Subject: Missing values in NDHS 2003, 2008 and 2013
Posted by mertmert on Thu, 23 Jul 2015 14:04:20 GMT
View Forum Message <> Reply to Message

I am a Masters studentand I am currently working on my thesis about the changes in HIV/AIDS prevention knowledge in Nigeria. I have received and I am working with the Nigerian DHS data of 2003, 2008 and 2013 and I have a few questions concerning missing values and I was wondering if you could help me understand the data. Thank you very much in advance.

I have combined the men's and women's DHS data sets of 2003, 2008 and 2013 in Nigeria and find myself with a data set of 115,144 individuals. I have three questions concerning missing values in this data set:

1. Many of the demographic variables do not have any missing values: age, region, urban/rural residence, highest educational level and wealth. The variable "marital status" also has a very low number of missing values (out of 115,144 only 1 value of missing). Because the data set is so large, it is very surprising that those variables have no missing values at all. I was wondering if the data sets that were provided to me already dealt with missing values for the demographic variables and if yes how? (imputing the missing data with replacement values?)

2. My second question is about two HIV/AIDS knowledge variables. The variable "knowledge about condoms to reduce the chance of getting the AIDS virus" (mv752cp for men and v752cp for women) has 12,903 missing values and the variable "knowledge about limiting oneself to one faithful and uninfected sexual partner" (mv754dp for men and v754dp for women) has 12,883 missing values. However the missing values of these two variables come mostly from the same individuals. Once the 12,903 missing values of "knowledge about condoms to reduce the chance of getting the AIDS virus" are eliminated, the variable "knowledge about limiting oneself to one faithful and uninfected sexual partner" has only 154 missing values. Since the same individuals seem to not have answered these two questions, I was wondering if maybe the question was not asked to some specific groups/sub-groups for some reasons? or at random?

3. Finally, I am concerned with the variable "used condom during last intercourse" (mv761 for men and v761 for women). This variable has 30,277 missing values which is a lot. I was wondering if there was some kind of selection when this question was asked. Was it not asked to certain kind of respondents? Why are the missing values so high?

Again, thank you very much in advance.

Subject: Re: Missing values in NDHS 2003, 2008 and 2013
Posted by datalearn on Sun, 01 May 2016 22:33:43 GMT
View Forum Message <> Reply to Message

I would be interested in this answer as well.
Thank you

Subject: Re: Missing values in NDHS 2003, 2008 and 2013
Posted by Liz-DHS on Mon, 09 May 2016 21:00:04 GMT
View Forum Message <> Reply to Message

Dear Users,
A good place to start is with the description of the various data sets:http://www.dhsprogram.com/data/Dataset-Types.cfm#CP_JUMP_6014, The Standard Recode Manual http:// www.dhsprogram.com/publications/publication-dhsg4-dhs-questi onnaires-and-manuals.cfm ,
The Guide to DHS Statistics http:// www.dhsprogram.com/publications/publication-dhsg1-dhs-questi onnaires-and-manuals.cfm,
The "Using Data for Analysis" section on our Website http://www.dhsprogram.com/data/Using-Datasets-for-Analysis.cfm and viewing various videos on data downloads and use. It's also important to look at the Core Questionnaires http:// dhsprogram.com/What-We-Do/Survey-Types/DHS-Questionnaires.cf m and for your own countries, the map files, frequencies (weighted and unweighted) and the questionnaires for your particular survey located in the back of the final reports. You can also download Model Datasets without registering. These will allow you to experiment with the model (not belonging to any country, but covering many tables)and get used to working with our data before attempting your own. http://dhsprogram.com/data/Download-Model-Datasets.cfm.
Thank you!

Subject: Re: Missing values in NDHS 2003, 2008 and 2013
Posted by Liz-DHS on Mon, 09 May 2016 21:36:23 GMT
View Forum Message <> Reply to Message

Dear User,
A response from data processing expert, Mr. Noureddine Abderrahim,
Quote:

During the data collection of the DHS survey, some questions can be missing such as the date of birth but some other questions can't be missing such as the marital status since these questions are used during the skip to define the base of individual to ask certain questions to. The respondent should be able to know whether he/she is in union or not. The questions where we don't allow missing values can be sometimes by accident be missing but these are very rare. To be able to identify which ones for which we allow missing and those for which don't allow this value, you need to look at the map distributed with the data files

In case the number of missing is considerable the one sure thing to check is whether the question is asked to the entire population or to a subset. In all cases, you need to check the base population to which the question is addressed. On the top of my head, we can't ask about "knowledge about condoms to reduce the chance of getting the AIDS virus" unless the question to which the person interviewed has a previous knowledge of the AIDS virus. This information is given, in general, in the DHS Recode Manual as the BASE of the question. The BASE is not given in all cases and you will have to review the skip pattern used for the survey in question.

Subject: Re: Missing values in NDHS 2003, 2008 and 2013
Posted by kingx025 on Sun, 22 Apr 2018 20:45:32 GMT
View Forum Message <> Reply to Message

Liz-DHS has already given a good explanation of why the number of cases with non-blank values is smaller for some variables. I'm just going to suggest an easy way for you to check and see who is appropriately included in the data for a given question: Check the "universe" tab for the description of that variable on the IPUMS-DHS website.

Here's the universe for the variable 761, used condom during last intercourse, for Nigerian women:
Nigeria 1999: Women age 10-49 who had sexual intercourse in the last 12 months.
Nigeria 2003: Women age 15-49 who had sexual intercourse in the last 12 months.
Nigeria 2008: Women age 15-49 who had sexual intercourse in the last 12 months.
Nigeria 2013: Women age 15-49 who had sexual intercourse in the last 12 months.

Here's the universe for the variable on whether the woman respondent thinks using a condom reduces the risk of AIDS
Nigeria 2003: Women age 15-49 who heard of HIV/AIDS and reported knowing a way to reduce the risk of contracting it.
Nigeria 2008: Women age 15-49 who have heard of HIV/AIDS.
Nigeria 2013: Women age 15-49 who heard of HIV/AIDS.

Obviously all of these preconditions for these questions will reduce the number of cases with meaningful responses.

Here's how to check universes from the IPUMS-DHS website:

1. Go to idhsdata.org
2. Click on Select Data
3. Choose the unit of analysis (women in this case)
4. If desired, restrict the samples shown (say, to just Nigeria), using the Sample selection tool
5. Enter the name of the variable of interest using the Search tool
6. Click on the variable name and read the online documentation for the variable, including the "Universe" tab
7. If you don't find the variable using the survey search tool (as I didn't for the variable about beliefs about condoms preventing AIDS, because the name of that variable changes across DHS surveys),
then look for a variable with the relevant content using the drop-down menu of variables organized by topic

Along with learning more about variables of interest, you can create a dataset with the variables and samples you need, in the format you prefer, if you log in using your DHS user name and password.

Miriam King