
Subject: Discrepancies in Observations (Children's File 2014 Ghana DHS)

Posted by [geniusmuso](#) on Tue, 07 Jun 2016 01:53:30 GMT

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

Dear Sir,

Please with regards to the children's recode file of the Ghana 2014 DHS, it can be seen that the total observation is 5884, however, a variable like b8 (child's age) has a total frequency of 5,595 which is less than 5,884 but Child's birth order has a total frequency of 5, 884 which is in line with the observation. Also this can be seen with regards to variables h4 (polio 1), h6 (polio 2) and h8 (polio 3) in which each had a total frequency of 5,592 or 5,588 excluding don't know responses which is less than 5,884 and h0 (polio 0) with a total frequency of 5, 594 or 5590 excluding don't know responses also less than 5,884. Also even there is a difference of between the total frequencies each for h4, h6, and h8 which were the same on one hand and h0 on another. Several of such situations can be seen for most variables like h34(Vitamin A), h42(Iron sprinkles), h43 (drugs for intestinal parasites) which all had total frequencies less than 5,884 and others. So please are these occurring due to missing data as i have assumed? please explain. Thank you

Best regards,

Mustapha Immurana

Subject: Re: Discrepancies in Observations (Children's File 2014 Ghana DHS)

Posted by [Bridgette-DHS](#) on Tue, 14 Jun 2016 14:29:13 GMT

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

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

The KR file (and the BR file) has a record for each child born in the five years before the interview, including children who have died. If you do "tab b5,m" you will see that of the 5884 children in the KR file, 289 died before the survey. If you do "tab b8,m" you will see that age in years is given for all 5595 children who are still alive; it is not given, i.e. is missing, for the 289 children who have died. I think that will explain the discrepancies you are finding. Those children are not missing, but current age is not defined for them.

Note also that all numbers in DHS reports, unless stated explicitly to the contrary, are weighted. If you do "tab b5 [iweight=v005/1000000]" you will see that the 5884 UNWEIGHTED children correspond with 5695 WEIGHTED children.

Subject: Re: Discrepancies in Observations (Children's File 2014 Ghana DHS)

Posted by [geniusmuso](#) on Thu, 16 Jun 2016 11:32:06 GMT

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

Dear Sir,

Thank you so much. Please i would want to find out if I do "tab b5 and any variable and i get something like this

tab b5 pn1

child is alive	received pneumococcal-1		Total
	no	yes	
yes	2,033	3,482	5,515
Total	2,033	3,482	5,515

Please does it mean for immunization status or specifically on the pneumococcal vaccine 1, it contains only information on children who were alive at the time of the survey since only yes appear under child is alive? and 2. if the total is not up to 5,595 (number of children who were alive), for instance in the above example where the total is 5,515, can we attribute 80 (5,595-5,515) as missing? and 3. Please can i apply the above responses (1 and 2) you will give to all variables? Thanks

Best regards,

Mustapha Immurana

Subject: Re: Discrepancies in Observations (Children's File 2014 Ghana DHS)
Posted by [Bridgette-DHS](#) on Thu, 16 Jun 2016 23:37:20 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

Data on immunizations/vaccinations is only collected for children who are alive at the time of the survey, that is, children with b5=1. If the child has not survived, that is, if b5=0, then all the data on immunizations/vaccinations is coded ".", which means "Not applicable".

If you enter "describe pn1" you will see that PN1 is the name of the label for pn1. (Usually the name of the label is the same as the name of the variable, except that the label may have upper case letters.) Then enter "label list PN1". This will give you the numeric codes and the complete labels for each code. It can be very helpful to list the label because often the label is incomplete in a "tab" command.

```
. label list PN1
PN1:
```

- 0 no
- 1 vaccination date on card
- 2 reported by mother
- 3 vaccination marked on card
- 8 don't know

All of the immunizations (for example h0, h2 ,..., h9, pn2, pn3) have this same label (repeated as H0, H2, etc.).

When I open the KR file for this survey and enter "tab pn1" I get all the codes, not just "no" and "yes". You must have done a recode. When you recode a variable you should always change the name of the variable so you do not overwrite the original. For this variable there are 80 cases with code 8, which means "don't know".

I recommend that you be very cautious with the term "missing". For this variable, for example, "." means "not applicable" and 8 means "don't know". There are not really any "missing" cases at all.

Finally, I would emphasize that if you plan to analyze the data you should use weights: "tab pn1 [iweight=v005/1000000]" (or you could use svyset and svy).

Subject: Re: Discrepancies in Observations (Children's File 2014 Ghana DHS)
Posted by [geniusmuso](#) on Fri, 17 Jun 2016 12:00:56 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Sir,

Thanks so much indeed i recoded it.

Best regards,

Mustapha Immurana

Subject: Re: Discrepancies in Observations (Children's File 2014 Ghana DHS)
Posted by [geniusmuso](#) on Tue, 21 Jun 2016 11:10:15 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Sir,

I am very much grateful since this is the first time i am using a DHS data. However I would please need assistance with regards to these questions

1. Can't I use unweighted descriptive statistics of variables since i am using these variables to run regressions?

2. Since variables like v131 (5,883), v481 (5,883), v714 (5,880), v701 (5479), v467b, v467c and v467d (5, 883) had their total frequencies (in parenthesis) not up to 5,884, how do I call these differences?

3. Also if for instance

tab b5 is

child is			
alive	Freq.	Percent	Cum.
-----+-----			
no	289	4.91	4.91
yes	5,595	95.09	100.00
-----+-----			
Total	5,884	100.00	

And tab h42 is

taking iron			
pills,			
sprinkles			
or syrup	Freq.	Percent	Cum.
-----+-----			
no	4,296	76.81	76.81
yes	1,220	21.81	98.62
don't know	77	1.38	100.00
-----+-----			
Total	5,593	100.00	

And tab h42 b5 is

taking		
iron		
pills, child is		
sprinkles alive		
or syrup yes Total		
-----+-----+-----		
no	4,296	4,296
yes	1,220	1,220
don't know	77	77
-----+-----+-----		
Total	5,593	5,593

Then please in the above example since b5 (yes) is 5, 595 and the total frequency of h42 is 5,593, please how do I term the difference of 2 since the h42 was applicable to only those who were alive? And please can I apply your response to similar situations?

4. Please is there any means in stata where for instance I can use the tab command to get me the maternal employment frequency distribution(not crosstab) of only children who were alive.
Thanks so much for your patience.

Best regards,

Mustapha Immurana

Subject: Re: Discrepancies in Observations (Children's File 2014 Ghana DHS)
Posted by [Bridgette-DHS](#) on Tue, 21 Jun 2016 14:07:19 GMT

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

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

You should use the weights for descriptive statistics as well as for regressions. Otherwise your estimates are biased. You do not need to use weights for data quality checks, calculating nonresponse rates, doing very preliminary analysis--things like that.

If you do, say, "tab h42, m" the tabulation will include the number of cases coded ".". The m option stands for "missing" but here the "." means "Not applicable". For those cases, the question or information is skipped because of some filter earlier in the questionnaire. You can go back to the questionnaire (an appendix in every main report) to find why those cases were coded "not applicable". In addition there may be a few codes "8" or "9" (or "88" or "99", etc.) which may mean "no response" or "don't know". Usually there are only a few such cases and they do not seriously affect the analysis, but you definitely must exclude them from regressions, or else the code will be treated as a numeric value.

Most people do listwise deletion, i.e. drop from a regression the cases that have 8 or 9 (etc.) on ANY of the variables in the model. You can accomplish that with an "if...." in the model OR you can recode the variables before doing the regression, such that those codes are converted to ".".

I don't understand "the maternal employment frequency distribution(not crosstab) of only children who were alive". The first part refers to the mother and the last part refers to her children, and she may have a combination of children who are alive and children who have died. Are you looking for women who have any living children, or any children at home? You may be able to get the information you want from v202, v203, or v218.

Subject: Re: Discrepancies in Observations (Children's File 2014 Ghana DHS)
Posted by [geniusmuso](#) on Tue, 21 Jun 2016 15:41:30 GMT

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

Dear Sir,

Please I read this "Use of sample weights is inappropriate for estimating relationships, such as

regression and correlation coefficients." in the guide to DHS statistics so i do not know what is it?

On mother's employment i meant if there is any approach to report descriptive statistics for let's say v714 whiles excluding mothers whose children died. Thus from the little understanding i have, a mother would appear as many times the number of children under five she has therefore if any of her children died, there wouldn't be any problem. Thanks so much

Best regards,

Mustapha Immurana

Subject: Re: Discrepancies in Observations (Children's File 2014 Ghana DHS)

Posted by [Bridgette-DHS](#) on Fri, 24 Jun 2016 13:51:48 GMT

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

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

Yes, the Guide to DHS Statistics does recommend against using weights in regressions and correlations, etc. This is an old recommendation and it will not appear in the next revision of the Guide. We now recommend that all models take account of the weights, clustering, and stratification with the `svyset` and `svy` commands (or whatever is appropriate if you are not using Stata).

For your question about analyzing the employment status of women, I think your research question is whether the number of living children under five affects a woman's employment status. If so, your cases would be women rather than children and should probably include women who have NO children under five as a comparison group. It may be easier to work with the IR file. The KR file will include women who had a birth in the past five years (with repetitions of women with more than one such birth, as you say). The BR file will include women who have had ANY births, with repetitions of women for each birth.

If you use the IR file, you will get all women, regardless of whether or when they ever had children. `v208` is number of births in the past five years. `v220` is number of living children (plus current pregnancy) and `v213` is current pregnancy status. I don't think there is a variable that is specifically the number of living children under five, although maybe I am missing it. You could construct such a variable, either working from the `b` variables that are on the woman's record or working from a merge with the KR or BR file. Please think about specifically what it is about the children that you want, and then try to construct it. Let me know if you have difficulty constructing it. I strongly recommend, however, that you include women with NO children, whether living or under 5, in your model. I could imagine other relevant categories, such as no children under five but children OLDER than five, or no children ever born, etc. Hope I have correctly understood what you are trying to do.
