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

Editor
Nicholas J. Cox
Geography Department
Durham University
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.

The Stata Journal (2005)
5, Number 2, pp. 280–281

Stata tip 20: Generating histogram bin variables

David A. Harrison
ICNARC, London, UK
david.harrison@icnarc.org

Did you know about `twoway__histogram_gen`? (Note the two underscores in the first gap and only one in the second.) This command is used by `histogram` to generate the variables that are plotted. It is undocumented in the manuals but explained in the online help. The command can be used directly to save these variables, enabling more complex manipulation of histograms and production of other graphs or tables.

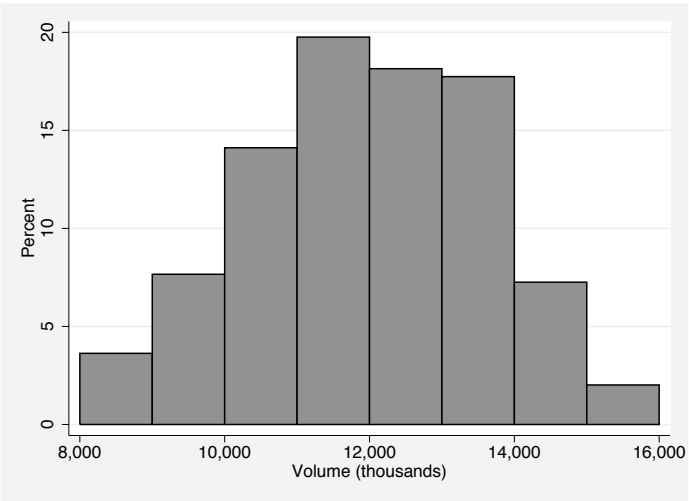
Consider the S&P 500 historical data that are used as an example for [R] **histogram**:

```
. use http://www.stata-press.com/data/r9/sp500
(S&P 500)
. histogram volume, percent start(4000) width(1000)
(bin=20, start=4000, width=1000)
```

(*output omitted*)

To display only the central part of this histogram from 8,000 to 16,000, we could use `if`, but this will change the height of the bars, as data outside the range 8,000 to 16,000 will be ignored completely. To restrict the range without altering the bars, we use `twoway__histogram_gen` to save the histogram and only plot the section of interest:

```
. twoway__histogram_gen volume, percent start(4000) width(1000) gen(h x)
. twoway bar h x if inrange(x,8000,16000), barwidth(1000) bstyle(histogram)
```



The `start()` and `width()` options above specified cutpoints that included 8,000 and 16,000. We could, alternatively, use the default cutpoints:

© 2005 StataCorp LP

```
. twoway__histogram_gen volume if inrange(volume,8000,16000), display
(bin=14, start=8117, width=525.08571)
. local m = r(start)
. local w = r(width)
. summarize volume, meanonly
. local s = `m' - `w' * ceil((`m' - r(min))/`w')
. twoway__histogram_gen volume, percent start(`s') width(`w') gen(h x, replace)
. twoway bar h x if inrange(x,8000,16000), barwidth(`w') bstyle(histogram)
  (output omitted)
```

Other uses of `twoway__histogram_gen` include the following:

- Overlaying or mirroring two histograms

```
. use http://www.stata-press.com/data/r9/bplong, clear
(fictional blood-pressure data)
. twoway__histogram_gen bp if sex == 0, frac start(125) w(5) gen(h1 x1)
. twoway__histogram_gen bp if sex == 1, frac start(125) w(5) gen(h2 x2)
. twoway (bar h1 x1, barw(5) bc(gs11))
> (bar h2 x2, barw(5) blc(black) bfc(none)),
> legend(order(1 "Male" 2 "Female"))
  (output omitted)
. qui replace h2 = -h2
. twoway (bar h1 x1, barw(5)) (bar h2 x2, barw(5)),
> yla(-.2 ".2" -.1 ".1" 0 .1 .2) legend(order(1 "Male" 2 "Female"))
  (output omitted)
```

- Changing the scale, for example, to plot density on a square-root scale

```
. twoway__histogram_gen bp, start(125) width(5) gen(h x)
. qui gen hsqrt = sqrt(h)
. twoway bar hsqrt x, barw(5) bstyle(histogram) ytitle(Density)
> ylabel(0 .05 ".0025" .1 ".01" .15 ".0225" .2 ".04")
  (output omitted)
```

- Plotting the differences between observed and expected frequencies

```
. twoway__histogram_gen bp, freq start(125) w(5) gen(h x, replace)
. qui summarize bp
. qui gen diff = h - r(N) * (norm((x + 2.5 - r(mean))/r(sd)) -
> norm((x - 2.5 - r(mean))/r(sd)))
. twoway bar diff x, barw(5) yti("Observed - expected frequency")
  (output omitted)
```

There are also two similar commands: `twoway__function_gen` to generate functions and `twoway__kdensity_gen` to generate kernel densities.