

## Understanding Society User Support - Support #1486

### Constructing a reference group using weights

01/20/2021 06:42 PM - Leilah Plant-Tchenguiz

<b>Status:</b>	Resolved	<b>Start date:</b>	01/20/2021
<b>Priority:</b>	Normal	<b>% Done:</b>	100%
<b>Assignee:</b>			
<b>Category:</b>	Weights		
<b>Description</b>			
<p>Hi, I am trying to construct an income reference group for each individual in each wave but am having issues incorporating the population weights into the Stata code. I would like each individual's reference income to be the mean income for people of the same sex, region and education level.</p> <p>The code I have written is:</p> <pre>svyset, clear svyset psu [pweight=g_indscus_lw], strata(strata) svy: bysort gor_dv sex hiqua1_dv: egen refgrs1 = mean(fimngrs_dv) if fimngrs_tc==0</pre> <p>However, Stata cannot run the last line because of the 'svy'. How I can resolve this issue?</p>			

#### History

##### #1 - 01/20/2021 08:10 PM - Gundi Knies

- Status changed from New to Resolved

Hi Leilah,

to generate the mean in a group you do not need to use the svy suite of commands.

Simply generate a new variable that contains the possible combinations of gor\_dv sex hiqua1\_dv using commands like egen group(), and then summarize the values by this new variable applying the weights:

```
gen x_refgr=.
forvalues v=1(1)N {
sum x [aw=weight] if refgroup==`N'
gen x_refgr=r(mean) if refgroup==`N'
}
```

##### #2 - 01/20/2021 08:17 PM - Leilah Plant-Tchenguiz

Great, thank you for your help!

##### #3 - 01/20/2021 08:21 PM - Gundi Knies

- Assignee set to Leilah Plant-Tchenguiz

Hi Leilah,

to generate the mean in a group you do not need to use the svy suite of commands.

Simply generate a new variable that contains the possible combinations of gor\_dv sex hiqua1\_dv using commands like egen group(), and then summarize the values by this new variable applying the weights:

```
gen x_refgr=.
lab var x_refgr "Mean value for reference group"
forvalues v=1(1)N {
sum x [aw=weight] if refgroup==`v'
replace x_refgr=`r(mean)' if refgroup==`v'
}
```

Each person in refgroup v (with values 1 through N) should then have the mean of their refgroup assigned to them in variable x\_refgr. Hope this helps,

Gundi

PS. Please ignore my previous reply. hit the wrong key and it was posted prematurely.

##### #4 - 01/20/2021 08:23 PM - Leilah Plant-Tchenguiz

Thanks

**#5 - 01/21/2021 10:09 AM - Alita Nandi**

- % Done changed from 0 to 100

There are multiple methods. For example, you can also use collapse:

```
Open your data
save temp, replace // collapse changes the dataset, so you will have to save the existing data first
use temp, clear
collapse fimngrs_dv [pweight=g_indscus_lw] if fimngrs_tc==0, by(gor_dv sex hiqua1_dv)
merge 1:m gor_dv sex hiqua1_dv using temp
```

**#6 - 01/21/2021 10:09 AM - Alita Nandi**

- Private changed from Yes to No

**#7 - 01/21/2021 10:40 AM - Alita Nandi**

- Assignee deleted (Leilah Plant-Tchenguiz)

**#8 - 01/30/2021 05:08 PM - Leilah Plant-Tchenguiz**

Hi Gundi, why do you use an aweight? I am trying to construct the mean income of a reference group weighted by population.

**#9 - 01/30/2021 05:41 PM - Leilah Plant-Tchenguiz**

In addition, how would I know which population weight to choose for the mean income calculation? When calculating mean income for each reference group, I would like reference groups to consist of people of the same sex and age in the same region and year. The dataset I am using to calculate this consists of only waves 1-7. Can I use a cross-sectional weight for a pooled sample from waves 1-7, or is this not valid because I am splitting the data by year?

**#10 - 02/01/2021 10:37 AM - Gundi Knies**

- Category set to Data analysis

- Assignee set to Leilah Plant-Tchenguiz

Hi Leilah,

as you are looking for the population mean, you can use summarize which allows aweight, but not pweight. Alternatively, use mean with pweight. The Stata user fora have quite a few queries and responses on this.

Regarding the reference income and which weights to use, most applications I have seen have used present (i.e., current year or quarter) or past (i.e., typically, last year or quarter) income as reference income. Assuming that you'd be looking to follow in the footsteps of that type of work, to predict incomes in a particular calendar year using Understanding Society, you'd be pooling data from two consecutive waves (e.g., W1Y2 and W2Y1 to get estimates for 2010). You'd use re-scaled cross-sectional weights for this. Please see <https://iserredex.essex.ac.uk/support/issues/494> for further guidance on pooled cross-sectional analyses.

Best wishes,  
Understanding Society User Support

**#11 - 02/01/2021 08:39 PM - Leilah Plant-Tchenguiz**

Thank you.

**#12 - 08/06/2021 11:20 AM - Understanding Society User Support Team**

- Category set to Weights