Statistics and Data Analysis II: The Basics of Regression
ICPSR Summer Program, 2024

Instructor
Shane P. Singh (University of Georgia)
Office: 4201 ISR Building and Canvas
Office Hours: 5:00-6:15PM (Eastern Time Zone)
Email: singh@uga.edu

Teaching Assistant
Poshan Dahal (Case Western Reserve University)
Office: ISR Building and Canvas
Office Hours: 9:00-11:00AM (Eastern Time Zone)
Email: pxd205@case.edu

Class Meeting Time: 1:30-4:30PM (Eastern Time Zone)

Class Location: 2306 Mason Hall

Description and Goal of the Course: This course introduces ordinary least squares (OLS) bivariate and multiple regression, and it assumes familiarity with introductory statistics, probability, and basic algebra. The course is open to students at any level, as well as professionals and faculty members. 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 strictly theoretical or mathematical) understanding of OLS regression. The course will also cover what OLS regression can and cannot do with regard to causal inference, and we will briefly discuss machine learning models for 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.

Readings:

Agresti, Alan. 2018. *Statistical Methods for the Social Sciences*. 5th ed. Boston: Pearson.

Lewis-Beck, Colin, and Michael Lewis-Beck. 2016. *Applied Regression: An Introduction*. 2nd ed. Thousand Oaks, CA: SAGE Publications.

The books are available for purchase online. Limited copies are also available from the Summer Program library.

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 three weeks. However, the readings and the in-class material will usually overlap. Sometimes, the readings will handle concepts differently from how

they are discussed in class. This divergence is often beneficial for students, as learning new statistical topics from multiple angles can facilitate comprehensive understanding.

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

Lab Sessions: The lab sessions, which will be fully online, will provide you the opportunity to get "hands-on" experience with course content and to receive guidance on in-lab practice exercises and the required homework assignments. These sessions will also serve as an opportunity to become familiar with statistical software. Information about the timing of the lab sessions will be provided.

Homework Assignments: In addition to the practice assignments covered in the labs, there will be three more substantial assignments, which will be graded. These are mandatory for participants taking the course for credit. For others, these are optional but encouraged. The assignments will require statistical software, which is freely available for use from the Summer Program and the University of Michigan. All assignments will be made available via Canvas.

The following is required for each homework assignment:
- Type up your homework. Hand calculations can be written out electronically or inserted into your homework as a scan or photograph.
- 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: 30%; **Due on Day 5**
- Assignment 2: 30%; **Due on Day 8**
- Assignment 3: 40%; **Due on Day 14**

A Note on Software: While the in-class data analysis and lab sessions 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 assistant(s) will be available to help students with the software during the lab sessions and in office hours.

Readings and Schedule:

WEEK 1

Day 1: Introduction and Review of Statistical Inference

Day 2: Bivariate Association and a First Look at Regression
- Agresti: Chapter 8 and Chapter 9, Sections 9.1-9.4
- Lewis-Beck and Lewis-Beck: Chapter 1

Day 3: Statistical Inference with OLS Regression and the OLS Assumptions
- Agresti: Chapter 9, Sections 9.5 and 9.6
- Lewis-Beck and Lewis-Beck: Chapter 2, pgs. 23-38

Day 4: Applying What We've Learned to This Point

WEEK 2

Day 5: Confounding Factors, Causality, and Control—on to Multiple Regression
- Agresti: Chapter 10; Chapter 11, pgs. 307-319; Chapter 13, Section 13.3
- Lewis-Beck and Lewis-Beck: Chapter 3, pgs. 55-64 and 72-73

Day 6: Comparing Coefficients and Categorical Independent Variables
- Agresti: Chapter 11, Section 11.7; Chapter 12, Section 12.1; Chapter 13, pgs. 387-392
- Lewis-Beck and Lewis-Beck: Chapter 3, pgs. 64-69; Chapter 4, pgs. 83-86

Day 7: Interactions between Variables
- Lewis-Beck and Lewis-Beck: Chapter 3, pgs. 69-71
- Agresti: Chapter 11, Section 11.4; Chapter 13, pgs. 392-397

Day 8: Applying What We've Learned to This Point

Day 9: Model Selection by Theory and by Machine
- Agresti: Chapter 11, pgs. 320-322, Section 11.5; Chapter 14, Section 14.1

WEEK 3

Day 10: Regression Diagnostics 1
- Agresti: Chapter 11, pgs. 319-320; Chapter 14, pgs. 426-429, Section 14.3, and pgs. 448-449
- Lewis-Beck and Lewis-Beck: Chapter 2, pgs. 44-49; Chapter 4, pgs. Chapter 4, pgs. 75-83

Day 11: Regression Diagnostics 2
- Agresti: Chapter 14, pgs. 430-433, Sections 14.5-14.6, and pgs. 449-450
- Lewis-Beck and Lewis-Beck: Chapter 2, pgs. 39-44 and 49-52; Chapter 4, pgs. 86-88

Day 12: Categorical Dependent Variables
- Agresti: Chapter 5, Section 5.5; Chapter 14, pages 435-436; Chapter 15, Sections 15.1-15.5

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

Day 14: OLS for Causal Inference, Wrap-Up, and Future Directions
- Agresti: Chapter 13, Section 13.5; Chapter 16, Sections 16.2 and 16.8
- Lewis-Beck and Lewis-Beck: Chapter 4, pgs. 93-95