

Theory of Statistics

John W. Emerson
Yale University
Spring 2003

Practical Information

- Although my name is John, I would prefer that you call me Jay, Mr. Emerson, or Dr. Emerson. No extra credit points will be awarded either way.
- E-mail is the best way to reach me: john.emerson@yale.edu
- Office: I'm in KGL 124 (Klein Geology Laboratory up on Science Hill), phone 432-3155. Because I will teach weekly review sessions, I will not have formal office hours. I can be available, though, as necessary -- please email me with questions and we can go from there.
- Lectures: Monday, Wednesday, and Friday, 9:30-10:20 in Becton 102.
- Review Sessions: as announced, but usually at night in the classroom of the Department of Statistics, 24 Hillhouse Avenue (the pink house with green shutters across from the Health Center).
- An optional lab (an introduction to S-Plus) will be taught this Sunday, January 19, in the StatLab (140 Prospect Street). We will set the time during the first class.
- TF: Jianfeng Yu, 3rd year Ph.D. student in Statistics. Jianfeng will have weekly office hours on Monday, 3-4 PM on the 3rd floor of the Department of Statistics, 24 Hillhouse Avenue. His first scheduled office hour will be Monday, January 20, even though no classes will be held. His email address is jianfeng.yu@yale.edu
- StatLab Information and Hours: in particular, note that students may obtain a free copy of S-Plus. Information on other software packages and hours of operation are posted on the web site: <http://statlab.stat.yale.edu> Consultants are regularly available for help using the software packages.

Objectives and Prerequisites

The course assumes students have studied first-year college calculus and probability theory at the level of Stat 241/541. Stat 242/542 is really the continuation of a full-year course – the part that addresses using probability models for making inferences, predictions, and decisions about practical problems that we encounter in many disciplines. Upon completion of this course, students should be able to think critically about data, to use graphical and numerical summaries, to apply standard statistical inference procedures and to draw conclusions from such analyses. This core material should provide adequate preparation for advanced courses on topics such as experimental design, regression, sample surveys, and quality control. The course is useful for professionals in many areas: statistics, actuarial work, medical research, public health, genetics, environmental studies, operations research, business, and econometrics.

Outline of Topics

- Descriptive statistics and exploratory data analysis
- Producing data: sampling and experimental design
- Estimation (emphasis on the method of moments and maximum likelihood)
- Statistical inference: confidence intervals, hypothesis testing, one and two-sample procedures, assessing goodness of fit
- Analysis of variance
- Linear least squares
- Logistic regression
- Decision theory and Bayesian inference

Textbooks

The required textbook is by John Rice, *Mathematical Statistics and Data Analysis*, 2nd Edition (1995). It is available at Barnes & Noble (the Yale Bookstore). For a much lighter (and fun) reference, you might consider:

Gonick and Smith (1993), *The Cartoon Guide to Statistics*. (optional and cheap!)

Other books that I often refer to include (in no particular order and often older versions):

Mendenhall, Scheaffer, and Wackerly (1986), *Mathematical Statistics with Applications*, 3rd Edition

Moore and McCabe (1998), *Introduction to the Practice of Statistics*, 3rd Edition.

Casella and Berger (1990), *Statistical Inference*.

Ross (1988), *A First Course in Probability*, 3rd Edition.

Cochran (1977), *Sampling Techniques*, 3rd Edition.

Homework

Homework really ought to be assigned and collected daily – it is critical that you stay on top of the material. However, I will generally collect homework once a week – I will try to keep the class web site (see <http://classes.yale.edu>) completely updated with respect to the assigned homework and due dates. Problem sets will be collected at the beginning of class.

I believe it is often possible to learn more by working together than by working individually. This requires that each member of a team actively participates, however. I wish to explore the possibility of some homework assignments (or parts of homework assignments) being completed in small groups of 2-4 students. Students who elect to do this would be required to complete a slightly longer assignment, but only one assignment would be submitted for the group. Students might find this option particularly attractive when there is computational work required.

Exams

I am inclined to give two shorter mid-term exams and a longer final exam. My exams are closed-book; I do, however, allow students to prepare a single-page (front and back) “crib sheet” of notes, formulae, and other pearls of wisdom. Tables will be provided as necessary.

Computing

Computing facilities are available to students at the Statlab. I encourage any and all use of computers, for word-processing, data analysis, simulation, anything you want. However, piles of unedited output will not be accepted. You must carefully document and explain your work. You are welcome to use any software packages you are comfortable with. However, my expertise is with S-Plus.

Grades

I would like to be a little vague about grading at this point. I don't want any single item to count too much. I expect the problem sets to be worth approximately 30%, the two mid-terms approximately 20% each, and the final exam to be worth the remaining 30%.