Syllabus for 401-3626-00L
Multivariate Statistics
Fall semester 2008

1 General information

Instructor:	Marloes Maathuis, LEO D 9.2, maathuis@stat.math.ethz.ch
Lectures:	Thursday 1-3pm, HG D 7.1 (**starting September 25th)**
Class website:	http://stat.ethz.ch/∼maathuis/teaching/fall08

2 Course description

The key element of multivariate statistics is that we analyze several variables simultaneously, without saying beforehand that one of them is of more interest than the other ones. This course introduces the basic concepts of multivariate statistics, and provides an overview of the available methods. All methods will be illustrated via real data sets, using the open source statistical software R (http://cran.r-project.org/). At the end of the course, you should be able to:

- Identify appropriate multivariate methods for a given research question
- Apply multivariate methods using R, and interpret their output correctly
- Interpret the results of multivariate statistical analyses done by others

3 Course format

There are 13 lectures (the first lecture is canceled) of 2 hours each. The lectures will be given in English. I will use the lectures to explain the statistical methods, and also to show how to apply them to solve real research problems.

4 Course material

There will be no official lecture notes for this course, so I encourage you to take careful notes. The class website is another important resource for the course, especially the part containing the class schedule. I will fill in this schedule during the semester: for each lecture, I will indicate suggesting readings, post the R-code that was used to do the examples, post relevant links, etc. The lectures will be mostly based on the following books:

- Brian F.J. Manly (1986, reprinted later), "Multivariate Statistical Methods - A Primer", Chapman and Hall (thin book, good to get an overview of the methods and an idea of the applications)
- K.V. Mardia, J.T. Kent and J.M. Bibby (2003 or 2008), "Multivariate Analysis", Academic Press, Amsterdam (more detailed and mathematical, good for really understanding the methods)

- Brian Everitt (2005), "R and S-plus companion to multivariate analysis". Available for free in electronic form via the ETH library, see http://www.springerlink.com/content/978-1-85233-882-4

5 Exercises

I will hand out a few exercise sets during the semester. I strongly encourage you to do these exercises, but it is not compulsory.

6 Exam

Everybody who needs ECTS credit points for this course needs to do an oral exam (20 minutes). Ph.D. students who only need a "Testat" should solve half of the exercises or give a short seminar or write a short report on the analysis of a data set or pass a semester end oral exam. The exercises or the report should be handed in before December 19.

The oral exam will take place in the regular exam period. I will schedule the exams at the very beginning of the exam period. If possible, I will conduct the exams myself. If that is not possible, another statistics professor will replace me. In that case, I will make sure to discuss the exam material with him/her.

The exam will cover all material we discussed in class. During the exam, you do not need to produce R-code yourself, but you should be able to understand R-code and interpret output from R. There will be questions on theoretical aspects and practical aspects. One of the questions on the exam will be directly taken from the exercise sets.

7 Tentative schedule

- Lecture 1: Basic statistical ideas:
  - Research design
  - Statistical models and inference
- Lecture 2: Introduction to multivariate statistics:
  - Example data sets
  - Graphical methods for multivariate data
  - Overview of methods for multivariate statistics
- Lectures 3,4,5,6: Data reduction techniques:
  - Principal component analysis
  - Factor analysis
- Lectures 7,8,9,10: Multivariate analysis of variance and multivariate regression
- Lectures 11,12,13: Classification techniques:
  - Discriminant analysis
  - Cluster analysis