Confirmatory Factor Analysis Part One[1]

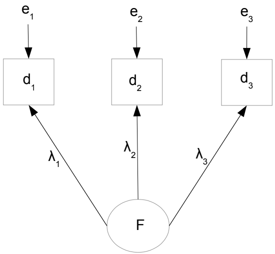
STA2053 Fall 2022

[1] See last slide for copyright information.

A confirmatory factor analysis model

One Factor: Starting simply

$$
\begin{aligned}
d_1 &= \lambda_1 F + e_1 \\
d_2 &= \lambda_2 F + e_2 \\
d_3 &= \lambda_3 F + e_3
\end{aligned}
$$



- $Var(F) = 1$
- $Var(e_j) = \omega_j$
- F, e_1, e_2, e_3 all independent.

Calculate $\boldsymbol{\Sigma}$

$$
\begin{aligned}
d_1 &= \lambda_1 F + e_1 \\
d_2 &= \lambda_2 F + e_2 \\
d_3 &= \lambda_3 F + e_3
\end{aligned}
$$

$\boldsymbol{\Sigma} =$

	d_1	d_2	d_3
d_1	$\lambda_1^2 + \omega_1$	$\lambda_1 \lambda_2$	$\lambda_1 \lambda_3$
d_2		$\lambda_2^2 + \omega_2$	$\lambda_2 \lambda_3$
d_3			$\lambda_3^2 + \omega_3$

Are the parameters identifiable? What if just one λ is zero?

Suppose no factor loadings equal zero

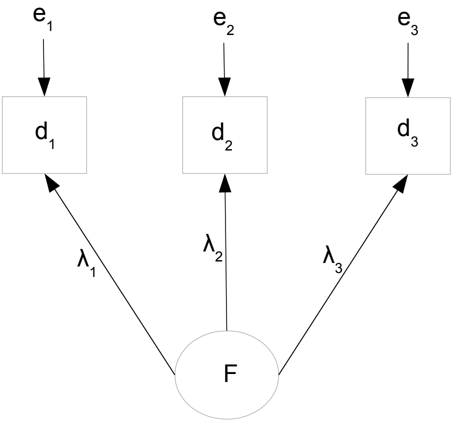
$$\boldsymbol{\Sigma} = \left(\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{array} \right) = \left(\begin{array}{ccc} \lambda_1^2 + \omega_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ & \lambda_2^2 + \omega_2 & \lambda_2\lambda_3 \\ & & \lambda_3^2 + \omega_3 \end{array} \right)$$

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

- Squared factor loadings are identifiable, but not the loadings.
- Replace all 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 math ability or math *inability*? You decide.
- It's just a matter of naming the factors.

If $\lambda_1 > 0$

$$\boldsymbol{\Sigma} = \left(\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{array} \right) = \left(\begin{array}{ccc} \lambda_1^2 + \omega_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ & \lambda_2^2 + \omega_2 & \lambda_2\lambda_3 \\ & & \lambda_3^2 + \omega_3 \end{array} \right)$$

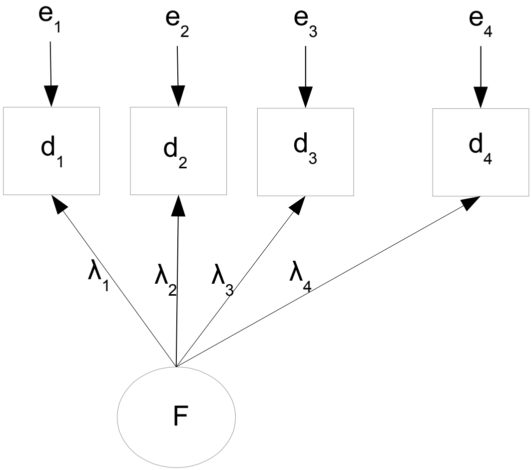
- Signs of λ_2 and λ_3 can be recovered right away from $\boldsymbol{\Sigma}$.
- And all the parameters are identified.

Add another variable: $d_4 = \lambda_4 F + e_4$

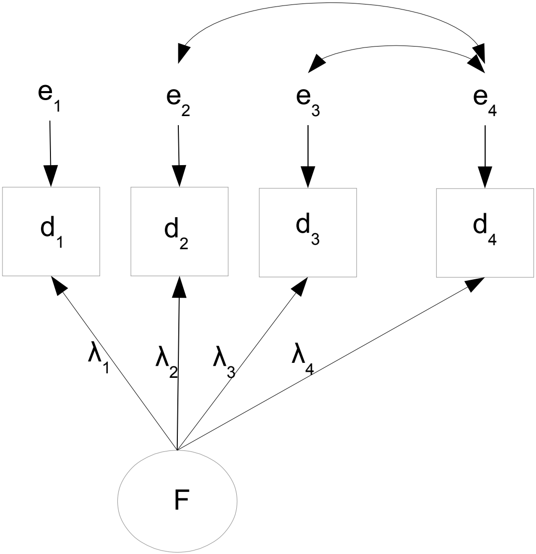
$$
\boldsymbol{\Sigma} = \left(\begin{array}{cccc}
\lambda_1^2 + \omega_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 & \lambda_1\lambda_4 \\
 & \lambda_2^2 + \omega_2 & \lambda_2\lambda_3 & \lambda_2\lambda_4 \\
 & & \lambda_3^2 + \omega_3 & \lambda_3\lambda_4 \\
 & & & \lambda_4^2 + \omega_4
\end{array} \right)
$$

- Parameters will all be identifiable as long as 3 out of 4 loadings are non-zero, and one sign is known.
- For example, if only $\lambda_1 = 0$ 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 with 4 variables? $6 - 4 = 2$.

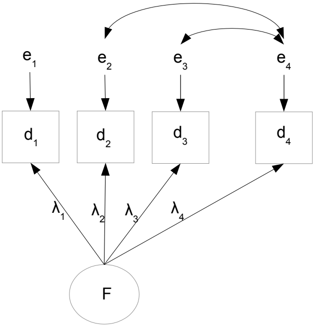
Four Observable Variables



Four Observable Variables



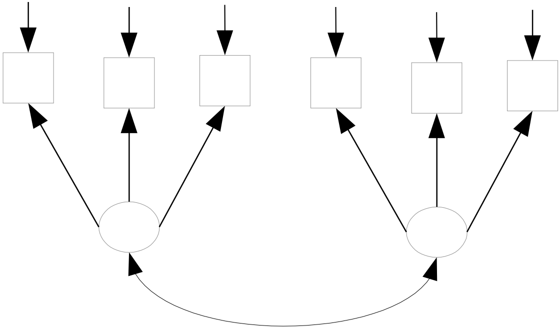
Identifying the covariances between error terms



$$
\begin{aligned}
d_2 &= \lambda_2 F + e_2 \\
d_4 &= \lambda_4 F + e_4 \\
\sigma_{24} = Cov(d_2, d_4) &= \lambda_2 \lambda_4 Var(F) + Cov(e_2, e_4) \\
&= \lambda_2 \lambda_4 + \omega_{24} \\
&\Rightarrow \omega_{24} = \sigma_{24} - \lambda_2 \lambda_4
\end{aligned}
$$

Now add another factor

$Var(F_1) = Var(F_2) = 1$



$$
\begin{aligned}
d_1 &= \lambda_1 F_1 + e_1 \\
&\vdots \\
d_6 &= \lambda_6 F_2 + e_6
\end{aligned}
$$

Covariance matrix of observable variables

$$\boldsymbol{\Sigma} = \left(\begin{array}{rrrrrr} \lambda_1^2+\omega_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 & \lambda_1\lambda_4\phi_{12} & \lambda_1\lambda_5\phi_{12} & \lambda_1\lambda_6\phi_{12} \\ & \lambda_2^2+\omega_2 & \lambda_2\lambda_3 & \lambda_2\lambda_4\phi_{12} & \lambda_2\lambda_5\phi_{12} & \lambda_2\lambda_6\phi_{12} \\ & & \lambda_3^2+\omega_3 & \lambda_3\lambda_4\phi_{12} & \lambda_3\lambda_5\phi_{12} & \lambda_3\lambda_6\phi_{12} \\ & & & \lambda_4^2+\omega_4 & \lambda_4\lambda_5 & \lambda_4\lambda_6 \\ & & & & \lambda_5^2+\omega_5 & \lambda_5\lambda_6 \\ & & & & & \lambda_6^2+\omega_6 \end{array} \right)$$

- Identify $\lambda_1, \lambda_2, \lambda_3$ from set One (assuming one sign is known).
- Identify $\lambda_4, \lambda_5, \lambda_6$ from set Two (lower right).
- Identify ϕ_{12} from any unused covariance.
- What if you added more variables?
- What if you added more factors?

Three-variable identification rule

For standardized factors

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

- Errors are independent of one another and of the factors.
- Variances of all 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.

Maybe we can do better.

Reference Variable Definition

A *reference variable* for a latent variable is an observable variable that is a function only of that latent variable and an error term. The factor loading is non-zero.

Three-variable identification rule for standardized factors: Version Two

- For a factor analysis model with a single standardized factor and three reference variables, the parameters will be identifiable provided that the errors are independent of one another and of the factor, and that the sign of one factor loading is known.
- Additional observed variables may be added to the model provided that for each additional variable, the error term of the additional variable has zero covariance with the error term at least one reference variable.
- Under these conditions, the error term of each additional variable may have non-zero covariance with the error terms of all the *other* (unused) variables already in the model.

Factor Model Combination Rule

- Suppose there are two factor analysis models A and B with identifiable parameters.
- Every factor has at least one reference variable.
- If the error terms of A have zero covariance with the error terms of B, the only additional parameters to be identified are the covariances between the A factors and the B factors.
- These may be identified from the covariances of the A reference variables and the B reference variables (one per factor).
- What covariances between error terms are permitted?
- Alternatively ...

Copyright Information