

Understanding Society User Support - Support #1728

Merging main survey wave to COVID-19 study wave

07/18/2022 10:58 PM - Parth Pandya

Status:	Resolved	Start date:	07/18/2022
Priority:	High	% Done:	100%
Assignee:			
Category:	COVID-19		
Description <p>Hi I am doing a Master's dissertation studying changing markers of adulthood from pre COVID-19 to the initial impact of COVID-19 (employment, leaving home, family and education). I have successfully merged the jk_ main survey wave supplied in the COVID-19 study to the cg_ wave (in long format).</p> <p>However I am finding a lot of the variable names and labels to be different and there are a lot of missing values. Is this a problem for data analysis (specifically pooled OLS regressions)? Are there any help guides/worksheets on going about doing this data analysis? What descriptive statistics would you recommend to show transitions?</p> <p>I see that the user guide for the COVID-19 study says you can do this but there is not much detail on how to actually go about doing this merging.</p>			

History

#1 - 07/19/2022 09:34 AM - Annette Pasotti

- Category set to COVID-19
- Status changed from New to Feedback
- Assignee set to Annette Pasotti
- % Done changed from 0 to 50

Dear Parth

Thank you for your email. The standalone syntax files explaining how to link variables across waves are available on our website here: <https://www.understandingsociety.ac.uk/documentation/mainstage/syntax> - take a look at the syntax files under "Merging individual files across waves into long format". These are in Stata. If you are using a different software we have these available via our Moodle course "Introduction to Understanding Society" which explains the steps you need to do and why. <https://www.understandingsociety.ac.uk/help/training/online/introduction-course>

The online course is free, you just need to register first. One of the sections deals with merging data into wide and long format. Please take a look at the worksheet (which guides you step by step through the process) and the accompanying do file and log file. If after doing that you still have questions, please let us know.

Searching the user forum can be useful in finding answers to questions which are similar, such as in this issue <https://iserredex.essex.ac.uk/support/issues/1648>

#2 - 07/20/2022 08:56 AM - Parth Pandya

- File Screen Shot 2022-07-20 at 08.56.30.png added

Thanks for your reply. I have been able to merge waves into long format as my message above (I can see two rows for each pidp in Data Editor of STATA). My specific query was that my variables are not aligning i.e. there are significant missing variables for one row compared to the other. This is because variables are different in the jk_ main survey wave and the cg_ COVID-19 study wave. Is there a way to fix this?

In the screenshot below you can see missing values indicated by a .

#3 - 07/21/2022 05:54 PM - Understanding Society User Support Team

- Assignee deleted (Annette Pasotti)
- % Done changed from 50 to 90
- Private changed from Yes to No

Dear Parth,

The missing values on the variables you included on the screenshot result from the way in which you created your dataset, for instance, there is no racel_dv variable in the jk_ dataset (only racel) hence there is no way Stata could match it with the racel_dv in cg_indresp. My advice would be to

revisit your data management plan:

- 1) check available variables in both datasets and decide which of these are comparable and could be matched, note that in some cases a straightforward equivalent might be not available. Additionally, remember that some baseline Covid-19 survey variables might give you more precise information about the pre-pandemic situation that the mainstage variable (so either from the jk_ dataset, or taken from source mainstage datasets, for example, j_ or k_indresp)
- 2) after you have identified the variables, rename them in both datafiles to match
- 3) link stable characteristics (racel_dv, sex_dv, bornuk_dv) at the end of data management, ideally from xwavedat, link it to your longfile using "merge m:1 pidp". This way you will avoid the kind of missing values as these on the screenshot.

I hope it helps.

Best wishes,

Understanding Society User Support Team

#4 - 11/30/2023 10:58 AM - Understanding Society User Support Team

- Status changed from Feedback to Resolved
- % Done changed from 90 to 100

Files

Screen Shot 2022-07-20 at 08.56.30.png	72.2 KB	07/20/2022	Parth Pandya
--	---------	------------	--------------