CHAPTER 22

Stage 1
Problem definition

Stage 2
Research approach developed

Stage 3
Research design developed

Stage 4
Fieldwork or data collection

**Stage 5**
**Data preparation and analysis**

Stage 6
Report preparation and presentation

# Factor analysis

## Objectives

After reading this chapter, you should be able to:

1 describe the concept of factor analysis and explain how it is different from analysis of variance, multiple regression and discriminant analysis;

2 discuss the procedure for conducting factor analysis, including problem formulation, construction of the correlation matrix, selection of an appropriate method, determination of the number of factors, rotation and interpretation of factors;

3 understand the distinction between principal component factor analysis and common factor analysis methods;

4 explain the selection of surrogate variables and their application with emphasis on their use in subsequent analysis;

5 describe the procedure for determining the fit of a factor analysis model using the observed and the reproduced correlations.

**Factor analysis allows an examination of the potential interrelationships among a number of variables and the evaluation of the underlying reasons for these relationships.**

## Overview

In analysis of variance (Chapter 19), regression (Chapter 20) and discriminant analysis (Chapter 21), one of the variables is clearly identified as the dependent variable. We now turn to a procedure, factor analysis, in which variables are not classified as independent or dependent. Instead, the whole set of interdependent relationships among variables is examined. This chapter discusses the basic concept of factor analysis and gives an exposition of the factor model. We describe the steps in factor analysis and illustrate them in the context of principal components analysis. Next, we present an application of common factor analysis. To begin, we provide some examples to illustrate the usefulness of factor analysis.

*example*
*GlobalCash Project*

### Factor analysis[1]

In the GlobalCash Project, the respondents' ratings of 11 service quality statements were factor analysed to determine the underlying service quality factors. Four factors emerged: close support, speed of activities, coping with errors and matching efficiency expectations. These factors, along with individual country characteristics, were used to profile market segments formed as a result of clustering. ■

*example*

### Personal alarms[2]

In a study of personal alarms, women were asked to rate eight personal alarms using the following 15 statements:

1 Feels comfortable in the hand
2 Could be easily kept in the pocket
3 Would fit easily into a handbag
4 Could be easily worn on the person
5 Could be carried to be very handy when needed
6 Could be set off almost as a reflex action
7 Would be difficult for an attacker to take it off me
8 Could keep a very firm grip on it if attacked
9 An attacker might be frightened that I might attack him with it
10 Would be difficult for an attacker to switch off
11 Solidly built
12 Would be difficult to break
13 Looks as if it would give off a very loud noise
14 An attacker might have second thoughts about attacking me if he saw me with it
15 I would be embarrassed to carry it around with me

The question was 'could these 15 variables be reduced to a smaller number of derived variables, known as factors, in such a way that too much information was not lost?' Factor analysis enabled these 15 variables to be reduced to four underlying dimensions or factors that women used to evaluate the alarms. Factor 1 seemed to measure a dimension of '*size*', on a continuum of small to large. Factor 2 tapped into aspects of the '*appearance*' of a personal alarm. Factor 3 revealed '*robustness*' characteristics, with factor 4 related to '*hand feel*'. ■

## Basic concept

**Factor analysis**
A class of procedures primarily used for data reduction and summarisation.

**Factor**
An underlying dimension that explains the correlations among a set of variables.

**Factor analysis** is a general name denoting a class of procedures primarily used for data reduction and summarisation. In marketing research, there may be a large number of variables, most of which are correlated and which must be reduced to a manageable level. Relationships among sets of many interrelated variables are examined and represented in terms of a few underlying factors. For example, bank image may be measured by asking respondents to evaluate banks on a series of items on a semantic differential scale or a Likert scale. These item evaluations may then be analysed to determine the **factors** underlying bank image.

**Interdependence technique**
A multivariate statistical technique in which the whole set of interdependent relationships is examined.

In analysis of variance, multiple regression and discriminant analysis, one variable is considered the dependent or criterion variable, and the others are considered independent or predictor variables. But no such distinction is made in factor analysis. Rather, factor analysis is an **interdependence technique** in that an entire set of interdependent relationships is examined.[3]

Factor analysis is used in the following circumstances:

1. To identify underlying dimensions, or factors, that explain the correlations among a set of variables. For example, a set of lifestyle statements may be used to measure the psychographic profiles of consumers. These statements may then be factor analysed to identify the underlying psychographic factors.[4]
2. To identify a new, smaller, set of uncorrelated variables to replace the original set of correlated variables in subsequent multivariate analysis (regression or discriminant analysis). For example, the psychographic factors identified may be used as independent variables in explaining the differences between loyal and non-loyal consumers.
3. To identify a smaller set of salient variables from a larger set for use in subsequent multivariate analysis. For example, a few of the original lifestyle statements that correlate highly with the identified factors may be used as independent variables to explain the differences between the loyal and non-loyal users.

Factor analysis has numerous applications in marketing research. For example:

- Factor analysis can be used in market segmentation for identifying the underlying variables on which to group the customers. New car buyers might be grouped based on the relative emphasis they place on economy, convenience, performance, comfort and luxury. This might result in five segments: economy seekers, convenience seekers, performance seekers, comfort seekers and luxury seekers.
- In product research, factor analysis can be employed to determine the brand attributes that influence consumer choice. Toothpaste brands might be evaluated in terms of protection against cavities, whiteness of teeth, taste, fresh breath and price.
- In advertising studies, factor analysis can be used to understand the media consumption habits of the target market. The users of frozen foods may be heavy viewers of satellite TV, see a lot of videos, and listen to country music.
- In pricing studies, factor analysis can be used to identify the characteristics of price-sensitive consumers. For example, these consumers might be methodical, economy minded and home centred.

## Factor analysis model

Mathematically, factor analysis is somewhat similar to multiple regression analysis in that each variable is expressed as a linear combination of underlying factors. The amount of variance a variable shares with all other variables included in the analysis is referred to as communality. The covariation among the variables is described in terms of a small number of common factors plus a unique factor for each variable. These factors are not overtly observed. If the variables are standardised, the factor model may be represented as

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \dots + A_{im}F_m + V_iU_i$$

where $X_i$ = $i$th standardised variable
$A_{ij}$ = standardised multiple regression coefficient of variable $i$ on common factor $j$
$F$ = common factor
$V_i$ = standardised regression coefficient of variable $i$ on unique factor $i$
$U_i$ = the unique factor for variable $i$
$m$ = number of common factors.

The unique factors are correlated with each other and with the common factors.[5] The common factors themselves can be expressed as linear combinations of the observed variables

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \ldots + W_{ik}X_k$$

where $F_i$ = estimate of $i$th factor
$W_i$ = weight or factor score coefficient
$k$ = number of variables.

It is possible to select weights or factor score coefficients so that the first factor explains the largest portion of the total variance. Then a second set of weights can be selected so that the second factor accounts for most of the residual variance, subject to being uncorrelated with the first factor. This same principle could be applied to selecting additional weights for the additional factors. Thus, the factors can be estimated so that their factor scores, unlike the values of the original variables, are not correlated. Furthermore, the first factor accounts for the highest variance in the data, the second factor the second highest, and so on. A technical treatment of the factor analysis model is presented in Appendix 22A.

The key statistics associated with factor analysis are as follows:

**Bartlett's test of sphericity**. Bartlett's test of sphericity is a test statistic used to examine the hypothesis that the variables are uncorrelated in the population. In other words, the population correlation matrix is an identity matrix; each variable correlates perfectly with itself ($r = 1$) but has no correlation with the other variables ($r = 0$).

**Communality**. Communality is the amount of variance a variable shares with all the other variables being considered. This is also the proportion of variance explained by the common factors.

**Correlation matrix**. A correlation matrix is a lower triangle matrix showing the simple correlations, $r$, between all possible pairs of variables included in the analysis. The diagonal elements, which are all 1, are usually omitted.

**Eigenvalue**. The eigenvalue represents the total variance explained by each factor.

**Factor loadings**. Factor loadings are simple correlations between the variables and the factors.

**Factor loading plot**. A factor loading plot is a plot of the original variables using the factor loadings as coordinates.

**Factor matrix**. A factor matrix contains the factor loadings of all the variables on all the factors extracted.

**Factor scores**. Factor scores are composite scores estimated for each respondent on the derived factors.

**Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy**. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is an index used to examine the appropriateness of factor analysis. High values (between 0.5 and 1.0) indicate that factor analysis is appropriate. Values below 0.5 imply that factor analysis may not be appropriate.

**Percentage of variance**. The percentage of the total variance attributed to each factor.

**Residuals**. Residuals are the differences between the observed correlations, as given in the input correlation matrix, and the reproduced correlations, as estimated from the factor matrix.

**Scree plot**. A scree plot is a plot of the eigenvalues against the number of factors in order of extraction.

We describe the uses of these statistics in the next section, in the context of the procedure for conducting factor analysis.

## Conducting factor analysis

The steps involved in conducting factor analysis are illustrated in Figure 22.1. The first step is to define the factor analysis problem and identify the variables to be factor analysed. Then a correlation matrix of these variables is constructed and a method of factor analysis is selected. The researcher decides on the number of factors to be extracted and the method of rotation. Next, the rotated factors should be interpreted. Depending on the objectives, the factor scores may be calculated, or surrogate variables selected, to represent the factors in subsequent multivariate analysis. Finally, the fit of the factor analysis model is determined. We discuss these steps in more detail below.[6]

### Formulate the problem

Formulating the problem includes several tasks. First, the objectives of factor analysis should be identified. The variables to be included in the factor analysis should be specified based on past research (quantitative or qualitative), theory, and judgement of the researcher. It is important that the variables be appropriately measured on an interval or ratio scale. An appropriate sample size should be used. As a rough guideline, there should be at least four or five times as many observations (sample size) as there are variables.[7] In many marketing research situations, the sample size is small, and this ratio is considerably lower. In these cases, the results should be interpreted cautiously.

To illustrate factor analysis, suppose that the researcher wants to determine the underlying benefits consumers seek from the purchase of a toothpaste. A sample of 237 respondents was interviewed using street interviewing. The respondents were asked to indicate their degree of agreement with the following statements using a seven-point scale (1 = strongly disagree, 7 = strongly agree):

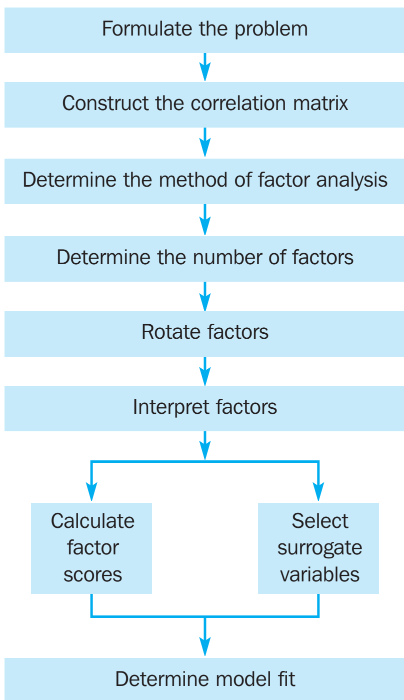


**Figure 22.1**
**Conducting factor analysis**

$V_1$ It is important to buy a toothpaste that prevents cavities.
$V_2$ I like a toothpaste that gives shiny teeth.
$V_3$ A toothpaste should strengthen your gums.
$V_4$ I prefer a toothpaste that freshens breath.
$V_5$ Prevention of tooth decay should be an important benefit offered by a toothpaste.
$V_6$ The most important consideration in buying a toothpaste is attractive teeth.

The data obtained are given in Table 22.1. A correlation matrix was constructed based on these ratings data.

**Table 22.1 Toothpaste attribute ratings**

| *Respondent number* | $V_1$ | $V_2$ | $V_3$ | $V_4$ | $V_5$ | $V_6$ |
|---|---|---|---|---|---|---|
| 1 | 7.00 | 3.00 | 6.00 | 4.00 | 2.00 | 4.00 |
| 2 | 1.00 | 3.00 | 2.00 | 4.00 | 5.00 | 4.00 |
| 3 | 6.00 | 2.00 | 7.00 | 4.00 | 1.00 | 3.00 |
| 4 | 4.00 | 5.00 | 4.00 | 6.00 | 2.00 | 5.00 |
| 5 | 1.00 | 2.00 | 2.00 | 3.00 | 6.00 | 2.00 |
| 6 | 6.00 | 3.00 | 6.00 | 4.00 | 2.00 | 4.00 |
| 7 | 5.00 | 3.00 | 6.00 | 3.00 | 4.00 | 3.00 |
| 8 | 6.00 | 4.00 | 7.00 | 4.00 | 1.00 | 4.00 |
| 9 | 3.00 | 4.00 | 2.00 | 3.00 | 6.00 | 3.00 |
| 10 | 2.00 | 6.00 | 2.00 | 6.00 | 7.00 | 6.00 |
| 11 | 6.00 | 4.00 | 7.00 | 3.00 | 2.00 | 3.00 |
| 12 | 2.00 | 3.00 | 1.00 | 4.00 | 5.00 | 4.00 |
| 13 | 7.00 | 2.00 | 6.00 | 4.00 | 1.00 | 3.00 |
| 14 | 4.00 | 6.00 | 4.00 | 5.00 | 3.00 | 6.00 |
| 15 | 1.00 | 3.00 | 2.00 | 2.00 | 6.00 | 4.00 |
| 16 | 6.00 | 4.00 | 6.00 | 3.00 | 3.00 | 4.00 |
| 17 | 5.00 | 3.00 | 6.00 | 3.00 | 3.00 | 4.00 |
| 18 | 7.00 | 3.00 | 7.00 | 4.00 | 1.00 | 4.00 |
| 19 | 2.00 | 4.00 | 3.00 | 3.00 | 6.00 | 3.00 |
| 20 | 3.00 | 5.00 | 3.00 | 6.00 | 4.00 | 6.00 |
| 21 | 1.00 | 3.00 | 2.00 | 3.00 | 5.00 | 3.00 |
| 22 | 5.00 | 4.00 | 5.00 | 4.00 | 2.00 | 4.00 |
| 23 | 2.00 | 2.00 | 1.00 | 5.00 | 4.00 | 4.00 |
| 24 | 4.00 | 6.00 | 4.00 | 6.00 | 4.00 | 7.00 |
| 25 | 6.00 | 5.00 | 4.00 | 2.00 | 1.00 | 4.00 |
| 26 | 3.00 | 5.00 | 4.00 | 6.00 | 4.00 | 7.00 |
| 27 | 4.00 | 4.00 | 7.00 | 2.00 | 2.00 | 5.00 |
| 28 | 3.00 | 7.00 | 2.00 | 6.00 | 4.00 | 3.00 |
| 29 | 4.00 | 6.00 | 3.00 | 7.00 | 2.00 | 7.00 |
| 30 | 2.00 | 3.00 | 2.00 | 4.00 | 7.00 | 2.00 |

## Construct the correlation matrix

The analytical process is based on a matrix of correlations between the variables. Valuable insights can be gained from an examination of this matrix. For factor analysis to be meaningful, the variables should be correlated. In practice, this is usually the case. If the correlations between all the variables are small, factor analysis may not be appropriate. We would also expect that variables that are highly correlated with each other would also highly correlate with the same factor or factors.

Formal statistics are available for testing the appropriateness of the factor model. Bartlett's test of sphericity can be used to test the null hypothesis that the variables are uncorrelated in the population; in other words, the population correlation matrix is an identity matrix. In an identity matrix, all the diagonal terms are 1, and all off-diagonal terms are 0. The test statistic for sphericity is based on a chi-square transformation of the determinant of the correlation matrix. A large value of the test statistic will favour the rejection of the null hypothesis. If this hypothesis cannot be rejected, then the appropriateness of factor analysis should be questioned. Another useful statistic is the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. This index compares the magnitudes of the observed correlation coefficients with the magnitudes of the partial correlation coefficients. Small values of the KMO statistic indicate that the correlations between pairs of variables cannot be explained by other variables and that factor analysis may not be appropriate.

The correlation matrix, constructed from the data obtained to understand toothpaste benefits, is shown in Table 22.2. There are relatively high correlations among $V_1$ (prevention of cavities), $V_3$ (strong gums) and $V_5$ (prevention of tooth decay). We would expect these variables to correlate with the same set of factors. Likewise, there are relatively high correlations among $V_2$ (shiny teeth), $V_4$ (fresh breath) and $V_6$ (attractive teeth). These variables may also be expected to correlate with the same factors.[8]

**Table 22.2 Correlation matrix**

| Variables | $V_1$ | $V_2$ | $V_3$ | $V_4$ | $V_5$ | $V_6$ |
|---|---|---|---|---|---|---|
| $V_1$ | 1.00 | | | | | |
| $V_2$ | −0.053 | 1.00 | | | | |
| $V_3$ | 0.873 | −0.155 | 1.00 | | | |
| $V_4$ | −0.086 | 0.572 | −0.248 | 1.00 | | |
| $V_5$ | −0.858 | 0.020 | −0.778 | −0.007 | 1.00 | |
| $V_6$ | 0.004 | 0.640 | −0.018 | 0.640 | −0.136 | 1.00 |

The results of factor analysis are given in Table 22.3. The null hypothesis, that the population correlation matrix is an identity matrix, is rejected by Bartlett's test of sphericity. The approximate chi-square statistic is 111.314 with 15 degrees of freedom which is significant at the 0.05 level. The value of the KMO statistic (0.660) is also large (> 0.5). Thus factor analysis may be considered an appropriate technique for analysing the correlation matrix of Table 22.2.

## Determine the method of factor analysis

Once it has been determined that factor analysis is an appropriate technique for analysing the data, an appropriate method must be selected. The approach used to derive the weights or factor score coefficients differentiates the various methods of factor analysis. The two basic approaches are principal components analysis and common factor analysis. In principal components analysis, the total variance in the

**Principal components analysis**
An approach to factor analysis that considers the total variance in the data.

**Common factor analysis**
An approach to factor analysis that estimates the factors based only on the common variance. Also called principal axis factoring.

data is considered. The diagonal of the correlation matrix consists of unities, and full variance is brought into the factor matrix. **Principal components analysis** is recommended when the primary concern is to determine the minimum number of factors that will account for maximum variance in the data for use in subsequent multivariate analysis. The factors are called *principal components*.

In **common factor analysis**, the factors are estimated based only on the common variance. Communalities are inserted in the diagonal of the correlation matrix. This method is appropriate when the primary concern is to identify the underlying dimensions and the common variance is of interest. This method is also known as *principal axis factoring*.

Other approaches for estimating the common factors are also available. These include the methods of unweighted least squares, generalised least squares, maximum likelihood, alpha method and image factoring. These methods are complex and are not recommended for inexperienced users.[9]

Table 22.3 shows the application of principal components analysis to the toothpaste example.

**Table 22.3 Results of principal components analysis**

Bartlett test of sphericity
Approximate chi-square = 111.314, $df$ = 15, significance = 0.00000
Kaiser-Meyer-Olkin measure of sampling adequacy = 0.660

**Communalities**

| *Variable* | *Initial* | *Extraction* |
|---|---|---|
| $V_1$ | 1.000 | 0.926 |
| $V_2$ | 1.000 | 0.723 |
| $V_3$ | 1.000 | 0.894 |
| $V_4$ | 1.000 | 0.739 |
| $V_5$ | 1.000 | 0.878 |
| $V_6$ | 1.000 | 0.790 |

**Initial eigenvalues**

| *Factor* | *Eigenvalue* | *Percentage of variance* | *Cumulative percentage* |
|---|---|---|---|
| 1 | 2.731 | 45.520 | 45.520 |
| 2 | 2.218 | 36.969 | 82.488 |
| 3 | 0.442 | 7.360 | 89.848 |
| 4 | 0.341 | 5.688 | 95.536 |
| 5 | 0.183 | 3.044 | 98.580 |
| 6 | 0.085 | 1.420 | 100.000 |

**Extraction sums of squared loadings**

| *Factor* | *Eigenvalue* | *Percentage of variance* | *Cumulative percentage* |
|---|---|---|---|
| 1 | 2.731 | 45.520 | 45.520 |
| 2 | 2.218 | 36.969 | 82.488 |

**Table 22.3 Continued**

Factor matrix

| | *Factor 1* | *Factor 2* |
|---|---|---|
| $V_1$ | 0.928 | 0.253 |
| $V_2$ | –0.301 | 0.795 |
| $V_3$ | 0.936 | 0.131 |
| $V_4$ | –0.342 | 0.789 |
| $V_5$ | –0.869 | –0.351 |
| $V_6$ | –0.177 | 0.871 |

Rotation sums of squared loadings

| *Factor* | *Eigenvalue* | *Percentage of variance* | *Cumulative percentage* |
|---|---|---|---|
| 1 | 2.688 | 44.802 | 44.802 |
| 2 | 2.261 | 37.687 | 82.488 |

Rotated factor matrix

| | *Factor 1* | *Factor 2* |
|---|---|---|
| $V_1$ | 0.962 | –0.027 |
| $V_2$ | –0.057 | 0.848 |
| $V_3$ | 0.934 | –0.146 |
| $V_4$ | –0.098 | 0.854 |
| $V_5$ | –0.933 | –0.084 |
| $V_6$ | 0.083 | 0.885 |

Factor score coefficient matrix

| | *Factor 1* | *Factor 2* |
|---|---|---|
| $V_1$ | 0.358 | 0.011 |
| $V_2$ | –0.001 | 0.375 |
| $V_3$ | 0.345 | –0.043 |
| $V_4$ | –0.017 | 0.377 |
| $V_5$ | –0.350 | –0.059 |
| $V_6$ | 0.052 | 0.395 |

Reproduced correlation matrix

| *Variables* | $V_1$ | $V_2$ | $V_3$ | $V_4$ | $V_5$ | $V_6$ |
|---|---|---|---|---|---|---|
| $V_1$ | 0.926* | 0.024 | –0.029 | 0.031 | 0.038 | –0.053 |
| $V_2$ | –0.078 | 0.723* | 0.022 | –0.158 | 0.038 | –0.105 |
| $V_3$ | 0.902 | –0.177 | 0.894* | –0.031 | 0.081 | 0.033 |
| $V_4$ | –0.117 | 0.730 | –0.217 | 0.739* | –0.027 | –0.107 |
| $V_5$ | –0.895 | –0.018 | 0.859 | 0.020 | 0.878* | 0.016 |
| $V_6$ | 0.057 | –0.746 | –0.051 | 0.748 | –0.152 | 0.790* |

The lower left triangle contains the reproduced correlation matrix; the diagonal, the communalities; and the upper right triangle, the residuals between the observed correlations and the reproduced correlations.

Under initial statistics, it can be seen that the communality for each variable, $V_1$ to $V_6$, is 1.0 as unities were inserted in the diagonal of the correlation matrix. The table labelled 'Initial eigenvalues' gives the eigenvalues. The eigenvalues for the factors are, as expected, in decreasing order of magnitude as we go from factor 1 to factor 6. The eigenvalue for a factor indicates the total variance attributed to that factor. The total variance accounted for by all the six factors is 6.00, which is equal to the number of variables. Factor 1 accounts for a variance of 2.731, which is (2.731/6) or 45.52% of the total variance. Likewise, the second factor accounts for (2.218/6) or 36.97% of the total variance, and the first two factors combined account for 82.49% of the total variance. Several considerations are involved in determining the number of factors that should be used in the analysis.

## Determine the number of factors

It is possible to compute as many principal components as there are variables, but in doing so, no parsimony is gained, i.e. we would not have summarised the information nor revealed any underlying structure. To summarise the information contained in the original variables, a smaller number of factors should be extracted. The question is, how many? Several procedures have been suggested for determining the number of factors. These included *a priori* determination and approaches based on eigenvalues, scree plot, percentage of variance accounted for, split-half reliability, and significance tests.

***A priori* determination**. Sometimes, because of prior knowledge, the researcher knows how many factors to expect and thus can specify the number of factors to be extracted beforehand. The extraction of factors ceases when the desired number of factors have been extracted. Most computer programs allow the user to specify the number of factors, allowing for an easy implementation of this approach.

**Determination based on eigenvalues**. In this approach, only factors with eigenvalues greater than 1.0 are retained; the other factors are not included in the model. An eigenvalue represents the amount of variance associated with the factor. Hence, only factors with a variance greater than 1.0 are included. Factors with variance less than 1.0 are no better than a single variable because, due to standardisation, each variable has a variance of 1.0. If the number of variables is less than 20, this approach will result in a conservative number of factors.

**Determination based on scree plot**. A scree plot is a plot of the eigenvalues against the number of factors in order of extraction. The shape of the plot is used to determine the number of factors. Typically, the plot has a distinct break between the steep slope of factors, with large eigenvalues and a gradual trailing off associated with the rest of the factors. This gradual trailing off is referred to as the scree. Experimental evidence indicates that the point at which the scree begins denotes the true number of factors. Generally, the number of factors determined by a scree plot will be one or a few more than that determined by the eigenvalue criterion.

**Determination based on percentage of variance**. In this approach, the number of factors extracted is determined so that the cumulative percentage of variance extracted by the factors reaches a satisfactory level. What level of variance is satisfactory depends upon the problem. It is recommended that the factors extracted should account for at least 60% of the variance.

**Determination based on split-half reliability**. The sample is split in half, and factor analysis is performed on each half. Only factors with high correspondence of factor loadings across the two sub-samples are retained.

**Determination based on significance tests**. It is possible to determine the statistical significance of the separate eigenvalues and retain only those factors that are statistically significant. A drawback is that with large samples (size greater than 200) many factors are likely to be statistically significant, although from a practical viewpoint many of these account for only a small proportion of the total variance.

In Table 22.3, we see that the eigenvalue greater than 1.0 (default option) results in two factors being extracted. Our *a priori* knowledge tells us that toothpaste is bought for two major reasons. The scree plot associated with this analysis is given in Figure 22.2. From the scree plot, a distinct break occurs at three factors. Finally, from the cumulative percentage of variance accounted for, we see that the first two factors account for 82.49% of the variance and that the gain achieved in going to three factors is marginal. Furthermore, split-half reliability also indicates that two factors are appropriate. Thus, two factors appear to be reasonable in this situation.

The second column under the 'Communalities' heading in Table 22.3 gives relevant information after the desired number of factors have been extracted. The communalities for the variances under 'Extraction' are different from those under 'Initial' because all of the variances associated with the variables are not explained unless all the factors are retained. The 'Extraction sums of squared loadings' table gives the variances associated with the factors that are retained. Note that these are the same as those under 'Initial eigenvalues'. This is always the case in principal components analysis. The percentage variance accounted for by a factor is determined by dividing the associated eigenvalue by the total number of factors (or variables) and multiplying by 100. Thus, the first factor accounts for $(2.731/6) \times 100$ or 45.52% of the variance of the six variables. Likewise, the second factor accounts for $(2.218/6) \times 100$ or 36.967% of the variance. Interpretation of the solution is often enhanced by a rotation of the factors.

## Rotate factors

An important output from factor analysis is the factor matrix, also called the *factor pattern matrix*. The factor matrix contains the coefficients used to express the standardised variables in terms of the factors. These coefficients, the factor loadings, represent the correlations between the factors and the variables. A coefficient with a large absolute value indicates that the factor and the variable are closely related. The coefficients of the factor matrix can be used to interpret the factors.

Although the initial or unrotated factor matrix indicates the relationship between the factors and individual variables, it seldom results in factors that can be interpreted, because the factors are correlated with many variables. For example, in Table

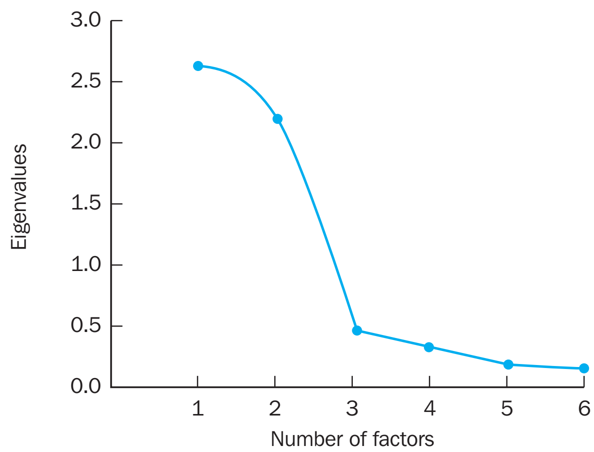


**Figure 22.2**
**Scree plot**

*Rotating factors allows facets of the dataset to be viewed from different perspectives.*

22.3, factor 1 is at least somewhat correlated with five of the six variables (absolute value of factor loading greater than 0.3). How should this factor be interpreted? In such a complex matrix, it is difficult to interpret the factors. Therefore, through rotation, the factor matrix is transformed into a simpler one that is easier to interpret.

In rotating the factors, we would like each factor to have non-zero, or significant, loadings or coefficients for only some of the variables. Likewise, we would like each variable to have non-zero or significant loadings with only a few factors, and if possible with only one. If several factors have high loadings with the same variable, it is difficult to interpret them. Rotation does not affect the communalities and the percentage of total variance explained. The percentage of variance accounted for by each factor does change, however. This is seen in Table 22.3. The variance explained by the individual factors is redistributed by rotation. Hence, different methods of rotation may result in the identification of different factors.

**Orthogonal rotation**
Rotation of factors in which the axes are maintained at right angles.

**Varimax procedure**
An orthogonal method of factor rotation that minimises the number of variables with high loadings on a factor, thereby enhancing the interpretability of the factors.

**Oblique rotation**
Rotation of factors when the axes are not maintained at right angles.

The rotation is called **orthogonal rotation** if the axes are maintained at right angles. The most commonly used method for rotation is the **varimax procedure**. This is an orthogonal method of rotation that minimises the number of variables with high loadings on a factor, thereby enhancing the interpretability of the factors.[10] Orthogonal rotation results in factors that are uncorrelated. The rotation is called **oblique rotation** when the axes are not maintained at right angles, and the factors are correlated. Sometimes, allowing for correlations among factors can simplify the factor pattern matrix. Oblique rotation should be used when factors in the population are likely to be strongly correlated.

In Table 22.3, by comparing the varimax rotated factor matrix with the unrotated matrix (entitled factor matrix), we can see how rotation achieves simplicity and enhances interpretability. Whereas five variables correlated with factor 1 in the unrotated matrix, only variables $V_1$, $V_3$ and $V_5$ correlate highly with factor 1 after rotation. The remaining variables – $V_2$, $V_4$ and $V_6$ – correlate highly with factor 2. Furthermore, no variable correlates highly with both the factors. The rotated factor matrix forms the basis for interpretation of the factors.

## Interpret factors

Interpretation is facilitated by identifying the variables that have large loadings on the same factor. That factor can then be interpreted in terms of the variables that load high on it. Another useful aid in interpretation is to plot the variables, using the factor loadings as coordinates. Variables at the end of an axis are those that have high loadings on only that factor and hence describe the factor. Variables near the origin have small loadings on both the factors. Variables that are not near any of the axes are related to both the factors. If a factor cannot be clearly defined in terms of the original variables, it should be labelled as an undefined or a general factor.

In the rotated factor matrix of Table 22.3, factor 1 has high coefficients for variables $V_1$ (prevention of cavities) and $V_3$ (strong gums), and a negative coefficient for $V_5$ (prevention of tooth decay is not important). Therefore, this factor may be labelled a health benefit factor. Note that a negative coefficient for a negative variable ($V_5$) leads to a positive interpretation that prevention of tooth decay is important. Factor 2 is highly related with variables $V_2$ (shiny teeth), $V_4$ (fresh breath) and $V_6$ (attractive teeth). Thus factor 2 may be labelled a social benefit factor. A plot of the factor loadings, given in Figure 22.3, confirms this interpretation. Variables $V_1$, $V_3$, and $V_5$ (denoted 1, 3, and 5, respectively) are at the end of the horizontal axis (factor 1), with $V_5$ at the end opposite to $V_1$ and $V_3$, whereas variables $V_2$, $V_4$ and $V_6$ (denoted 2, 4 and 6) are at the end of the vertical axis (factor 2). One could summarise the data by stating that consumers appear to seek two major kinds of benefits from a toothpaste: health benefits and social benefits.

## Calculate factor scores

**Factor scores**
Composite scores estimated for each respondent on the derived factors.

Following interpretation, **factor scores** can be calculated, if necessary. Factor analysis has its own stand-alone value. If the goal of factor analysis is to reduce the original set of variables to a smaller set of composite variables (factors) for use in subsequent multivariate analysis, however, it is useful to compute factor scores for each respondent. A factor is simply a linear combination of the original variables. The factor scores for the *i*th factor may be estimated as follows:

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \dots + W_{ik}X_k$$

where the symbols are as defined earlier in the chapter.

The weights or factor score coefficients used to combine the standardised variables are obtained from the factor score coefficient matrix. Most computer programs allow you to request factor scores. Only in the case of principal components analysis is it possible to compute exact factor scores. Moreover, in principal components analysis, these scores are uncorrelated. In common factor analysis, estimates of these scores are obtained, and there is no guarantee that the factors will be uncorrelated with each other. Factor scores can be used instead of the original variables in subsequent multi-

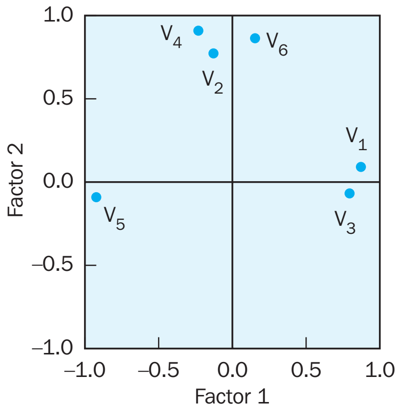


**Figure 22.3**
**Factor loading plot**

variate analysis. For example, using the factor score coefficient matrix in Table 22.3, one could compute two factor scores for each respondent. The standardised variable values would be multiplied by the corresponding factor score coefficients to obtain the factor scores.

## Select surrogate variables

**Surrogate variables**
A subset of original variables selected for use in subsequent analysis.

Sometimes, instead of computing factor scores, the researcher wishes to select surrogate variables. Selection of substitute or **surrogate variables** involves singling out some of the original variables for use in subsequent analysis. This allows the researcher to conduct subsequent analysis and to interpret the results in terms of original variables rather than factor scores. By examining the factor matrix, one could select for each factor the variables rather than factor scores. By examining the factor matrix, one could select for each factor the variable with the highest loading on that factor. That variable could then be used as a surrogate variable for the associated factor. This process works well if one factor loading for a variable is clearly higher than all other factor loadings. The choice is not as easy, however, if two or more variables have similarly high loadings. In such a case, the choice between these variables should be based on theoretical and measurement considerations. For example, theory may suggest that a variable with a slightly lower loading is more important than one with a slightly higher loading. Likewise, if a variable has a slightly lower loading but has been measured more precisely, it should be selected as the surrogate variable. In Table 22.3, the variables $V_1$, $V_3$ and $V_5$ all have high loadings on factor 1, and all are fairly close in magnitude, although $V_1$ has relatively the highest loading and would therefore be a likely candidate. However, if prior knowledge suggests that prevention of tooth decay is a very important benefit, $V_5$ would be selected as the surrogate for factor 1. Also, the choice of a surrogate for factor 2 is not straightforward. Variables $V_2$, $V_4$ and $V_6$ all have comparable high loadings on this factor. If prior knowledge suggests that attractive teeth are the most important social benefit sought from a toothpaste, the researcher would select $V_6$.

## Determine the model fit

The final step in factor analysis involves the determination of model fit. A basic assumption underlying factor analysis is that the observed correlation between variables can be attributed to common factors. Hence, the correlations between the variables can be deduced or reproduced from the estimated correlations between the variables and the factors. The differences between the observed correlations (as given in the input correlation matrix) and the reproduced correlations (as estimated from the factor matrix) can be examined to determine model fit. These differences are called residuals. If there are many large residuals, the factor model does not provide a good fit to the data and the model should be reconsidered. In Table 22.3, we see that only five residuals are larger than 0.05, indicating an acceptable model fit.

The following example further illustrates principal components factoring in the context of trade promotion.

*example*

### Manufacturing promotion components[11]

The objective of this study was to develop a comprehensive inventory of manufacturer-controlled trade promotion variables and to demonstrate that an association exists between these variables and the retailer's promotion support decision. Retailer or trade support was defined operationally as the trade buyer's attitude towards the promotion.

Factor analysis was performed on the explanatory variables with the primary goal of data reduction. The principal components method, using varimax rotation, reduced the 30 explanatory variables to eight factors having eigenvalues greater than 1.0. For the purpose of interpretation, each factor was composed of variables that loaded 0.40 or higher on that factor. In two instances, where variables loaded 0.40 or above on two factors, each variable was assigned to the factor where it had the highest loading. Only one variable, ease of handling/stocking at retail, did not load at least 0.40 on any factor. In all, the eight factors explained 62% of the total variance. Interpretation of the factor-loading matrix was straightforward. Table 1 lists the factors in the order in which they were extracted.

**Table 1 Factors influencing trade promotional support**

| *Factor* | *Factor interpretation (% variance explained)* | *Loading* | *Variables included in the factor* |
|---|---|---|---|
| *F*1 | Item importance (16.3%) | 0.77 | Item is significant enough to warrant promotion |
| | | 0.75 | Category responds well to promotion |
| | | 0.66 | Closest trade competitor is likely to promote item |
| | | 0.64 | Importance of promoted product category |
| | | 0.59 | Item regular (non-deal) sales volume |
| | | 0.57 | Deal meshes with trade promotional requirements |
| | | | *Buyer's estimate of sales increase on the basis of:* |
| *F*2 | Promotion elasticity (9.3%) | 0.86 | Price reduction and display |
| | | 0.82 | Display only |
| | | 0.80 | Price reduction only |
| | | 0.70 | Price reduction, display and advertising |
| | | | *Manufacturer's brand support in the form of:* |
| *F*3 | Manufacturer brand support (8.2%) | 0.85 | Coupons |
| | | 0.81 | Radio and television advertising |
| | | 0.80 | Newspaper advertising |
| | | 0.75 | Point of purchase promotion (e.g. display) |
| *F*4 | Manufacturer reputation (7.3%) | 0.72 | Manufacturer's overall reputation |
| | | 0.72 | Manufacturer's cooperation in meeting trade's promotional needs |
| | | 0.64 | Manufacturer's cooperation on emergency orders |
| | | 0.55 | Quality of sales presentation |
| | | 0.51 | Manufacturer's overall product quality |
| *F*5 | Promotion wearout (6.4%) | 0.93 | Product category is over-promoted |
| | | 0.93 | Item is over-promoted |
| *F*6 | Sales velocity (5.4%) | –0.81 | Brand market share rank[a] |
| | | 0.69 | Item regular sales volume[a] |
| | | 0.46 | Item regular sales volume |
| *F*7 | Item profitability (4.5%) | 0.79 | Item regular gross margin |
| | | 0.72 | Item regular gross margin[a] |
| | | 0.49 | Reasonableness of deal performance requirements |
| *F*8 | Incentive amount (4.2%) | 0.83 | Absolute amount of deal allowances |
| | | 0.81 | Deal allowances as per cent of regular trade cost[a] |
| | | 0.49 | Absolute amount of deal allowances[a] |

[a] Denotes objectives (archival) measure.

Stepwise discriminant analysis was conducted to determine which, if any, of the eight factors predicted trade support to a statistically significant degree. The factor scores for the eight factors were the explanatory variables. The dependent variable consisted of the retail buyer's overall rating of the deal (rating), which was collapsed into a three-group (low, medium and high) measure of trade support. The results of the discriminant analyses are shown in Table 2. All eight entered the discriminant functions. Goodness-of-fit measures indicated that, as a group, the eight factors discriminated between high, medium and low levels of trade support. Multivariate *F* ratios, indicating the degree of discrimination between each pair of groups, were significant at $p < 0.001$. Correct classification into high, medium and low categories was achieved for 65% of the cases. The order of entry into discriminant analysis was used to determine the relative importance of factors as trade support influencers, as shown in Table 3.

**Table 2 Discriminant analysis results: analysis on rating and performance ($n = 564$)**

| | | *Standardised discriminant coefficients Analysis of rating* | |
|---|---|---|---|
| *Factor* | | *Function 1* | *Function 2* |
| *F*1 | Item importance | 0.861 | –0.253 |
| *F*2 | Promotion elasticity | 0.081 | 0.398 |
| *F*3 | Manufacturer brand support | 0.127 | –0.036 |
| *F*4 | Manufacturer reputation | 0.394 | 0.014 |
| *F*5 | Promotion wearout | –0.207 | 0.380 |
| *F*6 | Sales velocity | 0.033 | –0.665 |
| *F*7 | Item profitability | 0.614 | 0.357 |
| *F*8 | Incentive amount | 0.461 | 0.254 |
| Wilks' $\lambda$ (for each factor) | | All significant at $p < 0.001$ | |
| Multivariate *F* ratios | | All significant at $p < 0.001$ | |
| Percentage of cases correctly classified | | 65% correct ($t = 14.4$, $p < 0.001$) | |

**Table 3 Relative importance of trade support influencers (as indicated by order of entry into the discriminant analysis)**

| *Analysis of rating* | |
|---|---|
| *Order of entry* | *Factor name* |
| 1 | Item importance |
| 2 | Item profitability |
| 3 | Incentive amount |
| 4 | Manufacturer reputation |
| 5 | Promotion wearout |
| 6 | Sales velocity |
| 7 | Promotion elasticity |
| 8 | Manufacturer brand support |

## Applications of common factor analysis

The data of Table 22.1 were analysed using the common factor analysis model. Instead of using unities in the diagonal, the communalities were inserted. The output, shown in Table 22.4, is similar to the output from principal components analysis presented in Table 22.3. Beneath the 'Communalities' heading, below the 'Initial' column, the communalities for the variables are no longer 1.0. Based on the eigenvalue criterion, again two factors are extracted. The variances, after extracting the factors, are different from the initial eigenvalues. The first factor accounts for 42.84% of the variance, whereas the second accounts for 31.13%, in each case a little less than what was observed in principal components analysis.

**Table 22.4 Results of common factor analysis**

Bartlett test of sphericity
Approximate chi-square = 111.314, *df* = 15, significance = 0.00000
Kaiser-Meyer-Olkin measure of sampling adequacy = 0.660

**Communalities**

| *Variable* | *Initial* | *Extraction* |
|---|---|---|
| $V_1$ | 0.859 | 0.928 |
| $V_2$ | 0.480 | 0.562 |
| $V_3$ | 0.814 | 0.836 |
| $V_4$ | 0.543 | 0.600 |
| $V_5$ | 0.763 | 0.789 |
| $V_6$ | 0.587 | 0.723 |

**Initial eigenvalues**

| *Factor* | *Eigenvalue* | *Percentage of variance* | *Cumulative percentage* |
|---|---|---|---|
| 1 | 2.731 | 45.520 | 45.520 |
| 2 | 2.218 | 36.969 | 82.488 |
| 3 | 0.442 | 7.360 | 89.848 |
| 4 | 0.341 | 5.688 | 95.536 |
| 5 | 0.183 | 3.044 | 98.580 |
| 6 | 0.085 | 1.420 | 100.000 |

**Extraction sums of squared loadings**

| *Factor* | *Eigenvalue* | *Percentage of variance* | *Cumulative percentage* |
|---|---|---|---|
| 1 | 2.570 | 42.837 | 42.837 |
| 2 | 1.868 | 31.126 | 73.964 |

**Factor matrix**

| | *Factor 1* | *Factor 2* |
|---|---|---|
| $V_1$ | 0.949 | 0.168 |
| $V_2$ | –0.206 | 0.720 |
| $V_3$ | 0.914 | 0.038 |
| $V_4$ | –0.246 | 0.734 |
| $V_5$ | –0.850 | –0.259 |
| $V_6$ | –0.101 | 0.844 |

**Rotation sums of squared loadings**

| *Factor* | *Eigenvalue* | *Percentage of variance* | *Cumulative percentage* |
|---|---|---|---|
| 1 | 2.541 | 42.343 | 42.343 |
| 2 | 1.897 | 31.621 | 73.964 |

**Table 22.4 Continued**

Rotated factor matrix

| | *Factor 1* | *Factor 2* |
|---|---|---|
| $V_1$ | 0.963 | –0.030 |
| $V_2$ | –0.054 | 0.747 |
| $V_3$ | 0.902 | –0.150 |
| $V_4$ | –0.090 | 0.769 |
| $V_5$ | –0.885 | –0.079 |
| $V_6$ | 0.075 | 0.847 |

Factor score coefficient matrix

| | *Factor 1* | *Factor 2* |
|---|---|---|
| $V_1$ | 0.628 | 0.101 |
| $V_2$ | –0.024 | 0.253 |
| $V_3$ | 0.217 | –0.169 |
| $V_4$ | –0.023 | 0.271 |
| $V_5$ | –0.166 | –0.059 |
| $V_6$ | 0.083 | 0.500 |

Reproduced correlation matrix

| *Variables* | $V_1$ | $V_2$ | $V_3$ | $V_4$ | $V_5$ | $V_6$ |
|---|---|---|---|---|---|---|
| $V_1$ | 0.928* | 0.022 | –0.000 | 0.024 | –0.008 | –0.042 |
| $V_2$ | –0.075 | 0.562* | 0.006 | –0.008 | 0.031 | 0.012 |
| $V_3$ | 0.873 | –0.161 | 0.836* | –0.051 | 0.008 | 0.042 |
| $V_4$ | –0.110 | 0.580 | –0.197 | 0.600* | –0.025 | –0.004 |
| $V_5$ | –0.850 | –0.012 | 0.786 | 0.019 | 0.789* | –0.003 |
| $V_6$ | 0.046 | 0.629 | 0.060 | 0.645 | –0.133 | 0.723* |

The lower left triangle contains the reproduced correlation matrix; the diagonal, the communalities; and the upper right triangle, the residuals between the observed correlations and the reproduced correlations.

The values in the unrotated factor pattern matrix of Table 22.4 are a little different from those in Table 22.3, although the pattern of the coefficients is similar. Sometimes, however, the pattern of loadings for common factor analysis is different from that for principal components analysis, with some variables loading on different factors. The rotated factors matrix has the same pattern as that in Table 22.3, leading to a similar interpretation of the factors.

We end with another application of common factor analysis, in the context of consumer perception of rebates.

*example*

### 'Common' rebate perceptions[12]

Rebates are effective in obtaining new users, brand switching and repeat purchases among current users. A study was undertaken to determine the factors underlying consumer perception of rebates. A set of 24 items measuring consumer perceptions of rebates was constructed. Respondents were asked to express their degree of agreement with these items on five-point Likert scales. The data were collected by a one-stage area telephone survey conducted in the Memphis metropolitan area. A total of 303 usable questionnaires was obtained.

The 24 items measuring perceptions of rebates were analysed using common factor analysis. The initial factor solution did not reveal a simple structure of underlying rebate perceptions. Therefore, items that had low loadings were deleted from the scale, and the factor analysis was performed on the remaining items. This second solution yielded three interpretable factors. The factor loadings and the reliability coefficients are presented in the table below. The three factors contained four, four and three items, respectively. Factor 1 seemed to capture the consumers' perceptions of the efforts and difficulties associated with rebate redemption (efforts). Factor 2 was defined as a representation of consumers' faith in the rebate system (faith). Factor 3 represented consumers' perceptions of the manufacturers' motives for offering rebates (motives). The loadings of items on their respective factor ranged from 0.527 to 0.744. ■

**Factor analysis of perceptions of rebates**

| | *Factor loading* | | |
|---|---|---|---|
| *Scale items*[a] | *Factor 1* | *Factor 2* | *Factor 3* |
| Manufacturers make the rebate process too complicated | 0.194 | **0.671** | –0.127 |
| Postal rebates are not worth the trouble involved | –0.031 | **0.612** | 0.352 |
| It takes too long to receive the rebate cheque from the manufacturer | 0.013 | **0.718** | 0.051 |
| Manufacturers could do more to make rebates easier to use | 0.205 | **0.616** | 0.173 |
| Manufacturers offer rebates because consumers want them[b] | **0.660** | 0.172 | 0.101 |
| Today's manufacturers take real interest in consumer welfare[b] | **0.569** | 0.203 | 0.334 |
| Consumer benefit is usually the primary consideration in rebate offers[b] | **0.660** | 0.002 | 0.318 |
| In general, manufacturers are sincere in their rebate offers to consumers[b] | **0.716** | 0.047 | –0.033 |
| Manufacturers offer rebates to get consumers to buy something they do not really need | 0.099 | 0.156 | **0.744** |
| Manufacturers use rebate offers to induce consumers to buy slow-moving items | 0.090 | 0.027 | **0.702** |
| Rebate offers require you to buy more of a product than you need | 0.230 | 0.066 | **0.527** |
| Eigenvalues | 2.030 | 1.344 | 1.062 |
| Percentage of explained variance | 27.500 | 12.200 | 9.700 |

a The response categories for all items were strongly agree (l), agree (2), neither agree nor disagree (3), disagree (4), strongly disagree (5) and don't know (6). 'Don't know' responses were excluded from data analysis.
b The scores of these items were reversed.

In this example, when the initial factor solution was not interpretable, items which had low loadings were deleted and the factor analysis was performed on the remaining items. If the number of variables is large (greater than 15), principal components analysis and common factor analysis result in similar solutions. Principal components analysis is less prone to misinterpretation, however, and is recommended for the non-expert user.

## Internet and computer applications

### SPSS[13]

The program FACTOR may be used for principal components analysis as well as for common factor analysis. Some other methods of factor analysis are also available and factor scores are available.

### SAS

The program PRINCOMP performs principal components analysis and calculates principal components scores. To perform common factor analysis, the program FACTOR can be used. The FACTOR program also performs principal components analysis.

### Minitab

Factor analysis can be accessed using Multivariate>Factor analysis. Principal components or maximum likelihood can be used to determine the initial factor extraction. If maximum likelihood is used, specify the number of factors to extract. If a number is not specified with a principal component extraction, the program will set it equal to a number of variables in the data set.

### Excel

At the time of writing, factor analysis was not available.

## Summary

Factor analysis is a class of procedures used for reducing and summarising data. Each variable is expressed as a linear combination of the underlying factors. Likewise, the factors themselves can be expressed as linear combinations of the observed variables. The factors are extracted in such a way that the first factor accounts for the highest variance in the data, the second the next highest, and so on. Additionally, it is possible to extract the factors so that the factors are uncorrelated, as in principal components analysis.

In formulating the factor analysis problem, the variables to be included in the analysis should be specified based on past research, theory, and the judgement of the researcher. These variables should be measured on an interval or ratio scale. Factor analysis is based on a matrix of correlation between the variables. The appropriateness of the correlation matrix for factor analysis can be statistically tested.

The two basic approaches to factor analysis are principal components analysis and common factor analysis. In principal components analysis, the total variance in the data is considered. Principal components analysis is recommended when the researcher's primary concern is to determine the minimum number of factors that will account for maximum variance in the data for use in subsequent multivariate analysis. In common factor analysis, the factors are estimated based only on the common variance. This method is appropriate when the primary concern is to identify the underlying dimensions and when the common variance is of interest. This method is also known as principal axis factoring.

The number of factors that should be extracted can be determined *a priori* or based on eigenvalues, scree plots, percentage of variance, split-half reliability or significance tests. Although the initial or unrotated factor matrix indicates the relationships between the factors and individual variables, it seldom results in factors that can be interpreted, because the factors are correlated with many variables. Therefore, rotation is used to transform the factor matrix into a simpler one that is easier to interpret. The most commonly used method of rotation is the varimax procedure, which results in orthogonal factors. If the factors are highly correlated in the population, oblique rotation can be used. The rotated factor matrix forms the basis for interpreting the factors.

Factor scores can be computed for each respondent. Alternatively, surrogate variables may be selected by examining the factor matrix and selecting a variable with the highest or near highest loading for each factor. The differences between the observed correlations and the reproduced correlations, as estimated from the factor matrix, can be examined to determine model fit.

## Questions

1 How is factor analysis different from multiple regression and discriminant analysis?

2 What are the major uses of factor analysis?

3 Describe the factor analysis model.

4 What hypothesis is examined by Bartlett's test of sphericity? For what purpose is this test used?

5 What is meant by the term communality of a variable?

6 Briefly define the following: eigenvalue, factor loadings, factor matrix and factor scores.

7 For what purpose is the Kaiser-Meyer-Olkin measure of sampling adequacy used?

8 What is the major difference between principal components analysis and common factor analysis?

9 Explain how eigenvalues are used to determine the number of factors.

10 What is a scree plot? For what purpose is it used?

11 Why is it useful to rotate the factors? Which is the most common method of rotation?

12 What guidelines are available for interpreting the factors?

13 When is it useful to calculate factor scores?

14 What are surrogate variables? How are they determined?

15 How is the fit of the factor analysis model examined?

## Appendix: Fundamental equations of factor analysis[14]

In the factor analysis model, hypothetical components are derived that account for the linear relationship between observed variables. The factor analysis model requires that the relationships between observed variables be linear and that the variables have non-zero correlations between them. The derived hypothetical components have the following properties:

1 They form a linearly independent set of variables. No hypothetical component is derivable from the other hypothetical components as a linear combination of them.
2 The hypothetical components' variables can be divided into two basic kinds of components: common factors and unique factors. These two components can be distinguished in terms of the patterns of weights in the linear equations that derive

the observed variables from the hypothetical components' variables. A common factor has more than one variable with a non-zero weight or factor loading associated with the factor. A unique factor has only one variable with a non-zero weight associated with the factor. Hence, only one variable depends on a unique factor.

3 Common factors are always assumed to be uncorrelated with the unique factors. Unique factors are also usually assumed to be mutually uncorrelated, but common factors may or may not be correlated with each other.

4 Generally, it is assumed that there are fewer common factors than observed variables. The number of unique factors is usually assumed to be equal to the number of observed variables, however.

The following notations are used.

$\boldsymbol{X}$ = an $n \times 1$ random vector of observed random variables $X_1, X_2, X_3, \ldots, X_n$

It is assumed that

$$E(\boldsymbol{X}) = \mathbf{0}$$
$$E(\boldsymbol{XX'}) = \boldsymbol{R}_{xx}, \text{ a correlation matrix with unities in the main diagonal}$$

$\boldsymbol{F}$ = an $m \times 1$ vector of $m$ common factors $F_1, F_2, \ldots, F_m$

It is assumed that

$$E(\boldsymbol{F}) = \mathbf{0}$$
$$E(\boldsymbol{FF'}) = \boldsymbol{R}_{ff}, \text{ a correlation matrix}$$

$\boldsymbol{U}$ = an $n \times 1$ random vector of the $n$ unique factor variables, $U_1, U_2, \ldots, U_n$

It is assumed that

$$E(\boldsymbol{U}) = \boldsymbol{O}$$
$$E(\boldsymbol{UU'}) = \boldsymbol{I}$$

The unique factors are normalised to have unit variances and are mutually uncorrelated.

$\boldsymbol{A}$ = an $n \times m$ matrix of coefficients called the factor pattern matrix

$\boldsymbol{V}$ = an $n \times n$ diagonal matrix of coefficients for the unique factors

The observed variables, which are the coordinates of X, are weighted combinations of the common factors and the unique factors. The fundamental equation of factor analysis can then be written as

$$\boldsymbol{X} = \boldsymbol{AF} + \boldsymbol{VU}$$

The correlations between variables in terms of the factors may be derived as follows:

$$\begin{aligned}
\boldsymbol{R}_{xx} &= E(\boldsymbol{XX'}) \\
&= E\{(\boldsymbol{AF} + \boldsymbol{VU})(\boldsymbol{AF} + \boldsymbol{VU})'\} \\
&= E\{(\boldsymbol{AF} + \boldsymbol{VU})(\boldsymbol{F'A'} + \boldsymbol{U'V'})\} \\
&= E(\boldsymbol{AFF'A'} + \boldsymbol{AFU'V'} + \boldsymbol{VUF'A'} + \boldsymbol{VUU'V'}) \\
&= \boldsymbol{AR}_{ff}\boldsymbol{A'} + \boldsymbol{AR}_{fu}\boldsymbol{V'} + \boldsymbol{VR}_{uf}\boldsymbol{A'} + \boldsymbol{V}^2
\end{aligned}$$

Given that the common factors are uncorrelated with the unique factors, we have

$$\boldsymbol{R}_{fu} = \boldsymbol{R}_{uf}' = \mathbf{0}$$

Hence,

$$\boldsymbol{R}_{xx} = \boldsymbol{AR}_{ff}\boldsymbol{A'} + \boldsymbol{V}^2$$

Suppose that we subtract the matrix of unique factor variance, $\boldsymbol{V}^2$, from both sides. We then obtain

$$\boldsymbol{R}_{xx} - \boldsymbol{V}^2 = \boldsymbol{AR}_{ff}\boldsymbol{A'}$$

$R_{xx}$ is dependent only on the common factor variables, and the correlations among the variables are related only to the common factors. Let $R_c = R_{xx} - V^2$ be the reduced correlation matrix.

We have already defined the factor pattern matrix $A$. The coefficients of the factor pattern matrix are weights assigned to the common factors when the observed variables are expressed as linear combinations of the common and unique factors. We now define the factor structure matrix. The coefficients of the factor structure matrix are the covariances between the observed variables and the factors. The factor structure matrix is helpful in the interpretation of factors as it shows which variables are similar to a common factor variable. The factor structure matrix, $A_s$, is defined as

$$\begin{aligned} A_s &= E(XF') \\ &= E[(AF + VU)F'] \\ &= AR_{ff} + VR_{uf} \\ &= AR_{ff} \end{aligned}$$

Thus, the factor structure matrix is equivalent to the factor pattern matrix $A$ multiplied by the matrix of covariances among the factors $R_{ff}$. Substituting $A_s$ for $AR_{ff}$, the reduced correlation matrix becomes the product of factor structure and the factor pattern matrix:

$$\begin{aligned} R_c &= AR_{ff}A' \\ &= A_sA' \end{aligned}$$

## Notes

1 Birks, D.F. and Birts, A.N., 'Service quality in domestic cash management banks', in Birks, D.F. (ed.), *Global Cash Management in Europe* (Basingstoke: Macmillan, 1998), 175–205. See also Mels, G., Boshof, C. and Nel, D., 'The dimensions of service quality: the original European perspective revisited', *Service Industries Journal* (January 1997), 173–89.

2 Alt, M., *Exploring Hyperspace* (New York: McGraw Hill, 1990), 74.

3 For a detailed discussion of factor analysis, see Tacq, J., *Multivariate Analysis Techniques in Social Science Research* (Thousand Oaks, CA: Sage, 1996); Dunteman, G.H., *Principal Components Analysis* (Newbury Park, CA: Sage, 1989). For an application, see Aaker, J.L., 'Dimensions of brand personality', *Journal of Marketing Research* 34 (August 1997), 347–56.

4 See, for example, Bo Edvardsson, S., Larsson, G. and Setterlind, S., 'Internal service quality and the psychosocial work environment: an empirical analysis of conceptual interrelatedness', *Services Industries Journal* 17(2) (April 1997), 252–63; Taylor, S., 'Waiting for service: the relationship between delays and evaluations of service', *Journal of Marketing* 58 (April 1994), 56–69.

5 See Gaur, S., 'Adelman and Morris factor analysis of developing countries', *Journal of Policy Modeling* 19(4) (August 1997), 407–15; Lastovicka, J.L. and Thamodaran, K., 'Common factor score estimates in multiple regression problems', *Journal of Marketing Research* 28 (February 1991), 105–12; Dillon, W.R. and Goldstein, M., *Multivariate Analysis: Methods and Applications* (New York: Wiley, 1984), 23–99.

6 For applications of factor analysis, see Ittner, C.D. and Larcker, D.F., 'Product development cycle time and organizational performance', *Journal of Marketing Research* 34 (February 1997), 13–23; Ganesan, S., 'Negotiation strategies and the nature of channel relationships', *Journal of Marketing Research* 30 (May 1993), 183–203.

7 Basilevsky, A., *Statistical Factor Analysis and Related Methods: Theory and Applications* (New York: Wiley, 1994); Hair Jr, J.E., Anderson, R.E., Tatham, R.L. and Black, W.C., *Multivariate Data Analysis with Readings*, 5th edn (Englewood Cliffs, NJ: Prentice Hall, 1999), 364–419.

8 Factor analysis is influenced by the relative size of the correlations rather than the absolute size.

9 See Roberts, J.A. and Beacon, D.R., 'Exploring the subtle relationships between environmental concern and ecologically conscious behavior', *Journal of Business Research* 40(1) (September 1997), 79–89; Chatterjee, S., Jamieson, L. and Wiseman, F., 'Identifying most influential observations in factor analysis', *Marketing Science* (Spring 1991), 145–60; Acito, F. and Anderson, R.D., 'A Monte Carlo comparison of factor analytic methods', *Journal of Marketing Research* 17 (May 1980), 228–36.

10 Other methods of orthogonal rotation are also available. The quartimax method minimises the number of factors needed to explain a variable. The equimax method is a combination of varimax and quartimax.

11 Zemanek Jr, J.E., 'Manufacturer influence versus manufacturer salesperson influence over the industrial distributor', *Industrial Marketing Management* 26(1) (January 1997), 59–66; Curhan, R.C. and Kopp, R.J., 'Obtaining retailer support for trade deals: key success factors', *Journal of Advertising Research* (December 1987–January 1988), 51–60.

12 Bulkeley, W.M., 'Rebates' secret appeal to manufacturers: few consumers actually redeem them', *Wall Street Journal* (10 February 1998), B1–B2; Lichtenstein, D.R., Ridgway, N.M. and Netemeyer, R.G., 'Price perceptions and consumer shopping

behaviour: a field study', *Journal of Marketing Research* 30(2) (May 1993), 234–45; Tat, P., Cunningham III, W.A. and Babakus, E., 'Consumer perceptions of rebates', *Journal of Advertising Research* (August–September 1988), 45–50.

13 Einspruch, E.L., *An Introductory Guide to SPSS for Windows* (Thousand Oaks, CA: Sage, 1998); Spector, P.E., *SAS Programming for Researchers and Social Scientists*, (Thousand Oaks, CA: Sage, 1993); Norat, M.A., Software reviews, *Economic Journal: The Journal of the Royal Economic Society* 107 (May 1997), 857–82; Seiter, C., 'The statistical difference', *Macworld* 10(10) (October 1993), 116–21.

14 The material in this Appendix 22A has been drawn from Stanley A. Muliak, *The Foundations of Factor Analysis* (New York: McGraw-Hill, 1972).