

PHP 1510/2510: Principles of Biostatistics and Data Analysis

Adam J Sullivan, PhD

Fall 2019

E-mail: adam_sullivan@brown.edu

Office Hours: Check Course Website

Office: School of Public Health - Room 717

Web: php-1510-2510.github.io

Class Hours: T/TH 9:00 am - 10:20am

Class Room: School of Public Health - Room 245

Course Description

This course is intended to provide a basic foundation in the methods and applications of biostatistics, and is geared towards the students whose fields of study include a substantial statistical or quantitative component. Ideally, this course is the first in a two-part sequence (the sequel being PHP 1511/2511: Applied Regression), designed to provide students in the public health, biological and life sciences with broad-based exposure to modern methods of biostatistical inference, in addition to an understanding of underlying mathematical principles and motivations.

Applications to real data from a variety of studies in public health and clinical research are used throughout the course to illustrate the materials, but the emphasis is on the principles of inference and the underlying theory.

In this course we also give students experience manipulating and analyzing data using the R statistical software package with RStudio interface. Students will work with statistical software during the weekly lab sessions, as well as in the homework.

Pre-requisites

No prior coursework in statistics or probability is needed. Prior experience in college level math such as Calculus will prove helpful but is not necessary.

Required Textbooks

Chihara, Laura and Tim Hesterberg (2018). *Mathematical Statistics with Resampling and R*. John Wiley & Sons.

Wickham, Hadley and Garrett Grolemund (2017). *R for Data Science*. O'Reilly Media, Inc. <https://r4ds.had.co.nz/>.

Speegle, Darrin (2018). *Foundations of Statistics with R*. Bookdown. <https://bookdown.org/speegled/foundations-of-statistics/>.

Irizarry, Rafael A. (2019). *Introduction to Data Science*. Bookdown. <https://rafalab.github.io/dsbook/>.

list()

Course Objectives

After successful completion of this course you will understand and be able to use probability, statistical graphics and hypothesis tests. In particular these include the following capabilities:

1. Manipulate and analyze data in R.
2. Students will use RStudio with proficiency.
3. Students will understand how to graph, test and interpret results for given probability distributions and data.

Overall Course Expectations

Students in this course will be expected to do the following:

1. Attend all lectures and actively participate in discussion.
2. Read all assigned material *prior* to coming to class and actively participate in class discussions.
3. Complete and turn in all assignments on time. Solutions to homework must be clearly written with appropriate tables and figures included.
4. Demonstrate an understanding on material on examinations.
5. Respect each other, each others questions and each others discussion.

Evaluation

Students will be evaluated based on:

Grade Category	Percentage
Participation	5%
Homework	20%
Exam 1	25%
Exam 2	25%
Final Take Home	25%

Evaluation Category Details

Participation

This course will move very fast and it is crucial to success in the course that students attend and participate. Many classes will have polls or quizzes that will not be graded for having the most correct or best answer but for participating. Unexcused absences will result in a loss of percentage points.

Homework

There will be 10 homework assignments during the semester, together worth 25% of course grade. Homework will consist of both writing and computing exercises. *Late homework will not be accepted.* Students are expected to work independently on their homework, but discussion on general aspects of the course content is encouraged.

Exam 1 & 2

There will be two in class exams: one midterm and one final. The midterm exam will be in class, and will cover materials presented in lectures, labs, readings and assignments. The final exam will cover material from the entire semester, but will emphasize material since the midterm exam. All in-class exams are closed-book but you can bring a cheat-sheet (A4 size) to the exams. Only pen or pencil, plus a calculator, will be permitted. As a general rule, make up exams and advance exams will not be given, with the exception when the dates conflict with religious observances or truly exceptional circumstance. Please let me know ahead of time if you need to make alternative arrangements for this purpose.

Final Take Home

There will also be a take-home final exam, focusing on data analysis from a real dataset.

Differences between PHP 1510/2510

Given the nature of this course with multiple levels of students from Undergraduate to PhD, it is important to discuss the differences of expectations and how students will be graded.

PHP 1510

Grade Category	Comments
Participation	Graded the same as all students, Must be in class and prepared to work in groups.
Homework	Students will be expected to complete a portion of the material with the exception of some more difficult problems which may be attempted but do not have to be complete.
Exam 1 & 2	Students will be expected to complete a portion of the exam.
Take Home Exam	This will be a complete data analysis. No part of this is allowed to be discussed with any other student.

PHP 2510

Grade Category	Comments
Participation	Graded the same as all students, Must be in class and prepared to work in groups.
Homework	Students will be expected to complete the entire assignment.
Exam 1 & 2	Students will be expected to complete the entire exam.
Take Home Exam	This will be a complete data analysis. No part of this is allowed to be discussed with any other student.

Semester Hours

Over the course of the semester students will spend at least the amounts of time shown below:

Task	Hours Spent on Task
Class Time	42
Labs	14
Out of Class Work	112
Take Home Exam	12
Total	180

Class Schedule

Class Schedule

Readings under each class are expected to be complete by class. Class discussion will then highlight parts of reading but not all things. Important: class readings are subject to change, contingent on mitigating circumstances and the progress we make as a class. Students are encouraged to attend lectures and check the course website for updates.

Note the schedule below is the planned schedule. See course website for the most up-to-date schedule and readings.

Week	Date	Topic	Readings
Week 1	5-Sep	Course Introduction	No Reading Prior to Class
Week 2	10-Sep	Basic Probability	CHEST: Appendix A; SPEEGLE Ch. 2; RAFA Ch. 1
	12-Sep	Basic Probability	CHEST: Appendix A; SPEEGLE Ch. 2; RAFA Ch. 1
Week 3	17-Sep	Random Variables	CHEST: Appendix B; SPEEGLE Ch. 3; RAFA Ch. 15
	19-Sep	Random Variables	CHEST: Appendix B; SPEEGLE Ch. 3; RAFA Ch. 15
Week 4	24-Sep	Simulations with Random Variables	SPEEGLE Ch. 4.1-4.4
	26-Sep	Manipulating Data with R	R4DS Ch. 4-5; SPEEGLE Ch. 5
Week 5	1-Oct	Exploratory Data Analysis	CHEST Ch. 2; SPEEGLE Ch. 6; R4DS Ch. 7
	3-Oct	Exploratory Data Analysis	CHEST Ch. 2; SPEEGLE Ch. 6; R4DS Ch. 7
Week 6	8-Oct	Introduction to Hypothesis Testing	CHEST Ch. 3
	10-Oct	Introduction to Hypothesis Testing	CHEST Ch. 3
Week 7	15-Oct	Sampling Distributions	CHEST Ch. 4; SPEEGLE Ch. 4.5; RAFA Ch. 15.6-15
	17-Oct	Sampling Distributions	CHEST Ch. 4; SPEEGLE Ch. 4.5; RAFA Ch. 15.6-15
Week 8	22-Oct	Introduction to Confidence Intervals	CHEST Ch. 5; Modern Ch. 9
	24-Oct	***Exam 1***	
Week 9	29-Oct	Introduction to Confidence Intervals	CHEST Ch. 5; Modern Ch. 9
	31-Oct	Hypothesis Testing	CHEST Ch. 8; SPEEGLE Ch 7; Modern Ch 10
Week 10	5-Nov	Hypothesis Testing	CHEST Ch. 8; SPEEGLE Ch 7; Modern Ch 10
	7-Nov	Non-Parametric Testing	SPEEGLE Ch. 8
Week 11	12-Nov	Categorical Data Comparisons	CHEST Ch. 10; SPEEGLE Ch. 9
	14-Nov	Categorical Data Comparisons	CHEST Ch. 10; SPEEGLE Ch. 9
Week 12	19-Nov	Simple Linear Regression	CHEST Ch. 9.1-9.3.2; SPEEGLE Ch. 10-11
	21-Nov	Simple Linear Regression	CHEST Ch. 9.1-9.3.2; SPEEGLE Ch. 10-11
Week 13	26-Nov	Multiple Linear Regression	SPEEGLE Ch. 12; RAFA Ch. 19
	28-Nov	**Thanksgiving Day: No Class**	
Week 14	3-Dec	Multiple Linear Regression	SPEEGLE Ch. 12; RAFA Ch. 19
	5-Dec	Inference for Linear Regression	CHEST 9.3.3-9.5.2
Week 15	10-Dec	Many Models	R4DS Ch 25
	12-Dec	***Exam 2***	

Students with Special Needs

Brown University is committed to full inclusion of all students. Students who, by nature of a documented disability, require academic accommodations should contact the professor during office hours. Students may also speak with Student and Employee Accessibility Services at 401-863-9588 to discuss the process for requesting accommodations.

Diversity Statement

This course is designed to support an inclusive learning environment where diverse perspectives are recognized, respected and seen as a source of strength. It is our intent to provide materials and activities that are respectful of various levels of diversity: mathematical background, previous computing skills, gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture.

Note: When we get involved with human subjects data, there can be very sensitive topics. The reality is that much of data that has been collected and used in research leaves out various groups or identities. We will pay close attention to understanding biases that this can induce in the data. However, this may bring up stressful situations or difficult memories at times. Please contact Professor Sullivan if any of these data cause you any harm.

English Language Learners

Brown University welcomes students from around the world, and the unique perspectives international students bring enrich the campus community. To empower students whose first language is not English, an array of ELL support is available on campus including language and culture workshops and individual appointments. For more information about English Language Learning at Brown, contact the ELL Specialists at ellwriting@brown.edu.