MSG500/MVE190, Linear Statistical Models LP2 2018

Preliminary Course Outline

Instructor: Umberto Picchini, picchini@chalmers.se
Course homepage: http://www.math.chalmers.se/Stat/Grundutb/GU/MSG500/A18/

Send me an empty email with subject header "MSG500/MVE190" in the first week of classes, to be put on the mail-list (used for class-updates).

Plan: the approximate content and schedule is given in the table below. This is NOT strict and subject to change.

Week	Topics	Chapters	Exercises
w45	Introduction, Basic Stats, Linear models Diagnostics and matrix formulation	1:1-7,9. 11, 12:1-4	MiniAnalysis1
w46	Multiple regression. Diagnostics and testing	2, 3, 4, (6), 11, 12, 13.1	Mini2
w47	Dummy variables, ANCOVA. Model selection and testing.	9 + notes	Mini3
w48	Model Selection	7, 13 + notes.	Mini4
w49	Bootstrap. Cross-validation.	7 + notes.	Mini5
w50	Regularized regression	13 + notes.	Mini6
w2	Bayesian linear regression In-class presentations		

MiniAnalysis tasks make up 10 % of the final grade and it is mandatory to present at least once! *However I might have to re-evaluate this criterion if this year there are many more students. I will inform you in due time about this.*
You will present MiniAnalysis tasks in class on Fridays (I will usually hand the tasks out the week before).
On MiniAnalyses you can work in teams of 2 but not the same team for more than two tasks.

The in-class final (January 17th) makes up 50 %.
The project and in-class presentation is worth 40 % of the final grade (to be handed in the week of the final exam).

Literature:

- Lecture notes: PDF of lecture notes by Rebecka Jörnsten will be posted at the course webpage.
- Book: J.O. Rawlings, S.G. Pantula, D.A. Dickey. Applied Regression Analysis available online. I won't be following this book for my lectures, but you can use it as additional support if needed.

Software: R and Rstudio. Both are free - download for Windows, Linux and Mac available. See (this page) for instructions.
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.
Here's a good place to start: https://www.r-bloggers.com/how-to-learn-r-2/

Notice: this is not a course about R! This is a course in statistics. We will discuss *some* of the commands needed to produce the desired output and answer the relevant statistical questions. For everything else, Google is your friend! Also, we will not consider tips-and-tricks, good programming practice or any advanced use of such powerful computer language. R has a large and friendly user community and you will be able to find plenty of good guides, tutorials and answered questions by a simple Google search.

Linear Statistical Models **LP2 2018**

Project

Create groups of two students, and not more than two.
The project will be based on a dataset I will give you. The dataset will be distributed later during the course, as it does not make sense to start experimenting too early, that is not before you have acquired enough skills and studied enough tools.
You will be given a specified goal. But you will have to be creative on how to achieve this goal, by using methods illustrated during the course in a critical way. That is it will not be about answering a sequence of questions I pose.

Your projects will be run through URKUND to check for plagiarism.

Project report

Work with your project partner. The project report should be typed! Do not hand in hand-written material. Your project report will be handed in the week of your final exam.
The report should **be written in English** and contain the following:

a) Name and surname of both students.

b) Number the pages.

c) Description of the methods used. Be brief - don't repeat what's in the text, just the key elements.

d) Discuss your results. Results without discussion are not graded.

c) Include only the crucial plots and graphs, don't go for quantity.

e) key information may be better summarized in tables than by including the R printouts (e.g. it may be enough to give regression coefficients and p-values without all the accompanying information provided by R).

f) Label all plots and graphs. Reference to those in the text so the report is understandable and readable.

g) Conclusions: what is the take-home message.

h) 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
https://www.r-bloggers.com/how-to-learn-r-2/

http://www.uga.edu/strata/software/pdf/RTutorial.pdf
http://cran.r-project.org/doc/manuals/R-intro.html or google for more....

I use RStudio - highly recommended for ease of use BUT has a tendency to crash from time to time.

MiniAnalysis

MiniAnalysis is like a lab where I provide most of the code (or ask you to adapt demo codes) and the main purpose of the task is for you to apply the methods and discuss and interpret the results.

MiniAnalysis tasks will be presented in class. I will choose teams at random to present their results. **It is mandatory to prepare an in-class presentation (5-6 slides to be sent to me ahead of class or brought on a USB stick or available online through Box/Dropbox etc)**.
If you don't attend class when MiniAnalysis tasks are presented you will be required to write a report (which is a lot more work!).

It is mandatory to present at least once. ⇒ *However I might have to re-evaluate this criterion if this year there are many more students. I will inform you in due time about this.*

a) Prepare a few, say 5-6 slides.

b) Discuss your approach and results.

c) Include only the crucial plots and graphs, don't go for quantity.

d) Conclusions: what is the take-home message.

Don't think of MiniAnalysis as an exam or a test. The idea is for you and the rest of the class to focus on results and interpretation and to get a discussion going.