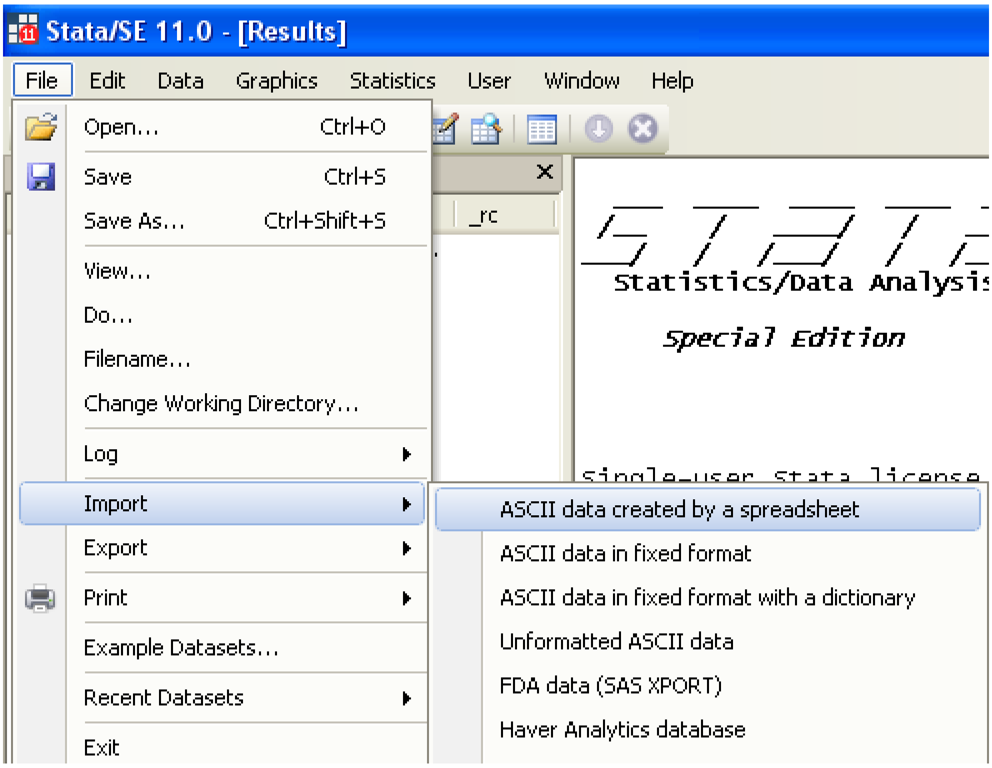
STATA Manual

By Tatre Jantarakolica

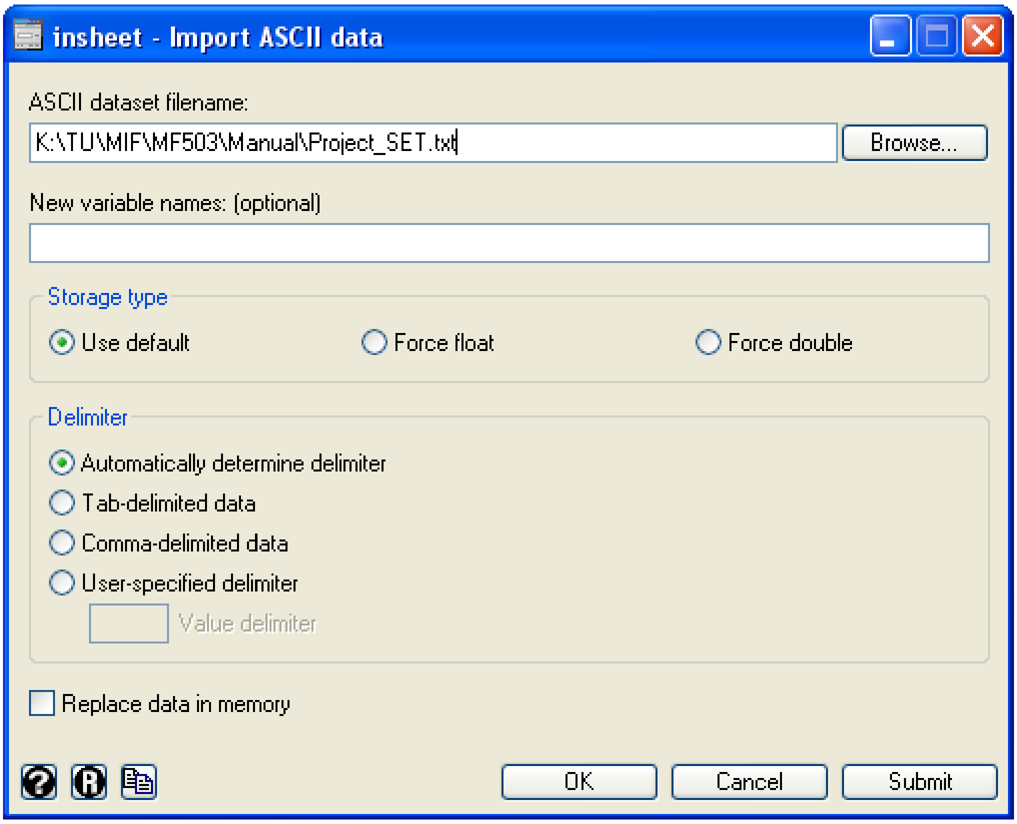
<u>Import Data</u>

To import data from Text file, we can begin by:

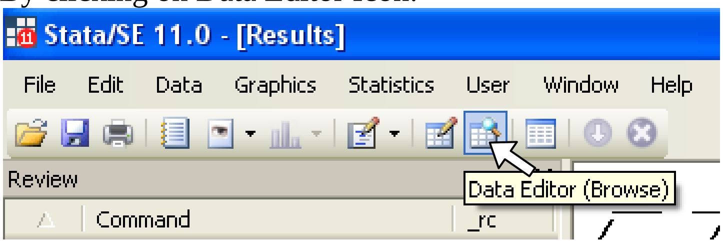
- From menu bar, go to File, select Import, choose ASCII data created by a spreadsheet



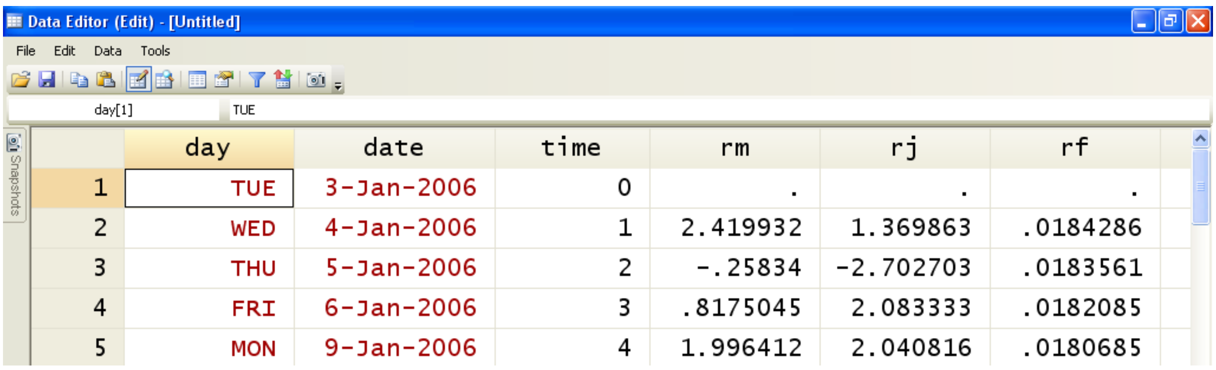
In insheet – Import ASCII data window, specify the data file (Text file). Then, click OK.



By clicking on Data Editor Icon:

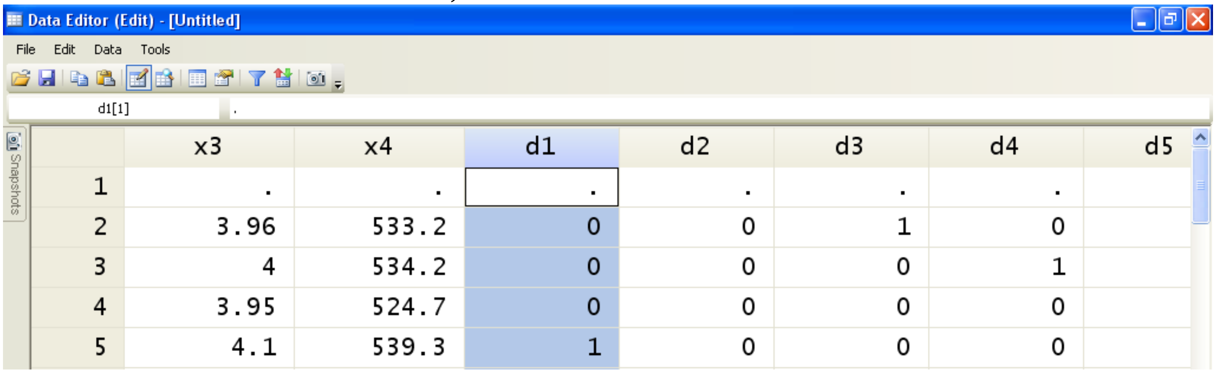


The data will be in Data Editor window as follows:

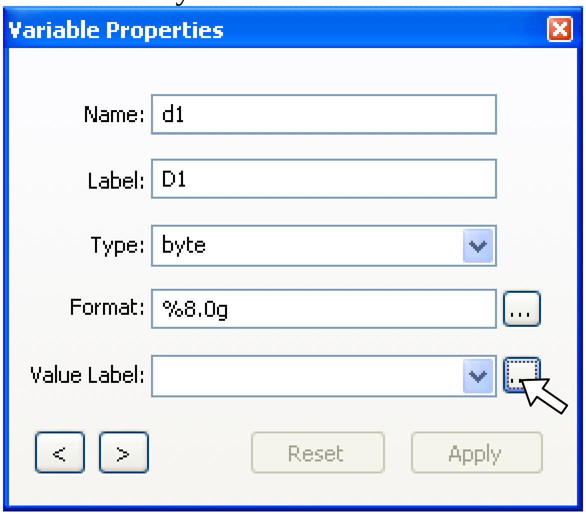


<u>Define Name and Definition of the Variables</u>

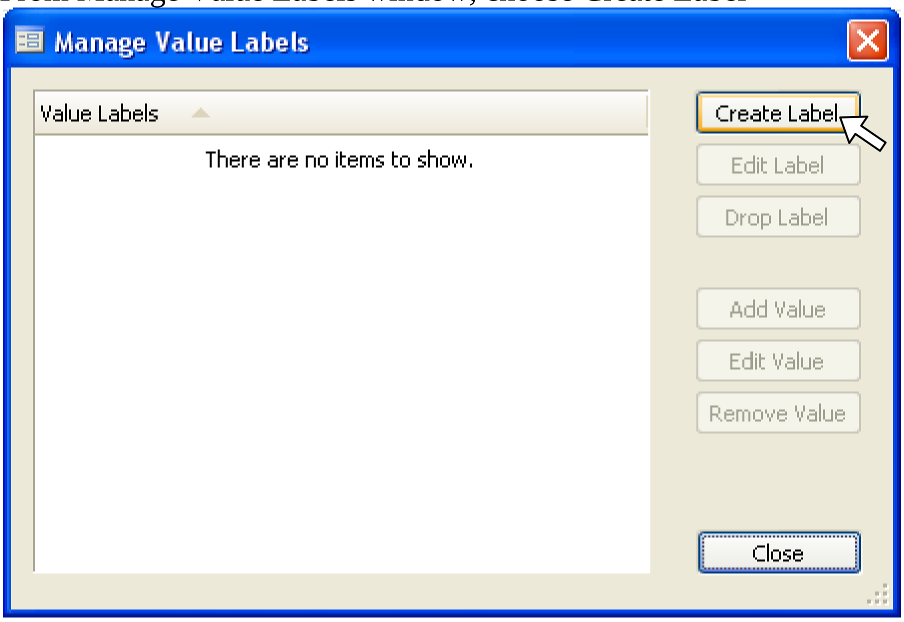
From Data Editor window, double click on the variable.



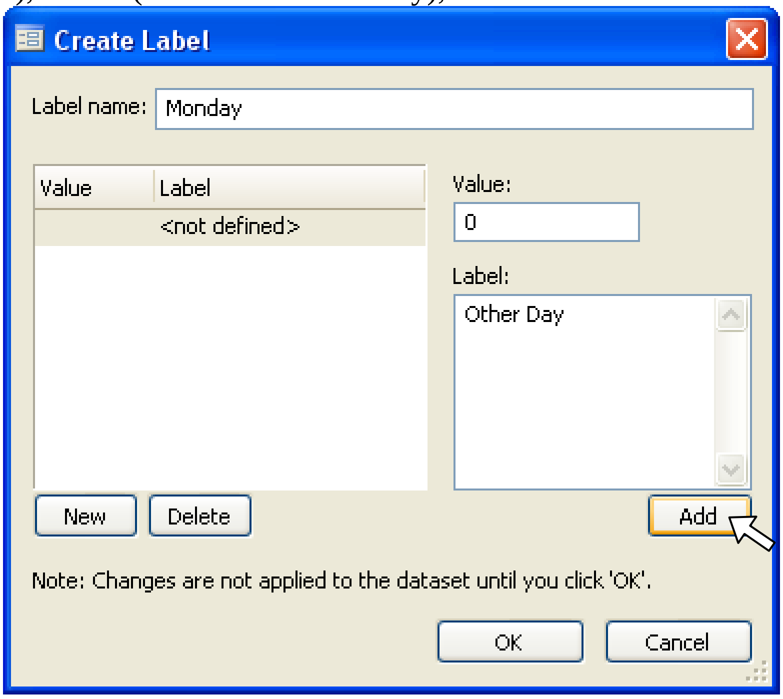
From Variable Properties window, specify Label of the variable, then, choose Define/Modify…



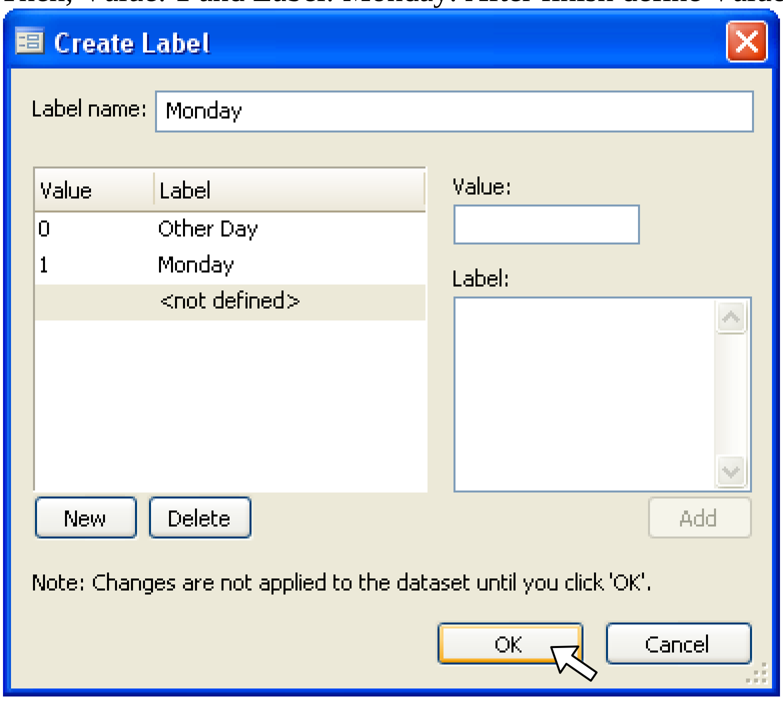
From Manage Value Labels window, choose Create Label



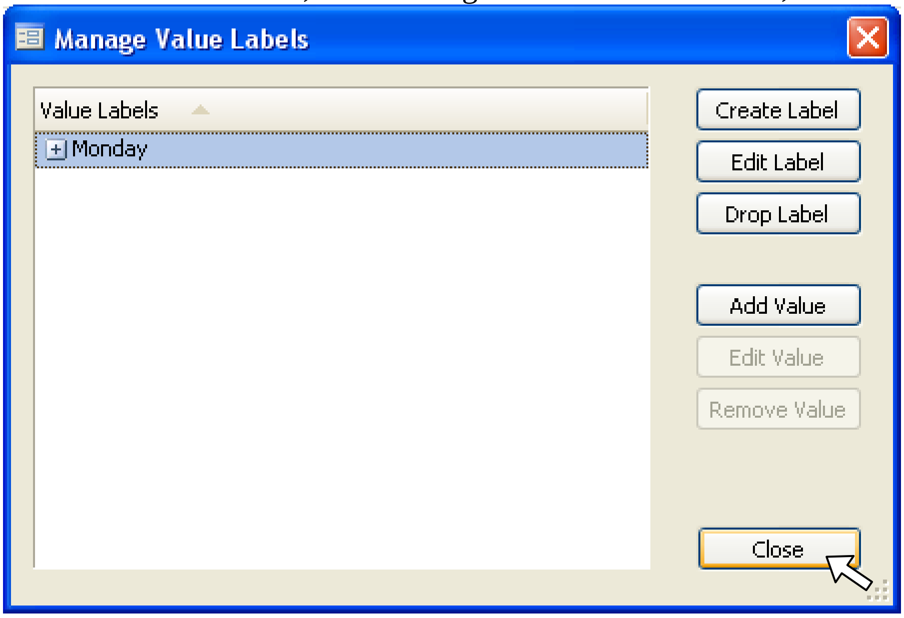
From Create Label window, define Label name: (in this case Monday), Value: (in this case 0), Label: (in this case Other Day), then click Add.



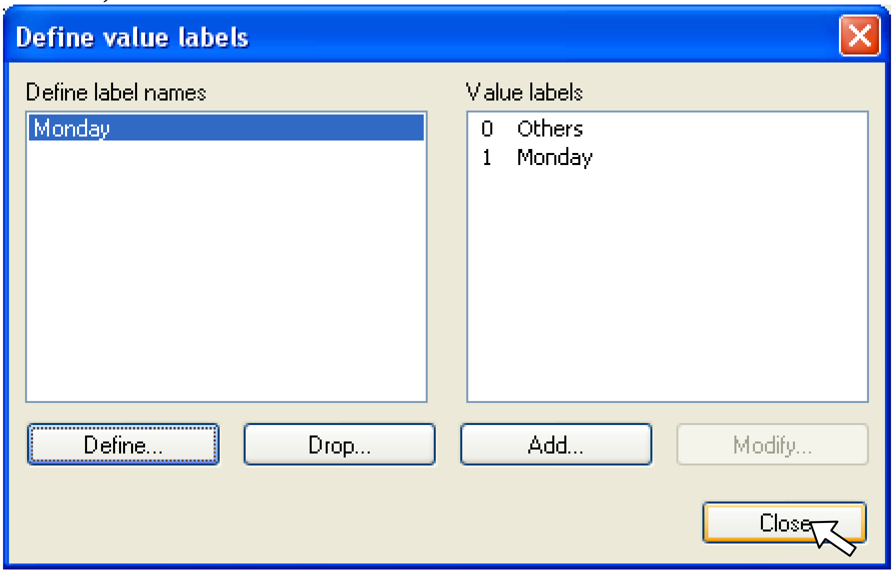
Then, Value: 1 and Label: Monday. After finish define Value and Label, click OK.



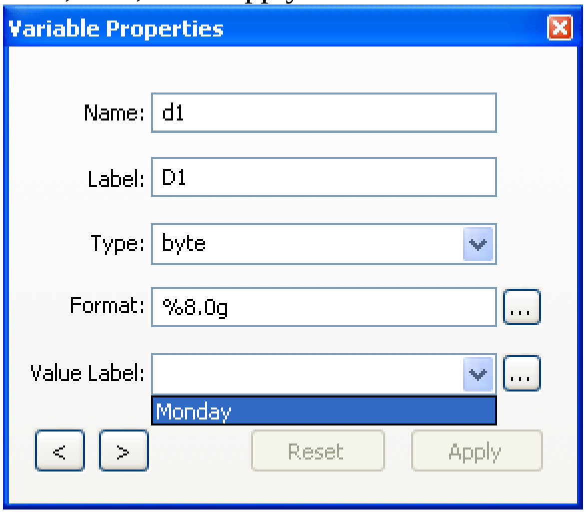
After finish create label, from Manage Value Labels window, click Close.

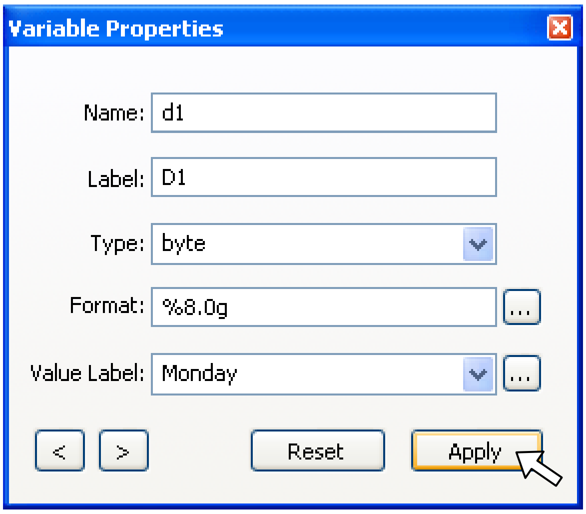


After finished inputing value of variables, click Cancel. Then, from Define value labels window, click Close.

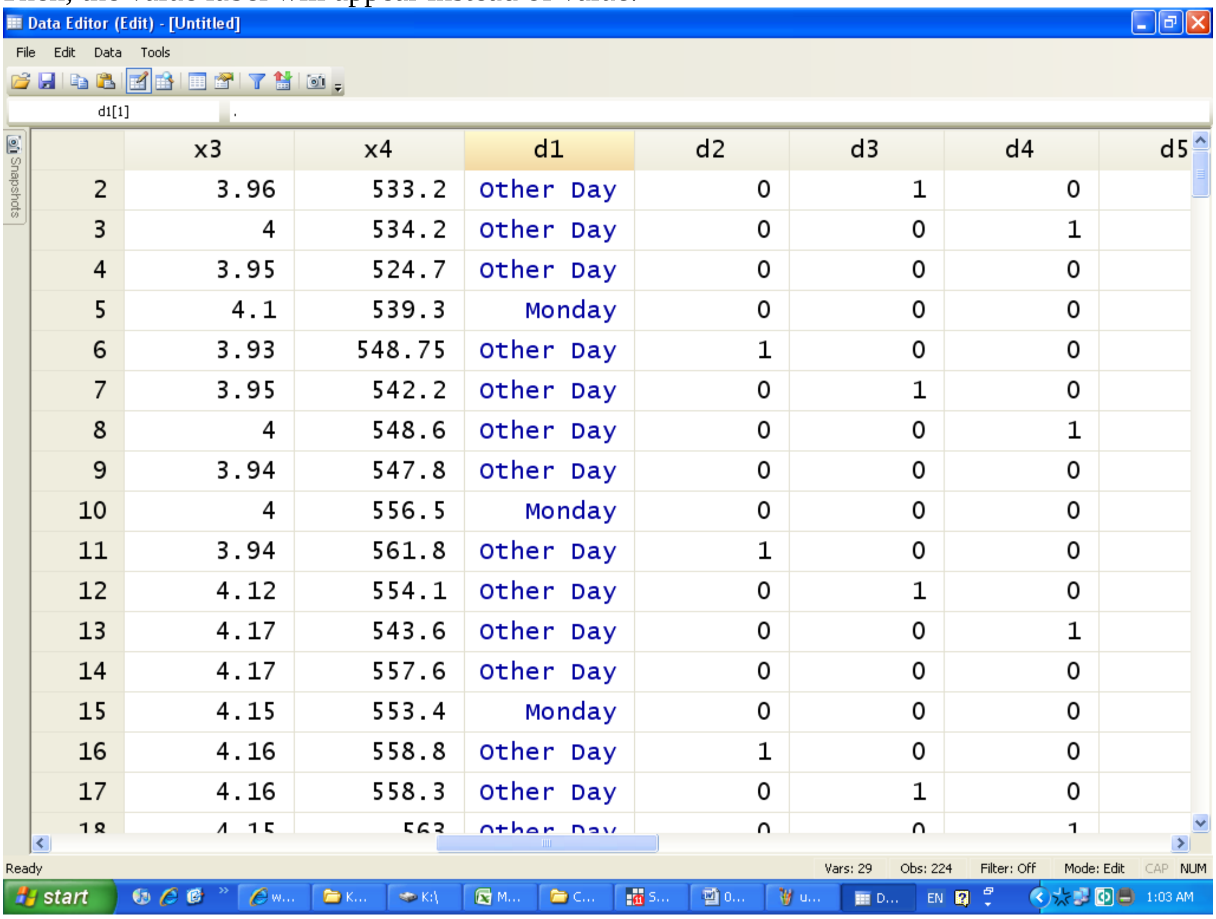


Then, from Variable Properties window, go to Value label box, choose the specify value label, then, click Apply.



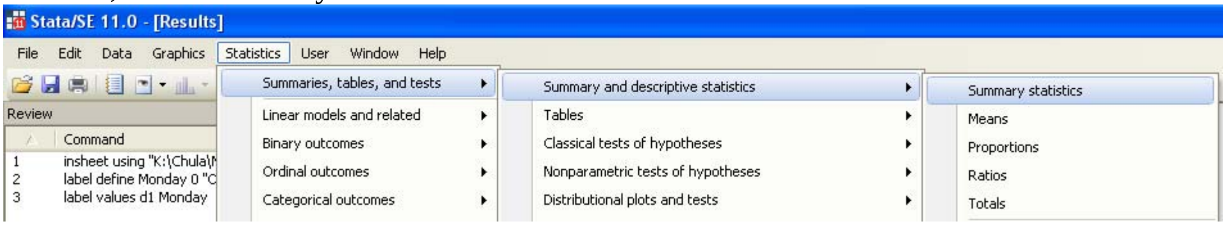


Then, the value label will appear instead of value.

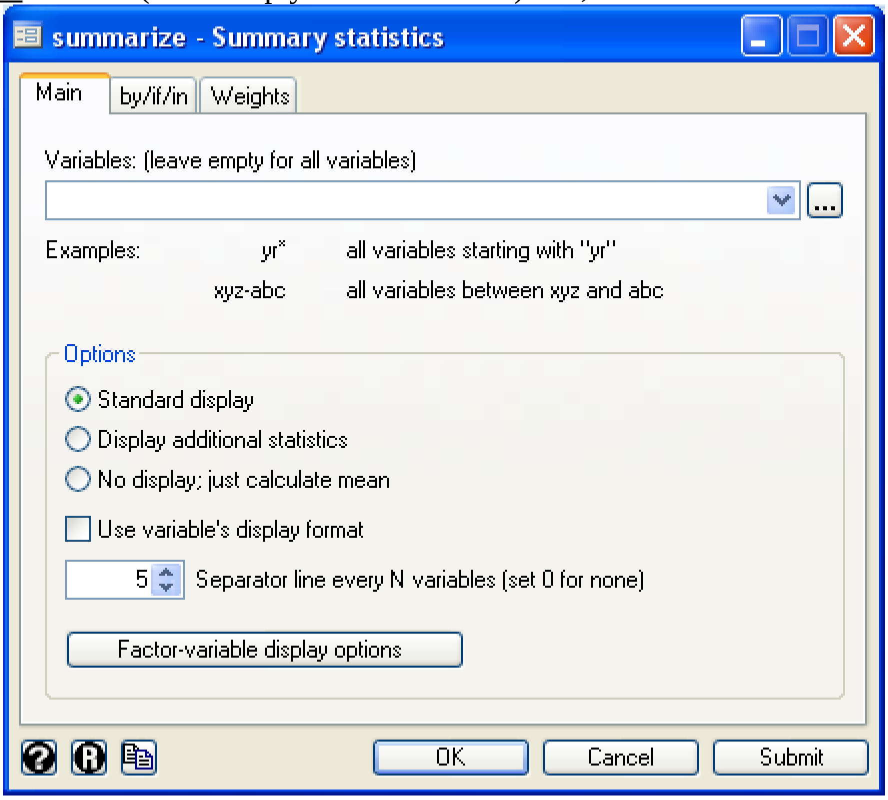


	x3	x4	d1	d2	d3	d4	d5
2	3.96	533.2	Other Day	0	1	0	
3	4	534.2	Other Day	0	0	1	
4	3.95	524.7	Other Day	0	0	0	
5	4.1	539.3	Monday	0	0	0	
6	3.93	548.75	Other Day	1	0	0	
7	3.95	542.2	Other Day	0	1	0	
8	4	548.6	Other Day	0	0	1	
9	3.94	547.8	Other Day	0	0	0	
10	4	556.5	Monday	0	0	0	
11	3.94	561.8	Other Day	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	

Descriptive Statistics can be analyzed by:
From Menu Bar, go to Statistics, choose Summaries, tables, & tests, select Summary statistics, select Summary statistics



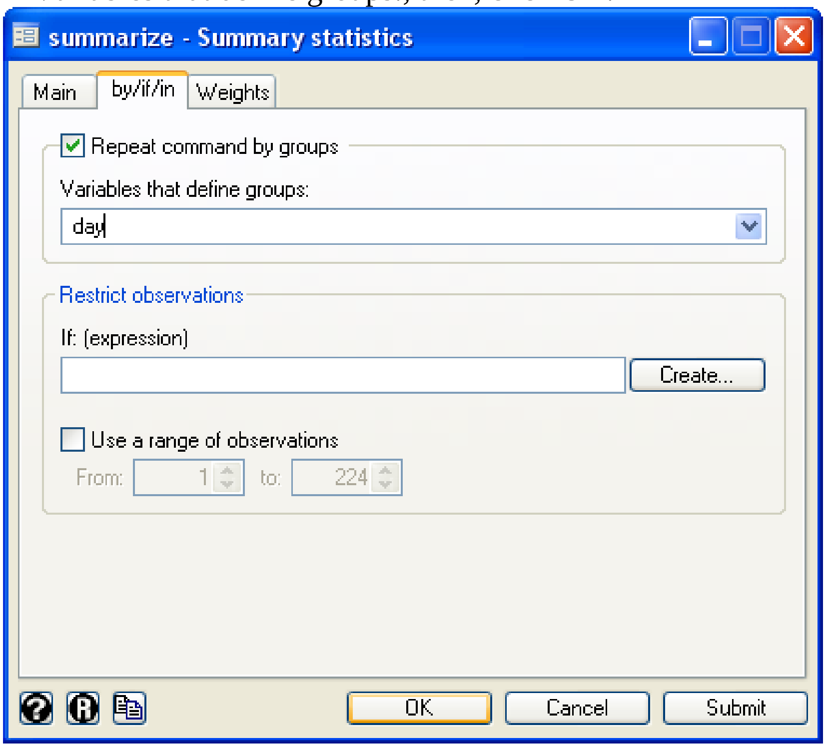
From summarize – Summary statistics window, specify the variables to be analyzed in Variables: (leave empty for all variables) box, then click OK



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.

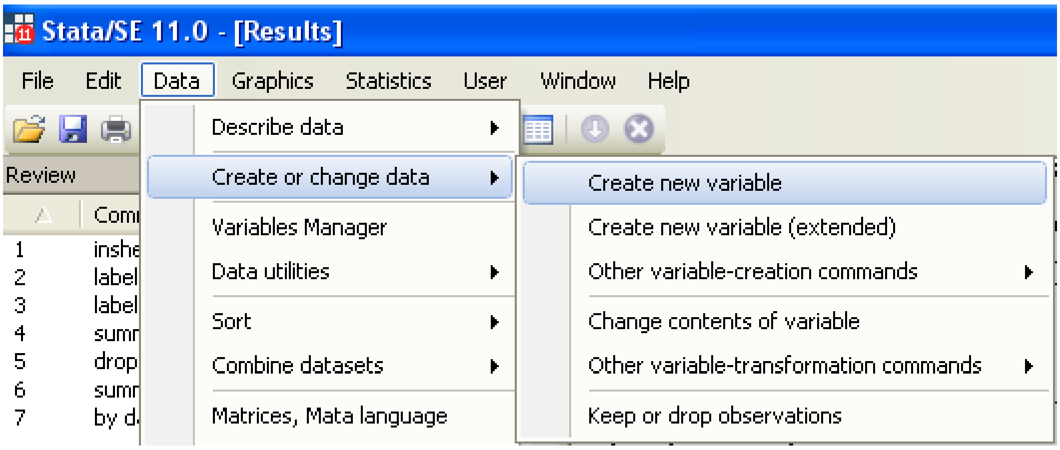


The results will be as follows:

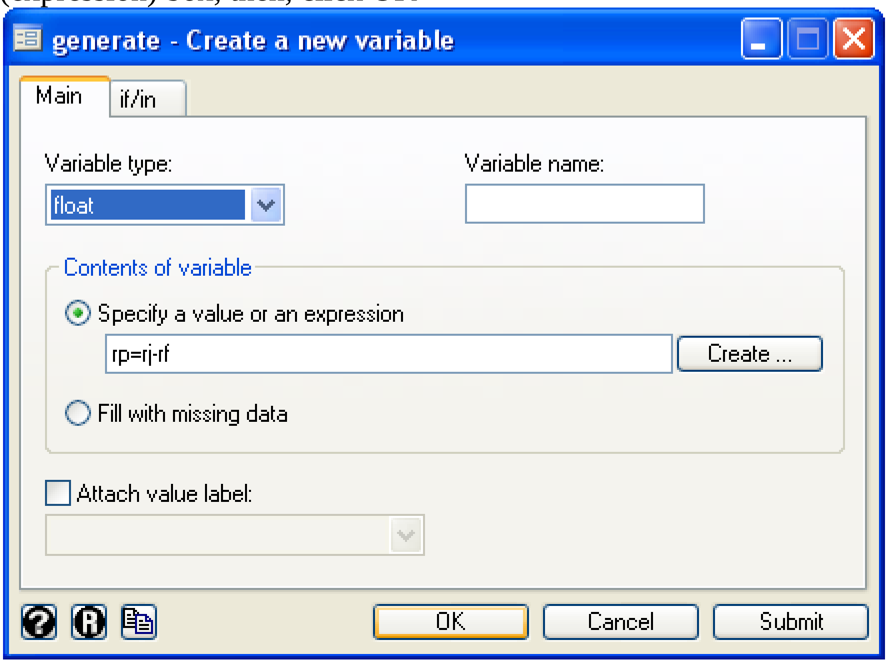
```
. by day, sort : summarize rm

-----------------------------------------------------------------------
-> 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



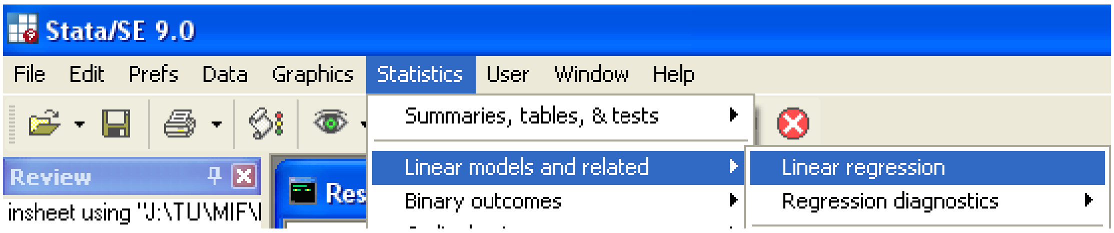
From generate – Create a new variable window, specify new variable name in New variable name: box, and expression of the new variable in Contents of new variable: (expression) box, then, click OK



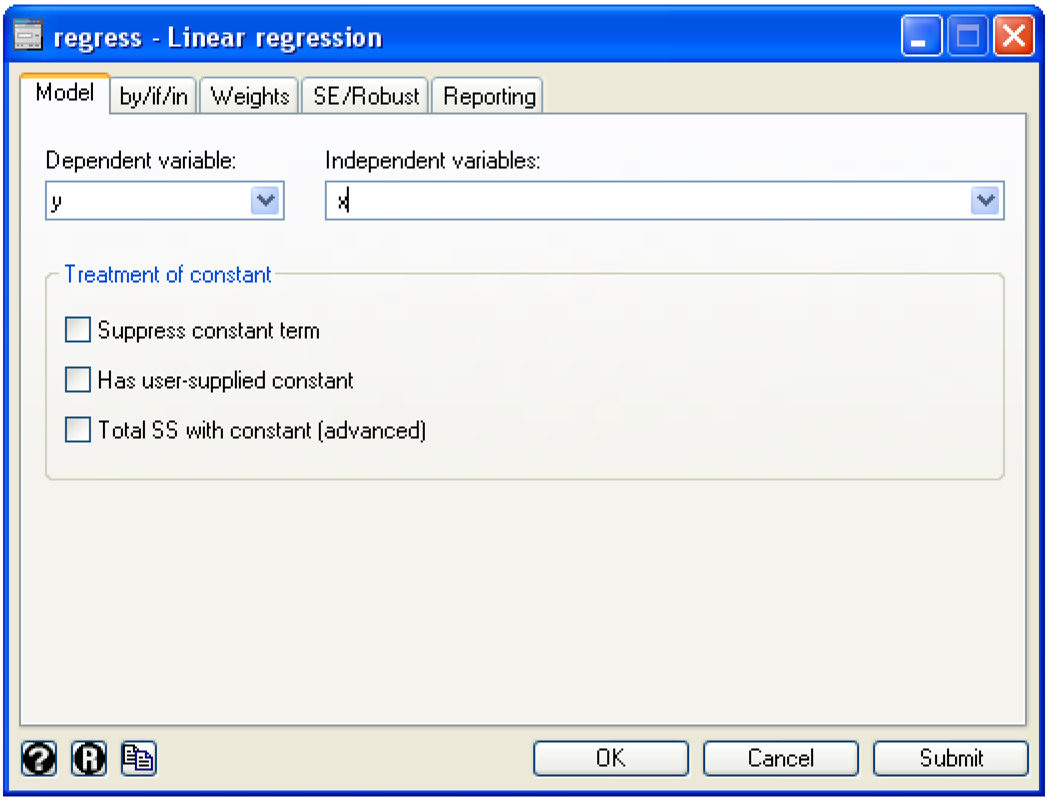
Simple Regression Model

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.



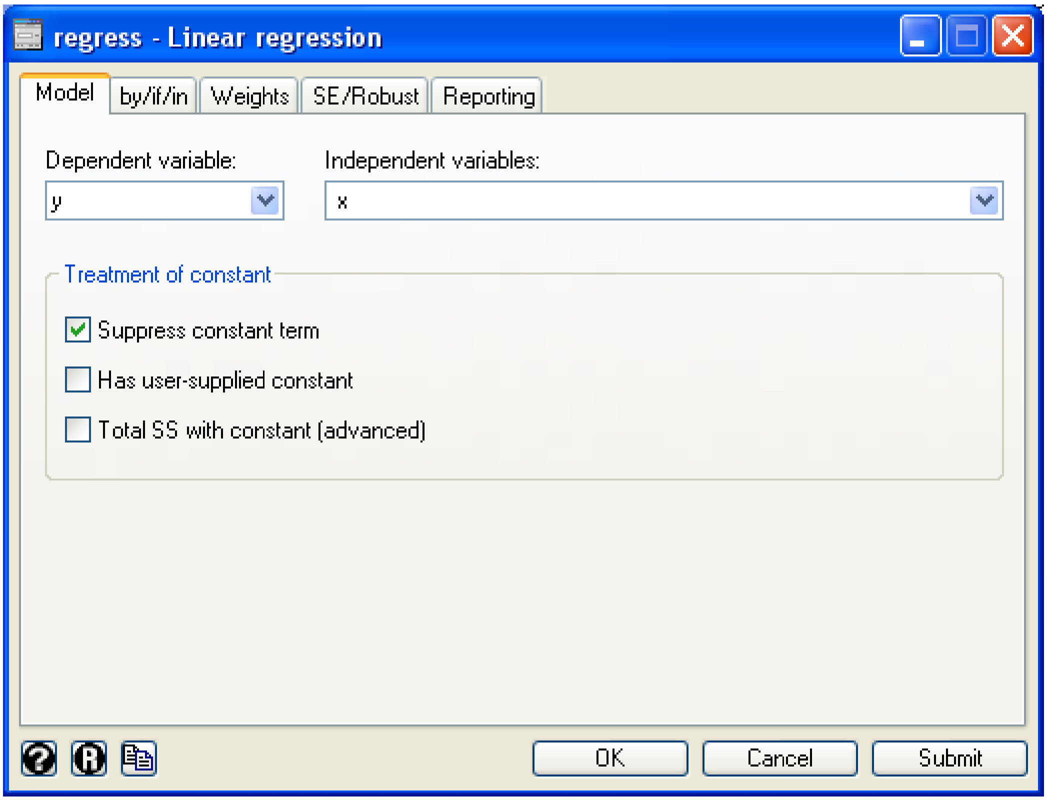
From regress – Linear regression window, specify dependent and independent variables in Dependent variable: and Independent variables: boxes, then, click OK.



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
```

<u>Multiply or Divide constant value to the variable</u>

Generate new variable by:

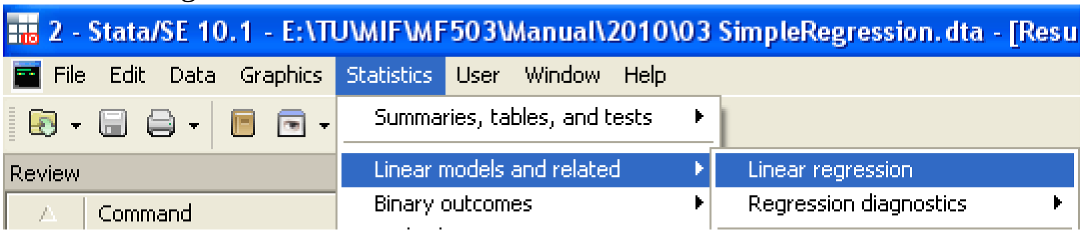
```
. g x10 = x*10

. 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
------------------------------------------------------------------------------
```

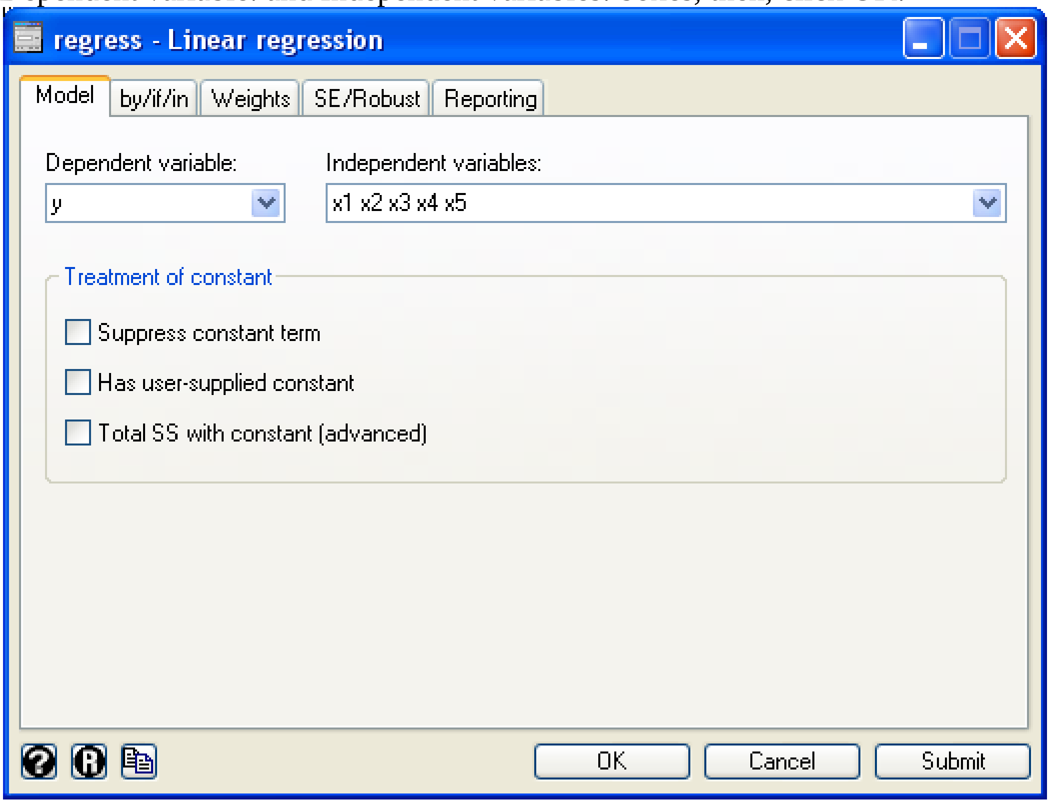
Multiple Regression Model

The model: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i$

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:

```
. reg y x1 x2 x3 x4 x5

      Source |       SS       df       MS              Number of obs =      23
-------------+------------------------------           F(  5,    17) =   57.63
       Model |  1129.30649     5  225.861299           Prob > F      =  0.0000
    Residual |   66.622243    17  3.91895547           R-squared     =  0.9443
-------------+------------------------------           Adj R-squared =  0.9279
       Total |  1195.92874    22  54.3603971           Root MSE      =  1.9796

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0048894   .0049619     0.99   0.338    -.0055794    .0153581
          x2 |  -.6518875   .1743999    -3.74   0.002    -1.019839   -.2839359
          x3 |   .2432418   .0895442     2.72   0.015       .05432    .4321636
          x4 |   .1043175   .0706436     1.48   0.158    -.0447275    .2533625
          x5 |  -.0711103   .0983811    -0.72   0.480    -.2786762    .1364556
       _cons |   38.59691   4.214487     9.16   0.000     29.70512     47.4887
------------------------------------------------------------------------------
```

Hypotheses Testing

To test whether $\beta_4 = \beta_5$.

```
. test x4 = x5

 ( 1)  x4 - x5 = 0

       F(  1,    17) =    1.27
            Prob > F =    0.2746
```

To test whether we should eliminate x_4 and x_5 out of the model, or H_0: $\beta_4 = \beta_5 = 0$.

```
. test (x4=0) (x5=0)

 ( 1)  x4 = 0
 ( 2)  x5 = 0

       F(  2,    17) =    1.17
            Prob > F =    0.3354

. regress y x1 x2 x3

      Source |       SS       df       MS              Number of obs =      23
-------------+------------------------------           F(  3,    19) =   93.65
       Model |  1120.17021     3  373.390069           Prob > F      =  0.0000
    Residual |  75.7585287    19  3.98729099           R-squared     =  0.9367
-------------+------------------------------           Adj R-squared =  0.9267
       Total |  1195.92874    22  54.3603971           Root MSE      =  1.9968

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0108762   .0023812     4.57   0.000     .0058922    .0158602
          x2 |  -.5410846   .1579697    -3.43   0.003    -.8717189   -.2104503
          x3 |   .1740546   .0625307     2.78   0.012     .0431764    .3049328
       _cons |    38.6472     3.6496    10.59   0.000      31.0085     46.2859
------------------------------------------------------------------------------
```

To test whether $\beta_2 + \beta_3 = 1$

```
. test x2+x3=1

 ( 1)  x2 + x3 = 1

       F(  1,    19) =  132.06
            Prob > F =    0.0000
```

Multicollinearity Problem

Detecting Serious Multicollinearity Problem

Model: $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$

```
. reg y x2 x3 x4 x5

      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  4,    11) =   13.89
       Model |  52249133.2     4  13062283.3           Prob > F      =  0.0003
    Residual |  10347222.8    11  940656.615           R-squared     =  0.8347
-------------+------------------------------           Adj R-squared =  0.7746
       Total |    62596356    15   4173090.4           Root MSE      =  969.87

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |  -2227.704   920.4659    -2.42   0.034    -4253.636    -201.772
          x3 |   1251.141   1157.021     1.08   0.303    -1295.445    3797.726
          x4 |   6.283002   30.62166     0.21   0.841    -61.11482    73.68083
          x5 |     -197.4   101.5612    -1.94   0.078    -420.9348    26.13479
       _cons |   10816.04    5988.35     1.81   0.098    -2364.229    23996.31
------------------------------------------------------------------------------
```

Signal of the problem – High F-test, but low t-test.

Checking Pair-wise Correlation:

```
. pwcorr y x2 x3 x4 x5

             |        y       x2       x3       x4       x5
-------------+---------------------------------------------
           y |   1.0000
          x2 |  -0.7842   1.0000
          x3 |  -0.0227   0.4725   1.0000
          x4 |  -0.4130   0.2892  -0.1044   1.0000
          x5 |  -0.8518   0.6531  -0.1273   0.5499   1.0000
```

Checking Partial Correlation:

```
. pcorr y x2 x3 x4 x5
(obs=16)

Partial correlation of y with

    Variable |    Corr.     Sig.
-------------+------------------
          x2 |  -0.5895    0.034
          x3 |   0.3100    0.303
          x4 |   0.0617    0.841
          x5 |  -0.5056    0.078
```

Checking Variance Inflation Factor (VIF):

```
. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
          x2 |      3.91    0.255901
          x5 |      3.73    0.268223
          x3 |      2.25    0.443702
          x4 |      1.46    0.686604
-------------+----------------------
    Mean VIF |      2.84
. reg x2 x3 x4 x5

      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  3,    12) =   11.63
```

```
       Model |  3.22830633     3  1.07610211           Prob > F      =  0.0007
    Residual |  1.11023708    12  .092519757           R-squared     =  0.7441
-------------+------------------------------           Adj R-squared =  0.6801
       Total |  4.33854341    15  .289236228           Root MSE      =  .30417

------------------------------------------------------------------------------
          x2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x3 |   .9306908   .2439007     3.82   0.002     .3992769    1.462105
          x4 |  -.0039507   .0095356    -0.41   0.686    -.0247269    .0168255
          x5 |   .0863716   .0198203     4.36   0.001     .0431868    .1295564
       _cons |  -.1080595   1.877798    -0.06   0.955    -4.199431    3.983312
------------------------------------------------------------------------------

. reg x3 x2 x4 x5

      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  3,    12) =    5.02
       Model |  .880977981     3  .293659327           Prob > F      =  0.0176
    Residual |  .702665775    12  .058555481           R-squared     =  0.5563
-------------+------------------------------           Adj R-squared =  0.4454
       Total |  1.58364376    15   .10557625           Root MSE      =  .24198

------------------------------------------------------------------------------
          x3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |   .5890315    .154364     3.82   0.002     .2527012    .9253618
          x4 |   .0015943   .0076262     0.21   0.838    -.0150217    .0182103
          x5 |  -.0539669   .0199846    -2.70   0.019    -.0975095   -.0104243
       _cons |   1.772722   1.403714     1.26   0.231    -1.285707    4.831152
------------------------------------------------------------------------------

. reg x4 x2 x3 x5

      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  3,    12) =    1.83
       Model |  457.888462     3  152.629487           Prob > F      =  0.1961
    Residual |  1003.16777    12  83.5973141           R-squared     =  0.3134
-------------+------------------------------           Adj R-squared =  0.1417
       Total |  1461.05623    15  97.4037487           Root MSE      =  9.1432

------------------------------------------------------------------------------
          x4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |   -3.56971   8.615968    -0.41   0.686    -22.34229    15.20287
          x3 |   2.276104    10.8876     0.21   0.838    -21.44595    25.99816
          x5 |    1.42305   .8648254     1.65   0.126    -.4612425    3.307343
       _cons |   171.7178   27.01301     6.36   0.000     112.8615    230.5741
------------------------------------------------------------------------------

. reg x5 x2 x3 x4

      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  3,    12) =   10.91
       Model |   248.80415     3  82.9347167           Prob > F      =  0.0010
    Residual |  91.1958498    12  7.59965415           R-squared     =  0.7318
-------------+------------------------------           Adj R-squared =  0.6647
       Total |         340    15  22.6666667           Root MSE      =  2.7567

------------------------------------------------------------------------------
          x5 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |    7.09464    1.62806     4.36   0.001     3.547402    10.64188
          x3 |  -7.004123   2.593707    -2.70   0.019    -12.65532   -1.352921
          x4 |   .1293665   .0786194     1.65   0.126    -.0419306    .3006635
       _cons |  -12.85998   16.61137    -0.77   0.454    -49.05304    23.33308
------------------------------------------------------------------------------
```

Solving Multicollinearity Problem by Dropping Variable

Before dropping variables, first, testing whether coefficients of those variables equals to zero.

```
. test x3 x4

 ( 1)  x3 = 0
 ( 2)  x4 = 0

       F(  2,    11) =    0.62
            Prob > F =    0.5551

. reg y x2 x5

      Source |       SS       df       MS              Number of obs =      16
-------------+------------------------------           F(  2,    13) =   28.83
       Model |  51080225.8     2  25540112.9           Prob > F      =  0.0000
    Residual |  11516130.2    13  885856.167           R-squared     =  0.8160
-------------+------------------------------           Adj R-squared =  0.7877
       Total |    62596356    15   4173090.4           Root MSE      =   941.2

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |  -1509.261   596.7261    -2.53   0.025    -2798.409   -220.1125
          x5 |  -254.1182   67.40739    -3.77   0.002     -399.743   -108.4934
       _cons |   14494.09   1559.877     9.29   0.000     11124.18       17864
------------------------------------------------------------------------------
```

Check by using Information Criteria:

Model with x_2 x_3 x_4 x_5:

```
. estat ic

------------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+----------------------------------------------------------------
           . |     16     -144.14   -129.7401      5     269.4803    273.3432
------------------------------------------------------------------------------
```

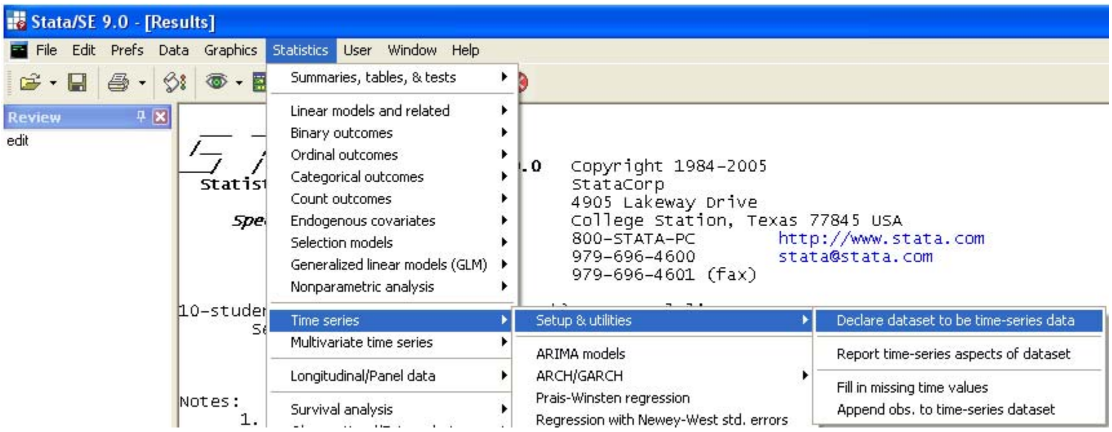
Model with x_2 x_5:

```
. estat ic

------------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+----------------------------------------------------------------
           . |     16     -144.14   -130.5964      3     267.1928    269.5105
------------------------------------------------------------------------------
```

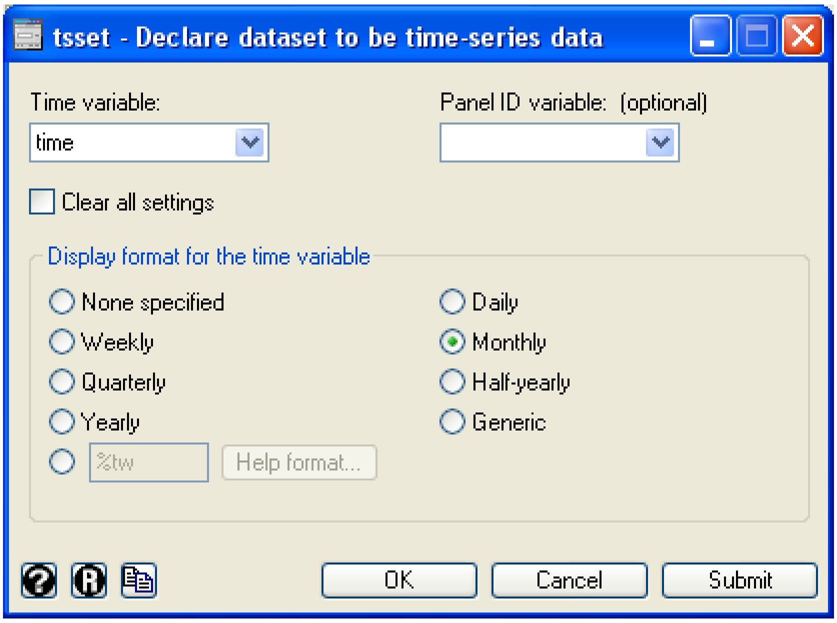
<u>Autocorrelation Problem</u>

Setting up data to be Time Series data

From menu bar, go to Statistics, choose Time series, select Setup & utilities, choose Declare dataset to be time-series data.



From tsset – Declare dataset to be time-series data window, specify variable determine time (in this case – time) in Time variable: box, and select time frequency in Display format for the time variable (in this case Monthly), then, click OK.




```
. tsset time, monthly
        time variable:  time, 1978m1 to 1987m12
```

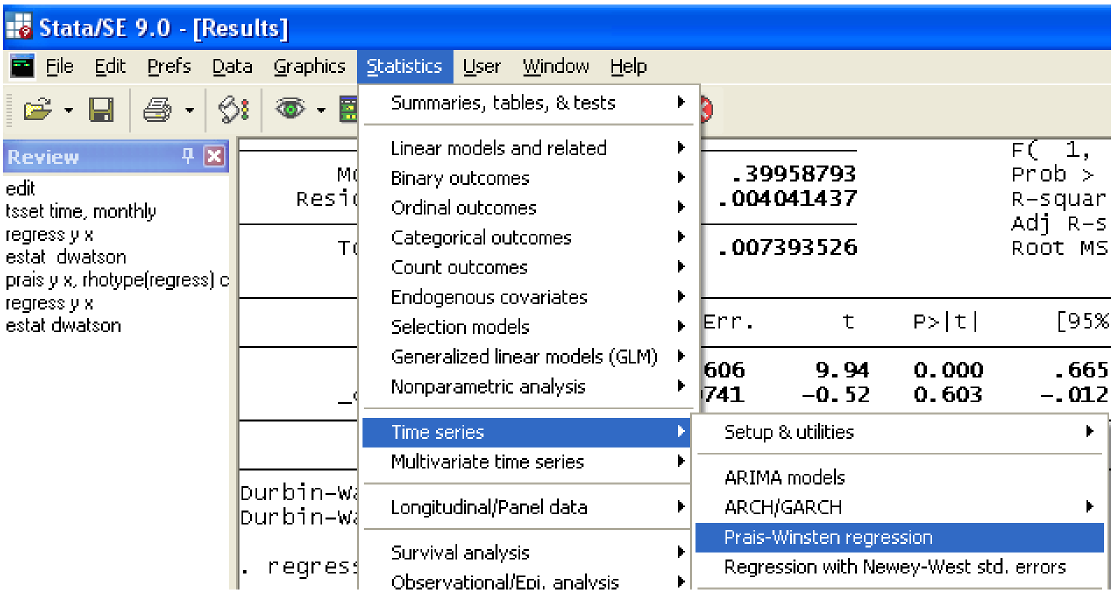
Detecting Autocorrelation Problem

Model: $y_t = \beta_0 + \beta_1 x_t + u_t$

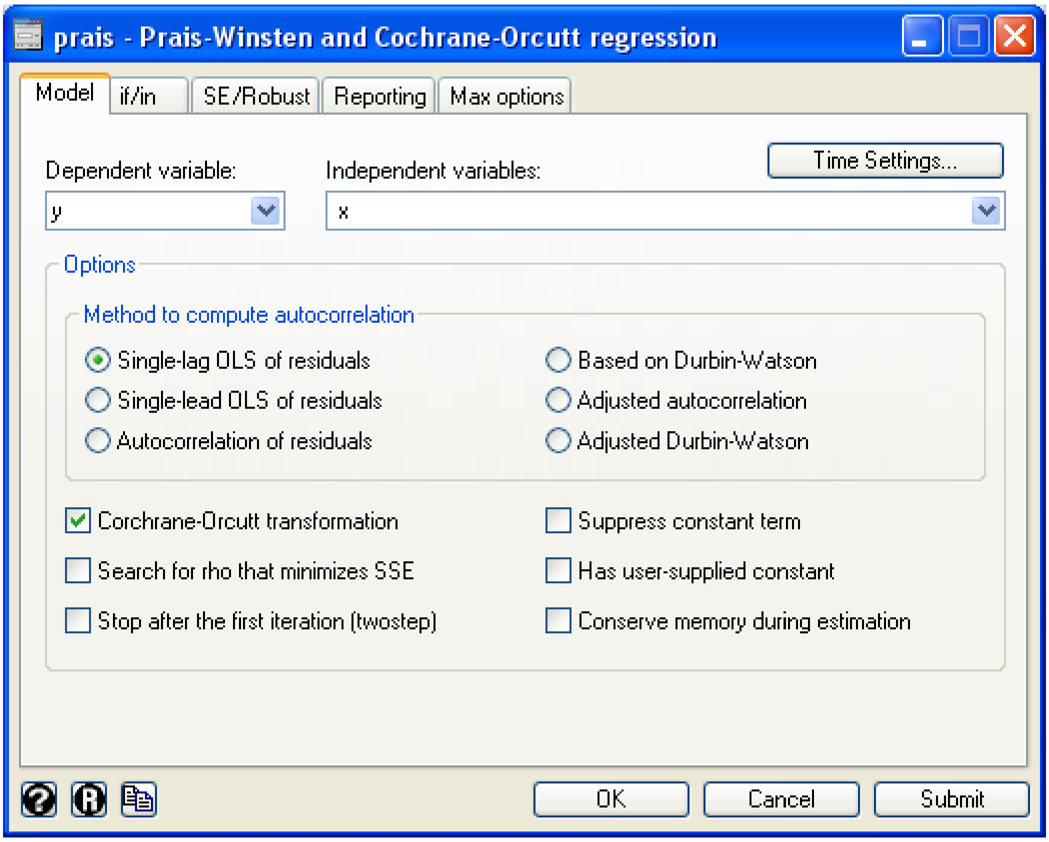
```
. reg y x

      Source |       SS       df       MS              Number of obs =     120
-------------+------------------------------           F(  1,    118) = 2609.35
       Model |  7284.31095     1  7284.31095           Prob > F      =  0.0000
    Residual |  329.411078   118  2.79161931           R-squared     =  0.9567
-------------+------------------------------           Adj R-squared =  0.9564
       Total |  7613.72203   119  63.9808574           Root MSE      =  1.6708

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .7370403   .0144286    51.08   0.000     .7084677    .7656129
       _cons |   4.542808   1.451023     3.13   0.002     1.669388    7.416228
------------------------------------------------------------------------------

. estat dwatson

Durbin-Watson d-statistic(  2,   120) =  3.424407
```

Solving Problem using Cochrane-Orcutt Technique

To solve problem using Cochrane-Orcutt technique, from menu bar, go to Statistics, choose Time series, select Prais-Winsten regression.



From prais – Prais-Winsten and Cochrane-Orcutt regression window, specify dependent and independent variables in Dependent variable: and Independent variables: boxes, and check on Corchrane-Orcutt transformation box, then, click OK.




```
. prais y x, rhotype(reg) corc

Iteration 0:  rho = 0.0000
Iteration 1:  rho = -0.7568
Iteration 2:  rho = -0.7582
Iteration 3:  rho = -0.7582

Cochrane-Orcutt AR(1) regression -- iterated estimates

      Source |       SS       df       MS              Number of obs =     119
-------------+------------------------------           F(  1,   117) = 8265.39
       Model |  10297.8332     1  10297.8332           Prob > F      =  0.0000
    Residual |  145.769995   117  1.24589739           R-squared     =  0.9860
-------------+------------------------------           Adj R-squared =  0.9859
       Total |  10443.6032   118  88.5051115           Root MSE      =  1.1162

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .7415161   .0081562    90.91   0.000     .7253631     .757669
       _cons |   4.112478    .816992     5.03   0.000     2.494469    5.730488
-------------+----------------------------------------------------------------
         rho |  -.7582272
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    3.424407
Durbin-Watson statistic (transformed) 2.061595
```

Cochrane-Orcutt Iterative Process

```
. reg y x

      Source |       SS       df       MS              Number of obs =     120
-------------+------------------------------           F(  1,   118) = 2609.35
       Model |  7284.31095     1  7284.31095           Prob > F      =  0.0000
    Residual |  329.411078   118  2.79161931           R-squared     =  0.9567
-------------+------------------------------           Adj R-squared =  0.9564
       Total |  7613.72203   119  63.9808574           Root MSE      =  1.6708

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .7370403   .0144286    51.08   0.000     .7084677    .7656129
       _cons |   4.542808   1.451023     3.13   0.002     1.669388    7.416228
------------------------------------------------------------------------------

. predict uhat1, resid

. reg uhat1 l.uhat1, noconst

      Source |       SS       df       MS              Number of obs =     119
-------------+------------------------------           F(  1,   118) =  147.01
       Model |  182.207794     1  182.207794           Prob > F      =  0.0000
    Residual |  146.249169   118  1.23939974           R-squared     =  0.5547
-------------+------------------------------           Adj R-squared =  0.5510
       Total |  328.456962   119  2.76014254           Root MSE      =  1.1133
------------------------------------------------------------------------------
       uhat1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       uhat1 |
         L1. |  -.7568481    .062421   -12.12   0.000    -.8804588   -.6332375
------------------------------------------------------------------------------

. g ys_1=y-(-0.7568481)*l.y
(1 missing value generated)

. g xs_1=x-(-0.7568481)*l.x
(1 missing value generated)

. g x0_1=1-(-0.7568481)

. reg ys_1 xs_1 x0_1, noconst

      Source |       SS       df       MS              Number of obs =     119
-------------+------------------------------           F(  2,   117) =       .
       Model |  2256381.42     2  1128190.71           Prob > F      =  0.0000
    Residual |  145.770628   117  1.24590281           R-squared     =  0.9999
-------------+------------------------------           Adj R-squared =  0.9999
       Total |  2256527.19   119  18962.4133           Root MSE      =  1.1162
------------------------------------------------------------------------------
        ys_1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        xs_1 |   .7415133   .0081615    90.86   0.000       .72535    .7576766
        x0_1 |   4.112747   .8175166     5.03   0.000     2.493698    5.731796
------------------------------------------------------------------------------

. g uhat2=y-(4.112747+0.7415133*x)

. reg uhat2 l.uhat2, noconst

      Source |       SS       df       MS              Number of obs =     119
-------------+------------------------------           F(  1,   118) =  148.00
       Model |  182.833426     1  182.833426           Prob > F      =  0.0000
    Residual |  145.769997   118  1.23533896           R-squared     =  0.5564
-------------+------------------------------           Adj R-squared =  0.5526
       Total |  328.603423   119   2.7613733           Root MSE      =  1.1115
------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       uhat2 |
         L1. |  -.7582268   .0623253   -12.17   0.000    -.8816478   -.6348057
------------------------------------------------------------------------------

. g ys_2=y-(-0.7582268)*l.y
```

```
(1 missing value generated)

. g xs_2=x-(-0.7582268)*l.x
(1 missing value generated)

. g x0_2=1-(-0.7582268)

. reg ys_2 xs_2 x0_2, noconst

      Source |       SS       df       MS              Number of obs =     119
-------------+------------------------------           F(  2,   117) =       .
       Model |  2259922.59     2   1129961.3           Prob > F      =  0.0000
    Residual |  145.770127   117  1.24589852           R-squared     =  0.9999
-------------+------------------------------           Adj R-squared =  0.9999
       Total |  2260068.36   119  18992.1711           Root MSE      =  1.1162

------------------------------------------------------------------------------
        ys_2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        xs_2 |    .741516   .0081562    90.91   0.000     .7253631     .757669
        x0_2 |   4.112483   .8169925     5.03   0.000     2.494472    5.730493
------------------------------------------------------------------------------

. g uhat3=y-(4.112483+0.741516*x)

. reg uhat3 l.uhat3, noconst

      Source |       SS       df       MS              Number of obs =     119
-------------+------------------------------           F(  1,   118) =  148.00
       Model |  182.833685     1  182.833685           Prob > F      =  0.0000
    Residual |  145.769993   118  1.23533892           R-squared     =  0.5564
-------------+------------------------------           Adj R-squared =  0.5526
       Total |  328.603678   119  2.76137545           Root MSE      =  1.1115

------------------------------------------------------------------------------
       uhat3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       uhat3 |
         L1. |  -.7582271   .0623253   -12.17   0.000    -.8816482   -.6348061
------------------------------------------------------------------------------

. g ys_3=y-(-0.7582271)*l.y
(1 missing value generated)

. g xs_3=x-(-0.7582271)*l.x
(1 missing value generated)

. g x0_3=1-(-0.7582271)

. reg ys_3 xs_3 x0_3, noconst

      Source |       SS       df       MS              Number of obs =     119
-------------+------------------------------           F(  2,   117) =       .
       Model |   2259923.4     2   1129961.7           Prob > F      =  0.0000
    Residual |  145.770195   117   1.2458991           R-squared     =  0.9999
-------------+------------------------------           Adj R-squared =  0.9999
       Total |  2260069.17   119  18992.1779           Root MSE      =  1.1162

------------------------------------------------------------------------------
        ys_3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        xs_3 |    .741516   .0081562    90.91   0.000     .7253631     .757669
        x0_3 |   4.112484   .8169926     5.03   0.000     2.494473    5.730495
------------------------------------------------------------------------------

. prais y x, rhotype(reg) corc

Iteration 0:  rho = 0.0000
Iteration 1:  rho = -0.7568
Iteration 2:  rho = -0.7582
Iteration 3:  rho = -0.7582

Cochrane-Orcutt AR(1) regression -- iterated estimates

      Source |       SS       df       MS              Number of obs =     119
-------------+------------------------------           F(  1,   117) = 8265.39
       Model |  10297.8332     1  10297.8332           Prob > F      =  0.0000
```

```
    Residual |  145.769995   117  1.24589739           R-squared     =  0.9860
-------------+------------------------------           Adj R-squared =  0.9859
       Total |  10443.6032   118  88.5051115           Root MSE      =  1.1162

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .7415161   .0081562    90.91   0.000     .7253631     .757669
       _cons |   4.112478    .816992     5.03   0.000     2.494469    5.730488
-------------+----------------------------------------------------------------
         rho |  -.7582272
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    3.424407
Durbin-Watson statistic (transformed) 2.061595
```

Heteroscedasticity Problem

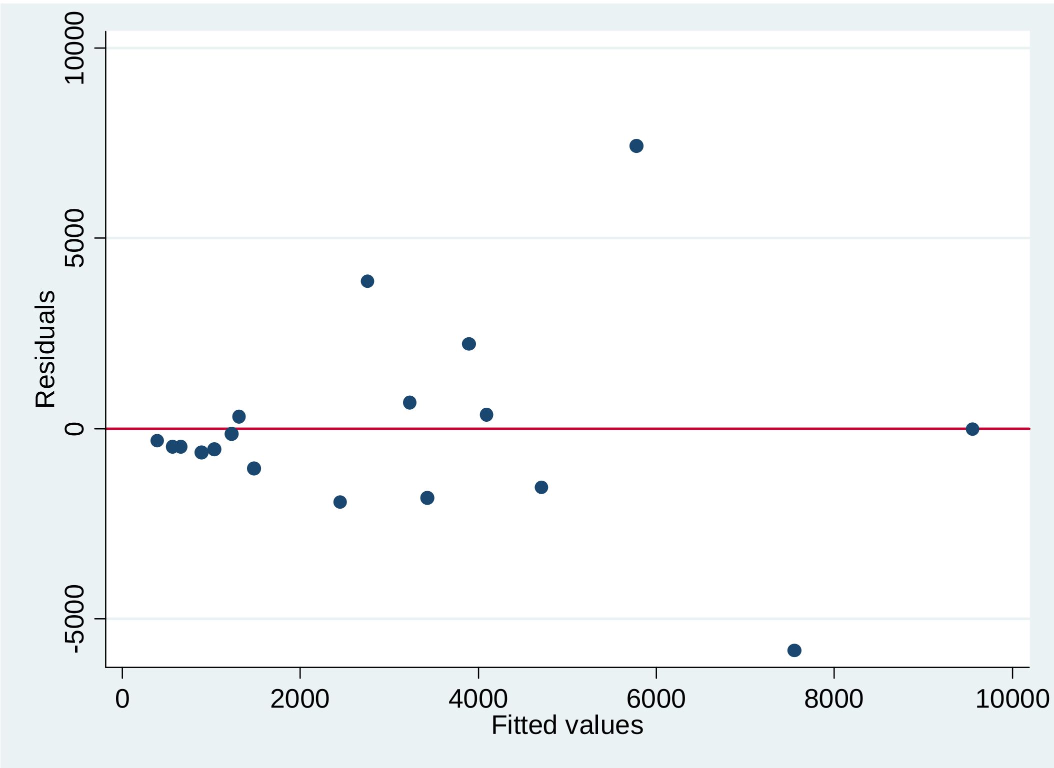
Detecting Heteroscedasticity Problem

Model: $RD = \beta_0 + \beta_1 SALE + u$

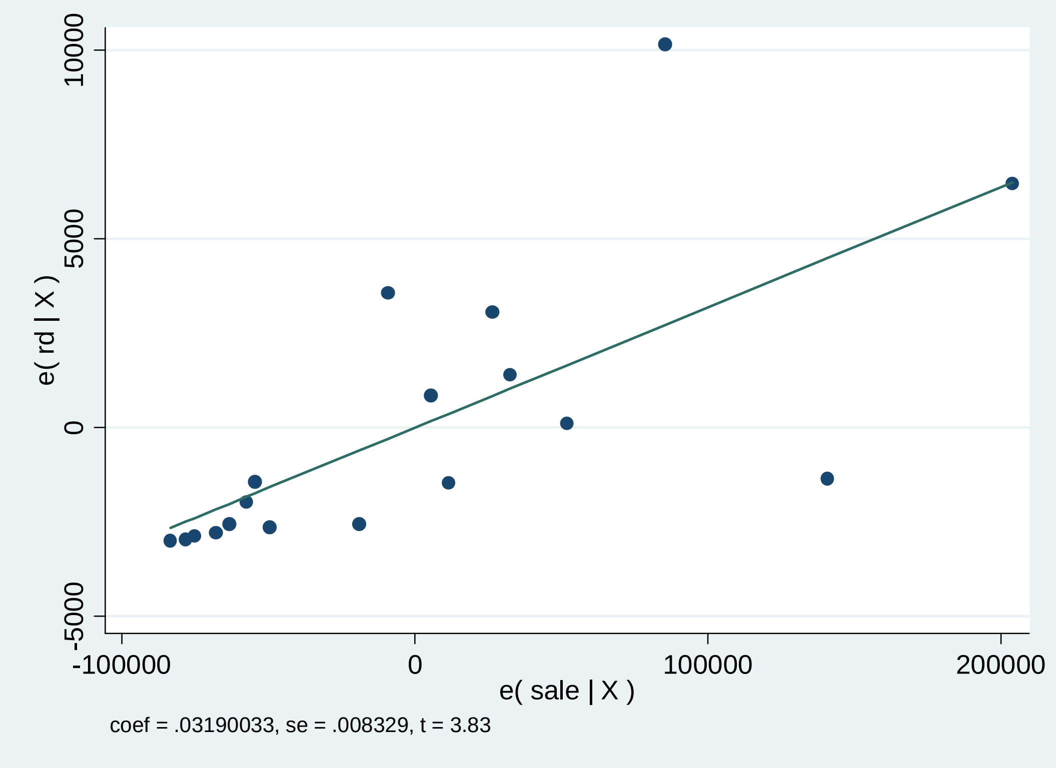
```
. reg rd sale

      Source |       SS       df       MS              Number of obs =      18
-------------+------------------------------           F(  1,    16) =   14.67
       Model |   111675212     1   111675212           Prob > F      =  0.0015
    Residual |   121806834    16  7612927.12           R-squared     =  0.4783
-------------+------------------------------           Adj R-squared =  0.4457
       Total |   233482046    17    13734238           Root MSE      =  2759.2

------------------------------------------------------------------------------
          rd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   .0319003    .008329     3.83   0.001     .0142436     .049557
       _cons |   192.9932   990.9858     0.19   0.848    -1907.803    2293.789
------------------------------------------------------------------------------
```

Informal Test:

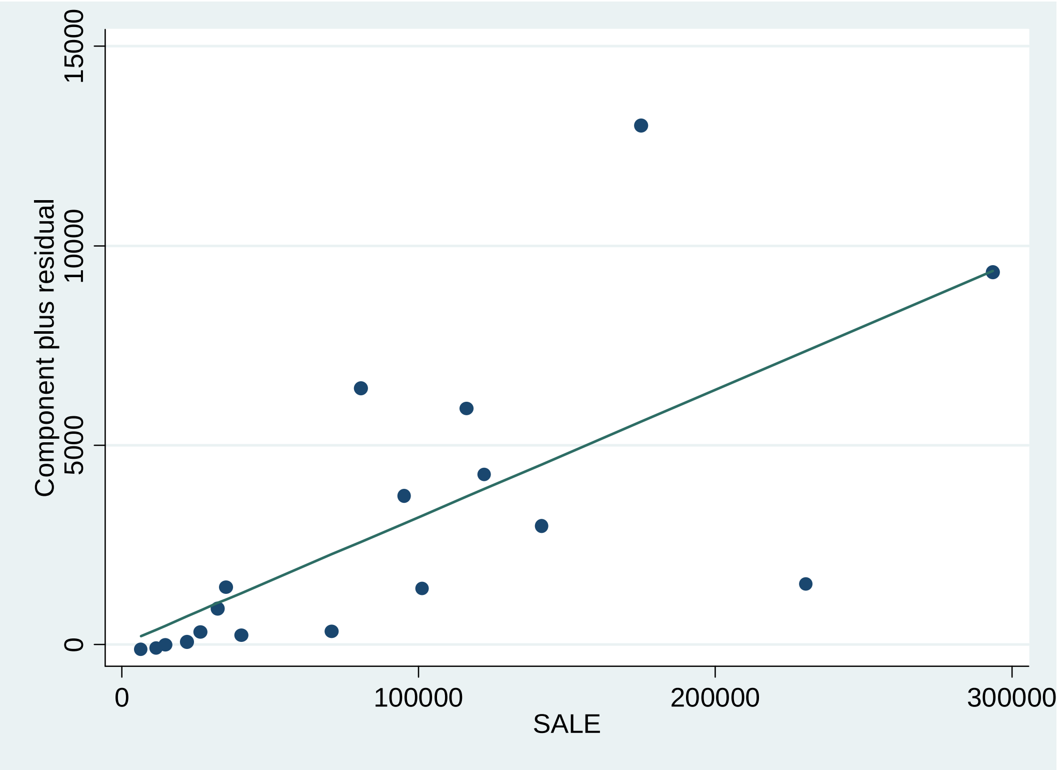
```
. rvfplot, yline(0)
```



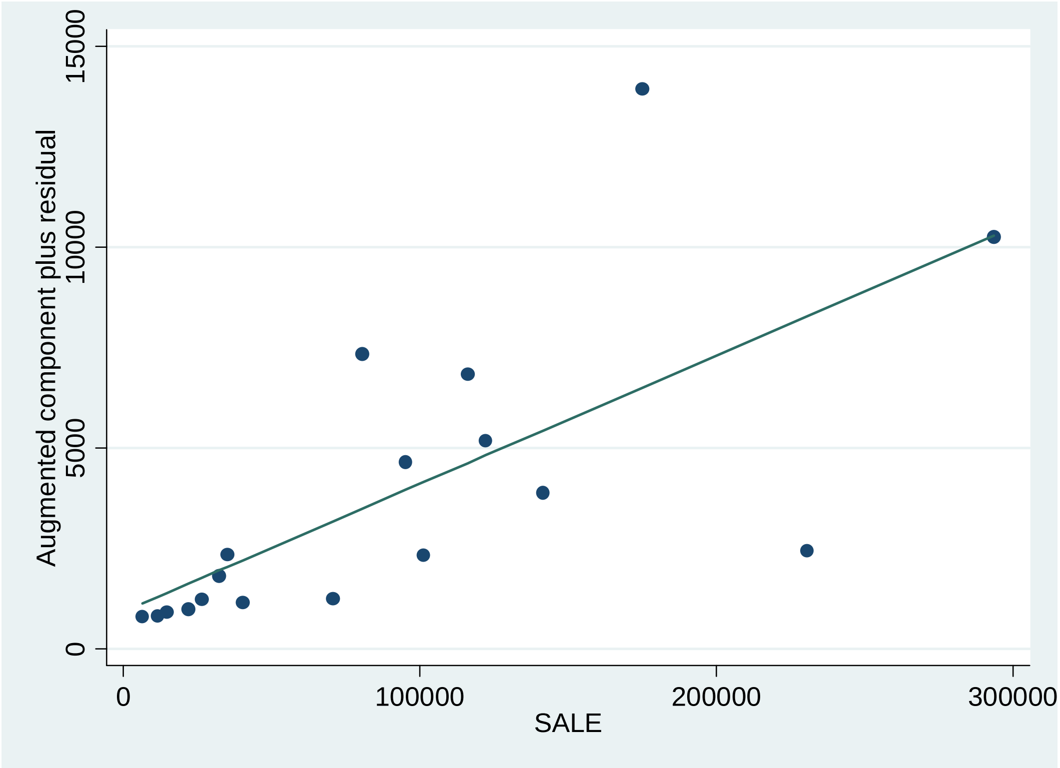
```
. avplot sale, recast(scatter)
```



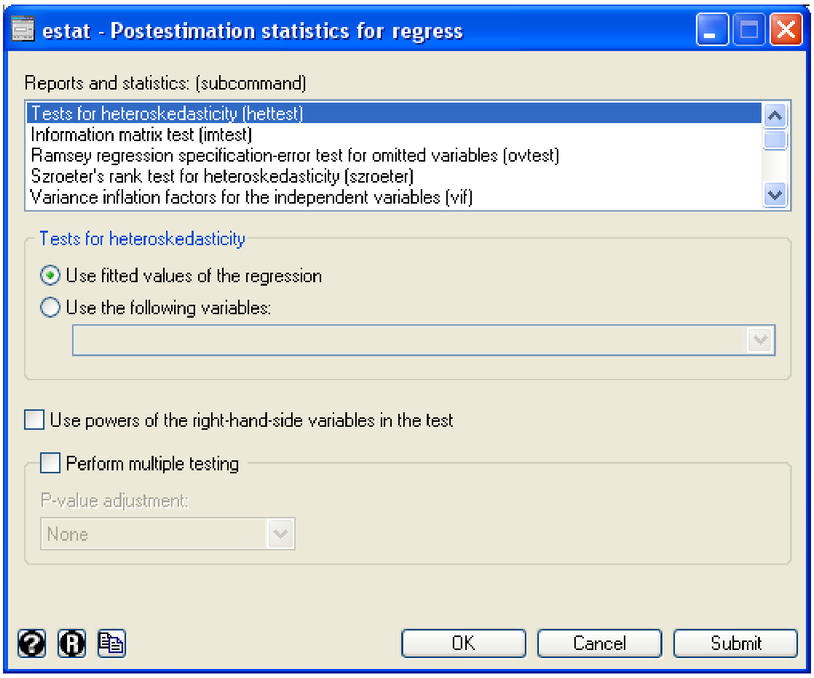

```
. cprplot sale, recast(scatter)
```



```
. acprplot sale, recast(scatter)
```



Formal Test:



Breusch-Pagan Test:

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of rd

         chi2(1)      =     8.91
         Prob > chi2  =   0.0028
```

White's Test:

```
. whitetst

White's general test statistic :  5.212492  Chi-sq( 2)  P-value =  .0738
```

Information Matrix Test (White's test):

```
. estat imtest

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |       5.21      2    0.0738
            Skewness |       0.94      1    0.3315
            Kurtosis |       3.59      1    0.0582
---------------------+-----------------------------
               Total |       9.74      4    0.0450
---------------------------------------------------
```

Szroteter's Test:

```
. estat szroeter sale

Szroeter's test for homoskedasticity

    Ho: variance constant
    Ha: variance monotonic in sale

         chi2(1)      =     9.56
         Prob > chi2  =   0.0020
```

Solving Heteroscedasticity Problem

From OLS Estimated Results:

```
. reg rd sale

      Source |       SS       df       MS              Number of obs =      18
-------------+------------------------------           F(  1,    16) =   14.67
       Model |   111675212     1   111675212           Prob > F      =  0.0015
    Residual |   121806834    16  7612927.12           R-squared     =  0.4783
-------------+------------------------------           Adj R-squared =  0.4457
       Total |   233482046    17    13734238           Root MSE      =  2759.2

------------------------------------------------------------------------------
          rd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   .0319003    .008329     3.83   0.001     .0142436     .049557
       _cons |   192.9932   990.9858     0.19   0.848    -1907.803    2293.789
------------------------------------------------------------------------------
```

Determine Variance equation

Functional form: $\hat{u}_i^2 \approx \sigma_i^2 = \alpha_1 + \alpha_2 SALE_i + v_i$

```
. predict uhat, resid

. g uhat2=uhat^2

. reg uhat2 sale

      Source |       SS       df       MS              Number of obs =      18
-------------+------------------------------           F(  1,    16) =    4.56
       Model |  8.1603e+14     1  8.1603e+14           Prob > F      =  0.0484
    Residual |  2.8605e+15    16  1.7878e+14           R-squared     =  0.2220
-------------+------------------------------           Adj R-squared =  0.1733
       Total |  3.6765e+15    17  2.1627e+14           Root MSE      =  1.3e+07

------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   86.23211   40.36254     2.14   0.048     .6673552    171.7969
       _cons |  -974468.8    4802344    -0.20   0.842    -1.12e+07     9206045
------------------------------------------------------------------------------
```

Weighted Least Squares:
Remedy by using Weighted Least Squares
Functional form: $w \cdot RD = \beta_1 w + \beta_2 w \cdot SALE + w \cdot u$

where $w = \dfrac{1}{\hat{\sigma}_i}$

```
. predict sigma2, xb

. g sigmahat=sqrt(sigma2)

. g w=1/sigmahat

. g wrd=w*rd

. g wsale=w*sale

. reg wrd wsale w, noconst

      Source |       SS       df       MS              Number of obs =      17
-------------+------------------------------           F(  2,    15) =   17.24
       Model |  24.9762465     2  12.4881233           Prob > F      =  0.0001
    Residual |   10.868054    15  .724536932           R-squared     =  0.6968
-------------+------------------------------           Adj R-squared =  0.6564
       Total |  35.8443005    17  2.10848827           Root MSE      =   .8512

------------------------------------------------------------------------------
         wrd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       wsale |   .0378137   .0068275     5.54   0.000     .0232613    .0523661
           w |   -347.254   168.4375    -2.06   0.057    -706.2701    11.76205
------------------------------------------------------------------------------
```

Using command Weighted Least Squares

```
. vwls rd sale, sd(sigmahat)

Variance-weighted least-squares regression           Number of obs   =      17
Goodness-of-fit chi2(15)   =   10.87                 Model chi2(1)   =   22.22
Prob > chi2                =  0.7619                 Prob > chi2     =  0.0000
------------------------------------------------------------------------------
          rd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   .0378137    .008021     4.71   0.000     .0220928    .0535346
       _cons |   -347.254    197.883    -1.75   0.079    -735.0977    40.58962
------------------------------------------------------------------------------
```

Relieve the Problem: White's Robust Standard Error

From OLS Estimated Results:

```
. reg rd sale

      Source |       SS       df       MS              Number of obs =      18
-------------+------------------------------           F(  1,    16) =   14.67
       Model |   111675212     1   111675212           Prob > F      =  0.0015
    Residual |   121806834    16  7612927.12           R-squared     =  0.4783
-------------+------------------------------           Adj R-squared =  0.4457
       Total |   233482046    17    13734238           Root MSE      =  2759.2

------------------------------------------------------------------------------
          rd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   .0319003    .008329     3.83   0.001     .0142436     .049557
       _cons |   192.9932   990.9858     0.19   0.848    -1907.803    2293.789
------------------------------------------------------------------------------
```

OLS with White's Robust Standard Error Estimated Results:

```
. reg rd sale, vce(robust)

Linear regression                                      Number of obs =      18
                                                       F(  1,    16) =    9.88
                                                       Prob > F      =  0.0063
                                                       R-squared     =  0.4783
                                                       Root MSE      =  2759.2

------------------------------------------------------------------------------
             |               Robust
          rd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sale |   .0319003    .010147     3.14   0.006     .0103896     .053411
       _cons |   192.9932   533.9317     0.36   0.722    -938.8916    1324.878
------------------------------------------------------------------------------
```