THE STATA JOURNAL

Department of Statistics
Texas A & M University
College Station, Texas 77843
979-845-3142; FAX 979-845-3144
jnewton@stata-journal.com

Department of Geography
University of Durham
South Road
Durham City DH1 3LE UK
n.j.cox@stata-journal.com

Associate Editors

Stata Press Production Manager   Lisa Gilmore

Stata tip 14: Using value labels in expressions

Kenneth Higbee
StataCorp
khigbee@stata.com

Did you know that there is a way in Stata to specify value labels directly in an expression, rather than through the underlying numeric value? You specify the label in double quotes (" "), followed by a colon (:), followed by the name of the value label. If we read in this dataset and see what it contains

```
. webuse census9
(1980 Census data by state)

. describe

Contains data from http://www.stata-press.com/data/r8/census9.dta
  obs:            50                          1980 Census data by state
 vars:             5                          16 Jul 2002 18:29
 size:         1,550 (99.9% of memory free)

              storage  display      value
variable name   type   format       label      variable label

state          str14   %-14s                   State
drate          float   %9.0g                   Death Rate
pop            long    %12.0gc                  Population
medage         float   %9.2f                    Median age
region         byte    %-8.0g       cenreg      Census region

Sorted by:
```

we notice that variable `region` has values labeled by the `cenreg` value label. The correspondence between the underlying number and the value label is shown by

```
. label list
cenreg:
           1 NE
           2 N Cntrl
           3 South
           4 West
```

[R] **regress** uses this dataset to illustrate weighted regression. To obtain the regression of `drate` and `medage` restricted to the "South" region, you could type

```
. regress drate medage [aweight=pop] if region == 3
```

But, if you do not remember the underlying region number for "South", you could also obtain this regression by typing

```
. regress drate medage [aweight=pop] if region == "South":cenreg
(sum of wgt is   7.4734e+07)
```

Source	SS	df	MS		Number of obs =	16
					F(1, 14) =	27.29
Model	1072.30989	1	1072.30989		Prob > F =	0.0001
Residual	550.163155	14	39.2973682		R-squared =	0.6609
					Adj R-squared =	0.6367
Total	1622.47305	15	108.16487		Root MSE =	6.2688

drate	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
medage	3.905819	.7477109	5.22	0.000	2.302139	5.509499
_cons	-29.34031	22.33676	-1.31	0.210	-77.2479	18.56727

Typing the value label instead of the underlying number makes it unlikely that you will obtain an unintended result from entering the wrong region number. An added benefit of using the value label is that, when you later review your results, you will quickly see that the regression is for the "South" region, and you will not need to remember what region was assigned number 3.

See [U] **16.9 Label values** for further information about specifying value labels in expressions.