

Diferença-em-Diferenças, Estudo de Evento e Pareamento

Heitor Lima

Métodos Quantitativos para Avaliação de Políticas Públicas - MPP - 2023

Professora: Letícia Nunes

Insper Instituto de Ensino e Pesquisa

`heitoraol@al.insper.edu.br`

Diferença-em-Diferenças

Diferença-em-Diferenças

O método de diferença-em-diferenças (*difference-in-differences*, *diff-in-diff*, *DiD*, *DD*) permite controlar, ao mesmo tempo, por dois fatores que podem afetar uma comparação de médias simples:

- **Viés de seleção**, que ocorre ao se comparar indivíduos diferentes
- **Efeito do tempo**, que ocorre ao se comparar momentos diferentes

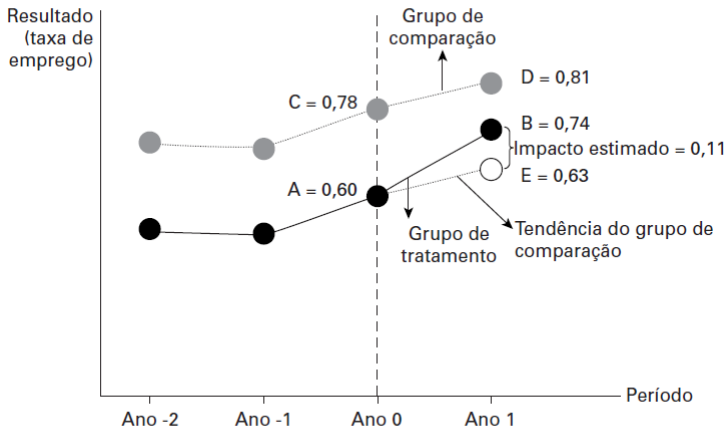
Pode ser implementado via diferenças de médias:

$$(\mathbb{E}[Y_{t=1}^T] - \mathbb{E}[Y_{t=0}^T]) - (\mathbb{E}[Y_{t=1}^C] - \mathbb{E}[Y_{t=0}^C])$$

- T e C indicam grupo de tratamento e controle, respectivamente
- $t = \{0, 1\}$ indica os períodos pré e pós-intervenção, respectivamente

Diferença-em-Diferenças

Figura 7.1 Método diferença em diferenças



Observação: todas as diferenças entre pontos devem ser lidas como diferenças verticais nos resultados medidos no eixo vertical.

Principal pressuposição: **tendências paralelas**

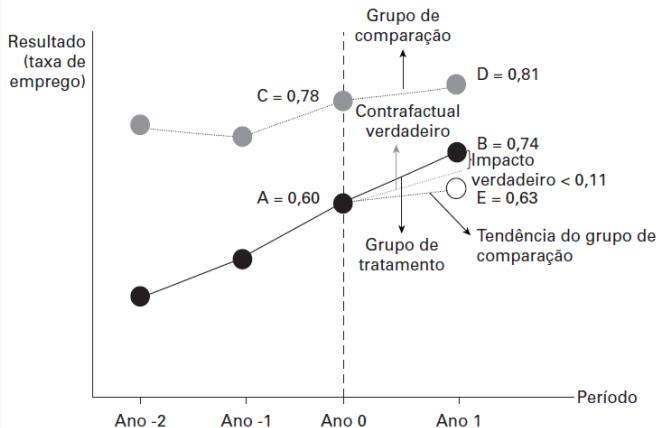
- Evolução do grupo controle é um contrafactual adequado para o grupo tratado
- Validade: antes do tratamento, tratado e controle se comportavam paralelamente

Esta pressuposição pode falhar, no contexto de políticas públicas, por diversas razões

- Indivíduos/locais tratados adotaram **outras práticas** que influenciam no resultado
- Locais tratados podem ser **melhor administrados** do que os não tratados
- Indivíduos/locais tratados tinham resultados **muito inferiores** no período pré-tratamento, tal que o resultado no período pós-tratamento é apenas uma regressão à média

Diferença-em-Diferenças

Figura 7.2 Diferença em diferenças quando as tendências de resultado diferem



A especificação mais comum utiliza regressões (em notação simplificada):

$$Y_{it} = \alpha + \rho \cdot D_{it} + \gamma_i + \delta_t + \varepsilon_{it}$$

- D_{it} indica se o tratamento é ativo para o indivíduo i no momento t
- γ_i são efeitos fixos de indivíduo
- δ_t são efeitos fixos de tempo

É possível incluir covariadas e tendências pré-intervenção, e também tendências temporais distintas para os indivíduos

Pontos para se verificar ao rodar um DD:

- Antes da intervenção, grupos de tratamento e controle se comportavam paralelamente?
 - Caso não, verificar se é possível incluir tendências temporais distintas e/ou controlar por observáveis
- Adoção do tratamento é endógena?
 - Se sim, avaliar quais fatores levam a adoção do tratamento, pois podem comprometer a suposição de tendências paralelas
- Há mudanças na composição dos grupos de controle e de tratamento?
 - Se sim, talvez seja interessante usar pareamento para obter um grupo controle similar ao tratado

Diferença-em-Diferenças: Exemplo

Card e Krueger (1994): usam DD para estudar se o aumento de salário mínimo em Nova Jérsei (NJ) em 1992, de \$4.25/hora para \$5.05/hora, aumentou o desemprego neste estado em relação à Pensilvânia (PA)

- **Grupo tratado:** restaurantes fast-food em NJ
- **Grupo controle:** restaurantes fast-food em PA
- **Tratamento:** exposição à nova lei, ou seja, estar localizado em NJ
- **Variável de interesse:** nível de emprego (*full-time employment*, FTE)
- **Amostra:** 410 unidades de redes de fast-food (Burger King, KFC, Wendy's e Roy Rogers)
 - 331 unidades em NJ
 - 79 unidades em PA

Diferença-em-Diferenças: Exemplo

Card e Krueger (2000)

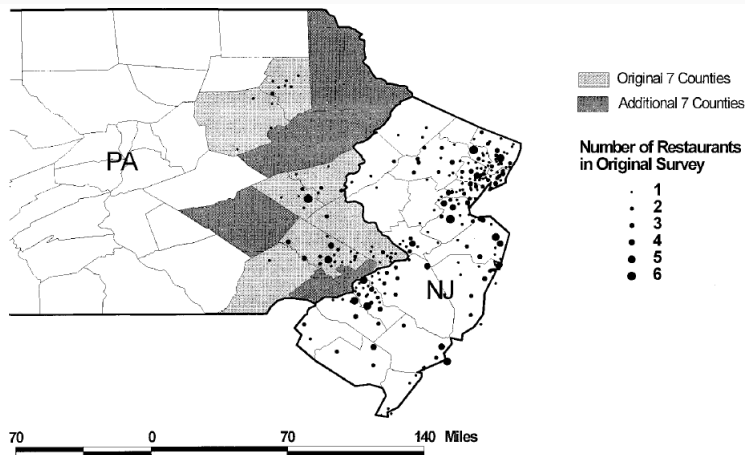


FIGURE 1. AREAS OF NEW JERSEY AND PENNSYLVANIA COVERED BY ORIGINAL SURVEY AND BLS DATA

Diferença-em-Diferenças: Exemplo

Card e Krueger (2000)

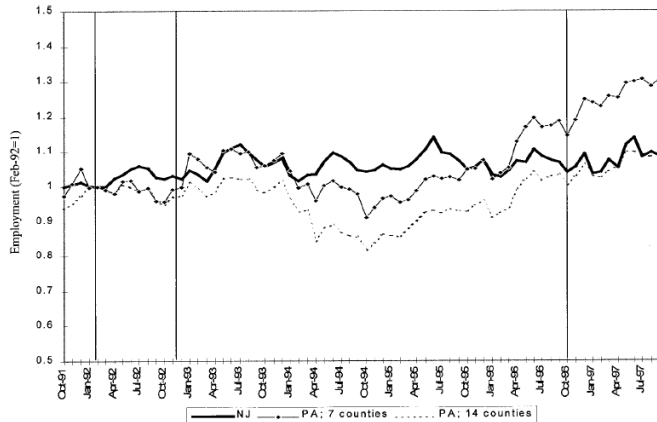


FIGURE 2. EMPLOYMENT IN NEW JERSEY AND PENNSYLVANIA FAST-FOOD RESTAURANTS, OCTOBER 1991 TO SEPTEMBER 1997

Note: Vertical lines indicate dates of original Card-Krueger survey and the October 1996 federal minimum-wage increase.

Source: Authors' calculations based on BLS ES-202 data.

Diferença-em-Diferenças: Exemplo

O principal modelo estimado possui duas versões

$$\Delta E_i = a + \mathbf{b}\mathbf{X}_i + c \text{ NJ}_i + \varepsilon_i$$

$$\Delta E_i = a' + \mathbf{b}'\mathbf{X}_i + c' \text{ GAP}_i + \varepsilon'_i$$

- ΔE_i : mudança no nível de emprego na loja i
- \mathbf{X}_i : conjunto de características da loja i
- NJ_i : igual a 1 para lojas em NJ
- GAP_i : *wage gap*, medida alternativa de impacto do tratamento, baseado no nível de salário no período pré-tratamento na loja i (W_{0i})
 - $\text{GAP}_i = 0$ para lojas em PA, ou lojas em NJ em que $W_{0i} \geq 5.05$
 - $\text{GAP}_i = (5.05 - W_{0i})/W_{0i}$ para outras lojas em NJ

Diferença-em-Diferenças: Exemplo

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

Independent variable	Model				
	(i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	2.33 (1.19)	2.30 (1.20)	—	—	—
2. Initial wage gap ^a	—	—	15.65 (6.08)	14.92 (6.21)	11.91 (7.39)
3. Controls for chain and ownership ^b	no	yes	no	yes	yes
4. Controls for region ^c	no	no	no	no	yes
5. Standard error of regression	8.79	8.78	8.76	8.76	8.75
6. Probability value for controls ^d	—	0.34	—	0.44	0.40

Notes: Standard errors are given in parentheses. The sample consists of 357 stores with available data on employment and starting wages in waves 1 and 2. The dependent variable in all models is change in FTE employment. The mean and standard deviation of the dependent variable are -0.237 and 8.825 , respectively. All models include an unrestricted constant (not reported).

^aProportional increase in starting wage necessary to raise starting wage to new minimum rate. For stores in Pennsylvania the wage gap is 0.

^bThree dummy variables for chain type and whether or not the store is company-owned are included.

^cDummy variables for two regions of New Jersey and two regions of eastern Pennsylvania are included.

^dProbability value of joint F test for exclusion of all control variables.

Conclusões

- O nível de emprego em restaurantes fast-food de NJ não foi impactado pelo aumento do salário mínimo, com relação à PA
 - Coeficiente estimado (2.33) não é o dobro do erro-padrão (1.19), i.e., não é estatisticamente significativo
 - Coeficiente estimado é positivo; se houve algum efeito, foi o de aumentar o nível emprego
- Esta é uma evidência contrária teoria econômica convencional, que afirma que aumentos de salário mínimo geram desemprego

Estudo de Evento

Em estudos com DD, é comum apresentar gráficos de **estudo de evento** (*event-study*)

- A diferença entre tratamento-controle é normalizada para zero no período anterior à adoção
- Efeito do tratamento é calculado para cada período pós-tratamento, e plotado juntamente com os intervalos de confiança

Estudo de Evento: Exemplo

Cornwell e Cunningham (2013): usam DD com *event-study* para estudar se legalização do aborto em alguns estados antes da legalização federal causou queda na infecção de adolescentes por gonorreia

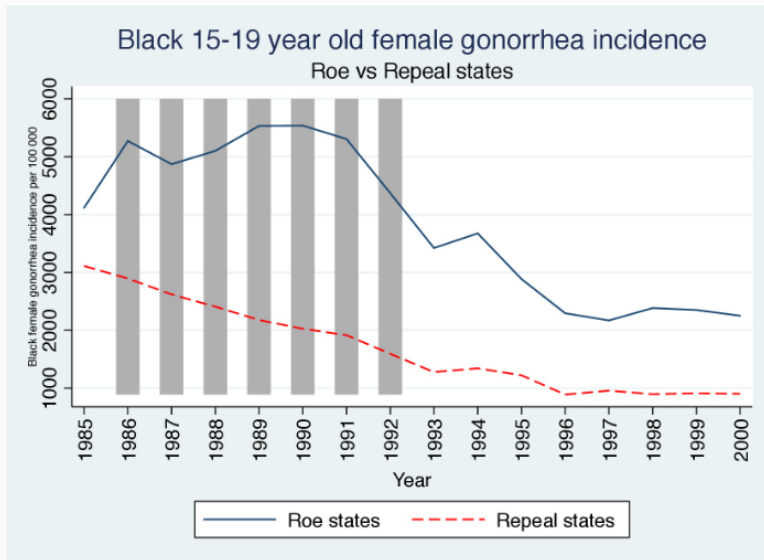
- **Tratamento:** legalização estadual do aborto em 1970 (*Alaska, California, Hawaii, New York, Washington*), antes da legalização federal (*Roe v. Wade*, 1973)
- **Grupo tratado:** indivíduos entre 15-19 anos nascidos em estados dos EUA que legalizaram o aborto antes da legislação federal
- **Grupo controle:** indivíduos entre 15-19 anos nascidos em estados que legalizaram o aborto somente após a legislação federal
- **Variável de interesse:** log da taxa de infecção por gonorreia
- **Amostra:** dados coletados pelo CDC para 5 faixas etárias

Estudo de Evento: Exemplo

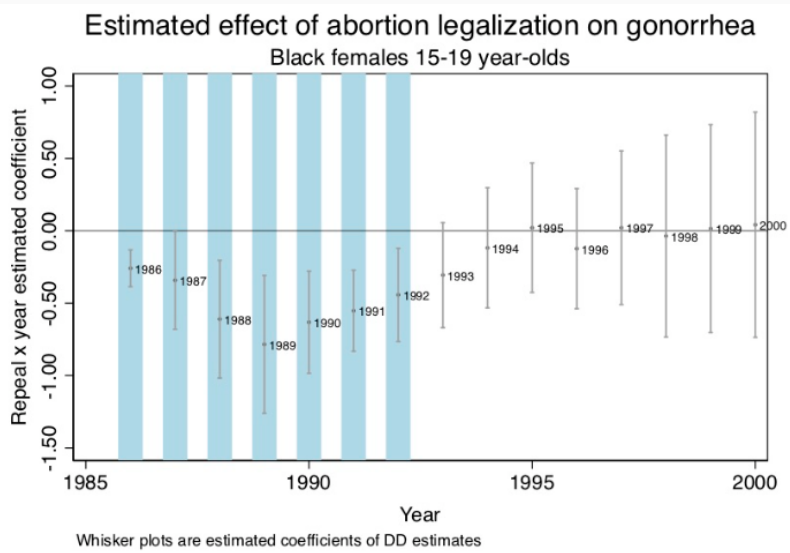
Age in calendar year	CDC Surveillance Data in Calendar Year															
	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
15	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
16	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
17	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
18	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82
19	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81
20	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
21	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
22	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78
23	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77
24	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76
25	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
26	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
27	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
28	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
29	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71
Number of cohorts (age 15-19) exposed, reforms in 71, 74	Repeal (1)															
	0	1	2	3	4	5	5	5	5	5	5	5	5	5	5	5
	No Repeal (2)															
	0	0	0	0	1	2	3	4	5	5	5	5	5	5	5	5
Difference (3)																
	0	1	2	3	3	3	2	1	0	0	0	0	0	0	0	0

Figure 1. Period-cohort diagram.

Estudo de Evento: Exemplo



Estudo de Evento: Exemplo



Pareamento

Pareamento

Pareamento (*propensity score matching*) é um método de avaliação não-experimental

Sem aleatorização, tratado e controle podem diferir em características que afetam tanto a **participação** no tratamento quanto a **variável de interesse**

- Possível viés de seleção

Os métodos de *matching* (ou pareamento) existem para permitir comparações entre unidades tratadas e unidades controle que sejam **semelhantes**

- “Semelhança”: é definida em termos do *propensity score*
- *Propensity score*: probabilidade de que uma unidade receba o tratamento, dado um conjunto de variáveis observáveis

Suposições necessárias:

- **Independência condicional:** $(Y_1, Y_0) \perp D \mid X$
 - Condicional às características observadas X , resultados potenciais são independentes do status de tratamento
 - É o mesmo que dizer que a alocação do tratamento D é aleatória, condicional a X
- **Suporte comum:** $0 < P(D = 1 \mid X) < 1$
 - Para cada valor de X , há uma probabilidade positiva de que indivíduos sejam tanto tratados quanto não tratados

1. Caracterizar o modelo de seleção do *propensity score*

- Escolher variáveis que afetem a probabilidade de participação (explicitamente ou não) e que **não sejam** afetadas pelo tratamento

2. Estimar o *propensity score*

- Usar um modelo de escolha binária (probit ou logit) e calcule os scores estimados
- De modo geral, a escolha entre logit probit não deve fazer muita diferença

3. Realizar o *matching*

- Escolher um algoritmo de pareamento: com reposição ou sem? Quantas unidades de tratamento para cada controle? Qual o critério para definir similaridade?

4. Estimar os resultados

- Como os grupos são comparáveis e aleatórios, o efeito do tratamento é obtido via teste de diferença de médias, como num RCT
- Avaliar os coeficientes estimados e calcular corretamente o erro-padrão
- Realizar testes de robustez com diferentes especificações, reavaliando os passos anteriores

Referências

Referências

ANGRIST, J., E J. PISCHKE (2015). *Mastering 'Metrics: The Path from Cause to Effect*. New Jersey, US: Princeton University Press.

GERTLER, P., S. MARTÍNEZ, P. PREMAM, L. RAWLINGS, E C. VERMEERSCH (2016). *Avaliação de Impacto na Prática* (2ª ed.). Washington, D.C.: Banco Mundial, 335 p.

CARD, D., E A. KRUEGER (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4), 772–793.

CARD, D., E A. KRUEGER (2000). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply. *American Economic Review*, 90(5), 1397–1420.

CORNWELL, C., E S. CUNNINGHAM (2013). The Long-run Effect of Abortion on Sexually Transmitted Infections. *American Law and Economics Review*, 15(1), 381–407.

Referências

HEINRICH, C., A. MAFFOLI, G. VAZQUEZ (2010). *A Primer for Applying Propensity-Score Matching*. Inter-American Development Bank.

NEUMARK, D., E W. WASCHER (2000). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Comment. *American Economic Review*, 90(5), 1362–1396.

Versão online do livro de Scott Cunningham: [https://https://mixtape.scunning.com/](https://mixtape.scunning.com/)

Notas de aula da Prof^a. Letícia Nunes.