Supplemental Notes on Standardized Coefficients

NOTE: Long and Freese's `spostado` programs are used in this handout; specifically, the `listcoef` command, which is part of `spostado`, is used. Use the `findit` command to locate and install `spostado`. See Long and Freese's book, Regression Models for Categorical Dependent Variables Using Stata, Revised Edition, for more information.

Overview. As noted before, one of the problems with standardized coefficients is that they are somewhat difficult to interpret. Long and Freese discuss some alternative ways of standardizing variables that may help with interpretation. They primarily talk about these techniques with regards to logistic, multinomial logistic, and ordinal regression models, but they may be useful for OLS regression as well. Their `listcoef` command illustrates these different alternatives:

```
. use http://www.nd.edu/~rwilliam/stats1/statafiles/reg01.dta

. reg  income educ jobexp, beta

      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  2,    17) =   46.33
       Model |  1538.22521     2  769.112605           Prob > F      =  0.0000
    Residual |  282.200265    17  16.6000156           R-squared     =  0.8450
-------------+------------------------------           Adj R-squared =  0.8267
       Total |  1820.42548    19  95.8118671           Root MSE      =  4.0743

------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
        educ |   1.933393   .2099494     9.21   0.000                 .8844385
      jobexp |   .6493654   .1721589     3.77   0.002                 .3622612
       _cons |  -7.096855   3.626412    -1.96   0.067                        .
------------------------------------------------------------------------------

. listcoef, help

regress (N=20): Unstandardized and Standardized Estimates

 Observed SD: 9.7883536
 SD of Error: 4.0743117

-------------------------------------------------------------------------------
      income |      b         t     P>|t|    bStdX    bStdY   bStdXY      SDofX
-------------+-----------------------------------------------------------------
        educ |   1.93339    9.209   0.000   8.6572   0.1975   0.8844     4.4777
      jobexp |   0.64937    3.772   0.002   3.5459   0.0663   0.3623     5.4606
-------------------------------------------------------------------------------
       b = raw coefficient
       t = t-score for test of b=0
   P>|t| = p-value for t-test
   bStdX = x-standardized coefficient
   bStdY = y-standardized coefficient
  bStdXY = fully standardized coefficient
   SDofX = standard deviation of X
```

In the `listcoef` output, the column labeled `b` (which the `regress` command labels as `Coef.`) gives the unstandardized (metric) coefficients. The columns labeled `t` and `P>|t|` are also the same as in the `regress` output. The other columns give information that is relevant to different types of standardization.

Full Standardization. With full standardization, both the X and the Y variables are standardized to have a mean of 0 and a standard deviation of 1. It is the same as the standardization we have already been talking about. In the `listcoef` output, the fully standardized coefficients are in the column labeled `bStdXY` (in the output from the regress command, these appear in the column labeled `Beta`).

X-Standardization. An intermediate approach is to standardize only the X variables. In the `listcoef` output, in the column labeled `bStdX`, the Xs are standardized but Y is not. What this tells you, then, is that a 1 standard deviation increase in education (i.e. 4.4777 years – you can tell this from the `SDofX` column) produces an average increase of $8,657.20 in income. Also, a standard deviation increase (5.4606 years) in job experiences produces an average gain of $3,545.90 in income. Hence, by standardizing the Xs only, you can see the relative importance of the Xs, while still keeping the dependent variable in its original and more meaningful metric.

Note that, since a 1 standard deviation increase of 4.4777 years of education produces an $8,657.20 average increase in income, this means that a 1 year increase in education produces an average income gain of $8,657.20/4.4777 = $1933. Similarly, a 1 year increase in job experience produces an average increase in income of $3,545.90/5.4606 = $649. These are the same as what the unstandardized coefficients said.

Y-Standardization. You can also standardize Y only. The `listcoef` column labeled `bStdY` tells you that a 1 year increase in education produces, on average, a .1975 standard deviation increase in education, while a 1 year increase in job experience produces, on average, a .0663 standard deviation increase in income.

Since, as the printout points out, the observed SD of income is 9.7883536, this means that a 1 year increase in education produces, on average, a .1975 * 9.7883536 = 1.933 increase in income, and a 1 standard deviation increase in job experience produces, on average, a .0663 * 9.7883536 = .649 in income. Note that this is the same as what the unstandardized coefficients said.

At least in the case of OLS regression, I don't find Y-standardization very useful, but it may be more helpful for things like logistic regression.

Other Comments. With the X variables, you could use combinations of standardized and unstandardized variables. For example, many people don't like to standardize dummy variables, which only have values of 0 and 1, because a "one standard deviation increase" isn't something that could actually happen with such a variable. Ergo, you might want to leave the dummy variables unstandardized while standardizing continuous X variables.

I'm still not a big fan of standardized coefficients, but I do probably prefer x-standardization over full-standardization. Long and Freese also argue that the different types of standardization can sometimes be useful, and give various examples, mostly for non-OLS techniques like logistic regression, multinomial logistic regression, and ordinal regression.