

Métodos Quantitativos IV - Lista 5 (resolução)

Departamento de Ciência Política da FFLCH-USP

Luiz Henrique da Silva Batista (Número USP: 12687228)

2023-11-19

Exercício 1

O pacote PNADcIBGE permite importar bases de dados diretamente para o environment do R. Primeiro, instale e ative o pacote. Depois, importe os dados do último trimestre de 2017 (variáveis selecionadas) por meio do código abaixo.

```
# Ver: https://cran.r-project.org/web/packages/PNADcIBGE/PNADcIBGE.pdf
# Instale o pacote
# install.packages("PNADcIBGE")

# Carregue o pacote
library(PNADcIBGE)

# Importe os dados desejados
data <- get_pnadc(year=2017,
                  quarter=4,
                  selected=FALSE,
                  vars=c("Ano", "Trimestre", "UF", "V2007", "VD4020", "VD4035"),
                  design=FALSE,
                  savedir=tempdir()
                  )

# Por razões didáticas, selecionamos "design=FALSE" para ignorar o plano amostral.
# Não faça isso em sua pesquisa.

# Selecione apenas as variáveis úteis para esta lista:
library(tidyverse)
library(tidylog)

data <- data |>
  select(Ano, Trimestre, UF, V2007, VD4020, VD4035)

# Renomeie as variáveis:
data <- data |>
  rename(Sexo = V2007,
         Renda = VD4020,
         Horas_trabalhadas = VD4035)
```

Exercício 2

Utilize uma regressão linear simples para estimar a correlação entre a renda (variável dependente) e o sexo (variável independente).

```
# Transformando a v. sexo em dummy (0 e 1)
data <- data |>
  mutate(sexo_dummy = case_when(
    Sexo == "Homem" ~ 1,
    Sexo == "Mulher" ~ 0
  ))

# Criando o modelo
modelo1 <- lm(Renda ~ sexo_dummy, data = data)
```

- Escreva a equação correspondente a essa regressão (**OPCIONAL**)

$$Renda = \alpha + \beta_1 X_{sexo}$$

- Interprete os coeficientes

```
modelo1 |>
  stargazer::stargazer(type = "latex")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: dom, nov 19, 2023 - 21:39:04

Table 1:	
	<i>Dependent variable:</i>
	Renda
sexo_dummy	357.173*** (13.127)
Constant	1,720.783*** (10.077)
Observations	228,116
R ²	0.003
Adjusted R ²	0.003
Residual Std. Error	3,084.290 (df = 228114)
F Statistic	740.367*** (df = 1; 228114)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

O α (alpha) representa o ponto em que a nossa reta de regressão corta o eixo Y . Já o β (coeficiente angular da reta) fornece a inclinação da reta. Assim, podemos interpretá-lo como o quanto em média Y deve aumentar ou diminuir para cada aumento de uma unidade em X .

Em nosso exemplo, a renda estimada das mulheres ($X = 0$) é igual a R\$ 1720.78 e a renda estimada dos homens ($X = 1$) é cerca de R\$ 357.17 maior do que a das mulheres.

- Apresente os resultados da sua regressão em uma tabela utilizando a função “stargazer”.

```
modelo1 |>
  stargazer::stargazer(type = "latex",
                        title="Modelo 1",
                        single.row = T,
                        keep.stat = c("n"),
                        dep.var.labels="Renda",
                        header=F,
                        dep.var.caption="")
```

Table 2: Modelo 1

	Renda
sexo_dummy	357.173*** (13.127)
Constant	1,720.783*** (10.077)
Observations	228,116
Note:	*p<0.1; **p<0.05; ***p<0.01

Exercício 3

Com base na questão anterior, qual é a renda média das mulheres? E a dos homens? Confirme que os resultados coincidem com o cálculo das médias para cada sexo utilizando a função “summarise”.

A renda média das mulheres é R\$ 1720.78 e a renda média dos homens é R\$ 2077.95 (R\$ 1720.78 + R\$ 357.17).

Confirmando esses valores com a função `summarise`, temos:

```
data |>
  drop_na() |>
  group_by(Sexo) |>
  summarise(renda_media = mean(Renda)) |>
  knitr::kable(caption = "Renda média por sexo",
               col.names = c("Sexo", "Total"),
               format.args = list(decimal.mark = ",", big.mark = "."),
               align = "c")
```

Table 3: Renda média por sexo

Sexo	Total
Homem	2.077,956
Mulher	1.720,783

Exercício 4

Utilize uma regressão linear simples para estimar a correlação entre a renda (variável dependente) e as horas trabalhadas (variável independente).

```
modelo2 <- lm(Renda ~ Horas_trabalhadas, data = data)
```

- Escreva a equação correspondente a essa regressão (**OPCIONAL**)

$$Renda = \alpha + \beta_1 X_{Horas\ trabalhadas}$$

- Interprete os coeficientes

```
modelo2 |>
  stargazer::stargazer(type = "latex")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: dom, nov 19, 2023 - 21:39:05

Table 4:

	<i>Dependent variable:</i>
	Renda
Horas_trabalhadas	28.901*** (0.453)
Constant	846.776*** (18.185)
Observations	228,116
R ²	0.017
Adjusted R ²	0.017
Residual Std. Error	3,062.150 (df = 228114)
F Statistic	4,061.663*** (df = 1; 228114)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

A renda estimada para um total de 0 (zero) horas trabalhadas é de R\$ 846.7765 e a nosso modelo estima cerca de R\$ 28.9 para cada hora a mais trabalhada.

- Qual é a renda prevista para uma pessoa que trabalha 40 horas por semana? Considere que a variável dependente refere-se à renda mensal

A renda prevista para uma pessoa que trabalha 40 por semana é de 2002.8125, conforme estimado pelo modelo:

$$Renda = 846.7765 + 28.9009 * 40$$

- Apresente os resultados da sua regressão em uma tabela utilizando a função “stargazer”

```
modelo2 |>
  stargazer::stargazer(type = "latex",
    title="Modelo 2",
    single.row = T,
    keep.stat = c("n"),
    dep.var.labels="Renda",
    header=F,
    dep.var.caption="")
```

Table 5: Modelo 2

	Renda
Horas_trabalhadas	28.901*** (0.453)
Constant	846.776*** (18.185)
Observations	228,116
Note:	*p<0.1; **p<0.05; ***p<0.01

Exercício 5

Calcule os intervalos de confiança para os coeficientes das duas regressões das questões anteriores ao nível de confiança de 95%.

Modelo 1: $Renda = \alpha + \beta_1 X_{sexo}$

Intervalo de confiança de 95% para α	Intervalo de confiança de 95% para β
(1700.64); (1740.96)	(330.91); (383.43)

Modelo 2: $Renda = \alpha + \beta_1 X_{Horas\ trabalhadas}$

Intervalo de confiança de 95% para α	Intervalo de confiança de 95% para β
(810.4073); (883.1457)	(27.9939); (29.8079)

- Explique o que representam os intervalos de confiança

Os intervalos de confiança representam o intervalo de valores que acreditamos conter o parâmetro populacional.

- O que eles informam a respeito da significância estatística (ao nível de 5%) dos coeficientes estimados?

Informam que existe uma probabilidade de 95% de que o parâmetro populacional está contido no intervalo. Podemos também dizer que existe uma probabilidade de 5% de que o parâmetro **não** está contido no intervalo de confiança.

- O que significa dizer que os coeficientes são estatisticamente significantes ou insignificantes?

Significa dizer

Exercício 6

Apresente seus resultados em um arquivo PDF. Garanta que seu arquivo esteja limpo, contendo as respostas, os gráficos e as tabelas, mas não eventuais mensagens e erros. O arquivo PDF pode ser gerado diretamente a partir do R por meio do RMarkdown ou do RSweave. Para os alunos de graduação, isso é recomendado, mas não obrigatório. Adicionalmente, forneça o script para replicação.

