

Micro Empírica - Trabalho 01

Samuel Barbosa

19 de janeiro de 2018

Introdução

Neste exercício vamos estimar os retornos à educação usando a equação de Mincer

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

e dados da PNAD 2015. Usaremos o software **R** para completar este exercício. Os dados da PNAD podem ser obtidos no R através do pacote **lodown**, por contribuição do pesquisador Djalma Pessoa (IBGE).

Download dos dados

```
install.packages("lodown")
library(lodown)
pnad_cat = get_catalog("pnad", output_dir = "data_raw/pnad/")
pnad_cat = subset(pnad_cat, year == 2015)
lodown("pnad", pnad_cat)
```

Leitura e tratamento

Os dados disponibilizados através do pacote **lodown** já combinam as variáveis de pessoas e domicílios. Vamos selecionar as variáveis de interesse e realizar as alterações necessárias.

VARIÁVEL	DESCRIÇÃO
V4718	Rendimento mensal do trabalho principal para pessoas de 10 anos ou mais de idade
V4803	Anos de estudo (todas as pessoas)
V8005	Idade do morador na data de referência
V0302	Sexo
V0404	Cor ou raça
V0504	Morou em outra Unidade da Federação ou país estrangeiro
V4729	Peso na amostra
v4617	Identificação de estrato
v4618	PSU - Unidade primária de amostragem

O código a seguir realiza este procedimento.

```
pnad_df = readRDS("data_raw/pnad/2015 main.rds")

pnad_df = pnad_df %>% select(v4718, v4803, v8005,          # principais
                           v0302, v0404,                # demograficas
                           v0504,                        # subpop. interesse
                           v4729, v4617, v4618) # amostra

save(pnad_df, file = "data/pnad/pnad2015_ex1.Rdata")
rm(pnad_df)
```

Vamos renomear as variáveis selecionadas, recodificar a variável anos de estudo (ver dicionário), incluir os rótulos das variáveis sexo, cor/raça e migração, e incluir idade ao quadrado e log da renda. Vamos também descartar observações em que a renda é nula (zero) ou não informada.

```
load("data/pnad/pnad2015_ex1.Rdata")

pnad_df = rename(pnad_df,
                  renda = v4718, anos_estudo = v4803, idade = v8005,
                  sexo = v0302, cor_raca = v0404,
                  migrou = v0504,
                  peso = v4729, estrato = v4617, psu = v4618)

pnad_df =
  mutate(pnad_df, anos_estudo = as.integer(anos_estudo)) %>%
  mutate(anos_estudo = ifelse(anos_estudo == 17, NA, anos_estudo)) %>%
  mutate(anos_estudo = anos_estudo - 1)

pnad_df =
  mutate_at(pnad_df, vars(sexo, cor_raca, migrou), as.factor)

levels(pnad_df$sexo) = c("Masc", "Fem")
levels(pnad_df$cor_raca) = c("Indigena", "Branca", "Preta", "Amarela", "Parda", "SemDecl")
levels(pnad_df$migrou) = c("Sim", "Nao")

pnad_df = mutate(pnad_df,
                  idade2 = idade^2,
                  log_renda = log(renda))

pnad_df = filter(pnad_df, renda > 0, !is.na(renda))
```

Desenho da amostra e estatísticas descritivas

Para trabalhar com amostras complexas, usaremos os pacote `survey` e `srvyr`.

```
library(srvyr); library(survey)

pnad_design = as_survey_design(pnad_df,          # define desenho amostral
                               ids = psu,
                               strata = estrato,
                               weights = peso,
                               nest = TRUE)
```

```
options(survey.lonely.psu="remove")           # remove estratos com PSU unico

pnad_design %>%
  group_by(anos_estudo) %>%                   # renda média por anos de estudo
  summarise(renda_media = survey_mean(renda))
```

As tabelas a seguir apresentam estatísticas descritivas da renda (média e erro padrão). Observe que temos uma relação crescente entre renda média e anos de estudo.

Tabela 1: Renda média

renda_media	renda_media_se
1799.46	16.49526

Tabela 2: Renda média por anos de estudo

anos_estudo	renda_media	renda_media_se
0	879.6120	27.62378
1	829.5682	24.53696
2	870.6848	25.63114
3	984.6438	28.83534
4	1135.2306	14.81840
5	1126.6333	16.24930
6	1120.1790	18.05032
7	1140.3473	18.48840
8	1315.0358	12.55239
9	1076.3695	17.27731
10	1232.0550	27.72555
11	1566.7961	10.13130
12	1764.8598	41.50910
13	2099.9411	48.20841
14	2362.9301	49.18460
15	4229.0781	65.71730

Estimação da Equação de Mincer

Vamos agora retornar à equação de Mincer, que desejamos estimar:

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

```
short_fit = svyglm(log_renda ~ anos_estudo + idade + idade2, pnad_design)
long_fit = svyglm(log_renda ~ anos_estudo + idade + idade2 + sexo + cor_raca, pnad_design)
```

Obtemos $\beta_1 \approx 0.106$ ao estimar a primeira equação. Conforme a especificação acima, interpretamos que um ano adicional de estudo está associado a um aumento na renda, em média, de aproximadamente 10,6%.

A inclusão dos controles demográficos pouco altera os resultados. Observamos, neste caso, $\beta_1 \approx 0.108$. Destaca-se também o valor do coeficiente relacionado à variável `sexo`, indicando renda média menor entre mulheres (aproximadamente 43,7%).

Tabela 3: Equação de Mincer

	log_renda	
	(1)	(2)
anos_estudo	0.106*** (0.001)	0.108*** (0.001)
idade	0.065*** (0.001)	0.069*** (0.001)
idade2	-0.001*** (0.000)	-0.001*** (0.000)
sexoFem		-0.437*** (0.005)
cor_racaBranca		0.265*** (0.083)
cor_racaPreta		0.025 (0.083)
cor_racaAmarela		0.505*** (0.091)
cor_racaParda		0.015 (0.082)
Constant	4.578*** (0.026)	4.544*** (0.088)
Observations	152,239	152,239
Log Likelihood	-181,483.300	-172,977.900
Akaike Inf. Crit.	362,974.500	345,973.800

Notes:

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Por fim vamos estimar a equação de Mincer para pessoas que migraram entre Unidades da Federação.

```
sub_pnad_design <- subset(pnad_design, migrou == "Sim")
fit_mig = svyglm(log_renda ~ anos_estudo + idade + idade2 + sexo + cor_raca, sub_pnad_design)
```

Tabela 4: Equação de Mincer (subgrupo Migração)

	<i>Dependent variable:</i>
	log_renda
anos_estudo	0.127*** (0.003)
idade	0.065*** (0.004)
idade2	−0.001*** (0.00004)
sexoFem	−0.524*** (0.016)
cor_racaBranca	0.046 (0.094)
cor_racaPreta	−0.206** (0.098)
cor_racaAmarela	0.244** (0.121)
cor_racaParda	−0.212** (0.093)
Constant	4.711*** (0.123)
Observations	12,993
Log Likelihood	−16,181.090
Akaike Inf. Crit.	32,380.180
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Observamos que, no caso dos migrantes, um ano adicional de escolaridade está associado a uma renda mais alta, em média, de aproximadamente 12,6%, indicando que o efeito da escolaridade sobre a renda é mais acentuado neste grupo.