

Understanding Society User Support - Support #1477

clusters not nested in b_strata (using R and svydesign object)

01/07/2021 11:16 AM - Marion Lieutaud

Status:	Resolved	Start date:	01/07/2021
Priority:	High	% Done:	100%
Assignee:			
Category:			
Description Dear Understanding Society team, I have an issue with the combination of clusters and strata in the Wave 2 data. I have tested the exact same code for Wave 1 data, and it works - it is only when using b_strata that the issue arises. Here is the R code. <code>*b_indresp <- read.dta("b_indresp.dta")</code> <code>library(survey)</code> <code>options(survey.lonely.psu = "adjust")</code> <code>options(na.action="na.pass")</code> <code>b_w <- svydesign(id = ~b_psu,</code> <code>strata = ~b_strata,</code> <code>weights = ~b_indinub_xw,</code> <code>data = b_indresp)*</code> this gives me the following error message: <i>"Error in svydesign.default(id = ~b_psu, strata = ~b_strata, weights = ~b_indinub_xw, : Clusters not nested in strata at top level; you may want nest=TRUE."</i> If I do add nest=TRUE: <code>b_w <- svydesign(id = ~b_psu,</code> <code>strata = ~b_strata,</code> <code>weights = ~b_indinub_xw,</code> <code>data = b_indresp,</code> <code>nest=TRUE)</code> I get no error message at this point, but an intractable list, not a svydesign object that would work for any {survey} function. The problem persisted when I tried using other cross-sectional weights for Wave 2, i.e b_indinus_xw. However, if instead of using Wave 2 data and variables, I use Wave 1 instead, it all works perfectly. <code>*a_indresp <- read.dta("a_indresp.dta")</code> <code>a_w <- svydesign(id = ~a_psu,</code> <code>strata = ~a_strata,</code> <code>weights = ~a_indinus_xw,</code> <code>data = a_indresp)*</code> Additionally, the code for Wave 2 works if I remove the strata element, i.e.: <code>b_w <- svydesign(id = ~b_psu,</code> <code>weights = ~b_indinub_xw,</code> <code>data = b_indresp,</code> <code>nest=TRUE)</code> I wonder whether there is an issue with b_strata and the way I use it, and where I have gone wrong..? I hope you can help! Thank you very much in advance for your time and support!			

History

#1 - 01/07/2021 02:26 PM - Alita Nandi

- Status changed from New to Feedback
- Assignee set to Marion Lieutaud
- % Done changed from 0 to 80
- Private changed from Yes to No

Hello,

This is possible because of single PSU strata. While this is not the case by design, for a particular analysis the sample that is selected could include strata with only single PSUs preventing the calculation of standard errors where you have specified a clustered and stratified design. In our training material, this issue is dealt with. Please sign up for the MoodleX course "Introduction to Understanding Society Using R". Instructions on how to do that is here: <https://www.understandingsociety.ac.uk/help/training/online/introduction-course>
When you enroll into this MoodleX course, look at the Module "Working with data collected from surveys with complex survey designs" where you will find the relevant worksheet with details on how to solve this issue as well as the accompanying R syntax file.

Hope this helps.
Best wishes,
Understanding Society User Support Team

#2 - 01/07/2021 02:32 PM - Marion Lieutaud

Dear Understanding Society team,

Thank you for your reply.
I am surprised because I had registered for and had looked through the module course you mention, and I had followed the instructions regarding single PSUs - namely to include this line in the code:

```
options(survey.lonely.psu = "adjust")
```

I included this in my code and in the example above, and I am not aware that another way of dealing with this was provided? Did I miss something? My apologies if I did!

Thank you very much again!

Marion

#3 - 01/07/2021 03:59 PM - Alita Nandi

- Status changed from Feedback to In Progress
- % Done changed from 80 to 10

Sorry I didn't notice that you had included the lonelypsu correction. We are looking into this.

#4 - 01/07/2021 04:19 PM - Marion Lieutaud

Thank you! Perhaps I should mention that I use the special license data. So the first line of the code is actually:
`b_indresp <- read.dta("b_indresp_protect.dta")`
it is probably of no relevance to the problem at hand, but just in case!

#5 - 01/08/2021 01:39 PM - Alita Nandi

- Status changed from In Progress to Feedback
- % Done changed from 10 to 80

Hello Marion,

My colleague has checked your code and that is ok. We have identified that for 2 cases in `b_indresp`, for which `b_psu=574`, the value of `b_strata` is 130 instead of 122 as for rest of the cases with `b_psu`. This is causing the error. We are looking into this and will resolve this at the next release. In the meantime please change the value of `b_strata` TO 122 if `b_psu=574` and `b_strata=130`.

We expect this to resolve the error you are encountering. Please let us know if this still does not resolve the problem.

Best wishes,
Understanding Society User Support Team

#6 - 01/08/2021 06:11 PM - Marion Lieutaud

Dear Understanding Society support team,
a thousand thanks for identifying the problem - that explains it all!
I have only one question left: when I checked the data, I found that in `b_indresp`, among the 10 individuals in `b_psu 574`, 8 are in `b_strata 130`, and 2

are in b_strata 122.

Should I be correcting the 2, and therefore make it so that all within b_psu 574 are in b_strata 130?

Or should I stick to your instructions, and change the 8 (the majority in the psu) to b_strata 122?

This is my last bit of confusion. I hope you can clear it, and I just want to add that I cannot thank you enough for your time and help thus far!

Best wishes,
Marion

#7 - 01/10/2021 02:22 PM - Alita Nandi

- % Done changed from 80 to 90

Hi Marion - We have identified that the value of strata=122 should be changed to 130 if horig=4. Please make this change and run your program. We will implement this change at the next release.

#8 - 01/19/2021 12:08 PM - Marion Lieutaud

Dear Alita,

I can confirm that this solved the problem completely!
Thank you again!

best wishes,
Marion

#9 - 01/20/2021 03:50 PM - Alita Nandi

- Status changed from Feedback to Resolved

- Assignee deleted (Marion Lieutaud)

- % Done changed from 90 to 100