



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.

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

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

Editor

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

Associate Editors

Christopher F. Baum
Boston College

Nathaniel Beck
New York University

Rino Bellocchio
Karolinska Institutet, Sweden, and
University of Milano-Bicocca, Italy

Maarten L. Buis
Tübingen University, Germany

A. Colin Cameron
University of California–Davis

Mario A. Cleves
Univ. of Arkansas for Medical Sciences

William D. Dupont
Vanderbilt University

David Epstein
Columbia University

Allan Gregory
Queen's University

James Hardin
University of South Carolina

Ben Jann
University of Bern, Switzerland

Stephen Jenkins
London School of Economics and
Political Science

Ulrich Kohler
WZB, Berlin

Frauke Kreuter
University of Maryland–College Park

Stata Press Editorial Manager
Stata Press Copy Editor

Peter A. Lachenbruch
Oregon State University

Jens Lauritsen
Odense University Hospital

Stanley Lemeshow
Ohio State University

J. Scott Long
Indiana University

Roger Newson
Imperial College, London

Austin Nichols
Urban Institute, Washington DC

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 McClellan

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.

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

<http://www.stata-journal.com>

The *Stata Journal* is indexed and abstracted in the following:

- CompuMath Citation Index®
- Current Contents/Social and Behavioral Sciences®
- RePEc: Research Papers in Economics
- Science Citation Index Expanded (also known as SciSearch®)
- Scopus™
- Social Sciences Citation Index®

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 websites, 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, Mata, NetCourse, and Stata Press are registered trademarks of StataCorp LP.

Stata tip 101: Previous but different

Nicholas J. Cox
Department of Geography
Durham University
Durham City, UK
n.j.cox@durham.ac.uk

Given a time series that changes intermittently, there may be interest in identifying the most recent value that differed from the present value. Given a history that was 36, 36, 36, 42 before a present value of 42, the value that was most recent but different is 36. That is easy enough to determine by eye, but how could we calculate that systematically in Stata? Such distinct values define runs or spells, as discussed in detail in [Cox \(2007\)](#), but here we examine the problem directly.

Adding some more detail, let us suppose first that we have a single panel with data starting with

```
time: 2, 3, 5, 7, 11
value: 36, 36, 36, 42, 42
```

The example is of irregularly spaced times, and in fact only one time step is 1 in whatever units we are using. These data were devised to emphasize that the method to be explained does not depend on anything other than the time variable being known and single-valued for a given panel and values being in `sort` order of time. We can always ensure that last condition by `sort`ing on the time variable. Not only do we not need to `tsset`, but that machinery and the related apparatus of time-series operators (such as `L.` and `D.`) would not help much, even for regularly spaced data.

Information on previous values comes from when the value changes. We can create a new variable containing previous values using subscripts:

```
. generate previous = value[_n-1] if value != value[_n-1]
```

Because `_n` means the present observation number, `_n - 1` means the previous observation number. If this is new to you, get more information by typing `help subscripting` into Stata.

The new variable in our example now includes missings in all but one observation: `., ., ., 36, .` are its values. The value of 36 clearly comes from the one point at which `value` changes, but we also need to understand the logic for the other observations.

Let us focus first on `value[1]`. Its previous value is `value[0]`, a hypothetical value beyond the dataset. Stata does not know what that is but does not regard a request to access `value[0]` as illegal; it merely returns missing. Now we have `value[1]` of 36, which differs from `value[0]` of missing, so `previous[1]` is returned as `value[0]`, namely, missing.

The reasoning for `value[2]`, `value[3]`, and `value[5]` is different. In each case, `value` is equal to the previous `value`, but the `generate` statement included no instruc-

tions for what to do if the inequality was not satisfied. So missing is also returned, the same result for a different reason.

We must now fill in missing values to the best of our ability. In particular, 36 is the previous but different value for observation 5 as well as observation 4. Exploiting the fact that `generate` and `replace` use the current sort order (Newson 2004), we can fill in missing values in a cascade:

```
. replace previous = previous[_n-1] if missing(previous)
```

That still leaves a block of missings before the first value change. Furthermore, if `value` never changed within our data, there would be no previous but different value to identify.

```
. list
```

	time	value	previous
1.	2	36	.
2.	3	36	.
3.	5	36	.
4.	7	42	36
5.	11	42	36

Now we need to extend this to any number of panels. Given also an identifier, the extension is just to do all using `by`: (Cox 2002):

```
. by id (time), sort: generate previous = value[_n-1] if value != value[_n-1]
. by id: replace previous = previous[_n-1] if missing(previous)
```

If there are missings in `value`, they would be better filled in first in a clone of `value` using the cascade device. The difference from the previous different value is then simply (regardless of panel context or time spacing)

```
. generate diff = value - previous
```

If we wanted to look forward in time to identify the next value that is different, it is easiest to reverse time temporarily:

```
. gsort id -time
. by id: generate next = value[_n-1] if value != value[_n-1]
. by id: replace next = next[_n-1] if missing(next)
. sort id time
```

References

Cox, N. J. 2002. Speaking Stata: How to move step by step. *Stata Journal* 2: 86–102.

———. 2007. Speaking Stata: Identifying spells. *Stata Journal* 7: 249–265.

Newson, R. 2004. Stata tip 13: generate and replace use the current sort order. *Stata Journal* 4: 484–485.