
Subject: BMI

Posted by [Nomvelo](#) on Tue, 07 May 2019 18:35:54 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hi

I am doing research on Obesity and Occupation. Using South African DHS 2016. On my sample, over 40% of the observations does not have BMI value. Please assist me what is the most appropriate way to deal with this missing data. ? Should I exclude the observations or impute the missing values. If imputation is the option, kindly advise on the technique that will be be most appropriate.

Kindly assist, I have red some similar questions but I didn't get quite a fitting answer for my problem..

Thanks for your assistance.

Subject: Re: BMI

Posted by [Liz-DHS](#) on Fri, 15 Nov 2019 17:12:19 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear User, the best place to begin is with The Guide to DHS Statistics. If you still need assistance, please feel free to post again. Thank you!

Subject: Re: BMI

Posted by [Frankfeng](#) on Wed, 03 Jun 2020 15:57:16 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear Liz,

I am also doing research on Obesity using South African DHS 2016.

I have recoded BMI for males and females. Males' BMI is matched with the SADHS report. But females' BMI is not matched. In my analyses, the percentage of obese for female is around 35% no matter I used weight or not, but it is 41% in the SADHS.

In the SADHS report, it says the "sample: Women age 15+ who are not pregnant and who have not had a birth in the 2 months before the survey in the 2 months before the survey" (on page 298).

I read the "Guide to DHS Statistics", the process is: Number of women age 1549, excluding women who are pregnant or who gave birth in the 2 months preceding the date of the interview

I used "individual recode" dataset for my analyses, and the STATA code is:

```
recode v445(min/1849=1 "Underweight") (1850/2499=2 "Normal") (2500/2999=3 "Overweight")  
(3000/5956=4 "Obese") (9998=.),gen(BMI_cate_F)
```

```
tab BMI_cate_F if v213!=1 & (v208 == 0 | b19_01 >= 2) & (v445>=1200 & v445<=6000)
```

Could you please point out that which step I missed?

Thank you.

Subject: Re: BMI

Posted by [Liz-DHS](#) on Wed, 03 Jun 2020 16:12:17 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear User, Check out this link. It contains Stata code to produce Demographic and Health Survey Indicators. Chapter 11 should provide guidance with your code. Thank you!

Subject: Re: BMI

Posted by [Frankfeng](#) on Thu, 04 Jun 2020 09:30:51 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear Liz,

Thank you very much for the link. It is very helpful.

I have checked the code, and compare the results. The percentage of obese from the code is 34.7% and I got 34.25% from my code. But both are not the same as the SADHS reported: 41% female is obese.

I think the differences would be the sample size. In SADHS it says:"sample: Women age 15+ who are not pregnant and who have not had a birth in the 2 months before the survey in the 2 months before the survey" (on page 298).

From the "Guide to DHS Statistics", the process suggests: Number of women age 1549, excluding women who are pregnant or who gave birth in the 2 months preceding the date of the interview

I also tried to limit the sample as v213!=1 & (v208 == 0 | b19_01 >= 2) & (v445>=1200 & v445<=6000) , but it doesn't work.

Could you please guide me where or how I can set such sample size for Figure 17.1 in the SADHS report?

Thank you.

Subject: Re: BMI

Posted by [Trevor-DHS](#) on Mon, 08 Jun 2020 21:34:12 GMT

[View Forum Message](#) <> [Reply to Message](#)

Calculating the BMI to match the South Africa report figure 17.1 is unfortunately not that simple. In this survey the BMI is calculated for women age 15+, not just woman age 15-49 (which is the standard approach used by DHS). In this survey an additional Adult Health module was included in a sample of roughly half of the households. The data for the adult health module can be found in the file ZAAH71DT.zip. In this file you will find two datasets - one for women (ZAAHW71FL.dta) and one for men (ZAAHM71FL.dta). You will need to use the women's file and link it to the IR file, and then again to the PR file to produce the estimate that you want. See the code below:

```
cd "your working directory"
```

```
* use the Adult Health module data
```

```
use "ZAAHW71FL.DTA", clear
```

```
* merge in data from IR file to get v213, v208 and b19_01 that are needed below
```

```
merge 1:1 caseid using "ZAIR71FL.dta"
```

```
rename _merge merge1
```

```
* create a household id and line number for matching to the PR file
```

```
gen hhid = substr(caseid,1,12)
```

```
gen hvidx = v003
```

```
* merge the PR file into the Adult Health module data
```

```
merge 1:1 hhid hvidx using "ZAPR71FL.DTA"
```

```
drop if _merge != 3
```

```
* calculate obesity for women who are not pregnant and have not had a birth in the prior 2 months  
(and have a valid BMI)
```

```
* for women older than 49 the assumption is that they are not pregnant and have not had a birth in  
the past 2 months
```

```
gen obese = (ha40>=3000 & ha40<=6000) if v213!=1 & (v208 == 0 | b19_01 >= 2) &  
(ha40>=1200 & ha40<=6000)
```

```
* tabulate obesity for the total sample of women and for women 15-49 alone.
```

```
tab obese [iw=sweight/1000000]
```

```
tab obese [iw=sweight/1000000] if v012 < 50
```

You will notice also that a different weight has been used here (sweight). This is a weight used for the subsample for the Adult Health module.

Subject: Re: BMI

Posted by [Frankfeng](#) on Tue, 09 Jun 2020 10:27:39 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear Trevor,

Thank you very much indeed for the coding, it works now from the code "tab obese [iw=sweight/1000000]" but not from the "tab obese [iw=sweight/1000000] if v012 < 50" (still 35% of females are obese). I don't know why, but at least I know how now.

May I ask two more questions?

1. In the "ZAIR71FL.dta", it already has "v445" for "body mass index", I was wondering why we cannot calculate obese directly in "ZAIR71FL.dta"? or because "ZAIR71FL.dta" does not have the sweight?

2. It seems most biomark information are stored in "ZAPR71FL.DTA". If I want to get the same results (Hypertension, BMI, anemia, or diabetes) as those shown in the SADHS report, do I need to merge the PR file into the Adult Health module data, and then do the analyses with sweight? Could you confirm that tables or figures in SADHS report are weighted? Or I can know it is weighted/unweighted?

Thank you very much.

Subject: Re: BMI

Posted by [Trevor-DHS](#) on Tue, 09 Jun 2020 17:13:26 GMT

[View Forum Message](#) <> [Reply to Message](#)

I think you are missing looking at table 17.1.1 which provides the data for figure 17.1. The bottom row of this table provides the percentage for women 15-49 (35.9), which is what I reproduced.

1. You can use just the ZAIR71FL.dta and restrict your analysis to women 15-49, and use v005 for the weight variable. You won't get exactly the same number as in the table, but it will be very, very close (differing just because of the separate weight variable).

2. If you want to match the results in the report for women or men age 15+ (including those age 50 and over), then, yes, you will likely need to perform similar merges of the Adult Health data, the IR file and the PR file, as shown for the obesity data, and using sweight as the weight variable. In some cases, though, it may be sufficient just to use the PR file alone (in this case using hv005 as the weight variable). I would suggest testing to ensure you can match the denominator first for each indicator, using both methods.
