

Regresiona i korelaciona analiza (prost linearni model)

Primer 1.

Regresiona j-na $\hat{y}_i = 35 + 0,18 \cdot x_i$, izražava zavisnost iznosa zakupa vozila (u evrima) od pređene kilometraže. Objasniti koeficijente regresionog modela i vezu između posmatranih pojava.

Rešenje:

Kako je $a=35$, to znači da početna vrednost zakupa vozila bez pređenih kilometara iznosi 35 evra..Kako je $b=0,18$, to znači da je toliko i prosečno povećanje iznasa zakupa vozila u odnosu na jedinično povećanje kilometara.

Primer 2.

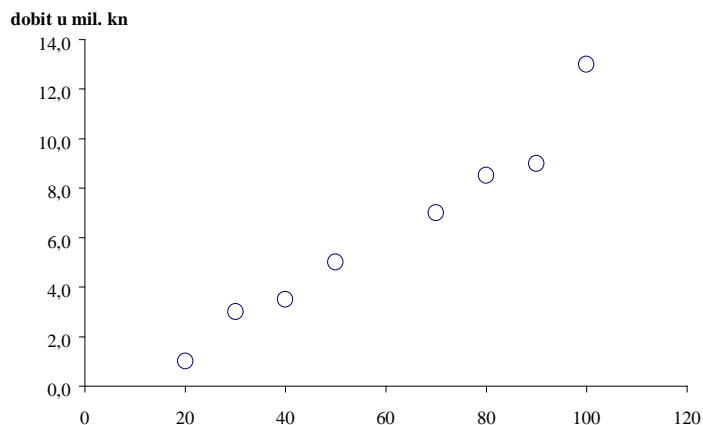
Neko preduzeće analizira podatke o ostvarenom prometu i dobiti (oboje u mil. dinara). U tabeli su prikazani podaci u 8 uzastopnih godina.

Promet (u mil. din.)	20	30	40	50	70	80	90	100
Dobit (u mil. din)	1	3	3,5	5	7	8,5	9	13

- Prikazati uzoračke podatke u dijagramu raspršenosti (rasturanja),
- Izračunaj jednačinu linearne regresije,
- Kolika se dobit može očekivati ako bi promet porastao na 110 mil.din,
- Izračunati standardnu grešku linearne regresije,
- Izračunati koeficijent determinacije u regresionom modelu i objasniti rezultat. (6)

Rešenje:

- Dijagram raspršenosti je



Sa slike dijagrama raspršenosti uočavamo da su tačke raspoređene približno na jednom pravcu, tj. da između posmatranih pojava (prometa i dobiti) postoji pozitivna linearna veza, a to znači da rast jedne promenljive prati rast druge promenljive. Veza je dosta jaka jer su tačke raspoređene veoma blizu zamišljene prave koja prolazi kroz tačku (\bar{x}, \bar{y}) .

b) Postupak računanja parametara linearne regresione jednačine je:

Promet x_i	Dobit y_i	$x_i y_i$	x_i^2	Regresijske vred. \hat{y}_i	y_i^2
20	1	20	400	1,05	1
30	3	90	900	2,35	9
40	3,5	140	1600	3,65	12,25
50	5	250	2500	4,95	25
70	7	490	4900	7,55	49
80	8,5	680	6400	8,85	72,25
90	9	810	8100	10,15	81
100	13	1300	10000	11,45	169
480	50	3780	34800	50,00	418,5

$$\bar{x} = \frac{\sum x_i}{n} = \frac{480}{8} = 60 ; \bar{y} = \frac{\sum y_i}{n} = \frac{50}{8} = 6,25 .$$

Ocenimo sada parametre linearne regresije a i b :

$$b = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} = \frac{3780 - 8 \cdot 60 \cdot 6,25}{34800 - 8 \cdot 60^2} = 0,13$$

$$a = \bar{y} - b \bar{x} = 6,25 - 0,13 \cdot 60 = -1,55$$

Jednačina linearne regresije sa ocenjenim parametrima glasi:

$$\hat{y}_i = a + b x_i = -1,55 + 0,13 x_i .$$

Za dati niz empirijskih podataka nezavisne promenljive X , regresijske vrednosti su izračunate i vide se u tabeli (kolona 5).

$$\hat{y}_1 = a + b x_1 = -1,55 + 0,13 \cdot 20 = 1,05;$$

$$\hat{y}_2 = a + b x_2 = -1,55 + 0,13 \cdot 30 = 2,35;$$

.

$$\hat{y}_8 = a + bx_8 = -1,55 + 0,13 \cdot 20 = 1,05.$$

c) Regresijske vrednosti nam mogu poslužiti za prognoziranje toka pojave.

Tako na osnovu dobijene regresije možemo izračunati očekivanu dobit ako promet poraste na 110 mil.din.:

$$\hat{y}_{(x=110)} = -1,55 + 0,13 \cdot 110 = 12,75 \text{ mil.din.}$$

$$d) S_e = \sqrt{\frac{\sum y_i^2 - a \sum y_i - b \sum x_i y_i}{n-2}} = 0,8756$$

Srednja mera odstupanja dobiti od linije regresije je 0,8756 mil. din.

e) Koeficijent determinacije u regresionom modelu je:

$$r^2 = b^2 \cdot \frac{\sum x_i^2 - n\bar{x}^2}{\sum y_i^2 - n\bar{y}^2} = 0,13^2 \cdot \frac{34800 - 8 \cdot 60^2}{418,5 - 8 \cdot 6,25^2} = 0,9566. \quad (6)$$

Dobijeni rezultat nam govori da je 95,66% ukupnog varijabiliteta dobiti preduzeća objašnjeno (determinisano) varijabilitetom nezavisne promenljive X tj. prometom u posmatranim godinama. Ostatak, 4,34% ukupnog varijabiliteta nije objašnjeno regresionom linijom, tj. pod uticajem je neidentifikovanih faktora.

Pošto je u našem primeru r^2 približno jednak jedinici, regresiona linija izuzetno dobro reprezentuje empirijske podatke.

Na osnovu podataka vrednost koeficijenta korelacije može se izračunati na sledeći način:

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum x_i^2 - n \bar{x}^2)(\sum y_i^2 - n \bar{y}^2)}}, \text{ sledi}$$

$$r = \frac{3780 - 8 \cdot 60 \cdot 6,25}{\sqrt{(34800 - 8 \cdot 60^2)(418,5 - 8 \cdot 6,25^2)}} = 0,9781$$

S druge strane, na osnovu izračunatog koeficijenta determinacije

$r^2 = 0,9566$, kao i znajući da je $b > 0$ može se, takođe, dobiti da je

$r = +\sqrt{0,9566} = 0,9781$ Na osnovu dobijenog rezultata koeficijenta korelacije zaključujemo da između posmatranih pojava (profita i dobiti) postoji jaka direktna korelaciona veza.

Napomena: Ovde se radi o školskom primeru, sa malim brojem parova vrednosti. Statistički zaključci, u praksi, se dobijaju na osnovi mnogo veće serije vrednosti obe promenljive.

Primer 3. (postupno uraditi za domaći)

Dati su uporedni podaci o dužini radnog iskustva i dužini odsustva sa posla. Oceniti jednačinu linearne regresije, ispitati smer i jačinu kvantitativnog slaganja varijacija posmatranih pojava.

Godine radnog isk.	4	10	12	18	23	28	36
Dužina ods. (u danima)	8	11	17	22	26	30	38

Rešenje:

$$\hat{y}_i = a + bx_i = 3,94 + 0,95 \cdot x_i, r = 0,9937, r^2 = 0,9874 = 98,74\%.$$

Primer 4. (postupno uraditi za domaći)

Proizvođač automobila je želeo da ispita kako se cena jednog od njegovih modela menja sa povećanjem starosti automobila. Izabran je uzorak od 8 automobila:

Starost (u god.)	8	3	6	9	2	5	6	3
Cena (u stotinama evra)	18	94	50	21	145	42	36	99

- Na osnovu dijagrama raspršenosti objasniti da li se može uočiti linearna veza između posmatranih pojava?
- Odredi regresionu pravu pri čemu je cena zavisna promenljiva i objasniti dobijene rezultate.
- Ucrtaj regresionu pravu u dijagram raspršenosti.
- Oceniti cenu automobila starosti 7 godina.

Rešenje:

$$b) \hat{y}_i = a + bx_i = 150,38 - 16,62x.$$

d) 3404 evra.

Primer 5. (postupno uraditi za domaći)

Osiguravajuća kompanija se interesuje da li iznos godišnje zarade pojedinaca utiče na iznos životnog osiguranja.

Godišnja zarada (000 evra.)	5	8	10	12	13	19	20
Životno osigu.(stotina evra)	7	10	15	17	18	21	25

e) Da li su ispunjeni uslovi za korišćenje prostog linearnog regresionog modela za predviđanje iznosa životnog osiguranja u zavisnosti od godišnje zarade pojedinca?

f) Odrediti j -nu regresionog modela.

g) Izračunati relativnu meru reprezentativnosti regresionog modela i objasniti rezultat.

Rešenje:

a) Može se zaključiti na osnovu grafika;

b) $\hat{y}_i = a + bx_i = 2,59 + 1,09x_i$;

c) $r^2 = 0,9394 = 93,94\%$.

Poštovani studenti,

Za primer 3, 4, 5 data su približna rešenja i potrudite se da zadatke uradite postupno u vašim sveskama nije potrebno da saljete radove. Cilj je uvežbati postupak dobijanja prostog linearnog regresionog i korelacionog modela, kao i pravilno tumačenje dobijenih rezultata. Možete se služiti primenom Excel-a za dobijanje dijagrama raspršenosti.

S poštovanjem,
Dr Slavica Dabetić