

1. Naloga: Analiza odvisnosti stopnje brezposelnosti od deleža prebivalcev z visokošolsko izobrazbo na podlagi vzorca slovenskih občin.

V vzorec smo zajeli 33 občin (slučajni vzorec). Spremenljivka VIS predstavlja število prebivalcev z visokošolsko strokovno izobrazbo, VIS_DOD število prebivalcev z univerzitetno izobrazbo in POD število prebivalcev s podiplomsko izobrazbo, spremenljivka BREZP predstavlja število registrirano brezposelnih oseb, DEL_AKT pa število delovno aktivnih prebivalcev. Da se v analizi izloči vpliv različne velikosti občin je smiselno vse analize delati na podlagi relativnih kazalnikov. Brezposelnost v posamezni občini izrazimo relativno z izračunom stopnje brezposelnosti, izobraženost prebivalstva pa z % prebivalcev s visokošolsko izobrazbo.

1) Oblikujemo relativne kazalnike (stopnje)

Postopek: Transform -> Compute...->v "Target Variable" vnesemo ST_BREZP->v polje "Numeric Expression" vpišemo $BREZP / (DEL_AKT + BREZP) * 100$ -> Potrdimo z OK.

```
COMPUTE ST_BREZP = BREZP / (DEL_AKT+BREZP)*100 .  
EXECUTE .
```

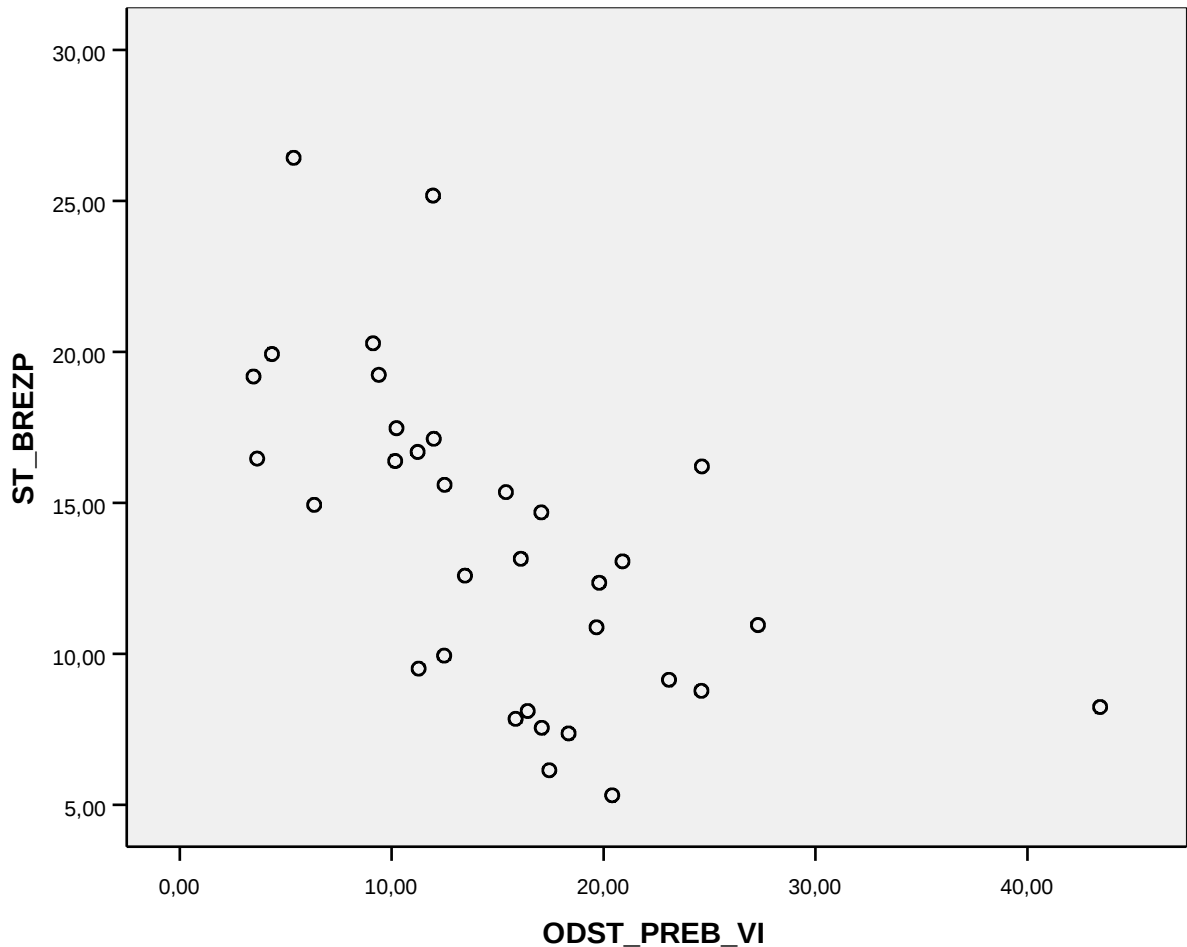
Transform -> Compute...->v "Target Variable" vnesemo ODST_PREB_VI->v polje "Numeric Expression" vpišemo $(VIS + VIS_DOD + POD) / (DEL_AKT + BREZP) * 100$ -> Potrdimo z OK.

```
COMPUTE ST_BREZP = (VIS+VIS_DOD+POD) / (DEL_AKT+BREZP)*100 .  
EXECUTE .
```

2) Podatke prikažemo v razsevnem diagramu s pomočjo programa SPSS

Postopek: Graphs -> Scatter/Dot...->Simple Scatter->Define-> na Y-axis določimo ST_BREZP in na X-Axis ODST_PREB_VI. Potrdimo z OK.

```
GRAPH  
/SCATTERPLOT(BIVAR)=ODST_PREB_VI WITH ST_BREZP  
/MISSING=LISTWISE .
```



3) Analiziramo razsevni diagram

- Na podlagi razsevnega diagrama lahko razberemo, da je odvisnost stopnje brezposelnosti od odstotka prebivalcev z visokošolsko izobrazbo negativna. V občinah z večjim odstotkom prebivalcev z visokošolsko izobrazbo je stopnja brezposelnosti v povprečju manjša.
- Odvisnost stopnje brezposelnosti od odstotka prebivalcev z visokošolsko izobrazbo ni povsem linearna (mi bomo pa vseeno ocenili linearni regresijski model).
- Odvisnost stopnje brezposelnosti od odstotka prebivalcev z visokošolsko izobrazbo je šibka (točke v razsevni diagramu so precej oddaljene od namišljene regresijske premice).

4) Ocenimo linearni regresijski model:

Splošen bivariatni linearni regresijski model: $y_i = \alpha + \beta x_i + \varepsilon_i$

Enačba regresijske premice: $y' = \alpha + \beta x$

V našem primeru (vzorčni podatki): $ST_BREZP = a + b * ODST_PREB_VI$

Postopek: Analyze -> Regression->Linear-> V polje Dependent vnesemo ST_BREZP in v polje Independent(s) pa ODST_PREB_VI. Potrdimo z OK.

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT ST_BREZP
/METHOD=ENTER ODST_PREB_VI .
```

Izpis:

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	ODST_ ^a PREB_VI	.	Enter

a. All requested variables entered.

b. Dependent Variable: ST_BREZP

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,595 ^a	,355	,334	4,32383

a. Predictors: (Constant), ODST_PREB_VI

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	318,407	1	318,407	17,031	,000 ^a
	Residual	579,561	31	18,696		
	Total	897,969	32			

a. Predictors: (Constant), ODST_PREB_VI

b. Dependent Variable: ST_BREZP

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	19,707	1,639		12,021	,000
	ODST_PREB_VI	-,393	,095	-,595	-4,127	,000

a. Dependent Variable: ST_BREZP

5) Zapišemo enačbo regresijske premice in komentiramo regresijske koeficiente

$$a = 19,707$$

$$b = -0,393$$

$$ST_BREZP = 19,707 - 0,393 * ODST_PREB_VI$$

Regresijske konstante a ni smiselno razlagati, saj ni realno predpostavljati občine brez prebivalcev z visokošolsko izobrazbo.

Razlaga regresijskega koeficienta b:

Na podlagi vzorca občin ocenjujemo, da se stopnja brezposelnosti v povprečju zmanjša za 0,39 odstotne točke, če se odstotek prebivalcev z visokošolsko izobrazbo poveča za 1 odstotno točko.

6) Izvedemo preskus neodvisnosti

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Na podlagi t-statistike ($t = -4,127$) oziroma natančne stopnje tveganja (Sig. = 0,000) lahko zavrnemo ničelno domnevo (poišči v tablicah kolikšna je kritična vrednost t –statistike za konkreten primer).

Na podlagi vzorčnih podatkov lahko pri zanemarljivi stopnji tveganja zavrnemo ničelno domnevo in sprejmemo sklep, da je stopnja brezposelnosti odvisna od odstotka prebivalcev z visokošolsko izobrazbo.

7) Komentiramo determinacijski in korelacijski koeficient

$$r_{yx}^2 = 0,355$$

Na podlagi ocenjenega determinacijskega koeficienta ugotavljamo, da je 35% variance stopnje brezposelnosti pojasnjene z linearnim vplivom odstotka prebivalcev z visokošolsko izobrazbo. Preostalih 65% variance povzročajo neznani, med njimi vsaj slučajni dejavniki.

$$r_{yx} = -0,595(\text{predznak!})$$

Na podlagi ocene korelacijskega koeficienta ocenjujemo, da je povezanost med stopnjo brezposelnosti in odstotkom prebivalcev z visokošolsko izobrazbo negativna in šibka.

2. Naloga: Analiza odvisnosti cene stanovanj od njihovih značilnosti (multipla linearna odvisnost)

Analizirali smo odvisnost cene stanovanj od njihove površine, starosti ter oddaljenosti od centra mesta. V vzorec smo zajeli 30 stanovanj.

1) Ocenimo multivariatni linearni regresijski model:

Splošen multivariatni linearni regresijski model: $y_i = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon_i$

Enačba regresijske hiperravnine: $y' = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$

V našem primeru (vzorčni podatki): $cena = a + b_1 * povm2 + b_2 * star + b_3 * odd$

Postopek: Analyze -> Regression->Linear-> V polje Dependent vnesemo Cena in v polje Independent(s) povm2, star in odd. Kliknemo 'Statistics' in odkljukamo 'Descriptives' (za izračun korelacijske matrike). Potrdimo s Continue in z OK.

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT CENA
  /METHOD=ENTER POV2 STAR ODD .
```

a) Korelacijska matrika

		Correlations			
		CENA	POVM2	STAR	ODD
Pearson Correlation	CENA	1,000	,925	,196	-,286
	POVM2	,925	1,000	,308	-,077
	STAR	,196	,308	1,000	-,006
	ODD	-,286	-,077	-,006	1,000
Sig. (1-tailed)	CENA	.	,000	,150	,063
	POVM2	,000	.	,049	,343
	STAR	,150	,049	.	,487
	ODD	,063	,343	,487	.
N	CENA	30	30	30	30
	POVM2	30	30	30	30
	STAR	30	30	30	30
	ODD	30	30	30	30

Korelacijski koeficient, ki kaže odvisnost med spremenljivkama površina stanovanja in oddaljenost stanovanja od centra znaša -0,077, kar pomeni, da je odvisnost med spremenljivkama negativna in zelo šibka.

Izvedemo preizkus neodvisnosti:

$$H_0 : \rho_{povm2,odd} = 0$$

$$H_1 : \rho_{povm2,odd} \neq 0$$

Natančno stopnjo tveganja pri kateri lahko zavrnemo navedeno ničelno domnevo lahko razberemo iz korelacijske matrike. Vendar je potrebno upoštevati, da je natančna stopnja tveganja v korelacijski matriki podana za enostranski preskus, mi pa želimo izvesti dvostranski preskus.

$$\text{Sig}(2\text{-tailed}) = 2 * \text{Sig}(1\text{-tailed})$$

Ničelno domnevo bi lahko zavrnili le pri stopnji tveganja 0,686 kar je povsem nesprejemljivo. **Ugotovitev:** Na podlagi vzorca stanovanj ni mogoče pri dovolj nizki stopnji tveganja zavrniti ničelne domneve, zato ne moremo trditi, da sta spremenljivki povezani.

b) Ocena regresijskega modela

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,954 ^a	,909	,899	1,75830

a. Predictors: (Constant), ODD, STAR, POV2

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	805,587	3	268,529	86,857	,000 ^a
	Residual	80,382	26	3,092		
	Total	885,969	29			

a. Predictors: (Constant), ODD, STAR, POV2

b. Dependent Variable: CENA

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3,207	1,078		2,976	,006
	POV2	,205	,014	,937	15,044	,000
	STAR	-,049	,033	-,094	-1,512	,143
	ODD	-,164	,045	-,214	-3,616	,001

a. Dependent Variable: CENA

Interpretacija rezultatov:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_1 : \text{vsaj en } \beta_i \neq 0$$

$$F = 86,857 > F_{\alpha=0,005, m_1=3, m_2=26} = 5,41$$

Na podlagi vzorca stanovanj lahko pri zanemarljivi stopnji tveganja zavrnilo ničelno domnevo in sprejmemo sklep, da je vsaj eden izmed regresijskih koeficientov različen od 0 (vsaj ena izmed vključenih spremenljivk dejansko vpliva na ceno stanovanja).

$$\bar{r}^2 = 0,899$$

Na podlagi vzorčnih podatkov ocenjujemo, da je (popravljeni) multipli determinacijski koeficient enak **0,899**, kar pomeni, da je 89,9% variance cene stanovanj pojasnjeno z linearnim vplivom površine stanovanj, starosti stanovanj in oddaljenosti od centra mesta.

Ocena regresijske funkcije:

$$\text{cena} = 3,207 + 0,205 * \text{povm2} - 0,049 * \text{star} - 0,164 * \text{odd}$$

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

$t = -1,512$ in natančna stopnja tveganja za zavrnitev ničelne domneve: 0,143. Ker je natančna stopnja tveganja višja od najvišje sprejemljive stopnje tveganja $\alpha = 0,05$, ničelne domneve ne moremo zavrniti.

Na podlagi vzorčnih podatkov ne moremo zavrniti ničelne domneve, da je regresijski koeficient β_2 enak nič. Ne moremo torej trditi, da starost stanovanj vpliva na njihovo ceno.

Glede na ugotovitev, ponovimo oceno modela, tokrat v obliki:

$$\text{cena} = a + b_1 * \text{povm2} + b_2 * \text{odd}$$

Postopek: Analyze -> Regression->Linear-> V polje Dependent vnesemo Cena in v polje Independent(s) povm2 in odd. Potrdimo z OK.

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT CENA
/METHOD=ENTER POV2 ODD .
```

Izpis:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,949 ^a	,901	,894	1,79966

a. Predictors: (Constant), ODD, POVM2

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	798,522	2	399,261	123,276	,000 ^a
	Residual	87,447	27	3,239		
	Total	885,969	29			

a. Predictors: (Constant), ODD, POVM2

b. Dependent Variable: CENA

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2,796	1,067		2,620	,014
	POVM2	,199	,013	,908	14,973	,000
	ODD	-,165	,046	-,216	-3,561	,001

a. Dependent Variable: CENA

Interpretacija rezultatov:

Vrednost multiplega determinacijskega koeficienta se je zmanjšala le za 0,005 (% pojasnjene variance se je zmanjšal za pol odstotne točke).

Na podlagi vrednosti F statistike lahko še pri nižji stopnji tveganja zavrnemo ničelno domnevo in sprejmemo sklep, da je vsaj en regresijski koeficient različen od 0 (vsaj ena izmed vključenih spremenljivk vpliva na ceno).

$$cena = 2,796 + 0,199 * povm2 - 0,165 * odd$$

Na podlagi vzorčnih podatkov ocenjujemo, da se cena stanovanja v povprečju poveča za 199.000 SIT, če se površina stanovanja poveča za kvadratni meter, oddaljenost od centra mesta pa ostane nespremenjena.

Na podlagi vzorčnih podatkov ocenjujemo, da se cena stanovanja v povprečju zmanjša za 165.000 SIT, če se oddaljenost od centra mesta poveča za en kilometer, površina stanovanja pa ostane nespremenjena.

Obe spremenljivki statistično značilno vplivata na ceno stanovanj.

3. Naloga: analiza nelinearne odvisnosti

Analiza odvisnosti deleža izdatkov za hrano v vseh izdatkih od višine razpoložljivih sredstev (dohodka).

1) Definiramo novo spremenljivko DEL, ki predstavlja delež izdatkov za hrano

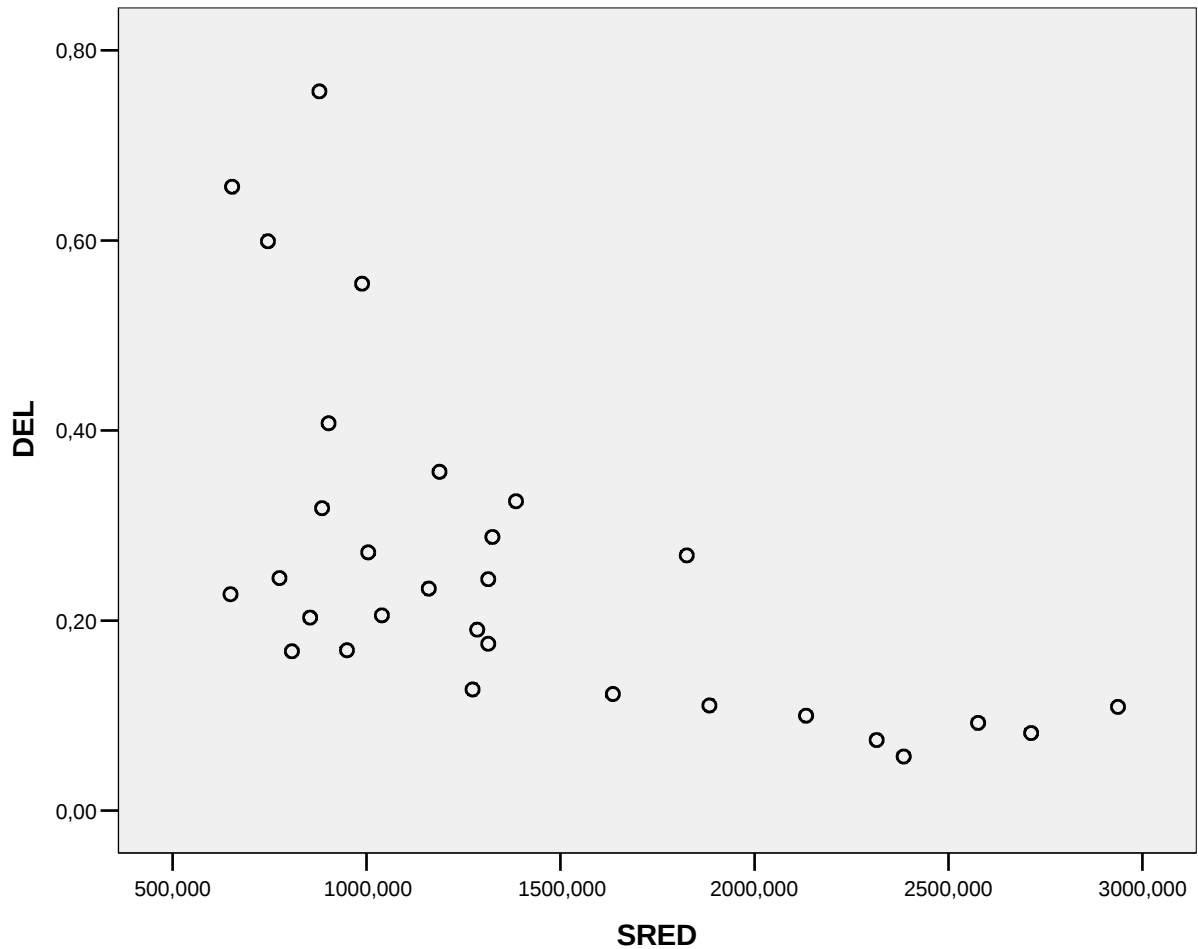
Postopek: Transform -> Compute->v polje Target Variable vpišemo DEL, v polje Numeric Expression pa vpišemo HRANA/SRED. Potrdimo z OK.

```
COMPUTE DEL = HRANA/SRED .  
EXECUTE .
```

2) Oblikujemo razsevni diagram:

Postopek: Graphs -> Scatter/Dot...->Simple Scatter->Define-> na Y-axis določimo DEL in na X-Axis SRED. Potrdimo z OK.

```
GRAPH  
  /SCATTERPLOT(BIVAR)=SRED WITH DEL  
  /MISSING=LISTWISE .
```



Na podlagi analize razsevnega diagrama lahko ugotovimo, da odvisnost ni linearna. V tem primeru si pomagamo z nelinearnim modelom, na primer eksponentno ali potenčno funkcijo.

Potenčna funkcija: $y' = \alpha x^\beta$

S pomočjo logaritmov jo transformiramo v funkcijo, ki je linearna po parametrih:

$$y' = \alpha x^\beta$$

$$\log y' = \log \alpha x^\beta$$

$$\log y' = \log \alpha + \log x^\beta$$

$$\log y' = \log \alpha + \beta \log x$$

Namesto osnovnih vrednosti spremenljivk moramo model oceniti na podlagi transformiranih vrednosti.

3) Postopek transformacije vrednosti:

Postopek: Transform -> Compute->v polje Target Variable vpišemo LDEL, v polje Numeric Expression pa vpišemo LG10(DEL). Potrdimo z OK.

Ponovimo še: Transform -> Compute->v polje Target Variable vpišemo LSRED, v polje Numeric Expression pa vpišemo LG10(SRED). Potrdimo z OK.

```
COMPUTE LDEL = LG10(DEL) .  
COMPUTE LSRED = LG10(SRED) .  
EXECUTE .
```

4) Ocenimo transformirani model, ki je linearen v parametrih:

$$LDEL = \log a + b \cdot LSRED$$

Postopek: Analyze -> Regression->Linear-> V polje Dependent vnesemo LDEL in v polje Independent(s) pa LSRED. Potrdimo z OK.

```
REGRESSION  
/MISSING LISTWISE  
/STATISTICS COEFF OUTS R ANOVA  
/CRITERIA=PIN(.05) POUT(.10)  
/NOORIGIN  
/DEPENDENT LDEL  
/METHOD=ENTER LSRED .
```

Izpis:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,756 ^a	,572	,557	,19202

a. Predictors: (Constant), LSRED

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,381	1	1,381	37,454	,000 ^a
	Residual	1,032	28	,037		
	Total	2,413	29			

a. Predictors: (Constant), LSRED

b. Dependent Variable: LDEL

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	2,873	,582		4,939	,000
LSRED	-1,145	,187	-,756	-6,120	,000

a. Dependent Variable: LDEL

Regressijske enačba: $DEL = 746,8 * SRED^{-1,145}$

Interpretacija rezultatov:

Na podlagi vzorca gospodinjstev ocenjujemo, da se delež izdatkov za hrano v povprečju zmanjša za 1,14% , če se vrednost razpoložljivih sredstev poveča za 1%.

Na podlagi izračunane t-statistike (-6,120) oziroma podane natančne stopnje tveganja (Sig.= 0,000) lahko pri zanemarljivem tveganju zavrnilo ničelno domnevo in sprejmemo sklep, da je regresijski koeficient različen od 0. Torej lahko trdimo, da je delež izdatkov za hrano odvisen od višine sredstev.

Na podlagi ocene determinacijskega koeficienta, ki znaša 0,572 ugotavljamo, da lahko 57% variance logaritmov deleža sredstev za hrano pojasnimo z linearno odvisnostjo od logaritmov vrednosti razpoložljivih sredstev.

Navedeni determinacijski koeficient ni primerljiv s tistim, kjer je odvisna spremenljivka izražena v običajnih merskih enotah.

5) Izračunamo primerljiv determinacijski koeficient:

$$s_y^2 = s_R^2 + s_e^2$$

$$\frac{\sum_{i=1}^n (y - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^n (y'' - \bar{y})^2}{k} + \frac{\sum_{i=1}^n (y - y'')^2}{n-k-1}$$

oziroma

$$k_y = k_R + k_e$$

$$r^2 = \frac{k_R}{k_y} = 1 - \frac{k_e}{k_y}$$

Oceniti moramo torej: $k_e = \sum_{i=1}^n (y - y'')^2$

Postopek sestoji iz naslednjih korakov:

- 1) Ocenimo vrednosti logaritmov odvisne spremenljivke v transformiranem modelu (vrednosti, ki ležijo na premici)

Analyze -> Regression->Linear-> V polje Dependent vnesemo LDEL in v polje Independent(s) pa LSRED. Kliknemo na Save... in izberemo Predicted Values – Unstandardized. Potrdimo s Continue in z OK.

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT LDEL
/METHOD=ENTER LSRED
/SAVE PRED .
```

- 2) Antilogaritmiramo ocenjene vrednosti logaritmov odvisne spremenljivke iz prvega koraka in dobimo ocenjene vrednosti odvisne spremenljivke v osnovnih merskih enotah (vrednosti, ki ležijo na krivulji)

Transform -> Compute->v polje Target Variable vpišemo OCDEL, v polje Numeric Expression pa vpišemo $10^{**}PRE_1$. Potrdimo z OK.

```
COMPUTE OCDEL = 10**PRE_1 .
EXECUTE .
```

- 3) Izračunam odklone dejanskih vrednosti od ocenjenih vrednosti (odkloni, ki jih povzročajo neznani dejavniki, med njimi vsaj slučajni).

Transform -> Compute->v polje Target Variable vpišemo ODKLON, v polje Numeric Expression pa vpišemo $DEL-OCDEL$. Potrdimo z OK.

```
COMPUTE ODKLON = DEL-OCDEL .
EXECUTE .
```

- 4) Kvadriramo odkone

Transform -> Compute->v polje Target Variable vpišemo KVODKLON, v polje Numeric Expression pa vpišemo $ODKLON^{**}2$. Potrdimo z OK.

```
COMPUTE KVODKLON = ODKLON**2 .
EXECUTE .
```

- 5) Seštejemo kvadrate odklonov, da dobimo k_e .

Analyze -> Descriptive statistics...-> Descriptives-> v polje Variables prenesemo KVODKLON. Kliknemo na Options in izberemo Sum. Potrdimo s Continue in z OK.

```
DESCRIPTIVES
VARIABLES=KVODKLON
/STATISTICS= SUM .
```

Izpis:

Descriptive Statistics

	N	Sum
KVODKLON	30	,58
Valid N (listwise)	30	

6) Ocenimo s_y^2 in izračunamo k_y .

$$k_y = s_y^2(n-1)$$

Analyze -> Descriptive statistics...-> Descriptives-> v polje Variables prenesemo DEL.
Kliknemo na Options in izberemo Variance. Potrdimo s Continue in z OK.

```
DESCRIPTIVES  
VARIABLES=DEL  
/STATISTICS= VAR .
```

Izpis:

Descriptive Statistics

	N	Variance
DEL	30	,032
Valid N (listwise)	30	

7) Izračunamo primerljiv determinacijski koeficient.

$$r_{yx}^2 = 1 - \frac{k_e}{k_y} = 1 - \frac{0,58}{0,928} = 0,375$$

8) Interpretiramo rezultat:

Na podlagi vzorca gospodinjstev ocenjujemo, da lahko 37,5% variance deleža izdatkov za hrano pojasnimo s potenčno odvisnostjo od višine razpoložljivih sredstev.