PHP0100: Statistics is everywhere
Fall 2016

Primary Instructor and Co-Instructors:
Zhijin Wu (Zhijin_wu@brown.edu)
Constantine Gatsonis (gatsonis@stat.brown.edu)
Chris Schmid (christopher_schmid@brown.edu)
Adam Sullivan (adam_sullivan@brown.edu)

Class Meeting Time:
Tue Thu 01:00 pm - 02:20 pm
Location: J. Walter Wilson 303

Instructors Office Hours:
By appointment

Course Description:
Analysis of data continually shapes our lives in many contexts that we may not recognize. How does Google decide how to rank pages for individuals? How does Amazon know how to target suggestions to customers? How can we predict hurricanes, heat waves and droughts? How do doctors determine if a medical treatment works or political strategists know which voters to target? All of these problems involve making inferences from data and statistics is the universal language that enables us to draw these conclusions. This seminar will expose students to a variety of problems encountered in the media, in science and in life for which solutions require drawing inferences from data. We will first introduce some basic concepts in probability and statistics such as randomness, probability, odds, variation, statistical significance, accuracy and precision that students have likely been exposed to in life and common media, but have not learned formally. The course will then alternate between two types of reading and discussion: 1) material from books and scientific news that illustrate statistical thinking and applications in politics, economics, forensics, physical sciences, social sciences and humanities; 2). student identified materials from the news or other reading. The instructors will provide background to help the students understand the technical aspects intuitively. Students will discuss interpretation of published analysis and will practice challenging or defending certain conclusions. The course will use simulation to illustrate basic concepts. Programming experience is not required but students will use simulation code provided in the course to assist their learning.

Though the topics have been pre-selected, students will participate in identifying reading material and steering the directions within each topic.

Grading:
The grades are Satisfactory/No Credit. Students are expected to do exercises assigned on canvas, lead discussion in one topic (identified from the topics in Week 5-13) with a partner, and participate in classroom discussion throughout the semester. Attendance is required.
Outline:

Week 1. (Sept 8; Instructor: Zhijin Wu) Introduction. Randomness and probability: Introduces the concept of randomness, probability and independence. We will illustrate using R actively in class and show how to simulate a simple experiment with a binary outcome with equal probability of success and failure.

Week 2 (Sept 13, 15; Instructor: Zhijin Wu) statistical significance: use and misuse. In this lecture we introduce the concept of hypothesis testing and statistical significance. We will illustrate repeated testing via R simulation, and have the students visualize the randomness and distribution of p-value. We will then discuss how it can be misused and/or misinterpreted.

Week 3 (Sept 20, 22; Instructor: Adam Sullivan) Estimation: Introduces the concepts of population quantities and sample quantities, and discusses the sources of randomness. Discusses the intrinsic variation of a random quantity and its relationship to, as well as difference from, the “error” in estimation. We will illustrate using R in class how estimates can be obtained, and how these vary. Student exercise and discussion: students will breakup in teams of two, with one generating data and knowing the truth and the other estimating with subsamples repeatedly. Discuss the costs and benefits of different strategies.

Week 4 (Sept 27, Sept 29; Instructor: Chris Schimid) Conditional probability: Is smoking not so bad because most smokers do not get lung cancer? Is some substance necessarily bad if we find it in 100% of cancer cells? This class introduces the concept of conditional probability, risk and relative risk, odds and odds ratio. We will provide a simulated dataset and code for students to exercise on estimating conditional probabilities and marginal probabilities. In class discussion we will interpret the results mathematically and interpret its implications in the context. Students will also observe how confounding occurs.

Week 5 (Oct 4, 6; Instructor: Chris Schimid) Election: Can you predict election results? What does a poll tell you and what is that margin of error? How do you make sense or use of multiple polling reports? We will discuss how statisticians approach these questions, the methodology involved, and share some past results – possibly new ones if good examples emerge.

Week 6: (Oct 11, 13; Instructor: Zhijin Wu) Birthdays: Are birthdays randomly and uniformly distributed? Is there association between the date of birth and your future? We will visualize, and analyze actual birth date data, and discuss how relationships between birthday and other outcome can be revealed and/or tested.

Week 7 (Oct 18, 20; Instructor: Zhijin Wu) Movies and statistics Are there general components of a good movie, or a blockbuster? Do online reviews affect box office? How did IMDB compile their list of best movies?

Week 8 (Oct 25, 27; Instructor: Constantine Gatsonis) Ranking and Comparing How do you rank athletes, schools or hospitals? What scores and statistics have been used in these comparisons?

Week 9 (Nov 1, 3; Instructor: Adam Sullivan) Data Visualization How to present and interpret information? What are the principles in data visualization?

Week 10 (Nov 8, 10; Instructor: Zhijin Wu) Big Data What is it? What can we do about it? Does “big data” mean no more uncertainty?
**Week 11** (Nov 15, 17; Instructor: Chris Schmid confirmed) **Sports Statistics** Are patterns of great streaks in athlete or team performance real or by chance? Can you measure it? Can you predict outcomes of a game? Are scoring systems in sports fair?

**Week 12, 13** (Nov 22, 29; Instructor: Constantine Gatsonis) **Forensics.** How are probability models and statistical techniques applied in law and order? How to quantify the weight of forensic evidence?

**Week 13, 14** (Dec 1, 6; Instructor: Adam Sullivan) **Causal Inference** What is the difference between causal connection and association? Can you establish causal relationship and how?

**Time expectations**
Time expectations for this Lecture Course: Over 14 weeks, students will spend 3 hours per week in class (42 hours total). Required reading and homework is expected to take up approximately 7 hours per week (98 hours total). In addition, approximately 40 hours of preparation is assumed necessary for the topic a student is leading the discussion.

**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 can also contact Student and Employee Accessibility Services (SEAS) by email at SEAS@brown.edu or by phone at 401-863-9588 to discuss the process for requesting accommodations.