Understanding Society User Support - Support #1378

Overlapping interview periods across waves?

07/15/2020 01:28 PM - Abigail Dumalus

Status:	Resolved	Start date:	07/15/2020
Priority:	High	% Done:	100%
Assignee:			
Category:	Survey design		

Description

Hello Alita,

I noticed from browsing the dataset that wave periods from wave 2 until wave 9 have been overlapping. My basis for this observation are the following variables: istrtdaty, istrtdatm, istrtdatd, and wave. To illustrate, let me focus on waves 22 and 23:

- wave 22 (UKHLS wave 5) starts 9 January 2013 [11 interviews] until 29 April 2014 [1 interview]
- wave 23 (UKHLS wave 6) starts 8 January 2014 [26 interviews] until 11 May 2015 [1interview]

From my perception, interviews done from 8 January 2014 until 29 April 2014 in wave 23 can also be assumed to have happened in the latter portion of wave 22. I am really puzzled because I have set xtreg command with wave as a time variable, but then interview periods appear to overlap into the next wave. I have been searching for fieldwork information per wave to find out about official interview timelines. Can you please clarify where I can confirm actual interview periods per wave, so that I can still use wave as a panel time variable? Would this be an issue as well with how the weighting variables have been constructed?

History

#1 - 07/15/2020 01:36 PM - Alita Nandi

- Status changed from New to Feedback

- % Done changed from 0 to 80

- Private changed from Yes to No

Yes, in Understanding Society the fieldwork period for any wave stretches over 24 months and is overlaps with the next wave. But this overlap is to make sure every person is interviewed at approx one year intervals. So, as far as the panel structure is concerned, it is ok - as for any person (pidp) the interval accross waves is one year (wave). THe weights are ok.
https://www.understandingsociety.ac.uk/documentation/mainstage/survey-timeline
https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/survey-timeline

Hope this helps.

Best wishes,
Alita

#2 - 07/15/2020 01:53 PM - Abigail Dumalus

Thanks so much, Alita, for your feedback. Does the 24-month wave period preclude me from using "year" as a panel time variable? I am wondering whether it makes sense to analyse waves 2-9 annually, when the weighting variable is based on these 24-month wave period? Is it possible to construct annual weighting variable or is this going to be more complicated? I am asking this because I need to generate 3-month rolling average / variance for life satisfaction from wave 6 of the BHPS until wave 9 of the UKHLS. For the monthly period, I am using the istrtdatm variable. How should I proceed in this situation? My supervisor (who's a professor) keeps asking me to be more precise with the time period on the x-axis on each chart.

#3 - 07/15/2020 02:14 PM - Alita Nandi

- Assignee changed from Alita Nandi to Abigail Dumalus

I see. Yes, you can analyse by calendar year. Take a look at item 11 in the weighting FAQ and see if that is appropriate for you. If not, I will forward your question to Olena.
https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/user-guides/mainstage/weighting_faqs.pdf

#4 - 07/15/2020 03:06 PM - Abigail Dumalus

I read item 11 in the weighting FAQ. It seems appropriate but I cannot be sure whether this applies for building 3-month rolling averages / variances of life satisfaction. Maybe, an example of how I can generate one data point/observation of a 3-month rolling average is a helpful guide. The example of analysing January 2014 alone entails information from 3 waves. Is there a snapshot or formula that I can look at to make sure that I am not excluding interviews across several waves for a single calendar month? I can only imagine that I would need to build custom monthly weights to ensure representativeness for each calendar month, right? Many thanks for your patience.

#5 - 07/15/2020 04:06 PM - Olena Kaminska

Abigail,

Yes, I can confirm that FAQ question 11 applies to 3-months analysis. We do not have a ready syntax for you, but all the principals are outlined in the FAQ. You will not need to build your own weights if you follow the example in FAQ (you will need to combine all waves that were asked at your relevant point of time) - just use the appropriate us_lw weight. The most relevant information for you in FAQ starts with 'Let's say you are interested in studying December 2014.'

To explain the fieldwork, each sample month is representative of the population with some exceptions outlined in the FAQ. The January month has most interviews completed in January, after which one month is taken as a break and two months (March, April) with follow ups of nonrespondents - therefore each sample month takes 4 months to complete. This is because we try hard to keep our response rates high and this approach (and extra time) helps. Having this information you should be able to figure which waves to use for each calendar month.

Hope this helps,
Olena

#6 - 07/15/2020 05:05 PM - Abigail Dumalus

Hello Olena,

I tried following item 11 for doing January 2014 as an example. When I filtered interviews from wave 5, sample month 1 (if wave==22 & month==1) using the data editor on Stata, istrtdaty indicated 2013, istrtdatm ranged from 1 to 6. I am confused why interviews that started in 2013 have to be added under "January 2014" monthly average. Am I doing the filtering in a completely wrong way? I am not asking for the exact syntax exactly but it would be helpful if an illustration of January 2014 can be provided, along with the appropriate weighting variable for this month. Does this mean that I should be generating separate variables for all 3-month rolling averages/variances when building the charts? How would the us_lw weight be used then if each panel is using indin91_lw as a panel weight? I apologise for still struggling with understanding the weighting FAQs.

#7 - 07/16/2020 12:50 PM - Abigail Dumalus

May I please get any advice on my queries above about item 11 in the weighting FAQ? Many thanks for your time and consideration.

#8 - 07/16/2020 04:39 PM - Alita Nandi

- Assignee changed from Abigail Dumalus to Olena Kaminska

#9 - 07/16/2020 05:03 PM - Olena Kaminska

Abigail,

I think you have a few questions that I will try to answer below.
First, why are you using indin91_lw. I would suggest that you use indinus_lw if you are only interested in 3 month rolling average.
Something seems to be not working with your set up. I suggest that you look at the date of the interview and select all interviews from the relevant months. You don't need to use interviews from 2013 if you are interested in 2014.
Try to follow these steps:
1) select the interviews relevant to your timeframe (three months of interest);
2) these should be selected from different waves (up to 3 waves)
3) for each interview keep the relevant indinus_lw weight
4) do your analysis as usual within the new data.

Hope this helps,
Olena

#10 - 07/16/2020 05:45 PM - Abigail Dumalus

Hello Olena,

I am using indin91_lw for longitudinal regressions over time from 1996 to 2017 with life satisfaction as an outcome (LHS) variable.

The 3-month rolling averages/variances are for plotting smoother trend lines of mean LS and variance LS over time.

I totally understand the 4 steps that you have laid out. I am looking at istrtdaty and istrtdatm to somehow come up with a calendar month variable (let us call this "mdate", e.g. 2014m1 for January 2014) that uniquely tags each interview under its respective interview month. I am trying to find a way to use the "month" variable which has either yr1 or yr2 in tandem with istrtdaty and istrtdatm to ensure that each month is nationally representative. Whenever I look at the Data Editor spreadsheet, it seems that I am still missing a step or two to generate average life satisfaction for January 2014 interviews because some interviews are finished much later than this calendar month. All I need is the correct month variable that is similar to how each BHPS wave represents a 12-month period, whereas UKHLS waves are stretched over a 24-month period. Is there another way of going about this process?

I have been explaining these overlaps to my supervisor/principal investigator but he cannot understand why we cannot simply disentangle one wave into annual points in time. As of the moment, we want to know which years have considerable spikes/amplitudes in LS, and within these particular years, which months are these spikes/amplitudes are occurring, so that we can make sense of social events/ policy changes that may have triggered these unusual spikes/amplitudes.

#11 - 07/16/2020 06:58 PM - Olena Kaminska

Abigail,

The method I described earlier does not correspond to BHPS data (before UKHLS started). BHPS is normally collected over 3 months (with few late respondents coming in later months). You can't combine them in a similar way to UKHLS as there is no parallel sample month. You have two options here:
- ignore that just a few late respondents respond later - the weight will be correct for this;
- delete late respondents and readjust the weight for their missingness.

The method I explained applies for UKHLS.

Hope this helps,
Olena

#12 - 07/16/2020 07:01 PM - Olena Kaminska

Oh, and with regard to your comment on 'it seems that I am still missing a step or two to generate average life satisfaction for January 2014 interviews because some interviews are finished much later than this calendar month' - what you need is to look at interviews that took place (finished) in January 2014 regardless of which sample month they come from. This corresponds to my step 1) earlier.

#13 - 07/16/2020 07:21 PM - Abigail Dumalus

Thanks for the suggestions, Olena.

Which variable would indicate those "interviews that took place (finished) in January 2014 regardless of which sample month they come from"? The other set of variables that I am guessing these would be are the following: intdatd_dv, intdatm_dv, and intdaty_dv. Am I correct? Or is there a specific variable that tags when the interview finished in January 2014, for example?

In the first option, what do you mean by "ignoring that just a few late respondents respond later - the weight will be correct for this"? Does this mean I can simply use the indinus_lw as is?

How is the second option above different from ignoring (as in the first option)? When you say readjust the weight for missingness, what would be the modification of this step based on item 11 in the weighting FAQs?

Since the BHPS is collected only over 3 months, it is impossible to do monthly analysis but annual analysis is possible? This tells me that the only comparable time period for both BHPS and UKHLS is a on a wave basis. How then can this wave analysis be disaggregated from wave 6-18 of BHPS and wave 2-9 of UKHLS for life satisfaction?

#14 - 07/17/2020 08:32 PM - Olena Kaminska

I will start with the second question:
- yes, you can use indinus_lw directly
- in the first option you assume that late respondents responded in your calendar month (which they didn't) - you use them in your analysis. In the second option you delete late respondents and don't use them in your analysis - you apply adjustment for this to our weight;
- If you are interested in comparability over time, you can use only BHPS sample over time through to the last released wave.

Hope this helps,
Olena

#15 - 07/17/2020 08:32 PM - Olena Kaminska

- Assignee changed from Olena Kaminska to Alita Nandi

#16 - 07/19/2020 04:46 PM - Nick Dret

Abigail Dumalus wrote:

> Thanks so much, Alita, for your feedback. Does the 24-month wave period preclude me from using "year" as a panel time variable? I am wondering whether it makes sense to analyse waves 2-9 annually, when the weighting variable is based on these 24-month wave period? Is it possible to construct annual weighting variable or is this going to be more complicated? I am asking this because I need to generate 3-month rolling average / variance for life satisfaction from wave 6 of the BHPS until wave 9 of the UKHLS. For the monthly period, I am using the istrtdatm variable. How should I proceed in this situation? My supervisor (who's a professor) keeps asking me to be more precise with the time period on the x-axis on each chart.

#17 - 07/20/2020 05:57 AM - Alita Nandi

- Assignee changed from Alita Nandi to Abigail Dumalus

Hi Abigail,

You are correct, The answer to your first question "Which variable would indicate those "interviews that took place (finished) in January 2014 regardless of which sample month they come from"?) is intdatd_dv, intdatm_dv, and intdaty_dv.

A few points to note:

- the intdat?_dv variables are the derived interview dates. These include imputed values for any missing date. These are only available for UKHLS and not BHPS
https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/intdaty_dv
- the non-imputed values are istrtdaty istrtdatm istrtdatd. These are available for UKHLS and BHPS
https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/istrtdaty

#18 - 10/13/2021 11:28 AM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- Assignee deleted (Abigail Dumalus)

- % Done changed from 80 to 100