
Subject: Comparing coefficients across years in the same country

Posted by [Ryan](#) on Mon, 27 May 2013 17:29:19 GMT

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

I am using four waves of male and female DHS data from Ghana (1993-2008), and my intention is to examine how different independent variables influenced education (v133) over time. I have 3 questions, which are in italics following a short explanatory statement.

First, after reading through the forum I am confused about if I need to denormalize my survey weights if I am comparing across time within the same country. If I do need to denormalize weights then I will use the formula: $\text{denormalized weight} = \text{weight} * \text{total size of survey population} / \text{size of population sampled}$. Can someone advise on if I need to denormalize the weights? If I do, is this formula correct?

Second, the stratification procedures in Ghanaian DHS surveys changed over time. In 1993, the survey was stratified by three ecological zones and then rural/urban. It was also self-weighting. From 1998 to 2008, I believe that the sample was stratified by region and then urban/rural. In the latter three waves, the northern regions were oversampled. I am uncertain as to how I should set up Stata (or R) to deal with these changes. I also don't know if I need to do anything special to the male recodes, as they were sampled similarly.

Third, to conduct the analysis I will add dummies for each DHS wave (e.g. all 1993 observations are $1993\text{Dummy}=1$, else 0) and a male dummy (all male observations are $\text{MaleDummy}=1$, else 0) and then I'll combine all of the individual datasets. This involves renaming the male variables, but that is all fine. To make the next part easier to explain, I'll pretend that I'm only merging and comparing the 1993 and 1998 portion of the data and I'll use only a few variables. To make valid comparisons over time, I would have a regression formula like:

$v133 \sim \text{MaleDummy} + \text{UrbanDummy} + 1998\text{Dummy} + 1998\text{Dummy} * (\text{MaleDummy} + \text{UrbanDummy})$

This will tell me if the effect of being male or urban on educational attainment is changing over time and will provide confidence intervals for the effect of time. I realize this final question is more statistical in nature, but can anyone foresee problems with using the DHS data to conduct this kind of analysis?

Thank you for any thoughts or recommendations.

Ryan

Information on DHS surveys:

1993: <http://www.measuredhs.com/pubs/pdf/FR59/FR59.pdf>

1998: <http://www.measuredhs.com/pubs/pdf/FR106/FR106.pdf>

2003: <http://www.measuredhs.com/pubs/pdf/FR152/FR152.pdf>

2008: <http://www.measuredhs.com/pubs/pdf/FR221/FR221%5B13Aug2012%5D.pdf>

Subject: Re: Comparing coefficients across years in the same country

I think this is just about right. A couple of quick points:

1 - Re-normalizing weights: In this case, I think you do want to re-normalize weights. You want each survey to represent the whole country at one point in time, and that is what each survey, re-normalized, will do. I would do it slightly differently than you do though (they will come out very similar). Here's my method:

*First - trim your sample to it's final set of observations, assuming you'll lose some for missing covariates and the like, then:

```
egen totalweight = sum(weight), by(survey) /*survey being survey round
```

```
gen pweight = DHSweight/totalweight
```

So now all your weights sum to 1 within a survey. Implicitly, you're ignoring changes in population size over time.

2 - Are you worried about the standard error/weighting problem here? As in, how to get svyset/svy set up right? Here's one way to think about it. You want to basically do what the DHS FAQs say, but you want to make sure that you have region/psu generated separately for each survey round. I usually do something like concatenate the year of the survey to the PSU, and use that as the "cluster" in my svyset. The weights you computed above would be the right weights. So the example code from the website is:

Example Stata code:

```
*generate weight  
generate weight = YOUR NEW WEIGHT
```

```
*make unique strata values by region/urban-rural (label option automatically labels the results)
```

```
egen strata = group(v024 v025), label
```

```
*check results
```

```
tab strata
```

```
*tell Stata the weight (using pweights for robust standard errors), cluster (psu), and strata:
```

```
svyset [pweight=weight], psu(v021) strata(strata)
```

...but you want to replace the "strata" line with something like: `gen strataUse = strata+surveyyear*10000` or something...just some way so that strata are separate by survey round. also, you'd want to replace the v021 in the "psu" bit with something that made sure the psu number wasn't repeated over rounds (like, you don't want v021=45 in two different rounds...I can't remember exactly what form v021 is in digit-wise).

3 - I think that is close, but not exactly right, but maybe I'm understanding it wrong. So, your idea is to know how urban males education levels have changed over time. One thing to do is just:

egen meanEd = mean(Education) if male==1 & urban==1, by(round)

but you want to use regression for standard errors and weighting and what not, so cool.... so suppose your model has all the year dummies, an urban and a male dummy, and then YearXUrbanXMale dummies too (but no YearXUrban and YearXMale dummies). The coefficient on YearXUrbanXMale would look at how education for urban men was different than education for the rest of the population (women, and rural men) in that year. I'm not sure that is exactly what you are going for. Maybe think about the problem like this: you have 4 potential groups of MaleXUrban - men in cities, rural men, and the corresponding women's groups. One of these groups, within each year, will always be a reference population (right now, implicitly I think you have are comparing urban men to all the other three groups), and the coefficient on the YearXcovariate dummies will be the difference between the mean education for the excluded group(s) in that year, and the group of interest.

But I think maybe you want to know, say, has the gap between rural/urban education levels for men been increasing or decreasing? In that case, first drop all the women, then regress Education on Year dummies, and dummies for SurveyXUrban (and no Urban dummy alone), then when you graph SurveyXUrban across survey round, you'll have the difference between rural men and urban men in education level for that survey year.

If you just wanted to know how urban men's education was changing over time....just drop all the non-urban men, and look at the survey dummies (for mean) or survey dummies interacted with X (for marginal effect of X in that year). Note, if you include both X itself and Survey*X, the effect in any year is the sum of the two coefficients.

That was a really long response, huh? Sorry, had a big cup of coffee with me. If I was unclear let me know.

Subject: Re: Comparing coefficients across years in the same country

Posted by [Ryan](#) on Mon, 27 May 2013 22:15:19 GMT

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

Your third point slightly missed what I was after, but I think that was due to my poor wording and my attempt to concoct an example that let me avoid getting too far into the weeds of my analysis. It doesn't matter, as your response still solved that issue.

I really appreciate your speed and attentiveness. Thank you.

Ryan

Subject: Re: Comparing coefficients across years in the same country

Posted by [bsayer](#) on Wed, 17 Jul 2013 16:26:22 GMT

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

If the strata are survey specific, there is no need to alter the PSU variable, as all calculations are

stratum specific. But it also doesn't hurt anything.

But don't subset the data - that may be incorrect unless you understand exactly how the calculations are made and can insure that you have enough of the correct observations to make those calculations.
