
Subject: Weighting of pooled country and year data
Posted by [Christiaan](#) on Fri, 17 Jan 2025 15:15:02 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear DHS community

I have been performing analyses using DHS data pooled over various countries and years. I would like to weight this data accordingly. I have been reading up on how this should be done and I came across the following thread:

<https://userforum.dhsprogram.com/index.php?t=msg&goto=9982&S=Google>

What interests me in particular is the first response by DHS Stata Specialist Shireen Assaf:

```
gen wt= v005/1000000
```

```
egen strata=group(v000 v025 ADM1_CODE) // strata also includes the survey (identified by v000)
in the group command
```

```
egen v001r = group(v000 v001) // cluster also includes the survey in the group command
```

```
svyset v001r [pw=wt], strata(strata) singleunit(centered)
svy: tab ADM1_CODE
```

Now, I would like to display mean values of certain outcome variables by birth cohort. How would I weight this appropriately, drawing on the code provided above?

```
gen wt= v005/1000000
```

```
egen strata=group(v000 country v007 birth_cohort) // Where v007 is the year in which the survey
was conducted and birth_cohort is the cohort in which the observation was born.
```

```
egen v001r = group(v000 country)
```

```
svyset v001r [pw=wt], strata(strata) singleunit(centered)
svy: sum outcome_var, by(birth_cohort)
```

Would this approach to weighting make sense? If not, how can I improve upon it? This is the first time that I am using weights and I would just like to make sure that I do it correctly.

Thank you so much for your kind assistance!