STA 111: Probability and Statistical Inference

Summer 2016 Session I – May 18, 2016 to June 30, 2016

Instructor: Olanrewaju (Michael) Akande

Office: Old Chemistry 222

Email: olanrewaju.akande@duke.edu

Course Website: Click here

Number of Units: 1

Lectures: Mon, Tues, Wed, Thurs and Fri, 11:00AM - 12:15PM, Social Sciences 311

Labs: Mon and Wed, 1:30PM - 2:45PM, Old Chemistry 101

Office Hours: Tues and Thurs 1:00PM - 2:00PM and by appointment, Old Chemistry 211A

Prerequisites: Students must 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 must understand the basic applications of calculus (finding areas under curves, maximizing/minimizing functions, etc.).

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

Other Required Materials: Calculator.

Exam Dates: Midterm Exam – Wednesday, June 8, 11:00AM – 12:15PM;
                Final Exam – Wednesday, June 29, 2:00PM – 5:00PM.

Holidays: Memorial Day Holiday – Monday, May 30.

Course Description: This class will provide the statistical background needed for students to study econometrics, financial statistics and more advanced methods of quantitative analyses in the natural and social sciences. We will learn about the basic laws of probability, random events, independence and dependence, expectations, and Bayes theorem. We will also cover discrete and continuous random variables, density, and distribution functions. Other topics include introduction to maximum likelihood estimation, Bayesian inference, one and two-sample mean problems, simple linear regression, multiple linear regression.

Learning Objectives:

- Develop tools for designing and analyzing basic statistical studies, and for understanding statistical methods.

- Provide the mathematical and statistical background needed to study econometrics, financial statistics, and other quantitative analyses in the natural and social sciences.

- Develop approaches to analyzing data in economics, public policy, social science, and the natural sciences.

- Use statistical software to summarize data and perform data analysis.

Course Format:

- Lectures will include a combination of theory and applications/examples. Notes/slides will "usually" be posted by 11:59PM the previous day. These notes are posted to allow you pay more attention in class so you should attempt to read them before class and expect to take notes in class as well. 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 in-class 15-min quizzes Tuesdays and Thursdays (the questions will be fairly similar to class examples and textbook exercises).

- Homework assignments will be posted immediately after class on Fridays and will be due beginning of class the following Tuesday. Homework assignments are to help you develop a

1

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. If you copy someone elses work, both parties will receive a 0 for the assignment and will be reported to the Undergraduate Conduct Board. 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. You can use RStudio by signing on to `https://vm-manage.oit.duke.edu/containers`.

Late Submission Policy:

- You will lose 25% if you submit after class and 100% if you submit the next day or later.

- You will lose 50% of the lab points for each lab if you submit the next day, and 100% if you submit later than that.

Grading:

- There will be no make-up for quizzes, homework assignments, and exams, thus class attendance is very (extremely!) important.

- Your final grade will be determined as follows:

Component	Percentage
Class Participation	5%
Homework	20%
Quizzes	20%
Lab Reports	10%
Midterm	20%
Final Exam	25%

- 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.

- See the instructor in advance of relevant due dates to discuss possible alternatives. Missed labs, homework, quizzes and exams are excused only by presenting an official Dean's excuse. You are not allowed to present more than 1 official excuse.

- 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.

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.

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

_____

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

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 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 non-academic endeavors, and to protect and promote a culture of integrity. 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.

General 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.

Important Dates:

Memorial Day Holiday . . . . Monday, May 30.

Midterm Exam . . . . Wednesday, June 8

Final Exam . . . . Wednesday, June 29