
Subject: Setting up pooled DHS data as panel
Posted by [dan7580](#) on Sun, 29 Jun 2014 14:36:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

I have pooled 2 rounds of data (appended) for several different countries separately and would like to set up as longitudinal data sets so that I can examine trends over time. Can you please share how DHS researchers have done this in the past using Stata? I have read about the xtset command but it would be helpful to see how this has been applied to DHS data and what steps are necessary to employ this command.

Thank you for your time.

Dan

Subject: Re: Setting up pooled DHS data as panel
Posted by [Reduced-For\(u\)m](#) on Sun, 29 Jun 2014 22:48:49 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dan,

I've done this in a few ways, but it really depends on the outcomes you are using and the right-hand-side variables of interest. Can you give me a little more detail?

1 - each country separately, or all countries pooled?

2 - what is the unit of analysis (and which recode)? Family, child, mother?

xtset (and xtreg/whatever) may be useful, but maybe not. It really depends on the kind of variation you are trying to grab and the questions you are asking.

If you just want to look at within-country changes over time, you can do the regressions separately by round and compare the coefficient estimates from first/second survey, and then use the "sureg" commands (seemingly unrelated regressions) to get standard errors/p-values/confidence intervals). That would essentially compare means from one survey to the next, and would not require adjusting the weights or re-setting new strata or the other kinds of problems that can arise when appending multiple rounds (see the discussions on merging/appending datasets that we have only partially managed to resolve on the forums).

Subject: Re: Setting up pooled DHS data as panel
Posted by [leo2015](#) on Mon, 30 Jun 2014 07:37:49 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks very much for your quick reply.

1 - I am looking at each country separately

2 - the unit of analysis is mother (I have merged the IR and HR data sets previously as I needed some variables from the HR data set)

My outcomes are all binary (ANC 4+ times, TT vaccination 2+ times, etc.) I'd like to compare within country and see if these changes are significant over time.

Dan

Subject: Re: Setting up pooled DHS data as panel

Posted by [Reduced-For\(u\)m](#) on Mon, 30 Jun 2014 19:34:27 GMT

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

Dan,

Then I suggest doing the regressions separately and using "suest"* to compute the standard errors of the differences. That would be the most straight forward. It works something like this:

```
eststo: svy: reg Y X if survey==1
      estimates store reg1
```

```
eststo: svy: reg Y X if survey==2
      estimates store reg2
```

```
suest reg1 reg2
      test X
```

- that should give you a comparison of the coefficients on X from the two regressions, along with standard errors, and accounting for survey design and all that. I haven't used this with the "svy" prefix, but I think it should work fine.

Documentation for "suest" here: <http://www.stata.com/manuals13/rsuest.pdf>

There are of course ways to do it using the pooled data, but I think this might be the simplest, most transparent way and the easiest to get all the weighting/stratification right (because you are only estimating parameters within-survey, and then combining those estimates across surveys and the "suest" command does all the work calculating the "simultaneous (co)variance matrix of the sandwich/robust type".

*I think I called it "sureg" in the previous post - my bad (its because it uses "seemingly unrelated regressions").

Subject: Re: Setting up pooled DHS data as panel
Posted by [dan7580](#) on Wed, 02 Jul 2014 09:46:07 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you for the information on using the suest command. This is really helpful.

Do you mind sharing a little more about some of the weighting/stratification limitations you mentioned with setting up the data as panel?

Thanks again,
Dan

Subject: Re: Setting up pooled DHS data as panel
Posted by [Reduced-For\(u\)m](#) on Thu, 03 Jul 2014 01:56:09 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi Dan,

I wish I could be super helpful, but probably only a little helpful.

A good place to start is with the document I'm attaching, which comes from this thread (the "weighting" thread in general has a lot of discussion on this topic):

[http://userforum.dhsprogram.com/index.php?t=tree&goto=82
&S=e3a92c3e8a765f0217181d74f8127581#msg_82](http://userforum.dhsprogram.com/index.php?t=tree&goto=82&S=e3a92c3e8a765f0217181d74f8127581#msg_82)

The major problem is that the DHS weighting design is intended for national (and sometimes sub-national) representativeness, but not for combining survey rounds - either within country over time, or across countries. The sample weights sum to the total sample size, and thus the meaning of some weight within some survey is lost when it is compared to an observation in another survey. The DHS way of handling that is to "de-normalize" the weights as described above. Then, in theory, the re-computed weights should work across survey rounds, and have implicit weighting for population size as well.

Unfortunately, if you read that document above, the "de-normalizing" process is complicated and requires you to bring in outside data that may or may not be good/useful. I've proposed other methods, such as forcing weights within a survey to sum to 1, and then multiplying them by some population size estimate to overlay population weights on the (DHS given) probability weights. I haven't done the math to see if this reduces to the formula provided by DHS, but I hope to get around to it at some point (I'm not a statistician/econometrician, so I'm worried I'm missing something with this method, even if the algebra seems reasonable).

You also have to re-define strata and cluster variables when using multiple survey rounds, so that cluster "101" in one survey is differentiated from cluster "101" in some other survey (and same with strata).

I think the "do all the surveys separately" thing might be easiest for you, since you are effectively looking at one survey as one observation - so all the weighting and p-value/standard-error problems reduce to the regular single-round DHS method. Then again, if the sample sizes are fairly constant within-country, you should be able to just use the DHS-provided weights and get a very similar answer. You could also ignore weighting altogether, cluster your standard errors using `cluster(clustervar)`, skip the `svy` part, and just estimate an effect within the sample population that is not scaled to be population representative (only a problem if there are heterogeneous effects of X on Y across the population - otherwise, in a lot of ways, one observation is as good as any another one).

It's up to you. That's most of what I know. I'm happy to follow up on anything, but don't know if I'll have much to offer.

Subject: Re: Setting up pooled DHS data as panel
Posted by [dan7580](#) on Mon, 07 Jul 2014 13:17:18 GMT
[View Forum Message](#) <> [Reply to Message](#)

Your feedback has been incredibly helpful -thank you.

I want to make sure that I am clear on the `suest` syntax example you provided previously. I read through the manual but wanted to quickly follow-up with you. It looks as though the data are appended with both survey rounds if you are setting each regression up with the `"if survey==1"` or `"if survey==2"` command. I had understood these would be done separately, is that correct?

Thanks again.

Subject: Re: Setting up pooled DHS data as panel
Posted by [Reduced-For\(u\)m](#) on Mon, 07 Jul 2014 17:06:45 GMT
[View Forum Message](#) <> [Reply to Message](#)

Yes - the survey rounds are actually appended before you do this procedure, you just aren't treating them as though they were appended when analyzing them. Having them in one dataset allows you to avoid problems with clearing the data in between regressions, and thus clearing out the information saved for the `suest` comparisons.

So you append the sets together, then run the regressions on each survey-round using an `"if"` command, then compare the regression results using `"suest"`. That is just the way I have done it, I'm sure there are other ways.

Subject: Re: Setting up pooled DHS data as panel
Posted by [dan7580](#) on Tue, 08 Jul 2014 11:48:06 GMT

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

OK. So in this case, you would not do anything to the weights or strata before using the svy command on the pooled data?

Subject: Re: Setting up pooled DHS data as panel
Posted by [Reduced-For\(u\)m](#) on Tue, 08 Jul 2014 23:56:49 GMT
[View Forum Message](#) <> [Reply to Message](#)

Correct - you do not need to adjust the weights or the strata, just use the "if" command.

But - one obvious way to check this is to run some estimate on just \one survey round, then run it again on the pooled dataset using the "if" command, and compare the point-estimates and standard errors. I think you should get exactly the same results - let me know if you try that, I'm curious. If you get different numbers, try using the "subpop" command on the pooled dataset (with the sub-population being the survey round) and see if that gets you back to the single-survey estimates.

I sort of hate the "svy" command, but that is because it is really hard for me to figure out exactly how it deals with things like clustering and stratification, and what exactly the "subpop" option is doing.

Subject: Re: Setting up pooled DHS data as panel
Posted by [dan7580](#) on Wed, 09 Jul 2014 08:29:37 GMT
[View Forum Message](#) <> [Reply to Message](#)

The "if" command does indeed work. Thank you for clarifying and for all of your help!
