

Experimentos e quase experimentos

Avaliação de Políticas Públicas A (DCP131)

04 e 09 de junho de 2025

Nota: este material é bem mais completo do que as nossas discussões em sala de aula.

Leitura básica:

JANNUZZI, Paulo de Martino. **Monitoramento e avaliação de programas sociais**. Campinas, SP: Editora Alínea, 2016. pp. 89-100.

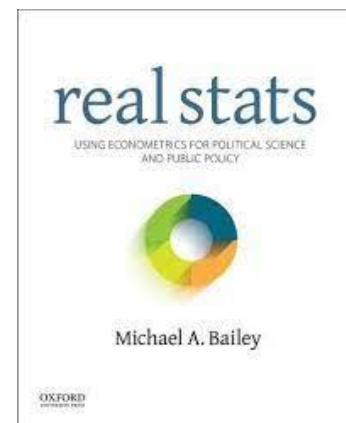
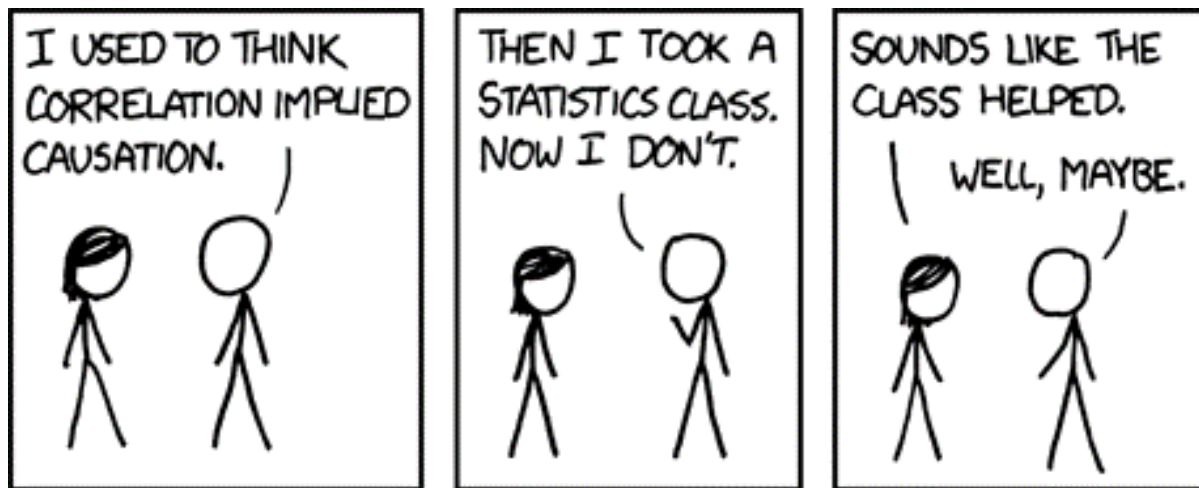
RAMOS, Marília Patta. Aspectos conceituais e metodológicos da avaliação de políticas e programas sociais. **Planejamento e Políticas Públicas**, n. 32, jan.-jun./2009, p. 95-114. Disponível em: <https://www.ipea.gov.br/ppp/index.php/PPP/article/view/11>

Leitura complementar:

GERTLER, Paul J. et al. **Avaliação de impacto na prática**. Washington: Banco Interamericano de Desenvolvimento; Banco Mundial, 2018. [Capítulo 3: Inferência causal e contrafactuais. Capítulo 4: Seleção aleatória. Capítulos 5-8, sobre abordagens quase experimentais, devem ser lidos conforme o interesse do/a discente.] Disponível em:

<https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464808890.pdf>

experimentos



Fonte: Bailey (2016, p. 3). Original em: <https://xkcd.com/552/>

OK. We get it. Correlação/ associação não implica causalção.

Mas, então, o que implica causalção? Resposta: Atendimento aos três requisitos para inferência causal: ANTERIORIDADE, ASSOCIAÇÃO E NÃO ESPURIEDADE.

Se conseguirmos identificar alguma variação da suposta causa (X) que seja independente de variáveis omitidas (i.e., variação exógena; neste caso, independente de Z) e ainda assim essa suposta causa se correlacionar/ associar com o efeito em questão, provavelmente essa correlação/ associação é causal (digo “provavelmente” pois a correlação/ associação ainda poderia ser fortuita).

Desenho experimental em avaliação

Indicação:



- Mensurar o efeito (resultado ou impacto) do programa sobre seus beneficiários ou comunidade atendida

Características:

- Inspirado no formato clássico do experimento nas **ciências naturais**
- Para o desenho **experimental** de pesquisa, a **atribuição ao tratamento** (i.e., definição de quem vai ter acesso ao programa) **deve ser aleatória**

Sobre o **desenho experimental** e a **comparabilidade presumida entre grupos formados aleatoriamente** a partir de uma **mesma população**:

*Como os **dois grupos** [de tratamento e de controle] foram expostos às mesmas condições ‘de laboratório’ e são **muito parecidos** [por conta da presunção de comparabilidade baseada na atribuição aleatória ao tratamento], a **eventual diferença** entre eles seria consequência do fato de um deles ter tido **acesso ao programa** (enquanto o outro não). Se a eventual diferença das medidas é significativa em termos estatísticos, estaria comprovada [melhor dizer, “apoiada pela evidência”] a tese de que o programa teria algum impacto sobre seus beneficiários.*

Jannuzzi (2016, p. 97)

Desenho experimental e atribuição aleatória ao tratamento

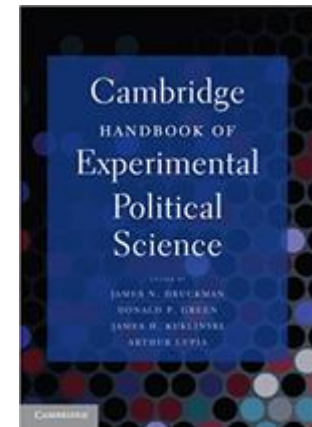
Um bom jeito de se ter variação exógena é criar tal variação

- Com base na **pergunta** de pesquisa, identifique uma **população** relevante e obtenha uma **amostra (aleatória ou não)**
- Use algum **procedimento aleatório para atribuir as unidades observadas a um de dois grupos**:
 - Grupo de **tratamento**: recebe a intervenção (e.g., política pública)
 - Grupo de **controle**: não recebe a intervenção

Amostragem vs. atribuição aleatória ao tratamento

*Um objetivo comum na amostragem aleatória é selecionar participantes de uma população mais ampla de modo a oferecer a cada potencial participante a mesma probabilidade de ser recrutado para o estudo. A atribuição aleatória é diferente. [...] O requisito principal é que um procedimento aleatório, como um sorteio com cara ou coroa, determine se os participantes do estudo receberão um certo tratamento. **Assim como um experimento não requer amostragem aleatória, um estudo com amostra aleatória nem sempre é um experimento.***

(Druckman et al., 2011, p. 17-18; tradução nossa)



Desenho experimental e atribuição aleatória ao tratamento

Um bom jeito de se ter variação exógena é criar tal variação

- **Após o tratamento, compare** o comportamento dos grupos de tratamento e controle no tocante ao **resultado de interesse**
- Se o grupo de tratamento **diferir substancialmente** do grupo de controle, **acreditamos que o tratamento teve um efeito**; caso contrário, estamos inclinados a acreditar que o tratamento não teve efeito
- Sabe-se com certeza que **nenhum dos condicionantes sistemáticos de Y afetaram a atribuição** ao tratamento
- A atribuição aleatória ao tratamento enseja uma **presunção de comparabilidade** entre grupos no momento logo anterior à atribuição ao tratamento.
- Em outras palavras, não há razão para acreditar que possíveis variáveis omitidas estejam associadas com o grau de exposição ao tratamento (já que este foi determinado aleatoriamente)
- Isso vale para as **variáveis (condicionantes) observáveis e não observáveis**. O grande valor da atribuição aleatória ao tratamento é a presunção de comparabilidade entre grupos nas dimensões não observáveis

Tipos de experimentos

É comum classificar experimentos de acordo com o ambiente (*setting*) onde eles ocorrem:

Experimento de laboratório

- Intervenção ocorre em contexto criado e controlado pelo(a) pesquisador(a)

Experimento de campo

- Intervenção ocorre num contexto “natural” (*naturally occurring*)

Experimento de *survey* (lista, vinheta)

- Intervenção ocorre no âmbito de uma pesquisa de opinião

Disclaimer:

Esta classificação não é perfeita, pois diferentes estratégias podem ser combinadas em um único experimento – e.g., lista após laboratório (Druckman, 2011, p. 17; título: “Experiments: an introduction to core concepts”, no livro “Cambridge Handbook of Experimental Political Science”).

Exemplo: experimento de campo

Attanazio et al. (2017, p. ii;
tradução nossa)



Resumo traduzido do relatório “Impact of free availability of childcare on labour supply and child development in Brazil”, datado de 2017 e disponível em: <http://www.3ieimpact.org/evidence-hub/publications/impact-evaluations/impact-free-availability-public-childcare-labour>.

Resumo

Neste relatório, examinamos os impactos do acesso a creche gratuita em período integral sobre o desenvolvimento da criança, a oferta de trabalho dos membros do domicílio e os ambientes domésticos. Usamos dados da cidade do Rio de Janeiro, Brasil, primariamente de domicílios localizados em favelas.

O programa de creches públicas do Rio de Janeiro é um programa integrado de desenvolvimento da primeira infância para crianças de 0-3 anos vivendo em bairros de baixa renda. O programa consiste de uma variedade de intervenções baseadas nas creches, incluindo oferta de creche em período integral, serviços de saúde, alimentação e provisão de brinquedos pedagógicos e materiais escolares para as crianças.

Espera-se que o acesso a essas creches bem equipadas e adequadamente administradas, ao propiciar um ambiente acolhedor e estimulante para crianças carentes, impulse a acumulação de capital humano pelos pobres. Porque os pais são encorajados a participar da vida de seus filhos na creche, e até recebem da equipe da creche informação sobre como cuidar das crianças, também poderíamos esperar uma mudança nos comportamentos parentais. Adicionalmente, é esperado que mães e outros membros do domicílio encarregados de cuidar das crianças passem a dispor do tempo para buscar emprego e aumentar sua renda, melhorando assim o bem-estar dos domicílios. Este seria outro mecanismo através do qual acesso à creche poderia levar a um melhor desenvolvimento da criança.

[Continua]

Exemplo: experimento de campo



Attanasio et al. (2017, p. ii; tradução nossa)

<http://www.3ieimpact.org/evidence-hub/publications/impact-evaluations/impact-free-availability-public-childcare-labour>

[Continuação]

Este estudo identifica a relação causal entre acesso a creche e resultados relacionados ao desenvolvimento infantil e participação maternal no mercado de trabalho, tirando proveito de um sorteio realizado pelo governo municipal do Rio de Janeiro em 2007. Todo ano, o governo do Rio oferece aproximadamente 10.000 novas vagas em creches para crianças de 0-3 anos. In 2007, como em anos anteriores, a demanda por essas vagas foi largamente superior à oferta. Mais de 25.000 famílias inscreveram-se para 10.000 novas vagas. Para garantir igualdade de oportunidade, um sorteio foi usado para atribuir as vagas aos interessados elegíveis (aproximadamente 24.000 dos 25.000). Novas crianças beneficiárias começaram a receber serviços em fevereiro de 2008.

Construímos um rico banco de dados que inclui várias medições das crianças e um questionário muito detalhado sobre o domicílio.

Encontramos impactos bastante fortes da frequência à creche pública sobre a altura e o peso das crianças, diversos anos após terem deixado as creches. Também, achamos evidência de que ter acesso à creche produz impactos substantivos e duradouros na renda domiciliar, gastos com bens duráveis e investimentos nas crianças – tanto em tempo como em bens.

Particularmente, há um aumento da oferta de trabalho e renda dos avós (principalmente das avós) residindo no mesmo domicílio que uma criança frequentando a creche. Notadamente, esses impactos estão presentes mesmo quatro anos e meio após a aleatorização inicial, num momento em que pouquíssimas crianças de nossa amostra ainda frequentam creches.

Autoria: Orazio Attanasio (Institute for Fiscal Studies); Ricardo Paes de Barros (Insper); Pedro Carneiro (University College London); David Evans (World Bank); Lycia Lima (Fundação Getulio Vargas, São Paulo); Pedro Olinto (World Bank); Norbert Schady (Inter-American Development Bank).

experimentos vs.
quase experimentos

Delineamentos experimentais nem sempre são aplicáveis em casos concretos de políticas públicas

Os grupos de tratamento e comparação acabam sendo definidos em função de atrasos ou do calendário de implantação de um programa social ou, ainda, pela cobertura deficiente de público-alvo deste. Jannuzzi (2016, p. 97)

São desafios do desenho experimental:

- Garantir **viabilidade**: há variáveis que não podem ser manipuladas (e.g., idade, cor/raça); também, não é possível manipular **eventos que ocorreram no passado**.
 - Conduzir **recrutamento** para o experimento (obter anuência à participação).
 - Vencer **problemas éticos** na escolha e justificação de quem será beneficiário e quem vai ficar de fora de um programa (supostamente benéfico).
 - Lidar com eventos que potencialmente geram **desbalanceamento dos grupos: noncompliance** com a atribuição ao tratamento (unidades designadas para o grupo de tratamento não o recebem, ou vice-versa); e **evasão** dos casos analisados.
 - Avaliar em que medida as **circunstâncias “artificiais”** que rodeiam o experimento conspiram contra a generalização dos resultados (validade externa do desenho).
- Garantir que os efeitos medidos sejam apenas do programa em foco, em meio a um **contexto crescente de intervenções sociais concomitantes**.
 - Escolher a **melhor medida para captar a dimensão supostamente impactada**:
 - O que se supõe como dimensão impactada guarda, pelo desenho lógico do programa, vinculação estreita com as ações e atividades deste?
 - O impacto deve ser medido ao nível dos beneficiários e/ ou da comunidade a que pertencem?

Estes desafios aplicam-se também a outros desenhos de avaliação.

Experimentos vs. Quase experimentos

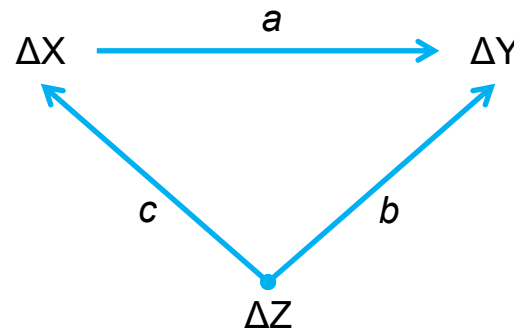
Estimação de efeitos das variáveis incluídas associadas com omitidas **confundirá a influência das duas**, resultando em viés – i.e., tendência a **superestimar ou subestimar** a importância da variável incluída na produção da variável dependente.

Diante da **atribuição aleatória ao tratamento** (típica de desenhos experimentais), presume-se que o nível da variável omitida é semelhante nos grupos tratado e não tratado, portanto neutralizando eventual associação entre variáveis incluídas e omitidas.

Desenhos **quase experimentais** não contam com atribuição aleatória ao tratamento, mas aplicam outras estratégias para aprofundar a comparabilidade entre grupos em dados observacionais (i.e., não experimentais).

Foco daqui em diante.

Associação espúria: use experimentos e quase experimentos para evitar essa cilada



Associação a é espúria (i.e., não é uma medida válida da associação causal entre X e Y) porque reporta uma combinação de dois efeitos: efeito de X sobre Y e efeito de Z sobre Y .

- Para que a não fosse espúria, para cada potencial fonte de espuriedade Z , as associações b e c não poderiam existir concomitante (ou seja, pelo menos uma delas teria de ser nula)
- A **atribuição aleatória ao tratamento** (desenho experimental) e alguns desenhos quase experimentais (como o **regression discontinuity design – RDD** e **desenhos com pareamento**) **eliminam variações de Z** ; se Z não varia, espera-se que não se associe com X ou Y na amostra
- O pareamento é uma solução inferior ao RDD, e o RDD é inferior à atribuição aleatória ao tratamento: para que seja internamente válido, o pareamento deve igualar todos os possíveis Z em cada par, e pode render poucos pares viáveis; o RDD tem pressupostos mais exigentes que a atribuição aleatória ao tratamento, e está também exposto ao problema de amostras pequenas

Ramos (2009): exemplo de experimento com pareamento para reforçar a comparabilidade (balance) entre grupos

Aspectos conceituais e metodológicos da avaliação de políticas e programas sociais

107

Um conceito importante dentro desta técnica é a ideia de um suporte comum, que se constitui na região onde o equilíbrio dos escores de propensão, entre os grupos de controle e de tratamento, se apresenta. Esta região de suporte comum é obtida a partir do descarte daqueles casos que estiverem muito abaixo ou muito acima da média dos escores. Entretanto, este procedimento de descarte dos casos que estiverem fora da região de suporte comum pode causar alguns problemas, como: bons pares podem ser perdidos próximo aos limites da região de suporte comum e a exclusão de casos pode mudar os parâmetros a serem estimados (redução do tamanho da amostra).

Pelo fato de esta técnica se basear nas características observáveis, ela apresenta certa limitação no que diz respeito às características não observadas que podem estar na base de processos de seletividade. Contudo, a literatura sobre esta técnica (HECKMAN; ISHIMURA; TODD, 1998) enfatiza que uma estratégia de avaliação que consegue controlar rigorosamente pelas características observadas e que consegue informações sobre o grupo de controle e o de tratamento de forma semelhante pode permitir estimadores confiáveis do impacto de um programa.

Como exemplo do uso do escore de propensão para pareamento, podemos citar a pesquisa de avaliação do Programa Nacional de Suporte ao Trabalho, desenvolvido em meados dos anos 1970 nos Estados Unidos para oferecer experiência de trabalho a trabalhadores que estivessem enfrentando problemas econômicos e sociais. Aqueles aleatoriamente designados para participar do programa tiveram treinamento em restaurantes e na construção civil. As informações relacionadas com o período anterior ao programa (ganhos, educação, estado civil e raça) foram medidas através de enquetes e de arquivos da administração pública. Em que pese os candidatos terem sido aleatoriamente selecionados, o fato de a avaliação ter sido feita num período de dois anos fez com que aqueles que entraram cedo no programa pudessem ter características diferentes daqueles que entraram mais tarde. Além disso não havia informação sobre se os participantes e não-participantes haviam estado no mercado de trabalho antes do programa, e assim a variável de interesse, a renda em 1978, poderia ser afetada caso tivessem sido empregados antes do programa. Então, havia a necessidade do procedimento de pareamento, a partir das características observadas antes do tratamento.

O que exatamente este procedimento faz é calcular a propensão (probabilidade) a participar baseado em características observáveis, após o que é feito o cálculo do valor estimado (y) na equação para todos os sujeitos (participantes e não-participantes), isto é, com base nos coeficientes da regressão logística somam-se as características de cada indivíduo (multiplicando-se pelo valor dos coeficientes da regressão) e cada indivíduo terá um valor de y diferente (um escore). Ressaltamos que para cada indivíduo no grupo participante do programa o pesquisador quer encontrar o indivíduo no grupo de controle (não-participante) que apresenta o

Aspectos conceituais e metodológicos da avaliação de políticas e programas sociais

109

Observando a tabela 5 podemos ver as diferenças nas variáveis observadas entre o grupo de tratamento e o de controle sem o pareamento e com o pareamento. Podemos observar que em todas as variáveis observadas, após o pareamento, não existe diferença estatisticamente significativa entre o grupo de controle e o de tratamento.

TABELA 5
Características antes do programa, grupo de controle e de tratamento

| Variáveis | Amostra | Médias | | | | t-test | |
|-------------|--------------|---------|----------|--------|---------|--------|-------|
| | | Tratado | Controle | % viés | [Viés] | t | p> t |
| T | Não-pareados | 1 | 0 | . | . | . | . |
| | Pareados | 1 | 0 | . | . | . | . |
| Idade | Não-pareados | 25.816 | 25.054 | 10.7 | | 1.12 | 0.265 |
| | Pareados | 25.809 | 25.75 | 0.8 | 92.3 | 0.08 | 0.934 |
| Educ | Não-pareados | 10.346 | 10.088 | 14.1 | | 1.50 | 0.135 |
| | Pareados | 10.328 | 10.392 | -3.5 | 75.1 | -0.34 | 0.731 |
| Negro | Não-pareados | 0.84324 | 0.82692 | 4.4 | | 0.45 | 0.649 |
| | Pareados | 0.85246 | 0.86475 | -3.3 | 24.7 | -0.36 | 0.720 |
| hisp | Não-pareados | 0.05946 | 0.10769 | -17.5 | | -1.78 | 0.076 |
| | Pareados | 0.06011 | 0.05464 | 2.0 | 88.7 | 0.24 | 0.811 |
| Casado | Não-pareados | 0.18919 | 0.15385 | 9.4 | | 0.98 | 0.327 |
| | Pareados | 0.19126 | 0.18852 | 0.7 | 92.3 | 0.07 | 0.944 |
| s/curso sup | Não-pareados | 0.70811 | 0.83462 | -30.4 | | -3.22 | 0.001 |
| | Pareados | 0.71585 | 0.70628 | 2.3 | 92.4 | 0.21 | 0.830 |
| Renda 1974 | Não-pareados | 2095.6 | 2107 | -0.2 | | -0.02 | 0.982 |
| | Pareados | 2081.5 | 1899.1 | 3.4 | -1493.1 | 0.39 | 0.697 |
| u74 | Não-pareados | 0.70811 | 0.75 | -9.4 | | -0.98 | 0.326 |
| | Pareados | 0.71038 | 0.73361 | -5.2 | 44.6 | -0.53 | 0.599 |
| Renda 1975 | Não-pareados | 1532.1 | 1266.9 | 8.4 | | 0.87 | 0.382 |
| | Pareados | 1488.3 | 1542.7 | -1.7 | 79.5 | -0.17 | 0.867 |
| u75 | Não-pareados | 0.6 | 0.68462 | -17.7 | | -1.85 | 0.065 |
| | Pareados | 0.60656 | 0.63525 | -6.0 | 66.1 | -0.60 | 0.548 |

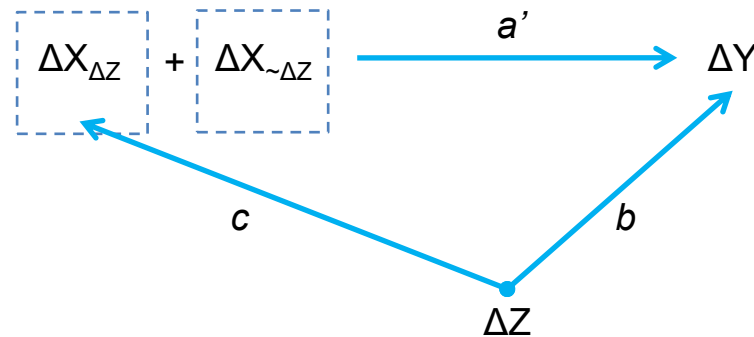
Fonte: Tabela obtida pelo teste de propensão para pareamento realizado no software Stata versão 9.1. Dados oriundos do Programa Nacional de Suporte ao Trabalho dos Estados Unidos, fornecidos pelo professor Doutor Chandler Stolp da Universidade do Texas.

6 VARIÁVEIS INSTRUMENTAIS

É sempre desejável procurar situações, quando trabalhamos com análise de regressão múltipla, nas quais seja razoável supor que as variáveis omitidas (aquelas não incluídas no modelo) não sejam correlacionadas com as variáveis de interesse. Tais situações podem surgir se o pesquisador supõe aleatoriedade ou em alguma situação próxima desta. Na economia, estas situações não são facilmente percebidas, o que

Associação espúria: use experimentos e quase experimentos para evitar essa cilada

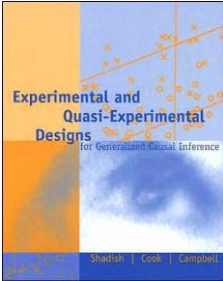
a é associação espúria; a' pode não ser espúria



Associação a' pode não ser espúria.

- Se associação a' (entre $\Delta X_{\sim \Delta Z}$ e ΔY) for não nula para todas as potenciais fontes de espuriedade Z , então a proposição de que variações em X causam variações em Y é válida
- Alguns desenhos **quase experimentais** (como variável instrumental, diferença em diferenças e efeitos fixos) e o **controle estatístico** (i.e., regressão múltipla em que se estima o efeito de X sobre Y controlando-se por Z) são caminhos para se **identificar a associação a'**
- Desenhos quase experimentais são superiores ao mero controle estatístico pois controlam por um agregado de variáveis Z não observáveis; porém, todas essas soluções são inferiores à atribuição aleatória ao tratamento

Elementos mínimos para caracterização de diferentes desenhos de pesquisa

|  Desenho de pesquisa | Atribuição ao tratamento | | | Comparação entre unidades | | Comparação no tempo |
|---|---|---|---|---|---|---|
| | Aleatória, sob comando do(a) pesquisador(a) | Aleatória, fora do comando do(a) pesquisador(a) | "Como se" aleatória, fora do comando do(a) pesquisador(a) | Grupo de tratamento vs. controle (atribuição aleatória) | Grupo de tratamento vs. comparação (atribuição não aleatória), ou outra comparação entre unidades com níveis diferentes de tratamento | Variável de resultado pré- e pós-tratamento |
| Experimento | X | | | X | | |

Notas: Todos os desenhos pressupõem que pelo menos parte da amostra foi exposta a algum tipo de tratamento. Atribuição “como se” aleatória pode ou não ser manipulável. Exemplos de atribuição manipulável: sorteio para Minha Casa Minha Vida e *charter schools*; mudança inesperada de política, com critérios rígidos de elegibilidade e sem que interessados tenham opção de “escolher” participar do grupo de tratamento ou comparação (e.g., influxo cubano em Miami). Exemplos de atribuição não manipulável: terremoto, ataque terrorista.

Fonte: Baseado em Shadish, Cook e Campbell (2002, p. 1-18, 156-161).

Elementos mínimos para caracterização de diferentes desenhos de pesquisa

| Desenho de pesquisa | Atribuição ao tratamento | | | Comparação entre unidades | Comparação no tempo |
|----------------------------------|---|---|---|---|---|
| | Aleatória, sob comando do(a) pesquisador(a) | Aleatória, fora do comando do(a) pesquisador(a) | "Como se" aleatória, fora do comando do(a) pesquisador(a) | Grupo de tratamento vs. controle (atribuição aleatória) | Grupo de tratamento vs. comparação (atribuição não aleatória), ou outra comparação entre unidades com níveis diferentes de tratamento |
| Experimento | X | Como nos exemplos dos currículos, Halloween e cotas na UnB. | | X | Variável de resultado pré- e pós-tratamento |
| Experimento natural "verdadeiro" | | X | | X | |
| | | Como nos exemplos das creches, vermífugos e transferência de renda. | | | |

Notas: Todos os desenhos pressupõem que pelo menos parte da amostra foi exposta a algum tipo de tratamento. Atribuição "como se" aleatória pode ou não ser manipulável. Exemplos de atribuição manipulável: sorteio para Minha Casa Minha Vida e *charter schools*; mudança inesperada de política, com critérios rígidos de elegibilidade e sem que interessados tenham opção de "escolher" participar do grupo de tratamento ou comparação (e.g., influxo cubano em Miami). Exemplos de atribuição não manipulável: terremoto, ataque terrorista.

Fonte: Baseado em Shadish, Cook e Campbell (2002, p. 1-18, 156-161).

Reflexões sobre o desenho experimental e experimentos naturais

O atributo comum a todos os experimentos é o **controle do tratamento (embora o controle possa assumir muitas formas diferentes)**. Assim, Mosteller (1990, p. 225) escreve: “Em um experimento, o **investigador controla a aplicação do tratamento**”; e Yaremko, Harari, Harrison e Lynn (1986, p. 72) escrevem, “uma ou mais variáveis independentes são **manipuladas** para observar seus efeitos em uma ou mais variáveis dependentes”. No entanto, ao longo do tempo, muitos subtipos experimentais diferentes se desenvolveram em resposta às necessidades e histórias de diferentes ciências (Winston, 1990; Winston & Blais, 1990).

(Shadish, Cook, Campbell, 2002, p. 12; tradução nossa)

Esses **experimentos naturais** incluem casos em que **loterias aleatórias** determinam quais homens são convocados para o serviço militar (por exemplo, Angrist, 1990) [...] O termo **experimento natural** às vezes é definido de forma mais ampla para incluir eventos que acontecem a algumas pessoas e não a outras, mas a casualidade não é aleatória. A adequação desta definição mais ampla é discutível; no entanto, **quando o mecanismo que determina se as pessoas estão expostas a um estímulo potencialmente relevante é suficientemente aleatório**, esses experimentos naturais podem fornecer aos estudiosos uma oportunidade de conduzir pesquisas sobre tópicos que normalmente estariam fora do alcance de um experimentador.

(Druckman et al., 2011, p. 22; tradução nossa)

Elementos mínimos para caracterização de diferentes desenhos de pesquisa

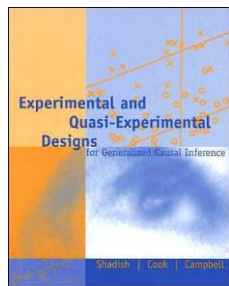
| Desenho de pesquisa | Atribuição ao tratamento | | | Comparação entre unidades | | Comparação no tempo |
|----------------------------------|---|---|---|---|---|--|
| | Aleatória, sob comando do(a) pesquisador(a) | Aleatória, fora do comando do(a) pesquisador(a) | "Como se" aleatória, fora do comando do(a) pesquisador(a) | Grupo de tratamento vs. controle (atribuição aleatória) | Grupo de tratamento vs. comparação (atribuição não aleatória), ou outra comparação entre unidades com níveis diferentes de tratamento | Variável de resultado pré- e pós-tratamento |
| Experimento | X | | | X | | |
| Experimento natural "verdadeiro" | | X | | X | | |
| Experimento natural "como se" | | | X | | | Pelo menos um destes elementos está presente |

Como numa mudança de política inesperada (vide Karruz, 2018, 2021 sobre cotas na UFMG).

Notas: Todos os desenhos pressupõem que pelo menos parte da amostra foi exposta a algum tipo de tratamento. Atribuição "como se" aleatória pode ou não ser manipulável. Exemplos de atribuição manipulável: sorteio para Minha Casa Minha Vida e *charter schools*; mudança inesperada de política, com critérios rígidos de elegibilidade e sem que interessados tenham opção de "escolher" participar do grupo de tratamento ou comparação (e.g., influxo cubano em Miami). Exemplos de atribuição não manipulável: terremoto, ataque terrorista.

Fonte: Baseado em Shadish, Cook e Campbell (2002, p. 1-18, 156-161).

Elementos mínimos para caracterização de diferentes desenhos de pesquisa



Desenho de pesquisa

| | Atribuição ao tratamento | | | Comparação entre unidades | | Comparação no tempo |
|----------------------------------|---|---|---|---|---|--|
| | Aleatória, sob comando do(a) pesquisador(a) | Aleatória, fora do comando do(a) pesquisador(a) | "Como se" aleatória, fora do comando do(a) pesquisador(a) | Grupo de tratamento vs. controle (atribuição aleatória) | Grupo de tratamento vs. comparação (atribuição não aleatória), ou outra comparação entre unidades com níveis diferentes de tratamento | Variável de resultado pré- e pós-tratamento |
| Experimento | X | | | X | | |
| Experimento natural "verdadeiro" | | X | | X | | |
| Experimento natural "como se" | | | X | | | Pelo menos um destes elementos está presente |
| Quase-experimento | | | | | | Pelo menos um destes elementos está presente |

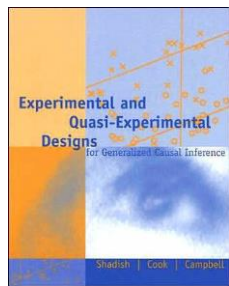
Não experimento

Notas: Todos os desenhos pressupõem que pelo menos parte da amostra foi exposta a algum tipo de tratamento. Atribuição "como se" aleatória pode ou não ser manipulável. Exemplos de atribuição manipulável: sorteio para Minha Casa Minha Vida e *charter schools*; mudança inesperada de política, com critérios rígidos de elegibilidade e sem que interessados tenham opção de "escolher" participar do grupo de tratamento ou comparação (e.g., influxo cubano em Miami). Exemplos de atribuição não manipulável: terremoto, ataque terrorista.

Fonte: Baseado em Shadish, Cook e Campbell (2002, p. 1-18, 156-161).

Elementos mínimos para caracterização de diferentes desenhos de pesquisa

abc Abordagem típica



Desenho de pesquisa

| | Atribuição ao tratamento | | | Comparação entre unidades | | Comparação no tempo |
|----------------------------------|---|---|---|---|---|---|
| | Aleatória, sob comando do(a) pesquisador(a) | Aleatória, fora do comando do(a) pesquisador(a) | "Como se" aleatória, fora do comando do(a) pesquisador(a) | Grupo de tratamento vs. controle (atribuição aleatória) | Grupo de tratamento vs. comparação (atribuição não aleatória), ou outra comparação entre unidades com níveis diferentes de tratamento | Variável de resultado pré- e pós-tratamento |
| Experimento | X | | | Diferença de médias | | |
| Experimento natural "verdadeiro" | | X | | | | |
| Experimento natural "como se" | | | X | | Especificações quase experimentais (e.g., RDD, VI, <i>interrupted time series</i> , DiD, controles sintéticos, VI, <i>matching</i> , efeitos fixos) | |
| Quase-experimento | | | | | | |

Não experimento

Notas: Todos os desenhos pressupõem que pelo menos parte da amostra foi exposta a algum tipo de tratamento. Atribuição "como se" aleatória pode ou não ser manipulável. Exemplos de atribuição manipulável: sorteio para Minha Casa Minha Vida e *charter schools*; mudança inesperada de política, com critérios rígidos de elegibilidade e sem que interessados tenham opção de "escolher" participar do grupo de tratamento ou comparação (e.g., influxo cubano em Miami). Exemplos de atribuição não manipulável: terremoto, ataque terrorista.

Fonte: Baseado em Shadish, Cook e Campbell (2002, p. 1-18, 156-161).

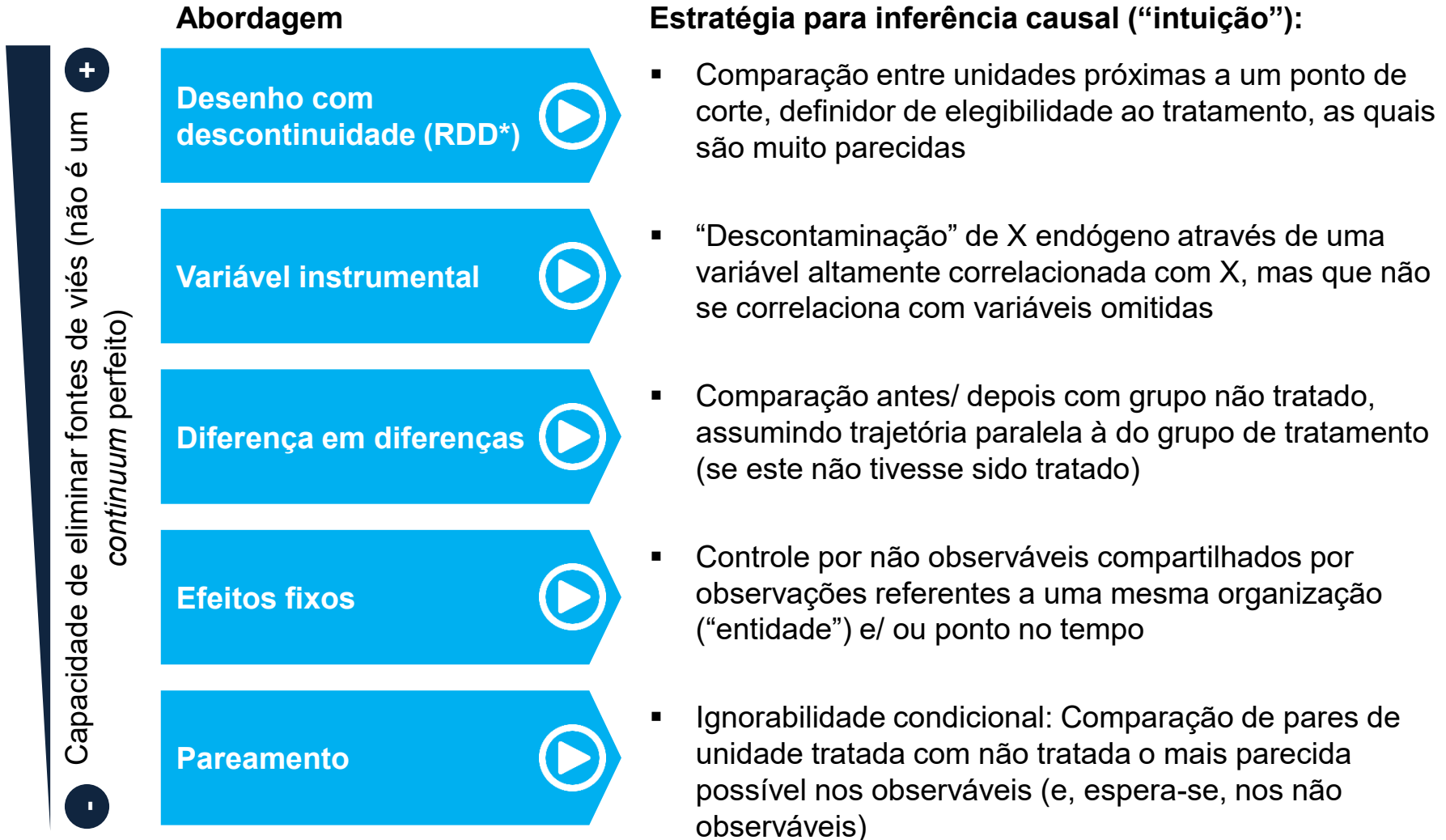
Atenção!

- Os **pressupostos** para produção de estimativas internamente válidas de um efeito causal vão ficando **mais demandantes (fortes) à medida que nos afastamos do contexto experimental**
- **Limitações** das abordagens quase experimentais estão **intimamente ligadas a esses pressupostos**:
 - Quanto **mais exigente o pressuposto**, mais improvável será o seu atendimento nos casos em tela
 - Quanto **mais restrito o foco da abordagem**, mais específica será sua utilidade (e.g., avaliação do efeito do Programa Bolsa Família via RDD, considerando os beneficiários próximos ao ponto de corte de elegibilidade pela idade do filho mais novo)
- Pode haver **combinação de abordagens** (e.g., DiD com pareamento; RDD com variável instrumental)

quase experimentos

De quais abordagens trataremos?

Experimentos naturais “como se” e quase experimentos frequentemente recorrem a uma ou mais destas abordagens:



* Regression Discontinuity Design.

De quais abordagens trataremos?

Abordagem

**Desenho com
descontinuidade (RDD*)**

Variável instrumental



Diferença em diferenças



Efeitos fixos



Pareamento



* Regression Discontinuity Design.



A natureza não dá saltos. [Natura non facit saltus.]

Carl Linnæus

[Carlos Lineu, zoólogo sueco considerado o
“pai” da taxonomia moderna (1707-1778)]

Regressão com descontinuidade: requisitos e intuição

Requisitos

- Unidades tratadas/ não tratadas podem ser **ordenadas** ao longo de uma **dimensão quantificável**
- Essa dimensão pode ser usada para computar um **índice contínuo (running variable)**
- A **elegibilidade** ao tratamento é definida por um **ponto de corte (cutoff)** na running variable
- O ponto de corte deve ser **determinado exogenamente**: unidades não podem ser capazes de definir onde será o ponto de corte da intervenção, nem de mudar seu status de forma a se tornarem elegíveis (ou não elegíveis)

Intuição

- As unidades logo acima do ponto de corte são muito parecidas com aquelas logo abaixo dele; na **vizinhança do cutoff**, assume-se que a **localização do cutoff** é “como se” aleatória
- **Comparamos resultados logo acima e logo abaixo do ponto de corte**

Regressão com descontinuidade: exemplo

Os alunos americanos do ensino médio recebem o **National Merit Scholarship Awards** com base nas pontuações do **PSAT**, um teste feito pela maioria dos alunos do penúltimo ano do ensino médio, especialmente aqueles que farão o SAT posteriormente. A **questão** que motivou as primeiras discussões sobre RD [regressão com descontinuidade] é **se os alunos que ganham esses prêmios têm mais probabilidade de terminar a faculdade** (Thistlewaithe e Campbell, 1960; Campbell, 1969). A **sharp RD** compara as **taxas de conclusão da faculdade de alunos com pontuações no PSAT logo acima e logo abaixo dos limites do Prêmio de Mérito Nacional**. [...] Neste exemplo, os **saltos na relação entre as pontuações do PSAT e a frequência à faculdade na vizinhança do limite do prêmio são tomados como evidência de um efeito do tratamento**. É esse salto nas linhas de regressão que dá o nome a RD.

Angrist e Pischke (2008, p. 189-190,
“Mostly harmless Econometrics: an empiricist’s companion”)

Regressão com descontinuidade: exemplo

Critérios do programa

- Elegibilidade baseada em índice de pobreza calculado a partir de características dos domicílios no momento pré-intervenção
- Índice varia de 1 a 100 (quanto mais baixo, maior a pobreza)
- Domicílios com índice ≤ 50 são considerados pobres e elegíveis ao programa
- Domicílios com índice > 50 são considerados não pobres e não elegíveis ao programa

Objetivos da avaliação

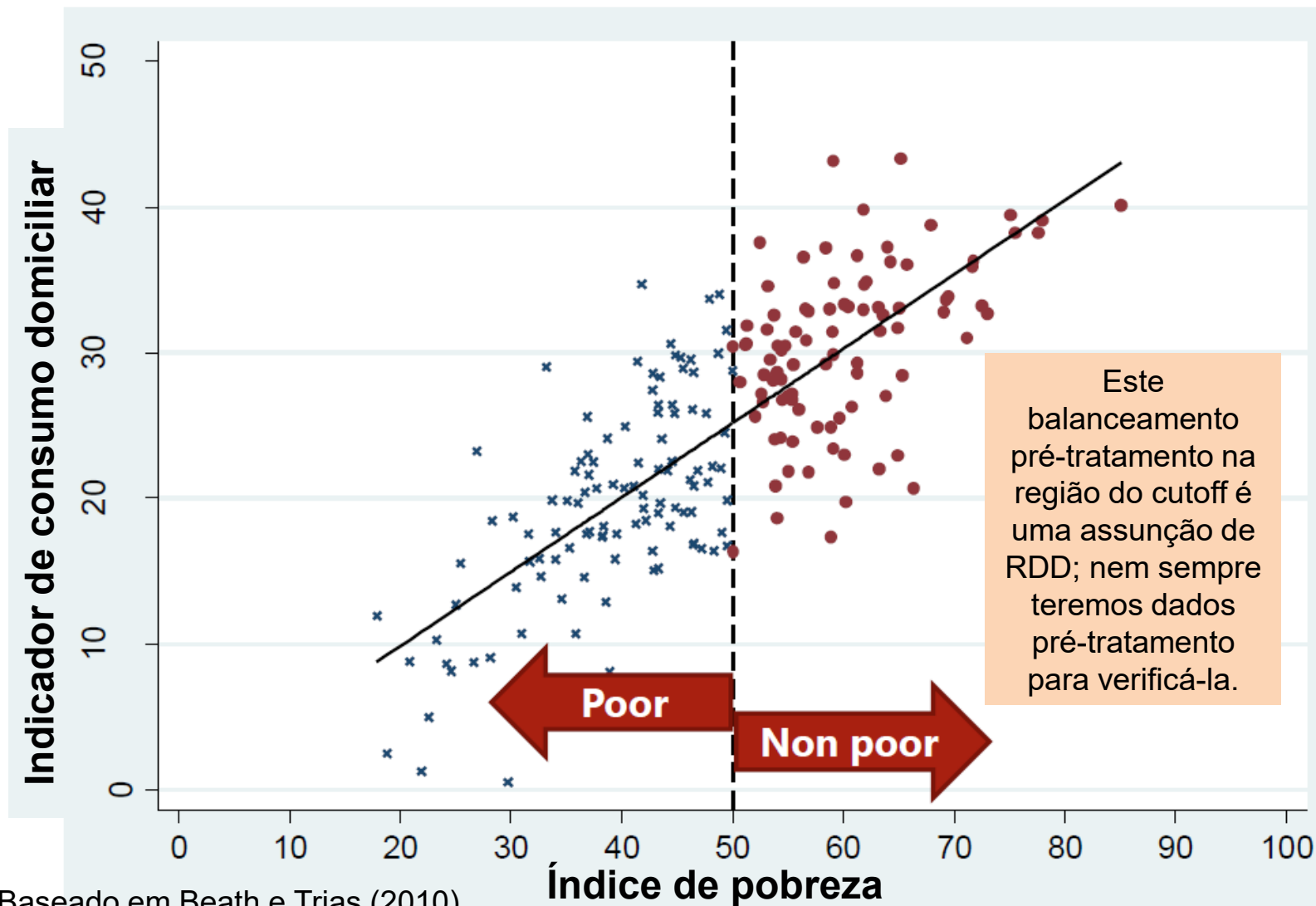
- Estimar efeitos da transferência de renda sobre o consumo domiciliar, comparando domicílios logo abaixo e logo acima do ponto de corte para elegibilidade ao programa (i.e., comparando domicílios localizados na vizinhança do valor 50 no índice de pobreza)

Baseado em Beath e Trias (2010)

<https://www.kdevelopedia.org/Resources/view/04201307240127417.do>

Regressão com descontinuidade: exemplo

PRÉ-INTERVENÇÃO: BASELINE

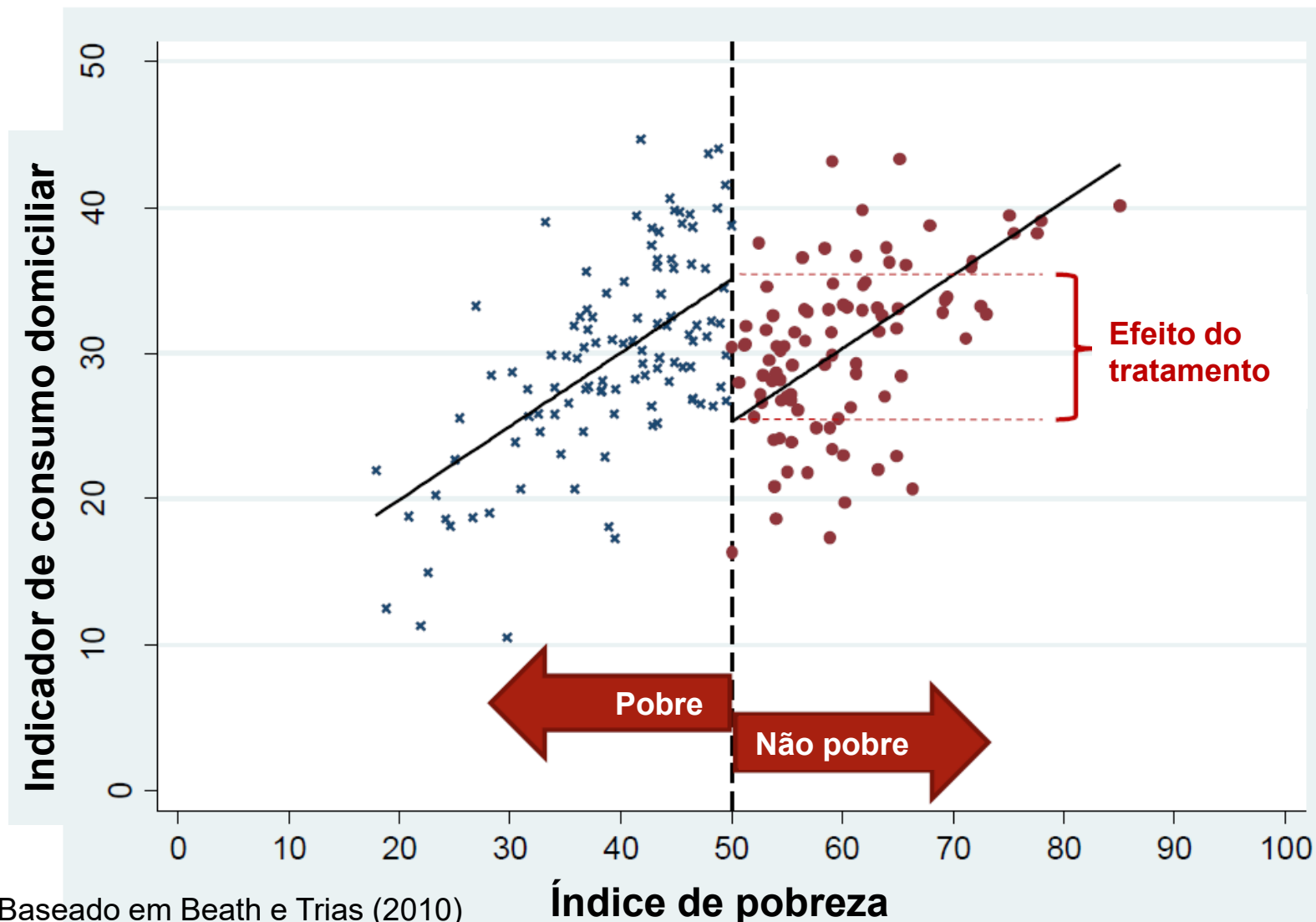


Baseado em Beath e Trias (2010)

<https://www.kdevelopedia.org/Resources/view/04201307240127417.do>

Regressão com descontinuidade: exemplo

PÓS-INTERVENÇÃO



Baseado em Beath e Trias (2010)

<https://www.kdevelopedia.org/Resources/view/04201307240127417.do>

Descontinuidade sharp vs. Descontinuidade fuzzy

Sharp



- **Elegibilidade = Participação:** a probabilidade de participação no tratamento (exposição ao tratamento) salta de 0 para 1 no ponto de corte (ou de 1 para 0)
- Elegibilidade exatamente determina o status de participação no tratamento: transferência ocorre se (e somente se) domicílio for elegível, e todos os domicílios elegíveis recebem transferência

Fuzzy



- **Elegibilidade \neq Participação:** a probabilidade de participação no tratamento não salta de 0 para 1 (ou de 1 para 0) no ponto de corte (mas salta, por exemplo, de 0,9 para 0,2, no caso da transferência de renda com descontinuidade fuzzy)
- Existe **noncompliance**:
 - Algumas unidades que não são elegíveis com base em sua pontuação no índice (running variable) recebem transferência (“always takers”)
E/ OU
 - Algumas unidades que são elegíveis com base em sua pontuação no índice não recebem transferência (“never takers”)
- Se aplicarmos o mesmo procedimento de estimação da descontinuidade sharp, teremos **viés de atenuação** (estimativa de efeito tende a ser mais próxima de zero que o respectivo parâmetro populacional)

Onde há erro de mensuração?* Se em Y, MQO funciona OK. Se em X, então teremos viés de atenuação

Erro de mensuração em Y: MQO é ok!

MQO funcionará bem se o erro de mensuração estiver apenas na variável dependente. Nesse caso, o **erro de mensuração é simplesmente parte do termo de erro geral**. Quanto maior o erro [de mensuração], maior a variância do termo de erro.

Bailey (2016, p. 220)

Se a variância do erro (i.e., a variância da regressão) crescer, então crescerá também o erro padrão dos β hat de inclinação

Erro de mensuração em X: Viés de atenuação (attenuation bias)

O truque aqui é pensar neste exemplo **como um problema de variável omitida onde v_i [o erro de mensuração] é a variável omitida**. Não observamos o erro de mensuração diretamente, certo? Se pudéssemos observá-lo, ajustaríamos nossa medida de X_1 . Então, o que fazemos é tratar o erro de mensuração como uma variável não observada que, por definição, devemos omitir e ver como essa forma particular de viés de variável omitida afeta o modelo.
[...]

Bailey está falando sobre β_1 hat, mas o mesmo é verdadeiro para outros coeficientes de inclinação

Referimo-nos a este exemplo particular de viés de variável omitida como viés de atenuação porque quando omitimos o termo de erro de medição do modelo, nossa estimativa β_1 hat se desvia do valor verdadeiro por um fator multiplicativo entre zero e um. Isso significa que β_1 hat tenderá a estar mais próximo de zero do que deveria estar quando X_1 for medido com erro. Se o verdadeiro valor de β_1 for algum número positivo, tendemos a ver valores de β_1 hat que são menores do que deveriam ser. Se o verdadeiro valor de β_1 for negativo, tendemos a ver valores de β_1 hat maiores (significando mais próximos de zero) do que deveriam ser.

Bailey (2016, p. 222-223, "Real Stats: Econometrics for Political Science, Public Policy, and Economics")

* Assume que o erro de mensuração (v_i) tem média zero e não está correlacionado com o valor verdadeiro (ou seja, a distância entre os valores reais e relatados não é sistematicamente diferente para diferentes níveis de X).

Convém transformar a running, apresentando-a como desvios em relação ao cutoff

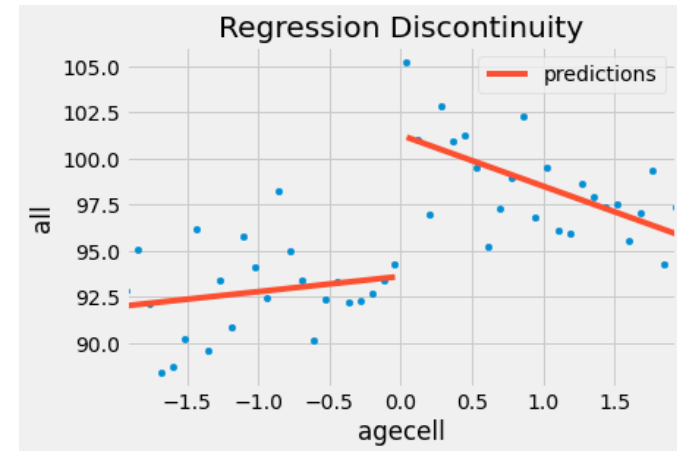
1

Facilita a interpretação do efeito estimado

*Aqui é onde entra o truque de centralizar a variável [running variable] em relação ao ponto de corte. Após essa etapa de pré-processamento, o **ponto de corte passa a ser zero**. Isso faz com que o intercepto seja o valor previsto no ponto de corte, para a regressão abaixo dele. (Facure, 2022, cap. 16; <https://matheusfacure.github.io/python-causality-handbook/landing-page.html>)*

2

Facilita a visualização dos dados (gráficos)



Fonte: Facure (2022, cap. 16)

3

É prática comum

Regressão com descontinuidade: exemplo SHARP



$$y_i = \beta_0 + \beta_1 T_i + \beta_2 \text{Índice}_i + \beta_3 T_i * \text{Índice}_i + \varepsilon_i$$

Embora não tenham sido incluídos controles aqui, a estimação de RDD pode incluir controles.

Essencialmente, isto é o mesmo que estimar uma regressão acima e outra abaixo do ponto de corte. β_0 é o intercepto da regressão se $T_i = 0$, e $\beta_0 + \beta_1$ é o intercepto da regressão se $T_i = 1$.

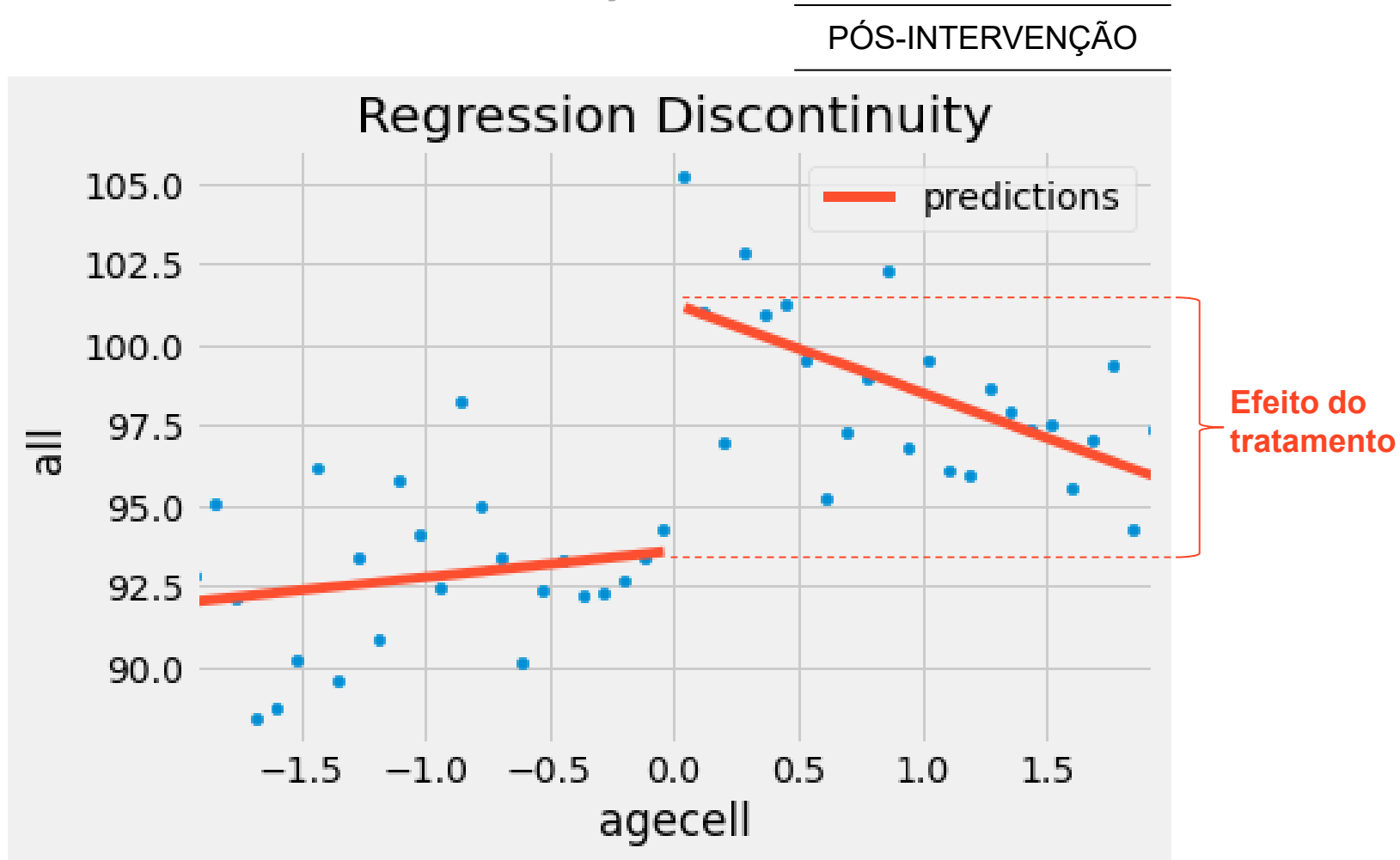
$$T_i = \begin{cases} 1 & \text{se domicílio } i \text{ é elegível (Índice}_i \leq 50\text{); por se tratar de RDD sharp, coincide exatamente com domicílios que participaram no tratamento (i.e., que receberam transferência);} \\ 0 & \text{se domicílio } i \text{ não é elegível (Índice}_i > 50\text{); por se tratar de RDD sharp, coincide exatamente com domicílios que não participaram no tratamento (i.e., que não receberam transferência).} \end{cases}$$

$$\text{Índice}_i = \text{Running variable (variável contínua utilizada para determinar elegibilidade dos domicílios), apresentada como desvios em relação ao cutoff.}$$

$$\beta_1 = \text{Descontinuidade no nível de consumo causada pelo recebimento da transferência de renda (i.e., o **efeito do tratamento**).}$$

Regressão com descontinuidade: exemplo de Facure (2022, cap. 16)

Note como o termo de interação dá mais flexibilidade à estimação do efeito, permitindo que as retas abaixo e acima do cutoff tenham não apenas interceptos distintos, mas também inclinações distintas



Ponderação usando função kernel triangular

- Dependendo da distribuição dos dados, **MQO pode focar em “encaixar” a linha estimada** de modo a **melhor prever** as observações (valores de Y) que estão **longe do ponto de corte**, em detrimento do “encaixe” na região próxima ao ponto de corte (que é a região de interesse em RDD)
- Um modo de solucionar esse problema é **dar peso maior para as observações próximas ao ponto de corte, usando a chamada função kernel triangular**:

$$K(R, c, h) = 1\{|R - c| \leq h\} * \left(1 - \frac{|R - c|}{h}\right)$$

Em que:

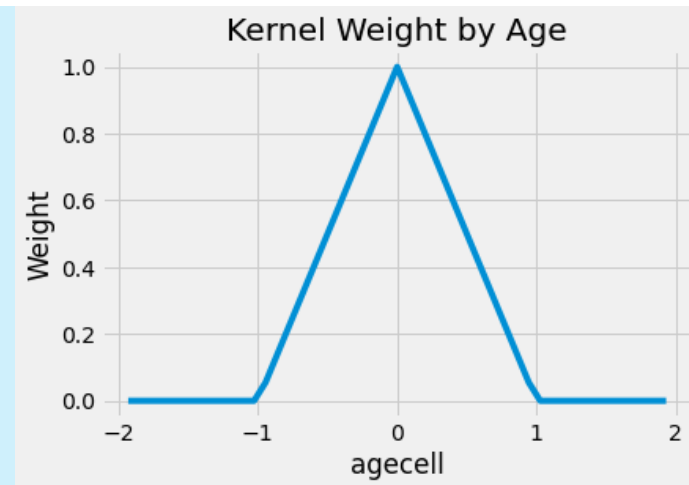
R = running Variable

c = ponto de corte

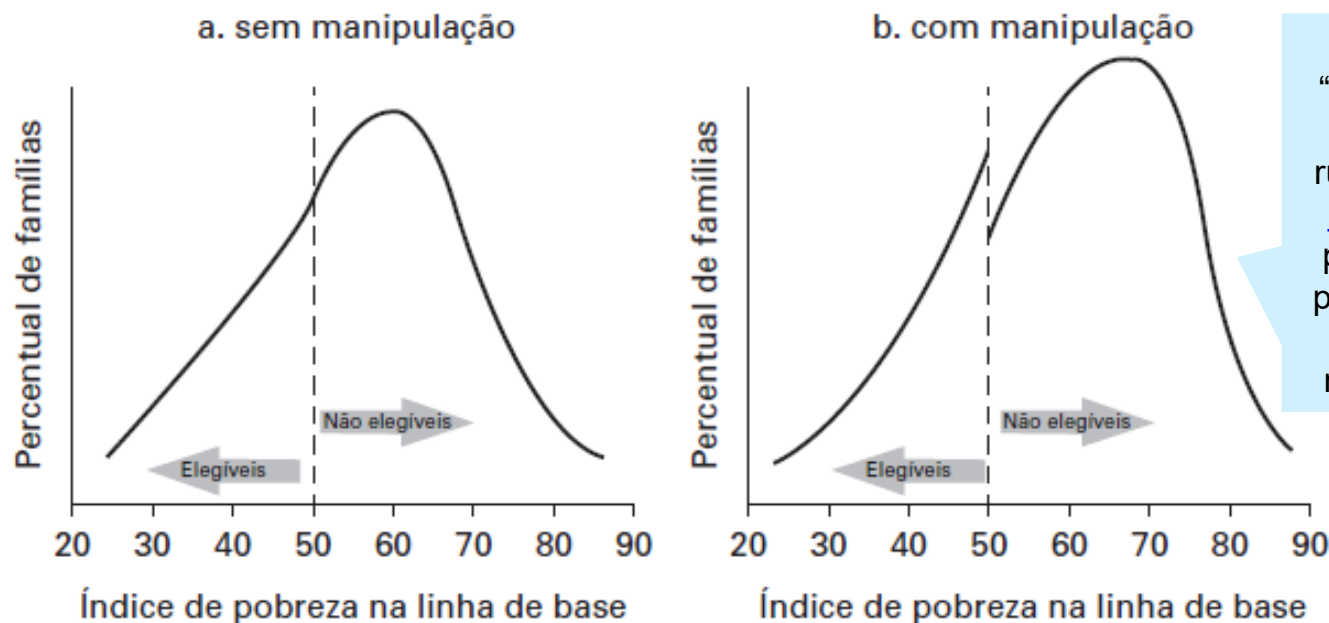
h = parâmetro referente à banda utilizada

“A primeira parte desse kernel é uma função indicadora de se estamos próximos ao limite. Quão próximos? Isso é determinado por um parâmetro de largura de banda (bandwidth). A segunda parte desse kernel é uma função de ponderação. À medida que nos afastamos do limite, os pesos ficam cada vez menores. Esses pesos são divididos pela largura de banda. Se a largura de banda é grande, os pesos diminuem mais lentamente. Se a largura de banda é pequena, os pesos rapidamente vão a zero.” (Facure, 2022, cap. 16;

<https://matheusfacure.github.io/python-causality-handbook/landing-page.html>)



Manipulação do índice de elegibilidade



Aqui temos “bunching on the density of the running variable”. [McCrary \(2008\)](#) propõe um teste para diferença de densidade na região do *cutoff*.

“O **painel a** mostra a distribuição dos domicílios de acordo com seu índice de elegibilidade [...] **quando não há manipulação**. A **densidade de domicílios situados em torno do ponto de corte (50)** é contínua (ou suave). O **painel b** mostra uma situação diferente: **um número maior de famílias parece estar “concentrado” logo abaixo do ponto de corte, enquanto uma quantidade relativamente pequena de famílias pode ser encontrada logo acima do ponto de corte**. Como não há uma razão a priori para acreditarmos que deveria haver uma grande alteração no número de domicílios em torno do ponto de corte, a ocorrência dessa **mudança na distribuição em torno desse ponto é evidência de que, de alguma maneira, as famílias podem estar manipulando suas pontuações para obter acesso ao programa.**” (Gertler et al., 2008, p. 132-133)

Teste de manipulação de McCrary (e aperfeiçoamentos posteriores):

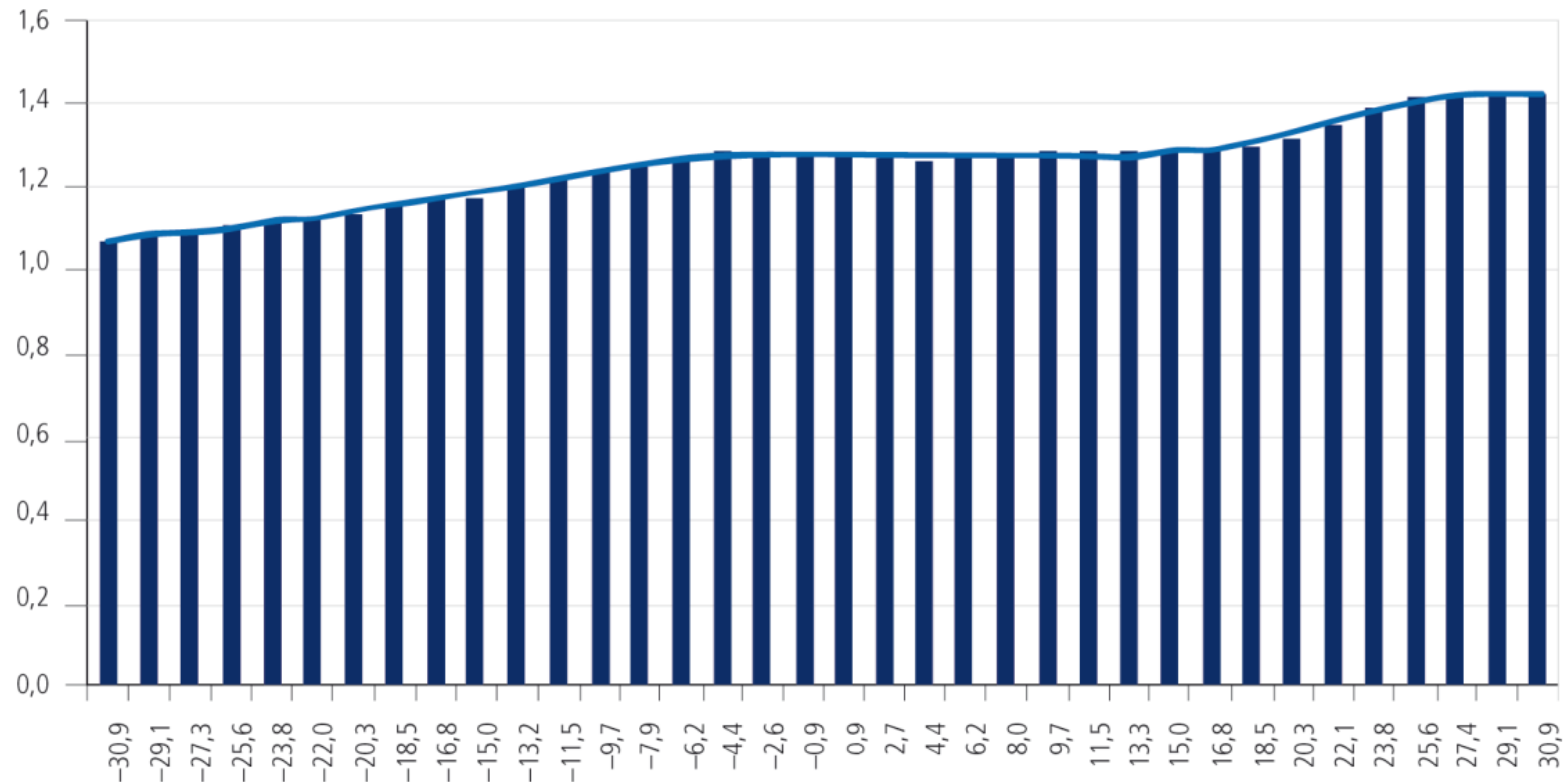
- **Python:** pacote “rddensity” (<https://rdpackages.github.io/rddensity/>)
- **R:** função “DCdensity” do pacote “rdd” (<https://cran.r-project.org/web/packages/rdd/index.html>)
- **Stata:** função user-written “DCdensity” (<http://eml.Berkeley.edu/~jmccrary/DCdensity/>)

Atenção! Nem todas as running variables apresentam distribuição bell-shaped (em formato de sino)

GRÁFICO 3

Estimativa da função de densidade do hiato em meses entre a idade do filho mais novo em 31 de dezembro de 2005 e 16 anos exatos

(Em %)



Fonte: PNAD de 2006.



Barbosa e Corseuil (2014, p. 26)

<https://repositorio.ipea.gov.br/handle/11058/3026>

Estimação com RDD fuzzy: diferentes abordagens

1

Estimador de Wald:

- Exemplificado em Facure (2022, cap. 16)
- É a mais direta, mas a menos flexível das opções aqui listadas pois não permite a inclusão de covariáveis; corresponde a uma **divisão de coeficientes estimados**

Equação do **numerador**: $y_i = \beta_0 + \beta_1 T_i + \beta_2 \text{Índice}_i + \beta_3 T_i * \text{Índice}_i + \varepsilon_i$

Equação do **denominador**: $D_i = \beta_0 + \beta_1 T_i + \beta_2 \text{Índice}_i + \beta_3 T_i * \text{Índice}_i + \varepsilon_i$

T_i = dummy indicando ou elegibilidade ao tratamento

D_i = dummy indicando participação no tratamento

$$\text{Estimador de Wald} = \frac{\widehat{\beta}_1 \text{ do numerador}}{\widehat{\beta}_1 \text{ do denominador}}$$

Abraham Wald (1902-1950) foi um matemático e estatístico húngaro-americano.

Foco das próximas páginas

Numerador: mesma especificação da RDD sharp.

Se houver muita noncompliance, $\widehat{\beta}_1$ do denominador será relativamente pequeno (porque elegibilidade não será um bom preditor de participação no tratamento) e estimador de Wald será maior que a estimativa naïve, dada pelo $\widehat{\beta}_1$ do numerador. Desta forma, o estimador de Wald ajustará a estimativa de efeito no sentido de reduzir o viés de atenuação.

2

Mínimos Quadrados em Dois Estágios (Two-Stage Least Squares – 2SLS)

- Opção mais conhecida e trabalhada nos manuais de Avaliação de Políticas Públicas e de Econometria
- Usa elegibilidade ao tratamento como variável instrumental para participação no tratamento
- Identifica o Local Average Treatment Effect (LATE) apenas:
 - i. na ausência de covariáveis (nas equações de 1º e 2º estágio), ou
 - ii. na presença de covariáveis, se efeitos para compliers e noncompliers forem totalmente independentes das covariáveis (o que é improvável)

3

Função de Resposta Média Local (Local Average Response Function – LARF)

- Opção mais sofisticada das três aqui apresentadas; referências: Abadie (2003); An e Wang (2016)
- Baseia-se em pesos calculados com base na probabilidade de elegibilidade ao tratamento, dadas características das unidades
- Relaxa o pressuposto de efeitos constantes (i.e., independentes do nível das covariáveis)

Estimation for fuzzy discontinuity

Equação de identificação de efeitos no contexto de RDD sharp; neste exemplo, Beath e Trias (2010) preferiram não incluir interação, mas poderiam ter incluído.

$$y_i = \beta_0 + \beta_1 D_i + \delta(score_i) + \varepsilon_i$$

D_i = dummy indicando participação no tratamento.

Estimation for fuzzy discontinuity

Equação de identificação de efeitos no contexto de RDD sharp; neste exemplo, Beath e Trias (2010) preferiram não incluir interação, mas poderiam ter incluído.

$$y_i = \beta_0 + \beta_1 D_i + \delta(score_i) + \varepsilon_i$$

D_i = dummy indicando participação no tratamento (também podemos dizer participação no tratamento).

IV estimation

- First stage:

$$D_i = \gamma_0 + \gamma_1 I(score_i > 50) + \eta_i$$

Dummy variable

$I(score_i > 50)$ = dummy indicando elegibilidade ao tratamento.

- Second stage:

$$y_i = \beta_0 + \beta_1 \hat{D}_i + \delta(score_i) + \varepsilon_i$$

Continuous function

\hat{D}_i = variação de D_i explicada pela elegibilidade ao tratamento (i.e., variação exógena de D_i).

Variável instrumental em RDD fuzzy: exemplo

CAPÍTULO 21

BOLSA FAMÍLIA, ESCOLHA OCUPACIONAL E INFORMALIDADE NO BRASIL*

Ana Luiza Neves de Holanda Barbosa
Carlos Henrique Leite Corseuil

O objetivo deste trabalho é o de investigar o impacto do PBF na composição do mercado de trabalho no Brasil. Em particular, é o de analisar em que medida o PBF distorce a escolha ocupacional dos beneficiários no sentido de torná-los mais propensos a optar por uma ocupação informal. A análise do impacto do PBF será realizada tanto no mercado de trabalho dos chefes de domicílio como também para o que aqui serão chamados de trabalhadores secundários do domicílio, isto é, que geram a segunda maior fonte de rendimento do trabalho ao domicílio, após o rendimento do trabalho principal do chefe. A nossa investigação será feita com base em uma metodologia que leva em conta uma possível autoseleção, baseada em características não observáveis, no PBF. Este arcabouço utiliza uma descontinuidade presente na regra de elegibilidade do PBF, que vem a ser um limite para a idade das crianças. Essa metodologia é conhecida como desenho de regressão descontínua – RDD (*regression discontinuity design*). Em particular, como a elegibilidade não coincide com a participação no programa, a metodologia explorada trata do caso conhecido como *fuzzy* RD.

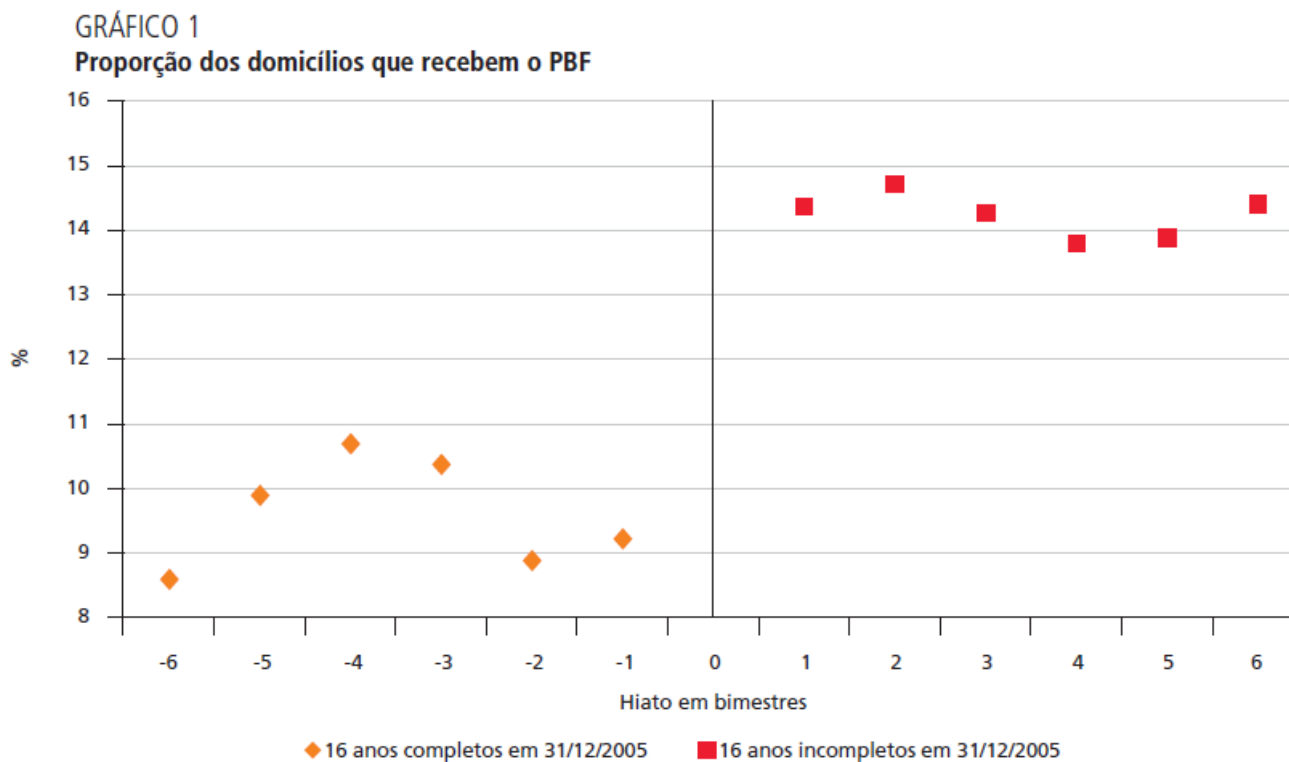
Barbosa e Corseuil (2013, p. 329)
<http://repositorio.ipea.gov.br/handle/11058/2082>

Variável instrumental em RDD fuzzy: exemplo

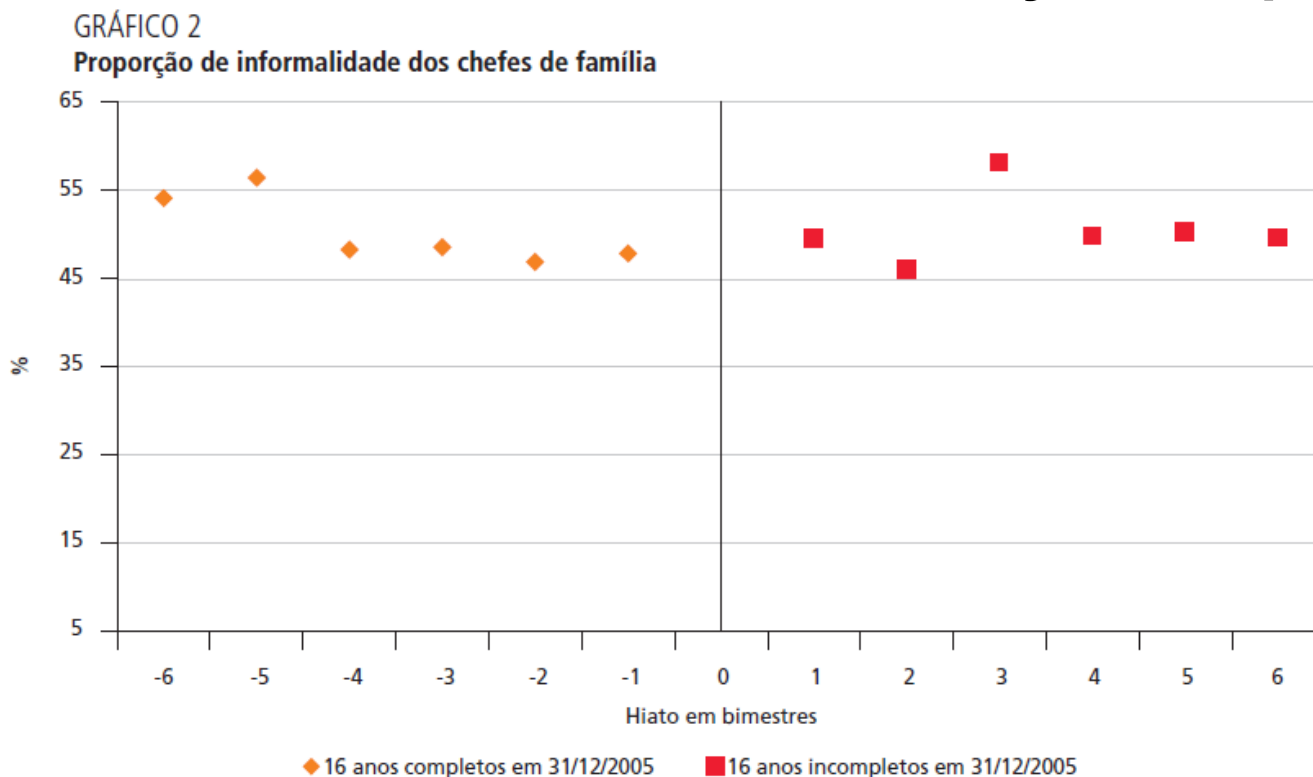
Há um grande desafio para a investigação dos efeitos de programas de transferência de renda na oferta de trabalho do adulto. A participação nesse tipo de programa é voluntária. Logo, o grupo de beneficiários pode diferir em certos determinantes da escolha ocupacional em relação ao grupo de não beneficiários. Isso dificulta a tarefa de encontrar um grupo de comparação que lembre o grupo tratado. O relatório técnico de Brauw *et al.* (2012) e o estudo de Teixeira (2010) usam métodos que tentam delimitar os grupos de forma a harmonizá-los no que se refere a características observáveis dos indivíduos. No entanto, nesse tipo de contexto é muito provável que seja necessário também harmonizar certas características não observáveis.

Variável instrumental em RDD fuzzy: exemplo

O gráfico 1 mostra claramente que a participação no programa cai abruptamente ao redor da marca correspondente ao limite para um dos critérios de elegibilidade. Ou seja, a comparação entre domicílios com filhos mais novos com 16 anos recém-completos, ou a serem completados pouco tempo depois de 31 de dezembro de 2005, traz diferenças muito marcantes na probabilidade de participação no programa em 2006. Sendo razoável supor que tudo mais deve ser muito parecido entre esses dois grupos de domicílios (incluindo valores de ε), tem-se na idade do filho mais novo em torno dos 16 anos a variação exógena que desloca T mantendo tudo mais constante e, portanto, permitindo a identificação do efeito do Bolsa Família.



Variável instrumental em RDD fuzzy: exemplo



Fonte: PNAD 2006.

Nota-se que o emprego no setor informal é bastante semelhante entre os dois grupos de comparação. Ou seja, entre os chefes ocupados a proporção no setor informal parece não variar com a elegibilidade do domicílio ao PBF. Vale ressaltar que o grupo elegível mostrava no gráfico 1 uma participação consideravelmente maior no PBF que o grupo não elegível. Em outros termos, elegibilidade influencia participação que por sua vez não parece influenciar informalidade. Esse resultado será confirmado nas estimativas aqui pelo método de variáveis instrumentais, adaptado ao contexto de regressão com descontinuidade difusa.

Variável instrumental em RDD fuzzy: exemplo

TABELA 2

Balanceamento – características dos elegíveis e não elegíveis

| | Amostra total Todos os domicílios que recebem PBF | | Filtro I Domicílios com RDPC < R\$ 700 | | Filtro II Domicílios com RDPC > R\$ 50 e RDPC < R\$ 700 | |
|---|---|---------------|--|---------------|---|---------------|
| | Elegíveis | Não elegíveis | Elegíveis | Não elegíveis | Elegíveis | Não elegíveis |
| Características individuais | | | | | | |
| Sexo | 77,2 | 75,5 | 78,2 | 77,6 | 78,5 | 77,1 |
| (% de homens) | | | | | | |
| Anos de escolaridade | 6,8 | 7,2 | 6,0 | 6,0 | 6,1 | 6,1 |
| Idade | 45,3 | 44,4 | 44,8 | 43,7 | 44,8 | 43,6 |
| Características do domicílio | | | | | | |
| Número de pessoas na família | 3,6 | 3,5 | 3,6 | 3,5 | 3,6 | 3,5 |
| Residentes nas regiões Norte e Nordeste (%) | 38,7 | 40,6 | 39,9 | 42,4 | 42,0 | 38,4 |
| Renda total <i>per capita</i> (%) | 508,1 | 544,1 | 306,3 | 304,4 | 313,3 | 310,1 |
| Número de observações | 284 | 310 | 238 | 250 | 232 | 245 |

Fonte: PNAD de 2006.

Barbosa e Corseuil (2014, p. 24)

<https://repositorio.ipea.gov.br/handle/11058/3026>



“A tabela 2 apresenta uma comparação de algumas características observáveis do chefe e da família registradas na PNAD de 2006 entre domicílios em que a idade do filho mais novo é ligeiramente menor e ligeiramente maior do que 16 anos em 31 de dezembro de 2005. **Para delimitar os grupos demarcados como elegíveis e não elegíveis, foi usado um raio de dois meses, tal como feito na seção anterior.** Também seguiu-se a tabela 1 no que diz respeito à delimitação de três amostras distintas de acordo com o uso da renda domiciliar *per capita* para esse fim.” (Barbosa; Corseuil, 2014, p. 23)

Variável instrumental em RDD fuzzy: exemplo

TABELA 1

Impacto do PBF sobre a escolha ocupacional do chefe de família

Não conseguimos localizar informação sobre a escolha desses filtros.

| | Amostra total todos os Domicílios que recebem PBF | | Filtro I Domicílios com RDPC < R\$ 700,00 | | Filtro II Domicílios com RDPC > R\$ 50,00 & RDPC < R\$ 700,00 | |
|--|---|-----------------|---|-----------------|---|-----------------|
| | Coef. | Janela ótima | Coef. | Janela ótima | Coef. | Janela ótima |
| Bolsa Família (erro-padrão) | -3,237 (3,239) | 1,608 | -2,417 (1,460) | 1,780 | -2,429 (1,485) | 1,759 |
| Bolsa Família (<i>bandwidth</i> = 110) (erro-padrão) | -2,339 (1,612) | 1,769 | -2,845 (1,797) | 1,958 | -2,629 (1,580) | 1,935 |
| Bolsa Família (<i>bandwidth</i> = 125) (erro-padrão) | -1,996 (1,387) | 2,011 | -2,607 (1,875) | 2,225 | -2,697 (1,747) | 2,199 |
| Bolsa Família (<i>bandwidth</i> = 150) (erro-padrão) | -1,122 (0,973) | 2,413 | -2,539 (1,614) | 2,670 | -2,648 (1,562) | 2,639 |

Fonte: PNAD 2006.

Barbosa e Corseuil (2013, p. 337)

<http://repositorio.ipea.gov.br/handle/11058/2082>

Colunas “Coef.”: “Trata-se dos valores estimados para o efeito do programa Bolsa Família sobre a probabilidade da ocupação principal do chefe ser **informal** [...]” (Barbosa; Corseuil, 2013, p. 336)

Sobre a “janela ótima” para RDD, vide [Calonico, Cattaneo e Titiunik \(2014\)](#): “Robust nonparametric confidence intervals for regression-discontinuity designs”.

Não conseguimos localizar informação sobre qual estimação gerou esses coeficientes (e.g., se logit, probit), nem os autores os interpretam.

Regressão com descontinuidade: limitações

- Fornece estimativas do **efeito médio local de tratamento** (local average treatment effect – LATE); efeitos estimados podem não ser válidos para unidades em outras regiões da running variable
- **Efeitos estimados** podem ser **sensíveis à forma funcional da regressão** usada na modelagem da relação entre o índice de elegibilidade e o resultado de interesse, bem como à **amplitude da banda/janela** considerada em torno do cutoff (quão perto é adequadamente perto do ponto de corte?)
- **Conexão entre elegibilidade e ponto de corte** deve ser **exclusiva do programa** de interesse – i.e., regra de elegibilidade não pode ser usada em outros programas potencialmente incidentes sobre Y; caso contrário, efeitos estimados não serão específicos do programa de interesse
- Em caso de **manipulação** do índice de elegibilidade (e.g., artifício para fazer domicílio parecer elegível, quando não é), cria-se uma descontinuidade no índice que **mina a condição básica para que o método funcione** – qual seja, que o índice de elegibilidade seja contínuo em torno do ponto de corte
- No caso de **noncompliance**, o que implica descontinuidade **fuzzy**, efeitos estimados via variável **instrumental** valem apenas para unidades que são sensíveis aos critérios de elegibilidade – ou seja, o tipo que participa se (e somente se) for elegível. **Em RDD sharp, calculamos o LATE por duas razões:** i) estimativa refere-se apenas ao efeito do tratamento na vizinhança do ponto de corte, e ii) a estimativa refere-se apenas ao efeito do tratamento para os compliers

De quais abordagens trataremos?

Abordagem

Desenho com
descontinuidade (RDD*)



Variável instrumental

Diferença em diferenças



Efeitos fixos



Pareamento



* Regression Discontinuity Design.



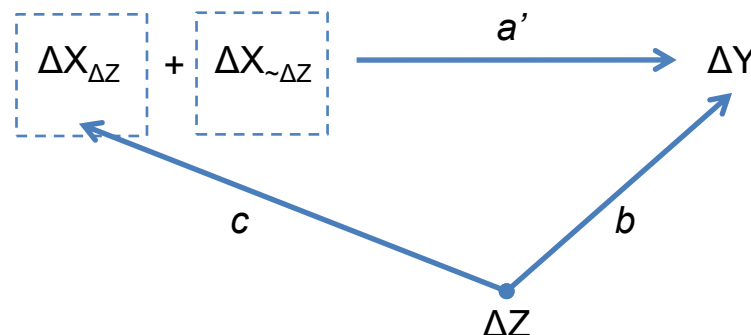
Variável instrumental: requisitos e intuição

Requisitos

- É possível usar uma *proxy* para o tratamento que atenda às seguintes condições:
 - **Relevância (good instrument):** deve correlacionar-se com a participação no tratamento (X); em outras palavras, deve afetar a probabilidade de as unidades serem tratadas
 - **Exogeneidade ou exclusão (valid instrument):** há diferentes maneiras de se declarar esta condição:
 - Instrumento afeta o resultado (Y) apenas através do tratamento que estamos interessados em avaliar
 - Instrumento não se correlaciona com as variáveis omitidas correlacionadas com participação no tratamento
 - Instrumento não se correlaciona com o erro (ε)

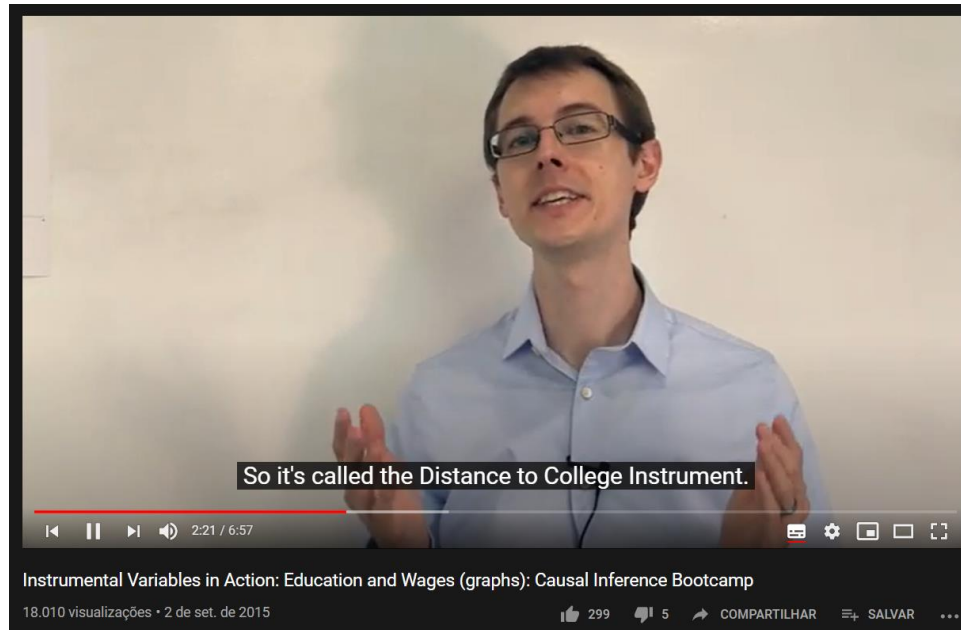
Intuição

- O instrumento identifica $\Delta X_{\sim \Delta Z}$: a porção da variação de X que é independente de Z



Variável instrumental: exemplo

Distância até a faculdade como instrumento para anos de estudo



<https://youtu.be/vacBsxBgFMY>

- Exemplo baseado em Card (1993): Using geographic variation in college proximity to estimate the return to schooling. <http://dx.doi.org/10.3386/w4483>
- Os anos de escolaridade são afetados pela habilidade inata, que também afeta os salários. Portanto, a habilidade inata é um potencial fator de confusão em nossa estimativa do efeito dos anos de estudo sobre os salários
- A distância até a faculdade tem correlação com frequentar faculdade (**good instrument**), mas não temos motivos para suspeitar que a distância até a faculdade tenha correlação com a habilidade inata (**valid instrument**)
- Portanto, a distância até a faculdade pode ser usada como um instrumento para a frequência à faculdade (e para anos de estudo)

Variável instrumental: exemplo

Distância até a faculdade como instrumento para anos de estudo

Este é o modelo “inicial”: modelo que se deseja estimar, mas que sofre de endogeneidade:

$$\text{Hourly_Wage} = \beta_0 + \beta_1 \text{Years_of_Education} + \beta \text{Controls} + \varepsilon$$

Em vez de estimar o modelo original, usamos a variável instrumental em uma estimação em 2 estágios:

1º estágio

Predizemos X (i.e., obtemos \hat{X}) com base na variável instrumental (exógena):

$$\text{Years_of_Education} = \alpha_0 + \alpha_1 \text{Distance_to_College} + \alpha \text{Controls} + \mu$$

2º estágio

Calculamos efeito de X em Y (i.e., estimamos γ_1) utilizando \hat{X} no lugar de X (i.e., utilizamos apenas a variação exógena de X para calcular o efeito de X):

$$\text{Hourly_Wage} = \gamma_0 + \gamma_1 \widehat{\text{Years_of_Education}} + \gamma \text{Controls} + \nu$$

Variável instrumental: exemplo

*Este artigo apresenta novas evidências do efeito causal do tamanho da família na **qualidade da infância** em um contexto de país em desenvolvimento. Estimamos o impacto do tamanho da família sobre o trabalho infantil e os resultados educacionais entre crianças e jovens brasileiros, explorando a variação exógena do tamanho da família impulsionada pela presença de gêmeos. Usando os dados do Censo Brasileiro de 1991, descobrimos que o aumento exógeno no tamanho da família está positivamente relacionado à participação de meninos e meninas na força de trabalho e às tarefas domésticas para mulheres jovens. Também encontramos efeitos negativos sobre os resultados educacionais de meninos e meninas e impactos negativos sobre a formação de capital humano para jovens do sexo feminino. Além disso, obtivemos evidências sugestivas de que as restrições de crédito e tempo enfrentadas pelas famílias pobres podem explicar os resultados.*

(Ponczek e Souza, 2012, p. 64; tradução nossa)

<https://doi.org/10.1353/jhr.2012.0006>

Unrelated
fact alert!



Cody Taylor
October 26, 2019 12:01 pm

Charlotte Hornets rookies Cody Martin and Caleb Martin became the first twins in NBA history on Friday to appear in the same game as teammates during their rookie seasons.

<https://therookiewire.usatoday.com/2019/10/26/cody-caleb-martin-first-twins-nba-history-same-game-rookie-teammates/>

Variável instrumental: exemplo

The presence of twins in a family has the two characteristics expected of a good IV. It is clearly correlated with family size and, because it is very likely to be a random occurrence, tends to be orthogonal to the error term in the main regression. A potential flaw of our strategy arises if there is any independent effect of the presence of twins on quality that does not operate through quantity. The presence of twins is directly associated with narrower spacing among sibling births. If average spacing is correlated with child-quality outcomes, then our instrument will not be valid. For instance, it is possible that breastfeeding twins may physically exhaust the mother, which may affect the raising of other children in the family. If this is the case, then the impact of family size on quality will be overestimated.

Our benchmark strategy consists of a Two Stage Least Square (2SLS) regression where, in the first step, we regress number of children (N_{ij}) on the presence of twins on the n th birth indicator variable (PT) and other predetermined variables (W):

$$(1) \quad N_{ij} = \alpha + \beta PT_j + \gamma' W_{ij} + \varepsilon_{ij}$$

The second step follows⁹:

$$(2) \quad Y_{ij} = \alpha + \hat{\beta} N_j + \gamma' W_{ij} + v_{ij}$$

where Y_{ij} is the outcome of interest of children i living with family j .¹⁰ The outcomes of interest for the 10–15-year-old age group are as follows: (i) labor force participation, defined as a binary variable that indicates whether the child participates in the labor market. Individuals are considered to be in the labor market if they have regularly or occasionally worked during the last 12 months or if they are currently searching for a job;¹¹ (ii) household chores, defined as a binary variable that indicates

Variável instrumental: exemplo

Dams

Esther Duflo and Rohini Pande

NBER Working Paper No. 11711

October 2005

JEL No. O21, O12, H43, H23

<http://dx.doi.org/10.3386/w11711>

ABSTRACT

A construção de grandes barragens é uma das formas mais caras e controversas de investimento em infraestrutura pública nos países em desenvolvimento, mas pouco se sabe sobre seu impacto. Este artigo estuda a produtividade e os efeitos distributivos de grandes represas na Índia. **Para levar em conta a localização endógena das barragens, usamos dados de GIS e o fato de que o gradiente do rio afeta a adequação de um distrito às barragens para fornecer estimativas de variáveis instrumentais de seu impacto.** Constatamos que, em um distrito onde uma barragem é construída, a produção agrícola não aumenta, mas a pobreza sim. Em contraste, os distritos localizados a jusante da barragem se beneficiam do aumento da irrigação e veem a produção agrícola aumentar e a pobreza diminuir. De modo geral, nossas estimativas sugerem que a construção de grandes barragens na Índia é um investimento marginalmente custo-efetivo com implicações distributivas significativas e que, em conjunto, aumentou a pobreza.

Variável instrumental: limitações

- É muito difícil encontrar uma variável instrumental boa (relevante) e válida (exógena)!
- Fornece estimativas do **efeito médio local de tratamento** (*local average treatment effect* – LATE) **apenas para** o tipo que participa se (e somente se) assim ditar o instrumento
- Estimador sofre de **imprecisão e viés em amostras pequenas**:

Em primeiro lugar, a suposição de que uma IV não tem um efeito líquido direto na variável de resultado costuma ser difícil de defender. Em segundo lugar, mesmo quando uma IV não tem um efeito líquido direto na variável de resultado, os estimadores de IV são enviesados em amostras finitas. Além disso, esse viés pode ser substancial quando um instrumento prevê apenas fracamente a variável causal.

[...]

Ao usar apenas uma parte da covariação entre a variável causal e a variável de resultado, os estimadores para IV usam apenas uma parte das informações nos dados. Isso representa uma perda direta no poder estatístico e, como resultado, os estimadores para IV tendem a exibir substancialmente mais variância amostral esperada do que outros estimadores.

(Morgan e Winship, 2007, p. 197-198; tradução nossa)

De quais abordagens trataremos?

Abordagem

Desenho com
descontinuidade (RDD*)



Variável instrumental



Diferença em diferenças

Efeitos fixos



Pareamento



* Regression Discontinuity Design.



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)

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

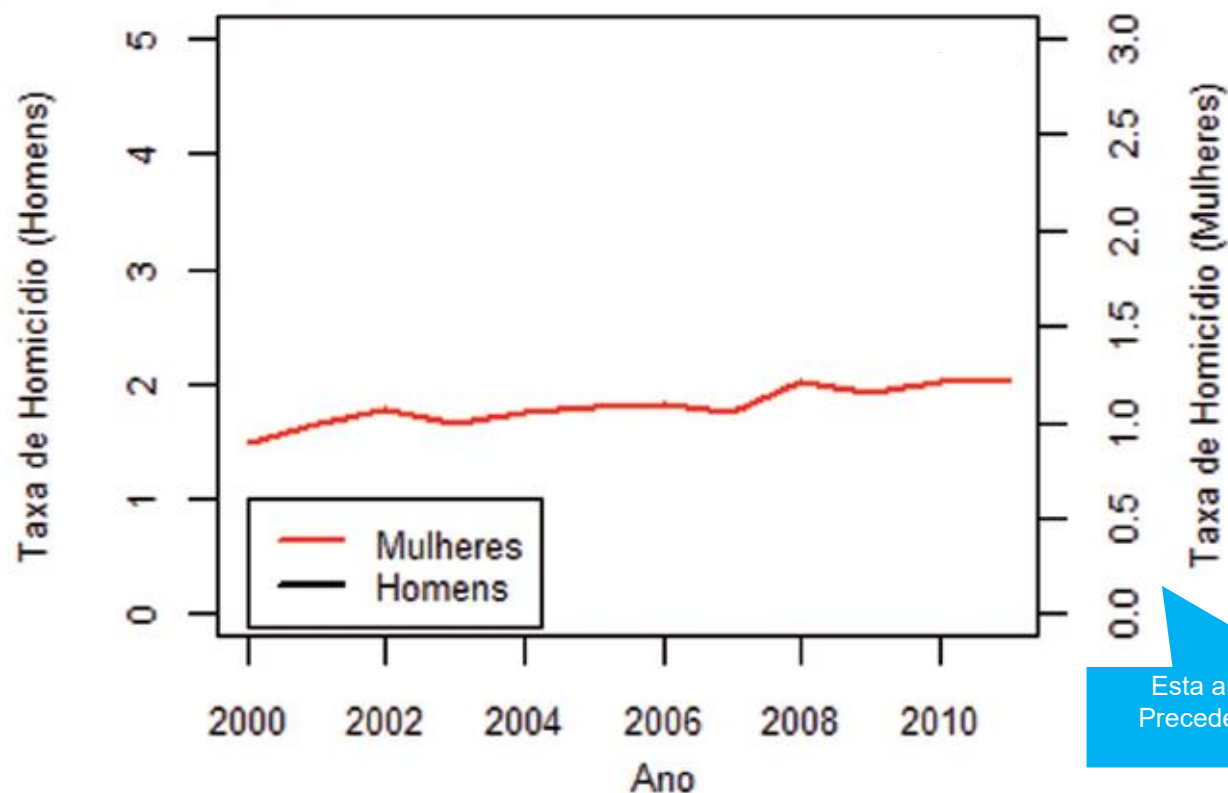
Jony Pinto Junior⁵

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)



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

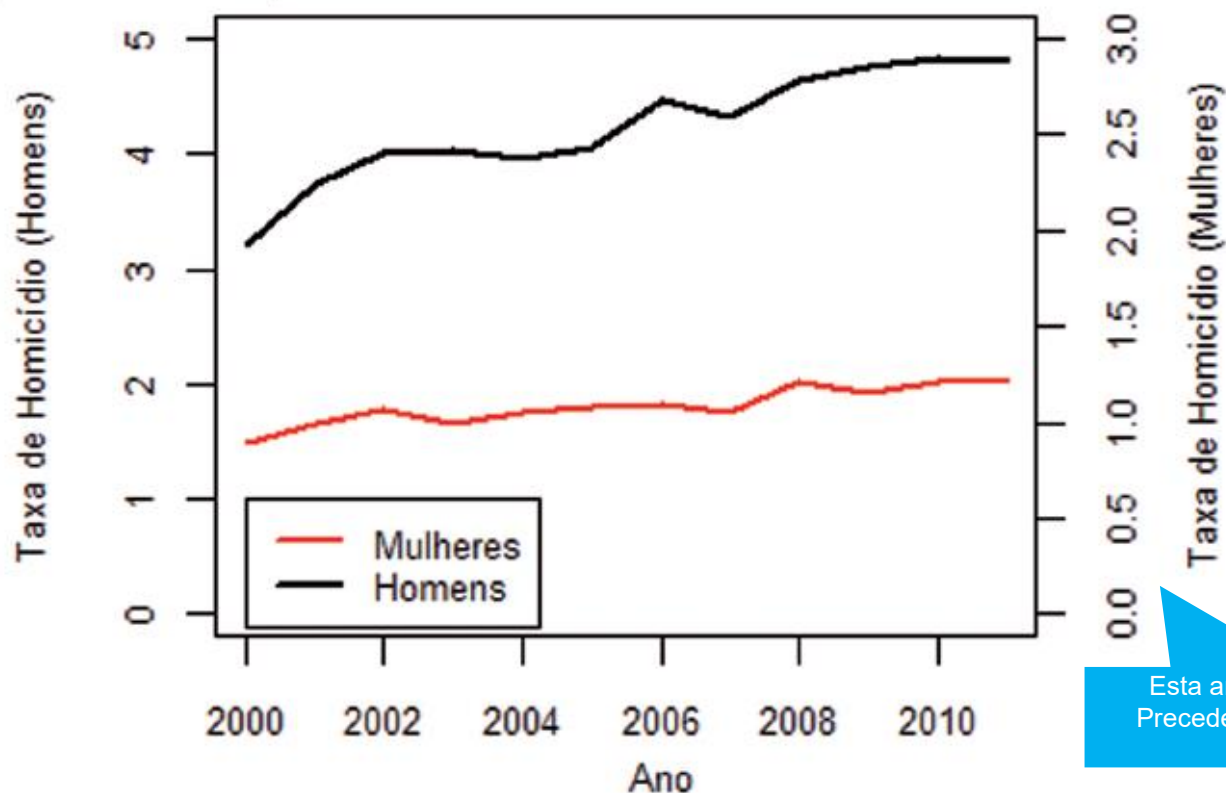
Daniel Cerqueira²
Mariana Vieira Martins Matos³
Ana Paula Antunes Martins⁴
Jony Pinto Junior⁵

<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)

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; DinD 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)

De quais abordagens trataremos?

Abordagem

Desenho com
descontinuidade (RDD*)



Variável instrumental



Diferença em diferenças



Efeitos fixos

Pareamento



* Regression Discontinuity Design.



Reflexões iniciais

- Em DiD, combinamos 2 comparações (depois vs. antes e tratamento vs. comparação) para neutralizar dois tipos de confounders (variáveis omitidas correlacionadas com o tratamento):
 - a_i : fatores constantes no tempo mas diferentes entre grupos (i);
 - d_t : fatores variáveis no tempo (t) que incidem sobre os dois grupos.
- Podemos aplicar essa lógica a dados longitudinais (e.g., de painel, de corte transversal empilhado) mesmo quando não temos dois grupos (tratamento e comparação) definidos, desde que tenhamos variação do tratamento no tempo para as unidades observadas.

Efeitos fixos: requisitos e intuição

Requisitos

- Dados **longitudinais** para **múltiplas unidades** observadas (**pelo menos duas medições por unidade**).
 - Idealmente, dados de painel; no mínimo, dados de corte transversal empilhados.
- **Fatores explicativos variantes no tempo** e entre **unidades**.
 - Fatores explicativos constantes no tempo serão absorvidos por efeitos fixos de entidade.
 - Fatores explicativos invariantes entre unidades serão absorvidos pelos efeitos fixos de tempo.

Intuição

- Modela as **singularidades não observadas** de cada agrupamento através de **interceptos próprios** aos grupos (nota: estimação nem sempre envolve dummies por grupos, mas a intuição permanece válida).
- Agrupamentos podem ser por **entidade** (grupo não temporal – e.g., municípios, estados) **e/ou por tempo** (momento da medida).

Intuição: efeitos fixos de entidade vs. tempo

Efeitos fixos de entidade

Interceptos próprios a cada entidade captam efeito combinado de variáveis não observadas que afetam o Y, variam entre entidades (mas não dentro delas), sendo constantes no tempo (pelo menos, no intervalo temporal observado).

Efeitos fixos de tempo

Interceptos próprios a cada momento de observação captam efeito combinado de variáveis não observadas que afetam o Y, não variam entre entidades (i.e., afetam todas as entidades da mesma forma num certo momento do tempo), e que são variáveis no tempo.

Exemplo: efeitos fixos de entidade e de tempo

A RELAÇÃO ENTRE O DESEMPENHO DA CARREIRA NO MERCADO DE TRABALHO E A ESCOLHA PROFISSIONAL DOS JOVENS

Otávio Bartalotti[§]

Naércio Menezes-Filho[□]

ECON. APLIC., SÃO PAULO, V. 11, N. 4, P. 487-505, OUTUBRO-DEZEMBRO 2007

<https://doi.org/10.1590/S1413-80502007000400002>

RESUMO

Este artigo examina como o desempenho relativo no mercado de trabalho de cada profissão afeta a escolha profissional dos futuros universitários. O desempenho no mercado de trabalho é medido pela média e pelo desvio padrão dos salários recebidos por profissão e pela sua taxa de desemprego no censo demográfico nos anos próximos ao vestibular. O número de pleiteantes a ingresso na carreira é medido pelo número de inscritos no exame da Fuvest. Utiliza-se dados em painel para os anos de 1991 e 2000 para controlar pelo efeito específico de cada profissão. Os resultados apontam para um efeito positivo e robusto do salário médio da profissão sobre a escolha profissional, que persiste na análise de painel, e para efeitos negativos da dispersão salarial da renda e do desemprego que, no entanto, não se mostraram significantes.

Palavras-chave: carreiras, profissões, mercado de trabalho, renda, Fuvest.

Exemplo: efeitos fixos de entidade e de tempo

Tabela 3 – Determinantes do número de inscritos nos cursos

| Variável Dependente | Logaritmo da média do número de inscritos apurada nos processos seletivos para ingresso em 1992-1993 e 2001-2002. | Efeitos Fixos | |
|--------------------------|---|---------------------------|---------------------------|
| | | Coeficientes | Coeficientes |
| Constante | | -6.810 (6.670) | -7.693 (6.602) |
| Ln (Renda) | | 2.086** (0.970) | 2.041** (0.964) |
| Desvio padrão Ln (Renda) | | -0.705 (1.018) | -0.654 (1.010) |
| Ln (Desemprego) | | -0.006 (0.085) | 0.010 (0.086) |
| <i>Dummy ano</i> | | -0.336 (0.329) | -0.321 (0.329) |
| Ln (Vagas) | | | 0.240 (0.220) |
| R ² | | 0.120 | 0.258 |

Notas: Estatística *t* entre parênteses. ***= coeficiente significativo a 1%; **= coeficiente significativo a 5%; *= coeficiente significativo a 10%.

Adaptado de Bartalotti e Menezes-Filho (2007, p. 500)

<https://doi.org/10.1590/S1413-80502007000400092>

Gestão local e seu impacto no desempenho fiscal: análise dos municípios mineiros

Ana Paula Karruz

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil

André Machado Moraes

Banco de Desenvolvimento de Minas Gerais (BDMG), Belo Horizonte, MG, Brasil

Karruz e Moraes (2020)

<https://doi.org/10.21874/rsp.v71i1.3777>

Resumo

O estudo busca elucidar e articular a gama de condicionantes do desempenho fiscal local, dimensionar a influência da gestão fiscal municipal nesse desempenho, e descrever como a qualidade da gestão fiscal se distribui entre municípios com diferentes características. Propõe um modelo conceitual e de regressão para o desempenho fiscal, tendo o Índice Firjan de Gestão Fiscal (IFGF) dos municípios de Minas Gerais como variável dependente. A identificação de efeitos apoia-se na natureza longitudinal dos dados (2007-2016) e na modelagem em painel. Transferências voluntárias da União e do governo estadual parecem favorecer o desempenho fiscal local, especialmente via investimentos. Um terço da variação do IFGF não pode ser atribuída às covariáveis e é tomada, por eliminação, como *proxy* para a qualidade da gestão fiscal. Esta apresenta moderada correlação com a população e mostra sinais de dependência espacial, com concentração de valores altos na região da capital.

Palavras-chave: gestão fiscal, municípios, Índice Firjan de Desempenho Fiscal (IFGF)

Efeitos fixos: exemplo

Quadro 04 | Principais influências sobre o desempenho fiscal municipal, por natureza de variação

| Varia entre municípios? | Varia frequentemente? | | | |
|-------------------------|--|---|--|--|
| | Não (fatores estruturais) | | Sim (fatores conjunturais) | |
| Não | Fatores (institucionais): <ul style="list-style-type: none">- Limites constitucionais da tributação municipal- Regras para transferências não discricionárias- Centralização do policy decision-making (e.g., percentuais mínimos de gastos em saúde e educação)- Determinações da Lei de Responsabilidade Fiscal (LRF) | Referências: <ul style="list-style-type: none">- Arretche (2004);- Baião, Cunha e Souza (2017); Veloso et al. (2011);- Arretche (2012);- Sistema Firjan (2017a). | Fatores: <ul style="list-style-type: none">- Contexto fiscal (federal, estadual)- Contexto econômico (nacional, estadual) | Referências: <ul style="list-style-type: none">- Sistema Firjan (2017a);- Sistema Firjan (2017a). |
| Sim | Fatores (estruturais locais): <ul style="list-style-type: none">- Ano de criação- Outros fatores estruturais locais, tais como:<ul style="list-style-type: none">° Recursos naturais° Principais atividades econômicas estabelecidas° Infraestrutura instalada° Extensão territorial° Status de capital° Proximidade com cidades-polo | Referências: <ul style="list-style-type: none">- Araújo Júnior (2016)- (Autores do presente trabalho reputam estes fatores como influentes sobre a capacidade de arrecadação própria e a demanda por serviços públicos) | Fatores: <ul style="list-style-type: none">- Contexto fiscal (municipal)- Potencial de arrecadação própria° Contexto econômico (municipal)° Tamanho populacional (base de arrecadação)- Demanda por serviços públicos (escala e complexidade)- Capacidade de atração de recursos públicos* (alinhamento partidário)- Transferências intergovernamentais (realizadas)° Compulsórias redistributivas (influxo de baixo custo)° Voluntárias- Ciclos políticos° Alternância ideológica° Ano eleitoral° Probabilidade de manutenção do incumbente ou seu grupo no poder- Qualidade da gestão fiscal | Referências: <ul style="list-style-type: none">- Sistema Firjan (2017a);- Soares e Melo (2016);- Veloso et al. (2011)- Gobetti e Klering (2007);- Bueno (2018); Puttomatti (2013); Stokes et al. (2013)- Baião, Cunha e Souza (2017); Brollo e Nannicini (2012); Bueno (2018);- Costa e Castelar (2015);- Mattos, Rocha e Arvate (2011); Meireles (2019);- Nunes (2015); Prado (2001); Puttomatti (2013);- Sakurai (2013); Soares e Melo (2016);- Alesina (1987); Hibbs Jr. (1977); Nordhaus (1975);- Rogoff e Sibert (1988);- Sakurai (2009); Sakurai e Gremaud (2007)- Batista (2015);- Cingolani (2013);- Cingolani, Thomsson e Crombrughe (2015);- Hou (2007); Ingraham (2007); Ingraham, Joyce e Donahue (2003); Marengo, Strohschoen e Joner (2017); Veloso et al. (2011) |

Nota: * Por exemplo, transferências voluntárias do estado ou da União para entidades privadas da assistência social e investimentos diretos dessas esferas em infraestrutura de transportes.

Fonte: elaboração própria.

Efeitos fixos: exemplo

O modelo empírico (Equação I)¹⁰ estrutura-se em torno de variações espaciais e temporais, podendo ser representado pelos mesmos quadrantes do Quadro 4. Tem o IFGF como variável dependente, e combinações município-ano (i - t) como unidade de análise.

$$y_{it} = \alpha + \beta C_{it,(t-1)} + a_i + b_t + \varepsilon_{it} \quad (I)$$

Fatores institucionais, incidentes sobre todos os municípios e invariantes na janela temporal considerada (2007-2016), não requerem operacionalização, justamente por não terem se alterado. Demais fatores estruturais, com diferenciação entre localidades, são captados via modelo para dados em painel, que considera o agrupamento das observações (tomadas ao longo do tempo) em municípios aos quais elas pertencem. O teste de Hausman indica que um modelo com efeitos fixos de município (a_i) é preferível em relação à especificação com interceptos aleatórios por município (p -valor = 0,0000). No modelo com efeitos fixos de município, um conjunto de *dummies* identificando cada município controla pelas diferenças entre essas unidades, desde que tais diferenças tenham permanecido constantes de 2007 a 2016, a exemplo do ano de criação e da condição de capital.

Condicionantes do desempenho fiscal que se modificam no tempo, porém atingindo igualmente todos os municípios, são operacionalizados por efeitos fixos de ano (b_t). Exemplos de fatores dessa natureza são a situação fiscal da União e do Estado de Minas Gerais; também, especificamente em 2016, as receitas municipais aumentaram perto de 4%, em média, como resultado da Lei de Repatriação (Lei Federal nº 13.254/2016, sobre regularização de recursos de origem lícita), já que IR e multas apuradas na repatriação foram distribuídos aos municípios seguindo o esquema de repartição do FPM (SISTEMA FIRJAN, 2017a).

IDB WORKING PAPER SERIES N° IDB-WP-679

Do Longer School Days Improve Student Achievement? Evidence from Colombia

Diana Hincapié

Inter-American Development Bank
Education Division

March 2016

Efeitos fixos: exemplo

Efeitos fixos: exemplo

Resumo: Este artigo analisa o impacto de jornadas escolares mais longas sobre o desempenho dos alunos na Colômbia. Para identificar esse impacto, o estudo explora variações plausivelmente exógenas na duração da jornada escolar dentro das próprias escolas. Utilizando as notas obtidas pelos alunos em um teste aplicado ao 5º e 9º anos nos anos de 2002, 2005 e 2009, juntamente com dados administrativos escolares, esta pesquisa emprega modelos com efeitos fixos por escola para estimar variações na média das notas entre diferentes coortes de escolas para escolas que passaram de meio turno para turno integral, ou vice-versa. Os resultados indicam que os alunos expostos ao turno integral apresentam notas aproximadamente um décimo de desvio padrão superiores às daqueles que frequentaram apenas meio turno. O impacto do turno integral é maior em matemática do que em linguagem, e é mais acentuado no 9º ano do que no 5º ano. Os efeitos são maiores nas escolas mais pobres e naquelas localizadas em áreas rurais. Os resultados sugerem que a ampliação da jornada escolar pode ser uma política eficaz para melhorar o desempenho dos alunos, especialmente entre os estudantes de baixa renda na Colômbia e em outros países em desenvolvimento.

Hincapie (2016; tradução nossa)
<http://dx.doi.org/10.18235/0000268>

Efeitos fixos: limitações

- Abordagem **não elimina *confounders*** que **variaram no período de tempo analisado e afetaram os grupos de maneira diferente.**
- **Remove toda a variação entre grupos** (pois esta é captada pelos efeitos fixos de entidade e ou tempo); calcula o **efeito de X considerando apenas as variações de X dentro de cada agrupamento.**
- **Diferentemente de efeitos aleatórios** (que têm suas próprias limitações), efeitos fixos **exigem dados longitudinais e não são capazes de calcular coeficientes para:**
 - Covariáveis de interesse se estas não variarem no tempo (no caso de efeitos fixos de entidade) ou
 - Covariáveis de interesse se estas variarem no tempo, mas não entre entidades (no caso de efeitos fixos de tempo). Dito de outro modo, efeitos fixos de tempo absorverão todas as variáveis que variam no tempo, mas não variam entre entidades.

De quais abordagens trataremos?

Abordagem

Desenho com
descontinuidade (RDD*)



Variável instrumental



Diferença em diferenças



Efeitos fixos

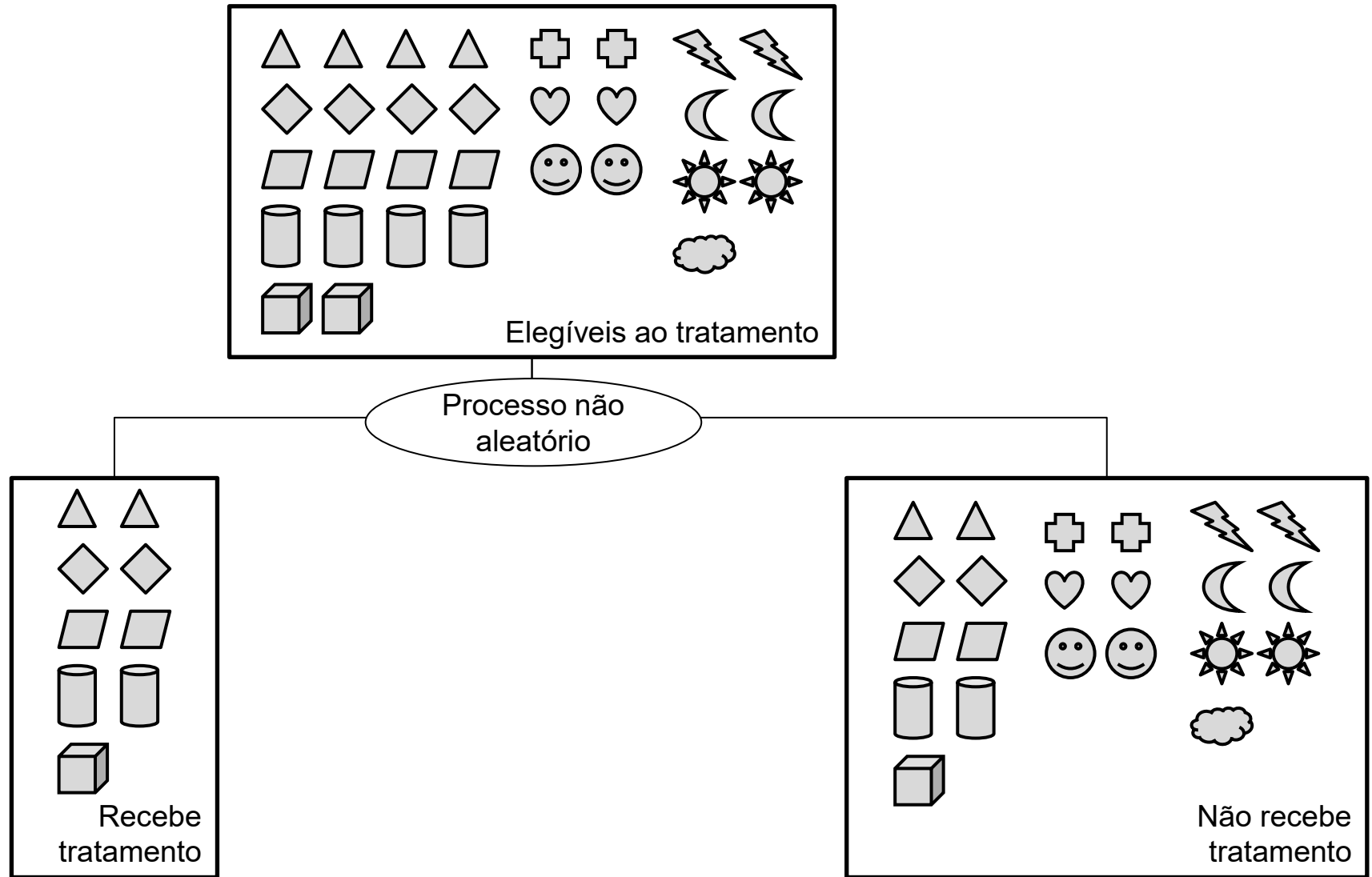


Pareamento

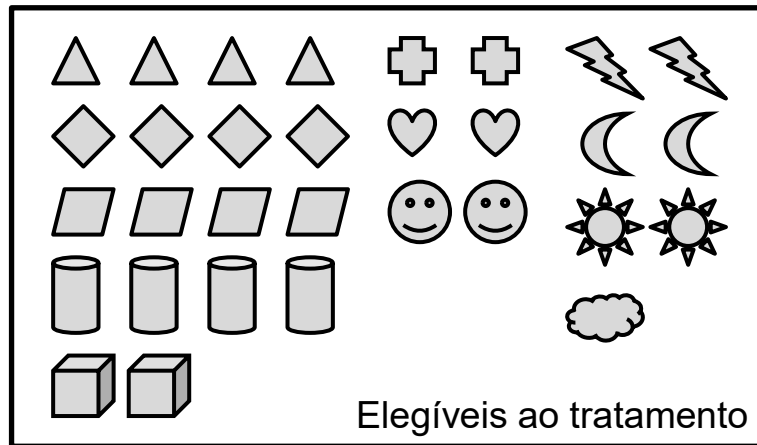
* Regression Discontinuity Design.



Pareamento: intuição ilustrada

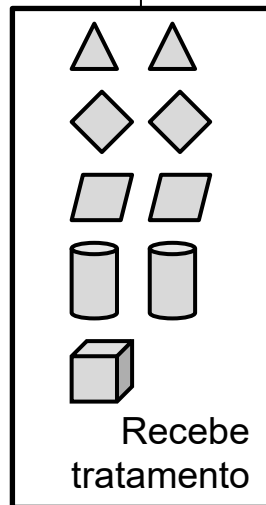


Pareamento: intuição ilustrada

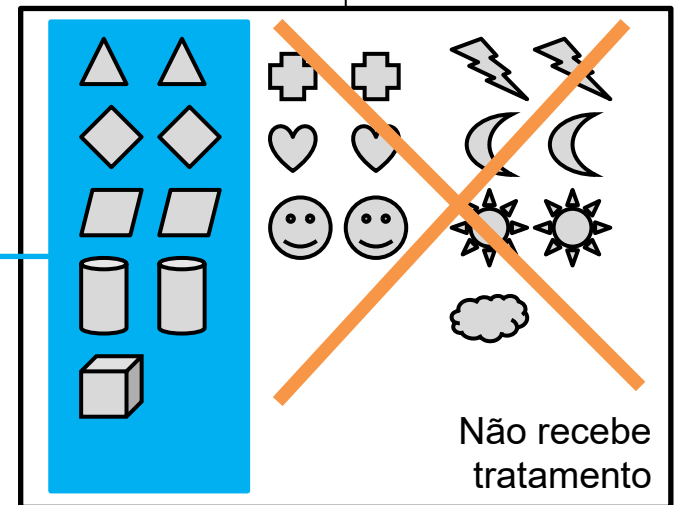


“É como se estivéssemos procurando um gêmeo não tratado para cada unidade tratada” (Facure, 2022, cap. 10; <https://matheusfacure.github.io/python-causality-handbook/landing-page.html>).

Processo não aleatório



Pareamento identifica as unidades mais similares àquelas que receberam o tratamento e forma pares (de unidade tratada com unidade não tratada). Efeito é calculado na amostra pareada.



Pareamento: intuição, assunção e requisito

Intuição

- Cria o melhor grupo de comparação possível com base nas características observáveis, identificando no conjunto de **unidades não tratadas** aquelas **que mais se parecem** com as unidades **tratadas**
- Estima o impacto do tratamento com base na **amostra pareada**

Assunção e requisito

- **ASSUNÇÃO:** **Independência condicional (a.k.a. conditional ignorability, selection on observables):** condicionada por características observáveis, a participação (i.e., recebimento do tratamento) é exógena
[A]fter controlling for X [matching variables], the treatment assignment is “as good as random”. (Heinrich; Maffioli; Vázquez, 2010, p. 16)
- **REQUISITO: Sobreposição (a.k.a. common support, positivity):** há no grupo de comparação casos com características similares às aquelas apresentadas pelos casos do grupo de tratamento

Propensity Score Matching (PSM): intuição básica

O PSM (Propensity Score Matching) utiliza informações de um grupo de unidades que não participam da intervenção para identificar o que teria acontecido com as unidades participantes na ausência da intervenção. [Note como esta descrição refere-se ao ATT.] Ao comparar como os resultados diferem para os participantes em relação a não participantes observacionalmente semelhantes, é possível estimar os efeitos da intervenção.

[...]

Uma das questões críticas na implementação de técnicas de pareamento é definir claramente (e justificar) o que significa “semelhante”. [...] Na prática, para que o processo de pareamento consiga mitigar com sucesso possíveis vieses, ele deve ser feito **considerando toda a gama de covariáveis nas quais as unidades tratadas e de comparação possam diferir**.

[...]

O escore de propensão é definido como a probabilidade de que uma unidade na amostra combinada de unidades tratadas e não tratadas receba o tratamento, dado um conjunto de variáveis observadas. Se todas as informações relevantes para a participação e os resultados [em conjunto] forem observáveis ao pesquisador, o escore de propensão (ou probabilidade de participação) produzirá pareamentos válidos para estimar o impacto de uma intervenção. Portanto, **em vez de tentar parear com base em todos os valores das variáveis, os casos podem ser comparados apenas com base nos escores de propensão**.

Heinrich, Maffioli, Vázquez (2010, p. 4)

PSM: exemplo

RESUMO

Nesta monografia realizamos a avaliação econômica do Programa Fica Vivo que é o pilar da política de prevenção e controle da criminalidade do Governo do Estado de Minas Gerais. Seu objetivo principal é a redução dos homicídios nas áreas de maior incidência, em geral favelas. A avaliação econômica envolve a apuração de dois elementos: o custo e a efetividade do programa. Os custos são apurados pelo método de contagem através das informações disponibilizadas pela Secretaria Estadual de Defesa Social e pela Polícia Militar de Minas Gerais (PMMG). A efetividade é mensurada pela metodologia Diferenças em Diferenças com Pareamento (*Double Difference Matching*) com base nas ocorrências georeferenciadas registradas pela PMMG e no Censo Demográfico 2000. Consideramos como variável de impacto do programa a taxa de homicídio por cem mil habitantes. Esta metodologia permite a mensuração da qualidade do investimento público através de dois indicadores de eficiência: razões custo-efetividade e custo-benefício. Neste sentido a presente monografia contribui à modernização da gestão governamental através da aplicação do método de avaliação econômica de projetos públicos baseado em registros administrativos e dados oficiais, possibilitando a sua replicação. Além disto, contribui ao desenvolver um método de análise de política na área de segurança pública que é carente de embasamento empírico. Os resultados mostram que o custo de um homicídio evitado pelo programa é de aproximadamente 244,6 mil reais o que implica em uma taxa de retorno do programa de aproximadamente 99%. A comparação destes resultados com avaliações internacionais de programas similares evidencia que o Fica Vivo tem um elevado retorno.

Palavras-chave: Qualidade do Gasto Público, Indicador de Eficiência, Avaliação Econômica do Fica Vivo.



1º Lugar - Qualidade do Gasto Público

Autora: **Betânia Totino Peixoto**
Belo Horizonte/MG

"Avaliação Econômica do Programa Fica Vivo: o caso piloto"

Em agosto de 2002, o programa Fica Vivo foi implantado na área piloto, favela denominada "Morro das Pedras". A escolha desta área como a primeira a receber o programa decorreu do fato desta ser, das seis áreas apontadas pelo diagnóstico, a que exibia maior taxa de homicídio por cem mil habitantes e elevado índice de vulnerabilidade social. Embora esses critérios tenham sido os mais relevantes há que se mencionar que essa área apresentava maior presença de aparelhos públicos locais e iniciativas privadas voltadas para a proteção social (SILVEIRA, 2007). Esse ambiente facilitava a implantação do programa.

Peixoto (2008, p. 27)

PSM: exemplo

O PAREAMENTO

Primeiramente, estimamos a probabilidade de participação no programa dos setores censitários. Esta estimação é realizada através do modelo PROBIT condicionado às características socioeconômicas, demográficas e às taxas de homicídio por cem mil habitantes dos setores censitários antes do programa²⁶. As variáveis utilizadas são ortogonais ao programa. Em seguida o grupo controle é selecionado pela metodologia de Pareamento por Vizinho mais Próximo (*Nearest Neighbor Matching*)²⁷.

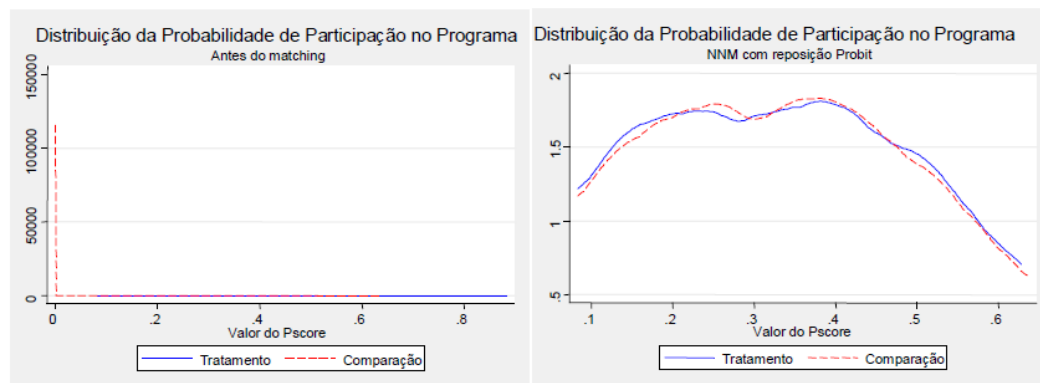
Na Figura 1, mostramos os gráficos da função de densidade da probabilidade de participação no programa para os grupos de tratamento e de comparação, antes e depois do pareamento. Antes do pareamento, a probabilidade de participação, de grande parte do grupo de comparação, está concentrada próxima à zero. Após o pareamento a probabilidade de participação do grupo de comparação passa a ter

²⁶ O modelo Probit estimado é apresentado na tabela B do anexo.

²⁷ Realizamos o pareamento pelas demais metodologias apresentadas no capítulo 1 e os resultados foram similares.

distribuição semelhante ao grupo de controle, sendo a curva de distribuição nos dois grupos quase sobreposta.

Figura 1: Distribuição da probabilidade de participação no programa Fica Vivo, Morro das Pedras e Controle, antes do pareamento e após o Pareamento por Vizinho mais Próximo.



Na tabela 6, apresentamos as médias das covariadas utilizadas na estimação da probabilidade de participação, antes e após o pareamento, entre os grupos de tratamento e comparação. As colunas “Dif-Médias” mostram o resultado do teste de diferenças nas médias das variáveis do grupo de tratamento e comparação. Em outras palavras, indica a semelhança entre as médias das variáveis nos grupos tratamento e controle, antes e após o pareamento. Podemos observar que o pareamento tornou as médias de todas as variáveis estatisticamente iguais. Antes do pareamento, as médias das variáveis socioeconômicas do grupo de tratamento e controle são diferentes.

Peixoto (2008, p. 42)

Tratamento e Comparação antes e após o Pareamento

| Variáveis | Antes do <i>Matching</i> | | | Dif-Médias após o <i>Matching</i> |
|-------------------------|--------------------------|-------------|------------|-----------------------------------|
| | Média Tratado | Média Comp. | Dif-Médias | |
| Txhoms1 | 26,031 | 7,647 | 18,384*** | -0,776 |
| Txhoms2 | 35,745 | 9,418 | 26,327*** | -41,939 |
| Txhoms3 | 56,377 | 9,293 | 47,084*** | -18,459 |
| Txhoms4 | 35,828 | 8,268 | 27,560*** | -2,671 |
| Txhoms5 | 47,691 | 11,137 | 36,554*** | -34,087 |
| P_1banho | 0,804 | 0,602 | 0,202*** | -0,021 |
| P_2banho | 0,083 | 0,210 | -0,127*** | 0,009 |
| P_3banho | 0,038 | 0,129 | -0,091** | 0,013 |
| P_4mbanho | 0,014 | 0,039 | -0,025 | -0,004 |
| P_lixo | 0,941 | 0,984 | -0,044*** | -0,018 |
| P_homem | 0,481 | 0,470 | 0,011** | -0,001 |
| p_09aa | 0,211 | 0,152 | 0,059*** | 0,001 |
| p_1014aa | 0,104 | 0,082 | 0,021*** | -0,002 |
| p_1519aa | 0,115 | 0,097 | 0,017*** | -0,006 |
| p_2024aa | 0,118 | 0,103 | 0,015*** | 0,002 |
| p_2529aa | 0,082 | 0,088 | -0,007* | -0,004 |
| p_30maa | 0,371 | 0,477 | -0,105*** | 0,010 |
| P_rend0 | 0,112 | 0,069 | 0,043*** | 0,007 |
| P_rend_1 | 0,252 | 0,112 | 0,140*** | 0,000 |
| P_rend1_3 | 0,438 | 0,268 | 0,170*** | -0,022 |
| P_rend3_5 | 0,100 | 0,148 | -0,048*** | -0,009 |
| P_rend5_10 | 0,047 | 0,188 | -0,141*** | 0,010 |
| População no semestre 1 | 781,190 | 879,410 | -98,220* | 5,180 |
| População no semestre 2 | 789,610 | 878,200 | -88,590 | 6,280 |
| População no semestre 3 | 798,110 | 877,460 | -79,350 | 7,380 |
| População no semestre 4 | 807,270 | 887,250 | -79,980 | 8,140 |
| População no semestre 5 | 816,540 | 898,440 | -81,900 | 8,890 |

Nota: ***significativa a 1%, **significativa a 5%, *significativa 10%.

Peixoto (2008, p. 43)

AVALIAÇÃO DA POLÍTICA DE COTAS NA UNIVERSIDADE FEDERAL DE VIÇOSA

Felipe Miranda de Souza Almeida¹

Cristiana Tristão Rodrigues²

Este trabalho visa avaliar a política de cotas na Universidade Federal de Viçosa (UFV), comparando o desempenho escolar e o número de reprovações entre os estudantes beneficiados ou não pela Lei de Cotas nos anos 2013, 2014 e 2015. Os resultados, encontrados por meio da metodologia *propensity score*, mostram que não há diferença significativa entre os dois grupos. Assim, o resultado vai ao encontro do objetivo da lei, que é ampliar o acesso às universidades e aos institutos federais para os jovens das escolas públicas, para os pretos e pardos e para os índios, sem diminuir a qualidade de ensino.

Palavras-chave: políticas de cotas; ensino superior; *propensity score*.

planejamento e políticas públicas | ppp | n. 53 | jul./dez. 2019

www.ipea.gov.br/ppp/index.php/PPP/article/view/868

Apresente múltiplas estimativas: exemplo

Diferentes critérios de *matching* podem ser usados para atribuir participantes a não participantes em função do PS. Isto implica calcular um peso para cada conjunto de participante e não participante. Como discutido a seguir, a escolha de uma técnica de correspondência particular pode, por conseguinte, afetar a estimativa resultante, por meio dos pesos atribuídos (Khandker, Koolwal e Samad, 2010).

- 1) Vizinho mais próximo (NN): é uma das técnicas de pareamento mais utilizadas, em que cada unidade de tratamento é comparada com a unidade de comparação com o PS mais próximo. Pode-se também escolher vizinhos mais próximos e fazer n pareamentos (normalmente é utilizado $n = 5$). O *matching* pode ser feito com ou sem substituição. Com a substituição, por exemplo, significa que o mesmo não participante pode ser utilizado como correspondente para diferentes participantes.
- 2) Calibre e radial: um problema com o *matching* NN é que a diferença no PS para um participante e o seu vizinho mais próximo não participante da política ainda pode ser muito alta. Esta situação resulta em comparações pobres e pode ser evitada mediante imposição de um limite ou “tolerância” na distância PS máxima (calibre). Este procedimento envolve pareamento com substituição apenas entre os PS dentro de um determinado intervalo. Uma maior redução de indivíduos não participantes provavelmente aumenta a chance de viés de amostragem.
- 3) Estratificação e intervalo: este procedimento particiona o suporte comum em diferentes estratos (ou intervalos) e calcula o impacto do programa em cada intervalo. Especificamente, dentro de cada intervalo, o efeito do programa é a diferença média nos resultados entre observações de tratados e de controle. A média ponderada destas estimativas de impacto produz o impacto global do programa, tendo a porcentagem de participantes em cada intervalo como os pesos.
- 4) Kernel e local linear: os algoritmos de pareamento discutidos até agora têm em comum que apenas algumas observações, a partir do grupo de comparação, são usadas para construir o resultado contrafactual de um indivíduo tratado. *Matching* Kernel (KM) e *matching* linear local (LLM) são estimadores não paramétricos de pareamento que utilizam ponderação das médias de todos os indivíduos no grupo de controle para construir o resultado contrafactual. Assim, uma vantagem principal destas abordagens é a mais baixa variância, devido à utilização de mais informação. Uma desvantagem é que, possivelmente, as observações são comparações ruins. Assim, a instituição adequada da condição de suporte comum é de grande importância para KM e LLM.

Avaliação da Política de Cotas na Universidade Federal de Viçosa

371

Na literatura que aborda as diferentes técnicas de pareamento, não há um consenso sobre qual técnica apresentaria melhores resultados, mas argumenta que a consideração conjunta das técnicas pode oferecer um método para avaliar se as estimativas são robustas. Seguindo esta linha, neste trabalho, serão utilizadas as quatro técnicas, visando uma estimativa robusta do possível efeito dos tratamentos.

Assim, o impacto da Lei nº 12.71/2012 sobre o rendimento acadêmico e o número de reprovações dos estudantes que ingressaram pelo sistema de cotas (ATT) foi estimado a partir da comparação entre os que ingressaram na universidade por meio da política e os que ingressaram pelo sistema de ampla concorrência, selecionados por suas características observáveis a partir da estimação do *propensity score* e pareados pelos algoritmos de vizinho mais próximo, radial, estratificação e de Kernel.

Almeida e Rodrigues (2019, p. 370-371)

www.ipea.gov.br/ppp/index.php/PPP/article/view/868

Apresente múltiplas estimativas: exemplo

Avaliação da Política de Cotas na Universidade Federal de Viçosa

375

Nas tabelas 4 e 5, são reportadas as estimativas do valor do efeito do tratamento sobre o coeficiente de rendimento acumulado e o número de reprovações, respectivamente.

TABELA 4
Cálculo do efeito do tratamento sobre o coeficiente de rendimento acumulado (2013-2015)

| | | NN | Radial | Estratificado | Kernel |
|------|-------------|--------|--------|---------------|--------|
| 2013 | ATT | 2,164 | 3,085 | 3,501 | 3,336 |
| | Erro-padrão | 1,969 | 1,399 | 1,566 | 1,896 |
| | T | 1,099 | 2,205 | 2,236 | 1,760 |
| 2014 | ATT | -1,291 | -2,629 | -1,005 | -1,91 |
| | Erro-padrão | 3,470 | 2,348 | 5,341 | 2,336 |
| | T | -0,372 | -1,120 | -0,188 | -0,817 |
| 2015 | ATT | -2,324 | -0,603 | -2,118 | -1,763 |
| | Erro-padrão | 3,100 | 1,948 | 1,826 | 2,160 |
| | T | -0,750 | -0,310 | -1,160 | -0,816 |

Elaboração dos autores.

Analisando os resultados da tabela 4, temos que, em todos os modelos, o efeito do tratamento sobre a variável rendimento acadêmico acumulado não foi significativo, indicando que não há diferença entre o desempenho acadêmico dos estudantes que ingressaram pelo sistema de cotas e pelo sistema de ampla concorrência.

Os resultados da tabela 5 são similares aos encontrados anteriormente, e são, de certa maneira, implícitos, dado que o cálculo do CRA leva em consideração a nota obtida nas disciplinas e possui relação positiva. Para um aluno ser reprovado, ele deve possuir uma nota inferior ou igual a 59,9 pontos na disciplina. Logo, uma menor nota implica um menor coeficiente de rendimento.

TABELA 5
Cálculo do efeito do tratamento sobre o número de reprovações – NReprovacoes (2013-2015)

| | | NN | Radial | Estratificado | Kernel |
|------|-------------|--------|--------|---------------|--------|
| 2013 | ATT | -0,381 | -0,468 | -0,4559 | -0,438 |
| | Erro-padrão | 0,459 | 0,326 | 0,351 | 0,379 |
| | T | -0,829 | -1,433 | -1,311 | -1,154 |
| 2014 | ATT | -0,317 | -0,020 | -0,086 | 0,002 |
| | Erro-padrão | 0,844 | 0,508 | 0,747 | 0,521 |
| | T | -0,375 | -0,040 | 0,521 | 0,003 |
| 2015 | ATT | -0,481 | 0,082 | 0,148 | 0,175 |
| | Erro-padrão | 0,353 | 0,231 | 0,244 | 0,230 |
| | T | 1,364 | 0,354 | 0,606 | 0,758 |

Elaboração dos autores.

Almeida e Rodrigues (2019, p. 375)

www.ipea.gov.br/ppp/index.php/PPP/article/view/868

Pareamento: limitações

- Somente reduz o viés causado por **características observáveis**: se o status do tratamento for influenciado por características não observáveis (quebra do pressuposto de independência condicional) e que não sejam altamente correlacionadas com as variáveis usadas no pareamento, os impactos **estimados serão enviesados (não por conta de viés e pareamento, mas por conta de viés de seleção na exposição ao tratamento)**
- Compromete validade interna se falta de área de suporte comum causar **descarte de um grupo de observações sistematicamente diferentes** daquelas retidas
- Requer uma grande **quantidade de variáveis de pareamento e de observações**

... quando os dados de linha de base [pré-teste] estão disponíveis, o pareamento baseado nas características socioeconômicas da linha de base pode ser muito útil se for combinado a outras técnicas, como diferença em diferenças, o que nos permitirá corrigir pelas diferenças entre os grupos que são fixas ao longo do tempo. O pareamento também é mais confiável quando a regra de seleção do programa e as variáveis subjacentes são conhecidas e, nesse caso, o pareamento pode ser realizado nessas variáveis.

Gertler et al. (2018, p. 1694)

considerações finais

Elegância e adequação do desenho

Segundo a visão (um tanto desencantada) de Jannuzzi (2016, p. 99), a perspectiva “randomista” incorre em erro ao defender que os métodos experimentais são os mais adequados e legítimos cientificamente para a avaliação de programas:

Afinal, apesar das advertências sobre as dificuldades de replicação das condições de controle laboratorial no contexto de operação dos programas sociais, a elegância e precisão do método experimental levaram a maioria dos avaliadores de programas a vê-lo como ideal [...]. Entretanto, uma vez que elegância é um critério de caráter estético, configura-se, antes, como um valor humano que como um atributo de legitimação científica.

Posição epistêmica de Jannuzzi (2016, 101):

“Não existe um método ou estratégia padrão-ouro para a produção de uma avaliação”

- *Ao contrário do que defendem os ‘quantitativistas’ e ‘randomistas’, avaliações conduzidas mediante **métodos quantitativos** ou, em particular, delineamentos quase experimentais **não dispõem de maior status de credibilidade e cientificidade que outras avaliações, com foco qualitativo** (Jannuzzi, 2016, 101)*
- **Abordagens quantitativas e qualitativas** não são mutuamente excludentes, mas **complementares**; a pesquisa qualitativa pode contribuir para a realização de um levantamento quantitativo posterior; achados quantitativos “difíceis de interpretar” podem ser investigados via pesquisa qualitativa
- Da mesma forma, **não existe um padrão-ouro para a realização de avaliações por grupos externos, especialistas ou pesquisadores acadêmicos**

Observar é interpretar:

*A **realidade social** não é passível de ser entendida de forma neutra ou objetiva com base nos dados empíricos coletados, mas, sim, a partir de **conhecimento anterior e da visão de mundo que se tem dela**. A observação não é, portanto, puramente passiva, pois **para observar é preciso sempre relacionar aquilo que se vê com noções que já se possuía anteriormente**. (Jannuzzi, 2016, 101)*

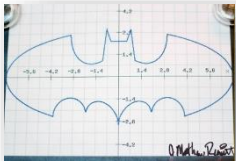
O melhor método é aquele que apresenta tempestividade e precisão compatíveis com a demanda em mãos



- Não adianta responder rápido e errado
- Não adianta responder precisa e acuradamente, porém após o momento de decisão



- **Triangular**, ou seja, abordar o objeto de pesquisa com técnicas diferentes de investigação (ou sujeitos entrevistados) é uma boa ideia, sempre que possível
- **Junto com uma perspectiva multidisciplinar e esforços combinados de avaliação interna e externa**, a triangulação garante **credibilidade e firmeza** para o aprimoramento da gestão e desenho das intervenções programáticas



- **Decisões metodológicas** quanto ao desenho e métodos de pesquisa, o tamanho e a intencionalidade da amostra das pesquisas de avaliação são **função de**:
 - Objetivos da avaliação/ perguntas que se quer responder
 - Natureza do programa/ complexidade do objeto investigado
 - Estágio em que o mesmo se encontra
 - Disponibilidade de tempo e outros recursos



- **A estratégia avaliativa deve ajustar-se** à complexidade do problema a ser investigado, e não o contrário
- **O problema de avaliação preside** a escolha dos métodos, sujeitos, e instituições a analisar

I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail – frase de autoria incerta, tendo sido atribuída a Mark Twain (1835-1910, do Tom Sawyer) e Abraham Maslow (1908-1970, da pirâmide).

Experimentos e quase experimentos

Avaliação de Políticas Públicas A (DCP131)

04 e 09 de junho de 2025

Nota: este material é bem mais completo do que as nossas discussões em sala de aula.

Leitura básica:

JANNUZZI, Paulo de Martino. **Monitoramento e avaliação de programas sociais**. Campinas, SP: Editora Alínea, 2016. pp. 89-100.

RAMOS, Marília Patta. Aspectos conceituais e metodológicos da avaliação de políticas e programas sociais. **Planejamento e Políticas Públicas**, n. 32, jan.-jun./2009, p. 95-114. Disponível em: <https://www.ipea.gov.br/ppp/index.php/PPP/article/view/11>

Leitura complementar:

GERTLER, Paul J. et al. **Avaliação de impacto na prática**. Washington: Banco Interamericano de Desenvolvimento; Banco Mundial, 2018. [Capítulo 3: Inferência causal e contrafactuais. Capítulo 4: Seleção aleatória. Capítulos 5-8, sobre abordagens quase experimentais, devem ser lidos conforme o interesse do/a discente.] Disponível em:

<https://openknowledge.worldbank.org/bitstream/handle/10986/25030/9781464808890.pdf>

extras

Circunstâncias, tais como seleção, podem afetar o tratamento e complicar a avaliação

AMEAÇAS À VALIDADE INTERNA DE UMA AVALIAÇÃO: Exemplos de fatores que podem afetar os resultados da aplicação de um tratamento (intervenção) e atrapalhar a avaliação:

- **Seleção:** Diferença sistemática entre o conjunto das unidades (p.ex., indivíduos) exposto ao tratamento e o conjunto de comparação
- **História:** Exposição a outras experiências durante o tratamento
- **Mortalidade:** Evasão de sujeitos durante o tratamento
- **Maturação:** Mudanças que ocorrem naturalmente com a passagem do tempo, como envelhecimento, cansaço
- **Teste:** Contaminação de respostas futuras pelo pré-teste
- **Instrumentação:** Mudança de procedimentos de medição
- **Regressão em direção à média:** Tendência de abrandamento de escores extremos (muito altos ou muito baixos)

*Idealmente falando, se queremos apontar causas, o mais correto seria que pudéssemos comparar o mesmo grupo (ou sujeitos, instituições, regiões) fazendo parte de uma política/ programa social e ele mesmo sem ter feito parte. Isto é, o ideal seria ter na pesquisa o sujeito atingido e seu **clone exato**. Como isso não é possível, um dos pontos mais delicados do desenho de avaliação diz respeito à tentativa de encontrarmos o grupo comparativo ou, como se chama na literatura de avaliação, o **contrafactual**.*

Ramos (2009, p. 99)

Randomized experiments are tricky

- Pessoas podem se **recusar a participar**
- Disponibilização do tratamento e controle do ambiente podem ser **custosos ou impraticáveis**; é arriscado afirmar que os efeitos medidos sejam apenas do tratamento em foco, em meio a um **contexto crescente de intervenções sociais concomitantes**
- Algumas variáveis explicativas **não podem ser ‘aplicadas’ como tratamento**; imagine perguntas de pesquisa envolvendo a relação entre corrupção e conflito civil, ou entre taxas de natalidade e crimes violentos; experimentos **não se adequam à pesquisa histórica**
- Experimentos podem ser **antiéticos** (e.g., deixar de disponibilizar um serviço social reputado como efetivo para poder rigorosamente estimar seus efeitos)
- Pode haver **evasão (attrition)** dos casos analisados, potencialmente gerando **desbalanceamento** dos grupos (grupos de tratamento e controle não são similares nos observáveis)
- Pode haver **noncompliance** com a atribuição ao tratamento: unidades designadas para o grupo de tratamento não o recebem, ou vice-versa
- **Especificidade e circunstâncias “artificiais” conspiram contra a generalização** dos resultados (validade externa do desenho); possivelmente, **pessoas dispostas a participar** de um experimento são, de alguma forma, diferentes da população em geral; pessoas em **diferentes comunidades** poderiam responder diferentemente ao tratamento, ou o mesmo tratamento funcionaria de forma diferente se aplicado em **larga escala**, ou em **outro momento no tempo**