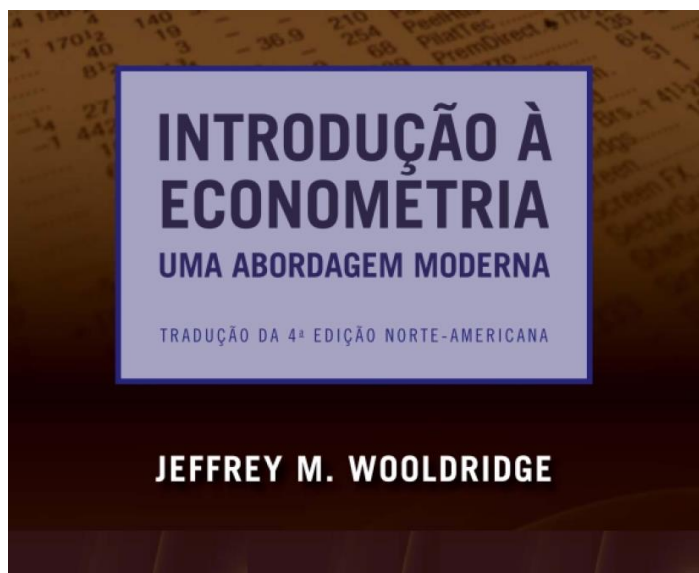


Atividade prática: dados em painel

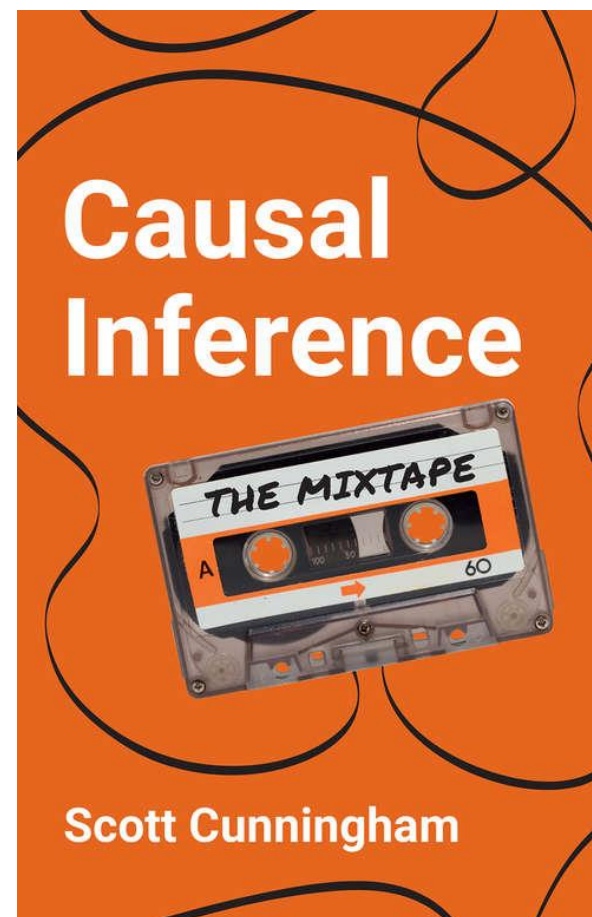
Daniel Leonel

Filipe Faeti

Referências



Capítulo 13 e 14



Capítulo 8 e 9

Dados em Painel

- Quando usamos o termo "dados em painel", o que queremos dizer?
- Queremos dizer um conjunto de dados em que observamos as mesmas unidades (por exemplo, indivíduos, empresas, países, escolas) em mais de um período (Cunningham, 2021).
 - Ou seja,
- Observamos uma amostra $i = 1, 2, \dots, N$ unidades transversais para $t = 1, 2, \dots, T$ períodos de tempo (um painel balanceado). Para cada unidade i denotamos as variáveis observáveis para todos os períodos de tempo t .

Nosso exemplo

Rodadas	Data de trabalho de campo	Tamanho da amostra
Rodada 1	Entre 6 e 21/07/2018	2.018
Rodada 2	Entre 10 e 19/09/ 2018	1.009
Rodada 3	Entre 21/09/2018 até 4/10/2018	1.224
Rodada 4	Entre 8 e 29/11/2018	957
Rodada 5	Entre 9 e 25/01/2019	817
Total de respostas		6025

Fonte: “**Democracy on the Ballot: Brazil 2018**”
(Mollie J. Cohen, Matthew L. Layton, Mason Moseley, and Amy Erica Smith)

Bibliotecas

```
# 2. Pacotes =====  
library(tidyverse)  
library(janitor)  
library(rio)  
library(Hmisc)  
library(plm)
```

ggplot; group_by; select; filter; mutate; ifelse; p.ex.

tabyl (calcular frequência)

Importar dados de todos os formatos

Para incluir uma função de ponderação
"weight".

Para incluir os estimadores do painel (a
documentação foi encaminhada no material)

Desenho de pesquisa e variáveis que escolhemos

Ondas 1 até 4 (desde o início da campanha até a semana posterior ao 1º turno).

```
# 3. Banco de dados =====
q = import("~/TRABALHOS/FAETI - DAN - TRAB/TRABALHO LEGO II/DemocracyOnTheBallot_Brazil2018v3.dta")

# 3.1. Filtro [recoltamos as ondas de 1 até 4] =====
q = q %>%
  filter(wave < 5)

# 4. Variáveis =====

# 4.1. polyarchy_1_1 = Tolerância sobre manifestações legais =====
#[1 = desaprova | 7 = aprova]

q = q %>%
  mutate(poli1 = polyarchy_1_1) # Seleccionamos esta variável no artigo

# 4.2. ing4 = democracia é a melhor forma de governo =====
#[1 = discorda muito | 7 = concorda muito]

q = q %>%
  mutate(democracia = ing4)

# 4.3. conf_votecounting = o resultado reflete o voto depositado nas urnas ====
q = q %>%
  mutate(eleições = conf_votecounting)

# 4.4. Dummy para identificar o período de tratamento =====
q = q %>%
  mutate(data_trat = ifelse(wave == 4, 1, 0))

# 4.5. Dummy para identificar tratamento 1 [voto em Bolsonaro] =====
q = q %>%
  mutate(votoBolsonaro = ifelse(cand_vote == 2, 1, 0))

# 4.6. Dummy para identificar tratamento 2 [voto em Haddad] =====
q = q %>%
  mutate(votoHaddad = ifelse(cand_vote == 1, 1, 0))
```

Uma breve observação

- No trabalho que compartilhamos, aplicamos um teste para analisar diferença de médias (diff-in-diff).

	Grupo 1	Grupo 2
Declaravam intenção de voto em Haddad	Eleitores que declaravam voto e votaram em Haddad no primeiro turno	Eleitores que declaravam voto, mas não votaram em Haddad no primeiro turno
Declaravam intenção de voto em Bolsonaro	Eleitores que declaravam voto e votaram em Bolsonaro no primeiro turno	Eleitores que declaravam voto, mas não votaram em Bolsonaro no primeiro turno

Ainda a breve observação

- Com os dados em painel, bastava calcular uma diferença de médias simples, como no exemplo abaixo:

```
# calcular diff in diff =====  
diferenças = q %>% filter(!is.na(votoBolsonaro)) %>%  
  group_by(data_trat, votoBolsonaro) %>%  
  summarise(democracia = wtd.mean(democracia, na.rm=T, weight))
```

	data_trat	votoBolsonaro	democracia
1	0	0	5.129875
2	0	1	5.136360
3	1	0	5.847840
4	1	1	6.071891

Antes do primeiro turno das eleições de 2018

Intenção de voto em Bolsonaro

Depois do primeiro turno das eleições de 2018

Como calcular a diferença de médias

```
> diff_diff  
[1] 0.2175654  
> |
```

Ou seja, numa escala de 1 até 7, verificamos um aumento de 0.22 (arredondando) de apoio a democracia entre os eleitores de Bolsonaro

```
# separar os quatro valores  
antes_tratamento = diferenças %>%  
  filter(data_trat == 0, votoBolsonaro == 1) %>%  
  pull(democracia)  
  
antes_controle = diferenças %>%  
  filter(data_trat == 0, votoBolsonaro == 0) %>%  
  pull(democracia)  
  
depois_tratamento = diferenças %>%  
  filter(data_trat == 1, votoBolsonaro == 1) %>%  
  pull(democracia)  
  
depois_controle = diferenças %>%  
  filter(data_trat == 1, votoBolsonaro == 0) %>%  
  pull(democracia)  
  
# calcular diff in diff  
diferença_tratamento_antes_depois = depois_tratamento - antes_tratamento  
diferença_controle_antes_depois = depois_controle - antes_controle  
  
diff_diff = diferença_tratamento_antes_depois - diferença_controle_antes_depois
```

Como calcular as diferenças usando regressão?

```
trat1 = plm(democracia ~ votoBolsonaro + data_trat + votoBolsonaro*data_trat,  
            weights = weight, model = "within",  
            index = c("idnumber", "wave"),  
            data = q)
```

Unidade
observada
= indivíduo

Ondas

Estimador =
efeitos fixos

Atenção!

ING4 [1, 2, 3, 4, 5]. A democracia tem alguns problemas, mas é melhor do que qualquer outra forma de governo.
[Responses noted on a 1-7 scale]

CAND_VOTE [2, 3]. Aqui tem uma lista de candidatos na eleição presidencial de 2018. Em qual candidato você pretende votar?

CAND_VOTE [4, 5]. Para começar, aqui tem uma lista de candidatos no primeiro turno da eleição presidencial de 2018. Em qual candidato você votou no primeiro turno?

*options needed

(1) Fernando Haddad (PT) 12

O modelo calcula as variáveis que se repetem conjuntamente. Do contrário, dá erro.

```
> summary(trat1)  
Oneway (individual) effect Within Model  
  
Call:  
plm(formula = democracia ~ votoBolsonaro + data_trat + votoBolsonaro *  
      data_trat, data = q, weights = weight, model = "within",  
      index = c("idnumber", "wave"))  
  
Unbalanced Panel: n = 1342, T = 1-3, N = 3095  
  
Residuals:  
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
-6.8332 -0.2376  0.0000  0.0002  0.2416  6.8332  
  
Coefficients:  
                Estimate Std. Error t-value Pr(>|t|)  
votoBolsonaro    0.067726   0.118598  0.5711    0.568  
data_trat        0.400405   0.064983  6.1617 8.912e-10 ***  
votoBolsonaro:data_trat 0.509990   0.109045  4.6769 3.136e-06 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Total Sum of Squares:    2382.5  
Residual Sum of Squares: 2131.2  
R-Squared:              0.083344  
Adj. R-Squared:        -0.62065  
F-statistic: 63.0148 on 3 and 1750 DF, p-value: < 2.22e-16  
> |
```

Agora é só aplicar a mesma lógica do calculo simples

```
> mean(fixef(trat1))  
[1] 5.189077  
> |
```

O modelo de efeitos fixos calcula uma média (intercepto) para cada indivíduo da amostra. Para aplicarmos a mesma estratégia do calculo simples que apresentamos anteriormente, basta calcular a média desses indivíduos.

```
Coefficients:  
votoBolsonaro      Estimate  
data_trat          0.067726  
votoBolsonaro:data_trat 0.509990
```

Diff-in-diff

```
# calculo da diferença de médias (Diff-in-Diff)
```

```
c0 = 5.189077  
t0 = 5.189077 + 0.067726  
c1 = 5.189077 + 0.400405  
t1 = 5.189077 + 0.509990 + 0.067726 + 0.400405
```

```
df_t = t1 - t0  
df_c = c1 - c0  
df_df = df_t - df_c
```

```
> df_df  
[1] 0.50999  
> |
```

Olha aqui a diferença da diferença!

Resultados

```
> c0  
[1] 5.189077  
> t0  
[1] 5.256803  
> c1  
[1] 5.589482  
> t1  
[1] 6.167198
```

	data_trat	votoBolsonaro	democracia
1	0	0	5.129875
2	0	1	5.136360
3	1	0	5.847840
4	1	1	6.071891

Observe a diferença entre os resultados: de um lado, temos o calculo com o modelo de regressão, de outro, o calculo de médias simples.

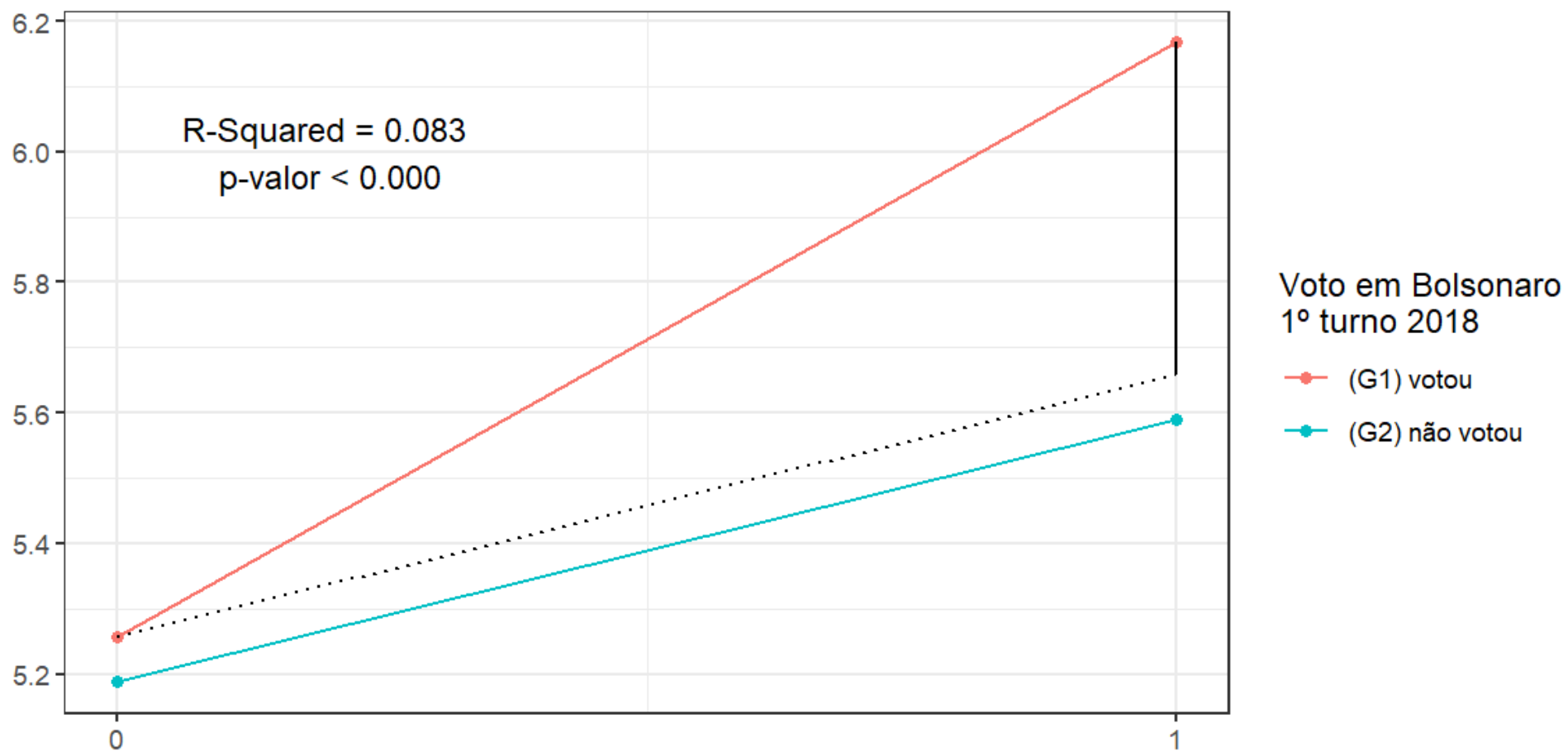
Apresentação gráfica dos resultados

```
x = c(0,0,1,1)
y = c(5.189077,5.256803,5.589482,6.167198)#segundo valor (y),último valor (yend)
voto = c("(G2) não votou","(G1) votou","(G2) não votou","(G1) votou")
objeto = data.frame(x, y, voto)

objeto %>% ggplot(aes(x = x, y = y,
                      color = voto)) +

  geom_point() +
  geom_line(aes(group = voto)) +
  annotate(geom = "segment", x = 0, xend = 1,
          y = 5.256803, yend = t1 - df_df,
          linetype = "dotted") +
  annotate(geom = "segment", x = 1, xend = 1,
          y = t1 - df_df, yend = 6.167198,
          color = "black") +
  annotate("text",y=6,x=0.2,label = "R-Squared = 0.083 \np-valor < 0.000") +
  scale_x_continuous(breaks = c(0,1)) +
  labs(
    y = "",
    x = "",
    color = "Voto em Bolsonaro \n1º turno 2018",
    caption = "Fonte: Democracy on the Ballot: Brazil 2018") +
  theme_bw()
```

Gráfico 1



Fonte: Democracy on the Ballot: Brazil 2018

OBRIGADO!