
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 systematique a probabilité egale) 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
