

Sociology 605/605L (Spring 2015)
Statistics for Regression Analysis
M 12:00-2:30 (Sakamaki B414)
Lab F 12:00-1:00 (Saunders 342)

Instructor: Yean-Ju Lee (yjlee@hawaii.edu)
Saunders 216
Office Hours: M 2:30-4:00 (or by appointment)
TA: Lin Liu and Sizhe Liu

Course Description

This course is designed to provide the basic knowledge and skills needed to analyze social science data using the multiple linear regression method. The course will also deal with the techniques of categorical data analysis, such as logistic regression. Regression is a powerful and flexible statistical method that can be adapted to nearly any social science research situation. The course emphasizes the application of the techniques to actual research. Lectures will discuss how to specify regression models to answer different research questions, and Lab sessions will focus on practices of data analysis, using data from the recent American Community Survey (ACS) for the state of Hawaii (or any other states of students' choice or any other data sets students have access to (e.g., data from ICPSR)) and **STATA** program.

Required Text:

Agresti, Alan and Barbara Finlay. 2009. *Statistical Methods for the Social Sciences*. Prentice Hall. (4th edition)

*We will read a few research papers from the current issues of sociology journals.

Reference Text:

Retherford, Robert and Minja Kim Choe. 1993. *Statistical Models for Causal Analysis*. New York: John Wiley & Sons, Inc. (RC)

Allison, Paul D. 1999. *Multiple Regression: A Primer*. Thousand Oaks: Pine Forge Press. (PA)

Learning Objectives:

- understand the principles of multivariate statistical analysis
- understand the statistical principles of linear regression models
- understand the statistical principles of logistic regression models (binomial, multinomial, and ordered logit models)
- understand how to apply the techniques of linear or logistic regression to empirical research
 - how to formulate research
 - dealing with data from a large-scale national survey
 - how to specify the regression models
 - using STATA
- can read and understand articles published in sociological journals that use the linear or logistic regression

Course Requirement

Attendance and Participation: Attendance to all classes and lab sessions is **required**. Active participation in discussion/presentation is expected.

Individual Consultation: Two individual consultations during the instructor's office hours are required, preferably one in the early weeks before submitting a research proposal and the other in the later part of the semester, and additional frequent visiting is highly encouraged. Group visits by students with the same issues are also fine.

Exam: There will be one in-class exam in the 9th week. The exam will comprise approximately 25 % of the final grade. One sheet of hand-written formulas/equations on both sides is allowed.

Assignments: There may be 2-3 short assignments from the regular class or LAB, related to the textbook exercises or individual research projects. The assignments will comprise 5 % of the final grade.

Research Proposal: A research proposal—3 or more pages—is due by the end of the 7th week. The proposal should: (1) clarify the research issue and its sociological significance (note: citing the literature is important) and (2) discuss the research design and methods as much as one can, although at this point they may be tentative. Students may use the ACS data (that are used at LAB classes) but are allowed to use any other data sets after consulting with the instructor. The proposal will comprise 10 % of the final grade.

Research Paper: A research paper—about 15 pages not including tables and figures—is due by the end of the semester. The paper will comprise 60 % of the final grade. Important evaluation criteria include the following:

- (1) whether appropriate statistical methods and models are chosen to answer the research questions,
- (2) whether the data analysis is done fully,
- (3) whether the interpretation of the results is appropriate,
- (4) whether the paper has an appropriate organization, and
- (5) whether the research has a potential to contribute to the field (of sociology).

Component	1. Emerging but Insufficient (standard not met)	2. Basic (standard met)	3. Proficient (standard met)	4. Exemplary (standard met)
1. Research questions and appropriate methods	Either research questions are not clear or choose inappropriate methods	Clearly state the research questions and find the appropriate statistical methods	Use the best possible methods that have been used in the field	Consider alternative methods and demonstrate the strength and weakness of the particular methods
2. Sufficient data analysis	Simplistic multivariate analysis; or bivariate analysis	Multivariate analysis with some non-linear covariates or interactions	Findings providing clear evidence to support or reject the hypothesis; usually short and long models	Findings providing new insights beyond the hypothesis or existing literature
3. Correct interpretation of the results	Incorrect or misleading interpretation	Interpret correctly but without relating to the hypotheses	Interpret correctly and test the hypotheses	Test the hypothesis and discuss the implications in relation to all other findings
4. Paper organization	Not have all the necessary components	Have all the components	Have all the components with a logical flow	Well organized to answer the research questions, linking the findings to the hypothesis/literature

Regular Class Content: Five Components and Q/A

- (1) Lecture: basic principles of the statistics
 - (a) Use of the white/chalk board
 - (b) Power-point files
 - (2) Textbook discussion: to link the lecture to the textbook, to identify important textbook sections
 - (3) STATA commands and output
 - (4) Example articles relevant to the topic
 - (5) Practice sheets
- Q & A any time at any component

Course Schedule

■It is essential to **READ** the relevant materials **BEFORE** each class!

(Make sure to read all the chapters listed in the following, including the review chapters.)

Week

1	Jan 12	<u>Overview</u> Introduction of Data for Individual Research Project: --2012 American Community Survey, Hawaii Sampling, Measurement, Descriptive Statistics Statistical Inference: Estimation and Significance Test --Each Chapter, Sections 1, 2, 3, 4 and CH 4, Section 5	CH 1, 2, 3 CH 4, 5, 6
2	Jan 19	(holiday)	
3	Jan 26	Simple Linear Regression and Correlation -- Simple Linear Regression --Least Squares Prediction Equation --Correlation and R^2 (Coefficient of Determination) --Inference for the Slope and Correlation --Model Assumptions and Violations	CH 9
4	Feb 2	Introduction to Multivariate Relationships Multiple Regression & Correlation --Multiple Regression Model --Multiple Correlation and R squared (Coefficient of Determination) --Inference for Multiple Regression Coefficients --Partial Correlation	CH 10 CH 11; PA

Course Schedule (continued)

■It is essential to READ the relevant materials BEFORE each class!

Week

5	Feb 9	Multiple Regression & Correlation -- Modeling Interaction between two continuous independent variables --Comparing Regression Models --Standardized Regression Coefficients	CH 11
6	Feb 16	(holiday)	
7	Feb 23	Combining Regression and ANOVA -- Categorical explanatory variables --Analysis of Covariance Model (ANACOVA)— permitting interaction between a categorical and a continuous independent variable --Inference for Analysis of Covariance Model -- Interaction between two categorical independent variables	CH 13; RC CH2 CH 12.3
		PROPOSAL DUE --Brief student presentations on research proposals in class	
8	March 2	Model Building with Multiple Regression -- Polynomial Models --Generalized Linear Models	CH 14
9	March 9	Model Building with Multiple Regression -- Exponential and Log Transforms --Model Selection --Regression Diagnostics --Multicollinearity Reading Published Research Papers What can go wrong?	CH 14 Handout PA Ch 3
10	March 16	EXAM	
	March 24	(Spring Break)	
11	March 30	Logistic Regression: Modeling Categorical Response Variables --Binary Response Variable --Linear Probability Model -- Logistic Regression	CH 15 RC CH5
12	April 6	Logistic Regression --Logit Models with Qualitative Explanatory Variables --Interaction models	CH 12, 13, 15; RC CH5

Course Schedule (continued)

■It is essential to **READ** the relevant materials **BEFORE** each class!

Week

13	April 13	Logistic Regression --Model Building with Multiple Logistic Regression	CH 15; RC CH5
14	April 20	Logistic Regression for Ordinal Response Variable	CH 15
15	April 27	Logistic Regression for Polytomous Response Variable -- Multinomial Logistic Regression	CH 15; RC CH6
16	May 4	Student Paper Presentation	
	May 11	PAPER DUE , Monday noon of the exam week (Send the paper through email)	