

Understanding Society User Support - Support #1198

Weighting data Wave 8 h_indresp

06/12/2019 08:47 AM - Sarah H

Status:	Resolved	Start date:	06/12/2019
Priority:	Normal	% Done:	100%
Assignee:	Sarah H		
Category:	Survey design		
Description			
Hi			
<p>The US provides instructions on pages 65 to 71. Section 3.3 to choose the correct Variable to Weight the data. I have selected n_indpxui_xw because I'm using the individual dataset and not excluding proxy responses. However, the naming convention on page 71 suggests that n_indpxus_xw would be the correct weight? This variable is not available in the dataset that I'm using. N_indpxus_lw is indeed available but LW stands for longitudinal data and I'm only doing analysis of one wave. therefore this is not longitudinal. Instead XW weights that stand for cross-sectional data, i.e. using one wave cross-sectional data.</p> <p>n_indpxui_wv is the only variable that is cross-sectional that is available in the dataset I have selected this as the correct weight to use? Can you confirm this please</p> <p>best wishes</p> <p>Sarah</p>			

History

#1 - 06/12/2019 10:37 AM - Stephanie Auty

- Category set to *Weights*
- Assignee set to *Olena Kaminska*
- Private changed from *Yes* to *No*

#2 - 06/12/2019 11:40 AM - Olena Kaminska

Sarah,

Yes, I can confirm that n_indpxui_xw is the correct weight to use in your analysis. Thank you for pointing this issue to us and apologies for the confusion.

Thanks,
Olena

#3 - 06/12/2019 11:48 AM - Stephanie Auty

- % Done changed from *0* to *80*

#4 - 06/12/2019 11:48 AM - Sarah H

Thank you Olena

Best wishes

Sarah

#5 - 06/12/2019 11:49 AM - Stephanie Auty

- Status changed from *New* to *Feedback*

#6 - 06/16/2019 05:38 PM - Sarah H

Hi

When I use this weight I get an error message >Warning # 3211

On at least one case, the value of the weight variable was zero, negative, or missing. Such cases are invisible to statistical procedures and graphs which need positively weighted cases, but remain on the file and are processed by non-statistical facilities such as LIST and SAVE.

however, the variables do not have zero in them necessarily. So do we assume that this data cannot be weighted and therefore cannot be generalizable ?

#7 - 06/18/2019 03:36 PM - Olena Kaminska

Dear Sarah,

Yes, this message is correct but you don't need to worry about it. Understanding Society has a much more complex design than most surveys - hence zero weights. Having said that, any analysis that uses weights will be generalizable to a population and you should just ignore this message.

Best,
Olena

#8 - 06/18/2019 03:41 PM - Sarah H

Hi Olena

I'm applying this weight to analysis conducted on data that respondents are resident in England only. Is this still correct?

Best wishes

Sarah

#9 - 06/18/2019 03:43 PM - Olena Kaminska

Sarah,

Yes, definitely.

Olena

#10 - 06/18/2019 06:45 PM - Sarah H

Hi Olena

when I apply the weight, shouldn't expect the counts to shift and the frequencies to be higher? They do not change when the weight is applied

Best wishes
Sarah

#11 - 06/19/2019 11:05 AM - Olena Kaminska

Sarah,

Which syntax are you using? Our weights are not frequency weights so please do not use fweights command. Our weights are probability weights and pweights should be used.

Hope this helps,
Olena

#12 - 06/20/2019 09:50 AM - Sarah H

Hi Olena

I am using `n_indpxui_xw` using the `h_indresp` database (individuals and proxies) looking at England geographical area only

#13 - 06/20/2019 10:54 AM - Olena Kaminska

Sarah,

I was referring to the syntax in Stata, for example `svy` command and `pw=n_indpxui_xw` within this command. Is it how you use our weights?

Thanks,
Olena

#14 - 06/20/2019 11:27 AM - Sarah H

I am using SPSS. Does this change anything that you have responded above in terms of error message and weighting variable?

#15 - 06/20/2019 12:49 PM - Olena Kaminska

Sarah,

I suggest that you use proportions, and not frequencies, that SPSS gives you. If you need to get population frequencies - just multiple the SPSS weighted proportions by the population total.

Hope this helps,
Olena

#16 - 06/26/2019 08:56 AM - Alita Nandi

- Assignee changed from Olena Kaminska to Sarah H

#17 - 06/27/2019 03:29 PM - Sarah H

Hi

I think this has gone on to a different topic, to confirm that if I use the `n_indpxui_xw` weight in SPSS, this makes the data representative? an I should ignore the error message 3211

#18 - 06/27/2019 03:39 PM - Alita Nandi

- Assignee changed from Sarah H to Olena Kaminska

#19 - 06/27/2019 03:48 PM - Olena Kaminska

Yes, to both questions.

#20 - 07/02/2019 11:06 AM - Sarah H

Thank you Olena.

To follow on from this, I've conducted some analysis using the weight and results I would expect to find are not coming out. e.g. ethnicity difference in terms of employment permanency. This proved to be significant for my sample when the data was analysed unweighted but weighted there is no significance. I am wondering if I have done something wrong such as clean the variable (throughout my analysis I've excluded non responses/ missing/na/don't know)

Please let me know should you require any further information
Best wishes,
Sarah

#21 - 07/02/2019 11:39 AM - Olena Kaminska

Sarah,

What was the sample size in your analysis before and after weighting, and what was the p-value?

Thank you,
Olena

#22 - 07/02/2019 11:47 AM - Sarah H

Hi

Fishers exact test performed.

Sample size before weighting $n=962$ $p=.027$

Sample size after weighting $n=1015$ $p=.631$

Best wishes

Sarah

#23 - 07/02/2019 11:54 AM - Olena Kaminska

Sarah,

This doesn't sound right: I expect the total with weights to be smaller than in unweighted analysis. I also don't expect such a large difference between p-values. Are you sure that the definition of the variables is the same between the models? You should not take out nonrespondents etc. with weights - the weights do the job for you. You should have the same coding of all the variables as well to make a fair comparison.

Best wishes,
Olena

#24 - 07/02/2019 12:09 PM - Sarah H

The definition of the variable is the same. However, the variable i'm using is ethnicity and as there are so many, I've created a new variable with only 3 ethnicities to so that a test could be conducted.

with regards to you second point, I've performed another test in SPSS with two variables: variable 1) samples of interest (2 groups), variable 2) contract type e.g. perm/temp which was not cleaned and the weighting variable had not removed the inapplicable, don't know, refusal from the table, hence why I've cleaned all the variables i'm using so these don't appear in the tests.

#25 - 07/02/2019 12:17 PM - Olena Kaminska

Try to exclude the variable reflecting samples of interests. The weight takes into account samples correctly on your behalf. It is very easy to get a non-representative results if you exclude some of the samples.

#26 - 07/02/2019 12:22 PM - Sarah H

Dear Olena

Have I missed something? I am only looking at a very specific sub sample of the dataset (e.g. specific age and caring responsibility). Are you suggesting that I cannot use weights for this? Thus it can't be generalizable to this group.

Best wishes

Sarah

#27 - 07/03/2019 11:49 AM - Olena Kaminska

Sarah,

I see. No, your subgroup definition is fine and yes, our data will represent this subgroup. I misunderstood thinking that you may have selected some of the samples (like general population sample or ethnic minority boost). So, this isn't a problem.

The difference between weighted and unweighted results must be explained by the difference in definitions of the variables then.

Hope this helps,
Olena

#28 - 07/03/2019 11:56 AM - Sarah H

Hi Olena

Thanks for your message. No I've used the exact same variables as I am using the same dataset but just applied the weights.

With regards to removing the responses, such as don't know, not applicable etc - could I enquire about this? I have cleaned the variables to remove this (by creating a new variable and coding these as missing). you suggested earlier that this would have been done automatically, however, it does not happen automatically.

Best wishes

Sarah

#29 - 07/03/2019 12:01 PM - Olena Kaminska

Sarah,

The weighting deals only with unit nonresponse. Don't knows and refusals can be treated as valid answers or can be recorded as missings, so yes you should record this yourself. Have you used the same recorded variables in your unweighted analysis too? It may be that this recording influences your analysis and p-value and the change is largely due to your coding of item missingness and has nothing to do with weighting.

Olena

#30 - 07/03/2019 04:38 PM - Sarah H

- *File weighting and ethnicity.docx added*

Hi Olena

Yes I have used all the same variables when I did the analysis weighted/unweighted.

however, I have done an experiment with my sub group and cross tabbed it with the ethnicity variable which I have not cleaned as a test. The results are very different, which I find surprising but would explain the difference in the P values? but does not explain the huge difference in percentages.

#31 - 07/04/2019 10:09 AM - Olena Kaminska

Sarah,

Unfortunately I don't think I know the reason for the difference between weighted and unweighted analysis that you observe. I can only say that the difference you observe is definitely wrong: there is no situation in which sample size can go up when you use a weight. The p-value chance looks very suspicious too. My guess is that weighting has little to do with the difference you observe.

I hope this helps,
Olena

#32 - 07/04/2019 12:11 PM - Olena Kaminska

Sarah,

I just want to clarify my previous comment. It relates to your earlier results of Fisher exact test.

I only now noticed your attachment of the weighted and unweighted distribution by ethnic group - and that's all fine and as expected. I am afraid this doesn't explain the earlier difference.

Thanks,
Olena

#33 - 07/04/2019 01:49 PM - Sarah H

Hi Olena

This is most strange and I've re-done the test again and appears the N does go up for weighted data in the Fisher exact test whereas for all my other analysis weighted it goes down (sample size drops). Do you have any other explanation as to why this is? I'm a bit concerned with the difference in weighting. To confirm this has nothing to do with me coding out non-responses?

Best wishes
Sarah

#34 - 07/04/2019 04:37 PM - Olena Kaminska

Sarah,

I am puzzled myself. But if you have a syntax that you can share I will be happy to look.

Thanks,
Olena

#35 - 07/05/2019 09:57 AM - Sarah H

- *File syntax understanding society ethnicity.docx added*

Hi Olena

Thank you. I've attached three pages in a word document. The first page has the syntax and the second and third the outputs with the Fishers Exact test

Best wishes

Sarah

#36 - 07/08/2019 04:09 PM - Olena Kaminska

Sarah,

I see the problem. The way you specified weights in SPSS they are assumed to be frequency weights. Our weights are probability weights, centred around 1. What SPSS seem to do is for those who have a weight over 1 it gives them the weight of 1, and for the other half of people with weights below 1, it gives the weight of 0. In other words your results come from an unweighted non-random half of a sample and the value you obtain is wrong.

You should use Complex Sample module in SPSS. Please specify cluster and strata in addition to the weight in your analysis. Here is the link to more information:

https://www.spss.ch/upload/1071150823_SPSS%2012%20Complex%20Samples.pdf

Best wishes,
Olena

#37 - 07/15/2019 01:03 PM - Sarah H

Thank you Olena for this information

Can I confirm that if I do not do hypothesis testing and only cross tabulation that weighting can still be correct in SPSS (e.g. not using complex sampling module)

Best wishes

Sarah

#38 - 07/15/2019 01:12 PM - Olena Kaminska

Sarah,

No, cross-tabulation will be wrong without weighting.

But if you are not presenting confidence intervals (though you should), you can ignore clustering and stratification.

Best wishes,
Olena

#39 - 07/18/2019 02:48 PM - Sarah H

Hi Olena

What is the cluster variable I should use? I have used the strata variable

Best wishes

Sarah

#40 - 07/18/2019 03:25 PM - Olena Kaminska

Sarah,

Cluster variable is called psu.

Best wishes,
Olena

#41 - 07/18/2019 04:15 PM - Sarah H

Hi

thank you. I'm coming up with an error message that says my weight is being ignored and a warning message on the output that 'the weight is being ignored'. would you be able assist me what I've done wrong? I have also copied the syntax

- Sampling Wizard.
CSPLAN SAMPLE
/PLAN FILE='C:\Users\sh33496\OneDrive - The Open University\Data\Secondary Data '+
'Analysis\Understanding Society\Datasets\understandingsocietyfinaltrial123.csplan'
/PLANVARS SAMPLEWEIGHT=SampleWeight_Final_ PREVIOUSWEIGHT=h_indpxui_xw
/PRINT PLAN
/DESIGN STAGELABEL='ussamplefinaltrial' STRATA=h_strata CLUSTER=h_psu
/METHOD TYPE=SIMPLE_WOR ESTIMATION=DEFAULT
/RATE VALUE=1
/STAGEVARS INCLPROB CUMWEIGHT.
CSSELECT
/PLAN FILE='C:\Users\sh33496\OneDrive - The Open University\Data\Secondary Data '+
'Analysis\Understanding Society\Datasets\understandingsocietyfinaltrial123.csplan'
/CRITERIA STAGES=1 SEED=RANDOM
/CLASSMISSING EXCLUDE
/PRINT SELECTION.

best wishes and many thanks for your assistance

Sarah

#42 - 07/22/2019 10:59 AM - Alita Nandi

- Assignee changed from Olena Kaminska to Sarah H

Dear Sarah,

Sorry for the delay in getting back to you. We generally provide support and advice on data (& weights) issues only. We provide guidance on syntax related to data management. Your query seems like a problem with SPSS syntax relating to how weights need to be specified.

You could look for solutions online. You could also send a query to our JISC mail group which has been set up for Understanding Society data users to discuss analysis (incl syntax related to analysis) issues. If you want to sign up please send an email to UKHLS-REQUEST@JISCMAIL.AC.UK

Best wishes,
Alita

#43 - 09/03/2019 12:17 PM - Sarah H

Dear Alita

I come to you with another question that I hope you can help with

I just want to confirm what Strata stands for, and PSU. i'm still struggling with complex sampling and going back to basics

best wishes

Sarah

#44 - 09/03/2019 04:14 PM - Alita Nandi

When a population is divided into mutually exclusive and exhaustive parts and then samples are chosen from each of these parts - known as strata - then we have stratified sampling. When a population is divided into mutually exclusive and exhaustive parts but samples are chosen from some of these parts - known as clusters or primary sampling units (PSU), then we have clustered sampling. These are very simplified explanations. There are different types of clustering and stratification. For example there are explicit and implicit stratification, multi-stage clustering where you will have primary and secondary clusters etc. You will find explanations of these concepts in many standard Statistics books. For example, Levy and Lemeshow "Sampling of Populations: Methods and Applications"

Hope this helps,
Alita

#45 - 10/17/2019 12:40 PM - Sarah H

Hi

what are the implications for the data analysis if I conduct the analysis without weighting the data using complex sampling method/ module?

Best wishes

Sarah

#46 - 10/24/2019 03:51 PM - Sarah H

Hi

what are the implications for the data analysis if I conduct the analysis without weighting the data using complex sampling method/ module?

Best wishes

Sarah

#47 - 03/02/2021 03:47 PM - Understanding Society User Support Team

Here is information on why weights should be used.

<https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/why-use-weights>

#48 - 08/10/2022 11:50 AM - Understanding Society User Support Team

- *Category set to Survey design*

- *Status changed from Feedback to Resolved*

- *% Done changed from 80 to 100*

Files

weighting and ethnicity.docx	17 KB	07/03/2019	Sarah H
syntax understanding society ethnicity.docx	108 KB	07/05/2019	Sarah H