Chapter 12 Factor Analysis

Applied Multivariate Statistical Analysis

Course Instructor: Shuen-Lin Jeng

Department of Statistics
National Cheng Kung University

April, 8, 2021

12.1 The Orthogonal Factor Model

- A frequently applied paradigm in analyzing data from multivariate observations is to model the relevant information (represented in a multivariate variable X) as coming from a limited number of latent factors.
- In a survey on household consumption, for example, the consumption levels, X, of p different goods during one month could be observed. The variations and covariations of the p components of X throughout the survey might in fact be explained by two or three main social behavior factors of the household.

12.1 The Orthogonal Factor Model

- We assume that there is a model (it will be called the "Factor Model") stating that most of the covariances between the p elements of X can be explained by a limited number of latent factors.
- The aim of factor analysis is to explain the outcome of p variables in the data matrix $\mathcal{X}$ using fewer variables, the so-called *factors*.
- The case just described occurs when every observed $x = (x_1, \ldots, x_p)^\top$ can be written as

$$x_j = \sum_{\ell=1}^{k} q_{j\ell} f_\ell + \mu_j, \ \mathrm{j} = 1, \ldots, \mathrm{p}. \tag{12.1}$$

12.1 The Orthogonal Factor Model

- Here f_ℓ, for $\ell = 1, \dots, k$ denotes the factors. The number of factors, k, should always be much smaller than p.
- For instance, in psychology x may represent p results of a test measuring intelligence scores. One common latent factor explaining $x \in \mathbb{R}^p$ could be the overall level of "intelligence".
- In marketing studies, x may consist measures could be explained by common latent factors like the attraction level of the product or the image of the brand, and so on.
- Indeed, it is possible to create a representation of the observations that is similar to the one in (12.1) by means of principal components, but only if the last $p - k$ eigenvalues corresponding to the covariance matrix are equal to zero.

12.1 The Orthogonal Factor Model

- Consider a p-dimensional random vector X with mean μ and covariance matrix $\mathbf{Var}(X) = \sum$.
- A model similar to (12.1) can be written for X in matrix notation, namely

$$X = QF + \mu, \tag{12.2}$$

where F is the k-dimensional vector of the k factors.
- When using the factor model (12.2) it is often assumed that the factors F are centered, uncorrelated and standardized: $\mathbf{E}(F) = 0$ and $\mathbf{Var}(F) = \mathcal{I}_k$.

12.1 The Orthogonal Factor Model

- The spectral decomposition of $\sum$ is given by $\Gamma\Lambda\Gamma^{\top}$.
- Suppose that only the first k eigenvalues are positive, i.e., $\lambda_{k+1} = \ldots = \lambda_p = 0$. Then the (singular) covariance matrix can be written as

$$\sum = \sum_{\ell=1}^{k} \lambda_\ell \gamma_\ell \gamma_\ell^{\top} = (\Gamma_1\ \Gamma_2) \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Gamma_1^{\top} \\ \Gamma_2^{\top} \end{pmatrix}.$$

where Γ_1 is a $p \times k$ matrix of eigenvectors with k nonzero eigenvalues and Γ_2 is a $p \times (p - k)$ matrix of eigenvectors with eigenvalue 0.

12.1 The Orthogonal Factor Model

- $$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \Gamma_1^\top \\ \Gamma_2^\top \end{pmatrix} (X - \mu), \text{ where } \begin{pmatrix} \Gamma_1^\top \\ \Gamma_2^\top \end{pmatrix} (X - \mu) \sim \left(0, \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix}\right).$$

  In other words, Y_2 has a singular distribution with mean and covariance matrix equal to zero.
- Therefore, $X - \mu = \Gamma_1 Y_1 + \Gamma_2 Y_2$ implies that $X - \mu$ is equivalent to $\Gamma_1 Y_1$. which can be written as

  $$X = \Gamma_1 \Lambda_1^{1/2} \Lambda_1^{-1/2} Y_1 + \mu$$

  Defining $Q = \Gamma_1 \Lambda_1^{1/2}$ and $F = \Lambda_1^{-1/2} Y_1$, we obtain the factor model (12.2).

12.1 The Orthogonal Factor Model

- It is common praxis in factor analysis to split the influences of the factors into common and specific ones. There are, for example, highly informative factors that are common to all of the components of X and factors that are specific to certain components.
- The factor analysis model used in praxis is a generalization of (12.2):

$$X = QF + U + \mu \tag{12.4}$$

where Q is a $(p \times k)$ matrix of the (non-random) *loadings* of the *common factors* $F(k \times 1)$ and U is a $(p \times 1)$ matrix of the (random) *specific factors*.

12.1 The Orthogonal Factor Model

It is assumed that:

$$\begin{aligned}\mathbf{E}F &= 0,\\ \mathbf{Var}(F) &= \mathcal{I}_k,\\ \mathbf{E}U &= 0,\\ \mathbf{Cov}(U_i, U_j) &= 0, i \neq j\\ \mathbf{Cov}(F, U) &= 0.\end{aligned} \tag{12.5}$$

Define

$$\mathbf{Var}(U) = \Psi = \operatorname{diag}(\psi_{11}, \dots, \psi_{pp}).$$

The generalized factor model (12.4) together with the assumptions given in (12.5) constitute the *orthogonal factor model.*

12.1 The Orthogonal Factor Model

Orthogonal Factor Model

$$\underset{(p\times 1)}{X} = \underset{(p\times k)}{Q}\ \underset{(k\times 1)}{F} + \underset{(p\times 1)}{U} + \underset{(p\times 1)}{\mu}$$

$$\mu_j = \text{ mean of variable } j$$

$$U_j = j\text{-th specific factor}$$

$$F_\ell = \ell\text{-th common factor}$$

$$q_{j\ell} = \text{loading of the } j\text{-th variable on the } \ell\text{-th factor}$$

The random factor F and U are unobservable and uncorrelated.

12.1 The Orthogonal Factor Model

Note that (12.4) implies for the components of $X = (X_1, \ldots, X_p)^\top$ that

$$X_j = \sum_{\ell=1}^{k} q_{j\ell} F_\ell + U_j + \mu_j, \; j = 1, \ldots, p. \tag{12.6}$$

Using (12.5) we obtain $\sigma_{X_j X_j} = \mathbf{Var}(X_j) = \sum_{\ell=1}^{k} q_{j\ell}^2 + \psi_{jj}$.

12.1 The Orthogonal Factor Model

The quantity $h_j^2 = \sum_{\ell=1}^{k} q_{j\ell}^2$ is called the *communality* and ψ_{jj} the *specific variance*. Thus the covariance of X can be rewritten as

$$\sum = \mathbf{E}(X-\mu)(X-\mu)^{\top}$$

$$= \mathbf{E}(\mathcal{Q}F+U)(\mathcal{Q}F+U)^{\top}$$

$$= \mathcal{Q}\mathbf{E}(FF^{\top})\mathcal{Q}^{\top} + \mathbf{E}(UU^{\top}) = \mathcal{Q}\mathbf{Var}(F)\mathcal{Q}^{\top} + \mathbf{Var}(U)$$

$$= \mathcal{Q}\mathcal{Q}^{\top} + \Psi. \tag{12.7}$$

12.1 The Orthogonal Factor Model

- If the assumptions are not met, the analysis could be spurious. Although principal components analysis and factor analysis might be related (this was hinted at in the derivation of the factor model), they are quite different in nature.
- PCs are linear transformations of X arranged in decreasing order of variance and used to reduce the dimension of the data set, whereas in factor analysis, we try to model the variations of X using a linear transformation of a fixed, limited number of latent factors. The objective of factor analysis is to find the loadings Q and the specific variance Ψ.

12.1 The Orthogonal Factor Model

- ***Interpretation of the Factors***

  To interpret F_ℓ, it makes sense to compute its correlations with the original variables X_j first. This is done for $\ell = 1, \dots, k$ and for $j = 1, \dots, p$ to obtain the matrix P_{XF}. The following covariance between X and F is obtained via (12.5),

$$\sum_{XF} = \mathbf{E}\left\{(QF + U)F^\top\right\} = Q.$$

  The correlation is

$$P_{XF} = D^{-1/2}Q \tag{12.8}$$

  where $D = \operatorname{diag}(\sigma_{X_1X_1}, \dots, \sigma_{X_pX_p})$.

- Returning to the psychology example where X are the observed scores to p different intelligence tests (the WAIS data set in Exercise 7.19 provides an example).

12.1 The Orthogonal Factor Model

- We would expect a model with one factor to produce a factor that is positively correlated with all of the components in X. For this example the factor represents the overall level of intelligence of an individual.
- A model with two factors could produce a refinement in explaining the variations of the p scores. For example, the first factor could be the same as before (overall level of intelligence), whereas the second factor could be positively correlated with some of the tests, X_j, that are related to the individual's ability to think abstractly and negatively correlated with other tests, X_i, that are related to the individual's practical ability.
- The second factor would then concern a particular dimension of the intelligence stressing the distinctions between the "theoretical" and "partical" abilities of the individual.

12.1 The Orthogonal Factor Model

Invariance of Scale

- What happens if we change the scale of X to $Y = \mathcal{C}X$ with $\mathcal{C} = \text{diag}(c_1, \ldots, c_p)$?
- If the k-factor model (12.6) is true for X with $\mathcal{Q} = \mathcal{Q}_X, \Psi = \Psi_X$, then, since

$$\mathbf{Var}(Y) = \mathcal{C} \sum \mathcal{C}^\top = \mathcal{C}\mathcal{Q}_X\mathcal{Q}_X^\top\mathcal{C}^\top + \mathcal{C}\Psi_X\mathcal{C}^\top$$

  the same k-factor model is also true for Y with $\mathcal{Q}_Y = \mathcal{C}\mathcal{Q}_X$ and $\Psi_Y = \mathcal{C}\Psi_X\mathcal{C}^\top$.
- In many applications ,the search for the loadings $\mathcal{Q}$ and for the specific variance Ψ will be done by the decomposition of the correlation matrix of X rather than the covariance matrix $\sum$.

12.1 The Orthogonal Factor Model

$$Y = D^{-1/2}(X - \mu)$$

$$\mathcal{Q}_X = D^{1/2}\mathcal{Q}_Y$$

$$\Psi_X = D^{1/2}\Psi_Y D^{1/2}$$

It should be noted that although the factor analysis model (12.4) enjoys the scale invariance property, the actual estimated factors could be scale dependent.

12.1 The Orthogonal Factor Model

Non-uniqueness of Factor Loadings

- The factor loadings are not unique!
- Suppose that $\mathcal{G}$ is an orthogonal matrix. Then X in (12.4) can also be written as

$$X = (\mathcal{Q}\mathcal{G})(\mathcal{G}^{\top}\mathcal{F}) + U + \mu$$

  This implies that, if a k-factor of X with factors F and loadings $\mathcal{Q}$ is true, then the k-factor model with factors $\mathcal{G}^{\top}\mathcal{F}$ and loadings $\mathcal{Q}\mathcal{G}$ is also true. In practice, we will take advantage of this non-uniqueness.
- It will be shown that choosing an appropriate rotation will result in a matrix of loadings $\mathcal{Q}\mathcal{G}$ that will be easier to interpret.

12.1 The Orthogonal Factor Model

It makes sense to search for rotations that give factors that are maximally correlated with various groups of variables.

Usually, we impose additional constraints where

$$\mathcal{Q}^{\top}\Psi^{-1}\mathcal{Q} \text{ is diagonal.} \tag{12.11}$$

or

$$\mathcal{Q}^{\top}\mathcal{D}^{-1}\mathcal{Q} \text{ is diagonal.} \tag{12.12}$$

12.1 The Orthogonal Factor Model

How many parameters does the model (12.7) have without constraints?

$$\mathcal{Q}(p \times k) \text{ has } p \cdot k \text{ parameters, and}$$

$$\Psi(p \times p) \text{ has } p \text{ parameters.}$$

Hence we have to determine $pk + p$ parameters! Conditions (12.11) respectively (12.12) introduce $\frac{1}{2}\{k(k-1)\}$ constraints, since we require the matrices to be diagonal.

12.1 The Orthogonal Factor Model

Therefore, the degrees of freedom of a model with k factors is:

$d =$ (#parameters for $\sum$ unconstrained) - (#parameters for $\sum$ constrained)
$= \frac{1}{2}p(p+1) - (pk + p - \frac{1}{2}k(k-1))$
$= \frac{1}{2}(p-k)^2 - \frac{1}{2}(p+k)$.

12.1 The Orthogonal Factor Model

- If $d < 0$, then the model is undetermined: there are infinitely many solutions to (12.7). This means that the number of parameters of the factorial model is larger than the number of parameters of the original model, or that the number of factors k is "too large" relative to p.
- In some cases $d = 0$: there is a unique solution to the problem (except for rotation).
- In practice we usually have that $d > 0$: there are more equations than parameters, thus an exact solution does not exist. In this case approximate solutions are used. An approximation of $\sum$, for example, is $QQ^{\top} + \Psi$.

12.1 The Orthogonal Factor Model

Example 12.1 Let $p = 3$ and $k = 1$, then $d = 0$ and

$$\sum = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} = \begin{pmatrix} q_1^2 + \psi_{11} & q_1 q_2 & q_1 q_3 \\ q_1 q_2 & q_2^2 + \psi_{22} & q_2 q_3 \\ q_1 q_3 & q_2 q_3 & q_3^2 + \psi_{33} \end{pmatrix}$$

with $Q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}$ and $\Psi = \begin{pmatrix} \psi_{11} & 0 & 0 \\ 0 & \psi_{22} & 0 \\ 0 & 0 & \psi_{33} \end{pmatrix}$. Note that here the constraint (12.11) is automatically verified since $k = 1$.

12.1 The Orthogonal Factor Model

We have

$$q_1^2 = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}}; q_2^2 = \frac{\sigma_{12}\sigma_{23}}{\sigma_{13}}; q_3^2 = \frac{\sigma_{13}\sigma_{23}}{\sigma_{12}}$$

and

$$\psi_{11} = \sigma_{11} - q_1^2;\ \psi_{22} = \sigma_{22} - q_2^2;\ \psi_{33} = \sigma_{33} - q_3^2.$$

In this particular case ($k = 1$), the only rotation is defined by $\mathcal{G} = -1$, so the other solution for the loadings is provide by $-\mathcal{Q}$.

12.1 The Orthogonal Factor Model

Example 12.2 Suppose now $p = 2$ and $k = 1$, then $d < 0$ and

$$\sum = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \begin{pmatrix} q_1^2 + \psi_{11} & q_1 q_2 \\ q_1 q_2 & q_2^2 + \psi_{22} \end{pmatrix}.$$

We have infinitely many solutions: for any $\alpha (\rho < \alpha < 1)$, a solution is provided by

$$q_1 = \alpha;\ q_2 = \rho/\alpha;\ \psi_{11} = 1 - \alpha^2;\ \psi_{22} = 1 - (\rho/\alpha)^2.$$

The solution in Example 12.1 may be unique (up to a rotation), but it is not proper in the sense that it cannot be interpreted statistically.

Even in the case of a unique solution ($d = 0$), the solution may be inconsistent with statistical interpretations.

12.2 Estimation of the Factor Model

In practice, we have to find estimates $\hat{\mathcal{Q}}$ of the loadings $\mathcal{Q}$ and estimates $\hat{\Psi}$ of the specific variances Ψ such that analogously to (12.7)

$$\mathcal{S} = \hat{\mathcal{Q}}\hat{\mathcal{Q}}^T + \hat{\Psi}$$

where $\mathcal{S}$ denotes the empirical covariance of $\mathcal{X}$. Given an estimate $\hat{\mathcal{Q}}$ of $\mathcal{Q}$, it is natural to set

$$\hat{\psi}_{jj} = s_{X_j X_j} - \sum_{\ell=1}^{k} \hat{q}_{j\ell}^2$$

We have that $\hat{h}_j^2 = \sum_{\ell=1}^{k} \hat{q}_{j\ell}^2$ is an estimate for the communality h_j^2.

12.2 Estimation of the Factor Model

Define $\mathcal{Y} = \mathcal{H}\mathcal{X}\mathcal{D}^{-1/2}$, the standardization of the data matrix $\mathcal{X}$, where $\mathcal{D} = \mathrm{diag}(s_{X_1X_1}, \dots, s_{X_pX_p})$ and the centering matrix $\mathcal{H} = \mathcal{I} - n^{-1}l_n l_n^T$ (recall from Chap.2 that $\mathcal{S} = \frac{1}{n}\mathcal{X}^{\mathcal{T}}\mathcal{H}\mathcal{X}$). The estimated factor loading matrix $\hat{\mathcal{Q}}_Y$ and the estimated specific variance $\hat{\Psi}_Y$ of $\mathcal{Y}$ are

$$\hat{\mathcal{Q}}_Y = \mathcal{D}^{-1/2}\hat{\mathcal{Q}}_X \text{ and } \hat{\Psi}_Y = \mathcal{D}^{-1}\hat{\Psi}_X.$$

For the correlation matrix $\mathcal{R}$ of $\mathcal{X}$, we have that

$$\mathcal{R} = \hat{\mathcal{Q}}_Y\hat{\mathcal{Q}}_Y^T + \hat{\Psi}_Y.$$

12.2 Estimation of the Factor Model

The Maximum Likelihood Method

Recall from Chap.6 the log-likelihood function ℓ for a data matrix $\mathcal{X}$ of observations of $X \sim N_p(\mu, \sum)$:

$$\ell(\mathcal{X}; \hat{\mu}, \mathcal{Q}, \Psi) = -\frac{n}{2}\left[\log\left\{\left|2\pi(\mathcal{Q}\mathcal{Q}^\top + \Psi)\right|\right\} + tr\left\{(\mathcal{Q}\mathcal{Q}^\top + \Psi)^{-1}\mathcal{S}\right\}\right]. \quad (12.13)$$

where $\hat{\mu} = \bar{x}$.

12.2 Estimation of the Factor Model

Likelihood Ratio Test for the Number of Common Factors

- Using the methodology of Chap.7, it is easy to test the adequacy of the factor analysis model by comparing the likelihood under the null (factor analysis) and alternative (no constraints on covariance matrix) hypotheses.
- Assuming that $\hat{\mathcal{Q}}$ and $\hat{\Psi}$ are the maximum likelihood estimates corresponding to (12.13), we obtain the following LR test statistic:

$$-2\log\left(\frac{\text{maximized likelihood under } H_0}{\text{maximized likelihood}}\right) = n\log\left(\frac{\left|\hat{\mathcal{Q}}\hat{\mathcal{Q}}^T+\hat{\Psi}\right|}{|\mathcal{S}|}\right), \tag{12.14}$$

which asymptotically has the $\chi^2_{\frac{1}{2}\{(p-k)^2-p-k\}}$ distribution.

12.2 Estimation of the Factor Model

- The χ^2 approximation can be improved if we replace n by $n - 1 - (2p + 4k + 5)/6$ in (12.14) (Bartlett 1954).
- Using Bartlett's correction, we reject the factor analysis model at the α level if

$$\{n - 1 - (2p + 4k + 5)/6\} \log\left(\frac{|\hat{\mathcal{Q}}\hat{\mathcal{Q}}^T + \hat{\Psi}|}{|\mathcal{S}|}\right) > \chi^2_{1-\alpha;\{(p-k)^2-p-k\}/2} \tag{12.15}$$

12.2 Estimation of the Factor Model

The principal Component Method

- The *principal component method* involves finding an approximation $\tilde{\Psi}$ of Ψ, the matrix of specific variances, and then correcting $\mathcal{R}$, the correlation matrix of X , by $\tilde{\Psi}$.
- The *principal component method* starts with an approximation $\hat{\mathcal{Q}}$ of $\mathcal{Q}$, the factor loadings matrix. The sample covariance matrix is diagonalized, $\mathcal{S} = \Gamma\Lambda\Gamma^T$. The the first k eigenvectors are retained to build

$$\hat{\mathcal{Q}} = \left(\sqrt{\lambda_1}\gamma_1, \ldots, \sqrt{\lambda_k}\gamma_k\right). \tag{12.16}$$

12.2 Estimation of the Factor Model

The estimated specific variances are provided by the diagonal elements of the matrix $\mathcal{S} - \hat{\mathcal{Q}}\hat{\mathcal{Q}}^{\top}$,

$$\hat{\Psi} = \begin{pmatrix} \hat{\psi}_{11} & & & 0 \\ & \hat{\psi}_{22} & & \\ & & \ddots & \\ 0 & & & \hat{\psi}_{pp} \end{pmatrix} \text{ with } \hat{\psi}_{jj} = s_{X_j X_j} - \sum_{\ell=1}^{k} \hat{q}_{j\ell}^2 \qquad (12.17)$$

12.2 Estimation of the Factor Model

A heuristic device for selecting the number of factors is to consider the proportion of the total sample variance due to the jth factor. This quantity is in general equal to

A. $\lambda_j / \sum_{j=1}^{p} s_{jj}$ for a factor analysis of $\mathcal{S}$,

B. λ_j / p for a factor analysis of $\mathcal{R}$.

12.2 Estimation of the Factor Model

Example12.5 This example uses a consumer-preference study from Johnson and Wichern (1998). Customers were asked to rate several attributes of a new product. The responses were tabulated and the following correlation matrix $\mathcal{R}$ was constructed:

$$
\begin{array}{ll}
\textit{Attribute (Variable)} & \\
\text{Taste} & 1 \\
\text{Good buy for money} & 2 \\
\text{Flavor} & 3 \\
\text{Suitable for snack} & 4 \\
\text{Provides lots of energy} & 5
\end{array}
\begin{pmatrix}
1.00 & 0.02 & \mathbf{0.96} & 0.42 & 0.01 \\
0.02 & 1.00 & 0.13 & 0.71 & \mathbf{0.85} \\
0.96 & 0.13 & 1.00 & 0.50 & 0.11 \\
0.42 & 0.71 & 0.50 & 1.00 & \mathbf{0.79} \\
0.01 & 0.85 & 0.11 & 0.79 & 1.00
\end{pmatrix}
$$

12.2 Estimation of the Factor Model

- The bold entries of $\mathcal{R}$ show that variables 1 and 3 and variables 2 and 5 are highly correlated. Variable 4 is more correlated with variables 2 and 5 than with variables 1 and 3. Hence, a model with 2 (or 3) factors seems to be reasonable.
- The first two eigenvalues $\lambda_1 = 2.85$ and $\lambda_2 = 1.81$ of $\mathcal{R}$ are the only eigenvalues greater than one.

12.2 Estimation of the Factor Model

- Moreover, $k = 2$ common factors account for a cumulative proportion

$$\frac{\lambda_1 + \lambda_2}{p} = \frac{2.85 + 1.81}{5} = 0.93$$

of the total (standardized) sample variance.

- Using the principal component method, the estimated factor loadings, communalites, and specific variances, are calculated from formulas (12.16) and (12.17), and the results are given in Table 12.1.

12.2 Estimation of the Factor Model

Table 12.1 Estimated factor loadings, communalities, and specific variances

Variables	Estimated factor loadings		Communalities	Specific variances
	$\widehat{q}_1$	$\widehat{q}_2$	$\widehat{h}_j^2$	$\widehat{\psi}_{jj} = 1 - \widehat{h}_j^2$
1. Taste	0.56	0.82	0.98	0.02
2. Good buy for money	0.78	−0.53	0.88	0.12
3. Flavor	0.65	0.75	0.98	0.02
4. Suitable for snack	0.94	−0.11	0.89	0.11
5. Provides lots of energy	0.80	−0.54	0.93	0.07
Eigenvalues	2.85	1.81		
Cumulative proportion of total (standardized) sample variance	0.571	0.932		

12.2 Estimation of the Factor Model

Take a look at:

$$\hat{Q}\hat{Q}^{\mathsf{T}} + \hat{\Psi} = \begin{pmatrix} 0.56 & 0.82 \\ 0.78 & -0.53 \\ 0.65 & 0.75 \\ 0.94 & -0.11 \\ 0.80 & -0.54 \end{pmatrix} \begin{pmatrix} 0.56 & 0.78 & 0.65 & 0.94 & 0.80 \\ 0.82 & -0.53 & 0.75 & -0.11 & -0.54 \end{pmatrix}$$

$$+ \begin{pmatrix} 0.02 & 0 & 0 & 0 & 0 \\ 0 & 0.12 & 0 & 0 & 0 \\ 0 & 0 & 0.02 & 0 & 0 \\ 0 & 0 & 0 & 0.11 & 0 \\ 0 & 0 & 0 & 0 & 0.07 \end{pmatrix} = \begin{pmatrix} 1.00 & 0.01 & 0.97 & 0.44 & 0.00 \\ 0.01 & 1.00 & 0.11 & 0.79 & 0.91 \\ 0.97 & 0.11 & 1.00 & 0.53 & 0.11 \\ 0.44 & 0.79 & 0.53 & 1.00 & 0.81 \\ 0.00 & 0.91 & 0.11 & 0.81 & 1.00 \end{pmatrix}.$$

12.2 Estimation of the Factor Model

- This nearly reproduces the correlation matrix $\mathcal{R}$.
- We conclude that the two-factor model provides a good fit of the data. The communalites (0.98, 0.88, 0.98, 0.89, 0.93) indicate that the two factors account for a large percentage of the sample variance of each variable.
- Due to the nonuniqueness of loadings, the interpretation might be enhanced by rotation.

12.2 Estimation of the Factor Model

Rotation

- The interpretation of the loadings would be very simple if the variables could be split into disjoint sets, each being associated with one factor.
- A well known analytical algorithm to rotate the loadings is given by the *varimax rotation method* proposed by Kaiser (1985). In the simplest case of $k = 2$ factors, a rotations matrix $\mathcal{G}$ is given by

$$\mathcal{G}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

  representing a clockwise rotation of the coordinate axes by the angle θ.
- The corresponding rotation of loadings is calculated via $\hat{\mathcal{Q}}^* = \hat{\mathcal{Q}}\mathcal{G}(\theta)$. The idea of the *varimax method* is to find the angle θ that maximizes the sum of the variances of the squared loadings $\hat{q}^*_{ij}$ within each column of $\hat{\mathcal{Q}}^*$.

12.2 Estimation of the Factor Model

Example 12.6

- Let us return to the marketing example of Johnson and Wichern (1998) (Example 12.5).
- The basic factor loadings given in Table 12.1 of the first factor and a second factor are almost identical making it difficult to interpret the factors.
- Applying the varimax rotation we obtain the loadings $\tilde{q}_1 = (0.02, \mathbf{0.94}, 0.13, \mathbf{0.84}, \mathbf{0.97})^T$ and $\tilde{q}_2 = (\mathbf{0.99}, -0.01, \mathbf{0.98}, 0.43, -0.02)^T$. The high loadings, indicated as bold entries, show that variables 2, 4, 5 define factor 1, a nutricional factor. Variables 1 and 3 define factor 2 which might be referred to as a taste factor.

12.3 Factor Scores and Strategies

- The factors F were considered to be normalized random sources of information and were explicitely addressed as nonspecific (common factors).
- The estimated values of the factors, called the *factor scores*, may also be useful in the interpretation as well as in the diagnostic analysis.

12.3 Factor Scores and Strategies

- To be more precise, the factor scores are estimates of the unobserved random vectors F_l, $l = 1, \ldots, k$, for each individual x_i, $i = 1, \ldots, n$.(Johnson and Wichern, 1998) describe three methods which in practice yield very similar results.
- Here, we present the regression method which has the advantage of being the simplest technique and is easy to implement.

12.3 Factor Scores and Strategies

- The idea is to consider the joint distribution of $(X-\mu)$ and F, and then to proceed with the regression analysis presented in Chap.5.
- Under the factor model (12.4), the joint covariance matrix of $(X-\mu)$ and F is:

$$\mathbf{Var}\begin{pmatrix} X-\mu \\ F \end{pmatrix} = \begin{pmatrix} \mathcal{Q}\mathcal{Q}^{\top}+\Psi & \mathcal{Q} \\ \mathcal{Q}^{\top} & \mathcal{I}_k \end{pmatrix} \tag{12.18}$$

- Note that the upper left entry of this matrix equals $\sum$ and that the matrix has size $(p+k)\times(p+k)$.

12.3 Factor Scores and Strategies

Theorem 5.1

$$\mathbf{E}(F \mid X = x) = \mathcal{Q}^{\top}\Sigma^{-1}(X - \mu) \tag{12.19}$$

$$\mathbf{Var}(F \mid X = x) = \mathcal{I}_k - \mathcal{Q}^{\top}\Sigma^{-1}\mathcal{Q} \tag{12.20}$$

The estimated individual factor scores:

$$\hat{f}_i = \hat{\mathcal{Q}}^{\top}\mathcal{S}^{-1}(x_i - \bar{x}) \tag{12.21}$$

12.3 Factor Scores and Strategies

- The same rule can be followed when using $\mathcal{R}$ instead of $\mathcal{S}$. Then (12.18) remains valid when standardized variables, i.e., $Z = \mathcal{D}_{\Sigma}^{-1/2}(X - \mu)$, are considered if $\mathcal{D}_{\Sigma} = \text{diag}\big(\sigma_{11}, \ldots, \sigma_{pp}\big)$.
- In this case the factors are given by

$$\hat{f}_i = \hat{\mathcal{Q}}^{\top} \mathcal{R}^{-1} (z_i) \tag{12.22}$$

where $z_i = \mathcal{D}_S^{-1/2} (x_i - \bar{x})$, $\hat{\mathcal{Q}}$ is the loading obtained with the matrix $\mathcal{R}$, and $\mathcal{D}_S = \text{diag}\big(s_{11}, \ldots, s_{pp}\big)$.

- If the factors are rotated by the orthogonal matrix $\mathcal{G}$, the factor scores have to be rotated accordingly, that is

$$\hat{f}_i^* = \mathcal{G}^{\top} \hat{f}_i \tag{12.23}$$

12.3 Factor Scores and Strategies

Practical Suggestions No one method outperforms another in the practical implementation of factor analysis. However, by applying a *tâtonnement* process, the factor analysis view of the data can be stabilized. This motivates the following procedure.

1. Fix a reasonable number of factors, say $k = 2$ or 3, based on the correlation structure of the data and/or screeplot of eigenvalues.
2. Perform several of the presented methods, including rotation. Compare the loadings, communalities, and factor scores from the respective results.
3. If the results show significant deviations, check for outliers (based on factor scores), and consider changing the number of factors k.

For larger data sets, cross-validation methods are recommended.

12.3 Factor Scores and Strategies

Factor Analysis Versus PCA

- Factor analysis and principal component analysis use the same set of mathematical tools (spectral decomposition, projections, …).
- One could conclude, on first sight, that they share the same view and strategy and therefore yield very similar results. This is not true.

12.3 Factor Scores and Strategies

- The biggest difference between PCA and factor analysis comes from the model philosophy.
- Factor analysis imposes a strict structure of a fixed number of common (latent) factors whereas the PCA determines p factors in decreasing order of importance.
- The most important factor in PCA is the one that maximizes the projected variance.
- The most important factor in factor analysis is the one that (after rotation) gives the maximal interpretation.
- Often this is different from the direction of the first principal component.

12.3 Factor Scores and Strategies

- From an implementation point of view, the PCA is based on a well-defined, unique algorithm (spectral decomposition), whereas fitting a factor analysis model involves a variety of numerical procedures.
- The non-uniqueness of the factor analysis procedure opens the door for subjective interpretation and yields therefore a spectrum of results.
- This data analysis philosophy makes factor analysis difficult especially if the model specification involves cross-validation and a data-driven selection of the number of factors.

12.4 Boston Housing

- The variable X_4 (Charles River indicator) will be excluded. As before, standardized variables are used and the analysis is based on the correlation matrix.
- For illustration, the MLM will be presented with and without varimax rotation.
- Table 12.2 gives the MLM factor loadings without rotation and Table 12.3 gives the varimax version of this analysis.
- The corresponding graphical representations of the loadings are displayed in Figs. 12.2 and 12.3.

12.4 Boston Housing

Table 12.2 Estimated factor loadings, communalities, and specific variances, MLM MVAfacthous

	Estimated factor loadings			Communalities	Specific variances
	$\widehat{q}_1$	$\widehat{q}_2$	$\widehat{q}_3$	$\widehat{h}_j^2$	$\widehat{\psi}_{jj} = 1 - \widehat{h}_j^2$
1. Crime	0.9295	0.1653	0.1107	0.9036	0.0964
2. Large lots	−0.5823	0.0379	0.2902	0.4248	0.5752
3. Nonretail acres	0.8192	−0.0296	−0.1378	0.6909	0.3091
4. Nitric oxides	0.8789	0.0987	−0.2719	0.8561	0.1439
5. Rooms	−0.4447	0.5311	−0.0380	0.4812	0.5188
6. Prior 1940	0.7837	−0.0149	−0.3554	0.7406	0.2594
7. Empl. centers	−0.8294	−0.1570	0.4110	0.8816	0.1184
8. Accessibility	0.7955	0.3062	0.4053	0.8908	0.1092
9. Tax-rate	0.8262	0.1401	0.2906	0.7867	0.2133
10. Pupil/Teacher	0.5051	−0.1850	0.1553	0.3135	0.6865
11. African American	0.4701	−0.0227	−0.1627	0.2480	0.7520
12. Lower status	0.7601	−0.5059	−0.0070	0.8337	0.1663
13. Value	−0.6942	0.5904	−0.1798	0.8628	0.1371

12.4 Boston Housing

Table 12.3 Estimated factor loadings, communalities, and specific variances, MLM, varimax rotation MVAfacthous

	Estimated factor loadings			Communalities	Specific variances
	$\hat{q}_1$	$\hat{q}_2$	$\hat{q}_3$	$\hat{h}_j^2$	$\hat{\psi}_{jj} = 1 - \hat{h}_j^2$
1. Crime	0.7247	−0.2705	−0.5525	0.9036	0.0964
2. Large lots	−0.1570	0.2377	0.5858	0.4248	0.5752
3. Nonretail acres	0.4195	−0.3566	−0.6287	0.6909	0.3091
5. Nitric oxides	0.4141	−0.2468	−0.7896	0.8561	0.1439
6. Rooms	−0.0799	0.6691	0.1644	0.4812	0.5188
7. Prior 1940	0.2518	−0.2934	−0.7688	0.7406	0.2594
8. Empl. centers	−0.3164	0.1515	0.8709	0.8816	0.1184
9. Accessibility	0.8932	−0.1347	−0.2736	0.8908	0.1092
10. Tax-rate	0.7673	−0.2772	−0.3480	0.7867	0.2133
11. Pupil/Teacher	0.3405	−0.4065	−0.1800	0.3135	0.6865
12. African American	−0.3917	0.2483	0.1813	0.2480	0.7520
13. Lower status	0.2586	−0.7752	−0.4072	0.8337	0.1663
14. Value	−0.3043	0.8520	0.2111	0.8630	0.1370

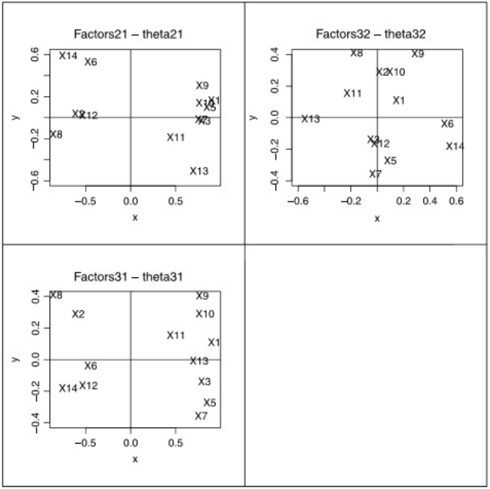
12.4 Boston Housing

- We can see that the varimax does not significantly change the interpretation of the factors obtained by the MLM.
- Factor 1 can be roughly interpreted as a "quality of life factor" because it is positively correlated with variables like X_1, X_9 and negatively correlated with X_8, both having low specific variances.
- The second factor may be interpreted as a "residential factor", since it is highly correlated with variables X_{13}, X_{14}, and X_6.

12.4 Boston Housing

- The most striking difference between the results with and without varimax rotation can be seen by comparing the lower left corners of Figs. 12.2 and 12.3.
- There is a clear separation of the variables in the varimax version of the MLM.
- Given this arrangement of the variables in Fig. 12.3, we can interpret factor 3 as an employment factor, since we observe high correlations with X_8 and X_5.

12.4 Boston Housing



12.4 Boston Housing

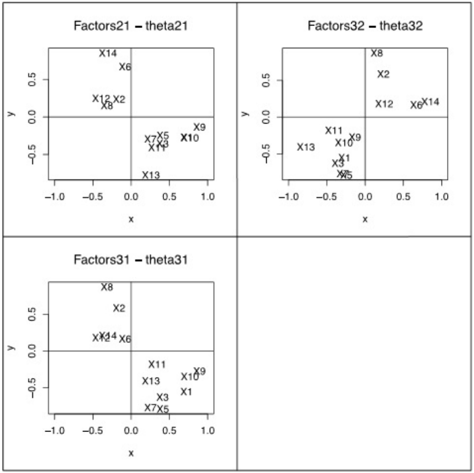


Fig. 12.3 Factor analysis for Boston housing data, MLM after varimax rotation MVAfacthous

12.4 Boston Housing

- We now turn to the PCM and PFM analyses. The results are presented in Tables 12.4 and 12.5 and in Figs. 12.4 and 12.5.
- We would like to focus on the PCM, because this 3-factor model yields only one specific variance (unexplained variation) above 0.5.

12.4 Boston Housing

- Looking at Fig. 12.4, it turns out that factor 1 remains a "quality of life factor" which is clearly visible form the clustering of X_5, X_3, X_{10} and X_1 on the right-hand side of the graph, while the variables X_8, X_2, X_{14}, X_{12} and X_6 are on the left-hand side.
- Again, the second factor is a "residential factor", clearly demonstrated by the location of variables X_6, X_{14}, X_{11}, and X_{13}.
- The interpretation of the third factor may be a "Accessibility". It is highly correlated with variables X_9 and X_{10}.

12.4 Boston Housing

Table 12.4 Estimated factor loadings, communalities, and specific variances, PCM, varimax rotation MVAfacthous

	Estimated factor loadings			Communalities	Specific variances
	$\widehat{q}_1$	$\widehat{q}_2$	$\widehat{q}_3$	$\widehat{h}_j^2$	$\widehat{\psi}_{jj} = 1 - \widehat{h}_j^2$
1. Crime	0.6034	−0.2456	0.6864	0.8955	0.1045
2. Large lots	−0.7722	0.2631	0.0270	0.6661	0.3339
3. Nonretail acres	0.7183	−0.3701	0.3449	0.7719	0.2281
5. Nitric oxides	0.7936	−0.2043	0.4250	0.8521	0.1479
6. Rooms	−0.1601	0.8585	0.0218	0.7632	0.2368
7. Prior 1940	0.7895	−0.2375	0.2670	0.7510	0.2490
8. Empl. centers	−0.8562	0.1318	−0.3240	0.8554	0.1446
9. Accessibility	0.3681	−0.1268	0.8012	0.7935	0.2065
10. Tax-rate	0.3744	−0.2604	0.7825	0.8203	0.1797
11. Pupil/Teacher	0.1982	−0.5124	0.3372	0.4155	0.5845
12. African American	0.1647	0.0368	−0.7002	0.5188	0.4812
13. Lower status	0.4141	−0.7564	0.2781	0.8209	0.1791
14. Value	−0.2111	0.8131	−0.3671	0.8394	0.1606

12.4 Boston Housing

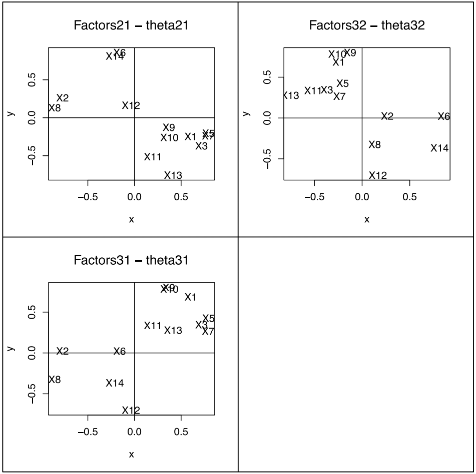


Fig. 12.4 Factor analysis for Boston housing data, PCM after varimax rotation MVAfacthous

12.4 Boston Housing

- Use equations(12.22) and (12.23) to calculate the factor scores for MLM varimax analysis.
- Which districts have with highest scores and lowest scores in "quality of life factor" ? What are the scores? They are districts 415 and 284 with "quality of life factor" scores 2.1 and -2.07 respectively.
- Which districts have with highest scores and lowest scores in "residential factor" ? What are the scores? They are districts 371 and 491 with "residential factor" scores 3.53 and -2.47 respectively.