# Stata Camp

## Day 1: Familiarize with Stata and Data preparation

**Agenda:**

Understand the layout of the interface and major functional areas

Learn how to open and save data files

Learn how to run commands and programs

Learn how to sort out data and construct panel data

Learn how to import data files in different formats (e.g., Excel, CSV)

Learn how to browse and describe datasets

Learn how to create and modify variables

1. **Why we choose to use Stata?**
2. **Stata Interface and Basic Operations**

Four main working zone: History, Command, Variables and Properties

1. **Know about variables**

Linear Regression Equation:

Y = β₀ + β₁X₁ + β₂X₂ + β₃X₃ + ε

Where:

Y is the dependent variable.

X₁ is the core independent variable.

X₂ and X₃ are control variables. (5 or more will be better)

β₀, β₁, β₂, β₃ are regression coefficients.

ε is the error term.

1. **How to obtain decent data?**

Official Public Reports/website:

IMF: <https://www.imf.org/external/datamapper/datasets>

OECD: <https://stats.oecd.org/>

World bank: <https://data.worldbank.org/>

National Bureau of Statistics data

Other government website

Credible Databases:

Datastream; Bloomberg; Winds/ CSMAR;Refinitiv Workspace

**Exercise Caution with unknown Sources**

1. **Sort out data**

Use excel to pre-processing data

understand the structure of panel data

Open the excel file “invest”

**Quiz:** sort out the variable “mvalue” and “kstock” into a panel data format

Import and merge data into the Stata:

**Command: merge**

--------------------------------------------------------------

merge 1:1 xx xx using “xxxx”

merge m:m xx using “xxxx”

merge 1:m

--------------------------------------------------------------

**Quiz:** merge the variables “mvalue” , “kstock” and “region” into stata

Replace the company name with non-string varibales

--------------------------------------------------------------

egen id = group( company )

help (xxx)

webuse grunfeld,clear

**Step1:** set panel data:

--------------------------------------------------------------

xtset company year

--------------------------------------------------------------

**Step2:** plot the relationship between variables

--------------------------------------------------------------

scatter invest mvalue, title("Scatter plot") xtitle("mvalue") ytitle("invest") scheme(s2manual) msymbol(Oh) msize(small)

aaplot invest mvalue, title("Scatter plot") xtitle("mvalue") ytitle("invest") scheme(s2manual) msymbol(Oh) msize(small)

tabstat invest mvalue kstock, s(N mean p50 sd min max skew kurt) f(%12.3f) c(s)

**logout,save(descriptive\_statistic)word replace:** tabstat invest mvalue kstock, s(N mean p50 sd min max skew kurt) f(%12.3f) c(s)

pwcor invest mvalue kstock

--------------------------------------------------------------

**Step3:** Choose suitable regression model

For linear regression, we have: OLS, fixed effect model and random effect model

--------------------------------------------------------------

reg invest mvalue kstock, r

est store ols\_result

vif

xtreg invest mvalue kstock,fe

xtreg invest mvalue kstock i.year,fe

xtreg invest mvalue kstock,re

xttest0

--------------------------------------------------------------

Choose between models:

--------------------------------------------------------------

xtreg invest mvalue kstock,fe

est store fe\_result

xtreg invest mvalue kstock,re

est store re\_result

hausman fe\_result re\_result

esttab ols\_result fe\_result re\_result using regression\_result.rtf, replace b(%12.3f) se(%12.3f) nogap compress s(N r2 r2\_a)star(\* 0.1 \*\* 0.05 \*\*\* 0.01)

--------------------------------------------------------------