

---

Subject: Re: Query on Cluster-Level Modeling with DHS Data and Sampling Weights

Posted by [Bridgette-DHS](#) on Mon, 23 Sep 2024 18:55:27 GMT

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

---

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

When I said "this" would not be a good analysis of the data, I was referring to my own example. I was not judging what you are doing!

In the glm command that I suggest, the outcome is the numerator of a proportion (which is the mean of a 0/1 variable) and the option "family(binomial cases)" specifies the denominator (as "cases"). This is equivalent to a model in which the cluster-level outcome is a proportion and it is weighted by the number of cases in the denominator. A fitted proportion for a cluster would be the fitted frequency divided by "cases".

This part of the collapse command: "(mean) v190\_\*" will construct five proportions that add to one. These will be the proportions of "cases" that are in wealth quintiles 1, 2, 3, 4, 5. On the right hand side of the estimation command you could include all five of those proportions (as I did); one will be aliased because of the linear constraint (they add to 1). However, I would recommend a recode to a single proportion, such as the proportion in the bottom two quintiles, which will give just one coefficient and be easier to interpret.

You could perhaps include other covariates after the "mean" portion of the collapse command, but you could also just use your geospatial variables.

Hope this helps but let us know if you have questions specifically for the geospatial team.

---