Using Stata
(notes adapted from Prof. Coulson's homeworks)

1. Click the *Start* Menu. Find STATA 8.0. It may be listed as INTERCOOLED STATA, and it may be listed under a "Statistical Analysis" directory. Click accordingly.

2. STATA should start up and display four windows and a drop-down menu. These four windows are:

    a. the lower right window is the command window. It is where you enter STATA commands, and it must be an active window for you to do that.

    b. the upper right window displays the commands and the output resulting from those commands.

    c. the upper left window displays the past history of commands. If you ever want to repeat a command, just click on it in this window. You can also scroll through commands in the command window by pressing the PgUp and PgDn keys.

    d. the lower left window is the list of variable names that you have loaded into STATA or have created internally. It should be empty now.

3. Under normal circumstances, you will want to create a log of your work, for later printout. Do this by entering the command

    log using *logfilename*

    where *logfilename* is a name you choose. STATA will create a file **logfilename.smcl** that will record everything that goes on in the output window. (*Note: You may have to specify a drive and/or directory name that you have writing permission for.*)

6. You must load data into STATA. For this demo, there is a prepared a STATA-ready file called Growth.dta. All STATA ready data files have the extension .dta To get the data for the next homework,

    go to: www.aw-bc.com/stock_watson
    choose: *View companion web site*, then
    Student resources, then
    Data for empirical exercises

    Select the relevant data set and download it. You can open a STATA session by clicking on it; alternatively you can save it to a file and open it from STATA by typing

    use "*filename***", *clear***

7. The variables window will now list six variables that are in the data set. One thing that we always do label the variables, not as Y or X1 or X2, but with easy mnemonics. Details on the variables in this data set can be obtained from the same web page you obtained the data from in the file Growth_description.pdf . That file contains the following information:

Documentation for Growth Data

Growth contains data on average growth rates over 1960-1995 for 65 countries, along with variables that are potentially related to growth. These data were provided by Professor Ross Levine of Brown University and were used in his paper with, Thorsten Beck and Norman Loayza "Finance and the Sources of Growth" *Journal of Financial Economics*, 2000, Vol. 58, pp. 261 -300.

Variable Definitions

Country_name Name of country

growth	Average annual percentage growth of real Gross Domestic Product (GDP)* from 1960 to 1995.
rgdp60	The value of GDP* per capita in 1960, converted to 1960 US dollars
tradehare	The average share of trade in the economy from 1960 to 1995, measured as the sum of exports plus imports, divided by GDP; that is, the average value of $(X + M)/GDP$ from 1960 to 1995, where X = exports and M = imports (both X and M are positive).
yearsshcool	Average number of years of schooling of adult residents in that country in 1960
rev_coups	Average annual number of revolutions, insurrections (successful or not) and coup d'etats in that country from 1960 to 1995
assasinations	Average annual number of political assassinations in that country from 1960 to 1995 (per million population)
oil	= 1 if oil accounted for at least half of exports in 1960 = 0 otherwise

To get a rough idea of the nature of the data, you can have STATA display some sample statistics. Type

summarize

To see the individual observations, click on the "**data**" tab at the top of the screen and scroll down to "**data browser**."

Whenever you are using new data, check for anomalies. Do you see any here?

You can also construct scatterplots for pairs of variables. Suppose you are interested in the relationship between trade shares and growth. Types

scatter growth tradeshare

8. Now you can do a regression. Suppose we want to look at the effect of openness on growth. We would estimate the following model:

$$growth = \beta_0 + \beta_1 tradeshare + u$$

We can do that very easily in STATA typing the command 'regress' followed by the name of the dependent variable and then the list of independent variables. In this case type the command:

regress growth tradeshare

and lots of stuff will appear:

```
      Source |       SS       df       MS              Number of obs =      65
-------------+------------------------------           F(  1,    63) =    8.89
       Model |  28.4885066     1  28.4885066           Prob > F      =  0.0041
    Residual |  201.851551    63  3.20399287           R-squared     =  0.1237
-------------+------------------------------           Adj R-squared =  0.1098
       Total |  230.340057    64  3.59906339           Root MSE      =    1.79

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  tradeshare |   2.306434    .773485     2.98   0.004     .7607473     3.85212
       _cons |   .6402653   .4899767     1.31   0.196    -.3388749    1.619405
------------------------------------------------------------------------------
```

How do we interpret this printout? Note the coefficients, the square roots of their estimated variances, the R^2, and the breakdown of the total sum of squares. (The rest we'll consider later.)

To see a scatterplot with the regression line superimposed, type

twoway (scatter growth tradeshare**) (lfit** growth tradeshare**)**

If we want to exclude the outlier, we can add the conditioning phrase: if tradeshare<1.5:

```
. regress growth tradeshare if tradeshare < 1.5

      Source |       SS       df       MS              Number of obs =      64
-------------+------------------------------           F(  1,    62) =    2.90
       Model |  9.28031557     1  9.28031557           Prob > F      =  0.0937
    Residual |  198.527844    62  3.20206201           R-squared     =  0.0447
-------------+------------------------------           Adj R-squared =  0.0292
       Total |   207.80816    63  3.29854222           Root MSE      =  1.7894

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  tradeshare |   1.680905   .9873624     1.70   0.094    -.2928046    3.654614
       _cons |   .9574107   .5803727     1.65   0.104    -.2027378    2.117559
------------------------------------------------------------------------------
```

Note that the variance of $\hat{\beta}_1$ increases because we have reduced the spread in our X variable.