STATA 7.0

Index:

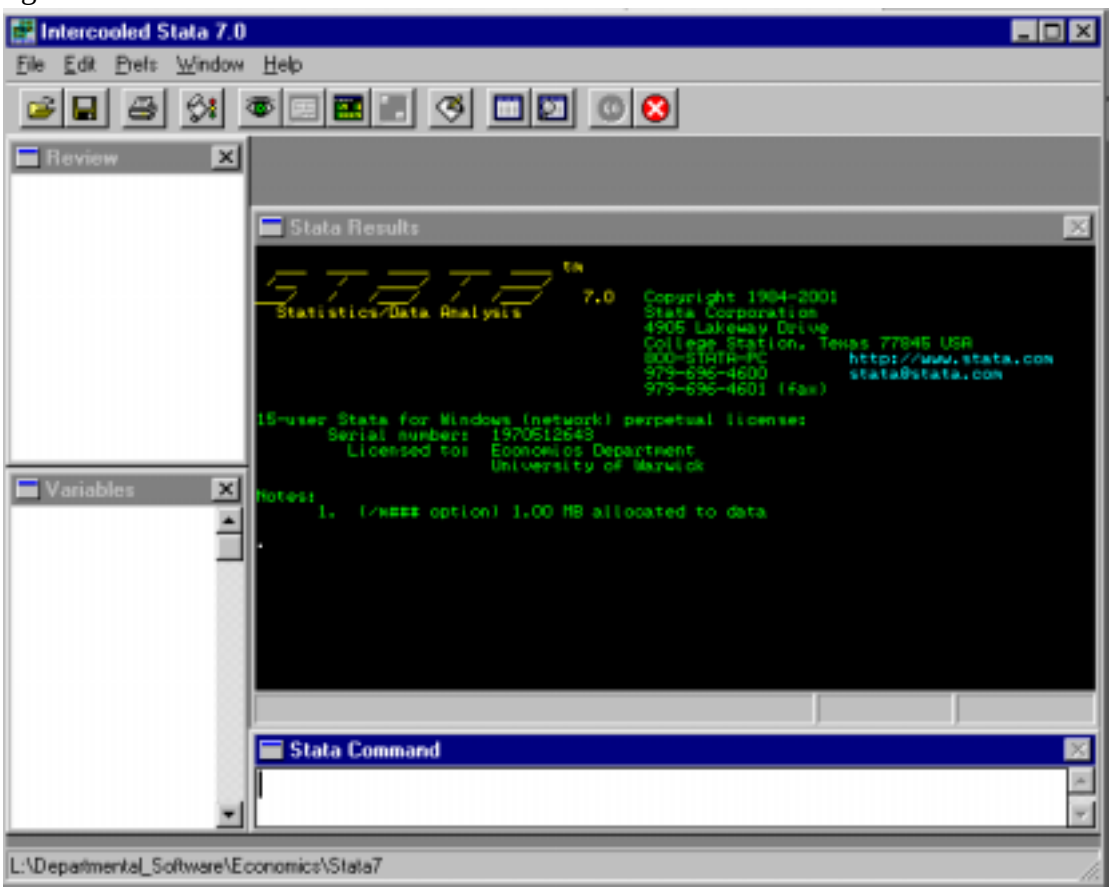
Introduction

This document introduces students to the interactive econometrics computer program Stata 7. For more details on various aspects of the program, users are directed towards the very extensive reference and user manuals "Stata 7", copies of which are in the SRC.

Stata is a windows based program and is on the Warwick network in the Departmental - Economics folder. From this folder select the Stata 7 folder and then click on Stata 7. Licences for this package are limited and you will need to be a declared user of Stata 7, before you will be able to see the Stata folder.

Clicking on the Stata you ought to get a window that looks like the one in Figure 1.

Figure 1: Stata window



You ought to see four windows:

- Stata Command: The is the window where you input commands that Stata will carry out.
- Stata Results: The results of your command will appear in the Results window. If you made an error in entering a command (for instance, misspelling or wrong syntax) a message will appear in this window in red. Note the window only has capacity for around 100 lines of results.

- Review: This gives a list of all the commands, both correct and incorrect, that you have entered in Stata.
- Variables: When you open a data set, the variables that the data set contains, along with a description of the variables (if available) will apear.

Stata is a command driven statistical package. The commands can be submitted either interactively by typing them in the Stata Coomand window and pressing return. Or a series of commands can be submitted as a BATCH job by creating a file (denoted as a Do file, written as *.do) with all the commands written down.

On opening up Stata the initial allocation of memory (1MB) to the program is usually very small and this will have to be increased. The typing the command below in the Stata Command window increases the memory to 100MB:

set memory 100000

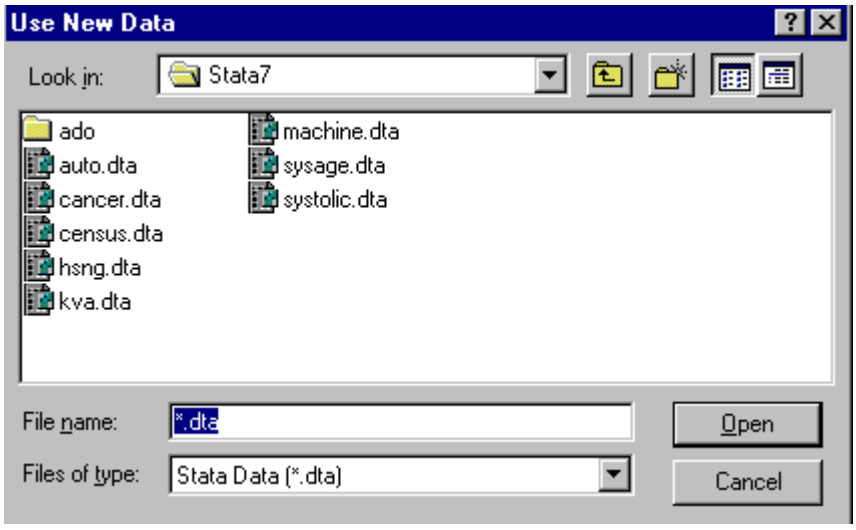
This will allow a very large data set to be read in.

1. Reading in data

If you have the data file as a Stata data file (denoted *.dta) then reading in of data is very easy.[1] There are one of two procedures:

(i) Click on File and then choose Open (or simply click  on). In Figure 2 below choose the drive and directory where you data is stored and click on OK after selecting the appropriate file.

Figure 2: File location



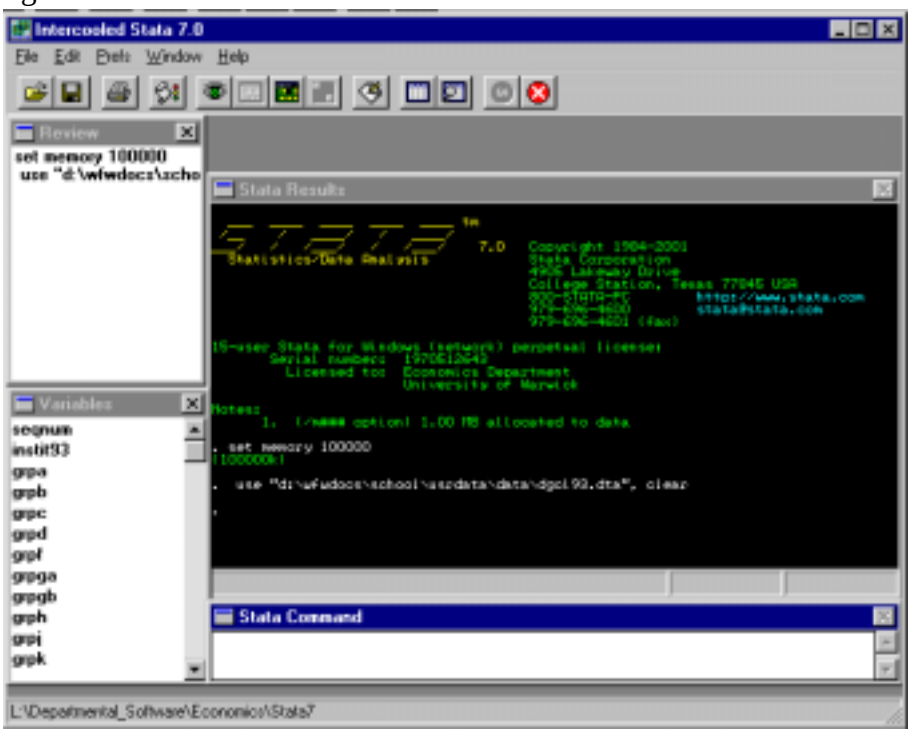
[1] You can always create Stata data files from Excel, SPSS, and other specified file types by using Stat/Transfer

(ii) In the Stata Command window, type the following command

use "d:\wfwdocs\school\usrdata\data\dgcl93.dta", clear

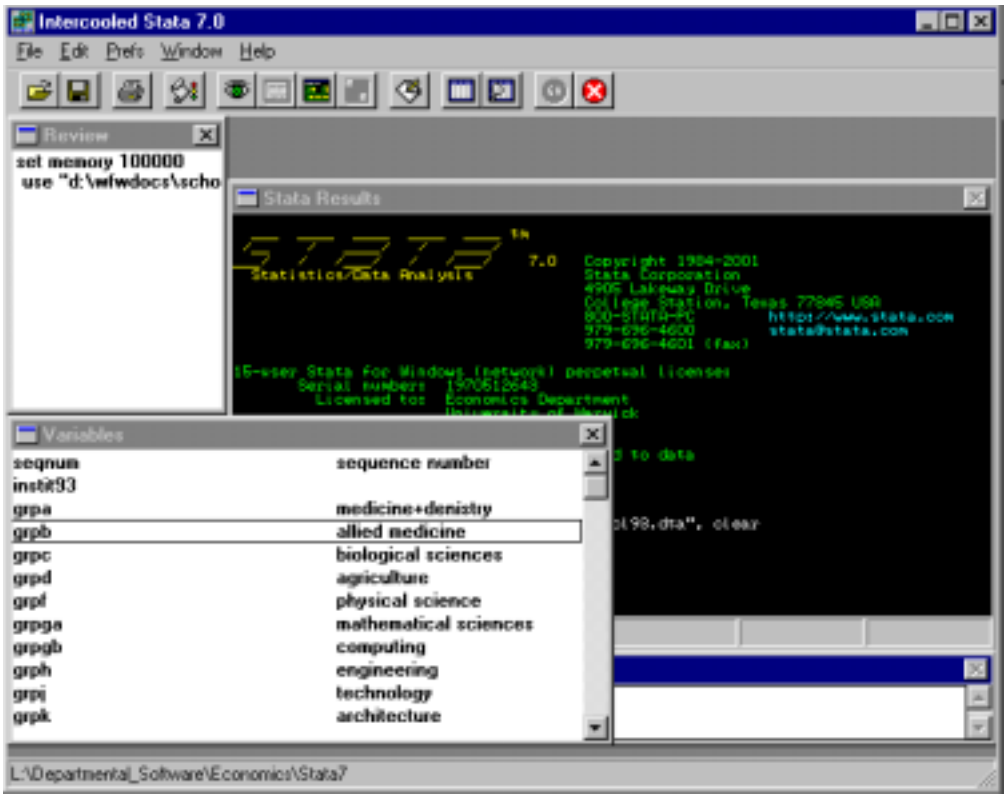
In either case your Stata window ought to look like Figure 3.

Figure 3: Stata window



The two commands now appear in the Review box and a list of variables (and their description) in the Variables box. Resizing the Variable box makes it clear the variables there along with the description (see Figure 4).

Figure 4: Variable window increased



1.1 Saving data

To save a data set type

save "d:\wfwdocs\school\usrdata\data\dgcl93.dta", replace

and this overwrites your old data file with the new data file (make sure this is really what you want to do before issuing this command).

2 Opening a results/log file

Stata allows you to create a file that keeps track of all that you do in a particular session. This file is called a results or log file. Before each session you ought to create a log file. To do this in the Stata Command window type:

log using d:\wfwdocs\school\occearn\results93.log

This creates a file called results93.log, which stores all the commands and the output of your Stata session. If you have created this file in an earlier session and you wish to add more results to that file type

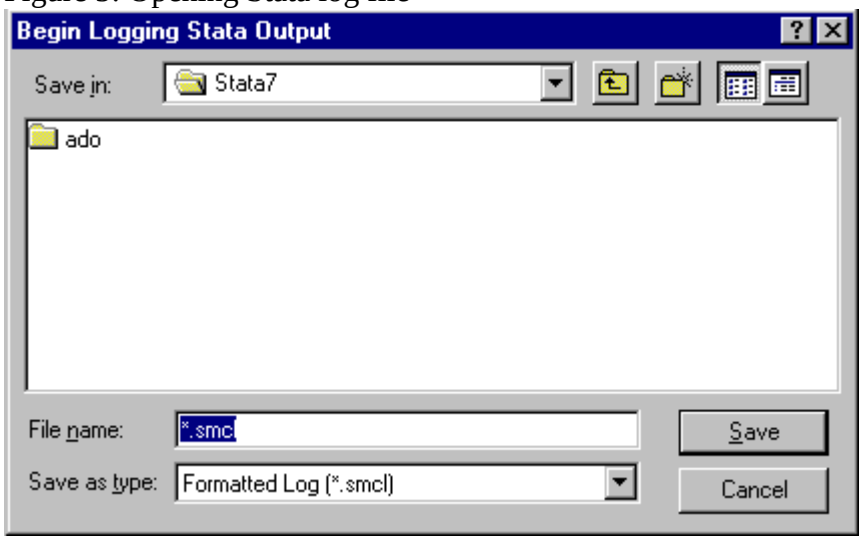
log using d:\wfwdocs\school\occearn\results93.log, append

To overwrite the file with new results type

log using d:\wfwdocs\school\occearn\results93.log, replace

Alternatively, clicking on the icon  you get the screen

Figure 5: Opening Stata log file



Which allows you to create formatted log or simple log files.

Log files can be edited in any text editor. Remember to close your log file in Stata before opening it in some other package.
To close a log file, either:

click on the icon  and you get the Figure 6, which enables you to View the log file in Stata, Close the log file or to Suspend the log file.

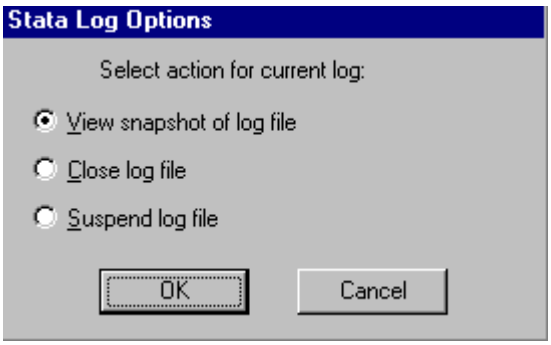


Figure 6: Stata log options

Alternartively, in the Stata Command window typing:

log close

closes the log file.

3 What is in the data files

It is possible to get some information on the data that you have in the data files. In the command window typing:

desc

(which is short for describe), gives a brief description of the data set.

```
Contains data from D:\WFWDOCS\SCHOOL\USRDATA\DATA\Dgcl93.dta
  obs:        99,569
 vars:            61                          18 Oct 2000 13:50
 size:    36,243,116 (64.6% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
seqnum          double %9.0g                  sequence number
instit93        int    %9.0g       instit93   university attended
achome          byte   %9.0g                  student living at home
mstpt           byte   %9.0g                  part-time student
age             float  %9.0g                  age of student
sex             byte   %4.0g       sex        sex of student
marital         byte   %9.0g       marital    married
resprio1        int    %9.0g       resprio1   residence prior to entry
ccode93         str4   %4s                    course code for 1993 (4 digit)
dgclass2        byte   %9.0g       dgclass2   classification of degree
nclassdg        byte   %9.0g                  classification of degree
os              byte   %9.0g                  non-uk students
aslscore        byte   %4.0g                  a-as level score best 3 passes
hlecount        byte   %4.0g                  h level count
hlescore        byte   %4.0g                  h level score best 5 passes
hlecouni        byte   %9.0g                  number of irish highers
oentq           byte   %9.0g       oentq      other entry qualifications
⋮                                              ⋮
social_c        byte   %9.0g       social_c   social class based upon parents
primm           byte   %9.0g       primm      fds: classifications
aslcount        float  %9.0g                  number of a (as) levels
hlescori        byte   %9.0g                  irish higher scores (in best 5)
hlescri2        byte   %9.0g                  irish higher scores
hlescor2        byte   %9.0g                  scottish higher scores
aslscre2        byte   %9.0g                  a-levels higher scores
lnearn          float  %9.0g                  log of weekly earninga
cstudy          str8   %8s                    Course of student (2-digit)
typework        int    %8.0g       typework   first occupation of students
cstudy2         str8   %8s                    Course of study (1 digit)
resleave        int    %8.0g       resleave   reason for leaving
enrol           long   %12.0g                 enrolment date
dalbiol         float  %9.0g                  Biology A-level
dalchem         float  %9.0g                  Chemistry A-level
dalengl         float  %9.0g                  English A-level
dalmath         float  %9.0g                  Maths A-level
dalphys         float  %9.0g                  Physics A-level
dalgens         float  %9.0g                  Gen. St. A-level
cca             str4   %4s                    course code of acceptance
-------------------------------------------------------------------------------
```

The output tells us we have 99,569 observations and 61 variables. All the variables are listed and their description is given beside the name. So that we know dgclass2 stands for classification of degree and social_c stands for social class of parents.

To look at the actual data you can click on either the data browser or the data editor icons.
The data browser icon is  clicking on this icon produces Figure 7, which lists the actual values of the variables.

The data editor icon is  and this yields a similar window, but in this window it is possible to change the actual values of the variables. You close both windows by clicking x in the top right hand corner.

Figure 7: Data browser

Stata Browser

Preserve | Restore | Sort | << | >> | Hide | Delete...

seqnum[1] = 934500009

	seqnum	instit93	grpa	grpb	grpc
1	9.35e+08	lse	0	0	0
2	9.35e+08	birkbeck	0	0	0
3	9.35e+08	queen ma	0	0	0
4	9.35e+08	durham	0	0	0
5	9.35e+08	queen ma	0	0	0
6	9.35e+08	queen ma	0	0	0
7	9.35e+08	queen ma	0	0	0
8	9.35e+08	queen ma	0	0	0
9	9.35e+08	queen ma	0	0	0
10	9.35e+08	queen ma	0	0	0
11	9.35e+08	queen ma	0	0	0
12	9.35e+08	queen ma	0	0	0
13	9.35e+08	queen ma	0	0	0
14	9.35e+08	queen ma	0	0	0
15	9.35e+08	queen ma	0	0	0
16	9.35e+08	queen ma	0	0	0

To observe the values of each of the variables and the corresponding descriptions of these values type:

label list social_c instit93

The output is:

```
social_c:
           1 professi
           2 intermed
```

```
           3 skilled
           4 skilled
           5 partly s
           6 unskille
           7 armed fo
           8 inadequa
           9 non-work
instit93:
         100 aberdeen
         200 edinburg
         300 glasgow
         400 st andre
         500 stirling
         600 strathci
         700 dundee
         800 heriot-w
        1000 birmingh
        1100 bristol
        1200 cambridg
             ⋮
        5000 warwick
        5100 umist
        5200 york
        5500 lbs
        5600 mbs
        6600 belfast
        6800 ulster
```

4 Data description

Possible to do basic descriptive analysis of some or all variables in your data set. Typing:

sum sex marital aslscore

Produces the output below:

```
Variable |    Obs       Mean  Std. Dev.      Min        Max
---------+--------------------------------------------------
     sex |  99569   .5520092  .4972902         0          1
 marital |  99569   .0463699  .2102858         0          1
aslscore |  99569   16.02156  10.77945         0         30
```

which is summary statistics (means and standard deviation, minimum and maximum) on the three specified variables. Typing:

sum

produces summary statistics on all variables. Typing

sum aslscore, detail

gives more information:

```
a-as level score best 3 passes
```

```
-------------------------------------------------------------
      Percentiles      Smallest
 1%             0             0
 5%             0             0
10%             0             0       Obs                99569
25%             0             0       Sum of Wgt.        99569

50%            20                     Mean           16.02156
                        Largest       Std. Dev.      10.77945
75%            24            30
90%            28            30       Variance       116.1965
95%            30            30       Skewness      -.4523598
99%            30            30       Kurtosis       1.730974
```

Typing:

sum t*

gives summary statistics on all variables that start with a m, that is, tctc, tfe, tgrant, tind, tsa, tspec, tvol.

To list all values for a particular variable type:

list social_c

and you get

```
social_c
    1.  skilled
    2.  skilled
    3.  non-work
    4.  professi
    5.  professi
    6.  skilled
    7.  partly s
    8.  skilled
    9.  professi
   10.  intermed
   11.  non-work
   12.  intermed
   13.  non-work
   14.  intermed
   15.  intermed
   16.  professi
   17.  skilled
   18.  skilled
   19.  professi
   20.  intermed
   21.  intermed
   22.  intermed
   23.  skilled
   24.  professi
```

[Note: At the bottom of the screen in blue is --more—this indicates the output is more than one screen long – press q to interupt the listing or enter to continue].

list social_c in 1/20 (list the first 20 observations).

Alternatively,

list sex if social_c==2 (list the sex of individuals, for social class intermediate).

Other relational logical or arithmetical symbols can be used in the if statement they are:

Relational	
>	Greater than
<	Less than
==	Equal to
>=	Greater than or equal to
<=	Less than or equal to
~=	Not equal to
Logical	
~	Not
|	Or
&	And
Arthimetical	
+	Addition
-	Subtraction
*	Multiplication
/	Division
^	Power

A more complicated statement may then be:

list sex if((social_c~=8 & social_c~=9) & age<23 & marital==0)

This will list the sex of all individuals who are not social class 8 or 9 and who are less than 23 years old and who are single.

For categorical variables (that is variables with more than two outcomes sum is not that informative and a tabulation of the number of individuals in each category is more informative. Typing:

tab instit93

Produces the frequency table:

```
   instit93 |      Freq.     Percent        Cum.
------------+-----------------------------------
   aberdeen |       1496        1.50        1.50
   edinburg |       2508        2.52        4.02
    glasgow |       2651        2.66        6.68
   st andre |       1031        1.04        7.72
   stirling |        864        0.87        8.59
   strathci |       2277        2.29       10.87
     dundee |        971        0.98       11.85
   heriot-w |        990        0.99       12.84
```

```
   birmingh |       2748        2.76       15.60
    bristol |       2374        2.38       17.99
   cambridg |       3041        3.05       21.04
                      :
    warwick |       1863        1.87       92.35
      umist |       1262        1.27       93.61
       york |       1140        1.14       94.76
    belfast |       2461        2.47       97.23
     ulster |       2759        2.77      100.00
------------+-----------------------------------
      Total |      99569      100.00
```

To do a cross tabulation of one categorical variable by another categorical variable type:

tab social_c sex

```
    social |
     class |
based upon |    sex of student
   parents |    female       male |     Total
-----------+----------------------+----------
  professi |      7579       9204 |     16783
  intermed |     16933      20206 |     37139
  skilled  |      4286       5443 |      9729
  skilled  |      4192       5880 |     10072
  partly s |      2684       3692 |      6376
  unskille |       406        658 |      1064
  armed fo |       352        464 |       816
  inadequa |      1221       1566 |      2787
  non-work |      6953       7850 |     14803
-----------+----------------------+----------
     Total |     44606      54963 |     99569
```

tab social_c if(age<23 & marital==0 & sex==1)

```
social |
class based |
       upon |
    parents |      Freq.     Percent        Cum.
------------+-----------------------------------
   professi |       5730       17.97       17.97
   intermed |      12634       39.62       57.59
   skilled  |       3499       10.97       68.56
   skilled  |       3723       11.68       80.24
   partly s |       2251        7.06       87.30
   unskille |        343        1.08       88.37
   armed fo |        297        0.93       89.31
   inadequa |        801        2.51       91.82
   non-work |       2609        8.18      100.00
------------+-----------------------------------
      Total |      31887      100.00
```

5 Graphing data

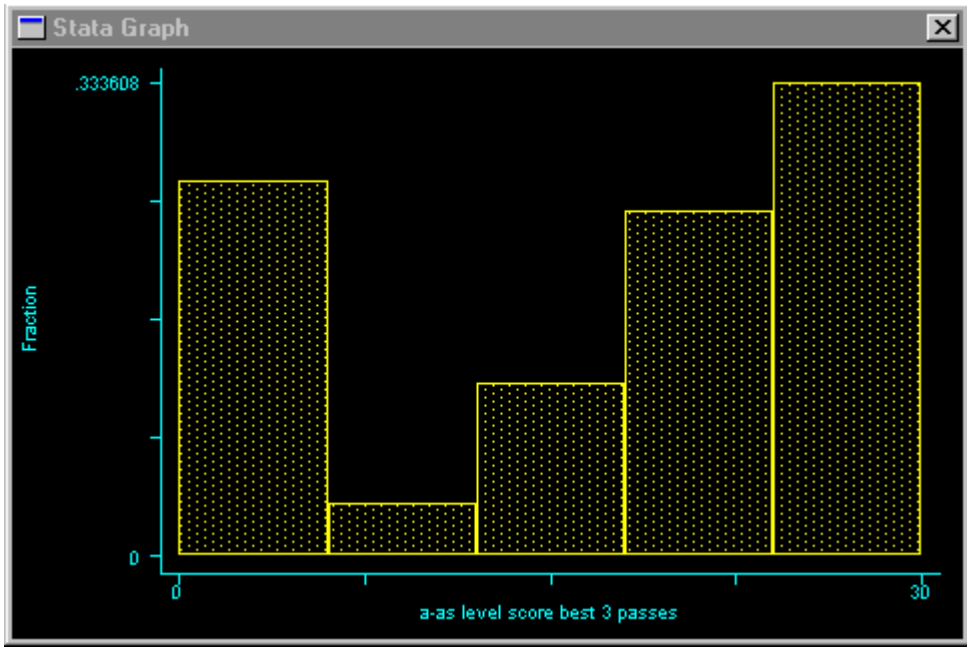
Stata has many different graphing options, which can be used to obtain visual information about the variables. If you only type a single variable name after the graph command

Stata assumes you want a histogram. If you supply a list of two or more variables after the graph command Stata assumes you want a scatter plot (the last variable list being on the x-axis).

gr aslscore

Stata brings up Figure 8.

Figure 8: Graphing window



It is possible to modify this graph in various ways:

gr aslscore if(aslscore>0), bin(8)

Produces a histogram only for A-level score>0 and has 8 bins along the x-axis, rather than the default of 5.

Typing:

gr aslscore if(aslscore>0), bin(10) normal

this overlays the histogram with a normal curve using the mean and standard deviation from the actual variable (aslscore).

Typing:

gr aslscore if(aslscore>0), bin(8) saving(c:\temp\aslscore.gph)

Saves the graph, alternatively graphs can be cut and pasted using Edit and Copy Graph. (see figure 8). To change the background colour in Stata on the graph click on Preference and Graph Preferences.

There are a large range of other types of graphs available.

6 Generating/Transforming variables

Stata generates variables using the generate command, examples of a few commands are written below:

gen tother=tsa+tvol+tctc+tgrant
gen dage1=(age>=24 & age<28)
gen dbiol=(albiol>0 | hbiol>0)

The first command simple generates a new variable *tother* as the sum of the variables for whether a student had previously attended one of the schools, tsa, tvol, tctc or tgrant. The second command generates a dummy variable equal to 1 if the expression in brackets is true and 0 otherwise. The last expression also generages a dummy variable if expression in brackets is true. Other mathematical expression available for undertaking data transformations are presented in Section 4.

To have a description along side each variable you need to type:

label var tother "Other school"
label var dage1 "Aged between 24-28"
label var dbiol "Has biology at A-level or Higher"

To construct dummy variables for each of the alternative outcomes in a categorical variable you type:

tab instit93, g(univ)
tab social_c, g(socc)

The first command generates 57 dummy variables, denoted univ1-univ57 for each of the 57 university categories. The second command generates 9 dummy variables socc1-socc9 for each of the 9 social class categories.

If a variable already exists and you wish to overwite it with a new variables you must use the replace command, for example,

replace classdg2=1 if (classdg1==1)

This command over-writes existing variable such that classdg2 now takes the value 1 if classdg1==1.

It is also possible to recode variables, for example,

recode typesch 0 6 9 20=3 3 4 8 21 23 36 86=1 5 41 42=2 15 16 93=4 87=5
label define typesch 1 "LEA" 2 "grammar" 3 "other" 4 "Indep" 5 "FE"

typesch is a categorical variable, but the 14 categories have neen recoded to be only 3 categories. The label define command defines each of the 5 category groups.

7. Regression analysis

Stata can undertake a variety of different types of regression analysis. In this section we are only looking at two types the simple OLS option and the limited dependent variables (logit/probit option).

Having undertaken the data transformations you wish, to run a regression type:

7.1 OLS

regress lnearn univ grp* socc* dage1 dage2 dage3 marital os classdg* aslscore hlescore tother tfe tind*

The dependent variable is the first variable in our case lnearn, which is the natural log or earnings. All the other variables are explanatory variables. The variable univ* are the series of 57 university dummy variables which start with the letters univ and will ultimately refer to univ1-univ57. This is similar for the variables socc* (which relates to the 9 social class groups) and grp* (this relates to the course of study for the student and there are 21 of these) and classdg* (which relates to the degree class the student obtained).

The output from this regression is reported below and includes basic summary statistics of the regression and a table of coefficients, standard errors, t-statistics, p-values and the 95% confidence interval on the coefficients.

```
      Source |       SS       df       MS              Number of obs =   48554
-------------+------------------------------           F(100, 48453) =  181.89
       Model |  1699.96778   100  16.9996778           Prob > F      =  0.0000
    Residual |  4528.39579 48453  .093459554           R-squared     =  0.2729
-------------+------------------------------           Adj R-squared =  0.2714
       Total |  6228.36357 48553  .128279685           Root MSE      =  .30571

------------------------------------------------------------------------------
      lnearn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       univ1 |    .089334   .0257009     3.48   0.001     .0389599    .1397081
       univ2 |   .0241331   .0247596     0.97   0.330     -.024396    .0726623
       univ3 |   .0513734   .0246336     2.09   0.037     .0030913    .0996555
       univ4 |   .0470262    .027111     1.73   0.083    -.0061117     .100164
                           ⋮
      univ42 |  -.0366063   .0248854    -1.47   0.141    -.0853819    .0121694
      univ43 |  -.1096254   .0300223    -3.65   0.000    -.1684694   -.0507813
      univ44 |  (dropped)
```

```
      univ45 |   .0723969   .0265071     2.73   0.006     .0204427    .1243511
      univ46 |   .0750642   .0258107     2.91   0.004     .0244749    .1256535
                                  ⋮
      univ57 |   .0313552   .0241309     1.30   0.194    -.0159417    .0786521
        grpb |  -.0125222   .0171536    -0.73   0.465    -.0461435    .0210991
        grpc |   -.131634   .0164425    -8.01   0.000    -.1638615   -.0994065
                                  ⋮
        grph |   .0837704   .0161306     5.19   0.000     .0521543    .1153866
        grpj |  (dropped)
        grpk |   .0044897   .0197299     0.23   0.820    -.0341811    .0431606
                                  ⋮
        grpy |   -.064244   .0177559    -3.62   0.000    -.0990458   -.0294422
       socc1 |   .0403259   .0151832     2.66   0.008     .0105667    .0700851
       socc2 |   .0366313   .0150195     2.44   0.015     .0071929    .0660698
       socc3 |   .0333278    .015475     2.15   0.031     .0029966     .063659
       socc4 |   .0335181   .0155896     2.15   0.032     .0029623    .0640738
       socc5 |   .0202335   .0160223     1.26   0.207    -.0111704    .0516375
       socc6 |   .0184877   .0215134     0.86   0.390     -.023679    .0606543
       socc7 |  (dropped)
       socc8 |   .0517817   .0171152     3.03   0.002     .0182356    .0853277
       socc9 |   .0487546   .0161011     3.03   0.002     .0171962    .0803129
       dage1 |   .0429626   .0047972     8.96   0.000       .03356    .0523651
       dage2 |   .0488095    .008598     5.68   0.000     .0319574    .0656617
       dage3 |   .0229254   .0103393     2.22   0.027     .0026602    .0431906
     marital |  -.0033197   .0096137    -0.35   0.730    -.0221627    .0155233
          os |   .0733923   .0082268     8.92   0.000     .0572677    .0895169
    classdg3 |  -.0073384   .0099033    -0.74   0.459     -.026749    .0120722
    classdg4 |  (dropped)
    classdg5 |   .0153796   .0071733     2.14   0.032     .0013197    .0294394
    classdg6 |   .0486205   .0071281     6.82   0.000     .0346492    .0625917
    classdg7 |   .1020855   .0084017    12.15   0.000      .085618     .118553
    aslscore |   .0020821   .0002326     8.95   0.000     .0016263     .002538
    hlescore |   .0036135   .0007503     4.82   0.000      .002143    .0050841
      tother |   .0038911   .0041845     0.93   0.352    -.0043106    .0120927
         tfe |  -.0114249    .005228    -2.19   0.029    -.0216719   -.0011779
        tind |   .0443413   .0037648    11.78   0.000     .0369623    .0517204
       _cons |   5.779109   .0318944   181.20   0.000     5.716596    5.841623
------------------------------------------------------------------------------
```

For each of the categorical dummy variables one has been excluded in the regression due to collinearity (as they together sum to the Intercept (denoted _Cons). R^2 is around 27% (typical for this type of model) Individuals with a 1st class degree (classdg7) have approximately 10.2% higher occupational earnings than individuals who obtained a 3rd class degree (classdg4). Individuals with an extra 1 pts at A-level get earnings of 0.21% higher and Independent school children have earnings of 4.4% higher.

Typing:

pre earnhat
pre earnres, re

The first command saves the predicted values into the variable earnhat, whereas the second command saves the residuals into the variable earnres.

7.2 Logit/Probit

In this section we talk about the use of the logit/probit command, we actually use logit, although there is essentially little difference between the two. Typing

logit classdg21 univ grp* socc* dage1 dage2 dage3 marital os aslscore hlescore tother tfe tind*

where classdg21 is a dummy variable where 1 indicates the student obtained at least an upper second class degree and 0 otherwise. You get the output:

```
note: univ43 dropped due to collinearity
note: grpt dropped due to collinearity
note: socc7 dropped due to collinearity
Iteration 0:   log likelihood = -68764.398
Iteration 1:   log likelihood = -61400.521
Iteration 2:   log likelihood = -61016.873
Iteration 3:   log likelihood = -60965.824
Iteration 4:   log likelihood = -60962.474
Iteration 5:   log likelihood = -60962.448
Iteration 6:   log likelihood = -60962.448

Logit estimates                                   Number of obs   =      99569
                                                  LR chi2(94)     =   15603.90
                                                  Prob > chi2     =     0.0000
Log likelihood = -60962.448                       Pseudo R2       =     0.1135

------------------------------------------------------------------------------
   classdg21 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       univ1 |  -.3801624    .214181    -1.77   0.076    -.7999493    .0396246
       univ2 |  -.4137712   .2112441    -1.96   0.050    -.8278019    .0002596
       univ3 |  -.5503153   .2120395    -2.60   0.009    -.9659051   -.1347254
       univ4 |  -.1534414   .2175108    -0.71   0.481    -.5797548    .2728719
                                 ⋮
      univ53 |  -.2715846   .2121095    -1.28   0.200    -.6873116    .1441424
      univ54 |  -.3992739   .2146989    -1.86   0.063     -.820076    .0215283
      univ55 |  -.4710884   .2153961    -2.19   0.029     -.893257   -.0489198
      univ56 |   -.726484   .2106743    -3.45   0.001    -1.139398   -.3135699
      univ57 |  -.1292077   .2097605    -0.62   0.538    -.5403307    .2819153
        grpb |   .1571008   .0786483     2.00   0.046      .002953    .3112486
        grpc |   .2573601   .0731861     3.52   0.000     .1139181    .4008022
                                 ⋮
        grpy |  -.5569456   .0789459    -7.05   0.000    -.7116767   -.4022145
       socc1 |   .0816127   .0766338     1.06   0.287    -.0685869    .2318123
       socc2 |   .0336371   .0756389     0.44   0.657    -.1146124    .1818865
       socc3 |  -.0404396    .077976    -0.52   0.604    -.1932698    .1123905
       socc4 |   -.371376   .0780382    -4.76   0.000    -.5243281   -.2184239
       socc5 |  -.4058435   .0797028    -5.09   0.000     -.562058   -.2496289
       socc6 |  -.5424652   .1003847    -5.40   0.000    -.7392156   -.3457149
       socc8 |  -.1035389   .0850949    -1.22   0.224    -.2703218     .063244
       socc9 |  -.6552374   .0786527    -8.33   0.000    -.8093939    -.501081
       dage1 |   .0607559   .0228935     2.65   0.008     .0158855    .1056262
       dage2 |    .395904   .0365193    10.84   0.000     .3243275    .4674806
       dage3 |   .2835188   .0444378     6.38   0.000     .1964223    .3706153
     marital |   .2502011   .0419534     5.96   0.000     .1679739    .3324283
          os |   .0304482   .0279176     1.09   0.275    -.0242692    .0851657
    aslscore |   .0478659   .0010293    46.50   0.000     .0458484    .0498834
```

```
    hlescore |   .0741026   .0037324    19.85   0.000     .0667873    .0814179
      tother |  -.0515832   .0211602    -2.44   0.015    -.0930564     -.01011
         tfe |  -.1445451   .0263634    -5.48   0.000    -.1962164   -.0928737
        tind |  -.3594523   .0192315   -18.69   0.000    -.3971453   -.3217593
       _cons |  -.1397614   .2304749    -0.61   0.544     -.591484    .3119612
------------------------------------------------------------------------------
```

Due to collinearity a number of variables are dropped from the analysis, as there were in the OLS model. These are the coefficient estimates and are not interpretable, however, typing

logistic classdg21 univ grp* socc* dage1 dage2 dage3 marital os aslscore hlescore tother tfe tind*

reports the Odds Ratio and the standard errors and t-statistics associated with this. If you instead typed:

probit classdg21 univ grp* socc* dage1 dage2 dage3 marital os aslscore hlescore tother tfe tind*

You get the probit results (assuming a normal error, rather than the logistical error term). Typing:

dprobit classdg21 univ grp* socc* dage1 dage2 dage3 marital os aslscore hlescore tother tfe tind*

Then reports the marginal effects of the probit results and the standard errors and t-ratios associated with these transformed coefifcients.

pre prob21

Saves the probabilities of getting an upper second or better into the variable prob21, to save the fitted values of the index function, xb, in a variable hat21, type:

pre hat21, xb

7.3 Hypothesis Testing

After running a regression (a linear regression or otherwise) it is possible to test the joint significance of a subset of variables. Typing:

test x1 x2 x3 x4

jointly tests the hypothesis that the coefficients on these 4 variables are equal to zero. Alternatively,

test x1=x2
test x2=x3, accum

test x3=x4, accum

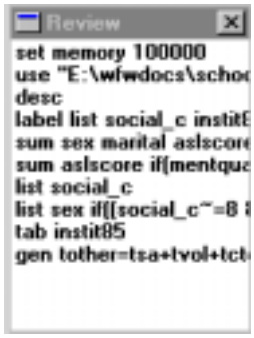
tests the hypothesis that the coefficients on the 4 variables are equal. This is done by first testing the coefficients on x1 and x2 are equal. It then reports whether the coefficients on x2 and x3 are equal and accumulates this with the previous test statistic. Finally it tests whether the coefficients on x3 and x4 are equal and again accumulates this in with the previous test statistics.

8 Repeating Commands

If you are issuing a number of commands that are similar in nature by pressing Page Up (or Page Down) when you are in the Stata Command box you can scroll through all of the commands you have written. These could then be edited (if needed) issued again.

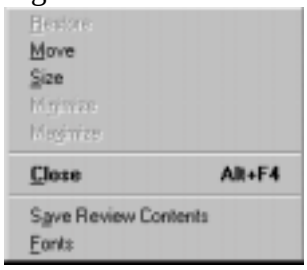
Alternatively, the Review box will also contain all of the commands that have been typed into the Stata Command box (see Figure 9).

Figure 9: Review box



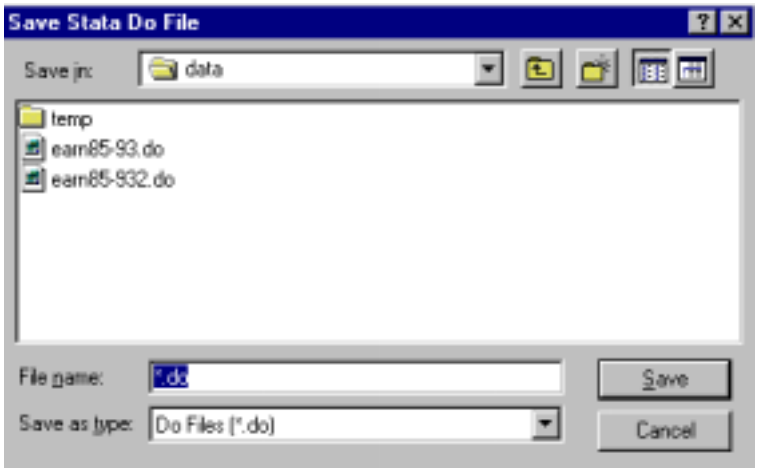
Using the mouse one click on any command in the Review box puts that command into the Stata Command box (and pressing enter issues that command again). Double clicking the command puts the command into the Stata Command box and tells Stata to execute that command. Left clicking on the icon in the top corner of the Review box produces Figure 10

Figure 10: Review box options



Choosing Save Review Contents – yields Figure 11 and requires you to supply the name and location of a DO file, to which all the commands in the Review box will be written.

Figure 11: Location of the *.DO file



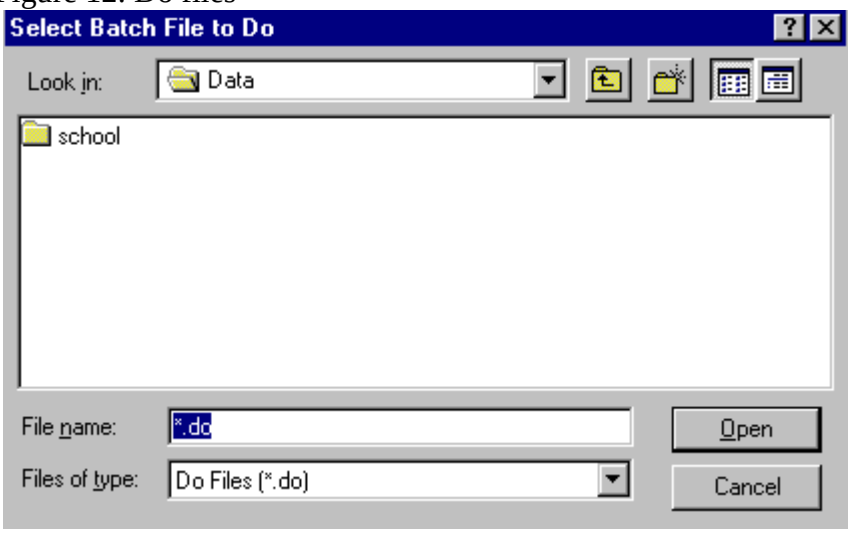
The do file is simply a syntax file of Stata commands. In our case the file we saved looks like this:

```
set memory 100000
use "E:\wfwdocs\school\Usrdata\data\usr93.dta", clear
desc
label list social_c instit93
sum sex marital aslscore
sum aslscore if(mentqual==5), detail
list social_c
list sex if((social_c~=8 & social_c~=9) & age<23 & marital==0)
tab instit93
gen tother=tsa+tvol+tctc+tgrant
```

9 Do files

All commands could have been written together into a DO file to start with. This file can be created in any text editor and then saved as *.do. To run the whole file click File and then Do. In Figure 9, you find the drive and directory where the file is located. Select the file and click Open.

Figure 12: Do files



The file will then run in Stata. Below is an example do file, written in a simple text editor. Stata will ignore all commands and notation between /* and **/.** If a single command line is too long put **/*** at the end of the line press enter and start the next line with */. Stata will then read the command as one long line.
[Note: I have continued to put actual Stata commands in italics to make reading easier].

/* This is needed to ensure there is enough space to read the data */
set memory 250000
/* This sets the size of the matrix to use, with a lot of explanatory variables the default matrix size was too small */
set matsize 750
/* This reads in the data file */
use "d:\wfwdocs\school\usrdata\data\dgcl93.dta", clear
/* This opens up a results file to which all output goes */
log using c:\wfwdocs\school\occearn\temp93.log, replace
/* This restricts the pasue command to 1 sec */
set more 1
/* These 3 commands generate dummy variables for university, social class and degree class */
tab instit93, g(univ)
tab social_c, g(socc)
tab dgclass2, g(classdg)
/* This does some basic data transformations */
replace classdg2=1 if (classdg1==1)
replace classdg3=1 if (classdg2==1)
gen classdg21=classdg6+classdg7
gen byte dage1=(age>=24 & age<28)
gen byte dage2=(age>=28 & age<34)
gen byte dage3=(age>=34)
gen tother=tsa+tvol+tctc+tgrant

/* This drops the following variables from the data set */
drop classdg2 classdg3 univ36 grpl socc2 classdg6

/* OLS regression */
regress lnearn univ grp* socc* dage1 dage2 dage3 marital os classdg* aslscore /* */ hlescore tother tfe tind if((aslscore>0 | hlescore>0) & grpa==0)*

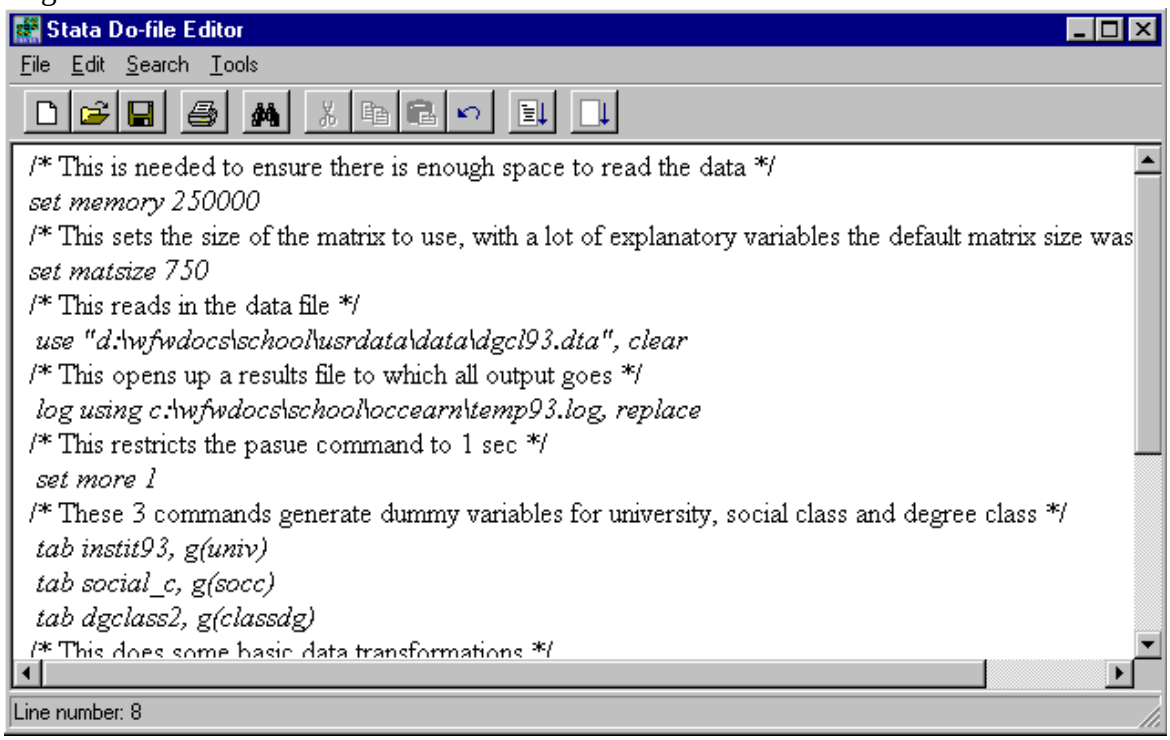
/* Logit regression */

logit classdg21 univ grp* socc* dage1 dage2 dage3 marital os aslscore hlescore /* */ tother tfe tind if((aslscore>0 | hlescore>0) & grpa==0)*

Stata has a DO file icon  clicking on this icon enables you to open up any DO file

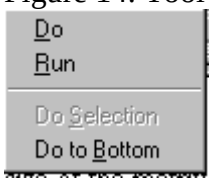
you have created. In Figure 13 we open up the do file written above.

Figure 13: Stata Do-file editor



This edito works in a similar fashion to all other text editors. However, under Tools you have the options:

Figure 14: Tool options



which enables you to run highlighted sections of your do file.