
Subject: District-level sex ratio and district-level age distribution

Posted by [Lea F](#) on Fri, 21 Jan 2022 14:54:08 GMT

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

Hi,

I am using the 2015-2016 DHS data for India (NFHS-4), and I would like to calculate the following:

1. District-level sex ratio (not at birth but in the total population)
2. District-level age distribution (e.g. in 5-year categories, or % below 15 years of age, or something similar)

I was wondering: a) how these would be calculated from the dataset, and b) which dataset would be best to use for these estimations - I assume the household survey would be best, but am not sure.

Thank you very much for your help!

Subject: Re: District-level sex ratio and district-level age distribution

Posted by [Bridgette-DHS](#) on Sun, 23 Jan 2022 21:07:02 GMT

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

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

This would be done with the PR file. The following lines use the "collapse" command to give you the sex distribution and the age distribution (in 5-year intervals). There is also a file that gives the age-sex distribution. You need to specify whether you want the de facto or de jure population. Even the collapsed files are large!

* You need to specify a workspace
cd e:\DHS\DHS_data\scratch

use hv005 hv102-hv105 shdistri using "C:\Users\26216\ICF\Analysis - Shared
Resources\Data\DHSdata\IAPR74FL.DTA" , clear

gen wt=hv005/1000000

* If you want the de jure population
keep if hv102==1

* If, instead, you want the de facto population
* keep if hv103==1

* drop if age (hv105) is missing (.02 of 1% of cases). hv105 is 95 for age 95+
drop if hv105>95

* age5 gives standard age intervals for age 0-4, 5-9, etc
gen age5=1+ int(hv105/5)

collapse (sum) wt, by(hv104 age5 shdistri)
save lAtemp_age_sex_district.dta, replace

use lAtemp_age_sex_district.dta, clear
collapse (sum) wt, by(sex district)

* this file gives the weighted numbers of males and females in each district in the NFHS-4
save lAtemp_sex_district.dta, replace

use lAtemp_age_sex_district.dta, clear
collapse (sum) wt, by(age district)

* this file gives the weighted numbers of males and females in each five-year age group in the NFHS-4
save lAtemp_age_district.dta, replace

Subject: Re: District-level sex ratio and district-level age distribution
Posted by [Lea F](#) on Mon, 24 Jan 2022 20:22:19 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks very much for this response. This looks like it's Stata code - I am working in R and have read the Stata data files into R. Would it be possible to explain the calculation steps so that I can write equivalent code for R (instead of running the provided Stata code)?

Thank you very much!
Lena

Subject: Re: District-level sex ratio and district-level age distribution
Posted by [Bridgette-DHS](#) on Tue, 25 Jan 2022 12:42:58 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

I just looked again and I see a couple of typos. I didn't actually test the program because the file is so large.

In these lines:

```
use lAtemp_age_sex_district.dta, clear
collapse (sum) wt, by(sex district)
```

* this file gives the weighted numbers of males and females in each district in the NFHS-4
save lAtemp_age_sex_district.dta, replace

```
use lAtemp_age_sex_district.dta, clear
collapse (sum) wt, by(age district)
```

"by(sex district)" should be replaced by "by(hv104 shdistri)" and "by(age district)" should be replaced by "by(age5 shdistri)".

I really think it is very clear what each step of the program does. Stata code is very intuitive. There are no complicated commands or options. I'll just clarify one command. "collapse (sum) wt, by(hv104 shdistri)" means that the file is reduced by adding up the values of "wt" within each combination of sex and district. There will be one line for each combination and it will include the total weighted number of cases.