【COURSE TITLE】
Statistical Analysis for Social & Behavioral Research

【INSTRUCTOR】
姓 名: Ce Shen
国 籍: U.S.A.
职 称: Associate Professor
来 自 学 校: Boston College
Email 地 址: shenc@bc.edu

【授课对象 STUDENTS】
First-year Graduate Students

【授课语言 TEACHING LANGUAGE】
English/Chinese

【先修课程 PREREQUISITES】
Basic knowledge of research methods and statistics in social sciences

【授课方式 METHODS OF INSTRUCTION】
An introductory lecture for each topic will be presented, followed by a review of example(s) of the syntax and output generated by STATA, and the interpretation of coefficients and other relevant statistics. The course will be taught in a lab with computer for each student and students are encouraged to actually analyze data in class from time to time.

【课堂规模 CLASSROOM CAPACITY】
Flexible (20 students)

【课程考核 EVALUATION】
Assignment 1: Linear multiple-regression analysis using data provided
Assignment 2: Multiple-regression analysis with categorical dependent variable using data provided.

【课程学分 CREDITS】
2 学分  2 credits

【教师简介 ABOUT THE INSTRUCTOR】
Ce Shen has a Ph.D. in sociology and currently an Associate Professor at Boston College. His work has been based on the cross-national comparative studies in the fields of social development, mental health services research, aging, and advanced statistical modeling.

【课程简介 COURSE DESCRIPTION】

This applied course is designed for graduate students with a prior background in basic statistics and research methods. This means that students must know the basic principles of statistical inferences: standard errors, confidence intervals, hypothesis tests, p-values, and so on. They also should have experience to perform basic statistical tests such as t-test, correlation analysis, analysis of variance, chi-square and other nonparametric test using some statistical software. This course will provide solid and comprehensive training in multiple regression-based methods for analyzing quantitative social and behavioral science data. Students will learn how to do ordinary linear regression with a statistical software package -- Stata. You will also learn to check on how well the data meet the assumptions of OLS regression and do relevant diagnostics. These assumptions include linearity, normality, homogeneity of variance (homoscedasticity), independence, model specification and other issues of great concern to data analysis. This course will also cover binary and ordinal logistic regression using Stata. The emphasis of this course is on application, focusing on estimating models and interpreting the results, rather than understanding in detail the mathematics behind the techniques. We will also discuss and critically evaluate published research based on the various techniques covered in this course. The goal of this course is to provide students with the knowledge and skills necessary to critically evaluate social research using OLS, to enable students to perform multiple regression analysis, and to be prepared for taking advanced courses such as structural equation modeling, hierarchical linear modeling and longitudinal data analysis.

【课程内容及教学进度 SYLLABUS】

SESSION 1: Introduction to STATA

Objectives:
To develop basic working techniques for using STATA to fit and interpret various statistical analysis models for this course. Students will become familiar with STATA interface, data format, basic syntax of STATA commands, and basic data management skills, including saving output into a log file, getting help for using STATA, recoding, creating new variables, labeling variables and values, creating simple graphs for univariate and bivariate analyses.

Required Readings:

SESSION 2: Bivariate Regression

Objectives:
To be able to perform correlation and simple regression analysis using STATA. Upon completion of this section of the course, students should understand such concepts as: Pearson R, factors influencing the magnitude and sign of Pearson R, slope, intercept, least squares (best fitting) line, R-square, adjusted R-square, non-standardized and standardized regression coefficients, beta weights, t values, residuals, basic assumptions.

Required Readings:

Chapter 2: Planning a quantitative research project with existing Data (pp. 24-38)
Chapter 5: Basic concepts of bivariate regression (pp. 94-151)
Chapter 9: Correlation and simple regression (pp. 197-222)

**Supplemental Readings:**

Chapter 2: Simple linear regression and correlation (pp. 15-42)

**SESSIONS 3: Multiple Regression Analysis with Two Continuous Predictors**

**Objectives:**
To be able to perform multiple regression analysis (MRA) with two continuous predictors, using STATA. Upon completion of this section of the course, students should understand such concepts as: partial regression coefficient, the coefficient of determination, partial correlation, and part (semipartial) correlation. You will also learn how to use Stata to check variables’ normality, do appropriate transformation and conduct regression with transformed variables.

**Required Readings:**

Chapter 6: Basic concepts of multiple regression (pp. 152-196)
Chapter 10: Indirect effects and omitted bias (pp. 337-359)

Chapter 10: Multiple regression (pp. 223-232)

**Supplemental Readings:**

Chapter 5: Summary statistics and tables (pp. 123-142)
Chapter 7: Linear regression analysis (pp. 163-214)

Chapter 13: Multiple linear regression (pp. 237-274)

Chapter 5: Elements of multiple regression analysis: Two independent variables (pp. 95-122)
Chapter 7: Statistical control: Partial and semi-partial correlation (pp. 156-194)
Chapter 8: Prediction (pp. 195-240)

SESSIONS 4-5: Multiple Regression Analysis with More Than Two Predictors Including Categorical Variables

Objectives:
To be able to perform multiple regression analysis (MRA) with continuous and categorical predictors using STATA. Upon completion of this section of the course, students should understand such concepts as: hierarchical multiple regression, stepwise multiple regression, forward inclusion, backward elimination, collinearity and multicollinearity, tolerance, residual analysis, the construction and use of dummy variables.

Required Readings:

Chapter 7: Dummy variables (pp. 197-247)

Chapter 10: Multiple regression (pp. 223-262)

[Note: This article illustrates the treatment of aggregate data analysis, especially with relative small N in multiple regression analysis.]

Supplemental Readings:

Chapter 11: A categorical independent variable: Dummy, effect, and orthogonal coding (pp. 340-409)
Chapter 12: Multiple categorical independent variables and factional designs (pp. 410-512)

SESSION 6: Analysis of Covariance (ANCOVA)

Objectives:
To be able to perform analysis of covariance (ANCOVA) using STATA. ANCOVA could be considered as a combination of multiple regression and ANOVA—making comparison between two or more group means, after statistically removing the effect of one or more variables (called covariates) on the dependent variable. Upon completion of this section of the course, students should understand such concepts as: covariates, adjusted means, homogeneity of regression across groups, and so on. Students should be able to select appropriate covariates, check assumptions of ANCOVA, and interpret the Stata output from ANCOVA.

**Required Readings:**


[Note: This article illustrates the treatment of categorical predictors and the treatment of analysis of covariance.]


Chapter 11: Analysis of covariance, multivariate ANOVA, and repeated multivariate analysis (pp. 263-276)

**SESSION 7: Multiple Regression Analysis with Interaction Effects**

**Objectives:**

To be able to perform multiple regression analysis (MRA) with interactive terms using STATA. Upon completion of this section of the course, students should understand such concepts as product interaction terms, main effects, and interaction effects. Students should be able to interpret the regression coefficients from the model and to use Stata to conduct basic OLS regression with interaction effects from two categorical variables, or one categorical and the other continuous, or two continuous variables.

**Required Readings:**


Chapter 8: Interactions (pp. 248-305)


[Note: This article illustrates the treatment of dummy coding and interaction effects in multiple regression analysis.]

**Supplemental Readings:**
SESSION 8: Multiple Regression Analysis with Nonlinear Relationships

Objectives:
Upon completion of this section of the course, students should understand such concepts as: the basic polynomial model and the exponential model. To be able to interpret and perform basic nonlinear multiple regression analysis (MRA) using STATA.

Required Readings:

Chapter 9: Nonlinear relationships (pp. 306-335)

[Note: This article illustrates the treatment of exploratory factor analysis and nonlinear regression analysis.]

Supplemental Readings:

Chapter 3: Levels and rates of change in socioeconomic development (pp. 21-42)
[Note: This chapter illustrates the treatment of non-linear regression analysis.]

Chapter 15: Questioning the greatness of straightness: Non-linear relationships
SESSION 9: Regression Diagnostics

Objectives:
Upon completion of this section of the course, students should be able to conduct basic OLS regression diagnostics, including detecting and dealing with problems such as multicollinearity, nonadditivity, and heteroscedasticity, influential observations and outliers using Stata.

Required Readings:

- Chapter 4: Multicollinearity (pp. 37-50)
- Chapter 5: Nonlinearity and nonadditivity (pp.51-72)
- Chapter 6: Heteroscedasticity and autocorrelation (pp.73-87)

- Chapter 11: Outliers, heteroskedasticity, and multcollinearity (pp.360-398)

- Chapter 10: Multiple regression (pp. 223-262)

Supplemental Readings:


- Chapter 3: Regression diagnostics (pp. 43-61)

SESSIONS 10, 11, 12, 13: Regression Models for Categorical Dependent Variables Using STATA

Objectives:
Logistic regression analysis includes outcomes that are binary, ordinal, nominal, or count
dependent variables. If students apply the linear regression model, estimates might be biased, inefficient, or simply inappropriate. Students will understand such concepts as: the logistic regression curve, maximum likelihood, Wald and likelihood ratio (LR) tests, logit coefficients, Wald statistic, odds ratio, and assessing goodness of fit of the model. Students will become familiar with Stata commands for performing logistic regression analysis and interpret the Stata output for the models.

**Binary outcomes**
Regression models for binary outcomes are the foundation from which more complex models for ordinal, nominal, and count models can be derived. Ordinal and nominal regression models are equivalent to the simultaneous estimation of a series of binary outcomes. Although the link between binary and count models is less direct, the Poisson distribution for count models can be derived as the outcome of many binary trials. For binary outcome models, we’ll cover

- Predicted probabilities and binary outcomes
- The linear probability model
- Problems with LPM
- Maximum likelihood estimation
- Odds ratio in logistic regression
- Classification in logistic regression
- Predictors in logistic regression: Continuous, categorical predictors and interaction terms
- Significance tests of overall model and individual predictors
- Relative importance of predictors
- Sample size and effect size
- Presentation of logistic regression in research report

**Ordinal outcomes**
The binary regression model can be viewed as a special case of the ordinal model in which the ordinal outcome has only two categories. Students will expand their understanding and grasp of Stata commands from binary model to ordinal models and be able to perform ordinal logistic regression analysis and interpret the Stata output for the models.

**Count outcomes**
Count variables indicate how many times something has happened. Although the linear regression model has often been applied to count outcomes, this can result in inefficient, inconsistent, and biased estimates. This topic will introduce Poisson (PRM) and negative binomial (NBRM) regression models, which are the foundation for other count models.

**Required Readings:**

Chapter 3: Estimation, testing, fit, and interpretation (pp.75-128)
Chapter 4: Models for binary outcomes (pp.131-181)
Chapter 5: Models for ordinal outcomes (pp.183-222)
Chapter 8: Models for count outcomes (pp.349-381)


Chapter 12: Logistic regression (pp. 306-332)


**Supplemental Readings:**


Chapter 9: Logistic regression (pp.251-281)


**【课程教材 Textbooks】**


**【参考读物 Reference Readings】**


