

AULA 1

Análise de Dados Legislativos e Eleitorais Utilizando o Programa Stata

Professor: Ernesto Friedrich de Lima Amaral

Email: eflamaral@gmail.com

Site do curso: www.ernestoamaral.com/stata20091.html

Data: 18/05/2009

Horário: 18:00–21:00

Entendendo o Stata

O Stata possibilita gerenciamento de dados, análise estatística e elaboração de gráficos. Existem programas para tornar o Stata mais amigável para aqueles que não gostam de trabalhar com comandos, mas o curso será baseado no estilo padrão de comandos.

Principais janelas do Stata: *Review*, *Variables*, *Results*, *Command* e *Do-file Editor*.

Bancos de dados em Stata possuem extensão ".dta", e programas (*syntax*) possuem extensão ".do".

O Stata trabalha com os dados copiando-os na memória RAM. Em virtude disso, bancos de dados grandes podem ser de difícil gerenciamento no Stata. Quando um banco é aberto, nenhuma mudança é realizada até que você salve o banco. O fato de usar uma cópia dos dados é importante porque:

- Quando se utiliza o comando "use arquivo", os dados são copiados para a memória do computador, e o arquivo original é fechado.
- Você pode fazer o que quiser com os dados na memória, e a cópia permanente continuará a mesma em seu disco.
- A única forma de mudar uma cópia permanente dos dados é utilizando o comando "save".
- Além disso, se algum erro é reportado, nenhuma mudança é realizada no banco que se encontra na memória.

Inserindo comentários

Para inserir comentários no Stata, simplesmente utilize asterisco (*) antes do texto:

```
*Isso é um tido como um comentário no Stata
```

Ou utilize os símbolos "/" e "*/", antes e depois do comentário:

```
/*Essa é uma outra forma de  
inserir comentário no Stata*/
```

Comandos que todos devem saber

O Stata tem vários comandos (por volta de 800). Fora os comandos estatísticos que interessam a muitas pessoas, aqui estão listados alguns comandos que todos devem saber do que se tratam:

Categoria	Comandos no Stata
1. Obtendo ajuda	search, findit, help
2. Atualizando o Stata pela internet	update, net, ado, news
3. Interface de sistema operacional	pwd, cd, sysdir, mkdir, dir, erase, copy, type
4. Começando, salvando e terminando o trabalho	clear, set mem, set more, log, notes, exit
5. Importando dados para o Stata	infix, input, infile, insheet
6. Usando e salvando dados do disco	use, save, append, merge, compress
7. Reportando dados básicos	describe, codebook, list, browse, edit, count, inspect, summarize, table, tabulate
8. Manipulação de dados	generate, replace, egen, rename, drop, keep, sort, encode, decode, order, by, reshape
9. Formatando	format, label
10. Conveniência	display

1. Obtendo ajuda

O comando "**findit**" procura informações de uma determinada palavra-chave. O "findit" faz uma procura completa, incluindo as procuras de "**help**" que procura por comando existente no Stata; "**search**" que procura pela palavra-chave na internet; e "**net search**" que procura por pacotes para instalação no site www.stata.com.

2. Atualizando o Stata pela internet

Para fazer atualização de comandos e procedimentos utilize o comando "**update**".

Para mostrar uma breve lista de recentes notícias e informações sobre o Stata, provenientes do site www.stata.com, digite "**news**".

Instalar um pacote de comandos:

```
net install nomepacote, from(diretório_ou_url)]
```

Utilize o comando "**ado**" para listar pacotes instalados.

Para descrever pacotes instalados, digite:

```
ado describe
```

3. Interface de sistema operacional

Alguns comandos básicos no Stata são importantes para lidar com o programa, e se assemelham aos comandos do DOS e UNIX:

- **pwd** Mostrar diretório em que se encontra.
- **cd** Mudar para diretório indicado.
- **sysdir** Mostrar diretório em que o Stata está instalado.
- **mkdir** Criar diretório em seu computador.
- **dir** Ver o conteúdo do diretório em que se encontra.
- **erase** Apagar arquivo no diretório especificado.
- **copy** Copiar arquivos para mesmo diretório ou diretório diferente.
- **type** Mostrar conteúdo de arquivo na tela do Stata.

Diretório em que o Stata está instalado:

```
. sysdir
  STATA:  C:\Programas\Stata10_SE\
  UPDATES: C:\Programas\Stata10_SE\ado\updates\
  BASE:   C:\Programas\Stata10_SE\ado\base\
  SITE:   C:\Programas\Stata10_SE\ado\site\
  PLUS:   c:\ado\plus\
  PERSONAL: c:\ado\personal\
  OLDPLACE: c:\ado\
```

- UPDATES Stata dá preferência a arquivos nesse diretório.
- BASE Diretório com comandos originais.
- PLUS Armazena tudo oriundo da internet, também chamado de STBPLUS.
- PERSONAL Arquivos pessoais.
- c:\ado Comandos são armazenados em sub-pastas com a primeira letra do comando.

Nunca é bom utilizar o diretório em que o Stata está instalado para trabalhar, já que arquivos de programas podem ser apagados, ou arquivos pessoais podem ser removidos em uma atualização do programa. O recomendado é utilizar um diretório para cada projeto.

Você pode criar diretórios com o comando **mkdir** no Stata, ou no próprio Windows Explorer. Vamos supor que queremos criar um diretório para esse curso no drive C:\. Esses são os procedimentos:

```
cd C:\
mkdir cursodcp
cd cursodcp
mkdir dados
cd dados
```

A partir de agora, assim que começar o Stata, mude para o diretório recentemente criado:

```
cd C:\cursodcp\dados
```

Se você criar um diretório com espaço entre as palavras, é necessário utilizar aspas:

```
cd "C:\curso dcp\dados"
```

Você também pode ver o conteúdo de diretórios, apagar arquivos, copiar arquivos, e examinar o conteúdo de um arquivo.

Listar o conteúdo do diretório atual:

```
dir
```

Listar os arquivos que possuem a extensão ".dta":

```
dir *.dta
```

Listar os arquivos que possuem a extensão ".dta" e nomes que começam com "prog" e possuem mais dois caracteres:

```
dir prog?? .dta
```

Apagar um arquivo no diretório atual:

```
erase meuarquivo.xyz
```

Copiar arquivo no diretório atual. Isso é importante para realizar cópia de segurança, antes de realizar mudanças em um arquivo:

```
copy meuarquivo.abc meuarquivo.bak
```

Copiar um arquivo para um diretório diferente:

```
copy D:\dados\meuarquivo.dta C:\cursodcp\dados\meuarquivo.dta
```

Ver o conteúdo de um arquivo:

```
type meusdados.raw
```

4. Começando, salvando e terminando o trabalho

Geralmente o comando "clear" inicia um programa ".do" para limpar a memória do Stata:

```
clear
```

Estabelecendo a quantidade de memória alocada para o Stata:

```
set mem 100m
```

Para que essa quantidade de memória seja permanente toda vez que abrir o Stata:

```
set mem 100m, perm
```

Se grandes tabelas ou regressões forem ser geradas pelos seus comandos, é bom digitar o comando abaixo para que o programa não paralise a tela:

```
set more off
```

Abrindo um arquivo ".log" para salvar o trabalho. O ideal é escrever esse comando no começo do arquivo ".do":

```
log using "C:\cursodcp\log\aula1.log", text replace
```

Salvando os comandos e tabelas geradas. O ideal é escrever esse comando no final do arquivo ".do":

```
log close
```

Para salvar somente os comandos, fazer um arquivo ".do" no "Do-file Editor". Se a janela "Review" tiver sido usada, clique com o botão direito do mouse para copiar o conteúdo e colar em um arquivo ".do".

Colocando avisos no banco de dados:

```
notes: criar rótulos em português para variáveis p501-p511
```

e

```
notes p201: verificar se variável foi codificada corretamente
```

Listar todos avisos criados no banco de dados:

```
notes
```

Encerrar o Stata:

```
exit
```

Se houver um banco de dados aberto no Stata, o ideal é digitar o seguinte comando para encerrar o programa sem salvar os dados:

```
exit, clear
```

5. Importando dados para o Stata

Importando dados de um arquivo texto que possui formato fixo para as colunas. Exemplos da "Demographic Health Survey" e do Censo:

```
infix v005 038-045 v012 062-063 v013 064 using "C:\DHS96\brir31f1.dat"
```

OU

```
infix v0001 001-005 v0002 006-008 v0003 009-010 using "C:\Censo\1960\rs.txt"
```

Importando dados manualmente para o Stata:

```
input nestu cuesti pais legis partido entrev
```

```
51 1 51 307 4 1
```

```
51 2 51 307 4 2
```

```
51 3 51 307 4 2
```

```
51 4 51 307 4 2
```

```
51 5 51 307 4 1
```

```
51 6 51 307 4 2
```

```
51 7 51 307 4 1
```

```
51 8 51 307 4 1
```

```
51 9 51 307 4 1
```

```
51 10 51 307 4 1
```

```
end
```

6. Usando e salvando dados do disco

Com o comando "use", você abre um banco de dados no Stata, mas não muda o diretório:

```
use c:\cursodcp\dados\Argentina51.dta
```

Se o diretório atual for c:\cursodcp\dados, simplesmente digite:

```
use Argentina51.dta
```

Já que os bancos de dados em Stata usam a extensão ".dta", você pode abrir o banco sem digitar a extensão:

```
use Argentina51
```

Se algum outro banco já estiver aberto, é preciso utilizar a opção "clear" para limpar a memória do Stata:

```
use c:\cursodcp\dados\Argentina51.dta, clear
```

Como vimos, o comando para abrir um banco é "use". Se não houver nenhum banco aberto, utiliza-se:

```
use arquivo.dta
```

Se o desejo for descartar tudo que estiver na memória, utiliza-se:

```
use arquivo.dta, clear
```

Para salvar um banco pela primeira vez, utiliza-se:

```
save arquivo.dta
```

Se o arquivo já existir, e você quiser gravar o banco por cima do anterior:

```
save arquivo.dta, replace
```

Ou seja, um banco é salvo somente com o comando "save", tornando difícil perder os dados originais. Mesmo se o comando "save" não for usado intencionalmente, o Stata recusará gravar o banco por cima do original, se a opção "replace" não for colocada.

7. Reportando dados básicos

Para mostrar o sumário do banco de dados, com nome, tipo e rótulo das variáveis:

```
describe
```

e

```
describe p501-p511
```

Para mostrar o sumário mais detalhado das variáveis do banco:

```
codebook
```

Outra forma de mostrar informações sobre as variáveis do banco, com ilustração de quantidade de números negativos, positivos e "missings", além de um pequeno gráfico de ramos e folhas (com distribuição da variável entre os seus valores) é com o seguinte comando:

```
inspect
```

Para contar quantos legisladores pensam que os riscos para consolidação da democracia são muito altos em decorrência da crise econômica (p502=4) e da dívida externa (p506=4):

```
count if p502==4 & p506==4
```

Para mostrar o banco na tela de resultados do Stata, utilize o comando "list". Para mostrar as variáveis que indicam a opinião dos legisladores do partido UCR (partido=2) sobre as possíveis vantagens de um regime democrático (p201 e p202 originárias da pergunta P2 na página 1 do questionário), digite o comando:

```
list p201 p202 if partido==2
```

```

      |      p201      p202 |
      +-----+-----+
21. | la posib  la posib |
22. | la prote  la resol |
23. | la prote  la posib |
24. | la prote  la posib |
25. | la prote  el respe |
      +-----+-----+
26. | la prote  la resol |
27. | la prote  la posib |
28. | la prote  el respe |
29. | la prote  la mayor |
30. | la prote  la posib |
      +-----+-----+
31. | la prote  la mayor |
32. | el creci  la posib |
33. | la prote  la posib |
34. | la prote  la posib |
35. | la posib  la prote |
      +-----+-----+
36. | la prote  la posib |
37. | la mayor  el creci |
97. | la resol  la mayor |
98. | el respe  el creci |
99. | el respe  la posib |
      +-----+-----+
100. | la posib  la prote |
101. | la posib  la resol |
102. | la prote  la posib |
103. | la prote  el respe |
      +-----+-----+

```

Se o rótulo da pergunta dificultar a visualização, utilize a opção "nolabel":

```
list p201 p202 if partido==2, nolabel
```

```

      | p201  p202 |
      +-----+-----+
21. |    9    3 |
22. |    2    8 |
23. |    2    9 |
24. |    2    6 |
25. |    2    5 |
      +-----+-----+
26. |    2    8 |
27. |    2    3 |
28. |    2    5 |
29. |    2    4 |
30. |    2    9 |
      +-----+-----+
31. |    2    4 |
32. |    1    9 |
33. |    2    9 |
34. |    2    9 |
35. |    9    2 |
      +-----+-----+
36. |    2    9 |
37. |    4    1 |
97. |    8    4 |
98. |    5    1 |
99. |    5    6 |
      +-----+-----+
100. |    3    2 |
101. |    6    8 |
102. |    2    6 |
103. |    2    5 |
      +-----+-----+

```

Para mostrar o banco em uma tela separada, utilize o comando "browse". Como no exemplo anterior:

```
browse p201 p202 if partido==2
browse p201 p202 if partido==2, nolabel
```

Para visualizar todo o banco, simplesmente digite:

```
browse
```

Para editar um banco, utilize o comando "edit" da mesma forma que o "list" e "browse". O comando "edit" pode ser acessado com o ícone "Data Editor" da barra de ferramentas.

Sem a utilização da opção "nolabel", as variáveis que aparecem em preto não possuem rótulos, aquelas que aparecem em azul possuem o rótulo visualizado, e as que apresentam a cor vermelha são variáveis nominais (*string* ou *character*).

É possível obter estatísticas básicas de variáveis com o comando "summarize" que é o mesmo que "sum". Podemos analisar as respostas quanto aos temas que podem representar uma ameaça ao risco de consolidação da democracia (p501 a p511):

```
summarize p501-p511
```

Variable	Obs	Mean	Std. Dev.	Min	Max
p501	104	1.596154	.6464125	1	4
p502	104	3.144231	.8409226	1	4
p503	104	2.048077	.9791993	1	4
p504	103	3.048544	.8561043	1	4
p505	104	2.509615	.9552633	1	4
p506	103	3	.9801961	1	4
p507	104	3.086538	.7896242	1	4
p508	103	2.203883	.7965517	1	4
p509	103	3.38835	.8311699	1	4
p510	103	2.84466	.9472073	1	4
p511	103	2	.8631906	1	4

Note acima que essas variáveis variam de 1 (Nada) a 4 (Muito). É bom estar ciente que poderia haver casos iguais a 8 (Não sabe=NS) e 9 (Não respondeu=NC), o que enviesaria a análise.

Uma análise de percentil poderia ser feita com a opção "detail" que é o mesmo que "d":

```
summarize p501-p511, detail
```

Variáveis nominais são automaticamente retiradas do sumário pelo Stata.

Para ordenar um banco de dados por uma variável ou conjunto de variáveis, utilize a opção "sort":

```
sort partido
```

Para realizar uma tabela de uma variável, utilize o comando "tabulate", que é o mesmo que "tab". As opções de "nolabel", "if" e "missing" também podem ser utilizadas:

```
tab p501 if p501!=1, nolabel missing
```

Para realizar um cruzamento entre partido político e a variável p501:

```
tab partido p501
```

Para gerar tabelas simples para cada uma das variáveis listadas:

```
tab1 partido p501-p511
```

Para gerar tabelas com cruzamentos entre duas variáveis para todas combinações possíveis:

```
tab2 partido p501 p502
```

Para obter a média, desvio padrão e frequência da variável p501 em cada partido:

```
tab partido, summarize(p501)
```

8. Manipulação de dados

Transformar variável numérica em variável nominal (*string*), utilize o comando "decode":

```
decode partido, generate(nomepart)
```

Transformar variável nominal em variável numérica, utilize o comando "encode":

```
encode nomepart, generate(numpart)
```

Verificando o que foi feito:

```
browse partido nomepart numpart
```

O comando "tostring" não exige a criação de uma nova variável para transformar uma variável numérica para nominal. Isso só funciona se a variável não tiver rótulo (*label*):

```
tostring nestu, replace
```

O comando "destring" pode converter todas variáveis nominais para numéricas em um único comando:

```
destring, replace
```

Alocar determinadas variáveis para o começo do banco:

```
order p501-p511
```

Mover uma variável para uma posição anterior à segunda variável indicada no comando:

```
move entrev pais
```

Colocar em ordem alfabética as variáveis listadas e movê-las para o começo do banco:

```
aorder legisbis partido departam
```

Para excluir uma variável ou conjunto de variáveis do banco, utilize a opção "drop". Foi escolhido o número do estudo (nestu), pois ele não varia no banco de dados:

```
drop nestu
```

Note que só excluimos a variável da cópia do banco que está na memória RAM, e não a que está originalmente no disco.

Para remover observações do banco, utilize o comando "drop" com indicação da primeira e última observações a serem retiradas. Por exemplo, vamos remover as observações de 5 a 10:

```
drop in 5/10
```

Essa remoção pode também ser feita de outras formas. Por exemplo, vamos remover aqueles que consideram que as relações entre as forças armadas e o governo (p501=1) não apresentam risco para a consolidação da democracia:

```
drop if p501==1
```

OU

```
drop if p501!=2 & p501!=3 & p501!=4 & p501!=.
```

OU

```
drop if p501<2
```

OU

```
drop if p501<2 & p501>=1
```

No Stata, o "missing" é o maior número (ao contrário do SAS), por isso o comando "drop if p501<2" não exclui os valores de p501 iguais a "missing".

É possível também escolher por manter determinadas variáveis no banco com o comando "keep":

```
keep if p501!=1
```

OU

```
keep if p501==2 | p501==3 | p501==4 | p501==.
```

Somente recapitulando:

== igual

!= diferente

> maior

>= maior/igual

< menor

<= menor/igual

& E

| OU