

# Regresní a korelační analýza

# Regrese a korelace

**Regrese** charakterizuje **průběh** závislosti mezi kvantitativními statistickými znaky pomocí matematického modelu (regresní funkce).

**Korelace** měří **těsnost** (sílu, míru, intenzitu) statistické závislosti mezi kvantitativními statistickými znaky pomocí koeficientů.

# Druhy závislostí

## Podle počtu kvantitativních znaků

- o závislost jednoduchá
- o závislost vícenásobná

# Druhy závislostí

## Podle typu regresní funkce

- lineární závislost
- nelineární závislost

## Podle směru změn kvantitat. znaků

- závislost **pozitivní** (kladná, přímá)
- závislost **negativní** (záporná, nepřímá)

# Regresní analýza

## Základní úkoly regresní analýzy

- získání statistických odhadů neznámých parametrů regresní funkce na základě výběru
- testování hypotéz o těchto parametrech
- ověřování předpokladů regresního modelu

# Základní model

## Základní model regresní závislosti

$$y'_i = f(x_i) + e_i,$$

$f(x_i)$  ...je regresní funkce,

$e_i$  ...jsou náhodné (reziduální) chyby (odchylky)

# Výběr regresní funkce

Logické posouzení daného vztahu

Vycházíme z grafické analýzy dat

Využití matematicko-statistický kritérií

# Jednoduchá lineární regrese

# Jednoduchá lineární regrese

## Model regresní přímky

$$y_i = \alpha + \beta x_i + e_i \quad i = 1, 2, \dots, n$$

$X$  .....nezávisle proměnná (vysvětlující, regresor)

$Y$  .....závisle proměnná (vysvětlovaná)

$\alpha, \beta$  .. neznámé parametry modelu v ZS

$e_i$  ..... náhodná chyba (reziduum; chyba predikce)  
odchylka naměřené hodnoty od hodnoty  
předpovídané vyrovnávací křivkou.

# Jednoduchá lineární regrese

**Metoda nejmenších čtverců** vychází z požadavku, aby součet čtverců odchylek pozorovaných hodnot (součet druhých mocnin reziduálních hodnot) byl minimální.

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n y_i - \underbrace{(a + bx_i)} = \min$$

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

# Jednoduchá lineární regrese

**Jednostranná závislost** – proměnná  $X$  je nezávisle proměnná a  $Y$  pak závisle proměnná.

$a_{yx}$  ... absolutní člen

$b_{yx}$  ... regresní koeficient

$y'_i$  ... vyrovnaná (teoretická) hodnota vysvětlované prom.

$$y'_i = a_{yx} + b_{yx} \cdot x_i$$

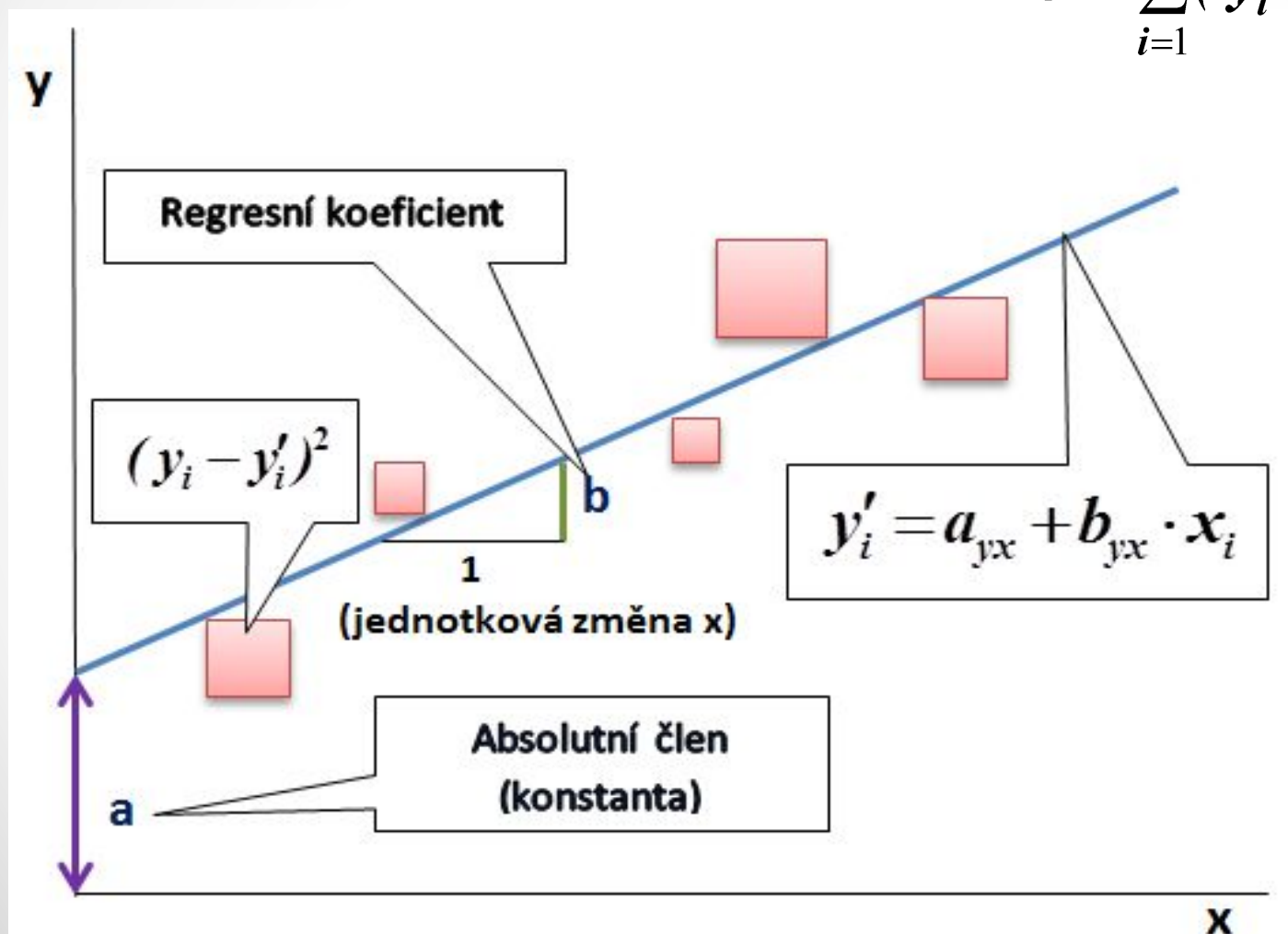
**Oboustranná závislost** – nelze rozhodnout, která proměnná je závislá a která nezávislá (sdružené fce.).

$$y'_i = a_{yx} + b_{yx} \cdot x_i$$

$$x'_i = a_{xy} + b_{xy} \cdot y_i$$

# Jednoduchá lineární regrese

$$MN\check{C} = \sum_{i=1}^n (y_i - y'_i)^2 = \min$$



# Korelační analýza

**Korelace** obecně označuje míru stupně (sílu) závislosti dvou proměnných  $X$  a  $Y$ .

**Měření těsnosti (síly)** závislosti - spočívá ve zjištění, jak těsně se jednotlivé skutečné napozorované hodnoty přimykají k regresní čáře, která vystihuje průběh závislosti.

# Pearsonův koeficient korelace

$$r_{yx} = r_{xy}$$

Platí  $\langle -1 \leq r \leq +1 \rangle$  dvě náhodné proměnné jsou tím více korelovány, čím blíže je hodnota korelačního koeficientu číslům  $+1$  nebo  $-1$ .

# Korelační analýza

$$0 \leq r^2 \leq 1$$

**Koeficient determinace  $r^2_{yx}$**  je druhou mocninou koeficientu korelace.

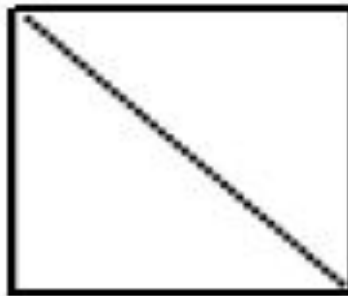
|                          |                      |
|--------------------------|----------------------|
| $r^2 < 10 \%$            | těsnost nízká        |
| $10 \% \leq r^2 < 25 \%$ | těsnost mírná        |
| $25 \% \leq r^2 < 50 \%$ | těsnost význačná     |
| $50 \% \leq r^2 < 80 \%$ | těsnost velká        |
| $80 \% \leq r^2$         | těsnost velmi vysoká |

# Korelační analýza

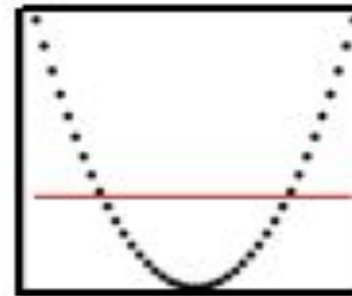
## Proložení regresní přímky korelačním polem



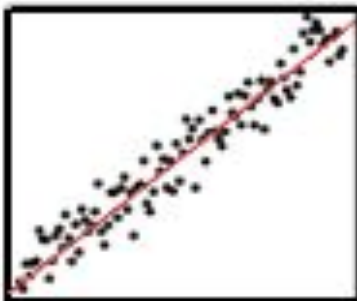
$$r = 1,000$$



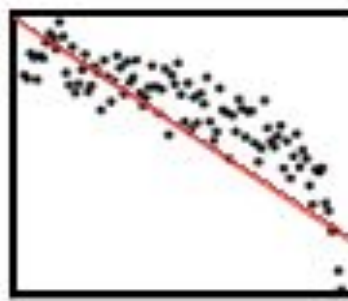
$$r = -1,000$$



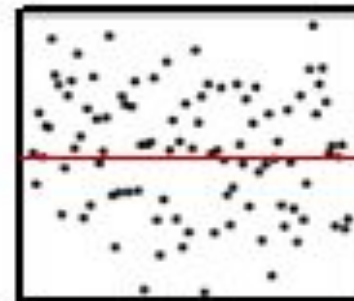
$$r = 0,000$$



$$r = 0,957$$



$$r = -0,824$$



$$r = -0,145$$

# Spearmanův koeficient pořadí

Spearmanův koeficient korelace  $r_s$  nabývá hodnot z intervalu  $(-1 \leq r_s \leq 1)$ .

# Souhrnný příklad

**Měření výšky a váhy u studentů druhého ročníku oboru PAA (datová matice v moodle)**

$X$  – výška;  $Y$  - váha

Vytvoříme graf – korelační pole

Vypočítáme rovnici regresní přímky

Určíme sílu závislosti