



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.*

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
979-845-3142; FAX 979-845-3144
jnewton@stata-journal.com

Editor

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

Associate Editors

Christopher Baum
Boston College
Rino Bellocco
Karolinska Institutet
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
Stephen Jenkins
University of Essex
Ulrich Kohler
WZB, Berlin
Jens Lauritsen
Odense University Hospital

Stanley Lemeshow
Ohio State University
J. Scott Long
Indiana University
Thomas Lumley
University of Washington, Seattle
Roger Newson
King's 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
Cornell University
Jeffrey Wooldridge
Michigan State University

Stata Press Production Manager

Lisa Gilmore

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, and Stata is a registered trademark of StataCorp LP.

Stata tip 16: Using input to generate variables

Ulrich Kohler
Wissenschaftszentrum Berlin für Sozialforschung
kohler@wz-berlin.de

Sometimes using `generate` is an untidy and long-winded way to generate new variables, particularly if the variable you want to create is categorical and there are many different categories. Thus rather than using

```
. gen iso3166_2 = "AT" if country == "Austria"  
. replace iso3166_2 = "BE" if country == "Belgium"  
. replace iso3166_2 = "TR" if country == "Turkey"  
and so on for say 28 countries  
  
. gen iso3166_3 = "AUT" if country == "Austria"  
. replace iso3166_3 = "BEL" if country == "Belgium"  
. replace iso3166_3 = "TUR" if country == "Turkey"  
and so on for say 28 countries  
  
. gen gername = "Österreich" if country == "Austria"  
. replace gername = "Belgien" if country == "Belgium"  
. replace gername = "Türkei" if country == "Turkey"  
and so on for say 28 countries
```

you can use `input` to produce a new dataset, save to a temporary file, and then merge:

```
. preserve  
. clear  
. input str15 country str2 iso3166_2 str3 iso3166_3 str15 gername  
Austria AT AUT Österreich  
Belgium BE BEL Belgien  
Turkey TK TUR Türkei  
and so on  
. end  
  
. sort country  
. tempfile foo  
. save 'foo'  
. restore  
. sort country  
. merge country using 'foo'
```

Among the benefits are less typing; a cleaner log file; in huge datasets, faster data processing; and arguably fewer errors.

See [R] `input` for the finer points on `input`.