

---

Subject: Re: Correct weight for a sub-sample

Posted by [Reduced-For\(u\)m](#) on Thu, 30 May 2013 03:03:23 GMT

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

---

Hi Adi,

You are right that normalizing each survey so the weights add to 1 would end up treating each country as an observation weighting-wise (and would still preserve the within-country sampling probabilities, so each country would represent a 1 that is a weighted average of it's population - man, this stuff is always a mouthfull).

As for the assumptions thing - this whole weighting bit is really about two different things in a regression context (as opposed to a tabulate means context). First is population weighting - to make the survey nationally representative. The second is efficiency - if you have observations that are like means (say, a state-by-year panel where states have different populations) you might want to weight up the populous states not for representativeness but for smaller standard errors (efficiency). There is a good paper called "What are we weighting for" which is here if you have access: <http://www.nber.org/papers/w18859>

Basically, before I suggest any weighting scheme, I just want to know why people are weighting. In this case, I think if you are happy treating each country as an observation (in the weighting sense) then you are fine. It gets a bit metaphysical at times, and I don't have all the answers by any stretch, but I've been trying to figure out some of these issues in my own cross-country stuff, so I'm also trying to figure out what other people are thinking when they weight. Somehow to me the idea that one survey is weighted the same as another survey seems reasonable enough, but there would be lots of people who think that they should be population weighted (a weighted average of heterogeneous treatment effects), and others who would say to weight those surveys by Pop to get more efficient estimators (supposing homogenous treatment effects). I think as long as you are clear, they are all fine (with the "weighting for efficiency" being the most suspect).

Here's a little thought experiment: if every single person in every single survey had the exact same response to your X of interest - how would you want to weight? I think at that point, I'd just weight everyone with 1, because a person is a person. This is totally different than tabulating population means.

---