Introduction to Statistics with R

University

Semester

Days and Times: TBD

Location: TBD

Instructor: Chris Newton

Email: cnewton2@buffalo.edu

Office Hours: TBD

Course Description: This class introduces statistical analysis for political research using the R programming language. By the end of this course, students should be able to import, merge and modify data, conduct multiple regression, and interpret, communicate and visualize their results. The lectures will cover statistical theory with code examples. Students will have weekly problem sets in R to gain practical command over the theoretical topics. In addition to problem sets, students will work on a project of their choosing throughout the semester.

Student Learning Outcomes:

- Gain a familiarity with the R programming language.
- Build an understanding of how statistical research is conducted, its limitations, and standard practices.
- Develop an ability to communicate statistical results to technical and nontechnical audiences.
- Use available data and R to conduct an original, multivariate analysis.

Course Requirements: Students are expected to complete the assigned readings and attend class each week. In addition, there will be:

Problem Sets: You will have a problem set due at the end of most weeks. Theses assignments practice the skills that we learn in class. Studying the theory is not enough. These problem sets will give you plenty of practice and help you find solutions to common issues. There are no problem sets for weeks when other assignments are due.

Research Design: Pick a question you have about politics that can be tested empirically (I highly recommended you first discuss this with me during office hours). Identify datasets that are freely available for one dependent variable, one independent variable, and one confounding variable. Describe a hypothesis based on your question that you can test using these variables. You must select at least two different datasets.

Data Summary: For your three variables, generate a table with the minimum, maximum, standard deviation, mean, median, and mode. Generate a variance/covariance matrix, and a correlation heat map. Commented R code must be submitted for full credit.

Initial Results: You will merge your datasets so you have one dataset with your three variables. Conduct regression to test your hypothesis and present your results with a regression table, coefficient plot, and scatter plot. The scatter plot should include the slope and uncertainty intervals. This will be presented in class.

Final Project: When you present in class, you will receive feedback from your classmates and me. Based on the feedback, you will make improvements to you initial results and submit everything as a final project. Commented R code is required.

Late Work: All assignments will lose 1/3 of a letter grade (A becomes A-) every 24 hours past the deadline.

Course Materials: We will be using David Diez, Mine Çetinkaya-Rundel, & Christopher Barr's *OpenIntro Statistics*, 4th ed. You can download the PDF for any price (including $0) from https://www.openintro.org/book/os/. All other materials will be distributed through the class' GitHub page.

Grading Policy:

Problem Sets	30%
Research design	15%
Data Summary	15%
Initial Results	15%
Final Project	25%

Grade	Percentage	Grade Points
A	93-100%	4.0
A-	90-92%	3.67
B+	87-89%	3.33
B	83-86%	3.0
B-	80-82%	2.67
C+	77-79%	2.33
C	73-76%	2.0
C-	70-72%	1.67
D+	67-79%	1.33
D	63-66%	1.0
F	0-63%	0.0

Learning Outcomes:

Learning Outcome	**Assessment Measure**
Gain a familiarity with the R programming language.	Problem Sets; Data Summary; Initial Results; Final Project
Build an understanding of how statistical research is conducted, it limitations, and standard practices.	Research Design; Data Summary; Initial Results; Final Project
Develop an ability to communicate statistical results to technical and nontechnical audiences.	Initial Results; Final Project
Use available data and the R programming language to conduct an original, multivariate analysis.	Problem Sets; Initial Results; Final Project

Incomplete Grades:

A grade of incomplete ("I") indicates that additional course work is required to fulfill the requirements of a given course. Students may only be given an "I" grade if they have a passing average in coursework that has been completed and have well-defined parameters to complete the course requirements that could result in a grade better than the default grade. An "I" grade may not be assigned to a student who did not attend the course.

Prior to the end of the semester, students must initiate the request for an "I" grade and receive the instructor's approval. Assignment of an "I" grade is at the discretion of the instructor.

The instructor must specify a default letter grade at the time the "I" grade is submitted. A default grade is the letter grade the student will receive if no additional coursework is completed and/or a grade change request is not submitted by the instructor. "I" grades must be completed within 12 months*. Individual instructors may set shorter time limits for removing an incomplete than the 12-month time limit. Upon assigning an "I" grade, the instructor shall provide the student specification, in writing or by electronic mail, of the requirements to be fulfilled, and shall file a copy with the appropriate departmental office.

Students must not re-register for courses for which they have received an "I" grade

Important Dates:

Research Design ..... Week 5
Data Summary ..... Week 9
Initial Results ..... Week 14
Final Project ..... Finals Week

..... University specific policies .....

Academic Integrity: All students should be aware of and follow the university Academic Integrity Polity https://catalog.buffalo.edu/policies/integrity.html. Cases of academic dishonesty will not be tolerated.

Counseling Services (Mental Health): As a student you may experience a range of issues that can cause barriers to learning or reduce your ability to participate in daily activities. These might include strained relationships, anxiety, high levels of stress, alcohol/drug problems, feeling down, health concerns, or unwanted sexual experiences. Counseling, Health Services, and Health Promotion are here to help with these or other concerns. You learn can more about these programs and services by contacting:

Counseling Services: 120 Richmond Quad (North Campus), phone 716-645-2720
202 Michael Hall (South Campus), phone: 716-829-5800
Health Services: Michael Hall (South Campus), phone: 716- 829-3316
Health Promotion: 114 Student Union (North Campus), phone: 716- 645-2837

Sexual Violence: UB is committed to providing a safe learning environment free of all forms of discrimination and sexual harassment, including sexual assault, domestic and dating violence and stalking. If you have experienced gender-based violence (intimate partner violence, attempted or completed sexual assault, harassment, coercion, stalking, etc.), UB has resources to help. This includes academic accommodations, health and counseling services, housing accommodations, helping with legal protective orders, and assistance with reporting the incident to police or other UB officials if you so choose. Please contact UB's Title IX Coordinator at 716-645-2266 for more information. For confidential assistance, you may also contact a Crisis Services Campus Advocate at 716-796-4399.

Useful Resources:

- Data wrangling and visualization in R
  - R for Data Science
  - R Graphics Cookbook
- Programming in R
  - Tidyverse style guide for clean code
  - Advanced R

Schedule

Week 1: Introduction to the Course and R

- R vs. RStudio
- Installing and loading packages
- Getting course materials from GitHub
- Code example: a project from start to finish

- How to look things up on the internet

Week 2: What is Data?

- Chapter 1
- The Tidyverse
- What is good code and why does it matter?

Problem Set 1 due before midnight on **Friday**

Week 3: Probability I

- Chapter 2
- Importing datasets
- Probability as frequency
- Probability as uncertainty

Problem Set 2 due before midnight on **Friday**

Week 4: Probability II

- Chapter 2
- Visualizing uncertainty
- Conditional probability
- Simulating probability

Problem Set 3 due before midnight on **Friday**

Week 5: Distributions

- Chapter 3
- Data generating processes
- Included datasets
- Generating and visualizing distributions
- Summarizing distributions

Research Designs due by midnight **Friday**

Week 6: Correlation

- Spurious correlation
- Visualizing correlation

Problem Set 4 due before midnight on **Friday**

Week 7: Inference

- Chapter 4
- Causation, correlation, and interpretation

Problem Set 5 due before midnight on **Friday**

Week 8: Analysis: Numerical Data

- Chapter 5
- Data types and data structures

Problem Set 6 due before midnight on **Friday**

Week 9: Analysis: Categorical Data

- Chapter 6
- More data types and data structures

Data Summaries due before midnight on **Friday**

Week 10: OLS I

- Chapter 7
- Experimental vs. observational data and design
- Error

Problem Set 7 due before midnight on **Friday**

Week 11: OLS II

- Chapter 7
- Visualizing and interpreting results
- When assumptions don't hold

Problem Set 8 due before midnight on **Friday**

Week 12: Multiple Regression

- Chapter 8
- Why we need multiple variables
- Problems with adding variables

Problem Set 9 due before midnight on **Friday**

Week 13: Logistic Regression

- Chapter 8
- What's wrong with OLS for a binary DV?
- Link functions and their problems
- Visualizing and interpreting results

Problem Set 10 due before midnight on **Friday**

Week 14: Debugging

- Reading error messages
- Common issues and solutions
- Traceback
- Documentation and source code

Initial Results due before midnight on **Friday**

Week 15: Presentations
Finals Week
Final Project due before midnight, Friday