



**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](mailto: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  
979-845-3142; FAX 979-845-3144  
jnewton@stata-journal.com

## Associate Editors

Christopher Baum  
Boston College

Rino Bellocco  
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

Joanne M. Garrett  
University of North Carolina

Allan Gregory  
Queen's University

James Hardin  
University of South Carolina

Ben Jann  
ETH Zurich, Switzerland

Stephen Jenkins  
University of Essex

Ulrich Kohler  
WZB, Berlin

## Stata Press Production Manager

## Stata Press Copy Editor

## Editor

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

Jens Lauritsen  
Odense University Hospital

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

Gabe Waggoner

**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, file servers, 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, and Stata is a registered trademark of StataCorp LP.

## Stata tip 35: Detecting whether data have changed

William Gould  
StataCorp  
College Station, TX  
wgould@stata.com

Included in the 17 May 2006 update is a new command that you may find useful, `datasignature`. If you have installed the update, type `help datasignature`. If you have not updated, or are unsure, type `update query` to find out, and type `update all` to install.

Here is the result of running `datasignature` on `auto.dta`:

```
. sysuse auto
(1978 automobile data)
. datasig
74:12(71728):3831085005:1395876116
```

The output is `auto.dta`'s data signature. If you change the data, even just a little bit, the last two numbers will change:

```
. replace mpg = mpg + 1 in 2
(1 real change made)
. datasig
74:12(71728):1616229321:1400086868
```

If you change the name of a variable, the last two numbers stay the same and the number inside the parentheses changes:

```
. rename mpg miles_per_gallon
. datasig
74:12(57876):1616229321:1400086868
```

`datasignature` is designed to help those who use data maintained by others and those who worry that they might themselves have accidentally changed their data.

In the latter case, you could save the signature in the dataset,

```
. datasig
74:12(57876):1616229321:1400086868
. note: 'r(datasignature)'
```

and check it later,

```
. notes
_dta:
1. from Consumer Reports with permission
2. 74:12(57876):1616229321:1400086868
```

Another idea is to include `datasignature` at the beginning of logs:

```
————— begin myfile.do ———  
log using myfile, replace  
sysuse auto, clear  
datasig  
...  
log close  
————— end myfile.do ———
```

You can specify a varlist and `if` and `in`, so if you have a large dataset and want to check just part of it, you can write

```
————— begin myfile.do ———  
log using myfile, replace  
sysuse auto, clear  
datasig mpg weight price if foreign  
...  
log close  
————— end myfile.do ———
```