Subject: Number of respondents mismatch in dataset and published report
Posted by gkibria1@jhu.edu on Mon, 02 May 2016 06:35:02 GMT

Hi
I was using the dataset of DHS 2014 of Bangladesh for a paper in my university. I already took permission of that. I was looking also at the published Bangladesh Demographic and Health Survey 2014 by the USAID. Wne I looked into the table 7.2 (page number 75) of the published report/book, I found that the number of respondents living in urban area was 4,709 and 12,149 in rural area. But when I was analyzing the dataset (in stata) BDIR70FL.DTA, I found the number was mismatch (6,1666 in urban and 11,693 in rural). Can you please help me to explain this?
Regards,
Kibria

Subject: Re: Number of respondents mismatch in dataset and published report
Posted by Bridgette-DHS on Fri, 06 May 2016 15:03:30 GMT

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

Two problems. First, you are not using weights. Virtually every number in DHS tables is weighted. Second, you are not using marital status. The table is restricted to currently married women. It is important to read the table titles carefully. If you run this line, "tab v501 v025 [iweight=v005/1000000]" in that file, the top line (rounded to the nearest integers) will give 4709 and 12149.

Subject: Re: Number of respondents mismatch in dataset and published report
Posted by Soleil Von on Mon, 15 May 2017 04:19:27 GMT

Hi,I also have the same question.I am using Bangladesh DHS survey in 2014 for one of my projects. I am using Individual Record(BDIR70DT). I am writing for difference between the results I summaried based on IR files and the DHS final reports.I will really appreciate your help in clarify my doubts.
  There are 17863 women in IR files.The final report shows 5047 women of them live in urban area.(P29,table 3.1) However,I analysis there are 6167 urban women using R. The number of every residence is also different.
  Now I know maybe I don't use weights to analysis,but I don't know what the weight is and how to use it.could you explain it for me?
  Thank you very much for your time.
Best regards,
Soleilvon

Subject: Re: Number of respondents mismatch in dataset and published report
Posted by Bridgette-DHS on Mon, 15 May 2017 15:22:20 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum and DHS Senior Research
Associate, Shireen Assaf:

The difference between the 6167 and the 5047 is completely due to the use of weights. Below I
will paste the results in Stata, first without weights and second with weights, using BDIR70FL.dta.
I do this with iweight but you can also do it with pweight, svyset, and svy.

To produce the weighted table in R, you would need to use the survey package. The following
lines should help you get started.

library(survey)
# install and load the survey package
options(survey.lonely.psu="adjust")
# to fix the issue with strata with single PSU
# adjust will center the stratum at the population mean

data$wt = data$v005/1000000
create the weight variable

mydesign<-svydesign(id=data$v021, data=data, strata=data$v023, weight=data$wt, nest=T)
# set the survey design using the uploaded data (named data), the cluster (v021), the strata
(v023) , and the weight (wt).
then use svymean or svytable to get the weighted proportions and frequencies of your variables.
 See the survey package documentation for further details.

File Attachments
1) tab-v025.jpg, downloaded 1079 times