

3.4 Počítačová analýza vícerozměrných dat

Prof. RNDr. Milan Meloun, DrSc.

Obsah předmětu: Praktické procvičování analýzy kategoriálních dat a analýzy vícerozměrných dat. Kategoriální data rozlišují tři základní typy proměnných: kvantitativní (číselné) proměnné, ordinální (pořadové) proměnné a nominální proměnné. Kontingenční tabulky umožňují výpočet statistiky 2 a dalších měř závislosti pro dvourozměrnou tabulku. Metoda hlavních komponent a faktorová analýza patří mezi metody skrytých (latentních) vztahů. Kanonická a diskriminační analýza jsou metody, zabývající se závislostmi mezi proměnnými. Shluková analýza a zobrazovací grafy zkoumají podobnost vícerozměrných grafů. Metody vícerozměrné analýzy dat jsou procvičovány na řadě chemických, fyzikálních a biologických příkladů.

Sylabus předmětu:

1. Systém STATISTICA nebo STATGRAPHICS (základy ovládání, nastavení systémového prostředí, příprava a uchování dat, import a export dat, editování dat, operace s datovými soubory).
2. Analýza kategoriálních dat vyšetřuje i nečíselné proměnné:
 - 2.1 Crosstabulation (vícenásobné třídění) a charakteristiky závislosti zkoumaných vztahů.
 - 2.2 Kontingenční tabulky a výpočet χ^2 pro dvourozměrnou tabulku.
 - 2.3 χ^2 -statistika dobré shody pozorovaných a očekávaných četností.
 - 2.4 Logaritmicke-lineární analýza hledá model četností ve vícerozměrných datech.
 - 2.5 Numerické kódování klasifikačních faktorů transformuje textový vektor.
 - 2.6 Překódování proměnných (obvykle textových) do nových proměnných (numerických).
3. Analýza hlavních komponent (graf komponentních vah, bodový diagram, dvojný graf).
4. Shluková analýza (eukleidovská metrika, 6 shlukovacích metod, graf shluků).
5. Faktorová analýza (graf faktorových zátěží, rotace faktorové matice, graf faktorových skóre).
6. Kanonické korelace (graf kanonických proměnných).
7. Diskriminační analýza (zobrazení (ne-)standardizovaných koeficientů, zobrazení těžiště skupin, zobrazení skupinových statistik, zobrazení skupinových korelací, klasifikace pozorování, graf hodnot diskriminačních funkcí).
8. Grafické metody vyšetřování podobnosti objektů:
 - 8.1 Sluníčkové grafy.
 - 8.2 Hvězdičkové grafy.
 - 8.3 Oblíčejoyé grafy.
 - 8.4 Profilové grafy.

Literatura povinná:

1. M. Meloun, J. Militký: *Statistické zpracování experimentálních dat*, PLUS Praha 1994 (1. vydání), EAST PUBLISHING Praha 1998 (2. vydání), Academia Praha 2004 (3. vydání).
2. M. Meloun, J. Militký: *Kompendium statistického zpracování dat*, Academia Praha 2002 (1. vydání), Academia Praha 2006 (2. vydání).

Literatura doporučená:

1. F. Koschin a kol.: *STATGRAPHICS aneb statistika pro každého*, GRADA Praha 1992.
2. J. Hanousek, P. Charamza: *Moderní metody zpracování dat - matematická statistika pro každého*, GRADA Praha 1992.
3. J. Hanousek, P. Charamza: *Moderní statistika a zpracování dat (průvodce úskalími matematické statistiky)*, Famulus Praha 1992.
4. P. Hebák, J. Hustopecský: *Vícerozměrné statistické metody s aplikacemi*, SNTL/ALFA Praha 1987.
5. P. Hebák, J. Hustopecský: *Průvodce moderními statistickými metodami*, SNTL Praha 1987.

Semestr/hodin: 21 hodin