S Y L L A B U S

PSY 513: MULTIVARIATE ANALYSIS

Spring 1989

0.0 Catalog Details

9:30-10:50 UH+Lab 154 Straub

[The time of the lab meetings will be worked out by the TAs and the students.]

Instructor: Ray Hyman

TA's: Robert Rocklin
Anton Tolman

1.0 Assumed Background of Students

I assume that you have had Psy 511 and Psy 512 or the equivalent. This would include some familiarity with matrix notation, the general linear model, linear and multiple regression using dummy coding, diagnostics, anova, manacova, as well as the SAS statistical package. See me if you are uncertain about your qualifications.

2.0 Texts

Chatfield, C.,& Collins, A.J. (1980). Introduction to multivariate analysis. New York: Chapman and Hall.

Here is what I wrote about this book in last year's syllabus when it was listed in the bibliography:

"Rather succinct, but excellent. Very good on issues such as preliminary data analysis. In addition to the topics covered by Stevens, includes chapters on multidimensional scaling and cluster analysis."

Chatfield and Collins freely use integrals and matrix notation. They also discuss distribution theory at a sophisticated level. However, if you read the words that are interspersed among the symbols you will find much that is both wise and helpful.

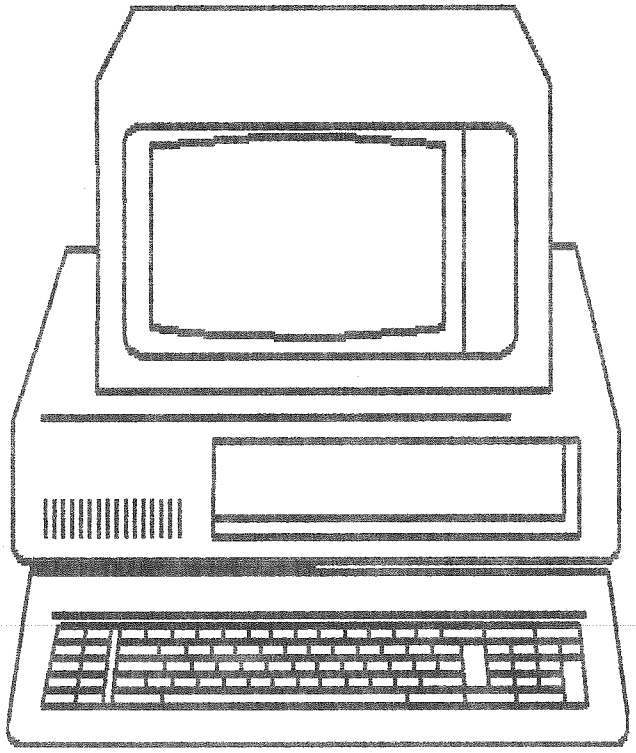
Hand, D. J., & Taylor, C.C. (1987). Multivariate analysis of variance and repeated measures. New York, NY: Chapman and Hall.

Do not be misled by the complete absence of mathematical symbolism and the authors' statement that they have written this book for the "mathematically naive." The authors delve into arcane and conceptually difficult aspects of manova and mancova. Many of their suggestions are controversial but

raise important questions. This book should be most helpful when you are working on your homework and examination problems on manova. Use the detailed examples in the appendices as models.

The preceding two books are the required texts for the course. I assume that you also have access to SAS User's Guide: Basics Version 5 (1985) and SAS User's Guide: Statistics Version 5 (1985).

Don't throw away your copy of Pedhazur! His two chapters on multivariate analysis can give you further insights--especially into canonical correlation and manova. His coverage of matrix algebra also provides a useful review. The graduate students from the preceding two classes should have copies of Stevens which they may be willing to lend to you. This too might be useful for some of you.



3.0 Calendar and Topics

WEEK	ASSIGNMENT	TOPICS
March 30	Chatfield Ch 1 Begin Homework 1 and 2 [Due Apr 11]	Matrix Notation as a device to overcome the Magic Number 7. Definition and classification of MV techniques.
April 4, 6	Chatfield Ch 2(skim) Ch 3	Preliminary data analysis. Eyeballing and graphic tools for exploring multivariate data.
April 11,13	Chatfield Ch 4 Begin Homework 3 [Due Apr 25]	Principal Components. Components of raw scores vs components of standardized scores. Uses of principal components in regression, data reduction and diagnostics. Spectral decomposition.
April 18,20	Chatfield Ch 5	Factor analysis. Numerology vs reality.
April 25,27	Chatfield Ch 9 [pp 162-173] See also Pedhazur Ch 18 Homework 4 [due May 2]	Canonical Correlation and Multivariate Regression. Canonical correlation as the embodiment of the multivariate general linear model. A priori vs post hoc linear composites. Why has canonical correlation failed to live up to its promise? Canonical axes. Planned and post hoc tests. Which test statistic to use?
May 2, 4	MIDTERM	Review plus takehome exam.
May 9,11	Chatfield Ch 6[skim] Ch 7 See also Pedhazur Ch 17 Homework 5 [Due May 23]	Multivariate analogues of the t-test. Two group discriminant function.

May 16,18	Chatfield Ch 8 Hand Chs 1-5 [Use appendices as guides] See also Pedhazur Ch 18	Manova. Issues and questions. Power.
May 23,25	Chatfield Pp 173-186. Hand. Read what is left.	Taking stock. Review, leftover issues, assumptions, misuse of statistics, and other issues that catch our fancy.
May 30, June 1	Chatfield Ch 11	Pattern seeking, classification and cluster analysis. Back to the basics. Multivariate Analysis and the New Age.
JUNE 9	FINAL EXAM due before 12:00pm.	**The final exam will be a takehome affair. The completed exam is due on Friday of the final exam week before noon.**

4.0 Data Sets for Homework, Exams and Labs

During the quarter we will use a small number of data sets. These will be placed in computer files and will be available to you for use in various SAS programs. The tentative data sets we will be using are:

1. STATES. This file contains the death rates for each of 50 states due to such causes as heart disease, cancer, stroke, accidents, and suicide. The same file also contains rates for such crimes as murder,rape, assault, auto theft, larceny and others. Additionally, the file contains information on various demographic variables such as age, poverty, income, population density, education, and the like. The states are organized in 8 census regions. Questions can be asked about the correlates of the various causes of death or the crime rates. One can use manova and mancova to ask questions about how some of these causes of death vary with census reasons.

2. HIGH SCHOOL AND BEYOND. This data base consists of values on 15 variables for 600 high school students. The data are part of a large-scale national longitudinal study conducted by the National Opinion Research Center. The information for each student includes sex, race, school type, high school program, career choice, and standardized scores on locus of control, self-concept, motivation, reading, writing, mathematics, science, and civics.

3. GRANT. This consists of subjects classified according to level of manifest anxiety and assigned to treatments of varying degrees of shock level. The within subjects variable was errors on successive trials on the Wisconsin Card Sorting Task. These data will serve for the variation of manova sometimes called profile analysis.

4. PSY 357. This consists of the scores that the students in Pseudopsychologies obtained on 3 reports, a term paper, a midterm, and the final examination. The 176 students who completed all the work are also classified according to sex, class, school, and grade option.

5. MACH. This consists of the 13x13 correlation matrix reported by Hunter et al (1982) in their study of Machievellian beliefs and personality. Scores were obtained on scales Mach IV, Deceit, Flattery, Immorality, Cynicism, Dogmatism, Self Esteem, Internal Locus of Control, Powerful Others, Chance, Religiosity, Socially Confident, and Competitiveness. We can use this matrix both for factor analysis and canonical correlation.

5.0 A Warmup Assignment [Assignment 1]

We will be using the data on death and crime rates by states for the first homework assignment involving principle

components. We will also use the same data set for one-way manova. Meanwhile, since I know how impatiently you are waiting for the opportunity to demonstrate your analytical flair, you can begin to ponder some aspects of this data base. The sampling unit is the individual state. The states are grouped into geographical regions. For each state we have statistics on rate of death by different causes, and on crime rates in different categories. We also have demographic data such as proportion of individuals below poverty level, proportion of individuals below age 17, proportion over age 65, proportion of college graduates, etc.

a) Before doing any formal statistical analyses it is useful to spend some time trying to eyeball the data. Look at Figure 1 which graphs the three most frequent causes of death for each of the eight census regions defined by the U.S. Census bureau. Notice, that deaths from heart disease and cancer dominate the total. By just studying this graph is there anything you can say about the distribution of death rates among the eight census regions?

b) Look at Figure 2 which graphs two additional causes of death for the same eight census regions. How does this pattern differ from that created by the deaths from cancer, heart disease, and cancer? Do you have any ideas to account for these patterns? How could you test your ideas? What data would you want to have?

c) Have you noticed something peculiar as you pondered questions about this data base? After having successfully survived Psy 511 and Psy 512 you almost certainly have become sensitized to violations of assumptions underlying the statistical models we use. The data on the states has several peculiarities that ought to bother us. Ponder. See how many different peculiarities you can think up. The student who comes up with the most non-trivial peculiarities will be awarded 10 psychic units towards his or her course grade.

d) Use some of the ideas from Chatfield and Collins as well as those discussed in class to further eyeball the data.

6.0 Useful References on Multivariate Analysis

Bernstein, I.H. (1988). Applied multivariate analysis. New York, NY: Springer-Verlag.
On the back cover, the author is quoted as saying, "I attempt to talk English to the reader, not mathematics." Covers both exploratory and confirmatory factor analysis. But coverage of Manova seems somewhat skimpy.

Barker, H.R.,& Barker, B.M. (1984). Multivariate analysis of variance (Manova): a practical guide to its use in scientific decision making. University of Alabama Press.

A brief, almost completely verbal--not a matrix nor complicated formula to be found--guide to understanding manova. I have not read it but it looks like it could provide a security blanket for anyone with phobic reactions to algebraic and matrix equations.

Cliff, N.(1987). Analyzing multivariate data. New York: Harcourt Brace Jovanovich.
The coverage of matrix algebra and its integration with the statistical material is much better than any of the other texts I have seen. Coverage is almost identical to that of Stevens' with the exception that Stevens deals with repeated measures in detail while Cliff has a chapter on factor analysis.

Dillon, W.R.,& Goldstein, M. (1984). Multivariate analysis: methods: methods and applications. New York: Wiley.
Marketing orientation, but worth consulting for a second opinion. A big book with coverage and illustrative applications of principle components, factor analysis, multidimensional scaling, cluster analysis, diagnostics, logistic models, canonical correlation, discriminant analysis, linear structural models (LISREL), and latent structure analysis.

Dunteman, G.H. (1984). Introduction to multivariate analysis. Beverly Hills, CA: Sage.
This is a companion volume to Dunteman's Introduction to linear models. Both are excellent books and include material not found in other books on the same subjects. For example, only Dunteman adequately explains the generalized inverse. A weakness, for our purposes, is that he does not cover manova and mancova.

Green,P.E. (1976). Mathematical tools for applied multivariate analysis. New York: Academic Press.
This is the book to consult when you finally decide you want to know the matrix algebra that will truly help you understand the multivariate procedures.

Hair, J.F., Anderson, R.E., & Tatham, R.L. (1987). Multivariate data analysis with readings(2nd edition). New York, NY: Macmillan.
"Today, just as you need not be a mechanical engineer or automobile mechanic to operate a sophisticated car, you do not have to be a statistician in order to successfully apply powerful multivariate statistical techniques to analyze data with a computer. In Multivariate Data Analysis with Readings we 'talk' our way through the fundamentals on which the family of multivariate statistical techniques is based...We specifically wrote this book for those who want a conceptual understanding of multivariate techniques--what they can do, when they should be used, and how they are interpreted--without becoming bogged down by symbols, formulas, or mathematical derivations...We believe that Multivariate Data

Analysis with Readings is the most practical and academically sound guide available to understanding and applying multivariate statistical techniques." The authors manage to cover the standard multivariate techniques including factor analysis, cluster analysis, multidimensional scaling, and conjoint analysis without even acknowledging the existence of matrix algebra.

Pedhazur, E. J. (1982). Multiple regression in behavioral research (Second edition). New York: Holt, Rinehart and Winston. Most of you already own this monster-sized book. Chapters 17 and 18 constitute a good introduction to our course.

Seber, G.A.F. (1984). Multivariate observations. New York: Wiley.
A valuable book, especially as a resource. Seber includes reviews of almost every simulation study to test the limits and robustness of the various techniques. Seber is especially good on diagnostics and alternatives to the standard algorithms. An excellent chapter on graphical and data-oriented techniques and one on data reduction exemplifies the spirit of his approach.

Stevens, James (1986). Applied Multivariate Statistics for the Social Sciences. Hillsdale, N.J: Lawrence Erlbaum.
This was the textbook I have used in this course for the past two years. I have abandoned it because several students have complained about it. I agree that the book does have some serious weaknesses. But it has two important virtues. Stevens covers multivariate analysis of variance in greater detail than do other multivariate texts. Stevens also takes pains to summarize all the known work on the effect of number of variates and sample size on power. It is sobering to discover the typical psychological study has two many variables compared with sample size to have realistic power.

Tabachnick, B.G., & Fidell, L.S. (1983). Using multivariate statistics. New York: Harper and Row.
A cookbook, but a good one as cookbooks go. Limited by its ties to the BMDP and SPSS packages. Still useful to consult when trying to understand and use one of the techniques.

Tatsuoka, M.M. (1988). Multivariate analysis: techniques for educational and psychological research (2nd edition). New York: Wiley.
The first edition appeared in 1971. While making some concessions to the changes in computing resources during the intervening 17 years, Tatsuoka still believes that the student can benefit from doing some of the calculations by hand. This is an excellent textbook for gaining helpful insights into what the multivariate procedures really do.

7.0 Figures

The next page contains the figures referred to under item 5 above.

FIGURE 1
DEATHS PER 100,000

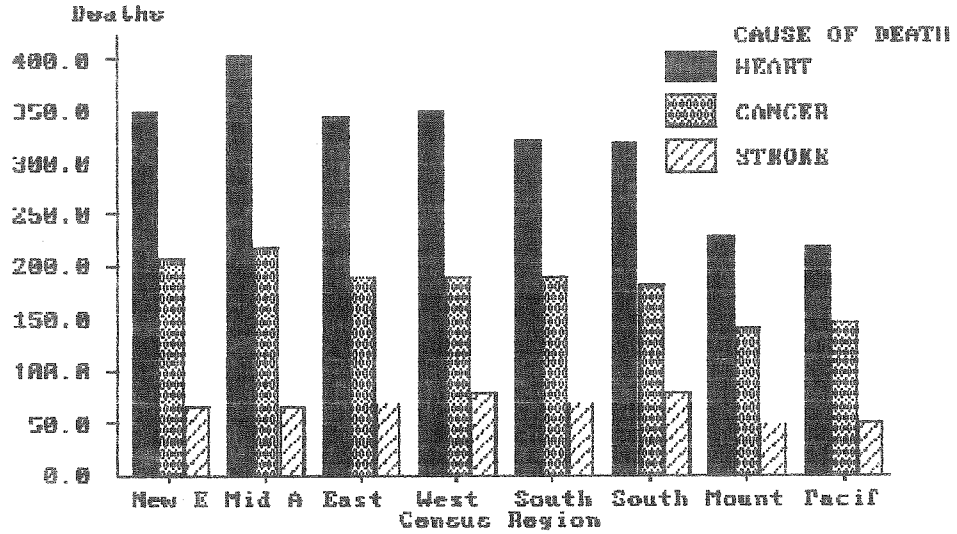
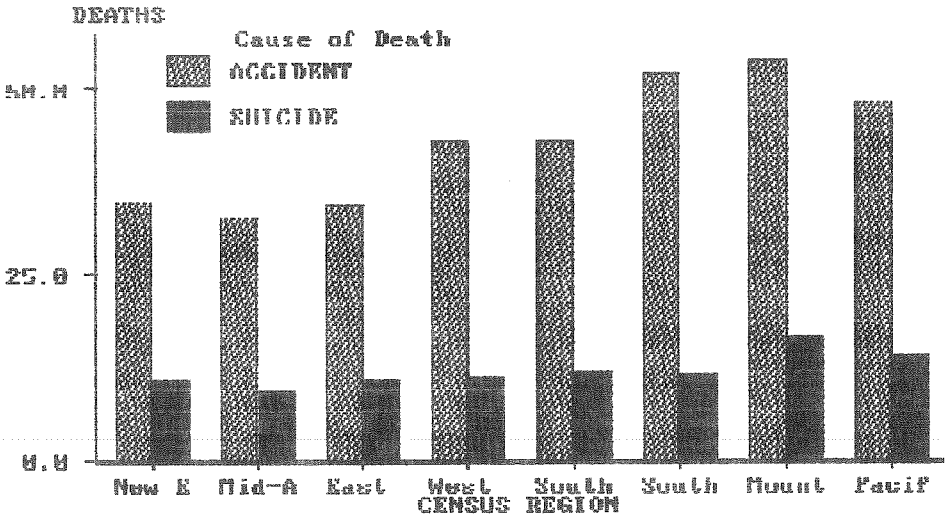


FIGURE 2
DEATHS PER 100,000



The six Census Regions from left to right are: New England, Mid-Atlantic, East North Central, West North Central, South Atlantic, South Central, Mountain, and Pacific states.
States included in each region:
NEW ENGLAND: ME NH VT MA RI CT
MID-ATLANTIC: NY NJ PA
EAST NORTH CENTRAL: OH IN IL MI WI
WEST NORTH CENTRAL: MN IA MO ND SD NB KS
SOUTH ATLANTIC: DE MD VA WV NC SC GA FL
SOUTH CENTRAL: KY TN AL MS AR LA OK TX
MOUNTAIN: MT ID WY CO NM AZ UT NV
PACIFIC: WA OR CA AK HI