

Využití exploratorní analýzy při kontrole kvality

Milan Meloun

Univerzita Pardubice, 532 10 Pardubice

***Souhrn:** Účelem průzkumové analýzy dat (EDA), jako prvního kroku v analýze jednorozměrných dat, je odhalení jejich statistických zvláštností pomocí 12 grafických diagnostik, ověření základních předpokladů a případná mocninná transformace. Při průzkumové analýze složitějších, nákladných nebo unikátních měření je totiž nutné posoudit zvláštnosti chování dat ještě před vlastní, rutinní statistickou analýzou. Jedině tak lze zabránit provádění numerických výpočtů bez hlubších statistických souvislostí. EDA je pak obzvláště důležitá u aplikace regulačních diagramů při kontrole kvality.*

1. Postup analýzy dat při kontrole kvality

V **prvním kroku** se v průzkumové (exploratorní) analýze dat vyšetřují *statistické zvláštnosti*, jako je lokální koncentrace dat, tvarové zvláštnosti rozdělení dat a přítomnosti podezřelých hodnot. Odhalí se tak anomálie a odchylky rozdělení výběru od předpokládaného rozdělení Gaussova. Interaktivní statistická analýza na počítači tento postup ulehčuje, většina statistického software totiž nabízí řadu diagnostických grafů a diagramů.

V **druhém kroku** se ověří *základní předpoklady*, kladené na výběr, jako jsou nezávislost prvků, homogenita výběru, dostatečný rozsah výběru a rozdělení výběru. Jsou-li závěry tohoto kroku optimistické, následuje vyčíslení klasických odhadů polohy a rozptýlení, tj. aritmetického průměru a rozptylu ve kroku čtvrtém. Sem patří i konstrukce intervalů spolehlivosti a případně testování hypotéz. V opačném případě následuje pokus o třetí krok, obsahující symetrizující transformaci dat.

Ve **třetím kroku** se provádí mocninná a Boxova-Coxova transformace, které mohou vést k symetričtějším rozdělení výběru a umožňují provedení korektnějšího odhadu polohy a rozptýlení. Transformace je vhodná především při nekontantnosti rozptylu a při asymetrii rozdělení původních dat.

Ve **čtvrtém kroku** se v konfirmatorní analýze nabízí paleta rozličných odhadů polohy, rozptýlení a tvaru, jež lze rozdělit do skupin: *klasické odhady* a *robustní odhady* (necitlivé na odlehlé prvky výběru, resp. další předpoklady o datech). Z nabídky odhadů parametrů vybírá uživatel uvážlivě ty, jež mají statistický smysl a odpovídají závěrům průzkumové analýzy dat a závěrům ověření předpokladů o výběru.

Postup statistické analýzy jednorozměrných dat lze shrnout do čtyř kroků:

A. Průzkumová (exploratorní) analýza dat (EDA)

1. Zkoumání zvláštností dat:

Odhalení stupně symetrie a špičatosti rozdělení;
Indikace lokální koncentrace dat a rozdělení výběru;
Nalezení vybočujících a podezřelých prvků ve výběru.

2. Ověření předpokladů o datech:

Ověření normality rozdělení; Ověření nezávislosti prvků výběru;
Ověření homogenity rozdělení výběru; Určení minimálního rozsahu výběru.

3. Transformace dat:

Mocninná transformace; Box-Coxova transformace.

B. Konfirmatorní analýza dat (CDA) - odhady parametrů

(polohy, rozptýlení a tvaru)

1. Klasické odhady (bodové a intervalové): - momentové;
2. Robustní odhady (bodové a intervalové): - kvantilové,
- uřezané,
- winsorizované.