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 http://www.stata-press.com/data/r15/census13
(1980 Census data by state)
. describe
    obs:    50  1980 Census data by state
   vars:    9  2 Dec 2016 14:01
   size: 1,250
```

```
<table>
<thead>
<tr>
<th>variable name</th>
<th>storage</th>
<th>type</th>
<th>format</th>
<th>value</th>
<th>label</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>byte</td>
<td>%13.0g</td>
<td>state1</td>
<td>State</td>
<td></td>
</tr>
<tr>
<td>brate</td>
<td>int</td>
<td>%10.0g</td>
<td></td>
<td>Birth rate</td>
<td></td>
</tr>
<tr>
<td>pop</td>
<td>long</td>
<td>%12.0gc</td>
<td></td>
<td>Population</td>
<td></td>
</tr>
<tr>
<td>medage</td>
<td>float</td>
<td>%9.2f</td>
<td></td>
<td>Median age</td>
<td></td>
</tr>
<tr>
<td>division</td>
<td>byte</td>
<td>%8.0g</td>
<td>division</td>
<td>Census division</td>
<td></td>
</tr>
<tr>
<td>region</td>
<td>byte</td>
<td>%-8.0g</td>
<td>cenreg</td>
<td>Census region</td>
<td></td>
</tr>
<tr>
<td>mrgrate</td>
<td>float</td>
<td>%9.0g</td>
<td></td>
<td>Marriage rate</td>
<td></td>
</tr>
<tr>
<td>dvcrate</td>
<td>float</td>
<td>%9.0g</td>
<td></td>
<td>Divorce rate</td>
<td></td>
</tr>
<tr>
<td>medagesq</td>
<td>float</td>
<td>%9.0g</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

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:

```
  mrgrate <--- dvcrate --- medage
```

1
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)

<table>
<thead>
<tr>
<th></th>
<th>mrgrate</th>
<th>dvcrate</th>
<th>medage</th>
</tr>
</thead>
<tbody>
<tr>
<td>mrgrate</td>
<td>1.0000</td>
<td>0.7700</td>
<td>-0.0177</td>
</tr>
<tr>
<td>dvcrate</td>
<td>0.7700</td>
<td>1.0000</td>
<td>-0.2290</td>
</tr>
<tr>
<td>medage</td>
<td>-0.0177</td>
<td>-0.2290</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
```

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
```

<table>
<thead>
<tr>
<th>Exogenous variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed: mrgrate dvcrate medage</td>
</tr>
</tbody>
</table>

Fitting target model:

- Iteration 0: log likelihood = 258.58985
- Iteration 1: log likelihood = 258.58985

Structural equation model

<table>
<thead>
<tr>
<th>Number of obs = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation method = ml</td>
</tr>
<tr>
<td>Log likelihood = 258.58985</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>OIM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
</tr>
<tr>
<td>mean(mrgrate)</td>
<td>.7332509</td>
</tr>
<tr>
<td>mean(dvcrate)</td>
<td>2.553791</td>
</tr>
<tr>
<td>mean(medage)</td>
<td>17.62083</td>
</tr>
<tr>
<td>var(mrgrate)</td>
<td>1</td>
</tr>
<tr>
<td>var(dvcrate)</td>
<td>1</td>
</tr>
<tr>
<td>var(medage)</td>
<td>1</td>
</tr>
<tr>
<td>cov(mrgrate, dvcrate)</td>
<td>.7699637</td>
</tr>
<tr>
<td>cov(mrgrate, medage)</td>
<td>-.0176541</td>
</tr>
<tr>
<td>cov(dvcrate, medage)</td>
<td>-.222932</td>
</tr>
</tbody>
</table>

LR test of model vs. saturated: chi2(0) = 0.00, Prob > chi2 = .
Note:

1. The correlations reported are

\[
\begin{align*}
\text{sem} & \quad \text{correlate} \\
\text{mrgrate and dvcrate} & \quad 0.7699637 \quad 0.7700 \\
\text{mrgrate and medage} & \quad -0.0176541 \quad -0.0177 \\
\text{dvcrate and medage} & \quad -0.222932 \quad -0.2229
\end{align*}
\]

Fitting the model with the Builder

Use the diagram above for reference.

1. Open the dataset.
   In the Command window, type
   
   \[. use http://www.stata-press.com/data/r15/census13\]

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.

   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
```

<table>
<thead>
<tr>
<th></th>
<th>Coef. Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>mean(mrgrate)</strong></td>
<td>.0186789 _b[/mean(mrgrate)]</td>
</tr>
<tr>
<td><strong>mean(dvcrate)</strong></td>
<td>.0079769 _b[/mean(dvcrate)]</td>
</tr>
<tr>
<td><strong>mean(medage)</strong></td>
<td>29.54 _b[/mean(medage)]</td>
</tr>
<tr>
<td><strong>var(mrgrate)</strong></td>
<td>.0006489 _b[/var(mrgrate)]</td>
</tr>
<tr>
<td><strong>var(dvcrate)</strong></td>
<td>9.76e-06 _b[/var(dvcrate)]</td>
</tr>
<tr>
<td><strong>var(medage)</strong></td>
<td>2.8104 _b[/var(medage)]</td>
</tr>
<tr>
<td><strong>cov(mrgrate, dvcrate)</strong></td>
<td>.0000613 _b[/cov(mrgrate,dvcrate)]</td>
</tr>
<tr>
<td><strong>cov(mrgrate, medage)</strong></td>
<td>-.0007539 _b[/cov(mrgrate,medage)]</td>
</tr>
<tr>
<td><strong>cov(dvcrate, medage)</strong></td>
<td>-.0011674 _b[/cov(dvcrate,medage)]</td>
</tr>
</tbody>
</table>

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] test — Wald test of linear hypotheses
[SEM] estat stdize — Test standardized parameters
[R] correlate — Correlations (covariances) of variables or coefficients