
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?