Dear Mr. Pullum,

Thanks a lot for your answer!

To clarify first: The first question in this thread (mentioning the egen command) was not posed by me. I just posed my question in this thread, since I thought to have a similar issue.

You are right, I have to be more precise about what I plan to do:
I want to evaluate the impact of a large employment programme in Ethiopia on certain outcomes, using diff-in-diff. Since the programme was implemented at the district-level, my identification strategy would consist in comparing changes between districts where the programme has been implemented and where it has not been implemented. Therefore, my unit of analysis is the district. For this purpose, I assigned district names to DHS clusters, using the DHS geographic dataset and GIS software (I am aware of possible issues with representativeness at the district-level).

Now, I want to run a diff-in-diff regression, aggregating individual observations at district-level and also cluster the standard errors at district-level.
Also, I wanted to account for stratification and weighting, following exactly the instructions of another thread:

**in the 2000 sample

egen stratum_ID_2000 = group(v024 v025)
gen tempvar=stratum_ID_2000

**in the 2011 sample

egen stratum_ID_2011 = group(v024 v025)
gen tempvar=stratum_ID_2011

**in the appended panel of both rounds

gen post =.
replace post =1 if v007 == 2003 & !missing(v007)
replace post =0 if v007 == 1992 & !missing(v007)

*construct unique identifiers for strata and clusters

egen stratum_ID = group(tempvar post)
drop tempvar

egen cluster_ID = group(v001 post)

```
svyset cluster_ID [pweight=v005], strata(stratum_ID) singleunit(centered)
```

In general, I was wondering if this procedure is still correct if I run my analysis at district-level? Or, should I code "egen Cluster_ID = group(district post) instead?

I would be very grateful, I you could give me some advice on how to proceed with the DHS weighting and stratification procedure in my case.

Thanks in advance!
Till