

Fundamentals of Statistics - ORF 245/ EGR 245 (Sp 16-17)

Syllabus

1 Lectures and precepts

Lectures: Mon., Wed., Fri., 10 am- 10:50 am, Friend Center 101. **By:** Samory Kpotufe.

Precepts (they all cover the same material):

- Mon., 3:30 pm to 4:20 pm, Sherrerd Hall 101. **By:** Huanran Lu.
- Mon., 7:30 pm to 8:20 pm, Sherrerd Hall 101. **By:** Wenyan Gong.
- Tues., 3:30 pm to 4:20 pm, Sherrerd Hall 101. **By:** Bachir El Khadir.
- Tues., 7:30 pm to 8:20 pm, Sherrerd Hall 001. **By:** Joe Zhong.

2 Contacts

Questions about the course: Samory Kpotufe (samory@princeton.edu).

Questions about grading: Huanran Lu (huanranl@princeton.edu).

Questions on assignments:

- Online on piazza: Jeff Zhang is primarily in charge of answering your piazza questions.
- In person, during office hours:
 - Tues. 9 am to 10:30 am: Huanran Lu (Sherrerd Hall, 107).
 - Tues. 1 pm to 2:30 pm: Joe Zhong (Sherrerd Hall, 107).
 - Wed. 1:30 pm to 3:00 pm: Jeff Zhang (Sherrerd Hall, 107).
 - Wed. 4:00 pm to 5:30pm: Samory Kpotufe (Sherrerd Hall, 327).
 - Thur. 9 am to 10:30 am: Wenyan Gong (Sherrerd Hall, 107).
 - Thur. 2:00 pm to 3:30 pm: Bachir El Khadir(Sherrerd Hall, 005).

3 What is the course about?

While some Statistics courses are meant to teach you how and when to apply various statistical methods, this course is different. The goal is for you to understand why Statisticians came up with those particular methods; at the end of the course you should be able to derive your own methods for various problems.

Reading Material

Required: *Jay Devore, Probability and Statistics, 9th Edition.*

Recommended: Sheldon M. Ross 5th Ed., Introduction to Probability and Statistics for Engineers & Scientists is a nice addition with different working examples.

4 Prerequisites

A good understanding of calculus, and some multivariate calculus. Basic counting (e.g., remember $n!$, $C(n, k)$ a.k.a $\binom{n}{k}$, $P(n, k)$, and such?) will be assumed. However, initial precepts will review counting.

The class will *not* teach programming (R). You will be expected to pick it up through the initial *guided* exercises and resources available on campus. Your first precepts will cover some of the various resources available to you (such as the McGraw center).

McGraw Center Tutoring: Time and location to be announced.

5 Course Evaluation

Exams 📖

There will be a midterm and final exam each worth 20% of your grade. Any material covered in lectures or in problem sets and programming assignments can make it into the exams.

Homework (problem sets) ☕

Weekly homeworks will be posted (usually by Friday eve) and due on Fridays at 5pm. Late homeworks will not be accepted, except under extreme circumstances.

A homework will consist of a number of problems¹, where each problem is worth 10 points. Together, homeworks will be worth 40% of a grade.

Collaboration: Homeworks can be discussed but must be written up individually. Homeworks will be *designed* to solidify your understanding of concepts seen in class.

Programming assignments 📄

Frequent programming assignments will be given. Each consists of a number of problems, each problem worth 10 points. Together, programming assignments will be worth 20% of a grade. Programming will be in R. No late programming assignment will be accepted.

Collaboration: Programming can be discussed but must be written up individually. You must write your own code, and turn in a printout of the code with your report. The aim of the assignments is to get concrete evidence of abstract concepts seen in class.

6 Overview of topics

We will *loosely* follow the order of topics below, keeping in mind that some concepts will often appear throughout the course and some are best learned together.

6.1 Introductory ideas [1 week]

- Sample (observations) vs. Population (underlying truth).
- Notions of typicality (simple ways to summarize samples or population): mean, median, mode.
- Proportion/relative-frequency, histograms (a more sophisticated summary).

Note: We'll also introduce concepts such as random variables, estimation, variance.

6.2 Probability (essentially, proportions in the population) [1 week]

- Events, and probabilities of events, union and intersection of events.
- Conditioning and Bayes rule. Independence of events. Independent trials.

6.3 Random variable (r.v.) and important distributions [3 weeks]

- Discrete or continuous r.v. – call it X – and its distribution (law governing the probability of events associated with X). Expectation and Variance of X .
- Sampling from a distribution (precept and programming).
- Joint distribution of r.v.s X, Y . Covariance of X, Y . Independent r.v.s. X, Y . Variance($X + Y$).
- Important distributions: Bernoulli, Binomial, Multinomial, Poisson, Uniform, Exponential, Normal (a.k.a Gaussian), χ^2 , t , \dots , will be discussed in lecture and assignments.
- Weak law of large numbers (using Chebyshev's inequality), Central Limit theorem.

Note: We'll also introduce concepts such as estimation error (mostly through programming).

¹Some problems will have solutions in the book. This is to encourage you to read the book.

6.4 Estimation (of distributions) and confidence [2 weeks]

- Maximum Likelihood principle: Bernoulli, Normal. (Bayesian MAP in precept and assignments).
- Bias, Variance principle: e.g. yields a different estimate of variance.
- Confidence intervals: Gaussian, Bernoulli.

6.5 Hypothesis testing [2 weeks]

- p -values, type-I and type-II errors, power of a test.
- Tests on a Normal population.
- Tests on a Bernoulli population.
- Large sample tests.

6.6 Regression [2 weeks]

- Linear regression (least-squares).
- NonLinear regression (polynomial, maybe kernel regression).

6.7 Classification (if time permits) [1 week]

Simple classification rules such as nearest neighbors, perceptron (a simple neural network).

... 