

Understanding Society User Support - Support #1951

Zero longitudinal weights

08/08/2023 09:33 AM - Isabel Hopwood

Status:	Resolved	Start date:	08/08/2023
Priority:	High	% Done:	100%
Assignee:	Understanding Society User Support Team		
Category:	Weights		
Description			
Hi, I am running some panel data regressions examining the change in mental health outcomes when an individual moves house. I have included longitudinal weights for this analysis. I am using the svydesign function in R to apply probability weights before I run the regressions. For the individuals that have zero weights, the probability weights are "INF" so when I run the regressions the p values are not defined. I have noticed that I can get round this issue by replacing the zero weights with a very small value such as "0.00000001". Will this make my analysis wrong if I do this?			

History

#1 - 08/08/2023 05:12 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 - 08/08/2023 05:14 PM - Understanding Society User Support Team

- Assignee changed from Understanding Society User Support Team to Olena Kaminska

#3 - 08/11/2023 11:53 AM - Understanding Society User Support Team

- Status changed from In Progress to Feedback
- Assignee changed from Olena Kaminska to Understanding Society User Support Team
- % Done changed from 10 to 50

Hello Isabel,
Our team responsible for weights provided us with the following response

To give you a comprehensive answer I would need some more information. It is unclear how you integrate the survey package into your analysis.

The survey package is ready to deal with zero weights. Therefore, if you use a function from the package (e.g., svyglm), it will ignore the observations with zero weights.

It seems you extracted the "prob" element from the svydesign list, a vector with the inverse weight for each observation in your dataset. This explains that the cases with zero weights are set to Inf in the "prob" vector of the svydesign object. However, the recommendation is to use the weights – not the inverse of the weights – in your analysis.

I hope this information is helpful.
Best wishes,

Roberto Cavazos
Understanding Society User Support Team

#4 - 08/13/2023 02:33 PM - Isabel Hopwood

Hi Roberto,

Thanks for your response. I integrate the survey design weights by extracting the longitudinal weight from the final wave of data that I am using. I then merge this weight with the rest of the waves of data for each pidp. Then I use the svydesign object to create a survey weighted dataset.

Here is the code:

```
svy_move_less_green <- svydesign(id = ~psu,  
strata = ~strata,  
weights = ~indin91_lw,  
data = move_less_green_complete2)
```

I then use the `svyglm` function to run regressions on this dataset. For example, `summary(svyglm(scghq2_dv~year_moved+log_income, design=svy_move_less_green, rescale=FALSE))`

How do I ensure that I am using the weights not the inverse of the weights? If the `svyglm` ignores the values with 0 weights, doesn't this reduce the sample size of individuals in the regressions by a lot? Also there are a lot of missing PSU values. I remove the observations with missing PSU values in order to apply weights to the dataset. Is there anything I can do to save these values?

I hope this all makes sense. Sorry for all the questions! I am new to using survey weights in analysis and my regression results are currently very strange so I want to ensure that I'm not doing something wrong.

Many thanks
Izzie

#5 - 08/22/2023 09:38 AM - Understanding Society User Support Team

Hello Isabel,

Our team responsible for weights provided us with the following response.

The R survey package can handle complex sample design and related issues such as zero-weights. To give you a complete answer, I would need to know more about the target population of your research, the objectives of your analysis, the time scope of your research (which waves it covers), and in which waves your outcome variable was collected.

I answer your questions below:

How do I ensure that I am using the weights not the inverse of the weights?

When fitting the model, the survey package uses the weight you specify in the `svydesign` function. In your example, the model would be fitted using the "indin91_lw" weight.

If the `svyglm` ignores the values with 0 weights, doesn't this reduce the sample size of individuals in the regressions by a lot?

The sample size for your analysis will depend on the definition of your target population and other analysis characteristics, such as the number of waves you intend to cover.

Sample members with a zero weight are partly excluded from the estimation. The survey R package ignores the observations with zero weights to compute the point estimates (e.g., model coefficients), but they can affect the variance estimation (e.g., standard errors of the model coefficients). It is important to note that sample members are given a zero weight for a reason (see pages 11-13 of the Weighting FAQs for a detailed explanation).

In your example above, you use the longitudinal weight for the GB sample recruited in 1991 (`indin91_lw`). One reason that explains the zero weights is that longitudinal weights in Understanding Society are developed for monotone attrition, meaning that sample members are given a longitudinal weight (`_lw`) when they respond to the adult interview in a number of consecutive waves. Thus, sample members who failed to respond to the adult interview at one or more waves are given a zero weight.

Also there are a lot of missing PSU values. I remove the observations with missing PSU values in order to apply weights to the dataset. Is there anything I can do to save these values?

Please, could you provide more details about this issue?

The PSU variable should be populated for all panel members. You can find the PSU variable in the `xwaveid` or `xwavedat` datasets or other wave-specific datasets (i.e., `hhsamp`, `indresp`, `indsamp`, `egoalt`, `hhresp`, `indall`...).

I hope this information is helpful.

Best wishes,

Roberto Cavazos
Understanding Society User Support Team

#6 - 11/30/2023 01:23 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 50 to 100