# Statistics 431: Statistical Inference

## Syllabus, Fall 2011

| | |
|---|---|
| **Classes:** | Section 001, Mon/Wed 1:30–2:50 p.m., in F55 JMHH |
| | Section 002, Mon/Wed 3:00–4:20 p.m., in F55 JMHH |
| | |
| **Instructor:** | Emily Fox |
| **Email:** | ebfox@wharton.upenn.edu |
| **Office:** | 466 JMHH |
| **Office hours:** | Thurs 10–11 a.m. |
| | |
| **Teaching Assistants:** | Yang Jiang (yajiang@wharton) and Ville Satopää (satopaa@wharton) |
| **Office:** | 434 JMHH and 431.3 JMHH |
| **Office hours:** | Mon 4:30–5:30 p.m., Tues 4:30–5:30 p.m. and 8:00-9:00 p.m. |

## Course overview

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 statements, and (3) simple and multiple linear regression. If time permits, we will also discuss variable selection and logistic regression. Students will experiment with these ideas on data examples using statistical software.

## Prerequisites

The official prerequisite is Statistics 430. The effective prerequisite is *fluency* with basic 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), together with the kind of mathematical maturity that often comes from taking at least one higher level undergraduate subject that has a significant mathematical component. Students are *not* expected to have knowledge of the statistical computing language R, though prior programming experience will be helpful.

When in doubt, students should consult the staff to determine if they have had subjects that are effectively equivalent to the official prerequisites.

## Recitations

In addition to office hours, the TAs will hold one *optional* recitation per week **Thurs 7:00-8:00 p.m.**, location TBA. This recitation will review key concepts from lecture, work through example problems, and provide essential tutorials on the statistical computing software used in the course.

**Textbook**

There is no textbook that exactly matches the content of this course and the style in which it will be taught. The first half of the course will closely correlate with portions of the required textbook:

[Required:] *Statistics and Data Analysis: from Elementary to Intermediate*, by A. C. Tamhane and D. D. Dunlop, Prentice Hall, 2000.

During the second half of the course our attention will turn to regression for which the following recommended textbook has a nice treatment:

[Highly Recommended:] *The Statistical Sleuth: A Course in Methods of Data Analysis*, by F. Ramsey and D. Schafer, Duxbury Press, 2002.

Due to the lack of one comprehensive text, it is strongly recommended that you utilize lectures and the associated notes. *Lecture notes will not be posted on the course website.* Students are expected to take notes during lecture or to get notes from another student in the course.

**Course Website and Email**

We will make announcements via email, and we will post various information and handouts on the course web site:

http://www-stat.wharton.upenn.edu/~ebfox/teaching/stat431/

In addition, there will be a Google Group for Statistics 431. Students are *required* to register for this moderated group. The group website is:

http://groups.google.com/group/stat-431

To register for the group, go to the above link and request membership by clicking on "Apply for Membership" and *include your* **full name** *in the "additional information" text box.* All course emails will be made via this group. Additionally, the group has a discussion board for posting questions to TAs and other students.

**Statistical computing software**

The statistical computing software R (version 2.10.0 or higher) will be used in the course. It is free, and can be downloaded at the R-project website:

http://www.r-project.org/.

The website also contains a list of manuals for using the software. Basic usage of R will be illustrated in recitation and through sample codes posted on the course website, and *no previous exposure to the software is required.*

**Homework assignments**

- There will be homeworks assigned weekly with a few exceptions.

- Homework assignments will be posted on the course website, and after the due dates, solutions will be posted.

- Howework will be due in class, *at the beginning of the class*. Typically, the due date will be during the Wednesday lecture, a week after being assigned.

- If for whatever reason you cannot attend lecture on the homework due date, you may hand in your solutions to the Statistics Department (JMHH, 4th floor) in the box marked with your section. This box will be collected by 12pm prior to lecture.

- *No late homework will be accepted.* However, the lowest score will not be counted toward your final grade.

- Regrade requests must be made in writing as a cover letter and submitted to one of the TAs. All problems on the resubmitted homework will be regraded, not just the problem(s) in question.

- Scores for homeworks are finalized one week after the graded copies are handed back. Thereafter there will be no changes and no regrading. Do not delay checking your graded homeworks to the end of the semester.

- In writing your solutions, make them complete. You should show your work and explain your conclusions clearly and precisely.

- Students may moderately collaborate in the form of joint problem solving with one or two classmates, but are expected to prepare the final write-up individually with acknowledgment of the help received.

**Exams**

- One in-class midterm exam: Wednesday, October 26. Location TBA.

- Final exam schedule: 6-8 p.m., Friday, December 16. Location TBA.

- Both exams will be closed book. However, you will be allowed to bring a certain number $8.5 \times 11$-inch sheets of hand-written notes. Details will be provided two weeks prior to each exam.

**Grading policy**

The final grade in the course is based upon our best assessment of your understanding of the material during the semester. Roughly, the weights used in grade assignment will be:

- Homework assignments: 20% (with the lowest score dropped)
- Midterm exam: 30%
- Final exam: 50%

However, as always, other factors such as contributions to the lecture discussion and other interactions can make a significant difference in the final grade.

**Syllabus and Schedule**

| | Date | Topic | HW due | HW out |
|---|---|---|---|---|
| W | 9/7 | L1: Introduction and Overview | | 1 |
| M | 9/12 | L2: Summarizing data – one variable | | |
| W | 9/14 | L3: Summarizing data – multiple variables | 1 | 2 |
| M | 9/19 | L4: Sampling distribution I | | |
| W | 9/21 | L5: Sampling distribution II | 2 | 3 |
| M | 9/26 | L6: Basic concepts of inference I | | |
| W | 9/28 | L7: Basic concepts of inference II | 3 | 4 |
| M | 10/3 | L8: Inferences for single samples I | | |
| W | 10/5 | L9: Inferences for single samples II | 4 | 5 |
| M | 10/10 | NO CLASS - FALL BREAK | | |
| W | 10/12 | L10: Inferences for two samples I | | |
| M | 10/17 | L11: Inferences for two samples II | | |
| W | 10/19 | L12: Simple regression I | 5 | 6 |
| M | 10/24 | L13: Review | | |
| W | 10/26 | L14: MIDTERM | | |
| M | 10/31 | L15: Simple regression II | | |
| W | 11/2 | L16: Simple regression III | 6 | 7 |
| M | 11/7 | L17: Simple regression IV | | |
| W | 11/9 | L18: Simple regression V | 7 | 8 |
| M | 11/14 | L19: Multiple regression I | | |
| W | 11/16 | L20: Multiple regression II | 8 | 9 |
| M | 11/21 | L21: Multiple regression III | | |
| W | 11/23 | L22: Multiple regression IV | | |
| M | 11/28 | L23: Multiple regression V | | |
| W | 11/30 | L24: Multiple regression VI | 9 | 10 |
| M | 12/5 | L25: Logistic and Log-Linear Regression | | |
| W | 12/7 | L26: Review | 10 | |