Introduction to Quantitative Methods with R

Course Overview

This course introduces basic quantitative methods used in political and social science, emphasizing understanding, interpretation, and application over mathematical derivations. Students learn to describe data, reason about probability and sampling, conduct basic hypothesis tests, and fit and interpret regression models, using R throughout the course. We will also engage with real social science research to connect methodology with substantive questions.

Course Objectives

- Use R to load, summarize, visualize, and analyze data.
- Read empirical social science articles and understand the main statistical arguments.
- Explain probability, sampling distributions, confidence intervals, and p-values in plain language.
- Conduct and interpret basic t-tests, chi-square tests, and simple/multiple regression models.

Design and execute a small empirical research project with existing political or social science data.

Assessment

- 10% - In-class participation
    - Active participation is required
- 40% - Bi-weekly take-away exercise *4
    - Solving specific tasks with methods learned in class
- 50% - Group Project (5% research prospectus; 15% presentation; 30% final report)
    - 2 – 3 students form a group
    - A group project that applies the methods learned in this class to solve a social science issue

Course Structure

Weekly class typically includes three 45-minute sessions. The first two sessions cover basic theoretical concepts and walk through applying relevant methods in the assigned readings (if any). The third session is a lab session in which all students follow the instructor to run programming code and do in-class coding exercises.

Required Texts

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2017). *Probability and statistics for engineers and scientists*. New York: Macmillan.

Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

Llaudet, E., & Imai, K. (2023). *Data analysis for social science: A friendly and practical introduction*. Princeton University Press.

Weekly Schedule and Readings

The schedule below outlines the issues covered in the first session, the lab content, and the required readings and applications for each week. Read the readings before class that day. Students are

expected to spend, on average, 3 hours per week reading the assigned materials and working on exercises and coding.

Part A: Introduction to Quantitative Analysis and Programming

Week 1 Introduction

- Issues: course overview; expectations; why statistics for social science.
- Lab: setup of RStudio, the basic operation of the R language in RStudio, write and read data
- Reading: L&I Chapter 1; Walpole Chapter 1

Week 2 Describing Social Science Data

- Issues: variable types; tables, graphs; measures of center and spread
- Lab: Descriptive statistics and basic plots in R
- Reading: L&I Chapter 3; Walpole: Chapter 1.3-1.6

Week 3 Probability & Uncertainty

- Issues: classical probability; events, randomness
- Lab: Simulations in R to illustrate probability and long-run frequency.
- Reading: L&I Chapter 6.1-6.3; Walpole Chapter. 2.1–2.5

*Exercise one due

Part B: Distributions and Sampling

Week 4 Distributions & Sampling

- Lecture: Discrete/continuous RVs, basic distributions; central limit theorem intuition
- Lab: Use R to simulate sampling distributions of sample means and illustrate normal approximation
- Readings: L&I Chapter 6.4-6.5; Walpole Chapter 5.2; 5.5; 6.1-6.3

Week 5 Sampling and Estimation

- Lecture: Estimators; population vs sample; confidence intervals for means and proportions
- Lab: Compute and interpret confidence intervals in R
- Readings: L&I Chapter 7.1-7.2; Walpole Chapter 8.1-8.4, 8.6; 9.2-9.4

* Exercise two due

Part C Introduction to Statistical Inference

Week 6 – Hypothesis Tests & p-values

- Lecture: Null/alternative hypotheses, test statistics, Maximum likelihood (MLE), p-values, and errors.
- Lab: Difference-in-means tests and MLE in R
- Reading: L&I Chapter 7.3-7.4; Walpole Chapter 9.14; 10.1-10.4

Week 7 – Comparing Two Groups (t-tests & Chi-square)

- Lecture: Independent t-tests as difference-in-means; chi-square tests for categorical association.
- Lab: Use R to compare groups and test for association between categorical variables.
- Readings: Walpole, Chapter 10.5-10.10

Week 8 – Correlation and Simple Regression

- Lecture: Correlation; basics of simple linear regression; prediction errors and fit.
- Lab: Fit and interpret simple linear regressions in R; visualize fitted lines and residuals.
- Readings: L&I Chapter 4; Walpole Chapter 11

* Exercise three due

<u>Part D Advanced Topics</u>

Week 9 – Multivariate Regression

- Lecture: Multiple regression; controlling for confounders; reading regression tables.
- Lab: Estimate multiple regression models in R; compare with bivariate models.
- Readings: Walpole Chapter 12.1-12.6

Week 10 – Diagnostics and Assumptions

- Lecture: Linearity, outliers, heteroskedasticity at a conceptual level; model fit.
- Lab: Residual plots and simple diagnostic checks in R; exploring influential observations.
- Readings: G&A Chapter 4

Week 11 – Generalized linear models

- Lecture: logistic regression, Poisson regression, multinomial regression
- Lab: Estimate multiple GLMs in R with comparison
- Readings
- Readings: G&A Chapter 5;6

Week 12 – Introduction to causal inference

- Lecture: Omitted variable bias; confounding; Randomized experiments vs observational studies; internal vs external validity
- Lab: Simple simulation in R to show confounding and the value of randomization
- Readings: L&I Chapter 5.2, 5.5, G&A Chapter 9

Week 13 – Project Presentations & Wrap-Up

Week 14 – Project Presentations & Wrap-Up