Syllabus for Multivariate Analysis
(26:630:670)

Rutgers Business School Newark● PHD in Management Program

Professor J. Douglas Carroll

Spring 2004

- PHONE: (973) 353-5814 (Office)
- E-MAIL: dcarroll@rci.rutgers.edu
- OFFICE HOURS: 4-5PM Tuesdays or by appointment. E-mail communication is strongly encouraged.
- OFFICE: 125 Management Education Center
- Class meets: Tuesdays, 6-9PM in MEC117
- Text: *Analyzing Multivariate Data* by James M. Lattin, J. Douglas Carroll, and Paul E. Green (Belmont, CA: Duxbury Press).
- Optional Supplementary Text: Carroll, J. D., & Green, P. E. (1997). *Mathematical Tools for Applied Multivariate Analysis*. San Diego, CA: Academic Press (with contributions by A. D. Chaturvedi).

Teaching Assistant: Ulas Akkucuk
Office Hours: By appointment
E-mail: ulas@pegasus.rutgers.edu

Class 1.
Chapter 1:   Introduction. Overview of book and course.

Chapter 2:   Review of basics of vector and matrix algebra and applicability to multivariate analysis (MVA).

Class 2.
Chapter 2 and supplementary material.

Review of Singular Value Decomposition (SVD) of general matrices, Eigenstructure (eigenvector and eigenvalue) decomposition of square matrices, and application of eigenstructure decomposition of square symmetric "product moment" matrices to calculation of SVD of a general matrix. Statistical applications of eigenstructure and SVD analysis of various matrices; use of SVD to determine rank, compute determinant, and define inverse (or various generalized inverses, including the right or left pseudo-inverse) of various special matrices.

Class 3.
Chapter 3:   Regression Analysis.

Overview of multiple linear regression analysis and its applications. Both ordinary regression analysis and partial regression analysis, in which the regression of one set of variables on another after the effect of a third set of variables has been "partialed out," or subtracted, will be dealt with, as well as methods of testing significance of various regression models, or of more complex models as compared to less complex ones.

Class 4.
Chapter 4:   Principal Components Analysis.

Overview of principal components analysis (PCA). Relationship to multivariate normal distribution and to eigenstructure of covariance or correlation matrix and to SVD of original (mean centered or standardized) matrix of observations by variables. The latter discussion will introduce the notion of component scores (closely related to that of factor scores in factor analysis).

Class 5.
Chapter 5:   Exploratory Factor Analysis.

Will primarily focus on principal axis form of exploratory factor analysis (FA) and its relationship to PCA. Will include a discussion of computation of factor scores as well as factor loadings, as well as various approaches to estimation of "communalities".

Class 6 & 7.
Chapter 7:   Multidimensional Scaling.

Metric and nonmetric models and methods of "two-way" multidimensional scaling (MDS). Three-way MDS; specifically the INDSCAL model and the SINDSCAL method of three-way (or individual differences) MDS, and its implementation. Methods of multidimensional analysis of preferential choice (or other "dominance") data. MDS as an alternative to PCA or exploratory FA, as well as a method of analysis of directly judged or otherwise measured proximities (similarities or dissimilarities, measures of association, concurrence, overlap, friendship or other indicators of affinity, or other measures of "closeness") of a set of stimuli or other objects (e.g., products or brands in marketing studies).

Please Note: Take home midterm exam will be distributed in Class 7. Will be due by time of meeting of **Class 9**.   Alternatively, some students may substitute a project in which a substantive data set relevant to that student's current research or planned dissertation research is analyzed via methods discussed in this course. Substitution of such a project will require permission of Prof. Carroll, and, if permitted, the project will be due not later than the meeting of **Class 11.**

Class 8.
Chapter 8:   Cluster Analysis.

Emphasis will be placed on hierarchical clustering, specifically single, complete and average linkage, Ward's method, and K^{th} nearest neighbor clustering, based on direct or derived measures of similarity or dissimilarity, and on K-means clustering for partitioning based on a standard objects by variables multivariate data matrix.

Class 9.
Chapter 9:   Canonical Correlation.

Canonical correlation analysis (CCA) will be presented as a very general technique for interrelating two (or more, in some generalizations) matrices of variables defined on the same objects by finding linear combinations of each that have maximum correlation. The method of computing the first set of canonical variates and their

canonical correlation is discussed and described, and then methods of computing a set of ordered canonical variates, the Kth set of variates constrained to be orthogonal to the first *K-1* (in a certain sense that will be defined) will be covered.

Classes 10 and 11.
Chapter 11:   Analysis of Variance, and generalizations.

A review of standard analysis of variance (ANOVA), plus an introduction of ANCOVA (analysis of covariance), MANOVA (Multivariate ANOVA) and MANCOVA (Multivariate ANCOVA) will all be covered. It will be shown that ANOVA can be viewed as a special case of multiple linear regression analysis, with "dummy" independent variables, while ANCOVA is a special case of partial regression analysis. Both MANOVA and MANCOVA will be shown to be formulatable as special cases of canonical correlation analysis.

Class 12.
Chapter 12:   Multiple Discriminant Analysis.

Multiple Discriminant analysis (MDA) is used when it is desired to find linear combinations of a set of given variables that best discriminate among two or more different groups. MDA will be formulated, and shown to be a special case of canonical correlation analysis, with the (usually continuous) given variables defining one matrix and a set of dummy variables encoding the groups comprising the second one.

Class 13.
Chapters 6:   Confirmatory Factor Analysis and
Chapter 10:   Structural Equations Models with Latent Variables.
OR,
Chapter 13:   Logit Choice Models.

As time permits, this class will deal with a synopsis of various approaches to confirmatory factor analysis (CFA) and Linear Structural Equation Modeling (LISREL and related approaches) as compared to and contrasted with exploratory factor analysis (EFA)

Alternatively, depending on preferences of class members, Class 13 will deal with Chapter 13 on Logit Choice Models.

Final exam: Will be given in last class.

Grading: Grades will be based on midterm exam or project (40%), final exam (40%), as well as class participation and homework (20%).