
Subject: Re: Matching most recent sexual activity (women) in India_2019/21

Posted by [Bridgette-DHS](#) on Fri, 19 Aug 2022 11:13:15 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

I'll answer in Stata. I don't use SPSS but I think you will be able to see the logic. This variable is based on v527, which is one of the DHS variables that has the units in the first column and the number in the next two columns. Here are the possible units of time in the first column: 1 for days, 2 for weeks, 3 for months, 4 for years. The code 998 is used for "don't know / refused to answer". "tab v527" and "label list V527" will help you to see how v527 is coded. You also have to look at v536 to get "never".

Here are the Stata lines that I would use:

```
gen time=6 if v536==0
replace time=1 if v527<107
replace time=2 if v527>=107 & v527<128
replace time=3 if v527>=128 & v527<400
replace time=4 if v527>=400 & v527<998
replace time=5 if v527==998
```

```
label define time 1 "<1 week" 2 ">1 week, <1 month" 3 ">1 month, <1 year" 4 "1+ years" 5
"Missing" 6 "Never"
label values time time
```

```
gen time12=0 if time<=6
replace time12=1 if time<=2
label variable time12 "< 1 month"
```

```
gen time123=0 if time<=6
replace time123=1 if time<=3
label variable time123 "< 1 year"
```

```
tab1 time time12 time123 [iweight=v005/1000000]
```

This includes the construction of time12, which combines categories time=1 and time=2 to get "within last 4 weeks". Also, time123 combines categories 1, 2, and 3 of time to get "within 1 year". The table is a little unusual because of how it combines categories within the distribution. The values of v527 and time are mutually exclusive, so the combinations of categories must be constructed. It would take some special formatting to get the columns as they are in the report.

When I do this, the n is slightly off. I get n=107,956, whereas the report has n=108,014. The difference of 18 cases is very small and I can't take the time to reduce it. The distribution I get differs from the report, but this has to do with how the values of v527 are interpreted. The number of days (leading digit 1) goes from 0 to 90. The number of weeks (leading digit 2) goes from 1 to

50. The number of months (leading digit 3) goes from 1 to 11 (which is ok!); the number of years (leading digit 4) goes from 1 to 90. There are inconsistencies and ambiguities. I don't have the original program for this table and don't know how those cases were handled, but however they were handled, there is underlying ambiguity in the data. It is likely that in some combinations of unit and number it is the number that is wrong, and in other cases it is the unit that has been entered incorrectly. You could tinker with how v527 is recoded into the categorical variable I call "time".
