CHAPTER 1

The Nature of Multivariate Data Analysis

1.1 INTRODUCTION

Stripped to their mathematical essentials, multivariate methods represent a blending of concepts from matrix algebra, geometry, the calculus, and statistics. In function, as well as in structure, multivariate techniques form a unified set of procedures that can be organized around a relatively few prototypical problems. However, in scope and variety of application, multivariate tools span all of the sciences.

This book is concerned with the mathematical foundations of the subject, particularly those aspects of matrix algebra and geometry that can help illuminate the structure of multivariate methods. While behavioral and administrative applications are stressed, this emphasis reflects the background of the author more than any belief about special advantages that might accrue from applications in these particular fields.

Multivariate techniques are useful for:

1. discovering regularities in the behavior of two or more variables;
2. testing alternative models of association between two or more variables, including the determination of whether and how two or more groups (or other entities) differ in their "multivariate profiles."

The former pursuit can be regarded as exploratory research and the latter as confirmatory research. While this view may seem a bit too pat, multivariate analysis *is* concerned with both the discovery and testing of patterns in associative data.

The principal aim of this chapter is to present motivational material for subsequent development of the requisite mathematical tools. We start the chapter off on a somewhat philosophical note about the value of multivariate analysis in scientific research generally. Some of the major characteristics of multivariate methods are introduced at this point, and specific techniques are briefly described in terms of these characteristics.

Application of multivariate techniques is by no means confined to a single discipline. In order to show the diversity of fields in which the methods have been applied, a number of examples drawn from the behavioral and administrative sciences are briefly described. Comments are also made on the trends that are taking place in multivariate analysis itself and the implications of these developments for future application of the methodology.

We next turn to a description of three small, interrelated problems that call for multivariate analysis. Each problem is described in terms of a common, miniature data

bank with integer-valued numbers. As simple as the problems are, it turns out that developing the apparatus necessary to solve them covers most of the mathematical concepts in multivariate analysis that constitute the rest of the book.

1.2 MULTIVARIATE METHODS IN RESEARCH

It is difficult to imagine any type of scientific inquiry that does not involve the recording of observations on one or more types of objects. The objects may be things, people, natural or man-made events. The selected objects–white rats, model airplanes, biopsy slides, x-ray pictures, patterns of response to complex stimulus situations, ability tests, brand selection behavior, corporate financial activities–vary with the investigator's discipline. The process by which he codifies the observations does not.

Whatever their nature, the objects themselves are never measured in total. Rather, what is recorded are observations dealing with *characteristics* of the objects, such as weight, wind velocity, cell diameter, location of a shadow on the lung, speed or latency of response, number of correctly answered questions, specific brand chosen, previous year's sales, and so on. It is often the case that two or more characteristics (e.g., weight, length, and heartbeat) will be measured at the same time on each object being studied. Furthermore, it would not be unusual to find that the measured characteristics were associated in some way; that is, values taken on by one variable are frequently related to values taken on by another variable.

As a set of statistical techniques, multivariate data analysis is strategically neutral. Techniques can be used for many purposes in the behavioral and administrative sciences–ranging from the analysis of data obtained from rigidly controlled experiments to teasing out relationships assumed to be present in a large mass of survey-type data. What can be said is that multivariate analysis is concerned with *association among multiple variates* (i.e., many variables).[1]

Raymond Cattell (1966) has put the matter well. Historically, empirical work in the behavioral sciences–more specifically, experimental psychology–has reflected two principal traditions: (a) the manipulative, typically bivariate approach of the researcher viewed as controller and (b) the nonmanipulative, typically multivariate approach of the researcher viewed as observer.

Cattell points out three characteristics that serve to distinguish these forms of strategic inquiry:

1. bivariate versus multivariate in the type of data collected,
2. manipulative versus noninterfering in the degree of control exercised by the researcher,
3. simultaneous versus temporally successive in the time sequence in which observations are recorded.

[1] Analysis of bivariate data can, of course, be viewed as a special case of multivariate analysis. However, in this book our discussion will emphasize association among more than two variables. One additional point–some multivariate statisticians restrict the term *multivariate* to cases involving more than a single criterion variable. Here, we take a broader view that includes multiple regression and its various extensions as part of the subject matter of multivariate analysis.

In recent years, bivariate analysis and more rigid forms of controlled inquiry have given way to experiments and observational studies dealing with a comparatively large number of variables, not all of which may be under the researcher's control. However, if one takes a broad enough view of multivariate data analysis, one that includes bivariate analysis as a special case, then the concepts and techniques of this methodology can be useful for either stereotype. Indeed, Cattell's definition of an experiment as:

> . . .A recording of observations, quantitative or qualitative, made by defined operations under defined conditions, and designed to permit non-subjective evaluation of the existence or magnitude of relations in the data. It aims to fit these relations to parsimonious models, in a process of hypothesis creation or hypothesis checking, at least two alternatives being logically possible in checking this fit. . . . (p. 9)

says quite a bit about the purview of multivariate analysis. That is, the process of scientific inquiry should embrace the search for naturalistic regularities in phenomena as well as their incorporation into models for subsequent testing under changed conditions. And in this book we shall be as much, if not more so, interested in using multivariate analysis to aid the process of discovery (hypothesis creation) as to aid the process of confirmation (hypothesis testing).

The heart of any multivariate analysis consists of the data matrix, or in some cases, matrices.[2] The data matrix is a rectangular array of numerical entries whose informational content is to be summarized and portrayed in some way. For example, in univariate statistics the computation of the mean and standard deviation of a single column of numbers is often done simply because we are unable to comprehend the meaning of the entire column of values. In so doing we often (willingly) forego the full information provided by the data in order to understand some of its basic characteristics, such as central tendency and dispersion. Similarly, in multivariate analysis we often use various summary measures–means, variances, covariances–of the raw data. Much of multivariate analysis is concerned with placing in relief certain aspects of the association among variables at the expense of suppressing less important details.

In virtually all applied studies we are concerned with variation in some characteristic, be it travel time of a white rat in a maze or the daily sales fluctuations of a retail store. Obviously, if there is no variation in the characteristic(s) under study, there is little need for statistical methods.

In multivariate analysis we are often interested in accounting for the variation in one variable or group of variables in terms of *covariation* with other variables. When we analyze associative data, we hope to "explain" variation according to one or more of the following points of view:

1. determination of the nature and degree of association between a set of *criterion* variables and a set of *predictor* variables, often called "dependent" and "independent" variables, respectively;
2. finding a function or formula by which we can estimate values of the criterion variable(s) from values of the predictor variable(s)–this is usually called the *regression* problem;

[2] Much of this section is drawn from Green and Tull (1975).

3. assaying the statistical "confidence" in the results of either or both of the above activities, via tests of statistical significance, placing confidence intervals on parameter estimates, or other ways.

In some cases of interest, however, we have no prior basis for distinguishing between criterion and predictor variables. We may still be interested in their interdependence as a whole and the possibility of summarizing information provided by this interdependence in terms of other variables, often taken to be linear composites of the original ones.

1.3 A CLASSIFICATION OF TECHNIQUES FOR ANALYZING ASSOCIATIVE DATA

The field of associative data analysis is vast; hence it seems useful to enumerate various descriptors by which the field can be classified. The key notion underlying the classification of multivariate methods is the *data matrix*. A conceptual illustration is shown in Table 1.1. We note that the table consists of a set of objects (the m rows) and a set of measurements on those objects (the n columns). Cell entries represent the value X_{ij} of object i on variable j. The objects are any kind of entity with characteristics capable of being measured. The variables are characteristics of the objects and serve to define the objects in any specific study. The cell values represent the state of object i with respect to variable j. Cell values may consist of nominal, ordinal, interval, or ratio-scaled measurements, or various combinations of these, as we go across columns.

By a nominal scale we mean categorical data where the only thing we know about the object is that it falls into one of a set of mutually exclusive and collectively exhaustive categories that have no necessary order vis à vis one another. Ordinal data are ranked data where all we know is that one object i has more, less, or the same amount of some variable j than some other object i'. Interval scale data enable us to say how much more one object has than another of some variable j (i.e., intervals between scale values are meaningful). Ratio scale data enable us to define a natural origin (e.g., a case in which

TABLE 1.1

Illustrative Data Matrix

	Variables				
Objects	1	2	3	j	n
1	X_{11}	X_{12}	X_{13}	X_{1j}	X_{1n}
2	X_{21}	X_{22}	X_{23}	X_{2j}	X_{2n}
3	X_{31}	X_{32}	X_{33}	X_{3j}	X_{3n}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
i	X_{i1}	X_{i2}	X_{i3}	X_{ij}	X_{in}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
m	X_{m1}	X_{m2}	X_{m3}	X_{mj}	X_{mn}

object i has zero amount of variable j), and ratios of scale values are meaningful. Each higher scale type subsumes the properties of those below it. For example, ratio scales possess all the properties of nominal, ordinal, and interval scales, in addition to a natural origin.

There are many descriptors by which we can characterize methods for analyzing associative data.[3] The following represent the more common bases by which the activity can be classified:

1. purpose of the study and the types of assertions desired by the researcher–what kinds of statements does he wish to make about the data or about the universe from which the data were drawn?
2. focus of research emphasis–statements regarding the objects (i.e., the whole profile or "bundle" of variables), specific variables, or both;
3. nature of his prior judgments as to how the data matrix should be partitioned in terms of the type and number of subsets of variables;
4. number of variables in each of the partitioned subsets;
5. type of association under study–linear in the parameters, transformable to linear, or "inherently" nonlinear in the parameters;
6. scales by which variables are measured–nominal, ordinal, interval, ratio, mixed.

All of these descriptors relate to certain decisions required of the researcher. Suppose he is interested in studying certain descriptive relationships among variables. If so, he must make decisions about how he wants to partition the set of columns (see Table 1.1) into subsets. Often he will call one subset "criterion" variables and the other subset "predictor" variables.[4] He must also decide, however, on the number of variables to include in each subset and on what type of functional relationship is to hold among the parameters in his statistical model.

Most decisions about associative data analysis are based on the researcher's "private model" of how the variables are related and what features are useful for study.[5] His choice of various "public models" for analysis–multiple regression, discriminant analysis, etc.–is predicated on his prior knowledge of the characteristics of the statistical universe from which the data were obtained and his knowledge of the assumption structure of each candidate technique.

1.3.1 Researcher's Objectives and Predictive Statements

We have already commented that the researcher may be interested in (a) measuring the nature and degree of association between two or more variables; (b) predicting the values of one or more criterion variables from values of one or more predictor variables; or (c)

[3] An excellent classification, based on a subset of the descriptors shown here, has been provided by M. M. Tatsuoka and D. V. Tiedeman (1963).

[4] As Horst (1961) has shown, relationships need not be restricted to two sets.

[5] To some extent this is true even of the scales along which the data are measured. The researcher may wish to "downgrade" data originally expressed on interval scales to ordered categories, if he feels that the quality of the data does not warrant the "strength" of scale in which it is originally expressed. In other cases he may "upgrade" data in order to use some statistical technique that assumes a type of measurement that is absent originally.

assessing the statistical reliability of an association between two or more variables. In a specific study all three objectives may be pursued. In using other techniques (i.e., those dealing mainly with factor and cluster analysis), the researcher may merely wish to portray association in a more parsimonious way without attempting to make specific predictions or inferential statements.

1.3.2 Focus of Research Interest

Some multivariate techniques (e.g., multiple regression) focus on association among variables; objects are treated only as replications. Other techniques (e.g., cluster analysis) focus on association among objects; information about specific variables is usually, although not necessarily, suppressed. In still other instances one may wish to examine interrelationships among variables, objects, and object–variable combinations, as well.

1.3.3 Nature of Assumed Prior Judgments or Presuppositions

In many cases the investigator is able to partition the data matrix into subsets of columns (or rows) on the basis of prior judgment. For example, suppose the first column of Table 1.1 is average weekly consumption of coffee by households, and the other columns consist of various demographic measurements of the m households. The analyst may wish to predict average weekly consumption of coffee from some linear composite of the $n - 1$ remaining variables. If so, he has used his presuppositions regarding how the dependence is to be described and, in this instance, might employ multiple regression.

In most cases the number of subsets developed from the data matrix partitioning will be two, usually labeled as criterion and predictor variable subsets. However, techniques have been designed to summarize association in cases involving more than two subsets of data.

Finally, we may have no reasonable basis for partitioning the data matrix into criterion or predictor variables. Our purpose here may be merely to group objects into "similar" subsets, based on their correspondence over the whole profile of variables. Alternatively, we may wish to portray the columns of the data matrix in terms of a smaller number of variables, such as linear combinations of the original set, that retain most of the information in the original data matrix. Cluster analysis and factor analysis, respectively, are useful techniques for these purposes.

1.3.4 Number of Variables in Partitioned Subsets

Clearly, the term "association" implies at least two characteristics–for example, a single criterion and a single predictor variable, usually referred to as bivariate data. In other cases involving two subsets of variables, we may wish to study association between a single criterion and more than one predictor. Or we may wish to study association between composites of several criterion variables and composites of several predictor variables. Finally we may want to study the relationship between several criterion variables and a single predictor variable.

Of course, we may elect not to divide the variables at all into two or more subsets, as would be the case in factor analysis. Furthermore, if we do elect to partition the matrix

and end up with two or more variables in a particular subset, what we are usually concerned with are various *linear composites* of the variables in that subset and each composite's association with other variables.

1.3.5 Type of Association

Most of the models of multivariate analysis emphasize linear relationships among the variables. The assumption of linearity, in the parameters, is not nearly so restrictive as it may seem.[6] First, various preliminary transformations (e.g., square root, logarithmic) of the data are possible in order to achieve linearity in the parameters.[7] Second, the use of "dummy" variables, coded, for example, as elementary polynomial functions of the "real" variables, or indicating category membership by patterns of zeroes and ones, will enable us to handle certain types of nonlinear relationships within the framework of a linear model. Third, a linear model is often a good approximation to a nonlinear one, at least over restricted ranges of the variables in question.

1.3.6 Types of Scales

Returning to the data matrix of Table 1.1, we now are concerned with the scales by which the characteristics are represented. Since all of the multivariate statistical techniques to be discussed in this book require no stronger form of measurement than an interval scale, we shall usually be interested in the following types: (a) nominal, (b) ordinal, and (c) interval. In terms of nominal scaling we shall find it useful to distinguish between dichotomies and (unordered) polytomies, the latter categorization involving more than two classes.

This distinction is important for three reasons. First, many of the statistical techniques for analyzing associative data are amenable to binary-coded (zero–one) variables but *not* to polytomies. Second, any polytomy can be recoded as a set of dichotomous "dummy" variables; we shall describe how this recoding is done in the next section. Third, when we discuss geometrical representations of variables and/or objects, dichotomous variables can be handled within the same general framework as interval-scaled variables.

Finally, mention should be made of cases in which the analyst must contend with *mixed* scales in the criterion subset, predictor subset, or both. Many multivariate techniques—if not modified for this type of application—lead to rather dubious results under such circumstances.

[6] By linear in the parameters is meant that the b_j's in the expression $y = b_1x_1 + b_2x_2 + \cdots + b_nx_n$ are each of the first degree. Similarly, $z = b_1x_1^2 + \cdots + b_nx_n^{n+1}$ is still linear in the parameters since each b_j continues to be of the first degree even though x_j is not.

[7] For example, the complicated expression $y = ax^be^{cx}$ (with both $a, x > 0$) can be "linearized" as $\ln y = \ln a + b \ln x + cx$ and, as shown by Hoerl (1954), is quite flexible in approximating many diverse types of curves. On the other hand, the function $y = 1/(a + b^{-cx})$ is inherently nonlinear in the parameters and cannot be "linearized" by transformation.

1.4 ORGANIZING THE TECHNIQUES

In most textbooks on multivariate analysis, three of the preceding characteristics are often used as primary bases for technique organization:

1. whether one's principal focus is on the objects or on the variables of the data matrix;
2. whether the data matrix is partitioned into criterion and predictor subsets, and the number of variables in each;
3. whether the cell values represent nominal, ordinal, or interval scale measurements.

This schema results in four major subdivisions of interest:

1. *single criterion, multiple predictor association,* including multiple regression, analysis of variance and covariance, and two-group discriminant analysis;
2. *multiple criterion, multiple predictor association,* including canonical correlation, multivariate analysis of variance and covariance, multiple discriminant analysis;
3. *analysis of variable interdependence,* including factor analysis, multidimensional scaling, and other types of dimension-reducing methods;
4. *analysis of interobject similarity,* including cluster analysis and other types of object-grouping procedures.

The first two categories involve dependence structures where the data matrix *is* partitioned into criterion and predictor subsets; in both cases interest is focused on the variables. The last two categories are concerned with interdependence—either focusing on variables or on objects. Within each of the four categories, various techniques are differentiated in terms of the type of scale assumed.

1.4.1 Scale Types

Traditionally, multivariate methods have emphasized two types of variables:

1. more or less continuous variables, that is, interval-scaled (or ratio-scaled) measurements;
2. binary-valued variables, coded zero or one.

The reader is no doubt already familiar with variables like length, weight, and height that can vary more or less continuously over some range of interest.

Natural dichotomies such as sex, male or female, or marital status, single or married, are also familiar. What is perhaps not as well known is that any (unordered) polytomy, consisting of three or more mutually exclusive and collectively exhaustive categories, can be recoded into dummy variables that are typically coded as one or zero. To illustrate, a person's occupation, classified into five categories, could be coded as:

	Dummy variable			
Category	1	2	3	4
Professional	1	0	0	0
Clerical	0	1	0	0
Skilled laborer	0	0	1	0
Unskilled laborer	0	0	0	1
Other	0	0	0	0

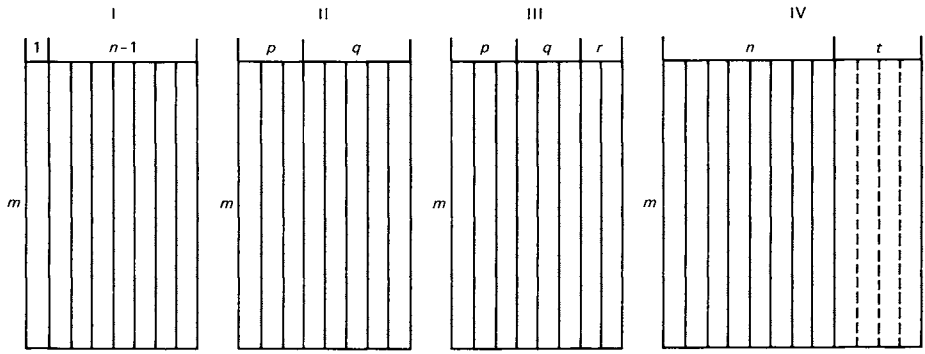

Fig. 1.1 Illustrative partitionings of data matrix.

For example, if a person falls into the professional category, he is coded 1 on dummy variable 1 and 0 on dummies 2 through 4. In general, a k-category polytomy can be represented by $k - 1$ dummy variables, with one category–such as the last category–receiving a value of zero on all $k - 1$ dummies.[8]

Multivariate techniques that are capable of dealing with some or all variables at the ordinally scaled level are of more recent vintage. With few exceptions our attention in this book will be focused on either continuous or binary-valued variables.[9]

Figure 1.1 shows some of the major ways in which the data matrix can be viewed from the standpoint of technique selection.

1.4.2 Single Criterion, Multiple Predictor Association

In Panel I of the figure we note that the first column of the matrix has been singled out as a criterion variable and the remaining $n - 1$ variables are considered as predictors. For example, the criterion variable could be average weekly consumption of beer by the ith individual. The $n - 1$ predictors could represent various demographic variables, such as the individual's age, years of education, income, and so on. This is a prototypical problem for the application of multiple regression in which one tries to predict values of the criterion variable from a linear composite of the predictors. The predictors, incidentally, can be either continuous variables or dummies, such as marital status or sex. Alternatively, the single criterion variable could represent the prior categorization of each individual as heavy beer drinker (coded one, arbitrarily) or light beer drinker (coded zero), based on some designated amount of average weekly beer consumption. If our purpose is to develop a linear composite of the predictors that enables us to classify each individual into either heavy or light beer drinker status, then we would employ two-group discriminant analysis. The critical distinction here is that the criterion variable is expressed as a single dummy variable rather than as a continuous one.

[8] Not only is greater parsimony achieved by using only $k - 1$ (rather than k) categories, but, as will be shown in later chapters, this type of coding device permits the matrix to be inverted by regular computational methods.

[9] While a classification *could* be coded as $1, 2, 3, \ldots, k$ in terms of a single variable, the resulting analysis would assume that *all classes are ordered and equally spaced*–a rather dubious assumption in most kinds of classificatory data.

Another possibility can arise in which the criterion variable continues to be average weekly beer consumption, but the predictor set consists of a classification of each individual into some occupational group, coded as a set of dummy or design variables. This represents an instance in which the technique of analysis of variance could be used. On the other hand, if the predictor set consists of a classification of occupations as well as annual income in dollars, then the latter variable could be treated as a covariate. In this case we would be interested in whether average weekly beer consumption differs across occupations once the effect of income is controlled for statistically.

1.4.3 Multiple Criterion, Multiple Predictor Association

In Panel II of Fig. 1.1 the $p = 3$ criterion variables could denote individual consumption of beer, wine, and liquor, and the remaining variables could denote demographic characteristics. If we were interested in the linear association between the two *batteries* of variables, we could employ the technique of canonical correlation.

Suppose, alternatively, that all individuals had been previously classified into one of four groups: (a) malt beverage drinker only, (b) drinker of spirits (liquor or wine) other than malt beverages, (c) drinker of both spirits and malt beverages, and (d) drinker of neither. We could then develop linear functions of the demographics that would enable us to assign each individual to one of the four groups in some "best" way (to be defined later). This is an illustration of multiple discriminant analysis; note that four mutually exclusive groups are classifiable in terms of $p = 3$ criterion dummies.

Alternatively, we could continue to let the criterion variables denote individual consumption levels of beer, wine, and liquor, but now assume that the predictors represent dummies based on an occupational classification. If so, multivariate analysis of variance is the appropriate procedure. If income is again included as a covariate, we have an instance of multivariate analysis of covariance.

Panel III of Fig. 1.1 shows a data structure involving association among three batteries of variables. Generalized canonical correlation can be employed in this type of situation. In this case we would be interested in what all three batteries exhibit in common and also in the strength of association between all distinct pairs of batteries as well.

1.4.4 Dimension-Reducing Methods

Panel IV of Fig. 1.1 shows a set of t appended columns, each of which is expressed as a linear composite of the original n variables. Suppose we want to portray the association across the m individuals in terms of fewer variables than the original n variables. If so, we might employ factor analysis, multidimensional scaling, or some other dimension-reduction method to represent the original set of n correlated variables as linear (or nonlinear) composites of a set of t ($t < n$) underlying or "latent" variables in such a way as to retain as much of the original information as possible. The composites themselves might be chosen to obey still other conditions, such as being mutually uncorrelated.

Thus, if the original n variables are various demographics characterizing a set of beer drinkers, we might be able to find a set of more basic dimensions—social class, stage in life cycle, etc.—so that linear composites of these basic dimensions account for the observable demographic variables.

1.4.5 Interobject Similarity

So far we have confined our attention to the columns of the matrices in Fig. 1.1. Suppose now that the n columns represent consumption of various kinds of alcoholic beverages—beers, ales, red wines, white wines, liquors, after-dinner cordials—over some stated time period. Each individual's consumption profile could be compared with every other individual's, and we could develop a measure of interindividual similarity with respect to patterns of alcoholic beverage drinking.

Having done so, we could then proceed to cluster individuals into similar groups on the basis of the overall similarity of their consumption profiles. Note here that information on specific variables is lost in the computation of interindividual similarity measures. Since our focus of interest is on the objects rather than on the variables, we may be willing to discard information on separate variables in order to grasp the notion of *overall* interobject similarity (and the "clusteriness" of objects) more clearly.

All of these techniques—and others as well—have been employed in the behavioral and administrative sciences. As suggested above, the tools of multivariate analysis form a unified set, based on a relatively few descriptors for distinguishing specific techniques.

1.5 ILLUSTRATIVE APPLICATIONS

Any empirically grounded discipline has need on occasion to use various types of multivariate techniques. Indeed, in some fields like psychometrics and survey research, multivariate analysis represents the methodological cornerstone.

Although multivariate analysis can be, and has been, used in the physical and life sciences, increasing applications are being made in the behavioral and administrative sciences. Even at that, the view is a broad one, as the following list of behavioral and administrative examples illustrate.

Example 1 Two economists, Quandt and Baumol (1966), were interested in predicting the demand for alternative modes of travel between various pairs of cities. They developed a characterization of each mode of travel (e.g., airplane, train, bus, private car) as a service profile varying in levels of cost, departure frequency, convenience to the traveler, speed, and so on.

A travel-demand forecasting model for each mode was then prepared which utilized a linear function of the logarithms of service profile levels. Traffic volumes, involving sixteen different city pairs, were available for each of the abovementioned modes to serve as criterion variables. The parameters of their demand forecasting model were estimated by multiple regression.

Example 2 A group of psychologists, Rorer *et al.* (1967), were concerned with the modeling of clinical judgment and, in particular, how subjects combined various information cues into an overall judgment. The subjects of their experiment were small groups of physicians, nurses, psychologists, and social workers. The judgment to be made concerned the subject's probability of granting a weekend pass to each of 128 (presumed real) patients. Each "patient" was described according to a six-component profile, involving such characteristics as (a) whether he had a drinking problem; (b) whether he

had abused privileges in the past; (c) whether his personal appearance was neat, and so on. The patient was described simply as to whether he displayed the characteristic or not.

The six characteristics used by the researchers were formulated in a 2^6 design of all possible combinations, and two replications were made up of each combination, leading to a total of 128 that were presented (in random order) to each subject. An analysis was made of each subject's response data separately, using an analysis of variance model applicable to a full factorial design. The researchers found evidence that subjects used cues interactively in arriving at an overall judgment. The relative importance of these interactions was measured as well as the separate main-effect contributions to the overall judgment regarding the subjective probability of granting a pass.

Example 3 A political scientist, R. J. Rummel (1970), was interested in a cross-national comparison of some 82 different countries, measured according to 230 characteristics. In particular, he wished to see what underlying factors or dimensions might account for various observed relationships across such characteristics as the nations' trade levels, memberships in international organizations, production of various commodities, and so on.

A variety of factor analyses and cluster analyses were performed on the data, leading to a set of underlying dimensions, identified principally as the nation's political orientation, economic development, and degree of foreign conflict.

Two mathematical psychologists, Wish and Carroll (1973), were also interested in national similarities and differences but, in this case, as subjectively perceived by U.S. and foreign students. Their methodology emphasized multidimensional scaling and, in particular, individual differences models of perception. They found that different subjects gave different importances to perceptual dimensions, depending upon the subject's attitude toward U.S. involvement in the Vietnam conflict.

Example 4 Two educational psychologists, Cooley and Lohnes (1971), were engaged in a massive sampling survey, called Project TALENT, involving measurement on a large number of personality and ability variables of a representative sample of American high school students. The purpose of the study was to examine interrelationships among these variables and various environmental variables in order to predict the students' motivations involving subsequent career and higher education activities.

A variety of multivariate techniques were employed in the analysis of these data. For example, one analysis used canonical correlation to examine the association between a set of eleven ability-type factors (e.g., verbal knowledge, mathematics, visual reasoning) and a set of eleven factors dealing with career motives (e.g., interest in science, interest in business). In this case the canonical correlation followed a preliminary factor analysis of each separate battery of variables.

Example 5 A group of survey researchers, Morgan, Sirageldin, and Baerwaldt (1965), were engaged in a large-scale survey in which the criterion variable of interest was hours spent on do-it-yourself activities by heads of families and their spouses. A sample size of 2214 households provided the data, and the predictor variables included a large number of demographic characteristics.

Not surprisingly, the authors found that marital status was the most important predictor variable. Single men and women spent relatively little time on do-it-yourself

activities. On the other hand, married couples with large families who had higher-than-average education, lived in single-family structures in rural areas, with youngest child between two and eight years, devoted a large amount of time to do-it-yourself activities. The researchers used a multivariate technique, known as Automatic Interaction Detection, to develop a sequential branching of groups. At each stage the program selects a predictor variable that accounts for the most variation in the criterion variable and splits the sample into two subgroups, according to their values on that predictor variable. The result is a sequential branching "tree" of groups that are most homogeneous with regard to the criterion variable of interest.

Example 6 A group of management scientists, Haynes, Komar, and Byrd (1973), were interested in the comparative performance of three heuristic rules that had been proposed for sequencing production jobs that incur setup changes. The objective of each of the rules was to minimize machine downtime over the whole production sequence. For example, Rule 1 involved selecting as the next job that one which has the least setup time relative to the job last completed, of all jobs yet unassigned. The researchers were interested in how the competing heuristics would perform under variations in setup time distributions and total number of jobs to be sequenced.

The researchers set up an experimental design in which application of the three rules was simulated in a computer under different sets of distribution times and numbers of jobs. The factorial design employed by the authors to test the behavior of the rules was then analyzed by analysis of variance procedures. The experiment indicated that a composite of the three heuristics might perform better than any of the three rules taken singly.

Example 7 Two marketing researchers, Perry and Hamm (1969), believed that consumers might ascribe higher importance to personal sources (in the selection of products involving high socioeconomic risk) than to impersonal sources of product information. They set up an experiment in which consumers rated a set of 25 products on degree of perceived social risk and degree of economic risk. Each respondent was also asked to rate the significance of various sources of influence (e.g., advertisements, *Consumer's Reports*, a friend's recommendations) on one's choice of brand within each product class.

The authors used canonical correlation to relate the two sets of measures. They found that the higher the risk, particularly the social risk, the greater the perceived importance of personal influence on brand choice. The authors concluded that in the advertising of high-risk products (e.g., color TV, automobiles, sports jackets), advertisers should try to reach prospective buyers through personal channels, such as opinion leaders, rather than through general media. Moreover, advertisers should emphasize the social, rather than the economic, benefits of the purchase.

As the preceding examples suggest, virtually any discipline in the behavioral and administrative sciences can find applications for multivariate tools. It is not surprising why this is so, given that multivariate techniques can be used in both controlled experiments and observational studies. And, in the latter case at least, data refuse to come in neat and tidy packages. Rather, the predictor variables are usually correlated themselves, and one needs statistical tools to assist one in finding out what is going on.

Even in controlled experiments it is usually not possible to control for *all* variables. Various experimental devices, such as blocking and covariance adjustment, are often used to increase precision as well as to reduce some of the sources of statistical bias.

Moreover, it seems to be the nature of things in both the behavioral and administrative sciences that the possible explanatory variables of some phenomenon of interest are myriad and difficult to measure (as well as interrelated). It should come as no surprise that methods are needed to reveal whatever patterns exist in the data as well as to help the analyst test hypotheses regarding his content area of interest. Thus, multivariate techniques are becoming as familiar to the marketing researcher, production engineer, and corporate finance officer as they are to the empirically oriented psychologist, sociologist, political scientist, and economist.

In addition to the diffusion among disciplines, multivariate techniques themselves are increasing in variety and sophistication. Researchers' past emphasis on multiple regression and factor analysis has given way to application of whole new classes of techniques—canonical correlation, multiple discriminant analysis, cluster analysis, and multidimensional scaling, to name a few. Methods are being extended to deal with multiway matrices and time-dependent observations. Computing routines have incorporated still-recent developments in nonlinear optimization and other forms of numerical analysis. Methods typically used for measured variables have been modified and extended to cope with data that are expressed only as ranks or in some cases only in terms of category membership.

In short, multivariate data analysis has become a vigorous field methodologically and a catholic field substantively. Indeed, it is difficult to think of any behavioral or administrative discipline in which multivariate methods have no applicability.

1.6 SOME NUMERICAL EXAMPLES

To a large extent, the study of multivariate techniques is the study of linear transformations. In some techniques the whole data matrix—or some matrix derived from it—may undergo a linear transformation. Other methods involve various transformations of submatrices obtained from partitioning the original matrix according to certain presuppositions about the substantive data of interest.

In the chapters that follow we shall be discussing those aspects of linear algebra and transformational geometry that underlie all methods of multivariate analysis. The arithmetic operations associated with vectors and matrices, determinants, eigenstructures, quadratic forms, and singular value decomposition are some of the concepts that will be presented.

As motivation for the study of these tools, let us consider three problems that could arise in an applied research area. While almost any field could supply appropriate examples, suppose we are working in the field of personnel research. In particular, imagine that we are interested in certain aspects of employee absenteeism. We shall assume that all employees are male clerks working in an insurance company.

Absenteeism records have been maintained for each employee over the past year. Personnel records also indicate how long each employee has worked for the company. In addition, each employee recently completed a clinical interview with the company psychologist and was scored by the psychologist on a 1-to-13 point rating scale, with "1"

TABLE 1.2

Personnel Data Used to Illustrate Multivariate Methods

	Number of days absent			Attitude rating			Years with company		
Employee	Y	Y_d	Y_s	X_1	X_{d1}	X_{s1}	X_2	X_{d2}	X_{s2}
a	1	−5.25	−0.97	1	−5.25	−1.39	1	−3.92	−1.31
b	0	−6.25	−1.15	2	−4.25	−1.13	1	−3.92	−1.31
c	1	−5.25	−0.97	2	−4.25	−1.13	2	−2.92	−0.98
d	4	−2.25	−0.41	3	−3.25	−0.86	2	−2.92	−0.98
e	3	−3.25	−0.60	5	−1.25	−0.33	4	−0.92	−0.31
f	2	−4.25	−0.78	5	−1.25	−0.33	6	1.08	0.36
g	5	−1.25	−0.23	6	−0.25	−0.07	5	0.08	0.03
h	6	−0.25	−0.05	7	0.75	0.20	4	−0.92	−0.31
i	9	2.75	0.51	10	3.75	0.99	8	3.08	1.03
j	13	6.75	1.24	11	4.75	1.26	7	2.08	0.70
k	15	8.75	1.61	11	4.75	1.26	9	4.08	1.37
l	16	9.75	1.80	12	5.75	1.53	10	5.08	1.71
Mean	6.25			6.25			4.92		
Standard deviation	5.43			3.77			2.98		

indicating an extremely favorable attitude and "13" indicating an extremely unfavorable attitude toward the company. (The 13 scale points were chosen arbitrarily.)

For purposes of illustration, a sample of 12 employees was selected for further study.[10] The "raw" data on each of the three variables are shown in Table 1.2. Figure 1.2 shows each pair of variables in scatter plot form.

From Fig. 1.2 we note the tendency for all three variables to be positively associated. That is, absenteeism increases with unfavorableness of attitude toward the firm and number of years with the company. Moreover, unfavorableness of attitude is positively associated with number of years of employment with the firm (although one might question the reasonableness of this assumed relationship).

Table 1.2 also shows the means and sample standard deviations of each of the three variables. By subtracting out the mean of each variable from that variable's original observation we obtain three columns of deviation (or mean-corrected) scores, denoted by Y_d, X_{d1}, and X_{d2} in Table 1.2. To illustrate:

$$\boxed{Y_{di} = Y_i - \overline{Y}}$$

where $\overline{Y}$, denoting the mean of Y, is written as

$$\boxed{\overline{Y} = \sum_{i=1}^{m} Y_i/m}$$

[10] Obviously, the small sample size of only 12 employees is for illustrative purposes only; moreover, all data are artificial.

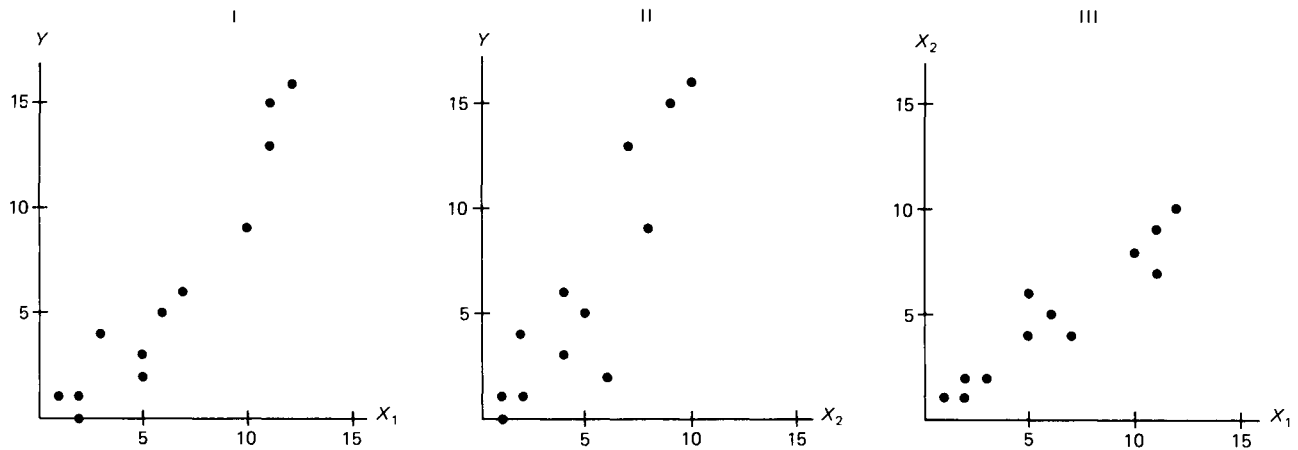

Fig. 1.2 Two-variable scatter plots of data from sample problem. Key: I, Y vs X_1; II, Y vs X_2; III, X_2 vs X_1.

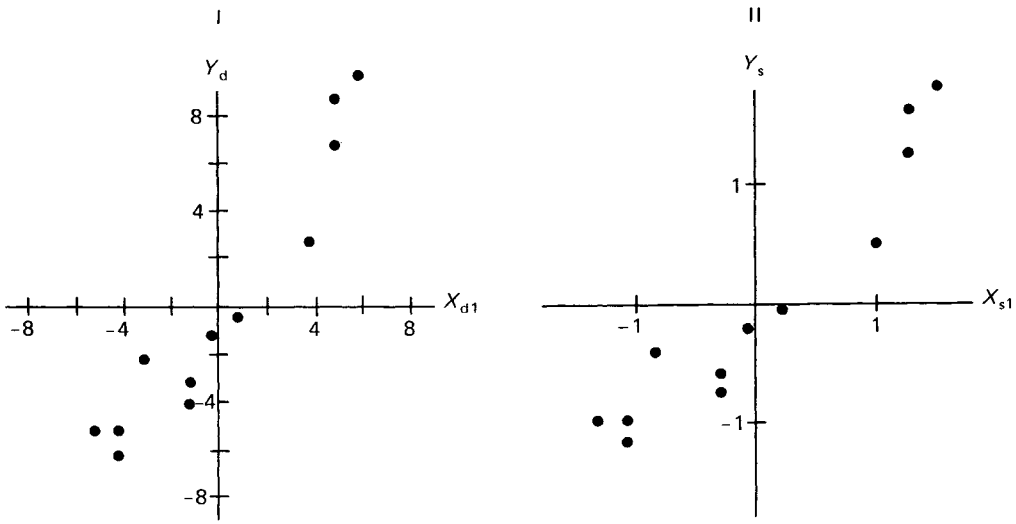

Fig. 1.3 Illustrative scatter plots of (I) mean-corrected and (II) standardized data.

Standardized scores, denoted by Y_s, X_{s1}, and X_{s2}, are obtained by dividing each mean-corrected score by the sample standard deviation of that variable.[11] To illustrate:

$$\boxed{Y_{si} = Y_{di}/s_y}$$

where s_y, in turn, is defined as

$$\boxed{s_y = \left[\sum_{i=1}^{m} (Y_i - \bar{Y})^2/m\right]^{1/2}}$$

Figure 1.3 shows, illustratively, the scatter plot of mean-corrected and standardized scores involving the criterion variable Y versus the first predictor X_1. We note that the computation of deviation scores merely changes the origin of the plot to an average of zero on each dimension. Interpoint distances do not change, and the configuration of points looks just like the configuration shown in the leftmost panel of Fig. 1.2.

Standardization of the data, however, does change the shape of the configuration, as well as shifting the origin. If the right-hand panel of Fig. 1.3 is compared to the left-hand panel, we see that the vertical axis, or ordinate, is compressed, relative to the horizontal axis, or abcissa. This is because the Y_d values are being divided by a larger constant (5.43) than the X_{d1} values, the latter being divided by 3.77, the sample standard deviation of X_1.

[11] The reader will note that s_y denotes the *sample* standard deviation rather than an estimate of the universe standard deviation. In this latter case the divisor would be $m - 1$, rather than m, as used here.

1.6.1 Research Questions

After this preliminary examination of the data, suppose the researcher raises the following questions:

1. How does Y relate to changes in X_1 and X_2?
   a. Can an equation be developed that will enable us to predict values of Y as a linear function of X_1 and X_2?
   b. How strong is the overall relationship of Y with X_1 and X_2?
   c. Is the overall relationship statistically significant?
   d. What is the relative influence of X_1 and X_2 on variation in Y and are these separate influences statistically significant?
2. Next, considering the relationship between the predictor variables X_1 and X_2, some further questions can be asked:
   a. Can the 12 scores on X_1 and X_2 be replaced by scores on a single variable that represents a linear composite of the two separate scores? That is, do X_1 and X_2 really reflect just a single underlying factor, or are there two separate factors operating?
   b. What is the association of X_1 and X_2, respectively, with this linear composite?
   c. How much of the total variation in X_1 and X_2 is accounted for by the single linear composite?
3. One additional thing that we might do is to split the sample of employees into three groups: Group 1—employees a, b, c, d; Group 2—employees e, f, g, h; Group 3—employees i, j, k, l. We could call these three groups low-, intermediate-, and high-absenteeism groups, respectively.[12]

Having classified the 12 respondents in this way, we would then raise the questions:

   a. How do we go about defining a linear composite of X_1 and X_2 that maximally separates the three groups?
   b. Is this linear composite statistically significant?
   c. How well does the linear composite assign individuals to their correct groups?
   d. What is the relative influence of X_1 and X_2 on group assignment and are their separate contributions statistically significant?
   e. How could we find a second linear composite, uncorrelated with the first, that does the next best job of separating the groups (and so on)?

Each set of questions describes a particular multivariate technique which we now consider.

1.6.2 Multiple Regression

The first set of questions pertain to a problem in multiple regression. This, in turn, involves the subproblems of developing an estimating equation, computing its strength of

[12] This assumes, of course, that specific numerical data on number of days absent are being supplanted by interest only in the three groups of the classification: low, intermediate, and high absenteeism. This disregard for numerical information on number of days absent is strictly for motivating the discussion of multiple discriminant analysis, although it could be rationalized on other grounds, such as possible nonlinear association.

relationship and statistical significance, and examining the contribution of each predictor to changes in the criterion variable.

Insofar as the first problem is concerned, we shall want to find parameter values for the linear equation:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

where $\hat{Y}$ denotes predicted values of Y; b_0 denotes the intercept term when X_1 and X_2 are both 0; and b_1 and b_2 denote the partial regression coefficients of X_1 and X_2, respectively. The partial regression coefficient measures the change in $\hat{Y}$ per unit change in some specific predictor, with other predictors held constant.

In terms of the data of Table 1.2 we shall want to find the parameter values b_0, b_1, b_2 and the 12 predicted values:

$$\begin{aligned}
\hat{Y}_1 &= b_0 + b_1(1) + b_2(1) \\
\hat{Y}_2 &= b_0 + b_1(2) + b_2(1) \\
\hat{Y}_3 &= b_0 + b_1(2) + b_2(2) \\
&\vdots \\
\hat{Y}_{12} &= b_0 + b_1(12) + b_2(10)
\end{aligned}$$

(where the numbers in parentheses are actual numerical values of X_1 and X_2, from Table 1.2). As will be shown in subsequent chapters, we shall find the specific values of b_0, b_1, and b_2, according to the *least-squares* principle. This entails minimizing the sum of the squared errors e_i :

$$\sum_{i=1}^{12} e_i^2 = \sum_{i=1}^{12} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{12} (Y_i - b_0 - b_1 X_{i1} - b_2 X_{i2})^2$$

The least-squares principle leads to a set of linear equations which, when solved, provide the desired parameter values.

The second question, concerning strength of the (linear) relationship, is answered by computing R^2, the squared multiple correlation. R^2 measures how much of the variation in Y, as measured about its mean $\bar{Y}$, is accounted for by variation in X_1 and X_2. R^2 can be expressed quite simply as

$$R^2 = 1 - \frac{\sum_{i=1}^{12} e_i^2}{\sum_{i=1}^{12} (Y_i - \bar{Y})^2}$$

where the denominator represents the sum of squares in Y as measured about its own mean. As can be observed, if the sum of squared errors is zero, then the $\hat{Y}_i$'s predict their respective Y_i's perfectly and $R^2 = 1$. However, if the inclusion of X_1 and X_2 in the estimating equation does no better than use of the $\bar{Y}$ alone, then the numerator of the fraction equals the denominator and $R^2 = 0$, denoting no variance accounted for, beyond using the criterion variable's mean.

The third question entails a test of the null hypothesis of no linear association between Y and X_1 and X_2, as considered together. This can be expressed either as

$$\boxed{R_p = 0}$$

where R_p denotes the population multiple correlation, or as

$$\boxed{\beta_1 = \beta_2 = 0}$$

where β_1 and β_2 denote population partial regression coefficients. In Chapter 6 these tests will actually be carried out in terms of the sample problem of Table 1.2.

The fourth question concerns what are called partial correlation coefficients. One interpretation of a partial correlation coefficient considers it as a measure of the linear association between the criterion variable and some predictor when both have been adjusted for their linear association with the remaining predictors. Although the question of determining the relative influence of predictors is an ambiguous one, we shall comment on partial correlations, and their associated tests of significance, in Chapter 6.

Multiple regression, aside from being the most popular multivariate technique in applied research, provides a vehicle for subsequent discussion of all basic matrix operations and, in particular, the topics of determinants, matrix inversion, and matrix rank. These aspects of matrix algebra are essential in understanding the procedures for solving simultaneous equations, as appearing in multiple regression and other multivariate procedures.

1.6.3 Factor Analysis

The second set of questions refers to a topic in multivariate analysis that is generically called factor analysis. In factor analysis we are interested in the interdependence among sets of variables and the possibility of representing the objects of the investigation in terms of fewer dimensions than originally expressed.

To illustrate, let us plot the mean-corrected scores of the predictors, X_2 versus X_1, as shown in Fig. 1.4. Also shown in the same figure is a new axis, labeled z_1, which makes an angle of 38° with the horizontal axis. Now suppose we drop perpendiculars, represented by dotted lines, from each point to the axis z_1. Assuming the same scale for z_1 as used for X_{d1} and X_{d2}, we could compute the variance of the 12 projected scores on z_1 as

$$\boxed{\text{Var}[z_{i(1)}] = \sum_{i=1}^{12} (z_{i1} - \bar{z}_1)^2/12}$$

where $\bar{z}$ is the mean (which, because the X's are in deviation form, is zero) of the 12 scores on z_1.

The idea behind this procedure—called principal components and representing one type of factor analysis—is to find the axis z_1 so that the variance of the 12 projections

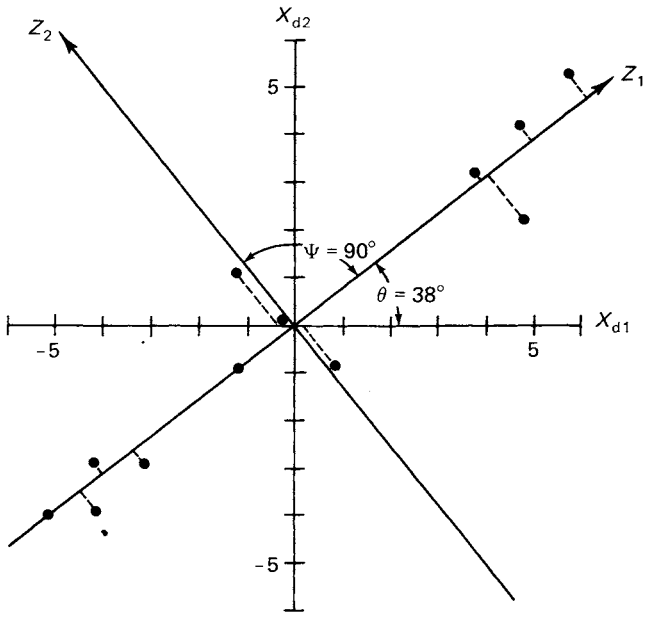

Fig. 1.4 Scatter plot of mean-corrected predictor variables.

onto it is maximal. As such the 12 mean-corrected scores, X_{d1} and X_{d2}, can be represented by a linear composite:

$$z_{i(1)} = t_1 X_{di1} + t_2 X_{di2}$$

and, hence, each pair of scores, X_{di1} and X_{di2} for each observation i, is replaced by a single score $z_{i(1)}$

In the present example, not much parsimony would be gained by merely replacing two scores with one score. In larger-scale problems, consisting of a large number of variables, considerable data reduction might be obtained. Moreover, principal components analysis allows the researcher to find additional axes, each at right angles to previously found axes and all with the property of maximum variance (subject to being at right angles to previously found axes).

This idea is also illustrated in Fig. 1.4 via the second axis z_2. Note that this axis has been drawn, as it should be, at right angles to z_1. One could, of course, project the 12 points onto this second axis and obtain a second set of scores. In this case, however, no parsimony would be gained, although the two axes z_1 and z_2 would be at right angles to each other and z_1 would contribute, by far, the greater variation in the pair of derived composites.

The second question, concerning the association of X_{d1} and X_{d2} with z_1, can be answered by computing product-moment correlations, X_{d1} with z_1 and X_{d2} with z_1. These are called component loadings and are measures of the association between each contributing (original) variable and the linear composite variable z_1 that is derived from them. Component loadings could also be computed for z_2.

The third question regarding how much of the total variation in X_{d1} and X_{d2} is accounted for by z_1 is also found from the principal components technique. In this example it happens to be 98 percent (leaving only 2 percent for z_2). That is, almost all of the original variation in X_{d1} and X_{d2} is retained in terms of the single composite variable z_1. This is evident by noting from Fig. 1.4 that the original 12 points lie close to the new axis z_1, and little information would be lost if their projections onto z_1 were substituted for their original values on X_{d1} and X_{d2}.

Subsequent chapters will discuss a number of important concepts from transformational geometry and matrix algebra–rotations, quadratic forms, eigenstructures of symmetric matrices–that pertain to solution procedures for principal components. Finally, in Chapter 6 the solution for the present problem will be described in detail.

1.6.4 Multiple Discriminant Analysis

Multiple discriminant analysis also entails a maximization objective. To illustrate, Fig. 1.5 shows a plot of X_{d2} versus X_{d1}. This time, however, each of the three groups–low, intermediate, and high absenteeism–is represented by different symbols. The first axis w_1 is the one, in this case, that maximizes among-group variation relative to average within-group variation.

That is, we wish to find a linear composite

$$w_{i(1)} = v_1 X_{di1} + v_2 X_{di2}$$

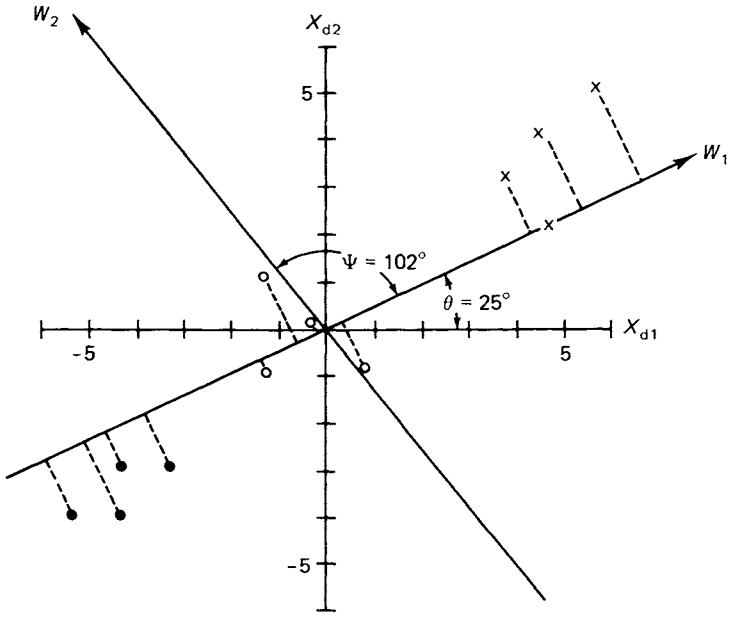

Fig. 1.5 A discriminant transformation of the mean-corrected predictor variables. Key: • Group 1; o Group 2; x Group 3.

with the property of maximizing the variation across group means relative to their pooled within-group variation.

Note that the first of these axes, w_1 in Fig. 1.5, makes an angle of 25° with the horizontal and, hence, is a different axis than the first principal component of Fig. 1.4. Note, also, that the second axis, w_2 in Fig. 1.5, is *not* at right angles to the first. In multiple discriminant analysis we can often find additional axes by which the groups can be discriminated, but these axes need not make right angles with each other.

In the case of multiple discriminant analysis, the points are projected onto w_1, and this axis exhibits the property of maximally separating the three group means, relative to their pooled within-group variation. Each point would be assigned to the closest group mean on the w_1 discriminant axis (i.e., the average of the four w_{i1}'s making up that group). As it turns out in this particular illustration, if this rule were followed, no point would be misclassified.[13]

The remaining questions dealing with statistical significance of the discriminant function(s), classification accuracy, and the relative importance of the predictor variables (X_{d1} and X_{d2}) are answered in ways analogous to the multiple regression case. We consider these, and related, questions in Chapter 6.

However, from the standpoint of transformational geometry and matrix algebra, multiple discriminant analysis involves such concepts as general linear transformations, simultaneous diagonalization of two different quadratic forms, and the eigenstructure of nonsymmetric matrices. Moreover, multiple discriminant analysis can also be related to principal component analysis in a space in which the points have previously been transformed (i.e., "spherized") to a pooled within-groups variance of unity.

In short, the three preceding problems provide sufficient motivational material for all of the remaining chapters, including the appendixes. Moreover, a full understanding of how these three problems can be solved will serve the applied researcher well insofar as understanding almost any other multivariate technique he may encounter.

1.7 FORMAT OF SUCCEEDING CHAPTERS

Succeeding chapters of the book are designed to develop the necessary concepts from transformational geometry and matrix algebra to deal with most problems in multivariate analysis, including the sample problems outlined in the preceding section. Obviously, no definitive treatment of the subject has been attempted. What we have tried to do is to select those aspects of matrix algebra that are most relevant for subsequent discussion of multivariate procedures.

Chapter 2 discusses definitions and operations on vectors and matrices. Here our emphasis is on the *mechanics* of working with vectors and matrices, rather than their geometric conceptualization. Elementary material on determinants is also presented as well as a demonstration of how various computations in multivariate analysis—for example, sums of squares and cross products—can be compactly expressed in matrix notation.

[13] Such would not be the case for the second discriminant w_2. As will be shown in Chapter 6, this second function is not statistically significant and would not be used for classification purposes anyway.

Chapter 3 is concerned with the conceptual aspects of vectors and matrices. Geometric representations are employed extensively as we discuss various operations on vectors and matrices, including scalar products and related topics in Euclidean distance geometry. Common statistical measures–standard deviation, covariance and correlation–are also portrayed from a geometric viewpoint.

Much of multivariate analysis is concerned with linear transformations, and this is the focus of Chapter 4. Each type of matrix transformation is described geometrically as we discuss such topics as rotations, stretches, and other transformations that have simple, geometric representations. Matrix inverses and the notion of matrix rank are also introduced here. We conclude the chapter with a description of the geometric effect of composite transformations that represent the matrix product of simpler ones.

Chapter 5 takes the opposite (and complementary) point of view. Here we are concerned with decomposing general matrix transformations into the product of simpler ones. The topic of matrix eigenstructure is introduced and related to the idea of changing basis vectors in order to bring out simpler geometric interpretations of the space. Matrix rank–first introduced in Chapter 4–is discussed more thoroughly in the context of the singular value decomposition of a matrix. Quadratic forms are also introduced and related to matrix eigenstructures. Again, geometric analogy is used wherever it can help illuminate the algebraic operations.

Chapter 6 completes the cycle by taking the reader back to multivariate methods per se and, in particular, to the three sample problems introduced in the present chapter. Each of these problems–centering around multiple regression, principal components analysis, and multiple discriminant analysis–is described from the standpoint of concepts developed in Chapters 2–5. Numerical solutions are obtained for each case, and several geometric aspects of the methods are illustrated. We conclude the chapter by presenting a complementary framework for multivariate technique classification in terms of the nature of the transformations characterizing the matching of one set of numbers with some other set or sets. As such, this classification descriptor serves as both a way to unify earlier material and as a prologue for various textbooks on the general topic of multivariate analysis.

The appendices cover more advanced material relevant to the general topic of multivariate analysis. Included here are such concepts as symbolic differentation, constrained optimization, generalized inverses, and other special topics of relevance to multivariate analysis.

1.8 SUMMARY

The purpose of this chapter has been to set the stage for later material dealing with aspects of transformational geometry and matrix algebra of interest to multivariate analysis. The topic of multivariate analysis was introduced as a set of procedures for dealing with association among multiple variables. A classification system based on aspects of the data matrix and the researcher's objectives and presuppositions was described as a way of matching problem with technique.

We then described briefly a number of substantive applications of multivariate methods so as to give the reader some flavor of their breadth and diversity of use.

Following this, three prototypical problems, calling for various types of multivariate analysis, were described in terms of a miniature and common data bank. These problems will serve to motivate subsequent discussion of algebraic and geometric tools, leading to the solution of the sample problems in the concluding chapter of the book.

REVIEW QUESTIONS

1. What other systems for classifying multivariate techniques can you find in the literature of your field?
   a. How would you compare these classifications with the one presented here?
   b. Criticize the present classification and indicate how you would modify it for purposes of research in your own discipline.
2. Examine the literature of your field and select a number of examples using multivariate analysis.
   a. In each example, what was the substantive problem of interest?
   b. How did the author's use of the technique(s) relate to the content side of the problem?
   c. Do other techniques suggest themselves for the specific problem examined by the author?
3. In terms of your own research try to formulate a problem that appears suitable for multivariate analysis.
   a. How would you classify the problem in terms of the system described in this chapter?
   b. What multivariate procedures are suggested by your classification of the problem?