```
/*******************************************************************************************************
Title: STATA Tutorial 1
*** Metrics
*** Purpose: Stata setup, basic data manipulation, Mean comparision test, Basic OLS regressions ***
*******************************************************************************************************
/

/* This is how you do a comment in Stata. Open with /* and close with */
You can also use just a star at the beginning of  the line, then the whole line will be ingored.
The advantage of /* */ is that you can put it inside the line that contains the command*/
        *forward slash

* E.g.: Comment
// E.g.: Comment

******************
****General Setup**
******************

*notice the colors

clear  /*This clears memory of all data*/

set mem 100m      /* Setting memory size - depending on your dataset you may need more or less*/

set matsize 500   /* Setting  number of RHS variables in a model */

/*set more on*/  /*This sets the 'more' at the bottom of the page on. I.e. if you have a model
with lots of output, it pauses the execution after it reaches pagesize limit until you press any
key for it to continue.*/

set more off  /*Turn the above thing off if you want it to do stuff continuously. Good if you're
leaving the program to run for hours*/

cap log close /* close the log file if any already open*/

pwd  /*shows us what directory Stata treats as working directory at the moment*/

     /* note to tutors: get students to place the datasets NHIS2009_clean in the working
directory at this point. */

// set to directory where NHIS2009_clean.dta is stored
cd "D:/Dropbox/lectures/ECO3211_2021Fall/STATA/Stata 1" /*cd - Change directory*/

* Sometimes, you want to keep a record of your STATA output. You can create a log file

log using tutorial1.log, text replace
/* log using - creates a new log file named tutorial1 in ?where */
/* log close - all STATA output will be saved */
/* I personaly like clicking */

*******************************
***Load and manipulate data*****
*******************************

use NHIS2009_tutorial, clear   /*Load data*/
                               /* I personaly like clicking*/

preserve    /* preserve the current state of your dataset, can be recalled with 'restore'*/

rename sex gender  /* rename sex as gender */

drop marstat       /* drop a var called marstat in this dataset */

keep if _n<=50     /* keep top 50 observations */

restore     /*back to the state before 'preserve' command */

```

```
*actively using help xx


**************************
**A first look at the data**
**************************
describe              /* description of data*/
** d                I personaly like cliking variables manager above output window

summarize             /* summary statistics */
**su
*see the difference in results

browse

browse in 20/30 /* open in a new window */
* click

list sex famsize in 20/30
/*open in the output window*/
/*list countable variables and display countable obs in a given range */

/*right click to change font size*/

label variable age2 "Age square" /*Example of labelling a variable*/
*see where the change is


***********************************
**Mean comparision test using ttest**
***********************************

* mean comparisons for husbands: use health index (hlth) as an example

keep if sex==1 /*keep only males */
* pay attention to ==, not =

* (9,395 observations deleted). be careful about the change in obs. When you are about to close down the dataset, don't save it.

sum hlth if hi==1 /*summarize health index for insured males. */

/*generate a new variable */
gen m0_hlth =r(mean) /*generate a variable = mean value of hlth in insured male group*/
gen sd0_hlth =r(sd) /*generate a variable = standard error of hlth in insured male group*/
* here only one =

sum hlth if hi==0 /*summarize health index for uninsured males*/
gen m1_hlth =r(mean) /*generate a variable = mean value of hlth in insured male group*/
gen sd1_hlth =r(sd) /*generate a variable = standard error of hlth in insured male group*/

ttest hlth, by(hi) /*two-sample t test using groups (divided by hinsurance) */

sum hlth [aw=perweight ] if hi==1 /*add basic annual weight: take the frequency a person entered into the panel surveys into account, perweight is usually given when data released*/

sum hlth [aw=perweight ] if hi==0


**************************
**Basic OLS regressions**
**************************
use NHIS2009_tutorial, clear  /*reload the full sample data*/

* use -regress- to do balancing test
reg yedu hi if sex==1, robust  /*OLS regression: use robust se, restrict to males only*/

outreg2 using Table1, excel replace /*generate an excel table which contains the regression
```

```
results*/
/*outreg2 using table2.xls, excelname replace 苹果电脑的话
*可能要install */


* use -regress- to do balancing test with weights
reg yedu hi[ w=perweight ] if sex==1, ro  /*add basic annual weight*/
outreg2 using Table1, excel

* add control variable
reg hlth hi sex inc age famsize [ w=perweight ], ro
outreg2 using Table1, excel
* using the same excel three times, append by order


* control for educational attainment
ta educ

/*               Educational attainment |      Freq.     Percent        Cum.
------------------------------------------+-----------------------------------
       Never attended/kindergarten only |         51        0.27        0.27
                                Grade 1 |         25        0.13        0.40
                                Grade 2 |         39        0.21        0.61
                                Grade 3 |         57        0.30        0.92
                                Grade 4 |         68        0.36        1.28
                                Grade 5 |         79        0.42        1.70
                                Grade 6 |        398        2.12        3.82
                                Grade 7 |         94        0.50        4.32
                                Grade 8 |        208        1.11        5.42
                                Grade 9 |        400        2.13        7.55
                               Grade 10 |        289        1.54        9.09
                               Grade 11 |        355        1.89       10.98
                 12th grade, no diploma |        324        1.72       12.70
                   High school graduate |      4,196       22.33       35.03
                      GED or equivalent |        394        2.10       37.13
                  Some college, no degree |      3,115       16.58       53.71
AA degree: technical/vocational/occupat |      1,399        7.45       61.15
             AA degree: academic program |        717        3.82       64.97
    Bachelor's degree (BA, AB, BS, BBA) |      4,269       22.72       87.69
Master's degree (MA, MS, Meng, Med, MBA |      1,732        9.22       96.91
 Professional degree (MD, DDS, DVM, JD) |        283        1.51       98.41
             Doctoral degree (PhD, EdD) |        298        1.59      100.00
------------------------------------------+-----------------------------------
                                  Total |     18,790      100.00
*/

ta empstat
/*
 Employment status in past 1 to 2 weeks |      Freq.     Percent        Cum.
------------------------------------------+-----------------------------------
        Working for pay at job/business |     14,685       78.15       78.15
    Working, w/out pay, at job/business |        106        0.56       78.72
               With job, but not at work |        678        3.61       82.33
                             Unemployed |        647        3.44       85.77
                 Not in labor force |      2,674       14.23      100.00
------------------------------------------+-----------------------------------
                                  Total |     18,790      100.00
*/

*xi:reg and reg, here, same results. When equations include more dummies, xi:reg is recommended.
xi:reg hlth hi sex inc age famsize i.educ [ w=perweight ], ro
outreg2 using Table1, excel

*i.educ generates dummy variables
*list i.educ in 1/10

cap log close    //results of all your commands will be saved in the log file named tutorial1 in
your working directory
```

```
196
197   * change color scheme in do-file editor, wx - "Stata 中 dofile 编辑器的配置 —— 来个漂亮的编辑器"
198
199
```