Understanding Society User Support - Support #1820

Cross-sectional vs longitudinal weights

11/29/2022 08:17 AM - Henrique Neves

Status:	Resolved		**Start date:**	11/29/2022
Priority:	High		**% Done:**	100%
Assignee:	Olena Kaminska			
Category:	Weights			

Description

Dear Understanding society support team,

Our research team is using data from the Understanding Society Main Annual Survey (waves 7 to 11) and the COVID-19 study (waves 1 to 9). In our analysis, we want to account for weights. However, we are unsure about which weighs to use.

Our main goal is to analyze gaps in mental health ($scghq1\_dv$) between a Muslim and a Non-Muslim population during COVID-19. We rely on a standard difference-in-differences design, comparing the average Muslim-Non-Muslim gaps in mental health during the pandemic (Covid Survey, waves 1 to 9) with the average pre-pandemic gaps (Main Survey, waves 7 to 11). Our treatment variable takes value 1 for Covid waves and 0 otherwise. Additionally, we run an event study design, comparing Muslim-Non-Muslim gaps in mental health in each wave (Waves 8 to 11 of the Main Survey and Covid waves) relative to wave 7 of the Main Annual Survey.

Unfortunately, the COVID-19 study does not ask about the participants' religion. To identify Muslims in the COVID-19 dataset we extract the last religion status reported on the Understanding Society Main Survey (based on the variable $oprlg1$) and link it with the Covid Survey through person identifiers.

Given our study design would you recommend we use cross-sectional or longitudinal weights?

Thanks in advance for your help!

Kind regards,
Henrique

History

#1 - 12/01/2022 06:36 PM - Understanding Society User Support Team

- Status changed from New to In Progress

- % Done changed from 0 to 10

- Private changed from Yes to No

Many thanks for your enquiry. The Understanding Society team is looking into it and we will get back to you as soon as we can.

We aim to respond to simple queries within 48 hours and more complex issues within 7 working days.

Best wishes,
Understanding Society User Support Team

#2 - 12/02/2022 04:30 PM - Olena Kaminska

Henrique,

Thank you for your question. Just to clarify:
1) do you use wave 7-11 of the mainstage in one model?
2) You mention "Our treatment variable takes value 1 for Covid waves and 0 otherwise". Is it 1 if a person present in a Covid wave? Can you clarify please what you mean by this?

Thank you,
Olena

#3 - 12/05/2022 10:01 AM - Henrique Neves

Thanks for your reply, and my apologies for not being clear. Responding to your questions:

1) We have combined the mainstage and Covid waves in a long format. We have a single model where we compare data from waves 7-11 of the mainstage and waves 1-9 of the Covid-19 Survey (we do not estimate separate models for the mainstage and Covid waves).

2) Our model includes a time dummy (our treatment variable) that takes value 1 for all observations in the Covid waves and 0 for all observations in the mainstage waves. For example: If a person is present in wave 9 of the mainstage and wave 1 of the Covid survey, he/she would have a value of 0 in our time dummy in wave 9 of the mainstage and a value of 1 in wave 1 of the Covid survey.

Let me know if this is clear.

Kind regards,
Henrique

#4 - 12/06/2022 02:17 PM - Olena Kaminska

Henrique,

Just to clarify, by treatment variable do you mean that you are interested in 0/1 effect of your treatment variable? Do you study nonresponse: so an effect of response to the mainstage and nonresponse to Covid survey? Or maybe you are looking into measurement error - a difference in response to the mainstage vs. Covid survey?

If not, could you clarify please as weights will depend on that?
Best,
Olena

#5 - 12/07/2022 01:40 PM - Henrique Neves

Thanks for your reply.

We are indeed interested in a 0/1 effect of our treatment variable. We are simply regressing a mental health variable on this binary (0–1) variable, and the coefficient is interpreted as the average causal effect of Covid-19 on mental health.

I am not sure I understood your question, but it seems that we are not studying non-response and measurement errors.

We are not studying the effect of response to the mainstage and nonresponse to the Covid survey. We are pooling observations across the mainstage and Covid waves and comparing outcomes before and during the pandemic. We don't consider changes in mental health from the mainstage to the Covid survey as measurement errors.

Let me know which further information I should provide.

Kind regards,
Henrique

#6 - 12/07/2022 02:20 PM - Olena Kaminska

Henrique,

Thank you. This clarifies it. Can you tell me how you are using Covid data - do you need a person to participate in all of the waves?

Thank you,
Olena

#7 - 12/07/2022 02:38 PM - Henrique Neves

No, a person just needs to participate in one of the Covid waves.

Kind regards,
Henrique

#8 - 12/07/2022 02:56 PM - Olena Kaminska

Ok, then your best option would be to use wave 11 longitudinal weight as your base weight, and predict response in your model.

Hope this helps,
Olena

#9 - 12/08/2022 01:51 PM - Henrique Neves

Thanks so much for your suggestion. So what you are suggesting would be to fit a conditional model of response, where the dependent variable would be a 0/1 indicator of whether a participant responded to every Covid wave (from 1 to 9)?

Thank you,
Henrique

#10 - 12/08/2022 02:07 PM - Olena Kaminska

Henrique,

Do you mean you want to model first nonresponse to a Covid wave, and then conditional on response your model of interest? It is a possibility, as

long as you understand what you are doing, and what you want to represent.

Hope this helps,
Olena

#11 - 12/09/2022 01:51 PM - Henrique Neves

Sorry, I am a bit confused. My wave-combination of interest is participants at waves 7 to 11 of the mainstage and 1 to 9 of the Covid survey. You suggested I use wave 11 longitudinal weight as the base weight and predict response in my model. But I am not sure how to fit a conditional model (logit) of response to my wave-combination of interest. Should I model nonresponse to both mainstage and Covid waves?

Thank you,
Henrique

#12 - 12/09/2022 02:51 PM - Olena Kaminska

Henrique,

Thank you. This clarifies the question. You should put the model together. If a person is in your model code response=1, if not, code response=0. If you have different models with different sets of people you may need different weights for each.

Hope this helps,
Olena

#13 - 12/16/2022 01:05 PM - Henrique Neves

Sorry for the late reply. Thanks so much for your suggestion, I've worked through it this week! I would just like to clarify two questions to make sure this is the optimal weight strategy for our study:

(1) Using wave 11 longitudinal weight as base weight imposes a relevant restriction to our sample: an individual must have participated in all waves from 1 to 11 of the mainstage to have non-zero longitudinal weight. This departs from our original sample of interest: participants in the COVID-19 survey who had responded to at least one wave of the mainstage from 7 to 11 (so not all from 1 to 11). I am just wondering if there is an alternative base weight closer better suited for our sample.

(2) If in our analysis we pool all observations from waves 7 to 11 of the mainstage and waves 1 to 9 of the COVID-19 survey (treating waves as repeated cross-sections instead of carrying longitudinal analysis) can we use cross-sectional weights?

Thanks so much again!

Kind regards,
Henrique

#14 - 01/19/2023 02:03 PM - Olena Kaminska

Henrique,

Thank you for your questions.

1) You could start at wave 7 longitudinal weight. The gain will be small, but you will have to take mortality into account.
2) In this situation you could just use cross-sectional weights provided with Covid data.

Hope this helps,
Olena

#15 - 02/09/2023 07:54 AM - Henrique Neves

Thanks for all your help!

#16 - 02/23/2023 11:16 AM - Understanding Society User Support Team

- Status changed from In Progress to Resolved

- % Done changed from 10 to 100