Subject: Re: Stunting Discrepancies with SL Reports
Posted by ADoggett on Thu, 20 Jan 2022 21:43:55 GMT
View Forum Message <> Reply to Message

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