Using Microsoft **Excel** to improve efficiency in working with large datasets in Stata

by:

Ahmad Khanijahani, Ph.D., CPH, CHDA

DUQUESNE UNIVERSITY

STata

Using Excel with Stata can save your time

- Big data and increased **number of variables** in datasets

- Hard to hold a **big picture** of the all variables available in datasets

- **Don't** want to **drop** (remove) the variables that might be useful later

- **Repeated tasks** (commands) can be cumbersome and time-consuming

- Challenging to merge and repeat the same/similar commands on **new versions of data. Cross-walking**

- and …

We can use excel in different ways

1- Interactive data dictionary

- Provided **originally in Excel** format by the data owner/provider

- Export from Stata: Export data as **xlsx**, Drop all observations and keep label

- Using Stata Do file: Copy do file into Excel and use **text to column** feature and some formulas

2- Batch commands in Stata: repeated or patterned commands

- Create a table or matrix in Excel and customize commands by **joining multiple columns**

- 3- Customizing Stata outputs: Different journal submission requirements, etc.

- Especially, if you have received the Stata output from someone else

Some useful Excel Formulas, Functions, or Features

Filter: Filtering

Sort: Sorting

FIND and **REPLACE**: new commands with new variables

&: Joining Columns

INDEX and **MATCH**: matching variables from different tables

TEXTJOIN: joining different cells with a delimiter of your choice

LEN, **RIGHT**, **LEFT**: length of the variable label, etc.

CLEAN, **TRIM**, **UPPER**, **LOWER**, **PROPPER**: cleaning extra spaces, switching between lower and upper case

ROUND, **ROUNDDOWN**, **ROUNDUP**, **MROUND**: working with Stata output

Let's work on a few examples

Some tips:

- Don't forget to have a **reference column** to keep the original order

- Don't leave empty columns between the columns of the table

- Don't leave a column without a header

- Always **Copy Excel Sheets** and Keep the original sheet intact

- Keep **backups** of your Stata file

- Run a few **pilot commands** to test the accuracy of the batch commends before running long list of commands

- If working with unreliable internet connection, **Copy your Stata file to local disk** and then work on it

Thank You!