17

Multivariate Analysis

LEARNING OBJECTIVES

After studying this chapter, you should be able to:

- Describe a multivariate normal distribution.
- Explain when a discriminant analysis could be conducted.
- Interpret the results of a discriminant analysis.
- Explain when a factor analysis could be conducted.
- Differentiate between principal components and factors.
- Interpret factor analysis results.

17–1 Using Statistics

A couple living in a suburb of Chicago, earning a modest living on salaries and claiming only small and reasonable deductions on their taxes, nonetheless gets audited by the IRS every year. The reason: The couple has 21 children. A formula residing deep inside the big IRS computer in West Virginia plucks taxpayers for audit based on the information on their tax returns. The formula is a statistical one and constitutes one of the advanced methods described in this chapter. The technique is called *discriminant analysis*. This example shows how not to use statistics, since the IRS has never been able to collect additional tax from this couple. And the IRS's discriminant analysis makes a variety of other errors with thousands of taxpayers.[1] Used correctly, however, discriminant analysis can lead to a reasonable breakdown of a population into two categories (in this example: the category of people who owe more tax and the category of people who do not owe more tax). This multivariate technique will be introduced in this chapter, along with a few others.

Multivariate statistical methods, or simply **multivariate methods,** are statistical methods for the simultaneous analysis of data on several variables. Suppose that a company markets two related products, say, toothbrushes and toothpaste. The company's marketing director may be interested in analyzing consumers' preferences for the two products. The exact type of analysis may vary depending on what the company needs to know. What distinguishes the analysis–whatever form it may take–is that it should consider people's perceptions of both products *jointly*. Why? If the two products are related, it is likely that consumers' perceptions of the two products will be correlated. Incorporating knowledge of such correlations in our analysis makes the analysis more accurate and more meaningful.

Recall that regression analysis and correlation analysis are methods involving several variables. In a sense, they are multivariate methods even though, strictly speaking, in regression analysis we make the assumption that the independent variable or variables are not random but are fixed quantities. In this chapter, we discuss statistical methods that are usually referred to as *multivariate*. These are more advanced than regression analysis or simple correlational analysis. In a multivariate analysis, we usually consider data on several variables as a single element–for example, an ordered set of values such as (x_1, x_2, x_3, x_4) is considered a single element in an analysis that concerns four variables. In the case of the analysis of consumers' preference scores for two products, we will have a consumer's response as the *pair* of scores (x_1, x_2), where x_1 is the consumer's preference for the toothbrush, measured on some scale, and x_2 is his or her preference for the toothpaste, measured on some scale. In the analysis, we consider the pair of scores (x_1, x_2) as one sample point. When k variables are involved in the analysis, we will consider the *k-tuple* of numbers $(x_1, x_2, \ldots, x_k)$ as one element–one data point. Such an ordered set of numbers is called a **vector.** Vectors form the basic elements of our analysis in this chapter.

As you recall, the normal distribution plays a crucial role in most areas of statistical analysis. You should therefore not be surprised that the normal distribution plays an equally important role in multivariate analysis. Interestingly, the normal distribution is easily extendable to several variables. As such, it is the distribution of *vector random variables* of the form $\mathbf{X} = (X_1, X_2, X_3, \ldots, X_k)$. The distribution is called the **multivariate normal distribution.** When $k = 2$, the *bivariate* case, we have a two-dimensional normal distribution. Instead of a bell-shaped curve, we have a (three-dimensional) bell-shaped *mound* as our density function. When k is greater than 2, the probability

[1]More on the IRS's use of statistics can be found in A. Aczel, *How to Beat the IRS at Its Own Game,* 2d ed. (New York: Four Walls, 1995).

function is a surface of higher dimensionality than 3, and we cannot graph it. The multivariate normal distribution will be discussed in the next section. It forms the basis for multivariate methods.

17–2 The Multivariate Normal Distribution

In the introduction, we mentioned that in multivariate analysis our elements are vectors rather than single observations. We did not define a vector, counting on the intuitive interpretation that a vector is an ordered set of numbers. For our purposes, a vector is just that: an ordered set of numbers, with each number representing a value of one of the k variables in our analysis.

> A k-dimensional random variable $\mathbf{X}$ is
>
> $$\mathbf{X} = (X_1, X_2, \ldots, X_k) \tag{17–1}$$
>
> where k is some integer.

A realization of the random variable $\mathbf{X}$ is a drawing from the populations of values of the k variables and will be denoted, as usual, by lowercase letters.

> A realization of a k-dimensional random variable $\mathbf{X}$ is
>
> $$\mathbf{x} = (x_1, x_2, \ldots, x_k) \tag{17–2}$$

Thus, in our simple example of consumer preferences for two products, we will be interested in the bivariate (two-component) random variable $\mathbf{X} = (X_1, X_2)$, where X_1 denotes a consumer's preference for the toothbrush and X_2 is the same consumer's preference for the toothpaste. A particular realization of the bivariate random variable may be (89, 78). If this is a result of random sampling from a population, it means that the particular sampled individual rates the toothbrush an 89 (on a scale of 0 to 100) and the toothpaste a 78.

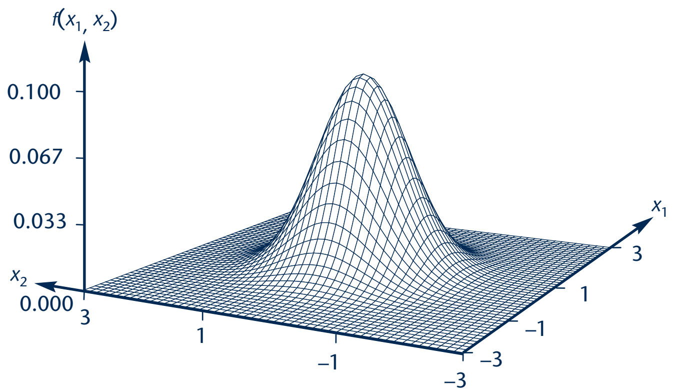
For the k-dimensional random variable $\mathbf{X} = (X_1, X_2, X_3, \ldots, X_k)$, we may define a cumulative probability distribution function $F(x_1, x_2, x_3, \ldots, x_k)$. This is a joint probability function for all k random variables X_i, where $i = 1, 2, 3, \ldots, k$.

> A joint cumulative probability distribution function of a k-dimensional random variable $\mathbf{X}$ is
>
> $$F(x_1, x_2, \ldots, x_k) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k) \tag{17–3}$$

Equation 17–3 is a statement of the probability that X_1 is less than or equal to some value x_1, *and* x_2 is less than or equal to some value x_2, *and* . . . *and* X_k is less than or equal to some value x_k. In our simple example, $F(55, 60)$ is the *joint* probability that a consumer's preference score for the toothbrush is less than or equal to 55 and that his or her preference score for the toothpaste is less than or equal to 60.

The multivariate normal distribution is an extension of the normal curve to several variables—it is the distribution of a k-dimensional vector random variable.

FIGURE 17–1 **The Bivariate Normal Probability Density Function**



A multivariate normal random variable has the probability density function

$$f(x_1, x_2, \ldots, x_k) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-(1/2)(\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{X}-\mu)} \tag{17–4}$$

where **X** is the vector random variable defined in equation 17–1; the term $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$ is the *vector of means* of the component variables X_j; and Σ is the *variance-covariance matrix*. The operations $'$ and $^{-1}$ are transposition and inversion of matrices, respectively, and $|\ |$ denotes the determinant of a matrix.

The multivariate normal distribution is an essential element in multivariate statistical techniques. Most such techniques assume that the data (on several variables), and the underlying multivariate random variable, are distributed according to a multivariate normal distribution. Figure 17–1 shows a bivariate (two-dimensional) normal probability density function.

17–3 Discriminant Analysis

A bank is faced with the following problem: Due to economic conditions in the area the bank serves, a large percentage of the bank's mortgage holders are defaulting on their loans. It therefore is very important for the bank to develop some criteria for making a statistical determination about whether any particular loan applicant is likely to be able to repay the loan. Is such a determination possible?

There is a very useful multivariate technique aimed at answering such a question. The idea is very similar to multiple regression analysis. In multiple regression, we try to predict values of a continuous-scale variable—the dependent variable—based on the values of a set of independent variables. The independent variables may be continuous, or they may be qualitative (in which case we use dummy variables to model them, as you recall from Chapter 11). In **discriminant analysis,** the situation is similar. We try to develop an equation that will help us predict the value of a dependent variable based on values of a set of independent variables. The difference is that the dependent variable is *qualitative.* In the bank loan example, the qualitative dependent variable is a classification: repay or default. The independent variables that help us make a classification of the loan outcome category may be family income, family assets, job stability (number of years with present employer), and any other variables we think may have an effect on whether the loan will be repaid. There is also an option where the algorithm itself chooses which variables should be included in the prediction equation. This is similar to a stepwise regression procedure.

If we let our dependent, qualitative variable be D and we consider k independent variables, then our prediction equation has the following form.

The form of an estimated prediction equation is

$$D = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k \tag{17–5}$$

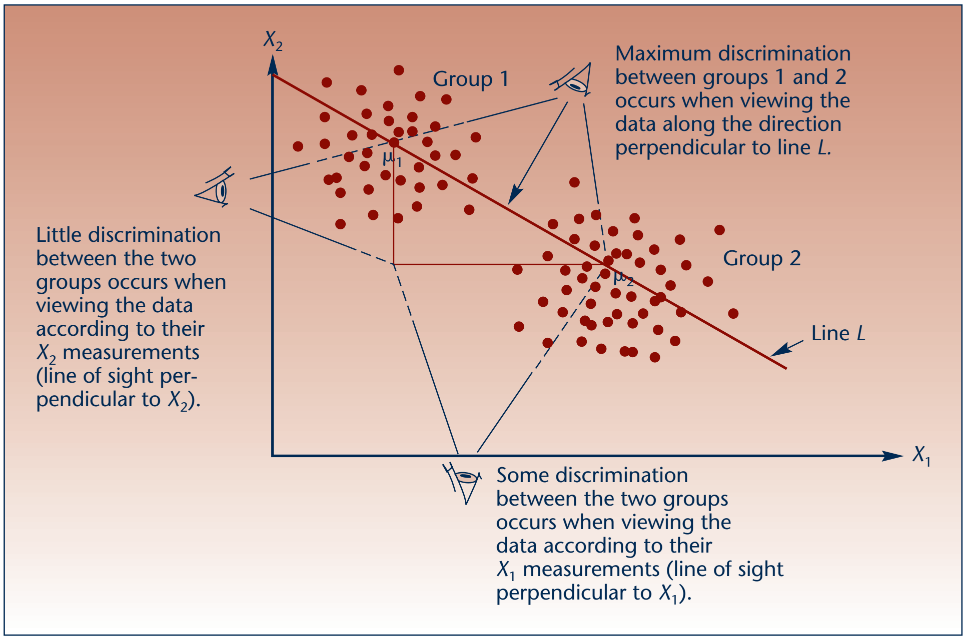
where the b_i, $i = 1, \ldots, k$, are the *discriminant weights*—they are like the estimated regression coefficients in multiple regression; b_0 is a constant.

Developing a Discriminant Function

In discriminant analysis, we aim at deriving the linear combination of the independent variables that *discriminates best* between the two or more a priori defined groups (the repay group versus the default group in the bank loan example). This is done by finding coefficient estimates b_i in equation 17–5 that maximize the among-groups variation relative to the within-groups variation.

Figure 17–2 shows how we develop a **discriminant function.** We look for a *direction in space,* a combination of variables (here, two variables, X_1 and X_2) that *maximizes the separation* between the two groups. As seen in the figure, if we consider only the X_2 component of every point in the two groups, we do not have much separation between the two groups. Look at the data in Figure 17–2 from the direction specified by having the eye located by the X_2 axis. As you see from your vantage point, the two groups overlap, and some of the upper points in group 2 look as if they belong in group 1. Now look at the data with the eye located below the X_1 axis. Here you have better separation between the two groups. From this vantage point, however,

FIGURE 17–2 Maximizing the Separation between Two Groups



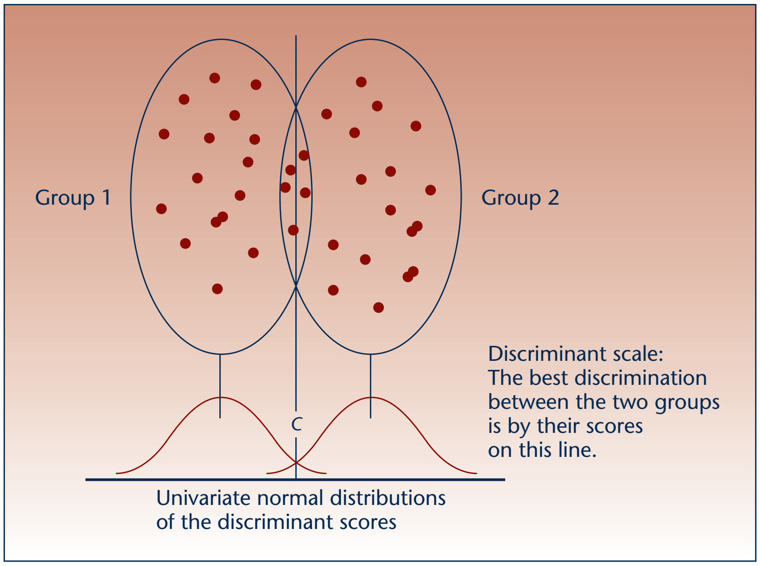
the points blend together into one big group, and you will still not be able to easily classify a point as belonging to a single group based solely on its location. Now look at the data with the eye above and perpendicular to line L. Here you have perfect separation of the two groups, and if you were given the coordinate along line L of a new point, you would probably be able to logically classify that point as belonging to one group or the other. (Such classification will never be perfect with real data because there will always be the chance that a point belonging to population 1 will somehow happen to have a low X_2 component and/or a large X_1 component that would throw it into the region we classify as belonging to population 2.) In discriminant analysis, we find the combination of variables (i.e., the direction in space) that maximizes the discrimination between groups. Then we classify new observations as belonging to one group or the other based on their score on the weighted combination of variables chosen as the discriminant function.

Since in multivariate analysis we assume that the points in each group have a multivariate normal distribution (with possibly different means), the marginal distribution of each of the two populations, when viewed along the direction that maximizes the differentiation between groups, is univariate *normal*. This is shown in Figure 17–3.

The point C on the discriminant scale is the *cutting score*. When a data point gets a score smaller than C, we classify that point as belonging to population 1; and when a data point receives a score greater than C, we classify that point as belonging to population 2. This assumes, of course, that we do not know which population the point really belongs to and we use the discriminant function to classify the point based on the values the point has with respect to the independent variables. In our bank loan example, we use the variables family income, assets, job stability, and other variables to estimate a discriminant function that will maximize the differences (i.e., the multivariate distance) between the two groups: the repay group and the default group. Then, when new applicants arrive, we find their score on our discriminant scale and classify the applicants as to whether we believe they are going to repay or default. Errors will, of course, occur. Someone we classify as a defaulter may (if given the loan) actually repay it, and someone we classify in the repay group may not.

Look at Figure 17–3. There is an area under the univariate normal projection of group 1 to the right of C. This is the probability of erroneously classifying an

FIGURE 17–3 The Discriminant Function



observation in population 1 as belonging to population 2. Similarly, the area under the right-hand normal curve to the left of the cutting score C is the probability of misclassifying a point that belongs to population 2 as being from population 1.

When the population means of the two groups are equal, there is *no discrimination* between the groups based on the values of the independent variables considered in the analysis. In such a case, the univariate normal distributions of the discriminant scores will be identical (the two curves will overlap). The reason the curves overlap is due to the model assumptions. In discriminant analysis, we assume that the populations under study have multivariate normal distributions with equal variance-covariance matrices and possibly different means.

Evaluating the Performance of the Model

We test the accuracy of our discriminant function by evaluating its success rate when the function is applied to cases with known group memberships. It is best to withhold some of our data when we carry out the estimation phase of the analysis, and then use the withheld observations in testing the accuracy of the predictions based on our estimated discriminant function. If we try to evaluate the success rate of our discriminant function based only on observations used in the estimation phase, then we run the risk of overestimating the success rate. Still, we will use our estimation data in estimating the success rate because withholding many observations for use solely in assessing our classification success rate is seldom efficient. A *classification summary table* will be produced by the computer. This table will show us how many cases were correctly classified and will also report the percentage of correctly classified cases in each group. This will give us the *hit rate* or *hitting probabilities*.

We assume that the cost of making one kind of error (classifying an element as belonging to population 1 when the element actually belongs to population 2) is equal to the cost of making the other kind of error (classifying an element as belonging to population 2 when the element actually belongs to population 1). When the costs are unequal, an adjustment to the procedure may be made.

The procedure may also be adjusted for prior probabilities of group membership. That is, when assigning an element to one of the two groups, we may account not only for its discriminant score, but also for its prior probability of belonging to the particular population, based on the relative size of the population compared with the other populations under study. In the bank loan example, suppose that defaulting on the loan is a very rare event, with a priori probability 0.001. We may wish to adjust our discriminant criterion to account for this fact, appropriately reducing our rate of classifying people as belonging to the default category. Such adjustments are based on the use of Bayes' theorem. We demonstrate discriminant analysis with the example we used at the beginning of this section, the bank loan example, which we will call Example 17–1.

EXAMPLE 17–1 The bank we have been discussing has data on 32 loan applicants. Data are available on each applicant's total family assets, total family income, total debt outstanding, family size, number of years with present employer for household head, and a qualitative variable that equals 1 if the applicant has repaid the loan and 0 if he or she has not repaid the loan. Data are presented in Table 17–1. The bank will use the data to estimate a discriminant function. The bank intends to use this function in classifying future loan applicants.

Solution The data, a random sample of 32 cases, are analyzed using the SPSS program DISCRIMINANT. The output of the analysis is given in the following figures. We use a stepwise procedure similar to stepwise multiple regression. At each stage, the computer chooses a variable to enter the discriminant function. The criterion for entering the equation may be specified by the user. Here we choose the Wilks lambda

TABLE 17–1 Data of Example 17–1 (assets, income, and debt, in thousands of dollars)

Assets	Income	Debt	Family Size	Number of Years with Present Employer	Repay/Default
98	35	12	4	4	1
65	44	5	3	1	1
22	50	0	2	7	1
78	60	34	5	5	1
50	31	4	2	2	1
21	30	5	3	7	1
42	32	21	4	11	1
20	41	10	2	3	1
33	25	0	3	6	1
57	32	8	2	5	1
8	23	12	2	1	0
0	15	10	4	2	0
12	18	7	3	4	0
7	21	19	4	2	0
15	14	28	2	1	0
30	27	50	4	4	0
29	18	30	3	6	0
9	22	10	4	5	0
12	25	39	5	3	0
23	30	65	3	1	0
34	45	21	2	5	0
21	12	28	3	2	1
10	17	0	2	3	1
57	39	13	5	8	0
60	40	10	3	2	1
78	60	8	3	5	1
45	33	9	4	7	0
9	18	9	3	5	1
12	23	10	4	4	1
55	36	12	2	5	1
67	33	35	2	4	1
42	45	12	3	8	0

criterion. The variable to enter is the variable that best fits the entry requirements in terms of the associated Wilks lambda value. Variables may enter and leave the equation at each step in the same way that they are processed in stepwise regression. The reason for this is that multicollinearity may exist. Therefore, we need to allow variables to leave the equation once other variables are in the equation. Figure 17–4 shows the variables that enter and leave the equation at each stage of the discriminant analysis.

We see that the procedure chose total family assets, total debt, and family size as the three most discriminating variables between the repay and the default groups. The summary table in Figure 17–4 shows that all three variables are significant, the largest p value being 0.0153. The three variables have some discriminating power. Figure 17–5 shows the estimated discriminant function coefficients. The results in the figure give us the following estimated discriminant function:

$$D = -0.995 - 0.0352\text{ ASSETS} + 0.0429\text{ DEBT} + 0.483\text{ FAMILY SIZE} \qquad (17\text{–}6)$$

FIGURE 17–4 SPSS-Produced Stepwise Discriminant Analysis for Example 17–1

```
------------------------- D I S C R I M I N A N T   A N A L Y S I S

ON GROUPS DEFINED BY REPAY

ANALYSIS NUMBER       1

STEPWISE VARIABLE SELECTION

SELECTION RULE:  MINIMIZE WILKS' LAMBDA

    MAXIMUM NUMBER OF STEPS. . . . . . . . . .      10
    MINIMUM TOLERANCE LEVEL. . . . . . . . . . 0.00100
    MINIMUM F TO ENTER . . . . . . . . . . . .  1.0000
    MAXIMUM F TO REMOVE. . . . . . . . . . . .  1.0000

CANONICAL DISCRIMINANT FUNCTIONS

    MAXIMUM NUMBER OF FUNCTIONS. . . . . . . . .  1
    MINIMUM CUMULATIVE PERCENT OF VARIANCE . . 100.00
    MAXIMUM SIGNIFICANCE OF WILKS' LAMBDA. . . 1.0000

PRIOR PROBABILITY FOR EACH GROUP IS 0.50000

------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP   0 -------------

                      MINIMUM
VARIABLE  TOLERANCE  TOLERANCE  F TO ENTER    WILKS' LAMBDA

ASSETS    1.0000000  1.0000000    6.6152          0.81933
INCOME    1.0000000  1.0000000    3.0672          0.90724
DEBT      1.0000000  1.0000000    5.2263          0.85164
FAMSIZE   1.0000000  1.0000000    2.5292          0.92225
JOB       1.0000000  1.0000000    0.24457         0.99191

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

AT STEP   1, ASSETS   WAS INCLUDED IN THE ANALYSIS.

                                DEGREES OF FREEDOM  SIGNIF.  BETWEEN GROUPS

WILKS' LAMBDA         0.81933    1     1       30.0
EQUIVALENT F          6.61516          1       30.0   0.0153
------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP   1 -------------

VARIABLE  TOLERANCE  F TO REMOVE   WILKS' LAMBDA

ASSETS    1.0000000     6.6152

                      MINIMUM
VARIABLE  TOLERANCE  TOLERANCE  F TO ENTER    WILKS' LAMBDA

INCOME    0.5784563  0.5784563  0.90821E-02      0.81908
DEBT      0.9706667  0.9706667   6.0662          0.67759
FAMSIZE   0.9492947  0.9492947   3.9269          0.72162
JOB       0.9631433  0.9631483   0.47688E-06     0.81933

F STATISTICS AND SIGNIFICANCES BETWEEN PAIRS OF GROUPS AFTER STEP   1
EACH F STATISTIC HAS   1 AND       30.0 DEGREES OF FREEDOM.

                 GROUP        0
   GROUP
       1                  6.6152
                          0.0153
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

AT STEP 2, DEBT       WAS INCLUDED IN THE ANALYSIS.
                             DEGREES OF FREEDOM   SIGNIF.  BETWEEN GROUPS
WILKS' LAMBDA    0.67759        2   1      30.0
EQUIVALENT F     6.89923            2      29.0   0.0035

--------------- VARIABLES IN THE ANALYSIS AFTER STEP   2 ----------------
```

FIGURE 17–4 (*Continued*)

```
VARIABLE  TOLERANCE  F TO REMOVE   WILKS' LAMBDA

ASSETS    0.9706667    7.4487          0.85164
DEBT      0.9706667    6.0662          0.81933

------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP   2 -------------

                      MINIMUM
VARIABLE  TOLERANCE  TOLERANCE  F TO ENTER    WILKS' LAMBDA

INCOME    0.5728383  0.5568120  0.17524E-01      0.67717
FAMSIZE   0.9323959  0.9308959  2.2214           0.62779
JOB       0.9105435  0.9105435  0.27914          0.67091

F STATISTICS AND SIGNIFICANCES BETWEEN PAIRS OF GROUPS AFTER STEP   2
EACH F STATISTIC HAS   2 AND       29.0 DEGREES OF FREEDOM.

                   GROUP        0
     GROUP
         1                6.8992
                          0.0035
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * 

AT STEP   3, FAMSIZE WAS INCLUDED IN  THE ANALYSIS.

                                DEGREES OF FREEDOM   SIGNIF.  BETWEEN GROUPS
WILKS' LAMBDA        0.62779       3     1    30.0
EQUIVALENT F         5.53369             3    28.0    0.0041
--------------- VARIABLES IN THE ANALYSIS AFTER STEP   3 ----------------

VARIABLE  TOLERANCE  F TO REMOVE   WILKS' LAMBDA

ASSETS    0.9308959    8.4282          0.81676
DEBT      0.9533874    4.1849          0.72162
FAMSIZE   0.9323959    2.2214          0.67759

------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP   3 -------------

                      MINIMUM
VARIABLE  TOLERANCE  TOLERANCE  F TO ENTER    WILKS' LAMBDA

INCOME    0.5725772  0.5410775  0.24098E-01      0.62723
JOB       0.8333526  0.8333526  0.86952E-02      0.62759

F STATISTICS AND SIGNIFICANCES BETWEEN PAIRS OF GROUPS AFTER STEP  3
EACH F STATISTIC HAS   3 AND       28.0 DEGREES OF FREEDOM.

                 GROUP        0
   GROUP
       1                5.5337
                        0.0041

F LEVEL OR TOLERANCE OR VIN INSUFFICIENT FOR FURTHER COMPUTATION

                                    SUMMARY TABLE

          ACTION      VARS  WILKS'
STEP ENTERED REMOVED   IN   LAMBDA    SIG.    LABEL

  1  ASSETS             1   .81933   .0153
  2  DEBT               2   .67759   .0035
  3  FAMSIZE            3   .62779   .0041
```

FIGURE 17–5
SPSS-Produced Estimates of the Discriminant Function Coefficients for Example 17–1

```
UNSTANDARDIZED CANONICAL
DISCRIMINANT FUNCTION
COEFFICIENTS

                 FUNC 1

ASSETS        -0.3522450E-01

DEBT           0.4291038E-01

FAMSIZE        0.4882695

(CONSTANT)    -0.9950070
```

The cutting score is zero. Discriminant scores greater than zero (i.e., positive scores) indicate a predicted membership in the default group (population 0), while negative scores imply predicted membership in the repay group (population 1). This can be seen by looking at the predicted group membership chart, Figure 17–6. The figure shows all cases used in the analysis. Since we have no holdout sample for testing the effectiveness

FIGURE 17–6 Predicted Group Membership Chart for Example 17–1

CASE SEGNUM	MIS VAL	SEL	ACTUAL GROUP	HIGHEST GROUP	PROBABILITY P(D/G)	P(G/D)	2ND HIGHEST GROUP	P (G/D)	DISCRIMINANT SCORES...
1			1	1	0.1798	0.9587	0	0.0413	−1.9990
2			1	1	0.3357	0.9293	0	0.0707	−1.6202
3			1	1	0.8840	0.7939	0	0.2061	−0.8034
4			1 **	0	0.4761	0.5146	1	0.4854	0.1328
5			1	1	0.3368	0.9291	0	0.0709	−1.6181
6			1	1	0.5571	0.5614	0	0.4386	−0.0704
7			1 **	0	0.6272	0.5986	1	0.4014	0.3598
8			1	1	0.7236	0.6452	0	0.3548	−0.3039
9			1	1	0.9600	0.7693	0	0.2307	−0.7076
10			1	1	0.3004	0.9362	0	0.0638	−1.6930
11			0	0	0.5217	0.5415	1	0.4585	0.2047
12			0	0	0.6018	0.8714	1	0.1286	1.3672
13			0	0	0.6080	0.5887	1	0.4113	0.3325
14			0	0	0.5083	0.8932	1	0.1068	1.5068
15			0	0	0.8409	0.6959	1	0.3041	0.6447
16			0	0	0.2374	0.9481	1	0.0519	2.0269
17			0	0	0.9007	0.7195	1	0.2805	0.7206
18			0	0	0.8377	0.8080	1	0.1920	1.0502
19			0	0	0.0677	0.9797	1	0.0203	2.6721
20			0	0	0.1122	0.9712	1	0.0288	2.4338
21			0 **	1	0.7395	0.6524	0	0.3476	−0.3250
22			1 **	0	0.9432	0.7749	1	0.2251	0.9166
23			1	1	0.7819	0.6711	0	0.3289	−0.3807
24			0 **	1	0.5294	0.5459	0	0.4541	−0.0286
25			1	1	0.5673	0.8796	0	0.1204	−1.2296
26			1	1	0.1964	0.9557	0	0.0443	−1.9494
27			0 **	1	0.6916	0.6302	0	0.3698	−0.2608
28			1 **	0	0.7479	0.6562	1	0.3438	0.5240
29			1 **	0	0.9211	0.7822	1	0.2178	0.9445
30			1	1	0.4276	0.9107	0	0.0893	−1.4509
31			1	1	0.8188	0.8136	0	0.1864	−0.8866
32			0 **	1	0.8825	0.7124	0	0.2876	−0.5097

of prediction of group membership, the results are for the estimation sample only. For each case, the table gives the *actual group* to which the data point (person, in our example) belongs. A double asterisk (**) next to the actual group indicates that the point was incorrectly classified. The next column, under the heading "Highest Probability: Group," gives the predicted group membership (0 or 1) for every element in our sample.

In a sense, the hit ratio, the overall percentage of cases that were correctly classified by the discriminant function, is similar to the R^2 statistic in multiple regression. The hit ratio is a measure of how well the discriminant function discriminates between groups. When this measure is 100%, the discrimination is very good; when it is small, the discrimination is poor. How small is "small"? Let us consider this problem logically. Suppose that our data set contains 100 observations: 50 in each of the two groups. Now, if we arbitrarily assign all 100 observations to one of the groups, we have a 50% prediction accuracy! We should expect the discriminant function to give us better than 50% correct classification ratio; otherwise we can do as well without it. Similarly, suppose that one group has 75 observations and the other 25. In this case,

FIGURE 17–7 Summary Table of Classification Results for Example 17–1

CLASSIFICATION RESULTS -

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP 0	1
GROUP 0	14	10	4
		71.4%	28.6%
GROUP 1	18	5	13
		27.8%	72.2%

PERCENT OF 'GROUPED' CASES CORRECTLY CLASSIFIED: 71.88%

we get 75% correct classification if we assign all our observations to the large group. Here the discriminant function should give us better than 75% correct classification if it is to be useful.

Another criterion for evaluating the success of the discriminant function is the *proportional chance criterion.*

The **proportional chance criterion** is

$$C = p^2 + (1 - p)^2 \qquad (17\text{–}7)$$

where p is the proportion of observations in one of the two groups (given as a decimal quantity).

In our example, the discriminant function passes both of these tests. The proportions of people in each of the two groups are $14/32 = 0.4375$, and $18/32 = 0.5625$. From Figure 17–7, we know that the hit ratio of the discriminant function is 0.7188 (71.88%). This figure is much higher than that we could obtain by arbitrary assignment (56.25%). The proportional chance criterion, equation 17–7, gives us $C = (0.4375)^2 + (0.5625)^2 = 0.5078$. The hit ratio is clearly larger than this criterion as well. While the hit ratio is better than expected under arbitrary classification, it is not great. We would probably like to have a greater hit ratio if we were to classify loan applicants in a meaningful way. In this case, over 28% may be expected to be incorrectly classified. We must also keep in mind two facts: (1) Our sample size was relatively small, and therefore our inference may be subject to large errors; and (2) our hit ratio is overestimated because it is based on the estimation data. To get a better idea, we would need to use the discriminant function in classifying cases not used in the estimation and see how well the function performs with this data set.

Figure 17–8 shows the locations of the data points in the two groups in relation to their discriminant scores. It is a *map* of the locations of the two groups along the direction of greatest differentiation between the groups (the direction of the discriminant function). Note the overlap of the two groups in the middle of the graph and the separation on the two sides. (Group 0 is denoted by 1s and group 1 by 2s.)

Discriminant Analysis with More Than Two Groups

Discriminant analysis is extendable to more than two groups. When we carry out an analysis with more than two groups, however, we have more than one discriminant function. The first discriminant function is the function that discriminates best among

FIGURE 17–8 **A Map of the Location of the Two Groups for Example 17–1**

```
SYMBOLS USED IN PLOTS
SYMBOL GROUP LABEL
------- ----- --------
  1        0
  2        1

                           ALL-GROUPS STACKED HISTOGRAM

                         CANONICAL DISCRIMINANT FUNCTION 1

      4 +
        |
        |
        |
      3 +                                  2
        |                                  2
F       |                                  2
R       |                                  2
E     2 +                     2            1           2
Q       |                     2            1           2
U       |                     2            1           2
E       |                     2            1           2
N     1 +                22  222  2  222 121 212112211 2 1   11   1
C       |                22  222  2  222 121 212112211 2 1   11   1
Y       |                22  222  2  222 121 212112211 2 1   11   1
        |                22  222  2  222 121 212112211 2 1   11   1
        X---------+---------+---------+---------+---------+---------+
      OUT      -3.0      -2.0      -1.0         0       1.0       2.0
 CLASS 222222222222222222222222222222222222222222221111111111111111111
CENTROIDS                               2               1
```

the *r* groups. The second discriminant function is a function that has zero correlation with the first and has second-best discriminating power among the *r* groups, and so on. With *r* groups, there are $r - 1$ discriminant functions. Thus, with three groups, for example, there are two discriminant functions.

For Example 17–1, suppose that the bank distinguishes three categories: people who repay the loan (group 1), people who default (group 0), and people who have some difficulties and are several months late with payments, but do not default (group 2). The bank has data on a random sample of 46 people, each person falling into one of the three categories. Figure 17–9 shows the classification probabilities and the predicted groups for the new analysis. The classification is based on scores on both discriminant functions. The discriminant scores of each person on each of the two discriminant functions are also shown. Again, double asterisks denote a misclassified case. Figure 17–10 gives the estimated coefficients of the two discriminant functions.

Figure 17–11 gives the classification summary. We see that 86.7% of the group 0 cases were correctly classified by the two discriminant functions, 78.6% of group 1 were correctly classified, and 82.4% of group 2 were correctly classified. The overall percentage of correctly classified cases is 82.61%, which is fairly high.

Figure 17–12 is a scatter plot of the data in the three groups. The figure also shows the three group means. The following figure, Figure 17–13, is especially useful. This is a *territorial map* of the three groups as determined by the pair of estimated discriminant functions. The map shows the boundaries of the plane formed by looking at the pair of scores: (discriminant function 1 score, discriminant function 2 score). Any new point may be classified as belonging to one of the groups depending on where its pair of computed scores makes it fall on the map. For example, a point with the scores 2 on function 1 and -4 on function 2 falls in the territory of group 0

FIGURE 17–9 Predicted Group Membership Chart for Three Groups (extended Example 17–1)

CASE SEGNUM	MIS VAL	SEL	ACTUAL GROUP	HIGHEST GROUP	PROBABILITY P(D/G)	P(G/D)	2ND HIGHEST GROUP	P(G/D)	DISCRIMINANT SCORES...	
1			1	1	0.6966	0.9781	2	0.0198	-2.3023	-0.4206
2			1	1	0.3304	0.9854	2	0.0142	-2.8760	-0.1267
3			1	1	0.9252	0.8584	2	0.1060	-1.2282	-0.3592
4			1	1	0.5982	0.9936	2	0.0040	-2.3031	-1.2574
5			1**	0	0.6971	0.8513	1	0.1098	0.6072	-1.3190
6			1	1	0.8917	0.8293	2	0.1226	-1.1074	-0.3643
7			1**	0	0.2512	0.5769	1	0.4032	-0.0298	-1.8240
8			1	1	0.7886	0.9855	2	0.0083	-1.9517	-1.1657
9			0	0	0.3132	0.4869	1	0.4675	-0.1210	-1.3934
10			0	0	0.4604	0.9951	2	0.0032	2.1534	-1.7015
11			0	0	0.5333	0.9572	1	0.0348	1.0323	-1.9002
12			0	0	0.8044	0.9762	2	0.0204	1.9347	-0.9280
13			0	0	0.6697	0.8395	1	0.1217	0.5641	-1.3381
14			0	0	0.2209	0.7170	2	0.2815	2.2185	0.6586
15			0	0	0.6520	0.9900	2	0.0075	2.0176	-1.3735
16			0	0	0.0848	0.9458	2	0.0541	3.2112	0.3004
17			0**	2	0.2951	0.7983	0	0.1995	1.6393	1.4480
18			1**	0	0.1217	0.6092	1	0.3843	-0.0234	-2.3885
19			0	0	0.6545	0.6144	2	0.3130	0.7054	-0.0932
20			1	1	0.7386	0.9606	2	0.0362	-2.1369	-0.2312
21			1	1	0.0613	0.9498	2	0.0501	-3.2772	0.8831
22			0	0	0.6667	0.6961	1	0.1797	0.3857	-0.7874
23			1	1	0.7659	0.8561	2	0.1320	-1.6001	0.0635
24			0**	1	0.5040	0.4938	2	0.3454	-0.4694	-0.0770
25			2**	1	0.9715	0.8941	2	0.0731	-1.2811	-0.5314
26			2	2	0.6241	0.5767	0	0.2936	0.2503	0.2971
27			2	2	0.9608	0.9420	0	0.0353	0.1808	1.5221
28			2	2	0.9594	0.9183	0	0.0589	0.3557	1.3629
29			2**	0	0.2982	0.5458	2	0.4492	1.6705	0.6994
30			2**	1	0.9627	0.9160	2	0.0462	-1.2538	-0.8067
31			2	2	0.0400	0.9923	0	0.0076	1.7304	3.1894
32			2	2	0.9426	0.9077	1	0.0620	-0.2467	1.3298
33			2	2	0.7863	0.7575	0	0.2075	0.6256	0.8154
34			2	2	0.3220	0.9927	0	0.0060	0.6198	2.6635
35			2	2	0.9093	0.8322	1	0.1113	-0.2519	0.9826
36			2	2	0.5387	0.5528	0	0.4147	0.8843	0.4770
37			2	2	0.7285	0.9752	1	0.0160	-0.1655	2.0088
38			2	2	0.7446	0.9662	1	0.0248	-0.3220	1.9034
39			0	0	0.6216	0.9039	1	0.0770	0.7409	-1.6165
40			2	2	0.9461	0.8737	1	0.0823	-0.2246	1.1434
41			1	1	0.7824	0.9250	2	0.0690	-1.8845	-0.0819
42			0	0	0.3184	0.9647	1	0.0319	1.0456	-2.3016
43			1	1	0.7266	0.7304	0	0.1409	-0.6875	-0.6183
44			2	2	0.8738	0.9561	1	0.0278	-0.1642	1.7082
45			0	0	0.6271	0.9864	2	0.0121	2.2294	-1.0154
46			2	2	0.2616	0.9813	1	0.0175	-0.8946	2.5641

FIGURE 17–10 Estimated Coefficients of the Two Discriminant Functions (extended Example 17–1)

UNSTANDARDIZED CANONICAL DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2
ASSETS	-0.4103059E-01	-0.5688170E-03
INCOME	-0.4325325E-01	-0.6726829E-01
DEBT	0.3644035E-01	0.4154356E-01
FAMSIZE	0.7471749	0.1772388
JOB	0.1787231	-0.4592559E-01
(CONSTANT)	-0.9083139	-3.743060

FIGURE 17–11 Summary Table of Classification Results (extended Example 17–1)

CLASSIFICATION RESULTS

ACTUAL GROUP		NO. OF CASES	PREDICTED GROUP MEMBERSHIP 0	1	2
GROUP	0	15	13	1	1
			86.7%	6.7%	6.7%
GROUP	1	14	3	11	0
			21.4%	78.6%	0.0%
GROUP	2	17	1	2	14
			5.9%	11.8%	82.4%

PERCENT OF 'GROUPED' CASES CORRECTLY CLASSIFIED: 82.61%

FIGURE 17–12 Scatter Plot of the Data (extended Example 17–1)

```
                  ALL-GROUPS SCATTERPLOT - * INDICATES A GROUP CENTROID
C
A                          CANONICAL DISCRIMINANT FUNCTION 1
N     OUT     -6.0      -4.0      -2.0          0      2.0       4.0       6.0        OUT
O       X---------+---------+---------+---------+---------+---------+---------+---------X
N OUT X                                                                                  X
I      |                                                                                 |
C      |                                                                                 |
A      |                                                                                 |
L      |                                                                                 |
       |                                                                                 |
D  4.0 +                                                                                 +
I      |                                                                                 |
S      |                                                     3                           |
C      |                                                                                 |
R      |                                   3       3                                     |
I      |                                                                                 |
M  2.0 +                                     33                                          +
I      |                                      3 3                                        |
N      |                                      3* 3        1                              |
A      |                        2             3                                          |
N      |                                           3       3 1                           |
T      |                                          3 3             1                      |
     0 +                             2  22   1      1                                    +
       |                              22  2                                              |
F      |                                 *3 2     1     *                                |
U      |                                2                    11                          |
N      |                              2       1    1         1                           |
C      |                                         2  1         1                          |
T -2.0 +                                             1                                   +
I      |                                         2   1                                   |
O      |                                                                                 |
N      |                                                                                 |
       |                                                                                 |
2      |                                                                                 |
   OUT X                                                                                 X
O       X---------+---------+---------+---------+---------+---------+---------+---------X
      OUT     -6.0      -4.0      -2.0          0      2.0       4.0       6.0        OUT
```

FIGURE 17–13 **Territorial Map (extended Example 17–1)**

```
                    TERRITORIAL MAP * INDICATES A GROUP CENTROID

                          CANONICAL DISCRIMINANT FUNCTION 1
     -8.0      -6.0      -4.0      -2.0         0       2.0       4.0       6.0      8.0
        +---------+---------+---------+---------+---------+---------+---------+--------+
C  8.0 +                                                                                +
A      |                                                                                |
N      |                                                                                |
O      |                                                                                |
N      | 3                                                                              |
I      | 333                                                                            |
C  6.0 + 22333    +         +         +         +         +         +         +         +
A      |  222333                                                                        |
L      |    222333                                                                      |
       |      222333                                                                    |
D      |        222333                                                                 3|
I      |          222333                                                            3333|
S  4.0 +          +222333   +         +         +         +         +         +  3333111+
C      |              222333                                                   3333111  |
R      |               222333                                               33331111    |
I      |                 222333                                           33331111      |
M      |                   222333                                      33331111         |
I      |                    222333                                  33331111            |
N  2.0 +          +         +  222333 +         +         +      33331111     +         +
A      |                         222333                       33331111                   |
N      |                           222333       *           33331111                     |
T      |                             222333              33331111                        |
       |                               222333        33331111                           |
F      |                                 222333   33331111                              |
U    0 +          +         +         +    2223333331111  +         +         +         +
N      |                                     22231111                                   |
C      |                                   *   221   *                                  |
T      |                                       211                                      |
I      |                                       21                                       |
O      |                                       21                                       |
N -2.0 +          +         +         +        21         +         +         +         +
       |                                       21                                       |
2      |                                       21                                       |
       |                                       21                                       |
       |                                       21                                       |
       |                                       21                                       |
  -4.0 +          +         +         +       221         +         +         +         +
       |                                      211                                       |
       |                                      21                                        |
       |                                      21                                        |
       |                                      21                                        |
       |                                      21                                        |
  -6.0 +          +         +         +       21          +         +         +         +
       |                                      21                                        |
       |                                      21                                        |
       |                                      21                                        |
       |                                      21                                        |
       |                                      21                                        |
  -8.0 +                                     221                                        +
        +---------+---------+---------+---------+---------+---------+---------+--------+
     -8.0      -6.0      -4.0      -2.0         0       2.0       4.0       6.0      8.0
SYMBOLS USED IN TERRITORIAL MAP
SYMBOL  GROUP  LABEL
-----   -----  --------------
  1       0
  2       1
  3       2
  *           GROUP CENTROIDS
```

(this group is denoted by 1 in the plot, as indicated). A group territory is marked by its symbol on the inside of its boundaries with other groups. Group means are also shown, denoted by asterisks.

Many more statistics relevant to discriminant analysis may be computed and reported by computer packages. These are beyond the scope of our discussion, but explanations of these statistics may be found in books on multivariate analysis. This section should give you the basic ideas of discriminant analysis so that you may build on the knowledge acquired here.

PROBLEMS

17–1. What are the purposes of discriminant analysis?

17–2. Suppose that a discriminant analysis is carried out on a data set consisting of two groups. The larger group constitutes 125 observations and the smaller one 89. The relative sizes of the two groups are believed to reflect their relative sizes within the population. If the classification summary table indicates that the overall percentage of correct classification is 57%, would you use the results of this analysis? Why or why not?

17–3. Refer to the results in Figure 17–5 and to equation 17–6. Suppose that a loan applicant has assets of $23,000, debt of $12,000, and a family with three members. How should you classify this person, if you are to use the results of the discriminant analysis? (Remember that debt and assets values are listed without the "000" digits in the program.)

17–4. For problem 17–3, suppose an applicant has $54,000 in assets, $10,000 of debt, and a family of four. How would you classify this applicant? In this problem and the preceding one, be careful to interpret the *sign* of the score correctly.

17–5. Why should you use a holdout data set and try to use the discriminant function for classifying its members? How would you go about doing this?

17–6. A mail-order firm wants to be able to classify people as prospective buyers versus nonbuyers based on some of the people's demographics provided on mailing lists. Prior experience indicates that only 8% of those who receive a brochure end up buying from the company. Use two criteria to determine the minimum overall prediction success rate you would expect from a discriminant function in this case.

17–7. In the situation of problem 17–6, how would you account for the prior knowledge that 8% of the population of those who receive a brochure actually buy?

17–8. Use the territorial map shown in Figure 17–13 to predict group membership for a point with a score of -3 on discriminant function 1 and a score of 0 on discriminant function 2. What about a point with a score of 2 on function 1 and 4 on function 2?

17–9. Use the information in Figure 17–10 and the territorial map in Figure 17–13 to classify a person with assets of $50,000, income of $37,500, debt of $23,000, family size of 2, and 3 years' employment at the current job.

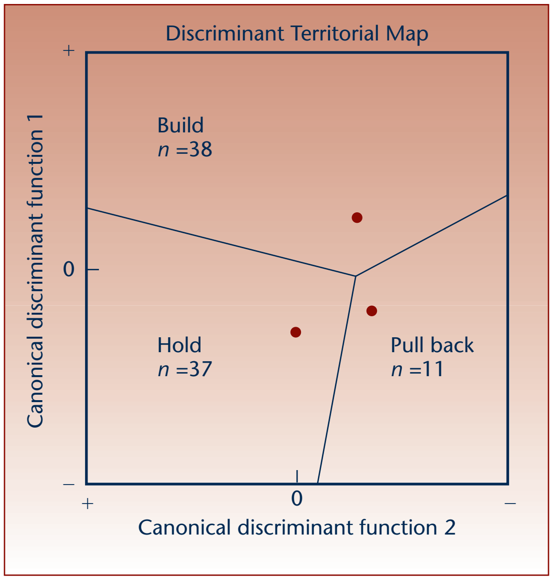
17–10. What are the advantages of a stepwise routine for selection of variables to be included in the discriminant function(s)?

17–11. A discriminant function is estimated, and the p-value based on Wilks' lambda is found to be 0.239. Would you use this function? Explain.

17–12. What is the meaning of $P(\text{G} \mid \text{D})$, and how is it computed when prior information is specified?

17–13. In trying to classify members of a population into one of six groups, how many discriminant functions are possible? Will all these functions necessarily be found significant? Explain.

17–14. A discriminant analysis was carried out to determine whether a firm belongs to one of three classes: build, hold, or pull back. The results, reported in an article in the *Journal of Marketing Research,* include the following territorial map. How would you classify a firm that received a score of 0 on both discriminant functions?



17–4 Principal Components and Factor Analysis

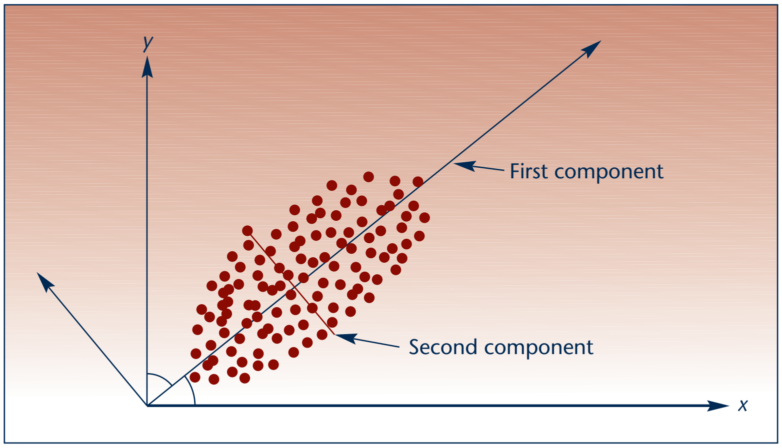
In this section, we discuss two related methods for decomposing the information content in a set of variables into information about an inherent set of latent components. The first method is called **principal component analysis.** Our aim with this method is to decompose the variation in a multivariate data set into a set of components such that the first component accounts for as much of the variation in the data as possible, the second component accounts for the second largest portion of the variation, and so on. In addition, each component in this method of analysis is *orthogonal* to the others; that is, each component is uncorrelated with the others: as a direction in space, each component is at right angles to the others.

In **factor analysis,** which is the second method for decomposing the information in a set of variables, our approach to the decomposition is different. We are not always interested in the orthogonality of the components (in this context, called *factors*); neither do we care whether the proportion of the variance accounted for by the factors decreases as each factor is extracted. Instead, we look for meaningful factors in terms of the particular application at hand. The factors we seek are the underlying, latent dimensions of the problem. The factors summarize the larger set of original variables.

For example, consider the results of a test consisting of answers to many questions administered to a sample of students. If we apply principal-components analysis, we will decompose the answers to the questions into scores on a (usually smaller) set of components that account for successively smaller portions of the variation in the student answers and that are independent of each other. If we apply factor analysis, on the other hand, we seek to group the question variables into a smaller set of meaningful factors. One factor, consisting of responses to several questions, may be a measure of raw intelligence; another factor may be a measure of verbal ability and will consist of another set of questions; and so on.

We start by discussing principal components and then present a detailed description of the techniques of factor analysis. There are two kinds of factor analysis. One is called *R*-factor analysis, and this is the method we will describe. Another is called *Q*-factor analysis. *Q*-factor analysis is a technique where we group the respondents, people or data elements, into sets with certain meanings rather than group the variables.

FIGURE 17–14 **Principal Components of a Bivariate Data Set**



Principal Components

Figure 17–14 shows a data set in two dimensions. Each point in the ellipsoid cluster has two components: X and Y. If we look at the direction of the data cluster, however, we see that it is not oriented along either of the two axes X and Y. In fact, the data are oriented in space at a certain angle to the X axis. Look at the two principal axes of the ellipse of data, and you will notice that one contains much variation along its direction. The other axis, at 90° to the first, represents less variation of the data along its direction. We choose that direction in space about which the data are most variable (the principal axis of the ellipse) and call it the *first principal component*. The *second principal component* is at 90° to the first–it is **orthogonal** to the first. These axes are shown in Figure 17–14. Note that all we really have to do is to rotate the original X and Y axes until we find a direction in space where the principal axis of the elliptical cluster of data lies along this direction. Since this is the larger axis, it represents the largest variation in the data; the data vary most along the direction we labeled *first component*. The second component captures the second-largest variation in the data.

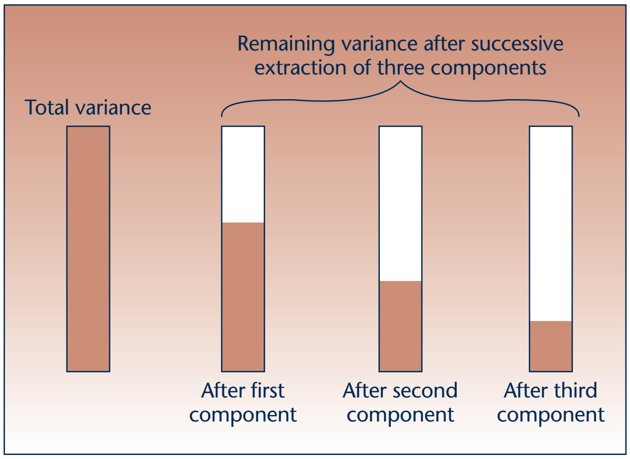
With three variables, there are three directions in space. We find that rotation of the axes of the three variables X, Y, and Z such that the first component is the direction in which the ellipsoid of data is widest. The second component is the direction with the second-largest proportion of variance, and the third component is the direction with the third-largest variation. All three components are at 90° to one another. Such rotations, which preserve the orthogonality (90° angle) of the axes, are called *rigid rotations*. With more variables, the procedure is the same (except that we can no longer graph it). The successive reduction in the variation in the data with the extraction of each component is shown schematically in Figure 17–15.

The Extraction of the Components

The fundamental theorem of principal components is a remarkable mathematical theorem that allows us to find the components. The theorem says that if we have any set of k variables $X_1, X_2, \ldots, X_k$, where the variance-covariance matrix of these variables, denoted Σ, *is invertible* (an algebraic condition you need not worry about), we can always transform the original variables to a set of k uncorrelated variables $Y_1, Y_2, \ldots, Y_k$ by an appropriate rotation. Note that we do not require a normal-distribution assumption.

Can you think of one very good use of principal-component analysis as a preliminary stage for an important statistical technique? Remember the ever-present problem of *multicollinearity* in multiple regression analysis? There the fact that k "independent" variables turned out to be dependent on one another caused many

FIGURE 17–15 **Reduction in the Variance in a Data Set with Successive Extraction of Components**



problems. One solution to the problem of multicollinearity is to transform the original k variables, which are correlated with one another, into a new set of k uncorrelated variables. These uncorrelated variables are the principal components of the data set. Then we can run the regression on the new set, the principal components, and avoid the multicollinearity altogether. We still have to consider, however, the contribution of each original variable to the dependent variable in the regression.

Equation 17–8 is the equation of the first principal component, which is a linear combination of the original k variables $X_1, X_2, \ldots, X_k$.

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1k}X_k \tag{17–8}$$

Similarly, the second principal component is given by

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2k}X_k \tag{17–9}$$

and so on. The a_{ij} are constants, like regression coefficients. The linear combinations are formed by the rotation of the axes.

If we use k new independent variables $Y_1, Y_2, \ldots, Y_k$, then we have accounted for all the variance in the observations. In that case, all we have done is to transform the original variables to linear combinations that are uncorrelated with one another (orthogonal) and that account for all the variance in the observations, the first component accounting for the largest portion, the second for less, and so on. When we use k new variables, however, there is no *economy* in the number of new variables. If, on the other hand, we want to reduce the number of original variables to a smaller set where each new variable has some meaning—each new variable represents a hidden *factor*—we need to use factor analysis. Factor analysis (the R-factor kind), also called *common-factor analysis,* is one of the most commonly used multivariate methods, and we devote the rest of this section to a description of this important method. In factor analysis, we assume a multivariate normal distribution.

Factor Analysis

In factor analysis, we assume that each of the variables we have is made up of a linear combination of common factors (hidden factors that affect the variable and possibly affect other variables) and a specific component unique to the variable.

The k original X_i variables written as linear combinations of a smaller set of m common factors and a unique component for each variable are

$$X_1 = b_{11}F_1 + b_{12}F_2 + \cdots + b_{1m}F_m + U_1$$
$$X_2 = b_{21}F_1 + b_{22}F_2 + \cdots + b_{2m}F_m + U_2$$
$$\vdots$$
$$X_k = b_{k1}F_1 + b_{k2}F_2 + \cdots + b_{km}F_m + U_k \tag{17–10}$$

The F_j, $j = 1, \ldots, m$, are the common factors. Each U_i, $i = 1, \ldots, k$, is the unique component of variable X_i. The coefficients b_{ij} are called *factor loadings.*

The total variance in the data in factor analysis is composed of the common-factor component, called the *communality,* and the *specific* part, due to each variable alone.

The Extraction of Factors

The factors are extracted according to the communality. We determine the number of factors in an analysis based on the percentage of the variation explained by each factor. Sometimes prior considerations lead to the determination of the number of factors. One rule of thumb in determining the number of factors to be extracted considers the total variance explained by the factor. In computer output, the total variance explained by a factor is listed as the *eigenvalue.* (Eigenvalues are roots of determinant equations and are fundamental to much of multivariate analysis. Since understanding them requires some familiarity with linear algebra, we will not say much about eigenvalues, except that they are used as measures of the variance explained by factors.) The rule just mentioned says that a factor with an eigenvalue less than 1.00 should not be used because it accounts for less than the variation explained by a single variable. This rule is conservative in the sense that we probably want to summarize the variables with a set of factors smaller than indicated by this rule. Another, less conservative, rule says that the factors should account for a relatively large portion of the variation in the variables: 80%, 70%, 65%, or any relatively high percentage of the variance. The consideration in setting the percentage is similar to our evaluation of R^2 in regression. There really is no absolute rule.

We start the factor analysis by computing a correlation matrix of all the variables. This diagonal matrix has 1s on the diagonal because the correlation of each variable with itself is equal to 1.00. The correlation in row i and column j of this matrix is the correlation between variables X_i and X_j. The correlation matrix is then used by the computer in extracting the factors and producing the factor matrix. The factor matrix is a matrix showing the factor loadings–the sample correlations between each factor and each variable. These are the coefficients b_{ij} in equation 17–10. Principal-component analysis is often used in the preliminary factor extraction procedure, although other methods are useful as well.

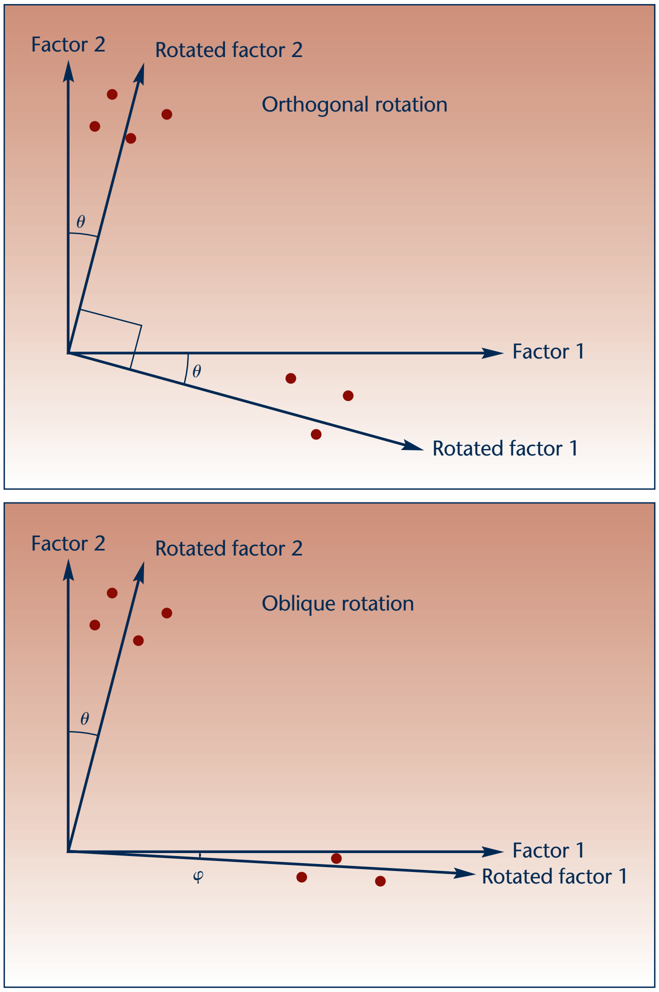
The Rotation of Factors

Once the factors are extracted, the next stage of the analysis begins. In this stage, the factors are *rotated.* The purpose of the rotation is to find the best distribution of the factor loadings in terms of the meaning of the factors. If you think of our hypothetical example of scores of students on an examination, it could be that the initial factors derived (these could be just the principal components) explain proportions of the variation in scores, but not in any meaningful way. The rotation may then lead us to find a factor that accounts for intelligence, a factor that accounts for verbal ability, a third factor that accounts for artistic talent, and so on. The rotation is an integral part of factor analysis and helps us derive factors that are as meaningful as possible. Usually, each of the initially derived factors will tend to be correlated with *many* of the variables. The purpose of the rotation is to identify each factor with only *some* of the variables—different variables with each factor—so that each factor may be interpreted in a meaningful way. Each factor will then be associated with *one* hidden attribute: intelligence, verbal ability, or artistic talent.

There are two classes of rotation methods. One is **orthogonal,** or **rigid, rotation.** Here the axes maintain their orthogonality; that is, they maintain an angle of 90° between every two of them. This means that the factors, once they are rotated, will maintain the quality of being uncorrelated with each other. This may be useful if we believe that the inherent, hidden dimensions in our problem are independent of one another (here this would mean that we believe intelligence is independent of verbal ability and that both are independent of artistic talent). The rigid rotation is also simpler to carry out than nonrigid rotation. A nonrigid rotation is called an **oblique rotation.** In an oblique rotation, we allow the factors to have some correlations among them. We break the initial 90° angles between pairs of axes (pairs of factors), and we seek the *best association* between factors and variables that are included in them, regardless of whether the factors are orthogonal to one another (i.e., at 90° to one another).

Figure 17–16 shows the two possible kinds of rotation. The dots on the graph in each part of the figure correspond to variables, and the axes correspond to factors. In

FIGURE 17–16 **An Orthogonal Factor Rotation and an Oblique Factor Rotation**



the first example, orthogonal rotation, look at the projections of the seven points (seven variables) along the two axes. These are the factor loadings. When we rotate the axes (the factors), maintaining their 90° angle, we find a better fit of the variables with the factors. The top four variables load highly on the shifted vertical axis, while the bottom three variables load highly on the shifted horizontal axis. In the lower figure, we see that an oblique rotation provides a better association of the factors with the variables in this different situation.

There are several algorithms for orthogonal rotation. The most commonly used algorithm is called VARIMAX. The VARIMAX rotation aims at finding a solution where a variable loads highly on one particular factor and loads as lowly as possible on other factors. The algorithm maximizes the sum of the variances of the loadings in the factor matrix; hence the name VARIMAX. When we use this method, our final solution will have factors with loadings that are high on some variables and low on others. This simplifies the interpretation of the factors. Two other methods are QUARTIMAX and EQUIMAX. Since they are less commonly used, we will not discuss them. Let us look at an example.

EXAMPLE 17–2

An analysis of the responses of 1,076 randomly sampled people to a survey about job satisfaction was carried out. The questionnaire contained 14 questions related to satisfaction on the job. The responses to the questions were analyzed using factor analysis with VARIMAX rotation of factors. The results, the four factors extracted and their loadings with respect to each of the original 14 variables, are shown in Table 17–2.

Solution The highest-loading variables are chosen for each factor. Thus, the first factor has loadings of 0.87, 0.88, 0.92, and 0.65 on the questions relabeled as 1, 2, 3, and 4, respectively. After looking at the questions, the analysts named this factor *satisfaction with information*. After looking at the highest-loading variables on the next factor, factor 2, the analysts named this factor *satisfaction with variety*. The two remaining factors

TABLE 17–2 Factor Analysis of Satisfaction Items

	Factor Loadings[a]			
	1	2	3	4
Satisfaction with Information				
1. I am satisfied with the information I receive from my superior about my job performance	$\boxed{0.87}$	0.19	0.13	0.22
2. I receive enough information from my supervisor about my job performance	$\boxed{0.88}$	0.14	0.15	0.13
3. I receive enough feedback from my supervisor on how well I'm doing	$\boxed{0.92}$	0.09	0.11	0.12
4. There is enough opportunity in my job to find out how I am doing	$\boxed{0.65}$	0.29	0.31	0.15
Satisfaction with Variety				
5. I am satisfied with the variety of activities my job offers	0.13	$\boxed{0.82}$	0.07	0.17
6. I am satisfied with the freedom I have to do what I want on my job	0.17	$\boxed{0.59}$	0.45	0.14
7. I am satisfied with the opportunities my job provides me to interact with others	0.18	$\boxed{0.48}$	0.32	0.22
8. There is enough variety in my job	0.11	$\boxed{0.75}$	0.02	0.12
9. I have enough freedom to do what I want in my job	0.17	$\boxed{0.62}$	0.46	0.12
10. My job has enough opportunity for independent thought and action	0.20	$\boxed{0.62}$	0.47	0.06
Satisfaction with Closure				
11. I am satisfied with the opportunities my job gives me to complete tasks from beginning to end	0.17	0.21	$\boxed{0.76}$	0.11
12. My job has enough opportunity to complete the work I start	0.12	0.10	$\boxed{0.71}$	0.12
Satisfaction with Pay				
13. I am satisfied with the pay I receive for my job	0.17	0.14	0.05	$\boxed{0.51}$
14. I am satisfied with the security my job provides me	0.10	0.11	0.15	$\boxed{0.66}$

[a]Varimax rotation. R^2 for each of the four factors is 41.0, 13.5, 8.5, and 7.8, respectively.

were named in a similar way. The key to identifying and interpreting the factors is to look for the variables with highest loadings on each factor and to find a common meaning: a summary name for all the variables loading high on that factor. The VARIMAX rotation is especially useful for such interpretations because it will make each factor have some variables with high loadings and the rest of the variables with low loadings. The factor is then identified with the high-loading variables.

The factor loadings are the standardized regression coefficients in a multiple regression equation of each original variable as dependent, and with the factors as independent variables. When the factors are uncorrelated, as is the case when we use an orthogonal rotation, the total proportion of the variance explained for each variable is equal to the sum of the proportions of the variance explained by all the factors. The proportion of the variance of each variable that is explained by the common factors is the communality. For each variable we therefore have

$$\text{Communality} = \%\text{ variance explained} = \sum_j b_{ij}^2 \qquad (17\text{–}11)$$

where b_{ij} are the coefficients from the appropriate equation for the variable in question from equation 17–11. In this example, we have for the variable "I am satisfied with the information I receive from my superior about my job performance" (variable 1):

$$\text{Communality} = (0.87)^2 + (0.19)^2 + (0.13)^2 + (0.22)^2 = 0.8583, \text{ or } 85.83\%$$

(See the loadings of this variable on the four factors in Table 17–2.) This means that 85.83% of the variation in values of variable 1 is explained by the four factors. We may similarly compute the communality of all other variables. Variable 1 is assigned to factor 1, as indicated in the table. That factor accounts for $(0.87)^2 = 0.7569$, or 75.69% of the variation in this variable. Variable 5, for example, is assigned to factor 2, and factor 2 accounts for $(0.82)^2 = 0.6724$, or 67.24% of the variation in variable 5.

Finally, we mention a use of factor analysis as a preliminary stage for other forms of analysis. We can assign factor scores to each of the respondents (each member of our data set) and then conduct whatever analysis we are interested in, using the factor scores instead of scores on the original variables. This is meaningful when the factors summarize the information in a way that can be interpreted easily.

PROBLEMS

17–15. What is the main purpose of factor analysis?

17–16. What are the differences between factor analysis and principal-components analysis?

17–17. What are the two kinds of factor analysis, and why is one of them more commonly used than the other?

17–18. What are the two kinds of factor rotation? What is the aim of rotating the factors? What is achieved by each of the two kinds of rotation?

17–19. What is achieved by the VARIMAX rotation, and what are two other rotation methods?

In the following problems, we present tables of results of factor analyses in different contexts reported in the marketing research literature. Give a brief interpretation of the findings in each problem.

17–20.

	Rotated Factor Loadings		
	Factor 1 (Scale 3)	Factor 2 (Scale 2)	Factor 3 (Scale 1)
1. Argument evaluation			
a. Supplier's argument	0.31	0.38	<u>0.76</u>
b. User's argument	0.15	0.35	<u>0.85</u>
2. Who-must-yield			
a. Who must give in	0.15	<u>0.85</u>	0.29
b. Who has best case	0.18	<u>0.78</u>	0.37
3. Overall supplier evaluation			
a. Overall impression	<u>0.90</u>	0.18	0.26
b. Buy from in future	<u>0.94</u>	0.14	0.12

17–21.

	Pattern Matrix			
	Factor 1 Price	Factor 2 Retailing/Selling	Factor 3 Advertising	Factor 4 Product
Price item 1	0.37964	−0.11218	0.21009	−0.16767
Price item 2	0.34560	−0.11200	0.18910	−0.09073
Price item 3	0.60497	0.07133	−0.04858	0.03024
Price item 6	0.81856	0.03963	−0.01044	0.01738
Price item 7	0.74661	0.03967	0.00884	−0.06703
Retailing/selling item 1	0.07910	0.74098	−0.02888	0.07095
Retailing/selling item 2	−0.13690	0.58813	0.15950	−0.14141
Retailing/selling item 3	0.01484	0.74749	−0.02151	0.02269
Retailing/selling item 6	−0.05868	0.56753	0.10925	−0.13337
Retailing/selling item 7	0.07788	0.69284	−0.02320	−0.00457
Advertising item 2	−0.03460	−0.03414	0.65854	−0.01691
Advertising item 3	−0.06838	0.01973	0.71499	−0.06951
Advertising item 4	0.01481	−0.00748	0.57196	−0.03100
Advertising item 5	0.20779	0.13434	0.38402	0.12561
Advertising item 7	0.00921	0.11200	0.64330	−0.02534
Product item 2	0.24372	0.16809	−0.05254	−0.33600
Product item 3	−0.00370	0.02951	−0.00013	−0.61145
Product item 5	−0.03193	0.00631	0.04031	−0.78286
Product item 6	0.02346	0.01814	0.09122	−0.73298
Product item 7	0.03854	0.08088	−0.05244	−0.33921

17–22. Name the factors; consider the signs of the loadings.

	Factor 1	Factor 2	Factor 3	Factor 4
Importance 1	0.59			
Importance 2	0.56			
Importance 3	0.62			
Importance 4	0.74			
Pleasure 1		−0.73		
Pleasure 2		−0.68		
Pleasure 3		−0.82		
Pleasure 4		−0.67		
Pleasure 5		−0.58		
Sign 1			0.78	
Sign 2			0.94	
Sign 3			0.73	
Sign 4			0.77	

	Factor 1	Factor 2	Factor 3	Factor 4
Risk importance 1	0.62			
Risk importance 2	0.74			
Risk importance 3	0.74			
Risk probability 1				0.76
Risk probability 2				0.64
Risk probability 3				0.50

Omitted loadings are inferior to 0.25.

17–23. Identify each of the variables with one factor only. Also find the communality of each variable.

Business Policy	Factor 1 Market Penetration Issues	Factor 2 Project Quality Issues
Pricing policies	0.331	0.626
Record and reporting procedures	0.136	0.242
Advertising copy and expenditures	0.468	0.101
Selection of sources of operating supplies	0.214	0.126
Customer service and complaints	−0.152	0.792
Market forecasting and performance standards	0.459	0.669
Warranty decisions	0.438	0.528
Personnel staffing and training	0.162	0.193
Product delivery scheduling	0.020	0.782
Construction/installation procedures	0.237	0.724
Subcontracting agreements	−0.015	0.112
Number of dealerships	0.899	0.138
Location of dealerships	0.926	0.122
Trade areas	0.885	0.033
Size of building projects	0.206	0.436
Building design capabilities	−0.047	0.076
Sales promotion materials	0.286	0.096
Financial resources	0.029	0.427
Builder reputation	0.076	0.166
Offering competitors' lines	0.213	0.111
Variance explained	3.528	3.479
Percentage of total variance	17.64	17.39
Reliability (coefficient σ)	0.94	0.83

17–24. Name the factors.

	Factor 1	Factor 2
Developing end-user preferences	−0.04	0.88
Product quality and technical leadership	0.19	0.65
Sales promotion programs and promotional aids	−0.11	0.86
Pricing policy	0.78	−0.03
Return-goods policy	0.79	−0.04
Product availability (delivery and reliability)	0.63	0.26
Cooperativeness and technical competence of its personnel	0.59	0.45

17–25. Telephone interviewing is widely used in random sampling, in which the telephone numbers are randomly selected. Under what conditions is this methodology flawed, and why?

17–5 Using the Computer

Using MINITAB for Discriminant and Principal Component Analysis

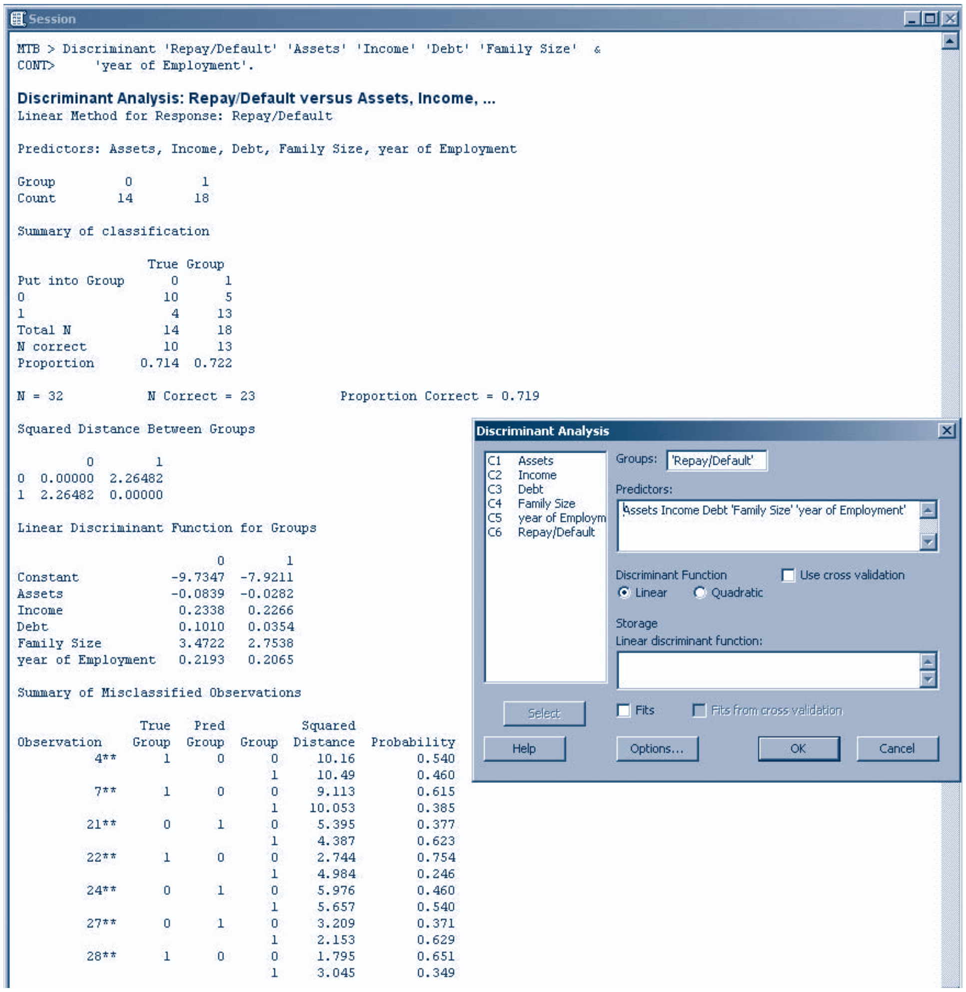
When you have a sample with known groups, you can use the MINITAB discriminant analysis tool to classify observations into two or more groups. The two available options in MINITAB are linear and quadratic discriminant analysis. With linear discriminant analysis all groups are assumed to have the same variance-covariance matrices and possibly different means. In order to start, choose Stat ▶ Multivariate ▶ Discriminant Analysis from the menu bar. When the corresponding dialog box appears, you need to choose the column containing the group codes in the Groups edit box. You can define up to 20 groups. The column(s) containing the measurement variables or predictors are entered in the Predictors dialog box. Then you can choose to perform linear discriminant analysis or quadratic discriminant analysis. Check the Use cross validation box to perform the discrimination using cross-validation. The cross-validation routine works by removing each observation at a time, recalculating the classification function using the remaining data, and then classifying the omitted observation. In the Linear discriminant function edit box enter storage columns for the coefficients from the linear discriminant function. MINITAB uses one column for each group. The constant is stored at the top of each column. By clicking on the Options button you can specify prior probabilities, predict group membership for new observations, and control the display of the Session window output.

As an example we used the data set of Example 17–1 to run a discriminant analysis using MINITAB. Figure 17–17 shows the MINITAB Discriminant Analysis dialog box, corresponding Session commands, and final results. As we can see, the Repay/Default column was defined as the column that contains the group codes. All other variables were entered as predictors. The linear discriminant analysis correctly identified 23 of 32 applicants, as shown in the Summary of classification table in the Session window. To identify new applicants as members of a particular group, you can compute the linear discriminant function associated with Repay or Default and then choose the group for which the discriminant function value is higher. The coefficients of the discriminant functions are seen in the Linear Discriminant Function for Groups table.

MINITAB is also used for principal component analysis. For this purpose, you need to set up your worksheet so that each row contains measurements on a single item. There must be two or more numeric columns, each of which represents a different response variable. To perform principal component analysis, start by choosing Stat ▶ Multivariate ▶ Principal Components from the menu bar. When the corresponding dialog box appears, enter the columns containing the variables to be included in the analysis in the Variables edit box. The Number of components to compute edit box will contain the number of principal components to be extracted. If you do not specify the number of components and m variables are selected, then m principal components will be extracted. Click the Correlation box if you wish to calculate the principal components using the correlation. This case usually happens when the variables are measured by different scales and you want to standardize variables. If you don't wish to standardize variables, choose the Covariance edit box. By clicking on the Graphs button, you can set to display plots for judging the importance of the different principal components and for examining the scores of the first two principal components. The Storage button enables you to store the coefficients and scores.

Another MINITAB tool that enables you to summarize the data covariance structure in a few dimensions of data is factor analysis. For this purpose, you can have three types of input data: columns of raw data, a matrix of correlations or covariances, or columns containing factor loadings. If you set your worksheet to contain the raw data, then each row represents the measurements on a single item.

FIGURE 17–17 Discriminant Analysis Using MINITAB



There must be two or more numeric columns, with each column representing a different response variable. To perform factor analysis with raw data, choose Stat ▸ Multivariate ▸ Factor Analysis from the menu bar. Enter the columns containing the variables you want to use in the analysis in the Variables edit box. Then you need to specify the number of factors to extract. As a Method of Extraction, choose Principal components to use the principal components method of factor extraction. Type of Rotation in the next section controls orthogonal rotations. If you want to use a stored correlation or covariance matrix, or the loadings from a previous analysis instead of the raw data, click Options. The corresponding dialog box allows you to specify the matrix type and source, and the loadings to use for the initial extraction. The Graphs button enables you to display a plot and score and loading plots for the first two factors. The Storage and Results buttons have the same functionality as before.

17–6 Summary and Review of Terms

There is a large body of statistical techniques called **multivariate methods.** These methods are useful in analyzing data in situations that involve several variables. The data and parameters are **vectors.** In this chapter, we discussed some of these methods. We began by describing the **multivariate normal distribution.** We then discussed **discriminant analysis,** a method of classifying members of a population into one of two (or more) groups. The analysis entailed the postulation and estimation of one or more **discriminant functions.** We then discussed **factor analysis,** a statistical technique for reducing the dimensionality of a problem by summarizing a set of variables as a smaller set of inherent, latent common factors. We also discussed a related technique often used as a first stage in factor analysis: **principal-components analysis.** We discussed the concept of independence in several dimensions: the concept of **orthogonality** of factors or variables. We discussed rotations used in factor analysis: **orthogonal rotations,** which maintain the noncorrelation of the factors, and **oblique rotations,** which do not.

ADDITIONAL PROBLEMS

17–26. What are the uses of the multivariate normal distribution? Why is it needed?

17–27. How many discriminant functions may be found significant in classifying an observation into one of four groups?

17–28. Is it possible that only one discriminant function will be found significant in discriminating among three groups? Explain.

17–29. What is a hit ratio? Will a hit ratio of 67% be sufficient when one group has 100 observations, another group has 200 observations, and the ratio of these groups is believed to reflect their ratio in the population? Explain.

17–30. What is achieved by principal-components analysis? How can it be used as a preliminary stage in factor analysis? What important stages must follow it, and why?

17–31. In a factor analysis of 17 variables, a solution is found consisting of 17 factors. Comment on the analysis.

17–32. When is an oblique rotation superior to an orthogonal one, and why?

17–33. What is a communality, and what does it indicate?

17–34. What is the communality of variable 9 listed in Table 17–2?

17–35. Name a statistical method of analysis for which principal components may be a first stage. Explain.

17–36. What are factor loadings, and what do they measure?

17–37. A television producer wants to predict the success of new television programs. A program is considered successful if it survives its first season. Data on production costs, number of sponsors, and the total amount spent on promoting the program are available. A random sample of programs is selected, and the data are presented in the following table. Production costs are in millions of dollars, and promotions in hundreds of thousands of dollars; S denotes success and F failure.

Success/Failure	Production Cost	Number of Sponsors	Promotions
S	2.5	1	2.1
F	1.1	1	3.7
S	2.0	1	2.8
F	2.5	1	1.8
F	2.5	1	1.0
S	3.7	2	5.5
S	3.6	3	5.0

Success/Failure	Production Cost	Number of Sponsors	Promotions
S	2.7	1	4.1
S	1.9	1	6.9
F	1.3	1	1.5
S	2.6	2	2.0
S	3.5	3	3.8
F	1.9	1	1.0
S	6.8	1	2.1
S	5.0	1	1.9
S	4.6	3	4.1
S	3.0	2	3.9
S	2.5	3	7.0
S	1.8	2	6.6
F	2.0	2	1.1
S	3.5	3	3.8
S	1.8	3	8.1
S	5.6	1	4.2
F	1.5	2	3.0
S	4.4	4	5.0

Conduct a discriminant analysis, and state your conclusions.

17–38. A factor analysis was conducted with 24 variables. The VARIMAX rotation was used, and the results were two factors. Comment on the analysis.

17–39. Suppose a major real estate development corporation has hired you to research the features of housing for which people will pay the most in making a home purchasing decision. Where would you start? Perhaps you would start with demographic data from the Department of Housing and Urban Development, www.huduser.org. From the Data Available area, locate and examine the American Housing Survey. Read the most recent Survey Quick Facts. Based on this information, design a principal-components demographic analysis for determining the purchase price of a new house. Would you be able to conduct the analysis entirely from data and other information available at the Hud User site? Why or why not?

www.exercise

CASE 22 Predicting Company Failure

The following article is reprinted in its entirety by permission from *Forbes* (April 1, 1991). Discuss the statistical method alluded to in this article. Could you reproduce (or improve upon) Professor Platt's results? Explain.

Is a former $40 stock now trading at $2.50 a bargain? Or an invitation to get wiped out? Depends.

How Cheap?

Harlan Platt, an associate professor of finance at Northeastern University, has written *Why Companies Fail,* a study that should be of considerable interest to bargain-hunting investors.

Platt's study can be useful in helping determine whether a stock that has fallen sharply in price is a bargain or a prospective piece of wallpaper.

Platt developed a mathematical model that predicts the probability of bankruptcy from certain ratios on a company's balance sheet.

Here's the thesis: Some companies trading very cheaply still have large sales, considerable brand recognition and a chance at recovery, or at least takeover at a premium. Their stocks could double or triple despite losses and weak balance sheets. Other borderline companies will land in bankruptcy court and leave common shareholders with nothing.

Even though it more than tripled from its October low, Unisys, with $10 billion in sales, is not a Wall Street favorite: At a recent 5½ its market capitalization is only $890 million. Will Unisys fail? Almost certainly not within the next 12 months, according to Platt.

For the list below, we found cheap stocks with low price-to-sales ratios. Then we eliminated all but the ones Platt says are highly unlikely to fail within a year. Platt put cheap stocks such as Gaylord Container, Masco Industries and Kinder-Care Learning Centers in the danger zone.

Among low-priced stocks, Unisys and Navistar, however, make the safety grade. So does Wang Laboratories. Says Platt, "They are still selling over $2 billion worth of computers and their $575 million in bank debt is now down to almost nothing."

Platt, who furnishes his probabilities to Prospect Street Investment Management Co., a Boston fund manager, refuses to disclose his proprietary formula. But among the ratios he considers are total debt to total assets, cash flow to sales, short-term debt to total debt, and fixed assets to total assets.

"Companies with large fixed assets are more likely to have trouble because these assets are less liquid," Platt says. But norms for a company's industry are also important. An unusually high level of such current assets as inventory and receivables may itself be a sign of weakness.

The low-priced stocks on the list may or may not rise sharply in the near future, but they are not likely to disappear into insolvency.

–STEVE KICHEN

Big Companies with Little Prices

These 10 companies are in poor financial shape, but not so poor, according to calculations by finance professor Harlan Platt, that they are likely to go bankrupt within the next year. Thus, these stocks are plausible bets for rebounds.

Company/Industry	Recent Price	Earnings per Share: Latest 12 Months	Earnings per Share: 1991 Estimated	Total Assets ($mil)	Total Debt/Total Assets	Sales ($mil)	Cash Flow/ Sales	Price/ Sales
Highland Superstores/consumer electronics stores	$2^1/_8$	$−0.89	NA	$320	57%	$892	−1.6%	0.04
Businessland/computer stores	$2^1/_2$	−1.65	$−1.02	616	72	1,306	−2.0	0.06
Jamesway/discount stores	$3^1/_2$	−0.06	0.18	435	61	909	1.6	0.06
Merisel/computer equipment wholesaler	$3^1/_8$	0.03	0.33	432	73	1,192	0.4	0.06
Unisys/computers	$5^1/_2$	−3.45	−0.56	10,484	65	10,111	3.1	0.09
National Convenience Stores/convenience stores	$5^1/_8$	−0.43	0.21	406	65	1,067	2.0	0.11
TW Holdings/restaurants	$4^7/_{16}$	−0.61	−0.43	3,531	79	3,682	3.9	0.13
Varity/farm and construction equipment	$2^3/_4$	0.35	0.32	3,177	61	3,472	6.4	0.20
Wang Laboratories CIB/minicomputers	$3^3/_8$	4.04	−0.29	1,750	72	2,369	−20.1	0.24
Navistar International/trucks	$4^1/_8$	−0.24	−0.08	3,795	60	3,810	−0.6	0.27

NA: Not available.

From "Big Companies with Little Prices," *Forbes*, April 1, 1991; Harlan Platt, Northeastern University; Institutional Brokers Estimate System (a service of Lynch, Jones & Ryan), via Lotus One Source; *Forbes*.