

Diferença em diferenças

Avaliação de Políticas Públicas B

04, 09 e 11 de junho de 2025

Leitura básica:

GERTLER, Paul J. et al. **Avaliação de impacto na prática**. Washington: Banco Interamericano de Desenvolvimento; Banco Mundial, 2018. [Capítulo 7: Diferença em diferenças.] Disponível em: <https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464808890.pdf>

Leitura complementar:

FACURE, Matheus. **Causal inference for the brave and true**. 2022. [Capítulo 13: Difference-in-differences.] Disponível em: <https://matheusfacure.github.io/python-causality-handbook/landing-page.html>

Interação com dummy

- Nossa discussão anterior sobre variáveis dummy tratou de casos em que a variável categórica afetava a dependente no intercepto (i.e., no nível), mas não na inclinação. Endereçávamos questões do tipo “Há uma diferença sistemática entre salários (Y) de mulheres e homens, para um mesmo nível de escolaridade (X)?”

$$\text{wage} = \beta_0 + \beta_1 \text{female} + \beta_2 \text{educ} + \varepsilon$$

Qual a representação gráfica desta equação?

- Adicionalmente, pode ser relevante saber se a relação entre a variável dependente (Y) e uma variável independente (X) difere para subgrupos da população. Isso corresponde a questões do tipo “Mulheres e homens obtêm o mesmo retorno do investimento em educação?”. Para endereçar esta questão é preciso interagir variáveis (i.e., multiplicá-las), como segue:

$$\text{wage} = \beta_0 + \beta_1 \text{female} + \beta_2 \text{educ} + \beta_3 (\text{female} * \text{educ}) + \varepsilon$$

- Na equação com interação, o sexo afeta tanto o intercepto como a inclinação
Salário estimado:

$$\text{Male: } \widehat{\text{wage}} = \widehat{\beta}_0 + \widehat{\beta}_2 \text{educ}$$

$$\text{Female: } \widehat{\text{wage}} = \widehat{\beta}_0 + \widehat{\beta}_1 + (\widehat{\beta}_2 + \widehat{\beta}_3) \text{educ}$$

Efeito estimado do aumento de uma unidade em educ:

$$\text{Male: } \widehat{\beta}_2$$

$$\text{Female: } \widehat{\beta}_2 + \widehat{\beta}_3$$

$\widehat{\beta}_3$, o coeficiente estimado do termo de interação, é nossa estimativa do efeito diferencial da educação no salário de mulheres em relação a homens

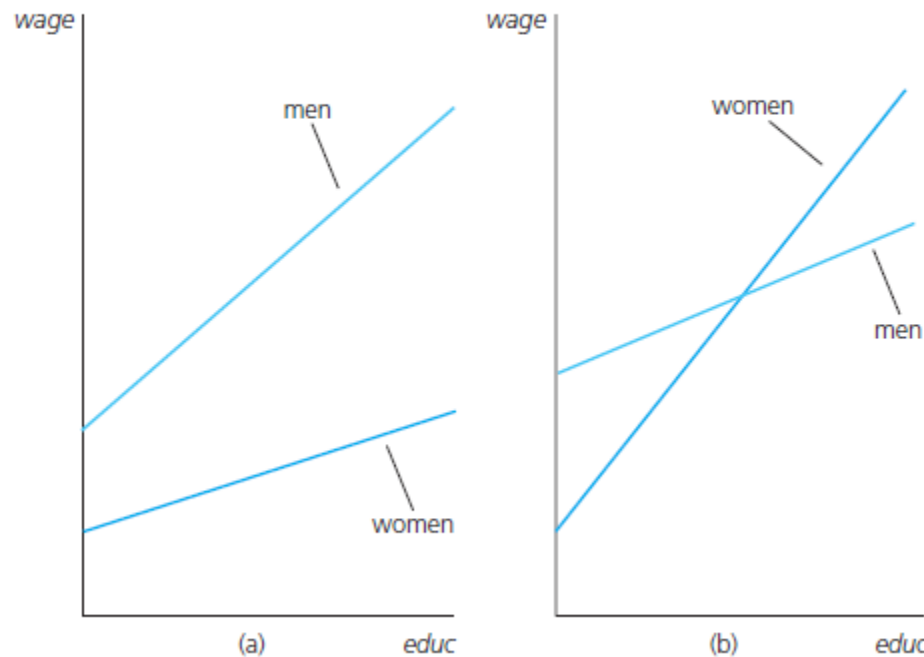
Interação com dummy: Representações gráficas

Equação populacional:

$$\text{wage} = \beta_0 + \delta_0 \text{female} + \beta_1 \text{educ} + \delta_1 \text{female} * \text{educ} + \varepsilon$$

FIGURE 7.2

Graphs of equation (7.16): (a) $\delta_0 < 0, \delta_1 < 0$; (b) $\delta_0 < 0, \delta_1 > 0$.



A inclusão de dummy e interação permite **testar** se os grupos possuem intercepto e inclinação próprios, respectivamente.

Fonte: Wooldridge (2009, p. 240).

Exemplo: Interação com dummy

$$\text{wage} = \beta_0 + \beta_1 \text{female} + \beta_2 \text{educ} + \beta_3 (\text{female} * \text{educ}) + \varepsilon$$

$$\text{femeduc} = \text{female} * \text{educ}$$

Quais coeficientes são estatisticamente diferentes de zero? O que isso implica para as conclusões desta análise?

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.539476	.0642229	8.40	0.000	.4133089	.6656432
female	-1.198523	1.32504	-0.90	0.366	-3.801589	1.404543
femeduc	-.085999	.1036388	-0.83	0.407	-.2895994	.1176014
_cons	.2004963	.8435616	0.24	0.812	-1.456696	1.857689

$$\widehat{\text{wage}} = 0,20 - 1,20\text{female} + 0,54\text{educ} - 0,09(\text{educ} * \text{female})$$

- Mulheres sem qualquer educação formal ganham, em média, 1,20 menos que os homens
- Um ano adicional de educação formal associa-se com um aumento de 0,54 no salário de homens
- Um ano adicional de educação formal associa-se com um aumento de 0,45 (0,54 – 0,09) no salário de mulheres
- O efeito diferencial de um ano adicional de educação formal para mulheres em relação a homens é de -0,09

Interação entre duas dummies

$$wage = \beta_0 + \beta_1 fem + \beta_2 mar + \beta_3 (fem \cdot mar) + \varepsilon$$

$$\hat{wage} = 5.2 - .56 fem + 2.8 mar - 2.9(fem \cdot mar)$$

Interpretação:

Group	Values	Predicted wage
Unmarried men	fem=0, mar=0	\$ 5.2
Unmarried women	fem=1, mar=0	\$ 4.64 (5.2-.56)
Married men	fem=0, mar=1	\$ 8.00 (5.2 +2.8)
Married women	fem=1, mar=1	\$ 4.54 (5.2-.56+2.8-2.9)

- Coeficiente de *fem*: Estima-se que o salário médio de mulheres solteiras seja 0,56 inferior ao de homens solteiros
- Coeficiente de *mar*: Estima-se que o salário médio de homens casados seja 2,8 superior ao de homens solteiros
- Coeficiente de *fem*mar*: Estima-se que o impacto médio do casamento no salário seja 2,90 inferior para mulheres em relação a homens

Como interpretar interações entre variáveis contínuas?

The UCLA logo consists of the letters "UCLA" in white, bold, sans-serif font, set against a solid blue rectangular background.

Institute for Digital Research & Education

<https://stats.idre.ucla.edu/stata/faq/how-can-i-explain-a-continuous-by-continuous-interaction-stata-12/>

<https://stats.idre.ucla.edu/r/faq/how-can-i-explain-a-continuous-by-continuous-interaction/>

A abordagem que demonstraremos é calcular inclinações simples, ou seja, as inclinações da variável dependente na variável independente quando a variável moderadora é mantida constante em diferentes combinações de valores de muito baixo a muito alto.

Estudos observacionais: estratégias e abordagens para enfrentar heterogeneidade não observada

Problema

Em estudos observacionais, não há presunção de comparabilidade entre grupo tratado e não tratado (antes da exposição ao tratamento). De fato, esses grupos podem diferir substancialmente, tanto nas características observáveis como nas não observáveis. Por isso, a associação observada entre T (tratamento) e Y (variável de resultado) não pode ser considerada uma estimativa do efeito causal de T em Y. Para a estimação de efeito, teremos de usar alguma estratégia para enfrentar a não comparabilidade entre os grupos.

Estratégia

1

Modelar o mecanismo de tratamento (T)

2

Modelar a variável de resultado (Y)

3

Uma combinação das estratégias anteriores

Exemplo de abordagem

- **Inverse Probability of Treatment Weighting (IPTW)** com probabilidade de tratamento definida como **propensity score (PS)**
- Matching (pode envolver PS)
- Controle estatístico (pode envolver PS)
- Diferença em diferenças
- Efeitos fixos
- Estimação do tipo double robust

Estudos observacionais: estratégias e abordagens para enfrentar heterogeneidade não observada

Problema

Em estudos observacionais, não há presunção de comparabilidade entre grupo tratado e não tratado (antes da exposição ao tratamento). De fato, esses grupos podem diferir substancialmente, tanto nas características observáveis como nas não observáveis. Por isso, a associação observada entre T (tratamento) e Y (variável de resultado) não pode ser considerada uma estimativa do efeito causal de T em Y. Para a estimação de efeito, teremos de usar alguma estratégia para enfrentar a não comparabilidade entre os grupos.

Estratégia

1

Modelar o mecanismo de tratamento (T)

2

Modelar a variável de resultado (Y)

3

Uma combinação das estratégias anteriores

Exemplo de abordagem

- **Inverse Probability of Treatment Weighting (IPTW)** com probabilidade de tratamento definida como **propensity score (PS)**
- Matching (pode envolver PS)
- Controle estatístico (pode envolver PS)
- **Diferença em diferenças**
- Efeitos fixos
- Estimação do tipo double robust

**Num reino distante, há alguns
anos...**



**Visão externa da sua casa
(com zoom potente daria para
ver você estudando lá dentro)**

Num reino distante, há alguns anos...



Visão externa da sua casa
(com zoom potente daria para
ver você estudando lá dentro).

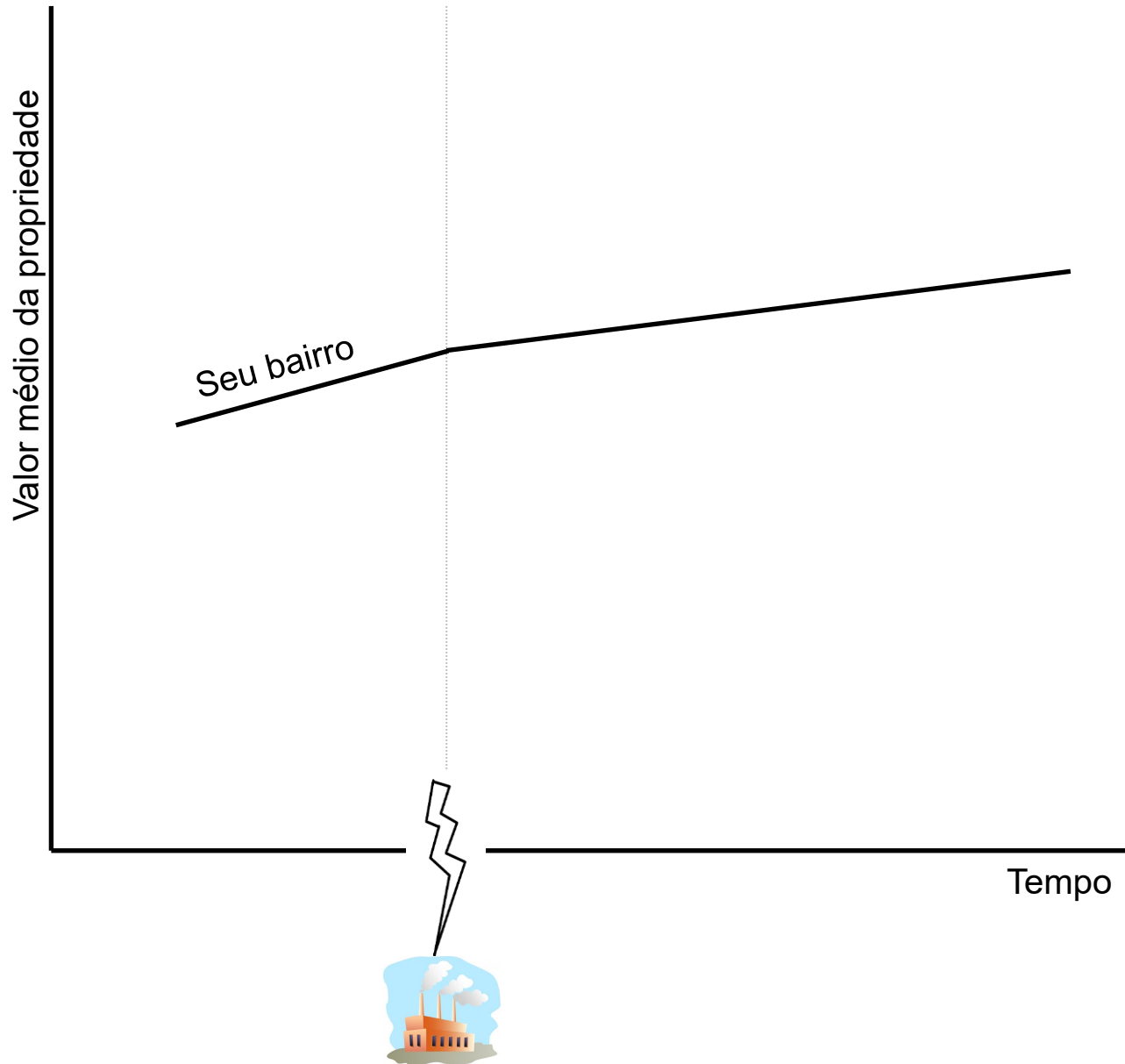
...um incinerador de lixo foi instalado no
seu bairro... (que legal!)



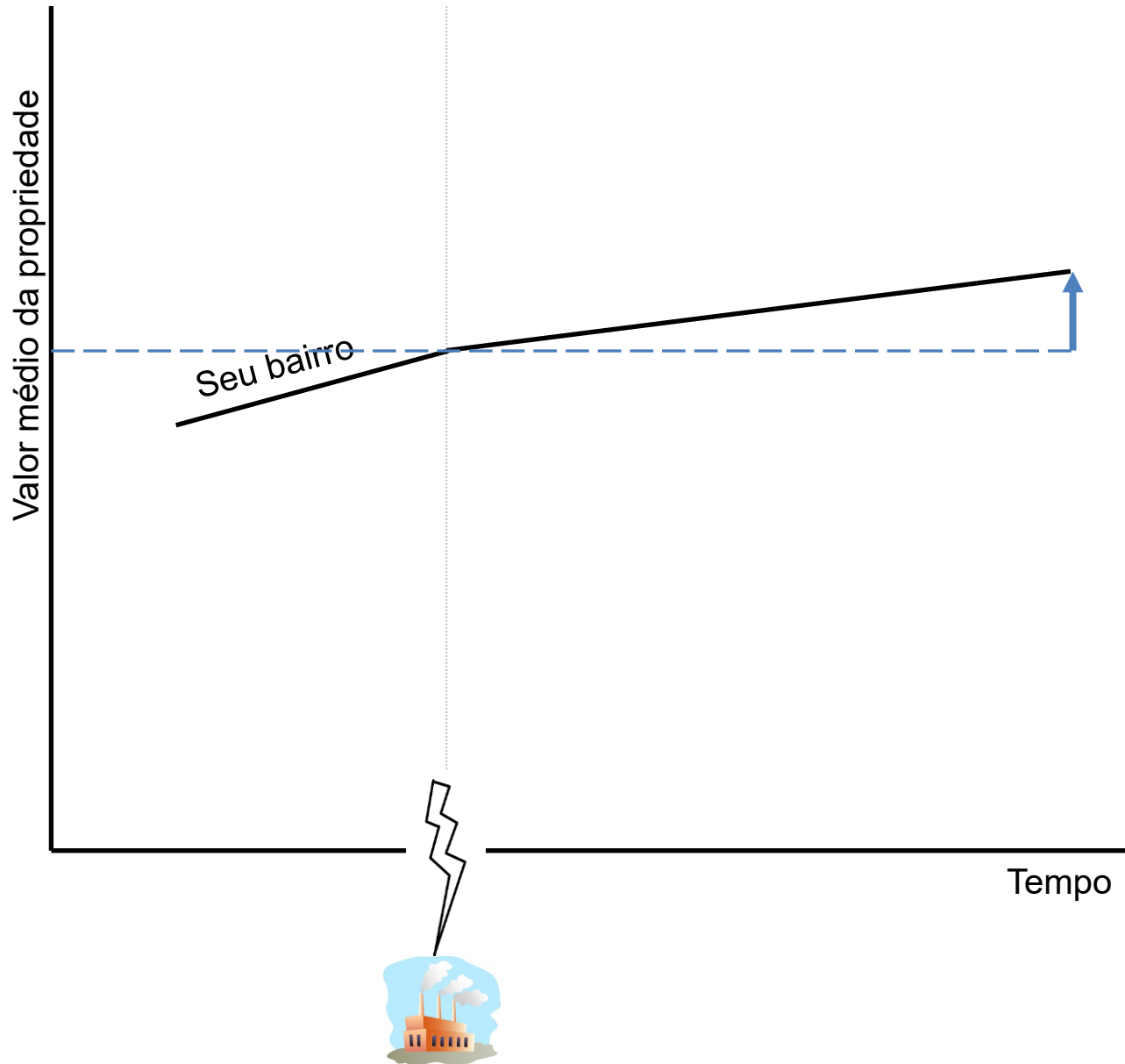
- ?** Quanto valor perdeu sua propriedade como resultado da instalação do incinerador?
- ?** Quanto de desconto no IPTU seria justo você e os vizinhos pleitearem?

Detalhe: Você acredita que o valor nominal da sua casa não necessariamente diminuiu, mas pelo menos deixou de crescer na mesma velocidade de casas comparáveis distantes do incinerador.

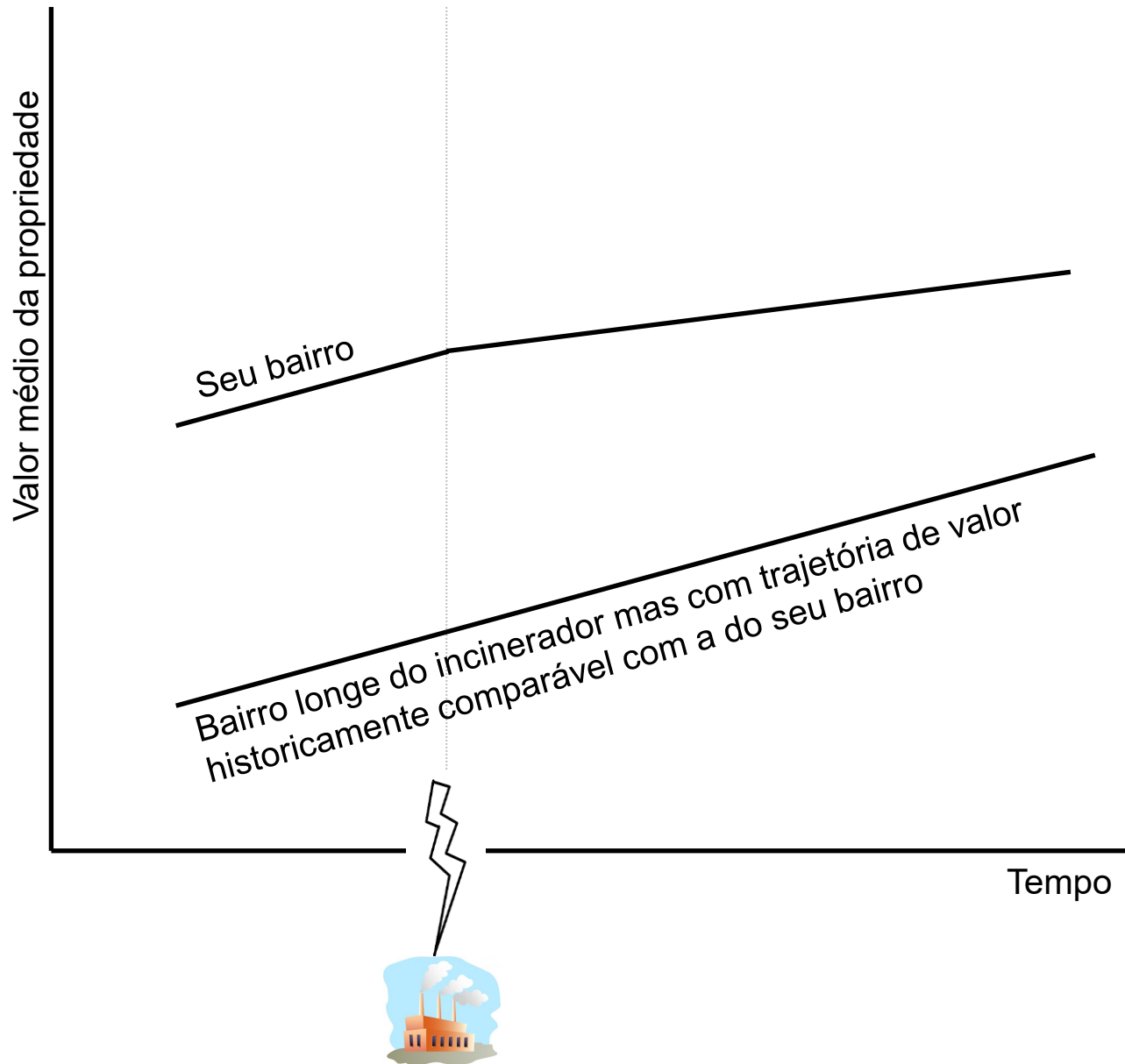
1ª Diferença: seu bairro depois e antes do incinerador



1ª Diferença: seu bairro depois e antes do incinerador

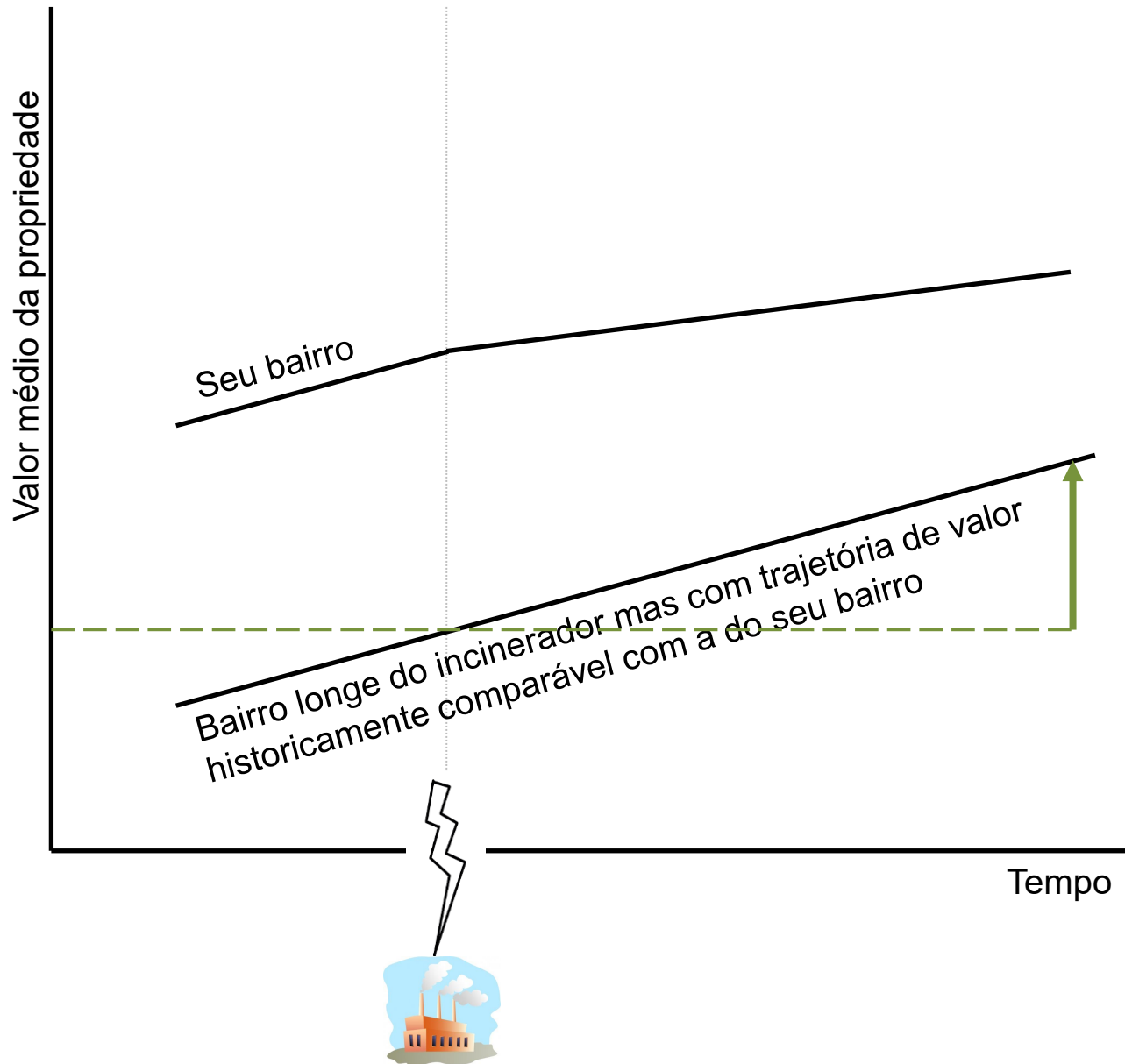


2ª Diferença: bairro comparável depois e antes do incinerador

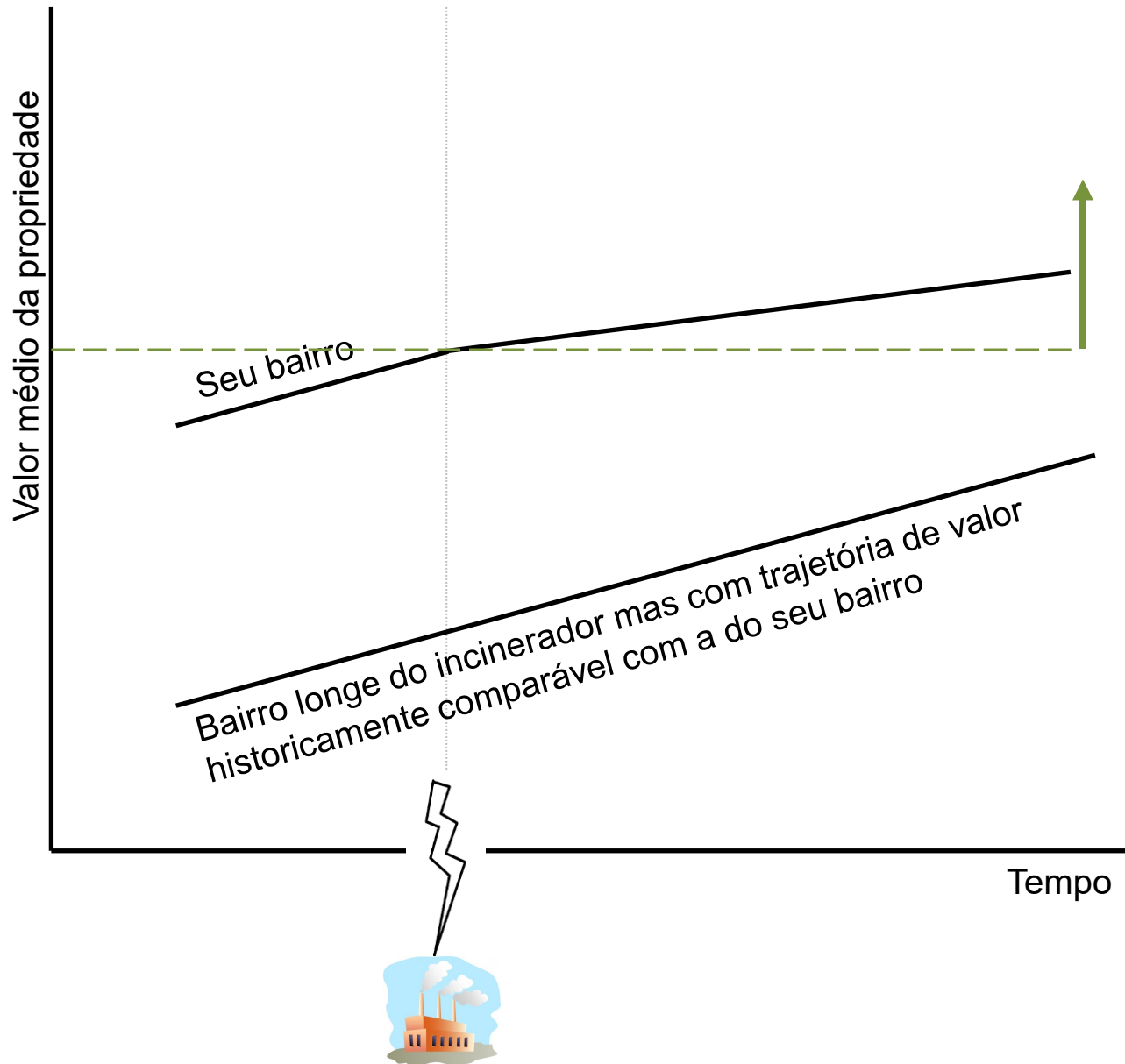


2ª Diferença:

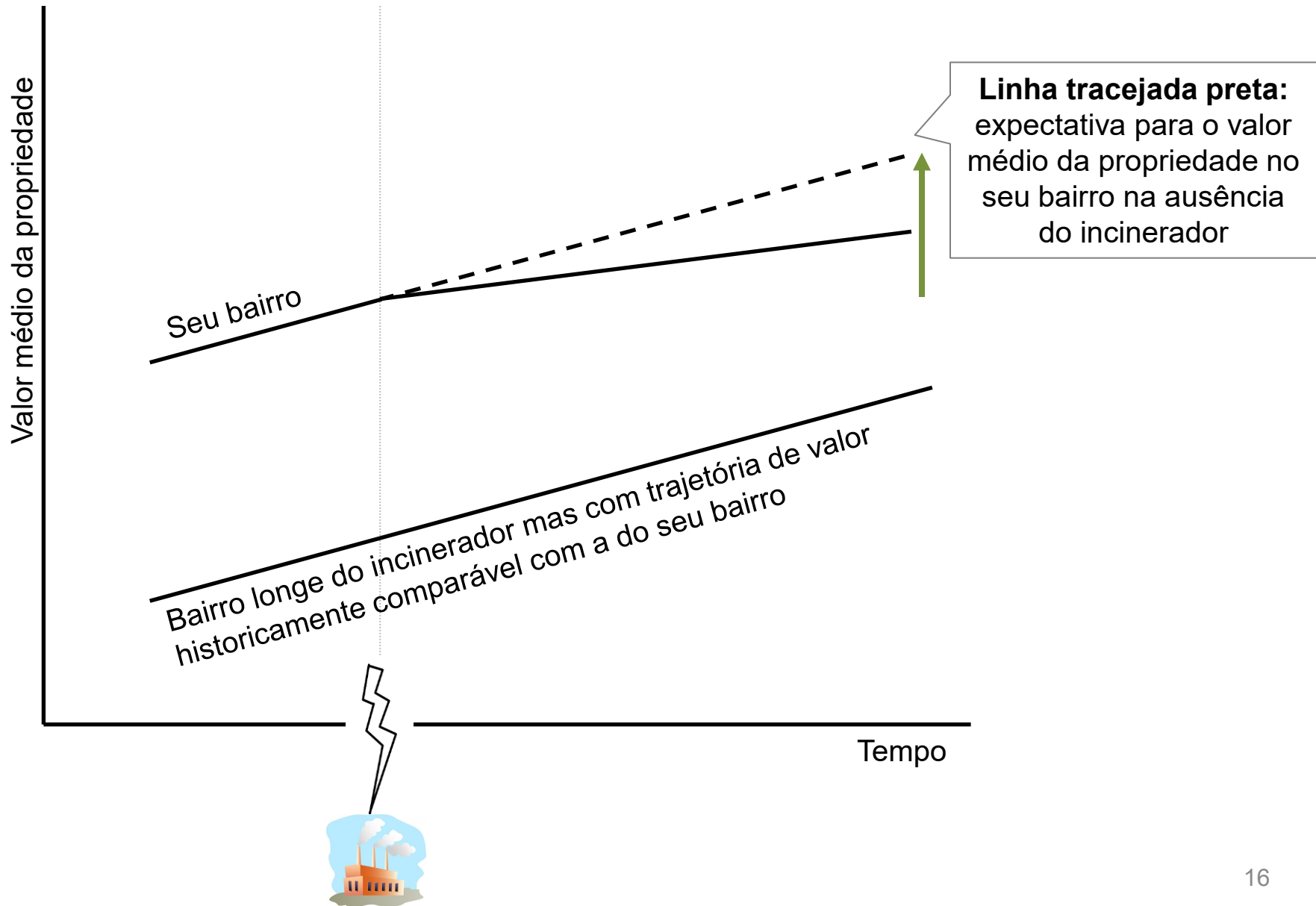
bairro comparável depois e antes do incinerador



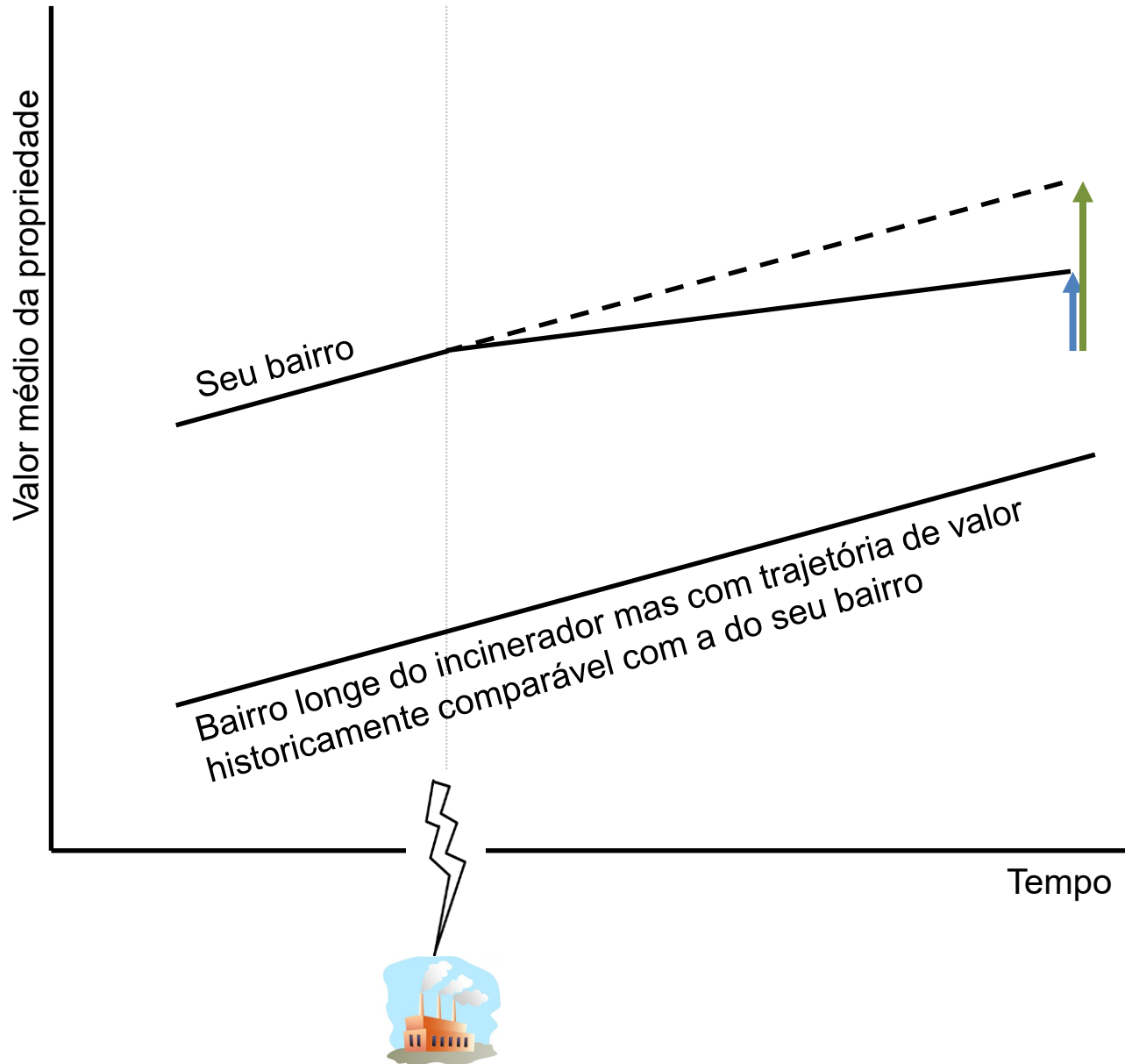
Projeção do contrafactual para seu bairro: valor da propriedade teria evoluído como no grupo de comparação



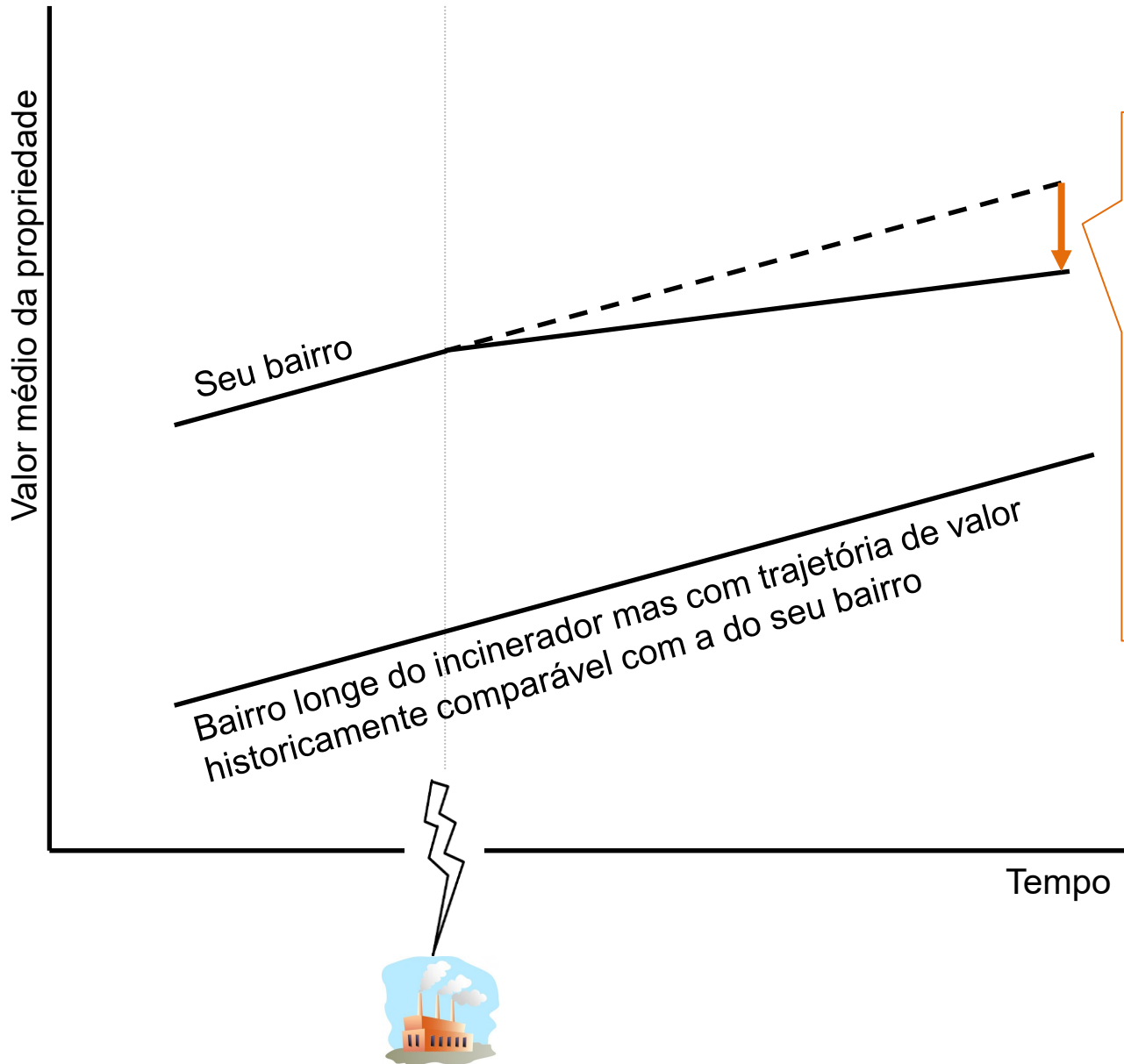
Projeção do contrafactual para seu bairro: valor da propriedade teria evoluído como no grupo de comparação



Diferença em diferenças: diferença entre 1ª e 2ª diferenças



Diferença em diferenças: diferença entre 1ª e 2ª diferenças



Efeito do incinerador no valor médio das propriedades localizadas no bairro do incinerador.

Nota: Se houver um efeito do incinerador, esse efeito só pode se apresentar:

- após a instalação do incinerador; e
- apenas na região próxima a ele.

Diferença em diferenças: requisitos e intuição

Requisitos

- Dados **longitudinais**, incluindo **pré- e pós-teste** (i.e., **pré- e pós-tratamento**)
- Existência de um **grupo de comparação** para o qual se espera uma evolução de Y igual à evolução de Y que seria esperada no grupo de tratamento na ausência de tratamento
 - **Atenção:** **não** se trata da existência de um grupo de comparação **igual (em nível)** ao grupo de tratamento, mas de um grupo de comparação com **trajetória** que ofereça um contrafactual plausível para a trajetória do grupo de tratamento (common trend assumption)

Intuição

- Compara mudanças nos resultados ao longo do tempo, entre o grupo de tratamento e o de comparação
 - **Diferenças depois vs. antes do tratamento:** eliminam *confounders* (Zs) constantes no tempo mas que variam entre grupos, pois comparam um grupo com ele mesmo (produzindo deltas temporais)
 - **Diferença entre deltas temporais:** elimina *confounders* (Zs) variantes no tempo e que afetam os dois grupos igualmente (condições ambientais)

Combinando duas estimativas falsas para o contrafactual

A abordagem da diferença em diferenças faz o que seu nome sugere. Ela combina as duas estimativas falsas do contrafactual (comparações antes e depois e comparações entre aqueles que optam por se inscrever e aqueles que optam por não se inscrever) para produzir uma melhor estimativa do contrafactual.

Gertler et al. (2018, p. 144)

DinD requer grupo de comparação + pré- e pós-teste

- Para analisarmos as diferenças entre os dois grupos, precisamos de dados de **antes** e de **depois** do evento/tratamento para os **dois grupos**
 - Idealmente, dados de painel; no mínimo, dados de corte transversal empilhados

APÊNDICE:

Tipos de dados



DinD requer grupo de comparação + pré e pós-teste

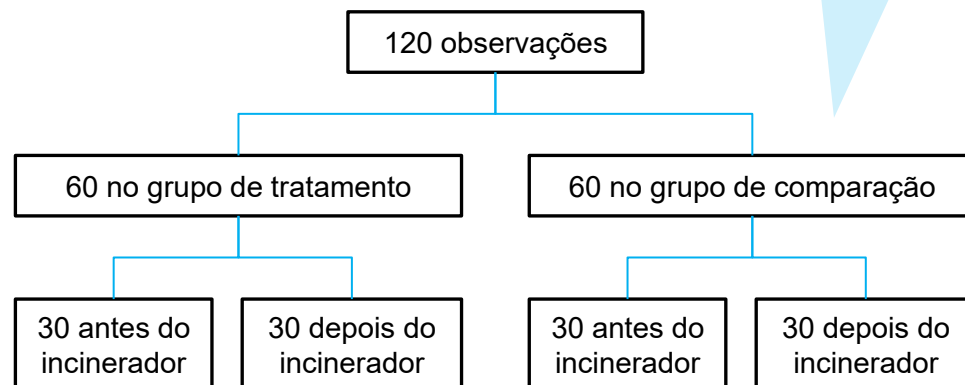
- Para analisarmos as diferenças entre os dois grupos, precisamos de dados de **antes** e de **depois** do evento/tratamento para os **dois grupos**
 - Idealmente, dados de painel; no mínimo, dados de corte transversal empilhados
- Assim, nossa amostra é dividida em **quatro subgrupos**:
 - i. Grupo de tratamento antes do tratamento
 - ii. Grupo de tratamento depois do tratamento
 - iii. Grupo de comparação antes do tratamento
 - iv. Grupo de comparação depois do tratamento

Ilustração: banco de dados em painel para o exemplo do incinerador

Observação	Propriedade	Grupo (=1 se tratamento)	Timing da observação (=1 se após incinerador)	Preço da propriedade (em milhares de reais)	Área construída (em metros quadrados)
1	Casa 01	1	0	75	50
2	Casa 01	1	1	80	50
3	Casa 02	0	0	75	45
4	Casa 02	0	1	82	50
5	Casa 03	0	0	90	60
6	Casa 03	0	1	90	60
7	Casa 04	1	0	100	70
8	Casa 04	1	1	105	70
9	Casa 05	1	0	76	55
10	Casa 05	1	1	80	50
11	Casa 06	0	0	80	70
12	Casa 06	0	1	99	70
...
113	Casa 57	0	0	80	50
114	Casa 57	0	1	79	50
115	Casa 58	0	0	90	70
116	Casa 58	0	1	120	70
117	Casa 59	1	0	100	75
118	Casa 59	1	1	125	75
119	Casa 60	1	0	97	60
120	Casa 60	1	1	103	65

Não é obrigatório ter grupos tratado e não tratado de mesmo tamanho

Não necessariamente serão observadas as mesmas 30 propriedades antes e depois do incinerador.



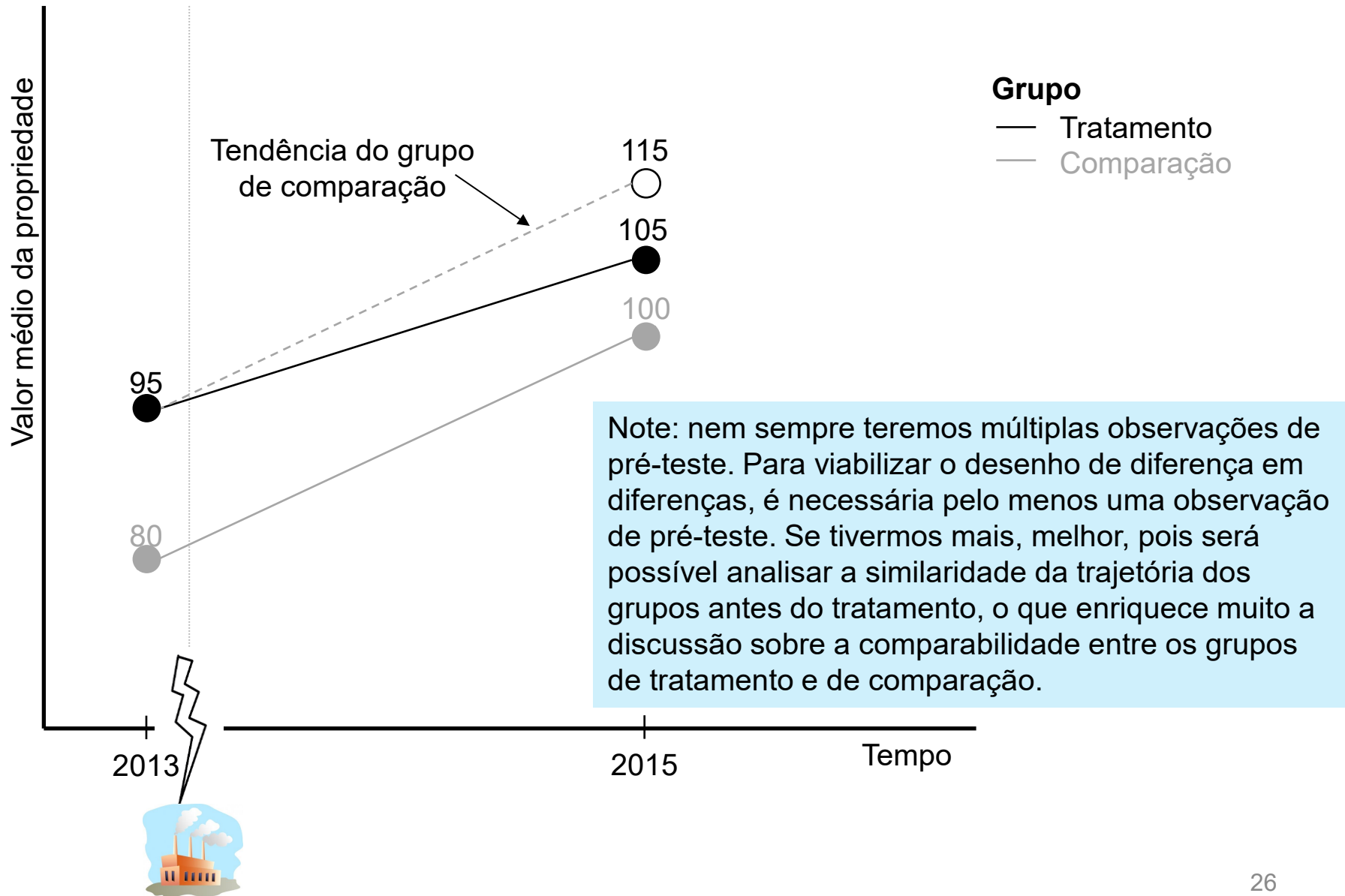
Diferença entre grupos no tempo: exemplo incinerador

$$valor = \beta_0 + \beta_1 tratado + \beta_2 depois + \beta_3(tratado * depois) + \varepsilon$$

- valor = valor da propriedade, em milhares de reais
- tratado = dummy para indicar exposição ao tratamento (1 se casa está no bairro próximo ao incinerador, 0 c. c.)
- depois = dummy para ano (1 se 2015; 0 se 2013); o incinerador foi construído em 2014

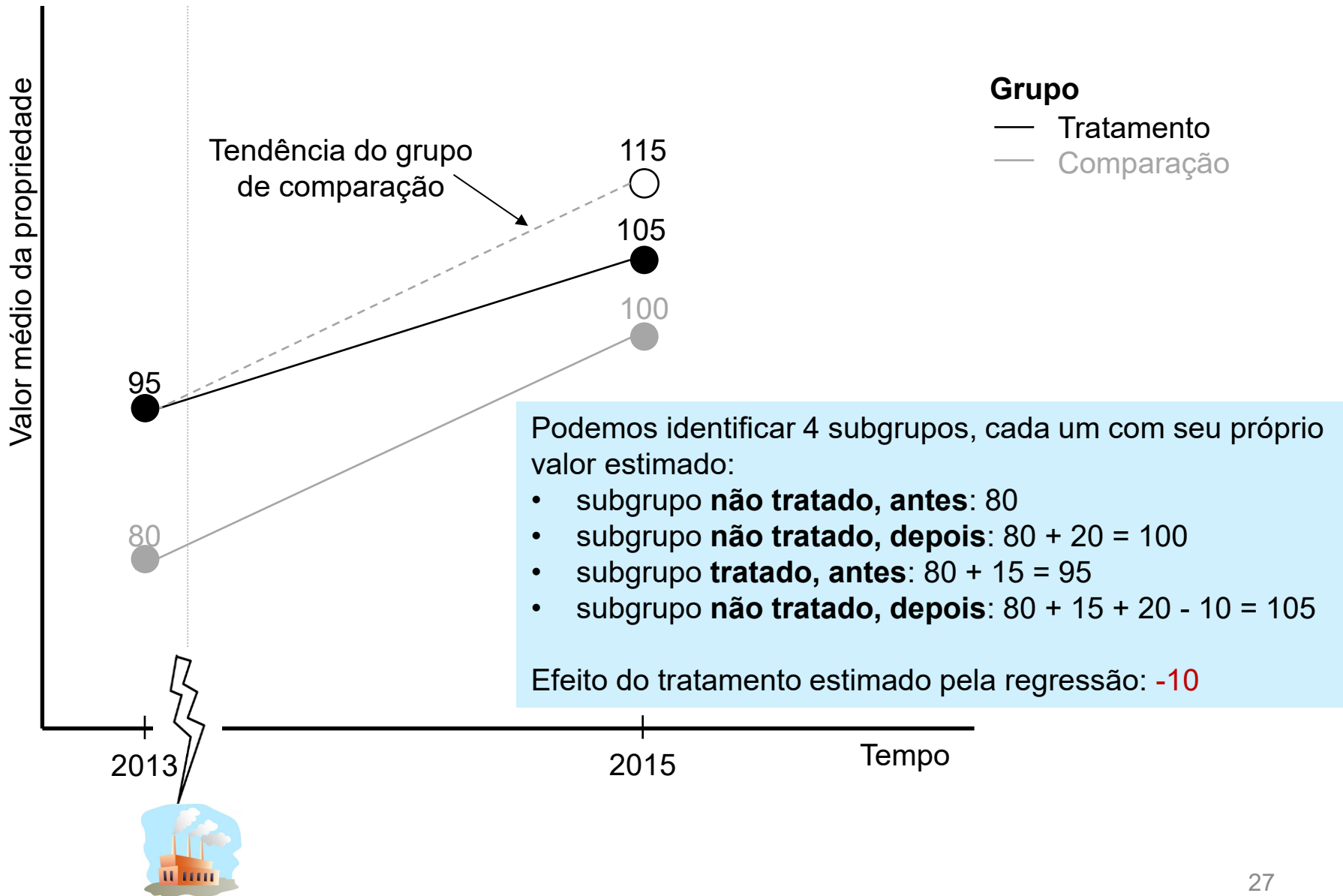
Diferença entre grupos no tempo: exemplo incinerador

$$\widehat{\text{valor}} = 80 + 15\text{tratado} + 20\text{depois} - 10(\text{tratado} * \text{depois})$$



Diferença entre grupos no tempo: exemplo incinerador

$$\widehat{\text{valor}} = 80 + 15\text{tratado} + 20\text{depois} - 10(\text{tratado} * \text{depois})$$



Calculando o efeito do incinerador sobre o valor da propriedade

Neste exemplo, não há controles, apenas a estrutura mínima da regressão de DiD

$$valor = \beta_0 + \beta_1 tratado + \beta_2 depois + \beta_3 (tratado * depois) + \varepsilon$$

Recordando o significado das variáveis:

- *valor* = valor da propriedade, em milhares de reais
- *tratado* = dummy para indicar exposição ao tratamento (1 se casa está no bairro próximo ao incinerador, 0 c. c.)
- *depois* = dummy para ano (1 se 2015; 0 se 2013); o incinerador foi construído em 2014

$$\widehat{valor} = 80 + 15tratado + 20depois - 10(tratado * depois)$$

Interpretando os coeficientes estimados:

- \$80.000: Valor médio de uma propriedade longe do incinerador em 2013
- \$15.000: Diferença média em 2013 entre o valor das casas localizadas no bairro do incinerador e o valor das casas localizadas longe do incinerador
- \$20.000: Aumento médio no valor das propriedades localizadas longe do incinerador, considerando o período 2013-2015
- -\$10.000: Variação do valor das propriedades atribuível ao incinerador (i.e., a “diferença em diferenças”)



É essencial checar a significância do β da interação

Estrutura do estimador de diferença em diferenças: regressão com interação e controles

Considere o modelo: $Y_{it} = \beta_0 + \beta_1 T_{it} + \beta_2 D_{it} + \beta_3 T_{it} * D_{it} + \beta_c \text{Controles}_{it} + \varepsilon_{it}$
em que:

Y_{it} = valor da propriedade

T_{it} = dummy para indicar exposição ao tratamento (T = 1 se propriedade fica perto do incinerador, T = 0 c.c.)

D_{it} = dummy para indicar momento da observação (D = 1 se depois da instalação do incinerador; D = 0 c.c.)

$T_{it} * D_{it}$ = termo de interação para identificar observações referentes a propriedades próximas ao local do incinerador e posteriores à implementação dele

Incluir nas células em azul os coeficientes dos termos de controle não alteraria o resultado da DinD

	Depois (D = 1)	Antes (D = 0)	Diferença
Grupo de tratamento (T = 1)	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_1$	$\beta_2 + \beta_3$
Grupo de comparação (T = 0)	$\beta_0 + \beta_2$	β_0	β_2
Diferença	$\beta_1 + \beta_3$	β_1	β_3

Preenchendo o quadro, percebemos que **β_3 corresponde à diferença entre as diferenças**, e pode ser descrito de 2 maneiras distintas, porém equivalentes:

- (depois – antes perto do incinerador) – (depois – antes longe do incinerador)
- (perto – longe depois do incinerador) – (perto – longe antes do incinerador)

Exemplo: Interação entre duas dummies (modelo de diferença em diferenças)

2048

AVALIANDO A EFETIVIDADE
DA LEI MARIA DA PENHA¹

TEXTOS PARA DISCUSSÃO

Brasília, março de 2015

SINOPSE

Daniel Cerqueira²

Mariana Vieira Martins Matos³

Ana Paula Antunes Martins⁴

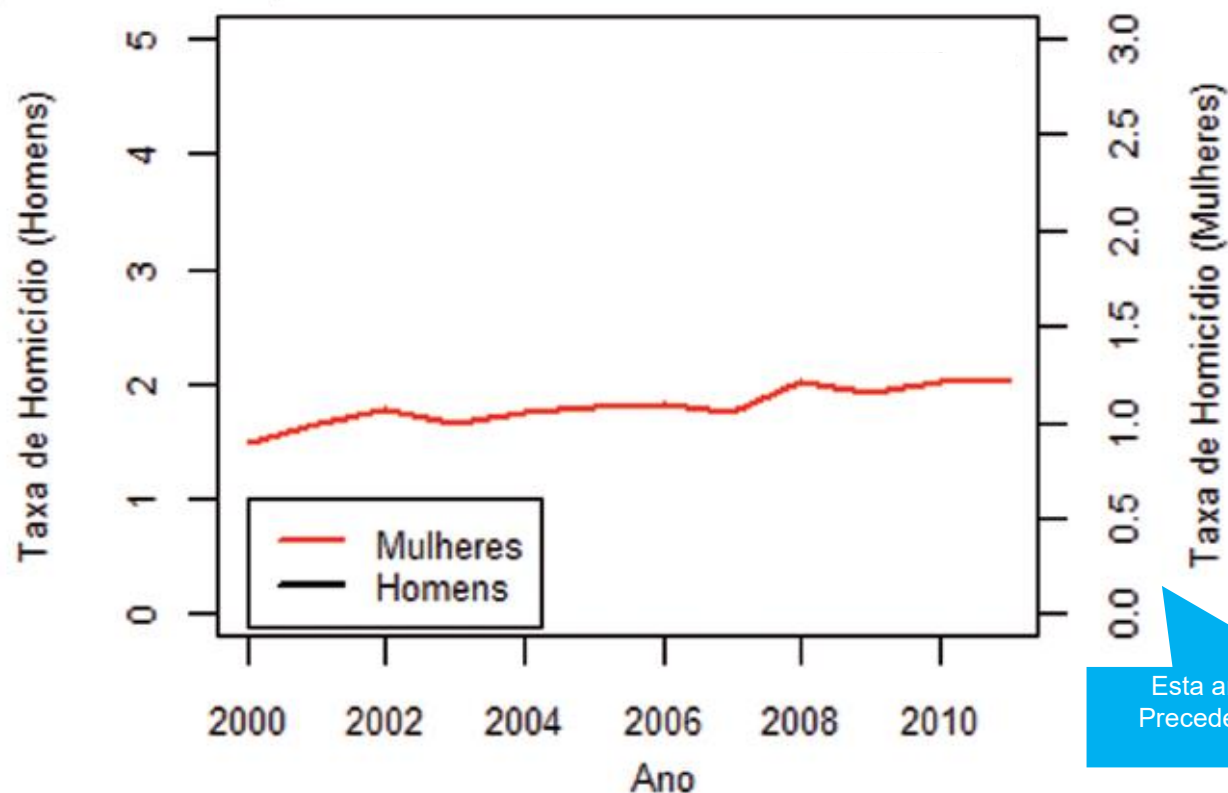
Jony Pinto Junior⁵

A superação da violência doméstica é um dos grandes desafios das políticas públicas no Brasil. A Lei nº 11.340, conhecida como Lei Maria da Penha (LMP), sancionada em 2006, representou um marco institucional importante nesse caminho, pois procurou tratar de forma integral o problema da violência doméstica, e não apenas da imputação de uma maior pena ao ofensor. Com efeito, a nova legislação ofereceu um conjunto de instrumentos para possibilitar a proteção e o acolhimento emergencial à vítima, isolando-a do agressor, ao mesmo tempo que criou mecanismos para garantir a assistência social da ofendida. Contudo, mais de oito anos após a sanção da LMP, há uma completa lacuna acerca de estudos empíricos que procurassem avaliar o seu efeito, que foi o objeto deste trabalho. Construimos um modelo de diferenças em diferenças, em que os homicídios contra mulheres que ocorreram dentro das residências foram confrontados com aqueles que acometeram os homens. Nossos resultados indicaram que a lei cumpriu um papel relevante para conter a violência de gênero, ainda que sua efetividade não tenha se dado de maneira uniforme no país, uma vez que a sua eficácia depende da institucionalização de vários serviços protetivos nas localidades, que se deu de forma desigual no território.

http://repositorio.ipea.gov.br/bitstream/11058/3538/1/td_2048.pdf

Exemplo: Interação entre duas dummies (modelo de diferença em diferenças)

GRÁFICO 2
Taxa de homicídios ocorridos em residência – Brasil (2000-2011)
(Por 100 mil habitantes)



2048

TEXTO PARA DISCUSSÃO

Brasília, março de 2015

AVALIANDO A EFETIVIDADE DA LEI MARIA DA PENHA¹

Daniel Cerqueira²
Mariana Vieira Martins Matos³
Ana Paula Antunes Martins⁴
Jony Pinto Junior⁵

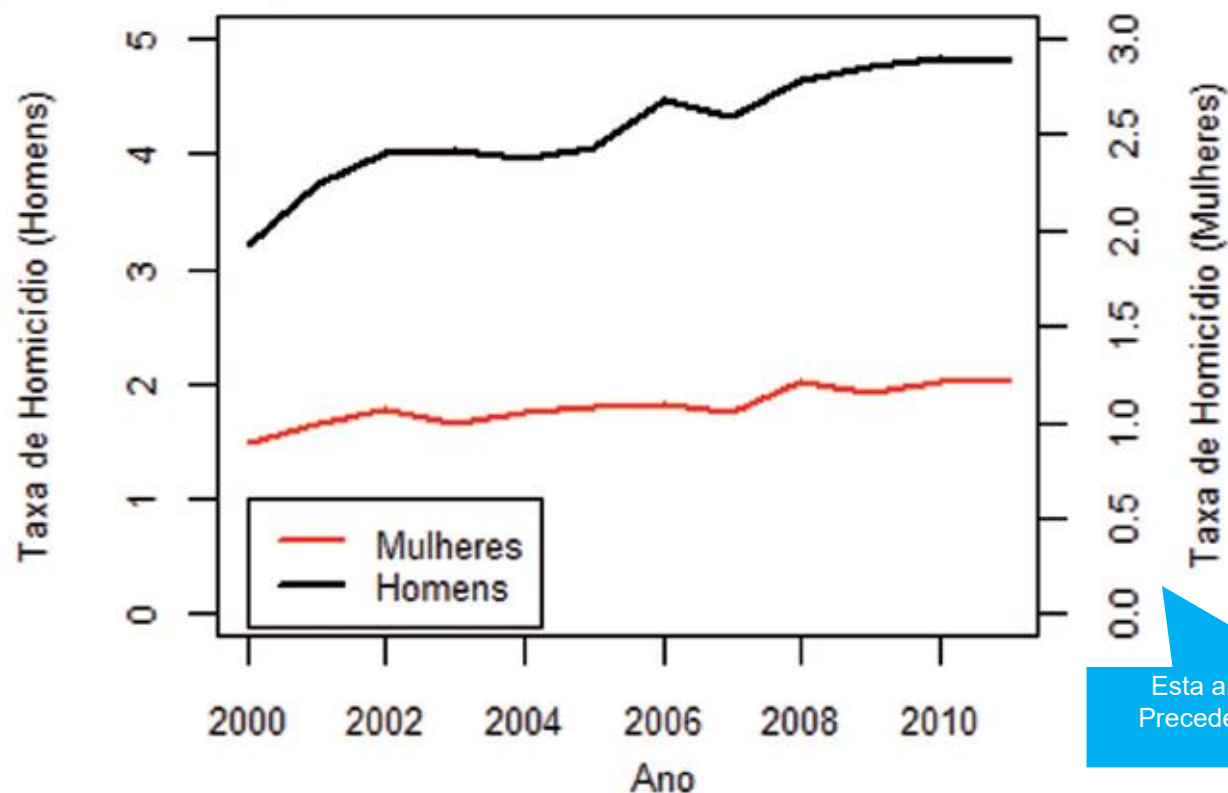
<https://repositorio.ipea.gov.br/handle/11058/3538>

Esta análise é apenas descritiva.
Precede o modelo de diferença em
diferenças.

Fonte: SIM.
Elaboração: Diest/Ipea.
Obs.: imagem cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais disponibilizados pelos autores para publicação
(nota do Editorial).

Exemplo: Interação entre duas dummies (modelo de diferença em diferenças)

GRÁFICO 2
Taxa de homicídios ocorridos em residência – Brasil (2000-2011)
(Por 100 mil habitantes)



Fonte: SIM.
Elaboração: Diest/Ipea.
Obs.: imagem cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais disponibilizados pelos autores para publicação (nota do Editorial).

2048

TEXTO PARA DISCUSSÃO

Brasília, março de 2015

AVALIANDO A EFETIVIDADE DA LEI MARIA DA PENHA¹

Daniel Cerqueira²
Mariana Vieira Martins Matos³
Ana Paula Antunes Martins⁴
Jony Pinto Junior⁵

<https://repositorio.ipea.gov.br/handle/11058/3538>

Esta análise é apenas descritiva.
Precede o modelo de diferença em
diferenças.

Exemplo: Interação entre duas dummies (modelo de diferença em diferenças)

Essa dinâmica diferenciada no movimento das taxas de homicídio entre homens e mulheres em diferentes localidades no Brasil,¹² no período considerado, justifica uma avaliação mais criteriosa, que inclua o efeito das especificidades regionais na análise, bem como controles temporais, que poderiam influenciar no efeito da LMP. Na próxima seção, passaremos a discutir a especificação do modelo empírico de *diferenças em diferenças*, discutido no início da seção 3.

4 ANÁLISE ECONOMETRICA COM BASE NO MODELO DE DIFERENÇAS EM DIFERENÇAS

Nesta seção, estimaremos o efeito da introdução da LMP para fazer diminuir os homicídios de mulheres motivados por questões relacionadas ao gênero, conforme sublinhado pela equação (11). Para tanto, considere a regressão apontada na equação (9).

$$\log(\text{Taxa de homicídio})_{smt} = \beta_0 + \beta_1 \lambda_s + \beta_2 \lambda_{LMP} + \beta_3 (\lambda_s * \lambda_{LMP})_i + \beta_4 \text{Tx_alcohol}_{mt} + \beta_5 \text{Tx_armas}_{mt} + \sum_{m=1}^{M-1} \alpha_m \text{Micro}_m + \sum_{t=1}^{T-1} t + \varepsilon_{mt}. \quad (12)$$

Onde: λ_s é a variável indicadora do sexo para observação e assume 0 para sexo masculino e 1 para feminino; λ_{LMP} indica a existência da lei e assume 0 para ano ≤ 2006 e 1 para os demais; Tx_alcohol_{mt} é a taxa de óbitos por uso de álcool referente à m – ésima microrregião e ao ano t ; Tx_armas_{mt} , a taxa de suicídio por armas de fogo referente à microrregião e ao ano; Micro_m , a variável indicadora de microrregião da m – ésima observação; e Ano_t , a variável indicadora de ano.

O número de microrregiões varia no intervalo 1, 2, ..., M . Nessa aplicação, $M = 558$. O número de anos varia no intervalo 1, 2, ..., T . Nessa aplicação, $T = 12$, pois o ano varia de 2000 a 2011. As taxas de álcool e a *proxy* de armas de fogo também foram consideradas em escala logarítmica.

12. No apêndice, apontamos os gráficos com as evoluções das taxas de homicídio dentro da residência para todas as Unidades da Federação.

2048

TEXTO PARA DISCUSSÃO

Brasília, março de 2015

AVALIANDO A EFETIVIDADE DA LEI MARIA DA PENHA¹

Daniel Cerqueira²
Mariana Vieira Martins Matos³
Ana Paula Antunes Martins⁴
Jony Pinto Junior⁵

<https://repositorio.ipea.gov.br/handle/11058/3538>

Exemplo: Interação entre duas dummies (modelo de diferença em diferenças)

TABELA 3

Resumo da estimação dos modelos de diferenças para a variável *logaritmo da taxa de homicídios em residência*

Variável dependente: $\ln(\text{taxa de homicídios em residência})$				
	(5)	(6)	(7)	(8)
Sexo (β_1)	-1,179 ($<0,001$)***	-1,181 ($<0,001$)***	-1,182 ($<0,001$)***	-1,183 ($<0,001$)***
Vigência da lei (β_2)	0,339 ($<0,001$)***	0,340 ($<0,001$)***	NA -	NA -
Sexo*lei (β_3)	-0,099 ($<0,001$)***	-0,097 ($<0,001$)***	-0,096 ($<0,001$)***	-0,096 ($<0,001$)***
$\ln(\text{Taxa_alcohol})$	-	-	-	0,135 ($<0,001$)***
$\ln(\text{Taxa_armas})$	-	-	-	0,091 ($<0,001$)***
Efeito fixo de microrregião	Não	Sim	Sim	Sim
Efeito fixo de tempo	Não	Não	Sim	Sim
Prob. > F	($<0,001$)***	($<0,001$)***	($<0,001$)***	($<0,001$)***
R – ajustado	0,453	0,695	0,705	0,710
Número de observações	13.392	13.392	13.392	13.358

Fonte: SIM.

Elaboração: Diest/Ipea.

Obs.: * $<0,05$; ** $<0,01$; *** $<0,001$; NA – não definido por causa de singularidade. A taxa de armas é uma *proxy* para a prevalência de armas de fogo nas microrregiões construída a partir da proporção de suicídios por armas de fogo em relação ao total de suicídios. A taxa de álcool é uma *proxy* para consumo de bebida alcoólica nas microrregiões, construída pela soma de óbitos ocasionados pelo envenenamento por bebidas alcoólicas, relativizados pela população residente na localidade. Para contabilizar apenas os homicídios que ocorreram em residências, utilizamos o terceiro dígito da CID-10.

2048

TEXTO PARA DISCUSSÃO

Brasília, março de 2015

AVALIANDO A EFETIVIDADE DA LEI MARIA DA PENHA¹

Daniel Cerqueira²

Mariana Vieira Martins Matos³

Ana Paula Antunes Martins⁴

Jony Pinto Junior⁵

<https://repositorio.ipea.gov.br/handle/11058/3538>

Estima-se que a Lei Maria da Penha tenha reduzido de 9,6% a 9,9% os homicídios de mulheres em residência. O efeito é estatisticamente diferente de zero ao nível de 1% de significância em todas as especificações estimadas. Nota: a interpretação em variação percentual é devida ao uso, como variável dependente, do logaritmo de homicídios de mulheres em residência.

Diferença em diferenças: limitações

- Não elimina *confounders* que **variaram no período de tempo analisado e afetaram os grupos de maneira diferente**
 - Por exemplo, concomitantemente à instalação do incinerador pode ter havido uma obra de infraestrutura viária servindo apenas as casas próximas do incinerador; DiD não conseguirá distinguir o efeito do incinerador do efeito conjunto do (incinerador + obra viária)
- **Dito de outra forma, violação da *common trend assumption*** (um pressuposto que não pode ser provado) **compromete a validade** do estimador de efeitos da diferença em diferenças

*Em geral, qualquer fator que afete desproporcionalmente um dos dois grupos e o faça concomitantemente ao recebimento do programa pelo grupo de tratamento – e não seja considerado na regressão – tem o potencial de invalidar ou **enviesar a estimativa** de impacto do programa. O método diferença em diferenças pressupõe que nenhum fator desse tipo estará presente.*

Gertler et al. (2018, p. 156)

Avaliando a plausibilidade do pressuposto de tendência paralela (common trend assumption)

- 1 Compare a evolução de Y (**deltas temporais**) nos grupos de tratamento e comparação **antes da implementação** (se tiver dados para isso); expectativa é encontrar trajetória paralela
- 2 Aplique DinD em **mais de um grupo de comparação** (e.g., um terceiro bairro longe do incinerador); expectativa é encontrar efeitos similares àqueles obtidos a partir do primeiro grupo de comparação
- 3 Execute **teste “placebo”**
 - Com **grupo de tratamento “falso”**: aplique DinD contrastando grupo de comparação com grupo de tratamento “falso”, sabidamente não afetado pelo tratamento (e.g., um terceiro bairro, longe do incinerador); expectativa é não encontrar efeito
 - Com **Y “falso”**: aplique DinD considerando Y sabidamente não afetado pelo tratamento (e.g., participação no mercado de trabalho); expectativa é não encontrar efeito do incinerador

APÊNDICE:

Tipos de dados

Referências

- Babbie, Earl. **Métodos de Pesquisas de Survey**. 1999. Belo Horizonte: Editora UFMG. pp. 93-111 (cap. 4: “Tipos de desenhos de pesquisas”).
- Wooldridge, Jeffrey M. **Introdução à econometria**: uma abordagem moderna. São Paulo: Cengage Learning, 2008. pp. 1 -17 (cap. 1: “A natureza da econometria e dos dados econômicos”).



Tipos de dados

✓ Atributo do tipo de dado

Atributo

Tipo de dados: Terminologia

Babbie	Wooldridge	Múltiplas unidades observadas	Múltiplas observações de uma mesma unidade	Múltiplas observações de unidades diferentes de uma mesma população	Múltiplas observações de unidades diferentes de uma mesma coorte ("turma")
Survey interseccional	Dados de corte transversal	✓			
-	Série temporal		✓		
Estudo de painel*	Dados de painel ou longitudinais	✓	✓		
Estudo de tendências*	Dados de corte transversal agrupados**	✓		✓	
Estudo de coorte*	-	✓			✓

* Considerados "longitudinais" na terminologia de Babbie. ** Também conhecidos como "dados empilhados".

EXEMPLO DE DADOS DE CORTE TRANSVERSAL

– Conjunto de dados de corte transversal para o ano de 1976 de 526 trabalhadores (Wooldridge 2008):

Número da observação	Salário por hora	Anos de escolaridade	Anos de experiência no mercado de trabalho	Feminino	Estado civil (casado)
1	3,10	11	2	1	0
2	3,24	12	22	1	1
3	3,00	11	2	0	0
4	6,00	8	44	0	1
5	5,30	12	7	0	1
...
525	11,56	16	5	0	1
526	3,50	14	5	1	0

*Ana Mayra A. de Oliveira
Eneida M. M. Cerqueira
Josenira da Silva Souza
Antonio César de Oliveira*

*Departamento de Saúde,
Universidade Estadual de
Feira de Santana (UEFS),
Feira de Santana, BA.*

RESUMO

Sobrepeso e obesidade infantil apresentam prevalência elevada e caráter multifatorial. Com o objetivo de determinar a influência de fatores biológicos e ambientais no seu desenvolvimento, foi realizado corte transversal com 699 crianças, de 5 a 9 anos, da rede de ensino público e privado da zona urbana de Feira de Santana, BA. Sobrepeso e obesidade foram definidos como Índice de massa corpórea \geq aos percentis 85 e 95 para idade e sexo, respectivamente. A análise de entrevistas com os responsáveis determinou a influência dos fatores em questão. Observados como de significância estatística para o desenvolvimento de ambas as condições: nível elevado de escolaridade e renda familiar, ser unigênito, freqüentar escola privada, possuir eletrodomésticos e utilizar computador. O grupo étnico branco relacionou-se apenas ao sobrepeso. Repetência escolar e realização de atividade física sistemática mostraram associação inversa. O *odds ratio* mostrou associação preditiva independente com o fato de estudar em escola privada e ser unigênito. Em conclusão, há influência de fatores biológicos e ambientais no desenvolvimento de sobrepeso e obesidade infantil, confirmando o caráter multifatorial. (Arq Bras Endocrinol Metab 2003;47/2:144-150)

Descritores: Sobrepeso; Obesidade; Infantil; Epidemiologia

A amostra foi constituída por crianças matriculadas na rede de ensino previamente citada, no ano letivo de 2001, compreendidas na faixa etária de 5 a 9 anos, cujo N foi calculado utilizando-se a fórmula recomendada por Daniel para população infinita (9). A

EXEMPLO DE DADOS DE SÉRIES DE TEMPO

- Conjunto de dados de séries de tempo sobre efeitos do salário mínimo em Porto Rico (Wooldridge 2008):

Número da observação	Ano	Salário mínimo médio no ano	Taxa de trabalhadores cobertos pela lei de salário mínimo	Taxa de desemprego	Produto Nacional Bruto (PNB)
1	1950	0,20	20,1	15,4	878,7
2	1951	0,21	20,7	16,0	925,0
3	1952	0,23	22,6	14,8	1.015,9
...
37	1986	3,35	58,1	18,9	4.281,6
38	1987	3,35	58,2	16,8	4.496,7

Mortalidade por queda em idosos: estudo de série temporal

Mortality caused by accidental falls among the elderly: a time series analysis

Danielle Ledur Antes¹
Ione Jayce Ceola Schneider²
Eleonora d'Orsi³

Resumo

Introdução: O aumento da população idosa em todo o mundo tem colocado as quedas e suas consequências em evidência. **Objetivo:** Verificar a tendência de mortalidade por quedas no município de Florianópolis, no Estado de Santa Catarina e no Brasil. **Método:** Trata-se de um estudo de séries temporais, com dados do Sistema de Informação sobre Mortalidade, que estimou a variação da mortalidade por queda de acordo com a Classificação Internacional de Doenças (CID-10), capítulo XX, códigos W00 a W15 e W17 a W19, no período de 1997 a 2010, por meio da regressão linear segmentada (*joinpoint regression*). **Resultados:** Observou-se que, no período mais recente (2005/2008; 2002/2008; 2003/2008), houve aumento significativo das taxas de mortalidade específica por quedas nas três regiões investigadas, respectivamente, e que as taxas aumentaram com o avançar da idade. **Conclusão:** Ações de prevenção de quedas para a população idosa devem visar, principalmente, ao grupo etário de 80 anos ou mais, faixa etária na qual a queda resulta em maior taxa de óbitos.

Palavras-chave: Acidentes por quedas; Idoso; Estudos Ecológicos; Distribuição Temporal; Taxa de Mortalidade; Causas Externas.

EXEMPLO DE DADOS DE PAINEL OU LONGITUDINAIS

- Conjunto de dados de painel sobre crime e estatísticas relacionadas em 1986 e 1990 em 150 cidades nos Estados Unidos (Wooldridge 2008):

Número da observação	Cidade	Ano	Homicídios	População	Desemprego	Polícia
1	1	1986	5	350.000	8,7	440
2	1	1990	8	359.200	7,2	471
3	2	1986	2	64.300	5,4	75
4	2	1990	1	65.100	5,5	75
...
297	149	1986	10	260.700	9,6	286
298	149	1990	6	245.000	9,8	334
299	150	1986	25	543.000	4,3	520
300	150	1990	32	546.200	5,2	493

“Não balanceado” significa que para pelo menos uma das variáveis analisadas, para pelo menos um dos momentos de observação, os pesquisadores não dispunham de informação para uma ou mais das unidades da federação.

OS EFEITOS DA MÍDIA SOBRE O SUICÍDIO: UMA ANÁLISE EMPÍRICA PARA OS ESTADOS BRASILEIROS*

Paulo R. A. Loureiro**
Tito Belchior Moreira***
Adolfo Sachsida****

* Os autores agradecem o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

** Professor adjunto da Universidade de Brasília (UnB). E-mail: pauloloureiro@unb.br.

*** Doutor em economia e professor da Universidade Católica de Brasília (UCB). E-mail: tito@ucb.br.

**** Técnico de Planejamento e Pesquisa da Diretoria de Estudos e Políticas Macroeconômicas (Dimac) do Ipea. E-mail: sachsida@hotmail.com.

No quadro 1, pode-se observar a lista de variáveis dependentes e independentes que participam do estudo. O banco de dados utilizado é um **painel não balanceado**, cobrindo uma amostra de 27 estados brasileiros para trinta anos, 1980-2009, totalizando cerca de 810 observações.

QUADRO 1

Brasil: descrição das variáveis utilizadas nas equações de Minimos Quadrados Ordinários (MQOs), Efeitos Fixos (EFs) e Diferença em Diferenças (DPD) (1980-2009)

Variáveis dependentes
TXSUMSC1529 = Taxa de suicídio de pessoas jovens do sexo masculino entre 15 a 29 anos de idade.
TXSUFEM1529 = Taxa de suicídio de pessoas jovens do sexo feminino entre 15 a 29 anos de idade.
TXSUC1529 = Taxa de suicídio de pessoas jovens entre 15 a 29 anos de idade.
TXSUMSC = Taxa de suicídio de pessoas do sexo masculino.
TXSUFEM = Taxa de suicídio de pessoas do sexo feminino.
TXSUC = Taxa de suicídio.
Variáveis explicativas
Mídia = Índice de mídia (criado com base no número de rádio e televisão por domicílio – PNADs).
DESEMP = Taxa de desemprego.
Violência = Taxa de homicídios.
TXDIV = Taxa de divórcio.
TXOBFET = Taxa do número de óbitos fetais.
Desmídia = Taxa de desemprego vezes o índice de mídia.
Theil = Índice de coeficiente de Theil.
PO = Índice de pobreza.
RMDD = Renda média real domiciliar.
TXINTER = Taxa de internação de paciente com perturbação psíquica.

Elaboração dos autores.

EXEMPLO DE CORTES TRANSVERSAIS AGRUPADOS

- Conjunto de dados sobre os preços da moradia em 1993 e 1995 nos Estados Unidos (Wooldridge 2008):

Note: conjuntos diferentes de propriedades foram observados em 1993 e 1995.

Imposto sobre propriedade

Metragem

Numero da observação	Ano	Preço comercializado	Imppro	Arquad	Quantidade de dormitórios	Quantidade de banheiros
1	1993	85.500	42	1.600	3	2,0
2	1993	67.300	36	1.440	3	2,5
3	1993	134.000	38	2.000	4	2,5
...
250	1993	243.600	41	2.600	4	3,0
251	1995	65.000	16	1.250	2	1,0
252	1995	182.400	20	2.200	4	2,0
253	1995	97.500	15	1.540	3	2,0
...
520	1995	57.200	16	1.100	2	1,5



ERNESTO DOS SANTOS VASCONCELOS

ENSAIOS SOBRE AVALIAÇÃO DO ENSINO FUNDAMENTAL NO BRASIL

RESUMO

A presente tese reúne capítulos, cujo tema central é a trajetória do desempenho escolar na educação básica no Brasil. Tem por objetivo identificar os fatores que explicam o desempenho escolar dos alunos do ensino fundamental analisando se existe um acúmulo de conhecimento entre os ciclos escolares, se os fatores que explicam o desempenho escolar persistem entre diferentes gerações e se as desigualdades de desempenho escolar podem ser devidas à diferenças socioeconômicas dos alunos. Com esse propósito, nos três capítulos, utilizou-se a

[...]

efeito negativo e chances menores para níveis elevados de proficiência. No segundo capítulo, em uma abordagem de gerações sobrepostas e de modelagem econométrica com dados em corte transversal agrupados, avaliou-se a relação e o efeito dos fatores: gênero, raça, classe social e região geográfica, sobre desempenho escolar, para fins de comparação entre as gerações. Os resultados sugerem não haver mudanças significativamente positivas entre as gerações em relação a desempenho escolar (nível de proficiência em leitura), além de evidenciar uma desigualdade de ordem educacional, racial, regional e de classe. Nessa perspectiva, no último

A tabela 10 mostra o tamanho das subamostras calculadas a partir dos quantitativos do quadro 7. Totalizando 3.164.282 estudantes do Ensino Fundamental, divididos por gerações/ano e fases do ensino básico. As fases do ensino básico estão assim discriminadas: 4ª série/5º ano do período de 2003 até 2013 e 8ª série/9º ano do período de 2007 a 2017.

Tabela 10 – Demonstrativo de tamanho das subamostras por geração e ano.

GERAÇÃO	ANO							
	2003	2005	2007	2009	2011	2013	2015	2017
1	5.066	-	215.451					
2		4.640	-	259.772				
3			390.369	-	265.102			
4				482.505	-	266.983		
5					394.638	-	259.230	
6						312.825		307.701
Total					3.164.282			

Fonte: SAEB – 2003/2017 (Elaborada pelo autor).

As variáveis utilizadas neste trabalho de investigação foram selecionadas a partir dos questionários aplicados aos estudantes nas edições de avaliação do SAEB no período de 2003 a 2017. As características de interesse foram tratadas, passando por um processo de [...]

Estudo prospectivo:
estudo com rodadas
futuras de observação

Riscos associados ao processo de desmame
entre crianças nascidas em hospital universitário
de São Paulo, entre 1998 e 1999: estudo
de coorte prospectivo do primeiro ano de vida

Risks associated with the weaning process
in children born in a university hospital:
a prospective cohort in the first year of life,
São Paulo, 1998-1999

Milena Baptista Bueno¹
José Maria Pacheco de Souza²
Sonia Buongiorno de Souza¹
Suzana Maria Rebelo Sampaio da Paz¹
Suely Godoy Agostinho Gimeno³
Arnaldo Augusto Franco de Siqueira⁴

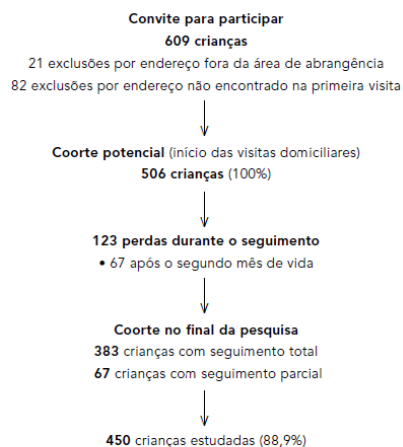
Resumo O objetivo do estudo foi descrever a prática do aleitamento materno de crianças nascidas em um hospital universitário em São Paulo, e identificar fatores associados com a duração do aleitamento materno e aleitamento materno exclusivo. Criou-se uma coorte potencial de 506 recém-nascidos, destas, foi possível utilizar informações de 450 que tinham dados sobre alimentação, pelo menos, até o segundo mês de vida. A mãe anotava diariamente, em um questionário de frequência alimentar, os alimentos consumidos pela criança. Técnicas de análise de sobrevivência (Kaplan-Meier e Cox) foram utilizadas para as análises. A mediana do aleitamento materno foi 205 dias e do aleitamento materno exclusivo 23 dias. Os fatores de importância na duração do aleitamento materno exclusivo foram idade da mãe (rh mais jovem/mais velha = 1,22; IC95% = 1,006-1,486) e escolaridade da mãe (rh fundamental/superior = 2,13; IC95% = 1,381-3,307 e rh médio/superior = 1,78; IC95% = 1,145-2,792), podendo estar relacionadas com o conhecimento dos benefícios do aleitamento materno exclusivo.

Palavras-chave Aleitamento Materno; Estudos de Coortes; Desmame

Cad. Saúde Pública, Rio de Janeiro, 19(5):1453-1460, set-out, 2003

Figura 1

Dinâmica da coorte.



No período de um ano, foram realizadas oito entrevistas com a mãe, sendo a primeira no hospital e as seguintes no domicílio, programadas para as idades de 15, 30, 60, 90, 180, 270 e 360 dias de vida da criança. Cada mãe recebeu, periodicamente, um questionário de frequência alimentar qualitativo, desenhado na forma de planilha de leitura ótica, para anotação diária da alimentação da criança. O primeiro questionário de frequência alimentar era deixado com a mãe na entrevista inicial no hospital, para que ela anotasse, diariamente, em sua casa, os alimentos e preparações consumidos pela criança. Nas visitas domiciliares, a entrevistadora revisava as anotações, recolhia os questionários e entregava um novo para o período seguinte; cada questionário permitia anotações de um período de 60 dias.

Por tratar-se de processo dinâmico, ocorreram 123 desistências e perdas (Figura 1). O principal motivo dessas perdas foi mudança de endereço (83,7%). Os resultados deste trabalho referem-se a 450 crianças, compreendendo as 383 que finalizaram o estudo após um ano e 67 que foram perdidas durante o seguimento mas que devolveram ao menos o primeiro questionário de frequência alimentar, com o registro dos primeiros 60 dias de vida.

<https://doi.org/10.1590/S0102-311X2003000500024>

[FIM DO APÊNDICE]:
Tipos de dados

Diferença em diferenças

Avaliação de Políticas Públicas B

04, 09 e 11 de junho de 2025

Leitura básica:

GERTLER, Paul J. et al. **Avaliação de impacto na prática**. Washington: Banco Interamericano de Desenvolvimento; Banco Mundial, 2018. [Capítulo 7: Diferença em diferenças.] Disponível em: <https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464808890.pdf>

Leitura complementar:

FACURE, Matheus. **Causal inference for the brave and true**. 2022. [Capítulo 13: Difference-in-differences.] Disponível em: <https://matheusfacure.github.io/python-causality-handbook/landing-page.html>