STA 111: Probability and Statistical Inference

Duke University, Summer Term II, 2018

Instructor: Dr. Shaobo Han

Office: Old Chemistry Bldg., Room 206B

Email: shaobo.han@duke.edu

Meeting Dates: 07/02/2018 - 08/09/2018

- **Lecture:** Mon, Tue, Wed, Thu, and Fri 11:00 am – 12:15 pm, Social Sciences 311
- **Lab:** Mon and Wed, 1:30 pm – 2:45 pm, Social Sciences 124

Office Hours: Tue and Thu 4:00 pm – 5:00 pm and by appointment, Old Chem 211A

Other Required Materials: Calculator

Course Website: https://shaobohan.net/courses/sta111su18/

Sakai Link: Click here

Course Description

This class will provide the necessary probability and statistical background needed for students to go on to study economics, financial statistics, engineering and more advanced methods of quantitative analyses in the natural and social sciences. We will learn about the basic probability laws, random events, independence and conditional independence, expectations, and Bayes theorem. We will also cover discrete and continuous random variables, density and distribution functions, point estimation, maximum likelihood estimation, confidence intervals, Bayesian inference, hypothesis testing, simple linear regression, multiple linear regression and much more. Labs will be taught using R and RStudio and will emphasize exploratory data analyis and implementation of inference procedures introduced in lecture. A detailed course schedule will be maintained throughout the semester.

Prerequisites:

Students should have some background in differential and integral calculus – i.e. you should be able to take derivatives and integrals of standard functions (exponential, polynomials, logarithms) and should understand the basic applications of calculus (finding areas under curves, maximizing/minimizing functions, etc.).

Textbook:

- Probability and Statistics (4th Edition), by Morris H. DeGroot and Mark J. Schervish
- (Optional) OpenIntro Statistics (3rd Edition), by David M. Diez, Christopher D. Barr, and Mine Çetinkaya-Rundel

Course Format

- **Lectures** will include a blend of theory and applications. Slides will be available on the website by 11:59pm the day before each class and are provided in advance to allow preparations before class. Make sure to read the slides before class and perhaps print them so you can focus better and take notes. You are responsible for all the material covered in class and assigned textbook readings. Ask questions in class, during office hours or send e-mails but do not wait until the last minute.
- **Homework assignments** are to help you develop a better understanding of the material covered in class and prepare for exams — you should take them seriously! For each question, you must show ALL work to receive credit. You are encouraged to work with each other on the homework problems, but you must turn in your own work. You will download homework assignments from Sakai. Homework must be turned in at the beginning of class on the date indicated.
- **Lab assignments** aim to provide you hands-on experience with data analysis using modern statistical software RStudio, which is a front end for the R statistical language with Rmarkdown and knitr support. You can use RStudio by signing on to `https://vm-manage.oit.duke.edu/containers`. The labs will facilitate the understanding of concepts and methods discussed in lectures. You are graded on lab reports that must be submitted by 11:59pm the same day on Sakai.
- **Exams:** The midterm exam will be held in class on Friday, 07/20/2018. A cumulative final exam will be held during our scheduled exam period on Saturday, 9:00AM – Noon, 08/11/2018. You can bring a handwritten cheat sheet prepared by you (8.5×11 inches, you may use both sides). There will be no make-up exams.

Class Materials: Lecture notes, labs and other reading resources will be posted on the course website while homework assignments will be posted on Sakai.

Evaluation:

- Your final grade will be computed as a weighted average of:

Component	Percentage
Class Attendance	5%
Homework	30%
Lab Reports	10%
Midterm	25%
Final Exam	30%

- Grades may be curved at the end of the semester. Cumulative averages of 90% – 100% are guaranteed at least an A-, 80% – 89% at least a B-, and 70% – 79% at least a C-, however the exact ranges for letter grades will be determined after the final exam.
- There will be approximately 10 homework assignments and 8 labs. The lowest homework score and the lowest lab report score will be dropped. There will be no labs on the sixth week to provide you with more time to prepare for the final exam.
- See the instructor in advance of relevant due dates to discuss possible alternatives. Missed lectures, labs, and homework are excused only by presenting an official Dean's excuse. You are not allowed to present more than 1 official excuse.

  Late Submission Policy:

  – You will lose 25% of points on each homework if you submit a day after it is due and 100% if you submit later than that.

  – You will lose 50% of points on each lab if you submit a day after it is due, and 100% if you submit later than that.

Course Schedule [1]**:**

Week 1:
Overview. Interpretations and definition of probability, experiments and events, summary statistics and histograms, permutations and combinations, conditional probability, independent events, law of total probability and Bayes' theorem.

Week 2:
Introduction to random variables, probability mass functions, cumulative distribution functions, discrete distributions, probability density functions, continuous distributions, marginal, joint and conditional distributions, expectations – mean, variance, covariance and correlation, and introduction to some special distributions – Bernoulli, Binomial, hyper-geometric, Poisson, negative binomial and normal distributions. Functions of random variable, linear combination, the law of large numbers, and central limit theorem.

Week 3:
Point estimation, bias and variance, unbiased estimators, maximum likelihood estimators and their properties. Experimental design, observational study, confounding factor, subgroup analysis and weighted average.

Week 4 :
Sampling distribution of a statistic, confidence intervals and interpretation, some specific confidence intervals, Bootstrap confidence interval, Introduction to Bayesian inference, prior and posterior distributions, conjugacy, credible intervals, hypothesis testing (one-group), the Student's t distribution.

Week 5:
Hypothesis testing (two-group), p-value, Type I and II errors, power calculations, and sample size determination. Contingency tables, Chi-square distribution, tests of independence and goodness of fit, Simpson's paradox. One-way ANOVA, the method of least squares and simple linear regression,

Week 6:
Residual analysis, transformation, introduction to multiple regression, model selection, multicollinearity and cross-validation.

Important Dates:

Independence Day Holiday (No class) . . . . . Wednesday, 07/04/2018
Midterm Exam . . . . . Friday, 11:00AM – 12:15PM, 07/20/2018
Final Exam . . . . . Saturday, 9:00AM – Noon, 08/11/2018

Both midterm and final exams will be held in Social Sciences 311.

[1] This is a tentative outline and it will be updated as we proceed. See the course website for a detailed schedule.

Academic Integrity

Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and non-academic endeavors, and to protect and promote a culture of integrity. Cheating on exams and quizzes, plagiarism on homework assignments and projects, lying about an illness or absence and other forms of academic dishonesty are a breach of trust with classmates and faculty, violate the Duke Community Standard, and will not be tolerated. Such incidences will result in a 0 grade for all parties involved as well as being reported to the University Judicial Board. Additionally, there may be penalties to your final class grade. Please review Duke's Standards of Conduct.

Disability Statement

Duke University is committed to providing equal access to students with documented disabilities. Students with disabilities may contact the Student Disability Access Office (SDAO) to ensure your access to this course and to the program. There you can engage in a confidential conversation about the process for requesting reasonable accommodations both in the classroom and in clinical settings. Students are encouraged to register with the SDAO as soon as they begin the program. Please note that accommodations are not provided retroactively. More information can be found online at `http://access.duke.edu/students/index.php` or by contacting SDAO at 919-668-1267, `SDAO@duke.edu`