Title

example 3 — Two-factor measurement model

Description Remarks and examples References Also see

Description

The multiple-factor measurement model is demonstrated using summary statistics dataset (SSD) `sem_2fmm.dta`:

```
. use http://www.stata-press.com/data/r15/sem_2fmm
(Affective and cognitive arousal)

. ssd describe

  Summary statistics data from
  http://www.stata-press.com/data/r15/sem_2fmm.dta
    obs:              216                Affective and cognitive arousal
   vars:               10                25 May 2016 10:11
                                         (_dta has notes)
  ------------------------------------------------------------------------

  variable name                 variable label
  ------------------------------------------------------------------------

  a1                            affective arousal 1
  a2                            affective arousal 2
  a3                            affective arousal 3
  a4                            affective arousal 4
  a5                            affective arousal 5
  c1                            cognitive arousal 1
  c2                            cognitive arousal 2
  c3                            cognitive arousal 3
  c4                            cognitive arousal 4
  c5                            cognitive arousal 5
  ------------------------------------------------------------------------

. notes

_dta:
  1.  Summary statistics data containing published covariances from Thomas O.
      Williams, Ronald C. Eaves, and Cynthia Cox, 2 Apr 2002, "Confirmatory
      factor analysis of an instrument designed to measure affective and
      cognitive arousal", _Educational and Psychological Measurement_, vol. 62
      no. 2, 264-283.
  2.  a1-a5 report scores from 5 miniscales designed to measure affective
      arousal.
  3.  c1-c5 report scores from 5 miniscales designed to measure cognitive
      arousal.
  4.  The series of tests, known as the VST II (Visual Similes Test II) were
      administered to 216 children ages 10 to 12.  The miniscales are sums of
      scores of 5 to 6 items in VST II.
```

See [SEM] **example 2** to learn how we created this SSD.

Remarks and examples

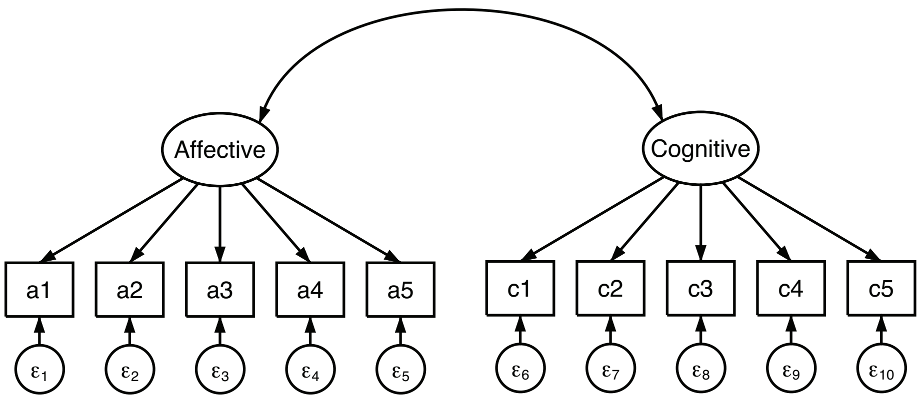
stata.com

Remarks are presented under the following headings:

See *Multiple-factor measurement models* in [SEM] **intro 5** for background.

Fitting multiple-factor measurement models

Below we fit the model shown by Kline (2005, 70–74, 184), namely,




```
. sem (Affective -> a1 a2 a3 a4 a5) (Cognitive -> c1 c2 c3 c4 c5)
Endogenous variables
Measurement:  a1 a2 a3 a4 a5 c1 c2 c3 c4 c5
Exogenous variables
Latent:       Affective Cognitive
Fitting target model:
Iteration 0:   log likelihood = -9542.8803
Iteration 1:   log likelihood = -9539.5505
Iteration 2:   log likelihood = -9539.3856
Iteration 3:   log likelihood = -9539.3851
```

```
Structural equation model                       Number of obs      =        216
Estimation method  = ml
Log likelihood     = -9539.3851

 ( 1)  [a1]Affective = 1
 ( 2)  [c1]Cognitive = 1

------------------------------------------------------------------------------
             |                 OIM
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Measurement  |
  a1         |
   Affective |          1  (constrained)
  -----------+----------------------------------------------------------------
  a2         |
   Affective |   .9758098   .0460752    21.18   0.000      .885504    1.066116
  -----------+----------------------------------------------------------------
  a3         |
   Affective |   .8372599   .0355086    23.58   0.000     .7676643    .9068556
  -----------+----------------------------------------------------------------
  a4         |
   Affective |   .9640461   .0499203    19.31   0.000      .866204    1.061888
  -----------+----------------------------------------------------------------
  a5         |
   Affective |   1.063701   .0435751    24.41   0.000     .9782951    1.149107
  -----------+----------------------------------------------------------------
  c1         |
   Cognitive |          1  (constrained)
  -----------+----------------------------------------------------------------
  c2         |
   Cognitive |   1.114702   .0655687    17.00   0.000     .9861901    1.243215
  -----------+----------------------------------------------------------------
  c3         |
   Cognitive |   1.329882   .0791968    16.79   0.000     1.174659    1.485105
  -----------+----------------------------------------------------------------
  c4         |
   Cognitive |   1.172792   .0711692    16.48   0.000     1.033303    1.312281
  -----------+----------------------------------------------------------------
  c5         |
   Cognitive |   1.126356   .0644475    17.48   0.000     1.000041    1.252671
-------------+----------------------------------------------------------------
    var(e.a1)|   384.1359   43.79119                      307.2194    480.3095
    var(e.a2)|   357.3524   41.00499                      285.3805    447.4755
    var(e.a3)|   154.9507   20.09026                      120.1795    199.7822
    var(e.a4)|   496.4594   54.16323                      400.8838    614.8214
    var(e.a5)|   191.6857   28.07212                      143.8574    255.4154
    var(e.c1)|   171.6638   19.82327                       136.894    215.2649
    var(e.c2)|   171.8055   20.53479                      135.9247    217.1579
    var(e.c3)|   276.0144   32.33535                      219.3879    347.2569
    var(e.c4)|   224.1994   25.93412                      178.7197    281.2527
    var(e.c5)|   146.8655    18.5756                      114.6198    188.1829
var(Affect~e)|   1644.463   193.1032                      1306.383    2070.034
var(Cognit~e)|   455.9349   59.11245                      353.6255    587.8439
-------------+----------------------------------------------------------------
cov(Affect~e,|
   Cognitive)|   702.0736   85.72272     8.19   0.000     534.0601     870.087
------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(34)  =     88.88, Prob > chi2 = 0.0000
```

Notes:

1. In [SEM] **example 1**, we ran `sem` on raw data. In this example, we run `sem` on SSD. There are no special `sem` options that we need to specify because of this.
2. The estimated coefficients reported above are unstandardized coefficients or, if you prefer, factor loadings.
3. The coefficients listed at the bottom of the coefficient table that start with `e.` are the estimated error variances. They represent the variance of the indicated measurement that is not measured by the respective latent variables.
4. The above results do not match exactly (Kline 2005, 184). If we specified `sem` option `nm1`, results are more likely to match to 3 or 4 digits. The `nm1` option says to divide by $N-1$ rather than by N in producing variances and covariances.

Displaying standardized results

The output will be easier to interpret if we display standardized values for paths rather than path coefficients. A standardized value is in standard deviation units. It is the change in one variable given a change in another, both measured in standard deviation units. We can obtain standardized values by specifying `sem`'s `standardized` option, which we can do when we fit the model or when we replay results:

```
. sem, standardized
Structural equation model                       Number of obs      =        216
Estimation method  = ml
Log likelihood     = -9539.3851
 ( 1)  [a1]Affective = 1
 ( 2)  [c1]Cognitive = 1
```

Standardized	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Measurement						
a1						
Affective	.9003553	.0143988	62.53	0.000	.8721342	.9285765
a2						
Affective	.9023249	.0141867	63.60	0.000	.8745195	.9301304
a3						
Affective	.9388883	.0097501	96.29	0.000	.9197784	.9579983
a4						
Affective	.8687982	.0181922	47.76	0.000	.8331421	.9044543
a5						
Affective	.9521559	.0083489	114.05	0.000	.9357923	.9685195
c1						
Cognitive	.8523351	.0212439	40.12	0.000	.8106978	.8939725
c2						
Cognitive	.8759601	.0184216	47.55	0.000	.8398544	.9120658
c3						
Cognitive	.863129	.0199624	43.24	0.000	.8240033	.9022547
c4						
Cognitive	.8582786	.0204477	41.97	0.000	.8182018	.8983554
c5						
Cognitive	.8930346	.0166261	53.71	0.000	.8604479	.9256212
var(e.a1)	.1893602	.0259281			.1447899	.2476506
var(e.a2)	.1858097	.0256021			.1418353	.2434179
var(e.a3)	.1184887	.0183086			.0875289	.1603993
var(e.a4)	.2451896	.0316107			.1904417	.3156764
var(e.a5)	.0933991	.015899			.0669031	.1303885
var(e.c1)	.2735248	.0362139			.2110086	.354563
var(e.c2)	.2326939	.0322732			.1773081	.3053806
var(e.c3)	.2550083	.0344603			.1956717	.3323385
var(e.c4)	.2633578	.0350997			.2028151	.3419733
var(e.c5)	.2024893	.0296954			.1519049	.2699183
var(Affect~e)	1	.			.	.
var(Cognit~e)	1	.			.	.
cov(Affect~e, Cognitive)	.8108102	.0268853	30.16	0.000	.758116	.8635045

```
LR test of model vs. saturated: chi2(34)  =     88.88, Prob > chi2 = 0.0000
```

Notes:

1. In addition to obtaining standardized coefficients, the `standardized` option reports estimated error variances as the fraction of the variance that is unexplained. Error variances were previously unintelligible numbers such as 384.136 and 357.352. Now they are 0.189 and 0.186.
2. Also listed in the `sem` output are variances of latent variables. In the previous output, latent variable `Affective` had variance 1,644.46 with standard error 193. In the standardized output, it has variance 1 with standard error missing. The variances of the latent variables are standardized to 1, and obviously, being a normalization, there is no corresponding standard error.
3. We can now see at the bottom of the coefficient table that affective and cognitive arousal are correlated 0.81 because standardized covariances are correlation coefficients.
4. The standardized coefficients for this model can be interpreted as the correlation coefficients between the indicator and the latent variable because each indicator measures only one factor. For instance, the standardized path coefficient `a1<-Affective` is 0.90, meaning the correlation between `a1` and `Affective` is 0.90.

Fitting the model with the Builder

Use the diagram above for reference.

1. Open the dataset.

   In the Command window, type

   ```
   . use http://www.stata-press.com/data/r15/sem_2fmm
   ```

2. Open a new Builder diagram.

   Select menu item **Statistics > SEM (structural equation modeling) > Model building and estimation**.

3. Change the size of the observed variables' rectangles.

   From the SEM Builder menu, select **Settings > Variables > All observed...**.

   In the resulting dialog box, change the first size to `.38` and click on **OK**.

4. Create the measurement component for affective arousal.

   Select the Add measurement component tool, , and then click in the diagram about one-third of the way down from the top and one-fourth of the way in from the left.

   In the resulting dialog box,

   a. change the *Latent variable name* to `Affective`;

   b. select `a1`, `a2`, `a3`, `a4`, and `a5` by using the *Measurement variables* control;

   c. select `Down` in the *Measurement direction* control;

   d. click on **OK**.

   If you wish, move this component by clicking on any variable and dragging it.

5. Create the measurement component for cognitive arousal.

   Repeat the process from item 4, but place the measurement component about one-third of the way down from the top and three-fourths of the way in from the left. Label the latent variable `Cognitive`, and select measurement variables `c1`, `c2`, `c3`, `c4`, and `c5`. Drag to reposition if desired.

6. Correlate the latent factors.

   a. Select the Add covariance tool, .

   b. Click in the upper-right quadrant of the `Affective` oval (it will highlight when you hover over it), and drag a covariance to the upper-left quadrant of the `Cognitive` oval (it will highlight when you can release to connect the covariance).

7. Clean up.

   If you do not like where a covariance has been connected to its variable, use the Select tool, , to click on the covariance, and then simply click on where it connects to an oval and drag the endpoint. You can also change the bow of the covariance by dragging the control point that extends from one end of the selected covariance.

8. Estimate.

   Click on the **Estimate** button, , in the Standard Toolbar, and then click on **OK** in the resulting *SEM estimation options* dialog box.

9. Show standardized estimates.

   From the SEM Builder menu, select **View > Standardized estimates**.

You can open a completed diagram in the Builder by typing

```
. webgetsem sem_2fmm
```

Obtaining equation-level goodness of fit by using estat eqgof

That the correlation between `a1` and `Affective` is 0.90 implies that the fraction of the variance of `a1` explained by `Affective` is $0.90^2 = 0.81$, and left unexplained is $1 - 0.81 = 0.19$. Instead of manually calculating the proportion of variance explained by indicators, we can use the `estat eqgof` command:

```
. estat eqgof
Equation-level goodness of fit
```

depvars	fitted	Variance predicted	residual	R-squared	mc	mc2
observed						
a1	2028.598	1644.463	384.1359	.8106398	.9003553	.8106398
a2	1923.217	1565.865	357.3524	.8141903	.9023249	.8141903
a3	1307.726	1152.775	154.9507	.8815113	.9388883	.8815113
a4	2024.798	1528.339	496.4594	.7548104	.8687982	.7548104
a5	2052.328	1860.643	191.6857	.9066009	.9521559	.9066009
c1	627.5987	455.9349	171.6638	.7264752	.8523351	.7264752
c2	738.3325	566.527	171.8055	.7673061	.8759601	.7673061
c3	1082.374	806.3598	276.0144	.7449917	.863129	.7449917
c4	851.311	627.1116	224.1994	.7366422	.8582786	.7366422
c5	725.3002	578.4346	146.8655	.7975107	.8930346	.7975107
overall				.9949997		

```
mc  = correlation between depvar and its prediction
mc2 = mc^2 is the Bentler-Raykov squared multiple correlation coefficient
```

Notes:

1. `fitted` reports the fitted variance of each of the endogenous variables, whether observed or latent. In this case, we have observed endogenous variables.
2. `predicted` reports the variance of the predicted value of each endogenous variable.
3. `residual` reports the leftover residual variance.
4. `R-squared` reports R^2, the fraction of variance explained by each indicator. The fraction of the variance of `Affective` explained by `a1` is 0.81, just as we calculated by hand at the beginning of this section. The overall R^2 is also called the coefficient of determination.
5. `mc` stands for multiple correlation, and `mc2` stands for multiple-correlation squared. `R-squared`, `mc`, and `mc2` all report the relatedness of the indicated dependent variable with the model's linear prediction. In recursive models, all three statistics are really the same number. `mc` is equal to the square root of `R-squared`, and `mc2` is equal to `R-squared`.

   In nonrecursive models, these three statistics are different and each can have problems. `R-squared` and `mc` can actually become negative! That does not mean the model has negative predictive power or that it might not even have reasonable predictive power. `mc2` = `mc`^{2} is recommended by Bentler and Raykov (2000) to be used instead of `R-squared` for nonrecursive systems.

In [SEM] **example 4**, we examine the goodness-of-fit statistics for this model.

In [SEM] **example 5**, we examine modification indices for this model.

References

Acock, A. C. 2013. *Discovering Structural Equation Modeling Using Stata*. Rev. ed. College Station, TX: Stata Press.

Bentler, P. M., and T. Raykov. 2000. On measures of explained variance in nonrecursive structural equation models. *Journal of Applied Psychology* 85: 125–131.

Kline, R. B. 2005. *Principles and Practice of Structural Equation Modeling*. 2nd ed. New York: Guilford Press.

Also see

[SEM] **example 1** — Single-factor measurement model

[SEM] **example 2** — Creating a dataset from published covariances

[SEM] **example 20** — Two-factor measurement model by group

[SEM] **example 26** — Fitting a model with data missing at random

[SEM] **example 31g** — Two-factor measurement model (generalized response)

[SEM] **sem** — Structural equation model estimation command

[SEM] **estat eqgof** — Equation-level goodness-of-fit statistics