Subject: Weithted analysis in R
Posted by Mahir on Tue, 23 Jul 2024 14:22:47 GMT
View Forum Message <> Reply to Message

Dear DHS team,

I have one more question regarding analysis of data using sampling weights.

I am using weighted data for analysis in R. I use r and I am using tbl_svysummary() in R to generate weighted table, how ever after accounting for weight using the following formula I get a weird error

library(survey)
tb1<-svydesign(ids = ~v021,weights = ~wt,data = tb_data, strata = ~v022, nest= T) %>%
tbl_svysummary(include = c(age,v025,WI),statistic = list(all_continuous() ~"{mean} ({sd})",all_categorical() ~ "{n} ({p}%)"), digits = all_continuous() ~ 2, label = c(age~"Age",v025~"Residence",WI~"Wealth Index"))

I am running this command and I have checked the syntax is ture. However I encounter the following error:

"Warning message:
There were 2 warnings in `mutate()`.
The first warning was:

Caused by warning in `svymean.survey.design2()`:
! Sample size greater than population size: are weights correctly scaled?

I am using the DHS_Benin data and I will do the same for seven other counties (Cameroon, Cote d'Ivoire, Ghana, Kenya, liberia, Nigeria and Uganda). Could you help me with this, it would be a big help. I have posted the question on DHS user forum already. Thank you!

Best
Mahir

Subject: Re: Weithted analysis in R
Posted by Janet-DHS on Fri, 26 Jul 2024 18:57:42 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Ali Roghani:
This warning likely stems from how the weights are currently scaled in your dataset. To eliminate this warning, you can adjust the weights column so that it represents the total population size rather than just the number of survey respondents.

Here's a step-by-step guide to making this adjustment:

First, define the total population size. Replace with your actual population size:
total_population <- XXXXX

Next, scale the weights in your dataset:
tb_data$wt_scaled <- tb_data$wt * total_population /sum(tb_data$wt)

Finally, create your survey design object using the scaled weights:
 tb1 <- svydesign(ids = ~v021, weights = ~wt_scaled, data = tb_data, strata = ~v022, nest = TRUE)

Using the adjusted weights in your code should remove the warning.

The best source of estimates of population size for different countries and years is the U.N. Population Division:https://population.un.org/wpp/.

Please let us know if this does not solve the problem or if you have other questions.

Subject: Re: Weithted analysis in R
Posted by Mahir on Mon, 05 Aug 2024 13:29:34 GMT
View Forum Message <> Reply to Message

Dear DHS team,

Thank you for your response.

Using the method you suggested removed the warning but the new scaled weight varialbe I created is actually in four digits for each observation. I have values of wt_scaled like (4061.151, 3859.654 etc). Is this normal or should I divide this by 1000?

Additionally when I create the table using tbl_svysummary() the overall sample size of the data increases exponentially. I am using Benin 2018 DHS KR data and the number of obersation increase from 3937 to total population that I have used for scaling the weight (14.080.072).

What should I do in this case?

Best
Mahir

Subject: Re: Weithted analysis in R
Posted by Janet-DHS on Wed, 07 Aug 2024 16:54:43 GMT
View Forum Message <> Reply to Message

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

You have re-scaled the weights so that the total weight for the sample matches the "total_population" that you entered. We thought that was what you wanted to do. If you want to do something else, please let us know what you want to do.

Subject: Re: Weithted analysis in R
Posted by Mahir on Thu, 08 Aug 2024 13:29:44 GMT
View Forum Message <> Reply to Message

Dear DHS team,

Thank you for your response.

In my case, I am using DHS data from eight countires, Benin 2017-18, Cameroon 2018, Ivory Coast 2021, Ghana 2022, Liberia 2018-19, Kenya 2022, Nigeria 2021 and Uganda 2016. I am interested in the IYCF indicators from KR file for children aged 0-23 months.From the nutrition data I have created a diet index, I want to look at the association of diet index with wealth index, mother's education, place od residence etc. I want to do this using the weighted data. I create a sub dataset from the KR file that contains only children aged 0-23 months. As I have already mentioned, when I use scaled weight, I get a talbe with total population of Benin. Should I only use total population of children aged 0-23 months to scale the weight ? In case of Benin, for example, there are total of 3937 children aged 0-23 months , when i use the weights togenerate a table, the table shows 2933 children (not the scaled weight, original weight variable). This should not be the case. Do you have any advise for this?

Secondly, I would also like to create a new population level weight by denormalzing the weight for the pooled data of all eight countires. Just to make camparasions between countires. Can you share tell me how I can do this? I am using R for data analysis. I can figure out the code if you can could just explain me how to conceptually do this. But if you can share the R code that would be also great. :)

Best
Mahir

Subject: Re: Weithted analysis in R
Posted by Bridgette-DHS on Fri, 09 Aug 2024 13:35:37 GMT
View Forum Message <> Reply to Message

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

You are making a bigger deal out of the weights than necessary. You are looking for some multiplier to re-scale the weights, but almost all DHS estimates--rates, means, proportions, regression coefficients, etc., are invariant with respect to such a multiplier. You can confirm this by, say, running an analysis with weight v005, as given in the data, and then multiplying v005 by 2 or 10 or 100, and re-running. Nothing will change, not even the standard errors.

The only exception would be if you are trying to estimate, say, the NUMBER of children age 0-23 months at the national level who were never breastfed. It that's what you want to do, you should re-scale to the population number of children age 0-23 months. You could get an estimate of that population number from Population Prospects 2024.

Weights for that purpose are called expansion weights. DHS reports, in both final reports and research reports, so far as I am aware, have never used expansion weights.

If you want to make pooled estimates, for example for "West Africa" (I put this in quotes because there are alternative lists of countries and in any list DHS has not had surveys in all countries) then you could multiply v005 by P/p, where P is the national population and p is the sum of the weights in the sample (without the factor of 1 million). But I would strongly recommend against such pooled estimates, because the pooled population is never well-defined and the results will be dominated by the largest country, such as Nigeria, obscuring differences between countries. If the goal of your pooling is to look at DIFFERENCES between countries, which is common, then you do not need to adjust the weights at all.

There have been many postings on weights on the forum. We have nothing to add to what has already been stated.

Subject: Re: Weithted analysis in R
Posted by Mahir on Wed, 14 Aug 2024 14:33:40 GMT
View Forum Message <> Reply to Message

Dear DHS team,

Thank you for your response.

I promise might be the last time I come back with question regarding sampling weight.

the reason why I am asking for denormalising weight because I want to do poolsed analysis of eight countires I have mentioned. I have gone through almost all the questions on sampling weight on this forum now. I also found a manual by DHS (attaching here) which recommends denormalizing weights for doing pooled analysis using the formula given. I using the exact same formula that the manual is sharing. I will share the example of DHS Benin KR recode (2017-18).

I apply fthe fooolwing code in

KRdata$denorm<-(KRdata$v005 * (2717666/13589)))/1000000
#2717666 are the total number of woemn aged 15-59 at the time of the survey. I got this data from this site (https://platform.who.int/data/maternal-newborn-child-adolesc
ent-ageing/indicator-explorer-new/MCA/women-of-reproductive-
age-(15-49-years)-population-(thousands)) and 13589 is the total number of respondents in the KR file

Now ideally with the sum of all the values of the new weight (denrom) should be 2717666 but that is not the case, the total comes out to be 2809377, this is a huge discrepency.

Would you be able to explain why this mgiht be happening? Is there a way to resolve this? I am afraid this same problem will occur with rest of the seven countires (Cameroon 2018, Cote d'Ivoire 2021, Ghana 2022, Kenya 2018-19, Liberia 2022, Nigeria 2021 and Uganda 2016)

Best
Mahir

File Attachments
1) Note+on+de-normalization+of+DHS+standard+weight.pdf,
downloaded 267 times

Subject: Re: Weithted analysis in R
Posted by Bridgette-DHS on Thu, 15 Aug 2024 14:09:40 GMT
View Forum Message <> Reply to Message

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

The weight of a child in the KR file is the same as the weight of the mother in the IR file. The cases in the KR file are not respondents, but are the births in the past 5 years to the women in the IR file.

If you are rescaling to the number of women in the population, then you are basically rescaling the weights in the IR file, NOT the weights in the KR file

For this survey, the total number of women in the IR file is 15928. The weights in the IR file, v005, are scaled so that their total matches 15928 (with a factor of 1000000).

You want to rescale so that the sum of the weights in the IR file would be 2717666. You do this by multiplying v005 by 2717666/15928.

If you multiply v005 in the KR file by this same ratio, the total will be (approximately) he number of births in the past 5 years to the women in the population. That's what you would want it to be. Note that the KR file includes children who died, as well as children who survived.

Let us know if you still have a problem.