
Subject: How to control for the year of sample
Posted by [olympiaca](#) on Thu, 07 Jan 2021 13:44:28 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello,

I'm using the domestic violence data to look at the likelihood of a woman receiving violence from her husband. Currently I'm combining three Jordan DHS: the 2007, 2012, and 2017. The data has structure in that it is grouped into regions and also grouped by cohort.

I'm interested in the difference in likelihood of receiving violence between cohorts as well as some other variables.

Is it appropriate to do a two-level model where I use Region as my grouping variable, and then have cohort as a predictor variable? Or should I be performing a three level model with individuals nested in cohorts nested in regions?

Statistically I'm not sure which one is appropriate?

Many thanks
O

Subject: Re: How to control for the year of sample
Posted by [olympiaca](#) on Thu, 07 Jan 2021 16:58:23 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi,

I received an email saying this had been replied to, but it seems it has since been deleted. By cohort what I meant was survey year. Essentially women surveyed in 2017 are much less likely than women in previous years to report domestic violence.

I want my regression to be: $\text{violencefromhusband} \sim \text{surveyyear} + \text{wealth} + \text{education} + \text{age} + \text{region}$

- 1) Given that I am combining 3 survey years do I need to renormalise the weights? I saw in another thread that I should perhaps do that?
- 2) using the R equivalent survey package to account for the survey structure I would then use V021 for the primary sampling unit and V022 as the strata? But I'm not sure how this works across survey years?
- 3) if I use the survey package then is there no need to control for region in my regression?

Many thanks
O

Subject: Re: How to control for the year of sample
Posted by [Bridgette-DHS](#) on Thu, 07 Jan 2021 18:38:14 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

It would be important to include survey adjustments with svyset and svy. In most surveys, the sampling strata are combinations of region and urban/rural residence. If svyset includes the strata() option then you have adjusted for the role of region in the sample design. You can also include region as a covariate. With multiple surveys you need to make the strata survey-specific. There have been postings that describe how to do this with "egen group".

Successive surveys are not cohorts. Date of survey is a time period indicator. You could include that as a covariate. Cohort is identified by year of birth. You can link birth cohorts across surveys if you construct a covariate based on year of birth.

I don't see how cohort, survey, or region are nested in any way. They are crossed. I can imagine having an interaction term for combinations of regions and cohorts, for example, although that would give a lot of terms.

Frankly, I don't think much will be gained by using a multi-level model--however it would be framed--in this context. I doubt that the theoretical justification (which I agree with) would translate into statistical insights. But it could be worth a try.

Subject: Re: How to control for the year of sample
Posted by [olympiaca](#) on Fri, 08 Jan 2021 09:52:29 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you so much for your response,

I have one more question. To use svyset the stratification seems to be different between 2017 and 2012/2007.

the Jordan 2017 DHS V023 is a combination of region and urban/rural residence and is numeric. however, for 2012 and 2007 V023 is simply the region and is categorical (the name of the region).

In order to make the strata survey specific do I convert the 2012 and 2007 into numbers and use:

```
egen strata_survey = group(V021, V023)
```

Or do I need to create a new variable for 2017 where the strata are just regions and then control for urban rural later as a covariate?

Many thanks

Olympia

p.s I am using R not Stata so apologies if i got the stata code a bit wrong.

Subject: Re: How to control for the year of sample
Posted by [Bridgette-DHS](#) on Fri, 08 Jan 2021 12:47:55 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

For each survey you can construct "stratumid" using whatever was specific to that survey. Then for the pooled file, "egen stratumid_all=group(stratumid survey)".

The clusterid is v021 (=v001) in each survey. You would have a separate command such as "egen clusterid_all=group(v021 survey)".

Then in the svyset command you put "clusterid_all" and "stratumid_all" in the appropriate position. Of course, the names you give these ID codes is up to you.

Subject: Re: How to control for the year of sample
Posted by [olympiaca](#) on Fri, 08 Jan 2021 14:30:01 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you very much, this is all extremely helpful.

V022 is correct for the 2017 Jordan DHS.

For the 2012 Jordan DHS V022 has 42 strata but the final report says it should have 43? Could this possibly be a typo?

For the 2007 I need to construct the variable for stratum - would a cross between V024 and V025 be the best thing to do?

Thanks
Olympia

Subject: Re: How to control for the year of sample
Posted by [Bridgette-DHS](#) on Fri, 08 Jan 2021 14:45:10 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

There have been many posts on this topic (pooling or comparing surveys), but mostly in terms of Stata because mostly we use Stata.

The strata in successive surveys should have different codes for the purpose of adjusting for the survey design, even when the strata are the same in each survey. You need some mechanism for assigning different ID codes in the different surveys. For example, if you number the surveys 1, 2, 3, you could construct a variable "stratumid" as "V022 + 1000*(survey-1)". In Stata we would use "egen stratumid=group(survey v022)" to construct distinct numbers.

There are different possibilities for the weights. Say that the number of cases in each survey is n_1 , n_2 , n_3 . You could construct a new weight "v005rev" that would be proportional to v005 within each survey but would add to $(n_1+n_2+n_3)/3$ within each survey. This is spelled out in other posts. Alternatively, you could leave the weights alone, but then the estimates would be biased toward the largest survey.

You should still include region as a predictor in the regression. Including it (via V022) in the survey adjustments will only adjust for the design. It does not control for region in the analysis.