

---

Subject: Re: Number of respondents

Posted by [Bridgette-DHS](#) on Fri, 16 Aug 2024 16:26:01 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from Senior DHS staff member, Tom Pullum:

The weights compensate for the over-sampling of clusters in smaller strata and the under-sampling of clusters in larger strata, as well as for variations in the number of vacant households and non-response. That's why the weighted numbers are more representative than the unweighted numbers. All households within a cluster have the same weight. All individuals within a cluster have the same weight.

In (a), there is NOT an assumption that all household members are in school. Don't worry about that.

For (b), I would say there are 2 ways to rank the counties. Both of them use weights. You can rank them in terms of the number of students currently attending, or in terms of the percentage of eligible students who are currently attending. The following lines do this, using the PR file for the 2014 survey. I use the "collapse" command. There are alternatives, but this seems the easiest.

```
use "...KEPR72FL.DTA", clear
```

```
* measure of school attendance: sh18, "attend school current year"
```

```
* county list: shregion
```

```
gen cases=1
```

```
gen inschool=1 if sh18==1
```

```
collapse (sum) cases inschool [iweight=hv005/1000000], by(shregion)
```

```
* sort the counties by number currently attending
```

```
sort inschool
```

```
gen cases_rank=_n
```

```
list, table clean
```

```
* sort the counties by current attendance rates
```

```
gen inschool_pct=100*inschool/cases
```

```
sort inschool_pct
```

```
gen rate_rank=_n
```

```
list, table clean
```