
Subject: Stratification and sampling in Haiti
Posted by [acolombo](#) on Wed, 28 Sep 2016 21:20:35 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear all,

I have some questions about the stratification and sampling strategy, and the conclusions I could draw from it, used for the 2005-06 DHS in Haiti. It looks a little bit different from the procedures I am used to.

The message will be a bit long, and I already apologize for that, but context and details are fundamental, I believe. Everything is based on the Household Recode dataset (hthr52dt).

In Haiti there are 10 departments (ADM1) and one metropolitan area (only urban). The stratification occurs at this level: each department is divided into urban and rural strata. In total there are 21 strata. For those interested and french-speaking (I can provide a translation in case), the text in the appendix of the final report explains:

L'échantillon de l'EMMUS-IV est un échantillon stratifié représentatif au niveau national tiré à deux degrés. Les onze départements sont stratifiés en parties urbaine et rurale pour former les strates d'échantillonnage. L'Aire Métropolitaine n'a qu'une partie urbaine. Donc, au total 21 strates d'échantillonnage ont été créées. L'échantillon au premier degré a été tiré indépendamment dans chaque strate, et l'échantillon au second degré a été tiré indépendamment dans chaque unité primaire tirée au premier degré.

The sampling strategy, instead, is the following. There are two stages: in the first stage a total of 339 clusters are selected from the strata proportionally to the number of household they host. This means that for each strata, highly populated clusters are oversampled. In the second stage, a "fix" number of households (tirage systématique a probabilité égale) is picked: 26 households from urban clusters and 34 households from rural clusters.

Now the questions:

1. why, if there are 21 strata, the variable hv022 (sample stratum number) assumes values from 1 to 163?
2. Given this sampling framework, can I:
infe the proportion of urban population per region (by computing the share of urban households in one region)? the proportion of people with access to electricity for the whole country each region for urban and rural population in each region

I would be really grateful if you could help me, as I've been breaking my head on this dilemma for several days.

Thanks,

Andrea

Subject: Re: Stratification and sampling in Haiti
Posted by [Bridgette-DHS](#) on Mon, 03 Oct 2016 12:10:14 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Specialists: Tom Pullum & Trevor Croft.

construct the strata variable in Stata with either of two options:

Option 1:

```
gen strata = v024 * 2 + v025 - 1  
replace strata = 1 if smetrop == 1
```

or Option 2:

```
egen strata = group(v024 smetrop)
```

They result in the same variable. Option 2 is close to "group(v024 v025)", but not quite the same, because v024, code 1, groups together two separate areas Aire metropolitan and the rest of the Ouest region.

Subject: Re: Stratification and sampling in Haiti
Posted by [acolombo](#) on Mon, 03 Oct 2016 13:01:40 GMT
[View Forum Message](#) <> [Reply to Message](#)

Excellent!

I managed to figure it out before, but you gave me a much appreciated confirmation.

The second question I had concerns the representativeness. On the website it's stated that surveys are generally representative at the national, regional and place of residence (urban vs rural) level. Is this true also for the case of Haiti? I'm wondering whether creating 21 strata (urban and rural for each of the 10 regions plus Port-au-Prince) allows me to compare urban and rural areas at the regional level.

For instance, can I infer anything about the access to electricity for urban households across regions? Or can I say anything only at the national level, comparing urban and rural households?

Thank you for your answer,

Andrea

Subject: Re: Stratification and sampling in Haiti
Posted by [acolombo](#) on Mon, 03 Oct 2016 14:21:40 GMT
[View Forum Message](#) <> [Reply to Message](#)

Also (and apologies for taking advantage of your kindness), could you please explain how do you

apply the formula we agreed on in order to retrieve the 26 strata specified in appendix A of the 2012 DHS survey for Haiti?

There are 10 departments + 1 metropolitan area. The then departments are divided in urban and rural. The metropolitan area (only urban) is stratified into its 6 municipalities. However, there is no variable disentangling the SDE from each metropolitan municipalities. In other words, I don't have enough information to use the code line you suggested before.

Another question is: how did you include the camps in the stratification strategy?

Thank you very much, once more.

Andrea

Subject: Re: Stratification and sampling in Haiti
Posted by [Bridgette-DHS](#) on Wed, 05 Oct 2016 10:02:56 GMT
[View Forum Message](#) <> [Reply to Message](#)

Here is a response from Trevor Croft and Tom Pullum:

Your first question is why, if there are 21 strata, does the variable hv022 (sample stratum number) assume values from 1 to 163? Here is a more complete answer. DHS previously used a procedure of constructing implicit strata (163 of them, for this survey) based on pairing (or in some cases groups of 3) clusters. These implicit strata were constructed within the explicit strata (the 21 strata) and were used to calculate sampling errors. DHS stopped using this procedure some years ago, but the dataset includes the constructed implicit strata variable.

Your question about representativeness below the national level often comes up. At the stratum level, there isn't really an issue. In essence, separate samples have been drawn within each stratum. Small strata tend to be over-sampled (conversely, large strata tend to be under-sampled) in order to have enough cases to be able to make good estimates of key indicators. "Representative" has two dimensions--bias and statistical uncertainty. Stratum level estimates are unbiased and have reasonable standard errors. Yes, and you can compare strata with one another as you described (the urban and rural parts of the same region), but you should check for statistical significance. If you go below the stratum level, for example to the second administrative level, generically called districts, the estimates are still unbiased, but the standard errors go way up. It is very important to include standard errors for these lower-level estimates, just as you would for categories of a covariate at the national level. If you compare two districts within the same region, it can be difficult to get a statistically significant difference because both estimates have high standard errors.

Subject: Re: Stratification and sampling in Haiti
Posted by [acolombo](#) on Wed, 05 Oct 2016 13:45:35 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you for your answer,

concerning DHS 2012 (household survey) could you please explain more in detail how the stratification was conducted?

There are 10 districts, the metropolitan area, and the camps. In the Appendix of the report it is specified that 26 strata were identified: rural and urban for each district and the six municipalities for the metropolitan area.

My questions are:

a) how can I build a strata variable as above, if I do not have any variable distinguishing SDEs extracted from each of the 6 metropolitan municipalities?

b) how were the camps considered? From page 355 of the report it looks like that they are a separate "domaine d'échantillonnage". Did you also distinguish camps along the urban/rural dimension?

Thank you,

Andrea

Subject: Re: Stratification and sampling in Haiti

Posted by [Bridgette-DHS](#) on Tue, 03 Jan 2017 19:31:35 GMT

[View Forum Message](#) <> [Reply to Message](#)

We reached out to the user via registered email, but there has been no response so far.
