Statistics 112 Fall 2015

Statistics

Linda Zhao: lzhao@wharton.upenn.edu, Office: 470-JMHH.
Office Hours: 3:00 – 5:00 pm Tuesdays or by appointment

Course Description: Data science is playing an increasingly crucial role in the current world, and its importance will continue to increase in the future. It requires a combination of domain knowledge, advanced technical and computational knowledge, and solid fundamental statistics. In this class we will focus on some statistical foundations and core techniques such as regression and logistic regression combined with applications in social science, business and medical science. A popular, powerful software R will be used. Students should be able to handle many of the data analyses encountered in current massive information world afterwards. A quick summary of the topics: inference for one population and two population means using t-tests and nonparametric tests such as Wilcoxon signed rank tests; simple and multiple regression; model building through the Cp/BIC criterion; logistic regression. The "Lasso", a method that can handle high dimensional scenarios, may be covered if time permits.

Textbook:

1. *The Statistical Sleuth,* Ramsey/Schafer, Third edition, 2013. Cover from Chapter 1-3 (a quick review), 4, 5, 7-13 and 20-21. (Browse from the Internet; you may find a cheaper, soft cover version.)

2. An Introduction to R: http://cran.r-project.org/doc/manuals/R-intro.pdf

3. (Option) *Introductory Statistics with R*, Peter Dalgaard, Second edition, 2008, Springer.
   OPTIONAL
   Available as a pdf
   http://www.academia.dk/BiologiskAntropologi/Epidemiologi/PDF/Introductory_Statistics_with_R__2nd_ed.pdf

Canvas: Most of the materials including announcements, lecture notes, homework, 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. This software is open-source, and freely available.

R tutorial: TBA at Wharton Computer Lab: JHMM375.

Homework: 5 – 6 regular homework assignments are given. The lowest grade will be dropped.

Exams/Projects:

One evening, open book midterm - a laptop is needed: basic questions plus an onsite data analysis with R

A take home final project: As a prelude/preparation for our final project. I will bring you a couple of data sets that you can choose from for your process of data analyses. A report is needed

which includes

i) Goal of the study and the findings
ii) Summary of the data
iii) Detailed analyses
iv) R-code

Five very short (10 minutes) in class quizzes: simple multiple-choice questions.

Exams/Projects schedules:

Quizzes:

Quiz 1: Monday,	**09/14**
Quiz 2: Wednesday,	**09/30**
Quiz 3: Monday,	**10/19**
Quiz 4: Monday,	**11/23**
Quiz 5: Tuesday,	**12/08**

Midterm: 6:00 pm – 8:00 pm, **Wednesday**, **Oct. 28th**

Final Project Due: By 11:59 pm **Sunday**, **Dec. 13th**

Grade allocation:

Homework: 28%
Quizzes: 12%
Midterm: 30%
Final Project: 30%

(The lowest homework score and lowest quiz score will be ignored in the final grade.)

Schedules: some adjustment maybe needed as semester proceeds.

Lecture	Date	Contents	Note
1	08/26/Wed	Review: Normality, sampling distribution, one population and two population means. Ch 1-3, **3 lects**	
2	08/31/Mon		
3	09/02/Wed		
	09/07/Mon	Labor Day	
4	09/09/Wed	Nonparametric tests Ch4, **2 lects**	
5	09/14/Mon		**Quiz 1**
6	09/16/Wed	Simple linear regression, ch 7-8, **4 lects**	
	09/18/Fri		**Homework 1 Ch1-3**
7	09/21/Mon		
8	09/23/Wed		
9	09/28/Mon		
10	09/30/Wed	Multiple reg, Ch 9, **3 lects**	**Quiz 2**
	01/02/Fri		**Homework 2 Ch4**
11	10/05/Mon		
12	10/07/Wed		
	10/08/Th	Fall Break	
13	10/12/Mon	Multiple reg, inference, Ch 10, **3 lects**	
14	10/14/Wed		
15	10/19/Mon		**Quiz 3**
16	10/21/Wed	Model Checking, Ch 11, **2 lects**	
	10/23/Fri		**Homework 3** Multiple reg
17	10/26/Mon		
18	10/28/Wed	Review	
	10/28/Wed	**Evening Midterm, open book, laptop needed**	**Midterm, 6-8pm**
19	11/02/Mon	Variable selection, Ch 12, **2 lects**	
20	11/04/Wed		
21	11/09/Mon	One way ANOVA, Multiple comparison Ch 6.4,	**Homework 4** Case study

		2 lects	
22	11/11/Wed		
23	11/16/Mon	Two way ANOVA, Ch 13, 3 lects	
24	11/18/Wed		
25	11/23/Mon		**Quiz 4**
26	11/25/Wed	Logistic Reg, Ch 20, **4 lects**	**Homework 5** One/Two way
	11/26/Th	Thanksgiving	
27	11/30/Mon		
28	12/02/Wed		
29	12/07/Mon		
	12/08/Tu	Last Day	**Quiz 5**
	12/13/Sun		**Final Project Due**