POLMATH QUESTIONS ANSWERED

Original Question

From: Paul Gronke <gronkep@reed.edu>

Quick question for the list: Lisa Bryant (CU Fresno) and I are preparing some “top lines” and “tabs” for a client with whom we conducted a survey.

If you have seen these before, they are usually organized so that categorial survey responses are reported on the rows, and the columns report the overall responses, then responses “tabbed” or “crossed” by various demographic and political categories. Roll your eyes if you will that this is just a big set of exploratory cross tabs, but a lot of folks expect to see them to help digest the survey results.

A typical “tab” looks like this:

VARIABLE Total GOP IND DEM

MEN WOMEN …

Category 1 N % N % N % N % N

% N %

Category 2 N % …

…

We’ve found a package in Stata, tabmult, that allows us to create “multiple tables” (hence “tabmult”) where the tables are accumulated as columns. It outputs XML format (bad) and only allows us to report percentages -or- case count, not both (although the documentation claims otherwise).

We haven’t found a way to get this done via other Stata procedures (eststore and esttab are pretty flexible but one thing they don’t do is produce multiple contingency tables along a horizontal, “long” dimension).

I’m wrestling with the puzzle in R, using the tables package, which outputs LaTeX and HTML (good) but is also relatively inflexible (and I’m fairly new to R). Seems like R should do this fairly easily but accumulating the results and kicking them out in a landscape print format is proving challenging.

We can produce this pretty easily using Tableau, but then we don’t have reproducible and flexible code.

Does anyone know of a way to do this, or has anyone seen a script for R or Stata that does this? I’m surprised; I talked to some local survey firms and they generally pay someone a fairly hefty fee to do this each survey, rather than paying just once for a flexible script.

Thanks!

--- Paul Gronke

TOP SOLUTION

Stata Special

From: Jonathan Mendelson

> Hi Paul,
>
> I posted a response to the list, but it hasn't gone through yet, so I thought I'd reply directly. I encountered the same issue as you and wrote a Stata package that essentially creates "tabs" in spreadsheet form. You can install it in Stata via "ssc install tabsheet" or view information at https://ideas.repec.org/c/boc/bocode/s458128.html; there are examples in the documentation so you should be able to get started fairly quickly.
>
> The program is not particularly flexible, but it is easy to use, and some colleagues at my survey firm have found it very useful. Although it doesn't currently output to anything other than tab-delimited file (which can be opened in a spreadsheet), with some clever formatting in Excel, you could print the resulting spreadsheet to PDF for something nicer looking.
>
> If you need something more flexible in Stata, I'd recommend tabout, although that may require more work to set up. If you find out about any R packages that do something similar, I'd be interested in hearing about it.
>
> Best,
> Jonathan

Other Solutions

R

R Dyplyr Xtable

From: Arndt Leininger

> Dear Paul,
>
> if I may, I would suggest the following solution. Prepare a data.frame which represents the table you need using the package "dplyr" and then save that data.frame to a table (as HTML oder LaTeX) using the "xtable" package.
> For automating report generation I also recommend looking into "ReporteR" which allows you to create Word or PowerPoint documents with R. You can even create figures.
>
> This is not a quick and easy solution but once the code is in place you can easily adapt it to future projects.
>
> Best,
> Arndt

From: Dag Tanneberg

Dear Paul,

> Judging from my experience with R, you will have to build everything up from repeated calls to whatever summary function you use and then rbind() (for each category) respectively cbind() (for each

column) the results into a single, big matrix. In my workflow, the call to a formatting function like xtable::xtable() is only the last step in the process. In other words, I would suggest that you have to generate each crosstab first, then bind them together into a single matrix object, and only afterwards translate the entire result into LaTeX or HTML.

If every column title was unique, then you could simply assign row and column names to the matrix. However, in your example table it looks like labels span multiple columns. Unfortunately, I cannot help you with that. Usually, I edit those cases manually, and that is apparently no option for you.

I'd be very interested to learn what other list members suggest.

HTH

Dag

R Dyplyr Stargazer

From: NaLette Brodnax

Hi Paul,

The "plyr" package in R has a count() function that you could use to create a frequency table for each variable, which you could then use to calculate the percentages.

For output, I recommend the "stargazer" package. Store your frequency table as a matrix and use the stargazer() function with the matrix as the object and the summary option set to FALSE. Stargazer can output to latex, text, or html. Jake Russ has a helpful cheat sheet for working with the various stargazer options.

Best regards,
NaLette

NaLette M. Brodnax
Joint PhD Candidate, Public Policy & Political Science
Indiana University Bloomington
nbrodnax@indiana.edu
www.nalettebrodnax.com

R Dyplr Special Script

From: David Armstrong

Hi Paul,

I thought I'd reply off list in the hopes that you might get it faster. Attached are a knitr file and the tex file that gets produced from knitting it along with the pdf output. Hopefully this will get you on your way.

Best,
Dave.

Dave Armstrong
Associate Professor of Political Science
Western University

CLICK HERE FOR FILES

Stata

Stata Special

From: Jonathan Mendelson

Hi Paul,

I posted a response to the list, but it hasn't gone through yet, so I thought I'd reply directly. I encountered the same issue as you and wrote a Stata package that essentially creates "tabs" in spreadsheet form. You can install it in Stata via "ssc install tabsheet" or view information at https://ideas.repec.org/c/boc/bocode/s458128.html; there are examples in the documentation so you should be able to get started fairly quickly.

The program is not particularly flexible, but it is easy to use, and some colleagues at my survey firm have found it very useful. Although it doesn't currently output to anything other than tab-delimited file (which can be opened in a spreadsheet), with some clever formatting in Excel, you could print the resulting spreadsheet to PDF for something nicer looking.

If you need something more flexible in Stata, I'd recommend tabout, although that may require more work to set up. If you find out about any R packages that do something similar, I'd be interested in hearing about it.

Best,
Jonathan

Stata Tabout

From: Thomas J. Leeper"

Paul,

Stata's _tabout_ add-on module should be able to do this. I've not used it for this exact purpose but the documentation includes examples of these kinds of summaries.

For R, I think you will need to do a fair amount of programming to get this working well. The simplest approach is probably writing a function that processes a given variable to a row vector of cells, then just rbind() the results together into a matrix or data frame; the. You can output using knitr::kable() or similar. Unfortunately, the best table production packages, for example stargazer, are very modelling focused.

Best,
Thomas

Stata Tabout Special

From: Eric A. Booth

Hi Paul -

I've never been able to get -tabmult- (from SSC) in Stata to produce percents (it implies it can in the help file but the adofile doesnt seem to have those options hardcoded), and IMO it's a poorly written program (lots of unexpected bugs =-o nand OS-dependent kludges in the adofile).

I do this type of survey output work quite frequently and essentially I use a custom-built wrapper to -tabout- to produce some tables we give to clients that are similar to what you describe. The limitation to -tabout- in providing a table as you have specified is that -tabout- doesnt operate across columns / horizontally - you will need to post-process -tabout- output to get what you want.
There is a new version of tabout that is coming out soon (Ian Watson announced the beta in Dec 2016) that may make the solution I provide below obsolete but currently I essentially produce the tabout tables, then reimport them from xls, manipulate them (to stack the panels side-by-side using -merge-) , and then re-export them to excel. The wrapper adofile I have is specific to my needs but the example below is a generalized version of the core loop I've used to do this task.

There are other non-tabout solutions including -contract- and then manipulating the results (see example 2 below) or even running separate -tabulate- commands, storing the results of each in matricies, and then using -putexcel- to export results (I tend to avoid this approach but there is a great overview here:
http://blog.stata.com/2013/09/25/export-tables-to-excel/) .

I'm assuming some things about your data structure based on the example table you provided, but if the fake data below dont match your data you'll need to either change the data structure to match mine or change the code to match your data structure. Also, I added a second item (each item has 5 categories) so that you can see how to extend this code to many items. You'll notice that table3 (from solution 1) and table 4 (from solution 2) match (if you run the do-file all together using the same fake, simulated data from the first section) and so those approaches are equivalent.

HTH,
Eric

```
************RUN ENTIRE FILE TOGETHER AS .DO FILE
//set up//

**create fake survey data
clear
set obs 1000
g item1 = int(1+runiform()*5)
g item2 = int(1+runiform()*5)
g partyid = int(runiform()*3)
g sex = rbinomial(1,.52)
lab def partyid 0 "DEM" 1 "IND" 2 "GOP", modify
lab def sex 0 "Male" 1 "Female", modify
forval ii = 1/5{
loc lv `"`lv' `ii' `"Category `ii'"'"'
}
lab def lv `lv', modify
lab val item* lv
lab val partyid partyid
lab val sex sex
inspect *
sa "master.dta", replace

**get -tabout- and -tabmult - from SSC
foreach x in tabout tabmult {
cap which `x'
if _rc ssc install `x', replace
}




//Here are 2 potential solutions//


**1. Use tabout to make table
u master, clear
tabout item1 item2 partyid using "table1.xls", replace ///
c(freq col) f(0c 1p)
tabout item1 item2 sex using "table2.xls", replace ///
c(freq col) f(0c 1p)
**the problem is that they dont horizontally combine, so use -merge-::
*first grab some values for later use*
levelsof partyid, loc(aa)
loc a : word count `aa' //number of levels in partyid
levelsof sex, loc(bb)
loc b: word count `bb'
di `"`a' & `b' used below..."'

*insheet, save and then merge files
forval n = 1/2 {
insheet using table`n'.xls, clear nonames
rename * =`n'
g match = _n
tempfile a`n'
sa `a`n'', replace
}
u `a1', clear
merge 1:1 match using `a2', nogen
drop match
**format table using stored macros with # of values/levels
drop if v11 == "Total" //unless you want these panel ttls
*values calculated off of # of categories above (loc a b)
drop v12  //always drop first col of panels after 1st (v12, v13...v1n)
drop v`=`b'*2+2'2 v`=`b'*2+3'2  //drops last 2 cols of panel after 1st (can
continue with v`=`c'*2+2'2 v`=`c'*2+3'2 ... to whatever letter you end with)
order v11 v`=`a'*2+2'1 v`=`a'*2+3'1 //moves last 2 cols of first panel to
front based on your posting example & v11 is always first
**export to excel:
export excel using table3.xlsx, replace

**2. Use contract & reshape to make table
*create table4 blank template for the loop below
clear
g item = .
g cat =.
```

```
sa table4.dta, replace emptyok
u master, clear
  forval n = 1/2 {  //item/category loop
foreach j in sex partyid  {  //characteristic loop
u master, clear
levelsof `j', loc(jj)
foreach x in `jj' {
u master, clear
  contract item`n' `j' if `j' == `x', freq(N) p(Pct)
 g Pct2 = string(Pct)+"%"
decode `j', g(`j'2)
drop Pct `j'
  reshape wide N Pct2, i(item`n') j(`j'2)  string
 sa item`n'_`j', replace
**merge within j loop
rename item`n' cat
g item = `n'
merge 1:1 cat item using table4.dta , nogen
sa table4.dta, replace
} //end x
} //end j
} //end n
**clean up
u table4, clear
sort item cat
egen Total = rowtotal(NDEM NIND NGOP)  //or base it off sex or something else
order cat item Total *DEM *IND *GOP *Male *Female
sa table4, replace


**other options include:
*3. use -tabulate [item1 partyid...], matcell() -  and -putexcel-  to build tables
*4. use -logout- to export results of -tabulate- or -tabdisp- or -tab3way-
```