Reproducible Research With Stata

E. F. Haghish

IMADA, University of Southern Denmark

This document summarizes the examples of the workshop. I have organized the examples in separate files and render them in a single document to provide an example of how a large document can be produced dynamically by calling smaller chunks of code.

*haghish@imada.sdu.de*

# Workshop Examples

This document includes some of the examples of the workshop. I have organized them in separate do files and the dynamic document will include all of the examples.

## Help

If you need help about working with **MarkDoc** package, probably the best place to begin is the [MarkDoc GitHub Wiki](https://github.com/haghish/MarkDoc/wiki), which is the package manual. If you have questions, post them on and I get back to you shortly.

Hellow World  
   
 . display "hello world"

# Header

sthing will appear

## Header 2

### Header 3

#### Header 4

##### Header 5

# Part 2

This is a text paragraph. Text can be *italic* or *italic*. Also, it can be **Bold** or **Bold**.

“the text paragraph can be indented for - typically -quoting”

* This is item 1
* Item 2
* item 3
* This is another item
* quite file I hope
  + sublist item 1
  + sublist 2
    - subsublist 1
    - item
    1. sth
    2. sth
  + continue with sublist
* continue with the list

# Numbered list

1. numbered item 1
2. item 2
   1. subitem 1
   2. subitem 2
      1. subsub item
   3. sth
3. back to item

Using the “—” sign will add a line in the document

## Breaking the line

Add 2 or more spaces at the end of the line to break it without starting  
a new  
paragraph

## Adding links

This is a link to MarkDoc’s GitHub Manual <https://github.com/haghish/MarkDoc/wiki>

You can also change the **hypertext** e.g. [MarDoc’s Manual](https://github.com/haghish/MarkDoc/wiki)

. sysuse auto, clear //load auto dataset   
 (1978 Automobile Data)  
   
 . summarize   
   
 Variable | Obs Mean Std. Dev. Min Max  
 -------------+---------------------------------------------------------  
 make | 0  
 price | 74 6165.257 2949.496 3291 15906  
 mpg | 74 21.2973 5.785503 12 41  
 rep78 | 69 3.405797 .9899323 1 5  
 headroom | 74 2.993243 .8459948 1.5 5  
 -------------+---------------------------------------------------------  
 trunk | 74 13.75676 4.277404 5 23  
 weight | 74 3019.459 777.1936 1760 4840  
 length | 74 187.9324 22.26634 142 233  
 turn | 74 39.64865 4.399354 31 51  
 displacement | 74 197.2973 91.83722 79 425  
 -------------+---------------------------------------------------------  
 gear\_ratio | 74 3.014865 .4562871 2.19 3.89  
 foreign | 74 .2972973 .4601885 0 1

. histogram price  
 (bin=8, start=3291, width=1576.875)  
   
   
 . graph export price\_hist.png, replace width(300)   
 (file price\_hist.png written in PNG format)

## Including a figure

Next, I include the histogram of the Price variable using the histogram command and the price variable.

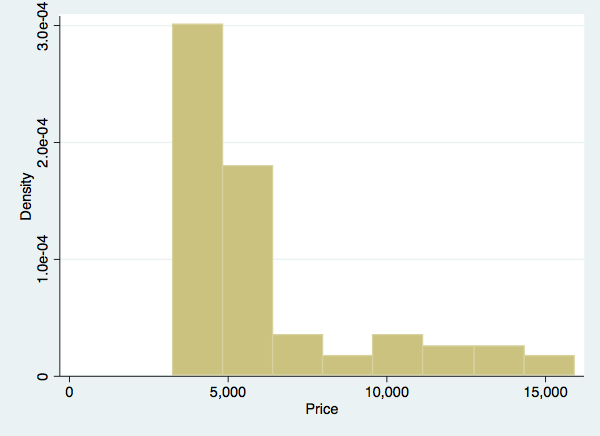


Figure 1. Histogram of the price variable

. histogram mpg  
 (bin=8, start=12, width=3.625)

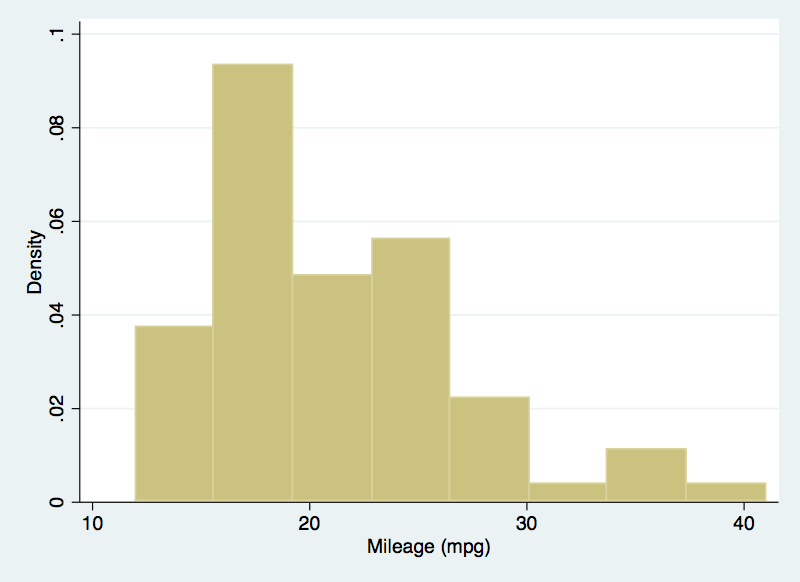


Figure 2. Histogram of the mpg variable

## Writing dynamic text

use the txt command to interpret values of macros and scalars:

. summarize price  
   
 Variable | Obs Mean Std. Dev. Min Max  
 -------------+---------------------------------------------------------  
 price | 74 6165.257 2949.496 3291 15906  
   
 . return list  
   
 scalars:  
 r(N) = 74  
 r(sum\_w) = 74  
 r(mean) = 6165.256756756757  
 r(Var) = 8699525.974268788  
 r(sd) = 2949.495884768919  
 r(min) = 3291  
 r(max) = 15906  
 r(sum) = 456229

The number of subject is 74 subjects in the darta set.

## creating table

|  |  |
| --- | --- |
| **Optopns** | **Discription** |
| markup(str) | specifies the markup language that is used for documentation |
| title(str) | displays the table description |
| width(int) | specifies the width of the table in HTML and LaTeX |
| height(int) | specifies the height of the table in HTML and LaTeX |
| center | aligns the table to the center of the document in HTML and LaTeX |
| left | aligns the table to the left of the document |

–

|  |  |  |
| --- | --- | --- |
| size | material | color |
| 9 | leather | brown |
| 10 | hemp canvas | natural |
| 11 | glass tran | sparent |

Mathematical notations ======================+

You can use markdoc to:

1. write inline notations (i.e. in the text paragraphs)
2. on a separate line

For example, $$ will appear inline whereas

willbe on another line.

# Writing mathematical notations

The text paragraph can include mathematical notations. For example, this formula will be displayed within the text paragraph, whereas this next formula will be placed on a separate line:

Use a single “$” sign for writing inline mathematical notations. For example, would be rendered inline with the text paragraph. Use double dollar signs “$$” for placing the notations on a separate lines:

1. Since the notations appear in comments, they will not be interpreted by Stata as global macros.
2. Place a backslash before the “$” if you are using them in the document, but not for rendering mathematical notations. The backslash will not appear in the dynamic document.
3. You can also write dynamic mathematical notations using the **txt** command.

* . local a = 10

Note that when you write inline mathematical notations, there should be **NO SPACE** between the dollar sign and the notation. However, if you are placing your notations on a separate line, there should be no problem.

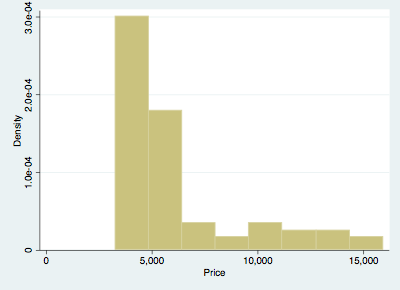
## Data preparation

1. clear the data from Stata
2. laod the auto dataset

* . clear  
    
   . sysuse auto  
   (1978 Automobile Data)

## Summarize the data

. summarize  
   
 Variable | Obs Mean Std. Dev. Min Max  
 -------------+---------------------------------------------------------  
 make | 0  
 price | 74 6165.257 2949.496 3291 15906  
 mpg | 74 21.2973 5.785503 12 41  
 rep78 | 69 3.405797 .9899323 1 5  
 headroom | 74 2.993243 .8459948 1.5 5  
 -------------+---------------------------------------------------------  
 trunk | 74 13.75676 4.277404 5 23  
 weight | 74 3019.459 777.1936 1760 4840  
 length | 74 187.9324 22.26634 142 233  
 turn | 74 39.64865 4.399354 31 51  
 displacement | 74 197.2973 91.83722 79 425  
 -------------+---------------------------------------------------------  
 gear\_ratio | 74 3.014865 .4562871 2.19 3.89  
 foreign | 74 .2972973 .4601885 0 1  
   
 . histogram price  
 (bin=8, start=3291, width=1576.875)



Histogram of the price variable

## Dealining with the returned values

Use the return list command to list the returned values. Most of the estimations are stored in a matrix. Therefore, we copy the matrix and use it to extract the values dynamically.

. regress price mpg foreign  
   
 Source | SS df MS Number of obs = 74  
 -------------+---------------------------------- F(2, 71) = 14.07  
 Model | 180261702 2 90130850.8 Prob > F = 0.0000  
 Residual | 454803695 71 6405685.84 R-squared = 0.2838  
 -------------+---------------------------------- Adj R-squared = 0.2637  
 Total | 635065396 73 8699525.97 Root MSE = 2530.9  
   
 ------------------------------------------------------------------------------  
 price | Coef. Std. Err. t P>|t| [95% Conf. Interval]  
 -------------+----------------------------------------------------------------  
 mpg | -294.1955 55.69172 -5.28 0.000 -405.2417 -183.1494  
 foreign | 1767.292 700.158 2.52 0.014 371.2169 3163.368  
 \_cons | 11905.42 1158.634 10.28 0.000 9595.164 14215.67  
 ------------------------------------------------------------------------------

return the rclass objects

. return list  
   
 scalars:  
 r(level) = 95  
   
 matrices:  
 r(table) : 9 x 3  
   
 . mat A = r(table)  
   
 . mat list A  
   
 A[9,3]  
 mpg foreign \_cons  
 b -294.19553 1767.2922 11905.415  
 se 55.691719 700.15797 1158.6345  
 t -5.2825724 2.5241336 10.275385  
 pvalue 1.333e-06 .01383634 1.085e-15  
 ll -405.24167 371.2169 9595.1638  
 ul -183.1494 3163.3676 14215.667  
 df 71 71 71  
 crit 1.9939434 1.9939434 1.9939434  
 eform 0 0 0

Stata returns the estimations in the **eclass**

. ereturn list  
   
 scalars:  
 e(N) = 74  
 e(df\_m) = 2  
 e(df\_r) = 71  
 e(F) = 14.07044507869964  
 e(r2) = .2838474630671753  
 e(rmse) = 2530.945641206649  
 e(mss) = 180261701.570873  
 e(rss) = 454803694.5507485  
 e(r2\_a) = .2636741521676592  
 e(ll) = -683.3599714095877  
 e(ll\_0) = -695.7128688987767  
 e(rank) = 3  
   
 macros:  
 e(cmdline) : "regress price mpg foreign"  
 e(title) : "Linear regression"  
 e(marginsok) : "XB default"  
 e(vce) : "ols"  
 e(depvar) : "price"  
 e(cmd) : "regress"  
 e(properties) : "b V"  
 e(predict) : "regres\_p"  
 e(estat\_cmd) : "regress\_estat"  
   
 matrices:  
 e(b) : 1 x 3  
 e(V) : 3 x 3  
   
 functions:  
 e(sample)   
   
 . mat B = e(b)  
   
 . mat list B  
   
 B[1,3]  
 mpg foreign \_cons  
 y1 -294.19553 1767.2922 11905.415

# Creating the dynamic table

In the workshop the following question came up. How can we use the returned values from Stata and create a dynamic table.

We are trying to create a **dynamic table** that includes the variables and . We need to get the stored values from returned matrices in Stata in the following procedure:

1. save the Stata matrix with a new name
2. extract the scalars of interest
3. get the name of the columns of the matrix
   1. parse the names of each column
   2. change the /\_const name to
4. use the **tbl** command to generate the dynamic table.

* . scalar scal1 = A[1,1]  
    
   . scalar scal2 = A[1,2]  
    
   . scalar scal3 = A[1,3]   
    
   . local colnms: coln A  
    
   . di "`colnms'"  
   mpg foreign \_cons  
    
   . tokenize "`colnms'"  
    
   . display "`1' `2' `3'"  
   mpg foreign \_cons  
    
   . local n = 1  
    
   . while "`1'" != "" {  
   . if "`1'" == "\_cons" {  
   . local m`n' "$\\beta\_0$"  
   . }  
   . else {  
   . local m`n' `1'  
   . }   
   . local n = `n' + 1  
   . macro shift  
   . }  
    
   . display "`m1' `m2' `m3'"  
   mpg foreign $\\beta\_0$

|  |  |  |
| --- | --- | --- |
| mpg | foreign |  |
| -294.20 | 1767.29 | 11905.42 |