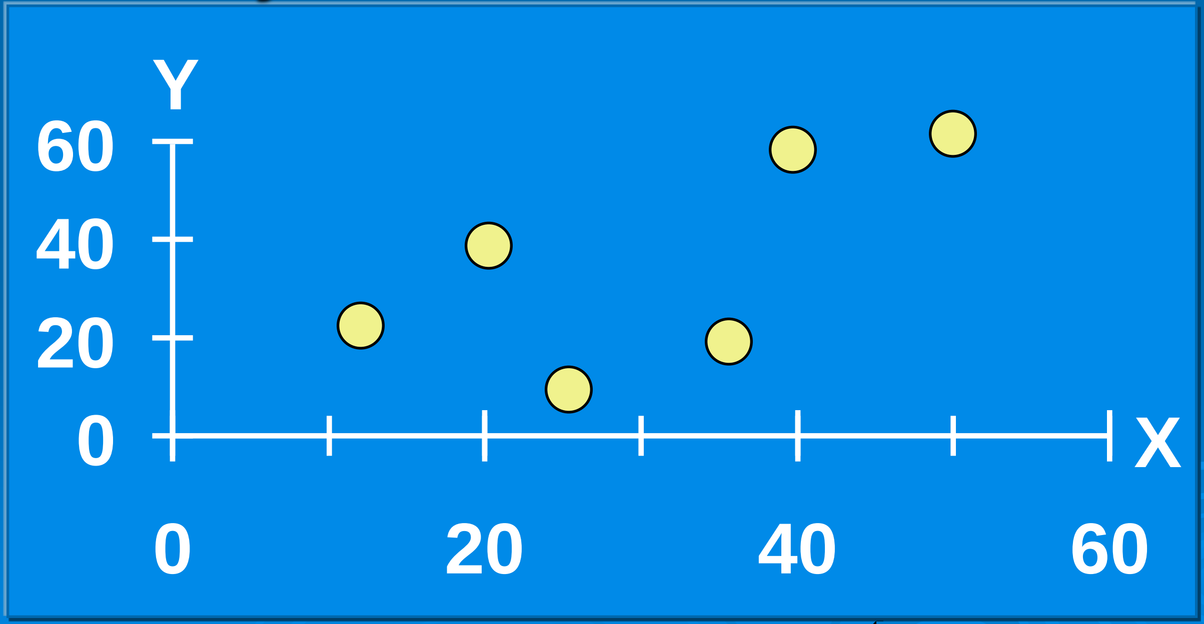
Linearna regresija u praksi

Ocenjivanje parametara: Metod najmanjih kvadrata

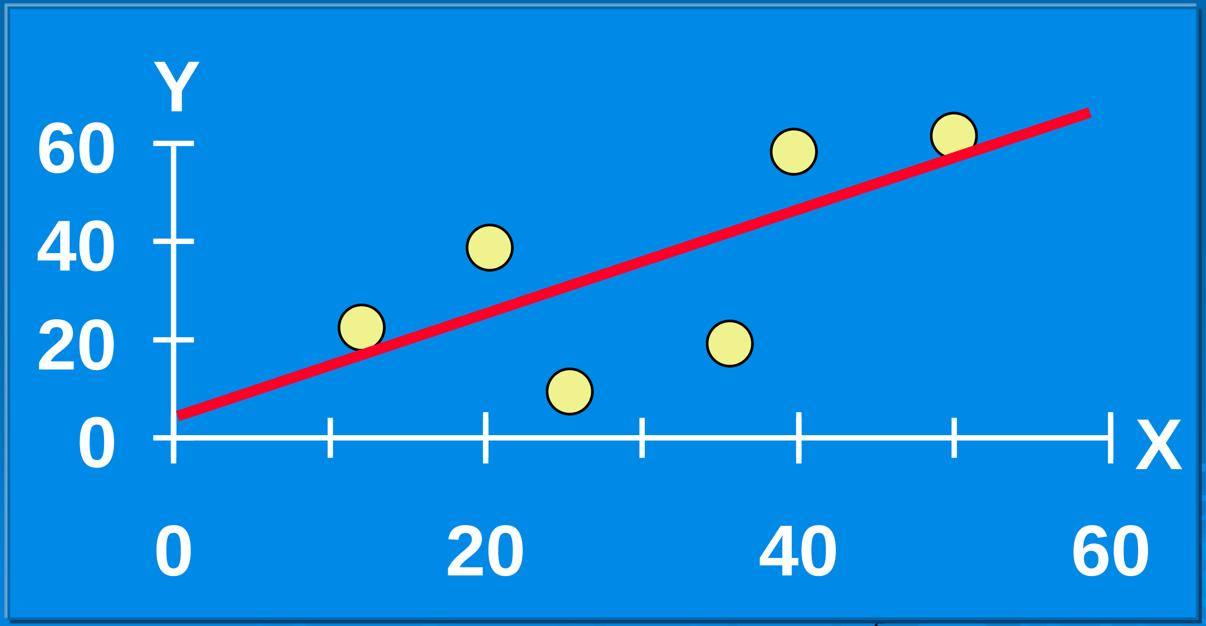
Dijagram rasturanja

- 1. Daje nam sliku svih (X_i, Y_i) parova
- 2. Pomaže nam da odaberemo optimalnu fit-funkciju



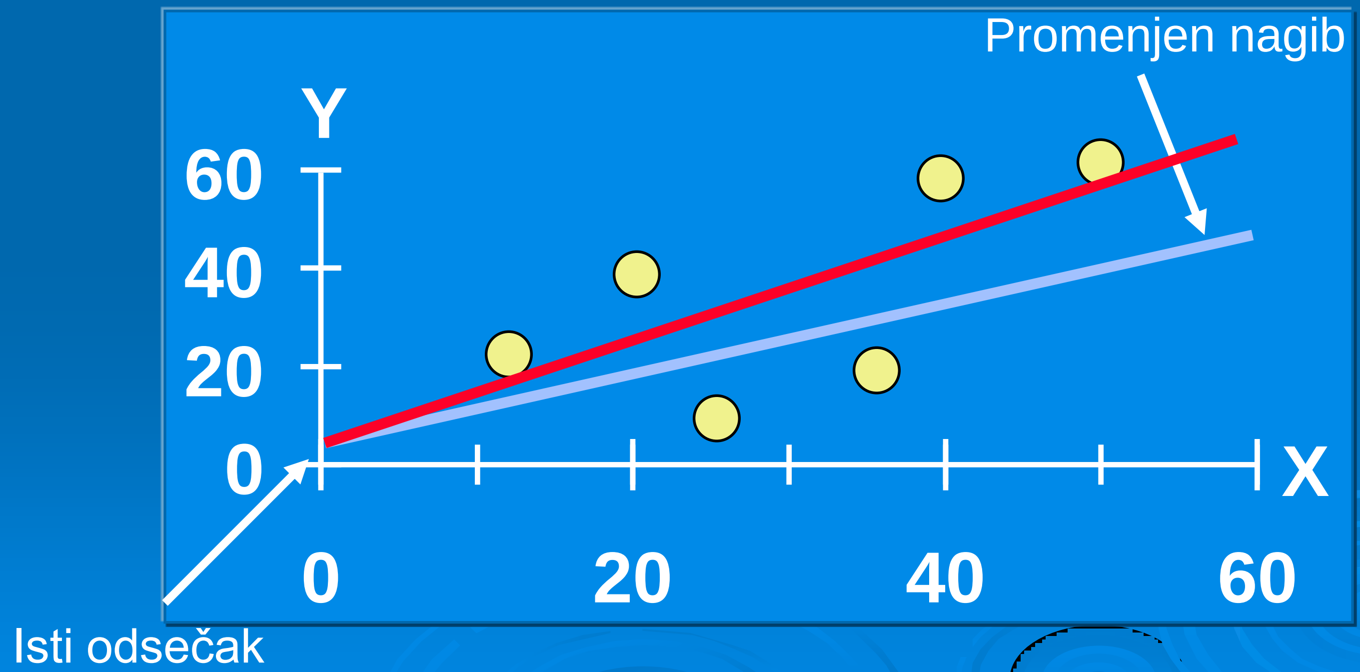
Izazov?

Kako biste nacrtali pravu liniju kroz date tačke? Kako biste odabrali onu koja najbolje „fituje“?



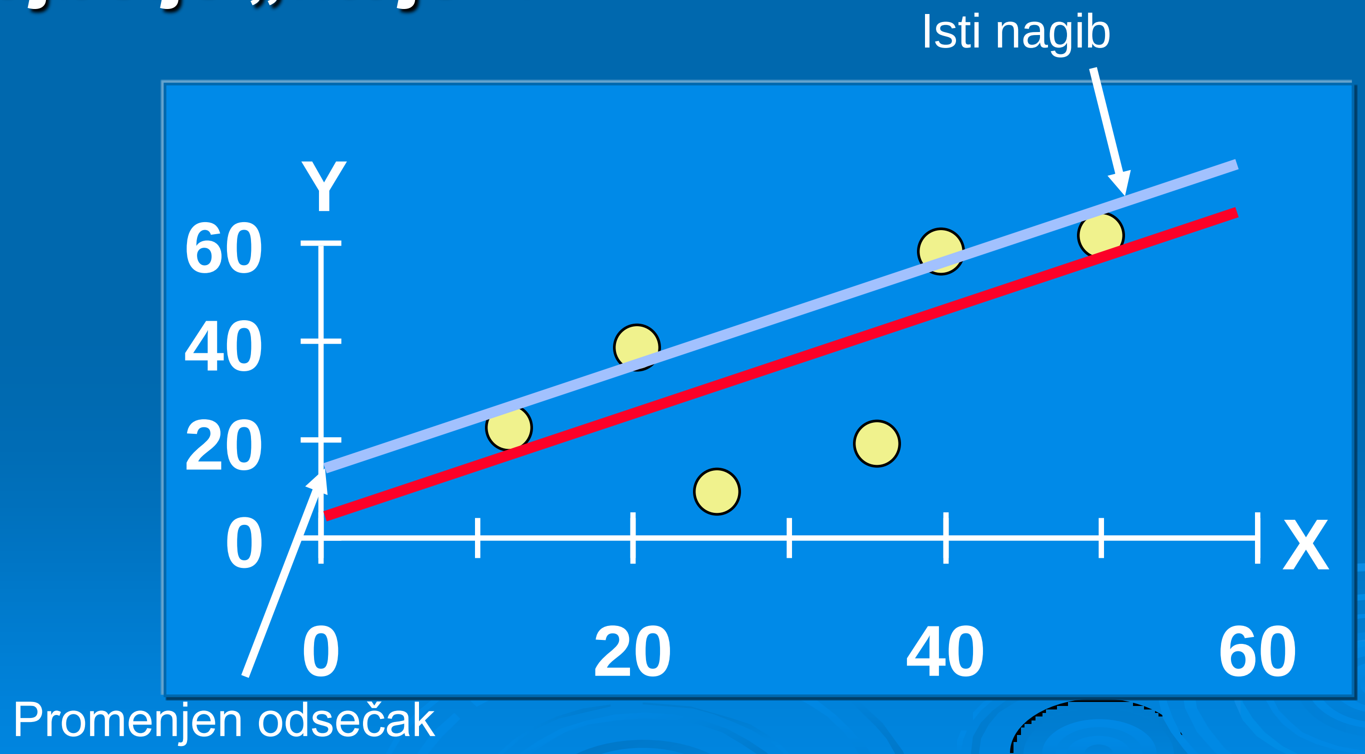
Izazov?

Kako biste nacrtali pravu liniju kroz date tačke? Kako biste odabrali onu koja najbolje „fituje“?



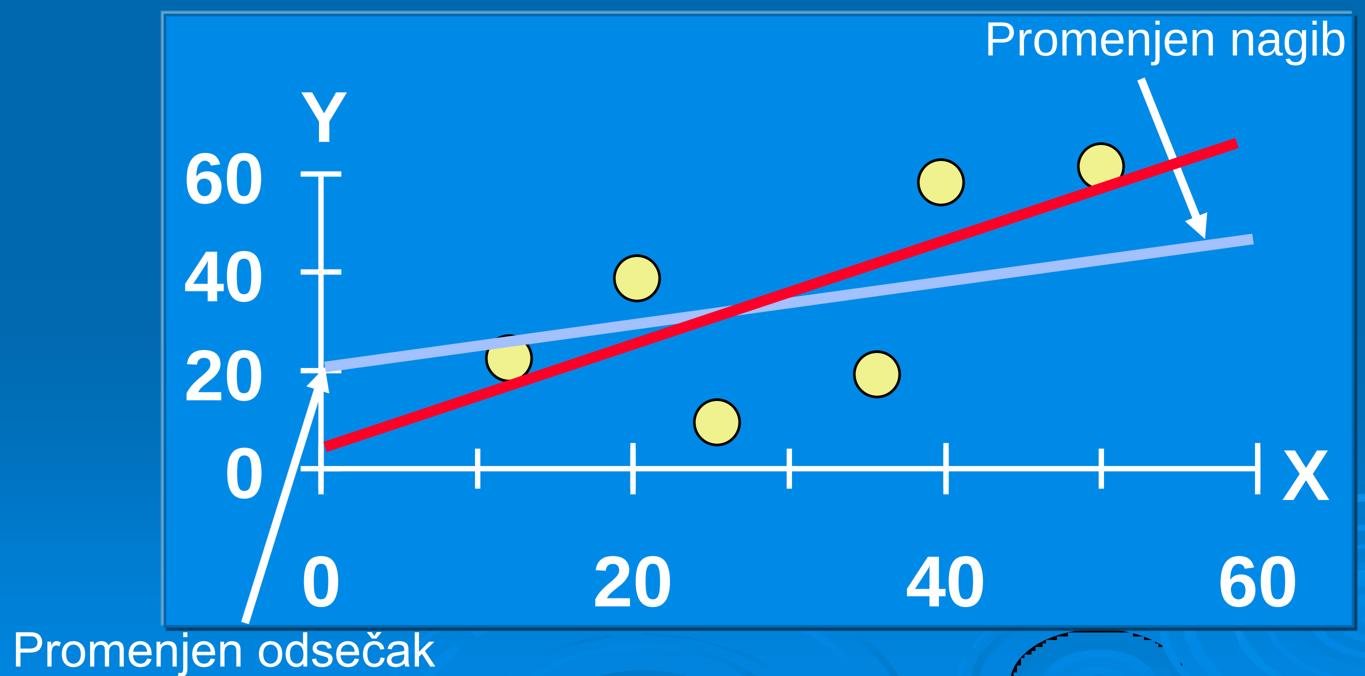
Izazov?

Kako biste nacrtali pravu liniju kroz date tačke? Kako biste odabrali onu koja najbolje „fituje“?



Izazov?

Kako biste nacrtali pravu liniju kroz date tačke? Kako biste odabrali onu koja najbolje „fituje“?



Najmanji kvadrati

- 1. „Najbolje fitovanje“ = razlika između stvarnih i prognoziranih Y vrednosti najmanja. Ali pozitivne razlike se često potiru sa negativnim.

Najmanji kvadrati

- 1. „Najbolje fitovanje“ = razlika između stvarnih i prognoziranih Y vrednosti najmanja. Ali pozitivne razlike se često potiru sa negativnim. Zato kvadriraj greške!

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

Najmanji kvadrati

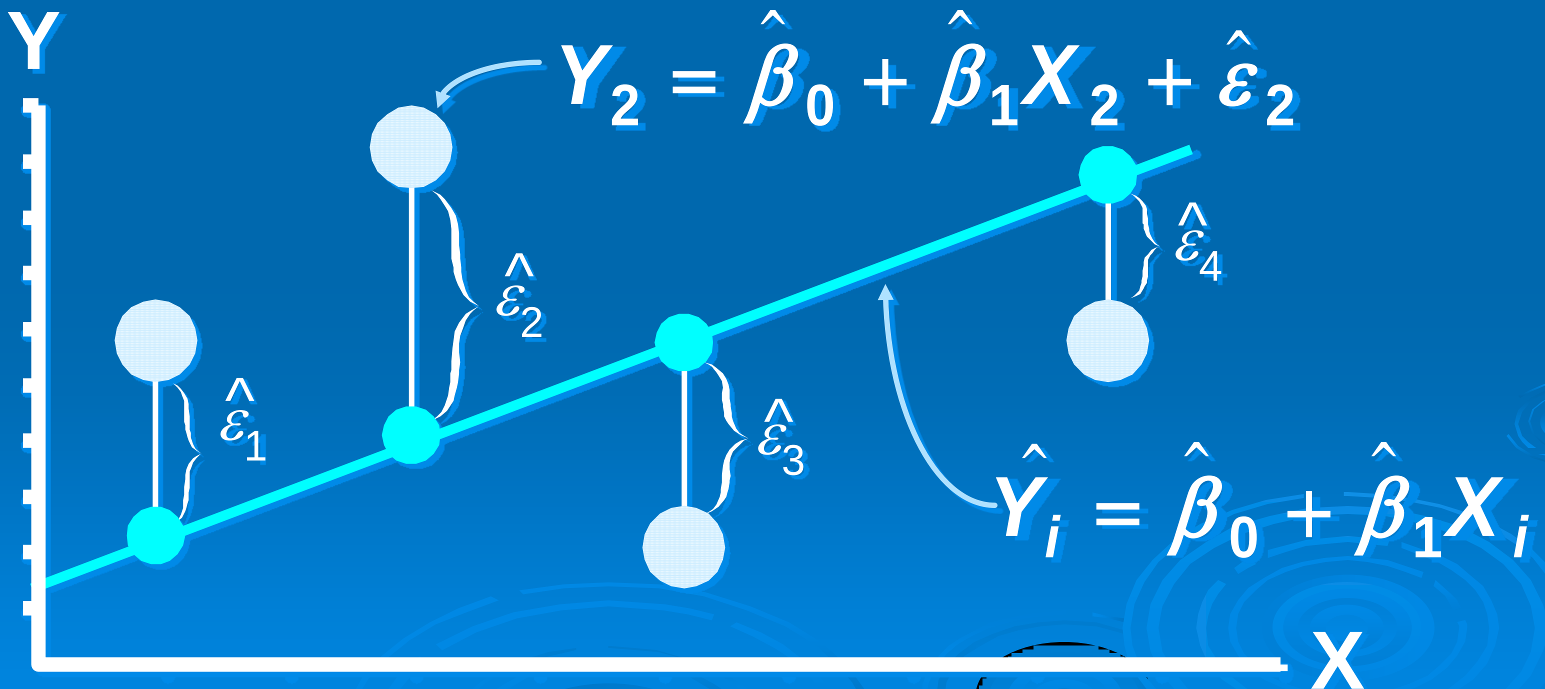
- 1. „Najbolje fitovanje“ = razlika između stvarnih i prognoziranih Y vrednosti najmanja. Ali pozitivne razlike se često potiru sa negativnim. Zato kvadriraj greške!

$$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

- 2. LS minimizira Sumu kvadrata (Squared) razlika (differences/Errors) (SSE)

Dijagram najmanjih kvadrata

LS minimizes $\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \hat{\varepsilon}_1^2 + \hat{\varepsilon}_2^2 + \hat{\varepsilon}_3^2 + \hat{\varepsilon}_4^2$



Jednačine parametara

- Prediction equation

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- Uzorački nagib

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

- Uzorački Y - odsečak

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Izvođenje parametara

- Najmanji kvadrati (L-S):

  Minimiziramo sumu kvadrata grešaka

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i \right)^2$$

$$0 = \frac{\partial \sum \varepsilon_i^2}{\partial \beta_0} = \frac{\partial \sum \left(y_i - \beta_0 - \beta_1 x_i \right)^2}{\partial \beta_0}$$

$$= -2\left(n\overline{y} - n\beta_0 - n\beta_1 \overline{x} \right)$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

Izvođenje parametara

- Najmanji kvadrati (L-S):

  Minimiziramo sumu kvadrata grešaka

$$0 = \frac{\partial \sum \varepsilon_i^2}{\partial \beta_1} = \frac{\partial \sum \left(y_i - \beta_0 - \beta_1 x_i \right)^2}{\partial \beta_1}$$

$$= -2 \sum x_i \left(y_i - \beta_0 - \beta_1 x_i \right)$$

$$= -2 \sum x_i \left(y_i - \overline{y} + \beta_1 \overline{x} - \beta_1 x_i \right)$$

$$\beta_1 \sum x_i \left(x_i - \overline{x} \right) = \sum x_i \left(y_i - \overline{y} \right)$$

$$\beta_1 \sum \left(x_i - \overline{x} \right)\left(x_i - \overline{x} \right) = \sum \left(x_i - \overline{x} \right)\left(y_i - \overline{y} \right)$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

Tablica međurezultata

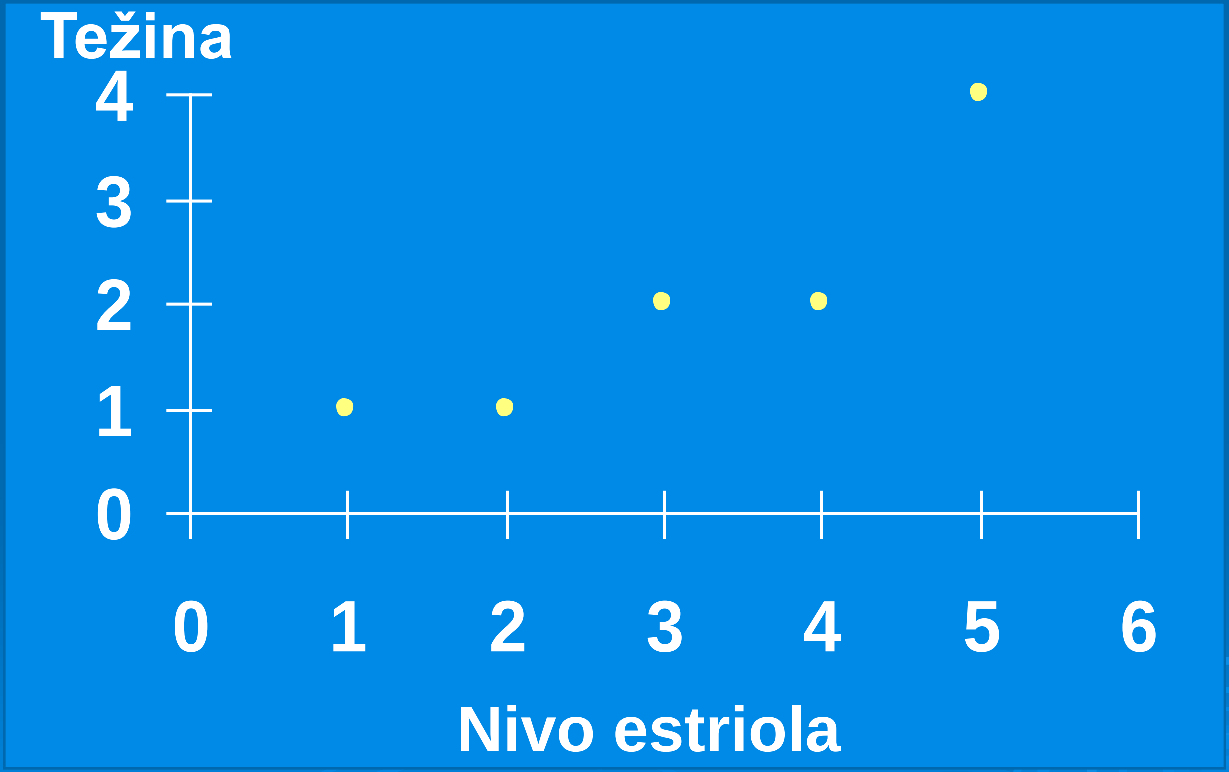
X_i	Y_i	X_i^2	Y_i^2	X_iY_i
X_1	Y_1	X_1^2	Y_1^2	X_1Y_1
X_2	Y_2	X_2^2	Y_2^2	X_2Y_2
:	:	:	:	:
X_n	Y_n	X_n^2	Y_n^2	X_nY_n
ΣX_i	ΣY_i	ΣX_i^2	ΣY_i^2	ΣX_iY_i

Primer linearne regresije

- Akušerstvo: Koja je veza između nivoa estriola majke i težine novorođene bebe, primenom sledećih podataka?

Estriol (mg/24h)	Težina (g/1000)
1	1
2	1
3	2
4	2
5	4

Dijagram rasturanja Težina x Nivo estriola



Tablica međurezultata

X_i	Y_i	X_i^2	Y_i^2	X_iY_i
1	1	1	1	1
2	1	4	1	2
3	2	9	4	6
4	2	16	4	8
5	4	25	16	20
15	10	55	26	37

Ocene parametara

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \frac{37 - \frac{(15)(10)}{5}}{55 - \frac{(15)^2}{5}} = 0.70$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 2 - (0.70)(3) = -0.10$$

Interpretacija rešenja

Interpretacija rešenja

- 1. Nagib ($\hat{\beta}_1$)
  - Očekuje se rast težine novorođenčeta (Y) za 0.7 jedinica za svaki jedinični porast nivoa Estriola (X).

Interpretacija rešenja

- 1. Nagib ($\hat{\beta}_1$)
  - Očekuje se rast težine novorođenčeta (*Y*) za 0.7 jedinica za svaki jedinični porast nivoa Estriola (*X*).
- 2. Odsečak ($\hat{\beta}_0$)
  - Prosečna težina (*Y*) je -0.10 jedinica kada je nivo Estriola (*X*) jednak 0.
    - Ovo je teško objasniti
    - Težina novorođene bebe je uvek pozitivna

SAS kôd za fitovanje podataka prostom linearnom regresijom

- **Data** BW; /*Reading data in SAS*/
- input estriol birthw@@;
- cards;
- 1 1 2 1 3 2
  4 2 5 4
- ;
- **run**;

- **PROC REG data=BW**; /*Fitting linear regression models*/
- model birthw=estriol;
- **run**;

SAS izlaz prethodnog programa

Parameter Estimates

<table>
<tr><th>Variable</th><th>DF</th><th>Parameter Estimate</th><th>Standard Error</th><th>t Value</th><th>Pr > |t|</th></tr>
<tr><td>Intercept</td><td>1</td><td>-0.10000</td><td>0.63509</td><td>-0.16</td><td>0.8849</td></tr>
<tr><td>Estriol</td><td>1</td><td>0.70000</td><td>0.19149</td><td>3.66</td><td>0.0354</td></tr>
</table>

$\hat{\beta}_0$

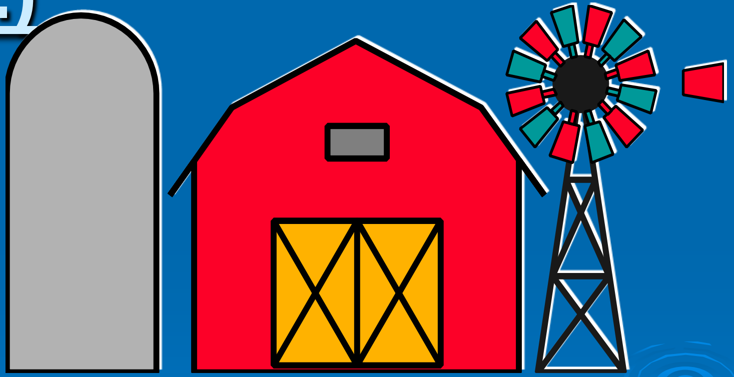
$\hat{\beta}_1$

Test (ne)značajnosti koeficijenta

Primer 2

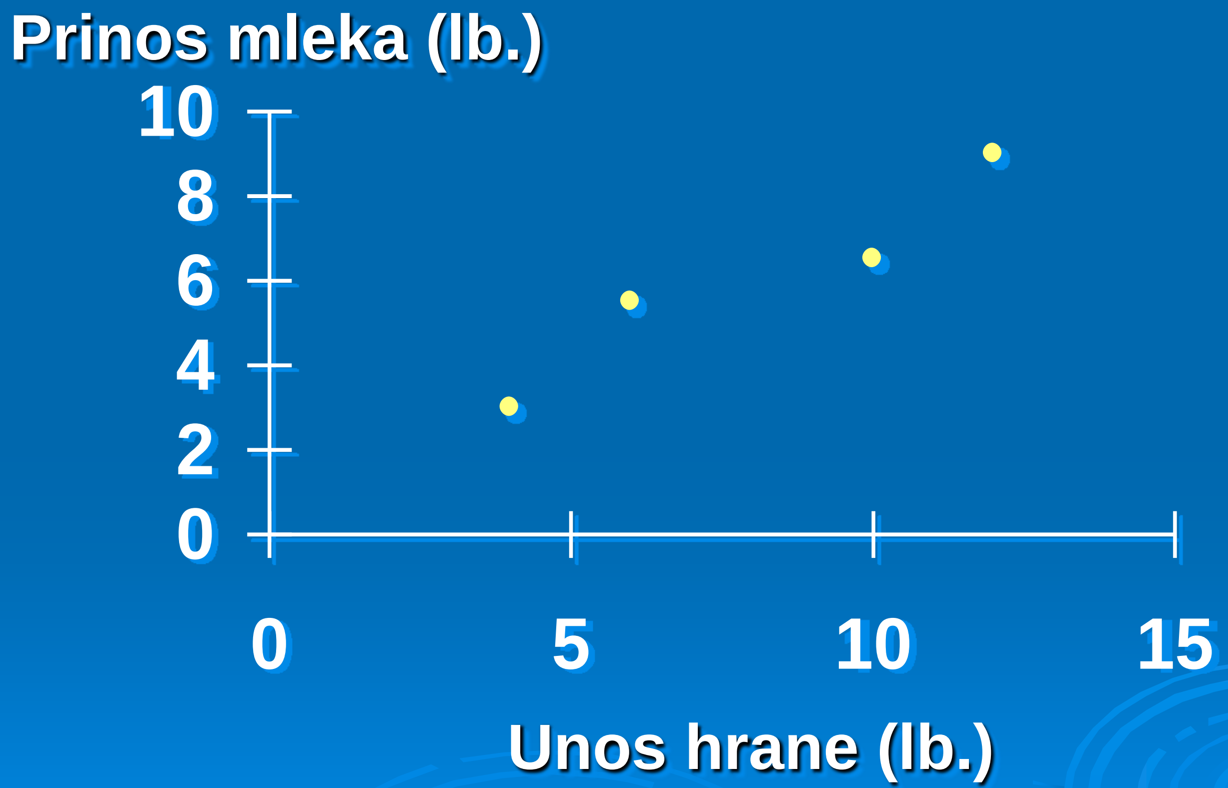
- Vi ste okružni veterinar-epidemiolog i dobili ste sa farmi sledeće podatke :
- **Hrana (lb.) Prinos mleka (lb.)**

Hrana (lb.)	Prinos mleka (lb.)
4	3.0
6	5.5
10	6.5
12	9.0



- Koja je **veza** izmedju unosa hrane i prinosa mleka krava?

Dijagram rasturanja



Tablica međurezultata

X_i	Y_i	X_i^2	Y_i^2	X_iY_i
4	3.0	16	9.00	12
6	5.5	36	30.25	33
10	6.5	100	42.25	65
12	9.0	144	81.00	108
32	24.0	296	162.50	218

Ocenjivanje parametara

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}} = \frac{218 - \frac{(32)(24)}{4}}{296 - \frac{(32)^2}{4}} = 0.65$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 6 - (0.65)(8) = 0.80$$

Interpretacija rešenja

Interpretacija rešenja

- 1. Nagib ($\hat{\beta}_1$)
  - Očekuje se da se prinos mleka (Y) uvećava za 0.65 lb, za svaki jedinični porast unosa hrane (X).

Interpretacija rešenja

- 1. Nagib ($\hat{\beta}_1$)
  - Očekuje se da se prinos mleka (Y) uvećava za 0.65 lb, za svaki jedinični porast unosa hrane (X).
- 2. Odsečak ($\hat{\beta}_0$)
  - Očekivani prosečni prinos mleka (*Y*) 0.8 lb. Kada je unos hrane (*X*) jednak 0.