Statistics and Data Analysis II: Intermediate

ICPSR Summer Program, 2016

Instructor
Shane P. Singh (University of Georgia)
Office: 211 Newberry Hall
Office Hours: 5:15-6:45PM
Email: singh@uga.edu

Teaching Assistants
Chase Meyer (University of Georgia)
Office: 212 Newberry Hall
Office Hours: TBA
Email: chasebm@uga.edu

Bryce Williams-Tuggle (Vanderbilt University)
Office: 214 Newberry Hall
Office Hours: TBA
Email: bryce.g.williams-tuggle@vanderbilt.edu

Class Meeting Time: 3:00PM-5:00PM

Class Location: TBA

Description and Goal of the Course: This course is an introduction to ordinary least squares (OLS) bivariate and multiple regression, and it assumes familiarity with introductory statistics, probability, and basic algebra. OLS, while useful in and of itself, also serves as the foundation of many more advanced data analytical techniques. The primary goal of the course is to develop a practical, applied, and intuitive (rather than theoretical or mathematical) understanding of OLS regression. Participants should leave with a solid working foundation of the application of OLS, the ability to use it in their own research, and the skills to build on it in their further study of more advanced statistical topics.

Required Readings:


The books are available at Ulrich’s Bookstore (549 E. University Ave.) and online.

Course Content:

Readings: Be sure to complete the readings prior to each class period. The content of the readings will not always be discussed in class due to the amount of material we
will cover in just four weeks. However, it is important that you comprehend the content because it is fundamental to your ability to meaningfully understand and use regression analysis.

**Daily Class Meetings:** Lectures will complement the readings. In some cases, material from the readings will not be covered. And, in many cases, we will cover material not in the readings. We will also frequently analyze real data.

**Weekly Lab Sessions:** The weekly lab sessions will provide you the opportunity to get “hands-on” experience with course content and to receive guidance on the practice assignments and homework assignments. These sessions will also serve as an opportunity to become familiar with statistical software. Information about the time and location of the lab sessions will be provided.

**Course Homework Assignments:** In addition to some short practice assignments, you will have four more substantial assignments, which will be graded. The assignments will require statistical software. If necessary, schedule time in ICPSR’s labs so you can access the software you need. All assignments will be available electronically on CTools.

The following is required for each homework assignment:

- Type and staple your homework (hand calculations can be written out).
- Present exercises in numerical order, and label your answers clearly.
- Embed any figures or tables within the homework; do not attach them at the end.

**Grade Breakdown:**

- Assignment 1: 20%; **Due on Day 8**
- Assignment 2: 30%; **Due on Day 13**
- Assignment 3: 30%; **Due on Day 16**
- Assignment 4: 20%; **Due on Day 19**

**A Note on Software:** While the in-class data analysis will be conducted with Stata, students are welcome to complete their assignments using Stata, R, or SPSS. (Please get our permission if you wish to use another software package.) The instructor and/or teaching assistants will be available to assist students with the software during the lab sessions.

**Readings and Course Schedule:** The course syllabus is a general plan for the course; deviations announced to the class by the instructor may be necessary.

**WEEK 1**

**Day 1: Introduction and Review of Statistical Inference**

**Day 2: Association between Two Variables**

- Agresti and Finlay, Chapter 8
Day 3: Constructing and Interpreting a Bivariate OLS Regression
  • Agresti and Finlay, Chapter 9, Sections 9.1-9.4
  • Lewis-Beck and Lewis-Beck, Chapter 1

Day 4: Statistical Inference with Bivariate OLS Regression
  • Agresti and Finlay, Chapter 9, pgs. 278-282
  • Lewis-Beck and Lewis-Beck, Chapter 2, pgs. 29-39

Day 5: The Assumptions of OLS Regression
  • Agresti and Finlay, Chapter 9, pgs. 276-277 and Section 9.6
  • Lewis-Beck and Lewis-Beck, Chapter 2, pgs. 23-28

WEEK 2

Day 6: Applying What We’ve Learned to This Point

Day 7: Confounding Factors, Causality, and Control
  • Agresti and Finlay, Chapter 10

Day 8: Multiple Regression
  • Agresti and Finlay, Chapter 11, pgs. 321-334
  • Lewis-Beck and Lewis-Beck, Chapter 3, pgs. 55-64 and 72-73

Day 9: Collinearity and Comparing the Effects of Variables in Multiple Regression
  • Agresti and Finlay, Chapter 11, pgs. 334-339 and Section 11.8
  • Lewis-Beck and Lewis-Beck, Chapter 4, pgs. 75-78 and 83-86

Day 10: Categorical Independent Variables
  • Agresti and Finlay, Chapter 12, pgs. 379-381; Chapter 13, Section 13.2
  • Lewis-Beck and Lewis-Beck, Chapter 3, pgs. 64-69

WEEK 3

Day 11: Interactions between Variables 1
  • Lewis-Beck and Lewis-Beck, Chapter 3, pgs. 69-71

Day 12: Interactions between Variables 2
  • Agresti and Finlay, Chapter 11, Section 11.5; Chapter 13, Sections 13.3-13.4

Day 13: Applying What We’ve Learned to This Point

Day 14: Model Selection
  • Agresti and Finlay, Chapter 11, Section 11.6; Chapter 14, Section 14.1
Day 15: Regression Diagnostics 1
- Agresti and Finlay, Chapter 14, pgs. 448-452 and Section 14.3
- Lewis-Beck and Lewis-Beck, Chapter 2, pgs. 44-49; Chapter 4, pgs. 79-83

WEEK 4

Day 16: Regression Diagnostics 2
- Agresti and Finlay, Chapter 14, pgs. 452-456 and Sections 14.5-14.6
- Lewis-Beck and Lewis-Beck, Chapter 2, pgs. 39-44 and 49-52; Chapter 4, pgs. 86-88

Day 17: Categorical Dependent Variables
- Agresti and Finlay, Chapter 15, Sections 15.1-15.5

Day 18: Applying What We’ve Learned to This Point

Day 19: Catch-Up (if Necessary), Wrap-Up, and What’s Next?
- Lewis-Beck and Lewis-Beck, Chapter 4, pgs. 93-95