MSG500/MVE190, Linear Statistical Models LP2 2019

Preliminary Course Outline

Instructor: Umberto Picchini, picchini@chalmers.se
Course homepage: https://chalmers.instructure.com/courses/7490

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.	
w50	Regularized regression	13 + notes.	
w51	Bayesian linear regression		

MiniAnalysis tasks make up 10 % of the final grade and **it is mandatory to attend**. Each Friday several randomly selected groups of students will present their MiniAnalysis results in front of the the class. Attendance will be recorded and those that do not attend will have to write a detailed report instead (which is a lot more work!). 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.

The project report (on different data than the ones considered for minianalyses) is worth 40 % of the final grade.
The in-class final exam makes up 50 %.

Literature:

- Lecture notes: PDF of lecture notes by Rebecka Jörnsten will be posted at the course webpage.
- No need to purchase any book, as I will closely follow the lecture notes. However here are some suggested resources that you may want to consider as additional support:
  1. J.O. Rawlings, S.G. Pantula, D.A. Dickey. Applied Regression Analysis available online.
  2. Alan Agresti. "Categorical Data Analysis", Wiley. available online.

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

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.

Submit you projects via Canvas as group work.

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

Project report (submitted via Canvas)

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.

Linear Statistical Models

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. Every time, I will choose several teams at random to present their results. You must work in a team, which I recommend to be of size two (as in the end the project will be written by 2 persons. If you are 3 persons doing the minis, it will be awkward for the third person having to find a project partner all of a sudden). **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!).

a) Prepare a few slides to illustrate your work in about 5-7 minutes.

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.
In class presentations are not graded.