

**Sociology 605/605L (Spring 2018)**  
**Statistics for Regression Analysis**  
M 12:00-2:30 (SAKAM B414)  
Lab F 12:30-1:20 (CR 220)

Instructor: Yean-Ju Lee (yjlee@hawaii.edu)  
Saunders 216  
Office Hours: M 2:45-4:15 (or by appointment)  
TA: Kitae Park

### Course Description

This course is designed to provide the basic knowledge and skills needed to analyze large-scale social survey data. The focus is on the multiple linear regression model, dealing with a continuous dependent variable, but we also discuss the logistic regression model analyzing a categorical dependent variable. 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 highlight how to manipulate the independent variables and specify the regression models to answer different research questions.

In lab sessions, we use STATA program to perform data analysis. We will use data from the recent American Community Survey (ACS) for the state of Hawaii. (For individual research projects, students are allowed to use ACS data from any other states or any other data sets that they have access to).

### 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

- 1) principles of multivariate statistical analysis
- 2) statistical principles of linear regression models
- 3) statistical principles of logistic regression (binomial, multinomial, & ordered logit models)
- 4) conducting empirical research using linear or logistic regression, using STATA
- 5) ability to read/understand published articles 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. With each absence, the final grade will be lowered by one level.

**Individual Consultation:**

Two individual consultations (most desirably, during the instructor’s office hours) are required: preferably, one in the early weeks before submitting research proposal, and the other in the later part of the semester. Additional, frequent visiting is highly encouraged. Group visiting by students with the same issues is also welcomed.

**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 **8th week (Feb 26)**. 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 in LAB classes) or 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 page long 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 question(s)
- (2) whether data analysis is done **fully**
- (3) whether interpretation of the results is appropriate
- (4) whether the paper has an appropriate organization
- (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 previous 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 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 presentation: to link the lecture to the textbook, to identify important sections
- (3) STATA commands and output: Discussing both STATA commands and output findings
- (4) Example articles that use the relevant techniques, mostly from recent ASR volumes
- (5) Some practice question sets: focusing on interpreting results
- (6) Students are encouraged to ask QUESTIONS 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 8	Overview  Introduction of Data for Individual Research Project: --2014 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 15	(holiday)	
3	Jan 22	Simple Linear Regression and Correlation -- <b>Simple Linear Regression</b> --Least Squares Prediction Equation --Correlation and $R^2$ (Coefficient of Determination) --Inference for the Slope and Correlation --Model Assumptions and Violations	CH 9
4	Jan 29	Introduction to Multivariate Relationships <b>Multiple Regression &amp; Correlation</b> --Multiple Regression Model --Multiple Correlation and $R^2$ (Coefficient of Determination) --Inference for Multiple Regression Coefficients --Partial Correlation	CH 10 CH 11; PA
5	Feb 5	Multiple Regression & Correlation -- <b>Modeling Interaction between two continuous independent variables</b> --Comparing Regression Models --Standardized Regression Coefficients	CH 11
6	Feb 12	Combining Regression and ANOVA -- <b>Categorical explanatory variables</b> --Analysis of Covariance Model (ANACOVA)— <b>permitting interaction between a categorical and a continuous independent variable</b> --Inference for Analysis of Covariance Model -- <b>Interaction between two categorical independent variables</b>	CH 13; RC CH2 CH 12.3

Course Schedule (continued)

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

Week

7	Feb 19	(holiday)	
8	Feb 26	Model Building with Multiple Regression -- <b>Polynomial Models</b> --Generalized Linear Models	CH 14
		<b>PROPOSAL DUE</b> --Brief student presentations on research proposals in class	
9	March 5	Model Building with Multiple Regression -- <b>Exponential and Log Transforms</b> --Generalized Linear Model (GLM) --Model Selection --Regression Diagnostics --Multicollinearity Reading Published Research Papers What can go wrong?	CH 14      Handout PA Ch 3
10	March 12	<b>EXAM</b>	
11	March 19	Logistic Regression: Modeling Categorical Response Variables --Binary Response Variable ( <b>Binomial logit</b> ) --Linear Probability Model --Logit Models with Qualitative Explanatory Variables --Interaction models	CH 15 RC CH5 CH12 CH13
	March 26	(Spring Break)	
12	April 2	Logistic Regression for Ordinal Response Variable ( <b>Ordered logit</b> )	CH 15
13	April 9	Logistic Regression for Polytomous Response Variable ( <b>Multinomial Logit</b> )	RC CH6
14	April 16	Logistic Regression --Model Building with Multiple Logistic Regression	CH 15; RC CH5, 6

Course Schedule (continued)

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

Week

15	April 23	Linear and Logistic Regression --Model Building with Multiple Regression	All readings
----	----------	---	--------------

16	April 30	Student Paper Presentation	
----	----------	----------------------------	--

	May 7	<b>PAPER DUE</b> , Monday noon of the exam week (via <b>email</b> )	
--	-------	--	--