

საქართველოს ტექნიკური უნივერსიტეტი

ელგუჯა მუხანიშვილი

მონაცემების სტატისტიკური  
დამუშავება  
**STATISTICA** სისტემის ბაზაზე

პრაქტიკული

თბილისი 2015

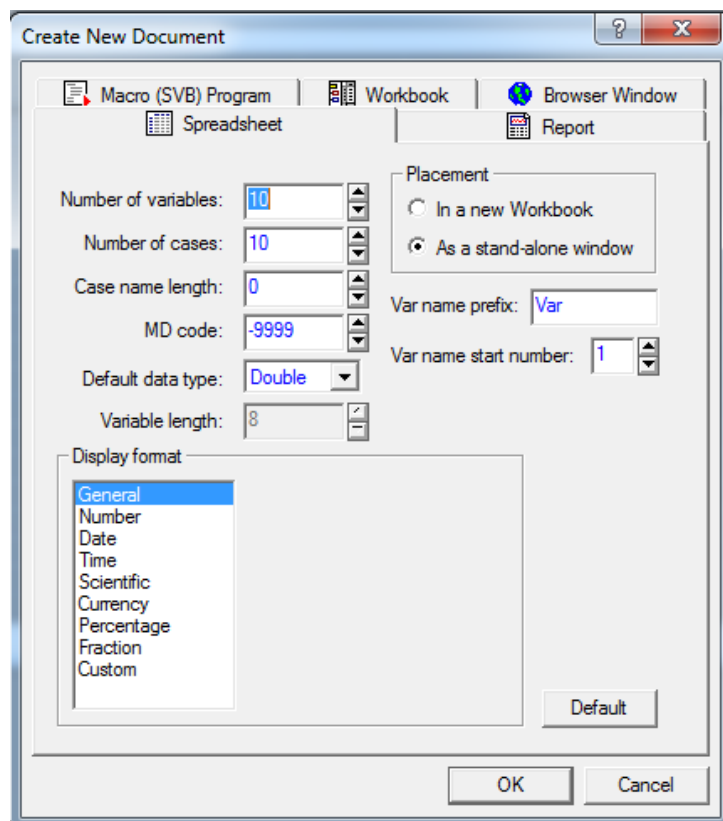
## პრაქტიკული სამუშაო 1

### **STATISTICA 7 სისტემაში მონაცემებთან მუშაობის მოკლე მიმოხილვა**

მონაცემები **Statistica7**-ში ორგანიზებულია ელექტრონული ცხრილების სახით. ცხრილები ინახებიან \*.sta გაფართოების ფაილების სახით. ცხრილის სვეტები წარმოადგენენ ცვლადებს (**Variables**), ხოლო სტრიქონები – დაკვირვებებს, შემთხვევებს (**Cases**). ყოველ ცვლადს გაჩნია სახელი, ფორმატი და სხვა ატრიბუტები. გაჩუმების პრინციპით ცვლადებს გაჩნიათ სახელები *Var1, Var2,...*

#### 1. ახალი მონაცემების ფაილის შექმნა.

გაგხსნათ მენიუ **File + New** ან ინსტრუმენტების პანელზე **click** (მარცხენა ღილაკზე დაწკაპუნება) **New (Ctrl+N)** ღილაკზე.



ეკრანზე გამოდის **Create New Document** ფანჯარა, სადაც **Number of variables:** ველში მიუთითოდ ცვლადების რაოდენობა, ხოლო **Number of cases** ველში – სტრიქონების რაოდენობა. **Display format** ველში შევარჩიოთ (მოვნიშნოთ) **General**. და შემდეგ **Ok**.

	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

ეკრანზე გამოდის ცარიელი ფანჯარა *Data Spreadsheet1* (ცხრილი) ფრჩხილებში ცვლადების და დაკვირვების რაოდენობა დასახელებითი. ფაილის სახელის დასარქმევად გავხსნათ მენიუ *File* და მოვნიშოთ *Save* ბრძანება ან ინსტრუმენტების პანელზე *click Save* (ან *Enter*) ღილაკზე, ეკრანზე გამოსულ *Save as* ფანჯრის *File name* ველში ჩავწერთ ახალი სახელი და შემდეგ *click Save* (ან *Enter*) ღილაკზე.

ცვლადების სახელების დასარქმევად ორჯერ *click* ცვლადის დასახელებაზე, ეკრანზე გამოსულ *Variable* ფანჯრის *Name* ველში შევიტანოთ ცვლადის ახალი დასახელება და შემდეგ *OK*. ეს პროცედურა გაიმეოროდ ყველა ცვლადისთვის. დაკვირვებების სახელების შესატანად ცხრილის ზედა მარცხენა არეში ორჯერ *click*, განხდება კურსორი და უკვე შეგვიძლია შევიტანოთ დასახელება (მაგ: პაციენტები, ქალაქები და სხვა). ცალკეული დაკვირვებების სახელების შესატანად ორჯერ *click* დაკვირვების ნომერზე და შემდეგ შევიტანოთ დასახელება. ფაილის დამატებით ინფორმაციის შესატანად (მაგ: გულით დაგადებული პაციენტების მაჩვენებლები) ცვლადების ზემოდ არსებული ცარიელ ველში ორჯერ *click* და შემდეგ შევიტანოთ ინფორმაცია.

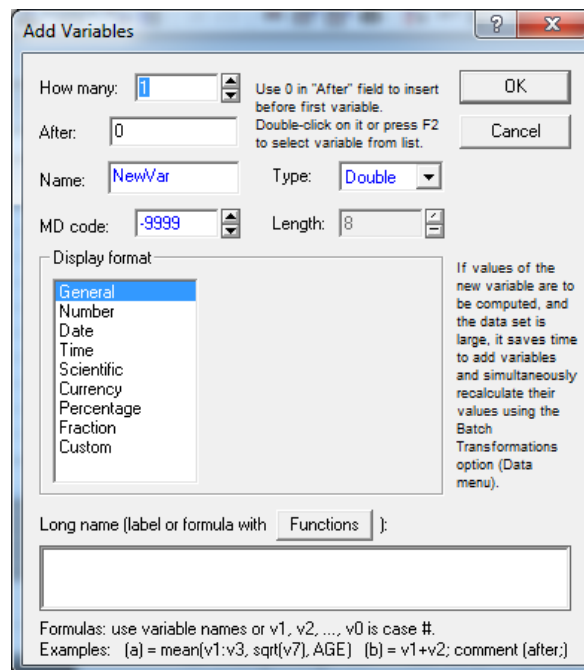
ცხრილის უჯრედებში მონაცემები შეგვყავს ჩვეულებრივ, ისევე როგორც *Excel*-ის ელექტრონულ ცხრილებში. უჯრედში ინფორმაციის შესამოწმებლად *click* და შევიყვანოთ ახალი მნიშვნელობა. თუ გვინდა, რომ უჯრედში ჩაწერილი ინფორმაცია არ წაიშალოს, მაშინ ორჯერ *click*, ამ დროს დაყენდება რედაქტირების რეჟიმი, კურსორი აღმოჩნდება უჯრედის შიგნით და უკვე შესაძლებელი იქნება შესწორებების შეტანა.

არსებული მონაცემების ფაილის გახსნა. *Statistica 7* გაშვებისას მას ავტომატურად გამოაქვს მონაცემების ის ცხრილი, რომელთანაც იგი ბოლო დროს მუშაობდა. აქვე გამოდის *Welcome to STATISTICA* ფანჯარა, სადაც *Most recently used files* ველის გაშლად სიაში წარმოდგენილია ადრე გამოყენებული ფაილები. ამ ფაილების ხელახლა გამოყენება შესაძლებელია თუ მის დასახელებაზე *click* და შემდეგ *OK*. გარდა ამისა, თუ ჩართულია *Open STATISTICA Data File* ოპცია, მაშინ *click OK*. ეკრანზე გამოდის *Datasets* ბიბლიოთეკის ფაილები.

## 2. ძირითადი ოპერაციები ცვლადებზე და დაკვირვებებზე

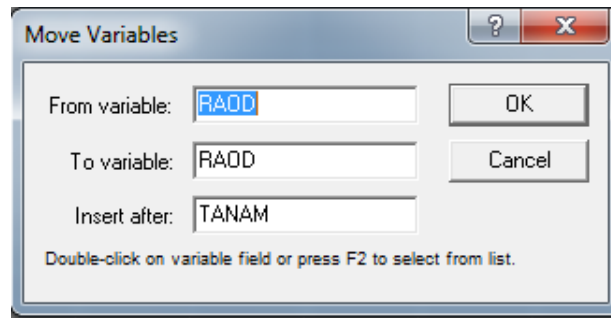
ცვლადებზე მოქმედებების ჩასატარებლად გამოიყენება **Data** მენიუ (**Data+Variables**) ან **click** ინსტრუმენტების პანელზე არსებული **Vars** ღილაკზე ან ცვლადის დასახელებაზე **click [R]** (მარჯვენა ღილაკზე დაწკაპუნება) და კონტექსტურ მენიუში შევარჩიოთ სასურველი ბრძანება.

**Add Variables** ახალი ცვლადის (ანუ ცარიელი სვეტის) დამატება იწვევს ცხრილის ზომის გაზრდას. ეკრანზე გამოდის **Add Variables** ფანჯარა, სადაც



- **How many:** (რამდენი) ველში უნდა მიუთითოდ დასამატებელი ცვლადების რაოდენობა (ეს რაოდენობა ცხრილისთვის არ არის შეზღუდული):
  - **After** (შემდეგ) აქ საჭიროა მიუთითოდ იმ ცვლადის დასახელება, რომლის შემდეგ გვინდა ახალი ცვლადის ან ცვლადების დამატება.
  - **Name** (სახელი) ველში შეგვიძლია ჩავწეროთ დასამატებელი ცვლადის სახელი.

**Move variables** (ცვლადის გადაადგილება)-ის ბრძანებით შეგვიძლია ცხრილში ცვლადის ან ცვლადების გადაადგილება. ეკრანზე გამოსულ **Move Variables** დიალოგურ სარკმელში (ფანჯარაში) **From variable** ველში მიუთითოდ გადასადგილებელი ცვლადის სახელი, ხოლო **Insert after** ველში იმ ცვლადის სახელი, რომლის შემდეგ გვინდა გადავიტანოთ ცვლადი.



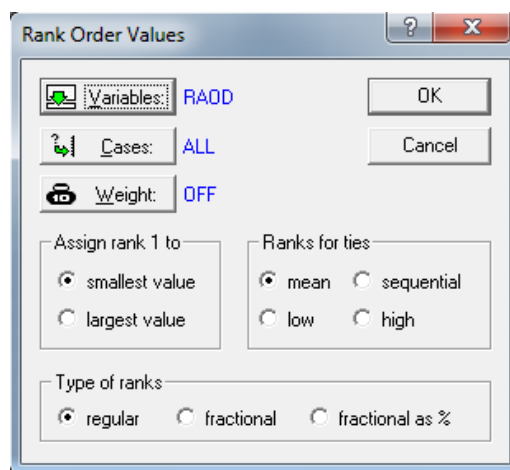
თუ გვინდა რამოდენიმე ცვლადის ერთდროულად გადაადგილება, მაშინ **To Variable** ველში მიუთითოდ პირველი გადასაადგილებელი ცვლადის სახელი, ხოლო **Insert after** ველში ბოლო გადასაადგილებელი ცვლადის სახელი.

**Copy variables** (ცვლადის კოპირება) გამოიყენება ცვლადის მნიშვნელობის კოპირებისთვის. ეკრანზე გამოდის **Copy Variables** ფანჯარა სადაც **From variable** ველში მიუთითებთ ცვლადის დასახელებას, ხოლო **Insert after** ველში იმ ცვლადის დასახელებას, რომლის შემდეგ გვინდა კოპირებული ცვლადის განთავსება. აქაც ერთდროულად შესაძლოა რამოდენიმე ცვლადის კოპირება.

**Delete variables** (ცვლადის წაშლა) ბრძანებით ხდება ცვლადის (ცვლადების) ცხრილიდან წაშლა. ეკრანზე გამოსულ **Delete Variables** ფანჯარის **From variables** ველში მოვნიშნოდ წასაშლელი ცვლადის (სვეტი) დასახელება და შემდეგ **OK**. რამოდენიმე ცვლადის წასაშლელად **To variable** ველში ჩავწერთ ბოლო წასაშლელი ცვლადის დასახელება.

ცვლადის მონაცემების წასაშლელად შევარჩიოთ **Clear** ან **Cut** ბრძანება.

**Rank** (რანგი) საშუალებით შესაძლებელია ცვლადის (ცხრილის სვეტში არსებული მონაცემები) მინიშვნელობებს მიენიჭოთ რანგული მნიშვნელობები. **Rank** ბრძანების შერჩევის შემდეგ ეკრანზე გამოდის **Rank Order Values** ფანჯარა,



სადაც განლაგებულია შემდეგი ღილაკები:

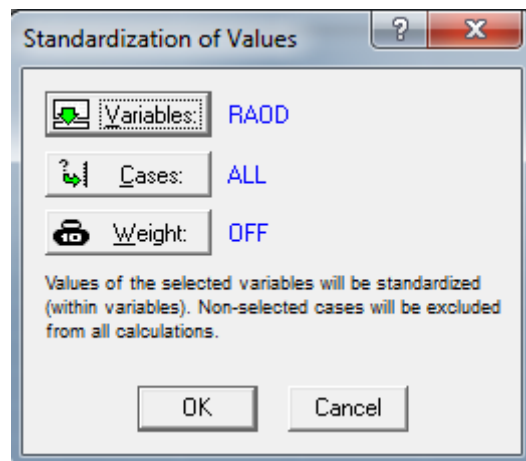
- **Variables** (ცვლადები), რომლის საშუალებითაც მონიშნება ცვლადი ან ცვლადები;

- **Sases** (დაკვირვება) გამოიყენება იმ დაკვირვებათა შესარჩევად, რომელთა რანჟირებაა საჭირო;
- **Weight** (წონის მინიჭება).

განვიხილოთ ფანჯარაში არსებული ოპერაციები.

- **Assign rank 1 to** (მივანიჭოთ რანგი 1). მნიშვნელობების რანჟირება შესაძლებელია ზრდადობით **Smallest value** (უმცირესი მნიშვნელობა) ოპციის ჩართვით დაწყებული 1-დან, ან კლებადობით **Largest value** (უდიდესი მნიშვნელობა) ოპციის ჩათვლით დაწყებული 1-დან;
- **Rank for ties** (ერთნაირი (თანხედომი) რანგები). **Mean** (საშუალო) ოპციის ჩართვით ერთნაირ რანგებს მიენიჭებათ ამ რანგების საშუალო მნიშვნელობა. **Sequential** (თანმიმდევრული) ოპერაციით ხდება ყოველი თანხედომი მნიშვნელობების სვეტში გამოჩენისთანავე მათი თანმიმდევრული რანჟირება;
- **Low** (ქვედა) და **High** (ზედა) ოპციებით ყოველი თანხედომი მნიშვნელობას მიენიჭება შესაბამისად ყველაზე მცირე ან ყველაზე დიდი რანგი თანხედომი მნიშვნელობების რანგებიდან;
- **Type of ranks** (რანგების ტიპი), **Regular** (ჩვეულებრივი) ოპციის საშუალებით რანჟირების დიაპაზონია 1-დან  $n$ -მდე ( $n$ -დაკვირვებათა რაოდენობა). **Fractional** (წილადი) ოპციის რანჟირების დიაპაზონია 0-დან 1-მდე. **Ractional as%** (წილადი %-ში) ოპციაში რანგები წარმოადგენენ % მნიშვნელობებს.

**Standardize** (სტანდარტიზაცია) ბრძანების შერჩევის შემდეგ ეკრანზე გამოდის **Standardize of Values** ფანჯარა.

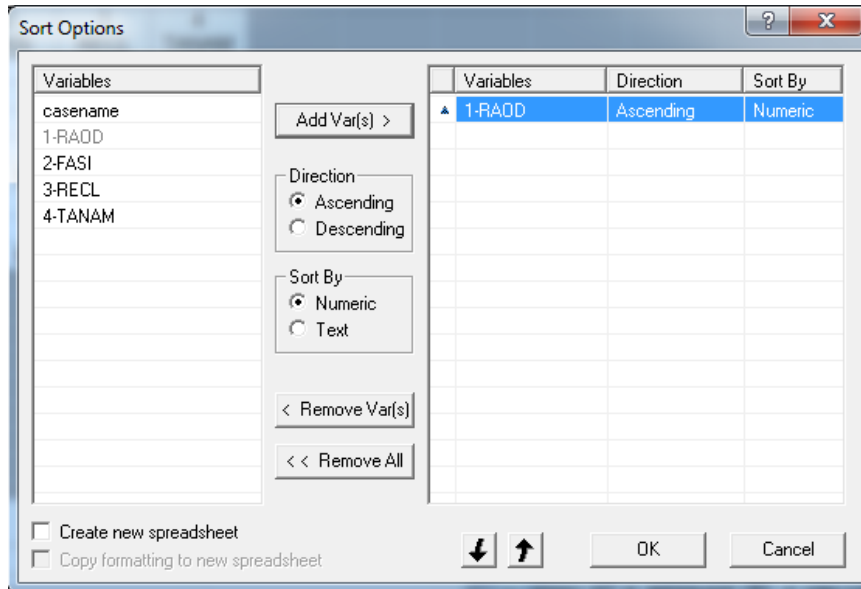


ცვლადის ან ცვლადების მოსანიშნათ *click Variables* ღილაკზე და შემდეგ *OK*. ვბრუნდებით **Standardize of Values** ფანჯარაში და კვლავ *OK*. ცხრილში ჩაიწერება სტანდარტიზირებული ნიშნელობები.

დაკვირვებებზე ძირითადი ოპერაციების ჩასატარებლად უნდა გამოვიყენოთ **Data** მენიუ (**Data+Cases**) ან ინსტრუმენტების პანელზე არსებული **Vars** ღილაკზე *click* ან დაკვირვების დასახელებაზე *click [R]* და კონტექსტურ მენიუში შევარჩიოთ სასურველი ბრძანება. **Add Cases, Move Cases,**

**Delete Cases, Copy Cases** ბრძანებების გამოყენების პროცედურები იგივეა, რაც ცვლადების დროს.

**Sort cases** (დაკვირვებების დახარისხება) ბრძანებით ეკრანზე გამოდის **Sort Options** ფანჯარა



სადაც **Variables** ჩამონათვალში მოვნიშნოთ ცვლადი (ცვლადები). **click Add var(s) >** ღლაკზე, **Direction** ჩარჩოში ჩავრთოდ **Ascending** (ზრდადობით) ან **Descending** (კლებადობით) ოპცია და შემდეგ **OK**.

**ფაილის იმპორტი.** შესაძლებელია სამი მეთოდით.

- ბუფერის საშუალებით;
- მონაცემების დინამიური გაცვლის ტექნოლოგიით;
- ფაილების იმპორტირება სპეციალური საშუალებით.

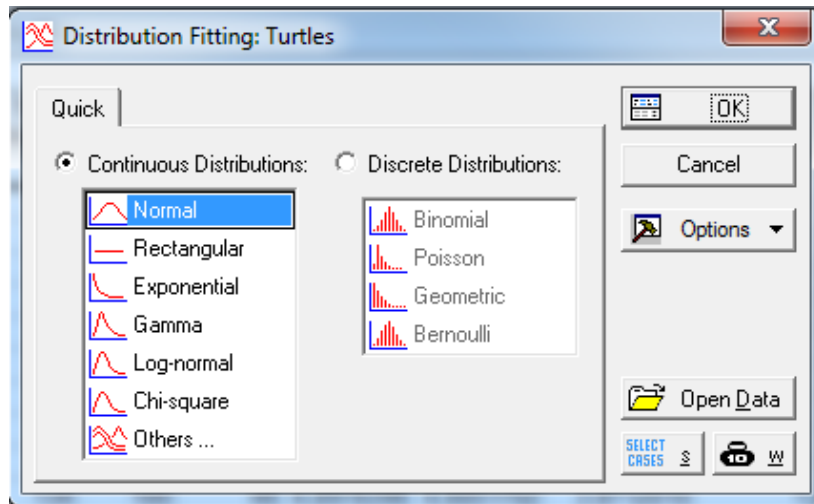
ამ სამიდან ყველაზე უფრო სწრაფი და პოპულარული მეთოდია ბუფერის გამოყენების მეთოდი. ამისთვის საწყის მონაცემებში მოვნიშნოთ ფაილი და მოვახდინოთ მისი კოპირება ჩვეულებრივი მეთოდით. კოპირების შემდეგ გავხსნათ ახალი ფაილი **Statistica** სისტემაში და მოვახდინოთ კოპირებული ფაილის ჩასმა. ამ დროს მონაცემები ელექტრული ცხრილში განლაგდებიან მონიშნული უჯრიდან მარჯვნივ და ქვემოთ.

### 3. განაწილების კანონის შერჩევა / შემოწმება

მონაცემების დამუშავებისას ხშირად საჭიროა ემპირიული განაწილების ფუნქციის აპროქსიმაცია ამა თუ იმ ცნობილი განაწილების კანონით. ამისათვის **Statistica** პაკეტში გათვალისწინებულია **Distribution Fitting** (განაწილების მორგება) მოდული. ამ მოდულის გასაცნობად განვიხილოთ **Examples** ბიბლიოთეკიდან **Turtles.sta** ფაილი.

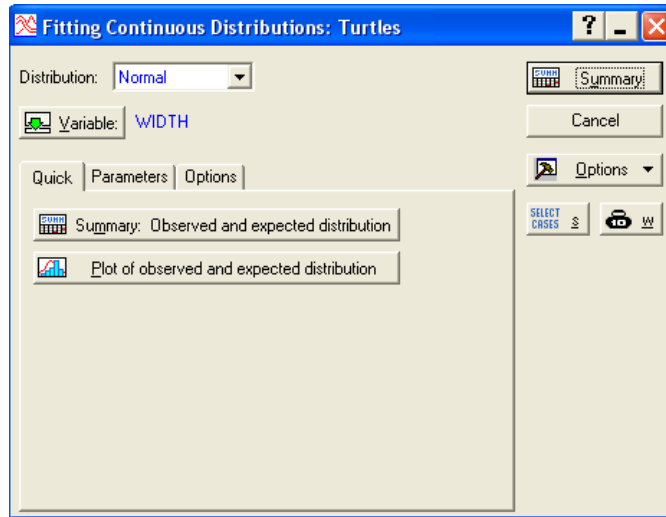
	1	2	3	4	5	6	7
	LENGTH	WIDTH	HEIGHT	LOG LEN	LOG WID	LOG HGHT	GENDER
1	98	81	38	4,5849675	4,3944492	3,6375862	1
2	103	84	38	4,6347290	4,4308168	3,6375862	1
3	103	86	42	4,6347290	4,4543473	3,7376696	1
4	105	86	42	4,6539604	4,4543473	3,7376696	1
5	109	88	44	4,6913479	4,4773368	3,7841896	1
6	123	92	50	4,8121844	4,5217886	3,9120230	1
7	123	95	46	4,8121844	4,5538769	3,8286414	1
8	133	99	51	4,8903491	4,5951199	3,9318256	1
9	133	102	51	4,8903491	4,6249728	3,9318256	1
10	133	102	51	4,8903491	4,6249728	3,9318256	1
11	134	100	48	4,8978398	4,6051702	3,8712010	1
12	136	102	49	4,9126549	4,6249728	3,8918203	1
13	138	98	51	4,9272537	4,5849675	3,9318256	1
14	138	99	51	4,9272537	4,5951199	3,9318256	1
15	141	105	53	4,9487599	4,6539604	3,9702919	1
16	147	108	57	4,9904326	4,6821312	4,0430513	1
17	149	107	55	5,0039463	4,6728288	4,0073332	1
18	153	107	56	5,0304379	4,6728288	4,0253517	1
19	155	115	63	5,0434251	4,7449321	4,1431347	1
20	155	117	60	5,0434251	4,7621739	4,0943446	1

გაგხსნათ მენიუ *Statisties* და შევარჩიოთ *Distribution Fiting*. ეკრანზე გამოსულ ამავე დასახელების ფანჯარაში უნდა მიუთითოდ შემთხვევითი სიდიდის ბუნება, ანუ ჩაეროთ *Continuous Distributions* (უწყვეტი განაწილება) ან *Discret Distrubutions* (დისკრეტული განაწილება) ოპცია. შემდეგ უნდა მოვნიშნოთ მოსალოდენელი განაწილების კანონი.



უწყვეტი შემთხვევითი სიდიდებისთვის წარმოდგენილია 6 განაწილების კანონი, ხოლო დისკრეტული შემთხვევითი სიდიდისთვის – ოთხი. შევარჩიოთ რომელიმე მათგანი მაგ:ლითად, ნორმალური და *OK*. გახსნილ *Fiting Continuous Distributions* ფანჯარაში ფაილის მოსანიშნად *click Variable* ღილაკზე და შემდეგ *OK*.





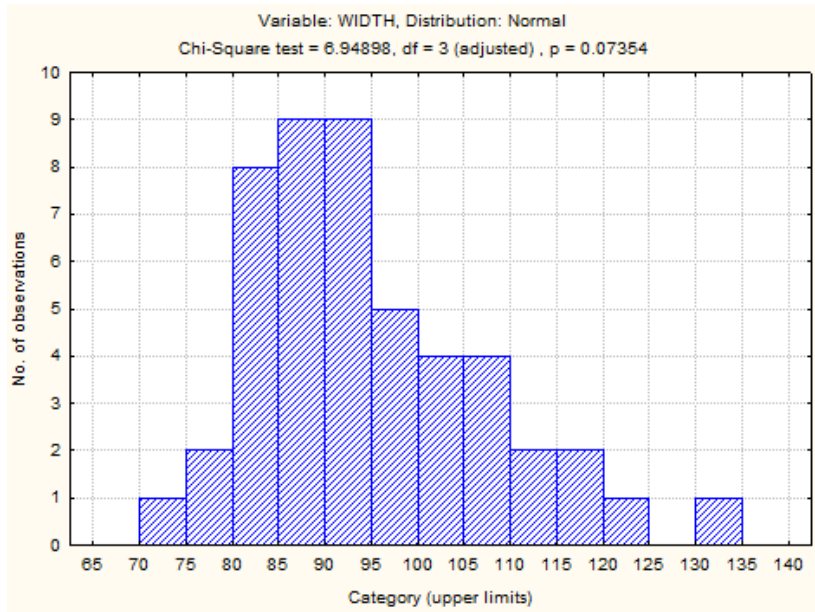
ვბრუნებით უკან ფანჯარაში, სადაც **Quick** ჩანართში წარმოდგენილია ორი ღილაკი. **Summary: Observed and expected distribution** (შედეგი: მონაცემები და მოსალოდნელი განაწილება) და **Plot of Observed and expected distribution** (მონაცემები და მოსალოდნელი განაწილების გრაფიკები).

პირველ ღილაკზე დაწკაპუნებით პროგრამას ეკრანზე გამოაქვს რიცხვითი მნიშვნელობის ცხრილი.

Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected
<= 70.00000	0	0	0,00000	0,0000	1,074578	1,07458	2,23870	2,2387	-1,07458
75.00000	1	1	2,08333	2,0833	1,490841	2,56542	3,10592	5,3446	-0,49084
80.00000	2	3	4,16667	6,2500	2,793159	5,35858	5,81908	11,1637	-0,79316
85.00000	8	11	16,66667	22,9167	4,487925	9,84650	9,34984	20,5135	3,51208
90.00000	9	20	18,75000	41,6667	6,184254	16,03076	12,88386	33,3974	2,81575
95.00000	9	29	18,75000	60,4167	7,308450	23,33921	15,22594	48,6233	1,69155
100.00000	5	34	10,41667	70,8333	7,407328	30,74653	15,43193	64,0553	-2,40733
105.00000	4	38	8,33333	79,1667	6,438676	37,18521	13,41391	77,4692	-2,43868
110.00000	4	42	8,33333	87,5000	4,799856	41,98507	9,99970	87,4689	-0,79986
115.00000	2	44	4,16667	91,6667	3,068687	45,05375	6,39310	93,8620	-1,06869
120.00000	2	46	4,16667	95,8333	1,682534	46,73629	3,50528	97,3673	0,31747
125.00000	1	47	2,08333	97,9167	0,791141	47,52743	1,64821	99,0155	0,20886
130.00000	0	47	0,00000	97,9167	0,319016	47,84644	0,66462	99,6801	-0,31902
135.00000	1	48	2,08333	100,0000	0,110314	47,95676	0,22982	99,9099	0,88969
< Infinity	0	48	0,00000	100,0000	0,043243	48,00000	0,09009	100,0000	-0,04324

ცხრილის ყოველი სტრიქონი ახასიათებს ინტერვალს, სადაც მოხვდებიან შესარჩევი ცვლადის მნიშვნელობები. პირველ სვეტში **Observed Frequency** (დაკვირვების სიხშირე) ნაჩვენებია ინტერვალის სიხშირე. მეორე სვეტში **Cuma lative observed** (დაგროვილი სიხშირეები), მესამე და მეოთხე სვეტებში ნაჩვენებია **Percent observed** (დაკვირვების პროცენტი) და **Cumul %** (ჯამური პროცენტი). მეხუთე სვეტში **Frequency Expected** (მოსალოდნელი სიხშირეები) წარმოდგენილია თეორიული სიხშირეების მნიშვნელობები.

მეორე ღილაკზე დაწკაპუნებით ეკრანზე გამოდის თეორიული მნიშვნელობის კანონის მრუდი და ჰისტოგრამა.



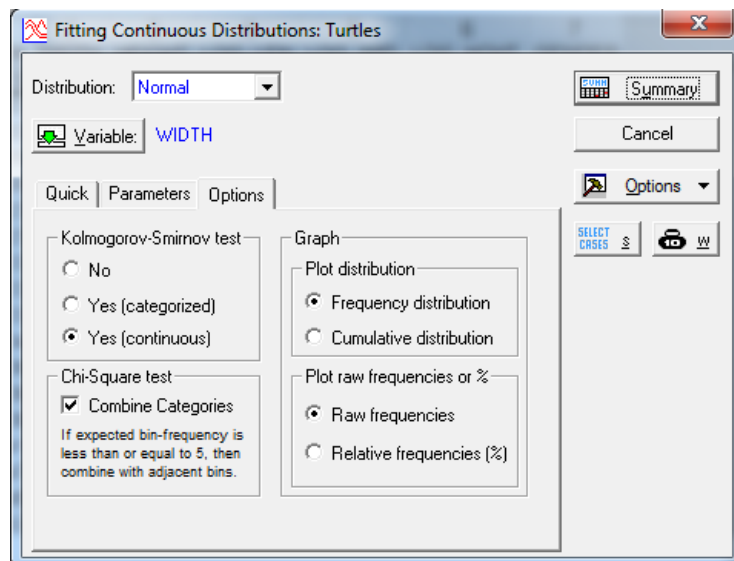
ჰისტოგრამის ზედა ნაწილში განთავსებულია ცვლადის დასახელება და მოსალოდნელი განაწილების კანონი. მეორე სტრიქონში პირველ პარამეტრს წარმოადგენს პირსონის  $\chi^2$  თანხმობის კრიტერიუმის მნიშვნელობა, მეორეს – თავისუფლების ხარისხი  $df$  და მესამეს –  $p$  - მნიშვნელობა. თუ  $p > 0.05$ , მაშინ ნულოვანი ჰიპოთეზა მიღებულია, ე.ი. მონაცემები განაწილებულია აღნიშნული კანონით.

დავბრუნდეთ *Fitting Continuous Distributions* ფანჯარაში *click Parameters* ჩანართზე. ეკრანზე გამოდის მოსალოდნელი განაწილების კანონის პარამეტრები, კერძოდ:

- **Set to default** - მონაცემები დავაყენოთ გაჩუმების პრინციპით.

აქ მოყვანილი პარამეტრებიდან სამი წარმოადგენს ძირითადს, რომელზეც დამახასიათებელია ყველა განაწილებლის კანონისთვის, ხოლო დანარჩენი დამახასიათებელი არიან შერჩეული განაწილების კანონისთვის.

- **Number of categories** (კატეგორიების რიცხვი) - ეს არის ამონარჩევის ინტერვალებად დაყოფის რაოდენობა;
- **Lower Limit, Uper Limit** (ქვედა და ზედა მნიშვნელობები). განუმეების პრინციპით აიღება ამონარჩევის მინიმალური და მაქსიმალური მნიშვნელობები;
- **Mean and Variance** (საშუალო და დისპერსია) - მხოლოდ ნორმალური განაწილების კანონისთვის. ამ პარამეტრებს პროგრამა ავტომატურად განსაზღვრავს, მაგრამ შესაძლებელია ხელითაც იყოს გამოთვლილი;
- **Options** ჩანართის ჩართვისას ეკრანზე გამოდის ოთხი ჩარჩო:



**Kologonov-Smirnov test** (კოლმოგონოვ-სმირნოვის ტესტი). აქ შეიძლება ავირჩიოთ სამი ოპცია: ტესტი არ გამოიყენება; ტესტი განისაზღვრება ინტერვალურად დაჯგუფებულ მონაცემებისათვის და ტესტი განისაზღვრება დაუჯგუფებელი მონაცემებისათვის.

**Chi - Square test** (პირსონის  $\chi^2$  თანხმობის ტესტი). თუ ჩართულია **Combine Categories** ალაში და ინტერვალში მოხვდა 5-ზე ნაკლები მნიშვნელობა, მაშინ ასეთი ინტერვალის გაერთიანება მეზობელ ინტერვალთან და ა.შ. მანამ სანამ ინტერვალში მოხვედრილი მნიშვნელობები 6-ზე მეტი არ იქნება. სხვა შემთხვევებში ინტერვალების გაერთიანება არ ხდება.

**Graph Plot Distributions** ოპციებით ხდება განაწილების გრაფიკების ტიპის დადგენა. თუ ჩართულია **Frequency distribution** ოპცია, მაშინ პროგრამა ააგებს განაწილების სიმკვრივის გრაფიკს, თუ ჩართულია **Cumulative distribution** ოპცია, მაშინ აგებული იქნება განაწილების ფუნქციის გრაფიკი.

**Plot raw frequencies or %** ოპციებში თუ ჩართულია **Raw frequencies** ოპცია, მაშინ გრაფიკის ვერტიკალურ ღერძზე გადაზომილი იქნება ფარდობითი სიხშირები, წინააღმდეგ შემთხვევაში ანუ როცა ჩართულია **Relative frequencies (%)** ოპცია - მათი პროცენტული ფარდობები.

## პრაქტიკული სამუშაო 2

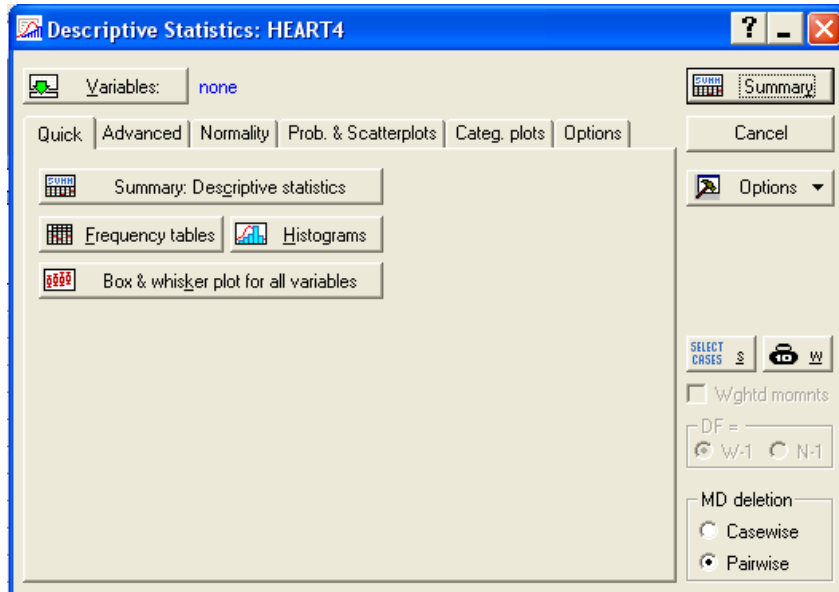
### ძირითადი სტატისტიკური მახასიათებლების განსაზღვრა. კორელაციური ანალიზი.

#### 1. აღწერითი სტატისტიკა.

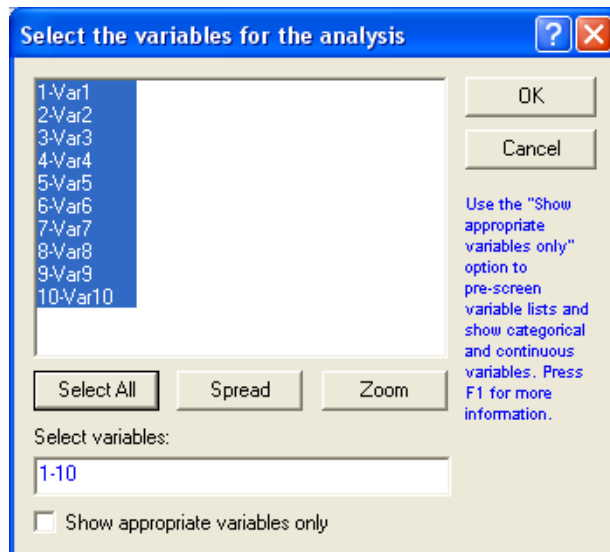
სტატისტიკური მახასიათებლების განსაზღვრისათვის გამოვიყენოთ *Examples* ბიბლიოთეკიდან შერჩეული *Heart 5* მონაცემთა ფაილი, სადაც წარმოდგენილია 6 ცვლადი: *AGE* (ასაკი), *SYST* (სისტოლური წნევა), *DIAST* (დიასტოლური წნევა), *HEIGHT* (სიმაღლე), *WEIGHT* (წონა) და *SER CH* (შრატში ხოლესტერინის რაოდენობა).

	1	2	3	4	5	6
	AGE	SYST	DTAST	HEIGHT	WEIGHT	SER CH
1	48	130	90	67	165	219
2	60	124	80	74	235	203
3	59	160	100	72	206	269
4	40	120	80	69	148	185
5	56	115	80	64	147	260
6	58	140	90	63	121	312
7	64	135	85	64	189	185
8	57	110	78	70	173	282
9	32	112	70	69	171	254
10	59	140	90	65	150	303
11	48	130	80	64	147	271
12	47	115	84	67	211	304
13	47	130	80	67	147	334
14	28	120	86	70	189	328
15	37	95	55	69	190	226
16	54	141	100	65	171	363
17	38	130	990	67	170	399
18	52	125	90	65	141	199
19	46	110	70	67	159	271
20	51	120	80	70	139	261
21	49	120	80	68	194	263
22	46	110	70	66	160	242
23	26	110	80	70	206	260
24	35	120	80	72	191	321

პროგრამის გასაშვებად *Statistics* მენიუში შევარჩიოთ *Basic Statistic Tables* (ძირითადი სტატისტიკები/ცხრილები). ეკრანზე გამოსულ *Basic Statistic and Tables* ფანჯრის მენიუში მოვნიშნოთ *Descriptive statistics* და შემდეგ *OK*.



ეკრანზე გამოდის *Descriptive Statistics* (აღწერითი სტატისტიკა)-ის სასტარტო ფანჯარა. ცვლადის ან ცვლადების მოსანიშნათ *click Variables* ღილაკზე და ეკრანზე გამოსულ ფანჯარაში მოვნიშნოთ ცვლადები. შემდეგ *OK*.

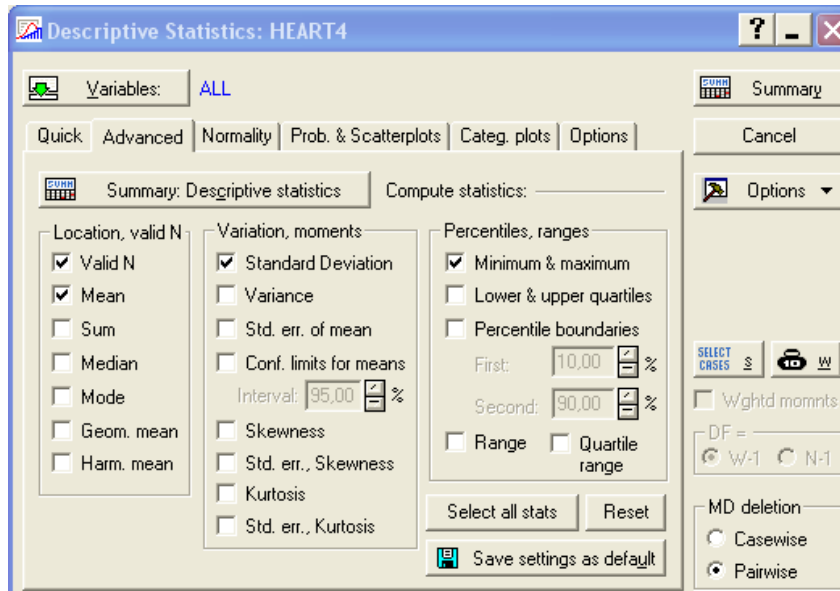


გბრუნდების სასტარტო ფანჯარაში და *click Summary: Descriptive Statistics* ღილაკზე. ეკრანზე გამოდის ცხრილი

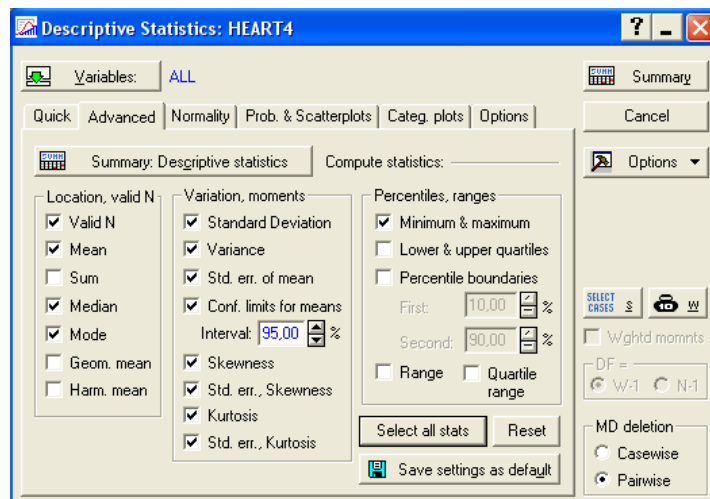
Variable	Descriptive Statistics (HEART 5)				
	Valid N	Mean	Minimum	Maximum	Std.Dev.
AGE	30	48,1333	26,0000	64,0000	9,7618
SYST	30	122,7333	95,0000	160,0000	12,6353
DTAST	30	110,7000	55,0000	990,0000	166,3524
HEIGHT	30	67,8000	63,0000	74,0000	2,8575
WEIGHT	30	171,8000	121,0000	235,0000	30,3513
SER CH	30	270,5667	185,0000	399,0000	52,6417

სადაც წარმოდგენილია ძირითადი სტატისტიკული მახასიათებლები: საშუალო, სტანდარტული გადახრა, ცვლადების მინიმალური და მაქსიმალური მნიშვნელობები და ცვლადების რაოდენობა.

თუ ჩვენ გვინტერესებს სხვა სტატისტიკური მახასიათებლები, მაშინ გავხსნათ **Advanced** ჩანართი.



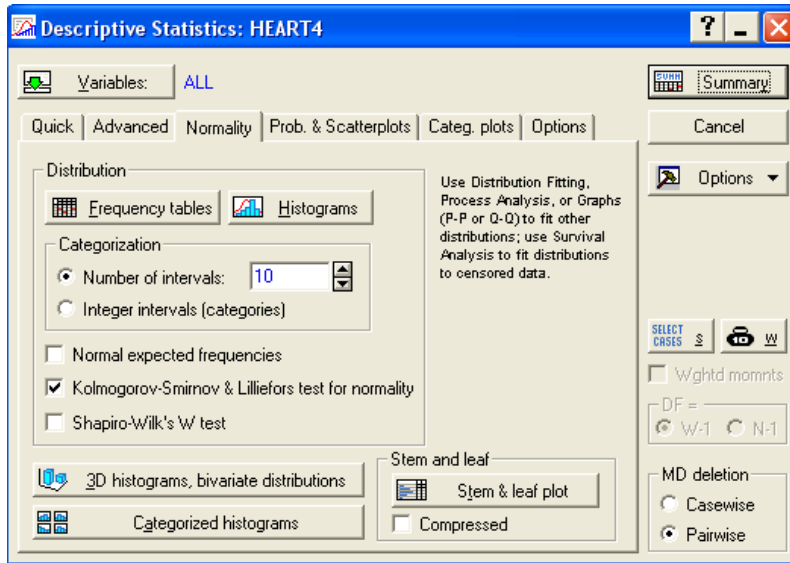
ეკრანზე გამოსულ ფანჯარაში ჩაერთოდ ჩვენთვის სასურველი სტატისტიკური მახასიათებლების ოპციები. **Select all stats** ღილაკით ჩაირთვება ყველა ოპცია.



შემდეგ **click Summary: Descriptive Statistics** ღილაკზე. ეკრანზე გამოდის მონიშნული სტატისტიკური მახასიათებლების შესაბამისი ცხრილი.

Variable	Descriptive Statistics (HEART 5)													
	Valid N	Mean	Median	Mode	Frequency of Mode	Minimum	Maximum	Variance	Std.Dev.	Standard Error	Skewness	Std.Err. Skewness	Kurtosis	Std.Err. Kurtosis
AGE	30	48,1333	48,5000	Multiple	3	26,0000	64,0000	95,29	9,7618	1,78225	-0,663815	0,426892	-0,17586	0,832746
SYST	30	122,7333	120,0000	120,0000	7	95,0000	160,0000	159,65	12,6353	2,30688	0,693418	0,426892	1,65355	0,832746
DTAST	30	110,7000	80,0000	80,00000	10	55,0000	990,0000	27673,11	166,3524	30,37165	5,447697	0,426892	29,77772	0,832746
HEIGHT	30	67,8000	67,5000	Multiple	5	63,0000	74,0000	8,17	2,8575	0,52171	0,177847	0,426892	-0,72189	0,832746
WEIGHT	30	171,8000	168,5000	147,0000	4	121,0000	235,0000	921,20	30,3513	5,54136	0,465243	0,426892	-0,66442	0,832746
SER CH	30	270,5667	266,0000	Multiple	2	185,0000	399,0000	2771,15	52,6417	9,61102	0,389789	0,426892	-0,07119	0,832746

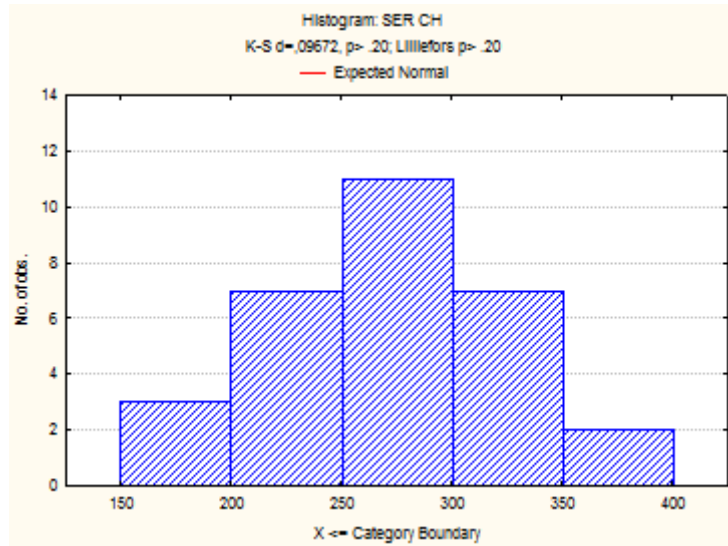
ცვლადის ან ცვლადების ნორმალური განაწილების კანონის გრაფიკულად შესამოწმებლად გამოიყენება **Descriptive Statistics** ფანჯრის **Normality** ჩანართი. ეკრანზე გამოდის ფანჯარა



თუ უწყვეტ შემთხვევით სიდიდესთან გვაქვს საქმე, მაშინ უნდა ჩავრთოდ **Number of intervals** ოპცია (დისკრეტული სიდიდისათვის – **Integer intervals** ოპცია). ჩავრთოდ **Kolmogorov-Smirnov&Liliefots test.....** ოპცია. **Frequency tables** დილაკით ეკრანზე გამოდის ცვლადების სისშირული ცხრილები.

Frequency table: SER CH (HEART 5)						
K-S d=.09672, p> .20; Lilliefors p> .20						
Category	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
150,0000<x<=200,0000	3	3	10,00000	10,0000	10,00000	10,0000
200,0000<x<=250,0000	7	10	23,33333	33,3333	23,33333	33,3333
250,0000<x<=300,0000	11	21	36,66667	70,0000	36,66667	70,0000
300,0000<x<=350,0000	7	28	23,33333	93,3333	23,33333	93,3333
350,0000<x<=400,0000	2	30	6,66667	100,0000	6,66667	100,0000
Missing	0	30	0,00000		0,00000	100,0000

**Histograms** დილაკით გამოდის ცვლადების ჰისტოგრამები.



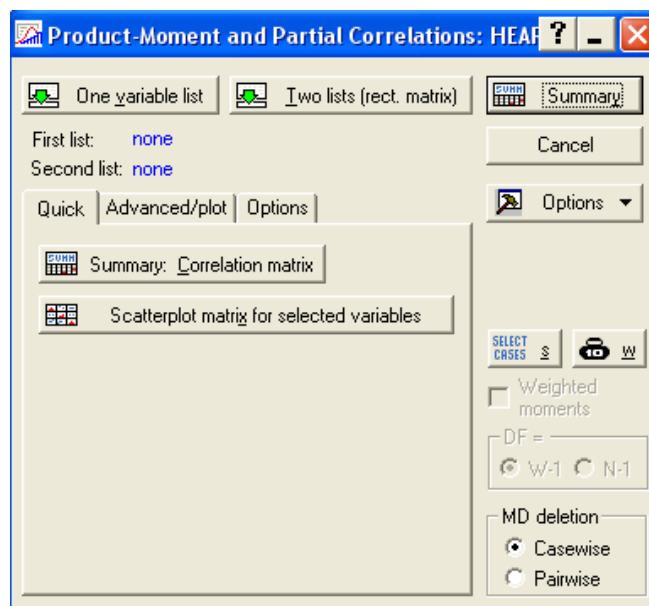
**Prob.&Scatterplots** ჩანართი ჩართვით ეკრანზე შესაძლებელია ცვლადების გრაფიკული გამოკვლევების ჩატარება.

**Categ.plots** ჩანართით შესაძლებელია ცვლადების კატეგორიზებული დაყოფის შედეგად მიღებული ჯგუფების გრაფიკული ანალიზის ჩატარება.

**Options** ჩანართით შესაძლებელია ცვლადების სტატისტიკური გამოკვლევების პარამეტრების შეცვლა.

## 2. კორელაციური ანალიზი.

**Descriptive Statistics** მოდულში განისაზღვრება პირსონის კორელაციის კოეფიციენტი. ამისათვის **Statistics** მენიუში შევარჩიოთ **Basic Statistic Tables** და ეკრანზე გამოსულ **Basic Statistic and Tables** ფანჯრაში მოვნიშნოთ **Correlation matrices** და შემდეგ **OK**. ეკრანზე გამოდის **Product Moment and Partial Correlations** სასტარტო ფანჯარა.



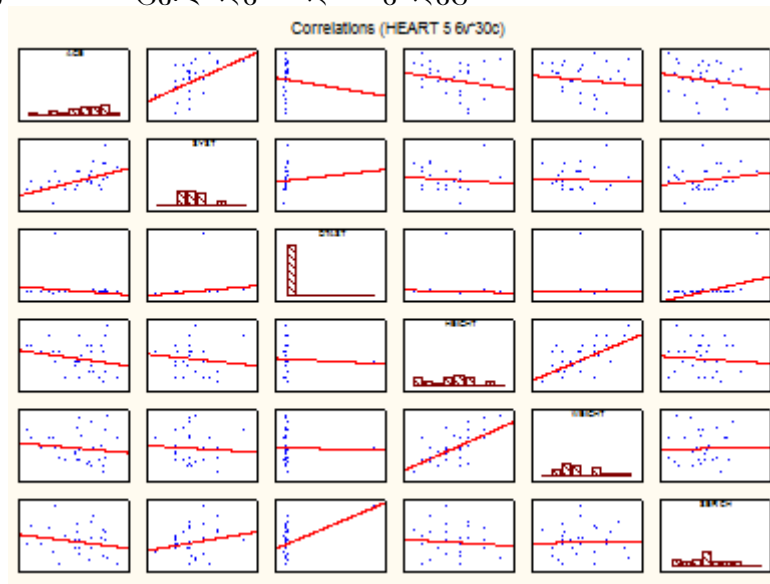


ცვლადების დასახელება შეიძლება მოცემული იყოს ერთი სიით (**One variable list**) ან ორი სიით (**Two fist(rect.matrix)**). პირველ შემთხვევაში აიგება კვადრატული კორელაციური მატრიცა, რომლის სტრიქონებს და სვეტებს წარმოადგენენ ცვლადები. მეორე შემთხვევის დროს აიგება მართკუთხა კორელაციური მატრიცა, რომლის სტრიქონები და სვეტები წარმოდგენილი არიან შესაბამისად ცვლადების პირველი და მეორე სიით.

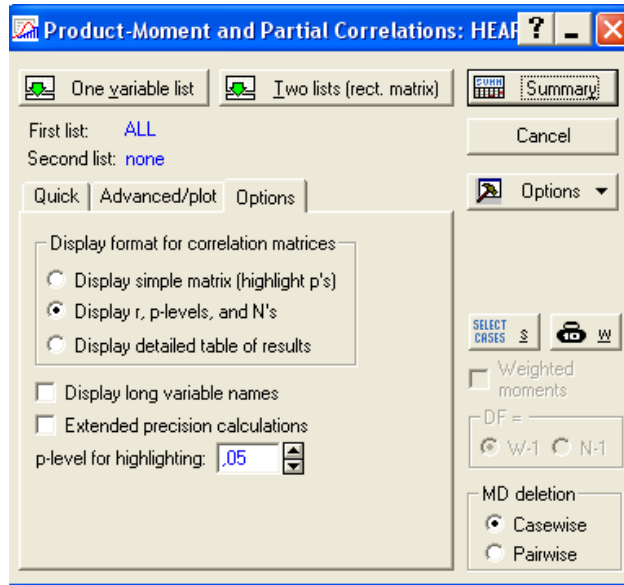
click **One variable list** ღილაკზე. ეკრანზე გამოსულ ფანჯარაში მოვნიშნოთ ცვლადები და შემდეგ **OK**. ვბრუნდებით საწყის ფანჯარაში. click **Summary: Correlations matrix** ღილაკზე. ეკრანზე გამოდის კვადრატული კორელაციური მატრიცა:

Correlations (HEART 5)						
Marked correlations are significant at p < ,05000						
N=30 (Casewise deletion of missing data)						
Variable	AGE	SYST	DTAST	HEIGHT	WEIGHT	SER CH
AGE	1,00	0,51	-0,17	-0,23	-0,14	-0,21
SYST	0,51	1,00	0,15	-0,12	-0,07	0,21
DTAST	-0,17	0,15	1,00	-0,06	-0,00	0,48
HEIGHT	-0,23	-0,12	-0,06	1,00	0,66	-0,10
WEIGHT	-0,14	-0,07	-0,00	0,66	1,00	0,03
SER CH	-0,21	0,21	0,48	-0,10	0,03	1,00

სადაც წითელი ფერით აღნიშნულია სარწმუნო კორელაციის კოეფიციენტები. **Scatterplot matrix for selected variables** ღილაკის საშუალებით შესაძლებელია ავაგოთ გაფანტვის ფუნქციის გრაფიკები და შერჩეული ცვლადების ჰისტოგრამები. ეკრანზე გამოდის ცვლადების მოსანიშნი ფანჯარა. მოვნიშნოთ ცვლადები და შემდეგ **OK**.



სასტარტო ფანჯარაში ჩავრთოდ **Options** ჩანართი



სადაც შესაძლებელია კორელაციური ანალიზის პარამეტრების შეცვლა. ასე მაგალითად, თუ ჩართულია **Display simple matrix** ოპცია, მაშინ ეკრანზე გამოდის უბრალო კორელაციური მატრიცა.

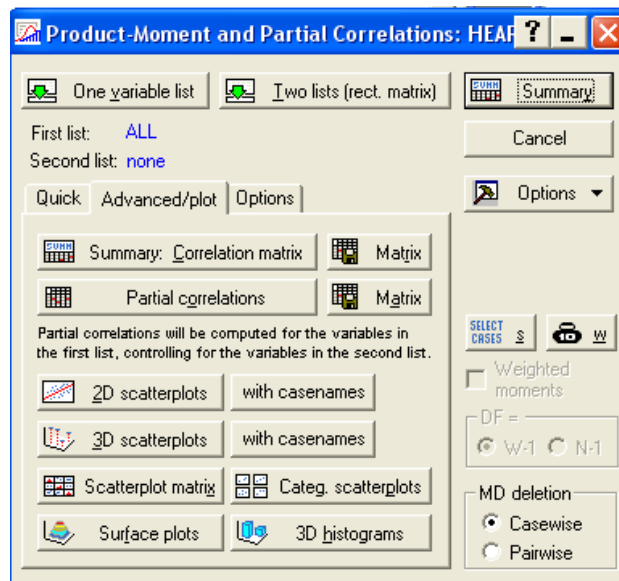
თუ ჩაერთავთ **Display r, p-levels, and Ns** ოპციას, მაშინ კორელაციური მატრიცის უჯრედში კორელაციის კოეფიციენტთან ერთად წარმოდგენილი იქნება მისი შესაბამისი **p** მნიშვნელობის დონე. *click Summary* ღილაკზე ეკრანზე გამოდის შემდეგი ცხრილი:

Correlations (HEART 5)						
Marked correlations are significant at $p < .05000$						
N=30 (Casewise deletion of missing data)						
Variable	AGE	SYST	DTAST	HEIGHT	WEIGHT	SER CH
AGE	1,0000	,5119	-,1748	-,2252	-,1398	-,2083
	p= ---	p=,004	p=,356	p=,231	p=,461	p=,269
SYST	,5119	1,0000	,1542	-,1238	-,0677	,2105
	p=,004	p= ---	p=,416	p=,515	p=,722	p=,264
DTAST	-,1748	,1542	1,0000	-,0559	-,0034	,4756
	p=,356	p=,416	p= ---	p=,769	p=,986	p=,008
HEIGHT	-,2252	-,1238	-,0559	1,0000	,6611	-,0971
	p=,231	p=,515	p=,769	p= ---	p=,000	p=,610
WEIGHT	-,1398	-,0677	-,0034	,6611	1,0000	,0316
	p=,461	p=,722	p=,986	p=,000	p= ---	p=,868
SER CH	-,2083	,2105	,4756	-,0971	,0316	1,0000
	p=,269	p=,264	p=,008	p=,610	p=,868	p= ---

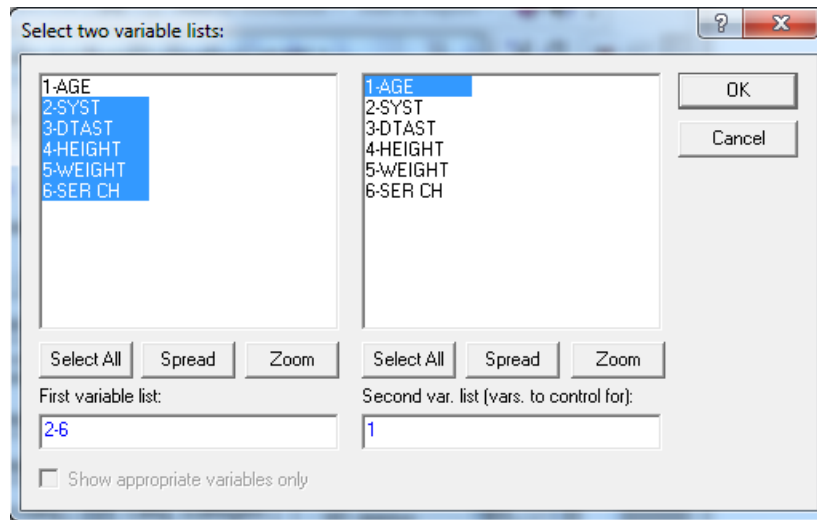
თუ ჩაერთავთ **Display detailed table of results** ოპციას, მაშინ კორელაციურ კოეფიციენტთან ერთად მოცემული იქნება ცვლადის სტატისტიკური ანალიზის შედეგები: საშუალო, სტანდარტული გადახრა, *t*-კრიტერიუმის მნიშვნელობა საშუალოების შესადარებლად და სხვა.

Correlations (HEART 5)											
Marked correlations are significant at $p < ,05000$											
(Casewise deletion of missing data)											
Var. X & Var. Y	Mean	Std.Dv.	r(X,Y)	r <sup>2</sup>	t	p	N	Constant dep: Y	Slope dep: Y	Constant dep: X	Slope dep: X
AGE	48,1333	9,7618									
AGE	48,1333	9,7618	1,000000	1,000000			30	0,000	1,00000	0,000	1,00000
AGE	48,1333	9,7618									
SYST	122,7333	12,6353	0,511909	0,262051	3,15325	0,003832	30	90,840	0,66260	-0,406	0,39549
AGE	48,1333	9,7618									
DTAST	110,7000	166,3524	-0,174757	0,030540	-0,93918	0,355671	30	254,044	-2,97807	49,269	-0,01025
AGE	48,1333	9,7618									
HEIGHT	67,8000	2,8575	-0,225232	0,050730	-1,22325	0,231432	30	70,974	-0,06593	100,300	-0,76943
AGE	48,1333	9,7618									
WEIGHT	171,8000	30,3513	-0,139802	0,019544	-0,74710	0,461233	30	192,722	-0,43467	55,858	-0,04496
AGE	48,1333	9,7618									
SER CH	270,5667	52,6417	-0,208307	0,043392	-1,12698	0,269318	30	324,636	-1,12332	58,585	-0,03863
SYST	122,7333	12,6353									
AGE	48,1333	9,7618	0,511909	0,262051	3,15325	0,003832	30	-0,406	0,39549	90,840	0,66260
SYST	122,7333	12,6353									
SYST	122,7333	12,6353	1,000000	1,000000			30	0,000	1,00000	0,000	1,00000
SYST	122,7333	12,6353									
DTAST	110,7000	166,3524	0,154188	0,023774	0,82576	0,415920	30	-138,448	2,02999	121,437	0,01171

*Advanced/plot* ჩანართით ხდება შერჩეული ცვლადების სტატისტიკური ანალიზის გრაფიკული წარმოდგენის გაფართოებული მომსახურება.



კორელაციის კერძო კოეფიციენტების მატრიცის გამოთვლისათვის *click Two fist(rect.matrix)* ღილაკზე.



ეკრანზე გამოსული ფანჯრის პირველ სვეტში მონიშნეთ ცვლადები, ხოლო მეორე სვეტში ერთი კონტროლირებადი ცვლადი. შემდეგ *click Partial correlations* ღილაკზე:

Partial Correlations (HEART 5)						
Marked correlations are significant at $p < ,05000$						
N=30 (Casewise deletion of missing data)						
Variable	AGE	SYST	DTAST	HEIGHT	WEIGHT	SER CH
AGE	1,00					
SYST		1,00	0,29	-0,01	0,00	0,38
DTAST		0,29	1,00	-0,10	-0,03	0,46
HEIGHT		-0,01	-0,10	1,00	0,65	-0,15
WEIGHT		0,00	-0,03	0,65	1,00	0,00
SER CH		0,38	0,46	-0,15	0,00	1,00

Partial Correlations (HEART 5)					
Marked correlations are significant at $p < ,05000$					
N=30 (Casewise deletion of missing data)					
Variable	AGE	SYST	DTAST	WEIGHT	SER CH
AGE	1,00	0,50	-0,19	0,01	-0,24
SYST	0,50	1,00	0,15	0,02	0,20
DTAST	-0,19	0,15	1,00	0,04	0,47
WEIGHT	0,01	0,02	0,04	1,00	0,13
SER CH	-0,24	0,20	0,47	0,13	1,00

### პრაქტიკული სამუშაო 3

#### მრავლობითი წრფივი რეგრესია

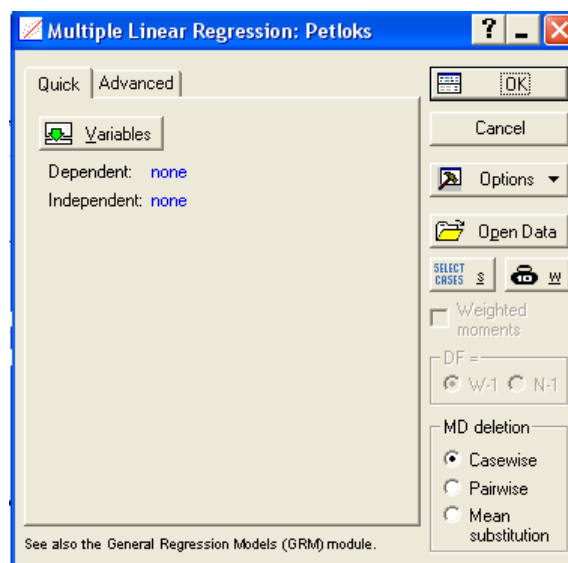
წრფივი რეგრესიის მოდელის მუშაობა განვიხილოთ *Petlocs* კომპანის მონაცემებისათვის, რომლებიც წარმოდგენილია *Petloks.sta* ფაილის სახით, სადაც წარმოდგენილია 4 ცვლადი: კომპანის თვიური გაყიდვების რაოდენობა (*RAOD*) (მლ.დოლარი), ერთი პროდუქტის ფასი (*FASI*), რეკლამის ხარჯი (*REKL*) და დასაქმებული თანამშრომლების რაოდენობა (*TANAM*).

mesiac	Data: Petloks			
	1 RAOD	2 FASI	3 REKL	4 TANAM
1	4	1	8	24
2	5,2	0,9	9	26
3	3,8	1,1	6	20
4	2,9	1,2	5	18
5	4,6	0,95	7	20
6	4,5	0,9	6	30
7	3,7	1	6	27
8	5	0,95	10	28

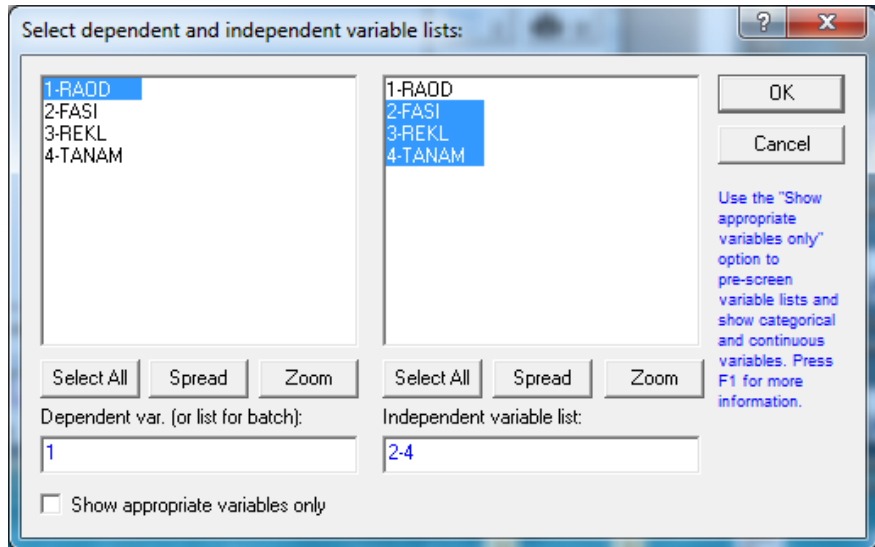
შევარჩიოთ წრფივი რეგრესიის განტოლების შემდეგი სახე:

$$RAOD = b_0 + b_1 FASI + b_2 REKL + b_3 TANAM$$

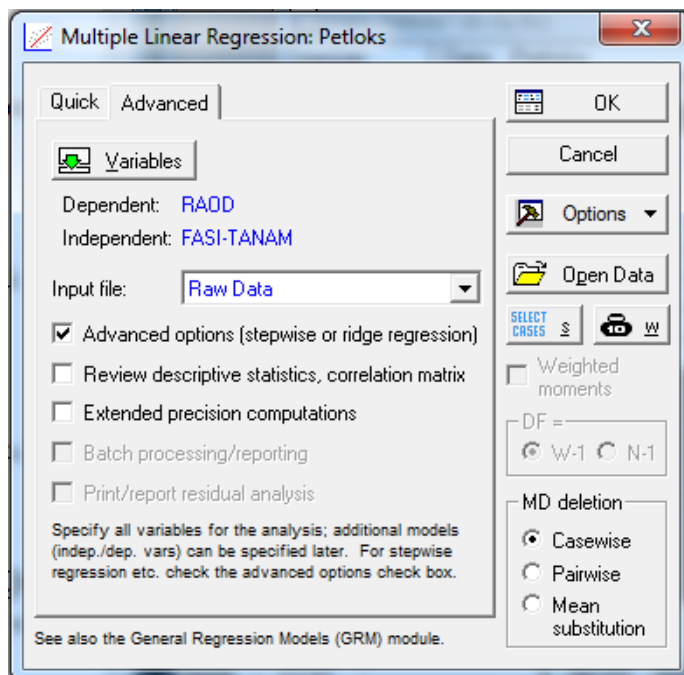
*Statistics* მენიუში შევარჩიოთ **Multiple Regression** (მრავლობითი რეგრესია) და ეკრანზე გამოსულ **Multiple Linear Regression** სასტარტო ფანჯარაში **click Variables** ღილაკზე.



– გაიხსნება *Select dependent and independent variable lists* (სიიდან შევარჩიოთ დამოკიდებული და დამოუკიდებელი ცვლადები) ფანჯარა. ფანჯარის მარცხენა მხარეს მოვნიშნოთ დამოკიდებული ცვლადი (მაგალითად, *RAOD*), ხოლო მარჯვენა მხარეს დამოუკიდებელი ცვლადები (*FASI*, *REKL*, *TANAM*). შემდეგ **OK**.



– ვბრუნდებით სასტარტო ფანჯარაში. ჩავრთოთ *Advanced* (დამატებითი) ჩანართი, სადაც წარმოდგენილია შემდეგი ოპერაციები:



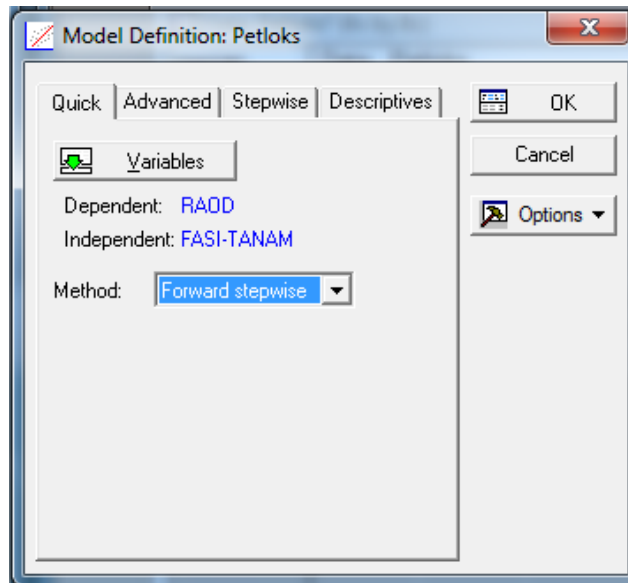
- *Advanced options (stepwise or ridge regression)* – ბიჯური რეგრესია.
- *Review descriptive statistics, correlation matrix* – აღწერითი სტატისტიკის და კორელაციური მატრიცის დათვალიერება.
- *Extended precision computations* – გამოთვლის სიზუსტის გაზრდა.
- *Batch processing/reporting* – მონაცემების პაკეტური დამუშავება/ბეჭვდა.

---

ე.ყუბანეიშვილი. მონაცემების სტატისტიკური დამუშავება

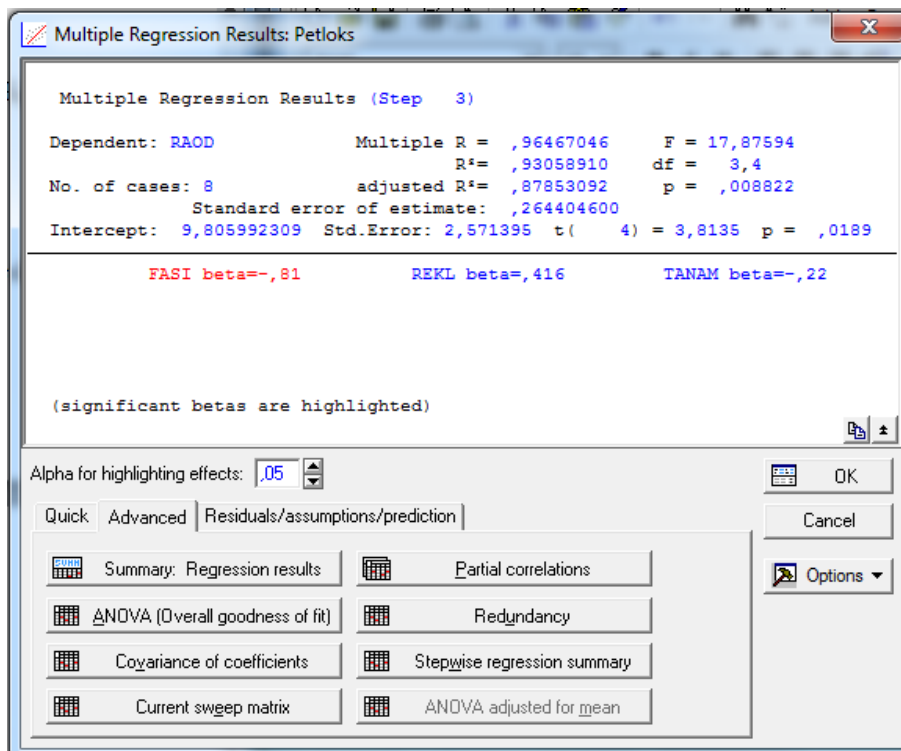
მოვნიშნოთ **Advanced options** და შეძღვოთ **OK**.

– ეკრანზე გამოდის **Model Definition** (მოდელის აგება) ფანჯარა. **Quick** ჩანართის **Method** ველში წარმოდგენილია შემდეგი მეთოდების ჩამონათვალი:



- **Standard** (სტანდარტული);
- **Forward stepwise** (ბიჯური ჩართვით);
- **Backward stepwise** (ბიჯური გამორიცხვით).

შეგარჩიოთ **Forward stepwise** მეთოდი. შეძღვოთ **OK**.



– ეკრანზე გამოდის **Multiple Regression Results** შედეგების ფანჯარა, სადაც ზედა სარკმელში წარმოდგენილია შედეგების შეფასების ძირითადი

ეკუბანეიშვილი. მონაცემების სტატისტიკური დამუშავება

ინფორმაცია, მეორე სარკმელში გამოტანილია რეგრესიის სტანდარტიზირებული კოეფიციენტები. ფანჯრის ქვედა ნაწილში განლაგებულია ფუნქციონალური დილაკები, რომლებიც საშუალებას იძლევიან ყოველმხრივ დავათვალიეროთ ანალიზის შედეგები.

ფანჯრის ინფორმაციულ ნაწილში წარმოდგენილია შედეგების მოკლე ინფორმაცია:

- **Dependet** (დამოკიდებული ცვლადი სახელი);
- **No.ofcases** (დაკვირვებათა რიცხვი);
- **Multiple R** = (მრავლობითი კორელაციის კოეფიციენტი);
- **R-sqaare R<sup>2</sup>** = (დეტერმინაციის კოეფიციენტი);
- **Adjusted R<sup>2</sup>** = (კორექტირებული დეტერმინაციის კოეფიციენტი);
- **Standard error of estimate** = (სტანდარტული ცდომილება);
- **Intercept** (რეგრესიის  $b_0$  თავისუფალი წევრის შეფასება);
- **Std. error** = ( $b_0$  თავისუფალი წევრის სტანდარტული შეცდომა);
- **t = , p = ,** ( $t$  კრიტერიუმის და  $p$  მნიშვნელობის სიდიდეები

$H_0 : b_0 = 0$  ნულოვანი ჰიპოთეზის შესამოწმებლად);

• **F = , df = , p =** ( $F$  კრიტერიუმის მნიშვნელობა,  $df$  - თავისუფალი ხარისხი,  $p$  - მნიშვნელობის სიდიდე). გამოიყენებიან რეგრესიის განტოლების ადეკვატურობის დასადგენად ანუ დამოკიდებული და დამოუკიდებელი ცვლადების დამოკიდებულების ჰიპოთეზის შესამოწმებლად.

მიღებული შედეგებიდან ჩანს, რომ დამოკიდებული და დამოუკიდებელი ცვლადებს შორის კავშირი ძლიერია ( $R^2 > 0,76$ ), ე.ი. რეგრესიის განტოლება კარგად აღწერს ცვლადების შორის ურთიერთკავშირებს.

– **Advanced** ჩანართში **click Summeery: regression results** დილაკზე. ეკრანზე გამოდის შედეგების ცხრილი, სადაც სტატისტიკები დეტალურად არიან წარმოდგენილნი.

Regression Summary for Dependent Variable: RAOD (Petloks)						
R= ,96467046 R <sup>2</sup> = ,93058910 Adjusted R <sup>2</sup> = ,87853092						
F(3,4)=17,876 p<,00882 Std.Error of estimate: ,26440						
N=8	Beta	Std.Err. of Beta	B	Std.Err. of B	t(4)	p-level
<b>Intercept</b>			9,80599	2,571395	3,81349	0,018883
FASI	-0,812418	0,228380	-5,95435	1,673837	-3,55730	0,023643
REKL	0,415881	0,164709	0,18270	0,072358	2,52494	0,065014
TANAM	-0,223980	0,202949	-0,03900	0,035339	-1,10362	0,331683

ცხრილი შეიცავს სტანდარტიზირებულ (**Beta**) და არასტანდარტიზირებულ (**B**) რეგრესიის კოეფიციენტებს, მათ სტანდარტულ შეცდომებს და მნიშვნელობების დონეებს. **Beta** კოეფიციენტებს განჩიათ ნულოვანი საშუალო სიდიდე, ხოლო სტანდარტული გადახრა 1-ის ტოლია. ეს საშუალებას გვაძლევს შევაფასოთ თითოეული დამოუკიდებელი ცვლადის წვლილი რეგრესიის მოდელის აგებაში. ასე მადალითად, ჩვენს შემთხვევაში ყველაზე დიდი წვლილი შეაქვს **FAST** ცვლადს, ხოლო ყველაზე მცირე - **TANAM**. კოეფიციენტების უარყოფითი მნიშვნელობა ამ მაგალითში იმას ნიშნავს, რომ ფასის ზრდასთან ერთად და მომსახურე პერსონალის რაოდენობის გაზრდა იწვევს გაყიდვის რაოდენობის შემცირებას. **FAST**



ცვლადს დადებითი ნიშანი იმას ნიშნავს, რომ რეკლამის ხარჯის ზრდასთან ერთად იზრდება გაყიდვების რაოდენობაც.

რეგრესიის განტოლების  $b_1, b_2$  და  $b_0$  კოეფიციენტები სარწმუნოანი არიან  $p = 0.1$  მნიშვნელობის დონით, ხოლო  $b_3$  კოეფიციენტი სტატისტიკურად არ არის სარწმუნო, რადგან  $p > 0.1$ .

– **click Partial Correlation** (კერძო კორელაცია) დილაკზე, ეკრანზე გამოდის ცხრილი რომელიც შეიცავს **Beta** კოეფიციენტებს, კორელაციის კერძო კოეფიციენტებს, ნახევრად კერძო კოეფიციენტებს (**semipartcor**), ტოლერანტობის (**tolerance**), დეტერმინაციის კოეფიციენტებს, (**R-dsquare**), **t(4)** - სტეიდენტის **t** კრიტერიუმის მნიშვნელობას კერძო კორელაციური კოეფიციენტების სარწმუნოების შესამოწმებლად და **p** -მნიშვნელობას.

Variable	Variables currently in the Equation; DV: RAOD (Petloks)					
	Partial Cor.	Semipart Cor.	Tolerance	R-square	t(4)	p-level
FASI	-0,871678	-0,468603	0,332698	0,667302	-3,55730	0,023643
REKL	0,783882	0,332610	0,639637	0,360363	2,52494	0,065014
TANAM	-0,483136	-0,145380	0,421301	0,578699	-1,10362	0,331683

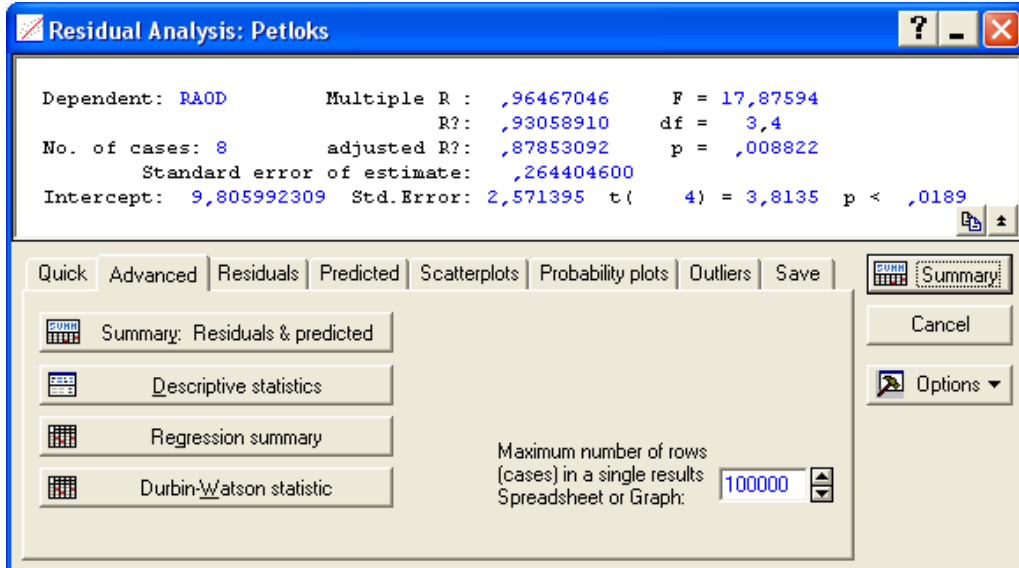
კორელაციის კერძო კოეფიციენტები, ისევე როგორც სტანდარტიზირებული **Beta** კოეფიციენტები, საშუალებას იძლევიან დამოუკიდებელი ცვლადების რანჟირებისა მათი დამოკიდებულ ცვლადთან ზეგავლენის ზომის მიხედვით. გარდა ამისა, კორელაციის კერძო კოეფიციენტები ფართოდ გამოიყენებიან დამოუკიდებელი ცვლადების შესარჩევად, კერძოდ რეგრესიის განტოლებაში მათი ჩართვის ან არჩართვისთვის. ცხრილიდან ჩანს, რომ დამოუკიდებელი ცვლადების რანჟირება, მათი დამოკიდებელი ცვლადთან ზეგავლენის ზომის მიხედვით, შემდეგია: **FASI**, **REKL**, **TANAM**, თანაც პირველი ორი ცვლადი იწვევს დამოკიდებელი ცვლადთან ძლიერ ზემოქმედებას, ხოლო მესამე ცვლადი – ზომიერ ზემოქმედებას.

**Semipart Cor** (ნახევრად კერძო კორელაცია) წარმოადგენს დამოკიდებელი და დამოუკიდებელ ცვლადებს შორის კორელაციის კოეფიციენტს, იმის გათვალისწინებით, რომ კონტროლირდება სხვა დამოუკიდებელი ცვლადების ზემოქმედება ამ დამოკიდებულ ცვლადზე და არ კონტროლირდება დამოუკიდებელი ცვლადების ზეგავლენა დამოკიდებულ ცვლადზე. თუ ნახევრად კერძო კორელაცია მცირეა, ხოლო კერძო კორელაცია შედარებით დიდია, მაშინ შესაბამის დამოუკიდებულ ცვლადს შეიძლება გააჩნდეს თავისი „ნაწილი“ დამოკიდებელი ცვლადის ცვალებადობაში, ანუ „ნაწილი“, რომელიც არ აიხსნება სხვა დამოუკიდებელი ცვლადებით. როგორც ცხრილიდან ჩანს, **FASI** და **REKL** ცვლადებს დამოკიდებელი ცვლადის ცვალებადობაში გაჩნიათ საკუთარი წილები.

ცხრილიდან ჩანს, რომ დეტერმინაციის ყველა კოეფიციენტი ზომიერი სიდიდისაა, მაგრამ **REKL** ურთიერთკავშირი დანარჩენ ორ დამოუკიდებელ ცვლადთან გაცილებით ნაკლებია, ვიდრე **FASI** და **TANAM** ცვლადებისა სხვა დანარჩენ ორ დამოუკიდებელ ცვლადთან.

ტოლერანტობა ტოლია **1 - R<sup>2</sup> (R-square)**.

რეგრესიული ანალიზის მნიშვნელოვან მახასიათებელს წარმოადგენს **Residuals** (ნაშთები). ამისათვის ჩაერთოთ **Residuals/assumptions/prediction** ჩანართი და **click Perform residuals analysis** ღილაკზე. ეკრანზე გამოდის **residuals analysis** (ნაშთების ანალიზი)-ის ფანჯარა.



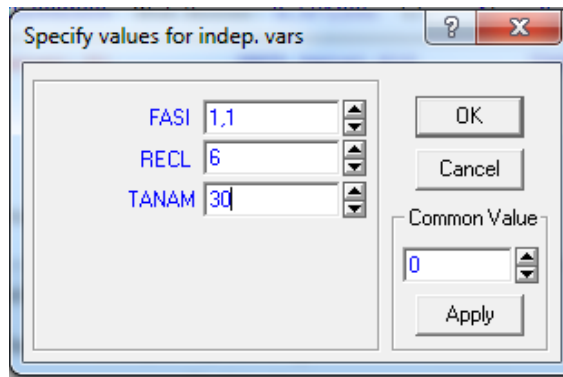
**click Advanced** ჩანართის **Durbin-Watson statistic** (დარბინ-უიტსონის სტატისტიკა)-ის ღილაკზე. ეკრანზე გამოდის ცხრილი, სადაც ნახვენებია დარბინ-უიტსონის კრიტერიუმის მნიშვნელობა და სერიალური ავტოკორელაციის მნიშვნელობა.

Durbin-Watson d (Petloks) and serial correlation of residuals		
	Durbin-Watson d	Serial Corr.
Estimate	1.920437	-0.250565

რადგან დარბინ-უიტსონის სტატისტიკას გააჩნია მცირე მნიშვნელობა (1,9204) და სერიალური ავტოკორელაციის მნიშვნელობა (-0,2505) ზომიერი სიდიდისაა, ამიტომ შეიძლება ითქვას, რომ ცვლადებს შორის სუსტი კორელაცია შეიმჩნევა, რაც მიგვანიშნებს იმაზე, რომ რეგრესიულ მოდელს გააჩნია ადეკვატურობის არც ისე დიდი სიდიდე.

რეგრესიის განტოლება შესაძლებელია გამოყენებული იყოს დამოკიდებული ცვლადის პროგნოზირებისთვის. ამისთვის უნდა დავბრუნდეთ **Multiple Regressior Results** ფანჯარაში, სადაც უნდა გავხსნათ **Residuds/assumption/prediction** (ნარჩენები/შეფასება/პროგნოზირება) და **click Predict dependent variable** (დამოკიდებული ცვლადის პროგნოზირება) ღილაკზე.

– ეკრანზე გამოდის **Specify values for indep. vars** ფანჯარა, სადაც **FACI, REKL TANAM** ცვლადების ველში მიუთითოდ ფასი (1,1), რეკლამაზე ხარჯი (6) და თანამშრომელთა რაოდენობა (30).

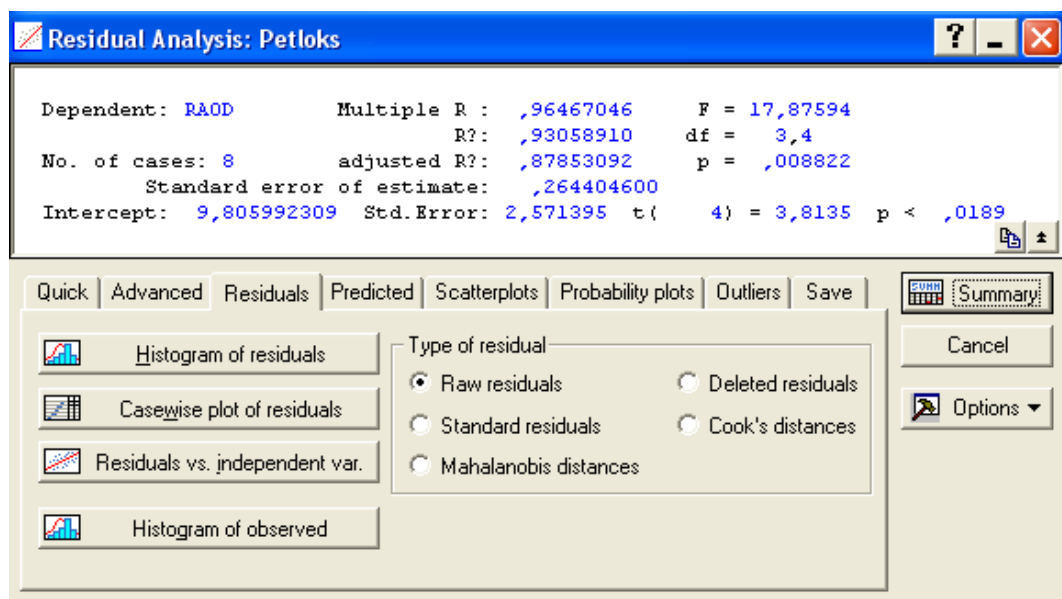


Click OK. ეკრანზე გამოდის პროგნოზის შედეგების ცხრილი.

Predicting Values for (Petloks) variable: RAOD			
Variable	B-Weight	Value	B-Weight * Value
FASI	-5,95435	1,10000	-6,54978
RECL	0,18270	6,00000	1,09621
TANAM	-0,03900	30,00000	-1,17002
Intercept			9,80599
Predicted			3,18240
-95,0%CL			2,21760
+95,0%CL			4,14719

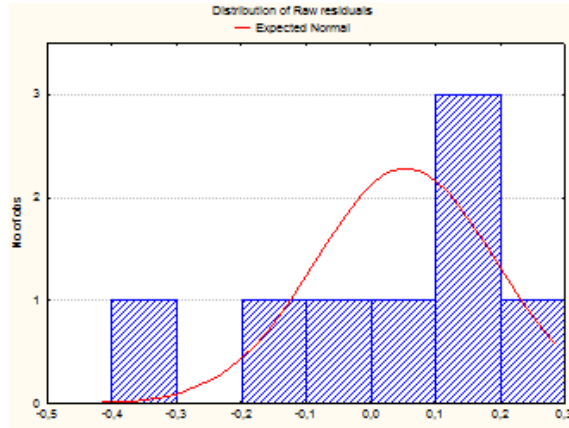
ცხრილიში ნაჩვენებია გაყიდვის პროგნოზი (*predicted*) – 3.1824 (მლ. დოლარი) 95%-ინი ნდობის ინტერვალით (2,2176; 4,1471).

რეგრესიული ანალიზის კორექტური გამოყენებისთვის ერთ-ერთ პირობას წარმოადგენს, ნაშთების (ნარჩენების) განაწილება შეესაბამება თუ არა ნორმალურს. ამისთვის **Multiple Regression Result** ფანჯარაში ჩავრთოთ **Residuals/assumptions/prediction** ჩანართ და click **Perform Residual analysis** ღილაკზე. ეკრანზე გამოდის **Residual Analysis** ფანჯარა.

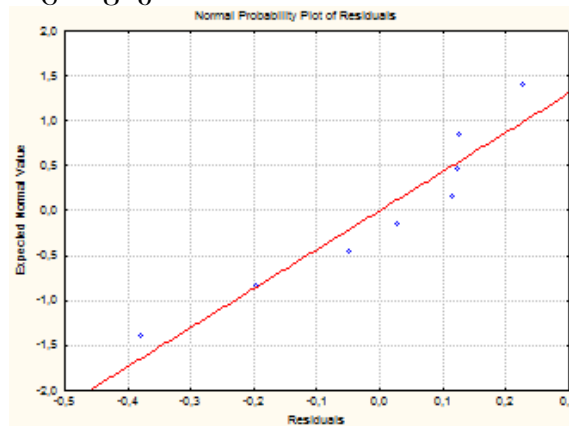


ე.ყუბანეიშვილი. მონაცემების სტატისტიკური დამუშავება

გავხსნათ **Residuals** ჩანართი და *click Histogram of residuals* ღლაკზე. ეკრანზე გამოდის ნაშთების ჰისტოგრამა და ნორმალური განაწილების სიმკვრივის ფუნქცია.



გრაფიკზე კარგად ჩანს, რომ დაკვირვებათა მცირე რაოდენობის (8) გამო ნარჩენების განაწილება არ შესაბამეობა ნორმალურს. იგივე ამოცანის გადასაწყვეტად შესაძლებელია გამოყენებული იყოს ნორმალური ალბათური გრაფიკი. ამისათვის **Residual Analysis** ფანჯარაში გავხსნათ **Probability** ჩანართი და *click Normal plot of residuals* ღლაკზე. ეკრანზე გამოდის ნორმალური ალბათობის გრაფიკი.



ჩატარებული წრფივი რეგრესიის ანალიზის საფუძველზე მივიღეთ შემდეგი რეგრესიის განტოლება

$$RAOD = 9.8059 - 5.9543 \cdot FASI + 0.1827 \cdot REKL - 0.039 \cdot TANAM$$

რადგან განტოლების ბოლო წევრის კოეფიციენტი 0,039 სტატისტიკურად არ არის სარწმუნო, ამიტომ მისი გამორიცხვა რეგრესიის განტოლებიდან შესაძლებელია.

პრაქტიკული სამუშაო 4

არაწრფივი რეგრესიული ანალიზი

*Statistica* პროგრამაში რეალიზირებული არაწრფივი რეგრესიული ანალიზის ორი მოდელი: **Fixed Nonlinear Regression** (ფიქსირებული არაწრფივი რეგრესია)-ის მოდელი, სადაც შესაძლებელია გაწრფივების პროცესის ჩატარება და **Nonlinear Estimation** (არაწრფივი შეფასება)-ის მოდელი სადაც გაწრფივება შეუძლებელია. განვიხილოთ **Fixed Nonlinear Regression** მოდელი.

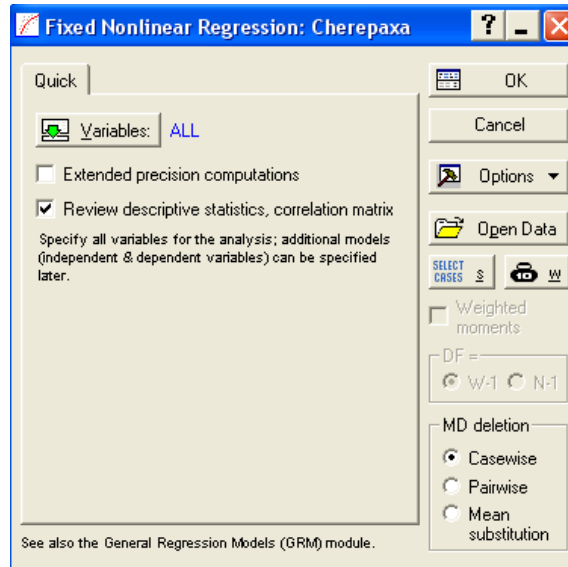
მოდელის მუშაობა განვიხილოთ **Cherepaxa.sta** მონაცემთა ფაილზე, სადაც წარმოდგენილია 20 კუს ზომები (**LENGTH**–სიგრძე, **WIDTH**–სიგანე, **HEIGHT** –სიმაღლე, **WEIGHT** –წონა).

Data: <i>Cherepaxa</i>				
	1	2	3	4
	LENGTH	WIDTH	HEIGHT	WEIGHT
1	98	81	38	155
2	103	84	38	160
3	103	86	42	164
4	105	86	42	165
5	109	88	44	170
6	123	92	50	180
7	123	95	46	180
8	133	99	51	190
9	133	102	51	193
10	133	102	51	190
11	134	100	48	190
12	136	102	49	192
13	138	98	51	192
14	141	99	51	193
15	147	105	53	200
16	149	108	57	210
17	153	107	55	206
18	149	107	56	210
19	155	115	63	220
20	155	117	60	220

განვიხილოთ კუს წონის კვადრატული დამოკიდებულება სიგრძეზე, სიგანეზე და სიმაღლეზე

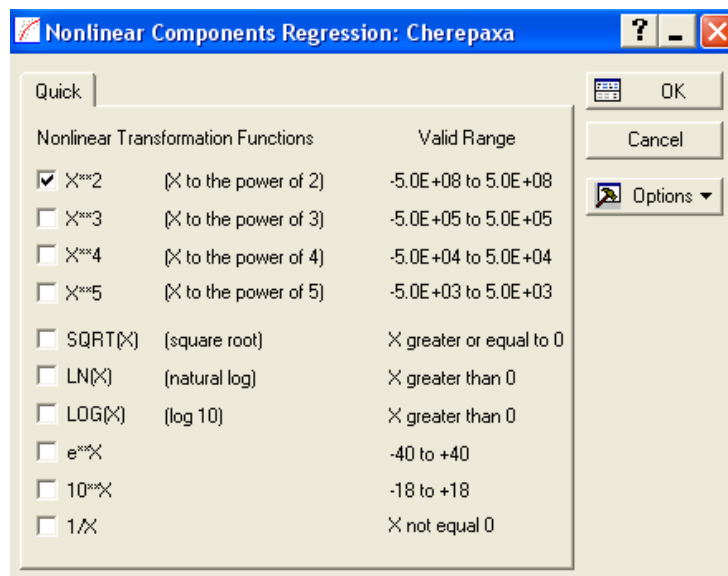
$$Weight = b_0 + b_1 + b_2 width + b_3 height + b_4 (length)^2 + b_5 (width)^2 + b_6 (height)^2.$$

*Statistica* მენიუში შევარჩიოთ **Advanced Linear/Nonlinear Modest** და **click Fixed Nonlinear Regression** ბრძანებაზე.

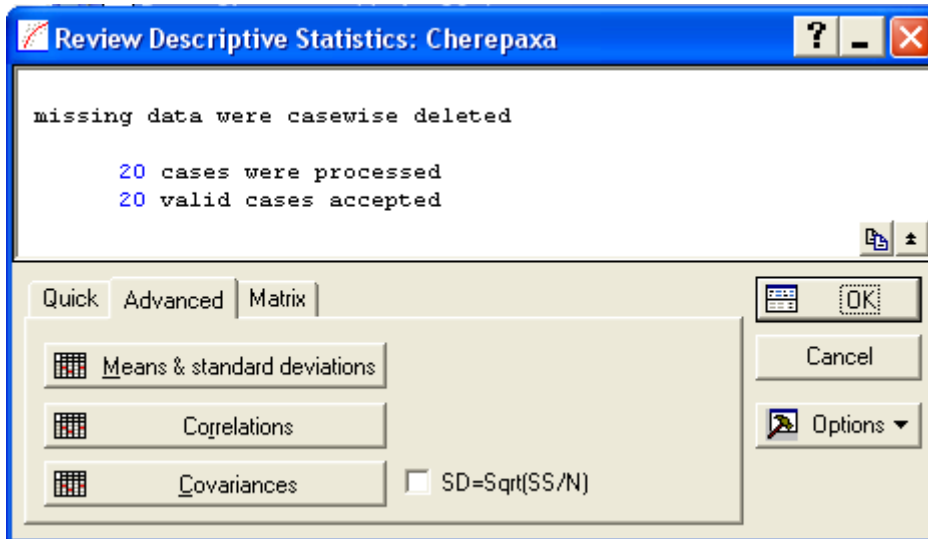


– გაიხსნება ამავე დასახელების ფანჯარა, *click Variables* ღილაკზე. შევარჩიოთ ყველა ცვლადი. შემდეგ *OK*.

– ჩაერთოთ ოპცია *Review descriptive statistics, correlation matrix* (ვაჩვენოთ აღწერით სტატისტიკა, კორელაციული მატრიცა). (ოპცია *Extended precision computations* ნიშნავს გაზრდილი სიზუსტით გამოთვლის რეჟიმს). შემდეგ *OK*.



– ეკრანზე გამოდის *Nonlinear Components Regression* ფანჯარა, რომელიც ითვალისწინებს ცვლადების გაწრფივების გარდაქმნას. *Nonlinear transformation functions* სვეტში მოყვანილია შესაბამისი გარდაქმნები, ხოლო *Valid Range* სვეტში ნაჩვენებია ცვლადების დასაშვები დიაპაზონი. რადგან ჩვენ ავირჩიეთ კვადრატული მოდული, ამიტომ ჩაერთოთ *X<sup>2</sup>* ოპცია. შემდეგ *OK*.



– გაიხსება *Review Descriptive Statistics* ფანჯარა. ჩაერთოდ *Advanced* ჩანართი და *click Correlations* ღილაკზე, ეკრანზე გამოდის კორელაციური მატრიცა, სადაც წინასწარ, მრავლობითი რეგრესიის მოდელის ჩართვამდე, შესაძლებელია დავათვალიეროთ პრედიქტორების (დამოუკიდებელი ცვლადები) ( $length, width, height, length^2, width^2, height^2$ ) და დამოკიდებული  $weight$  ცვლადის კორელაციები

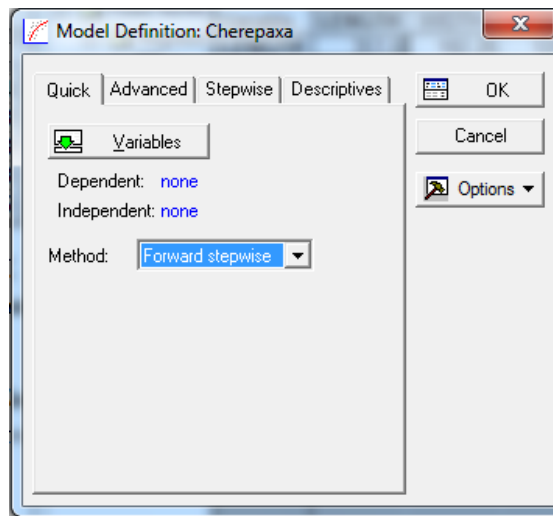
Variable	Correlations (Cherepaxa)							
	LENGTH	WIDTH	HEIGHT	WEIGHT	V1**2	V2**2	V3**2	V4**2
LENGTH	1,000000	0,965828	0,950057	0,981373	0,998101	0,956053	0,935917	0,973555
WIDTH	0,965828	1,000000	0,962186	0,989023	0,968673	0,998475	0,958296	0,988084
HEIGHT	0,950057	0,962186	1,000000	0,980127	0,953293	0,960685	0,996719	0,979675
WEIGHT	0,981373	0,989023	0,980127	1,000000	0,984571	0,986050	0,975006	0,998602
V1**2	0,998101	0,968673	0,953293	0,984571	1,000000	0,961899	0,943373	0,979778
V2**2	0,956053	0,998475	0,960685	0,986050	0,961899	1,000000	0,960924	0,987960
V3**2	0,935917	0,958296	0,996719	0,975006	0,943373	0,960924	1,000000	0,978637
V4**2	0,973555	0,988084	0,979675	0,998602	0,979778	0,987960	0,978637	1,000000

*Means & standard deviations* და *Covariances* ღილაკით შეგვიძლია დავათვალიეროთ ცვლადების საშუალოები, სასშუალო კვადრატული გადრახრები და კოვარიაციული მატრიცა.

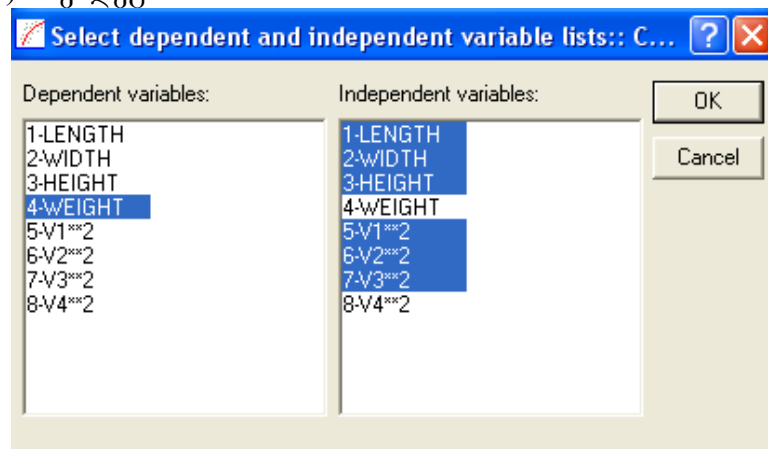
Variable	Means and Standard Deviations (Cherepaxa)		
	Means	Std.Dev.	N
LENGTH	131,00	18,750	20
WIDTH	98,65	10,064	20
HEIGHT	49,80	6,748	20
WEIGHT	189,00	19,178	20
V1**2	17495,00	4771,464	20
V2**2	9828,05	1981,724	20
V3**2	2523,30	671,689	20
V4**2	36070,40	7212,854	20

Variable	Covariances (Cherepaxa)							
	LENGTH	WIDTH	HEIGHT	WEIGHT	V1**2	V2**2	V3**2	V4**2
LENGTH	351,6	182,26	120,21	352,9	89297	35525	11787	131668
WIDTH	182,3	101,29	65,35	190,9	46518	19914	6478	71728
HEIGHT	120,2	65,35	45,54	126,8	30694	12847	4518	47684
WEIGHT	352,9	190,89	126,84	367,8	90094	37475	12560	138134
V1**2	89297,2	46517,53	30694,42	90094,5	22766872	9095449	3023452	33719913
V2**2	35525,2	19914,44	12847,12	37475,0	9095449	3927229	1279086	14121788
V3**2	11787,4	6478,22	4517,75	12559,6	3023452	1279086	451165	4741293
V4**2	131667,6	71728,04	47683,77	138133,6	33719913	14121788	4741293	52025265

– დავბრუნდეთ *Review Descriptive Statistics*-ის ფანჯარაში და *click OK*. გაიხსენბა *Model definition* ფანჯარა, სადაც *Method* არეში შევარჩიოთ *Forward stepwise* (ბიჯური ჩართვა)-ის მეთოდი.

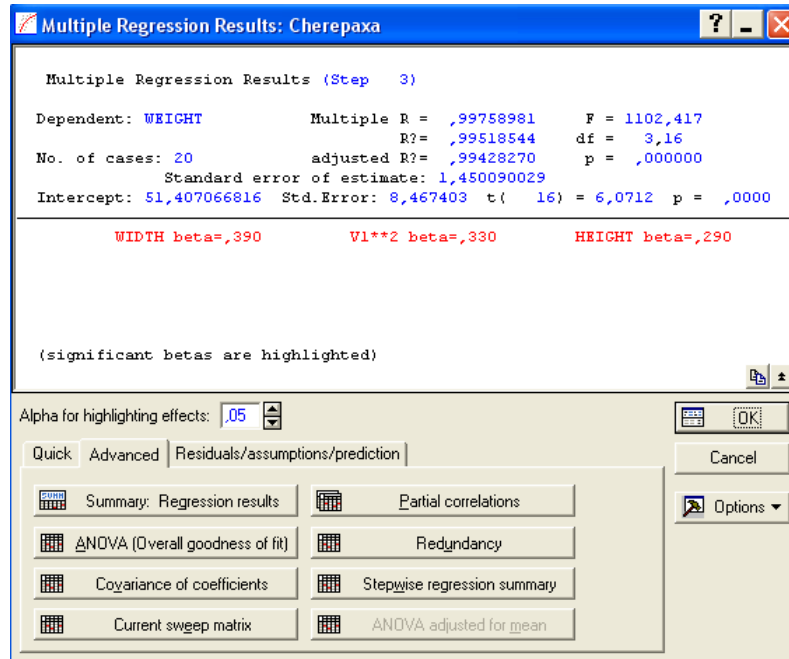


– *click Variables* ღილაკზე. გაიხსენბა *Select dependent and independent variable lists* ფანჯარა სადაც *Dependent variables* ველში მოვნიშნოთ დამოკიდებული ცვლადი *Weight*, ხოლო *Independent variables* ველში დამოუკიდებელი ცვლადი. (ჩვენი მაგალითის შემთხვევაში *Lenght, Width, Height, V1\*\*2, V2\*\*2*). შემდეგ *OK*.





– ვბრუნდებით **Model Definition** ფანჯარაში. *click OK*. ეკრანზე გამოდის **Multiple Regression Results** ფანჯარა. *click Quick* ჩანართის **Summary: regression results** ლილაკზე. ეკრანზე გამოდის შედეგების ცხრილი, სადაც წარმოდგენილია სტანდარტიზირებული (**Beta**) და არასტანდარტული (**B**) რეგრესიის კოეფიციენტები, მათი სტანდარტული შეცდომები და მნიშვნელოვნების დონეები. როგორც ცხრილიდან ჩანს მრავლობით კორელაციის და დეტერმინაციის კოეფიციენტებმა მიიღეს ერთთან მიახლოებული მნიშვნელობები (0,9968; 0,9935), რაც მიგვანიშნებს რეგრესიის განტოლების ადეკვატურობაზე.

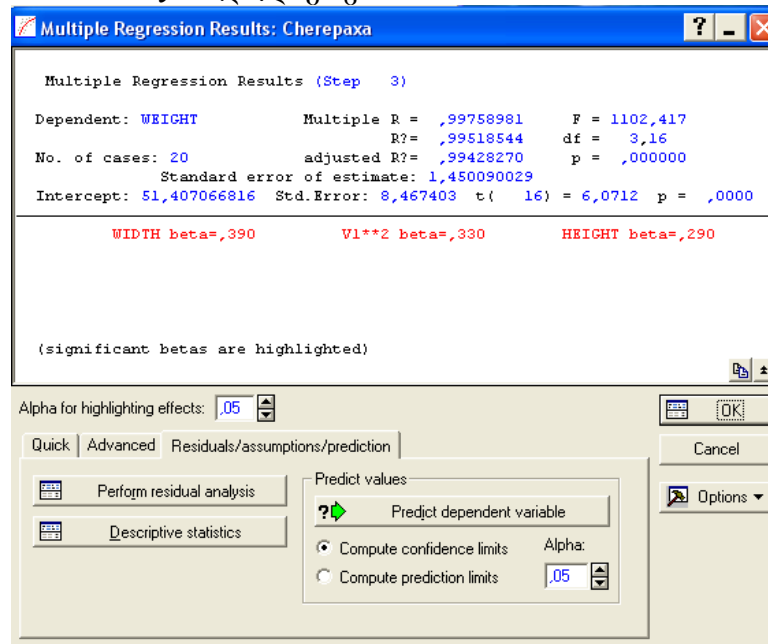


Regression Summary for Dependent Variable: WEIGHT (Cherepaxa)						
R = .99758981 R <sup>2</sup> = .99518544 Adjusted R <sup>2</sup> = .99428270						
F(3,16) = 1102.4 p < .00000 Std. Error of estimate: 1.4501						
N=20	Beta	Std. Err. of Beta	B	Std. Err. of B	t(16)	p-level
Intercept			51.40707	8.467403	6.071173	0.000016
WIDTH	0.390382	0.081585	0.74388	0.155460	4.784995	0.000202
V1**2	0.330428	0.073575	0.00133	0.000296	4.491027	0.000370
HEIGHT	0.289513	0.067078	0.82278	0.190633	4.316068	0.000533

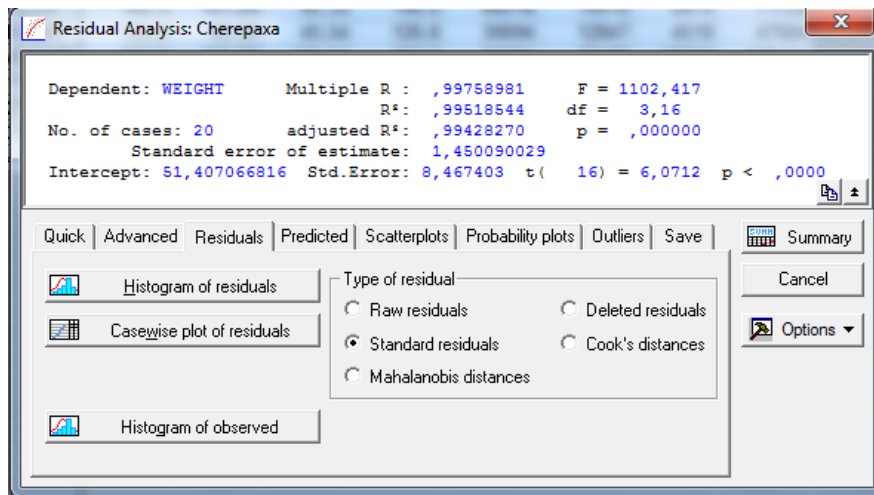
Variables currently in the Equation; DV: WEIGHT (Cherepaxa)							
Variable	Beta in	Partial Cor.	Semipart Cor.	Tolerance	R-square	t(16)	p-level
WIDTH	0.390382	0.767234	0.083004	0.045209	0.954791	4.784995	0.000202
V1**2	0.330428	0.746750	0.077905	0.055587	0.944413	4.491027	0.000370
HEIGHT	0.289513	0.733452	0.074870	0.066877	0.933123	4.316068	0.000533

ბიჯურმა რეგრესიამ არწრფივ რეგრესიულ განტოლებაში ჩართო მხოლოდ **Length<sup>2</sup>**, **Height<sup>2</sup>** და **Width** ცვლადი.

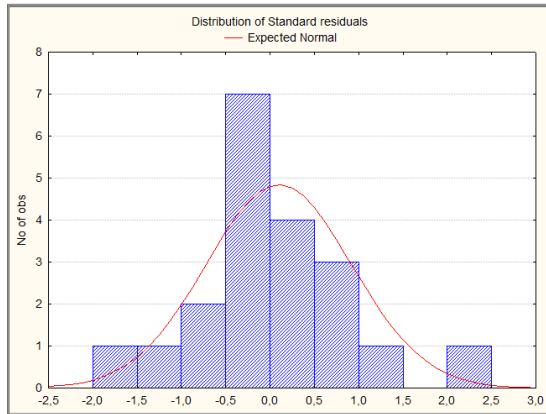
რეგრესიის მოდელის ადეკვატურობისთვის დავბრუნდეთ **Multiple Regression Results** ფანჯარაში გავხსნათ **Residuals/assumptions/prediction** ჩანართი და *click* **Perform residuals analysis** ღილაკზე.



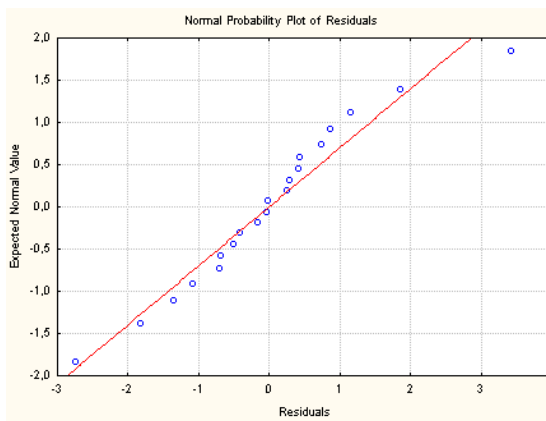
– ეკრანზე გახსნილ **Residual Analysis** ფანჯარაში ჩავრთოდ **Residuals** ჩანართს და **Type of residual** ოპციებში შევარჩით **Standart residuals**. შემდეგ *click* **Histogram of residuals**.



ეკრანზე გამოდის გრაფიკი შესაბამისი ნაშთების ჰისტოგრამით და განაწილების სიმკვრივის ფუნქციით.



როგორც ამ გრაფიკიდან ჩანს, მცირე დაკვირვების (20) გამო განაწილება დაახლოებით შეიძლება ჩავთვალოთ ნორმალურად. ნორმალური ალბათური გრაფიკის გამოყენებისას **Residual Analysis** ფანჯარაში გავსსნათ **Probability** ჩანართი და **click Normal plot of residuals** ღილაკზე. ეკრანზე გამოდის ნორმალური ალბათობის გრაფიკი.



ამრიგად რეგრესიის განტოლება ღებულობს შემდეგ სახეს:  
 $Y = b_0 + b_1V1^2 + b_2V2^2$  ანუ  $weight = 107,087 + 0,0024 \cdot length^2 + 0,0042 \cdot width^2$ .

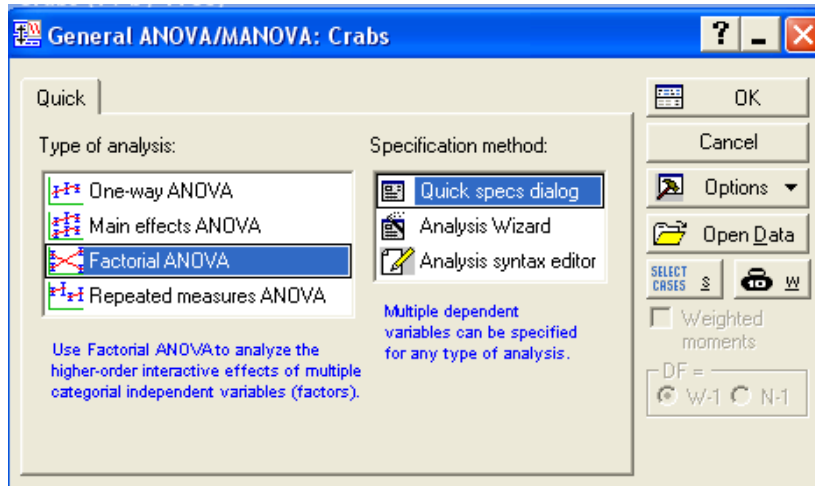
პრაქტიკული სამუშაო 6

დისპერსიული ანალიზი

დისპერსიული ანალიზის პროცედურის გაცნობისათვის გამოვიყენოთ *Examples* ბიბლიოთეკიდან შერჩეული *Crabs.sta* (კიბორჩხალები) მონაცემთა ფაილი, სადაც მოყვანილია კიბორჩხალების ფერი (*COLOR*), თანამგზავრის რაოდენობა (*SATELLTS*) – მამრობითი და მდედრობითი სქესის ინდივიდების მათი ფერიდან გამომდინარე, მარწუხის (*SPINE*) მდგომარეობა, ზომები (*CATWIDTH*), სიგანე (*WIDTH*), წონა (*WEGHT*).

Number of crab satellites by female's color, spine condition, width, and weight							
1	2	3	4	5	6	7	
Y	COLOR	SPINE	WIDTH	SATELLTS	WEIGHT	CATWIDTH	
1	1	medium	bothworn	28,3	8	3,05	28,75
2	0	darkmed	bothworn	22,5	0	1,55	22,75
3	1	lightmed	bothgood	26,0	9	2,30	25,75
4	0	darkmed	bothworn	24,8	0	2,10	24,75
5	1	darkmed	bothworn	26,0	4	2,60	25,75
6	0	medium	bothworn	23,8	0	2,10	23,75
7	0	lightmed	bothgood	26,5	0	2,35	26,75
8	0	darkmed	oneworn	24,7	0	1,90	24,75
9	0	medium	bothgood	23,7	0	1,95	23,75
10	0	darkmed	bothworn	25,6	0	2,15	25,75
11	0	darkmed	bothworn	24,3	0	2,15	24,75
12	0	medium	bothworn	25,8	0	2,65	25,75
13	1	medium	bothworn	28,2	11	3,05	27,75
14	0	dark	oneworn	21,0	0	1,85	22,75
15	1	medium	bothgood	26,0	14	2,30	25,75
16	1	lightmed	bothgood	27,1	8	2,95	26,75
17	1	medium	bothworn	25,2	1	2,00	24,75
18	1	medium	bothworn	29,0	1	3,00	28,75
19	0	dark	bothworn	24,7	0	2,20	24,75
20	1	medium	bothworn	27,4	5	2,70	27,75

გავხსნათ *Statistics* მენიუ და მოვნიშნოთ *ANOVA*. ეკრანზე გამოდის *General ANOVA/MANOVA* ფაჯარა, რომელსაც გაჩნია ორი სია *Type of analysis* (ანალიზის სახე) და *Specification method* (მეთოდების სპეციფიკაცია).



**Type of analysis** სია შედგება ოთხი ელემენტისგან, რომლებიც წარმოადგენენ დისპერსიული ანალიზის სხვადასხვა მოდელს:

- **One-way ANOVA** (ერთფაქტორიანი დისპერსიული ანალიზი);
- **Main effects ANOVA** (მთავარი ფაქტორების დისპერსიული ანალიზი);
- **Factorial ANOVA** (მარვალფაქტორიანი დისპერსიული ანალიზი);
- **Repeated measures ANOVA** (გამეორებითი გაზომვების დისპერსიული ანალიზი).

**Specification method** სიაში შეგვიძლია შევარჩოთ დისპერსიული ანალიზის ინტერფეისის სამი ტიპი:

- **Quick specs dialog** (დიალოგის სწრაფი სპეციფიკაციები);
- **Analysis Wizard** (ანალიზის ოსტატი);
- **Analysis syntax editor** (კოდის რედაქტორი).

**Quick specs dialog** დიალოგში შეგვიძლია დავადგინოთ დამოკიდებული და კატეგორიული (პრედიქტორები) ცვლადები. რიცხვების და ცვლადების ტიპის ვარიაცია დამოკიდებულია **Type of analysis**-ის სიაში შერჩეულ ანალიზის ტიპზე.

**Analysis Wizard** დიალოგი განკუთვნილია შერჩეულ მოდელში დავალებების ბიჯურად წარმოდგენისათვის. ანალიზის ბოლოს შესაძლებელია შედეგების გამოთვლა ან გამოყენებით **Analysis syntax editor** შემდგომი რეგულირებისთვის. **Analysis syntax editor** დიალოგი საშუალებას იძლევა მთლიანად დავარეგულიროთ როგორც გეგმის პარამეტრები, ასევე გამოთვლითი პროცედურების პარამეტრები. საჭიროების შემთხვევაში შესაძლებელი ფაილის შენახვა შემდგომი მისი გამოყენებისათვის.

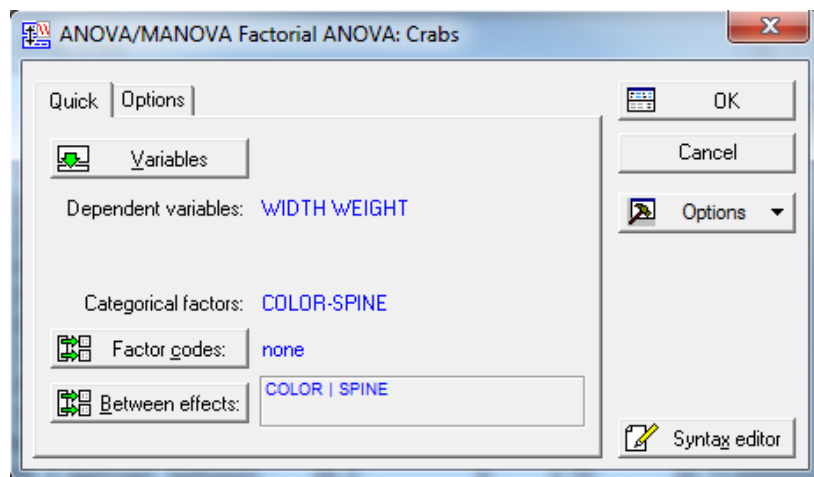
**Specification method** შერჩევის შემდეგ, შეგვიძლია შევარჩოთ **Type of analysis** ანალიზის ტიპი.

**One-way ANOVA** მეთოდი საშუალებას იძლევა შევადაროთ ერთი დამაჯგუფებელი ცვლადის ეფექტი ერთ ან რამოდენიმე დამოკიდებულ ცვლადზე.

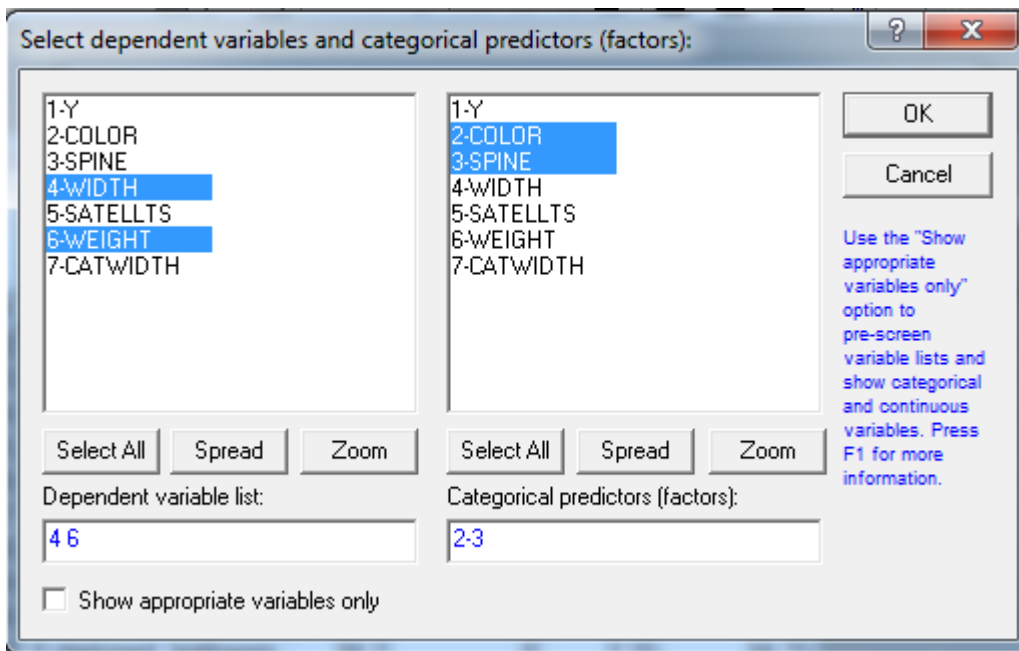
**Main effect ANOVA** ანალიზისათვის **Quick specs dialog** დიალოგში შეიძლება მოცემული იყოს ოთხამდე კატეგორიული (დამაჯგუფებელი) პრედექტორები. შემდეგ პროგრამა აწარმოებს მთავარი ეფექტების მოდელის შეფასებას. **Factorial ANOVA** ანალიზი ითვალისწინებს ფაქტორიალებს შორის ურთერთობას. ამ **Quick specs dialog** დიალოგის დროს შეიძლება მოცემული იყოს ოთხამდე კატეგორიული ცვლადი.

**Repeat measures ANOVA** ანალიზში დამოკიდებული ცვლადები შეიცავენ ერთი ფაქტორის განმეორებით გაზომვების მნიშვნელობებს. აქაც **Quick specs dialog** დიალოგის დროს შეიძლება მოცემული იყოს ოთხამდე კატეგორიული ცვლადები და ორი ან მეტი დამოკიდებული ცვლადი, რომლებიც პროგრამის მიერ ინტერპრეტირებული იქნება როგორც ერთი ფაქტორის განმეორებითი გაზომვები.

შევარჩოთ ანალიზის სახე (მაგალითად **Factorial ANOVA**) და სპეციალური მეთოდი (მაგალითად **Quick Specs Dialog**) და შემდეგ **OK**. ეკრანზე გამოდის **ANOVA/MANOVA Factorial ANOVA** ფანჯარა.

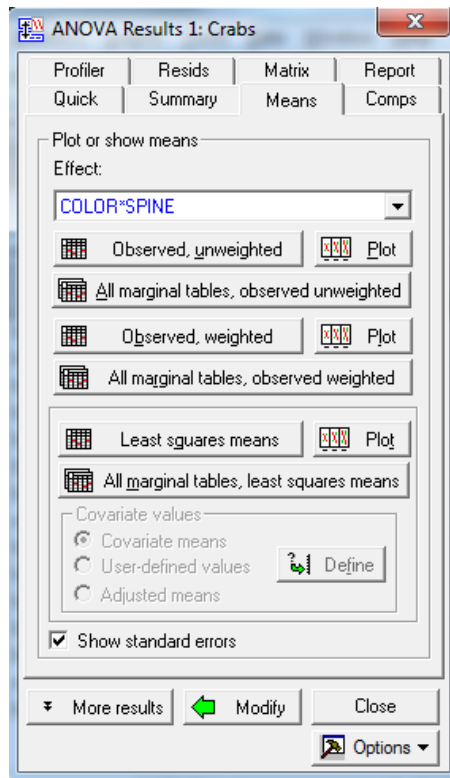


click **Variables** ღილაკზე და ეკრანზე გამოსულ ფანჯარაში მოვნიშნოთ დამაჯგუფებელი ცვლადები და დამოკიდებული ცვლადები. დამაჯგუფებელ ცვლადებად ავიღოთ **COLOR** და **SPINE**, ხოლო დამოკიდებულ ცვლადებად **WIDTH**, **WEIGHT**. სხვადასხვა ფერი და მარწუხების მდგომარეობა წარმოადგენენ ჯგუფთაშორისო ფაქტორებს. თუ დამოკიდებული ცვლადების რაოდენობა ერთზე მეტია, მაშინ პროგრამა შეასრულებს მრავალფაქტორიან დისპერსიულ ანალიზს.



შემდეგ *click OK*. ვბრუნდებით *ANOVA/MNNOVA Factorial ANOVA* ფანჯარაში. ჯგუფთაშორისო ფაქტორების კოდების მინიჭებისათვის *click Factor Codes* (ფაქტორების კოდები)-ის ღილაკზე. არ არის აუცილებელი ხელით კოდების მინიჭება, რადგან გაჩუმების პრინციპით პროგრამა ამას გააკეთებს ავტომატურად. *Options* ღილაკი გამოიყენება გამოთვლითი პროცესის პარამეტრების დასაყენებლად. *click OK*.

ეკრანზე გამოდის *ANOVA Results 1* (ანალიზის შედეგები) ფანჯარა, სადაც *Means* ჩანართში წარმოდგენილია საშუალოს გამოთვლის სხვადასხვა მეთოდები და მათი შესაბამისი გრაფიკები.



click **Quick** ჩანართის **All Effects/Graphs** (ყველა ეფექტი/გრაფიკები) ღილაკზე. ეკრანზე გამოდის **Table of All Effects** ფანჯარა, სადაც **Effect** სვეტში წარმოდგენილია დამაჯგუფებელი ცვლადების დასახელება და მათ ურთერთქმედება.

Effect	Wilk...	F	Eff...	Err...	p
COLOR	,925	2,119	6	320	,051
SPINE	,930	2,977	4	320	,019*
COLOR*SPINE	,880	1,766	12	320	,053

ის ეფექტები, რომლებიც მნიშვნელოვანია ანუ სარწმუნო არიან ( $p < 0,05$ ) აღნიშნულია \* სიმბოლოთი. წარმოდგენილ შედეგიდან ჩანს, რომ

---

ე.ყუბანეიშვილი. მონაცემების სტატისტიკური დამუშავება



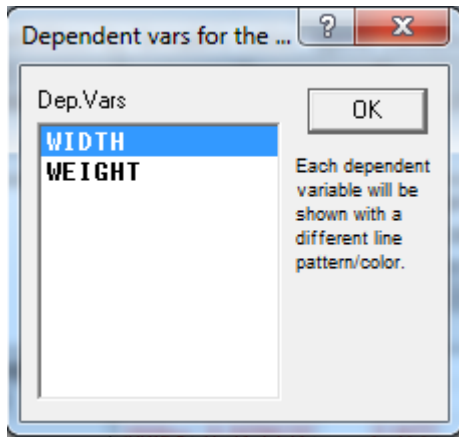
საშუალოების ტოლობის ჰიპოთეზა სარწმუნოა მხოლოდ **SPINE** პრედიქტორისათვის. **COLOR** პრედიქტორისათვის და **COLOR•SPINE** ურთიერთქმედების პრედიქტორებისათვის სარწმუნოების მნიშვნელოვნების დონე მეტია 0.05 სიდიდეზე.

ჩავრთოდ **Display** ველში **Spreadsheet** (ცხრილი)-ის ოპცია და ორჯერ **click SPINE** ეფექტზე ან მოვნიშნოთ **SPINE** ეფექტი და **OK**. ეკრანზე გამოდის ცხრილი,

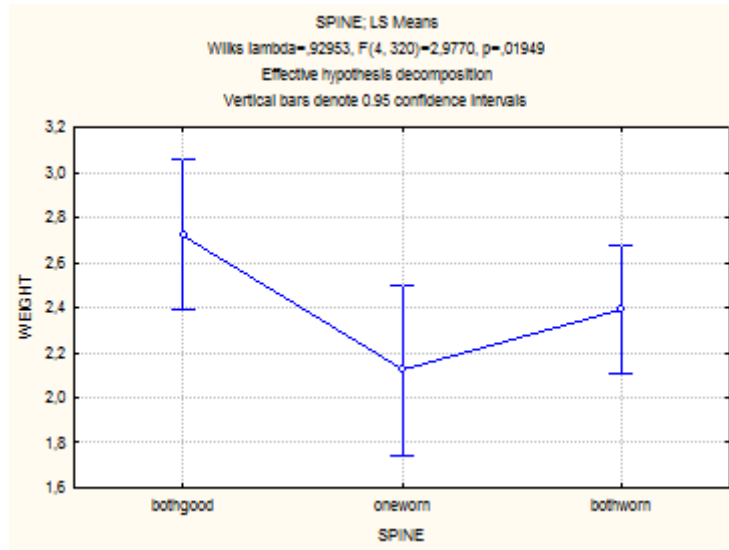
SPINE; LS Means (Crabs)										
Wilks lambda=.92953, F(4, 320)=2.9770, p=.01949										
Effective hypothesis decomposition										
Cell No.	SPINE	WIDTH Mean	WIDTH Std.Err.	WIDTH -95.00%	WIDTH +95.00%	WEIGHT Mean	WEIGHT Std.Err.	WEIGHT -95.00%	WEIGHT +95.00%	N
1	bothgood	26,80451	0,598481	25,62263	27,98640	2,721788	0,168747	2,388546	3,055030	37
2	oneworn	24,10625	0,672241	22,77870	25,43380	2,120313	0,189544	1,745999	2,494626	15
3	bothworn	25,93783	0,513233	24,92429	26,95137	2,391723	0,144710	2,105948	2,677498	121

სადაც მოცემულია ყველა დამოკიდებული ცვლადის და ჯგუფებში სხვა სტატისტიკების საშუალო მნიშვნელობები, რომლებიც შეესაბამებინან ჯგუფთაშორისო ფაქტორებს **SPINE-bothgood; oneworr; bothworr**.

– დაებრუნდეთ **Table of All Effects** ფანჯარაში. მოვნიშნოთ ცვლადი (მაგალითად **SPINE**). **Disply** ჩარჩოში ჩავრთოდ **Graph** ალამი და შემდეგ **OK**.

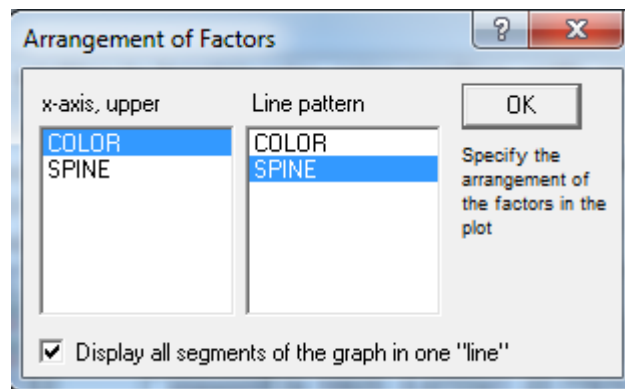


ეკრანზე გამოსულ **Dependat vars for the...** ფანჯარის **Dep.Var** ცვლადების ველში შევარჩოთ ცვლადი (მაგალითად **WIDHT**) და **OK**. ეკრანზე გამოდის **WIDHT** ცვლადის საშუალოების გრაფიკი

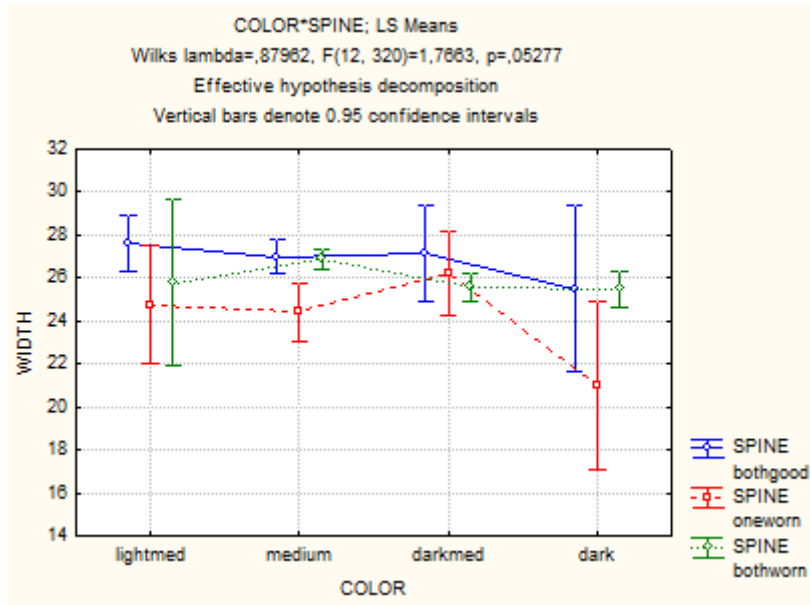


გრაფიკიდან ჩანს, რომ ორი მთელი მარწუხის მქონე კიბორჩხალების საშუალო სიგანე მეტია იმ კიბორჩხალების საშუალო სიგანეზე, რომელთაც გააჩნიათ დაზიანებული მარწუხები. როგორც ჩანს, უფრო დიდი სიგანის მქონე კიბორჩხალებს გააჩნიათ დიდი ძალა, რომელიც მათ საშუალებას აძლევს შეინარჩუნონ მარწუხი მთელი სახით.

ვბრუნდებით *Table of All Effects* ფანჯარაში, მოვნიშნოთ **COLOR\*SPINE**, ჩავრთოდ *Graph* ოპცია და შემდეგ *OK*. ეკრანზე გამოდის *Dependent vars for the...* ფანჯარა, სადაც მოვნიშნოთ **WIDHT** ცვლადი და შემდეგ *OK*. ეკრანზე გამოდის *Arrangement of Factors* (ფაქტორების განლაგება)-ის ფანჯარა,



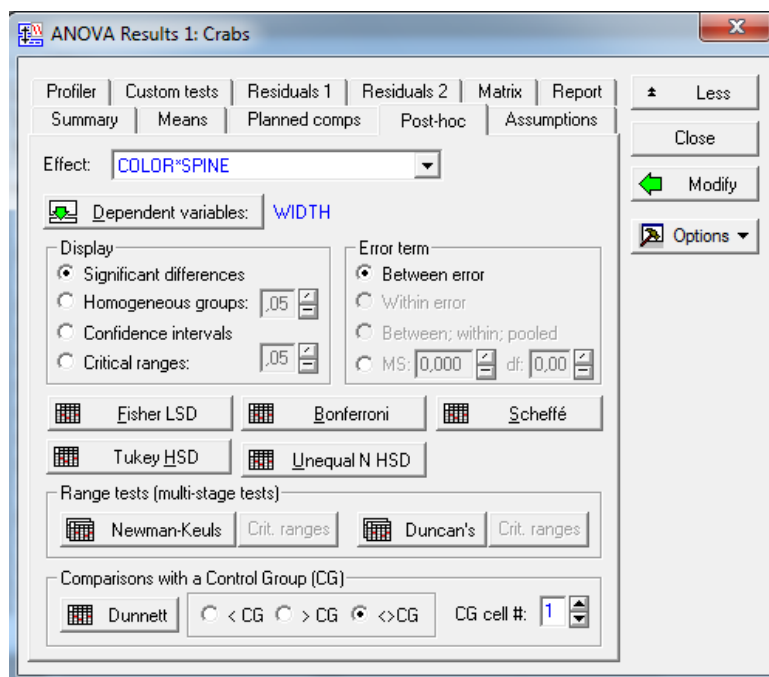
სადაც შესაძლებელია მიუთითოთ ურთიერთმოქმედი ფაქტორების შერჩევის რიგითობა. შევარჩიოთ **COLOR x-axis,upper** (x დერძის ზემოთ) ველში და **SPINE Line pattern** (ხაზის შაბლონი) ველში. შემდეგ *OK*. ეკრანზე გამოდის საშუალოების გრაფიკი.



მიღებული გრაფიკებიდან ჩანს, რომ ნათელი-რუხი ფერის ორი მთელი მარწუხის მქონე კიბორჩხალებს და მუქი ფერის ერთი დაზიანებული მარწუხის მქონე კიბორჩხალების საშუალო სიგანეები შესაბამისად დიდია და მცირე, ვიდრე სხვა ჯგუფებში. ფერის მიუხედავად ერთი დაზიანებული მარწუხის მქონე კიბორჩხალების საშუალო სიგანე ნაკლებია, ვიდრე ორი მთელი მარწუხის მქონე კიბორჩხალების საშუალო სიდიდეს.

მიღებული შედეგები გვიჩვენებს, რომ არსებობს განსხვავება ჯგუფების საშუალო სიდიდეებს შორის, მაგრამ ამ განსხვავების დასადგენად უნდა შევამოწმოთ საშუალოების ტოლობის ჰიპოთეზა.

**ANOVA Results 1** დიალოგურ ფანჯარაში გავხსნათ **Mears** ჩანართი და **Click More results** დილაკზე. ეკრანზე გამოსულ ფანჯარაში ჩავრთოდ **Post-hoc** ჩანართი,



ე.ყუბანეიშვილი. მონაცემების სტატისტიკური დამუშავება

სადაც მოცემულია სხვადასხვა აპოსტერიული კრიტერიუმები. ყველა ეს კრიტერიუმი საშუალებას გვაძლევს შევადაროთ საშუალოები. კრიტერიუმების უმრავლესობა ახდენს შემთხვევითი შედეგების ალბათობების მინიმიზაციას.

შევარჩიოთ დამოკიდებული **WIDTH** ცვლადი, **SPINE** ეფექტი და **Click Fisher LSD** დილაკზე. ეკრანზე გამოდის შედეგების ცხრილი,

LSD test; variable WIDTH (Crabs)				
Probabilities for Post Hoc Tests				
Error: Between MS = 3,8563, df = 161,00				
Cell No.	SPINE	{1}	{2}	{3}
		27,111	24,727	26,245
1	bothgood		0,000110	0,020209
2	oneworn	0,000110		0,005320
3	bothworn	0,020209	0,005320	

სადაც პირველ სტრიქონში მოყვანილია საშუალოების მნიშვნელობები, სვეტ 1-ში – ჯგუფების დასახელება, დანარჩენ უჯრედებში – მნიშვნელობების დონეები. ცხრილიდან გამომდინარეობს, რომ საშუალოების ტოლობის ჰიპოთეზა უარყოფილია, ე.ი. ყველა ჯგუფში კიბორჩხალების საშუალო სიგანე, რომლებიც შეესაბამებიან **SPINE** პრედიქტორს, სტატისტიკურად განსხვავდებიან ერთმანეთისაგან.

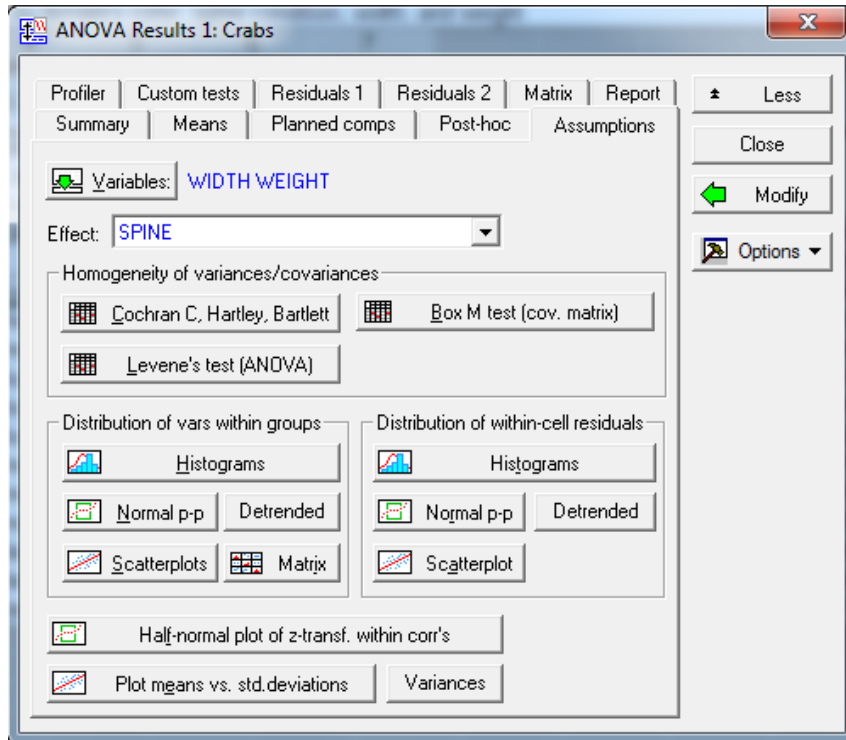
უფრო საინტერესო შედეგს ვღებულობთ, თუ კი **ANOVA Results 1** დიალოგურ ფანჯარაში შევარჩევთ **COLOR\*SPINE** ეფექტს. ამისათვის დავბრუნდეთ **Table of All Effects** ფანჯარაში. მოვნიშნოთ **COLOR\*SPINE** ეფექტი, ჩავართოდ **Spreadsheet** ოპცია და შემდეგ **OK**. ეკრანზე გამოდის შედეგების ცხრილი.

COLOR*SPINE; LS Means (Crabs)											
Wilks lambda=.87962, F(12, 320)=1,7663, p=.05277											
Effective hypothesis decomposition											
Cell No.	COLOR	SPINE	WIDTH Mean	WIDTH Std.Err.	WIDTH -95,00%	WIDTH +95,00%	WEIGHT Mean	WEIGHT Std.Err.	WEIGHT -95,00%	WEIGHT +95,00%	N
1	lightmed	bothgood	27,58889	0,654581	26,29622	28,88156	2,744444	0,184565	2,379965	3,108924	9
2	lightmed	oneworn	24,75000	1,388576	22,00783	27,49217	2,125000	0,391521	1,351822	2,898178	2
3	lightmed	bothworn	25,80000	1,963742	21,92199	29,67801	2,600000	0,553694	1,506561	3,693439	1
4	medium	bothgood	26,99583	0,400847	26,20424	27,78743	2,642708	0,113022	2,419511	2,865906	24
5	medium	oneworn	24,42500	0,694288	23,05391	25,79609	2,043750	0,195760	1,657161	2,430339	8
6	medium	bothworn	26,88254	0,247408	26,39396	27,37112	2,560587	0,069759	2,422827	2,698348	63
7	darkmed	bothgood	27,13333	1,133767	24,89436	29,37231	2,750000	0,319675	2,118703	3,381297	3
8	darkmed	oneworn	26,25000	0,981871	24,31099	28,18901	2,462500	0,276847	1,915781	3,009219	4
9	darkmed	bothworn	25,58378	0,322837	24,94624	26,22133	2,245054	0,091027	2,065294	2,424814	37
10	dark	bothgood	25,50000	1,963742	21,62199	29,37801	2,750000	0,553694	1,656561	3,843439	1
11	dark	oneworn	21,00000	1,963742	17,12199	24,87801	1,850000	0,553694	0,756561	2,943439	1
12	dark	bothworn	25,48500	0,439106	24,61785	26,35215	2,161250	0,123810	1,916750	2,405750	20

როგორც ცხრილიდან ჩანს, ორი მთელი მარწუხის მქონე რუხი ფერის კიბორჩხალის სიგანე (27,58) მნიშვნელოვნად მეტია, ვიდრე რუხი ფერის ერთი დაზიანებული მარწუხის მქონე კიბორჩხალების სიგანის საშუალო სიდიდეზე (24,75). შავი ფერის ერთი დაზიანებული მარწუხის კიბორჩხალის საშუალო

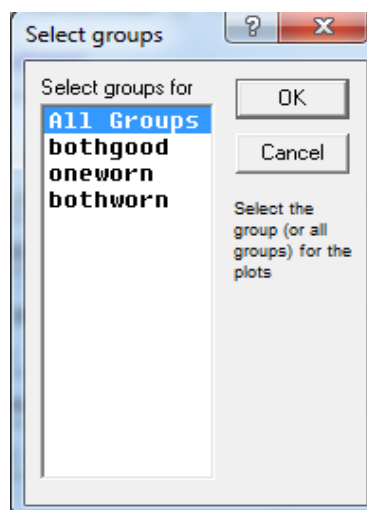
სიგანის მნიშვნელობა (27,58) მნიშვნელოვნად მეტია, ვიდრე რუხი ფერის ორი დაზიანებული მარწუხის მქონე კიბორჩხალების საშუალო სიგანეზე (26,99).

დისპერსიული ანალიზისებების შესამოწმებლად საჭიროა **ANOVA Results 1** დიალოგურ ფანჯარაში გავხსნათ **Assumptions** (დაშვებები) ჩანართი.

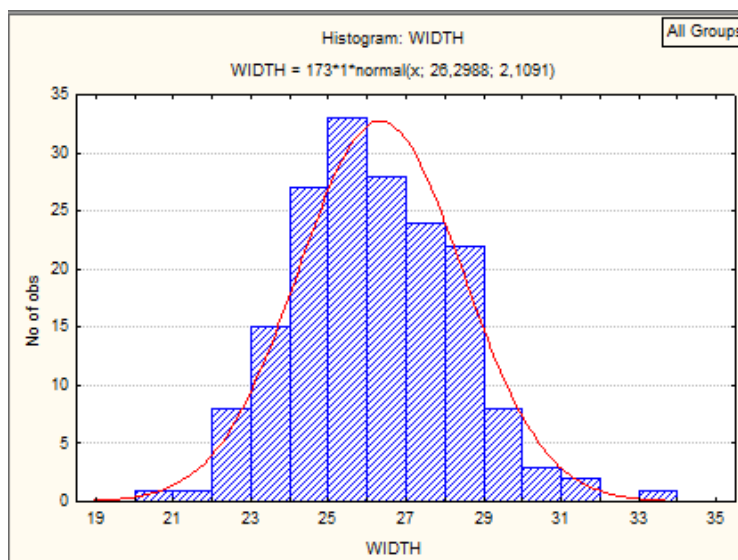
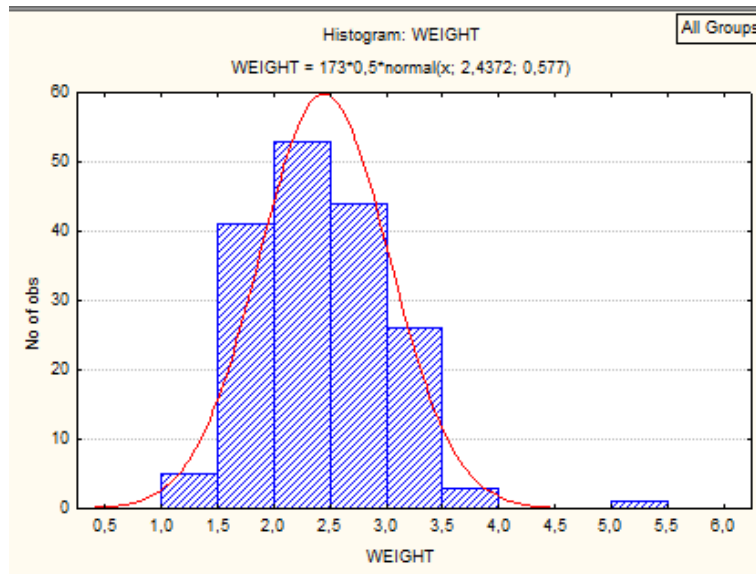


ეკრანზე გამოსულ ფანჯარაში წარმოდგენილია დისპერსიების ერთგვაროვნების რამოდენიმე კრიტერიუმი (კოჰრანის, ჰარტლის, ბარტლეს, ლევენის, ბოქსის  $M$  კრიტერიუმი); ცვლადების ნორმალური განაწილების კანონის შემოწმების გრაფიკული საშუალოებები (ჰისტოგრამა, გაფანტვის დიაგრამა, ნორმალური ალბათური გრაფიკი).

შევარჩიოთ **SPINE** ეფექტი და **Click Histograms** ღილაკზე. ეკრანზე გამოდის **Select groups** ფანჯარა,



სადაც მოვნიშნავთ ჯგუფს ან ყველა ჯგუფს. მოვნიშნოთ *All Group* და *OK*. ეკრანზე გამოდის ყველა ჯგუფის ჰისტოგრამა და განაწილების სიმკვრივის ფუნქციის გრაფიკები.

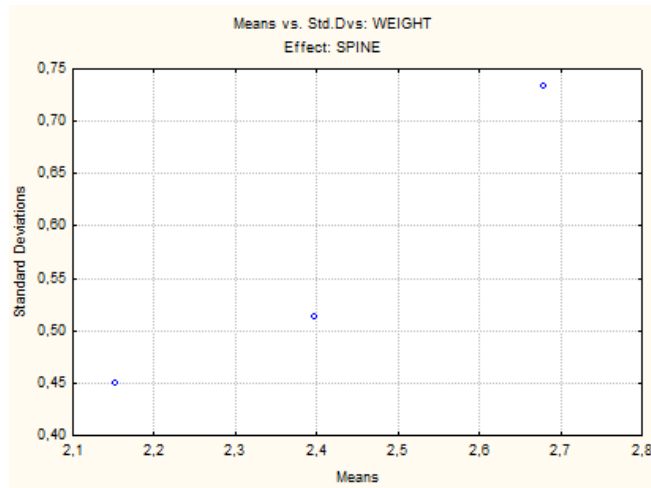


გრაფიკებიდან ჩანს, რომ საერთო განაწილება შეესაბამება ნორმალურს. *Click Levens test (ANOVA)* ღილაკზე. ეკრანზე გამოდის ცხრილი,

Levene's Test for Homogeneity of Variances (Crabs)				
Effect: SPINE				
Degrees of freedom for all F's: 2, 170				
	MS	MS	F	p
	Effect	Error		
WIDTH	2,952992	1,545181	1,911097	0,151082
WEIGHT	0,300225	0,113082	2,654939	0,073218

სადაც წარმოდგენილია **WIDTH** და **WEIGHT** დამოკიდებული ცვლადების დისპერსიების ერთგვაროვნების ჰიპოთეზის შემოწმების შედეგები. ცხრილიდან გამომდინარეობს, რომ ყველა ჯგუფის დისპერსიები, კატეგორიული **CPINE** პრედიქტორის დროს, ერთგვაროვნები არიან, ანუ დისპერსიების ტოლობის ჰიპოთეზა მიიღება.

დისპერსიული ანალიზის გამოყენების კიდევ ერთი პირობაა საშუალოსა და სტანდარტულ გადახრებს შორის კორელაციური კავშირის არ არსებობა. ამის შემოწმებისათვის **Click Plot means vs. std. deviations** დიაგრამა ეკრანზე გამოდის დიაგრამა,



სადაც ჩანს, რომ საშუალოები და სტანდარტული გადახრები უმნიშვნელოდ კორელირებენ.

პრაქტიკული სამუშაო 6

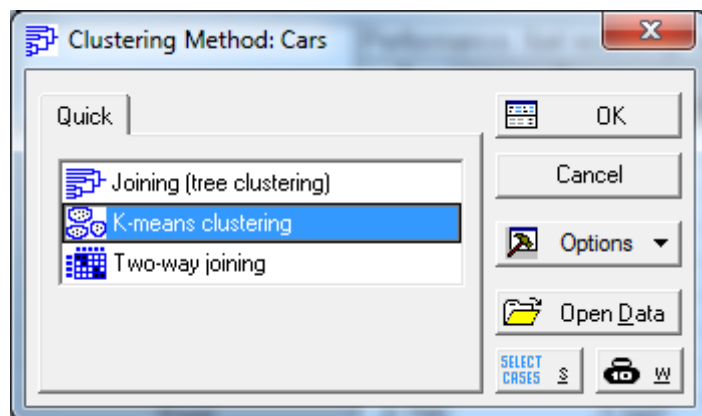
**კლასტერული ანალიზი**

პროგრამის მუშაობა განვიხილოთ **Examples** ბიბლიოთეკიდან შერჩეულ **Cars.sta** მონაცემთა ფაილზე, სადაც ცხრილის სტრიქონებში ჩაწერილია მანქანების დასახელება, სვეტებში პარამეტრები: **Price** - ფასი, **Acceleration** - გაქანების დრო წამებში ; **Braking** - სამუხრუჭე მანძილი, **Handling** - მოხვევის უმცირესი რადიუსი: **Mileage** - საწვავის ხარჯი.

	Performance, fuel economy, and approximate price for various automobiles				
	1 PRICE	2 ACCELERATION	3 BRAKING	4 HANDLING	5 MILEAGE
Acura	-0,521	0,477	-0,007	0,382	2,079
Audi	0,866	0,208	0,319	-0,091	-0,677
BMW	0,496	-0,802	0,192	-0,091	-0,154
Buick	-0,614	1,689	0,933	-0,210	-0,154
Corvette	1,235	-1,811	-0,494	0,973	-0,677
Chrysler	-0,614	0,073	0,427	-0,210	-0,154
Dodge	-0,706	-0,196	0,481	0,145	-0,154
Eagle	-0,614	1,218	-4,199	-0,210	-0,677
Ford	-0,706	-1,542	0,987	0,145	-1,724
Honda	-0,429	0,410	-0,007	0,027	0,369
Isuzu	-0,798	0,410	-0,061	-4,230	1,067
Mazda	0,126	0,679	-0,133	0,500	-1,724
Mercedes	1,051	0,006	0,120	-0,091	-0,154
Mitsub.	-0,614	-1,003	0,084	0,382	0,718
Nissan	-0,429	0,073	-0,007	0,263	0,997
Olds	-0,614	-0,734	0,409	0,382	2,114
Pontiac	-0,614	0,679	0,536	0,145	0,195
Porsche	3,454	-2,215	-0,296	0,618	-1,026
Saab	0,588	0,679	0,246	0,263	0,021
Toyota	-0,059	1,218	0,228	0,736	-0,851
VW	-0,706	-0,128	0,102	0,382	0,195
Volvo	0,219	0,612	0,138	-0,210	0,369

## 1. K-საშუალოს მეთოდი

*Statistics* მენიუში შევარჩიოთ *Multivariate Exporatory Teohniques* (მრავალგანზომილებიანი კვლევითი მეთოდები) და *click Cluster Analysis* (კლასტერული ანალიზი)-ის ბრძანებაზე. ეკრანზე გამოდის *Clustecring Metod* სასტარტო ფანჯარა,

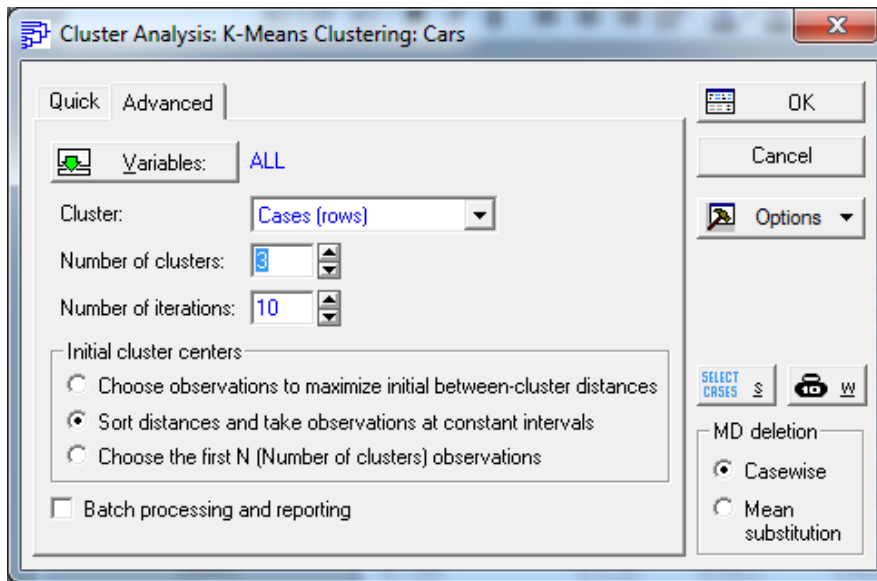


სადაც *Quick* ჩანართში წარმოდგენილია: *Joining (tree clustering)* (იერარქიული კლასტერისაცია (ხეს მსგავსი)); *K - means clustering* (*K*-საშუალოს მეთოდი), *Tow-way joining* (ორშესასვლიანი კლასტერისაცია).

-----  
ე.ყუბანეიშვილი. მონაცემების სტატისტიკური დამუშავება



მოენიშნოთ **K-means clustering** და შემდეგ **OK**. ეკრანზე გამოდის **Cluster Analysis: K-Means Clustering** ფანჯარა



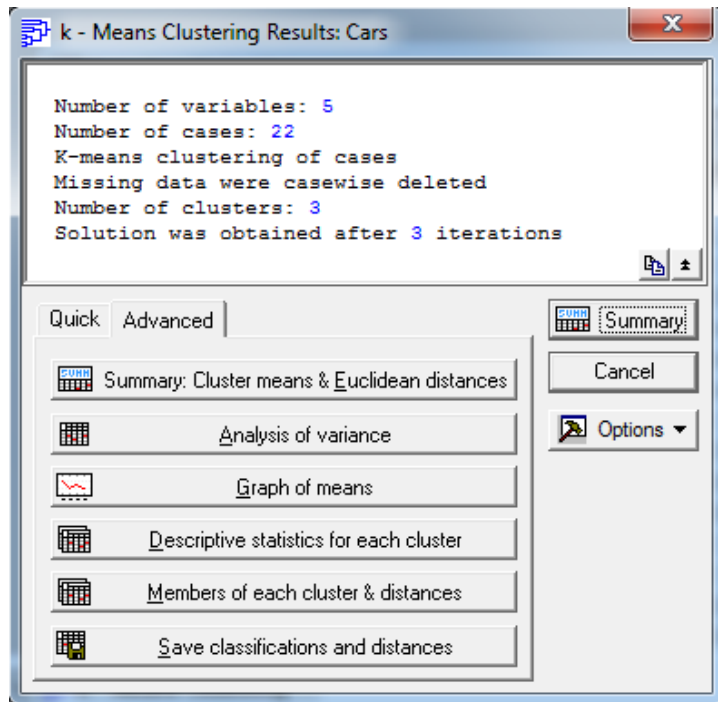
– ჩაერთო **Advanced** ჩანართი და **Variables** დილაკით მოენიშნოთ ყველა ცვლადი. შემდეგ **OK**. ვბრუნდებით წინა ფანჯარაში, სადაც **Cluster** ველის გაშლადს სიაში შევარჩოთ კლასტერიზაციისათვის ობიექტები, ჩვენ შემთხვევაში ავიღოთ **Cases(rows)**–დაკვირვება(სტრიქონი), რადგან აქ არის ჩაწერილი დასაჯგუფებელი მანქანების დასახელებები.

– **Number of clusters** ველში ჩაწეროთ კლასტერების რაოდენობა (მაგალითად-3). **Number of iteration** (იტერერაციის რაოდენობა)-ის ველში ჩაწეროთ იტერერაციის მაქსიმალური რაოდენობა, მაგალითად - 10.

**Initial cluster centers** არეში წარმოადგენილია შემდეგი ოპციები:

- **Choose observations to maximize initial between-cluster distauces** (შევარჩოთ დაკვირვებები, რომლებიც მოახდენენ კლასტერებს შორისო საწყისი მანძილების მაქსიმიზაციას);
- **Sort distances and take observatias at constant intervals** (მანძილების დახარისხება და შეირჩეს დაკვირვებები მუდმივ ინტერვალებიდან);
- **Choose the First N (Number of clusters) observations** (შეირჩეს პირველი **N** (კლასტერების რიცხვი) დაკვირვება).

ამ სამი ოპციიდან შევარჩოთ რომელიმე მათგანი: მაგალითად, მეორე. შემდეგ **OK**. გაიხსნება **K-Means Clustering Results** შედეგების ფანჯარა.



გაგეხსნათ *Advanced* ჩანართ. ფანჯრის ზედა საინფორმაციო კვლში წარმოდგენილია

- *Number of variables* (ცვლადების რაოდენობა);
- *Number of cases* (დაკვირვებათა რაოდენობა);
- *K-means clustering of cases* (*k*-საშუალოების მეთდი);
- *Missing data were casewise deleted* (გამოტოვებული მნიშვნელობების დამუშავება გამორიცხულია);
- *Number of clusters* (კლასტერების რაოდენობა);
- *Solution was obtained after • iterations* (ამოხსნა მოხდა • იტერაციის შემდეგ).

ფანჯრის ქვედა ნაწილში წარმოდგენილია შემდეგი ფუნქციონალური დანიშნულების დილაკები:

– *Summary : Cluster means & Euclidean distances* დილაკზე *click*. ეკრანზე გამოდის ცხრილები, სადაც ნაჩვენებია კლასტერების საშუალო მნიშვნელობები (გასაშუალოება ხდება კლასტერის შიგნით), კლასტერებს შორის ევკლიდეს მანძილები და ევკლიდეს მანძილების კვადრატები.

Variable	Cluster Means (Cars)		
	Cluster No. 1	Cluster No. 2	Cluster No. 3
PRICE	-0,393067	0,931687	-0,70597
ACCELERATION	0,296047	-0,782310	0,81378
BRAKING	0,274215	0,099270	-2,12984
HANDLING	0,190603	0,280264	-2,21984
MILEAGE	0,441901	-0,876397	0,19503

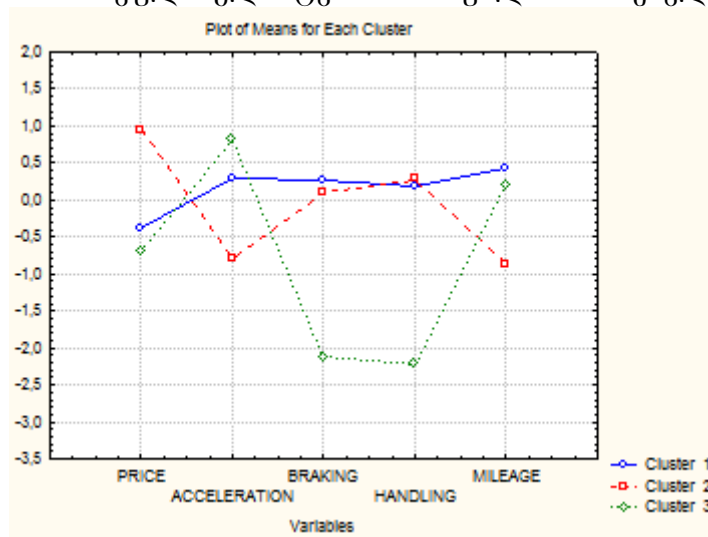
Euclidean Distances between Clusters (Cars)				
Cluster Number	Distances below diagonal			Squared distances above diagonal
	No. 1	No. 2	No. 3	
No. 1	0,000000	0,938877	2,403320	
No. 2	0,968957	0,000000	3,519362	
No. 3	1,550264	1,875996	0,000000	

– click **Analysis of variance** ღილაკზე. ეკრანზე გამოდის დისპერსიული ანალიზის ცხრილი,

Analysis of Variance (Cars)						
Variable	Between SS	df	Within SS	df	F	signif. p
PRICE	9,08159	2	11,91841	19	7,23881	0,004602
ACCELERATION	6,74790	2	14,25210	19	4,49794	0,025163
BRAKING	10,11892	2	10,88108	19	8,83457	0,001938
HANDLING	10,87750	2	10,12250	19	10,20857	0,000975
MILEAGE	7,99118	2	13,00882	19	5,83575	0,010573

სადაც მოყვანილია ჯგუფთაშორისი (**Between SS**) და შიგაჯგუფური (**Within SS**) ცვლადების დისპერსიები. რაც უფრო ნაკლებია შიგაჯგუფური დისპერსიები და მეტია ჯგუფთაშორისო დისპერსიები, მით უფრო „ხარისხიანია“ კლასტერიზაციის პროცესი. **F** და **p** პარამეტრები გვიჩვენებენ თითოეული ცვლადის წვლილს ობიექტების დაყოფაში კლასტრებად (ჯგუფებად). საუკეთესო კლასტერილიზაციას მიეკუთვნება **F**-ის დიდი, ხოლო **p**-ს მცირე მნიშვნელობები. ცვლადები, რომლებისთვისაც  $p > 0.05$  შეიძლება გამოირიცხოს კლასტერილიზაციის პროცედურაში.

– **Graph of means** ღილაკი საშუალებას იძლევა გრაფიკულად წარმოვადგინოთ თითოეული კლასტერის საშუალო მნიშვნელობა.



– **Descriptive statistics for each cluster** ღილაკზე დაწკაპუნებით ეკრანზე გამოდის ცხრილი, სადაც ნაჩვენებია თითოეული კლასტერის სტატისტიკური მახასიათებლები (საშუალო, დისპერსია და ა.შ.)

Descriptive Statistics for Cluster 3 (Cars) Cluster contains 2 cases			
Variable	Mean	Standard Deviation	Variance
PRICE	-0,70597	0,130742	0,017093
ACCELERATION	0,81378	0,571101	0,326156
BRAKING	-2,12984	2,926087	8,561983
HANDLING	-2,21984	2,842879	8,081959
MILEAGE	0,19503	1,233326	1,521094

– *Members of each cluster & distances* ღილაკით შეგვიძლია დავათვალიეროთ კლასტერებში ობიექტების განაწილების შედეგი. ცხრილებში ნაჩვენებია, აგრეთვე კლასტერის ცენტრთან ობიექტების მანძილები.

Members of Cluster Number 1 (Cars) and Distances from Respective Cluster Center Cluster contains 13 cases	
	Distance
Acura	0,754166
Buick	0,766466
Chrysler	0,356816
Dodge	0,384616
Honda	0,158199
Mitsub.	0,614239
Nissan	0,297823
Olds	0,889882
Pontiac	0,255611
Saab	0,508612
Toyota	0,766000
VW	0,284704
Volvo	0,362700

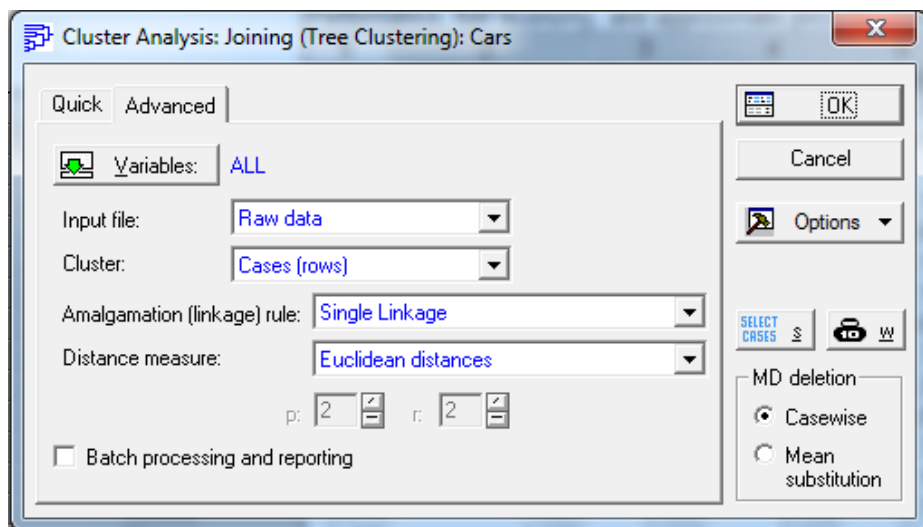
Members of Cluster Number 2 (Cars) and Distances from Respective Cluster Center Cluster contains 7 cases	
	Distance
Audi	0,492166
BMW	0,414537
Corvette	0,636028
Ford	0,978105
Mazda	0,849100
Mercedes	0,509201
Porsche	1,319699

Members of Cluster Number 3 (Cars) and Distances from Respective Cluster Center Cluster contains 2 cases	
	Distance
Eagle	1,360452
Isuzu	1,360452

– *Save classifications and distances* ღილაკის საშუალებით ხდება კლასტერიზაციის შედეგების დამახსოვრება მისი შემდგომი კვლევისათვის.

## 2. კლასტერიზაცია იერარქიის მეთოდით

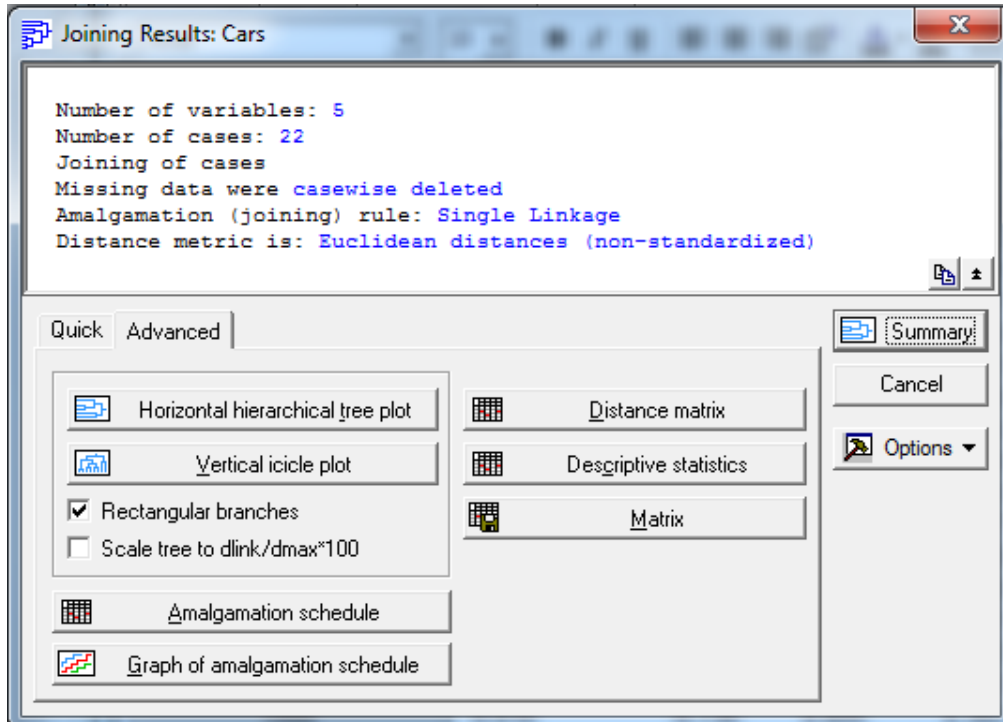
*Clarstering method* ფანჯარაში მოვნიშნოთ *Joining (tree clustering)* და შემდეგ *OK*. ეკრანზე გამოსულ *Cluster Analysis Joining (Tree Clustersing)*



ანალიზის ფანჯარაში *click Variables* ღილაკზე და *Select variables for the analysis* ფანჯარაში მოვნიშნოთ ყველა ცვლადი. შემდეგ *OK*.

– ვბრუნდებით წინა ფანჯარაში, სადაც *Cluster* გაშლად სიაში შევარჩიოთ *Cases(rows)*. რადგან საწყისად ვიღებთ მოცემული მონაცემების მატრიცას, ამიტომ *Input file*-ს ველში უნდა შევარჩიოთ *Raw data* და არა მანძილების მატრიცა.

– *Amalgamation (linkage) rule* (იერარქიული გაერთიანების წესი)-ის ველში შევარჩიოთ *Single linkage* (ერთეულგვანი კავშირის მეთოდი). შემდეგ *OK*. ეკრანზე გამოდის *Joining Results* შედეგების ფანჯარა,



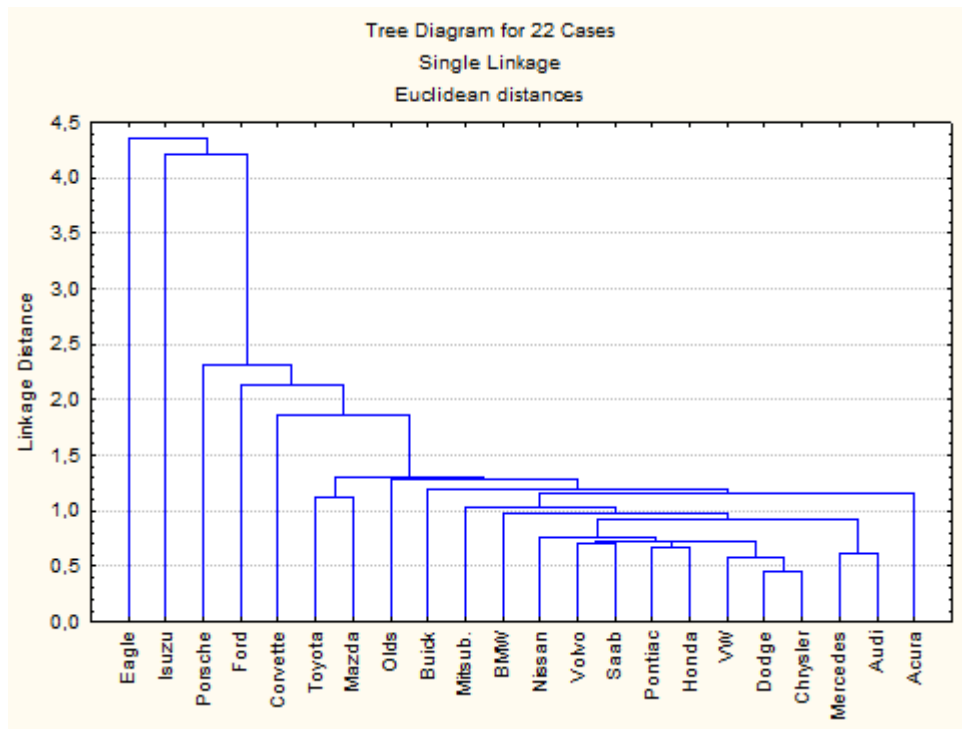
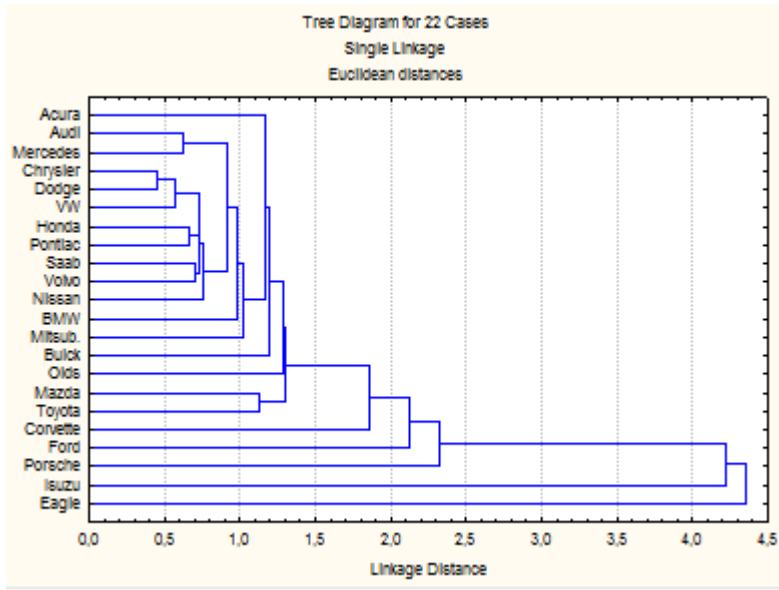
სადაც ზედა ველში მოყვანილია შემდეგი ინფორმაცია:

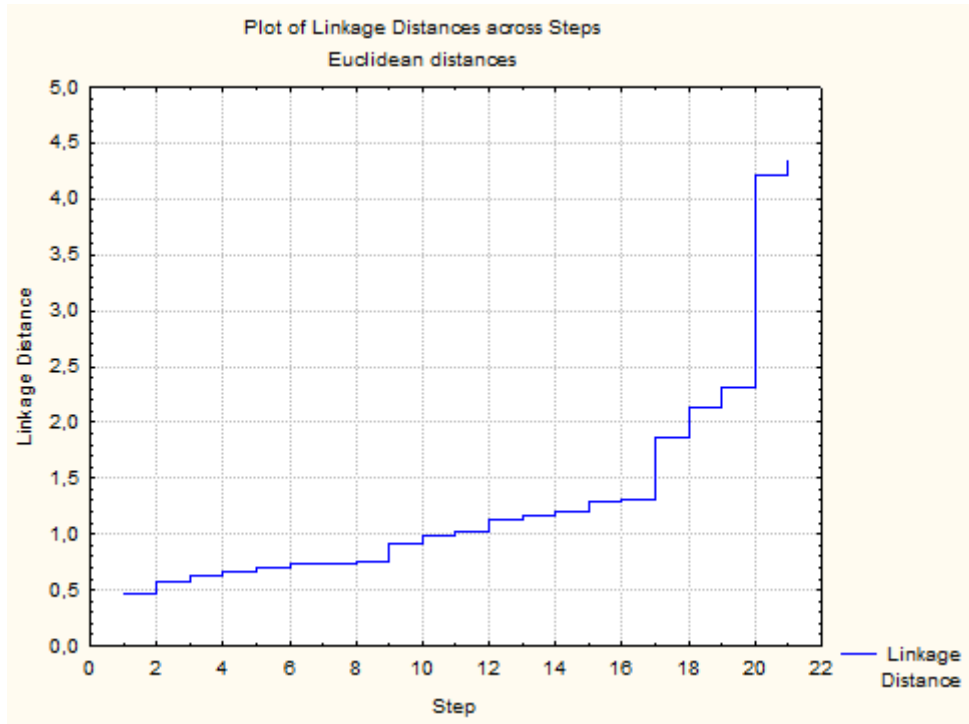
- **Number of variables** (ცვლადების რაოდენობა);
- **Number of cases** (დაკვირვებათა რიცხვი);
- **Joining of cases** (კლასტერიზაციის მეთოდი);
- **Missing data were** (შერჩეული მეტრიკა (მანძილი ობიექტებს შორის)).

ფანჯრის ქვედა ველში განლაგებულია შემდეგი დილაკები (**Advanced** ჩანართის ჩართვისას) :

- **Horizontal hierarchical tree plot** (ჰორიზონტალური ხეს მსგავსი დიაგრამა ანუ დენდოგრამა);
- **Vertical icicle plot** (ვერტიკალური დენდოგრამა);
- **Amalgamation schedule** (კლასტერებში გაერთიანების წესი);
- **Graph of amalgamation schedule** (გაერთიანების რიგის გრაფიკი);
- **Distance matrix** (მანძილების მატრიცა);
- **Descriptive statistics** (აღწერითი სტატისტიკა).

თითოეულ დილაკზე დაწკაპინებით გამოდის შესაბამისი შედეგების ცხრილები და გრაფიკები, შემდგომი ინტერპრეტაციისათვის.





Case No.	Euclidean distances (Cars)																					
	Acura	Audi	BMW	Buick	Corvette	Chrysler	Dodge	Eagle	Ford	Honda	Isuzu	Mazda	Mercedes	Mitsub.	Nissan	Olds	Pontiac	Porsche	Saab	Toyota	VW	Volvo
Acura	0.00	3.15	2.81	2.77	4.06	2.39	2.40	5.11	4.43	1.75	4.73	3.87	2.81	2.01	1.16	1.28	1.99	5.73	2.36	3.09	1.99	1.96
Audi	3.15	0.00	1.20	2.25	2.45	1.58	1.73	4.86	2.67	1.71	4.81	1.55	0.62	2.43	2.17	3.33	1.81	3.68	0.96	1.61	1.90	1.31
BMW	2.81	1.20	0.00	2.83	1.86	1.44	1.40	4.99	2.27	1.63	4.67	2.29	0.98	1.51	1.76	2.58	1.93	3.50	1.54	2.36	1.50	1.54
Buick	2.77	2.25	2.83	0.00	4.40	1.69	1.97	5.18	3.61	1.70	4.51	2.38	2.51	3.01	2.25	3.41	1.19	5.89	1.79	1.55	2.11	1.66
Corvette	4.06	2.45	1.86	4.40	0.00	3.08	2.88	5.26	2.80	3.15	6.27	2.98	2.26	2.59	3.14	3.68	3.48	2.32	2.86	3.38	2.84	3.13
Chrysler	2.39	1.58	1.44	1.69	3.08	0.00	0.46	4.79	2.35	0.82	4.25	2.05	1.70	1.55	1.33	2.48	0.79	4.87	1.45	1.74	0.79	1.16
Dodge	2.40	1.73	1.40	1.97	2.88	0.46	0.00	4.93	2.13	0.98	4.62	2.10	1.82	1.28	1.31	2.35	0.95	4.79	1.59	1.82	0.57	1.42
Eagle	5.11	4.86	4.99	5.18	5.26	4.79	4.93	0.00	5.98	4.41	6.08	4.36	4.81	5.06	4.68	5.76	4.86	6.66	4.71	4.56	4.63	4.58
Ford	4.43	2.67	2.27	3.61	2.80	2.35	2.13	5.98	0.00	3.04	5.64	2.65	2.96	2.67	3.33	3.97	2.97	4.49	3.20	3.12	2.55	3.27
Honda	1.75	1.71	1.63	1.70	3.15	0.82	0.98	4.41	3.04	0.00	4.33	2.24	1.63	1.51	0.75	2.16	0.67	4.93	1.16	1.68	0.73	0.73
Isuzu	4.73	4.81	4.67	4.51	6.27	4.25	4.62	6.08	5.64	4.33	0.00	5.58	4.72	4.84	4.52	4.89	4.51	7.27	4.83	5.44	4.73	4.21
Mazda	3.87	1.55	2.29	2.38	2.98	2.05	2.10	4.36	2.65	2.24	5.58	0.00	2.05	3.07	2.85	4.19	2.19	4.47	1.86	1.13	2.26	2.23
Mercedes	2.81	0.62	0.98	2.51	2.26	1.70	1.82	4.81	2.96	1.63	4.72	2.05	0.00	2.18	1.91	2.96	1.89	3.49	0.92	1.97	1.86	1.16
Mitsub.	2.01	2.43	1.51	3.01	2.59	1.55	1.28	5.06	2.67	1.51	4.84	3.07	2.18	0.00	1.14	1.46	1.83	4.61	2.19	2.80	1.02	1.94
Nissan	1.16	2.17	1.76	2.25	3.14	1.33	1.31	4.68	3.33	0.75	4.52	2.85	1.91	1.14	0.00	1.46	1.16	4.96	1.56	2.27	0.89	1.16
Olds	1.28	3.33	2.58	3.41	3.68	2.48	2.35	5.76	3.97	2.16	4.89	4.19	2.96	1.46	1.46	0.00	2.40	5.40	2.80	3.61	2.04	2.44
Pontiac	1.99	1.81	1.93	1.19	3.48	0.79	0.95	4.86	2.97	0.67	4.51	2.19	1.89	1.83	1.16	2.40	0.00	5.23	1.25	1.46	0.95	1.01
Porsche	5.73	3.68	3.50	5.89	2.32	4.87	4.79	6.66	4.49	4.93	7.27	4.47	3.49	4.61	4.96	5.40	5.23	0.00	4.25	4.94	4.83	4.61
Saab	2.36	0.96	1.54	1.79	2.86	1.45	1.59	4.71	3.20	1.16	4.83	1.86	0.92	2.19	1.56	2.80	1.25	4.25	0.00	1.30	1.55	0.71
Toyota	3.09	1.61	2.36	1.55	3.38	1.74	1.82	4.56	3.12	1.68	5.44	1.13	1.97	2.80	2.27	3.61	1.46	4.94	1.30	0.00	1.86	1.68



Case No.	Means and Standard Deviations (Cars)	
	Mean	Std.Dev.
Acura	0,481996	0,975266
Audi	0,124789	0,566370
BMW	-0,071731	0,482588
Buick	0,328976	0,952021
Corvette	-0,154868	1,258561
Chrysler	-0,095281	0,382895
Dodge	-0,085824	0,440538
Eagle	-0,896298	1,998454
Ford	-0,567803	1,143498
Honda	0,074216	0,339667
Isuzu	-0,722437	2,075560
Mazda	-0,110307	0,955851
Mercedes	0,186282	0,494108
Mitsub.	-0,086657	0,709643
Nissan	0,179791	0,522402
Olds	0,311310	1,141420
Pontiac	0,188268	0,501586
Porsche	0,107157	2,139045
Saab	0,359575	0,269914
Toyota	0,254400	0,787689
VW	-0,031192	0,419548
Volvo	0,225639	0,302917

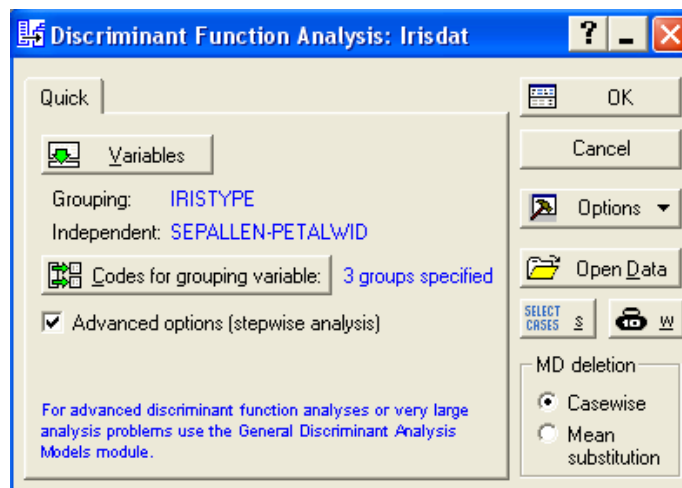
## პრაქტიკული სამუშაო 7

### დისკრიმინანტული ანალიზი

პროგრამის მუშაობა განვიხილოთ ფიშერის კლასიკურ მაგალითზე, კერძოდ ირისის ყვავილების ანალიზის მონაცემებზე, რომლებიც მოცემულია *Exemples* მონაცემთა ბიბლიოთეკაში *Irisadat.sta* ფაილის სახით. ფაილში მოცემულია 150 ირისის ყვავილისა გაზომვის შედეგები, კერძოდ: *Sepallen* (ჯამის ფოთლების სიგრძე); *Sepalwid* (ჯამის ფოთლების სიგანე); *Petalen* (ბუტკოს სიგრძე) და *Petalwid* (ბუტკოს სიგანე). ამოცანა მდგომარეობს ირისის ჯამის ფოთლის და ბუტკოს სიგრძე – სიგანის გაზომვის შედეგად იგი მივაკუთნოთ ერთ-ერთ რომელიმე *Setosa*, *Versicol*, *Virgenic* ირისის ტიპს. ჩვენგანით *Eexamples* ბიბლიოთეკიდან *Irisdat*. ფაილი., რომლის ფრაგმენტი წარმოდგენილია შემდეგ ნახაზზე:

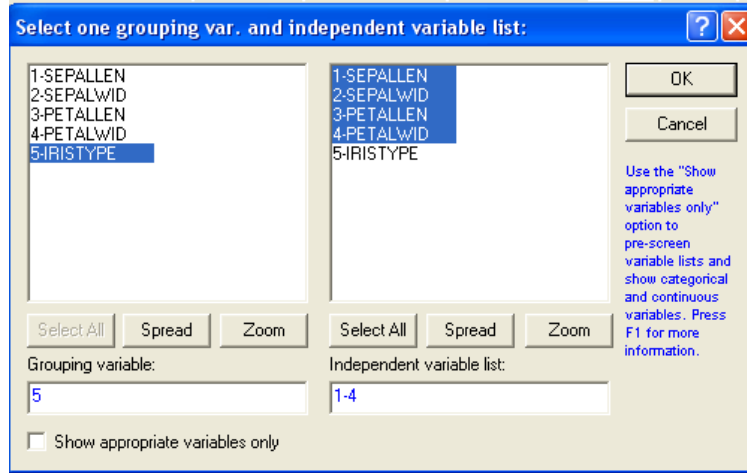
Fisher (1936) iris data: length & width of sepals and petals, 3 types of Iris					
	1	2	3	4	5
	SEPALLEN	SEPALWID	PETALLEN	PETALWID	IRISTYPE
1	5,0	3,3	1,4	0,2	SETOSA
2	6,4	2,8	5,6	2,2	VIRGINIC
3	6,5	2,8	4,6	1,5	VERSICO
4	6,7	3,1	5,6	2,4	VIRGINIC
5	6,3	2,8	5,1	1,5	VIRGINIC
6	4,6	3,4	1,4	0,3	SETOSA
7	6,9	3,1	5,1	2,3	VIRGINIC
8	6,2	2,2	4,5	1,5	VERSICO
9	5,9	3,2	4,8	1,8	VERSICO
10	4,6	3,6	1,0	0,2	SETOSA
11	6,1	3,0	4,6	1,4	VERSICO
12	6,0	2,7	5,1	1,6	VERSICO
13	6,5	3,0	5,2	2,0	VIRGINIC
14	5,6	2,5	3,9	1,1	VERSICO
15	6,5	3,0	5,5	1,8	VIRGINIC
16	5,8	2,7	5,1	1,9	VIRGINIC
17	6,8	3,2	5,9	2,3	VIRGINIC
18	5,1	3,3	1,7	0,5	SETOSA
19	5,7	2,8	4,5	1,3	VERSICO
20	6,2	3,4	5,4	2,3	VIRGINIC
21	7,7	3,8	6,7	2,2	VIRGINIC
22	6,3	3,3	4,7	1,6	VERSICO
23	6,7	3,3	5,7	2,5	VIRGINIC
24	7,6	3,0	6,6	2,1	VIRGINIC
25	4,9	2,5	4,5	1,7	VIRGINIC

*Statistics* მენიუში შევარჩიოთ *Multivariate Exploratory Technique* (მრავლობითი კვლევის მეთოდები) და *click Discriminant Analysis* ბრძანებაზე. ეკრანზე გამოდის *Discriminant Function Analysis* სასტარტო ფანჯარა.



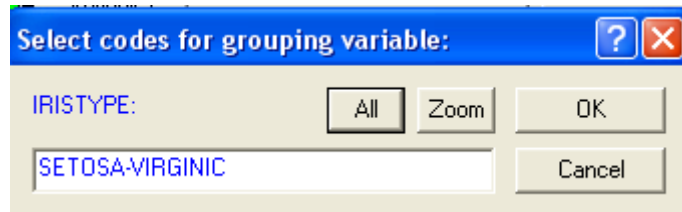
*click Variables* ღილაკზე. ეკრანზე გამოსულ *Select one groupin gvar....* ფანჯარაში საჭიროა შეირჩეს დამაჯგუფებელი (*Groupig*) და დამოუკიდებელი (*Independent*) ცვლადები. *Grouping variable* ველში მოვნიშნოთ დამოკიდებელი ცვლადი, მაგალითად *Iriastype*, რომელიც არ უნდა შედიოდეს დამოუკიდებელ

ცვლადების მონიშნულ სიაში. **Independent variable list** (დამოუკიდებელი ცვლადების სია)-ში მოვნიშნოთ პირველი ოთხი პარამეტრი და შემდეგ **OK**.



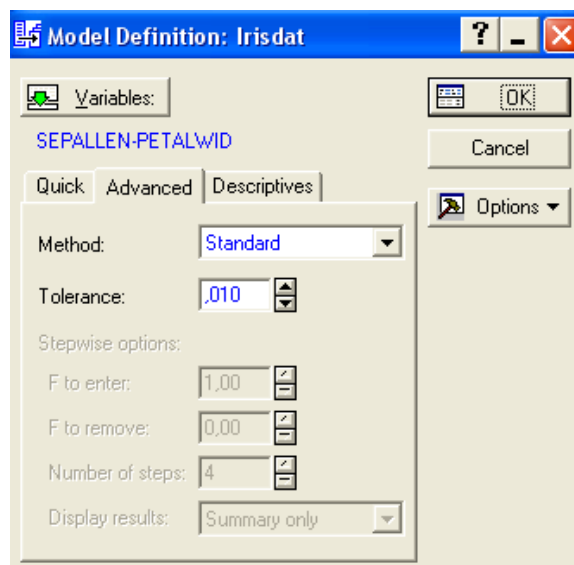
ვბრუნდებით სარსტარტო ფანჯარაში, სადაც უნდა მიუთოოდ დამაჯგუფებელი ცვლადების კოდები, ამისთვის **click Codes for grouping variable** ღილაკზე.

– ეკრანზე გამოდის **Select codes for grouping variable**-ს ფანჯარა. **Click All** ღილაკზე და **OK**.



ვბრუნდებით სასტარტო ფანჯარაში, ჩავრთოთ **Advanced options (stepwise analysis)** ოპცია და შემდეგ **OK**.

– გაიხსნება **Model Definition** ფანჯარა, ჩავრთოდ **Advanced** ჩანართ, და **Method** ველის გაშლად სიაში შევარჩოთ დისკრიმინანტული ანალიზის მეთდი.



- **Standard** (სტანდარტული) ამ დროს ხდება ერთდროულად ყველა შერჩეული ცვლადის ჩართვა დისკრიმინანტულ მოდელში;
- **Forward stepwise** (ბიჯური ჩართვა). ამ დროს პროგრამა ცვლადებს თანმიმდევრულად ჩართავს მოდელში;
- **Backward stepwise** (ბიჯური გამორთვა). დასაწყისში პროგრამა ყველა ცვლადს ჩართავს მოდელში და შემდეგ ყოველ ბიჯზე ხდება ცვლადის გამორიცხვა.

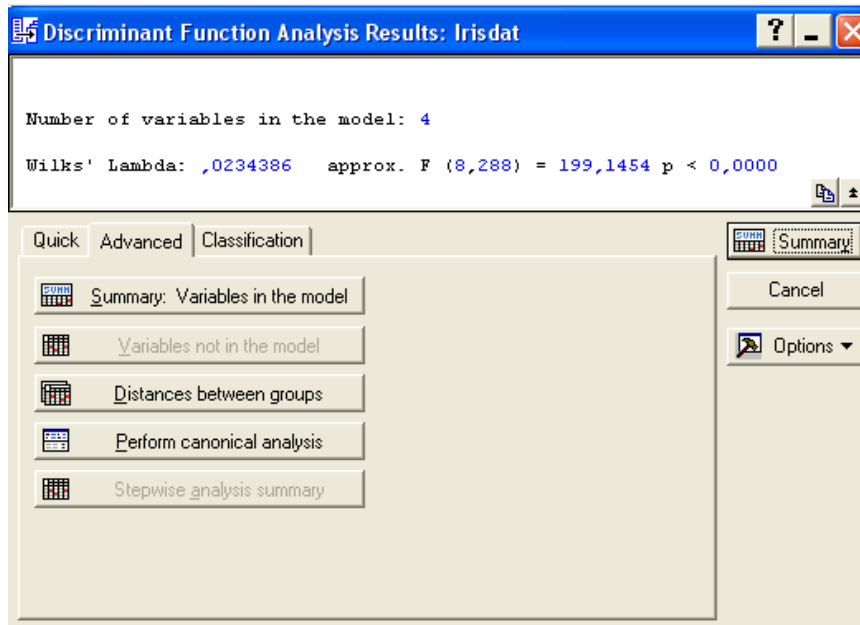
**Tolerance** (ტოლერანტობა)-ის ოპცია გვაძლევს ტოლერანტობის ქვედა ზღვრულ მნიშვნელობას. **ტოლერანტობა** არის ცვლადების სიჭარბის ზომა. რაც უფრო მცირეა ტოლერანტობა მოდელში მით უფრო გადაჭარბებულია ცვლადების რაოდენობა. ის ცვლადები, რომლებსაც გააჩნიათ ტოლერანტობა მოცემული ტოლერანტობის ზვრულ სიდიდეზე ნაკლები მოდელში არ ჩაითვებიან.

თუ **Method** ველის გაშლად სიაში შევარჩოთ **Forward stepwise** ან **Backward stepwise** ოპციებს, მაშინ ეკრანზე გააქტიურდება **Stepwise options** მაჩვენებლები:

- **F to enter** (**F** ჩართვა), **F to remove** (**F** გამორიცხვა). ბიჯური ანალიზის დროს ცვლადები ჩაირთვებიან დისკრეტულ მოდელში ის ცვლადები, რომელთა **F** მნიშვნელობა მეტია ვიდრე **F** - ჩართვის ველში ნაჩვენები სიდიდე. ცვლადები გამოირიცხებიან მოდელიდან, როცა შესაბამისი **F** - მნიშვნელობა ნაკლებია, ვიდრე **F** - გამორიცხვის. უნდა აღინიშნოს, რომ **F** - ჩართვის მნიშვნელობა ყოველთვის უნდა იყოს **F** - გამორიცხვის მნიშვნელობაზე მეტი, თუ გვინდა ბიჯების ანალიზის დროს ყველა ცვლადის ჩართვა, მაშინ **F to enter** ველში უნდა ჩავწეროთ მცირე მნიშვნელობა მაგალითად, 0,0001, ხოლო **F to remove** ველში – ნული. თუ გვინდა ბიჯური გამორიცხვის დროს მოდელიდან ყველა ცვლადის გამორიცხვა, მაშინ **F to enter** ველში უნდა ჩავწეროთ უდიდესი რიცხვი, მაგალითად, 9999, ხოლო **F to remove** ველში იგივე რიგის უმცირესი რიცხვი, მაგალითად, 9998.

- **Number of steps** (ბიჯების რაოდენობა)-ის ოპცია განსაზღვრავს ბიჯების მაქსიმალურ რაოდენობას, რომელიც შესრულებული იქნება;

- **Display of steps** (შედეგების გამოტანა). თუ ამ სიაში ავირჩევთ **Summary only** (მხოლოდ შედეგი), მაშინ პროგრამა დაასრულებს ბიჯურ ანალიზს და მხოლოდ მერე გამოიტანს შედეგებს. თუ შევარჩევთ **At each step** (ყოველ ბიჯზე) ოპციას, მაშინ პროგრამა შედეგებს გამოიტანს ყოველ ბიჯზე. **Model Definition** ფანჯარაში შევარჩოთ **Standard** მეთდი. შემდეგ **Ok**.



ეკრანზე გამოდის **Discriminant Function Analysis Results** (დისკრიმინანტული ანალიზის შედეგები)-ის ფანჯარა, სადაც ზედა ზოლში წარმოდგენილია ფანჯრის ინფორმაციული ნაწილი:

- **Number of variables in the model** (მოდელში ცვლადების რაოდენობა)
- **Wiles lambda** (უილკის ლამბდას მნიშვნელობა)
- **Approx F (8, 288)** (**F** სტატისტიკის მიახლოებითი მნიშვნელობა 8 და 288 თავისუფლების ხარისხებით), რომელიც ჩვენ შემთხვევაში ტოლია 199,1454.
- **p** (**F**-კრიტერიუმის მნიშვნელობების დონე). ჩვენ შემთხვევაში  $p < 0.0000$ .

უილკის  $\lambda$  (ლამბდა) სტატისტიკა განისაზღვრება, როგორც შიგაჯგუფური დისპერსიის/კოვარიაციის მატრიცის ფარდობა საერთო კოვარიაციული მატრიცის დეტერმინანტთან. ლამბდა იცვლის [0,1] ინტერვალში. თუ ლამბდა მნიშვნელობა ახლოსაა ნულთან, მაშინ ადგილი აქვს კარგ დისკრიმინაციას, ხოლო თუ ლამბდა ახლოს არის ერთან, მაშინ საქმე გვაქვს ცუდ დისკრიმინაციასთან.

- **click summary: variables in the model** (შედეგი: მოდელში ჩართული ცვლადები)-ის ღილაკზე, ეკრანზე გამოდის მონაცემთა ანალიზის შემაჯამებელი შედეგის ცხრილი,

Discriminant Function Analysis Summary (Irisdat)						
No. of vars in model: 4; Grouping: IRISTYPE (3 grps)						
Wilks' Lambda: ,02344 approx. F (8,288)=199,15 p<0,0000						
N=150	Wilks' Lambda	Partial Lambda	F-remove (2,144)	p-level	Toler.	1-Toler. (R-Sqr.)
SEPALLEN	0,024976	0,938464	4,72115	0,010329	0,347993	0,652007
SEPALWID	0,030580	0,766480	21,93593	0,000000	0,608859	0,391141
PETALLEN	0,035025	0,669206	35,59018	0,000000	0,365126	0,634874
PETALWID	0,031546	0,743001	24,90433	0,000000	0,649314	0,350686

სადაც პირველ სვეტში მოყვანილია ლამბდას მნიშვნელობები, რომელებიც წარმოადგენენ მოდელიდან შესაბამის ცვლადების გამორიცხვის შედეგს. რაც მეტია ლამბდა მით უფრო სასურველია ამ ცვლადის დატოვება მოდელში.

**Partial lambda** (კერძოდ ლამბდა) მნიშვნელობა წარმოადგენს ფარდობას უილკის ლამბდა მნიშვნელობისა, რომელიც გამოითვლება შესაბამისი ცვლადის დამატების შემდეგ, ცვლადის დამატებამდე არსებული უილკის ლამბდას მნიშვნელობასთან. კერძოდ ლამბდა ახასიათებს შესაბამისი ცვლადის წვლილს მოდელის დაყოფის ძალაზე. კერძოდ, რაც უფრო ნაკლებია ლამბდა, მით უფრო მეტია ამ ცვლადის წვლილი საერთო დისკრიმინაციაში. მაგალითად, ცხრილიში ჩანს, რომ **Petalen** ცვლადი გვაძლევს ყველაზე დიდ წვლილს, ცვლადი **Petalwid** მეორეა თავისი წვლილით, ცვლადი **Petalvvid** მესამეა და ა.შ. აქედან გამომდინარეობს, რომ **Petalen** ცვლადი წარმოადგენს მთავარ ცვლადს, რომელიც საშუალებას გვაძლევს ჩავატაროთ ირისების სხვადასხვა ტიპებზე დისკრიმინაცია.

**F-remove** (**F**-გამორიცხვა) არის **F** -კრიტიკული მნიშვნელობა, რომელიც დაკავშირებულია შესაბამის კერძოდ ლამბდაზე. **P-level** მნიშვნელობა წარმოადგენს **F - remove** კრიტერიუმის მნიშვნელოვნების დონეს.

ტოლერანტობა (**Toler**) განისაზღვრება, როგორ  $1-R^2$ , სადაც  $R^2$  - ამ ცვლადს დანარჩენ ცვლადებთან მრავლობითი კორელაციის კოეფიციენტი. დავბრუნდეთ დისკრიმინაციულ ანალიზის შედეგების ფანჯარაში და ჩავრთოთ **Advanced** ჩანართი. **Distances between groups** (ჯგუფებს შორის მანძილი) ღილაკის დანიშნულებაა ეკრანზე გამოიტანოს ჯგუფებს შორის მანძილების ცხრილებს

IRISTYPE	Squared Mahalanobis Distances (Irisdat)		
	SETOSA	VERSICOL	VIRGINIC
SETOSA	0,0000	89,86419	179,3847
VERSICOL	89,8642	0,00000	17,2011
VIRGINIC	179,3847	17,20107	0,0000

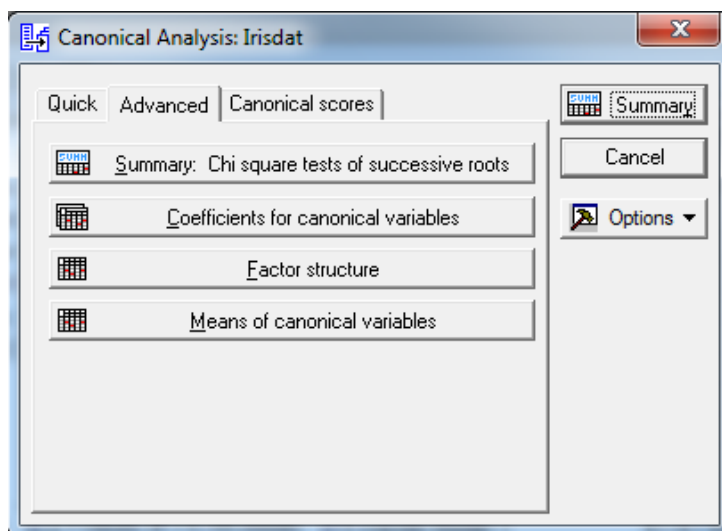
IRISTYPE	F-values; df = 4,144 (Irisdat)		
	SETOSA	VERSICOL	VIRGINIC
SETOSA		550,1889	1098,274
VERSICOL	550,189		105,313
VIRGINIC	1098,274	105,3127	

IRISTYPE	p-levels (Irisdat)		
	SETOSA	VERSICOL	VIRGINIC
SETOSA		0,00	0,00
VERSICOL	0,00		0,00
VIRGINIC	0,00	0,00	

დისკრიმინაციის შემდგომში ანალიზისთვის საჭიროა ჩატარდეს კანონიკური ანალიზი, რათა დავინახოთ, თუ როგორ დაიყოფიან დაჯგუფებული ცვლადები (ჩვენ მაგალითში ირისის ტიპები) შერჩეული დამოუკიდებელი ცვლადებით. ამისთვის უნდა განისაზღვროს დისკრიმინაციული ფუნქციები. ყოველ შემდგომ დისკრიმინაციულ ფუნქციას სულ უფრო ნაკლები წვლილი შეაქვს დისკრიმინაციის პროცესში.

დისკრეტული ფუნქციის მაქსიმალური რაოდენობა ტოლია ცვლადების რაოდენობას მინუს ერთი ან ჯგუფების რაოდენობას (ჩვენ შემთხვევაში 3) მინუს ერთ. იმისდა მიხედვით რომელია მინიმალური, ჩვენ შემთხვევაში გვექნება  $3-1 = 2$  დისკრიმინალტული ფუნქცია.

click **Perform canonical analysis** (კანონიკური ანალიზის შესრულება)-ის ღილაკზე. პროგრამა გამოთვლის დამოუკიდებელ (ორთოგონალურ) დისკრიმინანტულ ფუნქციებს.



ეკრანზე გამოსულ **Canonical Analysis**-ის ფანჯარის **Advanced** ჩანართში click **Summary: chi square tests of successive roots** (თანმიმდევრული ფესვების სიკვადრატ კრიტერიუმის შედეგები)-ის ღილაკზე.

Chi-Square Tests with Successive Roots Removed (Irisdat)						
Roots Removed	Eigen-value	Canonical R	Wilks' Lambda	Chi-Sqr.	df	p-level
0	32.19193	0.984821	0.023439	546.1153	8	0.000000
1	0.28539	0.471197	0.777973	36.5297	3	0.000000

ეკრანზე გამოდის ცხრილი, სადაც პირველ სტრიქონში მოცემულია ყველა ფესვის სარწმუნოების კრიტერიუმი, მეორე სტრიქონში მოცემულია ფესვების სარწმუნოების კოეფიციენტი, როცა პირველი ფესვი გამორიცხა ანალიზიდან და ა.შ. ამრიგად ცხრილი გვიჩვენებს თუ რამდენი ფესვია სარწმუნო. ჩვენი მაგალითის ცხრილიდან ჩანს, რომ ორივე დისკრეტული ფუნქცია სტატისტიკურად სარწმუნოა.

– თუ მოვახდენთ click **Coefficients for canonical variables** (კანონური ცვლადების კოეფიციენტები) ღილაკზე, მაშინ ეკრანზე გამოდის ორი ცხრილი დისკრიმინანტული ფუნქციის კოეფიციენტებით.

Variable	Raw Coefficients (Irisdat) for Canonical Variables	
	Root 1	Root 2
SEPALLEN	0,82938	0,02410
SEPALWID	1,53447	2,16452
PETALLEN	-2,20121	-0,93192
PETALWID	-2,81046	2,83919
Constant	2,10511	-6,66147
Eigenval	32,19193	0,28539
Cum.Prop	0,99121	1,00000

Variable	Standardized Coefficients (Irisdat) for Canonical Variables	
	Root 1	Root 2
SEPALLEN	0,42695	0,012408
SEPALWID	0,52124	0,735261
PETALLEN	-0,94726	-0,401038
PETALWID	-0,57516	0,581040
Eigenval	32,19193	0,285391
Cum.Prop	0,99121	1,000000

პირველ ცხრილში მოცემულია დისკრიმინანტული ფუნქციების საწყისი (არასტანდარტიზირებული) კოეფიციენტები, მეორე ცხრილში მოყვანილია დისკრიმინანტული ფუნქციების სტანდარტიზირებული კოეფიციენტები. ცხრილიდან ჩანს, რომ ყველაზე დიდი წვლილი პირველ (**Root 1**) დისკრიმინანტულ ფუნქციაში შეაქვს **Petalen**, **Petalwid** ცვლადებს, ხოლო მეორე (**Root2**) დისკრიმინანტულ ფუნქციაში – **Sepalwid**, **Petalwid**. ცხრილში აგრეთვე მოყვანილია თითოეული დისკრიმინანტული ფუნქციისთვის საკუთრივი მნიშვნელობები (**Eigenval**) და ყოველი ფუნქციისათვის დაგროვილი დისპერსიის ჯამური (კუმულატიური) (**Cum. Prop**) წილი. როგორც ცხრილიდან ჩანს, ფუნქცია **Root 1** შეესაბამება ჯამური დისპერსიის 99,1% ამიტომ ეს დისკრიმინანტული ფუნქცია ყველაზე „მნიშვნელოვანია“.

– დავბრუნდეთ **Canonical Analysis** ფანჯარაში და **click Means of canonical variables** (საშუალო კანონიკური ცვლადები) ღილაკზე. ეკრანზე გამოდის ცხრილი, სადაც მოყვანილია დისკრიმინანტული ფუნქციის საშუალო მნიშვნელობები, რომლებიც საშუალებას იძლევიან განსაზღვრონ ჯგუფები.

Group	Means of Canonical Variables (Irisdat)	
	Root 1	Root 2
SETOSA	7,60760	0,215133
VERSICOL	-1,82505	-0,727900
VIRGINIC	-5,78255	0,512767

როგორც ცხრილიდან ჩანს, პირველი დისკრიმინანტული ფუნქცია ძირითადად ახდენს **Setosa** ტიპის ირისების იდენტიფიკაციას, ხოლო მეორე დისკრიმინანტული ფუნქცია – **Varsicol** ტიპს..

- დავბრუნდეთ **Canonical Analysis** ფანჯარაში და ჩავრთოთ **Canonical**

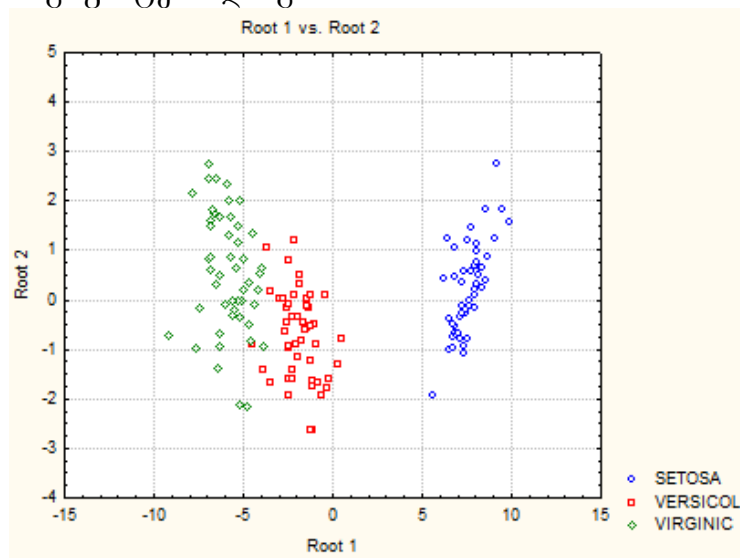


*scores* (კანონიკური მნიშვნელობები)-ის ჩანართი და *click Canonical scores for each case* (თითოეული დანაკვირის კანონიკური მნიშვნელობა)-ის დილაკზე. ეკრანზე გამოდის ცხრილი. სადაც წარმოდგენილია თითოეული დაკვირვებებისთვის დისკრიმინანტული ფუნქციის მნიშვნელობა.

Case	Unstandardized Canonical Scores (Irisdat)		
	Group	Root 1	Root 2
1	SETOSA	7,67197	-0,13489
2	VIRGINIC	-6,80015	0,58090
3	VERSICOL	-2,54868	-0,47220
4	VIRGINIC	-6,65309	1,80532
5	VIRGINIC	-3,81516	-0,94299
6	SETOSA	7,21262	0,35584
7	VIRGINIC	-5,10556	1,99218
8	VERSICOL	-3,49805	-1,68496
9	VERSICOL	-3,71590	1,04451
10	SETOSA	8,68104	0,87759
11	VERSICOL	-2,29249	-0,33286
12	VERSICOL	-4,49847	-0,88275
13	VIRGINIC	-4,96774	0,82114
14	VERSICOL	-1,09043	-1,62658
15	VIRGINIC	-5,06601	-0,02627

ამ ცხრილით რთულია ვიმსჯელოთ ირისების ჯგუფებში გადანაწილებაზე. უფრო მოსახერხებელია გაფანტვის დიაგრამის გამოყენება. ამისთვის:

– დავბრუნდეთ *Canonical Analysis ფანჯარაში* და *click Scatterplot of canonical scores* (კანონიკური მნიშვნელობების გაფანტვის დიაგრამა)-ის დილაკზე. ეკრანზე გამოდის ორი დისკრიმინანტული ფუნქციის კანონიკური მნიშვნელობების გაფანტვის დიაგრამა.



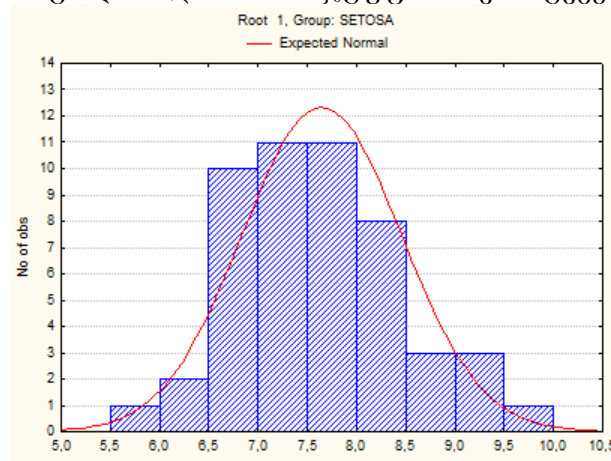
ცხრილიდან ჩანს, რომ ირისები, რომლებიც მიეკუთვნებიან ერთგვაროვან ჯგუფს (ჯიშს), ლოკალიზებული არიან გრაფიკის სიბრტყის გარკვეულ ნაწილში. ამავე დროს *Versicol* და *Virginic* ჯგუფების ცენტრებს შორის მანძლი გაცილებით ნაკლებია, ვიდრე *Setosa* და *Versicol* ჯგუფებს შორის და აგრეთვე *Setosa* და *Virginic* ჯგუფებს შორის. ეს იმას მიგვანიშნებს, რომ

ე.ყუბანეიშვილი. მონაცემების სტატისტიკური დამუშავება

*Versicol* და *Virginic* ირისის ჯგუფები მსგავსნი არიან, ხოლო *Setosa* ჯიში მკვეთრად განსხვავდება დანარჩენი ორი ჯიშისაგან.

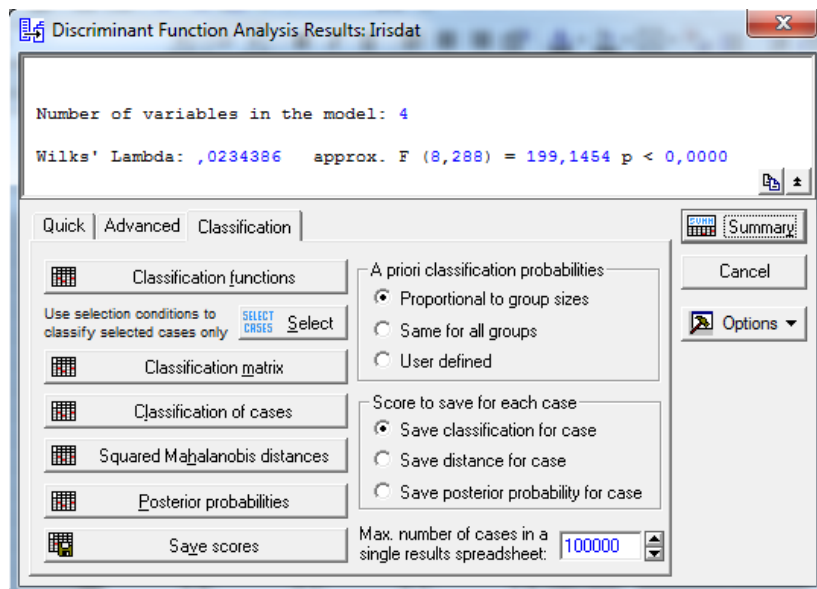
დისკრიმინანტული ფუნქცია 1 გააჩნია უარყოფითი კოეფიციენტები *Petalwid* და *Petalen* ცვალადებისთვის, ხოლო *Sepalwid* და *Sepallen* ცვალადებისთვის – დადებითი კოეფიციენტები. ამრიგად, რაც უფრო განიერია და გრძელი ყვავილის ბუტკო, ხოლო ფოთლის ჯამის სიგრძე და სიგანე მცირეა, მით უფრო მცირეა იმის ალბათობა, რომ ეს ყვავილი მიეკუთვნება *Setosa* ჯიშს,

**Canonical Analysis** ფანჯარაში არსებული **by group** ღილაკის დანიშნულებაა ჯგუფების კანონიკური მნიშვნელობების ჰისტოგრამების გამოტანა ეკრანზე. მაგალითად, *Setosa* ჯგუფისათვის გვექნება:



**All groups combined** ღილაკს ეკრანზე გამოაქვს ყველა ჯგუფის კომბინირებული ჰისტოგრამა.

დავბრუნდეთ **Discriminant Function Analysis Results** ფანჯარაში და ჩავრთოთ **Classification** ჩანართი. გაიხსნება კლასიფიკაციის შედეგების ფანჯარა



სადაც: *A priori classification probabilities* (კლასიფიკაციის აპიორული ალბათობები) ჩარჩოში მოყვანილ სხვადასხვა ოპციებში მოცემულია აპრიორული ალბათობები იმისა, რომ დაკვირვება კლასიფიკაციის დროს მოხვდება ერთ-ერთ ჯგუფში:

- *Proportional to groups sizes* (ჯგუფების პროპორციული ზომები);
- *Same for all groups* (ყველა ჯგუფებისთვის ერთნაირია);
- *User defined* (მომხმარებლის დანიშნულება).

გაჩუმების პრინციპით ჩართულია პირველი ოპცია.

*Score to save for each case* (დავიმახსოვროთ თითოეული დაკვირვებების) ჩარჩოში მოყვანილი ოპციები, რომელთა შერჩევის შედეგად შესაძლებელია ამა თუ იმ კლასიფიკაციის შედეგის შენახვა:

- *Save classification for case* (კლასიფიკაციის შედეგის შენახვა);
- *Save distance for case* (შევინახოთ მანძილები);
- *Save posterior probability for case* (აპოსტერიორული ალბათობის შენახვა).

გაჩუმების პრინციპით ჩართულია პირველი ოპცია.

*Max. number of cases in a single results spreadsheet* (შედეგების ცხრილში მასიმალურ დაკვირვებათა რიცხვი) ველში გაჩუმების პრინციპით დაყენებულია 100 000.

– *click Classification functions* (კლასიფიკაციის ფუნქციები)-ის ღილაკზე.

Variable	Classification Functions; grouping: IRISTYPE (Irisdat)		
	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
SEPALLEN	23,5442	15,6982	12,446
SEPALWID	23,5879	7,0725	3,685
PETALLEN	-16,4306	5,2115	12,767
PETALWID	-17,3984	6,4342	21,079
Constant	-86,3085	-72,8526	-104,368

ეკრანზე გამოსული ცხრილში წარმოდგენილია დისკრიმინანტული ფუნქციის თავისუფალი წევრი და კოეფიციენტები. მაგალითად, *Setosa* ჯგუფის დისკრიმინანტულ (ანუ კლასიფიკაციურ) ფუნქციას აქვს შემდეგი სახე:

$$Setosa = 23.54 \text{ sepalen} + 23.588 * \text{sepalwid} - 16.43 * \text{petallen} - 17.30 * \text{petalwid} - 86.31.$$

კლასიფიკაციური ფუნქცია წრფივი ფუნქციაა, რომელიც გამოითვლება ყოველი ჯგუფებისთვის და გამოიყენებინ დაკვირვებათა კლასიფიკაციისთვის. დაკვირვება მიეკუთვნება იმ ჯგუფს, რომელთანაც კლასიფიკაციის ფუნქციას გაჩნია უდიდესი მნიშვნელობა.

– *click Classification matrix* (კლასიფიკაციის მატრიცა)-ის ღილაკზე. ეკრანზე გამოსულ ცხრილში ნაჩვენებია თითოეულ ჯგუფში კლასიფიცირებული დაკვირვებათა რაოდენობა და პროცენტები.

Classification Matrix (Irisdat)				
Rows: Observed classifications				
Columns: Predicted classifications				
Group	Percent Correct	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
SETOSA	100,0000	50	0	0
VERSICOL	96,0000	0	48	2
VIRGINIC	98,0000	0	1	49
Total	98,0000	50	49	51

ცხრილის სტრიქონები – საწყისი კალსებია, ხოლო სვეტები – კლასებში მოხვედრილი დაკვირვებათა რაოდენობა.

– *click Classification of cases* (დაკვირვებათა კლასიფიკაცია)-ის ღილაკზე. პროგრამას ეკრანზე გამოაქვს თითოეული დაკვირვების კლასიფიკაცია.

Classification of Cases (Irisdat)				
Incorrect classifications are marked with *				
Case	Observed Classif.	1 p=,33333	2 p=,33333	3 p=,33333
1	SETOSA	SETOSA	VERSICOL	VIRGINIC
2	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
3	VERSICOL	VERSICOL	VIRGINIC	SETOSA
4	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
* 5	VIRGINIC	VERSICOL	VIRGINIC	SETOSA
6	SETOSA	SETOSA	VERSICOL	VIRGINIC
7	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
8	VERSICOL	VERSICOL	VIRGINIC	SETOSA
* 9	VERSICOL	VIRGINIC	VERSICOL	SETOSA
10	SETOSA	SETOSA	VERSICOL	VIRGINIC
11	VERSICOL	VERSICOL	VIRGINIC	SETOSA
* 12	VERSICOL	VIRGINIC	VERSICOL	SETOSA
13	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
14	VERSICOL	VERSICOL	VIRGINIC	SETOSA
15	VIRGINIC	VIRGINIC	VERSICOL	SETOSA

ცხრილის პირველ სვეტში მოცემულია ჯგუფი, რომლისთვისაც შესაბამის დაკვირვებას გააჩნია უდიდესი აპოსტერიორული ალბათობა და კლასიფიკაციის ფუნქციის მაქსიმალური მნიშვნელობა. დაკვირვებები, რომელთა კლასიფიკაცია ვერ მოხერხდა აღნიშნულ \* სიმბოლოთი.

– *click Squared Mahalanobis distances* (მახანალობის მანძლის კვადრატები)-ის ღილაკზე. ეკრანზე გამოდის ცხრილი,

Squared Mahalanobis Distances from Group Centroids (Irisdat)					
Incorrect classifications are marked with *					
Case	Observed Classif.	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333	
1	SETOSA	0,2419	90,6602	181,5587	
2	VIRGINIC	208,5713	27,3188	1,8944	
3	VERSICOL	105,2663	2,2329	13,0720	
4	VIRGINIC	207,9180	31,7492	4,4506	
* 5	VIRGINIC	133,0668	5,2529	7,2359	
6	SETOSA	1,3337	84,0118	170,0569	
7	VIRGINIC	173,1838	26,5620	11,0484	
8	VERSICOL	131,6617	8,4307	14,7647	
* 9	VERSICOL	130,8624	8,6697	6,5068	
10	SETOSA	2,2864	113,6509	210,0239	
11	VERSICOL	99,2338	1,2963	13,8174	
* 12	VERSICOL	149,0303	8,4393	4,8645	
13	VIRGINIC	158,9817	12,7512	1,2342	
14	VERSICOL	79,1079	1,4076	26,6531	
15	VIRGINIC	161,8529	12,1703	1,9781	

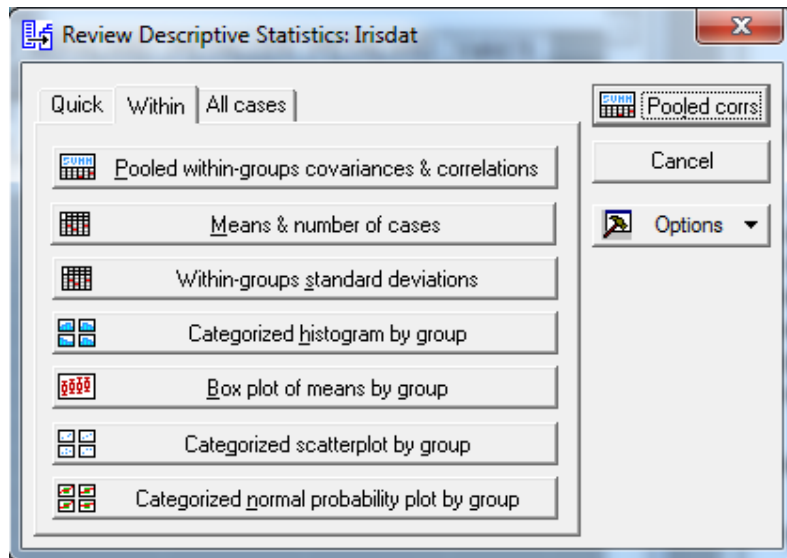
სადაც მოყვანილია თითოეული დაკვირვების მახანალობის მანძილი ჯგუფის ცენტრთან, ეს მანძილები ანალოგიურია ეკლიდეს მანძილის კვადრატისა, მაგრამ ის ითვალისწინებს მოდელში ცვლადებს შორის კორელაციას. დაკვირვება მიეკუთვნება იმ ჯგუფს, რომელთანაც მას გაჩნია მინიმალური მანძილი, კლასიფიკაციის გარეშე დარჩენილი დაკვირვებები აღნიშნება \* სიმბოლოთი.

- *click Posterior probabilities* (აპოსტერიული ალბათობა)-ის დილაკზე ეკრანზე გამოდის ცხრილი, სადაც თითოეული დაკვირვებისათვის ნაჩვენებია ჯგუფებში ჩართვის აპოსტერიული ალბათობა, რომელიც განსაზღვრულია აპრიორული ალბათობით და მახანალობის მანძილით. დაკვირვება მიეკუთვნება იმ ჯგუფს, რომელთანაც მას გაჩნია მაქსიმალური აპოსტერიული ალბათობა.

ახალი მონაცემების დისკრიმინაციისთვის საწყის მონაცემთა ცხრილს უნდა დაუმატოთ ახალი დაკვირვება. იმისთვის, რომ გავიგოთ თუ რომელ კლასს მიეკუთვნება ახალი დაკვირვება *click Posterior probabilities* დილაკზე ეკრანზე გამოდის აპოსტერიული ალბათობის ცხრილი დამატებული ახალი დაკვირვებებით.

თუ დავბრუნდებით *Madel definition* ფაჯნარაში (ამისათვის *click Cancel* დილაკზე) და ჩავრთავთ *Descriptive* ჩანართს, მაშინ პროგრამა საშუალებას იძლევა ფართდ გამოვიყენოთ აღწერითი სტატისტიკის მეთოდები.

- *click Review descriptive statistics* დილაკზე ეკრანზე გამოდის შესაბამისი დასახელების ფანჯარა.



**Quick** ჩანართში *click Pooled within-groups covariances & correlations* (გამეორებული ჯგუფშიგა კოვარიაცია და კორელაცია) და *Means & number of cases* (საშუალებები და დაკვირვებათა რაოდენობა) ღილაკებზე. ეკრანზე გამოდის შესაბამისი ცხრილები სტატისტიკური მახასიათებლებით.

Variable	Pooled Within-Groups Correlations (Irisdat)			
	SEPALLEN	SEPALWID	PETALLEN	PETALWID
SEPALLEN	1,00	0,53	0,76	0,36
SEPALWID	0,53	1,00	0,38	0,47
PETALLEN	0,76	0,38	1,00	0,48
PETALWID	0,36	0,47	0,48	1,00

IRISTYPE	Means (Irisdat)				Valid N
	SEPALLEN	SEPALWID	PETALLEN	PETALWID	
SETOSA	5,006000	3,428000	1,462000	0,246000	50
VERSICOL	5,936000	2,770000	4,260000	1,326000	50
VIRGINIC	6,588000	2,974000	5,552000	2,026000	50
All Grps	5,843333	3,057333	3,758000	1,199333	150

**Within** (შიგა) ჩანართის ჩართვით შესაძლებელია ეკრანზე გამოვიტანოთ შემდეგი ინფორმაცია:

- **Within-groups standard deviations** (შიგაჯგუფური სტანდარტული გადახრა);
- **Categorized histogram by group** (ჯგუფური კატეგორიზებული ჰისტოგრამა);
- **Box plot of means by group** (გაქანების დიაგრამა);
- **Categorized scatterplot by group** (გაფანტვის კატეგორიზებული დიაგრამები);
- **Categorized normal probability plot by group** (ჯგუფების კატეგორიზებული ნორმალური გრაფიკი).

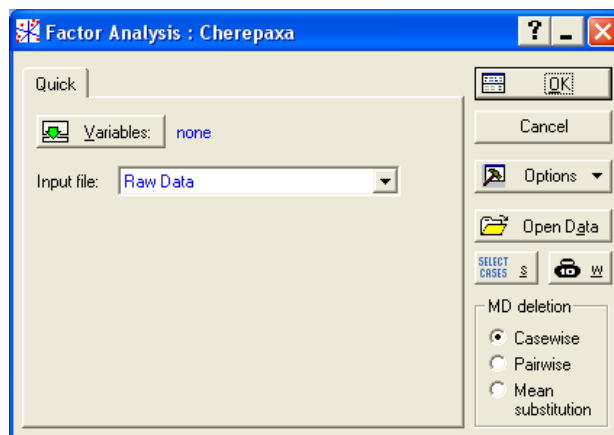
*All cases* (ყველა დაკვირვება) ჩანართი საშალებას იძლევა შემდეგი მონაცემების გამოსატანად:

- **Total covariances & correlations** (სრული (საერთ) კოვარიაცია და კორელაცია);
- **Plot of total correlations** (კორელაციის სრული გრაფიკი);
- **Box plot of means** (საშუალოებით გაქანების დიაგრამა).

## პრაქტიკული სამუშაო 8

### ფაქტორული ანალიზი

**Statistics** მენიუში შევარჩიოთ **Multivariate Exploratory Techniques** (მრავლობითი კვლევის მეთოდები) და **click Factor Analysis** (ფაქტორული ანალიზი)-ის ბრძანებაზე.



– ეკრანზე გამოდის **Factor Analysis** სასტარტო ფანჯარა. **Input File** (შემაავალი მონაცემების ფაილები) ველში უნდა შევარჩიოთ :

- **Raw data** (საწყისი მონაცემები);
- **Correlation Matrix** (კოლევაციური მატრიცა).

შევარჩიოთ **Raw Data**. ეს არის ჩვეულებრივი მონაცემთა ფაილი, სადაც სტრიქონებში ჩაწერილია ცვლადების მნიშვნელობები. ფანჯრის მარჯვენა ქვედა **MD deletion** (გამოტოვებულ მნიშვნელობების დამუშავების) ჩარჩოში უნდა ჩავრთოთ ერთ-ერთი რომელიმე ალმი:

- **Casewise** (გამოტოვებული შემთხვევების გამორიცხვის მეთოდი);
- **Pairwise** (გამოტოვებული მნიშვნელობების გამორიცხვა წყვილი მეთოდით);

• **Mean substitution** (გამოტოვებული მნიშვნელობების მაგივრად საშუალო მნიშვნელობის ჩასმა).

**Casewise** მეთოდის დროს ელექტრონულ ცხრილში იგნორირებული იქნება ყველა ის სტრიქონი, რომელიც შეიცავს თუნდაც ერთ გამოტოვებულ მნიშვნელობას.

**Pairwise** მეთოდის დროს გამოტოვებული მნიშვნელობის იგნორირება ხდება არა ყველა ცვლადისათვის, არამედ მხოლოდ შერჩეული წყვილი ცვლადისათვის. მაგალითად, როდესაც ხდება კორელაციურ მატრიცაში წყვილი ცვლადების ანალიზი.

**Mean substitution** მეთოდით ხდება გამოტოვებული მნიშვნელობის მაგივრად საშუალო სიდიდის ჩაწერა.

შევარჩით მაგ. **Casewise**. ჩავტვირთდ (თუ წინასწარ ჩატვირთლი არ არის) მონაცემთა ფაილი.

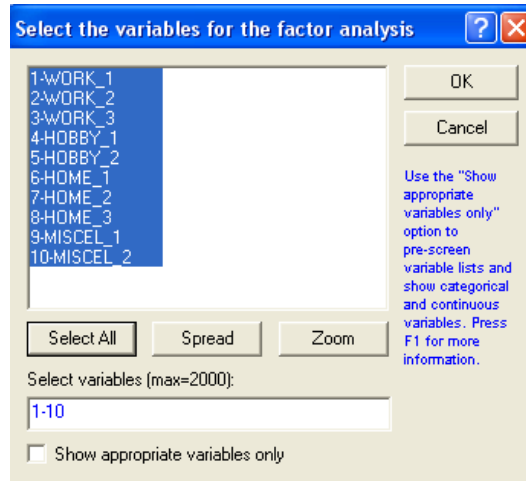
მაგალითისთვის განვიხილოთ **Examles** ბიბლიოთეკაში არსებული **Factor.sta** მონაცემთა ფაილი, სადაც თავმოყრილია 100 რესპონდენტის გამოკითხვის შედეგები ცხოვრების დონის კმაყოფილების ხარისხის შესახებ. ფაილი შედგება 10 ცვლადისგან:

- **Work 1** (სამუშაოთი დაკმაყოფილება), პირველი კომპონენტი;
- **Work 2** (სამუშაოთი დაკმაყოფილება), მეორე კომპონენტი;
- **Work 3** (სამუშაოთი დაკმაყოფილება), მესამე კომპონენტი;
- **Hobby 1** (თვისუფალი დროით დაკმაყოფილება), პირველი კომპონენტი;
- **Hobby 2** (თვისუფალი დროით დაკმაყოფილება), მეორე კომპონენტი;
- **Home 1** (საოჯახო ცხოვრებით დაკმაყოფილება), პირველი კომპონენტი;
- **Home 2** (საოჯახო ცხოვრებით დაკმაყოფილება), მეორე კომპონენტი;
- **Home 3** (საოჯახო ცხოვრებით დაკმაყოფილება), მესამე კომპონენტი;
- **Miscel 1** (საერთო დაკმაყოფილება), პირველი კომპონენტი;
- **Miscel 2** (საერთო დაკმაყოფილება), მეორე კომპონენტი;

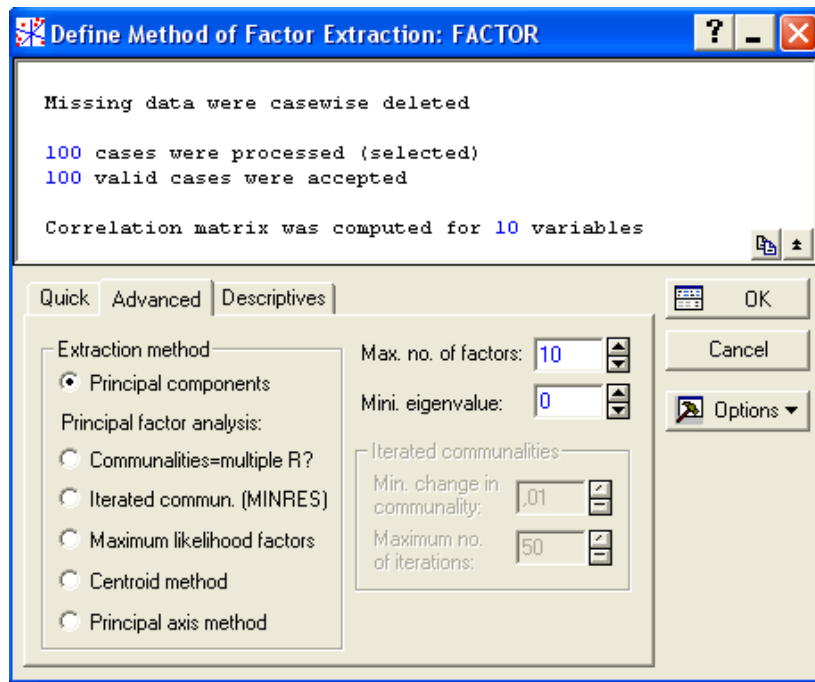
თითველი ცვლადის მრავლობითობა აისახება იმით, რომ განიხილება დაკმაყოფილებულობის სხვადასხვა ასპექტი, მაგალითად, შესაძლებელია პიროვნება დაკმაყოფილებული იყოს ჯამაგირით, მაგრამ არ იყოს კმაყოფილი კოლექტივით ან სამსახურის სახლიდან დაშორების მანძილის გამო და ა.შ. ცვლადების მნიშვნელობები ისეა გარდაქმნილი, რომ დაკმაყოფილების საშუალო ხარისხი შეესაბამება 100-თან ახლოს მდგომ მნიშვნელობას, დაბალი და მაღალი დაკმაყოფილებულობის ხარისხის მნიშვნელობა შესაბამისად - 100 ნაკლებს და 100 მეტს.

– **click Variables** დილაკზე. ეკრანზე გამოდის **Select the variables for the factor analasis** (ფაქტორული ანალიზისთვის ცვლადების შეჩვენა) ფანჯარა. ცვლადების მონიშვისათვის **click Select All** (ყველა შევარჩიოთ) დილაკზე და შემდეგ **OK**.



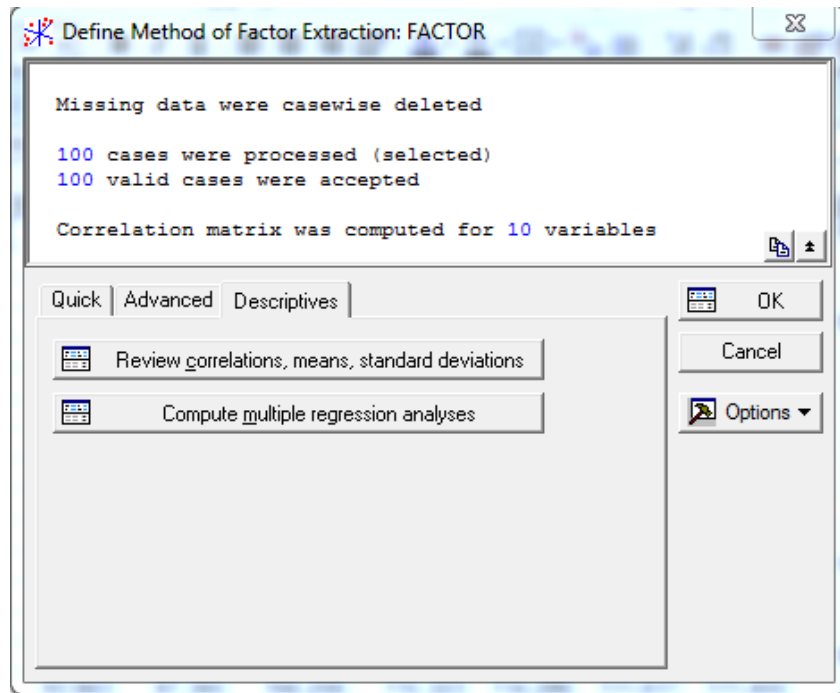


- ვუბრუნდებით **Factor Analysis** სასტატო ფანჯარაში. *click OK*.
- ეკრანზე გამოდის **Define Method of Factor Extraction** (ფაქტორების გამოყოფის მეთდის შერჩევა) ფანჯარა.



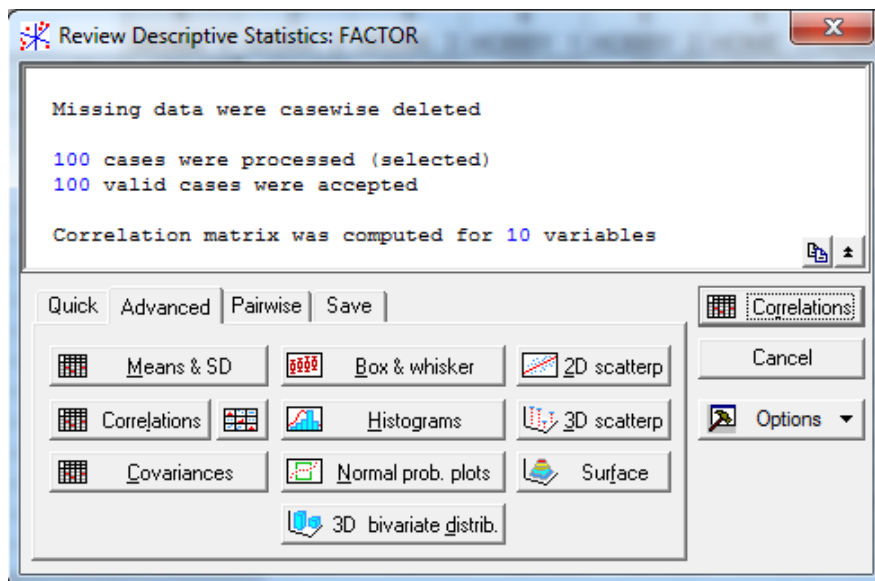
ფანჯრის ინფორმაციული ნაწილში ნაჩვენებია, რომ გამოტოვებული მნიშვნელობა დამუშავებულია **Casewise** მეთდით. კორელაციური მატრიცა განსაზღვრულია (მითითებული რაოდენობა 10) ცვლადებისთვის.

ფანჯრის ქვედა ნაწილი შედგება სამი ჩანართდან. ჩაერთოდ **Descriptives** ჩანართ, რადგან ფაქტორული ანალიზი უნდა დავიწყოთ კორელაციური მატრიცის განსაზღვრით.



კორელაციური ანალიზით შესაძლებელია ცვლადებს შორის კორელაციის ხარისხის დადგენა. თუ ეს ხარისხი (ანუ კორელაციის კოეფიციენტი) მაღალია, მაშინ ეს ცვლადები შეიძლება გავერთანოთ ერთ ფაქტორში. ეკრანზე გამოჩნდება ორი ღილაკი: **Review correlations, means, standard deviations** ღილაკი განკუთვნილია კორელაციური მატრიცის ასაგებად, საშუალოებისა და სატანდარტული გადახრების გამოსათვლელად. ხოლო **Computer multiple regression analyses** ღილაკის საშუალებით ხდება მრავლობითი რეგრესიული ანალიზის გაშვების პროცესი.

– *click Review correlations, means, standard deviations* ღილაკზე. გაიხსნება **Review Descriptive Statistics** (აღწერითი სტატისტიკის დამუშავება) ფანჯარა,



სადაც **Quick (Advanced)** ჩანართში **click Corelations** ღილაკზე. ეკრანზე გამოვა კორელაციული მატრიცა,

Correlations (FACTOR) Casewise deletion of MD N=100										
Variable	WORK_1	WORK_2	WORK_3	HOBBY_1	HOBBY_2	HOME_1	HOME_2	HOME_3	MISCEL_1	MISCEL_2
WORK_1	1.00	0,65	0,65	0,60	0,52	0,14	0,15	0,14	0,61	0,55
WORK_2	0,65	1,00	0,73	0,69	0,70	0,14	0,18	0,24	0,71	0,68
WORK_3	0,65	0,73	1,00	0,64	0,63	0,16	0,24	0,25	0,70	0,67
HOBBY_1	0,60	0,69	0,64	1,00	0,80	0,54	0,63	0,58	0,90	0,84
HOBBY_2	0,52	0,70	0,63	0,80	1,00	0,51	0,50	0,48	0,81	0,76
HOME_1	0,14	0,14	0,16	0,54	0,51	1,00	0,66	0,59	0,50	0,42
HOME_2	0,15	0,18	0,24	0,63	0,50	0,66	1,00	0,73	0,64	0,59
HOME_3	0,14	0,24	0,25	0,58	0,48	0,59	0,73	1,00	0,59	0,52
MISCEL_1	0,61	0,71	0,70	0,90	0,81	0,50	0,64	0,59	1,00	0,84
MISCEL_2	0,55	0,68	0,67	0,84	0,76	0,42	0,59	0,52	0,84	1,00

სადაც ცხადათ ჩანს, რომ **Work** და **Home** ცვლადებს შორის კორელაციის კოეფიციენტი დაბალი სიდიდისაა, მაშინ როცა სხვა ჯგუფების ცვლადებთან დებულბენ მაღალ მნიშვნელობებს.

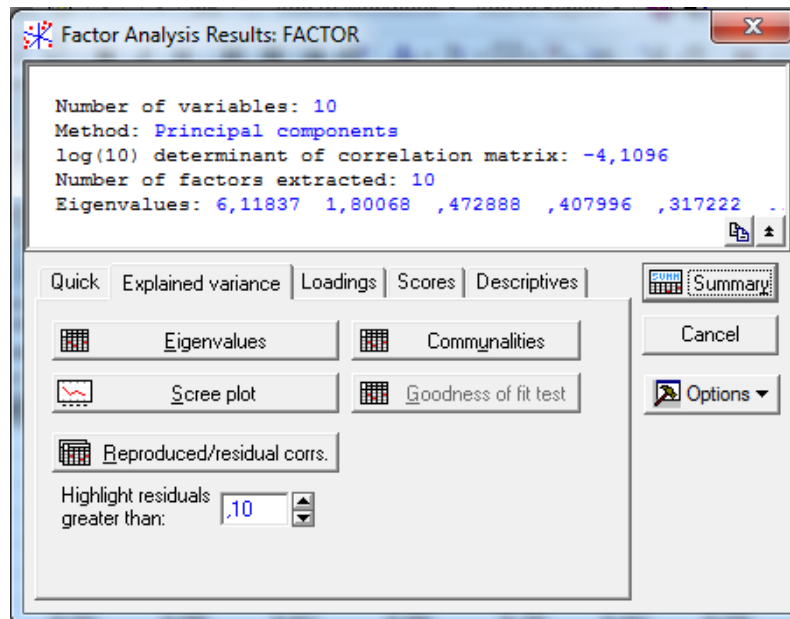
**Review Descriptive Statistics** ფანჯარაში **Clic Canccl** ღილაკზე. ვბრუნდებით **Define Method of Factor Extraction** ფანჯარაში. ჩავრთოდ **Advanced** ჩანართი, სადაც წარმოდგენილია შემდეგი ოპციები:

- **Max. no of factors** (ფაქტორების მაქსიმალური რაოდენობა);
- **Mini.engenvalue** (მინიმალური საკუთარი მნიშვნელობა), სადაც დაყენდება საკუთარი მნიშვნელობის მინიმალური მნიშვნელობა ე.ი. თუ რომელიმე საკუთარი მნიშვნელობა აღონჩნდება აქ დაყენებულ მნიშვნელობაზე ნაკლები, მაშინ ხდება მისი იგნორირება.

**Max.no of factors** ველში მომხმარებელი მიუთითებს იმ ფაქტორების რაოდენობას რომელთა გამოყოფაა საჭირო მოცემული მონაცემების ანალიზისთვის.

• გაჩუმების პრინციპით მთავარი კომპონენტების მეთოდში ყველა ცვლადის დისპერსიას იღებენ 1-ის ტოლად. მაშინ საერთო დისპერსია ცვლადების საერთო რაოდენობის ტოლია (ჩვენი მაგალითისათვის - 10). თითოეულ ფაქტორს შეესაბამება გარკვეული დისპერსია. დისპერსიებს, რომლებიც შეესაბამებიან ფაქტორებს, ეწოდებათ საკუთრივი მნიშვნელობები.

საკუთრივი მნიშვნელობების დათვალერებისთვის **Define Method of Factor Extractopn** ფანჯარაში დავაყენოთ **Max.on of factors = 10** და **Mini.engenvalue = 0**. შემდეგ **OK**.



– ეკრანზე გამოსულ **Factor Analysis Results** ფანჯარაში *click Eigenvalues* ღილაკზე. ეკრანზე გამოდის საკუთრივი მნიშვნელობების ცხრილი:

Eigenvalues (FACTOR)				
Extraction: Principal components				
Value	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	6,118369	61,18369	6,11837	61,1837
2	1,800682	18,00682	7,91905	79,1905
3	0,472888	4,72888	8,39194	83,9194
4	0,407996	4,07996	8,79993	87,9993
5	0,317222	3,17222	9,11716	91,1716
6	0,293300	2,93300	9,41046	94,1046
7	0,195808	1,95808	9,60626	96,0626
8	0,170431	1,70431	9,77670	97,7670
9	0,137970	1,37970	9,91467	99,1467
10	0,085334	0,85334	10,00000	100,0000

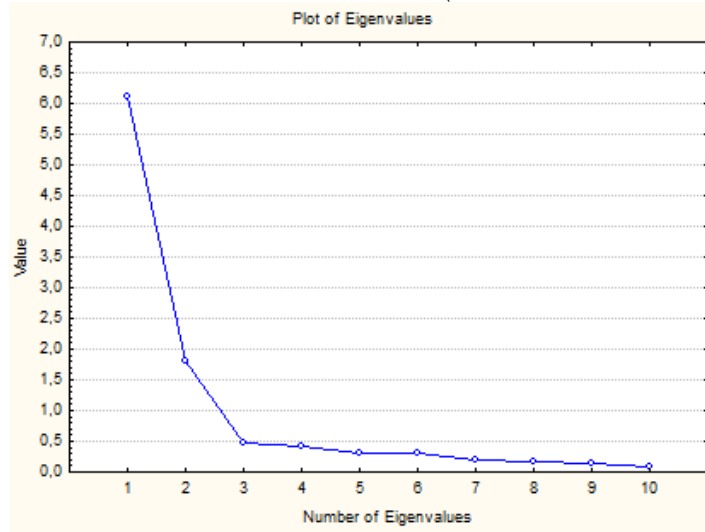
ცხრილის მეორე სვეტში წარმოდგენილია გამოყოფილი ფაქტორების დისპერსია - საკუთრივი მნიშვნელობები. მესამე სვეტში ყოველი ფაქტორისათვის მოცემულია საერთო დისპერსიდან პროცენტული მნიშვნელობა. როგორც ვხედავთ, პირველ ფაქტორზე მოდის საერთო დისპერსიის 61%, მეორე ფაქტორზე 18% და ა.შ. მეოთხე სვეტში წარმოდგენილია დაგროვილი ანუ კუმულატიური დისპერსიის მნიშვნელობა, ხოლო მეხუთე სვეტში მათი პროცენტული მნიშვნელობა.

შემდეგი ამოცანა მდგომარეობს რამდენი ფაქტორი უნდა დავტოვოთ ანალიზში. ამისთვის არსებობს, მაგალითად კაიხერის კრიტერიუმი.

**კაიხერის კრიტერიუმი.** დასაწყისში შეიძლება შევარჩიოთ მხოლოდ ის ფაქტორები, რომელთა დისპერსია ერთზე მეტია. ეს კრიტერიუმი, რომელიც კაიხერმა შემოგვთავაზა, ძალზე პოპულარულია. როგორც ცხრილიდან ჩანს,

ჩვენ მაგალითში გამოიყოფა მხოლოდ ორი ფაქტორი, რადგან მათი საკუთრივი მნიშვნელობები (დისპერსიები) ერთზე მეტია.

არსებობს გრაფიკული კრიტერიუმი, რომელიც შემოგვთავაზა კეტელმა. ამისთვის ჩავრთოთ **Explained variance** ჩანართი და **click Scree plot** ღილაკზე.



კეტელის მეთოდით გრაფიკზე ვპოულობთ ადგილს, სადაც საკუთრივი მნიშვნელობა გრაფიკის მარცხნიდან მარჯვნივ მაქსიმალურად მცირდება. როგორც ეკრანზე გამოსული გრაფიდიან ჩანს, კეტელის კრიტერიუმით შეიძლება გამოვყოთ 2 ან 3 ფაქტორი.

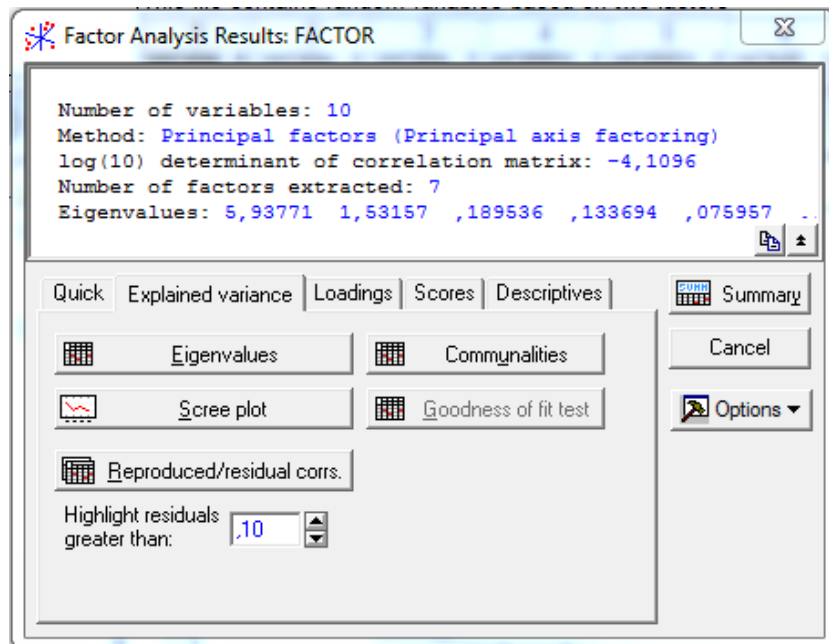
ფაქტორების გამოყოფის სხვადასხვა მეთოდები წარმოდგენილია **Define Method of Factor Extraction** ფანჯრის **Advanced** ჩანართის **Extraction method** (გამოყოფის მეთოდი)-ის ჩარჩოში.

ფაქტორების გამოყოფის ორი კრიტერიუმია: **Principal components** (მთავარი კომპონენტის მეთოდი) და **Principal factor analysis** (მთავარი ფაქტორების ანალიზი), სადაც წარმოდგენილია შემდეგი მეთოდები:

- **Communalities = multiple R<sup>2</sup>** (სპეციფიური დისპერსია, რომელიც ტოლია მრავლობით კორელაციის კოეფიციენტის კვადრატისა);
- **Iterated communalities (MINRES)** (იტერაციული სპეციფიური დისპერსია ან მინიმალური ნაშთები);
- **Maximum likelihood factors** (მაქსიმალური დასაჯერობა);
- **Centroid method** (ცენტროიდალური მეთოდი);
- **Principal axis method** (მთავარი ღერძების მეთოდი);

შევარჩოთ **Principal components** ოპცია, რათა კარგად გავეცნოთ ფაქტორულ ანალიზის ძირითად მომენტებს. წარმოვიდგინოთ, რომ ფაქტორების რაოდენობის შერჩევის კრიტერიუმი უცნობია და ამიტომ უნდა დავიწყოთ ფაქტორების მაქსიმალური რაოდენობით (10) და მინიმალური საკუთრივი მნიშვნელობით (ე.ი. უნდა ავიღოთ 0). ჩავრთოთ ოპცია და **OK**.

ეკრანზე გამოვა **Factor Analysis Results** ფანჯარა,



სადაც ზედა ინფორმაციული ნაწილში ნაჩვენებია:

- **Number of variables** (გასაანალიზირებადი ცვლადების რაოდენობა);
- **Method** (ანალიზის მეთოდი)
- **Log(10) determination of correlation matrix** (კორელაციული მატრიცის დეტერმინანტის ათობითი ლოგარითმი);
- **Number of factor extraction** (გამოყოფილი ფაქტორების რაოდენობა);
- **Eigenvalues** (საკუთრივი მნიშვნელობები);

– **click Summary** ღილაკზე. ეკრანზე გამოდის ფაქტორული დატვირთვის ცხრილი – გამოყოფილ ფაქტორებსა და ცვლადებს შორს კორელაციის მნიშვნელობები.

Variable	Factor Loadings (Unrotated) (FACTOR)									
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10
WORK_1	-0,652601	0,514217	0,301687	0,439108	-0,013701	0,127061	0,051149	-0,022306	0,080008	0,003894
WORK_2	-0,756976	0,494770	-0,078826	-0,211795	-0,090859	0,172261	-0,236057	0,163030	0,103633	0,012210
WORK_3	-0,745706	0,456680	-0,104749	0,030826	-0,204913	-0,422007	0,033818	0,019468	-0,017932	0,038980
HOBBY_1	-0,941630	-0,021835	0,012653	0,001861	0,120655	0,093740	-0,023837	0,001553	-0,243305	0,171990
HOBBY_2	-0,875615	0,051643	0,099675	-0,324541	-0,015852	0,091249	0,312973	-0,025504	0,088684	0,017996
HOME_1	-0,576062	-0,604977	0,490999	-0,114927	-0,112513	-0,114838	-0,145100	-0,023820	0,004027	-0,019576
HOME_2	-0,671289	-0,617962	-0,125776	0,159963	0,225012	-0,103741	0,028392	0,201046	0,145372	0,048318
HOME_3	-0,641532	-0,573925	-0,268572	0,152709	-0,362524	0,159987	0,011420	-0,079979	0,006890	0,000902
MISCEL_1	-0,951516	0,013513	-0,050164	0,026706	0,076795	0,012644	0,035938	0,095864	-0,156713	-0,223847
MISCEL_2	-0,900333	0,048154	-0,151805	-0,034832	0,226647	-0,050720	-0,120513	-0,292831	0,087690	-0,030324
Expl. Var	6,118369	1,800682	0,472888	0,407996	0,317222	0,293300	0,195808	0,170431	0,137970	0,085334
Prp. Totl	0,611837	0,180068	0,047289	0,040800	0,031722	0,029330	0,019581	0,017043	0,013797	0,008533

ცხრილიდან ჩანს, რომ პირველ და მეორე ფაქტორებს (**Factor 1, Factor 2**) გააჩნიათ სხვა ფაქტორებთან შედარებით კორელაციის კოეფიციენტის დიდი მნიშვნელობები. თანაც ფაქტორის რიგითი ნომრის გაზრდისას, კორელაციის კოეფიციენტის სიდიდე სწრაფად მცირდება. აქედან გამომდინარე,

არასასურველია განვიხილოთ ყველა 10 ფაქტორი. კაიზერის და კეტელის კრიტერიუმით უნდა დავტოვოთ მხოლოდ 2 ფაქტორი.

იმსითვის, რომ მივიღოთ ინტერპრეტირებული გადაწყვეტილება, საჭიროა დერძის გარშემო ფაქტორების ბრუნვა. პროგარაში გათვალისწინებულია ბრუნვის რამოდენიმე მეთოდი:

- *Varimax raw* (საწყისების ვარიმაქსი);
- *Varimax normalized* (ნორმალიზირებული ვარიმაქსი);
- *Biquartimax raw* (საწყისების ბიკვარტირმაქსი);
- *Biquartimax normalized* (ნორმალიზირებული ბიკვადრტიმაქსი);
- *Quartimax raw* (საწყისების კვარტიმაქსი);
- *Quartimax normalized* (ნორმალიზირებული კვარტიმაქსი);
- *Equamax raw* (საწყისების ექვიმაქსი);
- *Equamax normalized* (ნორმალიზირებული ექვიმაქსი);

ვარიმაქსის მეთოდი გამოიყენება საწყისი ფაქტორების დატვირთვების კვადრატების დისპერსიების მაქსიმიზაციისთვის, რაც ექვივალენტურია საწყისი ფაქტორული დატვირთვების კვადრატების მატრიცის სვეტების დისპერსიების მაქსიმიზაციისა.

ბიკვარტიმაქსი მეთოდის მიზანია საწყისი ფაქტორების დისპერსიების კვადრატების ჯამისა და საწყისი ფაქტორების ფაქტორული დატვირთვების ჯამის ერთდროული მაქსიმიზაცია. ეს ექვივალენტურია საწყისი ფაქტორული დატვირთვის მატრიცის სტრიქონების და სვეტების დისპერსიების ერთდროული მაქსიმიზაციისა.

კვარტიმაქსის მეთოდი ნიშნავს ფაქტორების მიხედვით თითოეული ცვლადისათვის ფაქტორული დატვირთვების კვადრატების დისპერსიების მაქსიმიზაციას, რაც ექვივალენტურია საწყისი ფაქტორული დატვირთვების მატრიცის სტრიქონის დისპერსიების მაქსიმიზაციისა.

ექვიმაქსის მეთოდი შეიძლება განვიხილოთ როგორც ვარიმაქსის და კვარტიმაქსის მეთოდების ნარევის შეწონილი ბრუნვა, რაც ექვივალენტურია საწყისი ფაქტორული დატვირთვის მატრიცის სტრიქონების და სვეტების დისპერსიების ერთდროული მაქსიმიზაციისა.

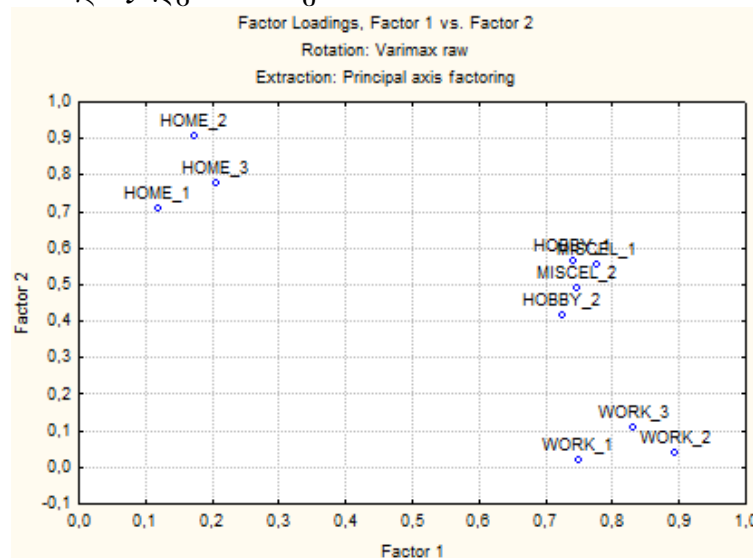
დამატებითი ტერმინი *Normalized* გვიჩვენებს იმას, რომ ფაქტორული დატვირთვები ნორმალიზირებულია ე.ი. ისინი იყოფიან შესაბამისი სპეციფიური დისპერსიიდან კვადრატული ფესვის მნიშვნელობაზე. ტერმინი *var* (საწყისები) ნიშნავს იმას, რომ მბრუნავი დატვირთვები არ არიან ნორმალიზირებულნი.

**Factor Analysis Results** ფანჯარის **Quick** ჩანართის **Factor rotation** გაშლად სიაში შევარჩიოთ ბრუნვის მეთოდი, მაგრალითად: **Varimax raw** და **click Summary** ღილაკზე.

Factor Loadings (Varimax raw) (FACTOR)							
Extraction: Principal axis factoring							
(Marked loadings are >,700000)							
Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
WORK_1	0,749741	0,018004	0,003387	0,299283	0,017650	0,015099	0,009042
WORK_2	0,895185	0,038609	0,059034	-0,097993	0,033573	0,054786	-0,047597
WORK_3	0,831912	0,105896	-0,031798	0,079777	-0,172209	-0,010121	-0,021807
HOBBY_1	0,741142	0,561829	0,102597	0,082015	0,218882	-0,031300	0,000477
HOBBY_2	0,725303	0,413904	0,367836	-0,081747	0,012236	-0,019083	0,024685
HOME_1	0,119248	0,708433	0,344782	0,106536	0,030750	0,029679	-0,028743
HOME_2	0,172653	0,904556	-0,009099	0,000970	-0,002921	-0,079990	0,040490
HOME_3	0,205369	0,778719	0,037318	-0,080662	-0,019138	0,101356	-0,062594
MISCEL_1	0,776034	0,556697	0,037671	0,049008	0,035952	0,030595	0,129208
MISCEL_2	0,747940	0,491087	0,000039	-0,053698	0,031000	-0,198604	-0,012244
Expl.Var	4,378750	2,977577	0,272106	0,142089	0,082728	0,062608	0,026661
Prp.Totl	0,437875	0,297758	0,027211	0,014209	0,008273	0,006261	0,002666

ეკრანზე გამოდის ფაქტორების დატვირთვის მატრიცა, სადაც ჩანს, რომ **Factor 1**-ს **Work** ცვლადისთვის გაჩნია მაღალი ფაქტორული დატვირთვები, ხოლო **Home** ცვლადისთვის – მცირე. **Factor 2** კი პირიქით: **Work** ცვლადისთვის გაჩნია დაბალი ფაქტორული დატვირთვები, ხოლო **Home** ცვლადისთვის – მაღალი.

ფაქტორული დატვირთვების ინტერპრეტაცია საკმაოდ რთული პროცესია. პროცედურა გაცილებით მარტივდება, თუ გამოვიყენებთ ფაქტორული დატვირთვების გრაფიკულ გამოსახულებას. ამისათვის *click Plot of factor loadings* (დატვირთვის ორგანზომილებიანი გრაფიკი) ღილაკზე. ეკრანზე გამოდის *Select two factors for the plot* ფანჯარა, სადაც მონიშნულია *x* ღერძისათვის **Factor 1** და *y* ღერძისათვის – **Factor 2**. *click OK*.



ეკრანზე გამოსულ გრაფიკზე ჩანს, რომ **Work** ცვლადების ჯგუფი, განლაგებულია გრაფიკის მარჯვენა ქვედა მხარეს, ხოლო **Home** ცვლადების ჯგუფი – გრაფიკის უკიდურეს ზედა მარცხენა მხარეს. შესაბამისად **Factor 1** პასუხობს სამუშაოთი დაკმაყოფილებას, ხოლო **Factor 2** – საოჯახო ცხოვრებით დაკმაყოფილების საზომია. ამიტომ შეგვიძლია



გამოვიტანოთ დასკვნა, რომ გამოსაკვლევი ჯგუფის პიროვნებების საერთო დაკმაყოფილება ძირითადად განისაზღვრება ორი ფაქტორით – სამუშაოთი და საოჯახო ცხოვრებით დაკმაყოფილებული.

**Factor Analysis Results** ფანჯარაში გავხსნათ **Scores** ჩანართი. **click Factor Scores coefficients** ღილაკზე. გაიხსნება წრფივი რეგრესიული განტოლების კოეფიციენტების ცხრილი, რომლის საშუალებითაც პროგრამა თითოეული რესპოდენტისათვის განსაზღვრავს ფაქტორების მნიშვნელობას.

Factor Score Coefficients (FACTOR)							
Rotation: Varimax raw							
Extraction: Principal axis factoring							
Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
WORK_1	0,105578	-0,099207	-0,016919	0,501790	-0,077616	0,006897	-0,046804
WORK_2	0,306347	-0,207779	-0,059265	-0,458734	0,015786	0,274193	-0,272165
WORK_3	0,192837	-0,074703	-0,145821	0,180846	-0,495108	-0,009470	-0,168772
HOBBY_1	0,133244	0,122264	0,059829	0,433698	1,333204	-0,076420	-0,369526
HOBBY_2	0,079565	-0,052671	0,940958	-0,347628	-0,199090	-0,088741	0,044900
HOME_1	-0,100549	0,141614	0,394804	0,220206	-0,080589	0,092093	-0,107190
HOME_2	-0,224548	0,537161	-0,268305	-0,031550	-0,236671	-0,208940	-0,023156
HOME_3	-0,066215	0,217493	-0,073178	-0,183061	-0,135933	0,309825	-0,243777
MISCEL_1	0,278136	0,174691	-0,393431	0,113285	-0,252092	0,557154	1,110826
MISCEL_2	0,154595	0,036153	-0,253427	-0,314608	-0,121699	-0,796642	-0,158533

– **click Factor Scores** ღილაკზე. ეკრანზე გამოდის ცხრილი, სადაც წარმოდგენილია თითოეული რესპოდენტის ფაქტორების მნიშვნელობა. ამ ცხრილის საშუალებით შეიძლება ვიმსჯელოთ თითოეული რესპოდენტის დამოკიდებულებაზე **Factor 1** და **Factor 2** მიმართ. ფაქტორის დადებითი მნიშვნელობა შეესაბამება რესპოდენტის ფაქტორთან დადებით დამოკიდებულებას, ხოლო უარყოფით – ნეგატიურ დამოკიდებულებას.

Factor Scores (FACTOR)							
Rotation: Varimax raw							
Extraction: Principal axis factoring							
Case	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
1	0,65593	-0,36168	-0,59458	0,59538	-0,36373	-0,63867	0,108107
2	-1,96699	-0,30579	-0,44971	0,38955	0,09108	-0,41132	0,083036
3	-1,44083	0,19199	-0,00233	-0,16812	-0,36391	0,05652	0,758473
4	0,36850	-1,05763	1,11494	-0,50783	-0,06311	1,00822	0,494972
5	-0,21146	-1,73441	1,39077	1,25730	-0,52568	0,20268	0,296924
6	-1,18915	0,27215	-0,13033	0,06385	1,37973	0,66111	0,267947
7	-0,27389	-0,31448	0,13544	0,12928	0,11449	0,02269	-0,122672
8	0,71867	-0,91731	-0,25964	0,13535	0,23506	-1,02855	-0,477260
9	-0,00516	-0,19930	-0,18252	-0,14310	-0,59991	-0,28664	0,281798
10	-0,87448	-0,63352	0,95000	0,40196	-0,04853	1,04838	-0,610835
11	-0,26194	-1,34359	-1,27836	0,69752	-0,21946	-0,31976	0,795036
12	0,14816	1,19081	0,63482	1,41094	0,64448	0,14501	0,349868
13	-1,34360	1,29712	0,96315	0,62823	-0,49004	0,10786	-0,318068
14	-0,59605	-0,43621	0,48314	-0,45268	-0,57236	-0,19416	0,261753
15	0,12899	0,20420	-0,72394	-0,30899	-0,02528	-0,42967	0,194101
16	-0,58579	0,99297	1,07077	-0,25943	0,42464	-0,45232	-0,472513
17	-2,64702	-0,56058	-1,00473	0,20899	-0,24430	-0,49448	-0,190092
18	0,33045	-1,19061	0,56329	0,70311	-0,43057	-0,02306	-0,152589
19	-0,27789	0,13076	-0,07327	0,97587	1,03547	-0,39878	0,622614
20	0,82068	-0,58071	0,90154	0,09436	0,52830	-0,76414	-0,119674

ამრიგად, რედუქციის პროცედურა საშვალეებს გვაძლევს გამოვეყოთ ორის სარწმუნო ფაქტორი **Factor 1** და **Factor 2** და ამით შევამციროთ ცვლადების რაოდენობა 10-დან 2-მდე.

## პრაქტიკული სამუშაო 9

### მთავარი კომპონენტების მეთოდი და კლასიფიკაცია

პრაქტიკაში ხშირად საჭირო ხდება დიდი განზომილების მონაცემთა ანალიზი. მთავარი კომპონენტების მეთოდით შესაძლებელია ასეთი ამოცანის გადაწყვეტა და ძირითადად გამოიყენება ორი მიზნის მისაღწევად:

- ცვლადების რაოდენობის შემცირება (მონაცემების რედუქცია), რათა მივიღოთ “მთავარი” და “არაკორერული” ცვლადები;
- აგებული ფაქტორული სივრცის საშვალეებით ცვლადებისა და დაკვირვებების კლასიფიკაცია,

ამოცანის დასმის თვალსაზრისით მთავარი კომპონენტების მეთოდი მსგავსია ფაქტორული ანალიზის მეთოდისა, მაგრამ გაჩნია მთელი რიგი მნიშვნელოვანი განსხვავება:

- მთავარი კომპონენტების მეთოდში არ გამოიყენება ფაქტორების შერჩევის იტერაციული მეთოდები;

- აქტიური ცვლადებთან და დაკვირვებებთან ერთად შესაძლებელია გამოყენებული იყოს დამატებითი ცვლადები და (ან) დაკვირვებები:

- ჩამოთვლილი შესაძლებლობები საშუალებას გვაძლევს გამოვიყენოთ მთავარი კომპონენტების მეთოდი, როგორც ერთ-ერთი მძლავრი საშუალება ცვლადების და დაკვირვებების ერთდროული კლასიფიკაციისათვის.

მთავარი კომპონენტების მეთოდის იდეოლოგიიდან გამომდინარე ცვლადები შეიძლება დავეყთ ორ ჯგუფად : აქტიურ და დამხმარე ცვლადებად. ორივე ჯგუფის ეს ცვლადები მიეკუთვნებიან ერთი და იგივე მონაცემებს და შესაბამისად ერთმანეთის მიმართ კორელირებულნი არიან.

მთავარი კომპონენტები (ფაქტორები) განისაზღვრებიან მხოლოდ აქტიური ცვლადებით (ანალიზის ცვლადები). დამხმარე ცვლადები გამოიყენებიან მხოლოდ შედეგების ინტერპრეტაციისათვის. უნდა აღინიშნოს, რომ ცვლადების ასეთი დაყოფა სავალდებულო არ არის და დამოკიდებულია ამოცანის არსზე.

ანალოგიურად, დაკვირვებებიც შეიძლება დავეყთ აქტიურ და დამხმარე დაკვირვებებად. ეს შეიძლება მოხდეს დამაჯგუფებელი ცვლადის საშუალებით, როცა მისი ერთ-ერთი მნიშვნელობა გამოიყენება კოდად. დანარჩენი დაკვირვებები ითვლებიან დამხმარე დაკვირვებებად. აქაც მხოლოდ ძირითადი დაკვირვებები გამოიყენებიან მთავარი კომპონენტების განსაზღვრაში.

მთავარი კომპონენტის მეთოდში მთავარი კომპონენტების განსაზღვისთვის გამოიყენება კორელაციის ან კოვარიაციის მატრიცები. თუ საწყისი მონაცემების ცხრილში მოცემულია სხვადასხვა განზომილების მონაცემები (მაგალითად, წონა, სიგრძე, ტემპერატურა და სხვა), მაშინ მთავარი კომპონენტების განსაზღვრისთვის უნდა გამოვიყენოთ კორელაციული მატრიცა.

კორელაციული ან კოვარიაციული მატრიცების საკუთრივი მნიშვნელობები თამაშობენ მნიშვნელოვან როლს მთავარი კომპონენტების განსაზღვრაში. საკუთრივი მნიშვნელობების გამოყენებით შემუშავებულია მრავალი კრიტერიუმი ფაქტორების ოპტიმალური რაოდენობის დასადგენად. რადგან საკუთრივი მნიშვნელობების ჯამი ტოლია “აქტიური” ცვლადების რაოდენობისა და მათი საშუალო მნიშვნელობი ტოლია 1, ამიტომ საერთო კრიტერიუმი იმაში მდგომარეობს, რომ პროცედურა უნდა დავიწყოთ იმ საკუთარი მნიშვნელობებით, რომლებიც ერთზე მეტია.

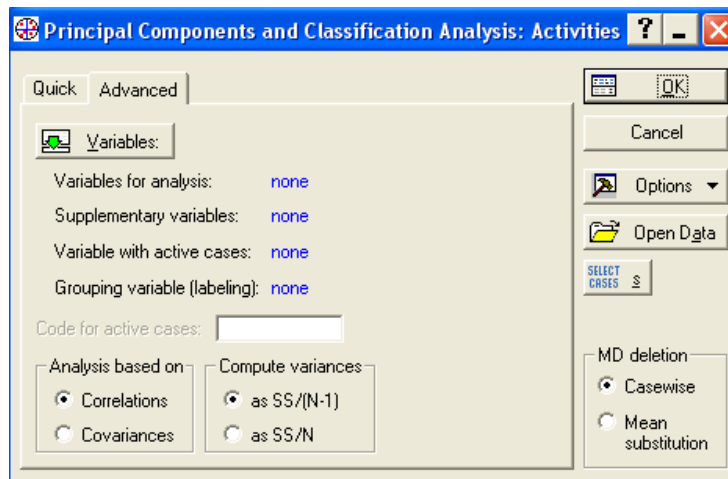
*Exmpl-ის Datasets* ბიბლიოთეკიდან გავხსნათ *Activities* მონაცემთა ფაილი, სადაც მოყვანილია ადამიანების 28 ჯგუფის ცხოვრების სხვადასხვა მახასიათებლები. აქტიურ ცვლადებად გამოვიყენოთ 7 სოციალური აქტიობის მაჩვენებელი: **WORK** (სამუშაო), **TRANSPORT** (ტრანსპორტი), **CHILDREN** (ბავშვები), **HOUSEHOLD** (ოჯახი), **SHOPPING** (შესყიდვები), **PERSONAL CARE** (პირადი დრო), **MEAL** (საჭმელი). ამ ცვლადების მაჩვენებლად აღებულია საერთო დრო საათებში, რომელიც საჭიროა შესაბამისი აქტიობისთვის. გამოტოვებული მონაცემები შევსებულია საშუალო სიდიდით.

Activities timetable data for 28 population groups; modified example data reported in Exploratory and Multivariate Data Analysis (Michel Jambu, 1991)												
	1	2	3	4	5	6	7	8	9	10	11	12
	WORK	TRANSPORT	HOUSEHOLD	CHILDREN	SHOPPING	PERSONAL CARE	MEAL	SLEEP	TV	LEISURE	GENDER	GEO.REGION
EMU	610	140	60	10	120	95	115	760	175	315	MALE	WEST
EWU	475	90	250	30	140	120	100	775	115	305	FEMALE	WEST
UWU	10		495	110	170	110	130	785	160	430	FEMALE	WEST
MMU	615	141	65	10	115	90	115	765	180	305	MALE	WEST
MWU	179	29	421	87	161	112	119	776	143	373	FEMALE	WEST
SMU	585	115	50		150	105	100	760	150	385	MALE	WEST
SWU	482	94	196	18	141	130	96	775	132	336	FEMALE	WEST
EMW	652	100	95	7	57	85	150	807	115	330	MALE	WEST
EWV	510	70	307	30	80	95	142	815	87	262	FEMALE	WEST
UWV	20	7	567	87	112	90	180	842	125	367	FEMALE	WEST
MMV	655	97	97	10	52	85	152	807	122	320	MALE	WEST
MWV	168	22	529	69	102	83	174	825	119	392	FEMALE	WEST
SMV	642	105	72		62	77	140	812	100	387	MALE	WEST
SWV	389	34	262	14	92	97	147	848	84	392	FEMALE	WEST
EME	650	142	122	22	76	94	100	764	96	334	MALE	EAST
EWE	578	106	338	42	106	94	92	752	64	228	FEMALE	EAST
UWE	24	8	594	72	158	82	128	840	86	398	FEMALE	EAST
MME	652	133	134	22	68	54	102	762	122	310	MALE	EAST
MWE	434	77	431	60	117	88	105	770	73	229	FEMALE	EAST
SME	627	148	68		88	92	86	770	58	463	MALE	EAST
SWE	433	88	296	21	128	102	94	798	58	379	FEMALE	EAST
EMY	650	140	120	15	85	90	105	760	70	365	MALE	EAST

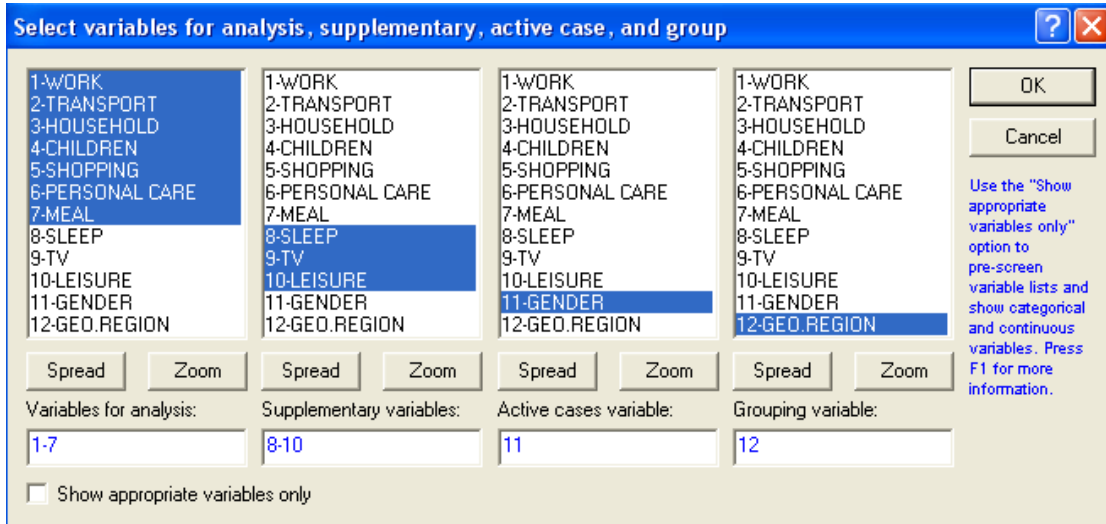
დამატებით ცვლადებად აღებულია 3 ცვლადი: **SLEEP** (ძილი), **TV** (ტელევიზორი) და **LEISURE** (მოცლა). იმისათვის რომ ძირითადი და დამატებითი დაკვირვებების წარმოდგენისთვის ფაილში დამატებულია დამატებითი დამაჯგუფებელი ცვლადი **GENDER** (სქესი). ეს იმას ნიშნავს, რომ ჯგუფის ერთი ნაწილი შედგება **FEMALE** ქალებისგან, ხოლო მეორე - **MALE** მამაკაცებისგან. გრაფიკული გამოსახულებისთვის დამატებულია **GEO.REGION** (რეგიონი), რომლის საშუალებითაც ხდება წერტილების მონიშვნა.

ანალიზის მიზანია – სოციალური აქტივობის მაჩვენებლების ურთიერთკავშირები, რათა კლასიფიკაციის მეთოდების გასამარტივებლად გამოვაგლინოთ ფარული ფაქტორები. ამ მიზნის მიღწევისთვის საჭიროა მცირე განზომილების სივრცეში განისაზღვროს ფაქტორების ღერძები, სადაც შესაძლებელი იქნება ცვლადების პროექტირება, შექმნილი ჯგუფების ვიზუალიზაცია.

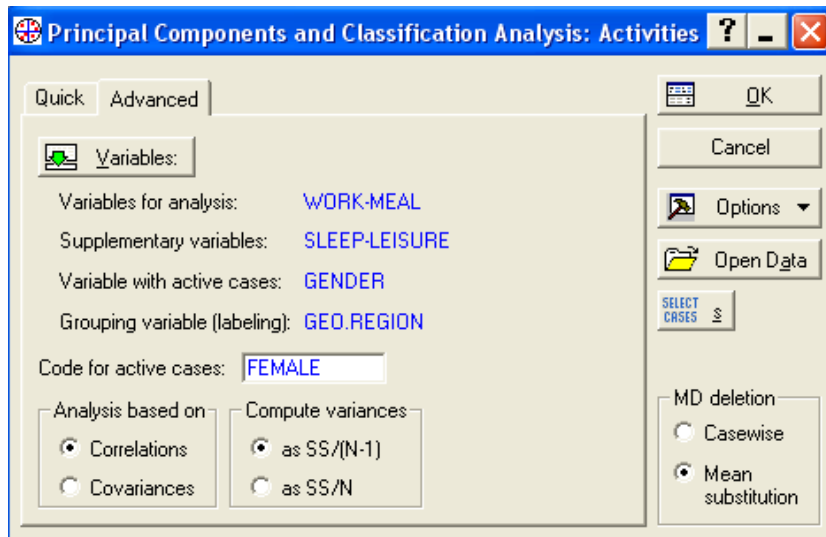
**Statistics** მენიუში *click Multivariate Exploratory Techniques* და შევარჩიოთ **Principal Components & Classification Analysis** ბრძანება. გაიხსნება სასტარტო **Principal Components & Classification Analysis** ფანჯარა:



– click **Variables** ღილაკზე. ეკრანზე გამოსახულ **Select variables for ...** ფანჯრის **Variable for analysis** (ცვლადების ანალიზისათვის) მოვნიშნოთ **WORK** (სამუშაო) – **MEAL** (კვება) ცვლადები, **Supplimentary varianbeles** (დამხმარე ცვლადების) ველში მოვნიშნოთ **SLEEP - LEASURE** ცვლადები, **Active cases variable** (ძირითადი დაკვირვებების ცვლადები) ველში მოვნიშნოთ **GENDER** ცვლადი, ხოლო **Grouping variable** ველში - **GEO REGION** ცვლადი და შემდეგ **OK**.



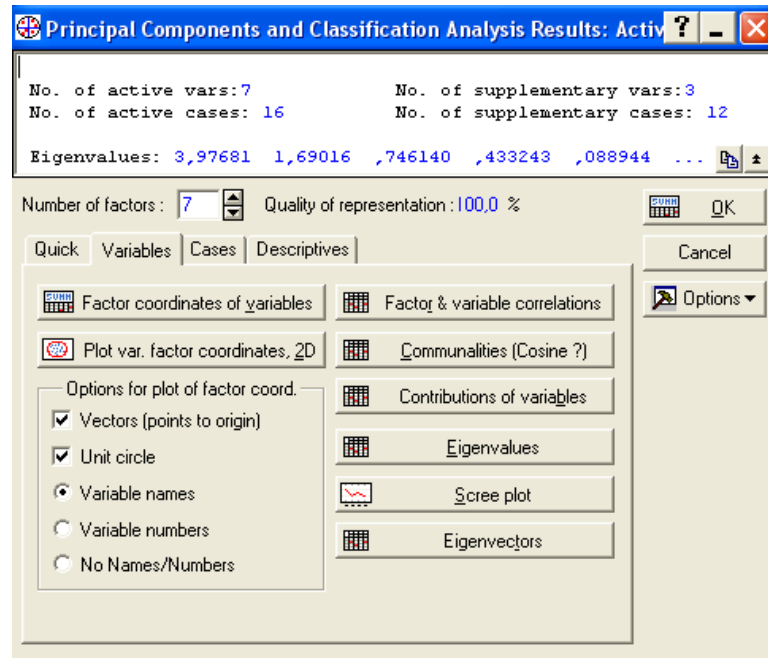
– ვბრუნდებით **Principal Components & Classification Analysis** ფანჯარაში სადაც **Code for active cases** ველში შევარჩიოთ დამაჯგუფებელი ცვლადი **FEMALE**.



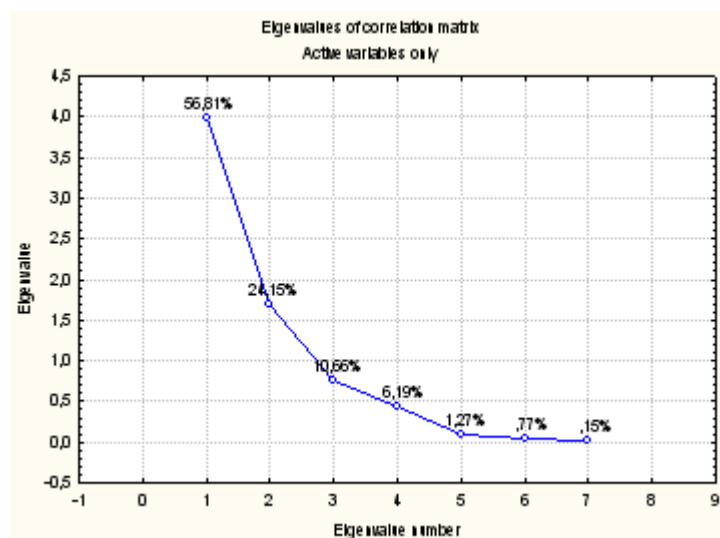
– ჩავრთოთ **Advanced** ჩანართი. **Analysis based on** (ანალიზი დაფუძნებულია) ჩარჩოში ჩავრთოთ **Correlation** ოპცია, რათა ანალიზი წარმართოს კორელაციური მატრიცის საშუალებით. **Complete variances** (დისპერსიის გამოთვლა) ჩარჩოში ჩავრთოთ **SS/(N-1)** ოპცია, რათა დისპერსიის შეფასება იყოს გადაუადგილებადი. **MD deletion** (გამოტოვებული მნიშვნელობის გამორიცხვა) ჩარჩოში ჩავრთოდ **Mean Substitution**(საშუალოთი შეცვლა) -ის ოპცია და **OK**.

-----  
ე.ყუბანეიშვილი. მონაცემების სტატისტიკური დამუშავება

– ეგრანზე გამოდის *Principal Components and Classification Analysis* შედეგების ფანჯარა, სადაც ინფორმაციულ ველში გამოტანილია ძირითადი და დამატებითი ცვლადების რაოდენობა და საკუთრივი მნიშვნელობები (*Eigenvalues*).



– ჩაერთოთ *Variables* ჩანართი და *click Scree plot* ღილაკზე. ეგრანზე გამოდის სამკუთრივი მნიშვნელობის გრაფიკი.



კეტელის კრიტერიუმის თანახმად (იხ. ფაქტორული ანალიზი) შეგვიძლია დავტოვოთ ორი ან სამი ფაქტორი. დავბრუნდეთ *Principal components* ფანჯარაში და *Number of factors* ველში დავაყენოთ რიცხვი 2.

– *click Eigenvalues* (საკუთრივი მნიშვნელობები) ღილაკზე. ეგრანზე გამოდის საკუთრივი მნიშვნელობების ცხრილი, სადაც თითოეული საკუთრივი მნიშვნელობისათვის მოცემულია *Total variance* (საერთო დიპერსია) საერთო

დისპერსიის პროცენტული მნიშვნელობა, კუმულატიური საკუთრივი მნიშვნელობა (*Cumulative Eigenvalue*) და საერთო დისპერსიის კუმულატიური პროცენტის მნიშვნელობა (*Cumulative %*).

Eigenvalues of correlation matrix, and related statistics (Activities) Active variables only				
Value number	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	3,976814	56,81163	3,976814	56,8116
2	1,690162	24,14518	5,666976	80,9568
3	0,746140	10,65914	6,413116	91,6159
4	0,433243	6,18918	6,846359	97,8051
5	0,088944	1,27063	6,935303	99,0758
6	0,054063	0,77233	6,989366	99,8481
7	0,010634	0,15191	7,000000	100,0000

კორელაციური მატრიცის დროს საკუთრივი მნიშვნელობის რაოდენობა ცვლადების რაოდენობის ტოლია. ფაქტორების შესარჩევად ყველაზე მარტივი მეთოდია – დავტოვოთ ის ფაქტორები, რომელთა საკუთრივი მნიშვნელობები 1-ზე მეტია. ჩვენ მაგალითში ასეთი მხოლოდ პირველი ორი საკუთრივი მნიშვნელობაა, რომელთა მნიშვნელობები 1-ზე მეტია და ჯამური (საერთო) დისპერსიიდან მათი წილი შეადგენს 82%. ე.ი. გამოყოფილი ფაქტორების რაოდენობა უნდა ავიღოთ ორის ტოლი.

– *click Factor coordinates of variables* (ცვლადების ფაქტორული კოორდინატები) დილაკზე. ეკრანზე გამოდის ცხრილი, სადაც ახალი გამოყოფილი ფაქტორების სივრცეში ნაჩვენებია საწყისი ფაქტორების კოორდინატები.

Factor coordinates of the variables, based on correlations (Activities) Active and Supplementary variables *Supplementary variable		
Variable	Factor 1	Factor 2
<b>WORK</b>	-0,941018	0,275054
TRANSPORT	-0,851971	-0,185457
HOUSEHOLD	0,912134	0,036525
CHILDREN	0,779245	-0,354216
SHOPPING	0,326204	-0,917236
PERSONAL CARE	-0,536329	-0,685359
MEAL	0,729504	0,377189
*SLEEP	0,590196	0,318393
*TV	0,280880	-0,568769
*LEISURE	0,476076	-0,318265

მთავარი კომპონენტების ინტერპრეტაცია უნდა გაკეთდეს კორელაციის ტერმინებით, ანუ საჭიროა გამოვყოთ ის ცვლადები (დაკვირვებები), რომლებსაც გაჩნიათ მოცემული ფაქტორის უდიდესი (აბსოლიტური) კოორდინატა.

როგორც ცხრილიდან ჩანს პირველი ფაქტორული ღერძი, რომლიც შეესაბამება საკუთრივი მნიშვნელობა 3,976, ყველაზე უფრო ძიერად არის კორელირებული **WORK**, **TRANSPORT** (ძლიერი ურთქოფითი კორელაცია), **PERSONAL CARE** (ზომიერი ურთქოფით კორელაცია), **MEAL**, **SLEEP** (ზომიერი დადებით კორელაცია), **HOUSEHOLD** და **CHILDREN** (ძლიერი დადებით

კორელაცია) ცვლადებთან, ამიტომ პირველი ფაქტორული ღერძი შეიძლება ჩავთვალოთ როგორც სოციალური აქტობის, რომელიც დაკავშირებულია სამუშაოსთან, სახლთან და ბავშვებთან. მეორე ფაქტორული ღერძი, რომელიც შესაბამება 1,69 მნიშვნელობის საკუთარ მნიშვნელობას, შეიძლება ჩავთვალოთ როგორც სოციალურად აქტიური, რომელიც დაკავშირებულია ისეთ მოღვაწეობასთან როგორცაა: შესყიდვები, პირადი დრო, ტელევიზია, (ძლიერი და ზომიერი კორელაციები **SHOPPING, TV, PERSONAL CARE**) ცვლადებთან.

**Cases** ჩანართში *click Cosine ?* ღილაკზე. ეკრანზე გამოდის ცხრილი, სადაც წარმოდგენილია დამატებითი ინფორმაცია დაკვირვების მიკუთვნებაზე ძირითად ან დამხმარე დაკვირვებებთან. თითოეულ დაკვირვებას ენიჭება დამაჯგუფებელი ცვლადის **GEO.REGION** მნიშვნელობა (**WEST** ან **FAST**).

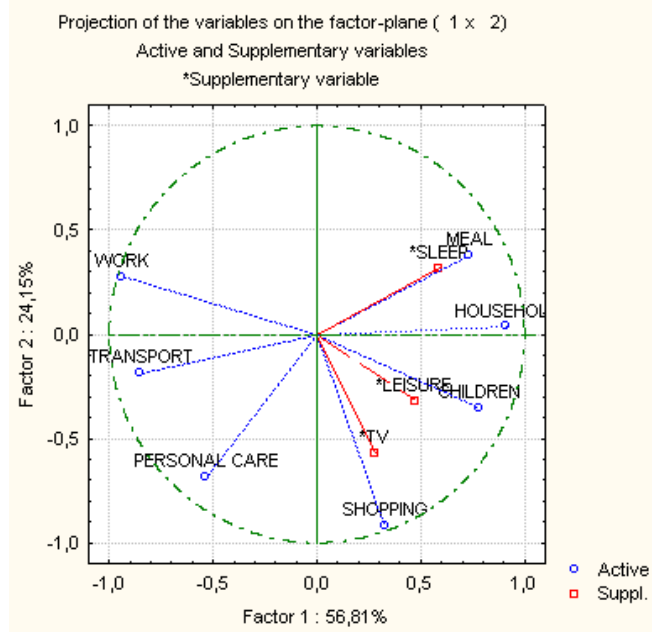
Cosine squares, based on correlations (Activities) Active cases variable: GENDER Labelling variable: GEO.REGION Code for active cases: FEMALE ; Suppl. cases highlighted				
Case	Factor 1	Factor 2	GENDER	GEO.REGION
EMU	0,812003	0,000212	MALE	WEST
EWU	0,594744	0,378262	FEMALE	WEST
UWU	0,269111	0,549224	FEMALE	WEST
MMU	0,772707	0,006230	MALE	WEST
MWU	0,292201	0,574430	FEMALE	WEST
SMU	0,596930	0,167057	MALE	WEST
SWU	0,575134	0,362167	FEMALE	WEST
EMW	0,645451	0,288717	MALE	WEST
EWV	0,561220	0,377164	FEMALE	WEST
UWV	0,819593	0,046281	FEMALE	WEST
MMW	0,616485	0,312694	MALE	WEST
MWV	0,552341	0,386092	FEMALE	WEST
SMW	0,470484	0,293301	MALE	WEST
SWV	0,317119	0,246930	FEMALE	WEST
EME	0,804535	0,074808	MALE	EAST
EWE	0,691459	0,074294	FEMALE	EAST
UWE	0,902008	0,000407	FEMALE	EAST
MME	0,295131	0,410025	MALE	EAST
MWE	0,028587	0,267624	FEMALE	EAST
SME	0,720830	0,027168	MALE	EAST
SWE	0,817444	0,028422	FEMALE	EAST
EMY	0,797867	0,080167	MALE	EAST
EWY	0,471734	0,270937	FEMALE	EAST
UWY	0,828791	0,013313	FEMALE	EAST
MMY	0,793998	0,071544	MALE	EAST
MWY	0,497039	0,348981	FEMALE	EAST
SMY	0,687629	0,003994	MALE	EAST

ფაქტორების ინტერპრეტაციის პროცედურას ძალზედ ეხმარება ფაქტორების და დაკვირვებების კოორდინატების გრაფიკი. ამისთვის:

– *click Variables* ჩანართში **Plot var. factor coordinates 2D** ღილაკზე. ეკრანზე გამოდის გრაფიკი, სადაც ყველა ცვლადი გამოსახულია ერთულოვან წრეზე წერტილის სახით, რადან დაკვირვებების ფაქტორებთან კორელაცია ღებულობს მნიშვნელობებს (0,1) ინტერვალშიდან. პორიზონტალური ღერძზე



გადაზომილია ფაქტორი 1 (**Factor1**), ხოლო ვერტიკალურ ღერძზე ფაქტორი 2 (**Factor2**). ძირითადი და დამხმარე ცვლადები აღნიშნული არიან შესაბამისად პატარა წითლ (მწვანე ფერი) და მარტკუთხედით (წითელი).



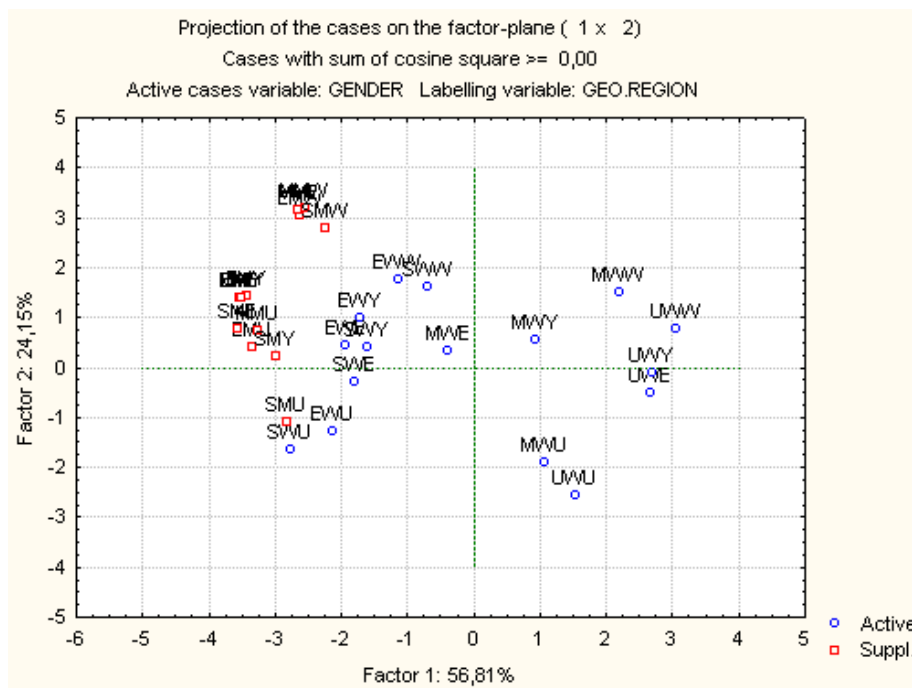
– **Click contributions of variables** (ცვლადების წვლილი) ღილაკზე. ეკრანზე გამოდის ცხრილი, სადაც წარმოდგენილია ძირითადი ცვლადების წვლილები, ცვლადის წვლილი წარმოადგენს ფაქტორულ ღერძის დისპერსიის ფაქტორულ წვლილს.

Variable	Variable contributions, based on correlations (Activities)	
	Factor 1	Factor 2
<b>WORK</b>	0,222669	0,044762
TRANSPORT	0,182522	0,020350
HOUSEHOLD	0,209210	0,000789
CHILDREN	0,152691	0,074235
SHOPPING	0,026757	0,497776
PERSONAL CARE	0,072331	0,277913
MEAL	0,133820	0,084176

– **click Communalities [cosine 2]** ღილაკზე. პროგრამა ეკრანზე გამოიტანს ცვლადების სპეციფიურ დისპერსიებს.

Communalities, based on correlations (Activities) Active and Supplementary variables *Supplementary variable		
Variable	From 1 factor	From 2 factors
<b>WORK</b>	0,885515	0,961170
TRANSPORT	0,725854	0,760248
HOUSEHOLD	0,831988	0,833322
CHILDREN	0,607222	0,732691
SHOPPING	0,106409	0,947731
PERSONAL CARE	0,287648	0,757366
MEAL	0,532177	0,674448
*SLEEP	0,348331	0,449705
*TV	0,078893	0,402391
*LEISURE	0,226649	0,327941

*Cases* ჩანართში *click Plor cases factor coordinates, 2D*(ფაქტორული სივრცეში დაკვირვებების გრაფიკი). ეკრანზე გამოდის გრაფიკი, სადაც გამოსახულია როგორც მთავარი (**FEMALES**) დაკვირვებები, რომლებიც გამოყენებული იყო ფაქტორების გამოსათვლელად (ცისფერი წრები), ისე დამხმარე (**MALES**) დაკვირვებები (წითელი ფერის კვადრატები).



როგორც გრაფიკიდან ცანს, ძირითადი და დამხმარე დაკვირვებები დაჯგუფებულნი არიან სიბრტყის სხვა და სხვა არეში, ე.ი. ისინი გაერთანებული არიან ერთგვაროვან ჯგუფებში – კალსტერებში. გრაფიკიდან ჩანს, რომ მამაკაცების ჯგუფები გაერთიანებული არიან **WORK** ცვლადის არეში. ეს იმას ნიშნავს, რომ მამაკაცთა ჯგუფების სოციალური აქტიობა ძირითადად კონცენტრირებულია სამუშაოზე, რაც შეეხება ქალთა ჯგუფის სოციალურ აქტივობას, მათ განნიათ უფრო მრავალმხრივი ხასიათი. ისინი დაჯგუფებულნი არიან მეტ-ნაკლებად თანაბრად მთელ სივრტყეზე, კერძოდ **PERSONAL CARE, CHILDREN, HOUSEHOLD, SHOPPING, MEAL, SLEEP, LEISURE** ცვლადების არეში.

პრაქტიკული სამუშაო 10

ბაღარჩენის ანალიზი. ბაღარჩენის ფუნქციის შეფასება

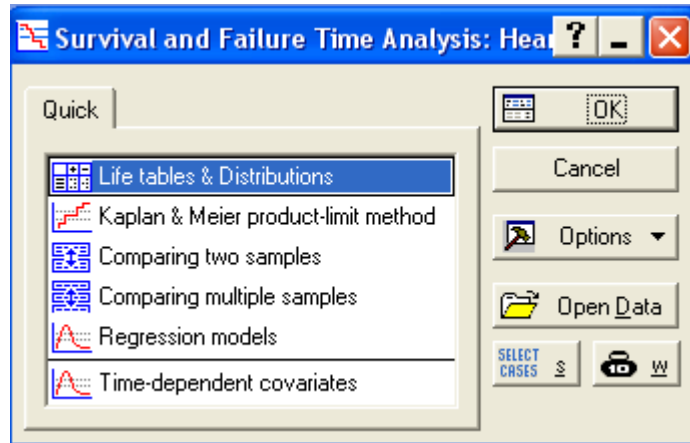
1. გადარჩენის ცხრილები

გადარჩენის ანალიზის მეთოდების განხილვისათვის *Statistica* სისტემის *Datasets* ბიბლიოთეკიდან გავხსნათ *Heart* მონაცემთა ფაილი, სადაც წარმოდგენილია 65 პაციენტის სიცოცხლის ხანგრძლივობა, რომელთაც ჩაუტარდათ გულის ტრანსპლანტაცია.

Heart transplant data from Crowley and Hu, stratified											
	1	2	3	4	5	6	7	8	9	10	11
	MONTH_1	DAY_1	YEAR_1	MONTH_2	DAY_2	YEAR_2	CENSORED	AGE	ANTIGEN	MISMATCH	HOSPITAL
1	JANUARY	6	68	JANUARY	21	68	CENSORED	54	0	1,11	HILLVIEW
2	MAY	2	68	MAY	5	68	CENSORED	40	0	1,66	HILLVIEW
3	AUGUST	31	68	MAY	17	70	COMPLETE	51	0	1,32	HILLVIEW
4	AUGUST	22	68	OCTOBER	7	68	COMPLETE	42	0	0,61	ST_AND
5	SEPTEMBER	9	68	JANUARY	14	69	CENSORED	48	0	0,36	ST_AND
6	OCTOBER	5	68	DECEMBER	8	68	COMPLETE	54	0	1,89	ST_AND
7	OCTOBER	26	68	JULY	7	72	COMPLETE	54	0	0,87	BINER
8	NOVEMBER	22	68	AUGUST	29	69	COMPLETE	49	0	1,12	BINER
9	NOVEMBER	20	68	DECEMBER	13	68	CENSORED	56	0	2,05	HILLVIEW
10	FEBRUARY	15	69	FEBRUARY	25	69	COMPLETE	55	1	2,76	HILLVIEW
11	FEBRUARY	8	69	NOVEMBER	29	71	COMPLETE	43	0	1,13	BINER
12	MARCH	29	69	MAY	7	69	COMPLETE	42	0	1,38	HILLVIEW
13	APRIL	13	69	APRIL	13	71	COMPLETE	58	0	0,96	ST_AND
14	JULY	16	69	NOVEMBER	29	69	COMPLETE	52	1	1,62	ST_AND
15	MAY	22	69	APRIL	1	74	CENSORED	33	0	1,06	ST_AND
16	AUGUST	16	69	AUGUST	17	69	CENSORED	54	0	0,47	BINER
17	SEPTEMBER	3	69	DECEMBER	18	71	COMPLETE	44	0	1,58	BINER
18	SEPTEMBER	14	69	NOVEMBER	13	69	COMPLETE	64	0	0,69	HILLVIEW
19	JANUARY	16	70	APRIL	1	74	CENSORED	49	0	0,91	BINER
20	JANUARY	3	70	APRIL	1	74	CENSORED	40	0	0,38	HILLVIEW
21	MAY	19	70	JULY	12	70	COMPLETE	49	0	2,09	HILLVIEW
22	MAY	13	70	JUNE	29	70	COMPLETE	61	1	0,87	ST_AND
23	MAY	9	70	MAY	9	70	CENSORED	41	0	0,87	ST_AND
24	JULY	4	70	APRIL	1	74	CENSORED	48	0	0,75	BINER
25	OCTOBER	15	70	APRIL	1	74	CENSORED	45	0	0,98	BINER

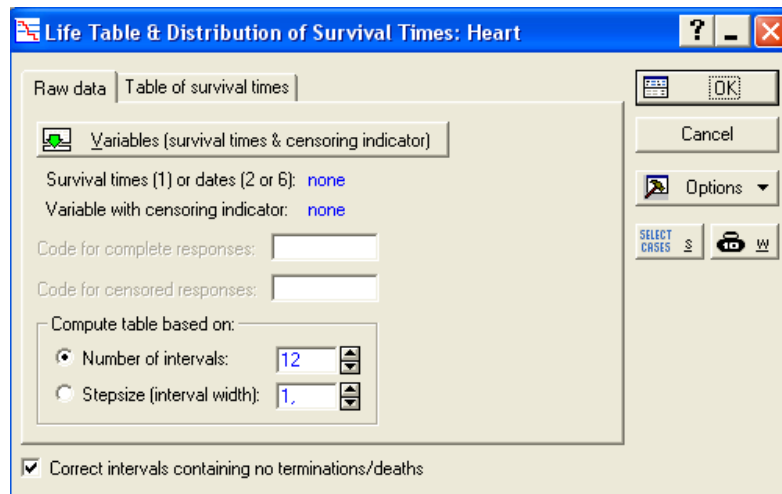
მონაცემების პირველი 6 ცვლადი წარმოადგენს თარიღს (თვე, დღე, წელი) და პაციენტის დაიღუპვის ან დაკვირვებიდან გამორიცხვის თარიღს. **CENSORED** ცვლადი წარმოადგენს ცენზურირების ინდიკატორის კოდებს, რომლებიც განსაზღვრავენ პაციენტის კონკრეტულ დაკვირვებას ან ცენზურირებულ დაკვირვებას (**COMPLETE** – 0, **CENSORED** – 1). **AGE** ცვლადი აღნიშნავს პაციენტის ასაკს. ცვლადები **ANTIGEN** და **MISMATCH** შეიცავენ სპეციალურ სამედიცინო ინფორმაციას ანტიგენურ და ქსოვილების შეუთავსებლობაზე. **HOSPITAL** (პირობითი ცვლადი) განსაზღვრავს სამი ჰოსპიტალიდან რომელს მიეკუთვნება პაციენტი.

**Survival Analysis** (გადარჩენის ანალიზი) მოდულის გაშვებისათვის **Statistics** მენიუში მოენიშნოთ **Advanced Linear/Nonlinear Modest** და შევარჩიოთ **Survival Analysis** ბრძანება. ეკრანზე გამოდის **Survival and Failure Time Analysis** სასტარტო ფანჯარა,



სადაც წარმოდგენილია ძირითადი მოდულები: **Line tables & Distributions** (გადარჩენის ცხრილები და განაწილება); **Kaplan & Meier product-limit method** (კაპლან-მეიერის მამრავლების მეთოდი); **Comparing two samples** (ორი ამონარჩევის შედარება); **Comparing multiple samples** (რამოდენიმე ამონარჩევის შედარება); **Regression models** (რეგრესიის მოდელები); **Time-dependent covariates** (დროზე დამოკიდებული კოვარიანტები).

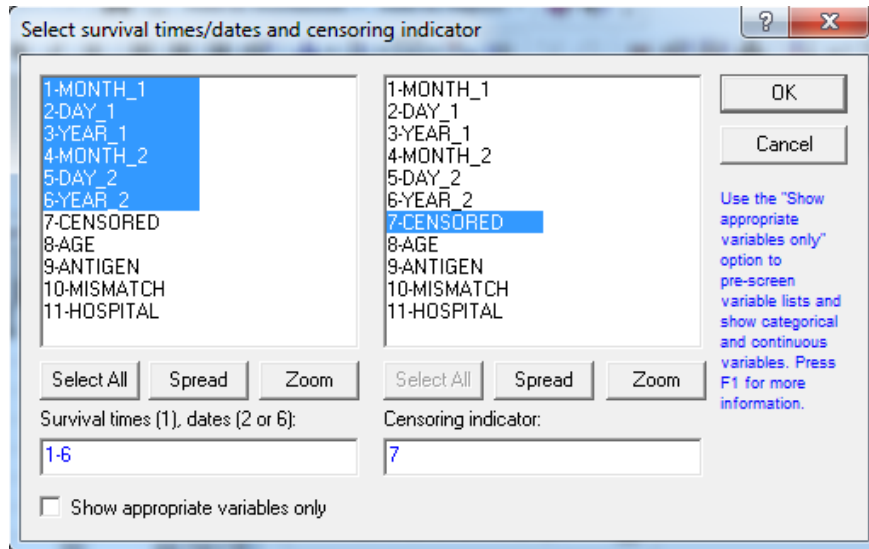
შევარჩიოთ **Line tables & Distributions** პროცედურა და click **OK** ღილაკზე. გაიხსნება **Line tables & Distributions of Survival Times** ფანჯარა,



რომელსაც გააჩნია ორი ჩანართი: **Raw data** (საწყისი მონაცემები) და **Table survival times** (გადარჩენის ცხრილები). პირველი ჩანართი შეესაბამება იმ შემთხვევას, როცა შემომავალ ინფორმაციას წარმოადგენს დაუმუშავებელი მონაცემები ანუ **Statistica** სისტემის ჩვეულებრივი ცხრილი (სტრიქონები – დაკვირვებები. სვეტები – ცვლადები). მეორე ჩანართი შეესაბამება იმ შემთხვევას, როცა

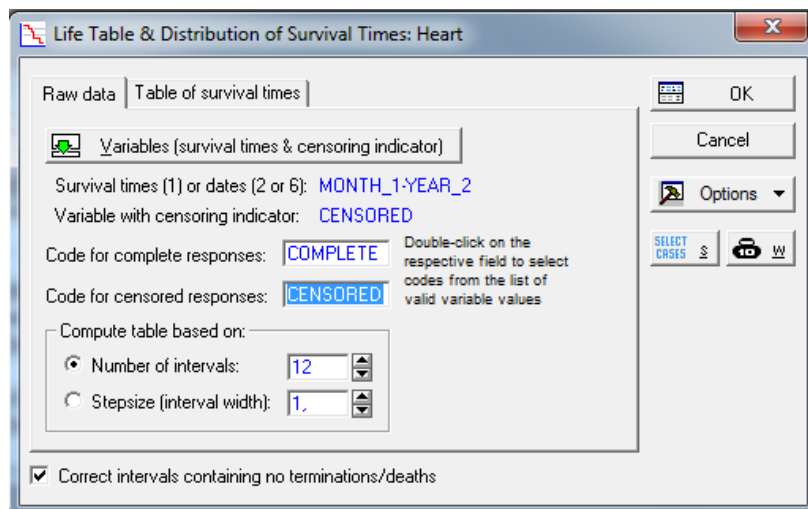
შემომავალ ინფორმაციას წარმოადგენს აღრე გამოთვლილი გადარჩენის მონაცემები.

შეგარჩიოთ **Raw data** და განვიხილოთ ფანჯრის დილაკების დანიშნულება. **Variables (survival times & censoring indicator)** (ცვლადები, ხანგრძლივობა და ცენზურირების ინდიკატორი) დილაკზე დაწკაპუნებით ეკრანზე გამოდის **Select survival times/dates...ფანჯარა**



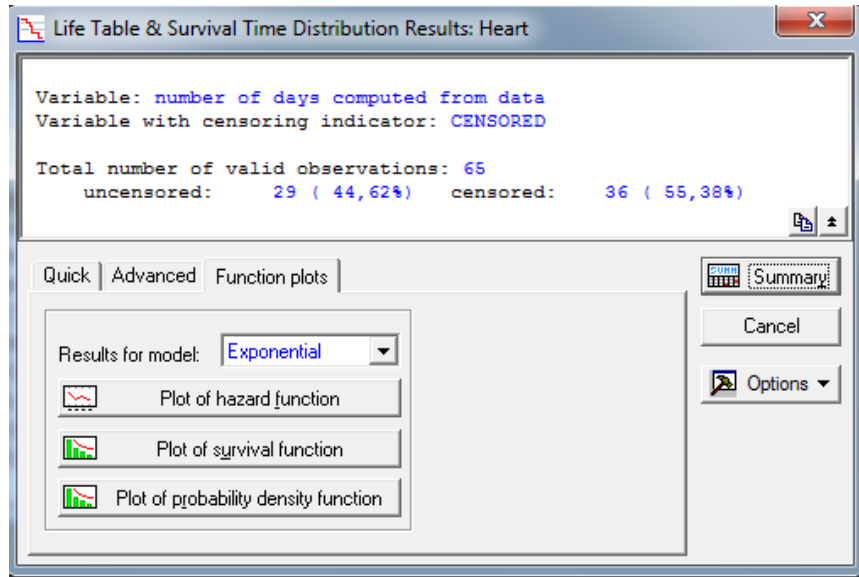
სადაც მოვნიშნოთ ფანჯრის მარცხენა არეში პირველი 6 ცვლადი, მარჯვენა არეში მე-7 ცვლადი. შემდეგ **OK**. ვბრუნდებით წინა ფანჯარაში.

ორჯერ **click Code for complete responses** (სრული დაკვირვებების კოდი) ველში და ეკრანზე გამოსულ **Variable 7** ფანჯარაში სრული დაკვირვებებისათვის მოვნიშნოთ **COMPLETE** და შემდეგ **OK**. ანალოგიურად ორჯერ **click Code for censored responses** (ცენზურირებული დაკვირვებების კოდი) ველში და ეკრანზე გამოსულ **Variable 7** ფანჯარაში შეგარჩიოთ **CENSORED** და შემდეგ **OK**.



ფანჯრის სხვა მაჩვენებლები დავტოვოთ უცვლელი. შემდეგ **OK**.

ეკრანზე გამოდის **Life Table & Survival Time Distribution Results Heart** (გადარჩენის ცხრილის შედეგების ფანჯარა),



სადაც ფანჯრის ინფორმაციულ ნაწილში წარმოდგენილია:

- **Variable** სტრიქონში – დღეების რაოდენობა, რომელიც გამოთვლილია მოცემული საწყისი მონაცემებიდან;
- **Variable with** სტრიქონში – ცენზურირების ინდიკატორი;
- **Total number of validi observations** – დაშვებული დაკვირვებების რაოდენობა, რომელიც ამ შემთხვევაში ტოლია 65;
- **uncesorend** სტრიქონში – არაცენზურირებული ობიექტების რაოდენობა 29;
- **cesorend** სტრიქონში – ცენზურირებული ობიექტების რაოდენობა 36.

ფრჩხილებში ნაჩვენებია ფარდობითი სიხშირე.

ჩაერთოდ **Advanced** ჩანართი და **click Summary: Life table** (გადარჩენის ცხრილის შედეგები)-ის ღილაკზე. ეკრანზე გამოდის შედეგების სრული გაშლილი ცხრილი

Life Table (Heart)											
Log-Likelihood for data: -68,6809											
Interval	Interval Start	Mid Point	Interval Width	Number Entering	Number Withdrwn	Number Exposed	Number Dying	Proportn Dead	Proportn Surviving	Cum.Prop Surviving	Problty Density
Intno.1	0,000	80,682	161,3636	65	14	58,00000	19	0,327586	0,672414	1,000000	0,002030
Intno.2	161,364	242,046	161,3636	32	4	30,00000	4	0,133333	0,866667	0,672414	0,000556
Intno.3	322,727	403,409	161,3636	24	4	22,00000	0	0,022727	0,977273	0,582759	0,000082
Intno.4	484,091	564,773	161,3636	20	4	18,00000	1	0,055556	0,944444	0,569514	0,000196
Intno.5	645,455	726,136	161,3636	15	1	14,50000	1	0,068966	0,931034	0,537875	0,000230
Intno.6	806,818	887,500	161,3636	13	3	11,50000	1	0,086957	0,913044	0,500780	0,000270
Intno.7	968,182	1048,864	161,3636	9	1	8,50000	2	0,235294	0,764706	0,457234	0,000667
Intno.8	1129,545	1210,227	161,3636	6	1	5,50000	0	0,090909	0,909091	0,349649	0,000197
Intno.9	1290,909	1371,591	161,3636	5	1	4,50000	1	0,222222	0,777778	0,317863	0,000438
Intno.10	1452,273	1532,955	161,3636	3	2	2,00000	0	0,250000	0,750000	0,247227	0,000383
Intno.11	1613,636	1694,318	161,3636	1	0	1,00000	0	0,500000	0,500000	0,185420	0,000575
Intno.12	1775,000			1	1	0,50000	0	1,000000	0,000000	0,092710	

სადაც წარმოდგენილია:

- **Interval Start** – ინტერვალის დასაწყისი;
- **Mid Point** – ინტერვალის საშუალო მნიშვნელობა;
- **Interval Width** – ინტერვალის სიგრძე;
- **Number Entering** – ობიექტების რაოდენობა, რომლებიც „ცოცხლები“ იყვნენ განსახილველი ინტერვალის დასაწყისში;
- **Number Exposed** – შესასწავლი ობიექტების რაოდენობა, ანუ განსახილველი ინტერვალის დასაწყისში არსებული „ცოცხალი“ ობიექტების რაოდენობას მინუს ცენზურირებული ან გამორიცხული ობიექტების ნახევარი;
- **Number Dying** – მოცემულ ინტერვალში დაღუპულთა რაოდენობა;
- **Number Dead** – მოცემულ ინტერვალში დაღუპულთა ფარდობითი სიხშირე, რომელიც ტოლია ინტერვალში დაღუპულთა რაოდენობის ფარდობისა ამ ინტერვალის შესასწავლი ობიექტა რაოდენობაზე;
- **Proportion surviving** – გადარჩენილთა ფარდობითი სიხშირე, რომელიც ტოლია ერთს გამოკლებული დაღუპულთა ფარდობითი სიხშირე;
- **Comulative Proportion surviving** – გადარჩენილთა კუმულატიური (დაგროვილი) ფარდობითი სიხშირე, რომელიც ტოლია წინა ინტერვალების ფარდობითი სიხშირეების ნამრავლისა. ფაქტიურად ეს არის გადარჩენის ფუნქციის შეფასება;
- **Probability density** – სიმკვრივის ფუნქციის შეფასება, რომელიც განისაზღვრება ფორმულით:

$$f_i = \frac{P_i - P_{i+1}}{h_i}$$

სადაც  $f_i$  –  $i$ -ურ მომენტში რისკის ფუნქციის შეფასებაა;

$P_i, P_{i+1}$  –  $i$  და  $i+1$  ინტერვალების დასაწყისში გადარჩენილთა კუმულატიური ფარდობითი სიხშირეების (გადარჩენის ფუნქციის) შეფასებებია;  $h_i$  –  $i$ -ური ინტერვალის სიგრძე.

- **Hazard rate** (მტყუნების ინტენსივობის ან მყისიერი რისკის ფუნქციის) შეფასებებია, რომლებიც განისაზღვრებიან როგორც შესაბამისი ინტერვალის დროის ერთეულზე მოსული მტყუნების რაოდენობა გაყოფილი ობიექტების საშუალო სიდიდეზე, რომლებმაც იცოცხლეს ინტერვალის საშუალო დრომდე;
- **Std.Err.Cum.surv.** – გადარჩენილთა კუმულატიური ფარდობითი სიხშირის სტარდარტული შეცდომა;
- **Std. Err.Prob.Den** – განაწილების სიმკვრივის სტარდარტული შეცდომა;
- **Std. Err.Har.Rate** – ინტენსივობის ფუნქციის სტარდარტული შეცდომა;
- **Median Life expected** – მოსალოდნელი გადარჩენის მედიანა;
- **Std. Err.Life expected** – მედიანის სტარდარტული შეცდომა.

უნდა აღინიშნოს, რომ სამი ძირითადი ფუნქციის (გადარჩენის, სიმკვრივის და ინტენსივობის) შეფასების და მათი სტარდარტული

შეცდომებისათვის ყოველ დროით ინტერვალში რეკომენდირებულია გამოყენებული იყოს არა ნაკლები 30 დაკვირვება.

**Results for modet** ველი გამოიყენება განაწილების კანონის დასადგენად: **Exponential** (ექსპონენციალური), **Linear Hazard** (წრფივი ინტენსივობის ოჯახი), **Gompertz** (გომპერტცის) და **Weibull** (ვეიბულის). მარტივი გარდაქმნების შედეგად ოთხივე განაწილება დაიყვანება  $y = ax + b$  წრფივ განტოლებამდე.

**Parameter estimates** დილაკის საშუალებით ხდება ელექტრონულ ცხრილებში შესაბამისი გადარჩენის განაწილების ფუნქციის პარამეტრების შეფასების შედეგების გამოტანა.

**Results for modet** ველის სიის საშუალებით მოცემული მონაცემებისათვის შეიძლება შეირჩეს ოთხი თეორიული განაწილების ფუნქცია. **Survival Analysis** მოდული თეორიულ განაწილების ფუნქციას ირჩევს უმცირეს კვადრატთა მეთოდით ანუ **Weight 1** – როცა წონითი კოეფიციენტები ერთის ტოლია და ორი შეწონილი კვადრატების მეთოდით **Weight 2** და **Weight 3** (ამიტომ, რომ ცხრილში გამოდის სამივე წონითი მნიშვნელობები)

$$WSS = \sum_i w_i (y_i - a - bx_i)^2 ,$$

სადაც  $w = 1$  (არაშეწონილი უმცირეს კვადრატები);

$w = 1/v_i$  (შეწონილი უმცირეს კვადრატები);

$w = n_i \cdot h_i$  (შეწონილი უმცირეს კვადრატები);

$v_i$  – ინტენსივობის დისპერსიის შეფასება;

$h_i$  და  $n_i$  –  $i$ -ური ინტერვალის სიგრძე და  $i$ -ური ინტერვალის დასაწყისში არსებული ობიექტების რაოდენობა.

ებრუნდებით **Life Table & Survival Time Distribution Results Heart** ფანჯარაში, სადაც **Results for modet** ველით განისაზღვრება: რომელი თეორიული განაწილების ფუნქცია საუკეთესოდ შეესაბამება ემპირიულ განაწილების ფუნქციას.

**Results for modet** ველში შევარჩიოთ ვეიბულის განაწილება და **click Parameter estimates** დილაკზე. ეკრანზე გამოდის ცხრილი

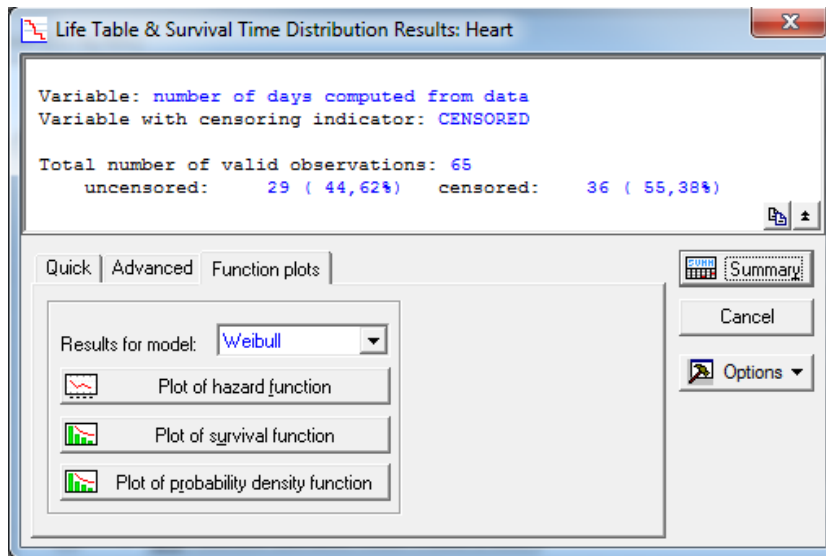
Parameter Estimates, Model: Weibull (Heart)											
Note: Weights: 1=1., 2=1./V, 3=N(I)*H(I)											
Estimate Method	Lambda	Variance Lambda	Std.Err. Lambda	Gamma	Variance Gamma	Std.Err. Gamma	Covariance Gam-Lamd	Log-Likelhd.	Chi-Sqr.	df	p
Weight 1	0,000311	0,000000	0,000574	1,142167	0,072583	0,269413	-0,000153	-84,3429	31,32401	9	0,000262
Weight 2	0,016002	0,000322	0,017952	0,644320	0,027270	0,165137	-0,002927	-75,4347	13,50757	9	0,141007
Weight 3	0,051100	0,005217	0,072230	0,427680	0,036929	0,192170	-0,013704	-72,5594	7,75703	9	0,558814

სადაც ჩანს, რომ საუკეთესოა **Weight 3**, რადგან ამ წონას შეესაბამება ხი-კვადრატ კრიტერიუმის უმცირესი მნიშვნელობა (7,757) და  $p$  მნიშვნელობა (0,5588), რომელიც მნიშვნელოვნად აღემატება 0,05 სიდიდეს.

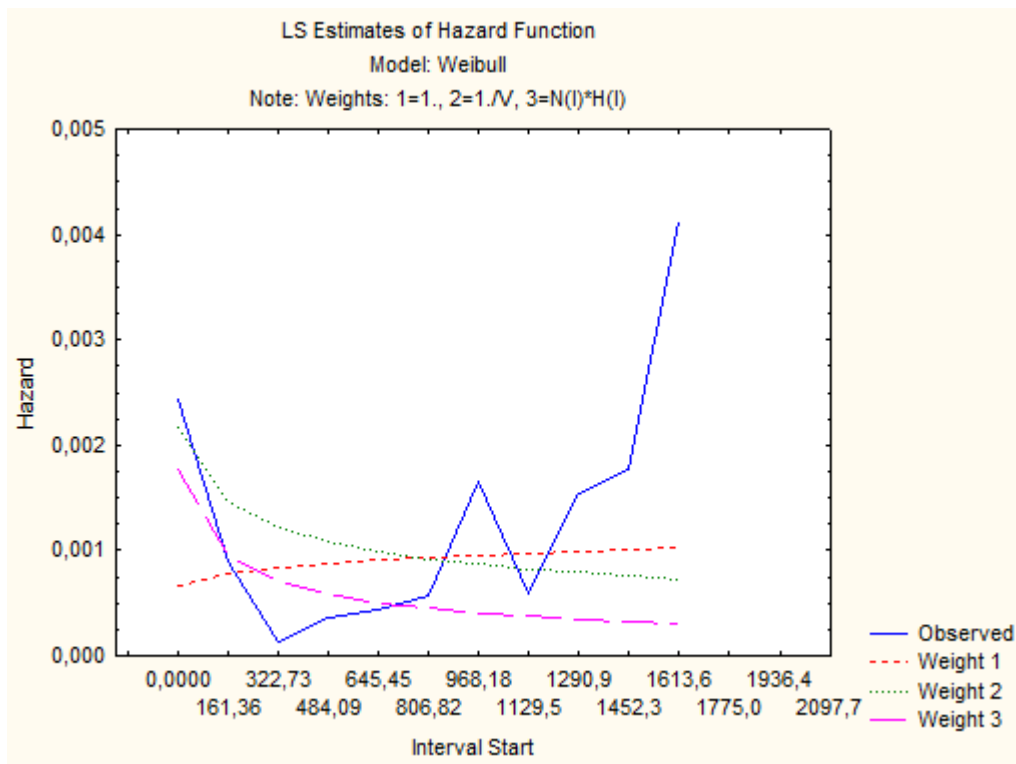
**Exstimates of survival function, Exstimates of hazard function, Exstimates of prob.density funct.** დილაკებით შესაძლებელია დამოუკიდებლად გადარჩენის, რისკის, განაწილების ფუნქციების შეფასებების ცხრილების გამოტანა. ამ ფუნქციების საწყისი მნიშვნელობები შეიძლება გამოვიტანოთ **Life table**

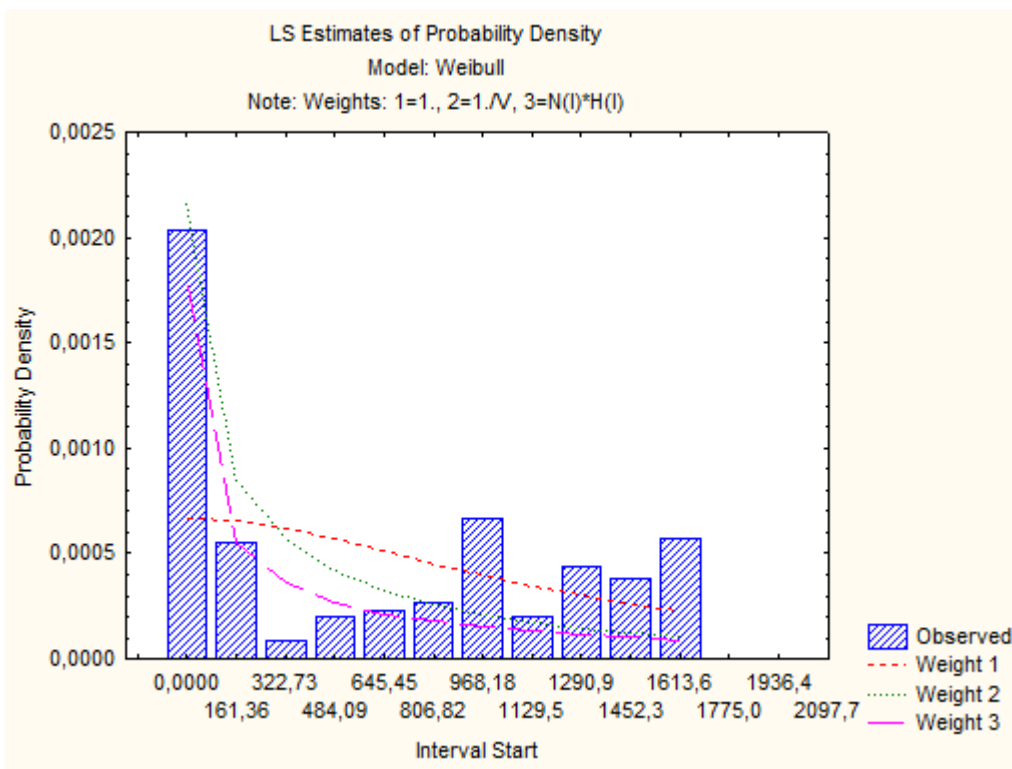
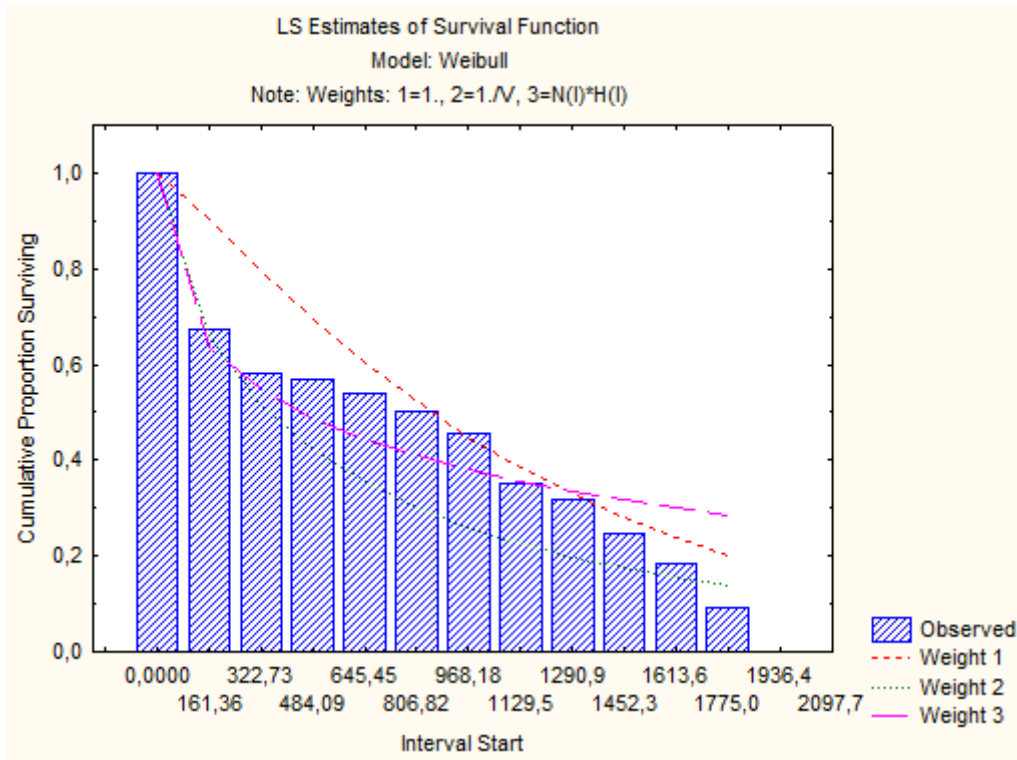


საერთო ცხრილში. ცხრილების ინფორმაციულ ნაწილებში გამოდის წონების გამოსათვლელი ფორმულები. ჩაერთოდ **Function plots** ჩანართი,



სადაც **Plot of hazard function**, **Plot of survival function**, **Plot of prob.density** დილაკებით შესაძლებელია გადარჩენის, რისკის და განაწილების ფუნქციების აგება.





გრაფიკებზე მთლიანი მრუდით წარმოდგენილია საწყისი (ემპირიული) განაწილება, ხოლო წყვეტილი ფერადი მრუდით – თეორიული განაწილებები, რომლებიც შეესაბამებიან სხვადასხვა წონებს.

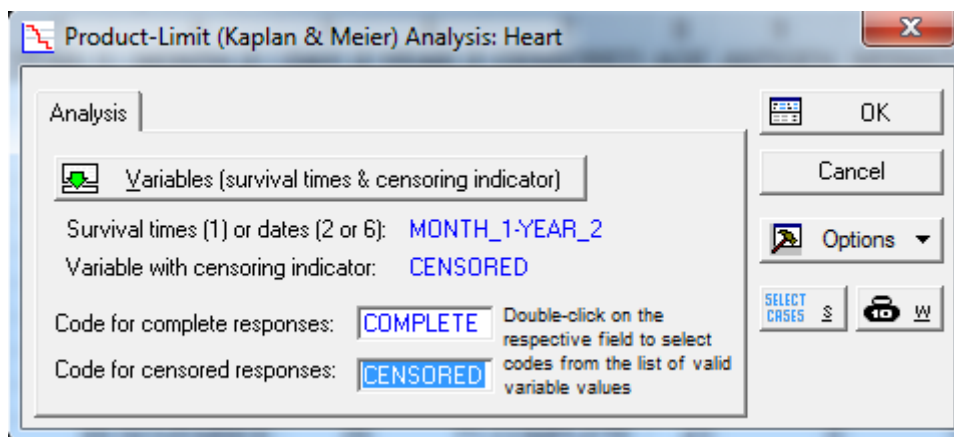
რისკის ფუნქციის გრაფიკიდან ჩანს, რომ დაღუპვის რისკის ალბათობა მკვეთრად მცირდება ოპერაციის პირველი დღიდან 322 დღემდე, შემდეგ ეს ალბათობა უმნიშვნელად იზრდება 806 დღემდე და მკვეთრად იზრდება 968 დღემდე. შემდეგ რისკის ალბათობა მცირდება 1129 დღემდე, რომლის შემდეგ იგი კვლავ მკვეთრად იზრდება. ყველაზე კარგი მიახლოება თეორიულ განაწილებასთან გვაძლევს **Weight 3** წონით შეწონილი უმცირეს კვადრატთა მეთოდი.

გადარჩენის ცხრილიდან და ჰისტოგრამიდან გამომდინარეობს, რომ ოპერაციიდან 161 დღის შემდეგ გადარჩენილების კუმულატიური ფარდობითი სიხშირე შეადგენს 67,2%, ხოლო 322 დღის შემდეგ – უკვე 58,3%. შემდგომ გადარჩენილების კუმულატიური ფარდობითი სიხშირის დაკლება გრძელდება, მაგრამ კლების ტემპი ეცემა. მკვეთრი შემცირება აღინიშნება 1129 დღის შემდეგ და შეადგენს 34,96%. განხილული დროის ინტერვალის ბოლოს (1775 დღის შემდეგ) გადარჩენილების ფარდობითი სიხშირე შეადგენს მხოლოდ 9,3%.

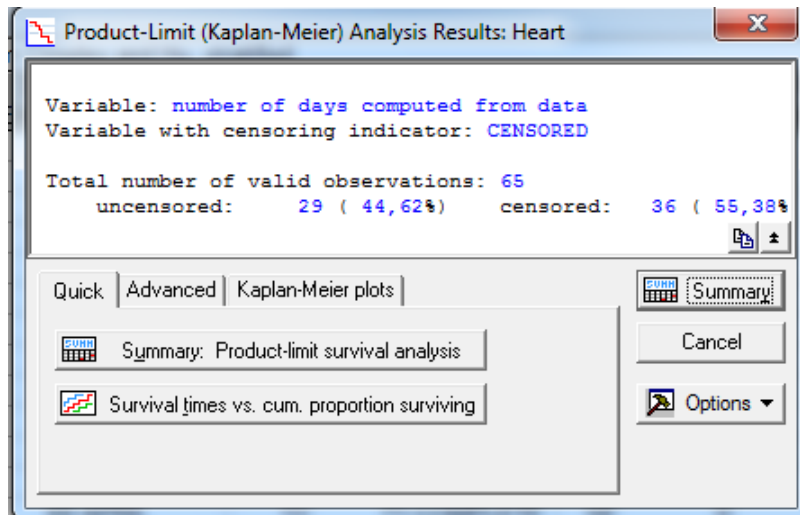
გადარჩენის ჰისტოგრამიდან ჩანს, რომ პაციენტების დაღუპვის ყველაზე დიდი რისკი არის ოპერაციიდან პირველი 161 დღე, უმცირესი 322 დღიდან 484 დღემდე.

## 2. კაპლან – მეიერის მამრავლების შეფასების მეთოდი

**Survival and Failure Time Analysis** სასტატო ფანჯარაში მოვნიშნოთ **Kaplan & Meier product-limit method** (კაპლან-მეიერის მამრავლების მეთოდი) და შემდეგ **OK**. ეკრანზე გამოდის **Product-Limit (Kaplan&Meier) Analysis** ფარჯარა,



სადაც მიუთითოდ ცვლადების კოდები ისე, როგორც **Line tables & Distributions of Survival Times** ფანჯარის დროს. **click OK**. ეკრანზე გადარჩენის ფუნქციის შეფასებისათვის გამოდის **Product-Limit (Kaplan&Meier) Analysis Results** ფარჯარა.



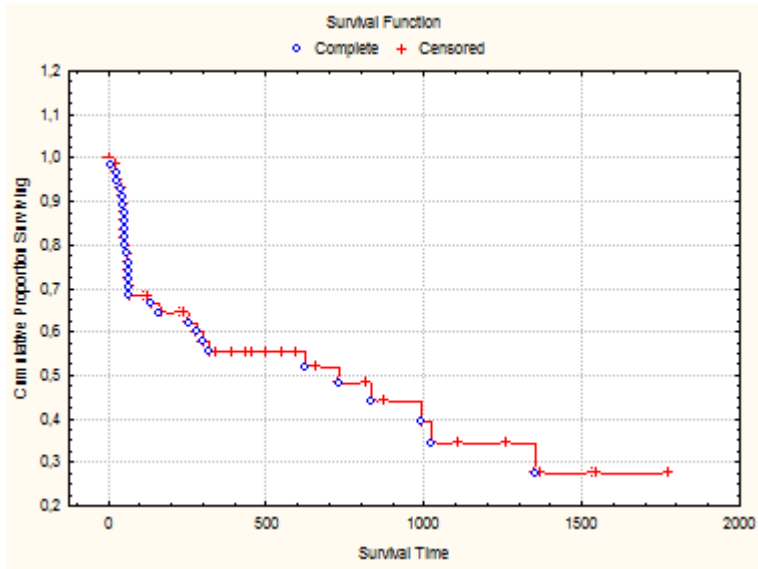
Click **Summary: Product-limit Survival analysis** ღილაკზე. ეკრანზე გამოდის შედეგების ცხრილი,

Kaplan-Meier (Product-limit) analysis (H)			
Note: Censored cases are marked with			
Case Number	Time	Cumulativ Survival	Standard Error
23+	0,000		
16+	1,000		
65+	1,000		
2+	3,000		
10	10,000	0,983607	0,016259
46+	12,000		
64+	13,000		
1+	15,000		
9+	23,000		
42	25,000	0,966042	0,023622
58+	26,000		
49	29,000	0,948152	0,029183

სადაც ცენზურირებული მონაცემები აღნიშნულია „+“ სიმბოლოთი. ცხრილის პირველ სვეტში ნაჩვენებია საწყის მონაცემთა ფაილში არსებული პაციენტების ნომრები, მეორე სვეტში – ჰოსპიტალში გატარებული დრო, რომელიც როგორც ცხრილიდან ჩანს იცვლება 0-დან 1775 დღემდე. ყველაზე მცირე დღე დაყო 23-ე პაციენტმა, რადგან მან ოპერაციის პირველ დღეს დატოვა საავადმყოფო. ყველაზე ბევრი დღე საავადმყოფოში დაყო მე-15 პაციენტმა, კერძოდ 1775 დღე.

თუ **Cumulative Survival** სვეტში მანაცემები არ არსებობენ, მაშინ პაციენტი გაეწერა ჰოსპიტალიდან (ცენზურირებული დაკვირვება), თუ მნიშვნელობა არსებობს – ე.ი. პაციენტი დაიღუპა და მან იცოცხლა **Time** სვეტში მითითებული დღეების რაოდენობა. მაგალითად, მე-10 პაციენტმა ოპერაციის შემდეგ იცოცხლა 10 დღე.

გადარჩენის ფუნქციის გრაფიკული გამოსახულების ასაგებად Click **Survival times vs. cum. proportion surviving** ღილაკზე. ეკრანზე გამოდის გადარჩენის ფუნქციის გრაფიკული გამოსახულება



სადაც ჩანს, რომ გულის ტრანსპლანტაციის შემდეგ გადარჩენის ფუნქციის მნიშვნელობა მკვეთრად ეცემა პირველ 100 დღეში. ამ მომენტიდან დაწყებული გადარჩენის ფუნქცია კლებულობს უფრო ნელა. აქედან შეიძლება გავაკეთოდ შემდეგი დასკვნა: გულის ტრანსპლანტაციის შემდეგ პირველ 100 დღე ყველაზე უფრო კრიტიკულია. ნახაზზე სრული დაკვირვებები წერტილებით არიან აღნიშნულნი, ხოლო არასრულები – ჯვრებით.

თუ მოვახდენთ **Advanced** ჩანართში **Click Percentiles of survival function** ღილაკზე, მაშინ გამოდის პროცენტების ცხრილი,

Percentiles of (Heart) the Survival Function	
Percentiles	Survival Time
25 <sup>th</sup> percentile (lower quartile)	63,5140
50 <sup>th</sup> percentile (median)	679,1255
75 <sup>th</sup> percentile (upper quartile)	

სადაც ნაჩვენებია, რომ პაციენტების 25% იღუპება პირველ 64 დღეში, ხოლო 50% (მედიანა) – პაციენტები ცოცხლობენ უფრო დიდხანს, არა ნაკლებ 679 დღისა. იღუპება პირველ 64 დღეში, ხოლო 50% (მედიანა) – პაციენტები ცოცხლობენ უფრო დიდხანს, არა ნაკლებ 679 დღისა.

## პრაქტიკული სამუშაო 11

### ბაღარჩენის ანალიზი. ბაღარჩენის ფუნქციების შედარება

#### 1. ორი გადარჩენის ფუნქციის შედარება

გადარჩენის ფუნქციების შედარებისათვის გამოიყენება ხუთი კრიტერიუმი: ვილკოქსონ – გეჰანის, კოქსის  $F$  – კრიტერიუმი, კოქსი – მენტელის კრიტერიუმი, ლოგარითმული რანგული კრიტერიუმი და ვილკოქსონ – პეტის კრიტერიუმი. ამ კრიტერიუმების უმეტესობა განსაზღვრავს სტანტარტიზებული ნორმალური განაწილების  $Z$  მნიშვნელობას. კრიტერიუმების ნაკლი ისაა, რომ მცირე ამონარჩევების დროს ისინი არც ისე საიმედოა არიან. აქედან გამომდინარე, სასურველია გადარჩენის ფუნქციასთან ერთად რიცხვითი კრიტერიუმების გამოყენება.

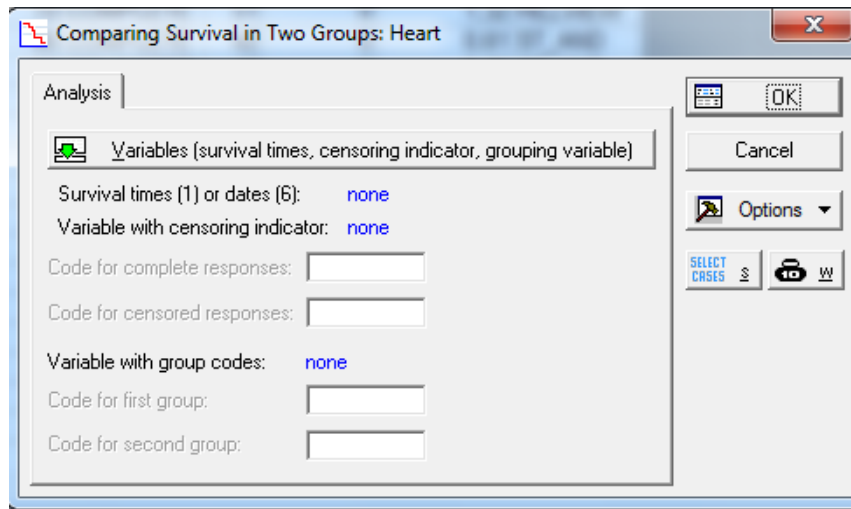
არ არსებობს მკაცრად განსაზღვრული კრიტერიუმების გამოყენების რეკომენდაციები. მაგრამ, ცნობილია, რომ ჩვეულებრივ კოქსის  $F$  – კრიტერიუმს გააჩნია დიდი სიმძლავრე, ვიდრე ვილკოქსონ – გეჰანის კრიტერიუმს, თუ სრულდება შემდეგი პირობები:

- ჯგუფების დაკვირვებათა რაოდენობა მცირეა (ანუ  $n < 50$ );
- მონაცემებს გააჩნიათ ექსპონენციალური ან ვეიბულის განაწილება;
- არ არსებობს ცენზურირებული დაკვირვებები.

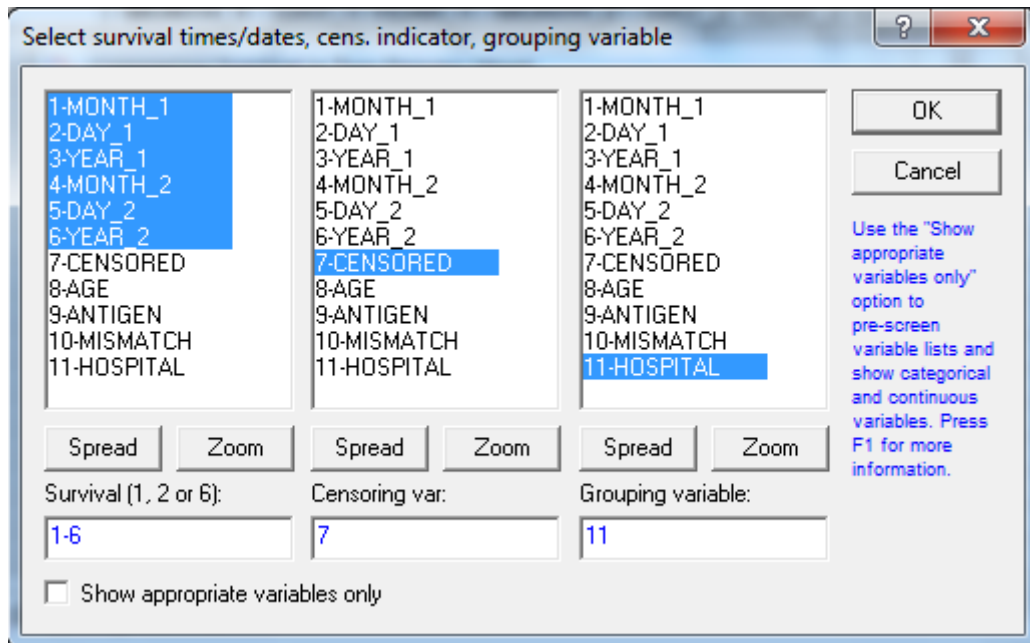
კოქსი – მენტელის და ლოგარითმული რანგული კრიტერიუმებს გააჩნიათ დიდი სიმძლავრე, თუ ამონარჩევები აღებულია ექსპონენციალური ან ვეიბულის განაწილების შემთხვევაში, მაშინ ამ ორ კრიტერიუმს შორის თითქმის არ არსებობს განსხვავება.

გადარჩენის ფუნქციის შედარებისათვის გამოვიყენოთ იგივე *Heart* მონაცემთა ფაილი. გულის გადანერგვის ოპერაციები ჩატარდა სამ კლინიკაში: *Hillview*, *St\_Andrea*, *Biner*. შევადაროთ *Hillview* და *St\_Andrea* კლინიკების გადარჩენის ფუნქციები.

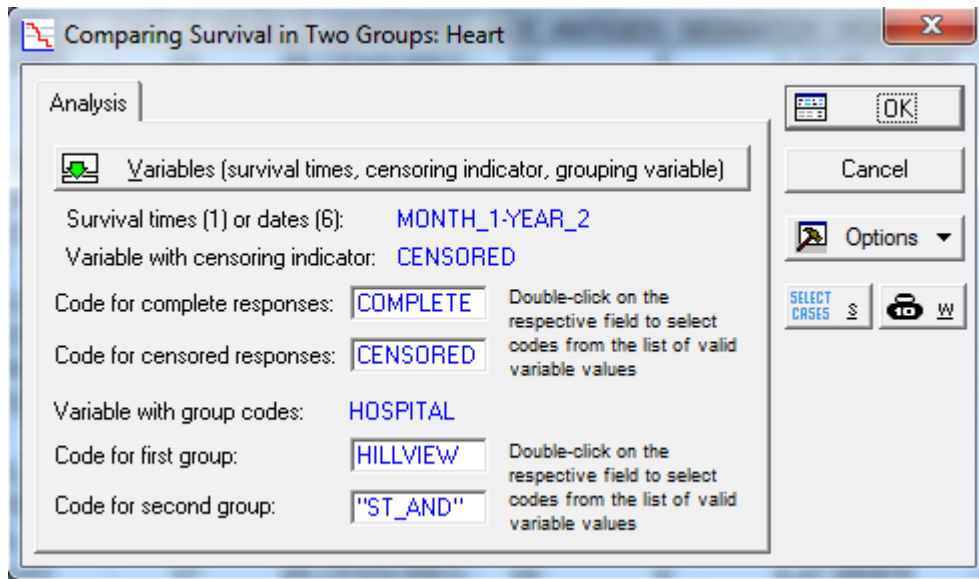
**Survival Analysis** (გადარჩენის ანალიზი) მოდულის გაშვებისათვის **Statistics** მენიუში მოვნიშნოთ **Advanced Linear/Nonlinear Modest** და შევარჩიოთ **Survival Analysis** ბრძანება. ეკრანზე გამოდის **Survival and Failure Time Analysis** სასტარტო ფანჯარა, სადაც **Click Comparing two samples** (ორი ჯგუფის შედარება)-ის სტრიქონზე და შემდეგ **OK**. ეკრანზე გამოდის **Comparing Survival in Two Groups** ფანჯარა.



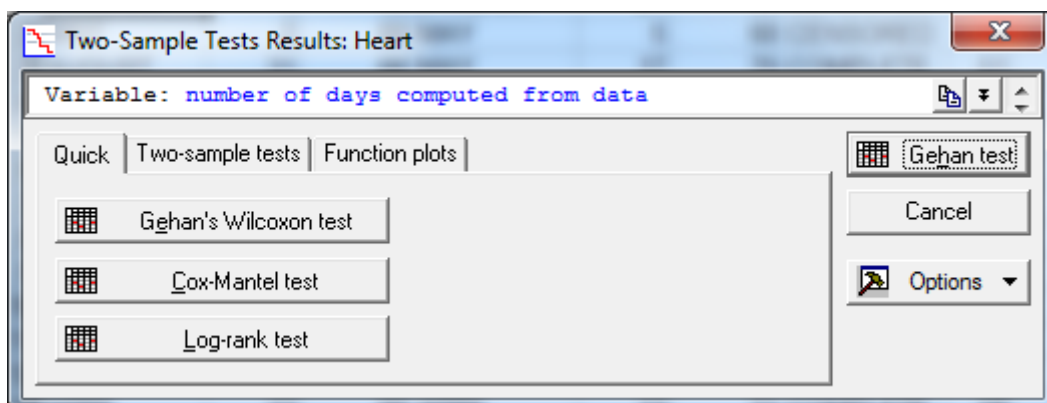
Click **Variables** ღილაკზე და ეკრანზე გამოსულ ფანჯარის პირველ სვეტში მოვნიშნოთ პირველი ექვსი ცვლადი, მეორე სვეტში მე-7 ცვლადი და მესამე სვეტში დამაჯგუფებელ ცვლადად შევარჩიოთ **Hospital**.



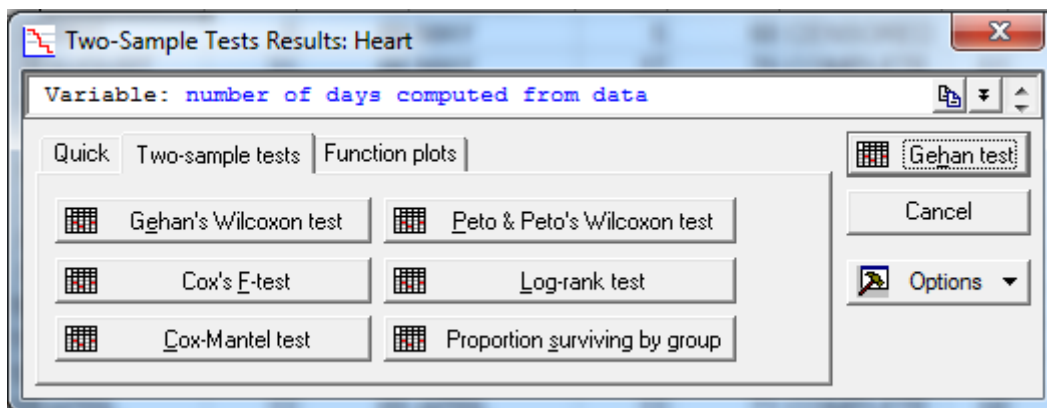
შემდეგ **OK**. ეკრანზე გამოდის **Comparing Survival in Groups** ფანჯარა,



სადაც ორჯერ *Click Code for complete responses, Code for censored responses* ველებში და შევარჩიოთ შესაბამისად *COMPLETE* და *CENSORED*. შემდეგ *OK*. გაიხსნება *Two-Sample Test Results* (ორი კრიტერიუმის შედარების შედეგები)-ის ფანჯარა.



ჩავრთოდ *Two-sample test* ჩანართი. ეკრანზე გამოდის ფანჯარა, სადაც პირველი 5 ლილაკი შეესაბამება კრიტერიუმების დასახელებას.





**Gehans Wilcoxon test** ღილაკზე დაწკაპუნებით ეკრანზე გამოდის შედეგების ცხრილი,

Gehan's Wilcoxon Test (Heart)			
WW = -2,000 Sum = 12456, Var = 3186,4			
Test statistic = -.026573 p = ,97880			
Survival Time	Group	R1	R2
0,0000+	ST_AND	1,00000	1,00000
1,0000+	ST_AND	1,00000	1,00000
3,0000+	HILLVIEW	1,00000	1,00000
10,000	HILLVIEW	1,00000	41,00000
12,000+	HILLVIEW	2,00000	1,00000
13,000+	HILLVIEW	2,00000	1,00000
15,000+	HILLVIEW	2,00000	1,00000
23,000+	HILLVIEW	2,00000	1,00000
25,000	ST_AND	2,00000	36,00000
26,000+	ST_AND	3,00000	1,00000
29,000	ST_AND	3,00000	34,00000
30,000+	ST_AND	4,00000	1,00000
39,000	HILLVIEW	4,00000	32,00000
44 000+	ST_AND	5 00000	1 00000

სადაც პირველ სვეტში დროში ზრდადობის მიხედვით ნაჩვენებია დაკვირვებათა ცხოვრების ხანგრძლივობა (ცენზურირებული დაკვირვებები აღნიშნულია (+) ნიშნით). მეორე სვეტში ნაჩვენებია ჯგუფის დასახელება, რომელსაც მიეკუთვნება დაკვირვება.

**Coxs F-testF** ღილაკზე დაწკაპუნებით ეკრანზე გამოდის შედეგების ცხრილი,

Cox's F-Test (Heart)				
T1 = 10,52661 T2 = 10,47339				
F( 20, 20) = 1,005081 p = ,49553				
Distinct Failures	R(I)	M(I)	M/R	Kap/Meir Estimate
10,000	41,00000	1,000000	0,024390	1,000000
25,000	36,00000	1,000000	0,027778	0,975610
29,000	34,00000	1,000000	0,029412	0,948510
39,000	32,00000	1,000000	0,031250	0,920612
46,000	30,00000	1,000000	0,033333	0,891843
47,000	29,00000	1,000000	0,034483	0,862115
50,000	28,00000	1,000000	0,035714	0,832387
51,000	27,00000	2,000000	0,074074	0,802659
54,000	25,00000	1,000000	0,040000	0,743202
60,000	24,00000	1,000000	0,041667	0,713474
64,000	23,00000	1,000000	0,043478	0,683746
65,000	22,00000	1,000000	0,045455	0,654018
66,000	21,00000	1,000000	0,047619	0,624290
68,000	20,00000	1,000000	0,050000	0,594562
136,00	18,00000	1,000000	0,055556	0,564834
253,00	15,00000	1,000000	0,066667	0,533454
322,00	13,00000	1,000000	0,076923	0,497891
624,00	7,00000	1,000000	0,142857	0,459591
730,00	5,00000	1,000000	0,200000	0,393935
				0,315148

სადაც ნაჩვენებია კოქსის კრიტერიუმის მნიშვნელობა. გარდა ამისა, ცხრილი შეიცავს *Distinct Failures* – ყველა სხვადასხვა დროის მტყუნების სიას (პირველი სვეტი); ობიექტების რაოდენობა, რომლებიც საფრთხის ქვეშ იყვნენ შესაბამისი მტყუნების მომენტებში ( $R(I)$ ); მტყუნების ჯერადობა მტყუნების დადგომის ყოველ მომენტში ( $M(I)$ ) და კაპლან-მეიერის გადარჩენის ფუნქციის შეფასებები.

*Cox-Mantel test* ღილაკზე დაწკაპუნებით ეკრანზე გამოდის შედეგების ცხრილი

Cox-Mantel Test (Heart)	
I = 4,955305 U = -,137266	
Test statistic = -,061663 p = ,95083	
Risk	A(I)
20,00000	0,487805
20,00000	0,555556
18,00000	0,529412
16,00000	0,500000
15,00000	0,500000
14,00000	0,482759
13,00000	0,464286
13,00000	0,481482
12,00000	0,480000
12,00000	0,500000
12,00000	0,521739
11,00000	0,500000
10,00000	0,476191
10,00000	0,500000
9,000000	0,500000
7,000000	0,466667
7,000000	0,538462
4,000000	0,571429
3,000000	0,600000

სადაც გარდა კოქსი-მენტელის სტატისტიკისა წარმოდგენილია ყოველ მტყუნების მომენტში რისკების სია (ობიექტების რაოდენობა, რომლებიც მტყუნების მომენტის წინ იყვნენ ცოცხალნი, სვეტი 1) და მე-2 ჯგუფის ობიექტების წილი, რომლებიც მიეკუთვნებიან შესაბამისი რისკების სიმრავლეს (სვეტი 2).

*Peto & Petos Wilcoxon test* ღილაკზე დაწკაპუნებით ეკრანზე გამოდის შედეგების ცხრილი,

Survival Time	Peto & Peto Wilcoxon Test (Heart)	
	Group	Score
	WW = ,02884 Sum = 10,624 Var = 2,7179 Test statistic = ,0174959 p = ,98604	
0,0000+	ST_AND	0,000000
1,0000+	ST_AND	0,000000
3,0000+	HILLVIEW	0,000000
10,000	HILLVIEW	0,975610
12,000+	HILLVIEW	-0,024390
13,000+	HILLVIEW	-0,024390
15,000+	HILLVIEW	-0,024390
23,000+	HILLVIEW	-0,024390
25,000	ST_AND	0,924119
26,000+	ST_AND	-0,051491
29,000	ST_AND	0,869122
30,000+	ST_AND	-0,079388
39,000	HILLVIEW	0,812455
44,000+	ST_AND	-0,108157
46,000	ST_AND	0,753958
47,000	ST_AND	0,694502
50,000	HILLVIEW	0,635046
51,000	ST_AND	0,545861
51,000	HILLVIEW	0,545861
54,000	HILLVIEW	0,456677
60,000	HILLVIEW	0,397221

სადაც განზოგადოებული ვილკოქსონი – პეტოს კრიტერიუმების გარდა წარმოდგენილია ყველა დაკვირვების სია (ცხოვრების დრო; ცენზურირებული დაკვირვებები აღნიშნულია (+) ნიშნით, სტრიქონი 1) და კრიტერიუმის გამოყენების წვლილების მნიშვნელობები (*Score*).

**Log-rank test** დილაკზე დაწკაპუნებით ეკრანზე გამოდის შედეგების ცხრილი,

Log-Rank Test (Heart)		
WW = ,13727 Sum = 18,825 Var = 4,8158		
Test statistic = ,0625504 p = ,95012		
Survival Time	Group	Score
0,0000+	ST_AND	0,00000
1,0000+	ST_AND	0,00000
3,0000+	HILLVIEW	0,00000
10,000	HILLVIEW	0,97561
12,000+	HILLVIEW	-0,02439
13,000+	HILLVIEW	-0,02439
15,000+	HILLVIEW	-0,02439
23,000+	HILLVIEW	-0,02439
25,000	ST_AND	0,94783
26,000+	ST_AND	-0,05217
29,000	ST_AND	0,91842
30,000+	ST_AND	-0,08158
39,000	HILLVIEW	0,88717
44,000+	ST_AND	-0,11283
46,000	ST_AND	0,85384
47,000	ST_AND	0,81935
50,000	HILLVIEW	0,78364
51,000	ST_AND	0,70957
51,000	HILLVIEW	0,70957
54,000	HILLVIEW	0,66957
60,000	HILLVIEW	0,62790

სადაც სტატისტიკის კრიტერიუმის გარდა ცხრილი შეიცავს ყველა დაკვირვების სიას (ცხოვრების დრო; ცენზურირებული დაკვირვებები აღნიშნულია (+) ნიშნით, სტრიქონი 1); ჯგუფი, რომელსაც მიეკუთვნება თითოეული დაკვირვება (სვეტი 2); კრიტერიუმის გამოყენების წვლილების მნიშვნელობები (Score).

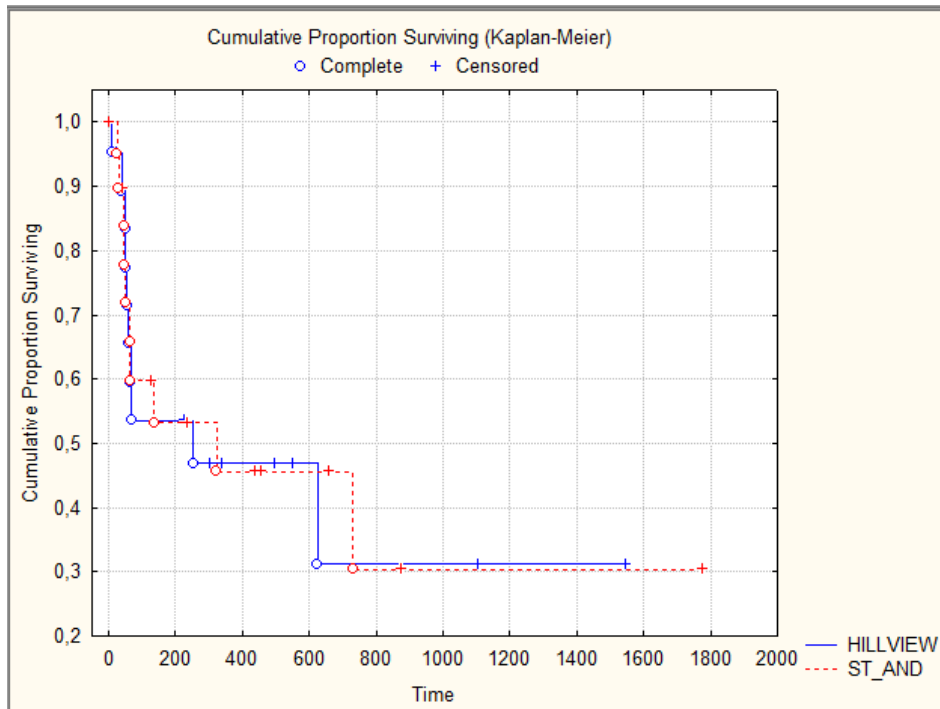
**Proportion Surviving by** დილაკზე დაწკაპუნებით ეკრანზე გამოდის შედეგების ცხრილი შესადარებელი მონაცემებით.

Life Table for Group 1 and Group 2 (Heart)										
Group 1: HILLVIEW Group 2: ST_AND										
Lower Limit	Group 1: No.Enter	Group 2: No.Enter	Group 1: No.Cnsrd	Group 2: No.Cnsrd	Group 1: No.Dying	Group 2: No.Dying	Group 1: % Srwng	Group 2: % Srwng	Group 1: Cum.% Sr	Group 2: Cum.% Sr
0,000000	22	22	5	6	8	8	58,9744	57,8947	100,0000	100,0000
197,2222	9	8	3	1	1	1	86,6667	86,6667	58,9744	57,8947
394,4445	5	6	2	2	0	0	100,0000	100,0000	51,1111	50,1754
591,6667	3	4	0	1	1	1	66,6667	71,4286	51,1111	50,1754
788,8889	2	2	0	1	0	0	100,0000	100,0000	34,0741	35,8396
986,1111	2	1	1	0	0	0	100,0000	100,0000	34,0741	35,8396
1183,333	1	1	0	0	0	0	100,0000	100,0000	34,0741	35,8396
1380,556	1	1	1	0	0	0	100,0000	100,0000	34,0741	35,8396
1577,778	0	1	0	0	0	0	0,0000	100,0000	34,0741	35,8396
1775,000	0	1	0	1	0	0	0,0000	100,0000	0,0000	35,8396

1 და 2 ჯგუფებისათვის მოცემულია ინტერვალის დასაწყისში დაკვირვებების რაოდენობა (No. Enter), ცენზურირებული რაოდენობა (No.Cnsrd), დაღუპულთა რაოდენობა (No.Dying), გადარჩენილთა პროცენტი (%Srwng), გადარჩენილთა კუმულატიური პროცენტი (Cum.%Sr).

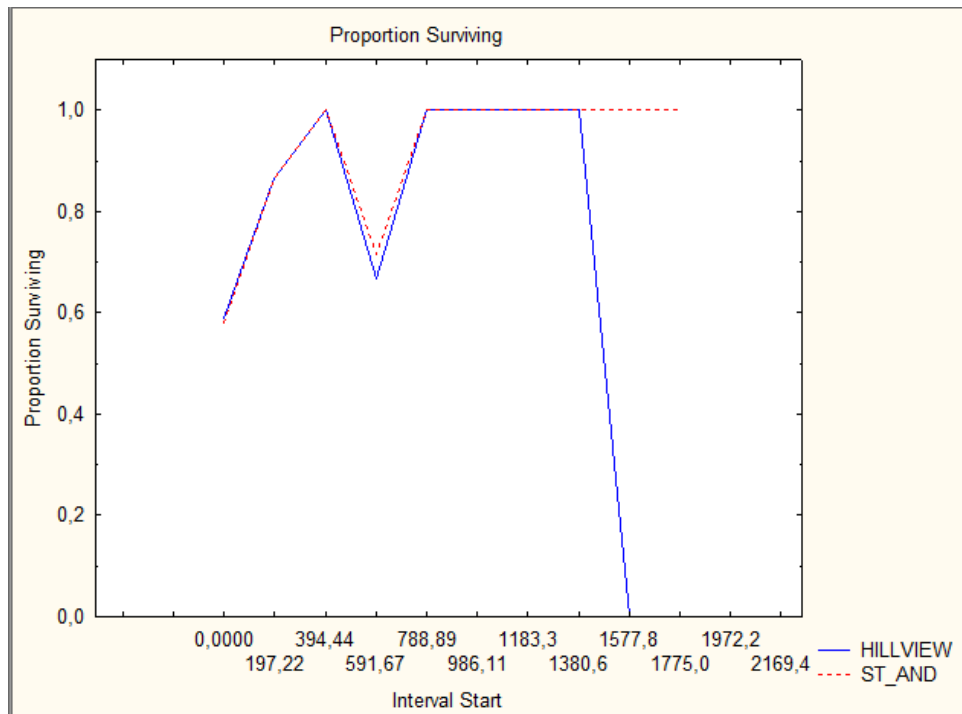
ზემოთ განხილული ყველა 5 კრიტერიუმის  $p$  - მნიშვნელობით საშუალო გადარჩენის ტოლობის ჰიპოთეზა მიიღება ( $p > 0,05$ ). ე.ი. ყველა კრიტერიუმით გამოძინარეობს, რომ ამ ორ ჰოსპიტალს შორის გადარჩენილთა პაციენტების ხანგრძლივობები არ განსხვავდებიან ერთმანეთისაგან.

შედეგების გრაფიკული გამოსახულებისათვის **Two-Sample Test Results** ფარჯარაში ჩაერთოდ **Function plots** ჩანართი და **Click Cum.prop.Surviving by group (Kaplan Meier)** (გადარჩენილთა კუმულატიური წილი) ღილაკზე. ეკრანზე გამოდის ორივე ჯგუფის გადარჩენის ფუნქციის კუმულატიური გრაფიკები.



გრაფიკზე სრული დაკვირვებები აღნიშნულია პატარა წრეებით, ხოლო ცენზურირებული – ჯვრებით. გრაფიკებიდან ჩანს, რომ ორივე ჯგუფისათვის გადარჩენის ფუნქციის კუმულატიური გრაფიკები უმნიშვნელოდ განსხვავდებიან ერთმანეთისაგან.

იმისათვის, რომ ავაგოთ გადარჩენილთა ფარდობითი სიხშირის გრაფიკები ამისათვის **Click Plot of proportion surviving for each group** (ყოველ ჯგუფში გადარჩენილთა წილის გრაფიკი) ღილაკზე.

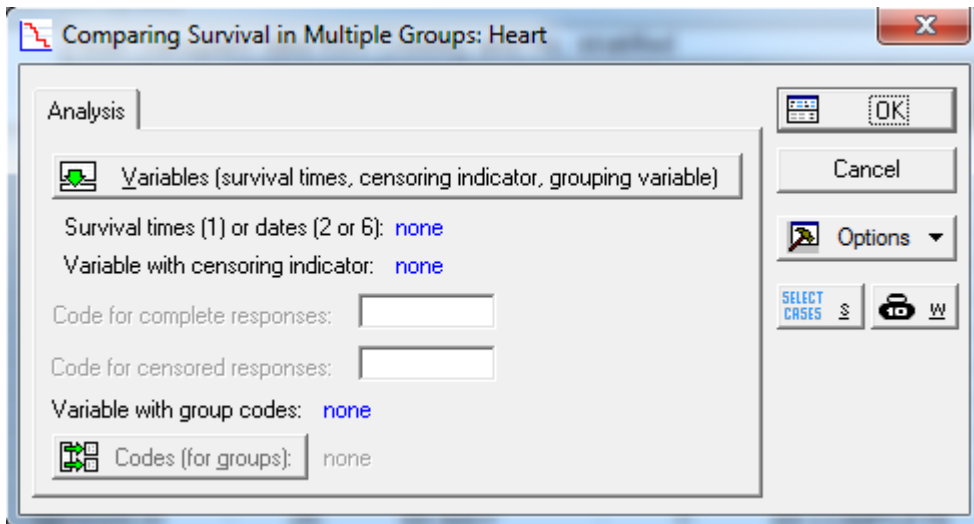


ეკრანზე გამოსულ ფრაფიკულ გამოსახულებიდან ჩანს, რომ აქაც ოპერაციის დღიდან 1380 დღემდე მნიშვნელოვანი განსხვავება არ შეიმჩნევა. ამის შემდეგ გადარჩენილთა წილი Hillview ჰოსპიტალში მკვეთრად მცირდება 1-დან 0-მდე, ხოლო St-Andreas ჰოსპიტალში რჩება უცვლელი და იგი ერთის ტოლია.

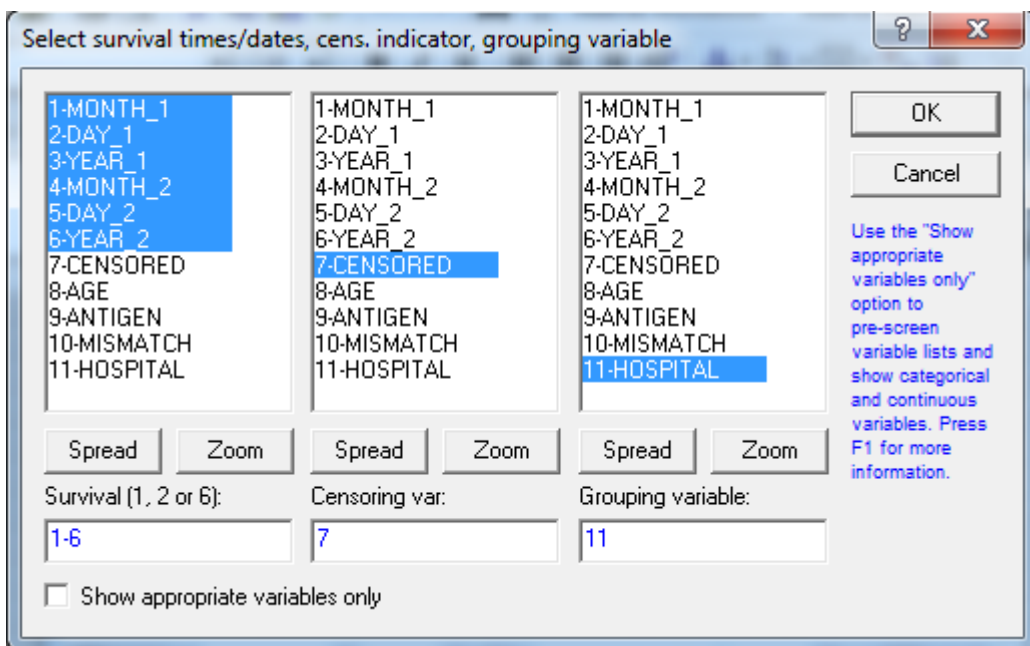
## 2. მრავლობითი გადარჩენის ფუნქციების შედარება

**Heart** მონაცემთა ფაილის საშუალებით ჩავატაროთ სამი კლინიკის **Hillview, St\_Andrea., Biner** გადარჩენის ფუნქციების შედარება.

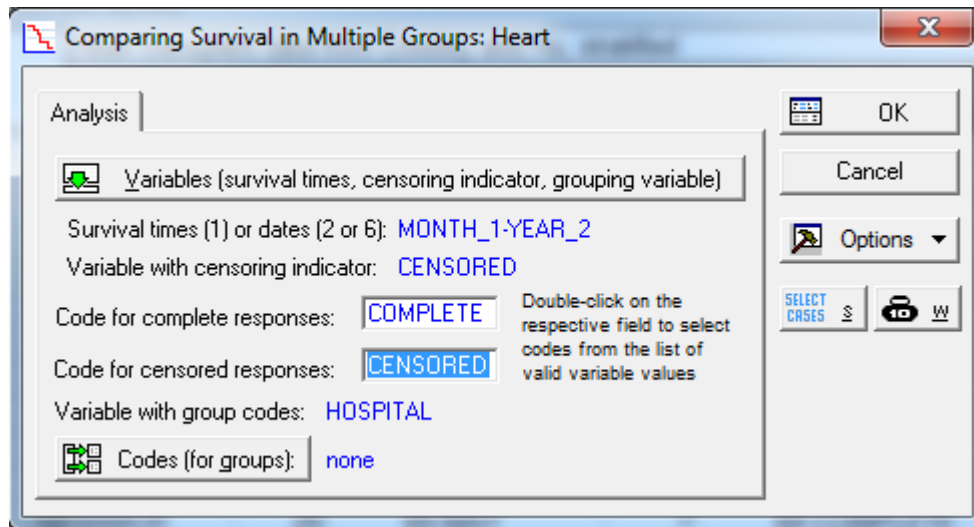
**Statistics** მენიუში მოვნიშნოთ **Advanced Linear/Nonlinear Modest** და შევარჩიოთ **Survival Analysis** ბრძანება. ეკრანზე გამოდის **Survival and Failure Time Analysis** სასტარტო ფანჯარა, სადაც **Click Comparing multiple samples** (რამოდენიმე ამონარჩევის შედარება)-ის სტრიქონზე და შემდეგ **OK**. ეკრანზე გამოდის **Comparing Survival in Multiple Groups** ფანჯარა



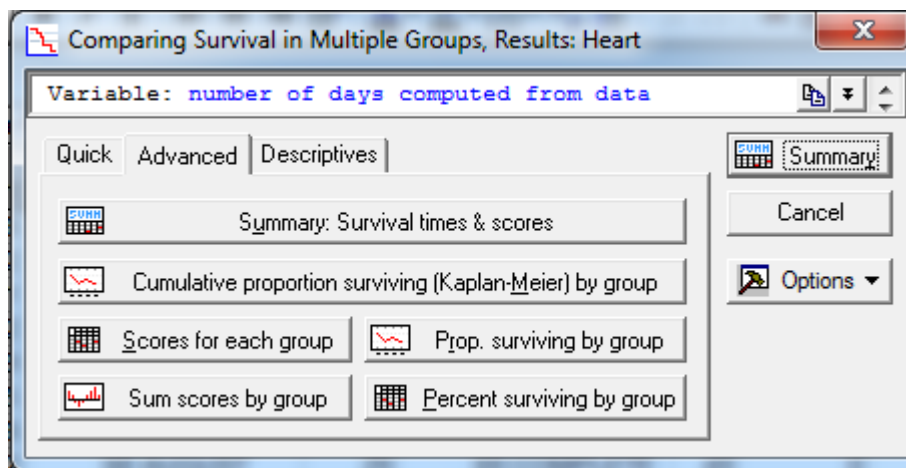
Click **Variables** ღილაკზე და ეკრანზე გამოსულ ფანჯარის პირველ სვეტში მოვნიშნოთ პირველი ექვსი ცვლადი, მეორე სვეტში მე-7 ცვლადი და მესამე სვეტში დამაჯგუფებელ ცვლადად შევარჩიოთ **Hospital**. შემდეგ **OK**.



ეკრანზე გამოდის **Comparing Survival in Multiple Groups** ფანჯარა,



სადაც ორჯერ *Click Code for complete responses, Code for censored responses* ველებში და შევარჩიოთ შესაბამისად *COMPLETE* და *CENSORED*. შემდეგ *OK*. გაიხსნება *Comparing Survival in Groups, Results* (ჯგუფების შედარების შედეგები)-ის ფანჯარა.

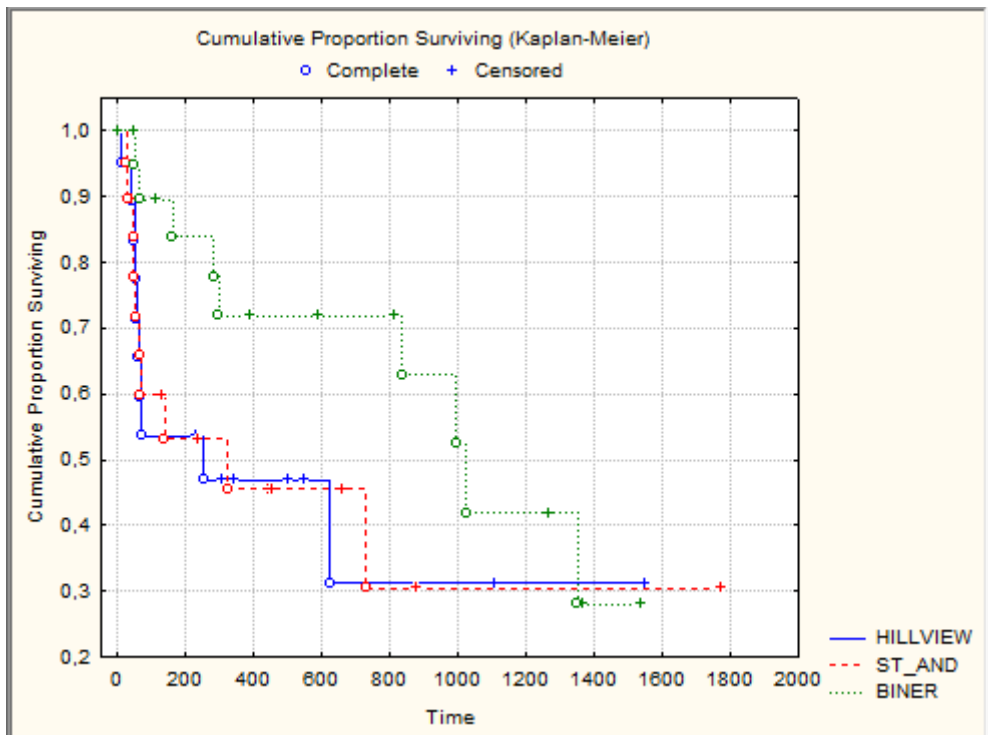


*Summary: Survival times & scores* (ხანგრ ძლივობის დროის და მათი წვლილის შედეგები)-ის დილაკით ეკრანზე გამოდის ყველა ხანგრძლივობა (ცენზურირებული დაკვირვებები აღნიშნულია (+) ნიშნით) და მათი შესაბამისი წვლილები, რომლებიც განისაზღვრებიან მენტელის პროცედურით და გამოიყენებიან მენტელის კრიტერიუმის სტატისტიკის გამოსათვლელად.



Survival Time		Group	Score
Variables: Survival times by HOSPITAL (3 groups) (Heart)			
Censoring var.: CENSORED (Censored cases are marked with +)			
Chi <sup>2</sup> = 5,73999 df = 3 p = ,05671			
0,0000+	ST_AND	0,0000	
1,0000+	ST_AND	0,0000	
1,0000+	BINER	0,0000	
3,0000+	HILLVIEW	0,0000	
10,000	HILLVIEW	-60,0000	
12,000+	HILLVIEW	1,0000	
13,000+	HILLVIEW	1,0000	
15,000+	HILLVIEW	1,0000	
23,000+	HILLVIEW	1,0000	
25,000	ST_AND	-54,0000	
26,000+	ST_AND	2,0000	
29,000	ST_AND	-51,0000	
30,000+	ST_AND	3,0000	
39,000	HILLVIEW	-48,0000	
44,000+	ST_AND	4,0000	
46,000	ST_AND	-45,0000	
47,000	ST_AND	-43,0000	
48,000+	BINER	6,0000	
50,000	HILLVIEW	-39,0000	

Click *Cumulative proportion surviving (Kaplan-Meier) by Group* (ჯგუფებში გადარჩენილთა ფარდობითი სისშირე)-ის ღილაკზე. ეკრანზე გამოდის სამივე ჯგუფის გადარჩენილთა კუმულატიური გრაფიკები. ამ გრაფიკებზე დასრულებული დაკვირვებები აღნიშნულია წრეთი, ხოლო ცენზურირებული დაკვირვებები ჯვრით.

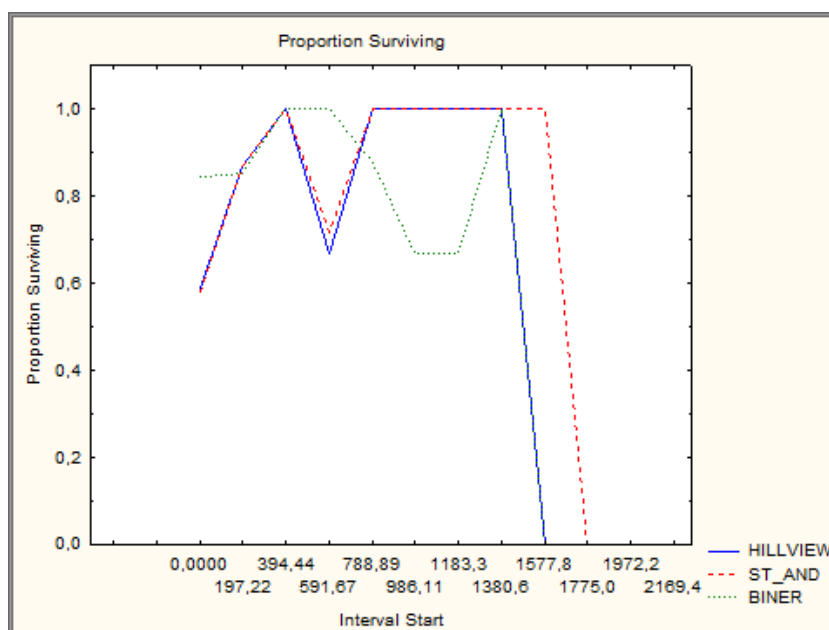


*Click Scores for each group* (ჯგუფების ფარდობითი სიხშირე)-ის ღილაკზე. ეკრანზე გამოდის ცხრილი,

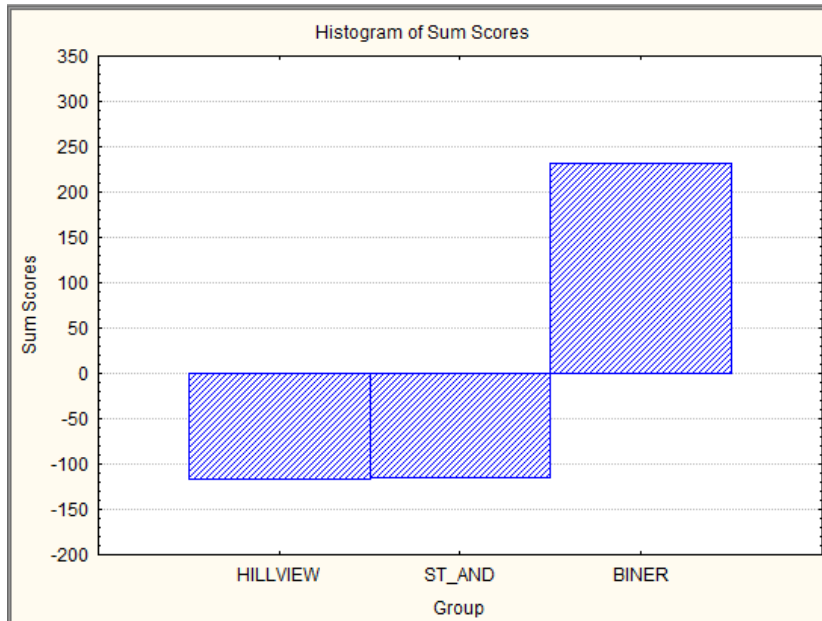
Summary statistics for each group (Heart)							
Group	Code for Group	Sum of Scores	N.uncsrd	Percent uncensrd	N.censrd	Percent censored	Total N
HILLVIEW	1	-116,000	10	45,45454	12	54,54546	22
ST_AND	2	-115,000	10	45,45454	12	54,54546	22
BINER	3	231,000	9	42,85714	12	57,14286	21

სადაც წარმოდგენილია თითოეული ჯგუფის ჯამური ფარდობითი სიხშირე, რომელებიც მიღებულია, მენტელის პროცედურით და გამოიყენებიან მენტელის კრიტერიუმის სტატისტიკის გამოსათვლელად. გარდა ამისა, ეკრანზე გამოდის თითოეული ჯგუფის დასრულებული და ცენზურირებული დაკვირვებების ჯამის რაოდენობა და პროცენტი.

**Prop. surviving by group** (ჯგუფებში გადარჩენილთა ფარდობითი სიხშირეთა გრაფიკები) ღილაკით ეკრანზე გამოდის სამივე ჯგუფის გადარჩენილთა ფარდობითი სიხშირის გრაფიკები.



**Sum scores by group** (ჯგუფებში გადარჩენილთა ფარდობითი სიხშირის ჰისტოგრამა) ღილაკით ეკრანზე გამოდის მენტელის პროცედურით გამოთვლილი ჯგუფების ფარდობითი სიხშირეების ჯამის ჰისტოგრამა.



**Click Percent surviving by group** (ჯგუფებში გადარჩენილთა ფარდობითი სიხშირე) ღილაკზე. ეკრანზე გამოდის სამივე ჯგუფის ხანგრძლივობა პროცენტებში.

Life Table for Group 3 (Heart)					
Group: BINER					
Lower Limit	No. Enter	No. Cnsrd	No. Dying	% Survng	Cum. % Sr
0,000000	21	4	3	84,2105	100,0000
197,2222	14	1	2	85,1852	84,2105
394,4445	11	1	0	100,0000	71,7349
591,6667	10	1	0	100,0000	71,7349
788,8889	9	2	1	87,5000	71,7349
986,1111	6	0	2	66,6667	62,7680
1183,333	4	2	1	66,6667	41,8454
1380,556	1	1	0	100,0000	27,8969
1577,778	0	0	0	0,0000	27,8969
1775,000	0	0	0	0,0000	0,0000

**Summary: Survival times & scores** შედეგების ფანჯრის ინფორმაციულ ნაწილში მოცემულია ხი-კვადრატ და *p* კრიტერიუმების მნიშვნელობები ( $Chi^2=5,73999$ ,  $p=0,05671$ ). რადგან *p*-მნიშვნელობა უმნიშვნელოდ მეტია 0,05 სიდიდეზე, ამიტომ შეგვიძლია გამოვიტანოთ დასკვნა იმის შესახებ, რომ გულის გადანერგვის ოპერაციების შედეგები სამივე ჰოსპიტალში, პაციენტის გადარჩენის თვალსაზრისიდან გამომდინარე, განსხვავდებიან ერთმანეთისაგან.

მიღებული ცხრილები და გრაფიკები გვიჩვენებენ, რომ პაციენტების ხანგრძლივობა **Hillview** და **St Andrea** კლინიკებში დაახლოებით თანაბარია, მაგრამ ისინი განსხვავდებიან **Biner** კლინიკის გადარჩენისაგან. ამასთან, როგორც გადარჩენილთა ფარდობითი სიხშირის გრაფიკებიდან ჩანს **Biner** კლინიკის გადარჩენილთა კუმულატიური ფუნქცია საწყის პერიოდში უფრო ნაკლებად კლებულობს დანარჩენ ორ კლინიკასთან შედარებით, ხოლო შემდეგ დროის გრძელ პერიოდში გადარჩენის ფუნქციის მნიშვნელობა

აღმატება *Hillview* და *St\_Andrea* კლინიკების გადარჩენის ფუნქციებთან შედარებით. პირველი 787 დღის განმავლობაში გადარჩენილთა ფარდობითი სისშირის მნიშვნელობა *Biner* კლინიკაში გაცილებით მაღალია *Hillview* და *St\_Andrea* კლინიკებთან შედარებით.

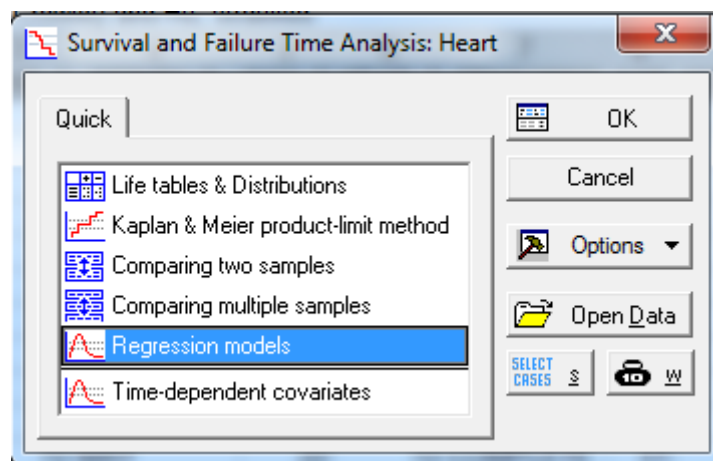
ამრიგად, შეიძლება გამოვიტანოთ დასკვნა იმის შესახებ, რომ რაღაც მიზეზების გამო *Biner* კლინიკის პაციენტებს გააჩნიათ გადარჩენის უფრო დიდი შანსი არა მარტო გულის ტრანსპლანტაციის პირველ კრიტიკულ დღეებში, არამედ შემდგომ პერიოდშიც.

## პრაქტიკული სამუშაო 12

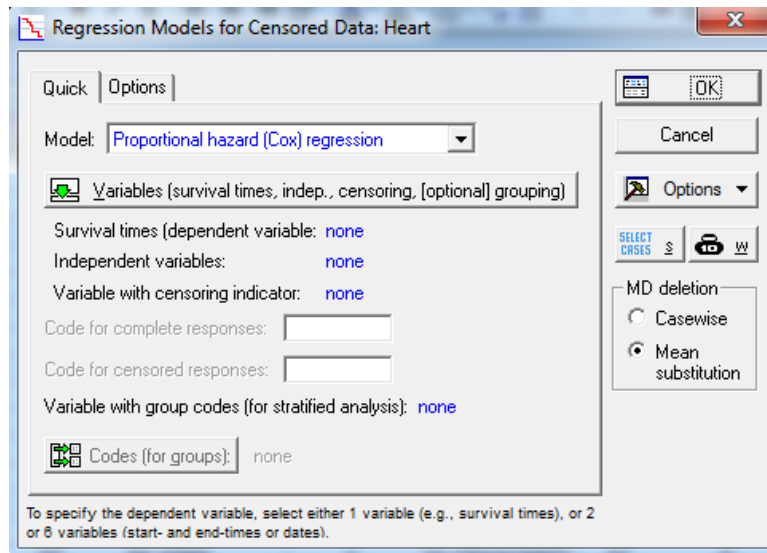
### გადარჩენის ანალიზი. რეგრესიული მოდელები

#### 1. კოქსის პროპორციული ინტენსიურობის მოდელი

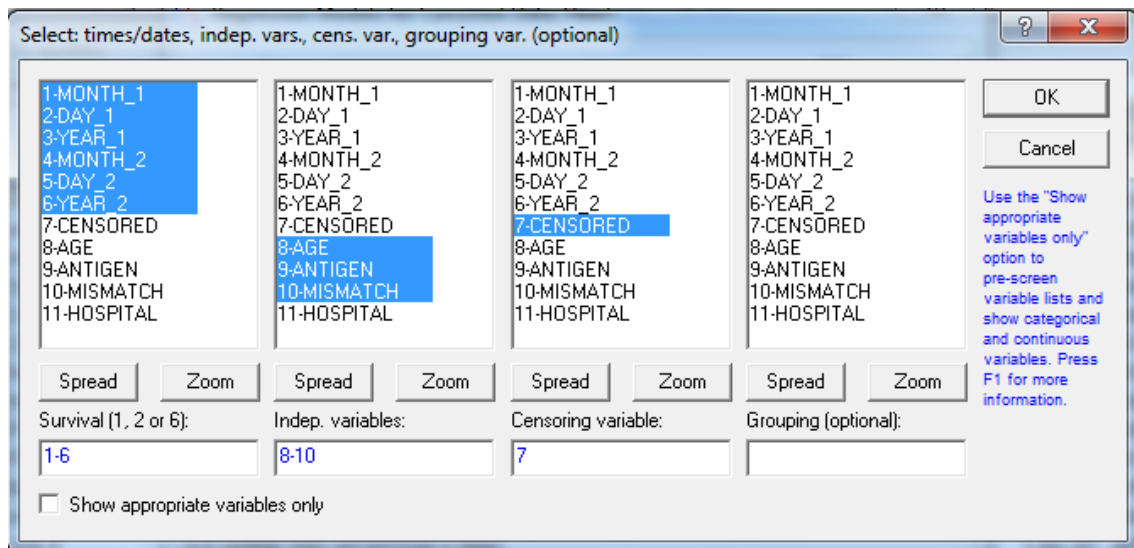
რეგრესიული ანალიზისათვის გამოვიყენოთ იგივე *Heart* მონაცემთა ფაილი. *Survival Analysis* (გადარჩენის ანალიზი) მოდულის გაშვებისათვის *Statistics* მენიუში მოვნიშნოთ *Advanced Linear/Nonlinear Modest* და შევარჩიოთ *Survival Analysis* ბრძანება. ეკრანზე გამოდის *Survival and Failure Time Analysis* სასტარტო ფანჯარა,



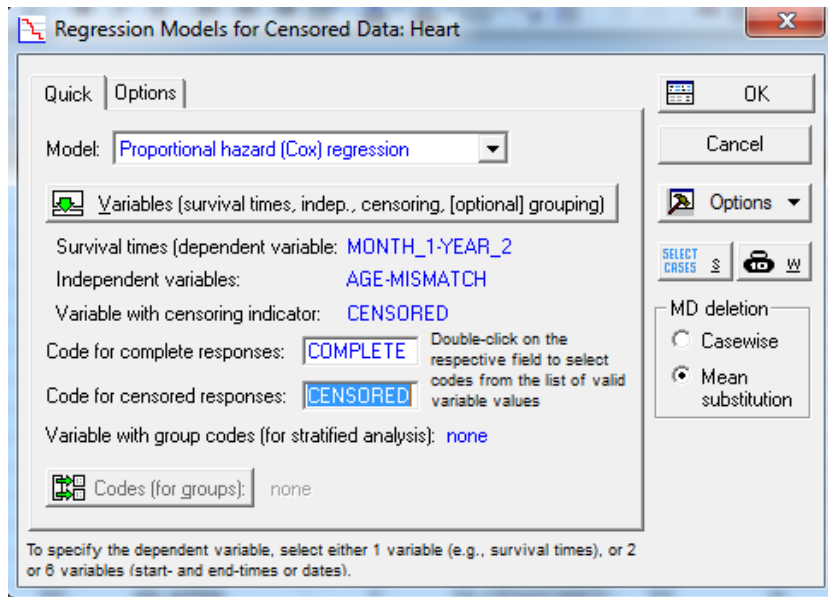
Click *Regression Models* ღილაკზე. ეკრანზე გამოდის ამავე დასახელების ფანჯარა



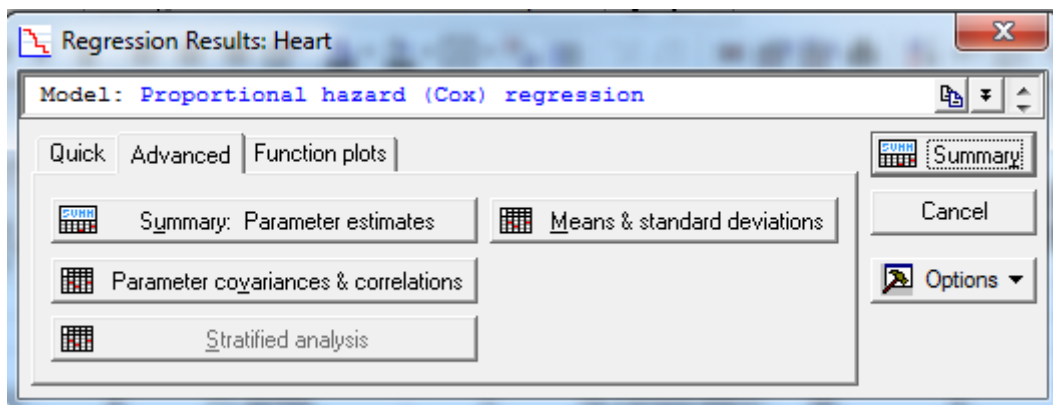
ცვლადების შესარჩევად *click Variables* ღილაკზე და ეკრანზე გამოსულ ფანჯარის პირველ ველში მოვნიშნოთ პირველი 6 ცვლადი, მეორე ველში – **AGE**, **ANTIGEN**, **MISMATCH** დამოუკიდებელი ცვლადები (კოვარიანტები) და ბოლოს მესამე ველში – **CENSORED** ცვლადი.



შემდეგ *OK*. ვბრუნდებით წინა ფანჯარაში და ორჯერ *click Code for complete responses*, *Code for censored responses* ველებში და შევარჩიოთ შესაბამისად **COMPLETE** და **CENSORED**.



შემდეგ *OK*. პროგრამა გაუშვებს პარამეტრების შეფასების იტერაციულ ციკლს, რომლის დამთავრების შემდეგ ეკრანზე გამოდის **Regression Results** ფანჯარა



ჩაერთოდ **Advanced** ჩანართი და პარამეტრების შეფასებისათვის *click Summary: Parameter estimates* ღილაკზე. ეკრანზე გამოდის ცხრილი,

Dependent Variable: Survival times in days (Heart)						
Censoring var.: CENSORED						
Chi <sup>2</sup> = 22,3193 df = 3 p = ,00006						
N=65	Beta	Standard Error	t-value	exponent beta	Wald Statist.	p
AGE	0,109096	0,033293	3,276836	1,115269	10,73766	0,001051
ANTIGEN	-0,048782	0,471644	-0,103431	0,952388	0,01070	0,917622
MISMATCH	1,063761	0,394599	2,695804	2,897246	7,26736	0,007026

სადაც ფანჯრის ინფორმაციული ნაწილიდან ჩანს, რომ ხი-კვადრატ სტატისტიკა სარწმუნოა ( $p \gg 0,05$ ), ამიტომ შეგვიძლია გავაკეთოდ დასკვნა იმაზე, რომ ზოგიერთი ცვლადი დაკავშირებულია ხანგრძლივობასთან. ცხრილის პირველ სვეტში (**Beta**) მოცემულია პარამეტრების შეფასებები

ე.ყუბანეიშვილი. მონაცემების სტატისტიკური დამუშავება

(რეგრესიის განტოლებაში შესაბამისი ცვლადების კოეფიციენტები), მეორე სვეტში (*Standard Error*) – სტანდარტული შეცდომები, მესამე სვეტში –  $t$ -კრიტერიუმის მნიშვნელობები, რომლებიც ტოლია პირველი და მეორე სვეტების მონაცემების ფარდობისა. მე-5 და მე-6 სვეტებში მოყვანილია ვალდის (*Wald Ststist.*) სტატისტიკები და  $p$  მნიშვნელობები. ჩვეულებრივ პრაქტიკაში სტატისტიკურად სარწმუნოდ ითვლება (როცა  $p < 0,05$ ) ისეთი პარამეტრების შეფასებები, რომლებიც ორჯერ აღემატება ამ პარამეტრის ცდომილებას ( $t > 2$ ).

ცხრილის მონაცემებიდან გამომდინარეობს, რომ ასაკი (*AGE*) და ქსოვილების შეუთავსებლობა (*MISMATCH*) წარმოადგენენ მნიშვნელოვან (სარწმუნო) ინტენსივობის (რისკის) ფუნქციის პრედიქტორებს. თანაც ქსოვილების შეუთავსებლობა რისკის უფრო მნიშვნელოვანი პრედიქტორია, ვიდრე ასაკი (*MISMATCH*-ის კოეფიციენტები 10-ჯერ უფრო დიდია, ვიდრე *AGE*-ის კოეფიციენტები), ხოლო *ANTIGEN* ცვლადის კოეფიციენტი კოქსის რეგრესიულ მოდელში შეიძლება ჩაითვალოს ნულის ტოლად, რადგან მისი შესაბამისი  $t$ -კრიტერიუმის მნიშვნელობა მნიშვნელოვნად მცირეა 2-ზე და  $p > 0,05$ .

დავბრუნდეთ რეგრესიის შედეგების ფანჯარაში და *Click Parameter covariances & correlations* ღილაკზე. ეკრანზე გამოდის პარამეტრების კორელაციური მატრიცა:

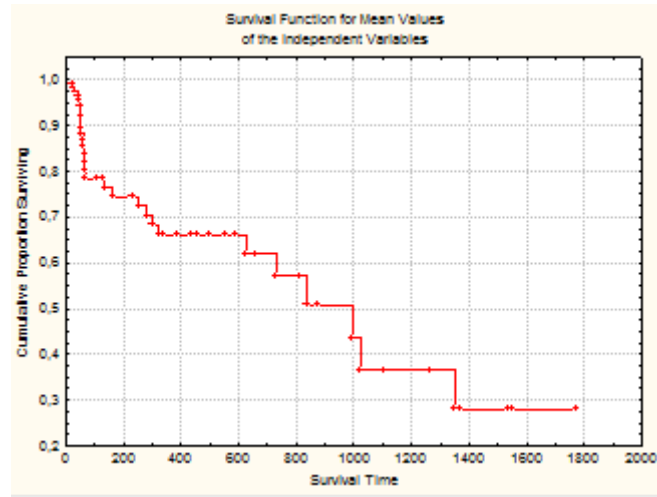
Variable	Parameter Correlations (Heart)		
	AGE	ANTIGEN	MISMATCH
AGE	1,000000	-0,043937	0,040090
ANTIGEN	-0,043937	1,000000	-0,425334
MISMATCH	0,040090	-0,425334	1,000000

*Click Means & standard deviations* ღილაკზე. ეკრანზე გამოდის შედეგების ცხრილი,

variable	Means and Standard Deviations (Heart)			
	mean	st. dev.	minimum	maximum
AGE	45,6769	9,1858	19,00000	64,000
ANTIGEN	0,2615	0,4429	0,00000	1,000
MISMATCH	1,1646	0,6233	0,00000	3,050
No.days	382,6769	463,2327	0,00000	1775,000

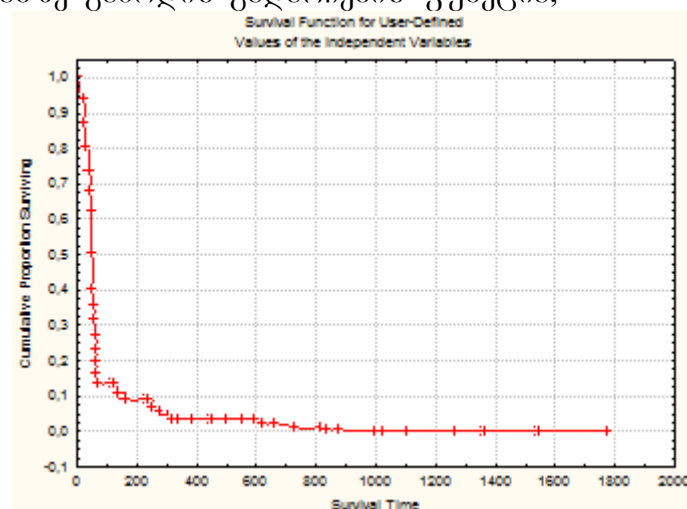
სადაც წარმოდგენილია დამოუკიდებელი (კოვარიანტების) საშუალო და სტანდარტული გადახრების შეფასებები.

გადარჩენის გრაფიკების ასაგებად ჩავრთოდ *Functions plots* ჩანართი და *click Graph survival function for means* ღილაკზე. პროგრამას ეკრანზე გამოაქვს *AGE*, *ANTIGEN*, *MISMATCH* ცვლადების საშუალო მნიშვნელობების გადარჩენის გრაფიკი.



პროგრამაში გათვალისწინებულია გადარჩენის ფუნქციის აგება, როცა კოვარიანტების მნიშვნელობებს მომხმარებელი თვითონ ადგენს. ამისათვის, *click Graph survival function for spec. vals.* ლილაკზე. ეკრანზე გაიხსნება *Independent Variable Values* ფანჯარა

სადაც გაჩუმების პრინციპით წარმოდგენილია ცვლადების საშუალო მნიშვნელობები. გაუზარლოთ *AGE* არეში ასაკი მაგალითად, 65 წლამდე და შემდეგ *OK*. ეკრანზე გამოდის გადარჩენის ფუნქცია,

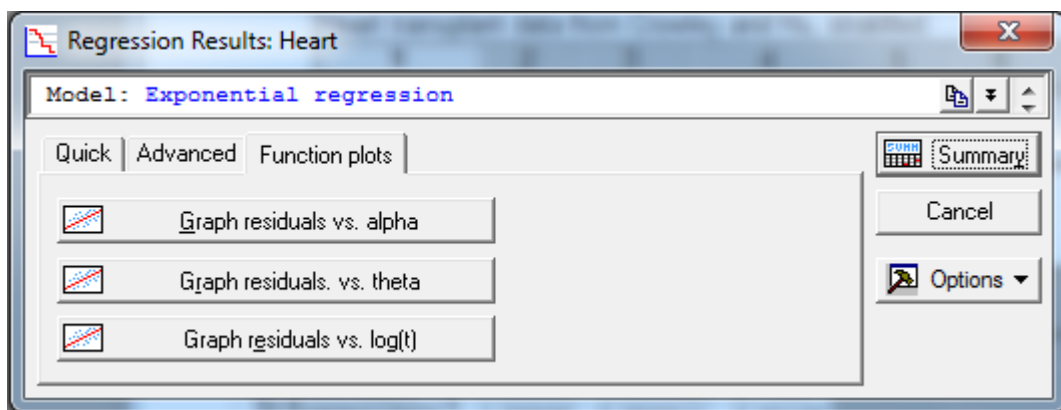




სადაც თვალნათლივ ჩანს, რომ გადარჩენის ფუნქცია მნიშვნელოვნად შემცირდა. ალბათობა იმისა, რომ პაციენტი იცოცხლებს ოპერაციიდან 50 დღეზე მეტად ტოლია 0,13, ხოლო გადარჩენის ალბათობა 800 დღის შემდეგ ნულის ტოლია, მაშინ როცა 45 წლის პაციენტისათვის ეს ალბათობა 0,5 ტოლია.

## 2. ექსპონენციალური რეგრესია

**Regression Models for Censored Data** ფანჯრის **Model** ველში დავაყენოთ **Exponential regression**. სხვა დაყენებული პარამეტრები დავტოვოთ უცვლელად. შემდეგ **OK**. ეკრანზე გამოდის **Regression Results** ფანჯარა



პარამეტრების შეფასების ცხრილის გამოსატანად *click Summary: Parameter estimates* ღილაკზე

Dependent Variable: Survival times in days (Heart)				
Censoring var.: CENSORED				
Chi <sup>2</sup> = 32,0812 df = 3 p = ,00000				
	Beta	Standard Error	t-value	
N=65				
AGE	-0,12418	0,032857	-3,77938	
ANTIGEN	-0,06363	0,459082	-0,13861	
MISMATCH	-1,22008	0,365911	-3,33436	
Constant	14,08542	1,686244	8,35313	

სადაც ჩანს, რომ **Exponential regression** მოდელი უფრო ადეკვატურია ვიდრე კოქსის მოდელი, რადგან მისი ხი-კვადრატის მნიშვნელობა გაცილებით მეტია ( $32,0812 > 22,0812$ ), ხოლო  $p$  მნიშვნელობა ნაკლებია. გარდა ამისა, აქაც ცხადად ჩანს, რომ ქსოვილების შეუთავსებლობა წარმოადგენს მნიშვნელოვან პრედიქტორს, ვიდრე ასაკი, ხოლო **ANTIGEN** ცვლადის კოეფიციენტი აქაც შეიძლება ჩაითვალოს ნულის ტოლად.

### პრაქტიკული სამუშაო 13

#### დროითი მწკრივების ანალიზი და პრობნოზირება

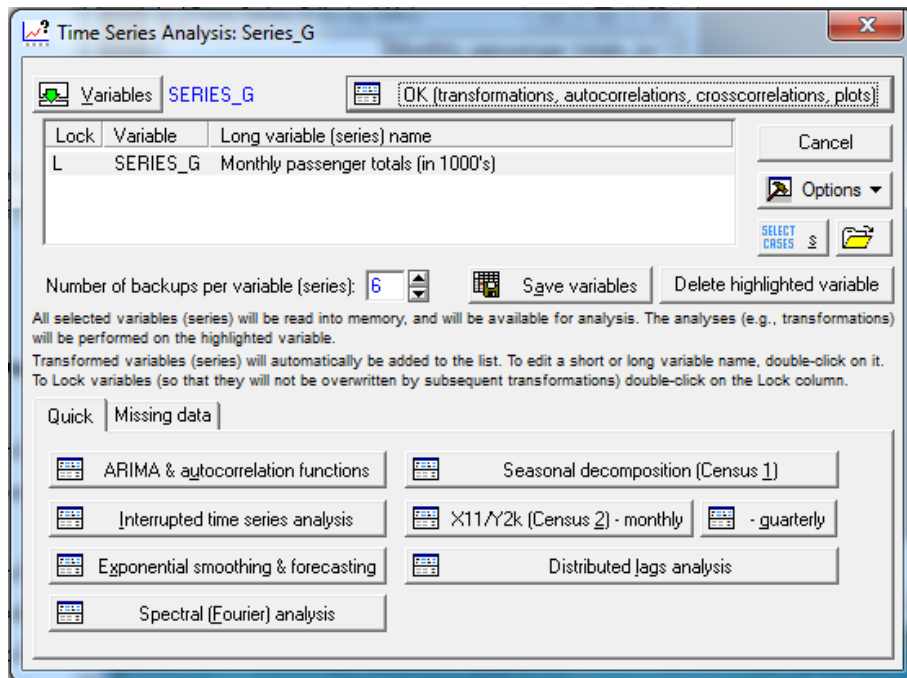
ჩვენ ვიყენებთ *Statistica*-ის *Exampfes* ბიბლიოთეკის *Dataset* მონაცემთა ბაზაში არსებული *Series.G* ფაილი, სადაც წარმოდგენილია თვეების მიხედვით ავიაგადაზიდვის მონაცემები 1949–1960წწ (სულ 144 დაკვირვება).

Monthly passenger totals (in 1000's) 1949-1960; Box & Jenkins, 1976; series G.	
1	SERIES_G
JAN 1949	112
FEB 1949	118
MAR 1949	132
APR 1949	129
MAY 1949	121
JUN 1949	135
JUL 1949	148
AUG 1949	148
SEP 1949	136
OCT 1949	119
NOV 1949	104
DEC 1949	118
JAN 1950	115
FEB 1950	126
MAR 1950	141
APR 1950	135
MAY 1950	125
JUN 1950	149
JUL 1950	170
AUG 1950	170
SEP 1950	158

#### დროითი მწკრივის ანალიზის მოდულის აღწერა

*Statistica* მენიუში მოვნიშნოთ *Advanced Linear/Nonlinear Models* და *Click Time Series/Forecasting* ბრძანებაზე,

ეკრანზე გამოდის *Time Series Analysis* (დროითი მწკრივების ანალიზი)-ის სასტარტო ფანჯარა.



click **Variables** ღიაკზე და თუ მონიშნული არ არის მოვნიშნოთ დროითი მწკრივი. შემდეგ **OK**. გბრუნდებით სასტატო ფანჯარაში.

– ფანჯრის საინფორმაციო ველში **Variables** ველში ნაჩვენებია ფაილის დასახელება, ხოლო **Lock** გრაფაში ნაჩვენებია **L** ნიშანი, რაც იმას ნიშნავს, რომ ცვლადები (ფაილები) დაკეტილია გასაღებით და მათი გამორიცხვა ანალიზიდან შეუძლებელია. **Long variable (series) name** ველში ნაჩვენებია ფაილის გაფართოებული სახელი.

მუშაობის პროცესში ხშირად საჭიროა სასურველი გარდაქმნების შერჩევა და იმისათვის, რომ არ შევინახოთ ზედმეტი ინფორმაცია, საჭიროა მისი წაშლა. ამისათვის გამოიყენება **Delete highlighted variable** (გამონათებული ცვლადების წაშლა) ღილაკი. თუ შემდგომი კვლევებისათვის საჭიროა ზოგიერთი გარდაქმნის შენახვა (მაგ. **Statistica**-ის სხვა მოდულისათვის) ამისათვის გამოიყენება **Save variables** (ცვლადების შენახვა)-ის ღილაკი.

**Number of backups per variable** (ცვლადების სარეზერვო რაოდენობა) ველში მითითებულია გარდაქმნის რაოდენობა. თუ გარდაქმნის რაოდენობა აღემატება ველში წარმოდგენილ რიცხვს, მაშინ პროგრამა შეგვეკითხება შევინახოთ თუ არა აღნიშნული გარდაქმნა.

**Select cases** (დაკვირვებების შერჩევა) ღილაკი გამოიყენება ანალიზისათვის დაკვირვებების შესარჩევად.

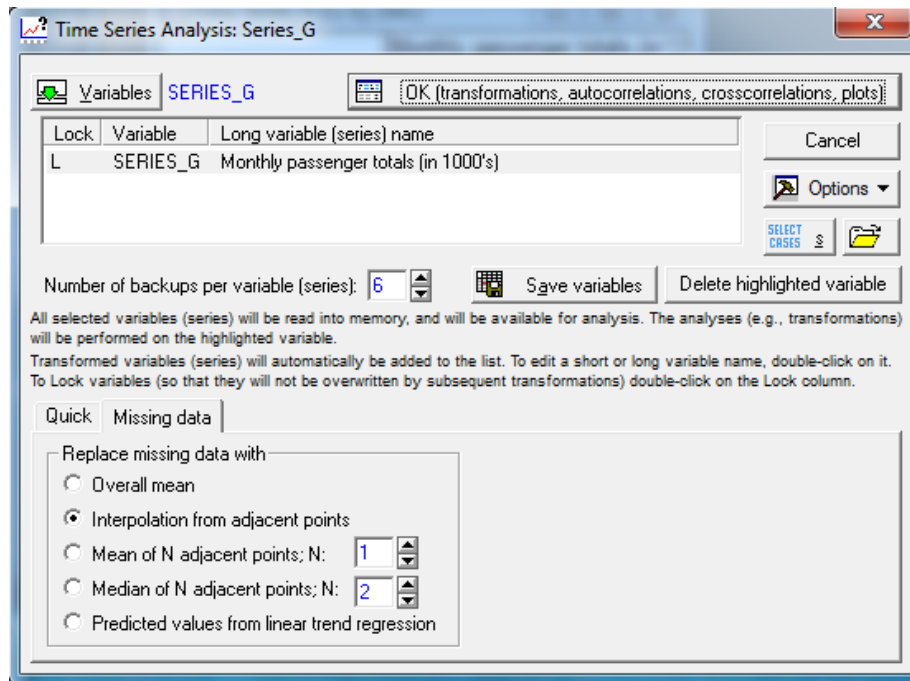
ღილაკი **OK (transformations, autocorrelations, plots)** (გარდაქმნა, ავტოკორელაცია, გრაფიკები) დროითი მწკრივის გარდაქმნისათვის გახსნის სპეციალურ ფანჯარას.

სასტატო ფანჯრის **Quick** ჩანართში განლაგებულია შემდეგი ფუნქციონალური ღილაკები:

- **ARIMA & autocorrelation functions** (ავტორეგრესიის და მცოცავი საშუალოს გაინტეგრირებელი მოდელი);
- **Interrupted time series analysis** (შეწყვეტილი დროითი მწკრივების ანალიზი ან **ARIMA** მოდელი ინტერვეციით);

- **Exponential smoothing & forecasting** (ექსპონენციალური გაგლუვება და პროგნოზირება);
- **Seasonal decomposition (Census 1)** (სეზონური დეკომპოზიცია);
- **XIIy2k (Census 2)-monthly** (12 თვიანი სეზონური დეკომპოზიცია);
- **Quarterly** (კვადტალური სეზონური კორექტირება);
- **Distributed Lags Analysis** (ლაგების (ძვრა) განაწილების ანალიზი);
- **Spectral (fourier) Analysis** (ფურიეს სპექტრული ანალიზი).

**Missing data** ჩანართში წარმოდგენილია გამოტოვებული მონაცემების შევსების სხვადასხვა საშუალებები:

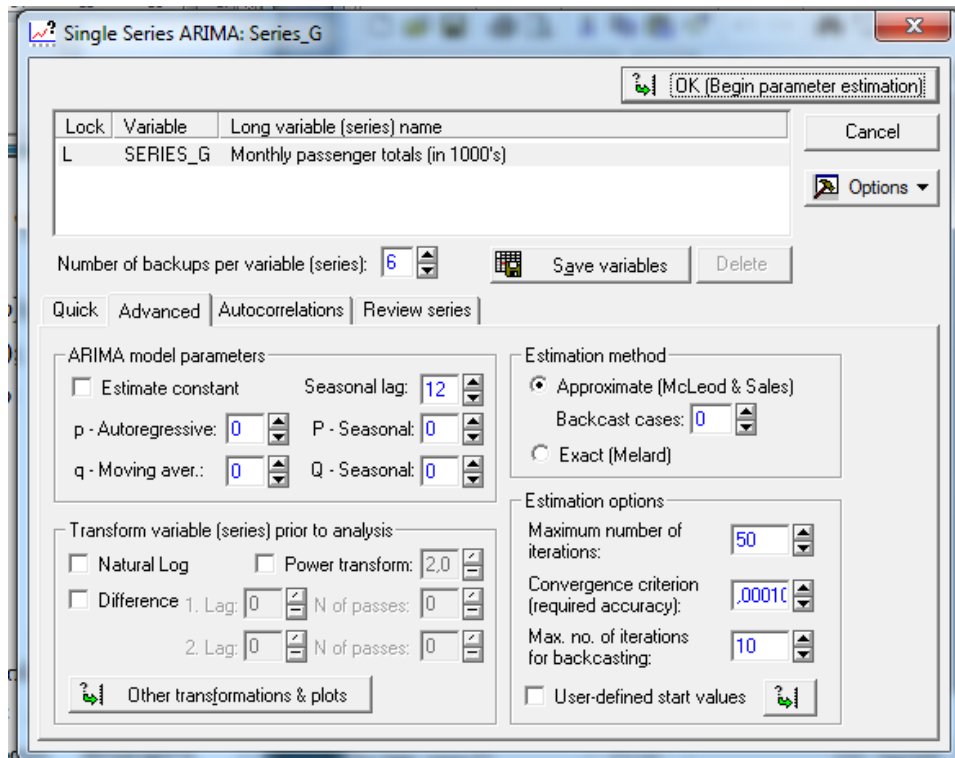


- **Overall mean** (საერთო საშუალო);
- **Interpolation from adjacent points** (მეზობელი წერტილებით ინტერპოლაცია);
- **Mean of N adjacent points** (მეზობელი წერტილების საშუალო);
- **Median of N adjacent points** (მეზობელი წერტილების მედიანა);
- **Predicted values from linear trend regression** (წრფივი ტენდენციის რეგრესიის საშუალებით მონაცემების პროგნოზირება).

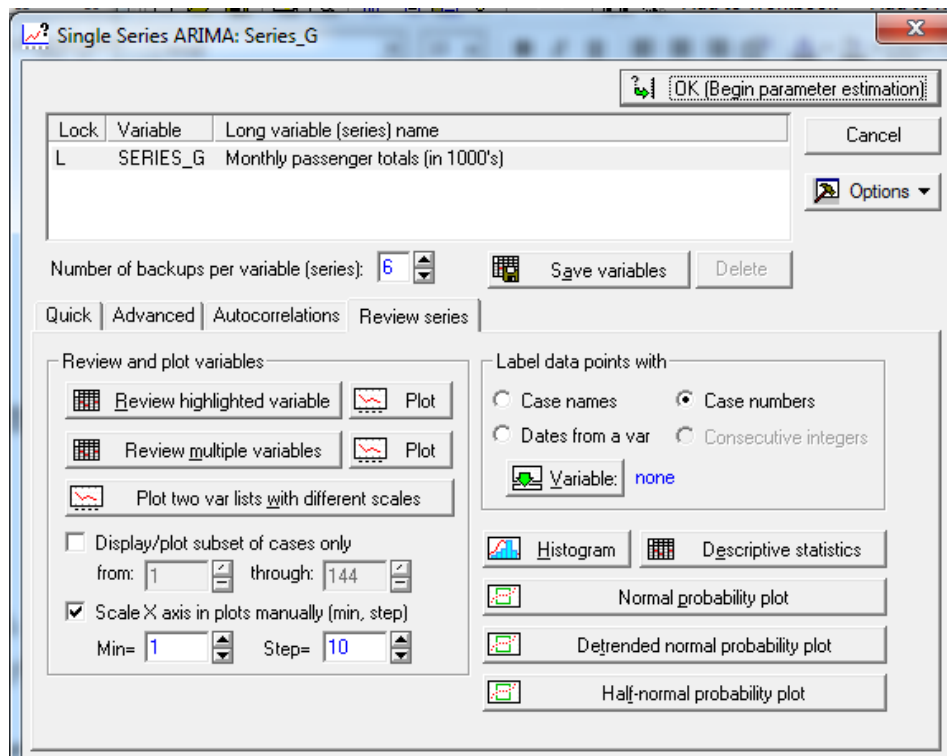
## 1. პროგნოზირების ARIMA მოდელი

**ARIMA** წარმოადგენს პარამეტრულ მეთოდს, რომელიც გამოიყენება როგორც სტაციონარული, ისე არასტაციონარული მწკრივებისათვის. ამიტომაც, რომ ამ მეთოდს გააჩნია დიდი პრაქტიკული გამოყენება. **Statistica** პროგრამაში **ARIMA** რეალიზირებულია ბოკსი და ჯენკინსის მეთოდოლოგიით. მოდელს შეუძლია ჩართოს მუდმივი სიდიდე.

**ARIMA** მოდელის გააქტიურებისათვის სასტარტო ფანჯარაში *click ARIMA & autocorrelation functions* ლილაკზე. ეკრანზე გაიხსნება *Single Series ARIMA* სასტარტო ფანჯარა.



საზოგადოდ, დროითი მწკრივის ანალიზი იწყება მისი გრაფიკული წარმოდგენით. ამისათვის ჩაერთოთ *Review Series* (მწკრივის დათვალიერება)-ს ჩანართი.



ე.ყუბანეიშვილი. მონაცემების სტატისტიკური დამუშავება

ფანჯრის ზედა ინფორმაციულ ველში ჩაწერილია მწკრივის სახელი და მისი გარდაქმნა. ფანჯრის **Review and plotvariables** (ცვლილების დათვალიერება და გრაფიკების აგება)-ის ჩარჩოში განთავსებულია შემდეგი ღილაკები და ოპციები:

- **Review highlighted variable** (გამონათებული ცვლადების დათვალიერება);
- **Review multiple variables** (რამოდენიმე ცვლადის დათვალიერება);
- **Plot** – გრაფიკი;
- **Plot two var list with different scales** (სხვადასხვა სკალის ორი სიიდან ცვლადების გრაფიკები);
- **Display/plot subset of cases only** (მხოლოდ ქვესიმრავლის ეკრანზე ჩვენება/გრაფიკის აგება)

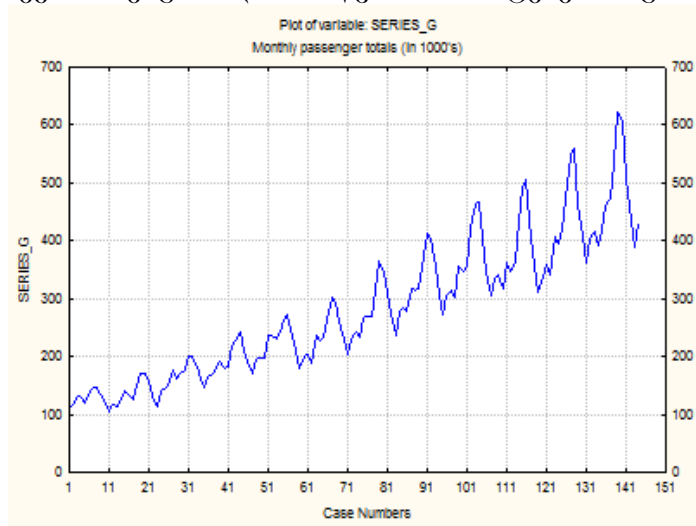
ფანჯრის მარჯვენა მხარეს **Label data points with** (მონაცემთა წერტილების ნიშანი) ჩარჩოში განთავსებულია შემდეგი ოპციები:

- **Case names** (დაკვირვების სახელი)
- **Case numbers** (დაკვირვების ნომერი)
- **Dates from a var** (იმ ცვლადის მნიშვნელობა, რომლის სახელი ნაჩვენებია **Variable** ღილაკის საშუალებით)

და სხვადასხვა ტიპის ალბათური გრაფიკების აგების ღილაკები.

გრაფიკების აგებისას საჭიროა **Scale X axis in plots manual** ოპციის ჩართვა და **Min** = ველში მიუთითოთ მწკრივის ის ნომერი საიდან იწყება გრაფიკის აგება, **Step** = ველში უნდა მიუთითოთ X დერძისათვის ბიჯის სიდიდე.

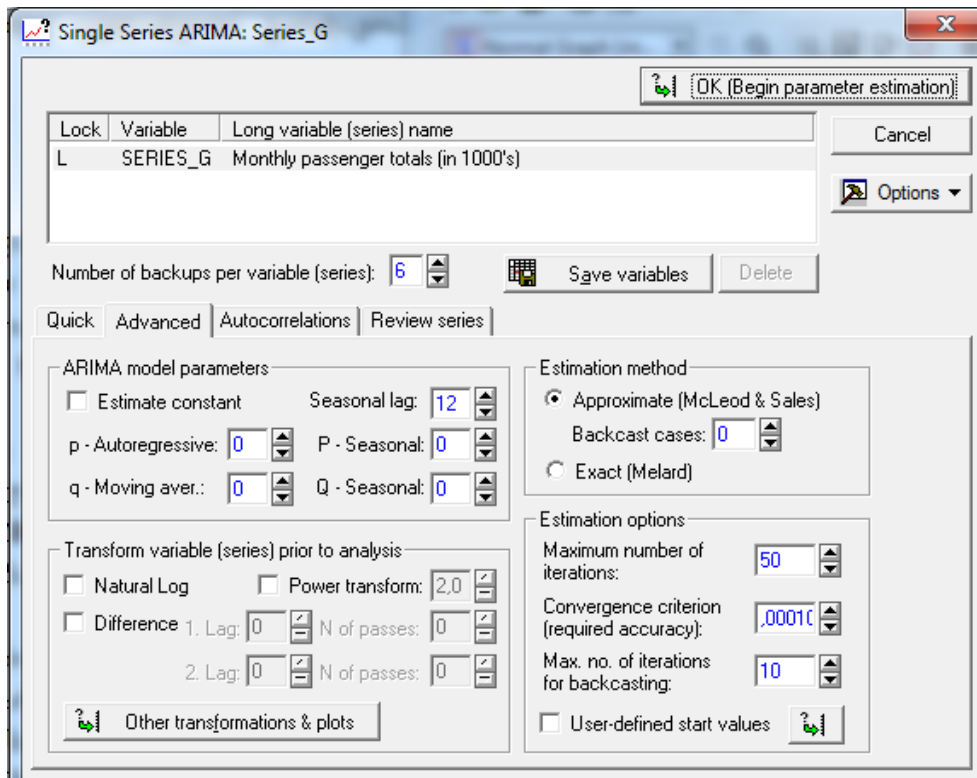
**Label data points with** ჩარჩოში ჩაერთოთ **Case names** ან **Case number** ოპცია. შემდეგ **click Plot** ღილაკზე. რომელიც მდებარეობს **Review highlighted variable** ღილაკის გვერდით. ეკრანზე გამოდის საწყისი მონაცემების გრაფიკი:



“ხარისხიანი” პროგნოზირებისათვის მწკრივი უნდა აკმაყოფილებდეს გარკვეულ კრიტერიუმებს. ასე მაგალითად, ის უნდა იყოს შედარებადი და ერთგვაროვანი, გარდა ამისა ტრენდის გამოსავლენად იგი უნდა იყოს მდგრადი და სრული ანუ დაკვირვებათა რაოდენობა უნდა იყოს საკმარისად დიდი.

მონაცემების ერთგვაროვნება გულისხმობს მწკრივის განვითარების ტენდენციაში მკვეთრი ცვლილებების და აგრეთვე არატიპიური და ანომალური დაკვირვებების გამორიცხვას.

დავუბრუნდეთ **Single Series ARIMA** სასტარტო ფანჯარას და ჩავრთოთ **Advanced** ჩანართი, სადაც **ARIMA model parameters** ჩარჩოში მოთავსებულია შემდეგი ოპციები:



- **p-Autoregressive** (ავტორეგრესიის პარამეტრი (რეგულარული));
- **P-Seasonal** (ავტორეგრესიის სეზონური პარამეტრი);
- **q -Moving aver** (მცოცავი საშუალოს პარამეტრი (რეგულარული));
- **Q-Seasonal** (მცოცავი საშუალოს სეზონური პარამეტრი)

ყოველი ამ პარამეტრისათვის მოიცემა პარამეტრების რიცხვი. უკიდურეს შემთხვევაში რომელიმე ერთი პარამეტრი განსაზღვრული უნდა იყოს. ოპციებში მითითებული პარამეტრები ნიშნავს იდენტიფიკაციას ანუ მოდელის განსაზღვრას.

**Transform variable prior to analysis** (ანალიზამდე ცვლადების გარდაქმნა) ჩარჩოში წარმოდგენილია შემდეგი ოპციები:

- **Natural log** (ნატურალური გალოგარითმება);
- **Diference** (გამოკლება);
- **Power transform** (ხარისხში აყვანა).

ამ ფანჯრიდან გამოსვლის შემდეგ აღნიშნული გარდაქმნები არ შეინახება.

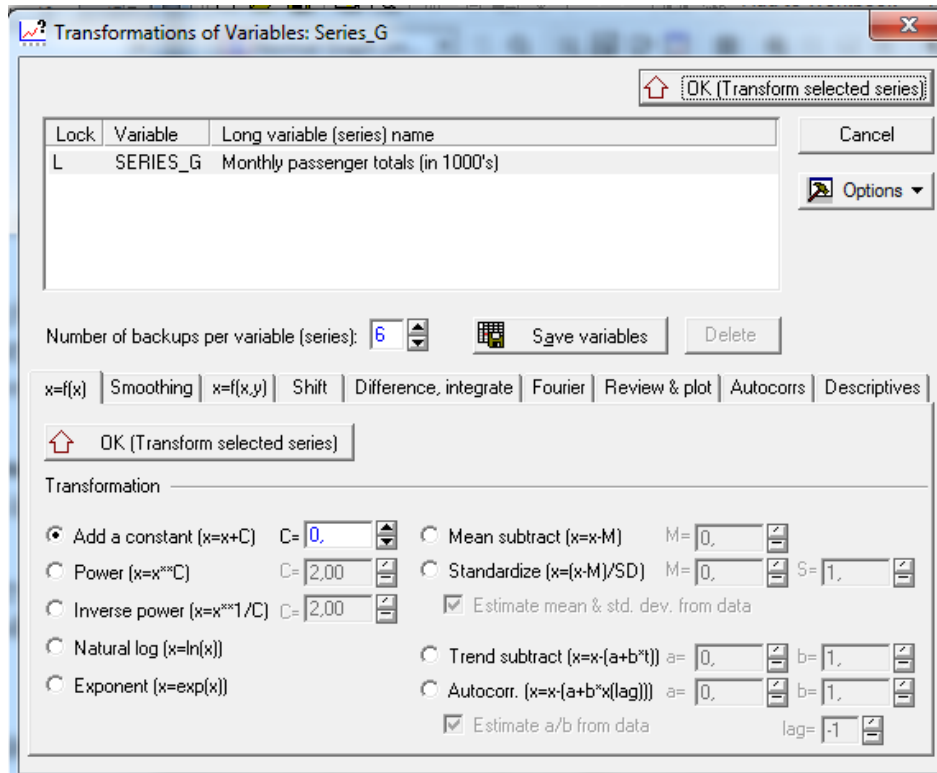
**Estimation method** (შეფასების მეთოდი) ჩარჩოში წარმოდგენილია შეფასების ორი გამოთვლითი მეთოდი: **Approximate** (აპროქსიმაცია) და **Exact** (ზუსტი).

**Estimation options** (შეფასების პარამეტრები) ჩარჩოში მოიცემა უცნობი პარამეტრების საწყისი მიახლოება; მიეთითება იტერაციის მაქსიმალური

რაოდენობა და დაყენება შეფასების პროცედურის კრებადობის კრიტერიუმის პარამეტრი. ყველა ეს მნიშვნელობები შესაძლებელია მოგვცეს სისტემამ.

ფანჯრის ზედა მარჯვენა მხარეს არსებული **OK(Begin parameter estimation)** (პარამეტრების შეფასების დაწყება) ღილაკით იწყება პარამეტრების შეფასების პროცედურა.

**Other transformations & plots** (სხვა გარდაქმნები და გრაფიკები) ღილაკით გაიხსნება **Transformations of Variables** (ცვლადების გარდაქმნა) ფანჯარა რომლის საშუალებითაც ხდება გარდაქმნის შედეგების დამახსოვრება.



ეკრანზე გამოსულ ამ ფანჯრის  $x = f(x)$  ჩანართში წარმოდგენილია შემდეგი გარდაქმნები:

- **Add a constant** ( $x = x+c$ ) (მწკრივის მნიშვნელობებს დაემატოთ მუდმივა);
- **Power** ( $x = x**c$ ) (ხარისხში აყვანა);
- **Inverse power** ( $x = x**1/c$ ) (შებრუნებულ ხარისხში აყვანა);
- **Natural log** [ $x = \ln(x)$ ] (ნატურალური ლოგარითმის აღება);
- **Exponent** [ $x = \exp(x)$ ] (ექსპონენციალური გარდაქმნა);

ამ ფანჯარაში არის შემდეგი ოპციები:

- **Mean subtract** [ $x = x - M$ ] (საშუალოს გამოკლება);
- **Standardize** [ $x = (x - M) / SD$ ] (სტანდარტიზირება). მწკრივის მნიშვნელობებს აკლდება საშუალო და იყოფა სტანდარტულ  $SD$  გადახრაზე.
- **Trend subtract**  $t$  (ტრენდის გამორიცხვა). მწკრივიდან გამოირიცხება წრფივი ტრენდი, რომლის პარამეტრები ან შეფასდებიან ან მოიცემა  $a, b$  ზოლში;



- **Autocorr** [ $x = x - (a + b * x(lag))$ ] (ავტოკორელაცია). ეს არის წრფივი გარდაქმნა, რომელიც საშუალებას იძლევა გავანულოთ ავტოკორელაცია გარკვეულ ლაგზე, რომელიც მოცემულია *lag* ველში

**Smoothing** (გაგლუვება) ჩანართში შესაძლებელია მწკრივის შემდეგი გარდაქმნები.

- **N-pts mov. averg.** (*N* წერტილიანი სრიალა საშუალო);
- **N-pts mov. median** (*N* წერტილიანი სრიალა მედიანა);
- **Weighted** (გასაშუალება არათანაბარი წონებით);
- **Drior** (გამოთვლები წარმოებს მწკრივის წინა მნიშვნელობებით);
- **Simple exponential** (მარტივი ექსპონენციალური გაგლუვება);
- **4253H filter**

$x=f(x,y)$  ჩანართში შესაძლებელია მწკრივის შემდეგი გარდაქმნები:

- **Difference** (სხვაობა). გამოითვლება მწკრივის ახალი მნიშვნელობები შემდეგი ფორმულით  $x = x - y(lag)$ , სადაც *lag* (დაყოვნება) მოიცემა *lag* ველში.
- **Residualizing**. გამოითვლება მწკრივის ახალი მნიშვნელობები შემდეგი ფორმულით:  $x = x - (a + by(lag))$ , სადაც *a* და *b* პარამეტრები წინასწარ მოცემულია ან განისაზღვრება უმცირეს კვადრატთა მეთოდით. ბოლო შემთხვევისას საჭიროა ჩავრთოთ **Estimate a and b from data** (*a* და *b* პარამეტრების შეფასება) ოპცია.

ამ ჩანართის ოპციები გამოიყენება იმ შემთხვევაშიც, როცა ორი დროითი მწკრივის ანალიზია საჭირო.

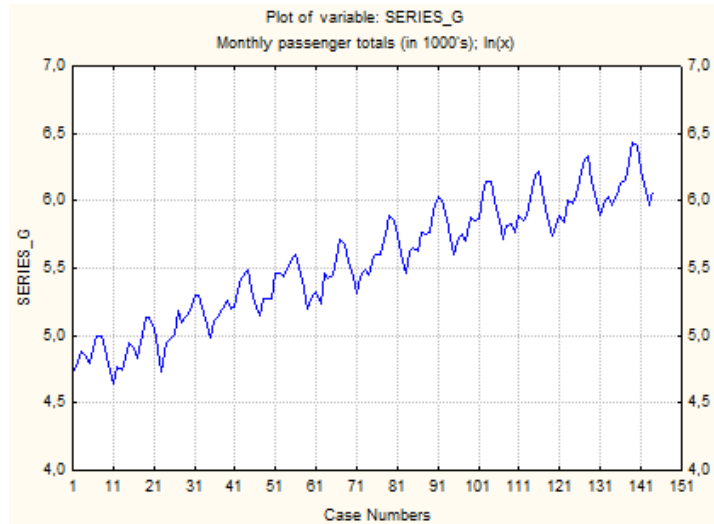
**Shift** (ძვრა) ჩანართს გააჩნია შემდეგი ოპციები:

- **Shift (lag) series forward** (დაყოვნების დაძვრა წინ);
- **Shift (lag) series back** (დაყოვნების დაძვრა უკან).

**Difference, integrate** (გამოკლება, მიმატება). ჩანართით განისაზღვრება ახალი მწკრივის მნიშვნელობები ფორმულით:  $x = x - x(log)$  ან  $x = x + x(log)$ .

დანარჩენი ჩანართებით ხდება სხვადასხვა გრაფიკების აგება, ავტოკორელაციის გამოთვლა და გარდაქმნილი მწკრივის აღწერითი სტატისტიკების განსაზღვრა.

გასაანალიზირებელი მწკრივის დისპერსიის შესამცირებლად გამოიყენება  $x = f(x)$  ჩანართის **Natural log** ოპცია. **click OK (Transform selected series)** ღილაკზე. ეკრანზე გამოდის საწყისი მწკრივის გალოგარითმებული მნიშვნელობები, რომლის დისპერსია გაცილებით ნაკლებია, ვიდრე საწყისი მწკრივისა.



ამის შემდეგ საჭიროა **ARIMA** მოდელის პარამეტრების იდენტიფიკაცია. **ARIMA** მოდელს გააჩნია შემდეგი პარამეტრები:  $p$ -ავტორეგრესიის რიგი,  $d$ -სხვაობის რიგი,  $q$ -სრიალა საშუალოს რიგი. ე.ი **ARIMA** ( $p, d, q$ ) მოდელის იდენტიფიცირება ნიშნავს განისაზღვროს ეს სამი პარამეტრი.

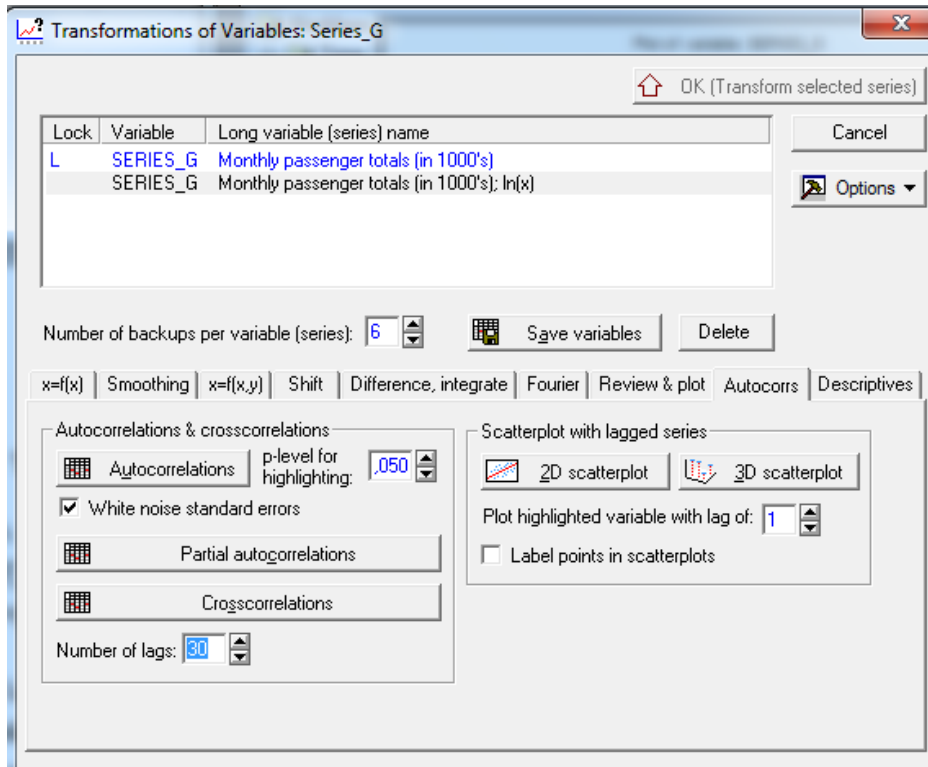
იდენტიფიკაცია საკმაოდ უხეში პროცედურაა, რომლის საშუალებითაც მიიღება მოდელის რიგის მიახლოებითი მნიშვნელობა. იდენტიფიკაციის ძირითად კრიტერიუმად მიიღება მწკრივის ავტოკორელაციური და კერძო ავტოკორელაციური ფუნქციების ქცევა. სინამდვილეში ეს ფუნქციები უცნობია და ამიტომ ჩვენ საქმე გვაქვს ამ ფუნქციების შეფასებებთან, რომლებსაც შერჩევითი ფუნქციები ეწოდებათ.

კერძო ავტოკორელაციური ფუნქცია წარმოადგენს ჩვეულებრივი ავტოკორელაციური ფუნქციის სიღრმისეულ ცოდნას. მოცემული ლაგის (ძვრის) კერძო ავტოკორელაციური ფუნქცია ანალოგიურია ჩვეულებრივი ავტოკორელაციური ფუნქციისა, იმ განსხვავებით, რომ მისი გამოთვლისას გამოირიცხება მცირე ძვრების ავტოკორელაციური ფუნქციის ზეგავლენა. როცა ძვრა 1-ის ტოლია, მაშინ კერძო ავტოკორელაციური ფუნქცია ჩვეულებრივი ავტოკორელაციური ფუნქციის ტოლია.

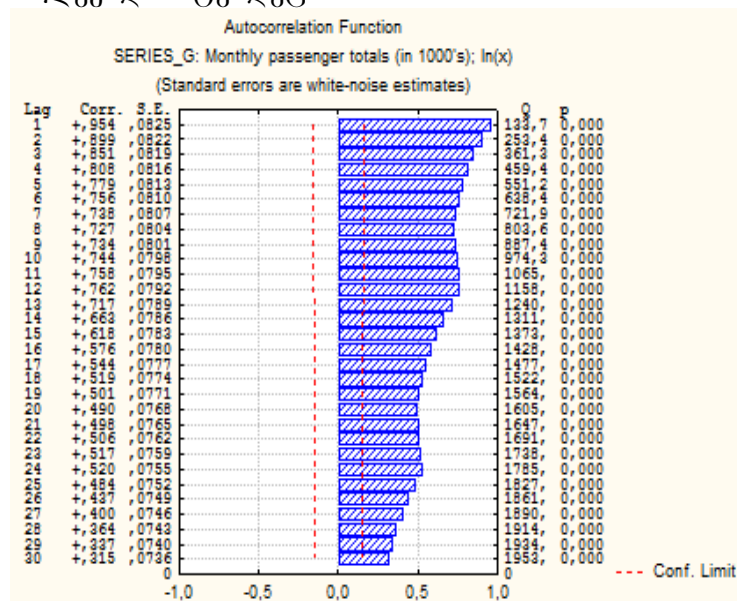
დავუშვათ  $d$  მოდელის უცნობი რიგია, რომლის შეფასებაც გვინდა. დასაწყისში ხდება დროითი მწკრივის ვიზუალიზაცია და ვადგენთ, მწკრივი სტაციონარულია თუ არა. მწკრივის არასტაციონარობა ხშირად ადვილად ჩანს მწკრივის გრაფიკულ გამოსახულებაზე, მაგალითად, თუ მწკრივს გააჩნია ცხადად გამოკვეთილი ტრენდი. განსაკუთრებით ადვილია ვიზუალურად განისაზღვროს მონოტონური ტრენდი: ლოგარითმული, ექსპონენციალური, წრფივი, პარაბოლა და სხვ. (ჩვენი მაგალითის შემთხვევები **Series G** მწკრივს გააჩნია ზრდის ცხადი ტენდენცია. ე.ი მონოტონური ტრენდი). ამრიგად, თუ მწკრივს გააჩნია ცხადად შესამჩნევი ტრენდი, მაშინ ასეთი მწკრივი არასტაციონარულია. თუ ტრენდი ცხადად არ ჩანს და არ გაგვაჩნია რაიმე მოსაზრება, რომელიც მიუთითებს დროითი მწკრივის არასტაციონარობაზე, მაშინ უნდა განვიხილოთ ავტოკორელაციური ფუნქცია, კერძოდ შერჩევითი ავტოკორელაციური ფუნქცია.

თუ ავტოკორელაციურ ფუნქციას არ გაჩნია რხევის მიღევის (ჩაქრობის) ტენდენცია, მაშინ შეიძლება ითქვას, რომ მწკრივი არასტაციონარულია.

ამრიგად, არასტაციონარობის კრიტერიუმი გამოიხატება შერჩევითი ავტოკორელაციური ფუნქციის რხევის მიღების ტენდენციაში. ამისათვის **Autocorr** ჩანართში **Number of lag** ოპციის საშუალებით, სადაც ლაგების რიცხვი უნდა გაიზარდოს 30-მდე, **click Autocorrelations** ღილაკზე.



ეკრანზე გამოდის ავტოკორელაციური ფუნქციის გრაფიკი, სადაც ნათლად ჩანს, რომ ავტოკორელაციურ ფუნქციას გააჩნია რხევის სუსტი არამონოტონური მიღევის ტენდენცია.

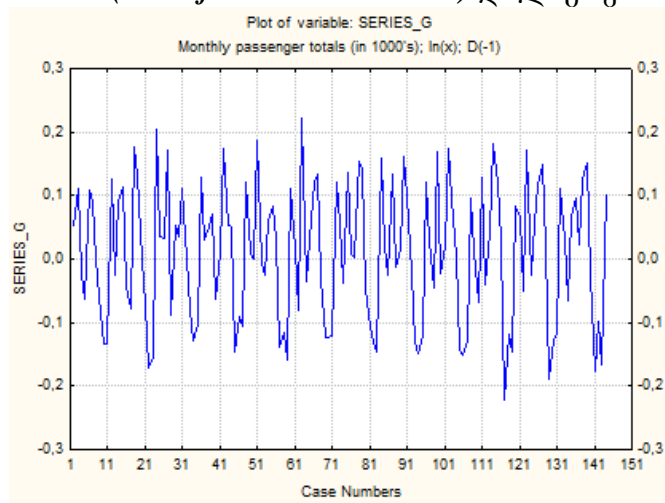


ამრიგად, ავტოკორელაციური ფუნქციის დათვალიერების შედეგად მივიღივართ დასკვნამდე: მწკრივი სტაციონარულია ან არასტაციონალურია.

როცა მწკრივი სტაციონარულია, მაშინ  $d = 0$  და გადავდივართ სხვა  $p$  და  $q$  პარამეტრების განსაზღვრაზე.

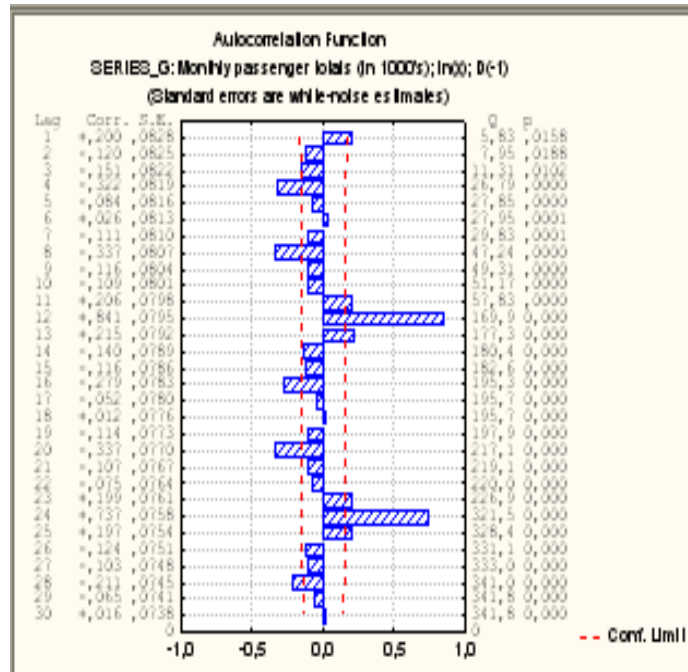
თუ მწკრივი არასტაციონარულია, მაშინ უნდა განვიხილოთ მოცემული მწკრივის პირველი რიგის სხვაობა. თუ აღმოჩნდება, რომ პირველი რიგის სხვაობის მწკრივი არასტაციონარულია, მაშინ კვლავ იღებენ პირველი რიგის სხვაობას და ხდება მწკრივის სტაციონარობის შემოწმება, რადგან პირველი რიგის სხვაობა გამოიყენება ორჯერ, ეს იმას ნიშნავს, რომ მოცემული მწკრივის მიმართ გამოყენებულია მეორე რიგის სხვაობითი ოპერატორი. პრაქტიკაში იშვიათად გამოიყენება ორზე მეტი სხვაობითი პროცედურა. ე-ი პროცედურა მთავრდება  $k$ -ურ ბიჯზე, თუ გარდაქმნილი მწკრივი სტაციონარული გახდა, მაშინ  $d = k$ . (**Series G** მწკრივისათვის ავიღოთ  $k = 1$  ე-ი  $d = 1$ ).

**Differencing, Integral** ჩანართის **Differencing** ( $x=x-x(\text{lag})$ ) ოპციაში მიუთითოთ  $\text{lag} = 1$  და შემდეგ **click OK (Transform selected series)** ღილაკზე.

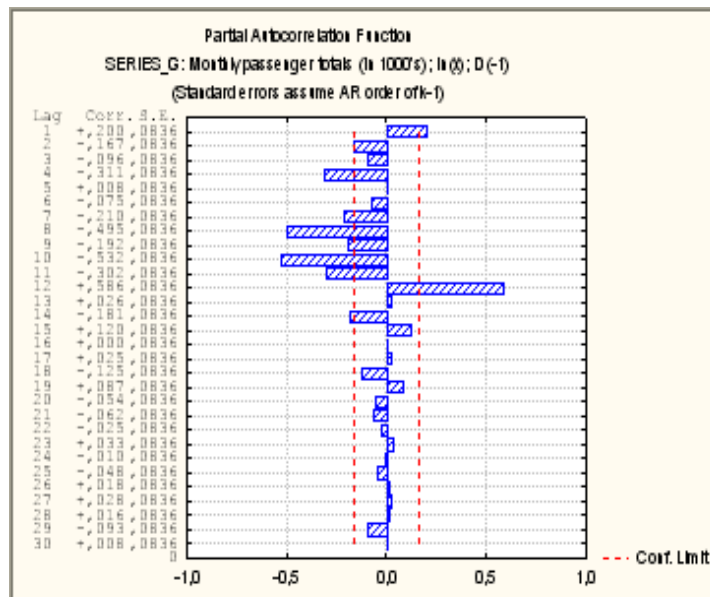


ეკრანზე გამოსულ გრაფიკზე ჩანს, რომ მწკრივი გახდა სტაციონარული, რადგან ტრენდი არ შეიმჩნევა.

ავტოკორელაციური ფუნქციის განსაზღვრისათვის **Autocorre** ჩანართში **click Autocorrelations** ღილაკზე. ეკრანზე გამოდის ავტოკორელაციური ფუნქციის გრაფიკი



რაც შეეხება ავტოკორელაციურ ფუნქციას, მას გააჩნია  $lag = 1$ -ზე უმნიშვნელო ამოგდება და ჩაქრობის სუსტი ტენდენცია, თუ არ ჩავთვლით იმ პიკებს, რომლებიც წარმოქმნილია 12 თვიანი სეზონური ციკლით. კერძო ავტოკორელაციის ფუნქციის ეკრანზე გამოსატანად, **click Partial autocorrelations** ღილაკზე:



როგორც გრაფიკიდან ჩანს კერძო ავტოკორელაციის ფუნქციის ცვალებადობა ექსპონენციალურად უახლოვდება ნულს. ამრიგად, ორმაგი გარდაქმნის შედეგად მივიღეთ სტაციონარული მწკრივი.

$p$  და  $q$  პარამეტრების შერჩევისათვის განიხილება შერჩევითი ავტოკორელაციური და კერძო კორელაციური ფუნქციების ქცევა.

დავუშვათ გვაქვს  $p$  რიგის ავტორეგრესიის მოდელი. მაშინ ამ პროცესის კერძო ავტოკორელაციური ფუნქცია წყდება  $p$  ლაგზე, ხოლო

ავტოკორელაციური ფუნქცია მდორედ მცირდება. დავუშვათ გვაქვს  $p$  რიგის მცოცავის საშუალო პროცესი, მაშინ კერძო ავტოკორელაციური ფუნქცია მდორე ტენდენციით მცირდება, ხოლო ავტოკორელაციური ფუნქცია  $p$  ლაგზე წყდება. თუ მოდელის ორივე პარამეტრი ნულისაგან განსხვავებულნი არიან, მაშინ ავტოკორელაციური ფუნქცია წარმოდგენილია როგორც ექსპონენტის და მილევადი სინუსოიდის ჯამი.

პრაქტიკა გვიჩვენებს, რომ დროითი მწკრივების უმეტესობა, რომელებიც აღწერილნი არიან ავტორეგრესიის და მცოცავი საშუალოს შერეული მოდელით, გარკვეული სიზუსტით შეიძლება მივაკუთნოთ რომელიმე ქვემოთ მოყვანილ კლასს:

- ავტორეგრესიის მოდელი ერთი პარამეტრით:  $p = 1, q = 0$
- ავტორეგრესიის მოდელი ორი პარამეტრით:  $p = 2, q = 0$
- მცოცავი საშუალოს მოდელი ერთი პარამეტრით:  $p = 0, q = 1$
- მცოცავი საშუალოს მოდელი ორი პარამეტრით:  $p = 0, q = 2$
- ავტორეგრესიის მოდელი ერთი პარამეტრით და მცოცავი საშუალოს მოდელი ერთი პარამეტრით:  $p = q = 1$ .

მოდელის ამა თუ იმ კლასში მოხვედრის პრაქტიკული კრიტერიუმები ეფუძნება ავტოკორელაციურ და კერძო ავტოკორელაციური ფუნქციების გამოყენებას.

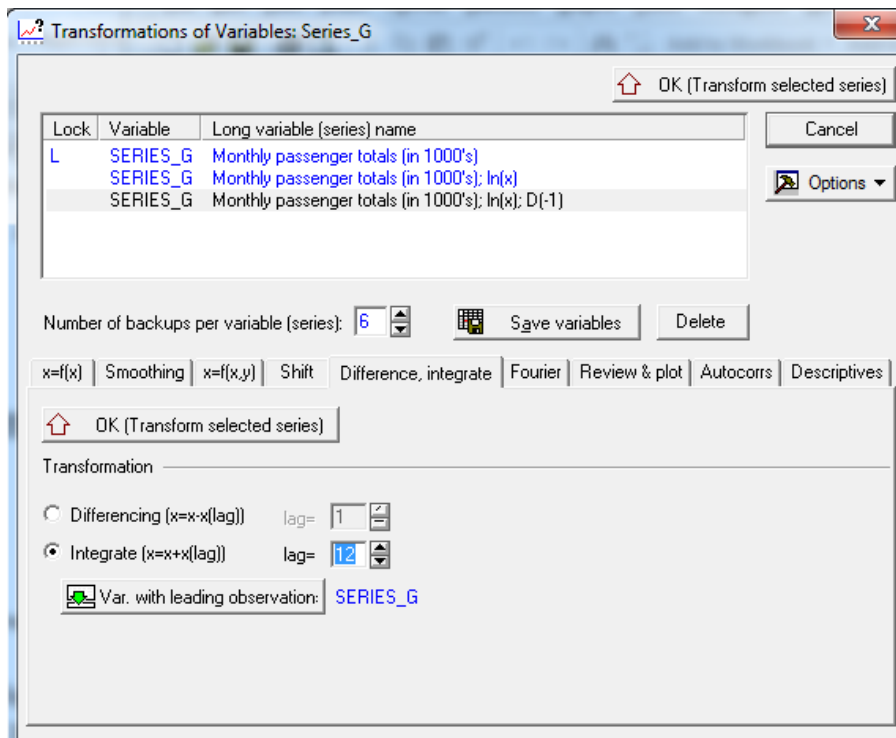
- **ავტორეგრესიის ერთი პარამეტრი:** ავტოკორელაციური ფუნქციის ჩაქრობა ხდება ექსპონენციალური კანონით, კერძო ავტოკორელაციურ ფუნქციას  $lag=1$ -ზე გააჩნია ამოვარდნა (სხვა ძვრებისათვის კორელაცია არ არსებობს);
- **ავტორეგრესიის ორი პარამეტრი:** ავტოკორელაციურ ფუნქციას გააჩნია სინუსოიდალური ან ექსპონენციალური მილევადი რხევის ფორმა; კერძო ავტოკორელაციურ ფუნქციას გააჩნია ამოვარდნები მხოლოდ  $lag = 1$  და  $lag = 2$  (სხვა ძვრებისათვის კორელაცია არ არსებობს);
- **მცოცავი საშუალოს ერთი პარამეტრი:** ავტოკორელაციურ ფუნქციას  $lag = 1$ -ზე გააჩნია ამოვარდნა (სხვა ძვრებისათვის კორელაცია არ არსებობს); კერძო ავტოკორელაციურ ფუნქციას გააჩნია ექსპონენციალურად მილევადი ფორმა – ან მონოტონურად, ან ოსცილაციის (ე.ი. ნიშანცვლადი) სახით;
- **მცოცავი საშუალოს ორი პარამეტრი:** ავტოკორელაციურ ფუნქციას გააჩნია 1 და 2 ძვრის დროს ამოვარდნები (სხვა ძვრებისათვის კორელაცია არ არსებობს); კერძო ავტოკორელაციურ ფუნქციას გააჩნია სინუსოიდალური ტალღის ფორმა ან ექსპონენციალურად მილევადია;
- **ავტორეგრესიის ერთი პარამეტრი და მცოცავი საშუალოს ერთი პარამეტრი:** ავტოკორელაციური ფუნქცია ექსპონენციალურად მილევადია, დაწყებული პირველი ძვრიდან (პირველი მნიშვნელობა ნული არ არის), მილევადობა შეიძლება იყოს მონოტონური და რხევითი; კერძო ავტოკორელაციურ ფუნქციაში დომინირებს ექსპონენციალური მილევადის წევრი – ან მონოტონური, ან ოსცილირებულ (პირველი მნიშვნელობა ნული არ არის).

როგორც პრაქტიკა გვიჩვენებს, იდენტიფიკაციის ეტაპზე მიზანშეწონილია განისაზღვროს რამოდენიმე შესაფერისი მოდელი და შემდეგ, როცა შეფასებული იქნება მათი პარამეტრები, ნაშთების გამოკვლევა, მოდელის აღეკვებურობა, უკვე შეიძლება საუკეთესო მოდელის შერჩევა.

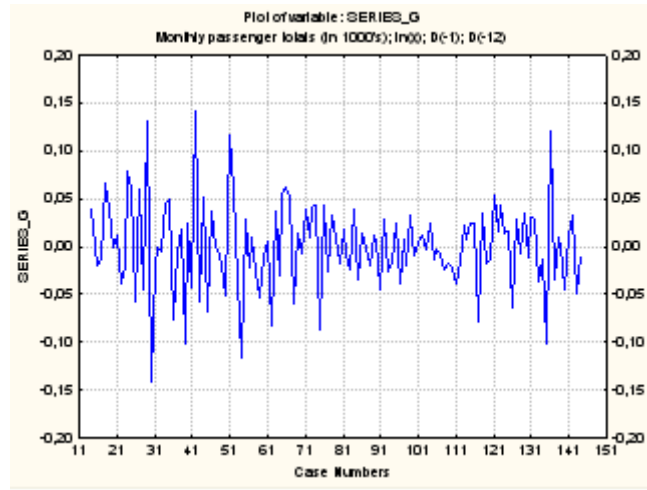
ავტოკორელაციური და კერძო ავტოკორელაციური ფუნქციების ანალიზის შედეგად შეიძლება დავასკვნათ, რომ **Series G ln(x)D(1)** მწკრივისათვის მისაღებია მოდელი – მცოცავი საშუალოს ერთი პარამეტრი  $p = 0, q = 1$ . თუ გავითვალისწინებთ, რომ  $d = 1$ , მაშინ გვექნება არასეზონური მოდელი (0,1,1).

რადგან მწკრივს გააჩნია ცხადად წარმოდგენილი სეზონური მდგენელი 12 თვის პერიოდით, საჭიროა მოდელში შეტანილი იყოს სეზონური კორექტირება. ამ შემთხვევაში მოდელი წარმოდგენილია **ARIMA (p,d,q) (Ps,Ds,Qs)** სახით, სადაც დამატებულია სეზონური პარამეტრები: **Ps**–ავტორეგრესიის სეზონური პარამეტრი, **Ds**–სეზონური სხვაობა, **Qs**–მცოცავი საშუალოს სეზონური პარამეტრი. ავტოკორელაციური და კერძო ავტოკორელაციური ფუნქციების ქცევა სეზონურ ძვრაზე ხდება იგივე სტანდარტული მეთოდებით ანუ ზემოთ განხილული პრაქტიკული კრიტერიუმები ძალაში რჩება.

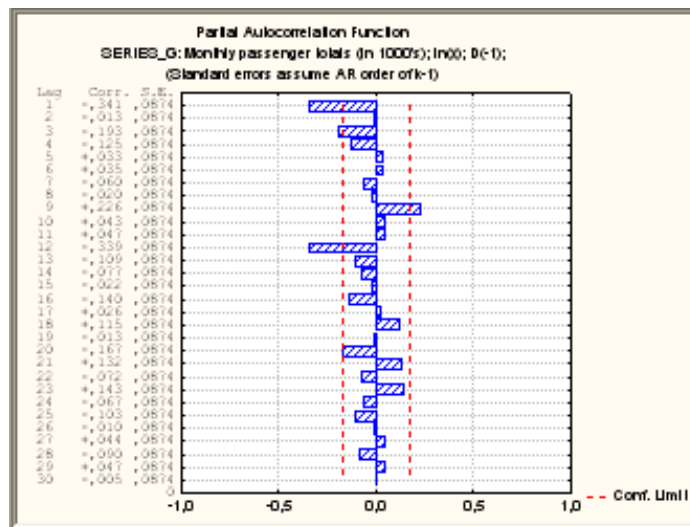
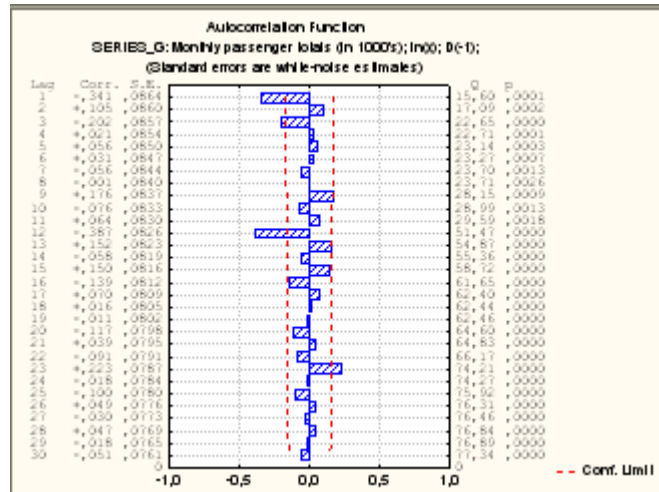
იმისათვის, რომ მხედველობაში მივიღოთ სეზონური რხევები 12 თვის პერიოდით. საჭიროა ავიღოთ **Series G ln(x)D(-1)**. ამისათვის დავბრუნდეთ **Transformation of Variables** ფანჯარაში და მოვნიშოთ **Series G ln(x)D(-1).Differencing, Integrate** ჩანართში ჩაერთოთ **Differencing (x=x-x(lag))** ოპცია და მიუთითოთ  $lag = 12$ .



შემდეგ **click OK(Transform selected series)** ღილაკზე. ეკრანზე გამოდის გრაფიკი.



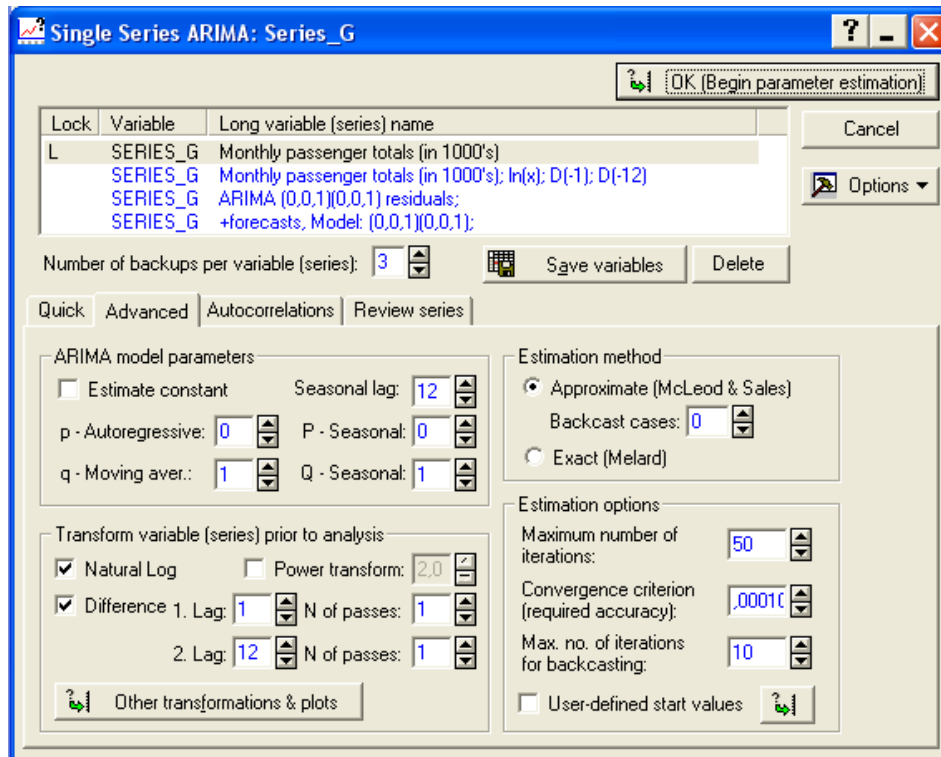
ეკრანზე გამოვიტანთ დროითი მწკრივის ავტოკორელაციური და კერძო კორელაციური ფუნქციების გრაფიკები.





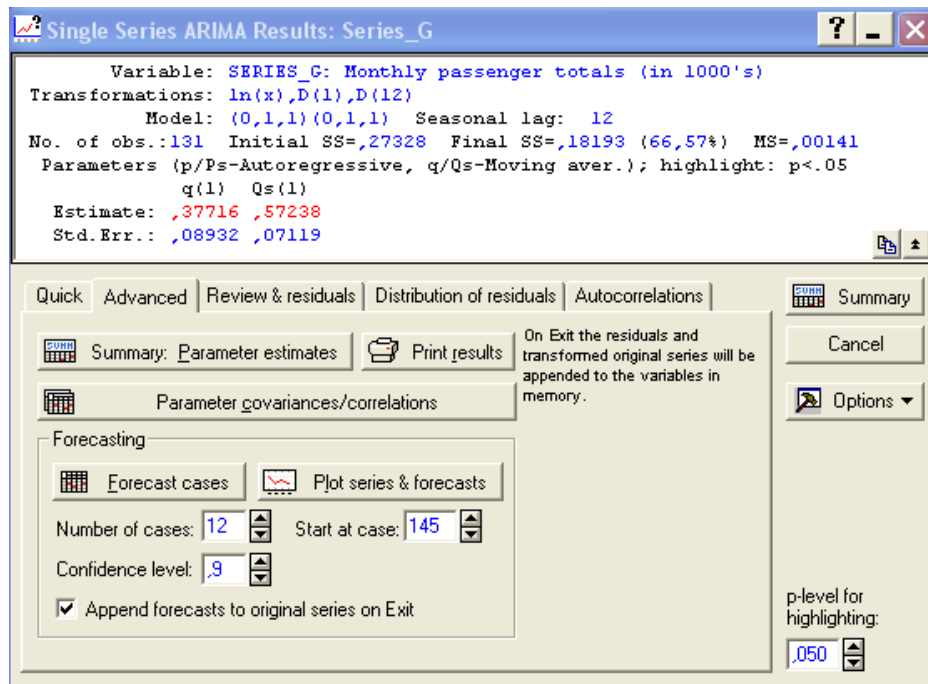
ამ გრაფიკებიდან ჩანს, რომ მწკრივი სტაციონარულია, ავტოკორელაციური ფუნქცია ექსპონენციალურად მიღევადია, ხოლო კერძო ავტოკორელაციური ფუნქციაც ექსპონენციალურად მიღევადია და გააჩნია ამოვარდნა 1 ძვრახე. აქედან გამომდინარე, სეზონური პარამეტრი  $P_s = 0$ ,  $D_s = 1$ ,  $Q_s = 1$ . და ჩვენ გვაქვს საქმე **ARIMA** (0,1,1) მოდელთან. ამრიგად ავტორეგრესიის სრულ მოდელს აქვს **ARIMA** (0,1,1), (0,1,1) სახე.

ვბრუნდებით **Single Series ARIMA** ფანჯარაში. ამისათვის **Transformation of Variables** ფანჯარაში *click Cancel* ღილაკზე. კრანზე გამოდის საწყისი **Series G** ცვლადი. **Arima model parameters** ჩაჩოში დავაყენოთ შესაბამისი პარამეტრების მნიშვნელობები. **p-Autoregressive-** 0; **q-Moving aver-** 1, **P-seasonal-** 0, **Seasonal lag-** 12, **Q-seasonal-** 1.



**Estimation method** ჩაჩოში ჩავრთოთ **Approximate** ოპცია. შემდეგ **OK**.

– გაიხსნება **Single Series ARIMA Results** (ერთგანზომილებიანი **ARIMA** შედეგები)-ის ფანჯარა.



click **Summary: parameter estimates** (პარამეტრების შეფასება) ღილაკზე. ეკრანზე გამოდის პარამეტრების შეფასების ცხრილი.

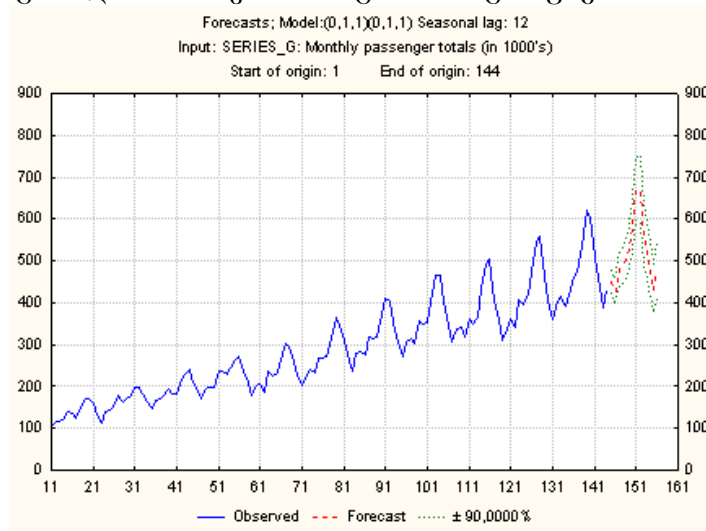
Input: SERIES_G: Monthly passenger totals (in 1000's) (Series_G)						
Transformations: ln(x),D(1),D(12)						
Model:(0,1,1)(0,1,1) Seasonal lag: 12 MS Residual= ,00141						
Paramet.	Param.	Asympt. Std.Err.	Asympt. t( 129)	p	Lower 95% Conf	Upper 95% Conf
q(1)	0,377162	0,089318	4,222697	0,000045	0,200445	0,553880
Qs(1)	0,572379	0,071189	8,040233	0,000000	0,431529	0,713229

ცხრილიდან ჩანს, რომ ორივე პარამეტრის შეფასება მეკეთრად სარწმუნოა ( $p$  გაცილებით ნაკლებია 0.05-ზე). გაჩუმების პრინციპით პროგრამა ერთი სრული სეზონის ციკლისათვის გამოთვლის პროგნოზის მნიშვნელობებს, დაწყებული მწკრივის ბოლო მნიშვნელობიდან.

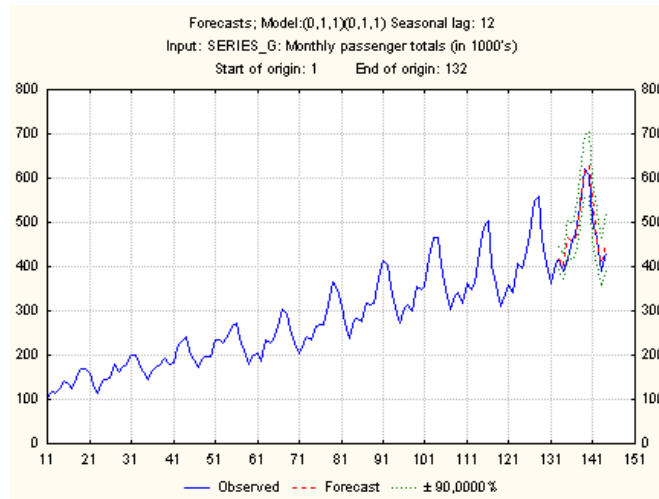
– click **Advanced** ანართის **forecast cases** (პროგნოზი)-ის ღილაკზე. ეკრანზე გამოდის შედეგების ცხრილი, სადაც წარმოდგენილია პროგნოზი და მათი ნდობის ინტერვალები, დაწყებული მწკრივის ბოლო მნიშვნელობიდან (ჩვენ შემთხვევაში 145-დან). ე.ი. ხდება 12 მნიშვნელობის პროგნოზი.

Forecasts; Model:(0,1,1)(0,1,1) Seasonal lag: 12 (Series_G) Input: SERIES_G: Monthly passenger totals (in 1000's) Start of origin: 1 End of origin: 144			
CaseNo.	Forecast	Lower 90,0000%	Upper 90,0000%
145	450,1171	422,9655	479,0117
146	425,6620	395,5777	458,0341
147	479,5240	441,3696	520,9766
148	492,0412	449,0088	539,1979
149	508,5479	460,4357	561,6874
150	583,0166	524,0264	648,6473
151	669,1520	597,3584	749,5742
152	666,4152	591,1003	751,3264
153	557,9980	491,9233	632,9478
154	496,7552	435,3899	566,7696
155	429,6965	374,5207	493,0009
156	477,1535	413,6613	550,3910

– click **Plot series & forecasts** (მწკრივის და პროგნოზი გრაფიკები) ღილაკზე. ეკრანზე გამოდის 12 თვის პროგნოზის გრაფიკი.



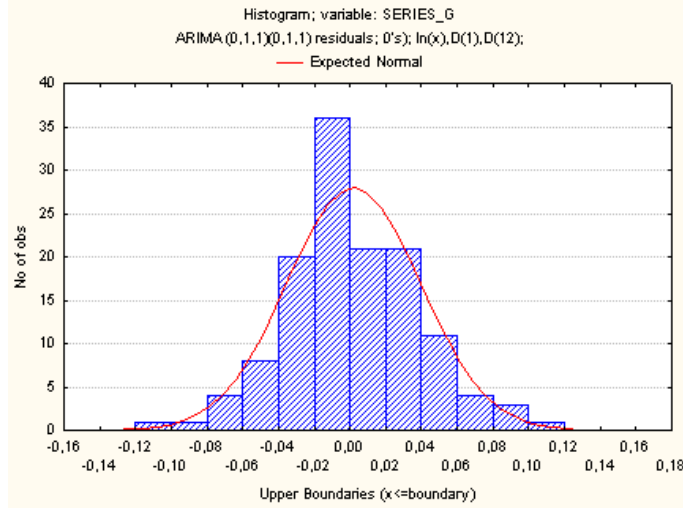
პროგნოზირების მოდელის შემოწმებისათვის. **Start at case** (დაიწყოს დაკვირვებიდან) ველში მიუუთითოთ, მაგ. 133 და კვლავ **click Plot series & forecasts** ღილაკზე.



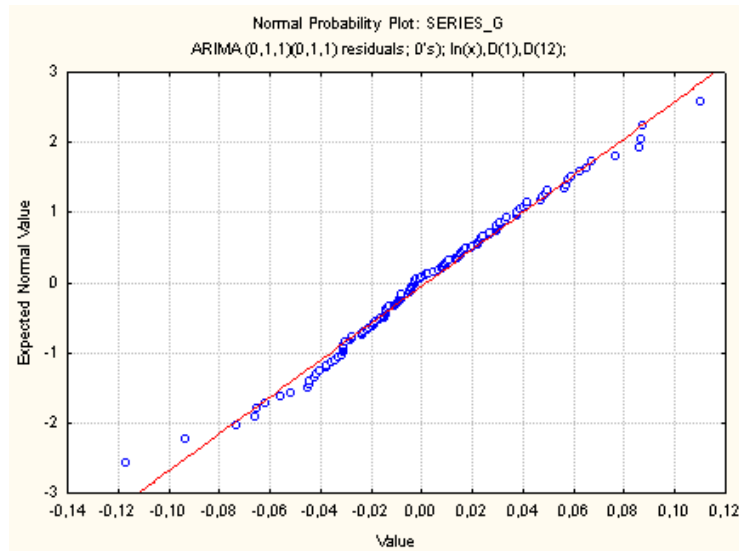
ეკრანზე გამოსულ გრაფიკში ნათლად ჩანს, რომ პროგნოზის მრუდი პრაქტიკულად არ განსხვავდება საწყისი მწკრივის გრაფიკიდან და მწკრივის ყველა მნიშვნელობა ხვდება ნდობის ინტერვალში.

მოდელის ადეკვატურობის შესამოწმებლად ჩავატაროთ ნაშთების ანალიზი ამისათვის:

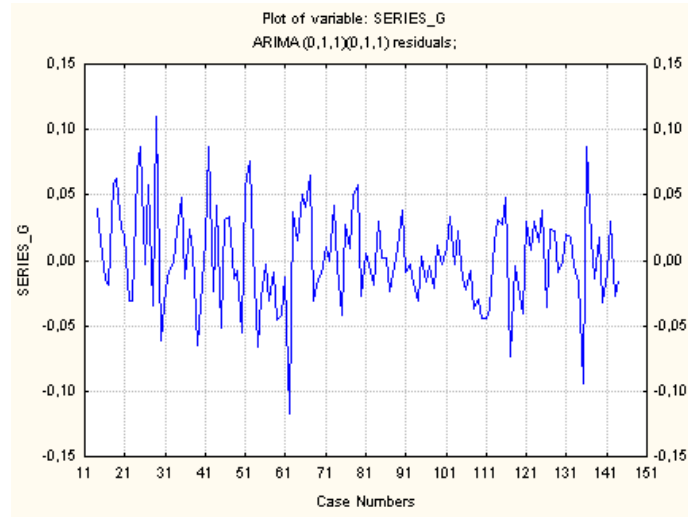
– **Distribution of residuals** ჩანართში *click Histogram* ღილაკზე. ეკრანზე გამოსულ გრაფიკზე ჩანს, რომ ნარჩენების ემპირიული განაწილების სიმკვრივის აპროქსიმაცია ნორმალური განაწილების კანონით წარმატებულია, რაც იმას მიგვანიშნებს, რომ პროგნოზის მოდელი ადეკვატურია.



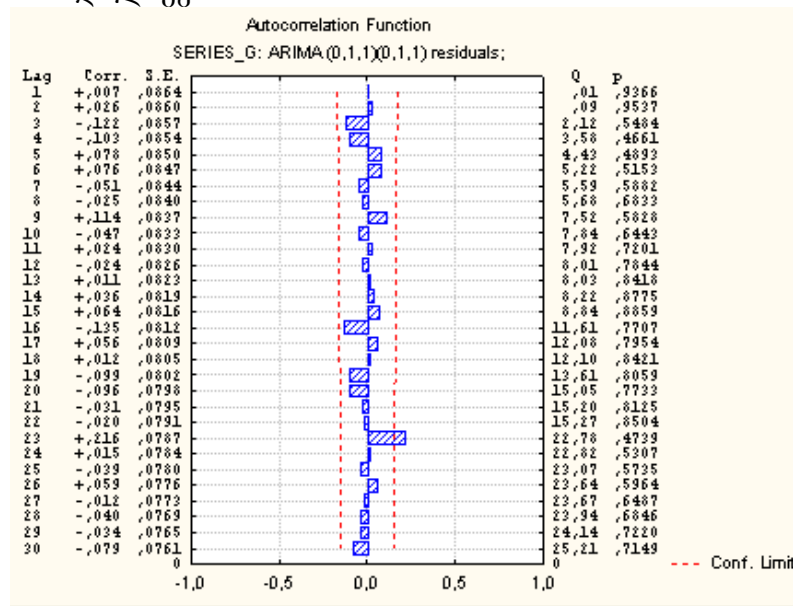
ნარჩენების ნორმალური განაწილების დასადგენად გამოიყენება აგრეთვე **Normal probability plot** (ნორმალური ალბათური გრაფიკები). თუ დაკვირვები მკვეთრად განსხვავდება ნორმალური განაწილებისაგან, მაშინ მათი სტანტარტიზირებული მნიშვნელობა მკვეთრად გადახრილი იქნებიან სწორი ხაზიდან.

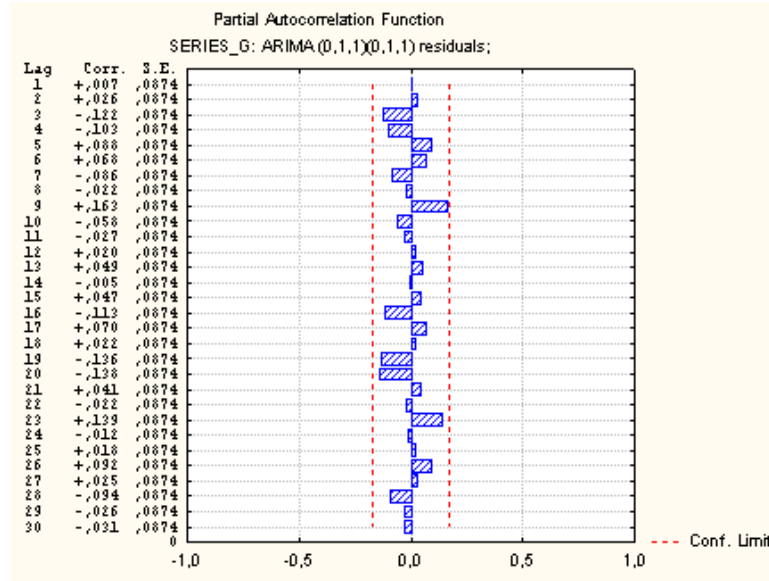


ნარჩენების გრაფიკის ასაგებად გავხსნათ **Review & residuals** ჩანართი, სადაც **Review & plot variables** (დავათვალიეროთ ცვლადები და ავაგოთ გრაფიკები) ოპციაში *click Plot (3)* (ნარჩენების გრაფიკი) ღილაკზე. პროგრამა აგებს ნარჩენების გრაფიკს, საიდანაც ჩანს, რომ ნარჩენებს გააჩნიათ დაახლოებით თანაბარი ვარიაცია და არ იკვეთება ტრენდი. ამრიგად, შეგვიძლია ჩავთვალოთ, რომ ნარჩენები ნორმალურად არიან განაწილებულნი.



კარგად შერჩეულ მოდელში ნარჩენები ძალზე წააგავენ თეთრ ხმაურს. მასში არ იქნება პერიოდული რხევები, სისტემური წანაცვლება, მათ შორის არ იქნება ძლიერი კავშირები. *Autocorrelations* ჩანართში *click autocorrelations* და *Partial autocorrelatins* ღილაკები.





ეკრანზე გამოსულ გრაფიკებში ჩანს, რომ ნარჩენები პრაქტიკულად წარმოადგენენ თეთრ ხმაურს.

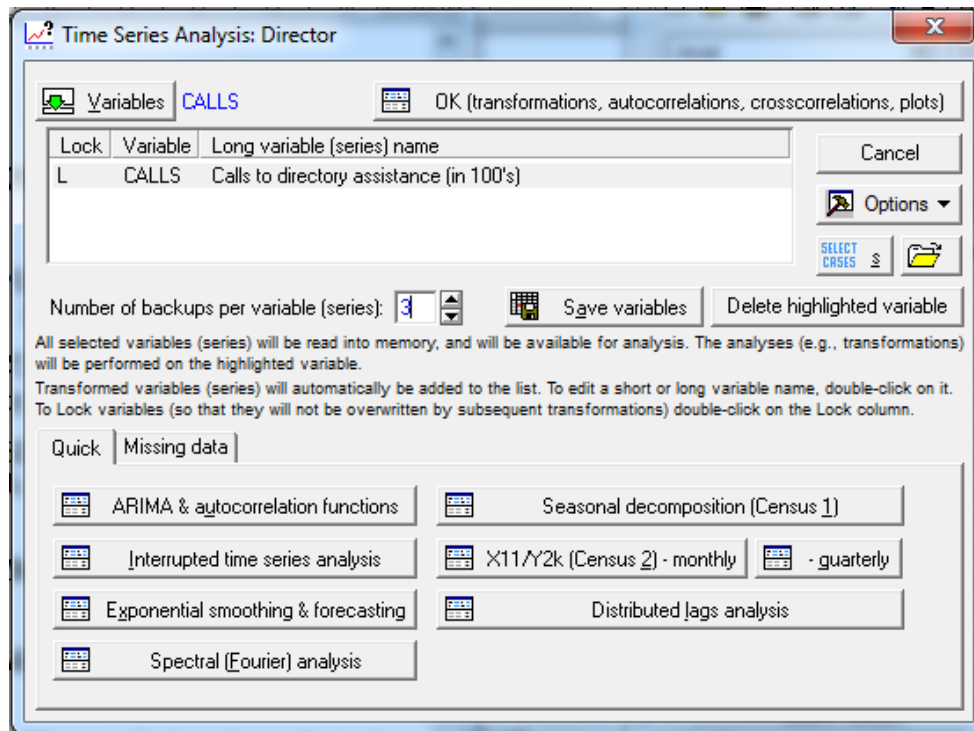
## 2. ინტერპენციის ARIMA მოდელი

ინტერპენციის მოდელი გამოიყენება იმ შემთხვევებში, როცა დროითი მწკრივის ქცევა ანუ ტრენდი გარე ფაქტორების ზემოქმედებით გარკვეული მომენტიდან მკვეთრად იცვლება. გარე ზეგავლენა შეიძლება იყოს მოკლევადიანი (იმპულსური) და გრძელვადიანი (მდგრადი). ზემოქმედების მომენტში მწკრივის ტრანსფორმაცია მკვეთრად იცვლება, მაგრამ შემდგომში კვლავ აღიწერება ARIMA მოდელით. დროითი მწკრივის მკვეთრი ცვლილების მომენტს ინტერპენცია ეწოდება.

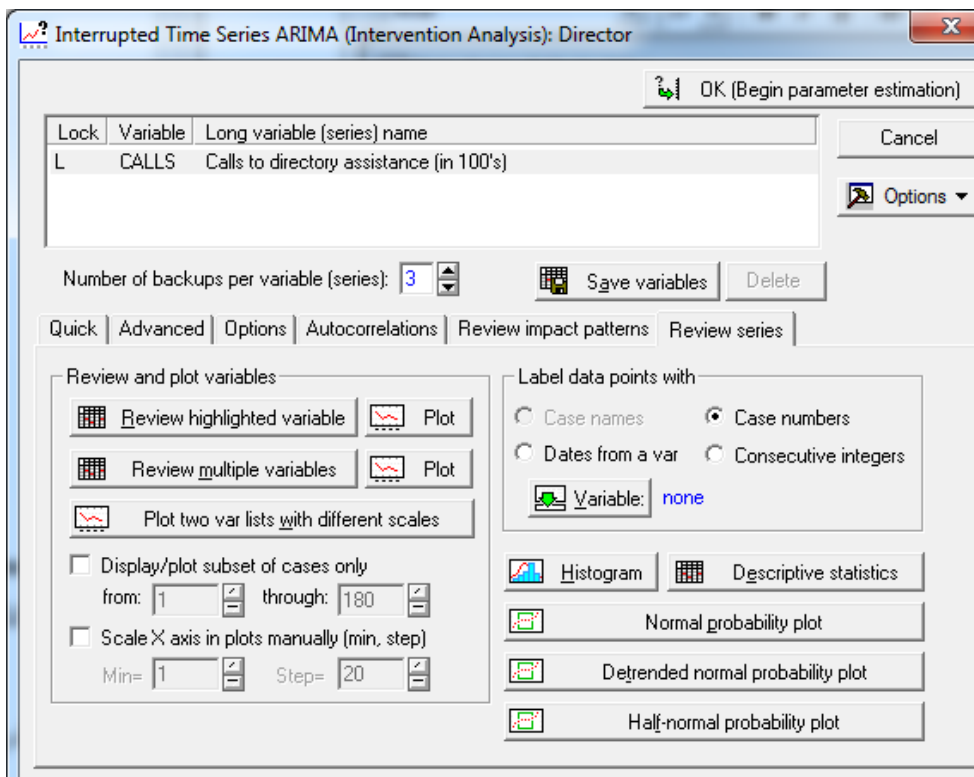
განვიხილოთ *Examles* ბიბლიოთეკაში არსებული *Dataset* მონაცემთა ბაზაში არსებული *Director* ფაილი.

*Statistica* მენიუში მოვნიშნოთ *Advanced Linear/Nonlinear Models* და *click Time Series/Forecasting* ბრძანებაზე

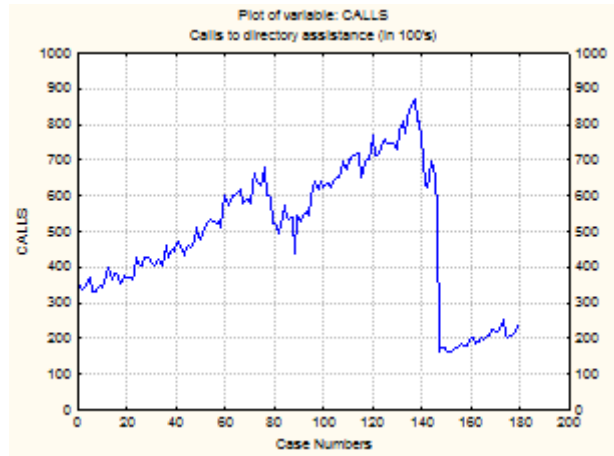
ეკრანზე გამოდის *Time Series Analysis* (დროითი მწკრივების ანალიზი)-ის სასტარტო ფანჯარა. *click Variables* ღილაკზე და ეკრანზე გამოსულ ფანჯარაში მოვნიშნოთ *Calls* ცვლადი. შემდეგ *OK*.



ეკრანზე გამოდის *Time Series Analysis* ფანჯარა. *click Interrupted time series analysis* (ინტერვენციული დროითი მწკრივების ანალიზი)-ის ღილაკზე. ეკრანზე გამოდის *Interrupted Time Series ARIMA* სასტარტო ფანჯარა.



– *Review series* ჩანართში *click Plot* ღილაკზე. ეკრანზე გამოდის ინტერვენციული დროითი მწკრივის გრაფიკი.



– *Advanced* ჩანართში გაიხსნება სასტატო პროცედურის ფანჯარა.

სადაც *ARIMA* ფანჯრიდან განსხვავდება მარჯვენა *Specify times and type of interventions* (დავაყენოთ ინტერვენციის დრო და ტიპი) ჩარჩოთი. ამ ჩარჩოს ოპციები შემდეგია:

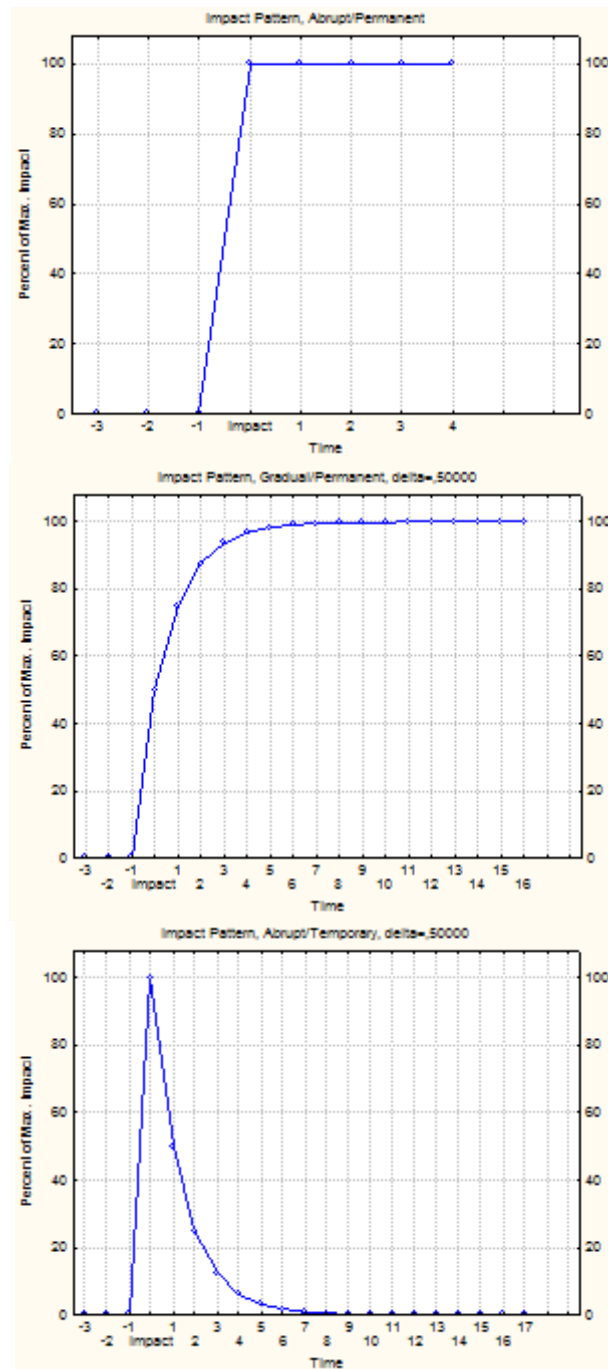
- **Intervention** – ინტერვენციის ნომერი (შეიძლება 6 ინტერვენცია);
- **At case number** – ინტერვენციის დაწყების შემთხვევის ნომერი;
- **Type of intervention** – ინტერვენციის ტიპი.

პროცედურა ითვალისწინებს ინტერვენციის სამ ტიპს. იმისათვის, რომ დავათვალიეროთ მათი გრაფიკები შევარჩიოთ *Review impact patterns* ჩანართი. სადაც *Type of intervention* გაშლად ველში წარმოდგენილია:

- **Abrupt, Permanent** (ნახტომისებური, მდგრადი);
- **Gradual Permanent** (თანდათანობით, მდგრადი);
- **Abrupt, Temporary** (ნახტომისებური, დროებითი).



შეგარჩიოთ თითოეული მათგანი და *click Review type of impact pattern* ღილაკზე.



სამი ტიპიდან უნდა შეირჩეს ჩვენთვის ყველაზე უფრო მიახლოებითი ინტერვენცია. ამის შემდეგ შეგვიძლია გადავიდეთ **ARIMA** მოდელის პარამეტრების შერჩევაზე ისე, როგორც აღწერილი იყო **ARIMA** ზოგადი მოდელის დროს.

## პრაქტიკული სამუშაო 14

### პროგნოზირება ექსპონენციალური გაბლუვებით

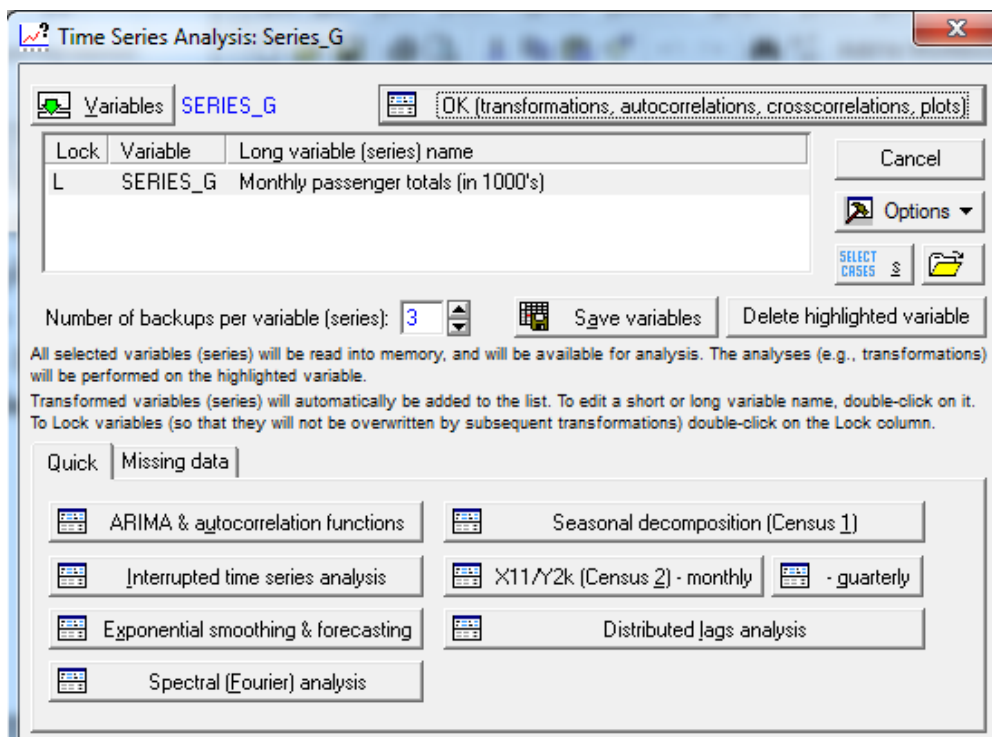
ექსპონენციალური გაგლუვების მეთოდის არსი იმაში მდგომარეობს იმაში, რომ დროითი მწკრივის გაგლუვება ხდება აწონილი მცოცავი საშუალოთი, რომლის კოეფიციენტებს გააჩნიათ სხვადასხვა წონა და ის დამოკიდებულია ექსპონენციალურ კანონზე. აქედან გამომდინარე, დროითი მწკრივის ის მნიშვნელობები, რომლებიც ახლოს არიან  $t$  პროგნოზირებად დროსთან გააჩნიათ დიდი წონითი კოეფიციენტები, ხოლო რაც უფრო დაშორებულნი არიან  $t$  პროგნოზირების დროსთან – მით უფრო ნაკლები წონები.

მარტივი ექსპონენციალური გაგლუვების ფორმულას აქვს შემდეგი სახე

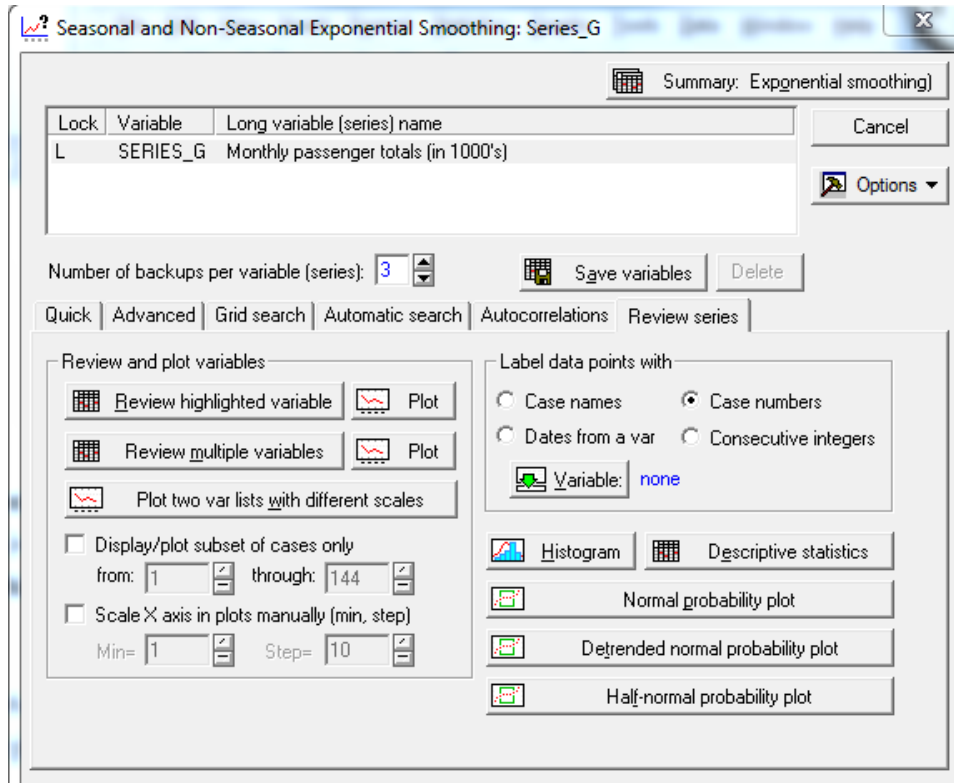
$$S_t = \alpha X_t + (1 - \alpha) S_{t-1} ,$$

სადაც  $X_t$  -საწყისი დროითი მწკრივია,  $S_t$  – გაგლუვებული დროითი მწკრივი. გასაგებია, რომ გაგლუვების შედეგი დამოკიდებულია  $\alpha$  პარამეტრზე. თუ  $\alpha = 1$  მაშინ წინა დაკვირვებები იგნორირებულნი არიან. თუ  $\alpha = 0$ , მაშინ იგნორირებულნი არიან მიმდინარე დაკვირვებები. აქედან გამომდინარე,  $\alpha$  სიდიდე იცვლება  $0 < \alpha < 1$ . მომხმარებელს შეუძლია დააფიქსიროს გაგლუვების საწყისი პარამეტრები, ტრენდის საწყისი მნიშვნელობა და თუ საჭიროა სეზონური ფაქტორები. მოდულის აღწერისათვის გამოვიყენოთ **Series G** დროითი მწკრივი.

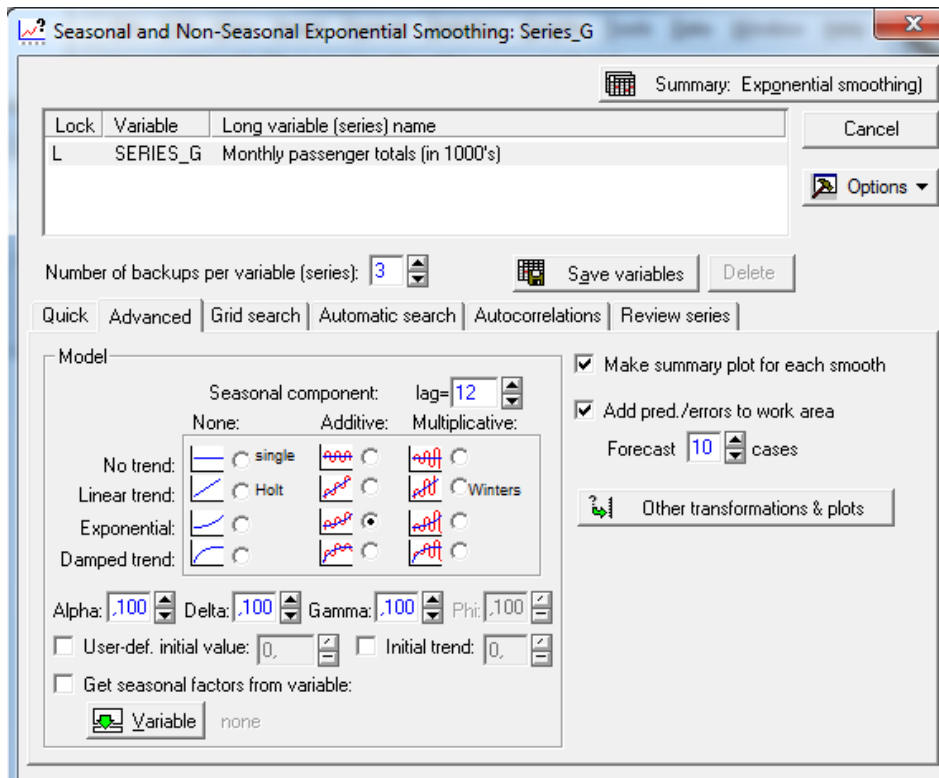
**Statistica** მენიუში მოვნიშნოთ **Advanced Linear/Nonlinear Models** და **click Time Series/Forecasting** ბრძანებაზე



– ეკრანზე გახსნილ *Time Series Analysis* ფანჯარაში *click Exponential smoothing & forecasting* (ექსპონენციალური გაგლუვება და პროგნოზირება)-ის ღილაკზე. ეკრანზე გამოდის *Seasonal and Non-Seasonal Exponential smoothing* (სეზონური და არასეზონური ექსპონენციალური გაგლუვება)-ის ფანჯარა.



ჩაერთოდ *Advanced* ჩანართი:



ექსპონენციალური გაგლუვების მოდელის განსაზღვრისათვის საჭიროა სეზონური კომპონენტის ტრენდის და გაგლუვების პარამეტრების დაფიქსირება. ეს შესაძლებელია შემდეგი ოპციებით:

- **Seasonal component** (სეზონური კომპონენტი);
- **None** (არა);
- **Additive** (ადიტიური);
- **Multiplicative** (მულტიპლიკატიური);
- **No trend** (ტრენდი არ არის);
- **Linear trend** (წრფივი ტრენდი);
- **Exponential** (ექსპონენციალური);
- **Damped trend** (დემფირებული (ჩაქრობადი) ტრენდი).

**User-def. Initial value** ოპცია გამოიყენება  $S(0)$  საწყისი მნიშვნელობის მოცემისათვის.

**Initial trend** (საწყისი ტრენდი) ოპცია გამოიყენება ტრენდის საწყისი მნიშვნელობის დასაფიქსირებლად. თუ ეს ოპცია არ გამოიყენება, მაშინ ტრენდის საწყისი მნიშვნელობა პროგრამაში შეფასდება.

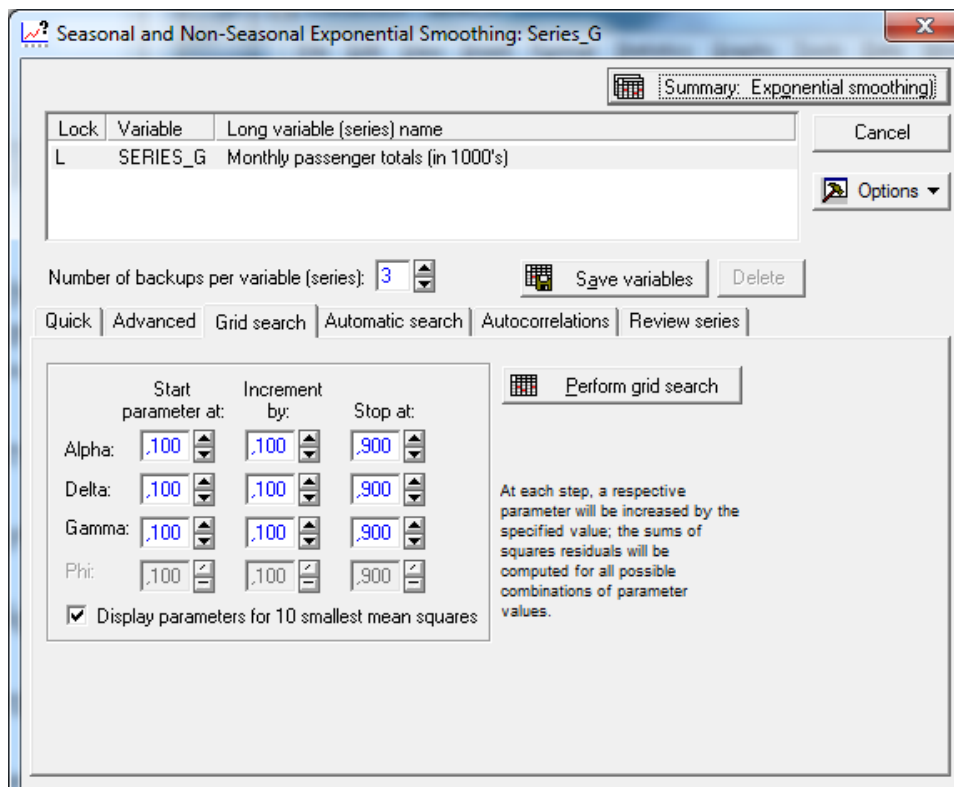
**Alpha, Delta, Gamma, Phi** ველებით მოცემულია გაგლუვების პარამეტრებით. მაგალითად, **Alpha** საჭიროა ყველა ექსპონენციალური გაგლუვებისათვის. სხვა მაჩვენებლები საჭიროა სპეციალური მოდელებისათვის. **Gamma** პარამეტრი გამოიყენება წრფივი და არაწრფივი ტრენდებისათვის და აგრეთვე ისეთი მოდელებისათვის, რომელთაც გააჩნიათ ჩაქრობადი ტრენდი და არ გააჩნიათ სეზონური კომპონენტები.

შემდეგი ორი ოპცია გამოიყენება შედეგების გრაფიკულად წარმოდგენისათვის.

- **Make summary plot for each smooth** (ყოველი გაგლუვებისათვის ავტომატურად შედგებოდა გრაფიკები);
- **Add pread/errors to work area** (გაგლუვებული მწკრივის დამატება/ნარჩენები სამუშაო არეში).

**Forecast cases** (შემთხვევების პროგნოზი) არეში მიეთითება რამდენი მნიშვნელობისათვის გვინდა პროგნოზის გაკეთება.

**Additive** ოპციების საშუალებით ხდება მოცემული მწკრივისათვის შესაფერისი ტრენდის შერჩევა. **Seasonal component** არეში მიუთითოთ **lag = 12**. ჩაერთოთ **Grid search** ჩანართი. ეკრანზე გამოჩნდება პარამეტრების მნიშვნელობები, სადაც ნაჩვენებია საწყისი, ზღვრული პარამეტრები და ბიჯი.



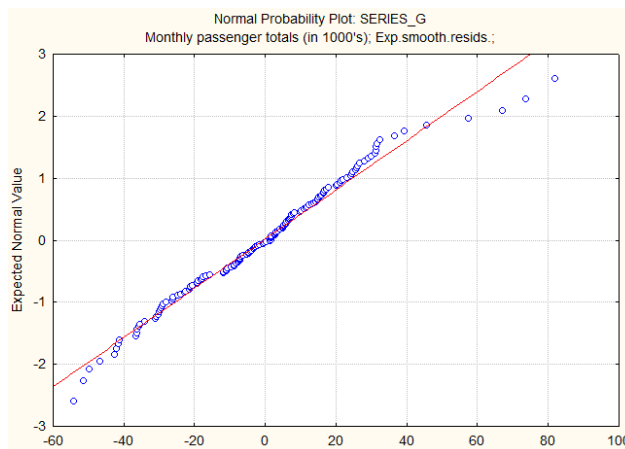
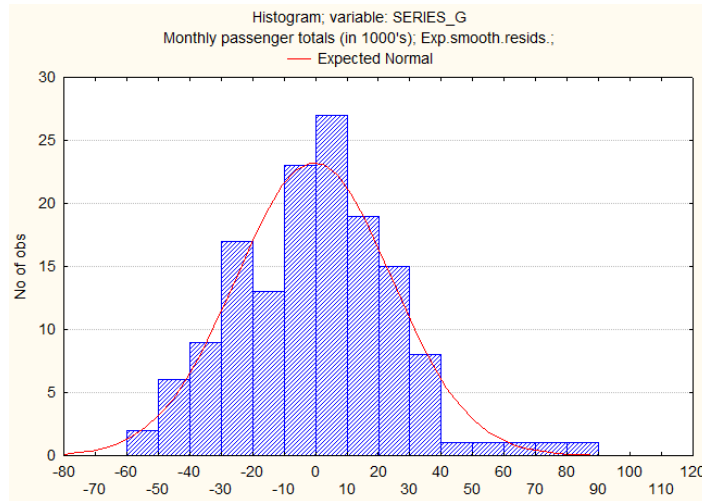
– **click Perform grid search** ღილაკზე. ამ ბრძანებით პროგრამა მოახდენს პარამეტრების შესაძლო მნიშვნელობების გადარჩევას და ეკრანზე გამოსულ ცხრილის პირველ სრიქონში მიუთითებს პარამეტრების საუკეთესო მნიშვნელობებს.

Parameter grid search (Smallest abs. errors are highlighted) (Series_G) Model: Expon. trend, add.season (12); S0=120,6 T0=1,008 SERIES_G: Monthly passenger totals (in 1000's)									
Model Number	Alpha	Delta	Gamma	Mean Error	Mean Abs Error	Sums of Squares	Mean Squares	Mean % Error	Mean Abs % Error
154	0,200000	0,900000	0,100000	-0,539034	12,02541	34948,92	242,7009	-0,276526	5,237676
145	0,200000	0,800000	0,100000	-0,525827	12,25777	36472,68	253,2825	-0,284180	5,305715
73	0,100000	0,900000	0,100000	-0,625175	12,12782	36852,55	255,9205	-0,237975	5,119473
64	0,100000	0,800000	0,100000	-0,595069	12,17511	37643,18	261,4110	-0,240639	5,098652
155	0,200000	0,900000	0,200000	-0,621882	12,59774	38492,36	267,3080	-0,342267	5,559599
136	0,200000	0,700000	0,100000	-0,513936	12,60913	39135,24	271,7725	-0,293995	5,419627
55	0,100000	0,700000	0,100000	-0,563808	12,51606	39416,10	273,7229	-0,245444	5,236786
235	0,300000	0,900000	0,100000	-0,589378	13,22174	39448,90	273,9507	-0,354056	5,844570
74	0,100000	0,900000	0,200000	-0,588828	12,70447	39531,69	274,5256	-0,276314	5,399838
146	0,200000	0,800000	0,200000	-0,610063	12,80810	39545,50	274,6215	-0,351569	5,613842

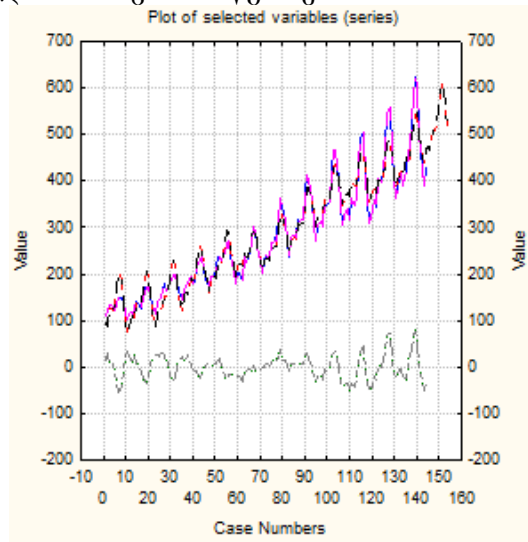
– click **Summary: Exponential smoothing** ღილაკზე. ეკრანზე გამოდის ცხრილი, სადაც წარმოდგენილია საწყისი, პროგნოზის მნიშვნელობები, ნაშთები და პროგნოზირების ახალი 12 მნიშვნელობა.

Exp. smoothing: Additive season (12) S0=120,6 T0=1,008 (Series_G) Expon.trend,add.season; Alpha= ,100 Delta=,100 Gamma=,100 SERIES_G: Monthly passenger totals (in 1000's)				
Case	SERIES_G	Smoothed Series	Resids	Seasonal Factors
133	417,0000	419,8070	-2,8070	
134	391,0000	412,3180	-21,3180	
135	419,0000	449,8559	-30,8559	
136	461,0000	440,9787	20,0213	
137	472,0000	449,8851	22,1149	
138	535,0000	495,6702	39,3298	
139	622,0000	539,8672	82,1328	
140	606,0000	548,4381	57,5619	
141	508,0000	506,2786	1,7214	
142	461,0000	471,8748	-10,8748	
143	390,0000	441,3714	-51,3714	
144	432,0000	468,1717	-36,1717	
145		475,2827		
146		467,6489		
147		507,9464		
148		508,5692		
149		517,3071		
150		563,8587		
151		608,9954		
152		607,2675		
153		553,8059		
154		517,5418		

მოდელის ადეკვატურობა დამატებით შეიძლება ვნახოთ **Review series** ჩანართის საშუალებით.



*click Plot two var lists with different scaies* გერანზე გამოდის საწყისი მწკრივი, პროგნოზის მწკრივი და ნაშთების მწკრივი.



როგორც ნახაზიდან ჩანს, ნაშთების მწკრივი სტაციონარულია.

## პრაქტიკული სამუშაო 15

## სეზონური დეკომპოზიცია (სეზონი 1)

*Time Series/Forecasting* მოდულში რეალიზირებულია სეზონური დეკომპოზიციის ორი სახე: კლასიკური დეკომპოზიცია (სეზონი 1) და ე.წ. *XII/Y2K(Census 1) - Montly* (12 თვიანი სეზონური კორექტირება). განვიხილოთ კლასიკური დეკომპოზიცია.

როგორც ცნობილია, დროითი მწკრივი შეიძლება წარმოვადგინოთ ოთხი ურთიერთდაკავშირებული კომპონენტის (ფაქტორების) საშუალებით: ტრენდი  $T_t$ , სეზონური  $S_t$ , ციკლური  $C_t$  და შემთხვევითი  $R_t$  არარეგულარული კომპონენტებით. ციკლურ და სეზონურ კომპონენტებს შორის სხვაობა იმაში მდგომარეობს, რომ სეზონურ კომპონენტს გააჩნია რეგულარული (სეზონური) პერიოდი, მაშინ როცა ციკლური ფაქტორს ჩვეულებრივ გააჩნია უფრო ხანგრძლივი ეფექტი და იცვლება ციკლიდან ციკლამდე. სეზონური დეკომპოზიციის მეთოდში ტრენდი და ციკლური კომპონენტა ჩვეულებრივ ერთიანდებიან ერთ ტრენდ-ციკლურ ( $T_t C_t$ ) კომპონენტაში.

არსებობს ადიტიური და მულტიპლიკატიური მოდელები. ადიტიური მოდელის შემთხვევაში დროითი მწკრივის მნიშვნელობა  $t$ -ურ მომენტში განსაზღვრულია ოთხივე კომპონენტის ჯამით:

$$y_t = T_t + C_t + S_t + R_t,$$

ხოლო მულტიპლიკატიური მოდელში – ოთხივე კომპონენტის გადამრავლებით:

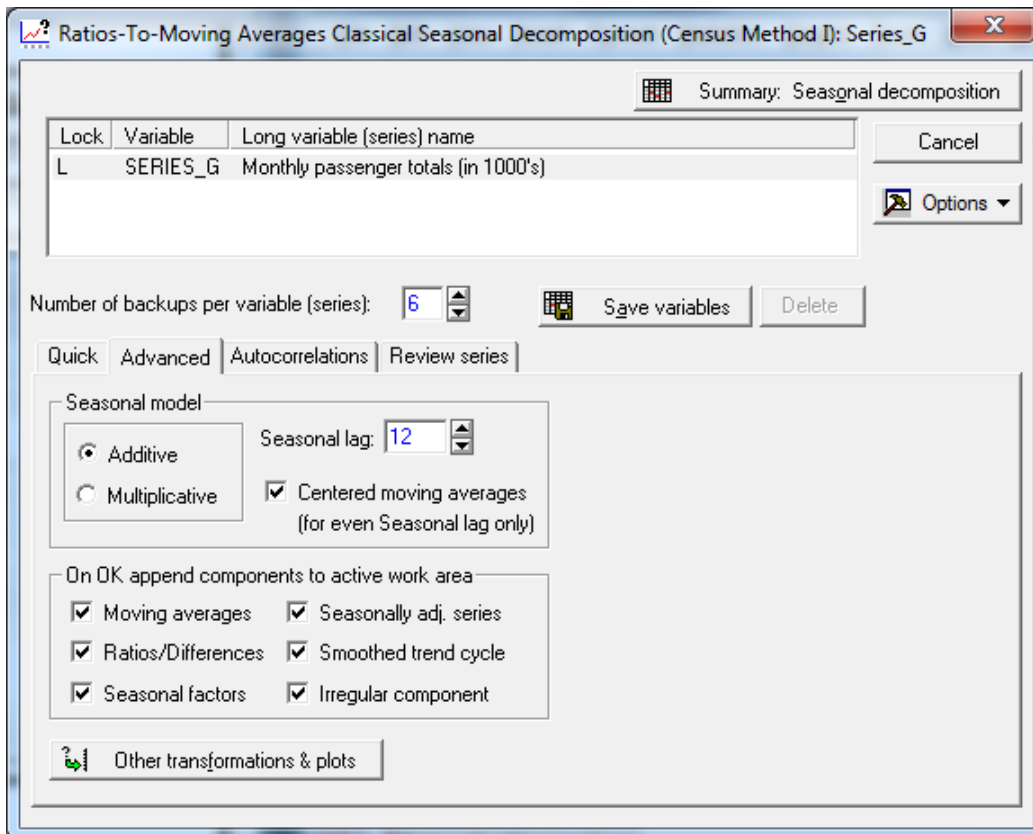
$$y_t = T_t \cdot C_t \cdot S_t \cdot R_t.$$

დროითი მწკრივის გრაფიკული გამოსახულებიდან გამომდინარე ამ ორ მოდელს შორის განსხვავება შემდეგია: ადიტიური მოდელის შემთხვევაში დროით მწკრივს გააჩნია მუდმივი სეზონური რხევები, რომელთა სიდიდე არ არის დამოკიდებული დროითი მწკრივის საერთო დონის მნიშვნელობაზე. მულტიპლიკატიური მოდელის შემთხვევაში სეზონური რხევის სიდიდე იცვლება დროითი მწკრივის საერთო დონის სიდიდესთან ერთად.

უნდა აღინიშნოს, რომ *ARIMA* მოდელს სჭირდება მინიმუმ 8 სრული სეზონური ციკლის მნიშვნელობები (ჩვენ შემთხვევაში  $8 \times 12 = 96$ ), ხოლო სეზონური დეკომპოზიციის მეთოდისთვის საკმარისია 5 სრული სეზონური ციკლი.

*Time Series/Forecasting* სასტარტო ფანჯარაში *click Seasonal decomposition (census1)* ღილაკზე. გაიხსნება *Ratius-To-Moving Averages classical Seasonal Decomposition (Census Method 1)* ფანჯარა:





სადაც *Seasonal model* ჩარჩოში უნდა შევარჩიოთ მოდელი:

- *Additive* (ადიტიური)
- *Multiplicative* (მულტიპლიკაციური)

*Seasonal lag* ოპციის ველში უნდა მიუთითოთ სეზონური პერიოდის სიგრძე.

*Advanced* ჩანართის *Centered moving averages (for even Seasonal lag only)* (ლუწი სიგრძის სეზონისათვის ცენტრირებული მცოცავი საშუალო) ოპცია საშუალებას იძლევა ლუწი *lag*-ისათვის შეარჩიოს ან ერთნაირწონიანი სრიალა საშუალო ან ისე, რომ პირველი და ბოლო დაკვირვებებს გააჩნდეს არათანაბარი წონითი კოეფიციენტები. ეს მეორე მეთოდი შეიძლება იმ შემთხვევაში, თუ ეს ოპცია ჩართულია.

*On Ok append components to active work area* ოპციების ჯგუფში განლაგებული შემდეგი ოპციები:

- *Moving averages* (მცოცავი საშუალო); რომლის საშუალებითაც ჯერ დროითი მწკრივისათვის განისაზღვრება მცოცავი საშუალო, ამ დროს ფანჯრის სიგრძე აიღება სეზონური პერიოდის სიგრძის ტოლი. თუ სეზონის პერიოდი ლუწი რიცხვია, მაშინ უნდა გამოვიყენოთ ან ერთწონიანი მცოცავი საშუალო ან არათანაბარწონიანი.
- *Ratios/Differences* (ფარდობა/სხვაობა). მცოცავი საშუალოს შერჩევის შემდეგ სეზონური ცვალებადობა გამოირიცხება და ამიტომ დაკვირვებებსა და გაგლუვებულ მნიშვნელობას შორის სხვაობა (ადიტიური მოდელის დროს) ან ფარდობა (მულტიპლიკაციური მოდელის დროს) გამოიწვევს სეზონური კომპონენტის გამოყოფას (პლიუს არარეგულარული კომპონენტა).

- **Seasonal factors** (სეზონური ფაქტორები). განისაზღვრება სეზონური კომპონენტა, როგორც საშუალო (ადიტიური მოდელისათვის) ან მედიანა (მულტიპლიკატიური მოდელისათვის) სეზონური მწკრივის ყველა მნიშვნელობებისათვის;
- **Seasonal agj. Series** (მწკრივი, სეზონური მდგენელით კორექტირებული). საწყისი მწკრივი შეიძლება კორექტირებული იყოს თუ მის მნიშვნელობებს გამოვაკლებთ (ადიტიური მოდელი) სეზონური მდგენელის მნიშვნელობებს ან გაგყოფთ (მულტიპლიკატიური მოდელი).
- **Smoothed trend cycle** (გაგლუვებული ტრენდი ციკლური კომპონენტა). ბოლო ბიჯზე გამოიყოფა შემთხვევითი ან არარეგულარული (ცდომილება) კომპონენტა მწკრივიდან ტრენდ-ციკლური კომპონენტის გამოკლებით (ადიტიური მოდელი) ან გაყოფით (მულტიპლიკატიური მოდელი).

დავაყენოთ სეზონური ძვრა lag = 12. ჩავრთოთ **On Okappend components to active work area**-ს ყველა ოპცია. ჩავრთოთ **Additive** ოპცია. **Number of backups ...** ველში დავაყენოთ 6.

– **click Summary: Seasonal decomposition** (შედგებები, სეზონური დეკომპოზიცია)-ის ღილაკზე. ეკრანზე გამოდის შედეგების ცხრილი, სადაც წარმოდგენილია მწკრივის მდგენელების მნიშვნელობები.

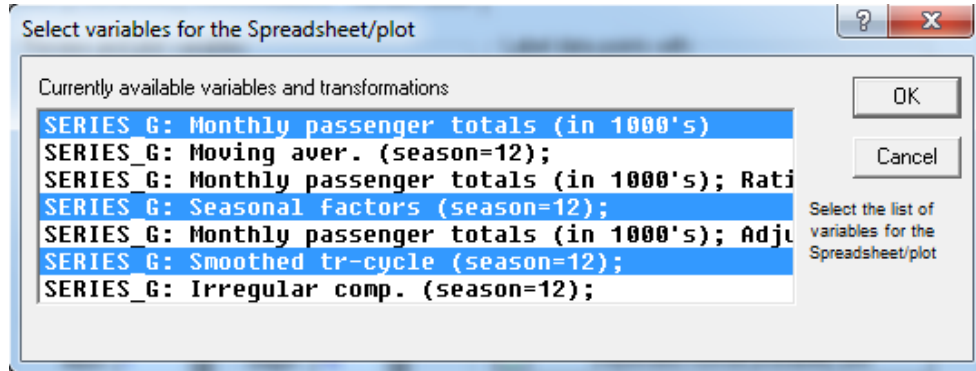
Seasonal Decomposition: Additive season (12); Centered means (Series_G)							
SERIES_G: Monthly passenger totals (in 1000's)							
Case	SERIES_G	Moving Averages	Diffnrcs	Seasonal Factors	Adjusted Series	Smoothed Trend-c.	Irreg. Compon.
2	118,0000			-36,1881	154,1881	141,7260	12,4621
3	132,0000			-2,2412	134,2412	138,6031	-4,3620
4	129,0000			-8,0366	137,0366	131,5989	5,4377
5	121,0000			-4,5063	125,5063	118,6886	6,8178
6	135,0000			35,4028	99,5972	104,4840	-4,8868
7	148,0000	126,7917	21,2083	63,8308	84,1692	96,3380	-12,1688
8	148,0000	127,2500	20,7500	62,8232	85,1768	100,2298	-15,0530
9	136,0000	127,9583	8,0417	16,5202	119,4798	116,6490	2,8308
10	119,0000	128,5833	-9,5833	-20,6427	139,6427	133,8746	5,7681
11	104,0000	129,0000	-25,0000	-53,5934	157,5934	144,9482	12,6452
12	118,0000	129,7500	-11,7500	-28,6199	146,6199	148,4861	-1,8662
13	115,0000	131,2500	-16,2500	-24,7487	139,7487	148,6330	-8,8843
14	126,0000	133,0833	-7,0833	-36,1881	162,1881	149,1334	13,0547
15	141,0000	134,9167	6,0833	-2,2412	143,2412	145,4920	-2,2508
16	135,0000	136,4167	-1,4167	-8,0366	143,0366	138,9322	4,1044
17	125,0000	137,4167	-12,4167	-4,5063	129,5063	127,9108	1,5955
18	149,0000	138,7500	10,2500	35,4028	113,5972	118,0396	-4,4423
19	170,0000	140,9167	29,0833	63,8308	106,1692	114,5602	-8,3910
20	170,0000	143,1667	26,8333	62,8232	107,1768	120,4520	-13,2753

ცხრილიდან ადვილად შეიძლება შევამოწმოთ, რომ ტრენდ-ციკლური, სეზონური და არარეგულარული კომპონენტების აჯამებით ვღებულობთ საწყის დროით მწკრივს, კორექტირებული მწკრივის და სეზონური მდგენელების ჯამი იძლევა, აგრეთვე, საწყის დროით მწკრივს.

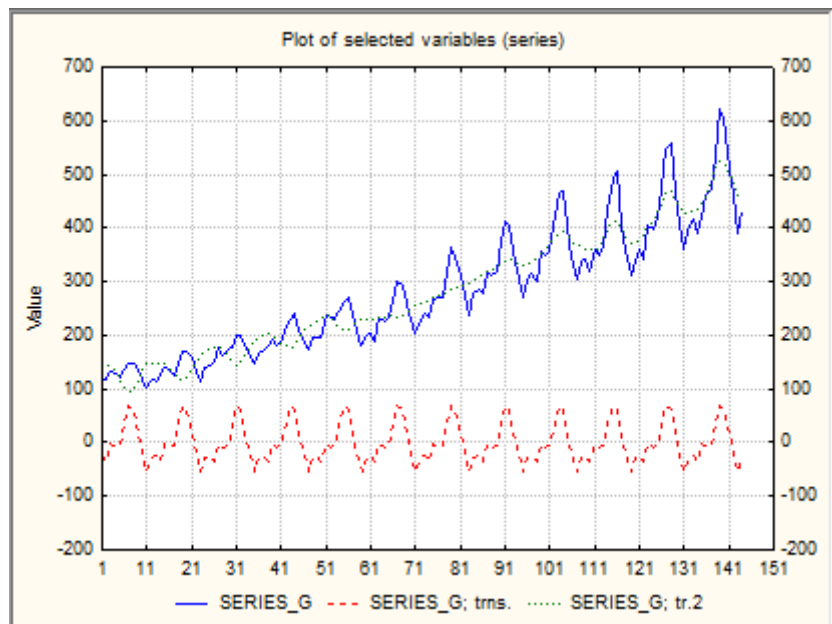
დეკომპოზიციის შედეგების გრაფიკული წარმოდგენისათვის ჩავრთოთ **Review series** ჩანართი. წინასწარ უნდა ჩავრთოთ **Display/plot subset of cases only**

და *Scale X axis in plots manually* ოპციები სათანადო მაჩვენებლების დაყენებით. მოვინიშნოთ *Scale X axis in plots manual* ოპცია. *click Review multiple variables* ღილაკის გვერდით არსებულ *Plot* ღილაკზე.

ეკრანზე გამოდის *Select variables for the Spreadsheet/plot* ფანჯარა



სადაც ჩამოთვლილია შედეგების მწკრივები. მოვინიშნოთ მწკრივები, მაგალითად *Monthly passenger totals*, *Seasonal factors*, *Smoothed tr-cycle* და შემდეგ *OK*. პროგრამა ააგებს შესაბამის გრაფიკებს.



## პრაქტიკული სამუშაო 16

### ფურიეს სპექტრული ანალიზი

სპექტრული ანალიზის საშუალებით დროით მწკრივებში შესაძლებელია პერიოდული რხევების გამოვლენა. **ARIMA** მოდელისაგან განსხვავებით სპექტრული ანალიზის მიზანია სეზონურობის ეფექტის გამოვლენა, მაშინ როდესაც **ARIMA** მოდელში სეზონურობის კომპონენტები წინასწარ ცნობილი უნდა იყოს.

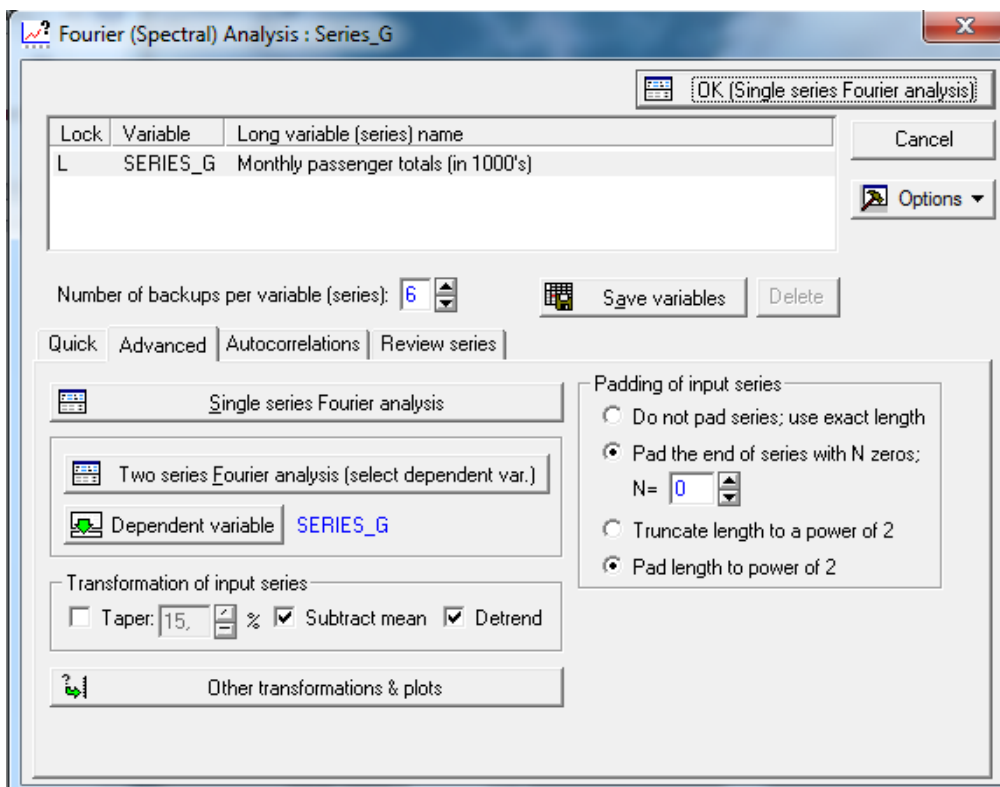
ფურიეს სპექტრული ანალიზის პროგრამა საშუალებას იძლევა ჩავატაროთ სტაციონარული დროითი მწკრივებისათვის სპექტრული ანალიზი, კერძოდ ავაგოთ პერიოდოგრამა, სპექტრული სიმკვრივის ფუნქციის შეფასება სხვადასხვა დროითი ფანჯრებისათვის. სპექტრული ანალიზის საშუალებით შესაძლებელია დროით მწკრივში ფარული პერიოდულობის გამოყოფა.

სპექტრული ანალიზისათვის წინასწარი სტანდარტული პროცედურებია: დროითი მწკრივების კოსინუსოიდალური ფანჯრით გაგლუვება, საშუალო მნიშვნელობის განსაზღვრა და ტრენდის გამორიცხვა.

ჩავტვირთოთ **Exmpl**-ის **Datasets** ბიბლიოთეკიდან **Series.G** ფაილი.

**Statistica** მენიუში მოვნიშნოთ **Advanced Linear/Nonlinear Models** და **click Time Series/Forecasting** ბრძანებაზე.

– ეკრანზე გამოსულ **Time Series Analysis** ფანჯარაში **click Spectral (Fourier) analysis** ღილაკზე.



ეკრანზე გამოდის *Fourier (Spectral) Analysis* ფანჯარა, სადაც ინფორმაციულ პანელში ჩაწერილია დროითი მწკრივის დასახელება. სიმბოლო  $L$  მიგვანიშნებს, რომ ცველადი გასაღებით არის დაკეტილია.

ფუნქციონალური ღილაკებიდან ყველაზე მნიშვნელოვანია შემდეგი:

- **Single series Fourier analysis** (ერთი მწკრივის ფურიეს ანალიზი);
- **Two series Fourier analysis** (ორი მწკრივის ფურიეს ანალიზი);

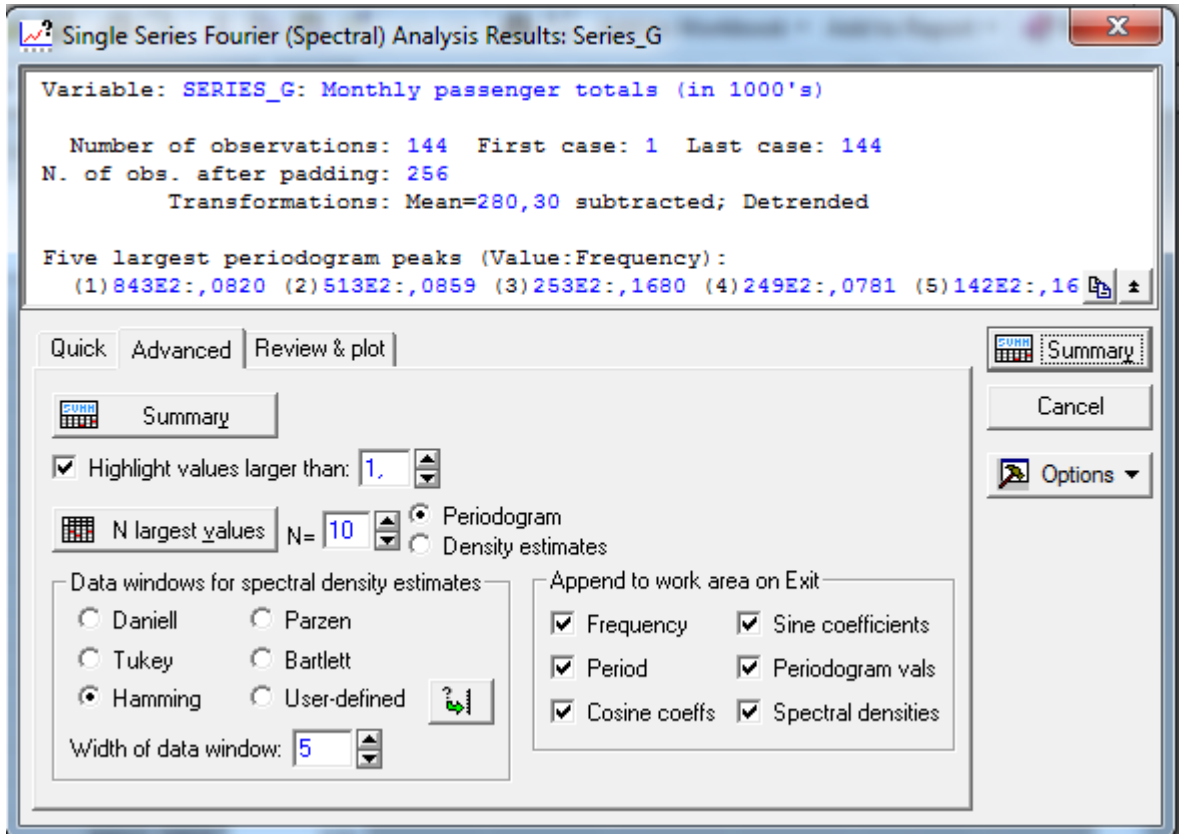
ჩავრთოთ **Advanced** ჩანართი. **Transformation of input series** (საწყისი მწკრივის გარდაქმნა) ჩარჩოში განთავსებული ოპციებით ხდება დროითი მწკრივის გარდაქმნა.

- **Toper** (ბოლოებში გაგლუვება). ეს პროცედურა რეკომენდირებულია პერიოდოგრამაში ცრუ პიკების გამოსარიცხად.
- **Subtract mean** (საშუალოს გამოთვლა). განისაზღვრება მწკრივის საშუალო სიდიდე;
- **Detrend** (ტრენდის გამორიცხვა). დროითი მწკრივიდან ხდება ტრენდის გამორიცხვა.

**Padding of input series** (საწყისი მწკრივის ბოლოში ნულების დამატება). ოპციების ჯგუფში შედიან:

- **Do not pad series: use exact leng** (ნულები არ დავამატოთ; გამოვიყენოთ მწკრივის ზუსტი სიგრძე). მწკრივში ნულების დამატება არ ხდება და ფორმალურად ფურიეს სწრაფი გარდაქმნის გამოყენება არ შეიძლება;
- **Pad the end of series with N zeros:** (მწკრივის ბოლოს დავუმატოთ  $N$  რაოდენობის ნული). ნულების რაოდენობა მიეთითება  $N=$  ველში;
- **Truncate Pength to a power of 2** (მწკრივს ჩამოვაჭრათ მნიშვნელობები მანამ, სანამ მწკრივი არ გახდება 2-ის ხარისხის ტოლი);
- **Pad length to power of 2** (გავზარდოთ მწკრივის სიგრძე უახლოესი 2 ხარისხის ტოლად). მწკრივს ბოლოში დავამატება იმდენი ნული, სანამ მისი სიგრძე არ გახდება 2-ის ხარისხის ტოლი.

ოპციების დაყენების შემდეგ **click OK (single series Fourier analysis)** ღილაკზე. პროგრამა ჩაატარებს სპექტრულ ანალიზს და ეკრანზე გამოდის **Single series Fourier (Spectral) Analysis Results** შედეგების ფანჯარა.



**Summary** ღილაკის საშუალებით შესაძლებელია სპექტრული ანალიზის შედეგის კომპაქტურად წარმოდგენა. ამ შემთხვევაში ეკრანზე გამოდის შედეგების ცხრილი, სადაც წარმოდგენილია სიხშირეები, პერიოდები, სინუსის და კოსინუსის კოეფიციენტები, პერიოდოგრამის მნიშვნელობები, სპექტრული სიმკვრივის შეფასებები და წონითი მნიშვნელობები, რომლებიც გამოიყენებიან სპექტრული სიმკვრივის შეფასებების მისაღებად.

Spectral analysis: SERIES_G: Monthly passenger totals (in 100 (Series_G))							
No. of cases: 256							
	Frequency	Period	Cosine Coeffs	Sine Coeffs	Periodogram	Density	Hamming Weights
0	0,000000		-0,0000	-0,0000	0,00	1875,08	0,035714
1	0,003906	256,0000	0,6037	-4,4609	2593,77	3403,80	0,241071
2	0,007813	128,0000	8,0473	1,8833	8743,13	4874,89	0,446429
3	0,011719	85,3333	0,4100	3,1286	1274,37	3120,04	0,241071
4	0,015625	64,0000	-0,3855	2,9032	1097,87	1896,10	0,035714
5	0,019531	51,2000	1,4951	-4,0743	2410,96	2776,81	
6	0,023438	42,6667	3,7712	5,5396	5748,45	3276,58	
7	0,027344	36,5714	-0,5641	-0,8195	126,69	1954,46	
8	0,031250	32,0000	0,5310	3,5637	1661,70	1198,75	
9	0,035156	28,4444	0,8173	-2,2121	711,88	1078,20	
10	0,039063	25,6000	1,8077	2,7482	1384,99	1006,55	
11	0,042969	23,2727	1,4378	1,6172	599,35	746,06	
12	0,046875	21,3333	-1,1414	1,2242	358,58	595,99	
13	0,050781	19,6923	1,3785	2,2941	916,88	737,32	
14	0,054688	18,2857	-1,5068	-1,5193	586,07	1032,97	
15	0,058594	17,0667	2,5324	3,2922	2208,18	1263,03	

განვიხილოთ ეს მნიშვნელობები. დროითი მწკრივის მნიშვნელობები განისაზღვრებიან ფორმულით:

$$x_i = \frac{a_0}{2} + \sum_k \{a_k \cos[2\pi f_k(t-1)] + b_k \sin[2\pi f_k(t-1)]\},$$

სადაც  $f_k$  სიხშირე განისაზღვრება როგორც დროით ერთეულში ციკლების რაოდენობა. **Time Series/Forecasting** მოდულში დროის ერთეულად აღებულია ერთი დაკვირვება (ანუ სიხშირე გამოისახება როგორც ციკლის ნაწილი ერთ დაკვირვებაზე). სიხშირეები განისაზღვრებიან როგორც,  $\frac{k}{n}$ ,  $k=1,2,\dots,\frac{n}{2}$ ,

სადაც  $n$ -მწკრივის განზომიერება. მაგალითად, რიცხვი 0.083 ნიშნავს იმას, რომ თითოეული დაკვირვება შედგება მთელი ციკლის 0.083-საგან ან 12 დაკვირვება შეადგენს ერთ ციკლს ( $0.083 \cdot 12=1$ ). ამრიგად, თუ მწკრივი შედგება რამოდენიმე წლის თვიური მონაცემებისაგან, მაშინ შესაბამისი პერიოდულობა განსაზღვრავს წლიურ ციკლს.

თუ სიგნალის დაქვანტის სიხშირე  $\Delta t \neq 1$ , მაშინ სიხშირეები განისაზღვრებიან როგორც  $\frac{k}{\Delta t n}$ ,  $k=1,2,\dots,\frac{n}{2}$ .

კოსინუსის კოეფიციენტები  $a_k$  წარმოადგენენ რეგრესიის კოეფიციენტებს, რომლებიც გვიჩვენებენ კოსინუს ფუნქციის კორელაციას შესაბამისი სიხშირულ მონაცემებთან. სინუსის  $b_k$  კოეფიციენტების ინტერპრეტაცია ანალოგიურია კოსინუს-კოეფიციენტებისა.

პერიოდოგრამის მნიშვნელობები განისაზღვრებიან როგორც ყოველი სიხშირის სინუსებისა და კოსინუსების კოეფიციენტების კვადრატების ჯამი.

სპექტრული სიმკვრივის ფუნქციის შეფასება განისაზღვრება პერიოდოგრამის გაგლუვებით. ამისათვის **Data windows for spectral density estimation** (სპექტრული სიმკვრივის შეფასების ფანჯარები)-ის ოპციებიდან განსაზღვრავენ ერთ-ერთი სპექტრულ ფანჯარას: დანიელის, ტიუკის, ჰემინგის, პარზენის, ბარტლეტის. შესაძლებელია საკუთარის ფანჯრის გამოყენება **User-defind** (განსაზღვრულია მომხმარებლის მიერ) ოპციის საშუალებით.

**Width of data window** (ფანჯრის სიგანე) ოპციით შესაძლებელია ფანჯრის სიგანის დაყენება.

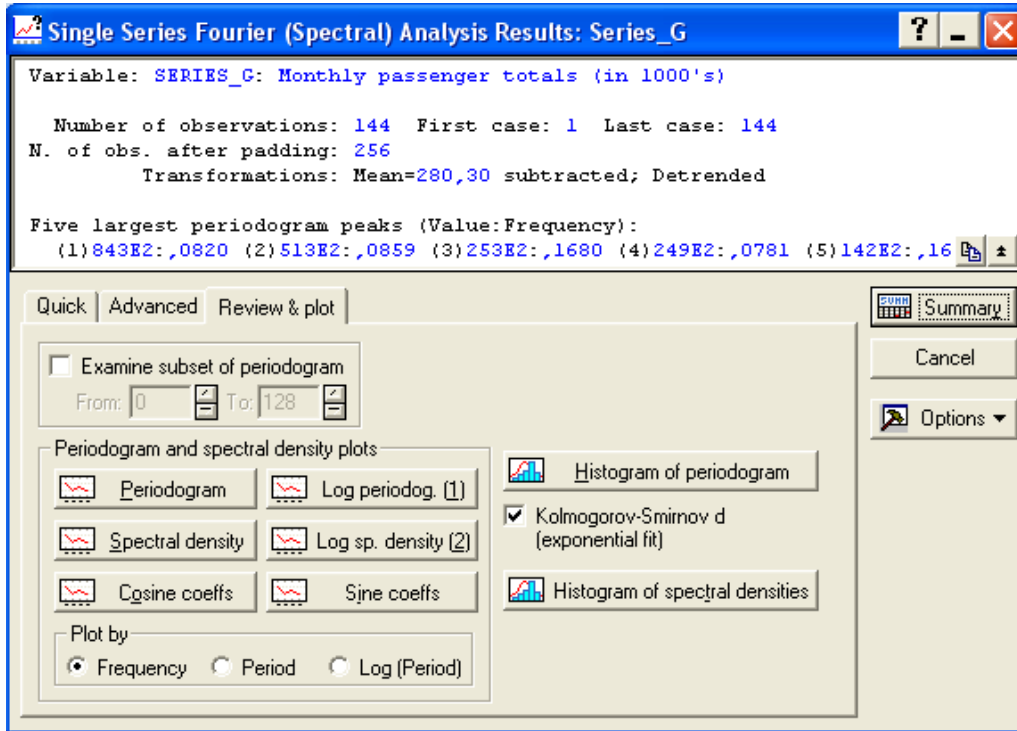
**Append to work area on Exit** (დაუმატოთ აქტიურ მუშა არეს) ჩარჩოში წარმოდგენილი ოპციები;

- **Frequency** (სიხშირე);
- **Period** (პერიოდი);
- **Cosine coeffs** (კოსინუსის კოეფიციენტები);
- **Sine coefficients** (სინუსის კოეფიციენტები);
- **Periodogram vals** (პერიოდოგრამის მნიშვნელობები);
- **Spectral densities** (სპექტრული სიმკვრივე);

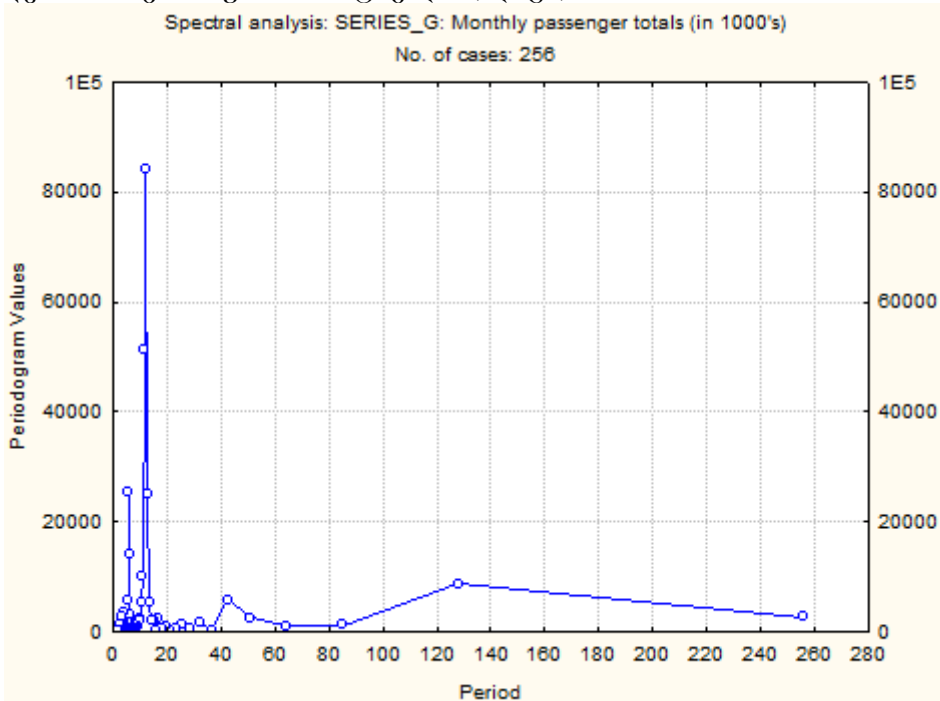
დოგორც წესი ყველა ეს ოპცია ჩართული უნდა იყოს.

**Review & plot** ჩანართის ჩართვით შესაძლებელია სხვადასხვა გრაფიკული შედეგების ეკრანზე გამოტანა: სპექტრული სიმკვრივის, პერიოდოგრამის, მათი ლოგარითმების, კოსინუს-სინუსური კოეფიციენტები. გრაფიკები შეიძლება აიგოს სიხშირით, პერიოდით ან ლოგ-პერიოდით, იმისდა

მიუხედავად თუ რომელი ოპცია იქნება ჩართული **Plot by** ოპციებიდან: **Frequency** (სიხშირე), **Period** (პერიოდი), **Log (Period)** (ლოგარითმული პერიოდი).

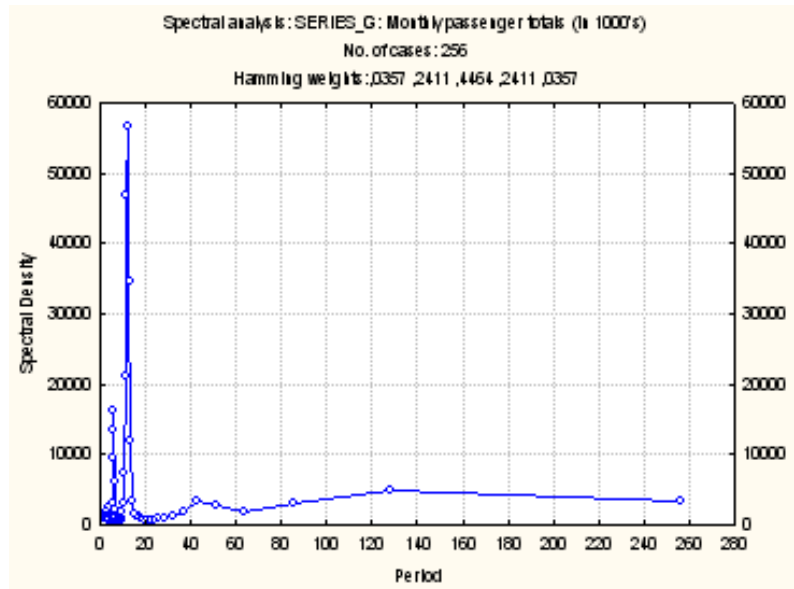


თუ ჩავრთავთ **Period** ოპციას და **click Periodogram** ღილაკზე, პროგრამა ააგებს პერიოდოგრამის გრაფიკს, რომლის საშუალებითაც შეგვიძლია ვიმსჯელოთ რეგულარული ციკლის არსებობაზე ან არარსებობაზე, განვსაზღვროთ სეზონურობის ციკლი (ლაგი).



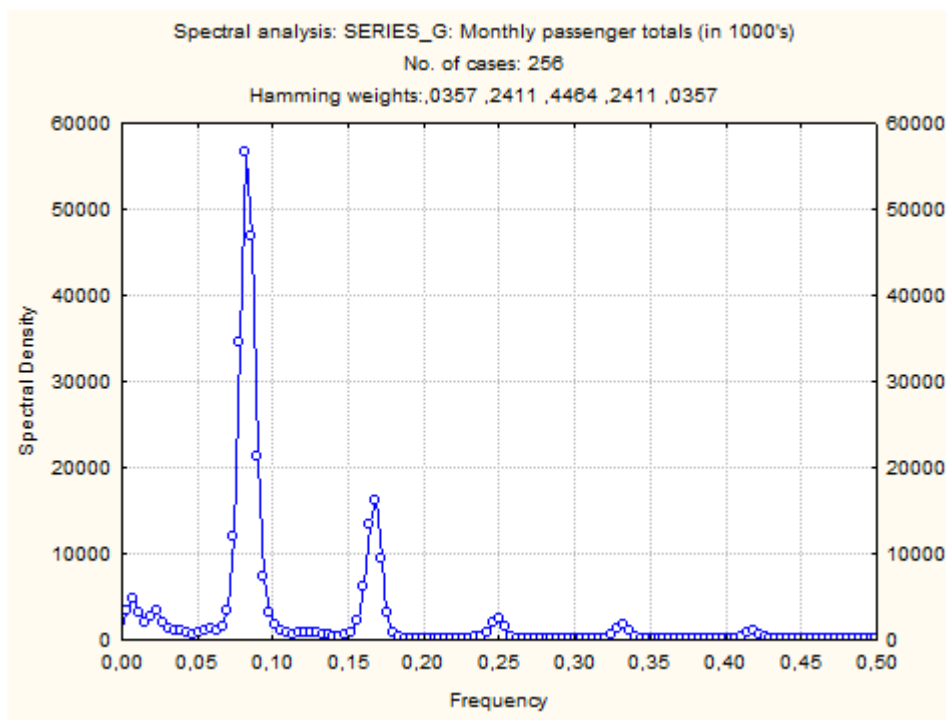
თუ ჩავრთავთ **Period** ოპციას და **click Spectral density** ღილაკზე, მაშინ პროგრამა დროით სივრცეში ააგებს სპექტრული სიმკვრივის გრაფიკს.



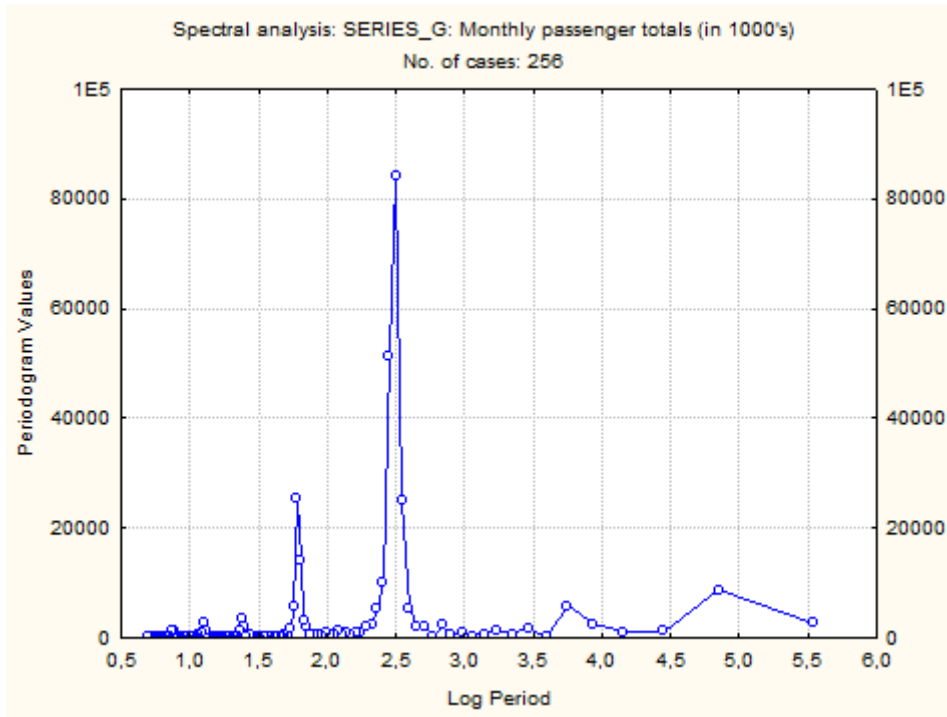


თუ ამ პერიოდოგრამებს ერთმანეთს შევადარებთ, ადვილი მისახვედრია, რომ სპექტრული სიმკვრივე წარმოადგენს პერიოდოგრამის გაგლუვების შედეგს.

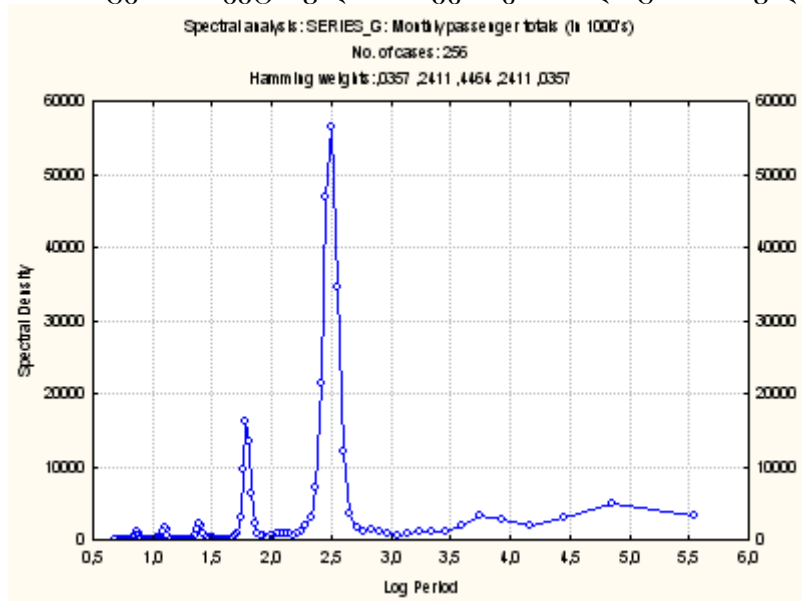
თუ ჩაერთავთ **Frequency** ოპციას და **click Spectral density** ღილაკზე, მაშინ პროგრამა ააგებს სისშირულ არეში სპექტრული სიმკვრივის გრაფიკს.



თუ ჩაერთავთ **Log [Period]** ოპციას და **click Periodogram** ღილაკზე, მაშინ პროგრამა ააგებს პერიოდოგრამის ლოგარითმულ გრაფიკს.

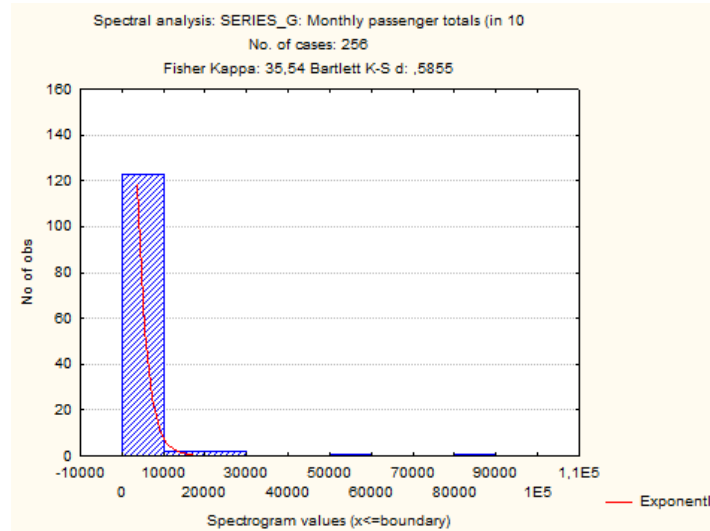


ხოლო თუ ჩავრთავთ  $\text{Log} [\text{Period}]$  ოპციას და *click Spectral density* ღილაკზე, მაშინ პროგრამა ააგებს სპექტრული სიმკვრივის ლოგარითმულ გრაფიკს.



თუ დროითი მწკრივის მნიშვნელობები ერთმანეთის მიმართ დამოუკიდებელნი არიან (ე.ი. პერიოდულობა არ გააჩნიათ) და ექვემდებარებიან ნორმალურ განაწილებას, მაშინ ასეთი დროითი მწკრივი შეიძლება იყოს თეთრი ხმაური. თუ საწყისი მწკრივი – თეთრი ხმაურია, მაშინ შესაბამისი პერიოდოგრამის მნიშვნელობებს გააჩნიათ ექსპონენციალური განაწილება.

**Histogram of periodogram** ღილაკით აიგება პერიოდოგრამის მნიშვნელობების ჰისტოგრამა და ექსპონენციალური განაწილების მრუდი. გარდა ამისა, მომხმარებელს შეუძლია მოითხოვოს კოლმოროვ-სმირნოვის  $d$ -სტატისტიკის განსაზღვრა.



### გამოყენებული ლიტერატურა

1. Халафян А.А. STATISTICA 6. Статистический анализ данных. Учебник М., Бином-Пресс, 2007.
2. Боровиков В.П.,Ивченко Г.И. Прогнозирование в системе STATISTICA в среде Windows. Учеб.пособие М.,Финансы и статистика, 2000.
3. Боровиков В.П. STATISTICA. Искусство анализа данных на компьютере. СПб,Питер,2003.
4. Валиев С.Г.,Клячкин В.Р. Практикум по прикладной статистике. Учеб.пособие,Ульяновск, 2008.
5. Вуколов Э.А. Основы статистического анализа. Практикум,Учеб.пособие, М.,ФОРУМ, 2008