Subject: Variable Definition
Posted by amvholloway on Sat, 14 Jan 2017 22:34:03 GMT
View Forum Message <> Reply to Message

Hello! I am having trouble deciphering some of the variables in the FGC module in the EDHS 2005. There are several variables with the following names: S808$1
S808$2
S808$3
S808$4
S808$5
S808$6
S808$7
Yet are all labeled "who performed the circumcision". The same goes for "age of daughter when she was circumcised" with the following names:
S809$1
S809$2
S809$3
S809$4
S809$5
S809$6
S809$7

I don't know what this means and cannot seem to find a definition or its significance. Can someone help define these for me or let me know what is the significance of the duplicate names with the dollar sign?
I would be eternally grateful! Thank you -

Subject: Re: Variable Definition
Posted by Bridgette-DHS on Tue, 17 Jan 2017 22:31:40 GMT
View Forum Message <> Reply to Message

Multiple or repeating records are placed one after the other, with the maximum number of occurrences of each section being represented in every case. Each variable in a repeating section is placed immediately after the preceding variable of the same occurrence, such that all variables for occurrence 1 precede all variables for occurrence 2 of a section. Each occurrence of every such variable must have its own name because statistical packages do not generally support the use of arrays or subscripts.

Subject: Re: Variable Definition
Posted by amvholloway on Sun, 22 Jan 2017 17:32:31 GMT
View Forum Message <> Reply to Message

Thank you for this definition. I have a question about how I analyze this in SPSS now. (please bear with me as I am VERY new to this and don't really know what I am doing). I just ran a process to clean the data for missing values and included these repeating variables. However, I'm

not sure this is the correct procedure as the output show that there are hundreds of missing values from these repeating variables. I don't understand how to analyze these values because if I replace each missing value with the calculated mean, this will grossly skew the data.

So my question to you is, how do I analyze this data in SPSS, accounting for missing data? I feel like I am missing a big piece of the puzzle and don't even know where to look for it. Any help you can offer would be amazing and I would be eternally grateful.

(I hope I have articulated myself well-enough).

Kindest regards,
Ashley

Subject: Re: Variable Definition
Posted by Bridgette-DHS on Mon, 23 Jan 2017 15:56:30 GMT
View Forum Message <> Reply to Message

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

Like many users, you are misinterpreting the symbol for "Not applicable" as a "Missing" code. For example, if $1 through $6 refer to children born in the past few years, with an allowance for a maximum of 6 children, and the woman only had one child in the past five years, then all of the $2 through $6 variables will be blank. It's not a data quality issue, such as, for example there being a real child whose height should have been measured but was not. In any kind of calculation, the blanks will be omitted entirely, from numerators, denominators, means, whatever. You do not have to do anything to remove them. They are already removed.

Some personal advice (not on behalf of DHS): do not use SPSS.

Subject: Re: Variable Definition
Posted by kingx025 on Sun, 22 Apr 2018 18:01:43 GMT
View Forum Message <> Reply to Message

Tom Pullum has explained the issue of "not applicable" cases very clearly; we use the term "universe" in IPUMS-DHS to express who is included in the coverage for a variable. For example, never-married women are not in the universe for questions about age at first marriage, and children who did not have recent diarrhea are not in the universe for a follow up question about whether they had blood in their stools during diarrhea.

If you want to check the universe (who was included in a question for a particular sample), try checking the variable-specific documentation on the IPUMS-DHS website. Use the Search tool on the Select data page, and then click on the name of a variable when it comes up. There will be documentation on various aspects of the variable, but one of the most useful is on the "universe" tab, which specifies for each sample who was covered by this variable. Not all of the DHS variables are included in IPUMS-DHS yet, but many thousands of them can be checked for their

universes in a particular sample.

Miriam King