

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/stata20092ufpe.html

Data: 21 e 24/11/2009

Alguns comentários

Copiando tabelas do Stata para Word ou Excel (texto, tabela, HTML, figura).

Utilizando os menus do Stata, quando não se sabe os comandos, e salvando a sintax.

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).

Limpendo o banco da Argentina e salvando em outro diretório:

```
use "C:\cursoufpe\dados\Argentina51.dta", clear
keep nestu cuesti pais legis partido departam entrev p501-p511 p701-p712
mkdir "C:\cursoufpe\dados\append"
save "C:\cursoufpe\dados\append\argentina.dta", replace
```

Limpendo o banco do Chile e salvando em outro diretório:

```
use "C:\cursoufpe\dados\Chile42.dta", clear
keep nestu cuesti pais legis partido departam entrev p501-p511 p701-p712
save "C:\cursoufpe\dados\append\chile.dta", replace
```

Limpendo o banco da Guatemala e salvando em outro diretório:

```
use "C:\cursoufpe\dados\Guatemala52.dta", clear
keep nestu cuesti pais legis partido departam entrev p501-p511 p701-p712
save "C:\cursoufpe\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:\cursoufpe\dados\append\argentina.dta", clear nolabel
append using "C:\cursoufpe\dados\append\chile.dta", nolabel
append using "C:\cursoufpe\dados\append\guatemala.dta", nolabel
save "C:\cursoufpe\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 (*cuesti*):

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

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

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

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 argetina 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:\cursoufpe\graf\med502-506.wmf", replace
```

– Histogramas:

Posicionamento ideológico de Carlos Menem, segundo legisladores argentinos:

```
gen menem=p1406
replace menem=. if p1406==99

histogram menem, ///
discrete frequency ytitle(Frequência) xtitle(Carlos Menem) ///
title("Histograma da escala esquerda/direita de Carlos Menem," ///
"segundo legisladores", size(medsmall))
graph export "C:\cursoufpe\graf\histmenem.wmf", replace
```

– Gráfico de caixa (box-plot):

Distribuição dos legisladores, segundo prática religiosa:

```
gen pratica=p60b
replace pratica=. if p60b==99
label variable pratica "Prática religiosa"

graph hbox pratica, ///
title("Distribuição dos legisladores, segundo prática religiosa" ///
, size (medsmall))
graph export "C:\cursoufpe\graf\boxpratrel.wmf", replace
```

– Gráfico de dispersão (scatterplot):

Idade do legislador pela prática religiosa:

```
gen idade=p63
replace idade=. if p63==99
label variable idade "Idade"

twoway scatter pratica idade, ///
title("Gráfico de dispersão da idade do legislador pela prática religiosa" ///
, size (medsmall))
graph export "C:\cursoufpe\graf\scatrelid.wmf", replace
```

Reorganização do arranjo (formato) dos bancos de dados

O comando "reshape" muda o formato dos bancos de dados, segundo a necessidade do estudo em questão. Este é o exemplo da ajuda do Stata:

```
(wide form)
i      ..... x_ij .....
id sex  inc80  inc81  inc82
-----
1   0   5000   5500   6000
2   1   2000   2200   3300
```

```
(long form)
i  j      x_ij
id year sex  inc
-----
1  80     0  5000
1  81     0  5500
1  82     0  6000
2  80     1  2000
2  81     1  2200
2  82     1  3300
```

O comando para mudar o formato dos bancos de "wide" (amplo) para "long" (dados em painel) é:
`reshape long inc, i(id) j(year)`

O comando para mudar o formato dos bancos de "long" (dados em painel) para "wide" (amplo) é:
`reshape wide inc, i(id) j(year)`

Suponha que queremos saber qual a média geral que os partidos argentinos receberam na escala de esquerda (1) à direita (10) por parte dos legisladores. As variáveis que indicam as escalas esquerda/direita dos partidos vão de "p1301" a "p1305". Vamos transformar essas seis variáveis em uma variável única (p), repetindo os deputados nas linhas:

```
keep cuesti p1301-p1305
reshape long p, i(cuesti) j(partido)
replace p=. if p==98 | p==99
```

O número de observações no novo banco é igual ao número de observações originais (105) multiplicado por cinco partidos:

```
display _N
display 105*5
```

Vamos estimar a média das escalas esquerda/direita de cada partido:

```
tab partido, sum(p) mean
```

partido	Summary of p Mean
1301	6.6041667
1302	5.8157895
1303	3.4042553
1304	8.2
1305	3.9569892
Total	5.5158151

Agora queremos colocar os valores da tabela acima em nosso banco como uma nova variável:

```
sort partido
by partido: egen esqdir=mean(p)
```

Lembrem-se que usando o comando "egen" podemos calcular a média de uma série de variáveis com a opção "rowmean" (média na linha). No caso acima, o "egen" foi utilizado para estimar a média (mean) de uma única variável (p) por categorias de outra variável (partido).

Somente como exercício, vamos retornar esse banco para o formato "wide":

```
drop esqdir
reshape wide p, i(cuesti esqdir) j(partido)
```

Utilização do comando foreach para recodificação e criação de grupos de variáveis

As variáveis "p2901" a "p2910" indicam a opinião do legislador sobre o grau de intervenção que o Estado deveria assumir frente a uma série de funções tradicionais. Com o intuito de reclassificar as respostas "não sabe" (8) e "não respondeu" (9) para "missing" (.), poderíamos usar todas essas linhas de comando:

```
replace p2901=. if p2901==8 | p2901==9
replace p2902=. if p2902==8 | p2902==9
replace p2903=. if p2903==8 | p2903==9
replace p2904=. if p2904==8 | p2904==9
replace p2905=. if p2905==8 | p2905==9
replace p2906=. if p2906==8 | p2906==9
replace p2907=. if p2907==8 | p2907==9
replace p2908=. if p2908==8 | p2908==9
replace p2909=. if p2909==8 | p2909==9
replace p2910=. if p2910==8 | p2910==9
```

No entanto, podemos simplesmente escrever:

```
foreach x of varlist p29?? {
  replace `x'=. if `x'==8 | `x'==9
}
```

Depois de recodificar os valores "não sabe" e "não respondeu" das variáveis p2901 a p2910, podemos gerar novas variáveis para informar as médias das opiniões dos legisladores (sobre o grau de intervenção do Estado em certas atividades) por partido:

```
sort partido
foreach x of varlist p2901-p2910 {
  by partido: egen med`x' = mean(`x')
}

browse partido p2901-p2910 med*
browse partido p2901-p2910 med*, nolabel
```

Criação de bancos de dados agrupados por categorias de variáveis (*collapse*)

Vamos criar e salvar um banco de dados que vai indicar o número de deputados por departamento e partido político. Abra o banco de dados original, antes de rodar os comandos abaixo:

```
gen ndeputado=1
sort departam partido
collapse (count) ndeputado, by(departam partido)
save "C:\cursoufpe\dados\ndep.dta", replace
browse departam partido ndeputado
```

A opção `peso` pode ser usada em uma tabela simples somente para certificar que o número de deputados foi calculado corretamente:

```
tab partido [fweight=ndeputado]
```

Em seguida, vamos abrir o banco de dados original novamente para calcular as médias das variáveis p2901–p2910 (opiniões sobre grau de intervenção do Estado) e p3101–p3114 (opiniões sobre relevância de problemas na Argentina) por departamento e partido. Antes de tudo, temos que reclassificar os "não sabe" e "não respondeu":

```
foreach x of varlist p29?? p31?? {
  replace `x'=. if `x'==8 | `x'==9
}

sort departam partido
collapse (mean) p29?? p31??, by(departam partido)
save "C:\cursoufpe\dados\p29p31.dta", replace

merge departam partido using "C:\cursoufpe\dados\ndep.dta"
tab _merge
drop _merge
save "C:\cursoufpe\dados\p29p31ndep.dta", replace
```

Alocação de informações entre registros

Vamos supor que queremos colocar algumas respostas do deputado mais velho de cada partido no registro dos outros deputados daquele mesmo partido. O pressuposto é de que o deputado mais velho teria influência sobre os demais deputados de seu partido em temas polêmicos. A informação de deputado mais velho poderia ser substituída por deputado mais rico, deputado líder do partido, e assim por diante (dependendo da disponibilidade dos dados e de suas hipóteses). Em nosso exemplo, testaremos se o deputado mais velho tem influência sobre os demais deputados de seu partido em temas polêmicos ("p34": opinião sobre imposto; "p35": opinião sobre privatizações; e "p36": opinião sobre Mercosul).

Esse tipo de exercício é muito utilizado em estudos que alocam informações dos pais (renda, educação e outros) para as crianças no mesmo domicílio.

Primeiramente vamos recodificar a variável idade e ordenar o banco de dados original de forma crescente por partido e decrescente por idade (p63) com o comando "gsort":

```
use "C:\cursoufpe\dados\Argentina51.dta", clear
replace p63=. if p63==99
gsort partido -p63
browse partido p63
```

Depois vamos criar uma variável dicotômica que terá valor igual a "1" para os deputados mais velhos em cada partido:

```
egen depvel=tag(partido)
browse partido p63 depvel
```

Vamos ordenar o banco por partido, já que mais adiante iremos juntá-lo a outro banco. É bom salvar esse banco com outro nome:

```
sort partido
save "C:\cursoufpe\dados\argoriginal.dta", replace
```

Agora vamos criar um pequeno banco que vai ter informações somente dos deputados mais velhos de cada partido ("depvel" igual a 1). Além disso, queremos somente as variáveis "p34" (opinião sobre imposto), "p35" (opinião sobre privatizações) e "p36" (opinião sobre Mercosul):

```
keep if depvel==1
keep partido p34 p35 p36
```

Como iremos juntar esse novo banco ao banco com todos os deputados, precisamos renomear as variáveis para que elas não sejam gravadas por cima das outras:

```
rename p34 p34vel
rename p35 p35vel
rename p36 p36vel
browse
```

Lembre-se de ordenar o banco por partido e salvá-lo com outro nome:

```
sort partido
save "C:\cursoufpe\dados\argvelho.dta", replace
```

Temos que abrir novamente o banco que salvamos anteriormente e juntá-lo (merge) ao banco dos deputados mais velhos:

```
use "C:\cursoufpe\dados\argoriginal.dta", clear
merge partido using "C:\cursoufpe\dados\argvelho.dta"
tab _merge
drop _merge
```

Por fim, seria interessante colocar valores "missing" para as novas variáveis ("p34vel", "p35vel" e "p36vel") quando se trata do deputado mais velho. Isso é feito porque as novas variáveis só fazem sentido de serem comparadas com as antigas ("p34", "p35" e "p36"), se o registro for dos demais deputados:

```
replace p34vel=. if depvel==1
replace p35vel=. if depvel==1
replace p36vel=. if depvel==1

save "C:\cursoufpe\dados\argalocado.dta", replace
```

Teste de Qui-quadrado (*Chi-square test*)

Antes de realizar o teste de qui-quadrado, é necessário recodificar as variáveis de interesse para colocar os valores "não sabe" e "não respondeu" como "missing":

```
*Opinião dos deputados sobre imposto
tab p34, missing nolabel
gen imposto=p34
replace imposto=. if p34==8 | p34==9
tab imposto, missing nolabel

*Opinião dos deputados sobre privatização
tab p35, missing nolabel
gen privatizar=p35
replace privatizar=. if p35==8 | p35==9
tab privatizar, missing nolabel

*Opinião dos deputados sobre Mercosul
tab p36, missing nolabel
gen mercosul=p36
replace mercosul=. if p36==8 | p36==9
tab mercosul, missing nolabel

*Escala ideológica (esquerda/direita) do próprio legislador
tab p58, missing nolabel
gen esqdir=p58
replace esqdir=. if p58==99
tab esqdir, missing nolabel
```

O teste de qui-quadrado é usado quando se deseja saber se há relação entre duas variáveis categóricas. A opção "chi2" é usada com o comando "tabulate" para obter o teste estatístico e seu valor "p":

```
tabulate esqdir imposto, chi2
tabulate esqdir privatizar, chi2
tabulate esqdir mercosul, chi2
```

Observamos que há correlação significativa no nível de 95% entre o posicionamento ideológico (esquerda/direita) dos legisladores e suas respostas sobre imposto (p34), privatização (p35) e Mercosul (p36).

Teste de "t" (*t-test*)

O teste de "t" (*t-test*) permite verificar se a média de uma variável com distribuição normal difere significativamente de um valor hipotético. Por exemplo, podemos testar se a média da auto-classificação dos deputados na escala esquerda/direita (esqdir) difere significativamente de 5,5:

```
histogram esqdir

ttest esqdir=5.5
```

A média da variável "esqdir" é 4,525, o que é estatisticamente significante diferente do valor de 5,5. Poderíamos concluir que nesse grupo de legisladores, eles se auto-classificam mais próximos à esquerda (ou centro-esquerda) do que a média geral da escala do questionário.

Correlação

O comando de correlação (`corr`) é útil para testar se há relação linear entre duas ou mais variáveis com intervalos normalmente distribuídos. Podemos testar nossas duas variáveis independentes ("imposto" e "privatizar"), mesmo sabendo que elas não têm distribuição normal:

```
hist imposto
hist privatizar

corr imposto privatizar
```

Também é possível estimar a significância dessa correlação:

```
pwcorr imposto privatizar, sig
```

Na realidade esse teste de correlação também pode ser realizado para variáveis dicotômicas (*dummy variables*). Isso seria mais aplicável ao nosso caso, já que "privatizar" é uma variável categórica que tem cinco valores, e "imposto" tem dois valores. Podemos realizar a correlação entre "imposto" e cada um dos cinco valores de "privatizar".

Primeiro, poderíamos criar manualmente variáveis dicotômicas para cada valor de "privatizar":

```
gen priv1=.
  replace priv1=0 if privatizar!=1 & privatizar!=.
  replace priv1=1 if privatizar==1
gen priv2=.
  replace priv2=0 if privatizar!=2 & privatizar!=.
  replace priv2=1 if privatizar==2
gen priv3=.
  replace priv3=0 if privatizar!=3 & privatizar!=.
  replace priv3=1 if privatizar==3
gen priv4=.
  replace priv4=0 if privatizar!=4 & privatizar!=.
  replace priv4=1 if privatizar==4
gen priv5=.
  replace priv5=0 if privatizar!=5 & privatizar!=.
  replace priv5=1 if privatizar==5

corr imposto priv1 priv2 priv3 priv4 priv5
pwcorr imposto priv1 priv2 priv3 priv4 priv5, sig
```

Também podemos usar as opções "xi" e "i." para calcular a correlação entre "imposto" e "privatizar" rapidamente. Nesse caso, uma categoria da variável "privatizar" é omitida:

```
xi: corr imposto i.privatizar
xi: pwcorr imposto i.privatizar, sig
```

O Stata também gera automaticamente variáveis dicotômicas com o uso do comando "tabulate", levando em consideração as categorias de uma determinada variável. No caso abaixo, serão criadas cinco variáveis ("privatizar1", "privatizar2", "privatizar3", "privatizar4" e "privatizar5"):

```
tab privatizar, gen(privatizar)
```

```
corr imposto privatizar1 privatizar2 privatizar3 privatizar4 privatizar5
pwwcorr imposto privatizar1 privatizar2 privatizar3 privatizar4 privatizar5, sig
```

	imposto	privat~1	privat~2	privat~3	privat~4	privat~5
imposto	1.0000					
privatizar1	0.2730	1.0000				
	0.0203					
privatizar2	-0.0973	-0.2607	1.0000			
	0.4159	0.0180				
privatizar3	-0.0157	-0.3397	-0.5715	1.0000		
	0.8959	0.0018	0.0000			
privatizar4	0.0732	-0.0767	-0.1291	-0.1682	1.0000	
	0.5413	0.4934	0.2479	0.1310		
privatizar5	-0.1612	-0.1294	-0.2178	-0.2837	-0.0641	1.0000
	0.1761	0.2465	0.0494	0.0098	0.5674	

Ao elevar ao quadrado a correlação e multiplicar por 100, podemos determinar a porcentagem de variabilidade conjugada entre as variáveis.

Podemos realizar esse cálculo para os deputados que pensam que "todos serviços públicos deveriam ser privatizados" ("privatizar" igual a 1):

```
display (0.2730*0.2730)*100
```

O resultado indica que a variável "imposto" tem uma variação de 7,45% com a opção "1" da variável "privatizar".

Vamos fazer esse cálculo para os deputados que pensam que o correto é "privatizar todos os serviços públicos, com exceção dos que tiveram uma incidência para a maioria da população" ("privatizar" igual a 3), em que a correlação com "imposto" foi de -0,0157:

```
display (-0.0157*-0.0157)*100
```

Isso significa que a variável "imposto" tem uma variação de apenas 0,025% com a opção "3" da variável "privatizar".

De uma forma geral, as correlações entre "imposto" e "privatizar" são pequenas, algo corroborado pela baixa significância dos testes estatísticos realizados com comando "pwwcorr". Em outras palavras, por não estarem correlacionadas, essas variáveis poderiam ser inseridas conjuntamente em um modelo de regressão como variáveis independentes.

A variável de opinião sobre imposto, dos que privatizariam todos os serviços públicos (privatizar=1) e dos que não privatizariam nenhum serviço público (privatizar=5) estão fortemente correlacionadas com a escala ideológica (a qual poderia ser a variável dependente do modelo):

```
pwcorr esqdir imposto, sig
```

```
pwcorr esqdir privatizar1 privatizar2 privatizar3 privatizar4 privatizar5, sig
```

Análise de Variância (ANOVA)

Podemos utilizar o comando de análise de variância (anova) para estimar a relação entre uma variável independente categórica e uma variável dependente com distribuição normal. Esse comando testa se há diferenças entre as médias da variável dependente nos diversos níveis da variável independente.

Nesse caso, usaremos a auto-classificação do deputado (esqdir) como variável dependente, e sua opinião sobre imposto (p34) como variável independente:

```
twoway scatter esqdir imposto
```

```
twoway (lfit esqdir imposto) (scatter esqdir imposto)
```

```
anova esqdir imposto
```

```
Number of obs =      82      R-squared      = 0.1475
Root MSE      = 1.27353    Adj R-squared = 0.1368
```

Source	Partial SS	df	MS	F	Prob > F
Model	22.4438782	1	22.4438782	13.84	0.0004
imposto	22.4438782	1	22.4438782	13.84	0.0004
Residual	129.751244	80	1.62189055		
Total	152.195122	81	1.87895212		

- *Partial SS* é o mesmo que "partial sum of squares", ou soma parcial dos quadrados.
- *Df* é o mesmo que "degrees of freedom", ou graus de liberdade.
- *MS* é o mesmo que "mean square", ou média dos quadrados.
- *F* é o mesmo que "F Value", ou "F-test", ou teste de "F".

A média de auto-classificação na escala esquerda/direita é significativamente diferente entre as duas opiniões sobre imposto.

O mesmo exercício pode ser usado para a variável independente que indica a opinião dos deputados sobre privatização (p35), e sua relação com a variável dependente "esqdir":

```
twoway scatter esqdir privatizar
twoway (lfit esqdir privatizar) (scatter esqdir privatizar)
```

```
anova esqdir privatizar
```

Source	Partial SS	df	MS	F	Prob > F
Model	38.0773827	4	9.51934568	5.62	0.0005
privatizar	38.0773827	4	9.51934568	5.62	0.0005
Residual	125.289706	74	1.69310413		
Total	163.367089	78	2.09444985		

A média da escala esquerda/direita também difere significativamente entre as cinco opiniões sobre privatização. Porém, não sabemos em quais níveis da variável de privatização essas médias variam.

Para saber a média da escala esquerda/direita para cada opinião de privatização, é possível combinar o comando "tabulate" com "summarize":

```
tabulate privatizar, summarize(esqdir)
```

OU

```
tab privatizar, sum(esqdir)
```

privatizar	Summary of esqdir		
	Mean	Std. Dev.	Freq.
1	5.9	1.3703203	10
2	4.4583333	1.2846643	24
3	4.6470588	1.2764143	34
4	4.3333333	.57735027	3
5	3	1.5118579	8
Total	4.5696203	1.4472214	79

Vemos que os deputados favoráveis a privatizar todos os serviços públicos (privatizar=1) são os que mais se classificam à direita (média=5,9).

Nos resultados gerados pelo comando "anova", os valores do teste "F" (*F-test*) dos modelos (*Model*) são os mesmos das variáveis independentes ("imposto" e "privatizar"). Isso aconteceu porque realizamos a análise com somente uma variável independente por comando. Se outras variáveis forem inseridas, o teste "F" será diferente no *Model* e nas variáveis independentes. Esse é o caso de:

```
anova esqdir imposto
anova esqdir privatizar
anova esqdir imposto privatizar
```

Passos para realização de inferência estatística

A intenção foi de mostrar como os diferentes passos para realização de inferência estatística podem ser realizados no Stata. A continuação desse exercício é justamente a exploração das possibilidades de realização de regressão estatística neste pacote. Estes foram os passos abordados:

- Organização do banco de dados.
- Recodificação de variáveis.
- Histograma.
- Gráfico de dispersão (variáveis independentes e variável dependente).
- Exploração de outros gráficos.
- Testes de significância: testes de correlação e significância entre variáveis.
- Regressão estatística.