VAJA 9 – STATISTIČNO SKLEPANJE O KORELACIJSKI - LINEARNA REGRESIJA

Kovarianca za populaciji *X* in *Y*:

$$C_{XY} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$$

$$C_{XY} = \begin{cases} > 0 & poz.lin.pov. \\ = 0 & ni\ lin.pov. \\ < 0 & neg.lin.pov. \end{cases}$$

Pearsonov koeficient korelacije:

$$\rho_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y} =$$

$$= \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{N}(x_i - \mu_x)^2}\sqrt{\sum_{i=1}^{N}(y_i - \mu_y)^2}} \in [-1,1]$$

$H_0 : \rho = 0$ spremenljivki nista linearno povezani

$H_1 : \rho \neq 0$ spremenljivki sta linearno povezani (dvostranski test)

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$$

$$m = n - 2$$

r- koef. korelacije v vzorcu

ρ - koef. korelacije v populaciji

Model linearne regresije:

$y' = a + bx$ prva regresijska premica

$x' = a' + b'y$ druga regresijska premica

Premici se sekata v točko (M_x, M_y)

Oceni parametrov regresijskega modela (*a* in *b*) izračunamo po metodi najmanjših kvadratov, ki privede do sledečih formul:

$$a = M_y - bM_x$$

$$b = \frac{C_{xy}}{\sigma_x^2}$$

Pearsonov keficient korelacije:

$$r = \frac{C_{xy}}{\sigma_x \sigma_y}$$

Determinacijski koeficient-delež pojasnjene variance v primeru linearne regresije:

$$r^2 = \frac{\sigma_{y'}^2}{\sigma_y^2}$$

Celotna varianca σ_y^2 = pojasnjena varianca $\sigma_{y'}^2$ +nepojasnjena varianca σ_e^2

Standardna napaka linearne regresijske ocene:

$$\sigma_e = \sigma_y \sqrt{1 - r^2}$$

Meri kakovost ocene z regresijsko premico; meri razpršenost točk okoli regresijske premice.

Včasih se podatkom bolje kot premica prilega kakšna krivulja. Tako npr. Excel pozna naslednje: logaritemska, polinomska, potenčna, eksponentna.

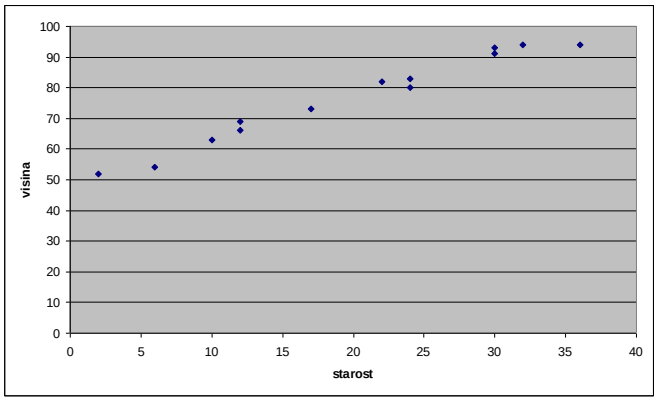
Primer 1

V naslednji tabeli so podatki o starosti (v mesecih) in telesni višini (v cm) za 13 otrok.

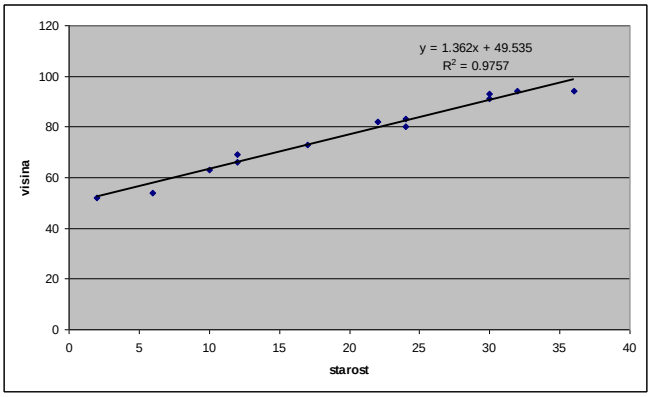
oseba	starost	višina
1	2	52
2	6	54
3	10	63
4	12	66
5	12	69
6	17	73
7	22	82
8	24	83
9	24	80
10	30	91
11	30	93
12	36	94
13	32	94

a. Grafično prikažite odvisnost telesne višine od starosti.
b. Ali je uporaba modela linearne regresije za te podatke upravičena? Obrazložite.
c. Izvedite regresijsko analizo najprej »peš«, nato pa uporabite ukaz *Orodja/Analiza podatkov, Regression*.

a.



Razsevni grafikon



Razsevni grafikon s regresijsko premico

b. in c.

Tabela za regresijsko analizo "peš":

Starost-x	Višina-y	$x-M_x$	$y-M_y$	$(x-M_x)^2$	$(y-M_y)^2$	$(x-M_x)(y-M_y)$

Rezultati uporabe ukaza: *Orodja/Analiza podatkov, Regression*.

SUMMARY OUTPUT

Regression Statistics		***Razlaga izračunov***
Multiple R	0.987786101	*koeficient korelacije, r*
R Square	0.975721382	*koeficient determinacije, r^2*
Adjusted R Square	0.973514235	*koeficient determinacije prilagojen na SP (stopnje prostosti)*
Standard Error	2.417194184	*standardna napaka regresije, s*
Observations	13	*število enot, n*

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2582.959664	2582.9597	442.07356	3.123E-10
Residual	11	64.27110497	5.8428277		
Total	12	2647.230769			
	stopinje prostosti	*vsota kvadriranih odklonov*	*srednji kvadrirani odklon*	*F statistika*	*p-vrednost*

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	
Intercept	49.53497238	1.445523168	34.267851	1.566E-12	46.353397	52.716547	*a*
starost	1.362044199	0.064780445	21.025545	3.123E-10	1.2194634	1.504625	*b*
	ocena koeficienta	*standardna napaka ocene*	*t - statistika*	*p-vrednost*	*spodnja meja intervala zaupanja*	*zgornja meja intervala zaupanja*	

Primer 2

Naslednja tabela prikazuje različne hitrosti vožnje v *km/h* istega avtomobila in pripadajoče porabe goriva v litrih.

Poraba goriva	6,8	7,5	8,0	8,2	8,4	8,5	9,0	9,3	9,8	10,0
Hitrost vožnje	60	65	70	75	80	85	90	95	100	105

a. Ugotovite, kako dobro sta linearno povezani poraba goriva in hitrost vožnje.
b. Zapiši enačbo linearne regresijske premice.
c. Grafično prikaži regresijsko premico v razsevnim grafikonu.

Primer 3

Preverite domnevo, da sta izobrazba (število priznanih let šole) in število ur branja dnevnih časopisov na teden povezana med seboj pri 5% stopnji značilnosti.

Izobrazba - *X*	**Branje -*Y***
10	3
8	4
16	7
8	3
6	1
4	2
8	3
4	1

Primer 4

V datoteki s podatki o letnem prometu Luke ugotovite kako dobro sta linearno povezani nosilnost in balast na ladji, (nosilnost in BRT, nosilnost in NRT).

a. Zapišite enačbo linearne regresijske premice. Nalogo rešujte najprej peš, uporabite prvih 10 parov.
b. Nato pa z Excelovim orodjem za določanje linearne regresije.
c. Grafično prikažite regresijsko premico v razsevnim grafikonu.