The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$ 98	1-year subscription	$138
2-year subscription	$165	2-year subscription	$245
3-year subscription	$225	3-year subscription	$345
1-year student subscription	$ 75	1-year student subscription	$ 99
1-year institutional subscription	$245	1-year institutional subscription	$285
2-year institutional subscription	$445	2-year institutional subscription	$525
3-year institutional subscription	$645	3-year institutional subscription	$765
Electronic only		**Electronic only**	
1-year subscription	$ 75	1-year subscription	$ 75
2-year subscription	$125	2-year subscription	$125
3-year subscription	$165	3-year subscription	$165
1-year student subscription	$ 45	1-year student subscription	$ 45

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The Stata Journal (2014)
14, Number 1, pp. 218–220

Stata tip 116: Where did my p-values go? (Part 3)

Maarten L. Buis
Wissenschaftszentrum Berlin für Sozialforschung (WZB)
Berlin, Germany
maarten.buis@wzb.eu

In a previous Stata tip (Buis 2007), I discussed how to recover t statistics, p-values, and confidence intervals for regression parameters by using the results that are returned by an estimation command. In a subsequent Stata tip (Buis 2011), I discussed how to recover parameter estimates for parameters that were estimated on a transformed scale. For example, if a likelihood function contains a standard deviation or a correlation, then many Stata commands will maximize the likelihood with respect to ln(standard deviation) and the Fisher's z transformation of the correlation. In this tip, I will discuss how to recover the standard errors for the back-transformed parameters, that is, the standard errors of the standard deviation and the correlation.

Stata often displays the back-transformed parameters and their standard errors, but it leaves behind only the estimates of the transformed parameters and their standard errors. In those cases, the delta method (for example, Feiveson 2005) was used to compute the standard errors of the back-transformed parameters. In its simplest form, the delta method means that if we apply a transformation $G(\cdot)$ to a parameter b, then we can approximate the standard error of the transformed parameter as

$$\text{se}\{G(b)\} \approx \text{se}(b) \times G'\left(\widehat{b}\right)$$

where $G'(\widehat{b})$ is the first derivative of $G(b)$ with respect to b evaluated at $\widehat{b}$. If Stata returned ln(standard deviation) and we wanted the standard deviation and its standard error, then $G(b) = \exp(b)$ and $G'(\widehat{b}) = \exp(\widehat{b})$. If Stata returned Fisher's z transformation of a correlation and we wanted the correlation and its standard error, then $G(b) = \tanh(b)$ and $G'(\widehat{b}) = \cosh(\widehat{b})^{-2}$. This is illustrated below using a model estimated with `heckman` (see [R] **heckman**). This model was chosen because it returns transformed parameters of both types.

 st0332

```
. use http://www.stata-press.com/data/r13/womenwk

. heckman wage educ, select(married children educ) nolog

Heckman selection model                         Number of obs      =      2000
(regression model with sample selection)        Censored obs       =       657
                                                Uncensored obs     =      1343

                                                Wald chi2(1)       =    403.39
Log likelihood = -5250.348                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
wage         |
   education |   1.099506   .0547435    20.08   0.000     .9922102    1.206801
       _cons |   7.042147   .8423253     8.36   0.000      5.39122    8.693074
-------------+----------------------------------------------------------------
select       |
     married |   .5420304   .0657798     8.24   0.000     .4131044    .6709564
    children |   .4409418   .0276093    15.97   0.000     .3868286     .495055
   education |   .0722993   .0105096     6.88   0.000     .0517007    .0928978
       _cons |  -1.473038   .1465476   -10.05   0.000    -1.760266    -1.18581
-------------+----------------------------------------------------------------
     /athrho |   .8081049   .1108545     7.29   0.000     .5908341    1.025376
    /lnsigma |   1.807547   .0291035    62.11   0.000     1.750506    1.864589
-------------+----------------------------------------------------------------
         rho |   .6685435    .061308                      .5304953     .772047
       sigma |   6.095479   .1773995                      5.757513    6.453283
      lambda |   4.075093   .4690025                      3.155865    4.994321
------------------------------------------------------------------------------
LR test of indep. eqns. (rho = 0):   chi2(1) =    47.02   Prob > chi2 = 0.0000

. tempname gprime

. scalar `gprime´ = exp(_b[lnsigma:_cons])

. display "se of sigma = " _se[lnsigma:_cons]*`gprime´
se of sigma = .17739954

. scalar `gprime´ = cosh(_b[athrho:_cons])^-2

. display "se of rho = " _se[athrho:_cons]*`gprime´
se of rho = .06130802
```

Alternatively, one can use `nlcom` (see [R] **nlcom**) to compute these standard errors, as follows:

```
. nlcom (rho: tanh(_b[athrho:_cons])) (sigma: exp(_b[lnsigma:_cons]))

         rho:  tanh( _b[athrho:_cons]  )
       sigma:  exp(  _b[lnsigma:_cons] )

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         rho |   .6685435    .061308    10.90   0.000      .548382     .788705
       sigma |   6.095479   .1773995    34.36   0.000     5.747782    6.443175
------------------------------------------------------------------------------
```

Notice that the confidence intervals do not correspond with those in the output of `heckman`. This is because `heckman` first computes the bounds of the confidence intervals for the transformed parameters and then back-transforms those bounds to the original metric, while `nlcom` uses the standard errors for the back-transformed parameters for computing these bounds. In most cases, computing the bounds on the transformed scale and then back-transforming those bounds to the original scale results in somewhat better bounds. This is because the sampling distribution of the transformed parameters is likely to be better approximated by a normal distribution than is the sampling distribution of the back-transformed parameters. For more discussion, see Sribney and Wiggins (2009). You can use the techniques discussed in Buis (2007) to recover the confidence intervals reported by `heckman`.

```
. display "confidence interval for rho: ["
> tanh( _b[athrho:_cons] - invnormal(.975)*_se[athrho:_cons] ) ", "
> tanh( _b[athrho:_cons] + invnormal(.975)*_se[athrho:_cons] ) "]"
confidence interval for rho: [.53049526, .77204703]
. display "confidence interval for sigma: ["
> exp( _b[lnsigma:_cons] - invnormal(.975)*_se[lnsigma:_cons] ) ", "
> exp( _b[lnsigma:_cons] + invnormal(.975)*_se[lnsigma:_cons] ) "]"
confidence interval for sigma: [5.7575126, 6.4532831]
```

Also note that `nlcom` returns the z statistic and p-value for the test of the null hypothesis that the standard deviation and the correlation are 0, which were not reported by `heckman`. This test is problematic in the case of the standard deviation, because this is a test "on the boundary of the parameter space". A standard deviation can only take values larger than or equal to 0. Thus the hypothesis that the standard deviation is equal to 0 is on the boundary of the possible values for the standard deviation, and standard tests do not tend to behave well in this extreme area (for example, Gutierrez, Carter, and Drukker [2001]).

References

Buis, M. L. 2007. Stata tip 53: Where did my p-values go? *Stata Journal* 7: 584–586.

———. 2011. Stata tip 97: Getting at ρ's and σ's. *Stata Journal* 11: 315–317.

Feiveson, A. H. 2005. FAQ: What is the delta method and how is it used to estimate the standard error of a transformed parameter? http://www.stata.com/support/faqs/statistics/delta-method/.

Gutierrez, R. G., S. Carter, and D. M. Drukker. 2001. sg160: On boundary-value likelihood-ratio tests. *Stata Technical Bulletin* 60: 15–18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 269–273. College Station, TX: Stata Press.

Sribney, W., and V. Wiggins. 2009. FAQ: Standard errors, confidence intervals, and significance tests for ORs, HRs, IRRs, and RRRs. http://www.stata.com/support/faqs/statistics/delta-rule/.