ABSTRACT

A susbtantial part of the data in the social, behavioral and life sciences are of a categorical nature. This is of course true for nominal level data, but also for ordinal or interval level measurements. There are two great traditions in analyzing categorical data. Either categorical data is implicitly or explicitly treated as realizations of underlying continuous variables (the 'Pearsonian tradition') or it is taken for what they are: 'just' categorical (the Yulean tradition). Especially due to the developments in loglinear (and logit) modeling, the scope of the Yulean approach has been expanded enormously. This course will treat and evaluate some of these developments.

A special place will be occupied by Latent Class Analysis (LCA) and loglinear models with latent variables. In general, latent variable modeling plays a crucial role in analyzing measurement models, separating true from random change, handling missing data, accounting for unobserved heterogeneity, correcting for many kinds of random and systematic measurement and survey errors, analysing multi-level or random effects models, etc.. In latent class analysis, the latent variables are considered to be categorical (nominal, ordinal or interval level) variables. In this course, it will be shown how to integrate categorical latent variables into loglinear models to fulfill the above mentioned roles of latent variable models.

In doing this, it is often very useful to take a 'causal' point of view in the form of Structural Equation Models – SEM's. In general, causal models in the form SEM's play an important part in research. The extended LCA can be used as a SEM for categorical (nominal, ordinal, interval level) data in much the same vein as LISREL, AMOS, etc. can be used for continuous data. It wil be discussed how to set up causal models for categorical data, also comparing the differences and commonalities between standard SEM and LCA. Graphical models provide the common ground for standard and 'nonstandard' SEM, as will a short discussion of the nature of causality.
This SEM version of LCA can then be used to investigate the nature of many missing data and selection problems and correct for them. Examples are selfselection, nonresponse, verification bias, and noncompliance bias. Also item bias in cross-cultural and group comparisons can be fruitfully investigated and corrected by means of LCA. LCA has also proven to be very useful for the analysis of change over time in categorical characteristics, especially regarding the disentanglement of the observed changes into true, real change and random change. A lot of seemingly systematic patterns of observed change might be explained by and just result from (measurement) error and random phenomena, such as 'mood'. Finally, LCA can be regarded as a mixture model where the population can be divided into several 'unobserved' distinct groups. It is then possible to investigate whether analysis results, such as from regression analysis, factor analysis, or loglinear models are the same in all 'unobserved' groups. In this way, unobserved heterogeneity can be taken into account. Standard random effects models (multi-level models) are essentially 'mixtures' based on a continuous variable. LCA provides an often to be preferred categorical alternative for continuous random effects.

The emphasis in this course will be on the underlying logic and application of the models. About 60% of the time will be spent in the class room on lectures, and 40% in computer rooms, applying the models to real world data.

Some (very) basic knowledge of loglinear modeling is strongly recommended.

Own data sets and problems that are relevant for this course can be used but must be submitted several weeks ahead to make preparation for everybody possible (when possible!).

Programs to be used: LEM (Vermunt), Latent Gold (Statistical Innovations), and/or Mplus (Muthen)

Literature:
Elementary
Knoke D. and P.J. Burke (1980) *Log-linear Models*. Sage University paper 07-020, Newbury park: Sage (or later edition)
McCutcheon a.L. (1987) *Latent Classs Analysis*. Sage University Paper, 07-64, Newbury park: Sage

Intermediate
Hagenaars J.A. (1990) *Categorical Longitudinal Data: loglinear panel, trend, and cohort analysis*, Chapters 2 and 3. Newbury Park: Sage
Hagenaars J. (1993) *Loglinear Models with Latent Variables*. Sage University Paper 94. Newbury park: Sage

Elementary .... Advanced:
Hagenaars, Jacques A. and Allan McCutcheon (Eds.) (2002), Applied Latent Class Analysis. New York: Cambridge University Press

Teachers:
Lectures: Jacques Hagenaars
Computer applications: Jacques Hagenaars
Ales Ziberna

Schedule 2008

Thursday, April 17: 3 hours
Basic loglinear modeling: odds (ratios), effect-, dummy coding, interpretation of parameters; (MLE)estimation, testing

Friday, April 18: 3 hours
Loglinear models for ordered data (ordinal, interval); Logit Models

Saturday, April 19: 6 hours
Latent Class Models; loglinear models with latent variables

Monday, April 21: 3 hours
Causal models: causality, (un)directed graphs, causal/directed loglinear models

Tuesday, April 22: 3 hours
SEM's for categorical data: directed loglinear models with latent variables

Wednesday, April 23: 3 hours

Comparative research, measurement/item bias

Thursday, April 24: 3 hours
Survey errors/nonexperimental designs: nonresponse, selection bias, noncompliance

Friday, April 25: 3 hours
Longitudinal research: True change, random change, systematic bias; misclassifications, regression to the mean/mode; (latent) markov chains

Monday, April 28: 3 hours
Longitudinal research (continued)

Tuesday, April 29: 3 hours
Mixture models; Categorical Multilevel models

Wednesday, April 30: 3 hours
Miscellaneous (topics from previous days, topics proposed by participants, additional (computer) applications, etc.)