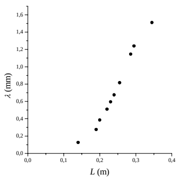
Regresija s transformiranim varijablama

nelinearnu ovisnost prikazati u linearnom obliku:

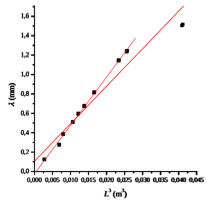
- Mogu se primijeniti jednadžbe za linearnu regresiju.
- Takav grafički prikaz zorno potvrđuje (ili odbacuje) ispravnost primijenjene teorije.

Primjer s prošlog predavanja (napredni praktikum 2):
Modul elastičnosti

$$\lambda = \frac{1}{4E}\frac{L^3}{ab^3}F$$



$x = L^3$



Logaritamski grafovi

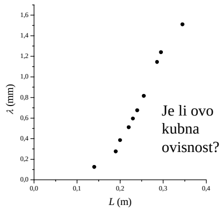
$Y = X^\alpha$ $\quad$ α ne znamo ili želimo provjeriti

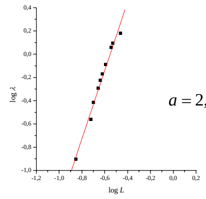
Isti primjer: Modul elastičnosti čelika.
Savijenost šipke je:

$$\lambda = \frac{1}{4E}\frac{F}{ab^3}L^\alpha = K \cdot L^\alpha$$

$$\log\lambda = \log K + \alpha \log L$$

$$x = \log L \qquad y = \log\lambda$$





$a = 2{,}9 \pm 0{,}1$

Nelinearne regresije

Zavisna varijabla nelinearno ovisi o nezavisnoj

Npr.: tjerani prigušeni harmonički oscilator (napredni praktikum 1)

$$y = \frac{A}{\sqrt{\left(\omega_0^2-\omega^2\right)^2 + (\omega/\tau)^2}}$$

parametri A, ω_0 i τ

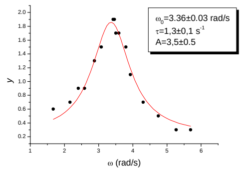
Suma kvadrata odstupanja minimalna:

$$f(A,\omega_0,\tau) = \sum_i \left(y_i - \frac{A}{\sqrt{\left(\omega_0^2-\omega_i^2\right)^2 + (\omega/\tau)^2}} \right)^2$$

Tri jednadžbe:

$$\frac{\partial f(A,\omega_0,\tau)}{\partial A} = 0 \quad ; \quad \frac{\partial f(A,\omega_0,\tau)}{\partial \omega_0} = 0 \quad ; \quad \frac{\partial f(A,\omega_0,\tau)}{\partial \tau} = 0$$

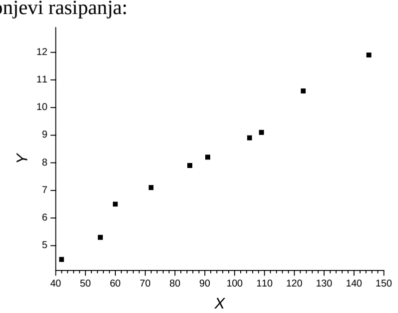
$\Rightarrow$ računalom!



$$y = \frac{A}{\sqrt{\left(\omega_0^2-\omega^2\right)^2 + (\omega/\tau)^2}}$$

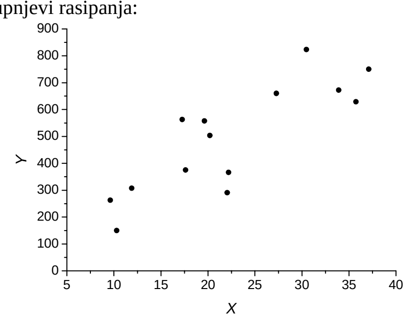
Korelacije

Razni stupnjevi rasipanja:



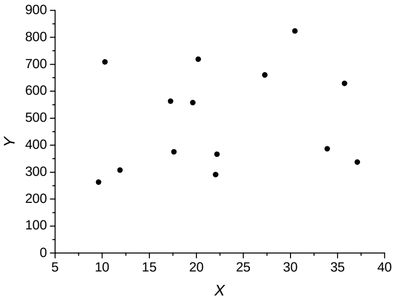
Korelacije

Razni stupnjevi rasipanja:



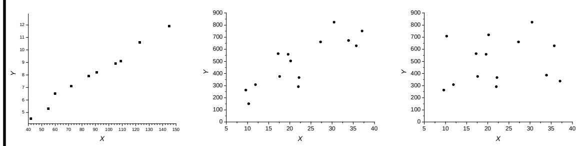
Korelacije

Razni stupnjevi rasipanja:



Korelacije

Razni stupnjevi rasipanja:



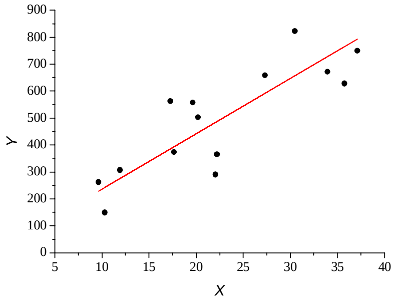
Proučavamo srednju sliku:

ovisnost Y o X:

$$y = ax + b$$

$$a = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - \bar{x}^2}$$

$$b = \frac{\overline{x^2}\,\bar{y} - \bar{x}\,\overline{xy}}{\overline{x^2} - \bar{x}^2}$$



$$a = \frac{m_{11} - m_{10}m_{01}}{m_{20} - m_{10}^2} = \frac{M_{11}}{M_{20}} = \frac{\sigma_{XY}}{\sigma_X^2}$$

kovarijanca :

$$M_{11} = \sigma_{XY} = \frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

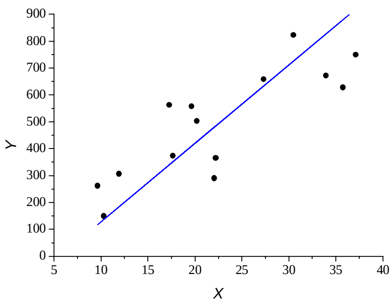
$$M_{11} = \frac{1}{n}\sum_i x_i y_i - \overline{x}\overline{y} = \frac{1}{n}\sum_i x_i y_i - \frac{1}{n^2}\sum_i x_i \sum_i y_i$$

Ovisnost X o Y:

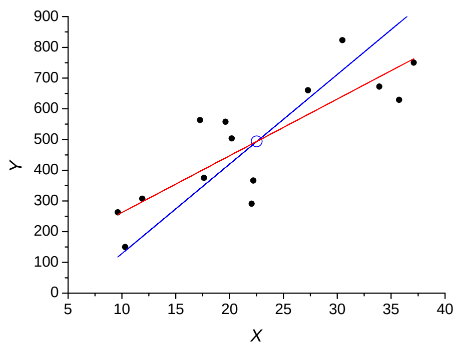
$$x = cy + d$$

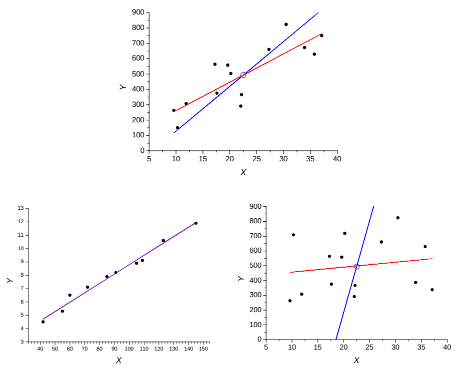
$$c = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{y^2} - \bar{y}^2}$$

$$d = \frac{\overline{y^2}\,\bar{x} - \bar{y}\,\overline{xy}}{\overline{y^2} - \bar{y}^2}$$



$$c = \frac{m_{11} - m_{10}m_{01}}{m_{02} - m_{01}^2} = \frac{M_{11}}{M_{02}} = \frac{\sigma_{XY}}{\sigma_Y^2}$$





Jednadžbe pravaca regresije pomoću kovarijance

Ovisnost Y o X:

jednadžba pravca: $(y-\bar{y})=\dfrac{\sigma_{XY}}{\sigma_X^2}(x-\bar{x})$

Ovisnost X o Y:

jednadžba pravca: $(x-\bar{x})=\dfrac{\sigma_{XY}}{\sigma_Y^2}(y-\bar{y})$

standardizirane jednadžbe pravaca:

$$\frac{(y-\bar{y})}{\sigma_Y}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}\frac{(x-\bar{x})}{\sigma_X}$$

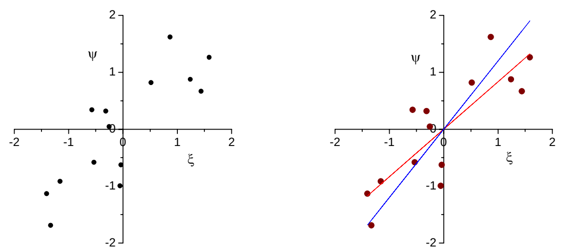
$$\frac{(x-\bar{x})}{\sigma_X}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}\frac{(y-\bar{y})}{\sigma_Y}$$

bezdimenzionalne varijable: $\xi=\dfrac{(x-\bar{x})}{\sigma_X}$ i $\psi=\dfrac{(y-\bar{y})}{\sigma_Y}$

novi pravci regresije: $\psi=\rho\xi$ $\quad\quad$ $\xi=\rho\psi$

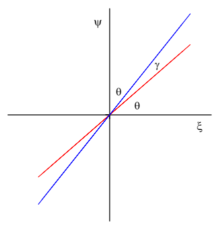
Def.: **koeficijent korelacije:** $\boxed{\rho=\dfrac{\sigma_{XY}}{\sigma_X\sigma_Y}}$

bezdimenzionalno
ne ovisi o jedinicama



Značenje koeficijenta korelacije

poprima vrijednosti: $-1\le\rho\le 1$



$\rho=\mathrm{tg}\,\theta$

$\gamma=\dfrac{\pi}{2}-2\theta$

$\mathrm{tg}\,\gamma=\dfrac{1-\rho^2}{2\rho}$

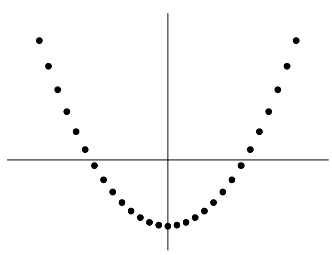
X i Y nezavisni $\Rightarrow$ $\gamma=\dfrac{\pi}{2}$, $\rho=0$

X i Y linearno deterministički vezani $\Rightarrow \gamma=0$, $\rho=-1$ ili $\rho=1$

Dogovor: dobra korelacija $\rightarrow |\rho|>0{,}5$

Napomena:
Kad su X i Y nezavisne varijable, koeficijent korelacije težit će nuli ($\rho\approx 0$). Međutim, činjenica da je $\rho=0$ ne znači nužno da su varijable nezavisne.

Primjer:



Provjeravanje (testiranje) hipoteza

Provjera hipoteze vrlo je bitan dio statističkog zaključivanja.

Da bi se takva provjera formulirala, potrebno je postaviti neku teoriju koja se želi dokazati.
Npr.:

- Novi lijek bolji je za liječenje određenih simptoma od starog.
- Igraća kocka ima pomaknuto težište
- U svijetu se rađa više muškaraca nego žena

U svakom takvom problemu postavljamo dvije tvrdnje (hipoteze) od kojih je točno jedna istinita:

H_0= nul-hipoteza
H_1= alternativna hipoteza

One se ne tretiraju ravnopravno:

U sudskom procesu: H_0: optuženi je nevin
H_1: optuženi je kriv

Hipoteza H_0 smatra se ispravnom dok se ne dokaže H_1.

Za H_0 obično se uzima stara, postojeća teorija:
Npr., za lijek: H_0= stari lijek jednako je dobar kao i novi
za kocku: H_0= kocka je poštena

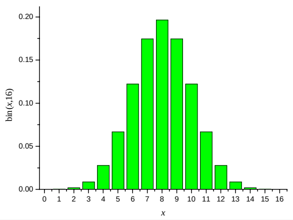
Ako ne odbacimo H_0, to ne znači da je ona ispravna, nego samo da nemamo dovoljno dokaza da ju odbacimo.

Primjer: ispitujemo simetričnost Galtonove daske sa 16 redova.

1. Postavljanje problema, izricanje hipoteza

H_0 : daska je simetrična, tj. X~Bin(16, ½)
H_1 : daska je nagnuta udesno

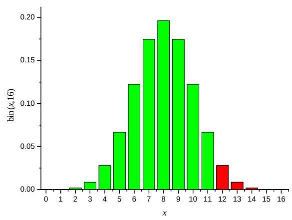
Raspodjela dana nul-hipotezom:



2. Postavljanje problema, kritično područje

Prije uzimanja uzorka podijelimo skup mogućih ishoda u dva područja:

A = područje prihvaćanja H_0
B = područje odbacivanja H_0 = **kritično područje**



Odlučimo se: A={0,1,2,…,11} B={12,13,…,16}

Vrste pogrešaka

		odluka	
		odbaci H_0	prihvati H_0
istina	H_0	pogreška I. vrste	ispravan zaključak
	H_1	ispravan zaključak	pogreška II. vrste

Smatra se da je pogreška I. vrste mnogo ozbiljnija od pogreške II. vrste.

Vjerojatnost pogreške I. vrste:

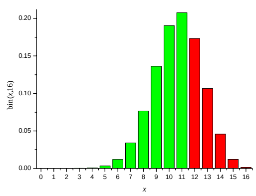
α = vjerojatnost odbacivanja H_0 kada je istinita = $P(B/H_0)$

Vjerojatnost pogreške II. vrste:

β = vjerojatnost prihvaćanja H_0 kada nije istinita = $P(A/H_1)$

U našem primjeru je $\alpha = 0{,}038 \approx 4\%$.

Međutim, čak i da je daska toliko nagnuta da je $p = 2/3$, raspodjela varijable X bi bila



pa bi vjerojatnost pogreške II. vrste uz naš odabir kritičnog područja bila $\beta = 0{,}6488$.

Veliki β se događa kada je uzorak premalen (n = 16).

Signifikantnost testa (važnost, značajnost)

Def:
Za postupak provjere kažemo da ima **razinu signifikantnosti** (važnosti) α ako je

$$P(\text{pogreška I. vrste}) \leq \alpha .$$

Kažemo da je to **test razine α**.

Naš primjer je test razine 0,04.

Tradicionalno se kao razine signifikantnosti uzimaju vrijednosti 0,01; 0,05 ili 0,10.
Za α = 0,05 kažemo da je test signifikantan, a za α = 0,01 kažemo da je test vrlo signifikantan.

Moć testa

Moć statističkog testa mjeri sposobnost testa da odbaci nul-hipotezu kad je uistinu pogrešna, tj. da učini ispravnu odluku.

$$P = 1-\beta$$

U našem primjeru je P = 0,34. Idealno je P = 1. Za snažniji test morali bismo imati veći uzorak.

Opća pravila odabira testa

Prije uzimanja uzorka:

1)Izreci nul-hipotezu i alternativnu hipotezu.
2)Razmotri odgovarajuću raspodjelu danu nul-hipotezom.
3)Odluči o razini signifikantnosti testa.
4)Odredi kritično područje (odluči o kriteriju odbacivanja nul-hipoteze)

Sada uzmi uzorak!

5)Očitaj rezultat testa (izračunaj vrijednost statistike testa)
6)Učini odluku:
Ako je vrijednost statistike testa u kritičnom području, odbaci H_0!
Ako vrijednost statistike testa nije u kritičnom području, nemoj odbaciti H_0!