

---

Subject: Re: How do I account for clustering within families?

Posted by [vega25](#) on Thu, 04 Dec 2014 01:33:54 GMT

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

---

Hi, Thanks very much for your earlier very helpful responses. I decided in the last analysis that I was running in April to go with the cluster(psu) option. But I have run into this problem again and in discussion with some colleagues.

The background is once again that I am trying to analyze women's and children's outcomes associated with household characteristics. The DHS interviews all eligible women in the household, not just one per household. Hence the individual women's dataset has some women who share household characteristics. In Bangladesh for example, the number is 4% - not large by any means. But I am curious to see what the best strategy to deal with this issue of some women sharing household characteristics should be.

Is this too small a number of multiple women per household that I should ignore it, or is another strategy advisable? Some of my colleagues suggested that I pick one woman per household at random and then perform my analysis on the smaller individual sample so that I solve the problem in one go. My concern with this is that I am losing valuable information, and that I am no longer certain that my sample will then be representative since I cannot prove that the presence of multiple women per household is a random event. I'd also be keen to know technically in STATA how to drop cases that share household characteristics.

The alternative strategy that I was considering is clustering at the psu level - the same strategy that Dr. Tom Pullum had recommended earlier. But as we discussed earlier, that would only address the standard errors, not the point estimates.

Thoughts?

---