Multivariate Statistical Analysis

STA 345 **SYLLABUS AND COURSE POLICIES** Spring, 2006

Instructor: Woncheol Jang
Old Chem 223B
wjang@stat.duke.edu
(919) 684-3437
Office Hours: By appointment

Course Web Page: `http://www.isds.duke.edu/courses/Spring06/sta345/`
Lectures: WF: 10:05 AM - 11:20 AM, Old Chem 116M

Textbooks

- Johnson, R.A. and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis, 5th ed.* Prentice Hall. **(Required)**.
- Everitt, B. (2005). *An R and S-plus companion to multivariate analysis.* Springer. **(Recommended)**.

Prerequisites

I assume you have taken STA 213 *Introduction to Statistical Methods* and STA 244 *Linear Models*. In other words, you should know linear algebra and statistical principles at the level of Casella and Berger (2002) or Wasserman (2004).

Computing

The examples and assignments will involve some computing. It will useful to learn one of the following programming language: R (recommended), Splus, SAS and MATLAB.

Homework, Final Project and Grading

There will be biweekly homework assignments. The assignments are due in class. Homework will be a mix of data analysis part and theoretical materials. Data sets will be provided as needed. You may work with other students on these problems or refer to other sources if you would like. However, the writeup of your homework should be your own. Each assignment will have equal wight on your grade.

The final project consists of a data analysis of a real data set using the techniques covered in class. You should choose a data set that is of interest to you. You are encouraged to find data yourself for the project or use data from your own research experiences. Thinking about this as soon as possible.

The grade will be determined based on the homework assignments (70%) and the final project (30%).

References

- Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis, 3rd ed.* Wiley.
- Hastie, T., Tibshirani, R. and Freedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Perdition.* Springer.
- Härdle, W. and Simar, L. (2003). *Applied Multivariate Statistical Analysis.* Springer.
- Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). *Multivariate Analysis.* Academic Press.

Course outline

This course covers methods for analyzing continuous multivariate data. Broadly, we discuss

1. Modeling and inference using the multivariate normal distribution
    - Multivariate data and models
    - Multivariate Normal distribution
    - Traditional inference: Multivariate Regression, MANOVA, etc
    - Links with mixed linear models and hierarchical modeling.
2. Exploratory techniques based eigenvalue and singular decomposition
    - SVD of a data matrix; special decomposition
    - Principle Component Analysis
    - Factor Analysis
    - Canonical Correlation
3. Classification and Clustering
    - Linear Discrimination
    - Classification Trees
    - Hierarchical Clustering
    - K-means Clustering
    - Multidimensional Scaling
4. Functional data analysis (if time permits)
    - Functional PCA
    - Functional Classification
    - Functional Clustering