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

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

Reference Text


**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)

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

**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.)

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan 8</td>
<td>Overview&lt;br&gt;Introduction of Data for Individual Research Project: --2014 American Community Survey, Hawaii&lt;br&gt;&lt;br&gt;Sampling, Measurement, Descriptive Statistics CH 1, 2, 3&lt;br&gt;&lt;br&gt;Statistical Inference: Estimation and Significance Test CH 4, 5, 6&lt;br&gt;&lt;br&gt;--Each Chapter, Sections 1, 2, 3, 4 and CH 4, Section 5</td>
</tr>
<tr>
<td>2</td>
<td>Jan 15</td>
<td>(holiday)</td>
</tr>
<tr>
<td>3</td>
<td>Jan 22</td>
<td>Simple Linear Regression and Correlation CH 9&lt;br&gt;&lt;br&gt;Simple Linear Regression&lt;br&gt;&lt;br&gt;--Least Squares Prediction Equation&lt;br&gt;&lt;br&gt;--Correlation and $R^2$ (Coefficient of Determination)&lt;br&gt;&lt;br&gt;--Inference for the Slope and Correlation&lt;br&gt;&lt;br&gt;--Model Assumptions and Violations</td>
</tr>
<tr>
<td>4</td>
<td>Jan 29</td>
<td>Introduction to Multivariate Relationships CH 10&lt;br&gt;&lt;br&gt;Multiple Regression &amp; Correlation CH 11; PA&lt;br&gt;&lt;br&gt;Multiple Regression Model&lt;br&gt;&lt;br&gt;Multiple Correlation and $R^2$ (Coefficient of Determination)&lt;br&gt;&lt;br&gt;--Inference for Multiple Regression Coefficients&lt;br&gt;&lt;br&gt;Partial Correlation</td>
</tr>
<tr>
<td>5</td>
<td>Feb 5</td>
<td>Multiple Regression &amp; Correlation CH 11&lt;br&gt;&lt;br&gt;--Modeling Interaction between two continuous independent variables&lt;br&gt;&lt;br&gt;--Comparing Regression Models&lt;br&gt;&lt;br&gt;--Standardized Regression Coefficients</td>
</tr>
<tr>
<td>6</td>
<td>Feb 12</td>
<td>Combining Regression and ANOVA CH 13; RC CH2&lt;br&gt;&lt;br&gt;Categorical explanatory variables CH 12.3&lt;br&gt;&lt;br&gt;--Analysis of Covariance Model (ANACOVA)—permitting interaction between a categorical and a continuous independent variable&lt;br&gt;&lt;br&gt;--Inference for Analysis of Covariance Model&lt;br&gt;&lt;br&gt;Interaction between two categorical independent variables</td>
</tr>
</tbody>
</table>
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  CH 14
        --*Polynomial Models*
        --Generalized Linear Models

**PROPOSAL DUE**
--Brief student presentations on research proposals in class

9   March 5  Model Building with Multiple Regression  CH 14
        --*Exponential and Log Transforms*
        --Generalized Linear Model (GLM)
        --Model Selection
        --Regression Diagnostics
        --Multicollinearity
        Reading Published Research Papers  Handout
        What can go wrong?  PA Ch 3

10  March 12 EXAM

11  March 19 Logistic Regression: Modeling Categorical Response Variables  CH 15
    --Binary Response Variable (*Binomial logit*)
    --Linear Probability Model  RC CH5
    --Logit Models with Qualitative Explanatory Variables  CH12
    --Interaction models  CH13

12  April 2  Logistic Regression for Ordinal Response Variable  CH 15
        (*Ordered logit*)

13  April 9  Logistic Regression for Polytomous Response Variable  RC CH6
        (*Multinomial Logit*)

14  April 16 Logistic Regression  CH 15; RC CH5, 6
    --Model Building with Multiple Logistic Regression
Course Schedule (continued)

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

Week

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

16 April 30 Student Paper Presentation

May 7 PAPER DUE, Monday noon of the exam week
       (via email)