
Subject: Missing data

Posted by [Dsisso](#) on Fri, 07 Feb 2014 06:39:07 GMT

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

Hello everybody,

I am working on an immunization dataset in which I like to impute missing values on DTP1,2 and 3 doses and to calculate the prevalence of unimmunized by each vaccine dose. Is someone experienced the same problem? In STATA, I am able to impute multiple imputed datasets but I am experiencing difficulties in combining all imputed datasets to one containing complete information per observation. So far, in my understanding, I can only combine estimates (e.g, coefficients by using logistic regression and Rubin rule or combination) while I want to rather generate categorical DTP1,2,3 with missing values filled by multiple imputation procedure.

Thank you for our help,

Dsisso

Subject: Re: Missing data

Posted by [Bridgette-DHS](#) on Wed, 26 Feb 2014 14:03:07 GMT

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

We are working on a response to your posting.

Thanks

Subject: Re: Missing data

Posted by [Bridgette-DHS](#) on Fri, 28 Feb 2014 15:20:06 GMT

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

Following are comments from DHS Specialist, Tom Pullum:

DHS does some imputation or reconciliation of incomplete or conflicting dates or ages, for example in birth histories. We may shift out-of-range codes into a not-stated code, as with the height and weight measurements. But we never, or hardly ever, make imputations for missing data and we don't have a policy on how users should make such imputations.

With multiple imputation procedures the interest is in the coefficients produced by a model. The model is run many times, with alternative individual-level imputations, to optimize the values of the coefficients, but multiple imputation is not a device for identifying optimal individual-level imputed values.

Subject: Re: Missing data

Posted by [Bridgette-DHS](#) on Wed, 05 Mar 2014 17:09:47 GMT

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

Following is another response from one of our DHS experts, Shea Rutstein.

In the DHS we do not impute whether each child received a vaccination. If a vaccine dose is not recorded on the the child's immunization card, then the mother is asked whether the child received the vaccination. For DPT, the questions asked are in the attached file.

The age at which the vaccine is given is assumed to be the same as that for which dates have been given (an aggregate assignment, not individual, done during tabulation). Since the outcome of whether a child is given a vaccination is dichotomous, either logistic or probit regression is appropriate. The predicted value will be the probability that the child was given the vaccination. I would combine the several imputations to get the average probability for each child and then randomly select a number so that you assign 1 or 0 (got vaccination or not) according to the number selected, e.g. if the probability of receiving DPT1 is 0.60 then a randomly selected number between 0 and 5 would indicate that the vaccine was given and between 6 and 9 would indicate that the vaccine was not given.

Another way to go about it would be to use hot deck imputation according to the characteristics correlated with each vaccination, such as child's sex, birth order, age, area of residence, province, place of birth, wealth quintile, etc. but the list of variables may be long.

As I understand it, Rubin's procedure for combining multiple imputations is to produce a more robust standard error by taking account of both interval variance and variance between the estimates. In this case, I would take the calculated probability of each child from the estimating equation and then randomly vary by selecting a deviation using the normal distribution of the standard error, and then apply the adjusted probability for each child to a randomly selected number to determine wether the vaccine was given.

Let me know if this helps.

Shea

File Attachments

1) [missing data.jpg](#), downloaded 873 times

Subject: Re: Missing data

Posted by [Dsisso](#) on Fri, 07 Mar 2014 06:55:06 GMT

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

Thank you every for your helpful replies. I will take i in consideration in order to rsolve this situation. Ihave ever tried the hotdeck procedure which reveals less robust and I seems more interesting to focus on estimates fro regressions rather than individual doses.

Many thanks

D. Sissoko
