

# Introdução ao Stata

Microeconomia Empírica

Laísa Rachter  
(laisarachter@gmail.com)

8 e 10 de Janeiro de 2017

## Material e Instalação

- Computadores da FGV.
- Instalar no Laptop versão teste.
- Download dos Microdados (site IBGE)

# Descrição

- Isso é uma **introdução** ao Stata.
- Vamos assumir nenhum/muito pouco conhecimento de Stata
- Objetivos:
  - 1 Familiarizar-se com a interface do Stata
  - 2 Importar dados
  - 3 Gerar novas variáveis e transformações
  - 4 Computar estatísticas, construir gráficos e rodar regressões

# Interface do Stata

The screenshot displays the Stata 12.1 interface with the following components:

- Review Window:** Shows the command history.
 

```

      1 use me...
      2 sum ha...
      3 tab sex
      4 hist age
      
```
- Results Window:** Displays the output of the commands.
 

```

      . use newgss.dta

      . sun happy
      
```

Variable	Obs	Mean	Std. Dev.	Min	Max
happy	217	1.886452	.6888996	1	3

```

      . tab sex
      respondents
      sex      Freq.    Percent    Cum.
      male     114      52.53      52.53
      female   103      47.47      100.00
      Total    217      100.00

      . hist age
      (bins=14, start=18, width=4.2142857)
      
```
- Variables Window:** Lists the variables in the dataset.
 

Name	Label
marital	marital status
age	age of respondent
educ	highest year of sc...
sex	respondents sex
inc	respondents income
happy	general happiness
region	
- Graph Window:** Displays a histogram of the variable 'age of respondent'. The x-axis is labeled 'age of respondent' and ranges from 20 to 80. The y-axis is labeled 'Density' and ranges from 0 to 0.04. The histogram shows a distribution peaking around age 35.

**Do-file editor:** The top right window shows the Stata do-file editor with the following code:

```

/*this is a preamble, always have it!*/
*
*Intro to stata
*Intro Zahn Fall 2011
*Revised: 10/21/11
*-----
*-----

use gss.dta

lookfor income

label define sexlabel 1 "Male" 2 "Female"
codebook sex
label value sex "sexlabel"

sun age
label variable region "Area of the United States"
rename region region1
codebook marital

**Start your Stata session
cd "~/StataIntro"
/*change directory/
/*cd "~/StataIntro"

```

## Do-Files

- Você pode digitar todos os comandos no do-file que você usaria na janela de comando.
- MAS...o Do-file permite **salvar** seus comandos
- Nunca usar a janela de comando ou menus para mudar o seu dado.
- Salvar os comandos no Do-file permite manter um histórico de tudo o que você fez com os dados:
  - Permite fácil replicação
  - Permite retornar e rodar novamente os comandos, análises e fazer modificações

## Help do Stata

- A forma mais fácil de obter ajuda - apenas digite **help** seguido do assunto ou comando, e.g., **help regress**
- Tente o **search** se o comando não for encontrado
- Um google em “Stata [Topic]” também ajuda!
- UCLA website: <http://www.ats.ucla.edu/stat/Stata/>
- Cameron e Trivedi (2009)

## Sintaxe Geral

- A maior parte dos comandos segue o mesmo princípio geral:
  - **Command varlist, options**
  - ATENÇÃO: em alguns casos, se você digitar o comando e não especificar a variável, o Stata vai realizar o comando para todas as variáveis na sua base de dados
- A melhor forma de encontrar a sintaxe específica de cada comando é em help-files

## Comentários

- Comece com um comentário descrevendo o seu Do-file e use comentários ao longo
- Comentário na linha ou em blocos

```
// comment  
describe var  
/*  
comment block comment block comment block comment  
block comment block comment block  
*/
```

- Use / para quebrar uma lista de variáveis em múltiplas linhas

```
// break commands over multiple lines  
describe var1 var2 var2 ///  
var4 var5 var6
```



## Primeiro Passo

- Abra um novo Do-file

```
// change directory  
cd "C://Users/dataclass/Desktop/StataIntro"  
// start a log file to record your stata session  
log using myStataLog.txt, text replace  
// Pause / resume logging with "Log on" / "Log off"  
// Close log with "Log close"
```

- How to start every do-file :
  - 1 Descreva o que o Do-file faz
  - 2 Mude o diretório
  - 3 Comece um log file
  - 4 Abra o dado
  - 5 Execute comandos: Manipulação de Dados, Estatísticas, Etc.
  - 6 Salve os dados com um novo nome (se aplicável)



## Comandos Data file

- Extensão: `.dta`
- Abrir/Salvar dados com `use` e `save`.

```
cd dataSets
// open the gss.dta data set
use gss.dta
// saving your data file:
save newgss.dta, replace
// the "replace" option tells stata it's OK to
// write over an existing file
```

```
cd dataSets
/nfs/home/I/izahn/StataIntro/dataSets

use gss.dta

save newgss.dta, replace
(note: file newgss.dta not found)
file newgss.dta saved
```

- Se o nome do path tem espaços, você precisa colocar " ".

## Browser: onde está o dado?

- Data Browser ([Browser](#))
- Data Editor ([Edit](#))
  - Usar o data editor não é recomendado
- Sempre faça qualquer mudança nos dados no seu Do-file
- Evite fazer mudanças manualmente no Editor

# Importar Dados

Stata/SE 13.0 - [Results]

File Edit Data Graphics Statistics User Window Help

Open... Ctrl+O  
Save Ctrl+S  
Save As... Ctrl+Shift+S  
View...  
Do...  
Filename...  
Change Working Directory...  
Log  
Import  
Export  
Print  
Example Datasets...  
Recent Datasets  
Exit

Excel spreadsheet ("xls","xlsx")  
Text data (delimited, ".csv", ...)  
Text data in fixed format  
Text data in fixed format with a dictionary  
Unformatted text data  
SAS XPORT  
Haver Analytics database  
ODBC data source  
XML data

Stata/SE 13.0  
Copyright 1995-2013 StataCorp LP  
StataCorp  
4905 Lakeway Drive  
College Station, Texas 77845 USA  
800-STATA-PC <http://www.stata.com>  
979-696-4600 [stata@stata.com](mailto:stata@stata.com)  
979-696-4601 (fax)

Variables

Variable	Label
There are no items to show.	

Properties

Variables

Name	Label	Type	Format	Value Label	Notes

Data

Filename	Label	Notes	Variables	Observations	Size
			0	0	0

Command

C:\Users\Sony\Documents

CAP | NUM | OVR

# Infile

The screenshot displays the Stata/SE 13.0 software interface. The main window shows the 'Review' tab with a command window containing the text 'There are no items to show.' and a command prompt with the command 'rc'. A dialog box titled 'infile - Import text data in fixed format' is open, showing the 'Main' tab. The dialog box contains the following fields and options:

- Dictionary filename: [ ] Browse...
- Text dataset filename: [ ] Browse...
- ☒ Replace data in memory
- Buttons: OK, Cancel, Submit

The 'Variables' window on the right shows 'There are no items to show.' The 'Properties' window on the bottom right shows the following information:

Variables	
Name	
Label	
Type	
Format	
Value Label	
Notes	

Data	
Filename	
Label	
Notes	
Variables	0
Observations	0
Size	0

The taskbar at the bottom shows the Windows Start button, a search bar with the text 'Digite aqui para pesquisar', and several application icons. The system tray on the right shows the date and time: 19:42, 06/01/2018.

# Infile

```
infile using "pnad2014pes.dct", using(PES2014.TXT)
sort uf v0102 v0103
save pnadPes_2014.dta, replace
clear
```

- OU DataZoom: <http://www.econ.puc-rio.br/datazoom/instal.html>

## Comandos Usuais

- Comandos para checar seus dados:
  - describe //labels, storage type etc.
  - sum //statistical summary (mean, sd, min/max etc.)
  - codebook //storage type, unique values, labels
  - list // print actual values
  - tab // (cross) tabulate variables
  - browse // view the data in a spreadsheet-like window

# Operadores

Operator	Meaning
==	equal to
!=	not equal to
>	greater than
>=	greater than or equal to
<	less than
<=	less than or equal to
&	and
	or



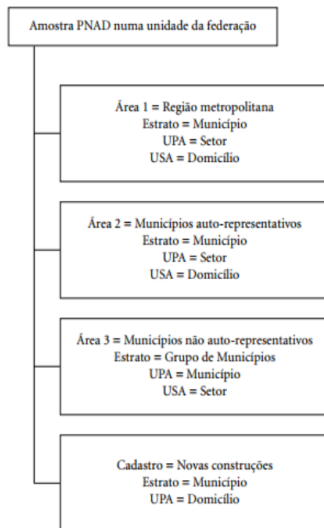
## Comandos Usuais

- Renomear: `rename`
- Codificar: `recode`
- Substituir: `replace`
- Gerar: `gen`
- combinar dados: `merge`
- Empilhar dados: `append`
- regressão OLS: `reg y X, r`
- Gerar Tabela: `outreg2`

## Amostras Complexas

- Quando trabalhamos com grandes populações, não conseguimos listar todos os componentes da população ou mesmo ter acesso a todos eles.
- Estágios:
  - 1 População existente é dividida em estratos (ex: região administrativa).
  - 2 Cada estrato contém USA (ex: municípios) retirados com probabilidades pré determinadas  $\pi_{1,s}$
  - 3 Em cada USA existem múltiplas UPAs (ex: domicílios) retiradas com probabilidade  $\pi_{2,s}$
- A probabilidade de uma UPA ser retirada é  $\pi_i$  e  $w_i = \frac{1}{\pi_i}$
- Cada observação  $i$  representa  $w_i \geq 1$  unidades da população.

# PNAD



# PNAD

- E daí?
- Dado das amostras das PNADs não podem ser tratados como se fossem observações independentes e identicamente distribuídas
- Procedimentos usuais de análise disponíveis nos pacotes estatísticos padrão assumem dados i.i.d.
- Pode afetar precisão das estimativas
- Alternativa: Usar o pacote [svy](#)

# SVY

- Definir o plano amostral:

*svyset[pw = v4729], strata(v4617)psu(4618)*

- Lidar com estratos com PSUs únicos (dropar ou outras opções Stata)
- Antes dos comandos usar:

*svy : comando*

- Subpopulação:

*svy, subpop(var) : comando*

## Equação Minceriana

- Vamos estimar os retornos a educação usando a Equação de Mincer (1974):

$$\log(y_i) = \beta_0 + \beta_1 \times S_i + \beta_2 \times E_i + \beta_3 \times E_i^2 + u_i \quad (1)$$

- $\log(y_i)$  log rendimentos
- $S_i$  Anos de Estudo
- $E_i$  Idade (proxy de experiência)

## Exercício:

- 1 Estime a equação acima por OLS usando a PNAD
- 2 Acrescente controles demográficos (cor, sexo...)
- 3 Estime (1) para alguma sub-população de seu interesse

## Passo a Passo

- Inicie um log file
- Escolha um ano da PNAD e importe os dados no Stata ([infile](#))
- Escolher e tratar variáveis de interesse e gerar estatísticas descritivas (Média, d.p., etc)
- Estime  $\beta$  por OLS
- Gere uma tabela com os resultados da regressão ([outreg2](#))
- Encerre o log
- Entregar: Do-file comentado, log file e resultados (tabela e um parágrafo interpretando os resultados obtidos).