
Subject: Spatial analysis

Posted by [David34](#) on Sun, 18 Aug 2024 15:58:24 GMT

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

For spatial analysis of intimate partner violence: In Stata, I 'svy' set the data and computed the cumulative proportion of intimate partner violence (IPV). Next, I computed the proportion of IPV at each cluster/GIS point (v001), for use (joining with) in the DHS Shapefile (variable: DHSClust) by using the Stata command:

```
svy:prop ipv, over (v001)
```

Is that the correct approach? Because v021 (same as v001) is also used for svy-setting as well. Thank you!

Subject: Re: Spatial analysis

Posted by [Janet-DHS](#) on Fri, 23 Aug 2024 15:58:16 GMT

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

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

What you did looks correct to me. However, for cluster-level proportions, I don't believe svyset and svy are actually needed. I suggest that you repeat this WITHOUT svyset and svy, and compare the cluster proportions with what you produced WITH svyset and svy. Can you let us know whether there is a difference?

Subject: Re: Spatial analysis

Posted by [David34](#) on Fri, 23 Aug 2024 17:17:50 GMT

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

Thank you so much for your reply.

I computed cluster-level proportions with and without svyset and svy and got different results. Out of 457 clusters, the proportions were the same for 166 clusters, and different for the remaining 291 clusters.

Subtracting the proportions obtained without the svy prefix from those obtained WITH the svy prefix, the mean of the 291 differences was 0.0072, with a standard deviation of 0.0808. The minimum difference was -0.26, and the maximum difference was 0.1687.

Would it be appropriate to weight the proportions by using only the 'frequency weights' without using 'stratum' and 'PSU'? I tried this approach using the following Stata code:

```
. proportion ipv [fweight=d005], over(v001)
```

Please note that I could not divide d005 with 1000000 (d005/1000000), because Stata doesn't accept non-integer frequency weights.

The results were perfect match for 447 out of 557 clusters, while for the 10 records the differences were minuscule.

Several articles using DHS data for the spatial analysis of IPV, state that they used weighted proportions for hotspots and kriging. I wrote to some for clarification, but didn't get a reply!

Please guide me.

Thank you indeed.

Subject: Re: Spatial analysis
Posted by [Janet-DHS](#) on Wed, 28 Aug 2024 13:29:43 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

You have raised an interesting issue. I did not expect to see variation in d005 with clusters. It took a while for me to understand it, but it makes sense.

The DV module is only administered to one woman in each household. If more than one woman in the household listing (the PR file) is eligible for the women's interview (the larger interview, not specifically the DV module), then one woman is selected at random using a "Kish grid". That is, before the interview with the women in the household actually begins, the eligible respondents (women with hv117=1) are listed and one of them is selected at random.

With this sampling scheme, if there are two eligible women, the weight for the one who is selected is approximately doubled. If there are three eligible women, the weight for the one who is selected is approximately tripled, and so on. I say "approximately" because there is an adjustment for nonresponse.

Below I will paste a short Stata program that calculates the number of eligible women from the PR file (I call it nelig), merges that onto the IR file, and then calculates the standard deviation of d005 within clusters and also within values of nelig. It shows that d005 IS constant within clusters, if you take account of nelig. You will also see this if you list v001 v002 v003 v005 d005 nelig within some representative clusters. (You will see that v005 is constant within clusters, regardless of nelig.) I use the Kenya 2022 survey for an illustration.

This means that the variation you are finding in d005 within clusters has nothing to do with svyset. You can get the cluster-level proportions just with "proportion ipv [fweight=d005], over(v001)", as you did. fweight is ok because the factor of 1000000 is in both the numerator and the denominator of the proportion and cancels out.

I recommend that you use the usual svyset command, with d005 in place of v005. This is our standard recommendation for analyses of the DV variables.

* The sampling weight for the DV respondents is constant within combinations of

* clusters and the number of women who are eligible

* nelig is the number of eligible women in the hh.

* it is the number of women in the household with hv117=1e

* Specify a workspace

```
cd e:\DHS\DHS_data\scratch
```

* Find the number of eligible women using the PR file

```
use "...KEPR8CFL.DTA" , clear
```

```
keep if hv117==1
```

```
collapse (sum) hv117, by(hv001 hv002)
```

```
rename hv001 v001
```

```
rename hv002 v002
```

```
rename hv117 nelig
```

```
label variable nelig "Number of women in hh eligible for DV module"
```

```
save temp.dta, replace
```

* Open the IR file, reduce to the women selected for DV, and add nelig to each woman

```
use "...KEIR8CFL.DTA" , clear
```

```
keep if d005<.
```

```
merge 1:1 v001 v002 using temp.dta
```

```
tab _merge
```

```
keep if _merge==3
```

drop _merge

* Show that there is d005 is constant within cluster, taking nelig into account

collapse (sd) v005 d005, by(v001 nelig)

summarize

Subject: Re: Spatial analysis

Posted by [David34](#) on Sat, 31 Aug 2024 18:42:48 GMT

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

Thank you so much for the clarification and guidance!

Just to confirm if I understand correctly: To obtain cluster-level proportions of IPV, I should use the following Stata command:

```
. proportion ipv [fweight=d005], over(v001)
```

Subject: Re: Spatial analysis

Posted by [Janet-DHS](#) on Mon, 09 Sep 2024 14:21:20 GMT

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

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

I assume that "proportion ipv [fweight=d005], over(v001)" is a generic kind of command. The data files do not include a variable "ipv" but that's something you have constructed. Yes, that will produce a cluster-level proportion, appropriately weighted.
