Everitt, Brian; Dunn, Graham
Exploratory Factor Analysis (Chapter 12) / from Applied Multivariate Data Analysis
Chichester, UK: Wiley, 2001. 2nd ed. (2012 printing) pp. 271-290.

Applied Multivariate Data Analysis

Second Edition

Brian S. Everitt
Institute of Psychiatry, King's College London, UK

and

Graham Dunn
School of Epidemiology and Health Sciences,
University of Manchester, UK

John Wiley & Sons, Ltd

First published in Great Britain in 2001 by Arnold
This impression printed by Hodder Education,
a part of Hachette Livre UK,
338 Euston Road, London NW1 3BH

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN 978-0-4707-1117-0

8 9 10

Typeset in 10/12pt Times by Academic & Technical Typesetting. Bristol

12

Exploratory factor analysis

12.1 Introduction

In many areas of psychology, sociology and the like it is often not possible to measure directly the concepts of primary interest. Two obvious examples are intelligence and social class. In such cases the researcher is forced to examine the concepts *indirectly* by collecting information on variables which can be measured and which are considered *indicators* of the concepts of real interest. The psychologist interested in assessing an individual's 'intelligence', for example, may record examination scores in a variety of different subjects. Again, a sociologist concerned with social class might pose questions about a person's occupation, educational background, whether or not they own their own house, and so on.

The concepts the researchers would like to measure are generally termed *latent variables*, and the method of analysis most often used to help uncover the relationships between the assumed underlying latent variables and the measured or *manifest* indicator variables is some form of *factor analysis*. The basic factor analysis model is essentially the same as that of multiple regression discussed in previous chapters, except that now the manifest variables are regressed on the unobservable latent variables, also often referred to as the *factor variables* or simply as *factors*.

Factor analysis can be used in at least two ways:

- as an exploratory technique to investigate the relationships between manifest variables and factors without making any prior assumptions about which manifest variables are related to which factors;
- as a technique for testing a specific factor structure in which particular manifest variables relate to particular factors.

In this chapter we shall concern ourselves with only *exploratory factor analysis*; discussion of the alternative approach, *confirmatory factor analysis*, will be left until the next chapter.

12.2 The basic factor analysis model

The basis of factor analysis is a regression model linking the manifest variables to a set of unobserved (and unobservable) latent variables. Details of the model are given in Box 12.1.

One problem that needs to be addressed before estimation of the parameters in the model can be considered is that, as formulated, the factor loading matrix $\mathbf{\Lambda}$ is not uniquely determined by the equations given in Box 12.1 that define the model. That this is so can be seen by introducing an orthogonal matrix $\mathbf{M}$ of order $k \times k$, and rewriting the basic regression equation linking the observed and latent variables as

$$\mathbf{x} = (\mathbf{\Lambda M})(\mathbf{M}'\mathbf{f}) + \mathbf{u}. \tag{12.1}$$

Box 12.1 The factor analysis model

- We assume that we have a set of observed or manifest variables, $\mathbf{x}' = [x_1, x_2, \ldots, x_p]$, assumed to be linked to a smaller number of unobserved latent variables, $f_1, f_2, \ldots, f_p$, where $k < p$, by a regression model of the form

$$x_1 = \lambda_{11}f_1 + \lambda_{12}f_2 + \ldots + \lambda_{1k}f_k + u_1,$$
$$x_2 = \lambda_{21}f_1 + \lambda_{22}f_2 + \ldots + \lambda_{2k}f_k + u_2,$$
$$\vdots$$
$$x_p = \lambda_{p1}f_1 + \lambda_{p2}f_2 + \ldots + \lambda_{pk}f_k + u_k.$$

- These equations may be written more concisely as

$$\mathbf{x} = \mathbf{\Lambda f} + \mathbf{u},$$

where

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & \vdots & \vdots \\ \lambda_{p1} & \cdots & \lambda_{pk} \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_k \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}.$$

(We have assumed that the mean vector of $\mathbf{x}$ is the null vector; this is of no practical consequence since we are only interested in the covariance or correlational structure of the variables.)

- We assume that the 'residual' terms $u_1, \ldots, u_p$ are uncorrelated with each other and with the factors $f_1, \ldots, f_k$. This implies that, given the values of the factors, the manifest variables are independent, that is, the correlations of the observed variables arise from their relationships with the factors. In factor analysis the regression coefficients in $\mathbf{\Lambda}$ are more usually known as *factor loadings*.
- Since the factors are unobserved we can fix their location and scale arbitrarily. We shall assume they occur in standardized form with mean zero and standard deviation one. We shall also assume, initially at least,

that the factors are uncorrelated with one another. With these assumptions the factor model given above implies that the variance of variable x_i, σ_i^2, is given by

$$\sigma_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2 + \psi_i,$$

where ψ_i is the variance of u_i.

- So the factor analysis model implies that the variance of each observed variable can be split into two parts. The first, h_i^2, given by

$$h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2,$$

is known as the *communality* of the variable and represents the variance shared with the other variables via the common factors. The second part, ψ_i, is called the *specific* or *unique* variance and relates to the variability in x_i not shared with other variables.

- In addition, the factor model leads to the following expression for the covariance of variables x_i and x_j:

$$\sigma_{ij} = \sum_{l=1}^{k} \lambda_{il}\lambda_{jl}.$$

- Summarizing, the factor analysis model implies that the covariance of two observed variables depends solely on their relationship with the common factors as measured by the sizes of the relevant factor loadings. (When the factors are assumed to be uncorrelated or *othogonal*, the factor loadings are simply the correlations between factors and observed variables.) The residual terms u_i, often known in factor analysis parlance as *specific variates*, play no part in determining the covariances of the observed variables.
- So the factor analysis model implies the following expression for the covariance matrix, Σ, of the observed variables:

$$\Sigma = \Lambda\Lambda' + \Psi,$$

where

$$\Psi = \mathrm{diag}(\psi_i).$$

- The converse also holds: if Σ can be decomposed into the form given above, then the k-factor model holds for $\mathbf{x}$.
- In practice, Σ will be estimated by the sample covariance matrix $\mathbf{S}$ (alternatively, the model will be applied to the correlation matrix $\mathbf{R}$), and we will need to obtain estimates of Λ and Ψ so that the observed covariance matrix takes the form required by the model.
- We will also need to determine the appropriate value of k, the number of factors.

This satisfies all the requirements of a k-factor model as outlined in Box 12.1 with new factors $\mathbf{f}^* = \mathbf{M}'\mathbf{f}$ and new factor loadings $\mathbf{\Lambda M}$. Consequently, (12.1) implies that

$$\mathbf{\Sigma} = (\mathbf{\Lambda M})(\mathbf{\Lambda M})' + \mathbf{\Psi}, \tag{12.2}$$

which reduces to the original $\mathbf{\Sigma} = \mathbf{\Lambda\Lambda}'$ since $\mathbf{MM}' = \mathbf{I}$. This implies that the factors $\mathbf{f}$ with loadings $\mathbf{\Lambda}$ and the factors $\mathbf{f}^*$ with loadings $\mathbf{\Lambda M}$, for any orthogonal matrix $\mathbf{M}$, are equivalent for explaining the covariance matrix of the observed variables. Consequently, to ensure a unique solution it becomes necessary to introduce some constraints on the parameters in the original model. In general, this is achieved by requiring the matrix $\mathbf{G}$, given by

$$\mathbf{G} = \mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}, \tag{12.3}$$

to be diagonal, with its elements arranged in descending order of magnitude. Such constraints mean that the first factor has maximal contribution to the common variance of the observed variables, the second has maximal contribution to this variance subject to being uncorrelated with the first, and so on. (Compare principal components analysis in Chapter 3.)

Although the constraints implied by (12.3) are necessary to achieve a unique solution for the factor analysis model, they are, essentially, arbitrary and it may be that a more interpretable solution can be achieved using the transformed model with loadings $\mathbf{\Lambda}^* = \mathbf{\Lambda M}$. Such a process is generally known as *factor rotation* and will be discussed in detail in Section 12.4.

12.3 Estimating the parameters in the factor analysis model

The estimation problem in factor analysis is essentially that of finding $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Psi}}$ satisfying the constraints (12.3) and for which

$$\mathbf{S} \approx \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Psi}}. \tag{12.4}$$

The type of solution possible depends on the difference between the number of independent elements of $\mathbf{S}$ and the number of free parameters in the factor analysis model – the former value is simply $\frac{1}{2}p(p+1)$ and the latter $p + pk - \frac{1}{2}k(k-1)$, a value which arises from counting the p specific variances and the pk factor loadings and subtracting the $\frac{1}{2}k(k-1)$ constraints imposed by (12.3). The difference between the two quantities, s, is therefore

$$\begin{aligned} s &= \tfrac{1}{2}p(p+1) - [p + pk - \tfrac{1}{2}k(k-1)] \\ &= \tfrac{1}{2}[(p-k)^2 - (p+k)] \end{aligned} \tag{12.5}$$

Three cases need to be considered:

- $s < 0$. Here there are fewer equations than free parameters, and so an infinite number of exact solutions are possible. The factor analysis model is of no practical interest in this case.
- $s = 0$. The factor analysis model contains as many parameters as elements of $\mathbf{S}$, and the model does not offer a parsimonious description of the relationships

between the observed variables. A unique solution may be found (see below) but not necessarily one where all the specific variances are greater than zero. Again, this case is of no practical importance.

- $s > 0$. This is the only situation of genuine practical interest. Here there are fewer parameters in the factor analysis model than there are independent elements in S. The model provides a more parsimonious explanation of the relationships among the manifest variables than is provided by the elements of S. An exact solution is now, of course, not possible and some form of estimation procedure is necessary (see Section 12.3.1).

But before looking at estimation in practice, it may be helpful to consider a simple example, originally discussed by Spearman (1904). In this example, $s = 0$ so an exact solution is possible.

Spearman considered a sample of children's examination marks in three subjects, Classics (x_1), French (x_2) and English (x_3), from which he calculated the following correlation matrix

$$\mathbf{R} = \begin{array}{l} \\ \text{Classics} \\ \text{French} \\ \text{English} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Classics} & \text{French} & \text{English} \end{array} \\ \begin{pmatrix} 1.00 & & \\ 0.83 & 1.00 & \\ 0.78 & 0.67 & 1.00 \end{pmatrix} \end{array}.$$

If we assume a single factor, then the appropriate factor analysis model is

$$\begin{aligned} x_1 &= \lambda_1 f + u_1, \\ x_2 &= \lambda_2 f + u_2, \\ x_3 &= \lambda_3 f + u_3. \end{aligned} \tag{12.6}$$

In this example the common factor, f, might be equated with intelligence or general intellectual ability, and the specific variates, u_1, u_2, u_3 will have small variances if their associated observed variable is closely related to f. The value of s in (12.5) is zero, and by equating elements of the observed correlation matrix to the corresponding values predicted by the single-factor model we will be able to find estimates of λ_1, λ_2, λ_3, ψ_1, ψ_2 and ψ_3 such that the model fits exactly. The six equations are:

$$\begin{aligned} \hat{\lambda}_1\lambda_2 &= 0.83, \\ \hat{\lambda}_1\lambda_3 &= 0.78, \\ \hat{\lambda}_1\lambda_3 &= 0.67, \\ \hat{\psi}_1 &= 1.0 - \hat{\lambda}_1^2, \\ \hat{\psi}_2 &= 1.0 - \hat{\lambda}_2^2, \\ \hat{\psi}_3 &= 1.0 - \hat{\lambda}_3^2. \end{aligned} \tag{12.7}$$

The solutions of these equations are

$$\hat{\lambda}_1 = 0.99, \qquad \hat{\lambda}_2 = 0.84, \qquad \hat{\lambda}_3 = 0.79,$$
$$\hat{\psi}_1 = 0.02, \qquad \hat{\psi}_2 = 0.30, \qquad \hat{\psi}_3 = 0.38.$$

Suppose now that the observed correlations had been

$$\mathbf{R} = \begin{array}{l} \\ \text{Classics} \\ \text{French} \\ \text{English} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Classics} & \text{French} & \text{English} \end{array} \\ \begin{pmatrix} 1.00 & & \\ 0.84 & 1.00 & \\ 0.60 & 0.35 & 1.00 \end{pmatrix} \end{array}$$

In this case the solution for the parameters of a single-factor model are

$$\hat{\lambda}_1 = 1.2, \qquad \hat{\lambda}_2 = 0.7, \qquad \hat{\lambda}_3 = 0.5,$$
$$\hat{\psi}_1 = -0.44, \qquad \hat{\psi}_2 = 0.51, \qquad \hat{\psi}_3 = 0.75.$$

Clearly this solution is unacceptable because of the negative estimate for the first specific variance.

In those cases of practical interest where $s > 0$ there are two main estimation techniques used, *principal factor analysis* and *maximum likelihood factor analysis*.

12.3.1 Principal factor analysis

Principal factor analysis is an eigenvalue and eigenvector technique similar in many respects to principal components analysis (see Chapter 3). But rather than using the sample covariance matrix, the procedure uses what is known as the *reduced covariance matrix*, $\mathbf{S}^*$, given by

$$\mathbf{S}^* = \mathbf{S} - \hat{\mathbf{\Psi}}, \tag{12.8}$$

where $\hat{\mathbf{\Psi}}$ is a diagonal matrix containing estimates of the ψ_i. The diagonal elements of $\mathbf{S}^*$ are thus estimated communalities – the parts of the variance of $\mathbf{x}$ which are explicable by the common factors. Two frequently used estimates of the communalities are:

- the square of the multiple correlation coefficient of x_i with the other observed variables;
- the largest of the absolute values of the correlation coefficients between x_i and another variable.

Each of these will give higher communality values when x_i is highly correlated with the other variables, which is what is required.

Having found initial communality estimates, a principal components analysis is performed on $\mathbf{S}^*$ and the first k components used to provide estimates of the loadings in the k-factor model. Revised estimates of the specific variances are then found from applying

$$\hat{\psi}_i = s_i^2 - \sum_{i=1}^{k} \hat{\lambda}_{ij}^2. \tag{12.9}$$

A principal factor analysis is considered to have provided a satisfactory solution if all estimated specific variances are non-negative. In many cases the investigator will accept this solution, but there is the possibility of now computing a revised $\mathbf{S}^*$ and repeating the process until the estimated factor loadings and specific variances converge to some final value. Difficulties can, however, arise with this iterative process if at any time a communality estimate exceeds the variance of the corresponding manifest variable, resulting in a negative estimate of a specific variance.

Some applications of this estimation procedure will be described later.

12.3.2 Maximum likelihood factor analysis

Maximum likelihood is regarded, by statisticians at least, as perhaps the most respectable method of estimating the parameters in the factor analysis model. It does, however, make the strong assumption that **f** and **u**, and hence **x**, have a multivariate normal distribution. With this assumption we can write down the likelihood function for the k-factor model as

$$L = -\tfrac{1}{2}n\{\ln|\mathbf{\Lambda\Lambda}' + \mathbf{\Psi}| + \mathrm{trace}(\mathbf{S}|\mathbf{\Lambda\Lambda}' + \mathbf{\Psi}|^{-1})\}. \tag{12.10}$$

For a number of reasons it is more convenient to minimize the function F given by

$$F = \ln|\mathbf{\Lambda\Lambda}' + \mathbf{\Psi}| + \mathrm{trace}(\mathbf{S}|\mathbf{\Lambda\Lambda}' + \mathbf{\Psi}|^{-1}) - \ln|\mathbf{S}| - p. \tag{12.11}$$

Minimizing F is equivalent to maximizing L, since L equals $-\tfrac{1}{2}nF$ plus a function of the observations. The function F takes the value zero if $\mathbf{\Lambda\Lambda}' + \mathbf{\Psi}$ is equal to $\mathbf{S}$ and values greater than zero otherwise. Details of how F is minimized to give the required estimates of factor loadings and specific variances are given in Lawley and Maxwell (1971), Mardia *et al.* (1979), and Everitt (1984; 1987). One of the advantages of using a maximum likelihood approach is that it provides a test for number of factors. The test statistic is

$$U = n' \min(F), \tag{12.12}$$

where $n' = n - 1 - \tfrac{1}{6}(2p+5) - \tfrac{2}{3}k$. If k common factors are adequate to account for the observed covariances or correlations of the manifest variables then U has, asymptotically, a chi-squared distribution with ν degrees of freedom, where

$$\nu = \tfrac{1}{2}(p-k)^2 - \tfrac{1}{2}(p+k). \tag{12.13}$$

In most exploratory studies k cannot be specified in advance and so a sequential procedure is used. Starting with some small value for k (usually $k = 1$), the parameters in the corresponding factor analysis model are estimated by maximum likelihood. If U is not significant the current value of k is accepted, otherwise k is increased by one and the process repeated. If at any stage the degrees of freedom of the test became zero then either no non-trivial solution is appropriate or alternatively the factor model itself with its assumption of linearity between observed and latent variables is questionable. Examples of the test are given later.

12.4 Rotation of factors

The constraints on the factor loadings imposed by (12.4) were introduced to make the parameter estimates in the factor analysis model unique; they lead to orthogonal factors that are arranged in descending order of importance. These properties are not, however, inherent in the factor model, and merely considering such a solution may lead to difficulties of interpretation. For example, two consequences of these properties of a factor solution are as follows:

- The factorial complexity of variables is likely to be greater than 1 regardless of the underlying true model; consequently, variables may have substantial loadings on more than one factor.
- Except for the first factor, the remaining factors are often *bipolar*, that is, they have a mixture of positive and negative loadings.

Both of these can make understanding a factor analysis solution somewhat difficult. Interpretation is more straightforward if each variable is highly loaded on at most one factor, and if all factor loadings are either large and positive or near zero, with few intermediate values. The variables are thus split into disjoint sets, each of which is associated with a single factor. This aim is essentially what Thurstone (1931) referred to as *simple structure*. In more detail such structure has the following properties:

- Each row or the factor loading matrix should contain at least one zero.
- Each column of the loading matrix should contain at least k zeros.
- Every pair of columns of the loading matrix should contain several variables whose loadings vanish in one column but not in the other.
- If the number of factors is four or more, every pair of columns should contain a large number of variables with zero loadings in both columns.
- Conversely, for every pair of columns of the loading matrix only a small number of variables should have non-zero loadings in both columns.

When simple structure is achieved the observed variables will fall into mutually exclusive groups whose loadings are high on single factors, perhaps moderate to low on a few factors, and of negligible size on the remaining factors.

The search for simple structure or something close to it begins after an initial factoring has determined the number of common factors necessary and the communalities of each observed variable. The factor loadings are then transformed by post-multiplication by a suitably chosen orthogonal matrix. Such

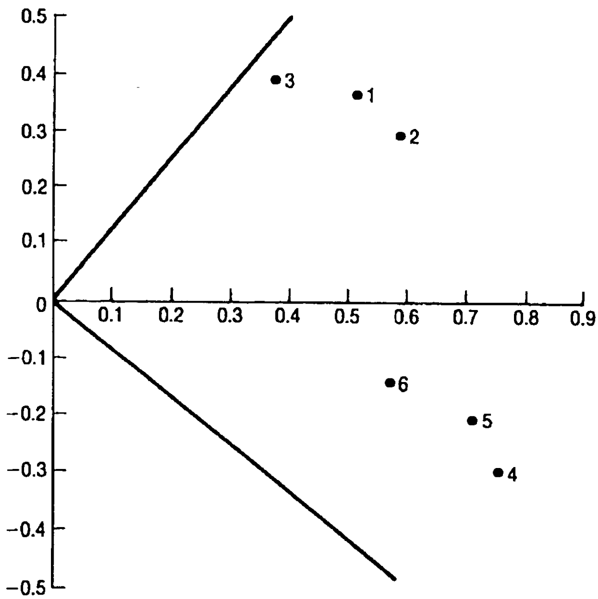

Figure 12.1 Plot of factor loadings for correlation matrix of six school subjects.

a transformation is equivalent to a rigid rotation of the axes of the originally identified factor space. For a two-factor model the process of rotation can be performed graphically. As an example, consider the following correlation matrix for six school subjects:

$$\mathbf{R} = \begin{array}{l} \\ \text{French} \\ \text{English} \\ \text{History} \\ \text{Arithmetic} \\ \text{Algebra} \\ \text{Geometry} \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{pmatrix} 1.00 & & & & & \\ 0.44 & 1.00 & & & & \\ 0.41 & 0.35 & 1.00 & & & \\ 0.29 & 0.35 & 0.16 & 1.00 & & \\ 0.33 & 0.32 & 0.19 & 0.59 & 1.00 & \\ 0.25 & 0.33 & 0.18 & 0.47 & 0.46 & 1.00 \end{pmatrix} \end{array}$$

The initial factor loadings are plotted in Figure 12.1. By referring each variable to the new axes shown, which correspond to a rotation of the original axes through about 40 degrees, a new set of loadings can be obtained which give an improved description of the fitted model. The two sets of loadings are given explicitly in Table 12.1.

When there are more than two factors more formal methods of rotation are needed. One which is very popular is the *varimax* technique originally proposed by Kaiser (1958). This has as its rationale the aim of factors with a few large loadings and as many near-zero loadings as possible. This is achieved by

Table 12.1 Two-factor solution for correlations of six school subjects

	Unrotated loadings		Rotated loadings	
Variable	1	2	1	2
French	0.55	0.43	0.20	0.62
English	0.57	0.29	0.30	0.52
History	0.39	0.45	0.05	0.55
Arithmetic	0.74	−0.27	0.75	0.15
Algebra	0.72	−0.21	0.65	0.18
Geometry	0.59	−0.13	0.50	0.20

iterative maximization of a quadratic function of the loadings – details are given in Mardia *et al.* (1979).

Varimax rotation leads to orthogonal factors, but in some cases a factor solution may become simpler to interpret if the factors are allowed to be correlated rather than independent. Various methods for the *oblique rotation* of factors are available, for example, *oblimin* (see Nie *et al.*, 1975). Because oblique rotations involve the introduction of correlations between the factors, both the factor correlation matrix and the loadings matrix need to be considered in interpreting the solution. An added complication is that the elements of the latter no longer represent correlations between manifest variables and the factors. Perhaps correlated factors become more worth considering if/when the analysis moves from the exploratory to the confirmatory (see Chapter 13).

Factor rotation is often regarded as controversial since it apparently allows the investigator to impose on the data whatever type of solution is required. But this is clearly *not* the case since although the axes may be rotated about their origin, or may be allowed to become oblique, *the distribution of the points will remain invariant.* Rotation is simply a procedure which allows new axes to be chosen so that the positions of the points can be described as simply as possible.

(It should be noted that rotation techniques are also often applied to the results from a principal components analysis in the hope that it will aid their interpretability. Although in some cases this may be acceptable it does have several disadvantages which are listed by Jolliffe, 1989. The main problem is that the defining property of principal components, namely that of accounting for maximal proportions of the total variation in the observed variables, is lost after rotation.)

12.5 Some examples of the application of factor analysis

The first example is taken from Calsyn and Kenny (1977), who collected data on the following six variables for 556 white eighth-grade students: x_1, self-concept of ability; x_2, perceived parental evaluation; x_3, perceived teacher evaluation; x_4, perceived friend's evaluation; x_5, educational aspiration; x_6, college plans.

The correlation matrix of the observed variables was calculated to be

$$\mathbf{R} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{array} \begin{array}{c} \begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{array} \\ \begin{pmatrix} 1.00 & & & & & \\ 0.73 & 1.00 & & & & \\ 0.70 & 0.68 & 1.00 & & & \\ 0.58 & 0.61 & 0.57 & 1.00 & & \\ 0.46 & 0.43 & 0.40 & 0.37 & 1.00 & \\ 0.56 & 0.52 & 0.48 & 0.41 & 0.72 & 1.00 \end{pmatrix} \end{array}$$

The two-factor solutions given by applying both principal factor analysis and maximum likelihood factor analysis to these correlations are shown in Table 12.2. The varimax rotated solution in each case is also shown. Here the first factor loading highly on variables 1–4 might be labelled 'ability', and the second, with high loadings on variables 5 and 6, 'aspiration'. We shall return to this example in Chapter 13.

The data for our second example of the application of factor analysis are from a survey of AIDS patients' reactions to their physicians. The survey was conducted by Van Servellen using a scale developed by Cope *et al.* (1986). The 14 items in the survey questionnaire measure patient attitudes about physician personality, demeanour, competence and prescribed treatment using a Likert-type scale from 1 to 5 for each item. Since seven of the items were stated negatively, they have been recoded (reflected) so that 1 represents the most positive response and 5 the least positive on all items. The items and the names we shall use are as follows:

1. My doctor treats me in a friendly manner.
2. I have some doubts about the ability of my doctor.
3. My doctor seems cold and impersonal.
4. My doctor does his/her best to keep me from worrying.
5. My doctor examines me as carefully as necessary.
6. My doctor should treat me with more respect.
7. I have some doubts about the treatment suggested by my doctor.
8. My doctor seems very competent and well trained.
9. My doctor seems to have a genuine interest in me as a person.
10. My doctor leaves me with many unanswered questions about my condition and its treatment.
11. My doctor uses words that I do not understand.
12. I have a great deal of confidence in my doctor.
13. I feel I can tell my doctor about very personal problems.
14. I do not feel free to ask my doctor questions.

The observed correlation matrix of the 14 variables is given in Table 12.3. Here we shall use only maximum likelihood factor analysis. First we shall try to determine the necessary number of factors by using the test described in Section 12.3.2. The results of this test for the one-, two- and three-factor

Table 12.2 Principal factor analysis and maximum likelihood results for ability and aspirations example

Principal factor analysis

	Unrotated loadings		Communality	Specific variance
x_1	0.83	−0.18	0.72	0.28
x_2	0.82	−0.24	0.73	0.27
x_3	0.77	−0.24	0.65	0.35
x_4	0.67	−0.19	0.49	0.51
x_5	0.67	0.48	0.68	0.32
x_6	0.76	0.45	0.78	0.22
Variance	3.440	0.61		

	Varimax rotated loadings		Communality	Specific variance
x_1	0.78	0.34	0.72	0.28
x_2	0.80	0.29	0.73	0.27
x_3	0.77	0.26	0.65	0.35
x_4	0.66	0.24	0.49	0.51
x_5	0.26	0.78	0.68	0.32
x_6	0.35	0.81	0.78	0.22
Variance	2.46	1.59		

Maximum likelihood factor analysis

	Unrotated loadings		Communality	Specific variance
x_1	0.56	0.65	0.74	0.26
x_2	0.52	0.68	0.73	0.27
x_3	0.48	0.65	0.65	0.35
x_4	0.41	0.57	0.49	0.51
x_5	0.72	0.09	0.56	0.44
x_6	1.00	0.00	1.00	0.00
Variance	2.50	1.63		

	Varimax rotated loadings		Communality	Specific variance
x_1	0.78	0.34	0.74	0.26
x_2	0.80	0.30	0.73	0.27
x_3	0.77	0.27	0.65	0.35
x_4	0.66	0.22	0.49	0.51
x_5	0.30	0.66	0.56	0.44
x_6	0.30	0.96	1.00	0.00
Variance	2.45	1.68		

models are as follows:

Factors	Chi-square	d.f.	*p*-value
1	139.7	77	0.000016
2	101.59	64	0.0019
3	68.96	52	0.058

Table 12.3 Correlation matrix for AIDS data

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.00													
2	0.56	1.00												
3	0.63	0.58	1.00											
4	0.64	0.46	0.35	1.00										
5	0.52	0.44	0.50	0.52	1.00									
6	0.70	0.51	0.49	0.52	0.54	1.00								
7	0.45	0.48	0.28	0.34	0.38	0.63	1.00							
8	0.61	0.68	0.44	0.43	0.56	0.64	0.49	1.00						
9	0.79	0.58	0.66	0.55	0.66	0.64	0.34	0.70	1.00					
10	0.57	0.63	0.40	0.55	0.54	0.58	0.65	0.62	0.62	1.00				
11	0.32	0.27	0.33	0.21	0.13	0.26	0.22	0.24	0.17	0.25	1.00			
12	0.55	0.72	0.51	0.49	0.63	0.62	0.47	0.75	0.70	0.67	0.31	1.00		
13	0.69	0.51	0.60	0.54	0.51	0.73	0.44	0.50	0.66	0.53	0.24	0.65	1.00	
14	0.62	0.42	0.33	0.47	0.38	0.58	0.51	0.49	0.53	0.56	0.23	0.51	0.56	1.00

These results suggest that we should use the three-factor solution. The estimated loadings and the varimax rotated loadings for the three-factor model are shown in Table 12.4. The rotated factors might be labelled 'trust in doctor', 'confidence in doctor's ability' and 'confidence in recommended treatment'.

12.6 Estimating factor scores

In most applications factor analysis stops with the estimation of the parameters in the model, the rotation of the factors and the attempted interpretation of the fitted model. There are occasions, however, when it may be required to find the values or scores of each individual observation on the derived factors. Such scores might be used in a variety of ways, similarly to the scores derived from a principal components analysis. But calculation of factor scores is more problematical than principal components scores. In the original equation defining the factor analysis model in Box 12.1, the variables are expressed in terms of the factors, whereas to calculate scores we require the relationship to be in the opposite direction. But if we make the assumption of normality, the conditional distribution of $\mathbf{f}$ given $\mathbf{x}$ can be found. It is

$$N[\mathbf{\Lambda}'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}), (\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}+\mathbf{I})^{-1}]. \tag{12.14}$$

Consequently, one plausible way of calculating factor scores would be to use the mean of this distribution,

$$\hat{\mathbf{f}} = \hat{\mathbf{\Lambda}}'\mathbf{S}^{-1}(\mathbf{x}-\bar{\mathbf{x}}). \tag{12.15}$$

Other possible methods for deriving factor scores are described in Rencher (1995).

Table 12.4 Three-factor maximum likelihood solution for AIDS data

Unrotated factor loadings

Item	F1	F2	F3	Communality	Specific variance
1	0.45	0.75	−0.38	0.91	0.09
2	0.48	0.56	0.24	0.60	0.40
3	0.28	0.63	−0.09	0.48	0.52
4	0.34	0.56	−0.14	0.45	0.55
5	0.38	0.57	0.14	0.49	0.51
6	0.63	0.52	−0.10	0.68	0.32
7	1.00	0.00	0.00	1.00	0.00
8	0.45	0.62	0.21	0.63	0.37
9	0.34	0.82	−0.04	0.79	0.21
10	0.65	0.45	0.14	0.64	0.36
11	0.22	0.25	0.00	0.11	0.89
12	0.47	0.69	0.45	0.90	0.61
13	0.44	0.64	−0.05	0.61	0.39
14	0.51	0.45	−0.13	0.48	0.52
Variance	3.66	3.59	0.556		

Varimax rotated loadings

Item	F1	F2	F3	Communality	Specific variance
1	0.89	0.25	0.25	0.91	0.09
2	0.36	0.62	0.29	0.60	0.40
3	0.58	0.36	0.10	0.48	0.52
4	0.57	0.30	0.18	0.45	0.55
5	0.41	0.53	0.19	0.49	0.51
6	0.57	0.38	0.47	0.68	0.32
7	0.17	0.23	0.96	1.00	0.00
8	0.42	0.64	0.29	0.63	0.37
9	0.71	0.53	0.11	0.79	0.21
10	0.37	0.52	0.48	0.64	0.36
11	0.23	0.20	0.15	0.11	0.89
12	0.32	0.86	0.23	0.90	0.10
13	0.60	0.44	0.25	0.61	0.39
14	0.51	0.28	0.37	0.48	0.52
Variance	3.68	3.15	0.96		

12.7 Factor analysis with categorical variables

In many areas of research – for example, psychology and psychiatry – the variable values will be on an ordinal rather than an interval scale, and, consequently, will not be ideally suited to analysis by, say, maximum likelihood factor analysis. For such data a number of more specialized and more complex factor analysis procedures have been suggested – see, for example, Muthén (1984) and Muthén and Christoffersson (1981). Details of these techniques are beyond the level of this text, but their central idea is that the values of an observed ordinal variable arise from thresholding an underlying continuous

Table 12.5 Data on female psychiatric patients

Anxiety	Depression	Sleep	Sex	Life
2	2	2	2	2
2	2	2	2	2
3	3	1	2	2
2	2	2	2	2
2	1	1	2	1
1	1	2	1	1
3	2	2	2	1
3	2	2	2	2
2	2	2	2	2
2	1	1	1	1
2	2	2	2	1
3	2	2	2	1
2	2	2	2	2
3	2	2	2	2
1	1	1	2	1
1	2	2	1	1
4	3	2	2	2
3	2	2	2	2
2	2	2	2	1
3	3	1	2	2
3	2	2	2	1
2	1	2	2	1
2	2	2	2	2
2	2	2	2	1
2	2	2	2	1
2	2	2	2	1
2	2	2	2	2
3	3	2	2	2
3	3	2	2	2

Anxiety and Depression are both scored 1 for none, 2 for mild, 3 for moderate, 4 for severe. Sleep is scored 1 for yes and 2 for no, in response to the question 'Can you sleep normally?'. Sex is scored 1 for no and 2 for yes, in response to the question 'Have you lost interest in sex?'. Life is also scored 1 for no and 2 for yes, in response to the question 'Have you thought recently about ending your life?'.

variable; that is, a categorical variable, y_j, with C ordered categories, arises from an underlying continuous variable, y_j^* as follows:

$$y_{ij} = c \qquad \text{if } \tau_{j,c} < y_{ij}^* \leq \tau_{j,c+1}, \tag{12.16}$$

for categories $c = 0, 1, \ldots, C-1$, and $\tau_0 = -\infty, \tau_C = \infty$.

As a simple illustration of this approach, we shall apply it to the data shown in Table 12.5 which concern a number of categorical variables observed on 29 female psychiatric patients. The results are shown in Table 12.6; for comparison, Table 12.7 shows the corresponding results from a maximum likelihood factor analysis. We see first that there are considerable differences in the two correlation matrices (that in Table 12.6 involves estimated polychoric correlations), and some considerable differences in the estimated factor loadings

Table 12.6 Results from factor analysis procedure appropriate for categorical variables on data in Table 12.5

Correlation matrix

	Anxiety	Depression	Sleep	Sex	Life
Anxiety	1.00				
Depression	0.87	1.00			
Sleep	0.12	0.20	1.00		
Sex	0.88	0.74	0.31	1.00	
Life	0.61	0.97	0.20	0.95	1.00

One- and two-factor solutions: varimax rotation

	One factor	Two factor	
	F1	F1	F2
Anxiety	0.83	0.84	0.04
Depression	0.94	0.94	0.08
Sleep	0.23	0.11	0.99
Sex	0.95	0.93	0.21
Life	0.94	0.93	0.11

of the two solutions, although the interpretation of the factors in each case would be similar. The example illustrates that careful consideration needs to be given to the question of applying maximum likelihood factor analysis to multivariate data involving categorical variables.

Table 12.7 Results from maximum likelihood factor analysis on data in Table 12.5

Correlation matrix

	Anxiety	Depression	Sleep	Sex	Life
Anxiety	1.00				
Depression	0.67	1.00			
Sleep	0.07	0.16	1.00		
Sex	0.47	0.39	0.14	1.00	
Life	0.43	0.59	0.11	0.35	1.00

One- and two-factor solutions: varimax rotation

	One factor	Two factor	
	F1	F1	F2
Anxiety	0.76	0.77	0.01
Depression	0.88	0.87	0.01
Sleep	0.16	0.01	0.99
Sex	0.50	0.50	0.10
Life	0.64	0.64	0.01

12.8 Factor analysis and principal components analysis compared

Factor analysis, like principal components analysis, is an attempt to explain a set of multivariate data using a smaller number of dimensions than one begins with, but the procedures used to achieve this goal are essentially quite different in the two approaches. Some differences between the two are as follows:

- Factor analysis postulates a model for the data – principal components analysis does not.
- Factor analysis tries to explain the covariances or correlations of the observed variables by means of a few common factors. Principal components analysis is primarily concerned with explaining the variance of the observed variables.
- If the number of retained components is increased, say, from m to $m + 1$, the first m components are unchanged. This is not the case in factor analysis, where there can be substantial changes in *all* factors if the number of factors is changed.
- The calculation of principal component scores is straightforward – the calculation of factor scores is more complex, and a variety of methods have been suggested.
- There is usually no relationship between the principal components of the sample correlation matrix and the sample covariance matrix. For maximum likelihood factor analysis, however, the results of analysing either matrix are essentially equivalent.

Despite these differences, the results from both types of analysis are frequently very similar. Certainly if the specific variances are small we would expect both forms of analysis to give similar results. However, if the specific variances are large they will be absorbed into all the principal components, both retained and rejected, whereas factor analysis makes special provision for them.

Lastly, it should be remembered that both principal components analysis and factor analysis are similar in one important respect – they are both pointless if the observed variables are almost uncorrelated. In this case factor analysis has nothing to explain and principal components analysis will simply lead to components which are similar to the original variables.

12.9 Summary

Factor analysis has probably attracted more critical comment than any other statistical technique. Because the factor loadings are not uniquely determined by the basic factor model many statisticians have complained that by rotating factors investigators can arrive at the answer they are looking for. Indeed, Blackith and Reyment (1971) suggest that the method has persisted *precisely* because it does enable users to impose their preconceived ideas of the structure of the observed correlations. Other criticisms have centred on whether the concept of underlying, unobservable variables is an acceptable one or not. In psychology, postulating latent variables to explain the correlations between manifest variables may be reasonable – in other areas it may not be so acceptable.

Hills (1977) has gone so far as to suggest that factor analysis is not worth the time necessary to understand it and carry it out. And Chatfield and Collins (1980) recommend that factor analysis should not be used in most practical situations. Such criticisms go too far. Factor analysis is simply an additional, and at times very useful, tool for investigating particular features of the structure of multivariate observations. Of course, like many models used in analysing data, the one used in factor analysis is likely to be only a very idealized approximation to the truth in the situations in which it is generally applied. Such an approximation may, however, prove a valuable starting point for further investigations.

Exercises

12.1 The correlation matrix given below arises from the scores of 220 boys in six school subjects: (1) French, (2) English, (3) History, (4) Arithmetic, (5) Algebra, (6) Geometry. The two-factor solution from a maximum likelihood factor analysis is shown in Table 12.8. By plotting the derived loadings, find an orthogonal rotation which allows easier interpretation of the results.

$$\mathbf{R} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & \;\;2 & \;\;3 & \;\;4 & \;\;5 & \;\;6 \end{array} \\ \begin{pmatrix} 1.00 & & & & & \\ 0.44 & 1.00 & & & & \\ 0.41 & 0.35 & 1.00 & & & \\ 0.29 & 0.35 & 0.16 & 1.00 & & \\ 0.33 & 0.32 & 0.19 & 0.60 & 1.00 & \\ 0.25 & 0.33 & 0.18 & 0.47 & 0.46 & 1.00 \end{pmatrix} \end{array}.$$

Table 12.8 Maximum likelihood factor analysis for school subjects data

Subject	Factor loadings F1	F2	Communality
1. French	0.55	0.43	0.49
2. English	0.57	0.29	0.41
3. History	0.39	0.45	0.36
4. Arithmetic	0.74	−0.27	0.62
5. Algebra	0.72	−0.21	0.57
6. Geometry	0.60	−0.13	0.37

Table 12.9 Maximum likelihood factor analysis for Exercise 12.3

F1	F2
0.789	−0.403
0.834	−0.234
0.740	−0.034
0.586	−0.185
0.676	−0.248
0.654	0.140
0.641	0.234
0.629	0.351
0.564	0.054
0.808	0.414

12.2 Find the exact one-factor solution for the following correlation matrix:

$$\mathbf{R} = \begin{pmatrix} 1 & & \\ \frac{1}{3} & 1 & \\ \frac{1}{3} & \frac{1}{10} & 1 \end{pmatrix}.$$

Is the solution acceptable?

12.3 A two-factor maximum likelihood analysis solution for a particular set of data yields the estimated loadings shown in Table 12.9. By making a plot of these loadings find the angle corresponding to the minimum rotation of the axes needed to remove those that are negative. Find the corresponding rotation matrix and the rotated loadings.

12.4 The rotation matrix for a varimax rotation of the factors in Exercise 12.3 is

$$\mathbf{H} = \begin{pmatrix} 0.7278 & -0.6858 \\ 0.6858 & 0.7278 \end{pmatrix}.$$

To what angle of rotation does this correspond?

12.5 The matrix below shows the correlations between ratings on nine statements about pain made by 123 people suffering from extreme pain. Each statement was scored on a scale from 1 to 6 ranging from agreement to disagreement. The nine pain statements were as follows:

1. Whether or not I am in pain in the future depends on the skill of the doctors.
2. Whenever I am in pain, it is usually because of something I have done or not done.
3. Whether or not I am in pain depends on what doctors do for me.
4. I cannot get any help for my pain unless I go to seek medical advice.
5. When I am in pain I know that it is because I have not been taking proper exercise or eating the right food.
6. People's pain results from their own carelessness.
7. I am directly responsible for my pain.

8. Relief from pain is chiefly controlled by the doctors.
9. People who are never in pain are just plain lucky.

$$
\mathbf{R} = \begin{pmatrix}
1.00 & & & & & & & & \\
-0.04 & 1.00 & & & & & & & \\
0.61 & -0.07 & 1.00 & & & & & & \\
0.45 & -0.12 & 0.59 & 1.00 & & & & & \\
0.03 & 0.49 & 0.03 & -0.08 & 1.00 & & & & \\
-0.29 & 0.43 & -0.13 & -0.21 & 0.47 & 1.00 & & & \\
-0.30 & 0.30 & -0.24 & -0.19 & 0.41 & 0.63 & 1.00 & & \\
0.45 & -0.31 & 0.59 & 0.63 & -0.14 & -0.13 & -0.26 & 1.00 & \\
0.30 & -0.17 & 0.32 & 0.37 & -0.24 & -0.15 & -0.29 & 0.40 & 1.00
\end{pmatrix}
$$

(a) Perform a principal components analysis on these data and examine the associated scree plot to decide on the appropriate number of components.

(b) Apply maximum likelihood factor analysis and use the test described in the chapter to select the necessary number of common factors.

(c) Rotate the factor solution selected using both an orthogonal and an oblique procedure, and interpret the results.

Logistic model for drugs:

Parameter	Estimate	Standard error
Const	−3.484	0.141
Sex(2)	0.697	0.093
Age(2)	0.683	0.158
Age(3)	1.173	0.145
Age(4)	1.565	0.158
Age(5)	1.580	0.185

The effect of Sex (coded 1 for male, 2 for female) is more marked for drugs than the GHQ. There appears to be a slight drop in the probability of being GHQ positive as one gets older; the effect of Age on drug prescription, on the other hand, is to increase the probability of being prescribed drugs with increasing age. These interactions are confirmed in bivariate logistic regression models (results not shown).

Chapter 11

11.2 The rule becomes: allocate observations to f_1 if

$$\pi_1 f_1(x) > \pi_2 f_2(x).$$

By substituting the correct forms for f_1 and f_2 this becomes

$$x \log\frac{p_1(1-p_2)}{p_2(1-p_1)} > \log\frac{\pi_1}{\pi_2} - n\log\frac{1-p_1}{1-p_2}.$$

Chapter 12

12.2 A single-factor model for the three variables is

$$x_1 = \lambda_1 f + u_1, \qquad x_2 = \lambda_2 f + u_2, \qquad x_3 = \lambda_3 f + u_3.$$

By equating elements of the observed correlation matrix to the corresponding values predicted by the single-factor model we will be able to find estimates of the parameters in the model. The six equations are:

$$\hat\lambda_1\hat\lambda_2 = 0.667, \qquad \hat\lambda_1\hat\lambda_3 = 0.667, \qquad \hat\lambda_2\hat\lambda_3 = 0.100,$$

$$\hat\psi_1 = 1.0 - \hat\lambda_1^2, \qquad \hat\psi_2 = 1.0 - \hat\lambda_2^2, \qquad \hat\psi_3 = 1.0 - \hat\lambda_3^2.$$

The solutions of these equations are

$$\hat\lambda_1 = 2.11, \qquad \hat\lambda_2 = 0.316, \qquad \hat\lambda_3 = 0.316,$$

$$\hat\psi_1 = -3.45, \qquad \hat\psi_2 = 0.9, \qquad \hat\psi_3 = 0.9.$$

Clearly the solution is unacceptable.