Subject: Estimation of percentage of 20-24 year old women who married before the age of 18
Posted by cpmavelikara18@gmail.com on Tue, 01 Nov 2022 16:30:05 GMT
View Forum Message <> Reply to Message

Hai All,

I have been trying to replicate the percentage of 20 to 24-year-old women who married before the age of 18, which is given in NFHS-5 factsheet for India and individual states. I used the IR (Individual Recode) file and used v511 (age at co-habitation) to replicate the results.

But I was not able to correctly match the results at national level. (Note: I weighted by iw=v005/1 million). The factsheet showed the figure as 23.3, while my estimate was 22.2.

Interestingly, I was able to replicate the state-level estimates for the same for a few states using iw=sweight/1 million.

I am adding the STATA codes I used to estimate these results.

    ** ***********************************
use "C:\Users\cpmav\Desktop\NFHS 5\Individual Recoded (Women 15-49 Years)\IAIR7AFL.DTA", clear

*Dividing women's individual weight by 1 million (national weight)

gen wt=v005/1000000

*Dividing women's individual weight by 1 million (state weight)

gen swt=sweight/1000000

*Marriage Before 18 Years

recode v511 (.=0) (0/17 = 1 "yes") (18/49 = 0 "no"), gen (ms_afm_18)
replace ms_afm_18 = . if v012<18
label var ms_afm_18 "First marriage by age 18"

*Tabulations

tab ms_afm_18 if v013==2 [iw=wt]   /// All India estimate

tab v024 ms_afm_18 if v013==2 [iw=swt], nofreq row /// State-level estimates.

For states like Bihar, Gujarat, Rajasthan and Meghalaya, I have experienced a deviation above 2 percentage points compared to NFHS-5 estimates.

I request scholars and DHS staff to kindly have a look at this.

Subject: Re: Estimation of percentage of 20-24 year old women who married before the age of 18
Posted by Bridgette-DHS on Fri, 04 Nov 2022 12:55:22 GMT
View Forum Message <> Reply to Message

Following is a response from DHS staff member, Tom Pullum:

Here is a Stata program that does what you want to do. It includes a translation of the CSPro instructions for coding age at marriage in this survey. I have included some lines (between "/*" and "*/") that give slightly different results. There were 2,487 women with a year of marriage but a missing month of marriage. Those lines assign month=6, which is often done when a month is missing, but the CSPro code treated those cases as NA. The Stata program does not give the rows for 20-49 and 25-49 but they could be added.

```
* Replication of Table 6.2, Age at First Marriage, in the India 2019-21 DHS (NFHS-5)

use "...IAIR7DFL.dta" , clear

gen never_married=0
replace never_married=1 if v501==0

/*
* Procedure that allocates cases with month missing to month 6 and uses the earlier of
*   the reported dates at first marriage and first cohabitation (if both are given)
gen     cmc_first_mar=s308c
replace cmc_first_mar=12*(s308y-1900)+6 if s308y<9998 & s308m==98
replace cmc_first_mar=. if s308y==9998

gen cmc_first_cohab=12*(v508-1900)+v507

gen afc=int((cmc_first_cohab-v011)/12)
gen afm=int((cmc_first_mar-v011)/12)

replace afm=min(afc,afm)
*/

* Procedure used in the CSPro construction of the table
gen v511x=.
replace v511x=s309 if s309>=0 & s309<=96
replace v511x=int((s308c-v011)/12) if s308c>=500 & s308c<=1500
replace afm=v511x
```

```
replace afm=99 if v501==0

* afm should be over-ridden as NA if v501==0

local lcutoffs 15 18 20 21 25
foreach lc of local lcutoffs {
gen     marr_by_`lc'=0
replace marr_by_`lc'=1 if afm<`lc'
}

tab v013 marr_by_15        [iweight=v005/1000000], row
tab v013 marr_by_18 if v013>1 [iweight=v005/1000000], row
tab v013 marr_by_20 if v013>1 [iweight=v005/1000000], row
tab v013 marr_by_21 if v013>2 [iweight=v005/1000000], row
tab v013 marr_by_25 if v013>2 [iweight=v005/1000000], row
tab v013 never_married      [iweight=v005/1000000], row
```

Subject: Re: Estimation of percentage of 20-24 year old women who married before the age of 18
Posted by cpmavelikara18@gmail.com on Sat, 05 Nov 2022 05:17:21 GMT
View Forum Message <> Reply to Message

How can a common person construct this table using information from DHS Guide to Statistics-7, without knowing this information about the coding for missing months? Where is this information in DHS Guide to Statistics-7?

The national and state level factsheets of India (NFHS-5) and Indian states contain 131 estimates. If this is the case with age at marriage, I suspect the case with all other estimates may be the same. Why should a person dedicate hours replicating these tables, when such crucial information is withheld?

So, could the DHS publish an accompanying do file/codes to replicate all these figures?

Subject: Re: Estimation of percentage of 20-24 year old women who married before the age of 18
Posted by Bridgette-DHS on Mon, 07 Nov 2022 17:00:23 GMT
View Forum Message <> Reply to Message

Following is a response from DHS staff member, Tom Pullum:

The tables in the reports and fact sheets were produced with a data processing package called

CSPro. It is very efficient for producing "camera-ready" tables, but it is not a statistical package such as Stata. It works off a hierarchical data file that is very different from the standard recode files that are available to users, including the researchers within DHS.

The Stata program I prepared implements the logic described in the Guide to DHS Statistics and will reproduce the numbers for any level of aggregation or covariate you want. There could be minor differences, probably mostly due to rounding error and minor modifications to the data files (now in version D).

The Stata output would have to be reformatted for presentation in tables. Stata includes commands such as "putexcel" and "putdoc" that can help you move results into Excel and Word. However, you don't need to go through that process if you just want to compare the numbers in the Stata output with the numbers in the fact sheets.

Subject: Re: Estimation of percentage of 20-24 year old women who married before the age of 18
Posted by cpmavelikara18@gmail.com on Tue, 08 Nov 2022 05:59:50 GMT
View Forum Message <> Reply to Message

Dear Bridgette/Tom Pullum,

First of all, my whole-hearted gratitude for your timely responses.

I understand the automated process of CSPro, which creates ready-made tables. However, most of the users of NFHS-5 (India DHS) who I know, use STATA. Therefore, we request the concerned team at DHS to publish just the STATA do files/codes for replicating the factsheet figures (excluding the putdocx command etc. to create tables). I use the word 'we', because I am writing this, on behalf of many of my friends in various departments of anthropology, sociology and other humanities stream at University of Hyderabad, who have faced difficulties, replicating these figures.

The factsheet contains only 131 indicators. I only require the codes only for these. The data for NFHS 5 and DHS Guide to Statistics are available. But we get stuck while preparing tables. In the previous messages, regarding the variable concerning child marriage, you will realize that it is because of certain adjustments in cmc (month, year) variables. This information is not available in DHS Guide to Statistics, which are just general instructions to generate the results.

My goal is to replicate the exact factsheet with extreme accuracy, taking into account all such minute details. I acknowledge that the revisions to data etc. may create deviations.

I insist on these do files for India and states in India, because in spite of being the largest DHS in the world in terms of sample size, it not widely used. One of the major reasons for this may be such data replication issues, which makes NFHS a black box for many scholars, requiring copious amount of time, to replicate even one figure.

The nodal agency in India, coordinating the survey, IIPS Mumbai, conducts only occasional data

workshops for NFHS. There are also no videos or helpful material, uploaded by IIPS, to replicate these figures. Some researchers, at IIPS, working in certain domains, possess the knowledge to replicate such figures. But they are inaccessible to students from other universities, due to their busy schedules.

So, in the interest of widening the use of NFHS data, I request the DHS team to disseminate the STATA do files, containing only the replication codes - the bare minimum (not tabulation codes) for all 131 indicators in the factsheets for India and Indian states.

Subject: Re: Estimation of percentage of 20-24 year old women who married before the age of 18
Posted by Varsha on Fri, 03 Mar 2023 11:43:12 GMT
View Forum Message <> Reply to Message

Hello,

I have a clarificatory question. When we are tabulating, we are also considering those women who are assigned ./99 in afm, right? The variables created through the loop assign these women a value of 0, so Stata will include them in the category of 0.

Subject: Re: Estimation of percentage of 20-24 year old women who married before the age of 18
Posted by Bridgette-DHS on Fri, 03 Mar 2023 13:34:07 GMT
View Forum Message <> Reply to Message

Following is a response from DHS staff member, Tom Pullum:

When calculating this kind of percentage, or the median age at marriage, you have to assign the never-married women to some numerical value, preferably one that is outside the age range of the data, such as 50 or 99. Otherwise, if they were treated as NA (age at marriage not applicable), the denominator for the calculation will be incorrect. These indicators are tricky to calculate in ever-married samples, because the women who have not (yet) married are not in the data.

Subject: Re: Estimation of percentage of 20-24 year old women who married before the age of 18

Posted by Varsha on Fri, 03 Mar 2023 19:20:37 GMT

Hello,

```
gen v511x = .
codebook s309, tab(9999)
replace v511x = s309 if s309>=0 & s309<=96
codebook s308c, tab(9999)
replace v511x = int((s308c-v011)/12) if s308c>=500 & s308c<=1500
replace v511x = 99 if v501==0

local lcutoffs 15 18 20 21 25
foreach lc of local lcutoffs {
gen    marr_by_`lc'=0
replace marr_by_`lc'=1 if v511x<`lc'
}
```

In the above codes, even if we do not assign the never-married women a value of 99 in place of ., they will anyway get a value of 0 when we run the loop. So how is the denominator going wrong? I couldn't get it.

Can you please explain it again?

Subject: Re: Estimation of percentage of 20-24 year old women who married before the age of 18
Posted by Bridgette-DHS on Mon, 06 Mar 2023 16:13:47 GMT

Following is a response from DHS staff member, Tom Pullum:

Below I will paste the full Stata code that I wrote to calculate table 6.2 in the NFHS-5 report. It borrows from the original CSPro code for this table. There are some small discrepancies from the published table (mostly only one-tenth of a percentage point) that I cannot account for.

In the procedure, afm (for age at first marriage) is set to 99 for women who are never-married. In your code you do this by setting v511x to 99. The point is that you cannot have "." or NA for afm for any women, because then you would be omitting unmarried women from the calculation. The never-married women have to be given a numerical value for afm that is greater than their current age. This is the only way you can calculate median age at marriage or the % married by an age that is less than current age. I don't know what you mean by "they will anyway get a value of 0 when we run the loop". Binary variables (such as "marr_by_18") are constructed, based on afm, but afm itself is NOT set to 0.

```
use "....IAIR7DFL.dta" , clear

gen never_married=0
replace never_married=1 if v501==0


* Procedure that allocates cases with month missing to month 6 and uses the earlier of
*   the reported dates at first marriage and first cohabitation (if both are given)
gen    cmc_first_mar=s308c
replace cmc_first_mar=12*(s308y-1900)+6 if s308y<9998 & s308m==98
replace cmc_first_mar=. if s308y==9998

gen cmc_first_cohab=12*(v508-1900)+v507

gen afc=int((cmc_first_cohab-v011)/12)
gen afm=int((cmc_first_mar-v011)/12)

replace afm=min(afc,afm)


* Procedure used in the CSPro construction of the table
gen v511x=.
replace v511x=s309 if s309>=0 & s309<=96
replace v511x=int((s308c-v011)/12) if s308c>=500 & s308c<=1500
replace afm=v511x

replace afm=99 if v501==0

* afm should be over-ridden as NA if v501==0

local lcutoffs 15 18 20 21 25
foreach lc of local lcutoffs {
gen    marr_by_`lc'=0
replace marr_by_`lc'=1 if afm<`lc'
}

tab v013 marr_by_15        [iweight=v005/1000000], row
tab v013 marr_by_18 if v013>1 [iweight=v005/1000000], row
tab v013 marr_by_20 if v013>1 [iweight=v005/1000000], row
tab v013 marr_by_21 if v013>2 [iweight=v005/1000000], row
tab v013 marr_by_25 if v013>2 [iweight=v005/1000000], row
tab v013 never_married     [iweight=v005/1000000], row
```

Subject: Re: Estimation of percentage of 20-24 year old women who married before the age of 18
Posted by Varsha on Wed, 08 Mar 2023 14:26:21 GMT
View Forum Message <> Reply to Message

Thank you, Tom.

Just one more clarificatory question. In the stata codes you have provided, you are replacing afm with v511x. So we can just replace v511x with 99 when v501==0 and get what we want. We do not have to generate afm and afc, right?

Subject: Re: Estimation of percentage of 20-24 year old women who married before the age of 18
Posted by Bridgette-DHS on Wed, 08 Mar 2023 18:40:14 GMT
View Forum Message <> Reply to Message

Following is a response from DHS staff member, Tom Pullum:

You may be able to simplify the code that way, but I recommend that you check to confirm that you get the same results.

Subject: Re: Estimation of percentage of 20-24 year old women who married before the age of 18
Posted by Varsha on Thu, 09 Mar 2023 09:08:29 GMT
View Forum Message <> Reply to Message

Sure, I'll do the check. Thank you.