

---

Subject: Re: age at the first birth

Posted by [Bridgette-DHS](#) on Wed, 13 Apr 2016 12:58:48 GMT

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

---

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

The difficulty with tables 7.4 and 7.5 is due to the fact that only ever-married women (EMW) were eligible for the women's interview in this survey. Therefore information about age at marriage is only available for women who are married. To describe age at marriage in the population of all women, you need to know how many women at each age are never-married. You might try to fill in those women by combining the IR and PR files, but that is not how DHS does it. (Actually I have tried to match the tables this way, and I think that's a legitimate strategy, but it will not give a match.) Instead, to calculate tables 7.4 and 7.5, DHS inflates the EMW sample with the all-woman factors (generically, "awfact"). The cases that are artificially added into the sample this way are considered to be never-married. I emphasize that this is very artificial. These are not real cases. The following Stata lines will produce everything in table 7.4 except for the median age at first marriage. Note that the inflation factor here is awfactt. The final "t" is for total. If you wanted to get this table separately for urban/rural women, for example, you would use awfactu. In table 7.5, the successive panels use awfactu, awfactr, awfacte, and awfactw.

```
use e:\DHS\DHS_data\IR_files\EGIR61FL.dta, clear
keep v001 v002 v003 v005 v012 v013 v511 awfactt
```

```
save e:\DHS\DHS_data\scratch\temp.dta, replace
gen weight=v005
gen never_married=0
```

```
save e:\DHS\DHS_data\scratch\emw_temp.dta, replace
use e:\DHS\DHS_data\scratch\temp.dta, clear
```

```
gen weight=round(v005*((awfactt/100)-1))
gen never_married=100
replace v511=50
```

```
append using e:\DHS\DHS_data\scratch\emw_temp.dta
```

```
tab v013 [iweight=weight/1000000]
```

```
format never_married %5.1f
tab v013 [iweight=weight], summarize(never_married) means noobs
```

```
gen before_15=0
replace before_15=100 if v511<15
```

```
gen before_18=0
replace before_18=100 if v511<18
```

```
gen before_20=0
```

```
replace before_20=100 if v511<20
```

```
gen before_22=0
```

```
replace before_22=100 if v511<22
```

```
gen before_25=0
```

```
replace before_25=100 if v511<25
```

```
format before* %5.1f
```

```
tab v013 if v012>=15 [iweight=weight], summarize(before_15) means noobs
```

```
tab v013 if v012>=18 [iweight=weight], summarize(before_18) means noobs
```

```
tab v013 if v012>=20 [iweight=weight], summarize(before_20) means noobs
```

```
tab v013 if v012>=22 [iweight=weight], summarize(before_22) means noobs
```

```
tab v013 if v012>=25 [iweight=weight], summarize(before_25) means noobs
```

```
mean before* never_married if v012>=25 [iweight=weight]
```

Getting the medians in the last column of table 7.4 is also a little complicated, because in Stata a median is always a value that actually occurs in the data. Stata would not give a median age of 21.3, for example, unless 21.3 was actually an age at marriage in the data. Instead, Stata would only give values such as 20, 21, 22, etc. Note also that a median age such as 21.3 implies that age is on a continuous scale, rather than being defined only at integer values.

A second complication with the calculation of the medians in table 7.4 is in the interpretation of, say, age at marriage = 20. It is DHS practice to replace the 20 with 21.0, which is exact age 21, that is, the 21st birthday. That is, continue the program with these lines:

```
gen v511r=v511+1
```

```
replace v511r=. if v511>v012
```

Then, to get the median age at marriage for women age 25-29 (that is, with v013=3), you do this line:

```
tab v511r [iweight=weight/1000000] if v013==3,m
```

Looking at the cumulative percentages in this one-way table, you see a cumulative percentage of 47.48% at v511r=21 and a cumulative percentage of 57.39% at v511r=22. Using linear interpolation between 21 and 22, you would get 50% at age 21.254289, i.e. at 21.3. That's the source of the 21.3 for age group 25-29 in the last column of table 7.4. Similarly for the other age intervals in table 7.4, and similarly for all the medians in table 7.5, but using the appropriate versions of awfact for the covariates as described above.

Of course, all of the calculations of medians could be automated in Stata, but here I'm just showing the basic steps.

I will add that I am personally not comfortable with the line "gen v511r=v511+1". Instead of adding 1, I would probably add .5, thinking of marriage at age 20 at last birthday, for example, as "on the average" marriage at exact age 20.5. However, this is just a matter of interpretation. Any

inferences about trends or differentials would be unaffected by whether you add 1 or .5.