ZÁKLADNÍ PŘEDPOKLADY O DATECH

Celá řada statistických postupů předpokládá, že analyzovaná data splňují určité předpoklady. Nejsou-li tyto předpoklady splněny, vede použití „klasických“ statistických postupů ke zkresleným výsledkům a analýza těchto souborů dat je složitá.

ZÁKLADNÍ PŘEDPOKLADY:

- Minimální velikost výběru
- Nezávislost
- Normalita (data pocházejí z Gaussova rozdělení)
- Homogenita výběru (nepřítomnost OB)

Velikost výběru

- Rozsah výběru n ovlivňuje přesnost odhadů parametrů polohy a rozptýlení ⇒ projeví se při konstrukci IS.
- U velmi malých výběrů může být výsledek (šířka IS, závěr testování hypotézy) více ovlivněn velikostí výběru než skutečnými daty (jejich variabilitou).
- Je-li potřeba ověřit minimální velikost výběru, pro normální rozdělení použijeme:

$$n_{\min} = \left(\frac{t_{(1-\alpha/2;\, n-1)}}{d} \right)^2 \cdot s^2$$

μ-d ≤ průměr ≤ μ+d

Nezávislost

- Závislost může být způsobena např. časovými změnami v měřícím procesu, nekonstantností podmínek, zanedbáním některých faktorů, nenáhodným výběrem vzorků.
- Závislost musí být proměřována před uspořádáním naměřených dat.
- Závislá data indikují přítomnost systematické chyby.
- Obvykle se ověřuje testováním významnosti autokorelačního koeficientu nějakým statistickým testem.

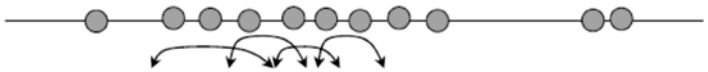
H_0: $\rho_A = 0$ H_1: $\rho_A \neq 0$

Lineární závislost prvků jednoho souboru - **AUTOKORELACE**

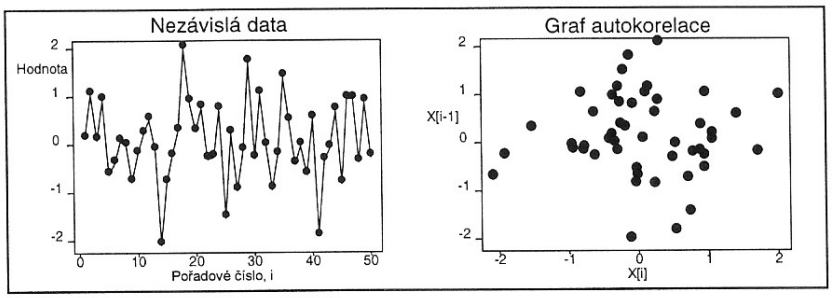
$$x_i = \rho_k \, x_{i-k} + e_i$$

ρ_k autokorelační koeficient k-tého řádu

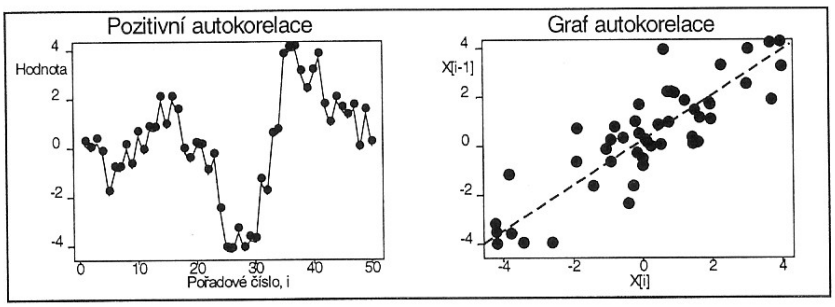
autokorelace I. řádu
sousední hodnoty



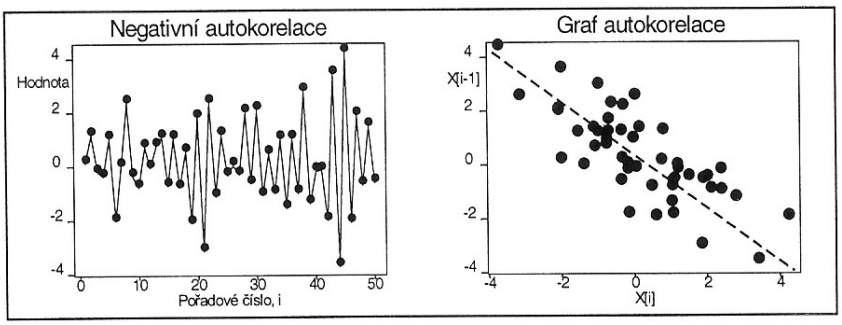
autokorelace II. řádu
hodnoty „přes jednu“



Obr. 3-10 A, B Nezávislá data, koeficient autokorelace $\rho_1 \approx 0$



Obr. 3-11 A, B Autokorelovaná data, $\rho_1 = +0.8$



Obr. 3-12 A, B Autokorelovaná data, $\rho = -0.8$

Ověření normality

- Normalita souboru dat patří k základním předpokladům, protože je na ní založena celá „klasická“ statistická analýza dat.
- Statistické testy jsou obecně méně citlivé než diagnostické grafy EDA.
- Pokud není normalita dat prokázána, je data třeba hlouběji analyzovat.
- ***Test kombinace výběrové šikmosti a špičatosti***

$$\chi^2_{\text{exp}} = \frac{g_1^2}{D(g_1)} + \frac{[g_2 - E(g_2)]^2}{D(g_2)}$$

Vypočtené χ^2_{exp} srovnáváme s $\chi^2_{\text{krit}}(1-\alpha;\ 2)$. Je-li $\chi^2_{\text{exp}} > \chi^2_{\text{krit}}$, předpoklad normality se zamítá.

- ***Shapiro-Wilkův test*** – podle ČSN 010225.
- ***D´Agostinův test*** – posuzuje výběrové momenty dat; považuje se za velmi citlivý i na malé odchylky od normality.
- ***Kolmogorov-Smirnovův test*** – založen na porovnání rozdílu teoretické a výběrové distribuční funkce.

Homogenita výběru

- Nehomogenita bývá způsobena přítomností OB, tj. hodnot, které se co do velikosti výrazně liší od ostatních a lze je běžně identifikovat v grafech EDA.
- Identifikace ***metodou modifikovaných vnitřních hradeb***:

$$B_D^* = \tilde{x}_{0,25} - K(\tilde{x}_{0,75} - \tilde{x}_{0,25}) \qquad B_H^* = \tilde{x}_{0,75} + K(\tilde{x}_{0,75} - \tilde{x}_{0,25})$$

kde K = 2,25 – 3,6/n

Body ležící mimo modifikované vnitřní hradby jsou OB.

- Řada dalších statistických testů: Dean-Dixonův, Grubbsův, …