Subject: Weighting Combined Individual Data for Logistic Regression Analysis
Posted by cudis on Mon, 02 Feb 2015 03:32:20 GMT
View Forum Message <> Reply to Message

We would like to conduct analysis of determinants of employment by estimating a logistic regression of the dichotomous employment outcome variable on various individual characteristics (e.g., education, age, age^2, sex, etc.).

From reading previous topics, our understanding is that DHS recommends weighting data before estimating regressions. However, although certain subpopulations are over- or under-represented in the sample, we cannot see how this would affect a regression onto individual characteristics. Could this please be explained in further detail? We also wonder whether the results of our regression analysis will be affected by the much different sample sizes for the two genders (but perhaps that is another issue entirely).

If we should weight the data, what would be the appropriate weight to use for a combined individual file (i.e., all men and women interviewed), where the unit of analysis is the individual?

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
View Forum Message <> Reply to Message

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).

Subject: Re: Weighting Combined Individual Data for Logistic Regression Analysis
Posted by Alanood on Tue, 28 Mar 2023 02:36:38 GMT
View Forum Message <> Reply to Message

Hello

Regarding your reply, " 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)."

I have a difficulty with my Instrument variable method results (weighted and unweighted) , the sata command for weigted is as follow

ivregress 2sls ....(outcome variable) (controls) (fixed effect) ..... (endogenous variable = instrument)[pweight = sampwt], vce(cluster v021) first

unweighted
ivregress 2sls ....(outcome variable) (controls) (fixed effect) ..... (endogenous variable = instrument), robust first

1- the results are differed between the weighted and unweighted, in the unweighted option: the main independent variable is significant at 3 out of 4 regressions outcome variables , where in the weighted option , i only got one outcome variable out of 4 regressions is significant, what do you think the reason?
2- how we can include the strata option in the ivregress command ? "svy" is not alowed with "ivreg" and i tried to include "strata" in the regression but I got an error.

Thank you in advanced
