
Subject: 2011 stata data file

Posted by [rsroche](#) on Sat, 24 Oct 2015 17:02:22 GMT

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

Hi,

I am using the 2011 Stata data set for Uganda and it appears to be missing any data for the HV022 Sample Strata variable. Without this I am not sure how I can apply the survey weighting, is there a reason why it is missing or should I be using another variable?

Thanks for your help,

Rachel

Subject: Re: 2011 stata data file

Posted by [Reduced-For\(u\)m](#) on Mon, 26 Oct 2015 18:47:35 GMT

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

Which recode are you using? Is the variable "V022" there?

Subject: Re: 2011 stata data file

Posted by [rsroche](#) on Tue, 27 Oct 2015 05:12:27 GMT

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

It's the Household recode. V022 is not there. I have been advised elsewhere that there has been confusion about HV022 as it is not the Strata variable but the sample strata used for sampling errors. I'm not sure now which variable I should have been using as Strata and maybe this is the wrong one, but I've used HV022 in my analyses for other data sets and it seemed to work. Any advice greatly appreciated.

Thanks for your help.

Rachel

Subject: Re: 2011 stata data file

Posted by [Liz-DHS](#) on Thu, 29 Oct 2015 18:54:49 GMT

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

Dear User,

Here is a response from Senior Statistician, Dr. Ruilin Ren,

Quote:

Variable V022 is at individual level, HV022 is at household level, but the two variables have the same contents. In older survey data sets, HV022/V022 might be codes for the paired clusters used for sampling error calculation, even the label of the variable is sampling stratum. Checking the frequency of the variable will help you understand if it is paired clusters (HV022/V022 has a large number of categories) or not. In recent surveys, HV022/V022 is labeled as "sample strata used for sampling errors", it is actually the sampling stratum used in sample selection. Another variable HV023/V023 is labeled as "sample strata used for sample selection", it is actually the same thing as HV022/V022 without exception. It might be the case that not all of the two variables are coded, use the one which is coded. Again, you can check the frequency of the variable. In any case, if you use HV025/V025 crossing HV024/V024, you will get the correct stratification in most of the cases, or it is close enough to the true stratification which will be good enough for all analysis purposes where stratification information is requested.

If you have additional questions, please feel free to post again. Thank you!

Subject: Re: 2011 stata data file

Posted by [luke](#) on Mon, 02 Nov 2015 18:26:22 GMT

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

I have the question as Rachel. In other data sets, I use the hv022 sampling strata to calculate confidence intervals and standard errors on binary variables, but in Uganda 6 hv022 consists only of null values. Rachel - it should be possible to recreate hv022 from sampling frame description in the final report, but I have not done this. Have you had any luck?

Subject: Re: 2011 stata data file

Posted by [luke](#) on Mon, 02 Nov 2015 19:36:23 GMT

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

Update: The DHS Final Report for Uganda 2011 (FR264) says "clusters, were selected from among a list of clusters sampled for the 2009/10 Uganda National Household Survey (2010 UNHS)." The UNHS 2010 report is available for download online, and that report says that ten regions compose the strata. UNHS did not divide each region into urban/rural for sampling, which we often see in other surveys and which had been done for an earlier UNHS.

So, I believe that you can pass region to svyset as the strata. That would certainly be true if you were analyzing UNHS data, but since the Uganda 2011 DHS clusters are a random sample of UNHS clusters, I'm not totally sure this is correct. Perhaps there is no clearly defined "strata" which is why hv023 is left as null. In any case, it might be good enough for your purposes.

Subject: Re: 2011 stata data file

Posted by [duke2015](#) on Mon, 23 Nov 2015 14:25:00 GMT

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

Hi Luke,

I looked online and found this link for the UNHS 2010 report:
<http://www.ubos.org/UNHS0910/unhs200910.pdf>. If you look at the sampling section, it says that :

The UNHS 2009/10 sample was designed to allow reliable estimation of key indicators for the Uganda, rural-urban, and separately for ten sub regions. A two-stage stratified sampling design was used. At the first stage, Enumeration Areas (EAs) were grouped by districts and rural-urban location; then drawn using Probability Proportional to Size (PPS). At the second stage, households which are the Ultimate Sampling Units were drawn using Systematic Sampling.

It looks like they used district (not sure if this is same as region?) and rural/urban for their stratification. Where were you looking where they said they didn't use rural/urban?

Subject: Re: 2011 stata data file
Posted by [jack.murphy](#) on Mon, 19 Dec 2016 15:40:20 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello all,

I am using the Individual Recode 2011 Uganda DHS data set in Stata, and have encountered a very high number of missing values (90-100% of the values) for the variables v023, v024, v101 and v139. Is there another variable that contains the region/district values? My goal is to make a "stratumid" variable from v025 (urban/rural) and region.

Thanks,
Jack
