

# Tutorial

## A. Import dataset

- The following is a brief tutorial using example data sets already installed in STATA
  - auto.dta is used for commands 1.1.1-1.2.2 & 4.1.1-4.3.1.3
  - uslifeexp.dta is used for commands 1.2.3-3.4.1

## 1. Basic commands

```
1 *Import data set
2 sysuse auto.dta
```

### 1.1.1 List variables

```
1 *List 1 variable
2 list price
3
4 *List 2 variables
5 list price mpg
```

### 1.1.2 List variables: by condition

```
1 *List only first 10 observations of variables
2 list price mpg in 1/10
3
4 *List variable by row by condition
5 list price mpg if make == "Subaru"
```

### 1.1.3 Rename variables

```
1 rename foreign location
```

### 1.1.4 Label variables

```
1 label variable location "Location of car production"
```

## 1.2 Basic statistics

### 1.2.1 Summary statistics

```
1 summarize price mpg
```

#### 1.2.1.1 Summary: by condition

```
1 summarize price mpg if length>200
```

### 1.2.2 Correlation matrix

```
1 correlate mpg price weight length
```

### 1.2.3 Cross tabulation

Note: 1.2.3-3.4.1 uses uslifeexp.dta

```
1 sysuse uslifeexp.dta
2 tabstat le_male le_female le_w, by(year)
```

## 2. Manipulating data

### 2.1 Basics

#### 2.1.1 Generating logs

```
1 gen ln_le_male = ln(le_male)
```

#### 2.1.2 First difference

```
1 gen fd_le_male = D.le_male
```

##### 2.1.2.1 First difference (Alternative method)

```
1 gen fd_le_male= le_male[_n]-le_male[_n-1]
```

#### 2.1.2.2 First difference by country (row condition)

```
1 *By country (row variable)
2 by country: gen fd_x1= x1[_n]-x1[_n-1]
```

#### 2.1.2.3 First difference logs

```
1 by country: gen x1_ln_diff = ln(x1[_n])-ln(x1[_n-1])
```

#### 2.1.2.4 First difference absolute value

```
1 gen x1_le_male = abs(le_male[_n]-le_male[_n-1])
```

## 2.2 Scale variables

#### 2.2.1 Scale variable by billions

```
1 gen divideby10_le_male = le_male/10
```

#### 2.2.2 Scale variables by other variables

```
1 gen male_female_ratio = le_male/le_female
```

## 2.3 Moving average

#### 2.3.1 Moving average 3-period

```
1 *Not an example shown with uploaded data set
2 *Use with panel data
3 rangestat (mean) yr_avg3 = x1, interval(time -1 1)
```

#### 2.3.2 Moving average 3-period by country

```
1 *Not an example shown with uploaded data set
2 *Use with panel data
3 rangestat (mean) yr_avg3 = x1, interval(time -1 1) by(country)
```

#### 2.3.3 Moving average by time

```
1 *Not an example shown with uploaded data set
2 *Use with panel data
3 bysort time: egen x1_yearly_avg =mean(x1)
```

## 2.4 Generate dummies

### 2.4.1 Dummy by country (row conditon)

```
1 gen long_life = 0
2 replace long_life = 1 if le_male>70
```

### 2.4.2 Dummy by time (By time period)

```
1 gen stock_crash = 0
2 replace stock_crash = 1 if year==1987
3 replace stock_crash = 1 if year==1997
```

### 2.4.2 Dummy by time (After time period)

```
1 gen after_eighty = 0
2 replace after_eighty =1 if year>1979
```

## 2.5 Creating lags

```
1 *One lag
2 gen lag_le = L.le
3 *Two lags
4 gen lag_le_2 = L2.le
```

## 3. Graphing

### 3.1 Scatter plot

```
1 twoway scatter le_male le_female
```

#### 3.1.1 Scatter plot with country labels

```
1 twoway scatter le_male le_female, mlabel(year)
```

### 3.1.2 Scatter plot with regression line

```
1 twoway lfit le_male le_female || scatter le_male le_female
```

## 3.2 Bar chart stacked

```
1 twoway (tsline le_b, recast(bar)) (tsline le , recast(bar)), ytitle(y title) title(Cool  
title)
```

### 3.3 Line chart

```
1 twoway (line le year)
```

#### 3.3.1 Line chart with two y-axis

```
1 twoway (line le year, lwidth(thick)) (line le_male year, yaxis(2))
```

## 3.4 Labeling & Saving

### 3.4.1 Saving graph

```
1 local directory : pwd  
2 local direct_graphs "`directory'/Graphs/"  
3 twoway scatter le le_male  
4 graph save Graph "`direct_graphs'_scatter", replace
```

## 4. Regression

### 4.1.1 Basic regression

```
1 sysuse auto.dta  
2 reg price mpg
```

#### 4.1.2 Regression while controlling for heteroskedasticity

```
1 reg price mpg, robust
```

### 4.1.3 Regression by country

```
1 reg price mpg if length>200, robust
```

### 4.1.4 Regression with many variables

```
1 local variable_list "price mpg weight length"  
2 reg foreign `variable_list', robust
```

## 5. Regression diagnostics

```
1 //Reject and heteroskedasticity  
2 xttest3
```

### 4.3.1.3 Serial correlation

```
1 //Reject and serial correlation  
2 xtserial price mpg
```

### 4.3.1.3 Durbin-watson

```
1 reg price mpg  
2 estat dwatson
```

## 5. Advanced commands

### 5.1 Save variable list

```
1 local variable_list "mpg weight length"  
2 summarize `variable_list'
```

### 5.2 For loop

```
1 local mcode 0  
2 local month jan feb mar apr  
3 foreach m of local month {
```

```
4 local mcode = `mcode' + 1
5 display "`mcode': `m'"}
```

## Appendix

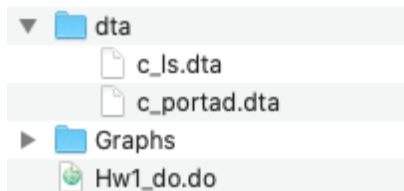
### A.1 Set local directory

Important: Sets the location where you can import from and export files to

```
1 local directory : pwd
2 mkdir Graphs
3 local direct_graphs "`directory'/Graphs/"
```

### A.2 Structure your dataset and folders for reproduction

- Many students turned in Stata code from HW1 as a word document.
- For future assignments please turn in code using a do file
- Ideally you should turn in a zipped folder so I can reproduce the assignment easier
- The following screenshot is one suggestion on how you can structure your code and dta files
- A.2.1 demonstrates how these folders can be created using the command line in Stata



#### A.2.1 Creating folders from the command line

```
1 * Created if you launch stata from desktop
2 local desktop_directory : pwd
3 *Create hw folder
4 mkdir hw1
5 *Change directory to hw folder
6 cd "`desktop_directory'\hw1"
7 * Set hw folder path as local variable
8 local hw_fol "`desktop_directory'\hw1"
9 *Create dta and Graphs folder
10 mkdir dta
11 mkdir Graphs
12 *set dta and graphs path as local variable
13 local dta_fol "`hw_fol'\dta\"
14 local graphs_fol "`hw_fol'\Graphs\"
```

```
15 *Return the current working directory to hw1
16 cd ..
```