Subject: Interpolated surfaces
Posted by David34 on Tue, 14 Jun 2022 16:58:08 GMT
View Forum Message <> Reply to Message

Hello,

Am I correct that in order to do GIS analysis I would need to first calculate the weighted (after svy-setting in Stata) proportion of my variable of interest e.g. prevalence of variable xyz (coded as: 1=yes/0=no), and then with cluster number and geographic coordinate data, merge in a statistical program?

So, once I have the weighted proportion of my variable of interest, I calculate proportions by cluster (PSU) number and then join the proportions with the cluster, using the following code:

.svy: prop variableXYZ by v021

Thank you

Subject: Re: Interpolated surfaces
Posted by Rose-DHS on Wed, 29 Jun 2022 13:00:45 GMT
View Forum Message <> Reply to Message

To get the displaced GPS coordinates of the clusters, you'll need to download the GE dataset(s) for the survey(s) of interest.

You can link the cluster number (V001) variable from the recode file with the DHSCLUST variable from the geographic dataset.

Could you provide us with more information to help us better answer your question?

Subject: Re: Interpolated surfaces
Posted by David34 on Fri, 01 Jul 2022 20:11:17 GMT
View Forum Message <> Reply to Message

Thanks, I have access to the GE dataset as well as the DHS country survey data. In order to create an interpolated surface for e.g. diarrhea in under 5 children; am I correct that:

1 I would need to first calculate the proportion of diarrhea
2 Assign/link the proportion of diarrhea to each cluster using the cluster number (V001) variable from the recode file with the DHSCLUST variable from the geographic dataset.
3 Create interpolated surface

I need help in calculating the proportion of diarrhea for each cluster please, specifically:

A Do I need to use 'svy' command in Stata for calculating prevalence of diarrhea at the PSU/cluster level or not?

B To calculate proportion (as a percentage) of diarrhea at the cluster level, which Stata command do I need to use i.e.
- svy: proportion diarrhea, over (v001)
Or
- proportion diarrhea, over (v001)

Thank you!

Subject: Re: Interpolated surfaces
Posted by Janet-DHS on Mon, 11 Jul 2022 14:15:40 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

You are correct. You need to calculate the percentage of children with diarrhea at the cluster level (along with the number of children in the denominator). Make sure you use weights. The weight for the cluster would be hv005 times the number of children in the denominator. Then merge the GPS coordinates to those clusters. After the indicators calculated at the cluster level have been merged with the GPS coordinates, you can apply an interpolation technique.

Subject: Re: Interpolated surfaces
Posted by David34 on Mon, 11 Jul 2022 16:09:02 GMT
View Forum Message <> Reply to Message

Thank you so very much indeed!

Would it please be possible to help me in understanding how to calculate percentage of children with diarrhea at the cluster level along with the number of children in the denominator, using weights in STATA?

Any tips/code using STATA would be a tremendous help!

Thanks indeed.

Subject: Re: Interpolated surfaces
Posted by Janet-DHS on Wed, 13 Jul 2022 13:31:30 GMT

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

The following Stata lines should get you started. I suggest restricting to children who are living with their mothers (b9=0) because the reporting is much more accurate for them. For the child illness variables (h11, h22, h31), treat responses 1 or 2 as "yes". The label for h11 is usually H11 but sometimes it is h11; "describe h11" will give you the label. You do not need to do anything more with the weights; svyset and svy will re-normalize them. Other variables of interest will be dropped in the "collapse" but you can keep cluster-level variables like v005, v024, v025, etc., by adding them into the collapse statement after "(first)". I have included "diarrhea" in variable names because you might want to do something similar for other illnesses or outcomes and you need a notation to distinguish between them. You could use "h11" rather than "diarrhea". Hope this works for you.

```
describe h11
tab h11
label list H11

gen nch_diarrhea_yes=1 if b9==0 & (h11==1 | h11==2)
gen nch_diarrhea_no =1 if b9==0 & h11==0

collapse (first) v005 (sum) nch_diarrhea*, by(v001)
gen nch_diarrhea=nch_diarrhea_yes + nch_diarrhea_no
gen prop_diarrhea=nch_diarrhea_yes/nch_diarrhea
gen wt_diarrhea=v005*nch_diarrhea

summarize
histogram prop_diarrhea
```

Subject: Re: Interpolated surfaces
Posted by David34 on Wed, 13 Jul 2022 21:14:00 GMT

Thank you so very much indeed for the generous help with the Stata code, I am very grateful! But I have few follow up questions please:

1 Could you help me better understand the statement "You do not need to do anything more with the weights; svyset and svy will re-normalize them" please? Does that mean, I would need to apply svyset and svy after running this Stata code to get weighted percentage of diarrhea at the cluster level?

2 After running this Stata code, I ended up with the variables: v001, v005, nch_diarrhea_yes, nch_diarrhea_no, nch_diarrhea, prop_diarrhea, and wt_diarrhea. I am somewhat unsure here as to which variable I would need to merge with GPS coordinates please? In order to have weighted percentage/roportions of diarrhea at the cluster (v001) level, should I use the newly created

variable 'prop_diarrhea', and merge GPS coordinates to those clusters. Or do I need to first svyset the datafile to get the weighted proportions/percentage of diarrhea? And if so, how do I need to do that using Stata please?

3 Am I correct that the variable 'nch_diarrhea' was created to correctly calculate the proportion of diarrhea at the cluster level? But 'prop_diarrhea' is probably not the weighted proportion of diarrhea at the cluster level?

4 In this Stata code, is there a need to divide the weight with 1000000 i.e. v005 / 1000000? Or is it not necessary?

5 The newly created variable 'wt_diarrhea' gives the weighted number of children at the cluster level. However, for creating interpolated surface, I would only need the weighted proportion/percentage of children with diarrhea at the cluster level? Am I correct here?

Thanks again, and please guide.

Subject: Re: Interpolated surfaces
Posted by Janet-DHS on Mon, 18 Jul 2022 14:08:07 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

The new (collapsed) file will have clusters as units. The two crucial variables are prop_diarrhea and wt_diarrhea. Prop_diarrhea would be the outcome variable and wt_diarrhea would be the weight for each cluster. It is a combination of the sampling weight for the cluster (every child in a cluster has the same value of v005) and the number of children for whom the individual-level outcome could be assessed. You don't need to do anything else with the weights, such as divide by 1000000. If you stay within Stata, and use pweights, Stata will renormalize so the mean weight is 1 for each unit (i.e. for each cluster).

I would expect you to carry along, in the collapse, other cluster-level variables such as place of residence (the same for everyone in a cluster), perhaps something related to wealth quintile, education level of the mothers, etc. You could have additional outcome variables, such as fever or cough, but because they would have different numbers of children in the denominator, they would have their own weights.

You would use the cluster GIS codes to merge with a file of cluster-level characteristics that are external to the DHS data. The clusters as geographic data points would be the basis for the interpolated surface. Good luck.

Subject: Re: Interpolated surfaces
Posted by David34 on Mon, 18 Jul 2022 21:19:20 GMT

Thanks!
One last question, if I may please...

So, am I correct that the interpolated surface would be based on the variable 'prop_diarrhea' for each cluster in the merged file?

Than you indeed!

Subject: Re: Interpolated surfaces
Posted by Janet-DHS on Thu, 21 Jul 2022 16:30:47 GMT

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

Yes, with the weight that was described. Good luck with your analysis.

Subject: Re: Interpolated surfaces
Posted by David34 on Wed, 03 Aug 2022 20:06:59 GMT

Thanks a lot, but I have been struggling to understand, how to use "weight that was described".

In order to create an interpolated surface, I only need the proportion of diarrhea for each cluster; as based on the diarrhea proportion one would create an interpolated surface using geostatistical methods. What I am struggling with is how to incorporate 'weight' and assign it to each cluster's diarrhea proportion please?

Could you kindly help with Stata code to do this i.e. how to calculate weighted proportion for each cluster please?

Thank you so much!

Subject: Re: Interpolated surfaces
Posted by Janet-DHS on Wed, 10 Aug 2022 20:25:55 GMT

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

On July 13 I gave you the Stata code to do this. I will repeat it below. Perhaps you do not know that every case in a cluster has exactly the same weight. There is no need to weight the

proportion WITHIN the cluster. Also, the lines I gave you will give weights that include a factor of 1000000. If you don't want that factor, you can insert a line "gen wt=v005/1000000" after the "label list" line and then replace "v005" with "wt" in the two places where "v005" appears.

If your question is about how to include the weight in the analysis of interpolated surfaces, that will depend on the software you are using. If the software does not have an option for weights, then you have no option other than ignoring the weights.

```
describe h11
tab h11
label list H11

gen nch_diarrhea_yes=1 if b9==0 & (h11==1 | h11==2)
gen nch_diarrhea_no =1 if b9==0 & h11==0

collapse (first) v005 (sum) nch_diarrhea*, by(v001)
gen nch_diarrhea=nch_diarrhea_yes + nch_diarrhea_no
gen prop_diarrhea=nch_diarrhea_yes/nch_diarrhea
gen wt_diarrhea=v005*nch_diarrhea

summarize
histogram prop_diarrhea
```

Subject: Re: Interpolated surfaces
Posted by David34 on Tue, 16 Aug 2022 18:34:42 GMT
View Forum Message <> Reply to Message

Thank you so, so very much; I am really very grateful!