

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

Executive 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

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

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

Jeffrey Wooldridge
Michigan State University

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 Technical Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press, and Stata is a registered trademark of StataCorp LP.

Stata tip 1: The `eform()` option of `regress`

Roger Newson, King's College London, UK
roger.newson@kcl.ac.uk

Did you know about the `eform()` option of `regress`? It is very useful for calculating confidence intervals for geometric means and their ratios. These are frequently used with skewed Y -variables, such as house prices and serum viral loads in HIV patients, as approximations for medians and their ratios. In Stata, I usually do this by using the `regress` command on the logs of the Y -values, with the `eform()` and `noconstant` options. For instance, in the `auto` dataset, we might compare prices between non-US and US cars as follows:

```
. sysuse auto, clear
(1978 Automobile Data)
. generate logprice = log(price)
. generate byte baseline = 1
. regress logprice foreign baseline, noconstant eform(GM/Ratio) robust
Regression with robust standard errors
```

Number of obs =	74
F(2, 72) =	18043.56
Prob > F =	0.0000
R-squared =	0.9980
Root MSE =	.39332

logprice	GM/Ratio	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
foreign	1.07697	.103165	0.77	0.441	.8897576	1.303573
baseline	5533.565	310.8747	153.41	0.000	4947.289	6189.316

We see from the `baseline` parameter that US-made cars had a geometric mean price of 5534 dollars (95% CI from 4947 to 6189 dollars), and we see from the `foreign` parameter that non-US cars were 108% as expensive (95% CI, 89% to 130% as expensive). An important point is that, if you want to see the baseline geometric mean, then you must define the constant variable, here `baseline`, and enter it into the model with the `noconstant` option. Stata usually suppresses the display of the intercept when we specify the `eform()` option, and this trick will fool Stata into thinking that there is no intercept for it to hide. The same trick can be used with `logit` using the `or` option, if you want to see the baseline odds as well as the odds ratios.

My nonstatistical colleagues understand regression models for log-transformed data a lot better this way than any other way. Continuous X -variables can also be included, in which case the parameter for each X -variable is a ratio of Y -values per unit change in X , assuming an exponential relationship—or assuming a power relationship, if X is itself log-transformed.