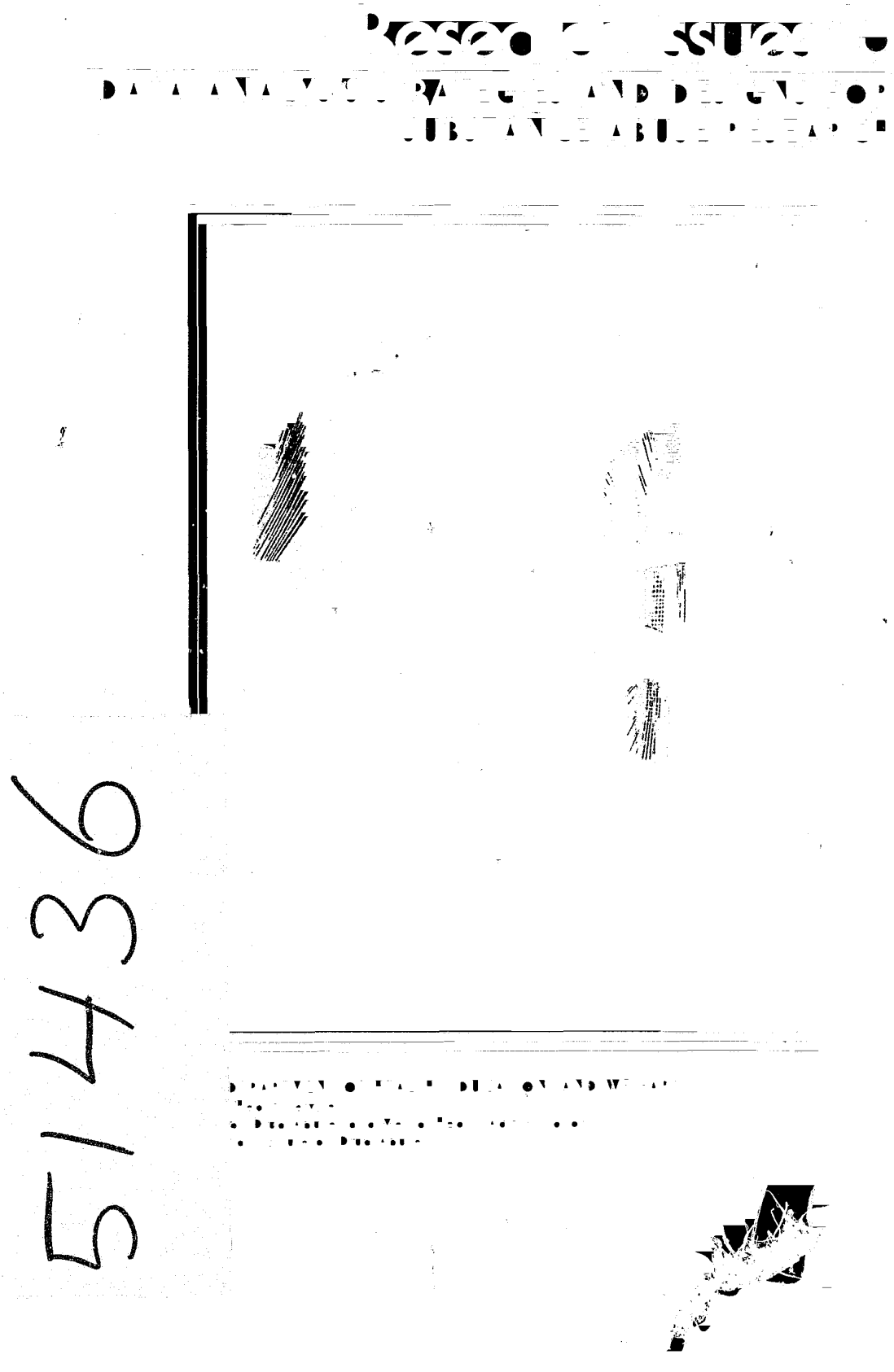


RESEARCH ISSUES SERIES

1. Drugs and Employment
2. Drugs and Sex
3. Drugs and Attitude Change
4. Drugs and Family/Peer Influence
5. Drugs and Pregnancy
6. Drugs and Death
7. Drugs and Addict Lifestyles
8. A Cocaine Bibliography – Nonannotated
9. Drug Themes in Science Fiction
10. Drug Themes in Fiction
11. Predicting Adolescent Drug Abuse
12. Drug Abuse Instrument Handbook
13. Data Analysis Strategies and Designs for Substance Abuse Research.
14. Drugs and Personality
15. Cocaine – Summaries of Psychosocial Research
16. The Lifestyles of Nine American Cocaine Users
17. Drugs and Crime

Cover Illustration

VAULTED STAIRCASE
M.C. Escher
Print Purchase Fund
National Gallery of Art, Washington, D.C.

DATA ANALYSIS STRATEGIES AND DESIGNS FOR SUBSTANCE ABUSE RESEARCH

Edited by

Peter M. Bentler, Ph.D.
University of California, Los Angeles

Dan J. Lettieri, Ph.D.
Division of Research
Psychosocial Branch
National Institute on Drug Abuse

and

Gregory A. Austin, C.Phil.
Documentation Associates

December 1976

National Institute on Drug Abuse
5600 Fishers Lane
Rockville, Maryland 20857

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Stock No. 017-024-00562-2

This volume, part of a Research Issues Series, was prepared for the National Institute on Drug Abuse by Documentation Associates Information Services Incorporated, 11716 West Pico Boulevard, Suite 201, Los Angeles, California 90064, under Contract Number 271-75-3071.

Contents

Foreword

The issues of psychosocial drug use and abuse have generated many volumes analyzing the "problem" and suggesting "solutions." Research has been conducted in many disciplines and from many different points of view. The need to bring together and make accessible the results of these research investigations is becoming increasingly important. The Research Issues Series is intended to aid investigators by collecting, summarizing, and disseminating this large and disparate body of literature. The focus of this series is on critical problems in the field. The topic of each volume is chosen because it represents a challenging issue of current interest to the research community. As additional issues are identified, relevant research will be published as part of the series.

Many of the volumes in the series are reference summaries of major empirical research and theoretical studies of the last fifteen years. These summaries are compiled to provide the reader with the purpose, methodology, findings, and conclusions of the studies in given topic areas. Other volumes are original resource handbooks designed to assist drug researchers. These resource works vary considerably in their topics and contents, but each addresses virtually unexplored areas which have received little attention from the research world.

The Research Issues Series is a group project of staff members of the National Institute on Drug Abuse, Division of Research, Psychosocial Branch. Special gratitude is due Dr. Louise Richards for her continued guidance and support.

Dan J. Lettieri, Ph.D.
Project Officer
National Institute on Drug Abuse

Preface

This volume contains ten original papers discussing methodolcgies applicable to performing psychosocial research on substance abuse, particularly abuse with drugs. The intent of the papers is to permit increased methodological sophistication in the field of drug abuse by making available basic information on some of the latest and most relevant research techniques. Each of the papers has been written by a prominent methodologist; each paper has been designed to assist drug researchers in the behavioral and social sciences who do not have an advanced background in research techniques and who are in need of introductory information. It is also hoped that this volume will provide a stimulus to drug researchers at large.

Eight data analysis strategies are discussed by the authors: automatic interaction detection, actuarial prediction, cluster and typological analysis, path analysis, factor analysis, general multiple regression and correlation analysis, multivariate analysis of variance, and discriminant analysis. In addition, two relevant research designs are dealt with: single-organism designs and longitudinal designs. Although many of the methods are complex, we have tried to keep the discussions as nontechnical as possible. Summaries of the papers are given in chapter 2.

Each paper includes a description of the rationale, procedures, assumptions, advantages, and disadvantages of the methodology. Practical illustrations show how the method has been applied in both nondrug and drug-related situations. References are provided to existing computer programs for performing the analysis, as well as to relevant documents for additional reading. These citations include more detailed discussions of mathematical derivations and descriptions of both drug and nondrug research that have employed the methodology. References are organized alphabetically by author; when more than one publication by a given author or set of authors is cited, publications are listed chronologically.

The content of this volume is the product of an unusual degree of cooperation on the part of a group of authors. All of the authors prepared their papers with great care and considerable effort. After initial drafts were independently generated, the National Institute on Drug Abuse invited the authors to convene in Washington, D.C., to jointly review their work and to discuss the interrelationships among the individual papers. In subsequent months, textual refinements were made. Credit for textual editing and production of the volume is due project staff members Mary Macari, Gayle Kleiman, and Garrie Bateson.

Gregory A. Austin
Project Manager
Documentation Associates

Contributors

Peter M. Bentler, Ph.D.
Professor of Psychology
University of California, Los Angeles

R. Darrell Bock, Ph.D.
Professor, Department of Behavioral Sciences and Education
University of Chicago
Chicago, Illinois

Jacob Cohen, Ph.D.
Professor of Psychology
New York University, New York, New York

Patricia Cohen, Ph.D.
Research Scientist
New York State Department of Mental Hygiene
New York, New York

Susan T. Davidson, B.A.
Research Specialist
Institute for Research in Social Behavior
Berkeley, California

Erich W. Labouvie, Ph.D.
Assistant Professor, Department of Educational Psychology
University of Wisconsin
Madison, Wisconsin

Dan J. Lettieri, Ph.D.
Division of Research
Psychosocial Branch
National Institute on Drug Abuse

Maurice Lorr, Ph.D.
Professor of Psychology
and Staff Member, Boys Town Center for the Study of Youth Development
Catholic University, Washington, D.C.

Glen D. Mellinger, Ph.D.
Research Specialist
Institute for Research in Social Behavior
Berkeley, California

Murray Naditch, Ph.D.
Assistant Professor of Psychology and Sociology
Cornell University
Ithaca, New York

Franklin C. Shontz, Ph.D.
Professor of Psychology
University of Kansas
Lawrence, Kansas

Jacob O. Sines, Ph.D.
Professor of Psychology
University of Iowa
Iowa City, Iowa

Robert H. Somers, Ph.D.
Research Specialist
Institute for Research in Social Behavior
Berkeley, California

Maurice M. Tatsuoka, Ph.D.
Professor of Educational Psychology and Sociology
Cornell University
Ithaca, New York

About the Contributors

PETER M. BENTLER is currently professor of psychology at UCLA, and is director of a National Institute on Drug Abuse research center on psychosocial aspects of drug and alcohol use. In 1972 he was a recipient of the Cattell Award for Distinguished Multivariate Research from the Society for Multivariate Experimental Psychology. He was also a postdoctoral fellow of the Personality Research Group at the Educational Testing Service, and the recipient of a NIDA career award. He has written extensively on research methodologies in various fields of psychology, and is a member of the editorial advisory board of Multivariate Behavioral Research, as well as a consulting editor of Psychometrika and the Journal of Educational Measurement.

R. DARREL BOCK is currently professor in the departments of behavioral science and education at the University of Chicago. He is past president of the Psychometric Society (1972-1973), a director of the Psychometric Laboratory at the University of North Carolina, and a former research fellow of Foundations' Fund for Research in Psychiatry at University College in London (1973-1974). He is an associate editor of Educational Statistics and Psychometrika, and is the author of Multivariate Statistical Methods in Behavioral Research.

JACOB COHEN has been professor of psychology and chairman of quantitative psychology at New York University since 1959. A specialist in statistical analysis and psychometric theory, he currently consults with the Columbia University Medical School, the National Institute of Mental Health, and the National Science Foundation. He is past president of the Society for Multivariate Experimental Psychology, and his publications include Statistical Power Analysis for the Behavioral Sciences, and Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.

PATRICIA R. COHEN is associate research scientist at the New York State Department of Mental Hygiene, and is also a faculty member in the School of Public Health at Columbia University. She has served as a consultant to businesses, universities, and other organizations, offering expertise in research design, statistics, measurement, and data analysis. She is the co-author of Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.

SUSAN T. DAVIDSON is a research analyst at the Institute for Research in Social Behavior in Berkeley, California. Her experience in social research includes studying the identity, styles and values of college students, as well as the patterns of drug acquisition and use. She has also written conference papers and articles for professional journals.

ERICH W. LABOUVIE is currently assistant professor in the Department of Educational Psychology at the University of Wisconsin. He received his M.A. and Ph.D. in developmental psychology and also has an M.S. in physics. His interests include psychometrics, personality development, and gerontology. He has authored several articles on research methodology published in Developmental Psychology, The British Journal of Mathematical and Statistical Psychology, Child Development, Psychological Bulletin, Human Development, and the Journal of Genetic Psychology.

DAN J. LETTIERI received his Ph.D. in social psychology and is currently social psychologist in the psychosocial branch of the National Institute on Drug Abuse. At NIDA, he has served as project editor for the Research Issues Series. His fields of interest include life threatening behaviors and, particularly, the methodological aspects of predicting such behavior. He has served as director of research for the NIMH Center for the Studies of Suicide and Suicide Prevention, and has written on the methodological aspects of interpreting mortality data. He is the editor of Predicting Adolescent Drug Abuse and the co-author of Prediction of Suicide.

MAURICE LORR, currently professor of psychology at Catholic University, Washington, D.C., has over 30 years of experience in the field of psychology and has developed eight psychological tests. He is a fellow of the division of evaluation and measurement and of the division of clinical psychology of the American Psychological Association, and is past president of the Society for Multivariate Experimental Psychology. He has also been an advisory editor of the Journal of Clinical Psychology and of Multivariate Behavioral Research.

GLEN D. MELLINGER, currently at the Institute for Research in Social Behavior, received his Ph.D. in social psychology. He has been principal investigator and researcher for several studies dealing with drug use patterns, the life styles and values of college students, the epidemiology of childhood accidents, and the design and analysis of mass communications networks. He has contributed to such works as Social Science Methods: A New Introduction (R.B. Smith, ed.), Sociological Aspects of Drug Dependence (C. Winick, ed.), and Epidemiology of Drug Use, (E. Josephson, ed.).

MURRAY P. NADITCH, associate professor in the departments of sociology and psychology at Cornell University, is currently conducting a research study on "Acute Adverse Reactions to Psychoactive Drugs" for the National Institute on Drug Abuse. His major areas of interest are personality and social systems, socio-cultural stress, social psychiatry, and social psychological aspects of alcoholism and psychoactive drug usage. He has writeen extensively on various aspects of locus of control, ego mechanisms, depression and anxiety in the Journal of Abnormal Psychology and the Journal of Personality, among others.

FRANKLIN C. SHONTZ is professor of psychology at the University of Kansas. He is also currently consulting on a representative case method research project on the lifestyles of cocaine and amphetamine users, the findings of which will appear in Volume 16 of the Research Issues Series, The Lifestyles of Nine American Cocaine Users. His specializations include research methodology, psychological testing, and rehabilitation psychology. He is the author of Research Methods in Personality, The Psychological Aspects of Physical Illness and Disability, and numerous articles in professional journals.

JACOB O. SINES, professor in the department of psychology at the University of Iowa, received his Ph.D. in clinical psychology. He has served as associate director of the Behavior Research Laboratory at the Washington University School of Medicine and as chief of the medical psychology section of the department of psychiatry at the University of Missouri School of Medicine. He has also consulted with the U.S. Veterans Administration, as well as the U.S. Public Health Service. Many of his articles have appeared in the Journal of Clinical Psychology, the Archives of General Psychiatry, the Journal of Neuropsychiatry, and the Journal of Genetic Psychology.

ROBERT H. SOMERS, of the Institute for Research in Social Behavior in Berkeley, California, is currently co-principal investigator of a study on identity development among university males. He received his Ph.D. in sociology and statistics, and has taught and conducted research at the University of California and the Wright Institute in Berkeley, the Langley Porter Neuropsychiatric Institute in San Francisco, and the Bureau of Social Science Research in Washington, D.C. His articles have appeared in such journals as Multivariate Behavioral Research, Biometrika, and the Journal of the American Statistical Association; he has contributed to Comparative Methodologies: Changes and Continuities (I. Vallier, ed.) and Social Science Methods: A New Introduction (R.B. Smith, ed.).

MAURICE TATSUOKA received his Ed.D. in educational measurement and statistics, and is currently professor of educational psychology at the University of Illinois in Urbana. His previous experience includes teaching mathematics and theoretical physics, as well as conducting research in statistics and psychometrics. Among his specializations are measurement theory and research methodology, including such topics as validation, multivariate analysis, discriminant analysis, experimental methods, significance tests and theories, and classification procedures. He is the author of many booklets in the Selected Topics in Advanced Statistics (STATS) series, and his contributions have appeared in Handbook of Measurement and Assessment in Behavioral Sciences (D.K. Witla, ed.), and Introductory Multivariate Analysis (D. Amick and H. Walberg, eds.).

Plates

1. PRINT GALLERY
   Maurits Cornelis Escher
   Gift of C.V.S. Roosevelt
   National Gallery of Art
   Washington, D.C.

2. VAULTED STAIRCASE
   Maurits Cornelis Escher
   Gift of C.V.S. Roosevelt
   National Gallery of Art
   Washington, D.C.

3. HAND WITH REFLECTING SPHERE
   Maurits Cornelis Escher
   Gift of C.V.S. Roosevelt
   National Gallery of Art
   Washington, D.C.

4. THREE SPHERES II
   Maurits Cornelis Escher
   Gift of C.V.S. Roosevelt
   National Gallery of Art
   Washington, D.C.

5. CUBIC SPACE-FILLING
   Maurits Cornelis Escher
   Rosenwald Collection
   National Gallery of Art
   Washington, D.C.

6. ASCENDING AND DESCENDING
   Maurits Cornelis Escher
   Rosenwald Collection
   National Gallery of Art
   Washington, D.C.

7. THREE SPHERES
   Maurits Cornelis Escher
   Gift of C.V.S. Roosevelt
   National Gallery of Art
   Washington, D.C.

8. SPIRALS
   Maurits Cornelis Escher
   Gift of C.V.S. Roosevelt
   National Gallery of Art
   Washington, D.C.

9. THE FIFTH DAY OF CREATION
   Maurits Cornelis Escher
   Gift of C.V.S. Roosevelt
   National Gallery of Art
   Washington, D.C.

10. DEPTH
    Maurits Cornelis Escher
    Rosenwald Collection
    National Gallery of Art
    Washington, D.C.

11. CUBE WITH MAGIC RIBBONS
    Maurits Cornelis Escher
    Gift of C.V.S. Roosevelt
    National Gallery of Art
    Washington, D.C.

12. RIND
    Maurits Cornelis Escher
    Gift of C.V.S. Roosevelt
    National Gallery of Art
    Washington, D.C.

Plate 1

THE DRUG RESEARCHER'S DILEMMA

This book contains an introductory description of a variety of data analytic methods that will assist the researcher to design, implement, analyze, and write up the results of meaningful psychosocial research in drug and alcohol abuse. Most of us have received some training in methodology--particularly, introductory statistical inference. However, in the field of substance abuse, as in many others, the growth of data analytic methods and designs has been so rapid that it is difficult, if not impossible, for the nonspecialist to keep on top of recent developments. In addition, many drug and alcohol researchers have come to their interest in the psychosocial aspects of abuse via a variety of educational pathways, many of which were not heavily laden with methodological training. It is the experience of confronting the many-faceted problem of substance abuse that has led many individuals to recognize the need for additional training in methodology as it is relevant to their interest.

Since none of us has the time or experience to tackle simultaneously the variety of problems requiring multifaceted skills, each of us has chosen a particular area of specialization. The drug and alcohol abuse researcher, of course, has become a specialist in understanding the development, maintenance, and modification of the destructive use of chemical substances. To study these phenomena in greater depth and with scientific precision requires the use of a variety of specialized quantitative techniques. The relevance of any given technique to a research problem may not be obvious to an individual until he has some overall conceptualization of its power to deal with a specific problem. Witnessing the growing sophistication of substance abuse research, we have become aware that certain techniques seem to be of particular value in clarifying important issues. In some instances, substance abuse researchers themselves have "discovered" these techniques, and we asked these researchers to share their understanding of the relevant techniques with the reader. In other instances, the relevance of certain methodological developments to drug and alcohol research has become apparent, even though there has been no actual application of the methodology to such research as yet. Consequently, we asked an expert to provide an introduction to the specific technique at a level that would be comprehensible and informative to the uninitiated. Of course, we have insured that drug-specific examples of the methods have been incorporated into the discussion of each method, so that the researcher can gain a more concrete picture of the use of the technique for his research goals.

It is unreasonable to expect the scientists committed to the content area of substance abuse to become methodological experts. If highly technical advice is needed, a specialist can and should be consulted. On the other hand, the researcher who has been stumped by certain types of problems, and who may wish to consider a new approach to these problems, may find it useful to explore this book with the goal of winnowing out the wheat from the chaff for his or her particular purposes. Similarly, the individual who has heard about a given technique but has not had the time or energy to find a suitable introductory presentation of its strengths as well as its limitations, the demands on quality and quantity of data that are made by it, and its ease in implementation by computer, may wish to browse through this volume to obtain enough information to be able to make a relatively informed decision about the methodology involved. As editors, we have attempted to make this exploratory attitude both rewarding and relevant to the individual.

EXPLORATION, CONFIRMATION, AND CLASSIFICATION

This volume does not discuss all of the recent developments in methodology and statistics that may be relevant to substance abuse research. In the broader sense, it is obvious that the entire areas of statistics, psychometrics, and sociological measurement would be relevant. How then did the editors arrive at a basis for selecting the methodologies which are discussed in the current volume? In part, the answer lies in relevance; in addition, we applied a theory of data that suggests that there may be three major purposes for the types of data manipulation suggested by the

writers of the subsequent chapters. These purposes include exploration, confirmation, and classification.

The first goal involves a rather preliminary exploration of data for purposes of formulating hypotheses and understanding the potential role of variables, types, or designs. Exploratory data analysis, as we present it, does not particularly involve statistical manipulation with its attendant hypothesis testing. Rather, it involves data fitting or partitioning of a less formal kind. This is useful in earlier stages of scientific development. Of course, several of the exploratory techniques we have included in the volume have hypothesis testing procedures available as well, so that it is easy to move to a more formal level in formulating and evaluating plausible alternative explanations of a given phenomenon. Other techniques cannot bridge this gap and can really only be used in an exploratory way.

After a given hypothesis has been formulated, it becomes necessary to use data observations to test or confirm whether the hypothesis is actually plausible or whether it must be rejected. Often such confirmatory studies are stated in the traditional method of statistical analysis: one states a null hypothesis that suggests that there is actually no effect, and then evaluates whether a given outcome could have occurred by chance under the null hypothesis. If not, one can conclude that some real effect had been operative to produce the result, and understanding the real effect is typically dependent on an adequate theory of the phenomenon in question. Confirmatory data analysis thus generally appears to be more formalized, and it tends to make more specific demands upon the investigator. It becomes relatively more crucial that the data meet certain assumptions; for example, that a data variable has a normal distribution, and that the scale of measurement is continuous. In exchange for the willingness to make these assumptions, the investigator gains in the ability to draw stronger conclusions. Of course, if the assumptions underlying the method cannot really be met, then the conclusions that are drawn are inappropriate. While in a certain sense the investigator can really never go beyond the data, in exploratory methods the bending of a few assumptions is not really so crucial as it is in the case of confirmatory methods.

The primary application of several methods in this volume is not described most clearly by the labels exploration or confirmation. These are methods aimed at the classification of individuals to groups, or drugs to homogeneous methods of action, or treatment successes to typologies. While in many cases such classification is purely exploratory in nature, and in others it is definitely of the hypothesis testing variety, nevertheless, as far as we can determine, clustering and classification seem to be a primary concern of a variety of substance abuse researchers. Can we predict relapse from treatment? Is the effect different for different types of individuals? Questions such as these assume that there is something distinct about the entities under investigation --individuals, drugs, etc.--so that it is not meaningful to talk about two entities as essentially the same except for a slight quantitative difference between them. Rather, there is a tendency to believe that differences are more of a qualitative nature rather than quantitative, with a difference in kind being more important than a difference in amount. The classificatory methods we have included in this volume tend to rely upon this notion, although, again, there is no hard and fast rule that is adhered to in every instance.

RESEARCH DESIGNS

Needless to say, every classification has some imperfections, and so it is with regard to our system for describing the contents of this volume. While we indeed could describe the two types of research designs contained in this volume--single-organism designs and longitudinal designs--in terms of their relevance to exploration, confirmation, or classification, it is meaningful to consider these techniques in their own right. The other eight contributions are more concerned with actual data analytic methodologies; the design chapters are more concerned with the structure of inquiry that generates the data in the first place. We have felt their inclusion was mandatory since these particular designs offer great promise to the drug researcher, though their applications to the area so far have been minimal.

Single-Organism Designs

Most researchers are quite aware of simple group comparative designs, or analysis of variance designs, that are relevant to the analysis of data in certain ways. However, the typical statistics class provides no overview of single-organism research. This probably occurs because such research is typically exploratory in nature. In a stricter sense, it is always necessary to go

beyond the individual case to other individuals to certify the results of such research. Nevertheless, in many contexts research at the level of the single organism is of extremely high quality and very likely to yield insights and evaluations of possible processes at work in a given case. For example, research of a time-series nature across an extended series of observations can provide valuable information about detailed intraindividual change and stability that is typically too expensive to obtain from many individuals. In some areas of drug research, it is almost impossible to propose alternative data gathering strategies. It should be noted that there are now scientific methodologies for drawing inferences that make it possible to evaluate the reliability of given results.

Longitudinal Designs

The chapter on longitudinal designs answers the need for an overview of methodological developments in research associated with changes across time in a group of individuals. Here we are not talking so much about the familiar problems associated with the repeated measurements analysis of variance technique, which is a particular formal statistical problem, but rather about the conceptualization of alternative explanations for given developmental processes. It seems as if drug and alcohol abuse research have recently discovered the longitudinal method, which seems to many to yield a more clearcut view of truth in this difficult research area. However, developmental psychologists and sociologists have made it clear that the longitudinal method is far from the royal road to truth it is sometimes made out to be. In longitudinal research there should not simply be a desire to see what happens to a given set of subjects across time, but rather to formulate data gathering methodologies that can answer the many methodological problems that simple longitudinal designs present to the unsuspecting researcher. Control groups, for example, can help evaluate the potential sources of invalidity in longitudinal research. The optimistic reader who had been hoping that "a followup study" might answer all his questions will have to read this chapter in detail.

THE ANALYTIC METHODS

The next chapter provides summaries of each contribution in this volume. These summaries enable the curious reader to quickly evaluate whether a given technique holds some promise of being relevant. In addition, the next page of this volume contains a figure that can and should be consulted to obtain an overview of the data requirements of a given technique. As pointed out there, each technique has a given number of variables and requires data of a particular level of measurement. In most cases, the methods make a distinction between independent variables and dependent variables, and the usefulness of the distinction must be established for one's own data purposes. We have attempted as well to describe each technique in terms of its primary applications to the areas of exploration, confirmation, or classification. Moving now beyond this summary table, but not quite to the level of the summaries presented in the next chapter, let us describe each technique in a paragraph.

Automatic Interaction Detection

When one has obtained numerous measures on nominal or categorical variables, and one wishes to study the interrelations of the variables to each other and the consequences one may have for another, it is not possible to fall back upon the simple correlation coefficient that one first learned about in an elementary statistics class. Correlation and regression analysis, to be mentioned in greater detail below, tends to require continuous and linearly related data, as well as fairly good understanding of the nature of the variables in order to be applied more effectively. Automatic interaction detection, in contrast, is an exploratory device that has been prepared for computer application to enable one to explore the possible nonlinear consequences of given variables on others. A given effect, for example, may be different for girls than for boys. While correlation and regression methodology would allow one to test hypotheses about interactions, automatic interaction detection is a computer program aimed to enable the investigator to search the data to find interactions. It is possible for theory and experience of the investigator to guide the search for interactions, just as it is possible to work in the absence of well-formulated constructs.

Actuarial Prediction

When conjuring up the word "prediction," the typical researcher remembers his statistical training and attempts to apply the model of linear regression and correlation, predicting one variable from

ANALYTIC METHODS	CHAPTER NO.	VARIABLES			
		INDEPENDENT (PREDICTOR)		DEPENDENT (CRITERIA)	
		NUMBER	MEASUREMENT LEVEL	NUMBER	MEASUREMENT LEVEL
AUTOMATIC INTERACTION DETECTION	5	SEVERAL	NOMINAL[6] ORDINAL[6] INTERVAL[6]	ONE	INTERVAL[1]
ACTUARIAL PREDICTION	6	SEVERAL	NOMINAL[2] ORDINAL INTERVAL	ONE	NOMINAL ORDINAL INTERVAL
CLUSTER AND TYPOLOGICAL ANALYSIS	7	SEVERAL	NOMINAL[2] ORDINAL INTERVAL	*There are no a priori dependent variables*[8]	
PATH ANALYSIS	8	SEVERAL	NOMINAL[2] ORDINAL[2] INTERVAL	*Generally there is no distinction between the independent and dependent variables*	
FACTOR ANALYSIS	9	SEVERAL	INTERVAL[9]	*There are no a priori dependent variables*[8]	
GENERAL MULTIPLE REGRESSION AND CORRELATION	10	SEVERAL	NOMINAL[2] ORDINAL[2] INTERVAL	ONE	ORDINAL INTERVAL
MULTIVARIATE ANALYSIS OF VARIANCE	11	ONE[7] OR SEVERAL	NOMINAL[2] ORDINAL INTERVAL	SEVERAL	INTERVAL[3]
CANONICAL CORRELATION	11	SEVERAL	NOMINAL[2,5] ORDINAL[5] INTERVAL[5]	SEVERAL	*See note #5*
DISCRIMINANT ANALYSIS	12	SEVERAL	NOMINAL[2] ORDINAL INTERVAL	SEVERAL	INTERVAL[4]

1. *May be used with dichotomous dependent variable.*
2. *Can be used if data is converted to the dummy variable format.*
3. *May be used with dichotomous (or polychotomous) dependent variables; however, interval level is statistically preferred.*
4. *Nominal or dichotomous data can be used if the discriminant analysis is employed for prediction or classification problems. Interval data should be employed if the discriminant analysis is employed as a form of MANOVA.*
5. *One of the sets (either the independent or dependent variables) can be nominal (as a dummy variable) or ordinal or interval. The other set must be interval. It doesn't matter which set is which.*
6. *Interval variables are treated by this program as ordinal. Nominal variables (unordered) can be used and do not have to be treated as dummy variables.*
7. *The case of one independent variable obtains with a one way multivariate analysis of variance.*
8. *Although there are no a priori dependent variables, the factors or clusters that are generated can be viewed as dependent variables.*
9. *Non-interval data can be used, but it is not recommended.*

GENERAL MODES OF USAGE AND PURPOSES			SAMPLE SIZES[11] n = number of subjects x = number of independent variables Minimum Sizes: *Italics = recommended minimum* Roman = mandatory minimum
EXPLORATORY	CONFIRMATORY	COMMENTS	
X		*Helps find non-linearities in data*	*n = 500, x = 10*
X		*An alternative to multiple regression. Useful in preliminary search for groups or types*	n = 100, x = 2 *n = 500, x = 10*
X	X	*Useful in generating groups*	n > 15 *(See note #12)*
	X	*Useful for testing causal hypotheses*	*As a guideline one can use notions for multiple regression*
X	X	*To establish basic dimensions*	n = 50, x = 5 *n = 200, x = 20*
X	X	*A valuable predictive method. One of the most general techniques*	$n = 60 + 10\sqrt{x}$ $n = 80 + 20\sqrt{x}$
	X	*A generalization of analysis of variance to several dependent variables*	*Difficult to specify without knowing error structure and magnitude of any fixed effects*
X	X	*A generalization of multiple regression*	
X		*Allows study of nature of group differences. Useful in differentiating existing groups*[10]	n = 3x *n = 10x*

10. *It can be used as a follow-up to MANOVA in describing the nature of group differences, or it can be used as a classificatory or predictive tool.*

11. *The reader is cautioned that the sample sizes suggested are intended as gross guidelines and not as dicta.*

12. *Depends on nature of the subjects and the specific method. Could be applied to 15 cases if they are relevant stimuli or objects.*

another. There are alternatives to linear prediction, although regression and correlation methodology can be extended to handle nonlinear prediction, as discussed in the regression chapter below. But across the years a separate methodology has been developed to deal with prediction that is more in the tradition of insurance research and population surveys than it is in the tradition of psychology and sociology. In these situations one develops actuarial tables in order to predict such attributes as probability of death at a given age, given that one smokes, for example. This actuarial methodology has been introduced into social sciences applications particularly through psychological testing, such as those that might be used in predicting a diagnostic classification from a series of test scores. The actuarial approach does not necessarily assume that the predictors are quantitative in nature or linearly related to each other. It uses a series of sequential empirical steps to develop homogeneous prediction groups, to identify patterns of scores on the predictors that relate to the criterion, and to establish the cross-validational validity of such approaches. Consequently, actuarial prediction is particularly applicable with multiple predictors that are of a nominal or categorial type rather than continuous, as might be the case in discriminant analysis.

Cluster and Typological Analysis

In many data situations, one has numerous scores on given entities such as individuals. These may be nominal, categorical, or continuous in nature. One suspects, however, that the representation of individuals on these scores is not smooth and continuous, as it might be if differing individuals simply differed from one another in slightly varying fashions. Rather, it is suspected that the entire set of individuals may consist of a discernible small number of groups of individuals, such that individuals within a group tend to be quite similar one to the other, and that across groups individuals tend to be relatively dissimilar to each other. It is the purpose of clustering and typological analyses to discover such natural groupings where they may occur. There is no single best approach to the problem of clustering individuals, but rather there is a family of approaches, each of which has advantages and drawbacks. In general, these techniques group individuals on the basis of some measure of similarity or dissimilarity. Computer programs try to find the partitioning of subjects that will yield the homogeneous categories referred to above. At a later stage, it will be necessary to develop a model for understanding the typology in terms of the original variables. Finally, it is possible to test hypotheses about the typologies, although more typically, cluster analysis is an exploratory data analytic technique.

Path Analysis

When one has scores on numerous entities and on numerous variables, and it is possible to conceive of the scores as essentially continuous in nature, it is possible to generate all possible intercorrelations or covariances among the variables. If the scores, in addition, have fairly nice distributions, it is possible to use relatively powerful statistical methods to test a variety of models or hypotheses about the interrelations of variables at a more advanced level than the simple evaluation of whether a given correlation is significantly different from zero or not. Path analysis is a technique for translating models of behavior, such as causal models, into diagrams and equations that represent faithful articulations of a given hypothesis describing the effect of one group of variables on another set of variables. For example, one might believe that the variable "peer influence" somehow leads to or causes drug use. Having translated a given model or miniature theory into a series of diagrams or equations, a number of consequences must be observed in the intercorrelations among the data if the model is a true representation of what actually occurs in the real world. Providing that one's model is correct, the data will be consistent with the model. On the other hand, and far more likely in practice, one's model may be incorrect and the correlational data will be inconsistent with the model as specified. It is in this sense that path analysis is a hypothesis testing, confirmatory procedure, since it allows one to test a given causal model.

Factor Analysis

The same multivariate data that can be studied by path analysis in a confirmatory context can be subjected to factor analysis of a confirmatory kind. In confirmatory factor analysis, one postulates the existence of certain factors that account for the interrelations among the variables. If one's hypothesis is correct, then removing the factors through computer-statistical means will make all variables become uncorrelated. If one's hypothesis is wrong, additional effects will remain. Unlike path analysis, factor analysis also has an exploratory role. Indeed, it is typically used as a means of discovering the possible underlying sources of variation in one's data. Given that one has measured many variables, there may well be far fewer latent sources of variance or factors. A well-known application of factor analysis has been to the area of intelligence,

where some have suggested that all variables of an intellectual nature intercorrelate only because they all measure a single construct "intelligence" (this is false--there are many intellectual factors). Exploratory factor analysis aims to discover as many underlying factors or dimensions as may be needed to account for all the interrelations among the variables the investigator has measured.

General Multiple Regression and Correlation Analysis

The concept of correlation as typically taught in statistics classes is applicable to many variables at once. In the generalization of the simple correlation coefficient between two variables, the most frequently occurring situation is one in which one has a variety of predictor variables and desires to predict a single dependent variable. For example, one may wish to predict drug use from a variety of personality and social variables. Multiple correlation and regression analysis is a technique concerned with the task of prediction itself, assessing the relative contributions of the various predictors to explaining variation in the dependent variable, and to testing hypotheses about whether given influences are truly greater than zero. In contrast to factor analysis, where there is no particular distinction made between independent and dependent variables, this categorization is of fundamental importance to multiple regression. Path analysis, discussed above, can be considered to be a series of simultaneous multiple regressions--where one not only wants to predict a given dependent variable from a set of independent variables--but one may also want to consider one of the independent variables as a criterion to be predicted by some other combination of variables. Thus, multiple regression analysis is of fundamental importance, not only in its own right, but in its implications for other methods.

Multivariate Analysis of Variance

It is hard to get through two semesters in statistics without learning something about analysis of variance. Typically associated with the analysis of continuous data on a single dependent variable according to a particular design or structure of independent variables, the analysis of variance is more appropriately described as a technique for the analysis of means or averages. By isolating sources of variance attributable to different independent variables, typically of a nominal or categorical nature, the analysis of variance attempts to evaluate the effects of these independent variables upon the mean dependent variable score. Although typically applied to data arising from experimental situations, the technique can also be applied where the data are obtained in quasi-experimental or nonexperimental research. The multivariate analysis of variance represents a conceptually very simple generalization of this simple idea. However, instead of having a single dependent variable, the investigator has scores on several dependent variables. The technique aims to determine whether any of the previously specified independent variables have any effect whatsoever upon any of the dependent variables, considered in combination. Obviously, if the independent variables have an important consequence for a single dependent variable, this can also be determined by the multivariate analysis of variance. Although one could perform individual univariate analyses of variance on each dependent variable in turn, repeating the analysis as many times as one had dependent variables, the multivariate analysis of variance performs this chore simultaneously as well as more appropriately from a statistical point of view. The technique tends to be more confirmatory in nature than most of the methods discussed previously, growing out of a hypothesis testing statistical tradition.

Discriminant Analysis

When one has multivariate quantitative data on numerous individuals, each of whom can be considered to be a member of a particular group, the question frequently arises as to whether the variables in question can be used to classify the individuals accurately according to their group membership. This is the basic problem of discriminant analysis. It attempts to use the information in the quantitative variables, considered as independent variables, to predict group membership. In reference to the previous discussion of cluster analysis, if all the individuals within each of the various groups is homogeneous in terms of their score profile, and the score profile in a given group were different from the score profile from another group, it would be immediately obvious that group membership would be perfectly predictable from the pattern of test scores. In the more typical situation, however, a statistical means of weighting the variables must be determined so as to optimally predict group membership from the variables. This procedure is useful not only in the initial problem of determining whether or not it is possible to classify individuals correctly into preexisting groups, but also in the future assignment of a new individual into one of the preexisting groups. The individual would be most accurately assigned, of course, if he was placed in the group whose scores he resembled most closely. This is the task confronted and solved by a discriminant analysis. It is the classification technique par excellence.

THE MINIMAL ASSUMPTIONS

In considering the relevance of one of the above-mentioned techniques to a particular problem, the investigator will immediately face the question of whether the data at hand meet all the assumptions required for the appropriate use of the technique. Some of the specific assumptions required by given methods are discussed in greater detail in the chapter of summaries that follows and, of course, each of the individual chapters goes into this question in great depth. The user always should be prepared to evaluate the given data relative to the requirements of the given technique. There is no point in worrying about the potential relevance of a technique if one's data simply are not in the appropriate form. Are the variables categorized into independent and dependent variables? Are the independent variables quantitative or simply categorical or nominal in nature? Is there one, or are there many dependent variables? Does one have enough subjects to be able to perform the analysis in question? Is the assumption of linearity a reasonable one in the investigator's situation? Are all the variables experimentally independent, or are some scores simple functions of others? Are complete data available for all subjects, or are there missing data? Are there enough subjects so that it is possible to divide all subjects into two groups, one on which exploratory data analysis can be performed, and another upon which a confirmatory, followup hypothesis testing procedure can be performed? Questions such as these will need to be answered by the investigator when exploring each given methodology.

THE VALUE OF COMPUTER PROGRAMS

It must be acknowledged immediately that the vast majority of techniques described in this volume are of such complexity that the untrained individual could never apply the techniques in a reasonable amount of time without the availability of standard computer programs for this purpose. The computer programs are valuable because of their standardized approach and implementation of a given methodology, and because, having been developed and distributed nationally, they also tend to have been tested and evaluated for accuracy and reliability. Thus, if an investigator's data meet the requirements of a particular technique, and if questions that can be answered by a given technique are indeed the ones the investigator wishes to pursue, it is only necessary to consult standard program sources for the implementation of the technique. Wherever relevant, each chapter in this volume lists, under "RESOURCES," the standardly available programs for performing the analysis or methodology described in the chapter. Where alternative programs are available, these are described. In all cases, an attempt is made to focus only upon standardly available programs, rather than esoteric and unreliable ones. Programs that are recommended have been well documented, so that the relative novice should be able to utilize them with accuracy and comprehension. Obviously each program will vary in the kind and nature of supporting information it provides the investigator, so that a certain amount of flexibility will be required by the user. Of particular importance will be the maintenance of a critical eye towards the output from an unfamiliar program, since the various programs usually provide error messages or other clues to possible problems in the computer reading of the investigator's data, in the computations, or in the inability to meet certain crucial assumptions. It is far better to be corrected during the analysis of a given set of data than to publish inaccurate results that may never be replicated by others.

THE IMPORTANCE OF CONSULTATION

As we have emphasized, multivariate and univariate statistical and quantitative techniques have been growing in variety, quantity, and sophistication at an ever increasing pace in recent years. A lack of acquaintance with the newer techniques would almost of necessity represent a universal condition rather than an inadequacy in an investigator. We have prepared this volume with the explicit goal of bringing the reader up-to-date in relevant methodological techniques, but the reader should not believe that an introductory presentation such as we are providing will suffice for all applications of the given techniques. We do believe that the reader will be able to judge from this volume the relevance of a given technique to a given problem, and, furthermore, that where there are computer programs, the investigator will be able to implement the technique. However, there may remain technical questions that are simply not answered in the presented materials. In this case it will be necessary for the reader to turn to the bibliographic references presented at the end of each chapter. These have been selected for their ability to instruct the reader at a more advanced level, as well as for their currentness. If these published works are unable to

satisfy the curiosity of the reader, or do not provide the specialized material required by the investigator, we urge that a consultant in methodology be engaged. The names of appropriate consultants will become obvious from the bibliography.

OTHER METHODS

We were unable to include in this volume as comprehensive a set of techniques as we had hoped. Limitations of resources, time, and space precluded the inclusion of a variety of other methods that have become popular and relevant to modern social science research. One might mention such techniques as discrete multivariate analysis, nonmetric multidimensional scaling, functional measurement, optimal scaling, and recent developments in more well-known fields such as the analysis of covariance. These techniques will have to be discussed in future volumes.

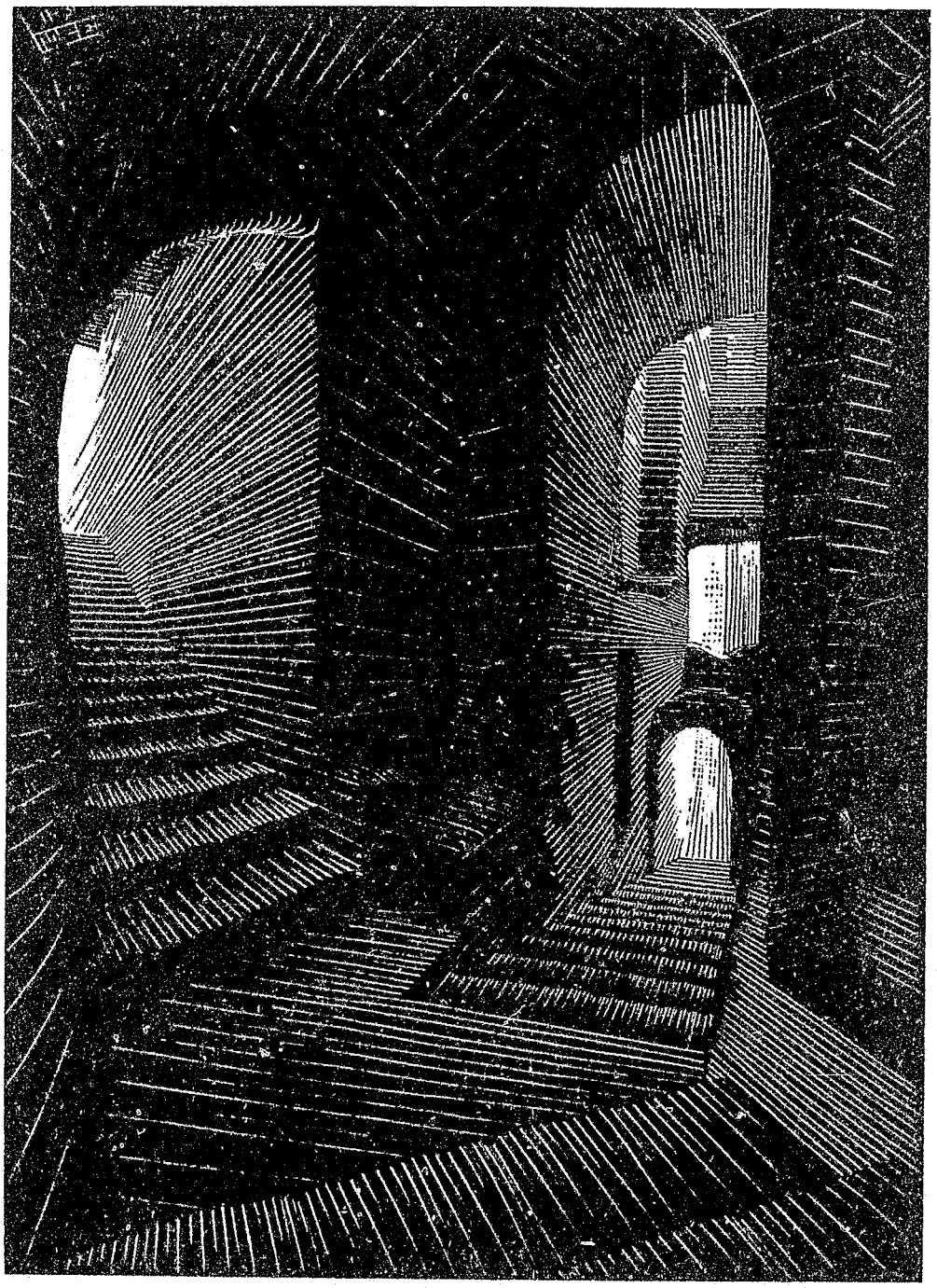


Plate 2

SINGLE-ORGANISM DESIGN BY FRANKLIN C. SHONTZ. CHAPTER 3

Whereas the conventional approach to psychological research advocates a direct search for general laws by studying large groups of subjects, single-organism strategy begins by studying individuals in order to discover valid principles for explaining the behavior of each. It regards the question of generality as one to be answered empirically, through replication of single-organism studies. Detractors argue that single-organism research uses samples that are too small, and, therefore, need not be taken seriously. However, large sample research fails to recognize that sometimes entirely different functional relations apply to group data than apply to data from individuals.

When properly employed, single-organism research is at least as demanding as large-sample methods, and it usually leaves less to chance. In laboratory experiments, single organisms can be more effectively and efficiently handled, more thoroughly known, and subjected to more completely and appropriately controlled conditions than can large groups. In exploratory research, individuals can be specifically selected for appropriateness to particular problems and treated not as "subjects" but as co-investigators, as active participants or even as expert consultants. For example, in representative case research, a single person may be deliberately sought out because he or she displays more clearly the precise characteristics an investigator wishes to study.

Single-organism strategy recognizes that tightly controlled experiments and loosely designed exploratory investigations are both useful in their own ways. Research designs may vary from those that adhere closely to the classical model of experiments conducted in controlled laboratory environments (as in operant conditioning), to quasi-experimental designs (such as time-series analysis in natural process research), and to formal case studies (as in the representative case method) which are as noninterfering as possible. An important feature at all levels is the prominence of investigations with a practical or therapeutic orientation.

Single-organism research assumes that a valid principle may apply to one individual, to only a few organisms, or to everyone. The strategy begins by studying individuals in order to generate valid principles for explaining the behavior of each. The validity of these principles is evaluated by how carefully and correctly each individual is studied. Because generality is a matter for investigation by replication, publication of single-organism studies is only warranted either to demonstrate innovative procedures or to report results after sufficient data have been collected from a sufficient number of organisms, examined under sufficiently well-controlled conditions, to justify the belief that reasonably general statements can be made.

LONGITUDINAL DESIGNS BY ERICH W. LABOUVIE. CHAPTER 4

In the study of behavior it has been common to rely primarily on static cross-sectional rather than longitudinal methods of design. However, it is abundantly clear that any social intervention aimed at modifying human behavior requires by necessity a direct assessment of intraindividual change and interindividual differences in individuals over time, i.e., longitudinal designs. Short-term cross-sectional designs have been preferred because longitudinal studies require greater investments of time and effort on the part of both subjects and researchers. But the internal validity of cross-sectional differences as indicators of intraindividual change is highly questionable. Cross-sectional data do not allow the researcher to trace individual change patterns or to relate earlier observations to later behaviors; and the obtained differences between groups are likely to confound time-related change. Methods that try to short-cut the more laborious and time consuming longitudinal measurement therefore sacrifice at least part or all of that information. The usefulness of simple or cross-sectional designs is limited primarily to initial explorations of behavioral change phenomena. Once a target pattern for a problem has been established, the application of longitudinal designs becomes necessary.

In Simple Longitudinal Designs, a researcher samples individuals from some target population and measures them repeatedly on two or more occasions. Sources of error in internal validity include the influence of testing effects and the possibility of unreliability in retrospective accounts. Methodological deficiencies affecting external validity are even more numerous and more difficult to control for. These include: (1) selective sampling, (2) selective survival, (3) selective drop-out, and (4) generation effects.

To more accurately describe age-related changes, developmental psychologists have introduced more sophisticated Extended Longitudinal Designs. Three types are time-sequential, cohort-sequential and cross-sequential designs. Time-sequential Design, although it does not yield longitudinal observations of intraindividual changes, can provide useful information about general cultural trends as a background against which to evaluate the impact of specific intervention programs. In Cohort-sequential Design, a set of cohorts is observed at different age levels providing a longitudinal series for each of several generations. In Cross-sequential Design, a fixed set of cohorts is observed on several occasions using repeated or independent observations. Because of its greater practicality, this last design has been employed more frequently in empirical studies.

Although all three sequential designs are strictly descriptive, they are preferable over the conventional cross-sectional and longitudinal designs. However, while they are useful to estimate the extent of cultural changes and generation effects, it is important to realize that other sources of error mentioned previously still have to be dealt with. A general strategy to cope with the potential sources of error includes: (1) the use of appropriate series of independent control groups; (2) an explicit attempt to describe various cohort samples in terms of relevant environmental and background variables; and (3) a posteriori comparison between drop-outs and "survivors."

The type of analytical procedures that may be utilized in longitudinal measurement greatly depends on issues concerning the type of dependent variable (quantitative or qualitative) and the particular aspect of change (quantitative or structural) assessed. Such procedures include variance or trend analysis, models that view time-series as stochastic processes, and factor analytic methods.

AUTOMATIC INTERACTION DETECTION BY ROBERT H. SOMERS, GLEN D. MELLINGER, AND SUSAN T. DAVIDSON, CHAPTER 5

Automatic Interaction Detection (AID) is one of the first computer programs developed specifically for the analysis of social science data that makes use of the decision-making capacity of a computer. AID is a multivariate method intended for analysis of a number of independent variables in relation to a single dependent variable. It is a useful preliminary screening or exploratory device to identify components of the sample where interaction occurs.

To identify and judge the import of interaction patterns is a major problem for survey analysts. AID accomplishes the former better than the latter; it is one of the few analysis techniques intentionally designed to identify interaction patterns. The ultimate aim of AID at each level of operation is to account for variation in the dependent variable. The program scans the relationship between predictors and a dependent variable and, on the basis of this scanning, selects the one best way to divide the sample into two groups so that a maximum reduction in variation on the dependent variable is accomplished. That is, it dichotomizes the sample so as to minimize the unexplained variance, then repeats the searching and splitting operation within the groups thus formed, and continues in this way until stopping criteria are reached.

Interaction presents special problems for analysis because it means the assumption of additivity of the effect of predictors on the criterion often required in multivariate analysis is violated. AID examines the relative importance of each of a set of independent variables in predicting a criteria without any assumptions of additivity or linearity. Especially by assuming additivity, other multivariate techniques overcome the need for making qualitative distinctions within the data. The elementary decision-making involved in AID incorporates the idea of making a selection at one level of data analysis, and then pursuing the implications of this and subsequent selections on increasingly deeper levels of analysis. Because it makes no assumptions about the data in terms of measurement properties or additivity, AID is employed usefully as an exploratory device prior to the utilization of multiple regression or partial correlation methods.

AID is intended for categorical predictors which may be unordered, or ranked, or measured on an interval scale. It also is intended principally for survey data, rather than data collected by more quantitative measurement procedures. A fairly large sample size is useful, although the program itself imposes no restrictions on sample size. In contrast to the flexibility regarding measurement assumptions in the predictors, AID requires interval measurement or dichotomization of the criterion. As AID makes few assumptions about the data, it takes literally each observed value that is presented to it, largely ignoring problems of sampling and measurement error. Even in the light of this limitation, AID is an extremely useful exploratory device.

ACTUARIAL PREDICTION BY JACOB O. SINES, CHAPTER 6

The actuarial approach is a set of methods for searching and identifying homogeneous subtypes or classes of individuals, and for predicting or understanding their behavior with a clinically and socially significant degree of precision. It enables the evaluation of the extent to which subtypes of drug users share relatively homogeneous etiologies, patterns of drug use or responses to specific treatment programs. Actuarial prediction is useful particularly in the identification of a set of taxonomic classes of drug users on the basis of psychological test scores.

Many assume, with cause, that some of the personality characteristics measured by one or another psychological test are related to clinically important characteristics of drug users. Psychological test variables also are "psychometrically tractible" and are able to be examined as useful predictors. For example, the four major actuarial systems developed for use with psychiatric patients have used the MMPI. Of course, there are times when the nature of the criterion to be predicted renders personality tests less appropriate than other types of predictors. Therefore, it is appropriate to collect as many types of predictor data as possible. It is also necessary to have far more than the usual number and kind of nontest information or criterion data about one's subjects. The largest practical array of clinically important information on each patient should be collected in the hope that some of the data may indeed be predictable from one or more of the patterns of test scores that may be identified. The relationship between test-defined groups and the criterion of interest can be empirically determined through grouping or clustering procedures such as r_p, profile correlation, D^2, and several nonstatistical methods. The D^2 technique is preferred.

The approach assumes that there are several distinguishable patterns of psychological test data that will define relatively homogeneous groups or taxonomic classes. However, the mere identification of psychometrically homogeneous subgroups is of relatively little clinical value unless members of such classes are found also to be homogeneous with respect to other clinically important nontest characteristics such as etiology, patterns of use, or response to treatment. Also, psychometrically highly homogeneous classes may only contain a few individuals; if one generates classes accommodating relatively large numbers of patients, they may be too heterogeneous. The appropriate narrowness of a test-defined group must be determined for each question.

It is erroneous to assume that personality variables assessed by a test must describe all drug users and must identify and distinguish between all the clinically meaningful subgroups of drug abusers. A few clinically quite important test-defined types of classes of drug users may be identified using one such measuring instrument, and yet scores or patterns of scores on that test may be unrelated to clinically meaningful characteristics of the remaining large proportion of drug users. If such is the case, another assessment instrument might identify clinically meaningful subgroups among the remainder of the drug users who had not already been classified.

While it is certainly hoped that some psychometrically homogeneous groups show greater-than-base-rate homogeneity in some clinically important respect, our present level of knowledge does not guarantee such a positive finding. But if one or more of the groups identified using one particular set of predictor variables shows a clinically important degree of homogeneity in terms of our criteria of interest, those are valuable data. In such a case, one should routinely collect those predictor data and make clinical decisions on the basis of membership in those groups while attempting to identify additional psychometrically homogeneous groups among the remaining patients using other domains of predictors.

CLUSTER AND TYPOLOGICAL ANALYSIS BY MAURICE LORR, CHAPTER 7

Cluster analysis of multivariate data groups together persons, objects, concepts or events into coherent classes on the basis of their measured similarities. The main goals of analysis are to recover or identify "natural clusters" of entities, generate a conceptual scheme reflecting their interrelationships, discover structure inherent in a body of data, and test hypotheses about groupings believed to be present in the data. The clustering process itself can be broken down into a number of steps as follows:

(1) Select a representative set of entities to be studied.
(2) Define the domain of similarity to be studied and select a representative set of attributes.
(3) Convert scores into a comparable metric if needed. Decide whether or not to include categorical as well as continuous variables.
(4) Decide whether to use factor analysis to reduce the number of descriptor variables.
(5) Select a suitable index of similarity or dissimilarity between pairs of entities.
(6) Choose a structural model for the clusters or types anticipated. The main models are the compact or homogeneous, the chained or continuously connected, and the hierarchical.
(7) Select an appropriate method of clustering, an efficient algorithm, and apply the procedure to the matrix of indices of similarity--dissimilarity.
(8) Determine the mean profiles of the various clusters found or convert into a tree-structure or dendogram.
(9) Interpret the results and choose some decision function (i.e., discriminant functions, multiple cutting scores, Bayesian analysis) to allocate new cases to the subgroup to which they belong.

There are several problems involved in the process of searching for groups or categories. Of considerable import are the variables and scales of measurement selected. In the social and behavioral sciences it is important to allocate variables to one of four kinds of measurement scales. The most rudimentary is the nominal or classificatory scale, whereby numbers or symbols are used to classify entities. A given collection of objects are partitioned into a set of mutually exclusive subsets. Ordinal scales reflect consistent rank orders. Objects in one category of the scale differ from objects in other categories of the scale by being greater than or less than. An interval scale is characterized by a constant or equal unit of measurement. The scale has all the characteristics of an ordinal scale but, in addition, provides a distance between any two objects. All of the parametric statistics such as means, standard deviations and correlations are applications to interval scale data. Finally, a ratio scale is an interval scale with a true zero point as its origin. The ratio of any two scale points is independent of the unit of measurement. This scale is extremely rare in the social or behavioral sciences.

Remember that similarity is not a general quality. It is necessary to specify the domain of similarity--difference in discussing the similarity of persons, objects or events. If a group of people are found to be similar on one set of scores, it is not justifiable to assume their similarity in general.

The three major structural models in typing and cluster analysis are: (1) compact or homogeneous, (2) chained or continuously connected, and (3) hierarchical. Members of the compact type are said to be similar or dissimilar, alike or different, close or far, etc. Within the chained type, ordinal (dominance) relations exist among objects within a type. The hierarchical scheme is usually represented by a hierarchical tree or dendogram. A hierarchy may be seen as a nested set of clusters in which each level is assigned a rank.

Cluster analyses procedures include: (1) density or mode seeking, (2) partitioning (3) clumping, and (4) hierarchical clustering. Density seeking searches for modes or regions of high density for entities in attribute space. Partitioning subdivides a collection or set of entities into mutually exclusive classes. Clumping groups objects into overlapping subsets. Finally, hierarchical clustering groups entities into clusters and merges the clusters at successive levels to form a tree. Merging of clusters can be done using, among others, single linkage, complete linkage, and average linkage analyses.

Ordination, or obtaining a low dimensional mapping of a set of data points, can be accomplished with principal components analysis, multidimensional scaling technique (MDS), and discriminant function analysis.

PATH ANALYSIS BY MURRAY P. NADITCH, CHAPTER 8

Path analysis is a mathematical modeling technique, based on multiple regression, that can be used to specify relations among a set of variables. When underlying assumptions are met, it represents a rather elegant way to express verbal theory in a diagram of causal paths, making implicit assumptions explicit and facilitating theory development. A set of structural equations, isomorphic to the causal path network, are used to estimate the magnitude of various parameters of the model. The first step is to hypothesize the important explanatory variables and then establish a temporal, theoretically appropriate ordering of the variables as they causally relate to the outcome being studied. The relationships among the variables are then exhibited in a path diagram, the presence or absence of causal arrows being based on theory and previous empirical research. (The diagram can thus be considered a statement of the author's hypothesis.) Numerical path coefficients are finally estimated from the statistical data using multiple regression techniques.

By estimating the path coefficients of this series of equations, the researcher can estimate the magnitude of parameters in the model. Often this enables researchers to reject aspects of the hypothesis which can then be reformulated in the light of empirical findings. Used in conjunction with longitudinal data, such a model facilitates analysis of the effects of possible intervention strategies or programs. The validity of any path model as a description of reality depends, however, both on the quality of the theoretical hypotheses constituting the model and also the representativeness and quality of the data from which the parameters are estimated. The most important prerequisite is a theoretically defensible specification of a model. Path coefficients will be biased to the degree and extent to which the equations estimated differ from hypothetical equations that "truly" describe the process being explained.

Path analysis assumes that a set of variables can be temporally ordered, and are asymmetrically related. Satisfying these assumptions may be especially difficult in drug research using cross-sectional data. This problem can sometimes be overcome with time-series data.

FACTOR ANALYSIS BY PETER M. BENTLER. CHAPTER 9

Factor analysis is the most widely used of all methods of multivariate analysis. Its major goal is the analysis and description of all sources of variance in the data when all the variables are mutually dependent. It is a means toward identification of important underlying variables in a given set of data. The factors of factor analysis try to account for the covariance or correlations among mutually dependent variables. When summarizing vast amounts of data, one may wish to find out only what it is that various variables share in common; specific aspects of a given variable that are not shared by other variables may be relegated to an irrelevant role. Typically, the part of a given variable that is shared by many other variables is called the common part; the part that is unique to a given variable, the specific and error part, is called the unique part. Each of the sources of variation in the common parts is called a common factor, or simply, a factor. There are also unique factors, but in factor analysis it is the common factors that are of special importance, since these represent independent variables that share variance among many dependent, given variables.

Factor analysis enables one to determine whether a single underlying variable (i.e., factor) can summarize all the information in a set of dependent variables (i.e., all the consistent differences among entities), such that the given variables are functions of the factor. This sets it apart from both analysis of variance, which seeks to determine the effects of independent variables on dependent variables, and principal components analysis, which seeks to obtain new observed variables as functions of the given variables such that the new variables account for as much variance on each and every variable as possible. Computer programs are available to perform the complicated mathematics.

In Data Reduction, factor analysis enables one to reduce masses of multivariate data to only a few factors. Exploratory Factor Analysis, the most frequent application, aids in acquiring a theoretical understanding of the nature of the factors. In purely exploratory work, without a well-developed theory nor enough previous data, one may not be able to predict with great accuracy what the various factors might be that account for the covariation observed among variables in a given domain. Here, factor analysis can be a viable alternative to stepwise regression in explaining the nature of underlying variables. Confirmatory Factor Analysis serves to cross-validate findings from a previous study or from a series of previous studies. It enables one to test the hypothesis that the given number of dimensions underlying the covariation among variables is some specific number, and the hypothesis that a given factor loading or beta weight has some specified value.

Factor analysis is a linear model. Dependability of results hinges strictly upon having an adequately large and random sample of entities (at least five times as many entities as variables), and having at least five variables for every factor. The samples of variables and entities must be adequate representations of the universe of variables and of the population of subjects. Missing data cannot be handled and it is assumed that data variables are experimentally independent.

GENERAL MULTIPLE REGRESSION AND CORRELATION BY JACOB COHEN AND PATRICIA COHEN. CHAPTER 10

Multiple regression and correlation (MRC) is a well-established data analytic procedure, long the method of choice when the relationship between one dependent (criterion) variable and a group of two or more independent variables (predictors) are studied. During the last decade, the scope and generality of MRC has greatly expanded. It is now known that virtually any information may be represented as independent variables and their bearing on a single dependent variable can be evaluated. The linear multiple regression equation can be used to estimate any individual's value on a criterion by entering the information regarding his predictor values. Applied to all subjects, the estimate is the best possible by the "least squares" criterion.

The numerical constants in the regression equation are not only error-minimizing values but have important interpretive properties. It is possible to determine the effect on the criterion caused by a change in a predictor when all other variables are held constant or partialled. This partialling is a centrally important feature of MRC, since it makes it possible to determine if any one variable has a sizeable and significant effect on the dependent variable when there is otherwise comparable standing on all other variables. In considering the association of each predictor on the criterion, MRC provides three different correlation coefficients whose squares are interpretable as proportions of variance. The ordinary squared product moment correlation gives the proportion of variance linearly accounted for by one predictor alone, ignoring any relationship this predictor may have with others, or the others' relationship to the criterion. Second, the squared semipartial correlation gives the proportion of variance accounted for by the part of a predictor which is unique to the predictor, i.e., the part which it does not share with others. Third, the squared partial correlation gives the expected value for the squared product moment correlation for subsets of cases, all of which share the same value on the other variables, i.e., are "held constant statistically". Thus, the relationship of one variable to the criterion can be estimated, uninfluenced by their relationship to the other variables.

All independent variables may be simultaneously regressed and correlated with the dependent variable. One result of so proceeding is that for each independent variable, all the others are partialled in the determination of partial regression and correlation coefficients. An alternative, hierarchical strategy enters each predictor successively in a predefined order, and determines for that hierarchical order how much each adds to the prior squared multiple correlation (R^2). The hierarchical strategy is the MRC method of choice in the analysis of the data of surveys and quasi experiments, and in the analysis of covariance and its generalization. A computer-defined hierarchical procedure ("stepwise regression analysis") can be used to select a small subset of predictors that predict that criterion well.

A research factor may be represented as a set of independent variables, and the set is the functional unit of analysis in general MRC. By using sets of predictors, one may bring into the MRC system group membership (nominal scale) information, nonlinear relationships, variables with missing data, and interactive information. Also, by using sets which function as control variables, one can greatly increase the scope and relevance of MRC to data analysis.

General MRC analysis offers a uniquely powerful device for the exploitation of data. By partialling a set A from set B and by using a single very general F test, significance testing is also possible. The null hypothesis tested throughout is that the population parameter value of the observed sample statistic equals zero; for example, that in the population, set B accounts for no criterion variance beyond what is accounted for by set A. The statistical power of the significance test, which is the probability that it will reject the null hypothesis, can also be evaluated.

The mere possibility of inclusion of almost any kind of information does not make all such possibilities equally desirable. One can have lesser confidence in the procedure as the number of hypotheses tested gets large. The larger the number of predictors, the more difficulty may be anticipated in interpreting the results. Finally, decrease in power occurs as the number of independent variables studied increases. These problems may be resolved in several ways. First, distinguish between variables whose function it is to test the validity of assumptions, and those representing real substantive hypotheses. Second, minimize the inclusion of redundant variables. Third, employ the hierarchical model to test variables.

MULTIVARIATE ANALYSIS OF VARIANCE BY R. DARRELL BOCK. CHAPTER 11

Univariate and multivariate analysis are methods for detecting and estimating, in sample data, differences between the means of populations. The populations may be naturally occurring and defined by attributes, or they may be created artificially by random assignment to experimental treatments.

It is both a strength and weakness of analysis of variance that it makes simplifying assumptions about the statistical structure of the data to be analyzed. It assumes that the variables under investigation are measured on a continuum with a uniform unit of scale; that the distributions of these measures in the populations differ only in the location of their central tendency and not in other aspects of shape such as dispersion, skewness, kurtosis, etc. For certain inferential purposes, it is in fact assumed that the distributions are normally distributed with unknown and possibly different means and unknown but constant variance. The strength of these assumptions is that they focus on the aspect of the distribution that is likely to be most sensitive to conditions of treatment or environment to which biological material might be exposed.

In most biological and behavioral studies, it is not possible to make observations under widely differing conditions without endangering the integrity of the organisms. As a result, most investigations deal with relatively small and essentially linear effects of different treatments or environments. These differences are expressed almost entirely in changes in the means of the distributions. By concentrating the inference on differences between means, the analysis of variance most effectively uses the information in the data to detect treatment or environment effects.

Like univariate analysis of variance (ANOVA), multivariate analysis of variance (MANOVA) focuses on means of continuously distributed variables, but, unlike ANOVA, does so jointly for more than one such variable. MANOVA is therefore especially suited to human behavioral studies, which typically involve a number of qualitatively distinct attributes or outcomes and for which no single index of value may be calculated. In the multivariate approach, the several variables are analyzed simultaneously, and the investigator or reader may decide for himself the overall meaning or importance of various differences that may be found.

Statistical methods allied to multivariate analysis of variance, and often included in the computer programs for the procedure, are the multivariate techniques of discriminant analysis, analysis of covariance, regression analysis and canonical correlation.

DISCRIMINANT ANALYSIS BY MAURICE M. TATSUOKA. CHAPTER 12

Broadly conceived, discriminant analysis is a system of multivariate statistical techniques that provides an integrated approach to the solution of three distinct but interrelated problems: (1) to determine whether or not significant differences exist among two or more groups of individuals in terms of several descriptor variables (Significance Testing); (2) if such differences exist, to try to "explain" them in terms of smaller numbers of "underlying factors" than the original descriptor (Explanation of Group Differences); and (3) to utilize the multivariate information from the samples studied in assigning a future individual to one of several groups studied (Classification).

As the first problem is precisely that addressed by multivariate analysis of variance (MANOVA) in its simplest form, discriminant analysis is often characterized as a follow-up or adjunct to MANOVA focusing on problem two--the explanation of group differences. This aspect in turn bears a certain resemblance to factor analysis, but factor analysis seeks to explain individual difference on a large number of attributes in terms of a small number of factors, while discriminant analysis seeks to do this for group differences. Whenever multiple criterion variables are used, MANOVA is the appropriate method for significance testing; explaining group differences parsimoniously is all but unique to discriminant analysis.

Historically, discriminant analysis has been associated with the problem of classification. It is probably the most important aspect in practical applications such as early detection of potential drug abusers with a view to offering them counseling and preventive treatment. It is necessary only to compute the discriminant function score for the individual to be classified (that is, the person of uncertain group membership, but who is known to be a member of one or the other of two groups), and then determine to which of the two group means on the discriminant function the individual's score is closer on the standardized scale.

The first phase requires one to look for the linear combination (i.e., a weighted sum) of the original variables such that the F-ratio for testing the significance of the differences among the several groups' mean on this linear combination is larger than that for any other linear combination of the original variables. To determine the weight for predictors that give rise to the largest possible value on the F-ratio is the task of discriminant analysis. The ideal situation is when the descriptor variables follow a multivariate normal distribution in each group. Furthermore, the mathematical model for the significance testing phase requires that the population covariance matrices of all groups be identical. The second phase does not require any distributional or equality-of-covariance-matrices assumptions. In classification, the multivariate normality assumption again becomes important if the numerical values of the likelihoods or probabilities of membership in the various groups are to be taken seriously. The equality-of-covariance-matrices assumption is not quite as crucial. Missing data always pose a problem when many variables are involved, and, of course, any method for supplying missing data is applicable only in the first two aspects of discriminant analysis. In the third phase, no individual with any missing data should be considered for classification.

As most empirical studies concerning drug abuse involve a comparison between users and nonusers, or among users of different types of drugs, in terms of demographic and/or personality variables, discriminant analysis could be used as one of their analytic tools. In reality, very few drug abuse studies seem to have employed this technique.

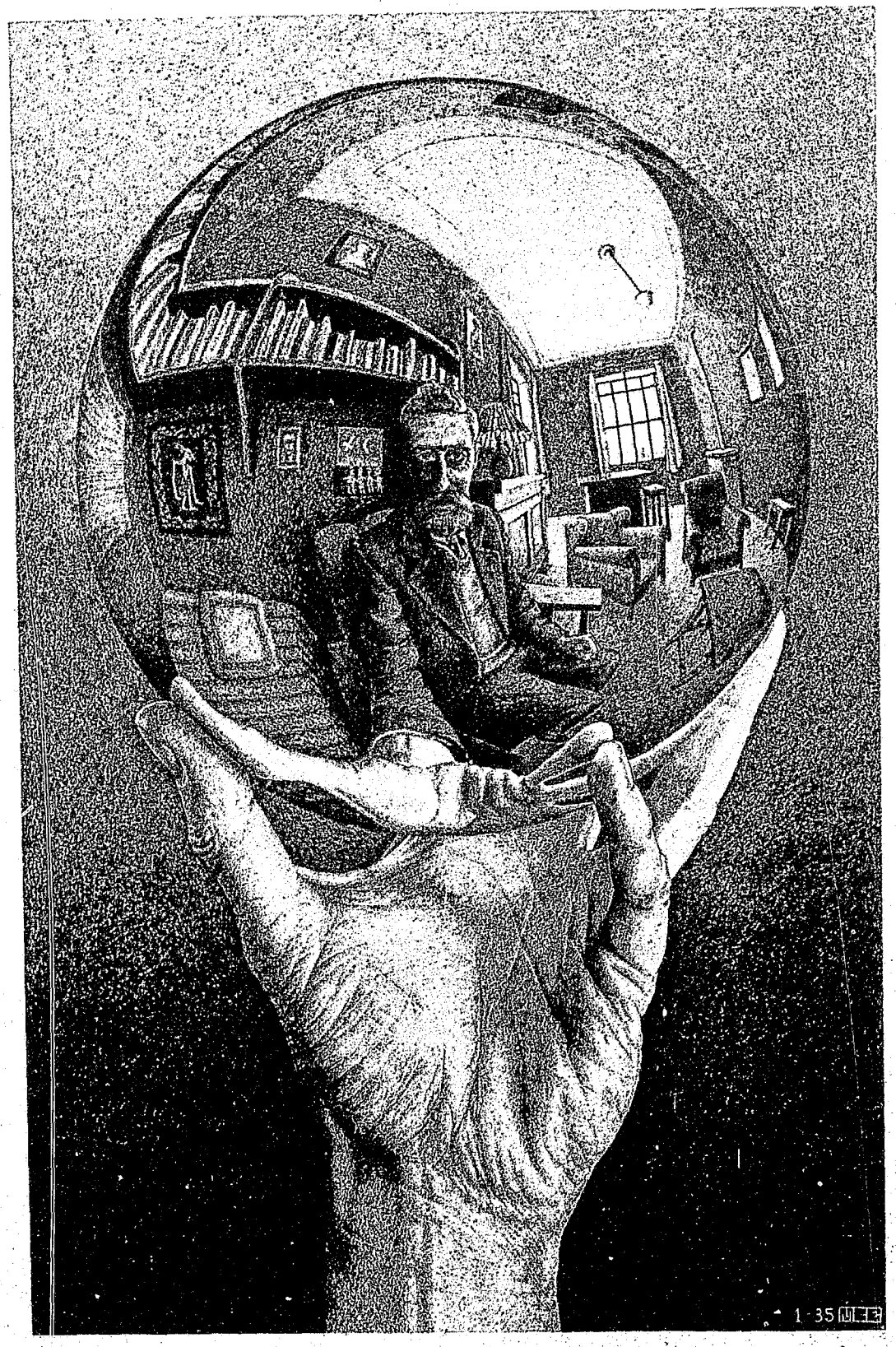


Plate 3

INTRODUCTION AND RATIONALE: SINGLE-ORGANISM VERSUS CONVENTIONAL RESEARCH

In recent years, research on single organisms or individuals has been advocated, directly or indirectly, by many authorities.[1] Dukes (1965) noted that the history of psychology contains a long and diversified list of influential studies of this type. He found 246 reports of research on single organisms, published between 1940 and 1965.[2] Indeed, the problems investigated and the research designs employed in single-organism studies are so varied that the situation may seem too chaotic to permit systematic integration. A common, but not very productive way out is to assert that research on single organisms cannot be used to test general laws of behavior and therefore need not be taken seriously.

The conventional approach to research in psychology is to study large groups of organisms, either under laboratory conditions that are deliberately simplified to approximate the ideal of univariate design, or under more complex conditions that permit elaborate statistical analyses of multivariate data. Examples of the latter are provided in most of the other chapters in this volume. The goal of such research is understood to be the discovery of laws that govern average behavior. Application of these laws to solve practical problems is not regarded as the primary purpose of the classical scientific enterprise. However, practitioners are permitted to apply the laws that scientists discover in order to understand, predict, and control the behavior of individual organisms.

A possible weakness in this approach is its assumption of isomorphism between the group and the individual. Not only practical experience, but also experimental data and mathematical logic show that this assumption is often unjustified (Bakan, 1954, 1955; Sidman, 1952). Statisticians might argue that the basis for failure of individual data to conform to group-derived functions is that the former contain larger components of error, and often that may be the case. Sometimes, however, the failure can be traced to the fact that entirely different functional relations apply to group data than apply to data from individuals. As a simple example, suppose that many organisms perform the same task, and that the average remains stable because the performances of half the group improve from practice while the performances of the other half deteriorate from fatigue. The statement, that performance on this task remains stable and is therefore unaffected by either practice or fatigue, is clearly untrue, whether that statement is applied to the group or to the individuals who compose it.

As a strategy for collecting data that will lead to the discovery of principles of behavior, the single-organism approach has much to recommend it. Properly employed, it is at least as demanding as large sample methods, and it usually leaves less to chance. Also, it requires that a distinction be recognized which both complicates and clarifies the research enterprise. The distinction is between the validity and the generality of a psychological principle.

VALIDITY AND GENERALITY

Single-organism research does not assume that a valid principle must apply to all organisms, but that it may apply to one individual, to only a few organisms, or to everyone. This assumption threatens to provoke knotty philosophical or methodological arguments about meaning in science; however, these are of no great concern, for the knots are easily cut. The liberating stroke is the realization that the limits of generalizability of a law or the description of conditions under which a principle applies are matters that are better resolved by systematic empirical investigation than by logical dispute.

The following diagram (Fig. 1) systematizes some characteristics of the single-organism approach. Both conventional methods and single-organism strategies are capable of increasing the store of valid principles (solid vertical arrow in center). However, the conventional approach (dashed vertical arrow on right) advocates attacking the universal by seeking general laws from the outset. It tends to reject the study of individuals as a romantic, or at best suggestive, exercise that may produce hypotheses but that cannot test laws (Holt, 1962). Conventional psychology maintains that valid universal (i.e., general) principles apply to individual cases, but the conventional scientist does not undertake the task of application himself. That is regarded as a job for technology or engineering, and it is left to practitioners. Consequently, in the diagram, no horizontal connection is indicated between the conventional approach and statements about individuals.

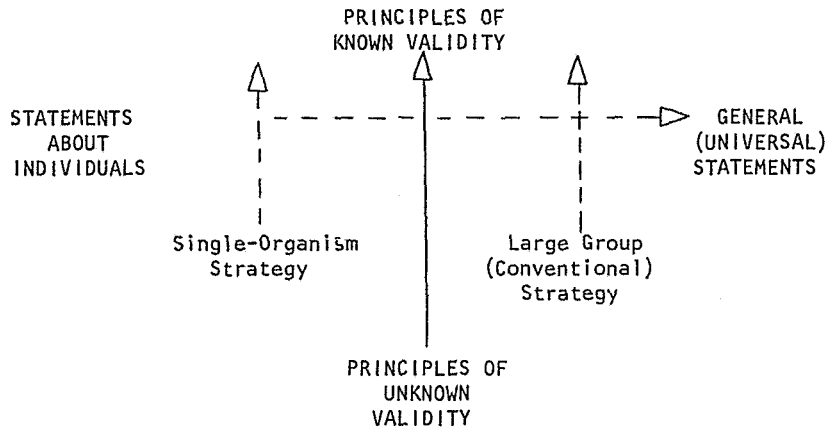


Figure 1. Comparison of single-organism and conventional research strategies.

Single-organism strategy (dashed vertical arrow on left) begins by studying individuals in order to generate principles for explaining the behavior of each. The validity of these principles is evaluated by judging how carefully and correctly each individual is studied, not how appropriate the laws of one organism's behavior are for others.

Generality is a separate question. It is a matter for investigation by replication, so that the limits of or conditions for validity may eventually be completely specified. Therefore, in the diagram, a horizontal (dashed) arrow, pointing right, is included to show that study of the particular is not the last step in the process but should lead to expansion of knowledge of the universal, through systematically developed generalizations.

CAUTIONS: WHAT SINGLE-ORGANISM RESEARCH IS NOT

Single-organism strategy does not advocate publication of research results every time a principle is discovered for a particular individual; such findings are often of only limited interest. Consequently, adoption of single-organism strategy need not flood the already overburdened scientific literature with a tidal wave of case studies. Publication is warranted only to describe procedural innovations of general interest or to present findings based on data collected from a sufficient number of organisms that have been examined, under sufficiently well controlled conditions, to justify the belief that reasonably generalizable statements can be made. This is not a simple matter but one requiring sound editorial judgment.

Neither does the single-organism approach advocate slipshod methodology, though it recognizes that tightly controlled experiments and loosely designed exploratory investigations are both useful in their own ways. Pioneering investigations often cannot be rigidly controlled, no

matter how many organisms are studied. Research of this sort is no more expected to produce valid statements of cause-effect relations in the single-organism approach than it is in investigations of large groups.

Single-organism strategy is also not anti-statistical, though some investigators advocate certain forms which are (Bakan, 1966, 1968; Sidman, 1960). Subsequent sections of this chapter show that highly sophisticated quantitative techniques, invoiving time-series analysis, mixed analysis of variance designs, and certain factor analytic procedures can often prove fruitful. The fact that such techniques are not often employed in single-organism research means only that, despite its potential, this strategy has not yet become popular with investigators who know how to use the techniques.

METHODS AND PROCEDURES: THREE TYPES OF DESIGNS

Methods of single-organism research are best illustrated by dividing them into three groups; these may be arranged along a continuum, according to the degree of control over experimental conditions exerted by the investigator. At one extreme are studies that adhere closely to the classical model of experimentation, which requires systematic manipulation of only a few independent variables in a closely controlled and constant laboratory environment. To illustrate these, examples from the literature on operant conditioning have been selected. In the mid-range are natural process research studies, some of which may use quasi-experimental designs (Campbell, 1969; Campbell and Stanley, 1963) and require time-series analysis or its equivalent for the evaluation of findings. At the other extreme are formal case studies, which are noninterfering but are nonetheless designed to be as objective and as explicit about procedures as possible. The representative case method, which stresses careful selection of each participant and the use of quantitative data, has been chosen to illustrate these. Informal case studies and anecdotal methods (with which single-organism research has been too often identified in the past) are not considered in detail in this chapter because they contribute little of a systematic nature to scientific knowledge.

An important feature of single-organism research at all levels of control is the prominence of investigations with a practical or therapeutic orientation. Typically, single-organism strategy does not stress the distinction between basic and applied science, which seems so important in conventional psychology. A principle that produces favorable change in an individual is just as valid and may be just as universal as a principle that applies to behavior that is apparently unrelated to problems of personal adjustment.

EXPERIMENTS WITH SINGLE ORGANISMS

Sidman (1960) made the strongest and most elaborate case for controlled experimentation on individuals. The approach he recommended requires tight control of all conditions that might affect outcomes, relatively simple operationally defined variables that can be manipulated or measured automatically, rapid output of results, and a succession of chained and logically interconnected investigations, each derived from the ones that have been already completed. The goal of a research program that follows these recommendations is to reduce variability of outcomes by the functional manipulation of the conditions under which they are produced. Single organisms are preferred because their use eliminates a major source of variability (individual differences) at the outset. Statistical evaluations are rejected because it is argued that they conceal variabilities which should not be ignored or regarded as error but brought under experimental control. Replication of investigations with individuals is recommended in preference to replication with groups, for a truly universal (general) law must apply not just to group averages but to every appropriate experimental subject.

The Principle of Reversibility

The essence of single-organism experimentation is reversibility of behavior effects. The experimenter tries to demonstrate that a certain behavior (b) appears only under a certain set of environmental conditions (B). To show this, the investigation may employ a simple ABA design. During the first administration of condition A, the experimenter records the base line or base rate of occurrence of b. Then, during the administration of condition B, the experimenter records changes, if any, in the rate of appearance of b. Following this, condition A is restored to show that b obediently returns to its base line level.

To overcome objections that behaviors other than b may also be affected by the experimental conditions, the investigator may establish multiple base lines. That is, behaviors c, d, and e may be measured throughout the experiment to show that they do not respond to changes in the environmental conditions. To overcome objections that the results may have been due to coincidence, the AB sequence may be repeated: ABABA. Or, the introduction of B may be randomized in a series of many trials, so that the organism cannot learn a regular sequence of events and must respond only to the independent variable.

Not all experimentally induced effects are readily reversible, but apparent irreversibility does not necessarily invalidate the method. For irreversibility implies a change of some kind in determining conditions (e.g., a change in habit patterns or in organization of neural pathways), and if these could be systematically manipulated (by introducing another appropriate condition), the behavior could be shown to return to its base line level again.

Illustrative Applications

Animal Research. A relatively simple illustration of how animal research may use single organisms is the procedure described by Sidman (1960), which was summarized later by Bachrach (1962) who presented it as an example of good experimental design. A single rat is placed in a compartment where it is given a brief electric shock unless it presses a lever. Pressing the lever delays the next shock for 20 seconds (avoidance conditioning; fixed interval schedule). Eventually, the rat learns to press the lever at a fairly constant rate and thereby avoids most shocks. The animal now retains a steady rate of lever pressing for about six hours.

If the experimenter is interested in learning per se, the rat's pre-shock level of bar pressing can be used as a base rate for evaluating post-shock levels of performance. However, the steady, learned performance may later become a base rate for evaluating the effects of other conditions, such as drugs. If amphetamine sulphate is administered after learning has stabilized, the rat's behavior shows a smooth acceleration in lever presses. The animal eventually reaches a level of performance at perhaps three to four times the base rate, where it stays for two to three hours. Then, its performance begins to slow down until it reaches a level below its original base line, where it stays for several hours.

A host of variations is possible within this paradigm. Most obvious are the possibilities of varying the type of behavior learned, the type of reinforcement schedule imposed, the types and dosages of drugs, and the types of organisms tested. If a more complex task were used, or if more measures were taken (heart rate, temperature, eyeblinks, etc.) multiple base lines could be established and differential effects of drugs on specific aspects of behavior could be evaluated. Still other possibilities would be to introduce several administrations of the same drug at randomly selected points in time and to include administrations of a neutral substance, or placebo, as well. Naturally, the experimenter should not know what substance is being administered on any trial. Such procedural refinements can eliminate objections that drug administration itself occurs on a fixed interval schedule or that the organism's behavior is under the control of cues systematically provided by the experimenter.

Therapeutic Research. The literature of behaviorism bulges with reports of therapeutic experiments on single organisms. Many of these experiments seem to be reported primarily to demonstrate the efficacy of operant techniques. In part, they serve the same functions as do testimonials at revival meetings. They provide reinforcement to those who are already true believers, they discourage backsliding, and perhaps they even inspire a few converts. However, they serve other more important purposes as well, for they provide explicit working models of how to manage every step of the therapeutic process, from assessment through outcome evaluation. In addition, they provide a means by which practitioners may exchange ideas and make suggestions to each other for improving procedures. (See, as examples, the report of a symposium on behavior modification in clinical psychology, edited by Neuringer and Michael, 1970, and Bandura's description of the uses of a variety of behavior modification techniques, including operant methods, in applied settings, 1969.)

On the level of applied research, Bandura (1969) pointed out that many treatment programs that use aversive drugs, like Antabuse, fail because the effects of the drugs are delayed too long after the behavior to be eliminated (e.g., alcohol ingestion) occurs. Because reinforcement must be immediate to be effective, the behavior one is trying to eliminate remains attractive to the patient, despite its adverse effects. Bandura also noted that recent evidence challenges both the assumption that alcohol addiction results from the reinforcement provided by stress reduction (Lester, 1961) and the assumption that total abstinence is the only feasible goal of

therapy (see also Lloyd and Salzberg, 1975). These conclusions are mentioned here because at some future time they may be found to apply to other drugs as well.

Several investigators have developed individualized treatment programs for alcoholics (examples are provided in Miller, 1972, and in Sobell and Sobell, 1973). A good illustration of therapeutic research on a single person is the study by Cohen, Liebson, and Faillace (1971) of reinforcement contingencies in chronic alcoholism. This study took a total of nine months and consisted of six experiments conducted on the alcohol research ward of a hospital. The patient's participation was voluntary, but he was paid in money only if he completed each experiment. The basic research plan was to make certain privileges (such as working for pay, using the telephone, eating the regular diet, having reading material, using the recreation room) contingent upon the patient's drinking less than a specified amount per day of 95 proof ethanol.

The first experiment lasted four weeks. During the first and third weeks, privileges were not granted at all and were thus independent of alcohol intake. During the second and fourth weeks, contingency conditions were imposed. That is, the patient obtained privileges only if he drank at a moderate rate (no more than 5 oz. per day). In this experiment, the patient went on drinking the maximum amount possible (10 oz. per day) regardless of conditions.

In the second experiment, the penalty for overdrinking was made more severe. In particular, this involved exchanging the regular diet for a pureed diet, removing reading material, removing the bedside chair, and extending the contingent deprivation period (the period of time deprivation was imposed if the patient overdrank during the contingency phases) from a variable period to a full 24 hours. The effectiveness of this regimen was amply demonstrated by the fact that, in two weeks of noncontingency, the patient consumed 10 oz. of 95 proof ethanol per day, while in three alternating weeks of contingency, he never once drank more than 5 ounces per day.

In the third experiment, the patient was allowed as much as 24 ounces of ethanol per day, but results remained therapeutically favorable. In the fourth experiment, the periods of noncontingency were enriched, to reduce the contrast between contingency and noncontingency conditions. Under the noncontingency condition, the patient was, in effect, allowed to go on binges without serious consequences for five days. The results continued to be stable, but a question arose as to whether the patient was merely "being good" during contingency periods for the reward of being able to go on binges the rest of the time.

In the fifth experiment, binges were eliminated from the noncontingent weeks by allowing the patient to drink only every other day. During this experiment, his responsiveness to the experimental contingencies broke down seriously. He overdrank on 5 out of 13 days.

Finally, in the sixth experiment, contingency conditions were imposed for five weeks in a row. During this time, the patient overdrank only twice. Thus it was established that moderate drinking can be maintained, if the environment is suitably controlled to provide and consistently apply appropriate contingent reinforcements. Perhaps the amount of time it takes to make such contingencies effective, and the expense involved in maintaining them, are so great as to make the whole idea economically unfeasible; but that is another issue.

Effectiveness of drugs. Bellak and Chassan (1964) reported a study that evaluated the effectiveness of a psychiatric drug, chlordiazepoxide, on eight variables in the behavior of a single patient. The study adhered to the usual precautions of double blind research: it incorporated a placebo at treatment points, which were not identified to either the investigator or the patient. Six administrations of the drug and four administrations of the placebo were included in the experimental design. By inspection, the data showed clear tendencies for the patient's behavior to improve when she was taking chlordiazepoxide and to become worse during periods when she was taking an identical appearing placebo. Bellak and Chassen recognized that a particular kind of statistic, interrupted time-series analysis, is required to provide appropriate quantitative analyses of their results. However, they did not evaluate their results by this procedure. Time-series analysis is outlined in the following section.

NATURAL PROCESS RESEARCH

All studies described up to this point utilize designs in which the experimenter controls the time of application and the intensity of the independent variable. However, single-organism research is often conducted in settings where events occur that may be thought of as altering the level of an independent variable, but that cannot be controlled by the experimenter. Examples

are the passage of new laws or the institution of social reforms (Campbell, 1969), administrative or therapeutic decisions in clinical settings, or fortuitous circumstances such as tornadoes or the winning of a lottery. Shontz (1965) identified studies of such phenomena as natural process research. When a degree of control can be exerted over measurements or over decision-making, such studies achieve the status of quasi experiments (Campbell, 1963, 1969; Campbell and Stanley, 1963).

An example of natural process research, conducted on a single person, is afforded by E.J. Murray's (1954) ratings of a patient's expressions of hostility and defensiveness over the course of 17 hours in psychotherapy. In this research, no systematic attempts were made to influence the therapist's activities. A reciprocal relationship was observed between ratings of the two types of behaviors; when one was high, the other was low. Furthermore, type of defense (intellectual defenses or physical complaints) was found to be related to the timing and type of interpretations offered by the therapist. For example, defenses tended to decrease following a punitive interpretation, but hostility and a subsequent return of defensiveness quickly followed.

Time-Series Analysis

Neither the research by Bellak and Chassan nor the study by Murray actually employed interrupted time-series analyses to quantify their results. Indeed, the requirements of this analytical technique are such as to make its application difficult in these particular investigations. As is true of all research, time-series studies are most effective when planned well in advance.

Time-series analysis is relatively new to psychology (Box and Jenkins, 1970; Glass, Willson, and Gottman, 1975; Gottman, McFall and Barnett, 1969; Harris, 1963; Holtzman, 1963; Jones, Crowell, and Kapuniai, 1969; Wold, 1965), and it offers many possibilities, especially for research on drugs and for natural process research on single persons. (For discussions of some other more or less closely related statistical approaches to data from single organisms see Chassan, 1960, 1961, 1965, 1967; Edgington, 1967; Luborsky, 1953; Shapiro, 1961a; Shontz, 1972; Stephenson, 1953; Wold, 1965.)

Score Dependencies. Conventional tests of statistical significance typically require that measures be independent. However, time-series analysis recognizes that when data are collected from the same organism over time, troublesome dependencies are introduced. For example, in a succession of scores that gradually increase in value, it is immediately apparent that later values are not independent of earlier scores. Anyone who knows the rate at which values are increasing in this series is in a position to anticipate later scores at a better-than-chance level. The higher the autocorrelation within a series, the greater is the dependence of later measures upon earlier measures in the series, and the less justified is the assertion that a later value or mean of values within the series is a random deviation from those that occur before it. (Changes in reliabilities of measures may also affect statistical judgments, but these are not considered in detail here.)

If a series were steadily increasing, an unknowing investigator might test a group of persons once at time t and once at time t' and find a mean difference large enough to permit rejection of the null hypothesis, on the basis of the assumption of independence of scores. However, in such a case, rejection of the null hypothesis is clearly inappropriate.

The same possibility exists in research on single organisms. A person may be on a dietary regimen that causes gradual loss of weight. If an investigator measured this person's weight one week before and one week after the experimental drug is administered, the investigator might conclude that the drug induced the weight loss. Fortunately, single-organism research is less likely to be subject to this type of error. Partly, that is because an investigator is likely to know more about the single person he studies than he would about individual members of a large group. Consequently, information about possible contaminating factors is more likely to be available in single-organism research. Also, and more importantly, the requirement of reversibility, especially if combined with multiple base line measures and with multiple, random presentations of the drug, may go a long way to obviate most confounding serial effects in single-organism studies. When a high degree of control is not possible, statistical adjustment of the data may be in order, and time-series analysis is clearly the technique of choice in such instances.

Types Of Changes Evaluated. In general, two types of changes are evaluated in interrupted time-series designs. One type is change in level; the other is change in direction. A meaningful change in level is, of course, one that could not have been anticipated on the basis of knowledge of the values of preceding observations. Changes in direction can be more complex.

For example, if a patient's anxiety was increasing until treatment began and then shifted direction to a steady plateau or even to a slightly less rapidly increasing trend, a change in direction but not in level might appear. Computer programs, described in a subsequent section, have been developed to test for both types of change.

Models. The first task in the analysis of time-series data is to establish which of several models best fits the data. Only the sketchiest notion of how models are identified can be provided here. Glass, Willson, and Gottman (1975) provide necessary details.

Two basic types of models are available: moving averages and autoregressive. An essential tool for deciding which type of model is appropriate is the correlogram: a series of autocorrelation coefficients. The lag 1 autocorrelation coefficient measures the correlation within a series when values at all data points are paired with the values at the data points immediately preceding them. The lag 2 autocorrelation coefficient increases the distance between paired data points by one interval; the lag 3 autocorrelation coefficient increases the distance by one more interval, and so on. A correlogram contains the array of autocorrelation values from lag 1 to lag k and the pattern of these values is usually diagnostic of the type of model that is appropriate.

According to Gottman (1973), the most practical model in psychological research is the integrated moving averages model of the first order. When this type of model is appropriate, the autocorrelation value is nonzero at lag 1 but drops immediately to zero thereafter. If autocorrelations are nonzero at lags 1 and 2 but zero thereafter, the model is of order 2, and so on.

Sometimes it is necessary to obtain autocorrelations of differences between values at predesignated pairs of data points. First order differencing subtracts the value at each data point from the one immediately following it. Second order differencing increases the interval by one data point, and so on. Differences are then lagged and autocorrelation coefficients determined. Differencing is necessary when a series is nonstationary, that is, when it does not stay at a steady mean level. Correlograms of nonstationary series contain unwanted correlations that make model identification difficult until after differencing has been performed. Sometimes more complex seasonal adjustments are also required to remove natural, but irrelevant, cycles from the series.

Autoregressive models are more difficult to identify than moving averages models. In general, the correlogram for an autoregressive model shows a gradual rather than a sudden decrease in autocorrelation values (after necessary differencing has been performed). However, when an autoregressive model is appropriate, a sudden decrease is shown in another set of values, the partial autocorrelation function. The order of the autoregressive component is specified by the size of the lag before which this drop occurs.

Further complications are added by the fact that a model may be of both the moving averages and autoregressive types and of different orders for each. In this case, both autocorrelations and partial autocorrelations are found to die out slowly. Still another possibility is that the process studied may itself change, thus requiring identification of different models at different stages in the series. Naturally, more complex models are more difficult to identify accurately.

Statistical Analysis. The final stage in the analysis of interrupted time-series is to perform desired statistical evaluations. This requires estimating optimal numerical values for parameters associated with the type of model identified. Typically, this is performed by a search technique. That is, successive parameter values are tried, and by inspection the one is chosen which produces minimum error variance. Statistical tests, based on these parameters, are interpreted in the usual way to decide whether significant changes in level or direction (or both) have occurred.

Significance Testing. Time-series analysis does not always require large mean differences for statistical significance. If measures are steadily increasing, even a small reduction in the mean of a group of measures, taken after experimental intervention, may be significant. Or, in the extreme, perhaps no mean change may deviate significantly from expectations. A great deal depends upon the characteristics of the autocorrelation function. When autocorrelations are zero, a before-after test of mean differences is identical to the standard t-test.

Computers/Software. What has been said about the quantitative aspects of time-series analysis in the preceding paragraphs may not be entirely clear. However, two points should be evident nonetheless. The first is that no investigator can expect to perform time-series analyses on a hand calculator. Access to large computers and to appropriate software is essential. The

second is that, even under the most favorable present circumstances, time-series analyses of the more sophisticated varieties are not "cook book" operations. Especially when data quantity and quality are not high, model identification is something of an art rather than a simple, automatic process. Unless and until that situation changes, most psychological investigators should employ an expert consultant if they intend to use time-series designs.

Number of Observations. One very practical problem remains. The accuracy of determination of a model depends upon the number of observations available, and no specific rules have been developed on this point. It is known that an interrupted time-series design uses its statistics most efficiently if the experimental intervention occurs half way through the observational series. If the data normally contain complex cyclical trends (seasonal variations), a large number of observations may be required to establish the conditions necessary for removing them and performing an adequate statistical test of induced changes. Simpler processes require fewer observations. Box and Jenkins (1970) recommended at least fifty observations to provide a useful estimate of the autocorrelation function. Glass, Willson, and Gottman (1975) agreed but pointed out that, while well-behaved data may be identified in 35 or 40 observations, data requiring seasonal adjustment will require many more than 50, at least enough to cover four or five cyles.

A study by Jones, Crowell and Kapuniai (1969) used only four prestimulus values as a base line for testing the effects of visual and auditory stimulation on the heart rates of infants. By contrast, a study by Holtzman (1963; summarized by Glass, Willson, and Gottman, 1975) measured a single patient's perceptual speed under base line conditions once a day for sixty days. Appropriate changes were shown to occur when the patient was placed under treatment with a psychiatric drug on the 61st day; electroshock treatment was added on the 121st day; and base line conditions were restored on the 181st day for the final 60 days of the investigation.

Complex Design Possibilities: Time-Series Designs

Holtzman's study, outlined above, is called a single-organism, multiple intervention design. It has many advantages, especially for evaluating treatment effectiveness in clinical observations, but it is not without problems of inference. Notice, for instance, that this experiment does not reveal whether a change that might occur following the introduction of electroshock is due to the electroshock alone or to a synergetic combination of electroshock and drug.

Obviously, the particular design Holzman used is not the only one possible. In fact, once one opens up the possibility for more than one interruption of a time-series and for tying several single-organism, time-series studies into an overall program of investigation, the number of research designs that could be developed staggers the imagination. Using a single intervention, or a simple reversal design, several organisms can be subjected to the same intervention at different times, with the data for each organism being analyzed separately, but the results accumulated for all participants; Gottman (1973) calls this "N-of-one-at-a-time" research (crediting Alexander Buchwald and Steven Shmurak for the term). To test the effects of two types of intervention and their interaction, the interventions may be introduced separately and in combination into the time-series; this requires a minimum of three interventions, with returns to base line intervals in between. Only a little ingenuity is required to expand the design possibilities for interrupted time-series experiments almost without limit.

Three final features of time-series designs deserve mention because they extend the potential of the method even further. The first is that concomitant variation may be evaluated among time-series. That is, one series may be used to predict another (to serve as a lead indicator). Also, in quasi experiments, covariates may be deliberately introduced to help adjust for certain forms of bias that may appear in complex natural process research. Finally, an intervention effect may be evaluated on the hypothesis that it is a one-time occurrence, or on the hypothesis that its effects are constant over many measurement periods, or on the hypothesis that its effects continue according to some specified mathematical function, e.g., a decay curve. Provisions for specifying hypotheses of this type are allowed in some computer programs now available.

Computer Programs and Resources

Two computer programs that proved very useful in a study of mood changes accompanying menstruation (O'Connell, 1975) are CORREL, which produces the correlograms (autocorrelations and partial autocorrelations for data in both differenced and undifferenced forms) needed for model identification, and TSX, which provides t-tests of level and direction for interrupted time-

series with specified model characteristics. Both programs are available in Bower, Padia, and Glass (1974) and should be satisfactory for most experimental purposes in psychology. In addition to information that appears in published articles, cited in the reference list of this article, Gottman (1973) cited two sources of computer programs that might be valuable to some users.[3]

REPRESENTATIVE CASE METHOD

As originally described, the representative case method included virtually all types of single-organism research (Shontz, 1965). However, developments over the past ten years indicate that a somewhat more precise application of the term is possible. Within the overall approach that studies single organisms, methodological subcategories have been emerging that merit separate consideration. Examples are operant methods, natural process designs (described in preceding sections), and informal case studies (dealt with only in passing in subsequent discussions).

As in conventional research, representative case studies may employ experimental manipulations of independent variables under controlled observational conditions, particularly if the studies test propositions about cause-effect relations. However, in conventional experiments, organisms are treated only as objects that are expected to contribute to the research by reacting passively to conditions, defined by a problem the investigator chooses. In representative case experiments, the person is an active, cooperative participant in the research process. The person's individuality is not forsaken (measures of dependent variables may be tailor-made to suit his or her individual modes of expression); the person is not deceived about the purposes of the study; and every opportunity that time and procedures can allow is provided for the person to comment on the validity of the data and (where possible) of the investigator's conclusions.

Representative case research may be purely descriptive. However, when it is so, it must be distinguished from informal case studies, which are not embedded in a systematic research program, which do not use explicit data collection methods, and which describe cases either because they are merely "interesting" or "rare," or because they provide material for demonstrating techniques of diagnosis or treatment (Neale and Liebert, 1973). As valuable as informal case studies may be for some purposes, they do not contribute as much as is possible under more carefully regulated conditions of data collection and replication. Furthermore, their lack of control has generally given single-organism research the poor reputation it now endures.

Cautions

Identification of Problems. Representative case research has three essential requirements. First, an appropriate problem must be clearly identified. On this matter, representative case research does not differ from any other form of scientific investigation. However, the representative case method is better suited to some types of problems than to others. For example, public opinion regarding a proposed change in tax laws would not be most efficiently assayed by studying single individuals intensively; conventional techniques, for polling samples drawn from large populations of persons, are clearly preferable. However, an attempt to discover the sources or implications for particular individuals of strongly held political beliefs would definitely proceed best through the intensive study of selected persons, known to hold relevant commitments. Indeed, research purposes of this type are often best served not by studying typical persons but by studying extreme cases.

This argument was suggested by William James as long ago as 1902 in his book, Varieties of Religious Experience (James, 1902/1958). He recognized that, in the mid-range of some complex dimensions, like religiosity, considerable disagreement may exist about the meaning of a term; but at the extremes disagreement vanishes. James argued that, if one studies a recognized saint, one is clearly studying a religious person, and what is learned from that study applies, to some degree, to the religiosity in us all. The saint is not chosen to typify the average of a population of people who vary in religious strength, but to represent religiosity in its most obvious, most researchable form. The saint provides the scientist with a view that enlarges for closer inspection a component of all people that is normally too obscure or too undeveloped to be clearly examined.

It follows that problems which are especially well suited to representative case research are those that can be identified with variables which display themselves in readily recognized behavioral, affective, cognitive, or physical states.

Selection of the Case. Second, as noted above, someone (the representative case) must be deliberately selected who has the properties that make that person a particularly suitable candidate for investigation. For example, an investigator who wishes to study anxiety might wish to select for study the most obviously anxious individual who can be found.

Nothing is more crucial than the careful selection of appropriate persons for study. For in representative case research the person must represent in clear, if necessary in exaggerated, form the exact condition that is under investigation. If this requirement is not fulfilled, subsequent findings will not be relevant.

Procedures. Third, the person selected for study must be examined by techniques appropriate to the problem at hand. Here again, the representative case method does not differ in principle from any other in science. But a few words of qualification are in order. Not every research that studies a single person is necessarily a representative case research.

Many case studies fall short because the investigator failed to ask clear questions before descriptive material was collected. Thus, the relation between the case described and the relevant theoretical or practical problem to which the research addresses itself is established after the fact, rather than before. Such case studies may raise questions, but they cannot answer them.

Other case studies fail because they lack objectivity and explicitness of procedure and data. In part, this failure may also be a consequence of failure to ask a clear question. If a problem is not clearly identified, the investigator cannot proceed in a systematic way to solve it. Often, the scientist remembers the case rather than selects it. Then, instead of finding another such case and examining it in a technically proficient way, he reports the one he recalls. Recollection is valuable, but it is not objective, and much of the current conventional bias against single subject research (e.g., Holt, 1962) probably stems, not from the small number of subjects it examines but from the unsystematic way it typically examines them.

These considerations should not be taken to suggest that procedural requirements are inflexible. Indeed, representative case research is ideally suited to the use of morphogenic measures (Allport, 1962). Such measures assess each individual in ways best suited to his or her own characteristics. Stephenson's Q-technique (1953) and Kelly's Rep Test (1955) are good examples of measurement formats that are well suited to morphogenic or semi-morphogenic measurement. It is even possible that the same variable (e.g., anxiety) would be indexed differently (heart rate, skin resistance, verbal report, etc.) in different studies, the appropriate index being selected on the basis of knowledge of how each representative person displays his or her inner state.

The Issue of Sample Size

The usual argument against research that studies individuals stems from conventional statistical doctrine. It asserts that a sample size of one is too small to support generalization. This argument is valid for research that regards persons as being sampled randomly from a large population. However, representative case research does not so regard its participants.

A close analogy to a psychological investigator using the method of the representative case is the chemist studying the properties of a specific substance. Although a geologist may concern himself with problems such as getting the best estimate of the average purity of samples, or determining the range of distribution of the substance in nature and the types of contaminants with which it is usually associated, the analytical chemist (i.e., the one who performs representative case research) prefers the purest supply of the substance he can get. He would rather have a gram of the compound, that is free of impurities, than a ton of ore straight from the mine.

Obviously, to achieve comparable purity in human research materials is nearly impossible. That is why many who use single-organism strategies prefer to study specially bred animals that can be developed to serve particular research needs. But, as every geologist knows, nature sometimes supplies small quantities of an unadulterated substance ready made, if only one knows where to look for it. That a close approximation to purity can sometimes be approached even in psychological research has already been indicated by the citation of James' early work on religiosity.

Statistical Analysis

A preceding discussion of time-series methods has shown that, for the most part, conventional t-tests and analysis of variance may be inappropriate in representative case research in which behavior from the same organism is measured many times. Of course, time-series techniques can be applied in any representative case research that supplies sufficient data. But such techniques are currently best suited to experiments, quasi experiments, or natural process studies, the outcomes of which can be indexed by a single value.

Because time-series analysis consists essentially of correlational procedures, multivariate longitudinal problems also can be handled by suitably elaborated time-series techniques. However, until time-series analyses reach the cook-book stage of development, the investigator might be well advised not to attempt more complex designs, unless he or she is prepared to develop the computer software necessary to solve associated problems of quantitative analysis.

Where multiple measures are used in the context of descriptive research (or even in some special instances that involve hypothesis testing), certain infrequently used factor analytic designs can be useful (Cattell, 1946; 1963). Suppose that a single individual is administered a battery of tests (all scored on the same scale) on several occasions. In P-design, factor analysis can be applied to correlations of scores from test to test (across occasions) in a way which is exactly analogous to conventional (R-type) factor analysis. This approach is often described as a way to validate factors, derived from large sample research, by replicating them in individuals (Cattell and Cross, 1952). Another use of P-design is illustrated by Lettieri's (1970) intensive study of four persons who differed from each other most essentially in degree of suicidal potential (High; Medium; Low; and Zero--i.e., this person was undergoing a personal crisis but was not suicidal). Four separate P-type factor analyses (one for each person) were performed on data from multiple measures taken over a period of 21 consecutive days following the onset of a crisis or a suicidal state. The extracted factors therefore represented psychological states that developed in persons at each level of risk. Lettieri found depression to be less prominent in the factor analysis of the person at high risk than in the analyses of data from persons at lower levels of risk. He attributed this to the likelihood that depression disappears once the decision to die is accepted as the solution to one's problems. High risk was also associated with dichotomous thinking (the tendency to think in terms of polarities such as bad or good, life or death) than was lower risk. Lettieri attributed this to suicidal persons' relative inability to accept partial solutions to life's problems.

In Q-design, correlation of scores from occasion to occasion (across tests) can be used to factor analyze occasions, i.e., to determine on which sets of testings the person produced similar patterns of test scores. At first glance, this may not seem a very exciting possibility. Suppose, however, that the "tests" were a series of items (perhaps as many as 60 or 80) describing psychological states (angry, sad, excited, etc.), and that the person's "scores" on these tests were ratings of each state, describing his or her feelings while taking a certain drug. On separate occasions, descriptions of feelings under several different drugs could be obtained, intercorrelated, and factor analyzed. This would produce quantitative descriptions of similarities among drugs (occasions), as measured by verbal reports of mental reactions. Hypotheses could be tested by specifying in advance how the investigator expects the substances to be grouped by this procedure. Examples of P and Q designs in research on drug usage are provided in a subsequent section of this chapter.

Illustrative Application

Treatment Outcome Evaluation. Outcome research is often hampered by the fact that uniform application of a single evaluative standard to all subjects is unrealistic and unfair; behavior that defines success in one case may define failure in another. Shontz (1972) advocated an individualized method of outcome evaluation which uses analysis of variance on data from single persons. In this method, individualized goals are established for each client. Statistical tests of before and after scores (or of change scores, or of after scores only, or of after scores adjusted for before scores, or of any other appropriate evaluative measure) provide the basis for evaluating outcomes in individual cases. In addition, the pooling of analyses with homogeneous estimates of error variances provides a basis for evaluating the treatment program as a whole. To illustrate the method, Shontz described studies of three children in the same rehabilitation agency. Each child was evaluated on items describing rehabilitation goals specified by the child's own parents, therapists, and teachers. Item contents described behaviors each judge thought should improve if rehabilitation were successful; 6 to 12 items were provided by each judge. The same persons who described goals for a child served as judges for that child. Each judge evaluated all items (goals) provided by all judges (including himself) and

described the child as he was when he entered the agency, as he is now, and as he probably will be when he leaves. Evaluations of the children were obtained as ratings on a 14-point scale from 0 (the behavior is impossible) to 13 (the behavior is easy). This scale provided a common metric and yielded data which could be treated statistically in the same way as would data from any standard rating instrument.

Because therapists and parents were not assigned randomly to clients, and because in this study goals were specified for each client according to his unique needs, judges and item sources as well as type of rating were fixed factors in this analysis of variance. Because the clients represented a larger population of children at the agency, this factor was treated as random. In this design the J x I x T within C interaction was the common estimate of error and all other terms were tested against it. The overall analysis of variance was calculated with all terms nested within clients. This procedure was virtually equivalent to calculating individual analyses for each child and pooling the separate analyses. Pooled analyses showed that raters were in overall agreement that rehabilitation was followed by significant overall improvement. However, individual analyses showed that only two of the three children were judged to have benefitted significantly. This approach to outcome evaluation has the merit of preserving client individuality and of promoting client participation (by having the client specify the goals he or she desires to achieve), and at the same time, preserving the capacity to assess treatment effectiveness on a client-by-client basis as well as in terms of aggregate results.

Drug Research. A particularly thorough plan to use representative case methods to study a clearly defined, though complex, problem is being applied in a project conducted under the auspices of the National Institute on Drug Abuse. This project began as a study of the life styles of nine persons who live in a large midwestern city and who use cocaine heavily. Data on this phase of the project have been collected and are currently being prepared for publication (Spotts and Shontz, in press). The second phase involves collecting comparable data from nine men who are similar to the cocaine users, except that these men are heavy users of amphetamines. Data collection for this phase of the program is currently under way. Plans call for additional research on users of opiates, barbiturates and alcohol, as well as on persons who use no drugs to excess.

What identifies this program as representative case research is the fact that each of its phases actually consists of nine separate studies. Participants in each phase were purposely not chosen as random samples of any population. Each was chosen for the value of the particular contribution he could make to the study as a whole. For example, participants were deliberately selected to represent modes of psychological adjustment that were as different as possible. One participant was a successful salesman, one was a professional thief, one was a millionaire's son, another was a pimp, and so on. The participants were not treated as "subjects," but as expert consultants who knew more about cocaine and its effects than the investigators (or almost anyone else). The intention behind this selection was to insure as broad a base as possible for knowledge about cocaine use and effects, employing a minimum number of persons in the process. All participants were well paid for their assistance.

Extensive data were collected from each participant. To provide general information, each was intensively interviewed about family history and psychosocial background. Each described the geneology and patterns of his drug use and provided data to describe in a standardized manner the physiological and psychological effects he had experienced from cocaine. All were administered standard psychological tests of intelligence and personality. Each participant then used the same sixty item Q-sort instrument to describe five personifications requested by the investigators. A sixty item Q-sort with individualized items was used to obtain nine more descriptions of specially requested personifications. Examples of descriptions (personifications) requested were: your usual self; yourself as you are when high on cocaine; your ideal self; the typical cocaine user; your bad self. Each participant also completed a special version of the Kelly Rep Test (grid form, Bannister and Mair, 1968; Kelly, 1955), as well as other special tests prepared for this investigation. All the Q-sorts and the Rep Tests were administered three times, at testing sessions at least one month apart.

Data from Q-sorts were collected in such a way as to fulfill the requirements of O-type factor analytic design. Each sorting provided by the participant within each testing session was identified as an occasion, and the resulting matrix of correlations for each participant reflected similarities among patterns of item values across all occasions.

Certain sorting instructions were repeated, both within and between testing sessions. This procedure served three purposes. First, it facilitated the examination and comparison of the

intrasession and intersession stabilities of especially important sortings (e.g., yourself as you usually are). Second, it facilitated specification of the effects of changes that sometimes took place between testing sessions. For example, important instabilities usually appeared if a participant decided during the course of the research to quit taking drugs. Third, it served a function similar to that of providing marker variables (see chapter 9). For instance, an investigator has little trouble identifying a factor as the self-concept when all of a participant's descriptions, of himself as he usually is, are heavily loaded on that factor.

As used in this investigation, Kelly's Rep Test required each participant to evaluate twenty persons he knew (representing specific interpersonal roles, such as mother, employer liked, person who dislikes you, some weirdo, etc.) on 21 dimensions (or constructs). Fifteen of these dimensions were created by the participant himself, to represent constructs that were important to him. Six additional, standard constructs were provided by the investigators and were used by all participants.

Data from the Rep Test were analyzed by both O-type and P-type factor analyses. In the O-type analyses, each participant's assessments of the twenty persons were intercorrelated to produce a matrix representing perceived similarities among people. Factor analysis of this matrix revealed the structure of the participant's interpersonal space. It showed which persons in his life he tended to group together as being similar and how many such groupings he recognized.

For P-type analyses, correlations were calculated among the constructs (dimensions) each participant used for describing the twenty persons. Factor analysis of this matrix showed the nature and complexity of the conceptual scheme by which each participant evaluated important people in his life. As noted above, the Rep Test included six standard constructs in addition to the fifteen provided by the participant himself. The standard constructs served as marker variables for the identification of factors. They were: kind, selfish, mean, strong, wise, and sexy.

Finally, both the Q-sort data and the Rep Test data were used to examine similarities among persons. This involved standard Q-type factor analysis, in which similarities among score patterns provide a basis for identifying groups of persons who respond to the test materials in homogeneous fashion. For example, participants who described their usual selves (or who used the six standard constructs on the Rep Test) in similar ways would tend to fall into the same factors.

In light of the high degree of individuality of the participants in this project, it was not surprising to find that summary statistics and attempts to group participants into types by quantitative means did not provide consistent results. Q-type factor analyses will become more useful when data are available from participants who use drugs other than cocaine. Factor analysis can then be used to determine whether persons who habitually take different drugs do indeed fall into separate factors, as identified by these measures.

In terms of the overall life style, the data seemed to indicate that persons who took relatively low levels of cocaine took it primarily to increase pleasure (i.e., to produce enjoyable experiences) and were relatively less intensely engaged in struggles to maintain their self-concepts or to succeed in a competitive world. Persons who more or less continuously took large amounts of the drug used it not for pleasure but to escape from intolerable internal states; they required it to support their own self-concepts and were actively engaged in desperate (but often self-destructive) struggles to assert themselves in a world they viewed as hostile and unsympathetic. When comparable data become available from persons who take other drugs, the end-product will be a veritable encyclopedia of information about life styles and drug usage.

CONCLUSIONS

Though not yet highly regarded by conventional psychology, single-organism research has proven its value in the past, and if properly used to study appropriate problems, it holds considerable promise for the future. The overview presented in this chapter has shown that studies of single organisms possess a variety of desirable characteristics for scientific investigations at all levels of control.

In laboratory experimentation, single organisms can be more effectively and efficiently handled, more thoroughly known, and subjected to more completely and more appropriately controlled conditions than can large groups. At the other extreme, in exploratory research, single organisms can be specifically selected for appropriateness to particular problems and treated not as

"subjects" but as expert consultants. For example, in exploratory representative case studies, a single person may be sought out because he or she displays precisely the characteristic an investigator wishes to study most closely.

As measurement procedures become more sophisticated, and as appropriate techniques of statistical analyses (in particular, time-series analysis) become more accessible, single-organism research will become progressively more effective as a tool of scientific investigation. It will contribute knowledge of principles of behavior that are valid not only for groups but for all individuals.

NOTES

[1]Allport, 1962; Bachrach, 1962; Bakan, 1968; Bellak and Chassan, 1964; Carlson, 1971; Chassan, 1960, 1961, 1965, 1967; Davidson and Costello, 1969; Dukes, 1965; Edgington, 1967; Gottman, 1973; Holtzman, 1963; Kelly, 1955, 1963, 1968; Mair, 1970a, 1970b; H.A. Murray, 1938; Ross, 1963; Schultz, 1969; Shapiro, 1961a, 1961b, 1966; Shontz, 1965; Sidman, 1960; Stephenson, 1953; White, 1952, 1963.

[2]The diversity of the problems that have been investigated by single-organism methods is shown by the fact that Dukes' list includes Ebbinghous' experiments on memory in 1885; Bryan and Harter's 1899 study of plateaus in learning; Stratton's 1897 studies of the effects of inverting lenses on perception; the Kelloggs' 1933 project (which was followed later by projects using a similar approach: e.g., Hayes and Hayes, 1952) in which a single chimpanzee was raised in a human environment; Cannon and Washburn's 1912 study of the relation between stomach contractions and hunger; Watson and Rayner's 1920 demonstration of conditioned emotional responses in a young boy; Jones' 1924 supplement to Watson and Rayner's research; Prince's 1905 description of a case of multiple personality: Breuer's famous case study of Anna O. (Breuer and Freud, 1895/1955); Yerkes' (1927) studies of a gorilla; Culler and Mettler's demonstration in 1934 of conditioning in a decorticate dog; and in 1932, Burtt's classic study of long term memory in his son (also Burtt, 1941).

[3]A manual for analyzing interrupted time-series experiments with the simplest integrated moving average model is reported to be available from John M. Gottman, Department of Psychology, Indiana University, Bloomington, Indiana 47401. As a source for programs for model fitting and forecasting, Gottman cited James R. Taylor, Project Administrator, University of Wisconsin, National Program Library and Inventory Service for the Social Sciences, Room 4430, Social Science Building, Madison, Wisconsin 53706.

REFERENCES

Allport, G. The general and the unique in psychological science. *Journal of Personality,* 30: 405-422, 1962.

Bachrach, A.J. *Psychological Research: An Introduction.* New York: Random House, 1962.

Bakan, D. A generalization of Sidman's results on group and individual functions. *Psychological Bulletin,* 51:63-64, 1954

Bakan, D. The general and the aggregate: A methodological distinction. *Perceptual and Motor Skills,* 5:211-212, 1955.

Bakan, D. The test of significance in psychological research. *Psychological Bulletin,* 66:423-437, 1966.

Bakan, D. *On Method: Toward a Reconstruction of Psychological Investigation.* San Francisco: Jossey-Bass, 1968.

Bandura, A. *Principles of Behavior Modification.* New York: Holt, Rinehart, and Winston, 1969.

Bannister, D., and Mair, J.M.M. *The Evaluation of Personal Constructs.* New York: Academic Press, 1968.

Bellak, L., and Chassan, J.B. An approach to the evaluation of drug effect during psychotherapy: A double-blind study of a single case. *Journal of Nervous and Mental Disease,* 139:20-30, 1964.

Bower, C.P.; Padia; W.L.; and Glass, G.V. *TMS: Two Fortran IV Programs for Analysis of Time Series Experiments.* Boulder: University of Colorado, Laboratory of Educational Research, 1974.

Box, G.E.P., and Jenkins, G.M. *Time-series Analysis: Forecasting and Control.* San Francisco: Holden Day, 1970.

Breuer, J., and Freud, S. Case histories. In: Strachey, J., ed. *The Standard Edition of the Complete Psychological Works of Sigmund Freud.* Vol. 2. London: Hogarth Press, 1955, pp.19-181. (Orig. publ. 1895.)

Bryan, W.L., and Harter, N. Studies on the telegraphic language. The acquisition of a hierarchy of habits. *Psychological Review,* 6:345-375, 1899.

Burtt, H.E. An experimental study of early childhood memory. *Journal of Genetic Psychology,* 40:287-295, 1932.

Burtt, H.E. An experimental study of early childhood memory: Final report. *Journal of Genetic Psychology,* 58:435-439, 1941.

Campbell, D.T. From description to experimentation: Interpreting trends in quasi-experiments. In: Harris, C.W., ed. *Problems in Measuring Change.* Madison: University of Wisconsin Press, 1963.

Campbell, D.T. Reforms as experiments. *American Psychologist,* 24:409-429, 1969.

Campbell, D.T., and Stanley, J.C. *Experimental and Quasi-experimental Designs for Research.* Chicago: Rand McNally and Co., 1963.

Cannon, W.B., and Washburn, A.L. An explanation of hunger. *American Journal of Physiology,* 29:441-454, 1912.

Carlson, R. Where is the person in personality research? *Psychological Bulletin,* 75:203-219, 1971.

Cattell, R.B. Personality structure and measurement. I. The operational definition of trait unities. *British Journal of Psychology,* 36:88-103, 1946.

Cattell, R.B. The structuring of change by P-technique and incremental R-technique. In: Harris, C.W., ed. *Problems in Measuring Change.* Madison: University of Wisconsin Press, 1963.

Cattell, R.B., and Cross, K.P. Comparison of the ergic and self-sentiment structures found in dynamic traits by R- and P- techniques. *Journal of Personality,* 21:250-271, 1952.

Chassan, J.B. Statistical inference and the single case in clinical design. *Psychiatry,* 23:173-184, 1960.

Chassan, J.B. Stochastic models of the single case as the basis of clinical research design. *Behavioral Science,* 6:42-50, 1961.

Chassan, J.B. Intensive design in clinical research. *Psychosomatics,* 6:289-294, 1965.

Chassan, J.B. *Research Designs in Clinical Psychology and Psychiatry.* New York: Appleton-Century-Crofts, 1967.

Cohen, M.; Liebsen, I.A.; and Faillace, L.A. The role of reinforcement contingencies in chronic alcoholism: An experimental analysis of one case. *Behavior Research and Theory,* 9:375-379, 1971.

Culler, E., and Mettler, F.A. Conditioned behavior in a decorticate dog. *Journal of Comparative Psychology,* 18:291-303, 1934.

Davidson, P.O., and Costello, C.G. *N=1: Experimental Studies of Single Cases.* New York: Van Nostrand Reinhold Co., 1969.

Dukes, W.F. N=1. *Psychological Bulletin.* 64:74-79, 1965.

Ebbinghaus, H. *Über das Gedächtnis.* Leipzig: Duncker & Humblot, 1885.

Edgington, E.S. Statistical inference from N=1 experiments. *Journal of Psychology,* 65:195-199, 1967.

Glass, G.V.; Willson, V.K.; and Gottman, J.M. *Design and Analysis of Time-Series Experiments.* Boulder: Colorado Associated University Press, 1975.

Gottman, J.M. N-of-one and N-of-two research in psychotherapy. *Psychological Bulletin,* 80:93-105, 1973.

Gottman, J.M.; McFall, R.M.; and Barnett, J.T. Design and analysis of research using time series. *Psychological Bulletin,* 72:299-306, 1969.

Harris, C.W., ed. *Problems in Measuring Change.* Madison: University of Wisconsin Press, 1963.

Hayes, K.J., and Hayes, C. Imitation in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology.* 45:450-459, 1952.

Holt, R.R. Individuality and generalization in the psychology of personality. *Journal of Personality,* 30:377-404, 1962.

Holtzman, W. Statistical models for the study of change in the single case. In: Harris, C.W., ed. *Problems in Measuring Change.* Madison: University of Wisconsin Press, 1963.

James, W. *The Varieties of Religious Experience.* New York: New American Library of World Literature, Inc., 1958. (Orig. pub. 1902).

Jones, M.C. A laboratory study of fear: The case of Peter. *Journal of Genetic Psychology,* 31:308-315, 1924.

Jones, R.H.; Crowell, D.H.; and Kapuniai, L.E. Change detection model for serially correlated data. *Psychological Bulletin,* 71:352-358, 1969.

Kellogg, W.N., and Kellogg, L. *The Ape and the Child.* New York: McGraw-Hill, 1933.

Kelly, G.A. *The Psychology of Personal Constructs.* Vols. 1 and 2. New York: Norton, 1955.

Kelly, G.A. *A Theory of Personality.* New York: Norton, 1963.

Kelly, G.A. Man's construction of his alternatives. In: Lindzey, G., ed. *The Assessment of Human Motives.* New York: Rinehart, 1968.

Lester, D. Self-maintenance of intoxication in the rat. *Quarterly Journal of Studies on Alcohol,* 22:223-231, 1961.

Lettieri, D.J. "Affect, Attitude and Cognition in Suicidal Persons." Doctoral dissertation. University of Kansas, Lawrence, Kansas, 1970.

Lloyd, R.W., Jr., and Salzberg, H.C. Controlled social drinking: An alternative to abstinence as a treatment goal for some alcohol abusers. *Psychological Bulletin,* 82:815-842, 1975.

Luborsky, L. Intraindividual repetitive measurements (P technique) in understanding psychotherapeutic change. In: Mowrer, O.H., ed. *Psychotherapy Theory and Research.* New York: Ronald Press, 1953. pp.389-413.

Mair, J.M.M. The person in psychology and psychotherapy: An introduction. *British Journal of Medical Psychology,* 43(3):197-205(a), 1970.

Mair, J.M.M. Experimenting with individuals. *British Journal of Medical Psychology,* 43(3):245-256(b), 1970.

Miller, P.M. The use of behavioral contracting in the treatment of alcoholism. *Behavior Therapy,* 3:593-596, 1972.

Murray, E.J. A case study in a behavioral analysis of psychotherapy. *Journal of Abnormal and Social Psychology,* 49:305-310, 1954.

Murray, H.A., ed. *Explorations in Personality.* New York: Oxford, 1938.

Neale, J.M., and Liebert, R.M. *Science and Behavior: An Introduction to Methods of Research.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Neuringer, C., and Michael, J.L. *Behavior Modification in Clinical Psychology.* New York: Appleton-Century-Crofts, 1970.

O'Connell, K.A. Interrupted time-series analysis of mood changes during the menstrual cycle. Masters thesis, University of Kansas, Lawrence, Kansas, 1975.

Prince, M. *The Dissociation of a Personality.* New York: Longmans, Green, 1905.

Ross, A.O. Deviant case analysis: A neglected approach to behavioral research. *Perceptual and Motor Skills,* 16:337-340, 1963.

Schultz, D.P. The human subject in psychological research. *Psychological Bulletin,* 72:214-228, 1969.

Shapiro, M.B. A method of measuring psychological changes specific to the individual psychiatric patient. *British Journal of Medical Psychology,* 34:151-155(a), 1961.

Shapiro, M.B. The single case in fundamental clinical psychological research. *British Journal of Medical Psychology,* 34:255-262(b), 1961.

Shapiro, M.B. The single case in clinical-psychological research. *Journal of General Psychology,* 74:3-23, 1966.

Shontz, F.C. *Research Methods in Personality.* New York: Appleton-Century-Crofts, 1965.

Shontz, F.C. Individuality in evaluations of treatment effectiveness. *Journal of Counseling Psychology,* 19:76-80, 1972.

Sidman, M. A note on functional relations obtained from group data. *Psychological Bulletin,* 49:263-269, 1952.

Sidman, M. *Tactics of Scientific Research: Evaluating Experimental Data.* New York: Basic Books, 1960.

Sobell, M.B., and Sobell, L.C. Individualized behavior therapy for alcoholics. *Behavior Therapy,* 4:49-72, 1973.

Spotts, J.V., and Shontz, F.C. *Life Styles of Nine American Cocaine Users: Trips to the Land of Cockaigne.* Washington, D.C.: NIDA (in press).

Stephenson, W. *The Study of Behavior: Q-technique and Its Methods.* Chicago: University of Chicago Press, 1953.

Stratton, G.M. Vision without inversion of the retinal image. *Psychological Review,* 4:341-360, 463-481, 1897.

Watson, J.B., and Rayner, R. Conditioned emotional reactions. *Journal of Experimental Psychology,* 3:1-14, 1920.

White, R.W. *Lives in Progress.* New York: Holt, Rinehart and Winston, 1952.

White, R.W., ed. *The Study of Lives: Essays on Personality in Honor of Henry A. Murray.* New York: Atherton Press, 1963.

Wold, H. *Bibliography on Time-Series and Stochastic Processes.* Cambridge, Mass.: M.I.T. Press, 1965.

Yerkes, R.M. The mind of a gorilla. *Genetic Psychology Monographs,* 2:1-193, 1927.

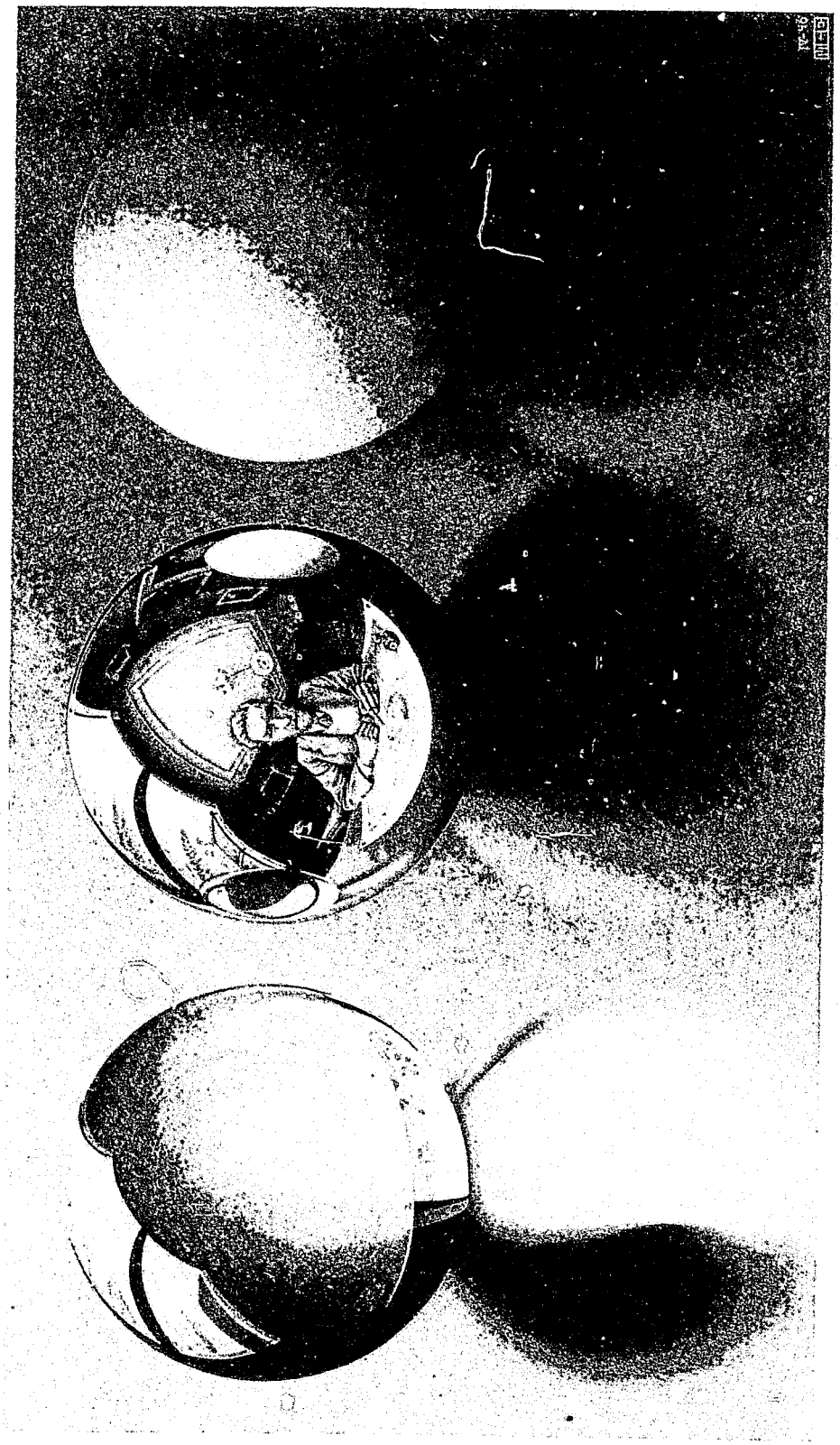


Plate 4

INTRODUCTION

According to Baltes (1973), and Baltes and Goulet (1970), the goal of behavioral sciences concerns the description, explanation, and modification of human behavior. However, while such a statement implies that the study of behavior emphasizes stability as well as change, actual empirical investigations of change, especially at relatively macroscopic levels and over longer time intervals, have been few and far between (Wohlwill, 1973a). That is, even in the area of developmental psychology, it has been common to rely primarily on static cross-sectional rather than longitudinal methods and designs. Only in longitudinal studies is the same sample of subjects followed over time and observed repeatedly at preselected age levels. In contrast, in cross-sectional studies independent samples of subjects from different age groups are observed only once at the same occasion.

It is abundantly clear that any social intervention aimed at modifying human behavior requires by necessity a direct assessment of intraindividual change, and interindividual differences in intraindividual change, through repeated observations of the same individuals over time. Only longitudinal data can provide information concerning: (a) the description of direction and shape of intraindividual changes; (b) the identification of individuals exhibiting exceptional changes; (c) the determination of relationships between earlier behaviors and later responses; (d) the determination of relationships between earlier life conditions and later behaviors; and (e) the assessment of differential changes for groups to whom different treatments have been administered. Methods that try to short-cut the more laborious and time consuming longitudinal measurement of individual as well as group patterns of change will, therefore, sacrifice at least part or all of that information.

Given the theoretical and practical importance of observing the course of behavioral events over time, the following discussion will first consider in greater detail the rationale for using longitudinal methods, particularly in comparison to cross-sectional methods. The next part, Rationale, will discuss the use of longitudinal methods in accurately describing change patterns and focus on problems of internal and external validity of simple, as well as extended, longitudinal designs. The experimental application of longitudinal methods and problems of causal inference will then be discussed in the sections on Methods and Procedures. The final section will present a practical and more concrete illustration of a research design and general considerations with regard to the choice of analytical procedures. No attempt will be made to outline specific statistical procedures, since they are described in other chapters of this volume.

RATIONALE: LONGITUDINAL VS. CROSS-SECTIONAL METHODS

The notion of behavioral change is operationally linked to the repeated observation of the same individuals on two or more occasions ordered along a dimension of chronological time. In developmental studies, time is usually conceptualized as chronological age (time passed since birth). In studies of drug use it may also be defined, depending on one's hypotheses, as time passed since first contact with a particular drug, as time passed since the onset of a specific treatment program, as time passed since the termination of a given treatment, etc.

it has been argued that time is not a psychological variable and should, therefore, be replaced by indices that are more closely related to psychological theories, such as mental age or social age (based on one's acquisition of certain social norms), for instance. However, there are at least two reasons favoring the use of chronological time in longitudinal designs. First, any other kind of index, since it has to indicate the temporal order of events, will by necessity be related to time, at least in a monotonic fashion. Second, in contrast to these derived measures, time itself represents a variable that is easily and reliably measured on an

interval scale and, therefore, easily replicable. The latter point is particularly important for comparative analyses of change patterns obtained for different groups or populations. In other words, if the time index against which behavior is plotted represents a scale with questionable properties, the specification of temporal patterns loses in precision and makes comparisons less meaningful. In the following, it will be assumed throughout that change is measured and plotted along a continuum of chronological time.

Since longitudinal studies require greater investments of time and effort on the part of both subjects and researchers, it has been common to employ cross-sectional designs that can be carried out over short periods of time. More specifically, temporally ordered, longitudinal series of observations are replaced by independent measurements that have no intrinsic sequentiality. For instance, in developmental studies different age groups are sampled and measured at the same point in time instead of having to wait until a given group of individuals has lived through a certain period of time. In drug research, one could assess the long-term effects of a given intervention program cross-sectionally by sampling from groups of individuals that have been involved in the program for different amounts of time.

However, it has become abundantly clear that the internal validity of cross-sectional differences as indicators of intraindividual changes is highly questionable (Baltes, 1968; Buss, 1973; Campbell and Stanley, 1963; Schaie, 1965, 1970, 1973). In other words, it is questionable whether differences between independent groups can be assumed to be valid estimates of changes over time in the same group of individuals. This consideration is particularly important when studying behaviors (e.g., drug use) that can be expected to be highly dependent upon cultural fads and trends. In general, cross-sectional differences between groups are likely to confound intraindividual changes with effects due to mechanisms such as selective sampling, selective survival, selective drop-out, generation differences, or any combination of them. At the same time, these error sources may also limit the external validity or the extent to which the obtained findings can be generalized beyond the limits of the specific study. For instance, if the participants in a given drug program drop out at different times and if that attrition is not random but systematically related to the dependent variable, the resulting cross-sectional 'trend' or pattern will be biased and misleading.

While the various deficiencies just mentioned jeopardize both the internal and external validity of cross-sectional differences, it should be realized that they are also a threat to the representativeness and generalizability of longitudinal change patterns. Therefore, they will be discussed more fully in the next section. Suffice it to say here that the usefulness of simple cross-sectional designs is limited primarily to initial explorations of behavioral change phenomena. Once a target pattern for a particular problem has been established, the application of longitudinal designs becomes necessary.

METHODS AND PROCEDURES: DESCRIPTION OF CHANGE

SIMPLE LONGITUDINAL DESIGNS

In the simplest case of a longitudinal series of observations, a researcher samples individuals from some target population and measures them repeatedly on two or more occasions. As might be expected, the internal and external validity of the obtained change patterns depend on the degree to which several design-related sources of error are controlled for (Baltes, 1968; Campbell, 1967; Campbell and Stanley, 1963). In addition, the role of these error factors will to some extent be related to the fact whether the data represent actual changes measured concurrently or whether retrospective and/or prospective methods are employed. The sources of error that will be discussed are: testing effects, selective survival, selective sampling, selective drop-out, and generation effects.

Internal Validity of Longitudinal Changes

A major factor jeopardizing the internal validity of longitudinal changes is related to the presence or absence of testing effects (Baltes, 1968; Campbell, 1967; Campbell and Stanley, 1963; Labouvie, Bartsch, Nesselroade and Baltes, 1974; Wohlwill, 1973a). In other words, being included in a study and being tested repeatedly may sensitize subjects and lead to practice effects that are performance-specific for the particular test, but not indicative of changes in the underlying characteristics the test is supposed to measure. For instance, when assessing

longitudinal patterns of drug use, repeated testing may lead, on the part of the subjects, to an increased awareness of socially accepted norms of drug-related behavior, thus leading to changes of test responses in a socially desirable direction without concomittant changes in actual drug use.

Although the influence of testing effects can be partially controlled for by a careful choice of unobtrusive and nonreactive measurement instruments (Wohlwill, 1973a), it should be realized that this problem is not necessarily eliminated by using behavior ratings. Even when subjects are quite unaware of being observed repeatedly, raters' perceptions of the same person may change over time because of increased familiarity with the observed individuals rather than because of actual changes in the observed behaviors. Regardless of whether behavior ratings or self-reports are used, measurement situations become less obtrusive if they become a more or less natural part of the subjects' environment.

To estimate the presence and magnitude of potential testing effects, a simple longitudinal series may be extended by adding a series of independent control groups, each measured only once (see Table 1). However, the quality of such a control series will be affected by the operation of several other factors to be mentioned later.

TABLE 1

Control for Testing Effects in Simple Longitudinal Designs

Group	Occasion			
	O_1	O_2	O_3	.....O_n
Longitudinal	X	X	X	X
Control 1	X			
Control 2		X		
Control 3			X	
.				
.				
.				
Control n				X

When an investigator chooses to use retrospective and/or prospective methods to assess actual change, the problem of internal validity is further complicated by questions of unreliability. It has been sufficiently documented that retrospective accounts of developmental changes are often systematically biased and distorted (Baltes and Goulet, 1971; Wohlwill, 1973a). Similarly, prospective accounts of expected changes may not reflect very often the actual changes that occur later. Of course, this issue of reliability does not apply if measures of retro- and prospective changes are used in their own right either as dependent variables or as possible determinants of actual changes (Baltes and Goulet, 1971; Thomae, 1970).

External Validity of Longitudinal Changes

While the internal validity of longitudinal changes is jeopardized mainly by testing effects and possible lack of reliability, methodological deficiencies affecting the external validity are more numerous and more difficult to control for. Among them and to be discussed in the following are: selective sampling, selective survival, selective drop-out, and generation effects.

Selective Sampling. Due to the requirement of repeated participation with its increased demands on subjects in terms of time and effort, longitudinal samples are usually biased from the very beginning (Baltes, 1968; Rose, 1965; Streib, 1966). That is, if volunteering for longitudinal studies is correlated with the dependent measures, the generalizability of the obtained individual

and group patterns of change is limited to selected subgroups. Furthermore, if selective volunteering or sampling is related to the anticipated number of participations, the use of independent control groups for the assessment of testing effects, as proposed above, becomes questionable.

Selective Survival. Due to death and disease of different individuals at different times/ages, a given population at birth (cohort) changes gradually over time in its composition. As has been shown empirically (Damon, 1965; Riegel, Riegel and Meyer, 1967; Riegel and Riegel, 1972), these changes are not random but selective. If correlated with the dependent variable, they will not only jeopardize the internal validity of cross-sectional differences, but also the external validity of longitudinal changes. Since the obtained pattern of change is based only on those subjects that survived all occasions, it is necessarily not representative for those individuals that died during the course of the study.

Selective Drop-out. Besides biological survival, attrition of subjects is also influenced by social and psychological factors. Because of loss of interest, change of residence, and similar reasons, some individuals will discontinue their participation during the course of a longitudinal study. As demonstrated empirically (Baltes, Schaie and Nardi, 1971; Labouvie, Bartsch, Nesselroade and Baltes, 1974; Riegel, Riegel and Meyer, 1968), this drop-out is likely to be selective and related to the dependent measure leading to an increasingly biased sample of retestees. If repeated participation itself is an important determinant of this selective experimental mortality (Campbell and Stanley, 1963), it will also limit the usefulness of independent control groups for the assessment of possible testing effects.

Generation Effects. One of the major factors that has been recognized as a source of internal invalidity of cross-sectional studies concerns the fact that different generations or cohorts of individuals grow up under different socio-cultural conditions (e.g., different educational systems); or they experience the same situations (e.g., economic depressions) at different ages (Baltes, 1968; Schaie, 1965). Because cultures are continuously changing and present changing environments for individuals to interact with, longitudinal changes obtained for one particular cohort may be rather specific and not generalizable to other generations. Therefore, it becomes necessary to replicate time-series of observations for different cohorts. The resulting extended designs will be discussed next.

EXTENDED LONGITUDINAL DESIGNS

After realizing that discrepancies between cross-sectional and longitudinal age curves of intellectual development were, at least in part, due to differences between generations (Baltes, 1968; Baltes and Labouvie, 1973), developmental psychologists introduced more sophisticated designs for the accurate description of age-related changes (Baltes, 1968; Buss, 1973; Cattell, 1970; Schaie, 1965, 1970, 1973). Initially, Schaie (1965) proposed a trifactorial model with the parameters of age, cohort (time of birth), and time of measurement to represent functionally different sources of behavioral variance. Partly because of the algebraic interdependence, and partly because of the assigned status of the three time variables, Baltes (1968) and Buss (1973) subsequently argued that a bifactorial Age X Cohort design was most useful and sufficient for strictly descriptive purposes. However, since the latter point of view is not quite satisfying either (Buss, 1975; Labouvie, 1975 a, b), the following discussion will consider all three bifactorial designs that can be derived from Schaie's general model.

Time-sequential Design

Although this design does not yield longitudinal observations of intraindividual changes, it can provide useful information about general cultural trends as the background against which to evaluate the impact of specific intervention programs. As illustrated in Table 2, a set of age levels is observed on several occasions (times of measurement).[1] Empirical applications of this design can be found in studies by Baltes, Baltes and Reinert (1970), Goulet, Hay and Barclay (1974) and Schaie and Strother (1968a).

The general purpose of time-sequential analyses is the detection and description of cultural changes and trends in the behaviors studied. For instance, it may be of considerable importance to be able to predict historical trends in drug-related behaviors before implementing specific intervention programs. If drug use among certain age groups varies from year to year, the effectiveness of a given program may depend upon appropriate adjustments of the planned intervention to such cultural trends.

TABLE 2

Time-sequential Design: Four Ages are Measured at Four Times

Ages	Time of Measurement			
	1972	1975	1978	1981
8	1964	1967	1970	1973
11	1961	1964	1967	1970
14	1958	1961	1964	1967
17	1955	1958	1961	1964
	Cohort/Time of Birth			

As seen in Table 2, the effects associated with age and time of measurement are confounded by cohort effects, that is, systematic differences between groups of individuals born at different times. Thus, without the assumption of nonexisting cohort differences (Schaie, 1965), it becomes difficult to draw conclusions that imply more than the presence or absence of certain cultural trends. Whether these trends reflect concurrent environmental changes (time of measurement), or cumulative effects of different past histories (cohorts), cannot be decided on the basis of such data. However, if a comparison of age and time effects in a "cohort-balanced" design reveals highly similar patterns (see Table 2), it would be reasonable to conclude that the relevant antecedents of the observed effects are most likely covarying with the cohort variable.

Cohort-sequential Design

In this design a set of cohorts is observed at different age levels, providing a longitudinal series for each of several generations (Table 3). Although this design is considered most appropriate by both Baltes (1968) and Buss (1973), it has been employed so far only in a study by Baltes and Reinert (1969). The major practical disadvantage of this design is the amount of time required for its completion. Depending upon the range of cohorts chosen, the life span of such a study may be considerably longer than the age range studied.

TABLE 3

Cohort-sequential Design: Four Cohorts are Measured at Four Ages

Cohort	Ages			
	8	11	14	17
1965	1973	1976	1979	1982
1962	1970	1973	1976	1979
1959	1967	1970	1973	1976
1956	1964	1967	1970	1973
	Time of Measurement			

Again, the effects of the independent variables (age, cohort) are confounded, in this case by time effects (see Table 3). However, a comparison of the various cohort-specific longitudinal patterns will at least provide information about the relative stability/instability of the observed trends.

Cross-sequential Design

Because of its greater practicality, this design has been employed most frequently in empirical studies (Baltes and Nesselroade, 1972; Nesselroade, Schaie and Baltes, 1972; Schaie, Labouvie and Buech, 1973; Schaie and Labouvie-Vief, 1974; Schaie and Strother, 1968a, b). Using repeated or independent observations, a fixed set of cohorts is observed on several occasions or times of measurement (see Table 4).

The effects of cohort and time are confounded by age effects. In an "age-balanced" cross-sequential design, the confounded age levels are symmetrically distributed along the diagonal from the lower left to the upper right corner in Table 4. In other words, the marginal age distributions covarying with cohort and time of measurement are identical. Therefore, a comparison of the cross-sectional and longitudinal age curves in an "age-balanced" design (Table 4) will reveal to what extent developmental trends are susceptible to changing environmental inputs. For instance, if

TABLE 4

Cross-sequential Design: Four Cohorts are Measured at Four Times

Cohort	Time of Measurement			
	1972	1975	1978	1981
1957	15	18	21	24
1960	12	15	18	21
1963	9	12	15	18
1966	6	9	12	15
	Ages			

a cross-sequential investigation of attitudes towards drugs suggests cohort-specific longitudinal patterns that differ from the corresponding time-specific cross-sectional age gradients, it is reasonable to conclude that the observed changes are not so much a developmental phenomenon but more the result of ever present changes in the sociocultural environment of individuals.

Cautions

Although all three sequential designs are strictly descriptive, they are nevertheless to be preferred over the conventional cross-sectional and longitudinal designs. Given the developmentalists' inclination to search for fixed and invariant age patterns, the application of sequential designs will at least guard against the premature acceptance of such change models. While the extended designs are useful to estimate the extent of cultural changes and generation effects as they affect behaviors, it is important to realize that the other sources of error mentioned previously still have to be dealt with. In fact, the picture is likely to become more complicated because of the possibility that the mechanisms underlying selective sampling, selective survival, selective drop-out, and testing effects may be subject to cultural changes too (Campbell and Stanley, 1963; Baltes, Schaie and Nardi, 1971). Therefore, a general strategy to cope with these problems will include: (a) the use of appropriate series of independent control groups (testing effects); (b) an explicit attempt to describe the various cohort samples in terms of relevant environmental and background variables (selective sampling); and (c) a posteriori comparisons between drop-outs and 'survivors' (selective drop-out and survival).

METHODS AND PROCEDURES: EXPERIMENTAL MANIPULATION OF CHANGE

The preceding section dealt with the issue of accurately describing longitudinal change patterns. The following discussion is concerned with the evaluation of the effects of experimental manipulations and programmed interventions in which an attempt is made to control the conditions and events to which individuals are exposed over longer time periods. In general, such efforts are intended to find out whether different programs have differential effects on members of the same target population, or whether a particular program affects different individuals in different ways. For instance, it may be important for educators to evaluate not only the efforts of different drug education programs, but also whether high school students with different levels of intelligence or different personality characteristics react differently to the same program. The utility and validity of two designs will be discussed--simple pretest-posttest and multiple time-series.

SIMPLE PRETEST-POSTTEST DESIGNS

In terms of time and effort involved, the simplest designs range from a one-group pretest-posttest design to the four-group design proposed by Solomon and Lessac (1968). As illustrated in Table 5, the latter one provides several controls to assess the internal validity of the experimentally induced changes. Nevertheless, like its simpler relatives, it is severely limited in its usefulness.

TABLE 5

Four-Group Design by Solomon and Lessac

Group	Time		
	Pretest	Treatment	Posttest
I	O_1	X	O_2
II		X	O_2
III	O_1		O_2
IV			O_2

Groups I and III are measured before and after the experimental treatment 'X'. Groups II and IV are included to control for potential testing effects.

It is somewhat of an irony to realize that pretest-posttest designs are essentially not change-oriented. As suggested in Figure 1, their use is really only justified if it can be assumed that: (a) the behaviors studied do not exhibit any systematic changes prior to the treatment; (b) the behaviors have reached a stable level after termination of the intervention; and (c) the rate of change is approximately the same during the duration of the treatment. If these assumptions are not valid, the rather arbitrary choice of times of measurement may lead to premature conclusions about the effects of different treatments (see Figure 1). Therefore, it seems that these designs are most appropriately employed in investigations of short-term changes.

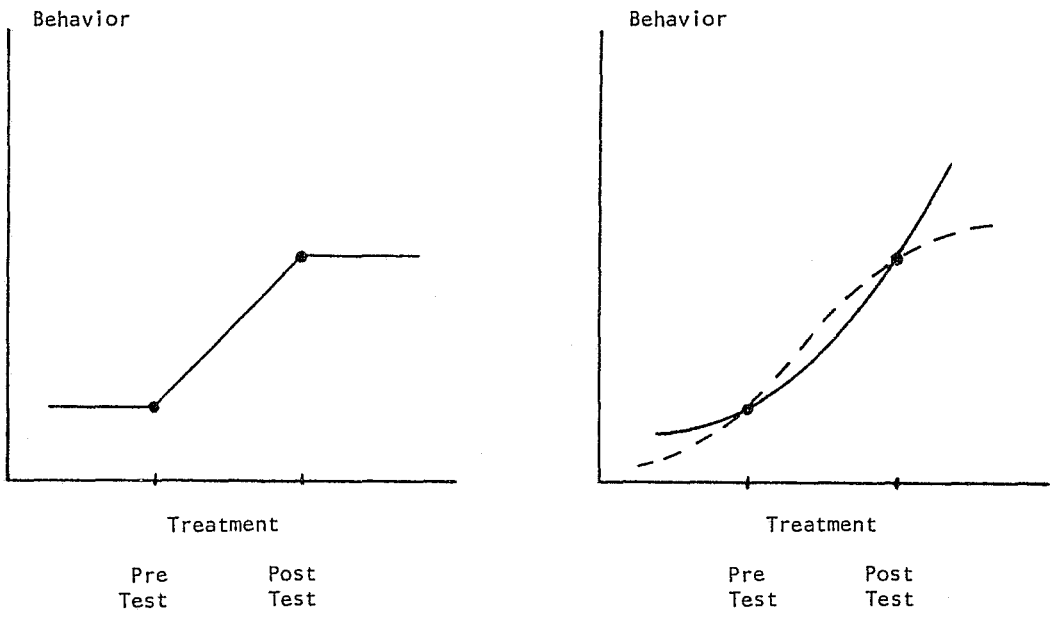


Figure 1: Assessment of behavioral trends by simple pretest-posttest designs: (a) the static assumptions implicit in these designs, (b) possible alternative trends compatible with the same observations

Since two-occasion longitudinal designs are used quite frequently, it is necessary to point out another deficiency jeopardizing the internal validity of these designs. Sometimes researchers are interested in determining the differential effects of a given treatment on persons differing on some psychological characteristic. However, when selecting subgroups of individuals with low or high scores on some measure, a second observation of these same individuals will usually indicate converging trends. This regression towards the mean (Campbell and Stanley, 1963) is a meaningful psychological phenomenon not restricted to measures with fallible scores (Furby, 1973). In order to distinguish between substantive changes and regression effects, it is, therefore, necessary to supplement the analysis of such designs by a time-reversed analysis in which subjects are classified into subgroups on the basis of their scores on the second occasion (Campbell and Stanley, 1963; Baltes, Nesselroade, Schaie and Labouvie, 1972). If the two analyses reveal opposing patterns of convergence, it is reasonable to assume the presence of regression effects.

MULTIPLE TIME-SERIES DESIGNS

In comparison to simple pretest-posttest assessments, these designs involve the repeated observation of individuals from two or more populations on numerous occasions before, during and after specified periods of intervention. Therefore, time-series provide much greater descriptive accuracy when studying trends over extended time intervals. Furthermore, in order to explicate the effects of the experimental manipulations, longitudinal series are obtained (usually simultaneously) for several experimental and control groups (Campbell, 1967).

The internal validity of multiple time-series designs in terms of differential trends for different groups depends on the presence of error sources similar to those mentioned earlier. Ideally, an experimenter assigns subjects randomly to the various treatment conditions to achieve internal validity. However, if volunteers for longitudinal studies are likely to represent biased samples, it is furthermore possible that they volunteer selectively for different types of interventions leading to a self-selection rather than randomization of subjects. For instance, volunteers for a particular drug education program may not be willing to be assigned to a control or no-treatment condition. In such a case, the problem of self-selection may be dealt with by using a time-lagged control group for which the intervention is merely temporarily delayed (Gottman, McFall and Barnett, 1969).

A second threat to the internal validity is differential selective drop-out. That is, even if initial randomization of subjects is achieved, different treatment conditions may cause different types of individuals to drop out of the study. Such a treatment by drop-out interaction makes the comparison of the various longitudinal patterns highly questionable. This is a problem particularly relevant to drug abuse treatment program evaluation. Finally, given the problem of testing effects, it is also possible that these effects may interact with the type of treatment to which subjects are exposed.

Considering the external validity of multiple time-series, it should be obvious that all behavioral research is embedded in a general background of ever changing socio-cultural conditions (Riegel, 1972, 1973). Therefore, a replication of the same design at different points in time may reveal different longitudinal patterns for the treatment as well as no-treatment conditions. Thus, it would seem useful to extend experimental multiple time-series designs to include features of the sequential designs discussed earlier.

While the aforementioned problems may be considered as error sources regardless of one's particular theoretical framework, there is also a more intrinsic issue of uncertainty involved in the explication of antecedent-consequent relationships in studies of long-term changes (Labouvie, 1975c; Wohlwill, 1973a, b). Depending on one's theoretical stance, the psychologically relevant aspects of a given intervention program may be defined either in terms of a series of specific stimulus events under the control of the experimenter, or in terms of the activities subjects engage in as a result of the intervention, or in terms of interactions between the former two. The first case can be realized only if the researcher is willing to sacrifice external validity to achieve internal validity by severely limiting the subjects' response repertoires. In the other two cases, a gain of external validity implies greater uncertainty in the explication of valid antecedent-consequent relationships. Since experimental control of specific behaviors becomes less effective the longer the time intervals studied, the uncertainty with regard to functional interpretations increases, and it seems most appropriate to describe programmed interventions not only in terms of manipulated stimulus conditions and certain target behaviors, but also in terms of each subject's responses and behaviors elicited by these events (Wohlwill, 1973a).

METHODS OF ANALYSIS

Since a number of analytical procedures is fully described in other chapters of this volume, it seems sufficient to limit the present discussion to some general considerations. Among these are issues concerning the type of dependent variable used--quantitative or qualitative--and the particular aspect of change--quantitative or structural--that a researcher may be interested in.

The most common and perhaps most preferred situation involves the measurement of quantitative changes in level on a variable measured by the same instrument on all occasions with the same reliability and validity. The analytical procedures employed in this case are analysis of variance or trend analysis (Kirk, 1968; Winer, 1962). The latter method becomes meaningful if more than two occasions are included and if one attempts to forecast behavioral trends beyond the last time of measurement. However, both methods are static in the sense that they require the variance-covariance matrix across occasions to be homogeneous (Kirk, 1968). In contrast, it is probably more likely that the empirical correlations between occasions decrease with increasing time intervals between them (Kagan and Moss, 1962), resulting in a positive bias in the corresponding F tests (Kirk, 1968). (See also chapters 10, 11, and 12.)

The time-related dependency of longitudinal observations can be used more directly in the case of multiple observations by estimating the parameters of models that view time-series as stochastic processes (Box and Jenkins, 1970; Glass, 1972; Gottman, McFall and Barnett, 1969). In these models interventions are represented by binary variables (0,1). Changes in slope and/or level of the series as a result of an experimental manipulation are assessed by estimating so-called generating functions (Gottman, McFall and Barnett, 1969). (See also chapter 6.)

If the emphasis is less on quantitative changes and more on changes in structural relationships, one may apply factor analytic methods (Baltes and Nesselroade, 1973; Bentler, 1973; Nesselroade, 1970) or path analytic procedures (Buss, 1974; Labouvie, 1974) or a combination of both to analyze relationships among multiple sets of response variables within and between occasions. These procedures can be meaningfully used even when different sets of dependent measures are employed at different occasions. (See also chapters 8 and 9.)

In the case of qualitative response variables, the dependent measure represents a minimal scale with two or more response classes or categories. In some cases, such variables may be quantified by measuring the age/time at which a particular response class (stage) is achieved. If such a procedure is not meaningful, the analysis of time-series with qualitative variables may investigate individual sequences of responses and compare their relative frequency of occurrence in the various experimental and control groups (Wohlwill, 1973a). For instance, a researcher may, on the basis of several characteristics, find it useful to distinguish between different 'stages' of drug use and abuse. Longitudinal observations will then yield an ordered string of stage designations for each subject in a form such as this: A-A-B-C-B (5 occasions, 3 stages A, B, and C). A comparison of different groups may then reveal differential frequencies of the various longitudinal patterns under differing conditions. (See also chapter 5.)

ILLUSTRATIVE APPLICATION: MULTIPLE TIME-SERIES IN DRUG RESEARCH

The recent increase in the popularity of drugs among large numbers of young people has become a matter of public concern and scientific interest (Josephson, 1974). Educators may want to know whether particular educational programs may lead to a more "enlightened" adolescent use of drugs in terms of frequency and amount of underlying motivations.

Previous observations suggest that age-related patterns of marihuana use during adolescence are highly susceptible to cultural trends (Josephson, 1974). Therefore, it is reasonable to assume that the effect of a given program may not only depend on the age at which it is administered, but also on the general historical context in which adolescents have grown up and/or are currently experiencing. To explicate the differential impact of such a planned intervention for different age levels and generations, a sequential multiple time-series design may be chosen, as illustrated in Table 6. Assuming that the measures used are sufficiently nonreactive, subjects within each of three age/cohort levels are randomly assigned to an experimental and a control condition and observed repeatedly over a period of five years (see Table 6). Using comparable sampling strategies, the first sequence is replicated after a delay of two years. It may also be mentioned here that a study by Jessor, Jessor and Finney (1973) on marihuana use represents an approximation to the extended designs discussed above. Their data on high school students correspond to a cross-sequential design. However, the design is incomplete in the sense that cohorts are dropped once students graduate from high school.

If subject attrition across occasions is found to be unrelated to the dependent variables and to be the same for all series, an analysis of the data may focus on the following comparisons. (a) Each cohort yields a longitudinal series of repeated observations of a control group covering a certain age range; at the same time, each time of measurement (within sequence A or B) provides cross-sectional age differences between the three control groups. This combination of longitudinal and cross-sectional age curves yields information concerning the presence of cultural trends. (b) For each cohort, the corresponding longitudinal series for the control and experimental group can be compared. (c) Since sequence B represents a replication of sequence A at a later time, longitudinal observations of experimental groups covering corresponding age ranges can be compared to indicate the presence of Treatment X Time interactions. Obviously, some of these comparisons will become questionable if the dropout of subjects is found to be selective and different for different series.

TABLE 6

Sequential Design for the Study of Adolescent Drug Use: Effects of a Programmed Intervention at Different Age Levels and Times of Measurement

Cohort Group		Time of Measurement[a]						
		1975	1976	1977	1978	1979	1980	1981
		Sequence A						
1967	exp.	8	9 X	10 X	11	12		
	cont.	8	9	10	11	12		
1965	exp.	10	11 X	12 X	13	14		
	cont.	10	11	12	13	14		
1963	exp.	12	13 X	14 X	15	16		
	cont.	12	13	14	15	16		
		Sequence B						
1969	exp.			8	9 X	10 X	11	12
	cont.			8	9	10	11	12
1967	exp.			10	11 X	12 X	13	14
	cont.			10	11	12	13	14
1965	exp.			12	13 X	14 X	15	16
	cont.			12	13	14	15	16

[a]Body entries represent age in years. 'X' marks periods of planned intervention. Within each cohort, subjects are randomly assigned to an experimental and a control group. Each row represents a series of repeated observations of the same group of subjects.

NOTES

[1]In Tables 2 to 4 it is assumed that age is given in years, while cohort and time of measurement are defined in terms of calendar years. Of course, a researcher may choose to use smaller or larger time units.

REFERENCES

Baltes, P.B. Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development,* 11:145-171, 1968.

_____. Prototypical paradigms and questions in life-span research on development and aging. *Gerontologist,* 13:458-467, 1973.

Baltes, P.B.; Baltes, M.M.; and Reinert, G. The relationship between time of measurement and age in cognitive development of children: An application of cross-sectional sequences. *Human Development,* 13:258-268, 1970.

Baltes, P.B., and Goulet, L.R. Status and issues of a life-span developmental psychology. In: Goulet, L.R., and Baltes, P.B., eds. *Lifespan Developmental Psychology. Research and Theory.* New York: Academic Press, 1970.

Baltes, P.B., and Goulet, L.R. Exploration of developmental variables by manipulation and simulation of age differences in behavior. *Human Development,* 14:149-170, 1971.

Baltes, P.B., and Labouvie, G.V. Adult development of intellectual performance: Description, explanation, modification. In: Eisdorfer, C., and Lawton, P.L., eds. *The Psychology of Adult Development and Aging.* Washington, D.C.: American Psychological Association, 1973.

Baltes, P.B., and Nesselroade, J.R. Cultural change and adolescent personality development: An application of longitudinal sequences. *Developmental Psychology,* 7:244-256, 1972.

Baltes, P.B., and Nesselroade, J.R. The developmental analysis of individual differences on multiple measures. In: Nesselroade, J.R., and Reese, H.W., eds. *Life-Span Developmental Psychology. Methodological Issues.* New York: Academic Press, 1973.

Baltes, P.B.; Nesselroade, J.R.; Schaie, K.W.; and Labouvie, E.W. On the dilemma of regression effects in examining ability-level-related differentials in ontogenetic patterns of intelligence. *Developmental Psychology,* 6:78-84, 1972.

Baltes, P.B., and Reinert, G. Cohort effects in cognitive development of children as revealed by cross-sectional sequences. *Developmental Psychology,* 1:169-177, 1969.

Baltes, P.B.; Schaie, K.W.; and Nardi, A.H. Age and experimental mortality in a seven-year longitudinal study of cognitive behavior. *Developmental Psychology,* 5:18-26, 1971.

Bentler, P.M. Assessment of developmental factor change at the individual and group level. In: Nesselroade, J.R., and Reese, H.W., eds. *Life-Span Developmental Psychology. Methodological Issues.* New York: Academic Press, 1973.

Box, G.E.P., and Jenkins, G.M. *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day, 1970.

Buss, A.R. An extension of developmental models that separate ontogenetic changes and cohort differences. *Psychological Bulletin,* 80:466-479, 1973.

Buss, A.R. A recursive-nonrecursive factor model and developmental causal networks. *Human Development,* 17:139-151, 1974.

Buss, A.R. More on the Age X Cohort developmental model: Reply to Labouvie. *Psychological Bulletin,* 82:170-173, 1975.

Campbell, D.T. From description to experimentation: Interpreting trends as quasi-experiments. In: Harris, C.W., ed. *Problems in Measuring Change.* Madison: The University of Wisconsin Press, 1967.

Campbell, D.T., and Stanley, J.S. Experimental and quasi-experimental designs for research on teaching. In: Gage, N.L., ed. *Handbook of Research on Teaching.* Chicago: Rand McNally, 1963.

Cattell, R.B. Separating endogenous, exogenous, ecogenic, and epogenic component curves in developmental data. *Developmental Psychology,* 3:151-162, 1970.

Damon, A. Discrepancies between findings of longitudinal and cross-sectional studies in adult life: Physique and physiology. *Human Development,* 8:16-22, 1965.

Furby, L. Interpreting regression toward the mean in developmental research. *Developmental Psychology,* 8:172-179, 1973.

Glass, G.V. Estimating the effects of intervention into non-stationary time-series. *American Educational Research Journal,* 9:463-477, 1972.

Gottman, J.M.; McFall, R.M.; and Barnett, J.T. Design and analysis of research using time series. *Psychological Bulletin,* 72:299-306, 1969.

Goulet, L.R.; Hay, C.M.; and Barclay, C.R. Sequential analysis and developmental research methods: Descriptions of cyclical phenomena. *Psychological Bulletin,* 81:517-521, 1974.

Jessor, R.; Jessor, S.L.; and Finney, J. A social psychology of marijuana use: Longitudinal studies of high school and college youth. *Journal of Personality and Social Psychology,* 26:1-15, 1973.

Josephson, E. Trends in adolescent marijuana use. In: Josephson, E., and Carroll, E.E., eds. *Drug Use. Epidemiological and Sociological Approaches.* Washington, D.C.: Hemisphere Publishing Co., 1974.

Kagan, J., and Moss, H.A. *Birth to Maturity: A Study in Psychological Development.* New York: Wiley, 1962.

Kirk, R.E. *Experimental Design: Procedures for the Behavioral Sciences.* Belmont, California: Brooks/Cole, 1968.

Labouvie, E.W. Developmental causal structures of organism-environment interactions. *Human Development,* 17:444-452, 1974.

Labouvie, E.W. Descriptive developmental research: Why only time? *Journal of Genetic Psychology,* 126:289-296, 1975a.

Labouvie, E.W. An extension of developmental models: Reply to Buss. *Psychological Bulletin,* 82:165-169, 1975b.

Labouvie, E.W. The dialectical nature of measurement activities in the behavioral sciences. *Human Development,* 18, in press, 1975c.

Labouvie, E.W.; Bartsch, T.W.; Nesselroade, J.R.; and Baltes, P.B. On the internal and external validity of simple longitudinal designs. *Child Development,* 45:282-290, 1974.

Nesselroade, J.R. Application of multivariate strategies to problems of measuring and structuring long-term change. In: Goulet, L.R., and Baltes, P.B., eds. *Life-Span Developmental Psychology. Research and Theory.* New York: Academic Press, 1970.

Nesselroade, J.R.; Schaie, D.W.; and Baltes, P.B. Ontogenetic and generational components of structural and quantitative change in adult behavior. *Journal of Gerontology,* 27:222-228, 1972.

Riegel, K.F. The influence of economic and political ideologies upon the development of developmental psychology. *Psychological Bulletin,* 78:129-141, 1972.

Riegel, K.F. Developmental psychology and society: Some historical and ethical considerations. In: Nesselroade, J.R., and Reese, H.W., eds. *Life-Span Developmental Psychology. Methodological Issues.* New York: Academic Press, 1973.

Riegel, K.F., and Riegel, R.M. Development, drop, and death. *Developmental Psychology,* 6:306-319, 1972.

Riegel, K.F.; Riegel, R.M.; and Meyer, G. A study of the drop-out rates in longitudinal research on aging and the prediction of death. *Journal of Personality and Social Psychology,* 4:342-348, 1967.

Riegel, K.F.; Riegel, R.M.; and Meyer, G. The prediction of retest-resisters in longitudinal research on aging. *Journal of Gerontology,* 23:370-374, 1968.

Rose, C.L. Representativeness of volunteer subjects in a longitudinal aging study. *Human Development,* 8:152-156, 1965.

Schaie, K.W. A general model for the study of developmental problems. *Psychological Bulletin,* 64:92-107, 1965.

Schaie, K.W. A reinterpretation of age related changes in cognitive structure and functioning. In: Goulet, L.R., and Baltes, P.B., eds. *Life-Span Developmental Psychology. Research and Theory.* New York: Academic Press, 1970.

Schaie, K.W. Methodological problems in descriptive developmental research on adulthood and aging. In: Nesselroade, J.R., and Reese, H.W., eds. *Life-Span Developmental Psychology. Methodological Issues.* New York: Academic Press, 1973.

Schaie, K.W.; Labouvie, G.V.; and Buech, B.U. Generational and cohort-specific differences in adult cognitive functioning: A fourteen-year study of independent samples. *Developmental Psychology,* 9:151-166, 1973.

Schaie, K.W., and Labouvie-Vief, G. Generational versus ontogenetic components of change in adult cognitive behavior: A fourteen-year cross-sequential study. *Developmental Psychology,* 10:305-320, 1974.

Schaie, K.W., and Strother, C.R. The effect of time and cohort differences on the interpretation of age changes in cognitive behavior. *Multivariate Behavioral Research,* 3:259-293, 1968a.

Schaie, K.W., and Strother, C.R. A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin,* 70:671-680, 1968b.

Solomon, R.L., and Lessac, M.S. A control group design for experimental studies of developmental processes. *Psychological Bulletin,* 70:145-150, 1968.

Streib, G.F. Participants and drop-outs in a longitudinal study. *Journal of Gerontology,* 21: 200-209, 1966.

Thomae, H. Theory of aging and cognitive theory of personality. *Human Development,* 13:1-16, 1970.

Winer, B.J. *Statistical Principles in Experimental Design.* New York: McGraw-Hill, 1962.

Wohlwill, J.F. *The Study of Behavioral Development.* New York: Academic Press, 1973a.

Wohlwill, J.F. The concept of experience: S or R? *Human Development,* 16:90-107, 1973b.

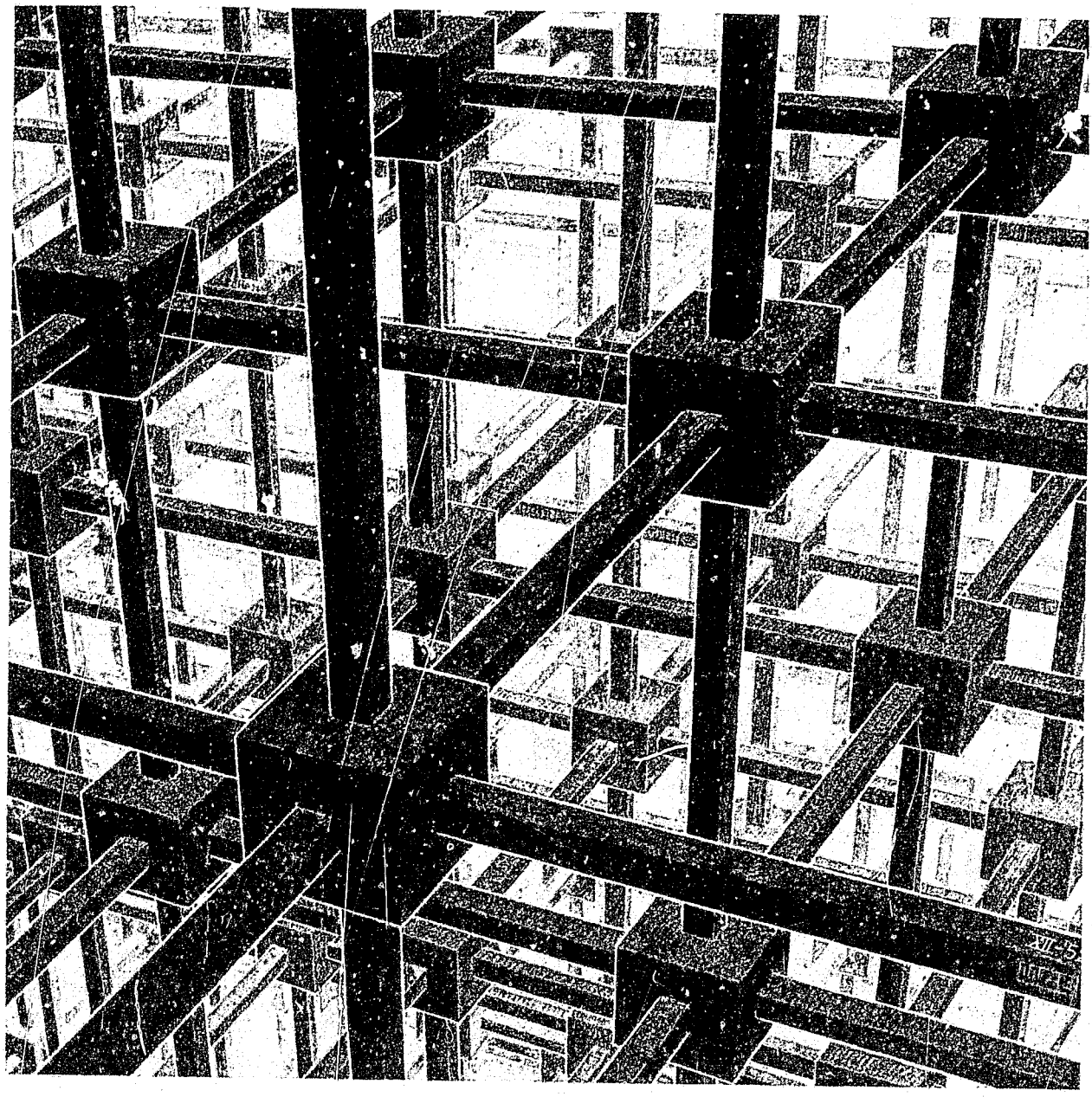


Plate 5

INTRODUCTION: USE AND ASSUMPTIONS[1]

Automatic Interaction Detection (AID) is one of the first analysis techniques designed for social science data that employs the decision-making capacity of a high-speed computer. AID is a computer program designed to scan, in a certain way to be described below, the relationship between a number of predictors and one criterion, or dependent variable. As is common in social science data, the predictors may be categorical in nature, but may also be, with grouping, ordinal or interval (see chapter 5 for a discussion of levels of measurement). The illustration included below demonstrates that AID may be used with longitudinal data (studies based on repeated measures) as well as with data collected at one point in time (see chapter 7). The purpose of AID is to identify a set of categories, defined in terms of combinations of predictors, that best explains variation in the criterion.

The remarkable capacity of the high-speed computer to act as an enormous storage and retrieval machine is well recognized, as is its ability to make lightning-fast calculations. Less well recognized by the user is its decision-making capability. As those who design computer programs to guide the operations of the machine know, without such capability the computer could not perform those more familiar functions.

The possibility of using this decision-making capacity in connection with matters other than the operations of the computer program, such as the analysis of a large body of data, has often been a subject of speculation, but to the present remains almost entirely unexplored. The innovative computer program known as Automatic Interaction Detection (AID) is one of the first of such programs, and apparently the only such program developed specifically for the analysis of social science data.

AID examines the relative importance of each of a set of independent variables in predicting a criterion, and conducts this examination without any assumptions of additivity or linearity. Especially when assuming additivity, other techniques that are commonly used in the social sciences to summarize a large body of multivariate data, such as factor analysis and multiple regression, overcome the need for making qualitative distinctions within the data. AID is an exploratory device to assess the homogeneity of the sample in the sense that relations between predictors and criterion are additive and not interactive. In subsequent regression analyses, the investigator can test the significance of interactions detected by AID (see chapter 10). The elementary decision-making involved in AID incorporates the idea of making a selection at one level of data analysis, and then pursuing the implications of this and subsequent selections on increasingly deeper levels of analysis.[2] Like many simulation models (which AID is not), and unlike factoring and regression techniques, the development of AID probably would not have occurred in the absence of stored-program, self-modifying computing machines.

Not surprisingly, in view of this approach, the decision-making capabilities reflected in AID are rudimentary. In comparison with the analogous capacities of the human analyst poring over a large body of data, they are extremely unsophisticated. This is especially true of the simpler version of AID discussed here.[3] Brief reference will be made to a more complex version, AID III, which incorporates the capacity for more sophisticated decision-making.[4] For purposes of this introduction, reliance on the simpler version is a better way to describe the basic logic and intent of both programs.

As a multivariate method, AID is intended for analyses of a number of independent variables (predictors) in relation to a single dependent variable or criterion. In most practical applications, an investigator will find it useful to have a fairly large sample size, such as more than 500 observations, although the program itself imposes no restrictions on sample size.

Note: The idea of the AID program described here originated with Prof. James N. Morgan and Prof. John Sonquist, then both at the Institute for Social Research, The University of Michigan, Ann Arbor, Michigan. References to their seminal work appear at the end.

Because AID makes no assumptions about the data in terms of measurement properties or additivity of the effects of predictors on the criterion, we find it useful to employ AID as an exploratory device prior to the utilization of multiple regression or partial correlation methods. With the latter, departures from additivity (interactions) must be anticipated in advance, while AID is expressly designed to identify important interactions. AID is sometimes thought to resemble stepwise regression analysis, since subsequent steps in both analyses are determined by the outcome of prior steps, but apart from this resemblance the methods are quite different. In the following, an example will be presented showing the utility of conducting an exploration with AID to suggest subgroups of the sample within which multiple regression, possibly of a stepwise nature, can be usefully employed.

AID analysis is intended principally for survey data, rather than data collected by more quantitative measurement procedures. Although there have been adaptations to categorical data with such procedures as "dummy variable" scoring in multiple regression, most multivariate procedures were developed with quantitative measurements in mind, and this is one of several ways in which AID is a most unusual multivariate method. While it is true that numerical information such as age in years and income in dollars is often obtained in surveys, investigators generally obtain such a predominance of categorical information that they find it convenient to cast even those quantitative variables in the form of a series of ordered categories.

In contrast to the flexibility regarding measurement assumptions in the predictors, AID requires interval measurement of the criterion. This includes the possibility of a dichotomous criterion, which may be analyzed as an interval variable utilizing (0,1) scoring. In many of our analyses, as in the one selected for illustration below, we employ a criterion that has been simplified in that way. If the criterion is ordinal, this means we need make no measurement assumptions. If the precise way in which that criterion should be dichotomized is unclear, we find it convenient to run two or more AID analyses, each using a different dichotomization, and compare the results.

In the following, we review in Section II, Rationale, the special problems of analyzing survey data which AID is designed to address. Then in Section III, Methods and Procedures, somewhat in the fashion of peeling layers from an onion, we consider the activities of the program and the basis for its decision-making operations in greater and greater detail. That section concludes with an example of the elegant tree-diagram, the final output of the version of the program that we use.[5] We then assess the elemental AID decision criterion from a number of different perspectives, and thereby gain a still more refined understanding of the method. In the fifth section we summarize some general limitations that should be kept in mind using this methodology. Finally, we illustrate the use of this program in a study of the correlates of marihuana use from one of the first general population surveys of that topic.

RATIONALE: PROBLEMS IN EXPLANATORY SURVEYS

The following discussion refers primarily to surveys that are intended to explain some aspect of social behavior or to assess its implications. Such surveys may be contrasted with those that are intended principally to describe the extent or character of some aspect of social reality. In explanatory surveys the major concern is to ascertain whether there is a relationship between one variable and another. While the aims of an explanatory survey are similar to those of a classical scientific experiment, the survey investigator does not have an opportunity to manipulate experimental conditions through random assignment. Consequently, this manipulation must necessarily be done through statistical procedures. Selection and use of these procedures presents special problems.

The first illustration presented below is drawn from a longitudinal survey that was explanatory in design.[6] As will be seen, the analysis capitalized on the longitudinal nature of the data. Even without the added complications of longitudinal data, explanatory surveys present many problems to the analyst.[7] Frequently it is necessary to observe a great variety of human behavior, presenting many variables for analysis. Often these variables are not quantitative measurements, but classifications which do not lend themselves easily to multivariate analysis, especially if they are rank-orderings.

Of special relevance for AID, survey analysts are often reluctant to assume that the behavior being studied is homogeneous in the following sense. Analyses of the sample as a unit, without allowing for the possibility that some subgroups within the sample behave differently with respect to particular variables, has often been found misleading. In studying the precursors of illicit drug use, for example, it is commonly found that the sources of illicit drugs are different for men than for women, and that variables describing the pattern of social influence

affecting illicit drug usage are correspondingly different according to the user's sex. This is one type of statistical interaction, and is one of the ways in which the sample may be heterogenous. Interaction presents special problems for analysis because it means the assumption of additivity often required in multivariate analysis is violated.

Interaction may be described in various ways and manifests itself with different degrees of complexity. In a careful discussion, different orders of interaction are distinguished. One way to describe interaction is to say that variables are not simply additive in their effect on a third variable, but that, instead, the effect is dependent upon the particular combination of values of the other variables. In some discussions of survey analysis, one of the variables involved in interaction is referred to as a "specifying variable." This designation is especially apt when second-order interaction is being discussed. In that case, one of the interacting variables specifies the conditions under which the other variable is related more or less strongly to the dependent variable. The version of AID being illustrated here is intended to locate only first-order interaction, which refers to variation in the mean value of the criterion among subgroups of the sample. Nevertheless, the example to be analyzed below will show that it is successful in locating this more complex kind of interaction as well.

To identify and judge the import of interaction patterns is a major problem for survey analysts. AID accomplishes the former better than the latter. Only rarely does a priori knowledge indicate which variables will be involved in interactions. Experience sugggests that sex and ethnic differences are specifying variables in many areas of social behavior, but other variables may play this role in a particular investigation. In a survey of college student populations, for example, academic major is likely to interact with other factors. AID is one of the few analysis techniques intentionally designed to identify interaction patterns such as this.[8]

METHODS AND PROCEDURES

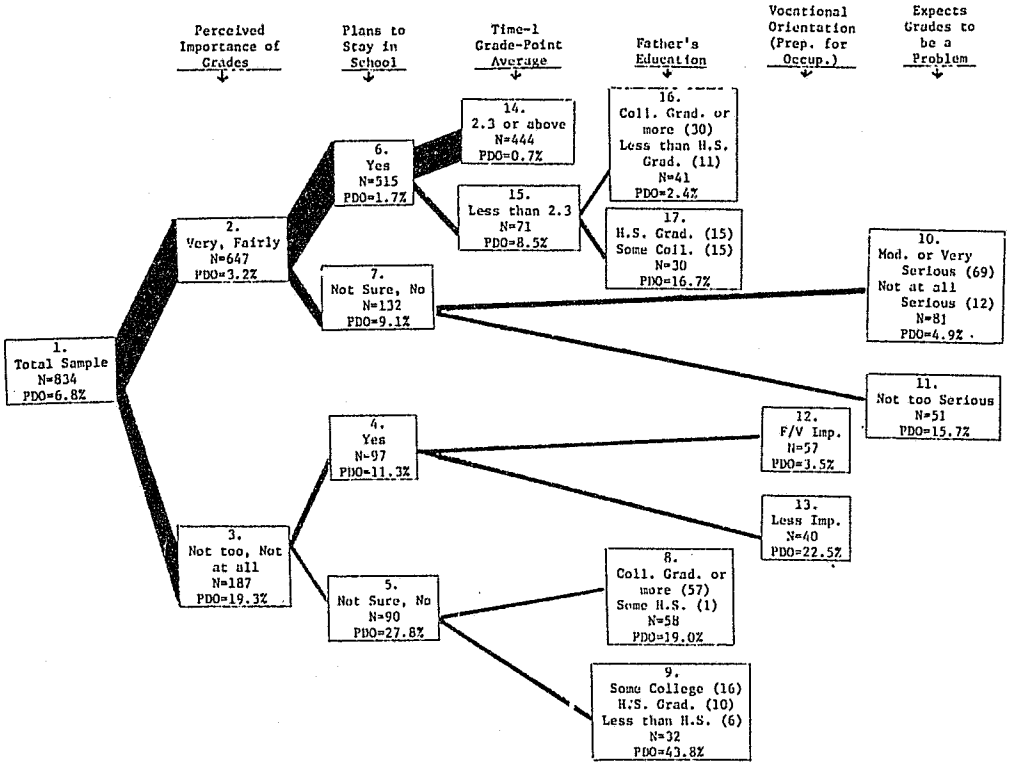
THE BASIC SEARCH PROCEDURE

On the basis of AID's scanning of the relationship between a number of predictors and one criterion, the program selects the one best way to divide the sample into two groups. "Best" means selecting a dichotomous partitioning which maximally explains variation in the criterion. Next, AID repeats this search and partitioning within each of the two subgroups, and continues operating in this fashion, generating and examining an increasing number of subgroups, until it reaches a preset indication to stop. In the following, the different repetitions of these twin operations of search and partition are referred to as levels of operation of the AID program. The outcome of one such analysis is presented in Table 1, and will be discussed in various ways in the next few pages.

The ultimate aim of AID at each level of operation is to account for variation in the dependent variable. Scanning all the predictors, AID identifies that predictor which permits the sample to be split into two subgroups in such a way that a maximum reduction in variation on the criterion is accomplished. Put differently, it splits the sample so as to minimize the unexplained variance. For example, in the analysis summarized in Table 1, we investigated factors that were predictive of male students becoming university dropouts. Our aim was to examine information obtained from them at the time they entered the university in 1970 to see what was predictive of their being in apparently permanent dropout (PDO) status two and one-half years later, in 1973. An AID analysis was conducted of this criterion in the hopes that it would identify subgroups in which the likelihood of becoming a permanent dropout was very high, in contrast to subgroups where the likelihood was very low. In the extreme, the analysis would have identified a subgroup--defined in terms of combinations of scores and categories of several predictors--that contained all of the dropouts. Had that happened, partitioning of the sample would have completely eliminated the residual variance. In the real world, of course, such perfect prediction is not to be expected. But AID did produce subgroups where the dropout rate ranged from 44 percent (group 9 in Table 1) to less than 1 percent (group 14).

At each level of operation, AID may be viewed as a variant of correlation analysis without the measurement assumptions that method requires. A common method of screening predictors is to examine their correlations with the criterion. At each level, AID does just that without requiring the predictors to be measured on an interval scale. As will be explained below, AID can be set to make no assumptions whatsoever about the measurement characteristics of those variables, or to maintain the categories in a particular rank order.

Table 1. Major Time-1 Predictors and Subgroups Emerging from AID Analysis of Permanent Dropout Status (PDO) 2½ Years after Entering University* (Nine other predictors screened in this analysis are shown in Table 6)



*Width of bars represents approximate proportion of total sample. Figures in parentheses in some boxes represent number of cases in indicated categories.

AID'S DECISION-MAKING STEPS

Using our dropout study as an example, we will now take a closer look at exactly how AID proceeds at one key element in those operations: where it is examining one predictor within one group. At the inception of the analysis, this would be the whole sample; at a later point, a previously isolated subgroup. Five steps are involved in operation:

1. Ordering the Categories
2. Selecting the Best Dichotomization
3. AID-Facilitated Review of an AID Decision
4. Moving to the Next Level
5. Termination of the Search

Ordering the Categories

The first step in the operation of the program for each predictor is to calculate, for each category of the predictor, the mean value of the criterion. In Table 2 are shown the categories of response to a question asking students, in the study cited above, how important it was to them to maintain a particular grade-point average. With the criterion employed here, whether or not a student became a permanent dropout two and one-half years after entering the university, the mean value for each category is simply the proportion of dropouts in that category. These proportions are shown in the form of percents in Table 2, where the categories are shown in the ordering arranged by the program and presented in the computer printout.

Table 2. Categories of a Predictor Ordered by AID in Terms of the Mean Level of the Criterion

Predictor: A question asked of entering male freshmen in 1970: "How important is it to you to maintain a particular grade-point average?"

Criterion: Percent Permanent Dropout (PDO) 2½ years After Entry into the University

Category Number	Response Category	% PDO	Number of Cases
4	Very Important	2.7%	222
3	Fairly Important	3.5%	423
2	Not Too Important	16.7%	138
1	Not At All Important	26.1%	46
5	No Answer	33.3%	3
	Total Sample	6.8%	834

Clearly there is a strong relationship between being sufficiently motivated to say that maintaining a particular grade-point average is very important and being enrolled in school (or only a temporary dropout) two and one-half years later. Fewer than 3 percent of those in the first category dropped out, whereas over one-fourth of those saying a particular grade-point average was not at all important had dropped out. As Table 1 shows, this variable provided the first split in the sample, suggesting AID made the choice on the strength of the relationship shown in Table 2. That is not the case, however. Intent on dichotomizing, the program explored this variable further.

Before turning to that next step, note that AID contains a convenient option whereby it may be instructed to maintain categories in rank order. Had that option been employed in this case, category 5 would not have been placed by the program at the bottom, but rather at the top, preceding category 4. In the present instance, we may infer from the high proportion of dropouts in category 5 that those who failed to answer the question were very unmotivated students, and it is clear that the category belongs at the bottom. Thus, AID may also be used to ascertain a plausible location for "no answer" categories.

Selecting the Best Dichotomization

An important aspect of AID's decision-making is that it does not select a variable for splitting a subgroup according to the strength of the overall relationship shown in Table 2, but rather the strength of the relationship after choosing the "best" dichotomization. According to AID, the best way of splitting the sample in two groups is that which maximally reduces residual variation. That residue is quantified by calculating the unexplained or aggregate within-group variance, which is the sum, over all observations, of the square of the distance separating each observation from the subgroup mean. Residual variance is zero if and only if all observations on the criterion are the same in each subgroup, in which case they coincide with the mean for that subgroup. In the present example, this would mean that all of the observations in each subgroup were either permanent dropouts (assigned a score of 1 for analysis purposes), or were not (in which case they were scored 0).

As is made clear in the analysis of variance, calculating the sum of squared deviations around a subgroup mean is quite different from calculating the sum of squared deviations around the mean value for the whole group. The latter quantity is called the total sum of squares (TSS). Taking a weighted average of the former (weighted by the size of the subgroups) yields the residual, unexplained or within-group sum of squares (WSS). The WSS is never larger than the TSS, and generally smaller. The difference between these two quantities is called the between-groups sum of squares (BSS). Hence, minimizing residual variation is the same as maximizing BSS. Discussions of AID are generally phrased in terms of the latter. When only two subgroups are being considered, this between-groups sum of squares can be simply expressed as a multiple of the square of the difference between the subgroup means. Designating subgroup sizes as N_1 and N_2, subgroup means as $\bar{X}_1$ and $\bar{X}_2$, with the whole group of $N_1 + N_2 = N$ observations,

$$BSS = \frac{N_1 N_2}{N} (\bar{X}_1 - \bar{X}_2)^2.$$

This calculation is illustrated in Table 3. It is apparent that BSS is in part a quantification of the extent to which the two subgroups differ in terms of the average value on the criterion $(\bar{X}_1 - \bar{X}_2)^2$, and in part a comment on the distribution of the predictor (N_1N_2/N). In the extreme, BSS would be maximized, and equal to the TSS, if all the dropouts were in one subgroup, and none in the other.

First ordering the categories by their mean values, AID investigates all pertinent dichotomizations of the predictor.[9] For each, it calculates BSS to ascertain the extent to which that way of subdividing the group accounts for variation in the criterion. Since there are five categories, there are four possible splits: group 4 from the remainder, groups 3 and 4 taken together and split from the remainder, and so on. Table 3 illustrates this calculation for the first two of the four possible splits for the predictor in Table 2.

The calculation assessing the merit of the first possible split is shown in the upper half of Table 3. Isolating group 4, containing 224 men, produces a BSS of .529, as illustrated. Calculations assessing the next possible split are shown in the lower portion of Table 3. Here groups 3 and 4 together are compared with the remaining groups, and the between sum of squares, with a value of 3.716, is much larger than before. AID continues this investigation over all possible splits. The results are shown in Table 4.

Table 3. Assessment by AID of the First and Second Possible Dichotomization of One Predictor (Figures are number of cases.)

Criterion:	First Possible Split: Group 4	Groups 1-3,5	Total
Permanent Dropout (PDO)	6	51	57
Not Permanent Dropout	218	559	777
Total number of cases	224	610	834
Proportion PDO	.0268	.0836	

$$BSS = \frac{N_1N_2}{N}(\bar{X}_1 - \bar{X}_2)^2$$

$$= \frac{(224)(610)}{834}(.0268 - .0836)^2 = 0.529$$

Criterion:	Second Possible Split: Groups 3,4	Groups 1-2,5	Total
Permanent Dropout (PDO)	21	36	57
Not Permanent Dropout	626	151	777
Total number of cases	647	187	834
Proportion PDO	.0325	.1925	

$$BSS = \frac{(647)(187)}{834}(.0325 - .1925)^2 = 3.716$$

Table 4. Between-Group Sum of Squares for Each Pertinent Dichotomization, for One Predictor

Dichotomized Between:	Between-Group Sum of Squares
Code 4 and Codes 1-3,5	0.529
Codes 3,4 and Codes 1-2,5	3.716 (maximum)
Codes 2-4 and Codes 1,5	2.020
Codes 1-4 and Code 5	0.211

Proceeding from this illustration as though this was the first predictor being examined, the remaining operations of the program can be briefly outlined. Since the second dichotomization in Table 4 (codes 3 and 4 versus codes 1-2,5) is that which maximally explains the criterion (BSS = 3.716), this information is stored by the program, which then investigates the next predictor in precisely the same way (in this case, plans to stay in school). For that predictor, the maximum between sum of squares, together with the particular split that maximized it, are again stored as summary information. Each predictor is examined in this way. When all predictors have been examined, the program selects the one with the largest maximum BSS, the group is split on this dichotomy, and the whole set of operations is repeated within each subgroup. This leads to four subgroups, and the whole set of operations is again repeated until certain stopping criteria are reached, as described below.

A remark is in order about the AID selection criterion, which thus far has been discussed as BSS, the between sum of squares. For some purposes it is useful to view the selection criterion as not simply BSS, but BSS/TSS, the between sum of squares taken relative to the total sum of squares on the criterion within the particular subgroup being examined. (In a more formal presentation, a subscript would be used to identify that subgroup.) Since the TSS for a subgroup is the same regardless of which predictor is being examined, AID would make the same selection regardless of whether BSS or BSS/TSS is taken as the selection mechanism.

Moreover, the ultimate goal of AID is to explain the total sum of squares for the whole sample, symbolized in our version of AID as TOTSS to distinguish it from the TSS for a subgroup. In the present example, with a dichotomous criterion, TOTSS may be easily calculated as $N_a N_b/N$ = (57)(777)/834 = 53.104, where N_a and N_b are the number of permanent dropouts and others, respectively. One might therefore consider BSS taken relative to this TOTSS as the AID selection criterion. Again, since TOTSS is a constant, it would not affect the selection process.

AID-Facilitated Review of an AID Decision

Most versions of AID print the information presented in Tables 2 and 4 for each predictor. While this is generally too much for the analyst to digest, it is helpful to have it available to review the decisions made by the program and to identify those that were made on the basis of a very small margin. For example, in Table 4 the maximum between sum of squares (3.716) is nearly twice as large as the next contender (2.020) which arose from combining groups 2, 3, and 4. Had the two values been nearly the same, the analyst might decide to override the decision of AID and instead choose the latter split. Such a decision to override might be based on the meaning of the categories. In the present instance the program made the decision to combine "very and fairly important", the same decision that an analyst is likely to make. In other instances such a happy coincidence might not occur.

Another procedure that we have found useful is to identify the set of predictors that were in contention for the one selected as the best split. We often annotate the AID tree-diagram with this auxiliary information. This is illustrated in Table 5. There are shown the five predictors presenting the largest maximum BSS when the group being analyzed is the whole sample. Thus a student's intention to stay continuously enrolled in college, plans expressed early in his freshman year, were second in importance in determining whether or not he dropped out.

Table 5. The Most Important Predictors Identified by AID in its Analysis of the Whole Sample.

Predictor Variables*	Max BSS	Max BSS/TOTSS
Importance of maintaining a particular grade-point average	3.7164	.06998
Plans to stay continuously in school	2.9246	.05507
Is at the University principally to prepare for an occupation	1.8045	.03398
Index of recent (past six months) illicit drug use	1.4659	.02760
Official freshman grade-point average	1.3210	.02488

*Observed soon after entry into the University.

An evaluation of the meaning of these results requires knowledge of the other variables entered as predictors. These are shown in Table 6. Quite clearly the analysis demonstrates that the student's expressed motivations were more important than either background factors or objective or subjective achievements and satisfactions. The two motivational variables--the importance of good grades and plans to stay in school--were the two most important predictors in the assessment of the whole sample by AID. Third in importance was his vocational orientation, fourth was his score on an index of recent (past six months) use of illicit drugs, and fifth, his freshman grade-point average as obtained from his university transcript.[10]

Table 6. Variables Entered as Predictors in the Illustration

Variables, by type	Number of categories
A. Family Background Factors	
1. Father's education	8
2. Ethnicity	7
3. Size of home community	8
B. Scholastic Ability	
4. Scholastic Aptitude Test score (verbal)	6
C. Experiences with Illicit Drugs	
5. Age when first used marihuana	6
6. Index of drug use in the year before college	5
7. Index of drug use in the freshman year	6
D. Academic Value Orientations	
8. Academic major	3
9. Importance of occupational preparation	4
E. Academic Achievement	
10. High school grade-point average	7
11. College grade-point average fall and winter quarter of freshman year	6
F. Academic Expectations and Satisfactions	
12. Expected to have problems with grades	5
13. Satisfied with quality of teaching	6
G. Academic Motivation	
14. Importance of grades	5
15. Planned to stay in school	4

Moving to the Next Level

As explained earlier, after achieving the best dichotomization of the whole sample, AID repeats exactly the same search and partition operations in each of the two resulting subgroups. It continues these operations in each such generated subgroup until reaching preset criteria for stopping. The continuation of this process of examination and generation of subgroups can become complex. Our version of AID helps to reduce this complexity by presenting, at the con-

clusion, the pictorial diagram shown in Table 1. The similarity between Table 1 and a common pictorial representation of a biological classification system is not surprising, since AID is pursuing a form of analysis that is logically identical to that type of differentiated classification.

Of particular importance to the analyst of social behavior, AID identifies interaction patterns by suggesting which variables play an important role within each of the subgroups that it previously identified, as well as in the whole sample. As always with AID, this assessment is made in terms of the role these variables play in predicting the criterion.

This may be illustrated by returning to Table 1. As will be explained below, the data presented in Table 1 are unusual in that AID selected the same variable on which to split groups 2 and 3. They were selected for illustration precisely because of this unusual outcome.

Suppose the outcome of the second level (assessment and splitting of groups 2 and 3 in Table 1) had shown father's education to be important among less motivated students but not in the remainder of the sample. Then an analyst who examined only within the sample as a whole the relation between father's education and permanently dropping out would have found a much smaller relationship between those two variables. This is especially true since relatively few students had low motivations. Without the proper specifying variable, in other words, the relationship would have been greatly attenuated, probably to such a degree that father's education would not have distinguished itself from a number of other candidates that were being considered as predictors.

Termination of the Search

Important to AID's decisions are its stopping criteria, which apply at several levels of operation of the program. Thus, an entire run may terminate for one of several reasons, or within a run AID may refrain from examining a subgroup that fails to meet certain criteria. Within a subgroup, it may fail to select a predictor that falls below certain standards. An entire run will be terminated either when a maximum number of groups, specified by the user, has been reached, or when none of the following criteria are satisfied.

Short of terminating an entire run, AID will refrain from examining a subgroup that contains too few cases or too small a proportion of the variance of the dependent variable. In employing this last criterion, AID utilizes a further partitioning of the sum of squares. We saw earlier that the total sum of squares may be partitioned into the within (WSS) and the between sum of squares (BSS). In addition, the WSS may be further partitioned into the sum of squares around each of the two subgroup means. The portion of the WSS associated with a particular subgroup becomes the TSS for that subgroup, and is tested for size relative to TOTSS.

Finally, if a subgroup satisfies these criteria, AID may refrain from dichotomizing if the maximum BSS for that subgroup, over all predictors and possible dichotomizations, fails to produce a significant difference by the conventional t-test (see below), or is too small a proportion of the total sum of squares for the whole sample. In addition, a split will not be made if either of the resulting subgroups is too small.

Experienced users of AID have their own conventions regarding these several criteria. We often set the maximum number of groups at 30, the significant level of the t-test at .05, and the minimum size of a subgroup to be examined at 30; but adjustments may be made in particular runs.

OPTIONAL MODIFICATION OF DECISION PROCEDURES

Especially with the newer version of AID referred to earlier, there are several options which may be employed to vary the procedures followed by the program.

Specification of Measurement Properties

Even with the simpler version of the program, measurement assumptions may be imposed on a variable in the fashion referred to above. Variables may be constrained to retain a particular ordering of the categories so that AID will never present combinations in which that ordering is violated, regardless of criterion means. Otherwise variables are considered unconstrained in their ordering, and the categories of each variable will be analyzed in the order that corresponds to the ordering of mean values of the criterion, as shown in Table 2. If variables are unconstrained, it may be well to avoid categories with a very small number of cases, since they do not provide reliable estimates of criterion means.

More Sophisticated Decision-Making Features

The newer version (AID III) incorporates features that help it make more sophisticated decisions. They can be only briefly mentioned here. For one, this version of the program permits the analyst to specify particular ways of subdividing the sample in the first few levels of its operations. In this way he can take advantage of prior knowledge about subgroup differences. It also permits the analyst to specify that splits be made from one set of predictors before they are made from a second set. Using the earlier version, it is necessary to pre-sort the observations and run separate AID analyses.

One especially impressive way in which the newer version of AID is more sophisticated is that it permits the investigator to indicate a preference for symmetric splits. In a symmetric split, two subgroups on one level are split on the same variable, suggesting the absence of this type of interaction. The example in Table 1 was selected in part because it is an unusual illustration of such symmetry, since groups 2 and 3 were both split on whether or not the student planned to remain continuously enrolled.

As noted earlier, interaction is commonly observed in survey data. However, some of this interaction, especially as subgroups get small, may be only chance fluctuation in criterion means, being a consequence of the particular sample that happens to have been drawn. By permitting the analyst to express a preference for symmetry, the more sophisticated version of AID allows one to override some of these idiosyncracies. The newer version of AID also allows an investigator to introduce an intervally-scaled (or dichotomous) "covariate," a predictor that is known to be strongly related to the criterion, in order to assess differences in the slope and intercept of subgroup regressions.

Finally, the newer version of AID incorporates a "lookahead" feature, which means that it automatically explores, according to certain specifications, alternative subdivisions of the sample and their implications. Since the later results of an AID analysis are heavily dependent on the choices it made at an earlier stage, this is obviously a useful feature. Both of these last two features of the newer version require a more extensive discussion than is possible here, and the reader is referred to the material on AID III cited earlier.

FURTHER DISCUSSION: THE BENEFITS OF AID

Earlier we remarked that AID is a useful preliminary screening device to identify components of the sample where interactions occur. The preceding illustration is a good example of this, since AID identified the fact that the second motivational variable (plans to stay in school) was important both in group 2 and in group 3. Thus the "plans" variable did not interact sharply with the first motivational variable. Had it done so, AID would probably not have split on the second variable in both of those groups. Among other things, this assured us that the use of an index of these two variables across the whole sample was not misleading.

On the other hand, AID identified the fact that the third variable (Time-1 grade-point average) did interact with the motivational variables. Low grade-point average (in contrast to an acceptable grade-point average) was especially predictive of dropping out for the most motivated student. This is the more typical result from AID, suggesting that regression analysis involving Time-1 grade-point is most suitably employed within the subgroup of more motivated students. We will examine this conclusion more explicitly in a moment.

Among other things, the economy of AID is impressive. Making a rough guess, it would have taken a clerk about a quarter of an hour to calculate the means for one predictor, order the groups, calculate the between sum of squares for each of the four possible partitionings, and check his work. For a set of calculations such as this, repeated for the whole sample, for the two resultant subgroups, then for two subgroups in each of those, and finally for two subgroups in each of those, one may estimate that it would take the clerk over fifty hours to analyze 15 subgroups. In contrast, the computer required only 32 seconds and (at our noncommercial processing rates) $3.38 to perform this analysis of 17 groups.

A DETAILED ASSESSMENT OF ONE AID DECISION

In the example presented above, AID concluded that freshman grade-point average was of major importance in determining whether or not a young man dropped out of the university, provided the young man was a relatively motivated student. For less motivated students, other factors came to the fore. While it is unlikely that we would have looked in advance for this particular inter-

action, once such a suggestion is obtained from AID, we can make a direct examination of the data to see what they show. Following this example a few steps further provides a useful opportunity to examine some of the fine details of the operations of AID.

Table 7 contains data constructed especially to check the above AID conclusion. The left panel of Table 7 shows the difference in dropout rate for motivated students according to whether or not their freshman grades were below 2.3. (A grade of 4.0 is A in this scheme, and a grade of 2.0 is C.) It is apparent that among motivated students there are essentially no permanent dropouts for the 444 whose grades were 2.3 or better as freshmen. Among those 71 whose grades were lower, 8.5 percent dropped out. This latter figure is not unusally high in comparison with less motivated students, shown in the center panel of Table 7, but it is remarkably high for motivated students.

Table 7. Relation of Freshman Grades to Permanent Dropout Status 2½ Years Later, Controlling Academic Motivation

	Motivated Students* GPA (Time-1)		Less Motivated Students GPA (Time-1)		Total Sample GPA (Time-1)	
	Less than 2.3	2.3 or greater	Less than 2.3	2.3 or greater	Less than 2.3	2.3 or greater
% Permanent Dropout	8.5%	0.7%	21.5%	12.9%	15.3%	5.0%
(Number of cases)	(71)	(444)	(79)	(240)	(150)	(684)

*Saying that they planned to stay continuously enrolled in school and that maintaining a particular grade-point average was fairly or very important.

The center panel for Table 7 shows that, contrary to what one might conclude from the AID results, even among less motivated students, there was a substantial difference in the dropout rate according to the level of freshman grades. Nearly twice as many of these men with grade-point averages less than 2.3 were permanent dropouts, as compared with those receiving better grades (21.5 percent vs. 12.9 percent).[11] Finally, in the right panel of Table 7 is shown the relation between Time-1 grades (as dichotomized here) and being a permanent dropout for the whole sample. Even in this last case, there is a fairly substantial relation. This suggests the fruitfulness of examining the operations of AID somewhat further to see why Time-1 grades were ascribed such important status as a predictor only among motivated students.

The Subgroups Examined by AID

Reference again to the tree diagram in Table 1 shows that AID never examined the large group of men presented in the center panel of Table 7 as less motivated (319 cases). Rather, AID examined these men in three separate subgroups, identified as groups 4, 5 and 7 in Table 1. Consequently our investigation of the operations of the program must consider these groups separately. Group 7 is shown first in Table 8. There data are presented for young men with the following motivational structure: they felt it fairly or very important to get good grades, but they were not certain of staying in the university continuously during the next few years. AID found the three predictors shown in Table 8 to be more important for this group than their Time-1 grade-point average. Those three predictors (all Time-1 observations) are: whether or not they expected that keeping their grades up would be a serious problem, the education level of their father, and whether or not they were satisfied with the quality of teaching at the university. The maximum relative between sum of squares (BSS/TSS) for each of those predictors can be seen from Table 8 to be larger than for Time-1 grades. (TSS here is for this subgroup, not the whole sample.)

Table 8. Predictors Found by AID to be Superior to Time-1 Grade-Point Average Among Three Groups of Less Motivated Students

Group Number	Motivational Structure: Grades	Motivational Structure: Plans	Predictors (Max BSS/TSS): Time-1 GPA	Predictors (Max BSS/TSS): More Important Predictors
7	Important	Time Off	.021	Expects Grade Problems (.033) Father's Education (.027) Satisfactory Quality Teaching (.024)
4	Not Important	Plans to Stay	.039	Preparing for Occupation (.087) Size Home Community (.083) Age First Used Marihuana (.057)
5	Not Important	Time Off	.046	Father's Education (.070) Drug Index, Year Before College (.058)

The situation is the same for groups 4 and 5. In other words, although AID did not ignore the relation that existed in these other groups between Time-1 grades and dropping out, it found other variables to have a stronger relationship with dropping out. Hence, the subsequent splits were made on the most important of these other variables--expectations regarding grades for group 7, vocational orientation in group 4, and father's education in group 5.

The Measure of Relationship Employed by AID

There is a related issue. The between-group sum of squares used as an assessment procedure by AID consists of two components, as we saw earlier. In part it is a function of the difference between criterion means in the two subgroups, the mean-difference. The other portion of BSS is basically a reflection of the structure of the sample, and requires special consideration. We first consider the mean-difference.

The Mean-Difference. The mean-difference is illustrated in Table 9 for group 7 of the example being analyzed. In Table 9 are shown the four most important variables for group 7, together with the particular split for each variable that produced a maximum reduction in variation in the criterion.[12] That is, the expectation by students that the problem of keeping up their grades will be "not too serious" identified a group with a dropout rate of 15.7 percent. The remainder of group 7 had a dropout rate of 4.9 percent, producing a difference of 10.7 percentage points, shown in the third column of Table 9.

Values of the mean-difference for the three predictors selected as next most important by AID are also shown in Table 9. The predictors are ordered in terms of the BSS, which corresponds to the ordering of the proportion of explained variance, BSS/TSS, since TSS is the same for each assessment in this group, as described above. (It has a value of 10.91 here.)

It is of special interest for the present discussion that the values of the mean-difference shown in Table 9 are ordered in the same way as the BSS/TSS shown in Table 8. This need not be the case, since BSS contains another factor. Before considering that factor, note that the mean-difference is the basic ingredient in two common forms of statistical analysis.

The Mean-Difference and Regression Analysis. From a regression viewpoint, if there are, as here, only two values of the independent variable but an array of observations on the dependent variable for each of those two values, then the difference in the array means on the dependent variable is the slope of the regression line. Hence, we may interpret the mean-difference, which in this example is a difference between proportions, as a regression slope. Had there been more than two values of the independent variable, as is usually the case in regression, then the line that minimizes residual or within variation would generally not, as here, coincide exactly with the array means, since they rarely lie on a straight line. A "standardized" regression coefficient is sometimes employed in multiple regression analysis. Its concept of standardization has relevance for this discussion, as will be seen in a moment.

The Mean-Difference in the t-test. A t-test is a statistical test used to determine whether the means of two independent samples from populations with the same variance are significantly different from zero. In the test, the mean-difference, $\bar{X}_1 - \bar{X}_2 = D$ is the observation to be tested. It has a sampling variance, σ_D^2, which is the sum of the variances of the two independent means,

$$\sigma_D^2 = \frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}$$
$$= \sigma^2 \left(\frac{1}{N_1} + \frac{1}{N_2}\right)$$
$$= \sigma^2 \left(\frac{N}{N_1 N_2}\right),$$

where σ^2, the population variance, is ordinarily estimated by a quantity, $\hat{\sigma}^2$, that can be expressed in analysis of variance terminology in keeping with the rest of this discussion:

$$\hat{\sigma}^2 = \frac{WSS}{N-2} .$$

The statistic for this test, which may be referred to a table of the t-distribution, is the ratio of the observation to its standard error,

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{WSS}{N-2}\right)\left(\frac{N}{N_1 N_2}\right)}}$$

One of the stopping criteria employed by AID is the nonsignificance of a t-test in relation to a particular split.

Table 9. Ingredients of the Between-Group Sum of Squares (BSS) for the Most Important Predictors in Group 7

		Proportion PDO	(N)	Ingredients of BSS: Mean-Difference*	Ingredients of BSS: Value of $\frac{N_1 N_2}{N}$
A.	Expects Grades to be a Problem				
	Not too serious	.157	(51)		
	Not at all, moderately, very serious	.049	(81)	.107	31.3
B.	Father's Education				
	H.S. grad., some college, college grad.	.135	(76)		
	Less than H.S. grad, post-graduate study, other, DK, NA	.036	(56)	.096	32.2
C.	Satisfied with Quality of Teaching				
	Fairly satisfied, somewhat dissatisfied	.121	(91)		
	Very dissatisfied, very satisfied, undecided	.024	(41)	.096	28.3
D.	Grade-Point Average, Time-1				
	Below 2.70	.138	(58)		
	2.70 or above	.054	(74)	.084	32.5

*The mean-difference, here a difference in proportions since the criterion is dichotomous, is squared before entering into BSS. The mean-difference here differs slightly from the difference in proportion PDO because of rounding error.

The Relation of BSS/TSS to the Mean-Difference. As we have seen, AID may be interpreted as employing BSS/TSS as a screening device in its selection of the best dichotomized predictor. Since it measures the proportion of variance on the criterion in that subgroup that is explained by that dichotomous predictor, BSS/TSS does not seem to require further interpretation. Yet because the mean-difference (or its square) alone appears to be a natural measure of relationship, the question arises as to what role is played by the ratio N_1N_2/N in BSS/TSS.

The overall aim of AID is to account for variance in the dependent variable. Hence, it employs a selection criterion that reflects not only the extent of relationship as measured by the regression slope, but also the distribution on the independent variable. A basic fact of least squares analysis is that, other things being equal, as the variance of the independent variable declines from maximum dispersion, represented in this case by a 50-50 split of the sample into the two categories of the independent variable, then a given regression slope accounts for less of the variation in the criterion. Put differently, to account for a given proportion of the variance, the mean-difference must be larger as a dichotomized independent variable increasingly departs from a 50-50 split.

That the ratio N_1N_2/N is the correction factor reflecting this may be seen as follows. Defining variance over the criterion variable as $TSS/N = \sigma_y^2$, then the proportion of explained variance may be expressed as

$$BSS/TSS = \sigma_x^2 \left[\frac{(\bar{X}_1 - \bar{X}_2)^2}{\sigma_y^2} \right]$$

where $\sigma_x^2 = (N_1/N)(N_2/N)$, the correction factor divided by N, may be seen as the variance of the predictor. (The extra factor of N in the denominator complements dividing TSS by N.) This variance is a maximum of $\frac{1}{4}$ for a 50-50 split on the independent variable, and declines to 0 if all observations fall in one category of that variable

The way this variance acts as a correction factor may be illustrated in connection with the variable "expects grades to be a problem" in Table 9. There the mean-difference is 0.107. With 12 permanent dropouts among these 132 men, the variance to be explained in the criterion is $(12)(120)/(132)^2 = .0826$. The ratio of the square of the mean-difference to this variance is 0.1385. Applying to this a correction factor of .2371 obtained from the observed values (see Table 9) of $N_1 = 51$ and $N_2 = 81$, the observed $BSS/TSS = .033$ is obtained.

Had there been less variance on the independent variable, such as $N_1 = 26$ and $N_2 = 106$ (only about half as many [illegible]ying "not too serious"), then the correction factor would be .1582, and the proportio[illegible]ained variance would be reduced to 0.022.

The Correlation Coefficient and the Correlation Ratio. We noted above that the mean-difference in this instance of a dichotomized predictor may be viewed as a regression slope. Designating that slope b_{yx} for independent variable X and dependent variable Y, a relationship ordinarily appearing only in multiple regression may be identified.

In multiple regression situations one examines the relation of X to Y with one or more variables, Z, controlled, and quantifies that relationship with a partial regression coefficient $b_{yx.z}$. Some analysts then standardize this coefficient for differences in the variance, σ_x^2 and σ_y^2, of the independent and dependent variables respectively, by calculating a standardized partial regression coefficient, also called a beta-weight, $\beta_{yx.z} = b_{yx.z}\sigma_x/\sigma_y$.

In the case, as above, where the relationship between X and Y is being measured without controlling other variables in a regression sense, the standardized regression coefficient reduces to the ordinary correlation coefficient. This may be seen by defining σ_{xy}, the covariance of X and Y, which is the numerator of both the correlation coefficient and the regression slope:

$$\beta_{yx} = b_{yx}\sigma_x/\sigma_y = \frac{\sigma_{yx}}{\sigma_x^2}\frac{\sigma_x}{\sigma_y}$$

$$= \frac{\sigma_{yx}}{\sqrt{\sigma_x^2\sigma_y^2}} = r_{xy}.$$

By the usual interpretation, r_{xy}^2 = BSS/TSS, hence the "correcting" of the mean-difference by the variance of the independent variable may be seen as analogous to the standardization of a partial regression coefficient.

Finally, for this situation where the independent variable is dichotomized, BSS/TSS may also be identified as the square of the correlation ratio, commonly designated ε^2. When the independent variable has only two categories, there is no reason to differentiate r^2, which measures linear correlation, and ε^2, which measures closeness of fit to a line of any form.

Thus AID may be examined from a number of different approaches to the statistical summarization of relationships. Except for the distinction between b_{yx} and β_{yx}, that is, between a measure of relationship (b_{yx}) which ignores variation in the independent variable, and one (β_{yx}) which heeds that variation, all of these approaches coincide in this simplest of situations where a relationship may be assessed--that in which the independent variable takes on only two distinct values.

The Relation to Analysis of Variance. By now it is apparent that there is no reason for a discussion of AID to employ analysis of variance terminology at any one point in the dichotomized predictor selection process. The phraseology of t-testing would be more appropriate, since a one-way analysis of variance ordinarily involves more than two categories in the independent variable. If such analysis is carried out for only two groups, the resulting F-ratio of between to within mean squares is simply the square of the t-statistic described above.

But the overall aim of AID is to select categories of combinations of predictors that maximally account for variance in the criterion. AID may be viewed as a technique that chooses the "best" set of categories for a one-way analysis of variance. In view of the overall intent of the program, analysis of variance terminology is desirable.

In the example of Table 1, nine "ultimate" categories were selected for this purpose. The AID program calculated that the final, cumulative BSS/TOTSS = .165, hence these nine ultimate groups account for 16.5% of the variance in the criterion variable. The accumulation of this explained variance is shown in Table 10. Considering all possible ways of dichotomizing this set of predictor variables, including subsequent re-dichotomizations of subsets of categories remaining in predictors dichotomized at an earlier stage, and all possible ways of combining these dichotomized categories, this BSS/TOTSS is close to a maximum possible value.[13]

Table 10. The Accumulation of Explained Variance (BSS/TOTSS) with Successive Partitionings*

		Individual Group			Cumulative	
Step	Parent Group	BSS	BSS/TSS	BSS/TOTSS	BSS	BSS/TOTSS
1	1	3.716	.0700	.0700	3.716	.0700
2	3	1.261	.0434	.0238	4.978	.0937
3	2	0.567	.0280	.0107	5.544	.1044
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
8	15	0.351	.0638	.0066	8.741	.1646

*The BSS shown here is the maximum between sum of squares for the indicated group and the basis for the splitting of that group. TSS is the criterion sum of squares for a particular subgroup, while TOTSS = 53.1043 is the same quantity for the whole sample. This run terminated at step 8 after reaching a maximum preset number of groups (16).

A Comment on AID's Ordering of Categories

One last remark is in order about the results presented in Table 9. The analyst in this case made the choice to enter these predictors without constraining the order of the categories, as is evident from the categories presented for the first three of these variables. That decision is sometimes made, as here, even when there is an obvious rank-ordering among categories, in order to learn whether there are monotonic relations between predictors and criterion.

Consider the behavior of father's level of education. There the highest dropout rate is found among students whose fathers had an intermediate level of education for this population, who graduated from high school, had some college, or were college graduates. It is lower among those young men whose fathers were less than high school graduates, as well as those whose fathers had post-graduate education. While the number of cases is too small to draw any definite conclusions, it is possible that there are systematic reasons for this lower level of dropping out at both extremes. Had the analyst instructed AID to keep father's level of education in rank order, the opportunity to examine this possibility would have been missed.

Moreoever, it is clear from an examination of groups 8, 9, 16 and 17 in Table 1 that AID provides many opportunities to search for independent replications of such suggestions. However, it is also clear that even with the relatively large sample used in this analysis, AID rather quickly yields subgroups that are small enough so that one suspects either measurement or sampling error is responsible for many of the differences observed within them.

CAUTIONS

The preceding discussion has somewhat idealized the analysis process. One should not be deceived by the neatness of the diagram presented in Table 1. In a rudimentary sense it is true that AID is "essentially a formalization of what a good researcher does, slowly and effectively, but insightfully on an IBM counter-sorter."[14] But the decision-making processes involved in AID are necessarily quite unsophisticated. Even with the additions incorporated in the newer version, it would be a mistake to assume that data analysis has now become mechanical. AID is an extremely useful and ingenious tool, but it is no substitute for intensive investigation and thoughtful exploration of the data. As has been shown, some of these further explorations are assisted by the AID printout.

We have already remarked that AID, as a decision-making entity, makes few assumptions about the data. While this is a refreshing contrast to most other multivariate methods, it has disadvantages. One of these is that it takes very literally every observed value that is presented to it. Thus if one predictor is found to be only a shade better as a basis for subdividing the sample than another predictor, it takes the first and disregards the second. In other words, at this point it ignores problems of sample and measurement error. If one were to repeat an AID analysis of two random samples from the same population, one could very easily get quite different results. This is especially true since the decision to subdivide one group has a crucial bearing on later decisions.[15] Thus, in the dropout study, the difference between the maximum BSS for the first two predictors, 3.7 and 2.9, may represent only random sampling fluctuations. In the present instance we were not particularly concerned with this question because the two variables were both indicators of academic motivations; and these results, together with those obtained from a factor analysis (see chapter 9), clearly suggested the usefulness of an index of academic motivations based on these and other items.

With a sufficiently large sample size, an investigator may wish to randomly divide his sample in two and conduct the same AID analysis in each random half. In this way he can obtain a crude estimate of the extent to which his results are subject to random fluctuation.

Earlier it was noted that an interaction detected by AID can be tested for significance in a subsequent regression analysis. Little or no information is available regarding the likelihood that an interaction deemed important by AID will be found significant by regression analysis, although it seems likely that this will be the case. Similarly, it is not clear that an interaction found significant in a regression analysis would be identified as an important "split" in AID, although again this seems likely to be the case.

In general, we do not feel overly constrained by the actual magnitude of the results of an AID analysis and do not consider the lack of sampling-error statements to be a major problem. We advocate exploratory analyses of alternative possibilities. Where the sample size permits,

parallel analyses in randomly selected halves of the sample may be informative. Some observers are more uneasy. Like stepwise regression, AID has been criticized on the grounds that it capitalizes on chance results presented by a particular sample.[16] We view AID as an extremely useful exploratory device which must be interpreted in the light of this limitation.

ILLUSTRATIVE APPLICATION IN DRUG RESEARCH

The AID program was found to be a useful analytic device in a report of two sample surveys that, in the context of questions about psychotherapeutic drugs, also asked members of the general public 18 years and over about their use of marihuana. These surveys, conducted in San Francisco and portions of nearby Contra Costa County, took place in the late 1960s and may have been the first to pose questions to the general public about their use of marihuana. At the time, it was not clear that people would be willing to respond, since the use of marihuana was illegal. A few years later, the question became much less sensitive.

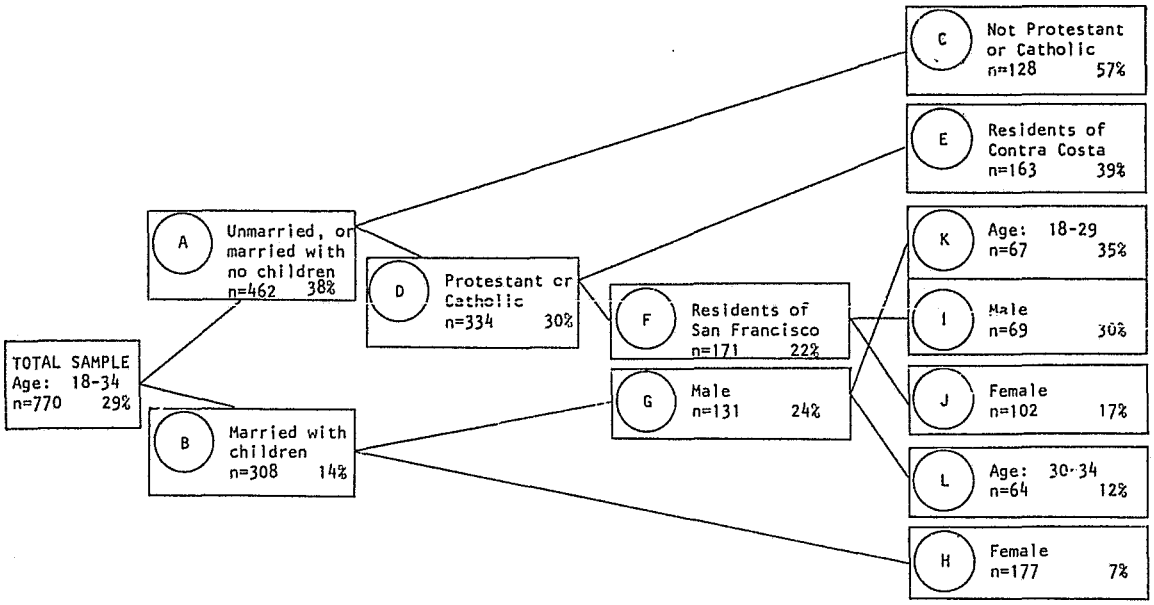
In contrast to the heterogeneous urban population sampled in San Francisco, the sampled portions of Contra Costa County contained mostly middle class suburbs. To the surprise of the investigators, the rates of reported marihuana usage were about the same in both locales. In the general population, those rates were 14% in San Francisco and 12% in Contra Costa. In the age group 18 to 34, 29% of both samples had used marihuana. The similarity in marihuana use rates in the two locales was especially surprising in view of the fact that social characteristics of users appeared to be quite similar in the two samples, while social characteristics of the general population in the two communities were quite different. For example, unmarried people were more likely to have used marihuana both in Contra Costa and San Francisco, yet a much smaller proportion of Contra Costa residents were unmarried. As a consequence, analysis of the data had two principal objectives, for both of which AID was used: (1) to determine what combination of variables discriminated best in terms of marihuana use; and (2) to search for an explanation of the apparent paradox of similar use rates in the two locales.

The first AID analysis was a search for that combination of demographic characteristics that would best differentiate among groups with various probabilities of marihuana use. San Francisco and Contra Costa were combined for this analysis, and locale was used as one of the potential predictor variables. A second AID analysis attempted to estimate the extent to which attitudes and values, characteristics less stable than those of a demographic nature, could add to the explanatory power of the latter characteristics. Thus, for example, respondents were asked about their attitudes concerning prescription drugs, whether they smoked cigarettes, how frequently they visited a doctor, questions about their use of alcohol, and questions which would permit them to be classified on a "stoicism" scale of personal values. In the following, only the first AID analysis will be reported. The reader interested in results of the second analysis may consult the original source.[17]

The results of the first analysis are shown in Table 11. The first split divided the sample into two primary groups: persons without children--in many cases because they were unmarried--and married persons with children. The former group had higher marihuana usage rates than the latter. The latter group then subdivided into males and females, with males having a higher rate of usage. The group of males subsequently split on age, with men under 30 having higher rates than older males. The other primary group, persons without children, first split on religious affiliation, with Protestants and Catholics having lower use rates than others. Among Protestants and Catholics, locale made a difference, with residents of Contra Costa County having higher use rates than those of San Francisco. Other details can be found in Table 11.

It is clear from these results that locale was one of the differentiating variables, but only within certain subgroups. The surprising equality of overall use rates was explained by examining particular subgroups. The key group of users comprised those adults who were childless and for the most part unmarried. Within this group, a major portion of the puzzle is untangled by noting that among San Franciscans the relationship of marihuana use to church membership was much more pronounced than among Contra Costans. Sixty percent of the unmarried or childless San Franciscans had used marihuana if they were not church affiliated (Protestant or Catholic), while only 22% of them had used marihuana if they were so affiliated. The comparable figures in Contra Costa were 50% and 39%. This is shown in Table 12. Table 12 also shows that among those persons with children, religious membership was related differently to marihuana use. Among these persons, religious membership showed a greater relationship to marihuana use in the suburbs rather than the city. This is still another example of interaction detected by AID--in this case interaction between locale and religious membership. But this particular interaction contributed only slightly to an understanding of the apparent contradiction, since relatively few persons with children were unaffiliated with either the Protestant or Catholic church.

Table 11. Patterns of Demographic Characteristics Associated with Marihuana Use (San Francisco and Contra Costa suburb)*



*The n's shown are the unweighted number of cases in each group. The percentages shown are based on the weighted n's; weights were based on (1) differential sampling rates of individuals within households and (2) standardization of the size of 18-to-34-year-old age groups.

Table 12. Interaction Between Religious Membership and Locale in Marihuana Use (from Table 11)

	San Francisco		Contra Costa	
	% Who Used Marihuana	% Of Cases+	% Who Used Marihuana	% Of Cases+
Married with children:				
Protestant or Catholic	12%	22%	11%	36%
Other or none	16%	5%	26%	11%
Unmarried or childless married:				
Protestant or Catholic	22%	48%	39%	42%
Other or none	60%	25%	50%	11%
Total percent		100%		100%
(Number of cases)		(346)		(424)

+Distribution of cases after weighting for differential sampling rates of individuals within households.

"In summary," the authors of this paper note, "utilization of the AID procedure clarified the information gained from looking solely at zero-order correlates of marihuana use. The patterns of relatively stable demographic characteristics that were associated with significantly varying levels of marihuana use reflected differences between city and suburb, as well as interactions among the major correlates." This is a good summary statement of the utility of AID, and a good illustration of its use in a practical research problem.

NOTES

[1]The authors are indebted to Professor Ira Cisin of the Social Research Group, The George Washington University, and the Human Population Laboratory, Department of Public Health, State of California for many helpful suggestions.

[2]The subroutine BREAKDOWN contained in the package of computer programs for social scientists known as SPSS (Norman H. Nie, et al., 1975, p. 249) is one example of what AID would look like if it did not incorporate decision-making capacities. BREAKDOWN provides essentially the same information that is presented by AID, but requires the analyst to determine in advance the variables and the partitionings of those variables that he wishes to see.

[3]See John Sonquist and James Morgan, *The Dectection of Interaction Effects*, Ann Arbor: Institute for Survey Research, Monograph No. 35, 1964.

[4]The newer version of AID is described in John Sonquist, et al., *Searching for Structure*, Ann Arbor: Survey Research Center, 1973.

[5]Not included in the original version of AID, credit for adding this option to the program belongs to Mr. Robin Room, the Social Research Group, School of Public Health, University of California, Berkeley, California. Various other modifications have been incorporated in other versions of the program.

[6]This study was supported by the National Institute on Drug Abuse (PHS Research Grant DA 00137). The research was conducted at the Institute for Research in Social Behavior, an independent, nonprofit organization located in Berkeley, California. For further details, see Manheimer (1972), Davidson (1976), and Mellinger (1976).

[7]These and other problems in the analysis of survey data are fully discussed in James N. Morgan and John A. Sonquist, Problems in the analysis of survey data, and a proposal, *Journal of the American Statistical Association*, 58:415-435, June 1963. This is the paper in which the reasoning underlying Automatic Interaction Detection was first described.

[8]Similar approaches are described in Belson (1958, 1959) and Hindelang (1974).

[9]It can be demonstrated that once the groups are ordered in terms of their mean values, none of the other possible combinations, such as taking groups 1 and 4 together and splitting them off from groups 3, 2, and 5, needs to be examined; cf. Sonquist, Baker, and Morgan (1973), pp. 209-215.

[10]We obtained permission from the men we interviewed to examine their records in the Registrar's Office. The grade-point average reported here is for the fall and winter quarter of the freshman year.

[11]The comparison of rates in this fashion by the construction of ratios can be misleading. Among the more motivated students, the ratio of 8.5 percent to 0.7 percent indicates a dropout rate of better than twelve times as great for the students with low grades. However, it is best to look upon this high ratio as inflated by the very small denominator (0.7 percent), on which small sampling fluctuations can have a large impact.

[12]We return to a brief discussion of these particular splits below.

[13]"The tree produced by the algorithm is not necessarily better (in terms of cumulative BSS/TOTSS) than all other possible trees," (Sonquist and Morgan, 1971, p. 133). Professor Ira Cisin (personal communication) refers to this as a "suboptimal" procedure, pointing out that it is sometimes possible to obtain a larger BSS/TOTSS by deleting a powerful predictor that is also powerfully related to several other predictors.

CONTINUED

1 OF 3

[14]Morgan and Sonquist, op. cit., p. 426.

[15]See John A. Sonquist, *Multivariate Model Building: The Validation of a Search Strategy*, Ann Arbor: Institute for Social Research, 1970, for further discussion.

[16]See, for example, Hillel Einhorn, Alchemy in the social sciences, *Public Opinion Quarterly*, 36:367-378, Fall, 1972 and James Morgan and Frank Andrews, A reply to Einhorn, loc. cit., 37: 127-129, Spring 1973.

[17]Ira H. Cisin and Dean I. Manheimer, Marihuana use among adults in large city and suburb, *Annals of the New York Academy of Sciences*, 191:222-234, December 31, 1971.

RESOURCES AND REFERENCES

Illustrations of the Use of AID and Similar Procedures

Anderson, P.R. *Discretionary and contractual saving in Canada, a cross-sectional study.* Unpublished Ph.D. dissertation, Harvard University, 1967.

Belson, W.A. Measuring the effects of television: A description of a method. *Public Opinion Quarterly*, 22:11-18, 1958.

Calahan, Don. *Problem Drinkers.* San Francisco: Jossey-Bass, 1970.

Carman, James M. *Multiple Classification Analysis Without Assumption of Interval Measurement, Linearity, or Additivity: A Comparison of Techniques.* Reprint No. 23, Institute of Business And Economic Research, University of California, Berkeley (no date). Reprinted from *Proceedings of the Social Statistics Section, American Statistical Association*, December 27-30, 1967.

Cisin, I.H., and Manheimer, D.I. Marijuana use among adults in a large city and suburb. *Annals of the New York Academy of Sciences*, 191:222-234, 1971.

Clark, W. Loss of control, heavy drinking, and drinking problems in a longitudinal study. In: Clark, W., and Knupfer, G., eds. *Drinking Patterns and Problems: A Decade of Studies in the San Francisco Area.* Special issue of *Journal of Studies on Alcohol*, forthcoming, 1976.

Hermann, R.O. Interaction effects and the analysis of household food expenditures. *Journal of Farm Economics*, 49, 1967.

Hindelang, M.J. Decisions of shoplifting victims to invoke the criminal justice process. *Social Problems*, 21(4):580-593, 1974.

Other References

Belson, W.A. Matching and prediction on the principle of biological classification. *Applied Statistics*, 8:65-75, 1959.

Davidson, S.O.; Mellinger, D.G.; and Manheimer, D.I. Changing patterns of drug use among university men. To be published in *Addictive Diseases*, 1976.

Einhorn, H. Alchemy in the behavioral sciences. *Public Opinion Quarterly*, 36:367-378, 1972.

Manheimer, D.I.; Mellinger, G.D.; Somers, R.H.; and Kleman, M.T. Technical and ethical considerations in data collection. In: Einstein, S. ed. *Student Drug Surveys; Proceedings of the First International Conference.* Farmingdale, N.Y.: Baywood Publishing Company, 1972.

Mellinger, G.D.; Somers, R.H.; and Manheimer, D.I. *Drug Use and Academic Attrition Among University Men.* Paper to be published in the *Proceedings of the Conference on the Social Psychology of Drug and Alcohol Abuse*, Los Angeles, May 1975, in press 1976.

Morgan, J.N., and Andrews, F.M. A comment on Mr. Einhorn's 'Alchemy in the behavioral sciences.' *Public Opinion Quarterly*, 37:127-129, 1973.

Morgan, J.N., and Sonquist, J.A. Problems in the analysis of survey data: And a proposal. *Journal of the American Statistical Association,* 58:415-434, 1963.

Nie, N.H.; Bent, D.H.; and Hull, C.H. *SPSS: Statistical Package for the Social Sciences.* 2nd ed. New York: McGraw-Hill, 1975.

Selvin, H.C., and Stuart, A. Data dredging procedures in survey analysis. *American Statistician,* 20(3):20-23, 1966.

Sonquist, J.A. Simulating the research analyst. *Social Science Information,* 6(4):207-215, 1967.

Sonquist, J.A. Finding variables that work. *Public Opinion Quarterly,* 33(1):83-95, 1969.

Sonquist, J.A. *Multivariate Model Building: The Validation of a Search Strategy.* Ann Arbor: University of Michigan, Institute for Social Research, 1970.

Sonquist, J.A.; Baker, E.L.; and Morgan, J.N. *Searching for Structure: An Approach to Analysis of Substantial Bodies of Micro-data and Documentation for a Computer Program.* Ann Arbor: University of Michigan, Institute for Social Research, 1971; and revised edition, 1973.

Sonquist, J.A., and Morgan, J.N. *The Detection of Interaction Effects: A Report on a Computer Program for the Selection of Optimal Combinations of Explanatory Variables.* Ann Arbor: University of Michigan, Institute for Social Research, 1964.

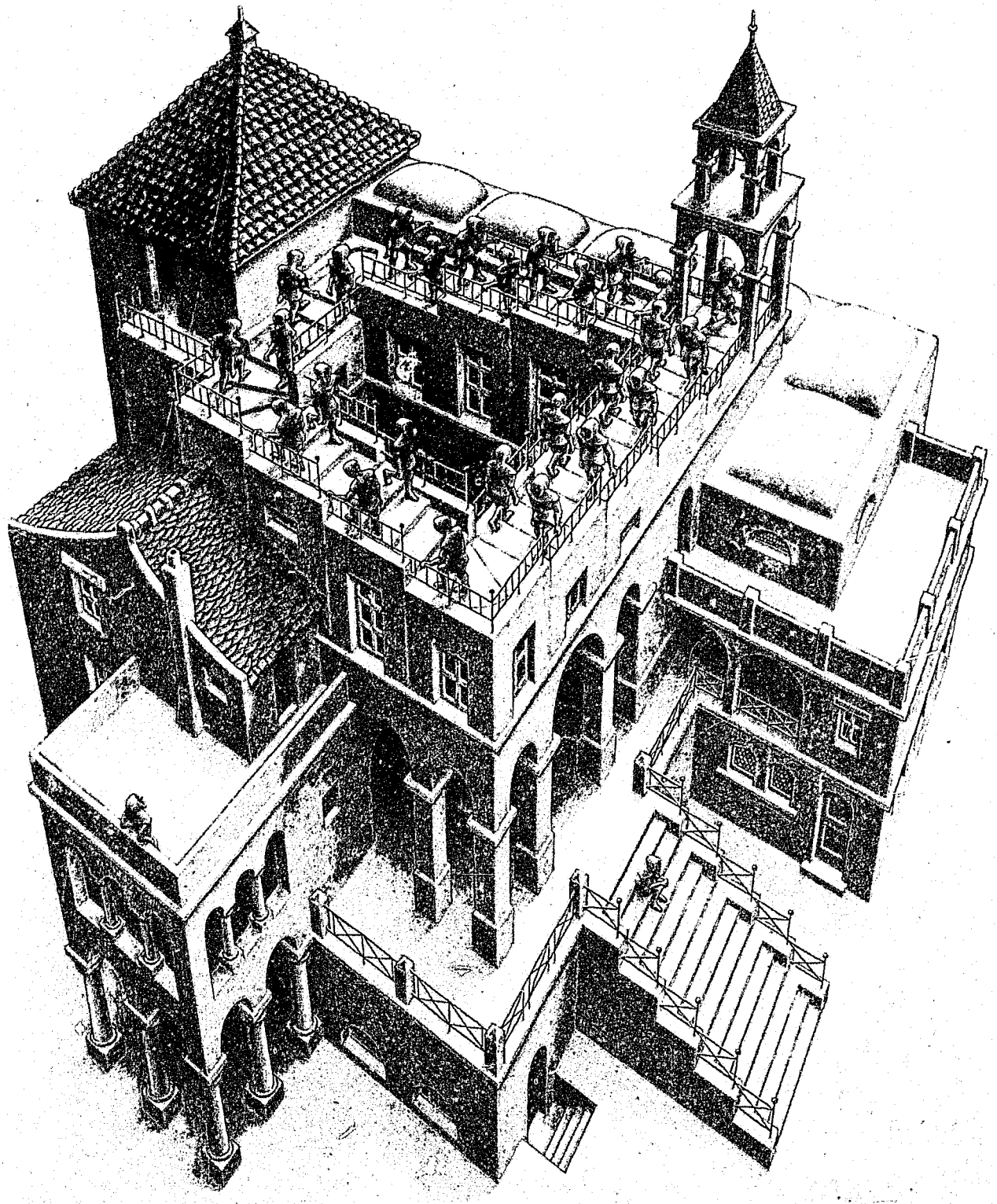


Plate 6

INTRODUCTION

This chapter describes one particular actuarial approach to the prediction of person-characteristics that are related to the use and abuse of drugs. This approach involves a set of procedures to identify a number of test-defined classes of drug users. More importantly, it involves several steps in evaluating the extent to which individuals who are members of each of these classes share relatively homogeneous etiologies, patterns of drug use, or responses to specific treatment programs. It is flexible enough to meet both the researcher's need for group data and the clinician's need to respect the considerable uniqueness of the individuals who use and abuse drugs. The methods described are neither mathematically elegant nor taxing, but they make statistical as well as clinical sense. Other actuarial procedures may be found in Robins (1972) and in Wiggins (1973). Wiggins particularly points up actuarial uses of patterns of either quantitative or qualitative indicators in the insurance industry. In any form, actuarial techniques are based on the use of empirically determined relations between a set of indicators (predictors) and another set of attributes that cannot be determined as economically or as quickly as we can obtain information about the predictors. Since psychological test data can be and often are collected from drug users economically, and early in our contact with them, the actuarial approach discussed here deals primarily with patterns of easily obtainable psychological test data.

The paper is intended to clarify the procedural steps involved in this approach and to encourage professionals working with drug abusers to consider using it to identify subgroups that may be relatively homogeneous in terms of several socially and clinically important attributes. Readers who are interested in more comprehensive discussions of the issues and procedures that are central to the development and use of actuarial methods should first consult Meehl (1954; 1956; 1960; 1973). In addition, there are four major actuarial programs using the MMPI with psychiatric patients that should be examined carefully (Gilberstadt and Duker, 1965; Gynther, Altman, Warbin, and Stetten, 1972; Gynther, Altman, and Sletten, 1973; Marks, Seeman, 1963; Marks, Seeman, and Haller, 1974), and several reviews of actuarial methods that are useful (Gough, 1972; Sawyer, 1966; Sines, 1966).

RATIONALE AND CAUTIONS

The approach is based on the assumption that there are several distinguishable patterns of psychological test data among drug users and that some of these patterns of test data will define groups (taxonomic classes) whose members are relatively homogeneous in terms of etiology, pattern of drug use, or response to treatment. These are not particularly revolutionary assumptions, as witnessed by the tendency of many experienced clinicians to develop some general working typology within which they classify individuals among their clientele. The procedures described below are suggested as ways to formalize the definitions of some of those types to evaluate critically the etiological or prognostic significance of their membership. The ideal typology or classification system, of course, would be one that accommodated all individuals and one in which all individuals assigned to a given class shared the same etiology, the same pattern of drug use, and the same response to treatment. In spite of the fact that we will probably not soon develop such an ideal typology, it seems prudent to try to subdivide the larger group of drug users into smaller groups. The members of such groups may be significantly more homogeneous than the total population of drug users in terms of either etiology or pattern of use or response to treatment.

This discussion is presented largely in terms of psychological test scores. One should keep in mind, however, the possibility and desirability of developing additional, quite possibly unrelated taxonomic classes on the basis of the patient's interpersonal behavior, demographic characteristics or information concerning the environments in which the drug user lives. The critical importance of environmental influences in the development of drug abuse is indicated in several reviews of the literature (Braucht, Brabarsh, Follingstad, and Berry, 1973; Ferguson et al., 1974). The history of psychology's recognition of the importance of environmental influences and some of the issues that must be dealt with in attempts to assess environments are considered in several recent reports (Bowers, 1973; Ekehammar, 1974; Endler, 1975; Insel and Moos, 1974).

It is finally necessary to keep in mind that while the identification of psychometrically homogeneous groups of persons is one step in the development and use of an actuarial system, the mere identification of such groups of drug users is of relatively little clinical value. The clinical value of any psychometrically defined group or class depends on the extent to which members of that class are found also to be homogeneous with respect to other clinically important nontest characteristics such as etiology or pattern of use, or response to treatment. If the personality characteristics or dimensions measured by our test are related to drug use, it seems reasonable to expect the clinical homogeneity of a test-defined subgroup to increase as we increase the psychometric homogeneity of that subgroup. As we increase the psychometric homogeneity within a group, we simultaneously reduce the number of persons who can be assigned to that group. This raises a number of problems that will be considered later in relation to the several steps involved in the development and use of actuarial prediction methods.

METHODS AND PROCEDURES

There are five steps involved in the development, evaluation and use of the actuarial procedures offered here for your consideration.

1. Selection of the criterion characteristics to be predicted (etiology, pattern of drug use, response to a particular treatment program).

2. Selection of the test variables thought to be related to the chosen criterion.

3. Identification of patterns of scores on the predictor variables.

4. Empirical determination of the relationship between each pattern of predictor scores and each criterion of interest.

5. Recognizing when the last drop of predictive blood has been squeezed out of one particular turnip and wisely turning one's attention to additional domains of predictors.

STEP 1: SELECTION OF CRITERIA TO BE PREDICTED

Type and Variety of Data

In considering the type and variety of criterion data to collect, it is helpful to note Gleser's (1963) important distinction between "fixed criteria" and "free criteria". An example of a "fixed criterion" is response to treatment when we can define a negative response, no response, or a positive response to a particular treatment program. Continuation in a treatment program, for instance, or "clean" urine, or abstinence from drugs at some specified later date may all be determined in a reasonably objective fashion and may constitute a fixed criterion. In this case, the focus is on the extent to which our test data will allow us to predict a specific predefined criterion characteristic.

We essentially cast a wider net when we talk about "free criteria." In this case the question is whether or not there are clinically or socially important characteristics that can be predicted using our test data. Or, in the context of this discussion of actuarial methods, the question is whether or not members of a test-defined group are relatively homogeneous in terms of any clinically or socially important nontest characteristics.

If the question involving "free criteria" is put in this form, the importance of collecting a wide variety of information about each of many individuals should be quite obvious. When we take this approach we are essentially asking whether patterns of the variables measured by some particular test are related to clinically or socially important characteristics of our patients. Since many of our tests may be significantly related to some but not all clinically important patient characteristics, it is important that we collect the largest practicable array of clinically important information on each of our patients in the hope that some of those data may indeed be predictable from one or more of the patterns of test scores that we may identify.

The choice of criterion data also bears on the generalizability of any actuarially derived descriptions and predictions that may result from a general research program. As will be noted below, the collection of criterion data and the actuarial development of descriptions in one particular clinical setting may seriously limit the extent to which even narrowly defined patterns of test

data provide accurate or clinically important descriptions or predictions about patients seen in other settings. The most energetic attempt to insure the generalizability of their actuarially-derived criterion descriptions was made by Marks, Seeman, and Haller (1974) when they collected a standard set of psychiatrically relevant criterion data about adolescents being seen at 74 different clinics and hospitals. But while the range and the relevance of the criterion data are important considerations, the quality or the accuracy of those data are even more critical. The issues raised here have generated an extensive literature concerning "the criterion problem" and no easy solutions can be offered. The interested reader should examine the discussion in Sines (1966) for a broad view of the issues involved.

The Limitation of Present Drug Research

Most reports in the psychological literature concerning drug use and abuse deal with a remarkably small number and variety of nontest characteristics of the persons studied. In order to justify an actuarial approach of the sort described here, one needs far more than the usual number and kinds of nontest information about one's subjects. The literature sampled and reviewed in previous NIDA Research Issues Series point up the varied types of antecedent events and experiences, patterns of use and natural history data that bear on drug use. None of the reports in the psychological literature includes even a reasonable sample of those several important domains of information. In view of this state of affairs, one of the first orders of business in further research in this area should be the development of a standard information-gathering form that would be available for use by clinical researchers. Such a form should be a reasonable compromise between the comprehensiveness of the armchair and the pragmatics of the clinic. In the absence of such a standard history, status, and follow-up form, each investigator would have to decide what items of history, present status and follow-up information are important in the study of drug abuse. The result would be a literature heavy on the psychometric characteristics and light on the socially and clinically important characteristics of drug abusers.

At this point it also should be noted that the serious study of the causes, correlates and consequences of drug abuse is a relatively recent undertaking and, as a result, we do not have the body of information or the broad perspectives that we have relative to psychiatric disorders. And this lack of information makes it very difficult for us to make wise choices of drug-relevant information to collect and to use as criterion data. In addition to the relative recency of our systematic investigation of drug abuse, the fact that the abuse of drugs has spread rapidly in the last few years from lower and marginal SES groups to all social groups and to younger individuals complicates our efforts to agree on the "important" or "relevant" criterion data to be collected.

It seems obvious, too, that the different theoretical orientations of clinicians working in different drug treatment centers will also generate wide differences in judgments about the important criterion data to be selected. To anticipate a point discussed later, we should develop a core list of characteristics of drug users that clinicians in several centers would agree are important or relevant to their understanding and work with such patients. Such a compilation would probably not be a comprehensively adequate set of data for workers in any one center but it would be a start on the tedious job of sorting fact from unfounded impressions.

Types of Criterion Data in Previous Research

Previous research using actuarial methods has made use of three types of criterion data. In their actuarial programs using the MMPI with adults (Marks and Seeman, 1963), and later with adolescents (Marks, Seeman, and Haller, 1974), Marks et al. made use of therapist Q-sorts of personality-descriptive statements, ratings of a large number of case history items and other psychometric test data. The criterion data used by Gilberstadt and Duker (1965) were derived from the ratings of case histories using a checklist of descriptive terms and statements. Sines (1964; 1966) and Davis and Sines (1971) used a system for recording the entire contents of the institutional records of patients, whose test data were also available. Gynther, Altman, Warbin, and Sletten (1972) used demographic data from the face sheet, and intake diagnosticians' ratings on a mental status form. In each of these instances, a wide variety of criterion information was available on each subject. When subgroups of those subjects were identified on the basis of various patterns of test data, several constellations of history, current status and outcome data were found to characterize some of the test-defined groups. It is important to note that the criterion data were collected independently of the test data. The nontest criterion characteristics were thus available for empirical study and were not later inferred from patterns of test data.

STEP 2: SELECTION OF PREDICTOR VARIABLES

Psychological Tests

There is, of course, no reason for us to consider only psychological test scores as our predictors. The available reports indicating that demographic characteristics, premorbid behavior and clinically rated observable behavior are related to clinically important characteristics of drug abusers or psychiatric patients should stimulate more intensive study of those domains of predictor variables. But the fact is that many of us assume, with cause, that some of the personality characteristics measured by one or another psychological test are related to clinically important characteristics of drug users. And since many psychological test variables are "psychometrically tractible" (Lindzey, 1965), we may efficiently and profitably examine test scores and patterns of test scores as potentially useful predictors of the criteria of interest.

Other Predictive Variables

Although we may be justified in asking whether a particular test will allow us to distinguish between various predefined criterion groups of drug users, the nature of the criterion to be predicted may sometimes render personality tests less appropriate than other types of predictors. For instance, if our criterion of interest is response to a treatment program that is closely controlled and that allows the patients rather little unplanned experience, we may not need to consider the possible impact of current living environment, frustrations at work, stresses of unemployment, poor diet or availability of drugs. Under such conditions we may reasonably expect personality characteristics to be significant determiners or predictors of response to treatment. If, on the other hand, our treatment program is used with persons who are living at home and are exposed to a number of the conditions mentioned above, we may reasonably expect personality characteristics, as measured by our tests, to account for much less of the variance in the criterion of interest. Under the second set of circumstances it may well be that members of psychometrically homogeneous classes may be distressingly heterogeneous on the criterion of interest. It therefore would have been more appropriate to examine the predictive value of the demographic or environmental characteristics of our patients.

We may face a similar set of circumstances when we attempt to identify addiction-prone persons, or persons who are vulnerable to the use of drugs. Certain personality characteristics or certain constellations of personality attributes may be predictive of future use of drugs. But when an investigator ignores home environments, school environments and neighborhood characteristics and attempts to assess risk of drug abuse using personality characteristics alone, he is essentially guessing that the predictor variables chosen for study are more important and more powerful than the influence of those other classes of potential predictors.

During the last 10 years a number of reports have appeared indicating differences between drug users and nonusers on several standardized multivariate personality tests such as the CPI (Hogan, Mankin, Conway, and Fox, 1970), the MMPI (Smart and Jones, 1970; Brill, Compton, and Grayson, 1971), the 16PF (Koslowsky and Deren, 1975), and the PRF (Holroyd and Kahn, 1974). Demographic, family and SES characteristics have also been found to discriminate users and nonusers of certain types of drugs (Ferguson et al., 1974). At this time it would seem appropriate, therefore, to collect as many of these types of predictor data as possible in order to evaluate clinical value and significance. It must be pointed out that our present state of knowledge does not justify the sole use of any one or another of these sources or types of potential predictors so our choices will often be dictated by the relative convenience, costs and feasibility of collecting the several types of predictor information. There are, to be sure, severe limits on the amount and variety of information that may be collected at any one center, but if even two or three types of information were to be collected in several reasonably comparable centers, the development of an effective prediction program would be greatly facilitated.

STEP 3: IDENTIFICATION OF PATTERNS OF SCORES ON THE PREDICTOR VARIABLES

Clustering Techniques

As Meehl pointed out in his Foreword to Marks and Seeman's *Actuarial Description of Abnormal Personality* (Marks and Seeman, 1963), "...there is no rigorous, straightforward actuarial 'searching' technique available for grouping similar but nonidentical (test) profiles into coarser classes or 'types' so as to combine the desiderata of stable sample size, high psychological homogenity, easy identifiability for routine clinical entrance, and quasi-complete coverage of the range of patterns empirically found."

As many of the papers in this volume indicate, we may use any of a large number of statistical methods to identify patterns of scores on our predictor variables. None of the four major actuarial systems that have been developed for use with psychiatric patients have used the statistical techniques described in this volume for generating psychometric types or classes. At least two procedures, however, have been used. Marks and Seeman (1963) and Gilberstadt and Duker (1965) developed MMPI profile patterns on the basis of their extensive clinical and research knowledge of that test. By refining their definitions of the profile patterns, those investigators developed a set of contingency rules that allowed relatively similar MMPI profiles to be assigned to one or another of 16-19 groups. Somewhat later, Gynther et al. (1972) and Marks, Seeman, and Haller (1974) developed patterns of test scores using only the highest two or three MMPI scale scores.

Regardless of the grouping or clustering procedures one uses, the rationale is the same, i.e., there are psychometrically relatively homogeneous classes of persons who are also relatively homogeneous in terms of some of our criteria of interest. Consider, for example, the procedures we might use to determine whether one or more test-defined groups exist of nondrug users who will subsequently become users, i.e., whether there are patterns of test scores that are predictive of later drug use or abuse. Here we might appropriately use one of the available statistical clustering techniques such as r_p (Cattell, 1949), profile correlation (Lorr, Bishop, and McNair, 1965) or the simple Euclidean distance function D^2 (Cronbach and Gleser, 1954) or the nonstatistical methods used by Marks and Seeman (1963) or by Gynther et al. (1972) to identify patterns of test scores (see chapter 7 for a discussion of several indices of similarity that may be used for these purposes). The subsequent use or nonuse of drugs by members of each of these test-defined groups could then be determined at appropriate later points in time. It seems highly likely that there are several patterns of predrug test scores on personality tests, each of which is predictive of higher or lower than base-rate use of drugs.

Even when we are concerned with predicting a fixed criterion, such as response to treatment defined in one of the ways noted earlier, we may profitably use one of the available clustering techniques in an attempt to identify several test patterns, each of which may be predictive of our criterion of interest. It seems clinically reasonable, for instance, to expect to find several rather different test-defined personality types among a group of drug users, all of whom will respond to treatment by terminating their use of drugs and by remaining abstinent for a year. To the extent that we have chosen to study a set of relevant predictor variables, we may efficiently distinguish those several "good prognosis" groups by applying one of the several clustering techniques to the pretreatment test data.

The D^2 Index of Profile Similarity (The Euclidean Distance Function)

The method recommended here involves grouping individuals on the basis of the patterns of their test scores. In this way we may identify several groups or classes whose members may be studied further to determine if they show greater than base-rate homogeneity in predrug behavior or response to treatment. Of the several clustering methods that have been described and recommended, I prefer to use the Euclidean distance function D^2. As discussed by Lorr (chapter 7 in this volume), this index is to be distinguished from the more rigorous generalized distance function developed by Mahalanobis and also designated by the symbol D^2. While the Mahalanobis generalized distance function provides a more accurate index of the similarity of two test profiles in terms of the basic factor structure underlying the several test scores, there are no data to suggest that persons grouped together by the Mahalanobis D^2 index are behaviorally more similar than persons grouped together on the basis of the much simpler Euclidean distance function D^2. At this point it appears that the Euclidean distance function (D^2) will serve adequately to identify groups of persons who are psychometrically highly similar. Hereafter the symbol D^2 will be used to designate the Euclidean distance function.

The steps involved in developing clusters of relatively homogeneous test data with the D^2 index of profile similarity can be illustrated using the L, F, and K, plus the 10 clinical scales of the MMPI. Each MMPI profile is compared to each other profile by calculating the D^2 value for each pair of profiles, where $D^2 = \Sigma\, d^2_L + d^2_F + d^2_K + d^2_{Hs} + d^2_D + \ldots . D^2_{Si}$. The component d^2 values are the squared differences in T score points between corresponding scales. The profile that generates D^2 values of 625 or less with the largest number of other profiles is chosen as the first "target" profile. A matrix is constructed of the target profile and all others that relate to it with D^2 values of 625 or less. The profiles that relate to the smallest number of other profiles in the matrix are successively eliminated until all remaining profiles relate to at least 60% of the remaining profiles with a D^2 of 625 or less.

In this process, the psychometric homogeneity of the cluster may be increased by setting the critical D^2 value at less than 625 or by requiring members of the final matrix to relate to more than 60% of the others with the critical D^2. Conversely, the number of profiles included in the final cluster may be increased by increasing the acceptable heterogeneity of the group by accepting a higher critical value of D^2 or by including profiles even if they relate to less than 60% of the others in the group with the critical D^2 value.

Once the final cluster of profiles is identified, the T scores on each of the scales are averaged and the mean profile for that group is set as Prototype #1. The prototypic profile is then compared to all of the profiles in the original sample, and any profile that relates to the prototype with a D^2 value of 484 or less is considered to be an instance of Prototype #1. The psychometric homogeneity of this final group may be increased or decreased by the appropriate shift of the critical D^2 value or by setting a smaller or larger limit on the largest acceptable d^2.

The critical D^2 values used in developing a number of MMPI defined classes were selected after several other values had been tried. The values chosen allowed us to construct several relatively homogeneous groups, each of which included at least five profiles. Furthermore, and perhaps more important, knowledgeable clinicians judge the MMPI profiles assigned to a particular group by this use of D^2 to be highly similar to one another.

Cautions

Homogeneity. The clinical homogeneity of any psychometrically defined subgroup of patients will not be assured by statistical evidence of the psychometric homogeneity of the group or cluster. In view of this fact, the availability of tests of the statistical significance of various indices of profile similarity does little to recommend the use of those clustering procedures (Lykken, 1968). If we choose to develop classes that are psychometrically highly homogeneous, we may end up with a large number of classes, each of which includes only a few individuals. If on the other hand, we generate classes or patterns that accommodate relatively large numbers of our patients, we will be dealing with psychometrically rather heterogeneous groups. And to the extent that different scores on any of our predictor variables are related to differences in any of the characteristics of clinical interest, a psychometrically heterogeneous group will also be clinically heterogeneous.

There is no ready answer to this problem. The researcher must decide which approach to take on the basis of the size of the available sample, the questions being studied, and prior knowledge about the variety of personality patterns that relate to the criterion. The appropriate narrowness of a test-defined group cannot be determined statistically; it is ultimately a clinical or, more accurately, an empirical question.

Class Composition. There is another problem that must be recognized by researchers who may consider using the actuarial approach recommended here. Regardless of the breadth or the narrowness of the method we use to define patterns of any specific multivariate predictor data, we will find that those classes will identify a disappointingly small proportion of our patients (Berzins, Ross, English, and Haley, 1974; Zuckerman, Sola, Masterson, and Angelone, 1975; Goldstein and Linden, 1969; Huff, 1965; Pauker, 1966; Gynther, et al., 1972; Marks and Seeman, 1963; Owen, 1970; Lorr, Bishop, and McNair, 1965; Payne and Wiggins, 1968; Sines, 1966). This failure to capture all or even most patients in one or another of the psychometrically-defined classes may lead some clinicians to either reject this actuarial approach entirely or to broaden the definition of the several classes so that each class will accommodate a larger proportion of the patients under study, even though by broadening the psychometric definition of the classes, one runs the serious risk of making those classes unacceptably heterogeneous clinically. I will discuss these issues in greater detail below in relation to knowing when to quit using a particular set of predictors (Step 5).

STEP 4: DETERMINING THE RELATIONSHIP BETWEEN TEST PATTERNS AND CRITERIA OF INTEREST

Cautions

The predictors we are now dealing with are patterns of test scores. It is important to note that none of the component scores in the patterns are to be dealt with as individual predictors, even though some discussions of mean profile patterns treat the various scale scores separately. In their discussion of the 2-7, 2-7-4 and 2-7-8 MMPI profile patterns, Gilberstadt and Duker (1963) have documented the considerable extent to which the clinical significance of a scale score depends on the configuration of the entire profile. We must determine, among drug users, the cri-

terion characteristics associated with each pattern rather than assuming that the background or prognostic characteristics observed in psychometrically similar but non-drug abusing psychiatric patients apply to this population as well.

This requirement emphatically points up the importance of a comprehensive system for collecting the criterion variables of interest. If we have chosen to search for patterns of test scores that relate to a fixed criterion, we will, of course, have that criterion information available. In those more likely instances where we may be interested in identifying groups that are relatively homogeneous in terms of background history, for instance, the success of the venture depends on the prior collection of the appropriate data--doubly difficult in that we do not yet know the several factors that may combine to put an individual at risk for drug abuse.

Analytic Techniques

Let us assume that we have administered the MMPI, or another multivariate test of our preference, to a sample of 500 patients about whom we have collected a preselected set of criterion information. Some of those criterion data are dichotomous or discrete items such as sex, race, urban-rural home background, SES level of parental family, marital status, etc., while other criterion data may be continuous such as age, highest school grade completed, IQ, or number of years of drug use. Let us also assume that we have identified a group of 25 persons whose MMPI profiles match our Prototype #1 profile with a D^2 value of 484 or less. We can determine the frequency of each of our discrete items in the test-defined group and in the remaining 475 subjects. These two frequencies for each discrete item may be compared using Chi-square with Yates' correction, and if the difference reaches a predetermined level of significance, we may conclude that our test-defined group is characterized by the appropriate criterion descriptor. When dealing with continuous criterion data such as number of years of education, a t-test may be used to compare the means of the members of the Prototypic group and the remainder of the sample.

In situations such as this, where one makes multiple comparisons, some unknown number of differences at, say, the .05 level should be found by chance alone. Since the usual expectation of 5% at the .05 level and 1% at the .01 level by chance appears to be overly conservative (Block, 1960), the importance of replicating one's findings is increased. Such replication should be done using totally independent samples of subjects but, as will be noted below, some investigators have divided the test-defined group and treated the two halves of that group as independent samples.

It is here also that the sources of data come to bear significantly on the generalizability of findings. The generalizability of the findings of this actuarial system will be greatly enhanced if the initial data were obtained from several clinical settings so that the patients and the theoretical orientations of the professionals may be as broadly representative as possible.

The critical point here is that the relationship between test-defined groups and the criteria of interest must be determined empirically. This has not often been done in the studies that have been reported in the available literature.

STEP 5: KNOWING WHEN TO MOVE ON TO OTHER PREDICTORS

We have gathered the potentially important data concerning knowable background events and experiences, patterns of drug use, responses to various treatment programs, etc. We have also selected a reasonable set of predictors and have identified psychometrically quite homogeneous classes that include at least a moderate number of subjects. We now will probably find that less than half of the original population is classifiable into one or another group (Berzins, Ross, and English, 1974; Goldstein and Linden, 1969; Holroyd and Kahn, 1974; Owen, 1970; Lorr, Bishop, and McNair, 1965). Even more distressing is the possibility that some of these psychometrically homogeneous groups of persons are no more homogeneous than the entire population in background characteristics, observable behavior, or response to treatment (Gynther, Altman, and Sletten, 1973).

While it is certainly to be hoped that some of those psychometrically homogeneous classes show greater than base-rate homogeneity in some clinically important respect, our present level of knowledge does not guarantee such a positive finding. But if one or more of the groups identified using one particular set of predictor variables shows a clinically important degree of homogeneity in terms of any of our criteria of interest, those are valuable data. In such a case, we should routinely collect those predictor data and make clinical decisions on the basis of membership in those groups while attempting to identify additional psychometrically homogeneous groups among the remaining patients using other domains of predictors.

As noted earlier, the fact that patterns of scores on any particular personality test, however typed or clustered, fail to capture or accommodate all or even most patients, has led some clinicians to consider either the test instrument or this actuarial approach to be of little clinical use (Huff, 1965; Gynther et al., 1972). This is an unfortunate reaction. If examined carefully, it appears to imply the unnecessary and erroneous assumptions that, in order to be significant, the personality variables that are assessed by a test, and patterns of those variables, must describe all drug users and must identify and distinguish between all of the clinically meaningful subgroups of drug abusers.

In view of the marked conceptual and methodological differences between the several multivariate instruments that are available for use, it seems quite reasonable to accept the fact that the variables assessed, for instance, by the 16 PF and the MMPI are not only substantively different, but that they may allow us to identify rather different groups of clinically homogeneous patients. If this is the case, we may reasonably expect to identify a few clinically quite important test-defined types or classes of drug users using one such measuring instrument, and yet find scores or patterns of scores on that test to be unrelated to clinically meaningful characteristics of the remaining large proportion of drug-abusers. It seems quite possible that the variables measured by any one of our testing instruments may be relevant only to a minority of drug users. If such is the case, we would be well advised, then, to use another assessment instrument or another type of instrument in an attempt to identify clinically meaningful subgroups among the remainder of the drug users who had not already been classified using the first test. If we proceed in this sequential fashion, it is probable that a very large proportion of those persons who use drugs can be typed or classified into one or another test-defined group.

ILLUSTRATIVE APPLICATIONS

NONDRUG RESEARCH

There are four reports of the use of actuarial methods with the MMPI in psychiatric populations (Gilberstadt and Duker, 1965; Gynther, 1972; Marks and Seeman, 1963; Marks, Seeman, and Haller, 1974). These reports illustrate the variety of ways in which each of the steps discussed above have been dealt with. The actuarial systems that have been developed raise a number of issues that emphasize the important decisions that must be made by an investigator who wishes to use actuarial methods.

Selection of Criteria to be Predicted

Each of these systems defined at the outset the domain of criterion information to be collected. Demographic characteristics, SES-related data, current behavior, mental status, treatment given, and response to treatment were recorded using specially prepared checklists or rating forms. These data were recorded by intake diagnosticians, by therapists, or by trained raters who carefully reviewed each patient's hospital record. Since these data constitute those clinically important attributes of the patients which we wish to predict, the initial choices are critical. The accuracy or the validity of these data are equally important, and most of the reports cited above describe the efforts that were made to ensure the accuracy of those basic data.

Marks and Seeman (1963) have reported the two main sources of personality and background information in sufficient detail for them to serve as illustrations of desirable procedures. A set of 108 personality-descriptive statements was selected on the basis of extensive prior research. These criterion descriptions were "selected for their representative coverage of the personality domain,...their applicability to both sexes, (their) clinical pertinence, interpatient variability" as well as their ratability. Each patient was described by his therapist using these 108 statements. The items were sorted into a nine-category rectangular Q-distribution so that the 12 statements placed in category 1 were judged by the therapist to be least characteristic of the patient, and those 12 statements placed in category 9 were judged to be most characteristic of the patient. It should be noted also that therapists made these Q-sorts only for patients they had seen for at least 15 hours of therapy. This last requirement was intended to enhance the accuracy or the validity of the ratings.

A serious disadvantage entailed by this method of collecting the criterion information is the great amount of skilled clinical personnel and time required to make such ratings. But since we are dealing with the pivotal information in any approach to understanding and predicting important characteristics of our patients, we must consider the alternatives. Low quality criterion information is worthless. The criterion information Gynther (1972) and his associates

used consisted primarily of 111-item mental status forms completed by psychiatric residents shortly after the patients' admission to a large state hospital system. The selection and method of collecting the criterion information involves a number of decisions that will affect the whole enterprise.

A second source of criterion information used by Marks and Seeman consisted of ratings of the patients' hospital records for the presence or absence, or degree of presence, of each of 225 case history variables chosen from lists used in previous research. Two out of three raters had to agree before any variable could be included as characteristic of a patient.

To summarize, the criterion information to be collected was selected on the basis of its clinical relevance, and procedures were developed to enhance the validity of the ratings of those variables.

Selection of the Test Variables

All of the actuarial systems I know of have made use of the MMPI. As noted above under Step #5, actuarial methods are applicable to other tests and to other types of predictor data. Furthermore, the use of only one personality test does not provide a comprehensive test of the underlying logic of actuarial methods. It is probable that the typologies that will ultimately allow us to predict patient behavior in a clinically useful fashion will require the use of several different tests as well as several different domains of information.

Identification of Patterns of Scores on the Predictor Variables

None of the four actuarial systems referenced above have used statistical procedures to identify patterns of test data. Marks and Seeman (1963) and Gilberstadt and Duker (1965) developed a set of MMPI profile types on the basis of extensive clinical experience with that test. By progressively refining the definitions of those test patterns, they developed a set of rules that are used to determine whether any given MMPI profile can be classified as an instance of any one of the test patterns.

Marks and Seeman reported that their 16 MMPI patterns accommodated the profiles of 78% of the patients who were seen by the Department of Psychiatry at the University of Kansas. Subsequent reports indicate that in other clinical settings the rules published by Marks and Seeman will allow only 20% to 30% of the MMPI profiles to be classified. A similarly low "hit rate" was found when Gilberstadt and Duker's rules were applied to MMPI profiles obtained in a variety of settings. This disappointingly limited coverage led Huff (1965) to conclude that the Marks and Seeman's classification procedure was unsatisfactory for use in other settings.

The small proportion of their patients that were classifiable using the published rules led Gynther et al. (1972) and Marks, Seeman, and Haller to define MMPI profile patterns in terms of the two highest scores on the 10 clinical scales. Using the two highest scale scores to classify profiles, Marks, Seeman, and Haller were able to classify virtually all of the 834 adolescents on whom they had data. Gynther (1972) was able to account for 60% and 64% of the MMPI profiles in a derivation sample and a replication sample respectively.

As I have pointed out elsewhere (Sines, 1966), neither Marks and Seeman nor Gilberstadt and Duker have reported the psychometric variability within each of the groups defined by their sets of rules. The variability within several of the Marks and Seeman and Gilberstadt and Duker classes was considerable (Sines, 1966). The variability that characterizes each of the test patterns defined by only the two highest scores must be even greater. To the extent that the test scores relate to clinically important nontest attributes of one's patients, groups that are psychometrically heterogeneous will be clinically heterogeneous. At this point one must decide: (1) either to identify a number of psychometrically homogeneous groups that are relatively small and account for relatively few of the patients of interest, (2) or to identify broad, psychometrically rather heterogeneous classes that will accommodate a large proportion of the patient sample.

Determination of the Relationship Between Test Patterns and Criteria of Interest

The procedures used by Marks, Seeman, and Haller illustrate the straightforward assessment of the extent to which adolescents whose MMPI profiles are classified in the same group are clinically distinct (in terms of the criterion data described above) from adolescents whose test profiles are not classified in that particular group. In order to determine whether a particular test-defined group differed from the remaining adolescents in terms of dichotomous criterion data, Chi-square with Yates' correction was used. Student's t-test was used to assess group differences

in means for continuous data. Differences that reached the .06 level were considered to be significant. Whenever the group defined by a particular test pattern included 20 or more subjects, that group was divided and each half was compared to the subjects not classified in that group (18 of the 29 profile-designed groups were large enough to allow this replication procedure).

Although the broad generalizability of their findings has not been established, there are two features of Marks, Seeman, and Haller's procedures that provide some basis for confidence in the general significance of their results. First, of course, is the replication when the test-defined groups were large enough. Second, those investigators collected data from 74 different agencies in 30 states and the criterion data were provided by 172 different psychotherapists. This broad base of subjects and data certainly should have increased the representativeness of the sample and should have reduced the likelihood that a single narrow theoretical orientation dominated. Unfortunately, Marks, Seeman, and Haller did not present their findings in a tabular form that would have allowed the reader to evaluate them closely. Instead, the investigators integrated their data into rather general narrative descriptions of each group.

None of the existing actuarially derived descriptions have been cross-validated with samples of patients other than the derivation samples or settings. Thus, even though Marks, Seeman, and Haller replicated their analyses in 18 of their 29 test-defined groups, no replications have been reported by other investigators.

DRUG RESEARCH

The available research on drug abuse emphatically points out the fact that the determinants, the patterns of use, and the long term consequences of drug use vary widely among various SES groups, personality types, urban and rural geographic areas, and among users of different drugs. The similarity of the findings concerning opiate addicts reported by Berzins, Ross, English, and Haley (1974), and Goldstein and Linden's (1969) results from their study of alcoholics, led Berzins et al. to suggest that there may be test patterns that are related to the abuse of a variety of substances rather than to either illicit drugs or alcohol alone. Braucht, et al. (1973), have made the same point in their review of research on drug use among adolescents. It seems eminently reasonable to hypothesize the existence of several distinguishable and relatively homogeneous types or groups among the obviously heterogeneous entire population of drug users. The actuarial approach described above appears to offer one set of methods that will allow us to search for and identify those more homogeneous subtypes of drug users. The following section outlines rather tersely a set of specific procedures that should be useful in the development and validation of an actuarial system for use in drug research.

Selection of Criterion Data

Unfortunately there is little uniformity in the criterion data that have been reported by drug researchers so far. Most investigators have collected a limited set of criterion data and have not sampled the several domains that appear to be important. I would like to suggest seven types of data to be collected with the understanding that several of those types of data will be examined later as potential predictors of the clinically and socially important attributes of our patients. The seven types of data and suggested ways in which they might be collected are the following:

1. Ratings of observable behavior and personality characteristics. The Mental Status Examination Record (Spitzer and Endicott, 1970), the set of 108 personality descriptive statements used by Marks and Seeman (1963), or the Interpersonal Behavior Inventory (Lorr, Bishop, and McNair, 1965) all provide systematic procedures for collecting data concerning the patient's current status. It should be noted that nonprofessionals can be trained to produce valid interview-based ratings of many of these items (Robbins and Braroe, 1964).

2. Demographic characteristics. A relatively simple form can be derived to ensure the routine collection of this information.

3. Environmental characteristics (predrug and current). Although there is a great deal of discussion of the influence of environmental conditions on the use of drugs, there are very few methods available to systematically quantify theoretically reasonable dimensions of environments. The use of the several environment scales developed by Moos (Moos, Insel, and Humphrey, 1974) should be seriously considered. These can be administered to the subjects themselves and to their families. These scales are designed to assess several dimensions of the home or work environment rather than focusing on individual events or attributes of those environments.

4. Predrug behavior pattern. Braucht et al. (1973) have noted that one of the major defects of much of the drug research is that this sort of data is most often retrospective information. While that may continue to be a serious problem in some settings, it can be overcome to some extent by reference to school and court records in some instances. It should also be noted that short-term prospective studies have profitably incorporated data of this sort (Jessor, Jessor, and Finney, 1973). This general class of data should certainly be included in any study of persons at risk for drug abuse.

5. Patterns of drug use. A number of investigators have referred to a standard form with which to record history and patterns of use, and a systematic method for collecting these data should be used (Holroyd and Kahn, 1974; Jessor and Finney, 1973).

6. Response to (specified) treatment. The psychological literature on the assessment of response to treatment provides no consensus on how this complex task is to be accomplished. Rather than join that ongoing debate, it seems advisable to define improvement in objective terms, such as continuation in a treatment program, or "clean" urine as Zuckerman, et al. (1975) have done, or in terms of the reduction in the frequency or intensity of drug use (analogous to Green, Gleser, Stone, and Seifert's (1975) definition of "improvement" as some detectable reduction in the target symptom).

7. Psychometric data. Although reports involving a wide variety of psychological tests are available in the literature, I would recommend the use of a brief intellectual evaluation such as the Shipley-Hartford Scale and the use of several objective personality tests such as the MMPI (Hathaway and McKinley, 1967), the CPI (Gough, 1957), the 16PF (Cattell, 1970) or the PRF (Jackson, 1967). The reasons for administering more than one objective personality test were discussed earlier and reflect the expectation that any single personality test will allow us to identify test patterns that will account for only a limited number of our patients.

While this may appear to be a rather large amount of data to be collected, I would estimate the total time required of each patient would be no more than 5 or 6 hours. No more than 3 or 4 hours of skilled professional time would be necessary and the remainder of the information could be collected by paraprofessionals or carefully trained volunteers.

In order to ensure a reasonably representative sample of patients, it would be highly desirable for these data to be collected on several hundred patients being seen or treated in a number of different agencies. If, for instance, 25 different centers were to collect these data on two new patients per month for as long as one year, we would have an invaluable pool of data. That pool of data would allow us to critically evaluate the assumptions underlying the actuarial approach described in this paper.

Selection of Predictors

The current literature clearly indicates that no single variable and no single class of variables can account for all or even most instances of drug abuse. Any attempt, therefore, to explain or predict drug abuse must recognize the fact that even a highly valid variable or set of variables will at best account for only a portion of the persons under study.

It seems highly likely that at least three rather different types of factors may describe and predict drug users. Those classes of variables are:

1. Environment (Predrug and current)
2. Ratable observable behavior and personality characteristics
3. Personality-relevant test scales

If one accepts this reasoning, it seems appropriate to examine each of those classes of information as the set of predictor variables. It will be necessary to collect each of those classes of data from each of the patients to be studied.

Identification of Patterns of Scores on the Predictor Variables

Since there is no clearly superior method for grouping or typing the predictor data, a number of methods should be used in several concurrent analyses of the data. We should be prepared to capture or account for only some of the patient population in the several groups efined by patterns of the predictor data. We must also recognize at the outset that the clinical value of test patterns may be limited if we deal with groups as large as those identified by Berzins, et

al. (1974), and Lorr et al. (1965), where as many as 20% to 30% of the patients are assigned to a single test-defined group. A number of different parameters will have to be examined empirically in order to determine how much psychometric heterogeneity is allowable with each predictor pattern while achieving a clinically acceptable degree of criterion homogeneity within the several types.

Determining the Relationships Between Predictor Patterns/Criterion Characteristics

Let us assume that we have identified a test-defined (Prototype #1) group of 25 drug users on the basis of the D^2 procedure discussed earlier in this paper. And let us assume that all of the other data are on punch cards or magnetic tape. We may then determine the frequency of each of the discrete variables in the Prototype #1 group and in the remaining sample. These frequencies can be compared using Chi-square, and those criterion variables that are significantly more frequent among members of Prototype #1 are considered to be characteristic of persons who generate that pattern of predictor data. In a comparable fashion, the means can be calculated for each of the continuous variables in the Prototype #1 group and in the remainder of the sample. Those continuous variables (such as age, years of education) on which there are significant differences can then be used to characterize the members of the Prototypic group.

While examining the relationships between our criterion data and patterns of the several predictor variables, it seems reasonable to expect rather different sets of criterion characteristics to be associated with narrowly defined MMPI patterns and narrowly defined patterns of scores on Moos' Family Environment Scale. At this point it should not be disconcerting to find such predictor-defined groups to overlap completely or to overlap not at all.

Moving on to Other Predictors

As research has accumulated it has become increasingly obvious that an exceedingly complex set of factors are related to drug use. In view of this fact, the appropriate question for us to ask is, "Can any specific pattern of scores on a relatively inexpensive instrument, such as our psychological tests, identify a subgroup of persons whose drug-related behavior can be predicted or understood with a clinically and socially significant degree of precision?" If some pattern of scores on one of our inexpensive predictors can identify even 5% of drug abusers for us, we have a good test, a valid test, and a useful test. It is not necessary that such a predictor account for all drug abuse. If we find that one of our sets of predictor data can only identify 5% of our patients, but does so in a manner that allows us to make clinically effective decisions about those patients, let us continue to use that predictor in order to identify that important group of patients. Our clinical and scientific task is then to identify additional subtypes among the remaining 95% of our patients using other predictors.

RESOURCES AND REFERENCES

REPORTS OF EXISTING ACTUARIAL SYSTEMS

Gilberstadt, H., and Duker, J. *A Handbook for Clinical and Actuarial MMPI Interpretation.* Philadelphia: Saunders, 1965.

Gynther, M. D.; Altman, H.; Warken, R. W.; and Sletten, I. W. A new actuarial system for MMPI interpretation: Rationale and methodology. *Journal of Clinical Psychology,* 28:173-179, 1972.

Gynther, M. D.; Altman, H.; and Sletten, I. W. Development of an empirical interpretive system for the MMPI: Some after-the-fact observations. *Journal of Clinical Psychology,* 29:232-234, 1973.

Marks, P. A., and Seeman, W. *An Atlas for Use with the MMPI: The Actuarial Description of Abnormal Personality.* Baltimore: Williams and Wilkins, 1963.

Marks, P. A.; Seeman, W.; and Haller, D. L. *The Actuarial Use of the MMPI with Adolescents and Adults.* Baltimore: Williams and Wilkins, 1974.

BASIC REFERENCES TO ACTUARIAL METHODS

Gough, H. G. Clinical versus statistical prediction in psychology. In: L. Postman, ed. *Psychology in the Making.* New York: Knopf, 1962.

Meehl, P. E. *Clinical versus Statistical Prediction.* Minneapolis: University of Minnesota Press, 1954.

Meehl, P. E. Wanted--a good cookbook. *American Psychologist,* 11:263-272, 1956.

Meehl, P. E. The cognitive activity of the clinician. *American Psychologist,* 15:19-27, 1960.

Meehl, P. E. *Psychodiagnosis: Selected Papers.* Minneapolis: University of Minnesota Press, 1973.

Sawyer, J. Measurement and prediction, clinical and statistical. *Psychological Bulletin,* 66:178-200, 1966.

Sines, J. O. Actuarial methods in personality assessment. In: B. A. Maher, ed. *Volume 3: Progress in Experimental Personality Research.* New York: Academic Press, 1966.

OTHER REFERENCES CITED

Berzins, J. I.; Ross, W. F.; English, G. E.; and Haley, J. V. Subgroups among opiate addicts: A typological investigation. *Journal of Abnormal Psychology,* 83:65-73, 1974.

Block, J. On the number of significant findings to be expected by chance. *Psychometrika,* 25:369-380, 1960.

Bowers, K. S. Situationism in psychology: An analysis and a critique. *Psychological Review,* 80:307-336, 1973.

Braucht, G. N.; Brakarsh, D.; Follingstad, D.; and Berry, K. L. Deviant drug use in adolescence: A review of psychological correlates. *Psychological Bulletin,* 79:92-106, 1973.

Brill, N. Q.; Compton, E.; and Grayson, H. M. Personality factors in marijuana use. *Archives of General Psychiatry,* 24:163-165, 1971.

Cattell, R. B. Rp and other coefficients of pattern similarity. *Psychometrika,* 14:279-298, 1949.

Cattell, R. B. *Handbook for the Sixteen Personality Factor Questionnaire.* Champaign, Ill.: Institute for Personality and Ability Testing, 1970.

Cronbach, L. J., and Gleser, G. C. Assessing similarity between profiles. *Psychological Bulletin,* 50:456-473, 1953.

Davis, K. R., and Sines, J. O. An antisocial behavior pattern associated with a specific MMPI profile. *Journal of Consulting and Clinical Psychology,* 36:229-234, 1971.

Ekehammar, B. Interactionism in personality from a historical perspective. *Psychological Bulletin,* 81:1026-1048, 1974.

Endler, N. S. The case for person-situation interaction. *Canadian Psychological Review,* 16:12-21, 1975.

Ferguson, P.; Lennox, T.; and Lettieri, D., eds. *Drugs and Family/Peer Influence.* NIDA Research Issues Series, vol. 4. Washington, D. C.: Government Printing Office, 1974.

Gilberstadt, H., and Duker, J. Case history correlates of three MMPI profile types. *Journal of Consulting Psychology,* 24:361-367, 1960.

Gleser, G. C. Projective methodologies. *Annual Review of Psychology,* 14:391-422, 1963.

Goldstein, S. G., and Linden, J. D. Multivariate classification of alcoholics by means of the MMPI. *Journal of Abnormal Psychology,* 74:661-669, 1969.

Gough, H. G. *Manual for the California Psychological Inventory.* Palo Alto: Consulting Psychologists Press, 1957.

Green, B. L.; Gleser, G. C.; Stone, W. N.; and Seifert, R. F. Relationships among diverse measures of psychotherapy outcome. *Journal of Consulting and Clinical Psychology,* 43:689-699, 1975.

Hathaway, W. R., and McKinley, J. C. *The Minnesota Multiphasic Personality Inventory Manual.* Revised. New York: The Psychological Corporation, 1967.

Hogan, R.; Mankin, D.; Conway, J.; and Fox, S. Personality correlates of undergraduate marijuana use. *Journal of Consulting and Clinical Psychology,* 35:58-63, 1970.

Holroyd, K., and Kahn, M. Personality factors in student drug use. *Journal of Consulting and Clinical Psychology,* 42:236-243, 1974.

Huff, F. W. Use of actuarial description of personality in a mental hospital. *Psychological Reports,* 17:224, 1965.

Insel, P. M., and Moos, R. H. Psychological environments: Expanding the scope of human ecology. *American Psychologist,* 29:179-188, 1974.

Jackson, D. N. *Personality Research Form Manual.* Goshen, N. Y.: Research Psychologists Press, 1967.

Jessor, R.; Jessor, S. L.; and Finney, J. A social psychology of marijuana use. *Journal of Personality and Social Psychology,* 26:1-15, 1973.

Kaslowsky, M., and Deren, S. A comparison of three procedures for classifying addicts. *Journal of Consulting and Clinical Psychology,* 43:433, 1975.

Lindzey, G. Seer versus sign. *Journal of Experimental Research in Personality,* 1:17-26, 1965.

Lorr, M.; Bishop, R. F.; and McNair, D. M. Interpersonal types among psychiatric patients. *Journal of Abnormal Psychology,* 70:468-472, 1965.

Lykken, D. T. Statistical significance in psychological research. *Psychological Bulletin,* 70:151-159, 1968.

Moos, R. H.; Insel, P. M.; and Humphrey, B. *Preliminary Manual for the Family Environment Scale, Work Environment Scale and Group Environment Scale.* Palo Alto: Consulting Psychologists Press, 1974.

Owen, D. R. Classification of MMPI profiles from non-psychiatric populations using two cookbook systems. *Journal of Clinical Psychology,* 26:79-82, 1970.

Pauker, J. D. Identification of MMPI profile types in a female, inpatient, psychiatric setting using the Marks and Seeman rules. *Journal of Consulting Psychology,* 30:90, 1966.

Payne, F. D., and Wiggins, J. S. The effects of rule relaxation and system combination on classification rates in two MMPI "cookbook" systems. *Journal of Consulting Psychology,* 32:734-736, 1968.

Robins, L. N., and Braroe, N. W. The lay interviewer in psychiatric research. *Journal of Nervous and Mental Disease,* 138:70-78, 1964.

Robins, L. N., and Taibleson, M. An actuarial method for assessing the direction of influence between two datable life events. *Sociological Methods and Research,* 1:243-270, 1972.

Sines, J. O. Actuarial methods as appropriate strategy for the validation of diagnostic tests. *Psychological Review,* 71:517-523, 1964.

Smart, R. G., and Jones, D. Illicit LSD users: Their personality characteristics and psychopathology. *Journal of Abnormal Psychology,* 75:286-292, 1970.

Spitzer, R. L., and Endicott, J. *Manual of Instructions: Mental Status Examination Record (MSER).* New York: Evaluation Section, Biometrics Research, New York State Department of Mental Hygiene and Department of Psychiatry, Columbia University.

Wiggins, J. S. *Personality and Prediction: Principles of Personality Assessment.* Reading: Addison-Wesley, 1973.

Zuckerman, M.; Sola, S.; Masterson, J.; and Angelone, J. V. MMPI patterns in drug abusers before and after treatment in therapeutic communities. *Journal of Consulting and Clinical Psychology,* 43:286-296, 1975.

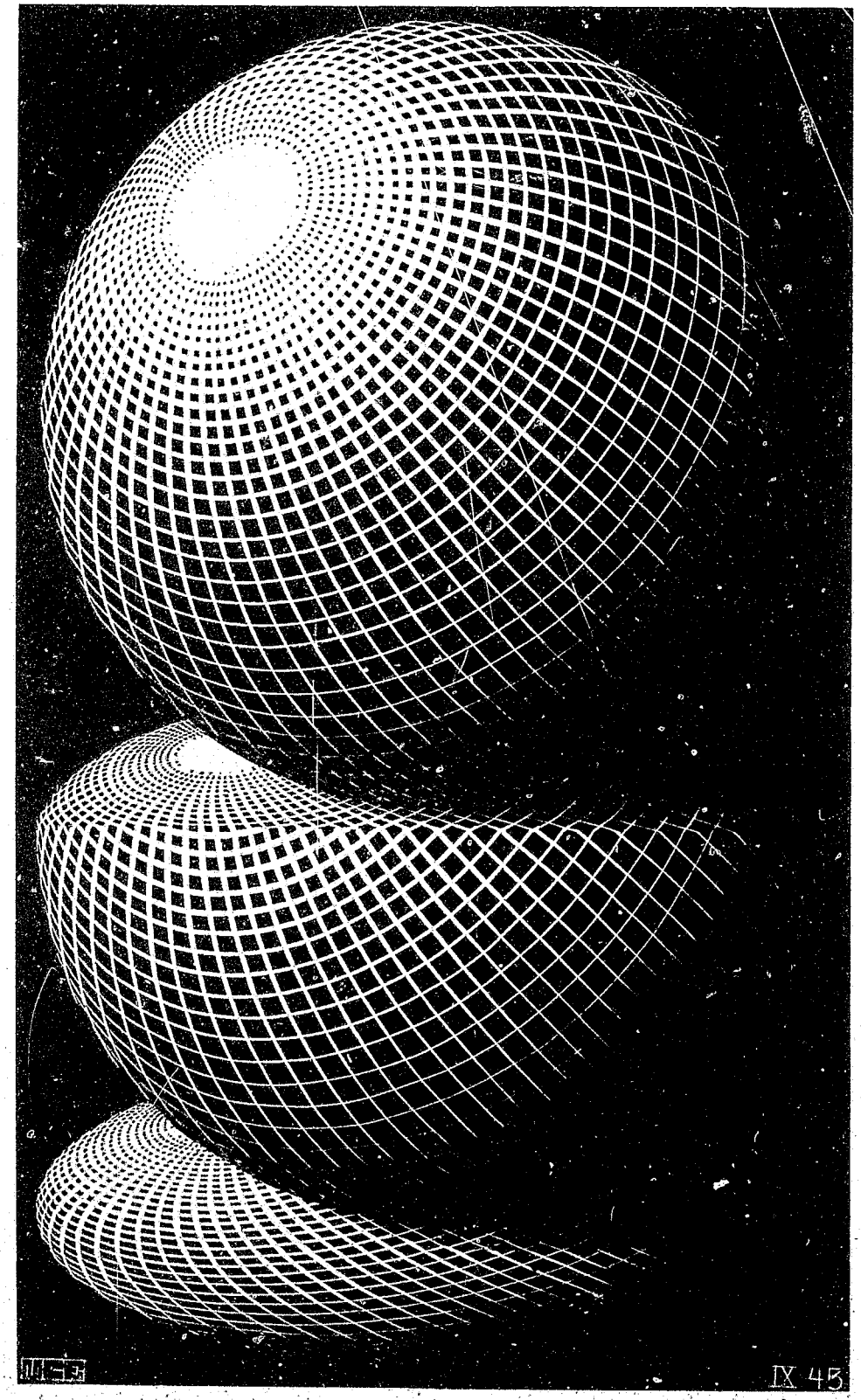


Plate 7.

INTRODUCTION

HISTORICAL BACKGROUND

Typing and taxonomic grouping have long been subjects of interest to scientists in many fields. Galen (circa 150 A.D.) defined nine temperamental types said to relate to a person's susceptibility to various diseases. Linneaus constructed a scheme for the classification of botanical specimens in the 18th century which had widespread impact on other fields. Most of the early work on classification was in botany and zoology. It led eventually to what is known as numerical taxonomy through the use of quantitative and statistical approaches to grouping objects. Among psychologists, some of the earliest techniques were proposed by Stephenson (1936), Zubin (1938), and Tryon (1939). The current surge of interest and publication began about 25 years ago, accelerated, no doubt, by the computer revolution. In 1953 Thorndike considered the problem of forming groups in terms of "who belongs in the family." In the same year, Cronbach and Gleser (1953) published their classic review of the problems of assessing similarity between objects. During the same period biologists like Sokal, Sneath, and Michener were developing parallel concepts and techniques for grouping biological specimens. Among psychologists, McQuitty was most active and productive, having published several dozen papers since 1956.

Since 1960 there have been literally hundreds of studies published by anthropologists, biologists, archeologists, geologists, information retrieval specialists, sociologists, and statisticians. Initially, these reports were published in technical journals by scientists concerned with different subject matter. Quite often these specialists were unaware of comparable development of terms, techniques, methods, and theories in border fields. The appearance of general methodological texts, however, has served to integrate the field and to provide a common language. There are general guides provided by Anderberg (1973), Everitt (1974), Jardine and Sibson (1971), Hartigan (1975), and Sneath and Sokal (1973). Sneath and Sokal as well as Jardine and Sibson focus primarily on the problems of biological taxonomists. Anderberg and Everitt are relatively general in approach and are best suited to social and behavioral scientists. Hartigan's text represents a statistician's compilation of algorithms (fixed procedures) for generating clusters and hierarchial structures. Cole (1969) includes a collection of papers presented at a conference on classification. Tryon and Bailey (1970) describe their BC TRY Program for factoring variables and grouping objects, but the focus is mainly on the method of factor analysis.

CLUSTER ANALYSIS AND RELATED PROCEDURES

Cluster and typological analysis is often confused with several related but distinctive procedures such as classification, discriminant function analysis, and identification (diagnosis). Classification, as a noun, is a systematic arrangement of objects into groups or categories according to known criteria. The term is also used to refer to the process of deciding into which of a number of categories, defined *a priori,* a new case should be allocated. Identification, assignment, and diagnosis refer to this same process of assigning a new observation or unit in an established set of categories. The defining or essential characteristics of each category are known. But the problem may be complicated and the process made uncertain if the criteria are vague, the characteristics subjectively defined, and the rules of assignment unspecified. A good example is the diagnostic system used to classify psychiatric patients. The symptoms are ill-defined, their number is not specified, and no rules are given for making decisions.

The procedure of discriminant function analysis is applied to known groups or categories on the basis of a set of measures not yet established. The investigator so weighs the predictor variables statistically as to maximize the separation of the groups. First, the number of dimensions of difference is ascertained, then the group means are tested for significant differences. Once these issues are determined, a set of weights is derived to allocate new cases to one of the groups in order to minimize the number of misclassifications. Discriminant analysis,

while operating in known groups, such as Protestant, Catholic, and Jewish, seeks to establish a new weighting scheme for assignment of new cases on the basis of a set of untried criteria.

In cluster and typological analysis little or nothing is known concerning the nature of the groups or categories, or their number. The defining characteristics of the unknown group are also unspecified. Cluster analysis is thus a search procedure for finding "natural groups", for discovering a conjectured structure, or for imposing a useful conceptual framework where none exists.

RATIONALE: AIMS AND USES

The general purpose of a cluster analysis of multivariate data is to group together persons, objects, concepts, or events into coherent classes as the basis of their measured similarities. Such analyses assume that the number and nature of the classes or groups are unknown *a priori*. The main goals of analysis are to:

(a) recover or identify "natural" clusters of entities within a mixture believed to be drawn from several populations sampled;
(b) generate a conceptual scheme for classifying entities which will reflect and summarize their interrelationships;
(c) discover structure inherent in a body of data when the data do not represent a sample;
(d) test hypotheses about groupings believed to be present in the data.

As a general practical matter, types are useful simply to facilitate communication. Types are easy to remember, to report, and to differentiate from objects in general. Categorization, including concept formation, provides direction for instrumental behavior. A depressive is treated differently from a person known to be a paranoid or psychopath. To categorize an entity is to bring to mind immediately an associated set of defining characteristics.

In biology and the social sciences, the researcher is often faced with a mixture of observations from several populations. The problem is one of allocating individual cases to an unknown number of categories representing different families, classes, or genera. To the extent that the search is successful, the groups present are recovered. In addition, knowledge and understanding of a domain may be substantially increased. New laws may be discovered and relations hidden or obscured may be found.

In many fields, researchers are often confronted by a very large mass of data involving numerous measures and observations. The test variables can be reduced by applying the techniques of factor analysis to isolate a smaller number of descriptive dimensions. Cluster analysis techniques are similarly useful for reducing the number of subjects or cases. Indeed, comprehension is substantially facilitated if numerous individual cases can be organized into a smaller number of cohesive groups or into a hierarchical arrangement of classes. An example of such use is the application of hierarchical grouping to air force jobs. Christal and Ward (1968) used information on job functions to arrange the jobs into meaningful families of increasing generality. Frank and Green (1968) grouped 88 cities on the basis of 14 variables such as city size, per capita income and newspaper circulation in order to discover potential test markets. Rice and Lorr (1968) describe grouping 150 ceramic pots from the Smithsonian Institute into three prototypes on the basis of 17 measures.

The investigator may have hypotheses regarding the existence of several subgroups within a specified domain. Cluster analysis can be useful in testing out such hunches. Or, observers may have established through clinical or field observation that certain subgroups can be discerned and want to confirm their findings by more objective means. For example, Kretschmer and Sheldom had hypothesized three body types: ectomorphs, endomorphs, and mesomorphs: Lorr correlated the score profiles of some 70 male subjects and established that three prototypic body types could be found. Everitt et al. (1971) applied cluster analysis toward validation of traditional psychiatric classes with a fair degree of success. In an effort to differentiate depressed patients, Paykell (1971) applied cluster analysis to patient symptoms. In another study, Lorr et al. (1973) used interview ratings as a basis for testing the hypothesized existence of three depressive subtypes: anxious, retarded, and hostile. A cluster analysis of the correlations among members of two large samples served to confirm these conjectures.

Cluster analysis techniques can also be applied to discover structures inherent in a body of data when the observations really do not represent a sample from a known population. Miller (1969) used complete linkage analysis to investigate verbal concepts. Fifty students divided 48 common nouns into piles with similar meaning. The analysis suggested five clusters of nouns: living things, nonliving things, quantitative terms, social interaction, and emotions. Clustering techniques were applied by Manning and Watson (1966) to 99 patients with heart disease who were described in terms of 129 items. The three clusters agreed substantially with physician's diagnosis. Hodson (1969), an archaeologist, applied cluster technique to 50 assemblages of stone tools from France and Europe. The tools were divided by expert judges into 46 classes. An average linkage analysis suggested eight clusters closely related to those hypothesized.

A series of related typological studies of alcoholics constitute a good illustration of the usefulness of typological techniques. Goldstein and Linden (1969) reported a study that yielded four inpatient MMPI alcoholic profile types that matched known actuarial types with addictive problems. Subsequently, Berzins et al. (1974) found two MMPI types for male and female addicts which were highly convergent with those of Goldstein. Nerviano (1976) then established seven common personality patterns among alcoholic males on the basis of a typological analysis of 366 subjects described by 12 personality inventory scales. He concluded that these were common personality patterns characteristic of psychiatric patients generally and not unique to alcoholics.

METHODS, PROCEDURES AND ASSUMPTIONS

THE PROCESS OF ANALYSIS

The process of clustering can be broken down into a number of steps. A listing of the usual choices and assumptions made provides insight into the nature of the analysis. The sequence is usually as follows:

(a) Select a representative set of entities to be studied, or select appropriate samples from populations of interest. The entities may be people, pots, plants, documents, languages, legislators, bees, birds, micro-organisms, etc.

(b) Define the domain of similarity to be studied and select a representative set of qualitative or quantitative attributes.

(c) Convert scores into a comparable metric if this seems needed. Decide whether or not to include categorical as well as continuous variables.

(d) Decide whether to conduct a dimensional analysis (factor analysis) in order to reduce the number of descriptor variables to simplify them into composite scales.

(e) Select a suitable index of similarity or dissimilarity between pairs of entities.

(f) Choose a structural model for the clusters or types anticipated. The main models are the compact or homogeneous, the chained or continuously connected, and the hierarchical. The distinction is based primarily on the nature of the relationship among entities (symmetric and transitive vs. asymmetric and transitive).

(g) Select an appropriate method of clustering, an efficient algorithm, and apply the procedure to the matrix of indices of similarity--dissimilarity. The algorithm chosen usually determines the number of clusters found.

(h) Determine the mean profiles of the various clusters found or convert into a tree-structure or dendogram.

(i) Interpret the results and choose some decision function to allocate new cases to the subgroup to which they belong. To this end, methods such as discriminant functions, multiple cutting scores, and Bayesian analysis may be applied to assign new cases to one of the clusters found.

The major steps in the process of searching for groups or categories have been sketched. Each step is associated with a problem and several decisions to be made. The discussion that follows will be concerned with these problems.

Choice of Entities to Be Studied

As was indicated earlier, the entities may be persons, stimuli, events, concepts, animals, plants, or languages. The entities are denoted as subjects, observations, cases, and data units.

The selection of a set of entities for study may be complicated in various ways. Suppose the aim is to identify "natural groups." Then it becomes important to select a random sample. It follows that existing groups are likely to be represented in the sample in proportion to their relative size in the population. Consideration should then be given to enlarging the size of the suspected smaller group. If an entire universe is analyzed, then no important group sources should be omitted.

Another type of study is concerned with stimuli such as nouns and verbs, or the data units may be skulls, artifacts from an area, symptoms of a disease, or varieties of ships. Here sampling of a population is not involved, but the entities must represent the domain of interest. It becomes important not to overlook any source of variation.

Choice of Variables

Here definition of the domain of similarity becomes important. If an investigator is studying psychotics, depressives will not appear if none of the symptoms or behavior characteristics of depressives is included. As far as possible, all possible sources of individual differences must be included. When relevant discriminating variables are left out of an analysis, some groups will merge and remain undifferentiated and confused.

Two related problems are discriminant validity and reliability. A measure has discriminant validity if it differentiates among members of a group. If all subjects are included or excluded, agree or disagree, if all say true or false, then the item is worthless. A variable must also be dependable over time and generalizable to comparable observations. To augment the stability of variables and their generalizability, it is a common practice to combine several additively for greater reliability.

Choice of Metric

In many studies the variables selected may vary markedly in metric and scale. For instance, in studying a body type the variables may be descriptive of weight, height, length of nose, degree of muscularity, and width of head. Or, the variables included may be demographic such as age, sex, level of education, religion, and marital status. Inspection of these variables makes it clear that some transformation is needed to express the variables in comparable units. The usual recommended procedure is to standardize all measures so that each scale has a mean of zero and a standard deviation of one.

Data Reduction

Each subject receives a set of scores on the descriptor variables called a score profile. An estimate of the degree of similarity or proximity of two individuals is usually expressed in terms of distance or correlation. Both measures are meaningful only when the variables involved are relatively independent of each other. It follows from this that the profile elements ideally should represent independent dimensions of variation. Should it be necessary, then the investigator would do well to conduct a factor analysis of the variable intercorrelations. The composite scores that result will enhance reliability and facilitate interpretation of profiles.

VARIABLES AND SCALES OF MEASUREMENT

Many discussions in measurement of properties make the convenient assumption that all variables are of a single type. Usually, the variables are assumed to represent continuous and equal interval scales. However, in the practical world the variables by which people are described are a mixture. Some are continuous variables like age, degree of agreement, and total score. Others are qualitative or categorical like religion, color of eyes, or occupation. A special variety of categorical variable is binary or dichotomous and takes on only two values like 0 and 1. Examples are statements or questions that are answered True or False, Yes or No.

Measurement is the assignment of numerals to events or objects according to rules. In the social and behavioral sciences, particularly, it is important to recognize and to allocate variables to one of four kinds of scales of measurement. The most rudimentary is the nominal or classificatory scale. Numbers or symbols are used to classify a person, object, or property. When numbers or symbols are used to identify groups to which objects or persons belong, these symbols constitute a nominal scale. The psychiatric classificatory systems of diagnostic groups constitute such a scale, as do the numbers assigned to football players on a team.

All scales have certain formal properties that provide fairly precise definitions of the scales, the operations of scaling, and the relations among the objects. In a nominal scale, the operation is one of partitioning a given collection of objects into a set of mutually exclusive subsets. The relation between objects is one of equivalence, meaning that members of any subset must be equivalent in the property being scaled.

In any nominal scale, the subsets may be represented equally well by any set of symbols. Thus, the nominal scale is said to be "unique up to a one-to-one transformation." The symbols or numbers designating the subsets may be interchanged providing this is done consistently and completely. In fact, the principle of permissible transformations for any scale type is that it does not change any implications about the empirical system it represents. For classificatory scales, the only kind of descriptive statistics are those which would be unchanged by such transformation. Included here are the mode and the frequency count.

Ordinal scales reflect consistent rank orders. Objects in one category of the scale differ from objects in other categories of the scale by being greater than or less than. Examples of this relation are: higher, more difficult, preferred to. Moh's scale of hardness represents an ordinal scale. One mineral is harder than another if the first scratches the second but not vice versa. The scale also reflects a transitive relation because mineral X scratches Y, Y scratches Z, and X must scratch Z. Military rank, social status, and most personality inventories and tests of ability yield scores that represent ordinal scales. The ordinal scale differs from the nominal by incorporating the relation "greater than" in addition to the relation of equivalence. Ordinal relations are irreflexive, asymmetric, and transitive. Irreflexive means that for any X, X is not greater than itself. Asymmetric means that if X is greater than Y, then Y is not greater than X. Any order-preserving transformation will not change the information given on an ordinal scale. Therefore, an ordered scale is "unique up to a monotone transformation." The monotone transformations must preserve the order of the numbers assigned say, to minerals, or to persons rated as to social status. It does not matter what numbers are given a pair of subsets just as long as the higher number is given to the members of the class which is "greater than" or "preferred to." The term "monotone" means that the variable increases or decreases systematically.

An interval scale is characterized by a constant or equal unit of measurement. A real number is assigned to all pairs of objects on the ordered set. In other words, an interval scale assigns a measure of the difference between two objects. The scale has all of the characteristics of an ordinal scale but in addition provides a distance between any two objects. It is then possible to say not only that A is greater than B, but also that A is so many units different from B on variable X. Temperature is an interval scale since equal intervals of temperature correspond to equal volumes of mercury expansion. Now any change in the numbers assigned to the positions of the objects measured on an interval scale must preserve not only the ordering of the objects, but also the relative difference between them. Therefore, an interval scale is "unique up to a linear transformation." The information in the scale will not be affected if each number is multiplied by a constant, and a constant is added to this product. For example, Centigrade degrees of temperature can be multiplied by 9/5 and a constant of 32 added to yield a Fahrenheit degree. All of the familiar parametric statistics such as means, standard deviations, and correlations are applications to interval scale data.

A ratio scale is an interval scale with a true zero point as its origin. In a ratio scale, the ratio of any two scale points is independent of the unit of measurement. Then if A is greater than B, it is possible to say that A is X_A/X_B times greater than B. Ratio scales are extremely rare in the social or behavioral sciences. Ratio scales are "unique up to multiplication by a positive constant." Examples are length, time intervals, loudness (sones), and brightness (brils).

MEASUREMENT OF SIMILARITY-DIFFERENCE

Before discussing measures of similarity or difference it is important to emphasize that similarity is not a general quality. When people are compared on independent dimensions, then those who are similar or close on one dimension need not be similar on other dimensions. People may be very much alike say, in social attitude, but different in food preference, educational background or other psychological or nonpsychological attributes. It is, therefore, always necessary to specify the domain of similarity-difference in discussing the similarity of persons, objects, or events. If a group of people are found to be similar on one set of scores, it is not justifiable to assume their similarity in general.

The basic data usually consist of a set of scores, called a profile, for each person or object. The score of person i on variate j can be symbolized as x_{ij}. Let the number of variates be P and the number of persons N. Then Person A's scores are denoted X_A: $[X_{A1}, X_{A2}, ----X_{AP}]$ and Person B's scores are denoted X_B: $[X_{B1}, X_{B2}----X_{BP}]$. It is common practice to visualize the N persons as N points or vectors in P-dimensional space, each variate being represented by a coordinate axis. Then the X_{ij}'s represent coordinates of the point or vector. The vector is usually defined as a quantity having magnitude and direction and is represented by a directed line segment whose length represents the magnitude and whose orientation in space represents the direction.

The more similar the scores of A and B, the closer their vectors in space, and conversely the more dissimilar their scores, the more distant their vectors. In two-dimensional space, the squared distance between A and B may be expressed by the Pythagorean theorem

$$D^2_{AB} = (X_{A1} - X_{B1})^2 + (X_{A2} - X_{B2})^2$$

Then, if the theorem is generalized to P-dimensional space, we have

$$D^2_{AB} = \sum_{j=1}^{P} (X_{Aj} - X_{Bj})^2 \qquad (1)$$

where Greek Sigma means sum of the squared differences between A and B on each of the variates.

The distance formula as a measure of dissimilarity between profiles may be applied to any type of scores. But the meaning of the distance measure depends on the nature of the scores. The original score set may be centered about the person's own mean, or it may be standardized. These transformations alter the meaning of the scores and reduce the number of degrees of freedom involved which is P, the number of variates. One source of variation is the level or mean ($\bar{X}_i$) of all scores for a person. For example, a bright person's score mean is likely to be high on intelligence tests while a dull person's score mean is likely to be low. Another source of variation is the scatter or dispersion of a person's scores. Scatter (S_i) is measured by the square root of the sum of squares of the person's deviation scores about his own mean. Scatter can be represented geometrically by the length of a person's score vector. The third characteristic of a score profile is its direction or orientation. In other words, the orientation indicates which scores are high and which are low.

To illustrate the loss of information, let us suppose persons A, B, and C were given scores on five tests 1, 2, 3, 4, and 5 as shown:

	1	2	3	4	5	Sum	$\bar{X}$
A	0	-1	2	4	0	5	1
B	-2	-3	0	2	-2	-5	-1
C	4	1	-3	1	-3	0	0

Then D^2_{AB} by formula #1 becomes 20 and the means are 1, -1, and 0. If score level or mean now is removed from each profile, the deviation scores are

	1	2	3	4	5	Sum	S_i^2	S_i
A	-1	-2	1	3	-1	0	16	4
B	-1	-2	1	3	-1	0	16	4
C	4	1	-3	1	-3	0	36	6

Since score level is removed, D^2_{AB} is now zero. Differences in scatter between profiles are eliminated by dividing each deviation score by the person's scatter, a process called standardization.

Geometrically, this operation stretches or reduces all vector profiles to unit length. For example, if A's deviation scores are divided by the scatter and then squared, their sum $(\frac{1}{16} + \frac{9}{16} + \frac{1}{16} + \frac{4}{16} + \frac{1}{16})$ equals unity.

Distance is the most common index of dissimilarity-similarity used. The correlation coefficient Q_{AB} is another index in frequent use. As may be seen from Table 1, a correlation is the sum of the crossproducts of the deviation scores of persons A and B, divided by the product of their scatter indices. Interpreted geometrically a correlation is the cosine of the angle of separation between two vectors, each of unit length. When the angle of separation is 90 degrees, the cosine is zero and there is no correlation. When the vectors coincide, the angle of separation is zero and the cosine (and thus the correlation) is 1.00. As may be seen from the formula for Q, variation due to mean level has been removed and the scatter has been made constant. Therefore, a Q correlation operates in K - 2 space.

THE SIMILARITY INDICES

The most commonly-used indices of similarity-dissimilarity are shown in Table 1. All of the general reference texts listed here discuss these indices and suggest computer programs. A separate set of references to journal articles on similarity indices is also given. A good general introduction may be found in Cronbach and Gleser (1953).

The first three indices are distance measures designed for use with continuous scores. Cattell's r_p is also based on distance but is designed to vary between +1 and -1 like a correlation coefficient. The correlations and congruency coefficients are measures of angular separation between profile vectors rather than distance. The last four indices are applicable to binary data such as given in a 2 by 2 table shown at the bottom of the table. The choice among them is dependent on whether or not negative matches are regarded as useful information. The disagreement index corresponds to D^2_{12} when the binary values are zero or one. The matching coefficient is the complement of D^2_{12} where a + d represent co-presence or co-absence of attributes. Jaccard's Connection index omits d, the number of negative matches. The Holley H index of agreement is the congruency coefficient for binary variables. It represents the proportion of matches minus the proportion of mismatches. The distance D_{12} and the correlation coefficient Q are related algebraically. If scores have been standardized around the person's own means, then $D^2_{12} = 2(1-Q_{12})$.

The Euclidean distance measure is the most widely used dissimilarity measure but there are other possible metrics. The city block or taxicab metric sums the absolute distances between objects on successive variables. The argument given for its use is that either of the two objects are described in terms of two variables with equal scale units. They should have the same distance whether they are two units apart on each variable, or they are one unit apart on one, and three units apart on the other.

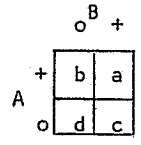
The generalized distance measure of Mahalanobis is designed to measure the distance between groups. This D measure is an index in which the independent (orthogonal) components of the original set of variates are assigned equal weights. Thus unreliable and unimportant factors are weighed equally with the first few. When the original variates are standardized and uncorrelated, $\bar{D}^2$ equals Euclidean D^2. A useful exposition is available (Overall, 1964) and the method is used in certain programs such as Freedman and Rubin (1967).

TYPOLOGICAL MODELS

There appear to be three major structural models in typing and cluster analysis. The first of these may be called compact or homogeneous, the second chained or continuously connected, and the third a hierarchical scheme. The first two structures reflect rather different relations among object pairs in the cluster. Members of the compact type are said to be similar or dissimilar, consonant or dissonant, alike or different, close or far, confusable or distinguishable. A term that has come to be used to convey all these meanings is proximity. The proximity relations between object pairs, as reflected in these terms are those of reflexivity, symmetry, and transitivity.

Within the chained type, ordinal (dominance) relations exist among objects within a type. One object is said to be greater than, dominant over, preferred to, or chosen over another. If the relation "closest to," or "nearest neighbor of," holds between a sequence of two or more objects, a partially ordered scale is implied. The dominance relation has the property of being irreflexive, asymmetric, and transitive. Ordinal relations are found in sociometric choices, communication networks, preferential orders, and competitive game rankings.

Table 1. Indices of Similarity and Dissimilarity.

Name	Formula	Reference
Euclidean distance	$\sum_{j}^{p}(X_{1j} - X_{2j})^2 = D^2_{12}$	Cronbach and Gleser
City-block metric	$\sum_{j}^{p}\lvert X_{1j} - X_{2j}\rvert =$	Johnson and Wall
Mahalanobis Distance	$d^1W^{-1}d = D^2$	Overall
Profile Similarity	$\frac{2\Sigma o_j^2 - \Sigma d_j^2}{2\Sigma o_j^2 + \Sigma d_j^2} \doteq r_p$	Cattell
Correlation Coefficient	$\frac{\sum_{j}^{p}(\bar{X}_{1j} - \bar{X}_1)(\bar{X}_{2j} - \bar{X}_2)}{[\Sigma(\bar{X}_{1j} - \bar{X}_1)^2\,\Sigma(\bar{X}_{2j} - \bar{X}_2)^2]1/2} = Q_{12}$	Pearson
Congruency Coefficient	$\frac{\sum_{j}^{p}X_{1j}X_{2j}}{[\Sigma X_{1j}^2 \Sigma X_{2j}^2]1/2} = C_{12}$	Cohen
Disagreement Index	$\frac{b+c}{a+b+c+d} = d^2_{12}$	
Matching Coefficient	$\frac{a+d}{a+b+c+d}$	Sneath and Sokal
Connection	$\frac{a}{a+b+c}$	Jaccard
Agreement Index	$\frac{(a+d)-(b+c)}{a+b+c+d} = H$	Holley and Guilford
Fourfold Table		

A helpful definition of a compact type is a subset of entities, each member of which is more like every other entity in the type than it is like entities in any other type. The chained type may be defined as a subset of entities such that every member is more like some one other member than it is like any other type. Figure 1 gives an illustration of these two kinds of clusters in two-dimensional space.

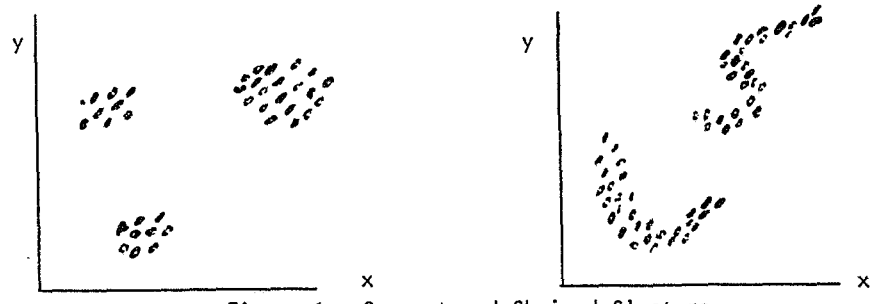


Figure 1. Compact and Chained Clusters

The third model, a hierarchical scheme, is usually represented by a hierarchical tree or dendogram, the term used by biologists. A hierarchy may be seen as a nested set of clusters in which each level is assigned a rank. Elements of the levels are called taxa (by biologists), and each taxon is assigned the rank of the cluster to which it belongs. Suppose N entities are given, and a sequence of N clusters is generated. As we move up the hierarchy, beginning with each entity as a cluster, each cluster (except the first) is obtained by a merging or union of clusters at the previous level. Thus the entire hierarchy is a family of clusters which also includes the set of all entities. Any two clusters in the hierarchy are discrete (disjoint) or one includes the other. As will be seen later in describing procedures for hierarchical schemes, the defining relations between entities result in slightly different structures. When the relations are reflexive, symmetric and transitive, the clusters are compact. When the relations are ordinal (asymmetric and transitive), the clusters are continuously connected chains (Johnson, 1967).

A CLASSIFICATION OF CLUSTER TECHNIQUES

Each of the general methodological texts available offer somewhat different schemes for classifying cluster analyses procedures. The same may be said for such reviewers as Cormack (1971), Ball (1971), and Bailey (1975). The distinctions offered here are the following:

(1) Density or Mode-Seeking is a process of searching for modes or regions of high density for entities in attribute space.

(2) Partitioning is the process of subdividing a collection or set of entities into mutually exclusive classes or subsets on the basis of a criterion.

(3) Clumping is the process of grouping objects into overlapping subsets which are called clumps.

(4) Hierarchical clustering is a process in which entities are grouped into clusters and the clusters themselves are in turn merged into clusters at successive levels to form a tree.

Density Search Techniques

A cluster may be represented by a swarm of points in P-dimensional space. The points are concentrated in some regions but not in others. Methods of cluster analysis which use this viewpoint search out regions of high density called modes. The structural model implied here is the compact, internally homogeneous cluster.

The density seeking techniques include complete and average linkage to be described shortly. Another procedure (Taxometric Maps) has been developed by Carmichael and Sneath (1969). In addition, there is the sophisticated method of mixtures developed by Wolfe (1970) and others. Members of a type or cluster in general differ from one another on most or all of the measures. Since members of a cluster differ, it is reasonable to assume the existence of a probability distribution of these characteristics. The combined population taken from all clusters will have a probability

distribution which is a mixture of distributions. The problem is to identify and describe the component distributions from a sample drawn from the mixture. Usually component distributions are unimodal, but in cluster analysis multimodal distributions must be resolved into simple components. Wolfe (1970) uses the method of maximum likelihood to estimate the mixture proportions, their means, and their covariance matrices. Each distribution, solved iteratively, indicates a separate group, and objects are assigned to the group for which their probability is greatest. The process is often begun with an initial set of clusters obtained through use of the K-means approach. Wolfe's general NORMIX program and his simpler NORMAP program have been applied to grouping occupations (1970) and to the classification of psychiatric patients (Everitt et al., 1971). Ideally the method requires large sets of data, and substantial amounts of computer time may be consumed.

Partitioning

The partitioning techniques usually seek to partition the set of entities so as to optimize some predefined criterion. Since a partition is a system of mutually exclusive subsets, there is nearly aways the need to decide on the number of groups present. Many of the methods do allow the number of groups to be changed during the course of analysis. Another characteristic of the partitioning methods is that they allow for corrections and relocations of the entities when the initial location was poor. The methods thus differ in the method of initiating clusters, the method of allocating entities to new clusters, and in procedures for reallocating entities to revised clusters. The majority of these techniques seek to minimize within-group distance among entities. They begin by finding K points in P-dimensional space, which serve as initial estimates of cluster centers. Entities are then allocated to the cluster whose mean they are nearest. Estimates of these centers are updated after each entity is assigned to a cluster. Once an initial classification has been found, a search is made for entities which should be reallocated. In general, relocation proceeds by considering each entity in turn for reassignment. Reassignment takes place if its addition improves the error term.

In MacQueen's "K-Means" method, the initial step is to take the first K entities in the data set as clusters of one member each. Each of the remaining entities are assigned to the cluster mean that is nearest. After each allocation, the mean of the cluster is recomputed. After all of the entities have been allocated, the revised cluster centroids are used as centers and the procedure is applied again. Of similar procedures proposed by Forgy and Jancey, the MacQueen procedure is simplest and least expensive.

The logic behind the K-means approach, simply stated, is to minimize the within group sum of squared differences of the partition. This is the same as maximizing the between cluster differences. The criterion for deciding if convergence has taken place is the stability of cluster membership. A convergent K-means process is offered in Wishart's RELOC and McRae's MIKCA computer programs, as well as in Anderberg. More complex and elaborate methods are found in Ball and Hall's (1965) program called ISODATA. Friedman and Rubin (1967) have a related optimizing cluster method which is available under the IBM SHARE system.

Clumping Techniques

In language studies, classifications are desired which permit an overlap between clusters. Since words may have several connotations, they may belong in several places. The techniques that allow for overlap are known as clumping techniques. Parker-Rhodes and Needham first introduced clumping methods. Jardine and Sibson (1968) have sought to construct several algorithms. Their method consists of representing each point by a node on a graph, and connecting all pairs of nodes which satisfy a specified inclusion criterion. Then a search is made for the largest subsets of entities for which all pairs of nodes are connected (maximally complete subgraphs). An algorithm for implementing this method may be found in Cole and Wishart's (1970) CLUSTAN program.

Hierarchical Clustering Methods

Hierarchical clustering techniques may be separated into the agglomerative and the divisive. The agglomerative methods begin with the N entities and successively merge or combine the two most similar. The divisive procedures successively subdivide or partition the entire collection into finer and finer subsets. The results of both techniques may be represented by a tree which is a two-dimensional diagram. The agglomerative methods build a tree from branches to root and the divisive begin at the root and subdivide the clusters into branches (see Figure 2).

Given an N by N matrix of similarity coefficients, (1) the process begins with N clusters each consisting of only one entity; (2) the matrix is searched for the most similar pair of clusters;

(3) the number of clusters is reduced to N-1 through a merger of the chosen pair; (4) steps (2) and (3) are followed N-1 times until all entities are in a conjoint cluster. At each stage the identity of the clusters combined are recorded as well as the similarity index between them. The process can be followed using correlation or distance measures.

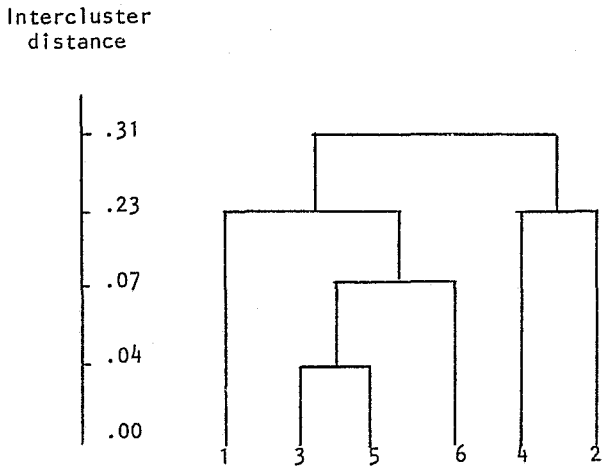


Figure 2. Hierarchical Tree Structure

There are six common bases for merging entities or clusters in a hierarchical analysis. These are as follows:

(1) Single Linkage (Nearest Neighbor). Single entities are merged on the basis of the distance between the two closest members of clusters. The distance between clusters is defined as the distance between their closest members (see Figure 3).

(2) Complete Linkage (Farthest Neighbor). Single entities are merged on the basis of the distance between the most distant pair. The distance between clusters is defined as the distance between their most remote pair. The value of this distance is the diameter of the smallest sphere which can enclose the combined clusters. Complete linkage, like single linkage, is invariant (unchanged) by monotonic transformations of the similarity measures (see Figure 3).

(3) Average Linkage method. The method defines the distance between groups as the average of the distances between all pairs of entities in two clusters.

(4) Centroid Cluster Analysis. The procedure clusters hierarchically by merging at each stage the two clusters with the most similar means or centroids. Sokal and Sneath (1963) describe this as the "pair group" method.

(5) Ward's (1963) procedure seeks to minimize the loss of information that results from combining entities into clusters. The error is measured by the total sum of squared deviations of every point from the mean of the cluster to which it belongs. At each stage the two clusters are merged that result in the minimum increase in the error seen. Clusters are combined on the basis of the minimum distance of pairs.

(6) Wishart's CLUSTAN 1B program was written to cover all of the above procedures. Veldman (1967) has a program called HGROUP which can handle 100 variables and 100 subjects using Ward's hierarchical grouping procedure. Johnson (1967) has developed two very rigorous hierarchical clustering schemes based on single and complete linkage. His Minimum Method corresponds to single linkage analysis and his Maximum Method corresponds to complete linkage. Both procedures satisfy what is known as the "ultrametric inequality." These procedures are especially good for small sets of entities.

METHODS OF LINKAGE ANALYSIS

Single Linkage Analysis

Perhaps the simplest and earliest of clustering methods developed is single linkage analysis. First introduced by Florek (1951), it was proposed independently by Sneath and by McQuitty in 1957. Lance and Williams (1967) call it the "nearest neighbor" technique. If correlations are involved, a link is defined as the largest index an object has with any other object. Should the index be a distance measure, then a link is the smallest distance an object has with all objects. The single linkage method thus generates a type or cluster in which every member is "closer to" or "more like" some one other member of the cluster than to members of other clusters. The object relations of "closer to" or "more similar to" are here asymmetric and transitive. The type is then a subset of continuously connected objects. Usually a type is begun by a reciprocal pair as a nucleus with other objects added if they are closest to at least one object in the cluster. A reciprocal pair exists if object A has its closest neighbor B and B has as its closest neighbor A.

The structural model appropriate to single link clusters is that of a chained or continuously connected subset. Ideally this procedure should be applied to dominance or order data. The investigator searching for compact clusters is likely to find that single linkage leads to long stringy-like groupings. This means that members of a cluster at one end will resemble each other but not members of the cluster at the other end. The technique is especially useful for isolating geometric figures like rings, circles, curves, or one-dimensional orders. Biologists have given this method priority (Jardine and Sibson).

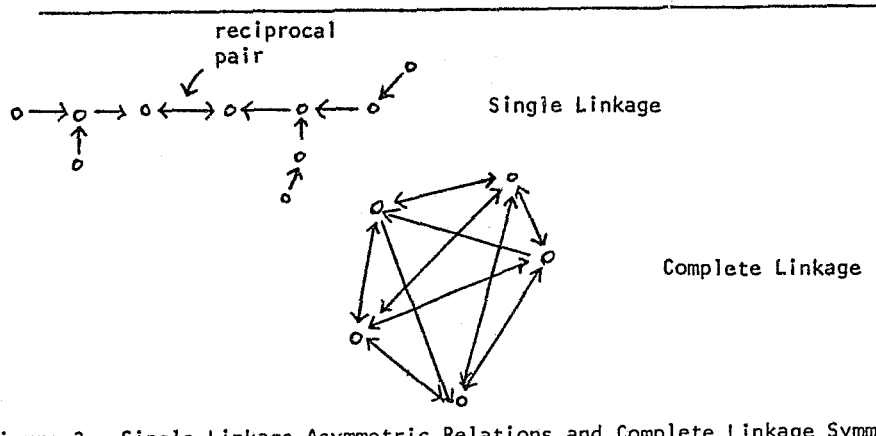


Figure 3. Single Linkage Asymmetric Relations and Complete Linkage Symmetric Relations

Complete Linkage Analysis

Complete linkage has been developed independently in many fields. Sorenson developed the method in 1948 for use in ecological studies. McQuitty (1961) suggested typal analysis for categorizing people or psychological test items. The method is also known as "farthest neighbor" clustering. An object considered for admission to a cluster has a similarity index equal to the farthest member within the cluster. The method generally generates or leads to tight hyperspherical discrete clusters. The subsets will vary in compactness as a function of the definition of a link. Complete linkage usually requires the investigator to set some arbitrary lower limit to the magnitude of the correlation coefficient for inclusion into a cluster. Or, if the measures represent indices of dissimilarity such as distance, some upper limit of difference must be set to qualify the object for entry into the cluster. The algorithm begins with the most highly correlated pair that satisfy the inclusion criterion. A third object is added to the nucleus if it is linked to both members, and a fourth object is added if it too is linked to all objects in the cluster. This means that the relations among cluster members are all reflexive, symmetric, and transitive, implying equivalency among object members of a subset.

Rice and Lorr (1968) have shown that the complete linkage procedure, when applied to correlations, yields sectors of equal area around the point of origin. Each sector subtends an angle whose cosine is equal to the inclusion limit. When the method is applied to distances, it generates circular zones around object pairs in centers of high density. Another property of the compact cluster is its convexity. A set of points in N-space is said to be convex if for every pair of

points A and B in the set, the line segment joining these points are also in the cluster. This means, of course, that a compact cluster cannot be hollow like a doughnut or shaped like a crescent.

McQuitty's rank order typal analysis calls for complete consistency within a cluster. Cluster members must not include a rank higher than the number of objects in the cluster. In graph theory, cliques correspond to such mutually consistent sets in which all relations are symmetric and transitive. In the previous link definition each member has a pair correlation with all other members at or above the arbitrary minimum.

Average Linkage Analysis

Because the requirements of complete linkage are stringent, the subsets found tend to be small. A less stringent and more realistic group of methods called average linkage take into account variation in similarity indices due to error. Admission of any object to a cluster is based on an arbitrary average correlation of an object with a cluster. The type is then defined as a subset of objects in which each member is, on the average, more like every other object than it is (on the average) like any other objects outside the cluster.

Sokal and Michener (1958) first outlined the procedure and called it the unweightened pair-group method. Both complete linkage and average linkage algorithms may be found in Rohlf's et al. (1971) NT-SYS program. It computes both average similarity or dissimilarity of a candidate object with an existing cluster. Sneath and Sokal (1973) give an extended discussion and set of illustrations for these methods.

The method of average linkage has been modified to exclude outliers and objects lying in the zone between two clusters. The procedure developed by Lorr and Radharkrishnar(1967) includes both an inclusion and an exlusion parameter. After a nucleus of three objects is established, objects that satisfy the inclusion criterion are added one at a time on the basis of their average correlation with cluster members. Next, all objects that correlate, on the average, above the exclusion criterion with members of the first subset are eliminated from the matrix. Then from among the remaining unclustered objects, another nucleus of three is sought whose similarity relationships are at least equal to the inclusion limit and the process is repeated. When no other clusters can be found the process stops.

METHODS OF ORDINATION

The term ordination, which originated in biology, has been used to refer to the process of obtaining a low dimensional mapping of a set of data points. One procedure is a principal component analysis of an N by N matrix of similarity indices. The space in which the N points are imbedded then can be examined visually to discover any groupings. A common mapping technique is to plot the data in the space of pairs of principal components.

Another procedure is to apply a multidimensional scaling technique (MDS) such as developed by Shepard-Kruskal (Shepard et al., 1972) to the similarity matrix. The general logic of the procedure is, when given some measure of similarity for every two objects, a configuration of N points is sought in a space of the smallest possible dimensions. The requirement is that to an acceptable degree of approximation, the resulting interpoint distances are monotonically related to the proximity data. However, MDS is really only useful for small sets of data (less than 60 objects) and is more useful in representing intercluster data. Further it should be clear that both principal component analyses and MDS leave the investigator with an essentially spatial representation. The assignment of points to clusters as such must be done after the analysis by eye or by some other objective method of identifying clusters as such.

A third helpful procedure for obtaining a graphical representation of the clusters found is to run a discriminant function analysis. The groups can then be plotted as points in canonical variate space.

CRITERIA FOR "NATURAL GROUPS"

One goal of typing is the identification of discrete homogeneous groups in a nonrandom population. When do the data indicate that some types exist? There are several important clues worth following. Generally, a frequency distribution of scores on a continuous variable will have a single mode or value with the highest frequency. If there are more than one, it is likely that the sample represents a mixture of several distinct types. Also, a scatterplot of cases in two or three dimensions may reveal a clustering of cases into two or more swarms. Again, such multidimensional multimodes are indicative of the existence of disjoint groups.

A second indication for the existence of types is strongly skewed or markedly asymmetric distributions. Johnson and Wall (1969) suggest a method of detecting polarization or skewedness. The distance matrix of between-subject distances is divided by the square root of the sum of all squared distances. This transformation yields squared entries that sum to unity. One may construct distributions of distances from each entity i to all other entities. An entity imbedded in a cluster will characteristically exhibit an early peak or modal hump along the abscissa near the origin. The key entities which are the centers of dense regions are used as nuclei for clusters.

Illustrations of high skewedness are found in score distributions of psychotic syndromes (Lorr, 1966). Plots of nearly every syndrome frequency distribution indicate asymmetry. Multidimensional scatterdiagrams also indicate multiple modes. Cluster analysis has revealed some seven psychotic types that have been replicated cross-nationally confirming the evidence from skewedness, asymmetry, and multimodality.

CAUTIONS: GENERAL DESIDERATA FOR TYPOLOGY

A number of *desiderata* can be offered for a classification scheme that results from a cluster analysis.

Objectivity. Independent researchers should reach similar conclusions. If the scaling process is explicit and a well-defined algorithm is applied, the results will be objective in the sense referred to.

Stability. The classification should be affected little by new data. If a scheme remains unaltered by the addition of other variables, the new variables are probably correlated highly with those already included. This requirement suggests the need for the widest possible input of relevant descriptors. All major sources of variation in a domain of similarity should be represented. If the clusters continue to change with added measures, then it is possible that the domain itself is poorly defined or that several domains are actually represented. Such an impasse calls for more extensive dimensional analysis of the variates represented and a consideration of the demands being made on the data. Suppose the problem is to identify distinctive schizophrenic subgroups but relevant symptom patterns of certain types of patients have been left out. The addition of such symptoms and signs can then lead to the identification of new groups and to changes in the patterns of groups already defined.

Replicability. The groupings generated should be replicable under changes in the samples of persons, stimuli, or objects studied. Any type identified should emerge when an adequate number are included in another sample. If, for example, an extrovert type were identified, then a subgroup of extroverts should be identifiable within a new sample.

Parsimony. Each type identified within a hierarchy or typological scheme should be definable in terms of relatively few of the classificatory variates. The logic in support of this principle is that surely not all descriptor variables are needed to define each type. Rather, the expectation is that most are sufficiently defined in terms of a relatively small number of descriptors. A useful basis for judgment are the mean standard score profiles of the subgroups and associated dispersion of the profile elements. Members of a cluster should agree closely on a few of the descriptors but vary substantially on nondefining variates. Consider, for example, the 10 syndromes that define psychotic behavior. An anxious depressive subgroup should be defined mainly by Anxious Depression and possibly by the Obsessive-Phobic syndrome. All other eight syndromes should be irrelevant descriptors. Indeed, this is what is found.

ILLUSTRATIVE APPLICATIONS

NONDRUG RESEARCH

One test of the effectiveness of a typological algorithm is its ability to recover groupings of known physical objects. The data used consist of 33 ships of the U.S. Navy, each measured with respect to length, displacement, beam, number of light, of medium, of heavy, and of very heavy guns, numbers of personnel, maximum speed, and submersibility. This set of data was taken by Cattell and Coulter (1966) from Jane's Fighting Ships. Johnson's (1967) hierarchical clustering scheme (maximum method) was applied to the data by the author.

The ship code numbers were as follows:

Light cruisers	01 to 05	Submarines	19 to 23
Heavy cruisers	06 to 08	Destroyers	24 to 28
Battleships	09 to 13	Frigates	29 to 33
Aircraft carriers	14 to 18		

Table 2 shows the merging of ships into clusters beginning at the far right where destroyers 25 and 28 join and then 26, 24, and 33 (a frigate). Next, to the left of the destroyers we see the frigates 29 to 32. To the left of the frigates we find the submarines 19 to 23. The battleships 10 to 13 merge next. The aircraft carrier group 14 to 18 may be found at the far left. The light and heavy cruisers are not as well differentiated. The column of values to the left of the figure gives the similarity values associated with each clustering in the hierarchical representation.

Table 2. The HCS Obtained on Ships' Data by the Maximum Method.

Ship Code Numbers

	1 5	1 6	1 7	1 4	1 8	0 6	0 7	0 3	0 2	0 8	0 1	0 9	1 2	1 0	1 1	1 3	2 0	1 9	2 2	1 3	2 0 4	0 5	3 2 0	3 9	2 3 1	3 2 4	2 2 7	2 6	2 5	3 8	3 3
.018	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X	X	X	.
.059	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X	X	X	.	.	.	X	X	X	.
.063	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X	X	X	.	.	.	X	X	X	.
.071	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X	X	X	.	.	X	X	X	.	.	.	X	X	X	.
.10	.	.	.	.	.	.	.	.	X	X	X	.	.	.	.	.	.	.	X	X	X	.	.	X	X	X	.	.	.	X	X	X	.
.10	.	.	.	.	.	.	.	.	X	X	X	.	.	.	.	.	.	.	X	X	X	.	X	X	X	X	.	.	.	X	X	X	.
.11	.	.	.	.	.	.	.	.	X	X	X	.	.	.	.	.	.	X	X	X	X	.	X	X	X	X	.	.	.	X	X	X	.
.13	.	.	.	.	.	.	.	.	X	X	X	.	.	.	.	.	.	X	X	X	X	.	X	X	X	X	.	.	X	X	X	X	.
.14	.	.	.	.	.	.	.	.	X	X	X	X	X	.	.	.	.	X	X	X	X	.	X	X	X	X	.	.	X	X	X	X	.
.15	.	.	.	.	.	.	.	.	X	X	X	X	X	.	.	.	.	X	X	X	X	X	X	X	X	X	.	.	X	X	X	X	.
.19	.	.	.	.	.	X	X	X	.	X	X	X	X	.	.	.	.	X	X	X	X	X	X	X	X	X	.	.	X	X	X	X	.
.26	.	.	.	.	.	X	X	X	.	X	X	X	X	.	.	.	.	X	X	X	X	X	X	X	X	X	X	.	X	X	X	X	.
.28	.	.	.	.	.	X	X	X	.	X	X	X	X	.	.	.	.	X	X	X	X	X	X	X	X	X	X	.	X	X	X	X	X
.29	.	.	.	.	.	X	X	X	.	X	X	X	X	.	.	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
.31	.	.	.	.	.	X	X	X	.	X	X	X	X	.	.	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
.42	.	.	.	.	.	X	X	X	.	X	X	X	X	.	.	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
.42	.	.	.	.	.	X	X	X	.	X	X	X	X	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
.47	.	.	.	.	.	X	X	X	.	X	X	X	X	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
.50	.	.	.	.	.	X	X	X	.	X	X	X	X	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
.56	.	X	X	X	.	.	X	X	X	.	X	X	X	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
.61	.	X	X	X	.	.	X	X	X	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
.65	.	X	X	X	X	X	X	X	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
.85	.	X	X	X	X	X	X	X	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
.95	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.1	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.2	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.3	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.4	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
1.8	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2.0	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2.7	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

DRUG RESEARCH

In order to delineate homogeneous subgroups among hospitalized opiate addicts, Berzins et al. (1974) applied a correlational cluster technique (Lorr and Radhakrishnar, 1967) to the MMPI profiles of 1,500 addicts. The total sample was subdivided into 10 subsamples (5 for each sex) of 150 cases representing four different types of admission to treatment. The 13 K-corrected profile elements were standardized and correlated. Application of the clustering techniques yielded two very similar homogeneous profile types in each subsample. The within-cluster homogeneity coefficients ranged from .61 to .74 indicating substantial similarity of type members. The

between-cluster correlations ranged from -.06 to .18 indicating little type overlap in profile. The mean score profiles were virtually identical across the five male samples, and across the five female samples.

Type I, which constituted 33% of all addicts, was characterized by elevated scores on Scales 2 (Depression), 4 (Psychopathic deviate), and 8 (Schizophrenic). These imply marked subjective distress, nonconformity, and disturbed thinking. Type II (about 7% of the addicts) showed a single peak on Scale 4. A discriminant function analysis of the two types on the basis of the 13 MMPI scales yielded one dimension of difference. The dimension appeared to refer to the general maladjustment of Type I subjects as inadequate, interpersonally alienated, confused, and hypersensitive. Type II subjects, in contrast, were characterized as adequate, poised, untroubled, outgoing, and optimistic.

In order to validate the type differences, the 14 scores of the Lexington Personality Inventory (LPI), also available, were used to compare the two types and the unclustered group. All univariate tests were significant indicating strong support for the type differentiations. Another feature of interest is that the two mean profile types bear striking resemblance to replicated types delineated for male alcoholics (Goldstein and Linden, 1969). Table 3 presents the mean raw scores of type members in the MMPI.

Table 3. Group Means on the Minnesota Multiphasic Personality Inventory

	Male			Female		
Measure set	Type I (n = 216)	Type II (n = 67)	Unclustered (n = 467)	Type I (n = 284)	Type II (n = 41)	Unclustered (n = 425)
No MMPI scales						
--F	9.77	3.81	6.36	11.27	3.88	6.57
--K	12.12	20.24	15.46	9.96	17.78	13.78
1 Hs	19.28	13.78	15.50	20.50	13.61	17.58
2 D	30.43	21.16	23.96	29.55	21.73	25.95
3 Hy	27.15	23.43	23.23	27.13	22.66	25.25
4 Pd	30.40	30.34	29.12	30.16	29.56	28.91
5 MF	26.26	24.28	25.68	34.64	36.37	35.48
6 Pa	14.06	10.66	11.30	14.57	10.59	11.84
7 Pt	35.10	26.40	28.68	35.87	26.98	30.23
8 Sc	35.82	27.27	29.70	38.40	26.66	29.96
9 Ma	22.90	20.51	23.00	24.75	22.29	22.15
10 Si	34.79	19.67	26.18	36.05	21.85	30.35
Es	37.75	50.88	45.60	33.61	48.68	40.32

RESOURCES AND REFERENCES

CLUSTER ANALYSIS PROGRAMS

Anderberg, M.R. *Cluster Analysis for Application.* New York: Academic Press, 1973.

Bailey, D. and Tryon, B.C. *BC-TRY Program.* Portland, Oregon: Tryon-Bailey Associates, Inc., c/o Peter Lenz, 222 S.E. Nehalem Street.

Hartigan, J.A. *Clustering Algorithms.* New York: Wiley & Sons, 1975.

McRae, D.J. MICKA, a Fortran IV iterative K-means cluster analysis program. *Behavioral Science,* 1971, pp. 423-424.

Rohlf, F.J.; Kishpangh, J.; and Kirk, D. *NT-SYS* (Program Package). New York: State University of New York at Stony Brook. Department of Ecology and Evolution.

Rubin, J., and Friedman, H. *Clus.* IBM SHARE system.

Veldman, D.J. *Fortran Programming for the Behavioral Sciences.* New York: Holt, Rinehart and Winston, 1967.

Wishart, D. *Clustan.* London: University College Computer Centre, 19 Gordon Street, London, Great Britain.

Wolfe, J.H. "NORMIX and NORMAP. FOCUS." Palo Alto: Control Data Corporation.

GENERAL REFERENCES

Anderberg, M.R. *Cluster Analyses for Applications.* New York: Academic Press, 1973.

Bailey, K.D. Cluster analysis. In: Heise, D.R., ed. *Sociological Methods.* San Francisco: Jossey-Bass, 1975. pp. 59-128.

Cole, A.J., ed. *Numerical Taxonomy.* New York: Academic Press, 1969.

Everitt, B. *Cluster Analysis.* London: Heinemann Educational Books, 1974.

Hartigan, J.A. *Clustering Algorithms.* New York: Wiley, 1975.

Jardine, N., and Sibson, R. *Mathematical Taxonomy.* New York: Wiley, 1971.

Sneath, P.H., and Sokal, R.R. *Numerical Taxonomy.* San Francisco: W.H. Freeman & Co., 1973.

JOURNAL ARTICLES ON METHOD

Ball, G.H. Data analysis in the social sciences: What about the details? *Proceedings of the Fall Joint Computer Conference,* 27(1):533-559, 1965.

Ball, G.H., and Hall, D.J. A clustering technique for summarizing multivariate data. *Behavioral Sciences,* 12:153-155, 1967.

Bonner, R.E. On some clustering techniques. *IBM Journal of Research and Development,* 8:22-33, 1964.

Cattell, R.B., and Coulter, M.A. Principles of behavioral taxonomy and the mathematical basis of the taxonome computer program. *British Journal of Mathematical and Statistical Psychology,* 19:237-269, 1966.

Cormack, R.M. A review of classification. *Journal of the Royal Statistical Society, Series A,* 134:321-353, 1971.

Fisher, W.D. On grouping for maximum homogeneity. *Journal of the American Statistical Association,* 53:789-798, 1958.

Fleiss, J.L., and Rubin, J. On the methods and theory of clustering. *Multivariate Behavioral Research,* 4:235-250, 1969.

Friedman, H.P., and Rubin, J. On some invariant criteria for grouping data. *Journal of the American Statistical Association,* 62:1159-1178, 1967.

Gower, J.C. A comparison of some methods of cluster analysis. *Biometrics,* 23:623-637, 1967.

_____. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika,* 53:325-328, 1966.

Gower, J.C., and Ross, G.J.S. Minimum spanning trees and single linkage cluster analysis. *Applied Statistics,* 18:54-64, 1969.

Hartigan, J.A. Representation of similarity matrices by trees. *Journal of American Statistical Association,* 62:1140-1158, 1967.

Jardine, C.J.; Jardin, N.; and Sibson, R. The structure and construction of taxonomic hierarchies. *Mathematical Bioscience,* 1:173-179, 1967.

Johnson, R.L., and Wall, D.D. Cluster analysis of semantic differential data. *Educational and Psychological Measurement,* 29:769-780, 1969.

Johnson, S.C. Hierarchical clustering schemes. *Psychometrika,* 32:241-254, 1967.

King, B.F. Step-wise clustering techniques. *Journal of American Statistical Association.* 62:86-101, 1967.

Kruskal, J.B. On the shortest spanning subtree of a graph and a traveling salesman problem. *Proceedings, American Mathematical Society,* 7:48-50, 1956.

Lance, G.N., and Williams, W.T. A general theory of classificatory sorting strategies. *Computer Journal,* 9:373-380, 1967.

Lorr, M., and Radhakrishnan, B.K. A comparison of two methods of cluster analysis. *Educational and Psychological Measurement,* 27:47-53, 1967.

McQuitty, L.L. Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. *Educational and Psychological Measurement,* 17:207-229, 1957.

_____. Typal analysis. *Educational and Psychological Measurement,* 17:207-229, 1957.

Rice, C.E., and Lorr, M. "An Empirical Comparison of Typological Methods." Contract No0014-67-A-0214 Task 005. Office of Naval Research, August 1969.

Rubin, J. Optimal classification into groups: An approach for solving the taxonomy problem. *Journal of Theoretical Biology,* 15:103-144, 1967.

Sawrey, W.L.; Keller, L.; and Conger, J.J. An objective method of grouping profiles by distance functions and its relation to factor analysis. *Educational and Psychological Measurement,* 20: 651-674, 1960.

Shepard, R.N.; Romney, A.K.; and Nerlove, S.B. *Multidimensional Scaling. V-I Theory.* New York: Seminar Press, 1972.

Ward, J.H., Jr., and Hook, M.E. Application of a hierarchical grouping procedure to a problem of grouping profiles. *Educational and Psychological Measurement,* 23:69-81, 1963.

Wishart, D. A generalization of nearest neighbors which reduces chaining effects. In: Cole, A.J., ed. *Numerical Taxonomy.* London: Academic Press, 1969.

Wolfe, J.H. Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research,* 5:329-350, 1970.

Wolfe, J.H. Comparative cluster analysis of patterns of vocational interest. *Technological Bulletin STB 72-3.* San Diego: Naval Pers. and Training Res. Lab., 1971.

SIMILARITY INDICES

Cattell, R.B. r_p and other coefficients of similarity. *Psychometrika*, 14:279-298, 1949.

Cohen, J. r_c: A profile similarity coefficient invariant over variable reflection. *Psychological Bulletin*, 71:281-284, 1969.

Cronbach, L.J., and Gleser, G.C. Assessing similarity between profiles. *Psychological Bulletin*, 50:456-473, 1953.

Goodman, L.A., and Kruskal, W.H. Measures of association for cross-classifications. *Journal of the American Statistical Association*, 49:732-764, 1954.

Gower, J.C. A general coefficient of similarity and some of its properties. *Biometrics*, 27:857-871, 1971.

Helmstadter, G.C. An empirical comparison of methods for estimating profile similarity. *Educational and Psychological Measurement*, 17:71-82, 1957.

Holley, J.W., and Guilford, J.P. A note on the G index of agreement. *Educational and Psychological Measurement*. 24:749-753. 1964.

Mosteller, F. Association and estimation in contingency tables. *Journal of the American Statistical Association*, 63:1-28, 1968.

Overall, J.E. Note on Multivariate methods for profile analysis. *Psychological Bulletin*, 61: 195-198, 1964.

TOPICAL REFERENCES

Miscellaneous

Boyce, A.J. Mapping diversity. In: Cole, A.J., ed. *Numerical Taxonomy*. London: Academic, 1969.

Christal, R.E., and Ward, J.H., Jr. *Use of an Objective Function in Clustering People and Things*. Lackland Air Force Base, Texas: Personnel Research Lab, Air Force Systems Command, 1966.

Frank, R.E., and Green, P.E. Numerical Taxonomy in marketing analysis. *Journal of Market Research*, 5:83-95, 1968.

Goronzy, F. A numerical taxonomy on business enterprises. In: Cole, A.J., ed. *Numerical Taxonomy*. London: Academic Press, 1969.

Grumm, J.G. The systematic analyses of blocs in the study of legislature behavior. *Western Political Quarterly*, 18:350-362, 1965.

Hudson, F.R. Cluster analyses and archaeology. Some new developments and applications. *World Archaeology*, 8:299-320, 1970.

Miller, G.A. A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology*, 6:168-191, 1970.

Medicine

Baron, D.M., and Frasar, P.M. Medical applications of taxonomic methods. *British Medical Bulletin*, 24:236-240, 1968.

Bouckaert, A. Computer diagnosis of goiters-I. Classification and differential diagnosis. *Journal of Chronic Disease*, 24:299-310, 1971.

Goldwyn, R.; Freedman, H.P.; and Siegel, J.H. Iterative and interaction in computer data bank analysis--Case study in physiological classification and assessment of the critically ill. *Comparative Biomedical Research*, 4:607-622, 1971.

Manning, R.T., and Watson, L. Signs, symptoms and systematics. *Journal of the American Medical Association*, 11:1180-1188, 1966.

Winkel, P., and Tygstrup, N. Classification of cirrhosis. The resolution of data modes and their recovery in an independent material. *Comparative Biomedical Research,* 4:417-426, 1971.

Psychiatry and Drug Abuse

Berzins, J.I.,; Ross, W.F.; English, G.W.; and Haley, J.V. Subgroups among opiate addicts: a typological investigation. *Journal of Abnormal Psychologgy,* 83:65-73, 1974.

Everitt, B.S.; Gourlay, A.J.; and Kendall, R.E. Attempt at validation of traditional psychiatric syndromes by cluster analysis. *British Journal of Psychiatry,* 119:399-412, 1971.

Goldstein, S.G., and Linden, J.D. A comparison of multivariate grouping techniques commonly used with profile data. *Multivariate Behavioral Research,* 4:104-114, 1969.

Lorr, M., ed. *Explorations in Typing Psychotics.* Oxford: Pergamon Press, 1966.

Lorr, M., and Gilberstadt, H. A comparison of two typologies for psychotics. *Journal of Nervous and Mental Diseases,* 155:144-148, 1972.

Lorr, M.; Pokorny, A.D.; and Klett, C.J. Three depressive types. *Journal of Clinical Psychology.* 29:290-294, 1973.

Nerviano, V.J. Common personality patterns among alcoholic males: a multivariate study. *Journal of Consulting and Clinical Psychology,* 44:104-110, 1976.

Rice, C.E., and Mattsson, N.B. Types based on repeated measurement. In M. Lorr, ed. *Explorations in Typing Psychotics.* Oxford: Pergamon Press, 1966.

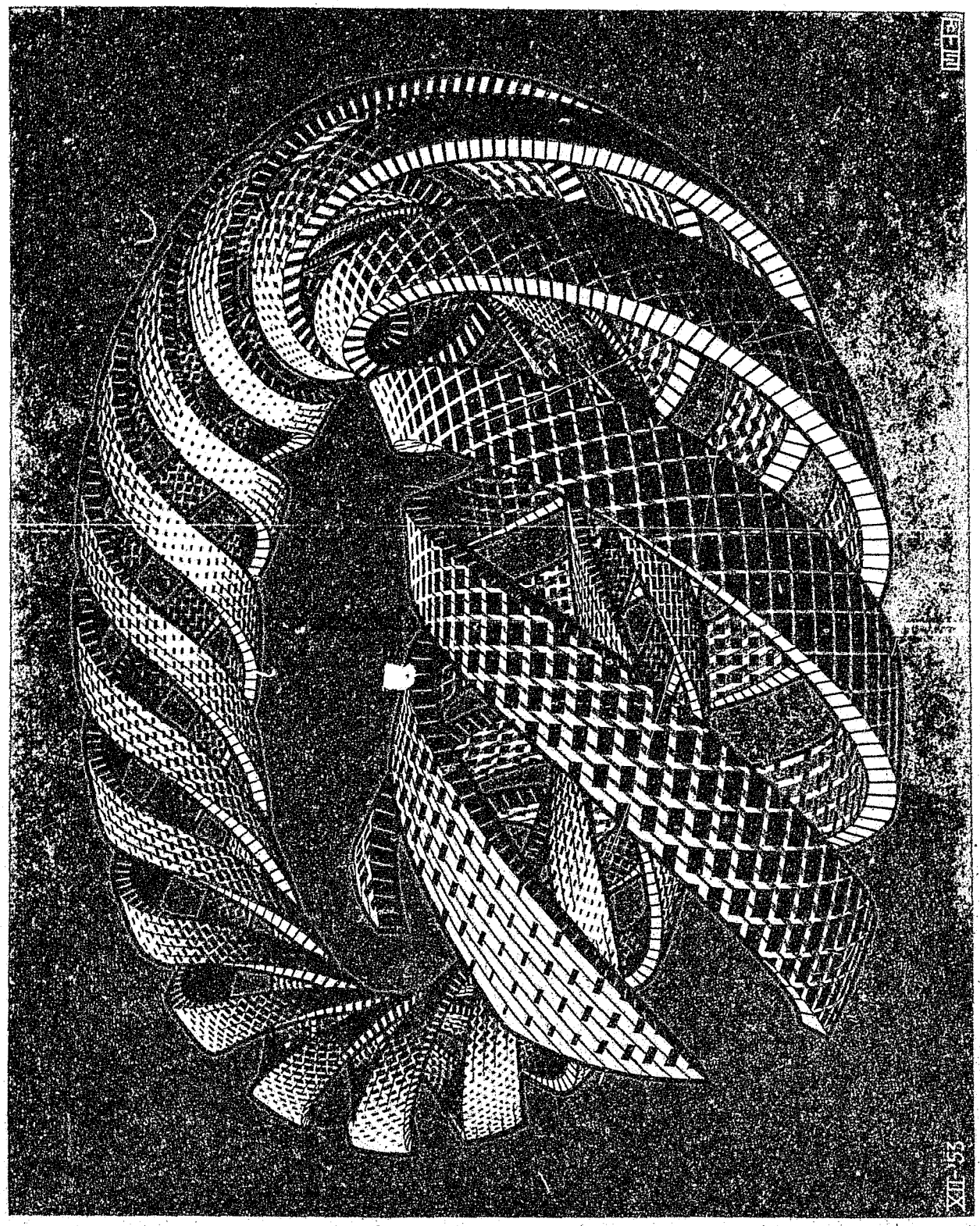


Plate 8

INTRODUCTION

A variety of techniques have been developed in order to make causal inferences in nonexperimental research (e.g., see Blalock, 1964). Path analysis has been used to exposit, empirically test, and develop theory in genetics (Wright, 1921, 1934, 1960; Li, 1975), economics (Wold, 1954), and sociology (Duncan, 1966; Land, 1969; Heise, 1969), and may be useful to researchers concerned with the formulation and verification of theories of drug use or abuse. This paper purports to add nothing original to the subject of path analysis, but rather it is an effort by one researcher concerned with drug research to communicate to other researchers what appear to be some of the major advantages, problems, and limitations of the method.[1] Readers interested in further consideration of the method are referred to more technical descriptions (Duncan, 1966; Land, 1969).

RATIONALE

Advantages

To understand a complex social behavior such as consequences associated with drug use and abuse, multivariate, multistage, interdisciplinary theories will have to be developed. There has traditionally been a disjuncture between verbal theories that have been created to explain social phenomenon, and the methods to test these theories using quantitative empirical research in a manner that could lead to theory growth through rejection and reformulation of hypotheses.

Path analysis is a mathematical modeling technique that can be used to specify relations among a set of variables. In cases where the assumptions underlying the method can be met, path analysis offers a rather elegant way to express a verbal theory in a diagram of causal paths. The development of causal diagram makes implicit assumptions explicit, and facilitates theory development. A set of equations isomorphic to the diagrammic path network can be used to estimate the magnitude of parameters in the model. Often this enables the researcher to reject aspects of the model which can then be reformulated in the light of empirical findings and perhaps tested. Although we can seldom be certain we have the right model, often we can be nearly certain we have the wrong one.

By using a series of equations rather than a single equation, the researcher can estimate what portion of the observed association between an exploratory variable and a dependent variable is attributed to direct causal effect, and what portion attributed to indirect effects through intervening variables. As in other mathematical structure models, in specifying the set of relations among a set of variables, one then observes how a change in one variable affects the other variables in the system. Used in conjunction with longitudinal data, such a model would facilitate analysis of the effects of possible intervention strategies or programs.

In those situations where path analysis can be appropriately applied, it offers a way to develop complex multivariate interdisciplinary theory that can be subject to rigorous empirical tests, and could lead to more comparable research findings, and perhaps to more cumulative scientific knowledge in this area. As Duncan (1975) has put it, such models have responded to a need for formalism that could help in maintaining order and coherence in increasingly complicated times of investigation and theorizing. This observation is particularly applicable in the study of complicated behavior patterns underlying drug use and abuse.

Assumptions

Use of path analysis with ordinary least squares estimation assumes that a set of variables can be temporally ordered, is asymmetrically related, is measurable on an interval scale, and that relations among the variables are linear and additive.[2] An additional assumption underlying the method is that the causal model be correctly specified. As Duncan notes:

> Each 'dependent' variable must be regarded explicitly as completely determined by some combination of variables in the system. In problems where complete determination by measured variables does not hold, a residual variable uncorrelated with other determining variables must be introduced.[3]

Appropriate use of the method is dependent on an accurate understanding of the ways in which the degree of satisfaction of various assumptions affects interpretations of results. The importance of satisfying various assumptions underlying use of the method will be discussed after the method is described in more detail.

METHODS AND PROCEDURES

The Model

Duncan (1966) suggested a series of simple notations useful in drawing path diagrams. The advantage of drawing diagrams in accordance with these explicit rules is that the system of equations will be isomorphic to the path diagrams.

> In path diagrams, we use one-way arrows leading from each determining variable to each variable dependent on it. Unanalyzed correlations between variables not dependent upon others in the system are shown by two-headed arrows, rather than straight, to call attention to their distinction from paths relating dependent to determining variables. The quantities entered on the diagram are symbolic or numerical values of path coefficients, or, in the case of the bidirectional correlations, the simple correlation coefficients.[4]

The basic model of the process of social stratification presented by Blau and Duncan (1967) provides a relatively simple, straightforward example with which to illustrate the method. Five variables were used in their model:

X_1: Father's educational attainment

X_2: Father's occupational status

X_3: Respondent's educational attainment

X_4: Status of respondent's first job

X_5: Status of respondent's occupation in 1962.

Having determined which explanatory variables these authors believe are more important in determining occupational prestige, they then argue for a temporal theoretically appropriate ordering of their variables as they causally relate to occupational prestige. Blau and Duncan argue that although father's education (X_1) and his occupational status (X_2) may not necessarily be ordered, these two variables precede son's educational attainment (X_3), which precedes the status of the son's first job (X_4), and the occupational status of the son's job at the time of the study (X_5). They also argue that the causal relations between these variables are sequential or assymmetrical, that is, the causal direction is one way.

The Path Diagram

Blau and Duncan exhibit the relations among their variables in a path diagram whose numerical path coefficients are estimated from the statistical data. For purposes of illustration, it may be easier to understand the rationale underlying the method if we first construct a path diagram with symbolic path coefficients. This diagram could be considered a statement of the author's hypotheses. The symbolic path coefficients correspond to terms in the estimation equations which constitute the model.

Blau and Duncan's hypothetical model is illustrated in Figure 1. Each straight line represents a causal assertion, and the path coefficient associated with that line estimates the magnitude of the effect that an explanatory variable has on the variable it is pointing to, independently of those other explanatory variables which are also represented by arrows pointing to the same dependent variable. Double headed curved arrows imply no assertions about causation in the relation between two variables. The magnitude of such associations is given by the simple

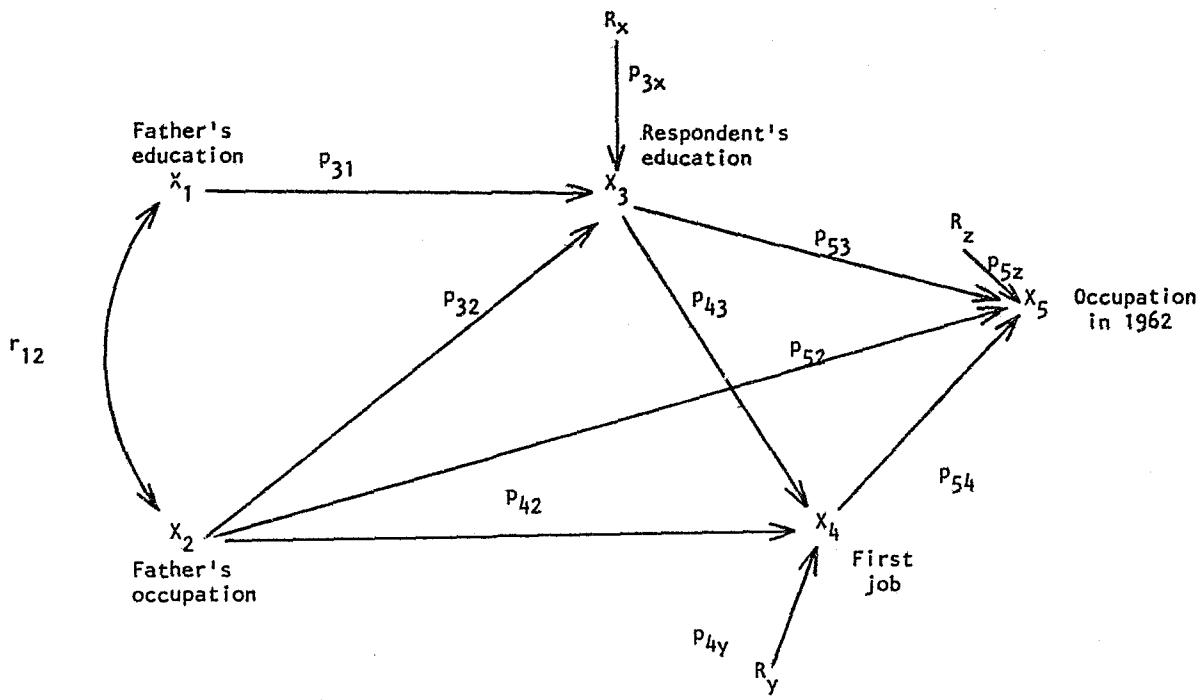


Fig. 1. The hypothetical path diagram of Blau and Duncan's (1967) basic model of the process of stratification.

---

correlation coefficients; R_x, R_y, and R_z are the residual variables. The method of estimating the path coefficients can be understood more easily after the equations describing the path model are further discussed.

Estimating Path Coefficients[5]

Each variable in the path diagram affected by other variables is called an endogenous variable. For each endogenous variable there is one equation. Since there are three endogenous variables, (X_3, education; X_4, first job; and X_5, occupation) there are three equations. Each straight arrow corresponds to a term in the equation whose dependent variable is at the point of the arrows.

Three equations describe this model:

$$X_5 = p_{52} X_2 + p_{53} X_3 + p_{54} X_4 + p_{5z} R_z \quad (1)$$

$$X_4 = p_{42} X_2 + p_{43} X_3 + p_{4y} R_y \quad (2)$$

$$X_3 = p_{31} X_1 + p_{32} X_2 + p_{3x} R_x \quad (3)$$

Each term in each equation corresponds to one straight arrow in the path diagram.

The path coefficient of the residual variable, in equation (1), p_{5z} is an estimate of the magnitude of the effects of error, chance and variables not included in the analysis. The multiple correlation coefficient squared, R^2, represents the amount of variance accounted for in equation (1). Consequently the residual variable, which must account for the remaining variance, has a path coefficient $p = \sqrt{1 - R^2}$.

Significance tests on the estimated parameters can be used to test hypotheses in the model, as long as the hypotheses are not dependent on extensive search of the data. For example, Blau and Duncan's model hypothesized that father's education would have no direct effect on son's occupation in 1962, when the indirect effects of the other intervening variables in the model were taken into account. This hypothesis was supported by regressing occupation in 1962 on the four explanatory variables in the analysis, and determining that father's education did not make a significant independent contribution to the variance in son's occupation in 1962.

The path diagram reported by Blau and Duncan is shown in Figure 2. Their results indicate that the direct effect of the respondent's education was more important than the status associated with the respondent's first job. The effect of father's occupation was considerably smaller, although father's education indirectly affected son's occupation in 1962 through its direct effect on the respondent's education, and its effects on the status of the respondent's first job.

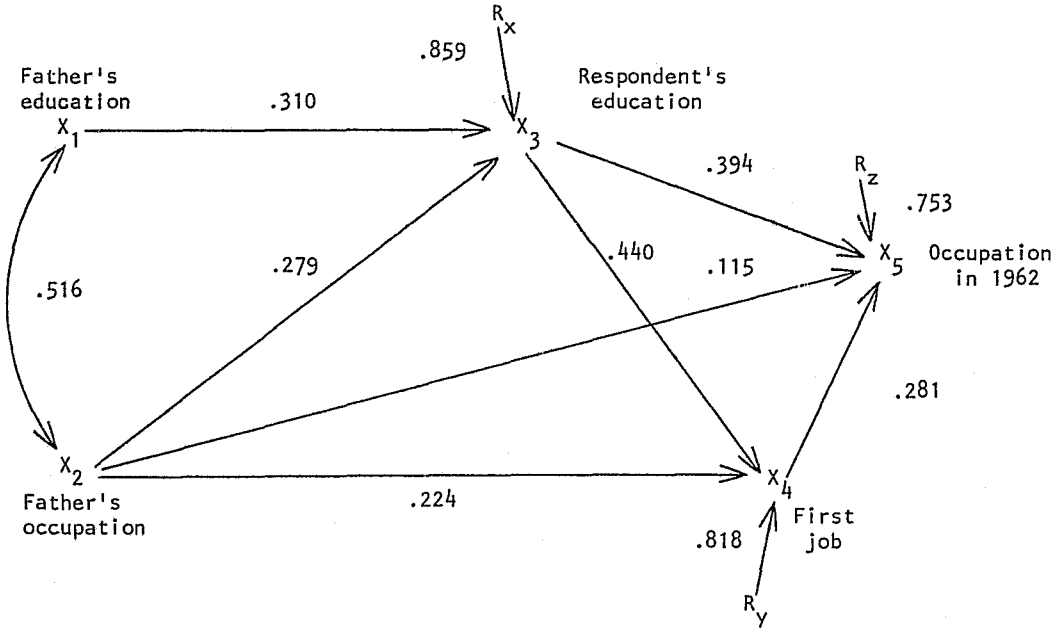


Fig. 2. Path coefficients in Blau and Duncan's (1967) basic model of the process of stratification.

ILLUSTRATIVE APPLICATION IN DRUG RESEARCH

In a study by Naditch (1975), three groups of explanatory variables were included in a model of acute adverse reactions to marihuana and LSD: psychopathology, motives underlying drug use, and drug usage experience. An hypothesized theoretical ordering of these variables was specified. The degree of psychopathology was hypothesized to affect the development of motives for use, which in turn was hypothesized to influence the degree of drug usage. Finally, drug usage was hypothesized to affect the development of an acute adverse reaction. Prior work suggested that three aspects of psychopathology be included: (1) a characteristic tendency to use defensive regression in the face of stress (X_{10}), (2) maladjustment (X_9), and (3) schizophrenic thought processes (X_8). Based on a factor analysis of a variety of motives reported by subjects for their drug use, three motive factors were included in the analysis: (1) use for pleasure (X_5), (2) reluctant use in response to peer pressure (X_6), and (3) use for self-therapy(X_9).

The hypothesized relations among these variables, exhibited in path diagramatic form, are shown in Figure 3. Two separate path models, one for acute adverse reactions to marihuana and one for acute adverse reactions to LSD, were drawn together in this diagram for purposes of comparison. Consequently, no associations between marihuana and LSD usage or between the two kinds of adverse reactions are shown on the diagram. Since this is a rather complex diagram, associations among the three motive variables were not shown on the diagram, although they were discussed in the text.

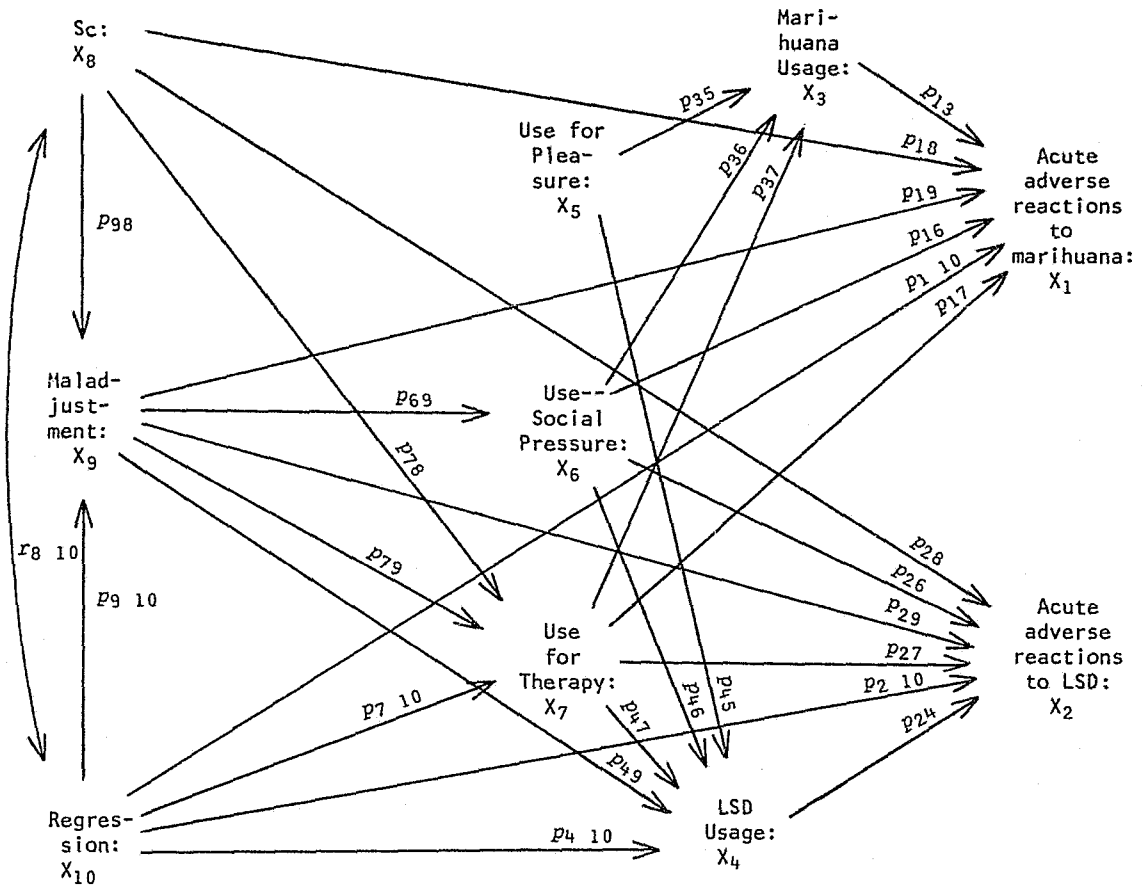


Fig. 3. Hypothesized path model of acute adverse reactions to LSD and marihuana.

The presence or absence of causal arrows was based on theory and previous empirical research in the area. For example, none of the three psychopathology variables were hypothesized to be related to increased use of psychoactive drugs for self-therapeutic motives. None of the three psychopathology variables was hypothesized to be related to marihuana use. Maladjustment and a characteristic tendency to use regression as an ego defense were hypothesized to be causally related to increased usage of LSD. Each of the three psychopathology variables was hypothesized to have direct causal influences on the development of both acute adverse reactions to LSD and marihuana, independently of the other explanatory variables in the analysis, as well as indirectly through the paths shown in the diagram. An explication of the rationale underlying these hypotheses can be found in Naditch (1975). Individual hypotheses taken together in the path diagram were interpreted as statements of both direct independent causal effects and also as indirect effects. A characteristic tendency to regress when faced with stress, for example, was hypothesized to have a direct causal effect on both acute adverse reactions to LSD and to marihuana, independently of the other variables in the analysis, and also hypothesized to indirectly affect the development of acute adverse reactions to LSD through a heightened motive to use the drug for self-therapeutic reasons, and through increased levels of LSD usage. For each dependent variable there should be a corresponding residual term to account for error, chance, and variables not included in the analysis. However, they were not shown in the diagram for purposes of readability and clarity.

The path coefficients in this model were estimated using ordinary least squares on standardized data. There are seven endogenous variables in this model, and consequently, seven equations describe the relations shown in the diagram. (The path coefficients are equivalent to standardized regression coefficients, here represented by B's. Had unstandardized data been used, the estimated parameters would be ordinary regression coefficients.)

$$X_1 = B_{13} X_3 + B_{18} X_8 + B_{19} X_9 + B_{16} X_6 + B_{110} X_{10} + B_{17} X_7 \quad (1)$$

$$X_2 = B_{28} X_8 + B_{26} X_6 + B_{29} X_9 + B_{27} X_7 + B_{210} X_{10} + B_{24} X_4 \quad (2)$$

$$X_3 = B_{35} X_5 + B_{36} X_6 + B_{37} X_7 \quad (3)$$

$$X_4 = B_{45} X_5 + B_{46} X_6 + B_{47} X_7 + B_{49} X_9 + B_{410} X_{10} \quad (4)$$

$$X_6 = B_{69} X_9 \quad (5)$$

$$X_7 = B_{78} X_8 + B_{79} X_9 + B_{710} X_{10} \quad (6)$$

$$X_9 = B_{98} X_8 + B_{910} X_{10} \quad (7)$$

For the purposes of illustration, the results of equation (1) in the model are shown in Table 1. The explanatory variables taken together in the model accounted for 40% of the variance in acute adverse reactions to marihuana. Each beta coefficient in equation (1) corresponds to a path coefficient associated with a causal arrow pointing toward acute adverse reactions to marihuana as shown in the path diagram in Figure 4. As can be seen from the results, a number of hypotheses shown as causal arrows in Figure 3 were rejected, and do not appear in the results diagram. For example, use of drugs as a response to social pressure did not make a significant independent contribution to the variance in either the degree of marihuana usage or LSD usage. From the path diagram, the reader can determine the relative importance of the independent contributions of the explanatory variables in explaining variance in any of the dependent variables, and also determine the extent to which variables prior in the system may indirectly affect dependent variables. For example, although each of the three psychopathology variables, two of the motive variables, and the degree of LSD usage each made significant independent contributions to the variance in acute adverse reactions to LSD, the path coefficients indicate that maladjustment problems and the degree of LSD usage had more direct effects of large magnitude than did other motive and psychopathology variables. Regression, in addition to a rather small direct effect, did indirectly affect acute adverse reactions to LSD through its effects on use for therapy and increased usage. The magnitude of the indirect effects can be estimated by multiplying along the appropriate paths. (For example, the indirect effect of regression through increased motives to use drugs for self-therapy in determining adverse reactions to LSD was .19 x .19 = .026.) A discussion of procedures used to calculate indirect effects can be found in Duncan (1966). It can also be determined from these results that the relative importance of some explanatory variables differed in determining adverse reactions to LSD as compared to adverse reactions to marihuana. For example, use for pleasure was a more important independent motive in determining marihuana use than in determining LSD use. More comprehensive interpretations of these results can be made by interested readers from discussions in Naditch (1975).

Table 1.
Path Equation of Acute Adverse Reactions to Marihuana
on Reasons for Use, Marihuana Usage, and Psychopathology

Independent variables	X_1 acute adverse reactions to marihuana			
	r	Equation (1) B	F	R^2chg
X_3 Marihuana usage	.26***	.10*	5.2	.01
X_5 Use for pleasure	.11*			
X_6 Use because of social pressure	.24**	.11*	7.1	.01
X_7 Use for therapy	.45***	.27***	34.1	.20
X_8 Schizophrenia	.44***	.21***	18.0	.12
X_9 Maladjustment	.43***	.19***	15.0	.04
X_{10} Regression	.40***	.14**	8.7	.01
R^2 [a]				.40

Note. R^2 = multiple correlation coefficient squared; r = zero-order correlation; B = standardized regression (path) coefficient; F = F statistic calculated for standardized regression (path) coefficient.

[a] df = 1,338
*p < .05, two-tailed test
**p < .01, two-tailed test
***p < .001, two-tailed test

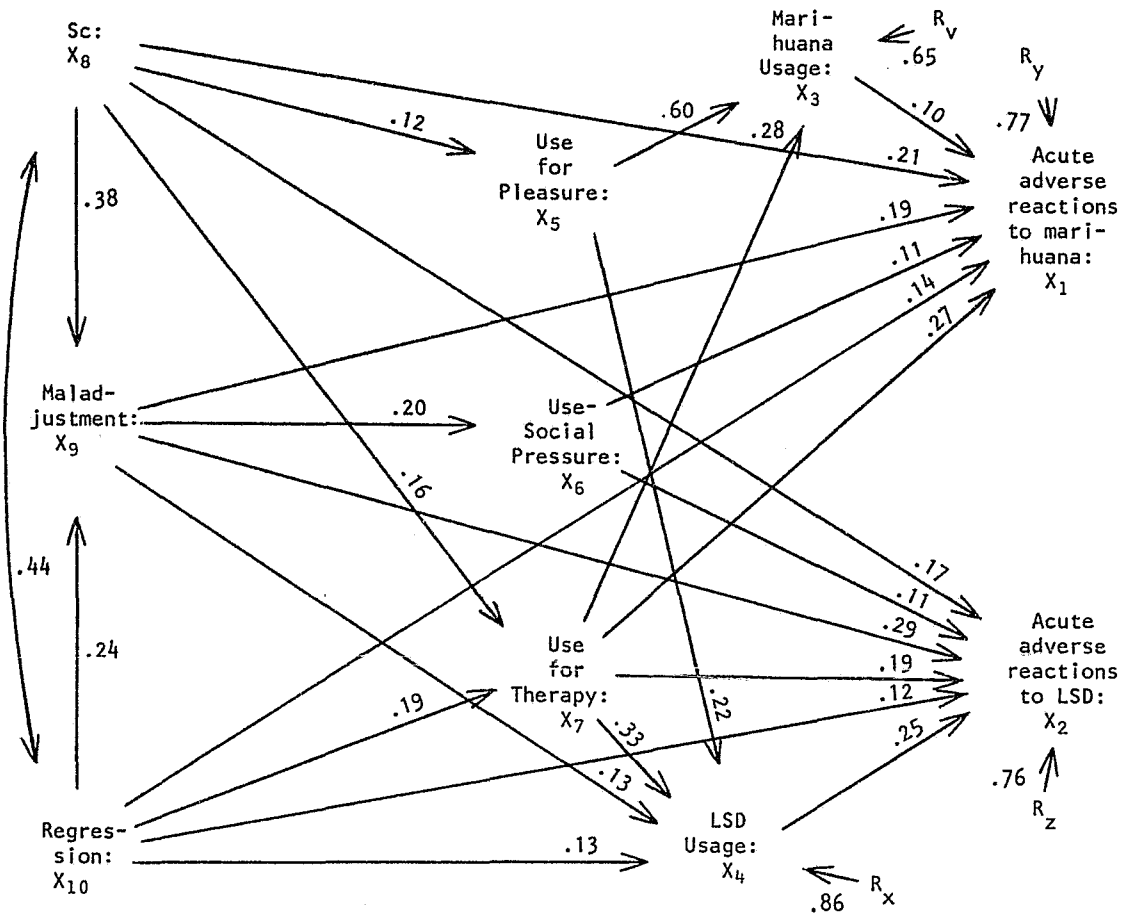


Fig. 4. Path coefficients in the model of acute adverse reactions to LSD and marihuana.

CAUTIONS

The major potential problem with the method is that the user may fail to take sufficient account of the inevitable violations of assumptions underlying the use of the method, and therefore may provide faulty interpretations of the results. The validity of any path model as a description of reality depends both on the quality of the theoretical hypotheses constituting the model and also the representativeness and quality of the data from which the parameters are estimated. An elegant path model describing a set of data may lead to erroneous conclusions if the assumptions underlying the model are not theoretically sound. Some of the assumptions underlying use of the method are more important than others, and an inability to adequately satisfy these assumptions may suggest that the technique is wholly inapplicable. The remainder of this section of the paper will be concerned with a discussion of assumptions underlying use of path analysis.

Specification of the Model

The most important prerequisite is a theoretically defensible specification of a model. Can the author specify what the important explanatory variables are in determining an outcome of interest?

Put somewhat differently, has theory and research in an area developed to the point where enough is known to be able to specify the most important underlying causes? Only insofar as path models rest on creative and substantive theories will they be contributions to scientific understanding. In accounting for variance unaccounted for by the explanatory variables in the model, residual variables represent the effects of variables not included in the analysis as well as measurement error and effects of chance. The implications of leaving relevant variables out of the analysis are not confined to the simple fact that the model will explain less of the variance. More importantly, to the extent variables left out of the analysis correlate with explanatory variables included in the analysis, the assumption that the residual variable not be correlated with any of the explanatory variables will be violated and the parameter estimates will be biased. Introducing an additional explanatory variable which is correlated with other explanatory variables will affect the regression coefficients of explanatory variables already in the equation. Path coefficients therefore will be biased to the degree and extent to which the equations estimated differ from a hypothetical equation that "truly" describes the process being explained.

In actual practice, given the level of sophistication of knowledge in the social sciences, one can rarely say that one understands a phenomenon of interest in sufficient depth that all the explanatory variables can be specified with certainty. Researchers should consider the extent to which their coefficients are biased because of failure to fully satisfy this assumption in interpreting the meaningfulness of path estimates, particularly during the early stages of theory development.

Cross-sectional Data

Use of path analysis assumes that the variables in the analysis are defensibly ordered in a causal sequence, and that causal relations among the variables are asymmetrical or unidirectional. Satisfying these assumptions may be especially difficult with cross-sectional data, in which the basis for unambiguous causal ordering and asymmetry are more tenuous.

Asymmetrical Causal Effects in Drug Research

The assumption of asymmetrical causal effects is often a difficult assumption to satisfy in drug-related research. Many of the variables of interest, e.g., problem drinking behavior, will often be affected by, and in turn causally affect, other variables. If one were studying problem drinking behavior, for example, it would be difficult to argue that an association between self-esteem and problem drinking behavior could be interpreted as singularly the effects of self-esteem in increasing problem drinking behavior to the exclusion of the hypotheses of increased drinking behavior leading to a loss of self-esteem. In cases where there is not sufficient basis to argue asymmetry of causal effects or justify causal ordering of the variables, the use of path analysis may be premature.

Reciprocal Causality

As mentioned, techniques have been developed in econometrics (e.g., Wright, 1960; Johnston, 1963) and used in sociology (Duncan, Haller and Portes, 1971) to estimate parameters given assumptions of reciprocal causality and feedback loops. The relatively simple least squares procedures used to estimate path coefficients in recursive models (one way causality) cannot be used to estimate parameters in nonrecursive models, and consequently more complex estimation procedures must be used.

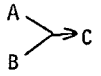
Time-Series Data

Although the asymmetry assumption is a major limitation in use of this method using cross sectional data, this assumption often becomes more tenable with time-series data. For example, the asymmetry and ordering problem in a study concerned with the relation of self-esteem to problem drinking behavior could be overcome by examining the relation of self-esteem in time one to problem drinking in time two, the relation of problem drinking in time one to self-esteem in time two, and relation of problem drinking in time two to self-esteem in time three. (For another example, see Kandel and Faust, 1975.)

Linearity, Additivity, Interval Strength

The assumptions of linearity, additivity and interval strength data are less severe limitations in use of the method. Nonlinear terms and interaction terms can be included in path models (e.g., Darlington, 1968; Cohen, 1968). When interaction effects are used in a path model, the following

notation (representing an interaction effect between variables A and B in determining C) can be used:



Ordinal and even nominal scaled data can be used in a multiple regression model using a procedure employing dummy variables, and these techniques have been employed in path models (e.g., Lyons, 1971).

Theory Building

A key distinction concerning the use of regression-related techniques such as path analysis concerns deductive versus inductive theory building. The statistical tests of significance employed assume that hypotheses have been deductively induced from the theory, as opposed to having been discovered by searching the data for significant associations and then ex post facto developing a theory to explain those relations. In actual practice, it would be a rare and uncurious scientist who would be content to calculate only a single set of parameter estimates from a data set. Most researchers prefer to further explore their data with less fully developed hypotheses and hunches eschewing the robust use of significance tests. One approach to this problem is to split the original data set into random halves, using one half for explorations and the second half for hypotheses tests.

NOTES

[1]The author would like to thank Steven Caldwell for reading the manuscript and offering critical comments.

[2]Reciprocally causal variables, nominal and ordinal variables and nonlinear, nonadditive relationships can be incorporated into causal models whose parameters are estimatable, but such topics are beyond the scope of this paper. See Duncan (1975) for discussion of these topics.

[3]Duncan, 1966, p. 3.

[4]Ibid.

[5]Path coefficients are usually estimated using multiple regression equations. Consequently any of the standard computer programs used for multiple regression may be used. Readers interested in more sophisticated and technically complex general programs may wish to refer to the LISREL program developed by Jöreskog and van Thillo listed in the reference.

REFERENCES

Blalock, H.M. *Causal Inferences in Nonexperimental Research.* Chapel Hill: University of North Carolina Press, 1964.

Blau, P.M., and Duncan, O.D. *The American Occupational Structure.* New York: John Wiley and Sons, 1967.

Cohen, J. Multiple regression as a general data-analytic system. *Psychological Bulletin,* 1968, 70:426-443, 1968.

Darlington, R.B. Multiple regression in psychological research and practice. *Psychological Bulletin,* 69:161-182, 1968.

Duncan, O.D. Path analysis: Sociological examples. *The American Journal of Sociology,* 72:1-16, 1966.

Duncan, O.D.; Haller, A.O.; and Portes, A. Peer influences on aspirations: A reinterpretation. In: Blalock, H.M., ed. *Causal Models in the Social Sciences.* Chicago: Aldine, 1971.

Duncan, O.D. *Introduction to Structural Equation Models.* New York: Academic Press, 1975.

Heise, D.R. Problems in path analysis and causal inference. In: Borgatta, E.F., ed. *Sociological Methodology 1969.* San Francisco: Jossey-Bass, Inc., 1969.

Johnston, J. *Econometric Methods.* New York: McGraw Hill, 1963.

Jöreskog, K.G., and van Thillo, M. LISREL: A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables. *Research Bulletin 72-66.* Princeton, N.J.: Educational Testing Service, Dec. 1972.

Kandel, D., and Faust, R. Sequence and stages in patterns of adolescent drug use. *Archives of General Psychiatry,* 32:923-932, 1975.

Land, K.C. Principles of path analysis. In: Borgatta, E.F., ed. *Sociological Methodology 1969.* San Francisco: Jossey-Bass, Inc., 1969.

Li, C.C. *Path Analysis - A Primer.* Pacific Grove, California: Boxwood Press, 1975.

Lyons, M. Techniques for using ordinal measures in regression analysis and path analysis. In: Costner, H.L., ed. *Sociological Methodology 1971.* San Francisco: Jossey-Bass, Inc., 1971.

Naditch, M.P. The relation of motives for drug use and psychopathology in the development of acute adverse reactions to psychoactive drugs. *Journal of Abnormal Psychology,* 84:374-385, 1975.

Wold, H. Causality and econometrics. *Econometrics,* 22:162-177, 1954.

Wright, S. Correlation and causation. *Journal of Agricultural Research,* 20:557-585, 1921.

Wright, S. The method of path coefficients. *Annals of Mathematical Statistics,* 5:161-215, 1934.

Wright, S. The treatment of reciprocal interaction, with and without lag, in path analysis. *Biometrics,* 16:423-445, 1960.

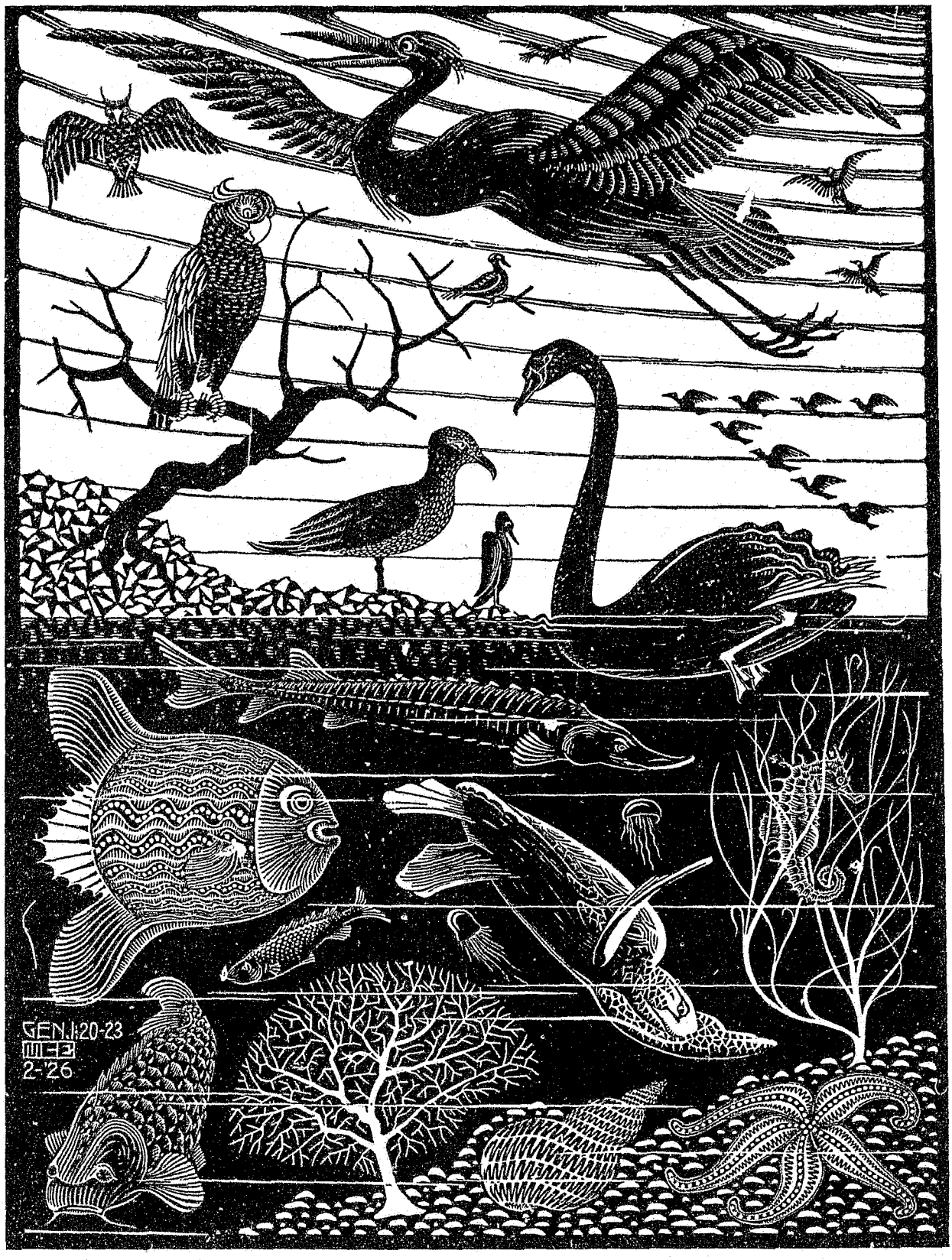


Plate 9

INTRODUCTION

Factor analysis is the most widely used of all methods of multivariate analysis; yet, its applications in the area of drug research have been minimal indeed. The purpose of this article is to provide a nontechnical description of the method, so as to make researchers in the drug and alcohol areas aware of the potential of the technique to their problem areas.

As a multivariate technique, factor analysis is concerned with the understanding of multiple variables measured on many entities. A given entity, such as an individual, has as many scores as there are variables. In any given application, there may be several dozen or several hundred variables, and anywhere from several hundred to several thousand entities. Illustrations of such data could include personality variables measured in some group of subjects, social and economic variables measured in a collection of societies, drug and alcohol attitudinal variables measured in a nationwide sample, or biochemical variables presumably related to drug and alcohol use in a set of subjects under a variety of experimental conditions. Factor analysis is but one of many techniques that might be applied to data of this sort. Its major goal is to analyze and describe sources of variation in the data. In the following section, we focus upon one possible use for factor analysis with data such as these--namely, data reduction. As will be pointed out below, however, there are other major reasons for wishing to undertake a factor analysis.

RATIONALE

USAGE

In experimental situations, statistical and mathematical techniques for analyzing sources of variation among scores are well known in the familiar term of analysis of variance. Analysis of variance is very useful when there exists a single dependent variable and known independent variables, and a generalization of the technique allows one to determine the effects of independent variables on multiple dependent variables. Actually, analysis of variance is really an analysis of means--or, more specifically, variation in means relative to other sources of variance. In the situation we are considering, there is not one but rather a very large set of dependent variables, and furthermore, no specific variables can be considered as independent. All the variables have the equivalent status of being mutually dependent. The concern is not with analyzing the variation among means, since these are typically quite arbitrary in this context, but rather with understanding the variation around the variables' various averages and the interdependence across variables.

What is the significance of analyzing sources of variance in a situation in which all variables take the status of mutual dependence? The answer is easiest to understand in simple, hypothetical situations. Suppose, for example, that all of the variables were perfectly correlated with each other. It is obvious that even though we may be measuring hundreds of variables, they are essentially completely redundant. Indeed, a single variable could summarize all the information in all variables except, as pointed out above, the information about the means of the variables; the means may be quite different. Generally, analysis of mutual dependence seeks to ignore the effects of these variable averages, and is concerned instead with analysis of deviations from means. Thus, since all entities have exactly equivalent standing on all variables, measured as deviations from means, we may as well discard the redundant variables and simply select any one of the variables to represent the entire set. A similar argument could be made when there exists a large set of variables that could be broken down into two subsets, such that in each subset all the variables are perfectly correlated. In this case, we might discard all but one variable for each subset.

Of course, no known data would conform to an ideal situation as described above. As a slightly more realistic situation, we might consider that all variables actually measure the same thing except for the fact that each variable has some random error component. If it were not for the error, the variables would still be perfectly correlated. Again, it seems reasonable to find a single variable that might summarize all the consistent differences among entities. In this case we could not pick any single variable arbitrarily, since different variables might have different reliabilities. Obviously, we would like to select the variable that was most reliable. Alternatively, as is well known from classical test theory and psychometrics, if all the variables are measuring the same thing, we might define a new variable as a composite based on all the somewhat unreliable variables or on some subset of these. This new composite variable will have greater reliability than any one variable selected to represent the entire set. Again, if there were truly two different subsets of variables with this error characteristic, we might be satisfied with two such newly created summary composite scores. The logic of this development can, of course, be extended to multiple subsets of variables.

If one knew how many subsets of variables there were, the task of summarizing significant sources of variance in terms of composite scores would be simple indeed. Take the two subset case again. If the correlation between the two composite scores is essentially zero, it is apparent that the two new scores, in addition to summarizing information within each subset of variables, are necessary and nonredundant to a complete description of the data. After all, the standing of a given entity on one composite variable cannot be predicted from its standing on the other composite variable. Consider now the opposite extreme. We have generated two new composite scores, but find that these scores correlate perfectly, or at least up to the maximum permitted by the error they contain. Apparently our concept about the existence of two different subsets of variables was wrong; they appear to be interchangeable since they are so highly interrelated. In this case, of course, the obvious remedy is simply to combine the two different composite scores into a single score which could capture the essence of all data on all variables. The task of deciding how many such composites are needed becomes more complicated as the variables contain greater amounts of error. For example, even if a new composite were relatively independent of others, but it was made up of component variables that are all very unreliable, the composite itself could contain so much error as to make it practically worthless. Later we shall see that this problem of deciding how many composites are necessary for a given set of data is somewhat subject to arbitrary decisions. Statistical tools will also be found helpful.

PRINCIPAL COMPONENTS VS. FACTOR ANALYSIS

We shall now become a bit more precise and distinguish between two different kinds of procedures. In one procedure, we obtain a new composite variable as a linear combination of the given variables (in the simplest case, for example, by simply adding up the scores a given entity has on all variables). This composite variable is a new dependent variable. In another procedure, we seek to determine independent variables such that our given variables can be considered to be linear combinations of these independent variables. The independent variables "explain" the given dependent variables. Of course, if we wish, we may try to obtain an estimate of this new independent variable as well.

The first procedure is known as principal components analysis. Principal components are simply new dependent variables created from a given set of variables. Of course, since there are many types of new variables possible, the principal component variables must also have a built-in restriction. This restriction is that the first new composite score, or first principal component, shall account for as much variance as possible among the total variance of all variables. In our simple example, where all variables were perfectly correlated, this component could just be the sum of all variables. As a new variable designed to summarize as much variation in the data as possible, the first principal component cannot be beaten. Suppose that variation among all entities cannot be summarized adequately by a single score. Then there are more principal components in the data. The second component is that linear combination of the given observed variables that accounts for as much variance as possible, subject to the restriction that this new component will be uncorrelated with the previous one. Thus, there would be two scores that are uncorrelated with each other, but in combination they may predict all variation in the observed variables. If not, the process is repeated until as many components are obtained as are required to predict all variation in the data. Of course, the last few components may be very small in nature, so that while they are nonzero, they may be practically insignificant. They may also be statistically unreliable, so that they may be discarded. Information from the components analysis may be summarized in a component loading matrix

representing the correlation of each given variable with the particular components. In addition, one could compute the actual component scores, which represent the scores of the entities on the components. For each individual there will now exist as many scores as there are components, in contrast to the original data, where there were as many scores as variables. To illustrate, if a study had 100 variables to begin with, there may only be a half-dozen important principal components. A tremendous amount of data reduction will have been realized.

The second procedure is known as factor analysis. Here the goal is not to obtain new variables (principal components) that are functions of the given variables but rather just the opposite. That is, one wants to determine new variables such that the given variables are functions of these new variables. If the given dependent variables are functions of these new variables, it is entirely appropriate to consider the new variables, or factors, as explanatory independent variables. In the example given above where all variables measured the same thing except for random errors, each given variable can be considered to be a function of two independent variables or factors: the "true" variable without error, and an error variable. Factor analysis, in contrast to components analysis, hypothesizes that each observed variable that a scientist must deal with will have a random error part. It does not believe that the strategy of adding up observed variables to generate a new variable is very profitable for this very reason, since such sums will also contain error. Factor analysis attempts to remove the error portion from each variable, so as to leave open to further analysis only the systematic and reliable part. Actually, factor analysis goes one step further. It recognizes that variables contain not only error, but something specific that a given variable may measure but that no other variable measures. This specific part of the given variable may be important for some purposes, but for many purposes it can be ignored. In particular, when summarizing vast amounts of data, one may wish to find out only what it is that various variables share in common; specific aspects of a given variable that are not shared by other variables may be relegated to an irrelevant role. In the typical factor analytic situation, this concept is accepted and defined in the following way: let the part of a given variable that is shared by many other variables be called the common part; the part that is unique to a given variable, its specific and error part, be called the unique part. The common parts, however, are considered much as principal components--there may be many sources of variation in the common parts. Each of these sources of variation is called a common factor, or simply, a factor. Of course, there are also unique factors--one for each variable. And so, in factor analysis, one hypothesizes that there are more factors than variables initially given. (Contrast this to principal components, where there are always fewer components than variables.) Of course, in factor analysis it is the common factors that are of special importance, since these represent independent variables that share variance among many dependent, given variables.

An excellent illustration of factor analysis comes from the area of intelligence. Indeed, factor analysis was born in the context of the study of intelligence. It was hypothesized that whatever set of intellectual variables one measured, each such variable might actually be generated by two independent processes: a general intelligence process, and a process unique to that variable. The unique process represents the combined effects of random error (that is, the score a person receives on the given variable depends in part upon chance) and a true specific ability or skill that is being measured by that particular variable. Thus, the maze tracing performance of a given subject may depend in part upon chance and in part upon his skill with this particular type of maze; these effects combine to generate the unique part of the actual observed score on a maze test. But, it was also hypothesized that maze performance depended in part upon general intelligence. Similar analyses were made of other verbal and quantitative intellectual tasks. According to this theory, whatever intellectual variables a scientist measured, performance on them would depend in part upon general intelligence and then also on a unique aspect. It was hypothesized, in other words, that there exists one (and only one) intelligence factor common to all variables. Needless to say, it turns out a half-century later that this theory is wrong. There appear to be several distinct intellectual factors, not only one.

Notice the phrasing of the previous discussion: general intelligence determines in part a given person's performance on a given intellectual variable. In other words, the scores on observed variables are assumed to be dependent variables generated by independent variables (here, general intelligence and a unique component). This is the major distinction between principal components and factor analysis. Components simply summarize data; they are new dependent variables. Factors are independent variables; they represent processes that generate the observed data. It is for this reason that many experts in multivariate analysis consider factor analysis as an explanatory tool and principal components as a descriptive tool.

It cannot be justified here, in an introductory presentation, but there are other reasons to preferring factors to components when explanation is desired. These have to do with the concepts of scale-freeness and factorial invariance. The idea of scale-freeness is that one can obtain the same factors no matter how one happens to scale the variables (e.g., measure in inches, miles, or meters). This is not true of components; components depend upon the particular choice of scale or variance for the observed variables. Factorial invariance refers to the fact that the same factors can be obtained in differing populations of entities. Only factor analysis can discover invariant factors.

A final major distinction between components and factors has to do with the issue of variance accounted for versus "covariance explained". Both methods try, analogously to anova, to "account" for variance. But the main goal of principal components analysis is to account for as much variance on each and every variable as possible. In contrast, the factors of factor analysis try to account mainly for the covariance or correlation among variables--it is the common variance that is considered important. If error variance is not accounted for by the common factors, all to the good. Common factors are truly covariance-explaining, or correlation-explaining, rather than variance explaining.

In many sections below, we shall generally ignore the crucial distinction between components and factors because many of the principles relevant to one are relevant to the other. For example, the loading matrices are interpreted equivalently. Nonetheless, it should be recognized that there is a crucial logical difference between the methods. We shall point out where this difference translates into a procedural and interpretive difference.

FACTOR SCORES AND FACTOR LOADINGS

As was pointed out above, the input to a factor analysis is the data of entities on variables. Actually, this data can and must be transformed to simpler form, since the procedure is most effectively applied to the intermediate correlation matrix generated from the data. The correlation matrix represents the intercorrelations among all given variables calculated across the entities. Although factor analysis is a data matrix analysis method, it is typically the correlations that are "factor analyzed", though it would be appropriate at times to use covariances or cross-products instead. The mathematics of factor analysis itself are quite complicated, and we shall assume that stardardly available computer programs are utilized to perform them. At this stage our concern is with the output from such an analysis.

Logically, of course, there are factor scores and factor loadings, since the scores refer to each entry's actual score on a factor, and the loading refers to the weight that the given factor has in generating the observed score. Please recognize, however, that the factor scores are really hypothetical scores, since they cannot be calculated exactly. Of course, in some sense these scores can be estimated, as will be discussed later. A typical factor analysis does not bother to estimate these factor scores, since the interest usually resides in attempting to understand the given variables in terms of the factors. This understanding must be obtained from the factor loading matrix.

The factor loading matrix is a matrix of multiple regression weights. The weights are applied to the factors to predict the observed variables. The convention is typically followed that the (unknown) factor scores have unit variance. Then the weights are standardized beta weights.

Up to now we have not discussed whether the factor scores are correlated or uncorrelated. Since factor analysis procedures allow the experimenter to specify this option at his discretion, in the general case the factors must be correlated. Thus there will exist also a factor correlation matrix, representing the intercorrelation among the factors. This situation is completely analogous to multiple regression. The predictor variables are the factors; the predictors may be correlated; the criterion variable is a given observed variable. Of course, it is well known in multiple regression that the predictors may be uncorrelated among themselves. Then the beta weights are simply correlations; specifically, correlations between the criterion and a given predictor. Analogously, in factor analysis, when the factors are uncorrelated the factor loading matrix contains correlations--correlations between latent factors and observed variables. The factor correlation matrix can then be ignored, of course, since different factors have zero correlation. When the factors are taken to be uncorrelated, they are known as orthogonal factors; when they are taken to be correlated, they are known as oblique factors. The loading matrix for oblique factors is sometimes called a factor pattern matrix.

THE NATURE OF A FACTOR

In principal components analysis, a given component is simply a linear combination of variables. What is the "meaning" of the component? Nothing more or less than the fact that it is a new variable made up in a particular fashion from old variables. How about a "factor"? It is not a linear combination of variables, so how could one determine what it actually is? The answer depends upon clarity in the factor loading matrix, a clarity that is often called simple structure, in which the ideal loading matrix contains many zeros and only a few large loadings. It is simplest to take the case of orthogonal (uncorrelated) factors first.

Suppose there are two factors, so that the loading matrix provides information about the correlation of each observed variable with each of the two factors. An understanding of the factor depends upon its pattern of correlations with the observed variables. If one could locate an observed variable that correlated almost perfectly with the factor then, obviously, the factor "is" whatever the observed variable "is". High scores on the observed variable are then essentially synonymous with high scores on the factor; and low scores on one imply low scores on the other. The problem becomes more complex as the factor correlates only to an average extent with some variables. Then it becomes more important to pay attention to what the factor does not correlate with. For example, suppose one has a correlation of .7 of a vocabulary test with a factor, and a correlation of .1 of that test with a second factor; also suppose one has a correlation of .1 for a quantitative variable with the first factor and a correlation of .7 of this variable with the second factor. It would appear as if the first factor measures verbal skills in some way, and the second factor measures quantitative skills. Obviously, any such interpretation must be tentative, subject to cross-validation and other deductive experimentation. The degree of certainty of interpretation while looking at the factor loading matrix depends, of course, upon how many of the correlations have a clear interpretation. Such interpretation is made easier if it is known that certain observed variables are marker variables for a given factor. For example, if a set of marker variables have been designed to measure verbal information processing, then a factor with consistently high loadings from these variables can be more confidently interpreted as an information processing factor.

If the factors are correlated, the factor pattern matrix is a matrix of beta weights, as previously discussed. Beta weights can be interpreted similarly in terms of their pattern. But it must be noted that these beta weights are not weights applied to variables to generate a factor (it is the reverse), so that interpreting the factor becomes somewhat more tentative. Nonetheless, the principle of interpreting high and low loadings as to a clue regarding the factor holds equally well.

More will be said later regarding the nature of factors. At this point, be aware that factors are still an abstraction. It will be imperative to make the abstraction concrete. We discuss the problem when talking about cross-validation, below.

ASSUMPTIONS AND LIMITATIONS

Score Distribution

As might be supposed, it is convenient to assume that the variables have a multivariate normal distribution. That is, that each variable considered singly and in combination with others shows the characteristic normal distribution. Actually, it turns out that this assumption is not really absolutely necessary; the procedures and interpretations of factor analysis are applicable even if the distributions are not perfectly normal. However, as in much of statistics, the adequacy of any statistical test of significance depends upon the extent to which this assumption is tenable. In exploratory work, and in data reduction, where there is no particular intention of testing a given hypothesis, the inappropriateness of the assumption may not matter much. The procedure is very robust.

Linearity

Even if the scores are not normally distributed, it must be remembered that the factor analytic model is a linear model. If it is believed that the variables relate in nonlinear ways, or that the underlying factors are nonlinearly related to the observed data, it is necessary to use alternative methods of analysis. In the case of binary data, for example, where the assumption of linearity is hardly ever met, one may certainly carry out factor analyses. The

problem is that the forced nonlinear relations among variables may generate artificial factors that are not true representatives of the underlying process that is operative. For example, it is well known that the Guttman scale consists of a set of binary items that measure a single underlying dimension. Factor analysis of items such as this can lead to the incorrect conclusion that there is more than a single dimension or factor, because more than one factor is needed to account for the correlations among items. The first and biggest factor would be a reasonable approximation to the true underlying single dimension, but additional factors would be developed that are simple artifacts. The degree of artifactuality with binary data would depend on the extent to which the various variables have splits that are unequal. If all variables have means in the .4 - .6 range, the degree of distortion is probably quite minor, but care must still be taken not to overfactor. In the case of binary variables, it is probably most reasonable to use an alternative method such as monotonicity analysis (Bentler, 1970).

When the variables are not binary, but have only a few response categories, the factor model will do a reasonable job at recovering the true underlying dimensionality providing the variables are not too badly skewed. It is sometimes proposed that the factor analytic model be abandoned because of its strict linearity assumptions. It is suggested that models that allow for nonlinear relations among factors would be preferable, because there may exist fewer nonlinear dimensions than linear ones. Methods based on these ideas, however, have not held out much promise to the practitioner (McDonald, 1967).

In recent years there has been a "nonmetric" revolution in psychometrics. This has suggested that one should not perform analyses that require one to make use of the strict interval nature of the raw data variables when, in general, variables as measured often represent little more than rank-order information. One proposal has been that a method of monotonic principal components be adopted, where the underlying components relate in a rank-order fashion to the observed variables rather than a strict linear fashion. Actually, it turns out that with the error typically found in social science variables, methods such as this recover the true underlying dimensions no better, and probably more poorly, than the factor analytic model (Kruskal and Shepard, 1974). Thus, the researcher who is cautious in his use of factor analysis will not find a better alternative method, even if the strict metric and linear assumptions cannot be met.

Ratio of Factors to Variables

It is not typically considered an assumption, but the ratio of the number of entities to number of variables to number of factors must be sufficiently favorable to allow one to draw inferences about the factors. The reliability of such inferences hinges strictly upon having an adequately large and random sample of entities. Several hundred individuals might be a good minimum, but far more are needed if there are also more than a hundred variables. A good rule of thumb might be that there should be at least five times as many entities as there are variables; but the more, the better. Similarly, the adequacy of the analysis will depend strongly upon the number of factors that exist. Again, the rule of thumb that there should be at least five variables for every factor is just an absolute minimum. Thus, if one has fifty variables, identifying and reliably measuring ten factors is about the outside limit that can be expected. The more marker variables per factor, the better are the chances of having the analysis reveal the true structure of the data. If one has a set of 20 variables measured on 60 subjects, it will generally not prove possible to have confidence in more than two or three factors. Obviously, the confidence one may have in the data will be mirrored in a significance test or, alternatively, in the reliability of the uncovered dimensions.

Missing Data

Missing data cannot be handled by the method. If there are only a few missing entries, then estimation of the missing data is possible by substituting mean values for missing data on given variables. Obviously, too much substitution will distort the picture dramatically. In the context of data on very many entities, a few missing entries will not matter. The fewer the entities, the more distortion will occur. Imagine the basic situation as one of the bivariate scatterplot of correlation. How many points can be missing or distorted without the correlation coefficient being distorted? To some extent this depends on the specific location of the missing data, but one or two percent error caused by data substitution is probably not generally harmful.

Scoring

It is assumed that data variables are experimentally independent. It is not possible to score the same questionnaire item in more than one variable, for example. Similarly, it is not appropriate to score one response alternative on a forced-choice item on one variable, and the other alternative on another variable. Similarly, it is not possible to include as one variable a score that is a linear combination of other variables already included; that is, if X is a variable, and Y is another variable, it is not possible to include the variable (X-Y) or (X+.3Y) in the same analysis.

Universe Representation

In general, it is assumed that the sample of variables and the sample of entities must be adequate representatives of the universe of variables and the population of subjects. It is difficult to give precise rules about this assumption. In the case of entities, the assumption is the rather typical one that is proposed in the theory of statistical inference. In the case of the universe of variables, there is no well-agreed upon definition of the universe for any given content domain that may be under investigation. Nonetheless, if certain types of "causative" factors are not being sampled because the variables chosen for analysis are systematically biased by avoiding these factors, there is certainly no opportunity for discovering these causative factors. Surprisingly, there is a similar constraint about including variables that are almost duplicates of one another. If a sufficient number of duplicates are included in the analysis, they are sure to form a factor--but quite possibly, an artifactual and trivial one, at that. An example of a duplicate would be alternative wordings of exactly the same question. The range of content included in the variables thus constrains the final factors to be discovered. It is in this context that the familiar phrase can be heard, that one does not get more out of a factor analysis than one puts into it. The quality of the end result depends on the quality of the input.

There are a number of other considerations in developing a competent factor analytic study. These will be discussed below, when the exploratory model will be discussed in further detail.

METHODS

There are three major purposes for factor analysis. The first is the one initially mentioned in association with principal components analysis--data reduction. The second is the one alluded to in the previous section, namely the exploration of data to formulate hypotheses about the nature of significant factors that generate the data. The third is relatively new but of major importance, namely, confirming or testing hypotheses about given factors.

DATA REDUCTION

Faced with masses of multivariate data, the investigator often faces the task of making the data more manageable and more easy to grasp. How can one comprehend the scores of 1000 individuals on 200 variables? The 200,000 data values are simply too overwhelming to process, and there is little relief gained by looking at the correlation matrix. The correlation matrix, representing the correlation of each variable with each other variable, has a potential 19,900 different entries. Suppose, on the other hand, that this mass of data could be reduced to one single important variable, with its 1000 scores, and its factor loading matrix of 200 entries (each variable correlated with the factor). Obviously, a great savings is obtained. Of course, in general there will be several factors, not only one. But the gain will still be substantial.

For purposes of data reduction, it does not matter much whether one is obtaining principal components or factors. If one or several linear combinations of variables (components) effectively exhaust all the important variance in the data, much is gained by the procedure. Then the components can replace the mass of data for further analysis or experimentation. Similarly, if a few factors account for all the covariance or correlation among variables, then reducing the mass of data to these few factors would be desirable.

This type of data reduction is useful in combination with other methods of analysis. For example, suppose one is attempting to build a prediction equation to predict some drug variable.

However, one may have over one hundred predictor variables. It is known that building a regression equation with that many variables is fraught with danger; it will typically not cross-validate well. Ideally, one would like to be able to select those few (say, 10) variables that give the highest prediction of the criterion. Stepwise regression procedures may be used, but such stepwise regression is itself highly problematical in application, since the procedure cannot guarantee the optimal selection of variables. Factor analysis (not principal components) is a viable alternative procedure. For example, one can intercorrelate all variables, including the criterion, factor analyze the matrix to be sure to extract enough factors to account for the criterion variable's covariance with all variables, and then rotate the matrix to obtain a factor loading matrix (rotation is discussed explicitly below) such that the criterion variable's loadings are in simple structure form (many zeros). If a given factor is involved in the criterion variable, any variable that measures the factor well can be selected as a predictor variable. If a given factor is not involved in the criterion variable, it can be ignored.

EXPLORATORY FACTOR ANALYSIS

Contrasting exploratory factor analysis with data reduction, here one desires a theoretical understanding of the nature of the factors. Arbitrary linear combinations no longer serve a purpose. Instead, it is desired to obtain a taxonomy, or to improve measuring instruments, or to develop criterion measures for some process. In purely exploratory work, one may not have a well-developed theory, nor enough previous empirical data, to be able to predict with great accuracy what the various factors might be that account for the covariation observed among variables in a given domain. It is hoped that the nature of these underlying variables can be clarified through the process of forming tentative hypotheses, exploratory factor analysis, reformulation of hypotheses, further exploratory work, as well as the beginnings of confirmatory experiments. Such an approach might be taken while developing a taxonomy of basic personality dimensions, understanding alcohol-related attitudes, or analyzing the dimensions of physiological responsiveness of the autonomic nervous system. In exploratory work one may be ignorant about the number of underlying dimensions as well as makeup of given variables in terms of the dimensions, and the task is to make an educated guess about these things. This procedure, the most frequently used, is discussed in further detail below.

CONFIRMATORY FACTOR ANALYSIS

At the opposite end of a continuum with exploratory factor analysis lies the confirmatory approach. Confirmatory factor analysis serves to cross-validate findings from a previous study or from a series of previous studies. They enable one to test the hypothesis that the given number of dimensions underlying the covariation among variables is some specific number k. For example, a test of the hypothesis that all intellectual variables are composed of a single general factor leads to the specific hypothesis that one common factor accounts for all correlations. Alternatively, the notion that all intellectual performance can be accounted for by two factors, one verbal and one quantitative, leads to the hypothesis that the correlations among the observed intellectual variables can be accounted for by two common factors. Furthermore, one may be able to specify that certain variables involve the verbal factor only, while other variables involve both factors, and still others involve the quantitative factor only. The statistical significance of these parameter estimates can be evaluated, and the correctness of the theory evaluated. If the theory is incorrect, this can be determined.

ALTERNATIVE FACTOR ANALYTIC DESIGNS

Up to this point we have been discussing the factor analysis of a set of variables, measured on a set of entities, in order to determine what the sources of variation underlying the variables are. In the typical social science application, variables are quantitative indices of one kind or another. The entities may be subjects, societies, animals, etc. A little reflection will make it obvious that almost all mathematical-statistical techniques are not sensitive to what a "variable" is nor to what the "entity" might be. The mathematics of the procedure, its assumptions, and the end result are legitimate products providing that the assumptions are met reasonably well. This freedom of choice has made it obvious that there are a number of other possible alternative factor analytic designs other than the standard one. The factors that result from any procedure depend upon an understanding of the correlation matrix that is being analyzed.

Consider only the possibilities opened up by considering, in addition to variables and entities, the possibility of obtaining assessments on several occasions (Cattell, 1952). Up to now it has been implicitly assumed that all scores were obtained on a single testing occasion. As one also considers multiple occasions, there are six possible designs. These are the following:

R -- correlate variables across entities
Q -- correlate entities across variables
O -- correlate occasions across variables
P -- correlate variables across occasions
S -- correlate entities across occasions
T -- correlate occasions across entities.

An understanding of the variables is afforded by the standard R design but also by the P design. In the R design, the occasion is fixed or held constant; in the P design, the entity is fixed or held constant. An understanding of how entities are grouped together is afforded by the Q and S designs, with occasion and variable being fixed respectively. An understanding of how occasions covary together is found in the O and T designs; in the first case, the entity is fixed, in the second, the variable is fixed. While these designs are not used often, they hold promise, providing that the data base is adequate to the task. Other sections of this book describe longitudinal designs in some detail, where occasions may vary (chapters 6 and 7). Similarly, methods of analysis related to Q designs are discussed elsewhere (chapter 5).

It may be appropriate to point out, however, that there exists a method for analyzing all sources of variance in a three-mode data matrix simultaneously. This is the method developed by Tucker (1966). Suffice it to say that one also requires complete data on all variables, all entities, and all occasions--but the final result is a set of factors for variables, for entities, and occasions, as well as an expression of how these factors interrelate one to another. While this is the ideal procedure for the analysis of three-mode data, incomplete data would make possible factor analysis according to one of the designs described above. If one has data only for a single person, for example, the O and P designs would be appropriate. When one has data for only a given variable, it may nonetheless be of interest to determine how persons or occasions covary according to underlying factors. Then S and T designs would be appropriate.

PROCEDURES: EXPLORATORY FACTOR ANALYSIS

Since the most frequent use of factor analysis occurs in exploratory situations in which the investigator is attempting to delineate the underlying variables that might account for the correlations among his observed data, particular attention might need to be given to the various steps involved in carrying out an exploratory factor analysis. These steps start with a theoretical analysis of the situation, and include obtaining the data sample, correlating the variables, extracting the factors, rotating the factors to a more meaningful position, interpreting the results, and cross-validating the results.

THEORETICAL ANALYSIS

Factor analysis can be undertaken in order to understand some particular domain or universe of variables. The first step thus should consist of trying to make as explicit as possible the nature of this domain of variables. In the drug area, for example, one might consider the domain of attitudes toward specific chemical agents. Alternatively, one might be concerned with the domain of physiological response to injections of drugs. Variables not particularly relevant to the domain should be excluded--they should not be "thrown into the analysis to see what might happen". The problem is that extraneous variables typically affect the results of a factor analysis. The more explicit one can be about the goals of the analysis, the better. Within the defined domain, previous empirical research or various theories might suggest that there exist logically or empirically distinct groups of variables. Such groups of variables might be potential candidates as factors, or as dimensions underlying all variables. Each of these groups should be spelled out as clearly as possible. Then marker variables might be selected for inclusion in the analysis to represent each of these a priori groups. Such marker variables will be particularly useful later in identifying a given factor. As pointed out before, numerous variables should potentially exist to measure any particular factor that might be anticipated.

It is important to know that if the variables are arbitrarily thrown together, the data has already been obtained, and a factor analysis is looked upon as a method of salvation for finding meaningful results from unplanned data, one's expectations may not be met. It is certainly true that factor analyses have at times turned up useful and important results when the analysis was completely post hoc; but in general, great care ought to be placed in planning the entire study before factor analysis is actually invoked as a procedure.

OBTAINING THE DATA SAMPLE

A part of the theoretical analysis consists of defining the subjects or entities to which one expects the factors to generalize. These entities must be sampled. Subjects that just happen to be available for a given study may be unusual in some way; if they do not truly represent the population one is attempting to generalize to, any result of an analysis will also not be representative of the appropriate population. Since the step of appropriate sampling of subjects is generally well explained in standard statistics books, most researchers are aware of the importance of random sampling if possible.

An important additional consideration in subject selection in factor analytic work is the issue of restriction of range on variables. Some people suggest that one should perform factor analysis only on a given subset of subjects, such as among males; on the other hand, others suggest that the analysis should be conducted in very heterogeneous groups, for example consisting of both males and females. There is no appropriate way of answering which of these opposing methods is the ideal procedure. This depends entirely on the aims of the investigator, specifically, his intention to generalize to some population. If his interest is to generalize to "people in general", his sample of subjects should include as much heterogeneity as is representative of "people in general".

Once having defined the population, a specific sample of subjects must be tested to obtain data on all relevant variables. As pointed out previously, the number of subjects should be very large, and particularly in relation to the number of variables. Indeed, if at all possible, the subject sample should be split into two halves--half for the purpose of the factor analysis itself, half for cross-validation of the results. This issue will be discussed further below.

CORRELATION

The next step in the analysis consists of calculating all the intercorrelations among all the variables in the analysis. The correlations represent the prime input for the factor extraction process, because factor analysis aims to account for the intercorrelation among variables. Obviously, computer programs as discussed will perform this step of calculating the correlations as well as the next two steps of factor extraction and rotation. If the correlation matrix is very small, it is sometimes possible to look at the intercorrelations among variables and get some idea of how the variables might be grouped. There is, however, no one-to-one correspondence between such a subjective view and the results of an analysis.

EXTRACTION

The next step is finding and extracting the factors from the correlation matrix. Many computer programs provide no choice whatsoever among methods of finding the factors. One must simply accept the method that they choose. Others provide the choice between principal components analysis and factor analysis. Often such a choice is couched in the question about "communalities". If communalities are to be estimated, these are numbers that will be placed into the main diagonal of the correlation matrix prior to the factor extraction process, or during it. They control whether error variance, as well as specific variance for each factor, will be included or excluded in the analysis. In the principal components method, the correlation matrix will not be modified, and no communalities need be estimated. In other methods of factor analysis, communalities will be estimated either as a product of the procedure, or one will be asked to provide an initial estimate of the communality. This estimate might be the highest correlation of a given variable with all others, or the squared multiple correlation of that variable with all other variables. The latter method is appropriate. Furthermore, one may have the choice between methods of extraction known as maximum likelihood, minimum residual, least squares, or other techniques. The maximum likelihood procedure has much to recommend it because it can provide statistical tests of goodness of fit of the model, while the other methods currently cannot do so.

Once the method of extraction has been determined, there is still the problem of determining the number of factors. Some programs have an automatic self-decision built into them. They choose the criterion known as "number of eigenvalues greater than 1" of the correlation matrix. This criterion simply says that any factor should account for as much variance as any single variable (a standardized variable has a variance of 1). Alternatively, one may be able to specify the number of factors to extract. This specification should be based upon the theoretical analysis performed prior to the study. Of course, a choice about the number of factors may be "wrong," in the sense that fewer factors really exist in the data, or in the sense that there may be more. One's decision here should be exploratory, and it may perhaps be necessary to perform several analyses, varying the number of factors in the range considered reasonable. Then, alternative solutions must be compared and the best one chosen.

ROTATING THE SOLUTION

Once the number of factors and method of extraction have been decided upon, there remains the problem of selecting a method of transforming or rotating the solution. This is an essential, but somewhat difficult to explain, concept. The factors of factor analysis actually define a dimensional space, with a variable being a point that has some location in that dimensional space. The problem of rotation is to select meaningful axes to describe the space. As an illustration, our daily life is governed by a geographical surface that is three dimensional. Considering the world as a globe, we can locate any point by a description in terms of the standard axes North-South and East-West. To be strictly accurate, we would also have to define the third dimension as distance from the center of the earth. Considering only the two dimensions, or factors, North-South (N-S) and East-West (E-W), one may ask the question: why do we accept these orientations as the basic ones? Suppose instead we defined the orientations of the axes as NE-SW and NW-SE. We could equally well locate any point relative to those axes--the location of the points remains fixed, but the axes have been changed. The problem of rotation is exactly the problem of rotating the axes (while keeping the variables in their same Euclidian location) so as to find a set of axes or factors that are meaningful and easy to comprehend and describe. On the globe, if we desire to differentiate between northern and southern hemispheres because of the seasons, for example, it makes sense to have an N-S factor. Any other rotation would not be as meaningful, though strictly mathematically legitimate.

The interpretation of a factor will depend upon the method of rotation chosen. For example, if the extracted factors are not rotated, the mathematics of the procedure tend to yield a large general factor that correlates with almost all variables, as well as other factors that are bipolar in nature, having some variables correlating positively with it and other variables correlating negatively with it. Such factors may make sense. For example, one may expect a social desirability factor in questionnaire data that might have high correlations with all variables. In a study of emotion, one may expect some bipolar factors, such as one dealing with pleasantness versus unpleasantness of emotion. On the other hand, and more typically, one may wish to break up a general factor and the bipolar factors so as to obtain factors that are more highly correlated with fewer variables. The higher the factor loading correlation, the more specific an understanding one can gain of a factor. One can also more easily interpret many zero correlations between various variables and factors, since such zero correlations clearly define what a factor does not measure. Rotation by standard computer programs such as varimax and orthosim (Bentler, 1977) produce these relatively easily interpretable "simple structure" factors. In general, rotation is to be recommended.

An additional option available in some rotation programs is one of allowing the factors to be correlated or uncorrelated. Procedures such as varimax or orthosim produce orthogonal or uncorrelated factors. It is also possible to request correlated factors via some transformation procedure such as oblimin, oblimax, Harris-Kaiser, oblisim, etc. If one has reason to expect that the basic underlying variables that will become factors should logically be correlated, it makes good sense to ask for "oblique" or correlated factors via such a transformation or rotation option. For example, in a factor analysis dealing with intellectual variables, if one hypothesizes a verbal and a quantitative factor as two distinct intellectual factors, one may nonetheless believe that these two factors may be correlated to some extent. The overlap of the factors, or the correlation between them, of course, might represent general intelligence. The oblisim procedure (Bentler, 1977) not only produces a meaningful simple structure, but it provides a coefficient to evaluate the degree of simplicity attained. It is also to be recommended because it is scale-free with respect to the arbitrary scale of the factor scores.

Another class of rotation procedures exists. These are the target rotations, or "procrustean" procedures. When an investigator has a specific hypothesis about the factors that might exist and knows which marker variables should define each of the factors, he may wish to combine the above "blind" rotation procedures with target rotation procedures. He will be asked to specify

which variables are expected to define a factor, and possibly which variables are expected not to define the factor. This definition is utilized through a mathematical optimization procedure to produce factors that are as close as possible to the ones hypothesized by the investigator. When dealing with large masses of data, or when theory is reasonably well advanced, target rotation procedures are to be strongly recommended. Many computer centers, however, do not have these procedures available.

INTERPRETATION

A final step in a factor analysis is interpreting the results. The investigator will have at his disposal the factor loading matrix. If the factors are orthogonal, entries in this matrix will represent the correlation of variables with factors. He can gain an understanding of what a given factor is by analyzing which variables seem to correlate highly with the factor, and which ones do not. He may wish to provide a temporary name for the factor, but this should be done with care, because such a name should be no more than a hypothesis about the factor. If the factors are correlated, he will have available a factor pattern matrix. This factor pattern matrix represents the weights attached to the factors in producing the variables. These should be evaluated as standardized beta weights in multiple correlation analysis. They will not have the same range as correlations, but the general principles of interpretation would be quite similar to that described above. The given row in the factor matrix will represent the weights attached to the various factors in producing a given variable. To understand a factor, one will want to look for an interpretable pattern of high and low factor loadings that makes sense, in terms of one's understanding of the variables.

Should the results of a factor analysis prove to be uninterpretable, it is possible that the investigator overfactored or underfactored (too many or too few factors). With overfactoring, there is a tendency to have many factors, each of which is defined by very few variables. With underfactoring, there may be only relatively large factors; one may suspect that further, smaller factors also exist. As pointed out previously, the number of factors must be interpreted in the light of the theory that one has about the factors. If a statistical test is available for the number of factors, it should be used as a guide but not an absolute criterion for decision.

CROSS-VALIDATION

This last step is unfortunately all too often ignored. The results of a factor analysis should not stop with a single study and a consequent factor loading matrix. Once an idea has been obtained about the nature of the factors, other means of gathering scientific evidence must be brought to bear upon this interpretation. If the original subject sample was large enough to be divided in two halves, there remains the possibility of validating the results in the data from the as yet unanalyzed sample. One way to perform such a cross-validation would be to perform a confirmatory factor analysis, using statistical methods of factoring (Jöreskog, 1969). As a poor alternative, one can perform an exploratory analysis in the new sample, fixing the number of factors based upon the previous analysis, and evaluating the similarity of the new results by comparing them to the old ones.

Another method can be recommended strongly. Since many people are suspicious of factor analysis, one may wish to use a non-factor analytic way of verifying the results of the previous study. This can be done by using the results of the factor analysis as a guide for how one might measure a given factor. There are complicated ways of "estimating factor scores," and the most appropriate way to do this has been described by Bentler (1976). An alternative and much simpler approach involves using the following simple expedient. Determine which variables are believed to clearly define a single factor only. Scores on all these variables can be added up to produce a new variable. Should the variances on these various variables be quite different one from another, one may first wish to convert raw scores into standard scores and add the standard scores. When adding the scores, they should be added in a consistent content-interpreted direction--for example, in the direction of "smartness" if one is measuring intelligence. Thus the scoring direction on a given variable may need to be reversed before adding the variable. Actually, the scoring direction will be given by the sign of the factor loading for that variable. When generating a total score from the variables believed to measure a given factor, one is essentially obtaining a composite score as in any psychological test. Consequently, it is possible to evaluate the reliability of the score. In this case, reliability must be based upon an internal consistency formula, such as a stepped up split-half correlation, coefficient alpha, or a dimension-free coefficient (Bentler, 1972a). If all the variables are indeed measuring something consistently, this internal consistency coefficient should be high. Equivalent results should be observed for all

factors, as scored in this fashion. In addition, these new total scores can be intercorrelated to generate a matrix of correlations. These correlations should be relatively low, in comparison to the internal consistencies, indicating that the different factors as measured indeed represent different entities. If the calculation of these total scores, analysis of internal consistency, and demonstration of intercorrelation among totals is based upon the new cross-validation sample, direct meaning can be attributed to the results. Factor analysis would have been used primarily as a means for grouping variables, but the final results do not depend in any way upon the factor analysis. Consequently, a skeptic of factor analysis would be convinced by this procedure.

ILLUSTRATIVE APPLICATION

NON-DRUG RESEARCH

An illustration of an exploratory factor analysis may be found in the domain of physical measurement. Mullen (1939) measured eight physical variables in a group of 305 girls. The variables were selected to deal with two distinct concepts of "lankiness" and "stockiness". The variables, as well as the intercorrelations among the variables, are presented below:

Correlation Matrix

Variable	1	2	3	4	5	6	7
1. Height	1.00						
2. Arm span	.85	1.00					
3. Length of forearm	.80	.88	1.00				
4. Length of lower leg	.86	.83	.80	1.00			
5. Weight	.47	.38	.38	.44	1.00		
6. Bitrochanteric diameter	.40	.33	.32	.33	.76	1.00	
7. Chest girth	.30	.28	.24	.33	.73	.58	1.00
8. Chest width	.38	.42	.34	.36	.63	.58	.54

This example, taken entirely from Harman (1967), can illustrate some of the concepts described in previous sections. Turning first to the correlation matrix, a close look at the pattern of correlations shows that variables 1-4 are very highly intercorrelated. Apparently these variables are measuring something in common. Similarly, variables 5-8 show high interrelations, suggesting they measure the same thing. On the other hand, the cross-correlations between these two sets of variables is relatively low, compared to the within-set correlations. An inspection thus reveals that there may well be two factors underlying the data, but that these two factors may be correlated. In this example there are only 28 different correlations, so that it is quite easy to pick out the grouping of variables. In an example with 100 variables there would be 4950 different correlations, far too many for visual inspection of any pattern. A technique, like factor analysis, would have to be involved to understand the possible latent independent variables.

In accord with the hypothesis, two factors were extracted from the correlation matrix by the minimum residual method. The unrotated loading matrix for this solution is presented below, as is the final rotated solution.

	Unrotated Factors			Orthogonal Rotated Solution	
	I	II	h^2	I'	II'
1	.86	-.32	.84	.87	.28
2	.85	-.41	.89	.92	.20
3	.81	-.41	.82	.89	.18
4	.83	-.34	.81	.86	.25
5	.75	.57	.89	.24	.91
6	.63	.49	.64	.19	.77
7	.57	.51	.58	.13	.75
8	.61	.35	.49	.26	.65

The un-otated solution is always an "orthogonal" solution; that is, the factors are uncorrelateu. As expected there is the first general factor, which has all variables correlating highly positively with it. The second factor, in contrast, is a bipolar factor. It has negative correlations with the first four variables and a positive correlation with variables 5-8. Apparently high scores on this factor (which we have not calculated, but can be interpreted in view of the correlations) go with having low scores on the lankiness variables as well as having high scores on the stockiness variables. Thus the factor seems to contrast stockiness versus lankiness. While the first factor makes theoretical sense as a "bigness" factor, the second one may be more difficult to understand. Consequently, a rotation was considered essential.

Before turning to the rotated solution, notice the column labeled h^2. These numbers represent how much variance both factors explain out of the total unit variance of each standardized variable. Thus variable one has 16% of the variance not accounted for by these two common factors--apparently, the remaining variance is not shared by other variables, and consists of random error and specific variance. One may calculate the quantities h^2 as the sum of squares of the elements in a given row of the left matrix.

The right part of the above table consists of a solution for the factors, after rotation, by an orthogonal simplicity method (orthosim). In contrast to the unrotated solution, the orthosim loadings made obvious the clustering of variables that was hinted at in the correlation matrix itself. This clustering becomes still more obvious in the diagram below.

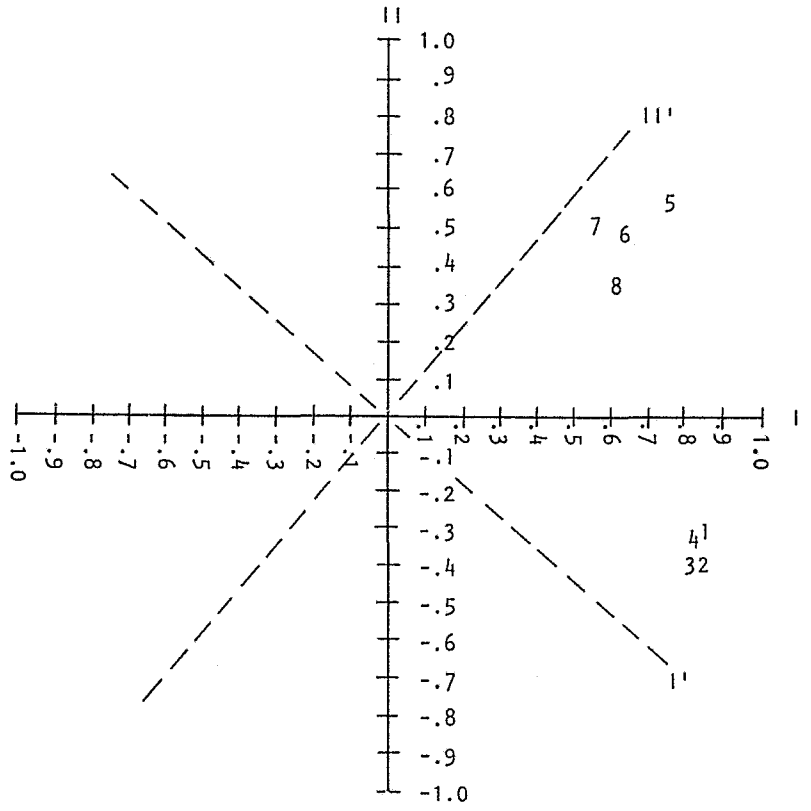


Figure 1.
Representation of Rotated and Unrotated Factor Solution.

The horizontal and vertical axes in the diagram represent the two 90° orthogonal dimensions in which the eight variables lie. Each variable can be exactly located by use of the numbers in the left loading matrix above. Thus, the position of variable one is given by a movement of +.86 units along the horizontal axis I, and then -.32 down the vertical axis II.

The orthosim rotation left the variables in their same exact location relative to one another and to the (0, 0) origin of the space. It simply moved the axes I and II to new positions I' and II', indicated by the dotted lines. The location of the points can thus be described relative to these new axes, and the orthosim loading matrix gives these geographical coordinates.

While it is obvious that the orthosim axes I' and II' are closer to the clusters of variables than the unrotated axes I and II, they are not as close as one might like. To find a better or more meaningful set of axes, it will be necessary to relax the idea that the axes must be at 90°. Thus it is necessary to use an oblique transformation. The one chosen for presentation is the oblisim solution (Bentler, 1977). The factor pattern matrix is presented below.

Oblisim Factor Pattern

	I'	II'
1	.88	.07
2	.96	-.03
3	.93	-.05
4	.88	.03
5	.01	.94
6	-.00	.80
7	-.06	.79
8	.11	.65

Factor correlation = .475
Factor simplicity index ≅ 1.000

The oblique solution makes clear that each factor seems to influence only the four variables in a given cluster of the diagram. Thus the meaning of a factor becomes still more clearcut in the oblique solution. To visualize the rotated solution, the reader may wish to consider the axes I' and II' to move closer together until they tend to move through the clusters of variables. Unfortunately, it is not possible to plot correlated factors by the procedures described above (please see more advanced descriptions of factor analysis for this purpose). What has happened, however, is that the factors were allowed to be correlated in the oblisim procedure. Actually, the factors correlate .475, not too far from what one might have guessed on the basis of the correlation matrix. The factor pattern matrix, it will be recalled, consists of weights (not correlations) applied to the factors in predicting the variables. Thus .88 is the weight for factor one in linearly predicting variable one. It will be seen that each variable is made up of essentially one factor in this solution; the other factor's weight is insignificantly small. This matrix, much more clearly than the unrotated solution, shows that the variables can be effectively grouped into two sets, as hypothesized. An output of the oblisim procedure is a coefficient (range 0-1) that summarizes the degree of simplicity in the factor pattern matrix. In this case, the index is an almost perfect 1.0.

If one were uncertain about the domain of variables, and had no clearcut theory about the variables, it might have been necessary to evaluate the relative merits of three solutions, one with a single factor, one with two factors, and one with three factors. Then one would have had to determine which solution made the most theoretical sense; and also, one would have to evaluate whether the factors account for the correlations quite closely. If two factors account for the correlations, there is no reason to extract a third one. There are no perfect rules for the number of factors, though maximum likelihood methods provide a statistical test that can be used as an aid.

The next step in verifying the results of this study would involve some kind of cross-validation. Simply publishing the above results would not satisfy critical readers. Had the sample of subjects been split initially, there would be the possibility of scoring variables 1-4 to generate a single lankiness score, and scoring variables 5-8 to generate a stockiness score. These scores would then have to be evaluated for internal consistency as well as for their intercorrelation.

DRUG RESEARCH

Segal (1975) wanted to determine the basic sources of variance that would account for the interrelations among a large set of daydreaming and inner process variables, self-report scales in the Murray need tradition, locus of control, sensation seeking, and a variety of self

ratings of extent, frequency, and duration of use of a variety of drugs and alcoholic beverages. There were eighty-one variables in all, far too many to understand by a visual inspection of the 3,240 intercorrelations. Data were obtained from 579 subjects of both sexes at two quite different universities. Since it was considered desirable to generalize the results to college students generally, rather than for a specific sex at a given college, one single analysis was undertaken. College and sex were included as coded variables to determine whether any factors were significantly associated with these variables; if this result were observed, it would suggest that the mean scores on a given factor would differ by sex or college. The main interest of the study was in verifying the imaginal process factors previously described by Singer and Antrobus (1972), and in relating these factors to possible drug use as well as to other personality dimensions.

The intercorrelations were factored to obtain a five factor solution. The specific method of extraction was not specified. A varimax rotation was performed, so that the resulting factors were orthogonal or uncorrelated. The factor loading matrix was quite large, of order 81 by 5, and will not be reproduced here, but the five factors will be described. Factors were interpreted and named in accord with the variables showing the highest correlations with the factor. The procedure recommended in this article, of splitting the sample, obtaining the factor solution in one half, and cross-validating the results in the other half, was not followed.

The first factor was a clearcut hard drug use factor, having correlations from .83 to .64 with various drug use variables such as hallucinogens, barbiturates, amphetamines, marijuana, cocaine, heroin, and other drugs. Although some personality variables and imaginal process variables correlated with this factor, these correlations were very small in nature (highest correlation .33). This factor indicates two things of interest to drug researchers: first, that virtually all drug use variables intercorrelate highly, and that this intercorrelation can be explained by a single, latent, independent variable or factor; and second, that drug use seems to be an entity pretty much to itself, not part of a larger constellation of personality attributes. It might have been possible, in contrast, that drug use was not a homogeneous entity, but rather a series of unconnected and uncorrelated activities. Actually, Segal also found another drug factor, his number three (the factor numbering is completely arbitrary). This factor did interconnect personality and drug use, but not drug use in general. Marijuana use only tended to define this factor along with experience seeking, adventure seeking, and autonomy measures from the personality domain. Thus marijuana use, in contrast to hard drug use, could be identified as a distinct entity, and different from drug use in general. Furthermore, its use was part of a pattern of exploration and autonomy. Frequent beer and wine drinking was only incidentally associated with this factor. This factor provides a reasonably coherent understanding of the nature of marijuana use, in contrast to hard drug use which appeared to be an isolated phenomenon.

Segal also reported three other factors that tended to be personality dimensions having little or no implication for drug use. One of the factors consisted of variables concerned with guilt or dysphoric daydreams. Another type of daydreaming factor was found, concerned with positively affected and vivid daydream variables. The final factor was defined by variables such as lack of endurance and achievement, mind-wandering and boredom, apparently a type of anxiousness and distractability in daydreaming.

FINAL CAUTIONS

Because factor analysis is so easy to use with canned computer programs, the method is easy to misuse. On the one hand, one may hope to salvage something useful out of an inadequately planned study, and on the other hand, one may believe that the procedure yields results of far greater importance than it is reasonable to expect. While it is important to recognize that factor analysis can be an extremely useful tool, exaggerations such as these should be avoided wherever possible. Factor analysis primarily provides a means towards an end, that of identifying the important, underlying, independent variables in a given set of data. As was pointed out in the previous section on cross-validation, there is no particular reason to rely exclusively on factor analysis to establish how well this goal can be met. Establishing the validity of the results in new samples, possibly by other techniques, is particularly important.

Many mistakes are possible in the use of factor analysis. In many instances the investigator is simply unaware of some consequences stemming from his decisions, from the nature of the

data, or some combination thereof. Some of the types of problems one might encounter can be listed here, but there is no space to discuss them in detail. Variables having badly skewed distributions, or extremely nonlinear interrelations with other variables, may cause problems. If several variables in the study are experimentally dependent, for example by having a given variable be a sum of other variables, the method cannot be used. Too few variables and too few subjects may be used in light of the number of factors. No marker variables may be included in the study. Almost identically equivalent variables may be included in the analysis. Failure to distinguish between principal components and factor analysis, evaluated relative to the goals of the investigation, can lead to errors. Too few factors may be extracted, or alternatively, too many factors may be obtained with the consequence of splitting the important factors. The rationale behind orthogonal or oblique rotation may not be evaluated carefully. Finally, the results may be overgeneralized. Errors such as these should be scrupulously avoided.

The reader interested in understanding more about the potentials and pitfalls of factor analysis should consult such sources as Comrey (1973), Gorsuch (1974), or the more sophisticated texts by Harman (1967) or Mulaik (1972). A complicated covariance structure model that includes factor analysis as a special case, but also subsumes univariate and multivariate analysis of variance, principal components, path analysis, and various other general methods (Jöreskog, 1973) is presented by Bentler (1976).

RESOURCES AND REFERENCES

COMPUTER PROGRAMS

Factor analytic computer packages exist at most university centers across the country. Programs also accompany various tests, such as Comrey (1973) or Horst (1965). Among the more well-known statistical packages, the BIMD series and the SPSS series contain factor analysis packages. Specifically, the reader may wish to use the BMD08M factor analysis program available in W. J. Dixon (Ed.), *BMD, Biomedical Computer Programs, 3rd Ed.*, University of California Press, 1973; the factor analysis procedure of N. H. Nie, C. H., Hull, J. G. Jenkins, K. Steinbrenner, and D. H. Bent, *SPSS, Statistical Package for the Social Sciences, 2nd Ed.*, McGraw-Hill, 1975; the general computer program *ACOVS* for analysis of co-variance structures prepared by K. G. Jöreskog, G. T. Gruvaeus, and M. van Thillo, available from the Educational Testing Service; or the package of factor analysis programs available in *SOUPAC*, distributed by the Computing Services Office of the University of Illinois. A program for exploratory factor analysis using the modern scale-free representations (Bentler 1972b, 1976, 1977) is available from the author. A target rotation procedure is also available from the author (Bentler, 1971).

REFERENCES

Bentler, P. M. A comparison of monotonicity analysis with factor analysis. *Educational and Psychological Measurement*, 30:241-250, 1970.

Bentler, P. M. Clustran, a program for oblique transformation. *Behavioral Science*, 16:183-185, 1971.

Bentler, P. M. A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research*, 1:343-357, 1972a.

Bentler, P. M. "Fast, Legitimate, Rank-independent Factor Analytic Procedures." Paper presented at the annual meeting of the Society of Multivariate Experimental Psychology, Fort Worth, Texas, 1972b.

Bentler, P. M. Multistructure statistical model applied to factor analysis. *Multivariate Behavioral Research*, 11:3-22, 1976.

Bentler, P. M. Factor simplicity index and transformations. *Psychometrika*, 1977, in press.

Cattell, R. B. *Factor Analysis*. New York: Harper, 1952.

Comrey, A. L. *A First Course in Factor Analysis*. New York: Academic, 1973.

Gorsuch, R. L. *Factor Analysis*. Philadelphia: Saunders, 1974.

Harman, H. H. *Modern Factor Analysis.* Chicago: University of Chicago Press, 1967.

Horst, P. *Factor Analysis of Data Matrices.* New York: Holt, 1965.

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika,* 34:183-202, 1969.

Jöreskog, K. G. Analysis of covariance structures. In: Krishnaiah, P. R., ed. *Multivariate Analysis - III.* New York: Academic, 1973. pp. 263-285.

Kruskal, J. B. and Shepard, R. N. a nonmetric variety of linear factor analysis. *Psychometrika,* 39:123-157, 1974.

McDonald, R. P. Nonlinear factor analysis. *Psychometric Monographs, No. 15,* 1967.

Mulaik, S. A. *The Foundations of Factor Analysis.* New York: McGraw-Hill, 1972.

Mullen, F. "Factors in the Growth of Girls Seven to Seventeen Years of Age." Unpublished Ph.D. thesis, University of Chicago, 1939.

Segal, B. Personality factors related tc drug and alcohol use. In: Lettieri, D. J., ed. *Predicting Adolescent Drug Abuse: A Review of Issues, Methods, and Correlates.* NIDA Research Issue Series, Vol. 11. Washington, D. C.: Government Printing Office, December, 1975.

Singer, J. L., and Antrobus, J. Daydreaming, imaginal processes and personality: a normative study. In: Sheehan, P., ed. *The Function and Nature of Imagery.* New York: Academic Press, 1972. pp. 175-202.

Tucker, L. R. Some mathematical notes on three-mode factor analysis. *Psychometrika,* 31:279-311, 1966.

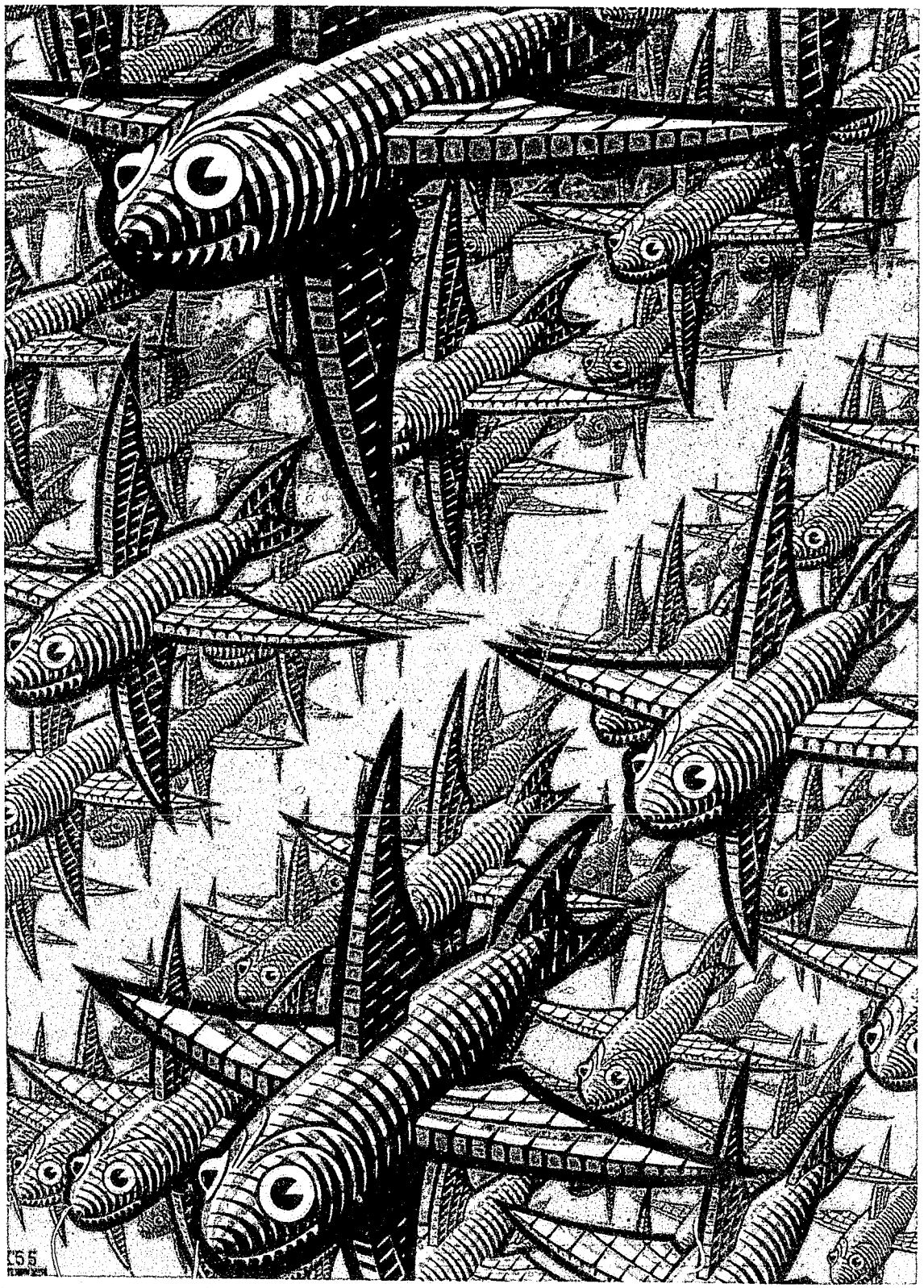


Plate 10

INTRODUCTION

Multiple regression and correlation is a data analytic procedure which is well established in psychology and the other social and behavioral sciences. When the relationship between one variable (the dependent or "criterion" variable) and a group of two of more variables (independent variables or "predictors") were to be studied, this has long been the method of choice. Indeed, it is difficult to find an applied general statistics textbook intended for graduate level work in the social sciences published in the last half century which does not include at least one chapter devoted to multiple regression and correlation. As exemplified and applied in practice until recently, the method tended primarily to be used in psychotechnological applications, for example, to predict future outcomes ("criteria" such as college grade-point average, rated job performance, length of hospitalization) by means of psychological test scores ("predictors" such as aptitude and personality test scores, ratings on psychiatric symptom scales) or other graduated quantitative variables (e.g., high school rank, years of experience, length of prior hospitalization). The purpose of such applications was usually to develop practical formulas for selection, classification or other decision-making functions of a forecasting type, and less often for the purpose of scientific understanding of behavioral phenomena.

During the last decade, however, the scope and generality of multiple regression and correlation analysis has so increased as to bear little resemblance to the method as represented in the standard textbooks. By the provision of appropriate methods of representation (coding) of information as independent variables, the method has been expanded to incorporate group membership variables (nominal or qualitative scales), nonlinearly related quantitative variables, and conditional relationships ("interactions"). Virtually any information (including the absence of information) may be represented as independent variables and its bearing on a single dependent variable studied. When thus expanded, many problems in data analysis are made tractable by this system, and some standard data analytic methods (analysis of variance and covariance, multiple partial correlation) are subsumed as special cases. It is this system of general multiple regression and correlation analysis (MRC) which will be described and illustrated in this chapter. Although we occasionally use the words "predictor" and "criterion" in conformity with the customary usage, our orientation is almost exclusively to the use of general MRC in the explanation of phenomena rather than to prediction in the narrow sense of forecasting.

The scope of this chapter is necessarily limited to an overview of the main features and possibilities of MRC; computational details, mathematical derivations, and extensive qualifications are unavoidably minimized or omitted. A full length, essentially nonmathematical, textbook treatment using the same concepts, terminology, and heuristics is provided by the authors in "Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences" (1975). (For the statistically sophisticated reader, we merely note that the general MRC system is effectively equivalent to the general univariate fixed linear model.)

METHODS AND PROCEDURES

General MRC is best described by first presenting the major features of conventional MRC and then showing how it generalizes via the utilization of sets of independent variables as units of analysis.

CONVENTIONAL MRC

For clarity of exposition, we will use a running concrete example and its numerical results, employing familiar variables. Assume a research investigation of factors determining the annual salaries of faculty members in a state university system. For a random sample of 100 (=n) cases,

we have data on salaries, the dependent variable (Y), and the following 4 (=k) independent variables (I.V.s): X_1 = sex (0 for men, 1 for women), X_2 = number of years since Ph.D. was awarded, X_3 = number of scholarly publications, and X_4 = number of citations in the literature during the preceding year. For the MRC analysis, the means, standard deviations, and the product-moment correlation coefficients among all pairs of these 5 (=k+1) variables are determined (see Table 1).

Table 1. Product-Moment Correlation Coefficients, Means and Standard Deviations of the Academic Salary Example

		Y	X_1	X_2	X_3	X_4
Y	Academic Salary	1.000	-.242*	.612**	.463**	.487**
X_1	Sex	-.242*	1.000	-.154	.049	-.006
X_2	Years since Ph.D.	.612**	-.154	1.000	.683**	.460**
X_3	No. of publications	.463**	.049	.683**	1.000	.297**
X_4	No. of citations	.487**	-.006	.460**	.297**	1.000
	Mean	18,029	.267	9.60	7.90	1.27
	Standard Deviation	7,481	.443	7.25	4.96	1.61

* P < .05 **P < .01

The Regression Equation

One of the fruits of an MRC analysis is the set of constants for a linear multiple regression equation of the following form:

$$\hat{Y} = B_1 X_1 + B_2 X_2 + B_3 X_3 + \ldots + B_k X_k + A. \qquad (1)$$

For our example, the equation derived from the information in Table 1 (after some heavy computation) is:

$$\hat{Y} = -3,266^* X_1 + 364^{**} X_2 + 224 X_3 + 1,296^{**} X_4 + 12,030.$$

The numerical constants determined for these data are the partial regression coefficients (the B_i, i=1,2,3,4) and the Y intercept (A); $\hat{Y}$ is the value of Y estimated for a subject by entering his X_i values in the equation. Now, if the $\hat{Y}$ values for all subjects were determined, it would be the case that these estimates are the best possible by the "least-squares" criterion. This means that if the error (residual) in estimating a subject salary as indexed by the discrepancy between the actual salary (Y) and his salary as estimated by the equation ($\hat{Y}$) were found and squared, i.e., $(Y-\hat{Y})^2$, and these squared "errors" were added for all n subjects, the resulting quantity, $\Sigma(Y-\hat{Y})^2$, would be smaller than that obtainable from the use of any other set of constants in a linear equation for these data.

The Numerical Constants: The Bs and A

The constants--the Bs and A--are not merely error-minimizing values, but have important interpretive properties. The partial regression coefficient B_i attached to a given independent variable X_i is the amount of change in the criterion Y associated with a unit change in X_i, given the presence of the other I.V.s in the equation. For example, an increase of one year since the Ph.D. (X_2) is associated with an increase in estimated salary of \$364 ($=B_2$) for any given combination of sex (X_1), number of publications (X_3) and number of citations (X_4). The latter qualification is important--it is what is meant by "holding constant" (or partialling) these other variables. Partialling is a centrally important feature of MRC, since it makes possible the determination of the net contribution of predictor over and above that of other predictors, i.e., holding these others constant statistically in research contexts where it is not possible to hold them constant by experimental manipulation. Below we will consider yet other ways of expressing a variable's partial (net, unique) association with a criterion.

Note that for sex, B_1 = -3266, with the predictor X_1 being arbitrarily scored 0 for men and 1 for women, thus a unit increase in X_1 implies going from men to women, and we see this is associated with a decrease in salary (since B_1 is negative) of $3,266, holding constant the other three variables. This in turn means that after one has allowed for the effects of salary on years since Ph.D., publications, and citations, there is a net mean salary difference of $3,266 favoring male faculty. Stated differently, these data show that for otherwise comparable standing on these other variables (which presumably reflect merit), there is nevertheless a sizable (and significant) sex difference in salary.

A, the Y intercept, is the estimated Y value when all of a subject's I.V. values (X_is) equal zero. In this example, A = 12,030 is interpretable, since zero values are meaningful for all four I.V.s. Thus, a male (X_1=0) faculty member, fresh from his Ph.D. (X_2=0), with neither publications nor (necessarily) citations (X_3=0, X_4=0), would have an estimated salary of $12,030. Assuming that the relationships are straight-line, this estimate would approximate the mean salary of such subjects.

The Standardized Partial Regression Coefficient: β

When the units in which the variables are represented are arbitrary or otherwise not meaningful, the analyst may prefer to work with the standardized partial regression coefficients, the βs. These are the regression weights which result when all the variables are rescaled to have a mean of 0 and a standard deviation of 1, i.e., z-transformed or standardized. So rescaled, the regression equation is:

$$\hat{z}_Y = \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + \cdots + \beta_k z_k. \qquad (2)$$

For the illustrative example, this equation reads

$$\hat{z}_Y = -.193^{*} z_1 + .352^{**} z_2 + .149\ z_3 + .279^{**}\ z_4.$$

The interpretation of a (standardized) β_1 remains the same as for a ("raw") β_i: a unit change in an I.V.'s standard score (z_i) is associated with a change of β_i in the dependent variable standard score (z_Y), but the units now are comparable in the sense that each unit is a standard deviation on the variable in question. Thus, for example, changes of one standard deviation in number of publications (z_3) and in number of citations (z_4) are associated respectively with changes of .149 and .279 of a standard deviation in salary.

Since the z-transformation of a variable is a simple linear transformation, no correlation values are affected. It follows not only that the product moment r of X_i and its z_i, and of Y with z_Y, is 1.00, but also that the r between $\hat{Y}$ and $\hat{z}_Y$ is 1.00.

The Multiple R and R^2

The multiple correlation (R) of a criterion with a group of k predictors (X_1, X_2,...X_k), symbolized as $R_{Y.12...k}$, is the simple product-moment correlation between the actual criterion value Y and the estimated criterion value Y (or $\hat{z}_Y$) obtained from the regression equation, explicitly

$$R_{Y.12...k} = r_{Y\hat{Y}}. \qquad (3)$$

It is thus the correlation between the actual dependent variable value Y and its best estimate (in the least squares sense) as obtained by the regression equation using the independent variables. Although the computation of R is not accomplished by the literal application of (3), that nevertheless is its definition and most straightforward interpretation.

In the analysis of relationship, it is very useful to work with squared correlations of all kinds, which may be interpreted as the proportion of the variance in one variable which is "accounted for" by the other. In what follows we will generally follow the practice of quantifying relationships in terms of proportions of Y variance variously accounted for. Thus, we interpret $R^2_{Y.12...k}$ as the proportion of Y variance accounted for by the group of predictors, i.e., via their optimal combination which produces $\hat{Y}$.

In the running example, $R^2_{Y.1234}$ = .4671** (and $R_{Y.1234}$ = .683**), indicating that when optimally weighted, the four I.V.s account for about 47% of the variance in salary in the sample, or, equivalently, yield estimated salaries which correlate .683 with the actual salaries.

Whole, Semipartial and Partial Correlations

In considering the bearing of each of the predictors on the criterion, MRC provides three different correlation coefficients, whose squares are interpretable as proportions of variance. Each offers a different aspect of the relationship of X_i to Y.

The squared product moment correlation: r^2_{Yi}. The simplest of these is the ordinary ("zero-order") product moment r between the given X_i and Y, given in Table 1 above. Its square, r^2_{Yi} gives the proportion of criterion variance linearly accounted for by the predictor X_i alone, ignoring any relationship X_i may have with the other predictors or their relationship to Y. For example, Table 2 shows that in our running problem, years since Ph.D. has the largest r^2_{Yi} among the four predictors, .3749 (r_{Y2} = .612), and sex the smallest, .0584 (r_{Y1} = -.242). Number of publications and number of citations are intermediate and about equal in this regard: r^2_{Y3} = .2144 and r^2_{Y4} = .2368. Unlike the next two correlation coefficients to be discussed, each of these values is in no way dependent upon the relationship of X_i and Y with any other variables.

Table 2. Squared Correlation Coefficients of X_i with Y: Whole, Semipartial and Partial

		r^2_{Yi}	sr^2_i	pr^2_i
X_1	Sex	.0584*	.0345*	.0608*
X_2	Years since Ph.D.	.3749**	.0531**	.0906**
X_3	No. of publications	.2144**	.0112	.0207
X_4	No. of citations	.2368**	.0609**	.1026**

$R^2_{Y\cdot1234}$ = .4671**

*P < .05
**P < .01

The squared semipartial correlation: sr^2_i. While r^2_{Yi} gives the proportion of Y variance accounted for by X_i, sr^2_i (the squared semipartial correlation) gives the proportion of Y variance accounted for by that part of the predictor X_i which is unique to X_i, i.e., the part of X_i which it does not share with the other predictors. Accordingly, it is the amount by which the multiple correlation R^2 would be reduced if X_i were omitted from the analysis and only the remaining I.V.s were used,

$$sr^2_i = R^2_{Y\cdot12..i..k} - R^2_{Y\cdot12..(i)..k}, \qquad (4)$$

where (i) symbolizes the omission of the given X_i. Conversely, of course, it follows that sr^2_i is the amount by which R^2 increases when X_i is added to a group of other specified variables. Clearly, then, as was the case for B_i and β_i, sr^2_i depends on what other I.V.s there are in the system.

The term "usefulness" has been used for sr^2_i and some computer programs designate it as the "unique" contribution of a predictor to a criterion, a term we prefer because it at least implies the context of other predictors which define it. The term "part" correlation and the notation $r_{Y(i\cdot12...k)}$ is also frequently used for sr_i.

The meaning of this is clarified by the illustrative problem. For example, the largest sr^2 for the four predictors is for the number of citations, sr^2_4 = .0609 (Table 2). The number of

literature citations (X_4), accounts uniquely (given the presence of the other predictors X_1, X_2, X_3) for 6% of the salary variance. Thus, were X_4 to be dropped, we would find that $R^2_{Y\cdot123}$ would be $R^2_{Y\cdot1234} - sr^2_4 = .4671 - .0609 = .4062$. Conversely, the addition of X_4 to the other three variables would increase R^2 by .0609 (from .4062 to .4671). On the other hand, dropping number of publications (X_3) from the set would reduce R^2 by only .0112 ($=sr^2_3$), an amount which is not significantly different from zero (see below). Apparently, almost all of the variation between faculty in number of publications which is relevant to salary ($r^2_{Y3} = .2144$) is not unique to itself--other variables share it, so that when they are partialled from X_3, the remainder accounts for only 1% of the salary variance. We can see from Table 1 why this sharp drop occurs from r^2_{Y3} to sr^2_3, for X_3 is largely redundant with X_2 (i.e., $r_{23} = .683$). This redundancy also accounts for why the highly salary-relevant years since Ph.D. ($r^2_{Y2} = .3749$) accounts uniquely for only .0531 ($=sr^2_2$) of the salary variance (but note also that $r_{24} = .460$, another source of redundancy in X_2).

In Table 2 all the squared semipartial values (sr^2_i) are small compared to their squared product-moment values (r^2_{Yi}). This circumstance is frequently encountered in social science data and reflects their tendency to be at least partially redundant. It is not, however, necessarily the case that sr^2_i must be smaller than r^2_{Yi}. One occasionally encounters an X_i whose sr^2_i exceeds its r^2_{Yi}. This phenomenon, called "suppression," can be understood as resulting from one or more other predictors removing ("suppressing") by partialling a portion of the variance of X_i which is irrelevant to the criterion. Thus partialled, the remaining X_i is more highly related to the criterion than X_i taken as a whole is; hence sr^2_i is greater than r^2_{Yi}.

The temptation to add sr^2 values of mutually correlated predictors must be resisted. They do not sum to R^2, nor does the sum of sr^2_i and sr^2_j give the amount by which R^2 would drop if X_i and X_j were simultaneously omitted.

The squared partial correlation: pr^2_i. The squared partial correlation coefficient (pr^2_i) of predictor X_i with Y estimates what the squared product-moment (r^2_{Yi}) would be for any subset of cases, all of whom have the same values on the other predictors. Whereas sr^2_i gives X_i's unique contribution as a proportion of all the criterion variance, pr^2_i gives it as a proportion of that part of the variance which is not related to the other predictors. Thus, sr_i is a semipartial since the other variables are partialled only from X_i, and pr_i is a (full) partial since the other variables are partialled from both the independent and dependent variables. They are thus related by

$$pr^2_i = \frac{sr^2_i}{1 - R^2_{Y\cdot12..(i)..k}} = \frac{sr^2_i}{1-(R^2_{Y\cdot12...k}-sr^2_i)}, \qquad (5)$$

the denominator literally being the proportion of Y variance not accounted for by the I.V.s other than X_i. A frequently used alternative notation for pr_i is $r_{Yi\cdot12...k}$, with everything following the dot understood as being partialled from both Y and X_i.

Because of its central importance in the application of MRC, we return to the core idea of pr^2_i, that it is the expected value for r^2_{Yi} for subsets of cases all of whom share the same values on the other variables. This is the sense in which we say that these other variables are "held constant statistically" or "statistically controlled" so that we can estimate the relationship of a predictor to the criterion uninfluenced by their relationship to other variables. Although the logical purity of a controlled manipulative experiment performed on randomly assigned subjects is sometimes available as a research method to the social scientist, more often he or she must observe phenomena as they exist, subject to variation and covariation due to extraneous and uncontrollable factors. Under these circumstances, all that is possible is the statistical control of such extraneous factors by the partialling process.

For example, in the illustrative problem, it was found that number of citations (X_4) accounted for .2368 ($=r^2_{Y4}$) of the salary variance in the total sample. Since $pr^2_4 = .1026$, however, we can estimate that for any subgroup with the same values for the other predictors, for example, males 12 years after their Ph.D. with nine publications, only about 10% of the salary variance is accounted for by number of citations. Since this value holds for this (or any other) subgroup which does not vary in these other regards, the 10% figure can be attributed to them, while the .2368 value for the total sample inevitably reflects, in part, the fact that faculty with more citations are inevitably older and have more publications.

That pr_i^2 need not be smaller than r_{Yi}^2 is seen from the relationship of sex to salary. For the whole sample, sex accounts for (r_{Y1}^2=) .0584 of the salary variance, while for a subsample with the same years since Ph.D., number of publications and number of citations, sex accounts for .0608 (=pr_1^2) of the salary variance. Lest this quantity seem small, recall that measured in dollars, it came to a sex salary differential of -$3,266(=$B_1$).

The Simultaneous and Hierarchical Strategies

In the illustrative problem above, all four independent variables were simultaneously regressed on and correlated with the dependent variable. One result of so proceeding was that for each predictor, all the others are partialled in the determination of partial regression and correlation coefficients.

An alternative strategy enters each predictor successively in a predefined order, and determines for that hierarchical order how much each adds to the prior R^2. The order selected is determined by assumptions of causal priority, or by centrality of research interest, or by certain structural properties of the predictors (Cohen and Cohen, 1975, pp. 98-102). The hierarchical strategy may be expressed as follows:

$$R^2_{Y\cdot123\ldots k} = r^2_{Y1} + sr^2_{2\cdot1} + sr^2_{3\cdot12} + \ldots + sr^2_{k\cdot12\ldots k-1}, \qquad (6)$$

where the predictors are numbered in their order of entry. The terms of this equation are squared semipartial correlations with Y, but only those predictors which have entered earlier (higher in the hierarchy) are partialled at each stage; not all the others as in the simultaneous model. In the illustrative example, the predictors are numbered in order of presumed causal priority: sex (X_1) is temporally prior to the others, length of career (X_2) is a necessary precondition for publications (X_3), which is in turn a necessary precondition for citations (X_4). Table 3 shows the derivation of the values necessary for the hierarchical analysis of Eq. (6). Each row of Table 3 represents a (simultaneous) MRC for the variables at each stage, showing the R^2 at the

Table 3. A Hierarchical Analysis of the Academic Salary Example

	Predictors	R^2 (cumulative)	Increment
Sex	X_1	$.0584 = R^2_{Y\cdot1} = r^2_{Y1}$	$.0584 = R^2_{Y\cdot1} - 0$
+Yrs. since Ph.D.	X_1, X_2	$.3967 = R^2_{Y\cdot12}$	$.3383 = R^2_{Y\cdot12} - R^2_{Y\cdot1} = sr^2_{2\cdot1}$
+No. Publications	X_1, X_2, X_3	$.4062 = R^2_{Y\cdot123}$	$.0095 = R^2_{Y\cdot123} - R^2_{Y\cdot12} = sr^2_{3\cdot12}$
+No. Citations	X_1, X_2, X_3, X_4	$.4671 = R^2_{Y\cdot1234}$	$.0609 = R^2_{Y\cdot1234} - R^2_{Y\cdot123} = sr^2_{4\cdot123}$
*P < .05 **P < .01	$R^2_{Y\cdot1234} = .0584* + .3383** + .0095 + .0609** = .4671**$		

stage, and the increment in R^2 over the previous stage due to the addition of a new variable. The increments are proportions of criterion variance added by the inclusion of the new variable, i.e., sr^2 values that partial (only) prior variables. Thus, sex, with nothing partialled, accounts for .0584 of the Y variance. When career length (X_2) is added, a total of .3967 of the Y variance is accounted for; the increment due to X_2 is thus .3383, which is identically the proportion of criterion variance accounted for by X_2 from which X_1 has been partialled (hence, $sr^2_{2\cdot1}$), etc. Years since Ph.D. (X_3) is of preeminent importance, its increment in Y variance being .3383 in this hierarchy. Obviously, the increment due to a predictor depends on the hierarchy that specifies at any stage which other predictors have already been partialled from the criterion. Were length of career (X_2) last in the hierarchy, its increment would be .0531, the sr^2 of the simultaneous analysis, hence $sr^2_{2\cdot134}$ (Table 3). Thus, although the hierarchical model provides an additive partitioning of the total Y variance accounted for by the k predictors in Eq. (6), the proportion which it attaches to each predictor is order-dependent. Since a different order would yield different values, it is clearly important that there be a defensible rationale for the order chosen. Some well publicized methods utilize a computer-defined hierarchical strategy, where variables are entered into prediction in a sequence according to how "important" they are to

predicting the criterion (measured, for example, by sr^2). These "stepwise" regression methods can be used to select a small subset of predictors from a larger set.[2]

The hierarchical strategy is the MRC method of choice in the analysis of the data of surveys and quasi experiments, and in the analysis of covariance and its generalization.

THE REPRESENTATION OF INFORMATION AS SETS OF PREDICTORS

As the preceding material illustrates, conventional MRC focuses its interpretation on single predictors (I.V.s), each a research factor. For reasons of structure, function or content, however, the representation of a single research factor of interest to the social scientist may well require multiple predictors. In fact, virtually any information in any form may be represented as a set of one or more predictors, and a set of predictors may be treated much as single predictors were in the previous section. We will see that these methods of representation bring into the MRC system: group membership (nominal scale) information, nonlinear relationships, variables with missing data, and interactive information. Also, by using sets which function as control variables, one can greatly increase the scope and relevance of data analysis.

Group Membership or Nominal Scales

Such research factors as diagnosis, type of drug abuse treatment group, place of birth, marital status, ethnic group and sex provide qualitative information by assigning the cases to one of g categories which are mutually exclusive and exhaustive. It is possible by any one of several coding methods to fully represent any such research factor G but it requires a set of g - 1 = k predictors (Cohen and Cohen, 1975, chapter 5). Table 4 illustrates one of the methods, effects coding, by a coding diagram for the representation of ethnic group membership (G) in one of the four groups: White, Black, Hispanic, and Other. The diagram indicates, for example, that all cases in the Hispanic category are coded (given values of) 0 on predictor X_1, 0 on X_2, and 1 on X_3. These artificial values are treated just like any other predictor values would be in the ensuing MRC analysis. Note that it takes only 3(=g-1) predictors to represent the 4(=g) ethnic groups. The order of the columns (as well as the groups) is quite arbitrary and does not matter, since X_1, X_2, and X_3 are treated simultaneously as a set G which completely carries the information as to ethnicity. Now assume that the criterion studied is length of prison sentence. One can then determine the square multiple correlation $R^2_{Y.123}$ (i.e., $R^2_{Y.G}$) as the proportion of Y variance accounted for by ethnicity (which is, incidentally, identically the squared correlation ratio which would be determined from an analysis of variance of these data).[3] Moreover, this set G can be combined with sets of predictors representing other research factors (F, H, etc.) to determine, among other things, the unique contribution of ethnicity, or the contribution of other factors holding ethnicity constant.

Table 4. Diagram for Effects Coding of Ethnicity (G: X_1, X_2, X_3)

		X_1	X_2	X_3
G_1	White	1	0	0
G_2	Black	0	1	0
G_3	Hispanic	0	0	1
G_4	Other	-1	-1	-1

In the analysis which produces $R^2_{Y.G}$, the partial coefficients for the individual effects-coded X_i have interpretive utility. For example, A equals the (unweighted) mean of the four groups' mean length of sentence (Y), and B_2 equals the Blacks' mean minus the mean of the groups' means, i.e., the "effect" of membership of G_2 (as the term is used in the analysis of variance). As another example, sr^2_2 is the proportion of the criterion variance accounted for by the Blacks' "effect", i.e., the amount by which R^2 would drop if the Blacks' mean prison sentence fell at the mean of the other three groups' means. The sr^2 (or sr) value thus provides a unit-free measure of the departure of one group relative to others with regard to the criterion, and therefore may be used to compare this departure among different criteria.

Effects coding is only one of several methods of expressing nominal scales; others include dummy, contrast, and nonsense coding, i.e., different patterns of artificial values from those of Table 4. The fact must be stressed that whichever coding method is used for a set of data, the $P^2_{Y \cdot G}$ value (as well as values for semipartial or partial R^2 involving G) will remain constant. In other words, however coded, the set of predictor values carry the group membership information. What changes is the meaning of the individual predictors, which, when partialled, carry different comparisons among group means or combinations of group means. The availability of alternate coding methods imparts a high degree of analytic flexibility and relevance. The analyst chooses the method whose single predictors carry the aspects of group membership (hypotheses, foci) which interest him while being assured that the set taken as a whole represents the research factor G.

Quantitative Scales and Curvilinear Relationships

Relationships for some variables, and particularly those which involve time, money, counts in general (e.g., of errors), and proportions, are frequently not well described by a straight line. When this is known (or suspected) to be the case, it is nevertheless possible to accomodate to and describe such curvilinearity using MRC. Here, too, several general and some specialized methods are available (Cohen and Cohen, 1975, chapter 6), of which we will here describe that of power polynomials.

It is fortunately the case that the relationship between the criterion (Y) and a nonlinear predictor (v) of any shape, no matter how complex, can be perfectly represented by an equation of the form,

$$Y = C + Dv + Ev^2 + Fv^3 + Gv^4 + \text{etc.} \qquad (7)$$

(where C, D, E etc. are constants), provided one goes out far enough. It is an even happier circumstance that in the social sciences, the relationships encountered are, with rare exceptions, well described by equations in the first two or three powers of the nonlinear predictor v. Thus, if we now let $v = X_1$, $v^2 = X_2$, and $v^3 = X_3$, Eq. (7) is in the form of the multiple linear regression equation, Eq. (1), the constants now being the B's and A. We have thereby used the multiplicity of MRC to represent a curvilinear relationship by means of a set V made up of linear (v), quadratic (v^2), and cubic (v^3) aspects (functions). For example, if in the academic salary problem we were concerned with the possibility that the relationship with number of publications (v) was not straight-line, we could represent it as a set V made up of three predictors, v, v^2, and v^3. Now using these three aspects of the set V, we might find $R^2_{Y \cdot V} = .2889^{**}$, in contrast with .2144 when only its linear aspect was used (see Table 2). The statement "number of publications accounts for 29% of the salary variance" is no longer qualified by the term "linearly"--whatever the shape of the relationship (within broad limits), it is likely to be captured by the set V.

In addition to being "covered" against the possibility of curvilinearity, power polynomials make possible an analysis of the shape of the relationship. One proceeds hierarchically, as was done in Eq. (6) and Table 3 above. However, here the necessity to proceed hierarchically arises from the fact that v^2 is not a pure measure of the nonlinear quadratic (parabolic) aspect of V; v^2 is correlated (usually highly so) with v, which therefore must be partialled from it. Similarly v^3 must have both v and v^2 partialled from it to measure purely the nonlinear cubic aspect. This is readily accomplished by determining the R^2 values cumulatively: first for v, then for v and v^2 combined, and then for v, v^2, and v^3. The increments are then found (as in Table 3), and the hierarchical partitioning equation (6) is produced:

$$R^2_{Y \cdot 123} = r^2_{Y1} + sr^2_{2 \cdot 1} + sr^2_{3 \cdot 12}$$

$$.2889^{**} = .2144^{**} + .0621^{**} + .0124$$

This is interpreted to mean that in addition to the fact that the linear component of number of publications accounts for about 21% of the salary variance, allowing for a (parabolic) bend in the line increases by (a significant) 6% the salary variance accounted for, while the further curvilinear complexity provided by the cubic term does not provide a significant further increase. Given this result, the analyst may then use the regression equation in v and v^2 to describe and plot the best fitting curve to the data.

Other methods for coping with curvilinear relationships include orthogonal polynomial coding, and breaking the set into intervals which are then treated like groups ("nominalization"). Some special circumstances invite representing the set V as a single variable by subjecting the predictor v to a rescaling by means of a suitably chosen monotonic nonlinear transformation (e.g., log v, $\sqrt{v}$, 1/v), and then using the rescaled variable in the analysis.

Missing Data

Although various methods are available for coping with the problem of missing data (omitting cases, omitting variables, estimating missing values), one which is both simple and effective proceeds by conceiving "missingness" as an aspect of the research factor (V) in question and incorporating it as a predictor in the set which represents that research factor V (Cohen and Cohen, 1975, chapter 7). This is accomplished by first creating a "missing data dummy variable," X_1, which is coded 1 for those cases where information on the research factor V is missing, and 0 where it is present. One then represents the other aspect(s) of V as X_2, X_3, etc., filling the blanks where no values are available on a predictor with the mean of the values which are present for the other cases on that predictor. We emphasize the fact that these means are in no way intended to be estimates of the missing values, but are merely a device to get on with the analysis. This gadget may be used either with quantitative or nominal research factors.

For concreteness, assume that the data for the academic salary study were obtained, in part, by a mailed questionnaire, and that in 24 of the 100 responses, the item requiring that the respondent list his or her publications was omitted. The research factor "number of publications" (V) would then be a set which includes the missing data dichotomy X_1, and, as X_2, the number of publications for the 76 cases where it is available. For the other 24 cases the blanks are plugged with the mean number of publications of the 76 cases. X_2 is the linear aspect of V. (If nonlinear aspects are to be represented by power polynomials, the values which are present are simply squared and cubed, their blanks being plugged with their respective means and constitute X_3 and X_4.) Several features of this procedure are worth noting.

First, the squared product-moment r^2_{Y1} gives the proportion of Y variance associated with the "missingness" of V. Only if r^2_{Y1} is nonsignificant is it reasonable to suppose that values are missing randomly relative to Y. In the example one might well find r_{Y1} to be negative, nontrivial and significant, indicating that omission of publications is associated with lower salaries (possibly because omission is more likely to occur with few publications). If the analyst had simply dropped those 24 cases, not only would the sample size (and therefore the statistical power) be reduced and the information on the other variables lost to the analysis, but the remaining 76 cases would no longer be representative of the target population.

Second, plugging with the mean has the effect of making a group of predictors so treated (X_j) correlate zero with the missing data dichotomy X_1. This in turn results in r^2_{Y1} being additive with the proportion of criterion variance due to the plugged variable(s), the latter carrying the information of the values which are present. Further, the B_j values and A are unaffected by the plugging--they are exactly the same and receive the same interpretation as they would were the cases with missing v omitted.

Third, this method may be used together with any of the methods of representing either quantitative or nominal scales. The resulting set V containing its missing data dichotomy can then be treated in MRC together with other sets, much as single predictors were above.

Finally, when the proportion of missing cases is small, the blanks may be plugged with means, but X_1 should generally be omitted since its inclusion would adversely affect the statistical power of the significance tests with no compensating gain. The results for the treated group of predictors X_j are interpreted normally.

Interaction Sets

An interaction between two research factors, for example, U and V, carries information about the conditionality of their relationship to the criterion (Y): the relationship of U to Y is conditioned by (varies with, is a function of) the specific characteristics of V. That is, for different standings on V, the Y-U regression differs. (The relationship is symmetrical--when the above holds, it must also hold with U and V interchanged.)

Many of the relationships studied in the behavioral sciences are not invariant over changes in other (contextual, moderator, conditioning) variables. For example, although one can determine

the overall relationship between income (Y) and education (U), it may well be found to differ (in degree, sign, or shape) as a function of race (V) or of age (W), in which case, an education by race (U x V) or education by age (U x W) interaction with regard to income is said to exist. In a quasi-experimental comparison of the rated success (Y) of three rehabilitation techniques (V), where it is desirable to control for a set of demographic characteristics, i.e., a covariate set (U), the finding that demographics relate differently in the three groups, a U x V interaction, invalidates the analysis of covariance attempt to statistically equate the groups on the demographic characteristics.

Note that U and V are represented as research factor sets, each of one or more predictors, and each either quantitative or nominal.[4] The interaction between sets U and V is contained in a set generated by multiplying each of the k_U predictors of set U by each of the k_V predictors of set V. The resulting UV product set of $k_U k_V$ predictors, after U and V are linearly partialled is the U x V interaction, i.e.,

$$U \times V = UV \cdot U,V. \tag{8}$$

The partialling is accomplished by using the hierarchical model with the three sets U, V, and UV. One finds the proportion of Y variance due to U x V from $R^2_{Y \cdot U,V,UV} - R^2_{Y \cdot U,V}$ (a squared multiple semipartial correlation) and can test its significance (see below). Further, each constituent predictor of the interaction set can be understood as a given aspect of U by aspect of V interaction, e.g., as a difference in slope or in curvature of the Y-U relationship between G_1 and G_4, and its B, sr, and pr interpreted. There is virtually no limit to the analytic specificity possible in the study of the conditionality of relationships via MRC. Also, the method illustrated for a two-way (U x V) interaction directly generalizes to interactions of any order.

SETS AS UNITS OF ANALYSIS

A case can readily be made for the proposition that the fundamental unit of analysis in MRC is a set of predictors which represents a research factor or a functional group of subsets of research factors. That any information may be represented as a set, and that sets may be combined, treated in tandem, and partialled from each other has already been suggested. In fact, the various measures attached to, and operations performed in, conventional MRC with single predictors have their analogues for sets of predictors; conversely, analysis with single predictors may be viewed as the special case where each set has only one predictor.

Consider the various measures of correlation and proportion of Y variance as discussed. If we replace the k single variables by h sets of variable A, B, C,...H (made up respectively of k_A, k_B, k_C, etc. single variables), we can define for any given set B, whatever its nature, its whole, semipartial or partial correlation with the criterion, and their squares as proportions of criterion variance. Set B's whole correlation with Y (analogous to r_{Yi}) is the multiple correlation $R_{Y \cdot B}$ (with k_B predictors), and $R^2_{Y \cdot B}$ is the proportion of Y variance accounted for by set B. If we wish to ascertain the proportion of total criterion variance accounted for by B over and above what is accounted for by another set A (however we choose to specify A), i.e., set B's unique Y variance, we determine the squared multiple semipartial correlation, $R^2_{Y \cdot (B \cdot A)}$ or sR^2_B (analogous to sr^2_i of Eq. 4), from

$$sR^2_B = R^2_{Y \cdot A,B} - R^2_{Y \cdot A}, \tag{9}$$

where A, B indicates the combination of the two sets of variables. Finally, parallel with pr^2_i of Eq. (5), if we wish to express set B's unique variance as a proportion of that part of Y's variance not accounted for by set A, we can determine the squared multiple partial correlation, $R^2_{YB \cdot A}$ or pR^2_B, from

$$pR^2_B = \frac{R^2_{Y \cdot A,B} - R^2_{Y \cdot A}}{1 - R^2_{Y \cdot A}} = \frac{sR^2_B}{1 - R^2_{Y \cdot A}} \tag{10}$$

Keep in mind that the analyst is completely free to define the content of sets A and B as he chooses. Either may be made up of one or more research factor sets, since the combination of sets is itself a set. In particular, set A may be made up of whatever research factors we wish to partial (control or hold constant statistically) in studying the Y-B relationship. In simultaneous setwise MRC, each research factor may in turn be designated set B and the remaining factors collectively set A. As another example, to measure an education (U) by race (V) interaction, U x V, let set B be the UV product set and set A made up of the U and V sets combined; then determine the sR^2_B or pR^2_B. The notion of hierarchical MRC also readily generalizes to sets. Replacing single variable by set designations, Eq. (6) becomes

$$R^2_{Y\cdot C,D,\ldots G,H} = R^2_{Y\cdot C} + sR^2_{D\cdot C} + sR^2_{E\cdot C,D} + \ldots + sR^2_{H\cdot C,D,\ldots G} \qquad (11)$$

The values are increments in R^2 as each set is added successively as with single I.V.s, in a predetermined order. Thus, one first finds the cumulative

$$R^2_{Y\cdot C},\ R^2_{Y\cdot C,D},\ R^2_{Y\cdot C,D,E,\ldots},\ R^2_{Y\cdot C,D,\ E,\ldots H},$$

and then the increments for Eq. (11) by subtraction, e.g.,

$$sR^2_{D\cdot C} = R^2_{Y\cdot C,D} - R^2_{Y\cdot C},\ sR^2_{E\cdot C,D} = R^2_{Y\cdot C,D,E} - R^2_{Y\cdot C,D},\ \text{etc.}$$

It is unnecessary to reiterate the analytic-interpretive possibilities made available by the use of research factor sets of predictors and their combinations, since they generalize from those described and exemplified for single variables. It may be apparent to the reader that, given our ability to represent any kind of information as sets of predictors, including aspects of group membership, curvilinear as well as straight-line relationships, the incorporation of missing data as positive information, and the representation of conditional relationships, and further, given the possibility of holding constant by partialling any research factors from any others in appraising relationships to a dependent variable, general MRC analysis offers a uniquely powerful device for the exploitation of research data along whatever line is defined by the logic of the research and the purposes of the investigator.

SIGNIFICANCE TESTING AND POWER ANALYSIS

Testing for Statistical Significance

This topic has been delayed to this point because, after the concept of partialling a set A from a set B is understood, significance testing for any of the statistics presented above can be presented most compactly and simply, indeed using a single very general F test.[5]

The formal assumptions underlying the (fixed model) F test are that the entities (e.g., subjects) are independently and randomly sampled, that subsets of entities in the population sharing the same set of values on the I.V.s have a normal distribution of Y values, and that the Y variances from subset to subset are equal. It is well known, however, that F (and t) tests are "robust," i.e., will tolerate a considerable amount of departure from the distribution assumptions without materially affecting their validity. This is particularly true when n is large, which is desirable on the even more important grounds of affording adequate statistical power (see below). The sampling assumptions (randomness, independence) must, however, be satisfied.

The null hypothesis under test throughout is that the population parameter value of the observed sample statistic equals zero, generally; that population $sR^2_B = R^2_{Y\cdot A,B} - R^2_{Y\cdot A} = 0$; or that set B accounts for no criterion variance beyond what is accounted for by set A in the population. The general formula applied to the values determined from the sample is

$$F = \frac{R^2_{Y\cdot A,B} - R^2_{Y\cdot A}}{1 - R^2_{Y\cdot A,B}} \times \frac{n - k_A - k_B - 1}{k_B}, \qquad (12)$$

with degrees of freedom (df) = k_B for the source (numerator) and $n - k_A - k_B - 1$ for error (denominator); n is the total sample size, and k_A and k_B are respectively the numbers of predictors in sets A and B. The computed F value is compared with the criterion value in standard F tables.[6] Note that although an F test formula can be written for pR^2_B, it is unnecessary since it must yield the same result as Eq. (12)--if sR^2_B is significant, so is pR^2_B, and to identically the same degree.

To specialize Eq. (12) for the test on the significance of an R^2, let set A be empty; then, nothing is being partialled from set B. Specialized equations exist to test the significance of an R^2, for the unique contribution to a criterion of a single predictor and for the sign of a simple (zero-order) r between two variables. The significance test for a single predictor may be equivalently performed as a t-test. Such tests of significance are typically available as output from a computer program.

We have suggested that while the set is an optimal unit for analysis, the predictors which constitute a set are individually interpretable as "aspects" of the research factor(s) which the set represents, and their coefficients can be tested for significance by an F (or $\sqrt{F} = t$) test. However, as the total number of predictors in a research increases, when all are tested for significance, the risk of one or more spuriously significant results (technically, the family-

wise or investigationwise Type I error rate) increases rapidly. To guard against this, it is recommended that, with the analysis organized into sets, only those predictors in sets which have made a significant contribution are tested for significance. This two-stage requirement for asserting the significance of the unique contribution of a single predictor is called the "protected t (or F) test" because the requirement that the set be significant protects the investigator from an unacceptably high risk of drawing spurious positive conclusions. At the same time, the protected t-test does not sacrifice statistical power (Cohen and Cohen, 1975, pp. 162-165).

Statistical Power Analysis

The statistical power of a significance test is the probability that it will reject the null hypothesis. For the tests in MRC, this will depend on the "effect size" (a function of the proportion of variance accounted for in the population) n, k_A, and k_B. When values for these are specified, it is possible to determine the power of the test. Since they can be estimated (effect size) or specified (the other parameters) before an investigation is undertaken, that is the optimal time to do so. If power is found to be low, one should want to reconsider one's plans. For example, all things equal, increasing the sample size (n) will increase power. As a practical matter, among those things under the investigator's control, it is the sample size which has the most important bearing on power.

By relatively simple methods, for the general F test of Eq. (12), and hence for its special cases, for any given population effect size, significance criterion, and number of variables in sets B and A, one can determine power as a function of the given sample size, or determine the necessary sample size for a desired amount of power.

Although the methods of power analysis in MRC are simple, their detailed exposition and special tables require more space than this short treatment of MRC justifies (see Cohen and Cohen, 1975, pp. 144-155). A rough rule of thumb has been suggested that for typical social science applications of MRC, adequate power (defined conventionally at .80) requires that the sample size be at least 25 times as large as the number of predictors to be employed. But such a guide is far too rough--the investigator is amply repaid for the modest effort required to do exact power analyses for the central issues of the investigation, and on the other facilitate the interpretation of the investigation's results.

ILLUSTRATIVE APPLICATION IN DRUG RESEARCH

RELATING DRUG USE TO ATTITUDINAL AND BEHAVIORAL MEASURES

The use of multiple regression techniques to provide a focused but thorough analysis among psychological, social, and other factors in drug use is elegantly exemplified in a recent study by Kendall (1975). The material for this study was supplied by extensive interviews of 823 high school and college students regarding their alcohol and drug usage and a number of potentially related attitudinal and behavioral issues. The overall goal was to describe the context within which students tended to abuse alcohol and/or drugs, and to explore some possible consequences. We present only a fragment of the Kendall study here.

Drug use was measured on a scale potentially ranging from 0 (never used drugs), to a maximum of 33 (never actually reached) which indicated frequent use of seven types of drugs. Because the scale units were essentially arbitrary, the independent variables were best described in terms of their contributions to Y variance accounted for in specified hierarchical sequences and the direction and shape of significant relationships. Had the variables been measured on scales with meaningful, commonly understood units, the regression coefficients would probably have been of primary interest.

Since it was the intention of the researcher to produce generalizations to students, any qualification of the findings associated with sex or school level needed to be discovered or ruled out. Therefore, the variables which first entered the hierarchical sequence reflected sex and school level. The three variables coded as contrasts are shown as X_1, X_2 and X_3 in Table 5. X_1 compares high school students to college students, and X_2 compares females to males. Because

Table 5. Variable Sets in Kendall Study of Drug Use

Set A
- X_1 = School level: -1 = high school, +1 = college
- X_2 = Sex : -1 = female, +1 = male
- $X_3 = X_1 \times X_2$: -1 = high school males and college females; +1 = high school females and college males

Set B
- X_4 = Religious involvement
- $X_5 = X_4^2$ = Religious involvement squared

Set C
- $X_6 = X_1 \times X_4$: $-X_4$ for high school students, $+X_4$ for college students
- $X_7 = X_2 \times X_4$: $-X_4$ for females; $+X_4$ for males
- $X_8 = X_3 \times X_4$ etc.
- $X_9 = X_1 \times X_5$
- $X_{10} = X_2 \times X_5$
- $X_{11} = X_3 \times X_5$

Set D
- X_{12} = Drinking behavior
- $X_{13} = X_{12}^2$

Set E
- $X_{14} = X_4 \times X_{12}$
- $X_{15} = X_5 \times X_{12}$
- $X_{16} = X_4 \times X_{13}$
- $X_{17} = X_5 \times X_{13}$

Set F
- $X_{18} = X_1 \times X_{12}$
- $X_{19} = X_2 \times X_{12}$
- $X_{20} = X_3 \times X_{12}$
- $X_{21} = X_1 \times X_{13}$
- $X_{22} = X_2 \times X_{13}$
- $X_{23} = X_3 \times X_{13}$

there were different numbers of respondents in each of the four groups, these two variables were somewhat correlated with each other. X_3 is the sex and school status variable product, literally obtained by multiplying X_1 times X_2 for each subject. The net contribution of X_3 partialling X_1 and X_2, sr_3, carries the interaction effect information, that is, answers the question, "Is the sex difference in drug use the same for college students as it is for high school students?" or, equivalently, "Is the drug use difference between high school and college students the same for males and females?"

As anticipated, females had lower average drug use levels than did males, and high school students used drugs less than did college students. In addition, the sex x school interaction was significant--the difference between high school and college students was greater for males than it was for females. Since the purpose of including these three variables as set A was to partial sex-school level group effects from variables more directly of interest (so as to produce within-group correlations), no detailed findings will be presented. However, as shown in Table 6, this set accounted for 2% of the drug use variance, a small but statistically significant amount for 3 and 819 degrees of freedom (Eq. 12).

One of the contextual variables included in this study was religious involvement, a scale based on five items concerned with religious attitudes and activities. Since no a priori case could be made for a linear relationship between religious involvement and drug use, this scale score (X_4) and its square (X_5) were both included in set B. When X_4 was added to the equation, R^2 increased to .068; the negative relationship between religious involvement and drug use was significant and not wholly accounted for by sex and school level differences of students on both. When X_5 was added, a further significant increase in R^2 to .084 was found, and with positive partial coefficients (see Table 6). A negative linear relationship combined with a positive quadratic relationship indicates a generally downward slope which is concave upward. Specifically, this relationship, when plotted, showed the primary differences in drug use to be between those who had no religious involvement and those with some, albeit minimal involvement; no further decrease in drug use was found among those students with above average religious involvement.

Table 6. Drug Use Variance Accounted For By Independent Variable Sets

	R^2	Increment	Set added
Set A	.020	.020**	sex and school level
A,B	.084	.064**	religious involvement
A,B,C	.087	.003	A x B interaction
Set A	.020	.020**	sex and school level
A,D	.154	.134**	drinking
A,D,B	.188	.034**	religious involvement
A,D,B,E	.192	.004	D x B interaction

As was mentioned, and is well known in the context of analysis of covariance, an assumption necessary for the validity of conclusions based on net or partial relationship is that the variables being partialled or covaried (set A) do not interact with the variables being examined (set B) in their relationship to Y; i.e., that the Y-B regression is homogeneous with regard to A. As a check on this assumption, an additional set C was created from the product of the set A and set B variables. The resulting six variables shown in Table 6 increased R^2 only .003, a far from significant amount. Thus, the drug use-religious involvement regression can be presumed to be the same in all four sex-school groups. The homogeneity assumption was therefore considered warranted, and set C was dropped from further consideration.

Since drinking and drug abuse frequently occur among the same students, a critical aspect of the study was developed in analyses which partialled the effects of drinking (set D) from the relationship between drug use and various independent variables. In this way, variables which were uniquely related to drug use could be distinguished from variables whose drug use relationship may be attributable to their relationship with drinking. Drinking and drug use were correlated .39, and drinking added .134 to $R^2_{Y\ A}$; drinking squared added a nonsignificant .004; thus, $R^2_{Y\cdot A,D} = .154$. A check on possible interactions between drinking behavior and the set A variables was now appropriate. When a set containing the relevant product terms (F = A x D) was added $R^2_{Y\cdot A,D,F} = .162$ was not significantly greater than $R^2_{Y\cdot A,D}$ (=.154). Set F was accordingly dropped.

When religious involvement and its square (set B) were now added to sets A and D, $R^2_{Y\cdot A,D,B} = .188$, a significant increase over $R^2_{Y\cdot A,D}$. Thus, although some portion of the relationship between drug use and religious involvement was redundant with the relationship of drug use and drinking, a significant portion was unique--$R^2_{Y(B\cdot A)} = .064$ and $R^2_{Y(B\cdot AD)} = .034$.

Finally, it would be appropriate to check on the proposition that the relationship between drug use and religious involvement did not vary for students at different levels of drinking. If set E, which includes these interaction terms, did not make a significant contribution, the assumption could be considered warranted. Table 6 shows a trivial contribution for this set, thus the effect of religious involvement on drug use is not conditional on the level of drinking behavior.

From these findings as a whole, we may conclude that the relationship between religious involvement and drug use is not "spurious"; that is, it is not wholly redundant with the effects of sex, school level, or drinking.

LONGITUDINAL STUDY OF CLINIC RECORDS

An important area of drug research lies in the investigation of possible changes over time in the patient population or in some criterion of success in therapy or rehabilitation. Since changes in real life rarely come only one or two at a time, an understanding of how these changes relate to each other may be sought.

As a fictitious example, suppose the director of a large drug clinic notices an apparent decrease in the average length of clinic contact, which in that context is suggestive of decreased treatment effectiveness. Because both staff and treatment concepts have evolved somewhat over this same period, the question arises as to whether the decline is attributable to these changes. It is also known that therapy is not equally successful with all types of clients, and that relevant changes in the client group may have taken place over the same period of time. A study is therefore undertaken to determine whether changes in the patient population account for the decrease over time in effectiveness as represented by length of clinic contact (Y).

Several sets of variables are taken from the clinic records. (See Table 7.) The first of these (set A) is a demographic set--age, sex, marital status, education, and ethnicity. This set itself may be subsets--e.g., age and age^2; single, married, divorced or separated represented as a nominal scale. In addition, it is probably appropriate to consider some interactions, such as age x education, or sex x marital status on the hypothesis that education will not have the same effect for young clients that it does for older clients or that the difference between married and single men will not be the same as the difference between married and single women.

Table 7. Longitudinal Study of Drug Clinic Records.

Set A:	Demographic	Subset A_1: age Subset A_2: marital status Subset A_3: ethnicity Subset A_4: sex Subset A_5: selected interactions among subsets
Set B:	Drug usage	
Set C:	Employment and Occupation	Subset C_1: occupation: highest, most recent, missing Subset C_2: employment: history, current, missing
Set D:	Interactions between Sets	
Set F:	Time of intake	
Set G:	Interactions with Time	

A second set of variables (set B) reflects drug type, dosages and the duration of the addiction. If appropriate, further interactions between variables within this set or across sets may be incorporated.

A third set (C), representing employment and occupation level and history, may be added. Perhaps this information is missing for a good many clients. Excluding these patients might well bias the sample, as patients who have never held jobs are more likely to be missing this information. On the other hand, it is not safe to assume that missing data necessarily indicate no employment. Therefore, one or more dichotomies reflecting the absence of data is added, and the variate blanks are plugged with their means.

Yet another set D of variables, which investigates the interactions of other variables with variables in this set, may be required. For example, the effects of employment status may differ systematically with the age of the client. The effects of missing data may be different for women than it is for men, if, for instance, it was caused by failure to obtain this information from women, but indicated an absence of employment for men.

Finally, the set F, which represents the time variable, and in which the interest is centered, is added. This factor might be represented in any of several ways--e.g., by numbering the dates of intake consecutively, by coding the years of intake with orthogonal polynominals, or by taking linear, quadratic and possibly higher powers of ordinal year of intake.

As in the Kendall study, interactions between the set of time variables and other sets should be investigated. If interactions were found to be significant, it would be clear that the effects of client characteristics on the duration of contact were not constant over time, or equivalently, that changes over time in effectiveness were not the same for all kinds of clients.

Please note that the rather large number of predictors to be employed presumes a rather large sample size for this study--at the very least, 1,000.

CAUTIONS: ISSUES IN VARIABLE SELECTION AND WAYS OF COPING

Lest the foregoing discussion misleadingly suggest to the reader that the flexibility of MRC makes a proliferation of variables desirable, let us set the record straight. The mere possibility of inclusion of almost any kind of information does not make all such possibilities equally desirable. On the contrary, automatic exploration of all the possibilities comes at very high cost indeed, and the cost is of several different kinds. First, although it is possible to control the Type I error for each test of a partial coefficient via the chosen significance criterion,

the investigationwise Type 1 error probability increases as a function of the number of hypotheses tested. As the likelihood that one or more variables will be spuriously significant increases, there is a corresponding decrease in the confidence in any given, apparently significant effect.

Secondly, a proliferation of independent variables is also likely to lead to increases in Type II error, i.e., decrease in power. This is particularly true when multiple variables are included to cover a single construct--as, for example, when several variables are used as a set to reflect socioeconomic status. As was seen, statistical significance and power are both functions of the unique contribution of a variable. If the construct is best represented by what these variables share, their unique effects may be small even in the population. Testing the sR^2 of the set of variables may appear to overcome this problem, since it includes Y variance redundantly accounted for by the predictors in the set. However, for example, if five variables are used when two would do--the five accounting for trivially more variance in the population than do the two--there is an unnecessarily large numerator df and, therefore, lower power.

Finally, the larger the number of variables, the more difficulty may be anticipated in interpretation. This is especially true when variables are highly redundant, and when there is an absence of specific a priori hypothetical bases for possible findings. Fortunately, there are several ways of coping with these general problems.

First, distinguish between variables whose function it is to test the validity of assumptions and those representing real substantive hypotheses. If, as in Kendall's study, some interactions are only included as a test of the uniformity of a relationship (the validity of the covariance assumption), nonsignificant terms may be dropped from later analyses. Similarly, nonsignificant power polynomial functions of variables and missing data dichotomies may be dropped if their sole purpose was to check on linearity or the unbiasedness of cases with missing data.

Second, minimize the inclusion of redundant variables by creating scales or summary indices, by prior factor-analyzing sets of variables to reduce to one or a few relatively distinct dimensions, or by judiciously selecting a priori among the available variables.

Third, employ the hierarchical model to test variables represented by clear and warranted hypotheses early in the sequence and more speculative variables later. The findings on the later variables may be considered more frankly exploratory in nature and in need of future substantiation.

NOTES

[1]The asterisks attached to numerical values designate the results of testing for each value the null hypothesis that it is zero in the population. One asterisk indicates statistical significance at the .05 level, and two asterisks at the .01 level. The methods for performing these tests will be discussed later.

[2]The hierarchical procedure should not be confused with "stepwise" MRC analysis. Some stepwise programs offer the option of entering the predictors in an analyst-specified order ("forced stepwise"), and thus can be used for hierarchical MRC.

[3]In the interest of brevity, we do not point out all of the many parallels with the analysis of variance/covariance throughout this chapter. In fact, the latter is merely a special case of general MRC, which therefore can produce all its results, and more.

[4]The conventional standard analysis of covariance model deals with quantitative covariates and the effects of a nominal scale (group membership), but the MRC method has no such constraints. Since any kind of information may be represented as a set, and any set may be partialled from any other, the method sketched above is better called the Analysis of Partial Variance (Cohen and Cohen, 1975, chapters 8 and 9), of which the analysis of covariance is a limited special case.

[5]An even more general F test, which allows for the exclusion from error variance due to sources other than sets A and B, called Model II error, is omitted from this brief account. It is treated in Cohen and Cohen (1975, pp. 141-144).

[6]The .01 criterion should usually be preferred to the .05 criterion when, as is often the case in MRC analysis, many tests are to be performed. This is to prevent the investigationwise (experimentwise) type I error rate from becoming unduly large.

REFERENCES

COMPUTER PROGRAMS FOR MRC

Computer programs for carrying out multiple regression/correlation analyses are widely available as part of larger packages such as SPSS, BMD, Data-Text, Omnitab, Osiris and PSTAT. Each of these packages also includes some provision for transforming, creating, or recod'ng variables which may be required for tests of the effects of categorical variables, curvilinearity, interactions, and missing data. Generally, machine transformation is to be preferred to creation of "special" variables prior to data punching, for the simple reason that errors in raw data are more difficult to check than are program instructions, especially if the sample is large. Naturally, the actual computer instruction is a job which should be handled by an experienced person. An ingenious user of any of the packaged programs will be able to accomplish all of the goals of the MRC analysis, although some of the desirable output may not be directly available. For example, many programs do not provide sr or pr for each variable. However, virtually all programs provide a t (or $F = t^2$) test for the partial regression coefficient for each variable from which one can determine sr_i and pr_i:

$$sr_i = t_i \sqrt{\frac{1 - R^2_{Y \cdot 12\ldots k}}{n-k-1}}, \text{ and } pr_i = \frac{t_i}{\sqrt{t_i^2 + n-k-1}}$$

Available programs may differ with regard to accuracy as determined by the number of digits carried in the calculations. This is usually a problem only when high multiple correlations among independent variables are present, as when power polynomial or product terms are included. Some programs also have relatively sharp limits on the number of variables and/or cases which may be considered. A check with computer personnel on these issues will usually yield a program with the characteristics necessary for any MRC problem.

BIBLIOGRAPHY OF METHODOLOGY

Cohen, J. Multiple regression as a general data-analytic system. *Psychological Bulletin,* 70: 426-443, 1968.

Cohen, J., and Cohen, P. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975. Distributed by Halsted Press Division, John Wiley & Sons, New York.

Darlington, R.B. Multiple regression in psychological research and practice. *Psychological Bulletin,* 69:161-182, 1968.

Draper, N.R., and Smith, H. *Applied Regression Analysis.* New York: Wiley, 1966.

Kerlinger, F.N., and Pedhazur, E.J. *Multiple Regression in Behavioral Research.* New York: Holt, Rinehart and Winston, 1973.

Overall, J.E., and Spiegel, D.K. Concerning least squares analysis of experimental data. *Psychological Bulletin,* 72:311-322, 1969.

BIBLIOGRAPHY OF APPLICATIONS

With the exception of the Kendall study, none of the studies which follow are in the drug field. They are nevertheless offered as good and varied examples of the application of general MRC analysis in behavioral science, which illustrate approaches applicable to research in the psychosocial aspects of drug abuse.

CONTINUED

2 OF 3

Anderson, G.J. Effects of classroom social climate on individual learning. *American Educational Research Journal*, 7:135-152, 1970.

Austin, A.W. Undergraduate achievement and institutional excellence. *Science*, 161:661-668, 1968.

Cohen, J. Prognostic factors in functional psychosis: A study in multivariate methodology. *Transactions of the N.Y. Academy of Sciences*, 30:833-840, 1968.

Coleman, J.S., et al. *Equality of Educational Opportunity*. Washington, D.C.: Office of Education, U.S. Government Printing Office, 1966. ("The Coleman Report.")

Cronbach, L.F. Intelligence? Creativity? A parsimonious reinterpretation of the Wallach - Kogan data. *American Educational Research Journal*, 5:491-511, 1968.

Kendall, R.F. *The Context and Implications of Drinking and Drug Use Among High School and College Students*. Doctoral dissertation, Department of Psychology, New York University, 1975.

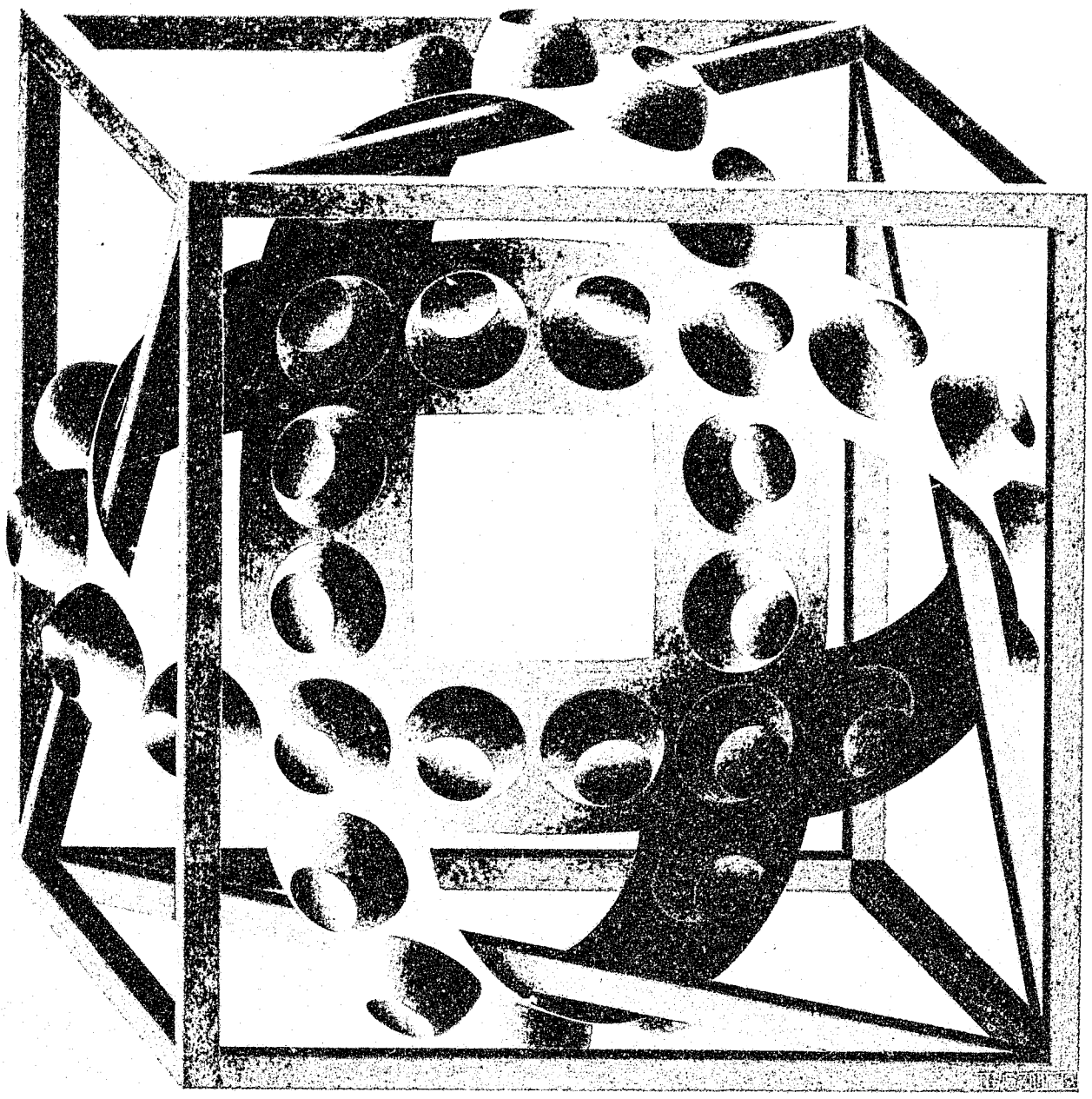


Plate 11

INTRODUCTION AND RATIONALE

ASSUMPTIONS AND ADVANTAGES

Univariate and multivariate analyses are basically methods for detecting and estimating, in sample data, differences between the means of populations. The populations may be naturally occurring and defined by attributes, or they may be created artificially by random assignment to treatments in an experiment. It is both a strength and a weakness of analysis of variance that it makes simplifying assumptions about the statistical structure of the data to be analyzed. It assumes first that the variables under investigation are measured on a continuum with a uniform unit of scale; it assumes further that the distributions of these measures in the populations differ only in the location of their central tendency and not in other aspects of shape such as dispersion, skewness, kurtosis, etc. For certain inferential purposes, it is in fact assumed that the distributions are normally distributed with unknown and possibly unequal means and unknown but equal variances.

The strength of these assumptions is that they focus on the aspect of the distribution that is likely to be most sensitive to conditions of environment or treatment to which biological material might be exposed. In most biological and behavioral studies, it is not possible to make observations under widely differing conditions without endangering the integrity of the organisms. As a result, most investigations are dealing with relatively small and essentially linear effects of the treatments or environments. These effects are expressed almost entirely in changes in the means of the distributions. By concentrating the inference on differences between means, the analysis of variance most effectively uses the information in the data to detect treatment or environment effects.

The focus on differences between means also makes sense from a practical point of view. In many applied studies, the economic value of the outcome of an experiment is described completely by the population mean. This was certainly true in the context of agricultural studies in which R.A. Fisher developed analysis of variance, for the economic value of a crop can always be computed from the price multiplied by the mean yield for the units multiplied by the number of units. Indeed, most of the commonly used indicators of social utility take the form of means (sometimes gratuitously so, as in the case of distributions, such as income, where the shape of the distribution may have as important an implication for policy as does the mean). The development of analysis of variance has been heavily influenced by this preoccupation with means in practical work (see the many examples in Cochran and Cox, 1957).

CAUTIONS

If analysis of variance has a weakness in biological and behavioral applications, it is the assumption that the variable of interest is measured continuously. Especially in medical and social experimentation, the outcomes of interest are often better described by success and failure than by a quantitative response measure. The statistical properties of these quantal or categorical variables demand a form of analysis which differs from analysis of variance in many ways. In the past, this has led to two major and distinct methodologies for data analysis--one based on least squares procedures for quantitative variables, and the other, based on chi-square procedures and applicable to qualitative data. Recently, these approaches have coalesced somewhat with the development of new methods of analysis, based on logistic or log-linear models, that are formally similar to univariate or multivariate analysis of variance (Bock, 1975; Bishop, Fienberg and Holland, 1975). But because these methodologies are based on different distributional assumptions, they remain distinct in the final analysis and are difficult to combine when studies involve both quantitative and qualitative outcomes. The present paper is therefore limited to data that can be considered quantitative (i.e., continuous) for purposes of analysis and can be represented on a continuum on which the units of measurement are well defined and everywhere comparable.

ANOVA VS. MANOVA

Like univariate analysis of variance (anova), multivariate analysis of variance (manova) focuses on means of continuously distributed variables; but unlike anova, does so jointly for more than one such variable. Manova is therefore especially suited to human behavioral studies, which typically involve a number of qualitatively distinct attributes or outcomes, and for which no single index of value may be calculated. In the multivariate approach, the several variables are analyzed simultaneously, and the investigator or reader may decide for himself the overall meaning or importance of various differences that may be found. Numerous examples of the application of multivariate analysis of variance to behavioral data may be found in Bock (1975).

METHODS AND PROCEDURES

A BRIEF REVIEW OF UNIVARIATE ANALYSIS OF VARIANCE

The main features of manova can most easily be understood by comparing it with univariate analysis of variance in the context of a typical application. Consider, for example, a clinical trial conducted in the form of a randomized experiment. For purposes of the trial, N patients in some diagnostic category are randomly assigned to n distinct courses of therapy. At the end of some period of trial, the patients are evaluated on multiple measures of outcome clinical-status (variates). For any one such measure, a univariate analysis of variance may be used to detect differences between the means of the treatments and to estimate the direction and magnitude of such differences. Formally, the analysis is based on the following model for the given variate:

$$y_{ij} = \mu + \alpha_j + \varepsilon_{ij}. \qquad (1)$$

The model states that the variate y_{ij}, representing the clinical status of subject i in treatment group j, is an additive combination of μ, the mean score before treatment; α_j, the effect of the j-th treatment; and a random error ε_j due to variation of the individuals within groups and to measurement error. For purposes of the least squares analysis--which treats (1) just like a regression equation--it is assumed that the error is independent from one subject to another, and that it is distributed with mean zero and constant variance σ^2 in all groups. For purposes of the subsequent statistical tests, it is assumed further that the error is normally distributed. In terms of this model, the investigation of differences between treatment means may be posed in the form of the null hypothesis H_0: $\alpha_1 = \alpha_2 = \ldots = \alpha_n$, stating that there are no differences among the n treatment effects. The test of this hypothesis under the assumptions of the model is carried out quite simply by the anova procedure. First the group means $y_{\cdot j}$ and the grand mean $y_{\cdot\cdot}$ are computed. From these and the original scores, the so-called analysis of variance table (Table 1) is prepared. Hypothesis H_0 is tested by the variance ratio

$$F_b = \frac{ssb/(n-1)}{ssw/(N-n)}$$

which, when H_0 is true, is distributed as a central F statistic with n-1 degrees of freedom in the numerator and N-n degrees of freedom in the denominator. It can be shown that this ratio increases monotonically as the sum of squared differences among the treatment effects increases. Hence, an observed value of the statistic more extreme than the $100(1-\gamma)$ percentile of the central F distribution is considered evidence of real differences among the treatment effects at the "γ-level of significance." Under the assumptions of the analysis, this test can be shown to be "uniformly most powerful," which means that for any value of the sum of squared differences, the probability of detecting a given departure from the null hypothesis is at least as great with this test as with any other test that might be proposed.

Significance of this F statistic is conventionally required if we wish to claim that we have established the direction of any of the treatment differences. Given their significance, we may then wish to go on to estimation of the treatment differences so as to determine their direction and magnitude. For this purpose, the analysis of variance provides the minimum-variance unbiased linear estimator of the treatment differences in the form of the differences between sample means

$$\widehat{\alpha_j - \alpha_{j'}} = y_{\cdot j} - y_{\cdot j'}. \qquad (2)$$

The sampling variance of this estimator is $\sigma^2\left(\frac{1}{N_j} + \frac{1}{N_{j'}}\right)$, which is the minimum attainable by any linear function of the data that unbiasedly estimates $\alpha_j - \alpha_{j'}$. An estimator of the unknown variance σ^2 in this expression can be obtained from the analysis of variance table by dividing the within-group sum of squares by its degrees of freedom:

$$\hat{\sigma}^2 = ssw/(N-n) .$$

From this estimate, the standard error of the estimated effect difference may be computed as

$$s_d = \hat{\sigma}\sqrt{\frac{N_{j'} + N_j}{N_j N_{j'}}}$$

and, thence, a confidence interval on the difference as

$$\widehat{\alpha_j - \alpha_{j'}} \pm t_{N-n}^{(\gamma)} s_d \quad , \tag{4}$$

where $t_{N-n}^{(\gamma)}$ is the 100 γ percent point of Student's t distribution for N-n degrees of freedom. This interval provides in one convenient form an expression of both our knowledge and our uncertainty about the true difference as inferred from the data. The probability is 1-2γ that (4) includes the unknown true difference being estimated.

TABLE 1

Analysis of Variance for the Simple Randomized Design

Source of Variation	Degrees of Freedom	Sum of Squares
Mean	1	$ssm = Ny_{\cdot}^2$
Between treatments	n-1	ssb = ssg - ssm
Treatment groups	n	$ssg = \sum_{j}^{n} N_j y_{\cdot j}^2$
Within groups	N-n	ssw = sst - ssg
Total	$N = \sum_{j=1}^{n} N_j$	$sst = \sum_{ji}^{nN} y_{ij}^2$

THE MULTIVARIATE GENERALIZATION OF ANOVA

In the multivariate extension of anova, a model such as (1) is specified for each of, say, p observed variables. Thus, the multivariate model may be expressed merely by indexing each term to denote the variable to which it applies. In (5) and thereafter, this index appears as the superscript "k" (in parentheses to distinguish it from an exponent), where k = 1,2,...,p.

$$y_{ij}^{(k)} = \mu^{(k)} + \alpha_j^{(k)} + \varepsilon_{ij}^{(k)} \tag{5}$$

For each of these terms, we may refer to an ordered set of variables specified by the range of k as a vector. Thus the left member of (5) is called a vector observation, the μ and α in the right member are called vector effects, and ε is called a vector error. In this model, the error distribution is specified by a multivariate normal distribution in which the vector mean has all zero components, and the errors have the same variances and covariances in all treatment groups. This specification is conventionally written $\underline{\varepsilon} \sim N(\underline{0},\Sigma)$, where $\underline{0}$ refers to the p x 1 vector mean and Σ, to the p x p variance-covariance matrix in which diagonal elements are variances and off-diagonal elements are covariances.

The computations of manova may be expressed compactly in terms of the superscripted components of the vector observations. Thus, the p x 1 vector mean of group j is

$$\left[y_{\cdot j}^{(k)}\right] = \frac{1}{N_j}\left[\sum_{i=1}^{N_j} y_{ij}^{(k)}\right], \text{ for } k=1,2,\ldots,p \tag{6}$$

and the p x 1 grand mean is

$$\left[y_{\cdot\cdot}^{(k)}\right] = \frac{1}{N}\left[\sum_{ji}^{nN_j} y_{ij}^{(k)}\right]. \tag{7}$$

In addition, p x p symmetric matrices of various sums of squares and cross-products of observations and of means are computed for all variates and pairs of variates. For example,

$$S_t = \left[\sum_{ji}^{nN_j} y_{ij}^{(k)} y_{ij}^{(\ell)}\right] \quad \begin{matrix} k=1,2,\ldots,p \\ \ell=1,2,\ldots,p \end{matrix},$$

which means that S_t is a matrix whose general (k-th row, ℓ-th column) element is the quantity standing inside the brackets.

In terms of these operations, the multivariate analysis of variance may be set up, as shown in Table 2, in line-by-line correspondence with the univariate analysis in Table 1. Table 2 differs only in the substitution of matrices of sums of squares and cross-products for the scalar sums of squares in Table 1. From these matrices, we obtain statistical tests of the multivariate null hypothesis H_ℓ: $\alpha_1^{(k)} = \alpha_2^{(k)} = \ldots \alpha_n^{(k)}$ for all k by computing a statistic sensitive to departure from equality of the vector effects.

TABLE 2

Multivariate Analysis of Variance for the Simple Randomized Design

Source of p-variate Variation	Degrees of Freedom	Sums of Squares and Cross-products (pxp) $\begin{matrix} k=1,2,\ldots,p \\ \ell=1,2,\ldots,p \end{matrix}$
Grand Mean	1	$S_m = [N y_{\cdot\cdot}^{(k)} y_{\cdot\cdot}^{(\ell)}]$
Between Groups	n-1	$S_b = S_g - S_m$
Group Means	n	$S_g = [\sum_j^n N_j y_{\cdot j}^{(k)} y_{\cdot j}^{(\ell)}]$
Within groups	N-n	$S_w = S_t - S_g$
Total	$N = \sum_{j=1}^{n} N_j$	$S_t = [\sum_{ji}^{nN_j} y_{ij}^{(k)} y_{ij}^{(\ell)}]$

A number of statistics have been proposed for this purpose, each with good properties from certain points of view. Their computation is simplified somewhat by the fact that each is a function of the so-called maximal invariant statistics given by the roots of the polynomial equation in λ obtained by expanding a determinant, defined in terms of two of the matrices in Table 2, and setting it equal to zero:

$$\left|S_b - \lambda S_w\right| = 0. \tag{8}$$

The alternative test statistics are computed from the $s = \min(n-1,p)$ non-zero roots of (8); i.e., s is the smaller of the two numbers n-1 and p. They are as follows:

Wilks' criterion $\quad \Lambda = \prod_{h=1}^{s} (1 + \lambda_h)^{-1}$

Roy's largest root criterion $\quad \theta = \lambda_1/(1 + \lambda_1)$

Hotelling's trace criterion $\quad \tau = (N-n) \sum_{h=1}^{s} \lambda_h$

C.R. Rao (1952) has shown that the distribution of a function of Wilks' criterion can be approximated with excellent accuracy by the F distribution with suitably defined degrees of freedom. Because percentage points of the F distribution are easy to approximate numerically, this criterion is perhaps the most widely used multivariate test statistic. But Roy's criterion has the advantage of slightly better power when departure from the null hypothesis is unidimensional (i.e., when the population group centroids are collinear). Because collinearity is common in behavioral applications, Roy's criterion deserves special attention. It also supports a system of confidence bounds that is sometimes used to judge the significance of all treatment contrasts and all variates simultaneously (see Bock, 1975).

Roy's criterion has been tabled for a wide range of arguments by Pillai, and a convenient form of Pillai's tables may be found in Bock (1975, Appendix A) with a discussion of the use of the statistics in tests of hypotheses and construction of confidence bounds. With respect to the latter, however, it must be pointed out that the Roy bounds tend to be overgenerous when the number of contrasts or number of variables is large. In these situations many workers prefer to employ so-called "protected" univariate statistics such as Fisher's F for testing individual variates, or Student's t for testing individual variates and contrasts, on condition that the overall multivariate test statistic is significant. This procedure has the experimentwise error rate of the multivariate test, and a separately specifiable variablewise error rate, but the comparisonwise error rate tends to be larger than its nominal value. If the latter is objectionable, the Tukey studentized-range test can be used for judging multiple differences. These protected F or studentized-range tests are useful in directing the investigator's attention to those variables and group differences that are most responsible for the significant multivariate effect.

In contrast to the multivariate interval estimates (confidence bounds), the multivariate point estimates are no more complex than in the univariate case. In fact, the multivariate estimator is given by the univariate estimator for each variable separately. The multiple univariate estimators are correlated, however, because the original observations are correlated, and hence their standard errors do not give a complete description of the sampling variability. The description must also include the correlations between estimators. The manova table provides a minimum-variance unbiased quadratic estimator of the error covariance matrix (9) from which these correlations may be estimated:

$$\hat{\Sigma} = \frac{1}{N-n} S_w \qquad (9)$$

Note that when there is just one group (n = 1), equation (9) represents the sample variances and covariances computed from deviations about the sample means. In that case, the corresponding correlation matrix, obtained by dividing the covariances by the standard deviations of the respective variables, contains the conventional sample correlations. When there is more than one group, however, the deviations are taken about the separate group means rather than the grand mean, and the correlation matrix then consists of what may be called the common within-group correlations. These correlations, and the corresponding standard deviations, provide an estimate of the across-samples association and variation among the response variables from sample to sample. The unique contribution of multivariate analysis of variance, as opposed to separate univariate analyses of variance of the data, is in the incorporation of correlations from this source into the statistical procedure.

STATISTICAL TECHNIQUES ALLIED TO MULTIVARIATE ANALYSIS OF VARIANCE

Allied to multivariate analysis of variance, and often included in the computer programs for the procedure, are the multivariate techniques of discriminant analysis, analysis of covariance, regression analysis and canonical correlation. These techniques are briefly described in this section.

Multiple-Group Discriminant Analysis

Discriminant analysis is, among other things, a method of assigning subjects, on the basis of their score vectors, to the multivariate population from which they are most likely to have arisen (see chapter 12). The technique was introduced by R.A. Fisher for the case of two groups (populations), but is readily generalized to n groups. Besides classification, discriminant analysis serves useful purposes in interpreting differences between groups.

The generalization is implicit in equation (8), which is the basis for the tests of multivariate hypotheses described in the previous section. Corresponding to each of the s = min(n-1, p) nonzero roots of (8), one obtains a solution to the linear homogeneous equations

$$\sum_{\ell=1}^{p}[s_b^{(k,\ell)} - \lambda_h s_w^{(k,\ell)}]a_{h\ell} = 0, \quad \begin{matrix} h=1,2,\ldots,s \\ k=1,2,\ldots,p \end{matrix} \tag{10}$$

where $s_b^{(k,\ell)}$ and $s_w^{(k,\ell)}$ are elements in the k-th row and ℓ-th column of S_b and S_w, respectively. Numerically, the solution of this system of equations is a so-called "two-matrix eigen problem," for which standard computer routines exist (Bock and Repp, 1974). The "eigenvectors" of the solution contain the coefficients $a_{h\ell}$ of the discriminant function or "canonical variate" corresponding to each of the nonzero roots. Thus, the h-th canonical variate is computed as

$$v_h = a_{h1}y^{(1)} + a_{h2}y^{(2)} + \ldots + a_{hp}y^{(p)} \tag{11}$$

Because of the invariant properties of canonical analysis, these variates have many interesting and useful properties. The best single score for classifying individuals as to their group membership is contained in the variate, v_1, corresponding to the largest root of (8). This variate maximizes the ratio of between-group to within-group sums of squares and in that sense is the best discriminator. In fact, when there are only two groups, there is just one nonzero root and the corresponding canonical variate is exactly R.A. Fisher's discriminant function. It can be shown in the multivariate normal case (Anderson, 1958) that when a critical value for v_1 is chosen that takes into account the population sizes, the assignment of subjects to groups according to their discriminant score above or below this value is a so-called Bayes procedure that minimizes expected errors of misclassification.

When there are more than two groups, there are additional discriminant functions, each of which maximizes between-group variation relative to within-group variation and is uncorrelated with the other canonical variates. These multiple discriminant functions are useful both in classifying subjects and in interpreting differences between groups. Because the functions are uncorrelated, the square of the distance between any two individuals, or between individuals and the group mean is simply the sum of squares of the differences of the canonical variate scores in within-groups standard deviation units. This so-called generalized (Mahalanobis) distance can be used to classify individuals by assigning them to the group for which the distance from group mean to their location in the multivariate space is smallest. It also can be used to screen for multivariate outliers in data by identifying those subjects whose generalized distance to the group mean is greater than, say, 3 sigma. The scores of those subjects can then be checked for clerical and other errors.

From point of view of interpretation, matters are simplified if the between-group variance of certain of the canonical variates, as measured by the corresponding root of (8), is statistically insignificant or small in practical terms. In that case, an economical summary of the data is achieved by representing the group means or individual subjects' scores in terms of the first, say, $s_0 \leq s \leq p$ canonical variates. If s_0 can be set as small as 2 or 3, the mean canonical scores may be plotted in 2 or 3 dimensions and their relative positions inspected for purposes of interpreting group differences. A simple instance of this will be seen in the first example cited below under Illustrative Applications. In more complex cases, it may be necessary to give some substantive meaning to the canonical variates in order to interpret the multivariate group differences. To the experienced investigator, interpretation is often apparent in the standardized coefficients of the canonical variate, taking into account the effects of correlation among variates and the effects of suppressor variables (see Bock, 1975, chap. 6). If the canonical variates can be characterized and named, a plot of the group mean canonical scores may then be inspected relative to these dimensions, in order to understand how the variables are acting to discriminate between groups in the multivariate space. Some good examples of this approach may be found in Jones (1966).

The same technique can be applied quite generally in multivariate analysis of variance to interpret main effects in complex designs and even the interactive effects. An illustration of the latter is presented in the second of the two examples under Illustrative Applications. The computer programs for multivariate analysis referred to later in this paper routinely compute canonical forms of the main class or interactive effects represented in each test of hypothesis.

Multivariate Analysis of Covariance

In studies where the experimental units are human or animal subjects, most of the sampling error is due to biological variation among subjects randomly assigned to the treatment groups. One strategy for controlling this variation is to sort the subjects into homogeneous "blocks" before assigning them to the treatments. The resulting "randomized block design" can then be analyzed effectively in a two-way "blocks x treatment" multivariate analysis of variance. If there are quan-

titative measures of attributes of the subject that are correlated with the sampling variation, this source of error can be reduced during the data processing by means of multivariate analysis of covariance. Measures used to reduce error in this way are called "control variables," "ancillary variables," or, briefly, "covariables." In drug trials, a measure of the subject's pretrial clinical status is an obvious choice for a covariable, although other medical or social background measures may also serve. In the data analysis, these pretrial measures are included as additional variables in the multivariate analysis of variance along with the posttrial response measures. The two sets of variables are analyzed together and the within-group correlation matrix for all pairs of measures is computed. If the regression of the posttrial measures on the pretrial measures proves to be nonzero, the analysis of variance will effect a reduction of error variance and thus enhance the sensitivity of the experiment. The resulting increase in sensitivity of randomized experiments can be appreciable if a good covariable can be found. In studies where pretrial measures are not available, it is often useful to obtain measures from first-degree relatives of the subjects, since if there is a social class or familial component of the variation among subjects, the responses of the subject and his relative will be correlated. In behavioral studies of younger subjects, information obtained from the mother, father, and older siblings is a potential source of powerful covariates.

Multivariate Multiple Regression Analysis

As a preliminary to multivariate analysis of covariance, it is common practice to perform a formal test of the hypothesis of no association between the response variables and the covariables. All multiple regressions in the set comprised of each response variable on the set of covariables are tested simultaneously. If these regressions are jointly null, the analysis of covariance may actually diminish the power of the test of treatment effects, because it produces no reduction of the error variance while reducing the degrees of freedom of the error estimate. For this reason, the covariates are not ordinarily included in the multivariate model if the hypothesis of no association is not rejected.

In the case of a simple randomized experiment with p response variables and q covariables, x_1, x_2, ..., x_q, the multivariate statistical model for analysis of covariance is (12):

$$y_{ij}^{(k)} = \mu + \alpha_j^{(k)} + \beta_1^{(k)} x_{i1} + \beta_2^{(k)} x_{i2} + \ldots + \beta_q^{(k)} x_{iq} + \epsilon_{ij}. \qquad (12)$$

The multivariate multiple regression analysis tests the hypothesis that the pq regression coefficients, $\beta_\ell^{(k)}$, in this model are jointly null. This hypothesis is tested after effects due to the general mean μ and treatments α_j have been accounted for. It follows that the test should depend upon the common within-groups error variation from which fixed effects have been excluded--hence, that the numerical calculations may be carried out entirely with elements of the corresponding matrix of correlations obtained from (9). This correlation matrix is partitioned into a part due to intercorrelations among the response variables, a part due to intercorrelations among the covariables, and a part due to the cross-correlations between the two sets. The least squares estimate from those correlations of the standardized beta weights corresponding to the regression coefficients for response variable k are obtained by solution of the systems of linear equations (13):

$$\begin{aligned}
&\beta_1^{(k)} + r_{12}\beta_2^{(k)} + \ldots + r_{1q}\beta_q^{(k)} = r_{1k} \\
&r_{21}\beta_1^{(k)} + \beta_2^{(k)} + \ldots + r_{2q}\beta_q^{(k)} = r_{2k} \\
&\qquad\vdots \qquad\qquad\quad \vdots \qquad\qquad\quad \vdots \\
&r_{q1}\beta_1^{(k)} + r_{q2}\beta_2^{(k)} + \ldots + \beta_q^{(k)} = r_{qk}
\end{aligned} \qquad (13)$$

If the solution of (13) is represented by $\hat{\beta}_\ell^{(k)}$, $\ell = 1,2,\ldots,q$, then the maximal invariant statistics for the test of the multivariate hypothesis H_0: $\beta_\ell^{(k)} = 0$, $\ell = 1,2,\ldots,q$, $k = 1,2,\ldots,p$ are the roots of the determinantal equation in ρ^2 expressed by (14). Comparable to (8), the alternative test statistics computed from the $s = \min(p,q)$ nonzero roots of (14),

$$\left|\left[\sum_{\ell,m}^{q} \hat{\beta}_\ell^{(g)} \hat{\beta}_m^{(k)} r_{\ell m}\right] - \rho^2 [r_{gk}]\right| = 0 \text{ , for } \{^{g=1,2,\ldots,p}_{k=1,2,\ldots,p}\} \qquad (14)$$

are given by,

Wilks' criterion	$\Lambda = \prod_{h=1}^{s} (1-\rho_h^2)$
Roy's largest root criterion	$\theta = \rho_1^2$
Hotelling's trace criterion	$\tau = (N-n-q) \sum_{h=1}^{s} \rho_h^2/(1-\rho_h^2)$

In practical work, it may be desirable to test not only the overall association between the two sets of variables, but also to test the multivariate partial contribution of each covariable as it is added to the generalized regression equation. Provided the order in which the covariables are to enter the equation is specified beforehand, these tests of partial contribution are a straightforward application of these same multivariate test criteria (see Bock, 1975, pp. 378-379). In analysis of covariance, it is advantageous to keep the number of covariables in the multivariate model as few as possible, thus holding to a minimum the number of degrees of freedom lost in the test of the treatment means.

Multivariate multiple regression analysis may, of course, be used in its own right as a technique for investigating asymmetrically the relationships between two sets of variables in a given population. The asymmetry consists in the fact that one set is treated as the independent variable set, often called "predictors," and the other as the dependent variable set or "criteria." The technique amounts to doing a univariate multiple regression analysis of each criterion variable, in turn, on the whole set of predictors. The analysis assesses the power of the multiple predictors to diminish dispersion in the conditional multivariate distribution of the criteria.

Canonical Correlation

When structural relationships between two sets of variables are investigated, there may be no sense in which one set may be regarded as predictors and the other criteria. In that case, a symmetrical concept of relationships between the two sets is needed, and one possibility is the canonical correlation defined by Hotelling (1936) [1] Given p variables in one set and q in the other, Hotelling defines $s = \min(p,q)$ pairs of linear combinations (one for each set of variables in each pair) that are maximally correlated within pairs and are uncorrelated outside the pairs, both within and across the two sets. The coefficients of these linear combinations are, in fact, just the eigenvectors associated with the solutions of (14), and the squares of their corresponding canonical correlations--as the maximal correlations are called--are just the solutions ρ_h^2 of the determinantal equation. Actually, these linear combinations are identical to the discriminant functions in the case where the response measures are regarded as one set of variables and any n-1 independent contrasts among the n treatment groups are regarded as the other. In an exact sense, all of the multivariate normal procedures based on maximal invariant statistics are special cases of canonical correlation.

Perhaps the most useful contribution of canonical correlation to data analysis is in providing a large sample test of the dimensionality of linear relationships between the two sets of variables. If the sample is large enough to justify treating the canonical variates corresponding to the larger canonical correlations as being equivalent to the population expressions, an application of Wilks' criterion to the remaining variation, with suitable adjustment of degrees of freedom, as given by Bartlett (1947), provides a test of the hypothesis that all of the significant associations between the two sets is contained in the excluded variates. In behavioral research, it is not unusual to find that significant association can be demonstrated, even between large sets of behavioral measures, only in one or two dimensions. If these dimensions can be conceptualized and named, a considerable simplification in the discussion of the results may be gained. For purposes of interpreting these dimensions, it is helpful to examine the first-order correlation structure between the corresponding canonical variates and the original variables, possibly after orthogonal rotation of this structure (for example, by Kaiser's Varimax procedure).

ILLUSTRATIVE APPLICATIONS

The first example is a comparison of populations of subjects defined by attributes (diagnosis); the data are taken from a study by Kahana (1968) as reanalyzed by Bock (1975, p. 289; 1976). The second example is based on a reanalysis of some of the data from a study of drug therapy reported by Hogarty, Goldberg and Schooler (1974).

A COMPARISON OF PSYCHIATRIC DIAGNOSTIC GROUPS

Data for this example are measures of psychiatric clinical status, based on the Kahn Mental Status Questionnaire (MSQ) and the Face-Hand test (F-H) administered to samples of confused, psychotic, and alcoholic geriatric patients in a study by E. Kahana. The study included an experimental treatment, but for present purposes, the experimental effect, which proved to be small, is ignored and the data are subjected to a one-way multivariate analysis of variance of the diagnostic classification. The group means $y._j^{(k)}$, the sample sizes N_j, for the three groups ($j = 1,2,3$), and the common within-group standard deviations s_k and correlations r_{gk}, between predictors for $g = 1,2$ and $k = 1,2$, are shown in Table 3.

TABLE 3

Diagnosis of group means and within-group standard deviations and correlations for the Mental Status Questionnaire (MSQ) and Face-Hand test

Diagnosis Group Means	N	Variable MSQ	F-H
Confused (Conf)	22	7.624	10.085
Psychotic (Psyc)	17	1.786	1.443
Alcoholic (Alch)	16	2.214	1.366
Standard deviations		2.466	2.613
Correlations		1.000 .490	.490 1.000

From the quantities in Table 3, the within-group sums of squares and cross-products (SSP) can be recovered by the calculation (for $n = 3$),

$$s_w^{(g,k)} = [\Sigma N_j - n]s_g s_k r_{gk} \quad ,$$

in which $r_{gk} = 1$ when $g = k$.

The between-group SSP is computed from the formula in Table 2. The results of these calculations appear in the partition of squares and products shown in abbreviated form in Table 4.

TABLE 4

Partition of sums of squares and cross-products

Source of Dispersion	Degrees of Freedom	Sums of Squares and Cross-products	
Between groups	2	419.981 644.797	644.797 994.416
Within groups	52	316.172 164.106	164.106 355.052
Total corrected for the grand mean	54		

According to (8), the maximal invariant statistics are obtained by solving the determinantal equation (15):

$$\begin{vmatrix} 419.981 - 316.172\lambda & 644.797 - 164.106\lambda \\ 644.797 - 164.106\lambda & 994.416 - 355.052\lambda \end{vmatrix} = 0 \qquad (15)$$

Expanding (15) gives the quadratic equation

$$85{,}326.724\lambda^2 - 251{,}891.552\lambda + 1{,}872.313 = 0$$

from which the quadratic formula yields two real roots

$$\lambda_1 = 2.945 \text{ and } \lambda_2 = .00745.$$

As a statistic for testing the null hypothesis of no multivariate differences between diagnostic groups, the likelihood ratio, for example, may be computed from the roots:

$$1/(1 + 2.945) \times 1/(1 + .00745) = .2516.$$

Corresponding to this value, C.R. Rao's F approximation for the distribution of the likelihood ratio (which is exact for the case p = 2), gives a p value less than .0001.

There is no doubt as to the significance of differences between the groups, but the Bartlett test of the contribution of the smaller root shows it to be nonsignificant (Table 5). Thus, of the two discriminant functions shown in Table 5, only the first is useful in characterizing the score distributions of the diagnostic group. This implies that the group centroids (vector means) are collinear with the axis of the first canonical variate, as is apparent in Figure 1 at the end of this chapter. It is clear that performance on the Mental Status Questionnaire and the Face-Hand serves only to distinguish the confused from the psychotic and alcoholic groups. Note that although the Face-Hand test has the larger standardized coefficient in the discriminant function, both tests contribute significantly to discrimination. This can be verified by computing, in an analysis of covariance, the between-group F statistic for the MSQ, eliminating variation due to F-H, treated as a covariate. The resulting F, 18.0 on 2 and 51 degrees of freedom, has a p value less than .0001. The fact that both variables contribute significantly to the discriminant function does not necessarily imply that the two tests measure different dimensions of variation; however, it may merely signify that they are both unreliable measures of the same dimension and that the improved discrimination due to adding either to the function is essentially a "test-lengthening" effect, analogous to improving the reliability and validity of a test by adding items of the same type.

TABLE 5

Discriminant functions, canonical variances, and Bartlett's chi-square

Variable	First Function (v_1) Coefficient (Standardized)	Second Function (v_2) Coefficient (Standardized)
MSQ	.1029 (.2539)	.4536 (1.1186)
F-H	.3256 (.8509)	-.2944 (-.7692
Canonical variance	2.945	.00745
X^2_B	71.06	.3823
d.f.	4	1
p	<.0001	.5364

EFFECTS OF ALTERNATIVE AFTERCARE PROGRAMS ON THE ADJUSTMENT OF NON-HOSPITALIZED, NON-RELAPSED SCHIZOPHRENIC PATIENTS[2]

This example is drawn from data obtained by Hogarty, Goldberg, Schooler, and others (1973, 1974) in a large-scale study of posthospitalization maintenance programs for schizophrenic patients. From an initial sample of 374 patients especially selected from populations in three clinics in the Baltimore area, approximately equal numbers of male and female subjects were randomly assigned to a 2 x 2 design of drug and sociotherapeutic aftercare treatments. The drug treatments consisted of a minimum of 100 mg/day of chloropromazine (Dr) vs. placebo (No-Dr); the sociotherapeutic treatments consisted of major role therapy (MRT) conducted by experienced social workers vs. no major role therapy (No-MRT).

The Data

The number of subjects of each sex initially assigned to the treatment combinations are shown in Table 6. Of these, the present analysis deals with measures of behavioral adjustment only for patients who continued in the study for 24 months without relapse or rehospitalization. The numbers of such patients, classified by sex, treatment combination, and clinic, are shown in Table 6.

Survival rates estimated from the data in Table 6 appear to indicate a drug effect and sex x drug interaction (females respond more favorably to the drug than males), and this impression is confirmed by log-linear analysis of the frequency data (see Bock, 1975, chap. 8; Bishop, Fienberg and Holland, 1975). The sex x drug interaction is required for a satisfactory fit of the log-linear model, but other two-factor and higher interactions are not. This suggests that interactions involving the sex factor should be examined carefully in analysis of the adjustment measures among the survivors.

TABLE 6

Initial and final sample composition

Group	Sex	Drug	Socio-therapy	Clinic	Initial Sample	Patients Unrelapsed at 24 mo.
1	Male	Dr	MRT	1	16	2
2				2	12	2
3				3	12	6
4			No-MRT	1	17	1
5				2	13	3
6				3	11	5
7		No-Dr	MRT	1	13	1
8				2	12	0
9				3	13	2
10			No-MRT	1	17	2
11				2	12	2
12				3	10	3
13	Female	Dr	MRT	1	17	9
14				2	20	12
15				3	18	12
16			No-MRT	1	17	5
17				2	22	13
18				3	17	7
19		No-Dr	MRT	1	21	2
20				2	20	3
21				3	16	1
22			No-MRT	1	17	1
23				2	15	0
24				3	16	3
					Total = 374	Total = 97

Although many different behavioral assessments were included in the study (see Hogarty, Goldberg and Schooler, 1974), only one measure typical of each of four independent sources of adjustment ratings will be used in this example. The retained measures are the following:

1. SCL FA - Fear Anxiety scale of the Johns Hopkins Symptom Check List (source: patient self-report)
2. IMPS AI- Anxious Intropunitiveness scale of the Inpatient Multidimensional Psychiatric Scale (source: psychiatric interview)
3. ADJUST - Course of adjustment global rating (source: social worker report)
4. KAS DISC-Katz Adjustment Scale, discrepancy between expected and actual behavior (source: relative of patient)

For each of these variables the scale of measurement is arbitrary, but the units should be meaningful to persons who are familiar with the rating instruments. In each scale, higher scores correspond to more psychopathology and indicate a less favorable adjustment.

In addition to the response measures, certain pretreatment information on the subjects was available and is potentially useful as covariables for reducing sampling variation. In the present analysis, the following covariables will be included:

1. ADM SCL - Neurotic Feelings scale of the SCL at hospital admission
2. ADM IMPS - Anxious Intropunitiveness scale of the IMPS at hospital admission
3. ONSET - Age of subject at onset of schizophrenia
4. MO EDUC - Mother's education (arbitrary scale)

For purposes of all subsequent computations in the multivariate analysis of variance, the data may be summarized in the statistics shown in Table 7. These include the means for each response variable and covariable in each of the twenty-two nonvacant cells in the design, and the within-cell correlation matrix and the standard deviations of the respective variables. As would be expected for scales assessing the same latent trait (adjustment) with reliability typical of rating scales, the intercorrelations of the response variables are all positive and moderate in size.

TABLE 7

Group means and common within-group correlations and standard deviations for all variables

1. Group means

Group	Response Variables				Covariables			
	SCL FA	IMPS AI	ADJUST	KAS DISC	ADM SCL	ADM IMPS	ONSET AGE	MO EDUC
1	15.0	2.0	3.0	6.5	17.0	18.0	17	7.0
2	6.0	2.0	1.0	5.5	19.5	9.0	19	4.5
3	8.0	2.7	7.7	5.2	18.3	14.0	26	5.3
4	10.0	2.0	5.0	5.0	19.0	4.0	12	6.0
5	28.0	3.3	8.3	7.3	17.3	12.0	26	5.3
6	20.0	2.8	13.2	6.4	20.6	22.0	24	6.0
7	40.0	6.0	23.0	9.0	15.0	34.0	21	6.0
--								
9	13.0	3.0	14.5	8.5	33.5	47.0	20	5.5
10	2.0	2.0	0.0	5.0	23.0	12.0	15	6.5
11	7.0	3.0	6.0	5.0	24.0	11.0	18	6.0
12	4.7	3.3	14.3	5.7	14.3	15.3	18	6.0
13	17.1	1.4	4.1	5.9	19.9	27.6	27	5.7
14	11.7	3.3	10.0	5.8	18.5	13.8	21	6.3
15	7.5	2.7	8.0	6.6	22.3	19.5	27	5.7
16	14.8	2.0	8.8	8.0	22.4	22.0	20	6.0
17	16.8	3.8	13.3	7.3	18.5	11.8	26	5.7
18	14.3	4.0	11.1	6.9	22.1	20.3	28	6.3
19	28.0	4.3	15.0	10.5	29.5	51.0	15	6.5
20	9.3	2.0	20.7	8.0	19.3	16.7	37	5.7
21	0.0	3.0	2.0	6.0	18.0	8.0	18	4.0
22	28.0	2.7	0.0	7.0	14.0	30.0	25	7.0
--								
24	11.3		11.0	6.0	17.7	24.0	19	6.0

(Table 7 continued on next page)

(Table 7 continued)

2. Correlations and Standard Deviations

	SCL FA	IMPS AI	ADJUST	KAS DISC	ADM SCL	ADM IMPS	ONSET AGE	MO EDUC
SCL FA	1.0000							
IMPS AI	.4652	1.0000						
ADJUST	.2651	.4141	1.0000					
KAS DISC	.4450	.2476	.4213	1.0000				
ADM SCL	.1995	.0936	.0534	.3554	1.0000			
ADM IMPS	.0391	.2504	.1180	.1395	.2250	1.0000		
ONSET	-.1708	-.1848	-.0656	-.1344	-.1000	-.0229	1.0000	
MO EDUC	-.0153	-.0657	.0734	-.0310	-.1525	-.0482	.4051	1.0000
S.D.	10.9339	0.8726	8.9233	2.1009	5.5450	12.9669	9.8719	1.1558

The Analysis of Regression

The first step in the analysis of these data consists of statistical tests of the contribution of the covariables to the multivariate linear model for the group means. For this purpose, a suitable statistic is the likelihood ratio (Wilks' criterion) for the multivariate hypothesis of no within-cell correlation between the response variables and the covariables versus a general alternative. The p-value for this statistic proves to be .186 on the null hypothesis, consistent with no association. However, if a similar statistic is computed for the successive partial contribution of each covariable (ordered in terms of their expected potential for error reduction), the corresponding p-values are:

Covariable added to the regression	Multivariate p-value (4 response variables)
ADM SCL	.026
ADM IMPS	.230
ONSET	.550
MO EDUC	.846

Although its effect is lost in the joint test, the ADM SCL, as the best single predictor, shows a p-value small enough to suggest that at least this scale be retained as a covariable in the analysis. Its inclusion is unlikely to have any great effect on the result, but since the cost is only one degree of freedom lost from the error estimate, the significant result ($p = .026$) indicates that some reduction of error will be gained. The remaining covariables make no additional contribution and may be omitted. The indication that onset and mother's education are negatively related to outcome pathology, although plausible, is not statistically significant in these data.

Tests of Alternative Models

The next step is to test the design-factor main effects and interactions after including the covariables in the model. In this multivariate analysis of covariance, the likelihood ratio statistic is again a convenient criterion for a joint test of effects in the four response variables. Because the subclass numbers for the surviving subjects are highly disproportionate, a nonorthogonal analysis, requiring a specification of the order of effects, is required. If the principle that factors with greater a priori expectation of effect appear first is followed, the order of elimination of effects shown in Table 8 is plausible. The effects tested include all two-factor interactions and the one three-factor interaction (Sex x Drug x MRT) that is of some interest. The remaining effects are tested jointly in the residual.

TABLE 8

Results of nonorthogonal multivariate analysis of covariance
(order of elimination from top downward)

Effect	Degrees of Freedom	Multivariate p-value (4 response variable)
Constant	1	Test suppressed
Clinics	2	<.001
Sex	1	.751
Drug	1	021
MRT	1	.650
Sex x Drug	1	.398
Sex x MRT	1	.644
Sex x Clinic	2	.144
Drug x MRT	1	.053
Drug x Clinic	2	.253
MRT x Clinic	2	.412
Sex x Drug x MRT	1	.162
Residual	6	.699
	Total = 22 (number of nonvacant cells)	

In the more extensive data analyzed by Hogarty, Goldberg and Schooler (1974), a significant Sex x Drug x MRT effect was found. In the present data, significance is not achieved (p = .162), but the p-value is sufficiently small to encourage some inspection of this aspect of the data. In particular, the univariate F-tests for this interaction show that any possible effect is confined to the SCL FA scale:

Effect	Variate	p
Sex x Drug x MRT,	SCL FA	.020
Eliminating two-factor	IMPS AI	.634
Interactions and main	ADJUST	.726
Effects	KAS DISC	.919

To locate the source of this complex interaction in the SCL FA scores--that is, to identify the particular cell(s) that contribute substantially to the interaction--it is helpful to compute residuals of the cell means after subtracting expected values from a model of lesser rank from which the Sex x Drug x MRT term is excluded. These residuals are shown in Table 9 in raw form and multiplied by the square roots of the subclass numbers to reflect their varying precisions. A large contribution to interaction arises from a single male subject in Group 7 under the NoDrug-MRT condition and a single female subject in Group 22 under the NoDrug-NoMRT condition. If the scores for these two subjects are excluded from the analysis, the Sex x Drug x MRT interaction is no longer significant (p = .226). Since the effect, if real, is limited to the patient self-report and may depend fortuitously on two particular subjects, it should perhaps not be interpreted on the basis of the present data. If this point of view is accepted, there are no significant effects due to sex in the adjustment data, and the statistical model for the data is considerably simplified.

Fitting the Reduced Rank Model

The foregoing considerations suggest that a rank 7 model, consisting of one covariate (ADM SCL), constant, clinic, Drug, MRT, and Drug x MRT effects, should give a good account of the data. The multivariate analysis of variance associated with the fitting of this model, shown in Table 10, supports this conclusion; the residual variation is not significantly greater than the within-cell variation.

TABLE 9

Residual symptom check list (SCL FA) means
after fitting two factor (rank 7) model

Group	Factor				N	SCL FA Adjusted Residuals	Residuals x $\sqrt{N}$
	Sex	Drug	MRT	Clinic			
1	1	1	1	1	2	.9	1.3
2				2	2	-5.5	-7.8
3				3	6	0.0	0.0
4			2	1	1	-11.2	-11.2
5				2	3	11.1	19.2
6				3	5	5.5	12.3
7		2	1	1	1	23.3	23.3
8				2	0	---	---
9				3	2	- 4.1	- 5.8
10			2	1	2	-12.7	-18.0
11				2	2	- 4.7	- 6.6
12				3	3	.1	.2
13	2	1	1	1	9	1.9	5.7
14				2	12	.5	1.7
15				3	12	- 1.8	- 6.2
16			2	1	5	- 7.6	-22.4
17				2	13	- .6	- 2.2
18				3	7	- 1.2	- 3.7
19		2	1	1	2	5.6	7.9
20				2	3	- 5.5	9.5
21				3	1	-11.1	-11.1
22			2	1	1	16.7	16.7
23				2	0	---	---
24				3	3	5.4	9.4

TABLE 10

Results of multivariate analysis of covariance for a rank 7 model

Effect	Degrees of Freedom	Likelihood Ratio p-value
Constant	1	---
Clinic	2	<.001
Drug	1	.008
MRT	1	.536
Drug x MRT	1	.010
Residual	16	.552
	Total = 22	

Estimated effects with standard errors under this model are shown in full-rank form in Table 11. There are, essentially, 68% confidence bounds for the contrasts listed. The two degrees of freedom for clinic effects are arbitrarily assigned for purposes of estimation to the simple contrasts of Clinics 1 and 2 with Clinic 3. Drug and MRT effects are estimated as treatment condition minus no treatment. The Drug x MRT interaction term is parameterized as

(Drug, MRT - Drug, NoMRT) - (NoDrug, MRT - NoDrug, NoMRT).

The only substantially interesting terms in Table 11 are, in fact, just these estimates of the Drug x MRT interaction. For all variables, they are obviously large relative to any other esti-

mated contrast. As presented in Table 11, their magnitudes depend on the units of measurement of the several scales and cannot be compared directly. If they are standardized as follows by dividing by the reduced common within-group standard deviations, however, the interactive effects are seen to be similar in magnitude in all variables:

	SCL FA	IMPS AI	ADJUST	KAS DISC
Drug x MRT	-1.046	-1.540	-1.312	-1.628

For a more intuitively understandable representation of the interactive effect, the estimated effects in Table 11 may be used to estimate means for populations represented by the Drug x MRT treatment combinations. Because the design is nonorthogonal, these means are not estimated by the corresponding subgroup sample means, but must be reproduced from the fitted model including the covariate adjustment and interaction term. The means computed in this way, shown for the present data in Table 12, are best estimates of the means that would be estimated directly by the Drug x MRT marginal means if the design had been orthogonal.

TABLE 11

Estimated effects for rank 7 model

Effects	Response Variables			
	SCL FA	IMPS AI	ADJUST	KAS DISC
Constant	5.78±4.82	2.73±0.39	8.64±4.01	4.07±0.88
Clinic 1 - Clinic 3	6.77±2.85	-0.63±0.23	-4.02±2.37	0.49±0.52
Clinic 2 - Clinic 3	3.29±2.60	0.68±0.21	0.82±2.17	0.34±0.48
Drug - NoDrug	2.21±2.75	-0.69±0.22	-3.14±2.29	-0.38±0.50
MRT - NoMRT	-0.62±2.79	0.24±0.23	1.98±2.23	0.48±0.51
Drug x MRT	-11.29±5.68	-1.35±0.46	-11.77±4.72	-3.22±1.04
Regression on ADM SCL	.393±.225	.015±.018	.086±.187	.135±.040

TABLE 12

Estimated Drug x MRT subclass means

Treatment Combination		Expected Means			
		SCL FA	IMPS AI	ADJUST	KAS DISC
Drug	MRT	11.71	2.47	6.83	6.08
	NoMRT	18.06	2.90	10.78	7.22
No Drug	MRT	15.94	3.90	16.21	8.15
(Placebo)	NoMRT	9.81	2.87	7.86	5.95

As would be expected if the response variables were fallibly measuring the same underlying trait (adjustment), the pattern of interaction revealed in the expected cell means is essentially the same in all measures; thus, reading down each column of Table 12 we find the pattern of "low-high-high-low" across the two factors for all dependent variables. This suggests that a still simpler characterization of the interaction could be obtained by computing a linear combination of variables that represents best, in some sense, the underlying interactive effect. The coefficients of such a combination are given by the discriminant function maximizing the squared interactive parameter relative to the within-group error estimate. In this instance the coefficients of this function, in both raw-score and standardized form, are as follows:

	Raw Coefficients	Standardized
SCL FA	.0006	.0063
IMPS AI	.5021	.4707
ADJUST	.0227	.1966
KAS DISC	.3195	.6104

The function weights all variables positively, but gives greater weight to the psychiatric and family ratings, presumably because they are the more reliable indices of the Drug x MRT interaction. The unit of scale of the function is, of course, arbitrary and has been chosen so that the mean discriminant scores shown in Table 12 are standardized (in sigma units) in the common within-group variation. When these scores are displayed graphically in Figure 2, they reveal clearly the nature of the interaction (which is significant at the 0.6% level): the unrelapsed patients maintained on chloropromazine combined with sociotherapy show less pathology, but those maintained on the drug alone are less well adjusted than the NoDrug, NoMRT controls. Similarly, patients receiving sociotherapy while not on drugs are less well adjusted than the controls.

Hogarty, Goldberg and Schooler (1974) discuss this interaction in some detail. The synergistic effect of drug and sociotherapy is extremely plausible and easy to accept. The apparent deleterious effect of the sociotherapy in the absence of the drug is not at all plausible, however, and suggests some tendency for selective survival of only the better adjusted patients among those receiving neither drug nor sociotherapy (see, however, Hogarty, Goldberg, and Schooler, 1974, p. 615).

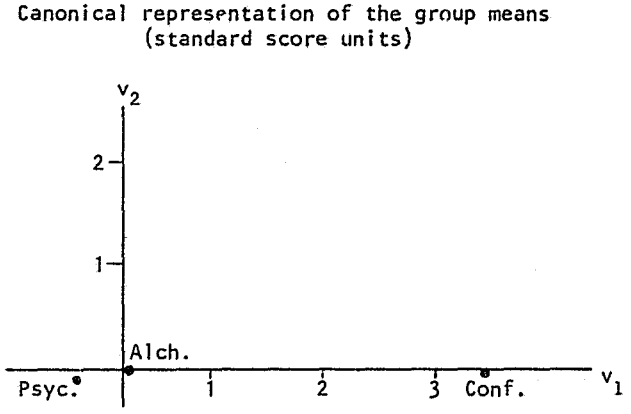


Figure 1. Canonical representation of the diagnostic group centroids.

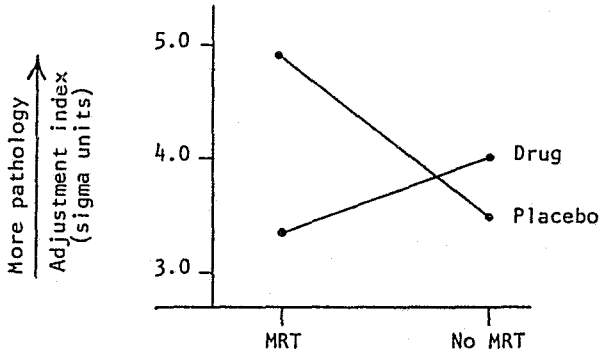


Figure 2. Canonical representation of the Drug x MRT interaction (p=.006)

NOTES

[1]Another is covariance structure analysis (Jöreskog, 1970).

[2]Data for this example, supplied by Dr. Solomon C. Goldberg, Psychopharmacology Research Branch, National Institute of Mental Health, are derived from Hogarty, G.E.; Goldberg, S.C.; and Schooler, N.R. Drug and Sociotherapy in the Aftercare of Schizophrenic Patients, III Adjustment of Non-relapsed Patients. Archives of General Psychiatry, 31:609-618, 1974.

RESOURCES AND REFERENCES

COMPUTER PROGRAMS

At the present time, the best known, most widely available computer programs for multivariate analysis of variance and allied techniques are MULTIVARIANCE (Finn, 1974) and MANOVA II (Cramer, 1975). These programs carry out univariate or multivariate analysis of variance for any design, balanced or unbalanced, complete or incomplete. At the option of the user and where the data permit, the calculations include multivariate multiple regression analysis, discriminant analysis, and canonical correlation.

The MULTIVARIANCE program also has special provisions for analysis of repeated measures. By "repeated measures" is meant measures of the same subject on more than one occasion with the same measuring instrument. Since the same instrument is used, it is assumed that the measurements obtained are commensurate for purposes of computing sums and differences. Thus it is meaningful to obtain averages for each subject over all occasions or to subtract scores from one occasion to another in order to compute gains. (A typical form of repeated measures data is that resulting from a clinical experiment in which "each subject serves as his own control," i.e., where the subject is measured with the same instrument both before and after a treatment intervention.) On somewhat restrictive assumptions, it is possible to analyze repeated measures data using a univariate "mixed-model" analysis of variance. But under much less restrictive assumptions the analysis may be carried out in the form of a one-way multivariate analysis of variance aimed at testing differential change among the treatment groups. Basically, the multivariate approach to repeated measures analysis consists of transforming the data into a number of a priori contrasts among the repeated measures and applying the multivariate analysis of variance to these contrasts. The MULTIVARIANCE program facilitates this type of analysis by providing for automatic generation of the transformations corresponding to these contrasts in experimental designs of any complexity. Both multivariate statistical tests and the conventional univariate mixed-model analysis can be extracted from the same MULTIVARIANCE run.

The MULTIVARIANCE and MANOVA II programs are distributed by International Educational Services (not-for-profit), P.O. Box A3650, Chicago, Illinois 60690.

Certain aspects of multivariate analysis, including discriminant analysis and canonical correlation can be carried out with the BMD routines distributed by the Health Sciences Computing Facility, University of California, Los Angeles, California 90024 (see Dixon, 1975). These routines have been extended by Wilkinson (1975) to include basic features of multivariate analysis of variance. The use of dummy variables and multiple regression techniques for multivariate analysis of variance have been suggested (Woodward and Overall, 1975), but this approach does not handle estimation of effects in a comprehensive way and cannot be recommended in general.

Examples in this paper were prepared with MULTIVARIANCE, with the exception of the preliminary log-linear analysis in Example 2, which was done with the MULTIQUAL program of Bock and Yates (1974).

REFERENCES

Anderson, T.W. *An Introduction to Multivariate Analysis.* New York: Wiley, 1958.

Bartlett, M.S. Multivariate analysis. *Journal of the Royal Statistical Society,* 9:176-197, 1947

Bishop, Y.M.M.; Fienberg, S.E.; and Holland, P.W. *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Mass.: M.I.T. Press, 1975.

Bock, R.D. *Multivariate Statistical Methods in Behavioral Research.* New York: McGraw-Hill, 1975.

Bock, R.D., and Repp, B.H. *MATCAL: Double-Precision Matrix Operations Subroutines for the IBM System/360-370 Computers.* Chicago: National Educational Resources, 1974.

Bock, R.D., and Yates, G. *MULTIQUAL: Log-Linear Analysis of Nominal or Ordinal Qualitative Data by the Method of Maximum Likelihood.* Chicago: National Educational Resources, 1974.

Cochran, W.G., and Cox, G.M. *Experimental Designs,* 2nd ed. New York: Wiley, 1957.

Cramer, E.M. *MANOVA II.* Chicago: National Educational Resources, 1975.

Dixon, W.J., ed. *BMDP: Biomedical Computer Programs.* Berkeley: University of California Press, 1975.

Finn, J.D. *MULTIVARIANCE: Univariate and Multivariate Analysis of Variance, Covariance and Regression.* Chicago: National Educational Resources, 1974.

Hogarty, G.E.; Goldberg, S.C.; and Schooler, N.R. Drug and sociotherapy in the aftercare of schizophrenic patients, III Adjustment of nonrelapsed patients. *Archives of General Psychiatry,* 31:609-618, 1974.

Hotelling, H. Relations between two sets of variates. *Biometrika,* 28:321-377, 1936.

Jones, L.V. Analysis of variance in its multivariate developments. In: Cattell, R.B., ed., *Handbook of Multivariate Experimental Psychology.* Chicago: Rand-McNally, 1966, pp. 244-266.

Jöreskog, K.G. A general method of analysis of covariance structures. *Biometrika,* 57:239-251, 1970.

Kahana, E. The effects of age segregation on elderly psychiatric patients. Unpublished Ph.D. Dissertation, Committee on Human Development, University of Chicago, 1968.

Rao, C.R. *Advanced Statistical Methods in Biometric Research.* New York: Wiley, 1952.

Wilkinson, L. Response variable hypothesis in the multivariate analysis of variance. *Psychological Bulletin,* 82:408-412, 1975.

Woodward, J.A., and Overall, J.E. Multivariate analysis of variance by multiple regression methods. *Psychological Bulletin,* 82:21-32, 1975.

Plate 12

INTRODUCTION AND RATIONALE: AN AID TO THREE PROBLEMS

Broadly conceived, discriminant analysis is a system of multivariate statistical techniques that provides an integrated approach to the solution of three distinct but interrelated problems often encountered by researchers in various fields--particularly the behavioral and social sciences. The three problems (or perhaps it would be better to call them three aspects of an overall research problem) are: (a) to determine whether or not significant differences exist among two or more groups of individuals in terms of several descriptor variables (significance testing); (b) if such differences exist, to try to "explain" them in terms of a smaller number of "underlying factors" than the original descriptor variables (explanation of group differences); and (c) to utilize the multivariate information from the samples studied in assigning a future individual to one of the several groups studied--assuming that the individual must be a member of one or another of these groups (classification).

The reader will recognize that (a) is precisely the problem addressed by multivariate analysis of variance (MANOVA) in its simplest form (one-factor design). For this reason, discriminant analysis is often characterized as a follow-up or adjunct to MANOVA, focusing on aspect (b), the explanation of group differences in terms of a small number of underlying factors. This aspect, which may be referred to as "discriminant analysis proper," bears a certain resemblance to factor analysis. The difference is that, whereas factor analysis seeks to explain individual differences on a large number of attributes in terms of a small number of factors, discriminant analysis seeks to do this for group differences.

The third aspect, (c), of discriminant analysis is more properly referred to as a classification procedure, but since this is often the ultimate goal of many practical research endeavors, we include it under the general rubric of discriminant analysis. In fact, this aspect was the primary focus when two-group discriminant functions were first developed by Fisher in 1935. The shift of emphasis to aspect (b) is a relatively recent development, as is the extension to situations for which aspect (a) corresponds to MANOVA of factorial and other designs.

METHODS AND PROCEDURES

THE GEOMETRIC APPROACH

As the starting point of discriminant analysis, we look for the linear combination (i.e., a weighted sum) of the original variables such that the F-ratio for testing the significance of the differences among the several group means on this linear combination is larger than that for any other linear combination of the original variables. The idea is perhaps best grasped by looking at the geometric representation of what is involved. For this purpose we take the simplest case of two groups (e.g., drug users and nonusers) and two variables (e.g., two personality attributes such as introversion and egocentrism).

Let us denote the two groups by U and NU, and the two variables by X_1 (= introversion) and X_2 (= egocentrism). A linear combination of X_1 and X_2 is any expression of the form

$$Y = v_1X_1 + v_2X_2,$$

where v_1 and v_2 are suitable weights applied to the two variables in forming their weighted sum. For instance, we might take $v_1 = 3$ and $v_2 = 4$, in which case

$$Y = 3X_1 + 4X_2.$$

To get a person's Y-score, we would multiply the X_1-score by 3, and to this add 4 times the X_2-score.

The reader should imagine finding the Y-score for everyone in the two groups in the above manner, then imagine calculating the F-ratio for testing the significance of the difference between $\bar{Y}_U$ and $\bar{Y}_{NU}$, the two group means on Y. Once the Y-scores have been calculated for everyone, the task is no different from the situation in which Y was the observed variable to begin with. (Of course, in the two-group case a t-ratio is ordinarily used instead of an F-ratio, but since $t^2 = F$ in this case, we may, for consistency, imagine calculating F rather than t.)

The value of the F-ratio will, of course, depend on what relative weights we choose for X_1 and X_2 in defining Y. For instance, consider another linear combination

$$Y' = 10X_1 + 3X_2.$$

This will give rise to a different F-ratio, say F', and we may compare F (resulting from the Y above) and F' to see which is larger. To determine the pair of weights v_1 and v_2 that give rise to the largest possible value of the F-ratio is the task of discriminant analysis.

To describe the above developments geometrically, it is necessary to associate the algebraic process of forming a linear combination with the geometric operation of determining a new axis in the plane defined by the original X_1 and X_2 axes. There is one modification we need to make in the definitions of Y and Y'. Namely, the weights must be such that the sum of their squares is unity, in order for the scale unit on the new axis (Y or Y' as the case may be) to remain the same as that on the X_1 and X_2 axes. Since only the relative weights for X_1 and X_2 matter in determining the resulting F-ratio value, such a modification is always possible. (We need only divide each weight by the square root of the sum of their squares.) For instance, the three linear combinations

$$Y_1 = 3X_1 + 4X_2$$
$$Y_2 = 1.5X_1 + 2X_2, \text{ and}$$
$$Y_3 = .6X_1 + .8X_2$$

will all lead to the same F-ratio value since the relative weights attached to X_1 and X_2, respectively, are in the ratio 3:4 in each case. Of these, Y_3 alone has the property that the squares of the weights sum to unity $[(.6)^2 + (.8)^2 = .36 + .64 = 1.00]$. This will therefore be taken as the "representative" of the class of linear combinations comprising Y_1, Y_2 and Y_3 among others, and will be denoted by Y without subscript.

The new axis Y corresponding to the above linear combination may be drawn by plotting any point whose (X_1, X_2) coordinates are proportional to .6 and .8 (e.g., 6 and 8), and connecting that

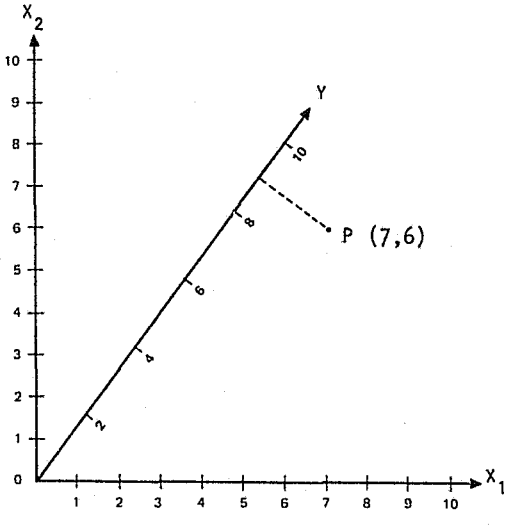


Figure 1. The axis corresponding to the linear combination $Y = .6X_1 + .8X_2$, the point P representing a person with scores 7 and 6 and X_1 and X_2, respectively, and its projection onto the Y axis.

point with the origin, as shown in Figure 1. Also shown in Figure 1 is the point P representing a person who scored 7 points on X_1 and 6 points on X_2. If we drop a perpendicular from P onto the Y axis--that is, if we project P onto the Y axis--we arrive at a point on the Y axis whose scale value is 9. This is precisely the Y-score obtained from the linear combination formula:

$$Y = (.6)(7) + (.8)(6) = 9.0.$$

This is what is meant when we speak of associating a linear combination with a new axis, or of drawing the axis corresponding to a given linear combination.

Now suppose that the two groups U (drug users) and NU (nonusers) had the following means on X_1 and X_2:

$$\bar{X}_{1,U} = 7.0 \qquad \bar{X}_{1,NU} = 5.0$$

$$\bar{X}_{2,U} = 6.8 \qquad \bar{X}_{2,NU} = 4.5$$

When the point (7.0, 6.8) in Figure 2--called the centroid of Group U on X_1 and X_2--is projected onto the Y axis, we get the Y-mean, $\bar{Y}_U$, for this group. Similarly, the projection of the Group NU centroid, (5.0, 4.5), onto the Y axis gives $\bar{Y}_{NU}$, the Y-mean for Group NU. The distance between $\bar{Y}_U$ and $\bar{Y}_{NU}$ gives a rough idea of how well the two groups are differentiated along the dimension represented by the Y axis. But further refinements are necessary before we can use such a distance as a measure of separation of the two groups.

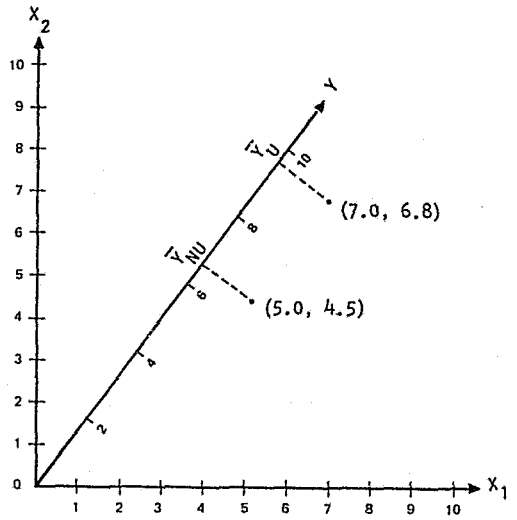


Figure 2. The projections of the Group U centroid (7.0, 6.8) and the Group NU centroid (5.0, 4.5) onto the axis $Y = .6X_1 + .8X_2$.

Before describing these refinements, however, let us see how well (or poorly) the two groups are differentiated along the dimension represented by the other linear combination cited above, $Y' = 10X_1 + 3X_2$, or its "representative,"

$$Y' = .96X_1 + .29X_2.$$

The Y' axis, the centroids of the two groups in the (X_1, X_2)-space, and their projections $\bar{Y}'_U$ and $\bar{Y}'_{NU}$ onto the Y' axis are shown in Figure 3. Comparing this with the preceding figure, it appears that $|\bar{Y}_U - \bar{Y}_{NU}|$ is greater than $|\bar{Y}'_U - \bar{Y}'_{NU}|$--that is, that the two groups are better differentiated along the Y axis than along the Y' axis. However, we must make the refinements alluded to above before coming to a definite conclusion.

The refinements consist in taking into consideration the variability of scores on the two linear combinations Y and Y' besides the means. If the standard deviations of Y are much larger than those of Y' in the two groups, this might offset the fact that $|\bar{Y}_U - \bar{Y}_{NU}|$ is larger than $|\bar{Y}'_U - \bar{Y}'_{NU}|$, since the magnitude order might be reversed when the differences are standardized.

(And it is the standardized difference that determines the magnitude of t, and hence of the F-ratio.) The variabilities of scores on each linear combination are reflected geometrically by the sharpness or diffuseness of the distribution of the projections of points (individuals) in each group

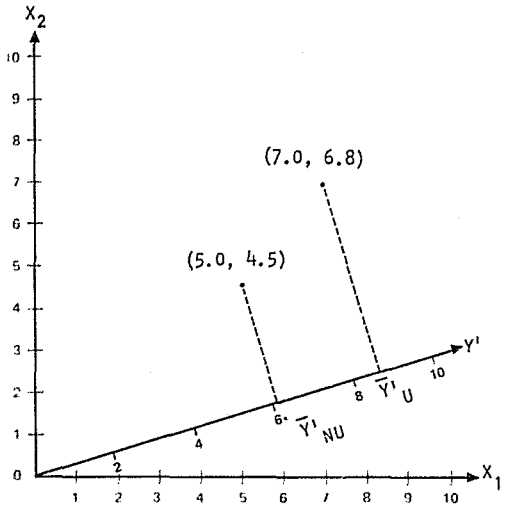


Figure 3. The projections of the Group U centroid (7.0, 6.8) and the Group NU centroid (5.0, 4.5) onto the axis $Y' = .96X_1 + .29X_2$.

onto that axis (Y or Y' as the case may be). Using just one group to avoid cluttering, we might have a scatter of points as shown in Figure 4. The projections of this scatter of points onto two axes Y' and Y'' have distributions with markedly different diffuseness: that for Y' is much more diffuse than that for Y''. (The previous axis Y has been replaced by a new axis Y'' in order to accentuate the difference in variability from Y'.) Thus, a numerically smaller difference on Y'' may represent a greater "real" difference than a numerically larger difference on Y'.

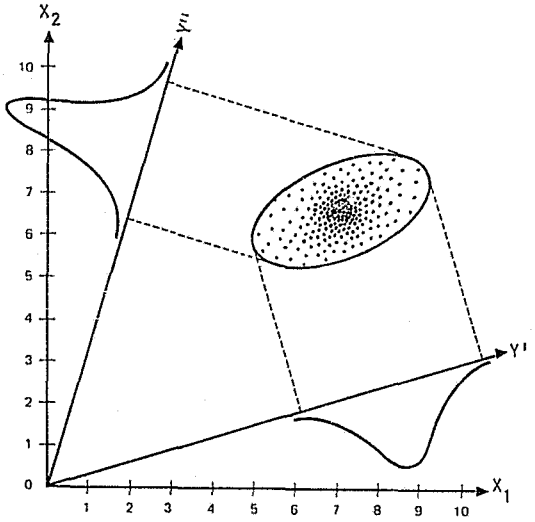


Figure 4. The projection of a scatter of points onto two axes, Y' and Y'', to illustrate the difference in diffuseness of the projected distribution depending on the orientation of the axis.

When the standard deviations besides the differences between the means are taken into account--that is, when standardized mean differences are compared--the degree of differentiation between two groups is measured by the amount of overlap of the two distribution curves: the smaller the overlap, the greater the differentiation. Thus, the problem of finding the linear combination of X_1 and X_2 that results in the largest possible F-ratio translates, geometrically, into the problem of finding a new axis such that the distributions of the projections of points in the two groups onto this axis have the smallest possible overlap. Figure 5 (in which the previous Y axis reappears) shows that, of the three axes Y, Y', Y'', the first one shows smallest overlap between the projected distributions of Groups U and NU.

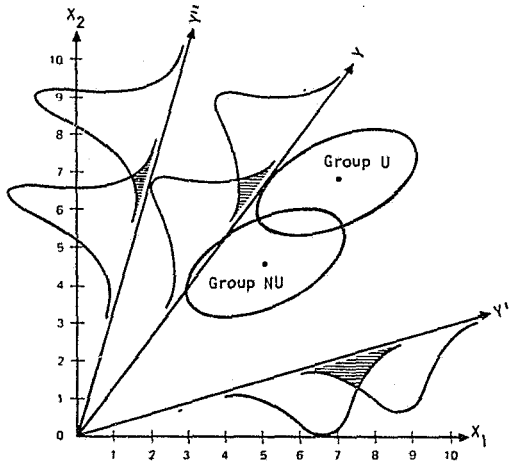


Figure 5. Projected distributions of Groups U and NU onto three axes, Y, Y' and Y'', showing the different degrees of overlap of the two distributions along the different axes.

THE ANALYTIC METHOD

Thus, in the simplest case of two groups measured on two variables, discriminant analysis can, in principle, be done geometrically by the "eyeballing" method. But, of course, in practice this would be very tedious and inefficient. It would be impossible if either the number of groups or the number of variables exceeds two. The analytic method which does the same thing as the geometric approach outlined above can be applied to cases with any number of groups and any number of variables. It calls for solving a matrix equation of the form

$$(\underline{W}^{-1}\underline{B} - \lambda \underline{I})\underline{v} = \underline{0},$$

where $\underline{W}$ is the within-groups sums-of-squares-and-cross-products (SSCP) matrix and $\underline{B}$ is the between-groups SSCP matrix. These matrices have as diagonal elements the within-groups sum of squares (SS_W) and the between-groups sum of squares (SS_B) in the usual ANOVA sense, respectively, for the several variables taken one at a time; their off-diagonal elements are the corresponding sums of cross products between pairs of variables. $\underline{W}^{-1}\underline{B}$ is obtained by finding the inverse of $\underline{W}$ and multiplying it by $\underline{B}$. The $\underline{v}$ is an unknown vector which, when solved for from the equation, gives the weights $v_1, v_2, \ldots, v_p$ to be applied to the original variables $X_1, X_2, \ldots, X_p$. The λ is an unknown scalar (an ordinary number) which, when solved for from the equation, gives a number proportional to the F-ratio--more specifically, the ratio SS_b/SS_w--for the linear combination defined by using $v_1, v_2, \ldots, v_p$ as the combining weights.

The $\underline{v}$ and λ obtained by solving the above equation are called an eigenvector and eigenvalue, respectively, of the matrix $\underline{W}^{-1}\underline{B}$. (Other terms used are characteristic vector and characteristic root; also latent vector and latent root.) Although we set out to find the linear combination resulting in the largest F-ratio, we end up getting several linear combinations, because the equation we solve yields several eigenvector-eigenvalue pairs. To be specific, the number of such solution pairs is equal to the number of original variables or one less than the number of groups, whichever is smaller; we denote this number by r. Thus, we have eigenvectors

$\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_r$ associated respectively with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$, which are arranged in descending order of magnitude. Consequently, the elements of $\underline{v}_1$, which we denote by $v_{11}, v_{12}, \ldots, v_{1p}$, form the weights of the linear combination

$$Y_1 = v_{11}X_1 + v_{12}X_2 + \ldots + v_{1p}X_p$$

that results in the largest possible F-ratio, or, equivalently, the largest SS_b/SS_w ratio, namely λ_1. This linear combination Y_1 is called the first discriminant function.

The second discriminant function,

$$Y_2 = v_{21}X_1 + v_{22}X_2 + \ldots + v_{2p}X_p,$$

using as combining weights the elements of $\underline{v}_2$, the eigenvector associated with the second largest eigenvalue λ_2, has the following property: it has the largest SS_b/SS_w ratio (λ_2) among all linear combinations of $X_1, X_2, \ldots, X_p$ that are uncorrelated with Y_1 in our total sample. The third discriminant function Y_3, which uses the elements of $\underline{v}_3$ as combining weights, has the largest SS_b/SS_w ratio (λ_3) among all linear combinations of the X's that are uncorrelated with both Y_1 and Y_2; and so on down the line.

Thus, solving the equation $(\underline{W}^{-1}\underline{B} - \lambda\underline{I})\underline{v} = \underline{0}$ yields r discriminant functions using as combining weights the elements of $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_r$, respectively. These eigenvectors and their associated eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$ (together with the $\underline{W}$ and $\underline{B}$ matrices) provide all the information necessary for the three aspects of discriminant analysis cited earlier.

THE THREE ASPECTS

Significance Testing

As mentioned earlier, this aspect is the same as that which is ordinarily treated under MANOVA. Therefore, we here confine ourselves to pointing out that all the information necessary for testing the null hypothesis that

$$\underline{\mu}_1 = \underline{\mu}_2 = \ldots = \underline{\mu}_K$$

(where the $\underline{\mu}$'s are the population centroids, and we have K populations) is contained in the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$ of $\underline{W}^{-1}\underline{B}$. The three commonly used test criteria--(i) Wilks' likelihood-ratio criterion, (ii) Roy's largest-root criterion, and (iii) Hotelling's trace criterion--are all simple functions of $\lambda_1, \lambda_2, \ldots, \lambda_r$. Thus, whenever we carry out a discriminant analysis, we automatically have the necessary quantities for carrying out the significance test of the corresponding MANOVA problem. It is for this reason that we can regard MANOVA as one aspect of discriminant analysis--although some authors prefer to speak of discriminant analysis as an adjunct to MANOVA. (The two views differ only in focus, and we shall not argue that one view is correct and the other, wrong.)

Explanation of Group Differences

By "explanation" here is not meant a causal or etiological explanation, but simply a parsimonious description in terms of the discriminant functions which constitute the "underlying factors" alluded to in the Introduction. As mentioned earlier, the number of discriminant functions is equal to the smaller of the two numbers, p (the number of original variables) and K-1 (where K is the number of groups). Usually the number of groups is much smaller than the number of variables, so that using K-1 discriminant functions to "explain" the group differences constitutes a considerable decrease of variables from the original p.

In practice, the number of discriminant functions that one needs to consider may be even smaller than K-1, because only the first few may have sufficiently large discriminant power--i.e., large SS_b/SS_w ratios. The procedure for determining the number of discriminant functions that are statistically significant is too involved to be described here.[1] For most practical purposes, a simple rule-of-thumb will suffice. This consists in examining what percentage of $T = \lambda_1 + \lambda_2 + \ldots + \lambda_r$ is accounted for by the first discriminant function by itself, the first two discriminant functions taken together, and so forth. That is, we may compute

$$\lambda_1/T, \ (\lambda_1 + \lambda_2)/T, \ (\lambda_1 + \lambda_2 + \lambda_3)/T, \text{ etc.}$$

and retain as many discriminant functions as are necessary to make this index adequately large (say .75 or greater).

The actual procedure for describing group differences in terms of the retained discriminant functions takes two forms. One is to examine the magnitudes and signs of the standardized discriminant function weights--that is, the elements of v_i each multiplied by the standard deviation of the particular variable--and thereby to determine what kind of person would tend to score high (and what kind, low) on each discriminant function. Then the groups which have large means on a given discriminant function are characterized as consisting predominantly of the kind of people who would score high on that function, and vice versa. (By "kind of person" here is meant a person with a particular pattern of scores on the descriptor variables.) The details of how this is done are best illustrated in the context of a real example, and are hence deferred to a subsequent section.

The second way for characterizing group differences more closely parallels the approach used in factor analysis to interpret the factors obtained. This is to examine the structure matrix, which is the matrix of correlations between the original variables and the retained discriminant functions. The interpretation of the structure matrix is described in chapter 9.

Classification

Historically, discriminant analysis has been associated with the problem of classification. Fisher first introduced two-group discriminant analysis as a tool for classifying an iris of doubtful species membership in one of two species on the basis of various botanical measurements. Since the number of discriminant functions is the smaller of the two numbers p and K-1, there is only one discriminant function in the two-group case. Classification in this case is a simple matter. We need only compute the discriminant function score for the individual to be classified (that is, the person of uncertain group membership, but who is known to be a member of one or the other of the two groups) and determine to which of the two group means on the discriminant function the individual's score is closer on the standardized scale.

The relevance of this phase of discriminant analysis to drug research is obvious. Assuming that we have antecedent measures on several variables for drug users and nonusers, we may construct a discriminant function to differentiate between these two groups. We may then obtain measures on these same variables for a new group of individuals, compute their discriminant function scores, and identify those who are likely to become users so that special, preventive steps can be taken for them.

When there are more than two groups (such as users of different kinds of drugs) among which we wish to differentiate and into one of which we want to classify an individual of uncertain group membership, the problem gets more complicated. Suppose there are three groups. We then have two discriminant functions, and must consider distances in a plane rather than along a single discriminant function axis. The two-dimensional counterpart of the standardized distance that was used in the two-group (single discriminant function) case to measure closeness of a given point to a group mean is called Mahalanobis' generalized distance. (In fact, this concept is applicable in any number of dimensions.) Although its algebraic definition is too technical for our purposes here, a general idea of what it means may be grasped through a geometric illustration.

In Figure 6 is shown a particular percent ellipse for some group--say the 90% ellipse--centered around the centroid of the group. This means that when the discriminant function score combinations of all individuals in that group are plotted as points on the (Y_1, Y_2) plane, 90 percent of these points will lie inside or on this ellipse. (Thus, the elliptical region is analogous to the corresponding percent interval in the univariate case, which extends from $\bar{Y} - 1.645s_Y$ to $\bar{Y} + 1.645s_Y$ when the distribution of Y is normal.) Given any percent ellipse, any two points on the ellipse are said to be equi-distant from the centroid M in the generalized-distance sense, regardless of the difference in their ordinary (Euclidean) distances from M. Thus, for example, points A and B are equi-distant from M in the generalized sense, even though A is closer to M than is B in the ordinary sense. Also, any point on or inside the ellipse is said to be closer to M in the generalized-distance sense than is any point outside the ellipse--regardless of which is close to M in the ordinary sense. Thus, point C is closer to M than is point D, in the generalized-distance sense, even though the opposite is true in the ordinary sense of distance.

Passing through any point A, there is one and only one applicable percent ellipse centered around the centroid M of a Group G, and this will enclose, say, P% of the points in that group. The same point A will lie on a unique P'% ellipse centered around the centroid M' of another group, G'. Then, depending on whether $P < P'$ or $P > P'$, point A is said to be closer (in the generalized-distance sense) to the centroid M of Group G or to the centroid M' of Group G'.

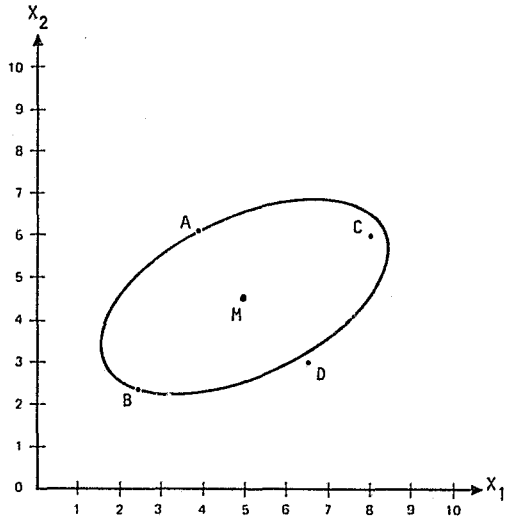


Figure 6. Illustration of proximity as measured by Mahalanobis' generalized distance: A and B are "equi-distant" from M (the center of the eliipse); C is "closer" to M than is D.

Using the measure of generalized distance just outlined, we can determine, for any point (representing an individual), which of three group centroids it is closest to in that sense. We would then classify the individual in that group. In practice all this is done algebraically without the need to construct ellipses, and the procedure generalizes to any number of discriminant functions.

There are other classification procedures besides that based on generalized distance--for example, a procedure based on probability of group membership. Discussion of these procedures would take us too far afield. The interested reader is referred to Overall and Klett (1972), Rulon et al. (1967), Tatsuoka (1971), or Tatsuoka (1974).

ADVANTAGES, LIMITATIONS, AND CAUTIONS

SIGNIFICANCE TESTING

The advantages of the first phase of discriminant analysis--significance testing--are described in the chapter on MANOVA. In brief, whenever multiple criterion variables are used (as will usually be the case in behavioral and social science research in general, and research on drug abuse in particular), MANOVA is the appropriate method for significance testing; using separate univariate F-tests (or t-tests in the case of two groups) for the separate criterion variables taken singly is to be avoided.

EXPLANATION OF GROUP DIFFERENCES

The second phase--explaining group differences parsimoniously--is all but unique to discriminant analysis, so it is difficult to discuss its advantages and disadvantages in comparison with alternative techniques. The only alternatives available, to my knowledge, are nonlinear and nonparametric extensions of the standard linear discriminant analysis--which assumes a multivariate normal distribution for the descriptor variables, although not in the second phase per se. While nonlinear discriminant analysis, utilizing higher-degree and product terms in the discriminant functions, will naturally improve group differentiation (i.e., yield a larger SS_b/SS_w ratio) in the sample at hand, the functions may not hold up as well on cross-validation. This point has been made by Bentler and Eichberg (1975) in connection with multiple regression analysis, and the comment holds with equal force in conjunction with discriminant analysis.

With regard to nonparametric discriminant analysis--i.e., discriminant analysis utilizing only ordinal data--it is obvious that this is a "last resort" with considerably decreased statistical power. It seems preferable to look very hard for transformations that will generate variables

that at least approximately follow a multivariate normal distribution. Such transformations are discussed in many textbooks on experimental design (e.g., Edwards, 1968; Kirk, 1969; Winer, 1971) for univariate analysis. Although transforming individual variables to univariate normality does not guarantee multivariate normality of their joint distribution, the chances of achieving approximate multivariate normality are improved.

CLASSIFICATION

The third aspect--classification--is probably the most important one in practical applications such as early detection of potential drug abusers, with a view to offering them counseling and preventive treatment. Unfortunately, it is also the phase that is most fraught with problems. Most of these problems, however, are not peculiar to classification procedures as an aspect of discriminant analysis, but are inherent in all classification methods whether or not they are preceded by the computation of discriminant functions. Indeed, the reduction of dimensionality by means of discriminant analysis is, to a large extent, inessential in this high-speed computer age, so long as classification is the sole purpose. (It was crucial at the time when Fisher first developed discriminant analysis as a tool for classification, for it was then practically infeasible to consider, say, 15 to 20 variables each time an individual was to be classified in one of two groups.) The computations for generalized distances can now be done almost as quickly using an original set of 15 to 20 predictor variables as they can be done with three or four discriminant function scores, once a set of preliminary, nonrecurring calculations (such as getting the inverses of the group covariance matrices) are done.

The most devastating of the problems relates not to classification procedures themselves (with or without discriminant functions), but rather to the design of the study. It will nevertheless be discussed here because there seems to be a tendency among many applied researchers to think that using a powerful analytic tool will somehow make up for a sloppily conducted experiment or carelessly gathered data. The problem in question concerns the chronological order of observing the "predictor" variables and defining the groups into which future individuals are to be classified. For instance, if personality measures are obtained on groups of drug users and non-users after they have been so identified, there is no guarantee that the users' personality pattern is conducive to drug usage rather than being a consequence thereof. If the latter is true, then the personality pattern found to be peculiar to the users' group may be totally useless in predicting before the fact who are likely to become drug users and who are not. Thus, measures antecedent to the individuals' subsequently becoming users or remaining nonusers are necessary as bases for prediction for future persons. Such data are admittedly difficult to collect, but nothing else will permit valid prediction. The seriousness of the problem of inadequate data bases is signalled by the fact that, in a recent survey of studies of drug abuse in adolescence, Braucht, Brakarsh, Follingstad and Berry (1973) could list only two (Jones, 1968; 1971) that actually used antecedent data in discussing personality differences between abusers (problem drinkers in these instances) and nonabusers.

The requirement of having descriptors measured prior to formation of the groups does not hold in the following situations: (1) when the attributes defining the groups accrue to the subjects at birth--e.g., race, nationality, parents' socioeconomic status, etc. In this case it is obvious that the pattern of descriptor-variable scores does not conduce to membership in the groups, but vice versa. (2) When we can be reasonably sure that membership in the different groups does not cause systematic differences in the descriptor variables to be used. (3) When the purpose of the discriminant analysis is simply to describe group differences (as in what is often called a status study), and there is no intention to use the discriminant functions for predictive purposes--i.e., to classify a future individual in one of the several groups on the basis of resemblance to current group members.

Other problems confronting classification procedures are more technical in nature, but the applied researcher needs to be aware of them in order to seek expert help when necessary. Broadly stated, these problems have to do with the choice of a "target function" to optimize (i.e., either to maximize or minimize). The generalized-distance approach outlined earlier minimizes the total proportion of cases misclassified--i.e., the percentage of false positives and false negatives all told. It may be desirable, instead, to minimize the total number of misclassifications. In that case, probabilities of group membership (rather than generalized distances from group centroids) have to be considered. If, furthermore, the relative costs of different types of misclassification (e.g., false positives versus false negatives) are to be taken into consideration, more complicated decision rules must be invoked. Thus, the researcher needs to have a clear idea of just what target function the researcher wishes to optimize, collect data accordingly, and select the appropriate classification rule for the purpose.

DATA FORMAT AND OTHER CONSTRAINTS

As indicated earlier, the ideal situation is when the descriptor variables follow a multivariate normal distribution in each group. Furthermore, the mathematical model for the significance testing phase requires that the population covariance matrices of all groups be identical. (This is the multivariate counterpart of the homogeneity of variances assumption in univariate ANOVA.) Fortunately, the significance tests are fairly robust (i.e., continue to be approximately valid) in the face of minor violations of these assumptions. When the departure from multivariate normality and/or equality of covariance matrices is drastic, suitable transformations need to be made.

The second phase of discriminant analysis, the construction of best-differentiating linear combinations of the original variables, does not require any distributional or equality-of-covariance-matrices assumption. The linear combination using as combining weights the elements of $\underline{v}_1$, the eigenvector associated with the largest eigenvalue of $\underline{W}^{-1}\underline{B}$, always will have the largest SS_b/SS_w ratio regardless of how the variables are distributed. Thus, for example, the inclusion of dichotomous variables poses no problem so far as this phase is concerned.

In the third phase, classification, the multivariate normality assumption again becomes important if the numerical values of the likelihoods or probabilities of membership in the various groups are to be taken seriously. The equality-of-covariance-matrices assumption is not quite as crucial, since classification rules that allow for unequal covariance matrices may be adopted. However, the meeting of this assumption (together with the multivariate normality assumption) does guarantee that classification results based on the discriminant functions will be identical with those based on the entire set of original variables.

Missing data always pose a problem when many variables are involved. Measures on the several variables are often taken at different testing sessions, and inevitably some people will be absent from some sessions. Persons with missing data on a substantial proportion of the variables (say 20% or more) should probably be eliminated from the sample. For those lacking scores on only a small proportion of the variables, several ways for supplying the missing scores are available. The simplest of these is to assign the mean of that variable in the group to which the individual belongs, and this is probably adequate for all practical purposes. Of course, any method for supplying missing data is applicable only in the first two aspects of discriminant analysis. In the third phase, no individual with any missing data should be considered for classification.

TERMINOLOGY

The field of discriminant analysis is unfortunately plagued with a lack of consistent terminology among its theoreticians and practitioners. Even the key term, "discriminant function," has two different usages. The sense in which we have been using this term here--as a linear combination which maximizes (absolutely or conditionally) the SS_b/SS_w ratio--is fairly standard within the behavioral and social sciences. However, mathematical statisticians and applied statisticians in the biological sciences tend to use the term in a different sense: as a linear (or quadratic) function which indexes the likelihood of an individual's being a member of a given group. Thus, there is one discriminant function, in this sense, for each group (instead of a total of K-1 functions in the sense we use the term). One calculates an individual's discriminant function value with respect to each group, and classifies the person in that group for which that person's score is the largest. Thus, this use of the term "discriminant function" focuses solely on the classification aspect of discriminant analysis. The reader is warned that the BMD computer program (Dixon, 1973) for discriminant analysis computes discriminant functions in this sense.

Another nonuniversal use of terms is that of "criterion" or "predictor" variables. This may be amazing (since the two terms are practically antonyms), but it becomes understandable when we realize that the descriptor variables play different roles in the different aspects of discriminant analysis. In the significance testing phase, it is natural to refer to the descriptor variables as "criterion variables," since they <u>are</u> the dependent variable in the analysis-of-various sense. On the other hand, in the classification phase, it is more natural to call these same descriptor variables "predictors," since they are used for predicting the group membership of a new individual. The reader should develop a flexible stance on seeing the same set of variables called by different names depending on the context.

Another problem in terminology is the likely confusion between "classification" and "taxonomy." Although the usage is not universal, most writers speak of "classification" when referring to the process of assigning an individual to one of several well-defined, preexisting groups. "Taxonomy" (or "typology"), on the other hand, usually refers to the process of <u>forming</u> groups where none were hitherto

recognized. Thus, when a psychiatrist diagnoses a patient to be a schizophrenic, he is engaging in classification, whereas if a researcher proposes to identify several distinct subtypes within what was hitherto treated as a single, undifferentiated class of schizophrenics, he is concerned with taxonomy.

ILLUSTRATIVE APPLICATIONS

NONDRUG RESEARCH: A NUMERICAL EXAMPLE[2]

Employees in three kinds of jobs in Trans-America Airlines were administered the Activity Preference Questionnaire (APQ) consisting of three bipolar scales: X_1 = Outdoor/Indoor preferences; X_2 = Gregarious/Solitary preferences; X_3 = Conservative/Liberal preferences. (A high score on each scale signifies that more activities of the first-named type were chosen compared to those of the second-named type.) The means of the three groups of employees on each of the three scales were as follows:

Group	n	Means on: X_1	X_2	X_3
Passenger Agents	85	12.59	24.22	9.02
Mechanics	93	18.54	21.14	10.14
Operations Control Persons	66	15.58	15.45	13.24

The within-groups SSCP matrix $\underline{W}$ (which requires the individual scores for computation) and the between-groups SSCP matrix $\underline{B}$ (which can be computed from the information given above) were:

$$\underline{W} = \begin{bmatrix} 3967.8301 & 351.6142 & 76.6342 \\ 351.6142 & 4406.2517 & 235.4365 \\ 76.6342 & 235.4365 & 2683.3164 \end{bmatrix}$$

and

$$\underline{B} = \begin{bmatrix} 1572.7441 & -773.0506 & 273.6214 \\ -773.0506 & 2889.3193 & -1405.9955 \\ 273.6214 & -1405.9955 & 691.6068 \end{bmatrix}$$

The next step is to compute the inverse $\underline{W}^{-1}$ of $\underline{W}$ and postmultiply it by $\underline{B}$ to get $\underline{W}^{-1}\underline{B}$, the matrix whose eigenvectors and eigenvalues we need. We find

$$\underline{W}^{-1}\underline{B} = \begin{bmatrix} .4133 & -.2462 & .0937 \\ -.2142 & .7063 & -.3418 \\ .1090 & -.5789 & .2851 \end{bmatrix}$$

The eigenvalues are obtained by solving the characteristic equation,

$$|\underline{W}^{-1}\underline{B} - \lambda\underline{I}| = 0,$$

which in this instance becomes

$$\lambda^3 - 1.4046\lambda^2 + .3502\lambda = 0.$$

The two (which is one less than the number of groups) nonzero roots of this equation are the desired eigenvalues:

$$\lambda_1 = 1.0805, \lambda_2 = .3241$$

The eigenvector associated with each of these eigenvalues is then computed by the method described, e.g., in Tatsuoka (1971, pp. 119-121). The results are:

$$\underline{v}_1 = \begin{bmatrix} -.3524 \\ .7331 \\ -.5818 \end{bmatrix} \quad \text{and} \quad \underline{v}_2 = \begin{bmatrix} .9145 \\ .1960 \\ -.3540 \end{bmatrix}$$

Hence, the two discriminant functions are

$$Y_1 = -.3524X_1 + .7331X_2 - .5818X_3$$

and

$$Y_2 = .9145X_1 + .1960X_2 - .3540X_3$$

Y_1 has the largest possible SS_b/SS_w ratio among all linear combinations of X_1, X_2 and X_3, the value being 1.0805 ($= \lambda_1$); Y_2 has the largest SS_b/SS_w ratio, .3241 ($=\lambda_2$), among all linear combinations of the X's that are uncorrelated with Y_1.

Next, we carry out the significance test. Normally, this would have preceded computing $\underline{v}_1$ and $\underline{v}_2$, because if no statistical significance is found, there would be no point in computing the discriminant functions. We computed $\underline{v}_1$ and $\underline{v}_2$ right after obtaining λ_1 and λ_2 to highlight the association between the eigenvalues and eigenvectors. We choose to use Wilks' likelihood-ratio criterion Λ among the three test criteria mentioned earlier. This is related to the eigenvalues λ_1 and λ_2 by the formula

$$\Lambda = 1/(1 + \lambda_1)(1 + \lambda_2),$$

so the numerical value for this example is $1/(2.0805)(1.3241) = .3630$. (Unlike most significance-test statistics, smaller values of Λ signify greater statistical significance.) The significance of this value may be tested by computing Bartlett's chi-square approximation:

$$V = -2.3026\left[N-1 - (p+K)/2\right]\log\Lambda$$

(where N is the total sample size, including all three groups; p is the number of variables, which is 3 in this example; and K is the number of groups). This is distributed approximately as a chi-square with p(K-1) degrees of freedom. For our numerical example, we have

$$\begin{aligned} V &= -2.3026\left[244 - 1 - (3 + 3)/2\right]\log.3630 \\ &= (-2.3026)(240)(-.4401) = 243.2, \end{aligned}$$

which, as a chi-square with p(K-1) = 6 degrees of freedom, is significant far beyond the .001 level.

The next question is whether to retain only the first discriminant function or both. The first function accounts for

$$\lambda_1/(\lambda_1 + \lambda_2) = 1.0805/(1.0805 + .3241 = .7693,$$

or about 77% of the total discriminatory power. It is probably a toss-up, whether to retain one or both functions. We choose to keep both for illustrative purposes. It will now be instructive to plot the centroids of the three groups on the two discriminant functions.

The means on the two discriminant functions for the three groups are calculated by substituting the means X_1, X_2 and X_3 for each group in the formulas for the discriminant functions given earlier. For example, the Group 1 mean on Y_1 is

$$\bar{Y}_{1,1} = -.3524(12.59) + .7331(24.22) - .5818(9.02) = 8.07.$$

Similar calculations for all three groups on both discriminant functions yield, as the three centroids in the (Y_1, Y_2) space, the following three points:

$$(8.07, 13.07), \qquad (3.06, 17.51), \qquad (-1.87, 12.59)$$

These are plotted in Figure 7 to give a visual impression of the way in which the three groups are differentiated (or separated) in the two-dimensional discriminant space. It is clear that the three groups are about equally separated along the first discriminant axis (Y_1), while Groups 1 and 3 are indistinguishable along the Y_2 axis, which sets these two groups apart from Group 2.

Our next task is to look for an interpretation of the two discriminant functions. Taking the option of examining the standardized discriminant weights, we need to multiply the weights earlier obtained (i.e., the elements of $\underline{v}_1$ and $\underline{v}_2$, respectively) by the within-groups standard deviations of the three variables. The latter are proportional to the square roots of the diagonal elements of the within-groups SSCP matrix $\underline{W}$ given earlier. We see that, in this example, the within-groups standard deviations for the three variables differ very little from one another (about 63:66:52), so we may take the raw score discriminant weights displayed earlier as they stand.

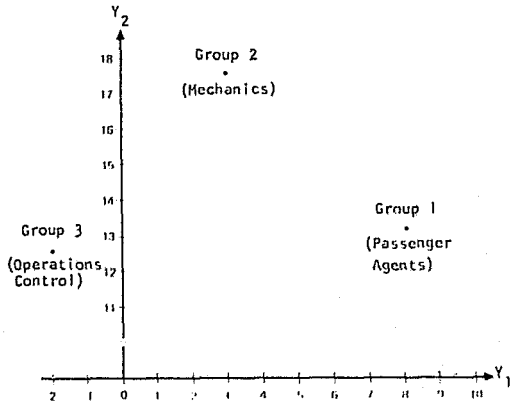


Figure 7. Centroids of the three groups in the discriminant function space.

Y_1 has a large positive weight (.73) for X_2 (Gregariousness) and substantial negative weights for X_3 (Conservativeness) and X_1 (Outdoor Interests). Thus, when we ask the question, "What type of person will score high on Y_1?", the answer is clearly, "A person who has gregarious tendencies, is liberal in outlook, and has indoor as against outdoor interests." However, since the weight for X_3 is about 1½ times as large (in absolute value) as that for X_1, we may conclude that Y_1 is essentially a "Gregarious-Liberal" dimension. The fact that the three groups go from high to low in the order, Passenger Agents, Mechanics, and Operations Control Persons (cf. Figure 7) on this dimension is consistent with our usual perception of the characteristics of these groups, given the above interpretation of the Y_1 dimension.

Examination of the relative weights for the three original variables on Y_2 shows that the latter is almost exclusively an "Outdoor Interests" factor. That the Mechanics (Group 2) are set quite apart from the other two groups on this dimension accords well with our stereotype concerning this group (remembering that we are talking about airline mechanics).

DRUG-RELATED RESEARCH

Most empirical studies concerning drug abuse involve (either solely or among other things) a comparison between users and nonusers or among users of different types of drugs in terms of demographic and/or personality variables. Hence, all of these studies could potentially have used discriminant analysis as one of their analytic tools. In reality, very few drug abuse studies seem to have employed this technique. In fact, in the limited search conducted in conjunction with this chapter, only one article was found that utilized discriminant analysis.

Krug and Henry (1974) administered the Sixteen Personality Factor Questionnaire (16 PF; Cattel, Eber and Tatsuoka, 1970), the Motivation Analysis Test (MAT; Cattel, Horn, Sweney and Radcliffe, 1964), and a questionnaire asking about the frequency and recency of usage of amphetamines, barbiturates, LSD, glues/aerosols, and marihuana, along with certain other pertinent biographical matters, to a total of 563 young men and women in their high-teens. Great care was taken to assure anonymity and immunity (under the Drug Abuse Prevention and Control Act of 1970) in order to secure frank, honest answers relating to drug use. (The authors suggest that they have been successful in this attempt, since a substantially larger percentage of the sample, 30% or 171 persons, admitted to using at least one drug than found in most other studies.)

The 16 PF measures 16 factor-analytically derived dimensions of the normal adult personality. The MAT, based on Cattell's work on the objective analysis of human motivation, measures drive level (the "unintegrated component," abbreviated U) and drive satisfaction (the "integrated component," or I) in each of ten behavioral areas such as Career Interest (Ca) and Attachment to Home (Ho). Thus, there were a total of 36 descriptor variables.

Among other things, the authors investigated sex differences in patterns of drug usage, correlations between use of different pairs of drug types (e.g., a female who uses glues and aerosols is more likely also to use amphetmaines than she is to use LSD as well). However, we shall here review only their discriminant analysis to differentiate the users from nonusers in terms of the personality and motivational variables.

The authors chose to use stepwise discriminant analysis (analogous to stepwise multiple regression), in which the computer program selects the predictor variables to include, one by one, on the basis of their contribution to increasing the SS_b/SS_w ratio. This method, as against the standard way of including all the variables at once, is preferable when the number of variables is very large, especially if the sample sizes are not commensurately large (say, if the smallest group size is less than three times the number of variables.)

The single discriminant function (since there are only two groups), with the combining weights suitably rescaled and an additive constant included so that the resulting scores range from 1 through 10 with a mean of 5.5 and a standard deviation of 2.0 in the total sample, was found to be

$$Y = -.16G + .15I + .22Q_1 - .12Ho(U) - .18Na(U) + .19SE(U) + .18Ho(I) + .17Na(I) - .21SE(I) + .16SS(I) + .30Pg(I) + 1.72.$$

The variables entering into this discriminant function (numbering 11, or only about one-third of the entire set of variables, it should be noted) are identified as follows: G = Conscientiousness; I = Tendermindedness; Q_1 = "Experimentingness" (socially); Ho(U) = Attachment to Home (Unintegrated component); Na(U) = Self-Indulgent Satisfaction (Narcism), Unintegrated; SE(U) = Conscience Development (Superego), Unintegrated; Ho(I) = Attachment to Home (Integrated component); Na(I) = Narcism, Integrated; SE = Superego, Integrated; SS = Concern for Social Reputation (Self Sentiment), Integrated; Pg = Destructive Drive (Pugnacity), Integrated. The first three of these are from the 16PF, and the rest are scales of the MAT. All variables are expressed on a standardized scale called the sten (for "standard ten") in which the mean is 5.5 and the standard deviation is 2.0 in the population. (The reason standardized rather than raw scores are used is that the sample includes both men and women, and a given raw score usually represents a different degree of extremeness on the personality dimension, depending on the sex of the subject. The sten scores are based on separate norm tables for the two sexes.)

Before going on to an interpretation of the discriminant function, let us make a few comments about its formal characteristics. First of all, it should be pointed out that the function is so oriented that drug users tend to get high scores on Y, and nonusers. low scores. Next, the presence of an additive constant (1.72) may have puzzled the reader, for in our previous discussion none was present. This is a completely arbitrary matter, since an additive constant does not affect either SS_b or SS_w (both being based on deviations from means), and hence also the ratio SS_b/SS_w. In this study the additive constant was included (besides the weights' being proportionally rescaled) to force the discriminant function scores to have a sten scale, just as the predictor variables do. Similar adjustments may be made if the researcher wishes to have the discriminant scores come out on a T-scale, with a mean of 50 and a standard deviation of 10 and so on.

Now to proceed with the interpretation, we recall that one way is to ask the question, "What kind of person--i.e., a person with what sort of personality pattern--will tend to score high on the discriminant function?" Since, as mentioned above, Y is so oriented that drug users tend to get high scores on it, this amounts to asking what sort of personality (and motivational) pattern tends to go along with drug use. Examination of the magnitudes and signs of the weights associated with the respective variables gives us an answer to this question. Note that, in this case, we need not convert the weights to standardized form since all variables already are measured on a standard (sten) scale, with common standard deviation 2.0 in the population.

Looking first at the three 16PF variables, we note the signs of the weights (which are nearly equal in magnitude) are -, +, + for scales G, I, and Q_1 respectively. Referring to the descriptions of the variables given earlier, we see that a person who scores high on Y (i.e., one who tends to be a drug user) is low on conscientiousness and high on "tendermindedness" and social experimentation (i.e., countering established social norms or mores). This syndrome seems to fit our stereotype of the drug user--except, perhaps, the "tenderminded" aspect. It should be pointed out that this term is used in the 16 PF in a sense not necessarily in complete agreement with the most common usage. It may help to note that the term is used as an antonym to "tough-mindedness" in the sense of "rejecting illusions." A person high on I tends to be attention-seeking and flighty.

Interpretation in terms of the MAT scales is more complicated in that eight scales are represented, and there are unintegrated (drive level) and integrated (drive satisfaction) components. Three behavioral areas are represented both on the unintegrated and integrated components, while two occur only in the integrated component. Let us take care of the latter first. According to the weights (both positive) for SS(I) and Pg(I), those whose drives for high "social reputation" and pugnacity (destructiveness) are satisfied tend to score high on Y (i.e., tend to be drug users). The second of these may fit our stereotype, but the first will probably run counter to it. However, if we recall that we are dealing exclusively with youths in their high-teens, we realize that "social reputation" in this stratum is possibly enhanced by a stance of rejecting the established mores.

We now come to the three behavioral areas that are represented both in their unintegrated and integrated components on the discriminant function (always with opposing signs for the two components, it should be noted). The high-Y person (who tends to be a drug user) is low on Ho(U), drive for home attachment, but high on Ho(I)--i.e., satisfaction of this drive. This is not as paradoxical as it may seem at first, for the person with a low drive for home attachment will easily be satisfied with the (low) level of attachment he has. In the narcistic area, high degree of satisfaction (i.e., a facile satisfaction of the drive) may go along with a low drive as such. Finally, in the superego (SE) area, a high level of aspiration is often associated with a low level of achievement (satisfaction of the drive), and may lead to drug use as an escape. The authors also cite evidence that a high unintegrated component score coupled with a low integrated score (a combination that would be conducive to a high Y-score in this instance) is indicative of dynamic conflict in that area.

In sum, it may be concluded that the discriminant analysis in this study yielded reasonable and intuitively appealing results, which in turn suggests that the definition of the user and nonuser groups was a valid one. In many ways, this study is a highly sophisticated one from the methodological standpoint, but unfortunately it, too, suffers from the almost universal problem of not having descriptors measured before the onset of drug use. Thus, intuitively appealing as the discriminant function is, it can only be regarded as descriptive, but not necessarily predictive, of differences between users and nonusers.

NOTES

[1]The interested reader may refer to Cooley and Lohnes (1971); Rao (1952); Rulon, Tiedeman, Tatsuoka and Langmuir (1967); or Tatsuoka (1971).

[2]Since real research problems using discriminant analysis usually involve a large number of variables, we first present a contrived example in a nondrug context, and defer a real example to the next section. Source of data: Rulon, Tiedeman, Tatsuoka and Langmuir (1967), by permission of the publishers, John Wiley and Sons.

RESOURCES AND REFERENCES

REFERENCES TO COMPUTER PROGRAMS TO PERFORM DISCRIMINANT ANALYSIS

Cooley, W.W., and Lohnes, P.R. *Multivariate Data Analysis*. New York: Wiley, 1971.

Dixon, W.J., ed. *Biomedical Computer Programs*. Berkeley, California: University of California Press, 1973.

Dixon, W.J., ed. *BMD P Biomedical Computer Programs*. Berkeley, California: University of California Press, 1973.

Finn, J.D. *MULTIVARIANCE: Univariate and Multivariate Analysis of Variance, Covariance, and Regression*. Ann Arbor, Michigan: National Educational Resources, Inc., 1972.

Nie, N.H.; Hull, C.H.; Jenkins, J.G.; Steinbrenner, K.; and Bent, D.H. *Statistical Package for Social Sciences (SPSS)*. New York: McGraw-Hill, 1975.

University of Illinois Computing Services Office. *SOUPAC Program Descriptions: Statistically Oriented Users Programming and Consulting*. Urbana, Illinois: University of Illinois, 1975.

Veldman, D.J. *FORTRAN Programming for the Behavioral Sciences.* New York: Holt, Rinehart and Winston, 1967.

DOCUMENTS CITED IN THIS CHAPTER

Bentler, P.M., and Eichberg, R.H. A social psychological approach to substance abuse construct validity: Prediction of adolescent drug use from independent data sources. In: Lettieri, Dan J., ed. *Predicting Adolescent Drug Abuse: A Review of Issues, Methods, and Correlates,* NIDA Research Issues Series, vol. 11. Washington, D.C.: Government Printing Office, December 1975.

Braucht, G.N.; Brakarsh, D.; Follingstad, D.; and Berry, K.L. Deviant drug use in adolescence: A review of psychological correlates. *Psychological Bulletin,* 79:92-106, 1973.

Cattell, R.B.; Eber, H.W.; and Tatsuoka, M.M. *Handbook for the 16 PF.* Champaign, Illinois: Institute for Personality and Ability Testing, 1970.

Cattell, R.B.; Horn, J.L.; Sweney, A.B.; and Radcliffe, J.A. *The Motivation Analysis Test.* Champaign, Illinois: Institute for Personality and Ability Testing, 1964.

Cooley, W.W., and Lohnes, P.R. *Multivariate Data Analysis.* New York: Wiley, 1971.

Dixon, W.J., ed. *Biomedical Computer Programs (BMD).* Berkeley, California: University of California Press, 1973.

Edwards, A.L. *Experimental Design in Psychological Research.* 3rd ed. New York: Holt, Rinehart and Winston, 1968.

Fisher, R.A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics,* 8: 376-386, 1936.

Jones, M.C. Personality correlates and antecedents of drinking patterns in adult males. *Journal of Consulting and Clinical Psychology,* 32:2-12, 1968.

Jones, M.C. Personality antecedents and correlates of drinking patterns in women. *Journal of Consulting and Clinical Psychology,* 36:61-69, 1971.

Kirk, R.E. *Experimental Design: Procedures for the Behavioral Sciences.* Belmont, California: Brooks/Cole, 1969.

Krug, S.E., and Henry, T.J. Personality, motivation, and adolescent drug use patterns. *Journal of Counseling Psychology,* 21:440-445, 1974.

Overall, J.E., and Klett, C.J. *Applied Multivariate Analysis.* New York: McGraw-Hill, 1972.

Rao, C.R. *Advanced Statistical Methods in Biometric Research.* New York: Wiley, 1967.

Rulon, P.J.; Tiedeman, D.V.; Tatsuoka, M.M.; and Langmuir, C.R. *Multivariate Statistics for Personnel Classification.* New York: Wiley, 1967.

Tatsuoka, M.M. *Multivariate Analysis: Techniques for Educational and Psychological Research.* New York: Wiley, 1971.

Tatsuoka, M.M. *Classification Procedures: Profile Similarity.* Champaign, Illinois: Institute for Personality and Ability Testing, 1974.

Winer, B.J. *Statistical Principles in Experimental Design.* Rev. ed. New York: McGraw-Hill, 1971.

SELECTED REFERENCES TO DOCUMENTS UTILIZING DISCRIMINANT ANALYSIS

Anderson, G.J.; Walberg, H.J.; and Welch, W.W. Curriculum effects on the social climate of learning: A new representation of discriminant functions. *American Educational Research Journal,* 6:315-328, 1969.

Baggaley, A.R., and Campbell, J.P. Multiple-discriminant analysis of academic curricula by interest and aptitude variables. *Journal of Educational Measurement,* 4:143-149, 1967.

Baggaley, A.R.; Isard, E.S.; and Sherwood, E.J. Discrimination of academic curricula by the runner studies of attitude patterns--College Form. *Measurement and Evaluation in Guidance,* 3:41-44, 1970.

Bledsoe, J.C. The prediction of teacher competence: A comparison of two multivariate statistical techniques. *Multivariate Behavioral Research,* 8:3-22, 1973.

Cohen, D. Differentiating motivations underlying vocational choice. *Journal of Educational Research,* 64:229-234, 1971.

Cornfield, J. Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: A discriminant function analysis. *Federation Proceedings,* 21:58-61, 1962.

Goldman, R.D., and Warren, R. Discriminant analysis of study strategies connected with college grade success in different major fields. *Journal of Educational Measurement,* 10:39-47, 1973.

Grizzle, J.E., and Allen, D.M. Analysis of growth and dose response curves. *Biometrics,* 25: 357-381, 1969.

Harper, A.E. Discrimination between matched schizophrenics and normals by the Wechsler-Bellevue Scale. *Journal of Consulting Psychology,* 4:351-357, 1950.

Keenan, C.B., and Holmes, J.E. Predicting graduation withdrawal and failure in college by multiple discriminant analysis. *Journal of Educational Measurement,* 7:91-95, 1970.

Krug, S.E., and Henry, T.J. Personality, motivation, and adolescent drug use patterns. *Journal of Counseling Psychology,* 21:440-445, 1974.

Loeb, J., and Bowers, J. Programs of study as a basis for selection placement and guidance of college students. *Journal of Educational Measurement,* 7:91-95, 1970.

Maw, W.H., and Magoon, A.J. The curiosity dimension of fifth-grade children: A factorial discriminant analysis. *Child Development,* 42:2023-2031, 1972.

McHugh, R.B., and Apostolakos, P.C. Methodology for the comparisons of clinical with actuarial prediction. *Psychological Bulletin,* 56:301-308, 1959.

McQuarrie, D., and Groteluеschen, A. Effects of verbal warning upon misapplication of a rule of limited applicability. *Journal of Educational Psychology,* 62:432-438, 1971.

Oyama, T., and Tatsuoka, M.M. Prediction of relapse in pulmonary tuberculosis: An application of discriminant analysis. *American Review of Tuberculosis,* 73:472-484, 1956.

Porebski, O.R. Discriminatory and canonical analysis of technical college data. *British Journal of Mathematical and Statistical Psychology,* 19:215-236, 1966.

Selover, R.B. A study of the sophomore testing program at the University of Minnesota. *Journal of Applied Psychology,* 26:296-307; 456-467; 587-593; 1942.

Stahmann, R.F. Predicting graduation major field from freshman entrance data. *Journal of Counseling Psychology,* 16:109-113, 1969.

Truett, J.; Cornfield, J.; and Kannel, W. A multivariate analysis of the risk of coronary heart disease in Framingham. *Journal of Chronic Disease,* 20:511-524, 1967.

SELECTED REFERENCES TO DOCUMENTS THAT DESCRIBE DISCRIMINANT ANALYSIS AND ITS DERIVATION

Bock, R.D. *Multivariate Statistical Methods in Behavioral Research.* New York: McGraw-Hill, 1975.

Cacoullos, T., ed. *Discriminant Analysis and its Applications.* New York: Academic Press, 1973.

Cooley, W.W., and Lohnes, P.R. *Multivariate Data Analysis.* New York: Wiley, 1971.

Harris, R.J. *A Primer of Multivariate Statistics.* New York: Acadmic Press, 1975.

Huberty, C.J. Discriminant analysis. *Review of Educational Research*, 45:543-598, 1975.

Lachenbruch, P.A. *Discriminant Analysis.* New York: Hafner, 1975.

Overall, J.E., and Klett, C.J. *Applied Multivariate Analysis.* New York: McGraw-Hill, 1972.

Press, S.J. *Applied Multivariate Analysis.* New York: Holt, Rinehart and Winston, 1972.

Tatsuoka, M.M. *Discriminant Analysis: The Study of Group Differences.* Champaign, Illinois: Institute for Personality and Ability Testing, 1970.

Timm, N.H. *Multivariate Analysis with Applications in Education and Psychology.* Belmont, California: Brooks/Cole, 1975.

Van de Geer, J.P. *Introduction to Multivariate Analysis for the Social Sciences.* San Francisco: Freeman, 1971.

☆U.S. GOVERNMENT PRINTING OFFICE: 1978-270-308/968

Volumes in the series are available for purchase from the Superintendent of Documents, Government Printing Office, Washington, D.C. 20402. When ordering, please give title and stock number of the publication.

DRUGS AND EMPLOYMENT. Attitudes toward drug use by employers and employees; the effects of drug use on work performance, hiring and firing practices; employer rehabilitation programs. Specific studies on drug use in medicine, sports, and aviation. 31 abstracts. Stock No. 1724-00424 ($1.80)

DRUGS AND SEX. The effects of drugs on sexual behavior, correlations and relationships among a variety of populations. Arranged by drug type: multi-drug, marihuana, amphetamines, LSD, heroin and methadone. 24 abstracts. Stock No. 1724-00425 ($1.45)

DRUGS AND ATTITUDE CHANGE. What people believe and how they learn about drugs. Users' and nonusers' attitudes toward drugs; sources of drug information; the role of the media, drug education programs, and communications, as they influence attitudes and attitude change. 44 abstracts. Stock No. 1724-00426 ($2.25)

DRUGS AND FAMILY/PEER INFLUENCE. Classification of peer group types; comparison of drug-using and other social "deviant" behavior among youth; prediction of drug use, particularly marihuana, and the family dynamics of heroin users. 36 abstracts. Stock No. 1724-00427 ($2.10)

DRUGS AND PREGNANCY. LSD: literature reviews, chromosome and teratogenesis studies. Heroin: the heroin-addicted mother-child dyad and neonatal withdrawal management. Overviews on genetics and mutational hazards. 52 abstracts. Stock No. 1724-00428 ($2.70)

DRUGS AND DEATH. Epidemiological studies of drug-related deaths, with an emphasis on opiates; methods of classifying and reporting drug-related deaths; and the relationship between drugs, suicide, and homicide. 49 abstracts. Stock No. 1724-00429 ($2.25)

DRUGS AND ADDICT LIFESTYLES. Lifestyle and motivation of the committed heroin user. Descriptive and comparative studies on the incidence, demographics, and characteristics of heroin-using populations; citations to books by and about addicts. 83 abstracts. Stock No. 1724-00430 ($3.30)

A COCAINE BIBLIOGRAPHY — NONANNOTATED. All aspects of cocaine use, including coca. Covers worldwide literature; research reports, journal articles, books, news sources and media. Arranged by decade, language, and subject. 1800 citations. Stock No. 1724-00431 ($2.00)

DRUG THEMES IN SCIENCE FICTION. By Robert Silverberg. Explores unreality as an indicator of reality; selected science fiction stories about drugs of the future and their impact on society. Includes an annotated bibliography for 1900-1973. Stock No. 1724-00432 ($1.20)

DRUG THEMES IN FICTION. By Digby Diehl. Surveys selected works of fiction with drug-related thematic content which reflect popular American attitudes toward drugs. Includes an annotated bibliography. Stock No. 1724-00433 ($1.05)

PREDICTING ADOLESCENT DRUG ABUSE. The most recent thinking on the problems and intricacies surrounding the prospect of predicting drug-abusing behaviors, particularly among adolescents. Discusses: general conceptual issues, nosological and clinical approaches, metodological strategies; intrapersonal, behavioral, and interpersonal variables and correlates; longitudinal design; and developmental models. 18 articles; 361 pp. Stock No. 1724-00496 ($4.90)

DRUG ABUSE INSTRUMENT HANDBOOK. A sourcebook containing over 2,000 representative items from 40 instruments used in drug abuse research. Includes demographic, interpersonal, intrapersonal, and drug items, plus detailed summaries describing the instruments utilized. A guide to obtaining and developing instruments. Stock No. 017-024-00533-9 ($4.65)

DATA ANALYSIS STRATEGIES AND DESIGNS FOR SUBSTANCE ABUSE RESEARCH. The latest data analysis and methodological strategies and their application in psychosocial substance abuse research. Includes: Automatic Interaction Detection, Actuarial Prediction, Cluster and Typological Analysis: Single-Organism Designs: Longitudinal Designs: Path Analysis: Factor Analysis: Multiple Regression and Correlation Analysis: Multivariate Analysis of Variance: and Discriminant Analysis. Written by noted authorities in each field. 10 articles. Stock No. 017-024-00562-2 ($3.00)

DRUGS AND PERSONALITY. Personality correlates and predicators of non-opiate drug use, particularly among adolescents. Discusses: psychological symptoms, locus of control, self-esteem, and the use of psychological tests, particularly the Minnesota Multiphasic Personality Inventory. 59 abstracts. Stock No. 017-024-00531-2 ($2.00)

COCAINE — SUMMARIES OF PSYCHOSOCIAL RESEARCH. Theories and research on human cocaine use from the turn of the century to the present in North America, South America and Europe. Also incidence studies noting the extent of cocaine use in contemporary society. 69 abstracts. Stock No. 017-024-00564-9 ($2.10)

DRUGS AND CRIME. The relationship of drug use and concomitant criminal behavior. Divided into sections on reviews and theories, non-opiate drug use, the opiates, delinquency, the female drug user, the impact of treatment modalities, and the economics of drugs and crime. 108 abstracts. Stock No. 017-024-00556-8 ($3.45)

DRUG USERS AND THE CRIMINAL JUSTICE SYSTEM. Research on drug-related offenses, and the enforcement of drug laws; the effect of the criminal justice system on the drug user. Divided into two sections: (1) research on drug laws — their effectiveness and their enforcement; (2) research on the use of compulsion in the treatment of addicts. 66 abstracts. Stock No. 017-024-00629-7 ($3.00)

DRUGS AND PSYCHOPATHOLOGY. The relationship of drug use to severe psychological disturbance and antisocial behavior. A companion volume to *Drugs and Personality*, this volume includes research on opiates and nonopiates in all age groups. Stock No. 017-024-00630-1 ($3.00)

DRUG USERS AND DRIVING BEHAVIORS. The use of nonmedical drugs and their effects on driving. Includes research on the effects of drug use on the physical functions associated with driving, incidence studies of driving arrests, and simulated driving studies. 76 abstracts. Stock No. 017-024-00640-8 ($3.25)

DRUGS AND MINORITIES. Research on the use of drugs among racial and ethnic minorities, particularly blacks; Mexican Americans, Puerto Ricans, and other Spanish surnamed Americans; Asian and Native Americans. Includes studies dealing with minority populations as well as with the racial/ethnic patterns of drug use among general populations. 93 abstracts. Available in December 1977.