
Subject: missing values in prenatal care

Posted by [kkkkkk](#) on Fri, 29 Jan 2021 00:49:08 GMT

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

Hi! I am looking into the variables regarding prenatal care. For some variables, it is entirely blank, representing by a dot (.) but some of the variables have responses and with a dot (.) as well. From the DHS manual, it is said that any category not used in a particular country is left blank. For some variables ie m2n_6, the responses of the entire column is (.), should I take that as missing value or an unanswered question? For some variables ie m2n_1, there are yes/no responses and a (.) as well, how should I interpret the dots? By the way, is there a way I can find questions I asked in this forum? Thank you!

Subject: Re: missing values in prenatal care

Posted by [Bridgette-DHS](#) on Fri, 29 Jan 2021 11:26:28 GMT

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

In general there are two types of missing. The ones caused by a skip or filter in the question, in which case the respondent was not asked the question at all. And the other type of Missing (coded 9, 99, 999....), where the information was omitted due to an interviewer's mistake. Whenever you see a lot of NA (".") codes, the only way to determine the type of missing information is to check the questionnaire, and observe the skip pattern if any. This will allow you to determine which ones are the result of a skip, and which ones are the result of omitted information. You should also look at the description of the survey, usually in the first chapter of the report. Lastly, look for variables with a title that includes "eligible" or "eligibility". These will help you figure out any subsampling. When there is subsampling, it usually means not everyone was interviewed, for instance: every person in every second or third household.

Subject: Re: missing values in prenatal care

Posted by [kkkkkk](#) on Sat, 30 Jan 2021 15:25:20 GMT

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

I am doing a research on the effects of maternal education on child mortality, and considering other factors such as breastfeeding practice, wealth index. I am using the IR file however I found that it is not a good thing to do as current wealth index has nothing to do with child mortality if the child died long time ago. Should I be using other files to only observe the mortality in the last 5 years or should I be merging datasets (ie waves) to create panel? Essentially, I am trying to avoid using information that happened after the child dead.

The solution I came up with is to merge the ir and kr datasets to obtain a variable where I will be able to know whether a mother experienced child mortality at least once in the past 5 years. What are the commands to do this?

By now, I merged the children and women's dataset so I have mortality in the past 5 years and constructed a variable for mortality. However, I got pretty weird results. I guess I made a mistake in the merging. Meanwhile, if I use the KR file to do it, my results become 'normal'. Therefore,

should I go for kr file since I am only looking into child mortality in the past 5 years (and the overlapping is not an issue). Can you please give me some advice on that? Thank you!

File Attachments

- 1) [ir results after merge.PNG](#), downloaded 356 times
 - 2) [kr results.PNG](#), downloaded 352 times
 - 3) [commands.PNG](#), downloaded 348 times
-

Subject: Re: missing values in prenatal care

Posted by [Bridgette-DHS](#) on Tue, 02 Feb 2021 14:05:29 GMT

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

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

I agree that you should be cautious about assuming that the wealth quintile of the household is constant over time. However, looking at earlier DHSs in the same country will not help with this issue. The households and clusters are completely different in successive surveys. Assuming that the household was in the same wealth quintile five years ago (or even earlier) that it is in at the time of the survey may be your best option.

It would be possible to use just the birth histories in the IR file to construct a variable like this: D is the number of child deaths at age<A months during the M months prior to the interview. For example, with A=12 and M=60, this would be the number of infant deaths in the past five years. I would construct D such that it would be a dot, or NA, if the woman did not have any children who had a risk of dying at age<A during the reference period. To calculate this you would loop through the b variables (subscripted with _01, _02,...,_20), focusing mainly on b3 and b7.

This is a non-trivial recode. Unfortunately I just don't have the time to do it--you will have to do it yourself. The main complication is that we don't know the cmc of death, for children who died. We know cmc of birth (b3), and age at death, in months (b7). After 23 months, b7 is in intervals, coded with the midpoint month of a 12-month interval.

I am not sure how this variable would be used in a model. It could be an interesting outcome variable, but you would have to be careful using it to predict something else.
