

ANALÝZA DAT – VÍCEROZMĚRNÉ STATISTICKÉ METODY (SOC418)

PS 2004/2005

Vyučující: Tomáš Katrňák a Ladislav Rabušic

počítačová učebna, středa 10.00-11.30

PRŮVODCE KURSEM

Zaměření a organizace kurzu a podmínky pro jeho absolvování

Kurz Analýza dat - *multivariační statistické metody* je určen pro studenty magisterského studia Fakulty sociálních studií MU. Vyžaduje znalost nezbytných základů práce v prostředí WINDOWS, zvládnutí manažera programů a textového editoru. Navazuje na kurz SOC108- *Statistická analýza dat*, přednášený v bakalářském studiu a předpokládá znalosti základů statistické analýzy a schopnost práce s programem SPSS pod WINDOWS. Kurz je jednosemestrový. Veden je v počítačové učebně a zahrnuje praktické semináře s výkladem. Kurz je zakončen zápočtem, pro jehož získání je třeba vypracovat samostatnou semestrální práci využívající aplikaci studentem vybrané vícerozměrné statistické metody na datech z reálných výzkumů při využití sociologického teoretického vhledu. Za absolvování kurzu je 12 kreditů. Výuka předpokládá individuální procvičování probrané látky a pravidelné řešení příkladů, které se průběžně odevzdávají vyučujícím.

Cíle a rámcová náplň kurzu

Cílem kurzu je seznámení studentů s aplikací vybraných vícerozměrných statistických metod v SPSS. Předmětem výuky bude regresní analýza (včetně logistické regrese), shluková, diskriminační, faktorová analýza, vícerozměrné škálování, korespondenční analýza, strukturní modelování a loglineární analýza. Při výuce bude kladen důraz na použití metod na reálných datech pocházejících z České republiky z let 1990-2001. Vždy bude vyložena základní podstata příslušné metody a základní požadavky na data vstupující do příslušné procedury. Poté bude metoda aplikována na datech z ČR v SPSS a budou vysvětleny základní výstupy z SPSS. Na závěr každé lekce bude zadána domácí úloha k samostatnému procvičení příslušné vícerozměrné statistické metody.

OBSAH

1. lekce (29. 9. 2004)

ÚVOD

(Katrňák)

Vícerozměrné statistické metody a jejich role v sociálních vědách, modelování v sociologii, logiky modelů, srovnání statistických programů SPSS a STATY

2. lekce (6. 10. 2004)

REGRESNÍ ANALÝZA

(Rabušic)

Regresní a korelační analýza. Dílčí, parciální a vícenásobná korelace. Regresní analýza dvou či více proměnných. Požadavky na data použítá v regresní analýze (včetně zařazení nominálních proměnných), hledání nejvhodnějšího regresního modelu, interpretace výsledků. Problémy v regresní analýze, jejich diagnostika a řešení.

Literatura:

Field, A. 2000. *Discovering Statistics using SPSS for Windows*. SAGE, London, str. 103–161. (v knihovně FSS).

Rabušic, L. 2003. *Regresní analýza*. Učební text umístěný v IS

Aplikace:

Řeháková, B. 2001. Determinanty vztahu k ochraně životního prostředí. *Sociologický časopis*, 37, č.4: 479-498.

Hamplová, D. 2000. Náboženství a nadpřirozeno ve společnosti. *Sociologické texty* č. 3, SoÚ AV ČR.

3. lekce (13. 10. 2004)

REGRESNÍ ANALÝZA – pokračování

(Rabušic)

4. lekce (20. 10. 2004)**ÚVOD DO LOGISTICKÉ REGRESE (Katrňák)**

Logika lineárního pravděpodobnostního modelu a modelu latentní proměnné, princip logistické regrese, způsob odhadu logitového modelu, logistická regrese pro závisle proměnnou o dvou kategoriích (binární logistická regrese a probitový model, test koeficientů - Wald test, Irtest, odhad nejpřesnějšího a zároveň nejúspornějšího modelu - fitstat, logika Occamovy břitvy, BIC, interpretace poměrů šancí - listcoef, predikce jako interpretační nástroj odhadnutého modelu - prvalue, grafická vizualizace)

Literatura

Agresti, Alan: *An Introduction to Categorical Data Analysis*. Wiley, New York 1996 (5. kapitola, 103-135)

Long, Scott J. & Jeremy Freese: *Regression Models For Categorical Dependent Variables using Stata*. Stata Corporation, College Station, 2000 (4. kapitola, str. 99-135)

Aplikace:

Katrňák, Tomáš: „Determinanty poklesu sňatečnosti v 90. letech 20. století v České republice.“ In. www.fss.muni.cz/~katrnak

5. lekce (27. 10. 2004)**LOGISTICKÁ REGRESE PRO ZÁVISLE PROMĚNNOU O VÍCE USPOŘÁDANÝCH KATEGORIÍCH (Katrňák)**

Ordinální logistická regrese a ordinální probitový model, způsob odhadu logitového modelu, , interpretace poměrů šancí - listcoef, predikce jako interpretační nástroj odhadnutého modelu - prvalue, grafická vizualizace

Literatura

Long, Scott J. & Jeremy Freese: *Regression Models For Categorical Dependent Variables using Stata*. Stata Corporation, College Station, 2000 (5. kapitola, str. 135-170)

Aplikace:

Příklady z četby *Regression Models For Categorical Dependent Variables using Stata*. Stata Corporation, College Station, 2000

6. lekce (3. 11. 2004)**LOGISTICKÁ REGRESE PRO ZÁVISLE PROMĚNNOU O VÍCE KATEGORIÍCH (MULTINOMICKÁ LOGISTICKÁ REGRESE A PODMÍNĚNÁ LOGISTICKÁ REGRESE) (Katrňák)**

Multinomická logistická regrese a podmíněná logistická regrese, způsob odhadu logitového modelu, , interpretace poměrů šancí - listcoef, predikce jako interpretační nástroj odhadnutého modelu - prvalue, grafická vizualizace

Literatura

Long, Scott J. & Jeremy Freese: *Regression Models For Categorical Dependent Variables using Stata*. Stata Corporation, College Station, 2000 (6. kapitola, str. 171-221)

Aplikace:

Příklady z četby *Regression Models For Categorical Dependent Variables using Stata*. Stata Corporation, College Station, 2000

7. lekce (10. 11. 2004)**VÍCEROZMĚRNÁ ANALÝZA ROZPTYLU (s opakováním t-testu a ANOVY) (Rabušic)**Literatura

Field, A. 2000. *Discovering Statistics using SPSS for Windows*. SAGE, London, str. 206–309. (v knihovně FSS).

Aplikace:**(17. 11. 2004) STÁTNÍ SVÁTEK****8. lekce (24. 11. 2004)****VÍCEROZMĚRNÁ ANALÝZA ROZPTYLU – pokračování (Rabušic)****9. lekce (1. 12. 2004)****FAKTOROVÁ ANALÝZA ANEB REDUKCE DAT (Rabušic)**

Základní pojmy faktorové analýzy: Faktorová matice, faktorové zátěže a vyčerpáný rozptyl. Rotace faktorů. Faktorové skóre a jeho využití jako další proměnné (pro regresní či diskriminační analýzu) a pro klasifikace. Faktorová

analýza jako podpora pro konstrukci škál. Grafické využití faktorové analýzy. Konfirmační faktorová analýza (stručně).

Literatura:

Field, A. 2000. Discovering Statistics using SPSS for Windows. SAGE, London, str. 423–470 (v knihovně FSS).

Hendl, J. 2004. Přehled statistických metod zpracování dat. Portál, Praha, str. 468–485.

Rabušic, L. 2003. Faktorová analýza. Učební text (v IS Muni).

Aplikace :

Matějů, P., Vlachová, K. a kol. 2000. *Nerovnost, spravedlnost, politika*, Slon, Praha, str. 17-28, (v knihovně FSS).

10. lekce (8. 12. 2004)

(Rabušic)

METODY SLOUŽÍCÍ KE KLASIFIKACI – SHLUKOVÁ (CLUSTER) ANALÝZA A MULTI-DIMENSIONÁLNÍ ŠKÁLOVÁNÍ

Vícerozměrný prostor a problém hledání podobných objektů, shlukování případů a proměnných. Hierarchická a K-means shluková analýza. Grafické prostředky - dendrogram. Hodnocení kvality klasifikace. Metody umožňující v grafické podobě znázornění vztahů více proměnných ve dvou popř. třírozměrném grafu. Ukázka základních výstupů a možnosti využití výstupu z SPSS pro kreslení vlastních grafů. Redukce informací při použití metod.

Literatura:

SPSS Categories 8.0 Chicago, SPSS Inc. Introduction to SPSS Optimal Scaling Procedures for Categorical Data, pp. 1-20, Correspondence Analysis pp. 43-52

Aplikace :

Rabušic, L.: Koho Češi nechtějí, Sociální studia 5: 67-84, 2000 (vícerozměrné škálování)

11. lekce (15. 12. 2004)

OPAKOVÁNÍ

(Rabušic, Katrňák)