

# Y604: Multivariate Analysis

## Information

### Course

- Y604, Fall 2006, Course Number 16437 (August 28–December 6, 2006).
- Y500, which is the laboratory component of the course, is required.
- Web Page <http://www.indiana.edu/~kenkel/courses>.
- Note: Y604 and Y500 require block enrollment (Course BE604).

### Instructor

- Ken Kelley, Ph.D.
- Email Address: KKIII@Indiana.Edu
- Office Location: 4040 W. W. Wright
- Office Phone Number: (812) 856-8330
- Office Hours: Monday 2:00–3:30 and by appointment.

### Times and Locations

#### *Lecture*

- Where: Room 1002 in W. W. Wright.
- When: 11:15AM–12:30PM Mondays and Wednesdays.
- Students are expected to attend all lectures.

#### *Laboratory*

- Where: Room 2015 in W. W. Wright.
- When: 1:00–2:00PM Wednesdays.
- Students are expected to attend all laboratory meetings.

### Prerequisites

Competency in basic algebra at the high school level is essential (knowledge of calculus is not required). Successful completion of a graduate univariate statistics class at the level of Y502 is required (completion of a research design course such as Y603 would be ideal but not necessary). Students are expected to have a working knowledge of descriptive statistics, the rationale and conceptual underpinnings of null hypothesis significance tests and confidence intervals,  $t$ -tests, one-way analysis of variance, simple regression analysis (one outcome and one regressor variable), and (ideally) chi-square goodness of fit and tests of independence.

## Course Description

This course is designed to introduce doctoral level students in the behavioral, educational, and social sciences to multivariate statistical techniques useful for addressing research questions that simultaneously involve multiple variables (i.e., questions that are multivariate in nature). The specific goals of the course are to provide a conceptual understanding of the selected multivariate methods, be able to implement the methods, and to determine the types of research questions that can and cannot be addressed with the multivariate methods. The overarching goal is that students will be able to appropriately address research questions that are inherently multivariate.

The General Linear Model (GLM) will be the basis of (roughly) half of the course topics. Many statistical techniques can be represented as a special case of the GLM, and thus a significant amount of our time will be devoted to this general topic and its special cases. We will also discuss methods of finding structure in data and modeling underlying latent variables. More specifically we will cover the following topics: introduction to matrix algebra for multivariate statistics, multiple regression, Hotelling's  $T^2$ , multivariate analysis of variance, discriminate analysis, finite mixture modeling, principal components analysis (briefly), exploratory factor analysis, confirmatory factor analysis, and briefly discuss structural equation modeling, time permitting. A central theme permeating the course will be linking specific questions of interest to the appropriate statistical technique, such that the specific question can be appropriately addressed. The rationale and conceptual underpinnings of the techniques covered will be stressed, as will limitations of the methods. The assumptions that the inferential tests are based will be emphasized and the consequences of failing to meet the assumptions will be discussed.

## Computing

Statistics is more than understanding conceptually the types of questions that can be addressed with different methods and interpreting the results of analyses. In order to effectively answer research questions, the use of computer programs is necessary. Although some analyses can easily be performed by hand when a data set is small, more complex models literally require the use of computer programs. At various points in the course we will use R, SPSS, and Mplus.

R is an Open Source and freely available language and environment for statistical computing and graphics. R is a very powerful program that is widely used in quantitative disciplines. One of R's biggest benefits is its ability to be customized. In fact, there are hundreds of add-on packages contributed by users, effectively ensuring R is one of, if not the most, up-to-date statistics programs available. You can learn more about *The R Project* here: <http://www.r-project.org/>.

SPSS is a general statistics program that performs all basic and many advanced analyses. SPSS is extremely easy to use due to the point-and-click nature of the program. The ridged structure imposed by the point-and-click design of the program, however, limits its usefulness for nonstandard and advanced analyses. Nevertheless, SPSS is the most popular statistics program within many domains in the behavioral, educational, and social sciences.

Mplus is a powerful (more specialized) program for modeling general latent variable models. We will use Mplus for factor analysis and structural equation models (time permitting). Mplus has many advanced features not available in other latent variable modeling programs (although there are some limitations). You can learn more about Mplus here: <http://www.statmodel.com/>.

The laboratory component of the course will use Microsoft Windows XP as the operating system. Although R and SPSS are available on Macintosh and Unix/Linux, Mplus is not (although other quality programs, such as Mx, are). If you are a Macintosh or Unix/Linux user and you have a laptop computer, feel free to bring your laptop to the laboratory sessions when R and/or SPSS is used.

## Laboratory Component

Students are expected to attend laboratory meetings and to work on that day's laboratory assignment(s). In addition to specific laboratory assignments, laboratory time will at times be used to clarify topics from the previous lecture, answer general questions on the assignments, and provide hands-on instruction of R, SPSS, and Mplus. 2

## References

### Required Textbook

Tabachnik, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). New York, NY: Allyn and Bacon.

### Required Articles and Chapters

Available on the course web page: <<http://www.indiana.edu/~kenkel/courses>>.

### Suggested Multivariate Supplemental Resources

Morrison, D. F. (2005). *Multivariate statistical methods* (4th ed.). Belmont, CA: Brooks/Cole.

Bray, J. H. and Maxwell, S. E. (1985). *Multivariate analysis of variance*. Thousand Oaks, CA: Sage.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates

Everitt, B. S. & Dunn, G. (2001). *Applied multivariate data analysis* (2nd ed.). New York, NY: Arnold

Fox, J. (2002). *An R and S-Plus companion to applied regression*. Thousand Oaks, CA: Sage

### Suggested Univariate Resources

Everitt, B. S. (2001). *Statistics for psychologists: An intermediate course*. Mahwah, NJ: Lawrence Erlbaum Associates

Hays, W. L. *Statistics* (5th ed). New York, NY: Harcourt Brace College Publishers.

Howell, D. C. (2007). *Statistical Methods for Psychology* (6th ed.). Pacific Grove, CA: Duxbury

Moore, D. S. & McCabe, G. P. (2006). *Introduction to the practice of statistics* (5th ed.). New York, NY: W. H. Freeman and Company.

Tamhane, A. C. & Dunlop, D. D. (2000). *Statistics and data analysis: From elementary to intermediate*. Upper Saddle River, NJ: Prentice Hall.

## Tentative Course Schedule

Date(s)	Topic(s)	Reading(s)
●8/28	●Welcome and Introduction to Y604	●Tabachnick and Fidell (T&F) Chap. 1
●8/30	●Overview of multivariate methods	●T&F Chap. 2.
●9/4–9/6	●Matrix Algebra	●T&F Appendix A ●: Morrison
●9/11–9/13	●Multiple Regression: General Model	●T&F Chap. 5
●9/18	●Multiple Regression: Interactions	●Cohen et al.

## Tentative Course Schedule

Date(s)	Topic(s)	Reading(s)
●9/20	●Multiple Regression: Sample Size Issues	●Kelley & Maxwell
●9/25	●Multiple Regression: Model Diagnostics	●Hair et al
●9/27	●Analysis of Covariance	●T&F Chap. 6
●10/2	●Hotelling's $T^2$	●Harris
●10/4–10/9	●Multivariate analysis of variance: General Model	●T&F Chap. 7
●10/11	●Multivariate analysis of variance: Comparisons	●Huberty & Morris
●10/16	●The General Linear Model	●T&F Chap. 17 ●Cohen
●10/18–10/25	●Discrimination and Classification	●T&F Chap. 9 ●Panel on Discriminant Analysis
●10/30–11/6	●Class presentations of advanced topics	●Readings made available by presenters
●11/8	●Principal components	●T&F Chap. 13
●11/13-11/15	●Exploratory Factor analysis	●Chapter 7 ●Fabrigar et al.
●11/20	●Flexible Day	●TBA
●11/22	●No class: Happy Thanksgiving!	

## Tentative Course Schedule

Date(s)	Topic(s)	Reading(s)
•11/27–12/4	•Confirmatory Factor Analysis	<ul style="list-style-type: none"> <li>•Boker &amp; McArdle</li> <li>•Brown</li> <li>•Hox &amp; Bechger (Option 1)</li> <li>•Fessinger (Option 2)</li> </ul>
•12/6	•Epilogue	

### Evaluation

Student evaluation will consist of an assignment for each topic covered, a summary of an applied article or chapter that used a method we discuss, a critique of an article or chapter that used a method we discuss, and an in-class presentation (see the course web page for details regarding the specifics of assignment). There will also be an optional extra credit assignment that will potentially add four percentage points to the final grade. Students are strongly encouraged to discuss lecture material, readings, assignments, and especially “big picture” issues with other students. Although the assignments are to be completed individually, this does not preclude students from discussing and offering assistance to one another. Students will generally have one week to complete assignments (although this can be flexible in some situations). Assignments are due at the *beginning* of class the week following completion of the topic.

The topical assignments will be worth 60%, the application 10%, the critique 15%, and the presentation 15% of the final grade. Thus, the way in which a numeric grade will be determined is governed by the following equation:

$$Grade = .60Assignments + .10Critique + .15Critique + .15Presentation.$$

Because numeric grades are reported as ordinal variables represented by letters, the way in which the numeric grade maps onto letter grades will be as follows:

<i>Numeric Score</i>	<i>Letter Grade</i>	<i>Description of Achievement</i>
96–100	A+	Incredible achievement
91–95	A	Outstanding achievement
86–90	A–	Excellent achievement
81–85	B+	Very good achievement
76–80	B	Good achievement
71–75	B–	Fair achievement
66–70	C+	Not wholly satisfactory achievement
61–65	C	Marginal achievement
56–60	C–	Unsatisfactory achievement
51–55	D	Significant lack of achievement
≤ 50	F	Complete lack of achievement

## **Academic Honesty and Intellectual Integrity**

Academic dishonesty of any kind (e.g., cheating, plagiarism, record altering, etc.) will not be tolerated. As stipulated in *General Principles and Policy* section of Indiana University's Academic Handbook (available here: <http://www.indiana.edu/~deanfac/acadhbk/>), academic dishonesty of any kind will be reported.

### **Syllabus Disclaimer**

The information provided on this syllabus is tentative and subject to change. In fact, it will almost certainly change from time to time. Major changes to the syllabus will be noted during lecture and/or laboratory meetings.