Subject: Compare two surveys
Posted by kmdshoyaib on Mon, 04 Jul 2022 05:53:52 GMT
View Forum Message <> Reply to Message

Hi all, I am planning to compare two Individual datasets from 2015 survey and 2019 survey. My dependent variable would be from Domestic Violence module and independent variables would be other factors such as occupation literacy etc.

I would like to know how to compare the variables affecting the domestic module.

Should i do the regression analysis on each dataset and then compare the odds ratio from one dataset to another dataset or is there any other way.

Any leads are highly appreciated.

Subject: Re: Compare two surveys
Posted by Janet-DHS on Thu, 07 Jul 2022 16:34:13 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

You can pool the IR files from the two surveys into a single file. In Stata, this is done with the "append" command. Define a variable "survey" that is 0 for the cases in the 2015 survey and 1 for the cases in the 2019 survey. Then you can do the same kind of regression that you would do for a single survey, but in the pooled file you include "survey" as a covariate or predictor. If the coefficient for "survey" is statistically significant, then the relationship of the outcome to the other covariates has changed. This is a simple description of a strategy we often use at DHS.

Subject: Re: Compare two surveys
Posted by kmdshoyaib on Fri, 29 Jul 2022 02:01:42 GMT
View Forum Message <> Reply to Message

Thank you for the reply. When comparing two different surveys, should i use complex analysis plan file for survey 1 or complex analysis plan file for survey 2 ?

Subject: Re: Compare two surveys
Posted by kmdshoyaib on Fri, 29 Jul 2022 02:03:21 GMT
View Forum Message <> Reply to Message

Thank you for the response. When comparing, should I use the analysis plan file of survey 1 or analysis plan file of survey 2 ?

Subject: Re: Compare two surveys
Posted by Janet-DHS on Mon, 01 Aug 2022 16:34:28 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

You need to construct new variables for the clusters and strata, like this:

egen cluster_ID=group(survey v001)
egen stratum_ID=group(survey v023)

and use these new variables in svyset. Note that even if v023 is defined the same way in both surveys, you still need to distinguish between the two surveys. You do not need to alter the weights. Good luck.

Subject: Re: Compare two surveys
Posted by Nikee on Wed, 30 Nov 2022 17:08:16 GMT
View Forum Message <> Reply to Message

Greetings Janet,

I want to ask what kind of regression models and techniques could be used while comparing the two datasets from these two surveys. I am assessing the impact of different socioeconomic development factors on domestic violence.

Subject: Re: Compare two surveys
Posted by Janet-DHS on Tue, 06 Dec 2022 14:19:48 GMT
View Forum Message <> Reply to Message

Following is a response from DHS staff member Tom Pullum:

As suggested on July 7, you would construct a pooled file with a binary variable that is coded S=0 for the first survey and S=1 for the second survey. Say that V is a scale constructed from the DV module. In Stata, enter "regress V S". Look at the coefficient for S. If it is positive and statistically significant, then the mean of V is greater in survey 1 than in survey 0. If V is also a 0/1 (binary) variable, then you enter "logit V S" and look at the coefficient for S. You can have more elaborate models that include interaction terms, controls, etc.

It sounds like you want to include macro-level indicators of development. There are limitations to this. Suppose you hypothesized that domestic violence declines as women's education improves. To show this, you add to your regression the national percentage of women who have achieved some level of education at the time of the first survey and the time of the second survey--that is, two numbers, one for each survey. You re-run the regression above, including those numbers. This will not work, because those national two percentages will be confounded with S. (There are other ways to describe this issue.) But suppose instead that you used a variable in the data files,

such as E=0 if the woman had a "lower" level of education and E=1 if she had a "higher" level. You could include E in your regression, AND you could include the interaction between E and S, to get at the effect of E on differences in V and differences in the trend. You could include cluster-level covariates from another source, such as the DHS spatial covariates files.

The models should include svyset and svy as described in earlier posts.