Subject: Using weights for small geographical locations
Posted by nobleman00 on Wed, 26 Nov 2014 01:29:12 GMT
View Forum Message <> Reply to Message

Hello,

I have a question regarding the sample weight (hv005) variable in the household dataset. I am currently analyzing household datasets (HR) for 40+ countries where GPS information is available. With two variables (hv201, hv205), I basically need to conduct the same analysis for two different levels: (1) sub-national boundary level of each country (hv024), (2) grid cell level (10km*10km) of each country. I created the grid-cell (10km*10km) level within a country in order to have smaller geographical units than sub-national boundaries.

When doing the analysis for (1) sub-national boundary level, I use the "hv005" variable whenever I produce some tables or collapse the datasets by sub-national boundary (hv024).

I was wondering if I can still use the same sample weight variable (hv005) when doing the analysis for (2) grid cell level (10km*10km) which shows smaller geographical units than sub-national boundary (hv024). I am asking this because I think I've read at some point that the "hv005" is designed to represent samples only at the sub-regional level (sub-national boundary) in a country. I wonder if I can still use hv005 when collapsing the dataset by the grid-cell (10km*10km) level which I intentionally created.

Thank you.

Subject: Re: Using weights for small geographical locations
Posted by Trevor-DHS on Mon, 01 Dec 2014 15:30:14 GMT
View Forum Message <> Reply to Message

You can and should still use hv005 as the sample weight, but doing your analysis with smaller geographic units is potentially problematic. The sample is designed to be representative at the region level, but not at the level of smaller units. As you disaggregate the data to smaller units the sample is less and less likely to be representative. The sample is also designed to provide a certain level of accuracy at the region level, and again as you disaggregate to smaller units the accuracy of those estimates gets worse and worse and the confidence intervals around the estimates quickly become very large and unreliable.

Subject: Re: Using weights for small geographical locations
Posted by nobleman00 on Tue, 02 Dec 2014 00:59:46 GMT
View Forum Message <> Reply to Message

Thank you for your response.

I was also concerned that the sample is designed to be representative at the region level, not at the level of smaller units.

Due to this reason, (1) the sub-national boundary level is the primary focus of this analysis. The purpose of (2) the grid cell level analysis is simply to show the raw outputs only for the households included in survey clusters within each grid cell, acknowledging that the results are no longer representative at any level (neither the sub-national boundary level nor the grid cell level).
In this regard, I was considering doing the second level analysis (grid cell) without using sample weights. Would this sound unreasonable?
Thank you for your time in advance.

Jung

---

Subject: Re: Using weights for small geographical locations
Posted by Trevor-DHS on Tue, 02 Dec 2014 16:07:02 GMT
View Forum Message <> Reply to Message

You could do your grid cell level analysis unweighted, particularly if there is only one cluster in each grid cell, but if you have more than one cluster in a grid cell then they are likely to have different weights and thus to combine them it would be better to weight the data. It is not clear though what the purpose and benefit of the grid cell level analysis would be if you know that the results are not representative.

---

Subject: Re: Using weights for small geographical locations
Posted by degreepupil on Thu, 10 Jul 2025 08:55:38 GMT
View Forum Message <> Reply to Message

nobleman00 wrote on Tue, 25 November 2014 20:29Hello,

I have a question regarding the sample weight (hv005 Geometry Dash) variable in the household dataset. I am currently analyzing household datasets (HR) for 40+ countries where GPS information is available. With two variables (hv201, hv205), I basically need to conduct the same analysis for two different levels: (1) sub-national boundary level of each country (hv024), (2) grid cell level (10km*10km) of each country. I created the grid-cell (10km*10km) level within a country in order to have smaller geographical units than sub-national boundaries.

When doing the analysis for (1) sub-national boundary level, I use the "hv005" variable whenever I produce some tables or collapse the datasets by sub-national boundary (hv024).

I was wondering if I can still use the same sample weight variable (hv005) when doing the analysis for (2) grid cell level (10km*10km) which shows smaller geographical units than sub-national boundary (hv024). I am asking this because I think I've read at some point that the "hv005" is designed to represent samples only at the sub-regional level (sub-national boundary) in a country. I wonder if I can still use hv005 when collapsing the dataset by the grid-cell (10km*10km) level which I intentionally created.

Thank you.
You can use hv005 at the grid-cell level, but with caution. It is designed for national and sub-national estimates (e.g., hv024), not for finer, non-standard units like 10x10 km grids. Using hv005 in grid-level analysis may introduce bias due to design not accounting for representativeness at that scale. For exploratory spatial patterns, it's acceptable, but for inference or policy decisions, results should be interpreted carefully.