Subject: IPUMS DHS Tip #7 - Comparing statistics over time
Posted by boyle014 on Thu, 18 Jul 2019 14:41:32 GMT
View Forum Message <> Reply to Message

It's very easy to review trends over time with IPUMS DHS. You can also see trends using StatCompiler, but IPUMS DHS has two advantages over that tool. First, in IPUMS DHS, the variables have been harmonized (although you should always check the COMPARABILITY and UNIVERSE information on the IPUMS DHS website about this). That's not the case for the StatCompiler numbers. Second, since IPUMS DHS uses microdata, you are able to calculate confidence intervals. That means you can see if differences are statistically significant over time.

Here's the Stata code.

```
************************************************************
*            *
* EXAMPLE DO FILE FOR USING IPUMS DHS FOR TREND ANALYSIS   *
*            *
************************************************************


/*
July 18, 2019
Stata version 14.2

*************************************
* CREATE THE MULTI-SAMPLE DATA FILE *
*************************************

This example uses the Rwanda data for all years.

1. Create an IPUMS DHS file.
Go to dhs.ipums.org.
Select all the Rwanda samples.
Find and select the agefrstmar "Age at first marriage" variable.
Find and select the age "Age" variable.
Find and select the currwork "Currently working" variable.
Submit the data extract.
Make sure to change the file format to Stata.

2. Download, unzip, save, and rename your IPUMS DHS file.

3. Open the data file in Stata.
*/

*******************
* WEIGHT THE DATA *
*******************
```

* The means and confidence intervals will be incorrect if you fail to do this.

svyset, clear

svyset [pw=perweight], psu(idhspsu) strata(idhsstrata) singleunit(centered)

****
* STATA CODE FOR COMPARING A CONTINUOUS VARIABLE OVER TIME *
****

* EXAMPLE: HAS MEAN AGE AT FIRST MARRIAGE DECREASED AMONG WOMEN 25-49 IN
RWANDA?

* The following tells Stata to provide both codes and value labels for
* variables. I find it helpful for properly recoding variables.

numlabel, add

* What is the distribution of the Age at First Marriage variable?

tabulate agefrstmar

* The tabulation shows that Not in Universe (NIU) is 99. If we leave that in, it will
* lead to incorrect calculations because Stata will think unmarried women were first married at
* the age of 99. We need to make Not in Universe a missing value instead:

replace agefrstmar=. if agefrstmar==99.

* Create a dummy age variable to use to limit analysis to younger women.

generate age20_35=0
replace age20_35=1 if age >= 20 & age <= 35

* We don't want to see both codes and labels in our table, so we take the numbers
* off:

numlabel, remove

* Run a regression predicting age at first marriage by survey year (sample) for
* younger women.

svy: regress agefrstmar i.sample if age20_35==1
margins sample
marginsplot

****
* STATA CODE FOR COMPARING THE DISTRIBUTION OF A DICHOTOMOUS VARIABLE

OVER TIME *
****

* EXAMPLE: HAS THE PERCENTAGE OF WOMEN WORKING IN RWANDA CHANGED SIGNIFICANTLY?

numlabel, add

tabulate currwork

* The tabulation shows a missing value code that needs to be replaced:

replace currwork=. if currwork==98

numlabel, remove

* Weights have been set; don't need to set again.

* Run a logistic regression looking at percent women working by survey year.

svy: logit currwork i.sample
margins sample
marginsplot