Subject: command for Cluster Average of Women's employment
Posted by Alanood on Sun, 09 Jul 2023 10:23:36 GMT
View Forum Message <> Reply to Message

Hi

I would greatly appreciate your guidance on calculating the cluster average of women's working status excluding women's own employment status to avoid in-built association using Stata.
I am applying instrumental variable approach and I need to use this variable as an instrument for maternal employment.

Thank you

Subject: Re: command for Cluster Average of Women's employment
Posted by Bridgette-DHS on Thu, 13 Jul 2023 16:49:34 GMT
View Forum Message <> Reply to Message

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

In Stata you can calculate a cluster-level mean with an "egen" command. For example, you could calculate the cluster-specific mean of v201 (children ever born) for all the women in a cluster using "egen v201_mean=mean(v201), by(v001)". This can be modified to construct a woman-specific variable that omits the woman herself from the calculation. Alternatively, the difference (the cluster mean minus the woman's value) will be statistically independent of the cluster mean.

Questions about instrumental variables are outside the scope of the assistance that DHS staff will provide.

Subject: Re: command for Cluster Average of Women's employment
Posted by Alanood on Thu, 13 Jul 2023 17:41:56 GMT
View Forum Message <> Reply to Message

Thank you for your replay

Regarding your point " This can be modified to construct a woman-specific variable that omits the woman herself from the calculation. "
How this can be done ?

If the average is done by this command

egen v201_mean=mean(v201), by(v001)

How can I exclude the women itself from v201_mean.

Subject: Re: command for Cluster Average of Women's employment
Posted by Bridgette-DHS on Mon, 17 Jul 2023 11:53:56 GMT
View Forum Message <> Reply to Message

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

I will illustrate with v201; you need to revise with your employment variable.

egen v201_total=total(v201), by(v001)
egen v201_n=count(v201),by(v001)

gen v201_mean=v201_total/v201_n
* v201_mean can also be obtained from "egen v201_mean=mean(v201), by(v001)"

* Now construct a mean for the cluster, for each respondent, that omits each respondent
gen v201_mean_adj=(v201_total-v201)/(v201_n-1)

correlate v201 v201_mean
correlate v201 v201_mean_adj
* The respondent's value of v201 is correlated with both of these means, although
* less so with the mean for which she is removed
