

# Using Stata's `esttab` and `outreg2` with MS-Word to make publication-style regression tables

*by*

*Brianne Barclay*  
Department of Sociology/Anthropology  
Florida International University  
University Park  
Miami, FL 33199

June 2007

*Note:* Those who prefer to use Excel may type the `esttab` option `csv` or the `outreg2` option `excel` to display the regression results in an Excel spreadsheet, then click the resulting link in the Stata results window to display the Excel output. Another alternative is to type the `tex` option to create LaTeX output. Other useful Stata commands for making regression tables include `estout`, `mktab`, `modl`, and `xml_tab`.

To download `hsb2.dta`:

<http://www.ats.ucla.edu/STAT/stata/notes/control.htm>

use `hsb2`, clear

findit `esttab` [download]      or: ssc install `esttab`  
findit `outreg2` [download]      ssc install `outreg2`

- `esttab`

```
reg science read write
eststo
reg science read write math female
eststo
```

```
esttab using science.rtf, se b(%9.2f) ar2(%9.2f) starlevels(* .05 ** .01 ***
.001) nodelvars nomtitles title("OLS Science Models") addnotes("Note: Put
your notes here.") replace
```

*Then click on the output link in the Stata results window.*

- *outreg2*

```
reg science read write  
outreg2 using science, ctitle(Model) bdec(2) tdec(2) rdec(2) adec(2)  
alpha(.001, .01, .05) addstat(Adj. R-squared, e(r2_a)) addnote(Put your  
notes here.) word replace
```

```
reg science read write math female  
outreg2 using science, ctitle(Model) bdec(2) tdec(2) rdec(2) adec(2)  
alpha(.001, .01, .05) addstat(Adj. R-squared, e(r2_a)) word append
```

If an error message appears, carefully check that the formatting of the *outreg2* command is correct. For example, while the above commands produce the same output whether or not the content of *ctitle* and *addnote* is contained within quotation marks, they do not work if *addnote* is specified as follows: *addnote*(Note: Put your notes here.) That is, typing the colon and "Note" does not permit the command to work.

Here are some other useful *outreg2* options: *ci*, *beta*, *eform* [exponentiated], *symbol*(+, \*, \*\*) [to change the p-value significance symbols], *onecol* [output displayed in one column only].

After typing the *outreg2* commands, click the *blue link* in Stata to see ('seeout') the table results in Stata's spreadsheet, or click the *rtf link* to see the Word formatted text. *outreg2* has created *txt* and *rtf* files that contain the regression results. These files are located in the folder in which you are working.

*txt* is easy to deal with: just tab things into alignment. *rtf*'s advantage is that it has already formatted an elegant output table, but its disadvantage is that fine-tuning the table often involves heavy MS-Word table-formatting.

- Here are step-by-step instructions for doing necessary table editing in Word. The example uses the table produced by *outreg2*.

## The Initial Outreg2 *rtf* Document

COEFFICIENT	(1) Model	(2) Model
read	0.26*** (0.07)	0.26*** (0.07)
write	0.30*** (0.07)	0.30*** (0.07)
math	0.28*** (0.08)	0.28*** (0.08)
female	-3.57*** (1.06)	-3.57*** (1.06)
Constant	9.29** (3.12)	9.29** (3.12)
Observations	200	200
R-squared	0.53	0.53
Adj. R-squared	0.52	0.52

Standard errors in parentheses  
\*\*\* p<0.001, \*\* p<0.01, \* p<0.05  
“Put your notes here.”

### Editing Steps

1. *Note:* It’s sometimes preferred that the decimals align with each other. This example will do so.
2. Edit column and row headings
  - a. Remove the numbers above the column headings (that is, delete “(1)” and “(2)”).
  - b. Merge the two (now empty) cells formerly occupied by “(1)” and “(2)” and then label appropriately.
    - i. Left click, hold down the cursor, and drag across both cells.
    - ii. Right click in the selected area and click *Merge Cells*.
    - iii. Enter a title such as “Coefficients” in the newly merged cell.
  - c. Rename the heading for the left-hand side’s row column, changing “COEFFICIENT” either to “VARIABLES” or, if you prefer to display the name of the dependent variable, “SCIENCE.” Alternatively, delete “COEFFICIENT” and leave the space blank.
  - d. Change the fonts of VARIABLES or SCIENCE, and Coefficients (located in the top right area of the table) to *Small Caps*.
    - i. Highlight the titles.
    - ii. Go to *Format*.
    - iii. Click *Font*.

- iv. Check *Small Caps*.
  - e. Rename the "Model" column headings to "Model 1" and "Model 2," respectively.
  - f. Merge the second row with the empty cells below in the third row.
    - i. Left-click on the "VARIABLES" or "SCIENCE" cell (if you chose not to leave that area blank), hold down the cursor, and drag downward selecting both cells
    - ii. Right click in the selected area and click *Merge Cells*
    - iii. Do the same individually with the "Model 1" cell and with the "Model 2" cell, merging each individually with the cell below.
3. Add lines
- a. Add a line above "COEFFICIENTS" (located in the top right area).
    - i. Left-click in the "COEFFICIENTS" cell, hold down the cursor, and drag across the cell to the left
    - ii. Right click
    - iii. Click *Borders and Shading*
    - iv. On the right side of the pop-up box click the button so the line option for the top of the cell is darkened
    - v. Under *Apply to:* select *Cell*
    - vi. Click *Ok*
  - b. Add a line below "Adjusted R-squared"
    - i. Left-click in the "Adjusted R-squared" cell, hold down the cursor, and drag across the next two cells
    - ii. Right click
    - iii. Click *Borders and Shading*
    - iv. On the right side of the pop-up box click the button so the line option for the bottom of the cell is darkened
    - v. Under *Apply to:* select *Cell*
    - vi. Click *Ok*
  - c. Add a line below "COEFFICIENTS."
    - i. Right-click in the "New Column Title" cell.
    - ii. Click *Borders and Shading*.
    - iii. On the right side of the pop-up box click the button so the line option for the bottom of the cell is darkened.
    - iv. Under *Apply to:* select *Cell*.
    - v. Click *Ok*.
4. Formatting the column numbers.
- a. Decimal-align the column numbers.
    - i. Select the variable names, names of diagnostic indicators, and the column figures.
    - ii. Format>Tab>set spacing, and select "decimal" alignment.

- iii. Tab column figures as necessary to align the decimals with each other.
  - b. Depending on your preference or format, you may want to delete the standard errors and their surrounding parentheses. If you do so, then remove the rows that include them.
    - i. Left click on the empty cell below "Read," hold down the cursor, and drag across the next two cells
    - ii. Right click in the selected area and click *Delete Cells*
    - iii. Click *Delete entire row*
    - iv. Click *Ok*
    - v. Do the same with the rows below "Write," "Math," "Female," and "Constant," deleting each row
- 5. Editing below the table.
  - a. Below the table, insert the word "Note," followed by a colon, in front of the sentence "Put your notes here." Perhaps italicize "Note" and the colon.
  - b. Delete the phrase "Standard errors in parentheses" which is located below the table, then paste it after "Note" and insert a period.
  - c. Use the toolbar's justification tools to left-justify the text below the table (that is, to align it with the left edge of the table).
  - d. If your format requires, delete the p-value symbols from their current placement and then paste them below the "Note" entry.
    - i. I added "(two-tailed tests)" after the p-value symbols.
  - e. If your format requires, remove the commas between the p-values; reorder the p-values from lowest level of significance to highest; and add three extra spaces between the equations; and remove the '0' before the decimal point for each of the three p-values.
- 6. Row name adjustments
  - a. Perhaps change "Observations" to "N".
  - b. Perhaps change "Adjusted R-squared" to "Adjusted R<sup>2</sup>" (by selecting "2," then clicking Format, Front, and superscript).
- 7. Final Touches
  - a. Add "Table" and its number, followed by a period, above the table, either lower-case or capitalized and either left-justified or centered, depending on your style format.
    - i. To type Table, you may have to click the box with the four arrows at the upper left corner of the table, right click to select all of the table and its associated text, cut, and then paste lower on the page.
  - b. Add a title either on the same line as the table number or under it, depending on the style format. I've italicized *Table 1. OLS Regression Model: Science*.

- c. Of course, there may be additional fine-tuning to do such as adding spacing and italics, again depending on the relevant style format as well as the complexity of your regression model.
  - d. If you need to shift the table and its associated text toward the left-hand side of the page, click the box with the four arrows at the upper left corner of the table, right click to select the entire table, and then click the toolbar's left-justify tool.
8. Save the table in the appropriate text format.

### The Edited Table

*Table 1. OLS Regression Model: Science*

VARIABLES	COEFFICIENTS	
	Model 1	Model 2
read	0.43*** (0.06)	0.26*** (0.07)
write	0.32*** (0.07)	0.30*** (0.07)
math		0.28*** (0.08)
female		-3.57*** (1.06)
Constant	12.51*** (3.16)	9.29** (3.12)
<i>N</i>	200	200
<i>R</i> <sup>2</sup>	0.46	0.53
Adj. <i>R</i> <sup>2</sup>	0.45	0.52

*Note:* Standard errors are in parentheses.

\*\*\* *p*<0.001, \*\* *p*<0.01, \* *p*<0.05 (two-tailed tests)

### Chicago Style Formatting Notes

From **The Chicago Manual of Style**. 15th Edition, 2003.

#### *13.2 Consistency*

Because a prime virtue of tables is easy comparison, consistency in style is indispensable both within one table and among several. Whatever style is chosen for titles, column heads, abbreviations, and the like for one table must be followed in all others in the same work. Similarly, the choice between horizontal rules and extra space or between indentation and font variation to distinguish between elements within a table must be made uniformly for all tables in a work. Certain tables, however, may require rules or other devices not needed in other tables in the same work.

### 13.4 *The variables*

The data in most tables include two sets of variables, dependent and independent. One set (traditionally the independent variables) appears in the first column, or stub), and the other appears in the column heads. In the printed work, the choice of location is sometimes limited by the physical dimensions of the table. If the same set of variables is included in two or more tables in the same work, it must appear in the same location in each.

### 13.5 *The population and the sample: "N" vs. "n"*

An italic capital *N* is used in many statistical tables to stand for the total group, or "population," from which data are drawn. An italic lowercase *n* stands for a portion of the total group, or a "sample." For example, if *N* refers to the total number of subjects (of both sexes) in a study, lowercase *n* might refer to the number of females in the study.

### 13.6 *"Percent" vs. "percentage"*

Despite changing usage, Chicago continues to regard *percent* as an adverb ("of each hundred") (or, less commonly, an adjective) and to use *percentage* as the noun form. The symbol %, however, may stand for either word.

### 13.7 *Number ranges*

Anyone preparing or editing a table must ensure that number ranges do not overlap, that there are no gaps between them, and that they are as precise as the data require. It must be clear whether "up to" or "up to and including" is meant. Dollar amounts, for example, might be given as "less than \$5, \$5–\$9, \$10–\$14, and \$15–\$19" (not "\$1–\$5, \$5–10," etc.). If greater precision is needed, they might be given as "\$1.00–\$4.99, \$5.00–\$9.99," and so forth. The symbols < and > must be used only to mean *less than* and *more than*. In a table including age ranges, >60 means "more than 60 years old" (not "60 and up," which would be typeset as  $\geq 60$ ).

### 13.10 *Form*

Every table should be given a number (arabic numerals are used) and should be cited in the text by the number, either directly or parenthetically. "The wide-ranging nature of the committee's discussions can be judged from the topics listed in table 14."

or

"Topics covered by the worker-management committee in three years of deliberations fell into five general categories (table 14)."

Note that the word *table* is lowercased in text references.

### 13.11 Tabular matter not requiring a number

A simple list or other tabular matter that requires only two columns can usually be left unnumbered and untitled.

"The prospective mountaineers were asked to bring the following equipment, in addition to toiletries:"

<i>ITEM</i>	<i>QUANTITY</i>
Boots	3 pairs
Down jackets	2
Pants	3 pairs
Sweaters	5
Underwear	5 sets
Socks	12 pairs
Bedroll	1

### 13.14 Length

The title should identify the table as briefly as possible. It should not furnish background information, repeat the column heads, or describe the results illustrated by the table. For example, a lengthy title such as "Effect of DMSO on arthritic rats and nonarthritic rats after 20, 60, and 90 days of treatment" should be pared down to "Effect of DMSO on arthritic and nonarthritic rats." From a title such as "High degree of recidivism among reform school parolees," eliminate "high degree of." Let the table give the facts; commentary can be offered in the text.

### 13.15 Syntax

The title should be a noun form. Use participles rather than relative clauses—for example, "Families subscribing to weekly news magazines" (rather than "Families that subscribe to weekly news magazines").

### 13.18 Number plus title

In a printed table, the title usually follows the number on the same line, separated by punctuation or by space and typographic distinction. Less commonly, the number appears on a line by itself, the title starting a new line. The number is always preceded by the word *table*.

"Table 2. Description of species pools used in the simulations"

*or*

"Table 6.4

Likelihood of poor weather for centennial parades (percentages)"

### 13.37 Vertical alignment: numerals

A column consisting of numerals without commas or decimal points is usually aligned on the last number, "ranged right." If the numerals have decimal points, the column is typically aligned on the decimal point. If the numerals have commas but no decimal points, alignment is made on the comma—the last comma if more than one. Where spaces rather than commas are used to separate groups of digits, alignment is made on the implicit comma. A column including different kinds of numerals is best aligned on the ones that occur most frequently. Ellipses and em dashes are centered. Where a column includes many figures in the thousands (or millions), "in thousands" (or "in millions") may be added to the column head, and the numbers in the column shortened accordingly.

### 13.42 Signs and symbols

In a column consisting exclusively of, for example, dollar amounts or percentages, the signs should be omitted from the cells and included in the column head. Mathematical operational signs preceding quantities in a column of figures are not necessarily aligned with other such signs but should appear immediately to the left of the figures they belong to.

### 13.44 Source notes: acknowledgment of data

If data for a table are not the author's own but are taken from another source or other sources, professional courtesy requires that full acknowledgment be made in an unnumbered footnote. The note is introduced by *Source* or *Sources* (often in italics and followed by a colon). "Sources: Data from Richard H. Adams Jr., "Remittances, Investment, and Rural Asset Accumulation in Pakistan," *Economic Development and Cultural Change* 47, no. 1 (1998): 155–73; David Bevan, Paul Collier, and Jan Gunning, *Peasants and Government: An Economic Analysis* (Oxford: Clarendon Press, 1989), 125–28."

If the sources are listed in the bibliography or reference list, a shortened form may be used:

"Sources: Data from Adams 1998; Bevan, Collier, and Gunning 1989."

### 13.45 Source notes: credit lines

Unless *fair use* applies, a table reproduced without change from a published work under copyright requires formal permission. Credit should be given in an unnumbered footnote, introduced by *Source*: (often in italics).

"Source: Reprinted by permission from *Scientific Style and Format: The CBE Manual for Authors, Editors, and Publishers*, 6th edition (Cambridge: Cambridge University Press, 1994), 202, table 11.3."

### 13.46 *General notes*

A note applying to the table as a whole follows any source note, is also unnumbered, and is introduced by *Note*: (often in italics).

*"Note: Since data were not available for all items on all individuals, there is some disparity in the totals."*

### 13.49 *Notes on significance levels*

If a table contains notes on significance levels (also called probability notes), asterisks may be used as reference marks. If two or three standard significance levels are noted, a single asterisk is used for the lowest level of significance, two for the next higher, and so on. If values other than these three are given, however, footnote letters are preferable to asterisks, to avoid misleading the reader. In the note, the letter *p* (probability) is usually lowercase and in italic. Zeros are generally omitted before the decimal point. Probability notes follow all other notes. (Since *p* stands for the probability of a Type I error, that of rejecting the null hypothesis when in fact it is true, the level of significance increases as the probability of this error decreases.)

\**p* < .05

\*\**p* < .01

\*\*\**p* < .001