The University of North Carolina at Chapel Hill
School of Social Work

SOWO 919: Applied Regression Analysis and Generalized Linear Models
Spring Semester, 2008

Instructor
Shenyang Guo, Ph.D.,
Room 524j, Tate Turner Kuralt Building
CB #3550,
School of Social Work, Chapel Hill, NC 27599-3550
Phone: (919) 843-2455
Email: sguo@email.unc.edu

Class Meeting Times & Office Hours
Class meets on Wednesdays 9:00-11:50 am
Office hours are Tuesdays 8:30 – 10:30 (Room 524j TTK)

Course Description:
This course introduces statistical frameworks, analytical tools, and social behavioral applications of OLS regression model, weighted least-square regression, logistic regression models, and generalized linear models.

Course Objectives:
At the completion of the course, students will be able to:

1. Understand the type and nature of research questions and data that are suitable for regression analysis;
2. Use Stata computing software package to manage and analyze data with the OLS regression model;
3. Understand the Gauss-Markov theorem and the BLUE property of OLS, especially conditions under which BLUE does not hold;
4. Have a solid understanding of the five assumptions embedded in the OLS regression;
5. Know how to conduct statistical tests detecting violations of OLS assumptions (i.e., multicollinearity, heteroskedasticity, influential data and outliers, etc.);
6. Know how to take remedial measures if harmful violations exist (i.e., weighted least-squares regression, etc.);
7. Understand the type and nature of research questions and data that are suitable for the generalized linear models;
8. Have a solid understanding of basic concepts of categorical data (i.e., odds ratio, relative risk, marginal probability, and conditional probability);
9. Use Stata computing software package to manage and analyze data with the binary, ordered, and multinomial logistic regressions;
10. Know how to interpret results of regression analysis and logistic regression analysis, and communicate findings to general audiences clearly and effectively in writing;
11. Understand limitations of the regression and logistic regression models, and common pitfalls in using these models;
12. Understand the basics of conducting a Monte Carlo study.

**Pre-requisite:**
Students are assumed to be familiar with descriptive and inferential statistics. They should have statistical and statistical software background at least equivalent to that provided by SOWO 911. Students without such prerequisites should contact the instructor to determine their eligibility to take the course.

**Statistical Software Package:**
This course will use Stata as the main software package.

**Required Textbook:**

**Required Articles**
All required journal articles are available on E-journals. Required book chapters will be distributed in class.

**Recommended Textbooks:**

<table>
<thead>
<tr>
<th>Assignments</th>
<th>Grade Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment 1</td>
<td>10%</td>
</tr>
<tr>
<td>Assignment 2</td>
<td>10%</td>
</tr>
<tr>
<td>Assignment 3</td>
<td>10%</td>
</tr>
<tr>
<td>Assignment 4</td>
<td>10%</td>
</tr>
<tr>
<td>Assignment 5</td>
<td>10%</td>
</tr>
<tr>
<td>Midterm Exam (take home)</td>
<td>25%</td>
</tr>
<tr>
<td>Final Exam (take home)</td>
<td>25%</td>
</tr>
</tbody>
</table>

**Grading System**
The standard of School of Social Work’s interpretation of grades and numerical scores will be used.

- H = 94-100
- P = 80-93
- L = 70-79
- F = 69 and below

**Policy on Incomplete and Late Assignments**
Assignments are to be turned in to the professor by 5pm of the due date noted in the course outline. Extensions may be granted by the professor given advance notice of at least 24 hours. Late assignments (not turned in by 5pm on the due date) will be reduced 10 percent for each day late (including weekend days). A grade of incomplete will only be given under extenuating circumstances and in accordance with University policy.
Policy on Academic Dishonesty
Students are expected to follow the UNC Honor Code. Please include the honor code statement along with your signature on all assignments:

“I have neither given nor received unauthorized aid on this assignment.”

Please refer to the APA Style Guide, the SSW Manual, and the SSW Writing Guide for information on attribution of quotes, plagiarism and appropriate use of assistance in preparing assignments.

If reason exists to believe that academic dishonesty has occurred, a referral will be made to the Office of the Student Attorney General for investigation and further action as required.

Policy on Accommodations for Students with Disabilities
Students with disabilities which affect their participation in the course may notify the instructor if they wish to have special accommodations in instructional format, examination format, etc., considered.

Course Outline (Topics, readings, and assignments):

Jan. 9 Introduction and Review
Course overview
Review of important concepts:
Association versus causation
Univariate, bivariate, and multivariate analyses
Statistical properties
Causal inferences
Summation operator Σ
Readings:
Neter et al: 1.1-1.2

Jan. 16 Simple Linear Regression
Optimization: minimization versus maximization
Least-square estimator
Interval estimation and hypothesis testing (t-test)
Readings:
Kutner et al: 1.3-1.7, Chapters 2 & 4

Jan. 23 Basic Matrix Algebra
Why matrix?
Matrix operations
Example: solve for \( b = (X'X)^{-1}(X'Y) \)
Stata Lab 1: Stata basics and running regression
Readings:
Assignment 1 out (Due: 1-30-08):
In this Assignment, you will solve problems pertaining to simple linear regression and basic matrix algebra. You will run Stata to solve some of the problems.

Jan. 30 Multiple Linear Regression
Model specifications
Estimation of regression coefficients
Inferences concerning $\beta$
Readings:
Kutner et al: Chapter 6 (the first half)

Feb. 6 The ANOVA Table and $R^2$
Review of variance, covariance and correlation
Decomposition of total sum of squares, F test
$R^2$ and adjusted $R^2$
Readings:
Kutner et al: Chapter 6 (the second half)

Assignment 2 out (Due: 2-20-08):
In this Assignment, you will solve problems pertaining to multiple linear regression models. You will run Stata to solve some of the problems.

Feb. 13 Properties of OLS and the Five Assumptions
BLUE criterion and properties of OLS
Maximum likelihood estimator
The five OLS assumptions
Stata Lab 2: Running multiple regression
Readings:
Kutner et al: 1.8

Feb. 20 Stata Lab 3: Diagnostic Tests in Regression Analysis
Residual analysis
Looking for influential data
Multicollinearity diagnostics – Variance Inflation Factor
A Monte Carlo study demonstrating OLS properties
Readings:
Kutner et al: Chapter 3

**Assignment 3 out (Due: 2-27-08):**
In this Assignment, you will conduct a Monte Carlo study that shows consequences of OLS regression under various conditions of data generation. Running Stata is required.

**Feb. 27**  
**Violating OLS Assumptions and Remedial Measures – I**  
Specification errors and selection of Predictors  
Influential data and outliers  
Multicollinearity  
Model validation  

**Readings:**  
Kutner et al: Chapters 9 and 10

**Midterm exam out (Due: 3-19-08):**
Use data sets provided by the course or data set you choose to run a multiple linear regression model. Write a paper (no more than 14 pages, double spaced) to present findings. The paper should include: (1) research questions and data that are suitable to a regression analysis; (2) methods and specification of the regression model; (3) diagnostics detecting at least two problems pertaining to violation of regression assumptions and discussion of remedial measures; (4) interpretation of findings; and (5) presentation of findings that answers research questions effectively and efficiently.

**Mar. 5**  
**Violating OLS Assumptions and Remedial Measures – II**  
Heteroskedasticity  
Weighted Least Squares  

**Readings:**  
Kutner et al: Chapter 11  

**Mar. 12**  
**Happy Spring Break! No Class**

**Mar. 19**  
**Other Topics in Regression Analysis -- I**  
Regression through origin  
Partial-correlation coefficient  
Standardized regression coefficient  
Functional form, curvilinear relationship and polynomial regression models  
$R^2$ increment: Hierarchical regression analysis  

**Readings:**  
Kutner et al: Chapter 7  
Other Topics in Regression Analysis – II

- Dummy variables as predictors:
  - Interpretation of regression coefficients
  - Comparison of several regression equations
- Testing interactions:
  - Mediator versus moderator
  - Interaction, joint effect and moderator
- ANCOVA
  - Experimental design
  - Concomitant variables

Stata Lab 4: Diagnostics of OLS regression

Readings:
- Kutner et al: Chapter 8

Assignment 4 out (Due: 4-9-08):
In this Assignment, you are required to solve problems pertaining to interpretation of standardized regression coefficients, diagnostics of heteroskedasticity, performing partial F test, and testing interactions. Some of the problems require running Stata.

Logistic Regression Analysis -- I

- Violations of OLS assumption when dependent variable is dichotomous
- Logistic response function
- Model specifications
- Wald statistic and goodness-of-fit indices
- Diagnostic methods

Readings:
- Kutner et al: Chapter 13

Logistic Regression Analysis -- II

- Odds ratio
- Relative risk
- Predicted probability

Stata Lab 5: Running logistic regression models

Readings:
- Kutner et al: Chapter 14

Assignment 5 out (Due: 4-16-08):
In this Assignment, you are required to solve problems pertaining to interpretation of odds ratio and running binary logistic regression. You will run Stata to solve some of the problems.

Logistic Regression Analysis -- III

- Multinomial logistic regression
- Ordinal logistic regression
Overview of the generalized linear models

Readings:


Final exam out (Due: 5-1-08):
Use data sets provided by the course or data set you choose to run a binary logistic regression model. Write a paper (no more than 14 pages, double spaced) to present findings. The paper should include: (1) research questions and data that are suitable to a logistic regression; (2) methods and specification of the logistic regression model; (3) diagnostics detecting potential problems in your data that might violate the logistic regression assumptions; (4) tests of interactions; (5) interpretation of findings; and (6) presentation of findings that answers research questions effectively and efficiently.

Apr. 23  
**A Critical Review of the Applications of Regression**

Common pitfalls in conducting and reporting regression analysis

Why Berk claims that “the practice of regression analysis and its extension is a disaster”?

What to do?

Where to go after this course?

Readings:


May 1  
Final paper due