
Subject: NFHS-5 data from STATA not matching the factsheets

Posted by [anshul.11](#) on Wed, 20 Dec 2023 19:32:07 GMT

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

Hi!

For a project, I am supposed to extract data from the Household data file. However, despite running commands which as per me are correct for tabulating, the results are not matching the fact sheets published by DHS. For example, following are the commands I run for defining svyset

```
gen pwt=.
replace pwt= hv005/1000000
tab hv206
egen cluster_id = group( hv021 hv024 )
egen stratum_id = group( hv023 hv024)
svyset cluster_id [pw=pwt], strata(stratum_id)
```

following which, I am trying to tabulate the percentage of households which have electricity in each state

```
svy: tabulate hv024 hv206, row
```

However, the results have a mismatch from the factsheets. as in, Bihar has 95.61% households with electricity as per the output in STATA. In the factsheet the percentage is 96.3% (accessed from: http://rchiips.org/nfhs/NFHS-5_FCTS/Bihar.pdf). Similarly, for the country, the percentage of households with electricity in the output is 96.53%; factsheet has 96.8% as the value. Same discrepancy exists with other indicators. Can anyone please let me know what am I doing wrong?

Subject: Re: NFHS-5 data from STATA not matching the factsheets

Posted by [Bridgette-DHS](#) on Thu, 21 Dec 2023 19:55:27 GMT

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

Following is a response from Senior DHS staff member, Tom Pullum:

I see that you used the HR file, in which households are units. Your calculation is correct for households. However, in the Bihar report, and elsewhere, the label is "Population living in households with electricity (%)". If you run the same lines on the PR file, in which individuals are units, you will match the report. Or, see below, you can match the report with the HR file if you multiply the weight by the household size.

For the percentages you only need the weights. You do not need the full svyset command. The adjustments for clusters and strata can be omitted. Also you do not need all the steps for the weights. You just need the following two lines in the HR file:

* The following line gives the percentages for households
tab hv024 hv206 [iweight=hv005/1000000], row

* The following line gives the percentages for individuals
tab hv024 hv206 [iweight=hv009*hv005/1000000], row

In the second command, I multiply the weight by hv009, which is the number of individuals in the household.

Subject: Re: NFHS-5 data from STATA not matching the factsheets

Posted by [anshul.11](#) on Fri, 22 Dec 2023 06:36:04 GMT

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

Thank you so much Tom and Bridgette! It solved my issue.

I have a follow-up query Since we do not svyset for percentages across the dataset. Can you please specify where do we need it? which usecase does the svyset command serve in respect to NFHS. For example, if we are using the wealth index estimates at a district level, will [iw=hv005/1000000] suffice?

Thanks in advance.

Subject: Re: NFHS-5 data from STATA not matching the factsheets

Posted by [Bridgette-DHS](#) on Fri, 22 Dec 2023 14:08:38 GMT

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

Following is a response from Senior DHS staff member, Tom Pullum:

In general, you should always adjust for weights, one way or another, in order to get unbiased estimates of population values. Adjustments for clusters and strata are only relevant for the calculation of standard errors, which are used for confidence intervals or statistical tests. I'm sure that many users include the full svyset adjustments even when they are not producing confidence intervals or test statistics, and it's ok to do that. In general, just as a matter of principle, I prefer simplicity over complexity and I don't like to include options that are not needed. I'll admit that's somewhat retro, in a world that is increasingly complex!

