Subject: Pooling men, women and household DHS from Haiti
Posted by acolombo on Wed, 26 Oct 2016 22:35:56 GMT
View Forum Message <> Reply to Message

Dear all,

I would like to pool three different waves of DHS carried in Haiti in 2000, 2005/06 and 2012. The main goal is to investigate the determinants of employment in the island using a probit model. In particular, I would like to understand how the likelihood of being employed is affected by: gender; level of education attained; place of residence (metro-area, other urban areas and rural areas); number of children; size of the household. Scott and Rodella (2016) already carried a similar analysis by pooling surveys coming from a difference source.

I dived into this thread all day and after having read several post (and numerous aspirins), I would like to sum up what I have learnt and to check whether I am on the right track about procedure to follow on Stata14.

"Denormalizing" weights: for each men (HTMR61DT) and women (HTIR61DT) survey and wave, I would have to:
multiply weights (mv005) and (v005) by the ratio of total men/female 15-49 population divided by the size of the male/female sample. In my case, the best I can get is estimation of male/female population aging 15-64 for every year. Is it enough? create a cluster including the survey id create a strata that also includes the survey (identified by v000) in the group command

Just for the sake of clarity, using as an example the 2012 women survey, the two previous steps would be implemented by coding the following:

replace v005=v005/1000000
gen v005_new=v005*wpop1564_2012/wsize1564_2012
egen v001r = group(v000 v001) // cluster also includes the survey in the group command
egen strata=group(v000 v025 sregnew) // strata also includes the survey (identified by v000) in the group command
svyset v001r [pw=v005_new], strata(strata) singleunit(centered)
In order to have a full overview over the labor market, I would like to append the men survey and the women surveys. I would end up with two weight variables: v005_new and mv005_new. Should I just "merge" them or should I further distinguish the weights coming from the men surveys from the weights from the women surveys? How would I do that? I would like then to include in this new dataset the respective household characteristics for every eligible men and women. Can I merge them without risk? Or should I take into account the weight assigned to the eligible individuals in the household dataset (HTPR61DT)? I repeat the above procedure for all the waves. At the end I should I have three dataset with eligible men, women, and their respective household characteristics for three years Append the three dataset from each wave, but:
should I have only one variable with the weights I derived in the previous steps? Or should I further manipulate them so that weights from different surveys can be distinguished from eachother? I will have only one variable distinguishing each strata (metroarea/urban, metroarea/rural, grandeanse/rural, grandeanse/urban, and so on and so forth) for each wave.
I would like to remark, though, that I have a different number of strata depending on the wave:
In 2000 there are 19 strata (9 districts*2 along the urban rural dimension, and one strata

metroarea-urban) In 2005/06 there are 21 strata (10 disticts*2, plus the metro area), since one district split in two in 2003 In 2012 there are 23 strata (10 districts without camps*2, metro area without camp, camps/rural, and camps/urban), since camps were built after the 2010 earthquake and the survey aims at investigating the living conditions of those households relocated after the disaster.

Is this a problem? If yes, how can I handle it? Finally, there is no unanimous consensus on whether I should re-normalize the weights or not. What would you suggest? How would you proceed? If everything I wrote above is correct, I should be able to carry my analysis by simply using the svy commands together with the new weight and strata variables.

What do you think? Thank you in advance for your help. I remain at your disposal if I was not clear enough and you need further details. Also, if you bumped into this post, have the same issue of mine, either you found a solution or not (or you think you have found it), please do not hesitate to contact me!

Andrea