The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–4

Stata tip 54: Where did my p-values go?

Maarten L. Buis
Department of Social Research Methodology
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
m.buis@fsw.vu.nl

A useful tool in the Stata toolkit is the returned result. For instance, after most estimation commands the parameter estimates are stored in a matrix called `e(b)`. However, these commands do not return the t-statistics, p-values, and confidence intervals for those parameter estimates. The aim of this tip is to show how to recover those statistics using the statistics that are returned. This is illustrated using the following OLS regression:

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

. regress price mpg foreign

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  2,    71) =   14.07
       Model |   180261702     2  90130850.8           Prob > F      =  0.0000
    Residual |   454803695    71  6405685.84           R-squared     =  0.2838
-------------+------------------------------           Adj R-squared =  0.2637
       Total |   635065396    73  8699525.97           Root MSE      =  2530.9

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         mpg |  -294.1955   55.69172    -5.28   0.000    -405.2417   -183.1494
     foreign |   1767.292    700.158     2.52   0.014     371.2169    3163.368
       _cons |   11905.42   1158.634    10.28   0.000     9595.164    14215.67
------------------------------------------------------------------------------
```

t-statistic

The t-value can be calculated using the formula $t = \frac{\widehat{b}-b}{se}$, whereby $\widehat{b}$ is the observed parameter, b is the parameter under the null hypothesis, and se is the standard error. The null hypothesis is almost always that the parameter equals zero, in which case the equation simplifies to $t = \frac{\widehat{b}}{se}$. The t-statistic for a single parameter (`foreign`) can be calculated as:

```
. di _b[foreign]/_se[foreign]
2.5241336
```

Alternatively, all the parameter estimates are returned in the matrix `e(b)`. A vector of all standard errors is a bit harder to obtain: They are the square root of the diagonal elements of the matrix `e(V)`. In `mata` that vector can be created by typing `diagonal(cholesky(diag(V)))`. Continuing the example above, a vector of all

 st0001

t-statistics can be computed in the following way:

```
: b = st_matrix("e(b)")'
: V = st_matrix("e(V)")
: se = diagonal(cholesky(diag(V)))
: b :/ se
                  1
    +----------------+
  1 |  -5.282572354  |
  2 |    2.52413358  |
  3 |   10.27538518  |
    +----------------+
```

p-value

The p-value can be calculated using the formula $p = 2 * (1 - T(df, |t|))$, whereby T is the cumulative distribution of the student's t-distribution, df the residual degrees of freedom, and $|t|$ the absolute value of the observed t-statistic. The t-statistic was calculated before and the residual degrees of freedom are returned as `e(df_r)`. The absolute value can be calculated using the `abs()` function, and $(1 - T(df, t))$ can be calculated using the `ttail(`*df*,*t*`)` function. So, the p-value for the parameter of `foreign` can be calculated as:

```
. local t = _b[foreign]/_se[foreign]
. di 2*ttail(e(df_r),abs(`t'))
.01383634
```

The vector of all p-values can be calculated as:

```
: df = st_numscalar("e(df_r)")
: t = b :/ se
: 2*ttail(df, abs(t))
                 1
    +---------------+
  1 |  1.33307e-06  |
  2 |  .0138363442  |
  3 |  1.08513e-15  |
    +---------------+
```

confidence interval

The lower and upper bounds of the confidence interval can be calculated using the formula $ci = \widehat{b} \pm t_{\frac{\alpha}{2}} se$, whereby $t_{\frac{\alpha}{2}}$ is the critical t-value given a significance level $\frac{1}{2}\alpha$. This critical value can be calculated using the `invttail(`*df*, $\frac{\alpha}{2}$`)` function. The upper and lower bound of the 95% confidence interval for the parameter of `foreign` can be calculated as:

```
. di _b[foreign] - invttail(e(df_r),0.025)*_se[foreign]
371.2169
. di _b[foreign] + invttail(e(df_r),0.025)*_se[foreign]
3163.3676
```

The vectors of lower and upper bounds for all parameters can be calculated as:

```
: b :- invttail(df,0.025):*se, b :+ invttail(df,0.025):*se
                  1              2
    +-------------------------------+
  1 |  -405.2416661   -183.1494001  |
  2 |   371.2169028    3163.367584  |
  3 |     9595.1638    14215.66676  |
    +-------------------------------+
```

Models reporting z-statistics

If you are using an estimation command that reports z-statistics instead of t-statistics the equivalent Stata command are:

- `_b[foreign]/_se[foreign]` for the z-statistic,
- `2*normal(-abs(‘z’))` for the p-value (the minus sign comes from the fact `normal()` looks at the lower tail of the distribution, while `ttail()` looks at the upper tail.),
- `_b[foreign] - invnormal(0.975)*_se[foreign]` for the lower bound of the 95% confidence interval, and `_b[foreign] + invnormal(0.975)*_se[foreign]` for the upper bound (Here .975 is used instead of .025 because `invnormal()` looks at the lower tail, while `invttail()` looks at the upper tail).

Conclusion

In many cases it is unnecessary to do these calculations. For instance if you are interested in creating custom tables of regression output the `estimates table` command or the tools developed by Jann (2005, 2007) and Wanda (2007) are much more convenient. Similarly, if the aim is to create graphs of regression output one should take a close look at the tools developed by Newson (2003) before attempting to use the methods described in this tip. This tip showed how to calculate these numbers in case none of these commands do what you want.

References

Jann, B. 2005. Making regression tables from stored estimates. *The Stata Journal* 5: 288 – 308.

———. 2007. Making regression tables simplified. *The Stata Journal* 7: 227 – 244.

Newson, R. 2003. Confidence intervals and p-values for delivery to the end user. *The Stata Journal* 3: 245 – 269.

Wanda, R. 2007. OUTREG2: module to arrange regression outputs into an illustrative table. `http://ideas.repec.org/c/boc/bocode/s456416.html`.