

---

Subject: Calculating Median Ages  
Posted by [hlantos](#) on Fri, 06 Jun 2014 15:37:31 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hello all, I'm trying to develop STATA code for median ages of variables that are age dependent (right now age at first sex, age at first marriage, and age at first birth). I've found the following explanation:

- 1) Age at first marriage or first union is calculated as the difference between date when woman began living with first husband or consensual partner and date of birth of woman in completed single years.
- 2) The numerators are the number of women within single year of age categories who have married or lived in a consensual union.
- 3) The denominator is the number of women of all marital statuses.
- 4) Numerators for each age category are divided by the corresponding age category denominator and multiplied by 100 to obtain percentages.
- 5) Once the percentages have been calculated within specific age group categories, medians are calculated from the cumulated single year of age percent distributions for the ages women were first married. The median is linearly interpolated between the age values by which 50 percent or more of the women were first married or lived in consensual union.

I understand each of the steps until step 5... a) what does it mean to calculate a median from the cumulated single year of age distributions - am i combining them all into a weighted sample essentially? and b) how would any of you think about doing that in STATA?

Thanks!

Hannah

---

---

Subject: Re: Calculating Median Ages  
Posted by [Liz-DHS](#) on Fri, 06 Jun 2014 16:01:48 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dear User,  
I have forwarded your question to one of our experts. Depending on availability, they will answer as soon as they can. In the meantime, please see the following resources on our website: Guide to DHS Statistics <http://dhsprogram.com/publications/publication-dhsg1-dhs-questionnaires-and-manuals.cfm>, Standard Recode Manual [http://dhsprogram.com/pubs/pdf/DHSG4/Recode6\\_DHS\\_22March2013\\_DHSG4.pdf](http://dhsprogram.com/pubs/pdf/DHSG4/Recode6_DHS_22March2013_DHSG4.pdf) and a link on our website called Using Datasets for Analysis <http://dhsprogram.com/data/Using-Datasets-for-Analysis.cfm>  
Thank you!

---

---

Subject: Re: Calculating Median Ages  
Posted by [hlantos](#) on Tue, 24 Jun 2014 19:07:50 GMT

---

Liz,

Thanks for this email. I still haven't heard from anyone - any more tips on where there might be STATA code for this written?

Thanks!

Hannah

---

---

Subject: Re: Calculating Median Ages  
Posted by [Liz-DHS](#) on Tue, 24 Jun 2014 19:27:04 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dear User,  
We're still working it out. We should have something for you within a day or two.  
Thank you!

---

---

Subject: Re: Calculating Median Ages  
Posted by [Liz-DHS](#) on Tue, 24 Jun 2014 19:54:28 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dear Hannah,  
Here is a response from one of our experts, Dr. Tom Pullum:  
\* Kenya 2008-09 survey report, page 83, table 6.3, median age at first  
\* marriage for women currently 25-29 is 20.2. Try to replicate.

use KEIR53FL.dta

set more off  
keep if v013==3

keep v000-v012 v507-v512

describe

\* afm is completed years of age at first marriage  
gen afm=int((v509-v011)/12)  
tab afm,m

tab afm [iweight=v005/1000000],m

\* look at the cumulative %; it is 48.37 by exact age 20 and 56.12 by exact age 21.

scalar mafm=20+1\*(50-48.37)/(56.12-48.37)

scalar list mafm

\* The value is 20.2

\* There are more automated ways to do this. If you try to do it with the cumul command,  
\* you must set the missing ages at marriage (i.e. not married) to some high numerical  
\* value such as 99

\* Note that the median as defined by stata would be 20, an integer.

---

Subject: Re: Calculating Median Ages  
Posted by [hiantos](#) on Tue, 24 Jun 2014 21:00:56 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Liz and Dr. Pullum,

This is very helpful. Thanks so much!

Hannah

---

---

Subject: Re: Calculating Median Ages  
Posted by [hiantos](#) on Tue, 08 Jul 2014 13:58:48 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Liz and Dr. Pullum,

I've been trying this code with a number of other countries (our project is on West and Central Africa so I didn't have Kenya and was trying to make it work there first) and was getting slightly different numbers than the final reports. Sometimes by exactly a year, which was strange. I was thinking maybe the samples are different from what I'm using and what they're using so I finally requested the Kenya data and am trying this again.

And, again, I am getting a number that is off by exactly a year from what your code says. Am I misinterpreting "exact age 20"? I have 48.37 by exact age 19 instead of exact age 20.

Also, just to confirm, your code uses a dataset called KEIR53FL.dta and the one on the website is KEIR52FL.dta. Are these the same?

Thanks,  
Hannah

---

---

Subject: Re: Calculating Median Ages  
Posted by [hiantos](#) on Tue, 08 Jul 2014 14:15:29 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Oh I just figured it out I think. Essentially 48.37% are cumulatively associated with 19. But that means any fraction between 19 and 20, yes? So by exact age 20 48.37% of people have been married because it's all the way up to 19.9999999 basically, yes?

thanks for all your help!

---

---

Subject: Re: Calculating Median Ages  
Posted by [Liz-DHS](#) on Tue, 08 Jul 2014 15:40:06 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dear User,  
On behalf of Dr. Pullum:  
Yes, that's how we are interpreting age. Thanks--Tom

---

---

Subject: Re: Calculating Median Ages  
Posted by [rbriceno](#) on Wed, 05 Nov 2014 21:04:18 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dear Liz, dear Dr. Pullum. I'm trying to use the stata command reference posted here. Since I do not have access to the survey I would like to ask you if is there a way to know what are those variables used in the proposed set of commands. I mean V509, V011, V005, etc.  
Thanks!

---

---

Subject: Re: Calculating Median Ages  
Posted by [Liz-DHS](#) on Wed, 05 Nov 2014 21:35:24 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Dear User,  
V509 Century month code of the date of start of first marriage or union (see note on century month codes).  
V011 Century month code of date of birth of the respondent (see note on century month codes).  
V005 Sample weight is an 8 digit variable with 6 implied decimal places. To use the sample weight divide it by 1000000 before applying the weighting factor. All sample weights are normalized such that the weighted number of cases is identical to the unweighted number of cases when using the full dataset with no selection. This variable should be used to weight all tabulations produced using the data file. For self-weighting samples this variable is equal to 1000000.

For a full description of recode variable names please refer to The Standard Recode Manual. You can download it from our website <http://>

Quote:Century Month Code

All dates in the data file are expressed in terms of months and years and also as century month codes. A

century month code (CMC) is the number of the month since the start of the century. For example, January

1900 is CMC 1, January 1901 is CMC 13, January 1980 is CMC 961, and September 1994 is CMC 1137.

The CMC for a date is calculated from the month and year as follows:

$CMC = (YY * 12) + MM$  for month MM in year 19YY.

To calculate the month and year from the CMC use the following formulae:

$YY = \text{int}((CMC - 1) / 12)$

$MM = CMC - (YY * 12)$

For Dates in 2000 and after the CMC is calculated as follows:

$CMC = ((YYYY-1900) * 12) + MM$  for month MM in year YYYY.

To calculate the month and year from the CMC use the following formulae:

$YYYY = \text{int}((CMC - 1) / 12) + 1900$

$MM = CMC - ((YYYY-1900) * 12)$

---

Subject: Re: Calculating Median Ages

Posted by [rbriceno](#) on Thu, 06 Nov 2014 15:38:17 GMT

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

Dear Liz-DHS. One more question, maybe to be referred to Dr. Pullum.

When I execute what the syntax express I got a table but the percentages expressed are instead located at 19 and 20 years.

afm	Freq.	Percent	Cum.
10	11.159992	0.77	0.77
11	6.016508	0.41	1.18
12	9.945302	0.68	1.87
13	27.479334	1.89	3.76
14	50.596559	3.48	7.24
15	80.935504	5.57	12.80
16	111.12714	7.64	20.45
17	124.129971	8.54	28.98
18	121.885099	8.38	37.37
19	159.949541	11.00	48.37
20	112.759464	7.76	56.12
21	119.70874	8.23	64.36
22	55.087768	3.79	68.15
23	67.350715	4.63	72.78
24	90.994014	6.26	79.04
25	34.253336	2.36	81.39
26	21.04997	1.45	82.84
27	16.490187	1.13	83.98

28	6.711959	0.46	84.44
29	.498598	0.03	84.47
.	225.767118	15.53	100.00

---

Total	1,453.8968	100.00
-------	------------	--------

Is there a way that the median age has an error in its calculation?

---



---

Subject: Re: Calculating Median Ages

Posted by [Liz-DHS](#) on Tue, 11 Nov 2014 16:33:23 GMT

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

---

Dear User,

Here is a response from Dr. Tom Pullum:

Quote: Say that  $x$  is the name of a variable, such as age at marriage, and  $X$  is a specific value of that variable, e.g. 19.

If you look at how cumulative percentages are calculated, for any variable, in Stata (or in general), the cumulative percentage for variable  $x$  and value  $X$  is the percentage of cases with  $x$  less than or equal to  $X$ . For example, 48.37% is the percentage of cases with  $x \leq 19$  and 56.12% is the percentage of cases with  $x \leq 20$ . I believe you were not including the "=" sign.

Age "19" is a one-year interval interpreted as age at last birthday. The upper boundary for age 19 is the exact day of the 20th birthday, which has an exact value is 20.00 (you can put as many zeroes after the decimal point as you want). That's why I said that 48.37% of women were married by exact age 20 and 56.12% by exact age 21. In terms of "exact" age, the median must be somewhere between 20.00 and 21.00. I gave the steps for finding the median, 20.2.

By the way, here's a trick that will save you some arithmetic. Paste the following lines into Stata:

```
input x P
20 48.37
. 50.00
21 56.12
end
```

```
regress x P
predict xhat
replace x=xhat if x==.
list x P, table clean
```

This will give the value of  $x$  for which  $P=50$ , i.e. the median.

---



---

Subject: Re: Calculating Median Ages

Posted by [akwarae@gmail.com](mailto:akwarae@gmail.com) on Tue, 26 Nov 2019 09:05:17 GMT

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

---

How does one calculate median age groups for sub-groups?

i.e. median age at first marriage by education status or wealth quintile- can you please assist with the STATA code to execute this?

Many thanks,  
Elsie

---