

## Understanding Society User Support - Support #1743

### Averaging regional data to obtain control variables for individuals

08/05/2022 11:14 AM - Carolin Schmidt

<b>Status:</b>	Resolved	<b>Start date:</b>	08/05/2022
<b>Priority:</b>	Normal	<b>% Done:</b>	100%
<b>Assignee:</b>	Olena Kaminska		
<b>Category:</b>	Weights		
<b>Description</b> Hi there,  I am using wave 6 to study household heads' homeownership probabilities. I am looking at native Brits and immigrants (I came up with an immigrant dummy for every household head).  I would now like to generate a control variable for each of my household heads: the variable should reflect the proportion of immigrants in the UK region where the person resides (that is, every household head in e.g London will have the same immigrant share attached, etc.). I am wondering how I should calculate that average: does it have to be weighted (i.e. <code>egen immishare = wtmean(immigrant), weight(indscui_xw) by(region)</code> using the <code>gwtmean</code> package which calculates weighted statistics)? I would think so, because without weighting it, I would have an average immigrant share based on the (not-per-se representative) raw data. However, if I calculate a weighted mean, then I would effectively double-weight the data because the regression itself would be weighted too, no?  I am unsure how to proceed and would appreciate any help.  Best wishes, Carolin			

#### History

##### #1 - 08/05/2022 03:05 PM - Olena Kaminska

Carolin,

You don't need our weight (probability weight) in creating variables. Sometime people use analitical weight in this situation when for example some groups should have a higher importance in the resulting variable than others. This is strictly research and theory driven and has nothing to do with sampling and nonresponse (for which our weights controls).

My suggestion for you is to ignore weighting in `egen` command and use our weight in any command that provides estimation. Or you could add analytical weight in `egen` command, but our weight would be unlikely suitable here - more likely you would need to use one of the variables in the dataset that corresponds directly to theoretical reasoning for this analytical weighting stage.

Hope this helps,  
Olena

##### #2 - 08/06/2022 08:33 AM - Carolin Schmidt

Thanks for the swift reply, Olena!

##### #3 - 08/08/2022 04:23 PM - Understanding Society User Support Team

- % Done changed from 0 to 80  
- Private changed from Yes to No

##### #4 - 08/10/2022 02:56 PM - Understanding Society User Support Team

- Status changed from New to Feedback

##### #5 - 11/30/2023 11:00 AM - Understanding Society User Support Team

- Status changed from Feedback to Resolved  
- % Done changed from 80 to 100