PHP 1501: Essentials of Data Analysis
Fall 2023

Instructor: Anarina Murillo, PhD
Email: anarina_murillo@brown.edu
Office: South Main Street 121, Room 713
Office Hours: TBD, see Canvas

Teaching Assistants: Carissa Chae, Email: carissa_chae@brown.edu
Office Hours: TBD, see Canvas

Tabib Chowdhury, Email: tabib_chowdhury@brown.edu
Office Hours: TBD, see Canvas

Devon Newman, Email: devon_newman@brown.edu
Office Hours: TBD, see Canvas

Lectures: S01: Tuesday and Thursday 1:00 PM - 2:20 PM
S02: Monday and Wednesday 8:30 AM - 9:50 AM

Class Room: S01: Bio Med Center, Room 202
S02: Friedman Hall, Room 201

Labs: L01: T 9-10:20 AM, Barus & Holley, Room 159
L02: Th 9-10:20 AM, Barus & Holley, Room 159
L03: F 1-2:20 PM, 121 South Main Street, Room 375
L04: M 3:40-5 PM, 121 South Main Street, Room 375

Course Overview

This course covers the basic concepts of statistics and the statistical methods commonly used in public health, biology, medicine, and the social sciences, with an emphasis on applications to real data. The first half of the course introduces descriptive statistics, common probability models, and basic methods of statistical inference such as parameter estimation, hypothesis testing, and confidence intervals. The second half introduces methods for understanding and modeling bivariate and multivariate relationships between variables, emphasizing regression analysis as a unifying theme for modeling and drawing inference about these relationships. Along the way we will draw connections to themes that are relevant to modern data science, such as the p-value controversy, causal inference, and basic machine learning.

The examples used in the course are primarily derived from public health and biomedical research. Statistics is an inherently mathematical discipline; we will make frequent use of basic algebra and will use the log and exponential functions frequently. We also will use summation notation quite often. A couple of basic ideas from calculus will be used to illustrate concepts but students will not be asked to use calculus-based tools for homework. The course uses the R

statistical software package, which will be reviewed during labs. The instructors will supply R code for some of the homework and in-class exercises.

Prerequisites

None. The course is intended for undergraduates who have no prior exposure to college-level statistics or that have only been exposed to AP statistics. It is designed to be a first course in statistics and is required for statistics concentrators and public health concentrators. The course can also be taken by those looking for a hands-on, introductory- level, one-semester survey course of commonly-used statistical methods. These students can register with instructor's permission.

Domains Covered by the Course

The School of Public Health's accrediting body, the Council on Education for Public Health (CEPH), specifies the domains in which all undergraduates in accredited programs are to receive instruction. The Brown SPH faculty designate the courses in which each domain will be introduced and those in which each will be covered in more depth. This course will cover the following public health domains:

Domain	**Teaching Activities**
Basic Concepts of Data Collection	• Lecture 1: Introduction • Lecture 2: Data Variable Types, Descriptive Statistics • Homework 1 • Lab 1, 2 • Midterm Exam
Basic Methods of Data Collection	• Lecture 3, 4, 5: Probability • Lecture 6, 7, 8, 9: Distribution of Random Variables • Homework 2, 3, 4 • Lab 1, 2, 3 • Midterm Exam

Basic Tools of Data Collection	• Lecture 1: Introduction • Lecture 2: Data Variable Types, Descriptive Statistics • Homework 1, 2, 3, 4, 5 • Lab 1, 2, 3 • Midterm Exam
Data Usage	• Lecture 10, 11: Foundations for Inference • Lecture 12, 13: Inference for Numerical Data • Lecture 14, 15: Simple Linear Regression • Lecture 16, 17, 18, 19: Multiple Linear Regression • Lecture 20, 21: Inference for Categorical Data • Lecture 22, 23: Special Topics • Homework 4, 5, 6, 7, 8 • Lab 4, 5, 6 • Midterm Exam, Quiz, Final Exam
Data Analysis	• Lecture 2: Data Variable Types, Descriptive Statistics • Lecture 12, 13: Inference for Numerical Data • Lecture 14, 15: Simple Linear Regression • Lecture 16, 17, 18, 19: Multiple Linear Regression • Lecture 20, 21: Inference for Categorical Data • Homework 2, 3, 4, 5, 6, 7, 8 • Lab 4, 5, 6, 7, 8, 9 • Quiz, Final Exam

Course Objectives

By the end of this course, students should be able to do the following:

1. Understand the research process including how to define a research question, study designs, data collection (sampling and randomization) and statistical analysis.
2. Identify an appropriate method for inferential and descriptive statistics to provide meaningful insights and assess assumptions and the uncertainty of the results.
3. Apply statistical software to organize, summarize, and present data using graphical and tabular representations.
4. Determine the appropriate methods for summarizing bivariate and multivariate associations decide on the appropriate use of multivariate analyses (e.g., regression analysis) based on the research question and data collected.
5. Be able to distinguish between association and causality, and critically review and analyze statistical arguments found in popular press and scholarly journals.

Books, Supplies, and Materials

Course Text and Other Readings

- Vu J. and Harrington D. (2020). Introductory Statistics for the Life and Biomedical Sciences (1st Edition). The PDF of the textbook is available online for free at `https://www.openintro.org/book/biostat/` A hardcopy can be purchased for $20 through the website, Amazon, and in other bookstores.
- Lab materials will be posted on Canvas and will include content from the VH textbook
- Additional article readings may be assigned during the semester.

Statistical Software This course uses R/RStudio for labs and homework assignments. Students should install software prior to the first day of lab.

- R Statistical Software and RStudio is available for free at `http://cran.r-project.org/` and `http://www.rstudio.com/products/rstudio/download/`. You may download the most recent version that is compatible with your device.
- In this course, we will use R/RStudio statistical software. You may use any programming language with which you're more comfortable with for assignments. However, the teaching team will not provide support for other programming languages.

Book/Course Material Support Pilot Program 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 Website (Canvas): The canvas site will contain all the information for this course including this syllabus, office hours, weekly course info, and posted grades. Make sure that you are enrolled on Canvas and that you check the site regularly. Please note that this course will share its Canvas site and other online materials for both sections of PHP 1501 (S01 and S02). Students in both courses will be able to see the roster of the other course. If you have concerns about this, please e-mail me as soon as possible to discuss your options.

Zoom: Office hours may be offered using Zoom. Please check Canvas for details.

Top Hat: This course will use Top Hat during lecture. Make sure to register for Top Hat through Canvas and check that you can log-in and click on the Top Hat application.

Method of Evaluation

Your course grade will be based on:

Item	Percentage
Homework Assignments (8 total)	30%
Midterm Exam	20%
Quiz	20%
Final Exam	30%

Homework assignments (HW)

There will be 6 assignments each worth 24 points and 2 assignments each worth 16 points. These problem sets will correspond to book readings and lectures. They may require a calculator or R/R Studio statistical software. Assignments must be submitted through Canvas no later than 11:59 PM on the due date. All assignments should be completed individually. You may discuss homework with your classmates. However, you must work on your assignments individually, submit your own assignment, and write your own analysis and conclusions. It is not acceptable to send or copy the homework solutions or codes for assignments to other classmates. If you have any questions about what constitutes plagiarism, please consult the Brown University Code of Conduct.

Midterm Examination

The midterm examination will consist of a set of problems similar to those in the assignments. This also will include interpreting output from data analyses but no coding in statistical software will be required. You will have the entire class session to complete the exam.

Final Examination

The final examination will be an in-person exam. The final exam will cover all the concepts covered during the course. It will include interpreting output from data analyses but no coding in R statistical software will be required. If you are unable to take the final exam in-person, please consult Brown's policy regarding these situations: Students unable to attend the final exam at the scheduled time are required to obtain an excuse from a Dean. The Registrar's office will issue the final exam to the student at the start of the following semester. The instructor will supply the final exam.

Quiz

There will be a take-home quiz and it will have a similar format to the exams. It will include

interpreting output from data analyses but no coding in statistical software will be required. The quiz is open-book and open-notes from the lecture and labs only. You may not browse the internet or use any resources online. You are required to finish the quiz independently. The quiz must be completed within four hours.

Participation

There will be no grade for participation. However, participation in-class and answering Top Hat questions is strongly encouraged.

Grading Policy

In accordance with Brown's grading system, ABC/NC or S/NC grades will be tentatively assigned based on the following cutoffs:
ABC/NC: A: 90%-100%, B: 80%-89%, C: 70%-79%, NC: 0%-69%
S/NC: S: 70%-100%, S with Distinction: 90%-100%, NC: 0%-69%
These ranges are tentative and the cutoffs for final grades will be determined at the end of the semester.

Tentative Schedule

S01: Tuesday and Thursday 1:00 PM - 2:20 PM
S02: Monday and Wednesday 8:30 AM - 9:50 AM

Week	Topic	(S02 date, S01 date)	Readings	Assignment
1	(9/6,9/7)	*Lecture 1:* Course Overview, Introductions	VH CH 1.1-1.3	
		Lab: No Lab		
2	(9/11,9/12)	*Lecture 2:* Data Variables, Descriptive Statistics	VH CH 1.4-1.8	
	(9/13,9/14)	*Lecture 3:* Probability	VH CH 2.1	
		Lab: No Lab		
3	(9/18,9/19)	*Lecture 4:* Probability	VH CH 2.1-2.2	
	(9/20,9/21)	*Lecture 5:* Probability	VH CH 2.3-2.4	
		Lab 1: Introduction to R/R Studio	VH CH 1	
4	(9/25,9/26)	*Lecture 6:* Distribution of Random Variables	VH CH 3.1	HW 1 Due
	(9/27,9/28)	*Lecture 7:* Distribution of Random Variables	VH CH 3.2-3.3	
		Lab 2: Probability	VH CH 2	
5	(10/2,10/3)	*Lecture 8:* Distribution of Random Variables	VH CH 3.4-3.5	HW 2 Due
	(10/4,10/5)	*Lecture 9:* Distribution of Random Variables	VH CH 3.6-3.7	
		Lab 3: Random Variables	VH CH 3	
6	(10/11,10/10)	*Lecture 10:* Foundations for Inference	VH CH 4.1-4.2	HW 3 Due
	(10/11,10/12)	*Lecture 11:* Foundations for Inference	VH CH 4.3-4.4	
		Lab 4: Foundations for Inference	VH CH 4	
		No Class on 10/9 Indigenous Peoples' Day.		

7	(10/16,10/17)	*Lecture 12:* Inference for Numerical Data	VH CH 5.1-5.2	HW 4 Due
	(10/18,10/19)	*Lecture 13:* Inference for Numerical Data	VH CH 5.3-5.6	
		Lab: Midterm Exam Review		
8	(10/23,10/24)	*Lecture:* Midterm Exam (In-Class)		
	(10/25,10/26)	*Lecture 14:* Simple Linear Regression	VH CH 6.1-6.3	
		Lab 5: Inference for Numerical Data	VH CH 5	
9	(10/30,10/31)	*Lecture 15:* Simple Linear Regression	VH CH 6.4-6.6	HW 5 Due
	(11/1,11/2)	*Lecture 16:* Multiple Linear Regression	VH CH 7.1-7.4	
		Lab 6: Simple Linear Regression	VH CH 6	
10	(11/6,11/7)	*Lecture 17:* Multiple Linear Regression	VH CH 7.5-7.7	HW 6 Due
	(11/8,11/9)	*Lecture 18:* Multiple Linear Regression	VH CH 7.8	
		Lab 7: Multiple Linear Regression	VH CH 7	
11	(11/13,11/14)	*Lecture 19:* Multiple Linear Regression	VH CH 7.10	HW 7 Due
	(11/15,11/16)	*Lecture 20:* Inference for Categorical Data	VH CH 8.1-8.2	
		Lab 8: Categorical Data	VH CH 8	
12	(11/20,11/21)	*Lecture:* Quiz (Take-Home), No Class		
	(11/22,11/23)	*Lecture:* No Class, Happy Thanksgiving!		
		Lab: No Lab		
13	(11/27,11/28)	*Lecture 21:* Inference for Categorical Data	VH CH 8.3-8.4	
	(11/29,11/30)	*Lecture 22*: (Special Topics) Logistic Regression, Regression in Machine Learning	See Canvas	
		Lab 9: Logistic Regression	See Canvas	
14	(12/4,12/5)	*Lecture 23:* (Special Topics) Statistical Rigor and Bias, Missing Data, Causal Inference, Final Exam Discussion	See Canvas	HW 8 Due
	(12/6,12/7)	*Lecture:* No Class, Additional Office Hours		
		Lab: Final Exam Review		
15	(12/11,12/12)	*Lecture:* No Class, Reading Period		
		Lab: No Lab, Reading Period		
16	(12/13)	Final Exam In-Class, 9:00 AM ET		

General Expectations

Readings

Please complete the required reading in advance of the lecture. This course will require dedicated time for full understanding of concepts.

Lecture and lab attendance

It is expected that students attend lectures and labs. This is important for understanding concepts and remaining engaged in the learning process throughout the semester. Students do not need to inform the instructor and/or TA if they will miss a lecture or lab. Students are encouraged to make arrangements with other students in the classroom to acquire missed lecture notes. Students with excused and/or long-term absences should discuss this with their academic advisor and instructor.

Late or missed homework assignments

Homework (HW) assignments must be turned in online at or before the posted due date which is Wednesday of the specified week. Should you turn your homework in late, you will lose 10% points for every additional late day up to two days total or Friday of the specified week. This means assignments will not be accepted after Friday of the specified week.

Credit hours and time expectations

Over 13 weeks, students are expected to spend 4 hours per week in class (lectures and lab), 6 hours on assigned reading, and 4 hours on homework (162 hours total). The quiz is expected to take up to 4 hours maximum. The midterm and final exams are expected to take up to 14 hours, for an overall total of 180 hours. The 13 weeks is out of the 16 weeks outlined in the tentative schedule where weeks 12, 15, and 16 are excluded due to no lecture and/or no lab.

Academic integrity Plagiarism will not be tolerated in this course.

- Review the Academic Code: (`https://college.brown.edu/design-your-education/academic-policies`).
- Plagiarism occurs when you deliberately use someone else's original material without acknowledging source (i.e., carelessly or inadequately citing ideas and words from another source). Other forms of plagiarism includes submitting an assignment or exam where some or all of the work is written by another person. You must properly document the sources used even if you are not directly quoting from another person's work. This includes citing online sources (e.g., Wikipedia, online books, Stack Exchange, Wolfram, ChatGPT and related tools, etc). If you have any questions about properly citing your work, please speak to the professor.
- **Additional comments on the use of artificial intelligence (AI) language models (e.g., ChatGPT):** Students may not use ChatGPT tools to complete homework assignments. ChatGPT may be used as a study guide for the course but not for completing homework or the take-home quiz. If you have any questions about using ChatGPT or other AI language models, please speak to the professor.
- All situations of suspected academic dishonesty will be handled in the manner as described in the Academic Code.

Communication about the course

Please check your Canvas regularly so that you can receive emergency communications and periodic updates about the class.

Communication with the teaching team

The preferred method of communication with the instructor is e-mail. Please use your

Brown email account to send correspondences. Your emails should be professional in nature. However, students should not wait until after Friday to send an email if a response is needed before the following Monday. Weekly office hours are the best time to ask the instructor questions regarding course materials.

Direct communication with the TA's will take place during office hours only. Please do not email the TA's with questions about class materials and do not request individual meetings with the TA's. Questions about class materials should be asked during office hours.

Class recording and distribution of materials

All lectures and course materials are copyrighted. Students are prohibited from reproducing, making copies, publicly displaying, selling, or otherwise distributing the recordings or transcripts of the materials. Disregard of the University's copyright policy and federal copyright law is a Student Code of Conduct violation.

Students with accommodations approved by Student Accessibility Services should speak with the instructor and TA's before recording lectures and/or meetings for their private use.

Course syllabus subject to change

The syllabus content and course schedule may change as the class progresses. Please check Canvas regularly for updated information.

Academic Support

Diversity and inclusion

A diversity of backgrounds, skills, and perspectives is critical to addressing public health issues. We all benefit from diversity in the classroom. It is the responsibility of all of us to maintain an environment that is respectful and supportive. Brown University is committed to full inclusion of all students. The instructor and TAs intend to present materials and activities that are respectful of diversity: gender identity, sexual orientation, disability age, socioeconomic status, ethnicity, race, culture, perspective, and other background characteristics. Please keep in mind that some studies that will be discussed in class that may have not considered diversity when the data were collected. In those cases, the instructor and TAs will do our best to provide context.

Accessibility and accommodations

Brown University is committed to full inclusion of all students. Please inform the instructor early in the semester 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.

Health and safety statement for course meetings

The class will be held in-person and office hours may be in person over Zoom. During both lecture and lab, a face mask is not required in university facilities. However, for in-person meetings with the instructor or TAs, I will ask students to wear a face mask when meeting with me if in close proximity.

Use of technology to support student learning in your course

This course will use the following technological platforms: Canvas, R/RStudio software, Top Hat, and Zoom. If you have any concerns or questions about access or the privacy of any of these platforms, please reach out to me. The IT Service Center (`https://it.brown.edu/get-help`) provides many IT Services including remote assistance, phones, tickets, and chat. Please also see the Online and Hybrid Learning Student Guide.