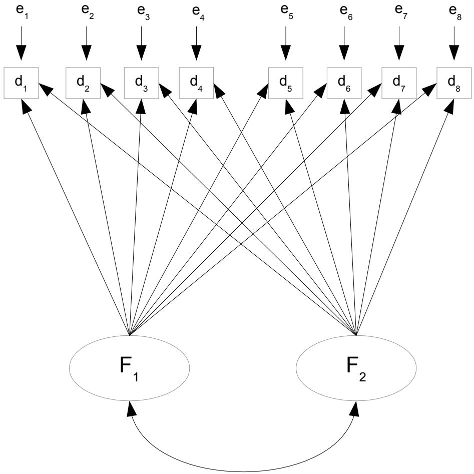
Exploratory Factor Analysis[1]

STA431 Spring 2023

[1] See last slide for copyright information.

Factor Analysis: The Measurement Model

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



Example with 2 factors and 8 observed variables

$$\begin{array}{cccccc} \mathbf{d}_i & = & \mathbf{\Lambda} & \mathbf{F}_i & + & \mathbf{e}_i \end{array}$$

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

Terminology

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

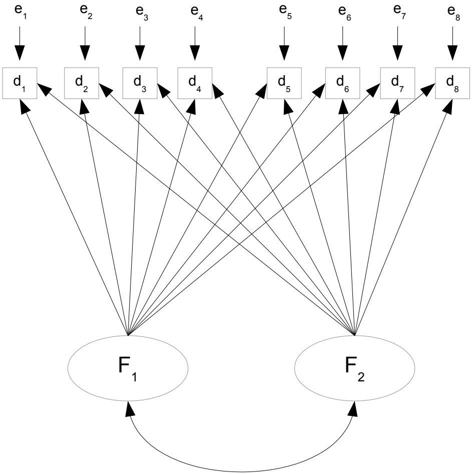
- The lambda values are called *factor loadings*.
- F_1 and F_2 are sometimes called *common factors*, because they influence all the observed variables.
- Error terms $e_1, \ldots, e_8$ are sometimes called *unique factors*, because each one influences only a single observed variable.
- The factors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

Two kinds of factor analysis

- **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. Arrows from all factors to all observable variables.
- **Confirmatory**: Estimation and hypothesis testing as usual.

Unconstrained **Exploratory** Factor Analysis

Arrows from all factors to all observed variables, factors correlated



The Model: $\mathbf{d} = \mathbf{\Lambda F} + \mathbf{e}$

$$
\begin{aligned}
cov(\mathbf{F}) &= \mathbf{\Phi} \\
cov(\mathbf{e}) &= \mathbf{\Omega} \text{ (usually diagonal)} \\
& \mathbf{F} \text{ and } \mathbf{e} \text{ independent (multivariate normal)} \\
cov(\mathbf{d}) &= \mathbf{\Sigma} = \mathbf{\Lambda \Phi \Lambda}^\top + \mathbf{\Omega}
\end{aligned}
$$

Identifiability

$$
\begin{aligned}
\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^\top + \boldsymbol{\Omega} &= \boldsymbol{\Lambda}\, \boldsymbol{\Phi}^{1/2}\mathbf{I}\boldsymbol{\Phi}^{1/2}\, \boldsymbol{\Lambda}^\top + \boldsymbol{\Omega} \\
&= (\boldsymbol{\Lambda}\boldsymbol{\Phi}^{1/2})\mathbf{I}(\boldsymbol{\Phi}^{1/2\top}\boldsymbol{\Lambda}^\top) + \boldsymbol{\Omega} \\
&= (\boldsymbol{\Lambda}\boldsymbol{\Phi}^{1/2})\mathbf{I}(\boldsymbol{\Lambda}\boldsymbol{\Phi}^{1/2})^\top + \boldsymbol{\Omega} \\
&= \boldsymbol{\Lambda}_2\mathbf{I}\boldsymbol{\Lambda}_2^\top + \boldsymbol{\Omega}
\end{aligned}
$$

$(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Omega})$ and $(\mathbf{I}, \boldsymbol{\Lambda}_2, \boldsymbol{\Omega})$ yield the same $\boldsymbol{\Sigma}$.

It's worse than that

Let $\mathbf{Q}$ be an arbitrary positive definite covariance matrix for $\mathbf{F}_i$.

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

So by adjusting the factor loadings, the covariance matrix of the factors could be *anything*.

Parameters are not identifiable

- 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.

Solution: Restrict the model

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

- Fix $\boldsymbol{\Phi} = \mathbf{I}$.
- 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).

Standardize the observed variables too

This is one version

- For $j = 1, \ldots, k$ and independently for $i = 1, \ldots, n$,

$$z_{ij} = \frac{d_{ij} - \mu_j}{\sigma jj}$$

- Each observed variable has variance one as well as mean zero.
- $\mathbf{\Sigma}$ is now a correlation matrix.
- Base inference on the sample correlation matrix.

Standardized Exploratory Factor Analysis Model

Implicitly for $i = 1, \ldots, n$

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

where

- $\mathbf{z}$ is a $k \times 1$ observable random vector. Each element of $\mathbf{z}$ has expected value zero and variance one.
- $\boldsymbol{\Lambda}$ is a $k \times p$ matrix of constants.
- $\mathbf{F}$ (F for factor) is a $p \times 1$ latent random vector with expected value zero and covariance matrix $\mathbf{I}_p$.
- The $k \times 1$ vector of error terms $\mathbf{e}$ has expected value zero and covariance matrix $\boldsymbol{\Omega}$, which is diagonal.
- $\mathbf{F}$ and $\mathbf{e}$ are independent

Factor Loadings are Correlations

$$
\begin{aligned}
corr(\mathbf{z}, \mathbf{F}) &= cov(\mathbf{z}, \mathbf{F}) \\
&= cov(\boldsymbol{\Lambda}\mathbf{F} + \mathbf{e}, \mathbf{F}) \\
&= \boldsymbol{\Lambda} cov(\mathbf{F}, \mathbf{F}) + cov(\mathbf{e}, \mathbf{F}) \\
&= \boldsymbol{\Lambda} cov(\mathbf{F}) + \mathbf{0} \\
&= \boldsymbol{\Lambda}\mathbf{I} \\
&= \boldsymbol{\Lambda}
\end{aligned}
$$

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

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

$$\begin{aligned}
z_1 &= \lambda_{11}F_1 + \lambda_{12}F_2 + \cdots + \lambda_{1p}F_p + e_1 \\
z_2 &= \lambda_{21}F_1 + \lambda_{22}F_2 + \cdots + \lambda_{2p}F_p + e_2 \\
\vdots & \qquad\qquad\quad \vdots \\
z_k &= \lambda_{k1}F_1 + \lambda_{k2}F_2 + \cdots + \lambda_{kp}F_p + e_k
\end{aligned}$$

$$\begin{aligned}
Var(z_1) &= \lambda_{11}^2 + \lambda_{12}^2 + \cdots + \lambda_{1p}^2 + \omega_1 \\
Var(z_2) &= \lambda_{21}^2 + \lambda_{22}^2 + \cdots + \lambda_{2p}^2 + \omega_2 \\
\vdots & \qquad\qquad\quad \vdots \\
Var(z_k) &= \lambda_{k1}^2 + \lambda_{k2}^2 + \cdots + \lambda_{kp}^2 + \omega_k
\end{aligned}$$

$Var(z_j) = 1$, so $\omega_j = 1 - \lambda_{j1}^2 - \lambda_{j2}^2 - \cdots - \lambda_{jp}^2$

Communality and Uniqueness

$Var(z_j) = \lambda_{j1}^2 + \lambda_{j2}^2 + \cdots + \lambda_{jp}^2 + \omega_j = 1$

- The explained variance in z_j is $\lambda_{j1}^2 + \lambda_{j2}^2 + \cdots + \lambda_{jp}^2$. It is called the *communality*.
- To get the communality, add the squared factor loadings in row j of $\mathbf{\Lambda}$.
- $\omega_j = 1 - \lambda_{j1}^2 - \lambda_{j2}^2 - \cdots - \lambda_{jp}^2$ is called the *uniqueness*. It's the proportion of variance that is *not* explained by the factors.

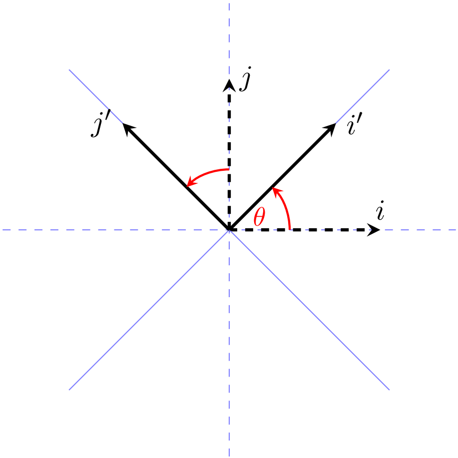
If we could estimate the factor loadings

- We could estimate the correlation of each observable variable with each factor.
- We could easily estimate reliabilities.
- We could assess 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.

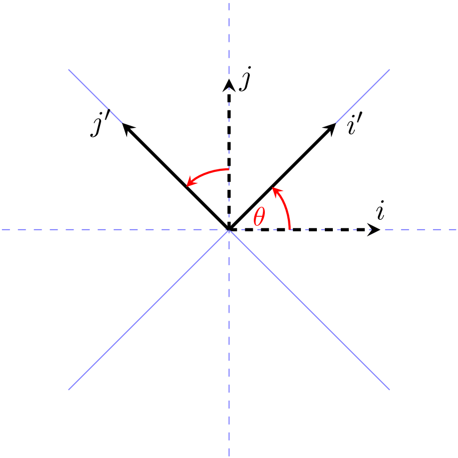
Unfortunately, we still can't estimate the factor loadings.

Rotation Matrices

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



Equations of Rotation



If a point on the plane is denoted in terms of $\vec{i}$ and $\vec{j}$ by (x, y), its position in terms of the rotated basis vectors is

$$
\begin{aligned}
x' &= x\cos\theta + y\sin\theta \\
y' &= -x\sin\theta + y\cos\theta.
\end{aligned}
$$

In Matrix Form

The equations of rotation

$$\begin{aligned} x' &= \quad x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta. \end{aligned}$$

May be written

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{R} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Using the identities $\cos(-\theta) = \cos\theta$ and $\sin(-\theta) = -\sin\theta$, rotate back through an angle of $-\theta$.

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = \mathbf{R}^\top \begin{pmatrix} x' \\ y' \end{pmatrix}.$$

Verifying that $\mathbf{R}^\top = \mathbf{R}^{-1}$

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

In higher dimension as well, pre-multiplication by an orthogonal matrix corresponds to a rotation or possibly a reflection.

Another source of non-identifiability

Returning to the standardized factor model

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

Infinitely many rotation matrices produce the same $\boldsymbol{\Sigma}$, even though the factor loadings in $\boldsymbol{\Lambda}_2 = \boldsymbol{\Lambda}\mathbf{R}^\top$ can be very different for different $\mathbf{R}$ matrices.

Rotating the Factors

Recall $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}^\top + \mathbf{\Omega} = \mathbf{\Lambda}\mathbf{R}^\top\mathbf{R}\mathbf{\Lambda}^\top + \mathbf{\Omega}$

Post-multiplication of $\mathbf{\Lambda}$ by $\mathbf{R}^\top$ is often called "rotation of the factors."

$$\begin{aligned} \mathbf{z} &= \mathbf{\Lambda}\mathbf{F} + \mathbf{e} \\ &= (\mathbf{\Lambda}\mathbf{R}^\top)(\mathbf{R}\mathbf{F}) + \mathbf{e} \\ &= \mathbf{\Lambda}_2\mathbf{F}' + \mathbf{e}. \end{aligned}$$

- $\mathbf{F}' = \mathbf{R}\mathbf{F}$ is a set of *rotated* factors.
- All rotations of the factors produce the same covariance matrix of the observable data.

Same Explained Variance

When factors are rotated

- Communality of variable i is $\sum_{j=1}^p \lambda_{ij}^2$.
- Add up the squares of the factor loadings in row i of $\boldsymbol{\Lambda}$.
- This equals the *ith* diagonal of element of $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$.

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

Ouch.

Strategy

1. 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 $j > i$.
2. Generally, the restrictions may not make sense in terms of the data. Don't worry about it.
3. Estimate the loadings, perhaps by maximum likelihood.
4. All (orthogonal) rotations result in the same maximum value of the likelihood function. That is, the maximum is not unique. Again, don't worry about it.
5. Pick a rotation that results in a simple pattern in the factor loadings, one that is easy to interpret.

Simple Structure

Something like this would be nice

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0.87 & 0.00 \\ -0.95 & 0.03 \\ 0.79 & 0.00 \\ 0.00 & 0.88 \\ 0.01 & 0.75 \\ 0.02 & -0.94 \\ 0.00 & -0.82 \end{pmatrix}$$

Rotation to Simple Structure

Rotation means post-multiply $\mathbf{\Lambda}$ by a rotation matrix

- Used to be subjective, and done by hand!
- Now it's objective and done by computer.
- 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

- The original idea was to maximize the variability of the *squared* loadings in each column.

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0.87 & 0.00 \\ -0.95 & 0.03 \\ 0.79 & 0.00 \\ 0.00 & 0.88 \\ 0.01 & 0.75 \\ 0.02 & -0.94 \\ 0.00 & -0.82 \end{pmatrix}$$

- The results weren't great, so they fixed it up, expressing each squared factor loading as a proportion of the communality.
- Note that the criterion depends on the factor loadings only through the λ_{ij}^2.
- In practice, varimax rotation tends to maximize the squared loading of each observable variable with *just one underlying factor*.

Everybody Loves Varimax

- Estimate the factor loadings with some crazy restrictions.
- Apply a varimax rotation.
- Interpret the results.

Note that rotation does not affect communalities (explained variance).

The Missing Ingredient: Number of Common Factors

- Number of common factors is generally not known in advance. This is *exploratory* factor analysis.
- There are *lots* of ideas and suggestions.
    - At least three variables per factor.
    - At least five variables per factor.
    - ...

Wisdom

From Kaiser (Mr. Varimax)

- There are probably hundreds of common factors.
- Including them all in the model is out of the question.
- The objective should be to come up with a model that includes the most important ones, and captures the essence of what is going on.
- Simplicity is important. Other things being more or less equal, the fewer factors the better.

Estimating Number of Factors

The three most popular ideas?

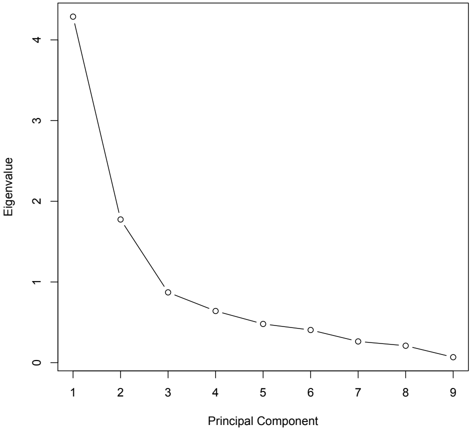
- Number of eigenvalues (of the sample correlation matrix) greater than one.
- Scree plots.
- Testing.

Scree Plots

- In geology, "scree" is the pile of rock and debris often found at the foot of a mountain cliff or volcano.
- Scree slopes tend to be concave up, steepest near the cliff and then tailing off.
- In factor analysis, a scree plot shows the eigenvalues of the correlation matrix, sorted in order of magnitude.

Scree Plot of the Body-Mind Data

See textbook



- It is very common for the graph to decrease rapidly at first, and then straighten out with a small negative slope for the rest of the way.
- The point at which the linear trend begins is the estimated number of factors.

Testing

- If the model is fit by maximum likelihood, carry out the likelihood ratio test for goodness of fit.
- If we really insist that the error terms are independent of the factors and have a diagonal covariance matrix, the only way that the model can be incorrect is that it does not have enough factors.
- Thus, any test for goodness of fit is also a test for number of factors.
- So if a model fails the goodness of fit test, increase the number of factors and try again.
- However …

Can you ever have too much statistical power?

- In reality, there are probably hundreds of factors.
- The power of the likelihood ratio test increases with the sample size
- For large samples, significant lack of fit may be expected for any model with a modest number of factors.
- Even if it's a good model.
- So while formal testing for lack of fit may be useful, one should not rely on it exclusively.

Consulting Advice

- 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, its the observed variables that arise from the factors.
- So principal components analysis is kind of like backwards factor analysis, though the spirit is similar

Copyright Information