

SYLABUS

SOC414: Analýza kategorizovaných dat v sociologii

Vyučující: PhDr. Tomáš Katrňák Ph.D.

Anotace: Kategorizovaná data tvoří většinu dat v sociologických šetřeních. Tento kurs je praktický, není zaměřen na teorii, ale na analýzu a interpretaci kategorizovaných dat. Studující se naučí pracovat se dvěma technikami, které se k analýze kategorizovaných dat od druhé poloviny osmdesátých let minulého století v sociologii používají. K analýze vztahu mezi nezávisle a závisle proměnnou budeme pracovat s jednotlivými variantami logistické regrese. V případě, že nelze ve vztahu proměnných určit nezávisle a závisle proměnnou, budeme pracovat s loglineární analýzou. Studenti a studentky se naučí pracovat v současnosti s velmi populárními statistickými programy STATA a LEM. Kurz je dimenzován na malou počítačovou učebnu FSS, a proto je jeho kapacita omezena na 11 studujících.

Literatura ke kursu:

Agresti, Alan: *An Introduction to Categorical Data Analysis*. Wiley, New York 1996 (dále jen Agresti)

Long, Scott J. & Jeremy Freese: *Regression Models For Categorical Dependent Variables using Stata*. Stata Corporation, College Station, 2000 (dále jen L&F)

Powers, Daniel A. & Yu Xie: *Statistical Methods for Categorical Data Analysis*. Academic Press. San Diego 2000 (dále jen P&X)

Podmínky absolvování kurzu:

Kurz je ukončen zápočtem. K jeho získání je nutné absolvovat semináře, splnit průběžné úkoly zadávané z hodiny na hodinu a vypracovat během semestru 3 rozsáhlejší úlohy, které musejí být odevzdané vždy v daném termínu (viz sylabus níže). Za splněnou úlohu může studující získat max. 10 bodů. Ke získání zápočtu je nutné dosáhnout alespoň 16 bodů. Při práci na úlohách nemusejí studující využít pouze statistické programy, které budou v kursu představeny, ale mohou využít i jiné statistické programy, např. SPSS nebo Excel, jež jsou na FSS k dispozici.

- 1) **Úvod: kategorizovaná data v sociologii** (transformační přístup k nim a přístup k nim jako k latentním proměnným, základní operace v programu STATA a LEM).
literatura: P&X (*Introduction*, str. 1-13); *Getting started with Stata for Windows* (knihovna FSS)
- 2) **Analýza dvojrozměrných tabulek v sociologii** (podíly, šance, poměry šancí, statistická inference pro poměry šancí, chí-kvadrát test, Pearsonův chí-kvadrát a poměr maximální věrohodnosti, adjustované reziduály, koeficienty asociace)
literatura: Agresti (2. kapitola: *Two-Way Contingency Tables*, str. 16-45)
zadání úlohy č. 1
- 3) **Analýza vícerozměrných tabulek v sociologii** (parciální asociace, podmíněné a marginální poměry šancí, homogenní asociace)
literatura: Agresti (3. kapitola: *Three-Way Contingency Tables*, str. 53-66)
- 4) **Cvičení: standardní analýza kontingenčních tabulek ve STATĚ a úvod do logistické regrese** (logika lineárního pravděpodobnostního modelu a modelu latentní proměnné)
literatura: Agresti (5. kapitola: *Logistic Regression*, str. 103-135)
odevzdání úlohy č. 1

* * *

5) **Logistická regrese pro závisle proměnnou o dvou kategoriích** (binární logistická regrese a probitový model, test koeficientů - Wald test, lrtest, 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: L&F (4. kapitola: *Models for Binary Outcomes*, str. 99-135)

6) **Logistická regrese pro závisle proměnnou o více uspořádaných kategoriích a pro závisle proměnnou o více kategoriích** (ordinální logistická regrese a ordinální probitový model, multinomická logistická regrese a podmíněná logistická regrese)
literatura: L&F (5. kapitola: *Models for Ordinal Outcomes*, str. 135-170 & 6. kapitola: *Models for Nominal Outcomes*, str. 171-221)
zadání úlohy č. 2

7) **Týden pro samostudium**

V tomto týdnu probíhá výuka na katedře sociologie ve formě samostudia. Studující mají za úkol znovu projít dosavadní četbu k kursu, nastudovat případnou zameškanou četbu, vypracovat a dopracovat zadané úkoly.

8) **Cvičení: odhady všech typů logitových modelů ve Statě** (logistic, logit, bilogit, mlogit, ologit a clogit)
odevzdání úlohy č. 2

* * *

9) **Hierarchické log-lineární modely** (logika loglineárního modelování, výstavba modelů, vztah mezi úsporností a přesností, chí-kvadrát test a poměr maximální věrohodnosti, BIC)
literatura: Agresti (6. kapitola: *Loglinear Models for Contingency Tables*, str. 145-167)

10) **Nehierarchické log-lineární modely, topologické modely a úvod do modelů asociace** (normalizace parametrů effect coding a dummy coding, interpretace, výstavba nehierarchických modelů, ordinální znaky jako předpoklad asociace)
literatura: Agresti (7. kapitola: *Building and Applying Logit and Loglinear Models*, str. 174-199)
zadání úlohy č. 3

11) **Modely asociace pro dvojrozměrné tabulky a vícerozměrné tabulky** (uniformní asociace, R model, C model, RCI model, RC II model, RCII homogenní model, vzorec asociace a efekt asociace, homogenní, uniformní, multiplikatívni a heterogenní mezitabulkový efekt asociace)
literatura: P&X (4. kapitola: *Loglinear Models for Contingency Tables*, str. 87-146)

12) **Cvičení: modely asociace ve Statě a LEMu**

literatura: P&X (4. kapitola: *Loglinear Models for Contingency Tables*, str. 87-146)
odevzdání úlohy č. 3