



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

THE STATA JOURNAL

Editor

H. Joseph Newton
Department of Statistics
Texas A & M University
College Station, Texas 77843
979-845-8817; FAX 979-845-6077
jnewton@stata-journal.com

Associate Editors

Christopher F. Baum
Boston College
Rino Bellocchio
Karolinska Institutet, Sweden and
Univ. degli Studi di Milano-Bicocca, Italy
A. Colin Cameron
University of California–Davis
David Clayton
Cambridge Inst. for Medical Research
Mario A. Cleves
Univ. of Arkansas for Medical Sciences
William D. Dupont
Vanderbilt University
Charles Franklin
University of Wisconsin–Madison
Allan Gregory
Queen's University
James Hardin
University of South Carolina
Ben Jann
ETH Zürich, Switzerland
Stephen Jenkins
University of Essex
Ulrich Kohler
WZB, Berlin
Jens Lauritsen
Odense University Hospital

Stata Press Production Manager

Stata Press Copy Editor

Editor

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

Stanley Lemeshow
Ohio State University

J. Scott Long
Indiana University

Thomas Lumley
University of Washington–Seattle

Roger Newson
Imperial College, London

Marcello Pagano
Harvard School of Public Health

Sophia Rabe-Hesketh
University of California–Berkeley

J. Patrick Royston
MRC Clinical Trials Unit, London

Philip Ryan
University of Adelaide

Mark E. Schaffer
Heriot-Watt University, Edinburgh

Jeroen Weesie
Utrecht University

Nicholas J. G. Winter
University of Virginia

Jeffrey Wooldridge
Michigan State University

Lisa Gilmore

Deirdre Patterson

Copyright Statement: The Stata Journal and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the Stata Journal.

The articles appearing in the Stata Journal may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the Stata Journal.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the Stata Journal, in whole or in part, on publicly accessible web sites, fileservers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the Stata Journal or the supporting files understand that such use is made without warranty of any kind, by either the Stata Journal, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the Stata Journal is to promote free communication among Stata users.

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press. Stata and Mata are registered trademarks of StataCorp LP.

Stata tip 53: 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 item in the Stata toolkit is the returned result. For example, after most estimation commands, parameter estimates are stored in a matrix `e(b)`. However, these commands do not return the t statistics, p -values, and confidence intervals for those parameter estimates. The aim here is to show how to recover those statistics by using the results that are returned. Consider the following OLS regression:

```
. sysuse auto
(1978 Automobile Data)
. regress price mpg foreign
```

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

1 t statistic

The t statistic can be calculated from $t = (\hat{b} - b)/se$, where \hat{b} is the estimated parameter, b is the parameter value under the null hypothesis, and se is the standard error. The null hypothesis is usually that the parameter equals zero; thus we have $t = \hat{b}/se$. The t statistic for one parameter (`foreign`) can be calculated by

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

All the parameter estimates are also returned in the matrix `e(b)`. A vector of all standard errors is a bit harder to obtain; they are the square roots 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, a vector of all t statistics can be computed within Mata by

```
: 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 |

2 p-value

The p -value can be calculated from $p = 2 * (1 - T(df, |t|))$, where T is the cumulative distribution function of Student's t distribution, df is the residual degrees of freedom, and $|t|$ is 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 by using the `abs()` function, and $(1 - T(df, t))$ can be calculated by using the `ttail(df, t)` function. The calculation is put together as follows:

```

. local t = _b[foreign]/_se[foreign]
. di 2*ttail(e(df_r),abs(`t`))
.01383634

```

Using Mata, the vector of all p -values is then

```

: 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 |

3 Confidence interval

The lower and upper bounds of the confidence interval can be calculated as $\hat{b} \pm t_{\alpha/2}se$, where $t_{\alpha/2}$ is the critical t -value given a significance level $\alpha/2$. This critical value can be calculated by using the `invttail(df, $\alpha/2$)` function. The lower and upper bounds of the 95% confidence interval for the parameter of `foreign` are thus given by

```

. 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

```

(Continued on next page)

The vectors of lower and upper bounds for all parameters follow suit in Mata 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 |

4 Models reporting z statistics

If you are using an estimation command that reports z statistics instead of t statistics, the values become

- `_b[foreign]/_se[foreign]` for the z statistic;
- `2*normal(-abs('z'))` for the p -value (where the minus sign comes from the fact `normal()` starts with the lower tail of the distribution, whereas `ttail()` starts with 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 (.975 is used instead of .025 for the same kind of reason).

5 Further comments

Often it is unnecessary to do these calculations. In particular, if you are interested in creating custom tables of regression-like output the `estimates table` command or the tools developed by Jann (2005, 2007) are much more convenient. Similarly, if the aim is to create graphs of regression output, take a good look at the tools developed by Newson (2003) before attempting to use the methods described here. This tip is for situations in which no such command does what you want.

References

- Jann, B. 2005. Making regression tables from stored estimates. *Stata Journal* 5: 288–308.
- . 2007. Making regression tables simplified. *Stata Journal* 7: 227–244.
- Newson, R. 2003. Confidence intervals and p-values for delivery to the end user. *Stata Journal* 3: 245–269.