



**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 F. 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 Zürich, Switzerland

Stephen Jenkins  
University of Essex

Ulrich Kohler  
WZB, Berlin

**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

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. Stata and Mata are registered trademarks of StataCorp LP.

## Stata tip 38: Testing for groupwise heteroskedasticity

Christopher F. Baum  
Department of Economics  
Boston College  
Chestnut Hill, MA 02467  
baum@bc.edu

A natural source of heteroskedasticity in many kinds of data is *group membership*: observations in the sample may be a priori defined as members of groups, and the variance of a series may differ considerably across groups. This concept will also apply to the errors from a linear regression. The assumption of homoskedasticity in the relationship may reasonably hold within each group, but not between groups. This assumption most commonly arises in cross-sectional datasets. In economic data, for instance, the groups may correspond to firms in different industries or workers in different occupations. It could also apply in a time-series context: for instance, the variance of daily temperature may not be constant over the four seasons. In any case, a test for heteroskedasticity of this sort should take this a priori knowledge into account.

How might we test for groupwise heteroskedasticity in a variable or in the errors from a regression? In the context of regression, if we can argue that each group's regression equation satisfies the classical assumptions (including that of homoskedasticity), the  $s^2$  computed by `regress` (see [R] `regress`) is a consistent estimate of the group-specific variance of the disturbance process. For two groups, an  $F$  test may be constructed, with the larger variance in the numerator; the degrees of freedom are the residual degrees of freedom of each group's regression. Conducting an  $F$  test is easy if both groups' residuals are stored in one variable, with a group variable indicating group membership (in this case 1 or 2). The third form of `sdtest` may then be used, with the `by(groupvar)` option, to conduct the  $F$  test.

What if there are more than two groups across which we wish to test for equality of disturbance variance, for instance, a set of 10 industries? We may then use the `robvar` command (see [R] `sdtest`), which like `sdtest` expects to find one variable containing each group's residuals, with a group membership variable identifying them. The `by(groupvar)` option is used here as well. The test conducted is that of Levene (1960) labeled as  $W_0$ , which is robust to nonnormality of the error distribution. Two variants of the test proposed by Brown and Forsythe (1974), which uses more robust estimators of central tendency (e.g., median rather than mean),  $W_{50}$  and  $W_{10}$ , are also computed.

We illustrate groupwise heteroskedasticity with state-level data: 1 observation per year for each of the six states in the New England region of the United States for 1981–2000. We first apply `robvar` to the state-level population series to examine whether the variance of population is constant across states.

```
. use http://www.stata-press.com/data/imeus/NEdata
. robvar pop, by(state)
```

state	Summary of pop		Freq.
	Mean	Std. Dev.	
CT	3276614.5	81452.212	20
MA	6030915.5	178354.76	20
ME	1212718.1	46958.538	20
NH	1094238.9	94362.302	20
RI	1000209.9	29548.701	20
VT	562960.65	31310.625	20
Total	2196276.3	1931629.4	120

W0 = 13.856324 df(5, 114) Pr > F = 0.00000000  
W50 = 11.820938 df(5, 114) Pr > F = 0.00000000  
W10 = 13.306895 df(5, 114) Pr > F = 0.00000000

All forms of the test clearly reject the hypothesis of homoskedasticity across states' population series: hardly surprising when the standard deviation of Massachusetts' (MA) population is six times that of Rhode Island (RI).

We now fit a linear trend model to state disposable personal income per capita, `dpipc`, by regressing that variable on `year`. The residuals are tested for equality of variances across states with `robvar`.

```
. regress dpipc year
```

Source	SS	df	MS			
Model	3009.33617	1	3009.33617	Number of obs =	120	
Residual	806.737449	118	6.83675804	F( 1, 118) =	440.17	
Total	3816.07362	119	32.0678456	Prob > F =	0.0000	
				R-squared =	0.7886	
				Adj R-squared =	0.7868	
				Root MSE =	2.6147	

  

dpipc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
year	.8684582	.0413941	20.98	0.000	.7864865	.9504298
_cons	-1710.508	82.39534	-20.76	0.000	-1873.673	-1547.343

```
. predict double eps, residual
. robvar eps, by(state)
```

state	Summary of Residuals		Freq.
	Mean	Std. Dev.	
CT	4.167853	1.3596266	20
MA	1.618796	.86550138	20
ME	-2.9841056	.93797625	20
NH	.51033312	.61139299	20
RI	-.8927223	.63408722	20
VT	-2.4201543	.71470977	20
Total	-6.063e-14	2.6037101	120

W0 = 4.3882072 df(5, 114) Pr > F = 0.00108562  
W50 = 3.2989851 df(5, 114) Pr > F = 0.00806751  
W10 = 4.2536245 df(5, 114) Pr > F = 0.00139064

The hypothesis of equality of variances is soundly rejected by all three `robvar` test statistics, with the residuals for Connecticut, Massachusetts, and Maine possessing a standard deviation considerably larger than those of the other three states.

## References

- Brown, M. B., and A. B. Forsythe. 1974. Robust tests for the equality of variances. *Journal of the American Statistical Association* 69: 364–367.
- Levene, H. 1960. Robust tests for equality of variances. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann, 278–292. Menlo Park, CA: Stanford University Press.