

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
jnnewton@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 42: The overlay problem: Offset for clarity

James Cui

Department of Epidemiology and Preventive Medicine

Monash University

Melbourne, Australia

james.cui@med.monash.edu.au

A common graphical problem often arises when one graph axis shows a discrete scale and the other shows a continuous scale. The discrete scale could, for example, represent distinct categories or a series of times at which data were observed. If we want to show several quantities on the continuous axis, matters may easily become confused—and confusing—when some of those quantities are close, especially if they are shown as confidence or other intervals. One answer to this overlap problem is to offset for clarity.

For example, in longitudinal studies, we often need to draw the mean response and 95% confidence intervals of a continuous variable for several categories over the follow-up period. However, the confidence intervals can overlap if the difference between the mean responses is small. Consider an example closely based on one in [Rabe-Hesketh and Everitt \(2004, 144–166\)](#). Mean and standard deviation of depression score, `dep` and `sddep`, have been calculated for each of five visits and two treatment groups, `visit` and `group`. The number of subjects in each combination of visit and group is also given as `n`, so that approximate 95% confidence limits `high` and `low` can be based on twice the standard error, `sddep` / \sqrt{n} . See table 1.

Table 1: Mean and standard deviation of depression score over visit

visit	group	dep	sddep	n	high	low
1	Placebo	16.48	5.28	27	18.51	14.45
1	Estrogen	13.37	5.56	34	15.28	11.46
2	Placebo	15.89	6.12	22	18.50	13.28
2	Estrogen	11.74	6.58	31	14.10	9.38
3	Placebo	14.13	4.97	17	16.54	11.72
3	Estrogen	9.13	5.48	29	11.17	7.09
4	Placebo	12.27	5.85	17	15.11	9.43
4	Estrogen	8.83	4.67	28	10.60	7.06
5	Placebo	11.40	4.44	17	13.55	9.25
5	Estrogen	7.31	5.74	28	9.48	5.14

To plot these results, we first use `clonevar` to make a copy of `visit` as `x`: that way, `x` inherits format and value labels as well as values from `visit`, not important here but useful in other problems. We copy so that the original `visit` remains unchanged. Adding and subtracting a small value depending on `group` offsets the two

groups. Clearly, the value here, 0.05, can be varied according to taste. If there had been three groups, we could have left one where it was and moved the other two. Because the number of groups is either even or odd, a symmetric placement around integer values on the discrete axis can thus be achieved either way.

```
. use depression
. clonevar x = visit
. replace x = cond(group == "Placebo", x - 0.05, x + 0.05)
x was byte now float
(10 real changes made)

. twoway (connected dep x if group == "Placebo", lpattern(solid) msymbol(D))
> (connected dep x if group == "Estrogen", lpattern(dash) msymbol(S))
> (rcap high low x if group == "Placebo")
> (rcap high low x if group == "Estrogen")
> , xlabel(1 2 3 4 5) ylab(5(5)20, format(%5.0f))
> xttitle("Visit") ytitle("Depression score")
> legend(pos(1) ring(0) col(1) order(1 "Placebo" 2 "Estrogen"))
```

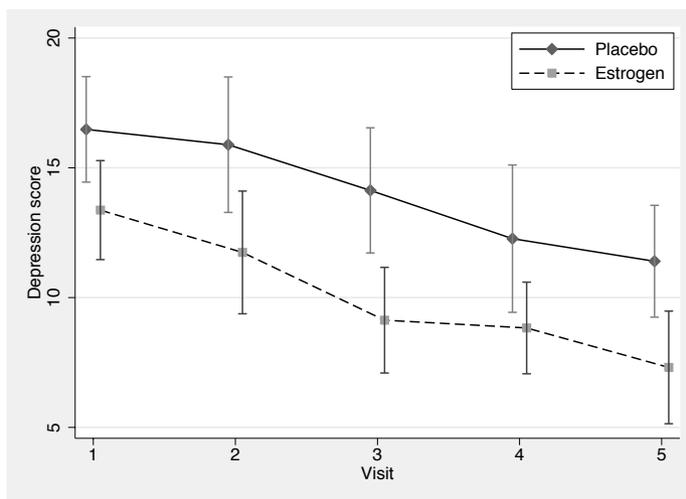


Figure 1: Mean depression score and 95% confidence intervals over visit

References

Rabe-Hesketh, S., and B. Everitt. 2004. *A Handbook of Statistical Analyses Using Stata*. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC.