Subject: Weights and Strata for Pooled Samples
Posted by Yawo on Mon, 10 Aug 2020 16:27:14 GMT
View Forum Message <> Reply to Message

Hello,

I've used IPUMS to create a combined dataset for men and women for 23 African countries.

As the listing below shows, the weights are not constant within each PSU as they should be. In fact we have two different weights within each PSU, one for males and one for females.

list sample sex weight weight2 psupool stratapool in 115431/115435

	sample	sex	weight	weight2	psupool	strata~l
115431.	Lesotho 2014	male	.515554	1	4753	336
115432.	Lesotho 2014	female	.48581	1	4753	350
115433.	Lesotho 2014	female	.48581	1	4753	350
115434.	Lesotho 2014	male	.515554	1	4753	336
115435.	Lesotho 2014	female	.48581	1	4753	350

This issue has been the source of "weights not constant within PSU" error I have been receiving when running svy: melogit models.

However, I thought I can adjust the strata and the psu by taking gender into account. Doing so results in constant weights within each psu, as the code and listing below shows.

egen psupool= group(idhspsu sex)
egen stratapool= group(strata sex)

list sample sex idhspid idhspsu idhsstrata psupool stratapool weight in 10/15

	sample	sex	idhspid	idhspsu	idhsstrata	psupool	strata~l	weight
10.	Angola 2015	male	2401 00010001	3	2401000001	2401000018	1	321.979475
11.	Angola 2015	male	2401 00010012	3	2401000001	2401000018	1	321.979475

```
 12. | Angola 2015     male       2401    00010026  1   2401000001   2401000018        1        321
.979475 |
 13. | Angola 2015   female       2401    00010001 02   2401000001    240100018        2         18
1.001989 |
 14. | Angola 2015   female       2401    00010002 03   2401000001    240100018        2         18
1.001989 |
     |-----------------------------------------------------------------------------------------------|
 15. | Angola 2015   female       2401    00010002 02   2401000001    240100018        2         18
1.001989 |
     +-----------------------------------------------------------------------------------------------
```

My question:

1. Is this necessary to adjust the strata, psu and even the weights when we append male and female data into one file.

2. And if so, is my <egen> code the right way to readjust the strata. If not, is there any other way?

3. I was able to run my melogit models successfully after applying the adjustment to the strata and the psu, but again, I want to be sure the adjustment is ok, else I would have to revert to running separate models for men and women.

thanks - cY

Subject: Re: Weights and Strata for Pooled Samples
Posted by Trevor-DHS on Mon, 10 Aug 2020 18:57:27 GMT
View Forum Message <> Reply to Message

Unfortunately you can't just merge the datasets together without adjusting the sample weights. The sample weights are relative weights and are normalized separately to the total number of women and total number of men in the sample. For example for Lesotho DHS 2014, 6621 women were interviewed, but only 2931 men (and that includes men age 15-59, not just 15-49). In Appendix A, it states that "In addition, in a subsample of households (every second household), all men age 15-59 who were usual residents of the households or stayed in the households on the night before the interview were eligible for interview". If you used the merged data without adjusting the weights you would be assuming that there were more than twice as many women in Lesotho than men!

Typically this adjustment is made by applying a constant factor to the weights for women and the weights for men. These constant factors are taken by using estimates of the male and female population age 15-49 in Lesotho from some external source, such as the UN's World Population Prospects or from census data, and dividing by the total sample size for women and for men age 15-49 from the survey.

This same issue applies when you pool data from multiple surveys or multiple countries, each of which have their own weights that will need adjusting.

As for the code adjusting the PSU and stata, I don't know melogit, but I cannot think of another solution if you need to include both women and men in the same model. Can you run separate models for women and for men? Otherwise I don't have a better option.

Subject: Re: Weights and Strata for Pooled Samples
Posted by Yawo on Tue, 11 Aug 2020 12:08:18 GMT
View Forum Message <> Reply to Message

Thanks so much, Trevor.

I think the best solution is to run separate models. I would be interested in knowing more about how these constant factors are applied. I have seen some studies using pooled DHS where male/females are combined. I will contact those authors to see how they handled the issue (just for curiosity sake). But I will separate the datasets by gender, and run separate models for each.

Thanks for a long and torturous detour !.

With much appreciation - cY

Subject: Re: Weights and Strata for Pooled Samples
Posted by Yawo on Fri, 14 Aug 2020 11:53:10 GMT
View Forum Message <> Reply to Message

Trevor: thanks so much, so would this be the correct svyset specification?

svyset psupool, strata(stratapool) weight(weight2) vce(linearized) singleunit(missing) || _n, weight(weight)

Note: Stage 1 is sampled with replacement; further stages will be ignored for variance estimation.

```
      pweight: <none>
          VCE: linearized
  Single unit: missing
     Strata 1: stratapool
         SU 1: psupool
        FPC 1: <zero>
     Weight 1: weight2
     Strata 2: <one>
         SU 2: <observations>
        FPC 2: <zero>
     Weight 2: weight
```

Thanks so much - cY

Subject: Re: Weights and Strata for Pooled Samples
Posted by Trevor-DHS on Fri, 14 Aug 2020 15:58:01 GMT
View Forum Message <> Reply to Message

This looks fine. A couple of notes:
1) vce(linearized) is the default, I believe, so you don't actually need to specify it.
2) singleunit - most people use singleunit(centered). Singleunit(missing) will result in missing values for the standard errors, whereas singleunit(centered) specifies that strata with one sampling unit are centered at the grand mean instead of the stratum mean. This doesn't usually matter as it is unusual to have only one sampling unit in a stratum, but in case there is such as case, using singleunit(centered) will give you an estimate.

Subject: Re: Weights and Strata for Pooled Samples
Posted by Yawo on Mon, 17 Aug 2020 21:08:25 GMT
View Forum Message <> Reply to Message

Thanks so much - changes made. Model runs well.

Gracias - cY