ICPSR Summer Program in Quantitative Methods of Social Research

Linear Regression Analysis in the Social Sciences

Instructors: Patrick Shea and M. C. Sunny Wong

Day and Time: 5/13 (Mondays) – 5/17 (Friday), 9:00 AM – 5:00 PM

Location: TBA

E-mails: pshea@central.uh.edu (Shea) and mwong8@central.uh.edu (Wong)

Teaching Assistants: TBA

Course Description

This course is designed for participants to develop quantitative research skills with applications to social science topics. Participants will gain an overview of research design, data management, and statistical analysis and interpretations of research findings.

The course will be centered around several main topics covering the basic analysis of ordinary least squares (OLS), the technique of estimating bivariate and multivariate regression models, the overall fitness of a regression equation, and the hypothesis and diagnostic testings, and more.

This course takes the "learning by doing" approach by discussing the major themes in regression analysis with detailed examples, which show how the subject works in practice using STATA.

Pre-requisites:

All participants in this course are expected to have a working knowledge of elementary statistical concepts.

Readings, Textbooks, and Statistical Software

These are the main textbooks for the course:


The following are the recommended texts for the course:


This course requires the use of computers for most assignments. We will use a statistical program called STATA and will provide you with STATA licenses so that you can install STATA in your computers.
Preliminary Schedule

DAY 1 - Statistics Review and Introduction to OLS

- Morning Session (9:00 PM - Noon)
  - Welcome and Introductions
  - Statistics Review (Printed Lecture Notes)
    - Scales of Measurement
    - Descriptive Statistics
    - Population vs. Sample
    - Measures of Location, Variability and Association between Two Variables
    - Normal Probability Distribution
    - Sampling and Sampling Distribution
    - Interval Estimation and Hypothesis Testing
  - STATA Lab #1

- Afternoon Session (1:00 PM - 5:00 PM)
  - Introduction to Regression Analysis – (Studenmund, Chapter 1)
  - Ordinary Least Squares (OLS) – (Studenmund, Chapter 2)
    - Characteristics of OLS
    - Calculating Coefficients for a bivariate OLS model
    - Interpreting Coefficients
    - Describing the Overall Fit of the Estimated Model
  - STATA Lab #2

DAY 2 - OLS Estimations, the Classical Model, Hypothesis Testing

- Morning Session (9:00 PM - Noon)
  - Learning to Use Regression Analysis – (Studenmund, Chapter 3)
    - Dummy Variables in OLS Regressions
    - Interaction terms in OLS Regressions
  - The Classical Model – (Studenmund, Chapter 4)
    - The Classical Assumptions and the Best Linear Unbiased Estimator (BLUE)
    - Sampling and Sampling Distribution
    - The Gauss-Markov Theorem
  - STATA Lab #3

- Afternoon Session (1:00 PM - 5:00 PM)
  - Hypothesis Testing and Statistical Inference – (Studenmund, Chapter 5)
    - The t-test
    - Confidence Intervals
    - The F-test
  - Specification – (Studenmund, Chapters 6 and 7)
    - Choosing the Independent Variables
    - Choosing a Functional Form
  - STATA Lab #4
DAY 3 - Potential Problems in Multiple Regression Analysis

- Morning Session (9:00 PM - Noon)
  - Multicollinearity – (Studenmund, Chapter 8)
    - Variance Inflation Factor (VIF)
  - STATA Lab #5
- Afternoon Session (1:00 PM - 5:00 PM)
  - Heteroskedasticity – (Studenmund, Chapter 9)
    - Breusch-Pagan Test for Heteroskedasticity
    - White Test for Heteroskedasticity (ssc install)
  - Serial Correlation – (Studenmund, Chapter 10)
    - Breusch-Godfrey LM Test for Serial Correlation
    - Generalized Least Squares - Prais-Winsten Estimator
    - Newey-West Standard Errors
  - STATA Lab #6

DAY 4 - Topics of Regression Analysis I

- Morning Session (9:00 PM - Noon)
  - Time-Series Models – (Studenmund, Chapter 12)
    - Distributed Lag Models and Dynamic Models
    - Granger Causality
    - Spurious Correlation and Nonstationarity
  - STATA Lab #7
- Afternoon Session (1:00 PM - 5:00 PM)
  - Forecasting – (Studenmund, Chapter 15)
    - ARMA Models - The Box–Jenkins (BJ) Methodology
    - Identification of an ARMA Series
  - STATA Lab #8

DAY 5 - Topics of Regression Analysis II

- Morning Session (9:00 PM - Noon)
  - Panel-Data Models – (Studenmund, Chapter 16)
    - Pooled OLS Regressions
    - Fixed-Effect Models
    - Random-Effect Models
  - STATA Lab #9
- Afternoon Session (1:00 PM - 5:00 PM)
  - Models of Discrete and Limited Dependent Variables – (Studenmund, Chapter 13)
    - The Linear Probability Model (LPM)
    - The Logit and Probit Models
    - Ordered Logit and Probit Models
  - STATA Lab #10