

UVOD

→ ekonomsko mjerenje
→ znanstvena disciplina koja se bavi empiričkim¹ dokazivanjem ekonomskih zakona koji su definirani u ekonomskoj teoriji
→ usko je vezana za discipline kao što su ekonomska teorija, matematička ekonomija i ekonomska statistika

Ekonomska teorija → povezanost između ekonomskih pojava; ne govore o jačini (veličini) već o smjeru

↻ Provjera da li u praksi postoje veze koju opisuje ekonomska teorija i da ih kvantificira
↻ Ekonometrija → potvrđuje odnosno ne potvrđuje ekonomsku teoriju

Matematička ekonomija → izražava ekonomsku teoriju u određenoj matematičkoj formi bez osvrtnja na empiričku osnovu u toj promjeni

Ekonometrija polazeći od jednadžbi matematičke ekonomije gradi ekonometrijske modele pogodne za empiričko testiranje ekonomskih teorija.

Ekonomska statistika → bavi se prikupljanjem, obradom i prezentiranjem ekonomskih pokazatelja u obliku tablica i slika. Ne bavi se provjerom ekonomskih teorija na temelju tako prikupljenih podataka, već to čini ekonometrija, koristeći se podacima prikupljenim od strane statistike koji ulaze u njezinu analizu.

Metodologija ekonometrije:

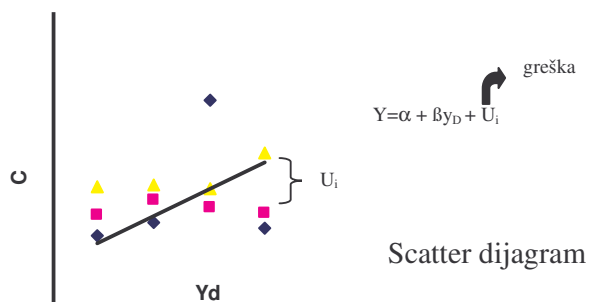
Ekonometrijska analiza slijedi određeni metodološki put:

1. Određivanje teorije ili hipoteze → preuzima iz ekonomske teorije
2. Specifikacija ekonometrijskog modela → iako je Keynes pretpostavio pozitivnu vezu između C i Y_D , nije specificirao precizni oblik funkcionalne veze između tih varijabli

$$C = \alpha + \beta Y_D \quad 0 < \beta < 1$$

C → zavisna varijabla (endogena varijabla)

Y_D → nezavisna varijabla (egzogena varijabla-zadana)



Stohastički ili empirički model

¹ iz iskustva

Stohastičke² varijable imaju određenu distribuciju vjerojatnosti.

3. Procjena parametara → dati vrijednosti temeljem neke metode parametrima (koeficijentima)

α ili $\beta_0=120$ β ili $\beta_1=0.8$
↓ ↓
odsječak na ordinati nagib pravca

Svojstvo pravca: minimizirati odstupanje

4. Provjera

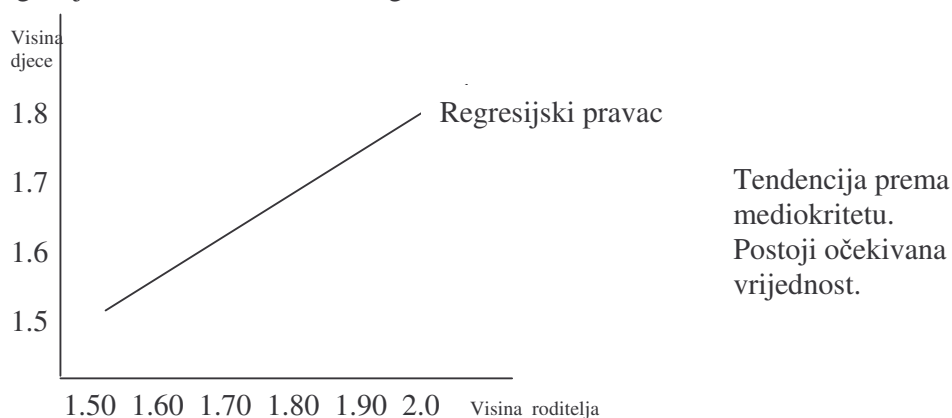
$H_0: \beta_1=0.9$

Ako je velika disperzija, prihvaćamo hipotezu.

5. Predviđanje (strukture modela)

PRIRODA REGRESIJSKE ANALIZE

„Regresija“ - Francis Galton 1886. godine



Regresija i uzročnost

Koeficijent korelacije³ → samo linearna veza/ mjera linearne korelacije

$$-1 < r < 1$$

žetva ↔ vrijeme

Ne govori se koja varijabla na koju utječe.

² Deterministička povezanost → za svaku vrijednost jedne pojave točno znamo vrijednost druge pojave
Stohastička povezanost → na osnovu vrijednosti jedne pojave ne možemo sa sigurnošću predvidjeti vrijednost druge pojave

³ Čemu služe korelacija i regresija? → Statistička sredstva za proučavanje povezanosti (odnosa) među pojavama. *Korelacija* proučava jakost (intenzitet, stupanj) povezanosti među pojavama. *Regresija* → povezanost se precizno opisuje pomoću modela (tzv. regresijski model). Koeficijent korelacije je mjera linearne korelacije.

žetva= f (meteorološke prilike, gnojivo,...)

Regresija pokazuje smjer uzročnosti, za razliku od korelacije koja je simetrična (obje se varijable tretira jednako).

Regresijska analiza proučava zavisnost varijable o nezavisnim varijablama. To je kretanje prema prosjeku. Služi nam za procjenu srednje vrijednosti zavisne varijable.

Stohastičke varijable mogu poprimiti bilo koju vrijednost uz zadanu vjerojatnost.

Terminologija:

$$Y = \beta_0 + \beta_1 X$$

Jednostavna regresija

$Y \rightarrow$ zavisna varijabla

$X \rightarrow$ nezavisna varijabla

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Višestruka regresija

$U \rightarrow$ Slučajna greška, rezidual

Vrste podataka:

1) Vremenski nizovi

\rightarrow sezonska kretanja

\rightarrow trendovi

\rightarrow slijedeća ovisi o prethodnoj

t	Y_D	C
1	500	350
2	350	100
3	600	550
4	1500	

2) Cross section

\rightarrow različite vremenske točke

\rightarrow nije vezan za vrijeme (nema vremensku komponentu)

IME	Y_D	C
Paola	350	100
Ivana	500	450

3) Panel podaci

\rightarrow vremenski nizovi + cross section

	Paola		Ivana	
	Y_D	C	Y_D	C
1	500	350		
2	350	100		
3	600	550		

Kvaliteta:

Ekonometrijska analiza vrijedi koliko i kvaliteta podataka koje koristimo. Podaci su loši jer:

- 1) Postoji opservacijska greška
- 2) Nema dovoljno podataka
- 3) Postoje greške koje se pojavljuju zaokruživanjem
- 4) Pristranost kod selekcije
- 5) Problem kad su podaci jako agregirani

Regresija s dvije varijable

$Y=f(X) \rightarrow$ jednostavna regresija

$Y \rightarrow$ potrošnja

$X \rightarrow$ raspoloživi dohodak

Dohodak		
80	100	120
55	65	79
55*	70	84
65	74	90*
70	80	94
75	85*	98
	88	

* uzeli smo kao uzorak

Uvjetna vjerojatnost \rightarrow zato što Y ovisi o $X \rightarrow Y$ će poprimiti neku vrijednost Y_i ako X poprimi X_i

$$P(Y=Y_i|X=X_i)$$

$$P(Y|X)$$

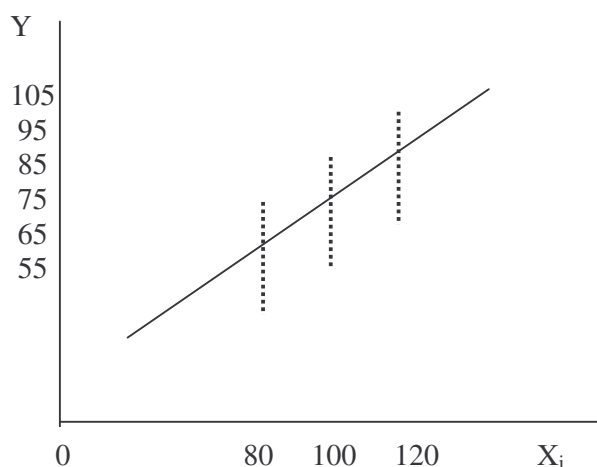
	Dohodak		
	80	100	120
P (Y X) *	2/5	1/6	1/5
		1/6	1/5
	1/5	1/6	1/5
	1/5	1/6	1/5
	1/5	1/6	1/5
E (Y X _i)	64	77	89

* uvjetna vjerojatnost Y uz dano X

Uvjetna sredina ili matematičko uvjetno očekivanje

Općeniti oblik linearne funkcije → E (Y|X_i)

$$E = 55 \times 2/5 + 65 \times 1/5 + 70 \times 1/5 + 75 \times 1/5 = 22 + 13 + 14 + 15 = 64$$



Regresijska funkcija
populacije

→ matematičko očekivanje ili
uvjetna sredina zavisne varijable
za fiksnu vrijednost nezavisne
varijable

$$E(Y|X_i) = f(X_i)$$

Očekivana vrijednost će ovisiti o visini dohotka.

$$E(Y|X_i) = \beta_0 + \beta_1 X_i \quad \text{Linearna populacijska regresija}$$

LINEARNA REGRESIJA → linearnost u parametrima, a ne linearnost u varijablama

$$\text{a) } E(Y|X_i) = \beta_0 + \beta_1 X_i \quad +$$

$$\text{b) } E(Y|X_i) = \sqrt{\beta_0} + \beta_1 X_i^2 \quad -$$

$$\text{c) } E(Y|X_i) = \beta_0 + \beta_1 \ln X_i \quad +$$

$$\text{d) } E(Y|X_i) = \beta_0 + \beta_1^2 X_i \quad -$$

Linearnost → odnos

→ PARAMETARA → odnos parametara i zavisne varijable $E(Y|X) = \beta_0 + \sqrt{\beta_1} X_1$
(nema linearnosti)

→VARIJABLI→ odnos varijable i zavisne varijable $E(Y|X)=\beta_0+\beta_1X_1^2$ (nema linearnosti)

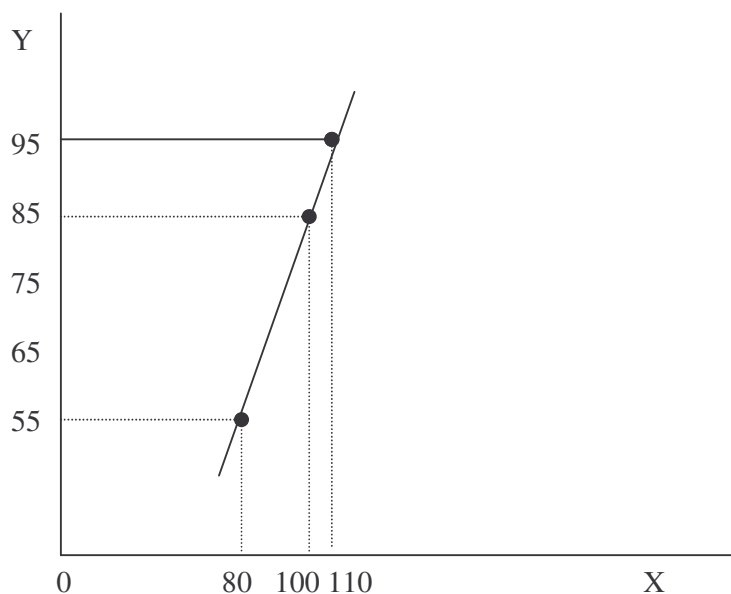
Odstupanje

$$U_i = Y_i - E(Y|X_i)$$

$$Y_i = E(Y|X_i) + U_i$$

Pojedinačna potrošnja će ovisiti o prosjeku grupacije \pm neka greška

Regresijska funkcija uzorka→ uzorak mora biti reprezentativan

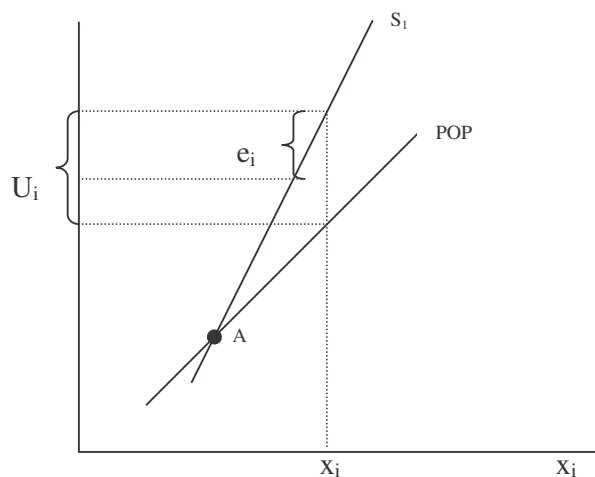


$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$$

$\hat{Y} \rightarrow$ procijenjena vrijednost Y-a

$\hat{\beta}_0 \rightarrow$ procijenjena vrijednost β

$\hat{\beta}_1 \rightarrow$ procijenjena vrijednost β_1



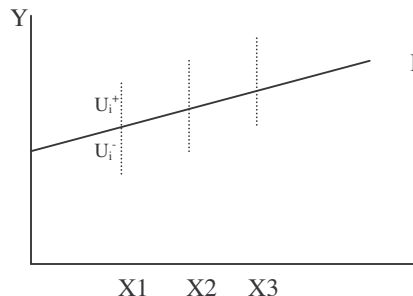
Regresijska funkcija uzorka podcjenjuje vrijednosti populacije→prije A

Regresijska funkcija uzorka precjenjuje vrijednosti populacije → nakon A

Pretpostavke koje moraju biti zadovoljene da bi prihvatili regresijsku funkciju uzorka kao procjenitelj regresijske funkcije populacije:

1. $E(U_i|X_i)=0$

Statističke greške regresijske funkcije populacije su 0.



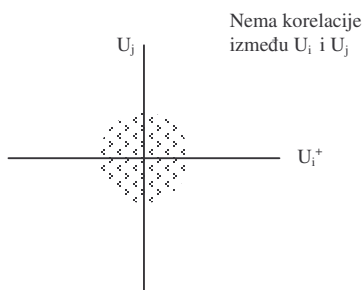
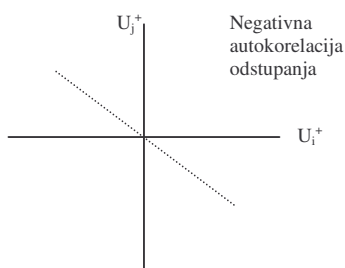
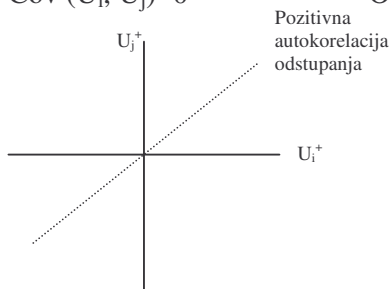
Regresijska funkcija populacije

→za svaku vrijednost varijable X,
odstupanje je normalno distribuirano oko 0
→sredina=0
→simetrično distribuirana oko njezine
sredine

Za svaki njegov X_i , njegov prosjek mora biti 0.

2. $Cov(U_i, U_j)=0$

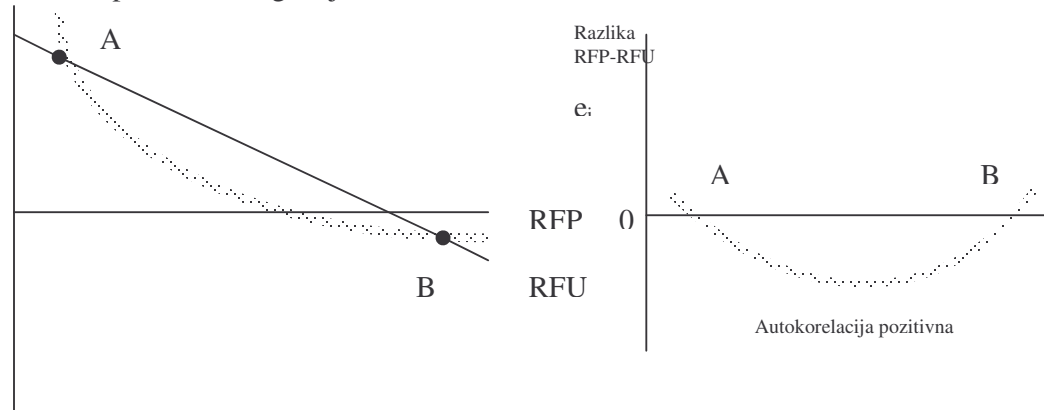
Odsutnost autokorelacije, odstupanja su nekorelirana.



3. Homoskedastičnost $Var(U_i|X_i)=\sigma^2$ → varijanca je jednaka

Svako odstupanje ima istu varijancu, nemogućnost raspršenja odstupanja za veće vrijednosti od malih.

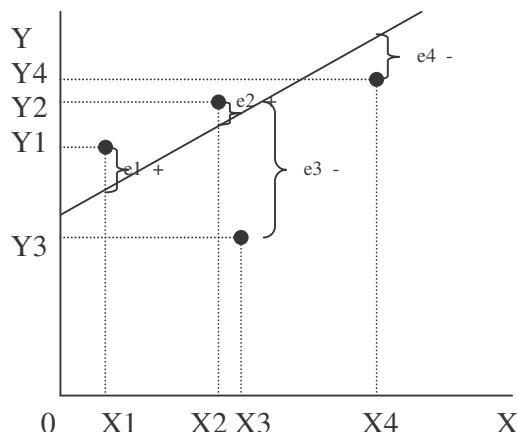
4. Pravilno specificiran regresijski model



Regresija s dvije varijable -Problem procjene parametara-

→ metoda najmanjih kvadrata (OLS)⁴

Karl Friedrich Gauss



$$\text{Min} \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$$

$$e_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

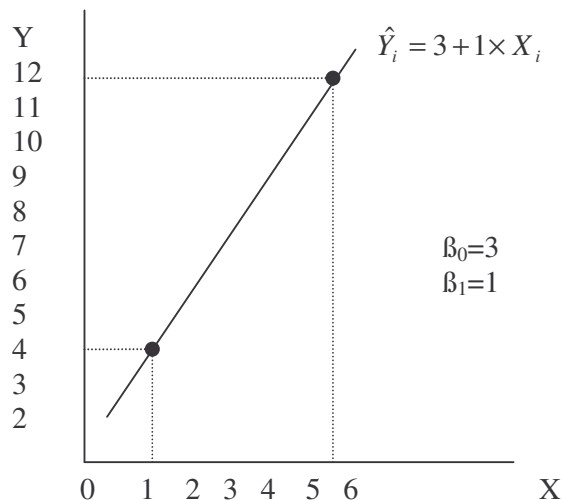
$$e_i = Y_i - \hat{Y}_i \quad Y_i \rightarrow \text{stvarni} \quad \hat{Y}_i \rightarrow \text{procijenjeni}$$

⁴ Ordinary Least Square → Dobivamo najmanja odstupanja od regresijske linije

Izabrati $\hat{\beta}_0$ i $\hat{\beta}_1$ da minimiziraju onu sumu

$$\sum e_i^2 = f(\hat{\beta}_0, \hat{\beta}_1)$$

Y_i	x_i	\hat{Y}_i	e_i	e_i^2
4	1	4	0	0
5	4	7	-2	4
7	5	8	-1	1
12	6	9	3	9
			0	14



npr. $\hat{Y} = 4 + 0.5X$

$$\sum e_i^2 = 26.5$$

Bolji je pravac jer minimizira sumu tih kvadrata.

Treba izabrati β_0 i β_1 takve da minimiziraju sumu kvadrata tih odstupanja.

$$\frac{\partial}{\partial \beta_0} \sum (Y_i - \underbrace{\hat{\beta}_0 - \hat{\beta}_1 X_i}_{e_i^2})^2 = 0$$

$$\frac{\partial}{\partial \beta_1} = \sum \dots = 0$$

$$\hat{\beta}_1 = \frac{\sum x_i \cdot y_i}{\sum x_i^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$x_i = (X_i - \bar{X})$$

$$y_i = (Y_i - \bar{Y})$$

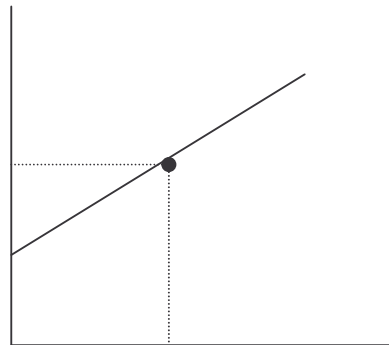
Y_i	X_i	\hat{Y}_i	e_i	e_i^2	x_i	y_i	$x_i \times y_i$	x_i^2	\hat{Y}_i	e_i	e_i^2
4	1	4	0	0	-3	-3	9	9	2.92	1.08	1.17
5	4	7	-2	4	0	-2	0	0	7	-2	4
7	5	8	-1	1	1	0	0	1	8.36	-1.36	1.85
12	6	9	3	9	2	5	10	4	9.72	2.28	5.2
$\bar{Y} = 7$	$\bar{X} = 4$			14			19	14	7		12.22

$$\hat{Y}_i = 1.56 + 1.36X$$

$$\hat{\beta}_1 = \frac{\sum X_i \times Y_i}{\sum X_i^2} = \frac{19}{14} = 1.36 \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 7 - 1.36 \times 4 = 1.56$$

Svojstva regresijske funkcije:

1. Prolazi kroz sredinu uzoraka



$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$$

2. Sredina procijenjenog Y-a jednaka je sredini stvarnog Y-a

$$\begin{aligned} \bar{\hat{Y}} &= \bar{Y} \\ \hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 X \\ &= (\bar{Y} - \hat{\beta}_1 \bar{X}) + \hat{\beta}_1 X \\ &= \bar{Y} + \hat{\beta}_1 (X - \bar{X}) \\ \hat{Y} &= \bar{Y} \end{aligned} \quad \longrightarrow \quad \text{DOKAZ!}$$

3. Sredina reziduala je 0

$$\bar{e} = 0 \quad \frac{0}{n}$$

4. Reziduali e_i su nezavisni od procijenjenog Y-a (\hat{Y}_i)

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_1 X_i \\ \sum \hat{Y}_i e_i &= \hat{\beta}_1 \sum X_i \times e_i \\ &= \hat{\beta}_1 \sum X_i (Y_i - \hat{\beta}_1 X_i) \\ &= \hat{\beta}_1 \sum X_i \times Y_i - \hat{\beta}_1^2 \sum X_i^2 \\ &= \hat{\beta}_1^2 \sum X_i^2 - \hat{\beta}_1^2 \sum X_i^2 \\ &= 0 \end{aligned}$$

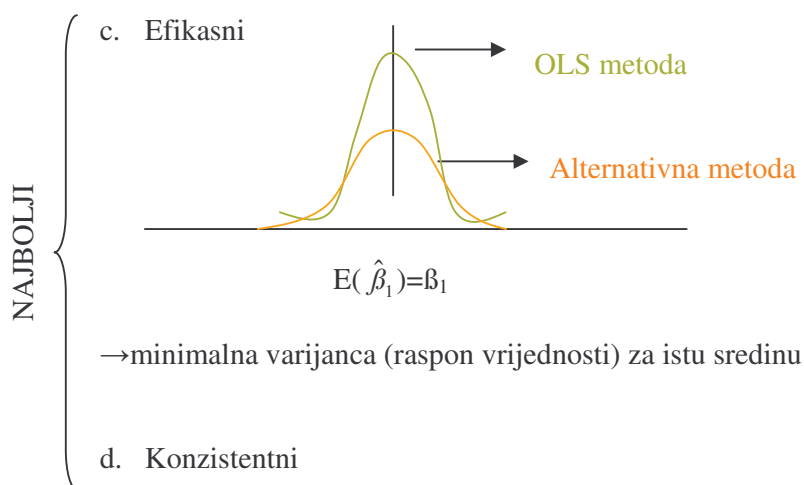
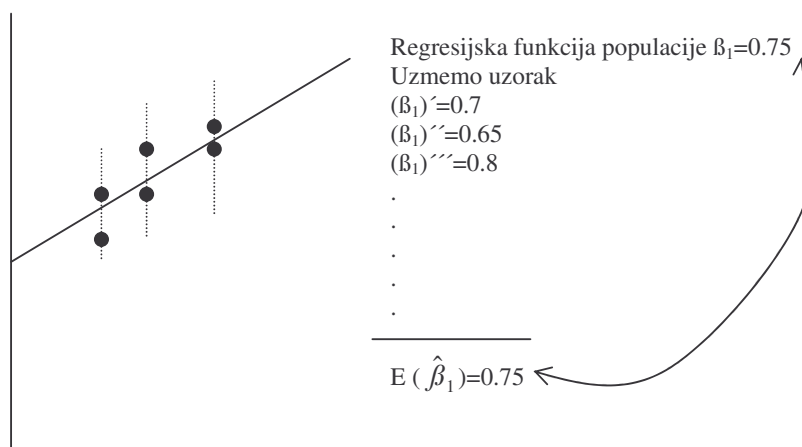
5. Reziduali e_i su nezavisni od X_i

Svojstva procjenitelja (parametara, koeficijenta regresije):

1. Oni su BLUE⁵

- Linearni → preduvjet za regresijsku funkciju
→ u linearnom odnosu prema zavisnoj varijabli
- Nepistrani → njihova očekivana vrijednost jednaka je stvarnoj vrijednosti

$$E(\hat{\beta}) = \beta$$



$E(\hat{\beta}_1)=0.7$ npr. uzeli smo prvih 20 uzoraka

Uzmemo još 20 i dobijemo očekivanu vrijednost za 40 opažanja

$$E(\hat{\beta}_1)=0.65$$

S povećanjem uzoraka, parametri konvergiraju populacijskoj sredini pa su konzistentni.

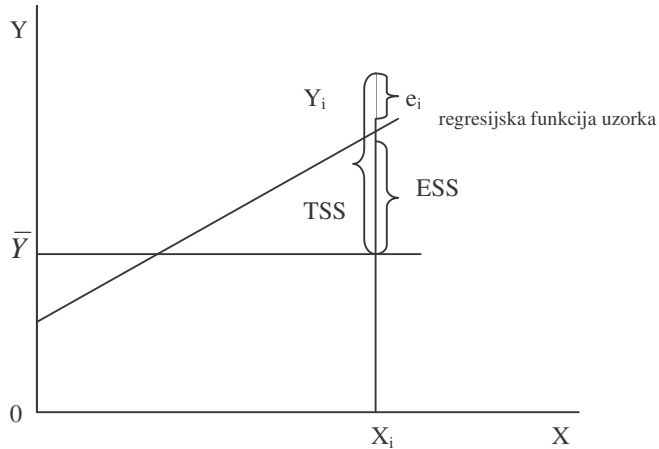
⁵ Best Linear Unbiased Estimator (Najbolji linearni nepristrani procjenitelji)

Koeficijent determinacije (koeficijent korelacije²)

$\mathcal{R}^2 \rightarrow$ govori koliko smo pomoću X protumačili varijaciju Y-a u odnosu na ukupnu varijancu Y-a

\rightarrow odnos između ESS i TSS

$r(x,y) \rightarrow$ koeficijent korelacije se radi između dvije varijable



e_i ili RSS (residual sum of square) \rightarrow neobjašnjeno odstupanje kvadrata

TSS \rightarrow total sum of square

ESS \rightarrow estimated sum of square \rightarrow objašnjeno odstupanje

$$\text{ESS ili SSE} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$\text{TSS ili SST} = \sum (Y_i - \bar{Y})^2$$

Što je pravac više nagnut, TSS će biti veći.

$$\text{RSS ili SSR} = \sum (\hat{Y}_i - \bar{Y})^2$$

$$\text{TSS} = \text{ESS} + \text{RSS}$$

$$\sum (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum (\hat{Y}_i - \bar{Y})^2 \quad /: \sum (Y_i - \bar{Y})^2$$

$$1 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / \sum (Y_i - \bar{Y})^2 + \sum (\hat{Y}_i - \bar{Y})^2 / \sum (Y_i - \bar{Y})^2$$

$$\mathcal{R}^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

$$\mathcal{R}^2 = \frac{\text{RSS}}{\text{TSS}}$$

$0 < \mathcal{R}^2 < 1$ Što je veći, to je bolji.

$Y = f(x_1, x_2, x_3, \dots, x_n)$ Što ima više varijabli, koeficijent determinacije će rasti.

Svojstva R^2 :

- nije nikada negativan (samo ako se radi o regresiji kroz ishodište)
- $0 \leq R^2 \leq 1$
- $r = \pm \sqrt{r^2}$

Stupnjevi slobode N-k (U softveru df.)

k → broj parametara koje ocjenjujemo

N=10

$$Y = \beta_0 + \beta_1 X_1 \quad df=8$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad df=7$$

Korigirani koeficijent determinacije

\bar{R}^2

$$\bar{R}^2 = 1 - \frac{\sum e_i^2 / (N - k)}{\sum (Y_i - \bar{Y})^2 / (N - 1)}$$

\bar{R}^2 ima u sebi trošak ubacivanja nove varijable (korigiran je za N-k)

Je li veći R^2 ili \bar{R}^2 ovisi o odnosu (N-k) i (N-1)

(N-k) < (N-1)

$\bar{R}^2 < R^2$ Vrijednost R^2 može ostati ista ili se povećati, ali nikako smanjiti ulaskom nove varijable u regresiju.

Uvođenjem nove varijable uma odstupanja će se smanjiti jer će nešto više objasniti. **Ali** gibe se i stupnjevi slobode. Ako je više objanila nego štetila, varijabla ostaje u modelu.

Testiranje nad koeficijentom

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

$$t = \frac{\hat{\beta} - \beta^*}{se(\hat{\beta})} = \frac{st\ var\ ni - novi}{se(\hat{\beta})} \quad \text{t-test}$$

se → standardna devijacija

T test → testiranje signifikantnosti regresijskih koeficijenata

→ pokazuje značajnost parametara

$$VAR(\hat{\beta}_1) = \frac{\sigma^2}{\sum X_i^2} \rightarrow \text{varijanca populacijskog modela (raspršenost)}$$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{(N - k)}$$

$$\sqrt{VAR(\hat{\beta}_1)} = se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum X_i^2}}$$

$$\text{VAR}(\hat{\beta}_0) = \frac{\sum X_i^2}{N \times \sum X_i^2} \times \sigma^2$$

$$\text{se}(\hat{\beta}_0) = \sqrt{\frac{\sum X_i^2}{N \times \sum X_i^2} \times \sigma^2}$$

STANDARDNA GREŠKA PROCJENITELJA

$$\hat{Y} = 0.5 + 2.5X_1 - 1.5X_2$$

se (0.1)

Testiram mogućnost da je $\beta_1=2$

$$H_0: \beta_1=2$$

$$t = \frac{2.5 - 2}{0.1} = \frac{0.5}{0.1} = 5$$

Ako je veći od 2 je značajan.

U Studentovoj t-distribuciji gledamo možemo li broj 5 prihvatiti ili odbaciti.

p → razina značajnosti testa → kolika je vjerojatnost da smo načinili grešku ako odbacujemo hipotezu

značajnost 0.05 5% p=1-z
pouzdanost 0.05 95% z=1-p

F-test

ANOVA tablica (analiza varijance)

Izvor odstupanja	SS	D.F.	MSS ⁶
ESS	$\sum_{i=1}^N (\hat{Y}_i - Y_i)^2$	1	SS:D.F.
RSS	$\sum_{i=1}^n e_i^2$	N-k	σ^2 → procijenjena varijanca modela (populacije)
TSS	$\sum_{i=1}^n (Y_i - \bar{Y})^2$		

Testiranje na više parametara

$$F = \frac{ESS}{M(RSS)} = \frac{\sum_{i=1}^N (\hat{Y}_i - Y_i)^2}{\hat{\sigma}^2}$$

$\frac{\text{objašnjena varijanca}}{\text{procijenjena varijanca}}$

⁶ sredina

$F_{(1, N-k)}$

$H_0: \beta_1 = \beta_2 = \dots = \beta_n = 0$ (nisu značajni za objašnjenje varijable Y)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

primjer $F = 35$ $p = 0,006$ odbacujemo
 primjer $F = 1.5$ $p = 0,15$

U F-testu je jedino bitan konstantni član.
 Često ćemo odbaciti H_0 .

Svi parametri u modelu koji se vezuju uz varijablu jesu 0. To F-test testira.

Testiranje na parametrima

$$F = \frac{(\mathcal{R}_{UR}^2 - \mathcal{R}_R^2) / m}{(1 - \mathcal{R}_{UR}^2) / (N - k)}$$

$\mathcal{R}_{UR}^2 \rightarrow$ neograničen

$\mathcal{R}_R^2 \rightarrow$ ograničen

$m \rightarrow$ ograničenja

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$ $\mathcal{R}_{UR}^2 = \dots$
 $H_0: \beta_2 = \beta_3 = 0$ ne spadaju

$Y = \beta_0 + \beta_1 X_1$ $\mathcal{R}_R^2 = \dots$ ograničen jer nismo uključili β_2 i β_3

$\mathcal{R}_R^2 < \mathcal{R}_{UR}^2$ te brojke koje dobijemo stavljamo u jednadžbu

$F_{(m, N-k)}$ $\left. \begin{matrix} \beta_2 = 0 \\ \beta_3 = 0 \end{matrix} \right\} m=2$ (ograničenja) $N-k \rightarrow$ stupnjevi slobode

Stavljamo UVIJEK stupnjeve slobode većeg modela, s više varijabli!!!

$H_0: \beta_1 + \beta_2 = 1$ $m=1$ $\beta_2 = 1 - \beta_1$

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$ $\mathcal{R}_{UR}^2 = 0,99$
 $Y = \beta_0 + \beta_1 X_1 + (1 - \beta_1) X_2 + \beta_3 X_3$
 $Y = \beta_0 + \beta_1 X_1 + X_2 - \beta_1 X_2 + \beta_3 X_3$
 $X_2 \rightarrow$ nema koeficijenta
 $Y - X_2 = \beta_0 + \beta_1 X_1 - \beta_1 X_2 + \beta_3 X_3$
 $Y - X_2 = \beta_0 + \beta_1 (X_1 - X_2) + \beta_3 X_3 \rightarrow$ konačni model iz kojeg ćemo dobiti $\mathcal{R}_R^2 \rightarrow$ gubimo stupnjeve slobode

$\mathcal{R}_R^2 = 0,98$

Da je velika razlika između \mathcal{R}_{UR}^2 i \mathcal{R}_R^2 , F će biti veliki broj, p će biti izrazito malen i hipoteza se odbacuje.

II. primjer:

$$H_0: \beta_1 - \beta_2 = 0 \quad m=1$$

$$\beta_1 = \beta_2$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

$$Y = \beta_0 + \beta_2 X_1 + \beta_2 X_2 + \beta_3 X_3$$

$$Y = \beta_0 + \beta_2 (X_1 + X_2) + \beta_3 X_3$$

ZADATAK:

Obiteljski dohodak X
Potrošnja Y

Y_i	X_i	x_i^7	y_i^8	$x_i \times y_i$	x_i^2	\hat{Y}_i^9	e_i^{10}	e_i^2	X_i^2	$(\hat{Y}_i - \bar{Y})^2$	$(Y_i - \bar{Y})^2$
70	80	-90	-41	3690	8100	65,1	4,9	24,01	6400	2106,81	1681
65	100	-70	.46	3220	4900	75,3	-10,3	106,09	10000	1274,49	2116
90	120	-50	-21	1050	2500	85,5	4,5	20,25	14400	650,25	441
95	140	-30	-16	480	900	95,7	-0,7	0,49	19600	234,09	256
110	160	-10	-1	10	100	105,9	4,1	16,81	25600	26,01	1
115	180	10	4	40	100	116,1	-1,1	1,21	32400	26,01	16
120	200	30	9	270	900	126,3	-6,3	36,69	40000	234,09	81
140	220	50	29	1450	2500	136,5	3,5	12,25	48400	650,25	841
155	240	70	44	3080	4900	146,7	8,3	68,89	57600	1274,49	1936
150	260	90	39	3510	8100	156,9	-6,9	47,61	67600	2106,81	1521
$\bar{Y} = 111$	$\bar{X} = 170$			16800	33000			337,3	322200	8583,3	8890

$$Y = \beta_0 + \beta_1 X$$

Izračunati koeficijente, R^2 (koeficijent determinacije), se, F i napraviti testiranje.

$$\beta_1 = \frac{\sum x_i \times y_i}{\sum x_i^2}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

$$\text{VAR}(\beta_1) = \frac{\sigma^2}{\sum x_i^2}$$

$$\text{VAR}(\beta_0) = \frac{\sum x_i^2}{N \times \sum x_i^2} \times \sigma^2$$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

$$F = \frac{ESS}{\sigma^2}$$

$$\hat{\beta}_1 = \frac{16800}{33000} = 0,51 \quad \beta_0 = 111 - 0,51 \times 170 = 24,3$$

⁷ $X_i - \bar{X}$

⁸ $Y_i - \bar{Y}$

⁹ zadano

¹⁰ $Y - \hat{Y}$

$$Y = 24,3 + 0,51X$$

$\begin{matrix} se(6,42) & (0,036) \\ t(3,79) & (14,17) \end{matrix}$

$$\sigma^2 = \frac{\sum e_i^2}{N-k} = 42,16 \quad \text{VAR}(\hat{\beta}_1) = 0,0013 \quad \text{VAR}(\hat{\beta}_0) = 41,16$$

$$\sqrt{\text{VAR}(\hat{\beta}_1)} = 0,036 \quad \sqrt{\text{VAR}(\hat{\beta}_0)} = 6,42$$

t-test $H_0: \beta_0 = 0$ $t = \frac{\hat{\beta} - \beta^*}{se(\hat{\beta})} = \frac{st\ var\ ni - novi}{se(\hat{\beta})}$

$H_0: \beta_1 = 0$

U apsolutnoj vrijednosti ako su veći od 2, odbacujemo. Softver daje i p vrijednosti.

$$t = \frac{0,51 - 0,8}{0,036} = -8,05 \quad |-8,05| > 2 \quad \text{odbacujemo}$$

ANOVA tablica

	SS	D.F.	MSS
ESS	8583,3	1	8583,3
RSS	337,3	8	42,16
TSS	8890	8	1111,25

$$F = \frac{ESS}{\sigma^2} = 203,59$$

$$\mathfrak{R}^2 = \frac{ESS}{TSS} = \frac{8583,3}{8890} = 0,966$$

$$\overline{\mathfrak{R}}^2 = 1 - \frac{\sum e_i^2 / N - k}{\sum (Y_i - \bar{Y})^2 / N - 1} = 1 - \frac{42,16}{8890 : 9} = 1 - 0,043 = 0,957$$

$$\overline{\mathfrak{R}}^2 < \mathfrak{R}^2$$

$$\hat{Y} = 24,3 + 0,51X$$

$\begin{matrix} (3,79) & (14,17) \end{matrix}$

Varijabla X je značajna za objašnjenje potrošnje.

F=203,59 Odbacujemo hipotezu da je X beznačajan za model

96% smo uspjeli s dohotkom objasniti kretanje potrošnje

Mjerne jedinice

GDP, M^{11} u milijunima

Y, C u milijunima

$$C = f(Y_D)$$

$$C = \beta_0 + 0,8 Y_D \quad R^2 = 0,89 \quad F = 200$$

(2,5)

$$C = \beta_0 + 800 Y_D \quad R^2 = 0,89 \quad F = 200$$

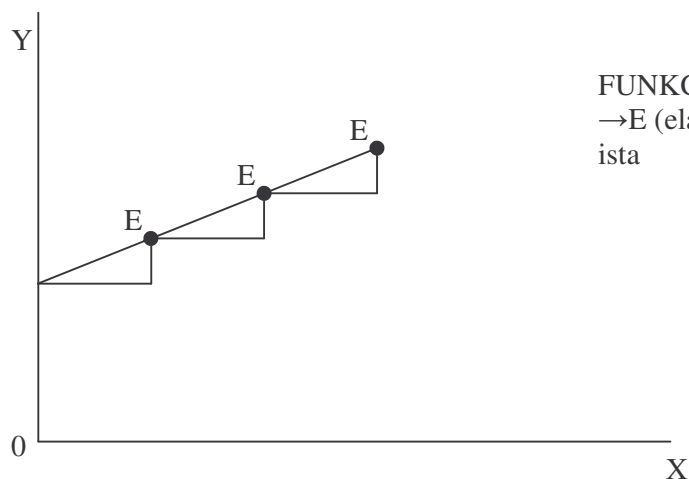
(2,5)

$$C = \beta_0 + 0,0008 Y_D \quad R^2 = 0,89 \quad F = 200$$

(2,5)

Y_D (milijuni)	C (tisuće)
2,5	1800
2,6	2000

Ako su varijable u različitim mjernim jedinicama, R^2 i t ostaju isti.



FUNKCIONALNE FORME

→ E (elastičnost) je u svakoj točki ista

Model:

1. LIN-LIN

Jednadžba

Nagib ($\frac{dY}{dX}$)

Elastičnost ($E_{Y,X} \frac{dY}{dX} \times \frac{X}{Y}$)

$$Y = \beta_0 + \beta_1 X$$

$$\beta_1$$

$$\beta_1 \times \frac{X}{Y}$$

2. LOG-LOG

$$\ln Y = \beta_0 + \beta_1 X$$

$$\beta_1 = \frac{d \ln Y}{d \ln X} = \frac{(\ln Y)' \times dY}{(\ln X)' \times dX}$$

$$\beta_1 \times \frac{Y}{X} \times \frac{X}{Y} = \beta_1$$

¹¹ novčana masa

$$\frac{\frac{1}{Y} \times dY}{\frac{1}{X} \times dX} = \frac{\frac{dY}{Y}}{\frac{dX}{X}} = \frac{dY}{dX} \times \frac{Y}{X}$$

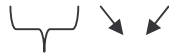
$Y = A \times K^\beta + L^\alpha$ Cobb-Douglasova funkcija proizvodnje

A → tehnološki napredak

K → kapital

L → rad

$$\ln Y = \ln A + \beta \ln K + \alpha \ln L$$



β_0 PARAMETRI

3. LOG-LIN Jednadžba

Nagib ($\frac{dY}{dX}$)

Elastičnost ($E_{Y,X} \frac{dY}{dX} \times \frac{X}{Y}$)

$$\ln Y = \beta_0 + \beta_1 X$$

$$\beta_1 = \frac{d \ln Y}{dX} = \frac{\frac{dY}{Y}}{\frac{dX}{X}} = \frac{dY}{dX} \times \frac{X}{Y}$$

$$\beta_1 \times Y \times \frac{X}{Y} = \beta_1 \times X$$

4. LIN-LOG $Y = \beta_0 + \beta_1 \ln X$

$$\beta_1 = \frac{dY}{d \ln X} = \frac{dY}{\frac{dX}{X}} = \frac{dY}{dX} \times \frac{1}{X}$$

$$\beta_1 \times \frac{1}{X} \times \frac{X}{Y} = \beta_1 \times \frac{1}{Y}$$

Stopa rasta: $\frac{dY}{Y}$

	Y
1991.	2
1992.	5
1993.	7
1994.	8

$$\left. \begin{array}{c} 1991. \\ 1992. \\ 1993. \\ 1994. \end{array} \right\} \frac{dY}{Y} = Y_2 - Y_1 = 5 - 2 = 3$$

1991.	1
1992.	5

$$\left. \begin{array}{c} 1991. \\ 1992. \end{array} \right\} \frac{dY}{Y} = Y_2 - Y_1 = \frac{5 - 1}{5} = \frac{4}{5}$$

→ od stope napravimo prosjeke → prosječna stopa rasta

→ izvodi se pomoću LOG-LIN modela

$$\ln Y = \beta_0 + \beta_1 X$$

β_1 → mjesečna stopa rasta $\times 12$ = godišnja stopa rasta

→ kvartalna stopa rasta $\times 4$ = godišnja stopa rasta

$$\beta_1 = \frac{d \ln Y}{dY} = \frac{\frac{dY}{Y}}{dT}$$

$\frac{dY}{Y}$ → stopa rasta

Mjesec	T(trend)	Kvartal
8	1	I
9	2	II
10	3	III
11	4	IV
12	5	V
1	6	VI

$$\ln Y = \beta_0 + \beta_1 X$$

(0,5)

znači da naša varijabla nema trenda



β je +



β je -

$$\beta_1 = \frac{\frac{dY}{dX}}{\frac{Y}{X}}$$

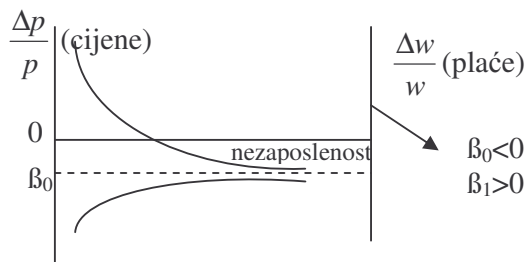
$$\frac{dY}{dX} = \beta_1 Y \longrightarrow \text{nagib: } \beta_1 Y$$

$$\frac{dY}{dX} \times \frac{X}{Y} = \beta_1 X \times \frac{X}{Y}$$

$$\beta_1 = \frac{dY}{d \ln X} = \frac{dY}{\frac{dX}{X}}$$

$$\frac{dY}{dX} = \beta_1 \frac{1}{X}$$

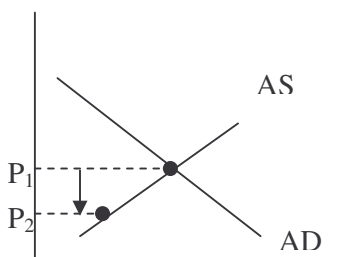
$$\frac{dY}{dX} \times \frac{X}{Y} = \beta \times \frac{1}{X} \times \frac{X}{Y}$$



Philipsova krivulja

$$\beta_0 < 0$$

$$\beta_1 > 0$$



$$Y = \beta_0 + \beta_1 \frac{1}{X}$$

Y	X
0,5	1
0,15	2
0,16	3

5. RECIPROČNI Jednadžba

$$Y = \beta_0 + \beta_1 \frac{1}{X}$$

$$\text{Nagib} \left(\frac{dY}{dX} \right)$$

$$\beta_1 = \frac{dY}{d\left(\frac{1}{X}\right)}$$

$$\text{Elastičnost} \left(E_{Y,X} \frac{dY}{dX} \times \frac{X}{Y} \right)$$

$$-\beta_1 \frac{1}{X^2} \times \frac{X}{Y} = -\beta_1 \times \frac{X}{X^2 Y}$$

$$\frac{d}{dX} \left(\frac{\beta_1}{X} \right) = \frac{0 - \beta_1 \times 1}{X^2}$$

$$= \frac{-\beta_1}{X^2}$$

DUMMY VARIABLE

→ koriste se kad želimo shvatiti neke nenumeričke varijable

→ mogu poprimiti dvije vrijednosti : 0 i 1

$$Y = \beta_0 + \beta_1 GS + \beta_2 D$$

Y → plaća, koja ovisi o GS (godinama staža)

→ pozitivna korelacija → Y raste s GS

→ možda ovisi i o stručnoj spremi

	GS	SS	
Ante	20	VSS	1
Miro	25	SSS	0
Anka	5	VSS	1
Marko	30	VSS	1

→ za kvalitativne modele → dummy varijable

D=0 → VSS

D=1 → SSS

SSS D=0 $Y = \beta_0 + \beta_1 GS$

VSS D=1 $Y = \beta_0 + \beta_1 GS + \beta_2$

$$Y = (\beta_0 + \beta_2) + \beta_1 GS$$

$$\underbrace{\hspace{1cm}} \quad \underbrace{\hspace{1cm}}$$

a

b

$$Y = 1000 + 50GS + 300D$$

Ante $Y = 1000 + 50 \times 20 + 300 = 2300$ kn

300 kn više radi VSS

Miro $Y = 1000 + 50 \times 25 = 2250$

Da smo za SSS imali D=1, a za VSS D=0, predznak bi za Miru ispred β_2 bio negativan.

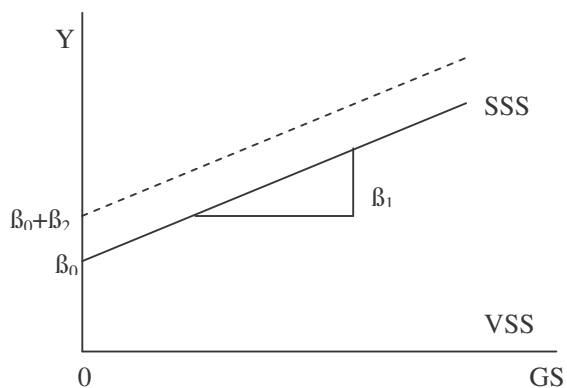
$R^2 \rightarrow$ biti će jednak

\rightarrow kvaliteta regresije ne ovisi o tome kojoj smo kvaliteti dali 0 ili 1

\rightarrow mijenja se parametar

$$Y = \beta_0 + \beta_1 GS \quad \text{bez D} \quad \text{SSS}$$

$$Y = \beta_0 + \beta_1 GS + \beta_2 \quad \text{s D} \quad \text{VSS}$$



INTERCEPT DUMMY
ostao je isti nagib, ali je
odsječak na ordinati
različit

pozitivna korelacija

S obzirom na godine staža, model procjenjuje Y za SSS i VSS.

	GS	SS	D1(SSS)	D2(VŠS)	D3(VSS)
Ante	20	VSS	0	0	1
Miro	25	SSS	1	0	0
Anka	5	VSS	0	0	1
Marko	30	VŠS	0	1	0

Jednu ćemo izbaciti. Svejedno je jer će β_0 poprimiti vrijednosti, npr. SSS

$$Y = \beta_0 + \beta_1 GS + \beta_2 V\check{S}S + \beta_3 VSS$$

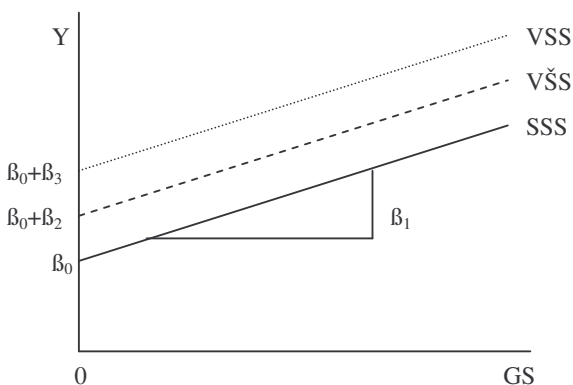
$$\text{SSS} \quad Y = \beta_0 + \beta_1 GS$$

$$\text{V}\check{S}\text{S} \quad Y = \beta_0 + \beta_1 GS + \beta_2 V\check{S}\text{S}$$

$$\text{VSS} \quad Y = \beta_0 + \beta_1 GS + \beta_3 VSS$$

$$Y = \beta_0 + \beta_1 GS + \beta_2$$

$$Y = \beta_0 + \beta_1 GS + \beta_3$$



pretpostavili $\beta_3 > \beta_2$

	GS	SS		SPOL	M→0 Ž→1
Ante	20	VSS	1	0	
Miro	25	SSS	0	0	
Anka	5	VSS	1	1	
Marko	30	VSS	1	0	

$$Y = \beta_0 + \beta_1 GS + \beta_2 SS + \beta_3 SPOL$$

(t) → pokazuje značajnost parametra

Ante $Y = \beta_0 + \beta_1 GS + \beta_2 = (\beta_0 + \beta_2) + \beta_1 GS$

Miro $Y = \beta_0 + \beta_1 GS$

Anka $Y = \beta_0 + \beta_1 GS + \beta_2 + \beta_3 = (\beta_0 + \beta_2 + \beta_3) + \beta_1 GS$

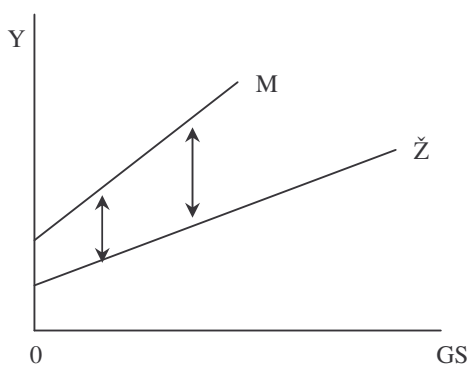
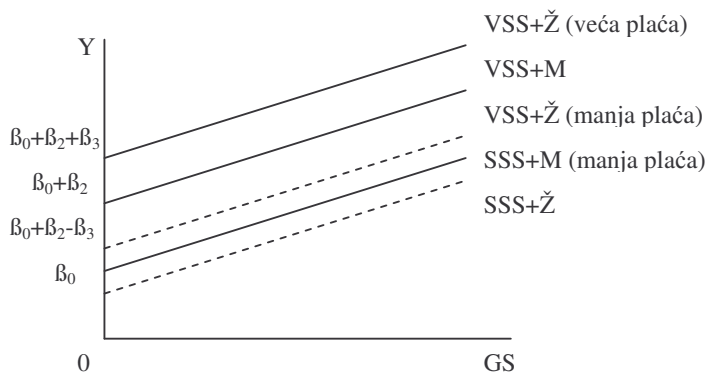
$H_0: \beta_3 = 0$ Ako odbacujemo $t > 2$, postoje razlike.

Ako je β_2 300, moramo gledati t vrijednost. Ako je npr. 0,25, onda ne igra veliku ulogu jer je t vrijednost mala.

* recimo na 95% pouzdanosti da testiramo, gledamo p vrijednost koja bi bila primjerice 0,7 što znači da ne možemo odbaciti hipotezu jer radimo 70% pogreške ako odbacimo.

* npr. t je 4,5 ispod β_3 → postoji diskriminacija

Marko $Y = \beta_0 + \beta_1 GS + \beta_2 = (\beta_0 + \beta_2) + \beta_1 GS$



SLOPE DUMMY → mijenja nagib

→ muškarcima se GS više uvažavaju nego ženama

→ ženama se onemogućava karijera

$$Y = \beta_0 + \beta_1 GS + \beta_2 D + \beta_3 GS \times D$$

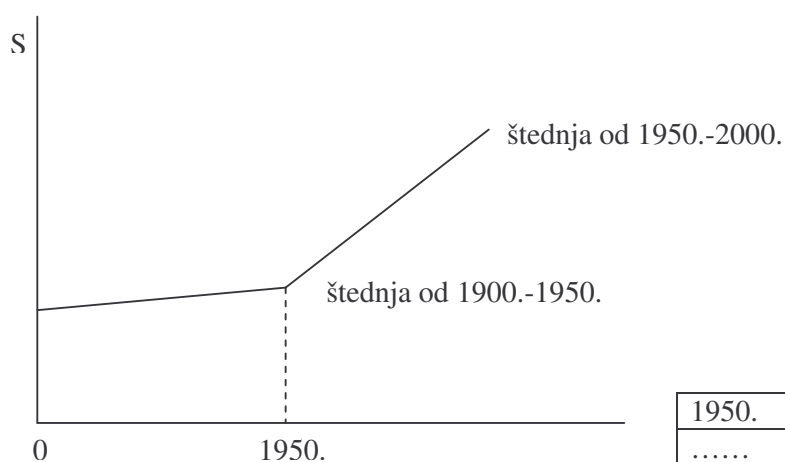
$\beta_2 D \rightarrow$ intercept $\beta_3 GS \times D \rightarrow$ slope

D=0 $Y = \beta_0 + \beta_1 GS$

D=1 $Y = \beta_0 + \beta_1 GS + \beta_2 + \beta_3 GS$
 $Y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) GS$

Razlika u odsječku na ordinati i u koeficijentu smjera

	Spol	GS	D	GS×D
Ante	M	20	0	0
Miro	M	25	0	0
Anka	Ž	5	1	5



nakon 1950.
prije 1950.

$$S = \beta_0 + \beta_1 Y_D + \beta_2 \times D + \beta_3 D \times Y$$

$$S = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) Y$$

$$S = \beta_0 + \beta_1 Y_D$$

	D	God×S
1950.	0
.....	0
	0
1950.	0
1951.	1
.....	1
	1
2000.	1

$$\beta_1 = \frac{dS}{dY} \quad \beta_1 + \beta_3 = \frac{dS}{dY}$$

Još jedan način izračunavanja razlike → raditi 2 regresije:

prije 1950. $S = 200 + 0,2 Y_D$

nakon 1950. $S = 150 + 0,3 Y_D$

usporediti

Nedostatak: moramo imati dovoljno opažanja

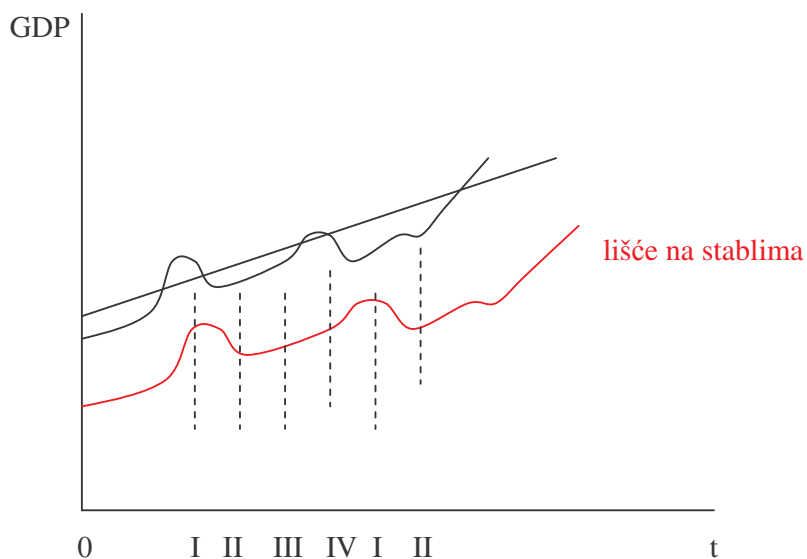
prije $S = \beta_0 + \beta_1 Y_D + \beta_2 \times D + \beta_3 D \times Y$

nakon 1950. $S = \beta_0 + \beta_1 Y_D$

nakon 1950. $S = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) Y$

Razlika u β i β_3 : ako je ta vrijednost velika, razlika u ponašanju je velika

→ dummy variјable se često koriste kod vremenskih nizova za desezoniranje



vremenski nizovi

→ trend-dugoročni

→ fluktuacije

→ druga varijabla koja nema veze

→ uvijek se koristi jedna dummy varijabla manje od broja varijabli

→ postoji korelacija između GDP-a i lišća na stablima

→ da bi izbjegli nelogičnosti korelacije vršimo desezoniranje

→ ubacujemo sezonske dummy varijable koje kupe na sebe pozitivna (sezonska) odstupanja

	S ₂	S ₃	S ₄
I	0	0	0
II	1	0	0
III	0	1	0
IV	0	0	1

$Y = \beta_0 + \beta_1 X + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4$ riješili smo sezonu; kod GDP-a bi se trebali riješiti trenda rasta

uključimo i TREND VARIJABLU

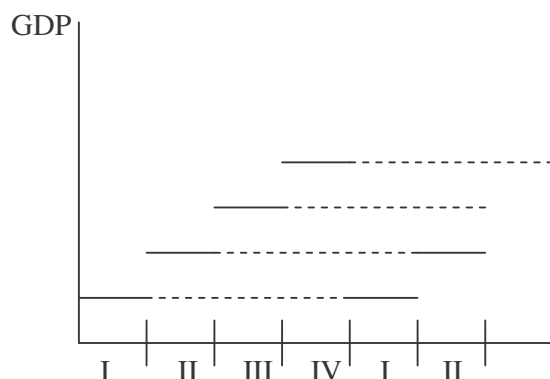
T
1
2
3
.
.
.
N

linearni

eksponencijalni trend T^2

$$Y = \beta_0 + \beta_1 X + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4 + \beta_5 T$$

β_2 pokazuje koliko je veći ili manji rast u II. kvartalu u odnosu na I. kvartal



→ dummy varijabla konstruira regresijske pravce

→ da izgledaju ravno omogućuje im TREND varijabla

$$Y = 20 + 5S_2 + 7S_3 - 4S_4$$

I 20

II 25 lin-lin → apsolutne vrijednosti

III 27

IV 16

$$\ln Y = 2 + 0,5S_2 + 0,7S_3 - 0,4S_4$$



50% veća u odnosu na I. kvartal

$$Y = \beta_0 + \beta_1 X + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4 + \beta_5 T$$



pokazuje realnu vezu između varijabli X i Y neovisno od sezone i trenda

$$\text{Tečaj kune} = \beta_0 + \beta_1 S_2 + \beta_2 S_3 + \dots + \beta_{11} S_{12} \quad \text{mjesečno}$$

(t) (t)

Gledamo t vrijednosti → ako su velike, špekulacija ima smisla, a ako je t oko 0, to nije sigurno

Tečaj

II. mjesec 7,4 7,4 kn za 1 €

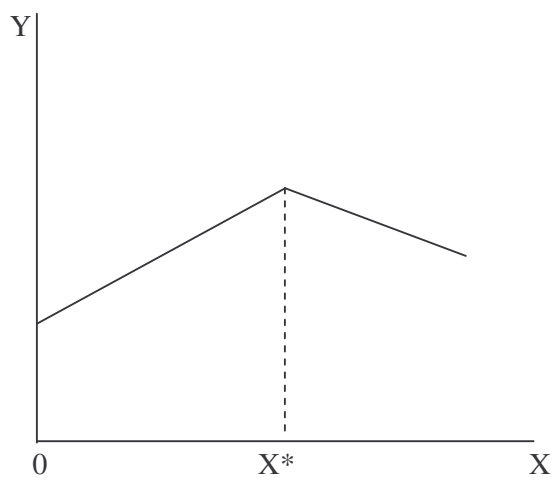
VII. mjesec 7,0 7 kn za 1 € → kupiti u VII. mjesecu da bi prodali u II. mjesecu



ako su visoke t vrijednosti, postoji velika vjerojatnost da će se to dogoditi

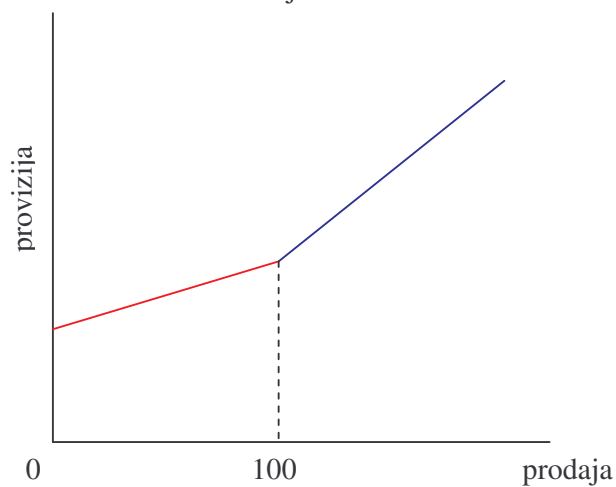
Postoji li sezona? Izbacimo dummy (\mathcal{R}^2_{UR}) i napravimo novu regresiju nakon čega rdimo F test.

PIECEWISE REGRESIJA



Pratimo regresijski pravac do neke vrijednosti nakon čega mijenjamo nagib

granična
vrijednost



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$Y = (\beta_0 + \beta_3) + \beta_1 X_1 + (\beta_2 + \beta_4) X_2$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 D + \beta_4 X_2 D$$

$$D = \begin{cases} 1, & X_2 > X^* \\ 0, & X_2 < X^* \end{cases}$$

Y	X ₁	X ₂	D	DX ₂
30	20	90	0	0
25	25	110	1	110
70	30	60	0	0
60	30	150	1	150
90	25	140	1	140

$$D=0 \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$D=1 \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 + \beta_4 X_2 = (\beta_0 + \beta_3) + \beta_1 X_1 + (\beta_2 + \beta_4) X_2$$

AUTOKORELACIJA

Autokorelacija odstupanja (reziduala)

→ veza između reziduala u vremenu t i reziduala $\rho_{e_{t-1}}$

$e_t = \rho_{e_{t-1}} + u_t$ → autokorelacija I. reda (jer stoji e_{t-1})

Imamo dvije serije → gledamo njihovu korelaciju

Y	\hat{Y}	e_t	e_{t-1}
30	28	2	-
25	32	-7	2
40	36	4	-7
			4

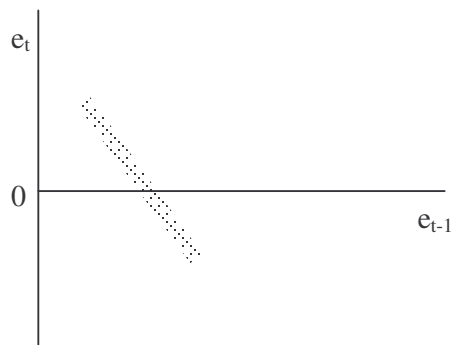
Ako postoji korelacija

$$\hat{Y} = \beta_0 + \beta_1 X + e_t$$

↓
 e_{t-1}

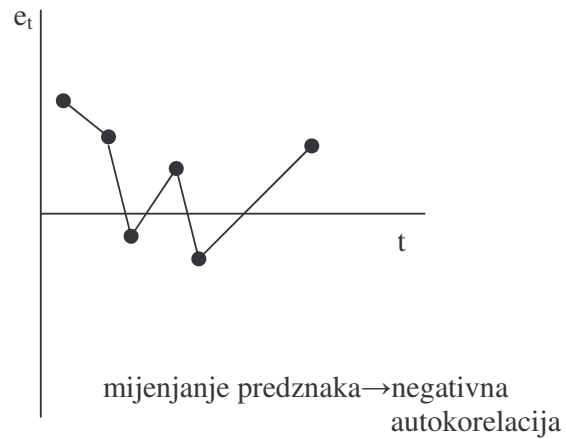
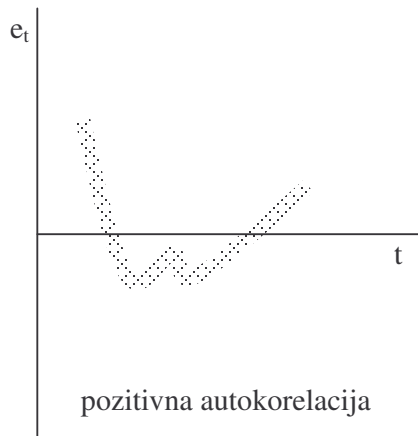
Narušena je pretpostavka o autokorelaciji¹²

- -
- -
+ +
- -

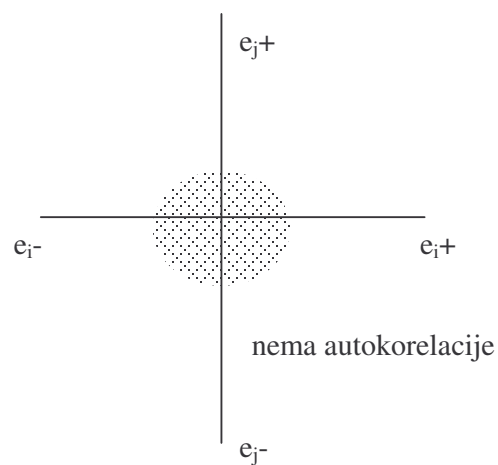
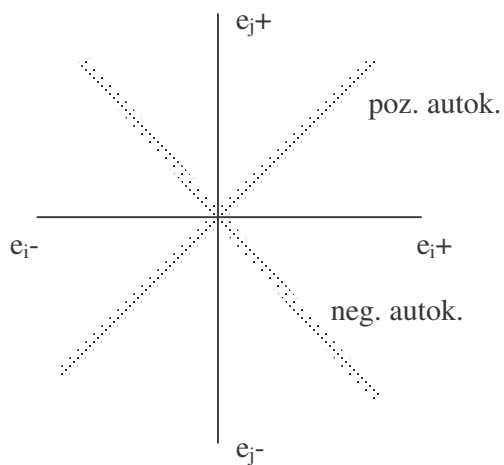


→ kad je pozitivna je pozitivna
→ u jednom trenutku se zadržava u negativnom segmentu

¹² Pretpostavka o autokorelaciji jest da nema autokorelacije



→ nema vidljivog uzorka u ponašanju



PRETPOSTAVKA O AUTOKORELACIJI

$$e_t = \rho e_{t-1} + u_t$$

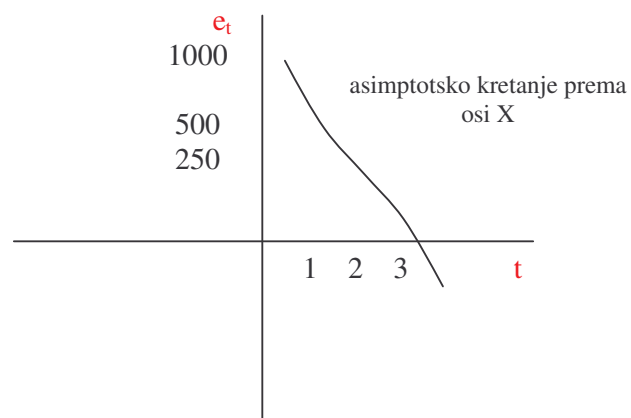
$$\rho = 0.5$$

$e_0 = 1000 \rightarrow$ početni uvjet

$$e_1 = \rho \times e_0 + u = 0.5 \times 1000 + 0 = 500$$

$$e_2 = 0.5 \times 500 + 0 = 250$$

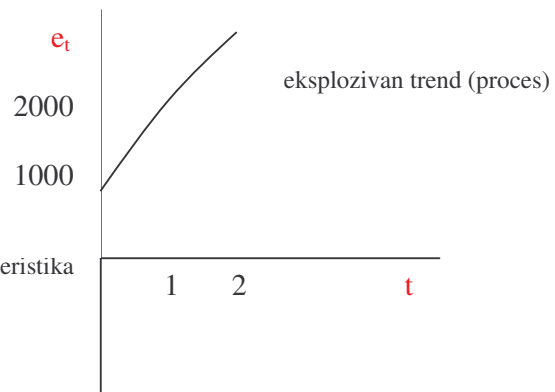
$$e_3 = 0.5 \times 250 + 200 = -75$$



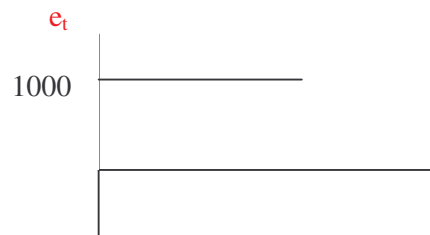
$$\rho = 2$$

$$e_0 = 1000$$

reziduali nemaju
eksplozivnih karakteristika
pa stavljamo $\rho < 1$



Da je $\rho = 1 \rightarrow$



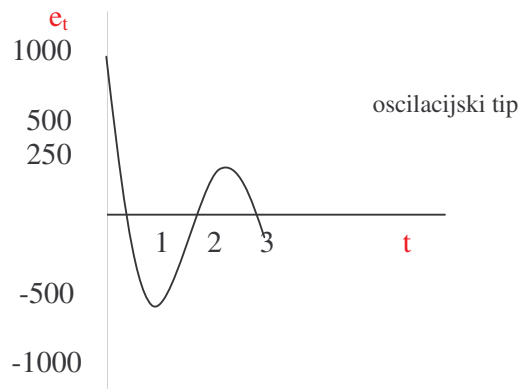
$$\rho = -0.5$$

$$e_0 = 1000$$

$$e_1 = -0.5 \times 1000 = -500$$

$$e_2 = -0.5 \times (-500) = 250$$

$$e_3 = -0.5 \times 250 = -125$$



$-1 < \rho < 1 \rightarrow$ stabilni procesi (ako ne želimo eksplozivne procese)

\downarrow
negativna
autokorelacija

\swarrow
pozitivna
autokorelacija

Zbog čega je prisutna autokorelacija? (UZROCI autokorelacije)

1. zbog zanemarene varijable

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

\rightarrow pravi proces

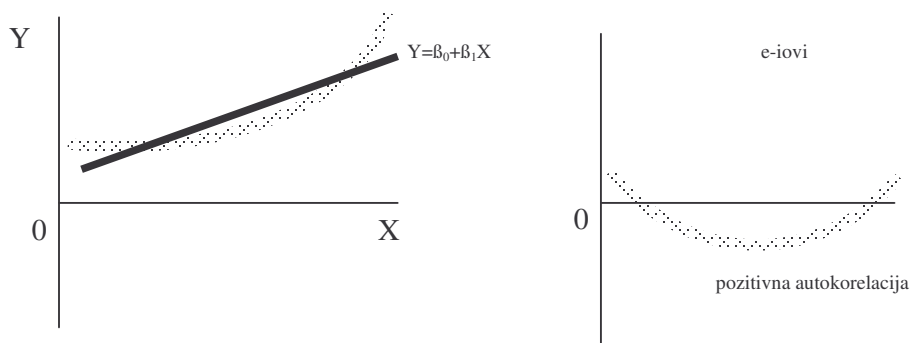
$$Y = \beta_0 + \beta_1 X_1 + e^*$$

\rightarrow izvedeno (mi napravili)

$$e^* = \beta_2 X_2 + e \rightarrow \text{u rezidualne otišli reziduali iz prvog modela s pravim procesom}$$

- povezano e i e* ako postoji kod X_2 vremenski uzorak (tada postoji i vremenski uzorak u rezidualima)
- postoji autokorelacija

2. zbog krive funkcionalne forme



Ako napravimo pravu funkcionalnu formu, rezidualni neće imati prepoznatljiv uzorak.

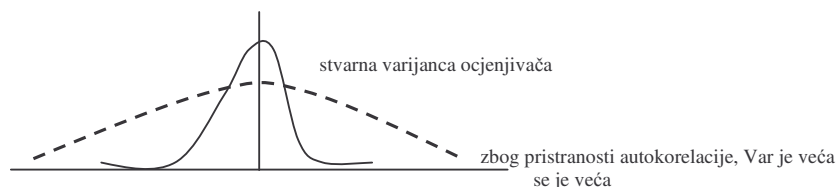
Posljedice autokorelacije:

1. Čista autokorelacija nema pristranost ocjenjivača → neprostranost
očekivani β (populacije) = procijenjeni β (uzorak)

$$E[\beta] = \hat{\beta}$$

Ne smije biti velika razlika između β i $\hat{\beta}$

2. Autokorelacija povećava varijancu ocjenjivača



3. OLS metoda će podcijeniti varijancu → precijenjene t vrijednosti

OLS će izvesti stvarnu varijancu
var s autokorelacijom se neće vidjeti
standardna devijacija će biti manja

$$Y = \beta_0 + 10X + e \rightarrow e\text{-iovi su autokorelirani pa to prebacuju na } X$$

se (5) → podcijenjena
t (2)

Otkrijemo da ima autokorelaciju
Var je značajna za model

$Y = \beta_0 + 10 X$ očistili smo od autokorelacije
se (5) Var X je neznačajna za model
t (1)

Kad očistimo, brojke u oba modela bi trebale biti slične
t vrijednosti bi, nakon čišćenja, trebale biti niže

Testiranje autokorelacije

DURBIN-WATSON TEST DW

1. Prikladan je kada regresija ima konstantni član
2. Prikladan je kada imamo autokorelaciju I. reda
3. Prikladan je kada model ne uključuje lagiranu varijablu

~~$$Y = \beta_1 X_1 + \beta_2 X_2$$~~
~~$$e = \rho_{et-1}$$~~
~~$$Y_{t-1}$$~~

$$Y_t = \beta_0 + \underbrace{\beta_1 Y_{t-1}}_{\text{lagirana varijabla}} + \beta_2 X_2 + \dots$$

$$DW = \frac{\sum_{i=2}^T (e_t - e_{t-1})^2}{\sum_{i=1}^T e_t^2}$$

obje sume su slične-razlikuju se u prvom članu

	e_t	e_{t-1}	$e_t - e_{t-1}$
1	20	-	-
2	-15	20	
3	30	-15	
	$\sum \approx \sum$		$\sum_{i=2}^T$

→ tiječ je o vremenskim varijablama; T je krajnje vrijeme

POZITIVNA AUTOKORELACIJA

$$\rho = 1 \quad \begin{aligned} e_t &= e_{t-1} \\ e_{t-1} &= e_t \end{aligned}$$

$$DW = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2} \rightarrow 0$$

NEGATIVNA AUTOKORELACIJA

$$\rho = -1 \quad \begin{aligned} e_t &= -e_{t-1} \\ e_t + e_{t-1} &= 0 \end{aligned}$$

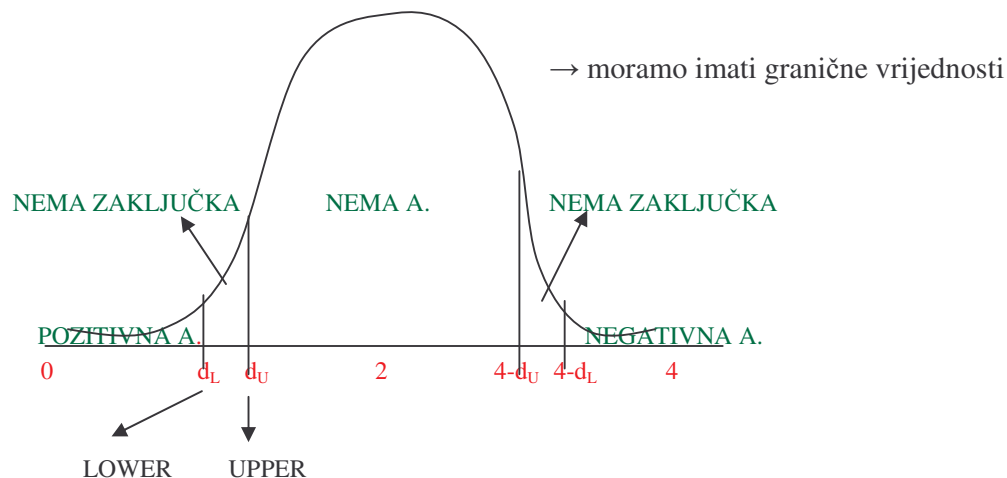
$$DW = \frac{\sum (2e_t)^2}{\sum e_t^2} = 4 \frac{\sum e_t^2}{\sum e_t^2} = 4$$

NEMA AUTOKORELACIJE

PRETPOSTAVKA $\rightarrow \text{Cov}=0$ (ako nema autokorelacije)

$$DW = \frac{\sum (e_t - 2e_t e_{t-1} + e_{t-1})}{\sum e_t^2} = \frac{\sum e_t^2 - 2 \sum e_t e_{t-1} + \sum e_{t-1}^2}{\sum e_t^2} = \frac{\sum e_t^2 + \sum e_{t-1}^2}{\sum e_t^2} = \frac{2 \sum e_t^2}{\sum e_t^2} = 2$$

Distribucija:



broj opservacija

n	k'=1	
	d _L	d _U
6		
7		
.....		
.....		
10	0.6	1.00

k'=k-1 Y=β₀+β₁X₁ k=2 k'=1
 k → broj varijabli

- ako dobijemo autokorelaciju, gledamo razlog
- loše specificirani model ili zanemarena varijabla
- ako ne znamo razlog, radimo tehnički zahvat (GLS)

Rješavanje autokorelacije :

GLS → generalizirana metoda najmanjih kvadrata

$$Y = \beta_0 + \beta_1 X_1 + \dots + \underbrace{(e_t - \rho e_{t-1})}_{u_t}$$

$$e_t = \rho e_{t-1} + u_t$$

$$u_t = e_t - \rho e_{t-1} \quad \text{ORCUTT-COHRANE}$$

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_n X_{nt} + e_t \quad / \times \rho$$

$$\rho Y_t = \rho \beta_0 + \rho \beta_1 X_{1t} + \rho \beta_2 X_{2t} + \dots + \rho \beta_n X_{nt} + \rho e_t \rightarrow \text{pomaknemo jednačbu razdoblje unazad}$$

$$\rho Y_{t-1} = \rho \beta_0 + \rho \beta_1 X_{1,t-1} + \rho \beta_2 X_{2,t-1} + \dots + \rho \beta_n X_{n,t-1} + \rho e_{t-1}$$

ODUZMEMO PRVU I TREĆU JEDNADŽBU

$$(Y_t - \rho Y_{t-1}) = (\beta_0 - \rho \beta_0) + \underbrace{\beta_1 (X_{1t} - \rho X_{1,t-1})}_{\substack{X_1 \text{ KO} \\ \text{nezavisna} \\ \text{varijabla}}} + \underbrace{\beta_2 (X_{2t} - \rho X_{2,t-1})}_{X_2 \text{ KO}} + \dots + \beta_n (X_{nt} - \rho X_{n,t-1}) + \underbrace{(e_t - \rho e_{t-1})}_{u_t}$$

$$e_t = \rho e_{t-1} + u_t \rightarrow \text{dobijemo } \rho \text{ pa će } \rho \text{ biti } \beta_1$$

I. način dobivanja ρ

$$\rho = 1 - \frac{DW}{2} \rightarrow \text{ako imamo DW}$$

II. način dobivanja ρ

β₁ će ostati isti, kao i β₂ i β₃..... → koeficijenti

Razlika će biti u β₀

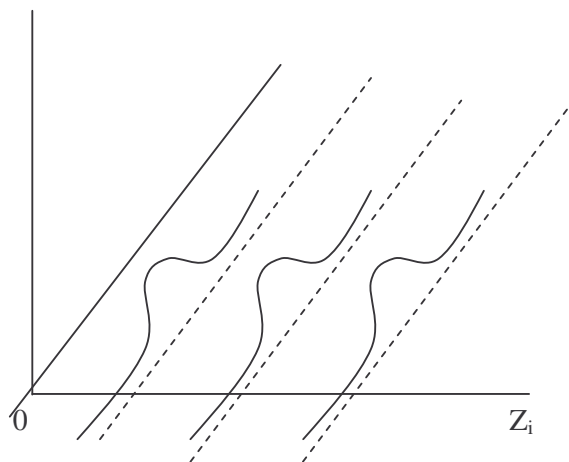
$$\beta_0^* = \beta_0 - \rho \beta_0$$

$$\beta_0^* = \beta_0(1 - \rho)$$

$$\beta_0 = \frac{\beta_0^*}{1 - \rho} \rightarrow \text{možemo iščupati } \beta_0 \text{ iz stvarnog modela}$$

Očekujemo manje t vrijednosti i $R^2 \rightarrow$ pa možda vidimo da i varijbla za koju smo smatrali da je bitna za model i nije baš bitna.

HETEROSKEDASTIČNOST



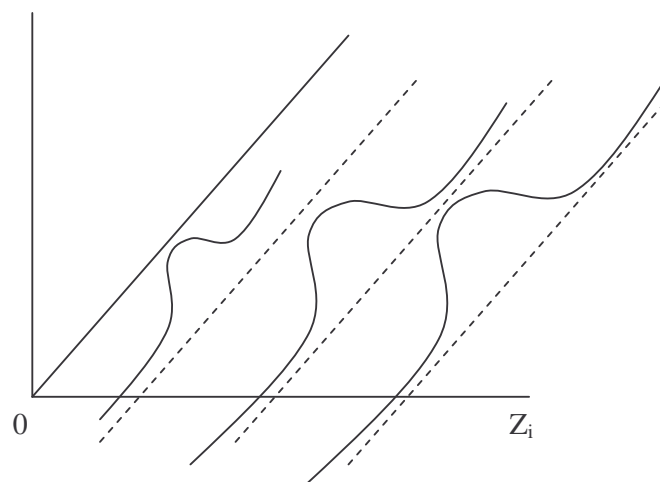
HOMOSKEDASTIČNOST

\rightarrow varijanca za različite Z_i -ove je jednaka

$$\text{Var}(e_i) = \sigma^2$$

$Z_i \rightarrow$ proporcionalni faktor

HETEROSKEDASTIČNOST \rightarrow kada se poveća neki X, poveća se i varijanca; postoji veza između varijance modela i nezavisne varijable



\rightarrow varijance se povećavaju
 \rightarrow različite su za svaku generaciju

$$\text{Var}(e_i) = \sigma^2$$

→ heteroskedastičnost se često pojavljuje u cross-sectionu (npr. različite zemlje, dohodak,...)

Može se javiti ako:

1. zaboravimo neku varijablu koja je važna

$$\text{UVOZ} = f(\text{GDP}, \text{PR})$$

$$\text{PR} = \frac{P_d}{P_w} \quad P_d \rightarrow \text{domaće} \quad P_w \rightarrow \text{strano}$$

$$\text{npr. uvoz} = \beta_0 + \beta_2 \text{PR} + e_j^*$$

$$e_j^* = e_j + \beta_1 \text{GDP}$$

→ ako je PR ↑, uvoz ↑

→ ako GDP ↑, uvoz ↑

2. ako nam je funkcionalna formula kriva → rijetko

Posljedice:

1. $E(\hat{\beta}) = \beta \rightarrow$ nema prostranosti koeficijenata
2. heteroskedastičnost povećava varijancu e_i -ova
3. potcjenjuje t i F statistiku

Testiranje:

1. PARK test → može se otkriti izgled varijance s obzirom na proporcionalan faktor
→ postupak:

a) regresija

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e$$

$Z_j \rightarrow$ neku od nezavisnih varijabli (X_1, X_2, \dots) mi moramo otkriti (unaprijed odrediti) koja je

- b) $\ln(e_i^2) \rightarrow$ logaritmiramo kvadrate reziduala

varijanca

$$\ln(e_i^2) = \alpha_0 + \alpha_1 \ln Z_i$$

→ navedeni postupak radimo za svaki X (nezavisnu varijablu) nakon čega uspoređujemo t vrijednosti (testiramo značajnost α pomoću t testa) → onaj koji ima najveću t vrijednost je proporcionalni faktor

2. Goldfeld-Quandt test

→ postupak:

I. poredati varijable (opažanja) po veličini proporcionalnog faktora

	C	Y	X ₁
Ante	1400	1500	35
Mira	9800	10000	36
Marko	4500	5000	40
Ive	999	1000	55

→ npr. shvatili smo da je Y proporcionalan faktor

	C	Y	X ₁
Ive	999	1000	55
Ante	1400	1500	35
Marko	4500	5000	40
Mira	9800	10000	36

II. → cijeli niz se dijeli na tri trećine (izbacimo 1/3 središnjih podataka) → radimo regresiju za prvu trećinu i treću trećinu i računamo njihove RSS-ove¹³ (RSS₁ i RSS₃)

III. $GQ = \frac{RSS_3}{RSS_1}$ računamo GQ koeficijent

Ako je $GQ=1$ → nema heteroskedastičnosti

Ako je $GQ>1$ → velika heteroskedastičnost

IV. $F(K, N''-K-1)$ N'' → broj N-ova u jednoj trećini

Ako je $GQ > F$ → onda nemamo homoskedastičnosti (postoji heteroskedastičnost)

H_0 : postoji homoskedastičnost (nul hipoteza kod ovog testa)

¹³ regresijske sume kvadrata

Kako se riješiti heteroskedastičnosti?

1. vidjeti jesmo li ispustili neku varijablu → redefiniranje varijabli
2. napraviti WLS¹⁴ → vagana sredina najmanjih kvadrata
→ najčešće za nečistu heteroskedastičnost

$$\circ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e \quad \rightarrow \frac{e}{Z}$$

npr. $Z = X_2 \rightarrow$ sve dijelimo sa Z-om (u ovom slučaju primjer Z-a je X_2)

$$\circ \frac{Y}{X_2} = \beta_0 \frac{1}{X_2} + \beta_1 \frac{X_1}{X_2} + \beta_2 \frac{X_2}{X_2} + \beta_3 \frac{X_3}{X_2} + \frac{e}{X_2}$$

$$\text{npr. dobijemo } YKO = 0.6 + 2.5 \frac{1}{X_2} + 0.5 \frac{X_1}{X_2} + 7 \frac{X_3}{X_2}$$

$\underbrace{\quad\quad\quad}_{\beta_2 \quad \beta_0}$

MULTIKOLINEARNOST

→ veza između nezavisnih varijabli

$$Y = f(X_1, X_2, \dots, X_n)$$

$$X_1 = g(X_2, X_3, \dots, X_n)$$

$$X_2 = z(X_1, X_3, X_4, \dots, X_n)$$

→ moguća je početnička greška radi multikolinearnosti

$$Y = f(i, \Pi^*)$$

Y je funkcija nominalnog kamatnjaka i inflacije

i već sadržava Π

$i = r + \Pi$ nominalni kamatnjak = realni kamatnjak + inflacija

→ slijedeća greška **DOMINANTNA VARIJABLA!**

pr. proizvodnja obuće = $f(L, K, \text{količina kože koja se troši})$

količina kože koja se troši → linearno vezana sa Q

→ deterministička veza usko vezana sa proizvodnjom

$$Q_t = \beta_0 + 0.5L_t + 0.3K_t + 0.9X_t \quad R^2 \approx 1 \quad \text{velik jer je X usko vezana s Q}$$

(200)

dominantna varijabla bi zasjenila druge nezavisne varijable

¹⁴ Weighted Least Square

→ dominantnu varijablu izbacimo iz modela

Posljedice multikolinearnosti:

1. procjenitelji će biti nepristrani
2. standardne greške koeficijenata će biti velike
3. imati ćemo male t vrijednosti → podcjenjuje t vrijednosti
4. β osjetljive na promjenu specifikacije i na promjenu nezavisnih varijabli u modelu

$$Q = \beta_0 + 0.5L + 0.3K$$

$$Q = \beta_0 + 0.5L + 0.3K + X$$

imamo multikolinearnost

ubacimo X

ako koeficijenti ostaju isti → robusni na promjene

ako se koeficijenti promijene → nisu robusni

Dvije promjene → ubacivanje/izbacivanje varijable
→ specifikacija modela

5. R^2 i F će biti nepromijenjeni
6. neće utjecati na procjenu nemultikolineiranih varijabli

Otkrivanje multikolinearnosti:

$$1. \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad R^2 = 0.98 \quad \text{veliki } R^2, \text{ male } t \text{ vrijednosti}$$

(0.6) (1.2) (0.8)

-98 % varijabli smo objasnili modelom

-ni jedna varijabla nije značajna

-postoji multikolinearnost

-neke nezavisne varijable su povezane

2. Koeficijenti korelacije između nezavisnih varijabli

$$r_1(X_1, X_2)$$

$$r_2(X_2, X_3)$$

$$r_3(X_1, X_3)$$

Problem se javlja ako imamo samo dvije varijable!

$$|r| > 0.8$$

3. Ako postoji multikolinearnost između više varijabli

VIF¹⁵ test → otkriva skrivene veze među varijablama; radimo regresiju između nezavisnih varijabli

-ako sumnjamo na kolinearnost, radimo druge regresije

¹⁵ Variance Inflationary Factor

$$\begin{aligned}\mathcal{R}^2_1 &= \text{VIF}_1 & X_1 &= \beta_0 + \beta_1 X_2 + \beta_2 X_3 \\ \mathcal{R}^2_2 &= \text{VIF}_2 & X_2 &= \alpha_0 + \alpha_1 X_1 + \alpha_2 X_3 \\ \mathcal{R}^2_3 &= \text{VIF}_3 & X_3 &= \gamma_0 + \gamma_1 X_2 + \gamma_2 X_1\end{aligned}$$

koeficijent determinacije \mathcal{R}^2 će pokazati vezu
-može li se X_1 izraziti kao linearna kombinacija između X_2 i X_3 varijable

$$\text{VIF} = \frac{1}{1 - \mathcal{R}^2}$$

\mathcal{R}^2 je 0 \rightarrow nema nikakve veze $\text{VIF}=1$
 \mathcal{R}^2 je 1 \rightarrow sve smo objasnili $\text{VIF}=\infty$
 $\text{VIF}>1$

Neki softveri, umjesto VIF vrijednosti \rightarrow TOLERANCIJA = $(1 - \mathcal{R}^2)$

Ako uključujemo više varijabli, shodno tome raste i \mathcal{R}^2

$$X_1 = \beta_0 + \beta_1 X_2 + \beta_2 X_3 + \dots$$

$\text{VIF}>5$ Prihvaćamo hipotezu o multikolinearnosti za mali broj nezavisnih varijabli
 (10-tak)

$\text{VIF}>10$ Za više varijabli

Kako se riješiti multikolinearnosti? Što učiniti?

1. Ništa

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad \mathcal{R}^2 = 0.8$$

(2.1) (2.2)

-nađemo multikolinearnost
 -ne stvara pristranost-ne remeti koeficijente
 -t vrijednosti oko 2 \rightarrow visoke

$$2. \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad \mathcal{R}^2 = 0.8$$

(2.1) (0.7) (0.8)

-multikolinearnost između X_2 i X_3

Izbaciti iz modela jednu (ili više) od multikolineiranih varijabli! (koja nije značajna) \rightarrow to utječe na pristranost koeficijenata

$$\text{npr. } C = \beta_0 + 0.8 Y_D + 0.6 S - 0.01 T \quad \mathcal{R}^2 = 0.9$$

(0.6) (0.9) (-2.5)

Samo porezi uzrokuju potrošnju?!

Nađemo vezu između Y_D i S

Koji izbaciti? \rightarrow t vrijednosti su poremećene radi multikolinearnosti

→ znanje iz ekonomske teorije → Y_D ostaje u modelu

$$C = \beta_0 + 0.75Y_D - 0.01T \quad \text{kad izbacimo } S$$

(7) (-2.5)

3. Transformacija nezavisnih varijabli

$$Y = f(i, \Pi)$$

$$r = i - \Pi$$

$$Y = f(r) \quad \text{gubimo neke informacije, ali nema multikolinearnosti}$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad \text{velika multikolinearnost između } X_1 \text{ i } X_2$$

ne smijemo izbaciti pa definiramo X_3 kao linearnu kombinaciju između dvije varijable

$$X_3 = X_1 + X_2 \rightarrow Y = \alpha_0 + \alpha_1 X_3$$

$$\begin{aligned} \text{pr. } X_3 &= \text{GDP} + \Pi = 3,200,000 + 0.08 = 3,200,000.08 \\ &= 3,200,000 + 0.15 = 3,200,000.15 \\ &= 3,400,000 + 0.15 = 3,400,000.15 \end{aligned}$$

Promjena inflacije ne mijenja ništa, dok promjena GDP-a itekako mijenja
→ veličine nisu ravnomjerne

Moramo ih prikazati ravnomjerno

Π pomnožiti s 1,000,000 npr.

$$Y = \beta_0 + 0.8X_3$$

\downarrow
 linearna kombinacija
 X_1 i X_2

→ ne možemo izdvojiti zasebne efekte

→ gubimo informacije

RESET TEST

-teško je izabrati nezavisne varijable, pogotovo ako prije nije bilo nikakvih istraživanja
-imamo dvije vrste mogućnosti:

1. da smo isključili varijablu koja je bitna
2. da smo isključili varijablu koja nije bitna

Ad 1. Isključenje bitne varijable

npr. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$
 $Y = \beta_0 + \underbrace{\beta_2 X_2 + e}_{e^*}$

$$E[\hat{\beta}_2] = \beta_2 + \beta_1 \times f(r_{1,2}) \longrightarrow \text{o tome ovisi hoće li } \beta_2 \text{ biti veći ili manji}$$

$$e^* = e + \beta_1 X_1$$

-stvara se pristranost koeficijenta

$$E[\hat{\beta}_2] \neq \beta_2$$

$$\beta_1 f(r_{1,2}) \rightarrow \hat{\beta}_2 > \beta_2$$

+ +

$$- + \rightarrow \hat{\beta}_2 < \beta_2$$

$$+ - \rightarrow \hat{\beta}_2 < \beta_2$$

$$- - \rightarrow \hat{\beta}_2 > \beta_2$$

-pristranosti nema kada je $\beta_1 = 0$ (varijabla X_1 nije značajna) ili kada ne postoji korelacija između X_1 i X_2 ($f(r_{1,2}) = 0$)

-kako ispraviti taj problem? Uključiti varijablu koja nedostaje (teško ju je otkriti)

-pristranost može biti mala, pa onda prihvaćamo model kakav jest

-početnici će često uključivati sve varijable

Ad 2. Uključivanje nebitne varijable

$$Y = \beta_0 + \beta_1 X_1 + e$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e^*$$

$$e^* = e - \beta_2 X_2$$

-ubacivanje nevažnih varijabli neće stvarati pristranost

Posljedice \rightarrow povećati će se varijanca ostalih varijabli, smanjiti će se njihove vrijednosti, smanjiti će se R^2 korigirani

Pravi odabir:

1. teorija
2. gledati t testove-jesu li varijable značajne
3. gledati R^2 korigirani-je li sveukupna kvaliteta regresije u redu
4. imamo li pristranost u koeficijentima ili mijenjaju li se bitno koeficijenti varijable kada uključujemo ili isključujemo neke od njih

Primjer:

Potražnja za brazilskom kavom $\rightarrow D_{BC}$

$$D_{BC} = f(P_{BC}, P_t, Y_D)$$

$P_{BC} \rightarrow$ cijena brazilske kave -

$P_t \rightarrow$ cijena čaja +

$Y_D \rightarrow$ dohodak +

$$D_{BC} = 8.1 + 7.8 P_{BC} + 2.4 P_t + 0.0035 Y_D$$

(0.5) (2.0) (3.5)

$$R^2 = 0.6$$

-cijena brazilske kave nema nikakav utjecaj na potrošnju kave

-izbacujemo varijablu \rightarrow cijena brazilske kave

$$D_{BC} = 9.3 + 2.6 P_t + 0.0036 Y_D \quad R^2 = 0.61$$

(2.6) (4.0)

-moguće je da je elastičnost za kavom mala ili nikakva u odnosu na cijenu kave

-uključujemo i cijenu kolumbijske kave

$$D_{BC} = f(P_{BC}, P_t, Y_D, P_{CC})$$

$P_{BC} \rightarrow$ cijena brazilske kave -

$P_t \rightarrow$ cijena čaja +

$Y_D \rightarrow$ dohodak +

$P_{CC} \rightarrow$ cijena kolumbijske kave +

$$D_{BC} = 10.0 + 8.6 P_{CC} - 5.6 P_{BC} + 2.6 P_t + 0.0030 Y_D$$

(2.0) (-2.8) (2.0) (3.0)

$$R^2 = 0.65$$

-cijena kolumbijske kave utječe na potražnju za brazilskom kavom

-teorija je zadovoljena, R^2 se povećao, predznaci su očekivani (pristranost), t vrijednosti su značajne

β_2 je precjenjivao stvarni β_2

-moramo se više oslanjati na ekonomsku teoriju nego na statističku značajnost

-način na koji se to radi \rightarrow TRAŽENJE SPECIFIKACIJE:

1. DATA MINING

\rightarrow neetična metoda, oslanja se na velik broj regresija s velikim brojem varijabli, tragova, različite funkcionalne forme

2. STEP WISE REGRESIJA

→ temelji se na tome da npr. uzmemo 6 varijabli s kojima se učine sve moguće kombinacije

$$Y = f(X_1)$$

$$Y = f(X_1, X_2)$$

$$Y = f(X_2)$$

$$Y = f(X_2, X_3)$$

$$Y = f(X_1, X_2, X_3, X_4)$$

-gleda koja kombinacija daje najveći R^2

$$Y = f(X_1, X_2, X_3, \dots, X_n) \quad \text{maksimalan } R^2$$

3. SEKVENCIJALNO TRAŽENJE SPECIFIKACIJE

→ istraživači teže sefiniranju početne jednačbe, ubacuju ili izbacuju varijablejednu po jednu i gledaju što će se dogoditi

→ sa što manje varijabli objasniti neki model

REMSEY TEST

→ njime testiramo funkcionalnu formu modela

→ temelji se na tome da uzima vrijednosti koje bi trebalo procijeniti kakve bi naše vrijednosti trebale biti

$$\hat{Y}_2, \hat{Y}_3, \hat{Y}_4$$

Izvođenje testa:

1. napravimo regresiju

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

→ dobivamo procijenjenu vrijednost Y

2. $\hat{Y}_2, \hat{Y}_3, \hat{Y}_4 \rightarrow$ potenciramo Y

$$3. \quad \hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \hat{Y}^2 + \beta_4 \hat{Y}^3 + \beta_5 \hat{Y}^4 + e$$

u $\hat{Y}_2, \hat{Y}_3, \hat{Y}_4$ ušle su sve greške

$$H_0 = \underbrace{\beta_3 = \beta_4 = \beta_5}_{=0} \rightarrow \text{da bi model bio značajan}$$

testiramo pomoću F testa

Cilj je dokazati da su $\beta_3, \beta_4, \beta_5$ jednake 0

ZADACI

OPĆE STVARI

Kada se pojavi A:/ upisati **sx** i pritisnuti enter

Sx.exe 2 puta kliknuti

Dobijemo BASE MENU

Na početni screen se vraćamo s **ESCAPE**

U **file management** idemo ako imamo sačuvane podatke

Inače u **Data management**

Ako pogriješimo broj pri upisivanju brojeva, samo na njegovom mjestu počnemo pisati drugi broj

Do opcija dođemo s **F3**

Linear Models- tu se nalazi regresija pod **Multiple Regression**

Moramo paziti na E-1 i slično, kao i na pozitivne predznake

P je razina značajnosti testa tj, greška koju smo načinili ako odbacimo hipotezu

R^2 ima problem jer pri ubacivanju novih varijabli ne kažnjava pa uzimamo korigirani tj. \bar{R}^2

Pod **SOURCE** imamo Anova tablicu, tj. analizu varijance

Regression (ESS) to smo uspjeli objasniti

Residual (RSS) nismo objasnili

Total (TSS)

D.F. stupnjevi slobode

Residual pod MS je σ^2

Standardna devijacija je σ koja kaže koliko u prosjeku jedinica odstupa naš uzorak

T- test radimo kada imamo jednu vrijednost

F-test radimo kada imamo zajedničke hipoteze

ZADATAK 1:

Obiteljski dohodak X
Potrošnja Y

Data management- Data entry

Enter the data set ID npr. zadatak 1
List the new variable names x (razmak) y

F1

Upisujemo listu za X pa za Y

Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

Kad upišemo **ESCAPE**

Na Base menu

Linear Models- Multiple Regression

Enter the name of the dependent variable Y
List the independent variables X

Dobijemo linearnu regresiju varijable Y

$$Y = 24.455 + 0.509X$$

(6.4138) (0.0357)

$$R^2 = 0.9621$$

$$\mathbb{R}^2 = 0.9573$$

$H_0: \beta = 0$ Dopuštamo 5%
Odbacujemo je pa β nije 0.

$$F\text{-test} \quad \frac{ESS}{RSS} \quad t = \frac{st \text{ var } ni - novi}{se}$$

$$\sigma^2 = 42.159$$

ESCAPE- Anova table- Residuals

Fitted Value YKAPA (bez razmaka)

Residual E

Data management- Edit data

Dobijemo tablicu s y kapa i e

Radimo t-test : može li β biti 0.8

$$t = \frac{0.54 - 0.8}{0.036} = -7.22$$

(0.54 je bilo zadano u zadatku)

Probability Distribution

Function ID: T2 (enter)

X: 7.22

Df: 8 (broj stupnjeva slobode: broj opažanja- broj varijabli)

Dobijemo vjerojatnost za vrijednost t

Možemo odbaciti jer činimo grešku manju od 5%

ZADATAK 2:

Potrošnja piletine (y), realni raspoloživi dohodak (X_1), realna cijena na malo piletine (X_2), realna cijena na malo svinjetine (X_3) i realna cijena na malo govedine (X_4) u SAD odb1960. do 1982. su sljedeće:

Godina	Y	X_1	X_2	X_3	X_4
1960	27,8	397,5	42,2	50,7	78,3
1961	29,9	413,3	38,1	52	79,2
1962	29,8	439,2	40,3	54	79,2
1963	30,8	459,7	39,5	55,3	79,2
1964	31,2	492,9	37,3	54,7	77,4
1965	33,3	528,6	38,1	63,7	80,2
1966	35,6	560,3	39,3	69,8	80,4
1967	36,4	624,6	37,8	65,9	83,9
1968	36,7	666,4	38,4	64,5	85,5
1969	38,4	717,8	40,1	70	93,7
1970	40,4	768,2	38,6	73,2	106,1
1971	40,3	843,3	39,8	67,8	104,8
1972	41,8	911,6	39,7	79,1	114
1973	40,4	931,1	52,1	95,4	124,1
1974	40,7	1021,5	48,9	94,2	127,6
1975	40,1	1165,9	58,3	123,5	142,9
1976	42,7	1349,6	57,9	129,9	143,6
1977	44,1	1449,4	56,5	117,6	139,2
1978	46,7	1575,5	63,7	130,9	165,5

1979	50,6	1759,1	61,6	129,8	203,3
1980	50,1	1994,2	58,9	128	219,6
1981	51,7	2258,1	66,4	141	221,6
1982	52,9	2478,7	70,4	168,2	232,6

- a) Odredite koeficijente regresije potražnje za piletinom kao funkcije dohotka, cijene piletine, govedine i svinjetine.
- b) Testirajte zajedničke hipoteze:
- b.a. $H_0: \beta_3 = \beta_4 = 0$
 - b.b. $H_0: \beta_1 + \beta_2 = 0$
 - b.c. $H_0: \beta_3 + \beta_4 = 1$
 - b.d. $H_0: \beta_2 + \beta_3 + \beta_4 = 0.5$
 - b.e. $H_0: \beta_1 - \beta_3 = 0.4$
- c) Testirajte posebno hipotezu da:
- c.a. $H_0: \beta_1 = 0.5$
 - c.b. $H_0: \beta_2 = -0.5$

Data management- Data entry unijeti podatke iz tablice

Linear models- Multiple regression

Dependent variable Y

Independent variables X1 X2 X3 X4

F1

Pod a)

$$Y = 37.232 + 0.00501X_1 - 0.611X_2 + 0.198X_3 + 0.0695X_4$$

(3.72) (0.004) (0.1628) (0.064) (0.051)
(10.01) (1.02) (-3.75) (3.11) (1.36)

Što regresija znači?

Bez obzira na x_1, x_2, x_3, x_4 , potrošnja piletine je 37,232 kg

Ako se poveća dohodak za 100 prodaja će porasti za 5.01

Rast cijene utječe na pad potrošnje

$$R^2 = 0.9426$$

$$\mathbb{R}^2 = 0.9298$$

Pod b)

b)a)

$$H_0: \beta_3 = \beta_4 = 0 \quad \text{Tu su dva ograničenja!!!!}$$

$$\text{Prvotna formula} \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

$$\text{S ograničenjima} \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Treba nam R^2 za F-test

ESCAPE

Idemo u **Linear Model- Multiple Regression**

Pobrisati X3 i X4

F1

$$R^2=0.9108$$

Kad smo maknuli varijable i koeficijent determinacije je manji

Na početku je $R^2=0.9426$ i to je neograničeni R tj. R_{UR} (unrestricted)

Sada je $R^2=0.9108$ i to je ograničeni R tj. R_R (restricted)

$$F = \frac{(R^2_{UR} - R^2_R) / m}{(1 - R^2_{UR}) / N - k} \quad \text{m-broj ograničenja N-broj opažanja k-broj varijabli}$$

$$F = \frac{\frac{0.9426 - 0.9108}{2}}{\frac{1 - 0.9426}{18}} = 6,82$$

Idemo na **Probability**

Function ID: FP (enter)

X: 6,82

Df num: 2 (ograničenja)

Df den: 20 (N-k)

b)b)

$$H_0 = \beta_1 + \beta_2 = 0 \quad ; m=1$$

$$\beta_1 = -\beta_2$$

Treba nam R^2

$$\text{Prvotna formula } Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

$$\text{Sadašnja formula } Y = \beta_0 - \beta_2 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

$$Y = \beta_0 + \beta_2 (-X_1 + X_2) + \beta_3 X_3 + \beta_4 X_4$$

Stavimo kao jednu varijablu $X_{12} = X_2 - X_1$

Idemo u **Data management- Trasformation**

Upišemo $X_{12} = X_2 - X_1$

F1

Idemo u **Linear Model- Multiple Regression**

Independent variables $X_{12} X_3 X_4$

F1

Treba nam R^2

$$R^2 = 0.8991$$

$$F = \frac{\frac{0.9426 - 0.8991}{1 - 0.9426}}{19} = 14.4$$

Idemo na **Probability**

Function ID: FP
X: 14.4
Df num: 1
Df den: 19

Dobili smo 0.00122 što znači da odbacujemo hipotezu jer činimo grešku manju od 5%

b)c)

$$H_0 = \beta_3 + \beta_4 = 1$$

$$\beta_4 = 1 - \beta_3$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + (1 - \beta_3) X_4$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + X_4 - \beta_3 X_4$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_3 - X_4) + X_4$$

Stavimo kao jednu varijablu: $X_{34} = X_3 - X_4$

$$Y - X_4 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_3 - X_4)$$

Opet stavimo kao posebnu varijablu: $YX_4 = Y - X_4$

Idemo u **Data management- Trasformation**

Upišemo $YX_4 = Y - X_4$

F1

Izbrišemo jer je računalo memoriralo tu transformaciju, a neće priznati dvije odjednom i

upišemo: $X_{34} = X_3 - X_4$

F1

Idemo u **Linear Model- Multiple Regression**

Dependent variable je sada YX_4

Independent variables $X_1 X_2 X_{34}$

F1

Zanima nas samo R^2 koji iznosi 0.9929

Izračunavamo F

$$F = \frac{\frac{0.9426 - 0.9929}{1}}{\frac{1 - 0.9426}{18}} = -15.77 \quad R_R > R_{UR}$$

Kad dobijemo negativan predznak F-a, što nije logično, u Probability pod x upisujemo vrijednost 0!!!!

Idemo na **Probability**

Function ID: FP (enter)

X: 0

Df num: 1

Df den: 18

Dobijemo da je p 1.0000 odnosno 100 % što značida ne možemo odbaciti hipotezu jer bi radili inače grešku od 100 %

b)d)

$$H_0 = \beta_2 + \beta_3 + \beta_4 = 0.5$$

Najjednostavnije je krenuti od zadnje varijable jer tek na kraju dobijemo transformaciju!

$$\beta_4 = 0.5 - \beta_2 - \beta_3$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + (0.5 - \beta_2 - \beta_3) X_4$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + 0.5 X_4 - \beta_2 X_4 - \beta_3 X_4$$

$$Y - 0.5 X_4 = \beta_0 + \beta_1 X_1 + \beta_2 (X_2 - X_4) + \beta_3 (X_3 - X_4)$$

Imamo tri transformacije, ali moramo paziti jesmo li koju već upisali!
Pošto smo zadnju već upisali, tj. $X_{34}=X_3-X_4$, moramo upisati dvije preostale!

Idemo u **Data management- Trasformation**

Upišemo $Y_{5X4}=Y-0.5*X_4$

F1

Izbrišemo i upisujemo drugu

$X_{24}=X_2-X_4$

F1

Idemo u **Linear Model- Multiple Regression**

Dependent variable je sada Y_{5X4}
Independent variables $x_1 x_{24} x_{34}$

F1

$R^2=0.9668$

$R_R > R_{UR}$, tako da ni ne trebamo računati F jer znamo da će doći negativan!

Idemo na **Probability**

Function ID: FP

X: 0

Df num: 1

Df den: 18

Dobijemo da je p ponovno 1.00, tj. 100% što znači da ne možemo odbaciti hipotezu.

b)e)

$H_0: \beta_1 - \beta_3 = 0.4$

$\beta_1 = 0.4 + \beta_3$

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$

$Y = \beta_0 + (0.4 + \beta_3) X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$

$Y = \beta_0 + 0.4 X_1 + \beta_3 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$

$Y - 0.4 X_1 = \beta_0 + \beta_2 X_2 + \beta_3 (X_1 + X_3) + \beta_4 X_4$

Imamo dvije transformacije pa opet gledamo što moramo upisati kao transformaciju, odnosno što nemamo prije zapisano.

Idemo u **Data management- Trasformation**

Upišemo $Y4X1=Y-0.4*X_1$

F1

Izbrišemo upisano i upisujemo drugu transformaciju

$X13=X_1+X_3$

F1

Idemo u **Linear Model- Multiple Regression**

Dependent variable je sada Y4X1

Independent variables jesu X2 X13 X4

F1

$R^2=0.9997$

$R_R > R_{UR}$, tako da ni ne trebamo računati F jer znamo da će doći negativan!

Idemo na **Probability**

Function ID: FP

X: 0

Df num: 1

Df den: 18

Dobijemo da je p ponovno 1.00, tj. 100% što znači da ne možemo odbaciti hipotezu.

Pod c)

c)a)

$H_0: \beta_1=0.5$

$$t = \frac{0.005 - 0.5}{0.00489} = -101.2$$

Idemo u **Probability**

Function ID: T2

X: 101.2

Df: 18

Dobili smo 0.00000

Odbacujemo hipotezu da je β_1 0.5 jer činimo grešku manju od 5%

c)b)

$H_0: \beta_2 = -0.5$

$$t = \frac{-0.611 - (-0.5)}{0.1628} = \frac{-0.611 + 0.5}{0.1628} = -0.68$$

Idemo u **Probability**

Function ID: T2

X: 0.68

Df: 18

Dobili smo 0.50516

Ne možemo odbaciti tu hipotezu jer činimo grešku 50% a dozvoljena je greška od 5%

Zadatak 3:

Procijenite Pfilipsovu krivulju za Veliku Britaniju

Godina	% povećanja plaća	% nezaposlenosti
1950	1,8	1,4
1951	8,5	1,1
1952	8,4	1,5
1953	4,5	1,5
1954	4,3	1,2
1955	6,9	1,0
1956	8,0	1,1
1957	5,0	1,3
1958	3,6	1,8
1959	2,6	1,9
1960	2,6	1,5
1961	4,2	1,4
1962	3,6	1,8
1963	3,7	2,1
1964	4,8	1,5
1965	4,3	1,3
1966	4,6	1,4

a) Postavite pravilan recipročan model i izračunajte značajnost parametara regresije

b) Ispod kojeg postotka plaće u Velikoj Britaniji ne mogu pasti?

c) Izračunajte koeficijent elastičnosti za nezaposlenost 1.5%

Varijabla W- % povećanje plaća

Varijabla U- % nezaposlenosti

Idemo u **Data management- Data entry**

List the variable names W U

Upisati vrijednosti

ESC

Recipročan model nam je

$$Y = \beta_0 + \beta_1 \frac{1}{x}, \text{ tj.}$$

$$W = \beta_0 + \beta_1 \frac{1}{U}$$

Idemo u **Transformation**

$$\text{INVU} = 1/U$$

F1

Idemo u **Linear Models- Multiple Regression**

Dependent variable W

Independent variable INVU

Enter

a)

$$R^2 = 0.3849$$

$$W = -1.4282 + 8.7243 \text{INVU}$$

(2.675) (2.8478)

b)

Ispod kojeg postotka plaće u Velikoj Britaniji ne mogu pasti nam govori β_0 .
Plaće u Velikoj Britaniji ne mogu pasti ispod 1.43

c)

Koeficijent elastičnost za nezaposlenost je $E_{W,U} = \frac{-\beta_1}{W \cdot U}$

$$U = 1.5\%$$

Trebamo izračunati W pomoću prvotne formule.

$$W = -1.428 + 8.72 \cdot \frac{1}{1.5} = 4.39$$

$$E_{W,U} = \frac{-8.72}{4.39 \cdot 1.5} = -1.324 \quad \text{Nezaposlenost od 1.5\% utječe na smanjenje plaća za 1.324.}$$

Zadatak 4:

Procijenite Cobb-Douglasovu funkciju proizvodnje $Y_i = \beta_0 x_{1i}^{\beta_1} x_{2i}^{\beta_2}$ za Tajvansko gospodarstvo u razdoblju 1958.-1972., ako su podaci sljedeći:

Godina	GDP	Radni dani	Realni kapital
	U milijunima \$ (Y)	U milijunima (X ₁)	U milijunima \$ (X ₂)
1958	16607,7	275,5	17803,7
1959	17511,3	274,4	18096,8
1960	20171,2	269,7	18271,8
1961	20932,9	267,0	19167,3
1962	20406,0	267,8	19647,6
1963	20831,6	275,0	20803,5
1964	24806,3	283,0	22076,6
1965	26465,8	300,7	23445,2
1966	27403,0	307,5	24939,0
1967	28628,7	303,7	26713,7
1968	29904,5	304,7	19957,8
1969	27508,2	298,6	31585,9
1970	29035,5	295,5	33474,5
1971	29281,5	299,0	34821,8
1972	31535,8	288,1	41794,3

- a) Prikažite model u uobičajenoj formi
- b) Izračunajte parcijalne koeficijente elastičnosti proizvodnje na promjenu rada i na promjenu kapitala
- c) Testirajte hipotezu da za Tajvansko gospodarstvo vrijede konstantni prinosi na opseg
- d) Testirajte hipotezu da su $\beta_1 - \beta_2 = 0$; te hipotezu da su $\beta_1 + \beta_2 = 2$

$$Y_i = \beta_0 X_{1i}^{\beta_1} X_{2i}^{\beta_2}$$

y-GDP
X1-radni dani
X2-realni kapital

Formulu pretvorimo u ovaj izraz:

$$\ln y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2$$

Idemo u **Data management-Data entry** i upišemo vrijednosti za Y X1 X2

ESC

Idemo u **Transformation**

$\ln Y = \ln(y)$ **F1** i obrišemo transformaciju
 $\ln X_1 = \ln(x_1)$ **F1** i obrišemo transformaciju
 $\ln X_2 = \ln(X_2)$ pritisnemo **F1** da umemorira transformaciju

ESC

Idemo u **Linear Models- Multiple Regression**

Dependent variable $\ln Y$
Independent variables $\ln x_1 \ln x_2$

a)

$$\ln Y = -3.385 + 1.4988 \ln x_1 + 0.48986 \ln x_2$$

(2.4495) (0.5398) (0.102)

$$R^2 = 0.889$$

b)

Parcijalni koeficijenti elastičnosti jesu β_1 i β_2 .

$$E_{y,x_1} = \beta_1 = 1.4988 \quad E_{y,x_2} = \beta_2 = 0.48986$$

c) $\beta_1 + \beta_2 = 1$

$$\beta_2 = 1 - \beta_1$$

$$\ln y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2$$

$$\ln y = \beta_0 + \beta_1 \ln X_1 + (1 - \beta_1) \ln X_2$$

$$\ln y = \beta_0 + \beta_1 \ln X_1 + \ln X_2 - \beta_1 \ln X_2$$

$$\ln y - \ln X_2 = \beta_0 + \beta_1 (\ln X_1 - \ln X_2)$$

Idemo u **Transformation** i upisujemo dvije transformacije

$$\ln y x_2 = \ln y - \ln X_2 \quad \text{F1} \quad \ln X_1 X_2 = \ln X_1 - \ln X_2 \quad \text{F1}$$

ESC

Idemo u **Linear Models- Multiple Regression**

Dependent variable $\ln y X_2$
Independent variable $\ln x_1 x_2$

$$\text{Novi } R^2 = 0.5696$$

Radimo F test

$$F = \frac{\frac{0.889 - 0.5696}{1}}{\frac{1 - 0.889}{15 - 3}} = \frac{0.3194}{\frac{0.111}{12}} = \frac{0.3194}{0.00925} = 34.53$$

Idemo u **Probability Distribution**

Dobijemo 0.00003

Hipotezu možemo odbaciti jer radimo pogrešku manju i od 1% i od 5%

d)

$$\beta_1 - \beta_2 = 0$$

$$\beta_1 = \beta_2$$

$$\ln y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2$$

$$\ln y = \beta_0 + \beta_1 \ln X_1 + \beta_1 \ln X_2$$

$$\ln y = \beta_0 + \beta_1 (\ln X_1 + \ln X_2)$$

Idemo u **Transformation** i radimo jednu transformaciju : $\ln x_1 x_2 = \ln x_1 + \ln x_2$ **F1**

ESC

Idemo u **Linear Models- Multiple Regression**

Dependent Variable $\ln y$

Independent Variable $\ln x_1 x_2$

$$R^2 = 0.8642$$

$$F = 2.68$$

Idemo u **Probability Distribution** i dobijemo 0.12242

Ako je razina značajnosti 1% ne možemo odbaciti hipotezu jer radimo pogrešku, ako je odbacimo, veću od 1%. A ako je razina značajnosti 5% onda odbacujemo hipotezu.

$$\beta_1 + \beta_2 = 2$$

$$\beta_2 = 2 - \beta_1$$

$$\ln y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2$$

$$\ln y = \beta_0 + \beta_1 \ln X_1 + (2 - \beta_1) \ln X_2$$

$$\ln y = \beta_0 + \beta_1 \ln X_1 + 2 \ln X_2 - \beta_1 \ln X_2$$

$$\ln y - 2 \ln X_2 = \beta_0 + \beta_1 (\ln X_1 - \ln X_2)$$

Idemo u **Transformation** i radimo jednu transformaciju : $\ln y_2 = \ln y - 2 \ln x_2$

Drugu transformaciju imamo umemoriranu od prije!

Idemo u **Linear Models-Multiple Regression**

Dependent Variable $\ln y_2$

Independent variable $\ln x_1 x_2$

$$R^2 = 0.9649$$

$$R^2_R > R^2_{UR}$$

Ne trebamo računati F pošto znamo da će doći negativan, a u Probability Distribution ćemo, nakon što ćemo pod x upisati 0, dobiti 1.00 odnosno 100% što znači da hipotezu nikako ne možemo odbaciti!

Zadatak 5:

Izračunajte funkciju potražnje za narančama u Hrvatskoj ako imate na raspolaganju sljedeće podatke:

Vrijeme (kvartali)		Prodaja naranča	Cijena naranča	Cijena mandarina	Dohodak po stanovniku
1999	1	1000,8	22,8	25,0	8076,9
	2	939,6	24,1	27,5	8461,5
	3	981,6	19,5	29,0	9423,1
	4	1000,8	21,5	31,0	10461,5
2000	1	997,2	23,4	32,0	11153,8
	2	873,6	26,0	35,0	11115,4
	3	916,8	27,3	34,1	11538,5
	4	891,6	30,6	37,0	12307,7
2001	1	823,2	33,8	39,0	11923,1
	2	975,6	34,5	35,0	13461,5
	3	990,0	37,1	37,0	13846,2
	4	867,6	38,4	38,1	11923,1
2002	1	895,2	39,0	41,0	13461,5
	2	918,0	40,3	43,0	14615,4

- 1) Postavite ekonometrijski model pogodan za izračunavanje koeficijenata elastičnosti. Komentirajte dobivene rezultate.
- 2) Testirajte model na mogućnost jedinične elastičnosti prodaje naranči na promjenu cijene naranči i na promjenu dohotka po stanovniku. Je li mandarina supstitutivno ili komplementarno dobro naranči?
- 3) Testirajte model na zajedničku hipotezu da cijena naranči i mandarina ne utječu na prodaju naranči
- 4) Testirajte zajedničku hipotezu da je zbroj elastičnosti cijene naranči i dohotka na prodaju naranči jednaka nuli, te nakon toga da je razlika elastičnosti cijene naranči i cijene mandarina jednaka nuli.
- 5) Izračunajte godišnju stopu rasta dohotka po stanovniku.

Y- prodaja naranči
X1- cijena naranči
X2- cijena mandarina
X3- dohodak po stanovniku

Data management- Data entry- unijeti vrijednosti

ESC

1)
Čim piše da je riječ o modelu pogodnom za izračunavanje koeficijenata elastičnosti, treba nam log-log.

$\ln y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3$ -to trebamo dobiti

Idemo u Transformation i upisujemo 4 transformacije

$\ln y = \ln(y)$
 $\ln x_1 = \ln(x_1)$
 $\ln x_2 = \ln(x_2)$
 $\ln x_3 = \ln(x_3)$

Idemo u Linear Models- Multiple Regression

Dependent variable $\ln Y$
Independent variables $\ln x_1 \ln x_2 \ln x_3$

$\ln y = 3.9713 - 0.037963 \ln x_1 - 0.90513 \ln x_2 + 0.66274 \ln x_3$
(0.71735) (0.063719) (0.15232) (0.12101)
(5.54) (0.0637) (-5.94) (5.48)

$R^2 = 0.8434$

2)
 $H_0: \beta_1 = -1$

$$t = \frac{\beta_1 - \beta^*}{se(\beta_1)} = \frac{-0.038 + 1}{0.064} = 15.03$$

Idemo u Probability

Upisujemo T2 (enter)
df num: 14-4=10

Dobijemo 0.00000

Hipotezu možemo odbaciti jer radimo malu pogrešku.

$H_0: \beta_3 = 1$

$$t = \frac{\beta_1 - \beta^*}{se(\beta_1)} = \frac{0.663 - 1}{0.121} = -2.78$$

Idemo u **Probability**

Radimo T2

Dobijemo 0.01945 što znači da na 95% pouzdanosti možemo odbiti hipotezu.

Mandarina je komplementarno dobro naranči(radi -)

3)

$$H_0: \beta_1 = \beta_2 = 0 \quad (m=2)$$

$$\ln y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3$$

$$\ln y = \beta_0 + \beta_3 \ln X_3$$

Idemo u **Linear Models- Multiple regression**

Dependent variable $\ln y$

Independent variable $\ln X_3$

$$\ln y = 7.8738 - 0.11097 \ln X_3$$

(0.90402) (0.096768)
(8.71) (-1.15)

$$R^2 = 0.0988$$

$$F = \frac{\frac{0.8434 - 0.0988}{2}}{\frac{1 - 0.8434}{14 - 4}} = 23.774$$

Idemo u **Probability**, ali je sad FP (enter)

Dobijemo 0.00016 što znači da hipotezu možemo odbaciti.

4)

$$a) H_0: \beta_1 + \beta_3 = 0 \quad (m=1)$$

$$\beta_1 = -\beta_3$$

$$\beta_3 = -\beta_1$$

$$\ln y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3$$

$$\ln y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 - \beta_1 \ln X_3$$

$$\ln y = \beta_0 + \beta_1 (\ln X_1 - \ln X_3) + \beta_2 \ln X_2$$

Idemo u **Transformation**

$$\ln x_1 x_3 = \ln x_1 + \ln x_3$$

Idemo u **Linear Models- Multiple Regression**

Dependent variable $\ln y$

Independent variables $\ln x_1 x_3 \ln x_2$

Dobijemo $R^2 = 0.4654$

$F = 24.16$

Idemo u **Probability** i dobijemo 0.00061 što znači da hipotezu možemo odbaciti

b)

$$H_0: \beta_1 - \beta_2 = 0 \quad (m=1) \\ \beta_2 = \beta_1$$

$$\ln y = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \beta_3 \ln x_3$$

$$\ln y = \beta_0 + \beta_1 \ln x_1 + \beta_1 \ln x_2 + \beta_3 \ln x_3$$

$$\ln y = \beta_0 + \beta_1 (\ln x_1 + \ln x_2) + \beta_3 \ln x_3$$

Idemo u **Transformation**

$$\ln x_1 x_2 = \ln x_1 + \ln x_2$$

Idemo u **Linear Models- Multiple Regression**

Dependent variable je $\ln y$

Independent variables $\ln x_1 x_2 \ln x_3$

Dobijemo $R^2 = 0.5131$

$F = 21.09$

Idemo u **Probability** i dobijemo 0.00099 što znači da možemo odbaciti hipotezu.

5)

Kad god imamo stopu rasta (bilo čega) kod Transformation upisujemo T=selcase!!!

Nakon što smo to napravili idemo u **Linear Models- Multiple Regression** i zavisna varijabla postaje ono što se traži!!!

$$\ln y = 9.0419 - 0.039808 \ln x_3$$

U zadatku imamo kvartale pa je po svemu kvartalni rast 3.98%

To pomnožimo s 4 da bi dobili godišnji rast, pa dobijemo 15.92% što je godišnji rast plaća.

Zadatak 6:

Poljoprivredni fakultet u Osijeku testira novo gnojivo koje bi trebalo povećati prinos krumpira po četvornome metru. Testiranja na probnim poljima dala su sljedeće rezultate:

Prinos u kg po četvornome metru	Novo gnojivo	Polje
12	Da	A
11	Ne	A
13	Da	A
13	Da	A
10	Ne	A
10	Ne	A
12	Da	A
13	Da	A
12	Da	A
13	Ne	B
17	Da	B
17	Ne	B
13	Ne	B
18	Da	B
13	Ne	B
14	Ne	C
15	Ne	C
18	Ne	C
21	Da	C
23	Da	C
35	Da	C

- Postavite lin-lin model prinosa krumpira koji ovisi o gnojivu i testirnim poljima tako da polje C bude obuhvaćeno konstantnim članom β_0
- Koliki je doprinos novog gnojiva prinosu krumpira i je li on značajan na razini pouzdanosti od 95%? Testirajte mogućnost (95%) da je doprinos novog gnojiva 10 kg po četvornome metru.
- Koliki je u prosjeku prinos krumpira po četvornome metru:
 - na polju A kada se koristi novo gnojivo a koliki kada se koristi staro?
 - na polju B kada se koristi novo gnojivo a koliki kada se koristi staro?
 - na polju C kada se koristi novo gnojivo a koliki kada se koristi staro?
- Testirajte zajedničku hipotezu da testirna polja nemaju utjecaja na prinos krumpira.
- Je li ekonomski isplativo koristiti novo gnojivo ako je cijena krumpira po kg 50 lipa, a novo gnojivo skuplje od starog 2.50 kn po četvornome metru polja?

a)

Y- prinos u kg
NG- novo gnojivo

D=0 staro gnojivo

D=1 novo gnojivo

Kod polja izbacujemo jedno jer kad bi imali tri dummy, došlo bi do kolinearnosti.
Polje C ne uključujemo u dummy varijablu

Polja P_A P_B

Unosimo na sljedeći način

Y	NG	P _A	P _B
12	1	1	0
11	0	1	0

itd.

Kod lin-lin modela nema transformacije. Riječ je o apsolutnim promjenama, tj. koliko apsolutnog nečeg utječe na nešto drugo.

Y=	18.414	+5.2NG	-10.1P _A	-5P _B
se	(1.62)	(1.58)	(1.85)	(2.02)
t	(11.34)	(3.27)	(-5.45)	(-2.46)

$R^2=0.68$

Novo gnojivo je povećalo prinos za 5.2 kg po m² na svim poljima

Ako se sadi krumpir imamo 10.1 kg manje od polja C

68% varijance smo uspjeli objasniti modelom

Kod testiranja hipoteza gledamo p!!

H₀: B₁=0 Kod NG imamo t=3.27, a p=0.0045, što znači da ako odbacimo hipotezu radimo 0,4% greške.

b) Doprinos novog gnojiva je u prosjeku 5.2 kg veći.

Pitanje je je li to značajno.

t=3.27

p=0.0045

Možemo odbaciti hipotezu da je B₁=0.

H₀= B₁=10

$$t = \frac{5.2 - 10}{1.58} = -3.04 > 2$$

Function ID: T2

$X=-3.04$

$Df=17$

Dobijemo $p=0.007$ što znači da možemo odbaciti hipotezu jer radimo grešku od 0.7%

c)

	Novo gnojivo	Staro gnojivo
P_A	$Y=18.4+5.2-10.1=13.5 \text{ kg/m}^2$	$Y=18.4+0-10.1=8.3 \text{ kg/m}^2$
P_B	18.6 kg/m^2	13.4 kg/m^2
P_C	23.6 kg/m^2	18.4 kg/m^2

d)

$R^2_{UR}=0.68$

Treba nam R^2_R

Napravimo model $Y=\beta_0+\beta_1NG$

Dobijemo da je $R^2_R=0.1169$

Radimo F test

$F=29.95$

Ako odbacimo hipotezu da su β_2 i β_3 0, radimo 6.4% greške što znači da hipotezu da testirna polja nemaju utjecaja na prinos krumpira ne možemo odbaciti.

e)

0.50 kn/kg cijena krumpira

2.5 kn više za novo gnojivo po m^2

Ako koristimo novo gnojivo $5.2 \text{ kg} \times 0.50 = 2.6 \text{ kn/m}^2$

Ako koristim novo gnojivo zaraditi ću 2.6. kn ali ću potrošiti 2.5 kn što znači da imamo zaradu od 0.10 kn.

Ako konstruiramo obrnuto tj. da je $D=1$ za staro gnojivo, a $D=0$ za novo gnojivo

Na tabelu gdje su nam svi podaci prikazani

F3 Idemo na INSERT, umjesto CASES upišemo VAR i upišemo NG1 i upisujemo obrnuto

Očekujemo istu reakciju

Promijeniti će se β_0

$$\beta_1 = -5.2$$

$$Y = 23.586 - 5.2KG - 10.1P_A - 5P_B$$

(1.62) (1.58) (1.85) (2.02)
(14.53) (-3.27) (-5.45) (-2.46)

Ako koristimo staro gnojivo, u prosjeku ćemo imati 5.2 kg/m² manje.

Ako bi zadatak tražio da napravimo model log-log

U transformacijama bi upisali sljedeće:

$$\ln y = \ln(y)$$

Nakon toga idemo u multiplu regresiju i dobijemo sljedeći model

$$\ln y = 2.85 + 0.297 \ln NG - 0.58 P_A - 0.24 P_B$$

$$R^2 = 0.81$$

Ako koristimo novo gnojivo prinos je za 29 % veći

Ne možemo apsolutno znati

58% manje u odnosu na polje C

24% manje u odnosu na polje C

ZADATAK 7:

Koristeći dummy varijable provjerite je li odnos štednje i dohotka bitno poremećen (različite regresije) nakon 1955. godine u Velikoj Britaniji. D=1 za razdoblje 1946.-1954.; D=0 za razdoblje 1955.-1963.

Godina	Štednja	Dohodak
1946	0,36	8,8
1947	0,21	9,4
1948	0,08	10,0
1949	0,20	10,6
1950	0,10	11,0
1951	0,12	11,9
1952	0,41	12,7
1953	0,50	13,5
1954	0,43	14,3
1955	0,59	15,5
1956	0,90	16,7
1957	0,95	17,7
1958	0,82	18,6

1959	1,04	19,7
1960	1,53	21,1
1961	1,94	22,8
1962	1,75	23,9
1963	1,99	25,2

Unesemo podatke na sljedeći način:

G	S	Y
1	0.36	8.8

Imamo malo opažanja za dvije regresije, tj. da bi radili lom.

$S = \beta_0 + \beta_1 Y$ do 55. godine- od te godine radimo uz još dvije varijable

$S = \beta_0 + \beta_1 Y + \beta_2 D + \beta_3 D \times Y$

$\beta_2 D$ - to je odsječak

$\beta_3 D \times Y$ -to je nagib

$D=0 \quad S = \beta_0 + \beta_1 Y$

$D=1 \quad S = \beta_0 + \beta_1 Y + \beta_2 + \beta_3 Y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) Y$

Ako su β_2 i β_3 0, nema razlike između razdoblja kad je D 1 ili 0.

Moramo napraviti DY , tj. idemo u Transformation

$GY = G \times Y$

F1

Model izgleda sljedeće

$S = -1.75 + 0.15Y + 1.48D - 0.1YD$
 (-5.77) (8.24) (3.15) (-3.11)

$R^2 = 0.95$

$S = -1.75 + 1.48 + (0.15 - 0.1)Y$

$S = -0.27 - 0.05Y$

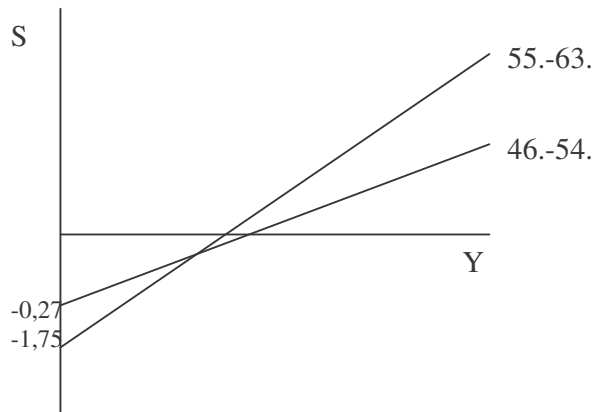
Granična sklonost štednji je 0.05 tj. na 1 funtu stavljaju sa strane 5 penija u razdoblju do 55. godine.

A za razdoblje od 55.-63. godine model izgleda sljedeće

$S = -1.75 + 0.15Y$

Granična sklonost štednji je 15 penija.

Je li razlika značajna ovisi o β_3 kojemu je $t=-3.11$, a $p=0.007$
 Možemo odbaciti hipotezu da je $\beta_3 = 0$.



Radimo sljedeću hipotezu $H_0: \beta_2 = \beta_3 = 0$
 Ako možemo odbaciti onda postoji značajna razlika između 2 pravca
 Radimo F test kojemu rješenje dođe 6.66, a $p=0.14$
 Ne možemo odbaciti hipotezu jer radimo veliku pogrešku.

TRIK: Stupnjeve slobode vidimo u Anova tablici pod RESIDUAL I df!!!!

Kad imamo samo jednu dummy varijablu (sociolozi, psiholozi), to se zove AOV model (analiza varijance)

$$Y = \beta_0 + \beta_1 D$$

ZADATAK 8:

Ocjenjivanje kod Bellula

	Ocjena	Spol	Mini suknja
Marko	3	0	0
Mira	4	1	0
Laura	5	1	1
Ive	2	0	0
Toni	4	0	0
Martina	4	1	0
Sonja	5	1	1
Vana	3	1	0
Maja	4	1	0
Darko	4	0	0

$$0 = \beta_1 + \beta_2 S + \beta_3 \text{MINI}$$

$$0 = 3.25 + 0.5S + 1.25\text{MINI}$$

Da je samo muškarac došao na ispit dobio bi ocjenu 3.25.

Da je došla ženska na ispit, dobila bi ocjenu 3.25+0.5

A da je došla i u Mini suknji dobila bi 3.25+0.5+1.25

$H_0: \beta_2 = 0$ p vrijednost je 0.35---ne možemo odbaciti hipotezu

$H_0: \beta_3 = 0$ p vrijednost je 0.08---na 90% testiramo pa možemo odbaciti hipotezu

Zaključak: Nije bitan spol koliko mini suknja

Zadatak 9:

Odlučili ste špekulirati na razlici u cijeni kupućih kostima tijekom godine. Dostupni podaci o cijeni kupućih kostima su:

Kvartal	Cijena u kn
1	24,26
2	44,74
3	36,05
4	28,59
1	23,26
2	44,79
3	37,59
4	29,55
1	26,72
2	44,87
3	34,37
4	29,57
1	24,78
2	45,46
3	35,49

a) Kolika je prosječna cijena kupućeg kostima u svakom kvartalu?

b) Kolika je maksimalna moguća potencijalna špekulacijska zarada po jednom kupućem kostimu?

c) U kojem bi kvartalu kupili a kada prodali kostime?

d) Kolike su relativne sezonske promjene?

e) Jesu li sezonske oscilacije značajne na razini pouzdanosti od 95%?

$$Y = \beta_0 + \beta_1 S_2 + \beta_2 S_3 + \beta_3 S_4$$

	S_2	S_3	S_4
I	0	0	0
II	1	0	0
III	0	1	0
IV	0	0	1
I	0	0	0
II	1	0	0

4 kvartala-3 dummy

12 mjeseci-11 dummy

F3
INSERT

VAR
S2 S3 S3

sve ispuniti, po dijagonali upisivati 0

a) linear regression

Y, S2 S3 S4
dobijemo model

$$Y = 24.755 + 20.21S_2 + 11.12S_3 + 4.481S_4$$

(0.537) (0.759) (0.759) (0.820)
(46.1) (26.61) (14.64) (5.46) $R^2 = 0.9862$

I 24.8 min
II 45 max
III 36.9
IV 29.3

b,c) kupiti u I. i prodati u II.

max. špekulacijska zarada 20,2 kn po kupačem kostimu

značajne t vrijednosti!! VELIKE-uvjet

d) log-log model! samo na lijevoj strani

Transformation $\ln Y = \ln(y)$

$$\ln y = 3.2078 + 0.5981s_2 + 0.372s_3 + 0.168s_4$$

(172.69) (22.77) (14.15) (5.91) $R^2 = 0.9811$

U II. kvartalu je veća prodaja za cca. 60%
U III. kvartalu je veća prodaja za cca. 40%
U IV. kvartalu je veća prodaja za cca 20%

I antilog $3.21 = 24.8 \quad e^x$

E) Gledali F(imamo samo β_0) i njegov p
Odbacujemo hipotezu da su $\beta_1, \beta_2, \beta_3 = 0$
Sezonske oscilacije su značajne

Zadatak 10:

Trgovački putnici dali su sljedeće podatke o prodaji i zaradi:

Prodaja	Zarada
45	1690
67	1768
34	1640
115	2400
236	3400
200	3100
45	1650

3	1500
88	1850
14	1550
180	2900
99	1900
44	1670
33	1640

1) Postavite ekonometrijski model pogodan za testiranje je li se zarada bitno razlikuje kada se prodaje do 100 komada i preko 100 komada

2) Koliko u prosjeku po komadu zarade trgovački putnici kada prodaju ispod 100 a koliko kada prodaju iznad 100 komada?

3) Je li ta razlika značajna na 95 % pouzdanosti?

Y-prodaja

X-zarada

D=0 do 100

D=1 preko 100

Y X D DX uvrstiti brojeve

$$D=0 \quad Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 DX \\ Y = \beta_0 + \beta_1 X$$

$$Y = -143.65 + 0.112X \quad R^2 = 0.9623$$

$$D=1 \quad Y = \beta_0 + \beta_1 X + \beta_2 + \beta_3 X \\ Y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X$$

$$Y = (-361.35 + 188.15) + (0.2423 - 0.1216)X = -173.2 + 0.1207X \quad R^2 = 0.9988$$

$$Y = 1493 + 4.1X - 54.4D + 4.2DX \quad \text{Da ništa ne proda 1493 kn + 4 kn po prodanom komadu}$$

Značajnu razliku vidimo u t vrijednosti.

$\beta_2 = 0$ ne možemo odbaciti hipotezu, nije značajna

Do 100 komada po komadu 4.1

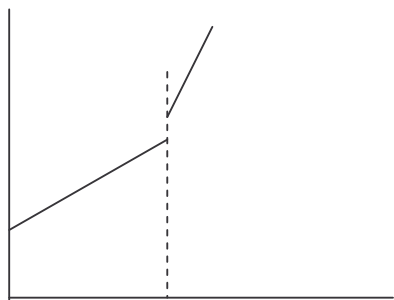
Iznad 100 kom. po kom. 8.3

Razlika u β_3 -značajan jer je $t = 7.97$

Postoji značajna razlika

Manje od 100 i više od 100

Nagib je duplo veći



df num:2
df den:10

$H:\beta_2=\beta_3=0$ RAZLIKA SE VIDI U NAGIBU

Zadatak 11. Autokorelacija

Procijenite funkciju potražnje za gliserima ako su vam poznati sljedeći podaci:

Vrijeme (kvartali)	Broj prodanih glisera	Cijena glisera	Cijena jedrilica	Dohodak po stanovniku
1999:1	79725	32	38	2100
2	88521	33,5	40	2200
3	93535	36	39	2450
4	113169	35	40	2720
2000:1	127596	36	43	2900
2	121190	38	44	2890
3	126748	38,5	47	3000
4	153355	39	49	3200
2001:1	131702	41	48	3100
2	146402	42	47	3500
3	149044	44	47,5	3600
4	142366	43	46	3100
2002:1	129357	47	47	3500
2	154928	48	46	3800
3	136047	50	47	3700

- 1) Postavite ekonometrijski model pogodan za izračunavanje koeficijenata elastičnosti.
- 2) Testirajte model na postojanje autokorelacije reziduala prvog reda na 95 % pouzdanosti.
- 3) Ako ne možete odbaciti hipotezu o postojanju autokorelacije, konstruirajte model kojim ćete se riješiti problema.
- 4) Testirajte zasebno hipoteze jediničnih cjenovnih elastičnosti (glisera i jedrilica) i jedinične dohodovne elastičnosti.
- 5) Testirajte zajedničku hipotezu da je zbroj cjenovnih elastičnosti (glisera+jedrilica) jednaka -0,5.

1)

$BPG = \beta_0 + \beta_1 PG + \beta_2 PJ + \beta_3 Y$ -za elastičnost

$\ln BPG = \beta_0 + \beta_1 \ln PG + \beta_2 \ln PJ + \beta_3 \ln Y$ -transformacije

$\ln BPG = \ln(BPG)$

$$\ln PG = \ln(PG)$$

$$\ln PJ = \ln(PJ)$$

$$\ln Y = \ln(Y)$$

zavisna $\ln BPG$

nezavisna $\ln PG, \ln PJ, \ln Y$

$$\ln BPG = 1.44 - 0.81 \ln PG + 0.78 \ln PJ + 1.28 \ln Y$$

$$\begin{array}{cccc} (1.76) & (-2.86) & (2.27) & (4.82) \\ & (0.28) & (0.34) & (0.26) \end{array}$$

$$R^2 = 0.95$$

$$F = 63.37$$

2)

DW = ?

Escape-Residuals-Pod Residuals upišemo E

Data Management-Edit Data-Vidimo da li se pojavilo E

DW = 2.96 → to vidimo u Durbin-Watson Statistics

Iz knjige:

$$dL = 0.81$$

$$dU = 1.75$$

Nema zaključka.

3)

Treba nam ρ

Residuals

Residual E

Transformation $E1 = \text{LAG}(E, 1)$

Radimo regresiju

zavisna E

nezavisna E1

Do you want the constant (intercept) fitted?..... no

$$\rho = -0.54$$

$$(-2.31)$$

Transformation:

$$Y_{KO} = \ln BPG + 0.54 \times \text{LAG}(\ln BPG, 1)$$

ili BPGKO

$$Y_{1KO} = \ln PG + 0.54 \times \text{LAG}(\ln PG, 1)$$

ili PGKO

$$Y_{2KO} = \ln PJ + 0.54 \times \text{LAG}(\ln PJ, 1)$$

ili PJKO

$$Y_3KO = \ln Y + 0.54 \times \text{LAG}(\ln Y, 1) \quad \text{ili} \quad YKO$$

Multiple Regression

zavisna YKO

nezavisne Y_1KO Y_2KO Y_3KO

Include.... YES

$$YKO = 2.52 - 0.72Y_1KO + 0.75Y_2KO + 1.24Y_3KO$$

(-3.72) (3.39) (6.47)

-sad moramo vidjeti imamo li autokorelaciju

-Data Management-Data-F3-DELETE CASES-ENTER

DW=2.32

Nema zaključka.

Zadatak 12: HETEROSKEDASTIČNOST

Potrošnja goriva, broj registriranih automobila, porez na gorivo i populacija u SAD-u 1982. godine su bile:

DRŽAVA	PCON	TAX	REG	POP
Maine	270	9	743	1136
New Hampshire	122	14	774	948
Vermont	58	11	351	520
Massachusetts	821	9,9	3750	5750
Rhode Island	98	13	586	953
Connecticut	450	11	2258	3126
New York	1819	8	8235	17567
New Yersey	1229	8	4917	7427
Pennsylvania	1200	11	6725	11879
Ohio	1205	11,7	7636	10772
Indiana	650	11,1	3884	5482
Illinois	1198	7,5	7242	11466
Michigan	760	13	6250	9116
Wisconsin	460	13	3162	4745
Minnesota	503	13	3278	4133
Iowa	371	13	2346	2906
Missouru	571	7	3412	4942
North Dakota	136	8	653	672
South Dakota	109	13	615	694
Nebraska	203	13,9	1215	1589
Kansas	349	8	2016	2408
Delaware	118	11	415	600
Maryland	487	13,5	2893	4270
Virginia	628	11	3705	5485

West Virginia	192	10,5	1142	1961
North Carolina	642	12	4583	6019
South Carolina	320	13	1975	3277
Georgia	677	7,5	3916	5648
Florida	1459	8	8335	10466
Kentucky	434	10	2615	3692
Tennessee	482	9	3381	4656
Alabama	457	11	3039	3941
Mississippi	325	9	1593	2569
Arkansas	300	9,5	1481	2307
Louisiana	1417	8	2800	4383
Oklahoma	451	6,58	2780	3226
Texas	3572	5	11388	15329
Montana	131	9	758	805
Idaho	105	7,5	873	977
Wyoming	163	8	508	509
Colorado	323	9	2502	3071
New Mexico	192	11	1193	1367
Arizona	291	10	2216	2892
Utah	169	11	1038	1571
Nevada	133	12	710	876
Washington	562	12	3237	4276
Oregon	364	8	2075	2668
California	2840	9	17130	24697
Alaska	155	8	319	444
Hawaii	214	8,5	586	997

- 1) Prikažite rezultate regresije gdje je potrošnja goriva u državama SAD-a zavisna o broju registriranih vozila i poreznoj stopi na gorivo
- 2) Testirajte model na 99 % pouzdanosti na mogućnost heteroskedastičnosti pomoću Park i Goldfeld-Quandt
- 3) Prikažite, u slučaju postojanja heteroskedastičnosti, rezultate vagane metode najmanjih kvadrata
- 4) Provjerite je li se korištenjem varijabli potrošnje goriva per capita, umjesto ukupne veličine, i broja registriranih automobila per capita, umjesto ukupnog broja automobila, moguće riješiti heteroskedastičnosti

1)

$$PCON = f(\overset{+}{REG}, \overset{-}{TAX}) \text{ -regresija}$$

$$PCON = 552,72 + 0,18REG - 53,67TAX$$

(2,97) (15,89) (-3,19)

$$R^2 = 0,87$$

2)

PARK TEST

$$\ln(e_i)^2 = \alpha + \alpha_1 \ln Z$$

REG=Z REG ima najveću t vrijednost!!

ESCAPE-Residuals
Residual E
F1

Transformation

$$E2 = \ln(\text{power}(e, 2))$$
$$\ln \text{REG} = \ln(\text{REG})$$

Regresija

Zavisna E2
Nezavisna $\ln \text{REG}$

$$\ln(e_i)^2 = 1,64 + 0,95 \ln(\text{REG})$$

(3,09)

Možemo odbaciti hipotezu da postoji homoskedastičnost → imamo heteroskedastičnost

GOLDFELD-QUANDT TEST

Data Management-Sort Cases
Ascending, REG, F1

Po redu smo svrstali proporcionalni faktor → REG

Radimo 3 trećine → prvih i zadnjih 15 (dogovor)

Data Management-Omit
OMIT(SELCASE>15) → to znači da smo izbacili sve osim prvih 15

F1

Regresija
Zavisna PCON
Nezavisna REG TAX

Zanima nas Residual pod SS=27398

Omit/Restore
Restore F1
OMIT(SELCASE<36)
RSS₃=1 551 300

$$GQ = \frac{RSS_3}{RSS_1} = 56,62$$

Probability FPROB
 X=56,62
 df num=3
 df den=11
 p=0,0000

„N“=15

→ možemo odbaciti hipotezu o postojanju homoskedastičnosti

3)

WLS

→ sve dijelimo s REG!! (s proporcionalnim faktorom)

Treba sve vratiti kako je bilo → RESTORE!!!

$$\frac{PCON}{REG} = \beta_0 \frac{1}{REG} + \beta_1 + \beta_2 \frac{TAX}{REG} + \frac{E}{REG}$$

Transformation

$$PCONREG = \frac{PCON}{REG}$$

$$INVREG = \frac{1}{REG}$$

$$TAXREG = \frac{TAX}{REG}$$

$$\frac{PCON}{REG} = 218,62 \frac{1}{REG} + 0,17 - 17,41 \frac{TAX}{REG} \quad \mathfrak{R}^2 = 0,36$$

(4,55) (12,28) (-3,72)

4)

$$\frac{PCON}{POP} = \beta_0 + \beta_1 \frac{REG}{POP} + \beta_2 TAX$$

Regresija

$$\frac{PCON}{POP} = 0,17 + 0,11 \frac{REG}{POP} - 0,01 TAX \quad \mathfrak{R}^2 = 0,20$$

(2,56) (1,53) (-2,96)

Opet radimo Park test da vidimo jesmo li se riješili heteroskedastičnosti

E1

Transformacija → E2 i lnREGPOP

Regresija

$$\ln(e_i)^2 = -7,25 + 1,33 \ln \text{REGPOP}$$

(0,63)

Zadatak 13: Multikolinearnost

Student	C	Y _d	S
Ante	200	250	2500
Maja	230	300	3100
Marija	280	350	3300
Tomislav	380	400	3900
Marko	350	450	4800
Ivica	500	500	5400
Ivan	450	550	5500
Martin	300	280	5000
Marica	500	300	6000
Martina	400	390	7000

Treba procijeniti funkciju potrošnje studenata. Teoretski je očekivati da će prosječna mjesečna potrošnja studenata ovisiti o njihovom raspoloživom dohotku i o veličini njihove štednje. Što su štednja i raspoloživi dohodak veći to bi njihova potrošnja morala biti veća, te očekujemo pozitivne predznake. Želimo provjeriti ovisi li potrošnja i o spolu.

- 1) Procijenite funkciju potrošnje studenata
- 2) Testirajte model na mogućnost multikolinearnosti pomoću koeficijenata korelacije i pomoću VIF testa
- 3) Riješite se multikolinearnosti pomoću izbacivanja jedne od varijabli i pomoću linearne kombinacije nezavisnih varijabli
- 4) Ovisi li o spolu razina potrošnje? Ako dam kolika je granična sklonost potrošnji žena, a koliko muškaraca?

1)

$$M=0 \quad \checkmark=1$$

$$C = -7,1398 + 0,382Y_d + 0,0478S - 0,0157SP \quad R^2 = 0.7337$$

(-0,07) (1,35) (2,57) (0)

Štednja je unutar područja značajnosti.

Spol ne bi trebao imati nikakav utjecaj.

R^2 je relativno visok.

Multikolineranost (ako postoji) ne treba biti samo između dvije varijable, već između više njih.

2)

Trebaju nam koeficijenti korelacije

Idemo na GOODNES OF FIT

Pearson

Između kojih varijabli?

$$S \text{ i } Y_d \rightarrow 0,4353$$

$$S \text{ i } SP \rightarrow 0,12$$

$$SP \text{ i } Y_d \rightarrow 0,3631$$

Nemamo multikolinearnosti. Mora biti broj veći od 0,8.
Jedan na jedna nema multikolinearnosti.

VIF TEST \rightarrow treba nam \mathfrak{R}^2

$$VIF = \frac{1}{1 - \mathfrak{R}^2}$$

U regresiji radimo 3 regresije

$$Y_d = \alpha_0 + \alpha_1 S + \alpha_{12} SP$$

$$S = \gamma_0 + \gamma_1 Y_d + \gamma_2 SP$$

$$SP = \delta_0 + \delta_1 Y_d + \delta_2 S$$

$$\text{Za } Y_d \quad \mathfrak{R}^2 = 0,3650 \quad VIF = 1,57 < 5$$

$$\text{Za } S \quad \mathfrak{R}^2 = 0,2793 \quad VIF = 1,39 < 5$$

$$\text{Za } SP \quad \mathfrak{R}^2 = 0,2280 \quad VIF = 1,3 < 5$$

Nema multikolinernosti .

Da multikolinearnost postoji (\mathfrak{R}^2 je veliki, a male su t vrijednosti, pa sumnjamo):

Izbacimo varijablu da bi dobili prostojnu regresiju.

4)

Razina potrošnje ne ovisi o spolu.

Granična sklonost potrošnji muškaraca i žena je jednaka.