

**Example 16 — Correlation**

[Description](#)[Remarks and examples](#)[Also see](#)

## Description

`sem` can be used to produce correlations or covariances between exogenous variables. The advantages of using `sem` over Stata's `correlate` command are that you can perform statistical tests on the results and that you can handle missing values in a more elegant way.

To demonstrate these features, we use

```
. use https://www.stata-press.com/data/r18/census13
(1980 Census data by state)
. describe
Contains data from https://www.stata-press.com/data/r18/census13.dta
Observations:      50          1980 Census data by state
Variables:         9           2 Dec 2022 14:01
```

Variable name	Storage type	Display format	Value label	Variable label
state	byte	%13.0g	state1	State
brate	int	%10.0g		Birthrate
pop	long	%12.0gc		Population
medage	float	%9.2f		Median age
division	byte	%8.0g	division	Census division
region	byte	%-8.0g	cenreg	Census region
mrgrate	float	%9.0g		Marriage rate
dvcrate	float	%9.0g		Divorce rate
medagesq	float	%9.0g		

Sorted by:

See [Correlations](#) in [\[SEM\] Intro 5](#) for background.

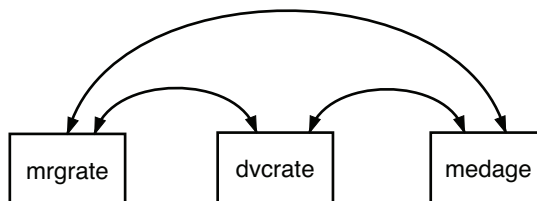
## Remarks and examples

Remarks are presented under the following headings:

- [Using sem to obtain correlation matrices](#)
- [Fitting the model with the Builder](#)
- [Testing correlations with `estat stdize` and `test`](#)

### Using sem to obtain correlation matrices

We fit the following model:



## 2 Example 16 — Correlation

This model does nothing more than estimate the covariances (correlations), something we could obtain from the `correlate` command by typing

```
. correlate mrgrate dvcrate medage
(obs=50)
```

	mrgrate	dvcrate	medage
mrgrate	1.0000		
dvcrate	0.7700	1.0000	
medage	-0.0177	-0.2229	1.0000

```
. correlate mrgrate dvcrate medage, covariance
(obs=50)
```

	mrgrate	dvcrate	medage
mrgrate	.000662		
dvcrate	.000063	1.0e-05	
medage	-.000769	-.001191	2.86775

As explained in *Correlations* in [SEM] **Intro 5**, to see results presented as correlations rather than as covariances, we specify `sem`'s `standardized` option:

```
. sem ( <- mrgrate dvcrate medage), standardized
```

Exogenous variables

Observed: mrgrate dvcrate medage

Fitting target model:

Iteration 0: Log likelihood = 258.58985

Iteration 1: Log likelihood = 258.58985

Structural equation model

Number of obs = 50

Estimation method: ml

Log likelihood = 258.58985

Standardized	OIM		z	P> z	[95% conf. interval]	
	Coefficient	std. err.				
mean(mrgrate)	.7332509	.1593002	4.60	0.000	.4210282	1.045474
mean(dvcrate)	2.553791	.291922	8.75	0.000	1.981634	3.125947
mean(medage)	17.62083	1.767749	9.97	0.000	14.15611	21.08556
var(mrgrate)	1	.			.	.
var(dvcrate)	1	.			.	.
var(medage)	1	.			.	.
cov(mrgrate, dvcrate)	.7699637	.0575805	13.37	0.000	.6571079	.8828195
cov(mrgrate, medage)	-.0176541	.1413773	-0.12	0.901	-.2947485	.2594403
cov(dvcrate, medage)	-.222932	.1343929	-1.66	0.097	-.4863373	.0404732

LR test of model vs. saturated:  $\chi^2(0) = 0.00$

Prob >  $\chi^2 = .$

Note:

1. The correlations reported are

	sem	correlate
mrgrate and dvcrate	0.7699637	0.7700
mrgrate and medage	-0.0176541	-0.0177
dvcrate and medage	-0.222932	-0.2229

## Fitting the model with the Builder

Use the diagram above for reference.

1. Open the dataset.


In the Command window, type

```
. use https://www.stata-press.com/data/r18/census13
```

2. Open a new Builder diagram.

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

3. Create the set of observed variables.

Select the Add observed variables set tool, , and then click in the diagram about halfway down from the top and a quarter of the way in from the left.


In the resulting dialog box,

- a. select the *Select variables* radio button (it may already be selected);
- b. use the *Variables* control to select the variables in this order: `mrgrate`, `dvcrate`, and `medage`;
- c. select **Horizontal** in the *Orientation* control;
- d. select the **Distances** tab;
- e. select `.5` (inch) in the *Distance between variables* control;
- f. click on **OK**.


If you wish, move the set of variables by clicking on any variable and dragging it.

Be sure you select the observed variables in the order indicated above; otherwise, the instructions below for creating covariances will not be correct.


4. Correlate each pair of variables.

- a. Select the Add covariance tool, .
- b. Click in the top of the `mrgrate` rectangle, slightly to the right of the center (it will highlight when you hover over it), and drag a path to the top of the `dvcrate` rectangle, slightly to the left of the center (it will highlight when you can release to connect the covariance).
- c. Click in the top of the `dvcrate` rectangle, slightly to the right of the center, and drag a path to the top of the `medage` rectangle, slightly to the left of the center.
- d. Click in the top of the `mrgrate` rectangle, slightly to the left of the center, and drag a path to the top of the `medage` rectangle, slightly to the right of the center.

## 5. 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.

## 6. Estimate.

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

## 7. 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_corr
```

## Testing correlations with estat stdize and test

We can test whether the correlations between median age and marriage and divorce rates are equal with `test` by typing

```
. estat stdize: ///
      test _b[/cov(medage,mrgrate)] = _b[/cov(medage,dvcrate)]
```

We must prefix `test` with `estat stdize` because otherwise we would be testing equality of covariances; see *Displaying other results, statistics, and tests (sem and gsem)* in [SEM] **Intro 7** and see [SEM] `estat stdize`.

That we refer to the two correlations (covariances) by typing `_b[/cov(medage,mrgrate)]` and `_b[/cov(medage,dvcrate)]` is something nobody remembers and that we remind ourselves of by redisplaying `sem` results with the `coeflegend` option:

```
. sem, coeflegend
Structural equation model                Number of obs = 50
Estimation method: ml
Log likelihood = 258.58985
```

	Coefficient	Legend
mean(mrgrate)	.0186789	_b[/mean(mrgrate)]
mean(dvcrate)	.0079769	_b[/mean(dvcrate)]
mean(medage)	29.54	_b[/mean(medage)]
var(mrgrate)	.0006489	_b[/var(mrgrate)]
var(dvcrate)	9.76e-06	_b[/var(dvcrate)]
var(medage)	2.8104	_b[/var(medage)]
cov(mrgrate, dvcrate)	.0000613	_b[/cov(mrgrate,dvcrate)]
cov(mrgrate, medage)	-.0007539	_b[/cov(mrgrate,medage)]
cov(dvcrate, medage)	-.0011674	_b[/cov(dvcrate,medage)]

```
LR test of model vs. saturated: chi2(0) = 0.00
```

```
Prob > chi2 = .
```

We can now obtain the test:

```
. estat stdize:
>      test _b[/cov(medage,mrgrate)] = _b[/cov(medage,dvcrate)]
( 1)  [/]cov(mrgrate,medage) - [/]cov(dvcrate,medage) = 0
      chi2( 1) =      4.78
      Prob > chi2 =      0.0288
```

Note:

1. We can reject the test at the 5% level.

## Also see

[SEM] [estat stdize](#) — Test standardized parameters

[SEM] [test](#) — Wald test of linear hypotheses

[R] [correlate](#) — Correlations of variables

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.

