FACTOR ANALYSIS

1 Introduction

Reference: Chapter 9, Johnson and Wichern

The general purpose of factor analysis is to find a way of summarising the information contained in a number of correlated random variables; $X_1, X_2, \cdots$, X_p into a smaller set of random variables called *factors*, with a minimum loss of information. Usually these factors are refer to as *common factors* and denoted by $F_1, F_2, \cdots, F_m$ for a suitably chosen $m(< p)$. The main aim of the factor analysis techniques is to search and construct the common factors using the all available information. Factor analysis can be considered as an extension of principal component analysis of which the basic idea is to develop a set of linear combinations from the original variables to condense the data. If the common factors are orthogonal (independent) then model is called *an orthogonal factor model* otherwise it is called *an oblique model*.

2 The Orthogonal Factor Model

Let $\boldsymbol{X}$ be a p-variate random vector with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Suppose there are m unobservable orthogonal factors (independent random variables), $F_1, F_2, \cdots, F_m$ then the factor model can be written as:

$$
\begin{aligned}
X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + \varepsilon_1 \\
X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \cdots + l_{2m}F_m + \varepsilon_2 \\
&\vdots \\
X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + \varepsilon_p
\end{aligned}
$$

In matrix notation,

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_i)_{p\times 1}$ is the random error vector, ε_i are called *Specific factors*. The coefficient l_{ij} is the loading for the i^{th} variable on the j^{th} factor and $\boldsymbol{L}$ is the matrix of *factor loadings*.

Assumptions

(a) $\mathbf{E}(\boldsymbol{F}) = \mathbf{0}$ and $\mathbf{Cov}(\boldsymbol{F}) = \boldsymbol{I}_m$. This ensures that $F_1, F_2, \cdots, F_m$ are independent standardised random variables.

(b) $\mathbf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathbf{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}_p = \text{diag}(\psi_1, \psi_2, \cdots, \psi_p)$. This implies that ε_i's are mutually independent random errors.

(c) $\boldsymbol{F}$ and $\boldsymbol{\varepsilon}$ are independent.

Theorem: For the orthogonal factor model given above,

(a) $\Sigma = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}_p$ and

(b) $\mathbf{Cov}(\boldsymbol{X}, \boldsymbol{F}) = \boldsymbol{L}$

provided the assumptions given above are valid.

Remark 1: $\Sigma = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}_p$ is equivalent to $\sigma_{ii} = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \psi_i$ for $i = 1, 2, \cdots, p$ and $\sigma_{ij} = l_{i1}l_{j1} + l_{i2}l_{j2} + \cdots + l_{im}l_{jm}$ for all $i \neq j$. That is, there are $\frac{1}{2}p(p+1)$ equations from the above matrix relationship.

Note that a factor model with m common factors and p original X variables has mp unknown l_{ij}'s and p unknown ψ_i's. That is, the model contains total of $p(m+1)$ unknown parameters, so the number of equations number of unknown parameters may be different, therefore this method may leads to multiple solutions.

Remark 2: $\mathbf{Cov}(\boldsymbol{X}, \boldsymbol{F}) = \boldsymbol{L}$ implies that $\mathbf{Cov}(X_i, F_j) = l_{ij}$ for $i, j = 1, 2, \cdots, p$.

Definition: The portion of the variance of X_i shared by the common factors in the model, is called the i^{th} communality and denoted by h_i^2. From the results given above,

$$\text{var}(X_i) = \sigma_{ii} = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \psi_i = h_i^2 + \psi_i$$

That is, the i^{th} communality is given by

$$h_i^2 = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2$$

Example 1: Let $\boldsymbol{X}^T = (X_1, X_2, X_3)$ be a trivariate random vector with mean $(5, 10, 8)^T$ and covariance matrix Σ is given by

$$\Sigma = \begin{pmatrix} 10 & 12 & 6 \\ 12 & 18 & 8 \\ 6 & 8 & 9 \end{pmatrix}.$$

Obtain an orthogonal factor model with one common factor.

Remark 3: In some cases, covariance matrices cannot be factored as $\Sigma = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}_p$ when the number of factors, m is much less than p.

Example 2: Given $\boldsymbol{\mu}^T = (50, 100, 75)$ and

$$\Sigma = \begin{pmatrix} 21 & 30 & 2 \\ 30 & 57 & 5 \\ 2 & 5 & 38 \end{pmatrix}.$$

Show that

(a) orthogonal two factor model with $\boldsymbol{L}^T = \begin{pmatrix} 4 & 7 & -1 \\ 1 & 2 & 6 \end{pmatrix}$ satisfy the information given in the above example where as

(b) a model with one common factor does not exists.

METHODS OF ESTIMATION

3 The Principal Factor Method

Let $\boldsymbol{S}$ be the sample covariance matrix and $\left(\widehat{\lambda}_i, \widehat{\boldsymbol{e}}_i\right)$ $(i = 1, 2, ..., p)$ be the i^{th} (eigenvalue,eigenvector) pair of $\boldsymbol{S}$ such that $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_p$. Take

$$\widehat{\boldsymbol{L}} = (\sqrt{\widehat{\lambda}_1}\widehat{\boldsymbol{e}}_1, \sqrt{\widehat{\lambda}_2}\widehat{\boldsymbol{e}}_2, \cdots, \sqrt{\widehat{\lambda}_m}\widehat{\boldsymbol{e}}_m)_{p\times m}$$

Then, using the spectral decomposition of $\boldsymbol{S}$, we can prove that $\boldsymbol{S} - \widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}^T$ is approximately a diagonal matrix. Therefore $\widehat{\boldsymbol{L}}$ is an estimator of $\boldsymbol{L}$. Hence,

$$\boldsymbol{S} \approx \widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}^T + \widehat{\boldsymbol{\Psi}}$$

where $\widehat{\boldsymbol{\Psi}} = \text{diag}(\widehat{\psi}_1, \widehat{\psi}_2, \cdots, \widehat{\psi}_p)$ and $\widehat{\psi}_i$ is the i^{th} diagonal element of $(\widehat{\lambda}_{m+1}\widehat{\boldsymbol{e}}_{m+1}\widehat{\boldsymbol{e}}_{m+1}^T + \widehat{\lambda}_{m+2}\widehat{\boldsymbol{e}}_{m+2}\widehat{\boldsymbol{e}}_{m+2}^T + \cdots + \widehat{\lambda}_p\widehat{\boldsymbol{e}}_p\widehat{\boldsymbol{e}}_p^T)$. Also, the communalities are estimated by

$$\widehat{h}_i^2 = \sum_{j=1}^{m} \widehat{l}_{ij}^2 = s_{ii} - \widehat{\psi}_i$$

3.1 Percentage of Variance Criterion

This is the most commonly used method to determine the number factors and this uses the cumulative percentages of the variances extracted by successive factors.

The cumulative percentages of the variances, κ extracted by $m(< p)$ successive factors, is given by

$$\kappa = 100\frac{\sum_{i=1}^{m}\widehat{\lambda}_i}{\sum_{i=1}^{p}\widehat{\lambda}_i}$$

where $\sum_{i=1}^{p}\widehat{\lambda}_i$ is the total variance and $\sum_{i=1}^{m}\widehat{\lambda}_i$ is the variance of first m successive factors. For any $m < p$, $0 < \kappa < 100$, therefore it is recommended to use a value closer to 100 to achieve a better model. Most analyst prefer to use very high values such as $\kappa = 95.0$ but it is not uncommon to use a value as low as 60.

3.2 Eigenvalue Criterion

This is the most simplest and easy apply criterion to determine the number of factors. Here m is set equal to the number eigenvalues of the correlation matrix, greater than or equal to 1. This criterion assumes that all the factors with eigenvalues less than 1 are insignificant and disregarded. The reliability of this criterion appears to be increasing with increasing number of original variables and if the number of variables are less than 20 then it is not recommended to use this criterion to determine the number of common factors.

Example 3: The sample mean $\overline{\boldsymbol{X}}$ and sample covariance matrix $\boldsymbol{S}$ for a multivariate population with unknown mean $\boldsymbol{\mu}$ and covariance matrix Σ is obtained using a random sample of size 25. They are:

$$\overline{\boldsymbol{X}} = \begin{pmatrix} 5 \\ 10 \\ 8 \end{pmatrix} \qquad \boldsymbol{S} = \begin{pmatrix} 19 & 30 & 2 \\ 30 & 57 & 5 \\ 2 & 5 & 38 \end{pmatrix}.$$

Estimate a suitable orthogonal factor model.

Example 4: A random sample of size 25 was obtained form multivariate population. The computed values for the sample mean and the sample covariance matrix are given below:

$$\overline{\boldsymbol{X}} = \begin{pmatrix} 5 \\ 2 \\ 8 \\ 8 \end{pmatrix} \qquad \boldsymbol{S} = \begin{pmatrix} 18.87 & 26.86 & 7.55 & -5.15 \\ 26.86 & 47.24 & 5.10 & -15.23 \\ 7.55 & 5.10 & 92.19 & 58.89 \\ -5.15 & -15.23 & 58.89 & 48.52 \end{pmatrix}.$$

Estimate a suitable orthogonal factor model.

Remark 4: The estimated factor loadings for a given factor, by this method, do not change as the number of factors is increased or decreased.

Remark 5: The principal component factor analysis of the sample correlation matrix is obtained by replacing the sample covariance matrix with the sample correlation matrix.

4 Maximum Likelihood Method

If the common factors, $\boldsymbol{F}$ and the specific factors (random errors), $\boldsymbol{\varepsilon}$ are normal random vectors, then $\boldsymbol{X}$ is also a normal random vector with mean $\boldsymbol{\mu}$ and covariance Σ. The maximum likelihood method can be used to estimate the factor model provided the normality assumptions on $\boldsymbol{F}$ and $\boldsymbol{\varepsilon}$ are valid.

Assumptions for the maximum likelihood method

(a) $\boldsymbol{F}$ is normal with $\mathbf{E}(\boldsymbol{F}) = \mathbf{0}$ and $\mathbf{Cov}(\boldsymbol{F}) = \boldsymbol{I}_m$. That is, $F_i \sim N(0,1)$ and mutually independent.

(b) $\boldsymbol{\varepsilon}$ is normal with $\mathbf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathbf{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}_p$= diag $(\psi_1, \psi_2, \cdots, \psi_p)$. This implies that $\varepsilon_i \sim N(0, \psi_i)$ and mutually independent.

(c) $\boldsymbol{F}$ and $\boldsymbol{\varepsilon}$ are independent

The above assumptions implies that $\boldsymbol{X}$ is multivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix Σ, that is, the probability density function of $\boldsymbol{X}$ is given by

$$f(\boldsymbol{x} : \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}p} |\Sigma|^{\frac{1}{2}}} \exp\left\{ \frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\}$$

for all $\boldsymbol{x}$.

Suppose $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ is a random sample from $\boldsymbol{X}$, then the likelihood function for the sample is given by

$$\mathcal{L}(\boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}p} |\Sigma|^{\frac{n}{2}}} \exp\left\{ \frac{1}{2} \sum_{j=1}^{n} (\boldsymbol{x}_j - \boldsymbol{\mu})^T \Sigma^{-1} \boldsymbol{x}_j - \boldsymbol{\mu}) \right\}$$

Since $\overline{\boldsymbol{x}}$ is the mle for $\boldsymbol{\mu}$ and $\Sigma = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}_p$, maximisation of $\mathcal{L}(\boldsymbol{\mu}, \Sigma)$ is equivalent to minimisation of

$$\mathcal{F}(\boldsymbol{L}, \boldsymbol{\Psi}_p) = \log|\boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}_p| + \mathrm{tr}\{(\boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi}_p)^{-1}\boldsymbol{S} - \log|\boldsymbol{S}| - p.$$

However, this minimisation may leads to multiple results for $\boldsymbol{L}$. To ensure the uniqueness of all the estimators, we impose the following uniqueness condition

$$\boldsymbol{\Gamma} = \boldsymbol{L}^T\boldsymbol{\Psi}_p^{-1}\boldsymbol{L}$$

where $\boldsymbol{\Gamma}$ is an unknown diagonal matrix. In this case the maximum likelihood estimators, $\widehat{\boldsymbol{L}}$ and $\hat{\boldsymbol{\Psi}}_p$ receptively for $\boldsymbol{L}$ and $\boldsymbol{\Psi}_p$. These estimators cannot be obtained algebraically.

These estimators can be obtained by numerical maximisation of (16.19) together with the uniqueness condition given in (16.20) provided m, the number of common factors are known and reasonably less than p. The computer packages such as SAS, SPSS and MINITAB version 10 (or higher) can be used to obtain these estimators numerically for a given value of m.

Example 5: The sample correlation matrix, $\boldsymbol{R}$ given of below is computed for seven descriptive characteristics observed on 30 residential properties. The characteristics are as follows:

X_1 = area in square feet,

X_2 = total number of rooms,

X_3 = number of bedrooms,

X_4 = number of bathrooms,

X_5 = age of the property,

X_6 = attached garage or car-port

(0=no garage or car-port, 1=single, 2= double or higher) and

X_7 = view(1=good, 0=poor).

Estimate a suitable factor model using the maximum likelihood procedure.

$$\boldsymbol{R} = \begin{pmatrix} 1.00 & 0.55 & 0.40 & 0.53 & 0.45 & 0.47 & 0.28 \\ 0.55 & 1.00 & 0.43 & 0.75 & 0.35 & 0.42 & 0.18 \\ 0.40 & 0.43 & 1.00 & 0.31 & 0.45 & 0.21 & 0.15 \\ 0.53 & 0.75 & 0.31 & 1.00 & 0.64 & 0.40 & 0.44 \\ 0.45 & 0.35 & 0.45 & 0.64 & 1.00 & 0.48 & 0.25 \\ 0.47 & 0.42 & 0.21 & 0.40 & 0.48 & 1.00 & 0.01 \\ 0.28 & 0.18 & 0.15 & 0.44 & 0.25 & 0.01 & 1.00 \end{pmatrix}.$$

Remark 6: In general, the maximum likelihood method gives a better estimation of the model than the principal component method, provided the normality assumption is valid. Further, it is recommended that the normality test should be performed with the data before using the maximum likelihood method.

5 A Statistical Test for the Number of Common Factors

The maximum likelihood procedure given in the previous section enables us to test the hypothesis that m common factors used in the model, are sufficient to describe the data.

The test statistic for testing the validity of the number of common factors, m is given by

$$U = k \min \left(\mathcal{F}(\boldsymbol{L}, \boldsymbol{\Psi}_p) \right)$$

where $k = n - \frac{1}{6}(2p + 4m + 11)$, n is the sample size. If m common factors are sufficient, then U is asymptotically χ^2 distributed with $\nu = \frac{1}{2}\{(p-m)^2 - (p+m)\}$ degrees of freedom provided the normality assumption is valid. If $U < \chi^2_\nu(\alpha)$ then accept m common factor model at α level of significance otherwise, increase the number of common factors by 1 and repeat the above test. Note that $m < \frac{1}{2}(2p + 1 - \sqrt{8p + 1})$ to use this hypothesis testing procedure. If these conditions are not satisfied, then the degrees of freedom, $\nu < 1$.

Example 6: Use the statistical testing procedure given above to obtain suitable number of common factors for the data given in Example 4.

6 Factor Rotation

In some cases, the factors obtained in the solution are difficult to identify and interpret. The process of adjusting the factor axes to achieve a simpler and more meaningful factors is called factor rotation. In general there are two types of rotations namely: orthogonal rotation and oblique rotation. The orthogonal factor rotation maintain the factor axes at 90 degrees, thus, the orthogonality (independence) of the common factors are preserved. The oblique rotation manipulate factor axes to achieve a preset correlations between the common factors. In other word, the oblique rotation produce factors which are not orthogonal.

A variety of algorithms is used for factor rotations to achieve a simple structure. Among them, the most popular rotations are varimax, quartimax and equamax. These are all orthogonal rotations, varimax attempts to minimise the number of variables that have high loadings on a factor; quartimax minimises number of factors needed to explain a variable; equamax is a combination of the varimax and quartimax rotations. That is varimax simplifies the common factors, quartimax simplifies the variables and equamax attempts to simplify both factors and variables.

Example 7: Let X_1, X_2, X_3, X_4, X_5 denote the weekly rate of return for five selected stocks listed in the Melbourne stock exchange. The correlation matrix between the stocks based on 50 observations is given below:

$$\boldsymbol{R} = \begin{pmatrix} 1.00 & 0.79 & 0.42 & 0.71 & 0.50 \\ 0.79 & 1.00 & 0.01 & 0.85 & 0.11 \\ 0.42 & 0.01 & 1.00 & 0.02 & 0.96 \\ 0.71 & 0.85 & 0.02 & 1.00 & 0.13 \\ 0.50 & 0.11 & 0.96 & 0.13 & 1.00 \end{pmatrix}.$$

Obtain two factor orthogonal model using maximum likelihood method. Apply varimax rotation to the results and explain the model.

Using MINITAB Version 10 Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Commnlty
Var 1	0.741	-0.531	0.832
Var 2	0.963	-0.141	0.947
Var 3	-0.134	-0.977	0.973
Var 4	0.861	-0.145	0.763
Var 5	-0.027	-0.979	0.959
Variance	2.2376	2.2362	4.4738

Rotated Factor Loadings and Communalities

Varimax Rotation

Variable	Factor1	Factor2	Commnlty
Var 1	0.818	-0.404	0.832
Var 2	0.973	0.018	0.947
Var 3	0.027	-0.986	0.973
Var 4	0.874	-0.003	0.763
Var 5	0.133	-0.970	0.959

Varimax rotation in the above example, makes the insignificant loadings numerically smaller than the corresponding unrotated values. Clearly, the companies X_1, X_2 and X_4 define the **Factor 1** and X_3 and X_5 define **Factor 2**. Note that the third and fifth loadings in Factor 1, are comparatively small. Also the first, second and fourth loading in Factor 2, are small.