Subject: Re: Weighting Combined Individual Data for Logistic Regression Analysis
Posted by Reduced-For(u)m on Mon, 02 Feb 2015 06:09:56 GMT

There is much debate about weighting data in regression contexts when the interest is in some particular causal effect as opposed to some population average. The usual DHS line is that you should weight all your regressions, but that is not always the advice in all academic fields.

If you want a population average, you have to use the weights. That is a general truth about representative sampling and the sampling structure of the DHS>

But, if you want a causal estimate, it gets a little murkier. If you believe (read: assume) that every person, regardless of their characteristics, will have the same response to some causal input, then you do not need to weight your regressions, because it doesn't matter who was in the sample.

That said, you are describing something somewhere in between. Without getting too into your interpretation of your model and/or your assumptions, I would say that this is a very good resource for thinking about when you do and don't want to weight your regressions.

http://www.nber.org/papers/w18859

If you don't have access, check around for a copy posted on the internet, or let me know.

In general, the most conservative thing to do would be to report both weighted and unweighted estimates. They really shouldn't vary too much - if they do, there is probably something weird going on with either your model or your basic assumptions (and their relationship with reality).

Regardless of your choice of weighting, you should cluster your standard errors by PSU (this is just a general point since often people conflate weighting and clustering, though I know you didn't ask about it).