
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
```