
Subject: Normalizing weight for region/province
Posted by [jcon](#) on Sun, 18 May 2014 06:57:58 GMT
[View Forum Message](#) <> [Reply to Message](#)

DHS normalizes weights so that the national unweighted n = weighted n.

The province/region sampling error tables show both unweighted and weighted n, with the weighted n normalized at the national level. This means that in some oversampled provinces the weighted n is very small. Are province/region level confidence intervals calculated with the weighted or unweighted n?

I am doing an endline evaluation for a project that covered three provinces in Lao PDR. The baseline is the 2011/12 MICS/DHS (combined). Currently, I'm trying to estimate power/sample size for comparing baseline to a future endline. I need to keep the sample weighted so that it is representative at the provincial level, but with the DHS national normalized weights I have an unweighted n of 2200 and a weighted n of 1300 (all of the provinces were oversampled). Can I re-normalize weights so that unweighted n = weighted n in these three provinces? There is no mention of this in the DHS manuals, which only state that provincial level estimates must use weights.

DHS does not normalize to provincial level in any of their tables; always showing the national normalized n at the provincial level. As the sample is representative at province level, it seems like it would make more sense to normalize weights at the provincial level when looking exclusively at specific provinces (something DHS reports are not designed to do).

Any suggestions would be greatly appreciated.

Subject: Re: Normalizing weight for region/province
Posted by [Reduced-For\(u\)m](#) on Mon, 19 May 2014 04:23:27 GMT
[View Forum Message](#) <> [Reply to Message](#)

I hesitate to offer an option, but here is one way to think about it. I'm supposing that those surveys are representative at the regional level - some aren't, and I don't know about this particular one.

Assuming they are regionally representative, you could normalize first within each region so that the region sums to 1. Then you could use outside data on regional populations to overlay population weighting on the within-region probability weights, doing that separately for each survey round.

The thinking is that even if the weights for some region sum to a number you aren't interested in, the relative size of the weights within a region contains all of the probability of selection information (within the region information). So by forcing those weights to sum to one, you preserve the probability weighting and lose the problem of the weights summing to the wrong number. Then, you multiply all those weights by the population of the region so that the region

itself has weight summing to its population. It would look something like this in Stata:

*region total sum of weights

```
egen region_tot_weight = total(weight), by(region)
```

*re-normalize within region

```
gen region_norm_weight = weight/region_tot_weight
```

*overlay population weight

```
gen final_weight = region_norm_weight * region_population
```

I've run this by some people who should know, and they generally seem to think it makes sense, but I wouldn't say that this is guaranteed right. I've never seen it used in a paper. Then again, most papers don't really address the weighting problem or at least don't provide any information on how they actually adjusted the weights. I think the widespread use of multiple-round DHS analysis is pretty recent and demand is a bit ahead of the technical expertise in this area.

Subject: Re: Normalizing weight for region/province

Posted by [jcon](#) on Sat, 14 Jun 2014 07:06:56 GMT

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

Yes, thanks, I think that works. I will also run by some biostat people.

These concepts apply not only for the growing demand for trend analysis (pooling multiple datasets), grouping regions/comparing countries, but also for looking within single country datasets.

For programme work (targeting and evaluation) people want to compare one province to another. First step is to make sure people are looking at confidence intervals when they do this. CIs must be calculated with unweighted n; i'll take a closer look to confirm that. Beyond using confidence intervals how do you compare two provinces. If you keep the weights, you lose your sample size; If you remove the weights the estimates will be off. For most surveys the second stage uses implicit stratification; urban and rural villages from one province are grouped together and then villages are selected PPS. So, at the province level it is a self weighting sample until there are corrections for mistakes in the sampling frame and non-response. As long as those two things aren't really bad, weighted and unweighted estimates at province level will be nearly identical. The most practical way to statistically compare provinces and have the right sample size is probably just to compare with no weights. However, it seems the right statistical methodology would be to renormalize based on the n of the two provinces so that you maintain corrections for sample frame/non-response.

Subject: Re: Normalizing weight for region/province

Posted by [Reduced-For\(u\)m](#) on Sun, 15 Jun 2014 19:03:44 GMT

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

RE: Comparing different provinces in same survey:

Wouldn't another way to do that be to use the full sample with the regular weights, and then use separate dummy variables for each province and do an F-test that all coefficients are the same? That might be a way to get around the re-weighting/re-normalizing problem, at least for some analyses.
