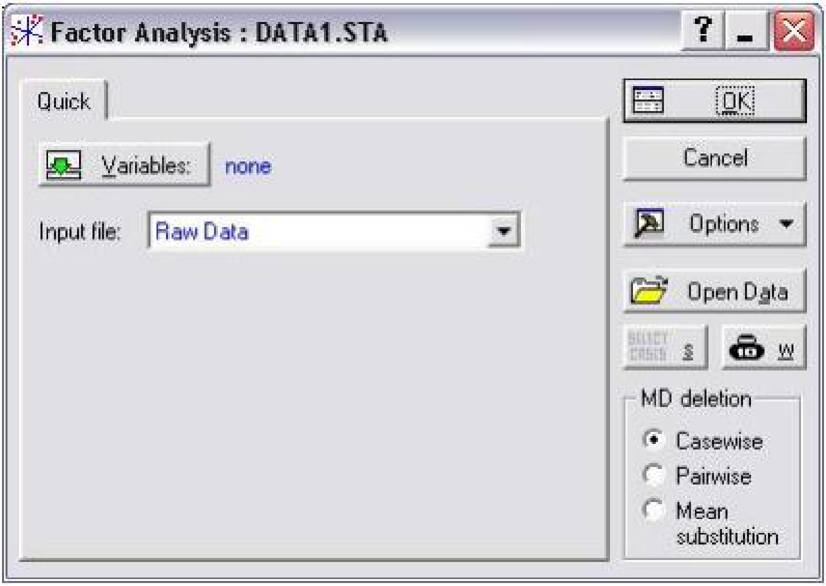
გვაქვს $Y_1,...,Y_3,X_4,\dots,X_{17}$ ცვლადთა სისტემა. საწყის მონაცემებად გამოვიყენოთ მონაცემები data.sta ფაილიდან.

თითოეული დაკვირვებისათვის ცვლადების ან ნიშან-თვისებების მნიშვნელობები ცნობილია. არსებული ინფორმაცია წარმოდგენილია მატრიცის სახით, რომლის თითოეული სტრიქონი შედგება ერთი მაჩვენებლის მნიშვნელობისაგან კვლევის თითოეული ობიექტისათვის. იგულისხმება, რომ ამ მატრიცის თითოეული ელემენტი წარმოადგენს ჰიპოთეზური საერთო ფაქტორების რაღაც რაოდენობისა და ერთი დამახასიათებელი ფაქტორის ზემოქმედების შედეგს, ანუ წარმოადგენს არადაკვირვებადი, ჰიპოთეზური, უშუალოდ არა გაზომვადი ფაქტორების წრფივ კომბინაციას. გამომდინარე, ფაქტორული ანალიზის ნებისმიერ მეთოდს აქვს ერთი მთავარი ამოცანა: შედეგობრივი ფაქტორის წარმოდგენა საერთო

ფაქტორთა რაღაც რაოდენობისა და ერთი დამახასიათებელი ფაქ-ტორის წრფივი კომბინაციის სახით.

განვიხილოთ სისტემა STATISTICA-ში ფაქტორული ანალი-ზის მეთოდით პრაქტიკული ამოცანის ამოხსნის პროცედურა.

- გახსენით ფაილი, რომელიც შეიცავს მონაცემებს ანალიზის-თვის. გამოვიყენოთ ფაილი data.sta (იხ. §5).
- Statistics/Multivariate Exploratory Technique/Factor Analysis მოდულის ბრძანებების დახმარებით გახსენით მოდული Factor Analysis (ფაქტორული ანალიზი). ეკრანზე გამონათდება Factor Analysis (ფაქტორული ანალიზი) მოდულის სასტარტო პანელი (ნახ. 10.1).



ნახ. 10.1.

მოდულში შესაძლებელია საწყისი მონაცემების შემდეგი ტიპების არსებობა:

Correlation Matrix (კორელაციური მატრიცა);

Raw Data (საწყისი მონაცემები).

- ამოირჩიეთ, მაგალითად, Raw Data. ეს მონაცემთა ჩვეულებრივი ფაილია, სადაც სტრიქონებად ჩაწერილია ცვლადების მნიშვნელობები.
- ველში MD (Missing Data - გამოტოვებული მონაცემები) მიუთითეთ გამოტოვებული მნიშვნელობების დამუშავების ხერხი.

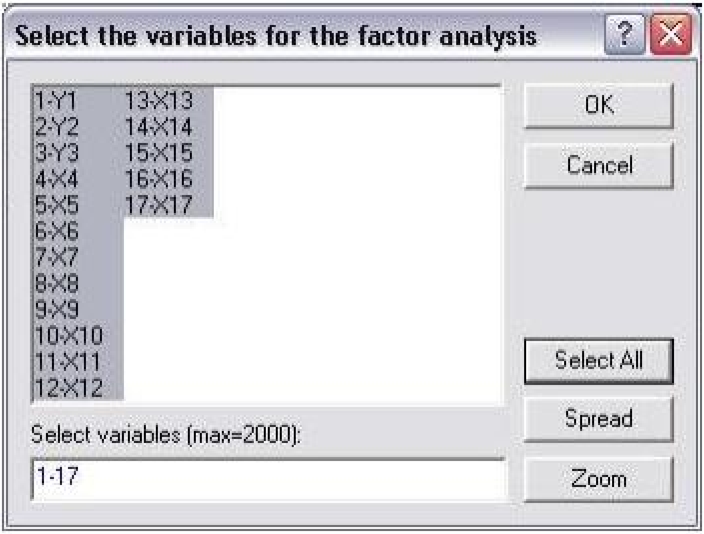
Casewise – გამოტოვებულ შემთხვევათა გამორიცხვის მეთოდი (იმაში მდგომარეობს, რომ მონაცემების შემცველ ელექტრონულ ცხრილში იგნორირდებიან ყველა ის სტრიქონები (შემთხვევები), რომლებშიც არის თუნდაც ერთი გამოტოვებული შემთხვევა. ეს ეხება ყველა ცვლადს. ცხრილში რჩებიან მხოლოდ ის შემთხვევები, რომლებშიც არ არის არცერთი გამოტოვება).

Pairwise - გამოტოვებული მნიშვნელობების გამორიცხვის წყვილური მეთოდი (გამოტოვებული შემთხვევები იგნორირდებიან არა ყველა ცვლადისათვის, არამედ მხოლოდ არჩეული წყვილისთვის. შემთხვევები, რომლებშიც არ არიან გამოტოვებულები, გამოიყენებიან დამუშავებებში, მაგალითად კორელაციური მატრიცის ელემენტების მიხედვით გამოთვლებში, როცა მიმდევრობით განიხილებიან ცვლადთა ყველა წყვილები). ცხადია, Pairwise ხერხში დამუშავებისათვის რჩება მეტი დაკვირვება, ვიდრე Casewise ხერხში. ოღონდ თავისებურება იმაში მდგომარეობს, რომ Pairwise-ში კორელაციის სხვადასხვა კოეფიციენტთა შეფასებები იგებიან დაკვირვებების სხვადასხვა რიცხვით.

Mean Substitution - საშუალოს ჩასმა გამოტოვებული მნიშვნელობის ნაცვლად. აირჩიეთ Casewise-გამოტოვებულ შემთხვევათა გამორიცხვის ხერხი.

- დააწექით ღილაკს Variables.
- ფანჯარაში (ნახ. 10.2) აირჩიეთ ცვლადები, ან წარმოდგენილ სიაში მაუსის მათი მონიშვნისას მარცხენა ღილაკის ჩაძირული მდგომარეობით ზემოთ ან ქვემოთ მოძრაობით, ან მა-

გალითად ქვედა სტრიქონში ცვლადების ნომრის აკრეფით. (ცვლადების მაქსიმალური რაოდენობაა 2000).



ნახ. 10.2.

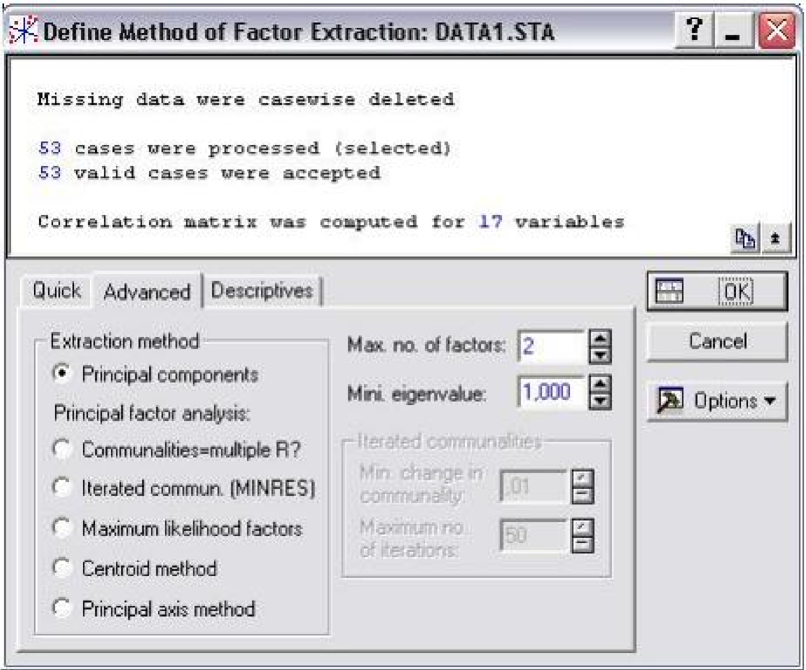
ღილაკი Sellect All (ყველას არჩევა) საშუალებას იძლევა ავირჩიოთ ყველა ცვლადი ერთად. ეს ყველაზე სწრაფი ხერხია.

- დააწკაპუნეთ ღილაკზე Spread (ფართოდ გაღება), დაათვალიერეთ ცვლადების გაფართოებული აღწერა. დააწკაპუნეთ ღილაკზე Sellect All (ყველას არჩევა) და აირჩიეთ ყველა ცვლადი.
- მოდულის სასტარტო ფანჯარაში OK ღილაკზე დაწკაპუნებით თქვენ დაიწყებთ არჩეული ცვლადების ანალიზს.

STATISTICA ამუშავებს გამოტოვებულ მნიშვნელობებს იმ ხერხით, რომელიც თქვენ მიუთითეთ, გამოითვლის კორელაციურ მატრიცას და შემოგთავაზებთ ასარჩევად ფაქტორული ანალიზის რამდენიმე მეთოდს.

კორელაციური მატრიცის გამოთვლა (თუ ის თავიდანვე არაა მოცემული) - ფაქტორული ანალიზის პირველი ეტაპია!

- დააწკაპუნეთ ღილაკზე OK, გამონათდება ფანჯარა Define Method of Factor Extraction (ფაქტორების გამოყოფის მეთოდის განსაზღვრა). მოცემულ ფანჯარას შემდეგი სტრუქტურა აქვს (ნახ. 10.3):



ნახ. 10.3.

ფანჯრის პირველი (ზედა) ნაწილი ინფორმაციულია: აქ გვატყობინებენ, რომ გამოტოვებული მნიშვნელობები დამუშავებულებია Casewise მეთოდით. დამუშავებულია 53 შემთხვევა და 53 შემთხვევა მიღებულია შემდგომი გამოთვლებისათვის. კორელაციური მატრიცა გამოთვლილია პირველი 17 შემთხვევისათვის.

დიალოგის ფანჯრის მეორე (ქვედა) ნაწილი - ფანჯარა Define Method of Factor Extraction (განისაზღვროს ფაქტორების გამოყოფის მეთოდი) - შეიცავს მეთოდის არჩევის ოფციებს, აგრეთვე ველებს,

რომლებშიც ხდება მომართვები ერთიანობათა იტერაციული გამოთვლებისათვის.

მიაქციეთ ყურადღება ველებს ფანჯრის მარჯვენა ნაწილში:

Max. no. of factors - ფაქტორთა მაქსიმალური რიცხვი;

Mini. eigen value – მინიმალური საკუთრივი მნიშვნელობა.

Extraction method (გამოყოფის მეთოდი) სათაურის ქვეშ გაერთიანებულ ოფციათა ჯგუფის საშუალებით ხდება დამუშავების მეთოდის არჩევა. ოპტიმალურობის კრიტერიუმიდან დამოკიდებულებით შესაძლებელია ანალიზის ჩატარება ან მეთოდით Principal components (მთავარი კომპონენტების მეთოდი), ანდა ერთ-ერთი იმ მეთოდით, რომლებიც გაერთიანებულებია ჯგუფში Principal factor analysis (მთავარი კომპონენტების ანალიზი).

სისტემა გვთავაზობს შემდეგ მეთოდებს ჯგუფში Principal factor analysis (მთავარი კომპონენტების ანალიზი):

Communalities = multiple R? (ერთობლიობები ტოლია მრავლობითი კორელაციის კოეფიციენტების კვადრატის);

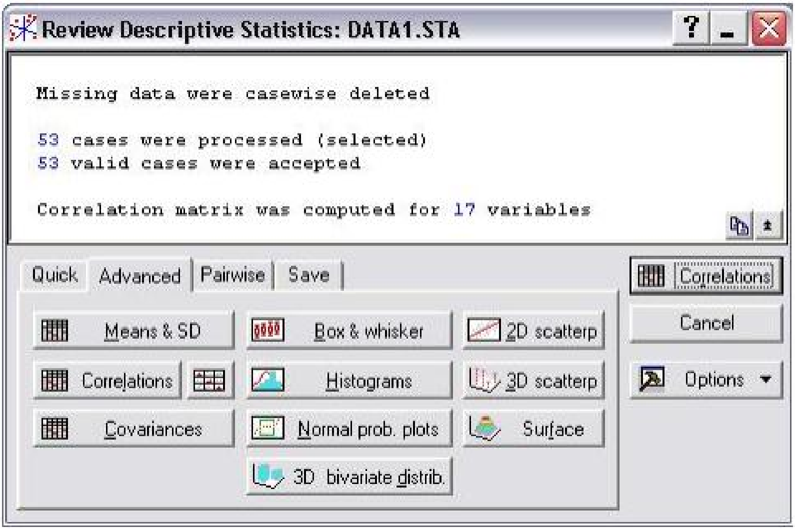
Iterated communalities (MINRES) იტერაციულ ერთობლიობათა მეთოდი (მინიმალური ნარჩენების);

Maximum likelihood factors - მაქსიმალური ალბათობის მეთოდი;

Centroid method - ცენტროიდული მეთოდი;

Principal axis method - მთავარ ღერძთა მეთოდი.

- გამოყავით ჩანართი Descriptives. Review correlation, means, standard deviations (კორელაციის, საშუალოს, სტანდარტული გადახრის დათვალიერება) ინიცირებით თქვენ გააღებთ Review Descriptive Statistics (აღწერითი სტატისტიკების დათვალიერება) ფანჯარას (ნახ. 10.4), სადაც შეგიძლიათ ნახოთ საშუალოები, სტანდარტული გადახრები, კორელაციები, კოვარიაციები, ააგოთ სხვადასხვა გრაფიკები.



ნახ. 10.4.

ღილაკის Compute multiple regression (მრავლობითი რეგრესიის გამოთვლა) დახმარებით შეიძლება მრავლობითი რეგრესიის გამოთვლა.

- დააწკაპუნეთ ღილაკზე Correlations (კორელაციები). თქვენ ეკრანზე დაინახავთ ადრე არჩეული ცვლადების კორელაციურ მატრიცას (ნახ. 10.5).
- File/Save correlations (კორელაციების დამახსოვრება) ღილაკის ინიცირებით დაიმახსოვრეთ კორელაციური მატრიცა. შემდეგ მოდულში **ფაქტორული ანალიზი** შემდგომი კვლევებისას პირდაპირ შეგიძლიათ იმუშაოთ კორელაციურ მატრიცასთან.
- Review Descriptive Statistics ფანჯარაში ღილაკზე Cancel დაწკაპუნებით დაბრუნდით Define Method of factor Extraction (განსაზღვრეთ ფაქტორების გამოყოფის მეთოდი) ფანჯარაში.

Correlations (DATA1.STA)
Casewise deletion of MD
N=53

Variable	Y1	Y2	Y3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17
Y1	1,00	0,55	0,13	0,10	0,28	0,03	-0,17	-0,06	-0,09	0,04	0,26	0,44	-0,10	0,20	-0,09	-0,25	0,02
Y2	0,55	1,00	0,04	-0,37	0,13	0,06	-0,12	0,13	-0,09	-0,08	0,24	0,83	-0,42	0,29	-0,30	-0,28	-0,06
Y3	0,13	0,04	1,00	0,08	-0,07	-0,19	0,22	0,41	0,04	0,08	0,07	0,02	0,06	-0,11	-0,20	-0,15	-0,33
X4	0,10	-0,37	0,08	1,00	-0,14	-0,05	-0,04	-0,07	0,11	-0,19	-0,07	-0,36	0,35	-0,01	0,22	-0,14	-0,01
X5	0,28	0,13	-0,07	-0,14	1,00	-0,10	-0,10	-0,05	-0,12	0,07	0,04	0,03	0,02	-0,09	0,01	0,11	-0,06
X6	0,03	0,06	-0,19	-0,05	-0,10	1,00	-0,20	-0,24	0,02	-0,05	0,23	-0,00	0,08	-0,17	0,28	-0,10	0,22
X7	-0,17	-0,12	0,22	-0,04	-0,10	-0,20	1,00	0,17	-0,23	-0,06	0,05	0,01	-0,06	0,07	0,16	-0,06	-0,16
X8	-0,06	0,13	0,41	-0,07	-0,05	-0,24	0,17	1,00	-0,10	-0,16	-0,01	0,14	-0,09	0,20	-0,13	0,08	-0,25
X9	-0,09	-0,09	0,04	0,11	-0,12	0,02	-0,23	-0,10	1,00	0,31	-0,24	-0,10	-0,07	-0,00	-0,02	0,07	-0,05
X10	0,04	-0,08	0,08	-0,19	0,07	-0,05	-0,06	-0,16	0,31	1,00	-0,05	-0,14	0,05	-0,27	-0,15	0,30	-0,08
X11	0,26	0,24	0,07	-0,07	0,04	0,23	0,05	-0,01	-0,24	-0,05	1,00	0,31	-0,05	0,14	-0,08	-0,28	-0,04
X12	0,44	0,83	0,02	-0,36	0,03	-0,00	0,01	0,14	-0,10	-0,14	0,31	1,00	-0,37	0,62	-0,29	-0,31	-0,05
X13	-0,10	-0,42	0,06	0,35	0,02	0,08	-0,06	-0,09	-0,07	0,05	-0,05	-0,37	1,00	-0,25	0,07	0,01	-0,09
X14	0,20	0,29	-0,11	-0,01	-0,09	-0,17	0,07	0,20	-0,00	-0,27	0,14	0,62	-0,25	1,00	-0,25	-0,31	-0,14
X15	-0,09	-0,30	-0,20	0,22	0,01	0,28	0,16	-0,13	-0,02	-0,15	-0,08	-0,29	0,07	-0,25	1,00	0,21	0,55
X16	-0,25	-0,28	-0,15	-0,14	0,11	-0,10	-0,06	0,08	0,07	0,30	-0,28	-0,31	0,01	-0,31	0,21	1,00	0,07
X17	0,02	-0,06	-0,33	-0,01	-0,06	0,22	-0,16	-0,25	-0,05	-0,08	-0,04	-0,05	-0,09	-0,14	0,55	0,07	1,00

ნახ. 10.5.

- აირჩიეთ ოპცია Principal components (მთავარი კომპონენტები) და დააწკაპუნეთ OK ღილაკზე.

სისტემა სწრაფად ჩაატარებს გამოთვლებს და ეკრანზე გამონათდება ფანჯარა Factor Analysis Results (ფაქტორული ანალიზის შედეგები) (ნახ. 10.6).

Factor Analysis Results (ფაქტორული ანალიზის შედეგები) ფანჯრის ზედა ნაწილში გამონათდება ინფორმაციული შეტყობინება:

Number of variables (გასაანალიზებელი ცვლადების რიცხვი) - 17;

Method (ანალიზის მეთოდი - მთავარი კომპონენტები);

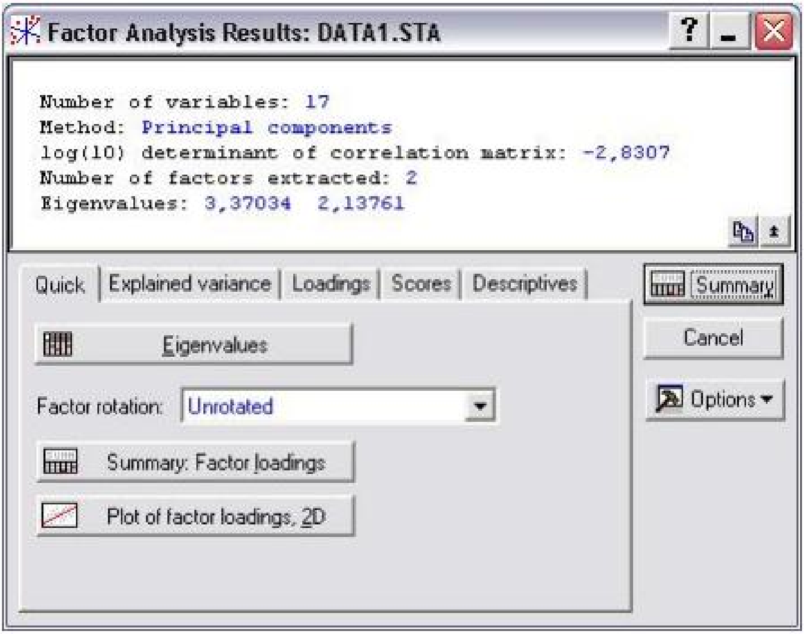
Log(10) determination of correlation matrix (კორელაციური მატრიცის დეტერმინანტის ათობითი ლოგარითმი) – 2,8307;

Number of Factor extraction (გამოყოფილი ფაქტორების რიცხვი) – 2;

Eigenvalues (საკუთრივი მნიშვნელობები) – 3,37034; 2,13761.

ფანჯრის ქვედა ნაწილში მოთავსებულია ფუნქციონალური ღილაკები, რომლებიც საშუალებას იძლევიან ყოველმხრივ დავათ-

ვალიეროთ ანალიზის შედეგები (როგორც რიცხობრივად, ისე გრაფიკულად).

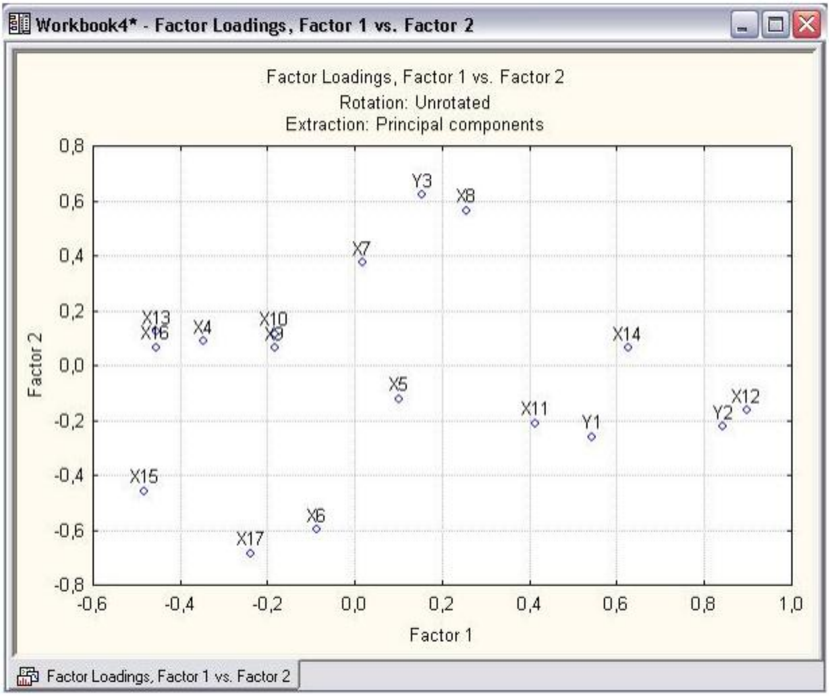


ნახ. 10.6.

- დააწკაპუნეთ ღილაკზე Plot of factor loadings, 2D (დატვირთვების ორობითი გრაფიკი) და ნახეთ ფაქტორული ანალიზის შედეგები გრაფიკზე (ნახ. 10.7).
- დაბრუნდით ფანჯარაში Factor Analysis Results და დააწკაპუნეთ ღილაკზე Summary: Factor Loadings (ფაქტორული დატვირთვები) და ნახეთ დატვირთვები რიცხობრივად (ნახ. 10.8).

ამონახსნის გრაფიკზე დათვალიერებისას, ყურადღება მიაქციეთ წითელი ფერით გამოყოფილი დატვირთვების პარამეტრებს. რთულია მათი ინტერპრეტირება, იმის კითხვა, რა აზრი მივანიჭოთ მეორე ფაქტორს. ამ შემთხვევაშია მიზანშეწონილია მივმართოთ

ღერძების მობრუნებას, იმ მიზნით, რომ მივიღებთ ამოხსნას, რომ-ლის ინტერპრეტირებაც შესაძლებელი იქნება საგნობრივ არეში.



ნახ. 10.7.

- პანელში Factor Analysis Results ველში Factor rolation (ფაქ-ტორების ბრუნვა) (ნახ. 10.9) თქვენ შეგიძლიათ აირჩიოთ ღერძების სხვადასხვა მობრუნებები. ფანჯარა გთავაზობთ რამდენიმე შესაძლებლობას შეაფასოთ და იპოვოთ საჭირო მობრუნება შემდეგი მეთოდებით:

  Varimax - ვარიმაქსი;

  Biquartimax - ბიკვარტიმაქსი;

  Equamax - ექვიმაქსი.

მობრუნების მიზანი მარტივი სტრუქტურის მიღებაა, რომლის დროსაც დაკვირვებების უმრავლესობა იმყოფებიან

კოორდინატთა ღერძების ახლოს. დაკვირვებების შემთხვევითი კონფიგურაციისას შეუძლებელია მივიღოთ მარტივი სტრუქტურა.

Workbook4* - Factor Loadings (Unrotate...

Factor Loadings (Unrotated) (DATA1.STA)
Extraction: Principal components
(Marked loadings are > ,700000)

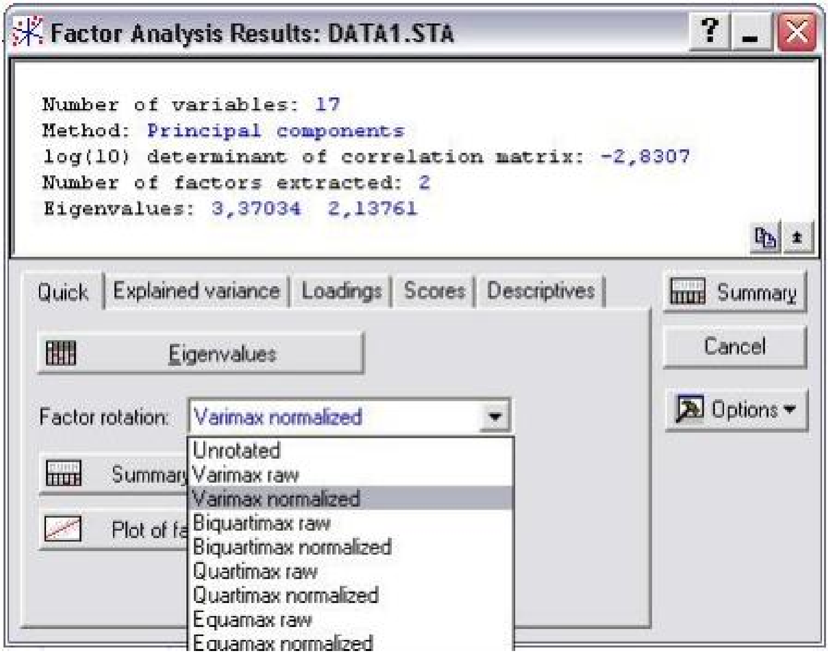
Variable	Factor 1	Factor 2
Y1	0,543757	-0,259269
Y2	0,843369	-0,224173
Y3	0,154152	0,621587
X4	-0,345795	0,087639
X5	0,100967	-0,122959
X6	-0,088632	-0,599487
X7	0,016999	0,374937
X8	0,256517	0,564900
X9	-0,181615	0,063141
X10	-0,183106	0,114819
X11	0,413336	-0,210444
X12	0,897339	-0,163027
X13	-0,453697	0,123133
X14	0,625475	0,062887
X15	-0,481775	-0,458884
X16	-0,455339	0,065198
X17	-0,238409	-0,685507
Expl.Var	3,370336	2,137613
Prp.Totl	0,198255	0,125742

Factor Loadings, Factor 1 vs. Factor 2 | Factor Loadings

ნახ. 10.8.

მეთოდის დასახელებაში დამატებითი ტერმინი - normalized (ნორმალიზაცია) - იმაზე მიუთითებს, რომ ფაქტორული დატვირთვები პროცედურაში ნორმალიზდებიან (ხდება მათი გაყოფა შესა-

ბამისი დისპერსიის კვადრატულ ფესვზე). ტერმინი raw (საწყისი) გვიჩვენებს, რომ საბრუნავი დატვირთვები არ არიან ნორმალიზე-ბულები.



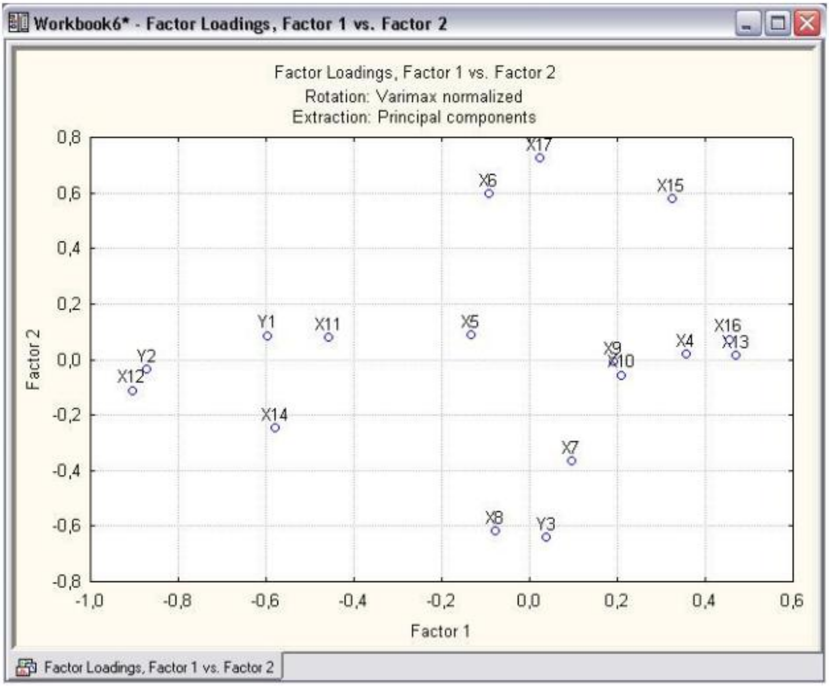
ნახ. 10.9.

მოახდინეთ Varimax normalized (ნორმალიზებული ვარიმაქ-სი) ინიცირება.

სისტემა მოახდენს ფაქტორების მობრუნებას ნორმალიზე-ბული ვარიმაქსის მეთოდით და ფანჯარა Factor Analysis Results (ფაქტორული ანალიზის შედეგები) ისევ გამონათდება ეკრანზე. ისევ მოახდინეთ ამ ფანჯარაში ღილაკის Plot of factor loadings, 2D (დატვირთვების ორგანზომილებიანი გრაფიკი) ინიცირება. თქვენ ისევ დაინახავთ დატვირთვების გრაფიკს (ნახ. 10.10).

რა თქმა უნდა, ეს გრაფიკი ცოტათი განსხვავდება წინასგან. ვნახოთ დატვირთვები რიცხობრივად Factor Analysis Results ფან-

ჯარაში Summary: Factor loadings (ფაქტორული დატვირთვები) ღილაკის ინიცირებით (ნახ. 10.11).



ნახ. 10.10.

ახლა უკვე შეიძლება ნაპოვნი გადაწყვეტილების ინტერპრეტირება. ფაქტორების ინტერპრეტირებას უფრო ხშირად ახდენენ დატვირთვების მიხედვით. პირველი ფაქტორი ყველაზე მჭიდროდაა დაკავშირებული Y2 - პროდუქციის თვითღირებულების შემცირების ინდექსთან და X12 - ძირითადი საწარმოო ფონდების საშუალო წლიურ ღირებულებასთან. ძირითადი საწარმოო ფონდების შენახვის საერთო ღირებულება შედგება პირველადი ინვესტიციური დანახარჯებისა და შემდგომი საექსპლოატაციო ხარჯებისაგან, რომლებიც დამოკიდებულებია წარმოების მოცულობაზე. მეორე

ფაქტორი - მხოლოდ ერთ მაჩვენებელთან - X17-თან (არასაწარმოო ხარჯებთან).

Workbook6* - Factor Loadings (Varimax normaliz...

Factor Loadings (Varimax normalized) (DATA1.STA)
Extraction: Principal components
(Marked loadings are > ,700000)

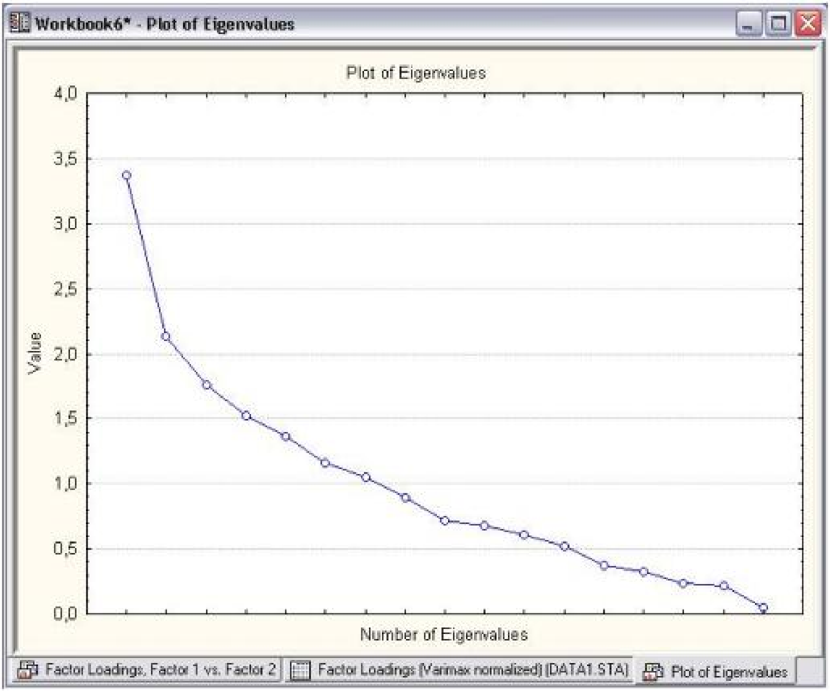
Variable	Factor 1	Factor 2
Y1	0,596182	-0,086366
Y2	0,871911	0,035995
Y3	-0,037101	0,639341
X4	-0,356230	-0,018842
X5	0,132888	-0,087489
X6	0,093122	-0,598806
X7	-0,094946	0,363114
X8	0,077469	0,615558
X9	-0,192170	0,006446
X10	-0,208918	0,055358
X11	0,457149	-0,078411
X12	0,905322	0,110395
X13	-0,469804	-0,016941
X14	0,578695	0,245532
X15	-0,324032	-0,581106
X16	-0,454193	-0,072757
X17	-0,024412	-0,725371
Expl.Var	3,261941	2,246008
Prp.Totl	0,191879	0,132118

Factor Loadings, Factor 1 vs. Factor 2 | Factor Loadings (Varimax no

ნახ. 10.11.

ჩნდება კითხვა: რამდენი ფაქტორითაა საჭირო შემოვიფარგლოთ პრაქტიკაში? ამისათვის პროგრამულ პაკეტ STATISTICA-ში ფანჯარა Factor Analysis Results ჩანართში Explained variance (ნახ.

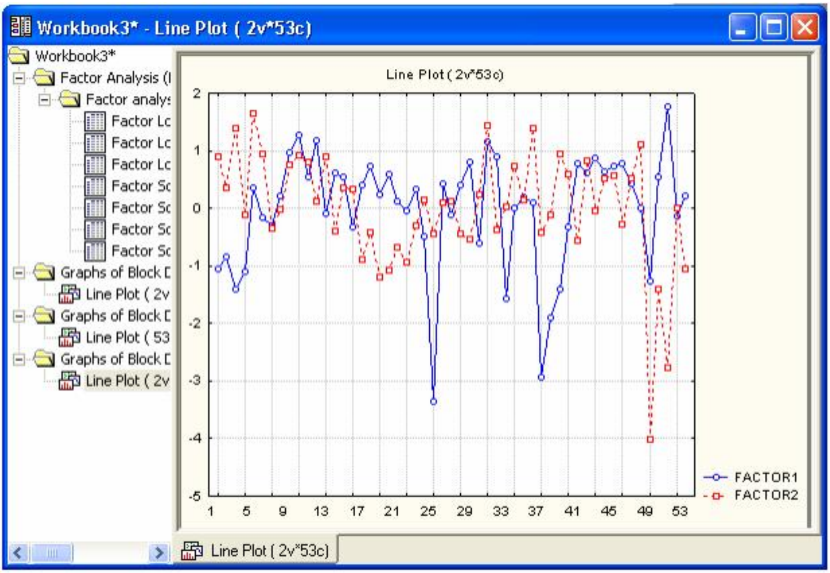
10.12) არსებობს კრიტერიუმი Scree plot (ქვიანი ჩამონაცვენის კრი-ტერიუმი).



ნახ. 10.12.

წერტილებში კოორდინატებით 1 , 2 ჩამოცვენა ნელდება ყველაზე მნიშვნელოვნად, გამომდინარე, შეიძლება შემოვისაზღვროთ ერთი ან ორი ფაქტორით.

სისტემა STATISTICA საშუალებას გვაძლევს განვსაზღვროთ ამ ფაქტორთა ნორმირებული მნიშვნელობები (ღილაკი Factor scores) და ავაგოთ გრაფიკები, ამ სვეტების წინასწარი გამოყოფით, თაგვის მარჯვენა ღილაკით გამოვიძახოთ კონტექსტური მენიუ და მივუთითოთ ბრძანება Graphs of BlockData/Line Plot: Entire Columns. ამ გრაფიკების (ნახ. 10.13) ინტერპრეტირება რჩება მოცემულ საგნობრივ არეში სპეციალისტის საქმედ.


Workbook3* - Line Plot (2v*53c)
Workbook3*
Factor Analysis (I
Factor analys
Factor Lc
Factor Lc
Factor Lc
Factor Sc
Factor Sc
Factor Sc
Factor Sc
Graphs of Block D
Line Plot (2v
Graphs of Block D
Line Plot (53
Graphs of Block D
Line Plot (2v
Line Plot (2v*53c)
2
1
0
-1
-2
-3
-4
-5
1
5
9
13
17
21
25
29
33
37
41
45
49
53
FACTOR1
FACTOR2
Line Plot (2v*53c)

ნახ. 10.13.

11. „გადაწყვეტილებათა ხე“
11.1. ზოგადი ცნობები

„გადაწყვეტილებათა ხე“ (კლასიფიკაცია) - ეს მეთოდია, რომელიც საშუალებას იძლევა ვიწინასწარმეტყველოთ დაკვირვებებისა და ობიექტების მიკუთვნება ცვლადის კატეგორიალური დამოკიდებულების ამა თუ იმ კლასს ერთი ან რამდენიმე პრედიქტორული ცვლადის მნიშვნელობათა შესაბამისად. „გადაწყვეტილებათა ხის“ იერარქიული აგება მისი ერთ-ერთი უმნიშვნელოვანესი თვისებაა. „ხის ტანი“ წარმოადგენს პრობლემას ან სიტუაციას, რომელიც გადაწყვეტას მოითხოვს. „ხის წვეროს“ წარმოადგენენ მიზნები ან ფასეულობები, რომლითაც ხელმძღვანელობს გადაწყვეტილებების მიმღები ადამიანი.

გადაწყვეტილებათა ხეების შექმნის პირველი იდეები დასაბამს იღებენ მე-20 საუკუნის 50-იანი წლების ჰოვლენდისა (Hoveland) და ჰანტის (Hunt) შრომებიდან. ოღონდ, ფუძემდებლური სამუშაო, რომელმაც იმპულსი მისცა ამ მიმართულების განვითარებას, გახდა ჰანტის (Hunt E.B.), მერინის (Marin J.) და სტოუნის (Stone P.J.) წიგნი „Experiments in Induction”, რომელიც 1966 წელს გამოვიდა.

11.2. ტერმინოლოგია

შემოვიტანოთ ძირითადი ცნებები თეორიიდან „გადაწყვეტილებათა ხე“(ცხრ. 11.1).

ცხრილი11.1.

დასახელება	აღწერა
ობიექტი	მაგალითი, შაბლონი, დაკვირვება

ატრიბუტი	ნიშან-თვისება, დამოუკიდებელი ცვლადი, თვისება
კლასის ნიშნული	დამოკიდებული ცვლადი, მიზნობრივი ცვლადი, ნიშან-თვისება, რომელიც განსაზღვრავს ობიექტის კლასს
კვანძი	ხის შიგა კვანძი, შემოწმების კვანძი
ფურცელი	ხის ბოლო კვანძი, გადაწყვეტილების კვანძი
შემოწმება	პირობა კვანძში

11.3. რა არის „გადაწყვეტილებათა ხე“ და გადასაწყვეტ ამოცანათა ტიპები?

„გადაწყვეტილებათა ხე“ – ეს წესების წარმოდგენის ხერხია იერარქიულ, მიმდევრობით სტრუქტურაში, სადაც თითოეულ ობიექტს შეესაბამება გადაწყვეტილების მომცემი ერთადერთი კვანძი.

წესის ქვეშ გულისხმობენ ლოგიკურ კონსტრუქციას, წარმოდგენილს „თუ ... მაშინ ...“ სახით.

დღეისათვის „გადაწყვეტილებათა ხის“ გამოყენების სფერო ფართოა, მაგრამ ამ აპარატით გადაწყვეტილი ყველა ამოცანა შეიძლება გაერთიანდეს შემდეგ სამ კლასში:

- **მონაცემების აღწერა:** „გადაწყვეტილებათა ხეები“ საშუალებას იძლევიან შენახულ იქნას ინფორმაცია კომპაქტურ ფორმაში, მათ ნაცვლად ჩვენ შეგვიძლია შევინახოთ „გადაწყვეტილებათა ხე“, რომელიც შეიცავს ობიექტების ზუსტ აღწერას;

- **კლასიფიკაცია:** „გადაწყვეტილებათა ხეები“ წარმატებით წყვეტენ კლასიფიკაციის ამოცანებს, ანუ ობიექტთა მიკუთვნებას წინასწარ ცნობილი კლასებიდან ერთ-ერთზე;

- **რეგრესია:** თუ მიზნის ცვლადს გააჩნია უწყვეტი მნიშვნელობები, მაშინ „გადაწყვეტილებათა ხეები“ საშუალებას იძლევიან დადგინდეს მიზნის ფუნქციის დამოკიდებულება დამოუკიდებელი (შესასვლელი) ცვლადებისგან. მაგალითად, ამ კლასს აკუთვნებენ რიცხვითი პროგნოზირების (მიზნის ფუნქციის მნიშვნელობათა წინასწარმეტყველების) ამოცანებს.

11.4. როგორ ავაგოთ „გადაწყვეტილებათა ხე“ ?

დღეისათვის არსებობს ალგორითმების მნიშვნელოვანი რაოდენობა, რომლებშიც რეალიზებულია „გადაწყვეტილებათა ხეები“ – CART, C4.5, NewId, ITrule, CHAID, CN2 და სხვა. მაგრამ ყველაზე ფართო გავრცელება და პოპულარობა შემდეგმა ორმა მიიღო:

- CART (Classification and Regression Tree) - ეს ბინარური „გადაწყვეტილებათა ხის“ დიქოტომიური კლასიფიკაციური მოდელის აგების ალგორითმია. ხის ყოველი კვანძს დაყოფისას გააჩნია მხოლოდ ორი შთამომავალი. როგორც ალგორითმის დასახელებიდან ჩანს, ის წყვეტს კლასიფიკაციისა და რეგრესიის ამოცანებს.

- C4.5 - ეს „გადაწყვეტილებათა ხის“ აგების ალგორითმია, რომელშიც კვანძის შთამომავლების რაოდენობა შეზღუდული არაა. არ შეუძლია იმუშაოს უწყვეტ მიზნობრივ ველთან, ამიტომ წყვეტს მხოლოდ კლასიფიკაციის ამოცანებს.

პაკეტ STATISTICA-ს ჩანართ Statistics-ის მოდულში „კლასიფიკაციის ხეები“ რეალიზებულია სამი მეთოდი:

- დისკრიმინანტული ერთგანზომილებიანი განშტოება კატეგორიალური და ხარისხობრივი პრედიქტორების მიხედვით;
- დისკრიმინანტული მრავლობითი განშტოება ხარისხობრივი პრედიქტორების წრფივი კომბინაციების მიხედვით;

- ერთგანზომილებიანი განშტოება CART მეთოდით.

პირველი ორი წარმოადგენს პაკეტ QUEST-ის შესაბამისი ალ-გორითმების (Quick, Unbiased, Efficient Statistical Trees) ადაპტაციას. QUEST - ეს „კლასიფიკაციის ხეების" პროგრამაა, დამუშავებული Loh და Shih-ის მიერ (1997), რომელშიც გამოიყენებიან რეკურსიული კვადრატული დისკრიმინანტული ანალიზის მეთოდის გაუმჯობე-სებული ვარიანტები და რომელიც შეიცავს ახალ საშუალებათა რიგს იმ „კლასიფიკაციის ხეების" საიმედობისა და ეფექტურობის ასამაღ-ლებლად, რომლებსაც ის აგებს. QUEST პაკეტის ალგორითმები საკ-მაოდ რთულებია.

მოდულში **კლასიფიკაციის ხეები** არის მეორე, კონცეპტუა-ლურად უფრო მარტივი, მიდგომა. აქ ერთგანზომილებიანი გან-შტოების CART მეთოდით რეალიზებული ალგორითმი წარმოად-გენს პაკეტ CART-ის ალგორითმების ადაპტაციას. CART (Classifi-cation and Regression Tree) - ეს პროგრამაა, რომელიც „ხის" აგებისას ახორციელებს *ერთგანზომილებიანი განშტოების* ყველა შესაძლო ვარიანტის სრულ გადარჩევას.

CART მეთოდი გამოიყენება კატეგორიზებული (ჩვეულებ-რივ ორდონიანი) და ხარისხობრივი პრედიქტორული ცვლადებისა-თვის. კატეგორიზებული პრედიქტორული ცვლადისათვის, რომე-ლიც მოცემულ კვანძში ღებულობს k მნიშვნელობას, არის მისი მნიშვნელობების სიმრავლის ორ ნაწილად დაყოფის ზუსტად $2^{(k-1)}$-1 ვარიანტი. ხარისხობრივი პრედიქტორისათვის, რომელსაც მოცემულ კვანძში აქვს k სხვადასხვა დონე, არის k-1 წერტილი, რომლებიც სხვადასხვა დონეებს ყოფენ. ვხედავთ, რომ განშტოების სხვადასხვა ვარიანტთა რაოდენობა, რომლებიც საჭიროა დავათვა-ლიეროთ, იქნება ძალიან დიდი: თუ ამოცანაში ბევრი პრედიქტო-რია, მაშინ მათ მნიშვნელობათა ბევრი დონე აქვთ, ე.ი. „ხეში" ბევრი ტერმინალური წვეროა.

ერთი შეხედვით *დისკრიმინანტული ანალიზისა* და „კლასიფიკაციის ხეების“ გადაწყვეტილებათა მიღების პროცედურები მსგავსებად გამოიყურებიან, ვინაიდან ორივეში ორივეში მონაწილეობენ „გადამწყვეტი“ განტოლებები და კოეფიციენტები. ოღონდ არსებობს პრინციპული განსხვავება *დისკრიმინანტულ ანალიზში* ერთდროულად და „კლასიფიკაციის ხეებში“ მიმდევრობით (იერარქიულად) მიღებულ გადაწყვეტილებებს შორის.

განსხვავება ამ ორ მიდგომას შორის უფრო ნათელი გახდება, თუ ვნახავთ როგორ სრულდება *რეგრესია* ამა თუ იმ შემთხვევაში. პროგნოზირება *დისკრიმინანტული ანალიზის* დახმარებით ხორციელდება *ერთდროული* მრავლობითი რეგრესიით ყველა პრედიქტორულ ცვლადზე. პროგნოზირება „კლასიფიკაციის ხეების“ მეთოდით შედგება მარტივი (ბიჯური) რეგრესიული ანალიზის ცალკეული ეტაპებისაგან.

„კლასიფიკაციის ხეების“ მეთოდის განმასხვავებელი ნიშანი - ესაა მისთვის დამახასიათებელი მოქნილობა - „კლასიფიკაციის ხეების“ შესაძლებლობა მიმდევრობით შეისწავლოს ცალკეული ცვლადების ზემოქმედების ეფექტი. არის კიდევ მთელი რიგი მიზეზებისა, რომლებიც *კლასიფიკაციის ხეებს* უფრო მოქნილ საშუალებად ქმნიან, ვიდრე ანალიზის ტრადიციული მეთოდებია.

„კლასიფიკაციის ხეების“ გამოყენების ფართო სფერო ქმნის მას მონაცემების ანალიზის ფრიად მიმზიდველ ინსტრუმენტად, მაგრამ არ უნდა ჩავთვალოთ, რომ რეკომენდირებულია მისი გამოყენება სტატისტიკის ტრადიციული მეთოდების ნაცვლად. პირიქით, თუ შესრულებულია უფრო მკაცრი თეორიული წინაპირობები, რომლებსაც ტრადიციული მეთოდები ითხოვენ და შერჩეულ განაწილებას გააჩნია ზოგიერთი სპეციალური თვისება, მაშინ უფრო შედეგიან იქნება სწორედ ტრადიციული მეთოდების გამოყენება. მაგრამ, როგორც დაზვერვითი ანალიზის მეთოდს ან როგორც უკანასკნელ საშუალებას, ყველა ტრადიციული მეთოდის მტყუნების შემ-

თხვევაში, ბევრი მკვლევარის თვალსაზრისით „კლასიფიკაციის ხეებს“ ბადალი არა აქვთ.

დამატებითი ინფორმაცია მონაცემების ანალიზის, მოპოვების, ვიზუალიზაციისა და პროგნოზირების მეთოდების შესახებ მოთავსებულია StatSoft-ის პორტალზე (http:// www.statsoft.ru/home/ portal/ default.asp).

გადაწყვეტილებათა ხეების გამოყენების უპირატესობები

„გადაწყვეტილებათა ხის“ მეთოდის ძირითად უპირატესობებს წარმოადგენენ:

- სწავლების პროცესის სისწრაფე;
- წესების იმ სფეროებში გენერაცია, სადაც რთულია თავისი ცოდნის ფორმალიზება;
- წესების ამოღება ბუნებრივ ენაზე;
- ინტუიციურ დონეზე გასაგები კლასიფიკაციური მოდელი;
- პროგნაზის მაღალი სიზუსტე, რომელიც შეჯერებადია სხვა მეთოდებთან (სტატისტიკა, ნეირონული ქსელები);
- არაპარამეტრული მოდელების აგება.

ამ და სხვა მრავალი მიზეზის გამო „გადაწყვეტილებათა ხის“ მეთოდოლოგია წარმოადგენს მნიშვნელოვან ინსტრუმენტს მონაცემების ანალიზით დაკავებული თითოეული სპეციალისტის საქმიანობაში, იმისგან დამოუკიდებლად პრაქტიკოსია ის თუ თეორეტიკოსი.

გადაწყვეტილებათა ხეების გამოყენების სფეროები

„გადაწყვეტილებათა ხეები“ წარმოადგენენ საუკეთესო ინსტრუმენტს გადაწყვეტილებათა მიღების, მონაცემების ინტელექტუალური ანალიზის (Data Mining) მხარდაჭერის სისტემებში. მონაცე-

მების ინტელექტუალური ანალიზის ბევრი პაკეტის შემადგენლობაში უკვე ჩართულია „გადაწყვეტილებათა ხეების“ აგების მეთოდები. იმ სფეროებში, სადაც მაღალია შეცდომის ფასი, ისინი იქნებიან საუკეთესო დამხმარე საშუალებები ანალიტიკოსისა თუ ხელმძღვანელისათვის.

„გადაწყვეტილებათა ხეები“ წარმატებით გამოიყენებიან პრაქტიკული ამოცანების გადასაწყვეტად შემდეგ დარგებში:

- საბანკო საქმეში ბანკის კლიენტების კრედიტუნარიანობის შესაფასებლად კრედიტების გაცემის დროს;
- მრეწველობაში პროდუქციის ხარისხის კონტროლის განხორციელებაში (დეფექტების გამოვლენა), გამოცდებში დარღვევების გარეშე (მაგალითად, შედუღების ხარისხის შემოწმება) და ა.შ.;
- მედიცინაში სხვადასხვა ავადმყოფობების დიაგნოსტირებისას;
- მოლეკულარულ ბიოლოგიაში ამინომჟავების აგებულების ანალიზის ჩატარებისას.

ეს სრულაც არაა სრული სია იმ დარგებისა, სადაც შეიძლება გამოყენებულ იქნას „გადაწყვეტილებათა ხეები“.

11.5. საკონტროლო მაგალითის შედეგების შემოწმება CLASSIFICATION TREE-ში

დავუშვათ, რომ ვთქვათ გაქვთ მონაცემები 37 ციკლონის კოორდინატების - გრძედის (Longibute) და განედის (Lotitude) - შესახებ (ცხრ. 11.2), რომლებიც აღწევენ გრიგალის ძალას, ციკლონების ორი კლასიფიკაციით - Bare და Trop. ქვემოთ მოყვანილი მონაცემების მოდულური ნაკრები გამოყენებული იყო

ილუსტრაციის მიზნებისათვის Elsner, Lehmiller, Kimberlain შრომაში (1996), რომელთა ავტორებიც იკვლევდნენ განსხვავებას ბაროსოლურ და ტროპიკულ ციკლონებს შორის ჩრდილოეთ ატლანტიკაში.

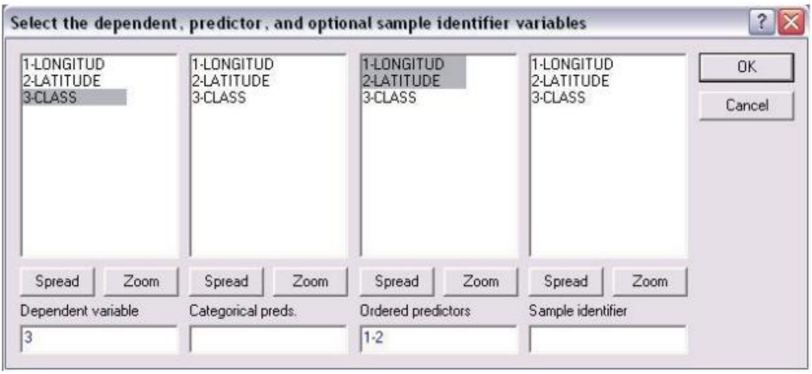
- შევიტანოთ ან გავხსნათ ფაილი საწყისი მონაცემებით, ეკრანზე გამონათდება ნახ. 11.1-ზე გამოსახული ფანჯარა.
- გავხსნათ მოდული Classification Tree ბრძანებით Statistics/ Multivariate Exploratory Technique/Classification Trees.
- ჩანართში Methods განაყოფში Split selection method აირჩიეთ მეთოდი CART (C&RT-style);

ცხრილი 11.2.

N/N	LONGITUD	LATITUDE	CLASS	№/№	LONGITUD	LATITUDE	CLASS
1	59.00	17.00	BARO	21	66.00	14.00	TROP
2	59.50	21.00	BARO	22	66.00	17.00	TROP
3	60.00	12.00	BARO	23	66.50	17.00	TROP
4	60.50	16.00	BARO	24	66.50	18.00	TROP
5	61.00	13.00	BARO	25	66.50	21.00	TROP
6	61.00	15.00	BARO	26	67.00	14.00	TROP
7	61.50	17.00	BARO	27	67.50	18.00	TROP
8	61.50	19.00	BARO	28	68.00	14.00	BARO
9	62.00	14.00	BARO	29	68.50	18.00	BARO
10	63.00	15.00	TROP	30	69.00	13.00	BARO
1	63.50	19.00	TROP	31	69.00	15.00	BARO
12	64.00	12.00	TROP	32	69.50	17.00	BARO
13	64.50	16.00	TROP	33	69.50	19.00	BARO
14	65.00	12.00	TROP	34	70.00	12.00	BARO
15	65.00	15.00	TROP	35	70.50	16.00	BARO
16	65.00	17.00	TROP	36	71.00	17.00	BARO
17	65.50	16.00	TROP	37	71.50	21.00	BARO
18	65.50	19.00	TROP				
19	65.50	21.00	TROP				
20	66.00	13.00	TROP				

ნახ. 11.1.

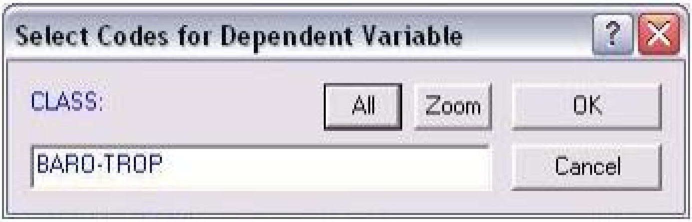
- ჩანართში Advanced დააწკაპუნეთ ღილაკზე Variables, გამოყავით ცვლადები, როგორც ეს ნაჩვენებია ნახაზზე და დააწკაპუნეთ OK-ზე. გამონათდება ნახ. 11.2-ზე წარმოდგენილი ფანჯარა.



ნახ. 11.2.

- განაყოფში Codes for variables (ჩანართში Advanced) დააწკაპუნეთ ღილაკზე Dep. Variable, ხოლო გამონათებულ ფანჯა-

რაში - შესაბამისად ღილაკს All-ზე, ხოლო შემდეგ OK-ზე (ნახ. 11.3).

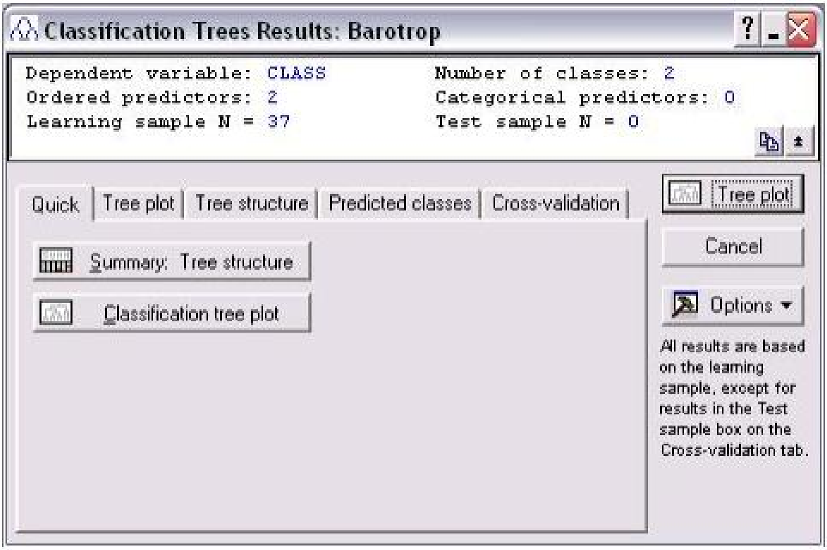


ნახ. 11.3.

- საბოლოოდ უნდა მივიღოთ შემდეგი ფანჯარა (ნახ. 11.4).

ნახ. 11.4.

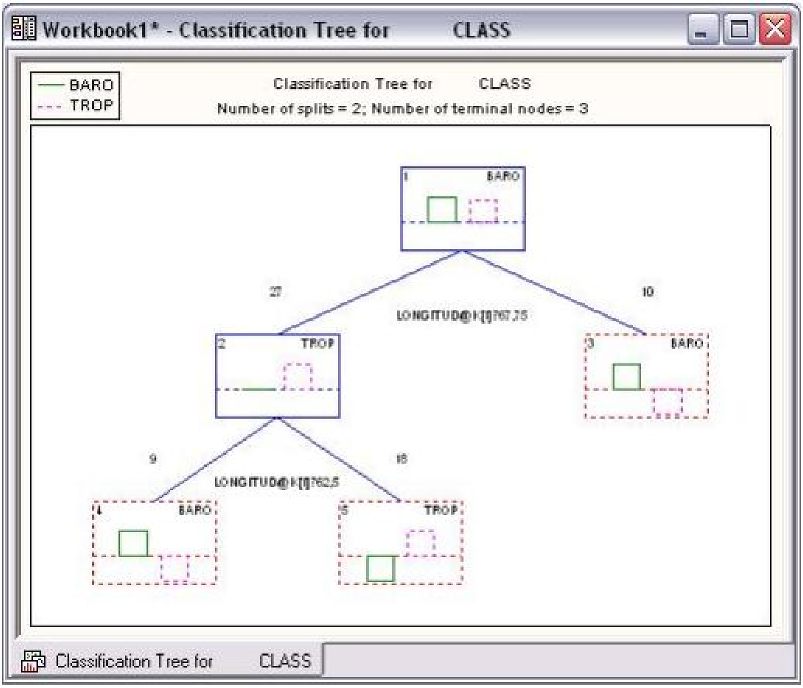
- OK-ზე დაწკაპუნებით მიიღებთ ნახ. 11.5-ზე გამოსახულ ფანჯარას.



ნახ. 11.5.

- გახსნილ დიალოგის ფანჯარაში ჩანართში Tree structure მორიგეობით დააწექით ღილაკებს შედეგებისა და მონაცემთა ანალიზის გამოსაყვანად: (OK (Tree plot), Classification tree structure, Importance plot) და შესაბამისი სურათების მისაღებად.

Class ცვლადისათვის „კლასიფიკაციის ხემ“, რომელიც იყენებდა ოპციას **ხეების სრული გადარჩევა ერთგანზომილებიანი განშტოება CART მეთოდით,** შეძლო ყველა 37 ციკლონის სწორად კლასიფიცირება. „ხის გრაფი“ ამ „კლასიფიკაციისთვის ხისთვის“ მოყვანილია ნახ. 11.6-ზე.



ნახ. 11.6.

„გრაფის“ დასათაურებაში მოყვანილია ზოგადი ინფორმაცია, რომლის მიხედვითაც მიღებულ„კლასიფიკაციის ხეს“ გააჩნია 2 განშტოება და 3 ტერმინალური წვერო. ტერმინალური წვეროები (ან, როგორც მათ ზოგჯერ ეძახიან, ფურცლები) - ეს „ხის“ კვანძებია, რომლებისგანაც დაწყებული აღარავითარი გადაწყვეტილებები აღარ მიიღებიან. ნახატზე ტერმინალური წვეროები ნაჩვენებია წითელი წყვეტილი ხაზებით, ხოლო დანარჩენები - მთლიანი შავი ხაზებით. „ხის“ დასაწყისად ითვლება ყველაზე ზედა გადამწყვეტი წვერო, რომელსაც ზოგჯერ ასევე „ხის“ ძირს ეძახიან. ნახაზზე ის მოთავსებულია მარცხენა ზედა კუთხეში და აღნიშნულია ციფრით 1. თავდაპირველად ყველა 37 ციკლონი მიეწერება ამ ძირეულ წვეროს და წინასწარ კლასიფიცირდება როგორც Baro - ამაზე მიუთი-

თებს წარწერა Baro წვეროს მარჯვენა ზედა კუთხეში. კლასი Baro არჩეული იყო საწყისი კლასიფიკაციისათვის იმიტომ, რომ Baro ციკლონთა რიცხვი ბევრად მეტია, ვიდრე Trop ციკლონების (იხ. ძირეული წვეროს შიგნით გამოსახული ჰისტოგრამა). „გრაფის“ ზედა მარცხენა კუთხეში არის წარწერა - ლეგენდა, რომელიც მიუთითებს თუ პიქტოგრამის რომელი სვეტები შეესაბამებიან Baro და Trop ციკლონებს.

ძირეული წვერო იტოტება ორ ახალ წვეროდ. ძირეული წვეროს ქვემოთ არის ტექსტი, რომელიც აღწერს მოცემული განშტოების სქემას. მისგან გამომდინარეობს, რომ ციკლონებს, რომელთა გრძედის მნიშვნელობა ნაკლებია ან ტოლი 67.75-ის, მიკუთვნებულებია წვეროზე ნომრით 2 და სავარაუდოდ კლასიფიცირებულებია როგორც Trop, ხოლო ციკლონები 67.75-ზე მეტი გრძედით მიწერილებია წვეროზე 3 და კლასიფიცირებულებია როგორც Baro. რიცხვები 27 და 10 წვეროებზე 2 და 3 შესაბამისად აღნიშნავენ დაკვირვებათა რიცხვებს, რომლებიც მოხვდნენ ამ ორ შვილობილ წვეროში მშობლიური ძირეული წვეროდან. შემდეგ ზუსტად ასევე განშტოვდება წვერო 2. შედეგად 9 ციკლონი გრძედის მნიშვნელობით ≤62.5, მიეწერება წვერო 4-ს და კლასიფიცირდება როგორც Baro, ხოლო დანარჩენი 18 ციკლონი, რომელთა გრძედიც მეტია 62.5-ზე მიკუთვნებულებია წვეროზე 5 და კლასიფიცირდებიან როგორც Trop.

„ხის გრაფზე“ მთელი ეს ინფორმაცია წარმოდგენილია მარტივ, აღსათქმელად მოხერხებულ სახეში. თუ ეხლა ჩვენ შევხედავთ „ხის“ ტერმინალური წვეროების ჰისტოგრამებს, განლაგებულებს ქვედა სტრიქონში, მაშინ დავინახავთ, რომ „კლასიფიკაციის ხემ“ შეძლო აბსოლუტურად სწორად მოეხდინა ციკლონების კლასიფიცირება. თითოეული ტერმინალური წვერო „სუფთაა“, ანუ არ შეიცავს არასწორად კლასიფიცირებულ დაკვირვებებს. „ხის გრაფში“ მოთავსებული მთელი ინფორმაცია დუბლირებულია შედეგების ცხრილში *ხის სტრუქტურა*, რომელიც მოყვანილია ნახ. 11.7-ზე.

Workbook3* - Tree Structure (Barotrop)

Tree Structure (Barotrop)
Child nodes, observed class n's,
predicted class, and split condition for each node

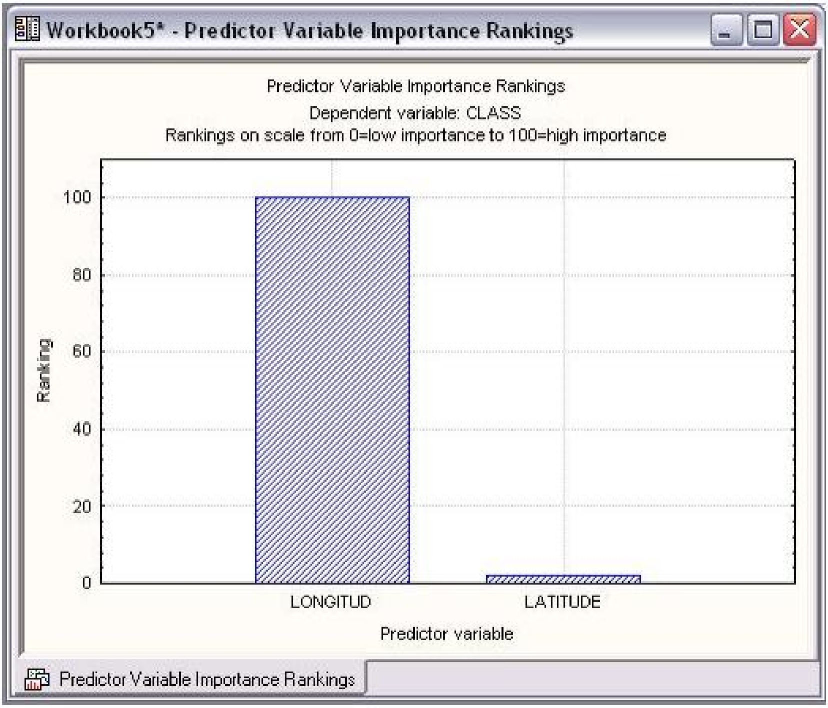
Node	**Left branch**	Right branch	n in cls BARO	n in cls TROP	Predict. class	Split constant	Split variable
1	2	3	19	18	BARO	-67,7500	LONGITUD
2	4	5	9	18	TROP	-62,5000	LONGITUD
3			10	0	BARO		
4			9	0	BARO		
5			0	18	TROP		

Tree Structure (Barotrop)

ნახ. 11.7.

მიაქციეთ ყურადღება იმას, რომ ამ შედეგების წვეროები ცხრილში 3- დან 5-ის ჩათვლით აღნიშნულებია როგორც ტერმინა-ლურები, ვინაიდან მათში არ ხდება განშტოება. მიაქციეთ აგრეთვე ყურადღება *განშტოების მუდმივების* ნიშნებს, მაგალითად, -67.75-ს წვეროსთვის 1.

თუ ხორციელდება ერთგანზომილებიანი განშტოება, მაშინ თითოეულ პრედიქტორულ ცვლადს შეიძლება მივაწეროთ რანგი შკალით 0-დან 100-მდე მისი ზემოქმედების ხარისხის მიხედვით დამოკიდებული ცვლადის ამოძახილზე. ჩვენს მაგალითში ცხადია, რომ გრძედს (Longitude) გააჩნია დიდი მნიშვნელობა, ხოლო განედს (Latitude) - შედარებით მცირე (ნახ. 11.8).

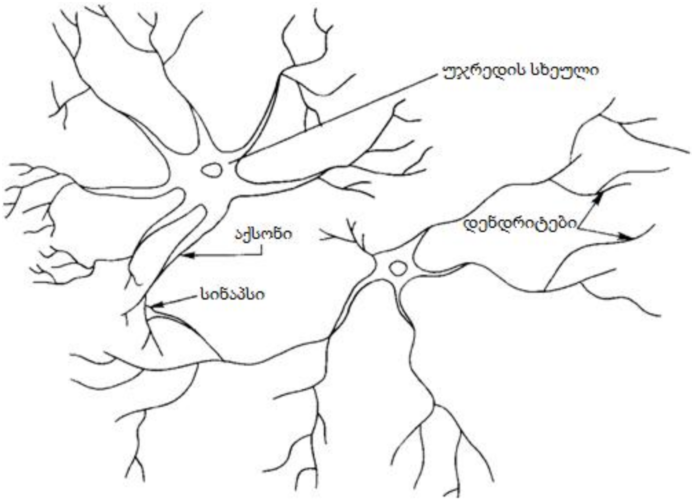

Workbook5* - Predictor Variable Importance Rankings
Predictor Variable Importance Rankings
Dependent variable: CLASS
Rankings on scale from 0=low importance to 100=high importance
Ranking
100
80
60
40
20
0
LONGITUD
LATITUDE
Predictor variable
Predictor Variable Importance Rankings

ნახ. 11.8.

12. ნეირონული ქსელები
12.1. ზოგადი ცნობები

ნეირონი წარმოადგენს ბიოლოგიურ უჯრედს, რომელიც ამუშავებს ინფორმაციას. თავის ტვინის ქერქი შეიცავს 10^{11} -მდე ნეირონს, რაც დაახლოებით ტოლია ირმის ნახტომის ვარსკვლავების რიცხვის. თითოეული ნეირონი დაკავშირებულია $10^3 – 10^4$ სხვა ნეირონებთან. მთლიანობაში ადამიანის ტვინი შეიცავს დაახლოებით 10^{14} -დან 10^{15} -მდე ნეირონთაშორის ურთიერთკავშირებს.

ნეირონი შედგება უჯრედის ტანისგან (cell body), ან სომისგან (soma), და ორი გარე ხისმაგვარი ტიპის განშტოებისაგან: აქსონის (axon) და დენდრიტებისგან (dendrites) (ნახ. 12.1). უჯრედის ტანი შეიცავს ბირთვს (nucleus), რომელიც შეიცავს ინფორმაციას შთამომავლობითი თვისებების შესახებ, პლაზმას, რომელსაც გააჩნია მოლეკულარული საშუალებები საჭირო ნეირონული მასალების წარმოებისთვის. ნეირონი იღებს სიგნალებს (იმპულსებს) სხვა ნეირონე-



ნახ. 12.1.

ბიდან დენდრიტების (მიმღებები) გავლით და გადასცემს უჯრედის ტანის მიერ გენერირებულ სიგნალებს აქსონის (გადამცემი) გასწვრივ, რომელიც ბოლოში იტოტება ბოჭკოებად (strands). ამ ბოჭკოთა დაბოლოებებზე იმყოფებიან სინაფსები (synapses).

მკაცრად რომ ვიმსჯელოთ, არაა დამტკიცებული, რომ ხელოვნური ნეირონული ქსელები, რომლებიც შექმნილებია როგორც ნერვული ქსოვილის მოდელი, სწორად ასახავენ ადამიანის ტვინის მუშაობას. ჩვენ ვისარგებლებთ ამ წარმოდგენით მხოლოდ უხეში შეფასებისათვის, რაც სრულებით საკმარისია პრაქტიკული მიზნებისათვის.

ადამიანის მეხსიერების უჯრედების ცალკეული ნეირონების ფიზიკური სწრაფქმედება სიდიდის 8 ხარისხით ნაკლებია თანამედროვე კომპიუტერის სწრაფქმედებაზე და აგებს სუპერ კომპიუტერებთან სიდიდის ათ ხარისხზე მეტად. ოღონდ ყველაზე მძლავრი კომპიუტერი „გადახურდება“ იმ ამოცანის გადაწყვეტისას, რომელსაც ადამიანი მყისიერად წყვეტს. გაიხსენეთ ადამიანის შესაძლებლობა ათეული წლების შემდეგ იცნოს ნაცნობი სახეები, რომლებთან ერთად ის შეიძლება ბავშვთა ბაღში დადიოდა. სხვა სიტყვებით რომ ვთქვათ, ასეთი რთული ამოცანებისათვის ტვინი „აამოქმედებს“ პარალელურ პროგრამებს, რომლებიც შეიცავენ დაახლოებით 100 ბიჯს. ეს ცნობილია, როგორც ასი ბიჯის წესი. ანალოგიურად თუ ვიმსჯელებთ, შეიძლება აღმოვაჩინოთ, რომ ერთი ნეირონიდან მეორესთვის გაგზავნილი ინფორმაციის რაოდენობა უნდა იყოს ძლიერ მცირე (რამდენიმე ბიტი). აქედან გამომდინარეობს, რომ ძირითადი ინფორმაციის უშუალოდ გადაცემა არ ხდება, არამედ ის მოიპოვება და ნაწილდება ნეირონებს შორის კავშირებში.

12.2. ტექნიკური ნეირონის მოდელი

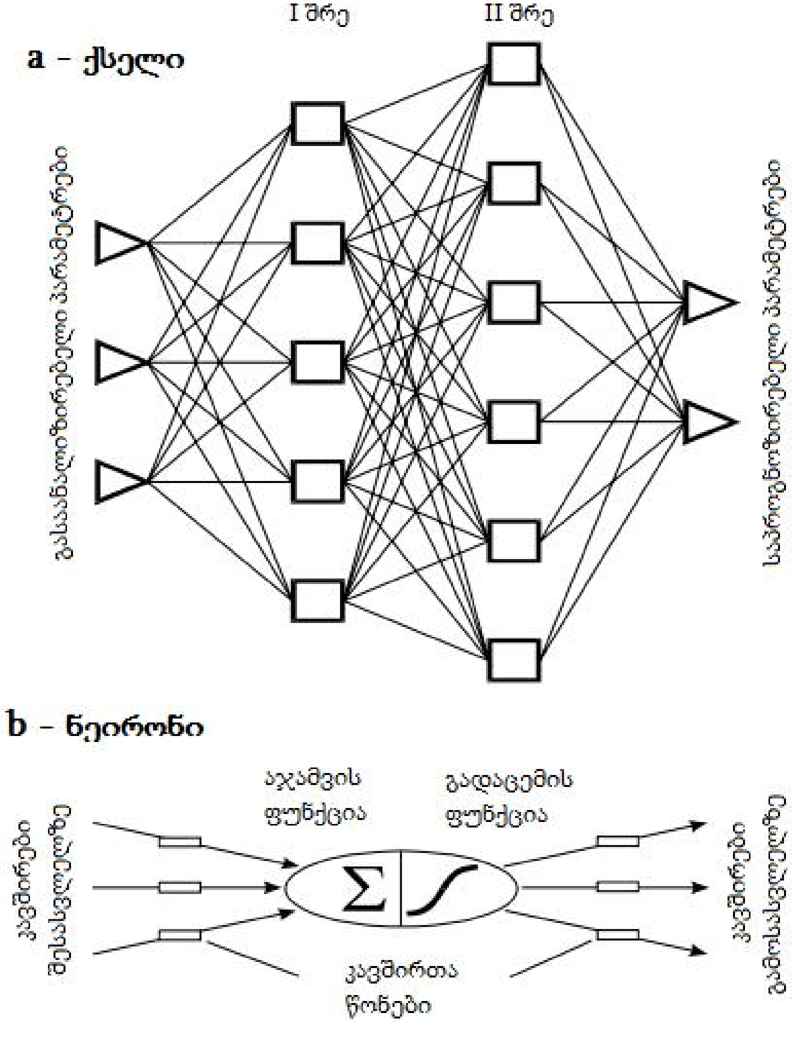
მაკ-კალოკმა და პიტსმა შემოგვთავაზეს ხელოვნური ნეირონის მოდელად ზღვრული ბინარული ელემენტების გამოყენება. ეს მათემატიკური ნეირონი ითვლის n შესასვლელი x_j , j=1,2,...,n სიგნალის შეწონილ ჯამს და გამოსასვლელზე აფორმირებს 1-ის სიგნალს, თუ ეს ჯამი აღემატება განსაზღვრულ u ზღვარს, და 0-ს წინააღმდეგ შემთხვევაში.

მოხერხებულია u-ს განხილვა როგორც წონითი კოეფიციენტისა, დაკავშირებულისა მუდმივ x_o=1 შესასვლელთან. დადებითი წონითი კოეფიციენტები შეესაბამებიან აგზნებულ კავშირებს, ხოლო უარყოფითები - დამუხრუჭებულებს. მაკ-კალოკმა და პიტსმა დაამტკიცეს, რომ შესაბამისი სახით შერჩეული წონებისას მსგავსი ტიპის პარალელურად მომუშავე ნეირონების ერთობლიობას შეუძლია შეასრულოს უნივერსალური გამოთვლები.

ნეირონული ქსელი შეიძლება განიხილებოდეს როგორც მიმართული გრაფი შეწონილი კავშირებით, რომელშიც ხელოვნური ნეირონები კვანძებს წარმოადგენენ. კავშირების არქიტექტურის მიხედვით ქსელები შეიძლება დაჯგუფდნენ ორ კლასად: პირდაპირი გავრცელების ქსელები, რომლებშიც გრაფებს არ გააჩნიათ მარყუჟები, და რეკურენტული ქსელები, ანუ ქსელები უკუ კავშირებით.

ყველაზე ფართო გავრცელება მოიპოვა პირველი კლასის ოჯახმა. მათ მრავალშრიანი პერსეპტრონები ჰქვიათ. მათში ნეირონები შრეებად არიან განლაგებულები და გააჩნიათ ერთმიმართულებიანი კავშირები შრეებს შორის. პირდაპირი გავრცელების ქსელები სტატიკურებია იმ აზრით, რომ მოცემულ შესასვლელზე ისინი გამოიმუშავებენ გამოსასვლელი მნიშვნელობების ერთ ერთობლიობას, რომელიც არაა დამოკიდებული ქსელის წინა მდგომარეობაზე. რეკურენტული ქსელები დინამიურებია, ვინაიდან

მათში უკუ კავშირებით მოდიფიცირდებიან ნეირონთა შესასვლე-ლები, რასაც ქსელის მდგომარეობის შეცლასთან მივყავართ.



ნახ. 12.2.

პირდაპირი გავრცელების ნეირონული ქსელი წარმოადგენს სტრუქტურას, რომელიც შედგება კვანძებისა და მათ შორის კავშირებისაგან (ნახ. 12.2). ქვედა დონის კვანძებს რეცეპტორები ჰქვიათ, ზედა დონისას - გადამწყვეტი კვანძები, ხოლო ზედა და ქვედა დონის კვანძებს შორის, როგორც წესი, არის ფარული შრეების რაღაც რაოდენობა. რეცეპტორების რაოდენობა ტოლია მონაცემების საკვლევი ბაზის ბაზის ველების რიცხვის, გადამწყვეტი კვანძების რიცხვი ტოლია საწინასწარმეტყველო პარამეტრების რიცხვის. სისტემის რეცეპტორებზე ეწოდება რაღაც ჩანაწერის მნიშვნელობა, გადამწყვეტი კვანძებიდან ამოიკითხება საწინასწარმეტყველო პარამეტრების მნიშვნელობები.

ქსელის თითოეულ კვანძს მიეწერება რაღაც რიცხვითი მნიშვნელობა. ქვედა დონის თითოეულ კვანძზე ან რეცეპტორზე მიწერილი მნიშვნელობა ტოლია ჩანაწერიდან რაღაც ველის (ცვლადის) მნიშვნელობის, რომელიც მოცემულ მომენტში ეწოდება ნეირონული ქსელის შესასვლელს. უფრო ზედა დონის ნეირონებთან დაკავშირებული მნიშვნელობები განისაზღვრებიან შემდეგი ალგორითმის მიხედვით. განსწავლულ ქსელში ნეირონებს შორის თითოეულ კავშირს მიეწერება რაღაც წონა, ნამდვილი რიცხვი, რომელიც როგორც წესი, მოთავსებულია -1-დან +1-მდე დიაპაზონში. განცალკევებული ნეირონი პირობითად ნაჩვენებია ქვედა ნაწილში (ნახ. 9.2). რომ განვსაზღვროთ რომელიმე ნეირონთან დაკავშირებული მნიშვნელობა, საჭიროა გამოითვალოს ნეირონის შესასვლელი დონე - ამ ნეირონთან დაკავშირებული წინა დონის ყველა ნეირონის მნიშვნელობათა შეწონილი ჯამი; მამრავლები შესაბამისი კავშირების წონების ტოლია. თუ a_i ნეირონის i-ურ შესასვლელთან დაკავშირებული მნიშვნელობაა, ხოლო w_i - შესაბამისი კავშირის წონა, მაშინ შესასვლელი დონე განისაზღვრება როგორც $p=\sum a_i w_i$.

თითოეულ ნეირონს გააჩნია ე.წ. გადამცემი f ფუნქცია, რომელიც ზღვარითი დონის თითოეული მნიშვნელობისათვის აბრუნებს

ნეირონზე მიწერილ რაღაც $t=f(p)$ რიცხვს. ჩვეულებრივ ეს მონოტონურად ზრდადი ასიმეტრიული ფუნქციაა $f(-\infty)=-1$, $f(0)=0$ და $f(\infty)=1$-ით. ხშირად ასეთ ფუნქციად იღებენ ჰიპერბოლურ ტანგენსს

$$f(p)=\frac{e^{kp}-e^{-kp}}{e^{kp}+e^{-kp}},$$

სადაც $k>0$ - დამოკიდებულების დახრილობის განმსაზღვრელი პარამეტრია $p=0$ - ის დროს. ამრიგად, თანმიმდევრულად გამოითვლებიან მნიშვნელობები სულ უფრო მაღალი შრეების ნეირონებისათვის, სანამ, ბოლოს, არ გამოითვლებიან მნიშვნელობები გადამწყვეტი კვანძებისათვის, ანუ საწინასწარმეტყველო პარამეტრები.

გამოყენების წინ საჭიროა ნეირონული ქსელის წინასწარი სწავლება. არსებობს ნეირონული ქსელების სწავლების ალგორითმების ბევრი ნაირსახეობა. მათი საერთო იდეა შემდეგში მდგომარეობს.

თავდაპირველად ნეირონებს შორის ყველა კავშირის წონებს გააჩნიათ ნებისმიერი, როგორც წესი, ერთნაირი მნიშვნელობები. შემდეგ მრავალჯერ ტარდება შემდეგი პროცესი. მონაცემთა ჯგუფიდან, რომელიც გამოიყენება სწავლებისთვის და რომლისთვისაც ცნობილია როგორც შესასვლელი, ისე გამოსასვლელი ცვლადების მნიშვნელობები, იღება ერთი ჩანაწერი და ეწოდება ნეირონული ქსელის შესასვლელს. აღწერილი ალგორითმით ითვლიან გამოსასვლელ მნიშვნელობებს. ცხადია, რომ მიღებული მნიშვნელობები ძლიერ შორსაა საძიებელი დამოკიდებული ცვლადების მნიშვნელობებისგან. მაშინ ხორციელდება ნეირონებს შორის კავშირების ისეთი ცვლილება, რომელიც ამცირებს ამ სხვაობას. ეს ოპერაცია მრავალჯერ ტარდება სასწავლო შერჩევის თითოეული ჩანაწერისათვის. როგორც წესი, სწავლების პროცესი იწყება ყველაზე მაღალი დონის ნეირონების კავშირებიდან და შემდეგ ვრცელდება ქვემოთ მდებარე შრეების ნეირონებზე, ამიტომ სწავლების ამ მეთოდს უწოდებენ შეც-

დომის უკუ გავრცელებას. სწავლის შედეგად ვღებულობთ ქსელს, რომელიც მეტ-ნაკლებად სწორად აღწერს გამოსასვლელი პარამეტრის მნიშვნელობას სწავლების მაგალითებისათვის, და უნდა ველოდოთ, რომ ის ასევე კარგად იწინასწარმეტყველებს გამოსასვლელ პარამეტრებს სხვა მონაცემებზეც.

ნეირონული ქსელების პროექტირებისა და სწავლისას საჭიროა დავიცვათ განსაზღვრული სიფრთხილე, ვინაიდან ძირითად ამოცანას წარმოადგენს განზოგადების, ან პროგნოზის მინიმალური შეცდომის მიღწევა და არა სწავლების. ხელოვნური ნეირონული ქსელის სწავლების ნეგატიური შედეგი შეიძლება იყოს ე. წ. „გადასწავლება“, რომლის დროსაც ქსელი, არსებითად, იმახსოვრებს სწავლის ფაქტებს; ამასთან მას გააჩნია მინიმალური წინასწარმეტყველების შესაძლებლობა, თუ მის შესასვლელს ეწოდება ახალი ფაქტები.

„გადასწავლულ“ ნეირონულ ქსელს არ ექნება მოქნილობა, განზოგადების შესაძლებლობა. „გადასწავლების“ ალბათობის შესამცირებლად, შესაძლებლობის ფარგლებში, საჭიროა არ გავზარდოთ შრეებისა და ნეირონების რაოდენობა ფარულ შრეებში და გამოვიყენოთ რაც შეიძლება მეტი ფაქტი სწავლის დროს. არსებობენ ემპირული წესები, რომლებიც გვეხმარებიან განვსაზღვროთ ფარული ნეირონებისა და განსასწავლი ფაქტების რიცხვის ცვლილების დიაპაზონები.

ნეირონულ ქსელებს შემდეგი ნაკლოვანებები აქვთ:

- *პ ი რ ვ ე ლ ი* ნაკლოვანება იმაში მდგომარეობს, რომ ძალიან რთულია სწავლის პროცესში მიღებული პროგნოზირების მოდელების სტატისტიკური მნიშვნელობების შეფასება. ქსელის სწავლება მიმდინარეობს ისე, რომ მან ძალიან კარგად იწინასწარმეტყველოს გამოსასვლელი პარამეტრების მნიშვნელობები ამ ჩანაწერებზე, მაგრამ ძალიან რთულია გავიგოთ, რამდენად მდგრადია ეს დამოკიდებულება, რომელიც განსაზღვრავს პროგნოზირებად მნიშვნელობებს, რამდენად მნიშვნელოვანია მიღებული კავშირი, და იმუშავებს თუ

არა ის ასევე კარგად სხვა მონაცემებზე. მთელი საქმე აქ თავისუფლების ხარისხის დიდ რაოდენობაშია. ფაქტიურად, თავისუფლების ხარისხს წარმოადგენს ნეირონებს შორის კავშირების თითოეული წონა, და თუ, მაგალითად, ჩვენი ქსელი შედგება რამდენიმე ათეული ნეირონისგან, მისი თავისუფლების ხარისხის რიცხვი შეადგენს რამდენიმე ასეულს. ცხადია, რომ თავისუფლების ამ ხარისხების მისადაგებით შეიძლება მივაღწიოთ ძლიერ ზუსტ წინასწარმეტყველებას სასწავლო შერჩევისთვის, მაგრამ სრულებით არაა ცხადი, რომ წინასწარმეტყველება ასევე მართებული იქნება ახალი მონაცემებისთვის, რომლებიც არ გამოიყენებოდნენ სწავლების პროცესში.

- *მ ე ო რ ე* ნაკლოვანება იმაში მდგომარეობს, რომ ნეირონული ქსელი ძალზე სპეციფიკური სტრუქტურაა; მას კარგად შეუძლია მოახდინოს იმიტაცია და გამოსახვა შესაძლო დამოკიდებულებების მხოლოდ საკმაოდ ვიწრო წრისთვის. ბუნებრივია მისი გამოყენება, ვთქვათ, სახეობათა გამოცნობის ზოგიერთი ანალოგური განაწილებული სისტემის ასაწერად, როცა კავშირთა წონები შეიძლება იყვნენ უშუალოდ ფიზიკურად ინტერპრეტირებულები როგორც ქსელის ცალკეულ ნეირონთა კავშირების ენერგიები. ოღონდ მსგავსი დამატებები ძლიერ იშვიათია. ხოლო, ვთქვათ, ძლიერ არაწრფივი დამოკიდებულებები, დამოკიდებულებები წყვეტებით, მახვილი პიკებით და სხვა თავისებურებებით შედარებით ცუდად აღითქმებიან ნეირონული ქსელით მისი უწყვეტი ბუნების გამო. მიუხედავად ამისა, ეს სისტემათა ძლიერ პოპულარული კლასია, რომელიც ხშირად გამართულად მუშაობს და გამოიყენება ბევრ გამოყენებით სფეროში.

- ნეირონული ქსელების პარადიგმის *მ ე ს ა მ ე* ნაკლოვანებაა შემსწავლელი შერჩევის ძალიან დიდი მოცულობის არსებობის აუცილებლობა. აქედან გამომდინარე ნეიროქსელური მეთოდები საკმაოდ მომთხოვნებია გამოთვლითი სიმძლავრეების მიმართ და ყველაზე მძიმე შემთხვევებში, როცა მოითხოვება მონაცემების დიდი რაოდენობის დამუშავება, აზრი აქვს პარალელური არქიტექტურის

გამოყენებას. საკმაოდ ეფექტურია ისეთი სისტემები, სადაც თითოეული პროცესორი რეალიზაციას უწევს ცალკეულ ნეირონს; ასეთი არქიტექტურის აპარატურული რეალიზაცია ფრიად ბუნებრივია. ოღონდ ასეთი პროგრამულ-აპარატურული კომპლექსები ძლიერ ძვირებია. სწორედ მათმა მაღალმა ღირებულებამ განსაზღვრა ისეთი პროგრამების ფართო გავრცელება, რომლებიც ახდენენ ნეირონული ქსელის მუშაობის ემულაციას ჩვეულებრივ კომპიუტერებზე.

ძირითადი ნაკლოვანება, რომელიც აფერხებს ნეირონული ქსელების გამოყენებას ცოდნის ამაღლებისათვის - მათი „გაუმჭვირვალეობაა“. აგებული მოდელი, როგორც წესი, წარმოადგენს „შავ ყუთს“ და არ გააჩნია მკაფიო ინტერპრეტაცია. ცოდნა, რომელიც დაფიქსირებულია როგორც რამდენიმე ასეულ ნეირონს შორის კავშირების წონები, სრულებით არ ექვემდებარება ადამიანის მიერ ანალიზსა და ინტერპრეტაციას. გარდა ამისა, სწორად აგებულ ნეირონულ ქსელს თუმცა შეუძლია ადექვატურად შეაფასოს მსგავსი სიტუაციები, ჩვეულებრივ ცუდად ატარებს პრინციპულად ახალი სიტუაციის ანალიზს, რომელიც არაა წარმოდგენილი მაგალითების სახით სწავლის დროს.

12.3. როგორ ამოცანებს ხსნიან ნეირონული ქსელები

ნეირონული ქსელები ყველაზე მეტად მომარჯვებულებია სახეობათა დამუშავებასთან ასე თუ ისე დაკავშირებული ამოცანების ფართო წრის გადაწყვეტასთან. ქვემოთ მოყვანილია ნეირონული ქსელებით გადაწყვეტის ტიპიური ამოცანების სია:

- ფუნქციათა აპროქსიმაცია წერტილების ნაკრების მიხედვით (რეგრესია);

- მონაცემების კლასიფიკაცია კლასების მოცემული ნაკრების მიხედვით;
- მონაცემების კლასტერიზაცია წინასწარ უცნობი პრეს-პროტოტიპების გამოვლენით;
- ინფორმაციის შეკუმშვა;
- დაკარგული მონაცემების აღდგენა;
- ასოციაციური მეხსიერების აგება;
- ოპტიმიზაცია, ოპტიმალური მართვა.

ეს სია შეგვეძლო გაგვეგრძელებინა. ოღონდ შევნიშნოთ, რომ ყველა ამ გარეგნულად განსხვავებული ამოცანების დასმებს შორის არსებობს ღრმა შინაგანი კავშირი. მათში ჩანს რაღაც ერთიანი პროტოტიპი, რომელიც საშუალებას მოგვცემს წარმოსახვის გარკვეული წილით დავიყვანოთ ისინი ერთმანეთამდე.

ავიღოთ, მაგალითად, წერტილების ნაკრების მიხედვით ფუნქციის *აპროქსიმაციის* ამოცანა. ეს არაკორექტული ამოცანის ტიპიური მაგალითია, ანუ ამოცანისა, რომელსაც არა აქვს ერთადერთი ამოხსნა. ერთადერთობის მისაღწევად საჭიროა ასეთი ამოცანების *რეგულარიზაცია* - მარეგულირებელი რაღაც ფუნქციონალის მინიმიზაციის მოთხოვნის დამატებით. სწორედ ასეთი ფუნქციონალის მინიმიზაცია არის ნეირონული ქსელის სწავლების მიზანი. *ოპტიმიზაციის* ამოცანებიც ასევე დადიან მიზნის ფუნქციათა მინიმიზაციასთან შეზღუდვების მოცემული ნაკრების დროს. მეორეს მხვრივ, *კლასიფიკაცია* - სხვა არაფერია, თუ არა ფუნქციის აპროქსიმაცია დისკრეტული მნიშვნელობებით (კლასების იდენტიფიკატორებით), თუმცა ის შეიძლება განვიხილოთ როგორც მონაცემთა ბაზებში გამოტოვებული მონაცემების შევსების კერძო შემთხვევა, მოცემულ შემთხვევაში - კლასების იდენტიფიკატორთა სვეტში. დაკარგული მონაცემების *აღდგენის* ამოცანა, თავის მხვრივ - ეს *ასოციაციური მეხსიერებაა,* რომელიც აღადგენს მის პირველსახეს მისი ნაწილით. ასეთი პირველსახეებით

კლასტერიზაციის ამოცანაში გამოდიან კლასტერთა ცენტ-რები. და ბოლოს, თუ შესაძლებელია ინფორმაციის აღდგენა მისი რაღაც ნაწილით, მაშინ ეს ნიშნავს, რომ ჩვენ მივაღწიეთ ამ ინფორმა-ციის *შეკუმშვას,* და ა. შ.

ნეიროკომპიუტინგი იძლევა პრაქტიკულად საინტე-რესო ამოცანათა ძალიან ფართო წრის ამოხსნის ერთიან მეთოდო-ლოგიას. ეს სხვა არაფერია, თუ არა მონაცემების ანალიზის ახალი ინსტრუმენტი და სხვებზე უკეთ ის შეიძლება გამოიყენოს სწორედ სპეციალისტმა მის საგნობრივ სფეროში. ძირითადი სიძნელეები ნე-იროტექნოლოგიების კიდევ უფრო ფართო გავრცელების გზაზე - ესაა პროფესიონალთა ფართო წრის უუნარობა ჩამოაყალიბოს თავი-სი პრობლემები ტერმინებში, რომლებიც დაუშვებენ მარტივ ნეირო-ქსელურ გადაწყვეტას.

12.4. ნეირონული ქსელები პაკეტში STATISTICA

12.4.1. აპროქსიმაცია

მაგალითის მიზანია მოახდინოს ნეირონულ-ქსელური მიდ-გომის რეალიზაციის ძირითადი ეტაპების დემონსტრირება აპროქ-სიმაციის ამოცანის გადასაწყვეტად. მაგალითის სახით განვიხილოთ შემდეგი ამოცანა: მოცემულია მასივი, შემდგარი $y=a+bx$ ფუნქცი-ის რამდენიმე მნიშვნელობისაგან.

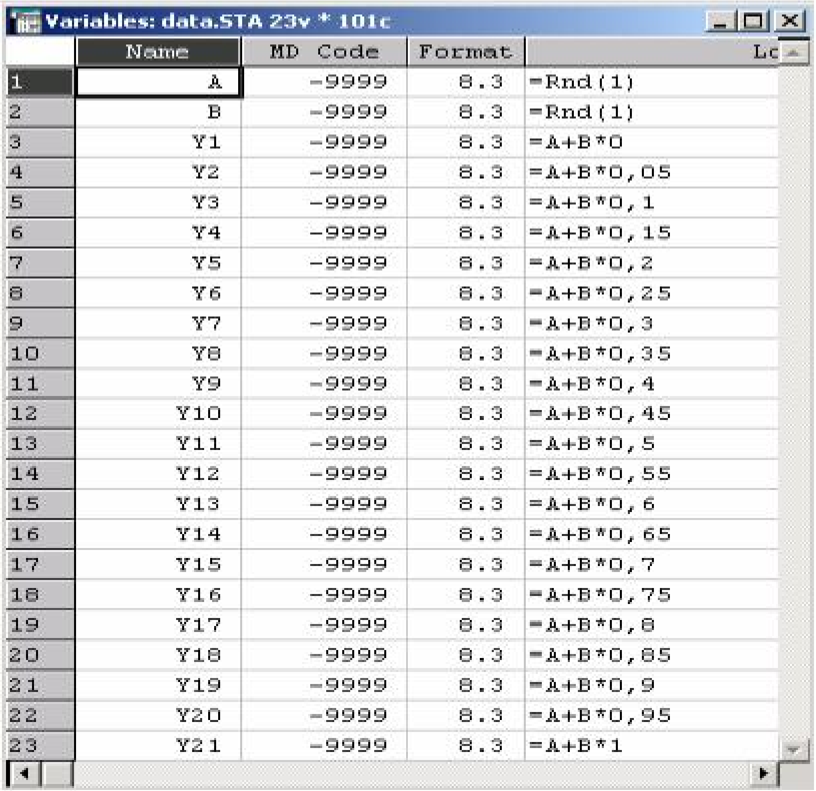
უნდა შევქმნათ ისეთი ნეირონული ქსელი, რომ ამ მნიშვნე-ლობების ქსელის შესასვლელებზე მიწოდებით გამოსასვლელებზე მივიღოთ a და b პარამეტრების მნიშვნელობები. პირველ რიგში სა-ჭიროა განვსაზღვროთ შესასვლელი მასივის განზომილება. y ფუნქ-ციის მნიშვნელობათა რაოდენობად ავირჩიოთ $N=21$, ანუ მასივის შესასვლელ ვექტორებად გამოვიყენოთ ფუნქციის მნიშვნელობები

x=0; 0,05; ... ,1,0 წერტილებში. ქსელის სწავლებისთვის აუცილებელია შესასვლელ ვექტორთა მასივის ფორმირება *a* და *b* პარამეტრების სხვადასხვა ნაკრებისათვის. შემთხვევითი სახით ავირჩიოთ *a* და *b* -ს მნიშვნელობები (პარამეტრების ცვლილების დიაპაზონი ავიღოთ ავიღოთ *0,0*-დან *1,0*-მდე და გამოვითვალოთ შესასვლელი ვექტორების შესაბამისი კომპონენტები. ვიმეორებთ ამ პროცედურას *M=100* -ჯერ და ვღებულობთ შესასვლელ ვექტორთა მასივს მატრიცის სახით, რომლის განზომილებაა (N+2)*M. მიღებული მასივი ჩვენ შეგვიძლია გამოვიყენოთ ქსელის სწავლებისთვის. ამ მასივის ფრაგმენტი მოყვანილია ცხრ. 12.1-ში.

ცხრილი 12.1.

1 A	2 B	3 Y1	4 Y2	5 Y3	6 Y4	7 Y5	8 Y6	9 Y7	10 Y8	11 Y9
,408	,409	,408	,428	,449	,469	,490	,510	,530	,551	,57
,538	,020	,538	,539	,540	,541	,542	,543	,544	,545	,54
,947	,684	,947	,981	1,015	1,049	1,084	1,118	1,152	1,186	1,22
,239	,580	,239	,268	,297	,326	,355	,384	,413	,442	,47
,556	,590	,556	,585	,615	,644	,674	,703	,733	,762	,79
,284	,869	,284	,327	,371	,414	,458	,501	,545	,588	,63
,792	,390	,792	,811	,831	,850	,870	,889	,909	,928	,94
,505	,406	,505	,526	,546	,566	,587	,607	,627	,648	,66
,205	,162	,205	,213	,222	,230	,238	,246	,254	,262	,27
,428	,347	,428	,445	,462	,480	,497	,514	,532	,549	,56
,763	,337	,763	,779	,796	,813	,830	,847	,864	,881	,89
,342	,528	,342	,368	,395	,421	,448	,474	,500	,527	,55
,584	,568	,584	,613	,641	,670	,698	,726	,755	,783	,81
,220	,812	,220	,261	,301	,342	,382	,423	,464	,504	,54
,811	,884	,811	,855	,899	,944	,988	1,032	1,076	1,120	1,16
,281	,083	,281	,285	,290	,294	,298	,302	,306	,310	,31
,427	,272	,427	,441	,455	,468	,482	,495	,509	,522	,53

შესასვლელი მასივისთვის საწყისი მონაცემები მოხერხებულია მოვამზადოთ სხვა მოდულში, მაგალითად, Basic Statistics/Tables - ში. ცვლადების სპეციფიკაცია წარმოდგენილია ნახ. 12.3-ზე.



Variables: data.STA 23v * 101c

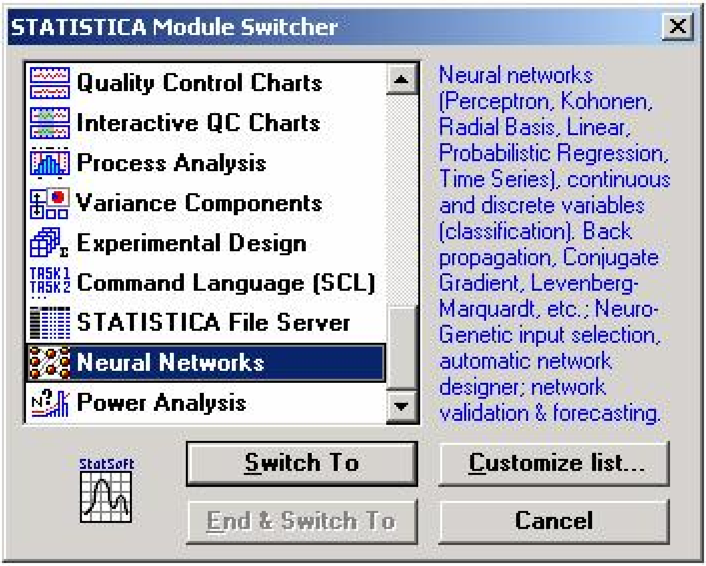
	Name	MD Code	Format	Lc
1	A	-9999	8.3	=Rnd(1)
2	B	-9999	8.3	=Rnd(1)
3	Y1	-9999	8.3	=A+B*0
4	Y2	-9999	8.3	=A+B*0,05
5	Y3	-9999	8.3	=A+B*0,1
6	Y4	-9999	8.3	=A+B*0,15
7	Y5	-9999	8.3	=A+B*0,2
8	Y6	-9999	8.3	=A+B*0,25
9	Y7	-9999	8.3	=A+B*0,3
10	Y8	-9999	8.3	=A+B*0,35
11	Y9	-9999	8.3	=A+B*0,4
12	Y10	-9999	8.3	=A+B*0,45
13	Y11	-9999	8.3	=A+B*0,5
14	Y12	-9999	8.3	=A+B*0,55
15	Y13	-9999	8.3	=A+B*0,6
16	Y14	-9999	8.3	=A+B*0,65
17	Y15	-9999	8.3	=A+B*0,7
18	Y16	-9999	8.3	=A+B*0,75
19	Y17	-9999	8.3	=A+B*0,8
20	Y18	-9999	8.3	=A+B*0,85
21	Y19	-9999	8.3	=A+B*0,9
22	Y20	-9999	8.3	=A+B*0,95
23	Y21	-9999	8.3	=A+B*1

ნახ. 12.3.

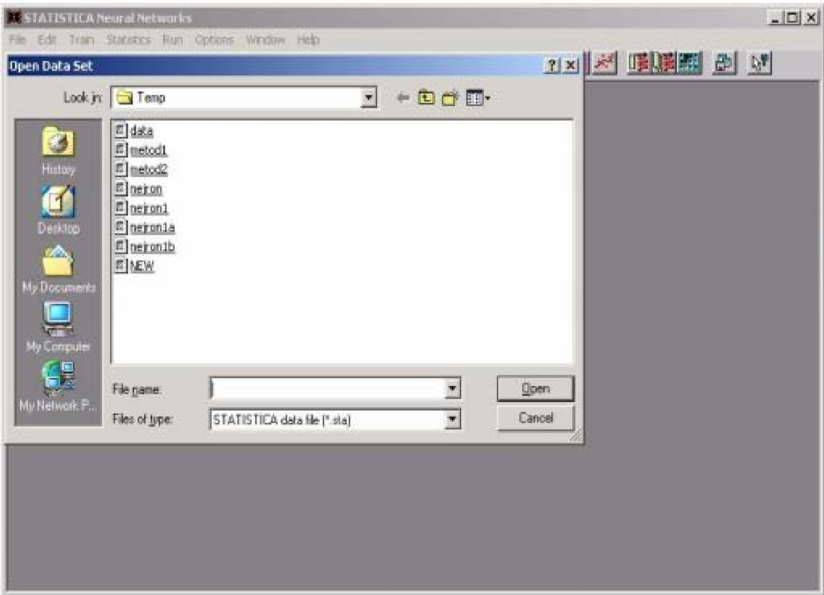
სისტემაში STATISTICA ფანჯარაში Module Switcher ავირჩიოთ მოდული Neural Networks (ნახ. 12.4).

უნდა ავღნიშნოთ, რომ ეს მოდული პაკეტის სტანდარტულ კონფიგურაციაში არ შედის და ცალკე ინსტალირდება!

გავხსნათ data.sta ფაილი მონაცემებით შესასვლელი მასივისთვის (ნახ. 12.5).

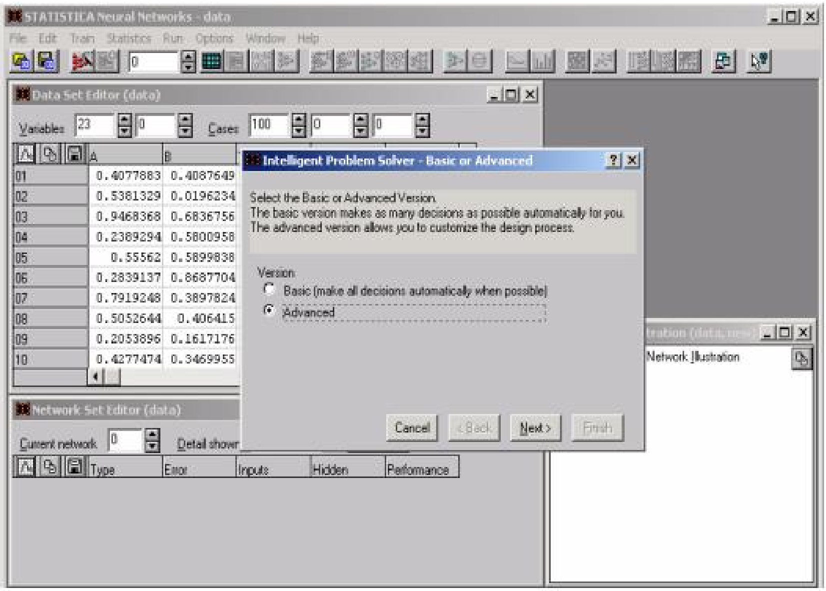


ნახ. 12.4.



ნახ. 12.5.

შევეცადოთ ვისარგებლოთ Inelligent Problem Solver-ით და მისი Advansed ვერსიით, რომელიც საშუალებას იძლევა ვმართოთ ქსელის პროექტირების პროცესი (ნახ. 12.6).

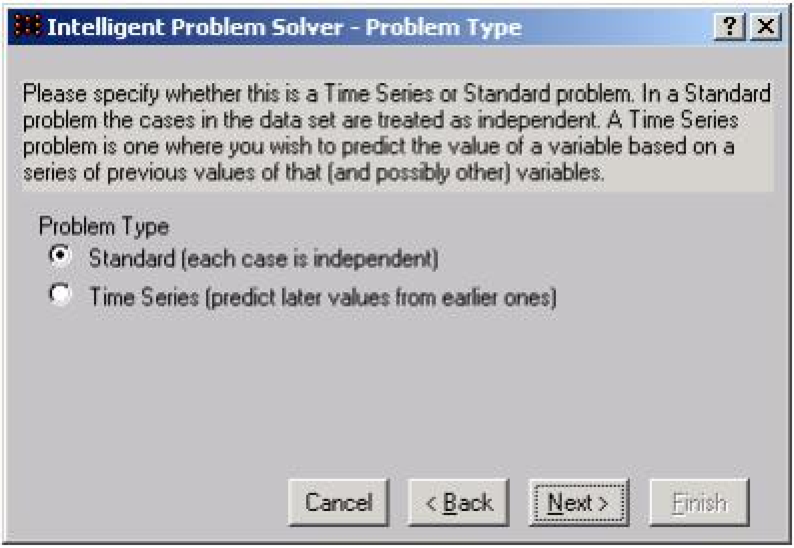


ნახ. 12.6.

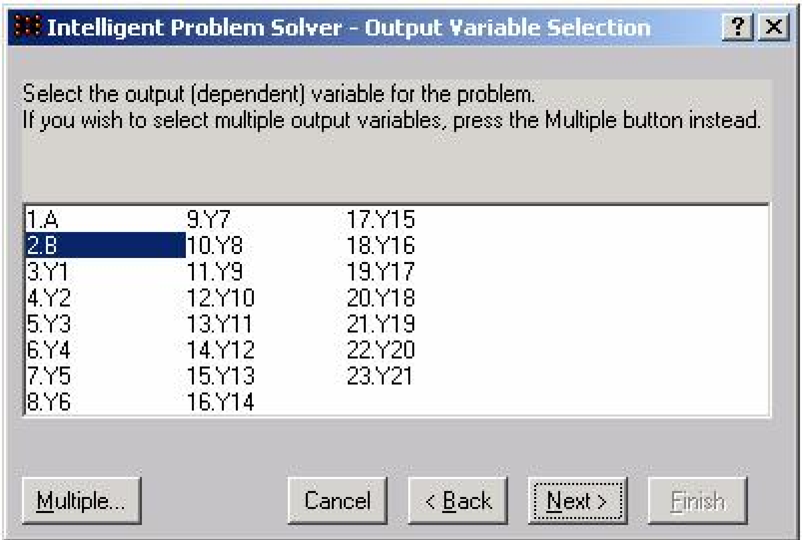
ავირჩიოთ Problem Type Standart…,ვინაიდან ჩვენ არ ვფლობთ მონაცემების დინამიურ რიგს, სადაც ყოველი მომდევნო მნიშვნელობა დამოკიდებულია წინასგან, არამედ გვაქვს უნიკალური დამოუკიდებელი მონაცემები (ნახ. 12.7).

ვინაიდან ჩვენს მომავალ ქსელს ექნება ორი გამოსასვლელი, ამიტომ დავაწვეთ ღილაკს Multiple … (ნახ. 12.8). განვსაზღვროთ დამოკიდებული და დამოუკიდებელი ცვლადები (ნახ. 12.9 და ნახ. 12.10).

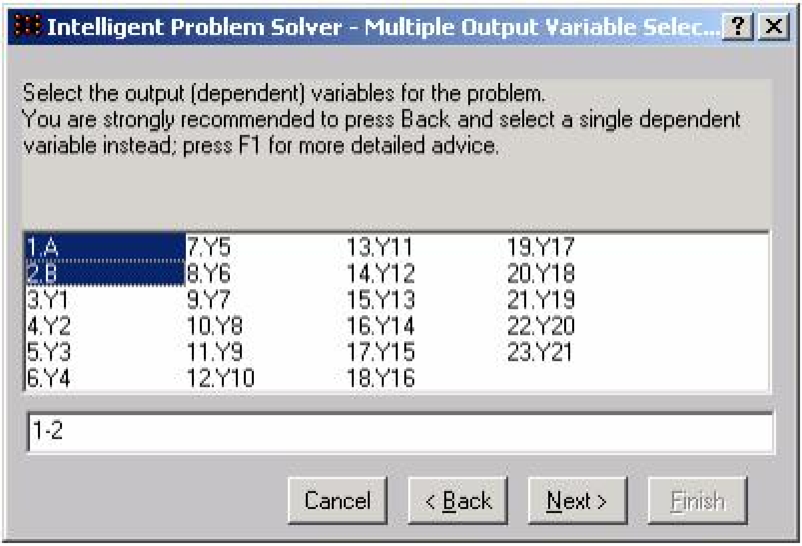
დამოუკიდებელი ცვლადების არჩევის დროს მოვხსნათ ალამი Search for an effective subset of the specified variables-ზე. წინააღმდეგ შემთხვევაში დაფორმირდება ქსელი ერთი შესასვლელით!



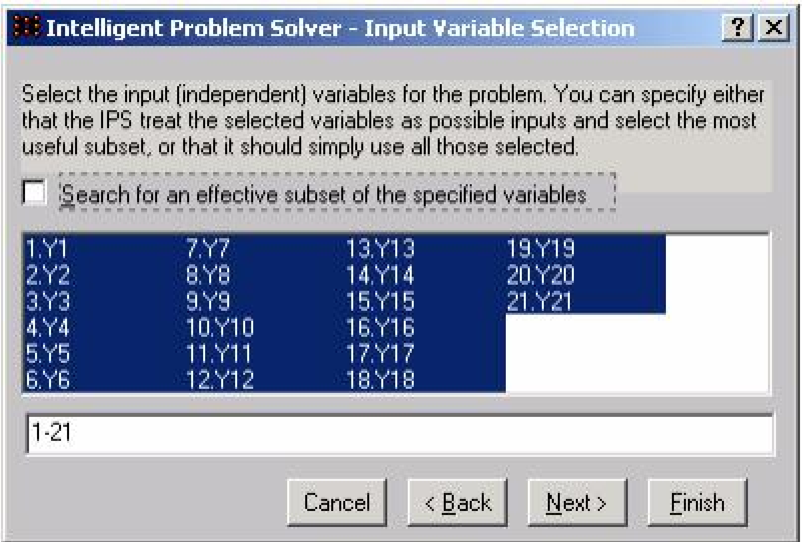
ნახ. 12.7.



ნახ. 12.8.

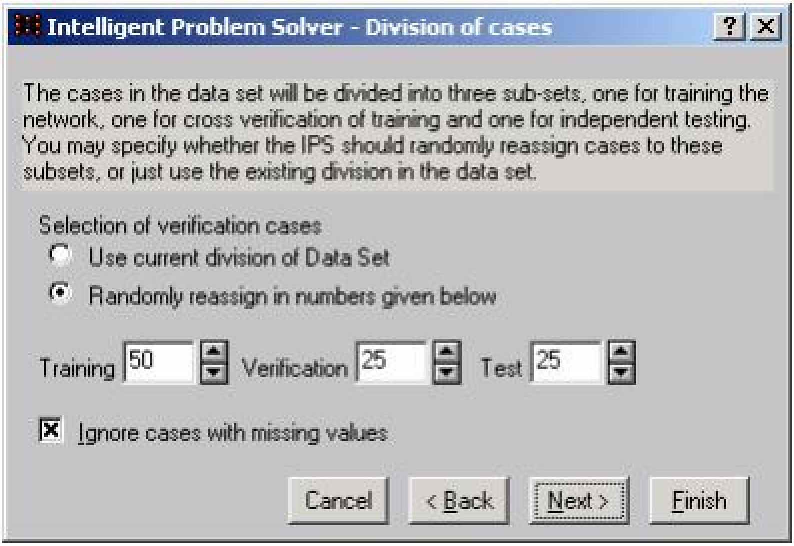


ნახ. 12.9.



ნახ. 12.10.

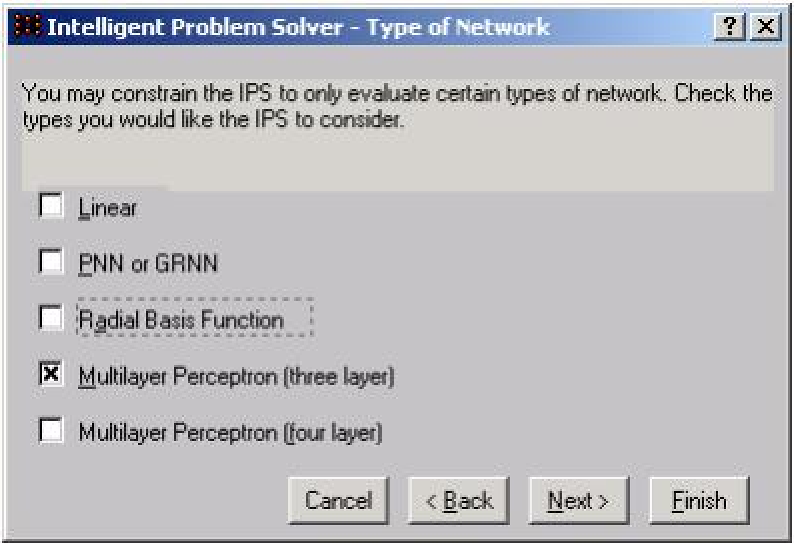
მთელი შესასვლელი მასივი დაიყოფა სამ მონაცემთა ნაკრებად: 50 მნიშვნელობა სწავლებისთვის (Traning), 25 საკონტროლო დაკვირვება შემოწმებისთვის (Verification) და 25 - დამოუკიდებელი ტესტირებისთვის (Test). ოპცია Randomly reassign … საშუალებას იძლევა შევუერთოთ სასწავლო და საკონტროლო დაკვირვებები მონაცემთა ფაილში (ნახ. 12.11). ავირჩიოთ Multilayer Perceptron (tree layer) (ნახ. 12.12). კონკრეტული ამოცანის გადასაწყვეტად ქსელის არქიტექტურის არჩევა დაფუძნებულია დამმუშავებლის გამოცდილებაზე. ამიტომ ქვემოთ შემოთავაზებული ქსელის არქიტექტურა წარმოადგენს ერთ ვარიანტს შესაძლო კონფიგურაციათა სიმრავლიდან.



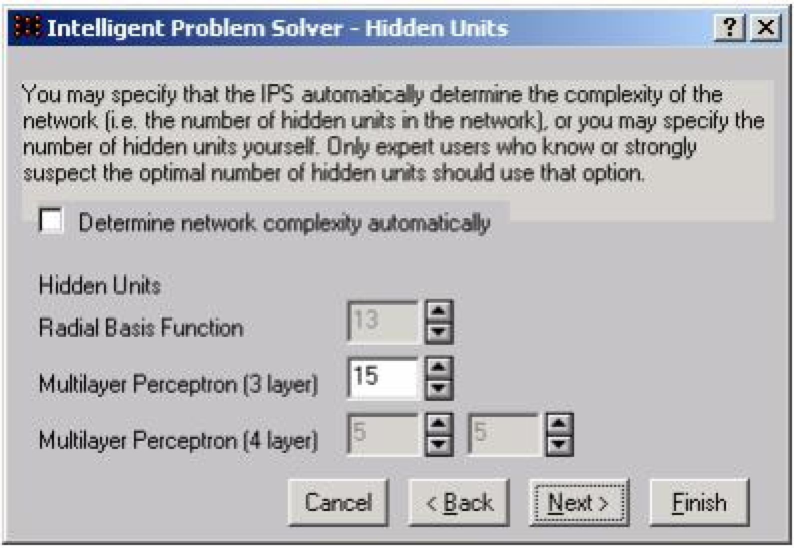
ნახ. 12.11.

გათიშეთ ოპცია Determine network complexity automatically ნახ. 12.13-ზე და მეორე დონის კვანძების რიცხვად მიუთითეთ 15.

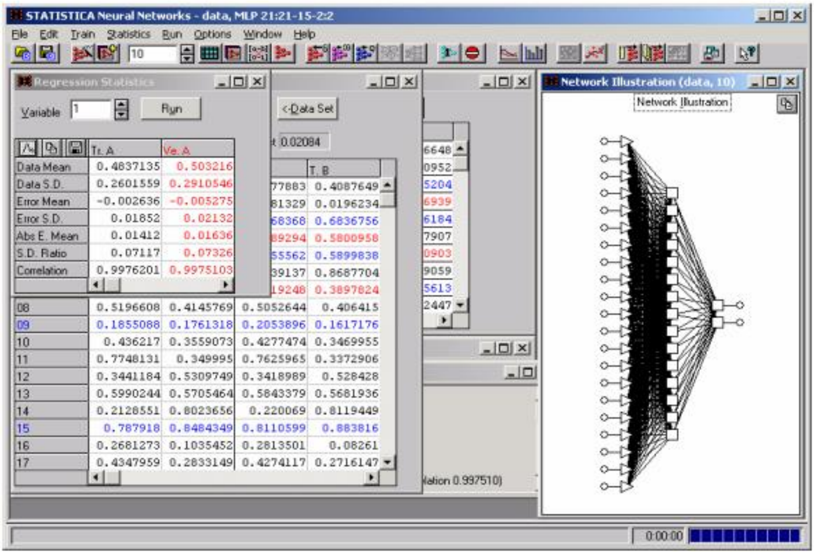
Next ღილაკზე სამჯერ დაწოლით და ოპციების შეუცვლელად, Finish-ზე დაწოლის შემდეგ მივიღებთ შედეგს(ნახ. 12.14).ამ ნახატზე



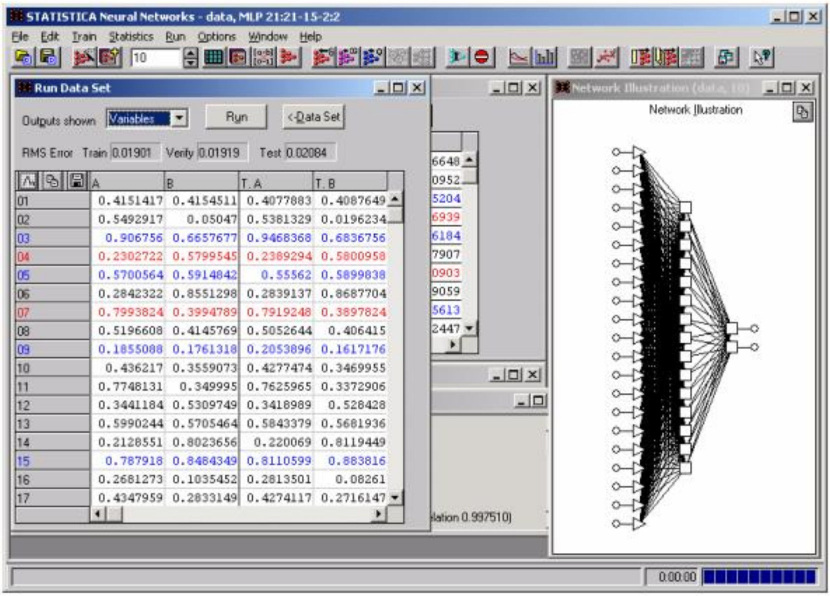
ნახ. 12.12.



ნახ. 12.13.

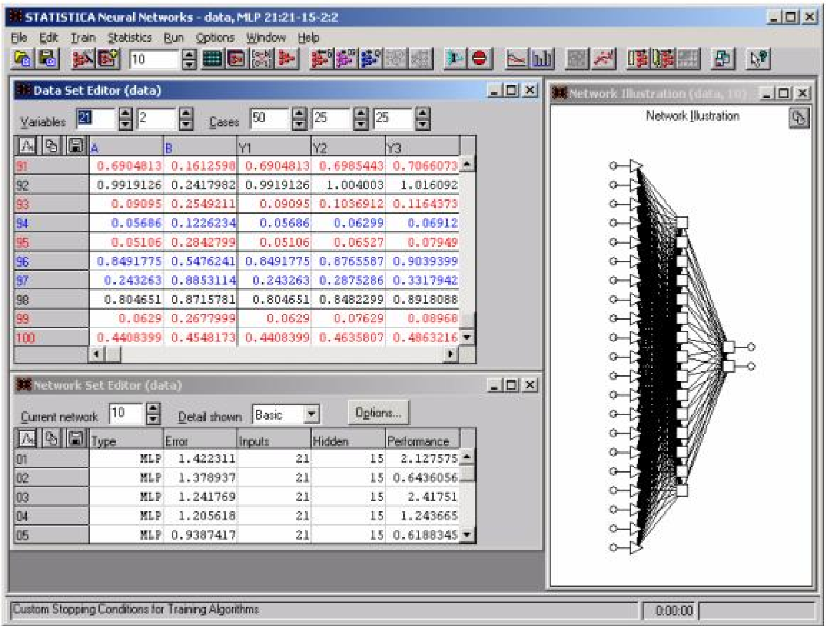


ნახ. 12.14.



ნახ. 12.15.

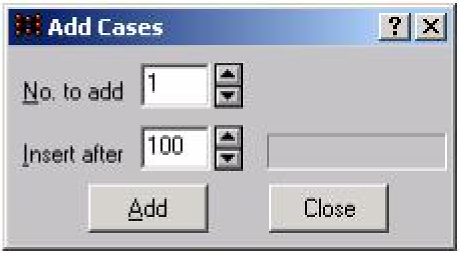
შეგვიძლია თავად ქსელისა და მუშაობის შედეგების დანახვა, ხოლო იმის შესახებ, თუ რამდენად კარგებია ისინი, შეგვიძლია ვიმსჯელოთ ბრძანებით Run/Data Set (ნახ. 12.15).შავი ფერით გამოყოფილია მონაცემთა ის ნაკრები, რომელზეც შესრულდა სწავლება, წითელი ფერით - ვერიფიკაცია, ხოლო ლურჯით - დამოუკიდებელი ტესტირება. A და B სვეტებში წარმოდგენილია *a* და *b* პარამეტრების შეფასებები, ხოლო სვეტებში T.A და T.B ამ პარამეტრების „რეალური (target)" მნისვნელობები. საუკეთესო ნეირონულ ქსელს აქვს *0,9975*-ის ტოლი კორელაციის კოეფიციენტი.



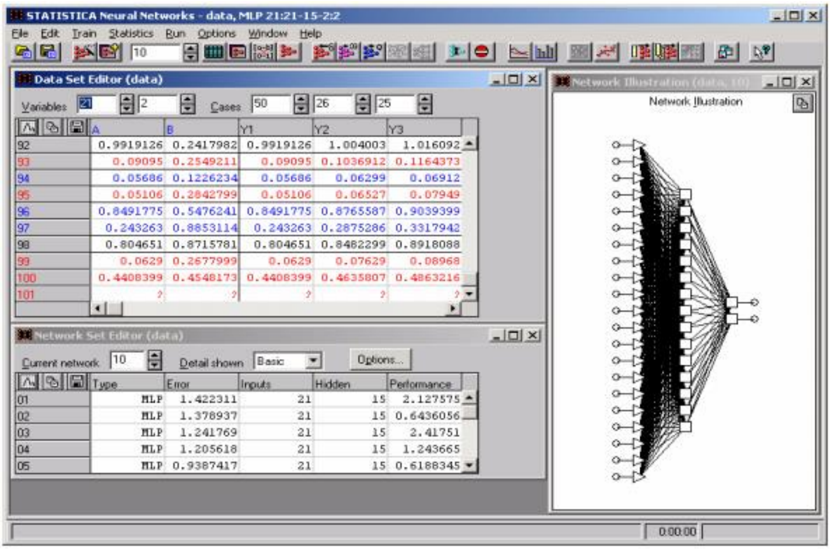
ნახ. 12.16.

შევეცადოთ გამოვიყენოთ მიღებული მოდელი. ამისათვის ჯერ შევინახოთ ჩვენი ქსელი ბრძანებით Save/Network Set სახელით data.bnt. თავიდან გავუშვათ პაკეტი STATISTICA და Neural Networks მოდულში გავხსნათ ფაილი data.sta, შესაბამისი ნეირონული ქსე-

ლით (ნახ. 12.16). მივუთითოთ ბრძანება Edit/Cases/Add (ნახ. 12.17) და დავუმატოთ ერთი სტრიქონი, მივიღებთ შედეგს (ნახ. 12.18).



ნახ. 12.17.



ნახ. 12.18.

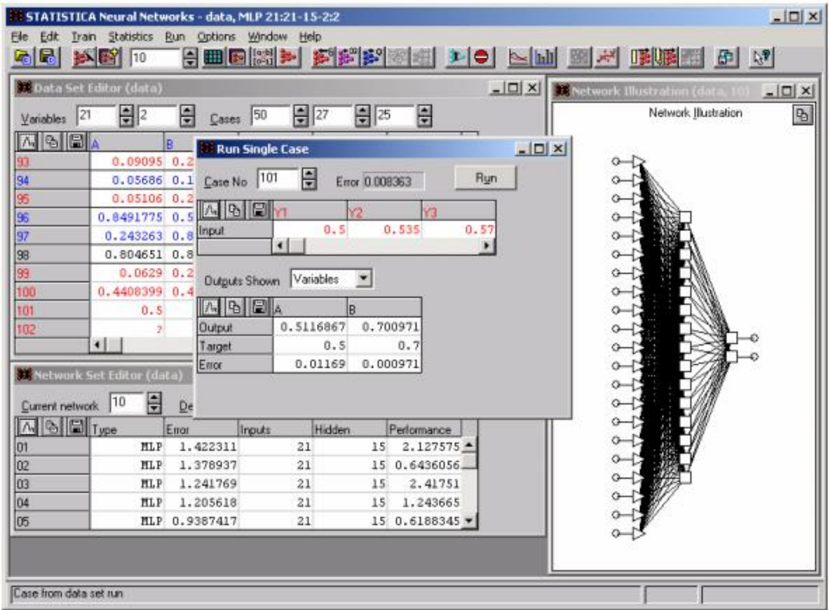
ვთქვათ A=0.5 და B=0.7, მაშინ შესაბამის შესასვლელ ვექტორს ექნება შემდეგი მნიშვნელობები:

0.500 0.535 0.570 0.605 0.640 0.675 0.710 0.745 0.780 0.815 0.850 0.885 0.920 0.955 0.990 1.025 1.060 1.095 1.130 1.165 1.200.

შევიტანოთ ეს მნიშვნელობები ცხრილში. შევასრულოთ ბრძანება Run/Single Case და მივუთითოთ სტრიქონის ნომერი 101; მივი-

ღებთ შემდეგ შედეგს (ნახ. 12.19), რაც საკმაოდ ახლოსაა სწორ შედეგთან:

Output	0.516867	0.700971
Target	0.5	0.7
Error	0.01169	0.000971



ნახ. 12.19.

12.4.2. კლასიფიკაცია

კლასიფიკაციის ამოცანა წარმოადგენს რამდენიმე წყვილ-წყვილად გადაუკვეთავი სიმრავლიდან ერთზე ნიმუშის მიკუთვნების ამოცანას. ასეთი ამოცანების მაგალითი შეიძლება იყოს, მაგალითად, ბანკის კლიენტის კრედიტუნარიანობის განსაზღვრის ამოცანა,

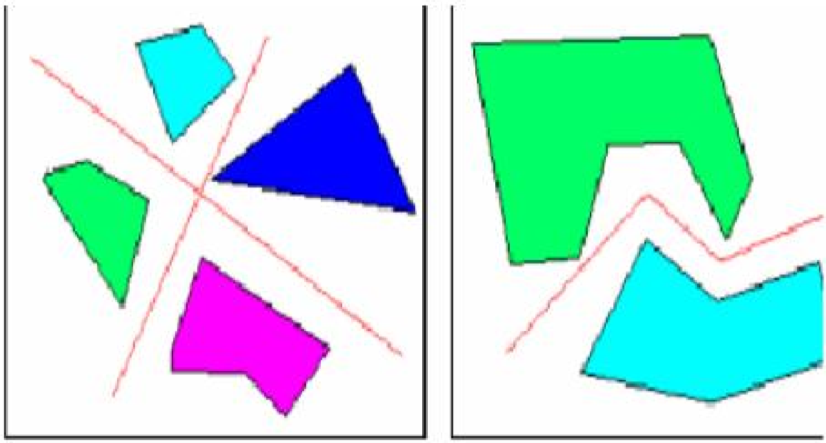
ფასიანი ქაღალდების პორტფელის მართვის ამოცანის გადაწყვეტა (აქციების გაყიდვა, ყიდვა ან შენახვა ბაზარზე სიტუაციიდან დამოკიდებულებით), სიცოცხლისუნარიანი და გაკოტრებისადმი მიდრეკილების მქონე ფირმების განსაზღვრის ამოცანა.

შესაძლებელია შესასვლელი მონაცემების წარმოდგენის რამდენიმე ხერხი. ყველაზე გავრცელებულია ხერხი, რომლის დროსაც ნიმუში ვექტორით წარმოიდგინება. ამ ვექტორის კომპონენტები წარმოადგენენ ნიმუშის სხვადასხვა მახასიათებლებს, რომლებიც ზემოქმედებენ გადაწყვეტილებების მიღებაზე იმის შესახებ, თუ რომელ კლასს შეიძლება მივაკუთვნოთ მოცემული ნიმუში. კლასიფიკატორი აკუთვნებს ობიექტს ერთ-ერთ კლასს N-განზომილებიანი სივრცის განსაზღვრული დაყოფის შესაბამისად, რომელიც იწოდება შესასვლელების სივრცედ, და ამ სივრცის განზომილება წარმოადგენს ვექტორის კომპონენტების რაოდენობას.

უპირველეს ყოვლისა, საჭიროა განვსაზღვროთ სისტემის სირთულის დონე. რეალურ ამოცანებში ხშირად ჩნდებიან სიტუაციები, როცა ნიმუშების რაოდენობა შეზღუდულია, რაც ართულებს ამოცანის სირთულის განსაზღვრას. შესაძლებელია გამოიყოს სირთულის სამი ძირითადი დონე. პირველი (ყველაზე მარტივი), როცა კლასები სწორი ხაზებით შეიძლება დაიყოს (ან ჰიპერსიბრტყეებით, თუ შესასვლელთა სიბრტყეს ორზე მეტი განზომილება გააჩნია) - ე.წ. *წრფივი დაყოფადობა.* მეორე შემთხვევაში შეუძლებელია კლასების დაყოფა ხაზებით (სიბრტყეებით), მაგრამ ისინი შესაძლებელია გამოვყოთ უფრო სრული დაყოფით - *არაწრფივი დაყოფადობა* (ნახ. 8.20). მესამე შემთხვევაში კლასები იკვეთებიან და შეიძლება ვისაუბროთ მხოლოდ *ალბათურ დაყოფადობაზე.*

იდეალურ ვარიანტში წინასწარი დამუშავების შემდეგ ჩვენ უნდა მივიღოთ წრფივად დაყოფილი ამოცანა, ვინაიდან ამის შემდეგ მნიშვნელოვნად მარტივდება კლასიფიკატორის აგება. სამწუხაროდ, რეალური ამოცანების ამოხსნისას გვაქვს ნიმუშების შეზღუ-

დული რაოდენობა, რომელთა საფუძველზეც წარმოებს კლასიფიკა-ტორის აგება. ამასთან ჩვენ არ შეგვიძლია ჩავატაროთ მონაცემების ისეთი წინასწარი დამუშავება, რომლის დროსაც მიიღწევა ნიმუშე-ბის წრფივი დაყოფადობა.



ნახ. 12.20.

შემდეგ აუცილებელია განვსაზღვროთ შესასვლელი მონაცე-მების წარმოდგენის ხერხი ნეირონული ქსელისათვის, ანუ განვსა-ზღვროთ ნორმირების ხერხი. ნორმირება აუცილებელია, ვინაიდან ნეირონული ქსელები მუშაობენ მონაცემებთან, რომლებიც წარმოდ-გენილებია რიცხვებით *0 ... 1* დიაპაზონიდან, ხოლო საწყის მონაცე-მებს შეიძლება ქონდეთ ნებისმიერი დიაპაზონი ან საერთოდ არა რიცხვითი მონაცემები იყვნენ. ამასთან შესაძლებელია სხვადასხვა ხერხი, დაწყებული მოთხოვნილ დიაპაზონში მარტივი წრფივი გარ-დაქმნიდან და პარამეტრების მრავალგანზომილებიანი ანალიზითა და პარამეტრების ერთმანეთზე ზემოქმედების დამოკიდებულებაში არაწრფივი ნორმირებით დასრულებული.

მაგალითად ავიღოთ ანალოგიური მაგალითი კლასტერული ანალიზიდან ინვესტიციურ ფონდებზე მათი მდგომარეობის შეფა-

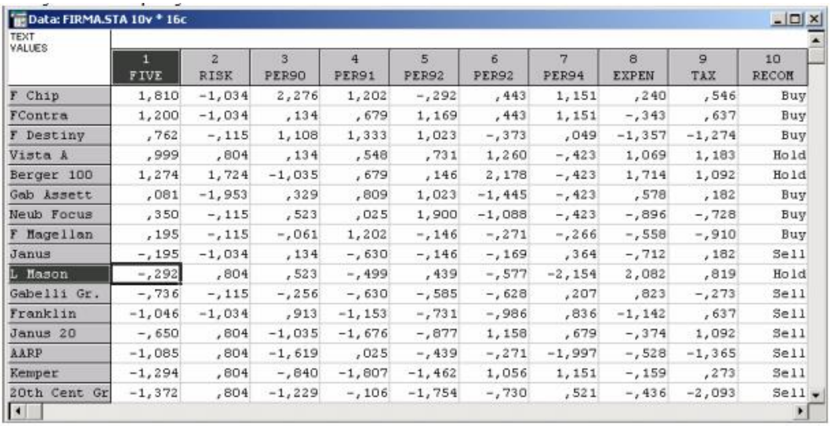
სების მიზნით. მოხერხებულობის მიზნით თავიდან მოვიყვანოთ საწყისი მონაცემები (ცხრ. 12.2).

ცხრილი 12.2.

Fund	Five	Risk	Per90	Per91	Per92	Per93	Per94	Expens	Tax	Recom
Chip	16476	2	10	25	6	55	4	1,22	89	Buy
FContra	15476	2	-1	21	16	55	4	1,03	90	Buy
F Destiny	14757	3	4	26	15	39	-3	0,7	69	Buy
Vista A	15145	4	-1	20	13	71	-6	1,49	96	Hold
Berger 100	15596	5	-7	21	9	89	-6	1,7	95	Hold
Gab Assett	13640	1	0	22	15	18	-6	1,33	85	Buy
Neub Focus	14081	3	1	16	21	25	-6	0,85	75	Buy
F Magellan	13827	3	-2	25	7	41	-5	0,96	73	Buy
Janus	13187	2	-1	11	7	43	-1	0,91	85	Sell
L Mason	13029	4	1	12	11	35	-17	1,82	92	Hold
Gabelli Gr.	12301	3	-3	11	4	34	-2	1,41	80	Sell
Franklin	11793	2	3	7	3	27	2	0,77	90	Sell
Janus 20	12441	4	-7	3	2	69	1	1,02	95	Sell
AARP	11728	4	-10	16	5	41	-16	0,97	68	Sell
Kemper	11386	4	-6	2	-2	67	4	1,09	86	Sell
20^{th} Cent Gr	11258	4	-8	15	-4	32	0	1	60	Sell

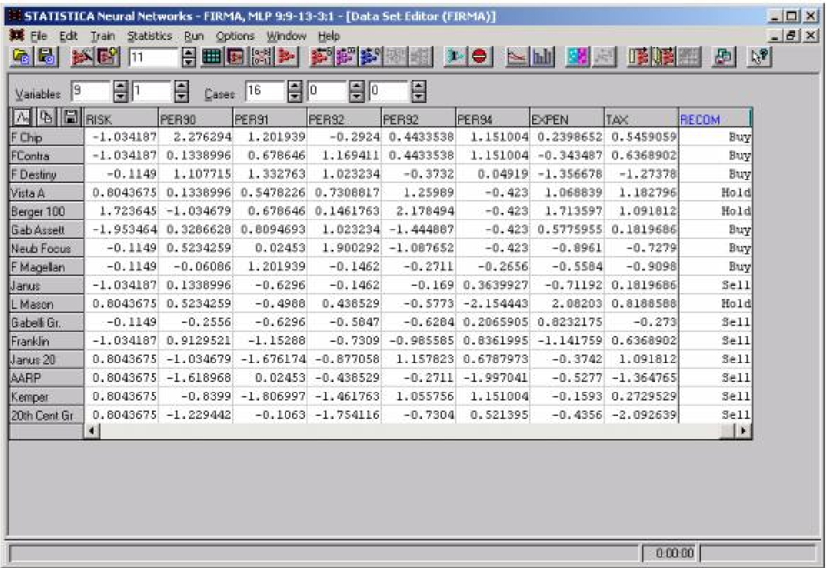
- გაუშვით პროგრამა STATISTICA.
- გამონათდება დიალოგის ფანჯარა Statistica Module Switcher.
- გამოყავით მოდული Data Management/MFM და დააწკაპუნეთ Switct To ღილაკზე.
- გაიღება ფანჯარა Data Management/MFM. ცვლადებისათვის შეიტანეთ საწყისი მონაცემები სვეტებში VAR1 და VAR9 შემდეგი სახით (საჭიროა 6 Cases დამატება) (ცხრ. 12.2).
- შეასრულეთ ბრძანება Analysis/Standartize … საჭიროა აირჩიოთ ყველა ცვლადი (გარდა RECOM-ის) და ყველა სტრიქონი. მივიღებთ საწყისი მონაცემების სტანდარტიზებულ მნიშვნელობებს (ნახ. 12.21) და დავიმახსოვროთ მიღებული შედეგები.

სისტემა STATISTICA-ში ავირჩიოთ მოდული Neural Networks და გავხსნათ ადრე შენახული ფაილი (ნახ. 12.22).



TEXT VALUES	1 FIVE	2 RISK	3 PER90	4 PER91	5 PER92	6 PER92	7 PER94	8 EXPEN	9 TAX	10 RECOM
F Chip	1,810	-1,034	2,276	1,202	-,292	,443	1,151	,240	,546	Buy
FContra	1,200	-1,034	,134	,679	1,169	,443	1,151	-,343	,637	Buy
F Destiny	,762	-,115	1,108	1,333	1,023	-,373	,049	-1,357	-1,274	Buy
Vista A	,999	,804	,134	,548	,731	1,260	-,423	1,069	1,183	Hold
Berger 100	1,274	1,724	-1,035	,679	,146	2,178	-,423	1,714	1,092	Hold
Gab Assett	,081	-1,953	,329	,809	1,023	-1,445	-,423	,578	,182	Buy
Neub Focus	,350	-,115	,523	,025	1,900	-1,088	-,423	-,896	-,728	Buy
F Magellan	,195	-,115	-,061	1,202	-,146	-,271	-,266	-,558	-,910	Buy
Janus	-,195	-1,034	,134	-,630	-,146	-,169	,364	-,712	,182	Sell
L Mason	-,292	,804	,523	-,499	,439	-,577	-2,154	2,082	,819	Hold
Gabelli Gr.	-,736	-,115	-,256	-,630	-,585	-,628	,207	,823	-,273	Sell
Franklin	-1,046	-1,034	,913	-1,153	-,731	-,986	,836	-1,142	,637	Sell
Janus 20	-,650	,804	-1,035	-1,676	-,877	1,158	,679	-,374	1,092	Sell
AARP	-1,085	,804	-1,619	,025	-,439	-,271	-1,997	-,528	-1,365	Sell
Kemper	-1,294	,804	-,840	-1,807	-1,462	1,056	1,151	-,159	,273	Sell
20th Cent Gr	-1,372	,804	-1,229	-,106	-1,754	-,730	,521	-,436	-2,093	Sell

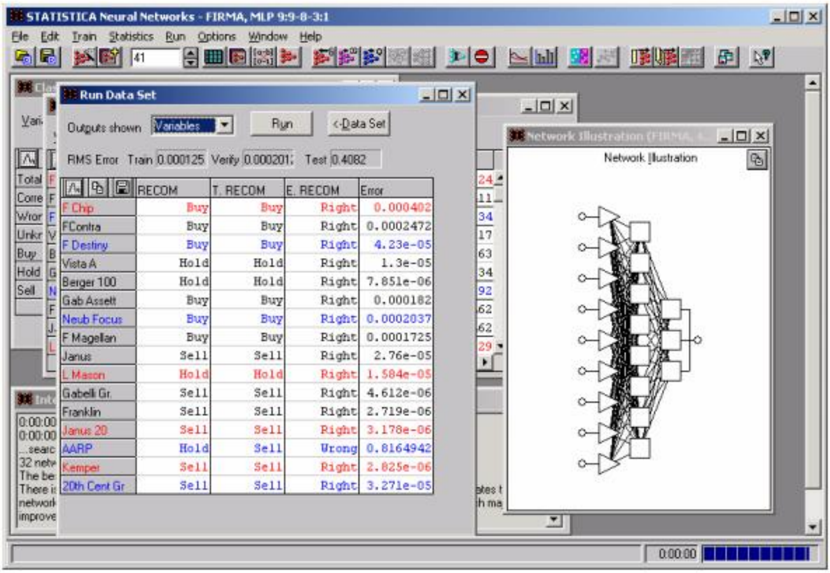
ნახ. 12.21.



	RISK	PER90	PER91	PER92	PER92	PER94	EXPEN	TAX	RECOM
F Chip	-1.034187	2.276294	1.201939	-0.2924	0.4433538	1.151004	0.2398652	0.5459059	Buy
FContra	-1.034187	0.1338996	0.678646	1.169411	0.4433538	1.151004	-0.343487	0.6368902	Buy
F Destiny	-0.1149	1.107715	1.332763	1.023234	-0.3732	0.04919	-1.356678	-1.27378	Buy
Vista A	0.8043675	0.1338996	0.5478226	0.7308817	1.25989	-0.423	1.068839	1.182796	Hold
Berger 100	1.723645	-1.034679	0.678646	0.1461763	2.178494	-0.423	1.713597	1.091812	Hold
Gab Assett	-1.953464	0.3286628	0.8094693	1.023234	-1.444887	-0.423	0.5775955	0.1819686	Buy
Neub Focus	-0.1149	0.5234259	0.02453	1.900292	-1.087652	-0.423	-0.8961	-0.7279	Buy
F Magellan	-0.1149	-0.06086	1.201939	-0.1462	-0.2711	-0.2656	-0.5584	-0.9098	Buy
Janus	-1.034187	0.1338996	-0.6296	-0.1462	-0.169	0.3639927	-0.71192	0.1819686	Sell
L Mason	0.8043675	0.5234259	-0.4988	0.438529	-0.5773	-2.154443	2.08203	0.8188588	Hold
Gabelli Gr.	-0.1149	-0.2556	-0.6296	-0.5847	-0.6284	0.2065905	0.8232175	-0.273	Sell
Franklin	-1.034187	0.9129521	-1.15288	-0.7309	-0.985585	0.8361995	-1.141759	0.6368902	Sell
Janus 20	0.8043675	-1.034679	-1.676174	-0.877058	1.157823	0.6787973	-0.3742	1.091812	Sell
AARP	0.8043675	-1.618968	0.02453	-0.438529	-0.2711	-1.997041	-0.5277	-1.364765	Sell
Kemper	0.8043675	-0.8399	-1.806997	-1.461763	1.055756	1.151004	-0.1593	0.2729529	Sell
20th Cent Gr	0.8043675	-1.229442	-0.1063	-1.754116	-0.7304	0.521395	-0.4356	-2.092639	Sell

ნახ. 12.22.

შევეცადოთ სრულად გამოვიყენოთ Inelligent Problem Solver შესაძლებლობები, დაყენებებში ცვლილებების შეტანის გარეშე. მივიღებთ შემდეგ შედეგს (ნახ. 12.23).



ნახ. 12.23.

პროგრამამ საკმაოდ კარგად ისწავლა წარმოდგენილ შერჩევაზე, მხოლოდ ერთ შემთხვევაში ის რეკომენდაციას გიწევთ „შეინახოთ - Hold” AAPP ფონდის აქციები, ნაცვლად იმისა, რომ „გავთავისუფლდეთ - Shel” მათგან.

მიღებული ქსელი მომავალში შეიძლება წარმატებით იქნას გამოყენებული ახალი ობიექტების (ფონდების) კლასიფიკაციისთვის. დამატებითი შეტანილი ფონდებისთვის შეიძლება მივიღოთ შემდეგი შედეგები (ცხრ. 12.3).

ცხრილი 12.3.

Fund	Five	Risk	Per90	Per91	Per92	Per93	Per94	Expens	Tax	Recom
F.OTC	13129	4	-5	49	15	8	-3	0,88	75	Buy
ColumbiaGr.	13399	3	-3	34	12	13	-1	0,83	71	Buy
TRP Capital	13449	1	-1	22	9	16	4	1,1	76	Buy
NEUB	13336	2	-5	22	18	16	-2	0,81	70	Buy

12.4.3. პროგნოზირება

პროგნოზირება ანალიზის ამოცანებიდან ერთ-ერთი ყველაზე მოთხოვნადი, მაგრამ ამავე დროს ყველაზე რთულია. მისი გადაწყვეტისას წარმოქმნილი პრობლემები ბევრი მიზეზით არიან განპირობებულები: საწყისი მონაცემების არასაკმარისი ხარისხი და რაოდენობა, იმ გარემოს ცვლილება, რომელშიც მიმდინარეობს პროცესი, სუბიექტური ფაქტორების ზემოქმედება. მაგრამ სწორედ ხარისხიანი პროგნოზი წარმოადგენს გასაღებს ისეთი ბიზნეს ამოცანების გადასაჭრელად როგორიცაა ფინანსური ნაკადების ოპტიმიზაცია, ინვესტიციური მიმზიდველობის შეფასება და ბევრი სხვა.

ზოგად შემთხვევაში პროგნოზირების ამოცანა დადის დროში მოწესრიგებული მომავალი მონაცემების შეფასებათა მიღებაში უკვე არსებული მონაცემების ანალიზის საფუძველზე. არსებობენ არსებულ მონაცემებში კანონზომიერებათა ძიების სხვადასხვა ალგორითმები. სტანდარტულ მეთოდებთან ერთად, რომლებიც პარამეტრულ მოდელებს იყენებენ, ბოლო დროს ამ მიზნებისათვის გამოყენება ჰპოვეს სხვა მიდგომებმაც, კერძოდ, ნეიროქსელურმა მეთოდებმა.

ნებისმიერ შემთხვევაში საჭიროა ორი საკითხის გადაჭრა: რა არის საპროგნოზირებელი სიდიდე და რას წარმოადგენენ შესასვლელი მონაცემები. უმეტეს შემთხვევაში საპროგნოზირებელ სიდიდეს წარმოადგენენ დროითი მწკრივის მნიშვნელობები [T(n+1), T(n+f)] ინტერვალზე, სადაც T(n)-დროის მიმდინარე მომენტია, ხოლო f - პროგნოზირების ინტერვალი. ზოგჯერ ჩნდება აუცილებლობა ვიწინასწარმეტყველოთ დროითი მწკრივის არა მნიშვნელობები მოცემულ ინტერვალზე, არამედ იმის ალბათობა, თუ როგორ მოიქცევა ის (გაიზრდება, შემცირდება, იქნება რაღაც საზღვრებში და ა. შ.).

რაც შეეხება საწყის მონაცემებს, პირველი რაც აუცილებელია ესაა არსებული მონაცემებიდან მნიშვნელოვანი ფაქტორების მაქსი-

მალური რიცხვის არჩევა. ეს ნიშნავს დაკვირვების ინტერვალის არჩევას, ანუ დროითი მწკრივის წინა მნიშვნელობების რა რაოდენობით ხორციელდება პროგნოზი და დამატებითი ფაქტორების განსაზღვრა, რომლებიც ზემოქმედებენ საპროგნოზირებო სიდიდის ქცევაზე. უკანასკნელებს ზოგჯერ უწოდებენ ეკზოგენურ (გარე) ფაქტორებს.

შემდეგ, შესასვლელი მონაცემებიდან საჭიროა არა მნიშვნელოვანი და იშვიათი ფაქტორების გამორიცხვა. საკმაოდ ხშირად ჩნდება მონაცემების წინასწარი დამუშავების ჩატარების აუცილებლობა: გამოტოვებული მონაცემების აღდგენა, ანომალური გამოვარდნების მოცილება, მაღალსიხშირული ხმაურის მოცილება. ამ პროცესს მონაცემების წინასწარი დამუშავება ჰქვია. ხარისხიანი წინასწარი დამუშავება საშუალებას იძლევა მნიშვნელოვნად გაუმჯობესდეს პროგნოზის ხარისხიც.

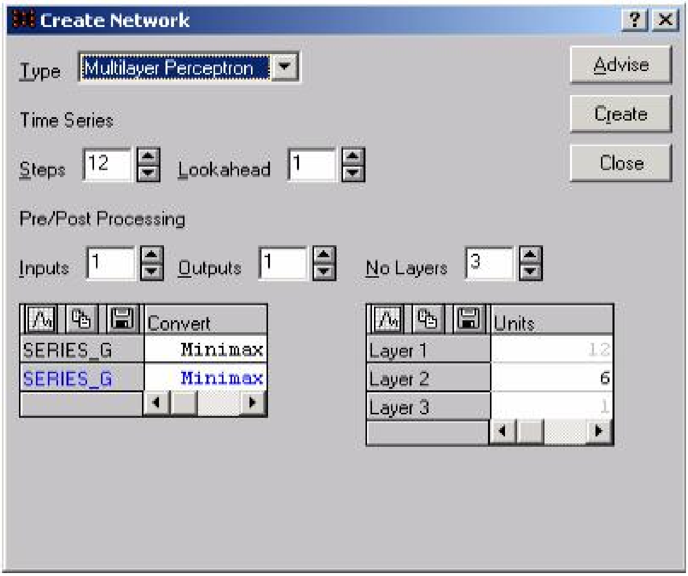
შემდეგი ნაბიჯია ნეიროქსელური მოდულის აგება - ნეირონული ქსელის სწავლება. ამ ბიჯზე საჭიროა რიგი სპეციფიკური ქვეამოცანების ამოხსნა: ნეირონული ქსელის სტრუქტურის, სწავლების ალგორითმის და სხვა არჩევა. აგებული ნეიროქსელური მოდელის საფუძველზე ხორციელდება პროგნოზირება.

მაგალითის სახით განვიხილოთ მონაცემთა ტესტური ფაილი Series_g.sta პაკეტ STATISTICA-ს Exemple საქაღალდიდან. მონაცემები შეიცავენ ერთი ცვლადის მნიშვნელობებს - 1949-დან 1960 წლის ჩათვლით პერიოდში გაერთიანებული სამეფოს ავიო კომპანიის მიერ გადაყვანილი მგზავრების შესახებ ყოველთვიური მონაცემები. მივუთითოთ ცვლადის ტიპად *შესასვლელ/გამოსასვლელი (input /output).* ამისათვის მონაცემების გახსნილ ფაილში გამოყავით ცვლადი და დააწკაპუნეთ თაგვის მარჯვენა ღილაკი, გამონათებული მენიუდან აირჩიეთ პუნქტი *შესასვლელ/გამოსასვლელი (input /output).* ცვლადის სახელი გამონათდება მწვანე ფერით. ამ ფაილში მონაცემების რაოდენობა 144-ია (ნახ. 12.24).

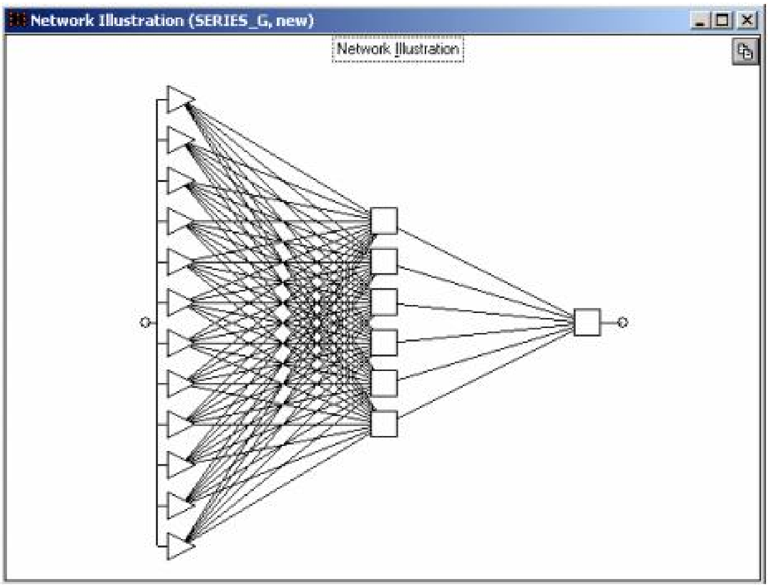
ნახ. 12.24.

ბრძანებით File/New/Network შევქმნათ ახალი ქსელი, გამონათდება ნახ. 12.25-ზე გამოსახული ფანჯარა.

ველში Type ავირჩიოთ ქსელის ტიპი: Multilayer perceptron (მრავალშრიანი პერსეპტრონი) და დავაყენოთ შემდეგი პარამეტრები: შრეების რაოდენობა 3-ის ტოლი, პარამეტრი Steps=12 (მონაცემები წარმოადგენენ ყოველთვიურ გადაზიდვებს მათში არსებული სეზონური შემადგენლით), ხოლო პარამეტრი Lookahead (ჰორიზონტი) 1-ის ტოლი. ეკრანზე გამონათდება სამშრიანი პერსეპტრონი (ნახ. 12.26) 12 შესასვლელითა და ერთი გამოსასვლელით.

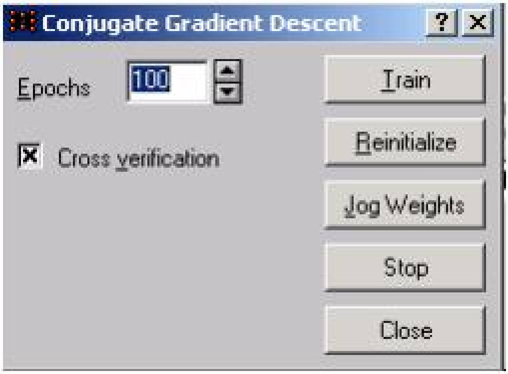


ნახ. 12.25.



ნახ. 12.26.

ახლა საჭიროა ქსელის სწავლება. მიუთითეთ 66 *სასწავლო (Traning)* და 56 *საკონტროლო (Verification)* დატვირთვა (ნახ. 12.24). ქსელის სწავლებისათვის ვისარგებლოთ შეუღლებულ გრადიენტთა მეთოდით ბრძანებით Train/Multilayer perceptrons/conjugate Gradients … , გამონათდება შეუღლებული გრადიენტების მეთოდით მინიმიზირების ფანჯარა (ნახ. 12.27).

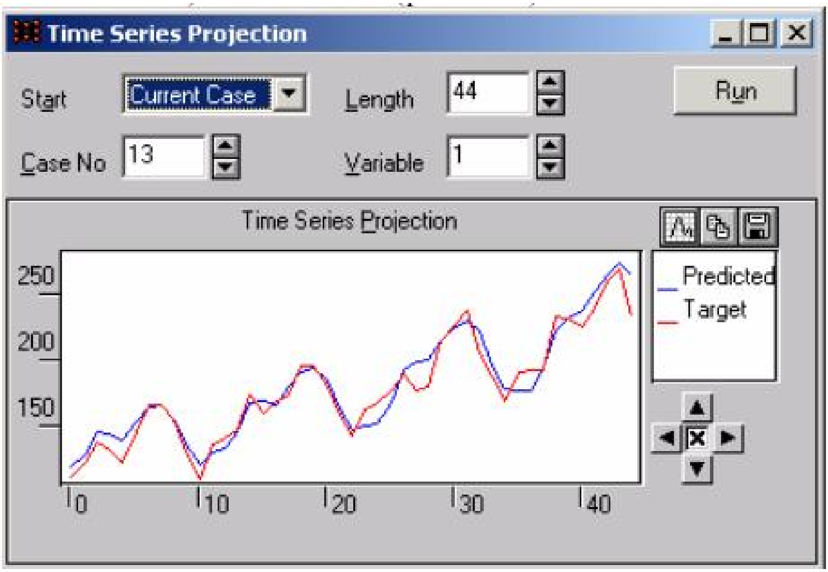


ნახ. 12.27.

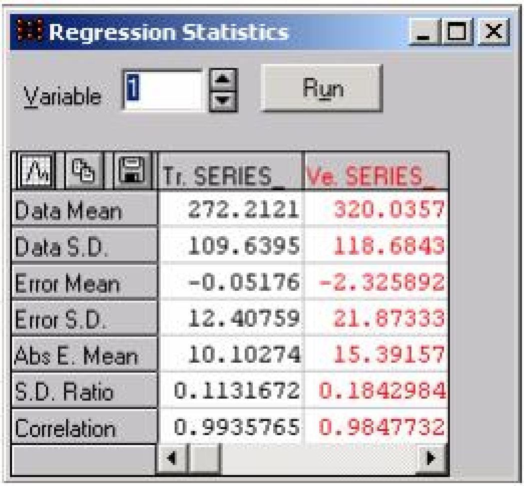
პროგნოზის ასაგებად მენიუ Run-ში (ნახ. 12.28) გავხსნათ ფანჯარა Times Series Proection (დროითი მწკრივის პროექცია).

ქსელის მუშაობის ხარისხის შესაფასებლად ვისარგებლოთ მენიუ Statistics-ის ფანჯრით Regression Statisties (რეგრესიის სტატისტიკა) (ნახ. 12.29).

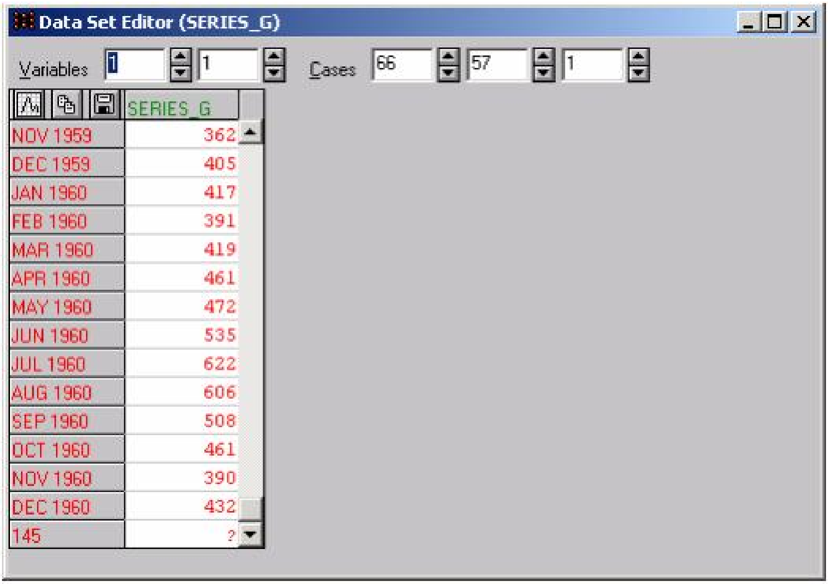
ავაგოთ პროგნოზი ერთი ბიჯით წინ განსწავლული ქსელის დახმარებით, ამისათვის ბრძანებით Edit/Cases/Add დავუმატოთ ერთი ცარიელი ჩანაწერი (№=145) „მომავალი" საპროგნოზირებო მნიშვნელობებისათვის (ნახ. 12.30) და შევასრულოთ ბრძანება Run /Single Cases, გამონათდება ნახ. 12.31-ზე გამოსახული ფანჯარა. ეკრანზე ველში შეიტანეთ Case No (დაკვირვების ნომერი), რომლისთვისაც საჭიროა პროგნოზის შესრულება და დააწექით ღილაკს Run. სტრიქონში Output გამონათდება 435,2634-ის ტოლი მნიშვნელობა.



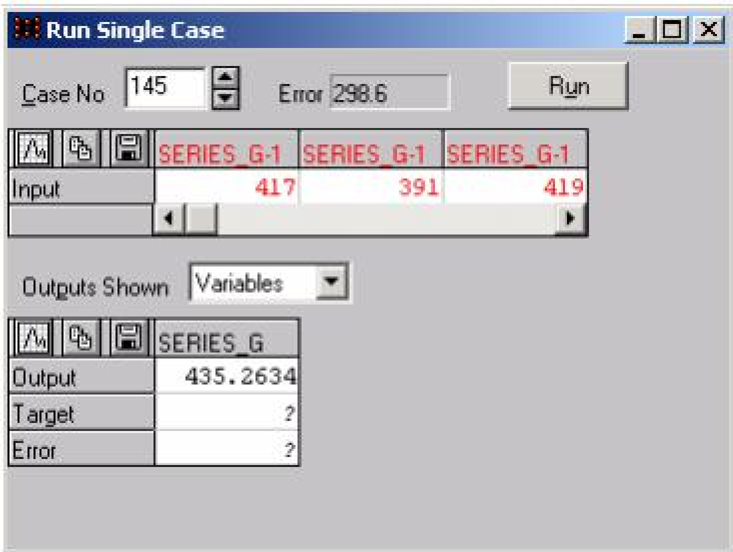
ნახ. 12.28.



ნახ. 12.29.



ნახ. 12.30.



ნახ. 12.31.

დ ა მ ა ტ ე ბ ე ბ ი

დამატება 1

ფიშერის განაწილება.

კვანტილების მნიშვნელობები f_1 და f_2 თავისუფლების ხარისხებისა და λ=0,05 ალბათობისთვის

$f_2 \backslash f_1$	1	2	3	4	5	6	8	12	24
1	161.45	199.50	215.72	224.57	230.17	233.97	238.89	243.91	249.04
2	18.512	18.999	19.163	19.248	19.298	19.329	19.371	19.414	19.453
3	10.129	9.552	9.276	9.118	9.014	8.941	8.844	8.744	8.638
4	7.710	6.945	6.591	6.388	6.257	6.164	6.041	5.912	5.774
5	6.607	5.786	5.410	5.192	5.050	4.950	4.818	4.678	4.527
6	5.987	5.143	4.756	4.388	4.284	4.147	4.000	3.841	3.669
7	5.591	4.737	4.347	4.121	3.972	3.866	3.725	3.574	3.410
8	5.317	4.459	4.067	3.838	3.688	3.580	3.438	3.284	3.116
9	5.117	4.256	3.863	3.633	3.482	3.374	3.230	3.073	2.900
10	4.965	4.103	3.708	3.478	3.326	3.217	3.072	2.913	2.737
11	4.844	3.982	3.587	3.357	3.204	3.094	2.948	2.778	2.609
12	4.747	3.885	3.490	3.259	3.106	2.999	2.848	2.686	2.505
13	4.667	3.805	3.410	3.179	3.025	2.915	2.767	2.604	2.420
14	4.600	3.739	3.344	3.112	2.958	2.848	2.699	2.534	2.349
15	4.543	3.683	3.287	3.056	2.901	2.790	2.641	2.475	2.288
16	4.494	3.634	3.239	3.007	2.853	2.741	2.591	2.424	2.235
17	4.451	3.592	3.197	2.965	2.810	2.699	2.548	2.381	2.190
18	4.414	3.555	3.160	2.928	2.773	2.661	2.510	2.342	2.150
19	4.381	3.522	3.127	2.895	2.740	2.629	2.477	2.308	2.114
20	4.351	3.493	3.098	2.866	2.711	2.599	2.447	2.278	2.083
21	4.325	3.467	3.072	2.840	2.685	2.573	2.421	2.250	2.054
22	4.301	3.443	3.049	2.817	2.661	2.549	2.397	2.226	2.028
23	4.279	3.422	3.028	2.795	2.640	2.528	2.375	2.203	2.005
24	4.260	3.403	3.009	2.777	2.621	2.508	2.355	2.183	1.984
25	4.242	3.385	2.991	2.759	2.603	2.490	2.337	2.165	1.965
26	4.225	3.369	2.975	2.743	2.587	2.474	2.321	2.148	1.947
27	4.210	3.354	2.961	2.728	2.572	2.459	2.305	2.132	1.930
28	4.196	3.340	2.947	2.714	2.558	2.445	2.292	2.118	1.915
29	4.183	3.328	2.934	2.702	2.545	2.432	2.278	2.104	1.901
30	4.171	3.316	2.922	2.690	2.534	2.421	2.266	2.092	1.887
60	4.001	3.151	2.758	2.525	2.368	2.254	2.097	1,918	1.700
120	3.920	3.072	2.680	2.447	2.290	2.175	2.016	1.834	1.608

დამატება 2

სტიუდენტის განაწილება.

კვანტილების მნიშვნელობები f თავისუფლების ხარისხებისა და λ=0,05 ალბათობისთვის

F	1	2	3	4	5	6	7	8	9
T	12.71	4.303	3.182	2.776	2.571	2.447	2.365	2.306	2.262
F	10	11	12	13	14	15	16	17	18
T	2.228	2.201	2.179	2.160	2.145	2.131	2.120	2.110	2.101
F	19	20	21	22	23	24	25	26	27
T	2.093	2.086	2.080	2.074	2.069	2.064	2.060	2.056	2.052
F	28	29	30	40	50	60	80	100	200
T	2.048	2.045	2.042	2.021	2.009	2.000	1.990	1.984	1.972

დამატება 3

დარბინ-უოტსონის კრიტერიუმი.

კრიტერიუმის ზედა და ქვედა საზღვრები λ=0,05 ალბათობისთვის

F_1	f_2=1		f_2=2		f_2=3		f_2=4		f_2=5		f_2=6		f_2=7	
	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$
6	0.61	1.40												
7	0.70	1.36	0.47	1.90										
8	0.76	1.33	0.56	1.78	0.37	2.29								
9	0.82	1.32	0.63	1.70	0.46	2.13	0.30	2.59						
10	0.88	1.32	0.70	1.64	1.53	2.02	0.38	2.41	0.24	2.82				
11	0.93	1.32	0.76	1.60	0.60	1.93	0.44	2.28	0.32	2.65	0.20	3.01		
12	0.97	1.33	0.81	1.58	0.66	1.86	0.51	2.18	0.38	2.51	0.27	2.83	0.17	3.15
13	1.01	1.34	0.86	1.56	0.72	1.82	0.57	2.09	0.45	2.39	0.33	2.69	0.23	2.99
14	1.05	1.35	0.91	1.55	0.77	1.78	0.63	2.03	0.51	2.30	0.39	2.57	0.29	2.85
15	1.08	1.36	0.95	1.54	0.81	1.75	0.69	1.98	0.56	2.22	0.45	2.47	0.34	2.73
16	1.11	1.37	0.98	1.54	0.86	1.73	0.73	1.94	0.62	2.16	0.50	2.39	0.40	2.62
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.66	2.10	0.55	2.32	0.45	2.54
18	1.16	1.39	1.05	1.54	0.93	1.70	0.82	1.87	0.71	2.06	0.60	2.26	0.50	2.46
19	1.18	1.40	1.07	1.54	0.97	1.69	0.86	1.85	0.75	2.02	0.65	2.21	0.55	2.40
20	1.20	1.41	1.10	1.54	1.00	1.68	0.89	1.83	0.79	1.99	0.69	2.16	0.60	2.34
25	1.29	1.45	1.21	1.55	1.12	1.65	1.04	1.77	0.95	1.89	0.87	2.01	0.78	2.14
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83	1.00	1.93	0.93	2.03
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79	1.18	1.85	1.12	1.92
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77	1.29	1.82	1.25	1.88
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77	1.37	1.81	1.34	1.85
70	1.58	1.64	1.55	1.67	1.53	1.70	1.49	1.74	1.46	1.77	1.43	1.80	1.40	1.84
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77	1.48	1.80	1.45	1.83
90	1.64	1.68	1.61	1.69	1.59	1.73	1.57	1.75	1.54	1.78	1.52	1.80	1.49	1.83
100	1.65	1.69	1.63	1.70	1.61	1.74	1.59	1.76	1.57	1.78	1.55	1.80	1.53	1.83
150	1.72	1.75	1.71	1.76	1.69	1.77	1.68	1.79	1.67	1.80	1.65	1.82	1.64	1.83
200	1.76	1.78	1.75	1.79	1.74	1.80	1.73	1.81	1.72	1.82	1.71	1.83	1.70	1.84

ლიტერატურა

1. Боровиков В.П. STATISTICA: искуство анализа данных на компьютере. Для профессионалов. 2-е изд. - СПб.: Питер, 2003.
2. Куприенко Н.В., Пономарева О.А., Тихонов Д.В. Статистика. Методы анализа распределений. Выборочное наблюдение. 3-е изд. – СПб.: Изд-во Политехн. ун-та, 2009.
3. Лабоцкий В.В. Управление знаниями (технологии, методи и средства представления, извлечения и измерения знаний). – Мн.: Соврем. шк., 2006.
4. Боровиков В.П. Популярное введение в программу STATISTICA.- М.: "Компьютер Пресс", 1998.
5. Электронный учебник по STATISTICA.-М.:StatSoft, 2001. (http://www.statistica.ru).

ს ა რ ჩ ე ვ ი

STATISTICA

პროგრამული პაკეტის გამოყენების საფუძვლები

ავტორი:

თენგიზ ხინიკაძე - ტექნიკურ მეცნიერებათა კანდიდატი, შოთა რუსთაველის სახელმწიფო უნივერსიტეტის განათლებისა და მეცნიერებათა ფაკულტეტის კომპიუტერულ მეცნიერებათა დეპარტამენტის ასოცირებული პროფესორი

რედაქტორები:

მიხეილ დონაძე, გრიგოლ კახიანი