

Desenho de Pesquisa

Glauco Peres da Silva

Brasília
Enap
2017

Ficha Catalográfica por: Keicielle Schimidt de Oliveira – CRB1 2392

S5864d SILVA, Glauco Peres da.
Desenho de pesquisa / Márcia Miranda Soares e José Ângelo
Machado. -- Brasília: Enap, 2017.
112 p. : il.

ISBN:

1. Ciências Sociais. 2. Pesquisa. 3. Método de Pesquisa.
I. Título.

CDU 3:001.8

Catalogado na fonte pela Biblioteca Graciliano Ramos da Enap

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista da Escola Nacional de Administração Pública (Enap). É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

Enap Fundação Escola Nacional de Administração Pública
SAIS – Área 2-A
70610-900 – Brasília, DF
Telefones: (61) 2020 3096 / 2020 3102 – Fax: (61) 2020 3178
Sítio: www.enap.gov.br
Tiragem: 1000 exemplares

situações particulares do mundo que sirvam para os propósitos do trabalho. Ao longo da investigação, a pesquisa demandará que alguns conceitos sejam repensados e ressignificados, o que se torna, inclusive, um dos achados e uma contribuição daquela pesquisa. Essas diferenças não exauram as principais questões que distinguem ambas tradições de pesquisa nas Ciências Sociais e, por isso, parte-se para a apresentação e discussão dos desenhos de pesquisa em cada uma delas.

Cabe antes apresentar uma maneira um tanto distinta de se considerar o trabalho de investigação de um cientista social. Em uma discussão provocativa, Abbott (2004) argumenta que as construções metodológicas dos trabalhos em Ciências Sociais, diferentemente das visões estanques entre abordagem qualitativa e quantitativa, são mais bem representadas pela analogia aos fractais. Ao lidar com dados empíricos, o pesquisador será forçado a tomar decisões que estão associadas a uma ou outra forma de tratar os dados. Por exemplo, em um trabalho de observação participante, uma pesquisadora pode quantificar certas ocorrências como forma de melhor compreender determinada situação. Em outros casos, um pesquisador pode focar em compreender quais são as características particulares de um caso desviante em uma regressão³⁷.

É diante desse contexto de organização da ciência e seus reflexos sobre a prática empírica que pesquisadores avaliam situações concretas ou testam decorrências das teorias.

3.1 Pesquisa com N-grande

Como dito, a ênfase das pesquisas com N-grande está primordialmente em modelos de “causas dos efeitos”³⁸. Essa abordagem vai, então, partir para a pesquisa empírica com um modelo de causa dos efeitos proveniente de trabalhos teóricos e empíricos anteriores e, então,

³⁷ Essa constatação de Abbott é similar a um argumento recorrentemente utilizado por pesquisadores que se veem diante do debate quali-quant: presenciei mais de uma vez pesquisadores associados ao lado quantitativo dizendo que também fazem pesquisa qualitativa. Inclusive “porque seus trabalhos estão repletos de variáveis *dummies*”.

³⁸ Novamente, essa apresentação inicial está baseada em Morton e Williams, 2010, cap. 2.

avaliar aquele modelo com os dados disponíveis, sejam observacionais ou experimentais. Em geral, os modelos apresentam várias previsões causais que são consistentes entre si e com a hipótese decorrente. “A causalidade do modelo é comumente condicional a dadas situações, ou seja, algumas variáveis podem estar determinadas simultaneamente. A avaliação do modelo conduz a pesquisas futuras tanto teóricas quanto empíricas” (MORTON; WILLIAMS, 2010, p. 34). As investigações empíricas tendem a usar versões aplicadas do modelo, ainda que experimentos permitam que o pesquisador se desloque para além do mundo observado, assim como no caso teórico. Mas os testes empíricos se referem sempre às decorrências das teorias e não às teorias em si. Como visto na Figura 3, os testes se dão em plano distinto.

A seleção de dados em uma pesquisa com N-grande possui particularidades que precisam ser destacadas. Em um modelo ideal, as observações são escolhidas aleatoriamente. Esse procedimento é necessário para não gerar nenhum tipo de viés na análise, ao garantir que a amostra possua as características relevantes da população. Não há, assim, nenhuma observação que mereça destaque. Todas são analisadas de maneira idêntica entre si. A pesquisa com N-grande tem como uma hipótese fundamental a homogeneidade dos casos observados. Essa homogeneidade se dá em relação às características mais importantes dos casos tratados, e as diferenças consideradas podem ser controladas. A utilização de controles é bastante importante nesse tipo de abordagem, pois é o que permite à pesquisadora a possibilidade de encontrar os resultados causais desejados. Esse aspecto básico vai se desdobrar em questões sobre as quais se desenvolverão técnicas estatísticas e sofisticadas metodológicas. Também se pode apontar como característica desse tipo de pesquisa a escolha de observações a partir das variáveis independentes, ou seja, daquelas que supostamente causam o efeito desejado. Por essa razão, no conjunto de observações tratados, há observações em que o fenômeno analisado não ocorreu. Isso é necessário para que se possa mensurar o efeito da variável explicativa sobre o fenômeno que se deseja estudar. Essa é uma característica

marcante dessa forma de se fazer pesquisa, que a distingue das pesquisas com N-pequeno.

Nas pesquisas com N-grande, os dados que são utilizados são classificados de duas formas: dados observacionais ou experimentais. Esses nomes referem-se à maneira como foram coletados: os dados experimentais foram obtidos por meio de um experimento e os observacionais, não. Em um experimento, o pesquisador possui controle preciso sobre a interferência dos fatores que causam ou modificam o resultado obtido. Já em um estudo observacional, esses controles não são estabelecidos *a priori*. A pesquisadora precisará lidar com eles após coletar os dados. A lógica ideal de trabalho está mais próxima aos estudos experimentais, mas a grande maioria das pesquisas utiliza dados observacionais. Em nosso exemplo sobre o PBF, a avaliação em uma lógica experimental é entendida como mais adequada, embora nem sempre seja possível³⁹. Nesse sentido, é preciso explorar as diferenças nessa lógica de análise.

3.1.1 Experimentos⁴⁰

Para ilustrar a importância da estruturação da pergunta com relação ao raciocínio que ela implica e retomar a distinção entre as tradições das pesquisas, vale a pena notar que a lógica da busca pela causa dos efeitos em economia, por exemplo,

explica porque muitos experimentalistas que se baseiam na tradição econômica não parecem interessados em usar seus experimentos para estudar uma particular relação única entre causa e efeito isoladamente, mas ao contrário, tipicamente estudam uma série de relações

³⁹ Há, inclusive, comprometimentos éticos no que tange a criação de experimentos nas Ciências Sociais, sejam eles em laboratórios ou não. A preocupação com o tema nas ciências humanas e sociais é crescente. A título de exemplo, a Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo organizou um seminário em 2012 para tratar o tema e o resultado está resumido em dois documentos, disponíveis em: <<http://pesquisa.fflch.usp.br/node/38>>.

⁴⁰ Apesar de haver um volume sobre experimentos na coleção à qual este livro faz parte, entende-se como necessária uma breve apresentação sobre a lógica experimental para que se entenda a constituição de um desenho de pesquisa. O foco estará apenas na lógica mais geral, sem nenhuma pretensão de apresentar elementos de sua execução, por exemplo.

previstas a partir de alguma teoria existente. Estes experimentalistas usualmente começam com um modelo formal de algum processo, derivam um número de previsões daquele modelo e então consideram se o comportamento das unidades estudadas está de acordo (ou não) com as previsões em seu experimento (MORTON; WILLIAMS, 2010, p. 34, tradução nossa).

Ou seja, os pesquisadores estão voltados a aspectos dedutivos de suas pesquisas e à verificação de suas ocorrências no mundo empírico. Pergunta-se sobre a causa dos efeitos, portanto.

Uma hipótese da qual uma pesquisa com N-grande parte é a da associação teórica entre dois conceitos. O primeiro, aquele que será explicado, é chamado de dependente ou resposta. O segundo, aquele que explica a ocorrência, é chamado de independente. Variáveis correspondem a cada um desses conceitos, chamadas de variáveis dependente (Y) e independentes (X). É bastante comum encontrarmos como variável independente um conjunto de variáveis e não apenas uma. Para simplificar, esse conjunto de variáveis será representado por X, apenas, indicando esse conjunto completo. A relação entre as variáveis indicadas acima pode ser representada da seguinte forma:

$$X \rightarrow Y \quad (1)$$

indicando que as variáveis X explicam ou causam a variável Y.

A questão premente em uma pesquisa está em como observar a relação indicada? A preocupação central para a determinação da relação entre X e Y está na capacidade de isolar esse efeito de outras influências. A relação (1) indicada assume que existe apenas X como fator de causa ou de influência sobre Y. Seguindo o nosso exemplo de avaliação do PBF, um pesquisador preocupado com os impactos do programa nas condições materiais de vida das famílias beneficiárias poderia dizer que X é uma variável que mensura se a família é ou não beneficiária do programa⁴¹ e

⁴¹ Nesse caso, X assumiria valor igual a 0, quando a família não recebe o benefício, e igual a 1, quando recebesse. Variáveis dicotômicas desse tipo são chamadas de variáveis *dummy*.

Y é uma variável que avalia as condições materiais dessas famílias. Essa pode ser operacionalizada pela renda familiar, por exemplo.

Em pesquisas nas Ciências Sociais, é muito difícil encontrar teoricamente, inclusive, uma relação tão simples e unívoca como representada em (1). Em geral, Y será resultado de diferentes tipos de interação entre X, que seriam as variáveis explicativas que teoricamente se deseja observar, e outro conjunto de variáveis, denominadas Z, que também afetam Y, mas sobre as quais não se tem interesse analítico. Estabelecer qual é a associação entre as três variáveis é fundamental para que a relação entre X e Y seja isolada e medida. É possível estabelecer diferentes relações entre as três variáveis. A mais simples delas pode ser apresentada da seguinte forma:

$$X \rightarrow Y \leftarrow Z \quad (2)$$

Em uma situação como essa, o conjunto de variáveis Z também afeta Y ao mesmo tempo que X. Suponha que as relações indicadas na expressão (2) ocorram simultaneamente com sinais trocados; ou seja, X afeta Y positivamente e Z afeta Y negativamente. Se as magnitudes dessas relações forem semelhantes, erroneamente diremos que X não afeta Y, quando essa conclusão está equivocada. Suponha agora que Z afete Y no mesmo sentido que X o faz: se X afeta Y positivamente, Z também afeta Y positivamente; se a relação entre X e Y for negativa, Z também afeta Y negativamente. Em situações como essa, se observarmos apenas a relação entre X e Y, superestimaremos a relação entre essas variáveis, pois o efeito de Z sobre Y não estará considerado. Em nosso exemplo sobre a avaliação do PBF, o pesquisador pode considerar que não foi apenas o fato de as famílias estarem submetidas ao programa (X) que provocou alterações nas suas condições materiais de vida (Y). Seria possível elencar um outro conjunto de fatores que tenham afetado Y que não tenham a ver com X, como, por exemplo, as condições macroeconômicas do país (Z). O ambiente macroeconômico do país está evidenciado pelos índices de desemprego, de inflação, pela variação do produto interno bruto etc.

A consideração da existência de Z e os procedimentos que decorrem dessa constatação remetem à importante distinção entre correlação e causalidade. Correlação é uma estatística que captura a simultaneidade das oscilações de duas variáveis em relação às suas respectivas médias⁴². Sua medida reflete, assim, uma propriedade dos valores assumidos para cada uma das variáveis e não considera a existência de outras variáveis. Essa observação sugere que se X e Y são independentes entre si, a correlação seria teoricamente igual a zero, mas o contrário não é verdadeiro; correlações iguais a zero não indicam independência. Nesse mesmo sentido, elevadas correlações podem simplesmente ser espúrias⁴³, ou seja, não indicam nenhuma relação entre as variáveis. Se o que se pretende é identificar causalidade, o cálculo da correlação não contribui nesse sentido⁴⁴.

Nos casos acima indicados, precisamos encontrar uma maneira de lidar com Z. Caso contrário, não seremos capazes de isolar o efeito de X em Y. O processo pelo qual se faz esse isolamento de efeitos é chamado de criação de controles. A inserção de controles nas análises permite que as pesquisadoras busquem observar apenas a relação de interesse, sem outros efeitos simultâneos. Dessa forma, a variável Z pode ser dividida em dois subgrupos: as variáveis que são observáveis e as que são não observáveis. As variáveis observáveis são aquelas que podemos considerar diretamente na análise empírica e de alguma forma identificar seu efeito sobre Y. Na expressão (2) acima, seria equivalente a encontrar a magnitude do efeito entre Z e Y para isolar a variação de Y que não

⁴² A fórmula da correlação é dada por: $\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}$, em que ρ é o índice de correlação, σ é o desvio padrão tanto da variável x quanto da variável y. Deve-se notar que a correlação é calculada a partir das diferenças entre cada uma das observações das variáveis x e y em relação às suas respectivas médias.

⁴³ Para demonstrações de casos de correlação espúria, veja <<http://www.tylervigen.com/spurious-correlations>>.

⁴⁴ A referência principal para a análise experimental nas Ciências Sociais é o modelo de Rubin que, com as contribuições posteriores de Neyman e Holland, também é chamado modelo de Neyman-Rubin ou de Neyman-Rubin-Holland. *Vide* Sekhon (2007). Deve-se mencionar também a importância recente dos trabalhos de J. Pearl a respeito do tempo. *Vide*: <<http://bayes.cs.ucla.edu/BOOK-2K/>>.

é causada por X. Dessa forma, seria possível encontrar analiticamente a importância de X sobre Y. Já quando Z é não observável, o problema de interferência persiste. Ainda que se reconheça a possibilidade ou mesmo a existência teórica de Z, pode ser de fato difícil identificar a interferência de Z sobre Y e, conseqüentemente, a de X sobre Y. No caso da avaliação do Bolsa Família, é possível dizer que as variáveis macroeconômicas de desemprego, inflação e PIB são observáveis pelo pesquisador e podem ser controladas na pesquisa. Há componentes macroeconômicos, porém, que são não observáveis, como, por exemplo, as expectativas sobre o desempenho econômico, que afetam diretamente as decisões de investimento. As expectativas dos agentes econômicos não são observáveis no momento de avaliação do PBF. Em um caso como esse, seria preciso considerar qual o desenho de pesquisa adequado que lide com uma variável não observável relevante.

Em um experimento, o problema de se isolar efeitos de X e Z sobre Y está resolvido. Por definição, o experimento se dá em uma situação em que as variáveis de contexto, que podem interferir na relação que se deseja estudar, estão controladas. A imagem de uma situação vivida em um laboratório é a melhor analogia para a compreensão de um experimento. Todas as variáveis que podem influenciar no valor de Y são conhecidas e estão estáveis; estão controladas. Permite-se que apenas X varie. Em uma situação como essa, os efeitos de X sobre Y são identificados com precisão.

A questão que se coloca está em: quão fidedigna com a realidade social está essa imagem do laboratório controlado? Os estudos em Ciências Sociais se dividem aqui em um duplo caminho. No primeiro caminho, está a utilização de simulações em laboratórios com o intuito de observar os efeitos exclusivos de X sobre Y. A Economia, em bastante escala, e a Ciência Política, ainda em menor montante, têm utilizado esse tipo de expediente para a observação dos fenômenos de interesse. Em um texto famoso sobre o método experimental, Vernon Smith, vencedor do Prêmio Nobel de Economia, assim se refere à utilização dessa abordagem na economia:

A questão de se a economia é uma ciência experimental é discutível entre economistas experimentais que estão, e deveriam estar, muito ocupados se divertindo em seus trabalhos para refletir nas implicações metodológicas sobre o que eles fazem. Mas quando nós o fazemos, como em introduções abrangentes sobre o campo, o que dizemos? Duas citações vindas de fontes impecáveis servirão para introduzir conceitos desenvolvidos em seguida. A primeira enfatiza que uma categoria importante do trabalho experimental ‘... inclui experimentos desenhados para testar as previsões vindas de teorias formais bem articuladas e para observar regularidades não previstas, em um ambiente controlado que permite que essas observações sejam interpretadas sem ambiguidade em relação à teoria’ (KAGEL; ROTH, 1995, p. 22). Economistas experimentais acreditam fortemente, me parece, que esta é a defesa científica mais forte do método experimental: apoiamos a nossa investigação experimental no firme alicerce da teoria econômica (de jogos). Uma segunda vantagem crucial, reconhecendo que os testes de campo envolvem testes de conjuntos perigosos de hipóteses múltiplas, é que ‘os métodos de laboratório permitem uma dramática redução no número de hipóteses auxiliares envolvidas em examinar uma hipótese básica’ (DAVIS; HOLT, 1993, p. 16) (SMITH, 2002, p. 97, tradução nossa).

Embora haja sérias críticas sobre a validade externa dos experimentos controlados em laboratório, esse tipo de desenho de pesquisa vem sendo amplamente aplicado. Uma crítica importante para esse tipo de desenho de pesquisa está na proximidade das condições de laboratório àquelas que os indivíduos estudados encontram em seu ambiente social, de forma a vincular a reação desses indivíduos do laboratório às suas condições sociais cotidianas.

A outra possibilidade de análise experimental está na busca por criar experimentos em condições naturais, fora de um laboratório. Há aqui também uma dupla possibilidade: a coleta de dados em uma determinada situação similar ao ambiente de experimento. São chamados de experimentos naturais. A outra possibilidade é a de estimular indivíduos em seus ambientes sociais regulares, cotidianos, e observar as suas reações a esses estímulos.

Antes de entrarmos nessa distinção, é importante apresentar as características importantes de um experimento. O pesquisador está preocupado com o isolamento do efeito de outras variáveis sobre a

variável dependente e o experimento é, por definição, um meio poderoso de estabelecer esse isolamento. Para que isso ocorra, o pesquisador precisa criar dois grupos idênticos em tudo, menos na ocorrência de X. O ambiente ideal é aquele em que se poderia comparar dois cenários idênticos, sendo que em uma das situações X ocorreu e na outra X não ocorreu. Em um cenário como esse, teríamos para cada observação:

$$\delta_i = (Y_i|X = 1) - (Y_i|X = 0) \quad (3)$$

em que δ é o efeito observado para cada observação i , $(Y_i|X = 1)$ é o valor encontrado em Y na observação i quando X ocorre e $(Y_i|X = 0)$ é o valor de Y para a observação i quando X não ocorre. Essa situação indicaria o efeito de uma forma simples para cada indivíduo. Se cada um dos termos do lado direito da equação é observado exatamente no mesmo ambiente, não precisaríamos nos preocupar com Z, já que Z teria o mesmo valor em ambos e o efeito da variação de X sobre Y, se existisse, decorreria apenas da própria variação de X.

Porém, essa situação é impossível. Ela é um cenário teórico, somente. Não há como termos dois mundos iguais nos quais apenas X seja diferente. Esse exercício mental é similar ao contrafactual já apresentado e é a maneira básica de proceder para a criação dos controles tão necessários à pesquisa quantitativa. Weber é uma referência histórica para esse tipo de elaboração, aplicada em sua obra *A ética protestante e o espírito do capitalismo*, mas é com a abordagem proposta por James Heckman que é amplamente utilizado em Economia e ganha projeção nas Ciências Sociais. Uma análise contrafactual é um exercício abstrato em que uma situação hipotética não existente é criada para servir de comparação a um contexto observado. O exercício abstrato se dá supondo o que teria ocorrido na realidade se algo anterior não tivesse acontecido. Nesse caso, é conhecido que esse algo anterior ocorreu e tudo que se observa depois desse fato está de alguma forma relacionado a evento prévio. Por isso, não há meios para observar o mundo considerando que o evento não tenha ocorrido, a não ser por meios indiretos. A análise contrafactual, então, pressupõe a comparação com um contexto não realizado e só pode

ser feita empiricamente por aproximações. A análise estatística contribui para essa avaliação, como ficará claro em seguida.

Assim, a mera diferença nos valores observados de Y , entre duas situações em que X ocorre e outra em que X não ocorre, se observada, não controlaria para o impacto de Z sobre Y . Ou seja, o resultado continuaria contaminado pela existência de Z , como na expressão (2). A maneira de evitar esse tipo de problema é raciocinar em termos de efeitos médios. Ao invés de observarmos apenas a diferença como indicado em (3), introduzimos uma modificação nos dois termos da equação, a partir dos efeitos médios:

$$\begin{aligned} E[\delta] &= E[(Y|X = 1) - (Y|X = 0)] \\ \delta_i &= (Y_i|X = 1) - (Y_i|X = 0) \end{aligned} \tag{4}$$

Como a média dos efeitos da diferença ($E[\delta]$) é igual à própria diferença (δ), pode-se considerar a diferença entre as médias de cada uma das situações hipotéticas, uma em que X ocorre contra a que X não ocorre. Como a média, tal como apresentada na equação (4), nos ajuda?

Se não existe a situação na qual apenas X varie em um mundo de *ceteris paribus*⁴⁵, trabalha-se com efeito médio. $E[(Y|X = 1)]$ significa a média dos valores de Y quando X ocorre, enquanto $E[(Y|X = 0)]$ significa a média dos valores de Y quando X não ocorre. A primeira expressão seria observada no grupo em que X ocorreu e esse grupo recebe o nome de grupo de tratamento. A segunda expressão determina o grupo de controle, aquele em que X não ocorreu. Em que condições pode-se dizer que essa diferença média seja aquela em que δ é o valor que se deseja analisar? Isso ocorrerá se o grupo de tratamento e de controle forem idênticos em Z . Como se sabe se essa é a situação? Se os indivíduos que participam de um ou de outro grupo forem designados de maneira aleatória, garante-se que Z é idêntico, em média, nos dois grupos. O jargão utilizado para

⁴⁵ Expressão latina amplamente utilizada por economistas, significa “todo o mais constante”. É utilizada para se referir ao caso em que as comparações abstratas são realizadas com a variação apenas de um fator específico, mantendo-se os demais constantes.

essa mensuração dos impactos de um tratamento é “efeito médio do tratamento” (do inglês *average treatment effect*, comumente abreviado como ATE).

Isso implica que a dimensão da atribuição aleatória do tratamento em um dos grupos é passo fundamental para que a avaliação do impacto de X sobre Y ocorra. A aleatoriedade é importante aqui porque se há um conjunto de indivíduos a quem o tratamento é autoatribuído, as razões para esses indivíduos decidirem receber o tratamento podem omitir algum componente de Z que torna os grupos diferentes. Nesse caso, não saberemos se a diferença entre as médias é um efeito de X ou de algum fator não observado.

Por exemplo, a aplicação de um experimento para a avaliação do PBF seria idealmente conduzida se as famílias que foram submetidas ao programa tivessem sido aleatoriamente selecionadas para receber o complemento de renda, a partir do universo das famílias elegíveis ao programa. Partindo desse ponto, seria possível assumir que as diferenças observadas nas condições materiais de vida das famílias decorreriam tão somente do fato de o grupo de tratamento ter recebido o benefício, enquanto o grupo de controle não. Porém, raramente o pesquisador tem essa possibilidade de atribuir o tratamento aleatoriamente. Com o PBF, não foi diferente. As famílias não foram separadas em dois grupos de maneira aleatória e, por essa razão, a avaliação do programa não poderia ser feita de maneira simples. A comparação das condições materiais das famílias que fizeram parte do programa *versus* daquelas que não fizeram não basta, pois estaríamos agindo como se o mundo estivesse de acordo com a expressão (3) e sabemos que não é verdade. Assim, a mera diferença entre a renda das famílias não é suficiente. É preciso que os grupos sejam em média idênticos. Se cada família pode decidir se faria ou não parte do programa ou se algum órgão governamental decidiu quem receberia o programa e quem não, a partir de um critério não observável pelo pesquisador, pode-se argumentar que essa decisão quanto à atribuição do tratamento esconde algum viés que pode contaminar a pesquisa.

Exemplo 9

Heckman e Hotz (1989) aplicaram essa lógica experimental à avaliação dos efeitos de programas de treinamento sobre a produtividade de indivíduos em seus trabalhos. Como medir esse efeito? Pode-se pensar que o grupo que recebe o treinamento é o grupo de tratamento, enquanto aqueles que não receberam formam o grupo de controle. Era preciso considerar a forma como os indivíduos se separaram entre os grupos, pois, se não fosse aleatória a divisão, seria possível imaginar que esses indivíduos que tenham feito o treinamento já fossem os mais produtivos de qualquer forma.

Cabe observar aqui que há casos em que a atribuição de tratamento não é dada de forma aleatória. Os indivíduos se autosselecionaram. É possível ainda assim lidar com a lógica experimental nesse caso, desde que se acredite que a escolha de fazer parte do grupo de tratamento ou de controle não esconda qualquer viés. Esses métodos, chamados de quase-experimentos, serão tratados mais adiante.

Há outras duas possíveis relações entre Y, X e Z com as quais o experimento pode lidar. A primeira delas está representada na expressão a seguir:

$$X \rightarrow Y \leftarrow Z \quad (5)$$


De acordo com a expressão (5), Z não só interfere em Y, como também interfere em X. Nesse caso, Z é conhecida como *confounder*. Em uma situação desse tipo, é necessário que se busque criar estratégias de identificação. Isso implica em buscar isolar o efeito de Z, tanto em X quanto em Y para ser capaz de determinar qual a importância de um fator sobre o resultado de Y. Quando não é calculado, diz-se que as estimativas da relação entre X e Y sofrem de viés de variável omitida. O pesquisador deve lidar com esse tipo de situação, embora não exista uma maneira padrão de fazê-lo. Os trabalhos acadêmicos variam bastante com relação a esse tópico, inclusive porque depende especificamente do objeto de estudo⁴⁶.

⁴⁶ Os trabalhos de Clarke sobre esse tema são bastante reveladores ao indicar que a inclusão de controles não é suficiente para garantir que tenhamos resolvido o problema. Vide Clarke (2005, 2009).

As estratégias de identificação para isolar o efeito de Z são as mais diversas. É importante notar que a identificação é um passo central na busca pela mensuração da relação entre X e Y. A pesquisadora deve ter claro onde ou de que forma essa relação pode ser observada e medida. A identificação passa por isolar os efeitos endógenos e simultâneos que ocorrem diante daquilo que se deseja estudar.

Exemplo 10

Por vezes, o efeito observado entre X e Y não existe. Ele apenas é resultado da presença de Z. Esse, sim, produz tanto X quanto Y. Um exemplo trivial a respeito de uma situação semelhante é o cálculo da correlação entre o número de ataques de tubarões a pessoas (Y) em relação ao consumo de sorvetes por habitantes (X). Não há razão alguma para supor que exista uma correlação desse tipo entre as variáveis, mas, em termos práticos, ela deve ser alta. Nesse caso, pode-se supor que as cidades nas quais ocorrem os ataques de tubarões são cidades litorâneas, com grande incidência solar. O elevado calor dessas regiões (Z) explica também o alto consumo de sorvetes.

A outra relação entre as variáveis X e Y está expressa a seguir:

$$X \rightleftarrows Y \quad (6)$$

Em uma situação como a indicada em (6), vemos que ao mesmo tempo em que X afeta Y, Y afeta X. É uma situação na qual múltiplos componentes estão ocorrendo simultaneamente e em vários casos se reforçando e, pior, afetando-se mutuamente. O jargão para se referir a problemas desse tipo é o de endogeneidade. Esse conceito está associado à hipótese de que a variável explicativa está correlacionada com a variável dependente por meio de uma dupla relação: qualquer uma das variáveis pode causar a outra. Em um caso como esse, a mera mensuração da relação entre X e Y não fornece informação sobre o impacto de X em Y. Um exemplo de uma situação em que há endogeneidade é uma análise que busca verificar a relação entre o número de crimes por municípios e o contingente de policiais nessas cidades. A relação de causalidade pode se dar nos dois sentidos: o número de policiais afeta negativamente o

número de crimes cometidos em determinada cidade, mas o número de crimes em um município pode provocar aumento no número de policiais alocados para aquela localidade. Ou seja, a variável explicativa, suponha que nesse caso seja o número de crimes de uma cidade, afeta a variável dependente, o número de policiais. Porém, a relação entre ambas não resolve o problema teórico de endogeneidade. É preciso adaptar a análise para que esse problema não interfira no resultado.

Vale dizer que a endogeneidade não é sempre uma propriedade dos conceitos ou das variáveis em si, mas quase sempre surge a partir da análise que se implementa. Voltando ao exemplo da relação entre número de crimes e contingente policial, a introdução da questão temporal pode amenizar ou até contornar a endogeneidade. Pode-se avaliar se é o número de crimes ocorridos em um momento t do tempo que explica o contingente policial no momento $t+1$ ou se é o contrário, o contingente policial em t que explica o número de crimes em $t+1$. Se houver razões para achar que ambos efeitos ainda ocorrem simultaneamente, mesmo que defasados temporalmente, a análise precisa explicitamente lidar com essa característica do fenômeno em questão.

Exemplo 11

Ann-Kristin Kölln (2016) afirma que a representação política é um processo que tem como momentos importantes o ato de votar e o acompanhamento do representante. A autora afirma que a pesquisa anterior aponta para uma dupla relação de causalidade nesse processo: de um lado, indivíduos decidem votar (em lugares em que o voto não é obrigatório) quando percebem que seus representantes exerceram suas funções de acordo com suas expectativas e, assim, sentem-se representados; de outro, por terem decidido votar, cidadãos voltam maior atenção às ações de seus representantes, restringindo as escolhas desses, pois os eleitos buscarão atender às suas demandas. Cria-se, portanto, um ciclo: vota-se em maior número quando se sente representado, mas sente-se representado após votar. A autora lida com esse problema através de uma análise dinâmica via um modelo de equações estruturais.

Há, então, duas possibilidades de que um experimento seja utilizado em Ciências Sociais. Como dito, a primeira delas faz uso de simulações em laboratório. Consiste em um ambiente controlado, em que os *confounders* e demais variáveis que afetam o estudo são conhecidos e não possuem efeito sobre as magnitudes observadas de Y. É o caso defendido por Vernon Smith na citação do início do capítulo. A outra possibilidade é a de criar um experimento fora do laboratório. Nessa situação, há duas alternativas: em uma, são situações nas quais indivíduos são expostos a informações distintas em um ambiente em que a pesquisadora possui controle sobre os estímulos a que as pessoas estão sujeitas, ou mesmo em que os grupos de tratamento e controle são de fato produzidos. Os trabalhos de Josh Angrist são possivelmente os melhores exemplos sobre a lógica do uso de dados observacionais para a aplicação de lógicas experimentais, nos assim chamados experimentos naturais⁴⁷. A aplicação de um método desse tipo seria o caso ideal apontado sobre a avaliação do PBF. Se as famílias fossem atribuídas ao grupo de tratamento e de controle de forma aleatória, os resultados observados estariam associados ao programa. A outra alternativa é a de estimular os indivíduos em seus ambientes convencionais com informações diferentes para a averiguação sobre quais são suas reações. Nesse caso, os estímulos são aleatoriamente associados a um ou a outro indivíduo e a comparação entre as reações fornece os dados para avaliação do efeito de determinado tratamento.

Exemplo 12

Os estudos que buscam avaliar as razões para a baixa participação feminina como representantes políticas eleitas sugerem uma série de causas possíveis. Entre elas, está a de que os eleitores não escolhem candidatas mulheres por preconceito. Rosario Aguilar *et al.* (2015) conduzem um experimento no Brasil para avaliar esse fenômeno. Os autores apresentam candidatos para eleitores por meio de fichas com o perfil biográfico de cada um. Havia perfis equivalentes entre si nos quais a única alteração entre eles era exatamente o gênero: Fernando era substituído por Fernanda, por exemplo. O resultado é contundente:

⁴⁷ Para uma revisão a respeito da carreira de Angrist e sua importância nos estudos experimentais, veja <<https://www.technologyreview.com/s/508381/the-natural-experimenter/>>

entre candidatos de mesmo nível, o eleitor brasileiro prefere mulheres. Não só não há preconceito, como há leve preferência a candidatas em relação a candidatos. A baixa participação feminina deve decorrer por outras razões.

Todos os problemas apresentados entre as relações X, Y e Z também devem ser tratados em uma pesquisa que utilize dados observacionais. O grande desafio nesse caso é o de lidar com dados cujas variáveis precisariam ser controladas de fato e não foram. Não houve aleatoriedade na atribuição do tratamento, nem controle sobre *confounders*. Inclusive, X e Y podem afetar-se simultaneamente e o pesquisador deverá corrigir os problemas de endogenia presentes. Há, entretanto, um conjunto de desenhos de pesquisa que buscam lidar com esses problemas e são, a essa altura, consagrados na literatura das Ciências Sociais. Os desenhos a seguir, brevemente apresentados, estão agrupados sob o título de métodos quase-experimentais, com o intuito de indicar ao leitor que seu uso se dá quando a pesquisadora se vê diante de dados observacionais, mas há a possibilidade da aplicação de um desenho de pesquisa que busque simular um experimento.

3.1.2 Métodos quase-experimentais

Uma maneira de distinguir os desenhos de pesquisa a seguir está no fato de que todos eles buscam lidar com dados observacionais, de forma a se aproximar do desenho de pesquisa ideal de um experimento. Ou seja, a intenção em cada um deles é a de tentar reproduzir a capacidade de estender os controles de tal forma a identificar a relação entre X e Y somente. A sua aplicabilidade varia de acordo com o problema a ser estudado e com a disponibilidade de dados. Não há preferência entre eles, apesar de haver tendências na literatura que enfatizam a maior adequação de um desenho em face de outros no intuito de replicar um experimento, mas essas tendências variam ao longo do tempo e mesmo as mais antigas ainda são utilizadas atualmente. Cinco desenhos distintos serão considerados aqui: Regressão Múltipla, Regressão Descontínua,

Propensity score matching, Controle Sintético e *Difference-in-differences*. É importante dizer que o objetivo aqui é apresentar cada um desses tipos a partir das suas características como desenho de pesquisa, ou seja, não serão apresentadas as maneiras como são estimados os parâmetros em cada um deles, mas como tentam solucionar o problema de adequar dados observacionais à lógica experimental. O leitor deve notar que, entre os critérios quanto à escolha desses modelos, estão o número de unidades nas quais o tratamento foi atribuído e aspectos temporais, como o tempo de duração desse tratamento ou o momento em que as medições são realizadas. Nesse sentido, a pesquisadora terá de lidar com restrições, como o momento em que as medições foram realizadas e sobre quantas unidades receberam o tratamento. Como os dados são observacionais, não puderam ser controlados ou determinados pela pesquisadora e, assim, acabam influenciando a escolha do desenho. A discussão a seguir deve tornar mais claros esses aspectos.

3.1.1.1 Regressão múltipla

Esta seção tem a pretensão de apenas expor a apresentação mais básica envolvendo análise de regressão múltipla, já que há um livro sobre o tema na coleção da qual este manual faz parte. Serão abordados tópicos apenas para adequar a linguagem e para que os modelos posteriores se tornem mais inteligíveis.

O objetivo de um modelo de regressão linear em sua forma mais simples é quantificar a relação entre as variáveis dependente (Y) e explicativa (X), estabelecendo explicitamente os controles de outras variáveis, Z . Em sua expressão mais geral, a relação entre essas variáveis pode ser indicada por:

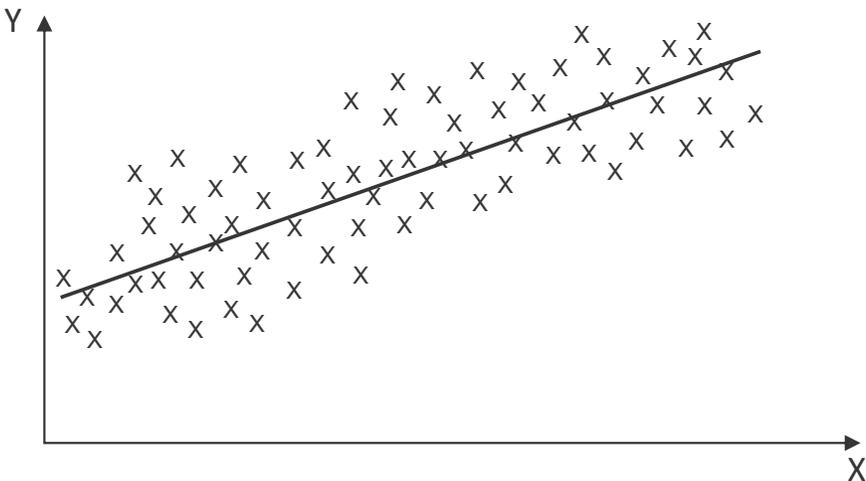
$$Y_i = \alpha + \beta X_i + \theta Z_i + \varepsilon_i \quad (7)$$

Os parâmetros α , β e θ são estimados e apontam os efeitos médios de variações em X e Z sobre Y . É interessante notar que a interpretação geral de uma equação de regressão se dá em termos de efeitos médios

e também de que sua aplicação permite que se encontrem os efeitos das variáveis líquidas da interferência de outras variáveis. Essa observação é importante, pois ela justifica por que, por um lado, se considera que uma regressão múltipla por si só não se refere à causalidade entre X e Y, mas sim deve ser interpretada como uma correlação. Porém, ela não é idêntica ao simple cálculo de correlação, porque na regressão estão considerados os efeitos dos controles sobre a variável dependente, o que não ocorre no cálculo de uma correlação. Assim, apesar de ser mais sofisticada na avaliação da relação entre X e Y, não se refere à relação causal entre elas. Para uma interpretação visual de uma regressão linear, o Gráfico 1 ilustra o resultado de uma regressão de X em Y.

No Gráfico 1, cada uma das marcas “x” representa uma observação de acordo com os valores de X e Y. A reta escura que atravessa essas observações representa a reta de regressão, construída a partir das estimativas dos parâmetros da equação 7. Essa reta, tal como está representada, apenas considera a relação entre X e Y, sem apresentar a relação de Y e Z. Se os parâmetros foram estimados a partir de uma equação tal como a 7, eles estão identificados com o efeito líquido de X sobre Y.

Gráfico 1 – Representação de uma reta de regressão de X em Y

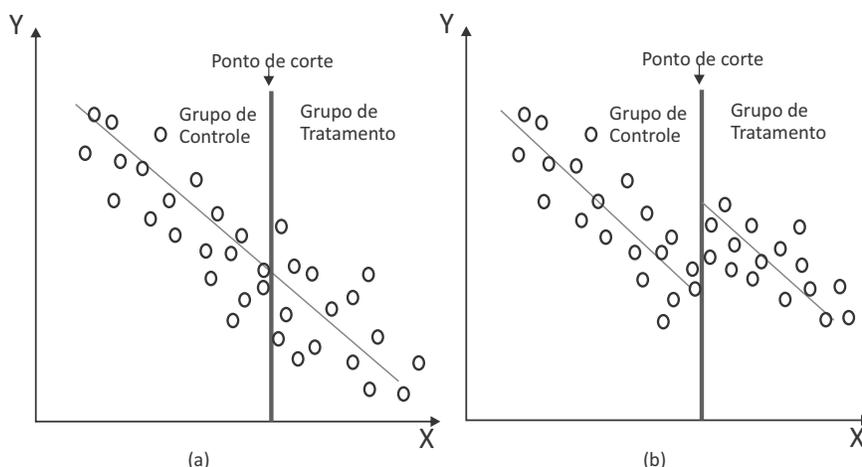


Fonte: Elaboração do autor.

Apenas para explicitar o que se quer dizer aqui, considere a estatística de correlação de Pearson, tal como apresentada na nota 28. Note que nesse índice de correlação apenas estão considerados Y e X. Não há possibilidade de que o valor de Z afete o cálculo da correlação, ainda que ele possa existir na prática. Ou seja, a regressão possibilita contornar esse problema, mas ainda assim, a compreensão mais comum com relação à interpretação de seus resultados é a de uma correlação em que os controles são considerados. Por essa razão, a estimação dos parâmetros, usando o modelo de regressão, é importante para entender os efeitos entre X e Y, mas não nos permite falar de causalidade. Além disso, os modelos comumente utilizados tomam por base o arcabouço do modelo de regressão, como ficará claro com os exemplos trazidos aqui.

3.1.1.2 Regressão descontínua

Em um modelo de regressão descontínua (RD), – em inglês, refere-se comumente a esse modelo como RDD: *regression discontinuity design* –, a avaliação do tratamento é feita tomando por base uma distinção exógena em relação a uma variável contínua qualquer, criando um grupo de tratamento e outro de controle entre as observações. Um indivíduo ou uma unidade de análise se torna parte de um dos grupos de maneira independente de suas intenções. Se cada uma das unidades não tem capacidade de intervir na característica pela qual os grupos serão formados, a intervenção é exógena. Então, aproveita-se desse fato para observar as diferenças encontradas nas unidades que estão bastante próximas ao ponto de corte. Teoricamente, espera-se que essas unidades sejam semelhantes entre si e que a única diferença no resultado observado seja decorrência do tratamento. Para uma maneira de entender como se estrutura o modelo de RD, considere a figura a seguir:

Figura 4 – Representação de um modelo de regressão descontínua

Fonte: Elaboração do autor.

Na Figura 4, em ambos os gráficos, a relação entre X e Y é negativa: à medida em que X aumenta, Y diminui. No gráfico (a), a reta que se ajusta aos dados passa continuamente pelo corte atribuído ao tratamento, mostrando não haver diferenças entre o grupo de tratados e o de não tratados. Pode-se dizer que o gráfico (a) representa uma situação em que o tratamento não produziu efeitos. Já no gráfico (b) a linha de ajuste aos dados apresenta uma descontinuidade no ponto de corte. Essa diferença é um indicativo do efeito do tratamento.

O modelo de RD é comumente associado a duas interpretações distintas. Na primeira, a RD é pensada como uma “descontinuidade no ponto de corte” (HAHN; TODD; VAN DER KLAUW, 1999). De acordo com essa interpretação, a quebra entre as linhas à esquerda e à direita do ponto de corte dá características do efeito do tratamento, como direção e magnitude, para observações próximas ao corte. Na segunda interpretação, a RD é associada a uma aleatorização local (LEE, 2008). Esse processo de aleatorização sugere que o valor da variável Y para unidades que estão próximas ao ponto de corte é aleatório. Isso significa que se considera que os sujeitos próximos ao ponto de corte são em média idênticos e que a diferença apresentada no valor de Y está sujeita ao acaso.

Exemplo 13

Em um trabalho publicado em 2012, Avelino, Biderman e Barone buscam avaliar a importância da vitória em uma eleição para a prefeitura sobre o resultado do partido vencedor na eleição para deputado federal. Os autores empregam uma RD da seguinte maneira: consideram a porcentagem de votos recebidos pelos dois primeiros partidos em uma eleição para prefeito. Quanto menor essa diferença, mais semelhantes são os partidos em relação a sua força no município. Essa disputa tem influência sobre os resultados dos partidos nas eleições para deputado federal. Na observação entre as eleições de 2008 e 2010, os autores encontram efeito de 3,4% a 4,3%, a depender do tamanho da margem de vitória que consideram.

Há condições que precisam ser cumpridas para que a RD seja aplicada. Em primeiro lugar, a variável sobre a qual ocorre a divisão dos grupos não pode ser influenciada pelo próprio tratamento. Em segundo lugar, como já dito, o tratamento deve ser exógeno. As unidades de análise não podem ter condições de se autoatribuir o tratamento ou o controle. Em terceiro lugar, as unidades de cada lado do ponto de corte devem ser idênticas. Não pode haver diferenças em outras características a não ser na atribuição do tratamento. Essas características devem ser verificadas e somente com a sua confirmação é possível seguir com o uso da RD.

Ao implementar a RD, a pesquisadora deve testar o efeito de distintas magnitudes na variável X com relação ao ponto de corte entre os grupos. Considere o seguinte exemplo: de acordo com o desenvolvimento atual da teoria de Duverger sobre o número de partidos que disputam uma eleição, a teoria prevê que eleições com 2º turno apresentam número maior de partidos do que sistemas com eleição em um turno apenas. O Brasil apresenta uma situação interessante para esse caso (Izumi, 2016). Nas eleições municipais, os municípios do país que possuem segundo turno são aqueles com número de eleitores superior a 200 mil. O número de eleitores de uma cidade é uma variável contínua e na qual cada município não tem ingerência: a variação no tamanho do eleitorado depende de outros fatores.

Assim, é possível pensar que municípios de aproximadamente 200 mil eleitores sejam idênticos entre si e que seja aleatório a cidade apresentar quantidade pouco abaixo ou pouco acima desse número. Ou seja, a cidade não escolhe se terá 2º turno ou não; a atribuição do tratamento é, assim, aleatória. Porém, qual a diferença no número de eleitores que permite observar o efeito do tratamento, já que poderíamos considerar as cidades como homogêneas? Se considerarmos, por exemplo, cidades com um eleitor a mais e com um eleitor a menos do que o ponto de corte, o argumento de que são cidades idênticas seria bastante forte. Se considerarmos cidades com 50 mil eleitores a mais ou a menos, é possível pensar que esse argumento da semelhança perde força. Assim, há um claro *trade-off* aqui: quanto maior a distância ao ponto de corte que consideramos observações para realizarmos a comparação, maior o número de observações, mas menor a semelhança entre elas. Ao contrário, quanto menor a distância ao ponto de corte, maior a semelhança, mas menor o número de observações.

Por fim, vale observar que a reta estimada em um modelo de RD pode ser obtida por diferentes especificações. Elas podem ser paramétricas ou não paramétricas. Uma especificação usual é a de considerar polinômios do 4º grau. É comum, inclusive, que os trabalhos testem diferentes especificações da forma funcional da reta estimada como forma de testar a consistência dos resultados encontrados. Sendo questões voltadas à aplicação em si da técnica, não serão tratadas aqui.

3.1.1.3 *Propensity score matching*

A técnica de *Propensity score matching* (PSM) é bastante popular nos estudos a respeito de impactos de políticas públicas. Ela segue a mesma lógica de comparação de resultados entre um grupo de tratamento e um grupo de controle. Porém, por lidar com dados observacionais, a forma de constituição dos grupos se dá de maneira diferente. O modelo aqui se baseia no cálculo de um *score* que indica a probabilidade de que uma determinada unidade de análise seja tratada, dadas as suas características individuais. Esse *score* pode ser entendido como a probabilidade de que cada unidade seja tratada de acordo com o que se observa delas.

A partir dessa medida, tomam-se as unidades efetivamente tratadas e seus resultados são comparados com os das unidades sem o tratamento, mas com o mesmo *score* – consideradas aqui como controles.

Esse método tem como objetivo básico encontrar um grupo que não seja participante do tratamento, mas cujos membros desse grupo sejam similares aos tratados em todas as características observáveis anteriores ao tratamento. Assim, ao comparar os dois grupos com base nas características observáveis, quaisquer diferenças entre eles serão atribuídas ao efeito do tratamento. O PSM busca contornar um problema de autosseleção entre os indivíduos que participam de um ou de outro grupo, já que não é possível observar o mesmo indivíduo pertencendo aos dois grupos ao mesmo tempo. Para isso, é criado um grupo de controle cuja principal diferença está relacionada à não participação desses indivíduos no tratamento. O problema está em determinar grupos que sejam comparáveis entre si.

Para contornar esse problema, o PSM é baseado na seguinte estrutura: suponha um indivíduo i , cuja variável de interesse é Y . Associa-se a esses indivíduos uma variável *dummy* D cujos valores serão $D=1$, para os indivíduos que recebem o tratamento, e $D=0$, caso contrário. Y_{1i}^T é o valor observado para o indivíduo tratado e Y_{0i}^C refere-se aos indivíduos do grupo de controle. O impacto do tratamento sobre o indivíduo i é mensurado por:

$$\delta_i = Y_{1i}^T - Y_{0i}^C \quad (8)$$

E o efeito médio desse tratamento sobre todos os indivíduos seria:

$$\delta_i = E(Y_{1i}^T - Y_{0i}^C | D = 1) \quad (9)$$

A equação 9 é denominada na literatura de avaliação como “efeito do tratamento médio sobre os tratados” (em inglês, *average treatment effect on the treated*, ATT).

O problema no cálculo desse efeito decorre de não ser possível observar o resultado contrafactual, $E(Y_{0i}^C | D = 1)$, pois o indivíduo estará apenas em um dos grupos (nunca em ambos ao mesmo tempo). Por conta disso, é necessário encontrar outra maneira de estimar esse valor. A expressão $E(Y_{0i}^C | D = 0)$ não serve como substituta porque o

fato de os indivíduos se autosseleccionarem para o tratamento gera viés. A ideia do viés aqui decorre de que o tratamento já pode ser atribuído para um grupo, cujas características influenciarão o resultado, pois as distinguem dos demais, que optaram por não fazer o tratamento. Um exemplo seria imaginar que os funcionários de uma empresa que se dispõem voluntariamente a participar de um treinamento que melhore sua *performance* no trabalho já sejam aqueles com maior potencial. O viés ocorre em razão da diferença tanto das características observáveis quanto das não observáveis entre todos os envolvidos, seja no grupo de tratamento, seja no de controle.

Para resolver a possibilidade de existência do viés, deve-se considerar a hipótese de que, a partir de um conjunto de características observáveis X^{48} , os indivíduos têm a mesma probabilidade de serem associados a um dos grupos, de tratamento ou de controle. Com isso, pode-se afirmar que a seleção dos indivíduos se baseia somente em características observáveis. O PSM tem o objetivo exatamente de contornar esse problema. O método consiste basicamente em encontrar um grupo de comparação o mais similar possível em termos do *propensity score*, a partir de características observáveis dos indivíduos.

O argumento central⁴⁹ do PSM é o de que se os potenciais resultados do tratamento não dependem do grupo de indivíduos participantes condicionalmente às variáveis observáveis, X , os potenciais resultados também serão independentes do tratamento condicional. Portanto, é possível substituir X pelo *propensity score*. Esse pode ser colocado como a probabilidade condicional de uma unidade de análise ter recebido o tratamento, dados os valores de X . Se o tratamento for indicado pela variável D , com $D=1$ indicando o tratamento e $D=0$ indicando o não tratamento, a equação a seguir representa o valor do *score*:

⁴⁸ Deve-se dizer que há limites no número de variáveis introduzidas no conjunto X , já que a consideração de maior número de características observáveis dificulta a construção do pareamento.

⁴⁹ O argumento apresentado é originalmente desenvolvido por Rosenbaum e Rubin (1983).

$$s(x) = P(D = 1|X = x) = P(X) \quad (10)$$

A equação 10 mostra que a probabilidade de ocorrer o tratamento, dadas as características observáveis X , é igual à probabilidade de um indivíduo apresentar as características X . O valor $s(x)$ é um escalar, solucionando o problema das múltiplas dimensões de X , que se reduzem a um único valor⁵⁰. Assim, a estimativa do efeito médio do tratamento se torna uma probabilidade condicional ao tratamento e ao *score*. A equação a seguir apresenta essa estimativa:

$$E(Y_{1i}^T - Y_{0i}^C | D = 1, s(x)) = E(Y_{1i}^T | D = 1, s(x)) - E(Y_{0i}^C | D = 0, s(x)) \quad (11)$$

Cabe dizer que $s(x)$, o *score*, é comumente estimado por uma equação logística. As unidades são pareadas, então, entre unidades tratadas e unidades de controle de acordo com seu *score*. Se a variável que mensura o resultado for uma variável contínua, o efeito do tratamento pode ser obtido através da diferença entre a média do resultado encontrado nas unidades tratadas e a média das unidades de controle. Se a variável for discreta, o efeito do tratamento pode ser obtido pela diferença entre a proporção de unidades tratadas *versus* as de controle.

Exemplo 14

Duarte *et al.* (2009) utilizam esse método para realizar uma avaliação do impacto do programa Bolsa Família sobre o consumo de alimento das famílias beneficiadas. As variáveis explicativas utilizadas para a construção do pareamento, cujo resultado foi estatisticamente significativo a pelo menos 10%, foram: Renda familiar total, líquida dos benefícios, Idade do chefe da família, *Dummy* que indica se a família possui crianças até 17 anos de idade, *Dummy* que indica se o domicílio possui água encanada. Seus resultados indicam que os beneficiados direcionam 88% da renda anual recebida com o programa, ou seja, R\$ 246,31 do total de R\$ 278 com gastos de alimentação.

⁵⁰ De acordo com França e Gonçalves, “O *propensity score matching* é uma forma de *matching* baseado no *balancing score*, $b(X)$. O *balancing score* é uma função construída com base em todas as variáveis independentes relevantes de forma que $E(X|b(X))=0$, isto é, seja independente da escolha ou não para tratamento.” (2009, p.6)

É preciso observar que o pareamento entre as unidades tratadas e de controle ocorrerá se a probabilidade de o tratamento ocorrer, dado X não for igual a 0 ou a 1. Caso essa condição seja violada, ela produz viés, pois não haverá unidades comparáveis entre os grupos. O viés ainda se manterá caso características não observáveis influenciem a participação em um dos grupos e os resultados condicionais a X . Nesse sentido, como observam França e Gonçalves (2009, p. 7), “a extensão em que o viés é reduzido depende da riqueza e da qualidade das variáveis de controle com as quais o *propensity score* é calculado e o *matching* é realizado”.

3.1.1.4 Método de controle sintético

O método de controle sintético, como desenvolvido por Abadie, Diamond e Hainmueller (2010), é aplicado para os casos nos quais o tratamento ocorreu para uma única unidade e não se pode, portanto, falar de grupo de tratamento. Em sua apresentação, os autores trazem um argumento semelhante aos trabalhos que consideram os efeitos de uma política ou de um evento por meio da observação de um único caso. Buscam conjugar a lógica experimental para uma situação em que o tratamento ocorreu para uma única observação.

Qual é o efeito do tratamento de uma única unidade? Recuperamos a equação (2) acima, o efeito do tratamento é dado por:

$$\delta_i = (Y_i|X = 1) - (Y_i|X = 0) \quad (2)$$

No caso em questão, tem-se apenas uma única observação sujeita ao tratamento. Ou seja, $i = 1$, apenas. Além disso, deve-se considerar que $(Y_1|X = 1)$ é conhecido, já que a unidade $i = 1$ sofreu o tratamento. Não se conhece $(Y_1|X = 0)$, ou seja, não se sabe o valor de Y quando o tratamento não ocorreu. A tarefa está, então, em encontrar uma maneira de estimar esse valor.

Intuitivamente, o que os autores propõem é a criação de um controle sintético. Como esse controle não existe na prática, a ideia está em simular um controle a partir de observações de outras unidades

que não tenham sofrido o mesmo tratamento. Isso é feito a partir da comparação entre a observação tratada contra uma média ponderada das demais observações não tratadas. Adequando a notação para tornar a demonstração mais simples, a expressão $(Y_1|X = 0)$ será igual a Y_{1t}^N e $(Y_1|X = 1)$ será renomeada para Y_{1t}^I . Então, o efeito do tratamento seria dado por: $\hat{\delta}_{1t} = Y_{1t}^I - Y_{1t}^N$.

Para determinar, vamos assumir que vale a seguinte expressão:

$$Y_{it}^N = \alpha_t + \beta_t Z_i + \varepsilon_{it} \quad (12)$$

Em que α_t é um fator comum desconhecido para todas as unidades; Z_i é um vetor de variáveis explicativas observáveis⁵¹, β_t são parâmetros desconhecidos e ε_{it} são os termos de erro não observáveis, cuja média é igual a zero.

Para a construção do controle sintético, considere um vetor de pesos $W = (w_2, \dots, w_{J+1})'$, tal que $w_j \geq 0$, para $j = 2, \dots, J+1$ e $w_2 + \dots + w_{J+1} = 1$. O valor da variável resposta para cada controle sintético indexado por W é:

$$\sum_{j=2}^{J+1} w_j Y_{jt}^N = \alpha_t + \beta_t \sum_{j=2}^{J+1} w_j Z_i + \sum_{j=2}^{J+1} w_j \varepsilon_{it} \quad (13)$$

Vamos admitir quando o período pré-tratamento for longo o suficiente em relação à escala ε_{it} dos choques transitórios, temos como estimador do efeito do tratamento a relação:

$$\hat{\delta}_{1t} = Y_{1t}^I - \sum_{j=2}^{J+1} w_j^* Y_{jt}^N \quad (14)$$

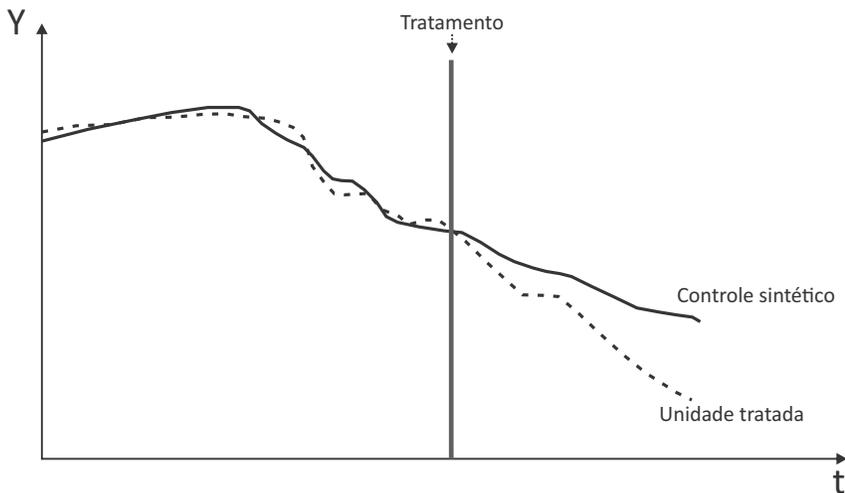
Em que w_j^* é um vetor de pesos que faz com que 13 exista. Isso por vezes ocorre apenas aproximadamente com os dados que se tem em mãos.

No Gráfico 2, representa-se uma relação possível entre o controle sintético e uma unidade tratada. Suponha que o gráfico represente uma intervenção que busque reduzir a incidência de uma doença em um

⁵¹ Seria possível incluir na equação variáveis não observáveis e, em termos práticos, o modelo se mantém o mesmo. Aqui, para efeitos de simplificação, decidiu-se manter a expressão similar a uma equação de regressão convencional.

estado qualquer do Brasil. No período anterior ao tratamento, tanto a unidade tratada quanto o controle apresentam trajetória semelhante em termos do número de casos (Y) ao longo do tempo. Em um dado instante do tempo, uma das unidades passa a adotar uma política que tem o intuito de reduzir a doença. A partir daquele momento, nesse caso, espera-se observar uma redução maior dos casos da doença em relação ao controle sintético, ainda que em ambos os casos a redução ocorra. Ou seja, em uma situação como essa, é possível argumentar que a política resultou nos efeitos no sentido esperado. É possível ainda implementar alguns outros testes de robustez, como o de supor que a política tivesse sido adotada em algum outro estado e comparada a evolução da incidência de doenças nessa unidade pseudotratada e o controle sintético. Um procedimento desse tipo possibilita que se avalie o impacto da política e, principalmente, da escolha do método sobre o resultado observado. Assim, ganha-se confiança de que o efeito do tratamento não decorre do acaso ou da escolha metodológica implementada.

Gráfico 2 – Representação do controle sintético



Fonte: Elaboração do autor.

Exemplo 15

O uso dessa metodologia ainda é recente no país. Os poucos trabalhos existentes se concentram na economia. Merece ser destacado como exemplo desse tipo de aplicação o trabalho de Corseuil *et al.* (2015), que avalia o efeito da adoção de pisos salariais por São Paulo e pelo Paraná no nível de emprego e na remuneração recebida em três setores de atividade. Os autores encontram impactos positivos em apenas um desses três setores, enquanto nos outros dois, os dados parecem indicar que a política foi inócua.

3.1.1.5 Difference-in-differences

O método também conhecido como *Diff-in-diff* tem sido amplamente utilizado nas Ciências Sociais como uma forma de avaliar o impacto de uma ocorrência exógena sobre as unidades em estudo. O nome se refere ao cálculo da diferença entre duas diferenças, com o intuito de observar o efeito de um tratamento em face do grupo de controle. A estrutura básica da análise se dá a partir da observação de determinada variável para dois grupos distintos em dois momentos temporais também distintos. Esses momentos são marcados pelo início do tratamento⁵² para um dos grupos. Ou seja, o grupo só passa a ser considerado sob efeito do tratamento no segundo momento, enquanto o outro grupo, chamado de controle, não recebe o tratamento em nenhum dos períodos. Nesse caso, não houve a atribuição aleatória dos indivíduos em cada grupo. A maneira de identificar o efeito do tratamento está em comparar a diferença entre o grupo de tratamento e de controle no primeiro momento do tempo, em que o tratamento não havia sido atribuído ainda, com a diferença entre os grupos no período posterior ao início do tratamento. Entende-se que, se garantida a comparabilidade entre os grupos, uma suposta nova diferença entre os grupos no segundo período (pós-tratamento), em relação à

⁵² Em geral, é importante observar se esse início é exógeno, ou seja, se os indivíduos puderam se autosselecionar. Problemas nesse aspecto precisam ser considerados.

diferença existente no período anterior ao início do tratamento, é efeito do próprio tratamento.

O exposto acima pode ser escrito da seguinte forma: seja Y a variável de interesse sobre a qual se deseja verificar o impacto de algum fenômeno – digamos que é a renda das famílias alvo do bolsa família. Pode-se dividir as unidades em dois grupos: tratamento (T) e controle (C). O horizonte temporal também é separado: há o momento anterior ao tratamento ($t=0$) e o momento posterior ao tratamento ($t=1$). Em sua estrutura mais simples, a construção do efeito é a seguinte: no primeiro momento do tempo, os dois grupos possuem alguma diferença em relação à Y , ou seja, $(Y_T - Y_C)|_{t=0}$. Essa é a diferença básica. A mesma diferença é calculada para o segundo período, pós tratamento: $(Y_T - Y_C)|_{t=1}$. O efeito do tratamento, δ , será dado por:

$$\delta = [(Y_T - Y_C)|_{t=1}] - [(Y_T - Y_C)|_{t=0}] \quad (15)$$

Se cada um dos grupos for composto por exatamente as mesmas unidades, o processo é exatamente como indicado acima em (15): o ganho do grupo de controle deve ser subtraído do ganho do grupo de tratamento. Essa diferença entre os grupos seria entendida como o efeito do tratamento. Esse caso, que pode ser entendido como um painel⁵³, não é o mais usual e por isso não será tratado aqui. Na sua forma mais comum, o pesquisador possui duas *cross sections*⁵⁴ repetidas em cada um dos períodos para cada grupo e, portanto, a mera diferença entre as médias do ganho de cada grupo não é suficiente. Os dados de cada um dos períodos podem ser diferentes entre si: os grupos de controle e tratamento podem ser distintos entre os momentos do tempo. Então, uma forma de controlar essa possibilidade de variação precisa ser considerada.

⁵³ Por painel se entende um banco de dados em que observações a respeito de um conjunto de unidades se repetem ao longo do tempo.

⁵⁴ Por *cross section* se entende um conjunto de dados aleatoriamente selecionados de uma população referente a determinadas unidades em um único momento do tempo.

De posse dessas *cross sections*, a formulação genérica para qualquer um dos membros de qualquer um dos grupos pode ser expressa da seguinte forma:

$$Y = \beta_0 + \beta_1 D_T + \delta_1 D_{t2} + \delta_2 D_T \cdot D_{t2} + \varepsilon \quad (16)$$

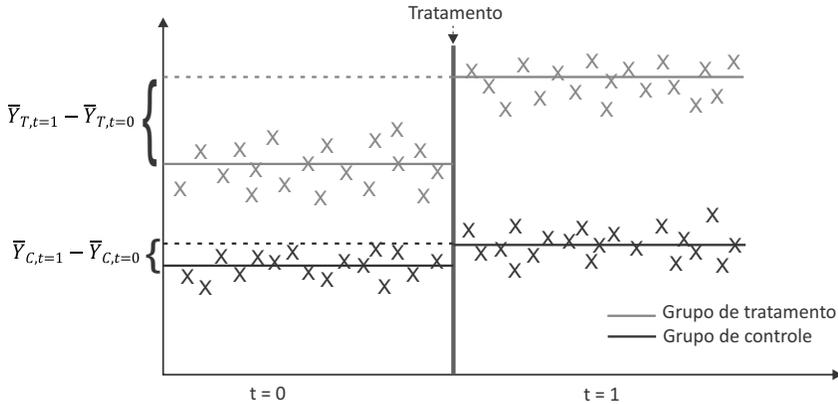
em que Y é a variável em que o efeito do tratamento se manifesta, D_T é uma variável *dummy* que indica se a observação é referente ao grupo de tratamento, D_{t2} é uma *dummy* que indica o segundo período, $D_T \cdot D_{t2}$ é a interação das duas *dummies*, que resulta em uma nova *dummy* indicando o grupo de tratamento no segundo período e $\beta_0, \beta_1, \delta_1, \delta_2$ são os parâmetros a serem estimados. A introdução da primeira *dummy* ocorre para que se distingam os dois grupos e se capturem efeitos decorrentes de diferenças dos próprios grupos a despeito do tratamento; a da segunda *dummy* tem a função de capturar variações que ocorreram entre os dois períodos e afetem ambos os grupos independentemente do tratamento; já a terceira *dummy* é a variável de interesse aqui. Seu parâmetro, δ_2 , é aquele que indica o efeito do tratamento e cuja estimativa pode ser escrita como:

$$\hat{\delta}_2 = (\bar{Y}_{T,t=1} - \bar{Y}_{T,t=0}) - (\bar{Y}_{C,t=1} - \bar{Y}_{C,t=0}) \quad (17)$$

Deve-se notar na expressão (17) que o parâmetro estimado, $\hat{\delta}_2$, depende dos valores médios, \bar{Y} , de cada um dos grupos, (T) e (C), nos diferentes momentos do tempo. Ou seja, o que se obtém como resultado do modelo aqui é um efeito médio do tratamento sobre o grupo em que é atribuído. É, portanto, importante que se avalie de que maneira os grupos de tratamento e controle podem estar desbalanceados, o que em um modelo como o de painel referenciado em (15) não seria um problema. A estimação de $\hat{\delta}_2$ se dá por um modelo de regressão linear convencional.

Graficamente, pode-se apresentar o resultado de um modelo como esse de acordo com a Figura 5:

Figura 5 – Representação de um modelo de Diff-in-Diff



Fonte: Elaboração do autor.

Na Figura 5, destacam-se as observações feitas para os dois grupos de tratamento e de controle antes e depois do início do tratamento. As barras horizontais indicam as médias em cada um dos períodos, para cada um dos grupos. Nota-se que o grupo de controle também apresentou variação na média de Y entre os dois períodos. Algum fator exógeno ao tratamento provocou essa variação. Por essa razão, a simples observação da variação da média de Y no grupo de tratamento entre os dois períodos não indica o efeito do tratamento. $\bar{Y}_{C,t=1} - \bar{Y}_{C,t=0}$ poderia ser grande o suficiente para fazer com que o efeito do tratamento fosse nulo, ou seja, $(\bar{Y}_{T,t=1} - \bar{Y}_{T,t=0}) = (\bar{Y}_{C,t=1} - \bar{Y}_{C,t=0})$, e portanto, $\hat{\delta}_2 = 0$.

Exemplo 16

Nishijima *et al.* (2011) investigam se o novo marco regulatório introduzido pela Lei nº 9.656/1998 afetou o comportamento dos portadores de planos de saúde privados em relação ao consumo de bens de saúde. Os autores identificam, inicialmente, que os segurados utilizam maior quantidade de bens médicos que os não segurados. Além disso, apontam que a legislação teve efeito nulo sobre o consumo de serviços médicos.

Esses desenhos exemplificam as possibilidades de utilização da pesquisa com N-grande e da lógica experimental a dados observacionais. As maneiras de estimar os parâmetros apontados seguem, via de regra, um modelo baseado em uma regressão de mínimos quadrados ordinários, que são discutidos em um outro volume desta coleção e foram sucintamente tratados aqui. Novamente, deve-se enfatizar alguns aspectos abordados anteriormente, que merecem consideração:

Número de observações em cada um dos grupos. Por vezes, o grupo de tratamento possui apenas uma observação; por vezes, muitas. O mesmo pode ocorrer com o grupo de controle. O leitor deve considerar essas diferenças nos desenhos acima. Por exemplo, o controle sintético é adaptado para uma única observação no tratamento. Essas condições importam e a pesquisadora pode considerar estratégias para a expansão do número de observações. Brambor e Ceneviva (2012), por exemplo, consideram pares de candidatos que disputam a eleição para prefeito nos municípios brasileiros. Em sua base de dados, são considerados os pares candidato A vs. candidato B e vice-versa. Isso os ajuda a encontrar maior número de observações.

Momento da medição. No cenário ideal, as medidas são feitas antes e depois do tratamento para permitir a comparação entre os grupos de tratamento e controle. Inclusive, deve-se atentar para que a mensuração seja feita depois que o tratamento tenha obtido efeito. Porém, como não há controle sobre quando são executadas, as medições restringem as alternativas de desenhos;

Número de medições. Para além do momento em que são realizadas, também é importante considerar o número de medições antes e depois do tratamento. Por exemplo, no desenho de regressão descontínua, a medição é feita em um único momento do tempo; já no controle sintético, são feitas diversas medições antes e depois do tratamento.

Esses aspectos são marcadores das diferenças entre os desenhos quase-experimentais. Como o pesquisador não possui condições para a construção dos dados desde o início da pesquisa e terá de lidar com medições realizadas em circunstâncias não ideais em termos da

construção de um experimento, essas particularidades se colocam como limitadoras das opções reais de desenhos de pesquisa. Porém, deve estar claro que essas opções aqui ocorrem quando há número elevado de observações. Seja no controle ou no tratamento, antes ou depois do tratamento, o N é grande. Mas, como já dito, há também as pesquisas que lidam com N-pequeno, em que o número total de observações é baixo. É para pesquisas desse tipo que nos voltamos na próxima seção.

3.2 Pesquisa com N-pequeno

A pesquisa com N-pequeno inverte a estrutura das perguntas de pesquisa em relação às pesquisas com N-grande e, assim, o tipo de explicação que busca prover. As perguntas estão voltadas ao “efeito das causas”, e esse procedimento permite que a pesquisa trate da explicação de casos individuais. As perguntas passam a ser construídas com o intuito de identificar qual a consequência de uma determinada ocorrência particular. Por exemplo, a questão de pesquisa pode ser feita como: qual o efeito da adoção de determinada política pública sobre as condições de vida da população? Uma pergunta desse tipo conduz a um desenho de pesquisa também diferente. Nesse sentido, a pesquisa será organizada de forma a fazer a avaliação da relação de determinada política pública que de fato aconteceu. Assim, o possível fator causal de alguma alteração em um objeto específico é o centro da pesquisa, e essa posição inicial será decisiva para a maneira como a pesquisa será conduzida, dados os princípios que norteiam esse tipo de questionamento. “Um propósito central da pesquisa é o de identificar as causas de resultados específicos para cada um dos casos que estão no escopo da teoria sob investigação” (MAHONEY; GOERTZ, 2006, p. 230). A partir dessa maneira de explicar, as teorias, deveriam ser capazes de lidar com os resultados de todos os casos da população.

Entretanto, essa caracterização da pesquisa de N-pequeno não é suficiente para aglutinar toda a sua variedade de métodos. O rol de técnicas e de procedimentos que podem se enquadrar nesse tipo de