

# Statistics and Data Analysis II: The Basics of Regression

## ICPSR Summer Program, 2026

### Instructor

Shane P. Singh (University of Georgia)

Office: ISR Building and Canvas

Office Hours: 4:15–5:30PM (Eastern Time Zone) and by appointment

Email: [singh@uga.edu](mailto:singh@uga.edu)

### Teaching Assistant

TBD (TBD)

Office: ISR Building and Canvas

Office Hours: TBD (Eastern Time Zone) and by appointment

Email: TBD

**Class Meeting Time:** 9:00AM–Noon (Eastern Time Zone), July 13–31

**Class Location:** TBD

**Description and Goal of the Course:** After a refresher on statistical inference and bivariate association, this course introduces ordinary least squares (OLS) bivariate and multiple regression. The primary goal of the course is to develop a practical, applied, and intuitive (rather than strictly theoretical or mathematical) understanding of OLS regression. There will be a heavy emphasis on the graphical presentation of results. 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 understanding of the theoretical underpinnings, assumptions, and principles of OLS, the ability to use OLS in their own research, and the skills to build on OLS in their further study of more advanced statistical topics. Familiarity with basic statistical inference, basic probability, and basic algebra will be beneficial. The course is open to students at any level and from any discipline, as well as professionals and faculty members.

### Books:

Andersen, Robert, and David A. Armstrong II. 2022. *Presenting Statistical Results Effectively*. London: SAGE Publications.

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

### Course Content:

**Readings:** Be sure to complete the readings prior to each class period. The full 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 frequently analyze real data.

**Homework Assignments:** 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 analyses 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 teaching assistant(s) will be available to help students with software during office hours.

## Schedule and Readings:

### WEEK 1

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

- Andersen and Armstrong: Chapter 1, Sections 1.3.1 and 1.3.3

#### Day 2: Bivariate Association and a First Look at Regression

- Andersen and Armstrong: Chapter 5, Sections 5.2.1–5.2.3 and 5.4.1
- Lewis-Beck and Lewis-Beck: Chapter 1

#### Day 3: Statistical Inference with OLS Regression and the OLS Assumptions

- Andersen and Armstrong: Chapter 1, Section 1.3.2; Chapter 5, Sections 5.2.4, 5.3, and 5.3.1
- Lewis-Beck and Lewis-Beck: Chapter 2, pgs. 23–38

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

#### Day 5: Confounding Factors, Causality, and Control—on to Multiple Regression

- Andersen and Armstrong: Chapter 5, Section 7.6.2
- Lewis-Beck and Lewis-Beck: Chapter 3, pgs. 55–64 and 72–73

### WEEK 2

#### Day 6: Comparing Coefficients and Categorical Independent Variables

- Andersen and Armstrong: Chapter 5, Sections 5.5.1–5.5.2; Chapter 6, Sections 6.2 and 6.2.1
- Lewis-Beck and Lewis-Beck: Chapter 3, pgs. 64–69; Chapter 4, pgs. 83–86

#### Day 7: Interactions between Variables

- Andersen and Armstrong: Chapter 9, Sections 9.2.1, 9.3, 9.4, 9.4.1, 9.4.3–9.4.4, 9.5, and 9.5.2–9.5.3
- Lewis-Beck and Lewis-Beck: Chapter 3, pgs. 69–71

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

#### Day 9: Model Selection by Theory and by Machine

- Andersen and Armstrong: Chapter 5, Sections 5.4.2 and 5.4.6; Chapter 6, Section 6.3.1

#### Day 10: Regression Diagnostics 1

- Andersen and Armstrong: Chapter 7, Sections 7.4–7.5, 7.6.1, and 7.6.3
- Lewis-Beck and Lewis-Beck: Chapter 2, pgs. 44–49; Chapter 4, pgs. 75–83

## **WEEK 3**

### **Day 11: Regression Diagnostics 2**

- Andersen and Armstrong: Chapter 7, Sections 7.1.1, 7.2, 7.2.1, and 7.3.6–7.3.8; Chapter 8, Sections 8.3, 8.4, 8.4.2, and 8.6.1
- Lewis-Beck and Lewis-Beck: Chapter 2, pgs. 39–44 and 49–52; Chapter 4, pgs. 86–88

### **Day 12: Categorical Dependent Variables**

- Andersen and Armstrong: Chapter 5, Sections 5.7.1–5.7.3; Chapter 11, Sections 11.2.1–11.2.2, 11.3.1–11.3.4, and 11.4.1

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

### **Day 14: OLS for Causal Inference, Future Directions, and Wrap-Up**

- Lewis-Beck and Lewis-Beck: Chapter 4, pgs. 93–95