PHP 2550 Syllabus

Instructor:
Alice Paul
Email: alice_paul@brown.edu
Office: SPH 717

Teaching Assistant:
Joanna Walsh
Email: joanna_walsh@brown.edu

Lectures: Mondays and Wednesdays 10:30-11:50 am

Course Goals

Analyzing data is a core skill for many scientists. In addition to familiarity with probability and statistics, good data analysis requires skill in computing and effective presentation of results and communication with scientific colleagues. The data analyst must be able to translate the scientific question and hypothesis into a testable form, advise about data to be collected, manipulate the data into a computable format to carry out appropriate analysis.
This course is designed for graduate students who will be analyzing data with scientific colleagues and who want to develop a practical hands-on toolkit and gain experience in distilling complex statistical information into formats understandable to colleagues. Topics including data collection, exploratory data analysis, missing data, fitting and checking models, simulation, predictive models, and presentation of reproducible results will be developed through a series of case studies based on different types of data requiring a variety of statistical methods. We will be using the R programming environment in class.

After you complete this course, you will be able to:

- Understand management and documentation of study data and the assumptions and limitations of the data.
- Use statistical programming to write reproducible, documented code to efficiently analyze data and develop simulation studies.
- Apply modern statistical techniques and principles using appropriate model selection.
- Evaluate research reports and proposals on the basis of their scientific integrity, validity, and the strength of quantitative analysis.
- Collaborate in teams to use biostatistical methods to solve substantive problems in public health.
- Present your analyses through convincing arguments to statistical and non-statistical colleagues.

Course Logistics

This course will be taught in a partially flipped format. Some weeks will consist of primarily lecture while other weeks will focus on discussion and class exercises to provide more depth and examples. Participation during these sessions is an integral part of the course and is reflected in the course grade. If you cannot make a class, you should notify me in advance. The deliverables for the class will include biweekly homework assignments and a final project.

Prerequisites: Students should have courses in probability and statistical inference at the level of PHP 2510 or PHP 2515 as well as regression analysis at the level of PHP 2511 or 2514. Familiarity with the R programming language is also required at the level of PHP2560.

Course Websites (Canvas): The canvas site will contain all the information for this course including this syllabus, assignments, office hours, weekly course info, and posted grades. Make sure that you are enrolled and that you check the sites regularly.

Reading: The reading for this course can be found in the two books below as well as supplementary materials that will be provided on Canvas. In some cases, readings from different books will complement each other and may appeal to students with different mathematical backgrounds. For example, the book by Hastie et al. is more mathematically intensive than the one by James et al. Students with a better mathematical foundation (especially PhD students in biostatistics and related fields) should read this material.

Hastie T, Tibshirani R and Friedman J. *The Elements of Statistical Learning*, 2nd ed. NY: Springer, 2009. (Available for free online.)

James G, Witten D, Hastie T and Tibshirani R. *An Introduction to Statistical Learning.* NY: Springer, 2013. (Available for free online.)

If your Brown undergraduate financial aid package includes the Book/Course Material Support Pilot Program (BCMS), concerns or questions about the cost of books and course materials for this or any other Brown course (including RISD courses via cross-registration) can be addressed to bcms@brown.edu. For all other concerns related to non-tuition course-related expenses, whether or not your Brown undergraduate financial aid package includes BCMS, please visit the Academic Emergency Fund in E-GAP (within the umbrella of "E-Gap Funds" in UFunds) to determine options for financing these costs, while ensuring your privacy.

Course Assessment

Your course grade will be based on:

Participation	10%
Homework Assignments	50%
Final Project	40%
Total	100%

Homework Assignments: Biweekly homework assignments will be a mixture of paper responses, data analyses, simulations and programming exercises. Students may work together on homework assignments but need to write them up individually. All students have a bank of seven late days that they may use at their discretion. After you use up your late days, all assignments must be turned in when due to receive full credit.

Final Project: The final project will require students to rigorously analyze data from a substantive problem in public health. The project grade will be determined based on intermediate deliverables, an in-class poster session, and a final written report. The report must carefully outline the methods, analyses and interpretations of the results using well-written prose and informative tables and figures, all accessible to the non-expert. Computer code sufficient to reproduce the analyses must be provided. Students will be allowed to work in pairs on the project.

Credit Hours and Time Expectations: Over 14 weeks, students will spend 3 hours per week in class (42 hours total), and should expect to do about 2 hours per week of class preparation including

reading (28 hours). Homework should take about 7 hours per assignment (35 hours) and the final project should take about 75 hours of work. The total amount of time expected is 180 hours.

The following schedule is tentative.

Week 1 (9/7)	Course Intro
Week 2 (9/12, 9/14)	Reproducible Research and Documentation
Week 3 (9/19, 9/21)	Data Exploration, Management, and Visualizations
Week 4 (9/26, 9/28)	Missing Data
Week 5 (10/3, 10/5)	Scientific Questions and Data Bias
Week 6 (10/12)	Regression Analysis
Week 7 (10/17, 10/19)	Sampling and Weighting
Week 8 (10/24, 10/26)	Metrics and Cross-Validation
Week 9 (10/31, 11/2)	Simulation
Week 10 (11/7, 11/9)	Simulation Applications
Week 11 (11/14, 11/16)	Shrinkage and Selection Methods
Week 12 (11/21)	Variable Selection Applications
Week 13 (11/28, 11/30)	Project Check-Ins
Week 14 (12/5, 12/7)	Reports and Posters
Week 15 (12/12)	Poster Session

Additional Information

Public Health Competency: Upon completion of the course, students should be able to manipulate, program, analyze, display and present data and statistical models so that they are comprehensible for the non-statistical expert scientific collaborator. This addresses the Biostatistics Masters Competency to Write a report of comprehensive and novel statistical analysis of public health data as well as the Biostatistics PhD competency to Explain statistical concepts and methods to a broad audiences (for example, undergraduate students, clinicians or medical researchers). These competencies are assessed through completion of the final project.

Academic Integrity You are expected to maintain the highest level of academic integrity in the course. Any violation will be penalized according to the Brown academic integrity code. In particular, sharing materials outside of this class will be treated as an academic code violation.

You are expected to communicate and interact with your classmates. You are welcome and encouraged to discuss projects and exercises with others as long as these interactions are consistent with your educational goals. For example, if you read someone else's code in order to learn a new technique and apply it to your own implementation, that is a good way to learn. But if you copy code or analysis you don't understand instead of learning to develop your own, you are violating the academic integrity code and undermining your own education.

Accessibility and Accommodations

Brown University is committed to full inclusion of all students. Please inform me early in the term if you may require accommodations or modification of any of course procedures. You may speak with me after class, during office hours, or by appointment. If you need accommodations around online learning or in classroom accommodations, please be sure to reach out to Student Accessibility Services (SAS) for their assistance (seas@brown.edu, 401-863-9588). Students in need of short-term academic advice or support can contact one of the academic deans in the College.

It is my intent that students from all diverse backgrounds and perspectives will be well served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. It is my intent to present materials and activities that are respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. I expect students to conduct themselves in a way that respects the experiences and identities of all members of the class. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups. In addition, if any of the class meetings conflict with your religious events, please let me know so that I can make arrangements for you.