Subject: DHS ZIMBABWE 2015
Posted by p.pajak on Wed, 29 Dec 2021 11:13:26 GMT

Good Morning,

I am in the process of analysing the 2015 DHS from Zimbabwe. I am using R (I am quite new to the programme) and I am currently stuck as my DEFFs are all coming out negative (it should not happen in my case) or not at all.
I am analysing the data from 1 region only and I already did the cleaning and subset of my data.
I set my survey design following the instructions I found online from other users:
DHSdesign<-svydesign(id= ~fin_df$PSU,   #V021
            strata= ~fin_df$STRATA,   #V022
            weights= ~fin_df$PERWEIGHT, #V005
            data=fin_df)

Now, I want to calculate the DEFF for the variable C_SEX2 from my fin_df dataframe and I type the following:
DEFF<- svymean(~fin_df$C_SEX2, design=DHSdesign, na = TRUE, deff = TRUE)

the output is the following:
Warning message:
In svymean.survey.design2(~fin_df$C_SEX2 == 1, design = DHSdesign, :
  Sample size greater than population size: are weights correctly scaled?

```
              mean      SE DEff
fin_df$C_SEX20 0.548656 0.021449   NA
fin_df$C_SEX21 0.451344 0.021449   NA
```

I also tried:
DEFF1 <- svytotal(~fin_df$C_SEX2, design=DHSdesign, na = TRUE, deff = TRUE)
```
            total    SE    DEff
fin_df$C_SEX20 144.736  12.150 -5.2710
fin_df$C_SEX21 119.065   9.295 -3.0851
```

Can anyone see a mistake or advise me?

Thank you very much in advance :)

Best wishes,
Patrizia

Subject: Re: DHS ZIMBABWE 2015
Posted by Bridgette-DHS on Tue, 04 Jan 2022 02:13:06 GMT

Following is a response from DHS Senior Sampling Specialist, Mahmoud Elkasabi:

To calculate DEFF with R, using the svydesign function, the weights must first be divided by 1,000,000. To analyze data for one region (or another subpopulation), the subset function should be used. Here is the R code:

```
DHSdesign<-svydesign(id= ~fin_df$PSU, #V021
strata= ~fin_df$STRATA, #V022
weights= ~fin_df$PERWEIGHT, #V005/1000000
data=fin_df)

DHSdesign_sub <-subset(DHSdesign,V024=xx)

# the DHSdesign_sub should be now used instead

DEFF<- svymean(~fin_df$C_SEX2, design=DHSdesign_sub, na = TRUE, deff = TRUE)
```

Subject: Re: DHS ZIMBABWE 2015
Posted by p.pajak on Fri, 07 Jan 2022 18:08:06 GMT
View Forum Message <> Reply to Message

Good Evening,

Thank you for providing me with the R code. My weights have already been divided by 1000 000 as I am using IPUMS DHS data and I have done the subset by the region of interest in fin_df. I think the issue is that r considers this survey design with replacement if I give the command as written above and therefore by putting DEFF=TRUE doesn't work. It only works if I write DEFF=replace.

Could you please let me know if this is correct.

Many thanks!

Subject: Re: DHS ZIMBABWE 2015
Posted by p.pajak on Fri, 07 Jan 2022 20:30:33 GMT
View Forum Message <> Reply to Message

Just to clarify further my answer please see below my code:
```
DHSdesign<-svydesign(id= ~fin_df$PSU,
          strata= ~fin_df$STRATA,
          weights= ~fin_df$PERWEIGHT,  # the weights have already been divided by
1000000
          data=fin_df)              # fin_df has already been subset by region of interest
```

This is the outcome I obtain by calling the summary of my survey design
summary(DHSdesign)

Stratified 1 - level Cluster Sampling design (with replacement)
With (36) clusters.

Therefore I need to use DEFF = "replace" to obtain the DEFF as below:
svymean(~fin_df$C_SEX2, design=DHSdesign, na = TRUE, deff = "replace")

I read in another comment that to change the design to "without replacement" I would need to
specify the fpc which is not provided by DHS and therefore doesn't need to be included.
Hope this helps clarify my answer! Thank you very much for the assistance.

Subject: Re: DHS ZIMBABWE 2015
Posted by Bridgette-DHS on Mon, 10 Jan 2022 14:50:22 GMT
View Forum Message <> Reply to Message

Following is another response from DHS Senior Sampling Specialist, Mahmoud Elkasabi:

Yes, this is correct, deff="replace" should be used instead.