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.