

Linearni regresioni model u SPSS softverskom paketu za Windows

Prof.dr. Nahod Vuković

Prof.dr. Milica Delić

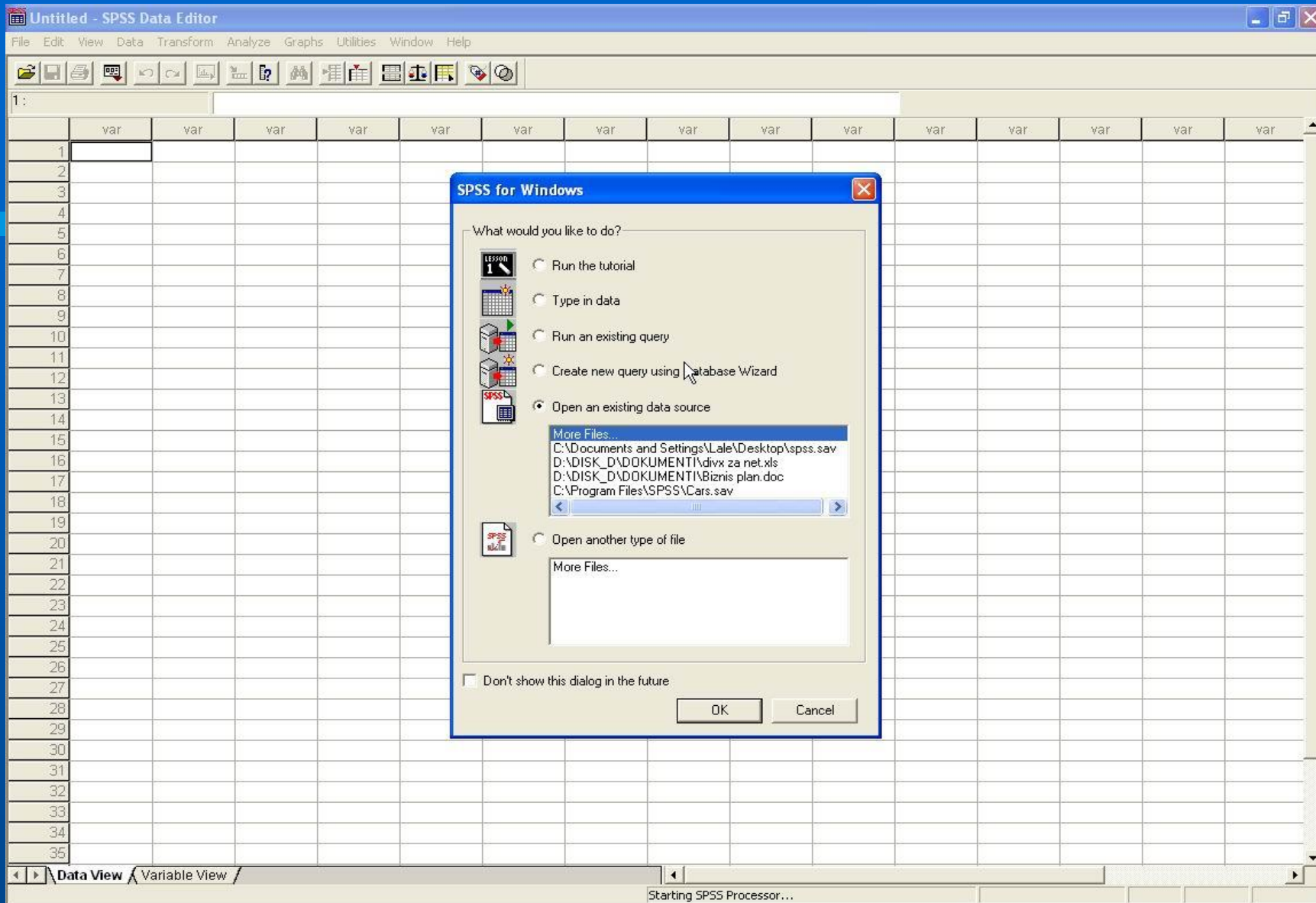
PLAN IZLAGANJA

- I. Osnovni koncepti SPSS-a
- II. Rad sa podacima
- III. Primer u Excel-u
- IV. Linearna regresija
- V. Ocene nelinearne regresije
- VI. Osvrt na izlaganje
- VII. Korišćena literatura

I. Osnovni koncepti SPSS-a

- o **primena informacione tehnologije u statistickoj praksi**
- o **razvoj raznih softverskih paketa (SPSS, SAS, SYSTAT, STATISTICA)**

- o SPSS predstavlja jedan od najrasprostranjenijih softverskih paketa u svetu (<http://www.spss.com>)
- o primena u statističkoj obradi podataka
- o radi pod Microsoft Windows okruženjem



SPSS softverski paket

spss - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

2: ime Borko

	IDENTIFIKACIONI BROJ	prezime	broj_ind	fakultet	smer	godina	br_ispit	status	pol	starost	var	var
2	176	Borko	Mikic	361/00	FON	zajednicke osnove	II	12	samofinansiranje	muski	21	
3	177	Ana	Peric	210/02	FPN	zumalistika	I	5	budzet	zenski	20	
4	178	Sanja	Lazic	23/99	PMF	opsti	IV	30	budzet	zenski	23	
5	179	Mina	Jovanovic	245/99	ETF	automatika	III	15	budzet	zenski	20	
6	180	Marko	Petrovic	65/01	ETF	telekomunikacije	I	2	budzet	muski	23	
7	181	Nemanja	Jevtic	548/01	FON	upravljanje kvalitetom	II	10	budzet	muski	21	
8	182	Jelena	Isajev	321/00	SF	zeleznicki saobracaj	II	10	samofinansiranje	zenski	19	
9	183	Vojkan	Nikolic	19/00	FON	informacioni sistemi	III	17	samofinansiranje	muski	25	
10												
11												

36

Data View Variable View

spss - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	8	0	IDENTIFIKACIONI BROJ	None	None	8	Right	Scale
2	ime	String	10	0	IME	None	None	8	Left	Nominal
3	prezime	String	15	0	PREZIME	None	None	9	Left	Nominal
4	broj_ind	String	8	0	BROJ INDEKSA	None	None	6	Right	Nominal
5	fakultet	String	16	0	FAKULTET	None	None	8	Left	Nominal
6	smer	String	25	0	SMER	None	None	22	Left	Nominal
7	godina	String	3	0	GODINA STUDIJA	None	None	5	Center	Nominal
8	br_ispit	Numeric	8	0	BROJ POLOZENIH ISPITA	None	None	6	Center	Scale
9	status	String	20	0	STATUS	None	None	17	Left	Nominal
10	pol	String	10	0	POL	None	None	9	Left	Nominal
11	starost	Numeric	8	0	STAROST	None	None	6	Center	Scale
12										
13										

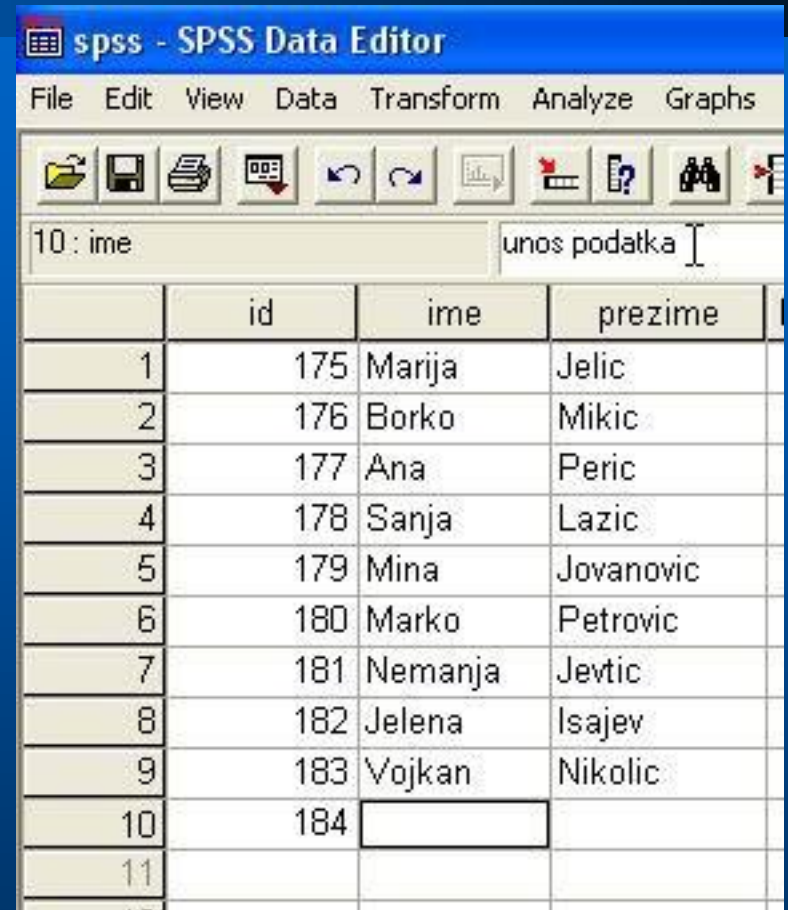
Data View Variable View

SPSS softverski paket

II. Rad sa podacima

Podaci se mogu unositi i menjati u

Data Editor-u, odnosno TextArea-i ispod toolbar-ova, kao i u Excel-u, ali se u njemu ne mogu vršiti izračunavanja ili unositi formule



spss - SPSS Data Editor

File Edit View Data Transform Analyze Graphs

10 : ime unos podatka

	id	ime	prezime
1	175	Marija	Jelic
2	176	Borko	Mikic
3	177	Ana	Peric
4	178	Sanja	Lazic
5	179	Mina	Jovanovic
6	180	Marko	Petrovic
7	181	Nemanja	Jevtic
8	182	Jelena	Isajev
9	183	Vojkan	Nikolic
10	184		
11			
12			

Čitanje file-a sa tabelom

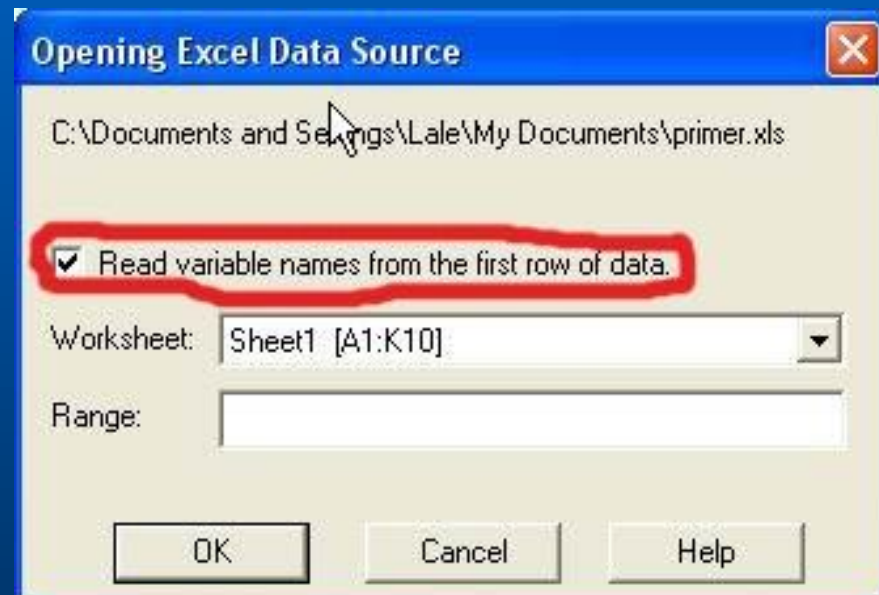
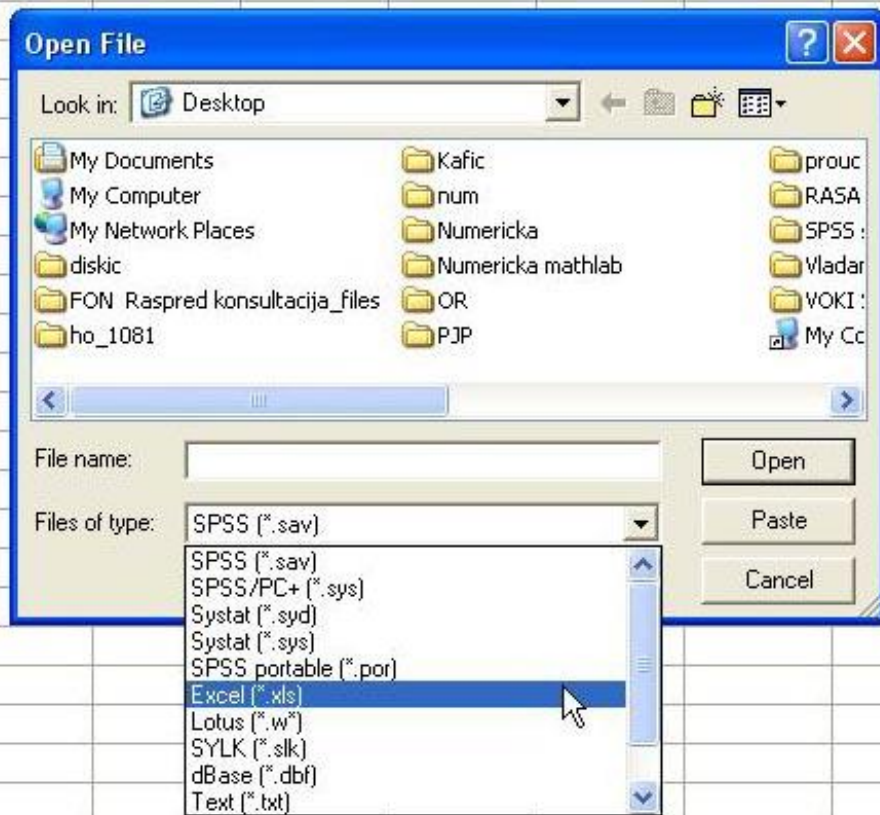
Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

A2 = 175

	A	B	C	D	E	F	G	H	I	J	K
1	id	ime	prezime	br_indexa	fakultet	smer	godina	br ispita	status	pol	starost
2	175	Marja	Jelic	546/99	FON	opsti menadzment	III	21	budzet	zensko	23
3	176	Borko	Mikic	361/00	FON	zajednicke osnove	II	12	samofinansiranje	muski	21
4	177	Ana	Peric	210/02	FPN	zurnalistika	I	5	budzet	zenski	20
5	178	Sanja	Lazic	23/99	PMF	opsti	IV	30	budzet	zenski	23
6	179	Mina	Jovanovic	245/99	ETF	automatika	III	15	budzet	zenski	20
7	180	Marko	Petrovic	65/01	ETF	telekomunikacije	I	2	budzet	muski	23
8	181	Nemanja	Jevtic	548/01	FON	upravljanje kvalitetom	II	10	budzet	muski	21
9	182	Jelena	Isajev	321/00	SF	zeleznicki saobracaj	II	10	samofinansiranje	zenski	19
10	183	Vojkan	Nikolic	19/00	FON	informacioni sistemi	III	17	samofinsiranje	muski	25
11											
12											
13											

Tip podataka za svaku promenljivu je definisan vrednostima u prvoj vrsti tabele



Microsoft Excel - primer

File Edit View Insert Format Tools Data Window Help

A2 = 175

	A	B	C	D	E	F	G	H	I	J	K
1	id	ime	prezime	br indexa	fakultet	smer	godina	br ispita	status	pol	starost
2	175	Marija	Jelic	546/99	FON	opsti menadzment	III	21	budzet	zensko	23
3	176	Borko	Mikic	361/00	FON	zajednicke osnove	II	12	samofinansiranje	muski	21
4	177	Ana	Peric	210/02	FPN	zurnalistika	I	5	budzet	zenski	20
5	178	Sanja	Lazic	23/99	PMF	opsti	IV	30	budzet	zenski	23
6	179	Mina	Jovanovic	245/99	ETF	automatika	III	15	budzet	zenski	20
7	180	Marko	Petrovic	65/01	ETF	telekomunikacije	I	2	budzet	muski	23
8	181	Nemanja	Jevtic	548/01	FON	upravljanje kvalitetom	II	10	budzet	muski	21
9	182	Jelena	Isajev	321/00	SF	zeleznicki saobracaj	II	10	samofinansiranje	zenski	19
10	183	Vojkan	Nikolic	19/00	FON	informacioni sistemi	III	17	samofinsiranje	muski	25
11											

Tipovi promenljivih će biti definisani na osnovu vrednosti koje se nalaze u drugoj vrsti

IV. Linearna regresija u SPSS softverskom paketu

U SPSS programu obezbedjen je postupak za analizu regresionog modela koristeći neki od 5 modula formiranja regresionih jednačina:

- selekcija unapred (forward selection)
- eliminacija unazad (backward elimination)
- selekcija korak po korak (stepwise selection)
- fiksirani ulazi (forced entry)
- fiksirana isključenja (forced removal)

Minimalni zahtevi za sprovođenje postupka analize regresionog modela u SPSS programu su:

- **Jedna zavisna promenljiva
(numerički izražena)**
- **Jedna ili više nezavisnih promenljivih
(numerički izraženih)**

Primer zadatka sa voćem

Untitled - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1: x 2

	x	dohodak po clanu	var	var	var	var
1	2	4				
2	3	3				
3	3	5				
4	4	6				
5	5	7				
6	5	8				
7	6	9				
8	7	12				
9	8	10				
10	10	14				

Untitled - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

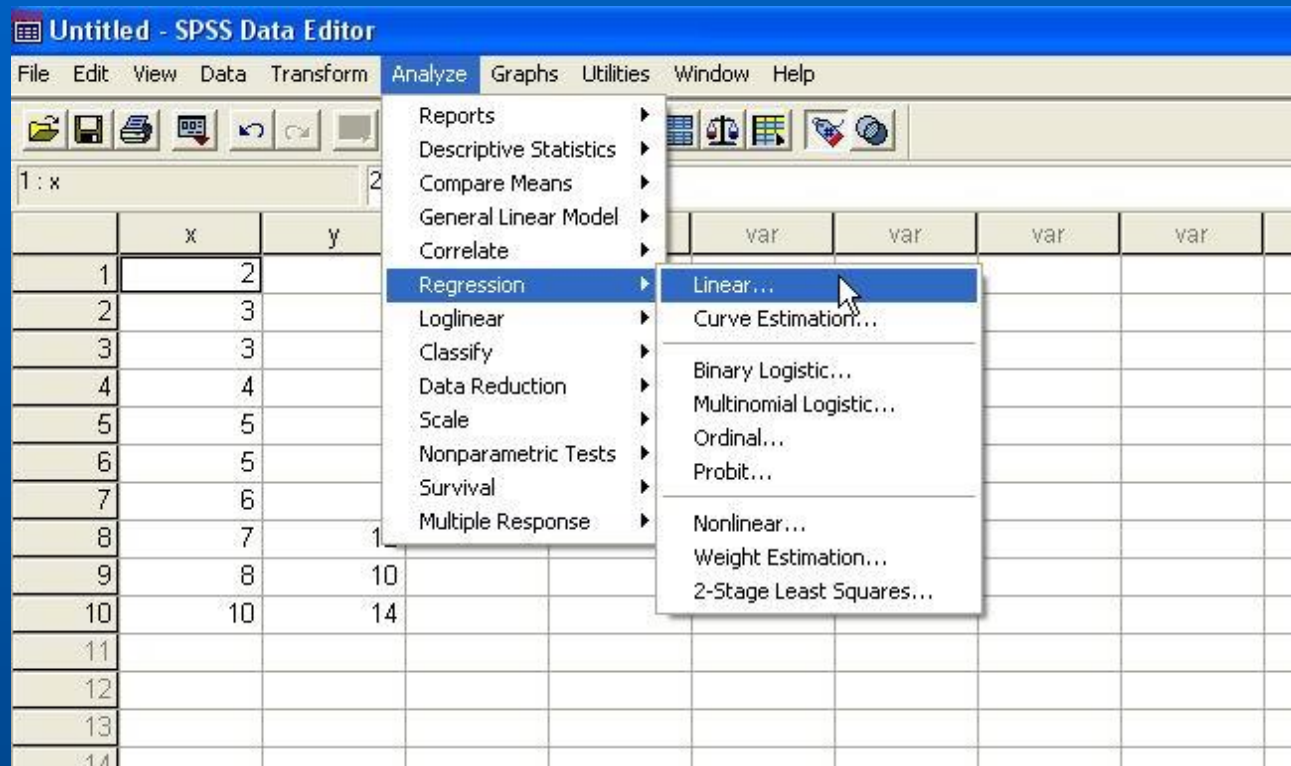
	Name	Type	Width	Decimals	Label	Value Labels
1	x	Numeric	8	0	dohodak po clanu	None
2	y	Numeric	8	0	potrosnja voca po clanu	None
3						
4						

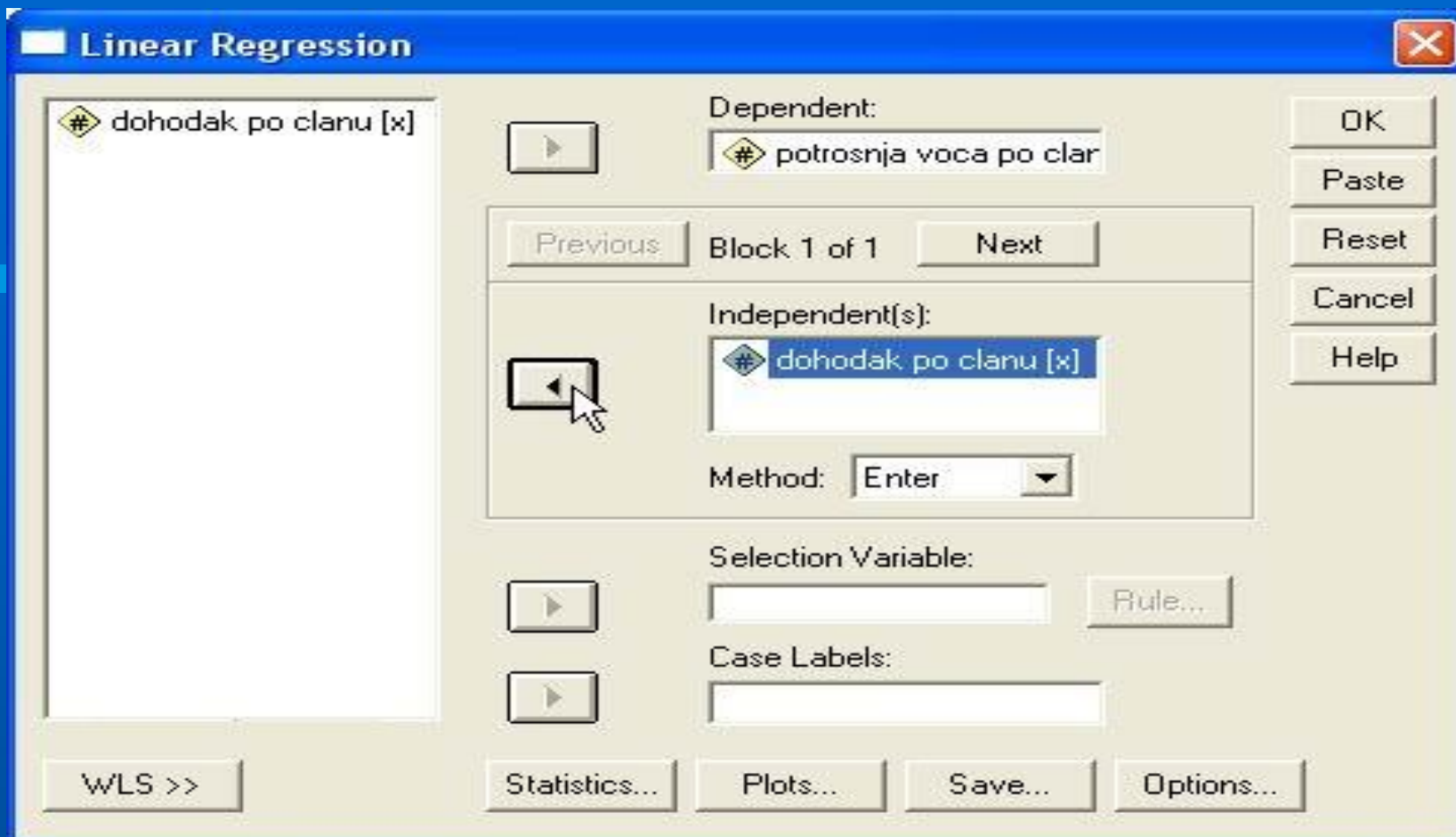
29			
30			
31			
32			

◀ ▶ Data View Variable View

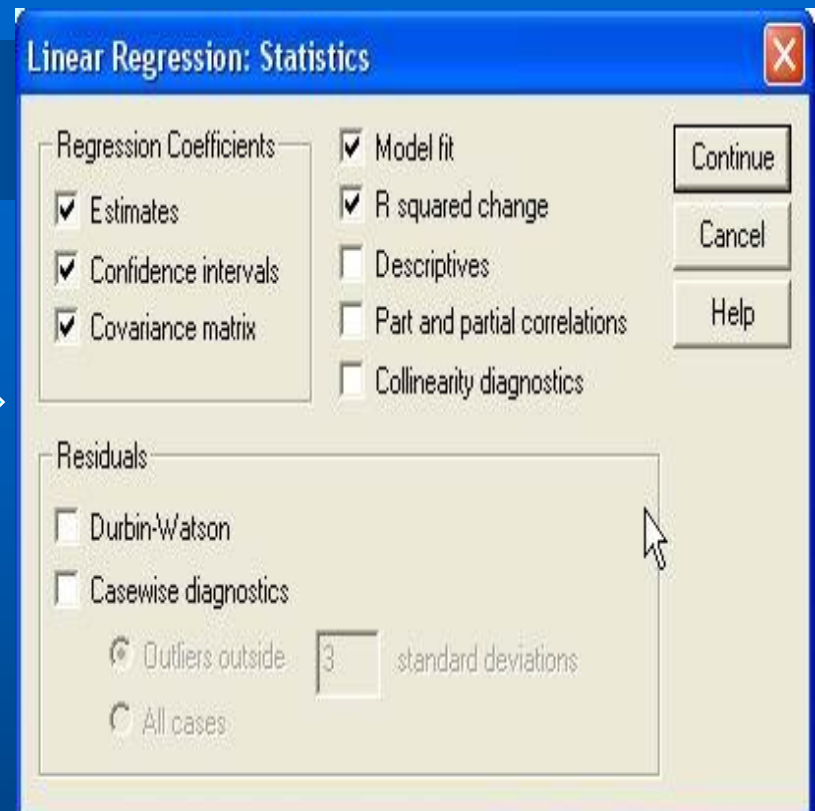
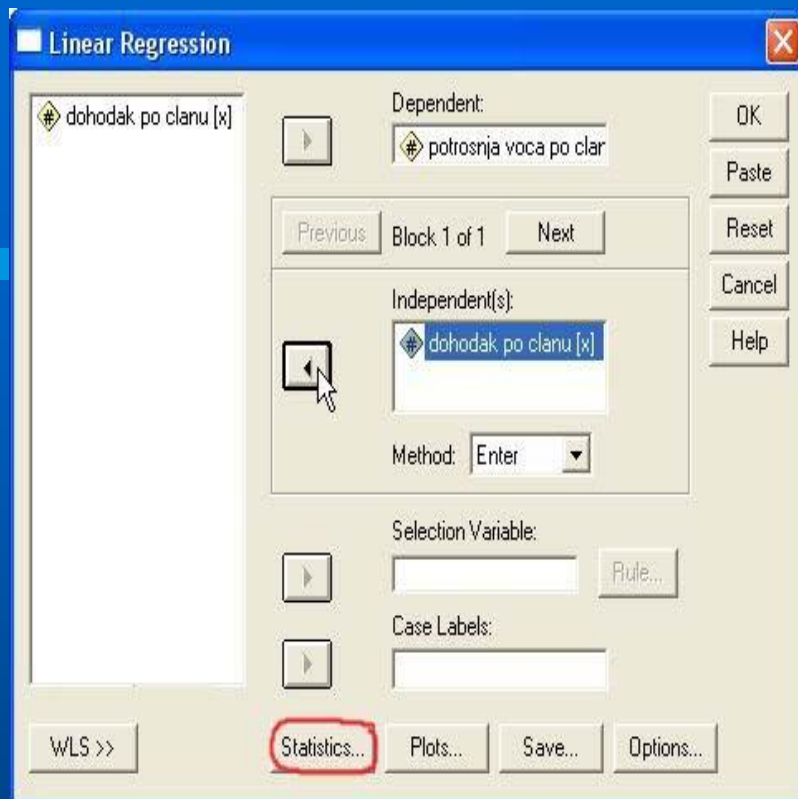
SPSS softverski paket

Da bismo počeli postupak regresione analize potrebno je u meniju izabrati: Analyze, Regression, Linear..





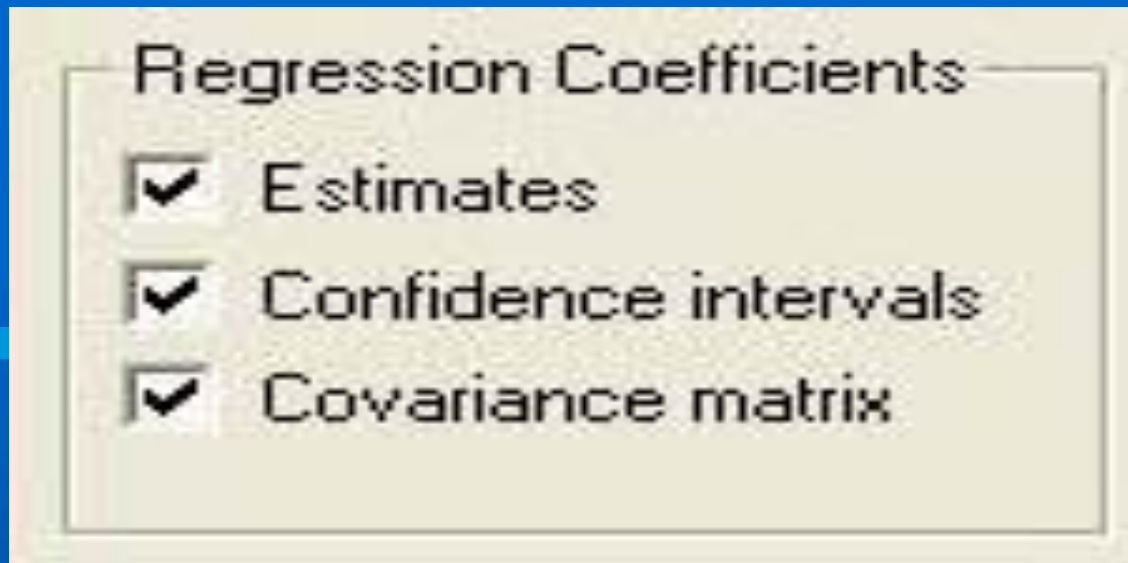
Otvoren je dialogbox: Linear Regression



Klikom na Statistics u prozoru Linear Regression
otvaramo

novi dialogbox Linear Regression: Statistics

SPSS softverski paket

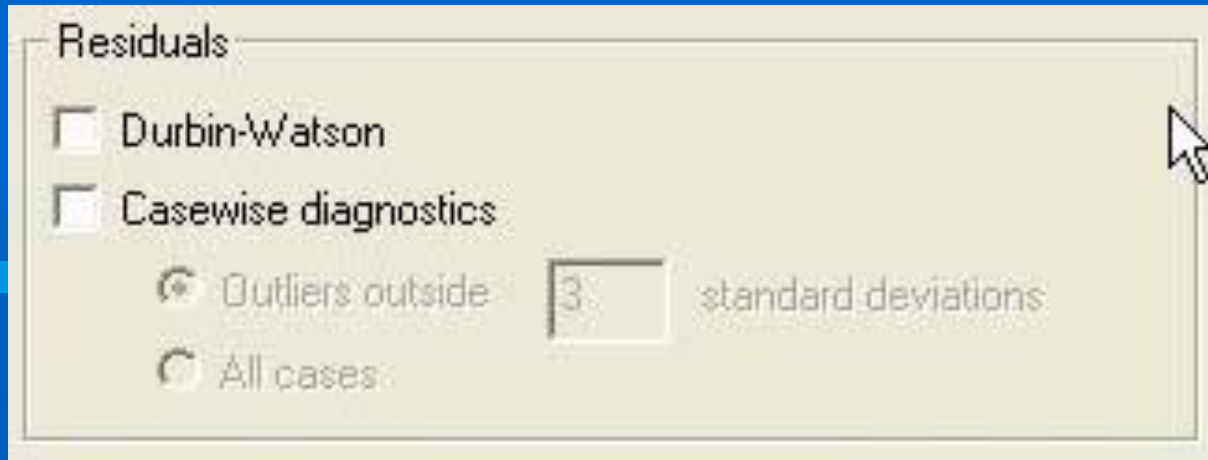


Regression Coefficients – koeficijenti regresije

- *estimates* – ispituju se koeficijenti regresije i sa njima povezane mere
- *confidence intervals* – prikazuje 95% interval poverenja za svaki nestandardizovani koeficijent regresije
- *covariance matrix* – matrica varijanse - kovarijanse nestandardizovanih koeficijenata regresije

- ☒ Model fit
- ☒ R squared change
- ☐ Descriptives
- ☐ Part and partial correlations
- ☐ Collinearity diagnostics

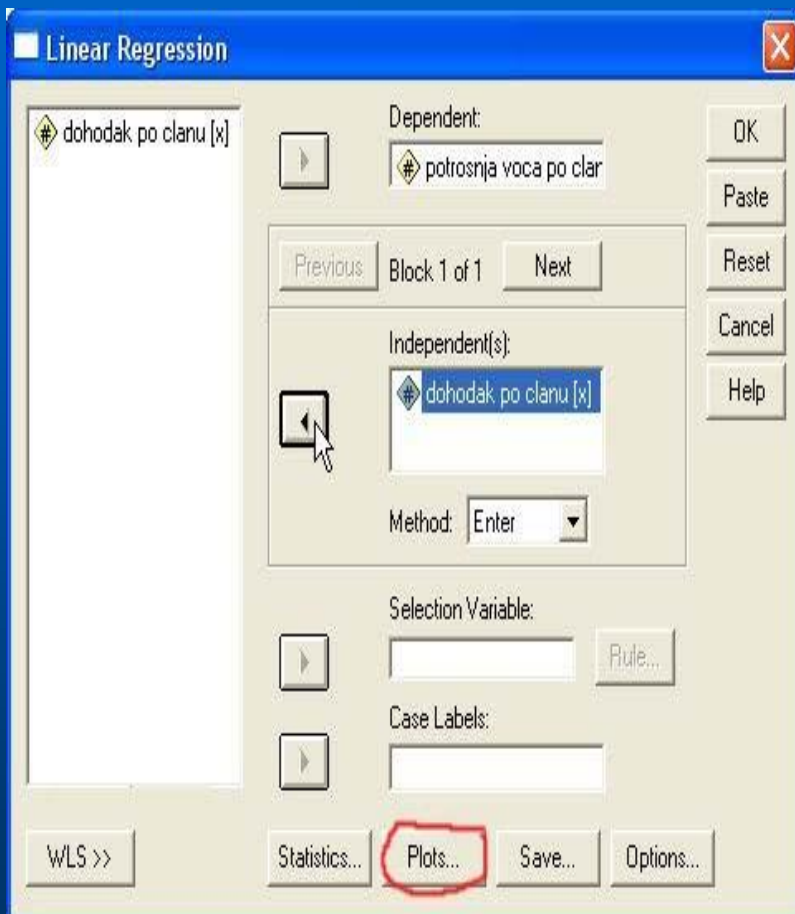
- ***model fit*** - R , R^2 , standardna greška, ANOVA tabela (prikazuje stepen slobode, sume kvadrata, srednje kvadrate, vrednost F i posmatranu verovatnoću za F)
- ***R squared change*** – podešeni koeficijenti višestruke determinacije
- ***descriptives*** – deskriptivna statistika
- ***part and partial correlations*** – parcijalna korelacija
- ***collinearity diagnostics*** – dijagnostika multikolinearnosti, (VIF) varijacioni inflacioni faktor

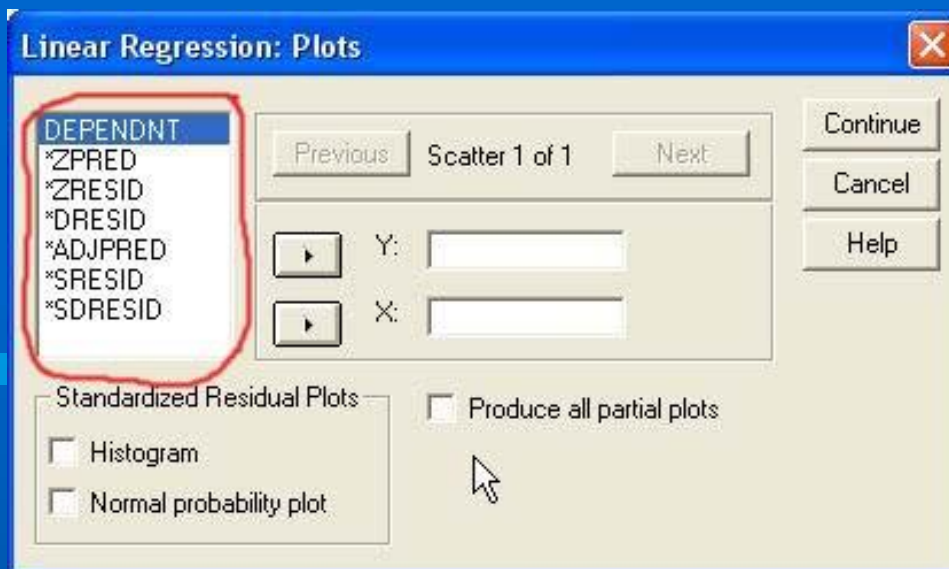


Residuals – Reziduali

- ***Durbin-Watson*** – test statistika, pokazuje odgovarajuće statistike za standardizovane i nestandardizovane rezidualne i predviđene vrednosti
- ***Casewise diagnostics*** - otkrivanje outlier-a

Grafike za promenljive u jednačini dobijamo klikom na dugme Plots u Linear Regression dialogbox-u





Zavisna promenljiva odgovarajuće predviđena iz regresionog modela i reziduali, pojavljuju se na izvornoj listi:

****ZPRED*** – standardizovane predviđene vrednosti

****ZRESID*** – standardizovani reziduali

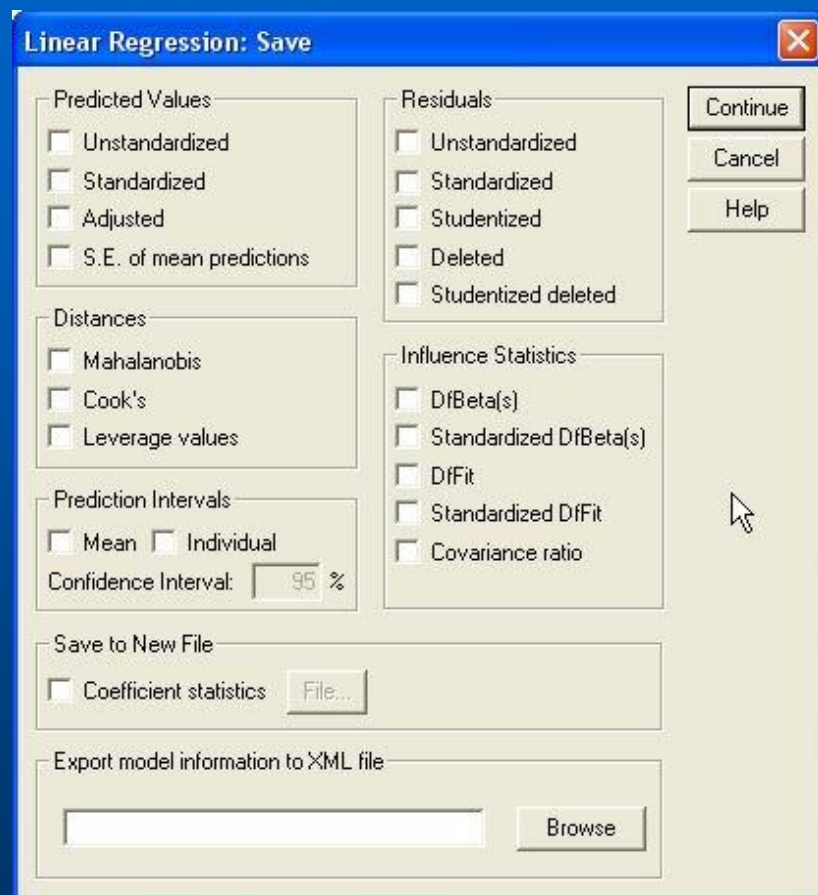
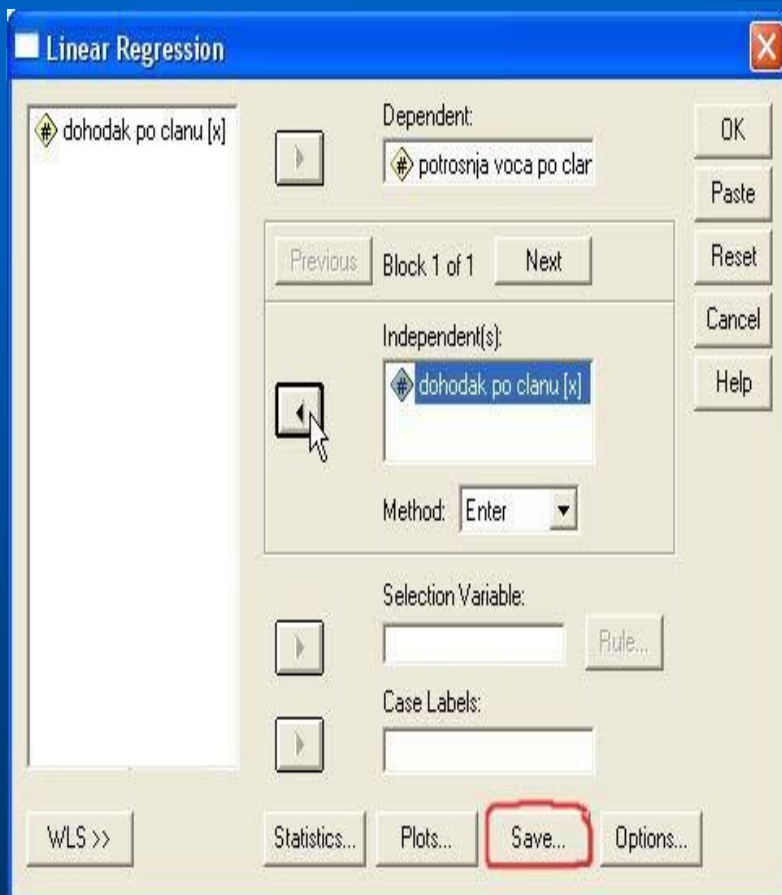
****DRESID*** – očišćeni reziduali

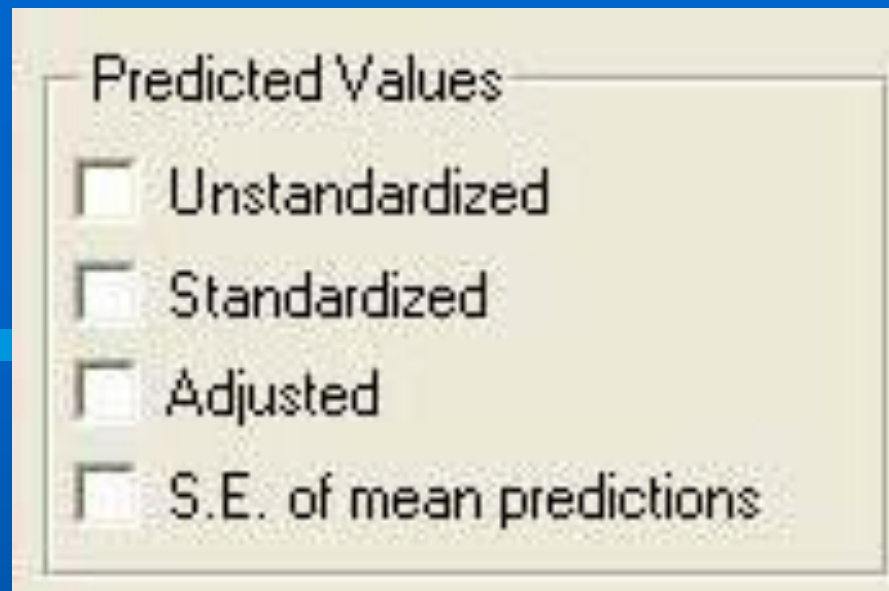
****ADJPRED*** – popravljene predviđene vrednosti

****SRESID*** – S - reziduali

****SDRESID*** – S - očišćeni reziduali

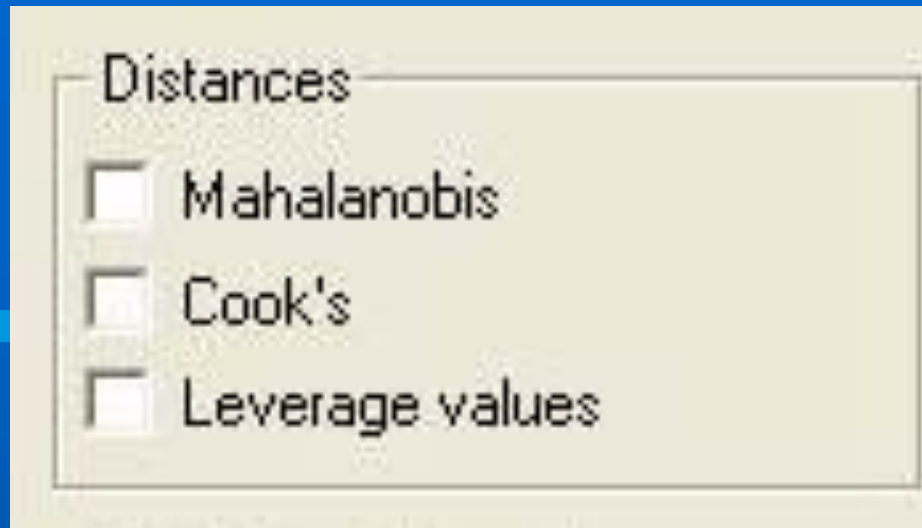
Da bi sačuvali rezidualne, predviđene vrednosti ili odgovarajuća merenja kao nove promenljive, kliknemo na Save





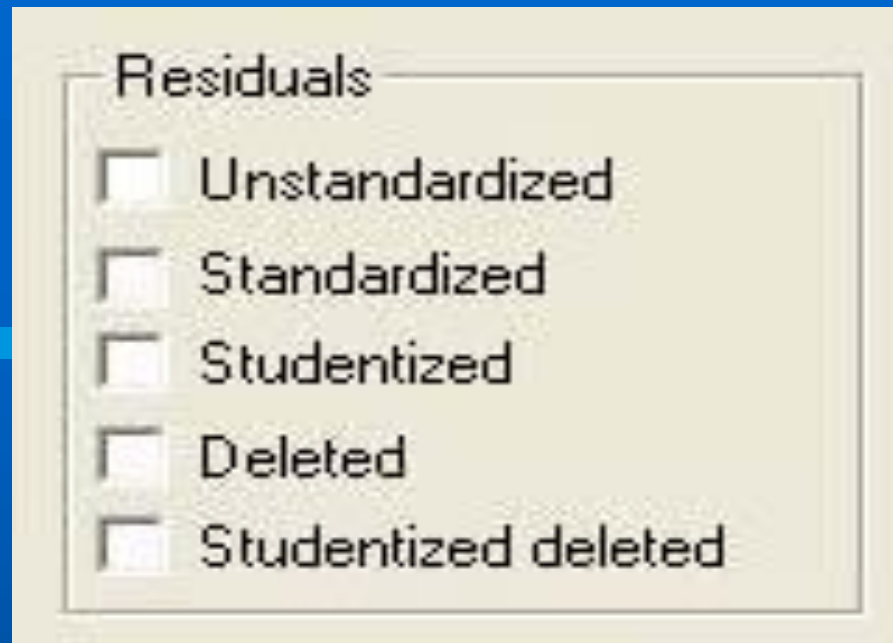
Predicted Values – predvidjene vrednosti

- ***Unstandardized*** – nestandardizovane predvidjene vrednosti
- ***Standardized*** – standardizovane predvidjene vrednosti
- ***Adjusted*** – popravljene predvidjene vrednosti
- ***S.E. of mean predictions*** – standardna greška predvidjenih vrednosti



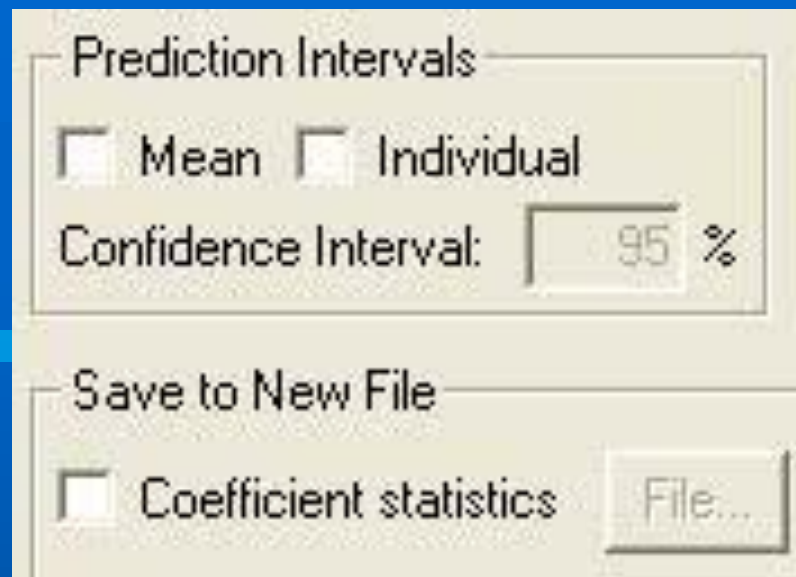
Distances – rastojanje

- ***Mahalanobis*** – Mahalanobisovo rastojanje
- ***Cook's*** – Kukovo rastojanje
- ***Leverage values*** – centrirane snažne (uporišne) vrednosti



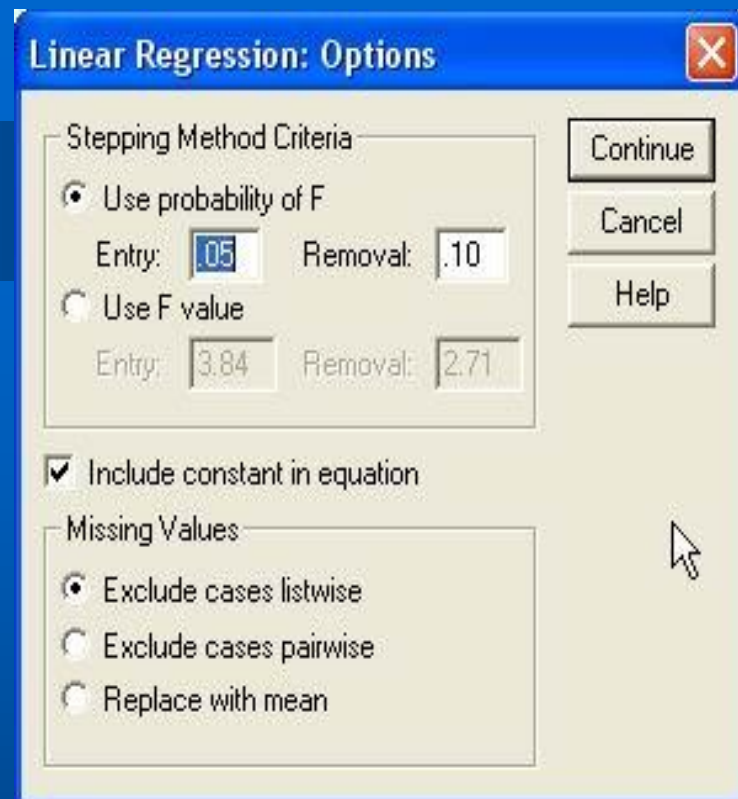
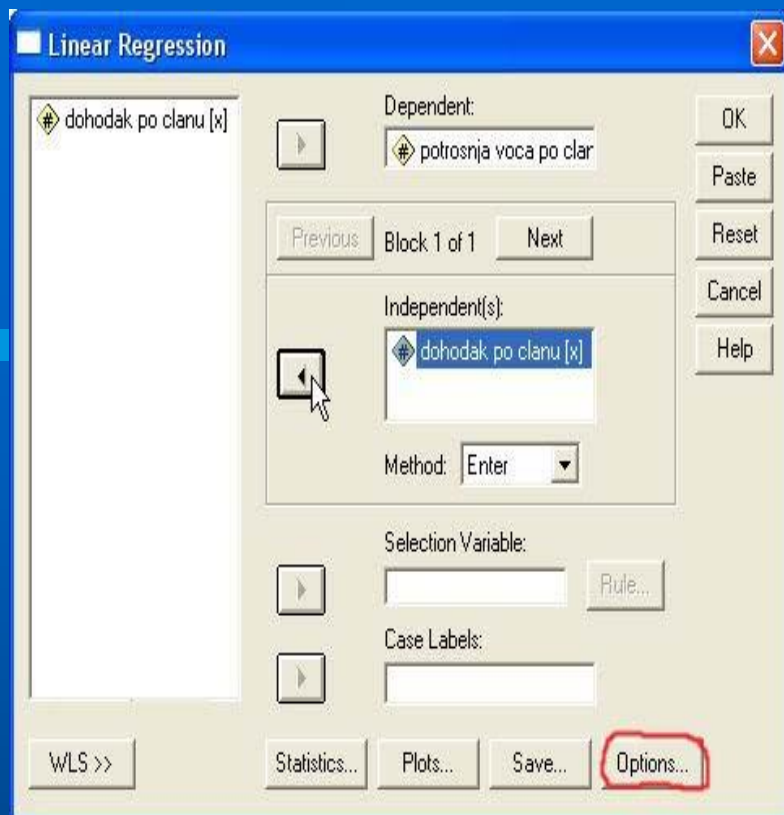
Residuals – Reziduali

- ***Unstandardized*** – nestandardizovani reziduali
- ***Standardized*** – standardizovani reziduali
- ***Studentized*** – S - reziduali (student)
- ***Deleted*** – isključeni reziduali
- ***Studentized deleted*** – S - isključeni reziduali

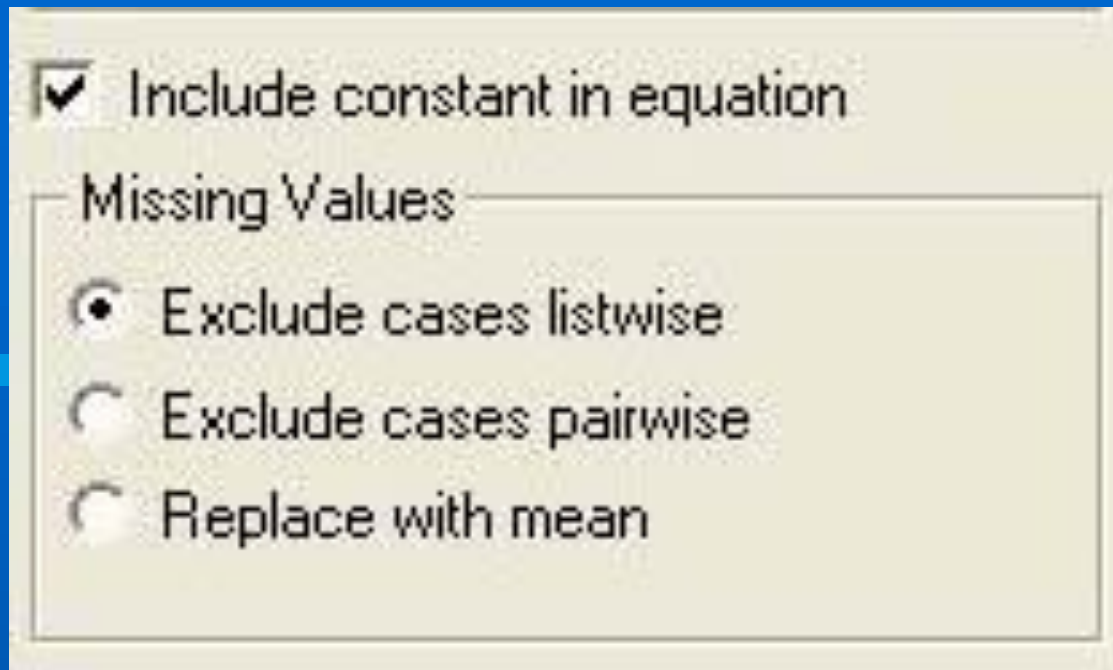


Prediction Intervals – interval poverenja za očekivanu i individualnu vrednost

- *Mean* – donja i gornja granica za interval predviđanja proseka predvidivih odgovora
- *Individual* – donja i gornja granica intervala predviđanja za jednu opservaciju
- *Confidence Interval* – odredjujemo intervale poverenja za prosečne individualne vrednosti



Za kontrolu kriterijuma po kojima se promenljive biraju za ubacivanje ili izbacivanje iz regresionog modela, za isključene konstante ili za kontrolu slučajeva kojima nedostaju vrednosti koristimo opciju Options u Linear Regression dialogbox-u



Missing Values – nedostajuće podaci

- *Exclude cases listwise* – samo slučajevi sa ispravnim vrednostima za sve promenljive se uključuju u analizu
- *Exclude cases pairwise* – slučajevi sa kompletnim podacima za par promenljivih koje se korelišu
- *Replace with mean* – zamenjuje se vrednost koja nedostaje sa aritmetičkom sredinom promenljive

Output sa traženim
izlaznim rezultatima
podeljen je na dva okvira

- o levi deo sadrži izgled
izlaznih rezultata u obliku
sadržaja

- o desni deo sadrži
statističke tabele,
dijagrame i izlazni text.

The screenshot shows the SPSS Output Viewer window with the following content:

Regression

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	dohodak po članu		Enter

a. All requested variables entered.
b. Dependent Variable: potrošnja voca po članu

Model Summary^b

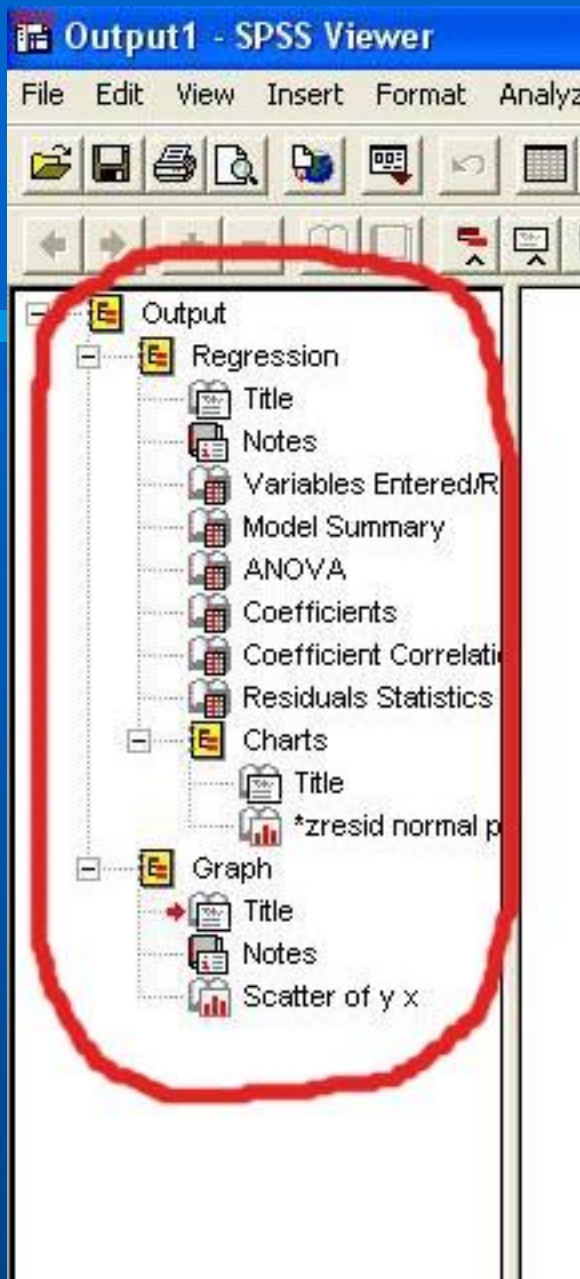
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.956 ^a	.913	.902	1.10	.913	83.834	1	8	.000

a. Predictors: (Constant), dohodak po članu
b. Dependent Variable: potrošnja voca po članu

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	101.878	1	101.878	83.834	.000 ^a
	Residual	9.722	8	1.215		
	Total	111.600	9			

a. Predictors: (Constant), dohodak po članu
b. Dependent Variable: potrošnja voca po članu



Pružá se mogućnost korišćenja scroll bars-a za pregled svih rezultata ili postoji mogućnost da kliknete na stavku u sadržaju i direktno ćete otići na odgovarajuću tabelu ili dijagram

Regression

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	dohodak _a po clanu		Enter

a. All requested variables entered.

b. Dependent Variable: potrosnja voca po clanu

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.955 ^a	.913	.902	1.10	.913	83.834	1	8	.000

a. Predictors: (Constant), dohodak po clanu

b. Dependent Variable: potrosnja voca po clanu

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	101.878	1	101.878	83.834	.000 ^a
	Residual	9.722	8	1.215		
	Total	111.600	9			

a. Predictors: (Constant), dohodak po clanu

b. Dependent Variable: potrosnja voca po clanu

Coefficients^a

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	.658	.854		.770	.464	-1.313	2.628
dohodak po clanu	1.348	.147	.955	9.156	.000	1.008	1.687

a. Dependent Variable: potrosnja voca po clanu

Coefficient Correlations^a

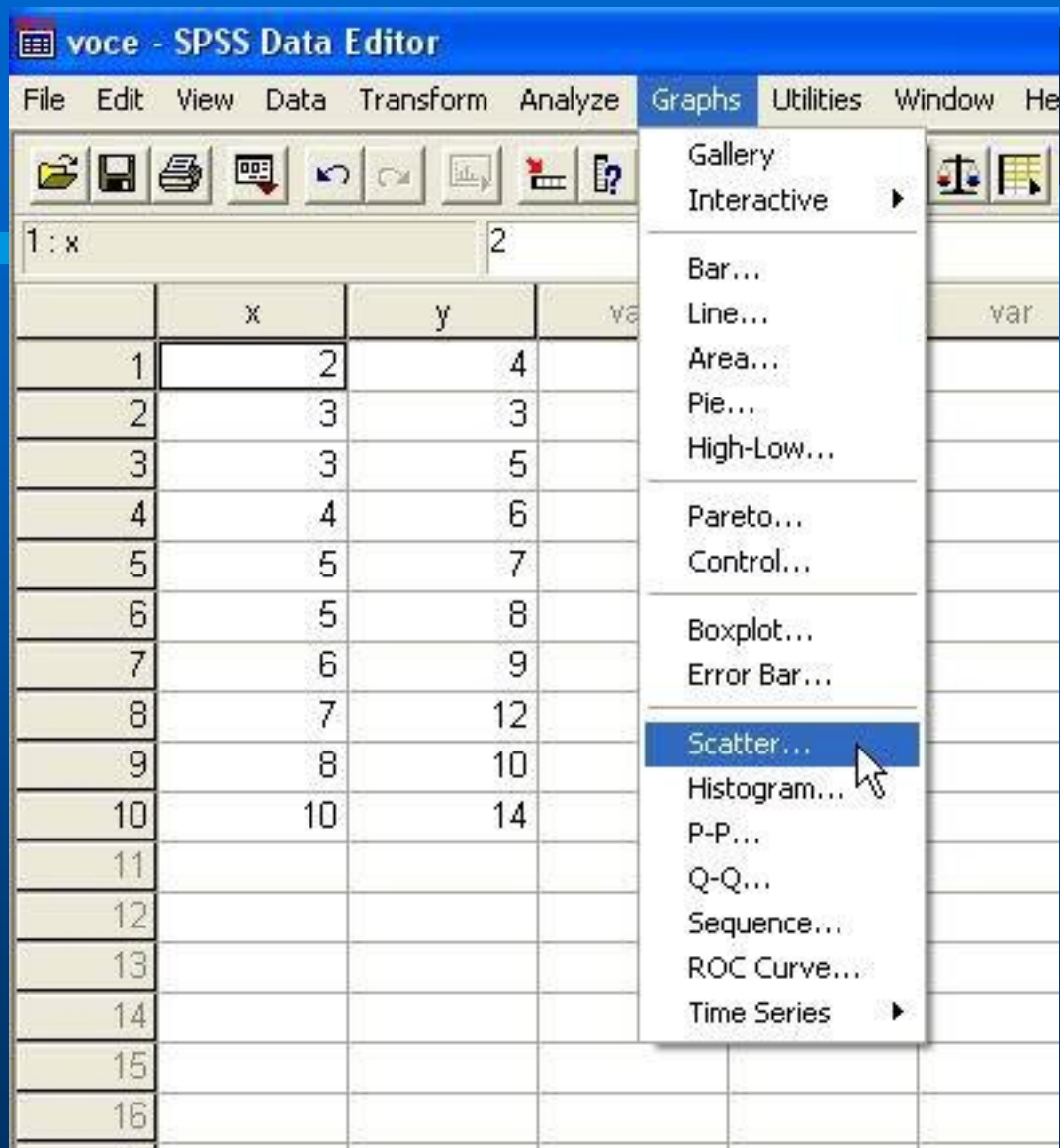
Model		dohodak po clanu
1	Correlations	1.000
	Covariances	2.166E-02

a. Dependent Variable: potrosnja voca po clanu

Residuals Statistics^a

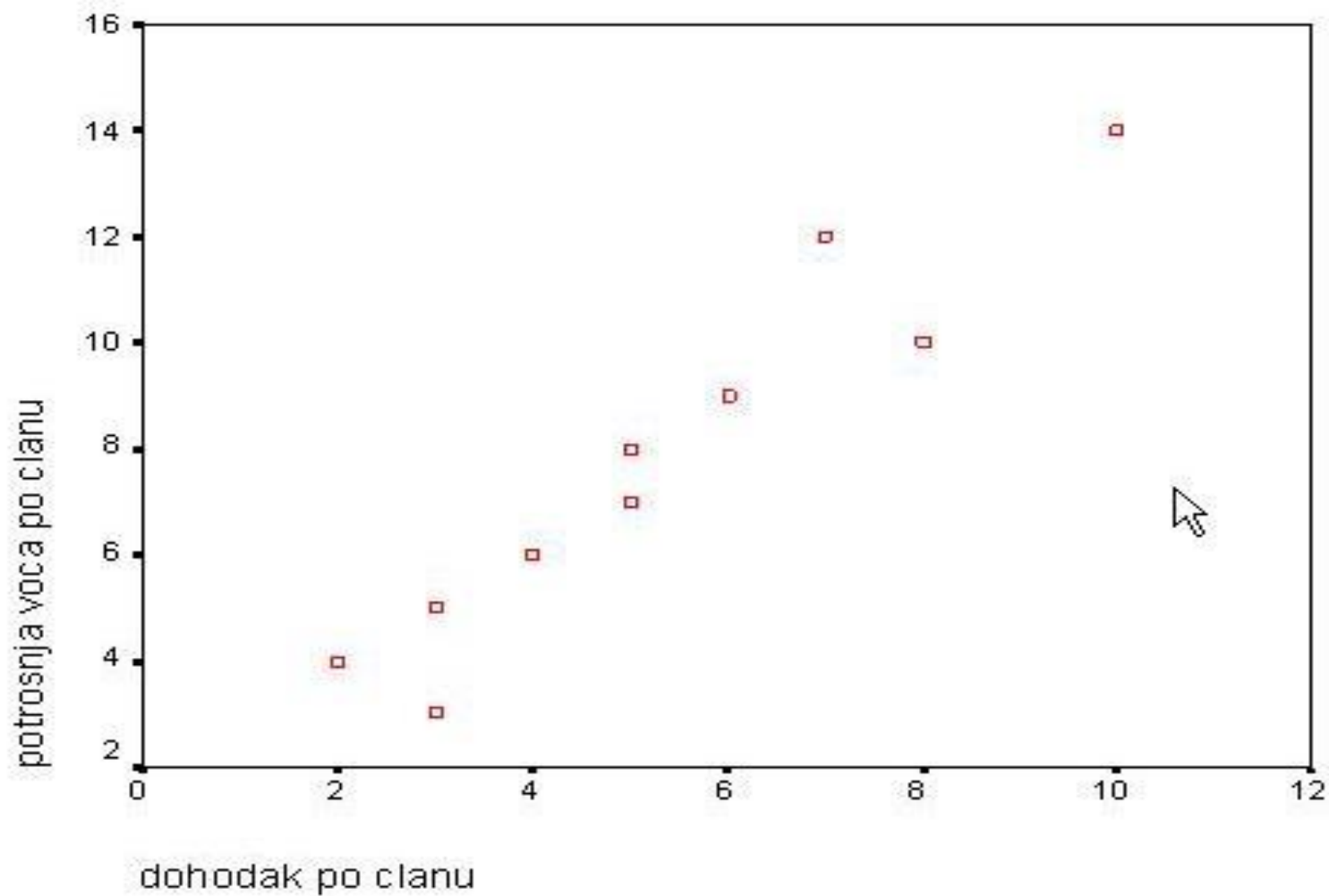
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	3.35	14.13	7.80	3.36	10
Std. Predicted Value	-1.322	1.883	.000	1.000	10
Standard Error of Predicted Value	.35	.77	.48	.13	10
Adjusted Predicted Value	3.08	14.26	7.82	3.43	10
Residual	-1.70	1.91	.00	1.04	10
Std. Residual	-1.543	1.732	.000	.943	10
Stud. Residual	-1.719	1.880	-.010	1.046	10
Deleted Residual	-2.11	2.25	-2.40E-02	1.28	10
Stud. Deleted Residual	-2.024	2.354	-.012	1.218	10
Mahal. Distance	.014	3.544	.900	1.086	10
Cook's Distance	.000	.356	.116	.153	10
Centered Leverage Value	.002	.394	.100	.121	10

a. Dependent Variable: potrosnja voca po clanu



SPSS softverski paket

Graph

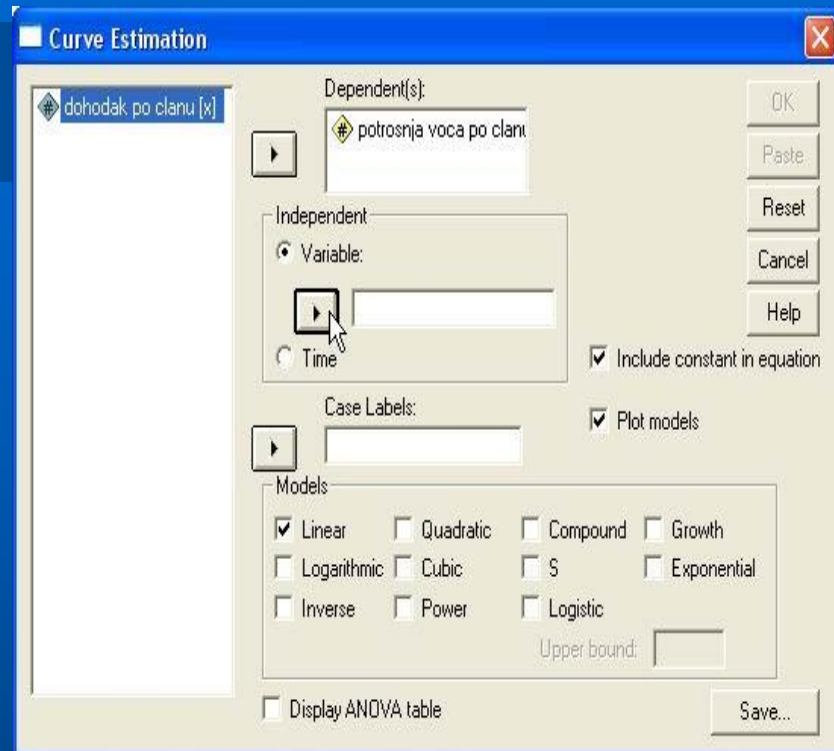
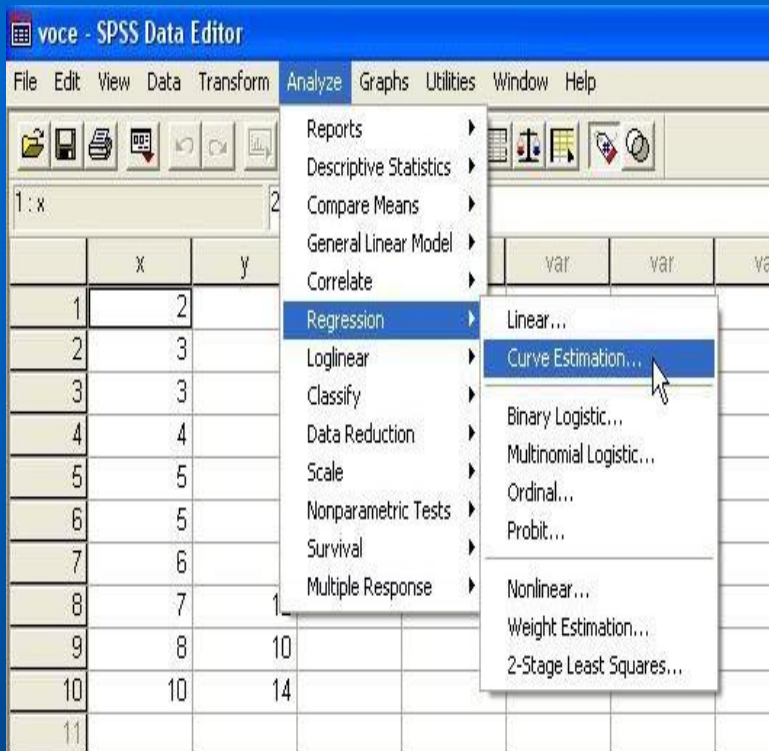


V. Ocene nelinearne regrersije

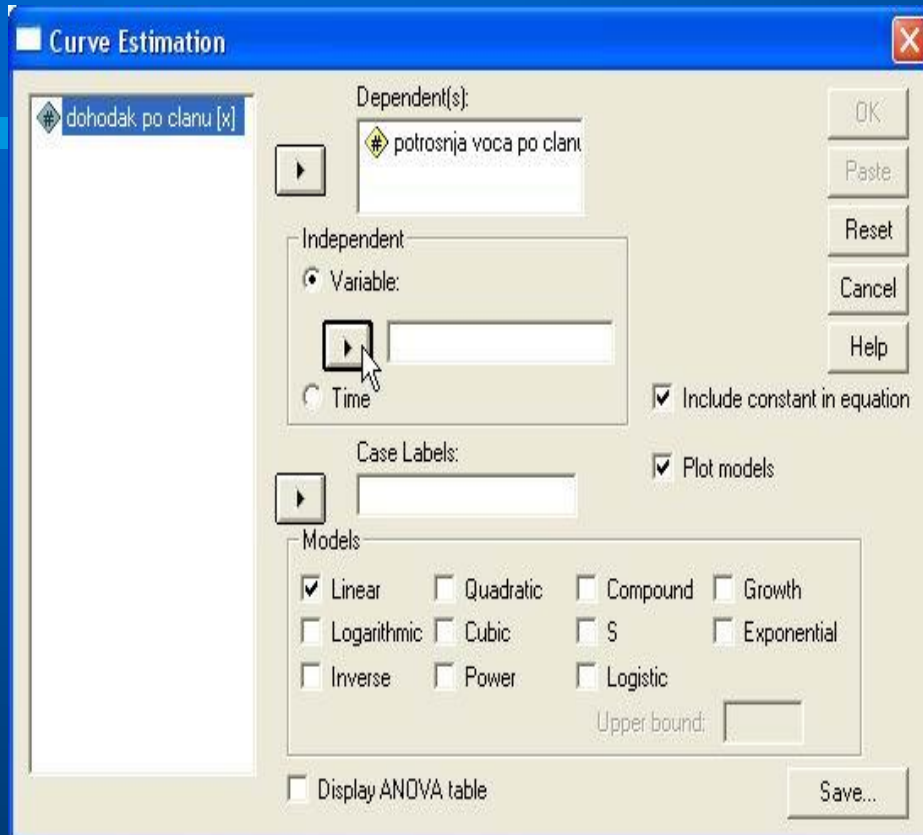
SPSS paket omogućava ocene nelinearnih (krivolinijskih) regresija zajedno sa grafičkim prikazima za 11 različitih regresionih modela.

Minimalni zahtevi su:

- jedna ili više zavisnih promanljivih
- nezavisna promenljiva koja moze biti promanljiva iz radne datoteke ili vreme

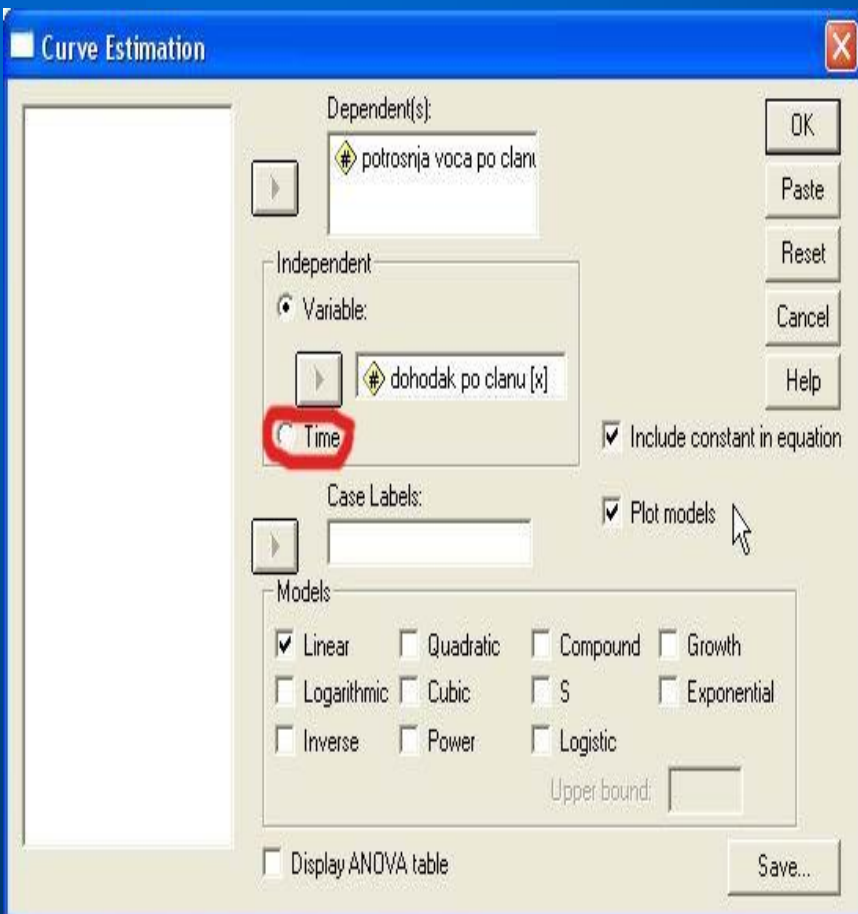


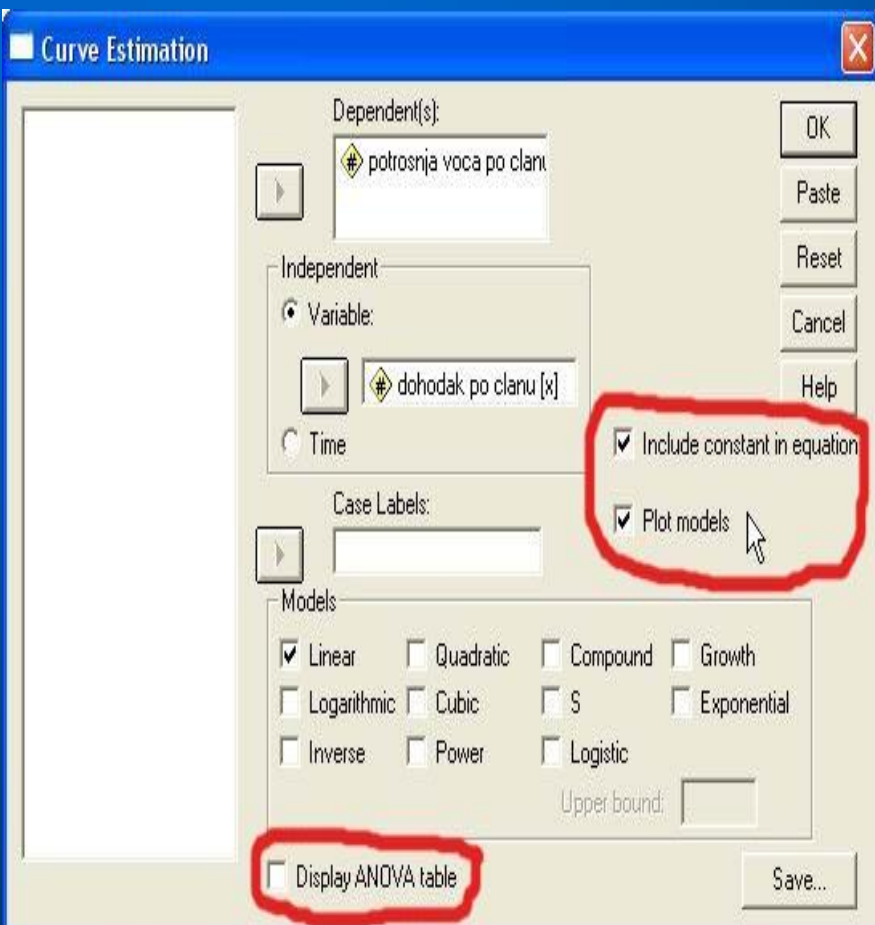
Za ovu vrstu analize iz menija Statistics odabiramo stavku Regression, a odatle stavku Curve Estimation.



**Numeričke
promenljive u radnoj
datoteci se prikazuju
na izvornoj listi
promenljivih.
Odabere se jedna ili
više nezavisnih
promenljivih i zavisna
promenljiva, a zatim
mišem klikom na OK,
da bi se sprovela
default opcija za
linearni model.**

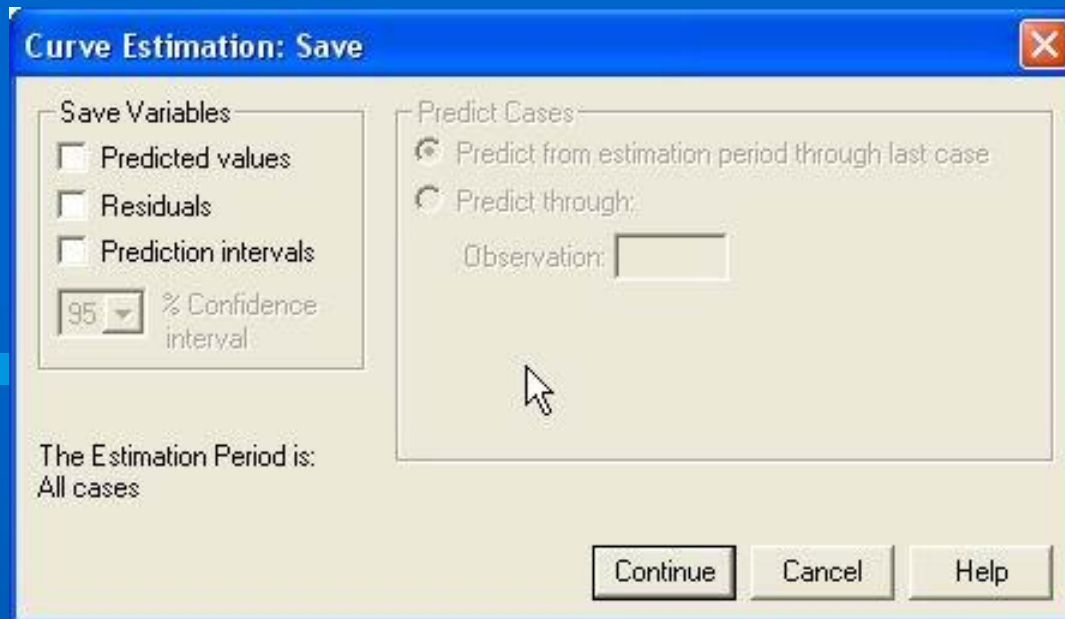
Ako se odabere Time (vreme) umesto promenljive iz radne datoteke za nezavisnu promenljivu, zavisna promenljiva treba da ima prirodu vremenske serije. Analiza vremenskih serija zahteva svaku strukturu datoteke u kojoj svaki slučaj (red) predstavlja skup opservacija u razlicitim vremenskim trenucima, dok je vremenska pauza izmedju slucajeva jednaka.





Ovde imamo dodatne opcije:

- Include constant in equation* - koja ocenjuje konstantu regresione jednacine. Ova opcija je uključena po default-u, a moze i da se isključi;
- Plot models* - koja iscrtava vrednosti zavisne promenljive i svakog izabranog modela nasuprot nezavisne promenljive. Rade se odvojeni grafici za svaku zavisnu promenljivu;
- Display ANOVA table* - koja prikazuje tabelu analize varijanse za svaki izabrani model



Da bi se snimile predviđene vrednosti, reziduali i intervali poverenja kao nove promenljive otvaramo prozor Curve Estimation: Save. U opciji Save Variables moguće je odabrati jednu ili više ponudjenih promenljivih. Nova imena promenljivih i opisani nazivi se pojavljuju u tabeli u izlaznom prozoru.

Curve Fit

MODEL: MOD_1.

Dependent variable.. Y Method.. LINEAR

Listwise Deletion of Missing Data

Multiple R .95545
R Square .91289
Adjusted R Square .90200
Standard Error 1.10238

Analysis of Variance:

	DF	Sum of Squares	Mean Square
→ Regression	1	101.87807	101.87807
Residuals	8	9.72193	1.21524

F = 83.83366 Signif F = .0000

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig T
X	1.347594	.147180	.955451	9.156	.0000
(Constant)	.657754	.854407		.770	.4635

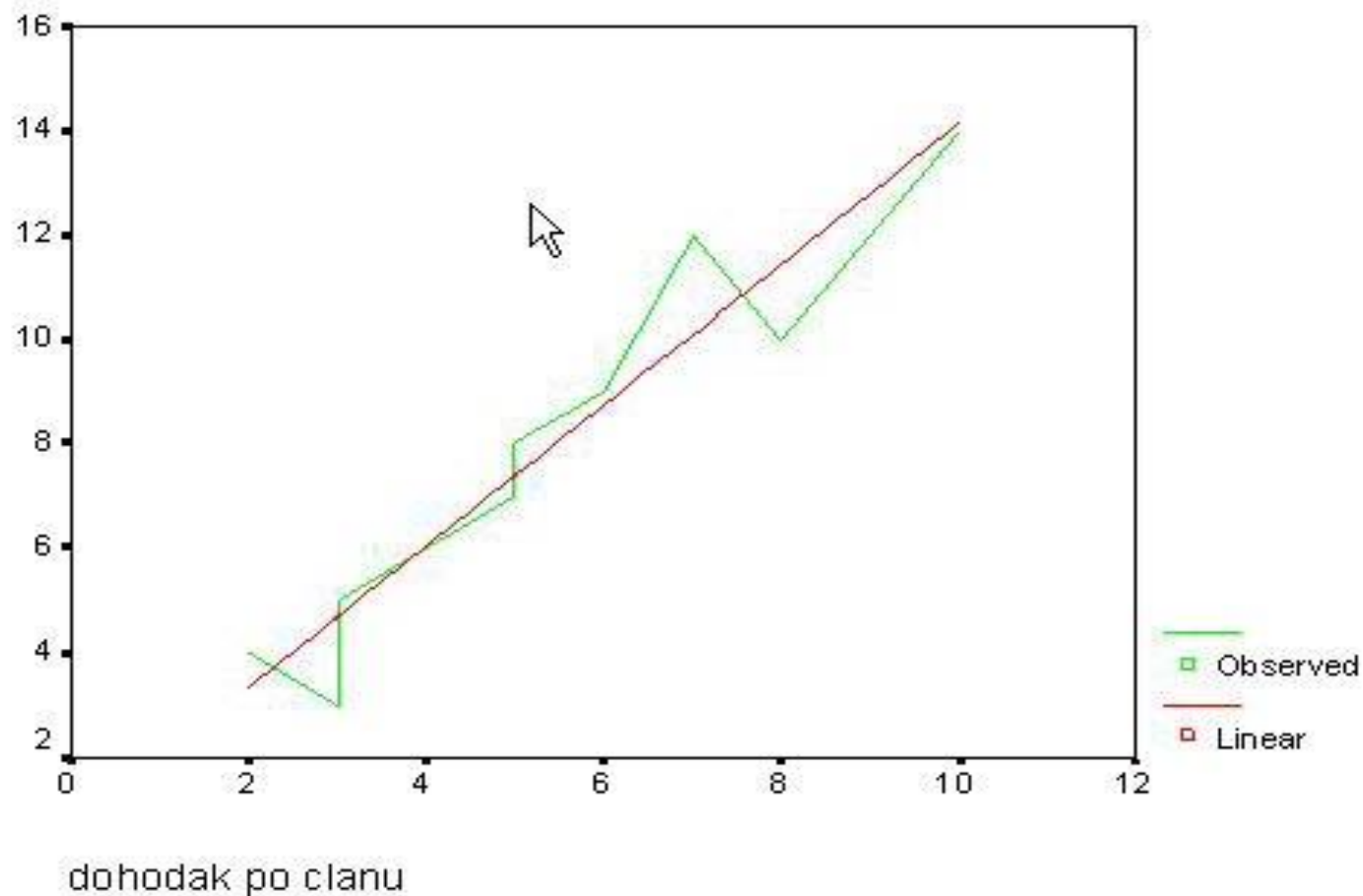
The following new variables are being created:

Name	Label
FIT_1	Fit for Y with X from CURVEFIT, MOD_1 LINEAR
ERR_1	Error for Y with X from CURVEFIT, MOD_1 LINEAR
LCL_1	95% LCL for Y with X from CURVEFIT, MOD_1 LINEAR
UCL_1	95% UCL for Y with X from CURVEFIT, MOD_1 LINEAR

Izlazni rezultati su prikazani u tekstualnom formatu.

SPSS softverski paket

potrosnja voca po clanu



Izlazni rezultati su prikazani na dijagramu zavisnosti

SPSS softverski paket

VII. Osvrt na izlaganje

Softverski paket SPSS koji smo mi obradili sadrži sve neophodne statističke metode za analizu linearnog regresionog modela.

Želeli smo da za većinu obezbedimo sasvim dovoljno praktičnih objašnjenja potrebnih za razumevanje samog statističkog paketa, mogućnosti njene primene i sa kratkim upustvom kako to radi.

Naše izlaganje je posvećeno problemima linearne regresije obradjenim podrškom SPSS paketa.

Obezbedili smo osnovu za razumevanje analize regresionog modela i interpretirali rezultate na brz i efikasan način korišćenjem softverskih alata.

VIII. Korišćena literatura

1. PC Statistika i Verovatnoća

Prof. dr. Nahod Vuković

Beograd, 2000

2. Ekonometrija – Skripta

Prof. dr. Milica Delić

Izlagачi

Spasić Aleksandra 352/01

E-mail: sasas@infosky.net

Nikolić Darko 135/01

E-mail: darkoni@net.yu