

## Preliminary Course Outline

**Instructor:** Rebecka Jörnsten, [jornsten@chalmers.se](mailto:jornsten@chalmers.se)

Class homepage: <http://www.math.chalmers.se/Stat/Grundutb/GU/MSG500/A12/>

Week	Topics	Chapters
w44	Introduction, Basic Stats, Linear models Diagnostics and matrix formulation	1:1-7,9. 11, 12:1-4
w45	Multiple regression. Diagnostics and testing	2, 3, 4, (6), 11, 12, 13.1
w46	Dummy variables, ANCOVA. Model selection and testing.	9 + notes
w47	Model Selection	7, 13 + notes.
w48	Bootstrap. Cross-validation.	7 + notes.
w49	Regularized regression	13 + notes.
w50	Weighted least squares, non-linear model, GLM In-class presentations	8, 10, 12, 15
w51	Final Exam	

Labs make up 10 % of the final grade.

The final makes up 60 %.

The project and in-class presentation is worth 30 % of the final grade.

Text: J.O. Rawlings, S.G. Pantula, D.A. Dickey. Applied Regression Analysis (available online).

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. However, I strongly urge you to use R since I will be providing demo codes from the lectures using this language.

## Project - Project Proposal

The project proposal is due Nov 20. 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. There are data set repositories on the web (see e.g. <http://lib.stat.cmu.edu/>, <http://www.statsci.org/datasets.html>). 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) You can discuss programming problems with your fellow classmates, but do all the work yourselves.

For help with computing; Modern applied statistics with Splus (Venables and Ripley) is a good text.

Some

online tutorials are available at <http://www.uga.edu/strata/software/pdf/RTutorial.pdf>, <http://cran.r-project.org/doc/manuals/R-intro.html> or google for more....