

---

Subject: Teenage pregnancies by year 2015 to 2022

Posted by [Melyn](#) on Sat, 24 Feb 2024 13:30:26 GMT

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

---

Dear DHS team,

I am using the Kenya DHS 2022 Stata dataset (IR) and would like to analyse teenage pregnancies (age group 15-19) by year for the years 2015 to 2022.

For the years, is p2\_01 - p2\_20 the right variables to consider? For the pregnancies, the DHS online guide states the variables v201, v213 and v245 to be the ones relating to teenage pregnancy. Is that right?

I am specifically stuck on how I should relate, generate and analyse these variables in Stata to come up with meaningful descriptive summary statistics, line plots and even trend analyses to assess teenage pregnancies by year. I want to evaluate whether teenage pregnancies increased after 2020.

Please assist me with the Stata codes for my analyses.

Thanks in advance.

---

---

Subject: Re: Teenage pregnancies by year 2015 to 2022

Posted by [Melyn](#) on Thu, 07 Mar 2024 20:50:07 GMT

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

---

My goal is to analyse teenage pregnancies by year (2017 to 2022). This is what I have managed to do on Stata. My reshape command, even though it produces results, does not give the desired transformation of the dataset. This is where I am stuck and need help. I imagine a transformed dataset having 3 variables: Year (2017-2022), Total teenage pregnancies and Age-groups.

```
use KEIR8BFL.DTA, clear
```

```
*TEENAGERS: age-groups 15-19 and 20-24  
keep if v013==1 | v013==2
```

```
*YEARS FOR ANALYSIS: keep only years 2017 to 2022 for the variables relating to year of  
pregnancy outcome (p2_01 to p2_20)  
foreach var of varlist p2_01-p2_20 {
```

```
  replace `var' = 0 if `var' < 2017 | `var' > 2022  
}
```

```
*assess which variables have missing observations (zero values)  
foreach var of varlist p2_01-p2_20 {  
  tabulate `var', missing
```

```
}
```

```
*drop pregnancy outcome variables with missing observations  
drop p2_10-p2_20
```

```
*TEENAGE PREGNANCY VARIABLES: v201 "Total children ever born", v213 "Currently  
pregnant", v245 "Pregnancy losses"
```

```
*keeping only variables required for analysis of 2017 to 2022 trend analysis of teenage  
pregnancies for the age groups 15-19 and 20-24  
keep v201 v213 v245 v013 p2_01-p2_09
```

```
*Recode "v213" into "preg_status_numeric" by generating a new variable "preg_status_numeric"  
based on "v213" such that "no or unsure" takes the value "0" and "yes" takes "1"  
recode v213 (0=0) (1=1), generate(preg_status_numeric)
```

```
drop v213
```

```
*Rename variable to original name for ease of referencing  
rename preg_status_numeric v213
```

```
*Generate a variable that sums up teenage pregnancies for the 3 related variables  
gen Total_TeenagePreg=v201+v213+v245
```

```
*TRANSFORMING THE DATA for ease of analysis
```

```
*Sort the dataset by the age-group variable  
sort v013
```

```
*Creating a new identifier variable named "id"  
gen id = _n
```

```
*reshape the variables p2_01 through p2_09 from wide to long format, creating a new variable  
named outcome_year  
reshape long p2_, i(id) j(outcome_year)
```

---

Subject: Re: Teenage pregnancies by year 2015 to 2022  
Posted by [Janet-DHS](#) on Wed, 27 Mar 2024 15:54:20 GMT  
[View Forum Message](#) <> [Reply to Message](#)

Thank you for submitting your question. We are working on a response, but the erratum is still not posted.

---

Subject: Re: Teenage pregnancies by year 2015 to 2022  
Posted by [Janet-DHS](#) on Wed, 10 Jul 2024 18:14:54 GMT

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

I would propose modifications to what you are doing. First, I believe you want age at the time of the pregnancy, not at the time of the survey. Second, I would date the pregnancy by the estimated month when it began, p3-p20, rather than the month when it ended. Also I would refer to age 20-24 as "youth" rather than "teenage".

The following Stata lines will give what I believe you are looking for. The table includes ages below 15, but there is incomplete reporting for those ages because the sample is limited to age 15+ at the time of the survey. You can include more variables for your analysis, of course. Please let us know if you have questions.

\*DHS has just released a new version of the data files for the Kenya 2022 survey. There are no changes to the variables you are using, but I recommend that you switch to the new files.\*

```
use " KEIR8BFL.DTA", clear
```

```
* Reduce to the minimum variables needed; you can keep more  
* Standard DHS approach uses months but not days
```

```
keep v001 v002 v003 v005 v011 p3_* p20_*
```

```
* Remove 0's in subscripts  
rename *_0* *_*
```

```
* Calculate cmc of conception OR just use p3 if you want
```

```
forvalues li=1/20 {  
gen pconception_`li'=p3_`li'-p20_`li'  
}
```

```
drop p3_* p20_*  
reshape long pconception_, i(v001 v002 v003) j(pidx)  
rename *_*
```

```
* Calculate the woman's age at the conception  
gen age=int((pconception-v011)/12)  
keep if age<=24
```

```
* Calculate the calendar year of the conception  
gen year=1900+int((pconception-1)/12)  
keep if year>=2015
```

```
tab age year
```

```
tab age year [iweight=v005/1000000]
```