

Understanding Society User Support - Support #2065

How to manage longitudinal data analysis after excluding sample based on date of interview

03/04/2024 03:09 PM - Marina Kousta

Status:	Feedback	Start date:	03/04/2024
Priority:	Normal	% Done:	50%
Assignee:	Understanding Society User Support Team		
Category:	Data documentation		
Description			
<p>I am conducting a (longitudinal) diff-in-diff analysis for a policy evaluation where the date of policy introduction is important. I have a few questions below:</p> <p>1) As my date of interest falls in the middle of a single wave, I could split up wave X into two parts indicating the before and after. Is this enough so that I can only use a single wave for the analysis, OR would you say it is preferable that I also use more waves to more accurately represent the year for the before and after treatment? (the reason i am asking is because i read the following on your website: "As some samples are fielded in the first 12 months (BHPS and General Population-Northern Ireland samples), some in months 13-24 (IEMB sample) and some across all 24 months (General Population-Great Britain and EMB samples), just using data from the same wave to compare the two consecutive years will result in comparing different samples. Similarly, just using data from year 1 or year 2 of a wave to conduct cross-sectional analyses of that year will result in analysing samples that are not-representative. So, to correctly do these types of analyses, data from two waves need to be combined. For example, for 2019, use data from year 2 of Wave 10 and year 1 of Wave 11."</p> <p>2) To split up any given wave into two separate waves etc, which variable would you recommend? I have seen many variables in the dataset indicating the month of interview, year, etc but there are also others relating to the sample, but I am unsure which variable would be the most accurate? Moreover, I am confused as some waves suggest they may extend across three calendar years but when I look at the year of interview variable, it only reflects year 1 and year 2, there is no mention of year 3.</p> <p>3) Which weights would you recommend using in this case?</p> <p>Many thanks in advance for any help you can provide.</p> <p>Best, Marina</p>			

History

#1 - 03/06/2024 02:59 PM - Understanding Society User Support Team

- Category changed from Data analysis to Data documentation
- Status changed from New to Feedback
- % Done changed from 0 to 50
- Private changed from Yes to No

Hello Marina,

The UKHLS sample is designed such that each sample month is a random representative (once weighted) sample of the population. There are some exceptions to this rule as Northern Ireland is only present in months 1-12 (first year of each wave), BHPS is only present in issue months 1-12 (first year of each wave) and the IEMB sample is only present in issue months 13-24 (second year of each wave).

We advise using the sample month/year (`w_month`) to identify the analysis sample rather than the month/year of the interview. For each sample month, interviews take place over 3-4 months, but the majority of interviews take place in the calendar month coinciding with the sample month. The interviews that come in later calendar months tend to be with sample members who are either hard to contact or reluctant to participate. Our weights are designed for each whole sample month to represent the population. If you omit the interviews from the calendar months following the sample months you will be excluding a category of respondents who tend to be very different to earlier respondents, so it is unlikely that your analysis sample will remain representative.

Please refer to [How to use weights – Analysis guidance for weights. PSU. Strata](#) section. It includes examples for analysing data monthly and for pooling data from different waves, which may offer some insight, if you intend to analyse data over 12-month periods.

I hope this information is helpful.

Best wishes,
Roberto Cavazos

