COUNTED & VISIBLE TOOLKIT

Using Stata to calculate "Proportion of women ages 18–49 who were married or in a union before age 18 (child marriage)" in Mongolia: A detailed guide

TUTORIAL GUIDE

ABOUT THIS GUIDE

- This guide serves as a complement to the video tutorial on calculating and assessing gender statistics using Stata, particularly, for those who wish to follow along with the video using the same data file.

- This will provide information on how to download the data file, description of variables utilized, and additional codes for further disaggregation, by type of location and wealth index.

- The video tutorial and guide are part of the Counted and Visible toolkit developed by Women Count, UN Women's gender data programme.

TO FOLLOW ALONG, YOU'LL NEED:

01

COMPUTER INSTALLED WITH STATA

- This guide is using a Windows-operated computer, but Stata can also be used in other operating systems, such as Mac and Linux.

- This guide is using Stata version 15, but the commands used are executable in either lower or higher versions.

WOMEN COUNT

TO FOLLOW ALONG, YOU'LL NEED:

02

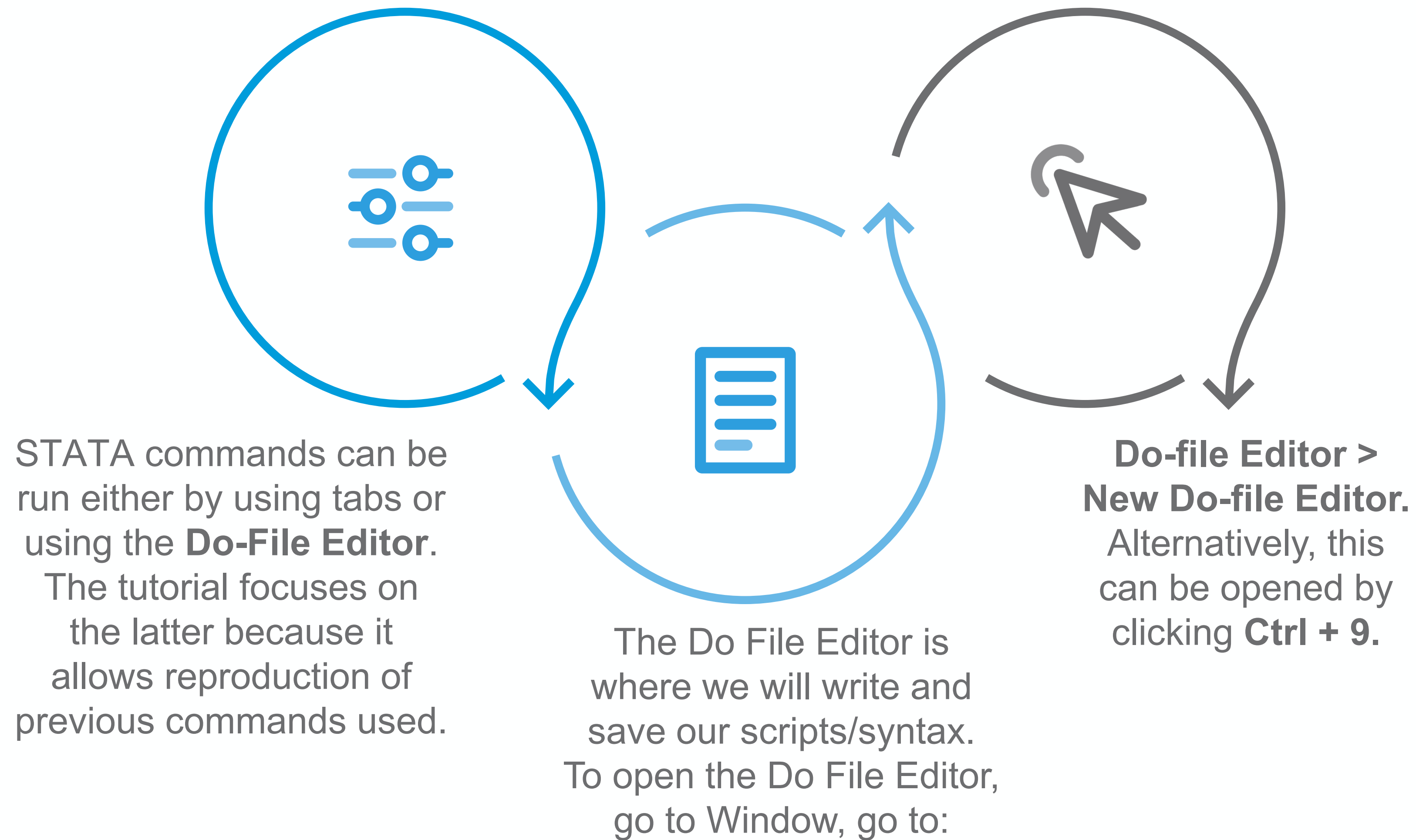### 📄 DATA FILE

- This guide makes use of Mongolia's Multiple Indicator Survey (MICS) 2018. MICS data sets can be downloaded for free at https://mics.unicef.org/surveys after registration.

- Upon downloading, you will find eight data files in the compressed zipped folder. For this exercise, we are specifically using the "Women in reproductive age (15–49 years of age)" data file, which is named "wm.sav".

- To facilitate this exercise, the "wm.sav" file was converted into "wm.dta", which is directly readable in STATA.

WOMEN
COUNT

STATA COMMANDS USED (SOURCE: STATA.COM)

Command	Description
use	Load Stata-format data set
keep if	Keep observations that satisfy specified condition
tabulate	Produce one-way table
generate	Create new variable
replace	Replace contents of existing variable
by	Repeat a command for each subset of the data
svyset	Declare survey design for data set
estat	Postestimation statistics
foreach	Executes the same command for each element of the list

WOMEN COUNT

DO-FILE EDITOR

STATA commands can be run either by using tabs or using the **Do-File Editor**. The tutorial focuses on the latter because it allows reproduction of previous commands used.

The Do File Editor is where we will write and save our scripts/syntax. To open the Do File Editor, go to Window, go to:

Do-file Editor > New Do-file Editor. Alternatively, this can be opened by clicking **Ctrl + 9.**

WOMEN
COUNT

STEP 1:
IMPORT YOUR DATA

```
use "file location\file name", clear
```

In the video tutorial, the "wm.dta" file is saved in the folder **"D:\Desktop\Toolkit".**

```
use "D:\Desktop\Toolkit\wm.dta", clear
```

VARIABLES USED FOR KEEPING RELEVANT CASES

- The "wm.dta" data file have 11,737 cases. This means that the data file consists of data from 11,737 women ages 15–49, irrespective of interview completion and marital status. However, we only need information for specific groups of women (i.e. women who completed the surveys, who are aged 18–49, who are currently or formerly married or in a union).

- The variables below will be used to limit the data files to the relevant cases:

Variable Name	Variable Label	Type	Values
WM17	Result of women's interviews	Categorical	1 – Completed 2 – Not at home 3 – Refused 5 – Incapacitated 6 – No adult consent 96 – Other
WB4	Age of woman	Numeric	15 – 49
MSTATUS	Marital/Union status of woman	Categorical	1 – Currently married/in union 2 – Formerly married/in union 3 – Never married/in union

WOMEN COUNT

```
keep if [conditions]
```

To keep only completed interviews:

```
keep if WM17 == 1
```

To keep only women ages 18–49:

```
keep if WB4 >= 18 & WB4 <=49
```

To keep only women who are currently or formerly married or in a union:
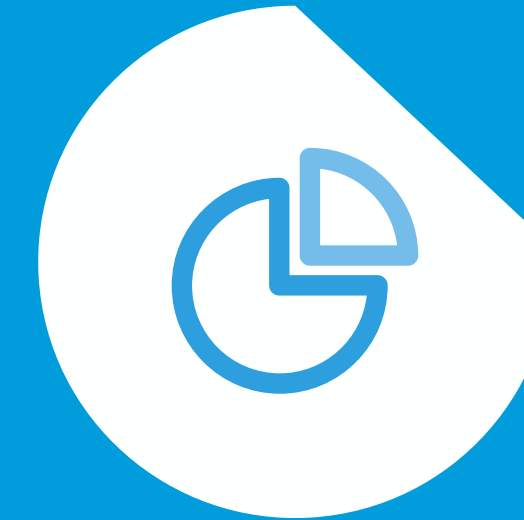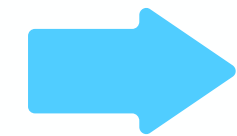
```
keep if MSTATUS==1 | MSTATUS==2
```

WOMEN
COUNT

STEP 3:
COMPUTE FOR VARIABLES OF INTEREST

INDICATOR

The indicator we will use is a variation of **SDG Indicator 5.3.1: Proportion of women aged 20–24 years who were married or in a union before age 15 and before age 18**

We will be calculating "Proportion of women ages 18–49 who were married or in a union before age 18 (child marriage)"

WOMEN
COUNT

VARIABLES GENERATING THE INDICATOR (1)

- The "WAGEM" variable will be used to generate a new variable – childm – a binary variable that identifies married women, whether or not they married as children (under 18 years old).

Variable Name	Variable Label	Type	Values
WAGEM	Age at first marriage/union of woman	Numeric	13–47
HH6	Area	Categorical	1 – Urban 2 – Rural
windex5	Wealth index quintile	Categorical	1 – Poorest 2 – Second 3 – Middle 4 – Fourth 5 – Richest

WOMEN
COUNT

VARIABLES GENERATING THE INDICATOR (2)

- The HH6 and windex5 variables are used as disaggregating variables to identify child marriage among:

women living in urban areas

women living in rural areas

poorest women

richest women

WOMEN COUNT

VARIABLES GENERATING THE INDICATOR (2)

- The combination of HH6 and windex5 variables as disaggregating variables will be used to identify child marriage among:

poorest women living in urban areas

poorest women living in rural areas

richest women living in urban areas

richest women living in rural areas

WOMEN COUNT

VARIABLES FOR APPLYING WEIGHTS

- As we are using a data file from a survey, we need to use weights to adjust for differences in probability of selection among cases, due to the survey design and other non-sampling instances.

- For this exercise, we are mainly using the following for weighing: wmweight, WM1, and HH7.

Variable Name	Variable Label	Weight type	Variable Name
wmweight	Woman's sample weight	Individual weights for women	wmweight
WM1	Cluster	Primary sampling unit	WM1
HH7	Region	Strata	HH7

```
generate [var] if [conditions]
tabulate [var] [weight], column
```

Generate new variable *childm* if age at first marriage is younger than 18

```
generate childm = 1 if WAGEM < 18
replace childm = 0 if childm != 1
```

```
label define yesno 1 "Yes" 0 "No"
label value childm yesno
```

Generate the frequency distribution of the variable childm

```
tabulate childm [iw=wmweight], column
```

```
tabulate [var1] [var2] [weight], column
by [var3] sort: tabulate [var1] [var2] [weight], column
```

Cross-tabulate *childm* by type of location

```
tabulate childm HH6 [iw=wmweight], column
```

Cross-tabulate *childm* by wealth index

```
tabulate childm windex5 [iw=wmweight], column
```

Cross-tabulate *childm* by wealth index and type of location

```
by HH6 sort: tabulate childm windex5 [iw=wmweight], column
```

WOMEN
COUNT

STEP 4:
REFORMAT VARIABLES FOR ANALYSIS

GENERATION OF BINARY VARIABLES

- STATA requires transformation to binary variables (coding 1 and 0) for post-estimation analysis. This will ensure that they can be individually assessed for precision and reliability.

- The variables needed to be generated are:

Variables	Definition
childm_urban	Child marriage among women living in urban areas
childm_rural	Child marriage among women living in rural areas
childm_poorest	Child marriage among poorest women
childm_richest	Child marriage among richest women
childm_poorest_urban	Child marriage among poorest women living in urban areas
childm_richest_urban	Child marriage among richest women living in urban areas
childm_poorest_rural	Child marriage among poorest women living in rural areas
childm_richest_rural	Child marriage among richest women living in rural areas

WOMEN
COUNT

```
generate urban = 1 if HH6 == 1
generate rural = 1 if HH6 == 2
generate poorest = 1 if windex5 == 1
generate richest = 1 if windex5 == 5
generate childm_urban = childm*urban
generate childm_rural = childm*rural
generate childm_poorest = childm*poorest
generate childm_richest = childm*richest
generate childm_poorest_urban = childm*poorest*urban
generate childm_richest_urban = childm*richest*urban
generate childm_poorest_rural = childm*poorest*rural
generate childm_richest_rural = childm*richest*rural

foreach v of varlist childm_* {
        label value `v' yesno
        }
```

WOMEN
COUNT

STEP 5:
GENERATE STANDARD ERROR AND COEFFICIENT OF VARIATION

STANDARD ERROR AND COEFFICIENT OF VARIATION

	Definition	**Computation**
Standard error (SE)	measure of precision or closeness of the estimates to each other	computed by dividing the square root of the ratio of the variance of the estimates and the number of observations used in the estimation
Coefficient of variation (CV)	measure of reliability or variability of the value of estimates	computed as a ratio of the standard error of the estimate and value of the estimate

WOMEN COUNT

STANDARD ERROR AND COEFFICIENT OF VARIATION

• In principle, preferred estimates are those with a low value of SE.

• For CV, it is important to note that there are no internationally agreed standards or recommendations as to the "acceptable" CV values for a certain type of estimator. In practice, CV thresholds vary country-to-country and in some cases survey-to-survey. For the purpose of the Toolkit, we are suggesting the classification in the table on the right.

VALUE OF CV	SUGGESTED CLASSIFICATION
CV less than 10%	Highly reliable
Between 10% and 20%	Sufficiently reliable
Between 20% and 33%	Still acceptable but should be used with caution
Greater than 33%	Caveats should be provided in terms of the level of reliability of the estimate

WOMEN COUNT

```
svyset sampling_unit [weight], str(strata)
svy: proportion variable
estat cv
```

Generate the SE and CV for *childm*

```
svyset WM1 [weight=wmweight], str(HH7)
svy: proportion childm
estat cv
```

Generate the SE and CV for other variables

```
foreach v of varlist childm_urban-childm_richest_rural {
    svy: proportion
    estat cv
    }
```

SUMMARY OF RESULTS

Indicators	Variable Name	Estimates	No. Of observations	Standard error	Coefficient of variation
1. Proportion of child marriage among women aged 18–49	childm	9.38	846	0.0053	5.61
2. Proportion of child marriage among poorest women aged 18–49	childm_poorest	11.42	280	0.0093	8.14
3. Proportion of child marriage among richest women aged 18–49	childm_richest	4.91	45	0.0096	19.51
4. Proportion of child marriage among women residing in urban areas aged 18–49	childm_urban	9.06	388	0.0070	7.78
5. Proportion of child marriage among women residing in rural areas aged 18–49	childm_rural	10.06	399	0.0067	6.69
6. Proportion of child marriage among poorest women residing in urban areas aged 18–49	childm_poorest_urban	11.41	14	0.0346	30.36
7. Proportion of child marriage among richest women residing in urban areas aged 18–49	childm_richest_urban	4.92	45	0.0096	19.51
8. Proportion of child marriage among poorest women residing in rural areas aged 18–49	childm_poorest_rural	11.42	266	0.0095	8.34
9. Proportion of child marriage among richest women residing in rural areas aged 18–49	childm_richest_rural	-	0	-	-

WOMEN COUNT

END

See the Counted and Visible Toolkit:
https://data.unwomen.org/resources/counted-and-visible-toolkit
Please visit the Women Count Data Hub:
https://data.unwomen.org/