

12 Korelační analýza

- PŘEDPOKLAD: DVOUROZMĚRNÁ NORMALITA DAT → Grafické ověření: **Tečkový graf** + elipsa spolehlivosti

12.1 Pořadová nezávislost

- Dvourozměrný náhodný výběr $(X_1, Y_1), \dots, (X_n, Y_n)$
- R_i ... pořadí hodnot náhodné veličiny X
- Q_i ... pořadí hodnot náhodné veličiny Y
- Spearmanův koeficient pořadové korelace R_S

$$R_S = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i - Q_i)^2$$

- $R_S \in \langle -1; 1 \rangle$
- `cor(X, Y, method='spearman')`

12.1.1 Testy o pořadové nezávislosti

- H_0 : X a Y jsou pořadově nezávislé. ... $H_0 : \rho = 0$
- H_1 : X a Y nejsou pořadově nezávislé. ... $H_1 : \rho \neq 0$

1. Exaktní test ($n \leq 30$)

- Testovací statistika $T_0 = |R_S|$
- Kritický obor: $W = \langle k; 1 \rangle \rightarrow$ tabulky

2. Asymptotický test ($n > 20$)

- Testovací statistika $T_0 = \frac{R_S \sqrt{n-2}}{\sqrt{1-R_S^2}}$
- Kritický obor: $W = (-\infty; t_{\alpha/2}(n-2)) \cup (t_{1-\alpha/2}(n-2); \infty)$
– $t_{\alpha/2}(n-2) \dots \text{qt}(\text{alpha}/2, n-2)$
- p -hodnota = $2 \min\{\Pr(T_0 \leq t_0), \Pr(T_0 > t_0)\}$
– $2 * \min(\text{pt}(t_0, n-2), 1-\text{pt}(t_0, n-2))$

3. Asymptotický test ($n > 30$)

- Testovací statistika $T_0 = R_S \sqrt{n-1}$
- Kritický obor: $W = (-\infty; u_{\alpha/2}) \cup (u_{1-\alpha/2}; \infty)$
– $u_{\alpha/2} \dots \text{qnorm}(\text{alpha}/2)$
- p -hodnota = $2 \min\{\Pr(T_0 \leq t_0), \Pr(T_0 > t_0)\}$
– $2 * \min(\text{pnorm}(t_0), 1-\text{pnorm}(t_0))$

12.2 Lineární nezávislost

- Výběrový (Pearsonův) koeficient korelace R_{12}

$$R_{12} = \frac{S_{12}}{S_1 S_2},$$

kde S_{12} je kovariance, S_1 je směrodatná odchylka výběru X , S_2 je sm. odchylka výběru Y

- $R_{12} \in \langle -1; 1 \rangle$
- `cor(X, Y, method='pearson')`

12.2.1 Testy o lineární nezávislosti

- H_0 : X a Y jsou lineárně nezávislé. ... H_0 : $\rho = 0$
- H_1 : X a Y nejsou lineárně nezávislé. ... H_1 : $\rho \neq 0$

1. Exaktní přístup

- Testovací statistika

$$T_0 = \frac{R_{12}}{\sqrt{1 - R_{12}^2}} \sqrt{n - 2}$$

- Kritický obor: $W = (-\infty; t_{\alpha/2}(n - 2)) \cup (t_{1-\alpha/2}(n - 2); \infty)$.
– $t_{\alpha/2}(n - 2)$... `qt(alpha/2, n-2)`
- Interval spolehlivosti

$$(dh, hh) = \left(\frac{t_{\alpha/2}(n - 2)}{\sqrt{t_{\alpha/2}^2(n - 2) + n - 2}}; \frac{t_{1-\alpha/2}(n - 2)}{\sqrt{t_{1-\alpha/2}^2(n - 2) + n - 2}} \right)$$

- **!!! H_0 zamítáme, pokud $R_{12} \notin IS$!!!**
- p -hodnota = $2 \min\{\Pr(T_0 \leq t_0), \Pr(T_0 > t_0)\}$
– $2 * \min(\text{pt}(t_0, n-2), 1 - \text{pt}(t_0, n-2))$

2. Asymptotický přístup

- Z -transformace $Z = \frac{1}{2} \ln \frac{1 + R_{12}}{1 - R_{12}}$
- Interval spolehlivosti

$$\text{tgh} \left(Z - \frac{u_{1-\alpha/2}}{\sqrt{n - 3}}; Z + \frac{u_{1-\alpha/2}}{\sqrt{n - 3}} \right)$$

- **!!! H_0 zamítáme, pokud $\rho = 0 \notin IS$!!!**

12.3 Test o dvou korelačních koeficientech

- dva nezávislé náhodné výběry o rozsazích n a n^* s korelačními koeficienty ρ a ρ^*
- $H_0 : \rho = \rho^*$
- $H_1 : \rho \neq \rho^*$

- Z -transformace $Z = \frac{1}{2} \ln \frac{1 + R_{12}}{1 - R_{12}}$

- Z -transformace $Z^* = \frac{1}{2} \ln \frac{1 + R_{12}^*}{1 - R_{12}^*}$

- testovací statistika

$$U = \frac{Z - Z^*}{\sqrt{\frac{1}{n-3} + \frac{1}{n^*-3}}}$$

- kritický obor $W = (-\infty; u_{\alpha/2}) \cup (u_{1-\alpha/2}; \infty)$

– $u_{\alpha/2} \dots \text{qnorm}(\alpha/2)$

12.4 Interpretační tabulka hodnot Spearmanova a Pearsonova korelačního koeficientu

Abs.hod. korel.koef.	Interpretace hodnoty
0	pořadová (lineární) nezávislost
(0; 0.1)	velmi nízký stupeň závislosti
[0.1; 0.3)	nízký stupeň závislosti
[0.30; 0.50)	mírný stupeň závislosti
[0.50; 0.70)	význačný stupeň závislosti
[0.70; 0.90)	vysoký stupeň závislosti
[0.90; 1)	velmi vysoký stupeň závislosti
1	úplná pořadová (lineární) závislost