Title

stata.com

example 26 — Fitting a model with data missing at random

Description Remarks and examples Also see

Description

`sem method(mlmv)` is demonstrated using

```
. use http://www.stata-press.com/data/r13/cfa_missing
(CFA MAR data)
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
id	500	250.5	144.4818	1	500
test1	406	97.37475	13.91442	56.0406	136.5672
test2	413	98.04501	13.84145	62.25496	129.3881
test3	443	100.9699	13.4862	65.51753	137.3046
test4	417	99.56815	14.25438	53.8719	153.9779
taken	500	3.358	.6593219	2	4

```
. notes
_dta:
  1.  Fictional data on 500 subjects taking four tests.
  2.  Tests results M.A.R. (missing at random).
  3.  230 took all 4 tests
  4.  219 took 3 of the 4 tests
  5.  51 took 2 of the 4 tests
  6.  All tests have expected mean 100, s.d. 14.
```

See [SEM] **intro 4** for background.

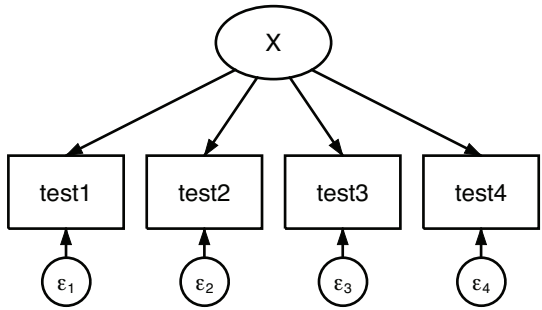
Remarks and examples

stata.com

Remarks are presented under the following headings:

Fitting the model with method(ml)

We fit a single-factor measurement model.




```
. sem (test1 test2 test3 test4 <- X), nolog
(270 observations with missing values excluded)
Endogenous variables
Measurement:  test1 test2 test3 test4
Exogenous variables
Latent:       X
Structural equation model                       Number of obs      =       230
Estimation method  = ml
Log likelihood     = -3464.3099
 ( 1)  [test1]X = 1
```

	OIM Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Measurement						
test1 <-						
X	1	(constrained)				
_cons	96.76907	.8134878	118.96	0.000	95.17467	98.36348
test2 <-						
X	1.021885	.1183745	8.63	0.000	.789875	1.253895
_cons	92.41248	.8405189	109.95	0.000	90.7651	94.05987
test3 <-						
X	.5084673	.0814191	6.25	0.000	.3488889	.6680457
_cons	94.12958	.7039862	133.71	0.000	92.7498	95.50937
test4 <-						
X	.5585651	.0857772	6.51	0.000	.3904449	.7266853
_cons	92.2556	.7322511	125.99	0.000	90.82042	93.69079
var(e.test1)	55.86083	10.85681			38.16563	81.76028
var(e.test2)	61.88092	11.50377			42.985	89.08338
var(e.test3)	89.07839	8.962574			73.13566	108.4965
var(e.test4)	93.26508	9.504276			76.37945	113.8837
var(X)	96.34453	16.28034			69.18161	134.1725

```
LR test of model vs. saturated: chi2(2)   =      0.39, Prob > chi2 = 0.8212
```

Notes:

1. This model was fit using 230 of the 500 observations in the dataset. Unless you use `sem`'s `method(mlmv)`, observations are casewise omitted, meaning that if there is a single variable with a missing value among the variables being used, the observation is ignored.
2. The coefficients for `test3` and `test4` are 0.51 and 0.56. Because we at StataCorp manufactured these data, we can tell you that the true coefficients are 1.
3. The error variance for `e.test1` and `e.test2` are understated. These data were manufactured with an error variance of 100.
4. These data are missing at random (MAR), not missing completely at random (MCAR). In MAR data, which values are missing can be a function of the observed values in the data. MAR data can produce biased estimates if the missingness is ignored, as we just did. MCAR data do not bias estimates.

Fitting the model with method(mlmv)

```
. sem (test1 test2 test3 test4 <- X), method(mlmv) nolog
Endogenous variables
Measurement:  test1 test2 test3 test4
Exogenous variables
Latent:       X
  (output omitted)
Structural equation model                       Number of obs      =       500
Estimation method  = mlmv
Log likelihood     = -6592.9961
 ( 1)  [test1]X = 1
```

	OIM Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Measurement						
test1 <-						
X	1	(constrained)				
_cons	98.94386	.6814418	145.20	0.000	97.60826	100.2795
test2 <-						
X	1.069952	.1079173	9.91	0.000	.8584378	1.281466
_cons	99.84218	.6911295	144.46	0.000	98.48759	101.1968
test3 <-						
X	.9489025	.0896098	10.59	0.000	.7732706	1.124534
_cons	101.0655	.6256275	161.54	0.000	99.83928	102.2917
test4 <-						
X	1.021626	.0958982	10.65	0.000	.8336687	1.209583
_cons	99.64509	.6730054	148.06	0.000	98.32603	100.9642
var(e.test1)	101.1135	10.1898			82.99057	123.1941
var(e.test2)	95.45572	10.79485			76.47892	119.1413
var(e.test3)	95.14847	9.053014			78.9611	114.6543
var(e.test4)	101.0943	10.0969			83.12124	122.9536
var(X)	94.04629	13.96734			70.29508	125.8225

```
LR test of model vs. saturated: chi2(2)   =      2.27, Prob > chi2 = 0.3209
```

Notes:

1. The model is now fit using all 500 observations in the dataset.
2. The coefficients for `test3` and `test4`—previously 0.51 and 0.56—are now 0.95 and 1.02.
3. Error variance estimates are now consistent with the true value of 100.
4. Standard errors of path coefficients are mostly smaller than reported in the previous model.
5. `method(mlmv)` requires that the data be MCAR or MAR.
6. `method(mlmv)` requires that the data be multivariate normal.

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/r13/cfa_missing
   ```

2. Open a new Builder diagram.

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

3. Create the measurement component for `X`.

   Select the Add Measurement Component tool,  , and then click in the diagram about one-third of the way down from the top and about halfway in from the left.

   In the resulting dialog box,

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

   b. select `test1`, `test2`, `test3`, and `test4` by using the *Measurement variables* control;

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

   d. click on **OK**.

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

4. Estimate.

   Click on the **Estimate** button,  , in the Standard Toolbar. In the resulting dialog box,

   a. select the *Model* tab;

   b. select the *Maximum likelihood with missing values* radio button;

   c. click on **OK**.

You can open a completed diagram in the Builder by typing

```
. webgetsem cfa_missing
```

Also see

[SEM] **intro 4** — Substantive concepts

[SEM] **sem option method()** — Specifying method and calculation of VCE