

# Data Analysis III: Multivariate Techniques Psychology 613

Spring 2005

Lecture: Tue, Thu 10:00-11:20, 303 Deady

Lab: Fri 10:00-11:50, 303 Deady

**Instructor:** Bertram Malle

Straub Hall 305. Phone: 346-0475.

Email: [bfmalle@uoregon.edu](mailto:bfmalle@uoregon.edu)

Office hours: Fri 1-3pm & by apptmt.

**Teaching Assistants:** Eric Olofson (.42) & Nathan Dieckmann (.42)

Straub 393 & 329. Phones: 346-1984 & 346-4963

Emails: [eolofson@uoregon.edu](mailto:eolofson@uoregon.edu) & [ndieckma@uoregon.edu](mailto:ndieckma@uoregon.edu)

Office hours: Mon 3-5 (EO) & Tue 12:30-1:30, Wed 1:00-2:00 (ND)

**Overview.** This course covers the basic multivariate techniques that are currently used in psychology and other social and life sciences. This includes principal components analysis, factor analysis, multivariate analysis of variance, discriminant function analysis, multiple and logistic regression, and log-linear analysis. Other techniques, such as structural equation modeling, hierarchical linear modeling, canonical correlation, or multidimensional scaling, are not covered in detail but are briefly introduced at the end of the term.

The learning goal for students is to have a conceptual understanding of each statistical technique, be able to apply the correct technique to any given set of data, properly interpret the output of statistical computer packages (primarily SPSS), and understand and critique scientific papers that use these techniques.

Conceptual understanding implies knowledge of the "functional character" of each technique -- what it is used for and how it works. Each technique is a tool that can immensely assist you in understanding your data and testing your hypotheses; but to apply these tools correctly you need to understand both your data and your tools.

The level of understanding you will reach in this course lies one step deeper than the level at which you will operate when using the techniques in years to come. You will learn to think about some mathematical underpinnings of the techniques, and this step will give you a deeper understanding of the workings of each technique. Even if you later forget the details of these underpinnings, you will not forget the conceptual implications they have. All mathematics introduced in this course is meant to serve your conceptual understanding.

Because most or all techniques will be new to you, we need to create maximal familiarity through redundancy. You are expected to read the relevant textbook chapters before each lecture. The lectures themselves will introduce the material from a conceptual and statistical perspective, with directions to computer analysis. The lab on Friday will briefly review the lectures and then introduce relevant SPSS commands to analyze actual data using the statistical techniques. The lab will also discuss the current homework and review past week's homework. In the homeworks, you will run analyses on provided data sets and write up the results just like you would for an empirical journal article. Finally, the exams will ask you to discuss important conceptual principles.

**Grading.** Your grade will be based on 9 homeworks (max. 100 pts each) and two take-home exams (300 points each), for a total of 1500 points. Tentative grade cut-offs are 1350 (A) and 1200 (B). The homeworks consist of data retrieval, data analysis, interpretation and write-up of results, all done electronically. The exams (web-posted 4/26 and 6/2) are more conceptual in nature; they consist of five to six questions on the abovementioned "functional character" of techniques and your understanding of central concepts.

**Readings:** Because there is no one all-around good and complete textbook of multivariate statistics, I have selected readings from various sources for this class. This is not an optimal solution because the various sources differ in notation and emphasis. However, no currently available textbook meets all the needs of this course. In addition to selected readings, lecture handouts are extensive and are posted on the course web page (<http://darkwing.uoregon.edu/~bfmalle/613.html>), sometimes complemented by references to additional web-based resources.

For those of you who would like to own a source book, I recommend [Tabachnick and Fidell \(2001, 4th ed.\)](#), which is also the optional textbook for this course. In addition to required chapter readings each week (available for individual copying with Lori Olsen), I typically designate some optional chapters, either from T & F or from other textbooks. You are not required to buy T & F. If you do buy it, you must still copy and read the required chapters, because for some topics T & F just don't provide the background you need.

**Educational Technology:** This course makes heavy use of electronic media through

- a) a World Wide Web page for announcements, resource material, extensive lecture handouts, homeworks, etc.
- b) Email to communicate, answer questions, and to turn in homeworks and exams;
- c) UNIX computers (darkwing) or desktop computers to run SPSS, which is the statistical package used in this course.

The requirements below must be met **by the end of the first week**. See the [Resources page](#), Computer Issues. For further questions, talk to me, your TA, or use the Computing Center as a resource.

1. Identify accessible computers (Mac or PC): at home, in the public computer labs on campus, or in the Psychology computer lab in Straub Hall. Computers must have network access--to read and write email, to run data analyses on the mainframe, and to perform file transfers.

2. If you don't already have one, get an account on darkwing, the preferred computing environment for this course. You can do all your work on a PC/Mac with resident SPSS software, but I want everybody to be in principle capable of using the mainframe system. Even if you use personal computer software, you will have to use the "syntax approach" to fully understand and execute the analyses we perform in this course.

3. Familiarize yourself with an Email software program, such as Eudora or Mail.app (Outlook Express is notoriously vulnerable to viruses): Make sure you attach files in the appropriate format that your TA can read.

4. Familiarize yourself with course resources available at <http://darkwing.uoregon.edu/~bfmalle/613/resources.html> and on the [Blackboard course page](#)

5. Learn to use at least the rudimentary operations of the text editor gnu emacs (on darkwing). You can learn emacs by typing emacs at your darkwing prompt and typing

ESC x help-with-tutorial

(type ESC first, then x, then the connected words).

*By Friday of the first week (4/1),* send an electronic "progress report" via email to **bfmalle@uoregon.edu**, covering each of the five points above. That way I can be confident that you are set up with the electronic tools needed in this course.

**Homework:** Each Thursday, a homework assignment will be posted on the web and is due the following Wednesday. Analyses relevant to the assignment will be practiced during lab hours on Friday (10-11:50 in Straub 180). Typically the assignment contains a data set and several instructions to analyze it. Your homework is complete if you run all required analyses, edit the output files down to the essential information, and annotate the output to demonstrate your understanding of what SPSS was doing. Finally, you write up a one-page summary of the results as you would for an empirical journal article. The annotated outputs and the written results must be sent to your TA via email by Tuesday, 5 p.m. For the written results, take Tabachnick & Fidell's result summaries as models and be sure to write clearly, concisely, and meaningfully.

The Tuesday deadline was chosen because I want to make sure you have time to read the material for the Thursday class and because the TAs need time to grade the exams and give you feedback.

## Topics and Readings

I do not require a textbook, and there is none currently on the market that I wholeheartedly recommend, but if you want to buy one, Tabachnick & Fidell (2001) will be the best resource for your statistics needs. For the readings in this class, I have selected chapters from various statistics textbooks, SPSS handbooks, and a few journal articles. In addition, I have listed the relevant Tabachnick & Fidell (2001) chapters where available. In some cases, a Tabachnick & Fidell chapter is required reading.

All required reading materials are available with the graduate program secretary (Lori Olsen, Straub 129) for checking out and copying. For psychology students, another set is available for copying in Straub 347. There are four SPSS resource books available with Lori Olsen, and I have

basic SPSS handbooks for checking out as well. Finally, the computing center documents room has a good set of SPSS books.

### **Textbook authors cited below:**

Cliff, N. (1987). *Analyzing multivariate data*. Harcourt Brace Jovanovich: San Diego.  
[not comprehensive, thorough explanations, thin on software guidance]

Dillon, W. R., & Goldstein, M. (1984). *Multivariate Analysis: Methods and Applications*. Wiley: New York. [comprehensive, very thorough but mostly lucid mathematics, very little software guidance]

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). New York: HarperCollins. [comprehensive, thin on mathematics, rich software guidance]  
References with the pdf marker at the end are available as pdf files on the Blackboard course page.

### **L 1: Tu 3/29 The multivariate approach: Introduction and overview.**

» Dillon & Goldstein (1984). Overview of multivariate techniques (pp. 19-22) pdf

» Tabachnick & Fidell (2001). ch. 1 + 2 pdf

[optional: SPSS command reference list; command syntax handouts; UNIX command list; gnu emacs command list]

### **L 2: Th 3/31 Data screening and exploratory data analysis**

» Tabachnick & Fidell (2001). ch. 4 pdf

» Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45, 1304-1312. pdf

### **Fri 4/1 Lab 1: Exploratory data analysis**

### **L 3: Tu 4/5 Basic matrix algebra**

» Cliff, N. (1987). Elements of matrix algebra for statistical applications (ch1.), Vectors (ch. 3).  
[very basic, slow pace, for beginners]

or

» Dillon & Goldstein (1984). Vector and matrix operations and selected statistical concepts (pp. 521-539). [faster pace, for intermediates]

[optional: Tabachnick & Fidell (2001). Appendix A]

### **L 4: Th 4/7 Matrix algebra for statistics**

» Cliff, N. (1987). Statistical formulas in matrix form (ch. 2), Variances and covariances of linear combinations (ch. 4), The inverse (ch. 5). [slower pace, very thorough]

### **OR**

» Dillon & Goldstein (1984). Statistical concepts and vector and matrix operations (pp. 6-18).  
[faster pace, for intermediates' review]

### **Fri 4/8 Lab 2: Matrix algebra**

### **L 5: Tu 4/12 Test Theory and Item Analysis**

» Nunnally, J. C. (1967). Theory of measurement error. In J. C. Nunnally, *Psychometric theory*

(ch. 6, pp. 172-205). New York: McGraw-Hill.  
» SPSS Base System User's Guide. ch. 26 (Procedure RELIABILITY)

**L 6: Th 4/14 Principal components analysis (PCA)**

» Dillon & Goldstein (1984). Principal components analysis -- ch. 2 (pp. 23-39, 47-52)  
» SPSS Base System User's Guide. ch. 21 (Procedure FACTOR)  
[optional: Cliff (1987). Components and principal components of variables (ch. 13) --  
challenging but thorough -- or  
Tabachnick & Fidell (2001). ch. 13]

**Fri 4/15 Lab 3: Item Analysis and PCA**

**L 7: Tu 4/19 Factor analysis (FA)**

» Dillon & Goldstein (1984). Factor Analysis (ch. 3)  
» (review from 4/14) SPSS Base System User's Guide. ch. 21 (Procedure FACTOR)  
[optional: Cliff (1987). The common factor model (ch. 15) -- as usual, challenging but thorough.]

**L 8: Th 4/21 Review and comparison of PCA and FA**

» review of readings from L5-L7  
» Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, 28,, 1629-1646. pdf

**Fri 4/23 Lab 4: Factor Analysis**

**Midterm exam posted Mon 4/25, due Mon 5/2, 4:00p.m.**

**L 9: Tu 4/26 (M)ANOVA preludes: Interactions, orthogonality, and significance testing**

» Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 44, 1276-1284. pdf  
» Rosnow, R. L., & Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychological Bulletin*, 105, 143-146. pdf [make sure you read this--it will prevent you from making one of the most wide-spread statistical errors in psychology]  
» Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1, 115-129. pdf

**L 10: Th 4/28 General multivariate analysis of variance (Manova)**

» Tabachnick & Fidell (2001). ch. 9  
» SPSS Advanced Statistics User's Guide. ch. 3 (Procedure MANOVA)  
» Categorical variable coding schemes (from old SPSS handbook)  
» SPSS Manual. Appendix B: Categorical variable coding schemes.  
[optional SPSS Keywords. Interpreting MANOVA parameter estimates.]

**Fri 4/29 Lab 5: General Manova**

**L 11: Tu 5/3 Discriminant function analysis (DFA)**

» Tabachnick & Fidell (2001). ch. 11  
» SPSS Advanced Statistics User's Guide. ch. 1 (Procedure » DISCRIMINANT)

**L 12: Th 5/5 Variations on DFA and Manova**

[optional: Tabachnick & Fidell (2001). ch. 8]

**Fri 5/6** *Lab 6: DFA*

**L 13: Tu 5/10 Univariate vs. multivariate repeated measures analysis of variance**

» [SPSS Advanced Statistics User's Guide. ch. 4 \(More on Procedure MANOVA\)](#)

**L 14: Th 5/12 Multifactor multivariate repeated measures analysis of variance (profile analysis)**

» [Tabachnick & Fidell \(2001\). ch. 10 \(pp. 441-483; pp. 503-505\)](#)

**Fri 5/13** *Lab 7: Repeated-measure analysis*

**L 15: Tu 5/17 (M)Anova expansions (doubly multivariate designs; simple effects)**

» [SPSS handouts on simple effects and on doubly multivariate analysis](#)

[optional: Tabachnick & Fidell (2001). ch. 3]

**L 16: Th 5/19 Logistic Regression**

» [SPSS Advanced Statistics User's Guide. ch. 2 \(LOGISTIC REGRESSION\)](#)

» [Tabachnick & Fidell \(2001\). ch. 12](#)

**Fri 5/20** *Lab 8: Logistic Regression*

**L 17: Tu 5/24 From chi2 to log-linear analysis**

» [Tabachnick & Fidell \(2001\). ch. 7](#)

**L 18: Th 5/26 Log-linear and logit analysis**

» [SPSS Advanced Statistics User's Guide. ch. 5 \(Procedure HILOGLINEAR\) and ch. 6 \(Procedure LOGLINEAR\)](#)

**Fri 5/27** *Lab 9: Log-linear analysis*

**L 19: Tu 5/31 Putting it all together: The multivariate tool box**

[optional: Tabachnick & Fidell (2001). ch. 15, 2]

Final exam posted Wed 6/1

**L 20: Th 6/2 Luxury tools: Canonical correlation, covariance structure analysis; cluster analysis; multidimensional scaling.**

[Tabachnick & Fidell (2001). ch. 14]

**Final exam due Tue 6/7, 12noon**