Factor Analysis

Analisis Multivariat Week 4

Factor Analysis

Factor Analysis is a method for modeling observed variables, and their covariance structure, in terms of a smaller number of underlying unobservable (latent) "factors." The factors typically are viewed as broad concepts or ideas that may describe an observed phenomenon. For example, a basic desire of obtaining a certain social level might explain most consumption behavior. These unobserved factors are more interesting to the social scientist than the observed quantitative measurements.

Factor analysis is generally an exploratory/descriptive method that requires many subjective judgments. It is a widely used tool and often controversial because the models, methods, and subjectivity are so flexible that debates about interpretations can occur.

The method is similar to principal components although, as the textbook points out, factor analysis is more elaborate. In one sense, factor analysis is an inversion of principal components. In factor analysis, we model the observed variables as linear functions of the "factors." In principal components, we create new variables that are linear combinations of the observed variables. In both PCA and FA, the dimension of the data is reduced. Recall that in PCA, the interpretation of the principal components is often not very clean. A particular variable may, on occasion, contribute significantly to more than one of the components. Ideally, we like each variable to contribute significantly to only one component. A technique called factor rotation is employed toward that goal. Examples of fields where factor analysis is involved include physiology, health, intelligence, sociology, and sometimes ecology among others.

Collect all of the variables *X*'s into a vector $\mathbf{X}$ for each individual subject. Let $\mathbf{X_i}$ denote observable trait *i*. These are the data from each subject and are collected into a vector of traits.

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = \text{vector of traits}$$

This is a random vector, with a population mean. Assume that vector of traits $\mathbf{X}$ is sampled from a population with population mean vector:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \text{population mean vector}$$

Here, $\mathrm{E}(X_i) = \mu_i$ denotes the population mean of variable *i*.

Consider *m* unobservable common factors $f_1, f_2, \dots, f_m$. The i^{th} common factor is f_i. Generally, *m* is going to be substantially less than *p*.

The common factors are also collected into a vector,

$$\mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix} = \text{vector of common factors}$$

Our factor model can be thought of as a series of multiple regressions, predicting each of the observable variables X_i from the values of the unobservable common factors f_i :

$$\begin{aligned} X_1 &= \mu_1 + l_{11}f_1 + l_{12}f_2 + \cdots + l_{1m}f_m + \epsilon_1 \\ X_2 &= \mu_2 + l_{21}f_1 + l_{22}f_2 + \cdots + l_{2m}f_m + \epsilon_2 \\ &\vdots \\ X_p &= \mu_p + l_{p1}f_1 + l_{p2}f_2 + \cdots + l_{pm}f_m + \epsilon_p \end{aligned}$$

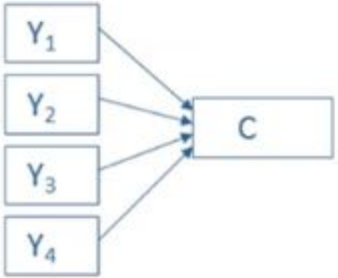
Here, the variable means μ_1 through μ_p can be regarded as the intercept terms for the multiple regression models.

The regression coefficients l_{ij} (the partial slopes) for all of these multiple regressions are called factor loadings. Here, l_{ij} = *loading* of the i^{th} variable on the j^{th} factor. These are collected into a matrix as shown here:

$$\mathbf{L} = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \vdots & \vdots & & \vdots \\ l_{p1} & l_{p2} & \dots & l_{pm} \end{pmatrix} = \text{matrix of factor loadings}$$

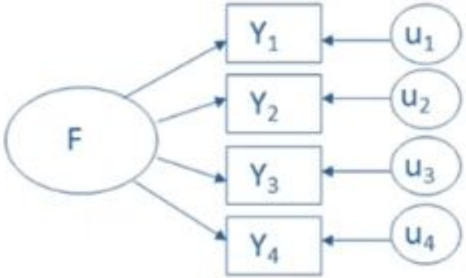
$$\boxed{\mathbf{X} = \boldsymbol{\mu} + \mathbf{L}\mathbf{f} + \boldsymbol{\epsilon}}$$

Principal Component Analysis



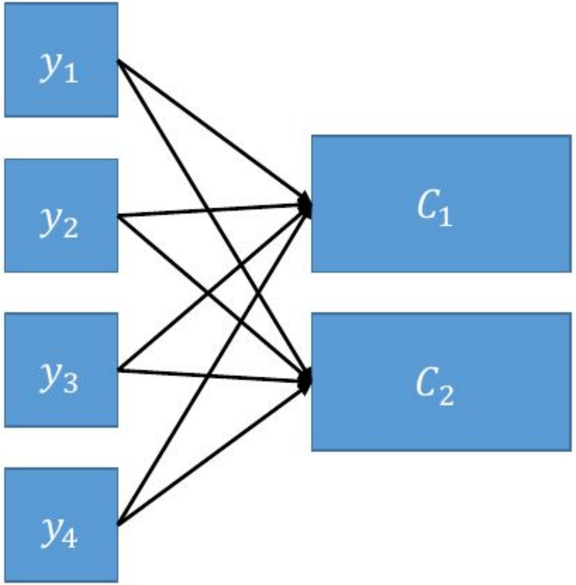
$$C = w_1(Y_1) + w_2(Y_2) + w_3(Y_3) + w_4(X_4)$$

Factor Analysis



$$Y_1 = b_1{*}F + u_1$$
$$Y_2 = b_2{*}F + u_2$$
$$Y_3 = b_3{*}F + u_3$$
$$Y_4 = b_4{*}F + u_4$$

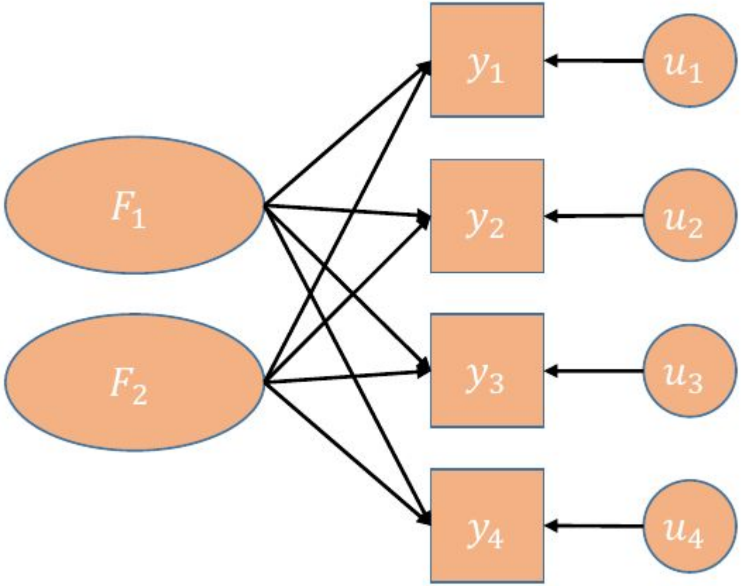
Principal Components Analysis (PCA)



$$C_1 = \lambda_{11} \cdot y_1 + \lambda_{12} \cdot y_2 + \lambda_{13} \cdot y_3 + \lambda_{14} \cdot y_4$$

$$C_2 = \lambda_{21} \cdot y_1 + \lambda_{22} \cdot y_2 + \lambda_{23} \cdot y_3 + \lambda_{24} \cdot y_4$$

Exploratory Factor Analysis (EFA)



$$y_1 = \lambda_{11} \cdot F_1 + \lambda_{21} \cdot F_2 + u_1$$
$$y_1 = \lambda_{12} \cdot F_1 + \lambda_{22} \cdot F_2 + u_2$$
$$y_1 = \lambda_{13} \cdot F_1 + \lambda_{23} \cdot F_2 + u_3$$
$$y_1 = \lambda_{14} \cdot F_1 + \lambda_{24} \cdot F_2 + u_4$$

Principal Component Method

Let X_i be a vector of observations for the i^{th} subject:

$$\mathbf{X_i} = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{pmatrix}$$

$\mathbf{S}$ denotes our sample variance-covariance matrix and is expressed as:

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{X_i} - \bar{\mathbf{x}})(\mathbf{X_i} - \bar{\mathbf{x}})'$$

We have *p* eigenvalues for this variance-covariance matrix as well as corresponding eigenvectors for this matrix.

Eigenvalues of $\mathbf{S}$:

$$\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p$$

Eigenvectors of $\mathbf{S}$:

$$\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \dots, \hat{\mathbf{e}}_p$$

Recall that the variance-covariance matrix can be re-expressed in the following form as a function of the eigenvalues and the eigenvectors:

Spectral Decomposition of Σ

$$\Sigma = \sum_{i=1}^{p} \lambda_i \mathbf{e_i e_i'} \cong \sum_{i=1}^{m} \lambda_i \mathbf{e_i e_i'} = \begin{pmatrix} \sqrt{\lambda_1}\mathbf{e_1} & \sqrt{\lambda_2}\mathbf{e_2} & \dots & \sqrt{\lambda_m}\mathbf{e_m} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1}\mathbf{e_1'} \\ \sqrt{\lambda_2}\mathbf{e_2'} \\ \vdots \\ \sqrt{\lambda_m}\mathbf{e_m'} \end{pmatrix} = \mathbf{LL'}$$

The idea behind the principal component method is to approximate this expression. Instead of summing from 1 to *p*, we now sum from 1 to *m*, ignoring the last *p* - *m* terms in the sum, and obtain the third expression. We can rewrite this as shown in the fourth expression, which is used to define the matrix of factor loadings $\mathbf{L}$, yielding the final expression in matrix notation.

Note! If standardized measurements are used, we replace $\mathbf{S}$ with the sample correlation matrix $\mathbf{R}$.

This yields the following estimator for the factor loadings:

$$\hat{l}_{ij} = \hat{e}_{ji}\sqrt{\hat{\lambda}_j}$$

This yields the following estimator for the factor loadings:

$$\hat{l}_{ij} = \hat{e}_{ji}\sqrt{\hat{\lambda}_j}$$

This forms the matrix $\mathbf{L}$ of factor loading in the factor analysis. This is followed by the transpose of $\mathbf{L}$. To estimate the specific variances, recall that our factor model for the variance-covariance matrix is

$$\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$$

in matrix notation. Ψ is now going to be equal to the variance-covariance matrix minus $\mathbf{LL}'$.

$$\boldsymbol{\Psi} = \boldsymbol{\Sigma} - \mathbf{LL}'$$

This in turn suggests that the specific variances, the diagonal elements of Ψ, are estimated with this expression:

$$\hat{\Psi}_i = s_i^2 - \sum_{j=1}^{m} \lambda_j \hat{e}_{ji}^2$$

We take the sample variance for the *i*th variable and subtract the sum of the squared factor loadings (i.e., the commonality).

Maximum Likelihood Estimation requires that the data are sampled from a multivariate normal distribution. This is a drawback of this method. Data is often collected on a Likert scale, especially in the social sciences. Because a Likert scale is discrete and bounded, these data cannot be normally distributed.

Using the Maximum Likelihood Estimation Method, we must assume that the data are independently sampled from a multivariate normal distribution with mean vector μ and variance-covariance matrix of the form:

$$\mathbf{\Sigma} = \mathbf{LL'} + \mathbf{\Psi}$$

where $\mathbf{L}$ is the matrix of factor loadings and Ψ is the diagonal matrix of specific variances.

We define additional notation: As usual, the data vectors for *n* subjects are represented as shown:

$$\mathbf{X_1}, \mathbf{X_2}, \dots, \mathbf{X_n}$$

Maximum likelihood estimation involves estimating the mean, the matrix of factor loadings, and the specific variance.

The maximum likelihood estimator for the mean vector μ, the factor loadings $\mathbf{L}$, and the specific variances Ψ are obtained by finding $\hat{\mu}$, $\hat{\mathbf{L}}$, and $\hat{\mathbf{\Psi}}$ that maximize the log-likelihood given by the following expression:

$$l(\mu, \mathbf{L}, \mathbf{\Psi}) = -\frac{np}{2}\log 2\pi - \frac{n}{2}\log|\mathbf{LL'} + \mathbf{\Psi}| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{X_i} - \mu)'(\mathbf{LL'} + \mathbf{\Psi})^{-1}(\mathbf{X_i} - \mu)$$

The log of the joint probability distribution of the data is maximized. We want to find the values of the parameters, (μ, $\mathbf{L}$, and Ψ), that are most compatible with what we see in the data. As was noted earlier the solutions for these factor models are not unique. Equivalent models can be obtained by rotation. If $\mathbf{L'\Psi^{-1}L}$ is a diagonal matrix, then we may obtain a unique solution.

Computationally this process is complex. In general, there is no closed-form solution to this maximization problem so iterative methods are applied. Implementation of iterative methods can run into problems as we will see later.

Varimax Rotation

Varimax rotation is the most common. It involves scaling the loadings by dividing them by the corresponding communality as shown below:

$$\tilde{l}^*_{ij} = \hat{l}^*_{ij}/\hat{h}_i$$

Varimax rotation finds the rotation that maximizes this quantity. The Varimax procedure, as defined below, selects the rotation in order to maximize

$$V = \frac{1}{p}\sum_{j=1}^{m}\left\{\sum_{i=1}^{p}(\tilde{l}^*_{ij})^4 - \frac{1}{p}\left(\sum_{i=1}^{p}(\tilde{l}^*_{ij})^2\right)^2\right\}$$

Let's Code

Maternal Mortality di Jawa Barat Tahun 2006

Keterangan:
X1 : Persentase sarana kesehatan
X2 : Persentase pengeluaran biaya kesehatan per kapita sebulan
X3 : Persentase wanita yang menikah dibawah umur
X4 : Persentase penolong proses kelahiran tenaga non medis
X5 : Persentase penduduk miskin
X6 : Persentase wanita dengan pendidikan tertinggi SD
X7 : Persentase tenaga medis dan paramedis

Tugas

1. 2 orang. Cari data min 10 variabel harus kuantitatif (yang kualitatif dihapus)
2. Asumsi PCA dan FA
3. PCA dan FA
4. Format Word:

   https://docs.google.com/document/d/1aXLaoad-ty2P0Isi3yR3RiEjmQmWFZviTt9qv_Vcuy0/edit?usp=sharing