Subject: weights for ttest
Posted by MissFay on Wed, 26 Mar 2014 21:08:35 GMT
View Forum Message <> Reply to Message

Hi all,
I want to compare HIV knowledge between the 2005 and 2011 Ethiopia DHS. I have appended these files.I have to use weights but I don't think you can use it with the the ttest.
I think you can use svy by comparing means but I am not sure how. Can someone please explain how to compare means and incorporate survey weights.
thanks kindly
Fay

Subject: Re: weights for ttest
Posted by Reduced-For(u)m on Wed, 26 Mar 2014 21:59:03 GMT
View Forum Message <> Reply to Message

Here's one option. Append the two files together. Make separate strata/PSU variables for each survey, and a dummy variable indicating the later survey. Then:

svy: reg HIVOutcome Survey2011

The coefficient on Survey2011 should be the thing you are interested in, assuming you have a continuous measure of some sort. If you want to compare categories it gets a bit trickier, but basically by appending the surveys and using a regression with a dummy for later survey you should be able to use weights and appropriate standard error calculations.

*Note: there is some disagreement here about how to use weights in multiple survey rounds. Weights as given sum to sample size, so I usually standardize weights by survey round (divide the given weight by the sum of weights for that survey round) so that each survey has equal total weight, but that is a preference and shouldn't matter too much.

That sufficient?

Subject: Re: weights for ttest
Posted by Liz-DHS on Thu, 27 Mar 2014 18:56:20 GMT
View Forum Message <> Reply to Message

Dear User,
Here is another response from one of our experts, Dr. Shea Rutstein:

The following have sampling errors estimated in the 2005 Ethiopia DHS:
Has heard of HIV/AIDS Proportion All women 15-49

Knows about condoms Proportion All women 15-49
Knows about limiting partners Proportion All women 15-49
Had 2+ sex partners in past 12 months Proportion All women 15-49
High-risk sex Proportion All women 15-49 with sexual intercourse in past 12 months
Condom use at high-risk sex Proportion All women 15-49 with high-risk intercourse in past 12 months
Abstinence among youth Proportion Women 15-24
Sexually active in past 12 months among youth Proportion Women 15-24
Had an injection in past 12 months Proportion Women 15-24
Had HIV test and received results in past 12 months Proportion All women 15-49
Accepting attitudes towards people with HIV Proportion All women 15-49 who have heard of HIV/AIDS
HIV prevalence among tested for HIV 15-49 Proportion All women 15-49 with blood sample tested at lab

In the 2011 survey sampling errors were calculated for

Had 2+ sex partners in past 12 months Proportion All women 15-49
Abstinence among youth (never had sex) Proportion All never married women 15-24
Sexually active in past 12 months among never-married youth Proportion All never married women 15-24
Had an injection in past 12 months Proportion All women 15-49
Had an HIV test and received results in past 12 months Proportion All women 15-49
Accepting attitudes towards people with HIV Proportion All women 15-49 who have heard of HIV/AIDS
Knows about condoms Proportion All women 15-49
Knows about limiting partners Proportion All women 15-49
HIV prevalence among all women 15-49 Proportion All interviewed women with Dried Blood Sample (DBS) tested at the lab

Since the samples are independent to estimate the sampling error for the trend between the two surveys, the formula var(diff)=var(2005 survey)+var(2011 survey),

Where var is the variance or square of the sampling error of the indicator of interest, and diff is the difference of the indicators. The square root of var(diff) would be the sampling error of the difference between the two surveys, se(diff). The 95% confidence interval would then be calculated as diff-1.96*se(diff) to diff+1.96*se(diff). If the confidence interval includes the value zero, the difference is not significant.

For variables not covered by the report tables, you need to calculate the sampling errors using the appropriate STATA or SPSS (or SAS) commands. In SPSS, first use CSPLAN to describe the sampling plan of the data, then use CSTABULATE to tabulate the indicators and their sampling errors. Alternatively, means with bootstrap can be used (with proper definition of the strata). In STATA, use SVYSET to describe the sampling plan and SVY: MEAN to tabulate the indicators. Remember to use the sampling weights.

Subject: Re: weights for ttest
Posted by Reduced-For(u)m on Thu, 27 Mar 2014 20:40:14 GMT
View Forum Message <> Reply to Message

Thanks for that Liz/Shea.

I've always needed a reason to check the equivalence of these two methods using DHS data, so I did a quick comparison of changes in HAZ over time in Bangladesh as a test case. Let me just confirm that both of these methods generate the same point estimates, and very similar standard errors (probably just rounding differences, and/or something about my specification of strata/psu across survey rounds - but the regression form generated slightly larger se's).

I would add that the regression version took me like 1/10th the time, but that was because retrieving the variance estimates from "svy: mean" took me a minute as they are returned in a matrix.

So point is - Fay: they both work, and statistical theory is vindicated!

Subject: Re: weights for ttest
Posted by Sarah B on Tue, 14 Apr 2015 18:45:14 GMT
View Forum Message <> Reply to Message

Just a note to say thanks for telling us the result of this comparison, Reduced For(u)m. I've always assumed these approaches were equivalent; very nice to have that confirmed. Cheers!