Subject: Sampling weights with value zero
Posted by albena on Wed, 23 Oct 2024 21:17:10 GMT

Hi all,

I am using the sampling weights from the Pakistan 2006/2007 PR dataset. I am observing that around 55,000 of the household members have a sampling weight that is zero. I am very puzzled because this means that these observations are to be excluded from the analysis as their weight is zero. I also have another example with zero sampling weights, but this time in the IR file - Namibia 2013.

Could you clarify in what cases can the sampling weights in the PR and the IR file can be zero? And also how to handle these observations - should the sampling weights be set to a missing maybe? On this note, I was wondering whether from your experience there could be also missing sampling weights be it in the PR or the IR datafile?

Thank you!

Albena

Subject: Re: Sampling weights with value zero
Posted by Bridgette-DHS on Thu, 24 Oct 2024 19:03:49 GMT

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

The Pakistan 2006-07 survey DOES NOT have cases with v005=0. I believe you are looking at the Pakistan 2017-18 survey, which DOES have cases with v005=0. Two geographic areas are omitted. If you enter "tab v024, summarize(v005) means" you will see that GB (Gilgit Baltistan) and AJK (Ajad Jammu and Kashmir) are omitted. A special weight variable, sv005, applies to all areas, but estimates for All Pakistan should not include GB and AJK. The sample design is described in the final report and there have been several posts on the forum on this issue.

The Namibia 2013 survey included a subsample of women age 50-64. The age range for men (in the men's survey) was 15-64. For women age 50-64, v005 was set at 0 because the standard estimates are based on that age range. There is another weight variable, sweightw, which applies to the age range 15-64. The sample design is described in Appendix A of the final report.

Subject: Re: Sampling weights with value zero
Posted by albena on Fri, 25 Oct 2024 15:28:07 GMT

Thank you very much for the clarifications!

I would like to still go back to the Pakistan 2006/07 DHS as my question was on it, I was just not complete in my explanation why I indicated you that there are women with a sampling weight of zero. In fact, these zeros appear because of zeros in the hh member recode (PKPR53FL.dta). I merge the PR with the IR recode (PKIR53FL.dta) as the women data is based only on the ever-married women, but I also need the never married. When I do the merge between the PR and IR file, it works perfectly and I merge all 10,023 ever married women in the IR data. What I do is to assign the hv005 as a sampling weight for the never married women as they don't have a v005. This is how I got to the 55,000 women that have a value of zero for their sampling weight. So, I went back to check the hv005 in the PR data and what I saw is that 657,364 observations out of 727,493 have hv005 =0.

So, I was wondering why is this. I also checked in the report, but there seems to be no information on this. What I read though is that there was a long and a short household questionnaire, the long one was used to identify the eligible women, men and children (?). I thought that this could be the reason why I see these zeros as the PR file I am using might be not the one I need to use to merge with the IR data. Do you have more information on these two household questionnaires and if this could be the reason for the zeros I see in the hv005? In any case the reason should be different that the reason for the zeros in the 2017-18 data as there it was clear that two regions had to be excluded.

Thank you!

Albena

Here also my code for merging the PR and IR data:

use "dir/PKPR53FL.dta", clear

keep if hv104 ==2

keep if hv105 >=15 & hv105 <=49

// keep if hv103 ==1 // not considered for now

rename hv001 v001
rename hv002 v002
rename hvidx v003

replace hv005 = hv005/1000000

sort v001 v002 v003

```
save "dir/PakistanPR2006.dta", replace


use "dir/PKIR53FL.dta", clear

keep v001 v002 v003 v005 v006 v007 v009 v010 v012 v014 v016 v024 v025 v135 v149 v190
v211 v503 v505 v507 v508 v509 v510 v511 b1_01 b2_01

save "dir/PakistanIR2006.dta", replace


use "dir/PakistanPR2006.dta", clear

merge 1:1 v001 v002 v003 using "/PakistanIR2006"
```

Subject: Re: Sampling weights with value zero
Posted by Bridgette-DHS on Fri, 25 Oct 2024 16:57:19 GMT
View Forum Message <> Reply to Message

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

This was an "ever-married women" (EMW) survey. Other countries that restrict the women's interview to ever-married women are Bangladesh and Jordan. In such surveys, all women are included in the household survey, but if their marital status in the hh survey is reported as never-married, then they are not eligible for the women's interview. You will find much discussion of this issue on the forum if you search for "awfact" or related terms. Also see the Guide to DHS Statistics (https://www.dhsprogram.com/Data/Guide-to-DHS-Statistics/inde x.cfm).

The EMW distinction is completely different from other kinds of subsampling and the special treatment of the disputed regions in Pakistan.

DHS discourages the limitation to EMW, which was more common many years ago than it is now, but some countries still feel that it would be inappropriate to ask never-married women even the most basic screening questions about fertility, fertility preferences, or family planning that are in all-women surveys.

In an EMW survey we usually simply report results as being limited to ever-married women. Mixing the never-married women back in with the ever-married women, with a weight such as hv005, is a risky thing to do and you will not match any DHS results if you do that. For some outcomes, DHS will multiply v005 for the EMW by (awfactx/100), but only for aw factors awfactt, awfactu, awfacte, etc. (look them up). I recommend that you simply report results as being limited to ever-married women, even if there is a loss of comparability with other countries.