Understanding Society User Support - Support #270

Weighting

06/16/2014 06:34 PM - Victoria Peacey

Status:	Closed		**Start date:**	06/16/2014
Priority:	High		**% Done:**	100%
Assignee:				
Category:	Data analysis			

Description

Hi

I am using the first three waves of US data to look at job loss and associations with mental health and self-rated health. I think that I will restructure the three waves of data into 'time 1' and 'time 2' data (before job loss and after job loss, whether job loss occurs between waves 1 and 2 or between waves 2 and 3). By doing this I can pool together all the people who lost their job between wave 1 and wave 2, and between wave 2 and wave 3, and compare them with people who were continuously employed for two or more waves.

I want to do this to simplify the analysis so I can use logistic regression to look at the effect of job loss vs continuous employment on (dichotomised) health outcomes, taking health at baseline into account in the model. Perhaps you will think it is a terrible idea, in which case please let me know. My main worry though is that I don't know where to start with using weighting once the data has been reorganised in this way. I need the self-completion weights as the health outcome data is from the self-completion section, but do I need the longitudinal weight? And should I use the weight from the wave which is 'time 1', or something different?

Sorry this is rather complicated
Best wishes
Vicki

History

#1 - 06/17/2014 12:18 PM - Olena Kaminska

Vicky,

Your issue is not with weighting, but rather with the population that you are interested to represent and with units of analysis. This is a common issue with pooled analysis and needs a lot of attention if such is attempted.
Here is my suggestion for your situation, but there are a number of other possibilities:
- your unit of analysis can be change
- indeed, you can calculate change between waves 1-2 and waves 2-3. In this way you will have two datasets, first with the longitudinal weight for wave 2, and the second, for wave 3 (you need longitudinal weights as you are looking at change) and you can stack these together (pool);
- in the above situation you will have a variable that indicates change or no change for each person;
- following the above steps you can talk about changes at the time of waves 1-3.
- in my opinion, pooling people who changed for two waves, and those who didn't change for three waves - introduces serious interpretation problems for units of analysis. I would advice against this approach.
- importantly, pooled analysis requires correction for clustering (nesting) within people (you observe the change twice within the same person - and the statistical program needs to know about it).
- and finally, in your report / paper it is important to not talk about people (you are not studying people any more), but to talk about year-on year changes over a time period.

Overall, pooled analysis can complex both for implementation and interpretation. Studying change between waves 1-2, and separately between waves 2-3 is a much cleaner and easier to interpret approach.

Hope this helps,
Olena

#2 - 06/17/2014 12:56 PM - Redmine Admin

- % Done changed from 0 to 50

#3 - 06/18/2014 12:29 PM - Victoria Peacey

Dear Olena
Thanks so much for the quick reply.

Studying and reporting the waves separately is definitely a cleaner approach, I agree. I have been reluctant to do this because there are only 350 people who lost their jobs involuntarily between wave 1 and 2, and 379 between wave 2 and 3 (I am focusing on involuntary job loss without re-employment by the time of interview, as it makes less complicated). Analysing the waves separately shows a marginally significant association of job loss with subsequent declining self-rated health in both analyses, which suggests that if I analyse them together then the effect would be more

robust. However analysing the association of job loss with mental health change does give significant effects when the waves are separately analysed, which is probably not that surprising given that you might expect job loss to have more of an impact on mental health than overall general health.

I appreciate what you are saying about including the same person twice in the case of pooled analysis. This wouldn't happen in the case of people who lose their jobs, because if you lose your job between wave 1-2 you wouldn't be re-employed at the time of the wave2 interview, so you couldn't be continuously employed in wave 2-3, and people who lost their job in wave 2-3 could have any continuous employment in wave 1-2 disregarded. But where people are continuously employed in all three waves, there would be some people who were counted twice – as continuously employed both in waves 1-2 and waves 2-3. If I randomly allocated them to only be included once, as either wave1-2 or wave 2-3, would that work – because then no-one is counted twice when the two separate analyses are pooled? And could I then carry on talking about people, rather than year-on-year change?

On a slightly different issue, may I ask why you suggest modelling change in health, rather than health at follow-up including health at baseline?

Many thanks. I really appreciate your input.
Vicki

#4 - 06/23/2014 05:29 PM - Olena Kaminska

Vicky,

Yes, your situation is a clear example where pooling helps - i.e. with such a rare population that you study getting information from multiple waves improves your statistical power.

My suggestion of the units of analysis is only one of many possible ones. But it is important to define your population (people or events and the time range), and keep it very clear in your analysis. Another simple solution to your situation may be people who lost job involuntarily at least once between waves 1 and 3 versus people who continuously were employed. This way you should have 700 in the first category (if I understand it correctly). But again, whatever you decide to do, you have to be very clear about the unit of analysis - your suggested approach makes me wonder how you would describe the unit of analysis.

Finally, I didn't mean to comment on any substantive topic - so ignore my reference to health.

Best wishes,
Olena

#5 - 07/03/2014 10:40 AM - Redmine Admin

- Category set to Data analysis

- Status changed from New to Closed

- Target version set to X M

- % Done changed from 50 to 100