

Statistics 471 Spring 2015

Modern Data Mining

Tu, TH, 3:00-4:30pm

Linda Zhao: lzhao@wharton.upenn.edu, Office: 470-JMHH.

Office Hours: 3:00 – 5:00 pm Mondays or by appointment

TA: TBA

Office Hours: TBA

Course Description: Statistics has been evolving rapidly to keep up with the modern world. We will show how to modify and adapt simpler models to handle contemporary large and complex data with applications in finance, marketing, medical fields... you name it. The free but powerful software “R” will be used. At the end of the semester we hope that students not only learn the modern statistic methods but also become skilled in dealing with data of essentially any size.

Textbook: (Required and they are available online for free)

Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani,
An Introduction to Statistical Learning with Application in R (ISLR),
First Edition, 2013, Springer New York.

An e-version is available from the author’s website: <http://www-bcf.usc.edu/~gareth/ISL/>

We will cover Chapter 1 through Chapter 6 of ISLR.

Additional optional reading:

1. Peter Dalgaard,
Introductory Statistics with R,
Second Edition, 2008, Springer

Available as a pdf through Springer Link or
http://www.academia.dk/BiologiskAntropologi/Epidemiologi/PDF/Introductory_Statistics_with_R_2nd_ed.pdf

2. Trever Hastie, Robert Tibshirani, Jerome Friedman
The Elements of Statistical Learning: Data Mining, Inference, and Prediction (ESL)
Second Edition, 2008, Springer

A pdf version is available: <http://statweb.stanford.edu/~tibs/ElemStatLearn/>

Canvas: <https://canvas.upenn.edu>

Most of the materials including announcements, data, codes, homework/projects, solutions, etc. will be available on our Canvas site.

Computer package: The statistical computing language R will be used. It is available through www.R-project.org for all common computing platforms such as Windows, Mac and Linux.

R tutorial: TBA at Wharton Computer Lab: JHMM375.

Homework and Projects: Six homework/projects are given. The lowest grade will be dropped.

Exams: There will be an in class, open book midterm; a two-day short take home mini project involving R (individual) and a final project and two very short (10 minutes) in class quizzes that will be simple multiple-choice questions.

Midterm: TBA

Final Project: TBA

Two Quizzes:

TBA

TBA

Take Home mini project:

TBA

Grade allocation:

Homework and Projects: 30%

Quizzes and take home min project: 20%

Midterm: 30%

Final Project: 20%