STA 111: Probability & Statistical Inference

Summer 2018 Term I – May 16, 2018 to June 28, 2018

Duke University

Instructor: Olanrewaju Michael Akande

Email: olanrewaju.akande@duke.edu

Course Website: https://akandelanre.github.io/STA111-Summer2018-Course-Website/

Lectures: Mon, Tues, Wed, Thurs and Fri, 11:00am - 12:15pm, Allen 103

Labs: Mon and Wed, 1:30pm - 2:45pm, Allen 103

Office: Old Chemistry 222

Office Hours: Tues and Thurs 12:30pm - 1:30pm, Old Chemistry 211A

Textbook: Probability and Statistics (4th edition) by De Groot and Schervish

Other Required Materials: Calculator

Exam Dates: Midterm Exam – Wednesday, June 6, 11:00am – 12:15pm, Allen 103

Final Exam – Wednesday, June 27, 2:00pm – 5:00pm, Allen 103

Holidays: Memorial Day Holiday – Monday, May 28

Course Overview

Probability and statistical inference provide tools needed to answer a very broad range of interesting questions in various fields. For example, how likely was Loyola IL to make the final four at the beginning of the 2017/2018 season or what is the probability that you will make six or more digits annual salary after graduation if you move to New York? This is an introductory level class that aims to provide some of the theoretical probability and statistical inference background needed for students to go on to study advanced quantitative analyses in the natural and social sciences, financial statistics, and so on. You will learn the basic laws of probability, random events, independence, expectations and Bayes theorem. You will also learn discrete and continuous random variables, density and distribution functions, point estimation, confidence intervals, Bayesian inference, one and two-sample mean problems, simple linear regression, multiple linear regression, and much more. While this is an introductory class with most students being freshmen, higher level students can also take the class as a first-level theoretical refresher course for a more advanced class or simply out of interest. All students must have some background in calculus (see prerequisites) to be able to keep up with and understand the materials.

Learning Objectives

By the end of this class, students should be able to

- Define probability, random variables, probability density and mass functions, probability distribution functions and use probability results for statistical inference.
- Derive and verify some well-known statistical results, for example, maximum likelihood estimators for common distributions.
- Apply the probability results and statistical models developed in class to real data from economics, public policy, social science, and so on.
- Summarize and analyze data using R or Stata.

Prerequisites

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

Course Format

- Lectures will include a combination of notes/slides, working through theory in class and some group work to understand applications. Notes will be available on the website by 11:59pm the day before each class and are provided in advance to allow you pay more attention in class. Make sure to read the notes before class and perhaps print them so you can focus better and take notes instead. You are responsible for all the material covered in class and assigned textbook readings. Ask questions in class, during office hours or send an e-mail, but do not wait until the last minute.
- There will be 15-min in-class quizzes Tuesdays and Thursdays.
- Homework assignments will be posted immediately after class on Fridays and will be due at the beginning of class the following Wednesday. The assignments are to help you develop a better understanding of the material covered in class and prepare for exams, so take them seriously! 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. If you copy someone else's work, both

parties will fail the assignment and be reported to the Office of Student Conduct. If you have any questions about what constitutes plagiarism, do not hesitate to ask.

- Lab assignments MUST be submitted by 11:59pm the same day. The objective of the lab is to give you hands-on experience with data analysis using modern statistical software. We will use a statistical analysis package called RStudio, which is a front end for the R statistical language.

Class Materials

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

Grading

- There will be no make-up for quizzes, homework assignments, and exams.
- Your final grade will be determined as follows:

Component	Percentage
Class Participation	5%
Lab Reports	10%
Quizzes	15%
Homework	20%
Midterm	20%
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 8–10 quizzes, 5 homework assignments, and 8–10 labs. The two lowest quiz scores and the lowest homework score will be dropped (this should give you enough cover for genuine and unavoidable absences). There will be no labs on the sixth week to provide you with more time to prepare for the final exam.

Late Submission Policy

- You will lose 40% 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 *

Week 1 (Chapters 1-2)

Interpretations and definition of probability, experiments and events, summary statistics and histograms, permutations and combinations, conditional probability, independent events, and Bayes' theorem.

Week 2 (Chapters 3-5)

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, multinomial, gamma and normal distributions.

Week 3 (Chapters 6-7)

The law of large numbers, central limit theorem and continuity correction, Bayesian estimation and inference, prior and posterior distributions, conjugacy, maximum likelihood estimators and their properties, improving an estimator, sufficient statistics, distributions of linear combinations, and functions of random variables.

Week 4 (Chapters 8-9)

Sampling distribution of a statistic, confidence and credible intervals, interpreting confidence and credible intervals, some specific confidence intervals, unbiased estimators, the student-t and Chi-square distributions, simple hypothesis testing, type I and II errors, two-sided hypothesis testing, power calculations, and introduction to Bayesian hypothesis testing.

Week 5 (Chapters 10-11)

Tests of independence, goodness of fit tests, contingency tables, Simpson's paradox, the method of

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

least squares and simple linear regression, introduction to multiple, nonlinear and nonparametric regression, model validation and assessment tools, and one-way analysis of variance.

Week 6 (Chapters 11-12)

Two-way analysis of variance, simulation, bootstrap; REVISION.

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 nonacademic endeavors, and to protect and promote a culture of integrity. To uphold the Duke Community Standard:

- I will not lie, cheat, or steal in my academic endeavors;
- I will conduct myself honorably in all my endeavors; and
- I will act if the Standard is compromised.

Cheating on exams or plagiarism on homework assignments, 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. Additionally, there may be penalties to your final class grade along with being reported to the Undergraduate Conduct Board. Please review the academic dishonesty policies at `https://studentaffairs.duke.edu/conduct/z-policies/academic-dishonesty`.

Disability Statement

Students with disabilities who believe that they may need accommodations in the class are encouraged to contact the Student Disabilities Access Office at 919.668.1267 or disabilities@aas.duke.edu as soon as possible to better ensure that such accommodations are implemented in a timely fashion.

Other Information

For every lecture, you will need a simple calculator for quizzes, exams and homework. Graphical capability is not required, and questions are worded so that advanced calculators confer no advantage. I will provide you any other material needed for the quizzes and exam. Again, do not hesitate to come to my office during office hours or by appointment to discuss a homework problem or any aspect of the course, and

working in groups is highly recommended. Questions related to course assignments and honesty policy should be directed to me. DO NOT search for direct answers to homework questions online; ask me instead.