

## Understanding Society User Support - Support #2196

### Weighting

01/21/2025 02:14 PM - Evie Gates

<b>Status:</b>	Feedback	<b>Start date:</b>	01/21/2025
<b>Priority:</b>	High	<b>% Done:</b>	90%
<b>Assignee:</b>	Understanding Society User Support Team		
<b>Category:</b>	Weights		
<b>Description</b>			
<p>Hi there,</p> <p>I am analysing wave 1 and 2 of Understanding society data and my sample is restricted to ethnic minorities, those who completed the extra 5 minutes questions and those who have ever consumed alcohol. From reviewing the weighting guidance I am using the "b_ind5mus_lw" weight. I selected this weight as wave 2 (b) is the last wave in my analysis, I am analysing individuals (ind), the data I am using is from both the extra five minutes and self completion questionnaires and as the extra five minutes questionnaire is the lowest level of analysis the "ind5m" was selected, my sample is made up of GPS and EMB members (us) and the analysis in longitudinal (lw).</p> <p>However, when I have used the svyset commands in stata to account for PSU, strata and weight the population size that is reported is substantially lower than my sample size (sample size - 956, population size - 123). I have looked around on stata forums to see why this would be the case and it is stated that this is due to the wrong weighting being used, however as far as I can tell I am using the correct weight.</p> <p>Please could you inform me why the population size would be so low?</p> <p>Many Thanks,</p> <p>Evie Gates</p>			

### History

#### #1 - 01/21/2025 04:14 PM - Understanding Society User Support Team

- Category set to Weights
- Status changed from New to In Progress
- % Done changed from 0 to 10

Hi Evie,

Would you mind sending the code that reproduces the creation of your target group and weighting?

Thanks,  
Piotr Marzec  
UKHLS User Support Team

#### #2 - 01/21/2025 04:29 PM - Evie Gates

- File Do file for US support.do added

Understanding Society User Support Team wrote in [#note-1](#):

Hi Evie,

Would you mind sending the code that reproduces the creation of your target group and weighting?

Thanks,  
Piotr Marzec  
UKHLS User Support Team

Hi Piotr,

Thanks you for getting back to me so quickly!

Please see the attached stata do file which details the code I used to create my target sample and how applied the weighting to said sample.

Evie Gates

**#3 - 01/21/2025 04:45 PM - Understanding Society User Support Team**

Hi Evie,

Thanks. Could you also include the parts that open the data files and create/rename the variables you use? (These variables are not available in the released data under these names.)

Thanks,  
Piotr

**#4 - 01/21/2025 05:04 PM - Evie Gates**

- File Do file for US support.do added

Hi Poitr,

Apologies, please see the updated do file. If it would be helpful I can also attach my dataset.

Many Thanks,

Evie

**#5 - 01/27/2025 09:46 AM - Understanding Society User Support Team**

- Status changed from In Progress to Feedback

- % Done changed from 10 to 90

- Private changed from Yes to No

Dear Evie,

In Stata's SVY, the "Population size" is the sum of all individual values of the b\_ind5mus\_lw weight for respondents remaining in the analysis after you restricted the sample. Since this value is less than 1 for most respondents (as you retained only minorities who are oversampled and on average have lower weights), the total sum is much smaller than the actual number of people in the analysis.

You might also consider using the subpop option for your analysis.

Best wishes,  
Piotr  
UKHLS User Support

**#6 - 01/27/2025 12:20 PM - Evie Gates**

Hi Piotr,

Thanks so much for explaining this!

Will the fact that my 'population size' is so low have any impact on my analyses going forward?

I am currently re-cleaning my data using the subpop option instead of dropping non-eligible cases.

Many Thanks,

Evie Gates

**#7 - 02/04/2025 05:29 PM - Understanding Society User Support Team**

Hi Evie,

Sorry for the delay. No, the 'population size' in Stata doesn't matter in your case — all statistical testing is based on the 'number of obs' figure. The former is more relevant when weights are scaled to the population size, which isn't the case with UKHLS.

Best wishes,  
Piotr Marzec

**Files**

Do file for US support.do	3.61 KB	01/21/2025	Evie Gates
Do file for US support.do	9.44 KB	01/21/2025	Evie Gates