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.