

Syllabus for STAT 431, Sections 001–002
Statistical Inference
Spring 2016

Course information:

Lecture: TR 10:30 am–12 pm, 12 pm–1:30 pm, G60 JMHH

Professor: Po-Ling Loh

Office: 466 JMHH
Email: loh@wharton.upenn.edu
Office hours: 2-3 pm on Thursdays (or by appointment)

Teaching assistants:

Sameer Deshpande: dsameer@wharton.upenn.edu
Office hours: TBD

Laura Luo: pailuo@wharton.upenn.edu
Office hours: TBD

Course description: This course offers an advanced undergraduate level exploration of statistical techniques for data analysis, with an emphasis on developing computational tools and an understanding of when and how to use them. The latter will require a level of mathematical maturity as we examine the theoretical underpinnings of the explored methods. Interpretation of the results and analysis of assumptions is a key part of the course. As such, the course is appropriate for mathematically inclined students who wish to learn hands-on computational techniques for data analysis. Topics include (1) collection, summary, and display of data, (2) estimation, hypothesis testing, and confidence intervals, and (3) simple and multiple linear regression. Students will experiment with these ideas on data examples using statistical software.

Prerequisites: STAT 430. Students are expected to be fluent with quantitative probabilistic reasoning and analysis (e.g., probability distributions and densities; jointly distributed random variables; conditional probability; independence; correlation and covariance; normal and binomial distributions). Students are not expected to have knowledge of the statistical programming language R, although some prior programming experience is helpful.

Text: *Statistics and Data Analysis*, Tamhane & Dunlop, ISBN 0-13-744426-5.

Homework: Homework will be assigned fortnightly and due on Thursdays. Assignments must be handed in during class time or in a box in the main office (400 JMHH) by **1:30 pm** on the date the assignment is due. **No late homework will be accepted**, but the lowest score (of seven assignments) will not be counted toward the final grade.

Exams: There will be an one midterm and one final exam, scheduled for the following dates:

- Midterm: Thurs, Mar 3 (in class)
- Final: Thurs, May 5 (6–8 pm)

Students are expected to show up for both exams at the times specified above. In the case of a medical emergency (requiring a doctor's note), a make-up midterm may be provided. A make-up final will only be scheduled for Sep 1, 2016. The final exam will be cumulative, with an emphasis on the material covered since the midterm.

Written project: In addition to homework and exams, a short written project will be assigned concerning statistics in the news media (details will be provided in class). It will be due by e-mail at **midnight on March 20**. Students with exceptionally written projects may be invited to present their work and lead a short in-class discussion for extra credit.

Grading:

- Homework: 20% (lowest score dropped)
- Written project: 15%
- Midterm exam: 30%
- Final exam: 35%

Class participation: Although no official weight is attached to class participation, students are encouraged to participate actively in class discussions. Critical thinking and problem solving play an important role in statistical thinking, and hearing opinions offered by other classmates will be beneficial to everyone. Thus, making an effort to participate regularly in class discussions may have a positive effect on one's final grade.

Collaboration policy: Working together on homework is allowed and encouraged. However, students must write up their homework solutions by themselves. Names of collaborating students should be provided on the front page of each homework write-up.

Statistical computing: Each homework assignment will involve some component of statistical analysis on dataset(s) provided. It is recommended that students download and use R for this purpose. Other software such as Excel or Matlab may also be used for data analysis on homework sets, and knowledge of R (or other statistical software) will *not* be tested on exams. However, R will be the language referred to in class, so students who choose to use another statistical computing platform will need to figure out the equivalent commands on their own.

Course announcements: Official class announcements, assignments, etc., will be posted on Canvas. However, all students should also join the Piazza page, which will be used to answer questions in a more informal setting:

<http://piazza.com/upenn/spring2015/stat431>