Seminarski rad iz statističkog softvera 3

Ispitivanje baze: Survival from Malignant Melanoma

Radila: Tijana Molerović

Asistent: Marija Radičević

Matematički fakultet,Beograd, 2016.godina

Sadržaj:

Uvod

Baza korišćena za ovo istraživanje je Survival from Malignant Melanoma, odnosno preživeli od opake bolesti Melanom (raka kože) u Nemačkoj. Baza se sastoji 205 pacijenata, odnosno 205 opservacija i 7 promenljivih.

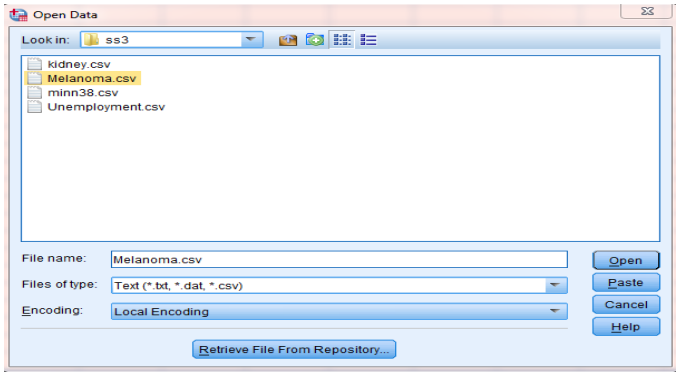
Atributi:

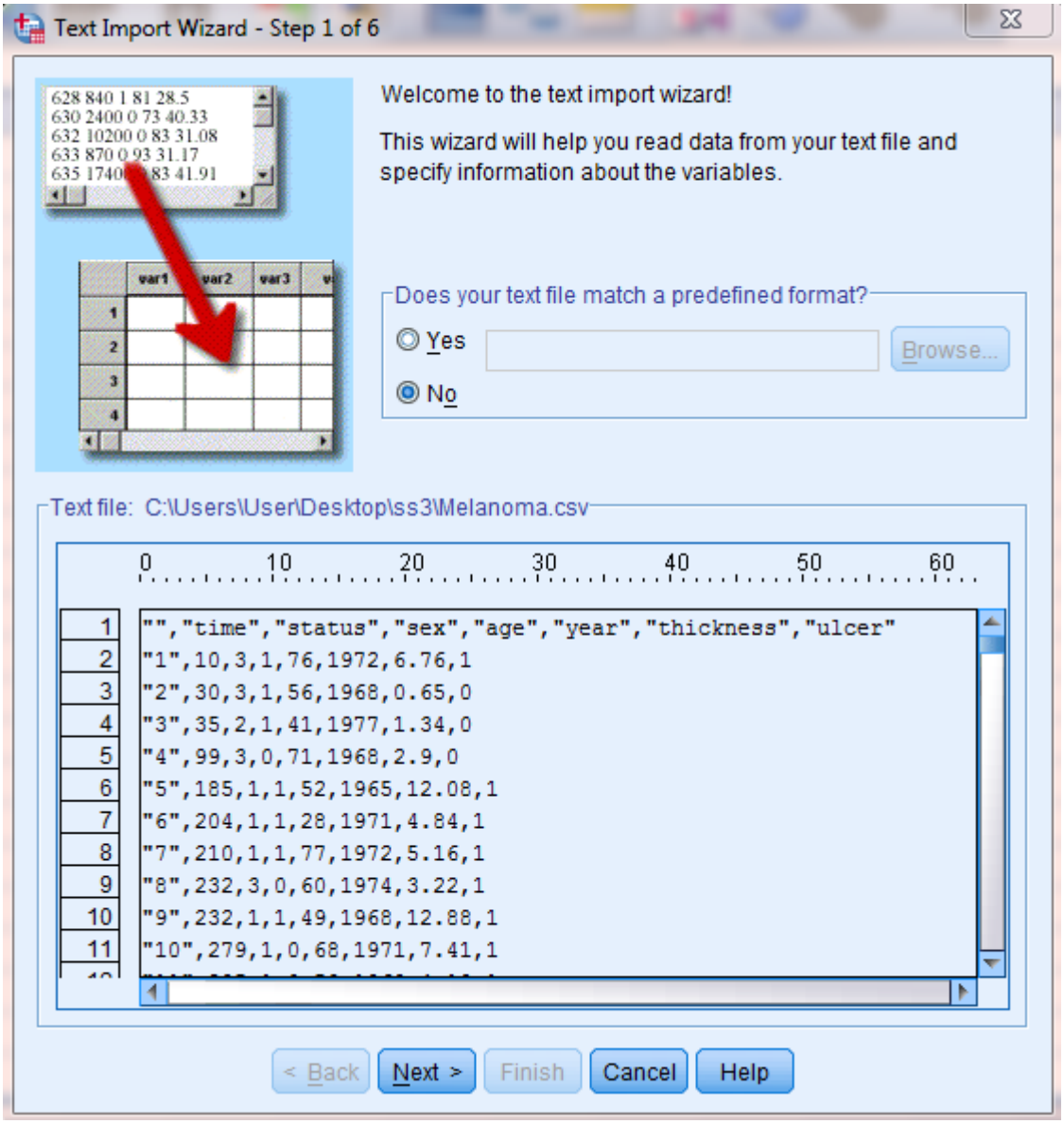
- Time- vreme lečenja u danima
- Status - 1 ako je umro od melanoma, 2 ako je živ, 3 mrtav iz drugih razloga
- Sex(pol) - 1 za muškarce, 0 za žene
- Age - godine ispitanika
- Year – godine rada(lečenja)
- Thickness – debljina tumora u milimetrima
- Ulcer(čir) – 1 prisustvo čira, 0 odsustvo čira

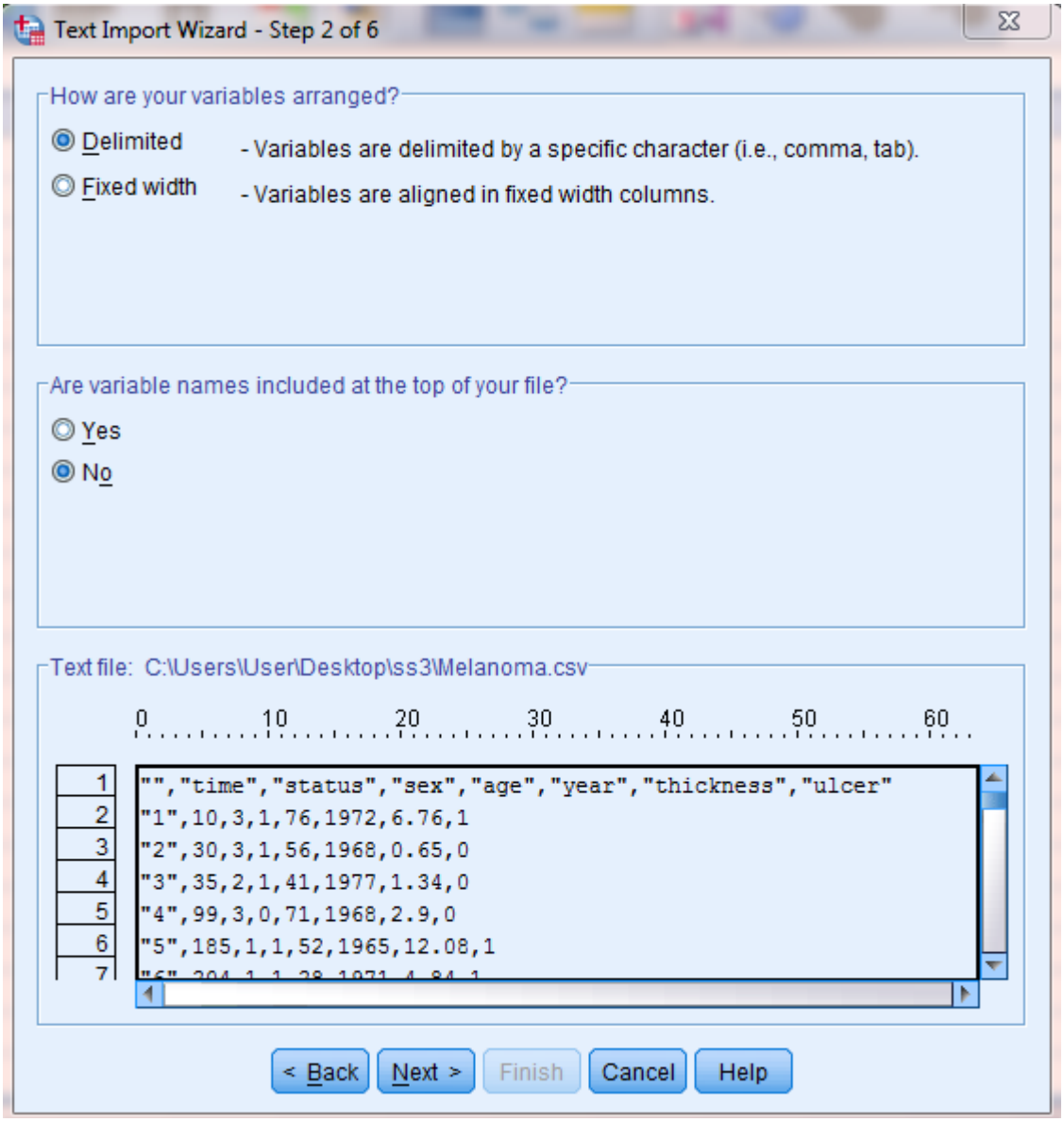
Naš cilj istraživanje jeste ispitivanje ove opeke bolesti,tj. preživljavanje, lečenje, posledice bolesti, uzroci bolesti, zastupljenost kod muškaraca i žena...

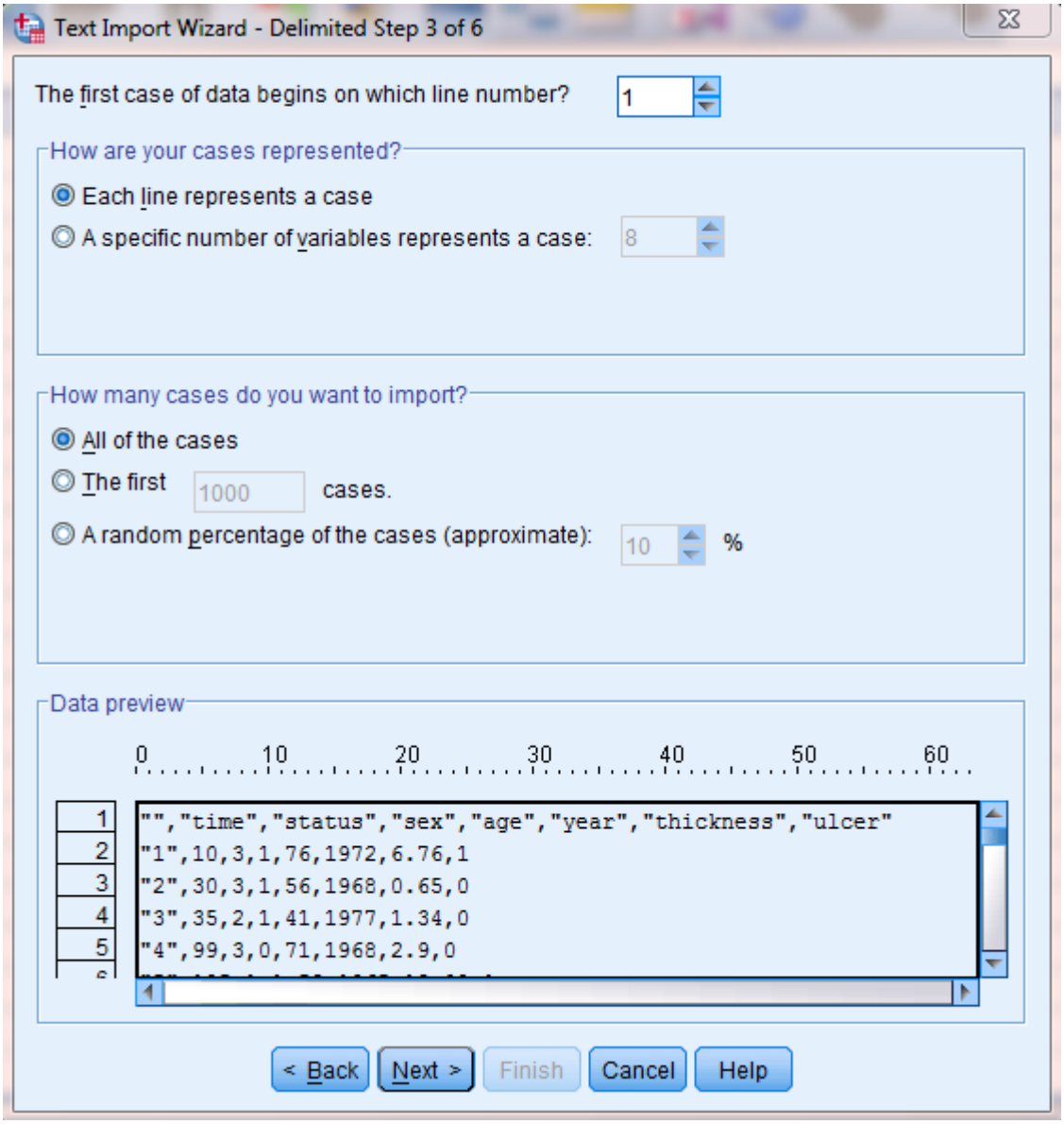
Učitavanje baze

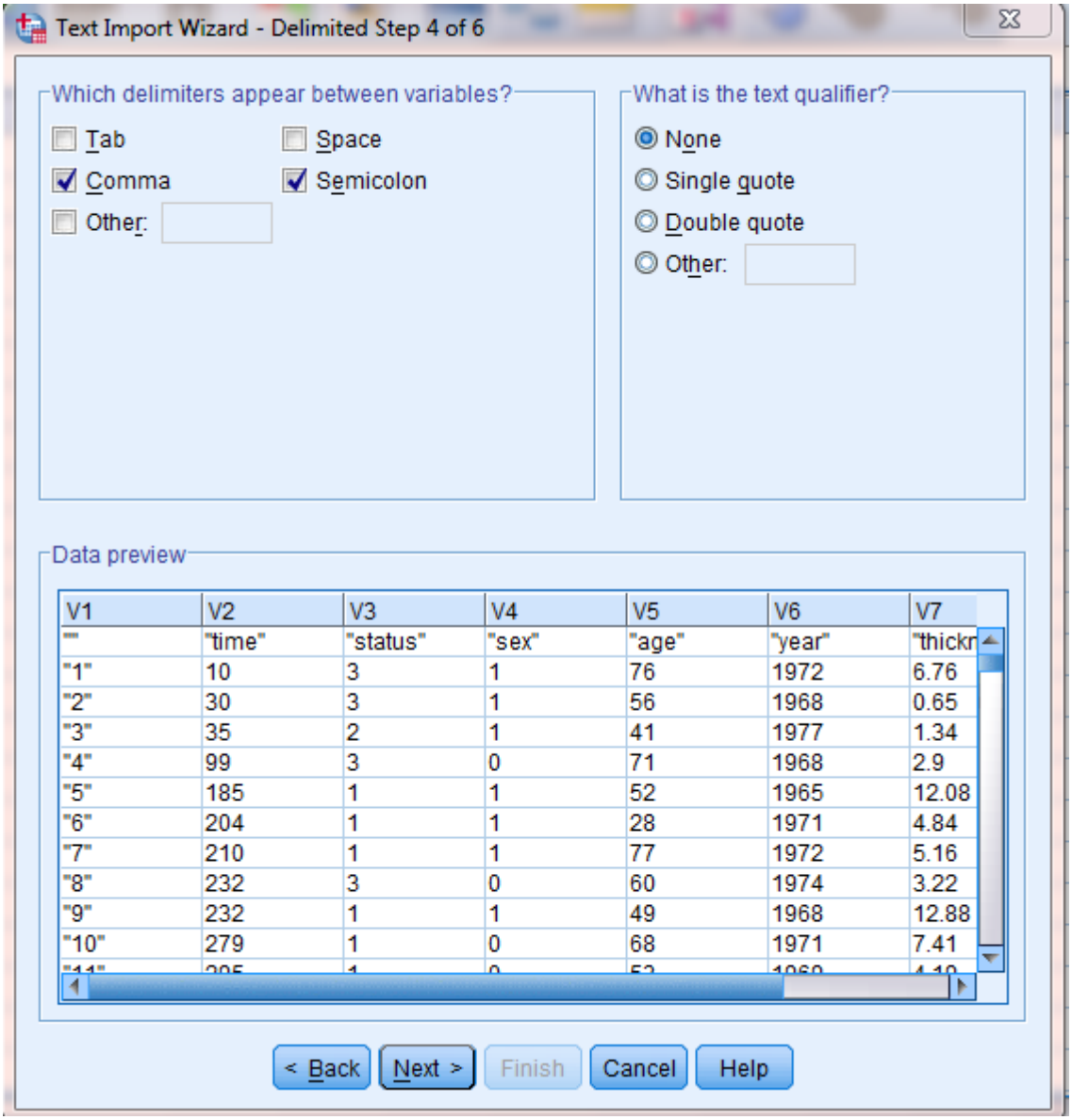
File→Read Text Data...

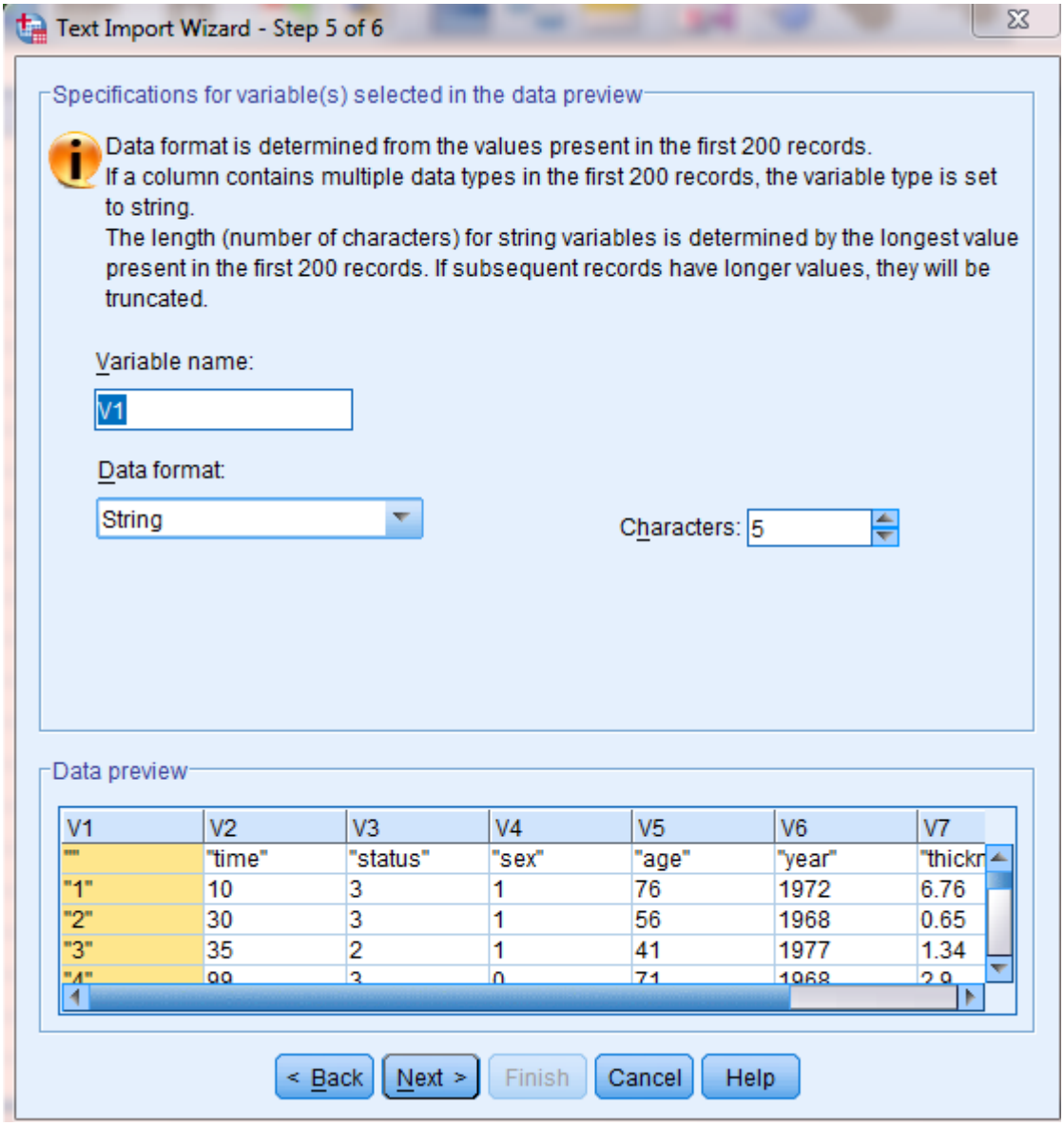


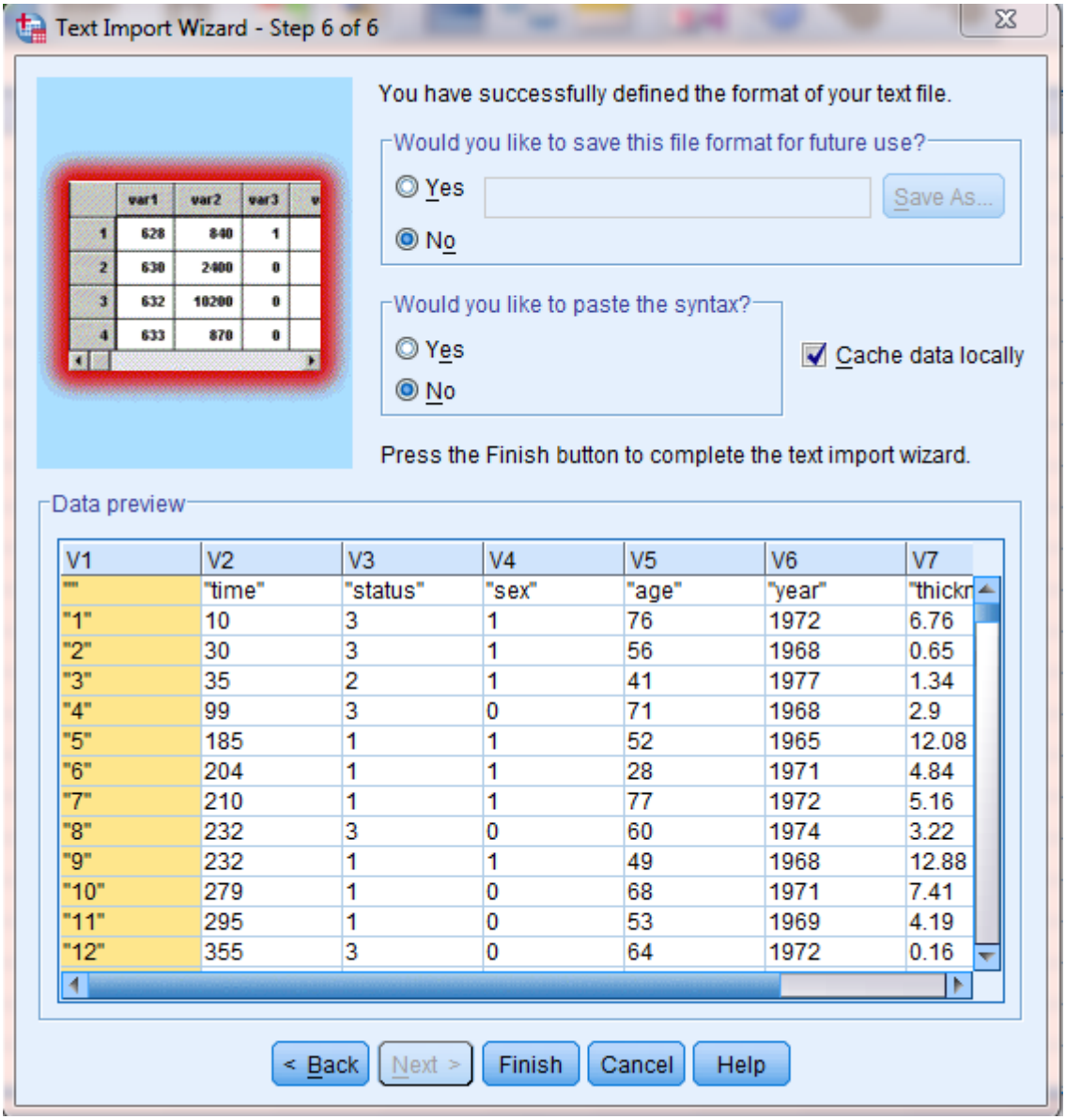

Text Import Wizard - Step 1 of 6
628 840 1 81 28.5
630 2400 0 73 40.33
632 10200 0 83 31.08
633 870 0 93 31.17
635 1740 0 83 41.91
var1
var2
var3
Welcome to the text import wizard!
This wizard will help you read data from your text file and specify information about the variables.
Does your text file match a predefined format?
Yes
No
Browse...
Text file: C:\Users\User\Desktop\ss3\Melanoma.csv
0
10
20
30
40
50
60
1 "","time","status","sex","age","year","thickness","ulcer"
2 "1",10,3,1,76,1972,6.76,1
3 "2",30,3,1,56,1968,0.65,0
4 "3",35,2,1,41,1977,1.34,0
5 "4",99,3,0,71,1968,2.9,0
6 "5",185,1,1,52,1965,12.08,1
7 "6",204,1,1,28,1971,4.84,1
8 "7",210,1,1,77,1972,5.16,1
9 "8",232,3,0,60,1974,3.22,1
10 "9",232,1,1,49,1968,12.88,1
11 "10",279,1,0,68,1971,7.41,1
< Back
Next >
Finish
Cancel
Help


Text Import Wizard - Step 2 of 6
How are your variables arranged?
Delimited
- Variables are delimited by a specific character (i.e., comma, tab).
Fixed width
- Variables are aligned in fixed width columns.
Are variable names included at the top of your file?
Yes
No
Text file: C:\Users\User\Desktop\ss3\Melanoma.csv
0 10 20 30 40 50 60
1 "","time","status","sex","age","year","thickness","ulcer"
2 "1",10,3,1,76,1972,6.76,1
3 "2",30,3,1,56,1968,0.65,0
4 "3",35,2,1,41,1977,1.34,0
5 "4",99,3,0,71,1968,2.9,0
6 "5",185,1,1,52,1965,12.08,1
7
< Back
Next >
Finish
Cancel
Help


Text Import Wizard - Delimited Step 3 of 6
The first case of data begins on which line number?
1
How are your cases represented?
Each line represents a case
A specific number of variables represents a case:
8
How many cases do you want to import?
All of the cases
The first
1000
cases.
A random percentage of the cases (approximate):
10
%
Data preview
0
10
20
30
40
50
60
1 ","time","status","sex","age","year","thickness","ulcer"
2 "1",10,3,1,76,1972,6.76,1
3 "2",30,3,1,56,1968,0.65,0
4 "3",35,2,1,41,1977,1.34,0
5 "4",99,3,0,71,1968,2.9,0
< Back
Next >
Finish
Cancel
Help


Text Import Wizard - Delimited Step 4 of 6
Which delimiters appear between variables?
Tab
Space
Comma
Semicolon
Other:
What is the text qualifier?
None
Single quote
Double quote
Other:
Data preview
V1 V2 V3 V4 V5 V6 V7
"" "time" "status" "sex" "age" "year" "thickn
"1" 10 3 1 76 1972 6.76
"2" 30 3 1 56 1968 0.65
"3" 35 2 1 41 1977 1.34
"4" 99 3 0 71 1968 2.9
"5" 185 1 1 52 1965 12.08
"6" 204 1 1 28 1971 4.84
"7" 210 1 1 77 1972 5.16
"8" 232 3 0 60 1974 3.22
"9" 232 1 1 49 1968 12.88
"10" 279 1 0 68 1971 7.41
< Back
Next >
Finish
Cancel
Help


Text Import Wizard - Step 5 of 6
Specifications for variable(s) selected in the data preview
Data format is determined from the values present in the first 200 records.
If a column contains multiple data types in the first 200 records, the variable type is set to string.
The length (number of characters) for string variables is determined by the longest value present in the first 200 records. If subsequent records have longer values, they will be truncated.
Variable name:
V1
Data format:
String
Characters: 5
Data preview
V1 V2 V3 V4 V5 V6 V7
"" "time" "status" "sex" "age" "year" "thickn
"1" 10 3 1 76 1972 6.76
"2" 30 3 1 56 1968 0.65
"3" 35 2 1 41 1977 1.34
< Back
Next >
Finish
Cancel
Help



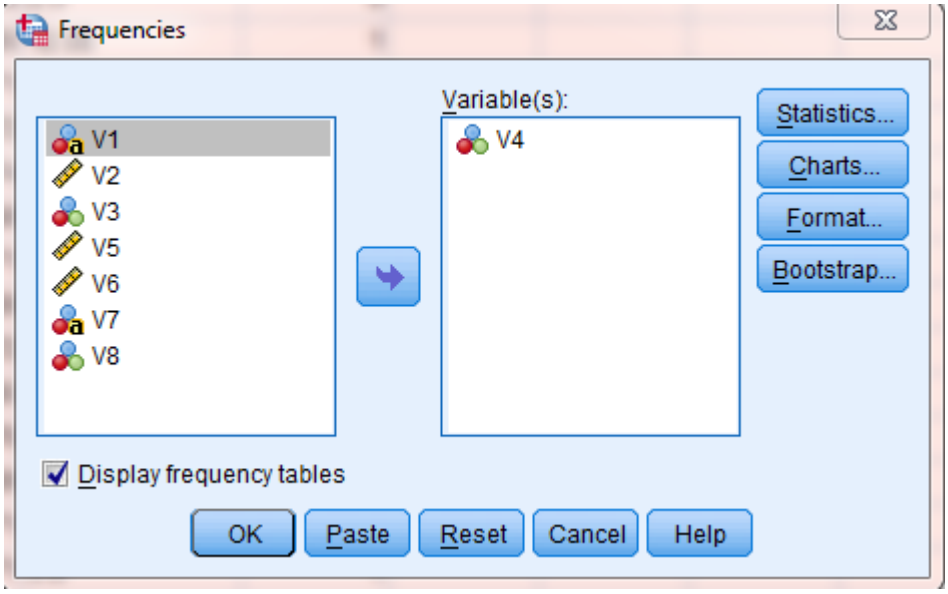
Deo učitane baze:

""	"time"	"status"	"sex"	"age"	"year"	"thickness"	"ulcer"
"1"	10	3	1	76	1972	6.76	1
"2"	30	3	1	56	1968	0.65	0
"3"	35	2	1	41	1977	1.34	0
"4"	99	3	0	71	1968	2.9	0
"5"	185	1	1	52	1965	12.08	1

Vidimo da su sve promenljive numeričkog tipa što i žeimo, ali ako kliknemo na Variable View piše da su tipa String . Promenićemo u tip Numeric, kao i Messure u Scale tamo gde je potrebno. Vreme, broj godina ispitanika, godine i debljina čira imaće intervalnu skalu, dok će

status, pol i prisustvo čira imati Nominalnu skalu(symbol predstavlja pripadnost određenoj grupi) . Takođe ćemo promenljivu V7 zaokružiti na ceo broj radi lakšeg rada(Variable View, Comma). Sada je naša baza spremna za rad.

Da bismo videli odnos žena i muškaraca *Analyze→Descriptive Statistic→Frequency*



Frequencies

[DataSet1]

Statistics

V4

N	Valid	205
	Missing	0

V4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	126	61,5	61,5	61,5
	1	79	38,5	38,5	100,0
	Total	205	100,0	100,0	

U ovom istraživanju učestvuje 126 žena i 79 muškaraca, što može ukazati da su žene podložnije ovoj bolesti u Nemačkoj. Ova bolest je inače više prisutnija kod žena nego kod muškaraca.

Promenićemo i imena promenljivih umesto V2 stavićemo time,V3 status... (radi lakšeg rada)

V1	String	5	0		None	None	5	Left	Nominal	Input
time	Numeric	6	0		None	None	6	Right	Scale	Input
status	Numeric	8	0		None	None	8	Right	Nominal	Input
sex	Numeric	5	0		None	None	5	Right	Nominal	Input
age	Numeric	5	0		None	None	5	Right	Scale	Input
year	Numeric	6	0		None	None	6	Right	Scale	Input
thickness	Comma	11	0		None	None	11	Right	Scale	Input
ulcer	Numeric	7	0		None	None	7	Right	Nominal	Input

Veličina tumora

Klasifikacija tumora prema debljini tumora po Breslovu i nivoa invazije po Klarku. Značajne tabele:

pTx – primarni tumor se ne može ispitati,
pT0 – nema dokaza o primarnom tumoru,
pTis – melanom *in situ* (atipična melanocitna hiperplazija ili teška melanocitna displazija), neinvazivna lezija Clark I,
pT1 – tumor debljine 0,75 mm ili manje i nivo invazije Clark II,
pT2 – tumor debljine 0,75 mm do 1,5 mm i/ili nivo invazije Clark III,
pT3 – tumor debljine 1,5 mm do 4,0 mm i/ili nivo invazije Clark IV,
pT3a – tumor debljine 1,5–3 mm,
pT3b – tumor debljine 3–4 mm,
pT4 – tumor debljine preko 4 mm i/ili nivo invazije Clark V i/ili satelit(i) do 2 cm od primarne promene,
pT4a – tumor debljine preko 4 mm i/ili nivo invazije Clark V,
pT4b – satelit(i) do 2 cm od primarnog tumora.

Klarkova klasifikacija označava nivoe invazije slojeva kože. Tako je I Klarkov nivo melanom in situ, II Klarkov nivo je invazija papilarnog sloja derma, III -invazija papilarnog i retikularnog sloja derma, IV- invazija retikularnog sloja derma, V-invazija supkutanog tkiva. Prema Klarku petogodišnje preživljavanje redom za stadiume I je veće od 95%, za stadium II 95%, za III 81%, za IV 68%, i V 47%.

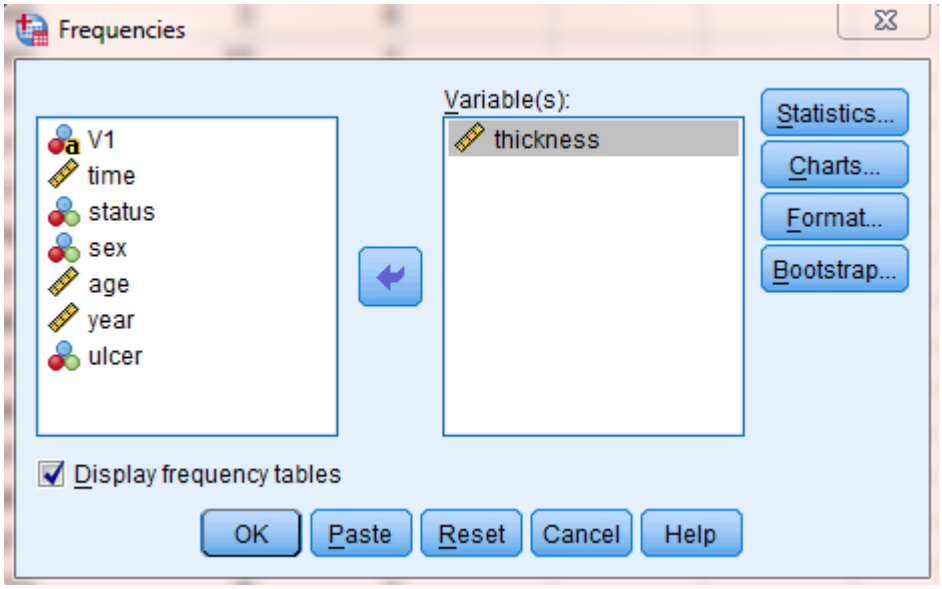
Aktuelna AJCC klasifikacija melanoma i očekivano preživljavanje bolesnika

Stadijum	Kriterijum klasifikacije			Petogodišnje preživljavanje
0	pTis	N0	M0	
IA	pT1	N0	M0	>95%
IB	pT2	N0	M0	>85%
IIA	pT3	N0	M0	>70%
IIB	pT4	N0	M0	50%
IIIA	svaki pT	N1	M0	<40%
IIIB	svaki pT	N2	M0	<30%
IVA	svaki pT	N2	M1a	<5–10%
IVB	svaki pT	svaki N	M1b	<5%

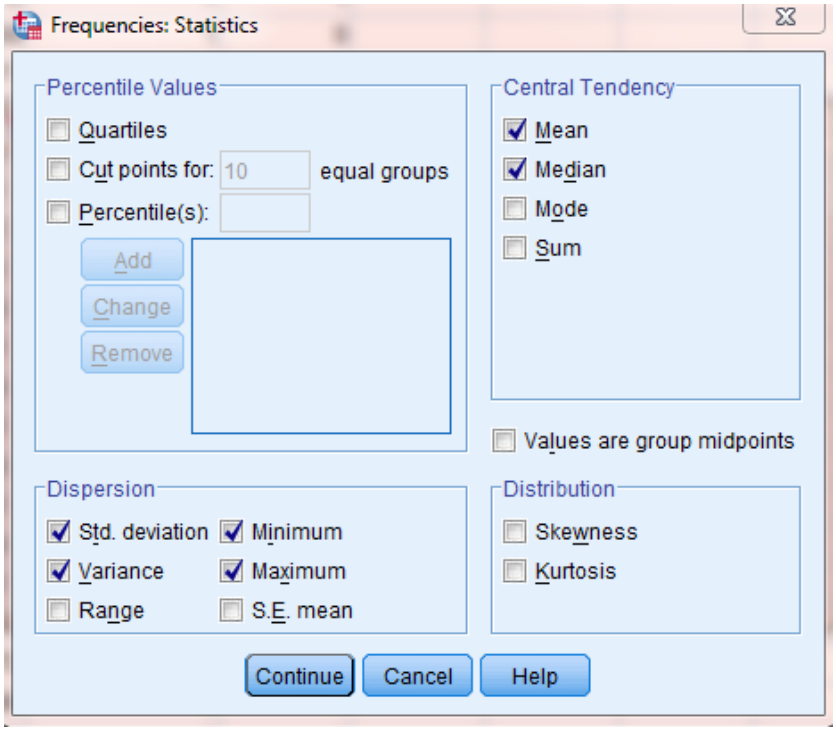
Prema Eggermontu (42)

Sada ćemo na osnovu naših podataka ispitati minimalnu, maksimalnu, srednju vrednost debljine tumora ispitanika.

Analyze→ Descriptive Statistic→Frequencies



Klikom na Statistic štikliramo Min,Max,Mean,Median, Sd, Variance i Continue. Čekiramo i display frequencies table.



Statistics

thickness

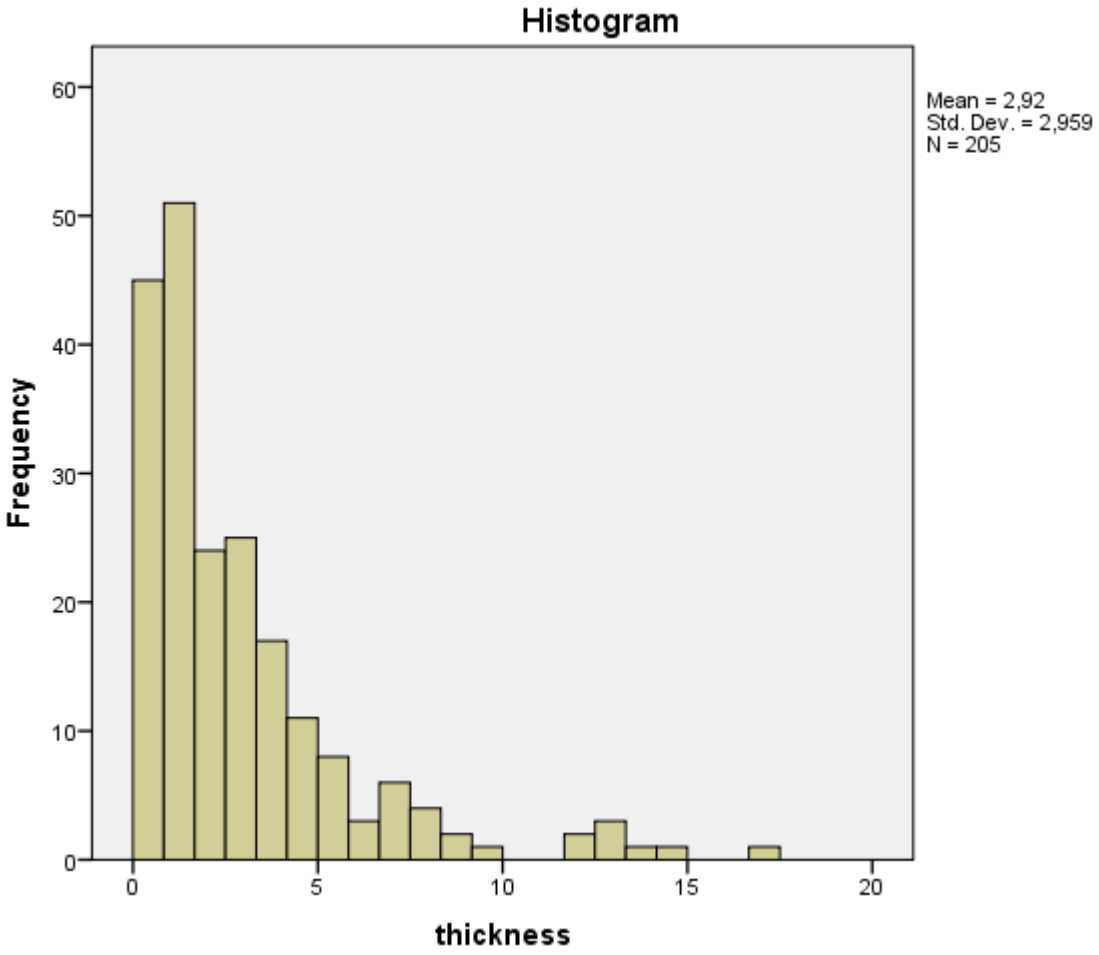
N	Valid	205
	Missing	0
Mean		2.92
Median		1.94
Std. Deviation		2.959
Variance		8,758
Minimum		0
Maximum		17

Vidimo da je najmanja debljina tumora 0mm, najveća 17mm (V stadium), standardna devijacija 2.959 a srednja vrednost 2.92 što pripada pt3 odnosno IV Klarkovom nivou gde je procenat petogodišnjeg preživljavanja 87%, a desetogodišnjeg 57%.

Crtamo odgovarajući histogram za ovo obeležje.

Analyze→ Descriptive Statistic→Frequencies

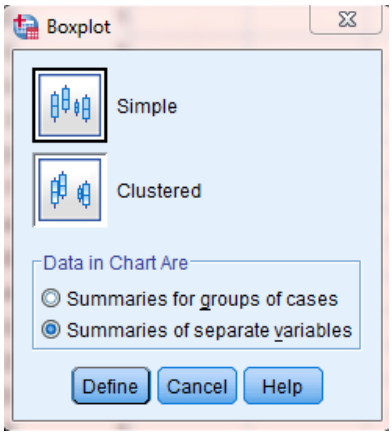
Odabereme thickness u prozoru Variable, i kliknemo na Charts pa zatim na Histograms.

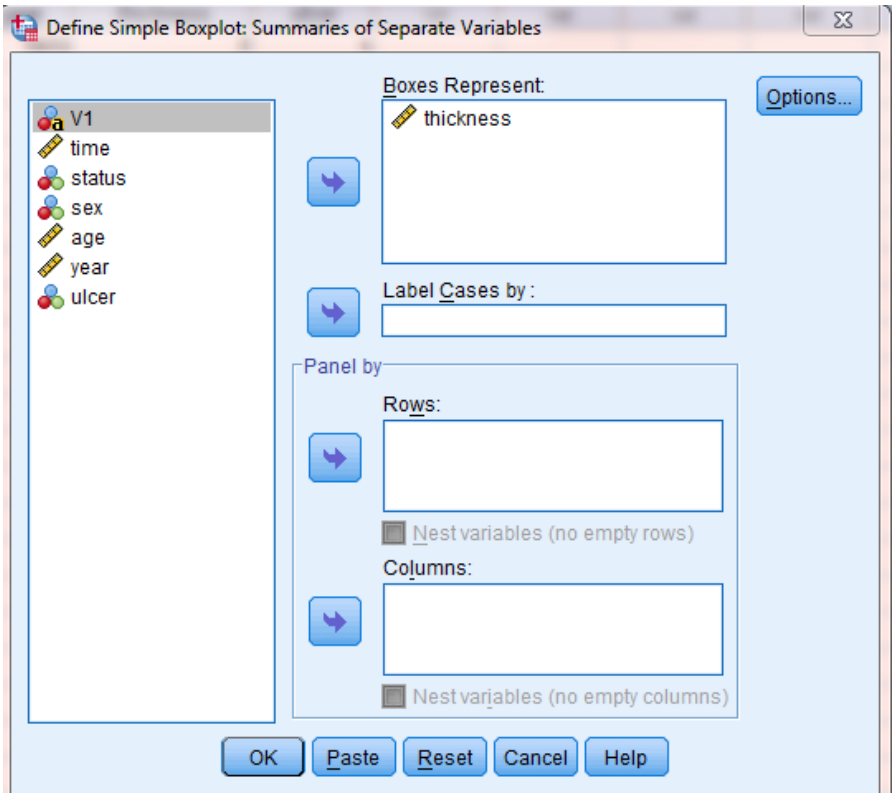


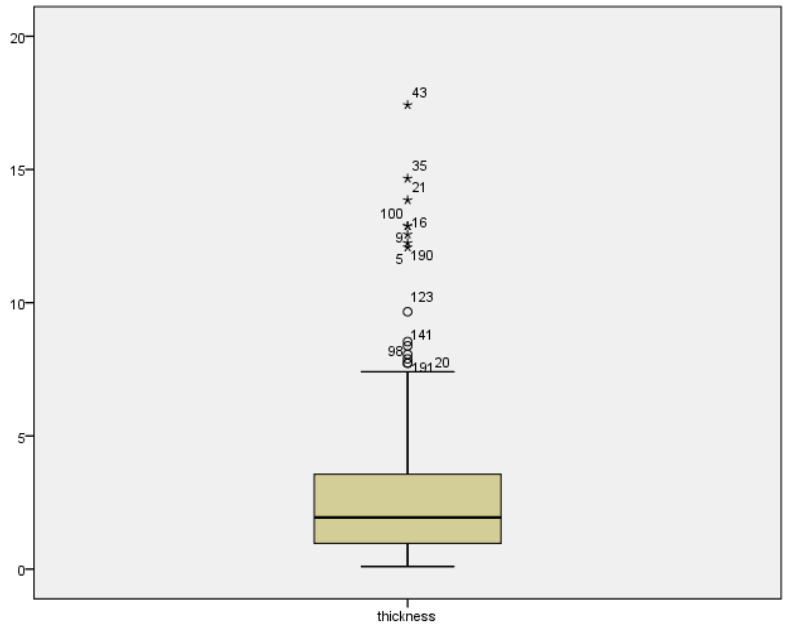
Sa histograma vidimo da najviše ispitanika ima tumor debljine između 0-2 mm.

Možemo nacrtati i Box-Plotove za debljinu tumora.

Graphs →Legacy Dialogs →Boxplot

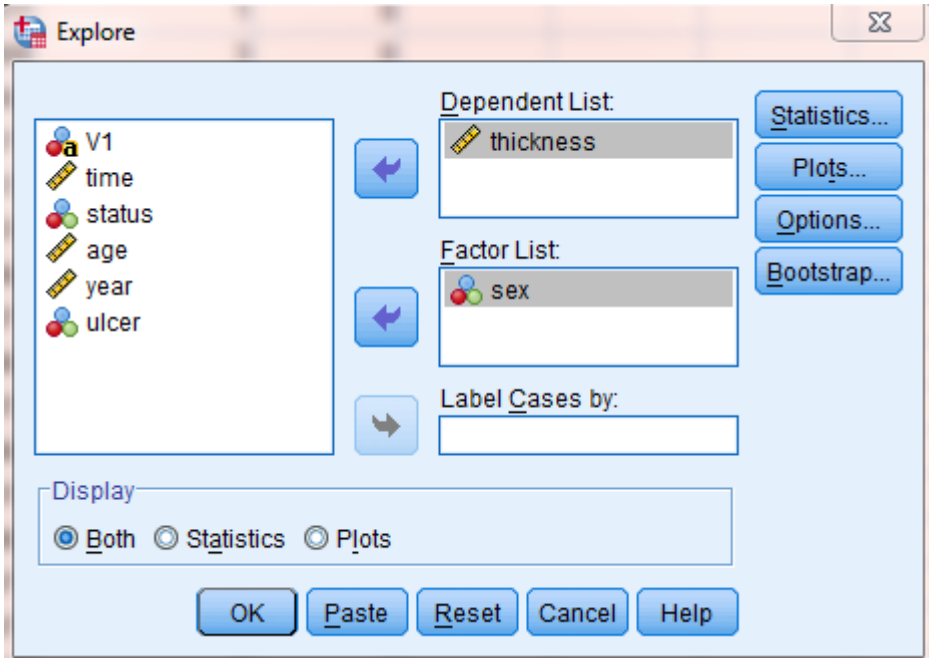


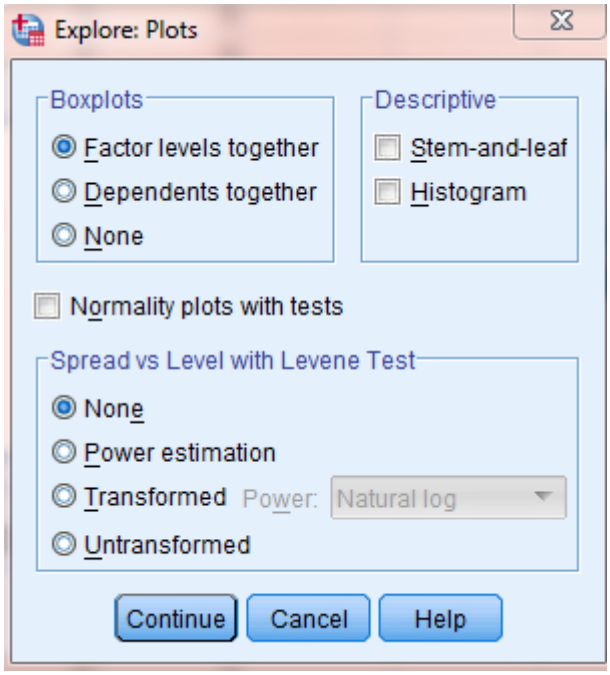




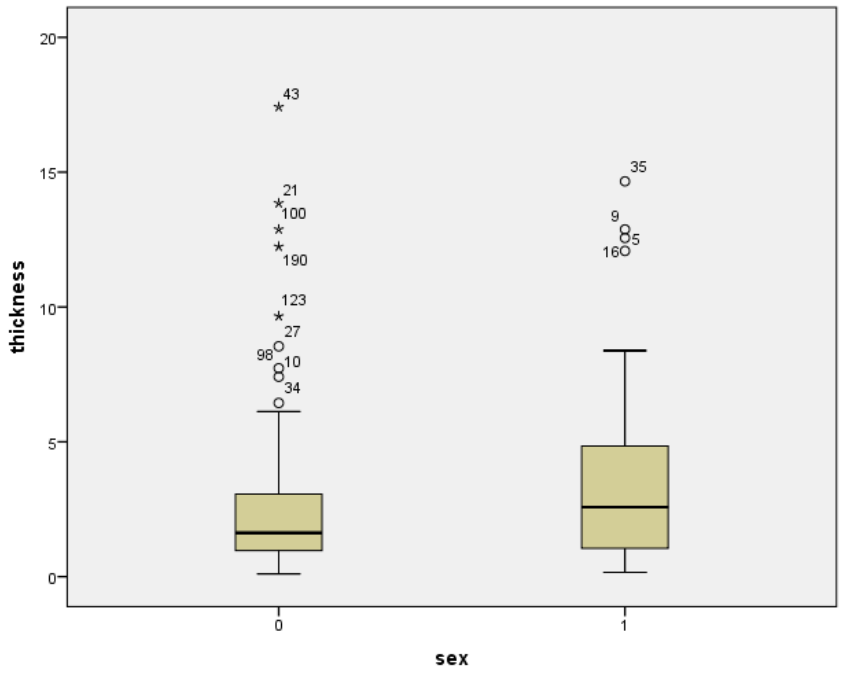
Imamo dosta autlajera, srednja vrednost je oko 2-3 mm što smo videli i na histogramu. Možemo prikazati Boxplotove posebno za muškarce a posebno za žene.

Analyze→ Descriptive Statistic→ Explore


Explore
V1
time
status
age
year
ulcer
Dependent List:
thickness
Factor List:
sex
Label Cases by:
Statistics...
Plots...
Options...
Bootstrap...
Display
Both
Statistics
Plots
OK
Paste
Reset
Cancel
Help



thickness



Vidimo da je srednja vrednost debljine tumora kod žena manja oko 2mm, a kod muškaraca veća oko 3 mm, kao i da više autlajera ima kod žena nego kod muškaraca.

Proveravamo da li su vrednosti obeležja thickness normalno raspoređene.

Analyze→ Descriptive Statistic→ Explore

U otvorenom prozoru kliknemo na Plots, i čekiramo Normality plots with tests.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	sex	Statistic	df	Sig.	Statistic	df	Sig.
thickness	0	,205	126	,000	,676	126	,000
	1	,137	79	,001	,863	79	,000

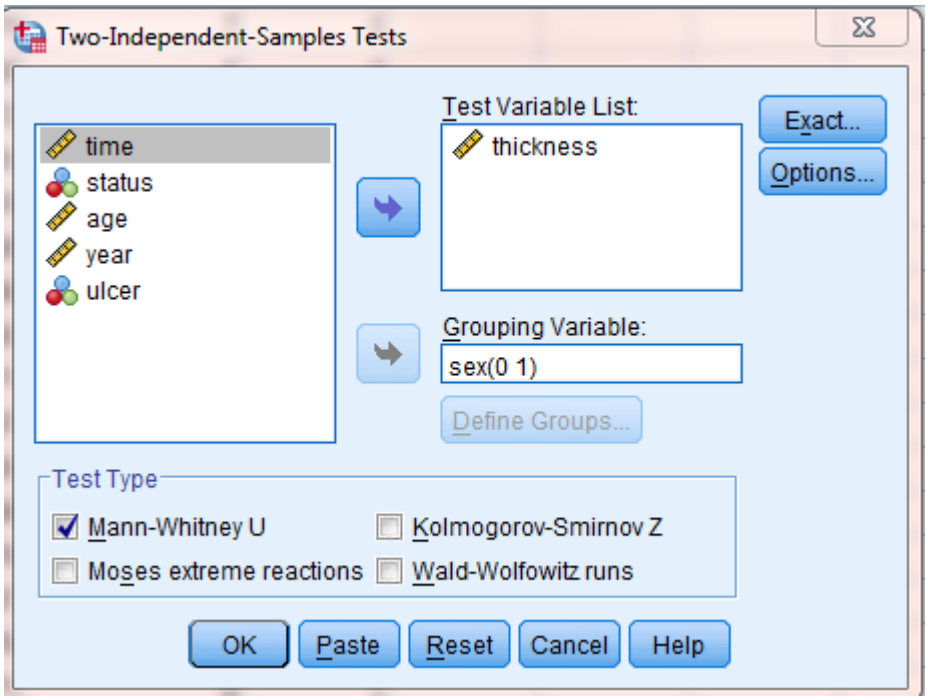
a. Lilliefors Significance Correction

P vrednost testa je manja od 0.001 zato se prikazuje u tabeli .000. Znači p vrednost testa je manja od 0.05 pa možemo zaključiti da vrednost obeležja thickness nije normalno raspodeljena. Za ispitivanje normalnosti ovde se koristio Kolmogorov-Smirnov test(za veliki obim uzorka) i Šapiro-Vilk(obim uzorka do 2000).

Narušene su prepostavke o normalnosti a postoje i autilajeri, pa nećemo moći primeniti t-testove i anovu.

Najpogodniji test kada ne važi pretpostavka o noralnosti je Mann Whutney. On poredi 2 promenljive na osnovu medijana.

Analyze→ Nonparametric Test→ Legacy dialogs→ 2 independet samples



Mann-Whitney Test

Ranks

	sex	N	Mean Rank	Sum of Ranks
thickness	0	126	93,62	11795,50
	1	79	117,97	9319,50
	Total	205		

Test Statistics[a]

	thickness
Mann-Whitney U	3794,500
Wilcoxon W	11795,500
Z	-2,863
Asymp. Sig. (2-tailed)	,004

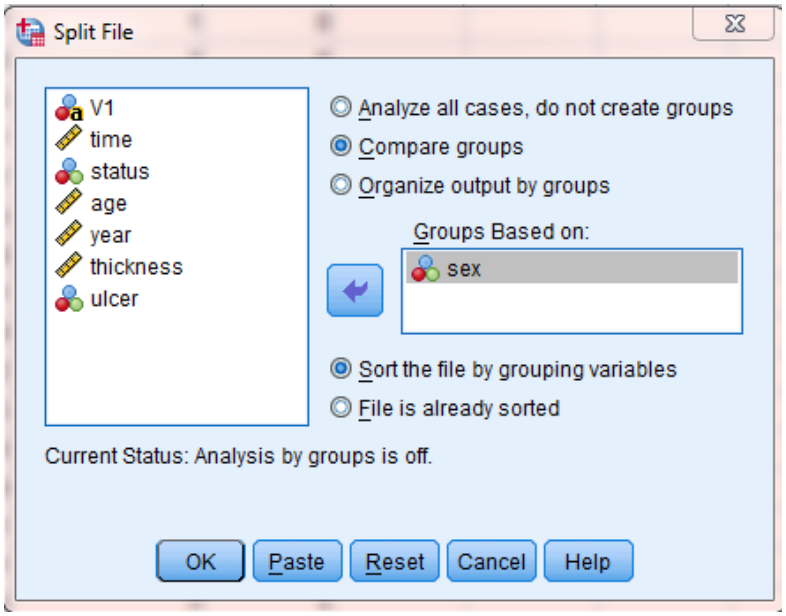
a. Grouping Variable: sex

Vidimo da je p vrednost testa 0.004 što je manje od 0.05. Nultu hipotezu odbacujemo, tj. debljina tumora ne zavisi od pola ispitanika.

Lečenje

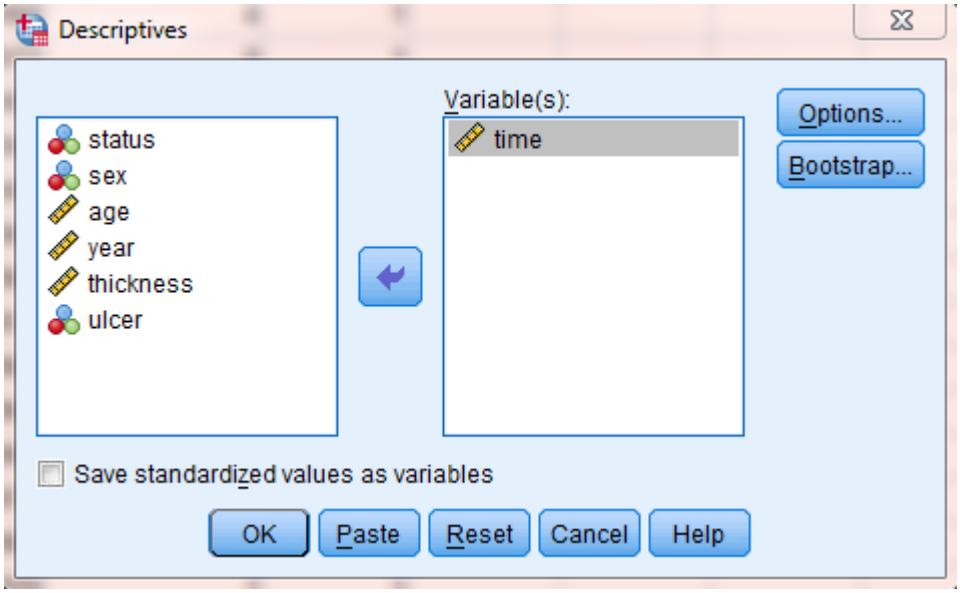
Prvo ćemo proveriti koliko se u principu leče žene a koliko muškarci. Za izvršavanje ovog zadatka potrebno je prvo spilitovati podatke.

Data →Split file



Sada smo splitovali fajl pa možemo da računamo.

Analyze→ Descriptive Statistics →Descriptives



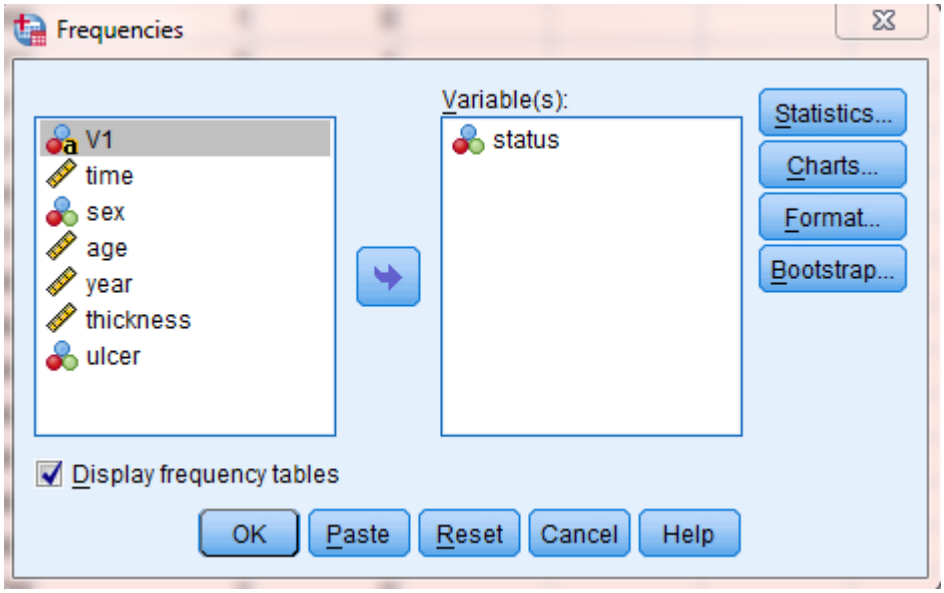
Descriptive Statistics

sex		N	Minimum	Maximum	Mean	Std. Deviation
0	time	126	99	5565	2282,64	1089,818
	Valid N (listwise)	126				
1	time	79	10	4492	1945,71	1148,382
	Valid N (listwise)	79				

Primećujemo da se žene duže leče, kod njih je minimum 99 dana lečenja dok je kod muškaraca samo 10. Takođe i maksimum je veći kod žena nego kod muškaraca. Srednja vrednost lečenja u danima kod žena je 2283 dana, a kod muškaraca 1946 dana. Zaključujemo da se žene duže leče ali da li to ima uticaja na stopu smrtnosti kod žena i muškaraca, kao posledica bolesti melanoma?

Ispitujemo kojoj kategoriji Statusa dužine lečenja pripada najviše ispitanika?

Analyze→ Descriptive Statistic→ Frequecies



status

sex			Frequency	Percent	Valid Percent	Cumulative Percent
0	Valid	1	28	22,2	22,2	22,2
		2	91	72,2	72,2	94,4
		3	7	5,6	5,6	100,0
		Total	126	100,0	100,0	
1	Valid	1	29	36,7	36,7	36,7
		2	43	54,4	54,4	91,1
		3	7	8,9	8,9	100,0
		Total	79	100,0	100,0	

Kod žena najviše njih prežive - 91, 28 žena umire od bolesti a 7 zbog nekih drugih faktora.

Kod muškaraca najviše njih preživi - 43, 29 umire zbog bolesti a 7 zbog nekog drugog fatora.

Kako je ispitano mnogo više žena nego muškaraca a broj umrlih od melanoma je sličan kod muškaraca (29) i kod žena (28), možemo zaključiti da je to posledica dužeg lečenja žena nego muškaraca (što smo ispitivali u prethodnom odeljku). Znači žene se u princiu duže leče i imaju duži životni vek.

Ispitaćemo i zavisnot dužine lečenja bolesti (zavisna promenljiva) u odnosu na pol (nezavisna promenljiva) uz pomoć One-way ANOVA testa.

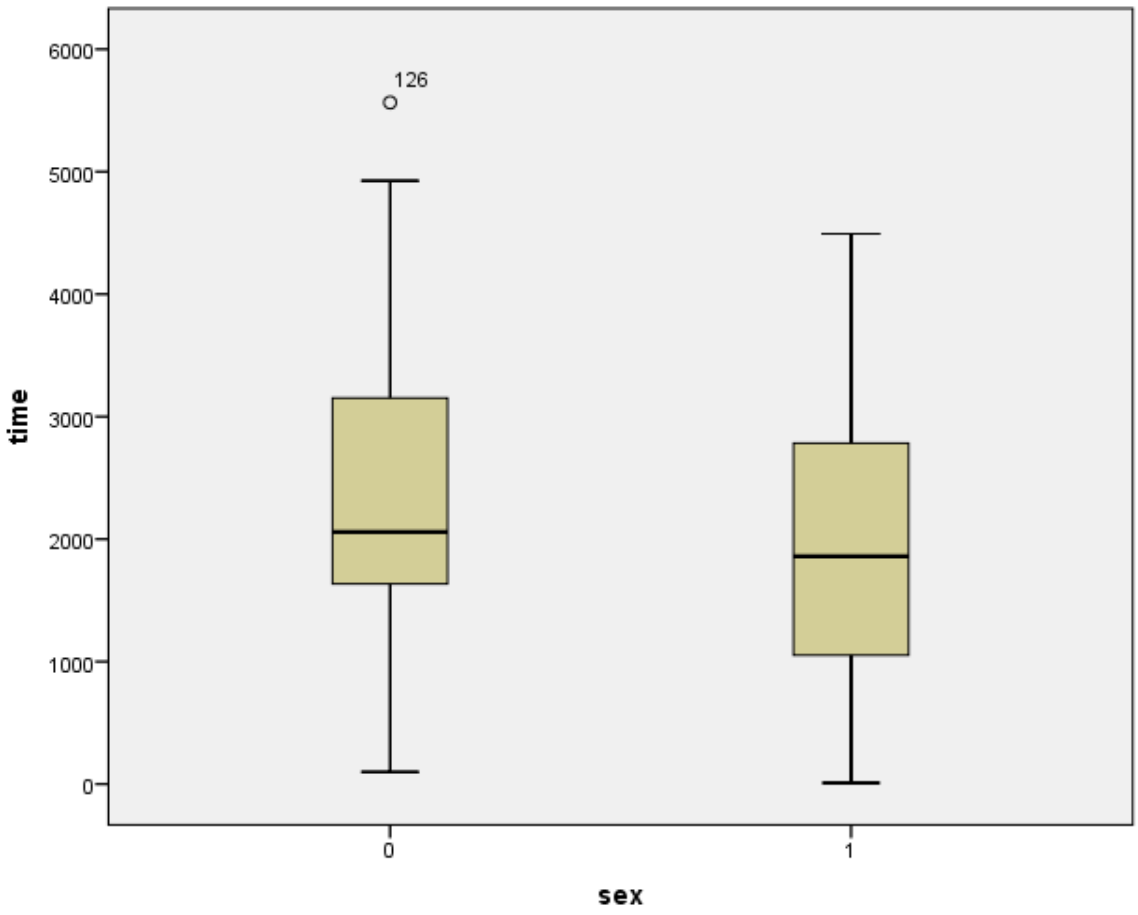
Uslovi koje treba ispuniti za ovaj test su:

-zavisna promenljiva je neprekidna (dužina lečenja u danima)

-faktor se sastoji od 2 ili više nezavisnih grupa(0-žena, 1-muškarac)

-ispitujemo autlajere: (narušena pp, postoji jedan autlajer kod žena prilikom 126 opservacije)

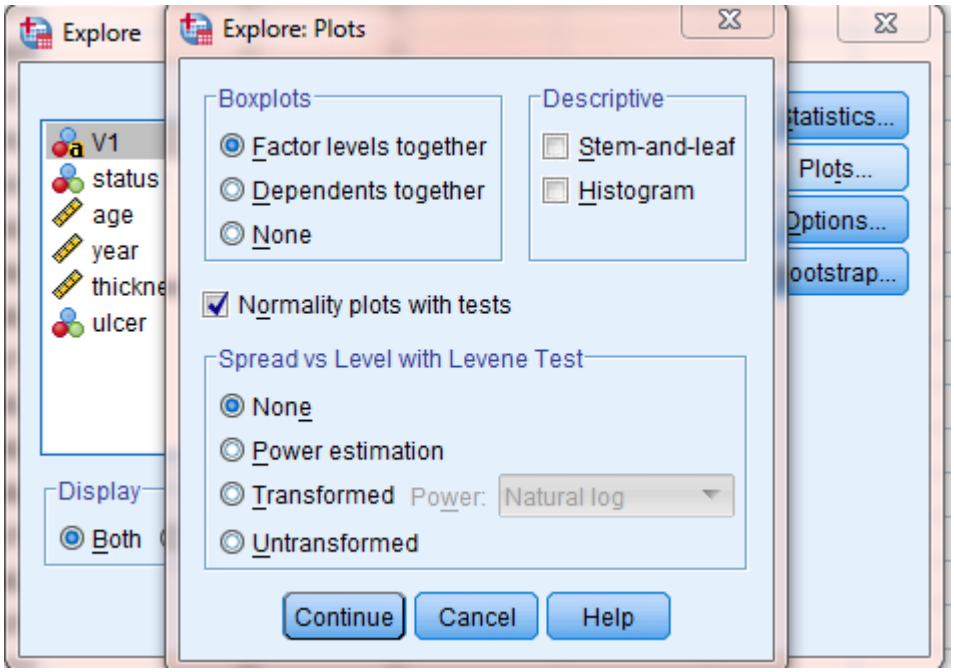
Analyze→ Descriptive Statistic→ Explore…



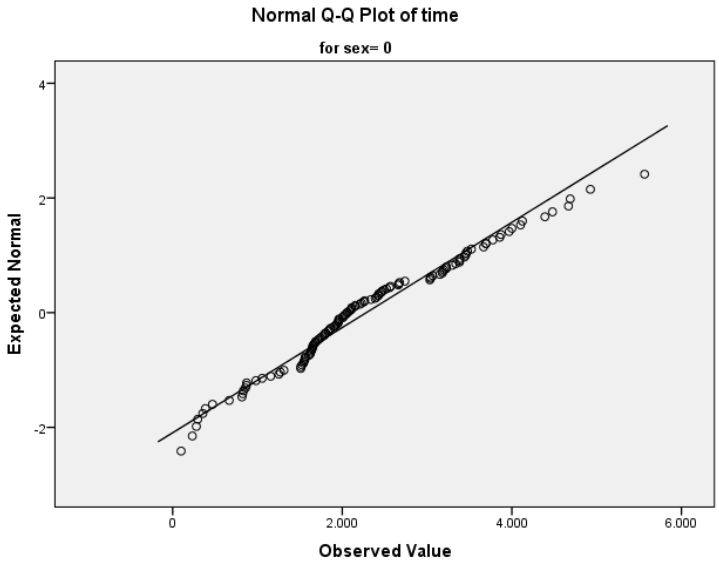
-zavisna promenljiva treba da ima bar aproksimalno normalnu raspodelu po grupama

Proveravamo da li dužina lečenja bolesti ima normalnu raspodelu u odnosu na pol.

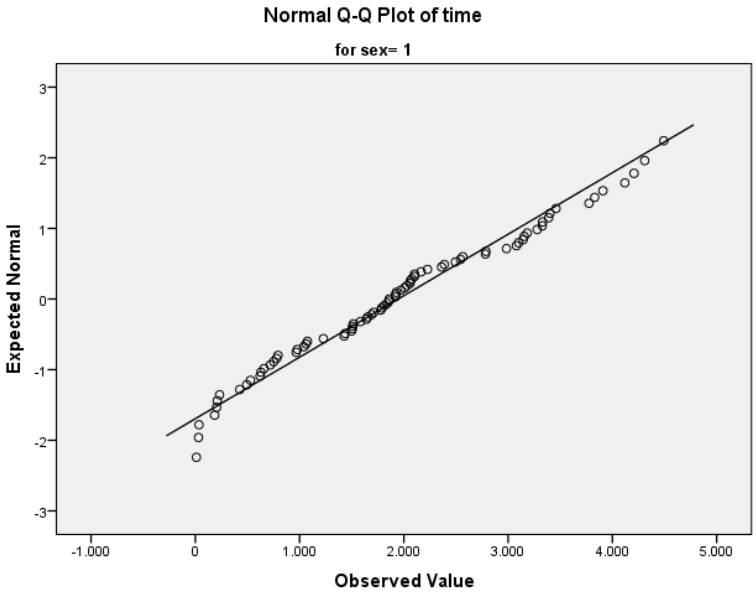
Analyze→Descriptive Statistic→Explore



Za žene:



Za muškarce:



Kriva na ovom grafiku se sastoji od tačaka koje se dobijaju oduzimanjem očekivanog od posmatranog kvantila. X-osa predstavlja empirijske vrednosti kvantila a y-osa očekivane vrednosti kvantila iz normalne raspodele. Ukoliko je raspodela normala tačke treba da se grupišu oko prave y=x. Vidimo da su tačke kod osoba muškog pola približnije ovoj liniji, ali i kod osoba ženskog pola ne štrče mnogo van linije. Sa ovih grafika možemo pretpostaviti da vreme lečenja u odnosu na pol ima normalnu raspodelu. Proverićemo i testovima za normalnost.

Tests of Normality

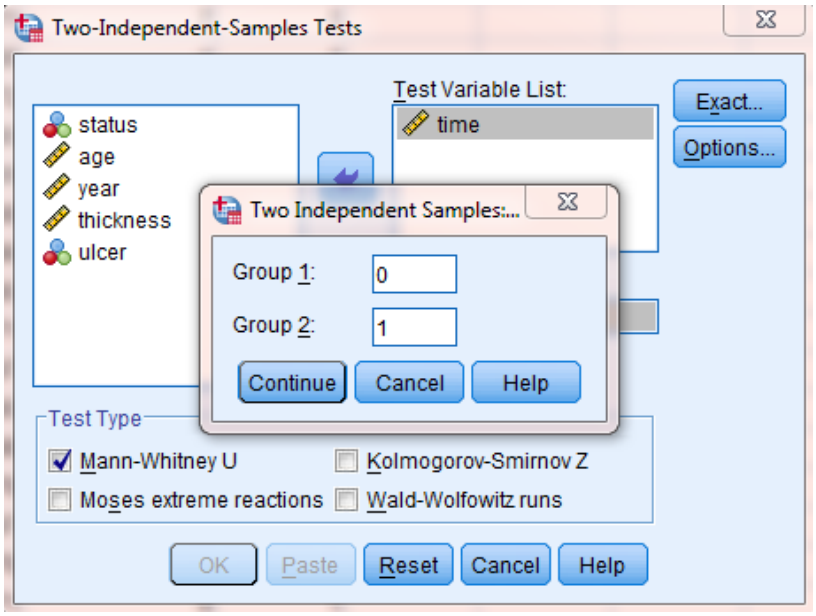
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	sex	Statistic	df	Sig.	Statistic	df	Sig.
time	0	,102	126	,003	,974	126	,015
	1	,091	79	,163	,971	79	,068

a. Lilliefors Significance Correction

Vrednost testa kod žena je 0.003 (Kolmogorov-Smirnov) i 0.015 (Šapiro-Vilk). Obe vrednosti su manje od 0.05 što nam ukazujem da odbacujemo nultu hipotezu koja glasi da je raspodela normalna. Dakle kod žena ne važi normalna raspodela, dok kod muškaraca važi (vrednosti su veće od 0.05 I prihvatamo nultu hipotezu).

Pošto je narušena normalnost ne možemo koristiti navedeni test. Najbolje rešenje kada ne važi normalna raspodela je Mann-Whutney test, koji poredi dve promenljive na osnovu medijana. Prvo otsplitujemo podatke a zatim:

Analyze→ Nonparametric Test→ Legacy dialogs→ 2 independet samples



Mann-Whitney Test

Ranks

	sex	N	Mean Rank	Sum of Ranks
time	0	126	109,73	13825,50
	1	79	92,27	7289,50
	Total	205		

Test Statistics[a]

	time
Mann-Whitney U	4129,500
Wilcoxon W	7289,500
Z	-2,050
Asymp. Sig. (2-tailed)	,040

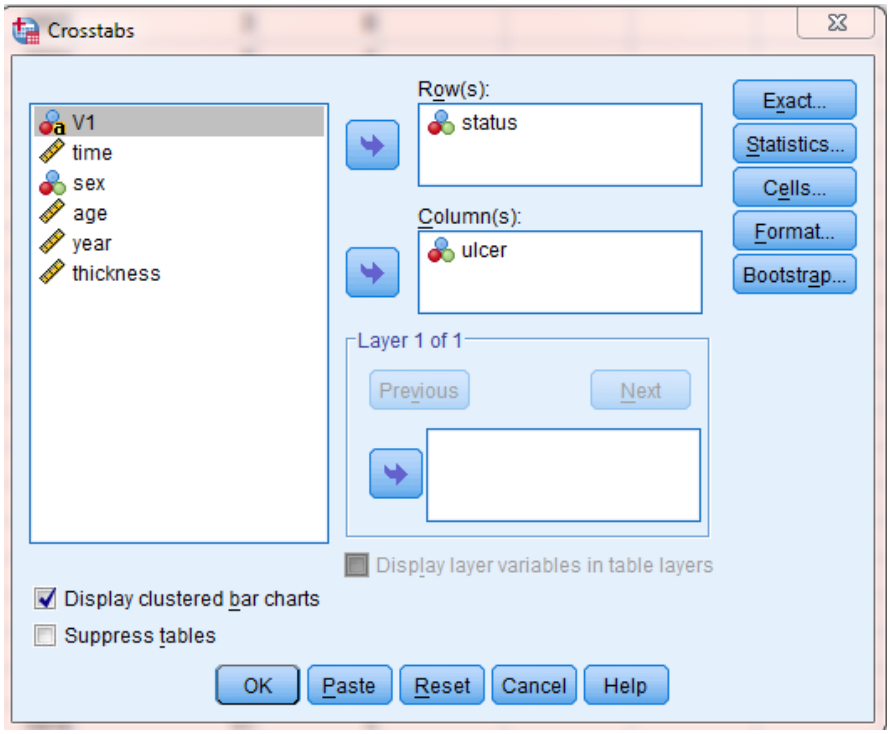
a. Grouping Variable: sex

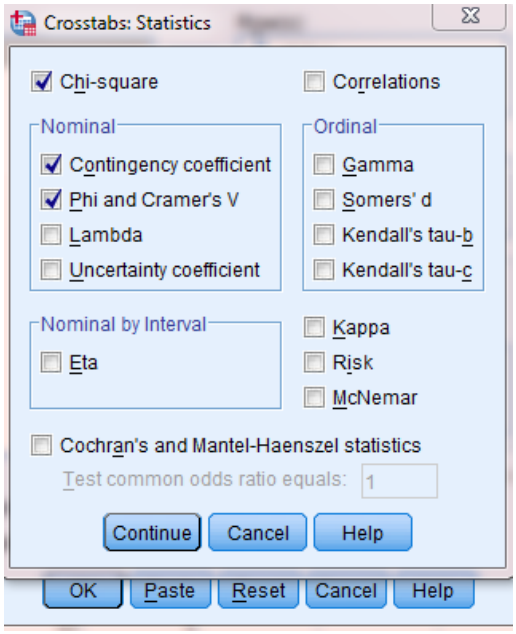
Vrednost testa je 0.04 što je blizu 0.05 (malo manje). Ovde odbacujemo nultu hipotezu, tj. možemo reći da dužina lečenja bolesti ne zavisi od pola ispitanika. (Logičnom diskusijom pre ovog ispitivanja smo zaključili da se žene duže leče od muškaraca, ali testom je pokazano da dužine lečenja melanoma kod žena i muskaraca nije povezano).

Ispitivanje zavisnosti preživljavanja od bolesti melanoma i postojanje čira

Da bismo ispitali zavisnost postojanja čira i preživljavanje koristimo hi-kvadrat test, pošto su obeležja kategorička. Hi-Kvadrat test nezavisnosti ispituje da li su obeležja nezavisna, pa je nulta hipoteza da su obeležja nezavisna, a alternativna da nisu. Podaci obeležja se svrstaju u tabelu kontigencije.

Analyze→ Descriptive Statistic→ Crosstabs





Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
status * ulcer	205	100,0%	0	0,0%	205	100,0%

status * ulcer Crosstabulation

Count

		ulcer		Total
		0	1	
status	1	16	41	57
	2	92	42	134
	3	7	7	14
Total		115	90	205

Najviše je onih koji su živi i nemaju čir (92), pa zatim oni koji su živi i imaju čir (42), mrtvi i imali čir (41). Pa na osnovu ovoga možemo naslutiti da postoji neka veza ali i da ona nije baš jaka.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	26,974[a]	2	,000
Likelihood Ratio	27,406	2	,000
Linear-by-Linear Association	14,907	1	,000
N of Valid Cases	205		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 6,15.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,363	,000
	Cramer's V	,363	,000
	Contingency Coefficient	,341	,000
N of Valid Cases		205	

U tabelei Chi-Square tests nema prekoračenja što je dobro. Vrednost test-statistike Hi-Kvadrat testa jeste vrednost iz prve vrste Pearson Chi-Square i iznosi 26,974, dok je p-vrednost manja od 0.01. Nivo značajnosti je 0.05 što znači da odbacujemo nultu hipotezu koja glasi da su obeležja nezavisna. Dakle, prihvatamo alternativnu hipotezu da su obeležja zavisna. Phi I Cramers V vrednosti su vrednosti koeficijenata korelacije i iznose 0.363 što je malo ispod umerene veze. Veza nije baš jaka ali nije ni mnogo slaba. Contingency Coefficient je koeficijent kontigencije i iznosi 0.341, dakle nije nešto jaka zavisnot između naših promenljivih, ali svakako postoji.

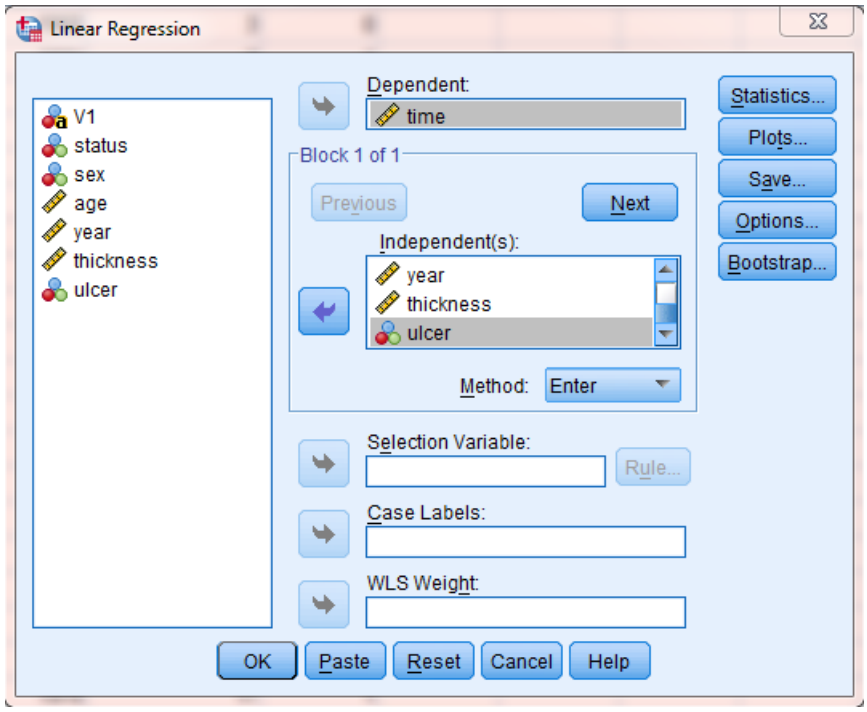
Pravljenje najboljeg linearnog model za određivanje vremena lečenja bolesti u odnosu na ostale promenljive

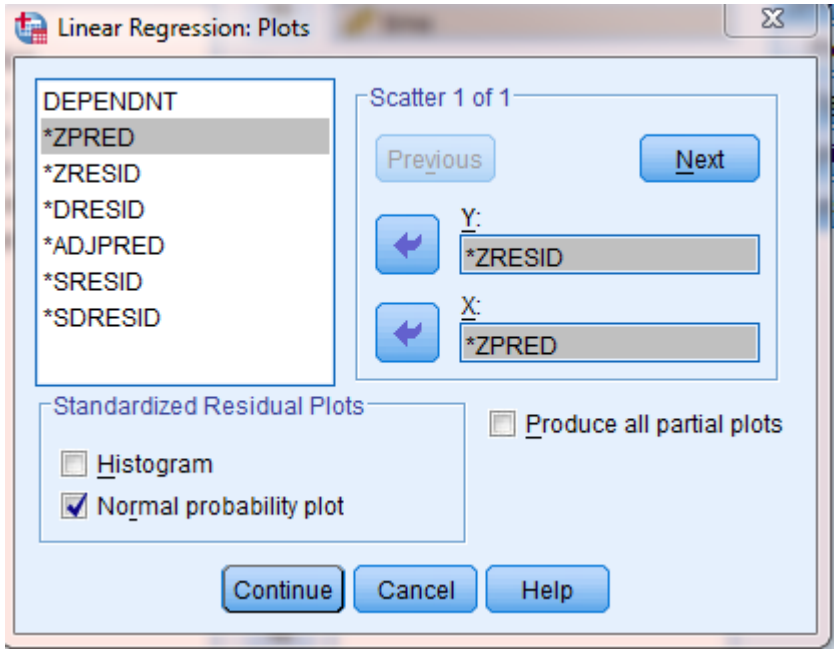
Da bi rešili ovaj problem primenjujemo linearnu regresiju. Za njenu primenu potrebno je da budu ispunjeni sledeći uslovi:

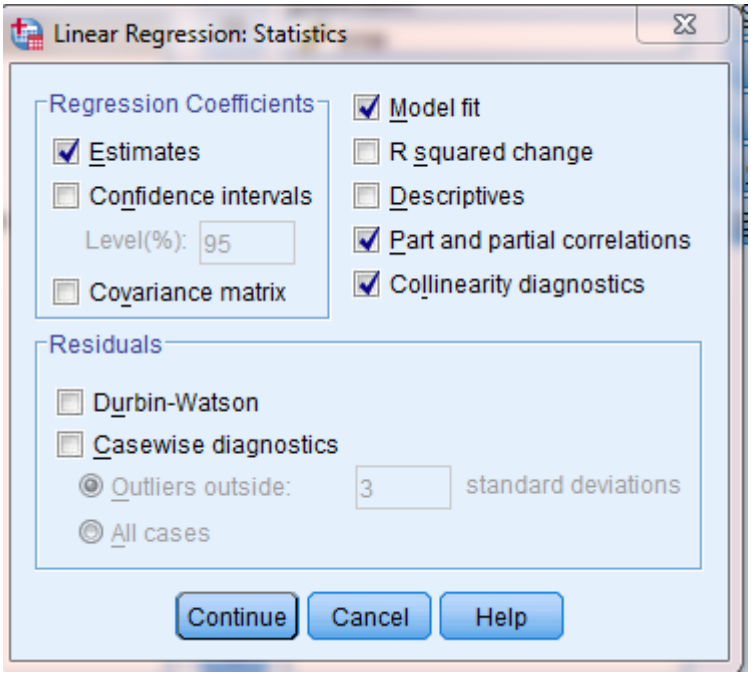
- Opservacije su nezavisne
- Greška ima normalnu raspodelu sa očekivanjem 0 i konstantnom disperzijom
- Greške su međusobno nekorelisane
- Broj podataka zanačajno je veći od broja parametara koji se ocenjuju

- Mora postojati linearna zavisnost između zavisne i bilo koje nezavisne promenljive

Analyze→ Regression→ Linear







Pri pokretanju dobijamo:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,690[a]	,477	,461	823,834

a. Predictors: (Constant), ulcer, year, sex, age, status, thickness

b. Dependent Variable: time

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	122457150,1	6	20409525,01	30,071	,000[b]
	Residual	134382958,7	198	678701,812		
	Total	256840108,8	204			

a. Dependent Variable: time

b. Predictors: (Constant), ulcer, year, sex, age, status, thickness

Na osnovu tabele zaključujemo da sve nazavisne promenljive objašnjavaju 47,7 % diperzije zavisne proemnljive(R square), što znači da je jačina veze jaka.

Koeficijent korelacije (R)-pokazuje linearnu korelaciju između vrednosti zavisne promenljive i regresijom predviđene vrednosti.

Koeficijent determinacije(R^2)-meri jačinu veze između zavisnih i nezavisnih promenljivih i predstavlja proporciju ukupnog varijabiliteta zavisne promenljive koja je objašnjena varijacijama nezavisne promenljive.

$$R^2 = \frac{objasnjeni\ varijavilitet}{ukupan\ varijabilitet}$$

Krigovani koeficijent determinacije-koeficijent determinacije korigovan prema broju nezavisnih promenljivih I veličini uzorka:

$$\hat{R}^2 = 1 - \frac{N-1}{N-k-1}(1 - R^2).$$

N-veličina uzorka

k-broj nezavisnih promenljivih

Posmatrajmo i koeficijente:

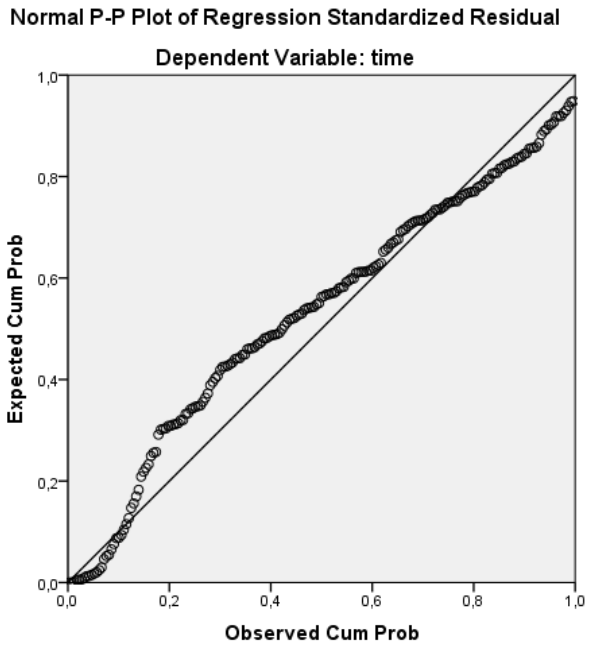
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	451853,225	45907,186		9,843	,000					
	status	666,505	110,346	,327	6,040	,000	,316	,394	,310	,899	1,112
	sex	-137,722	121,117	-,060	-1,137	,257	-,146	-,081	-,058	,953	1,049
	age	-10,722	3,640	-,159	-2,946	,004	-,302	-,205	-,151	,903	1,107
	year	-228,443	23,332	-,524	-9,791	,000	-,486	-,571	-,503	,921	1,085
	thickness	-57,327	22,457	-,151	-2,553	,011	-,235	-,179	-,131	,753	1,328
	ulcer	-223,965	131,685	-,099	-1,701	,091	-,265	-,120	-,087	,775	1,290

a. Dependent Variable: time

Za naš model značajni su samo oni koeficijenti čija je p-vrednost manja od 0.05. Iz tbele zaključujemo da promenljive sex i ulcer nisu značajne za naš model.

Proveravamo i normalnost reziduala.



Dijagram normalne raspodele za standardizovane reziduale ima približno normalnu raspodelu.

Model koji smo dobili je sledeći:

Y=451853,2+666,05X_1-137,7X_2-10,722X_3-228,4X_4-57,3X_5-223,9X_6

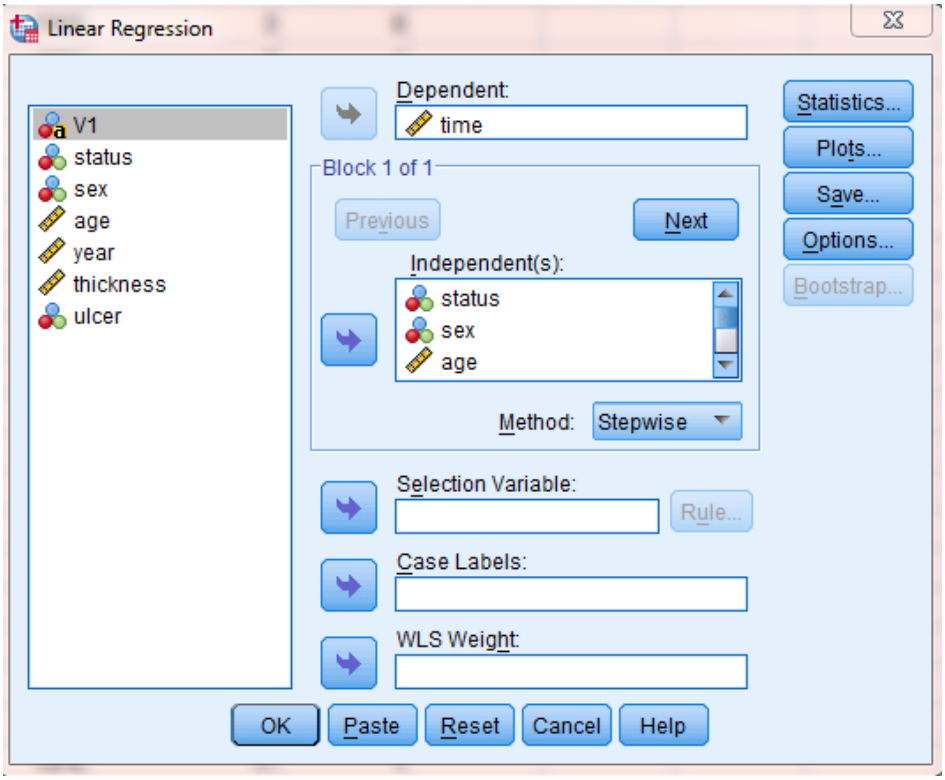
Ovo nije najbolji model jer promenljiva sex nije značajna za naš model. Najbolji model dobijamo izbacivanjem promenljive sex i ulcer.

Možemo pozvati i metodu Stepwise.

Forward-postepeno uključivanje promenljivih u model na osnovu p-vrednosti F-testa.

Backward-iz modela sa svim nezavisnim promenljivim se izbacuju jedna po jedna promenljiva na osnovu p-vrednosti F-testa.

Stepwise-postepeno se dodaju promenljive u model, pri čemu se u svakom koraku proverava p-vrednost F-testa. Ovo je kombinacija prethodne 2 metode.



Model	Variables Entered	Variables Removed	Method
1	year	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
2	status	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
3	thickness	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
4	age	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).

a. Dependent Variable: time

Dobili smo isti rezuletat, tj. za naš model nisu značajne sex i ulcer koje su izbačene. Iz sledeće tabele takođe uočavamo koje promenljive treba izbaciti, one koje imaju p-vrednost testa veću od 0.05.

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	status	,391[b]	7,015	,000	,443	,981	1,019	,981
	sex	-,148[b]	-2,438	,016	-,169	1,000	1,000	1,000
	age	-,218[b]	-3,587	,000	-,245	,965	1,037	,965
	thickness	-,306[b]	-5,249	,000	-,346	,982	1,018	,982
	ulcer	-,281[b]	-4,824	,000	-,321	,999	1,001	,999
2	sex	-,110[c]	-2,006	,046	-,140	,990	1,010	,971
	age	-,214[c]	-3,941	,000	-,268	,964	1,037	,946
	thickness	-,240[c]	-4,427	,000	-,298	,947	1,056	,946
	ulcer	-,191[c]	-3,421	,001	-,235	,927	1,079	,910
3	sex	-,072[d]	-1,342	,181	-,095	,961	1,041	,919
	age	-,165[d]	-3,039	,003	-,210	,906	1,104	,889
	ulcer	-,113[d]	-1,903	,059	-,133	,783	1,278	,783
4	sex	-,068[e]	-1,283	,201	-,091	,960	1,042	,866
	ulcer	-,105[e]	-1,805	,073	-,127	,781	1,280	,764

a. Dependent Variable: time

b. Predictors in the Model: (Constant), year

c. Predictors in the Model: (Constant), year, status

d. Predictors in the Model: (Constant), year, status, thickness

e. Predictors in the Model: (Constant), year, status, thickness, age

To su naravno sex i ulcer.

Ovom metodom smo dobili isti rezultat kao i metodom Enter(standardna linearna regresija).

Liteartura

- http://www.math.rs/p/marija-radicevic/kurs/324/%D0%A1%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%BA%D0%B8-%D1%81%D0%BE%D1%84%D1%82%D0%B2%D0%B5%D1%80-3/
- https://vincentarelbundock.github.io/Rdatasets/datasets.html
- https://stat.ethz.ch/R-manual/R-devel/library/MASS/html/Melanoma.html