Subject: How to treat missing data in analysis
Posted by India2222 on Tue, 09 Aug 2022 07:14:17 GMT
View Forum Message <> Reply to Message

Hello,

I am using data from the India 2019-20 NFHS survey for my thesis, concentrating on variables related to HIV knowledge, attitude and behaviour. For certain variables, for example, V769 'can get a condom' is missing over 30%.

I am relatively new to dealing with missing data. Should I run my analysis as normal (chi square, binary logistic regression) and address and state the missing values in my write up, perhaps discussing the reason for this missing data?

Do you know if this missing data is random or due to the surveying method for various HIV knowledge, attitude and behaviour variables?

EDIT: I believe that those missing are often due to the fact that they answered 'no' to 'ever heard of HIV/AIDS' hence no further questions were asked. Therefore, I believe I will only include those who answered 'yes' in my study?

Many thanks

Subject: Re: How to treat missing data in analysis
Posted by Janet-DHS on Fri, 12 Aug 2022 20:51:57 GMT
View Forum Message <> Reply to Message

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

By "missing" you are probably referring to a dot in Stata ("."), which is equivalent to a blank, and is the Not Applicable (NA) code. It means the question or variable did not apply to the case. If you go back to the questionnaire and find the question for "can get a condom" I expect you will find that it is preceded by filters and skips such that the question is only asked of some subpopulation of women. For other women, the question is not asked and is coded with a dot for NA.

Sometimes in DHS data files you will find a code "9" or "99" etc. that is not included in the label for the variable. When you find these codes, you will probably want to drop the case from a tabulation or calculation of a mean, etc., but they are actually very rare.