46880
Introduction to Probability and Statistics

Prof. Benjamin Moseley
4111 Tepper, (412) 268-2454, `moseleyb@andrew.cmu.edu`
Tepper School of Business
Carnegie Mellon University

Fall Mini 2, 2022

Description of the Course

This course introduces the fundamentals of probability theory and statistics. It covers the following topics: conditional probability and Bayes' theorem, decision trees, random variables, expectation and variance, discrete distributions, continuous distributions, the central limit theorem, sampling and sampling distributions, interval estimation, hypothesis testing, p-values, correlation and simple regression.

Learning Objectives

Broad objectives: Understand how to define and apply models of uncertainty in business, and how to incorporate them in decision making. Organize thinking about uncertainty starting with probability models describing a population, followed by analyzing data samples taken from a population and finally using statistical models to infer from these samples important parameters of the population.

Specific objectives:

- Basic Probability and Discrete Random Variables: Describe basic probability models and rules, specifically sample spaces and laws of probability, conditional probability, statistical independence and Bayes' rule. Define random variables that may take discrete values and their key properties such as expectation, variance and covariance; develop knowledge about commonly used distributions such as the Binomial and Poisson distributions.

- Decision Trees and Continuous Random Variables: As an application of discrete random variables, build decision trees to help find the best decisions when faced with discrete uncertainty and use them to estimate the value of sample information. Extend the definition of discrete random variables to those that may take continuous values, and as an example, understand the exponential distribution.

- Normal and Sampling Distributions: Define the Normal distribution and learn about its properties. Define sampling distributions. Understand the Central Limit Theorem and its importance in making statistical inference from samples.

- Interval Estimation and Hypothesis Testing: Derive point and interval estimates for the parameters of the population from which a sample is drawn; Do this specifically for the mean value and proportion in a population. Frame statistical claims as alternate hypotheses; Carry out a test of their validity using any of three equivalent methods: confidence intervals, critical values or p-values.

- Simple Linear Regression: Specialize the idea of hypothesis testing to correlation between two random variables via a simple one-variable linear regression; Implement a simple one-variable regression, interpret the results, define a goodness of fit measure. Use hypothesis testing to make inferences about the strength of the linear relationship; Use confidence intervals to make predictions. Apply simple transformations of the variables to check if it improves the fit.

- Advanced Hypothesis Tests Comparing Two Populations, Summary: Apply hypothesis testing to compare samples coming from two populations: carry this out for the difference between

their means and difference between proportions of binary events. Summarize the comprehensive lessons from the class.

Instructor

- Benjamin Moseley
  Tepper 4111, phone: (412) 268-2454, email: `moseleyb@andrew.cmu.edu`

Text

The instructor will present the lectures in the form of small videos. These videos will be available on canvas. As a supplement, students may want to work with a textbook (optional).

The recommended textbook in this course is Business Statistics: Communicating with Numbers, by Jaggia and Kelly, Second Edition 2016 (McGraw-Hill).

Software

This course will involve building models using the `Excel` spreadsheet program and probability functions. We will also use Excel's `Data Analysis` for statistical analysis.

Modules

The course is organized in six modules, each module covering a week of the class. Every module is tested by a homework, which will contain a set of problems each of which will have several parts of varying difficulty. The homework is available the weekend before the start of the module, and you have until the end of the week to work out the answers. During the following weekend, you must take an on-line quiz related to the homework. The quiz is administered on the course web page and lasts at most 90 minutes. Students take the on-line part individually, at a time of their choice in a window of two days (typically from Sunday noon EST to Tuesday noon EST).

You can work in groups to solve these homework sets but your work will be tested in individual online quizzes each week; you will need your own worked out answers for taking the quiz since you will not be able to collaborate while taking it. You are allowed to discuss and compare notes on the homework with other students until Sunday morning. However, while the window for taking the quiz is open, no collaboration or discussion related to the homework and quiz is allowed. You are allowed to use your notes, books, and laptop computer when you are taking the quiz.

There are actually two online quizzes per week, which we call first-quiz and second-quiz.

The first-quizzes are designed to keep you on track. You must complete the first-quizzes for each module by noon EST on Thursday of the corresponding week. Each first-quiz is worth 3% of the class grade and each second-quiz is worth 7%.

Grading policy

The course grade will be based on twelve quizzes and a final exam weighted as follows:

Each first-quiz	3%
Each second-quiz	7%
Final Exam	40%

The final exam is an individual assignment. The questions will be similar to those in the homework sets and quizzes, and to the problems discussed in class.

The exams will be **open notes, open book,** and **open notebook computer**.

Except for **provable** unforeseen circumstances, makeup exams will not be given. A student is required to inform the instructor as soon as possible in such a case. In case of an inevitable conflict, it is the student's responsibility to schedule a makeup exam with the instructor.

Academic integrity

Students at the Tepper School of Business, Carnegie Mellon are expected to abide by the University's and School's policies of academic integrity. Cheating and plagiarism are serious violations of these policies. They will be reported to the administration and subject to disciplinary sanctions.

Schedule Classes are on Tuesdays or Thursdays depending on the section.

Module	Schedule	Topic	Jaggia-Kelly Chapter	Readings
1	10/24	Probability models Discrete random variables	4.1-4.4 5.1-5.5	PROBA DISCRETE
2	10/31	Decision trees Continuous random variables	 6.1, 6.4	DECISION CONTINUOUS
3	11/7	Normal distribution Sampling distributions	6.2, 6.3 7.1-7.5	 SAMPLING
4	11/14	Estimation, Confidence intervals Hypothesis testing	8.1-8.5 9.1-9.4	 HYPOTHESIS
5	11/28	Simple regression Inference with regression	14.1, 14.2, 14.4 15.1, 15.3, 15.4	LINEARREG
6	12/5	Inference with two populations Inference about variance	10.1-10.3 11.1, 11.2	
	12/12	Review session	comprehensive	
	12/17-12/18	Final exam	comprehensive	

- PROBA: "A Brief Introduction to Probability".

- DISCRETE: "A Brief Introduction to Discrete Random Variables".

- DECISION: "A Brief Introduction to Decision Theory".

- CONTINUOUS: "A Brief Introduction to Continuous Random Variables".

- SAMPLING: "A Brief Introduction to Statistics, Sampling, the Central Limit Theorem and Confidence Intervals".

- HYPOTHESIS: "A Brief Introduction to Hypothesis Testing".

- LINEARREG: "A Brief Introduction to Linear Regression".