

---

Subject: How do I account for clustering within families?

Posted by [vega25](#) on Fri, 11 Apr 2014 04:01:33 GMT

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

---

Hello,

The children's dataset in DHS files may have some of the records for multiple children within families - for instance if the dataset has data for all children born in the five years preceding the survey and some of the female respondents had more than one children during that time. Thus many children in the dataset would be siblings.

When this is the case, how does one account for the fact that there are some siblings in the data and so they have the same set of background/household/parental variables? Is adding "cluster(caseid)" as an option at the end of the regression syntax in Stata a valid way of doing so? Would this be okay? Or just asking for robust standard errors? Alternately, is a family-fixed effects the best way of doing this? My concern with family-fixed effects is the loss of sample size.

Thank you for any advice that people may have.

---