

A publication to promote communication among Stata users

Editor

H. Joseph Newton
Department of Statistics
Texas A & M University
College Station, Texas 77843
979-845-3142
979-845-3144 FAX
stb@stata.com EMAIL

Associate Editors

Nicholas J. Cox, University of Durham
Joanne M. Garrett, University of North Carolina
Marcello Pagano, Harvard School of Public Health
J. Patrick Royston, UK Medical Research Council
Jeroen Weesie, Utrecht University

Subscriptions are available from Stata Corporation, email stata@stata.com, telephone 979-696-4600 or 800-STATAPC, fax 979-696-4601. Current subscription prices are posted at www.stata.com/bookstore/stb.html.

Previous Issues are available individually from StataCorp. See www.stata.com/bookstore/stbj.html for details.

Submissions to the STB, including submissions to the supporting files (programs, datasets, and help files), are on a nonexclusive, free-use basis. In particular, the author grants to StataCorp the nonexclusive right to copyright and distribute the material in accordance with the Copyright Statement below. The author also grants to StataCorp the right to freely use the ideas, including communication of the ideas to other parties, even if the material is never published in the STB. Submissions should be addressed to the Editor. Submission guidelines can be obtained from either the editor or StataCorp.

Copyright Statement. The Stata Technical Bulletin (STB) and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp. 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 STB.

The insertions appearing in the STB 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 STB. Written permission must be obtained from Stata Corporation if you wish to make electronic copies of the insertions.

Users of any of the software, ideas, data, or other materials published in the STB or the supporting files understand that such use is made without warranty of any kind, either by the STB, the author, or Stata Corporation. 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 STB is to promote free communication among Stata users.

The *Stata Technical Bulletin* (ISSN 1097-8879) is published six times per year by Stata Corporation. Stata is a registered trademark of Stata Corporation.

Contents of this issue

	page
dm65.1. Update to a program for saving a model fit as a dataset	2
dm82. Simulating two- and three-generation families	2
gr45. A turnip graph engine	5
sbe19.3. Tests for publication bias in meta-analysis: erratum	8
sbe39.1. Nonparametric trim and fill analysis of publication bias in meta-analysis: erratum	8
sg84.3. Concordance correlation coefficient: minor corrections	9
sg97.2. Update to formatting regression output	9
sg153. Censored least absolute deviations estimator: CLAD	13
sg154. Confidence intervals for the ratio of two binomial proportions by Koopman's method	16
sg155. Tests for the multinomial logit model	19
sg156. Mean score method for missing covariate data in logistic regression models	25
sg157. Predicted values calculated from linear or logistic regression models	27
snp15.2. Update to somersd	30
snp16. Robust confidence intervals for median and other percentile differences between two groups	30
sts15.1. Tests for stationarity of a time series: update	35
sxd2. Computing optimal sampling designs for two-stage studies	37
sxd3. Sample size for the kappa-statistic of interrater agreement	41

Acknowledgment

I am grateful to Dr. John Moran for alerting me to this error.

References

- Bradburn, M. J., J. J. Deeks, and D. G. Altman. 1998. sbe24: metan—an alternative meta-analysis command. *Stata Technical Bulletin* 44: 4–15. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 86–100.
- Duval, S. and R. Tweedie. 2000. A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association* 95(449): 89–98.
- Steichen, T. J. 2000. sbe39: Nonparametric “trim and fill” analysis of publication bias in meta-analysis. *Stata Technical Bulletin* 57: 8–14.

sg84.3

Concordance correlation coefficient: minor corrections

Thomas J. Steichen, RJRT, steicht@rjrt.com

Nicholas J. Cox, University of Durham, UK, n.j.cox@durham.ac.uk

Abstract: This insert fixes some bugs and corrects some defaults affecting detailed user control of graphical output in the program for concordance correlation. The numerical calculations were not changed.

Keywords: concordance correlation, graphics, measurement comparison.

Description

`concord` computes the concordance correlation coefficient for agreement on a continuous measure obtained by two persons or methods and provides optional graphical displays. A full description of the method and of the operation of the command was given by Steichen and Cox (1998a), with revisions and updates in Steichen and Cox (1998b, 2000).

This insert fixes some small bugs affecting detailed user control of graphical display through the `connect()`, `symbol()`, and `pen()` options. It also corrects a few of the corresponding default values.

Specifically, the default `graph_options` for `graph(1oa)` now include `connect(111.1)`, `symbol(iiii)`, and `pen(35324)` for the lower confidence interval limit, mean difference, upper confidence interval limit, data points, and regression line (if requested) respectively, along with default titles and labels. Further, the visual characteristics of the data points and line in the normal probability plot now follow those of the data points and regression line in the associated loa (limits-of-agreement) plot.

Acknowledgment

We are grateful to Sónia Dória Nóbrega for alerting us to the bugs.

References

- Steichen, T. J. and N. J. Cox. 1998a. sg84: Concordance correlation coefficient. *Stata Technical Bulletin* 43: 35–39. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 137–143.
- . 1998b. sg84.1: Concordance correlation coefficient, revisited. *Stata Technical Bulletin* 45: 21–23. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 143–145.
- . 2000. sg84.2: Concordance correlation coefficient: update for Stata 6. *Stata Technical Bulletin* 54: 25–26. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 169–170.

sg97.2

Update to formatting regression output

John Luke Gallup, Harvard University, jgallup@hiid.harvard.edu

Abstract: An update to the `outreg` command is described.

Keywords: regression output.

I have updated `outreg`, a program described in Gallup (1998, 1999) that writes regression output to a text file. New features allow including user-specified statistics and notes, 10% asterisks, table and column titles, scientific notation for coefficient estimates, and reporting of confidence interval and marginal effects. I have also fixed bugs in the reporting results from `dprobit`, `heckman`, and `heckprob` regressions, and `outreg` can now be used after `dlogit2`, `dprobit2`, and `dmlogit2`. Because `outreg` has been so extensively modified, we describe its full syntax.

Syntax

```

outreg [varlist] using filename [, nolabel title(textlist) ctitle(textlist) nonotes addnote(textlist)
      bdec(numlist) bfmt(textlist) coefastr {se | pvalue} ci tdec(#) noparen bracket
      {noaster | 3aster | 10pct} sig symb(textlist) nocons nonobs noni nor2 adjr2 rdec(#)
      addstat(text, # [, text, #, ... ]) adec(#) eform marginal onecol xstats
      comma append replace ]

```

`outreg` is available after any estimation command. A *textlist* is a list of text separated by commas. It is similar to a *numlist* or a *varlist*, but commas are required. Each text element in the list does not need to be enclosed in quotation marks unless the text contains commas or parentheses.

Description

`outreg` formats regression output as it is presented in most documents: *t* statistics or standard errors in parentheses under each coefficient, asterisks indicating coefficients statistically different from zero, and summary statistics like *R*-squared at the bottom. The formatted output is written to a tab- or comma-separated ASCII file, which can then be loaded into word processing or spreadsheet programs to be converted directly to a table. For example, the output file can be opened in Microsoft Word, the text selected, and then converted to a table by choosing “Table”, then “Convert Text to Table”. Note that when loading the output into a spreadsheet, the parentheses around the *t* statistics may convert to negative numbers.

`outreg` should work after any Stata estimation command. Like `predict`, `outreg` makes use of internally saved estimation results, so it should be invoked immediately after the estimation. In addition to coefficient estimates, by default `outreg` will report *t* statistics with asterisks for standard significance levels (1% and 5%), numbers of observations, true *R*-squareds (no pseudo *R*-squareds), and the number of groups in panel estimation. The user can add their own chosen titles (*title* and *ctitle*), statistics (`addstat`), and notes (`addnote`) to the table, and change many other aspects of the output.

`outreg` rewards the use of variable labels (and value labels for `mlogit`, `svylog`, and `dlogit2`). The variable labels are used in the output table (unless `nolabel` is chosen), providing more intelligible variable descriptions than 8-letter names. If different variables are assigned the same variable label (not usually done intentionally), and more than one regression is appended together, the coefficients and *t* statistics will not be properly ordered. The solution is to use distinct variable labels or the `nolabel` option.

If *filename* is specified without an extension, `.out` is assumed.

Several regressions with differing variables can be combined into a single table with the `append` option.

If a *varlist* is specified, only the regression coefficients corresponding to the variables in *varlist* will be included in the table. The intercept coefficient is included as well unless the `nocons` option is chosen. This is probably most useful for excluding numerous dummy variable coefficients. Time series prefixes are not allowed when using an explicit *varlist*.

Text-related options

`nolabel` specifies that variable names rather than variable labels be used to identify coefficients. It also suppresses the value labels of the dependent variable in `mlogit` and `svylog`.

`title`(*textlist*) specifies a title or titles at the top of the regression table. The maximum title length is 80 characters. Additional characters will be cut off. Longer titles can be put in two or more title lines. When regression results are appended together, the table title(s) must be specified in the first `outreg` call; titles specified in subsequent `outreg` calls using the `append` option will be ignored. Note that when converting the `outreg` text output to a table in a word processor or a spreadsheet, it is easier to leave the title row out of the text selected for conversion.

`ctitle`(*textlist*) specifies the regression title above the coefficient column. By default if no column title is specified, the label or name of the dependent variable is displayed. Multiple column titles are only appropriate for multi-equation regressions, using one title per equation, and then only if not `onecol`.

`nonotes` specifies that notes explaining the *t* statistics (or standard errors) and asterisks not be included.

`addnote`(*textlist*) specifies user-added notes to be displayed in new lines at the bottom of the `outreg` table. When regression results are appended together, `addnote` must be specified in the first `outreg` call; `addnotes` specified in subsequent `outreg` calls using the `append` option will be ignored. `addnote` is consistent with `nonotes`. A blank line can be inserted by including a blank within quotes as a note.

One technical note is in order. Text which includes quotation marks within the text (by means of double quotation) in `title`, `ctitle`, and `addnote` displays correctly in single regression tables but does not display correctly when subsequent regressions are appended using the `append` option.

Coefficient options

`bdec(numlist)` specifies the number of decimal places reported for coefficient estimates. It also specifies the decimal places reported for standard errors or confidence intervals if `se` or `ci` is chosen. The default value is 3. The minimum value is 0, and the maximum is 11. If one number is specified in `numlist`, it will apply to all coefficients. If multiple numbers are specified, the first number will determine the decimals reported for the first coefficient; the second number, the decimals for the second coefficient; and so on. If there are fewer numbers in the `numlist` than coefficients, the last number in the `numlist` will apply to all the remaining coefficients.

`bfmt(textlist)` specifies the format type for coefficient estimates (and standard errors or confidence intervals, if `se` or `ci` is chosen). Possible format types include `e` for scientific notation (for example, `1.00e+3`), `f` or `fc` for fixed format (with commas for thousands with `fc`), and `g` or `gc` for general format (with commas for thousands with `gc`). The default for `bfmt` is `fc`. If multiple format types are specified, they are applied to the coefficients the way that multiple `bdec` parameters are applied. This option is mainly to allow scientific notation. For an explanation of Stata numeric formats, see [U] 15.5.1 **Numeric formats**.

`coefastr` specifies that asterisks for significance levels are appended to regression coefficients rather than to t statistics or standard errors.

Options for statistics, standard error, etc.

`se` specifies that standard errors rather than t statistics are reported.

`pvalue` specifies that p -values (of t statistics) rather than t statistics are reported.

`ci` specifies that confidence intervals of coefficients rather than t statistics are reported.

`tdec(#)` specifies the number of decimal places reported for t statistics (or for p -values if `pvalue` is specified). It also specifies the decimal places reported for R -squared or adjusted R -squared if they are not specified in `rdec`. The default value for `tdec` is usually 2, but 3 if `pvalue` is specified. The minimum value is 0, and the maximum is 11.

`noparen` specifies that no parentheses be placed around t statistics or standard errors.

`bracket` specifies that square brackets be used rather than parentheses around t statistics or standard errors.

`noaster` specifies that no asterisks denoting 1% and 5% significance levels be reported.

`3aster` specifies 3 asterisks for 1%, 2 asterisks for 5%, and 1 asterisk for 10% significance levels.

`10pct` specifies a plus sign for 10% significance levels in addition to the default 2 asterisks for 1%, and 1 asterisk for 5% significance levels.

`sig symb(textlist)` specifies symbols for 1% and 5% significance levels (and 10% significance level if `10pct` is also chosen). The specified symbols replace the asterisks. Quotation marks around the new symbols are optional if the characters “,” and “)” are avoided. Omitting symbols will prevent the significance level from being labeled (see also `noaster`). For example, to display only 1% significance levels, one could use `outreg using table1, sig symb(*)`.

Options for statistics

`nocons` specifies that the intercept (constant) coefficient estimate not be reported.

`nonobs` specifies that the number of observations in the estimation not be reported.

`noni` specifies that the number of groups in a panel data regression not be reported (for example, the number of groups specified by the `i()` variable in `xtreg`).

`nor2` specifies that no R -squared (or adjusted R -squared) be reported. This option is only meaningful when Stata calculates a true R -squared.

`adjr2` specifies that the adjusted R -squared be reported rather than the regular R -squared.

`rdec(#)` specifies the number of decimal places reported for the R -squared or adjusted R -squared. The default value for `rdec` is the value for `tdec`. The minimum value is 0, and the maximum is 11.

`addstat(text, # [, text, #, ...])` specifies user-added statistics to be displayed in new lines below the *R*-squared (if shown). The user must specify both a name and a value for the statistic. Users can report significance levels of test statistics as a second statistic to be shown on the line below the first statistic.

`adec(numlist)` specifies the number of decimal places reported for user-added statistics (in `addstat`). The default value for `rdec` is the value for `tdec`. The minimum value is 0, and the maximum is 11. If one number is specified in `numlist`, it will apply to all statistics. If multiple numbers are specified in `numlist`, they are applied to the user-added statistics as in `bdec`.

`eform` specifies that the exponential form of coefficients be reported. This corresponds to the `or` option for `logit`, `clogit`, and `glogit` estimation, `irr` for `poisson` estimation, `rrr` for `mlogit`, `hr` for `cox` hazard models, and `eform` for `xtgee`, but it can be used to exponentiate the coefficients after any estimation; see *Methods and Formulas* in [R] `maximize`.

`marginal` specifies that the marginal effects rather than the coefficient estimates are reported. This is done automatically after `dprobit`.

`onecol` specifies that multi-equation models (for example, `mlogit`, `reg3`) be formatted in one column rather than the default of multiple columns, one column per equation. It also reports extra statistics included in the `e(b)` vector.

`xstats` specifies that the extra statistics included in the `e(b)` matrix be reported. Extra statistics for multi-equation models (for example, `heckman`, `heckprob`, and `biprobit`) are not reported; a user can use `addstat` or `onecol`. If there are no extra statistics in the `e(b)` matrix, `xstats` is ignored. This option is largely superseded by `addstat`.

Other options

`comma` specifies that the ASCII file output be separated by commas rather than by tabs. This can cause problems if any of the user-defined text has commas in it (such as variable labels, `title`, `ctitle`, `addstat`, or `addnote`).

`append` specifies that new estimation output be appended to an existing output file. In general, the same `outreg` options should be used in the original regression output and each appended regression. The notes at the bottom of the table explaining the *t* statistics or standard errors and asterisks are correct for the first estimation in the output file. If subsequently appended estimation results use different options (such as a switch to `noaster`, or changes the estimation's `robust` option), the notes will not be appropriate for all the columns. This problem can be addressed with a combination of `nonotes` and `addnote`.

`replace` specifies that it is okay to replace *filename* if it already exists.

Examples

We begin by using Stata's automobile data.

```
. generate weight2=weight^2
. regress mpg weight weight2 foreign
```

Source	SS	df	MS			
Model	1689.15372	3	563.05124	Number of obs =	74	
Residual	754.30574	70	10.7757963	F(3, 70) =	52.25	
				Prob > F =	0.0000	
				R-squared =	0.6913	
				Adj R-squared =	0.6781	
Total	2443.45946	73	33.4720474	Root MSE =	3.2827	

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
weight	-.0165729	.0039692	-4.175	0.000	-.0244892	-.0086567
weight2	1.59e-06	6.25e-07	2.546	0.013	3.45e-07	2.84e-06
foreign	-2.2035	1.059246	-2.080	0.041	-4.3161	-.0909003
_cons	56.53884	6.197383	9.123	0.000	44.17855	68.89913

```
. outreg using outreg1
. !more outreg1.out
      Mileage (mpg)
Weight (lbs.)  -0.017
               (4.18)**
weight2 0.000
           (2.55)*
Car type      -2.204
               (2.08)*
Constant      56.539
               (9.12)**
Observations  74
R-squared      0.69
Absolute value of t-statistics in parentheses
* significant at 5%; ** significant at 1%
```

Next we use some of the `outreg` options.

```
. outreg weight weight2 using outreg1, se bdec(4) bfmt(e) tdec(3) nonotes replace
. !more outreg1.out
      Mileage (mpg)
Weight (lbs.)   -1.6573e-02
                (3.9692e-03)**
weight2 1.5912e-06
                (6.2489e-07)*
Constant      5.6539e+01
                (6.1974e+00)**
Observations   74
R-squared      0.691
```

Finally, here is an example I used in my research which investigates the relationship of a dependent variable y on four X variables and shows how `outreg` can be used to produce the kinds of tables needed in publications.

```
. regress y x1-x2
. outreg using table4, title("Three Regression Variants")
> addnote("", "Run at $S_TIME, $S_DATE", Using data from $S_FN) replace
. regress y x1-x3
. outreg using table4, append
. regress y x1 x3-x4
. test x3 x4
. outreg using table4, append addstat("F test, x3=x4=0", r(F), Prob > F, r(p)) adec(2,3)
. type table4.out

      Three Regression Variants
      (1)      (2)      (3)
      y        y        y
x1      0.686    0.723   -0.388
      (4.47)**  (3.79)**  (0.62)
x2     -3.781   -3.809
      (25.66)** (22.21)**
x3              4.107   57.701
              (0.33)   (1.84)
x4              21.560
              (1.86)
Constant    1.330    0.224  -45.005
              (0.40)   (0.05)   (1.82)
Observations 100     100     100
R-squared    0.99    0.99    0.99
F test, x3=x4=0          1.78
Prob > F              0.174
Absolute value of t-statistics in parentheses
* significant at 5% level; ** significant at 1% level
Run at 12:10:19, 18 Apr 2000
Using data from xydata.dta
```

Acknowledgments

Thanks to many Statalist users for bug reports and suggestions. Mead Over's comments were especially useful.

References

- Gallup, J. L. 1998. sg97: Formatting regression output for published tables. *Stata Technical Bulletin* 46: 28–30. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 200–202.
- . 1999. sg97.1: Revision of `outreg`. *Stata Technical Bulletin* 49: 23. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 170–171.

sg153	Censored least absolute deviations estimator: CLAD
-------	--

Dean Jolliffe, Economic Research Service, U.S. Department of Agriculture, jolliffe@ers.usda.gov
 Bohdan Krushelnytsky, Czech Republic, bohdan.krushelnytsky@cerge.cuni.cz
 Anastassia Semykina, Czech Republic, anastassia.semykina@cerge.cuni.cz

Abstract: The command `clad` for estimating Powell's (1984) censored least absolute deviations estimator and bootstrap estimates of its sampling variance is introduced and illustrated.

Keywords: censored least absolute value deviations estimator, bootstrap.