Description of -tw3xls-

Hello.

Here is a new program for Stata – **tw3xls**. It's main purpose – export Stata three-way frequency tables into formatted Excel tables. It uses -putexcel- for formatting and saving into Excel. The minimal Stata version of executing -tw3xls- is 14.1.

It is designed to ease (or totally vanish) the pain of saving and formatting three-way frequency tables in Excel copied from Stata.

There are several options of the program:

- "show" – shows an output of built-in Stata -table- command. Useful when a user wants to see the output tables.
- "missing(value)" – replaces cells with zero frequencies to any other real value, like zero or -999 (blank by default).
- the "by(*varname*)" option produces a set of three-way tables for each unique value of the *varname,* allowing to produce a relative to a four-way table. Each output table has a header, titles for rows, columns and supercolumns. The number of output tables should be equal to the number of unique values of the *varname*.
- "stub(string)" – saves output table(s) into matrices named *string1*, *string2*, etc. There will be one table by default, and more tables if the *by()* option is specified. The output tables also available as matrices stored in r-class.
- "sheet(string)" – gives a name to an Excel worksheet where data are exported. The default name is "Data"
- "replace" or "modify" option replaces or modifies the output Excel file. The latter is useful when saving data into another worksheet using "*sheet(string)*" option. Default is "replace".
- "format" option sets borders across all cells in the Excel table, centers vertically contents of all cells, and merges table header cell with other empty cells.
- "mergecells" merges supercolumn cells and centers their value horizontally (super slow!). By default, each supercolumn header has its own value.

The -*tw3xls*- program handles both string and numeric data (including the "by()" option). If variables are encoded, their labels will be shown as numbers in the output Excel file. In order to get meaningful string labels, those variables should be converted to strings using -decode- command and used in the -tw3xls-.

Performance note

Adding options to the -tw3xls- can greatly improve an overall look of the output table but at the cost of performance. Below in table 1 you can find several performance tests using an "auto" dataset. To get a heavier load, the *by()* option is specified.
Each command was run ten times, so the table contains data of the average execution time. The encoded variable "foreign" was put into string using "decode" command.

A full list of commands is present below:

```
sysuse auto, clear
decode foreign, gen(foreign_str)
tw3xls mpg rep78 foreign_str using example
```

The command produces one three-way table with mileage (mpg) labeled at rows, repair record (rep78) at supercolumns, and car type (foreign_str) at columns. It takes less than 1 sec to execute in Stata 14.2 IC.
Then we add a by-variable, for example, *headroom*:

```
tw3xls mpg rep78 foreign_str using example, by(headroom)
```

The command produces eight three-way tables for each headroom value with the same rows, columns and supercolumns as in the previous table. Statistics capturing all possible runs is in the table 1. Each row in the table shows the program execution time with all previous and current options specified.

Table 1. Performance tests

No	Cumulatively added options	Mean run time, sec	Cumulative increase, sec	Cumulative increase, ratio	Increase from the base**, sec	Increase from the base**, ratio
1	*none*	0.55	0	-	0	-
2	by(headroom)	4.61	4.06	8.38	4.06	8.38
3	show	4.69	0.08	1.02	4.14	8.53
4	missing(0)	4.77	0.08	1.02	4.22	8.67
5	stub(mat)	4.85	0.08	1.02	4.3	8.82
6	format	36.46	31.61	7.52	35.91	66.29
7	mergecells	57.51	21.05	1.58	56.96	104.56

* mean time is calculated using 10 runs of the program with specified options
** The base run with no options is in the row 1. It presents the command: `tw3xls mpg rep78 foreign_str using example`

In general, the most time-consuming options are “format” and “mergecells”. In cost of execution time they automatically fix most of the formatting issues, making tables user-friendly and publication-ready. The greatest portion of that time is how Stata interacts with a spreadsheet file: opening, writing the data , calling procedures to format and merge cells, and then closing the workbook.

Working further on that part (speeding up formatting and merging) may greatly improve the performance keeping the same functionality.

An alternative to -tw3xls- is a user-written program -tab3way- in conjunction with -logout- (both available on ssc). However, it requires appropriate skills and additional time for post-formatting the output tables in Excel.
For one- and two-way tables use -tabout- or -tab2xl- and -tab2xl2-. The two former are available on ssc, the latter can be found on Stata blog: http://blog.stata.com/2013/09/25/export-tables-to-excel/ .

Andrey Ampilogov
Freelancer
a.ampilogov@gmail.com