

Statistika II z računalniško analizo podatkov

Bivariatna regresija,
tipi povezanosti

Regresijska analiza

Regresijska analiza je statistična metoda, ki nam pomaga analizirati odnos med odvisno spremenljivko in eno ali več neodvisnimi spremenljivkami.

Koraki regresijske analize

1. Najprej postavimo teoretične predpostavke o odvisnosti med spremenljivkami: **regresijski model**.
2. Izbrani model testiramo na vzorcu. Ocenimo **parametre** modela.
3. Regresijski model ima dvojno vlogo:
 - a) **Pojasnjevalno**: s pomočjo parametrov lahko opišemo, kako neodvisne spremenljivke vplivajo na odvisno.
 - b) **Napovedovalno**: S pomočjo regresijskega modela lahko na podlagi vrednosti neodvisnih spremenljivk napovemo vrednost odvisne spremenljivke.

Postavitev regresijskega modela

Pri postavitvi regresijskega modela je potrebno:

1. Izbrati odvisno spremenljivko in vse tiste spremenljivke, ki nanjo pomembno vplivajo ter izbrati način povezave. To je najbolje narediti na podlagi neke teorije. Statistična metoda ne more nadomestiti dobre teoretične osnove modela.
2. Ko izberemo spremenljivke, je potrebno ugotoviti še vrsto odvisnosti med njimi. Statistična metoda nato omogoča testiranje parametrov modela.
3. Ocenjevanje regresijskega modela temelji na določenih predpostavkah, ki jih je potrebno preveriti.
4. Neizpolnjevanje osnovnih predpostavk lahko vodi do statistično neveljavnih rezultatov.

Ponovitev linearne bivariatne regresije

Analizirati želimo linearno odvisnost spremenljivke Y od spremenljivke X . Iščemo torej premico, ki se čim bolj prilega točkam v dvorazsežnem prostoru, ki so določene z vrednostmi spremenljivk Y in X . Iščemo tako premico, da bodo odkloni točk (oz. vsota kvadriranih odklonov) v smeri odvisne (Y) spremenljivke čim manjša. Preveriti želimo, ali spremenljivka X vpliva na spremenljivko Y .

1. Predpostavimo torej linearno zvezo: $Y = \alpha + \beta X + \varepsilon$
2. Postavimo ničelno in osnovno hipotezo:

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

Ponovitev linearne bivariatne regresije

3. Izračunamo oceno regresijskega koeficientov α in β - se pravi a in b in standardno napako b ter odstotek pojasnjene variance r^2 .

$$b = \frac{C_{YX}}{s_X^2} \quad a = \bar{Y} - b\bar{X} \quad r^2 = \frac{C_{YX}^2}{s_Y^2 s_X^2} \quad SE(b) = \frac{s_Y \sqrt{1-r^2}}{s_X \sqrt{n-2}}$$

4. Zapišemo dobljeno linearno zvezo: $Y' = a + bX$
5. Izračunamo testno statistiko za preverjanje ničelne domneve in stopinje prostosti:

$$t = \frac{b}{SE(b)} \quad m = n - 2$$

6. Preverimo domnevo: Izračunamo, kakšna je verjetnost (statistična značilnost), da smo na konkretnem vzorcu dobili eksperimentalno vrednost testne statistike ob pravilnosti ničelne domneve. Če je ta verjetnost majhna (statistična značilnost manjša od 5% oz. 0,05), podvomimo v pravilnost ničelne domneve (jo zavrnamo) in trdimo, da spremenljivka X vpliva na spremenljivko Y .

$p < \alpha (0,05)$ H_0 zavrnamo, spremenljivka X vpliva na spremenljivko Y

$p > \alpha (0,05)$ H_0 ne zavrnamo, ne moremo trditi, da spremenljivka X vpliva na spremenljivko Y

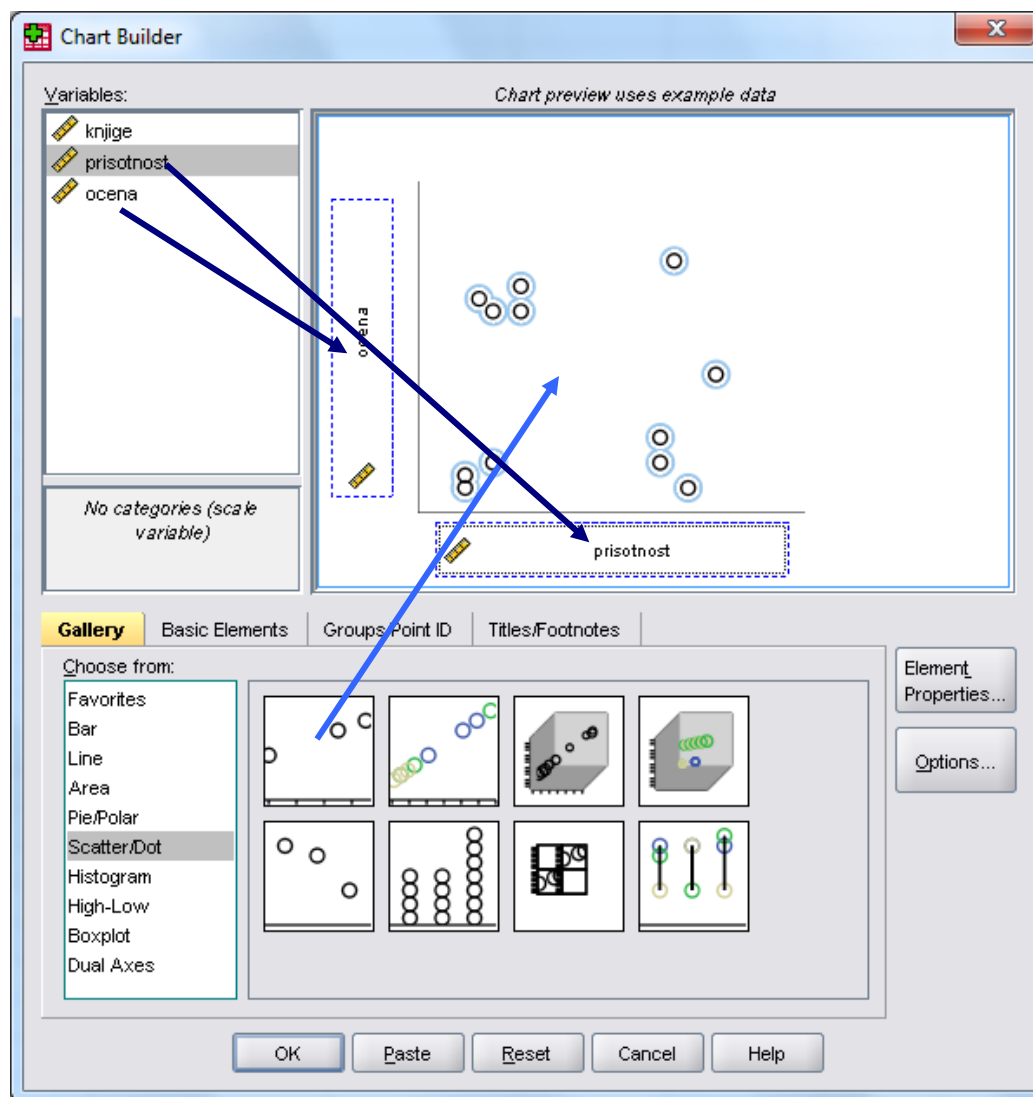
Primer in reševanje v SPSS-u

Na podlagi podatkov “primer_regresija_ocene.sav” bomo analizirali odvisnost spremenljivke ocene (končna ocena – število točk) od prisotnost “št. predavanja, ki jih je študent obiskal”. Zanima nas, ali prisotnost na predavanjih vpliva na oceno.

1. Predpostavimo linearno zvezo:
$$ocena = \alpha + \beta \cdot prisotnost + \varepsilon$$
2. Postavimo ničelno in osnovno hipotezo:
$$H_0: \beta = 0$$
$$H_1: \beta \neq 0$$

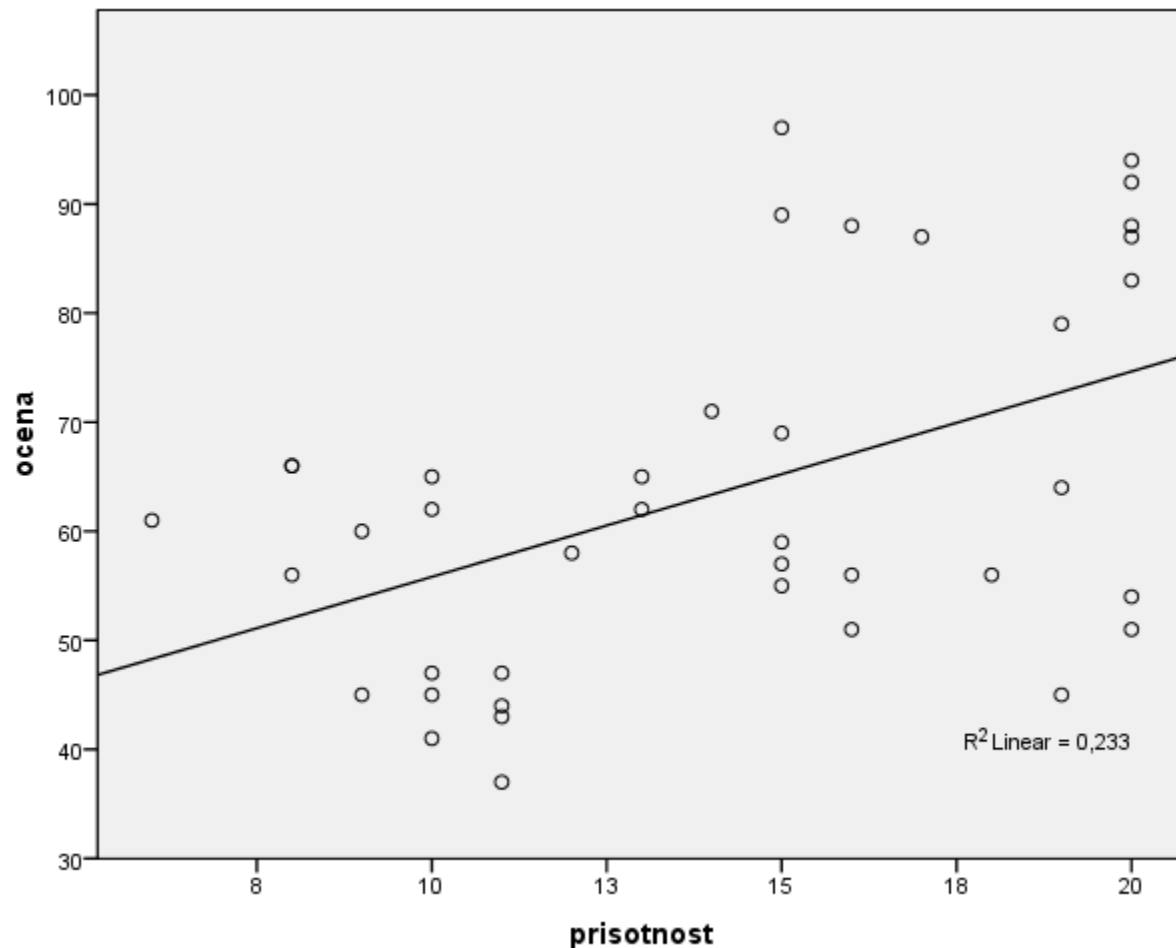
Primer in reševanje v SPSS - u

3. Narišemo razsevni grafikon (*Graphs – Chart Builder, Izberemo skupino grafov Scatter/Dot in v njej graf Simple Scatter*)



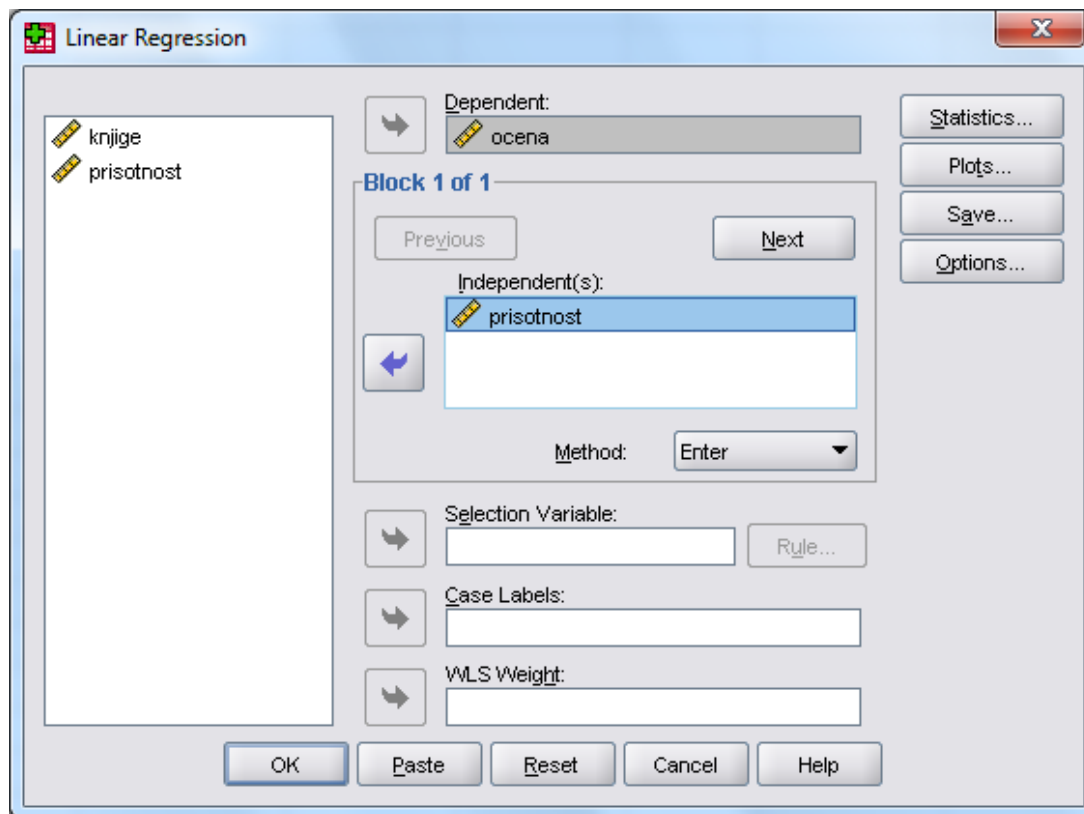
Primer in reševanje v SPSS - u

3. Pregledam o grafikon. Preverimo, če je linearna zveza ustrezna.



Primer in reševanje v SPSS - u

4. Zahtevamo linearno regresijo (*Analyze – Regression – Linear*). V polje *Dependent* prenesemo odvisno (*ocena*), v polje *Independents* pa neodvisno spremenljivko (*pristonost*)



Primer in reševanje v SPSS - u

5. Pregledamo in interpretiramo SPSS-ov output.
 - a) Odstotek pojasnjene variance: Neodvisna spremenljivke (*prisotnost*) pojasni 23,3% variabilnosti odvisne spremenljivke (*ocena*)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,482 ^a	,233	,212	14,826

a. Predictors: (Constant), prisotnost

Opozorilo: Pri statistiki 1 smo determinacijski koeficient označili z R . Tu R predstavlja (multipli) korelacijski koeficient, R^2 (R squared) pa (multipli) determinacijski koeficient.

Primer in reševanje v SPSS - u

5. Pregledamo in interpretiramo SPSS-ov output.
- b) Regresijski koeficienti:

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			
1	(Constant)	36,998	8,169		4,529	,000
	prisotnost	1,883	,555	,482	3,393	,002

a. Dependent Variable: ocena

Ocena regresijskega koeficienta b je 1,883. To pomeni, da ocenjujemo, da če se prisotnost poveča za eno enoto (en obisk predavnaj), se ocena (število točk) za 1,883 točk. Koeficient je statistično značilen pri statistični značilnosti 0,2%. Ker je manjša kot 5%, zavrnilo ničelno hipotezo in trdimo, da prisotnost na predavanjih vpliva na oceno pri predmetu (izraženo kot število točk).

Primer in reševanje v SPSS - u

5. Pregledamo in interpretiramo SPSS-ov output.
- b) Regresijski koeficienti:

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	36,998	8,169		4,529	,000
prisotnost	1,883	,555	,482	3,393	,002

a. Dependent Variable: ocena

Zapišemo lahko tudi sledečo linearno zvezo med spremenljivkama:

$$ocena' = 36,998 + 1,883 \cdot prisotnost$$

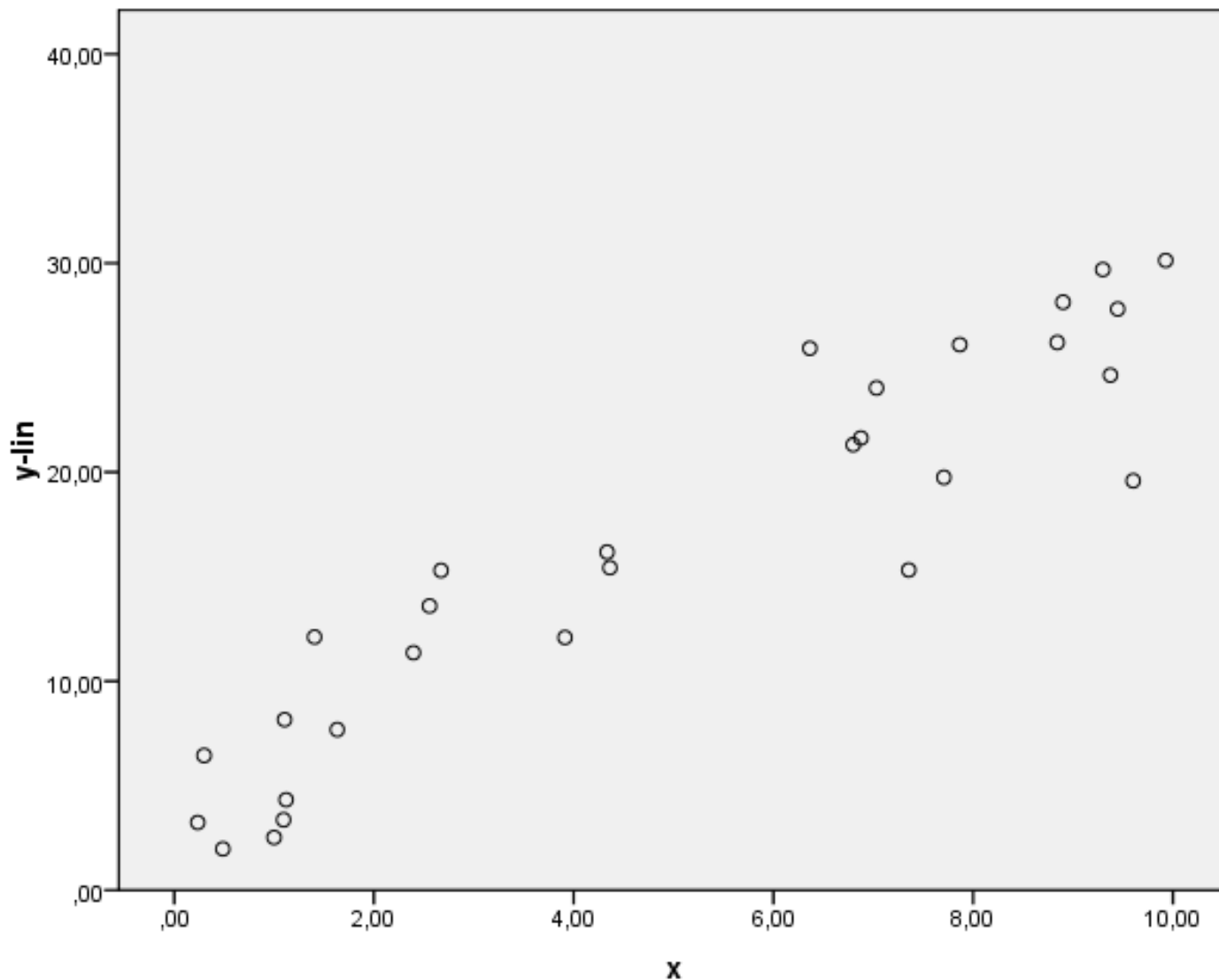
Tipi regresijskih povezanosti

Odvisna in neodvisna/e spremenljivka/e so lahko povezane na več načinov. Običajno predpostavljamo linearno povezanost, ki pa ni edina možna in vedno pravilna.

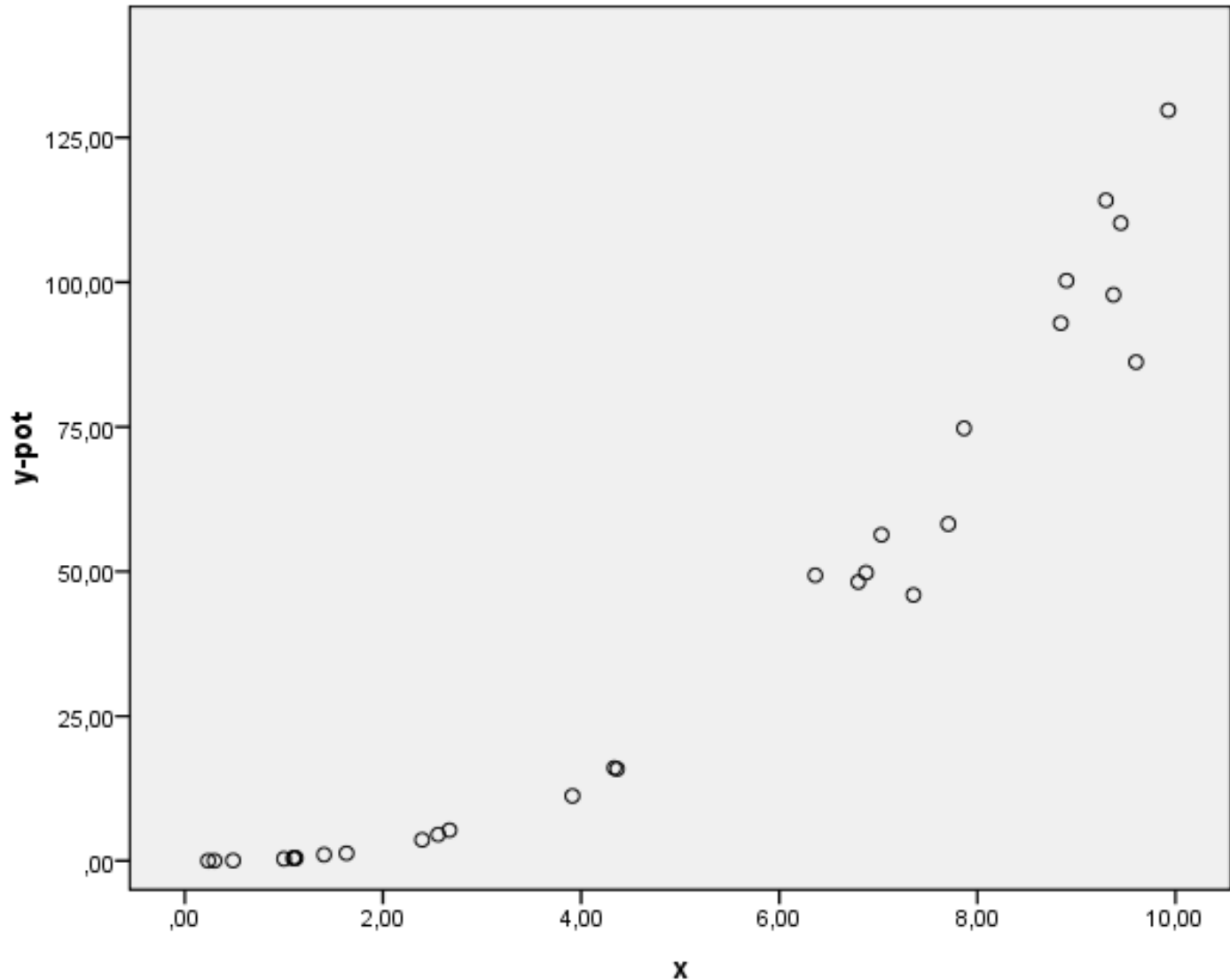
Tipi regresijskih povezanosti

- Linearna: $Y = \alpha + \beta X$
- Potenčna: $Y = \alpha X^\beta$
- Eksponentna: $Y = \alpha \beta^X$ ali $Y = \alpha e^{\beta X}$
- Recipročna: $Y = \alpha + \beta \frac{1}{X}$
- Drugo (logaritemska, korenska,)

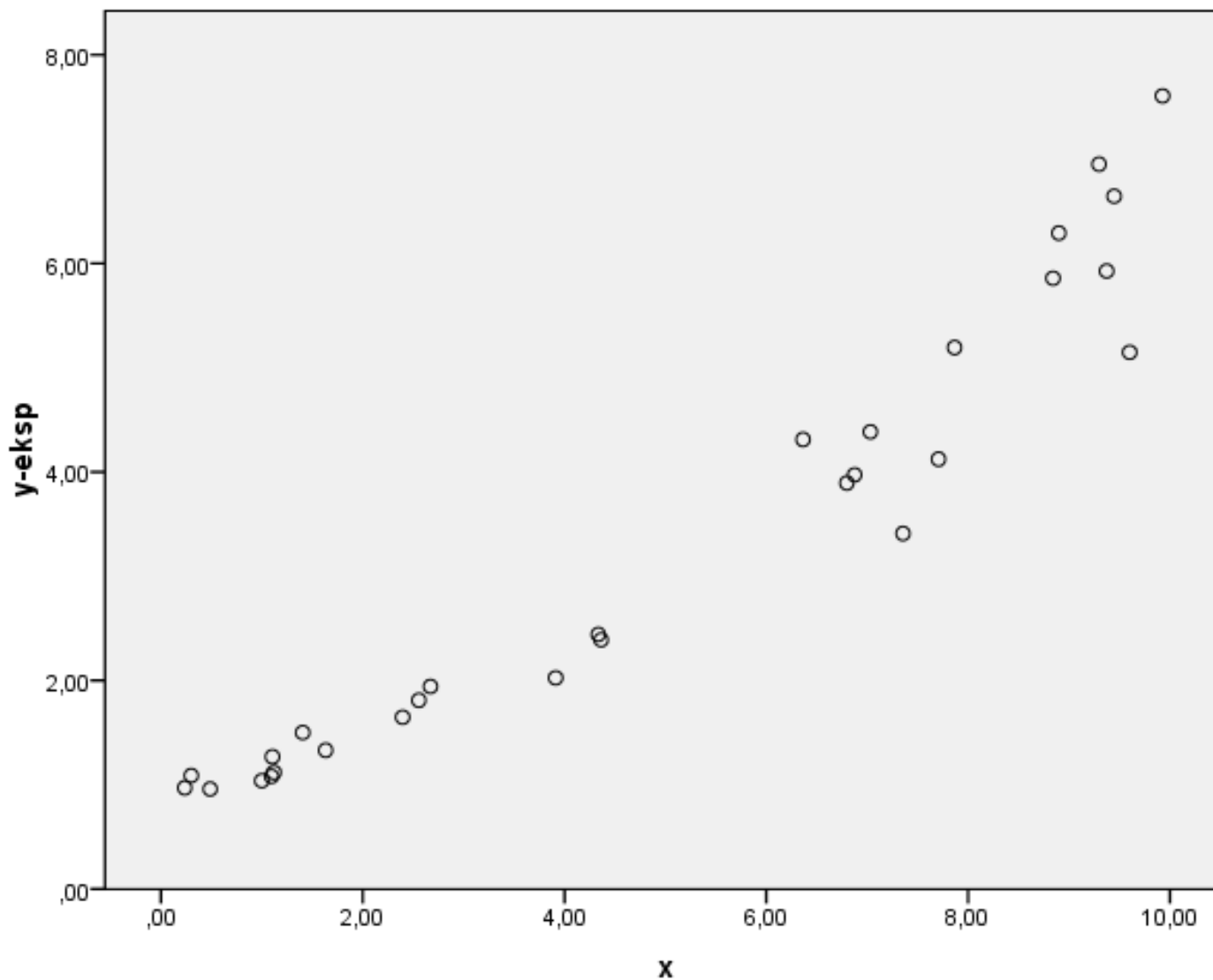
Primer 1: Linearna regresijska zveza



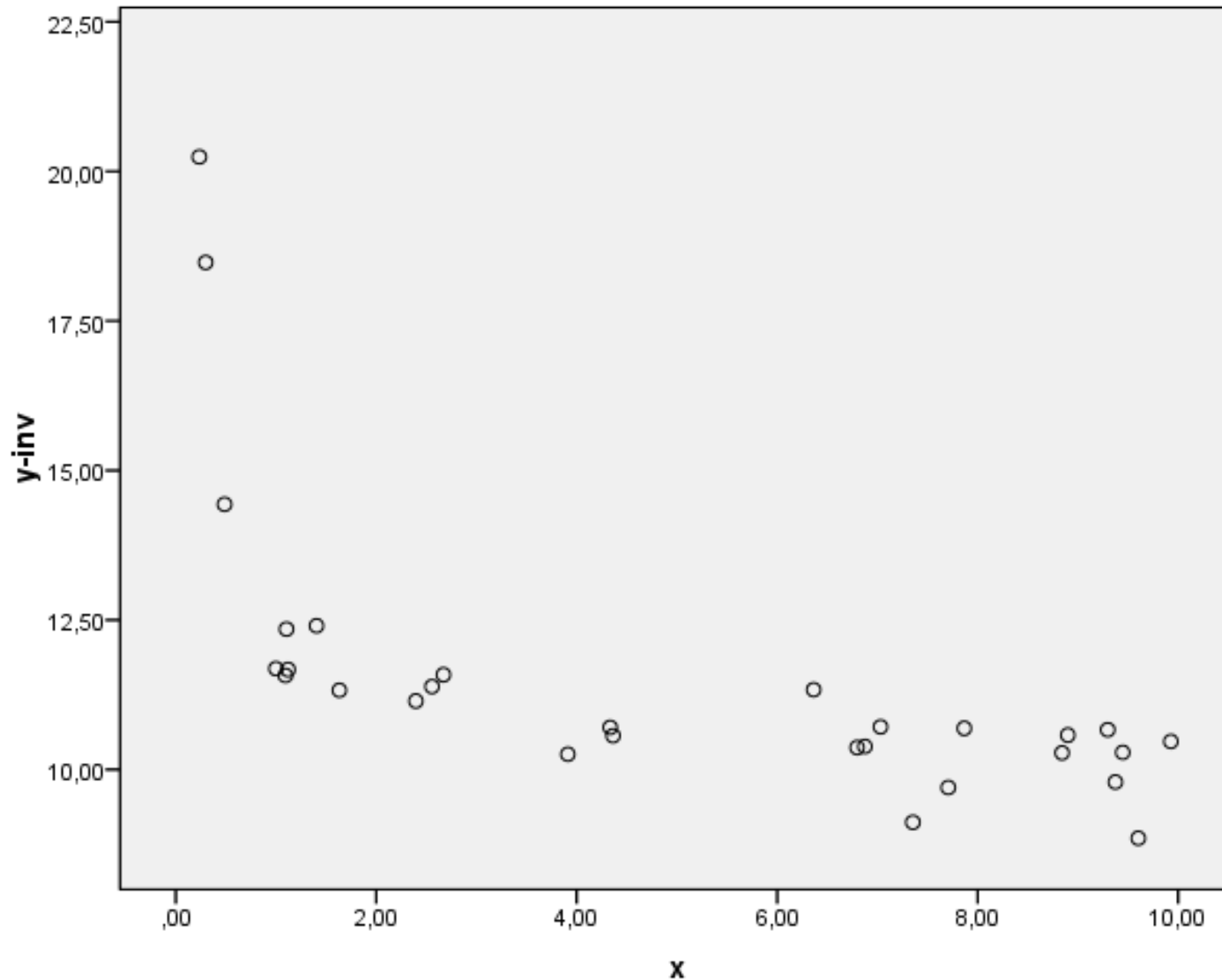
Primer 2: Potenčna regresijska zveza



Primer 3: Eksponentna regresijska zveza



Primer 4: Recipročna regresijska zveza



Transformacija spremenljivk

- Žal večina metod regresije deluje le za linearni regresijski model.
- Če regresijska zveza nelinearna, poskušamo s pomočjo transformacije spremenljivk dobiti linearno zvezo.
- Vse tri zveze nelinearne zveze iz prejšnjih prosojnic lahko prevedemo na linearni model s pomočjo transformacije spremenljivk. Nato ocenimo transformirani (linearni) model
- **Pozor:** Ocene parametrov, ki jih dobimo preko transformiranih modelov so le približne.
- Natančnejše ocene lahko dobimo z numeričnimi metodami.

Transformacija spremenljivk – potenčna zveza

- Originalna zveza: $Y = \alpha X^\beta$
- Način transformacije: logaritmujemo obe strani
- Dobljeni model: $\log Y = \log \alpha + \beta \log X$
- Novi spremenljivki: $Y_{\log} = \log Y$, $\alpha_{\log} = \log \alpha$ in $X_{\log} = \log X$
- Novi model: $Y_{\log} = \alpha_{\log} + \beta X_{\log}$
- Odvajanje po x :
$$\frac{dY}{dX} = \beta \alpha X^{\beta-1} = \beta \alpha \frac{X^\beta}{X} = \beta \frac{Y}{X} \implies \frac{dY}{Y} = \beta \frac{dX}{X}$$
- Interpretacija β koeficienta: pove nam za koliko % se spremeni Y , če se X poveča za 1 % - elastičnost.
- Postopku pravimo **popolna (dvojna) logaritemska transformacija**, regresijski modelu pa **log-log**, **dvojni log** ali **log-linearni** regresijski model.

Transformacija spremenljivk – eksponenta zveza

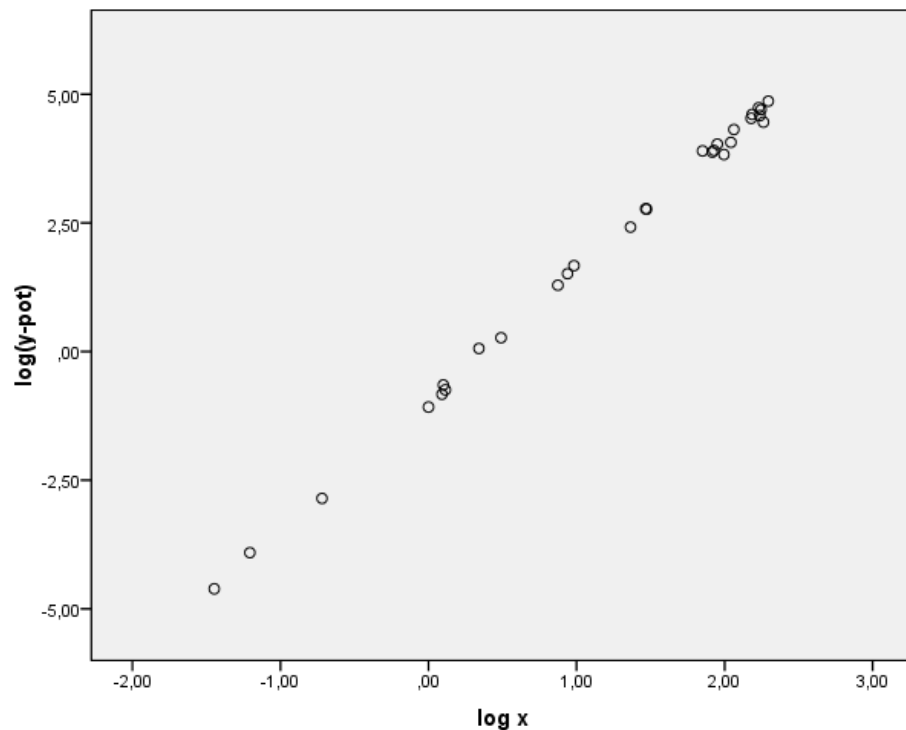
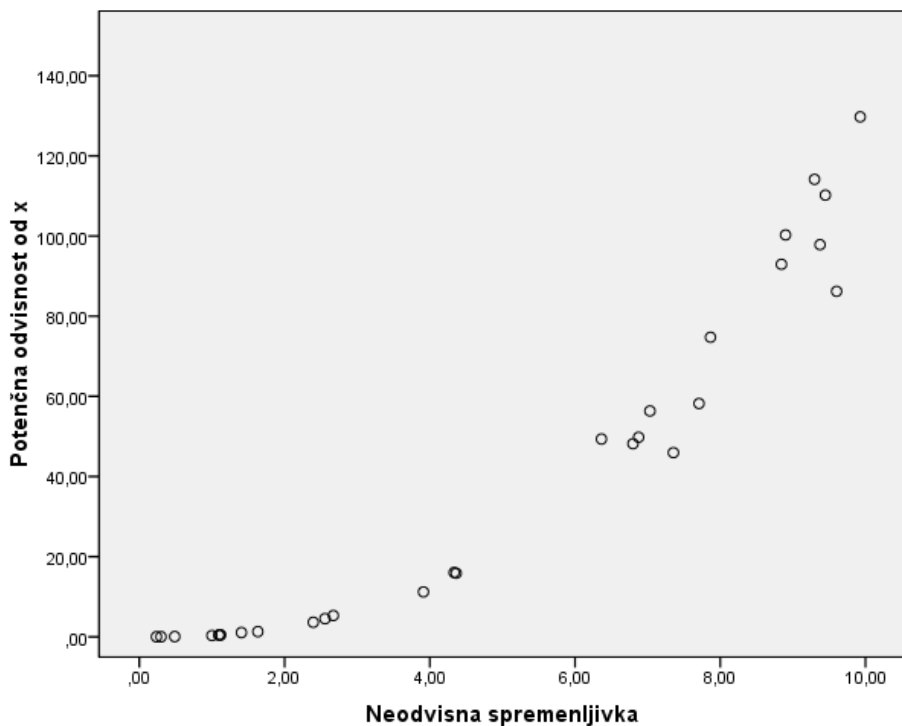
- Originalna zveza: $Y = \alpha e^{\beta X}$
- Način transformacije: logaritmujemo obe strani
- Dobljeni model: $\ln Y = \ln \alpha + \beta X$
- Novi spremenljivki: $Y_{\ln} = \ln Y$ in $\alpha_{\ln} = \ln \alpha$
- Novi model: $Y_{\ln} = \alpha_{\ln} + \beta X$
- Odvajanje po x : $\frac{dY}{dX} = \beta \alpha e^{\beta X} = \beta Y \implies \frac{dY}{Y} = \beta dX$
- Interpretacija β koeficienta: pove nam za kakšen delež (če pomnožimo s 100 dobimo %) se spremeni Y , če se X poveča za 1 enoto.
- Postopku pravimo **pollogaritemska transformacija**, ker smo logaritmerali samo odvisno spremenljivko, regresijski modelu pa **lin-log** model.

Transformacija spremenljivk – recipročna zveza

- Originalna zveza: $Y = \alpha + \beta \frac{1}{X}$
- Način transformacije: ni potrebna
- Dobljeni model: $Y = \alpha + \beta \frac{1}{X}$
- Nova spremenljivka: $X_{\text{rec}} = \frac{1}{X}$
- Novi model: $Y = \alpha + \beta X_{\text{rec}}$
- Regresijski modelu pravimo **recipročni** model.

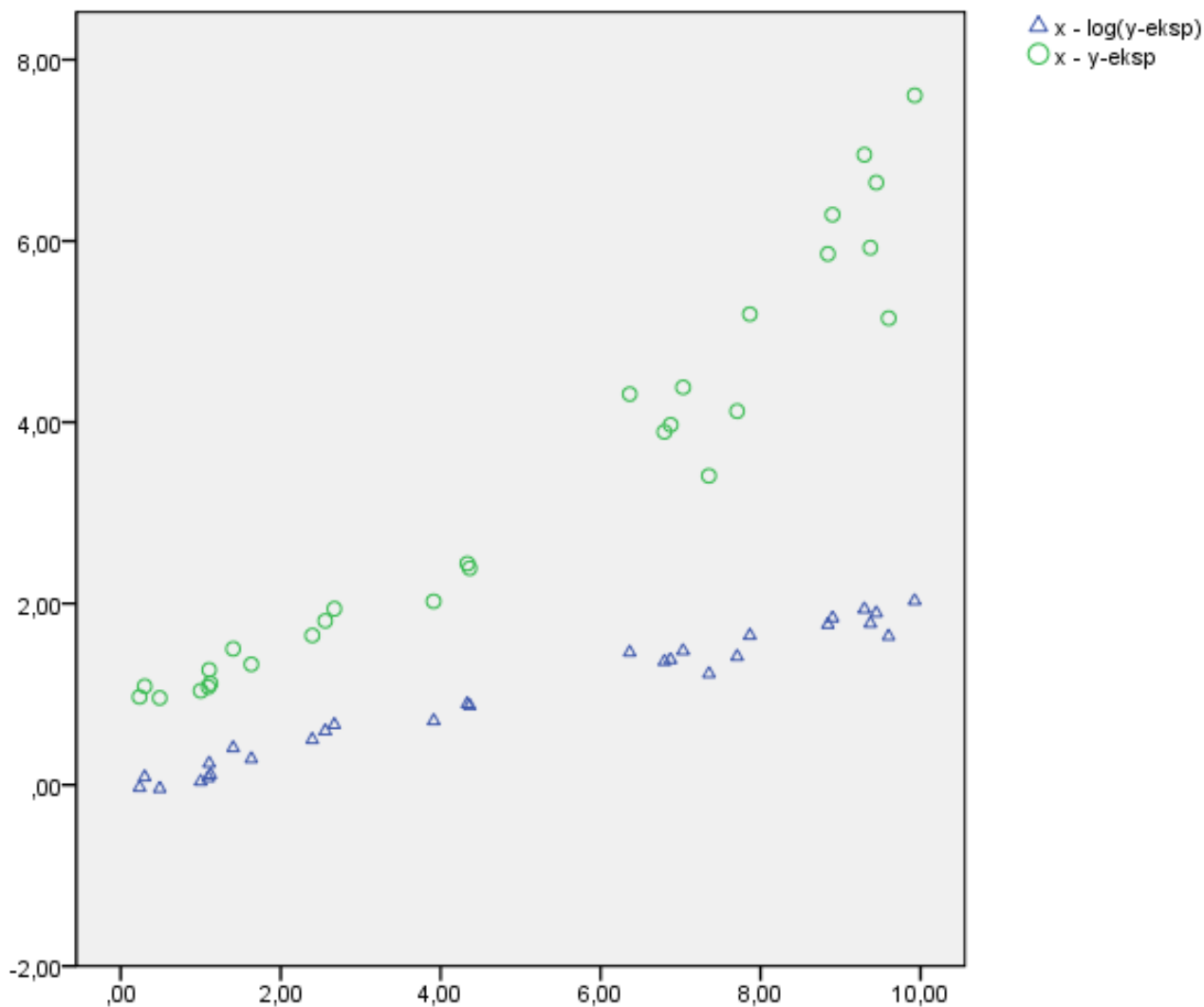
Primer: Učinek dvojne logaritemske transformacije

Obe strani enačbe prikazane na levem grafikonu smo logaritmirali

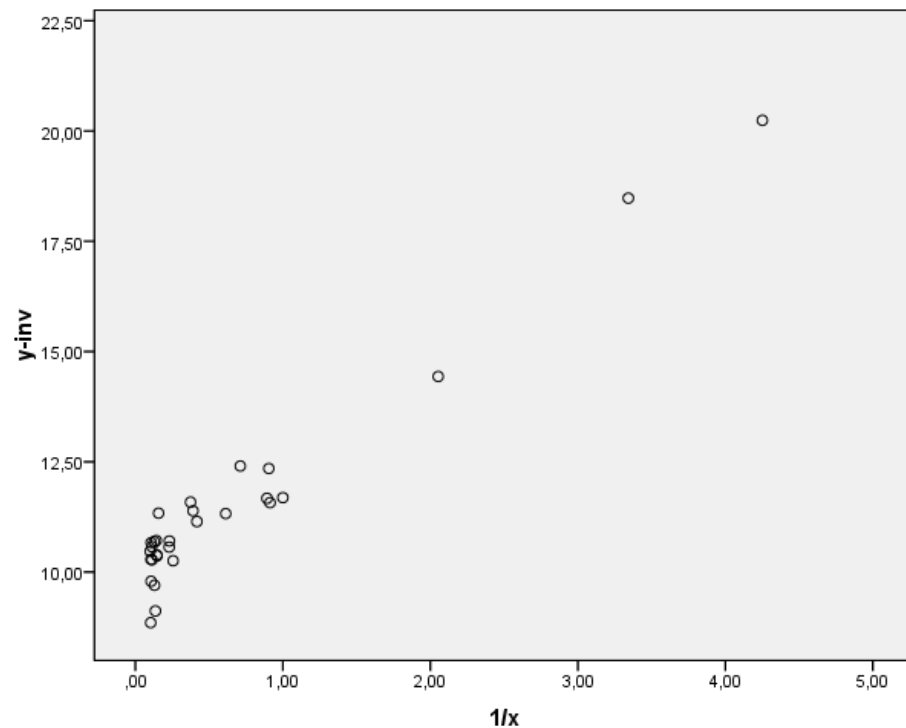
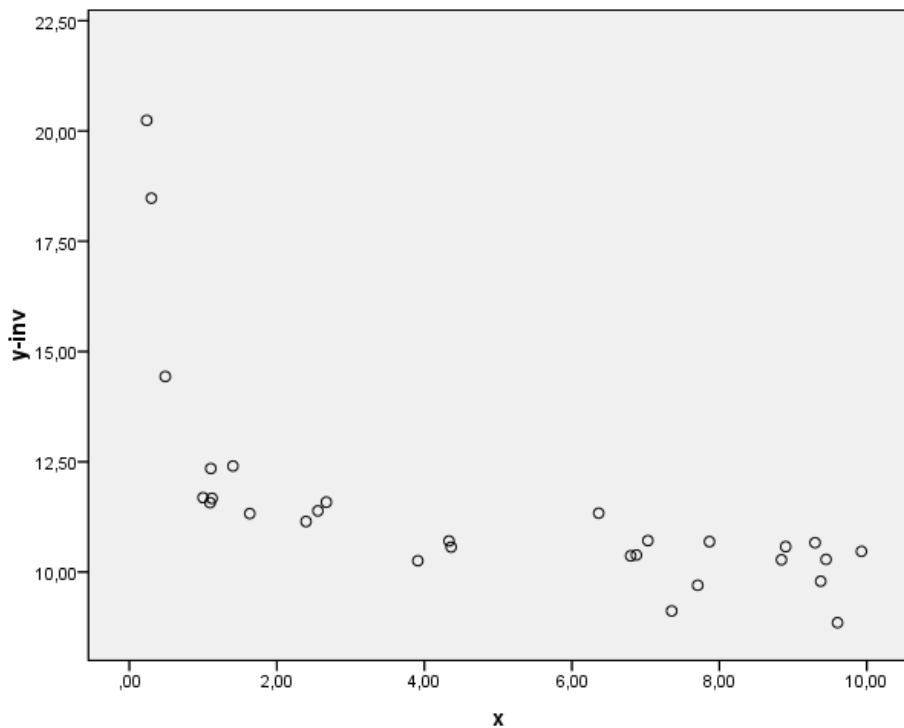


Primer: Učinek pollogaritemske transformacije

Logaritmiranje odvisne spremenljivke



Primer: Učinek recipročne transformacije



Primer: Ocena eksponentnega regresijskega modela

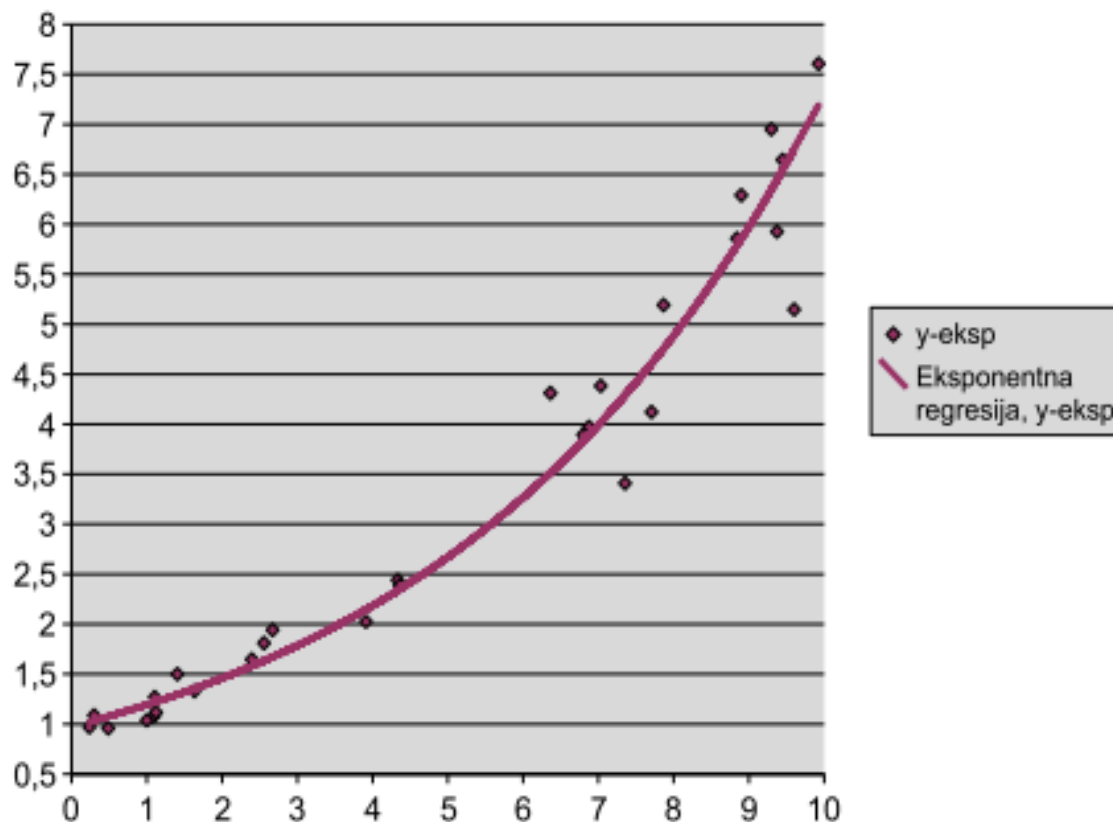
V datoteki “primeri.xls”, delovni list “Spremenljivke”, je podana ocena regresijskega modela iz Primera 3. Model ocenimo v naslednjih korakih:

1. Spremenljivki ustrezno transformiramo: v našem primeru logaritmujemo odvisno spremenljivko Y (y -eksp) in kot rezultat uvedemo novo spremenljivko $Y_{log} = \log Y$.
2. Upoštevamo linearno zvezo med spremenljivkama X in Y_{log} ter izračunamo regresijske koeficiente:
$$Y'_{log} = -0,026 + 0,2X$$
3. Z antilogaritmiranjem dobimo zvezo med originalnima spremenljivkama:
$$Y' = e^{-0,026} e^{0,2X} = 0,975 e^{0,2X}.$$

Primer: Napovejmo vrednost spremenljivke Y , če ima spremenljivka X vrednost 6

$$Y'(X = 6) = 0,975 e^{1,2} = 3,24.$$

Eksponentna regresijska krivulja



Interpretacija koeficientov v eksponentni regresijski zvezi

- Interpretacije koeficientov pri nelinearnih zvezah so zahtevnejše kot pri linearni zvezi. Spomnimo se, da naklon regresijske premice v linearni zvezi pomeni spremembo vrednosti odvisne spremenljivke, če se neodvisna spremenljivka poveča za eno enoto. To povečanje pa je neodvisno od trenutne vrednosti odvisne ali neodvisne spremenljivke.
- Pri eksponentni regresijski zvezi je sprememba odvisne spremenljivke, ko se neodvisna poveča za eno enoto, odvisna od še od tega, kolikšna je trenutna vrednost odvisne in neodvisne spremenljivke. Pri tem si lahko pomagamo z odvodom.
- Odvajajmo regresijsko zvezo $Y' = 0,975 e^{0,2X}$:

$$\frac{dY'}{dX} = 0,2 \cdot 0,975 e^{0,2X} = 0,2 Y' \implies \frac{dY'}{Y'} = 0,2 dX$$

- Naklonski koeficient tangente na regresijsko krivuljo je torej enak $0,2 \cdot Y$. Ugotovimo torej, da pomeni povečanje neodvisne spremenljivke za eno enoto, povečanje odvisne spremenljivke za približno 0,2-krat, oziroma za eno petino.

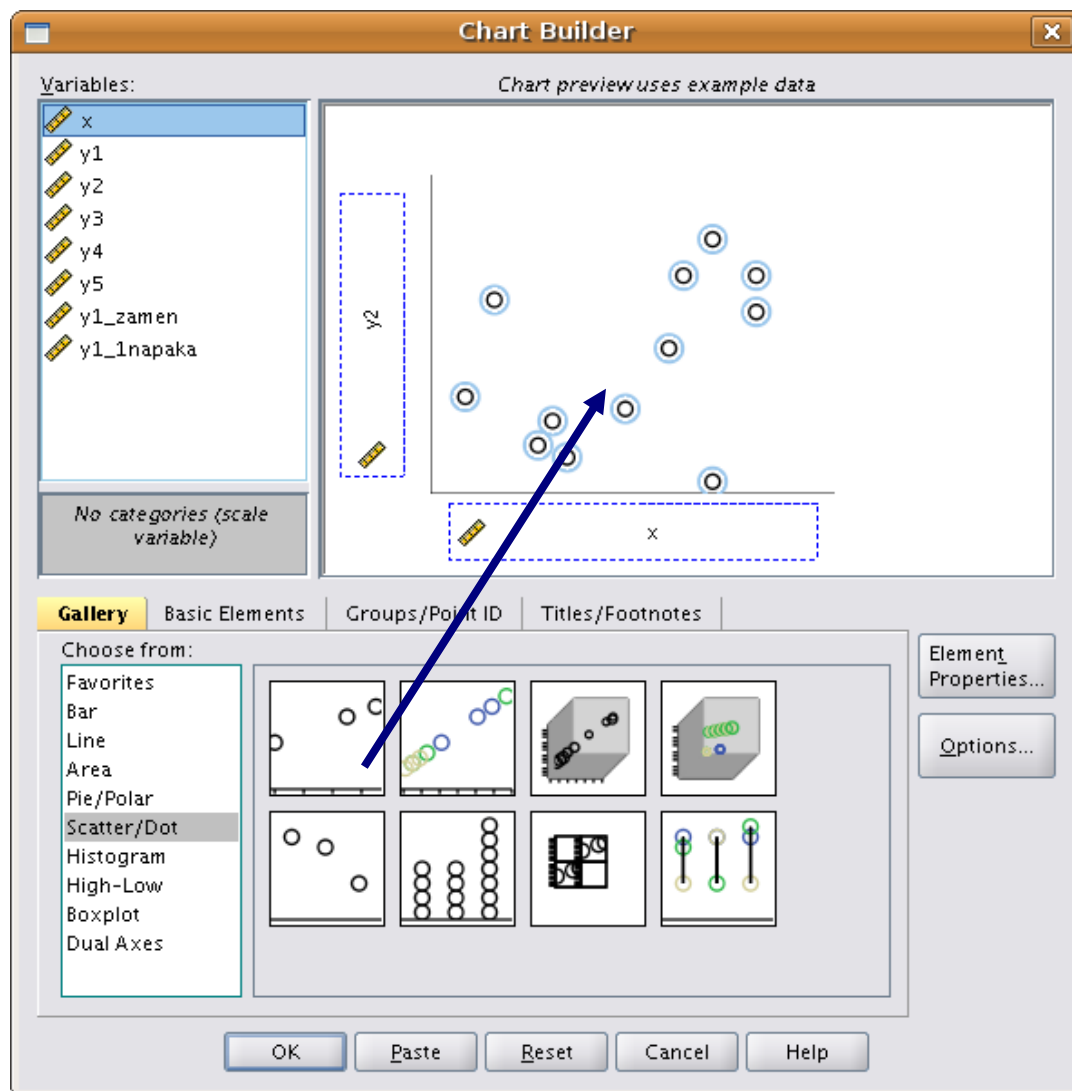
Ocenjevanje regresijskih parametrov s programom SPSS

1. Določimo odvisne in neodvisne spremenljivke in jih po potrebi transformiramo (ukaz: *Transform - Compute*) tako, da med njimi velja linearna zveza. Pri tem si lahko pomagamo z razsevnim grafikonom (ukaz: *Graphs – Chart Builder*).
2. V meniju izberemo metodo *Analyze – Regression - Linear*. V polje *Dependent* prenesemo odvisno, v polje *Independents* pa neodvisno spremenljivko. Kliknemo gumb OK in počakamo na rezultate izračuna.
3. Rezultati se izpišejo v novem oknu. V izpisu se zaenkrat omejimo na polje *Coefficients*. Regresijski koeficienti so zapisani v stolpcu *Unstandardized coefficients - B*.

Primer ocenjevanja nelinearne zveze

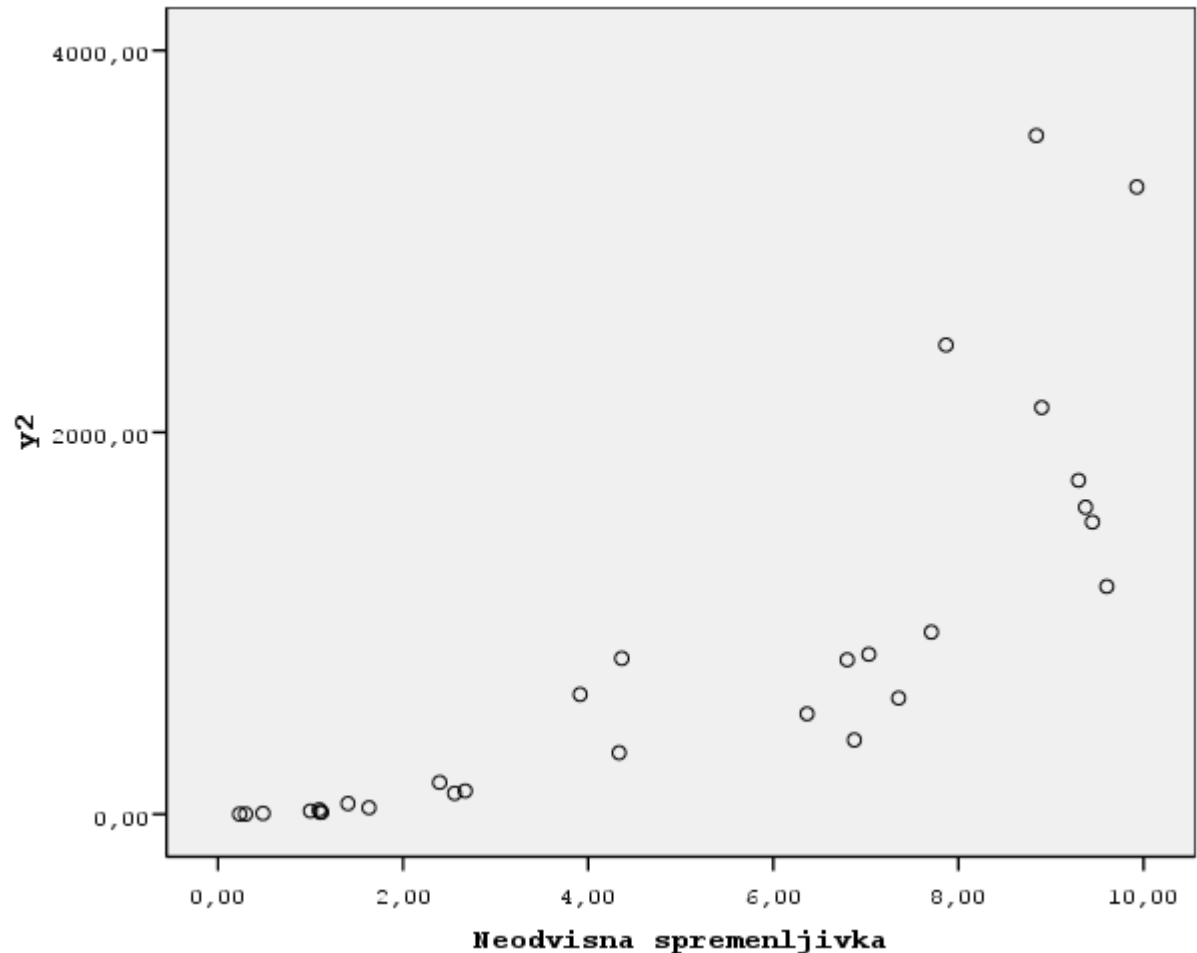
Uporabite datoteko nelinegr-podatki.sav in analizirajte zvezo med odvisno spremenljivko y_2 in neodvisno spremenljivko x . Napovejte vrednost spremenljivke y_2 pri $x = 2$

1. Ugotovimo obliko zveze med spremenljivkama. Pri tem si pomagamo z razsevnim grafikonom (ukaz: *Graphs – Chart Builder*, Izberemo skupino grafov *Scatter/Dot* in v njej graf *Simple Scatter*).



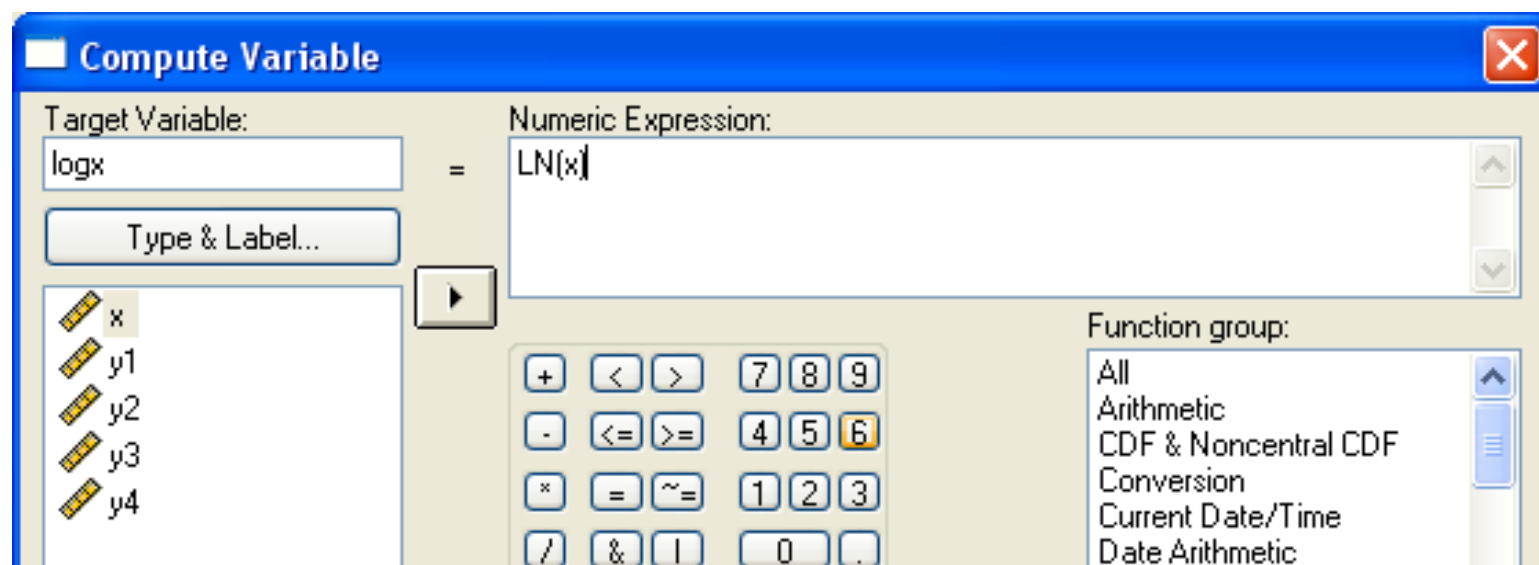
Primer ocenjevanja nelinearne zveze

- Zveza med spremenljivkama je očitno nelinearna.
- Glede na obliko lahko sklepamo, da gre za potenčno ali eksponentno zvezo.



Primer ocenjevanja nelinearne zveze

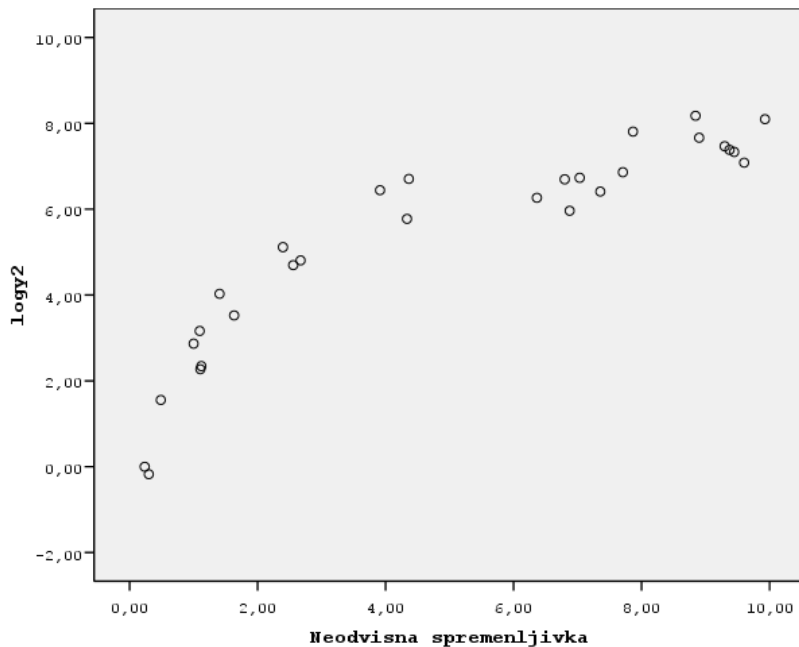
V primeru potenčne zveze moramo logaritmirati obe spremenljivki (odvisno in neodvisno), v primeru eksponentne pa samo odvisno. To storimo z ukazom *Transform – Compute*.



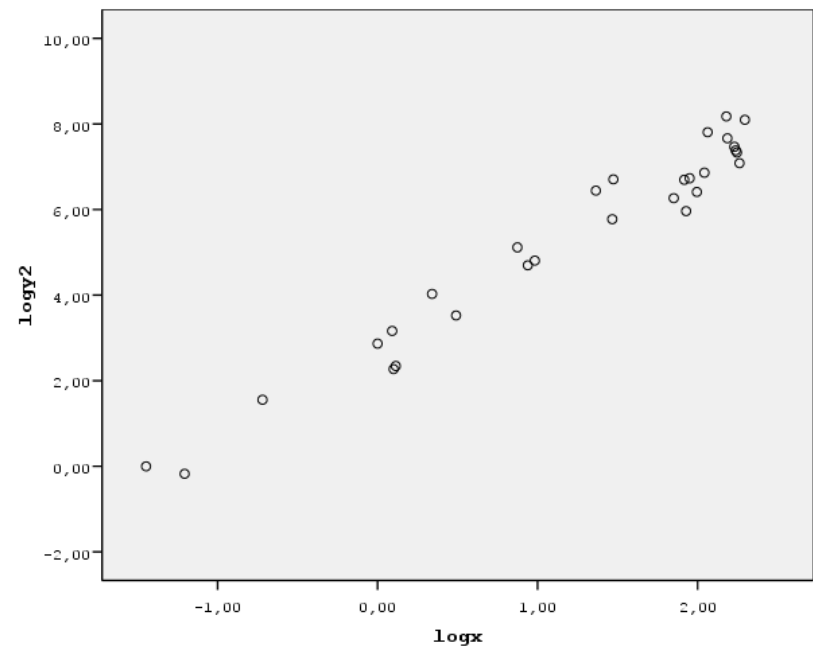
Primer ocenjevanja nelinearne zveze

Ponovno narišemo razsevna grafikona za obe zvezi (eksponentno in potenčno).

Eksponentna zveza



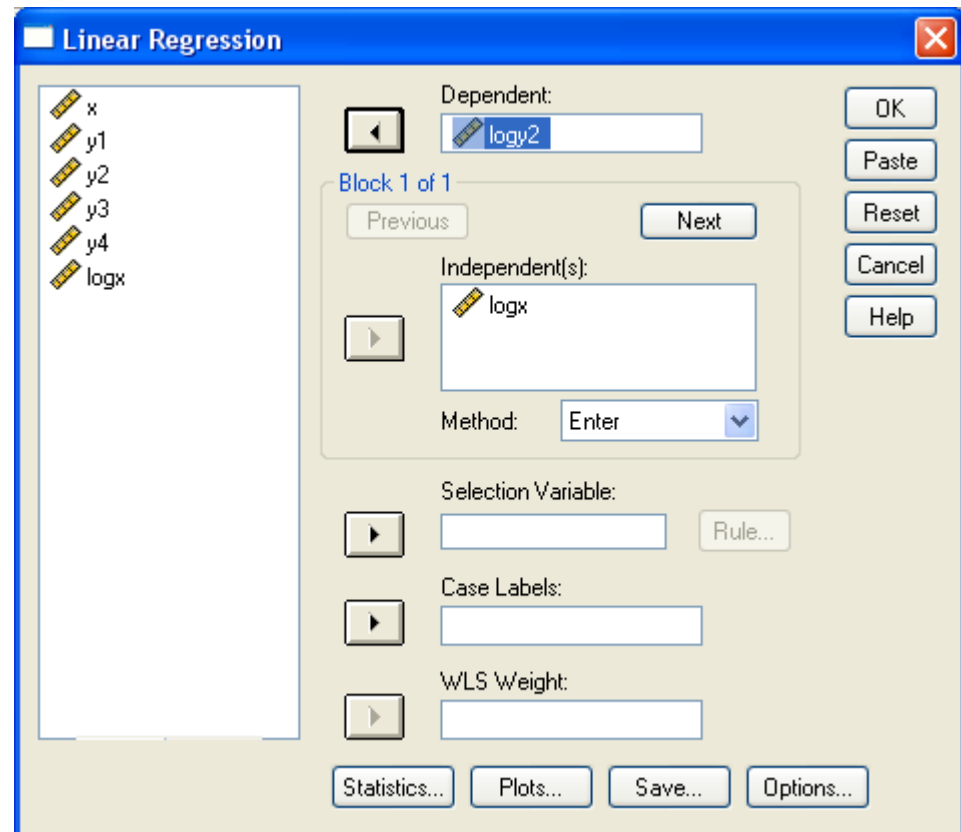
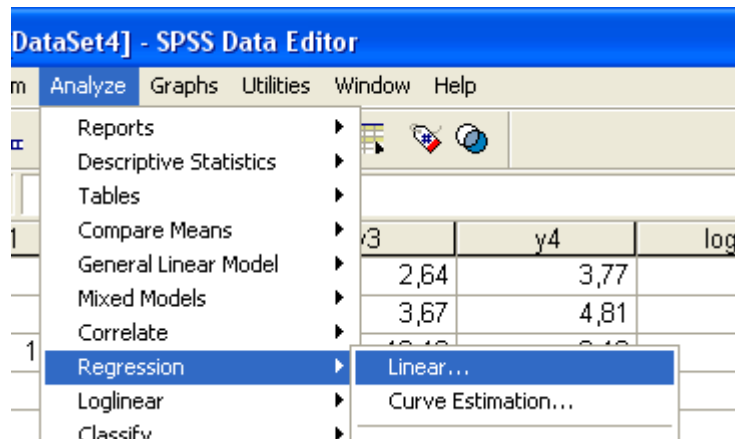
Potenčna zveza



Očitno je med spremenljivkama y_2 in x potenčna zveza (da smo dobili linearno povezavo smo morali logaritmerati obe spremenljivki)

Primer ocenjevanja nelinearne zveze

2. Ocenimo linearno regresijo na transformiranih spremenljivkah.



Primer ocenjevanja nelinearne zveze

3. Pregledamo rezultate (zaenkrat se omejimo na polje *Coefficients*) in zapišemo regresijsko enačbo.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2,773	,132		21,038	,000
	logx	2,121	,082	,980	25,915	,000

a. Dependent Variable: logy2

Enačba s transformiranimi spremenljivkama

$$\log y_2' = a + b \log x$$

$$\log y_2' = 2,773 + 2,121 \log x$$

Enačba z originalnima spremenljivkama:

$$y_2' = e^a e^{b \log x}$$

$$y_2' = e^{2,773} e^{2,121 \log x} = 16,01 x^{2,121}$$

Primer ocenjevanja nelinearne zveze

4. Napoved vrednosti spremenljive y^2 pri $x = 2$.

Vrednost vstavimo v prej nastavljeno enačbo:

Enačba z originalnima spremenljivkama:

$$y^2'(x=2) = 16,01 x^{2,121} = 16,01 \cdot 2^{2,121} = 69,64$$

Primer ocenjevanja nelinearne zveze

5. Interpretacija β koeficienta. Interpretacijo dobimo z odvajanjem enačbe z originalnimi spremenljivkami.

Enačba z originalnima spremenljivkama:

$$y_2' = 16,01 x^{2,121}$$

Odvajamo:

$$\frac{d y_2'}{d x} = 2,121 \cdot 16,01 x^{2,121-1} = 2,121 \frac{16,01 x^{2,121}}{x} = 2,121 \frac{y_2'}{x}$$

$$\frac{d y_2'}{y_2'} = 2,121 \frac{d x}{x}$$

Če se x poveča za 1 %, se y_2 v povprečju poveča za 2,121%.

Pogoji za uporabo transformacij

- Uporaba logaritemske transformacije je možna le na spremenljivkah, ki so strogo pozitivne (logaritem ni definiran za negativne vrednosti in vrednost 0)
- Logaritmiranje ima smisla le na razmernostnih spremenljivkah, saj drugače povečanje za določen delež trenutne vrednosti nima smisla.

Dodatno: numerično ocenjevanje

- Natančneje lahko vrednosti parametrov v nelinearnih modelih (regresijskih zvezah) ocenimo preko numerične aproksimacije.
- SPSS ponuja to možnost pod *Analyze – Regression – Nonlinear*.
- V polje *Dependent* prenesemo odvisno spremenljivko (y_2), v polje *Model Expression* pa vpišemo celotno desno stran regresijske enačbe brez člena napake (ax^b).
- Parametre (a in b) moramo posebej definirati v oknu *Parameters* ter jim nastaviti začetne vrednosti, ki jih običajno vzamemo iz transformiranega modela.

Dodatno: numerično ocenjevanje

The image shows two windows from a software application. The main window is titled "Nonlinear Regression" and contains the following elements:

- Dependent:** A dropdown menu showing "y2".
- Model Expression:** A text box containing the formula $a * x ** b$.
- Function group:** A list box with categories: All, Arithmetic, CDF & Noncentral CDF, Conversion, Current Date/Time, Date Arithmetic, Date Creation.
- Functions and Special Variables:** An empty text box.
- Buttons:** Loss..., Constraints..., Save..., Options..., OK, Paste, Reset, Cancel, Help.
- Calculator:** A numeric keypad with operators (+, -, *, /, **, ~, (), <, >, <=, >=, =, **), digits (0-9), and a Delete button.
- Variable List:** A list of variables: x, y1, y2, y3, y4, y5, y1_zamen, y1_1napaka.
- Parameters...** button: A button at the bottom of the variable list that opens the "Nonlinear Regression: Parameters" dialog.

The "Nonlinear Regression: Parameters" dialog box is open, showing:

- Name:** An empty text box.
- Starting Value:** An empty text box.
- Parameter List:** A list containing "a(16,01)" and "b(2,121)".
- Buttons:** Add, Change, Remove, Continue, Cancel, Help.
- Checkbox:** Use starting values from previous analysis.

A blue arrow points from the "Parameters..." button in the main dialog to the "Nonlinear Regression: Parameters" dialog.

Dodatno: numerično ocenjevanje

Iteration History^b

Iteration Number ^a	Residual Sum of Squares	Parameter	
		a	b
1.0	8609766,183	16,010	2,121
1.1	9317384,953	7,946	2,406
1.2	8161247,817	14,386	2,221
2.0	8161247,817	14,386	2,221
2.1	8259963,657	10,027	2,360
2.2	8155726,296	13,755	2,241
3.0	8155726,296	13,755	2,241
3.1	8147033,609	12,555	2,281
4.0	8147033,609	12,555	2,281
4.1	8146407,473	10,372	2,363
5.0	8146407,473	10,372	2,363
5.1	8133360,746	9,548	2,407
6.0	8133360,746	9,548	2,407
6.1	8132790,359	9,254	2,423
7.0	8132790,359	9,254	2,423
7.1	8132779,353	9,227	2,425
8.0	8132779,353	9,227	2,425
8.1	8132779,334	9,223	2,425

Derivatives are calculated numerically.

a. Major iteration number is displayed to the left of the decimal, and minor iteration number is to the right of the decimal.

b. Run stopped after 18 model evaluations and 8 derivative evaluations because the relative reduction between successive residual sums of squares is at most SCON = 1,00E-008.

Parameter Estimates

Parameter	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
a	9,223	13,129	-17,714	36,161
b	2,425	,652	1,087	3,762

Interpretacija β koeficienta.

Če se x poveča za 1 %, se y_2 v povprečju poveča za 2,425%.

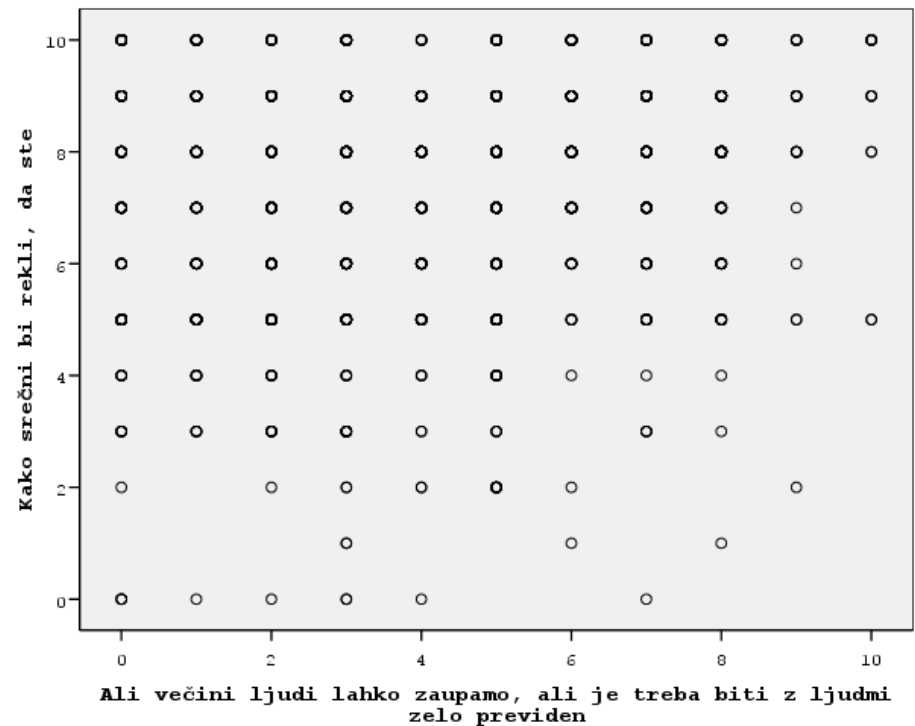
Vidimo torej, da se lahko rezultat precej spremeni. Ta ocena je bolj natančna, saj upošteva *pravi* model.

Dodatek za diskretne spremenljivke

- Razsevni grafikon ni primerne za prikaz podatkov oz. za ugotavljanje oblike povezanosti za diskretne spremenljivke z omejenim številom možnih vrednosti (recimo skala od 1-5, 0-10,...),
- Namesto njega uporabilo grafikon, kjer glede na vrednost neodvisne predstavimo povprečja odvisne spremenljivke.
- Poleg povprečji z “intervali” prikažemo tudi variabilnost vrednosti odvisne spremenljivke pri vsaki vrednosti neodvisne spremenljivke.

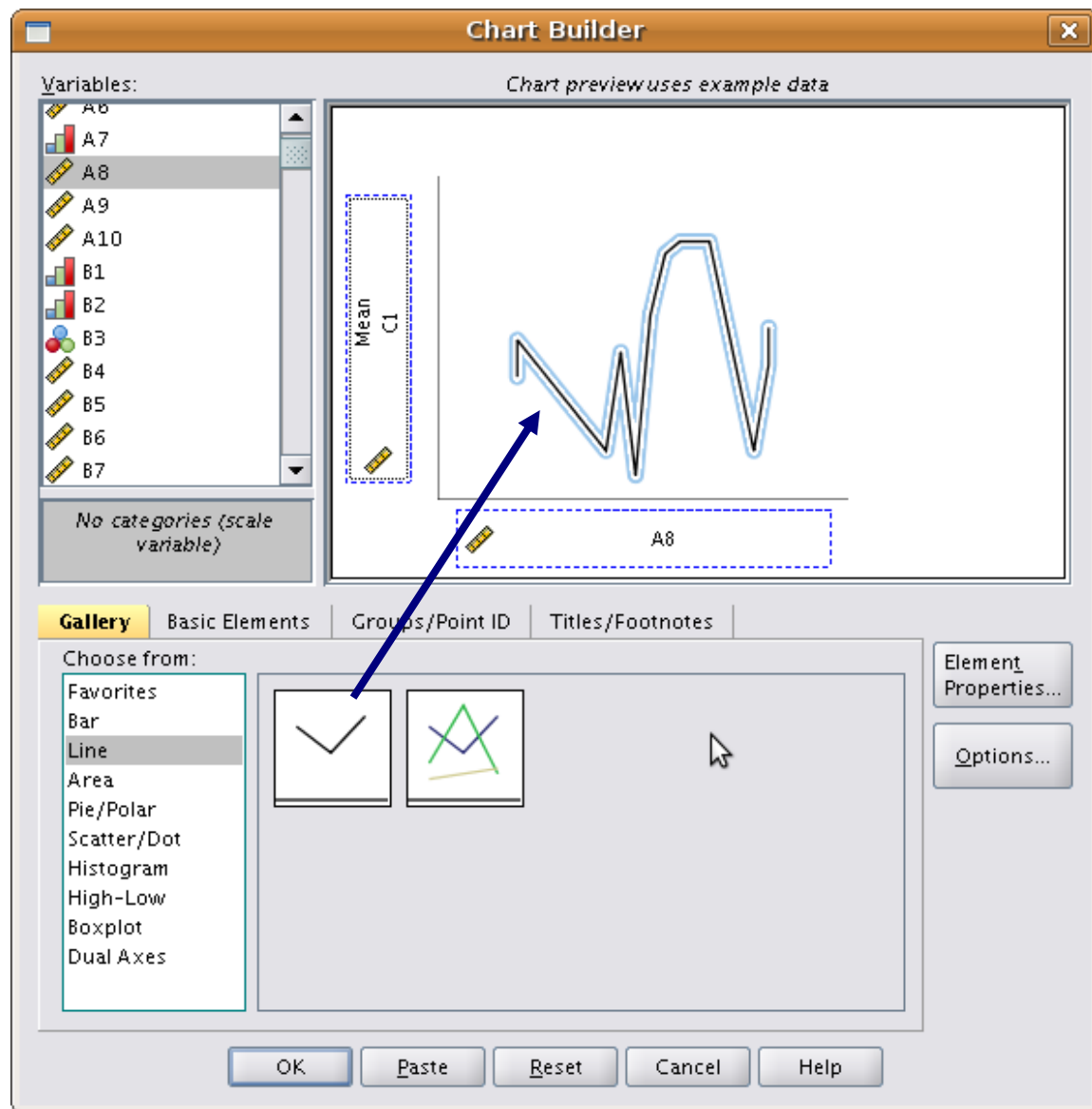
Primer grafičnega prikaza za diskretne spremenljivke

- Na podatkih ESS za leto 2004 za Slovenijo bomo grafično prikazali odvisnost sreče posameznika (C1) od zaupanja ljudem (A8).
- Predvidevamo, da so bolj zaupljivi ljudje tudi bolj srečni.
- Obe spremenljivki sta diskretni z mersko lestvico od 0-10. Razsevni grafikon zato ni primerne prikaz, kot se lahko prepričamo na levi sliki.
- Iz njega ne moremo ugotoviti oblike povezanosti.



Primer grafičnega prikaza za diskretne spremenljivke

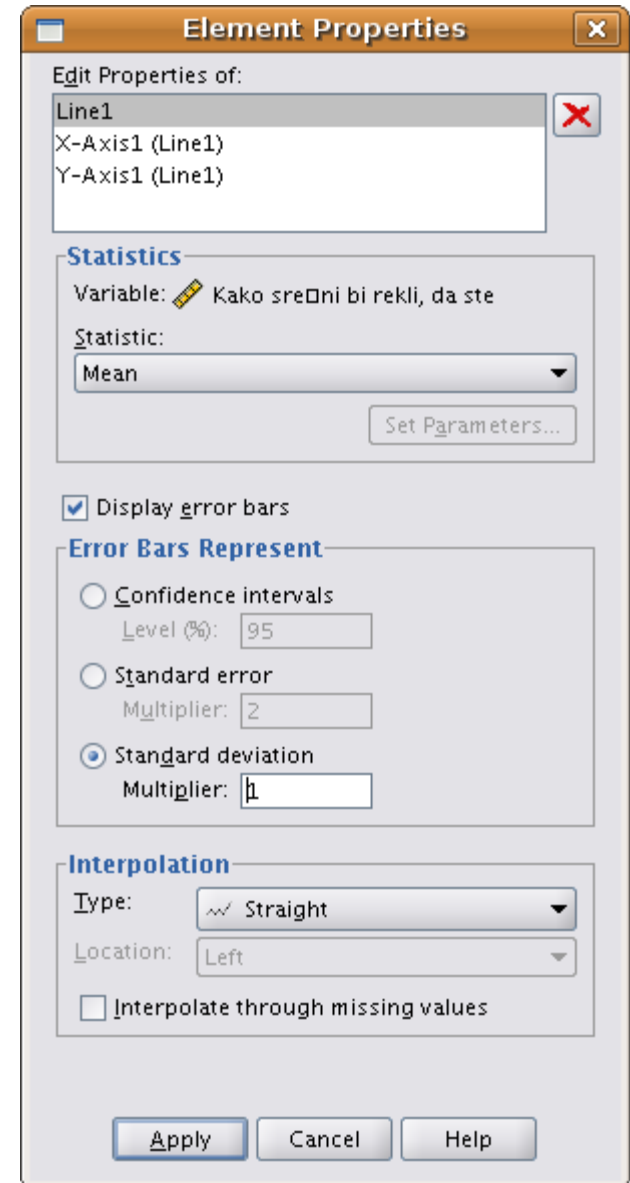
1. Ugotovimo obliko zveze med spremenljivkama. Pri tem si pomagamo grafikonom povprečij (ukaz: *Graphs – Chart Builder*):
 - Izberemo kategorijo *Line* in potem najenostavnejši graf (*Simple Line*)
 - Odvisno spremenljivko prenesemo damo na y os, neodvisno pa na x os.



Primer grafičnega prikaza za diskretne spremenljivke

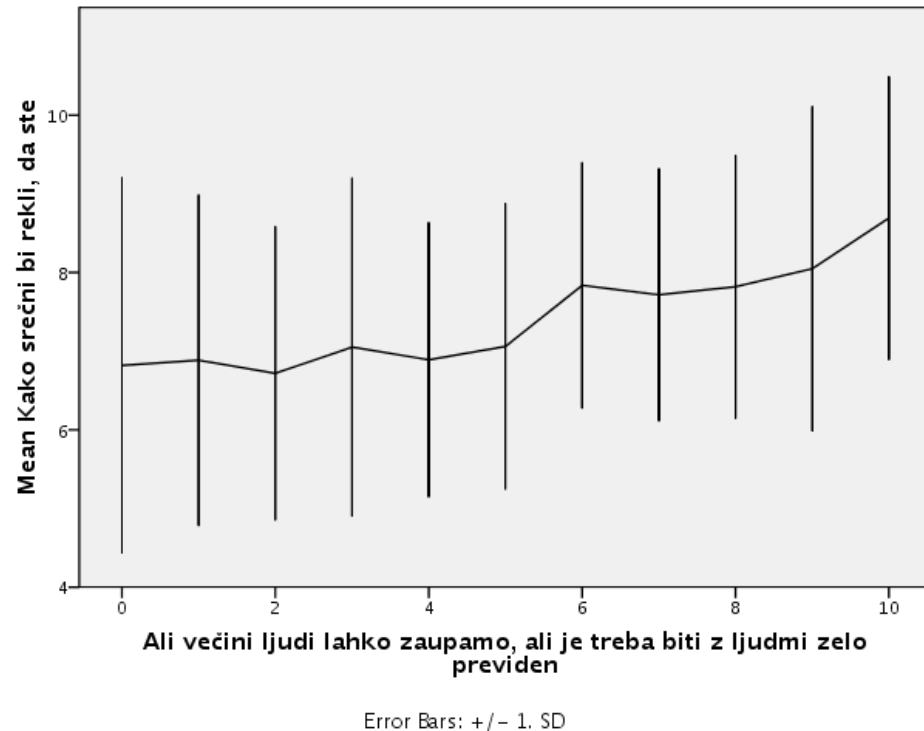
1. ... nadaljevanje

- V oknu *Element Properties* (če ni že odprto, ga odpremo s klikom na gumb *Element Properties* v oknu *Chart Builder*) izberemo v zgornjem oknu *Line1* (kar je sicer ponavadi že izbrano)
- Odključamo *Display Error Bars*
- Pod *Bar represents* izberemo *Standard Deviation*. V okence vnesemo neko vrednost, ki ni prevelika (npr 1.) Želimo namreč, da te kljuge (error bars) ne zavzamejo preveč prostora – po potrebi lahko to sicer tudi spreminjam potem v grafu.
- Kliknemo *Apply* in v glavnem oknu (*Chart Builder*) *OK*



Primer grafičnega prikaza za diskretne spremenljivke

- Na tem grafikonu je očitno, da je zveza med spremenljivkama pozitivna. Tisti ki bolj zaupajo so v povprečju tudi bolj srečni.
- Zveza med spremenljivkama je približno linearna.
- Variabilnost spremenljivke Y je pri vsaki vrednosti spremenljivke X približno enaka.



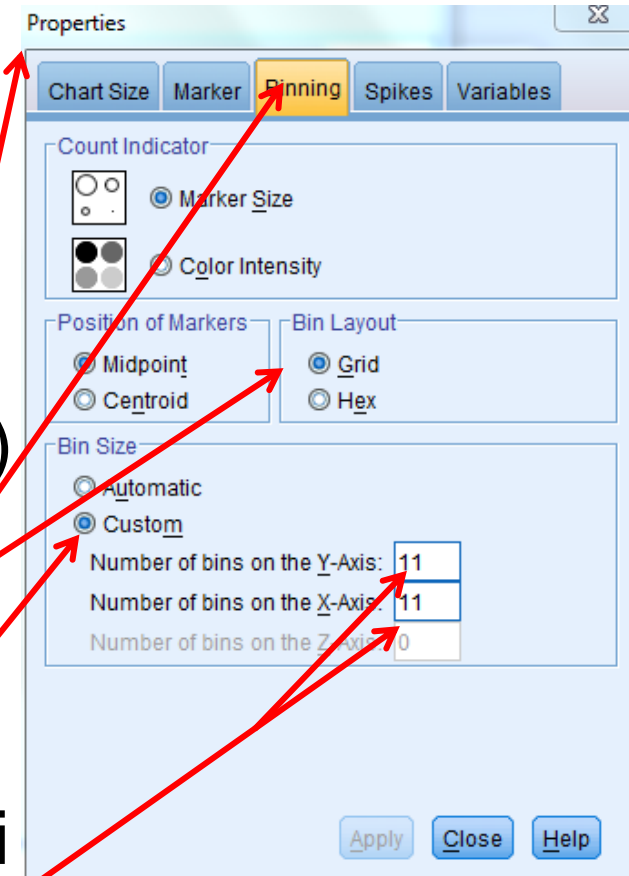
Razsevni prikaz za diskretne spremenljivke

- Alternativna možnost je tudi, da prilagodimo razsevni diagram tako, da se „vidi“, koliko točk je na posameznem delu.
- To lahko naredimo na več načinov:
 1. Velikost pike naredimo sorazmerno številu enot, ki so tam
 2. Zatresemo spremenljivke

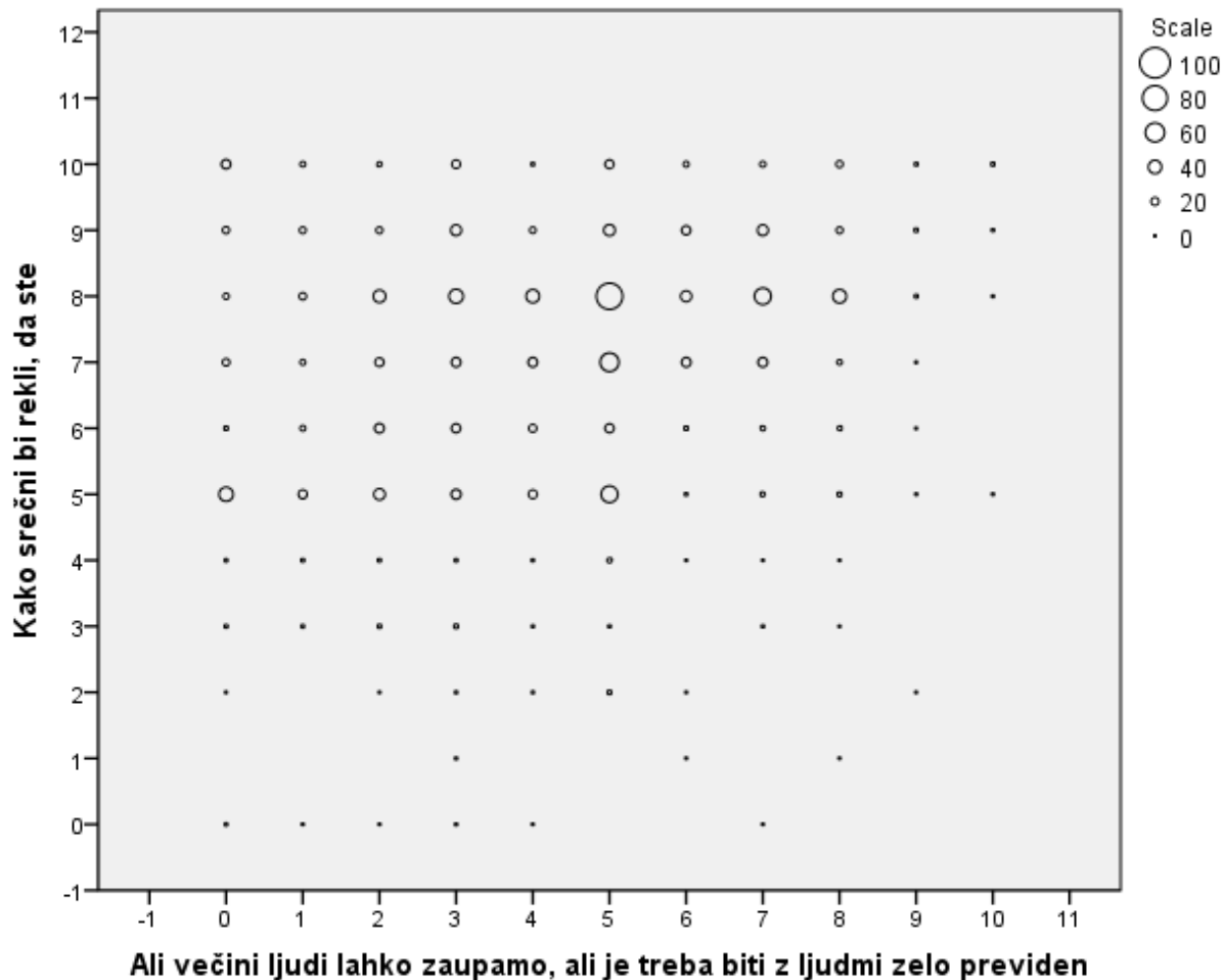
Velikost pike naredimo sorazmerno številu enot, ki so tam

- Narišemo razsevni grafikon kot običajno
- Dvokliknemo graf za urejanje
- Izberemo *Options-Bin Elements*
- V okenčku, ki se odpre (*Properties*) v zavihku *Binning* (privzeto odprto) nastavimo:

- *Bin Layout* na *Grid*
- *Bin Size* na *Custom* ter na vsaki osi število razredov (*bins*) na toliko, koliko je kategorij pri ustrezni spremenljivki.

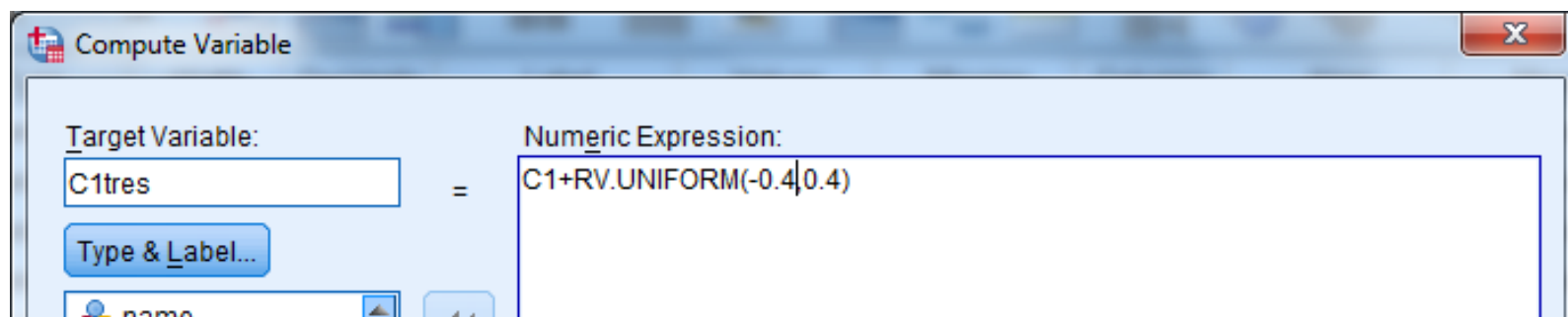


Velikost pike naredimo sorazmerno številu enot, ki so tam

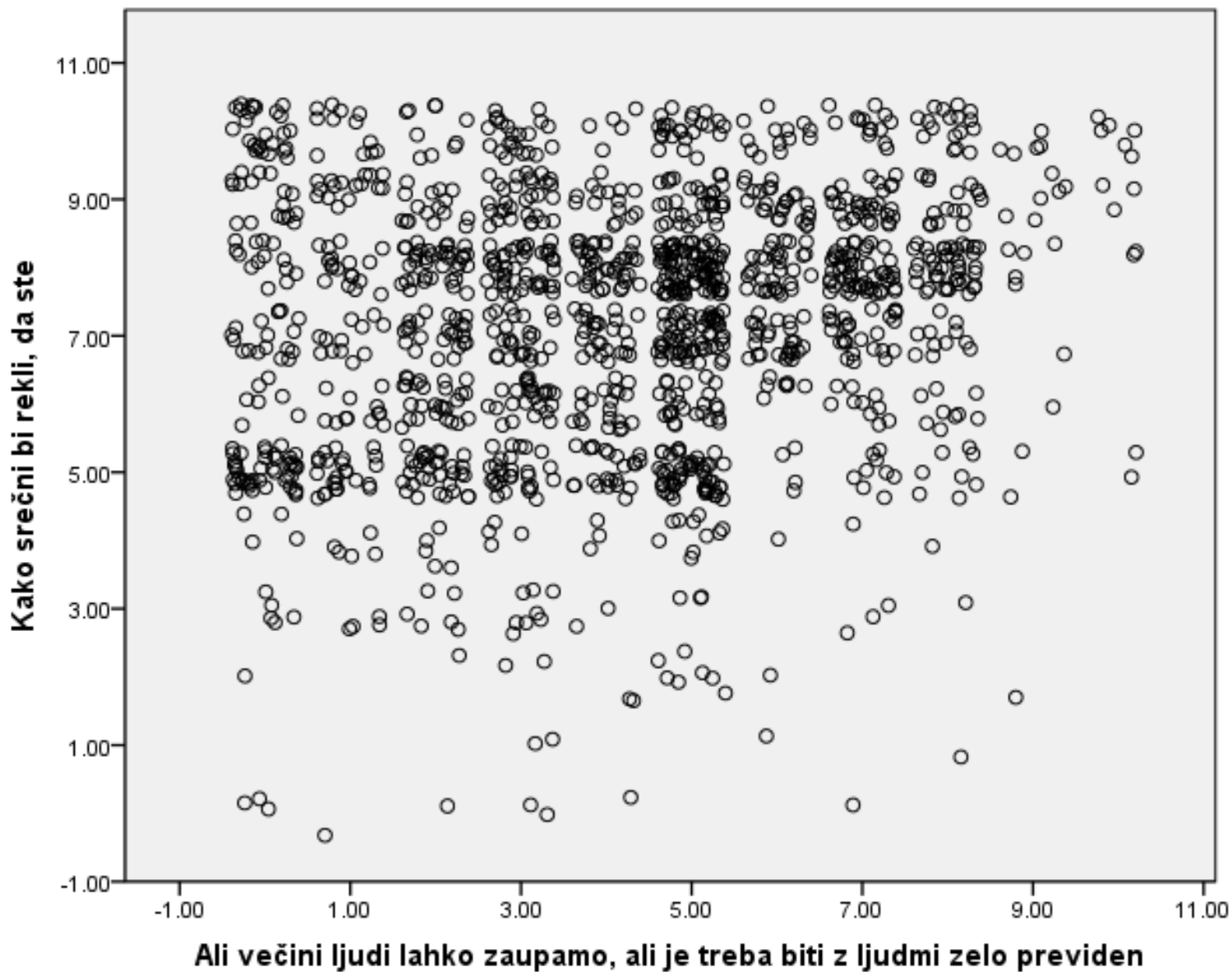


Zatresene spremenljivke

- Razsevni grafikon naredimo kot običajno, le da v njem uporabimo „zatresene“ spremenljivke
- Spremenljivke zatresemo tako, da vsaki enoti dodamo slučajno napako, recimo iz enakomerne porazdelitve, na intervalu, ki je centriran na 0 in malce ožji od razlike med dvema sosednjima vrednostnima.
- V našem primeru je razlike 1, tako da izberemo intreval $(-0.4, 0.4)$



Zatresene spremenljivke



Regresija in pogojno matematično upanje

Vzemimo spet klasični regresijski model: $Y' = \alpha + \beta X$, pri čemer z Y' označujemo regresijsko napoved spremenljivke Y . Zgornji model ima naslednjo statistično interpretacijo:

1. Vrednosti spremenljivke Y' so **pričakovane (povprečne) vrednosti** spremenljivke Y , za dane vrednosti spremenljivke X .
2. Pričakovano vrednost spremenljivke predstavlja njeno **matematično upanje**. Vrednosti spremenljivke Y' pri $X = X_i$ predstavljajo matematično upanje spremenljivke Y ob pogoju $X = X_i$. Tako matematično upanje imenujemo **pogojno matematično upanje** in označimo:

$$E(Y | X = X_i) = \alpha + \beta X_i.$$

Namesto $E(Y | X = X_i)$ bomo večkrat pisali kar $E(Y | X_i)$.

Člen napake v regresijskem modelu

Dejanske vrednosti spremenljivke Y se od regresijske napovedi bolj ali manj razlikujejo.

Označimo torej:

$$\varepsilon_i = Y_i - Y'(X_i).$$

Člen ε_i imenujemo člen napake ali stohastični člen napake.

Regresijsko enačbo sedaj lahko zapišemo kot

$$Y = \alpha + \beta X + \varepsilon.$$

Člen napake je slučajna spremenljivka, ki nadomešča vse nezajete vplive.

Člen napake v regresijskem modelu

Vzroki, zakaj bivaritnega modela ne nadomestimo z multiplim regresijskim modelom, ki bi vključeval tudi preostale spremenljivke, ki vplivajo na odvisno spremenljivko:

- Nekaterih vzrokov morda sploh sploh ne poznamo oziroma jih ne znamo ali ne moremo meriti.
- Na pojav lahko vpliva veliko število vzrokov, katerih posamezni vplivi so zanemarljivi.
- Notranja slučajnost v človeškem obnašanju: določenih oblik človeškega obnašanja ni mogoče napovedovati na podlagi značilnosti, ki jih je mogoče opazovati.
- Vzrok za večanje člena napake je lahko tudi nepravilna izbira funkcijske zveze med odvisno in neodvisno spremenljivko.
- Pogosto so vir napak tudi **merske napake** (še posebej, kadar imamo opravka s podatki iz anket).