Chapter 5: Exploratory Factor Analysis

Contents

1. Introduction

In many areas of psychology, and other disciplines in the behavioral sciences, often it is 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 that can be measured or observed directly and can also realistically be assumed to be indicators, in some sense, of the concepts of real interest. The psychologist who is interested in an individual's “intelligence”, for example, may record examination scores in a variety of different subjects in the expectation that these scores are dependent in some way on what is widely regarded as “intelligence” but are also subject to random errors. And a sociologist, say, concerned with people's “social class” might pose questions about a person's occupation, educational background, home ownership, etc., on the assumption that these do reflect the concept he or she is really interested in.

Both “intelligence” and “social class” are what are generally referred to as *latent variables*-i.e., concepts that cannot be measured directly but can be assumed to relate to a number of measurable or *manifest variables*.

The method of analysis most generally used to help uncover the relationships between the assumed latent variables and the manifest variables is *factor analysis*.

The model on which the method is based is essentially that of multiple regression, except now the manifest variables are regressed on the unobservable latent variables (often referred to in this context as *common factors*), so that direct estimation of the corresponding regression coefficients (*factor loadings*) is not possible.

A point to be made at the outset is that factor analysis comes in two distinct varieties:

1. ***Exploratory factor analysis***, which is used to investigate the relationship between manifest variables and factors without making any assumptions about which manifest variables are related to which factors.
2. ***Confirmatory factor analysis**,* which is used to test whether a specific factor model postulated a priori provides an adequate fit for the covariances or correlations between the manifest variables.

In this chapter, we shall consider only exploratory factor analysis.

2. A simple example of a factor analysis model

Consider a sample of children's examination marks in three subjects, Classics (x_1), French (x_2), and English (x_3), from which the following correlation matrix for a sample of children can be calculated:

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

If we assume a single factor, then the *single-factor model* is specified as follows:

$x_1 = \lambda_1 f + u_1,$

$x_2 = \lambda_2 f + u_2,$
$x_3 = \lambda_3 f + u_3. \qquad (5.1)$

We see that the model essentially involves the simple linear regression of each observed variable on the single common factor. In this example, the underlying latent variable or common factor, f, might possibly be equated with intelligence or general intellectual ability. The terms λ_1, λ_2, and λ_3 which are essentially regression coefficients are, in this context, known as factor loadings, and the terms u_1, u_2, and u_3 represent random disturbance terms and will have small variances if their associated observed variable is closely related to the underlying latent variable. The variation in u_i actually consists of two parts, the extent to which an individual's ability at Classics, say, differs from his or her general ability and the extent to which the examination in Classics is only an approximate measure of his or her ability in the subject. In practice no attempt is made to disentangle these two parts.

3. The *k*-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. In essence the model assumes that the observed relationships between the manifest variables (as measured by their covariances or correlations) are a result of the relationships of these variables to the latent variables. (Since it is the covariances or correlations of the manifest variables that are central to factor analysis, we can, in the description of the mathematics of the method given below, assume that the manifest variables all have zero mean.)

To begin, we assume that we have a set of observed or manifest variables, $\boldsymbol{x}' = (x_1, \dots, x_q)$, assumed to be linked to k unobserved latent variables or common factors $f_1, \dots, f_k$, where $k < q$, by a regression model of the form

$$\begin{gathered} x_1 = \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{1k} f_k + u_1 \\ x_2 = \lambda_{21} f_1 + \lambda_{22} f_2 + \cdots + \lambda_{2k} f_k + u_2 \\ \vdots \\ x_q = \lambda_{q1} f_1 + \lambda_{q2} f_2 + \cdots + \lambda_{qk} f_k + u_q \end{gathered}$$

In the context of factor analysis, the regression coefficients λ_js are known as the factor loadings.

The regression equations above may be written more concisely as

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

where

$$\Lambda = \begin{pmatrix} \lambda_{11} & \dots & \lambda_{1k} \\ \vdots & & \vdots \\ \lambda_{q1} & \dots & \lambda_{qk} \end{pmatrix}, \quad f = \begin{pmatrix} f_1 \\ \vdots \\ f_k \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_q \end{pmatrix}.$$

We assume that the random disturbance terms $u_1, \dots, u_q$ are uncorrelated with each other and with the factors $f_1, \dots, f_k$. (The elements of $\boldsymbol{u}$ are specific to each x_i and hence are generally better known in this context as specific variates.) The two assumptions imply that, given the values of the common factors, the manifest variables are independent; that is, the correlations of the observed variables arise from their relationships with the common factors.

Because the factors are unobserved, we can fix their locations and scales arbitrarily and we shall assume they occur in standardized form with mean zero and standard deviation one. We will also assume, initially at least, that the factors are uncorrelated with one another, in which case the factor loadings are the correlations of the manifest variables and the factors. With these additional assumptions about the factors, the factor analysis model 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. Consequently, we see that 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 :

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

We see that the covariances are not dependent on the specific variates in any way; it is the common factors only that aim to account for the relationships between the manifest variables.

The results above show that the k-factor analysis model implies that the population covariance matrix, $\boldsymbol{\Sigma}$, of the observed variables has the form

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi},$$

where

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

The converse also holds: if $\boldsymbol{\Sigma}$ can be decomposed into the form given above, then the k-factor model holds for $\boldsymbol{x}$. In practice, $\boldsymbol{\Sigma}$ will be estimated by the sample covariance matrix S and we will need to obtain estimates of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ so that the observed covariance matrix takes the form required by the model. We will also need to determine the value of k, the number of factors, so that the model provides an adequate fit for $\boldsymbol{S}$.

4. Scale invariance of the k-factor model

Rescaling the x variables is equivalent to letting $\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x}$, where $C = diag\ (c_i)$ and the $c_i; i = 1, \dots, q$ are the scaling values. If the k-factor model holds for x with $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_x$ and $\boldsymbol{\Psi} = \boldsymbol{\Psi}_x$, then

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{\Lambda}_x\boldsymbol{f} + \boldsymbol{C}\boldsymbol{u}$$

and the covariance matrix of $\boldsymbol{y}$ implied by the factor analysis model for $\boldsymbol{x}$ is

$Var(\boldsymbol{y}) = \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C} = \boldsymbol{C}\boldsymbol{\Lambda}_x\boldsymbol{\Lambda}_x{}'\boldsymbol{C} + \boldsymbol{C}\boldsymbol{\Psi}_x\boldsymbol{C}$.

So we see that the k-factor model also holds for $\boldsymbol{y}$ with factor loading matrix $\boldsymbol{\Lambda}_y = \boldsymbol{C}\boldsymbol{\Lambda}_x$ and specific variances $\boldsymbol{\Psi}_y = \boldsymbol{C}\boldsymbol{\Psi}_x\boldsymbol{C} = c_i^2\psi_i$. So the factor loading matrix for the scaled variables y is found by scaling the factor loading matrix of the original variables by multiplying the ith row of Λ_x by c_i and similarly for the specific variances. Thus factor analysis is essentially unaffected by the rescaling of the variables. In particular, if the rescaling factors are such that $c_i = \frac{1}{s_i}$, where s_i is the standard deviation of the x_i, then the rescaling is equivalent to applying the factor analysis model to the correlation matrix of the x variables and the factor loadings and specific variances that result can be found simply by scaling the corresponding loadings and variances obtained from the covariance matrix. Consequently, the factor analysis model can be applied to either the covariance matrix or the correlation matrix because the results are essentially equivalent.

5. Estimating the parameters in the *k*-factor analysis model

To use the above factor analysis model to a sample, we need to estimate factor loadings and variances. Essentially find $\widehat{\boldsymbol{\Lambda}}$ (the estimated factor loading matrix) and $\widehat{\boldsymbol{\Psi}}$ (the diagonal matrix containing the estimated specific variances), which, reproduce as accurately as possible the sample covariance matrix, $\boldsymbol{S}$. This implies

$$\boldsymbol{S} \approx \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Lambda}}^{'} + \widehat{\boldsymbol{\Psi}}$$

Given an estimate of the factor loading matrix, $\widehat{\boldsymbol{\Lambda}}$, it is clearly sensible to estimate the specific variances as

$$\widehat{\psi_i} = s_i^2 - \sum\nolimits_{j=1}^{k} \hat{\lambda}_{ij}^2, \qquad i = 1, \dots, q$$

so that the diagonal terms in $\boldsymbol{S}$ are estimated exactly.

Consider the simple single-factor model (1.5), the number of parameters in the model, 6 (three factor loadings and three specific variances), is equal to the number of independent elements in $\boldsymbol{R}$ (the three correlations and the three diagonal standardized variances), and so 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 derived from the matrix equality implied by the factor analysis model,

$$\boldsymbol{R} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} (\lambda_1 \quad \lambda_2 \quad \lambda_3) + \begin{pmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & \psi_3 \end{pmatrix}$$

are

$$\begin{cases} \hat{\lambda}_1\lambda_2 = 0.83 \\ \hat{\lambda}_1\lambda_3 = 0.78 \\ \hat{\lambda}_2\lambda_3 = 0.67 \end{cases} \qquad \begin{cases} \psi_1 = 1.0 - \hat{\lambda}_1^2 \\ \psi_2 = 1.0 - \hat{\lambda}_2^2 \\ \psi_3 = 1.0 - \hat{\lambda}_3^2 \end{cases}$$

The solutions of these equations are

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

Suppose now that the observed correlations had been

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

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

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

Clearly this solution is unacceptable because of the negative estimate for the first specific variance. In the simple example considered above, the factor analysis model does not give a useful description of the data because the number of parameters in the model equals the number of independent elements in the correlation matrix.

In practice, where the k-factor model has fewer parameters than there are independent elements of the covariance or correlation matrix, the fitted model represents a genuinely parsimonious description of the data and methods of estimation are needed that try to make the covariance matrix predicted by the factor model as close as possible in some sense to the observed covariance matrix of the manifest variables. There are two main methods of estimation:
1. Principal factor analysis
2. Maximum likelihood factor analysis

Principal factor analysis

Principal factor analysis is an eigenvalue and eigenvector technique similar in many respects to principal components analysis but operating not directly on $\boldsymbol{S}$ (or $\boldsymbol{R}$) but on what is known as the *reduced covariance matrix* , $\boldsymbol{S}^*$, defined as

$$\boldsymbol{S}^* = \boldsymbol{S} - \widehat{\boldsymbol{\Psi}}$$

where $\widehat{\boldsymbol{\Psi}}$ is a diagonal matrix containing estimates of the ψ_i. The "ones" on the diagonal of $\boldsymbol{S}$ have in $\boldsymbol{S}^*$ been replaced by the estimated communalities, $\sum_{j=1}^{k} \hat{\lambda}_{ij}^2$, the parts of the variance of each observed variable that can be explained by the common factors. Unlike principal components analysis, factor analysis does not try to account for *all* the observed variance, only that shared through the common factors. Of more concern in factor analysis is accounting for the covariances or correlations between the manifest variables.

To calculate $\boldsymbol{S}^*$ (or with $\boldsymbol{R}$ replacing $\boldsymbol{S}$, $\boldsymbol{R}^*$) we need values for the communalities. Clearly we cannot calculate them on the basis of factor loadings because these loadings still have to be estimated. We need to find a sensible way of finding initial values for the communalities that does not depend on knowing the factor loadings. When the factor analysis is based on the correlation matrix of the manifest variables, two frequently used methods are:

- Take the communality of a variable x_i as the square of the multiple correlation coefficient of x_i with the other observed variables.
- Take the communality of x_i as the largest of the absolute values of the correlation coefficients between x_i and one of the other variables.

Each of these possibilities will lead to higher values for the initial communality when x_i is highly correlated with at least some of the other manifest variables, which is essentially what is required.

Given the initial communality values, a principal components analysis is performed on $\boldsymbol{S}^*$ and the first k eigenvectors used to provide the estimates of the loadings in the k-factor model. The estimation process can stop here or the loadings obtained at this stage can provide revised communality estimates calculated as $\sum_{j=1}^{k} \hat{\lambda}_{ij}^2$, where the $\hat{\lambda}_{ij}^2$s are the loadings estimated in the previous step. The procedure is then repeated until some convergence criterion is satisfied. Difficulties can sometimes arise with this iterative approach if at any time a communality estimate exceeds the variance of the corresponding manifest variable, resulting in a negative estimate of the variable's specific variance. Such a result is known as a *Heywood case* and is clearly unacceptable since we cannot have a negative specific variance.

Maximum likelihood factor analysis

Under the assumption that the data being analyzed have a multivariate normal distribution, and assuming the factor analysis model holds, the logarithm of likelihood function L can be shown to be $-\frac{1}{2}nF$ plus a function of the observations where F is given by

$$F = \ln|\boldsymbol{\Lambda\Lambda}' + \boldsymbol{\Psi}| + trace(\boldsymbol{S}|\boldsymbol{\Lambda\Lambda}' + \boldsymbol{\Psi}|^{-1}) - \ln|\boldsymbol{S}| - q$$

The function F takes the value zero if $\boldsymbol{\Lambda\Lambda}' + \boldsymbol{\Psi}$ is equal to $\boldsymbol{S}$ and values greater than zero otherwise. Estimates of the loadings and the specific variances are found by minimizing F with respect to these parameters. A number of iterative numerical algorithms have been suggested. Initial values of the factor loadings and specific variances can be found in a number of ways, including that described in the above. As with iterated principal factor analysis, the maximum likelihood approach can also experience difficulties with Heywood cases.

6. Estimating the number of factors

An advantage of the maximum likelihood approach is that it has an associated formal hypothesis testing procedure that provides a test of the hypothesis H_k that k common factors are sufficient to describe the data against the alternative that the population covariance matrix of the data has no constraints. The test statistic is

$$U = N\min(F),$$

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

$$\nu = \frac{1}{2}(q-k)^2 - \frac{1}{2}(q+k).$$

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 using maximum likelihood. If U is not significant, the current value of k is accepted; otherwise k is increased by one and the process is repeated. If at any stage the degrees of freedom of the test become 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.

7. Factor rotation

In factor analysis model there is no unique solution for the factor loading matrix. We can see that this is so by introducing an orthogonal matrix $\boldsymbol{M}$ of order $k \times k$ and rewriting the basic regression equation linking the observed and latent variables as

$$\boldsymbol{x} = (\boldsymbol{\Lambda M})(\boldsymbol{M'f}) + \boldsymbol{u}.$$

This "new" model satisfies all the requirements of a k-factor model as previously outlined with new factors $\boldsymbol{f}^* = \boldsymbol{Mf}$ and the new factor loadings $\boldsymbol{\Lambda M}$. This model implies that the covariance matrix of the observed variables is

$$\Sigma = (\boldsymbol{\Lambda M})(\boldsymbol{\Lambda M})' + \Psi,$$

which, since $\boldsymbol{MM}' = \boldsymbol{I}$, reduces to $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Lambda}' + \boldsymbol{\Psi}$ as before. Consequently, factors $\boldsymbol{f}$ with loadings $\boldsymbol{\Lambda}$ and factors $\boldsymbol{f}^*$ with loadings $\boldsymbol{\Lambda M}$ are, for any orthogonal matrix $\boldsymbol{M}$, equivalent for explaining the covariance matrix of the observed variables. Essentially then there are an infinite number of solutions to the factor analysis model as previously formulated.

The problem is generally solved by introducing some constraints in the original model. One possibility is to require the matrix $\boldsymbol{G}$ given by

$$\boldsymbol{G} = \boldsymbol{\Lambda\Psi}^{-1}\boldsymbol{\Lambda}$$

to be diagonal, with its elements arranged in descending order of magnitude.

Such a requirement sets the first factor to have maximal contribution to the common variance of the observed variables, and the second has maximal contribution to this variance subject to being uncorrelated with the first and so on (cf. principal components analysis in Chapter 3). The constraint above ensures that $\boldsymbol{\Lambda}$ is uniquely determined, except for a possible change of sign of the columns. (When $k = 1$, the constraint is irrelevant.) The constraints on the factor loadings imposed by a condition such as that given above need to be introduced to make the parameter estimates in the factor analysis model unique, and 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 a factor solution found when applying the constraint above are:

- The factorial complexity of variables is likely to be greater than one 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 ; i.e., they have a mixture of positive and negative loadings.

It may be that a more interpretable orthogonal solution can be achieved using the equivalent model with loadings $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda M}$ for some particular orthogonal matrix, $\boldsymbol{M}$. Such a process is generally known as factor rotation, and we need to choose $\boldsymbol{M}$ (i.e., how to "rotate" the factors). Factor rotation can be a useful procedure for simplifying an exploratory factor analysis. Factor rotation merely allows the fitted factor analysis model to be described as simply as possible; rotation does not alter the overall structure of a solution but only how the solution is described. Rotation is a process by which a solution is made more interpretable without changing its underlying mathematical properties. Initial factor solutions with variables loading on several factors and with bipolar factors can be difficult to interpret. Interpretation is more straightforward if each variable is highly loaded on at most one factor and if all factor loadings are either large or 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 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. Medium-sized, equivocal loadings are to be avoided.

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 a transformation is equivalent to a rigid rotation of the axes of the originally identified factor space. And during the rotation phase of the analysis, we might choose to abandon one of the assumptions made previously, namely that factors are orthogonal, i.e., independent (the condition was assumed initially simply for convenience in describing the factor analysis model). Consequently, two types of rotation are possible:

- **orthogonal rotation**, in which methods restrict the rotated factors to being uncorrelated, or
- **oblique rotation**, where methods allow correlated factors.

As we have seen above, orthogonal rotation is achieved by post-multiplying the original matrix of loadings by an orthogonal matrix. For oblique rotation, the original loadings matrix is post-multiplied by a matrix that is no longer constrained to be orthogonal. With an orthogonal rotation, the matrix of correlations between factors after rotation is the identity matrix. With an oblique rotation, the corresponding matrix of correlations is restricted to have unit elements on its diagonal, but there are no restrictions on the off-diagonal elements.

So the first question that needs to be considered when rotating factors is whether we should use an orthogonal or an oblique rotation. There is no universal answer to this question.

There are advantages and disadvantages to using either type of rotation procedure. As a general rule, if a researcher is primarily concerned with getting results that “best fit” his or her data, then the factors should be rotated obliquely. If, on the other hand, the researcher is more interested in the generalizability of his or her results, then orthogonal rotation is probably to be preferred. One major advantage of an orthogonal rotation is simplicity since the loadings represent correlations between factors and manifest variables. This is not the case with an oblique rotation because of the correlations between the factors. Here there are two parts of the solution to consider;

- **factor pattern coefficients**, which are regression coefficients that multiply with factors to produce measured variables according to the common factor model, and
- **factor structure coefficients**, correlation coefficients between manifest variables and the factors.

Additionally there is a matrix of factor correlations to consider. In many cases where these correlations are relatively small, researchers may prefer to return to an orthogonal solution. There are a variety of rotation techniques, although only relatively few are in general use. For orthogonal rotation, the two most commonly used techniques are known as varimax and quartimax.

- **Varimax rotation**, 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 iterative maximization of a quadratic function of the loadings. It produces factors that have high correlations with one small set of variables and little or no correlation with other sets. There is a tendency for any general factor to disappear because the factor variance is redistributed.
- **Quartimax rotation**, forces a given variable to correlate highly on one factor and either not at all or very low on other factors. It is far less popular than varimax.

For oblique rotation, the two methods most often used are oblimin and promax.

- **Oblimin rotation**, attempts to find simple structure with regard to the factor pattern matrix through a parameter that is used to control the degree of correlation between the factors. Fixing a value for this parameter is not straightforward, but it is suggested that values between about -0.5 and 0.5 are sensible for many applications.
- **Promax rotation**, operates by raising the loadings in an orthogonal solution (generally a varimax rotation) to some power. The goal is to obtain a solution that provides the best structure using the lowest possible power loadings and the lowest correlation between the factors.

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 that allows new axes to be chosen so that the positions of the points can be described as simply as possible.

8. Estimating factor scores

The first stage of an exploratory factor analysis consists of the estimation of the parameters in the model and the rotation of the factors, followed by an (often heroic) attempt to interpret the fitted model. The second stage is concerned with estimating latent variable scores for each individual in the data set; such factor scores are often useful for a number of reasons:

1. They represent a parsimonious summary of the original data possibly useful in subsequent analyses (cf. principal component scores in Chapter 3).
2. They are likely to be more reliable than the observed variable values.
3. The factor score is a “pure” measure of a latent variable, while an observed value may be ambiguous because we do not know what combination of latent variables may be represented by that observed value.

Making the assumption of normality, the conditional distribution of $\boldsymbol{f}$ given $\boldsymbol{x}$ can be found. It is

$$N(\mathbf{\Lambda}'\mathbf{\Sigma}^{-1}\mathbf{x}, (\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}+\mathbf{I})^{-1}).$$

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

$$\hat{\boldsymbol{f}} = \hat{\mathbf{\Lambda}}'\mathbf{S}^{-1}\mathbf{x},$$

where the vector of scores for an individual, $\boldsymbol{x}$, is assumed to have mean zero; i.e., sample means for each variable have already been subtracted. In many respects, the most damaging problem with factor analysis is not the rotational indeterminacy of the loadings but the indeterminacy of the factor scores.

9. Examples

In this section we consider to examples of exploratory factor analysis.

Expectations of life

The data in Table 5.1 show life expectancy in years by country, age, and sex. The data relate to life expectancies in the 1960s for different countries by age and gender. (It is not here). R code for producing Table 5.1

```
life =  matrix(c(63, 51, 30, 13, 67, 54, 34, 15, 34, 29, 13, 5, 38, 32, 17, 6, 38, 30, 17, 7, 38, 34, 20, 7,
                 59, 42, 20, 6, 64, 46, 25, 8, 56, 38, 18, 7, 62, 46, 25, 10, 62, 44, 24, 7, 69, 50, 28, 14,
                 50, 39, 20, 7, 55, 43, 23, 8, 65, 44, 22, 7, 72, 50, 27, 9, 56, 46, 24, 11, 63, 54, 33, 19,
                 69, 47, 24, 8, 75, 53, 29, 10, 65, 48, 26, 9, 68, 50, 27, 10, 64, 50, 28, 11, 66, 51, 29, 11,
                 56, 44, 25, 10, 61, 48, 27, 12, 60, 44, 22, 6, 65, 45, 25, 9, 61, 45, 22, 8, 65, 49, 27, 10,
                 49, 40, 22, 9, 51, 41, 23, 8, 59, 42, 22, 6, 61, 43, 22, 7, 63, 44, 23, 8, 67, 48, 26, 9,
                 59, 44, 24, 8, 63, 46, 25, 8, 65, 48, 28, 14, 68, 51, 29, 13, 65, 48, 26, 9, 67, 49, 27, 10,
                 64, 63, 21, 7, 68, 47, 25, 9, 64, 43, 21, 6, 68, 47, 24, 8, 67, 45, 23, 8, 74, 51, 28, 10,
                 61, 40, 21, 10, 67, 46, 25, 11, 68, 46, 23, 8, 75, 52, 29, 10, 67, 45, 23, 8, 74, 51, 28, 10,
                 65, 46, 24, 9, 71, 51, 28, 10, 59, 43, 23, 10, 66, 49, 27, 12, 58, 44, 24, 9, 62, 47, 25, 10,
                 57, 46, 28, 9, 60, 49, 28, 11)
               ,31, byrow=TRUE)
colnames(life) = c('m0', 'm25', 'm50', 'm75', 'w0', 'w25', 'w50', 'w75')
rownames(life) = c('Algeria', 'Cameroon', 'Madagascar', 'Mauritius', 'Reunion', 'Seychelles',
                   'South Africa (C)', 'South Africa (W)', 'Tunisia', 'Canada', 'Costa Rica',
                   'Dominican Rep.', 'El Salvador', 'Greenland', 'Grenada', 'Guatemala', 'Honduras',
                   'Jamaica', 'Mexico', 'Nicaragua', 'Panama', 'Trinidad (62)', 'Trinidad (67)',
                   'United States (66)', 'United States (NW66)', 'United States (W66)',
                   'United States (67)', 'Argentina', 'Chile', 'Colombia', 'Ecuador')
```

To begin, we will use the formal test for the number of factors incorporated into the maximum likelihood approach. We can apply this test to the data, assumed to be contained in the data frame life using the following R code:

```
> sapply(1:3, function(f)  factanal(life, factors = f, method ="mle")$PVAL)
```

```
  objective    objective    objective
1.879555e-24 1.911514e-05 4.578204e-01
```

These results suggest that a three-factor solution might be adequate to account for the observed covariances in the data, although it has to be remembered that, with only 31 countries, use of an asymptotic test result may be rather suspect. The three-factor solution is as follows (note that the solution is that resulting from a varimax solution. the default for the factanal() function):

```
> factanal(life, factors = 3, method ="mle")

Call:
factanal(x = life, factors = 3, method = "mle")
Uniquenesses:
   m0   m25   m50   m75    w0   w25   w50   w75
0.005 0.362 0.066 0.288 0.005 0.011 0.020 0.146

Loadings:
    Factor1 Factor2 Factor3
m0  0.964   0.122   0.226
m25 0.646   0.169   0.438
m50 0.430   0.354   0.790
m75         0.525   0.656
w0  0.970   0.217
w25 0.764   0.556   0.310
w50 0.536   0.729   0.401
w75 0.156   0.867   0.280

               Factor1 Factor2 Factor3
SS loadings      3.375   2.082   1.640
Proportion Var   0.422   0.260   0.205
Cumulative Var   0.422   0.682   0.887

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 6.73 on 7 degrees of freedom.
The p-value is 0.458
```

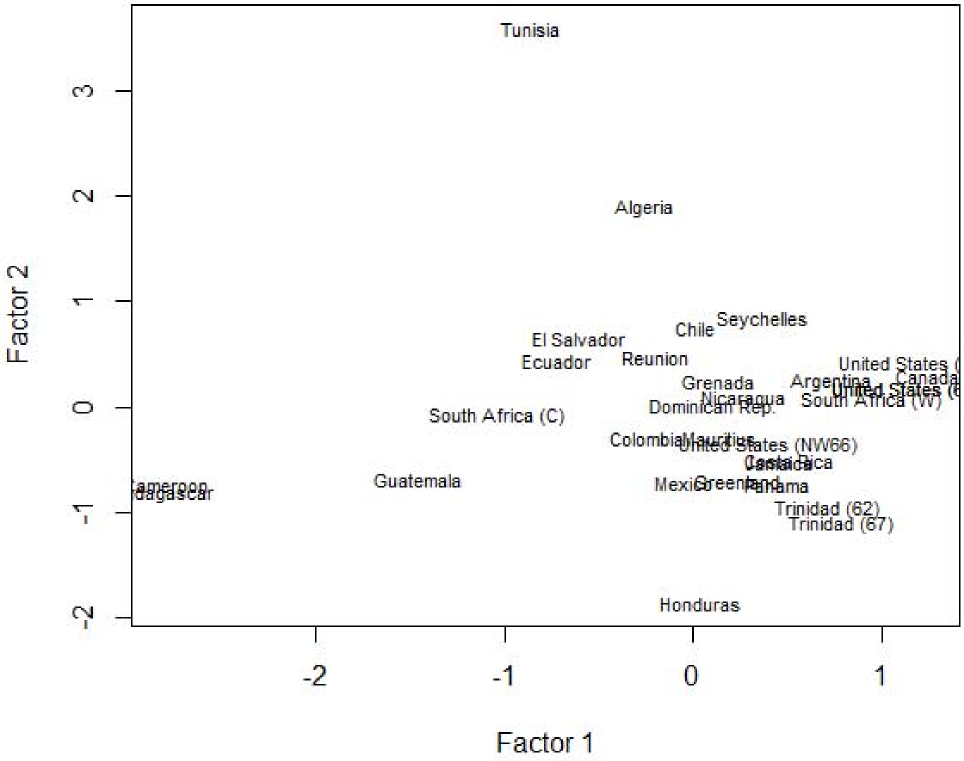
("Blanks" replace negligible loadings.) Examining the estimated factor loadings, we see that the first factor is dominated by life expectancy at birth for both males and females; perhaps this factor could be labeled "life force at birth". The second reflects life expectancies at older ages, and we might label it "life force amongst the elderly". The third factor from the varimax rotation has its highest loadings for the life expectancies of men aged 50 and 75 and in the same vein might be labeled "life force for elderly men". (When labeling factors in this way, factor analysts can often be extremely creative!) The estimated factor scores are found as follows;

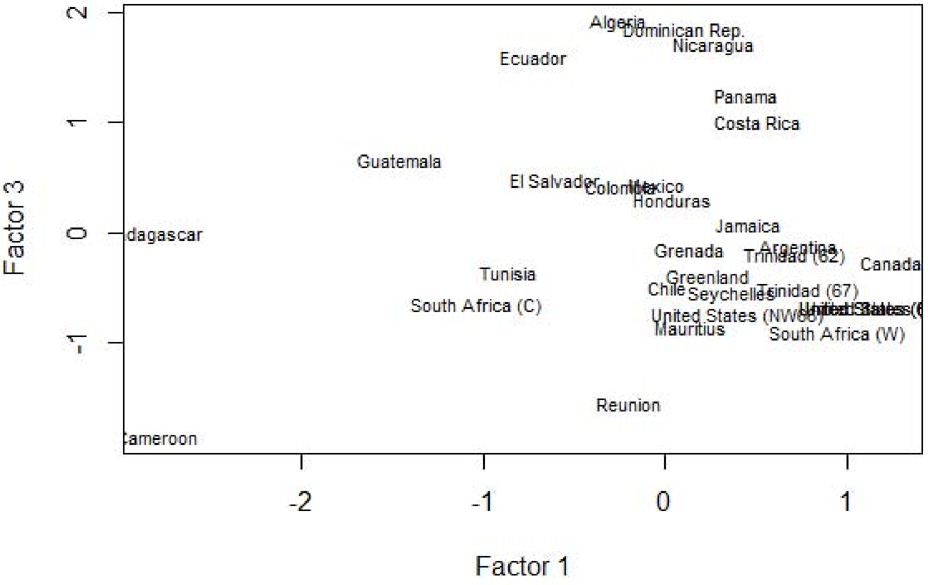
```
> (scores <- factanal(life, factors = 3, method = "mle", scores = "regression")$scores)

                       Factor1      Factor2     Factor3
Algeria            -0.258062561  1.90095771  1.91581631
Cameroon           -2.782495791 -0.72340014 -1.84772224
Madagascar         -2.806428187 -0.81158820 -0.01210318
```

```
Mauritius            0.141004934 -0.29028454 -0.85862443
Reunion             -0.196352142  0.47429917 -1.55046466
Seychelles           0.367371307  0.82902375 -0.55214085
South Africa (C)    -1.028567629 -0.08065792 -0.65421971
South Africa (W)     0.946193522  0.06400408 -0.91995289
Tunisia             -0.862493550  3.59177195 -0.36442148
Canada               1.245304248  0.29564122 -0.27342781
Costa Rica           0.508736247 -0.50500435  1.01328707
Dominican Rep.       0.106044085  0.01111171  1.83871599
El Salvador         -0.608155779  0.65100820  0.48836431
Greenland            0.235114220 -0.69123901 -0.38558654
Grenada              0.132008172  0.25241049 -0.15220645
Guatemala           -1.450336359 -0.67765804  0.65911906
Honduras             0.043253249 -1.85175707  0.30633182
Jamaica              0.462124701 -0.51918493  0.08032855
Mexico              -0.052332675 -0.72020002  0.44417800
Nicaragua            0.268974443  0.08407227  1.70568388
Panama               0.442333434 -0.73778272  1.25218728
Trinidad (62)        0.711367053 -0.95989475 -0.21545329
Trinidad (67)        0.787286051 -1.10729029 -0.51958264
United States (66)   1.128331259  0.16389896 -0.68177046
United States (NW66) 0.400058903 -0.36230253 -0.74299137
United States (W66)  1.214345385  0.40877239 -0.69225320
United States (67)   1.128331259  0.16389896 -0.68177046
Argentina            0.731344988  0.24811968 -0.12817725
Chile                0.009751528  0.75222637 -0.49198911
Colombia            -0.240602517 -0.29543613  0.42919600
Ecuador             -0.723451797  0.44246371  1.59164974
```

We can use the scores to provide the plot of the data shown in Figure 5.1 with the following codes

```
> plot(scores[, 2]~scores[, 1], type = "n", xlab = "Factor 1", ylab = "Factor 2", lwd = 2)
> text(scores[, 2]~scores[, 1], labels = rownames(life), cex = 0.7)
```


Tunisia
Algeria
Chile
Seychelles
El Salvador
Ecuador
Reunion
United States (
Grenada
Argentina
Canada
United States (6
Nicaragua
South Africa (W)
Dominican Rep.
South Africa (C)
Colombia
Mauritius
United States (NW66)
Costa Rica
Jamaica
Cameroon
Madagascar
Guatemala
Mexico
Greenland
Panama
Trinidad (62)
Trinidad (67)
Honduras
3
2
1
0
-1
-2
-2
-1
0
1
Factor 2
Factor 1


Algeria
Dominican Rep.
Nicaragua
Ecuador
Panama
Costa Rica
Guatemala
El Salvador
Colombia
Mexico
Honduras
Jamaica
dagascar
Grenada
Argentina
Trinidad (62)
Canada
Tunisia
Greenland
Chile
Seychelles
Trinidad (67)
South Africa (C)
United States (NW66)
United States (6
Mauritius
South Africa (W)
Reunion
Cameroon
2
1
0
-1
-2
-1
0
1
Factor 3
Factor 1

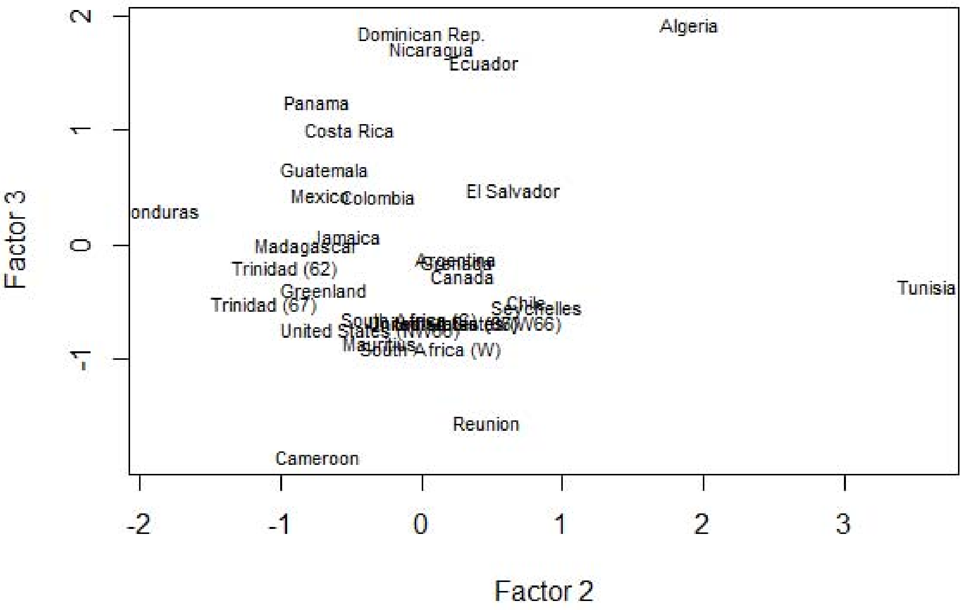


Fig. 5.1. Individual scatterplots of three factor scores for life expectancy data, with points labelled by abbreviated country names.

Ordering along the first axis reflects life force at birth ranging from Cameroon and Madagascar to countries such as the USA. And on the third axis Algeria is prominent because it has high life expectancy amongst men at higher ages, with Cameroon at the lower end of the scale with a low life expectancy for men over 50.

Drug use by American college students

The majority of adult and adolescent Americans regularly use psychoactive substances during an increasing proportion of their lifetimes. Various forms of licit and illicit psychoactive substance use are prevalent, suggesting that patterns of psychoactive substance taking are a major part of the individual's behavioral repertory and have pervasive implications for the performance of other behaviors. In an investigation of these phenomena, data on drug usage rates for 1634 students in the seventh to ninth grades in 11 schools in the greater metropolitan area of Los Angeles collected. Each participant completed a questionnaire about the number of times a particular substance had ever been used. The substances asked about were as follows:

- cigarettes;
- beer;
- wine;
- liquor;
- cocaine;
- tranquillizers;
- drug store medications used to get high;
- heroin and other opiates;
- marijuana;
- hashish;
- inhalants (glue, gasoline, etc.);
- hallucinogenic (LSD, mescaline, etc.);
- amphetamine stimulants.

Responses were recorded on a five-point scale: never tried, only once, a few times, many times, and regularly. The correlations between the usage rates of the 13 substances are shown in Figure 5.2. The plot was produced using the levelplot() function from the package **lattice** with the following somewhat lengthy panel function.

```
> require(ellipse)
> ord <- order.dendrogram(as.dendrogram(hclust(dist(druguse))))
> panel.corrgram = function(x, y, z, subscripts, at, level = 0.9, label = FALSE, ...)
  {
     require("ellipse", quietly = TRUE)
     x <- as.numeric(x)[subscripts]
     y <- as.numeric(y)[subscripts]
     z <- as.numeric(z)[subscripts]
     zcol <- level.colors(z, at = at, col.regions = grey.colors, ...)
     for (i in seq(along = z)) {
     ell <- ellipse(z[i], level = level, npoints = 50,
     scale = c(.2, .2), centre = c(x[i], y[i]))
     panel.polygon(ell, col = zcol[i], border = zcol[i], ...)
     }
     if (label)
```

```
          panel.text(x = x, y = y, lab = 100 * round(z, 2), cex = 0.8,
                    col = ifelse(z < 0, "white", "black"))
   }
> print(levelplot(druguse[ord, ord], at = do.breaks(c(-1.01, 1.01), 20),
        xlab = NULL, ylab = NULL, colorkey = list(space = "top"),
        scales = list(x = list(rot = 90)),
        panel = panel.corrgram, label = TRUE))
```

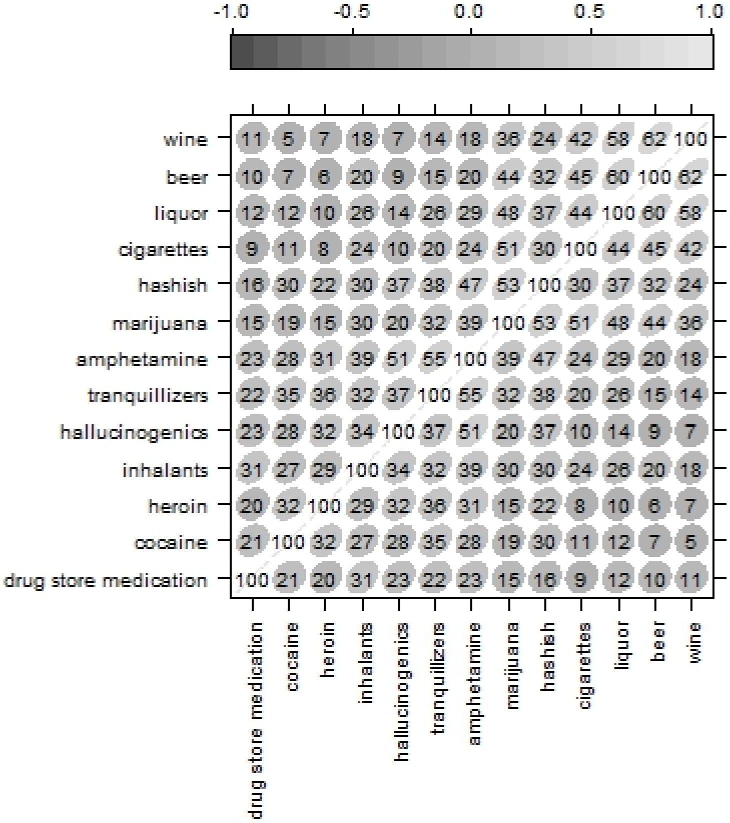


Fig. 5.2. Visualization of the correlation matrix of drug use. The numbers in the cells correspond to 100 times the correlation coefficient. The color and the shape of the plotting symbols also correspond to the correlation in this cell.

The figure depicts each correlation by an ellipse whose shape tends towards a line with slope 1 for correlations near 1, to a circle for correlations near zero, and to a line with negative slope −1 for negative correlations near −1.

In addition, 100 times the correlation coefficient is printed inside the ellipse, and a colourcoding indicates strong negative (dark) to strong positive (light) correlations.

We first try to determine the number of factors using the maximum likelihood test.

```
> sapply(1:6, function(nf)  factanal(covmat = druguse, factors = nf,  method = "mle", n.obs = 1634)$PVAL)
```

```
  objective    objective    objective    objective    objective    objective
0.000000e+00 9.786000e-70 7.363910e-28 1.794578e-11 3.891743e-06 9.752967e-02
```

These values suggest that only the six-factor solution provides an adequate fit. The results from the six-factor varimax solution are obtained from

```
> (factanal(covmat = druguse, factors = 6, method = "mle", n.obs = 1634))

Call:
factanal(factors = 6, covmat = druguse, n.obs = 1634, method = "mle")

Uniquenesses:
           cigarettes                  beer                  wine 
                0.563                 0.368                 0.374 
               liquor               cocaine        tranquillizers 
                0.412                 0.681                 0.522 
drug store medication                heroin             marijuana 
                0.785                 0.669                 0.318 
              hashish             inhalants       hallucinogenics 
                0.005                 0.541                 0.620 
          amphetamine 
                0.005 

Loadings:
                      Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
cigarettes              0.494                           0.407   0.110 
beer                    0.776                           0.112         
wine                    0.786                                         
liquor                  0.720   0.121   0.103   0.115   0.160         
cocaine                         0.519           0.132           0.158 
tranquillizers          0.130   0.564   0.321   0.105   0.143         
drug store medication           0.255                           0.372 
heroin                          0.532   0.101                   0.190 
marijuana               0.429   0.158   0.152   0.259   0.609   0.110 
hashish                 0.244   0.276   0.186   0.881   0.194   0.100 
inhalants               0.166   0.308   0.150           0.140   0.537 
hallucinogenics                 0.387   0.335   0.186           0.288 
amphetamine             0.151   0.336   0.886   0.145   0.137   0.187 

               Factor1 Factor2 Factor3 Factor4 Factor5 Factor6
SS loadings      2.301   1.415   1.116   0.964   0.676   0.666
Proportion Var   0.177   0.109   0.086   0.074   0.052   0.051
Cumulative Var   0.177   0.286   0.372   0.446   0.498   0.549

Test of the hypothesis that 6 factors are sufficient.
The chi square statistic is 22.41 on 15 degrees of freedom.
The p-value is 0.0975
```

Substances that load highly on the first factor are cigarettes, beer, wine, liquor, and marijuana and we might label it “social/soft drug use”. Cocaine, tranquillizers, and heroin load highly on the second factor- the obvious label for the factor is “hard drug use”. Factor three is essentially simply amphetamine use, and factor four hashish use. We will not try to interpret the last two factors, even though the formal test for number of factors indicated that a six-factor solution was necessary. It may be that we should not take the results of the formal test too literally; rather, it may be a better strategy to consider the value of k indicated by the test to be an upper bound on the number of factors with practical importance. Certainly a six-factor solution for a data set with only 13 manifest variables might be regarded as not entirely satisfactory, and clearly we would have some difficulties interpreting all the factors. One of the problems is that with the large sample size in this example, even small discrepancies between the correlation matrix predicted by a proposed model and the observed correlation matrix may lead to rejection of the model. One way to investigate this possibility is simply to look at the differences between the observed and predicted correlations. We shall do this first for the six-factor model using the following R code:

```
> pfun <- function(nf) {
        fa <- factanal(covmat = druguse, factors = nf,
        method = "mle", n.obs = 1634)
        est <- tcrossprod(fa$loadings) + diag(fa$uniquenesses)
        ret <- round(druguse - est, 3)
        colnames(ret) <- rownames(ret) <-
        abbreviate(rownames(ret), 3)
        ret
              }
> pfun(6)
```

```
       cgr    ber    win    lqr    ccn    trn    dsm    hrn    mrj hsh    inh    hll
cgr  0.000 -0.001  0.014 -0.018  0.010  0.001 -0.020 -0.004  0.001   0  0.010 -0.005
ber -0.001  0.000 -0.002  0.004  0.004 -0.011 -0.001  0.007  0.002   0 -0.004  0.005
win  0.014 -0.002  0.000 -0.001 -0.001 -0.005  0.008  0.008 -0.004   0 -0.007 -0.001
lqr -0.018  0.004 -0.001  0.000 -0.008  0.021 -0.006 -0.018  0.003   0  0.012 -0.005
ccn  0.010  0.004 -0.001 -0.008  0.000  0.000  0.008  0.004 -0.004   0 -0.003 -0.008
trn  0.001 -0.011 -0.005  0.021  0.000  0.000  0.006 -0.004 -0.004   0  0.002 -0.008
dsm -0.020 -0.001  0.008 -0.006  0.008  0.006  0.000 -0.015  0.008   0  0.004 -0.002
hrn -0.004  0.007  0.008 -0.018  0.004 -0.004 -0.015  0.000  0.006   0 -0.002  0.020
mrj  0.001  0.002 -0.004  0.003 -0.004 -0.004  0.008  0.006  0.000   0 -0.006  0.003
hsh  0.000  0.000  0.000  0.000  0.000  0.000  0.000  0.000  0.000   0  0.000  0.000
inh  0.010 -0.004 -0.007  0.012 -0.003  0.002  0.004 -0.002 -0.006   0  0.000 -0.002
hll -0.005  0.005 -0.001 -0.005 -0.008 -0.008 -0.002  0.020  0.003   0 -0.002  0.000
amp  0.000  0.000  0.000  0.000  0.000  0.000  0.000  0.000  0.000   0  0.000  0.000
    amp
cgr   0
ber   0
win   0
lqr   0
ccn   0
trn   0
dsm   0
hrn   0
mrj   0
hsh   0
inh   0
hll   0
amp   0
```

The differences are all very small, underlining that the six-factor model does describe the data very well. Now let us look at the corresponding matrices for the three- and four-factor solutions found in a similar way in Figure 5.3. Again, in both cases the residuals are all relatively small, suggesting perhaps that use of the formal test for number of factors leads, in this case, to overfitting. The three-factor model appears to provide a perfectly adequate fit for these data.

```
> pfun(3)
```

```
     cgr    ber    win    lqr    ccn    trn    dsm    hrn    mrj    hsh    inh
cgr  0.000 -0.001  0.009 -0.013  0.011  0.009 -0.011 -0.004  0.003 -0.027  0.039
ber -0.001  0.000 -0.002  0.002  0.002 -0.014  0.000  0.005 -0.001  0.019 -0.002
win  0.009 -0.002  0.000  0.000 -0.002 -0.004  0.012  0.013  0.001 -0.017 -0.007
lqr -0.013  0.002  0.000  0.000 -0.008  0.024 -0.017 -0.020 -0.001  0.014 -0.002
ccn  0.011  0.002 -0.002 -0.008  0.000  0.031  0.038  0.082 -0.002  0.041  0.023
trn  0.009 -0.014 -0.004  0.024  0.031  0.000 -0.021  0.026 -0.002 -0.016 -0.038
dsm -0.011  0.000  0.012 -0.017  0.038 -0.021  0.000  0.021  0.007 -0.040  0.113
hrn -0.004  0.005  0.013 -0.020  0.082  0.026  0.021  0.000  0.006 -0.035  0.031
mrj  0.003 -0.001  0.001 -0.001 -0.002 -0.002  0.007  0.006  0.000  0.001  0.003
hsh -0.027  0.019 -0.017  0.014  0.041 -0.016 -0.040 -0.035  0.001  0.000 -0.035
inh  0.039 -0.002 -0.007 -0.002  0.023 -0.038  0.113  0.031  0.003 -0.035  0.000
hll -0.017  0.009  0.004 -0.015 -0.030 -0.058  0.000 -0.005 -0.002  0.034  0.007
amp  0.002 -0.007  0.002  0.006 -0.075  0.044 -0.038 -0.049 -0.002  0.010 -0.015
       hll    amp
cgr -0.017  0.002
ber  0.009 -0.007
win  0.004  0.002
lqr -0.015  0.006
ccn -0.030 -0.075
trn -0.058  0.044
dsm  0.000 -0.038
hrn -0.005 -0.049
mrj -0.002 -0.002
hsh  0.034  0.010
inh  0.007 -0.015
hll  0.000  0.041
amp  0.041  0.000
> pfun(4)
       cgr    ber    win    lqr    ccn    trn    dsm    hrn    mrj    hsh    inh
cgr  0.000 -0.001  0.008 -0.012  0.009  0.008 -0.015 -0.007  0.001 -0.023  0.037
ber -0.001  0.000 -0.001  0.001  0.000 -0.016 -0.002  0.003 -0.001  0.018 -0.005
win  0.008 -0.001  0.000  0.000 -0.001 -0.005  0.012  0.014  0.001 -0.020 -0.008
lqr -0.012  0.001  0.000  0.000 -0.004  0.029 -0.015 -0.015 -0.001  0.018  0.001
ccn  0.009  0.000 -0.001 -0.004  0.000  0.024 -0.014  0.007 -0.003  0.035 -0.022
trn  0.008 -0.016 -0.005  0.029  0.024  0.000 -0.020  0.027 -0.001  0.001 -0.032
dsm -0.015 -0.002  0.012 -0.015 -0.014 -0.020  0.000 -0.018  0.003 -0.042  0.090
hrn -0.007  0.003  0.014 -0.015  0.007  0.027 -0.018  0.000  0.003 -0.037 -0.001
mrj  0.001 -0.001  0.001 -0.001 -0.003 -0.001  0.003  0.003  0.000  0.000  0.001
hsh -0.023  0.018 -0.020  0.018  0.035  0.001 -0.042 -0.037  0.000  0.000 -0.031
inh  0.037 -0.005 -0.008  0.001 -0.022 -0.032  0.090 -0.001  0.001 -0.031  0.000
hll -0.020  0.006  0.001 -0.010 -0.028 -0.028  0.008  0.005 -0.002  0.055  0.021
amp  0.000  0.000  0.000 -0.001  0.000  0.001  0.000  0.000  0.000 -0.001  0.000
       hll    amp
cgr -0.020  0.000
ber  0.006  0.000
win  0.001  0.000
lqr -0.010 -0.001
ccn -0.028  0.000
trn -0.028  0.001
dsm  0.008  0.000
hrn  0.005  0.000
mrj -0.002  0.000
hsh  0.055 -0.001
inh  0.021  0.000
hll  0.000  0.000
amp  0.000  0.000
```

Fig. 5.3. Differences between three- and four-factor solutions and actual correlation matrix for the drug use data.

10. 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 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, but 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 analyzing either matrix are essentially equivalent (which is not true of principal factor analysis).

Despite these differences, the results from both types of analyses are frequently very similar. Certainly, if the specific variances are small, we would expect both forms of analyses 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 that are similar to the original variables.