# Statistika: Dvodimenzionalne varijable

# Motivacija – Primjer 9.1 (6.2)

| Gustoća (kg/m$^3$) | Tlačna čvrstoća (MPa) |
|---|---|
| 2236 | 25.2 |
| 2244 | 25.2 |
| 2244 | 25.2 |
| 2244 | 25.3 |
| 2244 | 25.7 |
| 2253 | 25.4 |
| 2253 | 25.8 |
| 2262 | 25.4 |
| 2272 | 27.5 |
| 2280 | 27.5 |
| 2290 | 27.1 |

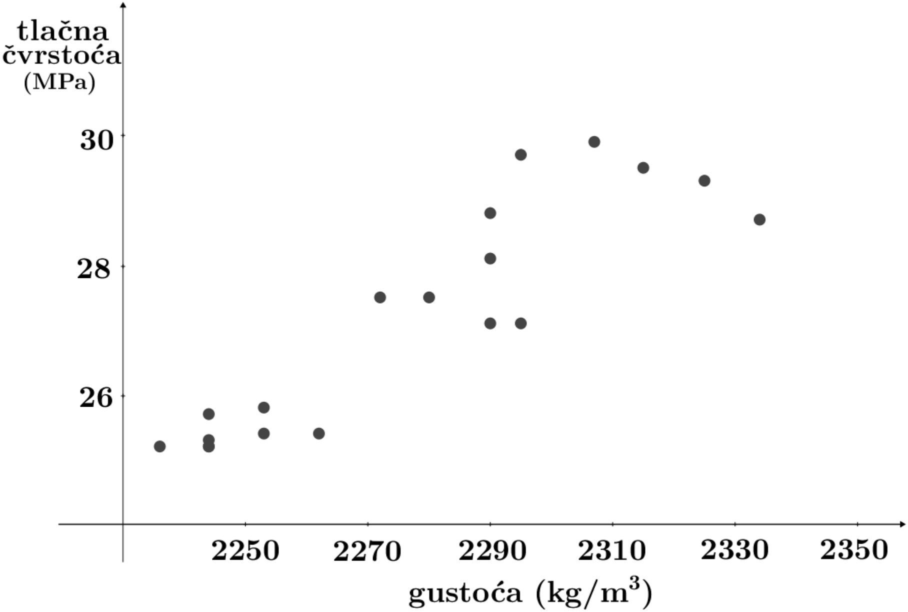
# Primjer 9.1 (Primjer 6.2)

| Gustoća (kg/m$^3$) | Tlačna čvrstoća (MPa) |
|---|---|
| 2290 | 28.1 |
| 2290 | 28.8 |
| 2295 | 27.1 |
| 2295 | 29.7 |
| 2307 | 29.9 |
| 2315 | 29.5 |
| 2325 | 29.3 |
| 2334 | 28.7 |

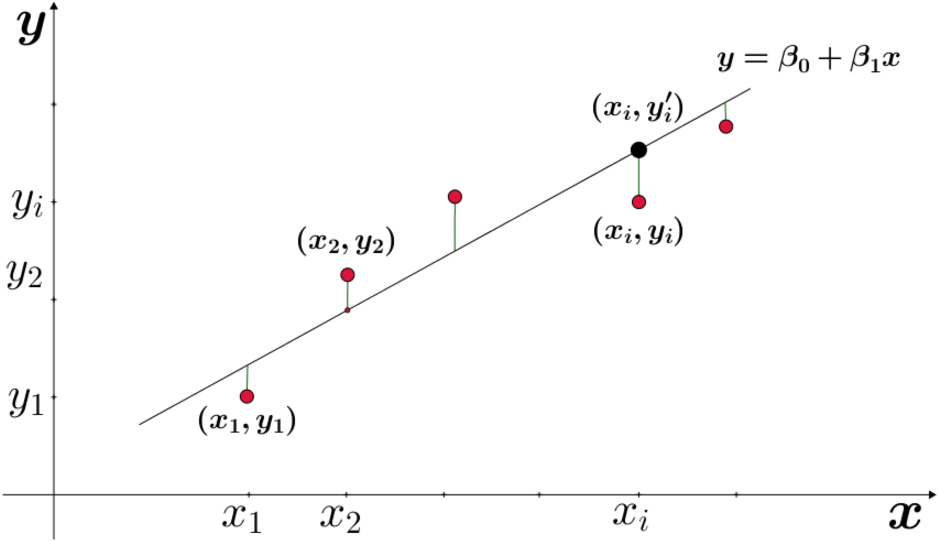
Tablica: Podatci o gustoći i tlačnoj čvrstoći betona [1]

[1] A. Mattaacchione i L. Mattacchione, "Correlation Between 28-Day Strength and Density", Concrete International, 1995 (3), str. 37-41

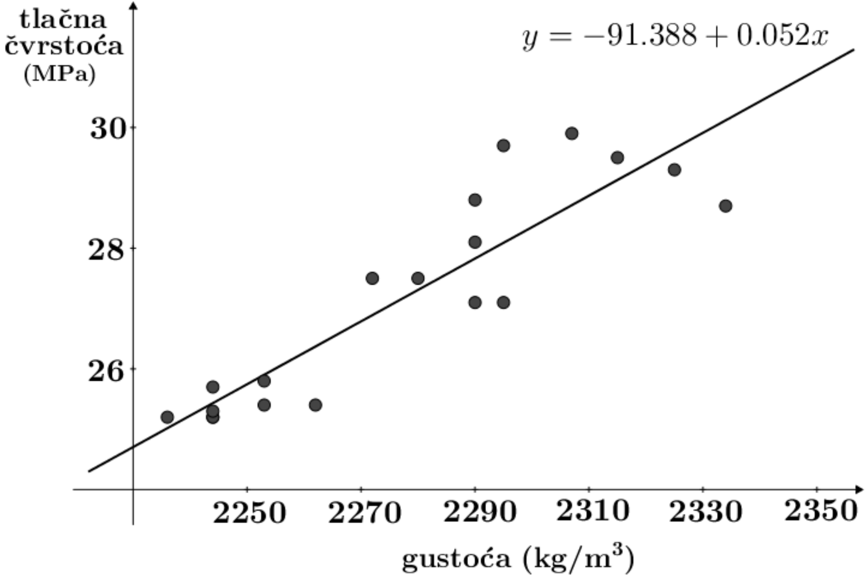
# Dijagram raspršenja (podataka) tlačne čvrstoće i gustoće



# Metoda najmanjih kvadrata



# Pravac najboljeg pristajanja



# Pearsonov koeficijent korelacije

- jedan od načina mjerenja zavisnosti dviju varijabli
- za opažene vrijednosti $(x_1, y_1), \ldots, (x_n, y_n)$ od $(X, Y)$ definiramo ga kao
$$r_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{YY}}},$$
gdje su
$$S_{XX} = \sum_{i=1}^{n}(x_i - \overline{x}_n)^2, \quad S_{YY} = \sum_{i=1}^{n}(y_i - \overline{y}_n)^2, \quad S_{XY} = \sum_{i=1}^{n}(x_i - \overline{x}_n)(y_i - \overline{y}_n).$$
- vrijedi
$$-1 \leq r_{XY} \leq 1$$

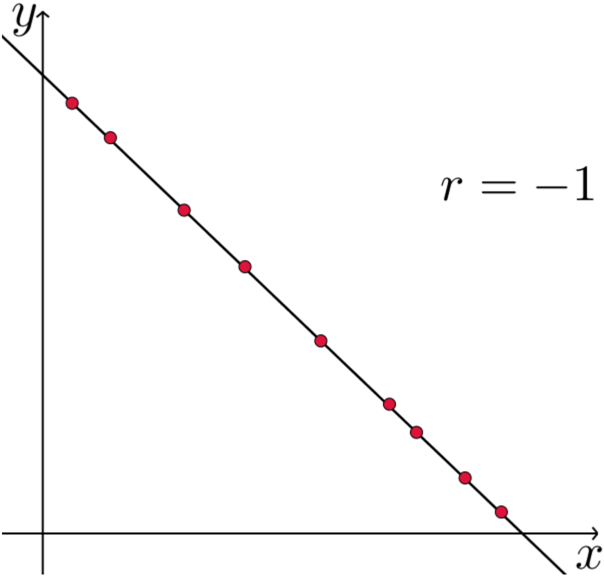
# Pearsonov koeficijent korelacije

(i) $|r_{XY}| = 1$ ako i samo ako opažanja $(x_1, y_1), \ldots, (x_n, y_n)$ leže na istom pravcu (i $S_{XX} > 0$, $S_{YY} > 0$)

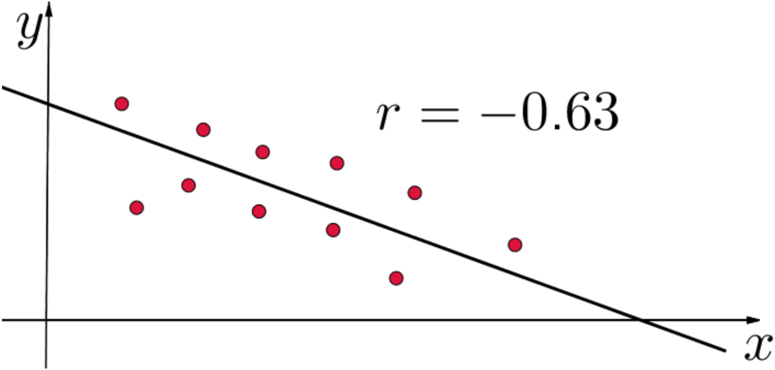
(ii) manji $|r_{XY}|$ znači "manji stupanj linearne zavisnosti" od opažanja $x_1, \ldots, x_n$ i $y_1, \ldots, y_n$

(iii) $r_{XY} = 0$ znači da su $x - \overline{x}_n$ i $y - \overline{y}_n$ "ortogonalni"

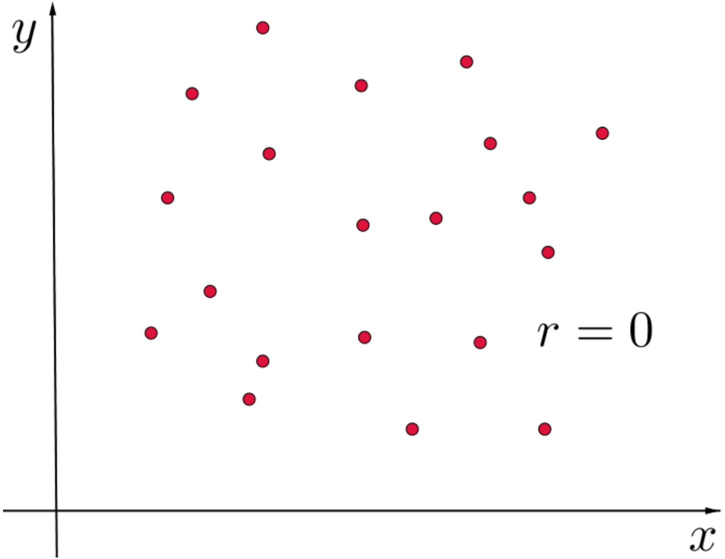
# Ilustracija podataka s različitim koeficijentima korelacije



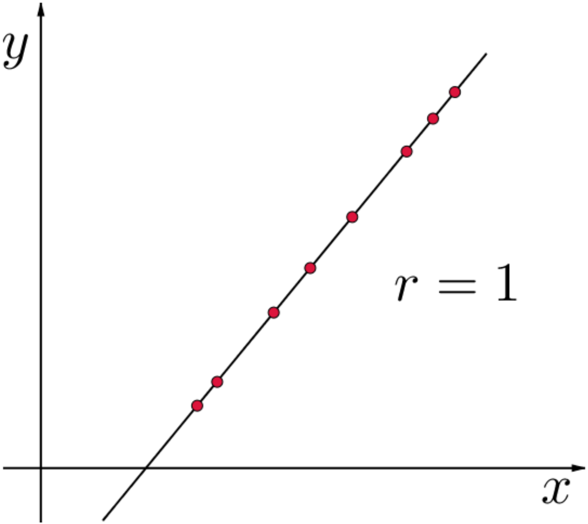
# Ilustracija podataka s različitim koeficijentima korelacije



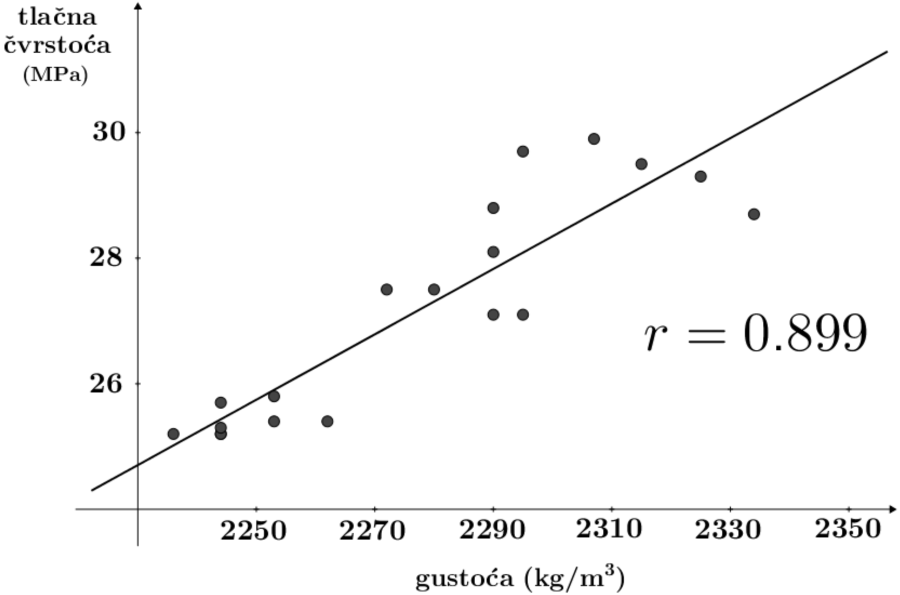
# Ilustracija podataka s različitim koeficijentima korelacije



# Ilustracija podataka s različitim koeficijentima korelacije



# Koeficijent korelacije tlačne čvrstoće i gustoće



# Jednostavna linearna regresija

- pretpostavlja povezanost između zavisne varijable $Y$ i jedne nezavisne varijable $X$
- pretpostavljamo linearni model (linearnu povezanost)

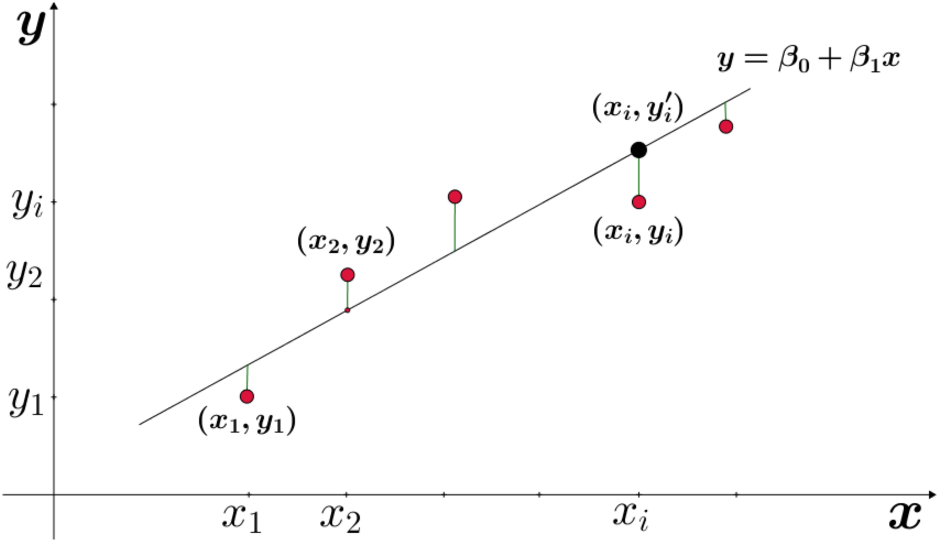
$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

gdje varijable $\varepsilon_1, \ldots, \varepsilon_n$ predstavljaju pogreške (nezavisne i jednako distribuirane s raspodjelom $N(0, \sigma^2)$)

# Jednostavna linearna regresija

- koeficijenti $\beta_0$ i $\beta_1$ su nepoznati **regresijski parametri** koje u postupku modeliranja procijenimo na osnovi opažanja
- $y_1, \ldots, y_n$ su stvarne izmjerene vrijednosti
- $y_i' = \beta_0 + \beta_1 x_i$ vrijednosti koje bismo očekivali za određeni $x_i$ na temelju postavljenog modela
- da bismo dobili pravac koji bi najbolje pristajao svim podacima mora se minimizirati udaljenost između $y_i$ i $y_i'$.

# Metoda najmanjih kvadrata



# Metoda najmanjih kvadrata

- procjena $\hat{\beta}_0$, $\hat{\beta}_1$ parametara $\beta_0$, $\beta_1$:

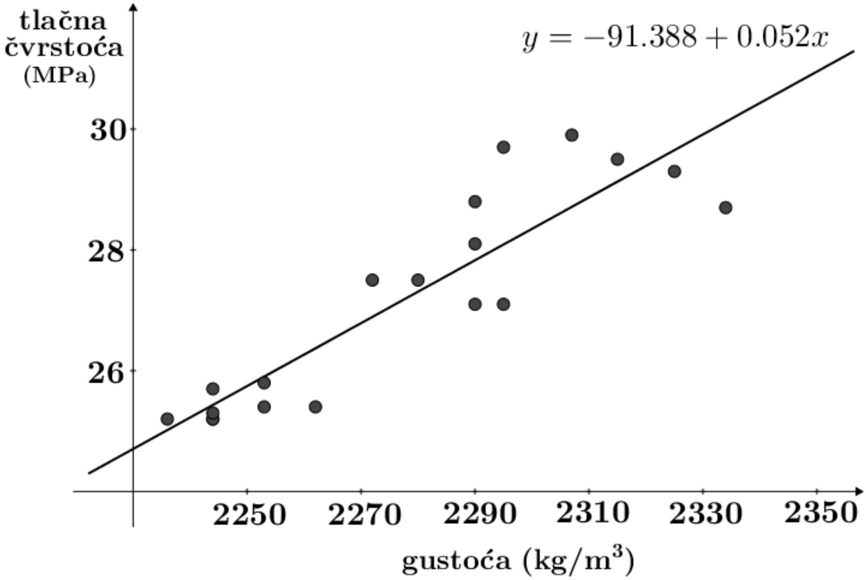
$$\hat{\beta}_0 = \overline{y}_n - \overline{x}_n \hat{\beta}_1, \quad \hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}$$

- u svrhu određivanja reprezentativnosti modela najčešće koristimo **koeficijent determinacije**:

$$r_{XY}^2 = \frac{S_{XY}^2}{S_{XX} S_{YY}}, \qquad 0 \leq r_{XY}^2 \leq 1$$

- što je $r_{XY}^2$ bliži 1, to je model reprezentativniji

# Pravac najboljeg pristajanja (tlačna čvrstoća i gustoća)



# Anscombeov kvartet - I

| $x$ | $y$ |
|---|---|
| 10 | 8.04 |
| 8 | 6.95 |
| 13 | 7.58 |
| 9 | 8.81 |
| 11 | 8.33 |
| 14 | 9.96 |
| 6 | 7.24 |
| 4 | 4.26 |
| 12 | 10.84 |
| 7 | 4.82 |
| 5 | 5.68 |

$$\overline{x}_{11} = 9, s_x^2 = 11, \overline{y}_{11} = 7.5, s_y^2 = 4.127, r_{XY} = 0.8164, y = 3 + 0.5x$$

# Anscombeov kvartet - II

| $x$ | $y$ |
|---|---|
| 10 | 9.14 |
| 8 | 8.14 |
| 13 | 8.74 |
| 9 | 8.77 |
| 11 | 9.26 |
| 14 | 8.10 |
| 6 | 6.13 |
| 4 | 3.1 |
| 12 | 9.13 |
| 7 | 7.26 |
| 5 | 4.74 |

$$\overline{x}_{11} = 9, s_x^2 = 11, \overline{y}_{11} = 7.5, s_y^2 = 4.128, r_{XY} = 0.8162, y = 3 + 0.5x$$

# Anscombeov kvartet - III

| $x$ | $y$ |
|---|---|
| 10 | 7.46 |
| 8 | 6.77 |
| 13 | 12.74 |
| 9 | 7.11 |
| 11 | 7.81 |
| 14 | 8.84 |
| 6 | 6.08 |
| 4 | 5.39 |
| 12 | 8.15 |
| 7 | 6.42 |
| 5 | 5.73 |

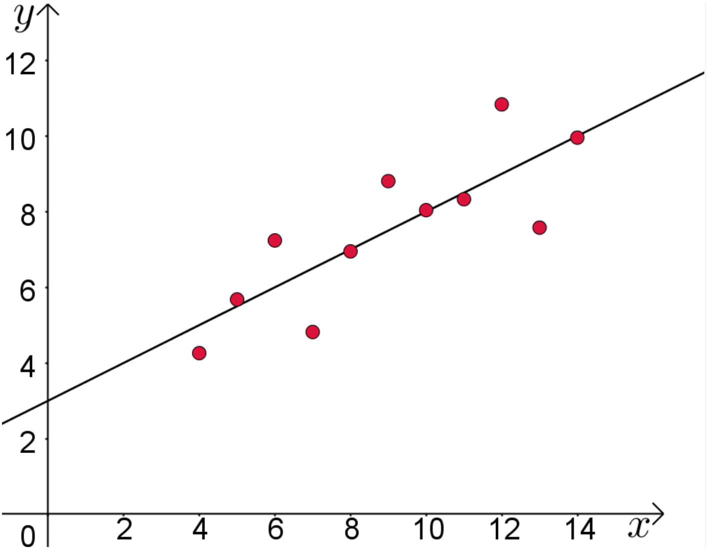
$$\overline{x}_{11} = 9, s_x^2 = 11, \overline{y}_{11} = 7.5, s_y^2 = 4.123, r_{XY} = 0.8163, y = 3 + 0.5x$$

# Anscombeov kvartet - IV

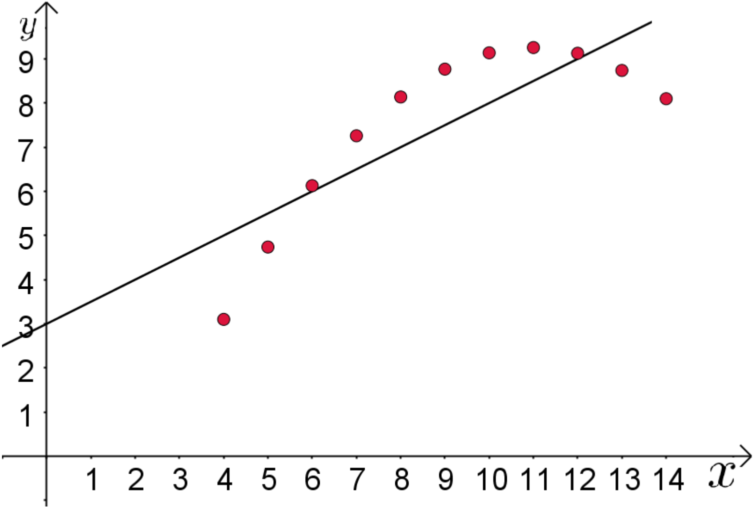
| $x$ | $y$ |
|---|---|
| 8 | 6.58 |
| 8 | 5.76 |
| 8 | 7.71 |
| 8 | 8.84 |
| 8 | 8.47 |
| 8 | 7.04 |
| 8 | 5.26 |
| 19 | 12.5 |
| 8 | 5.56 |
| 8 | 7.91 |
| 8 | 6.89 |

$$\overline{x}_{11} = 9, s_x^2 = 11, \overline{y}_{11} = 7.5, s_y^2 = 4.123, r_{XY} = 0.8165, y = 3 + 0.5x$$

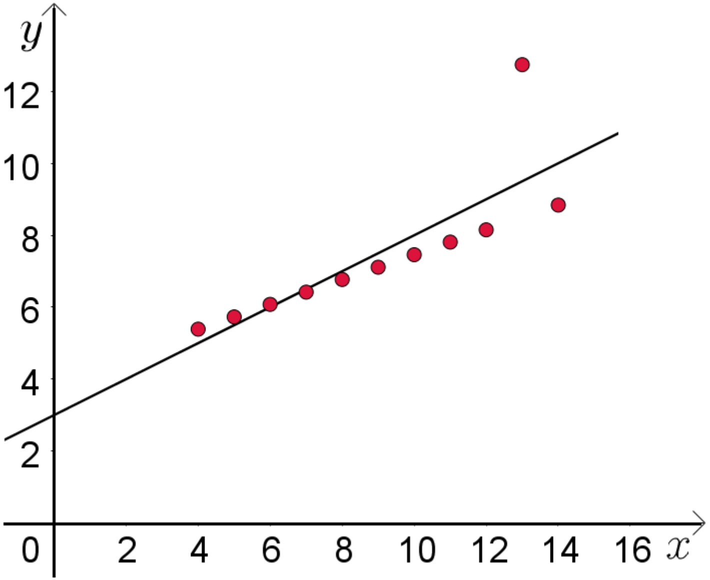
# Anscombeov kvartet - I



# Anscombeov kvartet - II



# Anscombeov kvartet - III



# Anscombeov kvartet - IV