

---

Subject: Proportions

Posted by [Khanjila](#) on Tue, 16 Aug 2016 10:22:23 GMT

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

---

I would like to create the proportion of households in a cluster that have access to electricity.

This is what I have done:

```
bys hv001: gen proportion= sum(hv206)/_N
bys hv001: replace proportion= proportion[_N]
```

However, I realized that since there can be duplicate households in a cluster as shown below. (This is because my study is focusing on impact of electricity on school attendance and so there can multiple children in a household who attend school)

hv001 hv002

2 76

2 76

2 76

3 7

3 7

3 23

3 39

3 39

3 39

3 47

3 56

3 80

3 80

I tried using duplicates drop household number

Before proceeding with the previous syntax. But my sample reduced drastically to around 147 from about 9000.

Is there something I am doing wrong?

Or is there a better method of getting the proportion of households in a cluster with access to electricity?

---

Subject: Re: Proportions

Posted by [Bridgette-DHS](#) on Tue, 16 Aug 2016 14:40:43 GMT

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

---

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

I suggest that you construct the proportion of households in the cluster that have electricity with a collapse command. You could do this from the HR file or from the PR file, selecting one record

per household. Then merge that back with the PR file or whatever other file you are using. The lines below show how to construct this proportion. You do not need to use the weights because the weights are constant within clusters. Let me know if this is not what you need.

```
use KEPR70FL.dta, clear
keep if hvidx==1
keep hv001 hv206
tab hv206
collapse (mean) hv206, by(hv001)
rename hv206 electricity
label variable electricity "Prop. of hh in cluster with elect."
sort hv001
* next merge this file with whatever other file you are using
```

---