Understanding Society User Support - Support #1904

Using weights when variables have some or many "missing value" codes or NAs due to missing household level data

05/17/2023 10:59 AM - Richard Belcher

Status:	Resolved		**Start date:**	05/16/2023
Priority:	Normal		**% Done:**	100%
Assignee:	Olena Kaminska			
Category:	Weights			

Description

Dear Olena,

I am running a pooled cross sectional individual level analysis (waves 1-9), using cross-sectional weights, but I am worried that by removing cases where sf-12 responses are -9, my sample is no longer nationally representative.

I am selected weights in my analysis that are appropriate for how the questions leading to the variables I want to use were administered. E.g. I am using self-completion questionnaire cross-sectional weights (waves 2+) as sf-12 is my dependent variable of interest in later models (weight appropriate for each wave). After aggregating the cross wave data there are a number of cases where cases with non-zero weights have missing value codes attached to them (or are NA due to me merging in household level data which is occasionally not collected). It is understandable that errors and non-response happens during the survey process. Am I safe to assume that some are random, e.g. the lack of household interviews being undertaken is random, so it wouldn't impact the weighting removing responses without that information. I am however worried that some may not be random and there may be some demographic or regional bias to -9 codes in the sf-12 variable, which prevent my sample from being nationally representative when weighted. I have 96% of the non-zero weighted samples remaining after removing those with errors, most of the reduction (3%) comes from "sf12mcs_dv" responses with the code -9.

Thanks for your help,

All the best,

Richard

History

#1 - 05/17/2023 01:19 PM - Olena Kaminska

Richard,

Thank you for your question. Indeed, there may be some part of nonresponse unaccounted due to missing responses to an item in a questionnaire. There are a few approaches how one would deal with it.
Firstly, one may consider missing responses as valid responses. For example, refusals and DKs may have a substantive meaning.
Secondly, one may report proportions with a missing category. In such situation the proportions will not be skewed, but missingness will be visible and may be informative for readers.
Third, one could impute item missingness. Due to richness of available information on other items, it is usually better to use imputation to deal with item missingness. Once imputed, the full information is then weighted to the population using weights.
It is best if the decision of the appropriate approach is taken by a researcher who understands the meaning of the item and its best correlates, if imputation is taking forward.

I hope this answers your question,
Olena

#2 - 05/17/2023 04:35 PM - Richard Belcher

Dear Olena,

Thank you for your quick reply and your ideas on what could be done.

Thank you for your suggestion on missing responses being valid, this is something I am going to account for in some variables e.g. in the commute time to work (many people don't work so have a -8 code, and that is important information, for me so I include it as a category). Also I think having just a "missing" category to preserve the weighting would be useful in other cases (e.g. for -9 codes).

I would like to keep my dependent variable (sf12mcs_dv) (and unfortunately the one with the most missing values), continuous. Thank you for pointing me in the right direction, it seems like multiple imputations could be the way forward for this variable.

More generally would you normally exclude cases where variables have a small amount of non-responses from mean calculations for example (in my case about ~1%)? Or replace using MI or other single imputation approaches. I did not see anything in the documentation, but even for some common

variables like Age_dv there are -9 values.

All the best,

Richard

#3 - 05/19/2023 12:38 PM - Olena Kaminska

Richard,

What to do with item missingness completely depends on the subject you study and research question you have, and the consequences for it. You could exclude it as well, as long as you describe how you did it. 1% is small, though may be crucial for some questions, e.g. a refusal to a drug use question, even 1% of it, may be important.

Hope this helps,
Olena

#4 - 05/19/2023 02:29 PM - Understanding Society User Support Team

- Status changed from New to Feedback

- % Done changed from 0 to 80

- Private changed from Yes to No

#5 - 11/30/2023 01:03 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 80 to 100