Title

> **example 6 —** Linear regression

Description

Linear regression is demonstrated using `auto.dta`:

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

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

Remarks and examples

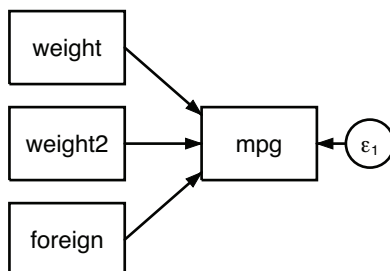Remarks are presented under the following headings:

> *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

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

This model corresponds to

To fit this model with `sem`, we type

```
. generate weight2 = weight^2
. sem (mpg <- weight weight2 foreign)
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
```

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf. Interval]
Structural					
mpg <-					
weight	-.0165729	.0038604	-4.29	0.000	-.0241392 -.0090067
weight2	1.59e-06	6.08e-07	2.62	0.009	4.00e-07 2.78e-06
foreign	-2.2035	1.03022	-2.14	0.032	-4.222695 -.1843056
_cons	56.53884	6.027559	9.38	0.000	44.72504 68.35264
var(e.mpg)	10.19332	1.675772			7.385485 14.06865

LR test of model vs. saturated: chi2(1)   =   0.00, Prob > chi2 = 1.0000

Notes:

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

2. Reported coefficients match those reported by `regress`.

3. Reported standard errors (SEs) differ slightly from those reported by `regress`. For instance, the SE for `foreign` is reported here as 1.03, whereas `regress` reported 1.06. SEM is an asymptotic estimator, and `sem` divides variances and covariances by $N = 74$, the number of observations. `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. `sem` reports z statistics whereas `regress` reports t statistics.

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

6. `sem` reports the point estimate of `e.mpg` as 10.19332. `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 `regress`, you specify the `beta` option. To obtain standardized coefficients from `sem`, you specify the `standardized` option.

example 6 — Linear regression 3

```
. sem, standardized
```

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

Standardized	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf. Interval]	
Structural						
mpg <-						
weight	-2.226321	.4950378	-4.50	0.000	-3.196577	-1.256064
weight2	1.32654	.498261	2.66	0.008	.3499662	2.303113
foreign	-.17527	.0810378	-2.16	0.031	-.3341011	-.0164389
_cons	9.839209	.9686872	10.16	0.000	7.940617	11.7378
var(e.mpg)	.308704	.0482719			.2272168	.4194152

LR test of model vs. saturated: chi2(1)   =   0.00, Prob > chi2 = 1.0000

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

   ```
   . use http://www.stata-press.com/data/r13/auto
   . 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, ⌐, 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, 🔲, 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