
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.
