Measurement error: Example

Data on the relationship between Heart (y) damage and Stress (x)

Heart = β₀ + β₁ Stressed

What happens if we add random error, N(0, δ × SD_Stress) to each x-value (δ = {0.75, 1.0, 1.5})?

- The grey ellipse and the regression line “0” show the original data
- Increasing measurement error makes the data ellipses wider
- Increasing measurement error biases β₁ towards zero!
- NB: Adding random error to Heart (y) would decrease precision but not introduce bias.
Measurement models

Now, consider a multiple regression model, with coffee as an additional predictor

\[ \text{Heart} = \beta_0 + \beta_1 \text{Stress} + \beta_2 \text{Coffee} \]

What is the effect of measurement error in Stress on both coefficients, \((\beta_1, \beta_2)\)?

- The coefficient \(\beta_1\) for Stress goes towards 0, as before
- The coefficient \(\beta_2\) for Coffee decreases towards its marginal value (Stress not included in the model)
- Thus, measurement error in even one \(x\) variable has effects throughout the model

Latent variables

In EFA, CFA & SEM, measurement error in observed variables is handled by positing an underlying latent variable (“factor”) responsible for producing the observed score \(x\)

\[ x_i = \lambda \xi_i + \delta_i \]

- \(\xi\) ("ksi" or "xi") is the true latent variable measured by \(x\)
- \(\lambda\) is the regression coefficient (“factor loading”) of \(x\) on \(\xi\)
- \(\delta\) is the error of measurement
- \(x\) is called an indicator of the latent variable \(\xi\)

There are usually multiple observed indicators, \(x_1, x_2, \ldots\) measuring a given (latent) construct

\[ x_1 = \lambda_1 \xi_1 + \delta_1 \]
\[ x_2 = \lambda_2 \xi_1 + \delta_2 \]
\[ x_3 = \lambda_3 \xi_1 + \delta_3 \]
\[ \vdots = \vdots \]

The General CFA model

The general CFA measurement model is

\[ x = \Lambda \xi + \delta \]

where
- \(x\) is the \(q \times 1\) vector of observed or measured variables
- \(\Lambda\) is the \(q \times k\) matrix of factor loadings
- \(\xi\) is the vector of latent variables
- \(i.e., \lambda_{ij}\) is the partial regression coefficient for \(x_i\) on \(\xi_j\) in the regression of \(x_j\) on \(\xi_1, \xi_2, \ldots, \xi_k\)
- \(\delta\) is the vector of errors of measurement or disturbance terms

This model, together with assumptions implies that the covariance matrix of \(x\) is

\[ \Sigma = \Lambda \Phi \Lambda^T + \Theta \]

where \(\Phi\) is the covariance matrix of the factors, \(\xi\), and \(\Theta\) is the covariance matrix of the errors, \(\delta\)
Test theory models

Testing Equivalence of Measures with CFA

Test theory is concerned with ideas of reliability, validity and equivalence of measures.
- The same ideas apply to other constructs (e.g., anxiety scales or experimental measures of conservation).
- Test theory defines several degrees of “equivalence”.
- Each kind may be specified as a confirmatory factor model with a single common factor.
- The CFA approach allows a more nuanced approach to these issues.

Path diagram:

\[
\Sigma = \begin{pmatrix}
\lambda_1 \\
\lambda_2 \\
\lambda_3 \\
\lambda_4
\end{pmatrix} +
\begin{pmatrix}
\theta_{11} & \theta_{22} & \theta_{33} & \theta_{44}
\end{pmatrix}
\]

Congeneric measurement model

- The single factor model is called the congeneric measurement model.
- It implies that the true scores, \( \tau_i = \lambda_i \xi \) are perfectly correlated.
- The true score variance in \( x_i \) is \( \lambda_i^2 \) — also called communality in EFA lingo.
- The reliability of \( x_i \) is
  \[ \rho_i = \frac{\lambda_i^2}{\text{var}(x_i)} = \frac{\lambda_i^2}{\lambda_i^2 + \theta_{ii}} = 1 - \frac{\theta_{ii}}{\lambda_i^2 + \theta_{ii}} \]
- Strictly speaking, the error term \( \delta_i \) (“unique factor”) is considered to be the sum of two uncorrelated components.
  \[ \delta_i = s_i + e_i \]
  unique = specific + error
- \( \rho_i \) is a lower bound on true reliability.

Kinds of equivalence

- **Parallel tests**: Measure the same thing with equal precision. The strongest form of “equivalence”.
- **Tau-equivalent tests**: Have equal true score variances (\( \lambda_i^2 \)), but may differ in error variance (\( \theta_{ii} \)).
  Like parallel tests, this requires tests of the same length & time limits.
  E.g., short forms cannot be \( \tau \)-equivalent.
- **Congeneric tests**: The weakest form of equivalence: All tests measure a single common factor, but the loadings & error variances may vary.
  These hypotheses may be tested with CFA/SEM by testing equality of the factor loadings (\( \lambda_i \)) and unique variances (\( \theta_{ii} \)).
Test theory models

Example: Votaw data

Example: Reliability in essay scoring

Essay exams present a challenge for standardized testing (SAT, LSAT, etc.).

An early study by Votaw (1948) analyzed scores for N=126 examinees given a 3-part English composition test:

- $x_1$: score on an original copy of the part 1 essay
- $x_2$: score on a hand-written copy of the part 1 essay
- $x_3$: score on a carbon-copy of the hand-written part 1 essay
- $x_4$: score on an original copy of the part 2 essay

Questions:

- Can these scores be used interchangeably— as strictly parallel or $\tau$-equivalent tests?
- If not, are the scores on original copies more reliable than those on copies?
- Are the scores for part 1 and part 2 originals equally reliable?

Other models

More restrictive models are specified simply by using the same parameter names for equal parameters.

$\tau$-equivalent model

parallel model

Example: Reliability in essay scoring

Read the covariance matrix:

```r
library(sem)
votaw <- readMoments(diag=TRUE, 
  names=c('orig1', 'hcpy1', 'ccpy1', 'orig2'), text="
")
```

Fit the congeneric model:

```r
votaw.mod1 <- specifyEquations(text="
orig1 = lam1 * Ability
hcpy1 = lam2 * Ability
ccpy1 = lam3 * Ability
orig2 = lam4 * Ability
V(Ability) = 1
")
```

An intermediate “semi-parallel” model specified two sets of equal loadings $\lambda_1$ for $orig1$ and $orig2$, $\lambda_1$ for $hcpy1$ and $ccpy1$

Summary of analyses:

<table>
<thead>
<tr>
<th>Model</th>
<th>Hypothesis</th>
<th>df</th>
<th>$\chi^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>congeneric</td>
<td>2</td>
<td>2.28</td>
<td>0.32</td>
</tr>
<tr>
<td>2</td>
<td>$\tau$-equivalent</td>
<td>5</td>
<td>40.42</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>parallel</td>
<td>8</td>
<td>109.12</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>semi-parallel</td>
<td>6</td>
<td>8.99</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Results for congeneric model:

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\hat{\lambda}_i$</th>
<th>s.e.$(\hat{\lambda}_i)$</th>
<th>$\hat{\rho}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>orig1</td>
<td>4.57</td>
<td>0.36</td>
<td>0.83</td>
</tr>
<tr>
<td>hcpy1</td>
<td>2.68</td>
<td>0.45</td>
<td>0.25</td>
</tr>
<tr>
<td>ccpy1</td>
<td>2.65</td>
<td>0.40</td>
<td>0.31</td>
</tr>
<tr>
<td>orig2</td>
<td>4.54</td>
<td>0.33</td>
<td>0.94</td>
</tr>
</tbody>
</table>

However, semi-parallel model is simpler, and fits well.
Several Sets of Congeneric Tests

For several sets of measures, the test theory ideas of congeneric tests can be extended to test the equivalence of each set. If each set is congeneric, the estimated correlations among the latent factors measure the strength of relations among the underlying “true scores.”

Example: Correcting for Unreliability
- Given two measures, x and y, the correlation between them is limited by the reliability of each.
- CFA can be used to estimate the correlation between the true scores, $\tau_x$, $\tau_y$, or to test the hypothesis that the true scores are perfectly correlated:

$H_0: \rho(\tau_x, \tau_y) = 1$

- The estimated true-score correlation, $\hat{\rho}(\tau_x, \tau_y)$ is called the correlation of x, y corrected for attenuation.

The analysis requires two “parallel” forms of each test, $x_1, x_2, y_1, y_2$. Tests are carried out with the model:

$$
\begin{bmatrix}
  x_1 \\
  x_2 \\
  y_1 \\
  y_2 \\
\end{bmatrix} = \begin{bmatrix}
  \beta_1 & 0 \\
  \beta_2 & 0 \\
  0 & \beta_3 \\
  0 & \beta_4 \\
\end{bmatrix} \begin{bmatrix}
  \tau_x \\
  \tau_y \\
\end{bmatrix} + \begin{bmatrix}
  e_1 \\
  e_2 \\
  e_3 \\
  e_4 \\
\end{bmatrix}
$$

with $\text{corr}(\tau) = \rho$, and $\text{var}(e) = \text{diag}\{\theta_1^2, \theta_2^2, \theta_3^2, \theta_4^2\}$. The model is shown in this path diagram:

Several Sets of Congeneric Tests

Hypotheses
The following hypotheses can be tested. The difference in $\chi^2$ for $H_1$ vs. $H_2$, or $H_3$ vs. $H_4$ provides a test of the hypothesis that $\rho = 1$.

- $H_1: \rho = 1$ and $H_2$
- $H_2: \begin{cases}
  \beta_1 = \beta_2 \\
  \beta_3 = \beta_4 \\
\end{cases}$ $\theta_1^2 = \theta_2^2$ $\theta_3^2 = \theta_4^2$
- $H_3: \rho = 1$, all other parameters free
- $H_4: \rho = 1$, all parameters free

$H_1$ and $H_2$ assume the measures $x_1, x_2$ and $y_1, y_2$ are parallel. $H_3$ and $H_4$ assume they are merely congeneric.

These four hypotheses actually form a $2 \times 2$ factorial
- parallel vs. congeneric: $H_1$ and $H_2$ vs. $H_3$ and $H_4$ and
- $\rho = 1$ vs. $\rho \neq 1$.

For nested models, model comparisons can be done by testing the difference in $\chi^2$, or by comparing other fit statistics (AIC, BIC, RMSEA, etc.)

- LISREL can fit multiple models, but you have to do the model comparison tests “by hand.”
- AMOS can fit multiple models, and does the model comparisons for you.
- With PROC CALIS, the CALISCMP macro provides a flexible summary of multiple-model comparisons.
- `sem()` provides an `anova()` method
Example: Lord’s data

Lord’s vocabulary test data:
- \(x_1, x_2\): two 15-item tests, liberal time limits
- \(y_1, y_2\): two 75-item tests, highly speeded

Analyses of these data give the following results:

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Free Parameters</th>
<th>(\chi^2)</th>
<th>p-value</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H_1): par, (\rho = 1)</td>
<td>4</td>
<td>6</td>
<td>37.33</td>
<td>0.00</td>
</tr>
<tr>
<td>(H_2): par</td>
<td>5</td>
<td>5</td>
<td>1.93</td>
<td>0.86</td>
</tr>
<tr>
<td>(H_3): cong, (\rho = 1)</td>
<td>8</td>
<td>2</td>
<td>36.21</td>
<td>0.00</td>
</tr>
<tr>
<td>(H_4): cong</td>
<td>9</td>
<td>1</td>
<td>0.70</td>
<td>0.70</td>
</tr>
</tbody>
</table>

- Models \(H_2\) and \(H_4\) are acceptable, by \(\chi^2\) tests
- Model \(H_2\) is “best” by AIC

\(\rho = 1\):

- Both tests reject the hypothesis that \(\rho = 1\),
- Under model \(H_2\), the ML estimate is \(\hat{\rho} = 0.889\).
- \(\Rightarrow\) speeded and unspeeded vocab. tests do not measure exactly the same thing.
- NB: The CFA/SEM approach is far more rigorous than usually applied to social measurements like anxiety, depression, etc.

SAS example: datavis.ca/courses/factor/sas/calislc.sas

Lord’s data: PROC CALIS

data lord(type=cov);
input_type_ $ _name_ $ x1 x2 y1 y2;
datalines;
1 n 649 649 649 649
2 cov x1 86.3937 . . .
3 cov x2 57.7751 86.2632 ...
4 cov y1 56.8651 59.3177 97.2850 ...
5 cov y2 58.8986 59.6683 73.8201 97.8192 ...
6 mean . 0 0 0 0
;

Model \(H_4\):

\(\rho_1, \rho_2, \rho_3, \rho_4 \ldots \rho = \text{free}\)

title "Lord's data: H4- unconstrained two-factor model";
proc calis data=lord
  cov summary outram=M4;
  lineqs x1 = beta1 F1 + e1,
   x2 = beta2 F1 + e2,
   y1 = beta3 F2 + e3,
   y2 = beta4 F2 + e4;
  std F1 F2 = 1 1,
  cov F1 F2 = rho;
run;

The SUMMARY output contains many fit indices:

- Lord’s data: H4- unconstrained two-factor model
- Covariance Structure Analysis: Maximum Likelihood Estimation
- Fit criterion
- Goodness of Fit Index (GFI)
- GFI Adjusted for Degrees of Freedom (AGFI)
- Root Mean Square Residual (RMR)
- Chi-square = 0.7033 df = 1 Prob>chi**2 = 0.4017
- Null Model Chi-square: df = 6 1466.5884
- Bentler's Comparative Fit Index
- Bentler's Comparative Fit Index
- Normal Theory Reweighted LS Chi-square
- Akaike's Information Criterion
- Consistent Information Criterion
- Schwarz's Bayesian Criterion
- McDonald's (1989) Centrality
- Bentler & Bonett's (1980) Non-normed Index
- Bentler & Bonett's (1980) Normed Index
- James, Mulaik, & Brett (1982) Parsimonious Index

...
Lord’s data: PROC CALIS

Model H3: H4, with $\rho = 1$

title "Lord’s data: H3- rho=1, one-congeneric factor";
proc calis data=lord;
cov summary outram=M3;
lineqs x1 = beta1 F1 + e1,
x2 = beta2 F1 + e2,
y1 = beta3 F2 + e3,
y2 = beta4 F2 + e4;
std F1 F2 = 1 1;
run;

Model H2: $\beta_1 = \beta_2, \beta_3 = \beta_4 \ldots, \rho=\text{free}$
title "Lord's data: H2- X1 and X2 parallel, Y1 and Y2 parallel";
proc calis data=lord;
cov summary outram=M2;
lineqs x1 = betax F1 + e1,
x2 = betax F1 + e2,
y1 = betay F2 + e3,
y2 = betay F2 + e4;
std F1 F2 = 1 1;
run;

Lord’s data: CALISCMP macro

Model comparisons using CALISCMP macro and the OUTRAM= data sets
%caliscmp(ram=M1 M2 M3 M4,
models=%str(H1 par rho=1/H2 par/H3 con rho=1/H4 con),
compare=1 2 / 3 4 /1 3/ 2 4);

Model Comparison Statistics from 4 RAM data sets

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>df</th>
<th>Chi-Square</th>
<th>p-value</th>
<th>Residual</th>
<th>GFI</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 par rho=1</td>
<td>4</td>
<td>6</td>
<td>37.3412</td>
<td>0.00000</td>
<td>2.53409</td>
<td>0.97048</td>
<td>25.3412</td>
</tr>
<tr>
<td>H2 par</td>
<td>5</td>
<td>5</td>
<td>1.9320</td>
<td>0.85847</td>
<td>0.69829</td>
<td>0.99849</td>
<td>-8.0680</td>
</tr>
<tr>
<td>H3 con rho=1</td>
<td>8</td>
<td>2</td>
<td>36.2723</td>
<td>0.00000</td>
<td>2.43565</td>
<td>0.97122</td>
<td>32.2723</td>
</tr>
<tr>
<td>H4 con</td>
<td>9</td>
<td>1</td>
<td>0.7033</td>
<td>0.40168</td>
<td>0.27150</td>
<td>0.99946</td>
<td>-1.2967</td>
</tr>
</tbody>
</table>

(more fit statistics are compared than shown here.)

Multi-factor congeneric models

- Multi-factor models are at the heart of CFA
- An important special case is when there are $G$ sets of (assumed) congeneric variables, each of which are indicators of a latent variable
- In EFA lingo, these are called non-overlapping factors
- The measurement models for the variables $x_g$ in set $g$ are of the form

$$x_g = \lambda_g \xi_g + \delta_g$$

- Then, the loadings $\Lambda$ for all variables can be represented as

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_G \end{bmatrix}$$

- The $0$s, of course, are fixed parameters. If this model does not fit, some of these can be set free (if there are good reasons!)
- More constrained models can be fit by imposing equality constraints to test stricter parallel or $\tau$-equivalent models
Multi-factor congeneric models

The covariance matrix \( \Sigma \) of \( x \) is again

\[
\Sigma = \Lambda \Phi \Lambda^T + \Theta
\]

where \( \Phi \) is the covariance matrix of the factors, \( \xi \), and \( \Theta \) is the covariance matrix of the errors, \( \delta \).

- In congeneric models, errors usually assumed to be uncorrelated: \( \Theta = \text{diagonal} \)
- (Some CFA models can allow correlated errors.)
- Model identification: in addition to the \( t \) rule,
  - It is necessary to set the scale for the latent \( \xi \) variables
  - Standardized solution: Set the diagonal entries of \( \Phi \) to 1, so \( \Phi \) is a correlation matrix
  - Reference variable solution: Set the loading \( \lambda_{ij} = 1 \) for one variable \( i \) in each column \( j \)

Example: Ability and Aspiration

Calsyn & Kenny (1971) studied the relation of perceived ability and educational aspiration in 556 white eighth-grade students. Their measures were:

- \( x_1 \): self-concept of ability
- \( x_2 \): perceived parental evaluation
- \( x_3 \): perceived teacher evaluation
- \( x_4 \): perceived friend's evaluation
- \( x_5 \): educational aspiration
- \( x_6 \): college plans

Their interest was primarily in estimating the correlation between “true (perceived) ability” and “true aspiration”.

There is also interest in determining which is the most reliable indicator of each latent variable.

The correlation matrix is shown below:

<table>
<thead>
<tr>
<th></th>
<th>S-C</th>
<th>Par</th>
<th>Tch</th>
<th>Frnd</th>
<th>Educ</th>
<th>Col</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-C Abil</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Par Eval</td>
<td>0.73</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tch Eval</td>
<td>0.70</td>
<td>0.68</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frnd Eval</td>
<td>0.58</td>
<td>0.61</td>
<td>0.57</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educ Asp</td>
<td>0.46</td>
<td>0.43</td>
<td>0.40</td>
<td>0.37</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Col Plan</td>
<td>0.56</td>
<td>0.52</td>
<td>0.48</td>
<td>0.41</td>
<td>0.72</td>
<td>1.00</td>
</tr>
<tr>
<td>x1</td>
<td>x2</td>
<td>x3</td>
<td>x4</td>
<td>x5</td>
<td>x6</td>
<td></td>
</tr>
</tbody>
</table>

The model to be tested is that
- \( x_1 \)-\( x_4 \) measure only the latent “ability” factor and
- \( x_5 \)-\( x_6 \) measure only the “aspiration” factor.
- i.e., two congeneric factors
- If so, are the two factors correlated?
- i.e., what is the true correlation \( \phi_{12} \) between the latent factors?

Specifying the model

The model can be shown as a path diagram:
Specifying the model

This can be cast as the congeneric CFA model:

\[
\begin{pmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4 \\
x_5 \\
x_6
\end{pmatrix} =
\begin{pmatrix}
\lambda_{11} & 0 \\
\lambda_{21} & 0 \\
\lambda_{31} & 0 \\
\lambda_{41} & 0 \\
0 & \lambda_{52} \\
0 & \lambda_{62}
\end{pmatrix}
\begin{pmatrix}
\xi_1 \\
\xi_2
\end{pmatrix} +
\begin{pmatrix}
Z_1 \\
Z_2 \\
Z_3 \\
Z_4 \\
Z_5 \\
Z_6
\end{pmatrix}
\]

If this model fits, the questions of interest can be answered in terms of the estimated parameters of the model:

- Correlation of latent variables: The estimated value of \( \phi_{12} = r(\xi_1, \xi_2) \).
- Reliabilities of indicators: The communality, e.g., \( h_i^2 = \lambda_{ii} \) is the estimated reliability of each measure.

The solution (found with LISREL and PROC CALIS) has an acceptable fit:

\[ \chi^2 = 9.26 \quad df = 8 \quad (p = 0.321) \]

The estimated parameters (standardized solution) are:

<table>
<thead>
<tr>
<th>LAMBDA X</th>
<th>Communality</th>
<th>Uniqueness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability</td>
<td>0.863</td>
<td>0.745</td>
</tr>
<tr>
<td>Aspirtn</td>
<td>0.849</td>
<td>0.721</td>
</tr>
<tr>
<td>Par Eval</td>
<td>0.805</td>
<td>0.648</td>
</tr>
<tr>
<td>Tch Eval</td>
<td>0.695</td>
<td>0.483</td>
</tr>
<tr>
<td>FrndEval</td>
<td>0</td>
<td>0.775</td>
</tr>
<tr>
<td>Educ Asp</td>
<td>0</td>
<td>0.929</td>
</tr>
<tr>
<td>Col Plan</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Thus,

- Self-Concept of Ability is the most reliable measure of \( \xi_1 \), and College Plans is the most reliable measure of \( \xi_2 \).
- The correlation between the latent variables is \( \phi_{12} = 0.67 \). Note that this is higher than any of the individual between-set correlations.

Using PROC CALIS

For SAS, a correlation matrix can be input as follows:

```sas
data calken(TYPE=CORR);
_TYPE_='CORR'; input _NAME_ $ V1-V6; %$
label V1='Self-concept of ability'
V2='Perceived parental evaluation'
V3='Perceived teacher evaluation'
V4='Perceived friends evaluation'
V5='Educational aspiration'
V6='College plans';
datalines;
V1 1. . . . . .
V2 .73 1. . . . .
V3 .70 .68 1. . .
V4 .58 .61 .57 1. .
V5 .46 .43 .40 .37 1.
V6 .56 .52 .48 .41 .72 1.
;```

The CFA model can be specified in several ways:

- With the FACTOR statement, specify names for the free parameters in \( \Lambda \) (MATRIX _F_) and \( \Phi \) (MATRIX _P_)

```sas
proc calis data=calken method=max edf=555 short mod;
factor n=2;
matrix _f_ / * loadings */
[ ,1] = lam1-lam4 , /* factor 1 */
[ ,2] = 4 * 0 lam5 lam6 ; /* factor 2 */
matrix _p_ / * factor correlation */
[1,1] = 2 * 1. ,
[1,2] = COR;
run;
```
Using PROC CALIS

- With the LINEQS statement, specify linear equations for the observed variables, using F1, F2, ... for common factors and E1, E2, ... for unique factors.
- STD statement specifies variances of the factors and errors
- COV statement specifies covariances

```plaintext
proc calis data=calken method=max edf=555;
  LINEQS
    V1 = lam1 F1 + E1 ,
    V2 = lam2 F1 + E2 ,
    V3 = lam3 F1 + E3 ,
    V4 = lam4 F1 + E4 ,
    V5 = lam5 F2 + E5 ,
    V6 = lam6 F2 + E6 ;
  STD
    E1-E6 = EPS: ,
    F1-F2 = 2 * 1. ;
  COV
    F1 F2 = COR ;
run;
```

Using cfa() in the sem package

In addition to specifyEquations(), in the sem package, CFA models are even easier to specify using the `cfa()` function.

```plaintext
library(sem)
mod.calken <- cfa()
F1: v1, v2, v3, v4
F2: v5, v6
fit.calken <- sem(mod.calken, R.calken, N=556)
```

Options allow you to specify reference indicators, and to specify covariances among the factors, allowing the factors to be correlated or uncorrelated.
- By default, all factors in CFA models are allowed to be correlated, simplifying model specification.
- sem includes `edit()` and `update()` functions, allowing you to delete, add, replace, fix, or free a path or parameter in a semmod object.

Example: Speeded and Non-speeded tests

If the measures are cross-classified in two or more ways, it is possible to test equivalence at the level of each way of classification.

Lord (1956) examined the correlations among 15 tests of three types:
- Vocabulary, Figural Intersections, and Arithmetic Reasoning.
- Each test given in two versions: Unspeeded (liberal time limits) and Speeded.

The goal was to identify factors of performance on speeded tests:
- Is speed on cognitive tests a unitary trait?
- If there are several type of speed factors, how are they correlated?
- How highly correlated are speed and power factors on the same test?

Hypothesized factor patterns ($\Lambda$):

1. 3 congeneric sets

\[
\Lambda_{15 \times 3} = \begin{bmatrix}
\beta_1 & 0 & 0 \\
0 & \beta_2 & 0 \\
0 & 0 & \beta_3 \\
\end{bmatrix}
\]

2. 3 congeneric sets + speed factor

\[
\Lambda_{15 \times 4} = \begin{bmatrix}
x & x & x & x \\
x & x & x & x \\
x & x & x & x \\
\end{bmatrix}
\]

Example: Speeded and Non-speeded tests
Example: Speeded and Non-speeded tests

Hypothesized factor patterns ($\Lambda$): Separate unspeeded and speeded factors

\[
\Lambda_{15 \times 6} =
\begin{bmatrix}
V & I & R & V & I & R \\
& X & & X & & X \\
& X & X & & X & X & X \\
Unspeeded & Speeded
\end{bmatrix}
\]

Models:
- (3) parallel: equal $\lambda$ & $\theta$ for each factor
- (4) $\tau$-equivalent: equal $\lambda$ in each col
- (5) congeneric: no equality constraints
- (6) six factors: 3 content, 3 speed

Results:

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Parameters</th>
<th>df</th>
<th>$\chi^2$</th>
<th>$\Delta\chi^2$ (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: 3 congeneric sets</td>
<td>33</td>
<td>87</td>
<td>264.35</td>
<td></td>
</tr>
<tr>
<td>2: 3 sets + speed factor</td>
<td>42</td>
<td>78</td>
<td>140.50</td>
<td>123.85 (9)</td>
</tr>
<tr>
<td>3: 6 sets, parallel</td>
<td>27</td>
<td>93</td>
<td>210.10</td>
<td></td>
</tr>
<tr>
<td>4: 6 sets, $\tau$-equiv.</td>
<td>36</td>
<td>84</td>
<td>138.72</td>
<td>71.45 (9)</td>
</tr>
<tr>
<td>5: 6 sets, congeneric</td>
<td>45</td>
<td>75</td>
<td>120.57</td>
<td>18.15 (9)</td>
</tr>
<tr>
<td>6: 6 factors</td>
<td>45</td>
<td>75</td>
<td>108.37</td>
<td>12.20 (0)</td>
</tr>
</tbody>
</table>

Notes:
- Significant improvement from (1) to (2) $\rightarrow$ speeded tests measure something the unspeeded tests do not.
- $\chi^2$ for (2) still large $\rightarrow$ perhaps there are different kinds of speed factors.
- Big improvement from (3) to (4) $\rightarrow$ not parallel

Higher-order factor analysis

- In EFA & CFA, we often have a model that allows the factors to be correlated ($\Phi \neq I$)
- If there are more than a few factors, it sometimes makes sense to consider a 2nd-order model, that describes the correlations among the 1st-order factors.
- In EFA, this was done simply by doing another factor analysis of the estimated factor correlations $\hat{\Phi}$ from the 1st-order analysis (after an oblique rotation)
- The second stage of development of CFA models was to combine these steps into a single model, and allow different hypotheses to be compared.

Second-order factor analysis: ACOVS model

- Start with a first-order CFA model for the observed variables, $y$ with factors $\eta$
  \[ y = \Lambda_y \eta + \epsilon \]
- Now, consider a 2nd-order model for the correlations among the factors $\eta$
  \[ \eta = \Gamma \xi + \zeta \]
- Combining these equations, we get
  \[ y = \Lambda_y (\Gamma \xi + \zeta) + \epsilon \]
- This is called the ACOVS model, for “analysis of covariance structures” Jöreskog (1970, 1974)
2nd Order models

Second-order factor analysis: ACOVS model

This gives the following model for the covariance matrix $\Sigma$:

$$
\Sigma = \Lambda_y (\Gamma \Phi^T + \Psi) \Lambda_y^T + \Theta_x
$$

where:

- $\Lambda_y(p \times k) = \text{loadings of observed variables on } k \text{ 1st-order factors.}$
- $\Omega(k \times k) = \text{correlations among 1st-order factors.}$
- $\Theta(p \times p) = \text{diagonal matrix of unique variances of 1st-order factors.}$
- $\Gamma(k \times r) = \text{loadings of 1st-order factors on } r \text{ second-order factors.}$
- $\Phi(r \times r) = \text{correlations among 2nd-order factors.}$
- $\Psi = \text{diagonal matrix of unique variances of 2nd-order factors.}$

The model is thus a nesting of a 2nd-order model for $\Gamma$ within the 1st-order model for $\Lambda_y$.

Example: 2nd Order Analysis of Self-Concept Scales

A theoretical model of self-concept by Shavelson & Bolus (1976) describes facets of an individual’s self-concept and presents a hierarchical model of how those facets are arranged. To test this theory, Marsh & Hocevar (1985) analyzed measures of self-concept obtained from 251 fifth grade children with a Self-Description Questionnaire (SDQ). 28 subscales (consisting of two items each) of the SDQ were determined to tap four non-academic and three academic facets of self-concept:

- physical ability
- physical appearance
- relations with peers
- relations with parents
- reading
- mathematics
- general school

Example: 2nd Order Analysis of Self-Concept Scales

The subscales of the SDQ were determined by a first-order exploratory factor analysis. A second-order analysis was carried out examining the correlations among the first-order factors to examine predictions from the Shavelson model(s).

Thurstone data

Data on 9 ability variables:

```r
R.thur <- readMoments(diag=FALSE, names=c('Sentences', 'Vocabulary', 'Sent.Completion', # verbal
                                          'First.Letters', '4.Letter.Words', 'Suffixes', # fluency
                                          'Letter.Series', 'Pedigrees', 'Letter.Group')) # reasoning
```

Thurstone & Thurstone (1941) considered these to measure three factors:

- Verbal Comprehension,
- Word Fluency,
- Reasoning

Figure 1. Higher order factor structures for Models 2 to 6.
sem package: Second-order CFA, Thurstone data

Using the `specifyEquations()` syntax:

```r
mod.thur.eq <- specifyEquations()
  Sentences  = lam11 * F1
  Vocabulary = lam21 * F1
  Sent.Completion = lam31 * F1
  First.Letters = lam41 * F1
  4.Letter.Words = lam51 * F1
  Suffixes = lam62 * F2
  Letter.Series = lam73 * F3
  Pedigrees = lam83 * F3
  Letter.Group = lam93 * F3
  F1 = gam1 * F4 # factor correlations
  F2 = gam2 * F4
  F3 = gam3 * F4
  V(F1) = 1 # factor variances
  V(F2) = 1
  V(F3) = 1
  V(F4) = 1
```

Each line gives a regression equation or the specification of a factor variance (V) or covariance (C)

Fit the model using `sem()`:

```r
(fit.thur <- sem(mod.thur.eq, R.thur, 213))
```

<table>
<thead>
<tr>
<th>Chisquare</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.2</td>
<td>24</td>
<td>0.033101</td>
</tr>
</tbody>
</table>

More detailed output is provided by `summary()`:

```r
summary(sem.thur)
```

<table>
<thead>
<tr>
<th>Model Chisquare</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>38.196</td>
<td>24</td>
<td>0.033101</td>
</tr>
</tbody>
</table>

Path diagram:

```r
pathDiagram(sem.thur, file="sem-thurstone", edge.labels="both")
```

Running `dot -Tpdf -o sem-thurstone.pdf sem-thurstone.dot`

---

The same model can be specified using `cfa()`, designed specially for confirmatory factor models

```r
mod.thur.cfa <- cfa(reference.indicators=FALSE,
                      covs=c("F1", "F2", "F3", "F4"))
```

`mod.thur.cfa`<ref>

- `F1`: Sentences, Vocabulary, Sent.Completion
- `F4`: F1, F2, F3

```r
sem.thur.cfa <- sem(mod.thur.cfa, R.thur, 213)
```
sem package: Other features

- With raw data input, sem provides robust estimates of standard errors and robust tests
- Can accommodate missing data, via full-information maximum likelihood (FIML)
- miSem() generates multiple imputations of missing data using the mi package
- bootSem() provides nonparametric bootstrap estimates by independent random sampling
- A given model can be easily modified via edit() and update() methods
- Multiple-group analyses and tests of factorial invariance: multigroupModel().
- Related: semPlot: lovely, flexible, pub. quality path diagrams

Path diagram from semPlot

```r
library(semPlot)
semPaths(sem.thur, what="std", color=list(man="lightblue", lat="pink"),
nCharNodes=6, sizeMan=6, edge.color="black")
title("Thurstone 2nd Order Model, Standardized estimates", cex=1.5)
```

Factorial Invariance

Multi-sample analyses:

- When a set of measures have been obtained from samples from several populations, we often wish to study the similarities in factor structure across groups.
- The CFA/SEM model allows any parameter to be assigned an arbitrary fixed value, or constrained to be equal to some other parameter. Constraints across groups provide the way to test these models.
- We can test any degree of invariance from totally separate factor structures to completely invariant ones.

Model

Let \( x_g \) be the vector of tests administered to group \( g, g = 1, 2, \ldots, m \), and assume that a factor analysis model holds in each population with some number of common factors, \( k_g \).

\[
\Sigma_g = \Lambda_g \Phi \Lambda_g^T + \Psi_g
\]

Factorial Invariance: Examples

- Arguably among the most important recent development in personality psychology is the idea that individual differences in personality characteristics is organized into five main trait domains: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness
  - One widely used instrument is the 60-item NEO-Five factor inventory (Costa & McCrae, 1992), developed and analyzed for a North American, English-speaking population
  - To what extent does the same factor structure apply across gender?
  - To what extent does the same factor structure applies in other cultural and language groups?
- The emerging field of cross-cultural psychology offers many similar examples.
Factorial Invariance: Hypotheses

We can examine a number of different hypotheses about how “similar” the covariance structure is across groups.

**Hypotheses**
- Can we simply pool the data over groups?
- If not, can we say that the same number of factors apply in all groups?
- If so, are the factor loadings equal over groups?
- What about factor correlations and unique variances?

**Software**
- LISREL, AMOS, and M Plus all provide convenient ways to do multi-sample analysis.
- PROC CALIS in SAS 9.3 does too.
- In R, the lavaan package provides multi-sample analysis and the measurementInvariance() function. The sem package includes a multigroupModel() for such models.

### Equality of Covariance Matrices

Equality of Covariance Matrices

$$H_{\Sigma} : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_m$$

- If this hypothesis is tenable, there is no need to analyse each group separately or test further for differences among them: Simply pool all the data, and do one analysis!
- If we reject $$H_{\Sigma}$$, we may wish to test a less restrictive hypothesis that posits some form of invariance.
- The test statistic for $$H_{\Sigma}$$ is Box’s test,

$$\chi^2_\Sigma = n \log |S| - \sum_{g=1}^{m} n_g \log |S_g|$$

which is distributed approx. as $$\chi^2$$ with $$d_{\Sigma} = (m - 1)p(p - 1)/2$$ df.

(This test can be carried out in SAS with PROC DISCRIM using the POOL=TEST option)

### Same number of factors (Configural invariance)

The least restrictive form of “invariance” is simply that the number of factors is the same in each population:

$$H_k : k_1 = k_2 = \cdots = k_m = \text{a specified value}, k$$

- This can be tested by doing an unrestricted factor analysis for $$k$$ factors on each group separately, and summing the $$\chi^2$$’s and degrees of freedom,

$$\chi^2_k = \sum_{g} \chi^2_k(g) \quad d_k = m \times [(p - k)^2 - (p + k)]/2$$

- If this hypothesis is rejected, there is no sense in testing more restrictive models

### Same factor pattern (Weak invariance)

If the hypothesis of a common number of factors is tenable, one may proceed to test the hypothesis of an invariant factor pattern:

$$H_\Lambda : \Lambda_1 = \Lambda_2 = \cdots = \Lambda_m$$

- The common factor pattern $$\Lambda$$ may be either completely unspecified, or be specified to have zeros in certain positions.
- To obtain a $$\chi^2$$ for this hypothesis, estimate $$\Lambda$$ (common to all groups), plus $$\Phi_1, \Phi_2, \ldots, \Phi_m$$, and $$\psi_1, \psi_2, \ldots, \psi_m$$, yielding a minimum value of the function, $$F$$. Then, $$\chi^2_\Lambda = 2 \times F_{\text{min}}$$.
- To test the hypothesis $$H_\Lambda$$, given that the number of factors is the same in all groups, use

$$\chi^2_{\Lambda|k} = \chi^2_\Lambda - \chi^2_k$$ with $$d_{\Lambda|k} = d_\Lambda - d_k$$ degrees of freedom
Factorial invariance

Hypotheses

The following hypotheses were tested:

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Model specifications</th>
</tr>
</thead>
</table>
| A. $H_S : \Sigma_1 = \Sigma_2$ | $\Lambda_1 = \Lambda_2 = I_{(4 \times 4)}$
$\Psi_1 = \Psi_2 = 0_{(4 \times 4)}$
$\Phi_1 = \Phi_2$ constrained, free |
| B. $H_{k=2} : \Sigma_1, \Sigma_2$ both fit with $k = 2$ correlated factors | $\Lambda_1 = \Lambda_2 = \begin{bmatrix} x & 0 \\ x & 0 \\ 0 & x \\ 0 & x \end{bmatrix}$
$\Phi_1, \Phi_2, \Psi_1, \Psi_2$ free |
| C. $H_{\Lambda} : H_{k=2} \& \Lambda_1 = \Lambda_2$ | $\Lambda_1 = \Lambda_2$ (constrained) |
| D. $H_{\Lambda, \Psi} : H_{\Lambda} \& \Psi_1 = \Psi_2$ | $\Psi_1 = \Psi_2$ (constrained)
$\Lambda_1 = \Lambda_2$ |
| E. $H_{\Lambda, \Theta, \phi} : H_{\Lambda, \Theta} \& \Phi_1 = \Phi_2$ | $\Phi_1 = \Phi_2$ (constrained)
$\Psi_1 = \Psi_2$
$\Lambda_1 = \Lambda_2$ |

Analysis

The analysis was carried out with both LISREL and AMOS. AMOS is particularly convenient for multi-sample analysis, and for testing a series of nested hypotheses.

Summary of Hypothesis Tests for Factorial Invariance

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Overall fit</th>
<th>Group A</th>
<th>Group N-A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>df</td>
<td>prob</td>
</tr>
<tr>
<td>A: $H_S$</td>
<td>38.08</td>
<td>10</td>
<td>.000</td>
</tr>
<tr>
<td>B: $H_{k=2}$</td>
<td>1.52</td>
<td>2</td>
<td>.468</td>
</tr>
<tr>
<td>C: $H_{\Lambda}$</td>
<td>8.77</td>
<td>4</td>
<td>.067</td>
</tr>
<tr>
<td>D: $H_{\Lambda, \Psi}$</td>
<td>21.55</td>
<td>8</td>
<td>.006</td>
</tr>
<tr>
<td>E: $H_{\Lambda, \Theta, \phi}$</td>
<td>38.22</td>
<td>11</td>
<td>.000</td>
</tr>
</tbody>
</table>

- The hypothesis of equal factor loadings ($H_{\Lambda}$) in both samples is tenable.
- Unique variances appear to differ in the two samples.
- The factor correlation ($\phi_{12}$) appears to be greater in the Academic sample than in the Non-Academic sample.
Data

Data for Academic and Non-academic boys:

```r
library(sem)
Sorbom.acad <- read.moments(diag=TRUE,
  names=c('Read.Gr5', 'Writ.Gr5', 'Read.Gr7', 'Writ.Gr7'))
281.349 184.219 182.821 216.739 171.699 283.289
198.376 153.201 208.837 246.069
Sorbom.nonacad <- read.moments(diag=TRUE,
  names=c('Read.Gr5', 'Writ.Gr5', 'Read.Gr7', 'Writ.Gr7'))
174.485 141.869 134.468 118.836 228.449
129.840 97.767 136.058 180.460
# make the two matrices into a list
Sorbom <- list(acad=Sorbom.acad, nonacad=Sorbom.nonacad)
```

Tests of measurement invariance I

Test all models of measurement invariance:

```r
library(semTools)
measurementInvariance(Sorbom.model, sample.cov=Sorbom,
  sample.nobs=c(373,249))
```

Tests of measurement invariance II

A fourth model also tests equality of means, but means are not available for this example.
Summary

- **measurement error** reduces precision, but worse— introduces **bias**
- CFA & SEM use latent variables in a **measurement model** to allow for this

\[
x = \Lambda\xi + \delta \quad \Rightarrow \quad \Sigma = \Lambda\Phi\Lambda^T + \Theta
\]

- One-factor models allow for testing various forms of “equivalence” within the SEM framework
  - An essential idea in CFA is allowing for **free** and **fixed** parameters and equality constraints
  - These ideas extend directly to more complex models, with multiple factors of possibly different types
- Higher-order CFA models take this a step further, allowing a factor structure for the 1st-order factors
- Multiple-group models allow for testing a variety of **measurement invariance** models