
Subject: Re-Numbering Survey_Year and Survey_Rounds

Posted by [Yawo](#) on Tue, 30 Apr 2019 14:44:16 GMT

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

Hello,

I have pooled data from about 10 countries, and ready to generate new strata/cluster variables. Consistent with advice from this board, I created a new variable, "survey_year" which takes the value of 1, 2, 3, 4 (to represent each of the survey rounds within a country).

Here are the two commands I intend to use:

```
egen cluster=group(survey_year v001)
egen stratum=group(survey_year v024 v0025)
```

But I am not sure if the survey_year variable is uniquely representing the survey rounds. For example, I have the following rounds of surveys, and their corresponding "survey_year" values:

Congo2007: survey_year=1
Congo2013: survey_year=2

Gabon2012: survey_year=1

Ghana2003: survey_year=1
Ghana2014: survey_year=2

So, survey_year=1 can refer to different years. Does this matter, for generating new strata and cluster variables?

What if I create two variables: "survey_round" (with values: 1, 2, 3, 4) - to capture the different rounds of survey, and a second variable: "survey_year" (to consecutively number the survey years: 2003, 2004, 2005, 2006,2017).

The two new strata variables could be:

```
egen cluster=group(survey_year survey_round v001)
egen stratum=group(survey_year survey_round v024 v0025)
```

Would these new approach be ok?

Thanks - Yy
