Factor Analysis

1 Introduction

In factor analysis the observable random vector $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$ is explained in terms minimal number of unobservable(latent) random variables, called factors. In this approach each component X_i is examined to see if it could be generated by linear combination of unobservable random variables, called common factors and a single variable, called specific factor variate.
The common factor will generate the covariance structure of X where specific factor will account for for the variance of component X_i.

Some example
In a survey of household consumption, the consumption levels of p different goods during 1 month could be observed. The *variability* present in the data might be explained by the two or three main social behaviour factors of the household. The latent factors could be basic desire for comfort or willingness to achive certain level or family income level etc. The unobserved factors are much more interesting to social scientist than observed quantiative measures themshelves because they give better idea of social behaviour of the households.
In a school examination pupils obtain marks in different subjects like english, bengali, history, geography, science and mathematics. The variability in marks among pupils can be explained by factors like creativity, interest in science and other unobservable factors.
Factor analysis was developed by Spearman for analysis of scores of mental

tests.

2 The Orthogonal Factor Model

Let $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$ be an observable random vector with $E(\mathbf{X}) = \mu$ and dispersion matrix $D(\mathbf{X}) = \Sigma = (\sigma_{ij})$, a positive definite matrix. Assuming each component X_i can be generated a linear combination of $m(m < p)$ unobservable variables, the factor analysis model can be written as

$$\begin{aligned} X_1 &= \mu_1 + l_{11}f_1 + l_{12}f_2 + \ldots + l_{1m}f_m + U_1 \\ X_2 &= \mu_2 + l_{21}f_1 + l_{22}f_2 + \ldots + l_{2m}f_m + U_2 \\ &\vdots \\ X_p &= \mu_p + l_{p1}f_1 + l_{p2}f_2 + \ldots + l_{pm}f_m + U_p \\ \text{or} & \\ \mathbf{X} &= \mu + \mathbf{Lf} + \mathbf{U} \end{aligned}$$

where $\mathbf{f} = (f_1, f_2, \ldots, f_m)'$. The vector $\mathbf{U} = (U_1, U_2, \ldots, U_p)'$ denotes the error vector (or vector of specific factors) and $\mathbf{L} = (l_{ij})$ is a $p \times m$ matrix of unknown coeffcients l_{ij} which is called *factor loading matrix*. The elements of $\mathbf{f}$ is called common factors.

It is assumed that $\mathbf{U}$ is distributed independently of $\mathbf{f}$ and with mean $E(\mathbf{U}) = 0$ and dispersion matrix $D(\mathbf{U}) = \Psi$ where Ψ is a diagonal matrix with diagonal elements $\psi_1, \psi_2, \ldots, \psi_p$; $var(U_i) = \psi_i$ is called specific factor variance of X_i.

In a test each component of X is a score in a test. The corresponding component μ is the average score of this test in the population. The elements of $\mathbf{f}$ is called common factors because they are common in several different tests. The component $\mathbf{U}$ is a part of test score not "explained" by the common factors. This is considered as the made up of error in measurement plus a specific factor.

The specification of a given component of $\mathbf{X}$ is similar to that in regression theory. Here, **f** plays the independent variable, is unobservable.

If **f** is a random vector we shall assume that

$$E(\mathbf{f}) = 0 \quad \text{and} \quad D(\mathbf{f}) = E(\mathbf{ff}') = \mathbf{I},$$

the identity matrix. As **U** and **f** are independently distributed we have $E(\mathbf{Uf}') = \mathbf{0}$. These assumptions and relations constitute *orthogonal factor model.*

So,orthogonal factor model is stated as

Orthogonal factor model with m common factors

$$\mathbf{X}_{p\times 1} = \mu_{p\times 1} + \mathbf{L}_{p\times m}\mathbf{f}_{m\times 1} + \mathbf{U}_{p\times 1}$$

$$\begin{aligned} \mu_i &= \text{mean of variable i} \\ u_i &= \text{ith specfic factor} \\ f_j &= \text{jth common factor} \\ l_{ij} &= \text{loading of ith variable in jth factor} \end{aligned}$$

The unobservable random vectors **f** and **U** satisfy the following conditions

f and **U** are independently distributed.

$E(\mathbf{f}) = 0$ and $D(\mathbf{f}) = I$

$E(\mathbf{U}) = 0$ and $D(\mathbf{U}) = \Psi$ where Ψ is a diagonal matrix

The orthogonal factor model implies the covariance structure for **X**. From the model,

$$\begin{aligned} (\mathbf{X}-\mu)(\mathbf{X}-\mu)' &= (\mathbf{Lf}+\mathbf{U})(\mathbf{Lf}+\mathbf{U})' \\ &= (\mathbf{Lf}+\mathbf{U})\left((\mathbf{Lf})'+\mathbf{U}'\right) \\ &= \mathbf{Lff}'\mathbf{L}' + \mathbf{UU}' + \mathbf{LfU}' + \mathbf{Uf}'\mathbf{L}' \end{aligned}$$

Hence the variance covariance matrix of $\mathbf{X}$ is

$$\begin{aligned}\mathbf{\Sigma} = D(\mathbf{X}) &= E(\mathbf{X}-\mu)(\mathbf{X}-\mu)' \\ &= \mathbf{L}E\left(\mathbf{ff}'\right)\mathbf{L}' + E\left(\mathbf{UU}'\right) + \mathbf{L}E\left(\mathbf{fU}'\right) + E\left(\mathbf{Uf}'\right)\mathbf{L}' \\ &= \mathbf{LL}' + \Psi\end{aligned}$$

Also, $(\mathbf{X}-\mu)\,\mathbf{f}' = (\mathbf{Lf}+\mathbf{U})\,\mathbf{f}'$ and hence

$$\text{Cov}(\mathbf{X},\mathbf{f}) = \mathbf{L}E\left(\mathbf{ff}'\right) + E\left(\mathbf{Uf}'\right) = \mathbf{L}$$

The covariance structure of $\mathbf{X}$ is given by
1. $\Sigma = D(\mathbf{X}) = \mathbf{LL}' + \Psi$
or

$$\begin{aligned}Var(X_i) &= l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2 + \psi_i \\ \text{cov}(X_i, X_k) &= l_{i1}l_{k1} + \ldots + l_{im}l_{km}\end{aligned}$$

2. $\text{cov}(X_i, f_j) = l_{ij}$

The portion of ith variable contributed in m common factors is called i^{th} communality. So,

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

is the communality and

$$Var(X_i) = \sigma_{ii} = h_i^2 + \psi_i \qquad i = 1, 2, \ldots, p$$

The purpose of factor analysis is the determination of L with the elements of Ψ such that

$$\Sigma - \Psi = LL'.$$

If the errors are small enough to be ignored we can take $\Sigma = \mathbf{LL}'$.

3 Estimation of factor loadings

1. Principal Component method

Let Σ have the eigen value-eigen vector pairs (λ_i, α_i) with $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p \geq 0$. Then

$$\begin{aligned}
\Sigma &= \lambda_1 \alpha_1 \alpha_1' + \lambda_2 \alpha_2 \alpha_2' + \ldots + \lambda_p \alpha_p \alpha_p' \\
&= \left(\sqrt{\lambda_1}\alpha_1 : \quad \sqrt{\lambda_2}\alpha_2 : \quad \cdots \quad : \sqrt{\lambda_p}\alpha_p\right) \begin{pmatrix} \sqrt{\lambda_1}\alpha_1 \\ \sqrt{\lambda_2}\alpha_2 \\ \vdots \\ \sqrt{\lambda_p}\alpha_p \end{pmatrix}
\end{aligned}$$

This fits the prescribed covariance structure for the factor analysis having as many factors as many variables(m = p) and $\psi_i = 0$ for all i.
If for $\mathbf{L}^{p \times m}$,

$$\begin{aligned}
\Sigma &= \mathbf{L}\mathbf{L}' + \Psi \\
&= \left(\sqrt{\lambda_1}\alpha_1 : \quad \sqrt{\lambda_2}\alpha_2 : \quad \cdots \quad : \sqrt{\lambda_m}\alpha_m\right) \begin{pmatrix} \sqrt{\lambda_1}\alpha_1 \\ \sqrt{\lambda_2}\alpha_2 \\ \vdots \\ \sqrt{\lambda_m}\alpha_m \end{pmatrix} + \begin{pmatrix} \psi_1 & 0 & \ldots & 0 \\ 0 & \psi_2 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & \psi_p \end{pmatrix}
\end{aligned}$$

where $\psi_i = \sigma_{ii} - l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2$ for $i = 1, 2, \ldots, p$.

For given data $\mathbf{x}^\alpha = (x_1^\alpha, x_2^\alpha, \ldots, x_p^\alpha)$ of size N, the vector population mean μ can be estimated by sample mean vector $\bar{\mathbf{x}}$, then we have centered observation

$$\begin{pmatrix} x_1^\alpha - \bar{x}_1 \\ x_2^\alpha - \bar{x}_2 \\ \vdots \\ x_p^\alpha - \bar{x}_p \end{pmatrix}$$

Let $\mathbf{S}$ is the sample variance covariance matrix. The eigen value-eigen vector pairs of $\mathbf{S}$ are $(\hat{\lambda}_1, \hat{\alpha}_1), (\hat{\lambda}_2, \hat{\alpha}_2), \cdots, (\hat{\lambda}_p, \hat{\alpha}_p)$ with $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \hat{\lambda}_p \geq 0$. Let $m < p$ be the number of common factors. Then the matrix of estimated factor loadings (l_{ij}) is given by

$$\tilde{\mathbf{L}} = \left(\sqrt{\hat{\lambda}_1}\hat{\alpha}_1 : \quad \sqrt{\hat{\lambda}_2}\hat{\alpha}_2 : \quad \cdots \quad : \sqrt{\hat{\lambda}_m}\hat{\alpha}_m\right)$$

The estimated specific variances are provided by the diagonal elements of the matrix $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$, so

$$\tilde{\Psi} = \begin{pmatrix} \tilde{\psi}_1 & 0 & \dots & 0 \\ 0 & \tilde{\psi}_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \tilde{\psi}_p \end{pmatrix} \qquad \text{with} \quad \tilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \tilde{l}_{ij}^2$$

Communalities are estimated as

$$\tilde{h}_i^2 = \tilde{l}_{i1}^2 + \tilde{l}_{i2}^2 + \ldots + \tilde{l}_{im}^2$$

We factor loadings l_{ij} can also be estimated from sample correlation matrix by principal component method.

The contribution to the sample variance $s_{ii} = Var(X_i)$ from the first common factor is $\tilde{\lambda}_{i1}^2$. The contribution to the total sample variance,$s_{11} + s_{22} + \cdots + s_{pp} = tr(S)$, from the first factor is then

$$\tilde{l}_{11}^2 + \tilde{l}_{21}^2 + \cdots + \tilde{l}_{p1}^2 = \left(\sqrt{\hat{\lambda}_1}\hat{\alpha}_1\right)' \left(\sqrt{\hat{\lambda}_1}\hat{\alpha}_1\right) = \hat{\lambda}_1$$

since the eigen vector α_1 has a unit length.

$$\text{proprtion of total variance due to } j^{th} \text{ factor} = \begin{cases} \frac{\hat{\lambda}_j}{s_{11}+s_{22}+\ldots+s_{pp}} & \text{for factor analysis of } \mathbf{S} \\ \frac{\hat{\lambda}_j}{p} & \text{for factor analysis of } \mathbf{R} \end{cases}$$

Example In a consumer preference study, a random sample of customers were asked to rate several attributes of a new product. The responses on a 7 point were tabulated for 5 variables “Taste”, “Good buy for mone”, “Flavour”,“Suitable for snack”, “Provides lots of energy” and the attribute correlation matrix is constructed. The correlation matrix is

$$\begin{bmatrix} 1 & 0.02 & 0.96 & 0.42 & 0.01 \\ 0.02 & 1.00 & 0.13 & 0.71 & 0.85 \\ 0.96 & 0.13 & 1.00 & 0.50 & 0.11 \\ 0.42 & 0.71 & 0.50 & 1.00 & 0.79 \\ 0.01 & 0.85 & 0.11 & 0.79 & 1.00 \end{bmatrix}$$

Find the factor loading matrix with number of common factors 2 and calculate the communality for each variable.