Example 6 — Linear regression

Description

Linear regression is demonstrated using auto.dta:

```stata
.sysuse auto
(1978 Automobile Data)
```

See *Structural models 1: Linear regression* in [SEM] Intro 5 for background.

Remarks and examples

Remarks are presented under the following headings:

- Fitting linear regression models
- Displaying standardized results
- Fitting the model with the Builder

Fitting linear regression models

The first two examples in [R] `regress` are

```stata
.regress mpg weight c.weight#c.weight foreign
.regress, beta
```

This model corresponds to

![Diagram](image)

weight

weight2

mpg

ε₁

foreign
To fit this model with \texttt{sem}, we type

\begin{verbatim}
. generate weight2 = weight^2
. sem (mpg <- weight weight2 foreign)
\end{verbatim}

Endogenous variables
Observed: mpg

Exogenous variables
Observed: weight weight2 foreign

Fitting target model:
Iteration 0: log likelihood = -1909.8206
Iteration 1: log likelihood = -1909.8206

Structural equation model Number of obs = 74
Estimation method = ml
Log likelihood = -1909.8206

<table>
<thead>
<tr>
<th></th>
<th>QIM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
</tr>
<tr>
<td>Structural</td>
<td></td>
</tr>
<tr>
<td>mpg</td>
<td></td>
</tr>
<tr>
<td>weight</td>
<td>-0.0165729</td>
</tr>
<tr>
<td>weight2</td>
<td>1.59e-06</td>
</tr>
<tr>
<td>foreign</td>
<td>-2.2035</td>
</tr>
<tr>
<td>_cons</td>
<td>56.53884</td>
</tr>
<tr>
<td>var(e.mpg)</td>
<td>10.19332</td>
</tr>
</tbody>
</table>

Note: The LR test of model vs. saturated is not reported because the fitted model is not full rank.

Notes:

1. We wished to include variable weight$^2$ in our model. Because \texttt{sem} does not allow Stata's factor-variable notation, we first had to \texttt{generate} new variable weight2.

2. Reported coefficients match those reported by \texttt{regress}.

3. Reported standard errors (SEs) differ slightly from those reported by \texttt{regress}. For instance, the SE for foreign is reported here as 1.03, whereas \texttt{regress} reported 1.06. SEM is an asymptotic estimator, and \texttt{sem} divides variances and covariances by $N = 74$, the number of observations. \texttt{regress} provides unbiased finite-sample estimates and divides by $N - k - 1 = 74 - 3 - 1 = 70$. Note that $1.03 \sqrt{74/70} = 1.06$.

4. \texttt{sem} reports $z$ statistics whereas \texttt{regress} reports $t$ statistics.

5. Reported confidence intervals differ slightly between \texttt{sem} and \texttt{regress} because of the $(N - k - 1)/N$ issue.

6. \texttt{sem} reports the point estimate of e.mpg as 10.19332. \texttt{regress} reports the root MSE as 3.2827. And $\sqrt{10.19332 \times 74/70} = 3.2827$.

**Displaying standardized results**

To obtain standardized coefficients from \texttt{regress}, you specify the \texttt{beta} option. To obtain standardized coefficients from \texttt{sem}, you specify the \texttt{standardized} option.
Example 6 — Linear regression

\[ \text{. sem, standardized} \]

Structural equation model

\[ \text{Number of obs} = 74 \]

Estimation method = ml

Log likelihood = -1909.8206

<table>
<thead>
<tr>
<th>Structural</th>
<th>OIM</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
</tr>
<tr>
<td>weight</td>
<td>-2.226321</td>
<td>.4950378</td>
<td>-4.50</td>
<td>0.000</td>
<td>-3.196577 -1.256064</td>
</tr>
<tr>
<td>weight2</td>
<td>1.32654</td>
<td>.498261</td>
<td>2.66</td>
<td>0.008</td>
<td>.3499662 2.303113</td>
</tr>
<tr>
<td>foreign</td>
<td>-.17527</td>
<td>.0810378</td>
<td>-2.16</td>
<td>0.031</td>
<td>-.3341011 -.0164399</td>
</tr>
<tr>
<td>_cons</td>
<td>9.839209</td>
<td>.9686872</td>
<td>10.16</td>
<td>0.000</td>
<td>7.940617 11.7378</td>
</tr>
<tr>
<td>var(e.mpg)</td>
<td>.308704</td>
<td>.0482719</td>
<td>.2272168</td>
<td>.4194152</td>
<td></td>
</tr>
</tbody>
</table>

Note: The LR test of model vs. saturated is not reported because the fitted model is not full rank.

regress simply reports standardized coefficients in an extra column. All other results are reported in unstandardized form. sem updates the entire output with the standardized values.

Fitting the model with the Builder

Use the diagram above for reference.

1. Open the dataset and create the additional variable weight2.

   In the Command window, type
   \[ \text{. use https://www.stata-press.com/data/r16/auto} \]
   \[ \text{. generate weight2 = weight^2} \]

2. Open a new Builder diagram.

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

3. Create the regression component for the mpg outcome.

   Select the Add regression component tool, \[ \text{Add} \], and then click in the center of the diagram.

   In the resulting dialog box,
   a. select mpg in the Dependent variable control;
   b. select weight, weight2, and foreign by using the Independent variables control;
   c. select Left in the Independent variables’ direction control;
   d. click on OK.

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

4. Estimate.

   Click on the Estimate button, \[ \text{Estimate} \], in the Standard Toolbar, and then click on OK in the resulting SEM estimation options dialog box.

5. 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_regress
```

**Also see**

- [SEM] **Example 12** — Seemingly unrelated regression
- [SEM] **Example 38g** — Random-intercept and random-slope models (multilevel)
- [SEM] **Example 43g** — Tobit regression
- [SEM] **Example 44g** — Interval regression
- [SEM] **sem** — Structural equation model estimation command