

Chapter 10

Factor Analysis



10.1 Factor Analysis

Factor analysis (Chattopadhyay and Chattopadhyay 2014) is a statistical method used to study the dimensionality of a set of variables. In factor analysis, latent variables represent unobserved constructs and are referred to as factors or dimensions. Factor analysis attempts to identify underlying variables or factors that explain the pattern of correlations within a set of observed variables. Factor analysis is often used in data reduction to identify a small number of factors that explain most of the variance that is observed in a much larger number of manifest variables.

Suppose the observable random vector X with p components has mean vector μ and covariance matrix Σ . In the factor model, we assume that X is linearly dependent upon a few unobservable random variables $F_1, F_2 \dots F_p$ called common factors and p additional sources of variation $\epsilon_1, \epsilon_2, \dots, \epsilon_p$ called the errors (or specific factors). Then the factor model is

$$X = \mu + L F + \epsilon \tag{10.1.1}$$

$$\begin{aligned} X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \epsilon_1 \\ X_2 - \mu_1 &= l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \epsilon_2 \\ &\vdots \\ X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \epsilon_p \end{aligned}$$

The coefficients l_{ij} s are called the loading of the i th variable on the j th factor so the matrix L is the matrix of factor loadings. Here ϵ_i is associated only with the i th response X_i . Here the p deviations $X_1 - \mu_1 \dots X_p - \mu_p$ are expressed in terms of

A significant part of ‘Chattopadhyay and Chattopadhyay (2014). Statistical Methods for Astronomical Data Analysis, Springer Series in Astrostatistics, Springer Science+Business Media New York’ is reproduced in this chapter.

$p + m$ random variables $F_1, F_2, \dots, F_m, \epsilon_1, \dots, \epsilon_p$ which are unobservable (but in multivariate regression independent variables can be observed).

With same additional assumption on the random vectors F and ϵ , the model w implies certain covariance relationships which can be checked.

We assume that

$$E(P) = 0^{m \times 1} \quad cov(F) = E(FP') = I^{m \times m}$$

$$E(\epsilon) = 0^{p \times 1} \quad cov(\epsilon) = E(\epsilon\epsilon') = \psi = \begin{pmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ 0 & 0 & \dots & \psi_p \end{pmatrix}$$

$$\text{and } cov(\epsilon, F) = E(\epsilon, F) = 0^{p \times m} \quad (10.1.2)$$

The model $X - \mu = LF + \epsilon$ is linear in the common factors. If the p response of X are related to the underlying in factors in a nonlinear form [$X_1 - \mu_1 = F_1 F_3 + \epsilon_1$] Then the covariance structure $LL' + \psi$ may not be adequate. The assumption of linearity is inherent here.

These assumption and the relation (10.1.1) constitute the orthogonal factor model.

The orthogonal factor model implies a covariance structure for X .

$$\begin{aligned} \text{Here } (X - \mu)(X - \mu)' &= (LF + \epsilon)(LF + \epsilon)' \\ &= (LF + \epsilon)((LF)' + \epsilon') \\ &= LF(LF)' + \epsilon(LF)' + LF\epsilon' + \epsilon\epsilon' \\ &= LFF'L' + \epsilon F'L' + LF\epsilon' + \epsilon\epsilon' \end{aligned}$$

$$\begin{aligned} \Sigma &= \text{covariance matrix of } X \\ &= E(X - \mu)(X - \mu)' \\ &= LE(FF')L' + E(\epsilon F')L' + LE(F\epsilon') + E(\epsilon\epsilon') \\ &= LIL' + \psi = LL' + \psi \end{aligned}$$

$$\text{Again } (X - \mu)F' = (LF + \epsilon)F' = LFF' + \epsilon F'$$

$$\text{or, } cov(X, F) = E(X - \mu)F' = E(LF + \epsilon)F' = LE(FF') + E(\epsilon F') = L$$

Now $\Sigma = LL' + \psi$ implies

$$\left. \begin{aligned} \text{var}(X_i) &= l_{i1}^2 + \dots + l_{im}^2 + \psi_i \\ \text{cov}(X_i X_k) &= l_{i1}l_{k1} + \dots + l_{im}l_{km} \end{aligned} \right\} \quad (10.1.3)$$

$$cov(XF) = L \Rightarrow cov(X_i F_j) = l_{ij}$$

$$\Rightarrow V(X_i) = \delta_{ii} = l_{i1}^2 + \dots + l_{im}^2 + \psi_i$$

Let i th communality = $h_i^2 = l_{i1}^2 + \dots + l_{im}^2$

Then $\delta_{ii} = h_i^2 + \psi_i$ ($i = 1 \dots p$)

h_i^2 = sum of squares of loadings of i th variable on the m common factors.

Given a random sample of observations $x_1^{b \times 1}, x_2 \dots x_p^{p \times 1}$. The basic problem is to decide whether Σ can be expressed in the form (10.1.3) for reasonably small value of m , and to estimate the elements of L and ψ .

Here the estimation procedure is not so easy. Primarily, we have from the sample data estimates of the $\frac{p(p+1)}{2}$ distinct elements of the upper triangle of Σ but on the RHS of (10.1.3) we have $pm + p$ parameters, pm for L and p for ψ . The solution will be indeterminate unless $\frac{p(p+1)}{2} - p(m + 1) \geq 0$ or $p > 2m$. Even if this condition is satisfied L is not unique.

Proof Let $T^{m \times m}$ be any \perp matrix so that $TT' = T'T = I$

Then (10.1.1) can be written as

$$X - \mu = LF + \epsilon = LTT'F + \epsilon = L^*F^* + \epsilon \tag{10.1.4}$$

where $L^* = LT$ and $F^* = T'F$

Since $E(F^*) = T'E(F) = 0$

and $\text{cov}(F^*) = T'\text{Cov}(F)T = T'T = I$

It is impossible to distinguish between loadings L and L^* on the basis of the observations on X . So the vectors F and $F^* = T'F$ have the same statistical properties and even if the loadings L and L^* are different, they both generate the same covariance matrix Σ , i.e.,

$$\Sigma = LL' + \psi = LTT'L' + \psi = L^*L^{*'} + \psi \tag{10.1.5}$$

The above problem of uniqueness is generally resolved by choosing an orthogonal rotation T such that the final loading L satisfies the condition that $L'\psi^{-1}L$ is diagonal with positive diagonal elements. This restriction requires L to be of full rank m . With a valid ψ viz. one with all positive diagonal elements it can be shown that the above restriction yields a unique L . □

10.1.1 Method of Estimation

Given n observations vectors $x_1 \dots x_n$ on p generally correlated variables, factor analysis seeks to verify whether the factor model (10.1.1) with a small number of factors adequately represent the data.

The sample covariance matrix is an estimator of the unknown covariance matrix Σ . If Σ appears to deviate significantly from a diagonal matrix, then a factor model can be used and the initial problem is one of estimating the factor loadings l_{ij} and the specific variances. ψ_i .

Principal Component Method¹

Let Σ has eigenvalue–eigenvector pairs (λ_i, e_i) with $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$. Then by specified decomposition

$$\begin{aligned} \Sigma &= \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \dots + \lambda_p e_p e_p' \\ &= (\sqrt{\lambda_1} e_1 \dots \sqrt{\lambda_p} e_p) \begin{pmatrix} \sqrt{\lambda_1} e_1' \\ \vdots \\ \sqrt{\lambda_p} e_p' \end{pmatrix} e_1 \sqrt{\lambda_1} \dots e_p \sqrt{\lambda_p} \quad (10.1.6) \\ &= \begin{matrix} p \times p & p \times p \\ L & L' \end{matrix} + 0^{p \times p} \end{aligned}$$

[in (10.1.6) $m = p$ and j th column of $L = \sqrt{\lambda_j} e_j$].

Apart from the scale factor $\sqrt{\lambda_j}$, the factor loadings on the j th factor are the pp^n j th principal component.

The approximate representation assumes that the specific factors ϵ are of minor importance and can be ignored in factoring Σ . If specific factors are included in the model, their variances may be taken to be the diagonal elements of $\Sigma - LL'$.

Allowing for specific factors, the approximation becomes

$$\begin{aligned} \Sigma &= LL' + \psi \\ &= (\sqrt{\lambda_1} e_1 \sqrt{\lambda_2} e_2 \dots \sqrt{\lambda_m} e_m) \begin{pmatrix} \sqrt{\lambda_1} e_1' \\ \sqrt{\lambda_2} e_2' \\ \vdots \\ \sqrt{\lambda_m} e_m' \end{pmatrix} + \begin{pmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{pmatrix} \quad (10.1.7) \end{aligned}$$

where $m \leq p$.

(we assume that last $p - m$ eigenvalues are small)

and $\psi_{ii} = \delta_{ii} - \sum_{j=1}^m l_{ij}^2$ for $i = 1 \dots p$.

¹A significant part of 'Chattopadhyay and Chattopadhyay (2014). Statistical methods for Astronomical Data Analysis, Springer Series in Astrostatistics, Springer' is reproduced in this part.

For the principal component solution, the estimated factor loadings for a given factor do not change as the number of factors is increased. If $m = 1$

$$L = (\sqrt{\widehat{\lambda}_1} \widehat{e}_1)$$

if $m = 2$

$$L = (\sqrt{\widehat{\lambda}_1} \widehat{e}_1, \sqrt{\widehat{\lambda}_2} \widehat{e}_2)$$

where $(\widehat{\lambda}_1, \widehat{e}_1)$ and $(\widehat{\lambda}_2, \widehat{e}_2)$ are the first two eigenvalue–eigenvector pairs for S (or R).

By the definition of $\widehat{\psi}_i$, the diagonal elements of S are equal to the diagonal elements of $\widehat{L}\widehat{L}' + \psi$. How to determine m ?

The choice of m can be based on the estimated eigenvalues.

Consider the residual matrix $S - (LL' + \psi)$

Here the diagonal elements are new and if the off-diagonal elements are also small we may take that particular value of m to be appropriate.

Analytically, we chose that m for which

$$\text{Sum of squared entries of } (S - (LL' + \psi)) \leq \widehat{\lambda}_{m+1}^2 + \dots + \widehat{\lambda}_p^2 \quad (10.1.8)$$

Ideally, the contribution of the first few factors to the sample variance of the variables should be large. The contribution to the sample variance s_{ii} from the first common factor is l_{ii}^2 . The contribution to the total sample variance $s_{11} + \dots + s_{pp} = h(S)$ from the first common factor is

$$\widehat{l}_{11}^2 + \widehat{l}_{21}^2 + \dots + \widehat{l}_{p1}^2 = (\sqrt{\widehat{\lambda}_1} \widehat{e}_1)' (\sqrt{\widehat{\lambda}_1} \widehat{e}_1) = \widehat{\lambda}_1$$

Since the eigenvectors \widehat{e}_1 has unit length.

In general,

$$\left(\begin{array}{c} \text{Proportion of total} \\ \text{sample variance due} \\ \text{to the factor} \end{array} \right) = \left\{ \begin{array}{l} \frac{\widehat{\lambda}_j}{s_{11} + \dots + s_{pp}} \text{ for a factor analysis of } S \\ \frac{\widehat{\lambda}_j}{p} \text{ for a factor analysis of } R \end{array} \right. \quad (10.1.9)$$

Criterion (10.1.9) is frequently used as a heuristic device for determining the appropriate number of common factors. The value of m is gradually increased until a suitable proportion of the total sample variance has been explained.

Other Rules Used in Package

No. of eigenvalue of R greater than one (when R is used)

No. of eigenvalue of S that are positive (when S is used).

10.1.2 Factor Rotation

If \widehat{L} be the $p \times m$ matrix of estimated factor loadings obtained by any method, then

$$L^* = \widehat{L}T \text{ where } TT' = T'T = I$$

is a $p \times m$ matrix of **rotated loadings**.

Moreover, the estimated covariance (or correlation) matrix remains unchanged since

$$\widehat{L}\widehat{L}' + \widehat{\psi} = \widehat{L}TT'\widehat{L}' + \widehat{\psi} = \widehat{L}^*\widehat{L}^{*'} + \widehat{\psi}$$

The above equation indicates that the residual matrix $S_n - \widehat{L}\widehat{L}' - \widehat{\psi} = S_n - \widehat{L}^*\widehat{L}^{*'} - \widehat{\psi}$ remains unchanged. Moreover, the specific variances $\widehat{\psi}_i$ and hence the communication \widehat{h}_i^2 are unaltered. Hence, mathematically it is immaterial whether \widehat{L} or L^* is obtained.

Since the original loadings may not be readily interpretable, it is usual practice to rotate them until a 'simple structure' is achieved.

Ideally, we should like to see a pattern of loadings of each variable loads highly on a single factor and has small to moderate loading on the remaining factors.

The problem is to find an **orthogonal rotation** which compounds to a 'simple structure.'

There can be achieved if often rotation the orthogonality of the factor still exists. This is maintained if we perform orthogonal rotation. Among these (1) Variance rotation, (2) Quartimax rotation, (3) Equamax rotation are important.

Oblique rotation does not ensure the orthogonality of factors often rotation. There are several algorithms like oblimax, Quartimax.

10.1.3 Varimax Rotation

Orthogonal Transformation on L

$$L^* = LT \quad TT' = I$$

L^* is the matrix of orthogonally rotated loadings and let $d_j = \sum_{i=1}^p l_{ij}^{*2}$ $j = 1 \dots m$

Then the following expression is maximized

$$\sum_{j=1}^m \left\{ \sum_{i=1}^p (l_{ij}^{*4} - d_j^2/p) \right\}$$

Such a procedure tries to give either large (in absolute value) or zero values in the columns of \mathbf{L}^* . Hence, the procedure tries to produce factors with either a stray association with the responses or no association at all.

The communality

$$h_i^2 = \sum_{j=1}^m l_{ij}^{*2} = \sum_{j=1}^m l_{ij}^2 \text{ remains constant under rotation.}$$

10.2 Quartimax Rotation

The factor pattern is simplified by forcing the variables to correlate highly with one main factor (the so-called G-factor of 1Q studies) and very little with remaining factors. Here all variables are primarily associated with a single factor.

Interpretation of results obtained from factor analysis. It is usually difficult to interpret. Many users should significant coefficient magnitudes on many of the retained factors (coefficient greater than $-.60$ — are often considered large and coefficients of $-.35$ — are often considered moderate). And especially on the first factor.

For good interpretation, factor rotation is necessary. The objective of the rotation is to achieve the most ‘simple structure’ though the manipulation of factor pattern matrix.

The most simple structure can be explained in terms of five principles of factor rotation.

1. Each variable should have at least one zero (small) loadings.
2. Each factor should have a set of linearly independent variables where factor loadings are zero (small).
3. For every pair of factors, there should be several variables where loadings are zero (small) for one factor but not the other.
4. For every pair of factors, a large proportion of variables should have zero (small) loading on both factors whenever more than about four factors are extracted.
5. For every pair of factors, there should only be a small number of variables with nonzero loadings on both.

In orthogonal rotation,

- (1) Factors are perfectly uncorrelated with one another.
- (2) Less parameters are to be estimated.

10.3 Promax Rotation

Factors are allowed to be correlated with one another.

Step I. Rotate the factors orthogonally.

Step II. Get a target matrix by raising the factor coefficients to an exponent (3 or 4). The coefficients secure smaller but absolute distance increases.

Step III. Rotate the original matrix to a best-fit position with the target matrix.

Here many moderate coefficients quickly approaches zero [$.3 \times .3 = .09$] then the large coefficients ($\geq .6$).

Example 10.1 Consider the data set related to the relative consumption of certain food items in European and Scandinavian countries considered in the chapter of principal component analysis.

If we do factor analysis with varimax rotation, then the output is as follows:

Rotated Factor Loadings and Communalities Varimax Rotation

Variable	Factor1	Factor2	Factor3	Factor4	Communality
coffee	0.336	0.807	0.018	-0.095	0.774
Tea	-0.233	0.752	0.330	0.370	0.866
Biscuits	0.502	0.124	0.712	-0.177	0.806
Powder	0.317	0.856	0.047	-0.230	0.889
Potatoes	0.595	0.047	0.060	0.485	0.595
Frozen fish	0.118	-0.100	0.050	0.918	0.869
Apples	0.832	0.284	0.251	0.097	0.846
Oranges	0.903	0.148	0.004	0.036	0.839
Butter	-0.004	0.089	0.900	0.172	0.847
Variance	2.3961	2.0886	1.4969	1.3480	7.3296
% Var	0.266	0.232	0.166	0.150	0.814

Factor Score Coefficients

Variable	Factor1	Factor2	Factor3	Factor4
coffee	0.038	0.408	-0.144	-0.040
Tea	-0.311	0.456	0.119	0.319
Biscuits	0.165	-0.141	0.506	-0.252
Powder	0.026	0.426	-0.109	-0.144
Potatoes	0.253	-0.047	-0.089	0.331
Frozen fish	0.006	-0.019	-0.072	0.692
Apples	0.339	-0.008	0.045	0.006
Oranges	0.431	-0.064	-0.129	-0.026
Butter	-0.132	-0.080	0.674	0.029

We see that according to percentage of variation about 80% variation is explained by first four components (as in case of PCA). But here the advantage is unlike PCA we can physically explain the factors. According to rotated factor loading, we can say that the first factor is composed of ‘apples, oranges, and potatoes,’ similarly the other three factors are composed of ‘coffee, tea, and powder soup,’ ‘butter and biscuits,’ and ‘potatoes and frozen fish,’ respectively.

Except Potato there is no overlapping, and the groups are well defined and may correspond to types of customers preferring ‘fruits,’ ‘hot drinks,’ ‘snacks’ and ‘proteins, vitamins, and minerals.’

The most significant difference between PCA and factor analysis is regarding the assumption of an underlying causal structure. Factor analysis assumes that the covariation among the observed variables is due to the presence of one or more latent variables known as factors that impose causal influence on these observed variables. Factor analysis is used when there exist some latent factors which impose causal influence on the observed variables under consideration. Exploratory factor analysis helps the researcher identify the number and nature of these latent factors. But principal component analysis makes no assumption about an underlying causal relation. It is simply a dimension reduction technique.

References and Suggested Readings

- Albazzas, H., & Wang, X. Z. (2004). *Industrial & Engineering Chemistry Research*, 43(21), 6731.
- Babu, J., et al. (2009). *The Astrophysical Journal*, 700, 1768.
- Chattopadhyay, A. K., & Chattopadhyay, T. (2014). *Statistical methods for astronomical data analysis.*, Springer series in astrostatistics New York: Springer.
- Chattopadhyay, A. K., Chattopadhyay, T., Davoust, E., Mondal, S., & Sharina, M. (2009). *The Astrophysical Journal*, 705, 1533.
- Chattopadhyay, A. K., Mondal, S., & Chattopadhyay, T. (2013). *Computational Statistics & Data Analysis*, 57, 17.
- Comon, P. (1994). *Signal Processing*, 36, 287.
- Dickens, R. J. (1972). *Monthly Notices of Royal Astronomical Society*, 157, 281.
- Fusi Pecci, F., et al. (1993). *Astronomical Journal*, 105, 1145.
- Gabriel, K. R. (1971). *Biometrika*, 5, 453.
- Hastie, T., & Tibshirani, R. (2003). In S. Becker, & K. Obermayer (Eds.). *Independent component analysis through product density estimation in advances in neural information processing system* (Vol. 15, pp. 649–656). Cambridge, MA: MIT Press.
- Hyvarinen, A., & Oja, E. (2000). *Neural Networks*, 13(4–5), 411.