STA 4702/5701  Intro to Multivariate Analysis  Spring 2007

Time: MWF 10:40-11:30am

Room: Griffin Floyd 100

Instructor: Machelle Wilson  
McCarty Hall C 430  
392-2545  
sta5701@gmail.com

Please send an email to the above address so that you are in the address book. Please also be sure that emails from this account are not allocated to your junk mail folder.

Office Hours: After class on Mondays/Wednesdays or by appointment

TAs: Herman Hooker  
Griffin Floyd 218  
392-1941 ext 231  
hhooker@stat.ufl.edu

Yingxia Wan  
McCarty Hall C 401  
392-1947 ext. 228  
yxwan@stat.ufl.edu

Office Hours: Tues/Thurs 9:30-10:30  
Tues/Thurs 4-5pm

Textbook: Multivariate Data Analysis by Hair, Anderson, Tatham & Black. You are expected to have read the material being covered in lecture before you come to class. This will greatly facilitate your learning the material as well as your enjoyment of lecture.

Course Webpage:  
http://www.stat.ufl.edu/STA%204702/Spring%2007/index.html

Grading:  
Homework: 50%  
Midterm Exam: 20%  
Final: 30%

Homework will be posted on the web and will be due in class on Fridays. There will be 5 homework assignments. Homework is not only a fairly substantial portion of your grade, but is vital to your success in this class. Working with other students on homework is allowed and even encouraged. But don't turn in anything you can't explain fully to me. Solutions to all problems will be provided.

Homework will consist of some problems from the end of each chapter, some simple math problems, and a data analysis. The data analysis portion of the homework will need to be written up like a report/journal article. This means that it should include:

Introduction: A brief description of the question of interest and why it’s interesting.

Methods: What statistical analysis methods (including exploratory and model validation techniques) you used and in what package. If the data set provided describes how the data were gathered, this should be briefly stated as well.

Results and Conclusion: This section will include statistical results, graphs, tables, and a summary paragraph which should include a statement and interpretation of the results in the context of the problem.
This is not just a statement of the statistical results such as significance, p-values, or anything using statistical jargon, rather it is what it means in a broader, real-life context.

As much as 15% of your homework grade will depend on how well you have presented the data analysis, so please do not ignore this. The remaining portion of your grade will be determined by accuracy and correctness. When solving mathematical problems, please be sure to show your work in complete and logical mathematical sentences so that your reasoning is obvious to the grader. If it doesn’t make sense, you will not receive full credit even if the number at the end is correct.

The first homework is due in class on Friday, January 26th.

Exams will be cumulative and closed book, except for 1 sheet (front and back) of notes you will make for yourself. All exams will be allowed only 1 sheet, front and back. If any probability tables or other standard references are needed for the exam, these will be provided for you and will be in addition to the page of notes you are allowed. You are also allowed a calculator, but be forewarned: if you don’t understand the material a calculator will not help you.

I usually provide a practice exam for you to use as a study guide. I do not provide solutions to the practice exams. If you have questions, come to office hours and we can work through the problems together.

No make-up exams will be given without a documented reason for missing the exam.

Midterm Exam: Friday, March 2nd.

Final exam: Thursday, May 3rd 7:30am-9:30am.

Course Material and Objectives: This course is intended as an introduction to multivariate statistical analysis techniques. Mastering a computer package is absolutely essential. We will be learning SAS. The university’s site license with SAS has recently changed. SAS will no longer be installed on classroom or lab computers. There is a unix-based version available on the server GROVE. If you want to use this version of SAS, talk to me and we will get you a log-in account. The best option for most of you will be to purchase your own student version of SAS. To purchase a student version of SAS go to http://www.software.ufl.edu/sas/

Because this course is offered both as an undergraduate course in Statistics and as a graduate course for non-majors, it is virtually impossible to make everyone happy with respect to the mathematical content. A deep understanding of multivariate techniques requires a thorough knowledge of matrix algebra. If I teach the math at a level appropriate for the stats majors, the non-majors will be lost. If I teach the math at a level appropriate for non-stats graduate students, the majors will be bored. To solve this problem, I’ve decided to dispense with the matrix algebra almost entirely and focus on computer implementation. An understanding of basic distribution theory (normal and binomial) and basic algebra is expected. I will present the underlying matrix algebra concepts graphically. This may prove to be unsatisfactory for everyone. Often without at least a basic understanding of the math, one can have a sort of ‘black box’ feeling of working in the dark. Please be patient. Everyone will learn many interesting and useful data analysis tools from the course and this is the real point of it all anyway. I will post notes to the web that cover introductory linear algebra for the non-majors that would like a deeper understanding. I will also post papers from the primary literature for those with a solid math background who are dying to get a sense the underlying mathematics in multivariate analysis.

Additional Course Information: The last day to add, or drop this course with a fee refund is January 12th. The last day to drop this course for a 25% refund is Feb 2nd. The last day to drop this course to receive a ‘W’ grade is April 13th.
Questions are strongly encouraged, both during class and at office hours. If you are lost or confused, please let me know.
Preliminary List of Topics to be Covered, as time allows:

1. Exploratory Analysis of Multivariate Data
2. Multiple Imputation of Missing Data
3. Multiple Regression
4. Factor Analysis
5. Multiple Discriminant Analysis
6. Principal Components and Partial Least Squares
7. Cluster Analysis
8. Multidimensional Scaling
9. Structural Equation Modeling
10. Other topics according to interest and time available