Subject: Merging data files differend years same IR
Posted by Davison on Thu, 02 Mar 2017 18:47:22 GMT

View Forum Message <> Reply to Message

Hello together,

first I want to say thank you for this great forum. It helps me a lot. I have read in this forums many questions about merging data files from

the same year but different questionaires.

I analyze the DHS data of Lesotho 2004 to 2014. I use SPSS. I want to merge the different Datasets, only the IR-Record together in one file to

make a chi square of significant differences. Can anyone help me? I must note somethings else because of weighting and my planfiles?

Thank you very much.

Many greeting from Germany!

Subject: Re: Merging data files differend years same IR
Posted by Bridgette-DHS on Fri, 03 Mar 2017 13:31:30 GMT

View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

In this situation you do not want to merge the files. You cannot, in fact, merge the files for different surveys, because you have different cases. Instead, you want to append the files. That is, you make one long file in which the records in one survey appear after the records for another survey. To do a test of changes or differences you must be sure to be consistent in the variable names and you must have a code that distinguishes one survey from another. I do not use SPSS, but I can provide an example of how to do this in Stata. The following lines include sub-programs called setup1, setup2, and analyze. The execution of the program begins after the multiple lines of asterisks. It is set up for two surveys but can include any number of surveys, with "use" and "setup1" lines inserted for each survey. The paths would have to be changed. The "analyze" routine could be modified to test differences between survey 1 and survey 2, survey 1 and survey 3 (if there is a 3rd survey), etc. You can add other covariates to the logit models, do chi-square tests, etc., within the analyze routine.

```
set logtype text
log using e:\DHS\programs\tests\diffs_between_surveys_log_22July2016.txt, replace

* Tom Pullum, tom.pullum@icfi.com, July 25, 2016

set more off
cd e:\DHS\DHS_data\KR_files
```

```
*****************************************************************

program define setup1

* Construct the indicator, number the surveys, save the needed variables

scalar ssurvey=ssurvey+1
local lsurvey=ssurvey
gen survey=ssurvey

* CONSTRUCT THE INDICATOR

* values other than 0 and 1 should be interpreted as .
replace g100=. if g100>1
replace g102=. if g102>1

gen y = .
replace y=0 if g100<.
replace y=1 if g102==1

keep v005 v021 v023 y survey
save temp_`lsurvey'.dta, replace

end

*****************************************************************

program define setup2

* Combine the surveys into one file

use temp_1.dta, clear
append using temp_2.dta

egen cluster=group(v021 survey)
egen stratum=group(v023 survey)

save temp.dta, replace

end

*****************************************************************

program define analyze

* Test whether the "survey" variable is statistically significant
```

```
svyset cluster [pweight=v005], strata(stratum) singleunit(scaled)

tab survey y
tab survey y [iweight=v005/1000000], row

* Test for significance of change or difference
svy: logit y i.survey
scalar p=e(p)
scalar list p

* p is the significance of a test of H0: in the population, there was no difference
*  in the prevalence of the outcome across the surveys

end

****************************************************************
****************************************************************
****************************************************************
****************************************************************
****************************************************************

* EXECUTION BEGINS HERE

* Example:  difference between two surveys in FGM prevalence

* Kenya 27.1% in 2008-09 vs 21.0% in 2014

scalar ssurvey=0

use e:\DHS\DHS_data\IR_files\KEIR52FL.dta, clear
setup1

use e:\DHS\DHS_data\IR_files\KEIR70FL.dta, clear
setup1

setup2
analyze
```

Subject: Re: Merging data files differend years same IR
Posted by bakerchowdhury on Mon, 06 Mar 2017 17:15:52 GMT
View Forum Message <> Reply to Message

I am working on a similar project (same IR files multiple years). I did manage to append data for Bangladesh 1999 to 2014 starting with the 2014 data and had year as an indicator variable for each survey using Stata.
I plan to do multiple linear regression, binary logistic, and multinomial logistic regression for three

different types outcomes.

Since I will be using "svy" command, I need to consider PSU (V021) and Strata (V022). However, I noticed the values (# of unique values and range) for V021 and V023 are not equal for each year.
I was wondering if that might be a problem for my analysis.

Thank you so much for your help
Baker

Subject: Re: Merging data files differend years same IR
Posted by Bridgette-DHS on Mon, 06 Mar 2017 22:32:54 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

You have appended the files and constructed a variable that identifies each survey. I will assume that variable is called "survey" and it takes the values 1999, ..., 2014, for example. You then need to construct new PSUs and strata. I suggest something like this in Stata:

egen psuid=group(survey v021) and
egen stratumid=group(survey v022) .

After that you do svyset and svy.

Subject: Re: Merging data files differend years same IR
Posted by bakerchowdhury on Tue, 07 Mar 2017 19:26:16 GMT
View Forum Message <> Reply to Message

Thank you so much for the response. Yes, my survey indicator variable is 'year' which takes values 1999, 2004,...2014.

Here are the steps I have done in my code

* create sample weight variable
gen wgt=v005/1000000
* generate new psuid and stratumid variable for svy command
egen psuid=group(year v021)
egen stratumid=group(year v022)
* set svy command
svyset[pw=wgt],psu(psuid) strata(stratumid)
*Example regression
svy: reg y x v012 i.v190

Could you please have a look if this looks okay.

Best regards,
Baker

Subject: Re: Merging data files differend years same IR
Posted by Bridgette-DHS on Wed, 08 Mar 2017 14:14:12 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:
Quote:
I recommend the following modification:

svyset psuid [pw=v005], strata(stratumid) singleunit(centered)

The psu is specified without "psu()", you do not need to divide v005 by 1000000, the comma goes after the specification of the psu and pweight, and you are likely to have a crash if you do not include singleunit.

Subject: Re: Merging data files differend years same IR
Posted by bakerchowdhury on Mon, 13 Mar 2017 19:25:46 GMT
View Forum Message <> Reply to Message

Thank you so much for the recommendation.The code worked. Could you please tell why we are using singleunit(centered) command? Also, for the SVY command is there anything I need to pay attention to the modeling (multiple and logistic regression) exercise.

Thanks again.
Baker

Subject: Re: Merging data files differend years same IR
Posted by Bridgette-DHS on Tue, 14 Mar 2017 14:38:18 GMT
View Forum Message <> Reply to Message

Another response from Tom Pullum:

There are three possible singleunit options: centered, scaled, and certainty. I have compared them and the differences are negligible (so far as I am concerned). If you do not include singleunit, with one of the options, there is a good chance that your run will simply stop because it does not have enough variation within one of the strata. As for the way you set up the model, I

cannot offer much assistance. Perhaps other forum users will help.

Subject: Re: Merging data files differend years same IR
Posted by bakerchowdhury on Wed, 15 Mar 2017 18:48:42 GMT

Great. Thank you so much for your assistance.