

Preliminary Course Outline

Instructor: Rebecka Jornsten, rebecka@stat.rutgers.edu

Office hours: 451 Hill Center, 4.30-5.30

Class homepage: <http://www.stat.rutgers.edu/~rebecka/Stat586>

Week/Date	Topics
w1, Jan 23	Introduction: Data summaries, Visualization, Linear Models
w2, Jan 30	Linear Models: Gauss-Markov, Maximum likelihood, Diagnostics
w3, Feb 6	Linear Models: Diagnostics continued, Model Selection
w4, Feb 13	Cross validation, Bootstrap, Regularization and Shrinkage
w5, Feb 20	Robust statistics
w6, Feb 27	ANOVA, ANCOVA.
w7, Mar 5	Mixed Effects models.
w8, Mar 12	Review.
Mar 19	Springbreak
w9, March 26	Classification.
w10, April 2	GLM: Diagnostics. Poisson and Binomial models
w11, Apr 9	Bagging and Boosting.
w12, Apr 16	Clustering.
w13, Apr 23	Review.
w14, Apr 30	Student Presentations

Labs make up 20 % of the final grade.

The final makes up 40 %.

The project is worth 40 % of the final grade.

Texts: T. Hastie, R. Tibshirani & J. Friedman (2001) *The Elements of Statistical Learning*. Springer. (See also suggested books on class homepage).

Software (suggested): R. R is free - download for windows, linux available. R is installed on the stat computers. If you're already familiar with another statistical computing package like SAS, feel free to use it. I will write the labs for R, but the data sets we will work with will be available in tab-del files.

Project - Project Proposal

The project proposal is due April 2nd. However, the sooner you get a proposal to me, the sooner I can give you feedback on it. Don't wait until the last minute.

The proposal should contain the following;

- a) a brief description of the data you intend to analyse, or the paper you intend to review.
- b) an indication of the source of the data set, or full references of the paper.
- c) the objectives of your investigation,
- d) an overview of the analyses you *anticipate* completing, or goals of your review.

If you can't think of a project, come see me.

Lab reports

Lab reports should be typed! Unless the lab contains an analytical problem, do not hand in handwritten material.

The report should contain the following;

- a) Description of the methods used. Be brief - don't repeat what's in the text, just the key elements.
- b) Discuss your results. Results without discussion are not graded.
- c) Include only the crucial plots and graphs, don't go for quantity.
- d) Label all plots and graphs.
- e) Conclusions: what is the take-home message.
- f) Do not collaborate! You can discuss programming problems with your fellow classmates, but do all the work yourselves. I will deduct marks if I notice that you are collaborating.

For help with computing; Modern applied statistics with Splus (Venables and Ripley) is a good text. Phil Spector (<http://www.stat.berkeley.edu/users/spector>) has an on-line introduction to Splus, and Splus and R are very similar. For help with R, check out <http://cran.us.r-project.org/>. Other online tutorials are available at <http://www-personal.engin.umich.edu/~bftsplyk/STAT500/R-intro.pdf>, <http://fisher.stats.uwo.ca/faculty/murdoch/458/r.htm>