Preliminary Course Outline

Instructor: Rebecka Jörnsten, jornsten@chalmers.se

Class homepage: http://www.math.chalmers.se/Stat/Grundutb/GU/MSG500/A09/

Week	Topics	Chapters
w44	Introduction, Basic Stats, Linear models Diagnostics and matrix formulation	0,1,2:1-6
w45	Multiple regression. Diagnostics and testing	4, 5:1-4, 6:1-2, 8:1-3
w46	Dummy variables, ANCOVA. Model selection and testing.	14, 23. 9:1 + notes
w47	Model Selection	11, 15 + notes.
w48	Bootstrap. Cross-validation.	26 + notes.
w49	Regularized regression	16:1-4, 16 + notes.
w50	Weighted least squares, non-linear model, GLM In-class presentations	9:2-3, 9:5, 18
w51	Final Exam Dec 17	

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.

Text: N. R. Draper and H. Smith, Applied Regression Analysis, Wiley series in probability and statistics.

Software: R. R is free - download for windows, linux and mac available. I will write the labs for R, but the data sets we will work with will be available in tab-del files so you can use another software package if you want.

Linear Statistical Models   **Fall 2009**

Project - Project Proposal

The project proposal is due Nov 24. 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.

   b) an indication of the source of the data set.

   c) the objectives of your investigation.

   d) an overview of the analyses you *anticipate* completing.

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. There are data set repositories on the web (see e.g. `http://lib.stat.cmu.edu/`)
Some data sets will not be accepted: these include the "body fat data", and "the boston housing data".

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.

Other

online tutorials are available at `http://cran.r-project.org/doc/manuals/R-intro.html`, `http://heather.cs.ucdavis.edu/ matloff/R/RProg.pdf`