Subject: Clubbing individual recode and mens recode file to calculate overall prevalence
Posted by sarizwan1986 on Fri, 12 Jan 2018 12:20:00 GMT
View Forum Message <> Reply to Message

Hi

I have been working on DHS datasets of India for quite some time now. The latest NFHS4 was released yesterday and i had crack at it.
I wanted to calculate the prevalence of any tobacco use among men, women and both combined.
I used the files IAMR71FL and IAIR71FL and mapped the common variables between them and created a separate combined file using the STATA append command.
The variables of interest were 463a thru 463g and sm609e sm609c (in the mens recode file) and s710c an s710e (in the individual recode file) and i created a single variable called anytobaccouse if the answer to any of the above was yes (1), else (0).

I calculated the proportions using the following STATA command:
proportion anytobaccouse [pweight=v005/1000000], over(sex)

The proportion was 44% in men, 6% in women (which are almost the same as reported in the India factsheet) but what i could not believe or understand is the proportion in both sexes combined which was 11% (which was not reported in the factsheet). I expected a proportion in the ballpark of half the sum of men and women values (more like (44+6)/2 ~ 25%).

I noticed that the sample size was about 700,000 for women and about 100,000 for men, which can explain the above 11% but i thought this was supposed to be corrected by applying the weights.
I have attached a file where i show the weighted and unwieghted numbers for your reference.

What am I missing?

1. Was combining the above mentioned files appropriate? if yes, does applying the weight correct for the sample size imbalances between men and women?
2. How do I get the combined sexes prevalence if combining the files is not appropriate?

Thanks for reading my query and really appreciate your time and effort.

File Attachments

1) dhs forum.docx, downloaded 462 times