Factor Analysis

Alexandros Korolis

July 2025

Introduction

What is factor analysis?

Factor analysis is a statistical method that aims to create new latent variables (factors) from the existing ones in a dataset, based on a somewhat subjective interpretation of their underlying relationships. By using these factors, we can reduce the dimensionality of the dataset and better explain the correlations among the original variables. Factor analysis differs from principal component analysis in that it aims to explain the underlying structure of the data rather than merely accounting for its variability. In factor analysis, there is not always a unique solution. Different estimation methods, or even the same method applied in slightly different ways, can lead to different results. Moreover, the extracted factors can have varying interpretations, which may even contradict each other.

Orthogonal Factor Model

In the orthogonal factor model, we assume that there are some underlying common factors, initially unknown,that we aim to estimate and that are responsible for the correlations among the observed variables. In this manner, if there are pp observed variables, each can be expressed as a linear combination of kk underlying factors, such that:

$$X - \mu = LF + \epsilon \ (1)$$

where: X : a vector of observed variables, with dimension px1
μ: a vector of means of each variable, with dimension px1
L : a matrix with dimension pxk, where L_{ij} is the loading of factor F_j in variable X_i
F : a kx1 matrix consisting of factors
ϵ : an error. ϵ_i represents the portion of the i-th observed variable that cannot be explained by the factors.

Assume that each variable has a mean of zero, and that k<p then :

$$\begin{gathered} X_1 = L_{11}F_1 + L_{12}F_2 + ... + L_{1k}F_k + \epsilon_1 \\ X_2 = L_{21}F_1 + L_{22}F_2 + ... + L_{2k}F_k + \epsilon_2 \\ ...... \\ X_p = L_{p1}F_1 + L_{p2}F_2 + ... + L_{pk}F_k + \epsilon_1 \end{gathered}$$

Note that the factors can also be written as a linear combination of the observed variables, which is particularly useful when creating new variables. In this context, the coefficients are called factor score coefficients, which are distinct from the factor loadings.

Assumptions of Orthogonal Factor Model

1. E(F)=0, each factor has a mean value of zero
2. Cov(F) = I, there is no correlation between each factor (Orthogonal Factors)
3. E(ϵ) = 0, each error term has a mean value of zero
4. Cov(ϵ = Ψ,

where $\Psi = \begin{pmatrix} \Psi_1 & 0 & .. & 0 \\ 0 & \Psi_2 & & \\ 0 & .. & .. & 0 \\ 0 & .. & .. & \Psi_p \end{pmatrix}$,

meaning that there is no correlation between the error terms.
5. $cov(\epsilon_i, F_j) = 0$ for i different than j, error terms and factor are unrelated.
6. If method of estimation is maximum likelihood, the we assume that the data follow a multivariate normal distribution.

Communality-Specificity

Based on the assumption of the orthogonal model, observe that: $\Sigma = cov(X) = cov(LF + \epsilon) = Lcov(F)L^T + cov(\epsilon) = LL^T + \Psi$

This means that the variance-covariance matrix consists of two components: the first corresponds to the common factors and is referred to as communality, while the second represents the portion that the model cannot explain and is known as specificity.

Loading refers to the correlation between each observed variable and each underlying factor.

Factor Analysis Steps

1.Examine the adequacy of correlations among variables in order to apply factor analysis.

2. Determine the appropriate number of factors and estimate the model parameters.

3. Examine the rotation method to enhance interpretability.

4. Estimate the score coefficients for further analysis.