Subject: When to use iweight and pweight in stata
Posted by chikhungulana on Fri, 29 Sep 2017 13:15:32 GMT
View Forum Message <> Reply to Message

Hello

I recently encountered differences in statistical significance for my SPSS and STATA analysis. One explanation is that I am using different types of weights. In SPSS I simply click on weight by the Sampleweight variable but in STATA I was using svyset [pweight=Sampleweight]. I see that the DHS manual recommends one to use iweight, but how can one use iweight for a cross tabulation and regression modelling?

Your help will be greatly appreciated.

Kind regards

Lana

Subject: Re: When to use iweight and pweight in stata
Posted by Bridgette-DHS on Mon, 02 Oct 2017 18:27:20 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

My rule is to always use pweight if it is accepted. Unfortunately there are some commands in Stata, such as tabulate and summarize, that will not accept pweight. Those commands will accept iweights, and for them I will use, say, iweight=v005/1000000. The division by 1,000,000 will give weights with an average value of 1. But if you want to use tabulate with an option such as chi2, you can't. Even if you use svyset and pweight, you cannot do tabulate and chi2. So far as I know, virtually all of the estimation commands will accept pweights. (There are some esoteric exceptions and I expect them to evolve to accept pweights in the future.)

You say that "the DHS manual recommends one to use iweight". Which manual are you referring to? I cannot find that recommendation.

I hope other users will add suggestions.

Subject: Re: When to use iweight and pweight in stata
Posted by chikhungulana on Mon, 02 Oct 2017 19:05:12 GMT
View Forum Message <> Reply to Message

Thanks Tom. I was trying to work out why I am getting different statistical results in SPSS and

STATA, Someone suggested that one the programmes should be using wrong weights. With respect to the recommendation to use iweight, I saw this on the DHS website: https://dhsprogram.com/data/Using-DataSets-for-Analysis.cfm# CP_JUMP_14042. I assumed one has to use iweight, but I suppose this for calculating simple percentages.

Attached are my chi square outputs from SPSS and STATA. The percentages are similar but the p values are different.

Kind regards

Lana

File Attachments
1) SPSS and STATA outputs.pdf, downloaded 987 times

Subject: Re: When to use iweight and pweight in stata
Posted by Reduced-For(u)m on Mon, 02 Oct 2017 19:49:13 GMT
View Forum Message <> Reply to Message

Maybe I'm totally wrong, but are you accounting for clustering at PSU in the SPSS version? If not, that could explain the differences in precision across the two estimates, since the point estimates look very similar.

Subject: Re: When to use iweight and pweight in stata
Posted by chikhungulana on Mon, 02 Oct 2017 20:12:32 GMT
View Forum Message <> Reply to Message

Thank you again. I am using the same weight variable: v005/1000,000 which I have defined as Sampleweight. In SPSS I simply choose weight cases by Sampleweight while in STATA I use the following command svyset [pweight=Sampleweight]. I dont think I am encountering for clustering in both SPSS and STATA.

Subject: Re: When to use iweight and pweight in stata
Posted by Reduced-For(u)m on Mon, 02 Oct 2017 22:33:40 GMT
View Forum Message <> Reply to Message

You are clustering your standard errors in Stata, because it says the number of clusters in your output. You probably set that up with your "svyset" command (not on printout how you set up svyset, but I'm guessing you set the PSU as a "cluster" which is what the DHS recommends).

But I'll bet your SPSS code does not do this. Essentially, "clustering" will allow your confidence intervals/p-values to account for within-group correlations in error terms (because people in the

same village are more similar to each other than two total strangers). The result is that you have to inflate your standard errors (make your p-values bigger) to account for the within-cluster similarities among people, otherwise you get p-values that are much too small relative to the "true" value.

So I think the difference is that in Stata you are (correctly) allowing for correlations among error terms and within-cluster heteroskedasticity (via your svyset command that references the PSUs) and this is generating (appropriately) larger standard errors and p-values than your SPSS code which is treating each observation as independent.

I'd look up "clustered standard errors" in SPSS, add that to your code (clustering at the PSU level) and I suspect at that point your p-values will become much closer (that is, the SPSS estimates will look more like the Stata estimates). I don't actually know how to do it in SPSS, but it should be relatively straight forward.

Again - this is not about weighting variables, that only affects point estimates. This is about accounting for clustering in your standard error/p-value calculations.

Subject: Re: When to use iweight and pweight in stata
Posted by chikhungulana on Tue, 03 Oct 2017 08:10:29 GMT
View Forum Message <> Reply to Message

Thank you. I will look up how to account for clustering in SPSS.

Kind regards

Lana

Subject: Re: When to use iweight and pweight in stata
Posted by Bridgette-DHS on Tue, 10 Oct 2017 14:34:03 GMT
View Forum Message <> Reply to Message

A response from DHS Stata Specialist, Tom Pullum:

Below I will insert some Stata lines that you can run after you have opened a KR file. If you want to adjust for weights, clustering, and stratification in a table, the best way (in my opinion!) is with logit (if one of the variables is binary) or mlogit (if both variables have more than two categories). You cannot get a chi square, but you can get the p-value for an F, which will be an equivalent test of the significance of the association. Note that the correspondence is with a likelihood ratio chi-square, rather than a Pearson chi-square. The correspondences between the following approaches with Stata add to my confidence in how Stata handles weights. You could check whether you get the same correspondences with SPSS.

gen stunted=.
replace stunted=0 if hw70<600

```
replace stunted=1 if hw70<-200

gen age=b8
replace age=. if v008-b3<6

tab stunted age, lrchi2
scalar pvalue=r(p_lr)
scalar list pvalue

logit stunted i.age
scalar pvalue=e(p)
scalar list pvalue

svyset v021 [pweight=v005], strata(v022) singleunit(centered)

svy: logit stunted i.age
scalar pvalue=e(p)
scalar list pvalue
```