

---

Subject: Regions and MultiLevel Models

Posted by [Yawo](#) on Tue, 19 May 2020 12:27:49 GMT

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

---

Hello,

I am involved in a research project examining HIV risk behaviors, pooling data from recent surveys in 14 sub-Saharan African countries.

I plan to fit a 3-level HLM model , with individuals nested in regions , which are in turn nested within countries (individuals --> regions --> countries). My concern is which variables to use for the level 2 (region) differences.

DHS provides two possible options:

v023: stratification

v024: regions.

The v023 stratification variable is already used to weight (svyset) the data, so I am not if I can use it again as a proxy for regional/community differences.

The v024 would have been appropriate, except when I pool the data, the regions codes get messed up, since each country numbers the regions sequentially 1, 2, 3, 4 with unique labels. Once pooled, these variable needs to be regenerated into a long a list of regions, each with unique codes. This is tricky, and can be accomplished with some coding. But before doing so, I wonder if:

- (1) there is/are other ways to recreate the region variable to ensure uniqueness across cases
- (2) there is any other variable to use to capture community differences in my HLM model (at level 2)
- (3) it is ok to use the v023 stratification variable as a proxy.

Thanks in advance for your attention.

cY

---