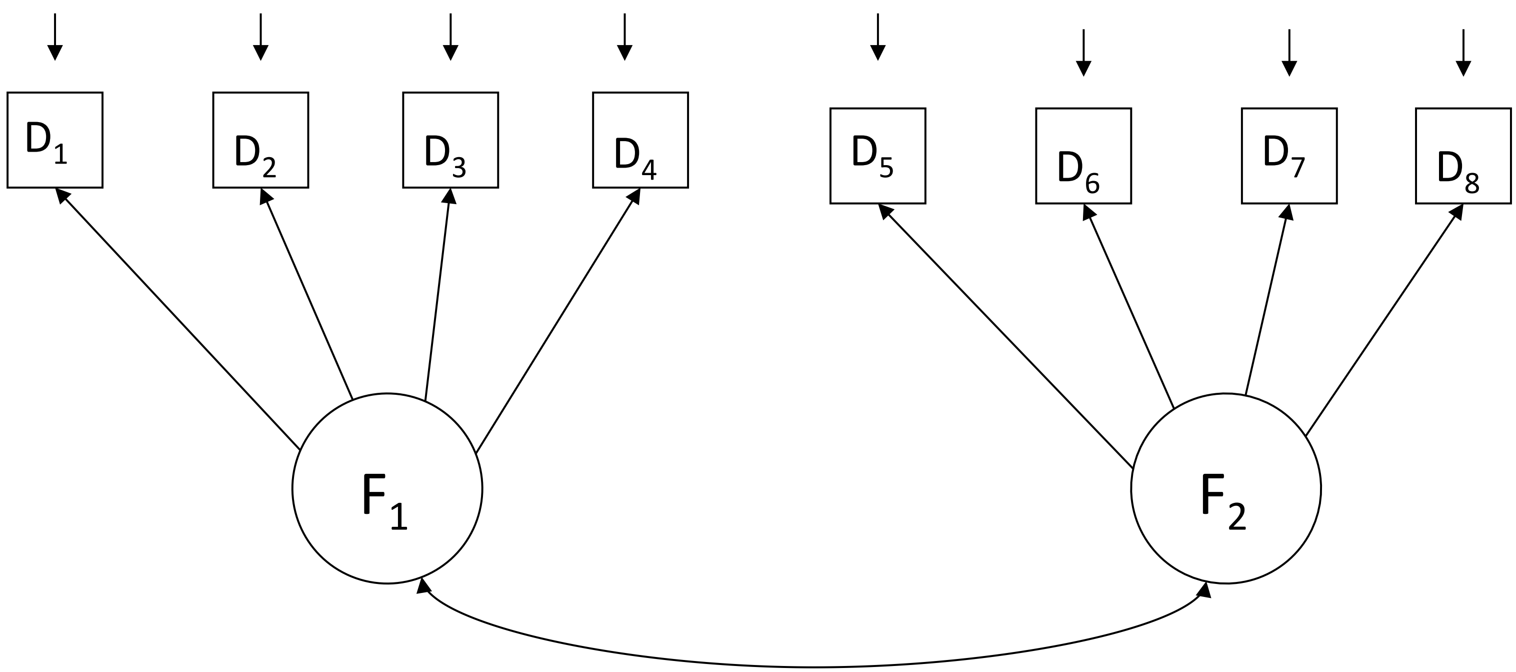
Exploratory Factor Analysis

STA2101: Fall 2019

Factor Analysis: The Measurement Model

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



Example with 2 factors and 8 observed variables

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

$$\begin{pmatrix} D_{i,1} \\ D_{i,2} \\ D_{i,3} \\ D_{i,4} \\ D_{i,5} \\ D_{i,6} \\ D_{i,7} \\ D_{i,8} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \\ \lambda_{71} & \lambda_{27} \\ \lambda_{81} & \lambda_{82} \end{pmatrix} \begin{pmatrix} F_{i,1} \\ F_{i,2} \end{pmatrix} + \begin{pmatrix} e_{i,1} \\ e_{i,2} \\ e_{i,3} \\ e_{i,4} \\ e_{i,5} \\ e_{i,6} \\ e_{i,7} \\ e_{i,8} \end{pmatrix}$$

$$\begin{aligned} D_{i,1} &= \lambda_{11}F_{i,1} + \lambda_{12}F_{i,2} + e_{i,1} \\ D_{i,2} &= \lambda_{21}F_{i,1} + \lambda_{22}F_{i,2} + e_{i,2} \ \text{ etc.} \end{aligned}$$

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

Terminology

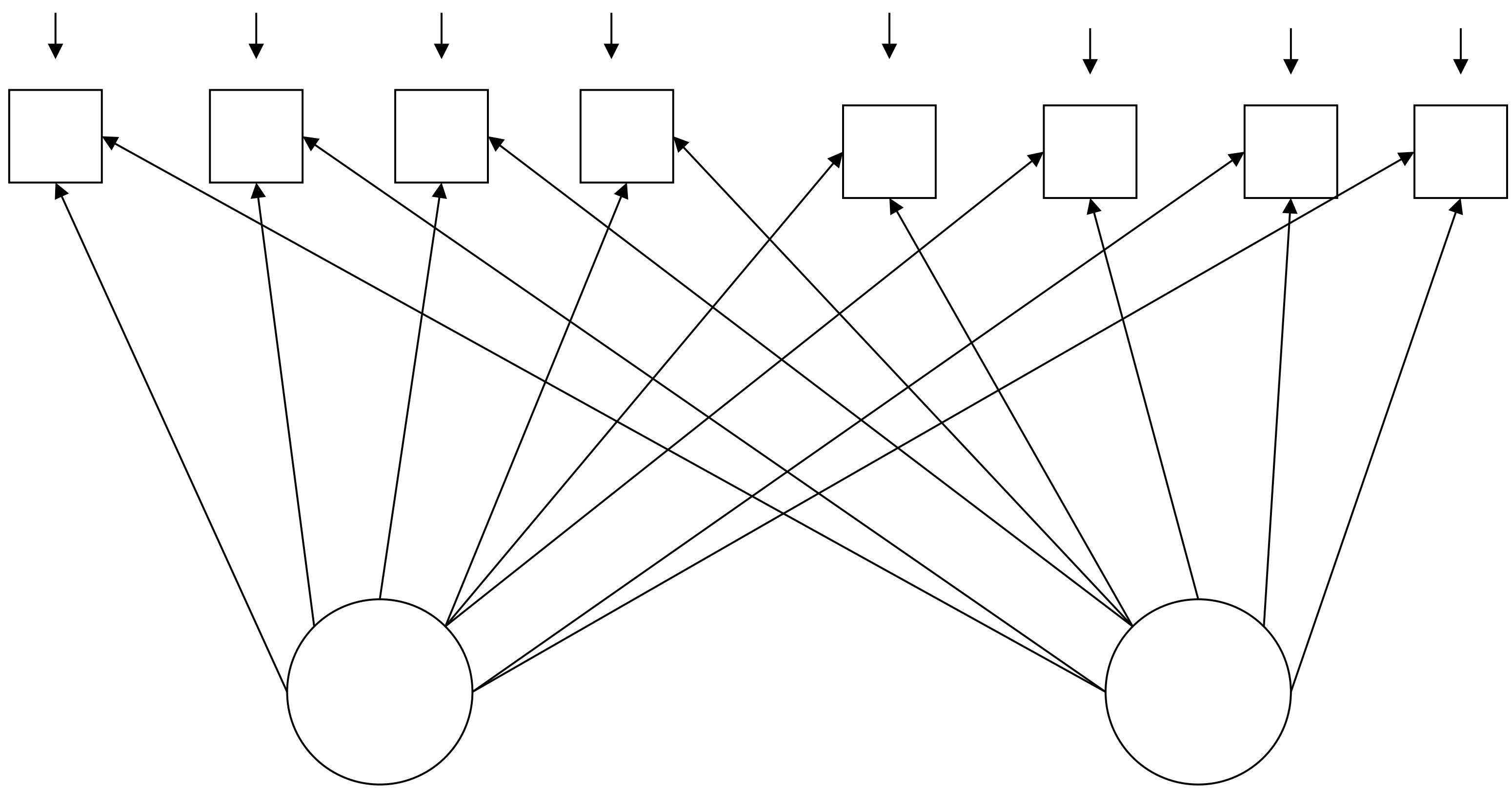
$$
\begin{aligned}
D_{i,1} &= \lambda_{11}F_{i,1} + \lambda_{12}F_{i,2} + e_{i,1} \\
D_{i,2} &= \lambda_{21}F_{i,1} + \lambda_{22}F_{i,2} + e_{i,2} \quad \text{etc.}
\end{aligned}
$$

- 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, …, 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



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

$$\begin{aligned} V(\mathbf{F}) &= \mathbf{\Phi} \\ V(\mathbf{e}) &= \mathbf{\Omega} \text{ (usually diagonal)} \end{aligned}$$

$\mathbf{F}$ and $\mathbf{e}$ independent (multivariate normal)

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

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

A Re-parameterization

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^\top + \boldsymbol{\Omega}$$

Square root matrix: $\boldsymbol{\Phi} = \mathbf{S}\mathbf{S} = \mathbf{S}\mathbf{S}^\top$, so

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

Parameters are not identifiable

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}^\top + \mathbf{\Omega} \qquad \mathbf{\Lambda\Phi\Lambda}^\top = \mathbf{\Lambda}_2\mathbf{I}\mathbf{\Lambda}_2^\top$$

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

$$\begin{aligned}
\mathbf{\Lambda}_2\mathbf{I}\mathbf{\Lambda}_2^\top &= \mathbf{\Lambda}_2\mathbf{Q}^{-\frac{1}{2}}\mathbf{Q}\mathbf{Q}^{-\frac{1}{2}}\mathbf{\Lambda}_2^\top \\
&= (\mathbf{\Lambda}_2\mathbf{Q}^{-\frac{1}{2}})\mathbf{Q}(\mathbf{Q}^{-\frac{1}{2}\top}\mathbf{\Lambda}_2^\top) \\
&= (\mathbf{\Lambda}_2\mathbf{Q}^{-\frac{1}{2}})\mathbf{Q}(\mathbf{\Lambda}_2\mathbf{Q}^{-\frac{1}{2}})^\top \\
&= \mathbf{\Lambda}_3\mathbf{Q}\mathbf{\Lambda}_3^\top
\end{aligned}$$

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.

Restrict the model

$$\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^\top = \mathbf{\Lambda}_2\mathbf{I}\mathbf{\Lambda}_2^\top$$

- Set Phi = the identity, so V(**F**) = 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).

Standardize the observed variables too

- For j = 1, …, k and independently for i=1, …,n
- $$Z_{ij} = \frac{D_{ij} - \overline{D}_j}{s_j}$$
- 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

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

$$\begin{aligned} V(\mathbf{F}) &= \mathbf{I} \\ V(\mathbf{e}) &= \boldsymbol{\Omega} \text{ (usually diagonal)} \end{aligned}$$

$\mathbf{F}$ and $\mathbf{e}$ independent (multivariate normal)

$$V(\mathbf{D}) = \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Omega}$$

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

Meaning of the factor loadings

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

- 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*.

Communality

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

- $\sum_{j=1}^{p} \lambda_{ij}^2$ 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?

If we could estimate the factor loadings

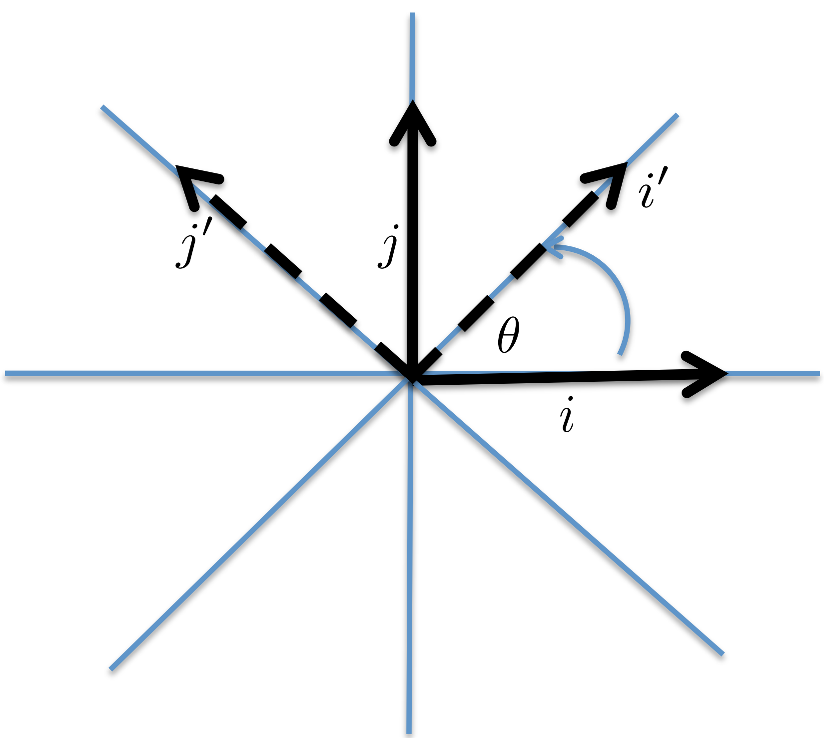
- We could estimate the correlation of each observable variable with each factor.
- We could easily estimate reliabilities.
- We could estimate how much of the variance in each observable variable comes from each factor.
- This could reveal what the underlying factors are, and what they mean.
- *Number* 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 series of studies suggested 16 personality factors, the basis of the widely used 16 pf test.

Rotation Matrices

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



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

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

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

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

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

In General

- A pxp matrix **R** satisfying **R**-inverse = **R**-transpose 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 **F** as a set of axies (underlying dimensions), then **RF** is a *rotation* of the factors.
- Call it an *orthogonal* rotation, because the factors remain uncorrelated (at right angles).

Another Source of non-identifiability

$$\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Omega} \\
&= \boldsymbol{\Lambda}\mathbf{R}\mathbf{R}^\top\boldsymbol{\Lambda}^\top + \boldsymbol{\Omega} \\
&= (\boldsymbol{\Lambda}\mathbf{R})(\mathbf{R}^\top\boldsymbol{\Lambda}^\top) + \boldsymbol{\Omega} \\
&= (\boldsymbol{\Lambda}\mathbf{R})(\boldsymbol{\Lambda}\mathbf{R})^\top + \boldsymbol{\Omega} \\
&= \boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_2^\top + \boldsymbol{\Omega}
\end{aligned}$$

Infinitely many rotation matrices produce the same Sigma.

New Model

$$\begin{aligned}\mathbf{Z} &= \mathbf{\Lambda}_2\mathbf{F} + \mathbf{e} \\ &= (\mathbf{\Lambda}\mathbf{R})\mathbf{F} + \mathbf{e} \\ &= \mathbf{\Lambda}(\mathbf{R}\mathbf{F}) + \mathbf{e} \\ &= \mathbf{\Lambda}\mathbf{F}' + \mathbf{e}\end{aligned}$$

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

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 (orthoganal) 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.

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