FACTOR ANALYSIS

Introduction

- The purpose of *factor analysis* is to describe the variation among many variables in terms of a few underlying but unobservable random variables called factors
- All the covariance or correlations are explained by the *common factors*
- Any portion of the variance unexplained by the common factors is assigned to residual errors terms which are called *unique factors*

Terminology

- Common Factor: 共同因素
- Factor loading: 因素負荷量
- Explanatory factor analysis: 探索性因素分析
- Confirmatory factor analysis: 驗證性因素分析

Concept

- Factor analysis can be viewed as a statistical procedure for grouping variables into subsets such that the variables with each set are mutually highly correlated, whereas at the same time variables in different subsets are relatively uncorrelated.

Model

- Common factor analysis is composed of 3 sets of variables
  - A set of p observed variables $X_1, X_2, \ldots, X_p$ with mean vector μ ($p\times1$) and covariance matrix $\boldsymbol{\Sigma}$ ($p\times p$)
  - A set of γ unobserved variables called common factor $F_1, F_2, \ldots, F_\gamma$ where γ □ p
  - A set of p unique but unobserved factors $U_1, U_2, \ldots, U_p$

$$(X_1 - \mu_1) = a_{11}F_1 + a_{12}F_2 + \ldots + a_{1r}F_r + U_1$$

$$(X_2 - \mu_2) = a_{21}F_1 + a_{22}F_2 + \ldots + a_{2r}F_r + U_2$$

$$\vdots$$

$$(X_p - \mu_p) = a_{p1}F_1 + a_{p2}F_2 + \ldots + a_{pr}F_r + U_p$$

Or

$$\boldsymbol{(X - \mu) = \mathbf{A}\mathbf{f} + U}$$

Notation

- $\boldsymbol{(X-\mu)}$ is the ($p\times1$) vector of elements ($X_i-\mu_i$), $i=1,2,\ldots,p$;
- $\boldsymbol{f}$ is the $\gamma\times1$ vector of linearly independent common factors, $F_{j,}$ $i=1,2,\ldots,\gamma$
- $\boldsymbol{A}$ is the $p\times\gamma$ factor pattern matrix (consisting of the unknown factor loadings) a_{ij}, $i=1,2,\ldots,p$; $i=1,2,\ldots,\gamma$; and
- $\boldsymbol{U}$ is the $p\times1$ vector of unique factors U_i, $i=1,2,\ldots,p$

$F_1, F_2, \ldots F_\gamma$ are common to all p X variables,

U_i is unique to X_i

Assumptions:

$$E[f] = 0$$

$$E[ff^{'}] = I, \gamma \times \gamma$$ identity matrix

$$E[u] = 0;$$

$$E[uu^{'}] = \Psi, p \times p$$, diagonal matrix with diagonal elements $\sigma_{u_i}^2$

$$E[uf'] = 0$$, no correlation between unique factors and common factors

Factor structure matrix: $Cov(x, f) = A$

If the **X** variables are standardized, the elements of **A** represent correlations between the **X** variables and the factors.

The variance of each X_i can be written as

$$\sigma_i^2 = \sum_{j=1}^{\gamma} a_{ij}^2 + \sigma_{u_i}^2$$

$\sum_{j=1}^{\gamma} a_{ij}^2$ is the variance explained by the common factors, and is usually called *communality*.

$\sigma_{u_i}^2$ is usually called *unique variance* or *specific variance.*

Estimation of the Factor Model Using the Principal Components

- Given an observed data matrix $\mathbf{X}$ $(n \times p)$, the factor model can be expressed as

$$\mathbf{X} = \mathbf{F}\mathbf{A}' + \mathbf{U}$$

where

$\mathbf{F}$ $(n \times \gamma)$ is the unobserved matrix of vales of the γ common factors for the n observational units;

$\mathbf{A}'$ is the $(\gamma \times p)$ unknown factor pattern or loading matrix; and

$\mathbf{U}$ is the $(n \times p)$ matrix of unobserved errors or values of unique factors for the n observational units.

Eigenvalues

- Consider equation $\mathbf{Au}=\lambda\mathbf{u}$,
  where $\mathbf{u}$ is a vector and λ is a scalar
- Question: under what conditions (other than $\mathbf{u}=0$) can u and λ exist so that the equation is true?
- Rewrite $(\mathbf{A}-\lambda\mathbf{I})\mathbf{u}=0$. If $(\mathbf{A}-\lambda\mathbf{I})$ is non-singular, the only solution is $\mathbf{u}=0$. But if it's singular, a non-null solution for u can be obtain by
  $\mathbf{u}=[(\mathbf{A}-\lambda\mathbf{I})^{-} (\mathbf{A}-\lambda\mathbf{I}) -\mathbf{I}]\mathbf{z}$, $\mathbf{z}$ is arbitrary.
- So we can use $|\mathbf{A}-\lambda\mathbf{I}|=0$ to find the solutions, this is called *characteristic equation* of $\mathbf{A}$.The roots $\lambda_1\ \lambda_2\ldots\ \lambda_n$ are called *latent roots*, *characteristic roots, eigenvalues*…

Example of Calculating Eignvalues

$$\mathbf{A} = \begin{bmatrix} -3 & 2\sqrt{2} \\ 2\sqrt{2} & -1 \end{bmatrix}$$

$$\mathbf{A} - \lambda \mathbf{I} = \begin{bmatrix} -3 & 2\sqrt{2} \\ 2\sqrt{2} & -1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$|\mathbf{A} - \lambda \mathbf{I}| = (-3-\lambda)(-1-\lambda) - (2\sqrt{2})^2$$

$$= \lambda^2 + 4\lambda - 5$$

$$= (\lambda+5)(\lambda-1)$$

$\lambda = -5$, $\lambda = 1$ are eigenvalues of $\mathbf{A}$

Calculating Eigenvectors

- Calculating an eigenvector corresponding to λ_k requires finding a non-null $\mathbf{v}$ to satisfy $\mathbf{A}\mathbf{v} = \lambda_k \mathbf{v}$, equivalent to solving $(\mathbf{A} - \lambda_k \mathbf{I})\mathbf{v} = 0$
- From the example, we have

$$\begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} \mathbf{v} = -5\ \mathbf{v}$$

Solve the equation, we will get $\mathbf{v}$.

Estimating Factors

- We use the relation for correlation matrix $\rho = \mathbf{A}\mathbf{A}' + \mathbf{\Psi}$
- Based on principle component, we can write $\mathbf{X}=\mathbf{Z}\mathbf{V}'$, where $\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{X}'\mathbf{X}$.
- Then, $\mathbf{X}=(\mathbf{Z}\mathbf{\Lambda}^{-1/2})(\mathbf{\Lambda}^{1/2}\mathbf{V}')$, let

$$\tilde{F} = (\mathbf{Z}\mathbf{\Lambda}^{-1/2}), \text{ and } \tilde{A} = (\mathbf{\Lambda}^{1/2}\mathbf{V}')$$

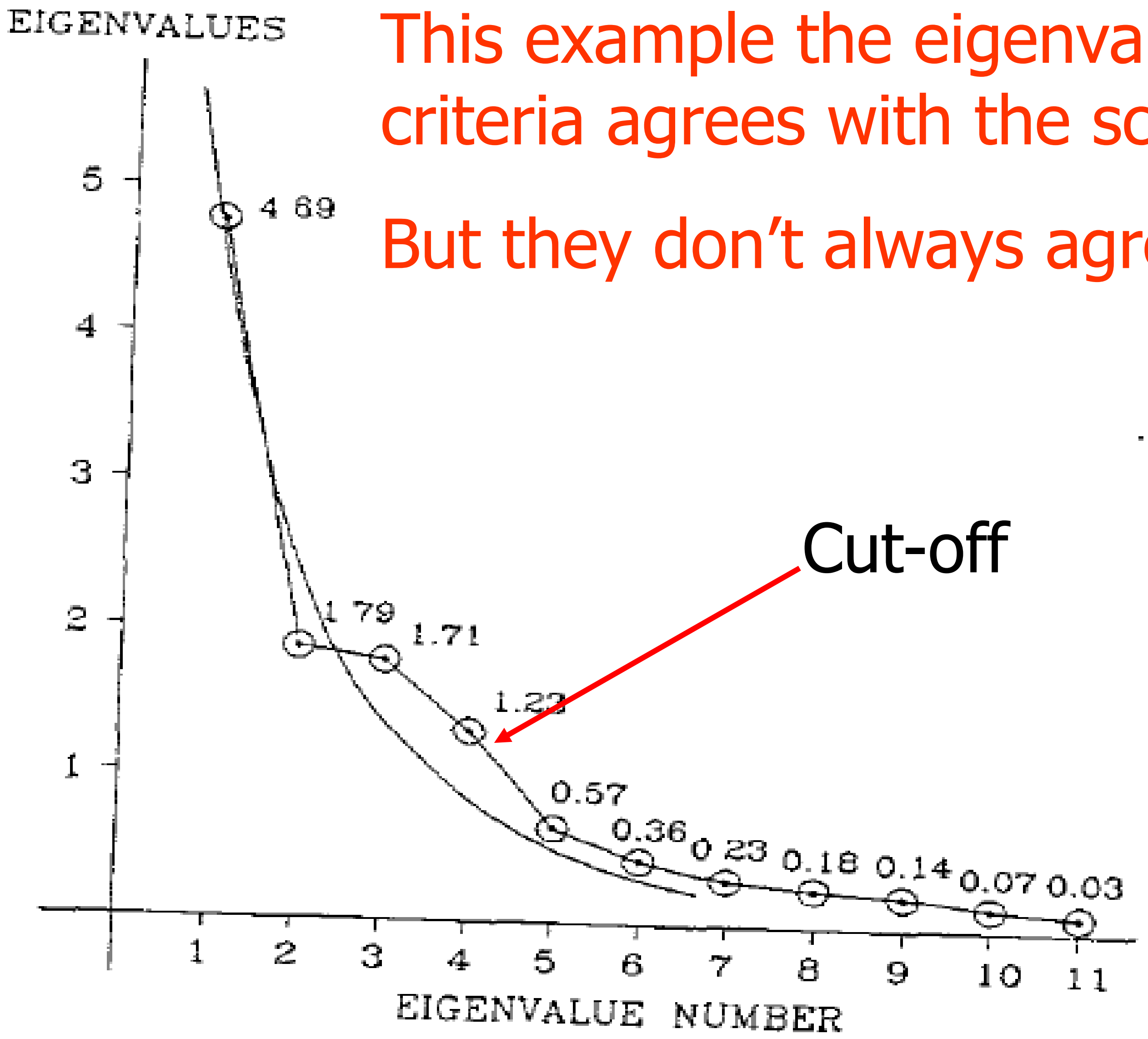
Since factors are of smaller dimension than the observed values, we can partition $\mathbf{Z}$ into r and p-r components.

Determining the Number of Factors

- Eigenvalues exceed 1: the eigenvalue of 1 is the arithmetic mean of the eigenvalues of a correlation matrix. It's also the variance of each of the $\mathbf{X}$ variables. Hence the eigenvalue-one-criterion suggests a factor re retained if it explains at least as much as a single variable.
- The test for zero correlation: if the correlation matrix is diagonal, there are no common factors
- Scree test: plot eigenvalues vs. eigenvalue number. Typically, this shape of a *scree graph* consists of two parts, a rapidly downward sloping followed by a second part which is almost horizontal. (example: 4 factors)

This example the eigenvalue-one-criteria agrees with the scree plot.

But they don’t always agree.



Factor Rotation

- Since there is no unique solution the the factor analysis, rotation can be used to obtain factors that are easily interpretable.
  - Orthogonal transformation: rigid
  - Varimax: the most commonly used method of rotation. It maximizes the squared ratio of each factor loading to communality of $\mathbf{X}$
  - Oblique: permits a minor amount of correlation among factors. However, there is no single popular method of this type of rotations. It requires considerable expertise.

Constructing factors

- Observed variables should fall into mutually exclusive categories in such a way that the variables in a given category exhibit loadings that are high on the same single factor, moderate to low on a very few factors and negligible on the remaining factors.
- Some use the criteria of factor loading greater than 0.3 or 0.4.