

Understanding Society User Support - Support #1678

Weights for the linked COVID-19 Youth data

04/07/2022 04:59 PM - Irina Kolegova

Status:	Resolved	Start date:	04/07/2022
Priority:	Urgent	% Done:	100%
Assignee:	Understanding Society User Support Team		
Category:	COVID-19		
Description			
Hello!			
Hope this email finds you well.			
I am currently working with COVID-19 youth data and I have a question about weights. I have already read the user guide you previously sent me, however, I still have a question.			
1) I have merged COVID youth wave 4 (July 2020) and wave 8 (March 2021). There are two weights available in this datasets (cd_betaryh_xw and ch_betaryh_xw). Do I need to use them both? Or just one of them? And do I need to use any other weights when I am merging to COVID waves for youth?			
2) Then I have another dataset - I have merged (1) COVID youth wave 4 (July 2020), (2) COVID youth wave 8 (March 2021) and (3) baseline UKLHS Wave 10 (2018-19). Which weight I need to use in this case?			
Thank you and looking forward to hearing from you!			
Kind regards, Irina			

History

#1 - 04/07/2022 05:36 PM - Understanding Society User Support Team

- Status changed from New to In Progress

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 - 04/08/2022 02:11 PM - Understanding Society User Support Team

Dear Irina,

Is your analysis cross-sectional or longitudinal?

Best wishes,
Understanding Society User Support Team

#3 - 04/08/2022 02:12 PM - Understanding Society User Support Team

- Status changed from In Progress to Feedback

- % Done changed from 0 to 10

- Private changed from Yes to No

#4 - 04/08/2022 02:12 PM - Understanding Society User Support Team

- Category changed from Data linkage and consents to COVID-19

#5 - 04/08/2022 03:17 PM - Irina Kolegova

As I am using the same respondents who took part in the pre-covid wave, the 4th Covid wave and the 8th Covid wave I think that this analysis is longitudinal. But please correct me if I am wrong.

Best wishes,
Irina

#6 - 04/08/2022 03:27 PM - Irina Kolegova

Let me show my steps:

1) I merged youth who participated in Wave 4 (July 2020) and Wave 8 (March 2021) and kept those who matched observations (N=821):
use `"/Users/irina/Desktop/Thesis/2022/Data/UKDA-8644-stata/stata/stata13_se/cd_youth_p.dta"`
merge 1:1 pidp_c using `"/Users/irina/Desktop/Thesis/2022/Data/UKDA-8644-stata/stata/stata13_se/ch_youth_p.dta"`, nogenerate keep(match))
2) then I merged it to the baseline (wave 10 UKHLS) and also kept matched observations (N=531):
merge 1:1 pidp using `"/Users/irina/Desktop/Thesis/2022/Data/UKDA-6614-stata/stata/stata13_se/ukhls_w10/j_youth.dta"`, nogenerate keep(match)
3) then I want to compare how SDQ score changed over time. For this I want to test absolute differences in SDQ scores between wave 10 UKHLS (2018/19) and wave 4(July 2020) samples using linear regression for continuous outcomes

#7 - 04/08/2022 03:31 PM - Understanding Society User Support Team

Could you please also include the syntax for the regression part?

Best wishes,
Understanding Society User Support Team

#8 - 04/08/2022 04:19 PM - Irina Kolegova

- File Screenshot 2022-04-08 at 16.16.50.png added

I have just realised that I do not know which dependent variable I should use for my regression. I was thinking to get a table like this. Do you know how can I do it?

#9 - 04/10/2022 08:04 PM - Irina Kolegova

Hello,

I have finally managed to figure out the syntax for the regression part. Here is my syntax:

```
// July 2020 + March 2021
use "/Users/irina/Desktop/Thesis/2022/Data/UKDA-8644-stata/stata/stata13_se/cd_youth_p.dta"
merge 1:1 pidp_c using "/Users/irina/Desktop/Thesis/2022/Data/UKDA-8644-stata/stata/stata13_se/ch_youth_p.dta", nogenerate keep(match)
rename pidp_c pidp
save "/Users/irina/Desktop/Thesis/2022/Data/youth data/cd_ch_youth_p.dta", replace

• Adding baseline Wave 10 UKHLS (2019) to July 2020 + March 2021
  use "/Users/irina/Desktop/Thesis/2022/Data/youth data/cd_ch_youth_p.dta"
  merge 1:1 pidp using "/Users/irina/Desktop/Thesis/2022/Data/UKDA-6614-stata/stata/stata13_se/ukhls/i_youth.dta", nogenerate keep(match)
```

```
//// Renaming variables and transforming wide data to long ////
rename cd_* 2020
rename ch_2021
rename i_*2018
```

```
//Reshaping data
reshape long ypsdqtd_dv ypsdqes_dv ypsdqcp_dv ypsdqha_dv ypsdqpp_dv ypsdqps_dv, i(pidp) j(year)
label variable ypsdqtd_dv "Total SDQ"
label variable ypsdqes_dv "Emotional Symptoms"
label variable ypsdqcp_dv "Conduct Problems"
label variable ypsdqha_dv "Hyperactivity/Inattention"
label variable ypsdqpp_dv "Peer Relationship Problems"
label variable ypsdqps_dv "Prosocial Behaviour"
```

```
// Testing changes in TOTAL mean scores
regress ypsdqtd_dv i.year if year==2020 | year==2018
```

```
// Testing changes in subscale mean scores
*Emotional S
regress ypsdqes_dv i.year if year==2020 | year==2018
*Conduct Problems
regress ypsdqcp_dv i.year if year==2020 | year==2018
*Hyperactivity/Inattention
regress ypsdqha_dv i.year if year==2020 | year==2018
*Peer Relationship Problems
regress ypsdqpp_dv i.year if year==2020 | year==2018
**Prosocial Behaviour
regress ypsdqps_dv i.year if year==2020 | year==2018
```

#10 - 04/10/2022 08:05 PM - Irina Kolegova

- File Screenshot 2022-04-10 at 20.05.16.png added

#11 - 04/12/2022 04:06 PM - Understanding Society User Support Team

- % Done changed from 10 to 80

Hi Irina,

It seems that you want to estimate trend in these SDQ summary scales, rather than within person change in these scores even though you have restricted the data to those who have responded to Wave 10, Covid Waves 4 & 8. But as you are not estimating within person change, you don't need to restrict the sample to those who have responded in all 3 waves and can use cross-sectional weights for the respective waves.

Also note that even though you are using "year" as an explanatory variable, the value of "2018" is actually "2018-20" as the interviews for Wave 10 were mostly in 2018 & 2019 plus a few in 2020. You could restrict the Wave 10 observations to only those who were interviewed in 2018 & 2019 so that these observations represented pre-pandemic.

When producing weighted estimates, you can use svyset to take into account weights as well as the complex survey design.

Best wishes,
Understanding Society User Support Team

#12 - 04/12/2022 04:27 PM - Irina Kolegova

Dear Understanding Society User Support Team,

Thank you for your response.

Just a quick remark - I am not using UKHLS Wave 10. For a baseline I use UKHLS Wave 9 (2017-2018).

It sounds good that I don't have to restrict the sample to those who have responded in all 3 waves. Could you please show me in my code where I should apply the "cross-sectional weights for the respective waves"? Do I need to apply weights only when I run regressions? Or should I apply weights in the beginning of the analysis once I've merged the 3 waves?

Could you please also explain me how can I "use svyset to take into account weights as well as the complex survey design"?

Thank you and looking forward to hearing from you

Best,
Irina

#13 - 04/12/2022 05:13 PM - Understanding Society User Support Team

Sorry about that - same principle applies for baseline using Wave 9.

After you have put the data together (as you have shown above) create a variable called weight and replace it with the different wave specific youth cross-sectional weights, depending on the wave.

As weights are about producing unbiased population estimates, apply them when producing population estimates. Here is some information about using weights and complex survey design in the main survey user guide:
<https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/how-to-use-weights-analysis-guidance-for-weights-psu-strata>

Hope this helps

#14 - 04/13/2022 12:57 PM - Irina Kolegova

- File Screenshot 2022-04-13 at 12.55.25.png added

Hello! I have applied the weights for the youth as you recommended, could you please tell me if I have done it correctly? (attached screenshot)

#15 - 04/13/2022 12:58 PM - Irina Kolegova

and do I need to keep only matched observations (when I am merging waves in the beginning) in this case?

#16 - 04/13/2022 05:30 PM - Understanding Society User Support Team

- Assignee changed from Irina Kolegova to Alita Nandi

#17 - 04/14/2022 03:26 AM - Understanding Society User Support Team

- Assignee changed from Alita Nandi to Understanding Society User Support Team

As your objective is to estimate trends, you don't need to restrict to cases that responded in all 3 waves.

#18 - 04/14/2022 03:31 AM - Understanding Society User Support Team

looked t the screenshop - yes, weights created and use of svyset is ok

#19 - 06/06/2022 08:18 PM - Understanding Society User Support Team

- *Status changed from Feedback to Resolved*

- *% Done changed from 80 to 100*

Files

Screenshot 2022-04-08 at 16.16.50.png	99.4 KB	04/08/2022	Irina Kolegova
Screenshot 2022-04-10 at 20.05.16.png	1.04 MB	04/10/2022	Irina Kolegova
Screenshot 2022-04-13 at 12.55.25.png	935 KB	04/13/2022	Irina Kolegova