

PNAD Contínua

Marcelo Prudente e Rafael Giacomini

10 de abril de 2018

- 1 PNAD CONTÍNUA
- 2 PNADC - Dados Educacionais
- 3 Modelagem com survey
- 4 Concentração de renda
- 5 APÊNDICE: BAIXAR DADOS PNADC COM LODOWN

PNAD CONTÍNUA

PNAD CONTÍNUA

- A lógica computacional da análise dos dados da PNAD Contínua é a mesma daquela aplicada à PNAD.
- Porém, há maior facilidade de acessar os seus dados.

PNAD Contínua

- Para a PNAD Contínua, é possível baixar os dados diretamente já com o tratamento dos dados para a estrutura de *survey*.
- Assim, é necessário baixar o pacote **PNADcIBGE**.
- Você pode ver exemplos instrutivos na página de exemplos do autor do pacote, Douglas Braga.

```
# instalar pacote
install.packages("PNADcIBGE")
# carregar pacote
library(PNADcIBGE)
```

PNADCs

- As PNADs Contínuas têm dois tipos de microdados:
 - Trimestrais
 - Anuais

Vantagens do PNADcIBGE

- 1 Os dados já são baixados com o desenho de survey.
- 2 É possível baixar dados anuais e trimestrais.
- 3 É possível baixar apenas algumas variáveis de interesse.
- 4 Variáveis carregam seus rótulos

Baixando os dados da PNADC

- Não há esforço para baixar a PNADC trimestral:

```
# Especifica o ano e o trimestre
pnadc = get_pnadc(year = 2017, quarter = 4)
class(pnadc)
```

- Tampouco a anual:

```
# Especifica o ano e a entrevista
pnadc_anual = get_pnadc(year = 2016, interview = 1)
```


Carregando dados Off-line [1]

- Entrar no (**ftp do ibge**)[`ftp://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Trimestral/Microdados/`]
- Baixar os arquivos da PNAD e o `Dicionario_e_input`

Carregando dados Off-line [2]

```
# diretório
setwd("C:/dados/PNADC")

# baixar dados
dados_pnadc <- read_pnadc("PNADC_042017.txt", "Input_PNADC_tri

# incluir rótulos
dados_pnadc <- pnadc_labeller(dados_pnadc, "dicionário_das_var

# desenho de survey
dados_pnadc <- pnadc_design(dados_pnadc)

# transformar em tbl_svy
dados_pnadc <- as_survey(dados_pnadc)
```

PNADC - Dados Educacionais

Estimativas educacionais

- A referência do IBGE para estatísticas educacionais é a PNAD Contínua do segundo trimestre de 2016.
- Portanto, se a intenção for reproduzir os dados oficiais sobre o tema, é necessário baixar os dados

```
# baixar dados
pnadc2 = get_pnadc(year = 2016, quarter = 2)
# transformá-los em tbl_svy
pnadc2 <- as_survey(pnadc2)
```

Exemplo e exercícios

```
pnadc2 <- pnadc2 %>%  
  mutate(regiao = as.factor(substr(UPA, 1, 1)))  
  
analfabetos_uf <- pnadc2 %>% filter(V2009 >=15) %>%  
  group_by(UF, V3001)%>%  
  summarise(analf = survey_total( na.rm = T), analf_perc =
```

- Exercícios

Baixar pacotes

```
# Pacotes exigidos
pacotes <- c("survey", "ggplot2", "srvyr")

# carregar lista de uma só vez
lapply(pacotes, require, character.only = TRUE)
```

- Para usar o **srvyr**, é necessário transformar o desenho de survey em **tbl_svy**.

```
# Pacotes exigidos
pnadc <- as_survey(pnadc)
```

Algumas análises

- Tamanho da população estimada no trimestre

```
# total
dadosPNADc %>% summarise(survey_total(one, na.rm = T))
# por uf
dadosPNADc %>%
  group_by(UF)%>%
  summarise(pop = survey_total( one ,na.rm = T))
```

Transformando tabelas

```
# renda por sexo e raca
tot_sexo_raca <- pnadc %>%
  group_by(V2007, V2010) %>%
  summarise(total = survey_mean(VD4016, na.rm = T))

# spread
tot_sexo_raca <- tot_sexo_raca %>%
  select(V2007:total) %>%
  spread(V2007, total)

# gather
tot_sexo_raca <- gather(r, V2007, value, - V2010)
```


Modelagem com survey

Teste de Hipóteses

- Vamos testar se a diferença salarial entre homens e mulheres tem significância estatística.

```
# VD4020 - renda  
# V2007 - sexo  
svytest(VD4020 ~ V2007, pnadc)
```

Regressão Linear

- A renda está associada ao nível educacional, à raça, à idade e ao sexo?
- Para regressão com surveys, utilizamos o *svyglm*.

```
# o ~ separa a variável dependente das independentes
modeloLin <- svyglm(VD4020 ~ VD3001 + V2010 + V2009 + V2007,
summary(modeloLin)
```

Regressões Logísticas

- O que está associado à conclusão de um curso de graduação.

```
modelo <- svyglm(V3007 ~ V2007 + V2010 + V2009 + regioao, pnado  
summary(modelo)
```

Concentração de renda

convey

- O Pacote convey permite estimar diversas medidas de concentração de renda para dados provenientes de pesquisas com planos amostrais complexos.

```
library(convey)
pnadc <- convey_prep(pnadc)
```

gini

- Para medir a concentração de renda no país por meio do índice de gini, podemos:

```
giniHab <- svygini(~VD4020, pnadc, na.rm = TRUE)
giniHab

giniUF <- svyby(~VD4020, by = ~UF,
                dadosPNADc, svygini, na.rm = TRUE)
giniUF

gini_regiao <- svyby(~VD4020, by = ~ regiao,
                    dadosPNADc, svygini, na.rm = TRUE)
giniUF
```

Curva de Lorenz

- A Curva de Lorenz é um gráfico utilizado para relacionar a distribuição relativa de renda pelas pessoas. A área entre essa curva e a reta identidade, é uma das formas de definir o coeficiente de Gini.

```
curvaLorenz <- svylorenz(~VD4020, pnadc,  
                        quantiles = seq(0, 1, .05), na.rm =
```


APÊNDICE: BAIXAR DADOS PNADC COM LODOWN

PNADC - criar catálogo

```
library(lodown)
# Pesquisa Nacional de Amostra de Domocílios Contínua
pnadc_cat <-
  get_catalog( "pnadc" ,
    output_dir = file.path( path.expand( "~" ) , "PNADC" ) )
```

PNADC - apenas o último trimestre disponível

```
str(pnadc_cat)
# selecionar periodo desejado
pnadc_cat <- pnadc_cat %>%
  filter(year == "2017" & quarter == "04")
```

PNADC - baixar

```
# baixar arquivos  
lodown("pnadc", pnadc_cat)
```

PNADC - ler dados no R

```
# é o diretorio correto
getwd()
# listar arquivos
list.files()

# ler arquivo
library(readr)
pnadc <- read_rds("pnadc 2017 04.rds")
```

Pre-estratificação

- Na PNADC as variáveis de estratificação estão explícitas. Confira o dicionário de variáveis.
 - **ids = unidades primária de amostragem (upa)**
 - **strata = estrato**
 - **weights = peso amostral (v1027)**

Pre-estratificação

- Lembre de **NÃO UTILIZAR AMOSTRA PRÉ-ESTRATIFICADA**

```
# pre-estratificação
pre_stratified <-
  svydesign(
    ids = ~ upa ,
    strata = ~ estrato ,
    weights = ~ v1027 ,
    data = pnadc ,
    nest = TRUE
  )
```

Pós-estratificação

- A pós-estratificação da PNADC também é distinta. Nela são computadas:
 - **27 capitais**
 - **26 restos da UF**
 - **20 regiões metropolitanas**
 - **4 regiões integradas de desenvolvimento econômico**

Pós-estratificação

```
# pós-estratificação
df_pos <-
  data.frame( posest = unique( pnadc$posest ) ,
              Freq = unique( pnadc$v1029 ) )

# desenho de survey
pnadc_design <- postStratify( pre_stratified ,
                             ~ posest , df_pos )

#remover objetos
rm( pnadc , pre_stratified )
gc()
pnadc_design <- as_survey(pnadc_design)
```