Subject: Zambia hiv dataset 2001-02
Posted by hannekeyserhegdahl on Thu, 23 Oct 2014 07:44:00 GMT
View Forum Message <> Reply to Message

Hi!

I would like to use the HIV dataset from the 2001-02 Zambia DHS for analysis, but I don´t
understand how to survey set this dataset in STATA. Do I have to use a psu and which variable is
it? I assume that the weight to be used is hiv_wgt.

Thanks for any answers.

Subject: Re: Zambia hiv dataset 2001-02
Posted by Trevor-DHS on Fri, 24 Oct 2014 15:34:57 GMT
View Forum Message <> Reply to Message

It is not possible to use this dataset in the conventional way with the svyset command as we have
no PSU identifier for this dataset. The best you can do is the following:

egen strata=group(hivprov hivresid)
svyset hivid [pweight=hiv_wgt], strata(strata)
Here we are using hivid in place of the psu, which will effectively give one case per "psu".

Subject: Re: Zambia hiv dataset 2001-02
Posted by Reduced-For(u)m on Fri, 24 Oct 2014 19:31:17 GMT
View Forum Message <> Reply to Message

Trevor,

  Given that, on its own, stratification should (weakly) reduce the size of standard errors, where
the clustering on PSU (weakly) inflates them - isn't this likely to produce standard error estimates
(and thus p-values and CIs) that are all too small?

  I would think that clustering on some geographic level greater than PSU would be the more
conservative way to do this. The AIS tabulation plan* seems to indicate that they keep region of
residence. Clustering on region-X-urban/rural might provide enough clusters to use the cluster
robust estimator, but there are small-cluster-number analogs that can be used as well.

*Maybe I'm looking at the wrong documentation:
http://dhsprogram.com/pubs/pdf/AISM9/HIV_Testing_Tabplan.pdf

Subject: Re: Zambia hiv dataset 2001-02
Posted by Trevor-DHS on Fri, 24 Oct 2014 20:43:09 GMT
View Forum Message <> Reply to Message

I asked our sampling specialist Ruilin Ren for his thoughts on this. Here is his reply:

"It is true that without declaring the cluster, the standard error of the estimators will be under estimated, especially for indicators with strong design effect. But taking the stratum (province x residence) as cluster will over estimate the standard error equally. So there is no perfect solution. For HIV prevalence, it may not be a major problem without declaring the cluster because HIV prevalence usually has weak design effect compared to other indicators."

Subject: Re: Zambia hiv dataset 2001-02
Posted by hannekeyserhegdahl on Tue, 28 Oct 2014 15:07:30 GMT
View Forum Message <> Reply to Message

Thank you!

Subject: Re: Zambia hiv dataset 2001-02
Posted by hannekeyserhegdahl on Mon, 05 Jan 2015 08:47:19 GMT
View Forum Message <> Reply to Message

Hi again!

I have the same problem with the Mali 2001 hiv dataset as with the Zambia 2001-02 hiv dataset. Can I use the same method on the dataset from Mali?

Subject: Re: Zambia hiv dataset 2001-02
Posted by Trevor-DHS on Mon, 05 Jan 2015 14:56:31 GMT
View Forum Message <> Reply to Message

Yes, the Mali 2001 HIV dataset can be treated the same way as the Zambia 2001-02 HIV dataset.

Subject: Re: Zambia hiv dataset 2001-02
Posted by CKAllen on Thu, 25 Aug 2016 19:19:42 GMT
View Forum Message <> Reply to Message

Because there is no PSU identifier, does this mean this dataset is impossible to merge with the IR dataset?

Subject: Re: Zambia hiv dataset 2001-02
Posted by Trevor-DHS on Fri, 26 Aug 2016 00:55:08 GMT
View Forum Message <> Reply to Message

Yes, it is not possible to merge the HIV test results for Zambia 2001-02, Mali 2001, and Dominican Republic 2002 with the Individual Recode datasets. These surveys were the first three to collect blood samples for HIV testing and at the time the protocol called for completely de-linked testing. However, these datasets do include a few characteristics of respondents such as age, sex, marital status, type of place of residence.