

STAT 418 – Analysis of Cross-Classified Data

1. Catalog Description

Stat 418 Analysis of Cross-Classified Data 4 units

Discrete multivariate statistics, including analysis of cross-classified data, log-linear models for multidimensional contingency tables, goodness of fit statistics, measures of association, model selection, and hypothesis testing. 4 lectures. Prerequisite: Two courses in statistics.

2. Required Background and/or Experience

Two courses in statistics

3. Expected Outcomes

The student should:

- a. be able to identify categorical variables
- b. be able to calculate and interpret odds ratios and relative risks
- c. understand and be able to calculate the different goodness of fit statistics
- d. be able to specify a generalized linear model and identify its components
- e. know what a logistic regression model is, be able to specify the correct model for a particular set of data, and know how to estimate the parameters and conduct tests on the model
- f. know what a loglinear model is, be able to specify the correct model for a particular set of data, and know how to estimate the parameters and conduct tests on the model
- g. be able to determine what models make up a hierarchical set of models for a particular data set, and understand the importance of using hierarchical sets
- h. know the methodology and be able to explain the rationale in building logit and loglinear models.

4. Text and References

Text: Agrest, Alan, *An Introduction to Categorical Data Analysis*, New York, Wiley, 1996.

Reference: Simonoff, Jeffrey, *Analyzing Categorical Data*, New York, Springer, 2003.

5. Minimum Student Materials

None.

6. Minimum University Facilities

Chalkboards for classroom use and computers equipped with suitable statistical computing software.

7. Expanded Description of Content and Method

CONTENT	Lecture hours
a. Introduction to the analysis of categorical data	1
b. One-way classification models	3
Sampling schemes	
Goodness of fit: X^2 and G^2 statistics	
c. Two-dimensional contingency tables.....	10
Sampling schemes	
Odds ratio	
The independence model	
Goodness of fit tests	
Multiple comparisons	
Tests for ordinal data	
Pooling two-way tables	
Residual analysis	
d. Three-dimensional contingency tables.....	6
Sampling schemes	
Models for three dimensional tables	
Cochran-Mantel-Haenszel	
Estimation of cell frequencies	
e. The generalized linear model and its components	3
Generalized linear models for binary data	
Generalized linear models for count data	
f. Logistic Regression	4
Model interpretation	
Model fitting	
g. Loglinear Models for contingency tables.....	5
Two-way tables	
Three-way tables	
Loglinear-logit connection	
h. Selected topics	4
	Total <hr style="display: inline-block; width: 100px; border: 0.5px solid black; vertical-align: middle; margin-left: 10px;"/>
	36