Subject: Stunting Discrepancies with SL Reports
Posted by ADoggett on Thu, 20 Jan 2022 17:55:36 GMT
View Forum Message <> Reply to Message

Hi all,

I am working with data from Sierra Leone in 2008 and 2013, and I am having trouble with my stunting prevalence rates as they differ from the report. I will focus on 2008 for this post because the issue seems to the similar for both years.

For Sierra Leone 2008, I get a stunting prevalence of 33.6% (should be 36.4% according to the report) and a severe stunting prevalence of 19% (should be 20.6% according to the report).

Some important notes:
- I am using the PR file to calculate the rates as opposed to the KR file (I am aware this is a common error when calculating anthropomorphic indicators in earlier DHS years)
- I have made sure to specify that individuals must have slept in the house the night before the survey (i.e. hv103==1)

I have attached my R script for these indicators - I am hoping someone can help shed some light on why my numbers are different from the report!

Thanks,

Amanda

File Attachments
1) StuntingCode_DHSForum.R, downloaded 442 times

Subject: Re: Stunting Discrepancies with SL Reports
Posted by Shireen-DHS on Thu, 20 Jan 2022 18:51:13 GMT
View Forum Message <> Reply to Message

Hello Amanda,

Using the PR file is correct for anthropometric indicators. Can you try the following R code below? We will be posting the R code for all nutrition indicators on our GitHub site soon. Here is what we have available in R so far (https://github.com/DHSProgram/DHS-Indicators-R). We have all the indicators coded in Stata and SPSS if you would like to check the code there as well.

Thank you.
Best,
Shireen Assaf
The DHS Program

libraries
library(tidyverse) # most variable creation here uses tidyverse

```r
library(haven)      # used for Haven labeled DHS variables
library(labelled)   # used for Haven labeled variable creation
library(expss)      # for creating tables with Haven labeled data
library(naniar)     # to use replace_with_na function
library(xlsx)       # for exporting to excel

# //Severely stunted
PRdata <- PRdata %>%
  mutate(nt_ch_sev_stunt =
       case_when(
         hv103==1 &  hc70< -300  ~ 1 ,
         hv103==1 &  hc70>= -300 ~ 0 ,
         hc70>=9996 ~ 99)) %>%
  replace_with_na(replace = list(nt_ch_sev_stunt = c(99))) %>%
  set_value_labels(nt_ch_sev_stunt = c("Yes" = 1, "No"=0  )) %>%
  set_variable_labels(nt_ch_sev_stunt = "Severely stunted child under 5 years")

# //Stunted
PRdata <- PRdata %>%
  mutate(nt_ch_stunt =
       case_when(
         hv103==1 &  hc70< -200  ~ 1 ,
         hv103==1 &  hc70>= -200 ~ 0 ,
         hc70>=9996 ~ 99)) %>%
  replace_with_na(replace = list(nt_ch_stunt = c(99))) %>%
  set_value_labels(nt_ch_stunt = c("Yes" = 1, "No"=0  )) %>%
  set_variable_labels(nt_ch_stunt = "Stunted child under 5 years")

PRdata <- PRdata %>%
  mutate(wt = hv005/1000000)

table_temp <-  PRdata %>%
  calc_cro_rpct(
    cell_vars = list(hc27, hv025, hv024, hv270, total()),
    col_vars = list(nt_ch_sev_stunt, nt_ch_stunt),
    weight = wt,
    total_label = "Weighted N",
    total_statistic = "w_cases",
    total_row_position = c("below"),
    expss_digits(digits=1)) %>%
  set_caption("Child's anthropometric indicators")
write.xlsx(table_temp, "Chap11_NT/Tables_nut_ch.xls", sheetName = "child_anthro",
append=TRUE)
```

Subject: Re: Stunting Discrepancies with SL Reports

Hi Shireen,

Thank you for your quick reply.

This code did not solve the error, but it has helped me solved the error and will share!

First, the reason the provided code did not produce correct %s (it lowered stunting prevalence to about 30% for SL 2013, which is 6% off what it is in the report) is the case_when function. Case_when executes each line in order, so your code needs to have "hc70>=9996 ~ 99" first, as below:

case_when(
hc70>=9996 ~ 99,
hv103==1 & hc70< -200 ~ 1 ,
hv103==1 & hc70>= -200 ~ 0
))

The way it was previously written (where hc70>=9996 ~ 99 was the last line) would categorize anyone with a value at or above 9996 who also slept in the house the night prior as a '1' because 9996 is larger than (-200), so the first line gets evaluated as TRUE for those people. Once these people are categorized as a '1' based on the execution of the first line, they won't be recategorized because case_when works similar to nested if/else statements (i.e. once someone is put into one group, they won't be put into a different group even if they meet the criteria.) I hope that helps with your R code for the GitHub repository!

For those reading, why my code was not matching was because of a very simple error. I had categorized missing values as below:

PR$hc70_n=ifelse(PR$hc70 %in% c(9996,9997,9998),NA_real_,PR$hc70)

According to the DHS labels in my dataset, 9996, 9997, and 9998 are supposed to correspond to 'height out of plausible limits', 'age in days out of plausible limits', and 'flagged cases', respectively, so this code should work. However somewhere the labels maybe got mixed up, or a label was missed, because 9997 does not exist in the dataset, but 9999 does, which my code was missing. If I change my code to recategorize missingness to:

PR$hc70_n=ifelse(PR$hc70 %in% c(9996,9997,9998,9999),NA_real_,PR$hc70)

Or more simply,

PR$hc70_n=ifelse(PR$hc70>=9996,NA_real_,PR$hc70)

Then it works no problem and I get #s that match the report for SL 2008 and 2013

Cheers,

Amanda

Subject: Re: Stunting Discrepancies with SL Reports
Posted by Shireen-DHS on Fri, 21 Jan 2022 13:17:45 GMT
View Forum Message <> Reply to Message

Hello Amanda,

I am glad you were able to locate your error.

The code I provided however does work. I just ran it for another survey to double check and match the report. I use replace_with_na function by first identifying children above or equal 9996 as 99 and then replace 99 as missing. Using ifelse is another way of course.

Thank you.

Best,
Shireen