





ESTADO DO RIO GRANDE DO SUL

Secretaria de Planejamento, Governança e Gestão
Departamento de Planejamento Governamental

ESTUDOS DEPLAN

Edição

Nº 10/2018

Avaliação de Impacto de Políticas Públicas
Conceitos, Metodologias e Experiências

Junho/2018



GOVERNO DO ESTADO DO RIO GRANDE DO SUL

Governador: José Ivo Sartori

Vice-Governador: José Paulo Dornelles Cairoli

SECRETARIA DE PLANEJAMENTO, GOVERNANÇA E GESTÃO

Secretário: Josué de Souza Barbosa

Secretário Adjunto: Melissa Guagnini Hoffmann Custódio

DEPARTAMENTO DE PLANEJAMENTO GOVERNAMENTAL (DEPLAN)

Diretor: Antonio Paulo Cargnin

Diretora Adjunta: Carla Giane Soares da Cunha

EQUIPE EDITORIAL

Antonio Paulo Cargnin

Juliana Feliciati Hoffmann

FICHA TÉCNICA:

Juliana Feliciati Hoffmann e Ana Júlia Possamai (Coordenação e elaboração)
César Stallbaum Conceição, Fernanda Rodrigues Vargas, Gisele da Silva
Ferreira, Juliana Feliciati Hoffmann, Sílvia Letícia Lorenzetti, Rayssa Miczewski
de Araújo, Vinícius Dias Fantinel (elaboração)

REVISÃO E TRADUÇÃO: Marlise Margô Henrich

CAPA: Laurie Fofonka Cunha

Estudos DEPLAN / Departamento de Planejamento Governamental - RS.
Porto Alegre : Secretaria de Planejamento, Governança e Gestão, 2015-

v. : il.

Semestral.

Título especial a cada edição, de acordo com o assunto predominante.
Publicado pela Secretaria de Planejamento, Planejamento, Governança
e Gestão, 2015-

ISSN 2447-4576

1. Desenvolvimento regional – Periódico – Rio Grande do Sul. I. Rio
Grande do Sul. Secretaria de Planejamento e Gestão. Departamento de
Planejamento Governamental.

CDU 332.1(816.5)(05)

Bibliotecário responsável: João Vítor Ditter Wallauer – CRB 10/2016

AVALIAÇÃO DE IMPACTO DE POLÍTICAS PÚBLICAS: MÉTODOS NÃO-EXPERIMENTAIS*

Beatrice Zimmermann¹

Resumo: O texto apresenta uma introdução aos métodos não-experimentais de avaliação de políticas públicas. Para tanto, ressalta-se a importância do desenho de um bom contrafactual, a fim de produzir evidências robustas a partir da avaliação. Qualquer grupo de comparação não é, necessariamente, um bom grupo de comparação, sendo necessário tomar cuidados para construí-lo. Nesse sentido, dois métodos não-experimentais serão apresentados: Diferenças-em-Diferenças (DD) e Regressão Descontinuada (RDD), sendo discutidas suas características e vantagens em relação aos métodos chamados “ingênuos” de definição do contrafactual.

Palavras-chave: Políticas públicas. Avaliação de impacto. Métodos não-experimentais. Contrafactual.

Abstract: The text presents an introduction to the non-experimental methods of public policy evaluation. The importance of designing a good counterfactual is emphasized, in order to produce robust evidences. It is argued that any comparison group is not necessarily a good comparison group, so it is required attention being in order to build it. In this sense, two non-experimental methods are presented: Differences-in-Differences (DD) and Regression Discontinuity (RD). It is discussed their characteristics and advantages in contrast to the so-called *naive* methods of defining the counterfactual.

Keywords: Public policies. Impact evaluation. Non-experimental methods. Counterfactual.

* Texto elaborado a partir da transcrição da fala e da apresentação em slides de palestras proferida no Workshop Avaliação de Impacto de Políticas Públicas, realizado nos dias 29 e 30 de novembro de 2017, no Auditório da Secretaria de Planejamento, Governança e Gestão (SPGG-RS), com o apoio do Banco Mundial.

¹ Coordenadora de Campo do *Development Impact Evaluation* (DIME), Banco Mundial.

INTRODUÇÃO

O presente artigo tem como objetivo fazer uma introdução aos métodos não-experimentais de avaliação de políticas públicas. Os métodos não-experimentais, apesar de não constituírem a primeira e mais adequada das opções em termos de avaliação de impacto, precisam ser conhecidos e dominados por avaliadores de políticas públicas. Considerando que nem sempre é possível planejar um programa antes de ele ser executado, é altamente provável que o analista de política, ao longo de sua carreira, precisará trabalhar com dados gerados *a priori*, quando o projeto objeto da avaliação já ocorreu. Além disso, algumas vezes há a necessidade de se utilizar métodos não-experimentais mesmo quando um experimento foi construído, pois alguns fatores ao longo do processo (implementação, desenho, etc.) podem fazer com que a comparação simples entre os grupos de tratamento e controle não gere resultados precisos. Nesses casos, métodos estatísticos não-experimentais são essenciais para contornar essas questões.

O texto a seguir se organiza da seguinte maneira: a primeira parte expõe a lição que fundamenta a necessidade de métodos robustos de avaliação de impacto; em seguida, discute-se a importância da situação contrafactual para avaliações de impacto e se examinam estratégias (boas e ruins) para construir grupos de comparação. Por fim, detalha-se dois métodos não-experimentais – Diferenças-em-Diferenças (DD) e Regressão Descontinuada (RDD), discutindo suas características e vantagens em relação ao que se chama de métodos “ingênuos” de construção do grupo de comparação (situação contrafactual).

AVALIAÇÃO DE IMPACTO: FATO CONCRETO – CORRELAÇÃO NÃO É CAUSALIDADE

A primeira lição fundamental aprendida por um analista de políticas públicas é que correlação não é causalidade. Como analista é preciso ter claro que aconselhar a formulação de políticas observando apenas relações não bem identificadas entre variáveis, considerando, por exemplo, correlações pouco estudadas entre elas, não é a melhor estratégia.

Para exemplificar essa questão, a Figura 1 expõe uma situação didática sobre a temática. Na figura é apresentada a evolução de duas variáveis, consumo de sorvetes e índice de ataques de tubarão, ao longo dos meses do ano. Vê-se, claramente, que essas duas variáveis possuem trajetórias bastante similares, o que nos indica que são duas variáveis correlacionadas. Entretanto, ao contrário do apresentado na mensagem de *warning*, não é possível afirmar que reduzir a quantidade de sorvetes vendida reduz o índice de ataques de tubarões. Tal análise não aponta nenhum sentido causal entre as variáveis.

Figura 1 – Correlação entre ataques de tubarão e vendas de sorvete



Fonte: Pinterest - <https://tinyurl.com/ycf3oofm>

É nesse sentido que técnicas de avaliação de políticas públicas são necessárias. O objetivo de uma avaliação de impacto é medir o efeito causal de intervenções sobre as variáveis de interesse. Para isso, o grande desafio é construir um excelente contrafactual, seja ele por método experimental ou por método não-experimental. O grupo de controle precisa ser bem construído para que mimitize o que teria acontecido com o grupo de tratados caso eles não tivessem recebido o tratamento (situação contrafactual). Precisamos comparar os resultados de grupos que possam efetivamente ser comparados (idênticos), de maneira que possamos argumentar que a única diferença entre eles é o tratamento.

Para se construir um bom contrafactual existem dois métodos: o experimental e o não-experimental. O método experimental constitui a melhor estratégia de avaliação, garantindo maior confiabilidade aos resultados ao controlar² como é definido o contrafactual. Mais especificamente, nele a definição dos grupos de tratamento e controle acontece através da utilização de algum método de aleatorização. É justamente essa capacidade que se tem de aleatorizar de alguma maneira (sorteio, ordem de chegada, etc.) que garante a comparabilidade entre os grupos e a identificação do efeito causal no método experimental.

Já no caso dos métodos não-experimentais, é necessário utilizar modelos estatísticos (sustentados em hipóteses) para poder construir um grupo de comparação que possa ser sustentado como uma boa aproximação da situação contrafactual. Dois desses métodos serão discutidos em mais detalhes na próxima seção.

² Essa capacidade de controlar como a situação contrafactual é definida faz com esse método seja conhecido pela alcunha de experimento controlado.

CONSTRUINDO UM BOM CONTRAFACTUAL

Como exposto na seção anterior, o grande desafio das análises de avaliação de impacto é a definição do contrafactual. Em termos conceituais, uma situação contrafactual representa aquilo que teria acontecido com a unidade de avaliação (seja ela pessoa, escola, município, etc.) na situação de ela não ter sido afetada pelo tratamento (projeto ou ação que está sendo avaliado). Entretanto, na prática, esse cenário comparativo perfeito é impossível de acontecer, considerando que não temos controle sobre o tempo, isto é, a habilidade de ir e voltar temporalmente para analisar diferentes possibilidades para a mesma unidade de avaliação. Diante dessa impossibilidade, o que analistas de política fazem para identificar ou construir um bom grupo de comparação é utilizar teoremas e modelos estatísticos.

No caso experimental, a aleatorização define os grupos de tratamento e controle (situação contrafactual) de forma que, na média, as características observáveis e não-observáveis dos grupos sejam idênticas. Com isso, ao se comparar as variáveis de resultado de ambos os grupos, estará se inferindo unicamente o efeito do programa ao qual o grupo de tratamento foi exposto, dado que a única diferença entre eles nesse cenário é o tratamento. Em outras palavras, pode-se afirmar que, no método experimental, a diferença entre as variáveis de resultado dos grupos tratado e controle é exclusivamente devida à intervenção (efeito causal).

No caso dos métodos não-experimentais, por sua vez, mesmo ao se utilizar de estratégias rigorosas para construir o contrafactual, é muito difícil sustentar o argumento de que se obteve grupos idênticos para a comparação. Nesses casos, possuem-se argumentos para defender que, na média, os grupos do estudo são similares entre si. Em comparação ao método experimental, a defesa da estratégia de identificação do efeito é mais difícil aqui, considerando que se demanda muito mais hipóteses para se estimar um efeito por um método não-experimental.

Para expor com clareza algumas das estratégias mais rigorosas não-experimentais, é importante discutir dois métodos de definição da situação contrafactual considerados ingênuos³. Tais métodos propõem as seguintes duas abordagens para a problemática:(i) comparar a situação do grupo dos tratados antes e depois da intervenção; ou então (ii) comparar a situação dos tratados com a de algum grupo não-tratado. Esses dois métodos, apesar de parecerem interessantes e adequados à questão, não são eficientes. Tais estratégias, como será discutido mais a frente, carregam vieses que impedem que se possa atribuir o efeito estimado exclusivamente ao programa avaliado.

A Figura 2 apresenta um resumo das estratégias abordadas aqui. Os quatro primeiros métodos serão discutidos em mais detalhes através de um estudo de caso na próxima seção. Os métodos ingênuos serão também abordados para se identificar onde recaem suas principais falhas.

³O termo ingênuo é utilizado no sentido de tais estratégias não fornecerem um bom contrafactual, mesmo que a princípio pareçam constituir boas abordagens. Tais procedimentos não são sequer considerados métodos de avaliação de impacto.

Figura 2 – Ferramentas para obter o contrafactual



Fonte: Elaboração própria.

MÉTODOS INGÊNUOS: QUALQUER GRUPO DE COMPARAÇÃO NÃO É, NECESSARIAMENTE, UM BOM GRUPO DE COMPARAÇÃO

Estudo de Caso

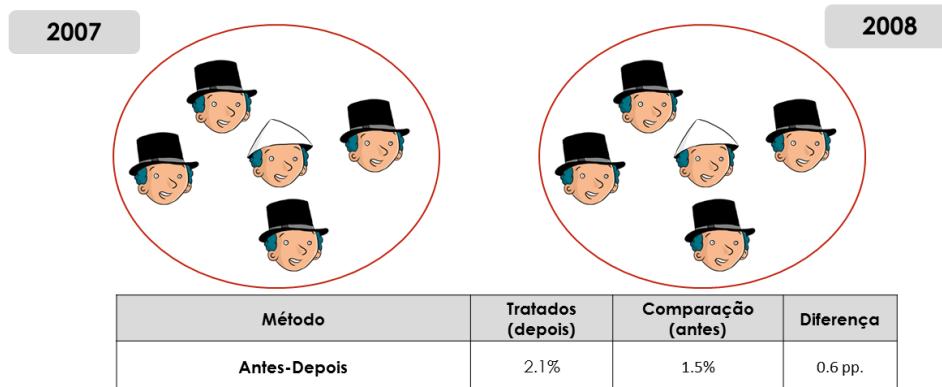
Em determinada cidade se observa que microempresas enfrentam restrições para acessar crédito. Diante disso, o gestor de política para o setor privado cria uma solução através de um programa de acesso a crédito. Por meio desse programa, o estado irá criar mecanismos para injetar capital nas microempresas que necessitem de crédito. O primeiro mecanismo criado estipula que todas as microempresas que se dirigirem à secretaria do estado e se cadastrarem no programa (preencham um formulário) terão à sua disposição R\$20.000,00 para investimento. Tal intervenção ocorreu no início de 2007. O objetivo final (variável final de interesse) desse gestor é aumentar a lucratividade dessas empresas. Após um ano da implementação da intervenção, o gestor deseja que o avaliador verifique se o programa alcançou o objetivo e estime qual foi a magnitude do impacto da intervenção.

A primeira ideia que ocorre ao analista é comparar a taxa de lucro das empresas participantes antes e depois de elas terem entrado no programa (Figura 3). Essa estratégia, à primeira vista, parece adequada para responder à pergunta do gestor, afinal as microempresas não tinham o tratamento antes e agora foram tratadas, logo comparar os dois tempos indicará o efeito do programa. Será?

Conforme mostrado na Figura 3, os indivíduos (microempresas) são os mesmos, observados no ano de 2008 e no ano de 2007 (exatamente antes da intervenção). Nessa análise, ao comparar os resultados de taxa de lucro antes e depois do programa, chega-se a um impacto de 0,6 p.p. sobre essa variável. Porém, será esse efeito resultado exclusivo da participação no programa? A resposta é não, esse não é um efeito factível e bem estimado. Temos, nesse caso, um problema conhecido como efeito tempo, isto é, inúmeras outras coisas (que não dizem respeito ao programa) podem ter afetado a variável de interesse, e o analista não as está levando em consideração. A economia, por exemplo, pode ter crescido, fazendo com que as empresas em geral tenham feito mais negócios e aumentado suas taxas de lucro.

Isso significa que o número estimado pode estar contaminado com outras coisas e não representa um bom indicativo do efeito do programa.

Figura 3 – Plano 1: comparar a taxa de lucro das empresas tratadas antes e depois do programa

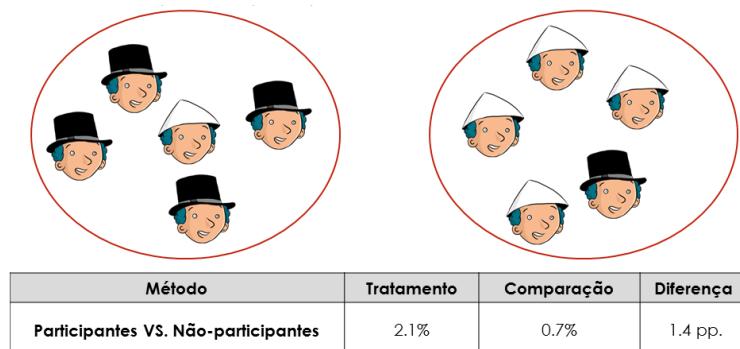


Fonte: Elaboração própria.

Sendo alertado para essa questão, o avaliador pensa, então, em outra maneira de estimar o impacto do programa. Agora ele imagina que comparar os participantes com os não-participantes seja uma forma eficiente de responder à pergunta. Ele tem à sua disposição uma base de dados que contém, além de todos os CNPJs das empresas tratadas, mais uma seleção de CNPJs de empresas que não foram tratadas. Quando aplicado esse método, comparando as médias de taxa de lucro das empresas que foram tratadas e das não que não foram, observa-se o resultado de 1,4 p.p. de diferença entre esses grupos.

Contudo, um observador cuidadoso percebe que os dois grupos comparados são bastante diferentes entre si (Figura 4). Considerando, por exemplo, que todas as características dos empresários estão resumidas nos chapéus que estão usando, na média, o grupo de tratamento teria um chapéu quase preto e o grupo que não recebeu o tratamento teria um chapéu quase branco. Isso significa que o grupo de controle não está fornecendo um bom contrafactual para o grupo de tratamento, e essa medida de impacto não é uma boa estimativa do efeito do programa. Essa situação é conhecida como problema de viés de seleção. Nesse caso, as empresas que se autosselecionaram para ser do grupo de tratamento (as características delas que as fizeram escolher ir até a secretaria se inscrever) e as que se autosselecionaram para ser do grupo controle (pois não foram até a secretaria) são muito diferentes entre si. Como por esse método essas diferenças não estão sendo levadas em consideração, a medida de efeito estimada não é acurada e não representa o impacto do programa.

Figura 4 – Plano 2: comparar a taxa de lucro das empresas que receberam crédito com a das que não receberam



Fonte: Elaboração própria.

É importante destacar que, além de características observáveis, podem existir ainda características não-observáveis que estejam afetando a variável de interesse e a decisão de ser tratado ou não. Nesse caso, por exemplo, uma característica não-observável que pode influenciar essas questões é a motivação dos empresários, indicador extremamente difícil de se medir. Nesse sentido, inclusive medidas não-observáveis podem estar impactando a estimativa de 1,4 p.p. sobre a taxa de lucro, tornando-a um resultado pouco confiável do impacto do programa.

Em resumo, os dois métodos acima tratados não conseguem estabelecer boas situações contrafactual, de forma que não são válidos para análises de avaliação de impacto. A estratégia do antes e depois incorre no problema de outras questões poderem ter acontecido ao longo do tempo, enquanto que o método de comparar participantes e não-participantes carrega o problema de viés de seleção. Ambos os métodos conduzem a estimativas viesadas do efeito do programa.

Por fim, a análise dos métodos ingênuos demonstra uma lição importante para avaliadores de política: qualquer grupo de comparação não constitui necessariamente um bom grupo de comparação. É necessário utilizar métodos mais consistentes e rigorosos para se garantir avaliações de impacto efetivas. A seguir, dois desses métodos serão discutidos.

MÉTODOS NÃO-EXPERIMENTAIS (I): DIFERENÇAS-EM-DIFERENÇAS (DD)

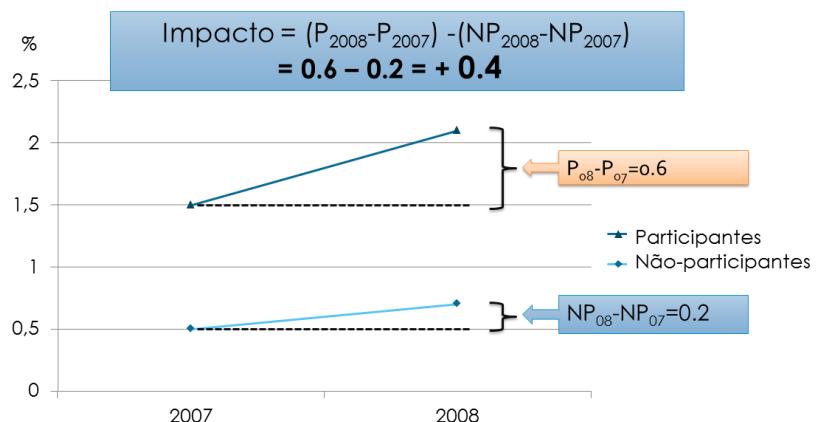
O método de Diferenças-em-Diferenças combina a dimensão temporal do antes-depois com a decisão de participação (participantes vs. não-participantes). Sobre algumas hipóteses extras, esse método consegue resolver tanto o problema do efeito tempo quanto o de viés de seleção.

Ele resolve o efeito tempo ao se assumir que o que afetou a variável de interesse ao longo do período afetou de forma “macro” ambos os grupos (tratado e não-tratado), de forma que, ao se tomar a diferença entre os grupos ($T - C$) esse efeito some. O problema de viés de

seleção é contornado por esse método ao se assumir que as variáveis (observáveis ou não) que afetam a definição dos grupos de tratamento e controle (e a variável de interesse) são constantes ao longo do tempo. Partindo dessa hipótese, quando é calculada a diferença da variável de interesse entre o tempo um (t_1) e o tempo zero (t_0) a parcela advinda dessas variáveis some, tendo em vista que elas afetam de forma idêntica a variável de interesse ao longo do tempo.

Utilizando o estudo de caso da seção anterior e empregando o método de Diferenças-em-Diferenças, o que se calcula, em resumo, são as seguintes diferenças (Figura 5): (i) a diferença da taxa de lucro dos tratados antes e depois da data do tratamento (0,6 p.p.); (ii) a diferença da taxa de lucro do grupo de controle antes e depois da data do tratamento (0,2 p.p.); (iii) a diferença dos resultados obtidos nas diferenças calculadas em (i) e (ii) (0,4 p.p.). É desse processo de cálculo que advém o nome do método.

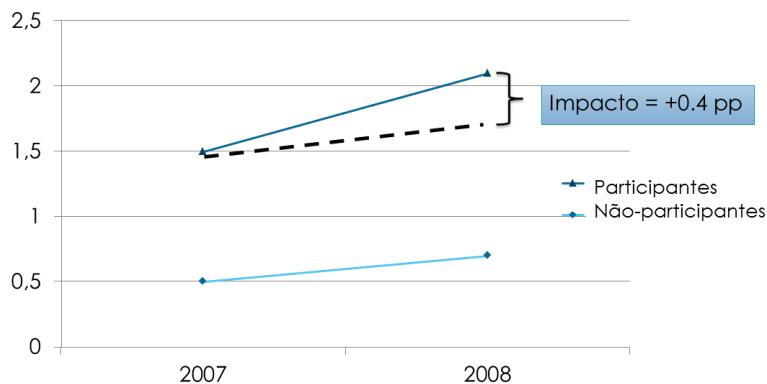
Figura 5 – Antes-depois + participantes X não-participantes: diferenças-em-diferenças



Fonte: Elaboração própria.

No entanto, além das hipóteses citadas para contornar os problemas de efeito tempo e de viés de seleção, existem ainda outras que precisam ser levadas em consideração para se usar esse método. Uma das suposições fundamentais se relaciona com o fato de, ao se aplicar o DD, estar se assumindo que os dois grupos observados são similares. Implicitamente, quando se utiliza o DD, se assume que a trajetória da variável de interesse observada para o grupo de controle seria a que se observaria para o grupo de tratamento na ausência do programa. Ou seja, assume-se que o grupo de controle representa um bom contrafactual para o grupo de tratamento. Nesse sentido, uma maneira de se entender o cálculo do efeito por esse método é imaginando que a trajetória dos não-tratados é deslocada até o nível da curva dos tratados e, então, a diferença entre essas duas trajetórias é realizada (Figura 6).

Figura 6 – Diferenças-em-Diferenças: deslocamento de trajetória

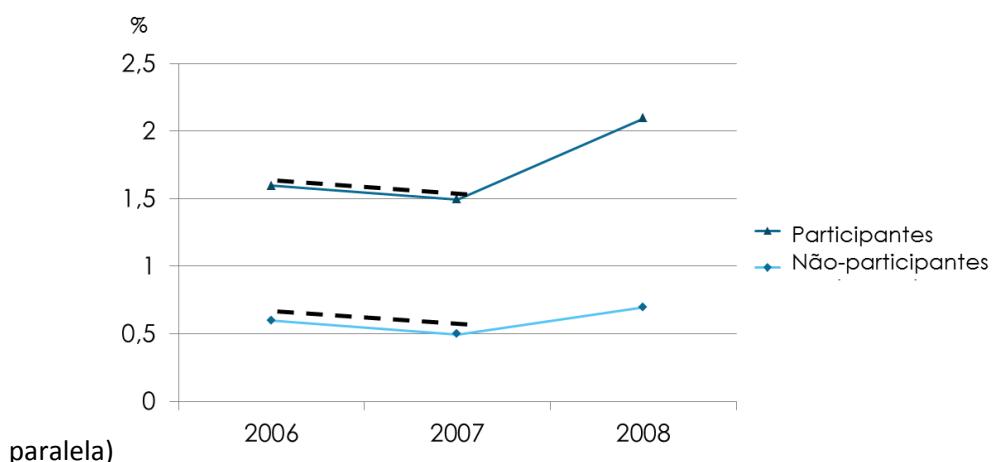


Fonte: elaboração própria.

Entretanto, para “deslocar” a trajetória do grupo de controle é necessário verificar a aderência dos dados ao que se chama de hipótese de tendência comum. Tal hipótese estabelece que o método de DD fornece uma estimativa confiável do impacto do programa se as trajetórias da variável de interesse para os grupos de tratamento e controle forem similares entre si antes da data do tratamento. Ou, de outra forma, imagina-se que, sendo os grupos bastante parecidos, a trajetória da variável de interesse fosse ser a mesma para os dois grupos caso o programa não existisse.

Para verificar a aderência das características do caso concreto a essa hipótese é fundamental ter acesso a dados históricos⁴. No exemplo (Figura 7), observa-se como eram as trajetórias de taxa de lucro tanto para participantes quanto para não-participantes antes de o programa acontecer (início de 2007). Nesse caso, as trajetórias são bem parecidas e, portanto, pode-se aplicar o método e assumir (defender) que o resultado expressará o efeito causal do programa.

Figura 7– Hipótese de tendência comum (ou



Fonte: elaboração própria.

⁴Ter acesso a dados de pelo menos um ponto anterior à intervenção é fundamental para a checagem da aderência da hipótese. Tal requerimento muitas vezes torna a aplicação do DD inviável (pouco defensável).

Por fim, cabe colocar que o método de DD pode ser ainda mais refinado se combinado com o método de *matching* ou pareamento. Esse método se baseia em um escore construído através de características observáveis para parear as observações de tratados e não-tratados, de forma a criar pares de análise similares e, portanto, comparáveis entre si. Dessa forma, a combinação do pareamento com o DD cria um grupo de controle eficiente (contrafactual defensável⁵) mesmo quando as trajetórias originais da variável de interesse não são completamente similares (hipótese de tendência comum pouco aderente).

MÉTODOS NÃO-EXPERIMENTAIS (II): REGRESSÃO DESCONTÍNUA (RDD)

O método de regressão descontínua é aplicado às situações onde a regra definidora do grupo de tratamento (e, complementarmente, do de controle) é bem estabelecida e clara. Tal regra deve respeitar três características: (i) ser baseada em uma variável contínua; (ii) sua definição ser exógena (definida sem influência lógica de fatores relacionados à problemática); e deve ser exclusiva e não-manipulável (as unidades às quais se destina o programa não podem ser capazes de manipular suas informações para se adequarem à regra de elegibilidade).

Através de tal regra de elegibilidade, o agente formulador define que as observações de um dos lados desse limiar (acima ou abaixo do valor estabelecido) serão elegíveis ao tratamento, e as do lado oposto não serão. Resumidamente, a intuição do RDD é que as observações que estão bem próximas desse limiar (acima ou abaixo dele) são bastante parecidas entre si. Sendo assim, é possível se argumentar que as observações não-tratadas à margem dessa fronteira constituem um bom contrafactual para o grupo de tratamento. Importante destacar que, no RDD, toda a análise é realizada utilizando somente as observações ao redor desse threshold.

Para exemplificar, suponha que um gestor público deseje realizar uma intervenção sobre os municípios carentes em termos de educação. Para isso, o gestor define que a regra de elegibilidade será derivada da variável (contínua) Índice de Desenvolvimento Humano (IDH). Ele estabelece, então, que a intervenção será aplicada para os municípios que possuírem IDH inferior a 0,7 (*threshold*). Em outras palavras, aqui a variável definidora do tratamento é o IDH e o limiar dessa intervenção é 0,7, de forma que todas as observações que estiverem abaixo de 0,7 receberão a intervenção, e todas as que estiverem acima não receberão. A intuição da regressão descontínua é que os municípios que têm 0,699 (ou um pouco menos) de IDH, e os que têm 0,701 (ou um pouco mais) de IDH são extremamente parecidos entre si, podendo o grupo não-tratado ser considerado um bom contrafactual para o grupo de tratamento.

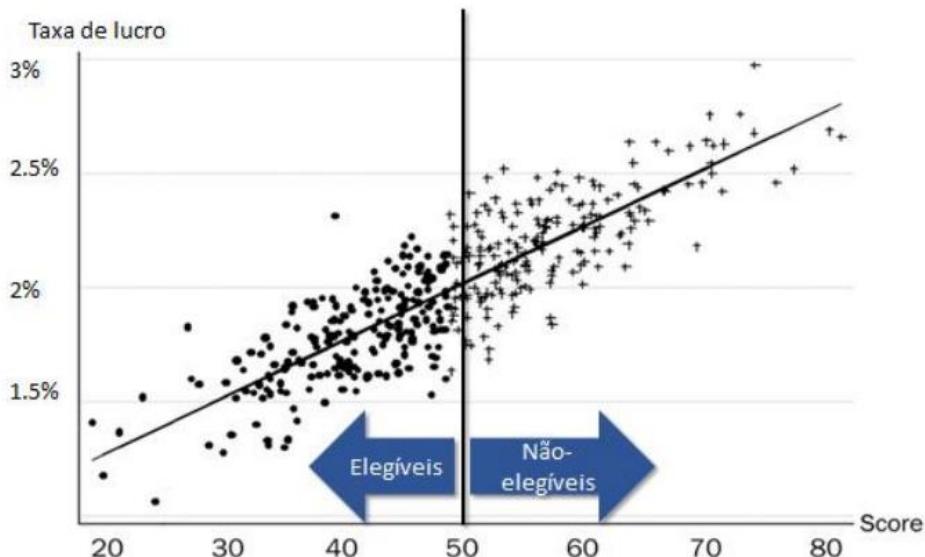
Para aplicar esse método ao estudo de caso anteriormente proposto, o formato em que o tratamento é ofertado precisa ser modificado. Nesse caso, o gestor decide alterar o

⁵Como em qualquer método, a defesa da aplicação do pareamento depende completamente da aderência de suas hipóteses ao caso concreto.

mecanismo para injetar capital nas microempresas. Agora ele decide ofertar o tratamento para as empresas conforme o escore de crédito delas. Tal escore foi calculado com base em características passadas da empresa, como idade, lucratividade, faturamento, número de empregados, etc. O escore varia de 0 a 100, onde 0 significa muita restrição a crédito, e 100 significa total acesso a crédito. O formulador da política determina discricionariamente que somente oferecerá o tratamento para as empresas que tiverem escore inferior a 50. Como o limiar escolhido é exógeno e exclusivo (somente esse programa usa), baseado numa variável contínua, e não pode ser manipulado (as empresas não conseguem trapacear dado que o escore foi calculado com base em dados passados), pode-se aplicar o RDD. A ideia, nesse caso, é comparar a taxa de lucro de empresas com escore um pouco abaixo de 50, elegíveis ao programa de crédito subsidiado, com empresas com escore um pouco acima de 50, não-elegíveis ao programa (Figura 7).

Como o nome do método alude, a estratégia de identificação no RDD se sustenta na existência de uma descontinuidade na variável de interesse após o tratamento no valor do limiar estabelecido. Entretanto, para se defender que essa descontinuidade (caso realmente verificada) decorre exclusivamente da intervenção, é importante verificar se ela não existia antes do tratamento. Se for observada alguma descontinuidade antes da intervenção, outros processos entre a variável definidora de tratamento (escore) e a variável de interesse (taxa de lucro) estão presentes e podem inviabilizar a defesa do método. No exemplo (Figura 8), percebe-se que a reta é bem suave e contínua em todo o espaço, inclusive no *threshold* 50.

Figura 8 – Regressão descontinuada: antes da intervenção

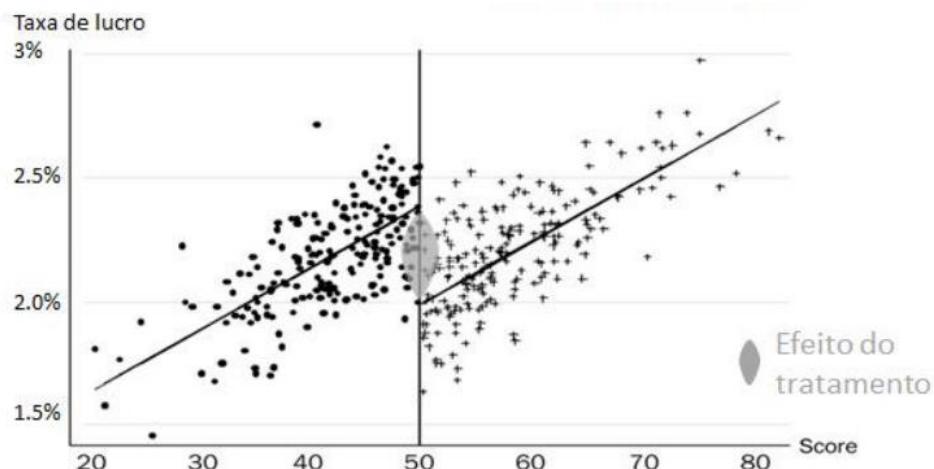


Fonte: Gertler et al. (2010).

Conforme apresentado na Figura 9, depois da intervenção se espera observar uma descontinuidade no valor de limiar, fator que permitirá avaliar o efeito do programa. É importante frisar que, no RDD, se estima um efeito local, pois somente as observações

localizadas próximas ao limiar são utilizadas na estimação. Tal definição de quão próximas do limiar as empresas devem estar para serem consideradas na estimação precisa ser avaliada com cautela. É importante garantir que as observações sejam parecidas entre si; entretanto, com o cuidado de não limitar a análise ao extremo, mantendo poucas observações e prejudicando o poder estatístico da estimação.

Figura 9 – Regressão descontinuada: depois da intervenção



Fonte: Gertler et al. (2010).

Em resumo, o método de Regressão Descontinua é muito poderoso e é o que mais se aproxima de um método experimental em termos de poder de inferência. Porém, a elegibilidade da intervenção precisa respeitar algumas características. Precisa-se de uma variável contínua que sustente a elegibilidade para o programa. A linha de corte deve ser claramente definida, exclusiva e não-manipulável. Por fim, reitera-se que o impacto será válido apenas para aqueles indivíduos que estão próximos à linha de corte. Logo, deve se ter claro qual pergunta está se buscando responder e se a análise com as empresas próximas ao limiar propicia a resposta mais acurada.

CONSIDERAÇÕES FINAIS

Como visto, métodos fracos geram resultados equivocados quando tratamos de Avaliação de Impacto. Não ter informação pode ser ruim; contudo, ter informação incorreta pode influenciar em sentido adverso a formulação de política pública. Resultados confiáveis só podem ser garantidos com a aplicação de métodos rigorosos, que possam ser defendidos à frente de qualquer audiência.

Em termos práticos, para decidir qual método é mais adequado ao caso concreto, é necessário analisar situação a situação. Nesse sentido, é fundamental conhecer em

profundidade o programa/intervenção que se pretende avaliar. São as regras do programa que definem em última instância qual o método mais interessante a ser utilizado. São elas que permitem argumentar a aderência do programa avaliado às hipóteses de cada método de avaliação. Muitas vezes, a complexidade dos programas faz com que seja necessário ser criativo para criar metodologias, requerendo que métodos sejam combinados a fim de encontrar resultados confiáveis, baseados em um contrafactual robusto. Cabe ao analista a responsabilidade de garantir que a identificação do efeito seja garantida através da utilização de métodos rigorosos.

REFERÊNCIAS

GERTLER, Paul J. et al. *Impact Evaluation in Practice*. Washington, DC: Inter-American Development Bank and World Bank, 2016. Disponível em: < <https://openknowledge.worldbank.org/handle/10986/25030> >. Acesso em: 07 de maio de 2018.