CEU Doctoral School of Political Science, Public Policy and International Relations

2014/2015 Fall

Categorical Data Analysis

Instructor

Tamas Rudas
Instructor e-mail: rudas@tarki.hu
Office hours: 10:40 – 12:40, Tuesdays or by appointment. Please send an e-mail if you want to see me
Class meets: 9:00 – 10:40, Tuesdays

Course description

Much of the data in political science (and in the social sciences, in general) are categorical, and overlooking this feature often leads to the application of inappropriate methods of analysis. While categorical data carry the possibility of revealing more complex association structures than those that may be discovered under the common assumption of normality, the application of inappropriate methods makes fine-tuning of the models impossible. The central theme of the class is the concept of interaction, which covers various forms of effects and associations. Methods for the analysis of interactions have proved useful across the social sciences, where the focus of interest is mostly the joint behaviour of the variables, or effects of some of them on others, as opposed to the individual analyses of the variables. The use of mathematics will be tailored to the level of the audience

Learning goals and outcomes

To gain a basic understanding of the central concepts of categorical data analysis, to be able to read and interpret such analyses, to be able to carry out certain analyses with categorical data.

Topics to be covered (timing subject to change)

Week 1
Levels of measurement
Contingency tables
Independence, tests of independence.

Week 2
Continuous and categorical data

Real data are always categorical, continuity is a simplifying assumption
Advantages and disadvantages of the assumption of normality
In reality, very few relationships are linear

Week 3
Sampling distributions for categorical data
Binomial distribution
Multinomial distribution
Poisson distribution

Week 4
Simpson's paradox
Relationship with causality
Confounding
Informative allocation

Week 5
The concept of interaction
Independence
Structural parameterizations
Variation indepence of the marginals and the odds ratio

Week 6
Conditional odds ratios
Measures of conditional association
General versions of the variation independence

Week 7
Log-linear models I
The structure of multidimensional contingency tables
Interpretation of log-linear models
The regression problem for categorical variables

Week 8
Log-linear models II
Log-linear parameters
Model selection

Week 9
Graphical models
Markov properties
Directed Markov properties
Path models f or categorical data

Week 10
Chain graphs
Marginal models
Data analysis with graphical models

Week 11
Asymptotic methods
Asymptotic normality of the multinomial distribution
The delta method

Week 12
Tests of model fit
Pearson chi-squared and likelihood ratio tests
The mixture index of fit
Resampling tests

Main Texts

Agresti, A (2010) *Categorical Data Analysis, 2nd ed*, Wiley

Rudas, T (1998) *Odds Ratios in the Analysis of Contingency Tables*, Sage

Some papers of the instructor which will be studied:

Rudas, T (2014) Log-linear and marginal models. To appear in Wright, J (ed)
International Encyclopedia of Social and Behavioral Sciences 2nd ed, Elsevier

Németh, R, Rudas, T (2013) On the application of discrete marginal graphical models.
Sociological Methodology, 43, 70-100.

Németh, R. Rudas, T (2013) Discrete Graphical Models in Social Mobility Research A
Comparative Analysis of American, Czechoslovakian and Hungarian Mobility before the
Collapse of State Socialism. *Bulletin of Sociological Methodology*, 118, 5-21.

Grading

Take-home assignment