



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
979-845-8817; 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 Bellocco
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

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

Stata Press Editorial Manager
Stata Press Copy Editors

Lisa Gilmore
Deirdre Skaggs

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

Stata tip 107: The baseline is now reported

Maarten L. Buis
Department of Sociology
Tübingen University
Tübingen, Germany
maarten.buis@uni-tuebingen.de

For a long time, Stata has had the capability to report exponentiated coefficients. Examples are the `or` option of `logit` and `ologit` (see [R] `logit` and [R] `ologit`); the `irr` option of `poisson`, `zip`, and `nbreg` (see [R] `poisson`, [R] `zip`, and [R] `nbreg`); and the `hr` and `tr` options of `streg` (see [ST] `streg`). Also see Newson (2003) and Buis (2010) for details. These exponentiated coefficients can be interpreted as odds ratios, incidence-rate ratios, hazard ratios, or time ratios. However, until Stata 12 the baseline odds, incidence rate, hazard, or time—that is, the exponentiated constant—was not reported. That was unfortunate because this baseline can be helpful for evaluating the size of the effects, and it provides a convenient way of discussing the exact interpretation of the coefficients.

As of Stata 12, this omission has been redressed. The usefulness of the baseline value and a couple of caveats are illustrated using the example below.

```
. sysuse nlsw88
(NLSW, 1988 extract)

. generate c_grade = grade - 12
(2 missing values generated)

. generate high_occ = occupation < 3 if occupation < .
(9 missing values generated)

. logit union c_grade i.high_occ, or nolog

Logistic regression                Number of obs   =       1867
                                   LR chi2(2)        =       49.44
                                   Prob > chi2         =       0.0000
                                   Pseudo R2           =       0.0237

Log likelihood = -1016.5579
```

union	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
c_grade	1.123325	.0248694	5.25	0.000	1.075624	1.173141
1.high_occ	.4651723	.0644307	-5.53	0.000	.3545803	.6102575
_cons	.3358115	.02213	-16.56	0.000	.2951218	.3821112

Odds ratios have a bad reputation for being hard to interpret. Part of the problem is that many people are not used to working with odds. Researchers rarely frequent race tracks or betting shops. Starting the results section of an article with interpreting the baseline odds is a nice way to remind the readers of the correct interpretation. This trick works well because it fits naturally within the normal format of an academic article. In this case, we expect to find 0.34 union members for every nonmember within the group of respondents that has 12 years of education (`c_grade` = 0) and a lower occupation (`high_occ` = 0). The odds ratios tell us that the odds increases by a factor of 1.12 or

12% $[(1.12 - 1) \times 100\% = 12\%]$ for every additional year of education, while the odds decreases 53% $[(0.47 - 1) \times 100\% = -53\%]$ when the respondent has a high occupation. Reporting the baseline odds in the results section of a paper allows you to translate the abstract concept of odds to the concrete situation that is being studied; in this case, it allows you to translate “the number of successes per failure” to “the number of union members per nonmember”.¹

A 53% decrease in the odds of being a union member sounds like a large effect. However, we can get a better understanding of the size of this effect by comparing it with the baseline odds. In this case, the odds changes from 0.34 union members per nonunion member for respondents with lower occupations to 0.16 $(0.47 \times 0.34 = 0.16)$ union members per nonmember, which is a substantively meaningful change. But what if being a union member was very rare? For example, assume that the baseline odds was 0.001 union member for every nonunion member. In that case, the odds would change from 0.001 to 0.00047 union members per nonmember when a respondent obtained a high occupation, which does not sound nearly as impressive as a change of -53% . So the baseline value can play an important role in evaluating how large an effect is.

There are, however, a couple of things you need to consider when interpreting these baseline values. First, the baseline value is the value when all explanatory variables are 0. So to get a meaningful baseline value, you need to make sure the value 0 is meaningful for all explanatory variables. In the example above, I did so by centering the variable `grade` at 12 years of education (obtaining high school). Second, the practice of reporting p -values or assigning stars to significant parameters needs a bit of thought in the case of baseline values. Stata automatically reports the results of the test of the null hypothesis that the coefficient is 0, and thus the exponentiated coefficient is 1. In the example, that would mean that the null hypothesis for the baseline odds is that there is 1 union member for every nonmember; that is, the probability of being a union member is 50%.

References

- Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. 2nd ed. Hoboken, NJ: Wiley.
- Buis, M. 2010. Stata tip 87: Interpretation of interactions in nonlinear models. *Stata Journal* 10: 305–308.
- Fienberg, S. E. 2007. *The Analysis of Cross-Classified Categorical Data*. 2nd ed. New York: Springer.
- Newson, R. 2003. Stata tip 1: The `eform()` option of `regress`. *Stata Journal* 3: 445.

1. More complete discussions of odds and odds ratios can be found in the textbooks by [Fienberg \(2007\)](#) and [Agresti \(2007\)](#).