Basic Data Analysis Course on Stata
Summarization of data in Stata

Md Monowar Hossain

Institute of Statistical Research and Training (ISRT)
University of Dhaka
mmhossain1@isrt.ac.bd

January 03, 2025

About the data

- We are going to use the national-scale dataset wm.dta from Bangladesh Multiple Indicator Cluster Survey (MICS) 2019.
- The data were collected by the Bangladesh Bureau of Statistics in cooperation with the UNICEF Bangladesh, as part of the global MICS programme.
- The survey employed a two-stage stratified cluster sampling approach where the 64 districts were the sampling strata.
- The 2011 national census enumeration areas (EAs) were defined as the primary sampling units (clusters) and households as the secondary sampling units.
- In the first stage of sampling, 3220 EAs were selected with probability proportional to size (PPS) method.
- In the second stage, a systematic sample of 20 households was obtained from each of the selected EAs, which led to a final sample of 64,400 households for the survey.

One-way table for qualitative data

```
use wm.dta, clear

tab HH7

tab HH7 [iweight=wmweight]
/* iweight means 'importance weight' */
```

One-way table for quantitative data

```
recode WB4 (15/20 = 1 "15-20") (20/25 = 2 "20-25") ///
(25/30 = 3 "25-30") (30/35 = 4 "30-35") ///
(35/40=5 "35-40") (40/45=6 "40-45") ///
(45/50=7 "45-50"), generate(agegroups) label(agegrp)

tab agegroups [iweight=wmweight]
```

Two-way table for qualitative data

```
tab HH6 welevel [iweight=wmweight]

tab HH6 welevel [iweight=wmweight], row

tab HH6 welevel [iweight=wmweight], col

tab HH6 welevel [iweight=wmweight], cell
```

Descriptive statistics

```
summarize WB4

summarize WB4, detail

summarize WB4 [aweight=wmweight]

summarize WB4 [aweight=wmweight], detail
```

Descriptive statistics

```
by HH6, sort: summarize WB4, detail

by HH6, sort: summarize WB4 [aweight=wmweight], detail

codebook HH6
codebook HH7

keep if HH6==1
summarize WB4

keep if HH6==1 & HH7==10
summarize WB4
```

Declare survey design for dataset: svyset

svyset manages the survey analysis settings of a dataset. You use svyset to designate variables that contain information about the survey design, such as the sampling units and weights.

```
svyset [pw=wmweight], psu(WM1) strata(HH7A)
/* WM1 is the cluster number
HH7A is the district */

tab HH7 [iweight=wmweight]
svy: tab HH7
/* both the commands give the same result */

svy: tab HH6 welevel, row pearson
/* pearson chi-square */
```

Multiple response analysis

```
use multipleresponse.dta, clear

ssc install mrtab

mrtab OTHER2_1 OTHER2_2 OTHER2_3 OTHER2_4 OTHER2_5 ///
OTHER2_6 OTHER2_7 OTHER2_8 OTHER2_9 OTHER2_10 ///
OTHER2_11 OTHER2_12 OTHER2_13 OTHER2_14 OTHER2_15 ///
OTHER2_16 OTHER2_17 OTHER2_18 OTHER2_19 OTHER2_20 ///
OTHER2_21 OTHER2_22 OTHER2_23 OTHER2_24 OTHER2_25 ///
OTHER2_26 OTHER2_27
```

Exporting results

- Copy table to a excel file
- Copy table to a word file
- Copy figure to a word file
- Create a pdf file of output

THANK YOU