

# PHP 1510/2510: Principles of Biostatistics and Data Analysis

Stavroula A. Chrysanthopoulou, PhD

Fall 2018

E-mail: [stavroula\\_chrysanthopoulou@brown.edu](mailto:stavroula_chrysanthopoulou@brown.edu)

Office Hours: TBD.

Office: School of Public Health - Room 717

Web: TBD

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

Class Room: School of Public Health - Room 375

---

## 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.

## Prerequisites

No prior coursework in statistics or probability is needed, but working knowledge of college-level calculus (derivatives, integrals, exponential and log functions, summations, etc.) is required. A short self evaluation test will be distributed on the first day of class. It will help you assess your level of comfort with various mathematical tools that are needed for this course. Ideally, you should be able to answer each of these questions either on your own, or with quick reference to an appropriate source to refresh your memory. We assume no prior knowledge of programming in this course.

## Required Textbooks

Rosner, Bernard (2015). *Fundamentals of biostatistics*. Nelson Education.

## Recommended Textbooks

Shahbaba, Babak (2012). *Biostatistics with R: An Introduction to Statistics Through Biological Data*. Springer. <http://link.springer.com.revproxy.brown.edu/book/10.1007%2F978-1-4614-1302-8> }

## 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	30%
Exam 1	20%
Exam 2	20%
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 6 homework assignments during the semester, together worth 30% 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 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	Students will be expected to complete a final project. Data as well as questions explored will be at a level appropriate of the background and other statistical courses taken.

### 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 S	tudents will be expected to complete a final project. Data as well as questions explored will be at a level appropriate of the background and other statistical courses taken.

## 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

Course Schedule	Week	Content	Reading	Notes
9/6	1	Introduction and overview of course	Chapters 1 and 2	introduction to R, data exploration
9/10 - 09/14	2	Exploring Relationships	Ch. 3	Lab 1
9/17 - 9/21	3	Discrete Probability Distributions	Ch. 4	Lab 2
9/24 - 9/28	4	Continuous Probability Distributions	Ch. 5	Lab 3
10/1 - 10/5	5	Estimation - Part 1	Ch. 6	Lab 4
10/8 - 10/12	6	Estimation - Part 2		Lab 5: Review
10/16		Midterm Exam		
10/15 - 10/19	7	Hypothesis Testing - Part 1	Ch. 7	
10/22 - 10/26	8	Hypothesis Testing - Part 2		Lab 6
10/29 - 11/2	9	Hypothesis Testing - 2 Variables	Ch. 8	Lab 7
11/5 - 11/9	10	Analysis of Variance (ANOVA)	Ch. 12	Lab 8
11/12 - 11/16	11	Analysis of Categorical Variables	Ch. 10	
11/19 - 11/22	12	Regression Analysis - Part 1	Ch. 11	No class on 11/24
11/26 - 11/30	13	Regression Analysis - Part 2		Lab 9
12/3 7- 12/	14	Review		Lab 10: Review

## **Brown Undergraduate Competencies**

### **Refresher Competencies**

1. Communicate in the language of mathematics
2. Perform theoretical calculations using multidimensional calculus and matrix algebra
3. Explain the interplay between mathematical derivations and statistical applications
4. Perform exploratory data analysis approaches and graphical data analysis methods
5. Describe design of studies (e.g., random assignment, random selection, data collection, and efficiency) and issues of bias, causality, confounding, and coincidence
6. Apply, describe and assess the validity of statistical models (e.g., variety of linear and non-linear parametric, semiparametric, and nonparametric regression models)
7. Demonstrate a foundation in statistical theory
8. Be able to use of one or more professional statistical software environment
9. Be able to manipulate data (possibly “big”) using software in a well-documented and reproducible way.
10. Communicate statistical results effectively
11. Identify and critique application of statistics in various fields

### **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.

### **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](mailto:ellwriting@brown.edu).