

Hong Kong Baptist University
Faculty of Science
Department of Mathematics

Title (Units): STAT 2120 CATEGORICAL DATA ANALYSIS (3,3,0)

Course Aims: To equip students with statistical methods for analyzing categorical data arisen from qualitative response variables which cannot be handled by methods dealing with quantitative response, such as regression and ANOVA. Some computing software, such as SAS, S-PLUS, R or MATLAB, will be used to implement the methods.

Prerequisite: STAT 2110 or STAT1620

Prepared by: S.N. Chiu

Learning Outcomes (LOs):

Upon successful completion of this course, students should be:

No.	Learning Outcomes (LOs)
	Knowledge
1	Able to understand the basic techniques for categorical data analysis
2	Able to understand the mathematical theory of generalized linear model
3	Able to understand the statistical inference for logistic regression and loglinear model
	Skills
4	Able to analyze categorical data in two-way and three-way contingency tables
5	Able to model binary response variable by logistic regression
6	Able to model count response variable by loglinear model
7	Able to model matched pair data by (quasi-)symmetry model
8	Able to implement the analyses in SAS, S-PLUS, R or MATLAB
	Attitude
9	Aware of the myriad uses for categorical data methods

Assessment:

No.	Assessment Methods	Weighting	Remarks
1	Continuous Assessment	30%	Continuous assessment is designed to measure how well the students learned the categorical data analysis skills and fundamental mathematical and statistical theories, as well as the use of statistical softwares.
2	Final Examination	70%	Final Examination is designed to see how far students have achieved their intended learning outcomes. The questions are set to assess the student's ability in formulating and solving a variety of problems involving categorical data. Some questions require the student to write computer codes and to understand the corresponding computer outputs. Students should have a thorough understanding of the knowledge and a good mastery of the skills in order to answer these questions satisfactorily.

Learning Outcomes and Weighting:

Content	LO No.	Teaching (in hours)
I. Two-way Contingency Tables	1, 4, 8-9	5
II. Three-way Contingency Tables	1, 4, 8-9	5
III. Generalized Linear Models	1-6, 8-9	8
IV. Logistic Linear Models	1-6, 8-9	8
V. Loglinear Models for Contingency Tables	1-6, 8-9	7
VI. Matched Pairs	7-9	7

Textbook:

A. Agresti, An Introduction to Categorical Data Analysis, 2nd edition, Wiley, 2007.

References:

A. Agresti, Categorical Data Analysis, 2nd Ed., Wiley, 2000.
T.J. Santner and D.E. Duffy, The Statistical Analysis of Discrete Data, Springer-Verlag, 1989.
R. Christensen, Log-Linear Models and Logistic Regression, Springer-Verlag, 1997.
S.E. Fienberg, The Analysis of Cross-Classified Categorical Data, MIT Press, 2nd Ed., 1980.
D.W. Hosmer and S. Lemeshow, Applied Logistic Regression, 2nd Ed., Wiley, 2000.

Software:

SAS, S-PLUS, R or MATLAB

Course Content in Outline:

	<u>Topic</u>	<u>Hours</u>
I.	Two-way Contingency Tables A. Comparing proportions B. Odds ratio C. Large sample inference D. Small sample exact inference	5
II.	Three-way Contingency Tables A. Partial association B. Cochran-Mantel-Haenszel methods C. Exact inference on conditional associations	5
III.	Generalized Linear Models A. Binary data B. Poisson counts C. Model inference and diagnostics D. Sample size and power consideration E. Empty cells and sparseness	8
IV.	Logistic Linear Models A. Interpretation and inference B. Model checking C. Qualitative predictors D. Multiple logistic regression E. Other logit functions	8
V.	Loglinear Models for Contingency Tables A. Two-way tables B. Three-way tables C. Tables of higher dimension D. Connection between loglinear and logit models	7
VI.	Matched Pairs A. Comparing dependent proportions B. Logistic regression for matched pairs C. Symmetry and Quasi-symmetry models D. Comparing marginal distributions E. Rater agreement	7