

Applied Multivariate Analysis	BST 660
T. Mark Beasley, Ph.D. Associate Professor, Department of Biostatistics E-Mail: mbeasley@uab.edu Website: http://www.soph.uab.edu/statgenetics/People/Beasley/tmbindex.html	RPHB 309-E Voice:(205) 975-4957 Fax: (205) 975-2540

Prerequisites: BST 601 and BST 602 (or equivalent), BST 603 and BST 706 (or equivalent) strongly recommended.

Course Description: This course covers multivariate general linear models, including multivariate regression, multivariate analysis of variance, discriminant analysis, multivariate analysis of covariance, multivariate analysis of repeated measures, and longitudinal data analysis for general and generalized linear models. The linear model basis of principal components and factor analysis will also be covered from the canonical correlation perspective.

Learning Objectives: Students will learn to use SAS and/or SPSS to conduct and interpret multivariate analyses. They will also gain skills necessary for reporting results in written form.

Text: (TF) Tabachnick, B, & Fidell, L. (2000). *Using multivariate statistics* (4th ed.). New York: Allyn & Bacon.

Supplementary Texts:

(GSA) Green, S. B., Salkind, N. J., & Akey, T. M. (2000). *Using SPSS for Windows: Analyzing and understanding data* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

(DE) Der, G., & Everitt, B.S. (2001). *Handbook of Statistical Analyses Using SAS* (2nd Ed.). London: CRC.

HANDOUTS FOR BST 660 ARE AVAILABLE in ACROBAT PDF FORMAT AT:

<http://www.soph.uab.edu/Statgenetics/People/Beasley/Courses/download.htm>

COURSE OUTLINE for APPLIED MULTIVARIATE ANALYSIS (BST 708 660)

WEEKS 1 - 2

I. Introduction to Multivariate Techniques: An Overview

A. Multivariate Research Questions vs. Multiple Univariate Research Questions

(TF1; Fish, 1988^[10]; Huberty, 1994^[15]; Huberty & Morris, 1989^[16])

B. Overview of Techniques (TF2)

C. Canonical Correlation as the most general analytic technique (Knapp, 1978^[19]; Thompson, 1991^[29]).

II. Matrix Algebra (TF Appendix A).

A. Notation and Terminology

1. Types of Matrices; Order (dimensions) of a Matrix

a. Vectors

b. Square and Symmetric Matrices

c. Identity Matrix the matrix equivalent of 1.

d. Null Matrix the matrix equivalent of 0.

2. Matrix addition and subtraction, Matrix conformability.

Exact order must be matched.

3. Trace and Transpose of a Matrix.

4. Matrix Multiplication, Matrix conformability

a. Column of prefactor must match row of postfactor

b. Row of prefactor and column of postfactor

define the order of the product matrix.

c. $\mathbf{A}_{(N \times K)} \mathbf{B}_{(K \times J)} = \mathbf{AB}_{(N \times J)}$

d. $\mathbf{A}'_{(K \times N)} \mathbf{A}_{(N \times K)} = \mathbf{A}'\mathbf{A}_{(K \times K)}$ = Sum of Squares and Cross Products Matrix.

5. Inverse of a Matrix; Matrix "Division"

a. $\mathbf{A}_{(N \times N)} \mathbf{A}^{-1}_{(N \times N)} = \mathbf{A}^{-1}_{(N \times N)} \mathbf{A}_{(N \times N)} = \mathbf{I}_{(N \times N)}$, if N=1 then $\mathbf{AA}^{-1}_{(N \times N)} = A/A = 1$.

b. Not all square matrices have an inverse, they are singular.

c. Singularity one or more column(s) is a linear combination of the other columns. This applies to rows as well.

c. Most symmetric matrices used in statistics do have an inverse.

d. Generalized inverse.

6. Eigenvalues and Determinants.

B. Univariate Procedures in Matrix Notation

WEEKS 3 - 5**III. Multiple Linear Regression**(TF5; DE4, GSA33; Cohen, 1968^[5])

- A. Review and Technical Issues
- B. Interpretation (Huyhn, 2000^[18]; Thompson & Borello, 1985^[30]; McNeil, 1992^[20]; Newman, et al., 1993^[21])
- C. Assumptions
- D. Random Effects vs. Fixed Effects Models
 - 1. Assumption that Y is continuous normally distributed
 - 2. X can be non-normal; Fixed effect models make no assumptions about the shape of X
- E. Regression Diagnostics
- F. Collinearity and Multicollinearity (Singularity).
- G. Nonlinear Regression, the use of exponents (Cohen, 1978^[6])
- H. Product Interaction Terms. (Cohen, 1978^[6])

WEEKS 6 - 7**IV. Exploratory Factor Analysis (TF13; DE13, GSA35; Comrey, 1988^[7])**

- A. Manifest vs. Latent Variables
- B. Factor Analysis as a Multiple Regression Model (Gorsuch, 1997^[36])
- C. Confirmatory vs. Exploratory (Cole, 1987^[32]; Gorsuch, 1997^[36])
- D. Principle Components (PCA) vs. Principal Axis Factoring or Common Factor Extraction (FA) (Velicer & Jackson, 1990^[31])
- E. FA and the choice of elements for the diagonal
 - 1. Squared Multiple Correlations (Tolerance values)
 - 2. Reliability estimates
- F. Eigenvalues and Eigenvectors (Cooper, 1983^[8]; Gorsuch, 1997^[36])
 - 1. For PCA, $\text{trace}(\mathbf{R}) = \sum \lambda = K$, the number of variables. Thus, (λ/K) and $(\sum \lambda/K)$ are proportions of common variance explained.
 - 2. For FA, the diagonal (prior communalities) of \mathbf{R} is replaced with values other than 1. Thus, $[\text{trace}(\mathbf{R}) = \sum \lambda] \neq K$, the number of variables.
- G. The number of factors Decisions (Scree Plot)
- H. Factor Rotation
 - 1. Orthogonal (Varimax); Factors are uncorrelated.
 - 2. Oblique; Factors are not assumed (or expected) to be independent.
 - 3. When do inter-factors correlations suggest a collapsed factor?
- I. Factor Loadings and Posterior Communalities (Cooper, 1983^[8])
- J. Factor Interpretation (Floyd & Widaman, 1995^[35]; Gorsuch, 1997^[36])
- K. Factor Scores (Floyd & Widaman, 1995^[35]; Gorsuch, 1997^[36])
 - 1. PCA or FA factor scores.
 - 2. Sum of variables that “load” onto a factor; loss of orthogonality.
 - 3. Internal Consistency Reliability of Factors (Gorsuch, 1997^[36])

WEEK 9**V. Multivariate Regression & Canonical Correlation Analysis (TF6; Knapp, 1978^[19]; Thompson, 1991^[29]).**

- A. Canonical Correlation as the most general analytic technique (Knapp, 1978^[19]; Thompson, 1991^[29]).
- B. Principal Components Regression.

WEEKS 10-11**VI. Statistical Models with Dichotomous Dependent Variable (Y).**

- A. Assumption that Y is continuous normally distributed
- B. Linear Probability Models
- C. Logistic Regression (TF12; DE8; Cizek & Fitzgerald, 1999^[4])
 - 1. Issues (Davis & Offord, 1997^[34])
 - 2. Interpretation
 - 3. Assumptions
 - 4. IRT as a logistic factor analysis
- D. Predictive Discriminant Analysis (PDA) (TF11; GSA34; DE15; Huberty 1984^[14]; Huberty & Barton, 1989^[17])
- E. PDA vs. Logistic Regression vs. Linear Probability (Huberty, 1972^[12]; Fan & Wang, 1999^[9])

WEEK 12**VII. Statistical Models with Categorical/Ordinal Dependent Variable (Y).**

- A. Contingency table analysis (TF7; Beasley & Schumacker, 1995^[2])
- B. Predictive Discriminant Analysis (Huberty & Barton, 1989^[17])
- C. Multinomial Logistic Models (TF12.8.3)
- D. Loglinear Analyses (TF7)
- E. Cumulative Logit Models (O'Connell, 2000^[22])

WEEKS 13-14**VIII. Multivariate Analysis of Variance (MANOVA) -****Turning Discriminant Analysis “sideways” (TF9; GSA27; Huberty, 1984^[14])**

- A. Experimental Designs with Multiple Continuous Dependent variables
- B. Quasi-Experimental Designs with Multiple Continuous Dependent variables
- C. Issues
 - 1. Multivariate Research Questions (Huberty & Morris, 1989^[16])
 - 2. Multiple Univariate Research Questions (Type I Error Control)
- D. Omnibus Test Statistics for MANOVA (Olson, 1976^[23]; Stevens, 1972^[26])
 - 1. Wilks' Lambda (W); Multivariate Generalization of $1 - \eta^2$.
 - 2. Hotelling's Trace (T); Multivariate Generalization of the F -ratio.
 - 3. Pillai's Trace (V); Multivariate Generalization of η^2 .
 - 4. Roy's Largest Root (R); Based on the same concept as Pillai's trace.
- E. Interpretation of Results and Follow-Up tests (TF9.5)
 - 1. Simultaneous Contrast Procedures (Bird & Hadzi-Pavlovic, 1983^[3]; Sheehan-Holt, 1998^[25])
 - 2. Step-Down Tests (TF9.5.2.2)
 - 3. Descriptive Discriminant Analysis (Huberty & Barton, 1989^[17]; GSA34)
- F. Assumptions of MANOVA
 - 1. Independence of Observations
 - 2. Homogeneity of Covariance Matrices
 - 3. Multivariate Normality
- G. Robustness & Statistical Power (Olson, 1976^[23], 1979^[24]; Stevens, 1979^[27], 1980^[28])
 - 1. Different Test Statistics have different statistical properties
 - 2. Effect of Heterogeneity of Covariance Matrices (V , W , T , R)
 - 3. Effect of Departures from Multivariate Normality (V , W , T , R)
 - 4. Power increases with r s approach 1.00.
 - 5. Concentrated Non-Centrality Structure (R , T , W , V).
 - 6. Diffuse Non-Centrality Structure, (V , W , T , R).
- F. Multivariate Approach to Repeated Measures Designs (TF10; GSA29; DE7)

WEEK 14-15**IX. Introduction to Longitudinal Data Analysis**

- A. Univariate Approach to Repeated Measures Designs (TF10; GSA29; DE7)
- B. Multivariate Approach to Repeated Measures Designs (TF10; GSA29; DE7)
 - 1. Profile Analysis vs. Repeated Measures.
 - 2. Difference Matrix.
 - 3. Profile MANOVA [$y_1 y_2 y_3 y_4$] = MANOVA on [($y_1 - y_2$) ($y_2 - y_3$) ($y_3 - y_4$)]
 - 4. Sphericity
- C. Mixed Linear Model Approach to Repeated Measures Designs (TF10; DE11)
 - 1. SAS PROC MIXED, SPSS 11.0 MIXED Module, & HLM Software
 - 2. Hierarchical Linear Models (HLM) (Wang, 1997) (O'Connell)
 - 3. Random Regression Models (Wang, 1997) (O'Connell)
- D. Structural Equation Model Approach to Repeated Measures Designs (TF10)
 - 1. Longitudinal Panels
 - 2. Latent Growth Curve Models
- E. Survival Analysis – (Time-to-Event/Event History) Longitudinal Models for Dichotomous States (TF15; DE12)

COURSE WORK and EVALUATION

- ◆ Multiple Regression Assignment (Due Week 6 – 15%)
- ◆ Factor Analysis Assignment (Due Week 8 – 15%)
- ◆ Mid-Term Exam (Week 8 – 10%)
- ◆ Logistic Regression Discriminant Analysis Assignment (Due Week 12 – 15%)
- ◆ MANOVA Discriminant Analysis Assignment (Due Week 14 – 15%)
- ◆ Longitudinal Models Assignment (Due Finals Week –15%)
- ◆ Final Exam (Finals Week 15%)

GRADING

- A** = 90% of total points
- B** = 80% of total points
- C** = 65% of total points
- F** < 65% of total points

Disability Student Services

Any student with a disability that may need accommodations in order to successfully complete all requirements for this course should visit the Office of Disability Support Services, located in Room 516 of the Hill University Center (205-934-4205). This office is responsible for registering students and ensuring the University's compliance with Section 504 of the Rehabilitation Act. Once registered, this office will then inform faculty members of all courses in which the student is enrolled, of the student's status, and the specific nature of any accommodations required. Any student requiring such accommodation should discuss this with the course master and assure that the appropriate correspondence is sent from the Office of Disability Support Services.

References

- 1) Baldwin, B. (1989). A primer in the use and interpretation of structural equation models. *Measurement & Evaluation in Counseling & Development*, 22, 100-112.
- 2) Beasley, T. M., & Schumacker, R. E. (1995). Multiple regression approach to analyzing contingency tables: Post-hoc and planned comparison procedures. *Journal of Experimental Education*, 64, 79-93.
- 3) Bird, K. D., & Hadzi-Pavlovic, D. (1983). Simultaneous test procedures and the choice of tests statistic in MANOVA. *Psychological Bulletin*, 105, 302-308.
- 4) Cizek, G. J., & Fitzgerald, S. M. (1999). An introduction of logistic regression. *Measurement & Evaluation in Counseling & Development*, 31, 223-245.
- 5) Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, 70, 426-443.
- 6) Cohen, J. (1978). Partialled products *are* interactions; Partialled powers *are* curve components. *Psychological Bulletin*, 85, 858-866.
- 7) Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, 56, 754-761.
- 8) Cooper, J. C. B. (1983). Factor Analysis: An Overview. *The American Statistician*, 37(2), 141-148
- 9) Fan, X., & Wang, L. (1999). Comparing linear discriminant function with logistic regression for the two-group classification problem. *Journal of Experimental Education*, 67, 265-286.
- 10) Fish, L. J. (1988). Why multivariate methods are usually vital. *Measurement & Evaluation in Counseling & Development*, 21, 130-40.
- 11) Henson, R. K. (2000). Demystifying parametric analyses: Illustrating canonical correlation analyses as the multivariate general linear model. *Multiple Linear Regression Viewpoints*, 26(1), 11-19.
- 12) Huberty, C. J. (1972). Regression analysis and 2-group discriminant analysis *Journal of Experimental Education*, 41, 39-41.
- 13) Huberty, C. J., & Smith J. D. (1982). The study of effects in MANOVA. *Multivariate Behavioral Research*, 17, 417-432.
- 14) Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. *Psychological Bulletin*, 95, 156-171.
- 15) Huberty, C. J. (1994). Why multivariable analyses? *Educational and Psychological Measurement*, 54, 620-627.
- 16) Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, 105, 302-308.
- 17) Huberty, C. J., & Barton, R. M. (1989). An introduction to discriminant analysis. *Measurement & Evaluation in Counseling & Development*, 22, 158-168.
- 18) Huynh, C. (2000). Extraneous variables and the interpretation of regression coefficients. *Multiple Linear Regression Viewpoints*, 26(1), 28-35.

- 19) Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance-testing system. *Psychological Bulletin*, 85, 410-416.
- 20) McNeil, K. (1992). Response to Lunneborg: The conditions for interpretation of regression weights. *Multiple Linear Regression Viewpoints*, 19(1), 37-43.
- 21) Newman, I., Harris, R. J. & McNeil, K. (1993). Both Sides Now: Interpreting Beta Weights. *Mid-Western Educational Researcher*, 6(1), 11-13.
- 22) O'Connell, A. A. (2000). Methods for modeling ordinal outcome variables. *Measurement & Evaluation in Counseling & Development*, 33, 170-193.
- 23) Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 83, 579-586.
- 24) Olson, C. L. (1979). Practical consideration in choosing a MANOVA test statistic: A rejoinder to Stevens. *Psychological Bulletin*, 86, 1350-1352.
- 25) Sheehan-Holt, J. K. (1998). MANOVA simultaneous test procedures: The power and robustness of restricted multivariate contrasts. *Educational and Psychological Measurement*, 58, 861-881.
- 26) Stevens, J. P. (1972). Four methods of analyzing between variation for the K-group MANOVA problem. *Multivariate Behavioral Research*, 7, 499-522.
- 27) Stevens, J. P. (1979). Comment on Olson: Choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 86, 355-360.
- 28) Stevens, J. P. (1980). Power of the multivariate analysis of variance tests. *Psychological Bulletin*, 88, 728-737.
- 29) Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. *Measurement & Evaluation in Counseling & Development*, 24, 80-93.
- 30) Thompson, B. & Borello, G. M. (1985). The importance of structure coefficients in regression research. *Educational and Psychological Measurement*, 45, 203-209.
- 31) Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25, 1-28.
- 32) Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology*, 55, 584-594.
- 33) Crowley, S. L., & Fan, X. (1997). Structural equation modeling: Basic concepts and applications in personality assessment. *Journal of Personality Assessment*, 68, 508-531.
- 34) Davis, L. J., & Offord, K. P. (1997). Logistic Regression. *Journal of Personality Assessment*, 68, 497-507.
- 35) Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299.
- 36) Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, 68, 532-560.