

SYLLABUS

PSY 546: MULTIVARIATE METHODS

0.0 Catalog Information

Psy 546: Multivariate Methods TLN 6990 03 credits
9:30-10:50 UH 180 Straub

1.0 Instructor

Ray Hyman 323 Straub 686-4910/4921

Office Hours: MF 2:30-3:30 PM

2.0 Assumed Background of Students

I assume each student has had the equivalent of the first-year sequence in statistics which is part of the graduate program in psychology. This means an acquaintance with analysis of variance designs and the multiple regression model.

It would also help to have some familiarity with the statistical packages such as SPSSX, SPSS, SAS, and BMDP. The computing center offers short courses each term which introduce the student to the various computers and facilities. They will also offer a short course on the use of SPSSX. I strongly recommend taking this course.

3.0 Textbooks

3.1 Required:

Marascuilo, L.A., & Levin, J.R. (1983). *Multivariate Statistics in the Social Sciences*. Monterey, CA: Brooks/Cole.

3.2 Recommended for those wanting more math background:

Green, P.E. (1978). *Mathematical Tools for Applied Multivariate Analysis*. New York: Academic Press.

[A few paperback copies should be available at the bookstore in the textbook department.]

4.0 Course Description and Objectives

We will begin with a review of Multiple Regression and the General Linear Model. This will provide an opportunity to introduce the matrix notation and concepts that we will use throughout the course. The Multiple Regression model is a general case that includes analysis of variance and covariance, Fisher's two-group discriminant function and Hotelling's T test.

Then we will take up Canonical Correlation and the General Multivariate Linear Model. In Canonical Correlation the investigator as a set of dependent and independent variables. The goal is to create two linear composites, one of the dependent variables and one of the independent variables, such that the correlation between the two composites is a maximum. When the independent variables are composed completely or partially of qualitative variables we have the multivariate analysis of variance or covariance. Formally, this is equivalent to the multiple-group discriminant function. The two techniques, however, differ in goals. In MANOVA we want to know if the

combinations of experimental conditions account for a significant amount of the variance among the individual profiles on the dependent variables. In discriminant function analysis we want to create a prediction formula which will enable us to classify new individuals into one of a number of discrete groups.

Our objective is to learn to use the multivariate models and methods as "tools for thought." The advantage of the multivariate methods is that they enable us to simultaneously deal with several dependent and independent variables. But this advantage often comes at too high a price. With too many variables or with ill-chosen combinations of variables, the output can be badly misleading. Most of the methods require unreasonable assumptions about multivariate normality and depend upon very large sample sizes even when the assumptions are reasonable. All of them require the use of statistical packages which force the user to make various compromises and which often use questionable options. Perhaps the worst consequence of this dependence upon prepackaged computer programs and complexity of underlying models, is cognitive and communicative impenetrability. The user not only is estranged from his or her data, but often lacks useful understanding of the various transformations which have transpired between input and output. And in those rare instances in which the user does have some understanding, he or she faces the difficult task of having to communicate the findings to others.

Hopefully, we will avoid these disadvantages by using the models and methods as aids to grasping the underlying patterns in our data. Towards this end we will emphasize various ways of getting a better feel for the data. Alternatives and supplements to the standard multivariate methods will be emphasized. Various graphical and tabular aids for detecting patterns will be discussed. The use of diagnostics, especially the study of residuals or discrepancies of the data from the fitted model will be highlighted. Preparation of the data for analysis; preliminary study of the data prior to formal analysis; the problems of missing data, outliers, collinearity, nonlinearity, peculiar distributions, and nominal variables; simultaneous testing and cross-validation; and the choice of variables to include in the analysis will be given special attention.

The overall goal will be *simplification* in conjunction with *understanding*.

5.0 Grading and Requirements

No examinations will be given. The students will be graded on the basis of a class project and general participation. The project is expected to yield both an oral presentation to the class as well as written report to be submitted at the last class meeting.

The class project, which must be negotiated between the student and myself, will be tailored to the needs and background of the individual student. The most obvious type of project will be the analysis of a given data base by several of the multivariate statistical programs. The data base can consist of one of the two or three that I will use during the course or one that the student has already collected for his or her own

purposes. A team of students can collaborate on the analyses of a given data base, but each will make his or her own individual summary and interpretation of the results.

Another alternative for a project is to survey the use of various multivariate methods in one or more areas of psychology or related fields. Here the report would emphasize both the uses and abuses of the various procedures and would indicate to what extent a research goal was advanced or not by the use of the multivariate tool.

6.0 Assignments and Topics*

WEEK	MEETINGS	READING	TOPIC
1	Apr 4	Chs 1,2	Overview. Pay attention to the method of least squares, multiple comparisons, missing data
2	Apr 9,11	Ch 3	Multiple Regression and the General Linear Model
3	Apr 16,18	Ch 4	Matrix Formulation of the Linear Model
4	Apr 23,25	Ch 5	Canonical Correlation and the General Multivariate Linear Model
5	Apr 30,May 2	Ch 6	Principal Components & Factor Analysis
6	May 7,9	Ch 7	Two-Group Linear Discriminant Analysis
7	May 14	Ch 8	Multiple-Group Discriminant Analysis
8	[NO MEETING ON MAY 16] May 21,23		One-Way Multivariate Analysis of Variance
9	May 28,30	Ch 9	Other Multivariate Analysis of Variance Designs
10	Jun 4,6		PROJECT PRESENTATIONS

*The class will not meet on Thursday, May 16, 1985.

COURSE IN MULTIVARIATE METHODS

INFORMATION AND DESCRIPTION

Psy 546: Multivariate Methods will be offered this Spring Term (1985) on Tuesdays and Thursdays in 180 Straub Hall at 9:30 to 11:20 AM. This memorandum will provide some relevant information for those who might be interested in taking the course.

1.0 Prerequisites and Assumed Background

I assume that students taking the course will have taken the first year sequence in statistics in the Psychology Graduate Program or its equivalent. That is, I assume you are familiar with the analysis of variance as covered by Edwards or Keppel and with multiple regression theory as covered by Cohen and Cohen.

1.1. Additional Assumptions

The course will not teach the details of using the various statistical packages such as SPSS, SAS, and BMDP. I assume that you already know how to use these packages or that you can readily learn how by using the relevant manuals and consulting with the appropriate personnel in the computing center.

Multivariate methods, of course, require the use of computers. And you will be expected to use such computational aids during the course [time on the university computers will be made available for this purpose]. Most of the multivariate methods require the use of statistical programs that currently are available mostly on the mainframe computers. Many software packages for personal computers have excellent programs for multiple regression but only a few have any truly multivariate capabilities. I know of 5 current such programs which can do factor analysis, 3 which can do multiple-group discriminant function analysis, 3 which can do cluster analysis, 1 which can do multidimensional scaling, and 1 which can do multivariate analysis of variance. Only SYSTAT currently has the capabilities to seriously compete with the mainframe packages.

2.0 Textbooks

The required textbook is *Multivariate Statistics in the Social Sciences: A Researcher's Guide* by L.A. Marascuilo and J.P. Levin.

A highly recommended, but not required, book is *Mathematical Tools for Applied Multivariate Analysis* by P.E. Green. If we had two terms for the course, it would be ideal to spend the first term going through this book. For those of you who want to get a deeper understanding of the mathematical underpinnings of multivariate analysis, this is an ideal book. I have ordered a few for those of you might want to supplement the regular text with the formal background.

3.0 Course Content

We will begin with a review of Multiple Regression Theory. This will provide an opportunity to introduce the matrix notation and concepts which we will use throughout the course. In

addition, Multiple Regression can be viewed as a general model which embodies most the classical univariate statistical procedures such as analysis of variance, analysis of covariance as well as some multivariate procedures such as Hotelling's T-test and the two-group discriminant function.

Then we will take up Canonical Correlation, both for its own sake and as a general model for the various multivariate techniques such as multivariate analysis of variance, multivariate analysis of covariance, and multiple-group discriminant function.

Some attention will also be given to Principal Component and Factor Analysis as ways of reducing the number of dimensions or variables as well as a tool for avoiding multicollinearity.

4.0 General Objectives

The general approach will be to emphasize the use of multivariate models as tools for thought and as aids to grasping the underlying patterns and structure of a body of data. Towards this end, we will explore the use of tabular and graphical methods for perceiving patterns in the data. The use of diagnostics to uncover departures from assumptions, highlight outliers, and extract information from residuals will be encouraged. Alternatives to the standard multivariate techniques will be considered. In general, the goal will be to use multivariate models as devices to clarify and simplify the relationships in the data rather than to follow the current tendency to obfuscate.

5.0 Grading and Student Requirements

No examinations, as such, will be given. Instead, each student will be graded on the basis of: 1) a written report on a class project; 2) an oral report on this same project to the class; and 3) general participation in the class. The Class Project can take a number of forms depending upon negotiations between the student and myself. Typically, the Class Project will consist of running a given data base through various programs in the statistical packages and reporting on the findings as they relate to the issues for which the data base was assembled. The student can use a database of his or her own choosing or one that will be supplied by the instructor. Another alternative, would be to survey and critique the use of multivariate methods in a given area of research.

Ray Hyman