How to Make Working & Publication-style tables in Stata

Gary Klass, “How to Construct Bad Charts and Graphs”
http://lilt.ilstu.edu/gmklass/pos138/datadisplay/badchart.htm

Gary Klass, “Presenting Data: Tabular and graphic display of social indicators”
http://lilt.ilstu.edu/gmklass/pos138/datadisplay/

Jon Kolko, “Edward Tufte” (on principles of data presentation)
http://facultypages.scad.edu/~jkolko/IACT370_InformationArchitecture/IACT370_05_TuftePrinciples.pdf

Here are examples of the table-making commands ‘tabout,’ ‘mkcorr,’ ‘estout,’ ‘outreg,’ ‘outreg2,’ and ‘xml_tab’. Get these by using ‘findit,’ ‘help,’ Stata listserv, ‘UCLA Resources for Learning Stata’, or ssc whatsnew and ssc install.

- ‘tabout’ is excellent for making publication-style summary and contingency tables.

- ‘mkcorr’ is helpful for making publication-style correlation matrices.

- ‘estout’ is powerful & flexible for making publication-style regression tables (see ‘estimates store’ and ‘estimates table’ below), which can be copied from Stata results-window to Word or Excel for editing. In addition, it can create LaTeX output.

- ‘estimates store’ followed by ‘estimates table’ is excellent for making working tables to compare regression models. Following them up, in turn, with ‘estout’ perhaps allows the greatest range of table options for publication-style regression and other tables.

- ‘xml_tab’, specified after ‘estimates store,’ is easy and excellent for creating publication-style regression tables in Excel. While it may be less flexible than most of the other table-making commands covered in this document, xml_tab may require the least amount of editing.

- ‘outreg2’ is also easy and excellent for make publication-style regression tables, and it has the advantage of directly creating such tables in Word or Excel (which then require editing). Like estout, it can create LaTeX output.

Note: Copying tables directly from the Stata Results-window to Word works reasonably well if Stata and Word are formatted to display Courier 11 point or less (which is a fixed-space font).
• **Publication-style summary & contingency tables (tabout):**

`findit tabout`  
`help tabout`

`use hsb2, clear`  
`la var female "Gender"`  
`la var prog "Program"`

`tabout female prog using gender_prog, cell(freq col) format(0 1) stats(chi2)`  
`layout(row) show(all) replace`

*Edit file ‘gender_prog’ (located in current folder) in Word or Excel.*

*Some options:*
- `cell`: freq, cell, row, col, cum [all can be specified]
- `format`: # decimal points; specify in order of types of cell entries specified
- `clab`: table’s title
- `layout`: how the columns are laid out
- `stats`: types of statistics specified

• **Publication-style correlation matrix (mkcorr):**

`findit mkcorr`  
`help mkcorr`

`mkcorr read write math science socst, log(hsb2_corr) sig means cdec(2) mdec(2)`  
`casewise`

*Note: The default is pairwise (pwcorr). You must specify ‘casewise’ if you want the equivalent results of ‘corr’.*

*Edit file ‘hsb2_corr’ (located in current folder) in Word.*

*Options:*
- `sig`: display significance test p-values
- `cdec( )`: # decimal points displayed in correlation coefficients
- `means`: display descriptive numerical summary
- `mdec( )`: # decimal points displayed in descriptive numerical summary
- `casewise`: display casewise results (instead of default pairwise results)

• **Heads up: Using ‘estout’ in profile.do to make superior working regression tables**

This document will address the ‘estout’ command later on (after ‘estimates store’ and ‘estimates table’), but this hint is worth addressing now. You can easily make easy-to-read, nearly full-feature regression tables if you include the following commands in profile.do (specifying any function keys that you wish):

* estout OLS regression table  
  `global F1 = ", cells(b(star fmt(%9.3f)) se(par)) starlevels(+ .10 * .05 ** .01 *** .001) stats(N p r2 r2_a bic, star(p)) legend style(fixed) "`
* estout logistic regression table
global F2 = ", cells(b(star fmt(%9.3f)) se(par)) starlevels(+ .10 * .05 ** .01 *** .001) stats(N chi2 ll bic, star(chi2)) legend style(fixed) "

* estout multinomial logistic regression table
global F3 = ", eform unstack cells(b(star fmt(%9.3f)) se(par)) starlevels(+ .10 * .05 ** .01 *** .001) stats(N chi2 ll bic, star(chi2)) legend style(fixed) "

For example, here’s out to use the first of these commands to make OLS regression working tables:

use hsb2, clear
regress science read write
estimate store m1

estout m1 [press F1, or other corresponding, function key to produce the table]

regress science read write math female
est store m2

estout m1 m2 [press F1, or other corresponding, function key to produce the table]
• Working OLS regression table (estimates store & estimates table)

help estimates

la var science “Science”
la var math “Math”
la var read “Read”
la var write “Write”

reg science math read write
est store m1

reg science math read write female white
est store m2
est table m1 m2, star         [To change defaults: star(.01 .05 .10)]

• Publication-style OLS regression table (estimates store & estimates table)—but does not display both significance stars & standard errors

la var socst “Social Studies”
la var female “Female”
la var white “White”

reg science math read write socst
est store m1

reg science math read write socst female white
est store m2
est table m1 m2, b(%9.2f) stats(N r2_a) drop(socst)
[See options such as label, drop, keep]

Note: using estimates store-estimates table is quick and easy for publication—if you are able to use stars to signify significance and if you don’t have to display standard errors. The procedure will not display both stars and standard errors.

• Publication-style multinomial logit regression table (estimates store & estimates table)—but does not display both significance stars & standard errors

la var ses “SES”
tab ses
tab ses, nolabel

mlogit ses math read, rrr nolog base(1)
est store m1

mlogit ses math read science socst, rrr nolog base(1)
est store m2
ximlogit ses math read science socst i.prog, rrr nolog base(1)
est store m3

est table m1 m2 m3, eform star(.01 .05 .10) stats(N ll df_m chi2) b(%9.2f)
   drop(_Iprog_2 _Iprog_3)

*Note: If using 'keep,' include _const or else it won’t be displayed. It’s often helpful
   to use the 'label' option. Remember: this procedure won’t display both stars and
   standard errors.

- **Publication-style OLS regression table (estout):**

reg science math read female
est store m1
la var math "Math"
la var read "Read"
la var female "Female"

estout m1, cells(b(star fmt(%9.2f)) se(par)) starlevels(+ .10 * .05 ** .01) stats(N p
   r2 r2_a bic, star(p) label("N" "Prob>F" "R2" "Adj. R2" "BIC") fmt(%9.0g %9.2f))
   mlabel(OLS Model) collabels(Model) nolabel abbrev varlabels(_cons Constant)
   varwidth(25) modelwidth(10) prefoot("") postfoot("") legend style(fixed)

*Note: See 'help estout' regarding optional specification of ‘replace’ or ‘append’. ‘p’
gives F-statistic model significance. See drop, keep, and order options, among
others.

*Note: To display two or more models in the same table – after each model: est store
   m1   est store m2 ; then to display: est m1 m2, cells... [do not type ‘replace’ or
   ‘append’]. Here’s an example, which builds upon our having done ‘estout m1’
above:

reg science math read write female ses
est store m2
la var write "Write"
la var ses "SES"

estout m1 m2, cells(b(star fmt (%9.3f)) se(par)) starlevels(+ .10 * .05 ** .01)
   stats(N p r2 r2_a bic, star(p) label("N" "Prob>F" "R2" "Adj. R2" "BIC") fmt(%9.0g
   %9.3f)) mlabel(OLS Model) collabels(Model) nolabel abbrev varlabels(_cons
   Constant) varwidth(25) modelwidth(10) prefoot("") postfoot("") legend style(fixed)

*Horizontal format:
reg science math prog
est store mh
estout m1, cells("b se t p") starlevels(+ .10 * .05 ** .01) stats(N p r2 r2_a bic,
   star(p) label("N" "Prob>F" "R2" "Adj. R2" "BIC") fmt(%9.0g %9.3f)) mlabel(OLS
   Model) collabels("coefs" "se" "t" "p") nolabel abbrev varlabels(_cons Constant)
   varwidth(25) modelwidth(10) prefoot("") postfoot("") legend style(fixed)

To display standardized (beta) coefficients (in place of standard errors): cells(beta)
To display confidence intervals:
estout m1 m2, cells(ci(star fmt(%9.2f)) se(par)) starlevels(+ .10 * .05 ** .01) 
stats(N p r2 r2_a, star(p) label("N" "R2" "Adj. R2") fmt(%9.0g %9.2f)) mlabel(OLS
95% CIs) collabels(95% CIs) nolabel abbrev varlabels(_cons Constant) varwidth(25) 
modelwidth(10) prefoot("") postfoot("") legend style(fixed)

- Publication-style logistic regression table (estout):

`xi3:logistic white math science i.prog, nolog` 
estore m1 
`la var white "White"` 
`la var science "Science"`

`estout m1, eform cells(b(star fmt(%9.2f)) se(par)) starlevels(+ .10 * .05 ** .01) 
stats(N chi2 ll bic, star(chi2) label("N" "chi2" "LL" "BIC") fmt(%9.0g %9.2f)) 
mlabel(Logistic Model) collabels(Model) nolabel abbrev varlabels(_cons Constant) 
varwidth(25) modelwidth(10) prefoot("") postfoot("") legend style(fixed)`

`Note:` eform to display odds ratios; star(chi2), instead of ‘p’, for model significance. 
See drop, keep, and order options, among others.

`Note:` See OLS example on displaying two or more models in the same table.

To display confidence intervals:
estout m1, eform cells(ci(star fmt(%9.2f)) se(par)) starlevels(+ 0.10 * 0.05 ** 0.01) 
stats(N chi2 ll bic, star(chi2) label("N" "chi2" "LL" "BIC") fmt(%9.0g %9.2f)) 
mlabel(Logistic Model) collabels(95% CIs) nolabel abbrev varlabels(_cons Constant) 
varwidth(25) modelwidth(10) prefoot("") postfoot("") legend style(fixed)

- Publication-style multinomial logit regression table (estout):

`tab prog` 
`tab prog, nolabel` 
mlogit prog math read science female, base(2) rrr nolog 
estore m1 

`estout m1, eform cells(b(star fmt(%9.2f)) se(par)) starlevels(+ .10 * .05 ** .01) 
stats(N chi2 ll bic, star(chi2) label("N" "chi2" "LL" "BIC") fmt(%9.0g %9.2f)) 
mlabel(Multinomial Logit Model) collabels(Model) nolabel abbrev varlabels(_cons Constant) 
varwidth(25) modelwidth(10) prefoot("") postfoot("") legend style(fixed) 
unstack`

`Note:` eform and star(chi2), as well as ‘unstack’ to display sub-equations in separate 
columns. See drop, keep, and order options, among others.

`Note:` See OLS example on displaying two or more models in the same table.

`Note:` As discussed above, you may want to specify a shortcut for the subcommands 
in C:/data/profile.do.
• **Add statistics such as standardized coefficients to a regression table (estout, estadd):**

```
reg science math read, beta
est store m1
estadd beta: m1
estout m1, cells(b beta(par)) stats(N chi2, star(chi2) label("N" "chi2") fmt(%9.0g %9.2f)) style(fixed)
```

• **How to create a table with the sign of estimated coefficients and have a double sign if the coefficient is significant (estout; from Stata listserv, March 2007)**

```
findit esto
help esto

clear
sysuse auto   [Stata pre-installed ‘automobile’ data]
reg price turn mpg if foreign==0  
esto   [est1 stored]
reg price turn mpg if foreign==1

esto   [est2 stored]
estout, cells(_sigsign) starlevels("" 1 * .05) style(fixed)
```

*Note:* The symbols specified in the starlevels() option do not really matter here, but they must not be omitted so that all coefficients get at least one sign.

Here’s the table produced:

<table>
<thead>
<tr>
<th></th>
<th>est1 _sigsign</th>
<th>est2 _sigsign</th>
</tr>
</thead>
<tbody>
<tr>
<td>turn</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>mpg</td>
<td>--</td>
<td>-</td>
</tr>
<tr>
<td>_cons</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

• **Publication-style OLS regression tables (xml_tab)**

```
findit xml_tab
help xml_tab

u hsb2, clear
la var science “Science”
la var read “Read”
la var write “Write”
la var math “Math”
la var socst “Social Studies”

reg science read write
est store m1
```
reg science read write math socst
est store m2

xml_tab m1, title("OLS Regression: Science") right label stats(N, r2, r2_a)
save("c:\data\ucla/science_out.xml") replace

Note: type 'ereturn list' to see stats options and codes for the model.
Note: see drop, keep, and rblank options, among others.

xml_tab m1 m2, title("OLS Regression: Science") below label stats(N, r2, r2_a)
stars(.001 .01 .05) font("Times New Roman” 10)
save("c:\data\ucla/science_out.xml") replace

Click on link to open Excel table.

xi3: logistic female read write
est store m1
xi3: logistic female read write math science
est store m2
xml_tab m1 m2, title("Logistic Regression: Gender") label stats(N, ll, aic, bic)
replace save("c:\data\ucla/female_out.xml") replace

Click on link to open Excel table.

xml_tab m1 m2, right stats(N, r2_a) title("Multinomial Logit Model") replace ['right,' i.e. standard errors to the right of the coefficients, is the default.
Click on link to open Excel table.

xml_tab m1 m2, below stars(.001 .01 .05) stats(N, r2_a) title("Multinomial Logit Models") replace [displays standard errors below the coefficients]

Click on link to open Excel table.

• Publication-style OLS regression table (outreg2):

reg science math read female
outreg2 using m1, replace  [Click ‘seeout’ in STATA results window.]
outreg2 using m1s, word replace  [Click ‘ols_science.rtf’.]
outreg2 using m1, excel replace  [Click ‘ols_science.xml’.
outreg2 using m1, alpha(0.001, 0.01, 0.05) symbol(***, **, *) adjr bdec(2)
title(Science) addnote(Do not try this at home.) onecol long replace
outreg2 using m1, ci replace
outreg2 using m1, beta replace  [betas in parentheses, in place of se’s]
outreg2 read female, replace [Displays only the specified explanatory variables.]

Click on ‘seeout’ to see STATA display & ‘ols.RTF’ to see Word display.
Table>Format>Hide gridlines

reg science math read
outreg2 using m1, onecol long replace
reg science math read female prog
outreg2 using m1, onecol long append  [to display results of both models]
Click on ‘seeout’ to see STATA display & ‘ols.RTF’ to see Word display.
Table>Format>Hide gridlines
Some options (after ','):
word: output text to a Word file, which is saved in the current directory
excel: output text to an Excel file
onecol: suppresses multi-column format if relevant
long: accompanies onecol, forcing Word of Excel to adopt one-column format
replace: replaces previous model or format
append: appends additional models (instead of ‘replace’)

ci: specifies confidence intervals instead of coefs
beta: specifies standardized coefs
bdec(): specifies # decimal places for regression coefs (default=3)
tdec(): if t-values are displayed, specifies # decimal places (default=3)
rdec(): if r2 or adj r2 is displayed, specifies # decimal places (default=3)
add(): if additional stats are displayed (see ‘e’ and addstat’)
bfmt(): specifies format for regress coefs – f, fixed; g, general

symbol(): specifies significance values; default is symbol(***, **, *)
alpha(): specifies significance levels; default is alpha(0.000, 0.01, 0.05)
adjr2: specifies adjusted r2
nor2: specifies not to display r2
e(): specify added stats, e.g., e(ll df m chi2 aic); or e(all); does not display publication-style stat names (see ‘addstat’). After estimating a model, type ‘ereturn list’ to obtain options and codes.

addstat(): to add other stats along with stat names. E.g., addstat(Log likelihood, e(ll), DF, e(df m), chi2, e(chi2), AIC, e(aic)). After estimating a model, type ‘ereturn list’ to obtain options and codes.

title(): specifies title
ctitle(): specifies column title
addnote(): to add notes

Note: title and note options give the same output whether or not the text is enclosed within quotation marks. outreg2 gives an error message if addnote is formatted as follows: addnote(Source: my data). So don’t include a colon in the text.

E.g.:
reg science math read
outreg2 using m1, bdec(2) bfmt(f) alpha(0.01, 0.05, 0.10) adjr2 addstat(AIC, e(aic))
title(OLS) addnote(Study data) word replace

Click on ‘seeout’ to see STATA display & ‘m1.RTF’ to see Word display.
Table>Format>Hide gridlines

Note: To display more than three levels of significance - alpha(0.001 0.01, 0.05, 0.10) symbol(***, **, *, +)

Note: To display two or more models in the same table – outreg2 using m1, ...... append
**Publication-style logistic regression table (outreg2)**

See the document 'Using Stata’s outreg2 with MS-Word to make publication-style regression tables'.

logistic female math read white
outreg2 using m1, eform nor2 replace
outreg2 using m1, eform nor2 e(ll df_m chi2 aic) word replace
outreg2 using m1, eform nor2 addstat(Log likelihood, e(ll), DF, e(df_m), Chi2, e(chi2)) word replace

[The option ‘eform’ specifies odds-ratios (i.e. exponentiated form) for display.]

See outreg2 options above.

**Publication-style multinomial logit regression table (outreg2)**

mlogit ses math read, base(1) nolog
outreg2 using m1, eform nor2 addstat(ll, e(ll), df_m, e(df_m), chi2, e(chi2)) word replace

mlogit ses math read write science, base(1) nolog
outreg2 using m1, eform nor2 e(ll df_m chi2) append

See outreg2 options above.