

# Analysis of Categorical Data

Wirtschaftswissenschaftliche Universität Wien

Summer 2007

Instructor: Alexander von Eye  
Phone: 355-3408  
e-mail: [VONEYE@MSU.EDU](mailto:VONEYE@MSU.EDU)

## Goals

The course on the Analysis of Categorical Data in Social Science Research pursues the following goals:

1. Provide participants with an overview of methods for analysis of categorical data.
2. Present sample methods in detail. Illustrate characteristics of these methods.
3. Practice computer application of such methods.
4. Application of methods to simulated and real data.

## Course Modules

This course involves four modules. The last one is optional.

1. **Coefficients for analysis of relationships in pairs of categorical variables:** sample coefficients include the  $\chi^2$  contingency coefficient, Cohen's  $\kappa$ , and Goodman and Kruskal's  $\lambda$ . Characteristics of these coefficients are illustrated.
2. **Log-linear models:** log-linear modeling will be presented from a general log-linear modeling perspective. The design matrix approach is taken. Log-linear modeling typically is applied for model fitting. Methods for parameter estimation are introduced. Hierarchical models, logit models and logistic regression, and non-hierarchical models are introduced. Parameter interpretation is illustrated. Models of specific concern to social science researchers are introduced, for example, models of symmetry, rater agreement, and longitudinal data. The general log-linear model is also used to come to a more in-depth understanding of the coefficients introduced in Module 1.
3. **Prediction Analysis (PA):** PA is a method for analysis of point predictions in categorical

variables. This module includes the following topics: types of point predictions; formulating point predictions in categorical variables; systems of point predictions; criteria for statistical evaluation of point predictions; estimation of expected values; measures of prediction success; statistical tests; computer applications. In addition, this module shows how to use log-linear models for analysis of systems of point predictions. *This module will be taught only if there is enough time left in the semester.*

### Readings

The course will not be based on a particular book. Rather, readings are selective, and a number of books and articles will be used. Here are examples:

- The book that covers models for longitudinal data is:  
von Eye, A., & Niedermeier, K.E. (1999). *Statistical analysis of longitudinal categorical data - An introduction with computer illustrations*. Mahwah, NJ: Lawrence Erlbaum.
- For odds-ratio analysis, the following text is recommended:  
Rudas, T. (1998). *Odds ratios in the analysis of contingency tables*. Thousand Oaks: Sage.
- A strong text that focuses on logistic regression and log-linear modeling is  
Agresti, A. (2002). *Categorical data analysis*, 2<sup>nd</sup> ed. Hoboken, NJ: Wiley.
- More technical in some passages and more computer-oriented is  
Lawal, B. (2003). *Categorical data analysis with SAS and SPSS applications*. Mahwah, NJ: Lawrence Erlbaum.
- Specifically useful for Configural Frequency Analysis is  
von Eye, A. (2002). *Configural Frequency Analysis - methods, models, and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Specifically useful for models of rater agreement is  
von Eye, A., & Mun, E.Y. (2005). *Analyzing rater agreement*. Mahwah, NJ: Lawrence Erlbaum.

### Course Grading

Each participant will produce one paper. These papers can be either theoretical in nature or present a data problem and an attempt at its solution. Theoretical papers discuss technical aspects of categorical data analysis. Applied papers present data and their analyses. Topics of the papers must be approved by the instructor and the class.

Course Schedule (part of the sessions early in the semester involve participants' brief presentations of project proposals)

---

Module	Topic	Reading
1	Introduction Measures of association in two-dimensional contingency tables	Rudas, Ch. 1 Liebetrau, 1983
2	Decomposition of variation in two- and higher dimensional tables	von Eye & Niedermeier, Ch. 4 Rudas, Ch. 1 and 2
2	The general log-linear model	Agresti, Ch. 8; Lawal, Ch. 6; von Eye & Niedermeier, Ch. 2
2	Log-linear sample models	Agresti, Ch. 8; Lawal, Ch. 6; von Eye & Niedermeier, Ch. 2
2	Log-linear models for longitudinal data	von Eye & Niedermeier, Ch. 3; Hagenaars, 1990
2	Parameter interpretation in log-linear modeling	Mair & von Eye, in press; von Eye & Niedermeier, Ch. 2
2	Computer programs and exercises	
3	Prediction analysis	von Eye & Niedermeier, Ch. 5

---

Additional References

- Fienberg, S.E. (1981). *The analysis of cross-classified categorical data*. Cambridge, MA: The MIT Press (2nd ed.)
- Gutiérrez-Peña, E., & von Eye, A. (2000). A Bayesian approach to Configural Frequency Analysis. *Journal of Mathematical Sociology*, 24, 151 - 174.
- Hagenaars, J.A. (1990). *Categorical longitudinal data*. Newbury Park: Sage.
- Hildebrand, D.K., Laing, J.D., & Rosenthal, H. (1977). *Prediction analysis of cross-classifications*. New York: Wiley.
- Indurkha, A., & von Eye, A. (2000). The power of tests in Configural Frequency Analysis. *Psychologische Beiträge*, 42, 301 - 308.
- Langeheine, R., & Rost, J. (Eds.)(1988). *Latent trait and latent class models*. New York: Plenum.
- Liebetrau, A.M. (1983). *Measures of association*. Beverly Hills: Sage.
- Long, J.S. (Ed.)(1988). *Common problems/proper solutions. Avoiding error in quantitative research*. Newbury Park: Sage.
- Mair, P., & von Eye, A. (2007). Application Scenarios for Nonstandard Log-Linear Models. *Psychological Methods*.
- Mun, E.-Y., von Eye, A., Fitzgerald, H.E., & Zucker, R.A. (2001). Using Mosaic Displays in Configural Frequency Analysis (CFA). *Methods of Psychological Research - online*, 6, 164 - 196.
- Rindskopf, D. (1990). Nonstandard log-linear models. *Psychological Bulletin*, 108, 150 - 162.
- von Eye, A. (2002). The odds favor antitypes - A comparison of tests for the identification of configural types and antitypes. *Methods of Psychological Research - online*, 7, 1-29.
- von Eye, A. (2004). Base models for Configural Frequency Analysis. *Psychology Science*, 46, 150 - 170.
- von Eye, A., & Brandtstädter, J. (1988). Formulating and testing developmental hypotheses using statement calculus and non-parametric statistics. In P. B. Baltes, D. Featherman, & R. M. Lerner (Eds.), *Life-span development and behavior* (Vol. 8, pp. 61-97). Hillsdale, NJ: Erlbaum.
- von Eye, A., & Brandtstädter, J. (1997). Configural Frequency Analysis as a searching device for possible causal relationships. *Methods of Psychological Research - Online*, 2, 1-23.
- von Eye, A., & Brandtstädter, J. (1998). The Wedge, the Fork, and the Chain - Modeling dependency concepts using manifest categorical variables. *Psychological Methods*, 3, 169 - 185.
- von Eye, A., Brandtstädter, J., & Rovine, M. J. (1993). Models for prediction analysis. *Journal of Mathematical Sociology*, 18, 65-80.
- von Eye, A., Brandtstädter, J., & Rovine, M.J. (1997). Models for prediction analysis in longitudinal research. *Journal of Mathematical Sociology*, 22, 355 - 371.
- von Eye, A., & Clogg, C.C. (Eds.)(1996). *Analysis of categorical variables in developmental research*. San Diego, CA: Academic Press.
- von Eye, A., & Gutiérrez Peña, E. (2004). Configural Frequency Analysis - the search for extreme cells. *Journal of Applied Statistics*, 31, 981 - 997.

- von Eye, A., Schauerhuber, M. & Mair, P. (2007). Significance tests for the measure of raw agreement. *InterStat*, <http://interstat.statjournals.net/YEAR/2007/abstracts/0701001.php> (January 2007)
- von Eye, A., & Mun, E.-Y. (2003). Characteristics of measures for 2 x 2 tables. *Understanding Statistics*, 2, 243 - 266.
- von Eye, A., & Mun, E.Y. (2005). *Modeling rater agreement - manifest variable approaches*. Mahwah, NJ: Lawrence Erlbaum.
- von Eye, A., Schuster, C., & Gutiérrez-Peña, E. (2000). Configural Frequency Analysis under retrospective and prospective sampling schemes - frequentist and Bayesian approaches. *Psychologische Beiträge*, 42, 428 - 447.
- von Eye, A., & Sörensen, S. (1991). Models of chance when measuring interrater agreement with kappa. *Biometrical Journal*, 33, 781 - 787.
- von Eye, A., & Spiel, C. (1996). Standard and non-standard log-linear symmetry models for measuring change in categorical variables. *The American Statistician*, 50, 300 - 305.
- von Eye, A., Spiel, C., & Rovine, M. J. (1995). Concepts of nonindependence in Configural Frequency Analysis. *Journal of Mathematical Sociology*, 20, 41 - 54.
- von Eye, A., & von Eye, M. (2005). Can one use Cohen's kappa to examine disagreement? *Methodology*, 1, 129 - 142.
- von Weber, S., Lautsch, E., & von Eye, A. (2003). On the limits of Configural Frequency Analysis: Analyzing small tables. *Psychology Science*, 45, 339 - 354.
- von Weber, S., von Eye, A., & Lautsch, E. (2004). The Type II error of measures for the analysis of 2 x 2 tables. *Understanding Statistics*, 3, 259 - 282.
- Wickens, T.D. (1989). *Multiway contingency tables analysis for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum.