Posted by MiFoo on Sat, 17 Apr 2021 01:27:14 GMT
View Forum Message <> Reply to Message

Hi everyone,

I am using the DHS from 2011 to assess diabetes prevalence and BMI. However, I found a very high number of missing BMI values. Based on the final report (p244/245), there should be 3812 women + 3721 men=7533 people with nonmissing/valid BMI. In contrast to this, I obtained only 5216.7 valid measurements. The numbers per BMI category also deviate, particularly for women. Note that I restricted the analysis to those with valid blood glucose measurements as this is case in the table of the final report.

This is my code (R), with sdata being the survey design object:

```
sdata11 %>%  filter(sh284a<994 &shbm<9998) %>%
summarize(total=survey_total(!is.na(shbm)))

sdata11 %>% filter(sh284a<994) %>%
   mutate(BMI_cat=case_when(shbm<1850~"underweight",
                shbm<2500~ "normal",
                shbm<3000~ "overweight",
                shbm<9998 ~ "obese")) %>%
  mutate(BMI_cat=factor(BMI_cat, levels=c("underweight", "normal", "overweight", "obese"))) %>%

        group_by(BMI_cat, sex) %>% summarize(n=survey_total())
```

	BMI_cat	sex	n	n_se
	<fct>	<fct>	<dbl>	<dbl>
1	underweight	men	1050.	37.0
2	underweight	women	582.	26.3
3	normal	men	2234.	48.0
4	normal	women	808.	30.5
5	overweight	men	303.	18.6
6	overweight	women	169.	15.1
7	obese	men	24.8	5.20
8	obese	women	45.7	7.76
9	NA	men	108.	14.7
10	NA	women	2218.	49.9

Height and weight are available for more people but not their combinations, so that a new calculation of the BMI leads to the same number of missing values and is not the reason for this deviation. I also did the same calculations for the DHS from 2017. For this survey, the numbers more or less match those in the final report. Therefore, I think my coding should not be the explanation for the missing values.

What can be the reason for this? Am I using a wrong variable?

Best wishes,
MiFoo

Hello,

You would need to code BMI within what is considered valid responses which are only values between 1200 and 6000. For women, we also exclude pregnant women and women with a birth in the last two months. Using our standard DHS code on GitHub in Chapter 11 (NT_WM_NUT) code file (https://github.com/DHSProgram/DHS-Indicators-Stata) I was able to match the Final report table 11.10.1 that reports the nutritional status for women. This is using the IR file. We do not have code yet for this chapter in R. However, using the same logic in the Stata or SPSS do file you can see that we only consider values in the range of 1200 to 6000 for both men and women. Perhaps this is the issue.

Thank you.

Best,
Shireen Assaf

Hi,
thank you very much for your reply. I was talking about table 15.5.1, as my anaylisis only includes respondents with both BMI and glucose measurements. I was not able to find the corresponding code on GitHub. Is it available and could you please provide a link?

I was using the PR, not the IR file to calculate the results in my last message, since the former does not include glucose measurements. And yes, I excluded BMI values coded as 9998. This resulted in valid BMI values (variable shbm) for 5217 people(3612 men and 1605 women) who also have valid glucose measurements. Based on the final report, it should be 3812 women and 3721 men=7533. Hence, I have much MORE missing BMI values.

I also tried to match the BMI values in the IR file (v445) to women in the PR files who have missing BMI values. This gave me 3642 nonmissing BMI values but still not the desired 3812 in the table 15.5.1.

Further advice would be great!
Sarah

Subject: Re: Missing BMI values in the DHS from 2011
Posted by Shireen-DHS on Fri, 23 Apr 2021 12:47:58 GMT
View Forum Message <> Reply to Message

Hello,

We do not have code on GitHub for glucose measurements.

To code the BMI in the PR file you can use the following CSPro code.

Best,
Shireen

```
  preg = 0;
    if SH231 = 2 then
      if pidx(SH230,2) <> 0 <=> SH234 >= 50 then errmsg("Age in HH=%d Age in
SH234=%d",HV105(SH230),SH234) endif;
      if pidx(SH230,2) <> 0 and HA65(pidx(SH230,2)) = 1 then
        if HA54(pidx(SH230,2)) = 1 then preg = 1; endif;     { currently pregnant for completed
interview }
      else
        if SH231 = 2 and SH232 = 1 then preg = 1; endif;     { if woman is not eligible/incomplete
interview }
      endif;
    endif;

{ BMI }
    if SH19(SH230) <> 0 then                 { If weight and height measurements were in the
women/men section }
      if SH231 = 2 and HA13(pidx(SH230,2)) = 0 then       { if woman measured }
        xbmi = HA40(pidx(SH230,2));
      elseif SH231 = 1 and HB13(pidx(SH230,1)) = 0 then   { if man measured }
        xbmi = HB40(pidx(SH230,1));
      endif;
    else
      if SHWH = 0 then xbmi = SHBM; endif;     { if weight and height measurements in
Biomarkers section }
    endif;

    recode xbmi    => nutrstat;
      1200-1849 => 0;     { Thin     }
      1850-2499 => 1;     { Normal    }
```

```
        2500-2999 => 2;     { Overweight }
        3000-6000 => 3;     { Obese     }
               => 9;
    endrecode;
    if SH234C in 35:49 and SH231 = 2 and preg = 1 then nutrstat = 4; endif;
```

Subject: Re: Missing BMI values in the DHS from 2011
Posted by MiFoo on Mon, 26 Apr 2021 05:51:36 GMT

View Forum Message <> Reply to Message

Hello Shireen,

thank you once more for your help! To be honest, I still have not managed to get to the results in the table of the final report. I have no experience with CSPro but I tried create a new BMI variable using similar code in R:

```
dataPR$BMI_new3 <- case_when(sh19!=0 & sh231==2 & ha13==0 ~ ha40,
               sh19!=0 &sh231==1 & hb13==0 ~ hb40,
               shwh==0 ~ shbm,
               TRUE~ NA_integer_)
```

Using
```
sdataPR %>%  filter(BMI_new<9998 & (sh284a<994 | sh259==1)) %>%
  group_by(hv104) %>%
  summarize(total=survey_total())
```
to restrict the analysis to those with both valid glucose or current diabetic medication and valid BMI, I ended up with 3612 observations for men and 3691 observations for women. Based on the final report (table 15.5.1 and 15.5.2 on p244/5), it should be 3812 women and 3721 men. Therefore, I still have too few observations for BMI (and these numbers still include pregnant women).

These are the estimated observations per category:

BMI_cat	hv104	n	n_se
<fct>	<int>	<dbl>	<dbl>
1 underweight	1	1050.	37.0
2 underweight	2	1088.	36.7
3 normal	1	2234.	48.0
4 normal	2	1949.	44.9
5 overweight	1	304.	18.6
6 overweight	2	521.	28.9
7 obese	1	24.8	5.20
8 obese	2	133.	12.8
9 NA	1	108.	14.7
10 NA	2	132.	13.4

Do you have any idea what my mistake might be? I ignored the pidx command in your code as I

don't know what exactly it does. I think referring to the household index is not necessary in R?

Secondly, I tried to create an indicator for pregnant women using
dataPR$preg <- case_when(dataPR$ha54==1 | dataPR$sh232==1 ~ 1,
                TRUE ~ 0)
This gave me 1140 pregnant women (unweighted) and 33 of these have valid glucose measurements and BMI. Strangely, all of these are above 50. According to table 15.5.1, there should be 6 pregnant women. Does the table assume a different cutoff than 50 to define a reasonable age for pregnant women?

Best,
Sarah

Subject: Re: Missing BMI values in the DHS from 2011
Posted by Shireen-DHS on Tue, 27 Apr 2021 12:30:30 GMT
View Forum Message <> Reply to Message

Hello Sarah,

Are you applying sampling weights? You can use the survey package in R to do that. The weight would be wt=hv005/1000000 .

Best,
Shireen

Subject: Re: Missing BMI values in the DHS from 2011
Posted by MiFoo on Thu, 29 Apr 2021 15:37:02 GMT
View Forum Message <> Reply to Message

Hello Shireen,

yes, I am using the survey package. I am able to reproduce the results in table 15.5.1 of the fianl report from 2011 for other variables than BMI (like education).
Besides, I can reproduce 13.5.6 in the final report from 2017, including the number of observations per BMI category.

Therefore, I think the problem is really the BMI variable in my dataset from 2011.
Do you have any other idea what might cause the higher number of missing values? Are some values imputed?

It would also be great if you commented on my question regarding the exclusion of pregnant women in my last message.

Best
Sarah

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

So-called "missing" cases are usually "Not Applicable" or "NA" cases. If a variable has a dot as a code (in Stata) it is NA, not missing. There should be other codes such as 9999 for refusals, for example. Sometimes extreme values are included and sometimes they are given a 9999 type of code. Sometimes extreme values are omitted during data processing, with arbitrary limits that are not specified anywhere, but this is rare. Such exclusions would usually be specified in the CSPro code. Imputations of biometric values are never made, so far as I know, although if there are multiple measurements, as with blood pressure, there are rules for coming up with a single value.

The analysis team at DHS cannot put further effort into this issue. We have sent you the relevant CSPro code. You have the standard recode files. We have nothing more to work with than you do. It is possible that there is an inconsistency or error in the CSPro program or in the data files. When we do our own reports, the analysis team sometimes (this is rare) has to accept that we cannot match an earlier value. We then proceed with our own value. You may have to do the same.