
Subject: Set up DRC data in Stata
Posted by [Ryan](#) on Thu, 04 Sep 2014 16:15:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello,

I would like to calculate means for a number of variables across provinces in the DRC (2007 dataset). I am not sure if the sampling procedure would preclude this and I am a little confused about how to set up my analysis (weights, strata, psu) in Stata. The main issue is that there are no reported strata (v022). I understand that the DRC had an uncommon, 3 stage sample design, but between my poor French and the new design, I'm not sure how to proceed.

Any advice would be greatly appreciated.

Thank you,

Ryan

Subject: Re: Set up DRC data in Stata
Posted by [Bridgette-DHS](#) on Fri, 05 Sep 2014 17:33:08 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

In this survey the stratum variable can be constructed as v024 x v025. In all the DHS surveys I know about, this will give the strata or be the best approximation if, say, v022 is not in the data. The summarize command, and most related commands, will usually not work with weights, but "mean(y), over(x)" will work. For example, for y=v201 (children ever born), here is the syntax:

- * Open CDIR50FL.dta
- * region or province: v024
- * children ever born: v201
- * stratum: v024 x v025
- * weight: v005
- * cluster: v001 or v021 (the same)

```
egen stratum=group(v024 v025)
svyset v001 [pweight=v005], strata(stratum)
svy: mean v201, over(v024)
```

Subject: Re: Set up DRC data in Stata
Posted by [Ryan](#) on Fri, 05 Sep 2014 17:38:00 GMT

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

Excellent. Thank you Tom (and Bridgette).

Subject: Re: Set up DRC data in Stata
Posted by [Bridgette-DHS](#) on Fri, 05 Sep 2014 17:39:25 GMT

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

You're welcome.

Subject: Re: Set up DRC data in Stata
Posted by [soniwe](#) on Tue, 07 Feb 2017 07:05:18 GMT

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

Hi there,

I have just seen this post and have come across the same problem with DRC 2007. However, the report definitely says that there were 34 sampling strata created using three strata (statutory city, city, rural area) for 10 of the 11 provinces (except Kinshasa). I can't find any variable that relates to these three strata, and even if there was, that would create 31 strata not 34! Using v024 and v025 gives 21 strata. I am now quite confused about how this stratification was done. Can you give any more clarification on this?

I'm having a similar problem with Uganda 2011. There is no data for sample strata in either v022 or v023. The report states that a stratified design was used, but does not explicitly say what the strata are. I assumed v024 and v025 (region and urban/rural area of residence), but just wanted to verify, as Uganda 2006 DHS seems to use region plus refugee camps, not urban/rural.

Thanks,

Sonia

Subject: Re: Set up DRC data in Stata
Posted by [Bridgette-DHS](#) on Wed, 08 Feb 2017 19:31:04 GMT

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

Following is a response from DHS Senior Sampling Specialist, Ruilin Ren:

Regarding the DRC 2007, the stratification was given in the appendix A, urban was split by two categories, and the rural, since Kinshasa has only Statutory city, so there were 34 sampling strata. But the data file may not give the two categories of urban. Crossing V024 and V025 was not correct for the stratification. If you want to know which stratum is which, produce a frequency table on number of clusters by using the v024 and the stratification variable, compare it with the sample allocation table, you will know exactly which stratum is which.

Regarding the Uganda 2011, it is true that the appendix B did not give the stratification details, but the sample allocation table in appendix A gives the stratification because the sample was allocated by stratum. V024 crossing V025 will give the correct stratification.

Subject: Re: Set up DRC data in Stata
Posted by [soniwe](#) on Thu, 09 Feb 2017 07:23:21 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks for your response, and the clarification on the Uganda 2011 sample. For DRC 2007, I have now understood why there are 34 strata, but I am still not sure what to use as the "stratification variable" as v022 has no data in it and I can't find anything else that looks relevant for the three strata ville statutaire, cite, secteur/chefferie, and the four districts of Kinshasa. How do I correctly create a strata variable for the svyset command?

Subject: Re: Set up DRC data in Stata
Posted by [Bridgette-DHS](#) on Fri, 10 Feb 2017 13:32:07 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is another response from Ruilin Ren:

I do not know why the stratification variable was not coded in the final data. However, if your aim is just for statistical analysis, and not intended to compare the difference between the urban categories (meaning you do not need to identify the stratum for restricted analysis), then using V024 crossed with V025 as stratification will be good enough, although it is not the exact stratification. Actually, the two different categories of the urban areas do not differ much in the Congo DR settings.
