Korelační analýza

- zabývá se především intenzitou vzájemného vztahu proměnných;
- je základní metodou měření síly lineární závislosti číselných proměnných;
- „correlatió“ = vzájemná souvislost (z lat.).

! Z výpočetních a interpretačních hledisek se regresní a korelační analýza prolínají, nelze mezi nimi stanovit ostrou hranici.

Sdružené regresní přímky

$Y = a_{yx} + b_{yx} x$ popisuje závislost *y* na *x*

$X = a_{xy} + b_{xy} y$ popisuje závislost *x* na *y*

$a_{yx} = \bar{y} - b_{yx}\bar{x}$ $\qquad b_{yx} = \dfrac{s_{xy}}{s_x^2}$

$a_{xy} = \bar{x} - b_{xy}\bar{y}$ $\qquad b_{xy} = \dfrac{s_{xy}}{s_y^2}$

1. $b_{yx} = b_{xy} = 0 \quad \Rightarrow$ *x* a *y* jsou korelačně nezávislé; sdružené regresní přímky svírají pravý úhel.
2. $b_{yx} = \dfrac{1}{b_{xy}} \quad \Rightarrow$ *x* a *y* jsou úplně závislé; sdružené regresní přímky svírají nulový úhel (splývají).

Míry těsnosti lineární závislosti

Koeficient determinace: $r_{yx}^2 = r_{xy}^2 = b_{yx} \cdot b_{xy} = \dfrac{s_{xy}}{s_x^2} \cdot \dfrac{s_{xy}}{s_y^2} = \dfrac{s_{xy}^2}{s_x^2 \cdot s_y^2}$ **;** $r_{xy}^2 \in \langle 0;1 \rangle$

Koeficient korelace: $r_{yx} = r_{xy} = \sqrt{r_{yx}^2} = \dfrac{s_{xy}}{s_x \cdot s_y} = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\left(\overline{x^2} - \bar{x}^2\right) \cdot \left(\overline{y^2} - \bar{y}^2\right)}}$ **;** $r_{xy} \in \langle -1;1 \rangle$

- měří sílu lineární závislosti, nikoli závislosti obecně (lineární závislost = korelovanost).
- koeficient je symetrický.

Interpretace:

1. znaménko +/– udává směr závislosti:

   $r_{xy} > 0 \Rightarrow$ přímá závislost

   $r_{xy} < 0 \Rightarrow$ nepřímá závislost

2. $|r_{xy}|$ udává sílu závislosti:

$r_{xy} = 0 \quad \Rightarrow$ lineární nezávislost

$|r_{xy}| = 1 \quad \Rightarrow$ funkční (úplná) závislost

$|r_{xy}| \to 0 \quad \Rightarrow$ slabá lineární závislost

$|r_{xy}| \to 1 \quad \Rightarrow$ silná lineární závislost

Test hypotézy o nulové hodnotě korelačního koeficientu

1) $H_0 : \rho_{yx} = 0$ (lineární nezávislost *x* a *y*)

$H_1 : \ non \ H_0$

2) ***Testové kritérium:***

$$t = \frac{r_{yx} \cdot \sqrt{n-2}}{\sqrt{1-r_{yx}^2}}$$; Statistika **t** má při platnosti H_0 rozdělení t s (n-2) stupni volnosti .

3) ***Kritický obor:***

$$W \equiv \left\{ t;\ t \le t_{\frac{\alpha}{2}}(n-2) \cup t \ge t_{1-\frac{\alpha}{2}}(n-2) \right\}$$

4) ***Závěr testu:***

Pokud leží hodnota testového kritéria v kritickém oboru, zamítáme H_0 a přijímáme H_1, tzn. prokázali jsme hypotézu o lineární závislosti proměnných *x* a *y*.

Pořadová korelace

- Pokud chceme získat rychlou představu o síle závislosti mezi 2 kvantitativními znaky nebo určit závislost mezi pořadími znaků, nahradíme původní hodnoty x_i a y_i jejich pořadovými čísly i_x a i_y podle toho, která místa hodnoty zaujímají v uspořádané řadě.

Spearmanův koeficient pořadové korelace: $r_s = 1 - \dfrac{6\sum_{i=1}^{n}(i_x - i_y)^2}{n(n^2-1)}$; $r_s \in \langle -1;1 \rangle$

- varianta korelačního koeficientu;
- měří sílu lineární závislosti dvou pořadí.

Interpretace: stejná jako u korelačního koeficientu.

Test hypotézy o nezávislosti pořadovou korelací

1) $H_0 : \rho_S = 0$ (nezávislost pořadí)

$H_1 : non\ H_0$

2) ***Testové kritérium:***

$$t = \frac{r_S \cdot \sqrt{n-2}}{\sqrt{1-r_S^2}}$$; Statistika **t** má při platnosti H_0 rozdělení t s (n-2) stupni volnosti .

3) ***Kritický obor:***

$$W \equiv \left\{ t;\ t \le t_{\frac{\alpha}{2}}(n-2) \cup t \ge t_{1-\frac{\alpha}{2}}(n-2) \right\}$$

4) Závěr testu:

Pokud leží hodnota testového kritéria v kritickém oboru, zamítáme H_0 a přijímáme H_1, tzn. prokázali jsme hypotézu o lineární závislosti obou pořadí.