VJEROJATNOST I STATISTIKA

TREĆE PREDAVANJE

REGRESIJA I KORELACIJA

REGRESIJSKA ANALIZA

- Ispitivanje ovisnosti jedne zavisne (regresand) varijable o jednoj (ili više) nezavisnih (regresorskih) varijabli s ciljem da se utvrdi analitički izraz takve povezanosti (model).

- Osnova regresijske analize je **regresijskih model**.

- Regresijski model - algebarski model kojim se analitički izražava *statistički* odnos među pojavama.

- Model: - *deterministički (funkcionalan)*
- - ***statistički (stohastički).***

- *Deterministički model* - za svaku vrijednost nezavisne varijable X jednoznačno je određena vrijednost zavisne varijable Y.

- ***Statistički model*** - vrijednost zavisne varijable Y nije jednoznačno određena za zadanu vrijednost nezavisne varijable X.

- pri čemu je : $$\boldsymbol{Y = f(X) + e}$$

- $f(X)$ funkcionalna (deterministička) komponenta
- ***e*** – stohastička varijabla koja predočuje nesistematske utjecaje na zavisnu varijablu.

- Korelacijska analiza - metode utvrđivanja pokazatelja *jakosti i smjera* statističkih veza među pojavama.

- Regresijski model koji sadrži jednu zavisnu, (Y) i jednu nezavisnu varijablu (X) je **model jednostavne regresije**, a model koji sadrži jednu zavisnu i dvije ili više nezavisnih varijabli je **model višestruke (multiple) regresije.**

- Model jednostavne regresije:

$$\boldsymbol{Y = f(X) + e}$$

- Model višestruke (multiple) regresije:

$$\boldsymbol{Y = f(X_1, X_2, \ldots X_j, \ldots X_k) + e}$$

- ***Deskriptivno statistička analiza*** regresijskog modela: - procjena nepoznatih parametara i utvrđivanje drugih statističko-analitičkih veličina. Dobiveni rezultati služe isključivo za opis danih podataka i ne generaliziraju se.
-
- ***Inferencijalno statistička analiza***: - procjena parametara i testiranje hipoteze o parametrima.
- ***Regresijska dijagnostika***: - ispitivanje kakvoće dobivenih rezultata, odnosno provjera teorijskih pretpostavki na kojima počivaju primijenjene metode analize.

JEDNOSTAVNA LINEARNA REGRESIJA

- **Empirijske vrijednosti**:

$$X : x_1, x_2, \dots x_i, \dots x_n$$
$$Y : y_1, y_2, \dots y_i, \dots y_n$$

dijagrama rasipanja

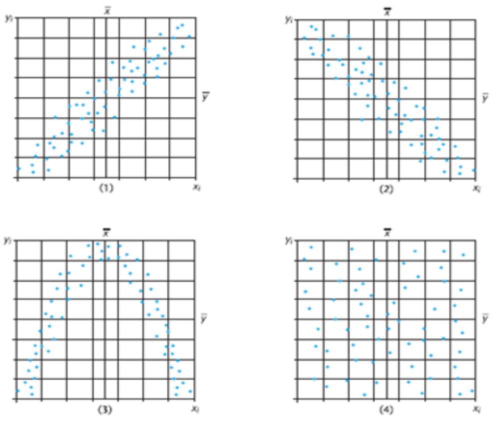
$$(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_n, y_n)$$

Važno je ne raspariti vrijednosti varijable X i Y (ne ih uređivati po veličini)!

- **Iz dijagrama rasipanja se može zaključiti o**:

- - postojanju odnosno nepostojanju povezanosti pojave
- - obliku veze
- - smjeru i jakosti veze.

Tipični dijagrami rasipanja:



Šošić,I.(2006). *PRIMIJENJENA STATISTIKA*. Zagreb, Školska knjiga, stranica 415

Model jednostavne linearne regresije

Model populacije:

$$y_i = \alpha + \beta x_i + e_i \qquad i = 1,2, \dots n$$

pri čemu je:

y_i - i-ta vrijednost zavisne varijable

x_i - i-ta vrijednost nezavisne varijable

α i β – nepoznati parametri

e_i - i-ta nepoznata vrijednost slučajne varijable (greške relacije)

e_i su nepoznate slučajne varijable za koje se pretpostavlja da su međusobno nezavisne i normalno distrubuirane slučajne varijable sa sredinom nula i varijancom σ^2 ,

$(e_i \sim N(0,\sigma^2),\ E(e_i, e_j) = 0,\ \forall i \neq j)$

- Metodom najmanjih kvadrata se radi procjena koeficijenata linearne regresije.
- Ta metoda minimizira kvadrate odstupanja

$$\sum \left(Y - Y_r\right)^2$$

- Jednadžba jednostavne linearne regresije s procijenjenim parametrima ima oblik:

$$Y_r = a + bX$$

- Procijenjeni parametri (koeficijenti) *a i b* jednostavne linearne regresije su:

$$b = \frac{\sum_i x_i y_i - N \cdot \overline{x} \cdot \overline{y}}{\sum_i x_i^2 - N \cdot \overline{x}^2}$$

$$a = \overline{y} - b \cdot \overline{x}$$

- Pri čemu su:

  $\overline{x}$ - srednja vrijednost za x

  $\overline{y}$ - srednja vrijednost za y

  N - broj parova xy

- **Interpretacija:**

- a – konstantni član : regresijska vrijednost zavisne varijable ako je nezavisna varijabla jednaka nuli.

- b - regresijski koeficijent: prosječna linearna promjena (povećanje ili smanjenje ovisno o predznaku koeficijenta) zavisne varijable za jedinično povećanje nezavisne varijable ili linearna promjena regresijske vrijednosti zavisne varijable za jedinično povećanje nezavisne varijable.

Mjera jakosti veze:
- koeficijent determinacije = r^2
- koeficijent linearne korelacije = r

- Koeficijent determinacije:

$$r^2 = \frac{\sum_i \left(Y_r - \overline{Y}\right)^2}{\sum_i \left(Y_i - \overline{Y}\right)^2}$$

ili

$$r^2 = \frac{a\sum_{i=1}^{n} y_i + b\sum_{i=1}^{n} x_i y_i - n\overline{y}^2}{\sum_{i=1}^{n} y_i^2 - n\overline{y}^2}$$

ili

$$r^2 = \frac{\sum_i x_i y_i - N\cdot\overline{x}\cdot\overline{y}}{\left(\sum_i x_i^2 - N\cdot\overline{x}^2\right)\cdot\left[\sum_i y_i^2 - N\overline{y}^2\right]}$$

Koeficijent linearne korelacije je:

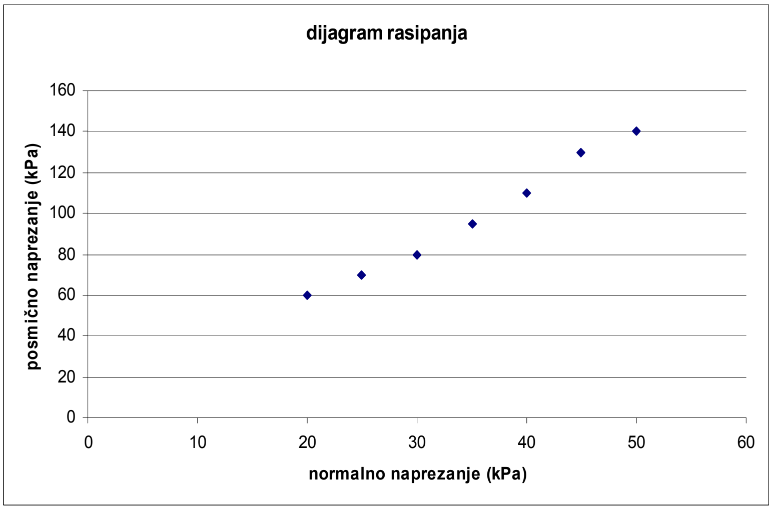
$$r = \sqrt{r^2}$$

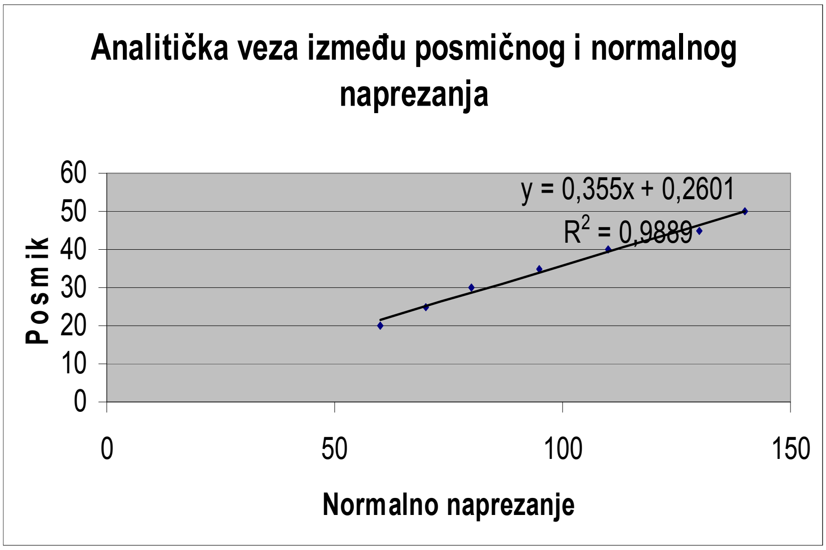
- Koeficijent linearne korelacije ima predznak koeficijenta b.

- Primjer:

  Odredite regresijsku jednadžbu i ispitajte korelaciju između posmičnog i normalnog naprezanja na nekom presjeku opterećenog tijela:

Y - posmik (kPa)	X - normalno naprezanje (kPa)
20	60
25	70
30	80
35	95
40	110
45	130
50	140




Analitička veza između posmičnog i normalnog naprezanja
y = 0,355x + 0,2601
R2 = 0,9889
Posmik
60
50
40
30
20
10
0
0
50
100
150
Normalno naprezanje