Subject: Number of strata in 1999/2000 and 2004 BD DHS
Posted by annadinnyc on Thu, 13 Mar 2014 20:11:56 GMT
View Forum Message <> Reply to Message

Hello,

I cannot find the number of strata used in the BD DHS 1999/2000 and 2004. I checked the final reports and they only provide the number of Enumeration Areas (341 and 361, respectively). The reports just mention choosing EAs from the list of EAs in the 1991 or 2001 census, respectively, but it is unclear how the two-stage selection was done.

When I tabulate the stratum variables (hv022 or v022, labeled "sample stratum number"), I get 168 for the 1999/2000 data and 177 for the 2004 data. In contrast, the 2007 and 2011 BD DHS have 22 and 20 strata, respectively.

I am trying to merge data from four BD DHS surveys and need to make sure that I get the survey design, weights, etc. correct.

Any suggestions would be appreciated. Thank you.

Anna

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by user-rhs on Thu, 13 Mar 2014 20:32:18 GMT
View Forum Message <> Reply to Message

Anna--the stratum that you specify in the complex sampling adjustment of the statistical software (e.g. -svyset- in Stata) is the stratification variable, not the actual stratum number. So for example, if it was a stratified two-stage cluster sampling design where it was stratified by urban/rural residence, then you would specify the vbl that indicates urban/rural (e.g. v025 in DHS) as the stratum instead of v022. Check the final reports on how the sampling was done. In other non-DHS surveys I've worked with, stratification was done by province, etc, so you will need to make the adjustment accordingly.

hth,
RHS

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by user-rhs on Thu, 13 Mar 2014 20:42:22 GMT
View Forum Message <> Reply to Message

I also suggest before merging to either create a cluster variable that is specific to each survey round (example: tack on "_2007" for all the obs'ns sampled in 2007, tack on "_2011" for all obs'ns sampled in 2011, etc.) OR using PSU-nested-within-survey year random effects if you do not want to create the survey round-specific cluster vbl in a regression model.

Example of a logit model using -xtmelogit- in Stata (v007 is year of interview, v001 is the PSU variable):
xtmelogit outcome predictor1 predictor2 blablabla ||v007:|| v001:

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by annadinnyc on Thu, 13 Mar 2014 21:13:06 GMT
View Forum Message <> Reply to Message

Thanks for the response.

The DHS reports for 1999/2000 (http://dhsprogram.com/publications/publication-fr119-dhs-fin al-reports.cfm) and 2004 (http://dhsprogram.com/publications/publication-fr165-dhs-fin al-reports.cfm) do not describe the stratification process explicitly. In the 2007 and 2011 reports, they explain that stratification is done by division and then rural / urban (with further classifications). The 2004 report suggests that stratification was done similarly to 2007 (by division, rural, three types of urban) but I would prefer to be certain. And the 2000 report does not provide enough information.

In the 2007 and 2011 data, the variables for strata (v022 and v023) are correct (based on division and city corp/other urban/rural.

I hope that a DHS person who knows the BD surveys will respond.

Thanks,

Anna

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by Reduced-For(u)m on Thu, 13 Mar 2014 22:19:45 GMT
View Forum Message <> Reply to Message

Anna,

I'm not a DHS person, but I've used the BD surveys a good bit. I used the code from the DHS FAQ: gen strata = group(v024 v025)

BUT - looking at the documentation for the 1996 survey (https://www.dhsprogram.com/pubs/pdf/FR88/FR88.pdf) page 3, they actually appear to have stratified at a different level ("Each division was stratified into three groups: 1) statistical metropolitan areas (SMAs) 2,

2) municipalities (other urban areas), and 3) rural areas." - so not just urban/rural-X-region, but urban/SMA/rural-X-region. So that would look to me like you should use:

gen strata = group(v023 v025) for 1996.

It looks like this system changed from round to round. Further, v023 is all 0 for 2004, meaning it isn't representative at sub-national levels, and for that one you'd need to use the group(v024 v025).

I'd suggest looking at the sampling design section (like the one I quoted) from each survey, generate the appropriate strata for each survey (by, say, tacking on the survey round, so 1_1996, 2_1996, 1_1997 - however you want to do that)and then use that new strata var in your svyset command. But I'm pretty sure v022 isn't the right one ever (something about for use with Taylor expansions? I dunno).

So I learned something, and I'd been doing it wrong. And that's always nice (well, the learning part, not the embarrassment part).

Hope this helps a bit. It helped me.

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by Sarah-DHS on Fri, 14 Mar 2014 12:32:12 GMT
View Forum Message <> Reply to Message

Dear Anna,

If the previous posts did not answer your question, please let us know and we will make sure you get the information you need.

Thanks,
Sarah at DHS

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by user-rhs on Fri, 14 Mar 2014 17:52:33 GMT
View Forum Message <> Reply to Message

Reduced-For(u)m wrote on Thu, 13 March 2014 18:19
So I learned something, and I'd been doing it wrong. And that's always nice (well, the learning part, not the embarrassment part).

^^^ Hoo boy, that makes two of us, Reduced-For(u)m. Thanks for clarifying this.

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by annadinnyc on Mon, 17 Mar 2014 14:57:01 GMT
View Forum Message <> Reply to Message

Hello Sarah,

I would like you to clarify please. I do not use round 1996 (which the other user refers to below). The DHS reports for 1999/2000 and 2004 are not explicit about stratification (only about the number of EAs), though the 2004 report seems to suggest a stratification procedure is similar to the 2007 data (the 2007 report was explicit).

Thanks,

Anna

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by Sarah-DHS on Tue, 18 Mar 2014 19:08:59 GMT
View Forum Message <> Reply to Message

Anna,

I'm working on getting an answer to your question.

Thanks,
Sarah at DHS

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by Sarah-DHS on Wed, 19 Mar 2014 13:33:57 GMT
View Forum Message <> Reply to Message

Dear Anna,

Please see the response below from Sampling Statistician Rulin Ren:
The first question about the variable HV022/V022, it is a typical question for many earlier surveys (earlier than 2005), where the variable HV022/V022 coded the paired clusters, not as its label said that it was sampling stratum. And the sampling stratum was not coded. To get the sampling stratum (the best approximation) one needs to use HV024 crossing HV025 (or V024 crossing V025). The BDHS 2007 and 2011 used more detailed residence type (four categories).

Please let us know if this does not answer your question.
Thank you,

Sarah at DHS

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by annadinnyc on Thu, 20 Mar 2014 15:57:37 GMT
View Forum Message <> Reply to Message

Thank you for clarifying. Much appreciated.

Anna

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by azucena2014 on Fri, 22 Aug 2014 22:09:57 GMT
View Forum Message <> Reply to Message

I have the same issue with the Uganda 2011 data. the variables v022 and v023 have no
observations. can we just create this variable ourselves using the region and urban/rural
variables?

Subject: Re: Number of strata in 1999/2000 and 2004 BD DHS
Posted by Trevor-DHS on Tue, 26 Aug 2014 01:25:51 GMT
View Forum Message <> Reply to Message

From the Uganda 2011 report (appendix A), the sample was stratified by urban rural within each
of the 10 regions. Thus you can create the strata variable as follows:
egen strata=group(v024 v025)
gen wt=v005/1000000
svyset v021 [pw=wt], strata(strata)