

Analizando a PNADC com o pacote **survey**

Solução

Guilherme Jacob

16/08/2021

O conjunto de dados que vamos utilizar neste exercício está guardado no arquivo `pnadc-recorte.Rds` na pasta `dados` e pode ser lido usando a função `readRDS`. Para ler estes dados e guardá-los no objeto `pnadc.df`, use o comando abaixo (verifique o endereço do arquivo no seu computador!):

```
pnadc.df <- readRDS( "/home/guilherme/GitHub/cursoIntroRenda/dados/pnadc-recorte.Rds" )
```

Este arquivo contém um recorte dos microdados da 1ª visita da PNAD Contínua. Com ele, realize as tarefas a seguir:

1. Crie um objeto de plano amostral;

```
# carrega library
library( survey )

# cria desenho amostral sem pós-estratificação
pnadc.design <-
  svydesign(
    ids = ~ upa + v1008 ,
    strata = ~ estrato ,
    weights = ~ v1032 ,
    data = pnadc.df ,
    nest = TRUE )
```

2. Estime as seguintes características:

2.a) Média do rendimento domiciliar per capita (`def.rdp`) do Brasil, usando a função `svymean`;

```
svymean( ~def.rdp , pnadc.design , na.rm = TRUE )
```

```
##           mean      SE
## def.rdp 1406.3 18.599
```

2.b) Média do rendimento domiciliar per capita da sua região usando a função `subset`;

```
svymean( ~def.rdp , subset( pnadc.design , regio == "Norte" ) , na.rm = TRUE )
```

```
##           mean      SE
## def.rdp 871.98 20.201
```

2.c) Média do tamanho dos domicílios (`vd2003`) por região, usando a função `svyby` e a variável `regiao`;

```
svyby( ~vd2003 , ~regiao , pnadc.design , svymean , na.rm = TRUE )
```

```
##           regio  vd2003      se
## Norte         Norte 4.289778 0.03357581
## Nordeste        Nordeste 3.740451 0.01261641
```

```
## Sudeste          Sudeste 3.454907 0.01554312
## Sul              Sul 3.341321 0.01428669
## Centro-Oeste    Centro-Oeste 3.579297 0.01986871
```

2.d) A soma dos rendimentos de todos os trabalhos (`def.vd4019`) por trimestre na sua região, usando `svyby` combinado com `svytotal` para o objeto de plano da sua região.

```
svyby( ~I( def.vd4019 / 10^9 ) , ~trimestre ,
       subset( pnadc.design , regiao == "Norte" ) ,
       svytotal ,
       na.rm = TRUE )
```

```
## trimestre I(def.vd4019/10^9)      se
## 1          1          2.679124 0.1873783
## 2          2          3.059524 0.2637918
## 3          3          3.019241 0.2266387
## 4          4          3.156726 0.2325679
```

3. Escolha uma região e crie um objeto de plano amostral para ela usando a função `subset`. Com ele:

```
pnadc.regiao <- subset( pnadc.design , regiao == "Norte" )
```

3.a) Estime a média do rendimento domiciliar per capita (`def.rdp`) por situação do domicílio, usando a função `svyby` e a variável `v1022`. Calcule também o coeficiente de variação e o intervalo de confiança (95%) das estimativas usando a opção `vartype` e o efeito do plano amostral com a opção `deff = TRUE`;

```
svyby( ~def.rdp , ~v1022 ,
       pnadc.regiao , svymean ,
       vartype = c( "cv" , "ci" ) ,
       deff = TRUE , na.rm = TRUE )
```

```
##          v1022 def.rdp      ci_l      ci_u      cv DEff.def.rdp
## Urbano Urbano 982.5081 931.9233 1033.0928 0.02626852      12.27593
## Rural  Rural 503.3545 463.8366 542.8723 0.04005637      17.52896
```

3.b) Usando a função `update`, crie a variável `p.trab = rtrab / rdom`, indicando a proporção do trabalho no rendimento do domicílio;

```
pnadc.regiao <- update( pnadc.regiao , p.trab = rtrab / rdom )
```

3.c) Agora, usando as técnicas utilizadas da questão 3, calcule a média desta proporção entre os domicílios na sua região;

```
svymean( ~p.trab , subset( pnadc.regiao , vd2002 == 1 ) , na.rm = TRUE )
```

```
##          mean      SE
## p.trab 0.69259 0.0044
```

3.d) Usando a função `svyplot`, faça um diagrama de dispersão comparando `log(rtrab)` vs. `p.trab`;

```
svyplot( p.trab~log(rtrab) ,
         subset( pnadc.regiao , vd2002 == 1 & rtrab > 0 ) ,
         basecol = "blue" , alpha = c( 0 , .6 ) ,
         style = "transparent" )
```

