

# Categorical Data Analysis

Thursday and Friday, starting at 5PM

Seminarraum (NT02064): 6, 20, 27 Nov; 4, 11, 12, 18 Dec; 8, 15, 29 Jan  
SR für Statistik (NT03098): 7, 14, 21, 28 Nov; 5, 19 Dec; 9, 16, 23, 30 Jan

## INSTRUCTOR:

Prof. Bernhard Klingenberg

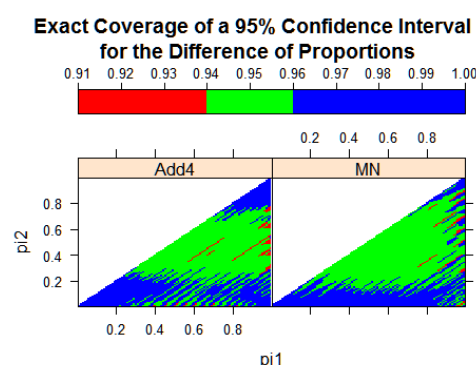
Dept. of Mathematics & Statistics

Williams College, MA

e-mail: [bklingen@williams.edu](mailto:bklingen@williams.edu)

Course Webpage:

<http://sites.williams.edu/bklingen/teaching/cda/>



## COURSE DESCRIPTION:

This course has two parts. In the first part we will discuss statistical inference for parameters of categorical distributions (Bernoulli, Binomial, Multinomial, Poisson) and for measures of association arising in contingency tables (difference and ratio of proportions and odds ratios). We will also discuss ordinal response data. Inferential methods covered include Wald, score and likelihood ratio tests and confidence intervals, as well as the bootstrap, exact methods (permutation approaches) and the Bayesian paradigm. The second part will focus on statistical inference for dependent observations and focus on analyzing matched pairs and the generalized estimating equation approach (GEE) for modeling clustered categorical responses. If time permits, we will also cover modern classification tools such as regression trees.

Topic	Summary
Introduction	Probability models for categorical data: Bernoulli, Binomial, Hypergeometric, Multinomial and Poisson distribution.
Maximum Likelihood Inference	General likelihood theory with special regard to confidence intervals and hypotheses testing for binomial, Poisson and multinomial parameters. Wald, Score, Likelihood Ratio tests and confidence intervals. Coverage probability. Exact and Bayesian inference for proportions.
Contingency tables	Sampling models for contingency tables. Effect measures in 2x2 and larger contingency tables. Comparing proportions in two groups. Chi-squared test. Exact small sample tests. Ordinal responses. Bradley-Terry Model.
Stratified tables	Mantel-Haenszel tests and estimators.
Marginal models	Comparing dependent proportions, Generalized Estimating Equations for modeling correlated binary responses.
Data mining tools for categorical responses	Classification and Regression trees (If time permits)

## LITERATURE:

Alan Agresti, Categorical Data Analysis, 2<sup>nd</sup> or 3<sup>rd</sup> edition, John Wiley & Sons