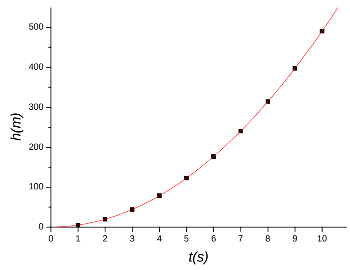
Linearna regresija

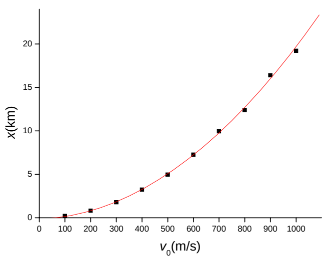
Deterministički povezane varijable:

Vrijeme i put u slobodnom padu:

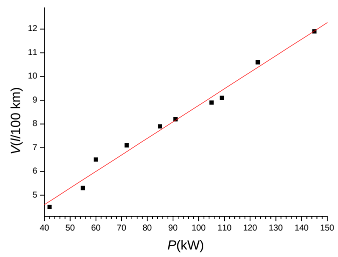


Nedeterministički povezane varijable:

Početna brzina i domet topovske granate:



Snaga automobila i potrošnja goriva



Linearan odnos

Deterministički: $y = ax + b$

Npr.: poznata opruga: $l = \frac{g}{K}m + l_0$

Nedeterministički:

Nezavisna varijabla X , vrijednosti x_i

Za određeni x_i, **zavisna varijabla** Y_i je slučajna varijabla.

Poprima vrijednost y_i

Npr.: opruge iz iste serije

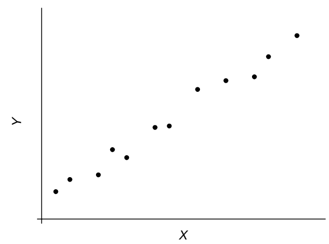
$x_1, x_2, \ldots, x_n$ - vrijednosti nezavisne varijable

Y_i slučajna varijabla pridružena x_i-u.

y_i opažena vrijednost pridružena x_i-u.

n opaženih parova $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

→ graf



linearno?

a=?

b=?

tražimo najvjerojatnije!

Određivanje koeficijenata metodom najmanjih kvadrata

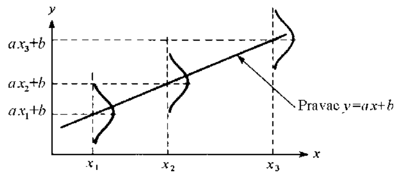
Pretpostavka:

Postoje parametri a i b takvi da za svaku vrijednost x_i nezavisne varijable X, zavisnu varijablu Y_i možemo pisati:

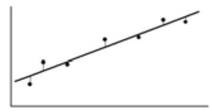
$$Y_i = ax_i + b + \hat{\varepsilon}$$

gdje je $\hat{\varepsilon}$ normalna slučajna varijabla s očekivanjem $E(\hat{\varepsilon}) = 0$ i varijancom $V(\hat{\varepsilon}) = \sigma^2$.

σ^2 je jednaka za sve vrijednosti x.



Za izmjerene (opažene) parove vrijedi:



$$y_i = ax_i + b + \varepsilon_i$$

Princip najmanjih kvadrata:

Od svih pravaca $y = ax + b$, najvjerojatniji pravac regresije jest onaj za koji je suma kvadrata odstupanja

$$f(a,b) = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2 = \sum_{i=1}^{n} \varepsilon_i^2$$

minimalna.

normalne jednadžbe

Suma kvadrata odstupanja je minimalna kada istodobno vrijedi:

$$\frac{\partial f(a,b)}{\partial a} = 0 \quad \text{i} \quad \frac{\partial f(a,b)}{\partial b} = 0$$

$$\sum_{i=1}^{n} y_i x_i - a\sum_{i=1}^{n} x_i^2 - b\sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i - nb = 0$$

rješenje:

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad b = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$N_z = n\sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad a = \frac{n}{N_z}\sum_{i=1}^{n} (x_i - \bar{x}) y_i \qquad b = \bar{y} - a\bar{x}$$

Napomene:

- Prije računanja pravca regresije treba u grafu provjeriti ima li smisla linearna regresija i jesu li podaci podjednako raspršeni.
- Rezultate sumiranja ne smije se zaokruživati jer pogreška zaokruživanja bitno utječe na razliku velikih sličnih brojeva.

Tražimo nepouzdanosti parametara a i b

$$V(\hat{a}) = \frac{n^2}{N_z^2}\sum_{i=1}^{n} (x_i - \bar{x})^2 V(Y_i)$$

$$\sigma_a^2 = \frac{n^2}{N_z^2}\sum_{i=1}^{n} (x_i - \bar{x})^2 \sigma^2 = n\frac{\sigma^2}{N_z}$$

Nepristrani procjenitelj za σ^2 dan je izrazom:

$$s^2 = \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - y_i^*)^2}{n-2} = \frac{\sum_{i=1}^{n} [y_i - (ax_i + b)]^2}{n-2}$$

Konačni rezultati:

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad b = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

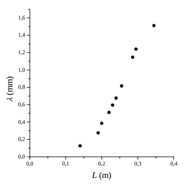
$$M_a = \sqrt{\frac{1}{(n-2)}\left[\frac{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} - a^2\right]} \qquad M_b = M_a\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

Regresija s transformiranim varijablama

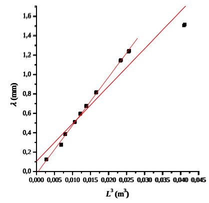
nelinearnu ovisnost prikazati u linearnom obliku:

- 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

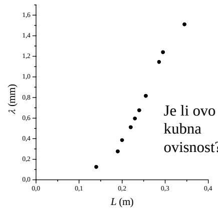
$Y = X^{\alpha}$ $\qquad$ α ne znamo ili želimo provjeriti

Isti primjer: Modul elastičnosti čelika.
Savijenost šipke je:

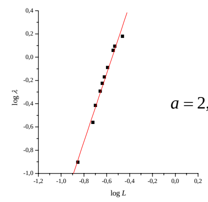
$$\lambda = \frac{1}{4E}\frac{F}{ab^3}L^{\alpha} = K \cdot L^{\alpha}$$

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

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



Je li ovo kubna ovisnost?



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

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

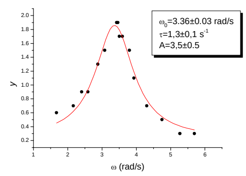
Suma kvadrata odstupanja minimalna:

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

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!



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