

AULA 2

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: 19/05/2009

Horário: 18:00–21:00

Semelhanças nas janelas do SPSS e Stata

SPSS e Stata possuem janelas para resultados (*Results*), variáveis (*Variables*), visualização dos dados (*Data Browser*, *Data Editor*), edição dos comandos (*Do-file Editor*).

O Stata ainda tem uma janela para digitar comandos rapidamente (*Command*) e comandos que foram digitados desde que o programa foi aberto (*Review*).

Organizando os diretórios

Antes de tudo vamos organizar os diretórios da aula na unidade C:\, utilizando o Windows Explorer:

C:\cursodcp	Diretório principal
C:\cursodcp\dados	Diretório para banco de dados
C:\cursodcp\doc	Diretório para documentos diversos
C:\cursodcp\graf	Diretório para gráficos
C:\cursodcp\log	Diretório para resultados dos comandos
C:\cursodcp\prog	Diretório para programas ".do"

Manipulação de dados (continuação da Aula 1)

Para gerar uma variável, utilize o comando "generate", que é o mesmo que "gen" ou "g":

```
gen p500=p501+p502+p503+p504+p505+p506+p507+p508+p509+p510+p511
```

ou

```
egen p500=rowtotal(p501 p502 p503 p504 p505 p506 p507 p508 p509 p510 p511)
```

Note que a opção "rowtotal" considera os valores "missings" iguais a zero. Como o comando "gen" simplesmente retira os casos que têm "missings" em algumas variáveis, o resultado é diferente entre os dois comandos acima. Para gerar o mesmo resultado, poderíamos fazer os seguintes comandos:

```
drop if p501==. | p502==. | p503==. | p504==. | p505==. | p506==. | p507==. |  
p508==. | p509==. | p510==. | p511==.
```

e

```
gen p500=p501+p502+p503+p504+p505+p506+p507+p508+p509+p510+p511  
tab p500
```

ou

```
egen p500=rowtotal(p501 p502 p503 p504 p505 p506 p507 p508 p509 p510 p511)  
tab p500
```

Construindo uma variável que agrega informações de partido e departamento do legislador:

```
sort partido departam
egen pardep=group(partido departam)
browse partido departam pardep
```

Calculando o número de partidos por departamento:

```
sort departam
by departam: egen npartido=max(partido)
browse partido departam npartido
```

Número de legisladores por departamento e por número de partidos no departamento:

```
tab departam npartido
```

circunscripción	npartido					Total
	1	2	3	4	5	
capita federal	0	0	7	0	0	7
provincia de buenos a	0	0	0	0	23	23
corrientes	0	1	0	0	0	1
entre ríos	0	0	0	7	0	7
chaco	0	3	0	0	0	3
santa fé	0	0	0	0	13	13
formosa	1	0	0	0	0	1
misiones	0	0	0	1	0	1
jujuy	0	4	0	0	0	4
salta	3	0	0	0	0	3
santiago del estero	0	2	0	0	0	2
tucumán	0	0	0	2	0	2
córdoba	0	0	0	7	0	7
la pampa	0	0	0	4	0	4
catamarca	0	2	0	0	0	2
la rioja	0	3	0	0	0	3
mendoza	0	0	0	5	0	5
san juan	2	0	0	0	0	2
río negro	0	0	0	2	0	2
chubut	2	0	0	0	0	2
neuquén	0	0	0	5	0	5
santa cruz	0	3	0	0	0	3
tierra del fuego	0	0	3	0	0	3
Total	8	18	10	33	36	105

Gerando uma variável que indica a quantidade de legisladores em cada partido:

```
tab partido
```

partido político	Freq.	Percent	Cum.
pdv	18	20.45	20.45
rn	16	18.18	38.64
udi	25	28.41	67.05
ppd	15	17.05	84.09
ps	9	10.23	94.32
prsd	5	5.68	100.00
Total	88	100.00	

A informação da tabela acima pode ser inserida no banco, seguindo os seguintes passos:

```
gen deputado=1
sort partido
by partido: egen ndeputado=count(deputado)
browse partido deputado ndeputado
tab ndeputado
```

ndeputado	Freq.	Percent	Cum.
5	5	5.68	5.68
9	9	10.23	15.91
15	15	17.05	32.95
16	16	18.18	51.14
18	18	20.45	71.59
25	25	28.41	100.00
Total	88	100.00	

Para renomear uma variável, digite "rename", o nome atual da variável, e o nome que deseja. Vamos renomear a variável que informa o número do questionário:

```
rename cuesti quest
```

Para criar uma nova categorização da variável que informa sobre o grau de confiança que o legislador tem sobre o poder judiciário (p701) na vida pública argentina, podemos agrupar os valores de "muito" (p701=4) e "bastante" (p701=3) em uma categoria, e os valores de "pouco" (p701=2) e "nenhuma" (p701=1) em outro grupo.

Primeiramente, vamos saber se essa variável tem valores de "não sabe", "não respondeu" ou "missing":

```
tab p701, missing
```

OU

```
tab p701, missing nolabel
```

Posteriormente, criamos uma variável em que todos valores são missing:

```
gen p701g=.
```

Depois, fazemos as substituições em "p701g", conforme os valores de "p701":

```
replace p701g=0 if p701==1 | p701==2
replace p701g=1 if p701==3 | p701==4
```

Verificar se a nova variável foi criada corretamente:

```
tab p701 p701g, missing
```

Formatando

O comando "format" estabelece o formato de determinadas variáveis no banco. Use "help format" para ver as diferentes opções. Mudando o ponto da variável peso para vírgula:

```
format peso %9,6fc
```

Colocando ou mudando o rótulo de uma variável:

```
label variable cuesti "Número do questionário"
```

e

```
label variable p701g "Poder judiciário (agrupado)"
```

Criando rótulos de categorias de variáveis:

```
label define escala 1 "Nada" 2 "Pouco" 3 "Bastante" 4 "Muito"
```

e

```
label define escalag 0 "Nada/Pouco" 1 "Bastante/Muito"
```

Colocando os rótulos das categorias para determinada variável:

```
label values p501-p511 escala
```

e

```
label values p701g escalag
```

Conveniência

"Display" mostra valores nominais e numéricos de expressões escalares.

Número de observações no banco de dados:

```
display _N
```

Fazendo cálculos manuais:

```
display 439/23
```

Utilizando os comandos "summarize" e "display" em conjunto:

```
summarize p501
```

```
display as text "Média de p501 = " as result r(mean)
```

Inclusão de observações de outros bancos (*append*)

Vamos selecionar algumas variáveis dos bancos da Argentina, Chile e Guatemala e depois utilizar o comando "append" para juntar todos esses bancos (inclusão de observações).

Limando o banco da Argentina e salvando em outro diretório:

```
use "C:\cursodcp\dados\Argentina51.dta", clear
```

```
keep nestu cuesti pais legis partido departam entrev p501-p511 p701-p712
```

```
mkdir "C:\cursodcp\dados\append"
```

```
save "C:\cursodcp\dados\append\argentina.dta", replace
```

Limando o banco do Chile e salvando em outro diretório:

```
use "C:\cursodcp\dados\Chile42.dta", clear
```

```
keep nestu cuesti pais legis partido departam entrev p501-p511 p701-p712
```

```
save "C:\cursodcp\dados\append\chile.dta", replace
```

Limando o banco da Guatemala e salvando em outro diretório:

```
use "C:\cursodcp\dados\Guatemala52.dta", clear
```

```
keep nestu cuesti pais legis partido departam entrev p501-p511 p701-p712
```

```
save "C:\cursodcp\dados\append\guatemala.dta", replace
```

Agora é preciso abrir o banco limpo da Argentina e juntar com os bancos limpos do Chile e Guatemala:

```
use "C:\cursodcp\dados\append\argentina.dta", clear nolabel
```

```
append using "C:\cursodcp\dados\append\chile.dta", nolabel
```

```
append using "C:\cursodcp\dados\append\guatemala.dta", nolabel
```

```
save "C:\cursodcp\dados\append\todos.dta", replace
```

Acima foi utilizada a opção "nolabel" para excluir os rótulos das categorias de todos os bancos. Isso foi feito porque determinadas variáveis possuem rótulos diferentes entre os países e utilizariam os rótulos da Argentina (primeiro banco aberto).

Você também poderia excluir a opção "nolabel" e, após fazer a junção dos bancos, você poderia gerar novas variáveis para que se saiba exatamente o que determinadas variáveis significam. Esse seria o melhor procedimento a ser adotado. Por exemplo, o entrevistador "1" na Argentina não é o mesmo que o entrevistador "1" no Chile. É possível gerar uma nova variável para que a informação de entrevistador (entrev) não seja interpretada incorretamente, já que ela possui informações de três países (pais). Isso também deveria ser feito para a variável partido, por exemplo.

No caso da variável de identificação do entrevistador, esse procedimento poderia ser feito de uma forma mais eficiente (menos comandos):

```
sort pais entrev
egen entrevistador=group(pais entrev)
```

Isso também pode ser feito de uma maneira menos eficiente (mais linhas de comandos), mas em que você saberá exatamente o que significam os códigos da nova variável:

```
gen entrevistador=.
replace entrevistador=1 if pais==42 & entrev==1
replace entrevistador=2 if pais==42 & entrev==2
replace entrevistador=3 if pais==42 & entrev==3
replace entrevistador=4 if pais==51 & entrev==1
replace entrevistador=5 if pais==51 & entrev==2
replace entrevistador=6 if pais==51 & entrev==3
replace entrevistador=7 if pais==52 & entrev==1
replace entrevistador=8 if pais==52 & entrev==2
replace entrevistador=9 if pais==52 & entrev==3
```

Inclusão de variáveis de outros bancos (*merge*)

Suponha que temos dois bancos da Argentina com variáveis distintas e gostaríamos de agrupá-los (*merge*). Na realidade, vamos montar dois bancos separados, e depois juntá-los pela variável de identidade (*questi*):

```
use "C:\cursodcp\dados\Argentina51.dta", clear
keep cuesti pais legis partido departam entrev
sort cuesti
mkdir "C:\cursodcp\dados\merge"
save "C:\cursodcp\dados\merge\argentina1.dta", replace

use "C:\cursodcp\dados\Argentina51.dta", clear
keep cuesti p1 p201-p202 p3 p401-p402 p501-p511
sort cuesti
save "C:\cursodcp\dados\merge\argentina2.dta", replace

use "C:\cursodcp\dados\merge\argentina1.dta", clear
merge cuesti using "C:\cursodcp\dados\merge\argentina2.dta"
tab _merge
drop _merge
save "C:\cursodcp\dados\merge\argentina3.dta", replace
```

A variável "_merge" é criada automaticamente pelo Stata e pode assumir os seguintes resultados:

```
_merge=1  observações somente do banco de dados mestre (master data)
_merge=2  observações somente do banco de dados secundário (using data)
_merge=3  observações dos dois bancos (master e using)
```

Quebra de linha

No caso de comandos muito longos (tais como os utilizados para gerar gráficos), é bom inserir quebras de linha que indicam ao Stata que o comando não foi finalizado na linha atual, e que continua na linha abaixo. Podemos usar três barras (*///*) no final da linha para fazer essa indicação ao Stata. Exemplos como este estão colocados no tópico seguinte sobre gráficos.

Gráficos

De uma forma geral, é bom olhar o menu *Graphics* para explorar os diversos tipos de gráficos elaborados pelo Stata. Aqui vou colocar alguns exemplos:

– Gráfico de barras:

Média de risco da crise econômica (p502) e da dívida externa (p506) para a democracia argentina por partido político:

```
label drop partido
label define partido 1 "PJ" 2 "UCR" 3 "ARI" 4 "Outros" 5 "PS"
label values partido partido

graph bar (mean) p502 p506, over(partido) ///
///
title("Média de risco da crise econômica (p502) e da dívida externa (p506)" ///
"para a democracia argentina por partido político, segundo legisladores" ///
, size(medsmall)) ///
///
yttitle("Média") ///
///
legend(title(Variáveis de risco para democracia, size(medsmall)) ///
label(1 "Crise econômica") label(2 "Dívida externa"))
graph export "C:\cursodcp\graf\med502-506.wmf", replace
```