Confirmatory Factor Analysis: The Measurement Model[1]

STA431 Winter/Spring 2013

[1] See last slide for copyright information.

A confirmatory factor analysis model

One Factor: Starting simply

$$\begin{aligned} Z_1 &= \lambda_1 F + e_1 \\ Z_2 &= \lambda_2 F + e_2 \\ Z_3 &= \lambda_3 F + e_3 \end{aligned}$$

$V(F) = V(e_1) = V(e_2) = V(e_3) = 1$

F, e_1, e_2, e_3 all independent

$$\begin{aligned} & V(Z_1) = 1 = \lambda^2 + V(e_1) \\ \Rightarrow \quad & V(e_1) = 1 - \lambda_1^2, \text{ etc.} \end{aligned}$$

Σ is a correlation matrix

$Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)}$

$$\begin{aligned} Z_1 &= \lambda_1 F + e_1 \\ Z_2 &= \lambda_2 F + e_2 \\ Z_3 &= \lambda_3 F + e_3 \end{aligned}$$

$$\boldsymbol{\Sigma} = \begin{array}{c|ccc} & Z_1 & Z_2 & Z_3 \\ \hline Z_1 & 1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ Z_2 & & 1 & \lambda_2\lambda_3 \\ Z_3 & & & 1 \end{array}$$

- $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \lambda_3)$
- The parameter space is an open cube.
- Are the parameters identifiable? What if just one is zero?

Suppose no factor loadings equal zero

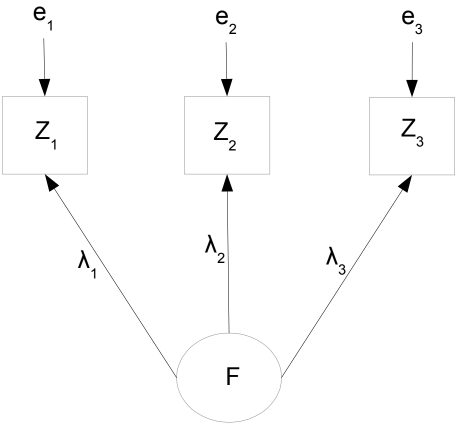
$$\boldsymbol{\Sigma} = \left(\begin{array}{ccc} 1 & \sigma_{12} & \sigma_{13} \\ & 1 & \sigma_{23} \\ & & 1 \end{array} \right) = \left(\begin{array}{ccc} 1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ & 1 & \lambda_2\lambda_3 \\ & & 1 \end{array} \right)$$

$$\begin{array}{rcl} \lambda_1^2 & = & \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} = \frac{\lambda_1\lambda_2 \, \lambda_1\lambda_3}{\lambda_2\lambda_3} \\ \lambda_2^2 & = & \frac{\sigma_{12}\sigma_{23}}{\sigma_{13}} \\ \lambda_3^2 & = & \frac{\sigma_{13}\sigma_{23}}{\sigma_{12}} \end{array}$$

- Squared factor loadings are identifiable, but not the loadings.
- Replace λ_j with $-\lambda_j$, get same $\boldsymbol{\Sigma}$
- Likelihood function will have two maxima, same height.
- Which one you find depends on where you start.

Solution: Decide on the sign of one loading

Based on *meaning*



- Is F_1 math ability or math *inability*? You decide.
- It's just a matter of naming the factors.

If $\lambda_1 > 0$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \sigma_{12} & \sigma_{13} \\ & 1 & \sigma_{23} \\ & & 1 \end{pmatrix} = \begin{pmatrix} 1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ & 1 & \lambda_2\lambda_3 \\ & & 1 \end{pmatrix}$$

- Signs of λ_2 and λ_3 can be recovered right away from $\boldsymbol{\Sigma}$.
- And all the parameters are identified.

Equality constraints

- Three parameters minus three correlations (equations) = ZERO.
- The inequality constraints are more slippery.

Inequality constraints
Slippery devils

$$\boldsymbol{\Sigma} = \left(\begin{array}{ccc} 1 & \sigma_{12} & \sigma_{13} \\ & 1 & \sigma_{23} \\ & & 1 \end{array} \right) = \left(\begin{array}{ccc} 1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ & 1 & \lambda_2\lambda_3 \\ & & 1 \end{array} \right)$$

- $\sigma_{12}\sigma_{13}\sigma_{23} = \lambda_1^2\lambda_2^2\lambda_3^2$, so $0 \leq \sigma_{12}\sigma_{13}\sigma_{23} < 1$.
- But $\sigma_{12}\sigma_{13}\sigma_{23} < 1$ is true of *any* positive definite correlation matrix, so that's one inequality, not two.
- How about $\lambda_1^2 = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}}$ so $0 \leq \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} < 1$?

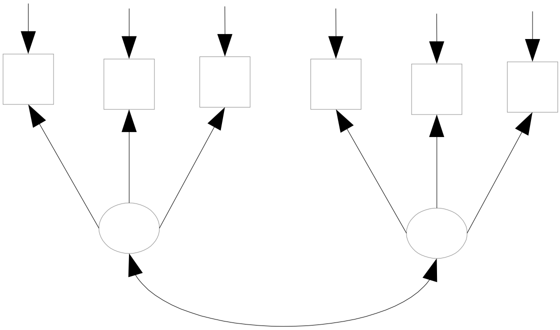
$$\begin{array}{rl} & \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} \frac{\sigma_{23}}{\sigma_{23}} < 1 \\ \Rightarrow & 0 \leq \sigma_{12}\sigma_{13}\sigma_{23} < \sigma_{23}^2 \end{array}$$

Add another variable: $Z_4 = \lambda_4 F + e_4$

$$\boldsymbol{\Sigma} = \left(\begin{array}{cccc} 1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 & \lambda_1\lambda_4 \\ & 1 & \lambda_2\lambda_3 & \lambda_2\lambda_4 \\ & & 1 & \lambda_3\lambda_4 \\ & & & 1 \end{array} \right)$$

- Parameters will all be identifiable as long as 3 out of 4 loadings are non-zero, and one sign is known.
- For example, if $\lambda_1 = 0$ then the top row $= 0$ and you can get $\lambda_2, \lambda_3, \lambda_4$ as before.
- For 5 variables, two loadings can be zero, etc.
- How many equality restrictions? $6 - 4 = 2$.
- Inequality restrictions? It's like an Easter egg hunt.

Now add another factor



$$Z_1 = \lambda_1 F_1 + e_1$$
$$\vdots$$
$$Z_6 = \lambda_6 F_2 + e_6$$

Correlation matrix of observable variables

$$\boldsymbol{\Sigma} = \left(\begin{array}{cccccc} 1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 & \lambda_1\lambda_4\phi_{12} & \lambda_1\lambda_5\phi_{12} & \lambda_1\lambda_6\phi_{12} \\ & 1 & \lambda_2\lambda_3 & \lambda_2\lambda_4\phi_{12} & \lambda_2\lambda_5\phi_{12} & \lambda_2\lambda_6\phi_{12} \\ & & 1 & \lambda_3\lambda_4\phi_{12} & \lambda_3\lambda_5\phi_{12} & \lambda_3\lambda_6\phi_{12} \\ & & & 1 & \lambda_4\lambda_5 & \lambda_4\lambda_6 \\ & & & & 1 & \lambda_5\lambda_6 \\ & & & & & 1 \end{array} \right)$$

- Identify $\lambda_1, \lambda_2, \lambda_3$ from set 1
- Identify $\lambda_4, \lambda_5, \lambda_6$ from set 2
- Identify ϕ_{12} from any unused correlation.
- What if you added more variables?
- What if you added more factors?
- What if observed variables were centered but not standardized?

Three-variable identification rule

For a factor analysis model, the parameters will be identifiable provided

- Errors are independent of one another and of the factors.
- Variances of factors equal one.
- Each observed variable is a function of only one factor.
- There are at least three observable variables with non-zero loadings per factor.
- The sign of one non-zero loading is known for each factor.

Copyright Information