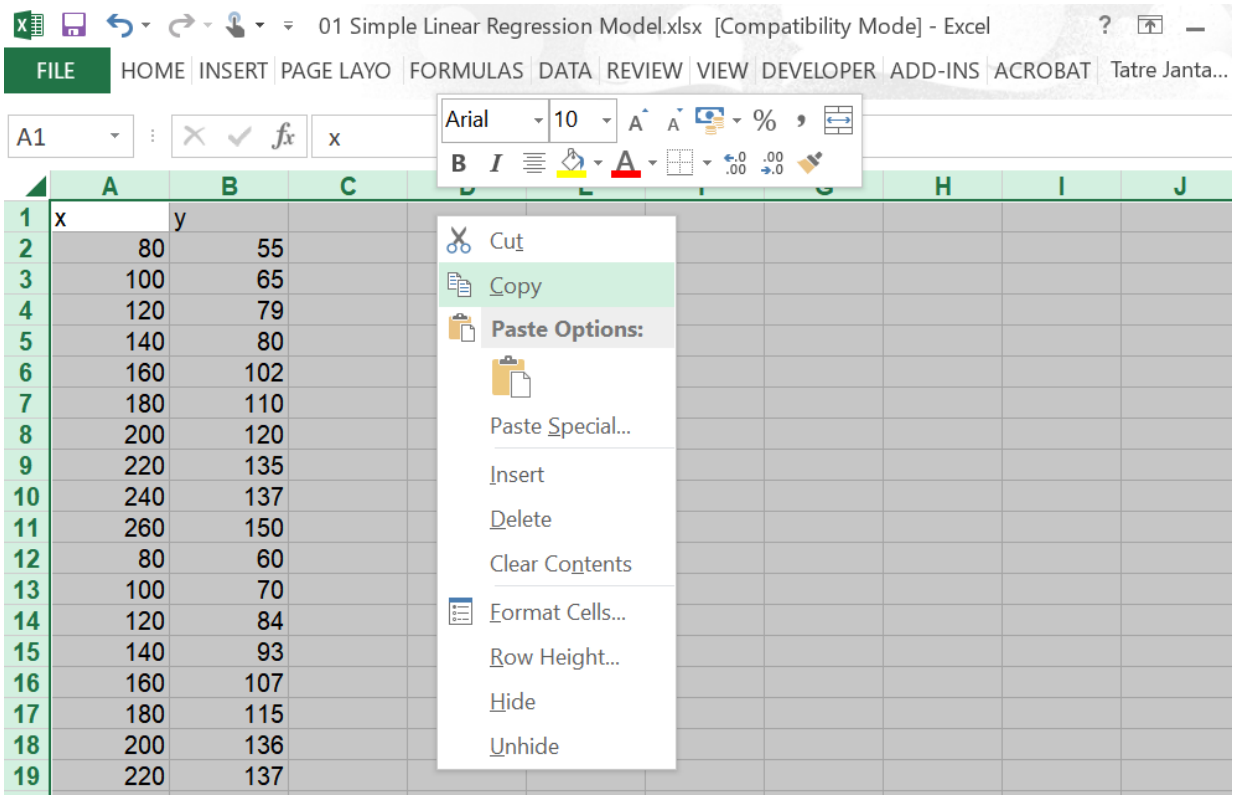
Instruction using STATA

By Tatre Jantarakolica

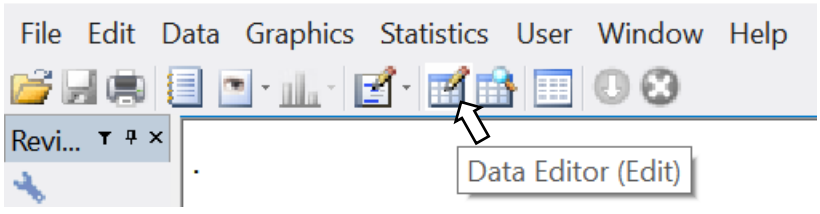
Import Data

From Excel, copy exporting data.

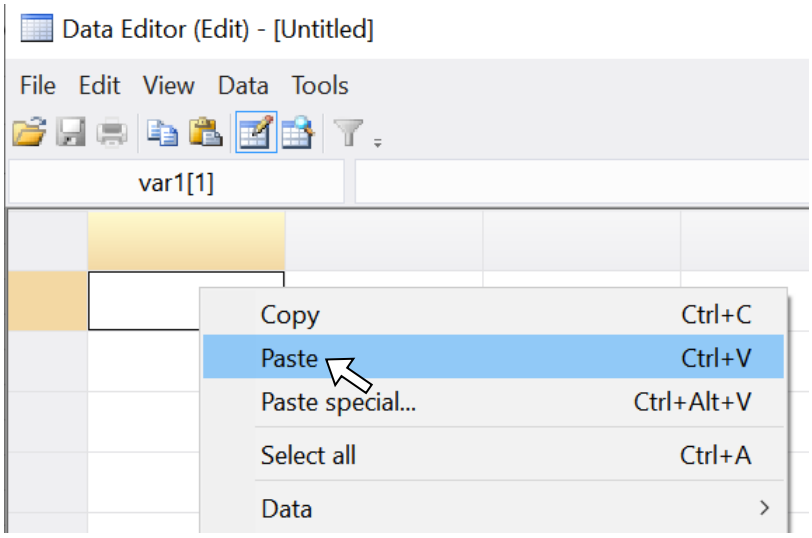


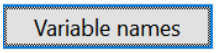
	A	B
1	x	y
2	80	55
3	100	65
4	120	79
5	140	80
6	160	102
7	180	110
8	200	120
9	220	135
10	240	137
11	260	150
12	80	60
13	100	70
14	120	84
15	140	93
16	160	107
17	180	115
18	200	136
19	220	137

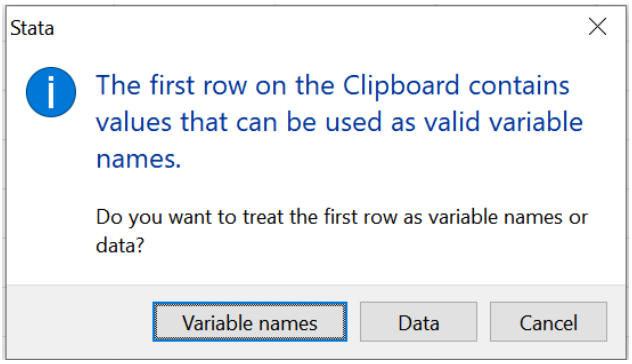
From STATA menu bar, click on icon Data Editor (Edit) 



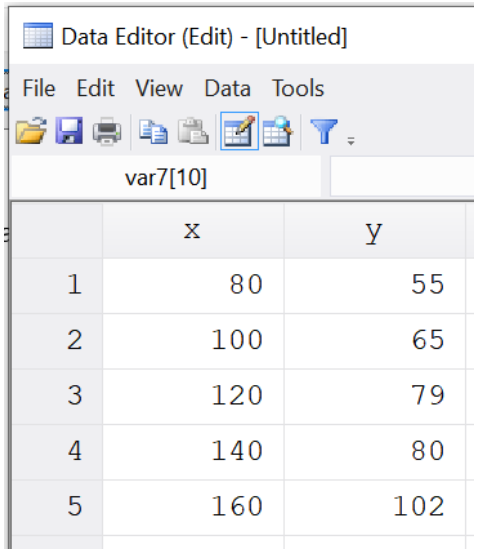
Then, Click Paste on Data Editor Window:



Then, Choose Variable names  button to set up the first row in Excel as variable names:



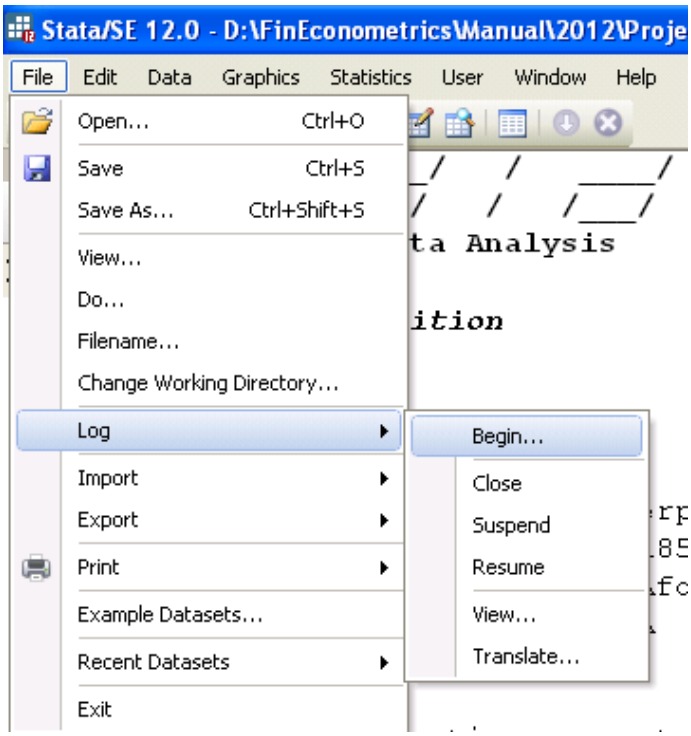
The data will be in Data Editor window as follows:



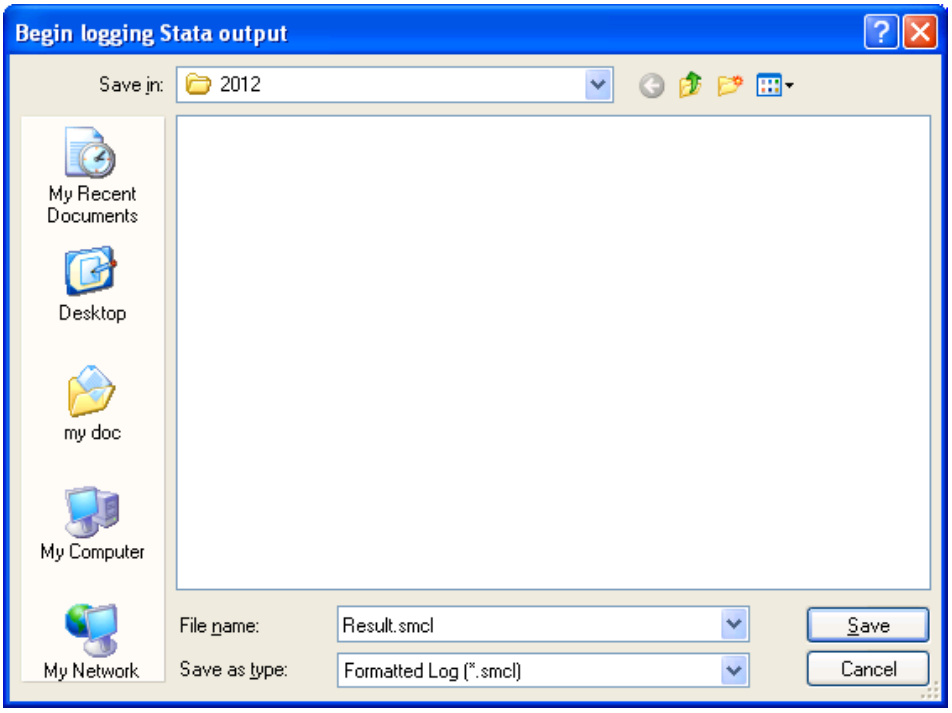
	x	y
1	80	55
2	100	65
3	120	79
4	140	80
5	160	102

Save Output File

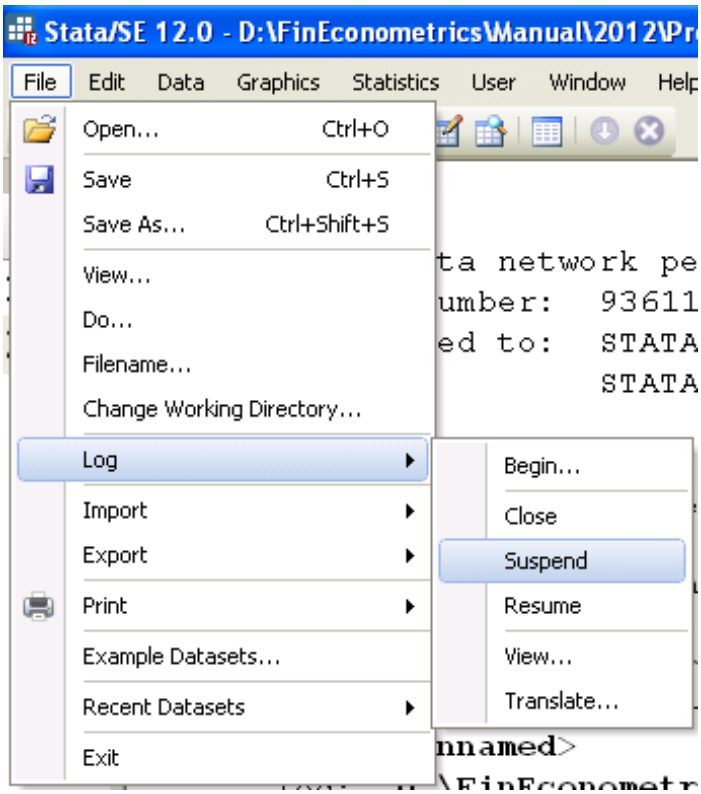
To save the output from data analysis in STATA, from menu bar go to File, select Log, and choose Begin…



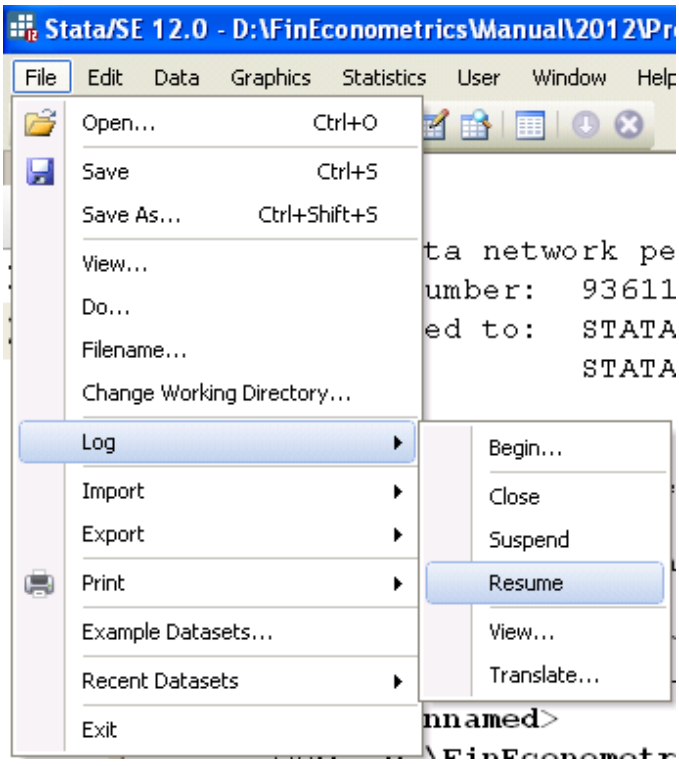
Then, in Begin logging Stata output window box, input file name that you would like to save and select the folder, then, click Save (as smcl (Stata Markup and Control Language) format).



If you would like to suspend some analysis, then, from Menu bar, go to File, select Log, and choose Suspend.

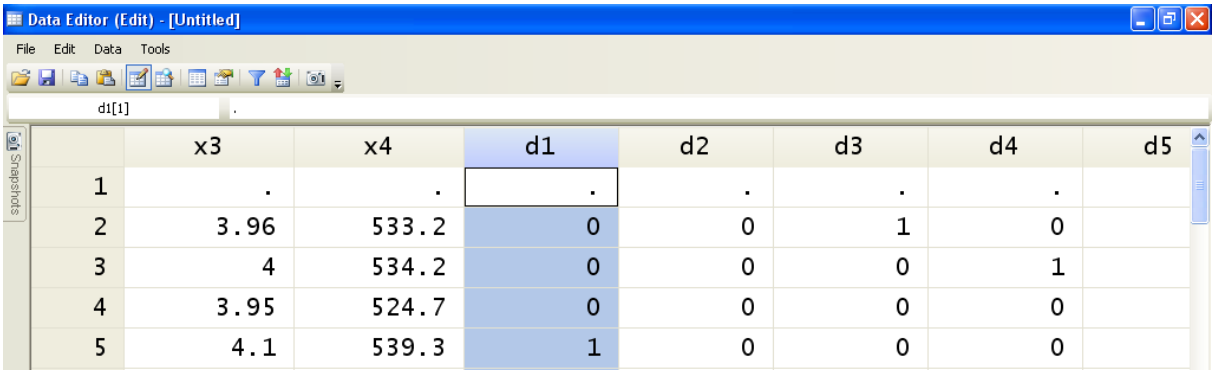


When you want to continue, go to Resume. If you finish, choose Close.



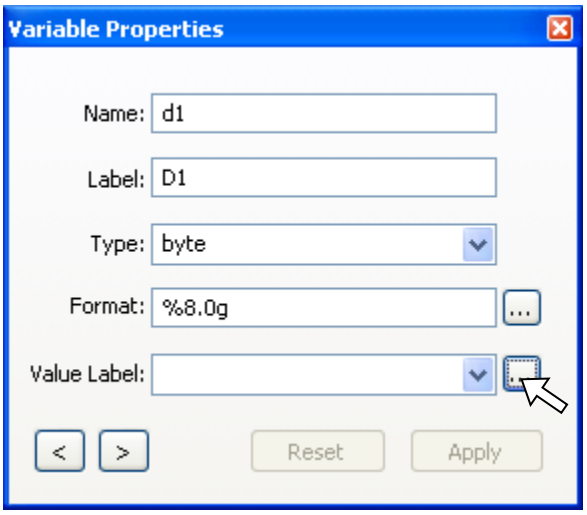
Define Name and Definition of the Variables

From Data Editor window, double click on the variable.

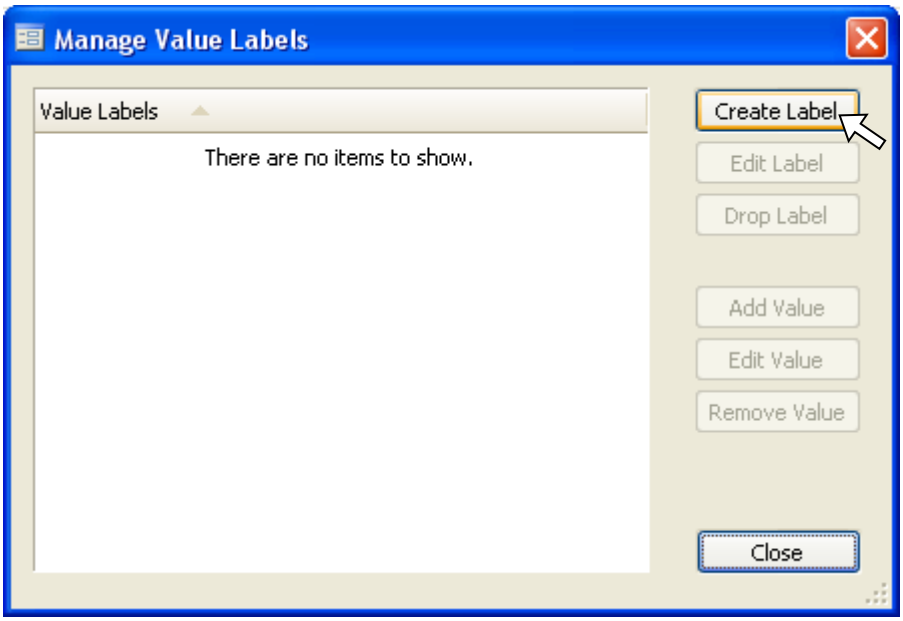


	x3	x4	d1	d2	d3	d4	d5
1	.	.	.	.	.	.	
2	3.96	533.2	0	0	1	0	
3	4	534.2	0	0	0	1	
4	3.95	524.7	0	0	0	0	
5	4.1	539.3	1	0	0	0	

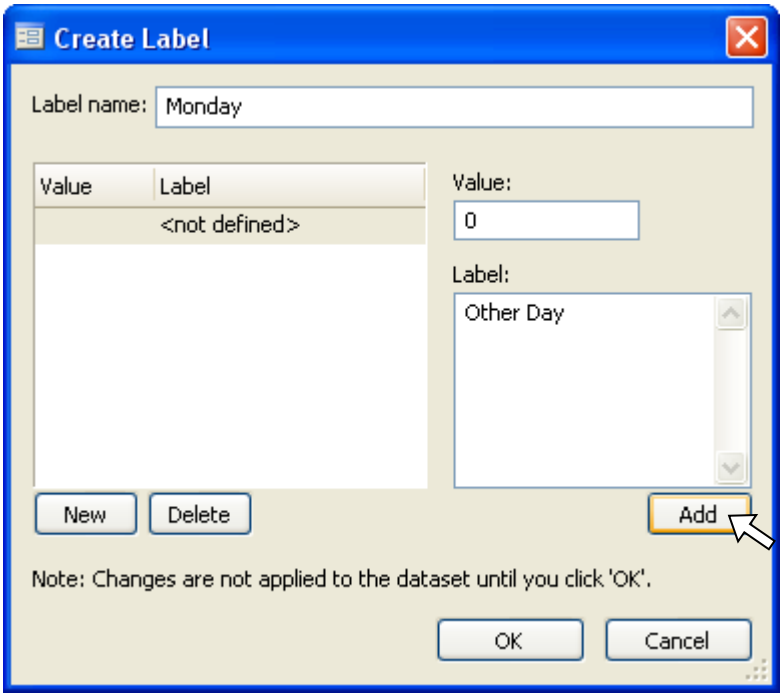
From Variable Properties window, specify Label of the variable, then, choose Define/Modify…



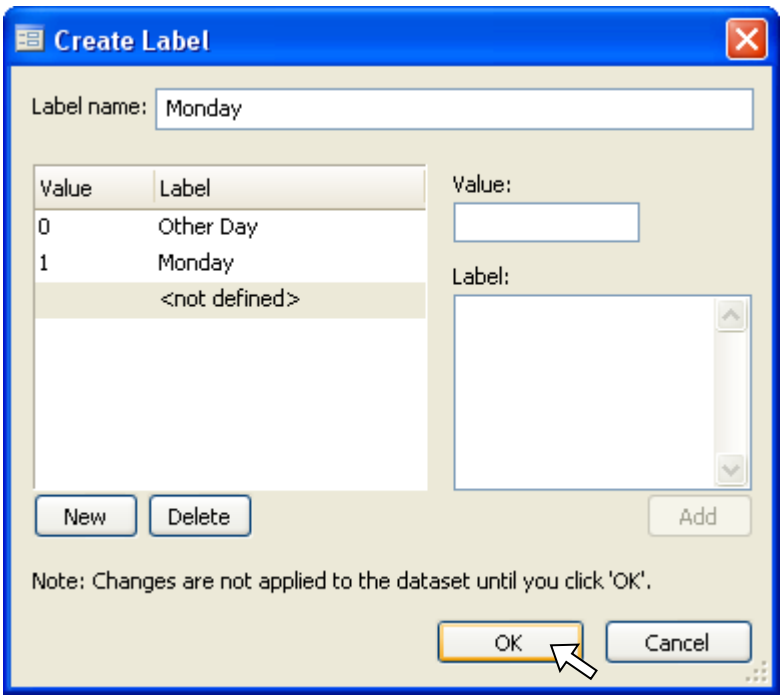
From Manage Value Labels window, choose Create Label



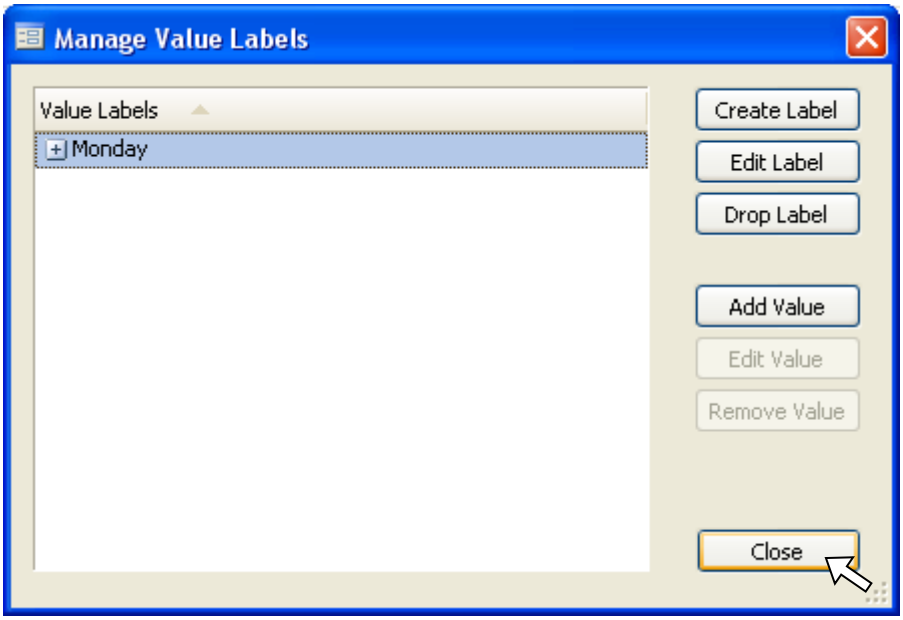
From Create Label window, define Label name: (in this case Monday), Value: (in this case 0), Label: (in this case Other Day), then click Add.



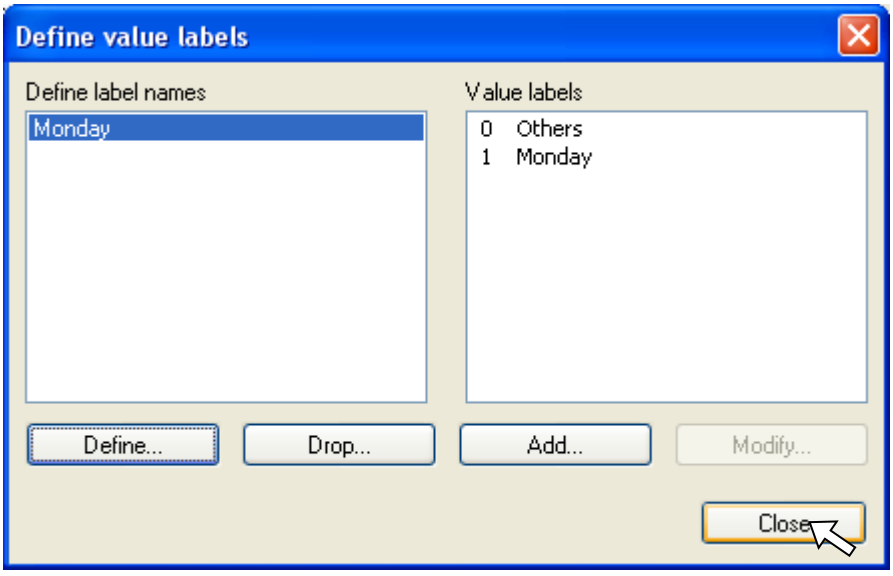
Then, Value: 1 and Label: Monday. After finish define Value and Label, click OK.



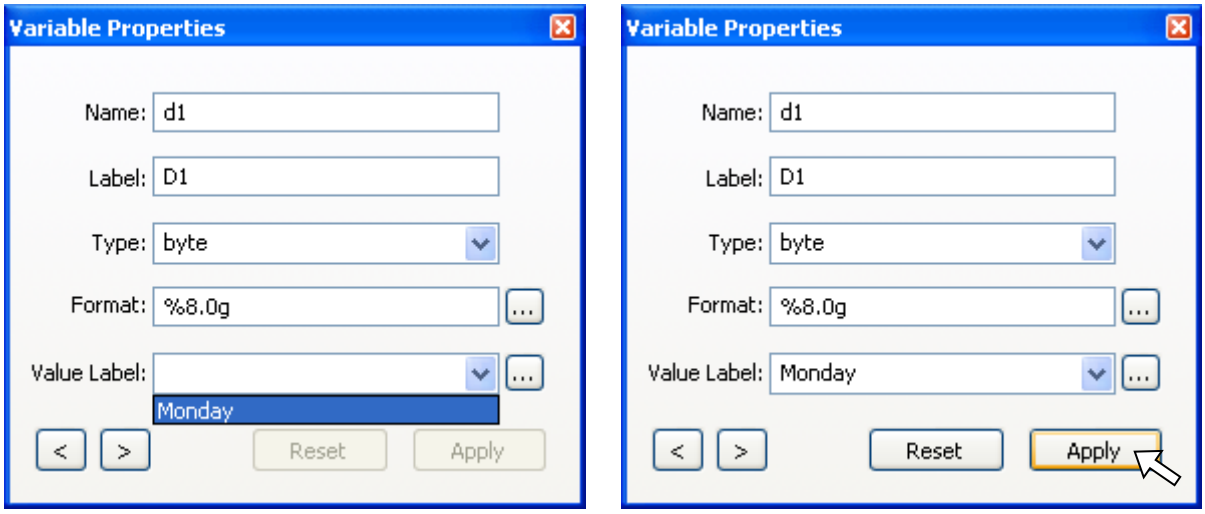
After finish create label, from Manage Value Labels window, click Close.



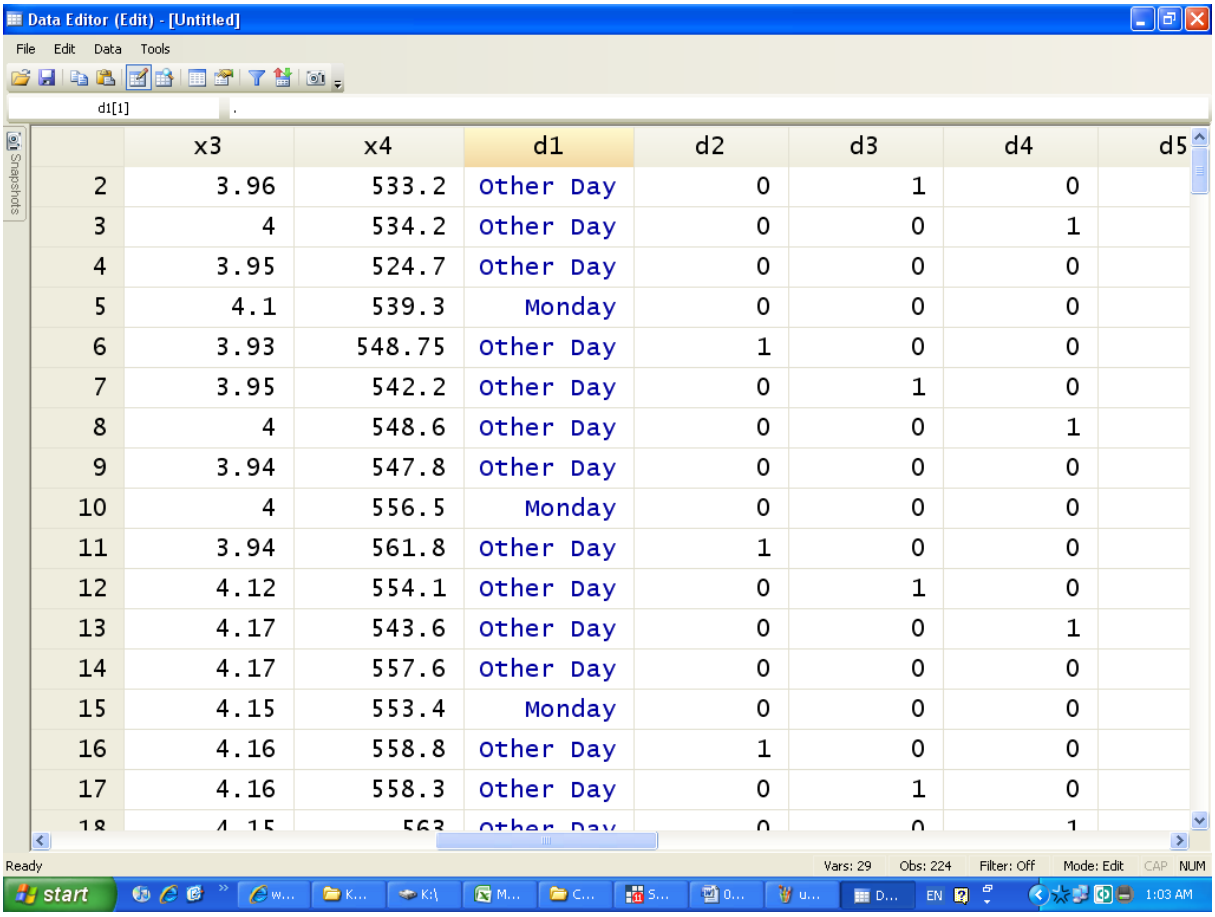
After finished inputing value of variables, click Cancel. Then, from Define value labels window, click Close.



Then, from Variable Properties window, go to Value label box, choose the specify value label, then, click Apply.



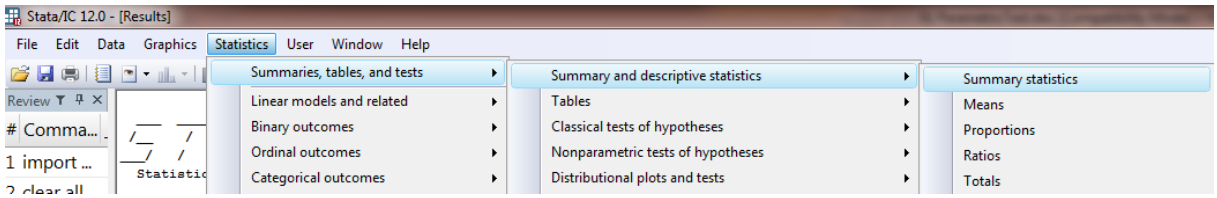
Then, the value label will appear instead of value.



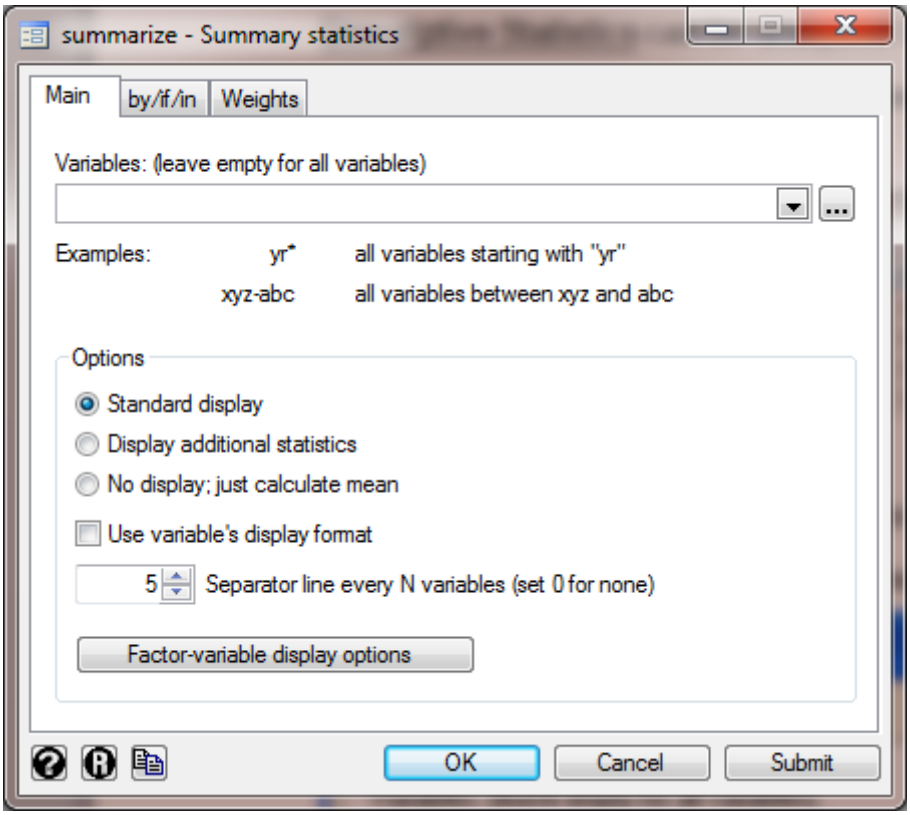
	x3	x4	d1	d2	d3	d4	d5
2	3.96	533.2	Other Day	0	1	0	
3	4	534.2	Other Day	0	0	1	
4	3.95	524.7	Other Day	0	0	0	
5	4.1	539.3	Monday	0	0	0	
6	3.93	548.75	Other Day	1	0	0	
7	3.95	542.2	Other Day	0	1	0	
8	4	548.6	Other Day	0	0	1	
9	3.94	547.8	Other Day	0	0	0	
10	4	556.5	Monday	0	0	0	
11	3.94	561.8	Other Day	1	0	0	
12	4.12	554.1	Other Day	0	1	0	
13	4.17	543.6	Other Day	0	0	1	
14	4.17	557.6	Other Day	0	0	0	
15	4.15	553.4	Monday	0	0	0	
16	4.16	558.8	Other Day	1	0	0	
17	4.16	558.3	Other Day	0	1	0	
18	4.15	563	Other Day	0	0	1	

Descriptive Statistics can be analyzed by:

From Menu Bar, go to Statistics, choose Summaries, tables, & tests, select Summary statistics, select Summary statistics



From summarize – Summary statistics window, specify the variables to be analyzed in Variables: (leave empty for all variables) box, then click OK



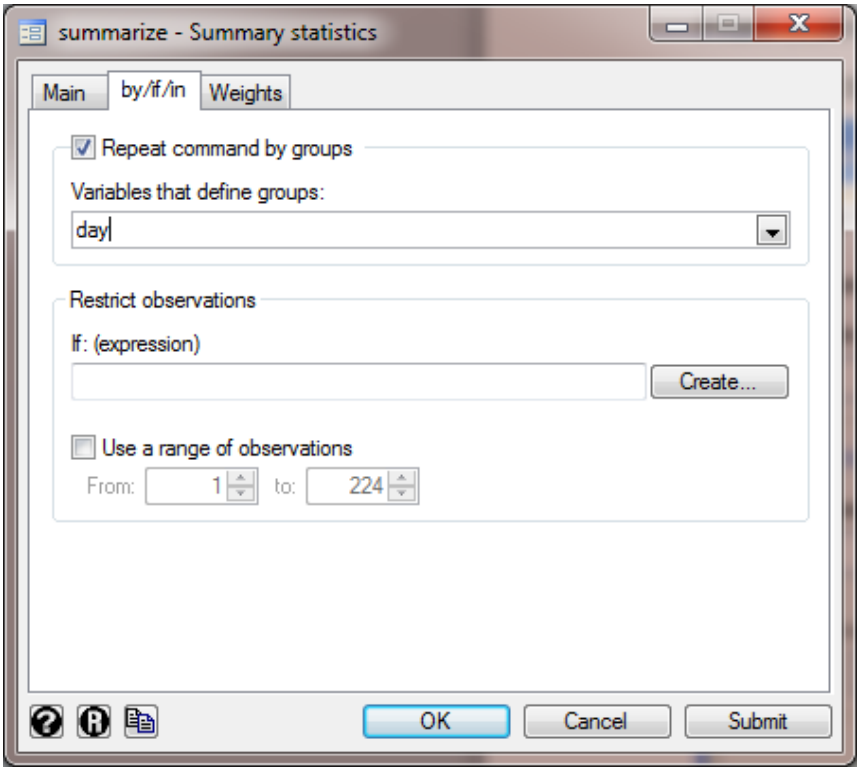
The result will be as follows:

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
day	0				
date	0				
time	224	111.5	64.80741	0	223
rm	223	.0134254	1.022241	-3.538129	3.050996
rj	223	.0012806	1.841034	-6.666667	4.580153
rf	223	.0201872	.0010914	.0180685	.0216338
ex	223	38.09048	.9104958	36.04	40.74
r	223	4.719708	.3435897	3.93	5.16
gp	223	601.4513	39.4314	524.7	718.8
y	223	-.0189067	1.840972	-6.68504	4.55939
x1	223	-.006762	1.022251	-3.55919	3.02984
x2	223	38.09048	.9104958	36.04	40.74
x3	223	4.719708	.3435897	3.93	5.16
x4	223	601.4513	39.4314	524.7	718.8
d1	223	.1883408	.3918637	0	1
d2	223	.2017937	.4022419	0	1
d3	223	.206278	.4055427	0	1
d4	223	.206278	.4055427	0	1
d5	223	.1973094	.3988632	0	1

Descriptive Statistics Classified by Groups can be done by

From summarize – Summary statistics window, choose by/if/in, specify grouping variable, in Variables that define groups:, then, click OK.

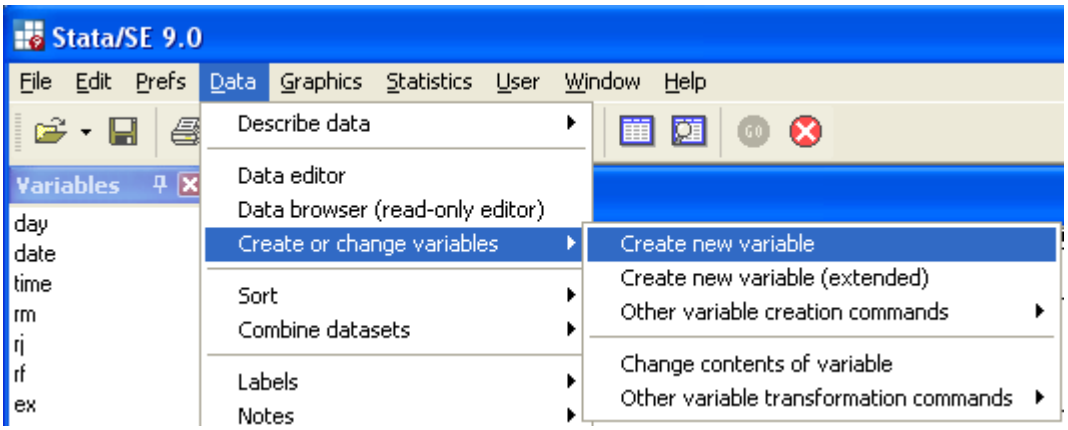


The results will be as follows:

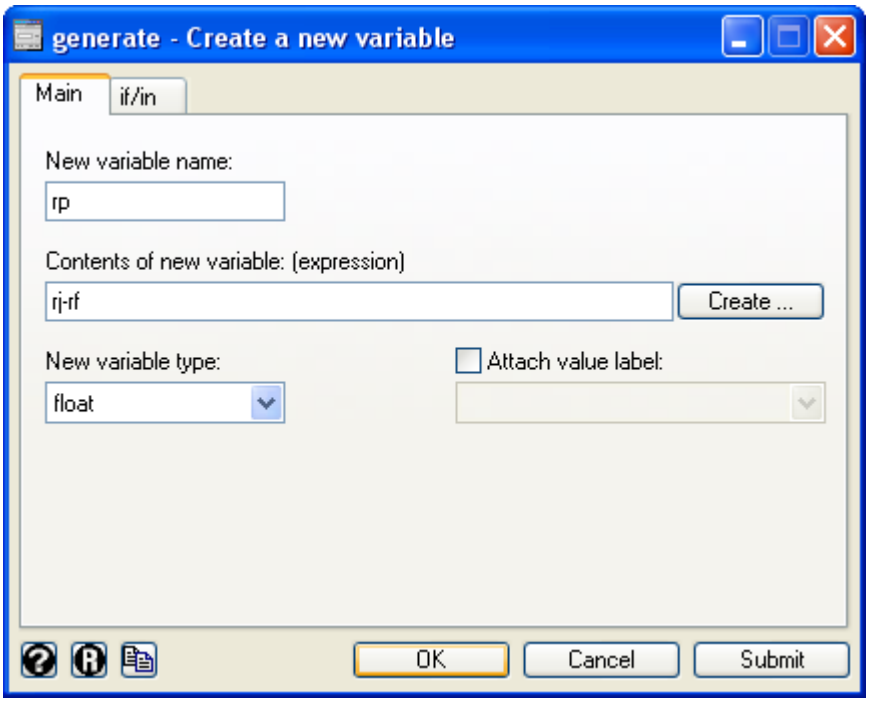
```
. by day, sort : summarize rm

-------------------------------------------------------------------------------
-> day = FRI

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          rm |        44    .2529098     .898997  -1.598893    2.63296

-------------------------------------------------------------------------------
-> day = MON

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          rm |        42   -.1194381    1.034139  -2.933021   1.996412

-------------------------------------------------------------------------------
-> day = THU

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          rm |        46   -.0067526    1.061755  -2.080594   1.698142

-------------------------------------------------------------------------------
-> day = TUE

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          rm |        45   -.1324269    .8035679  -1.955488   1.540987

-------------------------------------------------------------------------------
-> day = WED

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          rm |        46    .0685229    1.246002  -3.538129   3.050996
.
```

Create Variable

To create risk premium variable (stock return minus risk-free rate), from Menu Bar, go to Data, choose Create or change variables, select Create new variable



From generate – Create a new variable window, specify new variable name in New variable name: box, and expression of the new variable in Contents of new variable: (expression) box, then, click OK

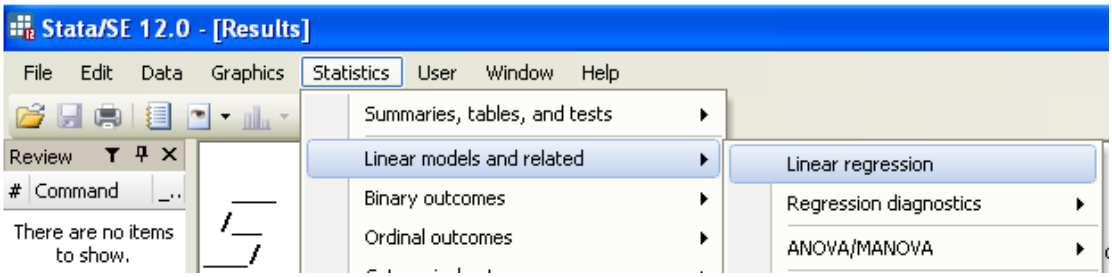


Introduction to the Regression Model

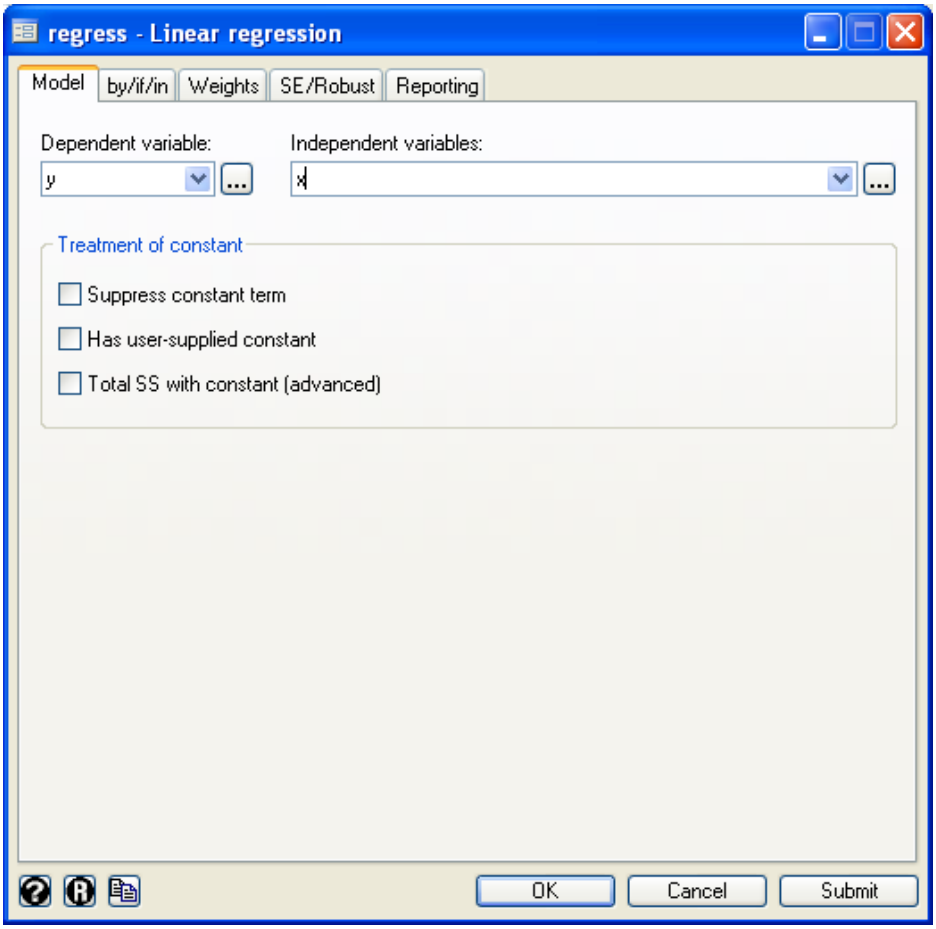
Simple Regression Model

The model: $Y_i = \beta_1 + \beta_2 X_i + u_i$

In this example, we estimate the least-squares regression of Y_i on X_i. To run a regression, from menu bar go to Statistics, choose Linear models and related, select Liner regression.



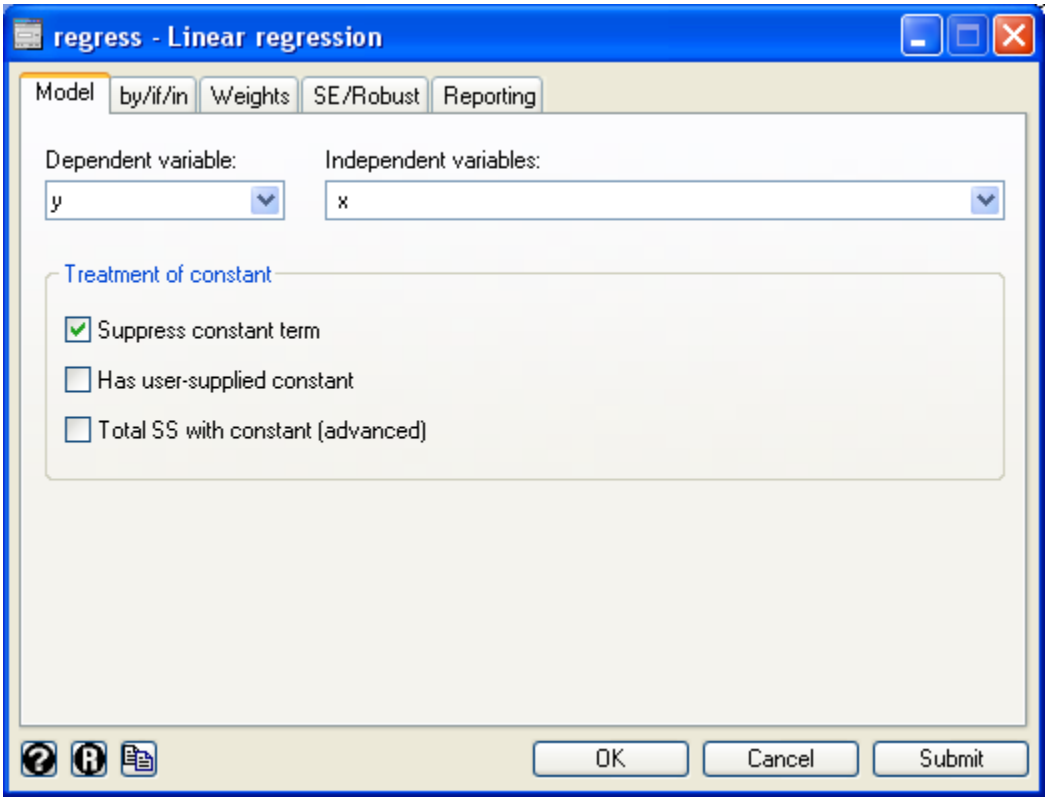
From regress – Linear regression window, specify dependent and independent variables in Dependent variable: and Independent variables: boxes, then, click OK.



The results will be as follows:

```
. regress y x

      Source |       SS       df       MS              Number of obs =      70
-------------+------------------------------           F(  1,    68) =  308.45
       Model |  84751.5429     1  84751.5429            Prob > F      =  0.0000
    Residual |  18683.8286    68  274.762185            R-squared     =  0.8194
-------------+------------------------------           Adj R-squared =  0.8167
       Total |  103435.371    69  1499.06335            Root MSE      =  16.576

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .6057143   .0344884    17.56   0.000     .5368938    .6745347
       _cons |   20.28571   6.188714     3.28   0.002     7.936327     32.6351
------------------------------------------------------------------------------

.est store withcon
```

To run the model without intercept term, check Suppress constant term box.




```
. regress y x, noconstant

      Source |       SS       df       MS              Number of obs =      70
-------------+------------------------------           F(  1,    69) = 3652.39
       Model |  1145262.03     1  1145262.03            Prob > F      =  0.0000
    Residual |   21635.967    69  313.564739            R-squared     =  0.9815
-------------+------------------------------           Adj R-squared =  0.9812
       Total |     1166898    70  16669.9714            Root MSE      =  17.708

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .7128128   .0117947     60.44   0.000      .689283    .7363425
------------------------------------------------------------------------------

. est store wocon
```

```
. est table withcon wocon, star(0.1 0.05 0.01) stat(N F r2)

----------------------------------------------
    Variable |    withcon          wocon
-------------+--------------------------------
           x |  .60571429***    .71281278***
       _cons |  20.285714***
-------------+--------------------------------
           N |         70              70
           F |  308.45417       3652.3942
          r2 |  .81936712       .98145856
----------------------------------------------
           legend: * p<.1; ** p<.05; *** p<.01
```

Multiply or Divide constant value to the variable

Generate new variable by:

```
. g x10 = x*10

. g y10 = y*10

. reg y x

      Source |       SS       df       MS              Number of obs =      70
-------------+------------------------------           F(  1,    68) =  308.45
       Model |  84751.5429     1  84751.5429           Prob > F      =  0.0000
    Residual |  18683.8286    68  274.762185           R-squared     =  0.8194
-------------+------------------------------           Adj R-squared =  0.8167
       Total |  103435.371    69  1499.06335           Root MSE      =  16.576

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .6057143   .0344884    17.56   0.000     .5368938    .6745347
       _cons |   20.28571   6.188714     3.28   0.002     7.936327     32.6351
------------------------------------------------------------------------------


. reg y x10

      Source |       SS       df       MS              Number of obs =      70
-------------+------------------------------           F(  1,    68) =  308.45
       Model |  84751.5429     1  84751.5429           Prob > F      =  0.0000
    Residual |  18683.8286    68  274.762185           R-squared     =  0.8194
-------------+------------------------------           Adj R-squared =  0.8167
       Total |  103435.371    69  1499.06335           Root MSE      =  16.576

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         x10 |   .0605714   .0034488    17.56   0.000     .0536894    .0674535
       _cons |   20.28571   6.188714     3.28   0.002     7.936327     32.6351
------------------------------------------------------------------------------


. reg y10 x

      Source |       SS       df       MS              Number of obs =      70
-------------+------------------------------           F(  1,    68) =  308.45
       Model |  8475154.29     1  8475154.29           Prob > F      =  0.0000
    Residual |  1868382.86    68  27476.2185           R-squared     =  0.8194
-------------+------------------------------           Adj R-squared =  0.8167
       Total |  10343537.1    69  149906.335           Root MSE      =  165.76

------------------------------------------------------------------------------
         y10 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   6.057143   .3448836    17.56   0.000     5.368938    6.745347
       _cons |   202.8571   61.88714     3.28   0.002     79.36327     326.351
------------------------------------------------------------------------------


. reg y10 x10

      Source |       SS       df       MS              Number of obs =      70
-------------+------------------------------           F(  1,    68) =  308.45
       Model |  8475154.29     1  8475154.29           Prob > F      =  0.0000
    Residual |  1868382.86    68  27476.2185           R-squared     =  0.8194
-------------+------------------------------           Adj R-squared =  0.8167
```

```
       Total |  10343537.1    69  149906.335           Root MSE      =  165.76

------------------------------------------------------------------------------
         y10 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         x10 |   .6057143   .0344884    17.56   0.000     .5368938    .6745347
       _cons |   202.8571   61.88714     3.28   0.002     79.36327     326.351
------------------------------------------------------------------------------

. g lny=ln(y)

. g lnx=ln(x)


. reg lny lnx

      Source |       SS       df       MS              Number of obs =      70
-------------+------------------------------           F(  1,    68) =  405.67
       Model |  6.35437211     1  6.35437211           Prob > F      =  0.0000
    Residual |  1.06515616    68  .015664061           R-squared     =  0.8564
-------------+------------------------------           Adj R-squared =  0.8543
       Total |  7.41952827    69  .107529395           Root MSE      =  .12516

------------------------------------------------------------------------------
         lny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnx |   .8148884   .0404589    20.14   0.000     .7341539    .8956228
       _cons |   .6302696   .2057428     3.06   0.003     .2197162    1.040823
------------------------------------------------------------------------------

. reg lny x

      Source |       SS       df       MS              Number of obs =      70
-------------+------------------------------           F(  1,    68) =  357.82
       Model |  6.23468848     1  6.23468848           Prob > F      =  0.0000
    Residual |  1.18483979    68  .017424115           R-squared     =  0.8403
-------------+------------------------------           Adj R-squared =  0.8380
       Total |  7.41952827    69  .107529395           Root MSE      =    .132

------------------------------------------------------------------------------
         lny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .0051952   .0002746    18.92   0.000     .0046471    .0057432
       _cons |   3.880016    .049283    78.73   0.000     3.781673    3.978359
------------------------------------------------------------------------------

. reg y lnx

      Source |       SS       df       MS              Number of obs =      70
-------------+------------------------------           F(  1,    68) =  274.84
       Model |  82919.4567     1  82919.4567           Prob > F      =  0.0000
    Residual |  20515.9147    68  301.704629           R-squared     =  0.8017
-------------+------------------------------           Adj R-squared =  0.7987
       Total |  103435.371    69  1499.06335           Root MSE      =   17.37

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnx |   93.08721   5.615039    16.58   0.000     81.88257    104.2918
       _cons |    -348.86   28.55377   -12.22   0.000    -405.8382   -291.8819
------------------------------------------------------------------------------
```