

Instructions on how to use the STATA code for Eozenou, Rivas and Schlag (2006): "Minimax Regret in Practice: Four Examples on Treatment Choice"

Patrick Eozenou, Javier Rivas and Karl Schlag
*European University Institute**

September 2006

Abstract

This is a short tutorial explaining how to use the STATA code with any given data set on treatment effects in order to implement the methodology developed in Eozenou, Rivas and Schlag (2006) : "Minimax Regret in Practice: Four Examples on Treatment Choice".

* To contact the authors: patrick.eozenou@iue.it, javier.rivas@iue.it, schlag@iue.it.

Department of Economics, European University Institute.

Villa San Paolo, Via della Piazzuola 43, 50133 Florence (Italy).

1. Preliminaries

The STATA routine is released in a *zip* file named **MINIMAX STATA.zip** and it is available for download at the authors websites:

- [Http://www.iue.it/Personal/Researchers/eozenou/Research.html](http://www.iue.it/Personal/Researchers/eozenou/Research.html) (Patrick Eozenou)
- [Http://www.iue.it/Personal/Researchers/JavierRivas/](http://www.iue.it/Personal/Researchers/JavierRivas/) (Javier Rivas)
- [Http://www.iue.it/Personal/Schlag/Welcome.html](http://www.iue.it/Personal/Schlag/Welcome.html) (Karl Schlag)

The package includes a folder called **STATA files** which contains five ado files: *minimax.ado*, *m_BAR.ado*, *m_CBAR.ado*, *m_BAR_PS.ado*, *m_CBAR_PS.ado*.

Beside this folder, the package also includes the *pdf* file containing the instructions (this file), and an Excel file (*Experiments_Round_1.xls*) containing a dataset which we use to illustrate the method¹.

The package can be unzipped anywhere. The first thing to do then is to copy the five ado files included in the folder **STATA files** and to paste them in the appropriate STATA directory. In most cases² the STATA directory path will be "**C:\Program Files\Stata9**". In this STATA directory you have to open a folder named **ado** (picture 1) where STATA stores all the ado files. In this ado folder you have then to open the folder **updates** (picture 2) in which you will find many folders indexed by alphabetical order (picture 3). The destination folder is the folder **m**. Open this folder and paste the files.

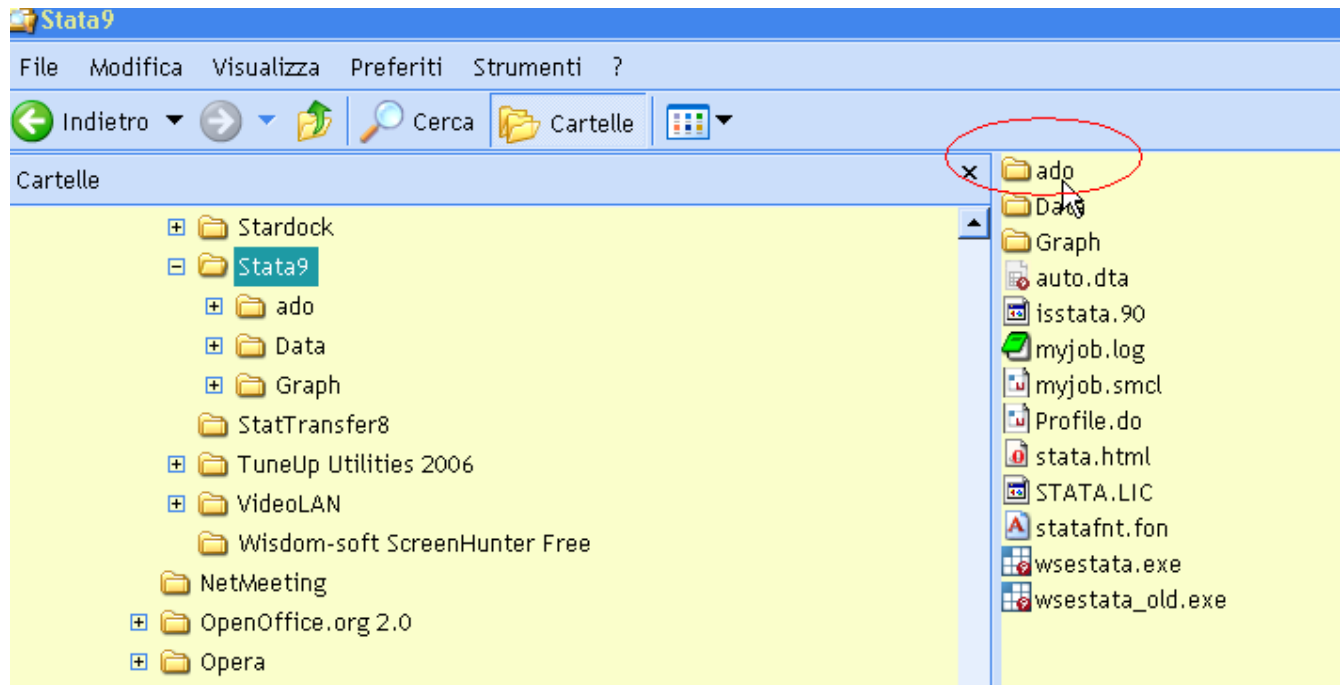
The second thing to do is to create a folder for our data. Go back to the STATA directory (**c:\Program files\Stata9**) and locate the folder where STATA stores the data (usually **c:\Program files\Stata9\Data**). If this folder doesn't exist, you can create it. In this

1 The STATA routine is a translation of the MATLAB code also available at the authors website. This tutorial echoes the instruction manual written for the MATLAB code and uses the same dataset for illustration.

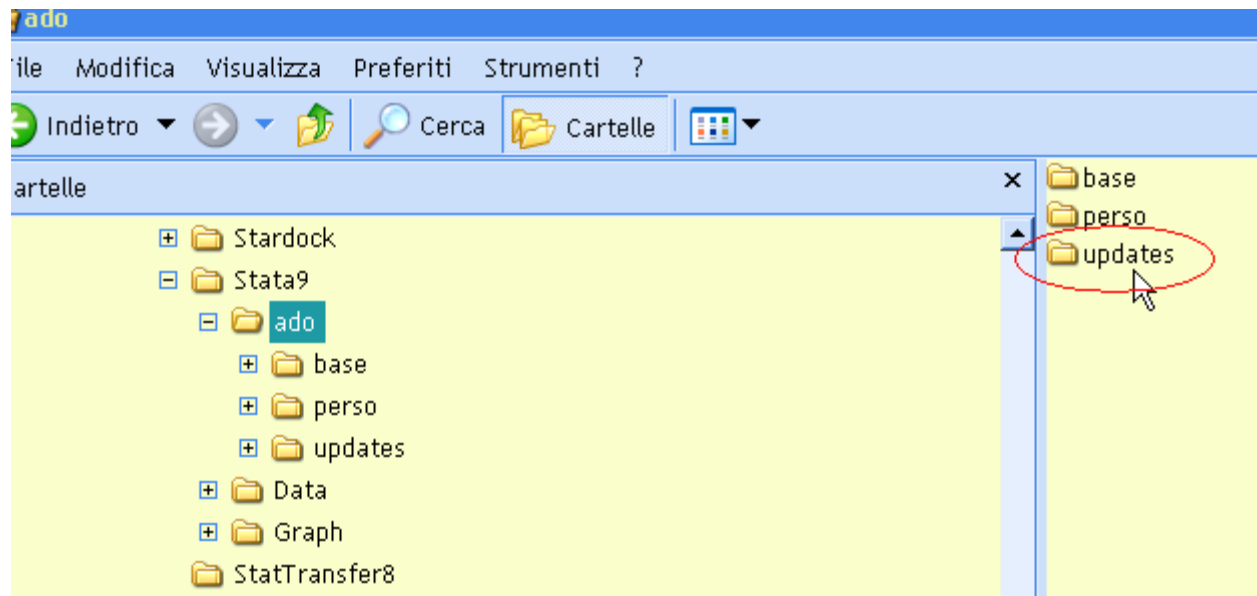
2 This path is valid for the version 9. If you own another version, then the version number should be different.

folder, create a new one which you can call *Data_MINIMAX*.

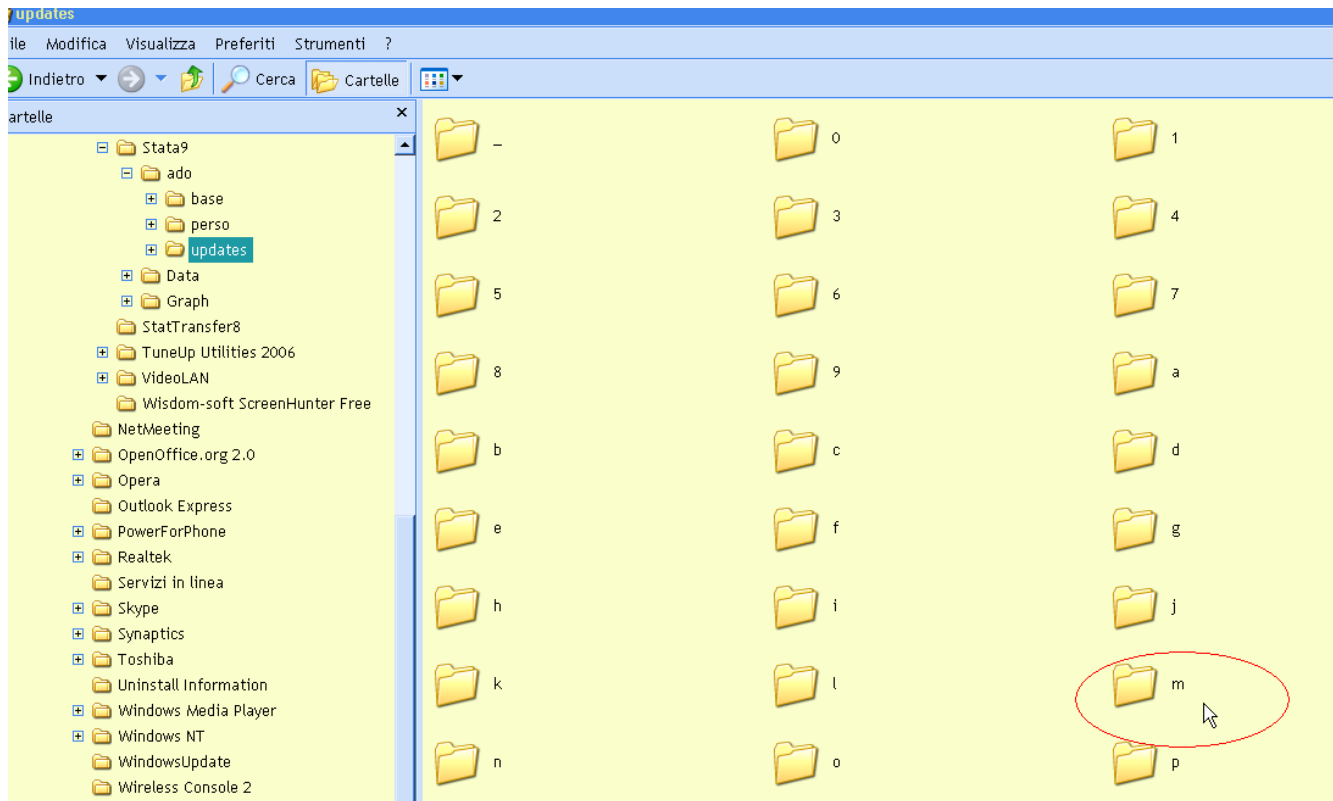
Picture 1



Picture 2



Picture 3



2. Preparing the data

The original dataset can be an Excel file or any other file.

2.1 Preparing the data from an Excel File

In our example case the dataset is the Excel file *Experiments_Round_1.xls*. This file contains the results on round 1 for all countries of the experiment run by Roth et al. (1991). For more information on this dataset see Roth et al. (1991) and Eozenou, Rivas and Schlag (2006).

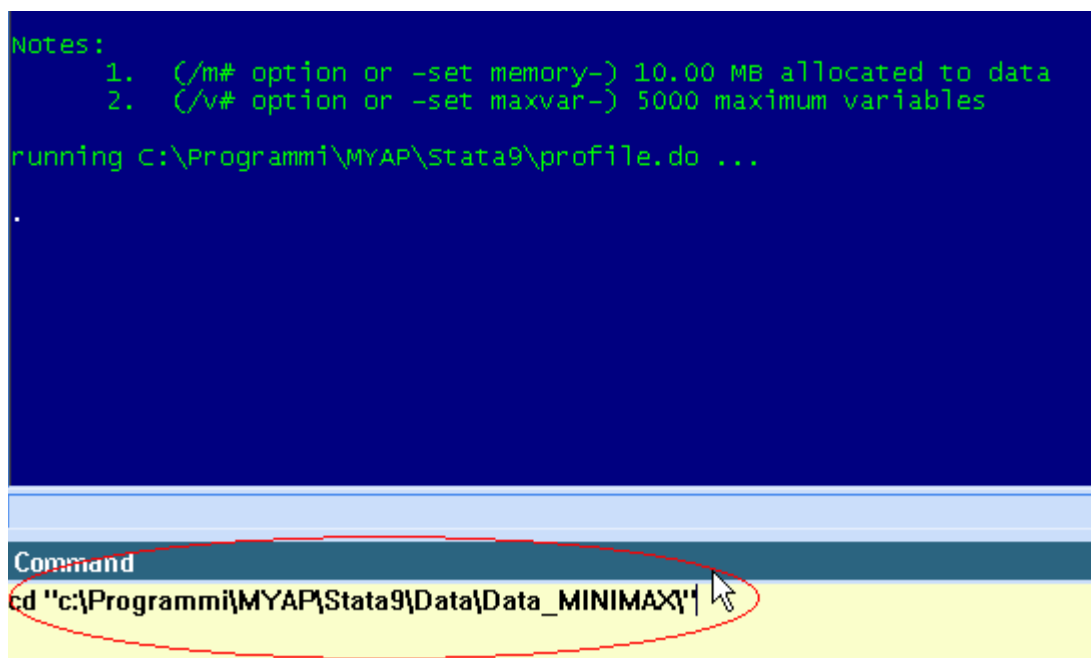
Each country (Israel, Japan, Yugoslavia and USA) represents a treatment. For each

country the first column (A, D, G and J) represents the ID of the offerer and the second column (B, E, H and K) represents the outcome of the treatment on a particular subject.

To prepare the data for **Israel**, proceed as follows:

1. Open STATA. Change the default directory by typing in the **Command window** (consistently with the folder previously created to store the data):

```
cd "c:\your directory path\Stata9\Data\Data_MINIMAX"
```

A screenshot of the STATA Command Window. The window has a dark blue background for the main area and a lighter blue header. The header contains the text "Notes:" followed by two numbered items: "1. (/m# option or -set memory-) 10.00 MB allocated to data" and "2. (/v# option or -set maxvar-) 5000 maximum variables". Below the header, the text "running C:\Programmi\MYAP\Stata9\profile.do ..." is displayed. At the bottom, there is a yellow bar labeled "Command" in a dark blue box. The command "cd \"c:\Programmi\MYAP\Stata9\Data\Data_MINIMAX\"|" is entered in the Command window, and a red oval highlights the entire command line. A mouse cursor is visible at the end of the command line.

```
Notes:
1. (/m# option or -set memory-) 10.00 MB allocated to data
2. (/v# option or -set maxvar-) 5000 maximum variables

running C:\Programmi\MYAP\Stata9\profile.do ...

Command
cd "c:\Programmi\MYAP\Stata9\Data\Data_MINIMAX\"|
```

(press enter after each command)

2. Open the file *Experiments_Round_1.xls*.
3. In the Excel window, select the data corresponding to the outcome (here the array B2:B31).

Experiments_round_1 - OpenOffice.org Calc

File Edit View Insert Format Tools Data Window Help

Arial 10 B I U

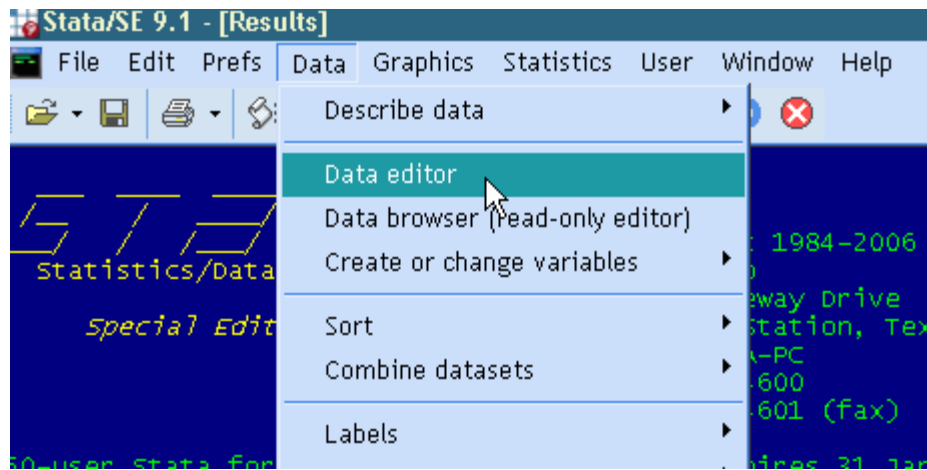
B2:B31 f(x) Σ = 4.85

	A	B	C	D	E	F	G	H	I	J	K	L
1	Israel			Japan			Yugoslavia			USA		
2	1	2		1	4.5		1	5		301	4	
3	2	4.75		2	2.5		2	5.5		302	2	
4	3	4.8		3	9.5		3	3.25		303	4	
5	4	5		4	5		4	1.5		304	5	
6	5	5		5	5		5	4.5		305	5	
7	6	3		6	5		6	4.5		306	4.5	
8	7	4.5		7	4.9		7	3.9		307	5	
9	8	5		8	0.6		8	3.2		308	4	
10	9	4		9	2.5		9	5		1	2.5	
11	10	5.5		10	4.9		10	5		2	5	
12	101	5		101	5		101	3		3	5	
13	102	4		102	2.5		102	3		4	5	
14	103	1		103	5.25		103	4.5		5	4	
15	104	4.95		104	2		104	4.45		6	4	
16	105	0.5		105	6.15		105	5		7	5	
17	106	3		106	4.5		106	5		8	4	
18	107	4.9		107	1		107	5		9	5	
19	108	3.75		108	10		108	4.9		10	3	
20	109	5		109	3.5		109	4.5		101	5	
21	110	4.05		191	4		110	5		102	5	
22	201	1		192	6.5		201	4.75		103	5	
23	202	3.5		193	3.8		202	4.7		104	3	
24	203	5		194	4		203	4.5		105	6.25	
25	204	1		195	5		204	4		106	5	
26	205	1		196	1.25		205	4.5		107	5	
27	206	2.5		197	5		206	5		108	5.5	
28	207	1.5		198	5.5		207	5		109	5	
29	208	4.8		199	5		208	5				
30	209	4		200	5		209	4.5				
31	210	4.85					210	5				
32												
33												

Take care that the data you read in Excel are *dot-separated* and not *comma-separated*, this will depend on your language settings in Excel. STATA does not read comma-separated values. You can change this settings in Excel by going to Format > Cell.

4. Copy the selected cells.

5. In STATA, open the Data editor



6. In the Data editor, select the upper left cell and paste the data

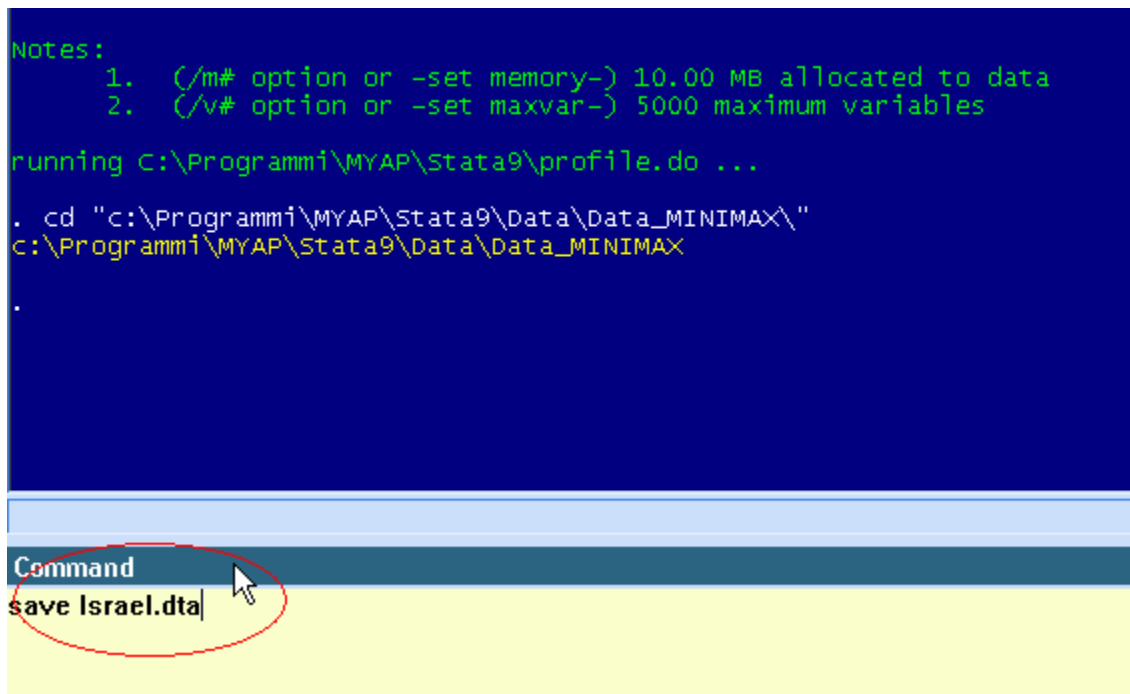
The screenshot shows the Data Editor window. The table has 10 rows and 7 columns. The first column is labeled 'var1'. The values in the first column are 2, 4.75, 4.8, 5, 5, 3, 4.5, 5, 4, 5.5. The second column is labeled 'var1' and contains values 2, 4.75, 4.8, 5, 5, 3, 4.5, 5, 4, 5.5. The third column is labeled 'var1' and contains values 2, 4.75, 4.8, 5, 5, 3, 4.5, 5, 4, 5.5. The fourth column is labeled 'var1' and contains values 2, 4.75, 4.8, 5, 5, 3, 4.5, 5, 4, 5.5. The fifth column is labeled 'var1' and contains values 2, 4.75, 4.8, 5, 5, 3, 4.5, 5, 4, 5.5. The sixth column is labeled 'var1' and contains values 2, 4.75, 4.8, 5, 5, 3, 4.5, 5, 4, 5.5. The seventh column is labeled 'var1' and contains values 2, 4.75, 4.8, 5, 5, 3, 4.5, 5, 4, 5.5.

	var1					
1	2					
2	4.75					
3	4.8					
4	5					
5	5					
6	3					
7	4.5					
8	5					
9	4					
10	5.5					

The variable created will be labeled **var1** (DO NOT change the name of this variable).
Close the Data editor.

7. Now save the data by typing in the **Command window**:

save Israel.dta

A screenshot of the Stata Command window. The window has a dark blue background for the main area and a lighter blue header. The text in the window is as follows:
Notes:
1. (/m# option or -set memory-) 10.00 MB allocated to data
2. (/v# option or -set maxvar-) 5000 maximum variables
running C:\Programmi\MYAP\Stata9\profile.do ...
. cd "c:\Programmi\MYAP\Stata9\Data\Data_MINIMAX\
c:\Programmi\MYAP\Stata9\Data\Data_MINIMAX
.
The Command window title bar is visible at the bottom, showing "Command" and "save Israel.dta". A red circle is drawn around the "save Israel.dta" text in the Command window, and a mouse cursor is pointing at it.

Note: If the data file Israel.dta was already created, type:

save Israel.dta, replace

8. Clear the Data editor by typing in the **Command window**:

clear

That's it. To prepare the data for the other countries, repeat step 2 to step 8. In the end, you should have four data files e.g *Israel.dta*, *Japan.dta*, *Yugoslavia.dta* and *USA.dta*.

You can check the presence of these files in

c:\your directory path\Stata9\Data\Data_MINIMAX.

2.2 Preparing the data from other types of files

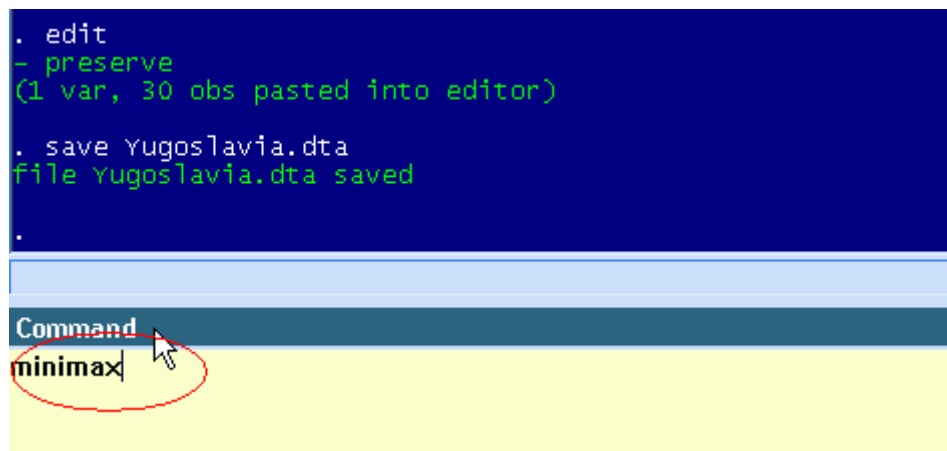
If the original dataset is not formatted as an Excel file you can either use a format transfer software (e.g *Stat Transfer*) to bring it into STATA format³, or convert the dataset into an Excel file (see the Excel manual).

3. Running the code

We explain now how to run the program. As an illustration we show how to choose the better treatment between Israel and Yugoslavia using the *Binomial Average Rule*, as described in Eozenou, Rivas and Schlag (2006).

1. Prepare the data for the two treatments as explained in the previous section.
2. In the **Command window**, type:

minimax

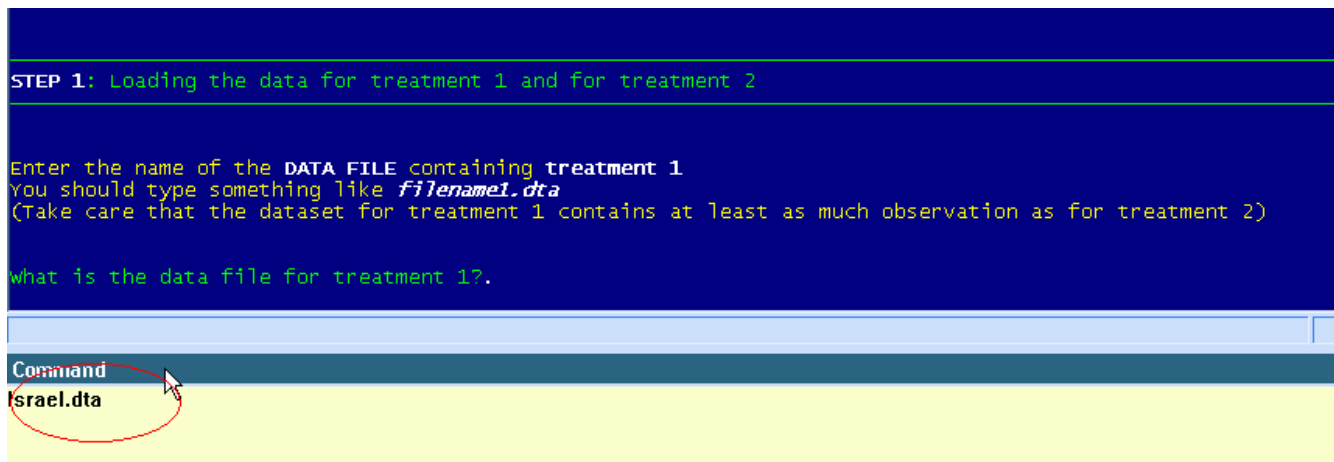


The screenshot shows the STATA Command Window with a dark blue background. The command history displays the following text: `. edit`, `- preserve`, `(1 var, 30 obs pasted into editor)`, `. save yugoslavia.dta`, and `file yugoslavia.dta saved`. Below this, the Command Window is highlighted in yellow, and the word `minimax` is entered at the prompt. A red circle is drawn around the `minimax` command, and a mouse cursor is pointing at it.

3 In this case it is important to have only one treatment per data file and to keep the variable name of your file labelled as `var1`.

3. The program then asks you the **name of the data file for treatment 1**. You should type here the name of the data file that contains the higher number of observation (in our example, both *Israel.dta* and *Yugoslavia.dta* contain 30 observations...)

Israel.dta



The screenshot shows a Stata command window with a dark blue background and yellow text. The text reads: "STEP 1: Loading the data for treatment 1 and for treatment 2", "Enter the name of the DATA FILE containing treatment 1", "You should type something like *filename1.dta*", "(Take care that the dataset for treatment 1 contains at least as much observation as for treatment 2)", and "What is the data file for treatment 1?". Below the text, there is a light blue horizontal bar. Underneath this bar, the word "Command" is written in white, and "Israel.dta" is written in black. A red circle is drawn around the text "Israel.dta", and a mouse cursor is pointing at it.

```
STEP 1: Loading the data for treatment 1 and for treatment 2

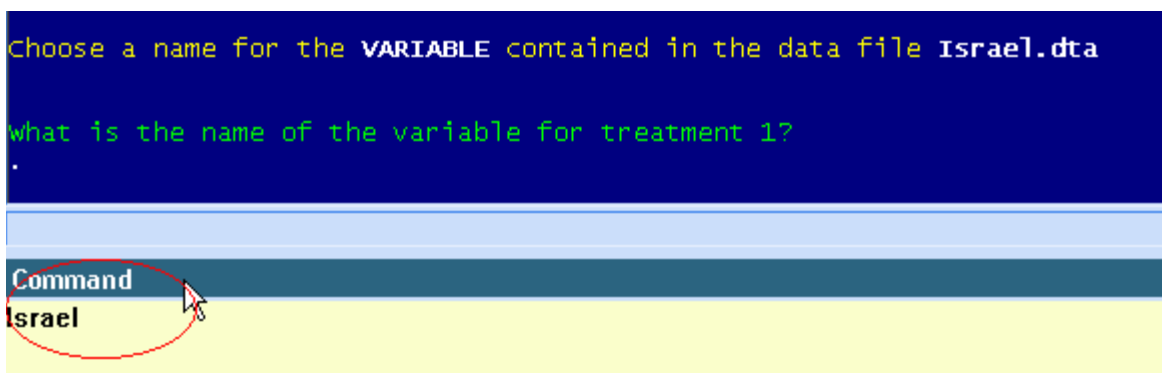
Enter the name of the DATA FILE containing treatment 1
You should type something like filename1.dta
(Take care that the dataset for treatment 1 contains at least as much observation as for treatment 2)

What is the data file for treatment 1?

Command
Israel.dta
```

4. Next, you have to choose a name for the variable of the previous data file, e.g type

Israel



The screenshot shows a Stata command window with a dark blue background and yellow text. The text reads: "Choose a name for the VARIABLE contained in the data file Israel.dta", "What is the name of the variable for treatment 1?", and ".". Below the text, there is a light blue horizontal bar. Underneath this bar, the word "Command" is written in white, and "Israel" is written in black. A red circle is drawn around the text "Israel", and a mouse cursor is pointing at it.

```
Choose a name for the VARIABLE contained in the data file Israel.dta

What is the name of the variable for treatment 1?
.

Command
Israel
```

5. After that, the program asks you to repeat step 3 and step 4 for treatment 2
6. Enter the **lowest possible value** of the outcome and the **highest possible value** to determine the range. In our example the range is [0 ; 10].
7. Now choose whether you want to implement the *Binomial Average Rule* (type 0) or the *Correlated Binomial Average Rule* (type 1).

```
STEP 3: Method

Choose here between the Binomial Average Rule or the Correlated Binomial Average Rule
If you want to implement the Binomial Average Rule, type 0
If you want instead to implement the Correlated Binomial Average Rule, type 1

Which method do you choose?.
```

```
Command
0
```

8. Tell the program whether the data come from a **paired sample** (1: yes / 0: no).

```
STEP 4: Paired Experiment

Indicate whether your data come from a paired experiment
(see section 2.2.4 in the paper)
(1:yes / 0:no)

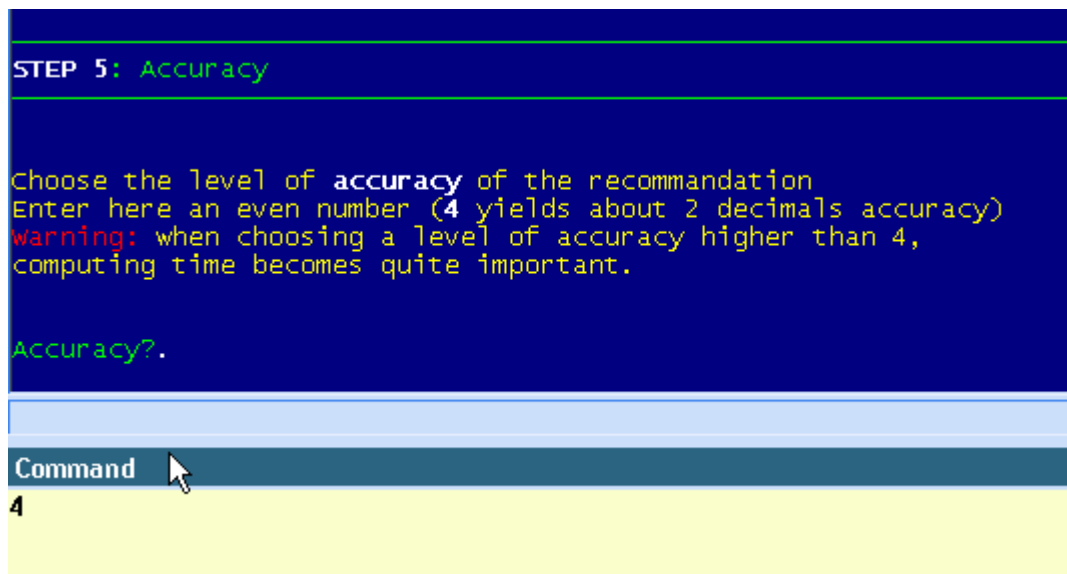
The data come from a paired experiment?.
```

```
Command
0
```

9. Choose the **level of accuracy** for the recommendation. You should enter here an even number. We suggest 4 here to get approximately 2 decimals accuracy.

Technical note:

The program is built on **two nested loops**. The level of accuracy determines the number of iterations for both the **inner loop** and the **outer loop**. When the sample is not balanced (i.e when the number of observations from treatment 1 differs from the number of observations for treatment 2) or when the data come from a paired sample, setting accuracy to 4 results in **100** iterations of the **outer loop**. Setting it to **6** results in **1000** iterations of the outer loop and hence implies a very important computing time. When the sample is balanced and when the data do not come from a paired experiment, it is ok to increase the level of accuracy. Otherwise, 4 is a good compromise between precision and computing time.



The screenshot shows a terminal window with a dark blue background. At the top, a green header reads "STEP 5: Accuracy". Below this, yellow text prompts the user to "Choose the level of accuracy of the recommendation" and "Enter here an even number (4 yields about 2 decimals accuracy)". A red "warning:" message follows, stating that accuracy higher than 4 increases computing time. The prompt "Accuracy?." is shown in green. Below the terminal window is a light blue bar with the label "Command" in white. A mouse cursor points to the input field, which contains the number "4" on a yellow background.

```
STEP 5: Accuracy

Choose the level of accuracy of the recommendation
Enter here an even number (4 yields about 2 decimals accuracy)
warning: when choosing a level of accuracy higher than 4,
computing time becomes quite important.

Accuracy?.

Command
4
```

10. Choose whether you want to include unitary costs for treatment 1 or for treatment 2. In our example we set both costs to zero.

STEP 6: Unitary Costs

Choose whether you want to include a cost for treatment 1

Cost for treatment 1?: 0

Choose whether you want to include a cost for treatment 2

Cost for treatment 2?:

Command
0

11. That's it. The program computes the recommendation and displays the results.

RESULTS		
PARAMETERS		
Regret:	.02213211	Cost for Treatment1: 0
Method:	Binomial Average Rule	Cost for Treatment2: 0
Paired sample:	No	
Range of values:	[0 ; 10]	
RECOMMENDATION		
Probability of choosing Treatment 1: 0.26300001		
Probability of choosing Treatment 2: 0.73699999		
.		

4. Errors

The program will detect most typical input mistakes and report an error message and a recommendation on how to correct it. However when a non standard error message is reported, the user can contact patrick.eozenou@iue.it for assistance. Any comments or suggestions on how to improve the code are also welcome.