

# K čemu jsou metody stanovení závislosti

- závislosti 1. **kvantitativního** znaku na 2. **kvantitativním** znaku (nebo více kvantitativních znacích) se zabývá **regresní a korelační analýza**
- závislost dvou znaků - **jednoduchá regrese (jednoduchá korelace)**
- závislost znaku na více znacích - **vícenásobná regrese**
- znalost závislosti umožňuje **předvídat chování** (prognózovat, predikovat) závislé veličiny

# Metoda nejmenších čtverců MNČ

Idea MNČ: minimalizovat reziduální součet čtverců:

$$S_R = \sum_{i=1}^n (y_i - Y_i)^2 = \sum_{i=1}^n (y_i - \underbrace{(b_0 + b_1 x_i)}_{Y_i})^2$$

**Příklad 1:**

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{462,1 - 14 \cdot 25,8}{230 - 14^2} = \frac{100,9}{34} = 2,97$$

$$b_0 = \bar{y} - b_1 \bar{x} = 25,8 - 2,97 \cdot 14 = -15,78$$

Regresní funkce:  $Y = -15,78 + 2,97x$

# Příklad 1: „Ruční výpočty“

$i$	$x_i$	$y_i$	$x_i^2$	$x_i y_i$	$Y_i$	$(Y_i - \bar{y})^2$	$(y_i - \bar{y})^2$
1	6	5	36	30	2,04	565,21	432,64
2	8	8	64	64	7,98	318,22	316,84
3	9	9	81	81	10,95	221,15	282,24
4	9	12	81	108	10,95	221,15	190,44
5	12	21	144	252	19,86	35,62	23,04
6	15	25	225	375	28,77	8,61	0,64
7	16	32	256	512	31,74	34,84	38,44
8	20	36	400	720	43,62	315,88	104,04
9	22	51	484	1122	49,56	562,08	635,04
10	23	59	529	1357	52,53	711,60	1102,24
<b>Součet</b>	140	258	2300	4621	258	2994,3	3125,6
<b>Průměr</b>	14	25,8	230	462,1			

# Koeficient determinace $R^2$

**Koeficient determinace** charakterizuje přiléhavost dat k regresnímu modelu (číslo mezi 0 a 1):

$$R^2 = \frac{S_T}{S_y}$$

$$S_T = \sum_{i=1}^n (Y_i - \bar{y})^2 - \text{teoretický součet čtverců: } S_y = S_R + S_T$$

$$S_R = \sum_{i=1}^n (y_i - Y_i)^2 - \text{reziduální součet čtverců}$$

$$\text{Pro malé soubory: } R_{adj}^2 = 1 - (1 - R^2) \frac{n-1}{n-2}$$

# Korelační analýza (KA)

- V KA není předem známo, které jsou vysvětlující a které vysvětlované proměnné!

**Příklad:** Závislost tržeb za zboží  $X$  na tržbách zboží  $Y$

Oboustranný vztah - 2 regresní přímky:

$$y = \alpha_0 + \alpha_1 x + \varepsilon_1 \quad x = \beta_0 + \beta_1 y + \varepsilon_2$$

Korelační koeficient:  $\rho = \pm \sqrt{|\alpha_1 \beta_1|}$

Odhad  $\rho$  :

$$r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]}}$$

## Příklad 2. Výsledky testů 10 studentů 1. ročníku OPF

Počet bodů z matematiky	56	79	50	84	63	91	46	56	74	76
Počet bodů z ekonomie	82	56	46	79	74	83	51	63	75	82

$$r = \frac{10 \cdot 47823 - 675 \cdot 691}{\sqrt{(10 \cdot 47687 - 675^2)(10 \cdot 49501 - 691^2)}} = 0,6112$$

$r > 0,6$  – vysoká hodnota korelace!