

Tema: Regresijska analiza

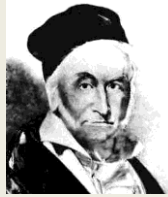


Regresijska analiza

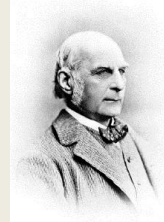
Uvod

U razvoju regresijske analize značajnu ulogu imaju radovi:

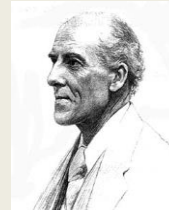
→ Gaussa



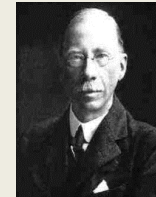
→ Galtona



→ Pearsona



→ Yulea



Regresijska analiza

Uvod

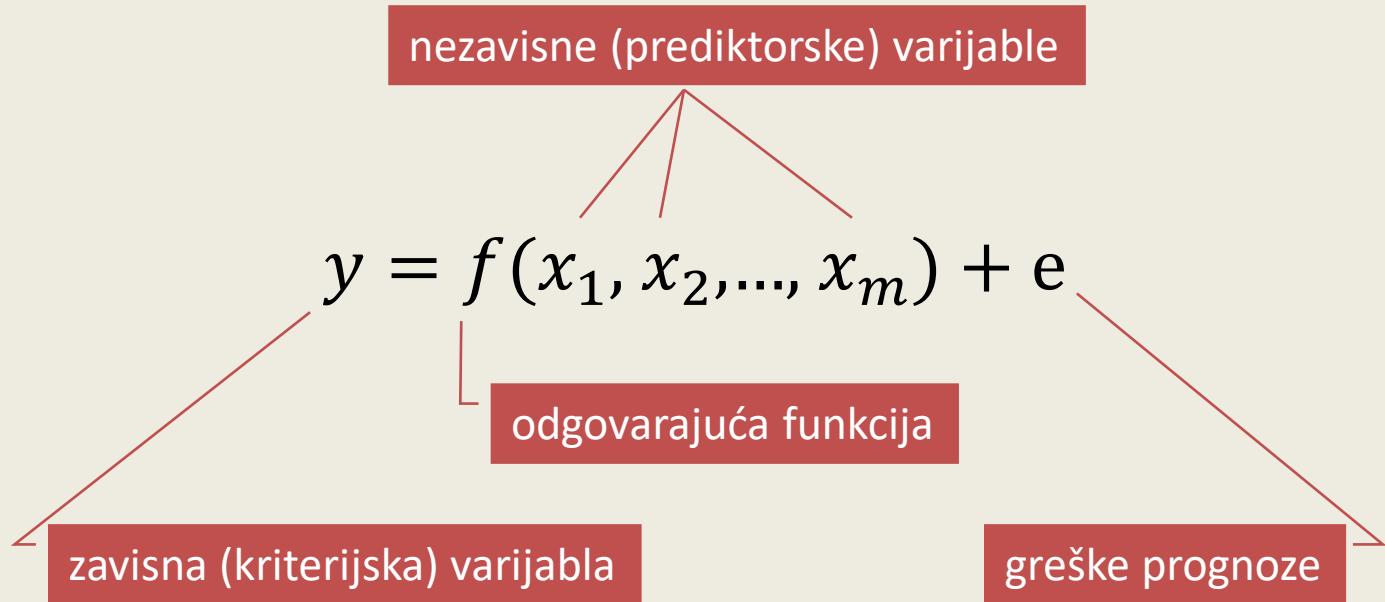
Regeresijska analiza predstavlja matematičko-statistički postupak kojim se utvrđuje odgovarajuća funkcionalna veza (relacija) između jedne zavisne varijable ili kriterijske varijable i jedne ili više nezavisnih varijabli ili prediktorskih varijabli.

Zavisna (kriterijska) varijabla predstavlja varijablu čije se varijacije objašnjavaju (prognoziraju) putem nezavisnih varijabli.

Nezavisne (prediktorske) varijable predstavljaju varijable na temelju kojih se objašnjavaju varijacije zavisne varijable.

Regresijska analiza

Uvod



Regresijska analiza

Uvod

Regresijske modele moguće je generalno podijeliti na temelju dvaju kriterija:

1. prema broju nezavisnih varijabli:

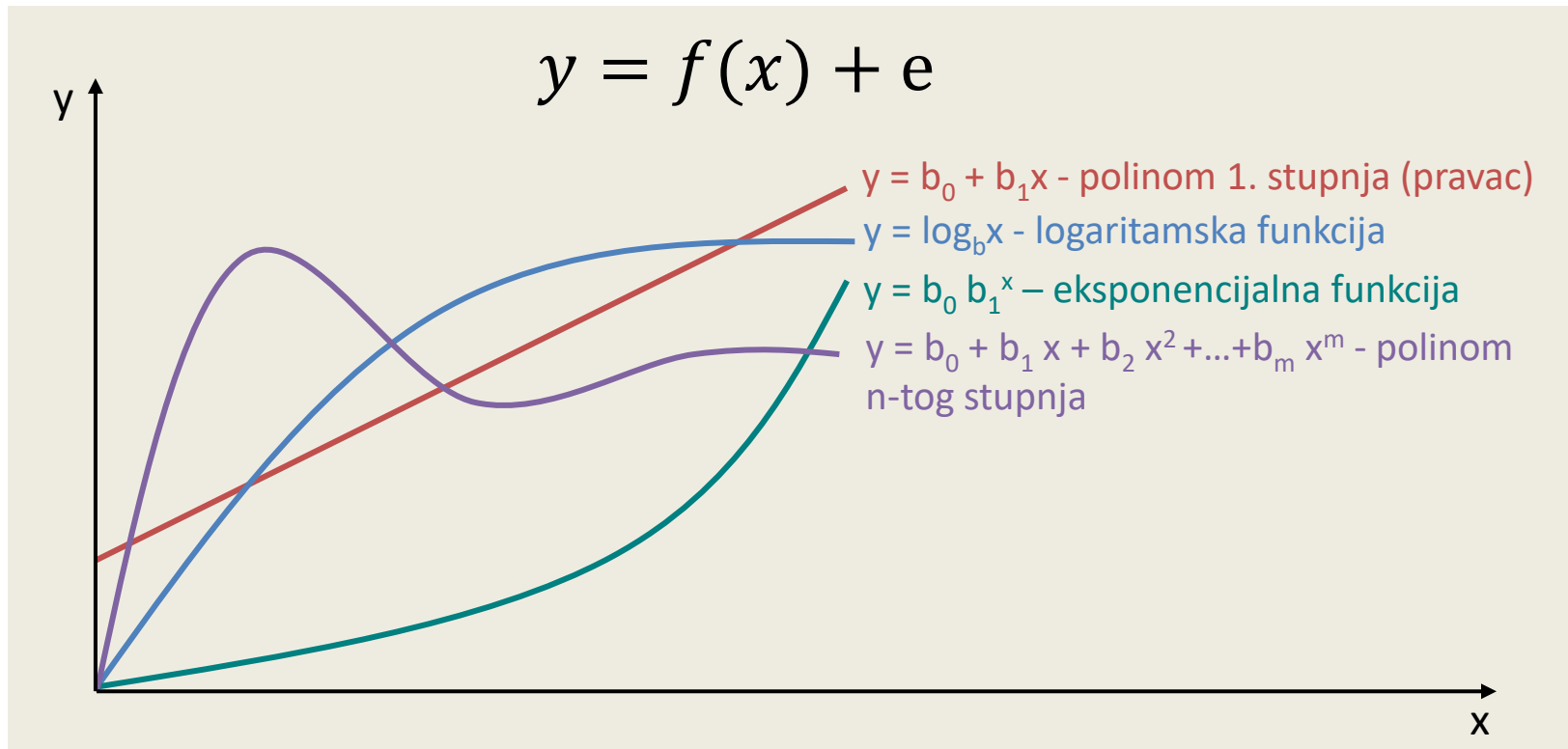
- jednostavne (simple) i
- višestruke (multiple) regresijske modele

2. prema odnosu između zavisne i nezavisnih varijabli:

- linearne i
- nelinearne regresijske modele.

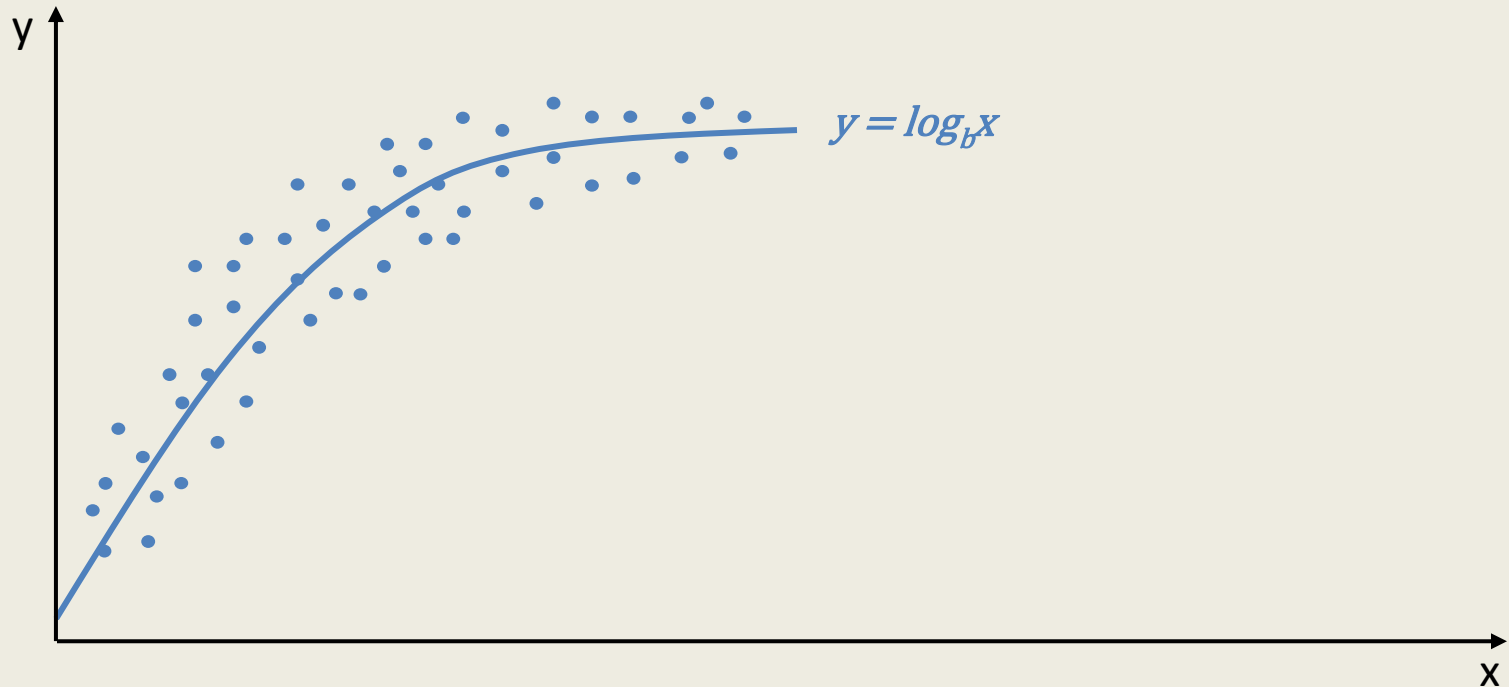
Regresijska analiza

Jednostavna regresijska analiza



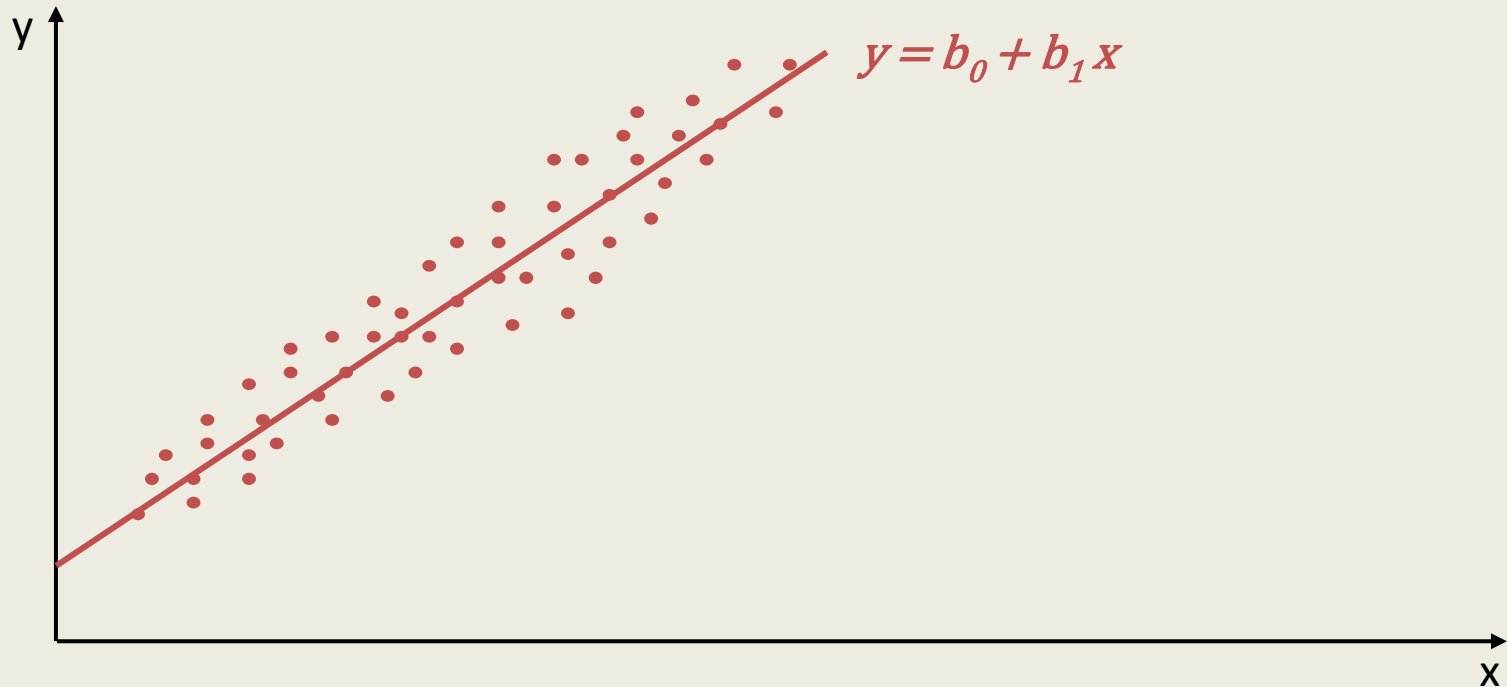
Regresijska analiza

Jednostavna regresijska analiza



Regresijska analiza

Jednostavna regresijska analiza



Regresijska analiza

Jednostavna regresijska analiza (linearni model)

$$y = b_0 + b_1 x + e$$

- y - zavisna (kriterijska) varijabla
- b_0 i b_1 - regresijski koeficijenti
- x - nezavisna (prediktorska) varijabla
- e - pogreške prognoze

Regresijska analiza

Jednostavna regresijska analiza (linearni model)

$$y_i = b_0 + b_1 x_i + e_i;$$

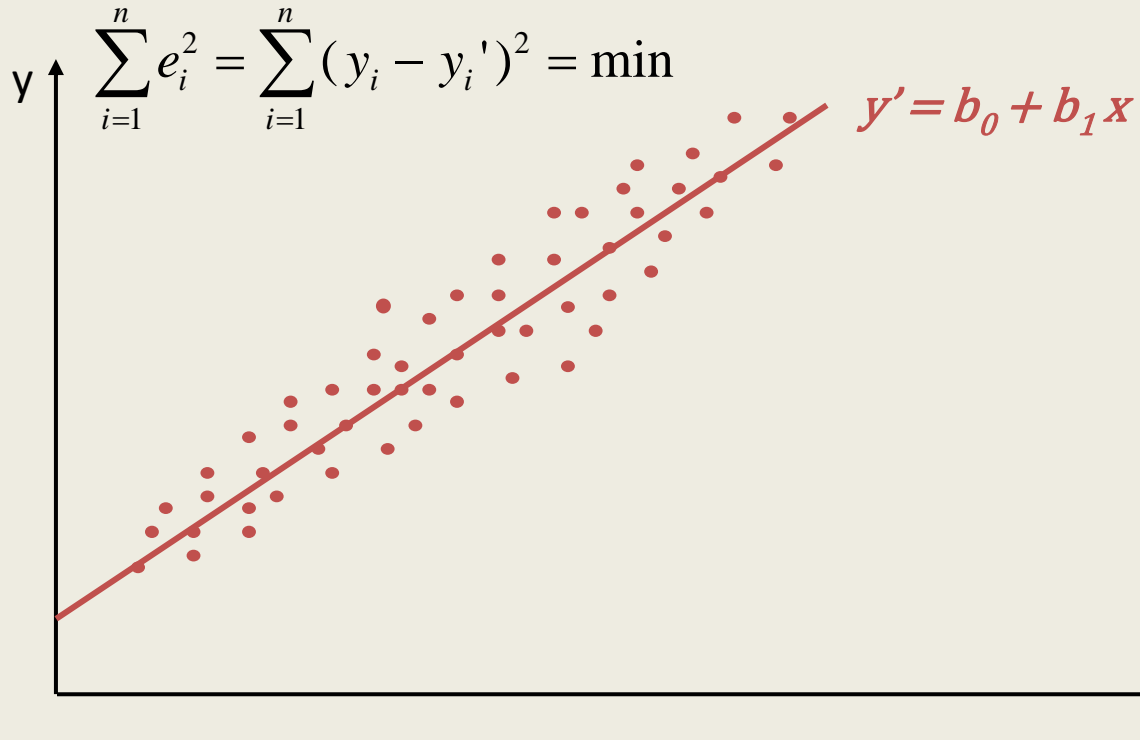
$$i = 1, 2, \dots, n$$

$$y = Xb + e$$

$$\begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix} \cdot \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} + \begin{pmatrix} e_1 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{pmatrix}$$

Regresijska analiza

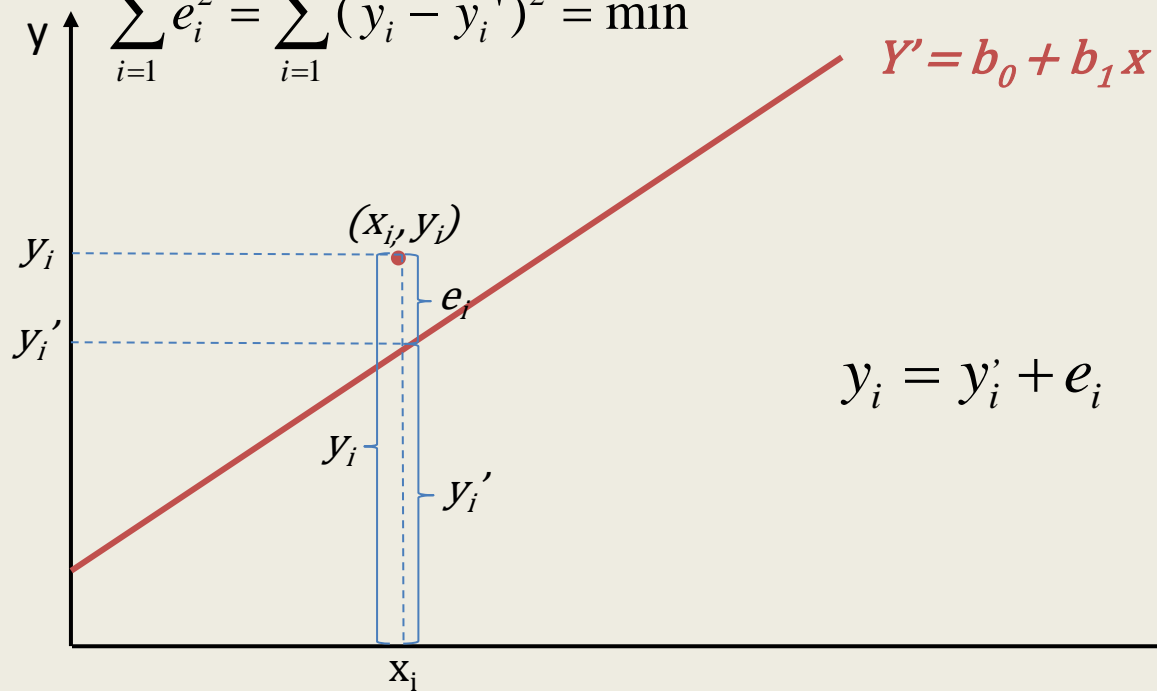
Jednostavna regresijska analiza (linearni model)



Regresijska analiza

Jednostavna regresijska analiza (linearni model)

$$y \quad \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - y_i')^2 = \min$$



Regresijska analiza

Jednostavna regresijska analiza (linearni model)

$$y = Xb \quad /X^T$$

$$X^T X b = X^T y$$

$$X^T X b = X^T y \quad / (X^T X)^{-1}$$

$$b = (X^T X)^{-1} X^T y$$

$$b_0 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$b_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

Regresijska analiza

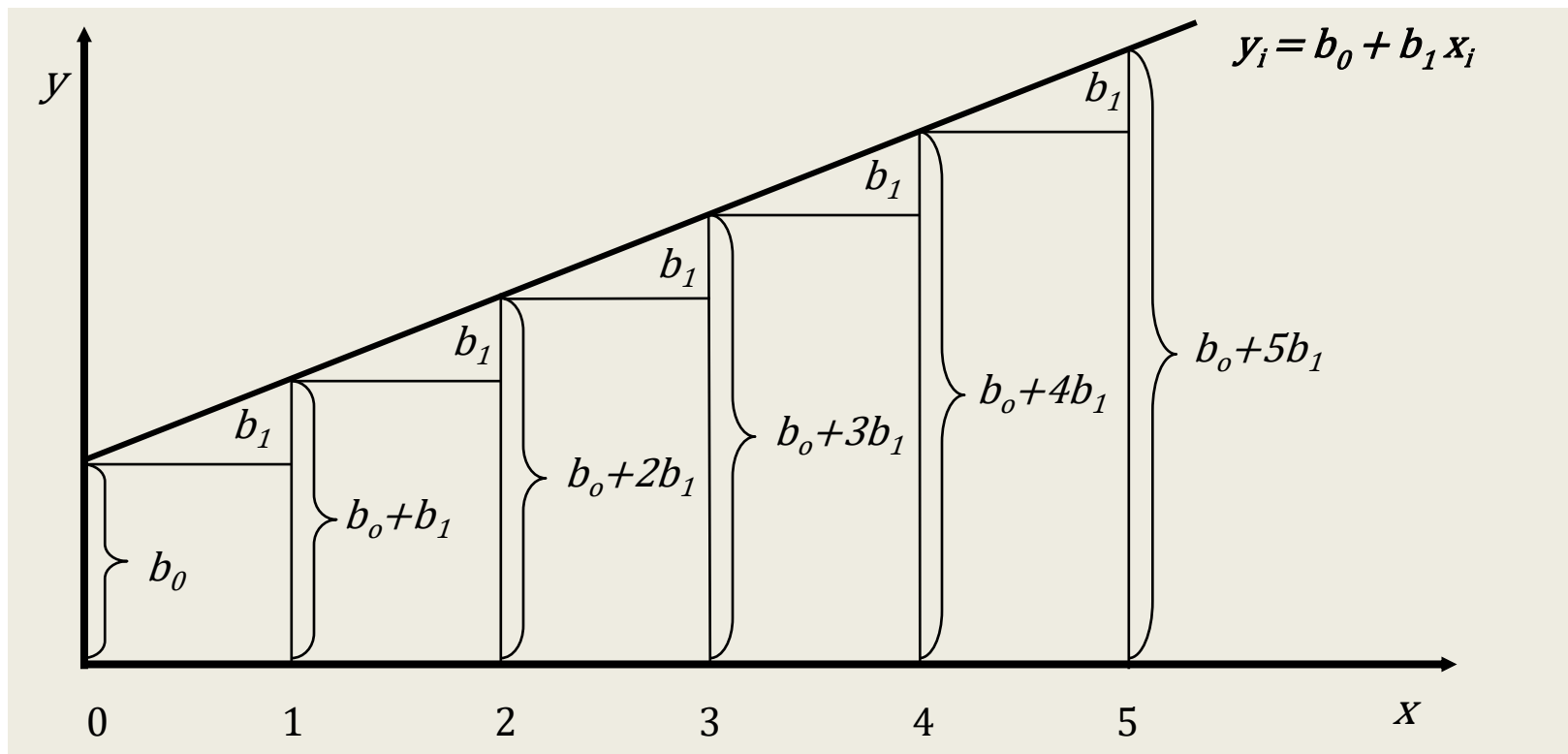
Jednostavna regresijska analiza (linearni model)

Regresijski koeficijent b_0 predstavlja odsječak na osi zavisne varijable y , odnosno, vrijednost zavisne varijable y ukoliko je vrijednost nezavisne varijable $x = 0$

Regresijski koeficijent b_1 određuje nagib pravca, odnosno, pokazuje koliko se u prosjeku linearno mijenja vrijednost zavisne varijable y za jedinični porast vrijednosti nezavisne varijable x .

Regresijska analiza

Jednostavna regresijska analiza (linearni model)



Regresijska analiza

Jednostavna regresijska analiza (linearni model)

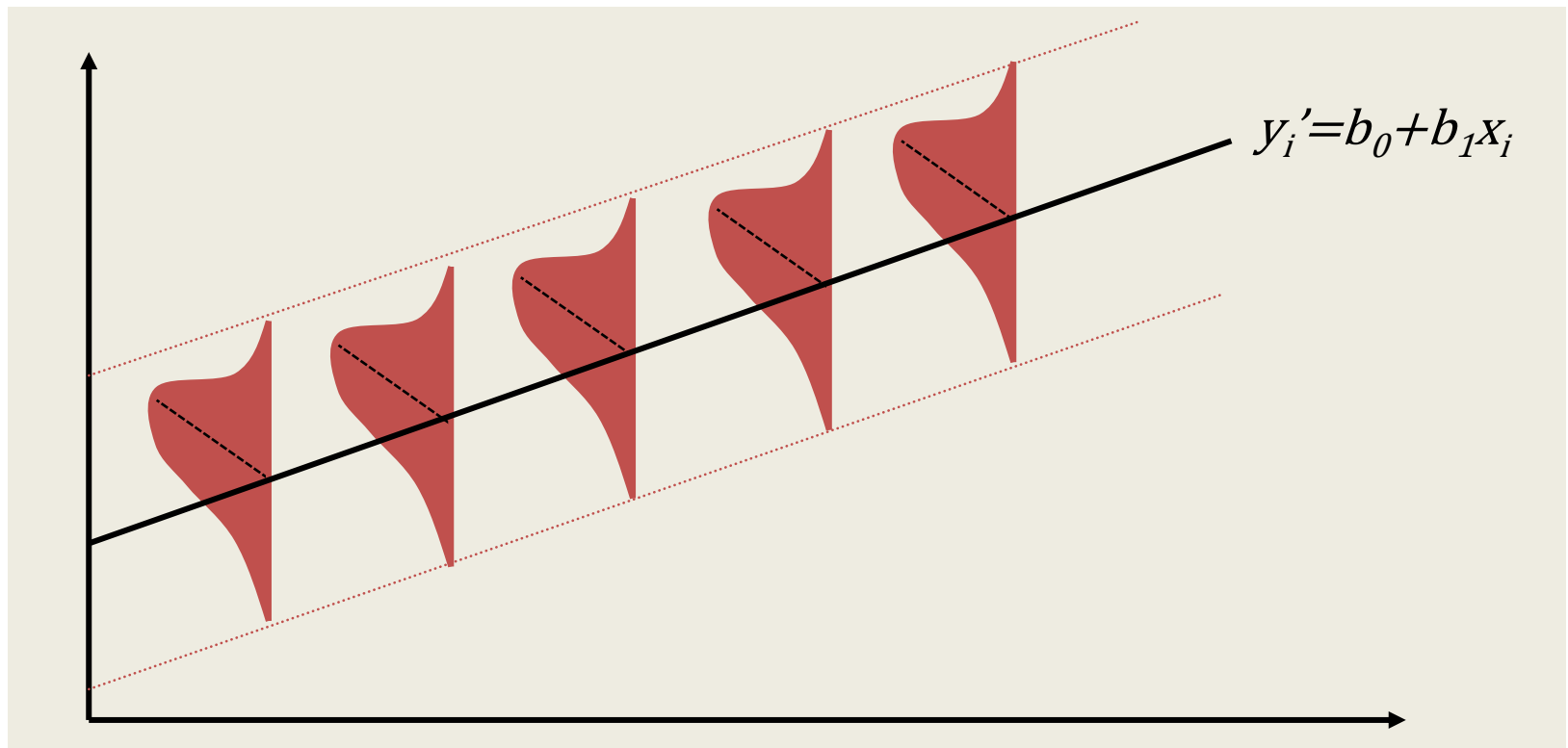
Rezidualni vektor e dobije se

$$e = y - y' = y - Xb$$

- očekivana vrijednost jednaka je nuli $E(e) = 0$
- varijanca distribucije reziduala e_i je konstantna za sve vrijednosti x_i
- distribucija vjerojatnosti reziduala e_i je normalna

Regresijska analiza

Jednostavna regresijska analiza (linearni model)



Regresijska analiza

Jednostavna regresijska analiza (linearni model)

Varijancu rezidualnih rezultata

$$\sigma_e^2 = \frac{r_{ss}}{df} = \frac{\sum_{i=1}^n (y_i - y_i')^2}{n-2}$$

Drugi korijen iz varijance rezidualnih vrijednosti predstavlja standardnu pogrešku prognoze (σ_e)

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i')^2}{n-2}}$$

Regresijska analiza

Jednostavna regresijska analiza (linearni model)

$$r_{ss} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - y'_i)^2$$

$$p_{ss} = \sum_{i=1}^n (y'_i - \bar{y})^2$$

$$t_{ss} = p_{ss} + r_{ss} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y'_i - \bar{y})^2 + \sum_{i=1}^n (y_i - y'_i)^2$$

Regresijska analiza

Jednostavna regresijska analiza (linearni model)

Koeficijent determinacije

$$r^2 = \frac{p_{ss}}{t_{ss}} \frac{\sum_{i=1}^n (y_i' - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$r^2 = \frac{p_{ss}}{t_{ss}} = \frac{t_{ss} - r_{ss}}{t_{ss}} = 1 - \frac{r_{ss}}{t_{ss}}$$

$$r = \sqrt{r^2} = \sqrt{\frac{p_{ss}}{t_{ss}}} = \sqrt{1 - \frac{r_{ss}}{t_{ss}}}$$

Regresijska analiza

Jednostavna regresijska analiza (linearni model)

Koeficijent korelacije između kriterijske i prediktorske varijable izražava njihovog jačinu funkcionalnog odnosa.

Kada je $r = 0$ to znači da na veličinu ukupnog varijabiliteta kriterijske varijable Y, nezavisna varijabla X nema nikakvog utjecaja.

Ako koeficijent korelacije ima maksimalnu vrijednost $r = 1$, znači da je cjelokupan varijabilitet varijable Y moguće pripisati utjecaju varijable X.

Regresijska analiza

Jednostavna regresijska analiza (linearni model)

Primjer:

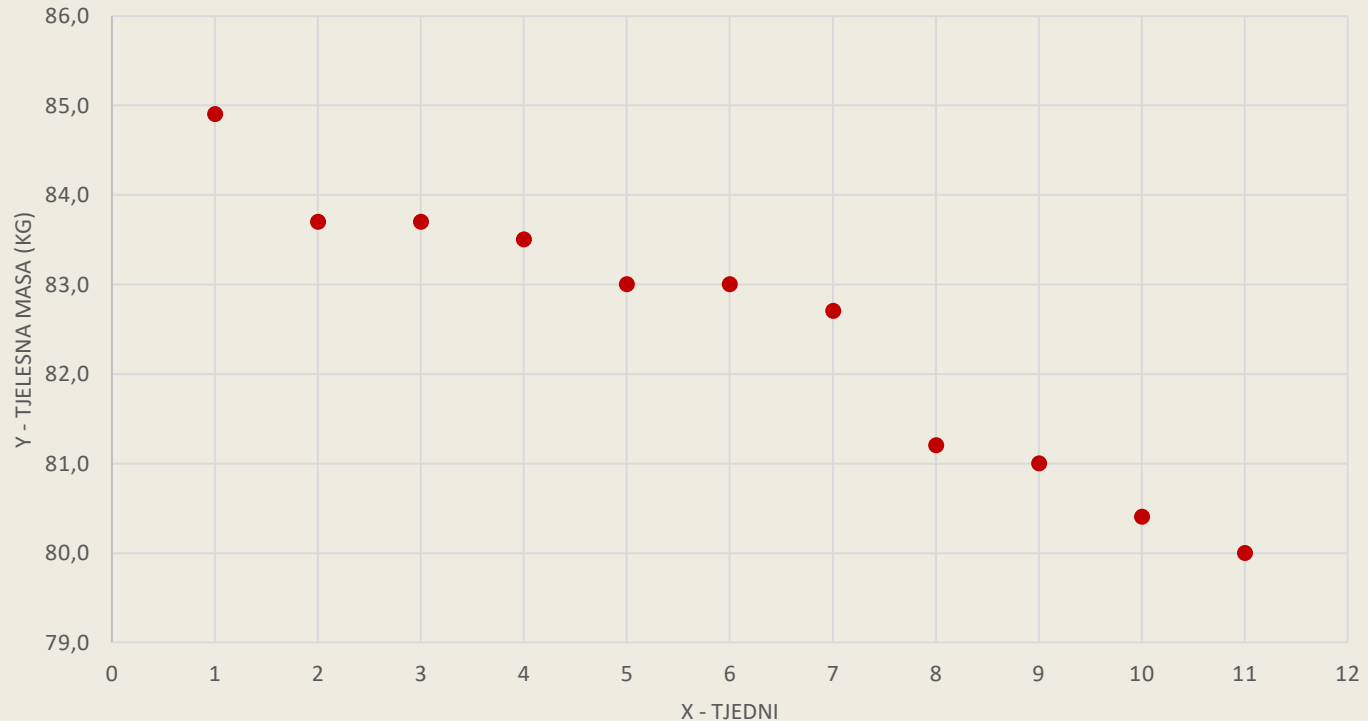
Ispitanik *TT* je bio podvrgnut 11 tjednom kineziološkom tretmanu s ciljem redukcije potkožnog masnog tkiva i poboljšanja funkcionalnih sposobnosti.

Nakon svakog tjedna tretmana mjerena je varijabla *tjelesna težina* (Y). Utvrdite linearnu funkcionalnu zavisnost između nezavisne varijable X - 11 tjedana kineziološkog tretmana i varijable Y - tjelesna težina.

X	Y
1	84,9
2	83,7
3	83,7
4	83,5
5	83,0
6	83,0
7	82,7
8	81,2
9	81,0
10	80,4
11	80,0

Regresijska analiza

Jednostavna regresijska analiza (linearni model)



Regresijska analiza

Jednostavna regresijska analiza (linearni model)

$$y = Xb + e$$

$$\begin{array}{c|cc|c} 84,9 & 1 & 1 & e_1 \\ 83,7 & 1 & 2 & e_2 \\ 83,7 & 1 & 3 & \cdot \\ 83,5 & 1 & 4 & \cdot \\ 83 & 1 & 5 & \cdot \\ 83 & 1 & 6 & \cdot \\ 82,7 & 1 & 7 & \cdot \\ 81,2 & 1 & 8 & \cdot \\ 81 & 1 & 9 & \cdot \\ 80,4 & 1 & 10 & \cdot \\ 80 & 1 & 11 & e_{11} \end{array} = \begin{array}{c} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{array} \begin{array}{c} |b_0| \\ |b_1| \end{array} +$$

Regresijska analiza

Jednostavna regresijska analiza (linearni model)

$$y = Xb$$

$$X^T y = X^T X b$$

$$/X^T$$

$$\begin{array}{c} \left| \begin{array}{cccccccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \end{array} \right| \begin{array}{c} 84,9 \\ 83,7 \\ 83,7 \\ 83,5 \\ 83 \\ 83 \\ 82,7 \\ 81,2 \\ 81 \\ 80,4 \\ 80 \end{array} = \left| \begin{array}{c} 907,1 \\ 5391,9 \end{array} \right| \end{array}$$

Regresijska analiza

Jednostavna regresijska analiza (linearni model)

$$y = Xb$$

$$/X^T$$

$$X^T y = X^T X b$$

$$\begin{array}{c} \left| \begin{array}{cccccccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 11 \end{array} \right| \\ \left| \begin{array}{c} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{array} \right| \end{array} = \begin{array}{c} \left| \begin{array}{cc} 11 & 66 \\ 66 & 506 \end{array} \right| \end{array}$$

Regresijska analiza

Jednostavna regresijska analiza (linearni model)

$$y = Xb \quad /X^T$$

$$X^T y = X^T X b \quad /(X^T X)^{-1}$$

$$\begin{vmatrix} 907,1 \\ 5391,9 \end{vmatrix} = \begin{vmatrix} 11 & 66 \\ 66 & 506 \end{vmatrix} \begin{vmatrix} b_0 \\ b_1 \end{vmatrix}$$

$$907,1 = 11b_0 + 66b_1$$

$$5391,9 = 66b_0 + 506b_1$$

Regresijska analiza

Jednostavna regresijska analiza (linearni model)

$$b = (X^T X)^{-1} X^T y$$

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 11 & 66 \\ 66 & 506 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 907,1 \\ 5391,9 \end{bmatrix}$$

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 0,4182 & -0,0545 \\ -0,0545 & 0,0092 \end{bmatrix} \cdot \begin{bmatrix} 907,1 \\ 5391,9 \end{bmatrix} = \begin{bmatrix} 85,23 \\ -0,461 \end{bmatrix}$$

$$b_0 = 85,23$$

$$b_1 = -0,461$$

Regresijska analiza

Jednostavna regresijska analiza (linearni model)

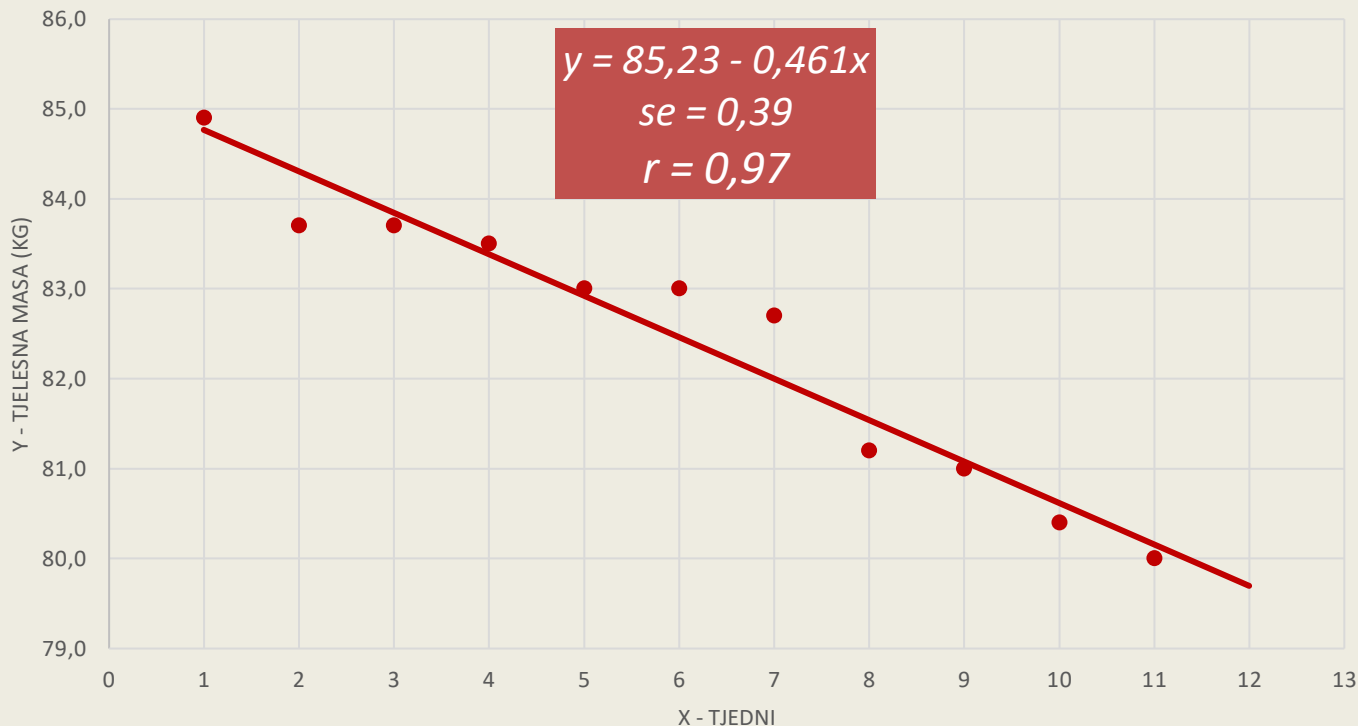
$$y - y' = e$$

84,9	84,77	0,13
83,7	84,31	-0,16
83,7	83,85	-0,15
83,5	83,39	0,11
83	82,92	0,08
83	82,46	0,54
82,7	82,00	0,70
81,2	81,54	-0,34
81	81,08	-0,08
80,4	80,62	-0,22
80	80,16	-0,16

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i')^2}{n - 2}} = 0,39$$

Regresijska analiza

Jednostavna regresijska analiza (linearni model)



Regresijska analiza

Višestruka regresijska analiza

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m + e$$

- y - zavisna (kriterijska) varijabla
- $b_0 \dots b_m$ regresijski koeficijenti
- $x_1 \dots x_m$ nezavisne (prediktorske) varijable
- e - pogreške prognoze.

Regresijska analiza

Višestruka regresijska analiza

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_mx_{im} + e_i$$

gdje je $i = 1 \dots n$

$$y = Xb + e$$

$$\begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdot & \cdot & x_{1m} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & \cdot & \cdot & x_{nm} \end{pmatrix} \cdot \begin{pmatrix} b_0 \\ b_1 \\ \cdot \\ b_m \end{pmatrix} + \begin{pmatrix} e_1 \\ \cdot \\ \cdot \\ e_n \end{pmatrix}$$

Regresijska analiza

Višestruka regresijska analiza

$$y = Xb \quad / X^T$$

$$X^T y = X^T X b \quad / (X^T X)^{-1}$$

$$b = (X^T X)^{-1} X^T y$$

$$y' = Xb$$

$$e = y - y'$$

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i')^2}{n - (m + 1)}}$$

$$\rho = \sqrt{\frac{p_{ss}}{t_{ss}}} = \sqrt{\frac{\sum_{i=1}^n (y_i' - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Regresijska analiza

Višestruka regresijska analiza

$$k = \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_m z_m + \varepsilon$$

$$k = Z \beta + \varepsilon$$

$$\begin{vmatrix} k_1 \\ \cdot \\ \cdot \\ k_n \end{vmatrix} = \begin{vmatrix} z_{11} & \cdot & \cdot & z_{1m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ z_{n1} & \cdot & \cdot & z_{nm} \end{vmatrix} \cdot \begin{vmatrix} \beta_1 \\ \cdot \\ \cdot \\ \beta_m \end{vmatrix} + \begin{vmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \varepsilon_1 \end{vmatrix}$$

Regresijska analiza

Višestruka regresijska analiza

$$k = Z\beta \quad /Z^T n^{-1}$$

$$Z^T k n^{-1} = Z^T Z n^{-1} \beta$$

$$\begin{array}{c|c|c|c|c} r_1 & 1 & r_{12} & \cdot & r_{1m} \\ \cdot & r_{21} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_m & r_{m1} & \cdot & \cdot & 1 \end{array}$$

Regresijska analiza

Višestruka regresijska analiza

$$r = R\beta \quad / R^{-1}$$

$$\beta = R^{-1}r$$

$$k' = Z\beta$$

$$\varepsilon = k - k'$$

$$\rho^2 = \sum_{i=1}^m \beta_1 r_1 + \beta_2 r_2 + \dots + \beta_m r_m = \sum_{i=1}^m \beta_i r_i = \sum_{i=1}^n p_i$$

$$\sigma_\varepsilon = \sqrt{1 - \rho^2}$$

Regresijska analiza

Višestruka regresijska analiza

Primjer:

12 ispitanik izmjereno je u varijablama *skok u dalj s mjesta* (SDM), *bench press* (BP) i *bacanju kugle* (BK).

Utvrđite utjecaj eksplozivne snage nogu procjenjene duljinom skoka iz mjesta i maksimalne jakosti procjenjene maksimalno podignutom težinom iz ležanja na ravnoj klupi (bench press) na uspješnost bacanja kugle.

SDM	BP	BK
245	75	14,8
270	85	15,35
240	87	15,45
295	90	16,9
275	80	17,05
295	100	18,5
265	70	15,1
225	65	14,7
240	80	15,6
285	82	16,3
230	75	14,6
260	90	17,2

Regresijska analiza

Višestruka regresijska analiza

$$\begin{array}{c} y \\ \hline 1480 \\ 1535 \\ 1545 \\ 1690 \\ 1705 \\ 1850 \\ 1510 \\ 1470 \\ 1560 \\ 1630 \\ 1460 \\ 1720 \\ \hline \end{array} = \begin{array}{c} X \\ \hline 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \hline \end{array} \begin{array}{c} \\ \\ 245 \\ 270 \\ 240 \\ 295 \\ 275 \\ 295 \\ 265 \\ 225 \\ 240 \\ 285 \\ 230 \\ 260 \\ \hline \end{array} \begin{array}{c} \\ \\ 75 \\ 85 \\ 87 \\ 80 \\ 90 \\ 100 \\ 70 \\ 65 \\ 80 \\ 82 \\ 75 \\ 90 \\ \hline \end{array} \cdot \begin{array}{c} \left| \begin{array}{c} b_0 \\ b_1 \\ b_2 \end{array} \right| \\ + \end{array} \begin{array}{c} e \\ \hline e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ e_{12} \\ \hline \end{array}$$

Regresijska analiza

Višestruka regresijska analiza

$$y = Xb$$

$$X^T y = X^T X b$$

$$/X^T$$

$$/(X^T X)^{-1}$$

$$\begin{vmatrix} 19155 \\ 5013875 \\ 1573400 \end{vmatrix} = \begin{vmatrix} 12 & 3125 & 979 \\ 3125 & 820375 & 256650 \\ 979 & 256650 & 80893 \end{vmatrix} \cdot \begin{vmatrix} b_0 \\ b_1 \\ b_2 \end{vmatrix}$$

$$19155 = 12b_0 + 3125b_1 + 979b_2$$

$$5013875 = 3125b_0 + 820375b_1 + 256650b_2$$

$$1573400 = 979b_0 + 256650b_1 + 80893b_2$$

Regresijska analiza

Višestruka regresijska analiza

$$b = (X^T X)^{-1} X^T y$$

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 12 & 3125 & 979 \\ 3125 & 820375 & 256650 \\ 979 & 256650 & 80893 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 19155 \\ 5013875 \\ 1573400 \end{bmatrix}$$

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 10,74 & -0,033327 & -0,024302 \\ -0,033327 & 0,000267 & -0,000445 \\ -0,24302 & -0,0004448 & 0,001718 \end{bmatrix} \cdot \begin{bmatrix} 19155 \\ 5013875 \\ 1573400 \end{bmatrix} = \begin{bmatrix} 484 \\ 2,1 \\ 6,9 \end{bmatrix}$$

Regresijska analiza

Višestruka regresijska analiza

$$b_0 = 484$$

$$b_1 = 2,1$$

$$b_2 = 6,9$$

$$y_i' = 484 + 2,1x_{1,i} + 6,9x_{2,i}$$

- y - bacanje kugle (cm)
- x_1 - skok u dalj s mjesta (cm)
- x_2 - bench press (kg)

Regresijska analiza

Višestruka regresijska analiza

$$k = Z\beta \quad /Z^T$$

$$Z^T k = Z^T Z \beta \quad /n^{-1}$$

$$\underbrace{Z^T k}_{r} n^{-1} = \underbrace{Z^T Z}_{R} n^{-1} \beta$$

$$r = R \beta$$

$$\begin{vmatrix} 0,78 \\ 0,82 \end{vmatrix} = \begin{vmatrix} 1 & 0,65 \\ 0,65 & 1 \end{vmatrix} \cdot \begin{vmatrix} \beta_1 \\ \beta_2 \end{vmatrix}$$

Regresijska analiza

Višestruka regresijska analiza

$$\beta = R^{-1} r$$
$$\begin{vmatrix} \beta_1 \\ \beta_2 \end{vmatrix} = \begin{vmatrix} 1 & 0,65 \\ 0,65 & 1 \end{vmatrix}^{-1} \cdot \begin{vmatrix} 0,78 \\ 0,82 \end{vmatrix}$$

$$\beta_1 = 0,42$$

$$\beta_2 = 0,55$$

$$\rho^2 = \beta_1 r_1 + \beta_2 r_2 = 0,42 \cdot 0,78 + 0,55 \cdot 0,82$$

$$\rho^2 = 0,33 + 0,45 = 0,78$$

$$\rho = \sqrt{0,78} = 0,88$$

Regresijska analiza

Testiranje značajnosti regresijskog modela

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

$$F = \frac{p_{ss} / df_p}{r_{ss} / df_r}$$

Varijanca	Stupnjevi slobode (df)	Suma kvadrata	Sredina kvadrata	F - vrijednost
Prognozirana	m	$p_{ss} = \sum_{i=1}^n (y'_i - \bar{y})^2$	$mp_{ss} = \frac{p_{ss}}{m}$	$F = \frac{mp_{ss}}{mr_{ss}}$
Rezidualna	$n - (m+1)$	$r_{ss} = \sum_{i=1}^n (y_i - y_i')^2$	$mr_{ss} = \frac{r_{ss}}{n - (m+1)}$	
Ukupna	$n - 1$	$t_{ss} = \sum_{i=1}^n (y_i - \bar{y})^2$		

Regresijska analiza

Testiranje značajnosti regresijskog modela

	Stupnjevi slobode	Suma kvadrata	Sredina kvadrata	F-vrijednost
Prognozirana	2	127744,6	63872,28	15,59
Rezidualna	9	36861,7	4095,74	

$$F = 15,59 > F_{(2,9; 0.05)} = 4,26$$

$$H_1: \rho \neq 0$$

Regresijska analiza

Testiranje značajnosti regresijskog modela

$$H_0 : b_j = 0,$$

$$H_1 : b_j \neq 0$$

$$0 \pm t_{\alpha/2} \cdot \sigma_{b_j}$$

$$\sigma_{b_j} = \frac{\sigma_e}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$t = \frac{b_j}{\sigma_{b_j}}$$

$$|t| < t_\alpha \Rightarrow H_0$$

$$|t| > t_\alpha \Rightarrow H_1$$