



Inspirar para Transformar

Introdução ao Stata

Érica Emilia Leite
Apoio à Pesquisa

Stata é um software de estatística usado geralmente para análise econométrica. O Stata pode realizar desde tarefas simples como calcular média, desvio padrão, testes de hipótese, intervalos de confiança, até estatísticas mais complexas como análise de variância, regressão linear múltipla, regressão não-linear, regressão logística, análise de sobrevivência, regressão de Cox, etc.

O Stata utiliza uma linguagem muito avançada e tem recursos para trabalhar através da Internet como quase nenhum outro pacote estatístico. É possível trabalhar com um arquivo de dados que esteja num servidor do outro lado do planeta, fazer atualizações ("updates"), tanto de código básico como de novos procedimentos, bem como baixar dados e procedimentos que sejam de utilidade para nossos interesses, sem custo adicional. Assim, é possível se utilizar programas desenvolvidos por outros pesquisadores sem dificuldade.

O Stata exige que todo o conjunto de dados a ser analisado esteja em sua memória. Isto traz uma vantagem de velocidade, mas implica que talvez você precise de mais memória RAM em seu computador.

A última versão do programa aloca automaticamente uma quantidade de memória compatível com a dimensão da base de dados trabalhada. Adiante veremos também alguns recursos com os quais podemos otimizar o armazenamento dos dados.

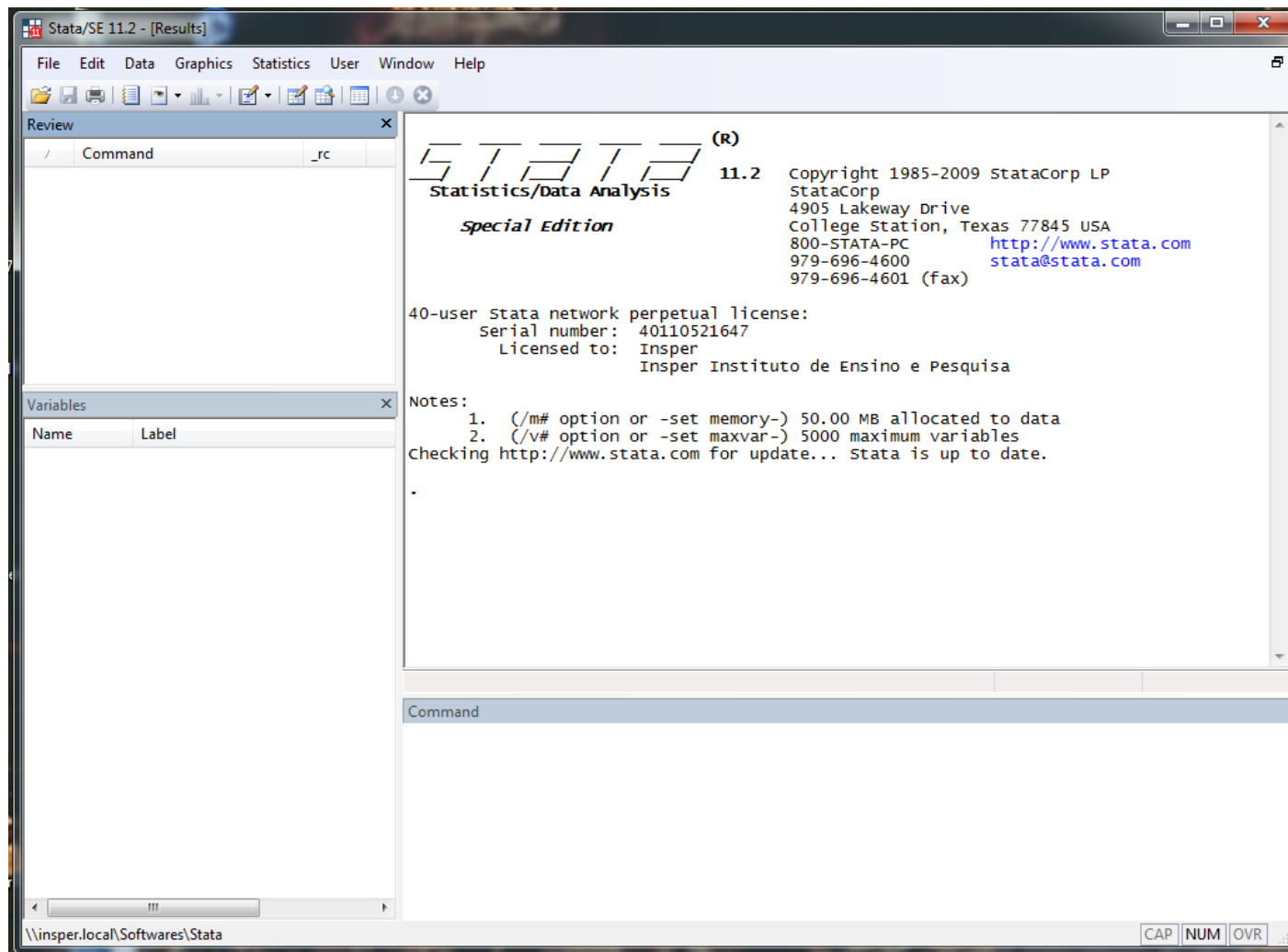
Para realocar mais memória, utilizamos o comando

```
set mem 700m
```

Acessando o programa

Caso o ícone do aplicativo Stata não esteja visível em sua área de trabalho, você pode acessá-lo através do endereço

```
\\insper.local\Softwares\Stata
```



Basicamente há 4 ambientes na janela do programa:

Review: Mostra os comandos recentes, para um acesso rápido

Variable: Mostra as variáveis dos dados utilizados

Results: Mostra as saídas dos comandos digitados

Command: Onde os comandos são digitados

A barra de ferramentas contém ícones que permitem abrir e salvar arquivos, imprimir resultados, e manipular as janelas. Algumas ferramentas importantes permitem-nos abrir os editores de Dados e Do-File. O Data Editor e o Data Browser fornecem uma visualização dos dados como numa planilha, não importando quão grande seja o conjunto de dados. O editor de Do-File, como discutiremos, permitirá construir um arquivo com comandos de Stata, e executá-los parcial ou inteiramente.

Diretório de trabalho

Para verificar qual o diretório atual, digite

```
pwd
```

Para mudar o diretório de trabalho e evitar digitar o caminho inteiro quando chamar por um arquivo, digite

```
cd "\\servidor-82\transfer\Intro Stata"
```

Utilize “ ” caso o novo diretório tenha espaços em branco.

E, para criar um novo diretório

```
mkdir seunome
```

Um problema

Estamos interessados em verificar qual o impacto dos anos de estudo sobre os ganhos (por hora trabalhada). Para tal, nosso arquivo (um exemplo retirado do livro *“Introdução à Econometria”*, de Jeffrey M. Wooldridge, contém dados referentes 526 trabalhadores.

Utilizaremos o Stata para obtermos algumas descrições estatísticas dos dados, incluindo um modelo de regressão linear simples.

Importando dados do Excel

O arquivo de nosso problema se encontra na pasta

```
\\servidor-82\transfer\Intro Stata
```

Podemos acessá-lo por duas vias

1. Utilizando o copy & paste através do Data Editor
2. Através do comando

```
insheet using wage1.csv
```

Salve o arquivo no seu diretório

```
cd seunome  
save wage
```

Explorando os dados

Para obter uma descrição geral dos dados, e o formato de cada variável, digite

```
describe  
list wage educ exper tenure in 1/5
```

Para algumas descrições estatísticas básicas, digite

```
summarize
```

Uma tabela de frequências pode ser obtida com o comando

```
tab female
```

Todos estes comandos possuem opções (designadas após uma vírgula), que permitem ampliar análise. Por exemplo

```
table female, contents(freq mean wage mean numdep)
```

Para facilitar a leitura das variáveis, podemos acrescentar-lhes descrições e etiquetas. Para adicionar a descrição, utilizamos o comando

```
label variable female "=1 if female"
```

Criei um Do-File com a lista das descrições das variáveis, [rotula_wage1.do](#) (está no diretório acima do criado por você); podemos abrí-lo e executá-lo com auxílio do editor de Do-File.

Podemos também renomear a variável female para gender, e colocar-lhe uma etiqueta:

```
rename female gender
```

```
label define genero 0 "Homens" 1 "Mulheres"
```

```
label values gender genero
```

```
tab gender
```

Editando os dados

Podemos criar novas variáveis através do comando gen

```
gen llwa= log(wage)  
edit llwa lwage
```

Podemos eliminar variáveis através do comando drop

```
drop llwa
```

Criando variáveis através da combinação de outras variáveis:

```
gen fem_constr = (gender==1 & construc==1)  
table fem_constr, contents(freq mean wage mean educ )
```

Explorando os dados

Você pode obter um sumário estatístico com o comando

```
tabstat wage educ exper tenure, s(mean sd var min max)
```

Digite **help tabstat** para uma lista completa da estatística descritiva.

Histogramas

```
histogram wage, frequency
```

Podemos visualizar gráficos de dispersão através do comando `twoway scatter` e suas opções

```
twoway scatter wage educ
```

Exemplo de uso com opções

```
twoway scatter lwage educ, by(gender)
```

■
Para realizarmos uma regressão, utilizamos o comando

```
regress [dependent variable] [independent variable(s)]  
regress y x
```

Para o nosso estudo de caso, queremos verificar se os ganhos por hora trabalhada relacionam-se a nossas variáveis (anos de educação, anos de experiência, anos no emprego atual).

$$wage = \beta_0 + \beta_1 educ + \beta_2 educ + \beta_3 exper + \beta_4 tenure + \varepsilon$$

Para estimarmos este nosso modelo, basta então digitarmos

```
regress wage educ exper tenure
```

Podemos visualizar a lista dos resultados da regressão através do comando

```
ereturn list
```

Podemos armazenar algum resultado de uma regressão executada gerando uma nova variável;

```
gen coef_reg1_educ = _b[educ]
```


Para exibir os resultados de várias regressões em uma tabela-resumo, podemos utilizar os comandos `eststo` e `esttab`:

```
regress wage educ exper tenure
eststo model1
regress wage educ exper tenure if gender==0 & married==1
eststo model2
regress wage educ exper tenure if fem_constr==1
eststo model3
regress lwage educ exper tenure
eststo model4
esttab, r2 ar2 se scalar(rmse)
```

Há dois modos de gerar valores previstos para Y

Opção 1: utilizando generate após a regressão

```
regress wage educ
generate wage_predict = _b[_cons] + _b[educ]* educ
```

Opção 2: utilizando predict após a regressão

```
regress wage educ
predict wage_predict
label variable wage_predict "wage predicted"

scatter wage wage_predict
```

Há alguns modos de guardar o registro do trabalho efetuado:

Através do arquivo log

```
clear  
cd "meu caminho"  
capture log close  
log using nome.log, replace  
...  
log close
```

Através do Do-File

Utilize o help do Stata para estender as possibilidades de uso de um comando, e mesmo para aprender novos comandos e funcionalidades. Para procurar a sintaxe de um comando ainda desconhecido, por exemplo, descobrir como fazer análises pos-estimatições; podemos começar tentando a palavra pós (em inglês, claro):

search post

Há muita informação na rede sobre funcionalidades do Stata, e na própria página do programa você pode encontrar informações úteis:

- <http://www.stata.com/training/>
- <http://www.ats.ucla.edu/stat/stata/whatstat/whatstat.htm>

Obrigada a todos.

Bom divertimento!