'standardized' coefficients

Sometime in your career you will encounter so-called 'standardized' regression coefficients. As a consumer of social statistics, you will encounter these in something you read or hear. Or, as a producer of social statistics, you will be asked to provide these, or — at least (and, in my experience, more likely) — asked to clarify whether the coefficients you have provided are standardized or not.

The logic of standardized coefficients is to re-express coefficients as the effect of a one-SD change in X_j as opposed to a *unit* change in X_j.

In other words, for a garden-variety OLS regression equation with N RHS variables:

$$E(Y|\mathbf{X}) = \alpha + \beta_1 X_1 + \cdots + \beta_N X_N + \epsilon$$

The β_j ($j = 1 \ldots N$) is the change in $E(Y)$ for a *unit change* in X_j holding all other X constant.

Standardized coefficients change this so that they are interpreted as a change in $E(Y)$ measured in units of SDs, for a *one SD change* in X_j holding all other X constant. The idea is that it makes the β_j's more comparable to each other, since (for example) a one-year change in age may be trivial but a one-category change in education may be profound. By placing everything on a common metric of SD's, we get comparability across β_j's. Or so the story goes. In other words it's meant to be an interpretation aid. *Statistical significance does not change.*

I for one am less persuaded that we are equipped to think about a one SD change in age as readily as a one-year change, but be that as it may, there is an infatuation, in some quarters, with using standardized coefficients.

Nomenclature.

Sometimes standardized coefficients are called "Betas". This is incredibly confusing, since this term is also widely used for *non*-standardized coefficients. SPSS uses

"Beta" for standardized coefficients and "B" for non-standardized coefficients (software changes constantly — check the latest version for current practice).

Estimation.

As Freedman points out, the name 'standardized' coefficients is a misnomer: "It is not coefficients that get standardized, but variables." (David A Freedman, *Statistical models: Theory and practice*, revised ed, Cambridge U Press, 2009, p 86). To get standardized coefficients, replace every variable (Y and all the X) in the regression with its standardized equivalent (subtract the mean off every observation, and divide that by the SD), and then just run the regression like you normally would.

note: This is not what Stata actually does. It just estimates OLS regression in the usual way, and then filters all the coefficients through this formula:

$$\hat{\beta}_j^s = \hat{\beta}_j \frac{\mathrm{SD}(x_j)}{\mathrm{SD}(Y)}$$

(see Eric Vittinghoff et al, *Regression methods in biostatistics: Linear, logistic, survival, and repeated measures models*, Springer, 2005, p 75).

Stata commands.

Use `regress` with the "`, beta`" option to obtain the standardized coefficients (these will appear under "Beta" where the CI normally is).

How to manually standardize a variable:

```
summ VARNAME
local a = r(mean)
local b = r(sd)
/* check: */
disp `a', `b'
gen std_VARNAME=(VARNAME -`a')/`b'
```

If you would prefer, you can standardize your variables like this, and then run `regress`.

No intercept.

There is no standardized intercept. That is to say, it is identically equal to zero. (*Why*?) If you are looking at a regression table that has an intercept, then you are not looking at standardized coefficients.