Subject: analysing antenatal care between 2017 & 2022 datasets
Posted by lulayct on Sun, 28 Jan 2024 15:54:48 GMT
View Forum Message <> Reply to Message

Hi. This is my very first tiem to use DHS datasets. My research question involves comparing aspects of antenatal care between 2 time periods (2017 & 2022) using multivariable logistic regression. I noticed that the 2022 DHS introduced a new Pregnancy file (GR) while the 2017 only has Individual and Birth Files. My questions are:

1) Will i be able to obtain the same details about antenatal care from the 2022 DHS Individual (IR) File that were present in the 2017 DHS?

2) How do i merge the 2017 and 2022 datasets in Stata so that logistic regression can be done?

3) I also noticed that there is a difference in the selection of women asked to provide information about antenatal care between the 2 datasets: for 2017, women with births in the past 5 years (B19<60) were selected but for 2022, only women with births in the past 2 years (P19<24). Is it possible for me to limit the 2017 data for my analysis on antenatal care to just those with B19<24?

Thank you for your help.

Subject: Re: analysing antenatal care between 2017 & 2022 datasets
Posted by Janet-DHS on Wed, 31 Jan 2024 14:39:26 GMT
View Forum Message <> Reply to Message

Following is a response from DHS staff member, Tom Pullum:

You did not say what country you are working with. Whatever country it is, the 2017 and 2022 surveys are independent cross-sections and you cannot to a case-by-case merge. You can append one to the other, but I don't think much will be gained, because the 2022 survey (as a DHS-8 survey) includes a pregnancy history, not just a birth history (plus calendar).

The same information about ANC is available in both surveys.

If you want to make your results from the two surveys more comparable, then yes, you can use the same time interval (the last 24 months, with p19<24 or b19<24) for both of them. But you have to be careful with births in one survey and pregnancies in the other. DHS-8 surveys have an important variable for each pregnancy, p32. If you select pregnancies with p32=1 then you will have live births, corresponding with birth histories in pre-DHS-8 surveys.

Subject: Re: analysing antenatal care between 2017 & 2022 datasets
Posted by lulayct on Thu, 01 Feb 2024 00:16:36 GMT
View Forum Message <> Reply to Message

Thank you for replying, Janet & Tom.

I am working with the Philippine National DHS data 2017 and 2022. The analysis plan is to compare the proportions or percentage of women having 4+ and 8+ antenatal care visits, so it will not be a case-by-case merge.

I have found the relevant files used for reporting Antenatal Care in both datasets: KR or IR file in 2017 and NR file in 2022. Is there a way I can obtain variables for antenatal care from different files to make this proposed comparative analysis? Because if this will be too complicated, I will change my project proposal and work on a research question focused purely on the 2022 Philippine NDHS dataset.

Again, Thank you very much for your help.

Subject: Re: analysing antenatal care between 2017 & 2022 datasets
Posted by lulayct on Thu, 01 Feb 2024 13:54:07 GMT
View Forum Message <> Reply to Message

Thank you for replying, Janet & Tom. I am revising my reply from what I sent a few hours ago. I received permission to access Philippine NDHS datasets and started exploring them to see how I can perform by comparative analysis with logistic regression.

From the Guide to DHS Statistics of 2017 & 2022, i noticed that the data i need for analysing features of antenatal care can be found in the IR file (Women's) and that observations to the ANC variables i intend to study are present only for women with midx_1=1 in both datasets.

I am planning to choose only women with midx_1=1 for my study. To combine these women from the two DHS datasets (2017 & 2022), is it possible to append them after I generate a variable that will allow me to group the 2017 observations from the 2022 observations. For this new variable, I was considering using the v008a variable (century day code of date of interview) which has a maximum value of 43029 for the 2017 dataset, and minimum value of 44683 for the 2022 dataset. For this new variable called "studygroup", i can assign studygroup=1 (pre-pandemic) if v008a<43030 and studygroup=2 (pandemic) if v008a>44680.

Is this a feasible strategy to be able to extract the data on variables i intend to analyse and compare one group with the other using multivariable logistic regression?

Thank you very much again for your help.

Subject: Re: analysing antenatal care between 2017 & 2022 datasets
Posted by lulayct on Thu, 01 Feb 2024 15:06:43 GMT
View Forum Message <> Reply to Message

To add to my last post, I am not planning to replicate the tables from the Final DHS reports. The

denominator for the proportions of women with 4+ and 8+ ANC visits (and for the other variables to be compared) will be women aged 15-49 yrs who had a livebirth in the 24 months prior to the survey interview.

Thanks.

Subject: Re: analysing antenatal care between 2017 & 2022 datasets
Posted by Janet-DHS on Mon, 05 Feb 2024 20:44:42 GMT
View Forum Message <> Reply to Message

Following is a response from DHS staff member, Tom Pullum:

It will be easier if you use the KR files, which have one record per child born in the past 5 years. The relevant variable on number of ANC visits is m14. It is only coded for children with bidx=1 (or midx=1) but you do not need to reduce the file.

You do not need to define the outcome in terms of 4+ visits or 8+ visits, because that would amount to throwing out some of the information. Instead of logit regression, you could use linear regression. You should use svyset and svy, and you could have a 2-category predictor that is 1 in the first time period and 2 in the second time period.

To illustrate, but not using svy, I opened the KR file in the 2022 survey and entered the following lines:

gen visits=m14
replace visits=. if visits==98
regress visits v025

```
     Source |    SS        df    MS    Number
of obs  =   7,974

-------------+---------------------------------  F(1, 7972)    =   214.64

     Model |  2188.613      1  2188.613  Prob >
F     =   0.0000

   Residual |  81287.9595   7,972 10.1966833  R-squared    =
0.0262

-------------+---------------------------------  Adj R-squared  =   0.0261

     Total |  83476.5725    7,973 10.4699075  Root MSE
=   3.1932
```

```
     ------------------------------------------------------------- -----------------

     visits |    Coef.  Std. Err.     t   P>|t|    [95% Conf.
Interval]

     -------------+----------------------------------------------- -----------------

       v025 |  -1.118605  .0763522  -14.65  0.000   -1.268275
-.9689345

       _cons |   8.384525  .1328004   63.14  0.000    8.124202
8.644849

     ------------------------------------------------------------- -----------------
```

Here I used a different variable, v025 (place of residence), just because it also takes the values 1 and 2. In this example, the coefficient for v025 is -1.12, and it is highly significant.

But you should be careful in your interpretation if the predictor is time. Other variables, not just Covid, could be associated with time, and it's risky to say that a change from time 1 to time 2 is due to Covid. Hope this helps.

Subject: Re: analysing antenatal care between 2017 & 2022 datasets
Posted by lulayct on Wed, 07 Feb 2024 13:11:35 GMT
View Forum Message <> Reply to Message

Thank you for your reply.

Will use the KR file for both datasets then. Will discuss whether to analyse number of antenatal care visits as a linear variable or a categorical one to exhibit compliance with WHO recommendations of at least 4 ANC visits (2012) and at least 8 visits (2016).

Lora