Multivariate Statistical Techniques

CLPS2908 | CRN 24571 | **Spring 2013** | **115 Metcalf** | **MW 10:30 am – noon**

Instructor: Bertram F. Malle
Professor of Psychology
bfmalle@brown.edu
Office hours: Tuesday 1-2 p.m. & by apptmt (Metcalf 357)

Teaching Assistant: Imri Sofer
imrisofer@gmail.com
Office hours: Monday 1-3 p.m. (Metcalf 354)
Occasional Lab: Friday, 10:30 a.m. -12 p.m.

Overview. This course covers the basic multivariate techniques currently used in social and cognitive sciences: multiple regression, principal components and factor analysis, multivariate analysis of variance (between-subjects and within-subject), discriminant function analysis, and log-linear analysis.

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

Even though much work will go into learning to conduct the various analyses, a primary goal is to gain a conceptual understanding of each technique, which refers to its function and capacity to *answer scientific questions* (not achieve *p* values). Each technique is a tool that can immensely assist you in understanding your data and testing your hypotheses; but to apply each tool correctly you need to understand both your data and the nature of the proper analysis 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. During this semester 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 of the 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. After lectures you will review the handouts that accompany them and review once more the corresponding textbook chapter. On Thursdays you need to start the week's homework and use offices hours (and the occasional lab) to ask questions. Homework will usually require you to run analyses on provided data sets and to write up the results, culminating in a one-page summary that resembles the report in an empirical journal article. Midterm and final exams are in take-home format and will ask you to show command over the conceptual underpinnings of each technique.

Grading. Your grade will be based on 12 homeworks (50 pts each) and two take-home exams (200 points each), for a total of 1000 points. Grade cut-offs are 900 (A) and 800 (B). As mentioned, the homeworks consist of data retrieval, analysis, interpretation and write-up of results, all done electronically. The exams (web-posted **Mar 22** and **Apr 26**) consist of five to six questions on the functional and conceptual characteristics of the covered techniques.

Readings: Because there is no all-around good, complete, and affordable textbook of multivariate statistics, I have not made any textbook obligatory but have selected readings from various textbooks and other sources for this class (available on MyCourses). You should read the relevant chapters before class to familiarize yourself with the terminology (but don't get hung up on details, because you will better understand them later on). After class you should read the chapter again to deepen you understanding, which will help with your homework and prepare you for the exams.

The recommended optional source book is Tabachnick and Fidell (T & F), *Multivariate Statistics.* If you want one comprehensive reference for multivariate statistics, this one is very useful. (Several of its chapters are part of our reading list.) The major drawback is that its latest edition costs $150. You may be able to find more affordable older editions, which are missing some chapters on recent high-level techniques, but we won't cover those techniques anyway.

Resources: Extensive lecture handouts are posted on the Canvas page (canvas.brown.edu). I also audio record each lecture and post it on Canvas. Before almost every lecture, a previous version of the lecture material can be inspected. However, some changes always emerge when I actually prepare the lecture, so I will update each week's material by Friday.

The Canvas page also contains the electronic reading library, homeworks, and additional resource material.

Please try to minimize printing the resource documents and work instead with electronic material whenever possible.

You will need to have reliable access to SPSS statistical software and be ready to exchange properly formatted electronic documents with the TA and me. Several computers on Brown campus have SPSS (including the computer classrom Metcalf 107), and a Windows version is free from the Brown software website (software.brown.edu/dist/sw-win.html). For Mac users, a 12-month SPSS license costs $80 (Standard Version) at www.studentdiscounts.com, and it can be installed on two computers. *Note:* You need at least the *Standard* version of SPSS, which includes "Advanced Statistics").

Homeworks: Each Thursday, a homework assignment will be posted that is due the following Wednesday at 1 p.m. Given the long time to complete the homework, extensions will be given only in extraordinary circumstances and only if you contact me *before* the due date. Points will be subtracted for lateness (-5 within 24 hours, -3 for each additional day) except for a **one-time** 1-day "grace period"—for the time when the dog (or computer) really ate your homework.

The assignments typically contain data sets and analysis instructions. SPSS analyses can be time consuming because the smallest mistakes lead to error messages, so make sure you read all the lecture handouts and readings carefully because they show sample syntax to run the programs. I also strongly recommend that you have the **SPSS Syntax command** reference document handy (available in the SPSS Help pull-down menu and on our Canvas page). Don't spend more than five minutes trying to fix an error; send us email with the syntax and we can probably figure it out quickly.

Your homework is complete if you run all required analyses, edit the output files down to the essential information (there will typically be a page limit), 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 Wednesday at 1 p.m. For the written results, you can model Tabachnick & Fidell's result summaries or two sample summaries on Canvas. Be sure to write concisely and clearly.

Lab office hours: We will offer an occasional "lab office hour" on Fridays 10:30 am-noon (Metcalf 107) that combines an office hour (to ask your TA questions) and a lab session (to practice the implementation of statistical techniques in SPSS syntax). The lab session does not cover new material, but it does assume that you know the material covered that week, so it is not meant to be remedial. Its primary purpose is to work on questions of general interest.

You can of course visit my own or the TA's regular office hours, but I encourage you to also take advantage of lab sessions when they are offered because knowing what other people find confusing often helps with removing your own confusion.

See next page for Schedule and Readings

Schedule and Readings

The readings in this class are selected chapters from various statistics textbooks, SPSS handbooks, and a few journal articles. In addition, I have listed corresponding Tabachnick & Fidell (2006) chapters where available. In some cases, a Tabachnick & Fidell chapter is required reading. Required readings are printed in blue font below.

All reading materials are available as pdf files on myCourses. Do not distribute these files beyond your use for this class.

Textbook authors cited below:

Cliff, N. (1987). *Analyzing multivariate data.* Harcourt Brace Jovanovich: San Diego.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis: Methods and applications.* Wiley.

Stevens, J. (1996). Applied multivariate statistics for the social sciences. Mahwah, NJ: Erlbaum.

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). New York: HarperCollins.

Orientation: Wed Jan 23

L 1: Mon Jan 28 Basic concepts, the multivariate approach, data habits
» Dillon & Goldstein (1984). Overview of multivariate techniques (pp. 19-22)
» Tabachnick & Fidell (2006). ch. 1 + ch. 2.
Optional
SPSS Syntax command reference

L 2: Wed Jan 30 Exploratory data analysis
» Tabachnick & Fidell (2006). ch. 4
» Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45,* 1304-1312.
Optional
SPSS EXAMINE complete chapter and SPSS EXAMINE syntax reference

---

L 3: Mon Feb 4 Matrix algebra (Introduction)
» [very basic, slow pace, for beginners] Cliff, N. (1987). Elements of matrix algebra for statistical applications (ch. 1), Vectors (ch. 3).
» [faster pace, for intermediates] Dillon & Goldstein (1984). Vector and matrix operations and selected statistical concepts (pp. 521-539).
Optional
Tabachnick & Fidell (2006). Appendix A

L 4: Wed Feb 6   Matrix algebra for statistics
» [faster pace] Dillon & Goldstein (1984). Statistical concepts and vector and matrix operations (pp. 6-18).
» [slower pace, thorough] Cliff, N. (1987). Statistical formulas in matrix form (ch. 2), Variances and covariances of linear combinations (ch. 4), The inverse (ch. 5).
Optional
Handout on determinants and inverses [advanced]

L 5: Mon Feb 11 (Multiple) Regression

» Regression equation: http://people.exeter.ac.uk/SEGLea/psy2005/simpreg.html
» Excerpts from N. Brace, R., Kemp, & R. Snelgar (2006). *SPSS for Psychologists* (Chapter 7, pp. 206-220). Taylor & Francis.
» Stephen Lea: Introduction to Multiple Regression (unpublished lecture notes). http://people.exeter.ac.uk/SEGLea/psy2005/basicmlt.html
» Tabachnick & Fidell (2006). Excerpts from ch. 5 (Multiple Regression, pp. 128-161).
» Muller, A. (2009). Education, income inequality, and mortality: a multiple regression analysis. *BMJ, 324.* Retrieved from bmj.com on 18 January 2009.
» SPSS REGRESSION syntax command chapter.

L 6: Wed Feb 13 Challenges in Multiple Regression

» Keith, T. Z. (2006). *Multiple regression and beyond.* Boston, MA: Pearson.
    Chapter 4: Three and more independent variables and related issues.
    Chapter 7: Categorical and continuous variables.
Introduction to suppressor effects, excerpted from R. Williams, Sociology Graduate Statistics course, University of Notre Dame. http://www.nd.edu/~rwilliam/stats2/l35.pdf
Optional
Rosenberg, M. (1973). The logical status of suppressor variables. *The Public Opinion Quarterly, 37*, 359-372.

No class on Monday February 18 (University long weekend)

L 7: Wed Feb 20 Logistic Regression
» SPSS Advanced Statistics User's Guide. ch. 2 (LOGISTIC REGRESSION)
» Tabachnick & Fidell (2006). ch. 12

L 8: Mon Feb 25 Statistical significance; meta-analysis
» Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1, 115-129.
» DeCoster, J. (2004). Meta-analysis. In Kempf-Leonard, K. (Ed.), *The Encyclopedia of Social Measurement*. San Diego, CA: Academic Press.

Optional
Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*, 183-192.
DeCoster, J. (2004). *Meta-analysis notes.* [Unpublished but very good guide to many aspects of meta-analysis, with formulas.]
Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170-177. [A little difficult, but with a very important message.]

L 9: Wed Feb 27 Test Theory and Item Analysis
» http://en.wikipedia.org/wiki/Classical_test_theory
» SPSS Base System User's Guide. ch. 26 (Procedure RELIABILITY)

» Rogosa, D. (2001). Shoe shopping and the reliability coefficient. *Educational Assessment*, 7, 255-258.

Optional

[very thorough explanation of reliability] Nunnally, J. C. (1967). Theory of measurement error. In J. C. Nunnally, *Psychometric theory* (ch. 6, pp. 172-205). New York: McGraw-Hill.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749.

L 10: Mon Mar 4 Basics of Principal components analysis (PCA)
» Dillon & Goldstein (1984). Principal components analysis (ch. 2, pp. 23-39, 47-52)

Optional

Cliff (1987). Components and principal components of variables (ch. 13)
Tabachnick & Fidell (2006). ch. 13

L 11: Wed Mar 6 PCA decisions and interpretation
» SPSS Base System User's Guide. ch. 21 (Procedure FACTOR)

L 12: Mon Mar 11 Factor analysis (FA)

» Dillon & Goldstein (1984). Factor Analysis (ch. 3)
» (review from Mar 4) SPSS Base System User's Guide. ch. 21 (Procedure FACTOR)

Optional

Cliff (1987). The common factor model (ch. 15)[challenging but thorough.]

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of personality assessment, 68*, 532-560.

L 13: Wed Mar 13 More FA: Extraction, rotation, confirmation
» 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.

L 14: Mon Mar 18 (M)ANOVA preludes: Interactions, orthogonality
» Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,* 1276-1284.
» Rosnow, R. L., & Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychological Bulletin, 105,* 143-146. [Make sure you read this--it will prevent you from making one of the most common statistical errors in psychology!]

No class on Wednesday March 20

Midterm take-home exam posted Thursday, March 21

L 15: Mon Apr 1   From Anova to General Manova
» Tabachnick & Fidell (2006). ch. 9
» SPSS Advanced Statistics User's Guide. ch. 3 (Procedure MANOVA)

Optional

SPSS Keywords: Interpreting MANOVA parameter estimates.

Huberty, C. J., and Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin, 105*, 302-308.

L 16: Wed Apr 3   Expansion of General Manova
» Tabachnick & Fidell (2006). ch. 9
» SPSS Manual: Appendix B–Categorical variable coding schemes.

Midterm take-home exam due Friday, Apr 5

No class on Monday April 8

L 17: Wed Apr 10   Discriminant function analysis (DFA)
» Dillon & Goldstein (1984). Multiple discriminant analysis and related topics
   (Ch. 11, pp. 394-416).
» SPSS Advanced Statistics User's Guide. ch. 1 (Procedure DISCRIMINANT)

Optional
Stevens (1996). Discriminant analysis (Ch. 7, pp. 261-290).
Tabachnick & Fidell (2006). ch. 11

L 18: Mon Apr 15 From univariate to multivariate within-subject analysis of variance
» SPSS Advanced Statistics User's Guide. ch. 4 (More on Procedure MANOVA)

L 19: Wed Apr 17 Within-subject analysis of variance (Profile analysis)
» Tabachnick & Fidell (2006). ch. 10 (pp. 441-483; pp. 503-505)

L 20: Mon Apr 22 Multi-factor designs in ws MANOVA
» SPSS Keywords: Testing simple effects in MANOVA.

L 21: Wed Apr 24 Special designs and review
» SPSS handout on doubly multivariate analysis
Optional
Tabachnick & Fidell (2006). ch. 10 excerpt

L 22: Mon Apr 29   From χ^2 to log-linear analysis

» Stevens (1996). Categorical data analysis: The log linear model (Ch. 14, pp. 518-521, 524-547).

L 23: Wed May 1   Log-linear and logit analysis

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

L 24: Mon May 6   The multivariate tool box/Luxury techniques, Part I (Canonical Correlation, Cluster anlaysis, MDS)

Optional

Tabachnick & Fidell (2006). ch. 14

Tabachnick & Fidell (2006). ch. 15, 2

Thompson, B. (1991, July). A primer on the logic and use of canonical correlation analysis. *Measurement and Evaluation in Counseling and Development, 24,* 80-95.

L 25: Wed May 8 Luxury techniques, Part II (Multi-level modeling, SEM),

Optional

Tabachnick & Fidell (2006). ch. 14

Tabachnick & Fidell (2006). ch. 15, 2

> **Final take-home exam posted Wednesday, May 8**

> **Final take-home exam due Wednesdy, May 15**