

**DSc 8450 Applications of Statistical Models  
for Decision Making  
Syllabus**

**Brian Schott / Spring '01**

<http://www.gsu.edu/~dscbms/dsc8450/>

Course Objectives:

Upon completion of the course the student should be able to:

1. Diagnose problems using exploratory data analysis procedures.
2. Select and be able to apply a variety of statistical models.
3. Use methodologies to determine and evaluate the attributes that decision makers use to select from among alternatives.
4. Apply computer software to the analysis of business problems.

Textbooks:

- a) Joseph F. Hair, Jr., Rolph E. Anderson, Ronald L. Tatham, & William C. Black, **Multivariate Data Analysis with Readings**, 1998, 5th edition, Prentice Hall.<sup>1</sup>
- b) A coursepack **Appl Stat Models/Dec Mkg** contains relevant SAS .log and .lst files (and other class materials) is available for sale at The Printshop at 6 Decatur Street **only**.

We will be using the Microsoft Windows version of SAS version 8. SAS is available in the student computer labs, or you may choose to purchase a copy that will not work after July 1, for use off campus. There is voluminous help available to you while you use SAS. You will find additional information at the DSc8450 web homepage which I have organized for this class.

Methods of Instruction:

DSc8450 will be a lecture course with a few assigned computer-related problem sets that will be used as a basis for class discussion. Also, students will design and carry out a project requiring data analysis and interpretation of a self-defined problem.

Alone or in pairs students can work on problem sets and on the class project. The project will include a research statement and a statistical analysis of data -- in most cases the data values are from existing databases. It is not usually practical for students to collect (primary) data themselves during the school term. Many databases are stored and available on the internet. Links to possible data sets are suggested in the "Web Sites regarding Statistical Data" web link at the class's home page.

Each of the Hair *et al* chapter examples have been coded as SAS command files on our Unix computer and we discuss these examples in class. The daily class schedule provides the file names containing each SAS command file available in my files directory. A coursepack containing the relevant SAS .log and .lst files (and other class materials) is available for sale at The Printshop at 6

---

<sup>1</sup>The course web homepage contains a list of errata for this textbook.

Decatur Street. Review these SAS files before designated classes so that we can better use class time.

The class will use the statistical analysis program SAS. SAS is one of the most respected programs of its type and affords very diversified and tested analysis programs. A key feature of the SAS program is that each analysis produces two files; a .log file which annotates the analysis performed all and a .lst file which reports the actual analysis results. All student projects must include written or emailed versions of both the .log and the .lst files. Please note that I do NOT wish to have WORD or other non-text files emailed to me. Please email only text files.

My goal is to help each of you to better learn, understand and use effectively powerful multivariate statistical techniques. This goal is challenging because of the number of topics and their complexity and interdependency. With the availability of excellent texts for the course and ample computer resources at GSU, I hope to support your achieving the course objectives by communicating with you frequently. Please communicate with me promptly when you run into a problem; your questions will not be considered in your grade. My preferred communication method is email, but I am available regularly during office hours for face to face consultation and am very available at the office via telephone. I also am available at home by phone, but only for more urgent situations, please.

#### Grading:

Problem Sets	20% <sup>2</sup>
Term Project (stages)	
1. Research problem & objectives	5% <sup>2</sup>
2. Analysis plan	5% <sup>2</sup>
3. Evaluate assumptions	5% <sup>2</sup>
4. Estimate model & assess fit	10% <sup>2</sup>
5&6. Interpret variate(s) & validate model	15% <sup>2</sup>
Completed oral and written project	15% <sup>2</sup>
Final Exam	25% <sup>3</sup>

#### Miscellaneous:

Academic misconduct such as cheating and plagiarism will result in a failing course grade.

The course syllabus and guidelines are approximate and may be subject to change.

---

<sup>2</sup>The course schedule indicates a due date for each problem set and project update. To receive full credit each must be submitted at the beginning of class on its due date. A 10% deduction will be applied for each class period that it is late.

<sup>3</sup> 33 single sided pages of handwritten notes can be brought to the exam.

Course content:

- I. Computer systems
  - A. Unix
    - 1. communication
    - 2. file management
    - 3. text editing
  - B. SAS
    - 1. data input
    - 2. program commands to label, manipulate, and create data
    - 3. executing statistical analyses
- II. Hair et al's "structured approach to multivariate model building"
  - A. Define the research problem, objectives, and multivariate technique to be used
  - B. Develop the analysis plan
  - C. Evaluate the assumptions underlying the multivariate technique
  - D. Estimate the multivariate model and assess overall model fit
  - E. Interpret the variate(s)
  - F. Validate the multivariate model
- III. Statistics (CAPITALIZED terms refer to SAS procedures)
  - A. General methods and examining data
    - 1. describing a variable and its data values
      - a. type
      - b. location
      - c. spread
      - d. shape, eg, normality, skewness (UNIVARIATE)
    - 2. statistical reliability and validity are important and different
    - 3. relationships between variables (COR, PLOT)
      - a. measures of association
      - b. (curvi)linearity
    - 4. hypothesis testing and p-values
    - 5. effect and power analysis
    - 6. data transformations
      - a. to achieve linearity
      - b. to achieve normality
      - c. to remove unwanted association
      - d. to include meaningful variables
    - 7. a statistical variate is a linear combination of multiple variables computed in multivariate analysis
    - 8. selecting a multivariate technique
  - B. Specific analyses
    - 1. Analyzing Relationships for Prediction
      - Regression (REG, SCORE)
    - 2. Analyzing Relationships for Classification
      - a. Discriminant (DISCRIM, CANDISC)
      - b. Cluster (CLUSTER)
    - 3. Examining Interrelationships
      - Principal components & common factor (PRINCOMP, FACTOR)

## Term Project:

On pages 25-27 in their textbook Hair et al suggest six clear stages for performing multivariable projects. This is one of the stronger features of our textbook and these steps have been incorporated into the course structure; separate update deadlines which correspond almost completely to the authors' six stages are given for each step in the daily schedule. These six stages include both conceptual and empirical issues listed on the course homepage (as originally shown in abbreviated form on the inside front cover pages of the text).

Each chapter contains an illustrative study based on the fictional firm HATCO and the database described on pages 27-29. Familiarize yourself with this database before looking at the later chapters. These illustrative studies provide project objectives corresponding to the four statistical analysis methods covered in this course (regression p 195, discriminant p 281, cluster p 502, and factor analysis p 120).

In the first stage, state your project objective with a particular individual or organizational user in mind. Consider the user's (imagined) application of the results in each stage but especially in stages 1 and 5. Write up the first stage as if you have developed a clear rationale for the need for the study you are proposing and that you can imagine the critical conceptual dimensions that define the statistical relations you will study. However, be advised that you in fact need a pretty clear idea that actual data are available for you to perform a multivariate statistical analysis using one of the four methodologies we explore during the school term. As a rule this means finding an existing data set which contains several concrete metric variables observed on a sample of at least 100 sampling units (people or companies or geographical regions, for example). Students have even added a key variable to an existing data set, but this is unusual. I recommend against studying time ordered data sets in this course.

Prior to stage three you will discuss plans and strategies for accomplishing these plans. Also, it is advisable to submit a description of the data base you will be using (and an example of a typical multivariate observation) so we can be confident that multivariate analysis is appropriate and achievable.

Beginning with stage three you will be expected to supply a concrete SAS analysis with your updates. This does not mean that by itself a SAS printed listing is sufficient. Rather a statistical clarification should accompany each submitted update.

The completed project will be assessed on its cohesiveness and on your ability to revise the earlier parts as I have recommended. Along with your final project writeup and the final SAS files, *submit your updates with my comments attached.*

If your project plan must be changed during the school term, keep me informed. There is no grade penalty for changing your project.