Subject: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Thu, 29 Nov 2018 02:40:37 GMT
View Forum Message <> Reply to Message

I am working with NFHS4 raw data. Many times my analysis is not matching with National Report. Tried to sort out this issue, discussed with various people, but couldn't succeed. For example the number given in the national report of children born in the two years preceding the survey who ever breasted itself is greater than the total number of children which includes both who breastfed and not. How come that number will be greater than this total number. This is only one example, there are many other issues as well. I have used weights and all, followed every statistical procedure explained by USAID people.

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by fredarnold on Fri, 30 Nov 2018 16:48:55 GMT
View Forum Message <> Reply to Message

In Table 10.4 in the national report, the second column shows the number of children born in the last two years (97,989). that's the denominator of the estimate, not the numerator. As the table footnote indicates, that number includes births in the last two years whether the children are living or dead at the time of the interview.

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Sat, 01 Dec 2018 02:37:55 GMT
View Forum Message <> Reply to Message

Dear sir
I am very gratitude andThanks for your response

If you are saying this number is (97,989) is the total number and the denominator, again it is not matching with our analysis, I have attached the stata code for the same, let me know whether I have made any mistakes.

These are my stata command used for

```
**********************************************************
clear all
use "/Users/muneerkalliyil/Desktop/UNICEF/NFHS4/IAKR74FL.DTA"

***APPLYING WEIGHTS***********************

*Survey Weights

des v005
```

gen survweight = v005/1000000
gen stratum = v023
svyset v021 [iweight=survweight], strata(stratum)

****AFTER APPLYING WEIGHTS, JUST WE ARE LOOKING HOW MANY CHILDRENS ARE THERE BELOW 24 MONTHS,

**total number of children under 24 months
svy: tab hw1 if hw1 < 24
****

The result says no of obs as 96782;

. svy: tab hw1 if hw1 < 24
(running tabulate on estimation sample)

Number of strata = 2513 Number of obs = 96782
Number of PSUs = 26472 Population size = 93053.216
Design df = 23959

NFHS4 National report gives the total number of children below 24 months as (97989), but our analysis shows as (96782)

Please let me know, waiting for your reply

Regards
Muneer

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by fredarnold on Sun, 02 Dec 2018 22:42:38 GMT
View Forum Message <> Reply to Message

It looks like you're only including living children. As indicated earlier, the denominator includes both living and dead children who were born in the last two years.

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Mon, 03 Dec 2018 03:53:52 GMT
View Forum Message <> Reply to Message

Dear Sir

Thanks for your reply

I will be very grateful if you could tell me how to include both living and dead children. As per my

understanding, the variable hw1 (age in months) includes both, I have gone through all other variables in Children file, but couldn't get a variable differentiating between dead and alive children. Please let me know if you know how to include.

Looking forward to hearing from you
Thanks in advance

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by fredarnold on Mon, 03 Dec 2018 21:25:04 GMT
View Forum Message <> Reply to Message

You should use the kids' file (IAKR74). To obtain the number of children born in the two years before the survey (including both living and dead children), use V008 - B3 < 24.

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Tue, 04 Dec 2018 03:36:02 GMT
View Forum Message <> Reply to Message

Dear Sir

very thankful to your replies

I tried using V008 - B3 < 24 for age, then I got the number of children below 24 as 101955. However, the national report says 95% of the total below 24 has been breastfed, if we take the 95% of this number (96857), again it is not matching with national report number (97989). Could you please let me know if I am making any mistakes

Looking forward to hearing from you
Regards

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by fredarnold on Tue, 04 Dec 2018 18:53:50 GMT
View Forum Message <> Reply to Message

You're using the unweighted data. You'll have to weight the data to match what's in the report. Once you weight the data, you should be able to match the total number of children born in the two years preceding the survey (97,989). As mentioned earlier, the number of children in the second column of Table 10.4 is the denominator (the total number of children born in the two years preceding the survey). It is not the numerator (the number of children born in the two years preceding the survey who were ever breastfed).

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Wed, 05 Dec 2018 04:28:46 GMT
View Forum Message <> Reply to Message

Dear Sir,

I have already applied survweight, tried both pweight and iweight. The number which I told earlier is coming after applying the weight. Secondly, If the number in the second column is denominator, then our estimated denominator is definitely coming larger than the given denominator. As I told before, in our estimation, the total number of children below 24 months is coming around 101955. However, the given number in the report is only around 97989.

Please let me know if I am making any mistakes.

Looking forward to hearing from you

Thanks

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by fredarnold on Wed, 05 Dec 2018 16:30:49 GMT
View Forum Message <> Reply to Message

The Stata code below matches the denominator in the NFHS-4 report:

```
gen age = v008  b3
gen xweight = v005/1000000
tabulate age if age < 24 [iw=xweight]
```

age	Freq.	Percent	Cum.
0	1,629.0776	1.66	1.66
1	3,670.6754	3.75	5.41
2	3,935.1411	4.02	9.42
3	4,126.8893	4.21	13.64
4	4,405.3671	4.50	18.13
5	4,464.9622	4.56	22.69
6	4,501.2567	4.59	27.28
7	4,415.1894	4.51	31.79
8	4,631.844	4.73	36.51
9	4,300.8026	4.39	40.90
10	4,114.4416	4.20	45.10
11	3,905.3321	3.99	49.09
12	4,205.2019	4.29	53.38
13	4,226.0275	4.31	57.69
14	4,074.5033	4.16	61.85
15	4,041.3627	4.12	65.97
16	4,252.8293	4.34	70.31

17	4,150.9773	4.24	74.55
18	4,489.1452	4.58	79.13
19	4,189.1484	4.28	83.41
20	4,192.2816	4.28	87.69
21	4,265.4102	4.35	92.04
22	4,003.8333	4.09	96.12
23	3,797.2187	3.88	100.00
Total	97,988.918	100.00	

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Thu, 06 Dec 2018 01:57:59 GMT
View Forum Message <> Reply to Message

Dear Fredarnold Sir,

Thank you so much, the code is working.

I was looking for this solution for many days.

Will keep in touch with

Regards
Muneer

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Tue, 18 Dec 2018 05:46:12 GMT
View Forum Message <> Reply to Message

Dear Fredarnold Sir,

First of all thank you so much for earlier clarification regarding mismatch in NFHS national report, Your solution was working perfectly.

However, now I have come across state-level reports. For example, I have taken the Jharkhand report. Table 61 of initial breastfeeding has been attached for your reference.

The table says that Number of last born children below 24 months is given as 4723.

However my estimation is not matching with state report, Stata coding is given below;

```
********************************************************** *******
****Jharkhand****
```

```
clear all
use "/Users/muneerkalliyil/Desktop/UNICEF/NFHS4/IAKR74FL.DTA"
gen survweight = v005/1000000
gen stratum = v023
svyset v021 [iw=survweight], strata(stratum)

gen age = v008 - b3

numlabel, add
 ****keeping only Jharkhand****
tab  v024

keep if  v024 == 15

 ****giving weight****
tabulate age if age < 24  & midx==1 [iw=survweight]

****without weight*****
tab age if age < 24 & midx==1
 ********************************************************** ********
```

I have estimated number of last born children below 24 months, giving weight and without giving as well.
Result says the number of children with weight = 2864, without weight = 4705.
Therefore, in both cases, the result is not matching with state report. It says the number is 4723.
And the gap becomes large when we use weight.

My queries are;
Is it correct to use weight in the state level, if yes, is it different from national weight?
Why the gap is large with and without weight
Could you please sort out my problem as soon as possible

Looking forward to hearing from you

Thanks in advance

File Attachments

1) Screenshot 2018-12-18 at 10.51.42 AM.png, downloaded 482 times

---

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by fredarnold on Tue, 18 Dec 2018 16:44:18 GMT
View Forum Message <> Reply to Message

For tables for individual states, you should use the state weight variable (SV005) instead of V005. SV005 is the weight used in all the state report for NFHS-4.

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Wed, 19 Dec 2018 01:21:47 GMT
View Forum Message <> Reply to Message

Thank You so much, sir
It is working perfectly

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Wed, 19 Dec 2018 16:09:51 GMT
View Forum Message <> Reply to Message

Dear Sir,

Sorry for spamming your inbox again and again.
This time again with earlier same table, Table 61 Initial breastfeeding (attached below).
I am looking to the second column, Percentage who started breastfeeding within one hour of birth.
The table says that 29.9% of Urban and 33.8% of Rural children have been put into breast within one hour of birth.

myStata code;
recode m34 (0/100 = 1) (101/236 = 0), gen(bf1hour)
svy: tab bf1hour if age < 24 & midx==1 & v102 == 1
svy: tab bf1hour if age < 24 & midx==1 & v102 == 2

and our results say

. svy: tab bf1hour if age < 24 & midx==1 & v102 == 1
(running tabulate on estimation sample)

Number of strata = 24 Number of obs = 828
Number of PSUs = 282 Population size = 858.55407
Design df = 258

RECODE of
m34 (when
child put
to
breast) proportions

0 .6929
1 .3071

Total 1

Key: proportions = cell proportions

. svy: tab bf1hour if age < 24 & midx==1 & v102 == 2
(running tabulate on estimation sample)

Number of strata = 72 Number of obs = 3735
Number of PSUs = 917 Population size = 3725.5765
Design df = 845

RECODE of
m34 (when
child put
to
breast) proportions

0 .6518
1 .3482

I agree it is a small difference, 30.71 in place of 29.9% and 34.82 in place of 33.8%, still, I am asking you because I got an exact matching solution from you for earlier problems.

Looking forward to hearing from you
Thanks in advance

File Attachments

1) Screenshot 2018-12-18 at 10.51.42 AM.png, downloaded 360 times

---

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by fredarnold on Wed, 26 Dec 2018 17:15:50 GMT
View Forum Message <> Reply to Message

It looks like you are trying to match the results for Jharkhand. In that case, you need to change the weight from v005 to sv005, which is necessary every time you are analyzing an individual state. As you can see below, the urban percentage for Jharkhand (29.9 percent) comes out exactly the same as the estimate in the Jharkhand state report.

. svy: tab bf1hour if age < 24 & midx == 1 & v102 == 1 & v024 == 15
(running tabulate on estimation sample)

Number of strata = 24 Number of obs = 852
Number of PSUs = 286 Population size = 881.91211
Design df = 262

```
----------------------
RECODE of |
m34 (when |
child put |
to        |
breast)   | proportions
----------+-----------
        0 |      .7011
        1 |      .2989
          |
    Total |           1
----------------------
  Key:  proportions  =  cell proportions
```

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Thu, 27 Dec 2018 14:07:11 GMT
View Forum Message <> Reply to Message

Thanks so much
The problem was in recoding, now it is matching perfectly

Thank
keep in touch with

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Pooja Arora on Sun, 17 Feb 2019 15:34:38 GMT
View Forum Message <> Reply to Message

Dear fredarnold,

When I use svyset command with state level weights for doing state level analysis for India, I use the weight variable sv005. In that, does my strata remain the same? That is v022?

Thank you,
Pooja

fredarnold wrote on Wed, 26 December 2018 12:15It looks like you are trying to match the results for Jharkhand. In that case, you need to change the weight from v005 to sv005, which is necessary every time you are analyzing an individual state. As you can see below, the urban percentage for Jharkhand (29.9 percent) comes out exactly the same as the estimate in the

Jharkhand state report.

```
. svy: tab bf1hour if age < 24 & midx == 1 & v102 == 1 & v024 == 15
(running tabulate on estimation sample)

Number of strata   =        24                  Number of obs      =       852
Number of PSUs     =       286                  Population size    = 881.91211
                                                Design df          =       262

----------------------
RECODE of |
m34 (when |
child put |
to        |
breast)   | proportions
----------+-----------
        0 |       .7011
        1 |       .2989
          |
    Total |          1
----------------------
  Key:  proportions  =  cell proportions
```

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Tue, 19 Feb 2019 13:58:04 GMT
View Forum Message <> Reply to Message

yes

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Bridgette-DHS on Tue, 19 Feb 2019 16:44:22 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

Yes, v022 is the stratum id at the state level as well as at the national level. The strata that are outside the state will be ignored.

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Fri, 22 Feb 2019 13:27:07 GMT

View Forum Message <> Reply to Message

Dear fredarnold,

As there are two different weights for the country and state, is there any third weight for district analysis in case of India. If not, what should we do, should we apply state weight or do without any weight.

Looking forward to hearing from you

Thanks

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by fredarnold on Fri, 22 Feb 2019 14:00:01 GMT
View Forum Message <> Reply to Message

For district level, or any sub-domain, either national weights or state weights can be used. Both will result in the same indicator values but not the same weighted N's. The national weights can be used for any analysis. The state weights can be used for data analysis on the state or lower levels.

Subject: Re: Mismatch between National report and NFHS4 raw data analysis
Posted by Muneer Kalliyil on Fri, 22 Feb 2019 14:57:39 GMT
View Forum Message <> Reply to Message

Thanks