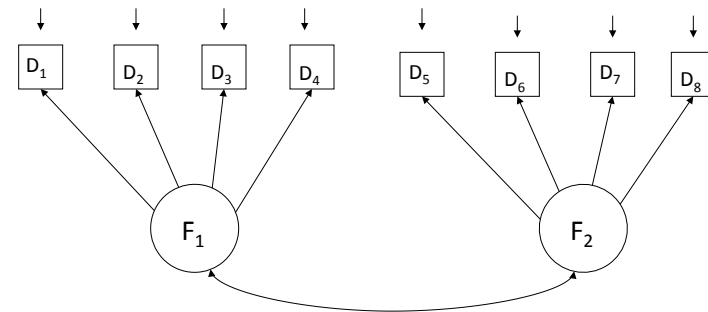


Factor Analysis: The Measurement Model

$$\mathbf{D}_i = \mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i$$



Exploratory Factor Analysis

STA431: Spring 2013

See last slide for copyright information

Example with 2 factors and 8 observed variables

$$\mathbf{D}_i = \mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i$$

$$\begin{bmatrix} D_{i,1} \\ D_{i,2} \\ D_{i,3} \\ D_{i,4} \\ D_{i,5} \\ D_{i,6} \\ D_{i,7} \\ D_{i,8} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \\ \lambda_{71} & \lambda_{72} \\ \lambda_{81} & \lambda_{82} \end{bmatrix} \begin{bmatrix} F_{i,1} \\ F_{i,2} \end{bmatrix} + \begin{bmatrix} e_{i,1} \\ e_{i,2} \\ e_{i,3} \\ e_{i,4} \\ e_{i,5} \\ e_{i,6} \\ e_{i,7} \\ e_{i,8} \end{bmatrix}$$

$$D_{i,1} = \lambda_{11}F_{i,1} + \lambda_{12}F_{i,2} + e_{i,1}$$

$$D_{i,2} = \lambda_{21}F_{i,1} + \lambda_{22}F_{i,2} + e_{i,2} \text{ etc.}$$

The lambda values are called **factor loadings**.

Terminology

$$D_{i,1} = \lambda_{11}F_{i,1} + \lambda_{12}F_{i,2} + e_{i,1}$$

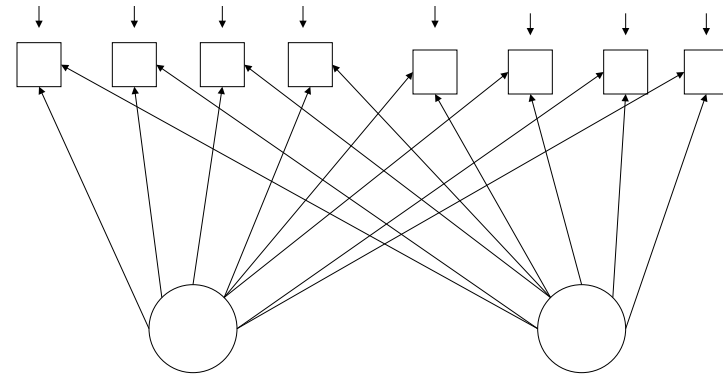
$$D_{i,2} = \lambda_{21}F_{i,1} + \lambda_{22}F_{i,2} + e_{i,2} \text{ etc.}$$

- The lambda values are called **factor loadings**.
- F_1 and F_2 are sometimes called **common factors**, because they influence all the observed variables.
- Error terms e_1, \dots, e_8 are sometimes called **unique factors**, because each one influences only a single observed variable.

Factor Analysis can be

- **Exploratory:** The goal is to describe and summarize the data by explaining a large number of observed variables in terms of a smaller number of latent variables (factors). The factors are the reason the observable variables have the correlations they do.
- **Confirmatory:** Statistical estimation and testing as usual.

Part One: Unconstrained (Exploratory) Factor Analysis



$$\mathbf{D} = \mathbf{\Lambda F} + \mathbf{e}$$

$$V(\mathbf{F}) = \mathbf{\Phi}$$

$$V(\mathbf{e}) = \mathbf{\Omega} \text{ (usually diagonal)}$$

\mathbf{F} and \mathbf{e} independent multivariate normal

$$V(\mathbf{D}) = \mathbf{\Sigma} = \mathbf{\Lambda \Phi \Lambda}' + \mathbf{\Omega}$$

Main interest is in the number of factors and the factor loadings $\mathbf{\Lambda}$.

A Re-parameterization

$$\mathbf{\Sigma} = \mathbf{\Lambda \Phi \Lambda}' + \mathbf{\Omega}$$

Square root matrix: $\mathbf{\Phi} = \mathbf{S S} = \mathbf{S S}'$, so

$$\begin{aligned} \mathbf{\Lambda \Phi \Lambda}' &= \mathbf{\Lambda S S' \Lambda}' \\ &= (\mathbf{\Lambda S}) \mathbf{I} (\mathbf{S' \Lambda}') \\ &= (\mathbf{\Lambda S}) \mathbf{I} (\mathbf{\Lambda S})' \\ &= \mathbf{\Lambda}_2 \mathbf{I} \mathbf{\Lambda}'_2 \end{aligned}$$

Parameters are not identifiable

$$\Sigma = \Lambda\Phi\Lambda' + \Omega \quad \Lambda\Phi\Lambda' = \Lambda_2\mathbf{I}\Lambda_2'$$

- Two distinct (Lambda, Phi) pairs give the same Sigma, and hence the same distribution of the data.
- Actually, there are infinitely many. Let \mathbf{Q} be an arbitrary covariance matrix for \mathbf{F} .

$$\begin{aligned} \Lambda_2\mathbf{I}\Lambda_2' &= \Lambda_2\mathbf{Q}^{-\frac{1}{2}}\mathbf{Q}\mathbf{Q}^{-\frac{1}{2}}\Lambda_2' \\ &= (\Lambda_2\mathbf{Q}^{-\frac{1}{2}})\mathbf{Q}(\mathbf{Q}^{-\frac{1}{2}}\Lambda_2') \\ &= (\Lambda_2\mathbf{Q}^{-\frac{1}{2}})\mathbf{Q}(\Lambda_2\mathbf{Q}^{-\frac{1}{2}})' \\ &= \Lambda_3\mathbf{Q}\Lambda_3' \end{aligned}$$

Restrict the model

$$\Lambda\Phi\Lambda' = \Lambda_2\mathbf{I}\Lambda_2'$$

- Set Phi = the identity, so $V(\mathbf{F}) = \mathbf{I}$
- All the factors are standardized, as well as independent.
- Justify this on the grounds of simplicity.
- Say the factors are “orthogonal” (at right angles, uncorrelated).

Parameters are not identifiable

- This shows that the parameters of the general measurement model are not identifiable without some restrictions on the possible values of the parameter matrices.
- Notice that the general unrestricted model could be very close to the truth. But the parameters cannot be estimated successfully, period.

Standardize the observed variables too

- For $j = 1, \dots, k$ and independently for $i=1, \dots, n$
- $Z_{ij} = \frac{D_{ij} - \bar{D}_j}{s_j}$
- Assume each observed variable has variance one as well as mean zero.
- Sigma is now a correlation matrix.
- Base inference on the sample correlation matrix.

Revised Exploratory Factor Analysis Model

$$\mathbf{Z} = \mathbf{\Lambda}\mathbf{F} + \mathbf{e}$$

$$V(\mathbf{F}) = \mathbf{I}$$

$$V(\mathbf{e}) = \mathbf{\Omega} \text{ (usually diagonal)}$$

\mathbf{F} and \mathbf{e} independent multivariate normal

$$V(\mathbf{D}) = \mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Omega}$$

$\mathbf{\Sigma}$ is a correlation matrix.

Communality

$$\begin{aligned} \text{Var}(D_i) &= \sum_{j=1}^p \lambda_{ij}^2 \text{Var}(F_j) + \text{Var}(e_i) \\ &= \sum_{j=1}^p \lambda_{ij}^2 \text{Var}(F_j) + \text{Var}(e_i) \\ &= \sum_{j=1}^p \lambda_{ij}^2 + \omega_i = 1 \end{aligned}$$

- $\sum_{j=1}^p \lambda_{ij}^2$ is the proportion of variance in variable i that comes from the common factors.
- It is called the **communality** of variable i .
- The communality cannot exceed one.
- $\omega_i = 1 - \sum_{j=1}^p \lambda_{ij}^2$ Peculiar?

Meaning of the factor loadings

$$\begin{aligned} \text{Corr}(D_6, F_2) &= \text{Cov}(D_6, F_2) = E(D_6 F_2) \\ &= E((\lambda_{61} F_1 + \lambda_{62} F_2) F_2) \\ &= \lambda_{61} E(F_1 F_2) + \lambda_{62} E(F_2^2) \\ &= \lambda_{61} E(F_1) E(F_2) + \lambda_{62} \text{Var}(F_2) \\ &= \lambda_{62} \end{aligned}$$

- λ_{ij} is the correlation between variable i and factor j .
- Square of λ_{ij} is the reliability of variable i as a measure of factor j .

If we could estimate the factor loadings

- We could estimate the correlation of each observable variable with each factor.
- We could easily estimate reliabilities.
- We could estimate how much of the variance in each observable variable comes from each factor.
- This could reveal what the underlying factors are, and what they mean.
- *Number* of common factors can be very important too.

Examples

- A major study of how people describe objects (using 7-point scales from Ugly to Beautiful, Strong to Weak, Fast to Slow etc. revealed 3 factors of connotative meaning:
 - Evaluation
 - Potency
 - Activity
- Factor analysis of a large collection of personality scales revealed 2 major factors:
 - Neuroticism
 - Extraversion
- Yet another study led to 16 personality factors, the basis of the widely used 16 pf test.

$$\begin{aligned} i' &= (\cos \theta)i + (\sin \theta)j \\ j' &= (-\sin \theta)i + (\cos \theta)j \end{aligned}$$

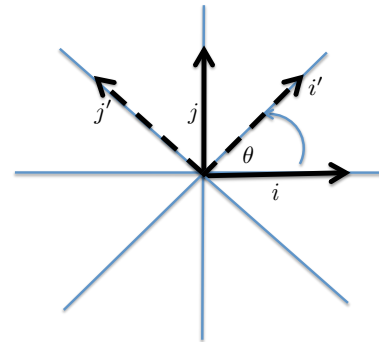
$$\begin{bmatrix} i' \\ j' \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} i \\ j \end{bmatrix} = \mathbf{R} \begin{bmatrix} i \\ j \end{bmatrix}$$

$$\begin{aligned} \mathbf{R}\mathbf{R}' &= \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \\ &= \begin{bmatrix} \cos^2 \theta + \sin^2 \theta & -\cos \theta \sin \theta + \sin \theta \cos \theta \\ -\sin \theta \cos \theta + \cos \theta \sin \theta & \sin^2 \theta + \cos^2 \theta \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I} \end{aligned}$$

The transpose rotated the axes back through an angle of minus theta.

Rotation Matrices

- Have a co-ordinate system in terms of \vec{i}, \vec{j} orthonormal vectors
- Rotate the axes through an angle θ .



$$\begin{aligned} i' &= i \cos \theta + j \sin \theta \\ j' &= -i \sin \theta + j \cos \theta \end{aligned}$$

In General

- A $p \times p$ matrix \mathbf{R} satisfying $\mathbf{R}^{-1} = \mathbf{R}^T$ is called an *orthogonal matrix*.
- Geometrically, pre-multiplication by an orthogonal matrix corresponds to a rotation in p -dimensional space.
- If you think of a set of factors \mathbf{F} as a set of axes (underlying dimensions), then $\mathbf{R}\mathbf{F}$ is a *rotation* of the factors.
- Call it an *orthogonal* rotation, because the factors remain uncorrelated (at right angles).

Another Source of non-identifiability

$$\begin{aligned}\Sigma &= \Lambda\Lambda' + \Omega \\ &= \Lambda\mathbf{R}\mathbf{R}'\Lambda' + \Omega \\ &= (\Lambda\mathbf{R})(\mathbf{R}'\Lambda') + \Omega \\ &= (\Lambda\mathbf{R})(\Lambda\mathbf{R})' + \Omega \\ &= \Lambda_2\Lambda_2' + \Omega\end{aligned}$$

Infinitely many rotation matrices produce the same Sigma.

A Solution

- Place some restrictions on the factor loadings, so that the only rotation matrix that preserves the restrictions is the identity matrix. For example, $\lambda_{ij} = 0$ for $j > i$
- There are other sets of restrictions that work.
- Generally, they result in a set of factor loadings that are impossible to interpret. Don't worry about it.
- Estimate the loadings by maximum likelihood. Other methods are possible but used much less than in the past.
- All (orthogonal) rotations result in the same value of the likelihood function (the maximum is not unique).
- Rotate the factors (that is, post-multiply the loadings by a rotation matrix) so as to achieve a simple pattern that is easy to interpret.

New Model

$$\begin{aligned}\mathbf{Z} &= \Lambda_2\mathbf{F} + \mathbf{e} \\ &= (\Lambda\mathbf{R})\mathbf{F} + \mathbf{e} \\ &= \Lambda(\mathbf{R}\mathbf{F}) + \mathbf{e} \\ &= \Lambda\mathbf{F}' + \mathbf{e}\end{aligned}$$

\mathbf{F}' is a set of *rotated* factors.

Rotate the factor solution

- Rotate the factors to achieve a simple pattern that is easy to interpret.
- There are various criteria. They are all iterative, taking a number of steps to approach some criterion.
- The most popular rotation method is varimax rotation.
- Varimax rotation tries to maximize the (squared) loading of each observable variable with just one underlying factor.
- So typically each variable has a big loading on (correlation with) one of the factors, and small loadings on the rest.
- Look at the loadings and decide what the factors mean (name the factors).

A Warning

- When a non-statistician claims to have done a “factor analysis,” ask what kind.
- Usually it was a principal components analysis.
- Principal components are linear combinations of the observed variables. They come from the observed variables by direct calculation.
- In true factor analysis, it’s the observed variables that arise from the factors.
- So principal components analysis is kind of like backwards factor analysis, though the spirit is similar.
- Most factor analysis (SAS, SPSS, etc.) does principal components analysis by default.

Copyright Information

This slide show was prepared by Jerry Brunner, Department of Statistics, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. These Powerpoint slides are available from the course website:

<http://www.utstat.toronto.edu/~brunner/oldclass/431s13>