

---

Subject: Calculating median age at first sex and percentage of respondents with sex before the age of 15

Posted by [chichi](#) on Tue, 14 Mar 2017 14:26:30 GMT

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

---

Hello!

I am using the Namibian DHS 2013. I created one dataset with the women's, the men's and the HIV data.

I want to calculate the median age at first sex intercourse by women and men in the age group 25-49. First I checked if I have the same number of women and men as in the final report of 2013.

This was my stata code:

```
tab v525 if v012 > 24 & v012 < 50 [iweight=wgt1], miss
```

I had 7.776 respondents like in the report. Then for calculating the median I used this code:

```
by gender, sort : summarize v525 if v012 > 24 & v012 < 50 [aweight = wgt1], detail
```

The median number which came out was the same as in the report. But the number of observations for men and women was different. Overall I had 7.746 respondents, 30 respondents fewer than in the report. Can someone tell what I made wrong?

Further I want to get out the percentage of women and men who had sex before the age of 15. my code was:

```
by gender, sort: tab v525 if v012 > 24 & v012 < 50 [iweight=wgt1], missing
```

My result was 7.9% for women and 12.5% for men, while the report indicated 5.0% for women and 10.1% for men in the age group 25-49 years.

It would be great if someone could help me or have a stata code for my intention! Thank you!

---

---

Subject: Re: Calculating median age at first sex and percentage of respondents with sex before the age of 15

Posted by [Bridgette-DHS](#) on Mon, 27 Mar 2017 13:28:24 GMT

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

---

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

Why do you keep using the "missing" option? In DHS surveys, a dot means "Not applicable". You should not include those cases in any of the analysis. Please try your commands again, excluding "missing". Let us know if you still have a problem.

---

---

Subject: Re: Calculating median age at first sex and percentage of respondents with sex before the age of 15

---

Posted by [cmhabito](#) on Mon, 07 Aug 2017 06:10:19 GMT

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

---

Hello,

I have been using the individual woman's dataset from the Philippines NDHS 2013, and am having some trouble replicating the median age at first sex by age group (Table 4.5) that appear in the final report. I have been using Stata 14 to analyze the data, and have read the relevant threads on this user forum ([http:// userforum.dhsprogram.com/index.php?t=msg&th=5960&start=0](http://userforum.dhsprogram.com/index.php?t=msg&th=5960&start=0) and [http:// userforum.dhsprogram.com/index.php?t=msg&th=1337&got o=2470&S=08c954804736d97c007deae6e959abed#msg\\_2470](http://userforum.dhsprogram.com/index.php?t=msg&th=1337&got o=2470&S=08c954804736d97c007deae6e959abed#msg_2470)), but have still not had any luck generating the same values that appear in the final report. My Stata code is as follows:

```
*** Survey design
```

```
gen weight = v005/1000000
```

```
svyset [pweight = weight], psu(v021) strata(v023)
```

```
*** First, checking the percent distribution of age at first sex for women 25-49 years old:
```

```
svy: tab v531, obs percent format(%9.1f) // v531 is age at first sex (imputed)
```

```
*** Computing for median age at first sex for respondents 25-49 years old, Option A:
```

```
_pctile v531 if v012 > 24 & v012 < 50 [pweight = weight], p(50) // v012 is current age of respondent
```

```
return list // returns median = 20
```

```
// OR
```

```
*** Computing for median age at first sex for respondents 25-49 years old, Option B:
```

```
summarize v531 if v012 > 24 & v012 < 50 [aweight = weight], detail // returns median = 20
```

According to Table 4.5 in the Philippines NDHS 2013 final report, the median age at first sex for this age group should be 21.5.

Could you please help me figure out the reason why I am not generating the correct median value? I am not sure if it has to do with my code or with the variables I am using (or both), but I would appreciate any help you could offer.

Thank you!

Cheers,  
Marie

---

---

Subject: Re: Calculating median age at first sex and percentage of respondents with sex before the age of 15

Posted by [Bridgette-DHS](#) on Mon, 07 Aug 2017 13:35:52 GMT

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

---

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

You need to use weights, but the rest of svyset is not relevant for calculating a point estimate. The main issue is that in Stata, medians (and other percentiles) are calculated as integers. You have to interpolate to get anything to the right of the decimal place. Another issue is that you need to recode 0, 98, and 99. The following program will match DHS calculations. It gives 21.46, which rounds to 21.5. To run it, you must change the path to the data file.

\* Calculation of median age at first sex

```
set more off
```

```
*****
```

```
program define calc_median_age
```

```
summarize age [fweight=v005] if v012>=25 & v012<=49, detail
```

```
scalar sp50=r(p50)
```

```
gen dummy=.
```

```
replace dummy=0 if v012>=25 & v012<=49
```

```
replace dummy=1 if v012>=25 & v012<=49 & age<sp50
```

```
summarize dummy [fweight=v005]
```

```
scalar sL=r(mean)
```

```
replace dummy=.
```

```
replace dummy=0 if v012>=25 & v012<=49
```

```
replace dummy=1 if v012>=25 & v012<=49 & age<=sp50
```

```
summarize dummy [fweight=v005]
```

```
scalar sU=r(mean)
```

```
drop dummy
```

```
scalar smedian=round(sp50+(.5-sL)/(sU-sL),.01)
```

```
scalar list sp50 sL sU smedian
```

```
* warning if sL and sU are miscalculated
```

```
if sL>.5 | sU<.5 {
```

```
ERROR IN CALCULATION OF L AND/OR U
```

```
}
```

```
drop age
```

```
end
```

```
*****
```

```
*****
```

```
*****
```

```
*****
```

```
*****
```

\*\*\*\*\*

\* EXECUTION BEGINS HERE

\* sp50 is the integer-valued median produced by summarize, detail;  
\* what we need is an interpolated or fractional value of the median.

\* In the program, "age" is reset as age at first cohabitation or age at first birth;  
\* with modifications, other possibilities would require modifications.

\* sL and sU are the cumulative values of the distribution that straddle the integer-valued median

\* v011 date of woman's birth (cmc)  
\* v211 date of first child's birth (cmc)  
\* v511 age at first cohabitation

```
set maxvar 10000  
use e:\DHS\DHS_data\IR_files\PHIR61FL.dta, clear
```

```
* age at first sex calculated from v531  
gen afs=v531  
replace afs=99 if v531==. | v531==0  
replace afs=. if v531==98 | v531==99  
gen age=afs  
calc_median_age  
scalar safs_median=smedian  
scalar list safs_median
```

---

Subject: Re: Calculating median age at first sex and percentage of respondents with sex before the age of 15

Posted by [cmhabito](#) on Thu, 10 Aug 2017 08:23:50 GMT

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

---

Dear Dr. Pullum and Bridgette,

Thank you for this! Much appreciated - I have been stuck on this for a while. I have been able to adjust the code to compute for median age at first sex for each of the 5-year age groups as well.

Good day to you, and thanks again!

Cheers,  
Marie