
Subject: Re: Turkey mother tongue variable
Posted by [Bridgette-DHS](#) on Thu, 04 Feb 2016 15:03:05 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

For tables you do not need the full svyset command. You only need to specify the weights. The results are more readable with something like "tab sh57_1 {iweight=hv005/1000000}".

I looked at the household questionnaire in the report on the 2003 survey. sh57 and other sh5* variables refer to never-married women (NMW) age 15-49 who are living in the household. The subscripts _1, _2, etc., are used to identify specific people (never-married women age 15-49) in the household. Values will be missing (".") if there are fewer than 8 eligible persons in the household.

Normally such a variable would also appear in the PR file, just as sh57, with no subscript, for the women 15-49 who were not eligible for the survey of women, which was limited to ever-married women. It would be easier for you to use the PR file rather than the HR file. Then you just have sh57 for the people in the household to whom it applied. The units in the HR file are households and the units in the PR file are individuals.

I looked for sh57 in the PR file. It is there, but all cases are coded ".", i.e. are missing. This is strange. There must have been an error in the construction of the PR file.

In the HR file, sh51a_*, "line number in household schedule", is included. I used it to merge sh57 with the PR file with the following steps (change path names, of course). Note that within this merge, the existing incorrect version of sh57 (with all values ".") must first be removed from the PR file, or it will override the new value.

```
* Copy sh57 from the HR file to the PR file
set more off
use e:\DHS\DHS_data\HR_files\TRHR4AFL.dta, clear
keep hv001 hv002 sh51a_* sh57_*
reshape long sh51a_ sh57_, i(hv001 hv002) j(nmw)
drop if sh51a_==.
rename sh51a hvidx
rename sh57_ sh57
drop nmw
```

```
sort hv001 hv002 hvidx
save e:\DHS\DHS_data\scratch\temp.dta, replace
```

```
use e:\DHS\DHS_data\PR_files\TRPR4AFL.dta, clear
```

```
* You must drop the existing imperfect copy of sh57 !!
drop sh57
sort hv001 hv002 hvidx
merge hv001 hv002 hvidx using e:\DHS\DHS_data\scratch\temp.dta
```

```
tab sh57 _merge,m  
keep if _merge==3
```

```
tab sh57  
tab sh57 [iweight=hv005/1000000]
```

The last two lines give the unweighted and weighted tabulations of sh57. The number of never-married women who are Kurdish speakers is 887 (unweighted) or 599 (weighted). I can confirm these numbers by adding up the numbers from sh57_1 through sh57_8 in the HR file. I do not get your number, 559.

I may not have directly answered your question, but I have shown you how to do much more, because once you have a valid version of sh57 in the PR file, you can actually do some analysis to compare the Kurdish NMWs with the Turkish NMWs, etc.

There are other questions about language as part of the household interview (the language of the interview) and in the survey of ever-married women (go to the HR or IR file and enter "lookfor language").