
Subject: Too many not-available (.) observations
Posted by [nwegbus](#) on Fri, 11 Dec 2015 17:29:32 GMT

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

Hello,

I'm a doctoral student working on my first academic paper and I'm using the Nigeria DHS 2008 dataset and StatC 13 software for analysis.

My population is 33,385 women (15-49years) but I'm interested in a subpopulation of these (23,954) who are married. However, when I weight the data, specify the subpopulation and run my logistic regression model, the output gives a subpopulation size of 15,449. I think this may have to do with the fact that my main predictor variable v511 (age at first marriage) has 8021 not-applicable (coded ".").

I initially had re-coded the not applicables as missing, but then ran into trouble when chief evaluator wanted to know why I wouldn't do multiple imputation since the missingness was so pronounced.

QUESTIONS:

- (1) How do you suggest that I handle the not-applicable since recoding them as missing does not seem to work
- (2) Would my sample size then be 15,449 (as specified by the regression output) or 23,954(the number of married women in my specified subpopulation)?

Any tips or links would be highly appreciated. Thanks in advance.

SN
