

Regression analysis Fall 2008

Preliminary Course Outline

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

Office hours: 451 Hill Center, 5.00-6.30

Class homepage: <http://www.stat.rutgers.edu/~rebecka/Stat563>

Week/Date	Topics
w1, Sep 4	Introduction, Basic Stats, Linear models
w2, Sep 11	Linear Models: Diagnostics and matrix formulation
w3, Sep 18	Multiple regression
w4, Sep 25	Multiple regression: diagnostics and testing
w5, Oct 2	ANOVA, ANCOVA
w6, Oct 9	Model selection I: testing, subset selection
w7, Oct 16	Model selection II: selection criteria
w8, Oct 23	Model selection III: bootstrap, cross-validation
w9, Oct 30	Regularized regression: PCreg, Ridge and Lasso. Bayesian methods.
w10, Nov 6	Nonlinear regression, Weighted LS
w11, Nov 13	Looking ahead: generalized linear models
w12, Nov 20	CART, non-parametric methods
w13, Nov 25 (OBS TUESDAY!!!)	Review
w14, Dec 4	Student Presentations
	FINAL DUE on Dec 15

Labs make up 20 % of the final grade.

The final makes up 40 %.

The project and in-class presentation is worth 40 % of the final grade.

Texts: Kutner, Nachtsheim and Neter, Applied linear regression, McGraw and Hill. W.N. Venables and B.D. Ripley, Modern applied statistics with R, Springer, 4th edition.

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 Oct 16. 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 analyze, 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.

Note, this is a regression class. Choose a data set that fits with the class syllabus. If you can't think of a project, come see me.

Some data sets will not be accepted: these include the "body fat data", and "the boston housing data". Try to pick something different, something that YOU are interested in.

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 and may take more serious disciplinary actions (Don't test me on this...).

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>