Avaliação de Políticas Públicas*

Lista de Exercícios

Alberson da Silva Miranda†

30 de dezembro de 2024

*Código disponível em https://github.com/albersonmiranda/politicas_publicas.

†Discente PPGEco/UFES.

SUMÁRIO

Q1

O banco de dados PIAA_2017-2018.xlsx contém informações desidentificadas sobre notas dos alunos e frequência na monitoria da disciplina Matemática I no departamento de economia nos anos 2017 e 2018. A tabela a seguir lista as variáveis contidas no banco de dados.

Variável	Descrição
ID	Identificação do aluno
Semestre	Semestre no qual a disciplina foi cursada
Nota	Nota do aluno
Faltas	Número de faltas na disciplina
Situação	Situação final do aluno na disciplina
Presença	Número de seções de monitoria que o aluno foi
Período	Número de períodos matriculado
Sexo	Masculino ou feminino

a

> Usando o pacote "TableOne" do R, faça um teste de balanceamento da amostra de acordo com uma variável que indica se o aluno foi a mais do que 4 seções de monitoria (tratamento). Faça esse teste sobre as seguintes variáveis: Faltas, Sexo Masculino, variáveis dummies para os semestres e variável dummy para alunos do primeiro período. O que você conclui sobre o balanceamento dos alunos entre tratados e não tratados?

Em relação às faltas, a amostra está desbalanceada. A média de faltas dos alunos que frequentaram mais de 4 seções de monitoria é de 3,82, enquanto a média dos que frequentaram menos de 4 seções é de 17,11 faltas. Formalmente, no teste de diferença de médias, com um p-valor menor que 0.001, podemos rejeitar a hipótese nula de que as médias são iguais (sendo a diferença apenas por uma questão de aleatoriedade amostral).

Em relação ao sexo, sob a hipótese nula de proporções iguais, com p-valor de 0.511, o teste indica que a amostra está balanceada. A proporção de homens (sexo = Masculino) é de 68,9% entre os que não frequentaram mais de 4 seções de monitoria e de 62,2% entre os que frequentaram.

Em relação ao semestre no qual a disciplina foi cursada, a amostra está desbalanceada (provavelmente, o semestre 20172 está puxando esse resultado). Para alunos do primeiro período, também está desbalanceada, com 100% dos alunos dentre os tratados no primeiro período.

```r
# pacotes
library(tableone)

# carregar dados
dados <- readxl::read_excel("lista/data/PIAA_2017-2018.xlsx") |>
 janitor::clean_names()

# feature engineering
dados <- within(dados, {
 tratamento <- ifelse(presenca > 4, 1, 0)
 semestre <- as.factor(semestre)
 periodo <- as.factor(periodo)
 sexo <- as.factor(sexo)
 primeiro_periodo <- as.factor(ifelse(periodo == 1, 1, 0))
})

# tabela
tableone <- CreateTableOne(
 vars = c("faltas", "sexo", "semestre", "primeiro_periodo"),
 strata = "tratamento",
 data = dados
)
print(tableone)
```

```
                        Stratified by tratamento
                         0             1             p      test
  n                        148            45
  faltas (mean (SD))     17.11 (16.65) 3.82 (3.95)   <0.001
  sexo = M (%)             102 (68.9)    28 ( 62.2)   0.511
  semestre (%)                                        0.037
     20171                  30 (20.3)    14 ( 31.1)
     20172                  47 (31.8)     5 ( 11.1)
     20181                  36 (24.3)    11 ( 24.4)
     20182                  35 (23.6)    15 ( 33.3)
  primeiro_periodo = 1 (%) 131 (88.5)    45 (100.0)   0.037
```

b

> Faça uma regressão da nota dos alunos contra uma variável indicativa de que o aluno foi a mais do que 4 seções de monitoria (tratamento). Interprete o resultado. O que seria necessário para que esse resultado possa ser interpretado como um efeito causal? Essas condições são válidas para esses dados? Por quê? Em seguida, repita essa regressão adicionando ao modelo as variáveis Faltas, variável dummy para Sexo Masculino, variáveis dummies para o semestre no qual a disciplina foi cursada e variável dummy para alunos do primeiro período. O que acontece com o valor do coeficiente da presença na monitoria em relação ao modelo do item "a"? Essa estimativa seria mais próxima de um efeito causal? Por quê? Para as regressões, use os desvios padrões robustos a heterocedasticidade.

Para o primeiro modelo, a regressão é conjuntamente significativa, com R^2 de 13,9% e coeficientes individualmente significativos. A nota média é de 2 para os alunos que não frequentaram mais de 4 seções de monitoria e de $2 + 2.6$ para os alunos que frequentaram. O coeficiente de 2.6 para a *dummy* de tratamento indica que, em média, os alunos que frequentaram mais de 4 seções de monitoria tiveram nota 2.6 pontos maior que os que não frequentaram.

Para que o resultado possa ser interpretado como um efeito causal, é necessário que o grupo de controle seja um contrafactual válido. Ou seja, que os alunos que frequentaram mais de 4 seções de monitoria sejam comparáveis aos que não frequentaram, exceto pelo tratamento. No entanto, como verificado em (a), a amostra está desbalanceada em relação às faltas, ao semestre e ao primeiro período. Portanto, o resultado não pode ser interpretado como um efeito causal. Além disso, isso resolveria apenas as questões em relação ao observáveis, não cobrindo possíveis variáveis não observadas, incorrendo em viés de seleção.

```r
# modelos
m1 <- lm(nota ~ tratamento, data = dados)
m2 <- lm(
 nota ~ tratamento + faltas + sexo + semestre + primeiro_periodo,
 data = dados
)

# matriz de covariância corrigida
cov_m1 <- sandwich::vcovHC(m1, type = "HC")
cov_m2 <- sandwich::vcovHC(m2, type = "HC")

# erros padrões robustos
robust_se_m1 <- sqrt(diag(cov_m1))
robust_se_m2 <- sqrt(diag(cov_m2))

# sumário
```

```r
stargazer::stargazer(
 m1,
 m2,
 single.row = TRUE,
 se = list(robust_se_m1, robust_se_m2),
 title = "Comparação de modelos",
 header = FALSE
)
```

Tabela 2: Comparação de modelos

	Dependent variable:	
	nota	
	(1)	(2)
tratamento	2.602*** (0.491)	1.271*** (0.408)
faltas		−0.081*** (0.010)
sexoM		0.851*** (0.293)
semestre20172		−2.135*** (0.459)
semestre20181		−3.318*** (0.393)
semestre20182		−1.743*** (0.465)
primeiro_periodo1		0.425 (0.542)
Constant	2.024*** (0.220)	4.339*** (0.664)
Observations	193	193
R^2	0.139	0.520
Adjusted R^2	0.134	0.502
Residual Std. Error	2.753 (df = 191)	2.088 (df = 185)
F Statistic	30.836*** (df = 1; 191)	28.686*** (df = 7; 185)
Note:		*p<0.1; **p<0.05; ***p<0.01

Com todas as variáveis, o segundo modelo resulta num coeficiente bem menor para a variável de tratamento, de 1.27. Isso indica que, controlando para as demais variáveis, o efeito da presença na monitoria é menor. Isso melhora o modelo em relação aos observáveis, porém não resolve o problema do desbalanceamento e nem em relação aos não observáveis e viés de seleção.

c

> Usando o pacote "MatchIt", faça o pareamento dos tratados e controles usando os seguintes critérios: pareamento exato, vizinho mais próximo por distância de Mahalanobis e vizinho mais próximo por escore de propensão usando o modelo logit.Para cada um deles, verifique o balanceamento entre tratados e controles e verifique se o pareamento foi bem-sucedido.

O pareamento exato obteve sucesso em balancear as variáveis entre tratados e controles, com apenas 7 observações dos tratados sem correspondência.

O pareamento por distância de Mahalanobis encontrou correspondência para toda a amostra tratamento (45 observações) e obteve sucesso no balanceamento para a maior parte das variáveis, exceto para faltas, que ainda apresenta diferença entre tratados e controles (eCDF Max de 11%, acima dos 5% considerados aceitáveis).

Já o pareamento por escore de propensão usando o modelo logit falhou claramento no balanceamento, com eCDF > 5% para faltas, dois dos semestres e a distância (probabilidade predita de pertencer ao grupo das tratadas).

```r
library(MatchIt)

m_exato <- matchit(
 tratamento ~ faltas + sexo + semestre + primeiro_periodo,
 data = dados,
 method = "exact"
)

m_mahalanobis <- matchit(
 tratamento ~ faltas + sexo + semestre + primeiro_periodo,
 data = dados,
 method = "nearest",
 distance = "mahalanobis"
)

m_logit <- matchit(
 tratamento ~ faltas + sexo + semestre + primeiro_periodo,
 data = dados,
 method = "nearest",
 distance = "logit"
)

# balanceamento
summary(m_exato)
```

```
Call:
matchit(formula = tratamento ~ faltas + sexo + semestre + primeiro_periodo,
    data = dados, method = "exact")

Summary of Balance for All Data:
                  Means Treated Means Control Std. Mean Diff. Var. Ratio
faltas                   3.8222       17.1149         -3.3651     0.0563
sexoF                    0.3778        0.3108          0.1381          .
sexoM                    0.6222        0.6892         -0.1381          .
semestre20171            0.3111        0.2027          0.2342          .
semestre20172            0.1111        0.3176         -0.6569          .
semestre20181            0.2444        0.2432          0.0028          .
semestre20182            0.3333        0.2365          0.2054          .
primeiro_periodo0        0.0000        0.1149         -0.4114          .
primeiro_periodo1        1.0000        0.8851          0.4114          .
                  eCDF Mean eCDF Max
faltas               0.2293   0.4874
sexoF                0.0670   0.0670
sexoM                0.0670   0.0670
semestre20171        0.1084   0.1084
semestre20172        0.2065   0.2065
semestre20181        0.0012   0.0012
semestre20182        0.0968   0.0968
primeiro_periodo0    0.1149   0.1149
primeiro_periodo1    0.1149   0.1149

Summary of Balance for Matched Data:
                  Means Treated Means Control Std. Mean Diff. Var. Ratio
faltas                   3.6842        3.6842              -0     0.9916
sexoF                    0.3421        0.3421               0          .
sexoM                    0.6579        0.6579               0          .
semestre20171            0.3684        0.3684              -0          .
semestre20172            0.1053        0.1053              -0          .
semestre20181            0.1842        0.1842              -0          .
semestre20182            0.3421        0.3421              -0          .
primeiro_periodo0        0.0000        0.0000               0          .
primeiro_periodo1        1.0000        1.0000               0          .
                  eCDF Mean eCDF Max Std. Pair Dist.
faltas                    0        0               0
sexoF                     0        0               0
sexoM                     0        0               0
semestre20171             0        0               0
```

```
semestre20172              0        0                0
semestre20181              0        0                0
semestre20182              0        0                0
primeiro_periodo0          0        0                0
primeiro_periodo1          0        0                0

Sample Sizes:
              Control Treated
All            148.        45
Matched (ESS)   29.03      38
Matched         44.        38
Unmatched      104.         7
Discarded        0.         0
```

```
summary(m_mahalanobis)
```

```
Call:
matchit(formula = tratamento ~ faltas + sexo + semestre + primeiro_periodo,
    data = dados, method = "nearest", distance = "mahalanobis")

Summary of Balance for All Data:
                  Means Treated Means Control Std. Mean Diff. Var. Ratio
faltas                   3.8222       17.1149         -3.3651     0.0563
sexoF                    0.3778        0.3108          0.1381          .
sexoM                    0.6222        0.6892         -0.1381          .
semestre20171            0.3111        0.2027          0.2342          .
semestre20172            0.1111        0.3176         -0.6569          .
semestre20181            0.2444        0.2432          0.0028          .
semestre20182            0.3333        0.2365          0.2054          .
primeiro_periodo0        0.0000        0.1149         -0.4114          .
primeiro_periodo1        1.0000        0.8851          0.4114          .
                  eCDF Mean eCDF Max
faltas               0.2293   0.4874
sexoF                0.0670   0.0670
sexoM                0.0670   0.0670
semestre20171        0.1084   0.1084
semestre20172        0.2065   0.2065
semestre20181        0.0012   0.0012
semestre20182        0.0968   0.0968
primeiro_periodo0    0.1149   0.1149
primeiro_periodo1    0.1149   0.1149
```

```
Summary of Balance for Matched Data:
                  Means Treated Means Control Std. Mean Diff. Var. Ratio
faltas                   3.8222        4.3556          -0.135     0.7643
sexoF                    0.3778        0.3778           0.000          .
sexoM                    0.6222        0.6222           0.000          .
semestre20171            0.3111        0.3111           0.000          .
semestre20172            0.1111        0.1111           0.000          .
semestre20181            0.2444        0.2444           0.000          .
semestre20182            0.3333        0.3333           0.000          .
primeiro_periodo0        0.0000        0.0000           0.000          .
primeiro_periodo1        1.0000        1.0000           0.000          .
                  eCDF Mean eCDF Max Std. Pair Dist.
faltas               0.0123   0.1111          0.2925
sexoF                0.0000   0.0000          0.0000
sexoM                0.0000   0.0000          0.0000
semestre20171        0.0000   0.0000          0.0000
semestre20172        0.0000   0.0000          0.0000
semestre20181        0.0000   0.0000          0.0000
semestre20182        0.0000   0.0000          0.0000
primeiro_periodo0    0.0000   0.0000          0.0000
primeiro_periodo1    0.0000   0.0000          0.0000

Sample Sizes:
          Control Treated
All           148      45
Matched        45      45
Unmatched     103       0
Discarded       0       0
```

```r
summary(m_logit)
```

```
Call:
matchit(formula = tratamento ~ faltas + sexo + semestre + primeiro_periodo,
    data = dados, method = "nearest", distance = "logit")

Summary of Balance for All Data:
                  Means Treated Means Control Std. Mean Diff. Var. Ratio
distance                 0.4190        0.1767          1.7540     0.4937
faltas                   3.8222       17.1149         -3.3651     0.0563
sexoF                    0.3778        0.3108          0.1381          .
```

```
sexoM                   0.6222        0.6892         -0.1381          .
semestre20171           0.3111        0.2027          0.2342          .
semestre20172           0.1111        0.3176         -0.6569          .
semestre20181           0.2444        0.2432          0.0028          .
semestre20182           0.3333        0.2365          0.2054          .
primeiro_periodo0       0.0000        0.1149         -0.4114          .
primeiro_periodo1       1.0000        0.8851          0.4114          .
                  eCDF Mean eCDF Max
distance             0.2971   0.5820
faltas               0.2293   0.4874
sexoF                0.0670   0.0670
sexoM                0.0670   0.0670
semestre20171        0.1084   0.1084
semestre20172        0.2065   0.2065
semestre20181        0.0012   0.0012
semestre20182        0.0968   0.0968
primeiro_periodo0    0.1149   0.1149
primeiro_periodo1    0.1149   0.1149

Summary of Balance for Matched Data:
                  Means Treated Means Control Std. Mean Diff. Var. Ratio
distance                0.4190        0.4067          0.0890     1.1192
faltas                  3.8222        3.6000          0.0563     0.9369
sexoF                   0.3778        0.3778          0.0000          .
sexoM                   0.6222        0.6222          0.0000          .
semestre20171           0.3111        0.4000         -0.1920          .
semestre20172           0.1111        0.1111          0.0000          .
semestre20181           0.2444        0.1778          0.1551          .
semestre20182           0.3333        0.3111          0.0471          .
primeiro_periodo0       0.0000        0.0000          0.0000          .
primeiro_periodo1       1.0000        1.0000          0.0000          .
                  eCDF Mean eCDF Max Std. Pair Dist.
distance             0.0096   0.1111          0.1039
faltas               0.0069   0.0889          0.3713
sexoF                0.0000   0.0000          0.3556
sexoM                0.0000   0.0000          0.3556
semestre20171        0.0889   0.0889          0.3840
semestre20172        0.0000   0.0000          0.0889
semestre20181        0.0667   0.0667          0.4654
semestre20182        0.0222   0.0222          0.8014
primeiro_periodo0    0.0000   0.0000          0.0000
primeiro_periodo1    0.0000   0.0000          0.0000
```

```
Sample Sizes:
          Control Treated
All           148      45
Matched        45      45
Unmatched     103       0
Discarded       0       0
```

d

> Para cada amostra pareada, faça uma regressão da nota dos alunos contra uma variável indicativa de que o aluno foi a mais do que 4 seções de monitoria(tratamento). Usando o pacote "Satargazer", organize os resultados das cinco regressões em uma tabela arrumada, interprete-os e compare os resultados da amostra pareada com o que você encontrou nas regressões sem pareamento. O que você conclui sobre o efeito da monitoria sobre as notas dos alunos?

Com as amostras pareadas, a regressão não é significativa (teste F) e o R^2 é reduzido drasticamente. O efeito do tratamento não é significativo em nenhuma das amostras pareadas. Isso significa que não há efeito causal da monitoria sobre as notas dos alunos (ou pelo menos que a amostra reduzida por conta do pareamento não é grande o suficiente para detectar o efeito, que é menos de 1 ponto em média).

```r
# modelos
m_exato_bal <- lm(
 nota ~ tratamento,
 data = match.data(m_exato),
 weights = weights
)
m_mahalanobis_bal <- lm(
 nota ~ tratamento,
 data = match.data(m_mahalanobis),
 weights = weights
)
m_logit_bal <- lm(
 nota ~ tratamento,
 data = match.data(m_logit),
 weights = weights
)

# matriz de covariância corrigida
cov_m_exato_bal <- sandwich::vcovHC(m_exato_bal, type = "HC")
cov_m_mahalanobis_bal <- sandwich::vcovHC(m_mahalanobis_bal, type = "HC")
```

```r
cov_m_logit_bal <- sandwich::vcovHC(m_logit_bal, type = "HC")

# erros padrões robustos
robust_se_m_exato_bal <- sqrt(diag(cov_m_exato_bal))
robust_se_m_mahalanobis_bal <- sqrt(diag(cov_m_mahalanobis_bal))
robust_se_m_logit_bal <- sqrt(diag(cov_m_logit_bal))

# sumário
stargazer::stargazer(
 m1,
 m2,
 m_exato_bal,
 m_mahalanobis_bal,
 m_logit_bal,
 single.row = FALSE,
 se = list(
 robust_se_m1,
 robust_se_m2,
 robust_se_m_exato_bal,
 robust_se_m_mahalanobis_bal,
 robust_se_m_logit_bal
 ),
 title = "Comparação de modelos",
 header = FALSE,
 font.size = "scriptsize",
 column.sep.width = "3pt"
)
```

Tabela 3: Comparação de modelos

	Dependent variable:				
	nota				
	(1)	(2)	(3)	(4)	(5)
tratamento	2.602***	1.271***	0.874	0.924	0.818
	(0.491)	(0.408)	(0.709)	(0.610)	(0.624)
faltas		−0.081***			
		(0.010)			
sexoM		0.851***			
		(0.293)			
semestre20172		−2.135***			
		(0.459)			
semestre20181		−3.318***			
		(0.393)			
semestre20182		−1.743***			
		(0.465)			
primeiro_periodo1		0.425			
		(0.542)			
Constant	2.024***	4.339***	4.084***	3.702***	3.809***
	(0.220)	(0.664)	(0.521)	(0.424)	(0.444)
Observations	193	193	82	90	90
R^2	0.139	0.520	0.022	0.025	0.019
Adjusted R^2	0.134	0.502	0.010	0.014	0.008
Residual Std. Error	2.753 (df = 191)	2.088 (df = 185)	2.924 (df = 80)	2.928 (df = 88)	2.994 (df = 88)
F Statistic	30.836*** (df = 1; 191)	28.686*** (df = 7; 185)	1.823 (df = 1; 80)	2.243 (df = 1; 88)	1.679 (df = 1; 88)
Note:					*p<0.1; **p<0.05; ***p<0.01

Q2

Para essa questão, use o banco de dados `minwage.xlsx`. Ele contém informações coletadas por Card & Krueger (1994) para restaurantes de fast food nos estados de New Jersey (NJ) e Pennsylvania (PA) em duas rodadas de entrevistas: em Março e Novembro/Dezembro de 1992. Em Abril daquele ano, NJ aumentou seu salário mínimo de $4,25 para $5,05 por hora. Em um artigo bastante discutido, Card e Krueger usaram esse experimento natural para avaliar o efeito de um aumento do salário-mínimob sobre o emprego (um dos exemplos clássicos de controle de preços usados em livros texto de economia). Você vai usar esse banco de dados para replicar parte do estudo de Card e Krueger. No que segue, variáveis cujo nome que termina em "2" se referem a segunda rodada da pesquisa. As variáveis `fte` e `fte2` se referem a emprego equivalente em horário integral, ou seja, a soma do número de empregados em horário integral com metade do número de empregados que trabalham em meio expediente, excluindo gerentes, `dfte` se refere a mudança em `fte` entre a primeira e a segunda entrevista (`dfte` = `fte2` – `fte`); `dw` se refere a mudança no salário inicial dos funcionários entre a primeira e a segunda entrevistas; `state` é uma variável dummy para lojas localizadas em NJ e `sample` é uma variável dummy que assume o valor 1 se dados de salário e emprego estavam disponíveis na primeira e segunda entrevista. Na análise a seguir, você deve usar apenas as observações para as quais sample é igual a 1.

a

Calcule o salário inicial médio (`wage_st`) separadamente para restaurantes em NJ e PA, em cada uma das rodadas de entrevistas.

1

Calcule a diferença nos salários médios para cada estado entre a primeira e a segunda entrevista.

```r
# carregar dados
dados <- readxl::read_excel("lista/data/minwage.xlsx")

# filtrando observações
dados <- subset(dados, sample == 1)
```

```r

# salário médio
salario_medio <- aggregate(
 cbind(wage_st, wage_st2) ~ state,
 data = dados,
 FUN = mean
)

# diferença
salario_medio <- transform(salario_medio, diff = wage_st2 - wage_st)

salario_medio
```

```
  state  wage_st wage_st2        diff
1     0 4.653636 4.618788 -0.03484848
2     1 4.612982 5.082141  0.46915807
```

2

Calcule a diferença entre as diferenças para NJ e PA que você calculou acima.

```r
diff_in_diff <- diff(salario_medio$diff)

diff_in_diff
```

```
[1] 0.5040066
```

3

> Qual a interpretação dessa estimativa de diferença em diferenças para o efeito sobre os salários? Sob que condições essa conta fornece uma estimativa válida do aumento do salário-mínimo sobre os salários nos restaurantes de fast food?

A estimativa de diferença em diferenças é de 0,5. Isso significa que, em média, o salário inicial dos funcionários dos restaurantes de fast food aumentou em 0.50 dólares por hora em NJ em relação a PA. Para que essa estimativa seja válida, é necessário que a diferença entre NJ e PA seja constante ao longo do tempo, exceto pelo aumento do salário mínimo em NJ. Isso é conhecido como a hipótese de tendências paralalelas: a diferença no resultado potencial sem tratamento $Y(0)$ para os indivíduos no grupo de tratamento $D = 1$ na mudança de $t = 0 \to t = 1$ deve ser igual à diferença no resultado potencial sem tratamento para os indivíduos no grupo de controle $D = 0$ no mesmo intervalo de tempo. Se essa hipótese for válida, a estimativa de diferença em diferenças é um estimador consistente do efeito do aumento do salário mínimo sobre os salários.

4

> Interprete seu resultado.

Já interpretado acima.

b

> Repita o mesmo exercício de (a) para a variável 'fte'. Qual o impacto do aumento do salário mínimo sobre o emprego nos restaurantes de NJ?

Referente à empregabilidade, a diferença em diferenças é de 2,30. Isso significa que, em média, o emprego nos restaurantes de fast food aumentou em 2,30 empregos equivalentes em horário integral em NJ em relação a PA. O mesmo pressuposto em relação à hipótese de tendências paralelas se aplica.

```r
# emprego médio
emprego_medio <- aggregate(
 cbind(fte, fte2) ~ state,
 data = dados,
 FUN = mean
)

# diferença
emprego_medio <- transform(emprego_medio, diff = fte2 - fte)

# diferença em diferenças
diff_in_diff_emprego <- diff(emprego_medio$diff)

emprego_medio
```

```
  state      fte     fte2       diff
1     0 20.11364 18.09848 -2.0151515
2     1 17.27544 17.56228  0.2868421
```

```
diff_in_diff_emprego
```

```
[1] 2.301994
```

c

A metodologia de diferenças em diferenças (DD) também pode ser implementada por meio da seguinte regressão:

$$Y_{ist} = \alpha + \beta_1 \text{TREAT}_{is} + \gamma POST_t + \delta_{DD} \text{TREAT}_{is} \times \text{POST}_t + \varepsilon_{ist}$$

onde Y_{ist} representa emprego no restaurante i, no estado s e período t, $\text{TREAT}_i s$ é um indicador para a área de tratamento (NJ ou restaurantes de baixo salário em NJ), POST_t é um indicador do período de tratamento (Novembro/Dezembro) e $\text{TREAT}_{is} \times \text{POST}_t$ é a interação entre essas duas *dummies*. Note que a regressão usa dados para restaurantes individuais, ao invés de dados para o estado, como em (a) e (b).

1

> Estime a regressão acima para salários e emprego. Como as estimativas diferem dos resultados que você encontrou em (a) e (b)?

Primeiramente, deve-se manipular o *dataset*, uma vez que os estados (salário e emprego) no tempo 1 e 2 estão sendo tratados em colunas ao invés de linhas. Com cada linha representando cada estado em cada tempo, é possível estimar a regressão proposta. Com isso, foram criadas uma linha adicional para cada observação, coopiando os dados fixos e alterando as informações que devem ser atualizadas no tempo 2. Além disso, foi criada uma variável `post` para indicar qual o tempo da observação. Também alterei os valores de `state` para facilitar a interpretação.

```
# pacotes
suppressPackageStartupMessages({
 library(tidyr)
 library(dplyr)
})

# alongando a tabela para tratar a mudança no tempo como
# linha ao invés de coluna
colunas_repetidas <- setdiff(
```

```r
  names(dados),
  grep("2$", names(dados), value = TRUE)
)

dados_long <- pivot_longer(
  data = dados,
  cols = matches(".*2$|.*[^2]$"),
  names_to = c(".value", "post"),
  names_pattern = "(.*?)(2?)$"
) |>
  mutate(
  post = factor(ifelse(post == "2", "nov_dez", "mar"), levels = c("mar", "nov_dez")),
  state = factor(ifelse(state == "1", "NJ", "PA"), levels = c("PA", "NJ"))
  ) |>
  fill(all_of(colunas_repetidas), .direction = "down")

dados_long
```

```
# A tibble: 702 x 13
   post    sheet chain co_owned state empft emppt wage_st   fte  dfte   gap    dw
   <fct>   <dbl> <dbl>    <dbl> <fct> <dbl> <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl>
 1 mar        56     4        1 PA       20    20    5     30   -12       0  0.25
 2 nov_d~     56     4        1 PA        0    36    5.25  18   -12       0  0.25
 3 mar        61     4        1 PA        6    26    5.5   19    10.5     0 -0.75
 4 nov_d~     61     4        1 PA       28     3    4.75  29.5  10.5     0 -0.75
 5 mar       445     1        0 PA       50    35    5     67.5 -43.5     0 -0.25
 6 nov_d~    445     1        0 PA       15    18    4.75  24   -43.5     0 -0.25
 7 mar       451     1        0 PA       10    17    5     18.5  12       0  0
 8 nov_d~    451     1        0 PA       26     9    5     30.5  12       0  0
 9 mar       455     2        1 PA        2     8    5.25   6     3       0 -0.25
10 nov_d~    455     2        1 PA        3    12    5      9     3       0 -0.25
# i 692 more rows
# i 1 more variable: sample <dbl>
```

Agora, é possível estimar a regressão proposta. Na primeira, em relação aos salários iniciais, o único coeficiente significativo foi o da interação `state:post`. Para interpretá-la, deve-se ter atenção aos valores de linha de base. Como a linha de base é PA e março, `state:post` representa o efeito quando o estado é NJ e o tempo é novembro/dezembro simultaneamente. Isso quer dizer que a diferença entre os estados em março não é significativa, assim como a diferença entre os tempos em PA. No entanto, a diferença entre os estados em novembro/dezembro é significativa, indicando que o salário em NJ aumentou em relação a PA. O coeficiente de 0.504

é consistente (na verdade, exato!) com o resultado de diferença em diferenças encontrado em (a).

Para o emprego, o coeficiente da interação `state:post` também é consistente com o resultado de diferença em diferenças encontrado em (b), porém não foi significativo, indicando que não podemos apartar essa diferença do puro acaso amostral. Essa é uma vantagem de se usar a regressão, pois ela permite testar a significância estatística do efeito.

```r
# regressão
m_salario <- lm(wage_st ~ state * post, data = dados_long)
m_emprego <- lm(fte ~ state * post, data = dados_long)

# sumário
stargazer::stargazer(
 m_salario,
 m_emprego,
 title = "Regressão de diferenças em diferenças",
 header = FALSE
)
```

2

A implementação por regressão permite que você inclua controles adicionais. Estime as regressões para salários e emprego incluindo uma variável idicativa de que a loja é própria ou franquia (co_owned) e dummies para as cadeias(chain) de restaurantes.

```r
# regressão
m_salario_2 <- lm(wage_st ~ state * post + co_owned + chain, data = dados_long)
m_emprego_2 <- lm(fte ~ state * post + co_owned + chain, data = dados_long)
```

3

Coloque os resultados que você obteve em 1 e 2 lado a lado na mesma tabela. Faça uma tabela para salário e outra para emprego.

```r
# sumário
stargazer::stargazer(
 m_salario,
 m_salario_2,
 title = "Regressão de diferenças em diferenças",
 header = FALSE
)
```

Tabela 1: Regressão de diferenças em diferenças

	Dependent variable:	
	wage_st	fte
	(1)	(2)
stateNJ	−0.041	−2.838**
	(0.038)	(1.225)
postnov_dez	−0.035	−2.015
	(0.048)	(1.561)
stateNJ:postnov_dez	0.504***	2.302
	(0.054)	(1.732)
Constant	4.654***	20.114***
	(0.034)	(1.104)
Observations	702	702
R^2	0.403	0.008
Adjusted R^2	0.400	0.004
Residual Std. Error (df = 698)	0.277	8.966
F Statistic (df = 3; 698)	156.907***	1.869
Note:	*p<0.1; **p<0.05; ***p<0.01	

Tabela 2: Regressão de diferenças em diferenças

	Dependent variable:	
	wage_st	
	(1)	(2)
stateNJ	−0.041	−0.039
	(0.038)	(0.037)
postnov_dez	−0.035	−0.035
	(0.048)	(0.047)
co_owned		0.064***
		(0.022)
chain		0.034***
		(0.010)
stateNJ:postnov_dez	0.504***	0.504***
	(0.054)	(0.053)
Constant	4.654***	4.559***
	(0.034)	(0.039)
Observations	702	702
R^2	0.403	0.425
Adjusted R^2	0.400	0.421
Residual Std. Error	0.277 (df = 698)	0.272 (df = 696)
F Statistic	156.907*** (df = 3; 698)	102.945*** (df = 5; 696)
Note:		*p<0.1; **p<0.05; ***p<0.01

```r
stargazer::stargazer(
 m_emprego,
 m_emprego_2,
 title = "Regressão de diferenças em diferenças",
 header = FALSE
)
```

Tabela 3: Regressão de diferenças em diferenças

	Dependent variable:	
	fte	
	(1)	(2)
stateNJ	-2.838^{**}	-2.871^{**}
	(1.225)	(1.216)
postnov_dez	-2.015	-2.015
	(1.561)	(1.550)
co_owned		-2.243^{***}
		(0.728)
chain		-0.243
		(0.322)
stateNJ:postnov_dez	2.302	2.302
	(1.732)	(1.720)
Constant	20.114^{***}	21.440^{***}
	(1.104)	(1.280)
Observations	702	702
R^2	0.008	0.025
Adjusted R^2	0.004	0.018
Residual Std. Error	8.966 (df = 698)	8.902 (df = 696)
F Statistic	1.869 (df = 3; 698)	3.552^{***} (df = 5; 696)
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

4

> Os seus resultados mudaram quando incluiu as dummies para restaurantes? Essa mudança era esperada? Explique por quê.

Os resultados não mudaram (o coeficiente de interação). O que mudou foi a qualidade do ajuste. Isso pode ser importante na hora de detectar um efeito, dado que o desvio-padrão dos coeficientes é menor e pode ser o suficiente para detectar um efeito que não seria detectado sem os controles adicionais.

d

Uma alternativa a comparação entre restaurantes em NJ e PA seria comparar restaurantes em NJ que pagam salários altos vs. restaurantes que pagam salários mais baixos antes do aumento do salário-mínimo. Restrinja sua amostra para os restaurantes de NJ apenas.

```r
# filtrando observações
dados_nj <- subset(dados_long, state == "NJ")
```

1

> Você esperaria que as suposições para a metodologia DD sejam mais fáceis de serem defendidas na comparação de restaurantes em NJ do que na comparação de restaurantes em NJ vs. restaurantes em PA?

Não, pois haveria apenas o grupo de tratamento (NJ) sem um grupo de controle que servisse como um contrafactual válido para a comparação. A hipótese de tendências paralelas não poderia ser testada, pois não haveria um grupo de controle para comparar a evolução do salário e do emprego. Aliás, tendo apenas um ponto no tempo antes da intervenção, a hipótese de tendências paralelas já não pode ser testada!

2

> Construa uma variável que indique os restaurantes que pagam menos que $5 antes do aumento do salário-mínimo. Use uma regressão para calcular a estimativa DD do efeito do aumento do salário-mínimo sobre emprego e salários. Qual impacto você encontra para cada uma dessas variáveis usando os restaurantes de NJ apenas?

A variável de interação `low_wage:post` é significativa para o salário, indicando que o salário aumentou em 0.62 dólares por hora para os restaurantes de baixo salário em NJ. Para o emprego, a interação não é significativa a 95% de confiança.

```r
# variável de baixo salário
dados_nj <- by(dados_nj, dados_nj$sheet, function(conjunto) {
  conjunto$low_wage <- ifelse(any(conjunto$wage_st < 5 & conjunto$post == "mar"), 1, 0)
  conjunto
}) |>
  do.call(what = rbind)

# regressão
m_salario_nj <- lm(wage_st ~ low_wage * post, data = dados_nj)
m_emprego_nj <- lm(fte ~ low_wage * post, data = dados_nj)

# sumário
stargazer::stargazer(
  m_salario_nj,
  m_emprego_nj,
  title = "Regressão de diferenças em diferenças",
  header = FALSE
)
```

3

> Compare as estimativas obtidas com o que você obteve anteriormente na parte (c). Os resultados são muito diferentes?

Os resultados para `low_wage` mostram um impacto maior no salário (aproximadamente 10 cents a mais). Para a empregabilidade, continua não significativo a 95% de confiança.

e

Repita a regressão em (d) usando agora os restaurantes em PA.

```r
# filtrando observações
dados_pa <- subset(dados_long, state == "PA")

# variável de baixo salário
dados_pa <- by(dados_pa, dados_pa$sheet, function(conjunto) {
```

Tabela 4: Regressão de diferenças em diferenças

	Dependent variable:	
	wage_st	fte
	(1)	(2)
low_wage	-0.651^{***}	-2.230^{*}
	(0.023)	(1.211)
postnov_dez	-0.004	-2.250
	(0.029)	(1.501)
low_wage:postnov_dez	0.616^{***}	3.301^{*}
	(0.033)	(1.713)
Constant	5.113^{***}	18.989^{***}
	(0.020)	(1.062)
Observations	570	570
R^2	0.776	0.008
Adjusted R^2	0.775	0.002
Residual Std. Error (df = 566)	0.164	8.625
F Statistic (df = 3; 566)	653.183^{***}	1.443
Note:	*p<0.1; **p<0.05; ***p<0.01	

```
  conjunto$low_wage <- ifelse(
  any(conjunto$wage_st < 5 & conjunto$post == "mar"),
  1,
  0
  )
  conjunto
}) |>
  do.call(what = rbind)

# regressão
m_salario_pa <- lm(wage_st ~ low_wage * post, data = dados_pa)
m_emprego_pa <- lm(fte ~ low_wage * post, data = dados_pa)

# sumário
stargazer::stargazer(
  m_salario_pa,
  m_emprego_pa,
  title = "Regressão de diferenças em diferenças",
  header = FALSE
)
```

1

> Compare os resultados que você encontrou para PA com os resultados que você encontrou para NJ.

Obtemos resultados semelhantes, com `low_wage:postnov_dez` significativo para salário, numa magnitude de 0.35 cents, quase metade de NJ, e não significativo para emprego.

2

> Faça um teste estatístico para a hipótese que o coeficiente para a variável de baixo salário tenha o mesmo valor em NJ e PA.

Com p-valor de 0.013, o teste de diferença de médias nos permite rejeitar a hipótese nula de que os coeficientes são iguais. Os coeficientes não têm o mesmo valor em NJ e PA.

```
# coeficientes e erro padrão
beta_1 <- coef(m_salario_nj)["low_wage:postnov_dez"]
beta_2 <- coef(m_salario_pa)["low_wage:postnov_dez"]

```

Tabela 5: Regressão de diferenças em diferenças

	Dependent variable:	
	wage_st	fte
	(1)	(2)
low_wage	-0.632^{***}	-0.893
	(0.071)	(2.666)
postnov_dez	-0.265^{***}	-3.848
	(0.081)	(3.043)
low_wage:postnov_dez	0.354^{***}	2.813
	(0.101)	(3.770)
Constant	5.065^{***}	20.696^{***}
	(0.057)	(2.152)
Observations	132	132
R^2	0.426	0.015
Adjusted R^2	0.412	-0.009
Residual Std. Error (df = 128)	0.275	10.318
F Statistic (df = 3; 128)	31.605^{***}	0.630
Note:	*p<0.1; **p<0.05; ***p<0.01	

```r
se_1 <- summary(m_salario_nj)$coefficients["low_wage:postnov_dez", "Std. Error"]
se_2 <- summary(m_salario_pa)$coefficients["low_wage:postnov_dez", "Std. Error"]

# estatística de teste
z <- (beta_1 - beta_2) / sqrt(se_1^2 + se_2^2)

# p-valor
p_valor <- 2 * (1 - pnorm(abs(z)))

p_valor
```

```
low_wage:postnov_dez
          0.01308211
```

3

Por que verificar se o aumento do salário-mínimo em NJ teve impacto em PA pode ser uma maneira de confirmar que a metodologia produz resultados sensatos? O que você pode concluir com essa comparação?

Há dois caminhos: ou é esperado que o aumento do salário-mínimo em NJ tenha um efeito de *spillover* em PA — causado, por exemplo, pela migração de trabalhadores de PA para NJ e consequente redução da oferta de trabalho em PA, aumentando o salário médio —, de forma que o resultado confirma a metodologia, ou é esperado que o aumento do salário-mínimo em NJ não tenha efeito em PA, e o resultado indica que, ao menos, parte do resultado em NJ seja devido a fatores não observáveis (que também afetaram PA, mas não foram capturados na análise).

Sem o conhecimento de economia do trabalho e da região, eu suponho que o segundo caso seja mais provável, ou seja, que o aumento do salário-mínimo em NJ não tenha efeito em PA e que a metodologia tenha sucesso em capturar o efeito causal do aumento do salário-mínimo em NJ mas que parte desse efeito seja devido a fatores não observáveis, registrados no coeficiente de `low_wage:postnov_dez` em PA.

Q3

Para essa questão, use o banco de dados `Guns.xlsx`. Uma descrição detalhada dos dados está contida no arquivo `Guns_Description.pdf`. Alguns estados dos EUA promulgaram leis que permitem que os cidadãos carreguem armas escondidas. Essas leis, conhecidas como "shall-issue laws", instruem as autoridades locais a emitirem uma permissão de armas ocultas a todos os requerentes que sejam cidadãos, sejam mentalmente competentes e não tenham sido condenados por crime doloso (alguns estados têm algumas restrições adicionais). Os proponentes argumentam que, se mais pessoas portarem armas ocultas, o crime diminuirá, porque os criminosos são dissuadidos de atacar outras pessoas. Oponentes argumentam que o crime aumentará por causa do uso acidental ou espontâneo da arma. Nessa questão, você usará os dados de Ayres & Donohue (2003) para estimar o efeito das leis de armas ocultas em crimes violentos.

```r
dados <- readxl::read_excel("lista/data/Guns.xlsx") |>
  transform(
    shall = as.factor(shall),
    stateid = as.factor(stateid)
  )
```

a

Estime uma regressão de `ln(vio)` contra `shall` e uma regressão de `ln(vio)` contra `shall`, `incarc_rate`, `density`, `avginc`, `pop`, `pb1064`, `pw1064` e `pm1029`.

```r
# regressões
m1 <- lm(log(vio) ~ shall, data = dados)
m2 <- lm(
  log(vio) ~ shall + incarc_rate + density + avginc + pop + pb1064 + pw1064 + pm1029,
  data = dados
)

# sumário
stargazer::stargazer(
 m1,
 m2,
```

```
  title = "Regressões",
  header = FALSE
)
```

i

> Interprete o coeficiente de shall na segunda regressão. Essa estimativa pode ser considerada um efeito causal? Por que?

O coeficiente de `shall` em ambas regressões são significativos e negativos, o que quer dizer que a presença de leis de armas ocultas está associada a uma redução no crime violento. No entanto, a estimativa não pode ser considerada um efeito causal, pois não foi estabelecido um contrafactual válido: Não há controle de indivíduos ou de tempo, de forma que não é possível erificar a hipótese de tendências paralelas. O efeito pode ser devido a fatores não observados associados à passagem do tempo ou não presente em todos os estados que implementaram as leis de armas ocultas.

ii

> As variáveis de controle adicionadas na segunda regressão mudam muito a magnitude do efeito das leis de armas ocultas na primeira regressão? Mudam a significância estatística do coeficiente estimado? Você espera que essa estimativa se aproxime mais de um efeito causal do que a anterior?

Na primeira regressão, o efeito de `shall` é de $e^{-0443} = 0.6421$, ou seja, uma redução de 35,8% no crime violento. Na segunda regressão, o efeito é de $e^{-0.368} = 0.692$, ou seja, uma redução de 30,8%. A diferença não é muito grande e também não muda a significância estatística do coeficiente, que continua significativo a 95% de confiança. A inclusão das variáveis de controle não torna a estimativa mais próxima de um efeito causal, pois os problemas citados anteriormente não foram resolvidos.

b

> Os resultados se alteram se você adicionar efeitos fixos para estados e períodos de tempo (‘TWFE‘)? Qual dos resultados é mais crível para estimar um efeito causal e por quê?

Tabela 1: Regressões

	Dependent variable:	
	log(vio)	
	(1)	(2)
shall1	-0.443^{***}	-0.368^{***}
	(0.042)	(0.033)
incarc_rate		0.002^{***}
		(0.0001)
density		0.027^{**}
		(0.013)
avginc		0.001
		(0.008)
pop		0.043^{***}
		(0.003)
pb1064		0.081^{***}
		(0.017)
pw1064		0.031^{***}
		(0.008)
pm1029		0.009
		(0.011)
Constant	6.135^{***}	2.982^{***}
	(0.021)	(0.543)
Observations	1,173	1,173
R^2	0.087	0.564
Adjusted R^2	0.086	0.561
Residual Std. Error	0.617 (df = 1171)	0.428 (df = 1164)
F Statistic	111.079^{***} (df = 1; 1171)	188.411^{***} (df = 8; 1164)
Note:		*p<0.1; **p<0.05; ***p<0.01

Partindo para um modelo do tipo *within* com efeitos fixos individuais e temporais (Equação 0.1), representados no código pelo argumento `effect = "twoways"`, temos que o coeficiente de `shall` deixa de ser significativo, ou seja, a implementação do porte de armas escondidas não tem efeito significativo no crime violento.

O uso de efeitos fixos para estados e períodos de tempo é mais crível para estimar um efeito causal, uma vez que foram adicionados os controle para estados e tempo, permitindo a comparação de cada estado consigo mesmo ao longo do tempo. Isso permite que a hipótese de tendências paralelas seja testada, o que não era possível nos modelos anteriores.

$$y_{it} = x'it\beta + \alpha_i + \theta_t + \varepsilon_{it} \tag{0.1}$$

```r
# pacote para dados em painel
suppressPackageStartupMessages(library(plm))

# organização dos dados em painel
painel <- pdata.frame(dados, index = c("stateid", "year"))

# modelo
m3 <- plm(
  log(vio) ~ shall,
  data = painel,
  model = "within",
  effect = "twoways"
)

m4 <- plm(
  log(vio) ~ shall + incarc_rate + density + avginc + pop + pb1064 + pw1064 + pm1029,
  data = painel,
  model = "within",
  effect = "twoways"
)

# sumário
stargazer::stargazer(
 m3,
 m4,
 title = "Regressões em painel",
 header = FALSE
)
```

Tabela 2: Regressões em painel

	Dependent variable:	
	log(vio)	
	(1)	(2)
shall1	0.002	−0.028
	(0.017)	(0.017)
incarc_rate		0.0001
		(0.0001)
density		−0.092
		(0.076)
avginc		0.001
		(0.006)
pop		−0.005
		(0.008)
pb1064		0.029
		(0.023)
pw1064		0.009
		(0.008)
pm1029		0.073***
		(0.016)
Observations	1,173	1,173
R^2	0.00001	0.056
Adjusted R^2	−0.066	−0.013
F Statistic	0.013 (df = 1; 1099)	8.151*** (df = 8; 1092)
Note:		*p<0.1; **p<0.05; ***p<0.01

c

> Usando o modelo de estudos de evento, tente fornecer alguma evidência da validade da hipótese de tendências paralelas entre tratados e controles.

Para testar a hipótese de tendências paralelas, é necessário que a diferença entre os grupos tratados e de controle seja constante ao longo do tempo, exceto pelo tratamento. Entretanto, vemos na Figura 2 que a inclinação da reta para o grupo tratado pré-tratamento é diferente da inclinação do grupo controle, indicando que a hipótese de tendências paralelas não é válida. Além disso, a inclinação do grupo controle é a mesma do grupo tratado pós-tratamento, o que pode ser um indício da ausência de efeito causal do tratamento.

A Figura 1 mostra que, em muitos dos estados, a quantidade de crimes violentos aumenta após o tratamento, o que é contrário à hipótese de que o porte de armas escondidas reduziria o crime violento.

```r
# ano de início do tratamento
anos_tratamento <- aggregate(
  year ~ stateid,
  data = subset(dados, shall == 1),
  FUN = min
)

# merge com dataset
dados <- merge(dados, anos_tratamento, by = "stateid", all.x = TRUE) |>
  (\(dadinho) sort_by(dadinho, dadinho$stateid, dadinho$year.x))()

# renomear variáveis
names(dados)[names(dados) %in% c("year.x", "year.y")] <- c("year", "year_tratamento")

# mais data engineering!
dados <- within(dados, {
  # anos relativos ao tratamento
  year_rel <- year - year_tratamento
  # indicador de pré ou pós tratamento
  tratamento <- ifelse(year_rel >= 0, "post", "pre")
  # indicador de controle
  tratamento <- ifelse(is.na(tratamento), "controle", tratamento)
})

# plotar quantidade de `vio` por ano relativo ao tratamento para ambos grupos
library(ggplot2)
dados |>
```

```r
subset(tratamento != "controle") |>
ggplot(aes(x = year_rel, y = vio, color = tratamento)) +
geom_point() +
geom_smooth(method = "lm", se = FALSE, formula = y ~ poly(x, 3)) +
facet_wrap(~ stateid, scales = "free") +
labs(
  x = "Ano relativo ao tratamento",
  y = "Violência"
) +
# removendo elementos para melhor visualização
theme(
  # legenda
  legend.position = "none",
  # removendo números de eixos
  axis.text.x = element_blank(),
  axis.text.y = element_blank(),
  # reduzindo strip text e largura
  strip.text = element_text(size = 6),
  strip.background = element_blank()
)
```

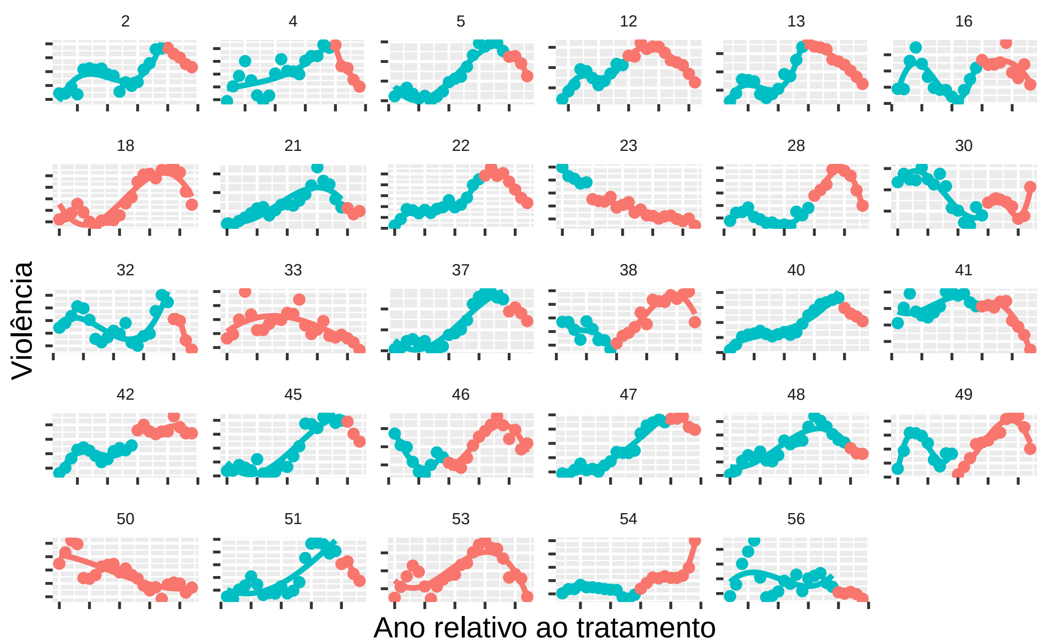

Figura 1: Violência por ano relativo ao tratamento

```r
# plotar quantidade de `vio` por ano, considerando apenas período sem tratamento
dados |>
  #subset(tratamento != "post") |>
  ggplot(aes(x = year, y = vio, color = tratamento)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Violência por ano",
    x = "Ano",
    y = "Violência"
  )
```

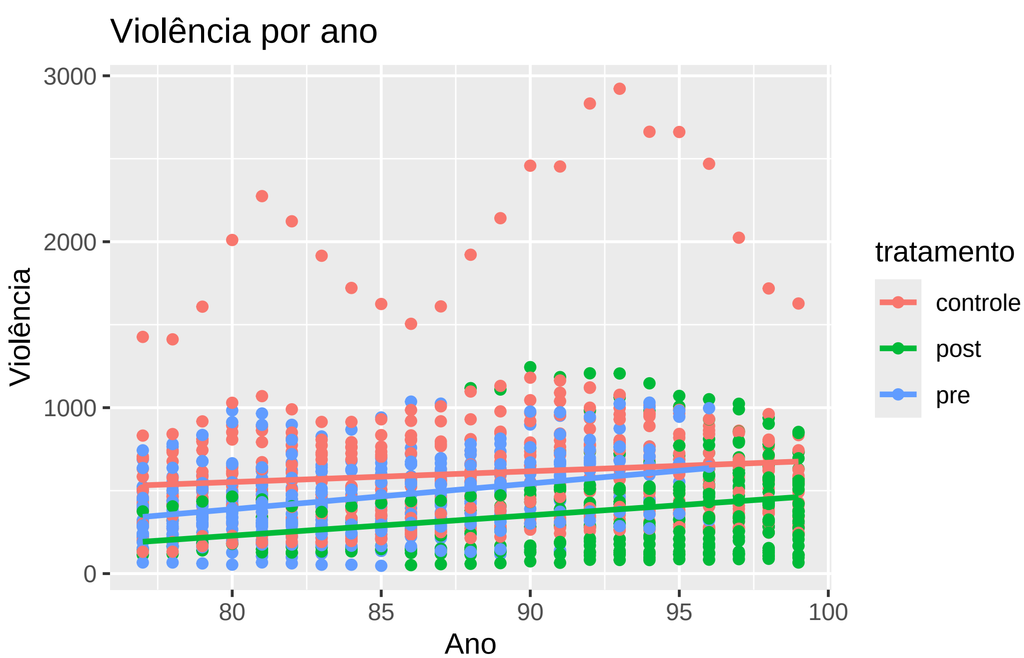

Figura 2: Violência por ano, grupos tratado e controle

d

> Usando a nomenclatura de Callaway e Sant'Anna (2021), quantos grupos de tratados existem no banco de dados?

Neste caso, a quantidade de grupos de tratados corresponde aos anos distintos de implementação do tratamento, que são 23.

```r
# quantidade de grupos de tratados
grupos_unicos <- unique(dados$year[dados$shall == 1])
grupos_unicos
```

```
 [1] 95 96 97 98 99 88 89 90 91 92 93 94 77 78 79 80 81 82 83 84 85 86 87
```

```r
length(grupos_unicos)
```

```
[1] 23
```

e

> Usando o método de Callaway e Sant'Anna (2021) estime os diferentes Efeitos Médios do Programa para Grupo-Período. Existe evidência de efeitos heterogêneos por grupo-período?

Sim. Na maior parte dos grupos, ou não há observações suficientes para estimação do modelo, ou não há efeito significativo, ou quando há é apenas pontual, em um ano. Apenas no grupo 95 há efeito significativo em vários anos consecutivos, sendo negativo para o ano de 1989 mas positivo de 1994 em diante.

```r
# mais features
dados <- within(dados, {
  G <- ifelse(is.na(year_tratamento), 0, year_tratamento)
  D <- ifelse(year >= year_tratamento & !is.na(year_tratamento), 1, 0)
  stateid_n <- as.numeric(stateid)
})

# Estimar os efeitos médios do programa para grupo-período
efeitos_medios <- did::att_gt(
  yname = "vio",
  tname = "year",
  idname = "stateid_n",
  gname = "G",
  xformla = ~ incarc_rate + pb1064 + pw1064 + avginc + density,
  data = dados,
  panel = TRUE
)

# sumário
summary(efeitos_medios)
```

```
Call:
did::att_gt(yname = "vio", tname = "year", idname = "stateid_n",
    gname = "G", xformla = ~incarc_rate + pb1064 + pw1064 + avginc +
        density, data = dados, panel = TRUE)

Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-Differences with Multiple Time Periods." Journ
230, 2021. <https://doi.org/10.1016/j.jeconom.2020.12.001>, <https://arxiv.org/abs/1803.09015>

Group-Time Average Treatment Effects:
 Group Time  ATT(g,t) Std. Error [95% Simult.  Conf. Band]
    82   78        NA         NA            NA          NA
    82   79        NA         NA            NA          NA
    82   80        NA         NA            NA          NA
    82   81        NA         NA            NA          NA
    82   82        NA         NA            NA          NA
    82   83        NA         NA            NA          NA
    82   84        NA         NA            NA          NA
    82   85        NA         NA            NA          NA
    82   86        NA         NA            NA          NA
    82   87        NA         NA            NA          NA
    82   88        NA         NA            NA          NA
    82   89        NA         NA            NA          NA
    82   90        NA         NA            NA          NA
    82   91        NA         NA            NA          NA
    82   92        NA         NA            NA          NA
    82   93        NA         NA            NA          NA
    82   94        NA         NA            NA          NA
    82   95        NA         NA            NA          NA
    82   96        NA         NA            NA          NA
    82   97        NA         NA            NA          NA
    82   98        NA         NA            NA          NA
    82   99        NA         NA            NA          NA
    86   78        NA         NA            NA          NA
    86   79        NA         NA            NA          NA
    86   80        NA         NA            NA          NA
    86   81        NA         NA            NA          NA
    86   82        NA         NA            NA          NA
    86   83        NA         NA            NA          NA
    86   84        NA         NA            NA          NA
    86   85        NA         NA            NA          NA
    86   86        NA         NA            NA          NA
    86   87        NA         NA            NA          NA
```

86	88	NA	NA	NA	NA
86	89	NA	NA	NA	NA
86	90	NA	NA	NA	NA
86	91	NA	NA	NA	NA
86	92	NA	NA	NA	NA
86	93	NA	NA	NA	NA
86	94	NA	NA	NA	NA
86	95	NA	NA	NA	NA
86	96	NA	NA	NA	NA
86	97	NA	NA	NA	NA
86	98	NA	NA	NA	NA
86	99	NA	NA	NA	NA
87	78	NA	NA	NA	NA
87	79	NA	NA	NA	NA
87	80	NA	NA	NA	NA
87	81	NA	NA	NA	NA
87	82	NA	NA	NA	NA
87	83	NA	NA	NA	NA
87	84	NA	NA	NA	NA
87	85	NA	NA	NA	NA
87	86	NA	NA	NA	NA
87	87	NA	NA	NA	NA
87	88	NA	NA	NA	NA
87	89	NA	NA	NA	NA
87	90	NA	NA	NA	NA
87	91	NA	NA	NA	NA
87	92	NA	NA	NA	NA
87	93	NA	NA	NA	NA
87	94	NA	NA	NA	NA
87	95	NA	NA	NA	NA
87	96	NA	NA	NA	NA
87	97	NA	NA	NA	NA
87	98	NA	NA	NA	NA
87	99	NA	NA	NA	NA
88	78	NA	NA	NA	NA
88	79	NA	NA	NA	NA
88	80	NA	NA	NA	NA
88	81	NA	NA	NA	NA
88	82	NA	NA	NA	NA
88	83	NA	NA	NA	NA
88	84	NA	NA	NA	NA
88	85	NA	NA	NA	NA
88	86	NA	NA	NA	NA

88	87	NA	NA	NA	NA	
88	88	NA	NA	NA	NA	
88	89	NA	NA	NA	NA	
88	90	NA	NA	NA	NA	
88	91	NA	NA	NA	NA	
88	92	NA	NA	NA	NA	
88	93	NA	NA	NA	NA	
88	94	NA	NA	NA	NA	
88	95	NA	NA	NA	NA	
88	96	NA	NA	NA	NA	
88	97	NA	NA	NA	NA	
88	98	NA	NA	NA	NA	
88	99	NA	NA	NA	NA	
90	78	16.8113	14.2867	-31.7113	65.3339	
90	79	2.2446	15.0760	-48.9587	53.4479	
90	80	-16.0294	10.2245	-50.7552	18.6964	
90	81	-29.3239	15.8919	-83.2985	24.6506	
90	82	-33.4837	25.3216	-119.4847	52.5172	
90	83	17.5104	10.1424	-16.9367	51.9575	
90	84	-4.3933	8.1878	-32.2019	23.4152	
90	85	-3.1396	7.9024	-29.9790	23.6998	
90	86	-9.0373	11.4653	-47.9774	29.9028	
90	87	30.9261	14.0169	-16.6802	78.5324	
90	88	9.9978	31.4636	-96.8634	116.8591	
90	89	11.1119	8.8935	-19.0936	41.3174	
90	90	-30.9216	9.0637	-61.7050	-0.1382	*
90	91	-64.4779	37.6898	-192.4856	63.5298	
90	92	-106.4677	37.2242	-232.8941	19.9586	
90	93	-108.3028	48.3704	-272.5857	55.9802	
90	94	-88.5237	45.0310	-241.4646	64.4173	
90	95	-59.2638	28.4436	-155.8682	37.3406	
90	96	-32.8639	66.2600	-257.9061	192.1783	
90	97	-56.1566	60.9141	-263.0421	150.7288	
90	98	-88.0526	92.4604	-402.0806	225.9754	
90	99	-3.0983	65.4142	-225.2679	219.0713	
91	78	NA	NA	NA	NA	
91	79	NA	NA	NA	NA	
91	80	NA	NA	NA	NA	
91	81	NA	NA	NA	NA	
91	82	NA	NA	NA	NA	
91	83	NA	NA	NA	NA	
91	84	NA	NA	NA	NA	
91	85	NA	NA	NA	NA	

91	86	NA	NA	NA	NA	
91	87	NA	NA	NA	NA	
91	88	NA	NA	NA	NA	
91	89	-36.1013	20.0528	-104.2076	32.0050	
91	90	-46.5126	12.0302	-87.3712	-5.6539	*
91	91	NA	NA	NA	NA	
91	92	NA	NA	NA	NA	
91	93	NA	NA	NA	NA	
91	94	NA	NA	NA	NA	
91	95	NA	NA	NA	NA	
91	96	NA	NA	NA	NA	
91	97	NA	NA	NA	NA	
91	98	NA	NA	NA	NA	
91	99	NA	NA	NA	NA	
92	78	28.1476	9.6637	-4.6737	60.9690	
92	79	-0.3729	23.8453	-81.3599	80.6141	
92	80	-40.9712	11.2696	-79.2466	-2.6958	*
92	81	NA	NA	NA	NA	
92	82	NA	NA	NA	NA	
92	83	NA	NA	NA	NA	
92	84	NA	NA	NA	NA	
92	85	NA	NA	NA	NA	
92	86	NA	NA	NA	NA	
92	87	NA	NA	NA	NA	
92	88	NA	NA	NA	NA	
92	89	NA	NA	NA	NA	
92	90	NA	NA	NA	NA	
92	91	NA	NA	NA	NA	
92	92	NA	NA	NA	NA	
92	93	NA	NA	NA	NA	
92	94	NA	NA	NA	NA	
92	95	NA	NA	NA	NA	
92	96	NA	NA	NA	NA	
92	97	NA	NA	NA	NA	
92	98	NA	NA	NA	NA	
92	99	NA	NA	NA	NA	
95	78	4.2739	21.8061	-69.7873	78.3352	
95	79	-18.2749	19.8239	-85.6037	49.0538	
95	80	-10.0527	30.9227	-115.0770	94.9716	
95	81	NA	NA	NA	NA	
95	82	NA	NA	NA	NA	
95	83	NA	NA	NA	NA	
95	84	NA	NA	NA	NA	

95	85	NA	NA	NA	NA	
95	86	-18.5775	24.9309	-103.2515	66.0966	
95	87	11.9033	28.4693	-84.7883	108.5948	
95	88	5.6073	18.5548	-57.4112	68.6259	
95	89	-60.1374	18.1752	-121.8666	1.5918	
95	90	-20.3368	25.8018	-107.9687	67.2951	
95	91	-30.9435	35.3330	-150.9467	89.0598	
95	92	3.9848	17.6780	-56.0559	64.0255	
95	93	63.0494	35.1258	-56.2501	182.3489	
95	94	52.7887	13.3864	7.3236	98.2538	*
95	95	61.8424	15.5472	9.0386	114.6462	*
95	96	75.6614	17.3788	16.6368	134.6860	*
95	97	94.9832	23.8858	13.8587	176.1078	*
95	98	54.3264	32.6799	-56.6661	165.3189	
95	99	71.2528	29.4585	-28.7987	171.3043	
96	78	NA	NA	NA	NA	
96	79	NA	NA	NA	NA	
96	80	-14.0876	24.9627	-98.8696	70.6944	
96	81	NA	NA	NA	NA	
96	82	NA	NA	NA	NA	
96	83	NA	NA	NA	NA	
96	84	NA	NA	NA	NA	
96	85	NA	NA	NA	NA	
96	86	NA	NA	NA	NA	
96	87	NA	NA	NA	NA	
96	88	19.4054	15.6467	-33.7364	72.5472	
96	89	-36.3295	56.9907	-229.8898	157.2307	
96	90	-60.0790	28.8244	-157.9767	37.8187	
96	91	-67.2576	23.4881	-147.0313	12.5161	
96	92	-5.5084	17.7897	-65.9283	54.9115	
96	93	104.8199	44.0335	-44.7332	254.3729	
96	94	105.6578	35.5282	-15.0085	226.3241	
96	95	33.8210	16.4616	-22.0885	89.7305	
96	96	-2.5534	16.3175	-57.9732	52.8663	
96	97	2.9448	20.6656	-67.2428	73.1323	
96	98	-12.8283	39.9534	-148.5240	122.8673	
96	99	-32.2859	50.2273	-202.8754	138.3036	
97	78	NA	NA	NA	NA	
97	79	NA	NA	NA	NA	
97	80	NA	NA	NA	NA	
97	81	NA	NA	NA	NA	
97	82	NA	NA	NA	NA	
97	83	NA	NA	NA	NA	

```
    97   84        NA         NA            NA           NA
    97   85        NA         NA            NA           NA
    97   86        NA         NA            NA           NA
    97   87        NA         NA            NA           NA
    97   88        NA         NA            NA           NA
    97   89        NA         NA            NA           NA
    97   90        NA         NA            NA           NA
    97   91  -51.3260    44.6489     -202.9692     100.3172
    97   92        NA         NA            NA           NA
    97   93   27.9079    62.6209     -184.7747     240.5904
    97   94        NA         NA            NA           NA
    97   95        NA         NA            NA           NA
    97   96        NA         NA            NA           NA
    97   97        NA         NA            NA           NA
    97   98        NA         NA            NA           NA
    97   99        NA         NA            NA           NA
---
Signif. codes: `*' confidence band does not cover 0

Control Group:  Never Treated,  Anticipation Periods:  0
Estimation Method:  Doubly Robust
```

f

> A partir da resposta anterior, obtenha o efeito médio agregado total e compare esse resultado como o que você obteve na letra "a".

O efeito médio agregado total é de -6,56, bem menor do que o obtido na letra "a", onde os coeficientes foram de -0.44 e -0.37.

```r
mean(efeitos_medios$att, na.rm = TRUE)
```

```
[1] -6.560892
```

```r
coef(m1)["shall1"]
```

```
    shall1
-0.4429646
```

```r
coef(m2)["shall1"]
```

```
    shall1
-0.3683869
```

g

> Repita a análise usando ‘ln(rob)‘ e ‘ln(mur)‘ no lugar de ‘ln(vio)‘. Coloque seus resultados em tabelas arrumadas.

O resultado da regressão com efeitos fixos para estados e períodos de tempo mostra que a implementação do porte de armas escondidas também não tem efeito significativo no número de roubos e assassinatos.

```r
# modelo
m5 <- plm(
  log(rob) ~ shall + incarc_rate + density + avginc + pop + pb1064 + pw1064 + pm1029,
  data = painel,
  model = "within",
  effect = "twoways"
)

m6 <- plm(
  log(mur) ~ shall + incarc_rate + density + avginc + pop + pb1064 + pw1064 + pm1029,
  data = painel,
  model = "within",
  effect = "twoways"
)

# sumário
stargazer::stargazer(
 m5,
 m6,
 title = "Regressões em painel",
 header = FALSE
)
```

Tabela 3: Regressões em painel

	Dependent variable:	
	log(rob)	log(mur)
	(1)	(2)
shall1	0.027	−0.015
	(0.024)	(0.025)
incarc_rate	0.00003	−0.0001
	(0.0001)	(0.0001)
density	−0.045	-0.544^{***}
	(0.105)	(0.110)
avginc	0.014	0.057^{***}
	(0.009)	(0.009)
pop	0.00002	-0.032^{***}
	(0.011)	(0.011)
pb1064	0.014	0.022
	(0.031)	(0.033)
pw1064	−0.013	−0.0005
	(0.011)	(0.011)
pm1029	0.105^{***}	0.069^{***}
	(0.022)	(0.023)
Observations	1,173	1,173
R^2	0.049	0.116
Adjusted R^2	−0.021	0.051
F Statistic (df = 8; 1092)	7.048^{***}	17.845^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

h.

> Baseado na sua análise, que conclusões você tiraria a respeito dos efeitos das leis de armas ocultas sobre as taxas de criminalidade? Use uma linguagem clara e acessível, de maneira que até o deputado Marcel Van Hatten (Novo/RS) entenda.

Eu tenho alguma experiência adestrando cachorros, mas asnos não são minha especialidade. Farei meu melhor.

Não existem evidências para se afirmar que as "shall issue law" têm qualquer efeito sobre crimes violentos, roubos e assassinatos. A análise estatística não encontrou efeitos causais entre a promulgação dessas leis e a criminalidade. Como podemos ver na Figura 1, em alguns estados a incidência de crimes violentos diminuiu e em outros aumentou. Isso quer dizer que existe uma grande variabilidade nos resultados, e a verdadeira causa dessas variações não foi capturada na análise. Se houvesse efeito causal das leis de armas ocultas, esperaríamos que a criminalidade diminuísse de forma consistente após a implementação dessas leis em todos os estados, o que não foi observado. Portanto, não podemos afirmar que as "shall issue laws" têm qualquer efeito sobre a criminalidade.

Além disso, a Figura 2 mostra que a tendência de crimes violentos entre os estados que nunca implementaram as leis de armas ocultas é a mesma que a dos estados que implementaram após a implementação. Isso também reforça a ideia de que a implementação das leis de armas ocultas não teve efeito sobre a criminalidade.