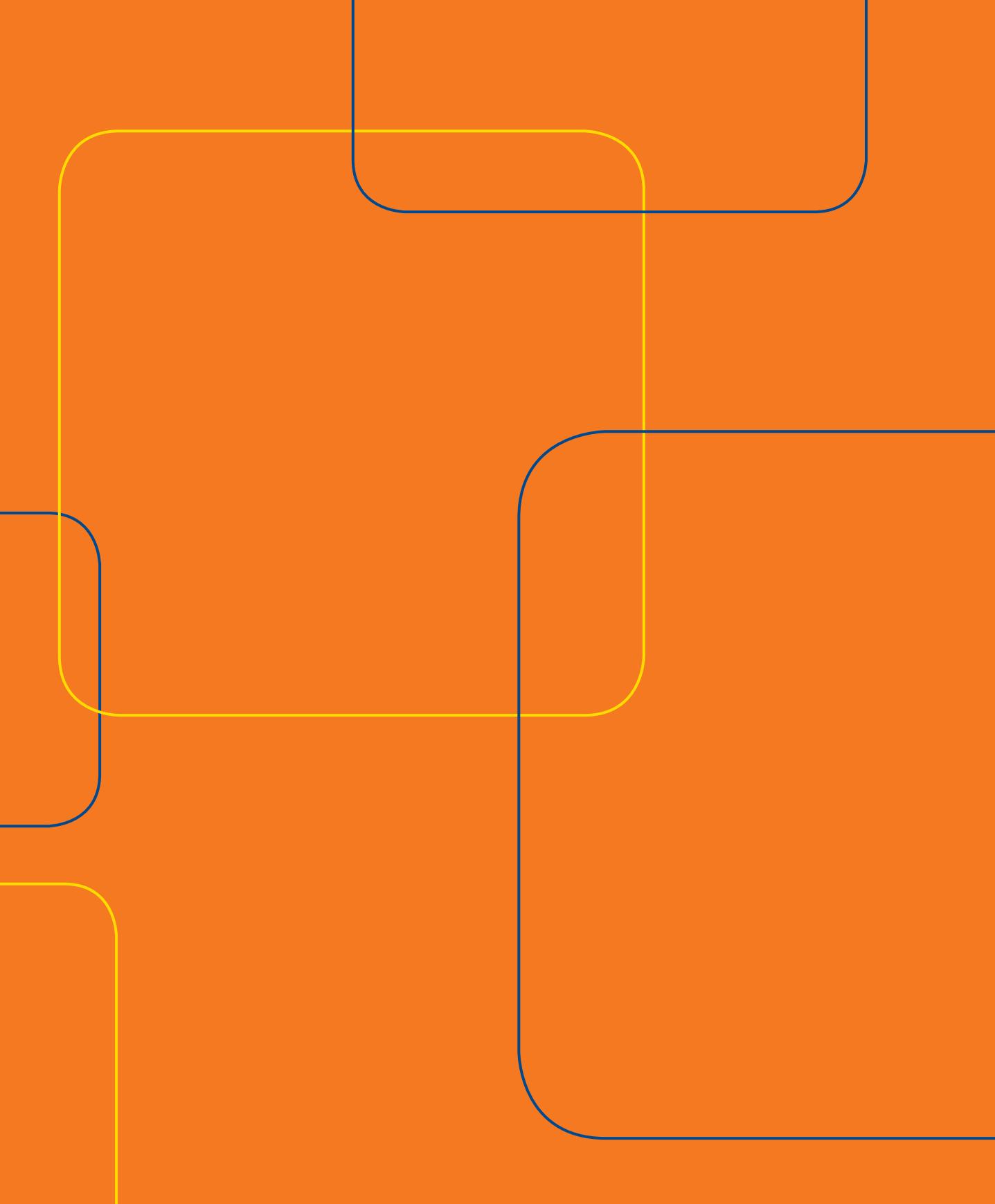


Avaliação Econômica de projetos sociais

Itaú Social



Avaliação Econômica de projetos sociais

São Paulo - SP
Itaú Social
2017



Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Avaliação econômica de projetos sociais /
[organizadores Naercio Aquino Menezes Filho,
Cristine Campos de Xavier Pinto]. -- 3. ed. --
São Paulo : Fundação Itaú Social, 2017.

Vários autores.

Bibliografia.

ISBN 978-85-66932-31-7

1. Avaliação de programas de ação social
 2. Avaliação econômica
 3. Cálculos
 4. Investimento social
 5. Retorno de investimentos
- I. Menezes Filho, Naercio Aquino. II. Pinto, Cristine Campos de Xavier.

17-10611

CDD-361.610981

Índices para catálogo sistemático:

1. Brasil : Avaliação econômica : Projetos sociais :
Bem-estar social 361.610981
2. Brasil : Projetos sociais : Avaliação econômica :
Bem-estar social 361.610981

Ficha Técnica

Iniciativa

Itaú Social

Vice-presidente

Fábio Colletti Barbosa

Superintendente

Angela Cristina Dannemann

Gerente de Pesquisa e**Desenvolvimento**

Patrícia Mota Guedes

Apoio

Alan Albuquerque Ribeiro Correia

Carlos Eduardo Garrido

Clérito Leonardo Moraes Rossati

Esmeralda Correa Macana

Karen Dias Mendes

Marina Brito Ferraz

Créditos da Publicação**Organizadores**

Naercio Aquino Menezes Filho

Cristine Campos de Xavier Pinto

Autores

Betânia Peixoto

Cristine Campos de Xavier Pinto

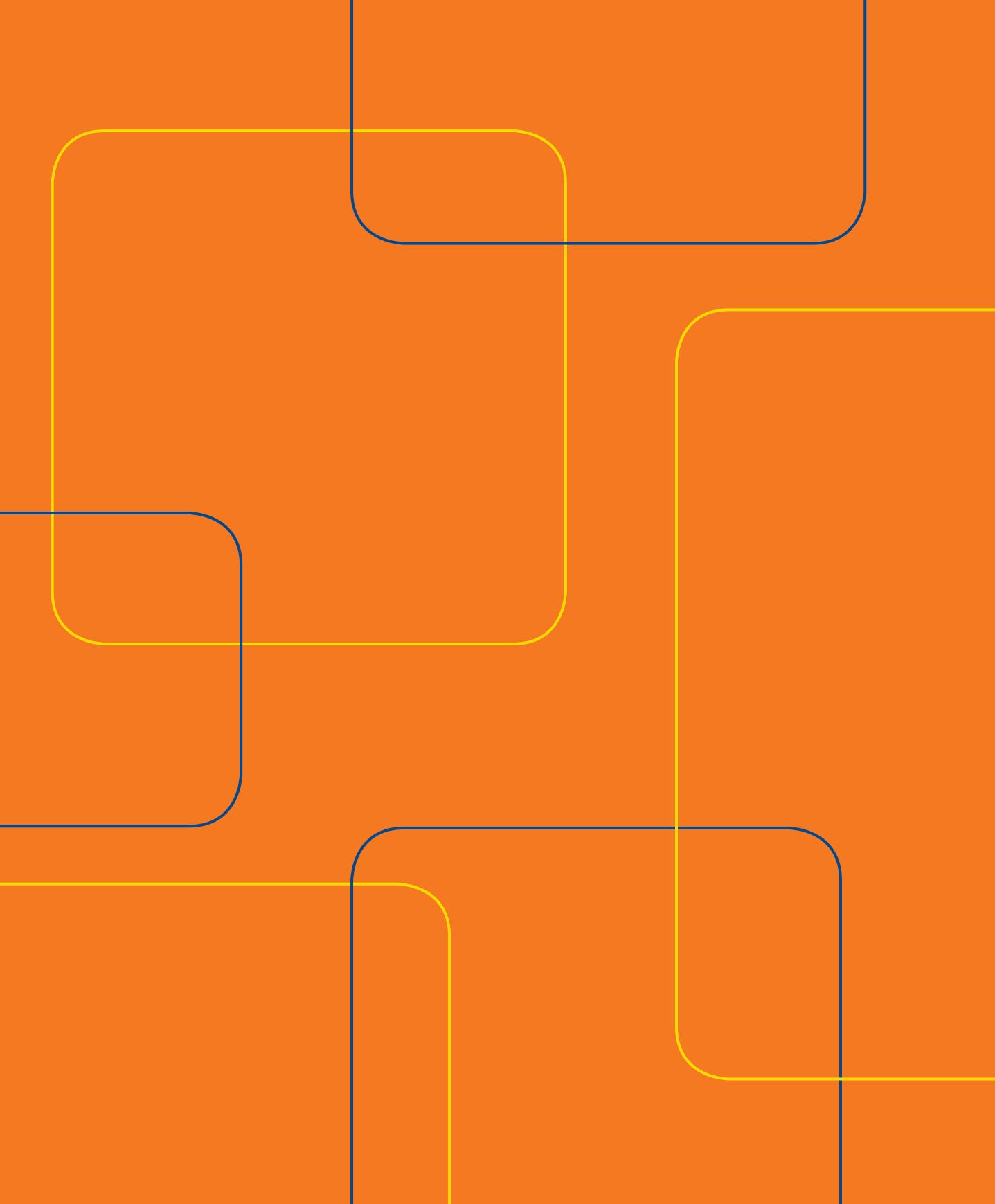
Lycia Lima

Miguel Nathan Foguel

Ricardo Paes de Barros

Projeto Gráfico e Diagramação

theSign



ÍNDICE

Apresentação	8
Prefácio	9
1. Avaliação de Impacto de Programas Sociais. Por que, para que e quando fazer?	13
2. Modelo de Resultados Potenciais	39
3. Método de Aleatorização.....	55
4. Diferenças em Diferenças	85
5. Pareamento.....	111
6. Variáveis Instrumentais	145
7. Regressão Descontínua	167
8. O Cálculo do Retorno Econômico	193
Referências Bibliográficas.....	226
Apêndice.....	237

APRESENTAÇÃO

O Itaú Social tem como atividade central a formulação, implementação e disseminação de tecnologias sociais que contribuam para a melhoria da educação pública brasileira. Sua atuação acontece em todo o território brasileiro, em parceria com governos, setor privado e organizações da sociedade civil.

O trabalho do Itaú Social na área de avaliação de impacto de projetos sociais iniciou-se em 2004. Desde então, são desenvolvidas diversas ações com o objetivo de fomentar a cultura da avaliação. Pautadas pela convicção de que essa prática constitui ferramenta fundamental de gestão no campo social, há mais de uma década, o Itaú Social oferece seminários, cursos e outros eventos formativos para gestores públicos, organizações da sociedade civil, fundações e institutos empresariais, assim como para alunos de graduação e pós-graduação de todo o país.

Esta publicação integra a estratégia de disseminação do conceito e da prática da avaliação de projetos sociais, facilitando o acesso a esse arcabouço metodológico. Ao apresentar diferentes exemplos de avaliação de iniciativas brasileiras, evidencia-se que a avaliação tem sido assumida, nos últimos anos, como elemento-chave para o aprimoramento contínuo, a aprendizagem e a prestação de contas à sociedade.

Com essa iniciativa de difundir as metodologias de avaliação, esperamos colaborar para o aperfeiçoamento dos projetos sociais e o aumento da eficiência das políticas sociais do nosso país.

Em nome do Itaú Social, agradeço aos autores e aos organizadores deste livro por sua dedicação e empenho, convencida da importância de difundir este conhecimento para a melhoria da ação social no Brasil.

Angela Cristina Dannemann

Superintendente do Itaú Social

PREFÁCIO

Há meros 15 anos, a avaliação econômica de projetos sociais, tema deste livro, era uma área praticamente inexplorada por economistas e cientistas sociais no Brasil. Os primeiros estudos brasileiros nesse campo, avaliando o impacto de programas de treinamento e emprego, foram realizados no início deste século. A atuação do Itaú Social na área de avaliação de impacto e retorno econômico iniciou-se em 2004, quando o seu programa “Raízes e Asas” foi avaliado economicamente.

Desde então, o Itaú Social oferece vários cursos para gestores de organizações não governamentais, do setor público e de fundações empresariais, além de seminários internacionais e regionais e cursos avançados para alunos de pós-graduação de todo o país. A partir do interesse pelos cursos, que quase sempre geram lista de espera, surgiu a ideia de lançar uma publicação que tratasse do tema com amplitude e rigor, para atingir o público que tem interesse pelo assunto, mas que não tem a possibilidade de presenciar os cursos e seminários. Dessa ideia nasceu este livro, publicado originalmente em 2012. Com os rápidos avanços nas técnicas disponíveis para avaliação de impacto, decidimos lançar uma nova edição com algumas revisões no texto da edição original.

A preocupação com a avaliação tem ganhado força uma vez que a sociedade brasileira sente a necessidade de usar seus recursos na área social da melhor forma possível. Afinal, vários programas e políticas em diversas áreas são lançados todos os anos nos vários níveis da administração pública, e existe a necessidade de sabermos o real efeito desses programas na vida das pessoas. Da mesma forma, o trabalho desenvolvido no terceiro setor, principalmente na área social, precisa prestar contas à sociedade e a seus financiadores, mostrando que os recursos estão sendo gastos da melhor forma possível.

Mas, o que significa avaliar economicamente os projetos sociais? Como os autores mostram nos capítulos que se seguem, a avaliação econômica é composta de duas partes: avaliação de impacto e cálculo do retorno econômico. A avaliação de impacto usa ferramentas estatísticas para estimar o efeito de um programa sobre os seus beneficiários, ou seja, se o programa atingiu ou não seus objetivos iniciais e qual a magnitude desse efeito. Para que essa análise seja feita de forma objetiva, a avaliação necessita ter indicadores objetivos e mensuráveis, que possam ser utilizados para aferir o sucesso do programa. Além disso, é necessário que haja um grupo de controle, ou seja, pessoas, escolas ou regiões que não foram atendidas pelo programa e que são parecidas com o grupo que sofreu a intervenção, chamado, seguindo a tradição da literatura médica, de grupo de tratamento.

Mas, para que a avaliação seja completa, é necessário também saber se os benefícios gerados a partir do impacto estimado superam os custos do programa, ou seja, se o projeto vale a pena do ponto de vista econômico. Muitas avaliações quantitativas existentes não chegam a essa parte, concludendo a avaliação com a estimativa do impacto. Entretanto, o cálculo do retorno econômico é essencial para verificar se os custos do programa são altos a ponto de inviabilizar a replicação do projeto. Além disso, essa metodologia permite o cálculo da taxa interna de retorno de diferentes projetos, que podem assim ser comparados. Para esse fim, são utilizadas técnicas simples de matemática financeira, que são muito usadas para calcular o retorno de outros tipos de investimentos.

Os autores deste livro estão entre os maiores especialistas em avaliação do país. No capítulo 1, à guisa de introdução, Ricardo Paes de Barros e Lycia Lima discutem por que a avaliação de impacto é necessária. Os autores explicam que a avaliação conduzida de forma rigorosa pode mostrar se os resultados esperados foram efetivamente alcançados, dada a imperfeição com que diversos programas são colocados em prática. Além disso, projetos que tiveram impacto comprovado para um grupo específico ou região podem ter um resultado completamente diferente quando implementados em outra região ou grupo de pessoas ou até em outro período de tempo.

Após a introdução, a primeira parte do livro trata dos métodos básicos de avaliação de impacto. No capítulo 2, Miguel Nathan Foguel introduz o modelo básico que permeia todas as metodologias de avaliação de impacto, chamado “Modelo de Resultados Potenciais”. Em seguida, no capítulo 3, Miguel explica a teoria por detrás do procedimento que é considerado o padrão-ouro nas avaliações de impacto, no qual os grupos de tratamento e controle são escolhidos aleatoriamente, ou seja, por sorteio. Mas, como na grande parte dos casos práticos a escolha dos tratados não se dá de forma aleatória, a maior parte do livro foca as metodologias não experimentais. Assim, no capítulo 4, Miguel explica a metodologia de diferenças em diferenças, que pode ser utilizada quando possuímos informações sobre a variável de resultado antes e depois da realização do programa. Por fim, no capítulo 5, Cristine Campos de Xavier Pinto explica o método de pareamento, o mais utilizado na prática pelas avaliações de projetos sociais, que procura comparar grupos de tratamento e controle que são parecidos em suas características.

A segunda parte do livro apresenta métodos mais sofisticados de avaliação de impacto. No capítulo 6, Cristine explica o método de variáveis instrumentais, que lida com o fato

de que a participação no programa pode ser explicada por fatores que não são observados pelo avaliador. Em seguida, a mesma autora apresenta o método de regressão descontínua, bastante utilizado atualmente em artigos acadêmicos, que é apropriado para os casos em que a probabilidade de uma pessoa ou escola ser tratada muda drasticamente dependendo do valor alcançado por uma variável específica.

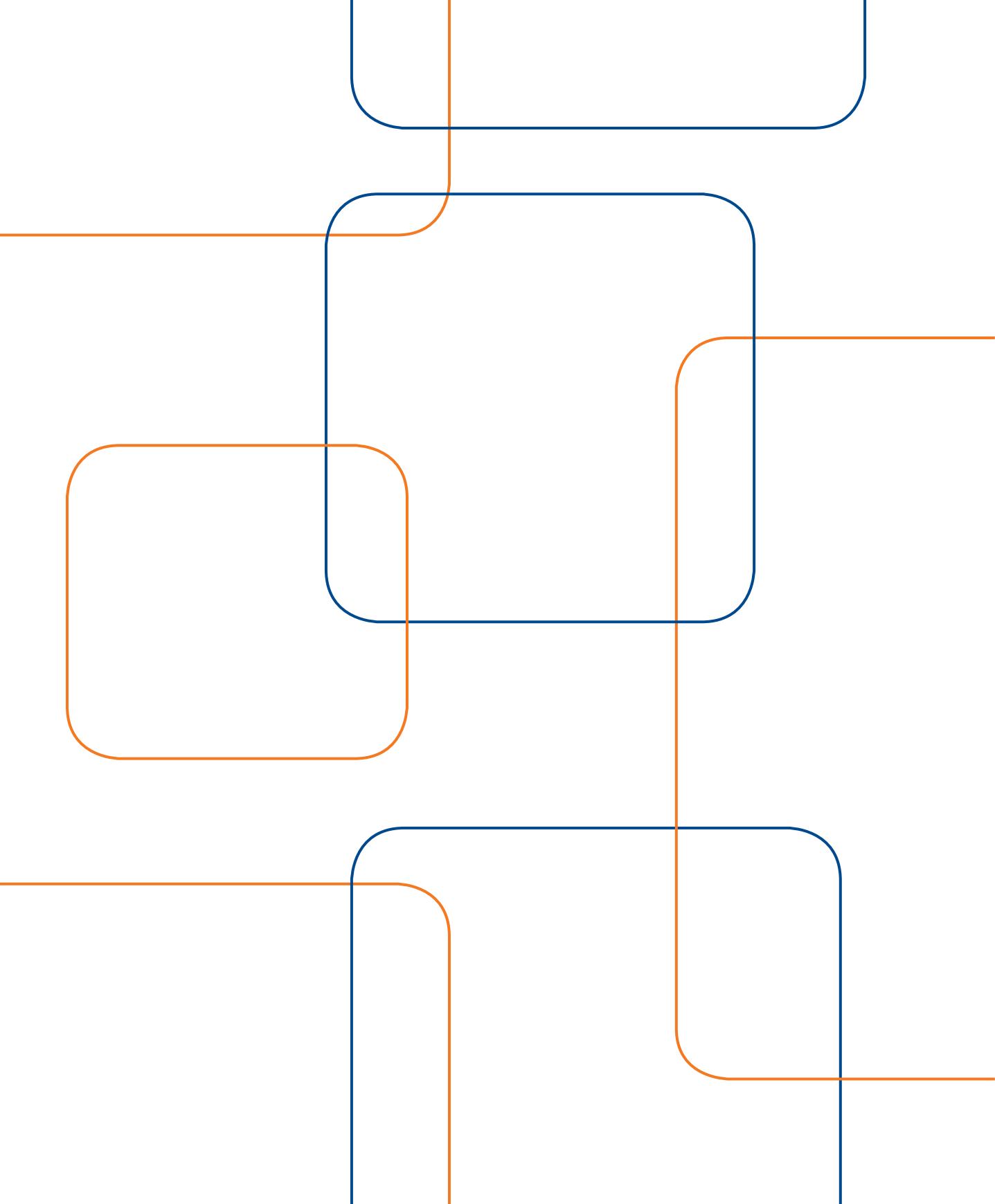
A terceira parte do livro lida com o cálculo do retorno econômico de um projeto social. No capítulo 8, Betânia Peixoto explica detalhadamente os conceitos e o instrumental necessários para o cálculo desse retorno, incluindo aplicações práticas e um exemplo para que os conceitos sejam sedimentados. Conceitos como valor presente líquido, taxa interna de retorno e custo-efetividade são discutidos nesse capítulo.

Este livro foi planejado tendo em vista o público formado por alunos de graduação de cursos na área de ciências sociais, que tenham passado por um bom curso de estatística básica e que tenham interesse por projetos sociais. Gestores de entidades do terceiro setor, fundações empresariais e dos vários níveis do governo também acharão o conteúdo bastante estimulante. Além disso, o instrumental desenvolvido nos capítulos avançados é ideal para os interessados em se aprofundar na metodologia.

Cada capítulo do livro traz exercícios específicos que ajudam a fixar o conteúdo, e a parte final traz exercícios práticos para serem resolvidos pelos leitores. As bases de dados e as respostas desses exercícios estão disponíveis na página da Fundação Itaú Social na internet. O apêndice comenta uma bibliografia complementar nas áreas de estatística e matemática financeira que pode ser utilizada como auxílio para entendimento dos capítulos do livro.

Tenho certeza de que esta nova edição do livro vai continuar estimulando a realização de avaliações de impacto, para que possamos conhecer e replicar os projetos sociais e as políticas públicas com maior retorno econômico, de forma a gerar o máximo de valor para a sociedade brasileira.

Naercio Menezes Filho



Capítulo 1

Avaliação de Impacto de Programas Sociais

Por que, para que e quando fazer?

Ricardo Paes de Barros
Lycia Lima

1.1. Introdução

O propósito do programa social é contribuir para que um objetivo coletivo seja alcançado. A definição de objetivos coletivos não é uma tarefa fácil, já que as escolhas coletivas são agregações de escolhas individuais. Escolhas individuais conscientes tipicamente requerem um acúmulo considerável de informação e bastante reflexão. Além de todos os requisitos informacionais e de racionalidade comuns à escolha individual, a escolha coletiva requer certa simultaneidade na formulação das opiniões individuais e ainda a existência de um processo para consulta e agregação dessas opiniões. Não é surpreendente, portanto, que a definição de objetivos coletivos demande esforço e tempo. Um exemplo notável foi o acordo mundial alcançado em 2015 sobre os 17 Objetivos de Desenvolvimento Sustentável (ODS) e suas 169 metas – foram mais de quatro décadas de trabalho, depois de considerado pela primeira vez em Estocolmo¹.

Uma vez definidos, objetivos coletivos, invariavelmente, são traduzidos em alguma forma de legislação, algum tipo de contrato social, que estabelece formalmente o compromisso com os objetivos escolhidos. Tipicamente um objetivo específico envolve a modificação de algum resultado de interesse ou seu progresso em direção a uma meta pré-definida. Por exemplo, a 1ª meta do 1º Objetivo de Desenvolvimento Sustentável consiste em assegurar que, até 2030, todas as pessoas, em todo o mundo, vivam em famílias com renda *per capita* de ao menos US\$ 1,25 por dia². Essa meta, portanto, pretende modificar o indicador de extrema pobreza, zerando-o até 2030.

Após a definição de objetivos coletivos e sua especificação em legislações, ou alguma forma de contrato social, programas sociais são criados para modificar os resultados de valor ou de interesse, implícitos nesses objetivos. É, em princípio, a busca por alcançar modificações nesses determinados resultados que justifica o desenho e a implementação de programas sociais.

1. Esse acordo começou a ser concebido em 1972 em Estocolmo, foi impulsionado pela Rio-92 e principalmente pelo documento “O Futuro Que Queremos” da Rio+20 que lançou as bases para a instituição do Grupo de Trabalho Aberto – GTA. Em 18 meses de trabalho extremamente intenso (entre março de 2013 e agosto de 2014) o GTA foi capaz de organizar uma proposta com 17 objetivos e submetê-la a Assembleia Geral da ONU. Essa proposta ainda passou por um ano de consultas e avaliação pelos países-membros até sua aprovação final em agosto de 2015, mais de quatro décadas depois de considerada pela primeira vez em Estocolmo.

2. Famílias que vivem com menos de US\$ 1,25 por dia estão abaixo da linha de extrema pobreza, definida pelo Banco Mundial e adotada pela ONU.

Para alcançar o progresso almejado nos resultados de interesse, programas sociais em geral recorrem à produção e à oferta de uma série de bens e serviços, frequentemente denominados de produtos. O acesso e a utilização, pelos beneficiários do programa, desses produtos deveria levar inexoravelmente à modificação do resultado de valor ou de interesse que o programa foi desenhado para modificar – tal conclusão deveria ser advinda da teoria de mudança em que o desenho do programa foi baseado. Exceto caso a teoria de mudança que tenha fundamentado o desenho do programa seja plenamente comprovada, o sucesso de um programa não pode ser medido simplesmente por sua capacidade de entregar os produtos que deveria oferecer. É preciso verificar diretamente se o programa foi efetivamente capaz de transformar o resultado de valor ou de interesse que pretendia modificar e se essa transformação foi da magnitude esperada.

Dessa afirmação, pode-se concluir dois pontos importantes, simples, mas que rotineiramente costumam ser esquecidos. Em primeiro lugar, todo programa social precisa estar muito bem definido no que se refere aos resultados de interesse coletivo que foi desenhado para modificar. Embora seja evidente que programas sociais devem ser desenhados para alcançar objetivos coletivos, não é incomum serem desenhados e implementados em função das atividades intrínsecas que realizam. Frequentemente, os resultados que o programa busca modificar são apenas explicitados quando, por exemplo, considera-se realizar uma avaliação. Nesse caso, há uma total reversão de papéis. Em princípio, é o desenho dos programas sociais que deve se adequar aos progressos desejados nos resultados de interesse coletivo e não progressos nos resultados de interesse coletivo que devem se adaptar ao desenho de programas sociais, intrinsecamente definidos a partir de outros critérios.

Em segundo lugar, não é suficiente definir quais os resultados de interesse coletivo que o programa foi desenhado para modificar. É necessário também especificar qual será a magnitude da contribuição desse programa para a modificação que se deseja provocar nos resultados de interesse. Assim, não basta estabelecer que um programa social deva contribuir para a erradicação da extrema pobreza (1^a meta do 1º ODS), é necessário também definir a magnitude dessa contribuição. Afinal, a justificativa da sua existência depende tanto do seu custo quanto da magnitude dessa contribuição: em princípio, para ser justificável, o valor da contribuição (o benefício) de um programa deve ser superior a seu custo – este é o princípio básico de qualquer análise de custo-benefício.

Uma vez definidos os objetivos coletivos e estabelecidas as legislações que formalizam esses compromissos, não é nada surpreendente que os resultados de interesse presentes nesses objetivos passem a ser cuidadosamente monitorados. É também esperado que, com vistas a alcançar tais objetivos, invariavelmente, mais de um programa seja desenhado e implementado.

Como o sucesso coletivo em alcançar objetivos é determinado pela magnitude da mudança alcançada *vis-à-vis* as metas estabelecidas, para avaliá-lo, um bom sistema de monitoramento é suficiente. É possível avaliar o cumprimento desses objetivos verificando se os resultados de interesse estão se modificando na direção desejada e à velocidade desejada. Assim, não há justificativa para avaliações de impacto quando a única preocupação coletiva é o cumprimento de objetivos coletivos.

Entretanto, o progresso de um determinado resultado de interesse pode ser consequência de vários fatores. Assim, o sucesso em relação ao alcance de dado resultado não significa que a contribuição de cada um dos programas implementados com o objetivo de promover progresso nesse resultado específico tenha sido significativa. Ou seja, o cumprimento de determinado objetivo não indica que os programas desenvolvidos para alcançar tal objetivo tenham sido bem-sucedidos. Da mesma forma, o fracasso social com relação ao progresso em dado resultado de interesse não significa que a contribuição de cada um dos programas desenhados e implantados com o objetivo de promover o progresso nesse resultado tenha sido insignificante.

Como toda parte de um todo, a contribuição de um programa social é apenas uma contribuição com limitado poder de definir a direção ou a velocidade do progresso total no resultado de interesse. Ou seja, um bom sistema de monitoramento, cujo objetivo maior é conhecer o que ocorreu com o todo, pode ser dramaticamente pouco informativo sobre a contribuição de cada programa isoladamente. Afinal, em geral, a sociedade não quer apenas saber se foi ou não capaz de alcançar seus objetivos coletivos. Mesmo tendo alcançado a meta, deseja-se saber se seria possível alcançá-la com número menor de programas e volume menor de recursos. Caso não se tenha cumprido a meta, deseja-se saber quais programas tiveram desempenho abaixo do esperado e, portanto, deveriam ser redesenhados ou descontinuados.

Para responder a essas questões, a simples disponibilidade de um sistema de monitoramento, por melhor que seja, não é suficiente. Para responder a essas perguntas, é necessário medir a contribuição de cada um dos programas ao progresso alcançado no resultado de interesse. É necessário,

assim, quantificar a contribuição de cada programa para cada resultado de interesse coletivo que, supostamente, este programa deveria influenciar. Para tal, é preciso recorrer a avaliações de impacto.

Assim, a avaliação do impacto de programas sociais tem precisamente a missão de isolar a contribuição de cada programa. Enquanto o monitoramento busca verificar se os resultados de interesse estão sendo alcançados e a que velocidade isso está ocorrendo, a avaliação de impacto visa a identificar e isolar a contribuição de cada programa para esse sucesso. Um programa pode, então, ter uma péssima avaliação de impacto em um ambiente de extremo sucesso social; ao mesmo tempo, outro pode ter uma ótima avaliação de impacto em um ambiente de pleno fracasso social. Afinal, em uma avaliação de impacto, o que importa não é a evolução do todo e sim a contribuição de cada parte.

Toda avaliação de impacto tem, portanto, uma natureza egocêntrica. Em um ambiente em que vários programas e ações contribuíram para alcançar um resultado de grande interesse coletivo, e em um momento em que todos deveriam estar celebrando o resultado alcançado pelo conjunto de programas, a avaliação de impacto procura isolar e quantificar a contribuição individual de um dos programas. Uma avaliação de impacto pode buscar responder qual a contribuição do programa na presença de todos os demais; pode, também, buscar responder qual teria sido o progresso no resultado de interesse, caso esse tivesse sido o único programa implantado.

Para uma avaliação de impacto, o contexto definitivamente importa, assim como a presença de outros programas complementares e substitutos. Mas o contexto e a presença de outros programas importam apenas conforme influenciam a contribuição do programa que está sendo avaliado para o progresso no resultado de interesse coletivo que foi desenhado para modificar.

1.2. O que significa impacto de um programa social?

Como definido na seção anterior, uma avaliação de impacto tem como objetivo estimar a magnitude da contribuição de um programa para o progresso de dado resultado de interesse, isto é, uma avaliação de impacto pretende indicar qual parcela do progresso alcançado foi consequência do impacto do programa. Mas o que significa exatamente o impacto ou a contribuição de um programa para o progresso em dado resultado?

O impacto de um programa deve ser entendido como a diferença entre o que aconteceu e o que teria acontecido na ausência de um programa. Mas o que significa aconteceu ou teria acontecido? Afinal aconteceu o quê e com quem? O impacto de um programa social é sempre definido com relação a dado resultado de interesse (por exemplo, taxa de mortalidade infantil) e dado grupo de referência (por exemplo, municípios beneficiados por dado programa). Assim, a magnitude do impacto de um programa é entendida como a diferença entre o resultado alcançado pelo grupo de referência na presença do programa e o resultado que esse mesmo grupo teria alcançado na ausência desse programa. Para tal, supõe-se implicitamente que o grupo de referência participou do programa. Essa hipótese, no entanto, não é necessária: em termos gerais, a magnitude do impacto pode ser definida como a diferença entre o resultado que o grupo de referência alcançou ou teria alcançado na presença do programa e o resultado que esse mesmo grupo alcançou ou teria alcançado na ausência desse programa.

Essa definição tem dois componentes essenciais: as duas situações analisadas referem-se a resultados alcançados pelo mesmo grupo de referência e a única diferença entre elas deve ser a presença ou ausência do programa e todas as consequências que essa diferença deve trazer. A diferença entre as duas situações mede o impacto do programa.

A partir dessa definição, a dificuldade fundamental para se medir a magnitude do impacto de um programa social pode ser facilmente entendida. Dado determinado grupo de referência e determinado programa, nunca é possível observar a situação do grupo na presença e na ausência do programa. Na prática, apenas uma das duas situações pode ser observada; a outra será sempre hipotética. Quando o grupo de referência é formado por beneficiários, apenas a situação na presença do programa pode ser diretamente observada. Quando o grupo de referência é formado por futuros beneficiários, apenas a situação na ausência do programa pode ser diretamente observada. Portanto, o impacto de um programa social é sempre medido a partir da comparação de duas situações, onde uma é sempre não observável.

Esta é a grande dificuldade em realizar qualquer avaliação de impacto, algumas vezes precipitadamente percebida como indicação de impossibilidade. Em sua *Insustentável Leveza do Ser*, Milan Kundera afirma que “*There is no means of testing which decision is better, because there is no basis for comparison*”. Frente à angustia de não saber a resposta, Kundera, ao longo de seu livro, encontra uma base para comparação para possibilitar a tomada de decisão. Ao longo desse livro,

será discutido como é possível realizar estatisticamente comparações envolvendo ao menos uma situação hipotética (não factual e por isso, denominada contra-factual) e obter estimativas da magnitude do impacto de programas sociais.

Caso o resultado de interesse não fosse também influenciado por uma variedade de outros fatores que variam ao longo do período da observação, seria simples avaliar o impacto de um programa social: bastaria contrastar a situação do grupo de interesse algum tempo após a implementação do programa com a situação, nesse mesmo grupo, antes da implantação do programa. Nesse caso, a situação antes da implantação do programa seria idêntica à que prevaleceria após o momento em que o programa foi implantado, caso ele não tivesse sido implantado.

Assim, pode-se reverter a argumentação e afirmar que a grande dificuldade da avaliação de impacto não é observar a situação do grupo de referência na ausência do programa, já que essa observação pode ser feita antes da implantação do programa. A dificuldade, nessa ótica, resulta da invariável evolução da situação do grupo de referência, determinada por uma série de fatores para além da influência do próprio programa, cujo impacto se deseja isolar. Por esse ângulo, o grande desafio de toda avaliação de impacto é isolar, da modificação da situação do grupo de interesse, a parcela que se deve exclusivamente à influência do programa.

Como já explicitado anteriormente, o ideal em uma avaliação de impacto é que a única diferença entre as duas situações analisadas seja a participação ou não do grupo de interesse em um dado programa. Apenas nesse caso, seria possível referir à avaliação como sendo a avaliação da magnitude do impacto do programa em questão. Via de regra, no entanto, a ausência de um programa não corresponde a um “vazio”: existem alternativas ao programa e que, na sua ausência, irão “ocupar seu espaço”. Afinal, em princípio, o objetivo de um programa é apenas garantir aos beneficiários a melhor opção e não ser a única opção. Com isso, o impacto é tipicamente apenas o contraste entre uma situação em que o grupo de referência foi beneficiado pelo programa cujo impacto se deseja avaliar e outra em que o grupo de referência foi beneficiado por algum outro programa ou influenciado por alguma outra condição. Nesse caso, não se está propriamente avaliando o impacto de um programa: o que está sendo avaliado é a diferença de impacto entre dois programas, ou seja, qual seria o valor adicionado do programa. Dessa forma, para que o impacto possa ser corretamente interpretado, é fundamental contar com uma descrição precisa não só do programa em análise, mas também da alternativa que estará disponível na sua ausência.

Outro ponto importante refere-se à definição do grupo de referência, cuja situação o programa pretende modificar. Na prática, a escolha desse grupo é, muitas vezes, feita com base em pura conveniência metodológica; escolhe-se o grupo de referência sobre o qual é mais fácil estimar a magnitude do impacto de um determinado programa. Mas, em princípio, a escolha do grupo de referência deve decorrer dos objetivos da avaliação e não de conveniências metodológicas. Assim, caso o objetivo seja prestar contas, o grupo de referência deve ser o conjunto dos beneficiários. Por outro lado, caso o interesse seja avaliar a expansão do programa, procura-se estimar o impacto do programa sobre os futuros beneficiários e nesse caso, o grupo de referência mais adequado é aquele formado pelos que deverão ser imediatamente beneficiados em caso de ampliação da escala da intervenção. Já quando o objetivo for avaliar o desempenho de longo prazo do programa que cobrir toda a sua população-alvo, o grupo de referência mais adequado será aquele formado por todos aqueles para os quais o programa foi desenhado para atender (atuais e potenciais beneficiários).

Embora o objetivo de uma avaliação de impacto seja obter o impacto de uma intervenção sobre a população de referência, via de regra, essa intervenção também tem impacto sobre populações que não foram, e muitas vezes que nunca serão, beneficiadas pelo programa. Existe, portanto, uma diferença entre o impacto de um programa sobre a população que se pretendia beneficiar, usualmente denominado impacto privado do programa, e o seu impacto sobre todo o conjunto da população que é direta ou indiretamente afetado pela intervenção, usualmente denominado de impacto social (no sentido de abranger o impacto sobre toda a sociedade). Assim, pode-se ter avaliações de impacto que sejam dirigidas a isolar e a mensurar impactos privados ou impactos sociais de um dado programa.

Na busca por um desenho progressivamente mais eficaz e adequado às necessidades dos beneficiários, é essencial que aqueles responsáveis pelo desenho de políticas públicas conheçam que mecanismos levam o programa alcançar um certo impacto. Em seu sentido mais estrito, o objetivo de toda avaliação de impacto se limita a isolar e mensurar a magnitude do impacto do programa. Em princípio, a determinação dos mecanismos que levam um programa a alcançar determinado impacto encontra-se fora do escopo da avaliação de impacto. No entanto, como será mencionado na sequência, o conhecimento da sensibilidade da magnitude do impacto a uma série de fatores pode ser de grande valia na busca por desenhos mais eficazes e adequados às necessidades dos beneficiários.

1.3. Heterogeneidade do impacto

Nesta seção, será discutida a heterogeneidade intrínseca do impacto de um programa social. A magnitude do impacto de um programa, por ser a diferença entre o que aconteceria na presença do programa e o que teria acontecido na sua ausência, varia naturalmente entre as unidades de análise. Nada na definição de impacto requer que a sua magnitude seja uniforme ou invariante. Assim, o impacto não tem uma magnitude fixa ou dada que varia entre as unidades beneficiadas. Ao final, existe uma distribuição de magnitudes entre todos os possíveis beneficiários do programa.

Em geral, avaliações de impacto preocupam-se em estimar a média dessa distribuição, denominada impacto médio. Entretanto, o conhecimento de toda a distribuição é certamente ainda mais indispensável. Conhecer certos aspectos da distribuição pode ser bem mais importante que conhecer o impacto médio. Por exemplo, conhecer a porcentagem dos beneficiados para os quais a magnitude do impacto é superior a um dado nível mínimo pode ser mais importante que conhecer a magnitude do impacto médio de um programa.

O conhecimento da distribuição dos impactos é de relevância prática maior quando é possível conhecer simultaneamente as características e a magnitude do impacto sobre os beneficiários do programa. Nesse caso, é, por um lado, possível identificar aqueles sobre os quais o programa tem maior impacto, para então focalizar a atenção nesse grupo. Por outro lado, também pode-se identificar os agrupamentos em que o impacto é limitado. Esse conhecimento pode, então, sugerir alterações no desenho do programa, de forma a promover maior adequação às necessidades desses beneficiários e, com isso, aumentar seu impacto médio.

Tipicamente, a magnitude do impacto de um programa também depende de sua adequação ao ambiente em que ele se insere. Assim, para promover programas que tenham maior impacto, é importante que as avaliações investiguem a sensibilidade do impacto ao ambiente socioeconômico. Esse conhecimento pode melhorar o foco dos programas, além de tornar possível o redesenho desses programas, de modo a torná-los mais adequados às condições socioeconômicas vigentes.

O impacto de um programa tampouco é constante com o tempo de exposição dele. Tipicamente, a magnitude do impacto deve crescer logo após a exposição ao programa, na medida em que os efeitos do programa madurem, seguindo-se de um declínio da magnitude do impacto, devido a

um natural processo de depreciação. O benefício de um programa, evidentemente, não é o vigente em um momento qualquer, mas o valor presente dos impactos em todos os momentos. Assim, para calcular o valor total dos benefícios de um programa, é necessário avaliar a magnitude do seu impacto em cada momento no tempo.

Até esse momento, essa seção discutiu a heterogeneidade intrínseca do impacto de um programa. Foi detalhado como a magnitude do impacto de um programa muito bem definido é diferente para diferentes beneficiários, variável ao longo do tempo e de acordo com o ambiente em que se insere, como também dependente de que outros programas estejam disponíveis. Mas os programas não tendem a ser intrinsecamente homogêneos. Assim, tipicamente o impacto varia entre beneficiários não apenas porque são influenciados de forma diferenciada pelo mesmo programa, mas também porque tendem a ser beneficiados por variações dentro do mesmo programa. Essas variações podem diferir do seu desenho, dos componentes que contemplam, da dosagem e do tempo de exposição. Mas podem também diferir com relação a escala da operação e da forma de implementação e operação. Dado que muito dessas variáveis estão sob controle dos gestores do programa, a utilidade de uma avaliação de impacto é muito ampliada quando a dependência da magnitude do impacto a essas variáveis é mapeada.

1.4. Quando avaliar o impacto de um programa social? Avaliações ex ante e ex post

Existem essencialmente três momentos na execução de um programa em que avaliações de impacto podem ocorrer: (i) antes do início do programa, denominadas avaliações *ex ante*; (ii) durante o período de execução do programa, denominadas avaliações *ex post* de percurso; e (iii) após a conclusão do programa, denominadas avaliações *ex post* de encerramento.

Em cada caso, a avaliação é caracterizada por metodologias e objetivos distintos, condicionados à disponibilidade de informações, relativas ao programa nos diferentes momentos do tempo.

A decisão pela implementação de um dado programa é geralmente precedida pela apresentação de argumentos que justifiquem a racionalidade da opção por aquele determinado desenho. Justificativas desse tipo baseiam-se, ou deveriam se basear, em estimativas de custos e das

magnitudes dos impactos esperados. Prever a magnitude dos impactos antes que o programa seja implantado é o objetivo das avaliações *ex ante* de impacto. Embora avaliações *ex ante* sejam absolutamente indispensáveis para justificar a adoção de um dado programa, ainda mais importante é utilizá-las para analisar e decidir sobre o melhor desenho de um novo programa.

Mesmo após terem sido elaborados e implementados, programas sociais muitas vezes permitem ajustes ao seu desenho durante a fase de operação. Avaliações de impacto realizadas durante o período de operação do programa, denominadas avaliações *ex post* de percurso, devem ser utilizadas para verificar a validade das previsões, realizadas pelas avaliações *ex ante*. Dessa forma, esse tipo de avaliação é fundamental para traçar recomendações sobre o futuro do programa, seja no sentido de interrompê-lo, mantê-lo ou aperfeiçoar seu desenho. Esse tipo de avaliação é muito comum durante a fase-piloto da implementação de um programa – mas não deve ser limitado a essa fase.

O grande desafio para a realização de uma boa avaliação *ex post* de percurso é o tempo. Por um lado, é importante que a avaliação seja realizada o quanto antes para que possa influenciar o desenho, a operação, a continuidade e a expansão do programa maximizada. Por outro lado, dado que alguns impactos podem demorar a se manifestar, quanto mais cedo realizada a avaliação de impacto, maior a probabilidade de se subestimar ou até ignorar impactos importantes, que não se manifestam em curto prazo. Esse problema é particularmente grave, pois frequentemente, os impactos de maior importância são os de longo prazo. Assim, toda avaliação *ex post* de percurso precisa analisar cuidadosamente o compromisso entre fornecer respostas rápidas que contribuem com o aperfeiçoamento e a expansão do programa e aguardar o tempo necessário para que parte substancial dos impactos do programa se manifeste. Por conta desses desafios, avaliações *ex post* de percurso são mais comuns em programas que permanecem em operação por um longo tempo e dificilmente são realizadas em projetos pontuais.

Por fim, também é justificável a realização de uma avaliação de impacto após o encerramento de um programa social. Neste caso, dois objetivos merecem atenção. Por um lado, a avaliação *ex post* de encerramento é útil para determinar quão adequada foi a decisão de alocação de recursos àquele programa; trata-se, assim, de uma avaliação *ex post* das decisões realizadas no passado e da qualidade da gestão do programa. Uma avaliação desse tipo não permite rever ou aprimorar ações, mas apenas verificar se os agentes envolvidos no processo decisório tomaram ou não decisões

adequadas, além de analisar se foi feita uma boa gestão do programa. Por outro lado, avaliações *ex post* tem uma importância que vai além dos benefícios diretos para o programa avaliado: a análise de avaliações *ex post* para um amplo leque de programas que buscam modificar o mesmo resultado permite identificar as melhores práticas e programas com a melhor relação custo-efetividade. Tais informações auxiliam muito gestores públicos empenhados em desenhar novos programas, ou aperfeiçoar programas já existentes que atuam sobre os mesmos resultados. Vale também ressaltar que as avaliações *ex post* de encerramento têm a grande vantagem de superar o conflito entre o momento da avaliação e o prazo de maturação do impacto do programa: não há restrição de tempo para a sua realização, o que torna este tipo de avaliação ideal para garantir a captação dos impactos verdadeiramente de longo prazo de um programa social.

Todos os três tipos de avaliação combinam teoria com evidência empírica e não há uma recomendação pré-estabelecida sobre os pesos que devem ser atribuídos à teoria ou à evidência empírica em avaliações de um modo geral. Nas avaliações *ex ante*, a análise precisa ser conduzida sem que ninguém jamais tenha sido beneficiado pelo programa e portanto, é natural que o papel da teoria passe a ser primordial. Nesse caso, não existem muitas alternativas e as estimativas de impacto são obtidas a partir de simulações sobre o comportamento hipotético dos beneficiários do programa com base em modelos teóricos. Os parâmetros desses modelos precisam ser estimados ou calibrados, em ambos os casos com base em hipóteses difíceis ou impossíveis de serem testadas.

Já uma avaliação *ex post* pode ser inteiramente planejada e contar com informações previamente desenhadas sobre a situação e desempenho de beneficiários e não beneficiários, antes e em vários momentos após sua exposição ao programa. Conforme informações naturalmente acumulam-se ao longo do tempo, avaliações *ex post* têm grande vantagem informacional sobre suas contrapartidas *ex ante*. Não é à toa, portanto, que as avaliações *ex post* tendem a ser muito mais empiricamente orientadas e muito menos baseadas em teoria que avaliações *ex ante*. Seria um erro, no entanto, ignorar o papel da teoria mesmo nas mais bem planejadas avaliações *ex post*. Embora, nessas avaliações, medidas da magnitude do impacto possam ser obtidas sem a necessidade do auxílio da teoria, o mesmo não se pode dizer da interpretação do impacto, da identificação dos mecanismos que o determinam e da sua generalização a outros contextos. Todos esses três importantes componentes da avaliação de impacto *ex post* só podem ser feitos após uma intensa utilização de teoria.

1.5. Por que avaliar o impacto de um programa social?

A decisão dos formuladores de políticas públicas de investir em um programa social é sempre baseada na compreensão de que o programa terá impacto sobre um conjunto de resultados de interesse coletivo. Mas será que, na prática, as políticas públicas e os projetos sociais de instituições de terceiro setor atingem os objetivos para os quais foram desenhados? Conforme já ressaltado, o principal propósito de uma avaliação de impacto é verificar se um determinado programa alcançou, ou está alcançando, os impactos esperados.

Entretanto, avaliações de impacto certamente consomem recursos que, alternativamente, poderiam ser investidos diretamente no próprio programa. A crescente expansão de estudos dessa natureza indica que a importância de dedicar esforços e recursos para este fim é cada vez mais reconhecida. Mas, por que realizar avaliações de impacto representa um uso adequado dos recursos disponíveis? Por que não seria melhor revertê-los em maior investimento no próprio programa? Afinal, os benefícios de uma avaliação de impacto superam seus custos?

A resposta a essas indagações evidentemente depende do custo, da qualidade e principalmente da utilização que será dada a avaliação, afinal nenhum estudo desse gênero tem valor intrínseco. Ao contrário, seu valor reside na sua utilização para a melhoria do desenho do programa que foi avaliado, para o desenho de outros programas subsequentes na mesma comunidade, ou para informar formuladores de políticas que pretendam desenhar programas com objetivos similares em outras comunidades. A seguir, a utilidade da avaliação de impacto em diferentes cenários será discutida.

O primeiro questionamento refere-se à utilidade de uma avaliação de impacto de um programa amplamente utilizado. Não é óbvia, nesse caso, a justificativa para a aplicação de recursos para a avaliação de impacto. Afinal, pode-se argumentar ser pouco provável que um programa com acesso voluntário não tenha impacto sobre aqueles que o procuram e dele se utilizam. Por que alguém dedicaria tempo e esforço a um programa que não lhe traz benefícios e, portanto, não tem qualquer impacto? A ampla utilização de um programa por ao menos um segmento da sociedade não seria, por si só, indício de existência de impacto? De fato, é pouco provável que a maioria dos programas existentes não traga benefício algum a seus beneficiários. Afinal, em sua maioria, os programas sociais são voluntariamente utilizados pelos seus beneficiários que, portanto, devem estar convencidos da sua utilidade.

Além disso, em sua maioria, os programas sociais não são desenhados sem uma base teórica e empírica. Em geral, o desenho de um programa social é baseado em teorias que preveem um elo entre o acesso ao programa e impactos positivos sobre um conjunto predeterminado de resultados. Portanto, os formuladores de políticas públicas, geralmente, contam com argumentos teóricos e, frequentemente, com evidência empírica, que sinalizam para a eficácia do programa em questão. Assim, investigar a existência de impacto de um programa cuidadosamente desenhado e voluntariamente utilizado pela sociedade não seria questionar a racionalidade dos usuários e a validade dos argumentos teóricos e empíricos adotados no desenho do programa? Existem duas possíveis respostas a essa indagação.

Por um lado, uma das motivações de uma avaliação de impacto é, precisamente, a validação das teorias que serviram de base ao desenho do programa. Por esse motivo, é tão comum e tão importante que sejam feitas avaliações de impacto de programas em escala-piloto. A avaliação de um piloto busca verificar se a ligação entre o acesso ao programa e os resultados esperados é válida, para que, então, o programa possa ser expandido com menos incerteza sobre o seu potencial de alcançar sucesso. Similarmente, uma avaliação de impacto também é útil para confirmar se as percepções dos usuários sobre a eficácia do programa coincidem com a realidade.

Por outro lado, pode-se argumentar que o objetivo de uma avaliação de impacto vai muito além de simplesmente determinar a existência ou não de impacto. Mesmo que haja convicção da existência de impacto de um programa, ainda assim será necessário conduzir uma avaliação para que este impacto possa ser melhor compreendido. Tanto para o gestor do programa, quanto para seus usuários, tão ou mais importante do que saber se há algum impacto, é conhecer a magnitude do impacto, assim como saber sobre quais dimensões o programa tem impacto e quais canais possibilitam a manifestação desse impacto. Esse conhecimento é indispensável para o aperfeiçoamento do desenho do programa e para a melhoria na sua adequação às necessidades de seus usuários.

Outro questionamento refere-se a programas sociais ofertados gratuitamente para a parcela mais pobre da sociedade: nesse caso, a adesão voluntária ao programa deve ser necessariamente interpretada como indício do impacto do programa? Pode-se argumentar que isso não é verdade, afinal é possível que os pobres tenham conhecimento limitado sobre a eficácia daquele programa e, como a oferta é gratuita, aderem a ele simplesmente por não perderem nada participando.

E o que se pode dizer do impacto de um programa, oferecido gratuitamente aos mais pobres, mas para o qual existe provisão privada para suprir a demanda dos mais ricos, que, por sua vez, pagam preços significativos pelo acesso a esse programa? Assumindo que a qualidade dos serviços oferecidos gratuitamente aos pobres seja igual à qualidade da oferta privada aos ricos, não seria a existência de demanda pelos mais ricos evidência suficiente de que o serviço tem impacto? Com base na teoria econômica, podemos argumentar que a existência de demanda por um serviço com preço significativo é sim um indício de impacto. Por que então realizar uma avaliação de impacto nesta situação?

Observadas as diferenças entre os pobres e os ricos, pode-se argumentar que o fato de um programa ter impacto sobre um grupo não implica necessariamente que existirá o mesmo impacto sobre o outro grupo. Dessa forma, havendo demanda dos mais ricos por um programa, ainda assim será necessário avaliar a magnitude do impacto sobre os mais pobres.

Além disso, mesmo que não haja dúvidas sobre a existência do impacto de um programa sobre os mais pobres, pode ser importante avaliar a sua magnitude. Sabe-se que, em geral, as prioridades dos indivíduos (e, portanto, o valor dos benefícios) variam com o nível de renda: benefícios que têm elevado valor para os mais ricos podem não ser classificados como prioritários para os mais pobres. Logo, nem mesmo evidência de que a eficácia do programa seja a mesma entre pobres e ricos é suficiente para justificar a implantação de um programa para os mais pobres. É também preciso avaliar quão elevado é o valor dos benefícios do programa para o grupo mais pobre.

Outra indagação refere-se à avaliação de programas que já tenham sido avaliados em outros contextos: qual seria a justificativa para realizar novas avaliações? É necessário verificar a validade externa das avaliações anteriores. Via de regra, programas sociais são desenhados visando a uma população-alvo específica em um ambiente preestabelecido. Assim, é natural que existam dúvidas sobre o impacto que o programa terá quando é implementado em outros locais, beneficiando populações para as quais não foi originalmente desenhado. Conforme já ressaltado anteriormente, uma das utilidades da avaliação de impacto é, precisamente, determinar em que medida a eficácia de um programa depende das características da população beneficiada e da natureza do ambiente em que é implementado.

Importante também lembrar que impacto de um programa não depende apenas do seu desenho, de sua adequação ao perfil dos beneficiários e ao ambiente socioeconômico em que se insere, mas

também de sua implementação. Mesmo existindo poucas dúvidas a respeito do impacto potencial de um programa bem desenhado e adequadamente focalizado, podem persistir grandes incertezas relativas ao efetivo impacto desse programa quando implementado de maneira deficiente. Nesse caso, o objetivo da avaliação de impacto não é investigar o impacto teórico ou potencial do programa, mas sim sobre o seu impacto efetivo, dada a implementação imperfeita. Assim, o impacto de um programa depende não apenas do seu impacto potencial, mas está também intrinsecamente ligado às condições de sua implementação. Ainda que um programa tenha comprovadamente impacto potencial, é fundamental avaliar o seu impacto efetivo e a avaliação de impacto pode ser útil para verificar qual parcela do impacto potencial foi desperdiçada devido a falhas no processo de implementação.

Aliás, esse dilema entre impacto efetivo e potencial é inerente à avaliação de impacto de políticas públicas. Em várias áreas de conhecimento, existem estudos teóricos que, consensualmente, apontam para a importância de determinadas ações; entretanto, muitas avaliações de impacto encontram impactos irrisórios dessas mesmas ações após sua implementação. Assim, o frequente descompasso entre os resultados teóricos e empíricos reforça ainda mais a necessidade de uma avaliação de impacto: esta é essencial para discernir entre o impacto do programa como originalmente desenhado (o impacto potencial) e o impacto do programa como de fato implementado (o impacto efetivo).

Pode-se, também, indagar por que realizar várias avaliações de um mesmo programa social, em diferentes pontos no tempo. Para justificar a importância do uso da avaliação nesse contexto, é importante indicar que o impacto de um programa tem caráter mutável ao longo do tempo, e por isso, a avaliação de impacto deve ser considerada uma ferramenta de utilização contínua, já que as conclusões auferidas a partir de sua utilização pontual não são necessariamente válidas ao longo do tempo.

Existem, pelo menos, quatro razões para justificar a potencial inconstância temporal do impacto de um programa social. Em primeiro lugar, o impacto de um programa está normalmente relacionado ao tempo de exposição do beneficiário, podendo tanto declinar rapidamente logo após o término do programa, como persistir por um longo período de tempo. Em segundo lugar, ao longo do tempo, podem ocorrer flutuações na qualidade da gestão do programa e, consequentemente, na eficácia da implementação, gerando oscilações na magnitude do impacto. Em terceiro

lugar, o impacto pode variar ao longo do tempo, caso seja sensível a mudanças no ambiente socioeconômico. Por fim, caso o perfil da população beneficiada pelo programa mude ao longo do tempo, é esperado que a magnitude do impacto também se altere.

Assim, até mesmo programas cujo impacto inicial é de magnitude conhecida devem ser continuamente avaliados. Por um lado, uma avaliação contínua é a única maneira de identificar os impactos de médio e longo prazos do programa e, portanto, a única forma de avaliar seu grau de maturação e persistência. Por outro lado, repetidas avaliações do mesmo programa podem revelar como o impacto do programa responde a mudanças no ambiente socioeconômico e no perfil dos beneficiários. Por fim, mesmo caso o perfil da população beneficiada e o ambiente socioeconômico permaneçam estáveis, avaliações de impacto periódicas podem ser de extrema utilidade como instrumento de gestão. Nesse caso, flutuações na magnitude do impacto podem indicar melhoria ou deterioração na forma de implementação e, portanto, na gestão do programa.

Avaliações de impacto também são importantes, pois possibilitam avaliar a relação custo-efetividade de um programa. Afinal, em um cenário onde os recursos são escassos e diferentes programas competem pelos mesmos recursos, a existência de impacto não é suficiente para justificar a alocação de recursos adicionais para um dado programa: é preciso mensurar sua magnitude e calcular o impacto por unidade de custo (seu custo-efetividade). Mesmo que o impacto de um dado programa seja inquestionável, este programa pode não ser aquele com a melhor relação custo-efetividade e, portanto, pode não ser o melhor candidato a receber os recursos disponíveis. Assim, para se avaliar a relação custo-efetividade de um programa, é necessário não apenas reconhecer a existência do impacto, mas também mensurar a sua magnitude, e por isso, a avaliação de impacto é tão importante.

Entretanto, para que recursos sejam alocados a um programa, não basta que o programa tenha impacto ou mesmo que seja aquele com a melhor relação custo-efetividade. É necessário também que o valor dos benefícios do programa supere seus custos, isto é, que o programa tenha uma relação custo-benefício favorável. Para tanto, realiza-se uma comparação entre os custos e o valor dos benefícios do programa, sendo o último obtido a partir do produto entre a magnitude do impacto e o valor atribuído a esse valor pelo conjunto de beneficiários. No cálculo dessa relação, deve-se considerar todos os custos e benefícios sociais, tanto aqueles que atingem os agentes diretamente envolvidos, quanto as externalidades sobre o restante da sociedade.

Assim, mesmo diante da evidência de impacto de um dado programa, é essencial estimar a magnitude desse impacto para que se possa verificar se o valor dos benefícios produzidos supera os custos demandados. Muitas vezes, diferentes ações com objetivos análogos não diferem substancialmente com relação ao seu custo. Nesses casos, são as diferenças na magnitude do impacto que irão determinar qual o programa que tem a melhor relação custo-benefício – e avaliações de impacto são centrais nessa análise.

Em princípio, é possível argumentar que todos os programas com relação custo-benefício favorável deveriam ser implementados. Essa certamente é a regra quando não existe interdependência entre os programas. De fato, se os benefícios e os custos de um programa não dependem da existência ou operação dos demais, não existe razão para que todos os programas com relação custo-benefício favorável não sejam simultaneamente implementados.

No entanto, em geral, os benefícios e custos de programas alternativos dependem da existência ou operação de seus concorrentes, e as relações de custo-benefício são estimadas supondo que não foram nem serão implementadas. Nesse caso, o procedimento decisório deve ser necessariamente sequencial. Primeiro, decide-se sobre a implementação do programa com a melhor relação custo-benefício. Em seguida, novas relações de custo-benefício são recalculadas para os programas não implantados, levando em consideração que aquele de melhor relação custo-benefício foi efetivamente implantado. Após essa reavaliação, caso ainda exista algum programa com relação custo-benefício favorável, o melhor deles deve ser implantado e o procedimento novamente conduzido. Note que essa análise sequencial requer a estimação do impacto de cada programa tanto na ausência quanto na presença dos demais. Novamente, pode-se perceber a importância de avaliações de impacto.

Entretanto, há limitações na aplicação da avaliação de impacto à investigação da relação custo-benefício de um programa. Primeiramente, um programa tem geralmente uma variedade de impactos – alguns intencionais e muitos outros colaterais. Nesse caso, para obter a relação custo-benefício, seria necessário estimar a magnitude de todos esses impactos, o que já seria um desafio, dado que muitos desses impactos são de difícil mensuração. Em segundo lugar, mesmo que fosse possível estimar a magnitude de todos os impactos, ainda seria preciso estimar o valor que cada beneficiário atribui a cada um deles. Ou seja, a avaliação de impacto é apenas um primeiro passo para se estimar a relação custo-benefício.

Caso a mensuração da magnitude de cada impacto, seguida de sua valoração, fosse a única alternativa para se obter o valor dos benefícios, sem dúvida, a avaliação de impacto seria essencial para a estimação da relação custo-benefício de um programa. No entanto, existe uma alternativa para estimar o valor dos benefícios: a “propensão a pagar” do beneficiário pelo serviço.

O valor total de um programa, benefício ou serviço para um beneficiário pode ser avaliado a partir da investigação do valor pelo qual ele estaria disposto a trocar o acesso ao programa, medido monetariamente ou conforme outros bens e serviços. As técnicas utilizadas para a investigação da “propensão a pagar” podem ser classificadas em dois grandes grupos: (i) comportamental e (ii) não comportamental.

Nos métodos não comportamentais, pergunta-se diretamente ao beneficiário quanto ele estaria disposto a pagar pelo acesso a determinado programa, benefício ou serviço. O desafio nesse caso é que, frequentemente, a resposta pode não ser fidedigna. Como a pergunta se refere a uma situação hipotética, tende-se a superestimar o valor que elas efetivamente estariam dispostas a pagar. Existem, entretanto, protocolos preestabelecidos que definem como investigações dessa natureza podem ser conduzidas visando a mitigar a possibilidade de resultados pouco confiáveis.

Por outro lado, nos métodos comportamentais, o valor do benefício é inferido a partir da observação direta do comportamento efetivo do beneficiário. No caso dos serviços para os quais existe um mercado, verificamos que o preço de mercado é o valor (ou a propensão a pagar) que o beneficiário marginal atribui ao serviço em questão. Quando não existe mercado, ainda assim o valor associado ao serviço pode ser inferido com base na observação de outros comportamentos. No caso de um posto de saúde, por exemplo, o valor pode ser avaliado a partir do raio de cobertura do serviço. Quanto mais valorizado o serviço, maior a distância que os beneficiários estarão dispostos a percorrer para acessá-lo.

Em resumo, a vantagem metodológica da utilização da propensão a pagar é que esse método estima diretamente o valor do benefício, evitando problemas potenciais causados pela necessidade de estimar múltiplos impactos e o valor atribuído a cada um deles. Ao contrário da avaliação de impacto, não é necessário primeiro decompor o impacto nos seus diversos componentes, para posteriormente valorar cada um deles individualmente. Sendo assim, dado que a avaliação de impacto se apresenta como um instrumento mais complexo e que via de regra exige maior esforço, por que utilizá-la?

Existem essencialmente três justificativas que sustentam o uso da avaliação de impacto, mesmo diante da possibilidade de estimar a propensão a pagar. Em primeiro lugar, quando o impacto de um programa ocorre via poucas dimensões de fácil mensuração, a magnitude do impacto pode ser avaliada de forma relativamente incontestável com base em métodos experimentais. Além disso, na ausência de mercados para o serviço ou benefício em questão, a utilização da propensão a pagar poderia levar a estimativas ainda menos confiáveis. No entanto, é importante ressaltar que um aumento do número de dimensões do impacto e o surgimento de mercado para o serviço ou benefício avaliado, tornaria o uso da propensão a pagar mais fundamentada *vis-à-vis* à utilização da avaliação de impacto.

Em segundo lugar, pode-se argumentar que a utilização da propensão a pagar tem sua confiabilidade reduzida quando existem externalidades e o programa avaliado impacta não beneficiários. Nesse caso, se o número de dimensões do impacto é limitado e o grupo de não beneficiários impactados pelo programa está bem definido, uma avaliação de impacto experimental é provavelmente capaz de produzir estimativas mais confiáveis.

Em terceiro lugar, o uso da propensão a pagar não permite a identificação dos motivos pelos quais os beneficiários valorizam aquele serviço. Assim, sabe-se quanto o programa é valorizado, mas nada se sabe sobre os mecanismos pelos quais o programa influencia o bem-estar dos beneficiários. Essa limitação é especialmente problemática no caso de programas multifacetados, já que se torna impossível distinguir qual componente do programa está contribuindo mais ou menos para a satisfação do usuário.

Por fim, mesmo que do ponto de vista dos beneficiários, seja suficiente saber o grau de satisfação dos beneficiários com o programa de um modo geral, para os gestores e aqueles que financiam o programa, essa informação pode ser insuficiente. É preciso, então, identificar os impactos específicos de um programa. Por um lado, é possível que os responsáveis pelo financiamento valorizem os impactos específicos de maneira distinta dos beneficiários. Em outras palavras, aqueles que financiam o programa podem querer se assegurar de que o programa está causando satisfação pela via pretendida e não por algum outro motivo qualquer. Por outro lado, identificar a magnitude dos impactos específicos é importante para avaliar a adequação do programa a seus objetivos específicos. Essa informação pode ser útil tanto para o redesenho do próprio programa quanto para o desenho de outros programas em contextos nos quais alguns impactos específicos são particularmente valorizados.

1.6. Como utilizar a avaliação de impacto de um programa social?

Avaliações de impacto têm, em geral, uma multiplicidade de usos que podem ser categorizados em dois grandes grupos: (i) uso interno e (ii) uso externo. O uso interno de uma avaliação diz respeito à sua utilidade para a tomada de decisões sobre o próprio programa avaliado. Já o uso externo de uma avaliação é relativo à utilização dos resultados da avaliação como instrumento para o desenho ou a melhoria do desenho de outros programas similares.

Em seu uso interno, a avaliação de impacto serve como insumo para a tomada de decisão sobre o próprio programa, podendo ser utilizada com vários objetivos, a depender do usuário a que se destina. Quando os usuários dos resultados são aqueles responsáveis pela continuidade ou descontinuidade do programa, a avaliação de impacto é tipicamente utilizada como veredito: resultados positivos justificam a permanência ou ampliação de um programa, já a ausência deles dá suporte à desativação progressiva ou imediata do programa. Assim, em primeiro lugar, o importante uso interno da avaliação de impacto é: (i) justificar a existência de um programa e (ii) decidir sobre sua desativação, continuidade ou expansão.

Outra utilidade interna potencial de uma avaliação de impacto é fornecer insumos para a promoção de melhorias no desenho do programa. Para tanto, é fundamental que a avaliação não apenas obtenha estimativas da magnitude do impacto, mas também identifique os fatores determinantes do impacto. Saber como a magnitude do impacto varia de acordo com os parâmetros que definem a intervenção é a única maneira pela qual uma avaliação pode contribuir para a reformulação dos princípios que embasam o programa, possibilitando, assim, o aperfeiçoamento do seu desenho e a adequação de seu marco lógico.

O impacto de um programa, entretanto, não é determinado apenas por seu desenho. A natureza dos beneficiários, além da forma e do grau com que eles utilizam as ações do programa, também influenciam a magnitude do impacto. Assim, conforme a avaliação revela como a magnitude do impacto do programa varia com a forma e grau de utilização dos serviços, e também com o tipo de beneficiário, ela se torna um instrumento de extrema utilidade ao gestor responsável pela operação do programa. Afinal, a partir dos resultados de uma avaliação de impacto, o gestor torna-se capaz

de realizar ajustes finos no programa, podendo, potencialmente, readequar a natureza dos serviços e bens oferecidos às necessidades dos beneficiários e, com isso, promover maior efetividade do programa em questão.

Por fim, é importante ressaltar que o impacto do programa também depende da forma como a gestão e operação do programa é conduzida. Com isso, uma avaliação de impacto que conteemplace informações sobre a sensibilidade da magnitude do impacto às variáveis relacionadas à gestão e operação do programa fornece informações valiosas para o aperfeiçoamento ou reformulação do sistema de gestão e operação.

Mas uma avaliação de impacto também tem importantes usos externos ao programa, relacionados com a sua utilidade em auxiliar no desenho, na implementação ou na gestão de outros programas similares em contextos distintos. Nesse sentido, avaliações de impacto funcionam como bens públicos, na medida em que beneficiam uma população muito maior que aquela que originalmente participou do programa e financiou a avaliação.

Dentre a grande variedade de usos externos de uma avaliação de impacto, talvez o de maior importância seja a identificação de melhores práticas. Gestores em toda parte estão continuamente em busca dos melhores programas, ações e práticas que permitam alcançar seus objetivos da forma mais eficaz possível. Em todo momento, dada a tecnologia social existente, há um conjunto de práticas com as melhores relações custo-efetividade. A identificação dessas intervenções requer avaliações de impacto de uma ampla variedade de programas com objetivos similares. Assim, é indiscutível que a identificação das melhores práticas e, portanto, a realização de avaliações de impacto é fundamental para o aprimoramento da eficácia dos programas sociais em vigor.

Além disso, a avaliação dos impactos de um programa, quando associada à avaliação de seu custo, permite comparar o programa avaliado com as alternativas disponíveis, referentes às suas relações de custo-efetividade. A sua posição relativa àqueles programas que representam as melhores práticas define o seu potencial de expansão e difusão. Programas com as relações custo-efetividade mais favoráveis estão em evidência como candidatos tanto para a expansão nos locais onde atuam quanto para a adoção em outras áreas, desde que seu impacto seja robusto a mudanças nos contextos socioeconômicos e culturais.

Assim, para que a utilização externa das avaliações de impacto seja a mais ampla possível, é essencial que elas investiguem não apenas o tamanho do impacto, mas também a interação dele com características do ambiente socioeconômico e cultural em que as avaliações se inserem. Esse é o conceito de validade externa que, em outras palavras, determina em que medida o impacto estimado de um programa pode ser extrapolado para diferentes contextos. Trata-se, fundamentalmente, de uma avaliação da robustez ou da sensibilidade do programa ao contexto.

Frequentemente, os programas avaliados se mostram sensíveis ao contexto em que se inserem. Essa sensibilidade, entretanto, não é de forma alguma uma indicação de que o programa não possa ser ajustado para outro contexto. A constatação da sensibilidade do impacto do programa ao seu ambiente deve ser utilizada apenas como alerta para que a expansão do programa seja feita de forma cuidadosa, adequando-o às características específicas do novo contexto. Nesses casos, é importante que as avaliações de impacto estabeleçam não apenas o grau de sensibilidade do impacto do programa como um todo ao contexto, mas fundamentalmente estabeleçam quais aspectos do programa determinam essa sensibilidade (ou, em outras palavras, como a magnitude do impacto depende do contexto).

Esse tipo de informação é fundamental para sinalizar quais cuidados devem ser adotados na difusão do programa, mas nada dizem sobre quais seriam os ajustes necessários no seu desenho para que o seu impacto seja invariante ao contexto. Para que esses ajustes possam ser identificados, seria necessário que as avaliações também indicassem como os parâmetros do programa devem ser ajustados em cada situação para que sua eficácia seja preservada nos mais distintos ambientes. Esse é um dos grandes desafios de uma avaliação de impacto: avaliar a sensibilidade do impacto ao contexto e identificar as adequações necessárias para que o programa seja apropriado às mais variáveis circunstâncias.

1.7. Metodologias e tipos de avaliação de impacto

Como já discutido em outras seções deste texto, a magnitude do impacto de um programa sobre dado resultado é a diferença entre o resultado que uma unidade de intervenção teria na presença comparado com o que teria na ausência do programa, e apenas uma das duas situações pode ser observada. Nesse caso, o impacto de um programa social poderia ser estimado? A resposta é simples: não poderia.

Entretanto, se em vez de estimar o impacto do programa sobre cada uma das unidades de intervenção for estimado apenas a média desses impactos sobre uma ampla população e, supondo que fosse possível manter um amplo segmento dessa população selecionado ao acaso, fora do programa, nesse caso, haveria, para todos os momentos após a intervenção, um estimador não enviesado do resultado médio que aqueles beneficiados pelo programa teriam caso não tivessem participado do programa, o resultado médio dentre aqueles não beneficiados pelo programa. Com isso, a simples diferença entre o resultado médio para os grupos beneficiados e não beneficiados seria uma estimativa não enviesada do impacto do programa.

Esse exemplo tem duas mensagens gerais e uma particular. A primeira geral é que tipicamente uma característica da distribuição de impacto possível de ser estimada é a média. A segunda característica geral é que alguma forma de seleção por sorteio é essencial.

A particular, que pode ser descartada, é o requerimento de que exista apenas um sorteio de subpopulação representativa da população total. Em geral, a população original pode ser arbitrariamente estratificada e, de cada estrato, uma subpopulação retirada ao acaso. Nesse caso, a média ponderada dos contrastes entre os resultados médios para beneficiários e não-beneficiários em cada estrato é uma medida do impacto médio do programa.

Vale ressaltar que a característica efetivamente requerida para que essas estimativas levem a medidas não enviesadas da magnitude do impacto é que, condicionado nos atributos utilizados na estratificação da população, o resultado médio na ausência do impacto seja o mesmo para beneficiários e não beneficiários. Assim, mantida essa hipótese, estimativas não enviesadas da magnitude do impacto podem sempre ser obtidas contrastando-se resultados para beneficiários e não beneficiários, controlando-se pelos atributos individuais.

Um caso extremo de seleção baseada em observáveis é o caso em que dependendo do estrato (baseado em atributos individuais observáveis) ou todos participam ou todos não participam do programa. Quando esses estratos podem ser ordenados e, em todos com ordenação abaixo de um dado valor, não há nenhum participante e, em todos acima desse valor, todos participam, uma medida do impacto pode ser obtida contrastando-se o resultado médio no primeiro grupo de participantes com menor escore com o resultado médio do último grupo de não participantes (com maior escore). Nesse caso, desconsiderando-se a diferença de escore entre os participantes

e não participantes, o resultado para o grupo não participante aponta para o que seria o resultado do grupo participante na ausência do programa.

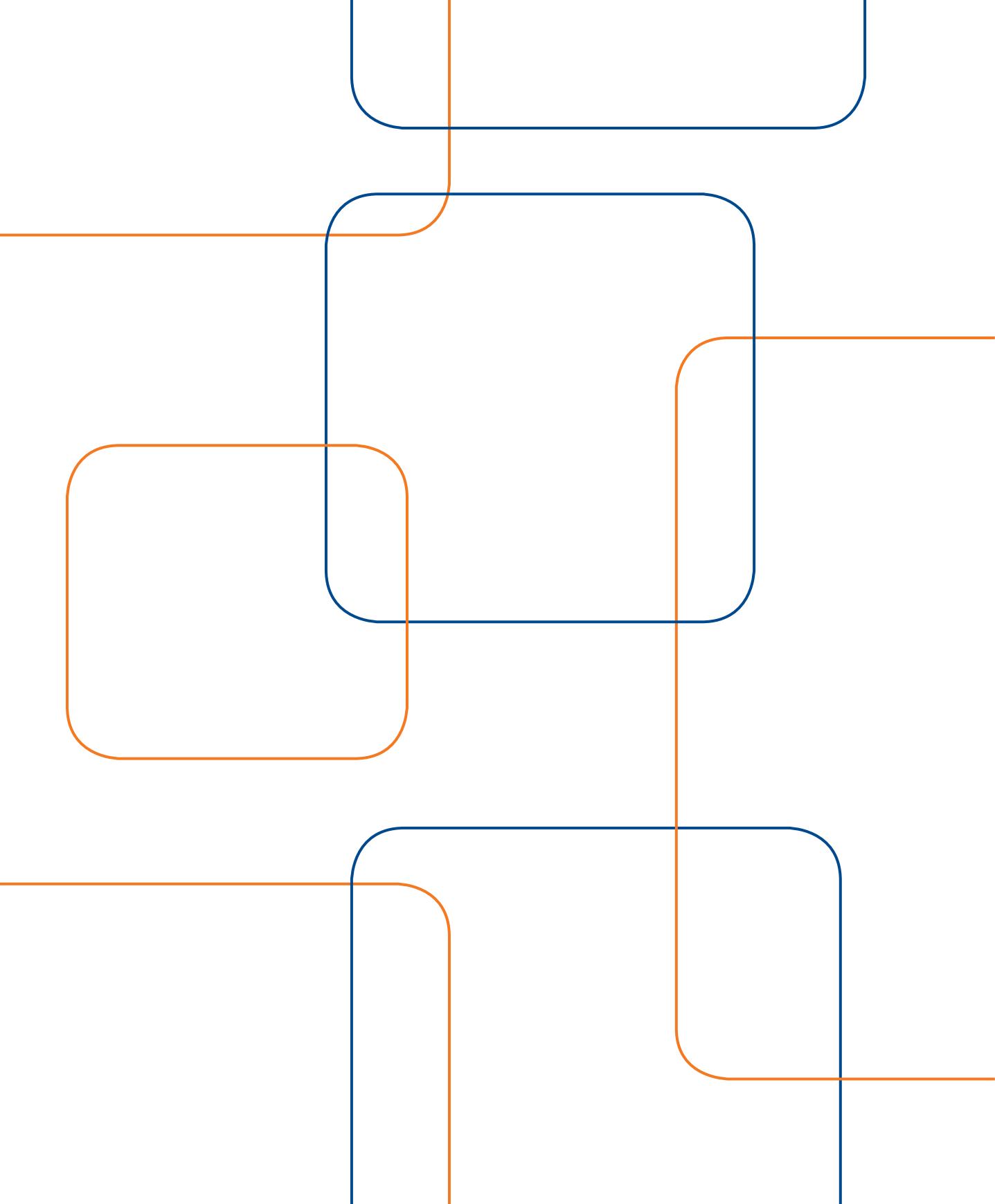
Toda essa discussão indica que o sucesso de uma avaliação de impacto depende de dois fatores. Por um lado, toda avaliação requer algum controle sobre o processo de seleção dos beneficiários. Ao final, para que estimativas não enviesadas estejam disponíveis, é necessário que o conjunto dos potenciais beneficiários seja estratificado e que a seleção dos beneficiários, no interior de cada estrato, seja por sorteio. Por outro lado, toda avaliação precisa de um bom sistema de monitoramento, uma vez que como ressalta Lord Kelvin, “aqueilo que não se mede, não se pode saber se melhorou”. É o sistema de monitoramento que irá oferecer à avaliação as medidas necessárias dos resultados de interesse nos grupos de tratamento e controle.

1.8. Considerações finais

Na nossa sociedade, formuladores de políticas continuamente propõem uma grande variedade de soluções visando a atender às mais diversas necessidades da população. Entretanto, dada a escassez dos recursos disponíveis, nem todas essas soluções podem ser efetivamente implementadas. É necessário, portanto, escolher bem. Quais das propostas apresentadas abrangem a maior parte das necessidades da população ao menor custo? Em outras palavras, quais dessas propostas representam o melhor uso do orçamento disponível?

Um dos instrumentos fundamentais para se obter a resposta a essas questões é a avaliação de impacto, já que esse instrumento permite isolar a contribuição de ação específica para o alcance dos resultados de interesse. No entanto, isolar o impacto de uma ação não é tarefa fácil. Este livro trata das metodologias disponíveis para identificar e estimar o impacto de intervenções implementadas por agentes governamentais ou não governamentais.

Neste capítulo inicial, procuramos apresentar ao leitor por que, para que e quando deve-se realizar uma avaliação de impacto. A nossa discussão sinaliza que a avaliação de impacto é uma ferramenta valiosa que serve a propósitos múltiplos, sendo assim relevante não apenas para aqueles diretamente envolvidos no programa, mas potencialmente para vários outros agentes da sociedade.



Capítulo 2

Modelo de Resultados Potenciais

Miguel N. Foguel

O problema central da área de avaliação de programas é construir o contrafactual do grupo tratado pelo programa. Na forma mais simples de apresentar esse problema, podemos pensar que qualquer indivíduo se encontra em uma de duas situações distintas: ter sido tratado ou ter sido tratado não pelo programa. Idealmente, o melhor grupo de comparação para os indivíduos tratados seria formado pelos mesmos indivíduos na situação em que eles não fossem expostos ao programa. Contudo, essas situações são mutuamente exclusivas: claramente não é possível observar os mesmos indivíduos na condição de tratados e de não tratados ao mesmo tempo. O desafio do avaliador, portanto, é encontrar um grupo de indivíduos que represente adequadamente a situação de não tratamento, ou seja, um grupo que funcione como um bom contrafactual do grupo tratado.

Uma série de pontos precisam ser levados em consideração para se encontrar um grupo de controle que possa replicar o grupo tratado caso ele não tivesse passado pela intervenção. Naturalmente, a escolha desse grupo de comparação envolverá o uso de procedimentos e hipóteses cuja finalidade básica é minimizar o que se costuma denominar de viés de seleção. Para entender melhor o que está por trás desse viés, apresentaremos neste capítulo o arcabouço mais utilizado na área de avaliação, a saber, o arcabouço de *resultados potenciais*. Um elemento crucial para o bom entendimento desse arcabouço – e da área de avaliação de impacto em geral – é ter sempre em mente a potencial existência de uma situação contrafactual.

Antes de apresentarmos formalmente esse arcabouço, vale a pena analisar dois métodos “ingênuos”, porém amplamente utilizados por não especialistas para construir o grupo de controle. O primeiro é baseado na comparação do que ocorre com o grupo de tratamento antes e depois da intervenção. O segundo consiste em comparar o grupo que escolhe passar pelo programa com o grupo que decide não participar da intervenção. Como veremos, esses métodos dificilmente são capazes de fornecer o impacto do programa.

2.1. Comparando os tratados antes e depois

Considere uma situação em que um programa de treinamento profissional foi oferecido para trabalhadores de baixa escolaridade numa certa localidade, por exemplo, um município ou uma área de uma cidade. Esse programa consistia em um curso ministrado por profissionais e tinha como objetivo aumentar o rendimento do trabalho dos participantes. Suponha que o total de inscritos tenha sido igual ao número de vagas oferecidas e que todos os inscritos tenham participado do programa.

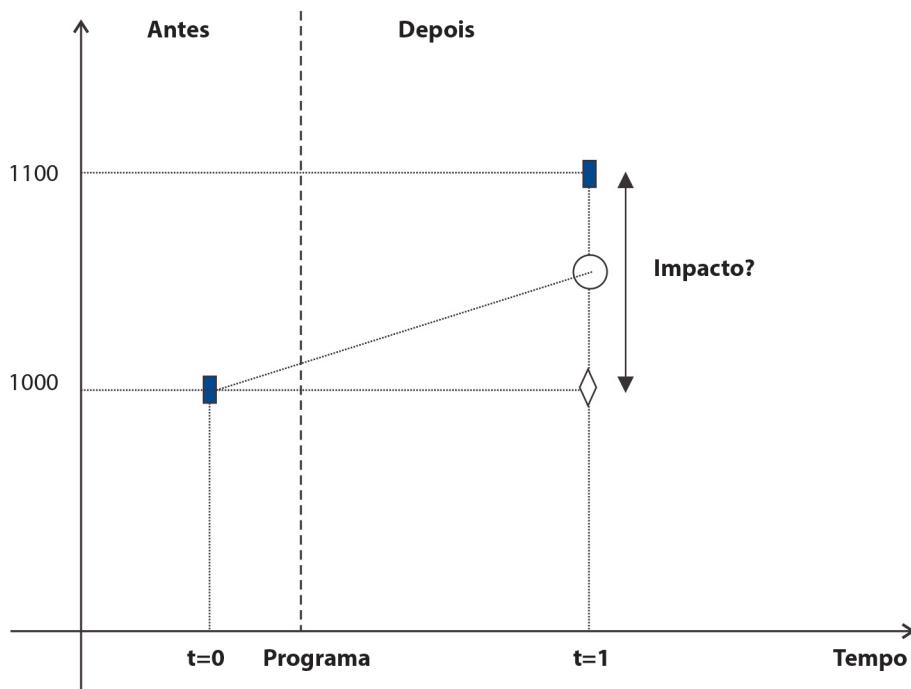
Um pouco antes do começo do programa (por exemplo, no dia da inscrição), todos os participantes responderam a um questionário que levantou informações sobre suas características socioeconômicas, incluindo a situação de emprego e rendimento laboral naquele momento. Suponha que o rendimento médio do trabalho para todos os indivíduos que entraram no programa de treinamento tenha sido de R\$ 1.000. Um ano depois, os mesmos tipos de informações foram novamente coletadas com todos os participantes do programa. O rendimento médio encontrado nesse segundo momento foi igual a R\$ 1.100.

A Figura 2.1 ilustra as informações obtidas para o rendimento médio para os períodos anterior e posterior ao programa¹. Os retângulos estão associados aos valores observados. Alguns diriam que o efeito do programa foi de R\$ 100 (R\$ 1.100 – R\$ 1.000), o que equivale a um aumento de 10% no rendimento médio inicial. Mas será que podemos afirmar que esse foi o verdadeiro efeito do programa?

1. A ocorrência do programa foi simplificada para se dar no momento demarcado pela linha vertical tracejada.

Figura 2.1: Rendimento médio dos participantes antes e depois do programa

Rendimento Médio (R\$)



Para responder a essa pergunta, temos que nos questionar se o grupo de tratamento *antes* do programa representa o contrafactual adequado para a situação de não tratamento *depois* do programa. Ou seja, quanto à Figura 2.1, será que o losango representa o valor que seria recebido pelo grupo tratado depois do programa, caso ele não tivesse passado pela intervenção? A resposta a essa pergunta seria normalmente negativa, já que há uma série de fatores que podem ter afetado o rendimento médio do grupo tratado além do programa. Por exemplo, é possível que a situação do mercado de trabalho da localidade onde residem os participantes do programa tenha melhorado ao longo dos doze meses entre as coletas das informações. Para ilustrar essa possibilidade, suponha que essa mudança nas condições do mercado de trabalho tenha sido responsável por levar o rendimento médio para o ponto associado à forma circular na Figura 2.1. Nesse caso, a variação

do rendimento médio que pode ser atribuída ao programa deixa de ser os R\$ 100 estimados inicialmente. Esse exemplo mostra que, a menos que possamos controlar a influência do mercado de trabalho e de todos os outros fatores que afetam o rendimento médio do trabalho, a simples comparação entre o que ocorre com o grupo de tratados antes e depois do programa não identifica corretamente o efeito causal de uma intervenção².

2.2. Comparando os tratados com os não tratados

Outra prática “ingênu” comumente usada pelos não especialistas é a que compara o grupo de tratados diretamente com um grupo de não tratados. Entretanto, apenas por não ter passado pela intervenção não significa que o grupo de não tratados representa bem o que ocorreria com o grupo de tratamento caso este não tivesse sido tratado. Em outras palavras, a ausência do tratamento para alguns não gera automaticamente o contrafactual de não tratamento para outros.

As principais razões para isso estão relacionadas com as diferenças nas características observáveis e não observáveis – pelo avaliador – entre os dois grupos. Por exemplo, é possível que, mesmo que os grupos sejam parecidos – em média – em atributos observáveis como gênero, idade, escolaridade, renda familiar, etc., eles difiram em habilidades, motivação, esforço, e outras características que tipicamente não estão disponíveis nas bases de dados utilizadas nas avaliações de impacto. Na medida em que muitas dessas características podem afetar a variável de resultado sobre a qual se quer medir o impacto da intervenção, o uso da comparação direta entre os dois grupos não necessariamente mede o efeito causal do programa. Afinal, as diferenças nas características não observáveis entre os grupos poderiam estar misturadas ao efeito do programa.

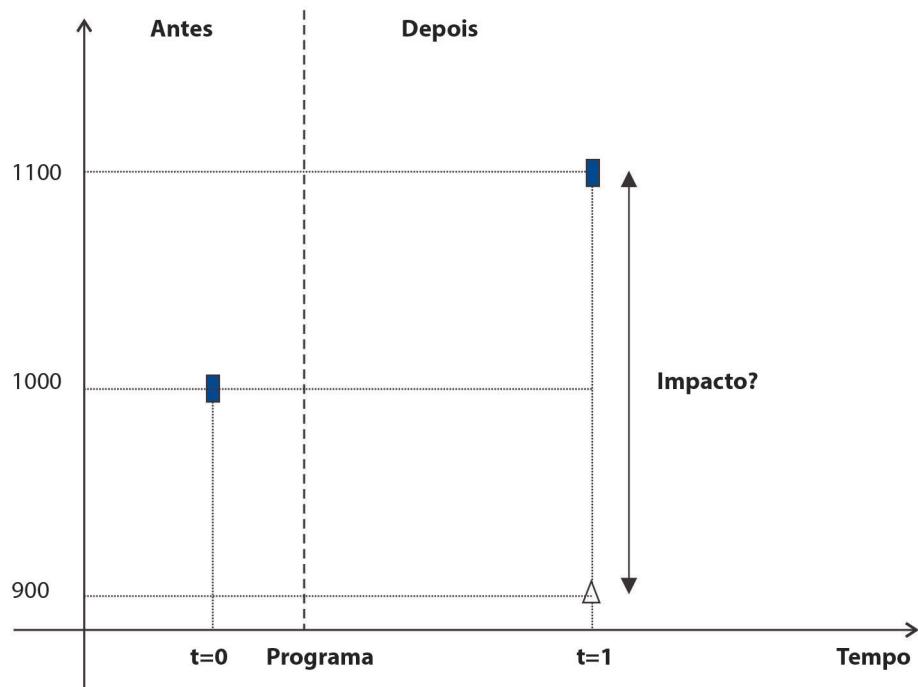
Para ilustrar esse problema, continuaremos com a situação hipotética da subseção anterior. Assim, admita que o segundo questionário usado com os participantes do programa de treinamento tenha sido aplicado a uma amostra de pessoas da mesma localidade que não

2. Vale observar que um conjunto de variáveis observadas nos dados referentes aos próprios indivíduos podem ter mudado entre os dois períodos. Essas mudanças podem afetar a diferença de rendimento médio observada. Por exemplo, como a idade tende a influenciar o rendimento médio das pessoas, o fato de que se passou um ano entre os períodos de comparação seria um fator adicional a explicar parte da diferença observada.

tenham participado do programa. Suponha que a média do rendimento do trabalho obtida com as informações coletadas com esse grupo de não tratados após o programa tenha sido de R\$ 900. Suponha que os dois grupos sejam idênticos, em média, nas características levantadas no questionário, que não incluiu nenhuma tentativa de medir a motivação das pessoas. A Figura 2.2 contém as informações para os dois grupos para o período depois do programa, onde o triângulo representa o rendimento médio do grupo de não participantes.

Figura 2.2: Rendimento médio dos tratados e não tratados

Rendimento Médio (R\$)



Embora as pessoas dos dois grupos sejam parecidas em média em diversas características, não podemos afirmar que os R\$ 200 de diferença no rendimento médio dos grupos (ou seja, a distância entre o retângulo e o triângulo em t=1) medem corretamente o impacto do programa.

de treinamento.³ Para isso, basta admitirmos que a motivação (que não foi medida) seja uma característica que afeta o desempenho das pessoas no mercado de trabalho, incluindo o seu salário. Se os dois grupos diferem em termos de motivação – diferença essa que pode explicar por que o grupo tratado decidiu participar do programa e o grupo não tratado preferiu não participar –, então é provável que parte dos R\$ 200 de diferença entre os grupos se deva ao fator motivacional. Se isso ocorrer, não podemos tratar essa quantia como o efeito causal do programa de treinamento.

Os próximos capítulos desta parte do livro serão dedicados a apresentar um conjunto amplo de metodologias que buscam isolar o efeito dos programas dos efeitos causados por outros fatores que afetam a(s) variável(eis) de resultado de interesse. O arcabouço conceitual que apresentaremos na próxima seção tem a finalidade de facilitar o entendimento dessas metodologias.

2.3. O arcabouço de resultados potenciais

Para apresentar o arcabouço de modo formal, utilizaremos o algarismo 1 para denotar a situação em que o indivíduo é tratado e o algarismo 0 para a situação contrafatual na qual esse mesmo indivíduo não é tratado. Estamos interessados em avaliar o impacto de um programa (por exemplo, um programa de qualificação de mão de obra) sobre uma variável de resultado (por exemplo, o rendimento do trabalho). Denotando por Y essa variável, podemos pensar que seu valor para indivíduo i será $Y_i(1)$ caso ele esteja na situação de tratado, mas poderia assumir, pelo menos potencialmente, o valor $Y_i(0)$ caso ele não seja tratado. A vantagem de apresentar o problema dessa forma é explicitar que cada uma das situações está associada a um resultado potencialmente distinto para um mesmo indivíduo. Assim, se fosse possível observar o indivíduo i nas duas situações, a diferença $\beta_i = Y_i(1) - Y_i(0)$ forneceria o impacto do programa para esse indivíduo. Naturalmente, essa impossibilidade de observar as duas situações se estende para todos os indivíduos da população.

3. Nesse exemplo, assumimos que não há informações para o grupo de não tratados antes da intervenção. Essas informações serão introduzidas no capítulo 4 dentro do contexto do método de diferenças em diferenças.

Cada indivíduo i possui um par de resultados potenciais ($Y_i(1)$, $Y_i(0)$) e, embora somente um deles seja observado na prática, é interessante pensar em cada um de forma separada. Uma maneira bastante geral de especificar cada resultado potencial é pelo par de equações:

- (1) $Y_i(0)=\alpha+\varepsilon_i$
- (2) $Y_i(1)=\alpha+\beta_i+\varepsilon_i$

onde α é o intercepto e ε_i é o componente não observável (pelo avaliador) que afeta os resultados potenciais do indivíduo i .⁴ Note que β_i é o que diferencia as duas equações, mostrando que os indivíduos para os quais o programa tem efeito positivo (negativo) apresentam resultado potencial na situação de tratamento, $Y_i(1)$, maior (menor) que o resultado que eles teriam caso não tivessem sido tratados, $Y_i(0)$.

Naturalmente, não é possível conhecer o valor de β_i para cada indivíduo, pois somente um dos dois resultados potenciais se realiza concretamente. Esse problema pode ser pensado como um problema incontornável de dados faltantes (*missing*, em inglês), pois não há um modo de conhecer a situação contrafáctual para cada indivíduo. Há, no entanto, objetos agregados que podem ser identificados sob certas condições que serão discutidas nos demais capítulos do livro. Dois dos objetos mais utilizados na área de avaliação são o efeito médio do programa (**EMP**) e o efeito médio do programa sobre os tratados (**EMPT**). O primeiro é definido por:

$$(3) \ EMP=E[Y_i(1)-Y_i(0)]=E[\beta_i],$$

e representa a média do efeito do programa para todos os indivíduos da população, independentemente de quem participou ou não do programa. Já o segundo foca apenas no subgrupo que participa do programa e é definido por:

$$(4) \ EMPT=E[Y_i(1)-Y_i(0)/T_i=1]=E[\beta_i/T_i=1],$$

4. As equações (1) e (2) poderiam incluir termos com as características observadas nos indivíduos. Não incluímos essas características aqui, pois elas não são necessárias para o entendimento do argumento principal deste capítulo e facilita a exposição. Essas características serão apropriadamente introduzidas nos modelos específicos dos demais capítulos.

onde T_i uma variável binária que assume valor igual a 1 se o indivíduo participa do programa e valor 0 caso ele não participe.⁵ Note que esse parâmetro capta a média do efeito do programa somente para a subpopulação de indivíduos tratados pelo programa e, nesse sentido, ele pode diferir do efeito médio do programa para a população como um todo. O parâmetro $EMPT$ é útil para medir o efeito de programas cuja participação é voluntária, ou seja, programas que tipicamente atendem um público com características diferenciadas em relação à população como um todo. Como a maior parte dos programas públicos tem essa característica, o $EMPT$ é um parâmetro de grande interesse na área de avaliação de impacto.⁶

Utilizando o par de resultados potenciais ($Y_i(1)$, $Y_i(0)$) e a participação ou não no programa ($T_i=1$, $1-T_i=0$), podemos expressar o resultado efetivamente observado para o indivíduo i pela equação:

$$(5) \quad Y_i = T_i Y_i(1) + (1-T_i) Y_i(0) = Y_i(0) + T_i(Y_i(1) - Y_i(0)).$$

Note que ela informa o resultado observado para qualquer indivíduo i na população, pois quando ocorre a participação no programa observamos o resultado na situação de tratado, $Y_i(1)$, ao passo que quando não há participação observamos o resultado na situação de não tratado, $Y_i(0)$.

Se substituirmos as equações (1) e (2) na equação (5), esta última se transforma em:

$$(6) \quad Y_i = \alpha + \beta_i T_i + \varepsilon_i,$$

que é uma equação de regressão linear cuja única diferença em relação àquelas tipicamente encontradas em livros de estatística e econometria básica é que o coeficiente que capta o efeito do programa, β_i , varia entre os indivíduos. Esse modelo, também chamado de modelo de regressão linear com coeficientes (lineares) aleatórios, é bastante geral e permite que o programa tenha efeitos heterogêneos entre os indivíduos da população.

5. Embora menos utilizado, há outro objeto que aparece na literatura que é o efeito médio do programa sobre os não tratados: $EMPNT = E[Y_i(1) - Y_i(0)|T_i=0] = E[\beta_i|T_i=0]$. Cabe observar que o EMP é uma média ponderada do $EMPT$ e do $EMPNT$, em que os pesos são dados pelas proporções de tratados e não tratados na população.

6. Em situações nas quais o gestor intenciona expandir significativamente o programa, o EMP (ou o $EMPNT$, ver nota de rodapé 5) passa a ser um parâmetro de grande interesse.

Em muitas avaliações, no entanto, supõe-se que o impacto do programa é igual para todos os indivíduos, ou seja, faz-se a hipótese de que o efeito de tratamento é homogêneo. Naturalmente, assumir homogeneidade do impacto é uma hipótese forte, pois parece mais razoável esperar que o efeito de grande parte dos programas seja heterogêneo entre as pessoas. Contudo, sob a hipótese de homogeneidade, $\beta_i = \beta$ para qualquer indivíduo da população, a equação (6) se especializa num modelo de regressão linear com coeficientes constantes:⁷

$$(7) \quad Y_i = \alpha + \beta T_i + \varepsilon_i.$$

O maior problema para identificar o impacto do programa por meio da equação (7) é que a variável T_i é potencialmente correlacionada com o componente não observável ε_i . Na realidade, é bastante provável que isso ocorra, já que a participação no programa normalmente não se dá de forma aleatória. Ao contrário, a participação geralmente decorre de uma decisão voluntária do indivíduo (ou da família), a qual pode ter sido tomada com base em informações que não são observadas pelo avaliador e, portanto, estão embutidas no componente ε_i da equação (7). Um exemplo foi apresentado na seção 2.2, no qual os indivíduos que participam de programas de treinamento são considerados mais motivados. Como motivação é uma dimensão do comportamento das pessoas que pode afetar Y_i (por exemplo, o rendimento do trabalho) e dificilmente é medida nos dados (ou seja, passa a ser incorporada em ε_i), aparece um problema de correlação entre T_i e ε_i que precisa ser resolvido para que o parâmetro de interesse, β , seja corretamente estimado.

Esse problema de estimação incorreta de β devido à correlação entre a participação no programa (T_i) e a parte não observável (ε_i) geralmente recebe o nome de problema de seleção, sendo sua magnitude chamada de *viés de seleção*. Sua denominação decorre do fato de que os próprios indivíduos, o governo ou a entidade gestora do programa selecionam as pessoas (ou famílias) que receberão as ações da intervenção. Se essa seleção estiver associada a características não observáveis dos indivíduos, esse grupo se torna um grupo “especial” da população (por exemplo, mais motivados), o que pode afetar diversas variáveis de resultado, tornando difícil a identificação do efeito causal do programa.

7. Claramente, a hipótese de que o efeito do programa é igual para todos na população torna o $EMP = E[\beta_i] = \beta = E[\beta_i | T_i = 1] = EMPT$.

Boa parte do esforço de pesquisa realizado até os dias de hoje na área de avaliação tem sido dedicado a encontrar soluções para contornar o problema da seleção. A fim de apresentá-lo de forma geral, vamos retornar ao caso de efeitos heterogêneos da intervenção. Para facilitar o entendimento, considere as seguintes médias populacionais da variável de resultado para os grupos de tratados e não tratados:

$$E_{11}=E[Y_i(1)/T_i=1]$$

$$E_{10}=E[Y_i(0)/T_i=1]$$

$$E_{01}=E[Y_i(1)/T_i=0]$$

$$E_{00}=E[Y_i(0)/T_i=0].$$

Note que somente E_{11} e E_{00} são médias que podem ser calculadas diretamente dos dados.⁸ Já E_{10} e E_{01} representam médias contrafutais, a primeira correspondendo à média para os tratados, caso eles não tivessem sido tratados; e a segunda, à média para os não tratados, caso eles tivessem sido tratados. Para ver como o viés de seleção aparece, considere a identificação do efeito médio do programa sobre os tratados (*EMPT*). Pela equação (4), temos:

$$(8) \quad EMPT=E[Y_i(1)-Y_i(0)/T_i=1]=E[Y_i(1)/T_i=1]-E[Y_i(0)/T_i=1]=E_{11}-E_{10},$$

ou seja, é a diferença entre a média da variável de resultado para os tratados na situação de tratados e a média para esses mesmos indivíduos na situação contrafactual. Como mencionado anteriormente, o desafio é encontrar um grupo que represente adequadamente o contrafactual do grupo tratado. É comum o uso de um grupo de indivíduos não tratados como o grupo que representa o contrafactual desejado – por exemplo, o grupo de não participantes do exemplo hipotético do programa de treinamento da seção 2.1. Quando isso ocorre, geralmente calcula-se a diferença:

$$(9) \quad R=E[Y_i(1)/T_i=1]-E[Y_i(0)/T_i=0]=E_{11}-E_{00},$$

8. Por exemplo, se tivermos uma amostra aleatória da população com N_1 indivíduos tratados e N_0 indivíduos não tratados, podemos estimar E_{11} por meio de $\hat{E}_{11} = N_1^{-1} \sum_{i=1}^{N_1} Y_{1i}$ e E_{00} via $\hat{E}_{00} = N_0^{-1} \sum_{i=1}^{N_0} Y_{0i}$, onde Y_{1i} e Y_{0i} correspondem às variáveis de resultado respectivamente para os indivíduos que participaram e não participaram do programa.

que é tomada como o impacto médio do programa. No entanto, isso pode ser um procedimento incorreto, uma vez que o grupo de não tratados não necessariamente funciona como um bom contrafactual para o grupo de tratados, caso esses não tivessem passado pela intervenção. Para ver isso formalmente, adicionaremos e subtrairemos da equação (9) a média contrafactual $E_{10}=E[Y_i(0)/T_i=1]$:

$$\begin{aligned} (10) \quad R &= \{E[Y_i(1)/T_i=1] - E[Y_i(0)/T_i=0]\} - \{E[Y_i(0)/T_i=1] - E[Y_i(0)/T_i=1]\} \\ &= \{E[Y_i(1)/T_i=1] - E[Y_i(0)/T_i=1]\} + \{E[Y_i(0)/T_i=1] - E[Y_i(0)/T_i=0]\} \\ &= \{E_{11} - E_{10}\} + \{E_{10} - E_{00}\} = EMPT + V. \end{aligned}$$

Como se pode ver, R não é igual ao $EMPT$ por conta da existência do termo $V=E[Y_i(0)/T_i=1]-E[Y_i(0)/T_i=0]$, que corresponde ao que denominamos viés de seleção. Note que ele corresponde à diferença de médias da variável de resultado na situação de não tratamento (isto é, $Y_i(0)$), quando os indivíduos participam ($T_i=1$) ou não ($T_i=0$) do programa. Para que V seja igual a 0, a média observada para os não tratados, E_{00} , deve ser igual à média do contrafactual que buscamos, E_{10} . O problema é que não observamos E_{10} e, portanto, não temos como saber a magnitude do viés de seleção embutido em R .

Nossa tarefa nos próximos capítulos será apresentar métodos que buscam solucionar o problema do viés de seleção. Cada método lança mão de certos procedimentos e hipóteses para tentar resolver esse problema. Embora cada metodologia será discutida em detalhes em cada capítulo, faremos aqui um sumário das hipóteses básicas utilizadas.

Considerado o “padrão-ouro” da área de avaliação, o primeiro método baseia-se na aleatorização de indivíduos, famílias ou até mesmo de comunidades inteiras para participar ou não do programa. Esse procedimento de aleatorização gera dois grupos experimentais, o de tratamento, formado pelos indivíduos que foram aleatorizados para participar do programa, e o grupo de controle, composto pelos que foram aleatorizados para ficar de fora. O fato de que a participação ou não no tratamento – ou seja, na notação introduzida neste capítulo, a variável T_i assumir valor 1 ou 0, respectivamente – é definida pelo procedimento de aleatorização, garantindo que os grupos de tratamento e controle sejam parecidos tanto nas características observáveis quanto nas não observáveis. O viés de seleção fica, dessa forma, contornado, permitindo que a comparação das médias da variável de resultado entre os grupos identifique o efeito causal do programa.

Outro método que discutiremos em detalhes é chamado diferenças em diferenças (*DD*). Esse método requer dados em painel das unidades tratadas e não tratadas, com períodos de observação antes e depois da intervenção. Ele recebe esse nome porque duas diferenças são calculadas: a diferença entre os períodos de tempo para cada unidade observada e a diferença entre as unidades para cada período de tempo. Como os dados são em painel, o método comporta a presença de um termo invariante no tempo para cada unidade de observação. Esse termo, que é eliminado no cômputo da diferença temporal para cada unidade, capta todas as características não observadas da unidade que são constantes no tempo. A principal hipótese desse método é que a trajetória da variável de resultado do grupo de controle reflita a do grupo de tratamento na ausência da intervenção. Embora essa hipótese não seja testável, uma indicação de sua validade ocorre quando os grupos possuem uma mesma trajetória temporal da variável de resultado antes da intervenção. A ideia é que, se as trajetórias dos grupos são semelhantes no período pré-programa, então aumentam as chances de que a trajetória do grupo de controle após o programa represente bem a trajetória que o grupo de tratamento teria na ausência da intervenção. Nesse caso, qualquer diferença entre os grupos que apareça após o programa pode ser interpretada como o impacto da intervenção. Quando as informações disponíveis cobrem um intervalo relativamente extenso antes do programa, o teste de trajetórias iguais entre os grupos pré-programa é passível de ser testada estatisticamente nos dados.

O método denominado “pareamento” busca tornar o grupo tratado semelhante a um grupo de não tratados com base em um conjunto amplo de características observadas dos indivíduos, normalmente denotadas pelo vetor X_i . Exemplos típicos de características observadas que são incluídas em avaliações desse tipo são o gênero, a idade, a escolaridade, a renda, entre outras características das pessoas que sejam consideradas importantes para determinar a participação no programa. A hipótese é que, ao levar em consideração essas características em X_i , a variável de resultado passa a independente da participação ou não dos indivíduos no programa – ou seja, as variáveis $Y_i(0)$ e $Y_i(1)$ tornam-se independentes de T_i condicionais em X_i –, fazendo com que o viés de seleção deixe de operar.

Existem diversas maneiras de tornar os grupos de tratamento e controle semelhantes em características observadas, sendo a mais utilizada conhecida como método do escore de propensão” (MEP). Nesse método, assume-se que a probabilidade de um indivíduo participar do programa depende apenas do conjunto de características observadas em X_i . Estimada essa probabilidade, o

pareamento dos grupos é realizado em termos das semelhanças dessa probabilidade estimada. A ideia por trás do MEP é que, quando a probabilidade de participação é similar entre os indivíduos tratados e não tratados, os dois grupos tornam-se comparáveis, o que permite identificar o efeito causal da intervenção. A hipótese de que apenas as características observáveis são suficientes para contornar o viés de seleção não é passível de ser testada diretamente nos dados.

Um método bastante geral que procura resolver o problema de viés de seleção é o de variáveis instrumentais. Trata-se de uma metodologia na qual se procura encontrar alguma variável (instrumento) que influencie a decisão de participar no programa e ao mesmo tempo não seja correlacionada com as características não observadas dos indivíduos que afetam a variável de resultado – ou seja, uma variável que se correlacione com T_i , mas não com ε_i na equação (6). Como é capaz de afetar a participação no programa, mas não está correlacionada com as características não observadas, essa variável corrige o problema de seleção, permitindo assim identificar o efeito causal de interesse. Naturalmente, a existência de uma variável com essas características vai depender do contexto no qual o programa está inserido e da disponibilidade dos dados. Ademais, tipicamente, não é possível testar a validade da hipótese de que o instrumento não é de fato correlacionado com as características não observáveis dos indivíduos. De todo modo, o método de variáveis instrumentais é uma abordagem bastante geral dentro da área de econometria, sendo também bastante empregada na área de avaliação de impacto.

Muitos programas e políticas definem sua população-alvo com base em critérios que estabelecem valores de corte para quem pode ser ou não elegível a receber ou participar das ações da intervenção. Por exemplo, o programa Bolsa Família estabelece que somente a família cuja renda *per capita* está abaixo de certa linha de corte pode receber a transferência monetária prevista pelo programa. Outro exemplo é o programa de Jovens Aprendizes do governo federal, que fixa uma idade abaixo da qual um jovem estudante pode ser contratado com benefício fiscal por uma empresa no Brasil. O método denominado “regressão descontínua” procura explorar a existência desse tipo de descontinuidade nos critérios de elegibilidade dos programas para criar os grupos de tratamento e controle. A ideia básica é que os indivíduos não elegíveis que estão próximos à linha de corte (por exemplo, acima dela) são comparáveis em características observáveis e não observáveis aos indivíduos elegíveis que também estão próximos a essa linha (por exemplo, abaixo dela). Como a única diferença entre esses dois grupos de cada lado do valor de corte é o fato de um poder participar do programa e outro não, o problema de seleção fica, pelo menos em princípio, resolvido.

Assim, sob a hipótese de que não há uma descontinuidade marcante na variável de interesse em torno da linha de corte para o grupo de não elegíveis, a diferença nessa variável entre os dois grupos fornece o impacto do programa. Embora seja um método que só pode ser empregado para avaliar o impacto de intervenções que possuam linhas de corte, ele é considerado bastante consistente para estimar o efeito de programas nessa categoria, pelo menos para os indivíduos que se situam em torno da linha de corte estabelecida pelas regras dos programas.

2.4. Exercícios:

- 1.** Um jornal de grande circulação apresentou em uma matéria um gráfico com a série histórica da incidência de diarreia em crianças de certo município. O gráfico mostrava que a incidência de diarreia permanecia inalterada, e a matéria dizia que o programa de saneamento do governo não teve nenhum impacto sobre a doença. Pode-se considerar essa afirmativa do jornal correta? Por quê?
- 2.** O prefeito de uma capital implantou um programa de alfabetização de adultos no terceiro ano de seu mandato. No ano seguinte, em campanha eleitoral, ele comparou a taxa de analfabetismo da sua cidade com a do interior do estado para mostrar que o programa era um sucesso.

 - a.** Defina formalmente o parâmetro que deveria ser utilizado para avaliar o programa de alfabetização na capital.
 - b.** Com base no item anterior, discuta sob que condições o interior do estado seria um bom grupo de controle para avaliar o programa implantado na capital?
- 3.** Qual é a mais importante função do grupo de controle em uma avaliação de impacto? Mostre como o grupo de controle resolve o problema do viés de seleção.
- 4.** Tipicamente, a escolha do curso universitário é uma decisão que envolve uma análise sobre as motivações, interesses, aptidões, os custos pessoais e financeiros de fazer o curso, além de uma previsão sobre a renda que a profissão escolhida pode oferecer no futuro. Tomando o curso universitário como um programa (de treinamento), responda aos seguintes itens:

- a.** Tomando o rendimento do trabalho como a variável de resultado, defina as equações dos resultados potenciais para indivíduos com e sem curso universitário.
- b.** Supondo que o efeito do treinamento universitário é heterogêneo, escreva a equação para a variável de resultado observada e aponte as fontes por onde pode ocorrer viés de seleção. Repita o exercício supondo que o efeito é homogêneo.
- c.** Discuta o principal tipo de problema que pode surgir se compararmos diretamente os rendimentos do trabalho com e sem curso universitário na população.
- 5.** Suponha que fosse possível conhecer os resultados potenciais para os indivíduos que passaram por certa intervenção de acordo com a tabela abaixo:
- | Indivíduo i | T_i | $Y_i(0)$ | $Y_i(1)$ | Y_i | $Y_i(1)-Y_i(0)$ |
|---------------|-------|----------|----------|-------|-----------------|
| 1 | 0 | 15 | 15 | 15 | 0 |
| 2 | 1 | 12 | 16 | 16 | 4 |
| 3 | 1 | 15 | 14 | 14 | -1 |
| 4 | 0 | 15 | 17 | 15 | 2 |
| 5 | 1 | 11 | 14 | 14 | 3 |
- a.** Calcule o efeito médio do programa (EMP) e o efeito médio do programa sobre os tratados ($EMPT$).
- b.** Calcule as médias dos resultados observados para quem passou e não passou pela intervenção.
- c.** Qual a diferença das médias do item (b) para quem passou e não passou pela intervenção?
- d.** Compare a diferença de médias do item (c) com o $EMPT$. Por que eles são diferentes?
- 6.** Usando a equação linear (6) no texto, calcule a diferença. Qual a expressão para o viés de seleção?

Capítulo 3

Método de Aleatorização

Miguel N. Foguel

O principal problema da área de avaliação é encontrar um grupo de controle que represente adequadamente o contrafactual do grupo tratado, ou seja, um grupo de comparação que seja o mais próximo possível do que teria ocorrido com o grupo tratado caso este não tivesse recebido a intervenção. Embora a situação contrafactual não tenha como ser concretamente observada, ela pode ser bem aproximada pelo uso do método de aleatorização. Esse método, muitas vezes chamado de método experimental, é baseado na seleção aleatória dos indivíduos, famílias ou até mesmo comunidades inteiras que farão parte dos grupos de tratamento e controle.¹ Quando bem implementado, o mecanismo da aleatorização fornece o balanceamento necessário tanto das características observadas quanto das não observadas das unidades que compõem os dois grupos. Portanto, por construção, o método permite criar uma situação na qual não há correlação entre ser ou não tratado e os atributos das unidades de observação. O problema do viés de seleção fica resolvido.

A avaliação aleatorizada é utilizada em diversos campos, sendo considerado o procedimento de referência para se estabelecer causalidade e medir o impacto de vários tipos de tratamento. Por exemplo, ele é amplamente empregado nas áreas de medicina e farmacêutica, em que medicações e procedimentos médicos são avaliados com base na aleatorização dos indivíduos que receberão ou não o tratamento em avaliação. Na realidade, tamanha é a importância do método de aleatorização que o seu uso é obrigatório para se comprovar a eficácia (e a existência de eventuais efeitos adversos) dos medicamentos que os laboratórios farmacêuticos pretendem disponibilizar para uso humano.

Além de ser o “padrão-ouro” para determinar a causalidade de um tratamento, o método de aleatorização tem outro apelo importante: sua simplicidade de cálculo. De fato, para obtermos o impacto da intervenção com base nesse método, basta subtrairmos a média da variável de interesse do grupo de tratamento da média correspondente calculada para o grupo de controle. Em condições ideais, essa simples diferença de médias fornece uma estimativa do efeito causal do tratamento. Esse atributo dos experimentos sociais é um atrativo importante, pois torna o método mais transparente para os gestores do programa e para o público em geral.

1. O método de aleatorização recebe vários nomes na literatura especializada, incluindo experimento social (controlado), experimento aleatório, avaliação experimental, avaliação aleatorizada ou método experimental. Usaremos indistintamente essas denominações ao longo deste capítulo e do livro.

Contudo, apesar de suas diversas vantagens, o método de aleatorização não está isento de problemas. Como veremos, a maior parte deles surge de situações práticas que ocorrem após a aleatorização dos indivíduos entre os grupos experimentais. Os principais exemplos são o não comparecimento às atividades do programa por parte de uma parcela do grupo selecionado para o tratamento e a substituição da intervenção por outros programas similares (ou até pelo mesmo programa) por parte de uma parcela do grupo de controle. Em ambos os casos, pode haver mudança de composição das características entre grupos experimentais, que foram selecionados *a priori* de forma aleatória, porém terminaram influenciados *a posteriori* por decisões de participação ou não no programa em avaliação ou em similares. Problemas desse tipo podem causar vieses não negligenciables na estimativa do efeito que se pretendia isolar com o experimento aleatório.

Neste capítulo, apresentaremos as principais propriedades do método de aleatorização, incluindo suas limitações e possíveis soluções. Antes, porém, discutiremos um ponto importante existente na área de avaliação de impacto: a validade das conclusões que podem ser inferidas a partir dos estudos de avaliação.

3.1. Validade interna e validade externa

A validade interna questiona se o desenho da avaliação identifica corretamente o efeito causal de um programa. Quando um estudo possui validade interna, seus resultados conseguem isolar o verdadeiro impacto do programa, isto é, estão livres da influência de outros fatores que não o próprio efeito da intervenção. Quando bem realizadas, as avaliações baseadas em experimentos aleatórios criam um grupo de controle que é estatisticamente equivalente ao grupo de tratamento antes do início do programa. Em uma situação ideal, qualquer diferença observada entre os grupos após a intervenção é atribuída ao efeito causal do programa. Tem-se então que a seleção aleatória dos grupos de tratamento e controle assegura a validade interna de uma avaliação. Ela garante que os resultados da avaliação sejam válidos dentro do contexto específico daquela avaliação.

Já a validade externa é a capacidade de estender os resultados para fora do contexto da avaliação. Em outras palavras, uma avaliação possui validade externa se for possível generalizar as estimativas do impacto do programa avaliado para outras populações e outros momentos no tempo. Um ponto importante para assegurar a validade externa dos resultados em certo ponto no tempo

é que a amostra utilizada na avaliação seja representativa da população-alvo do programa naquele momento. Normalmente, quando a amostra é obtida por métodos de amostragem aleatória da população-alvo, essa representatividade é assegurada.

3.1.1. Validade interna

A característica básica de um experimento aleatório é o uso de um mecanismo que seleciona aleatoriamente os indivíduos a participar ou não do programa. Quando essa seleção aleatória é feita de forma correta, cada indivíduo pertencente ao conjunto de potenciais participantes e possui a mesma probabilidade de ser selecionado a participar do programa, independentemente de suas características pessoais. Em outras palavras, quando o processo de aleatorização é bem realizado, devemos esperar que os grupos de tratamento e controle sejam semelhantes tanto nas suas características observadas quanto nas não observadas. A única diferença que se espera entre os dois grupos é que o grupo de tratamento participe do programa, ao passo que o de controle não passe pela intervenção.

A ideia de que deve haver um balanceamento de características entre os dois grupos aleatorizados é bastante natural. Para ver isso, considere uma situação hipotética em que há 2 mil indivíduos elegíveis para frequentar um programa de treinamento básico para o uso de planilhas eletrônicas em computadores. Suponhamos que, nesse total de duas mil pessoas, 1.200 (60%) sejam homens, 1.000 (50%) tenham cabelo preto e 400 (20%) sejam hábeis com números. Admita que esta última característica seja difícil de medir e, nesse sentido, torna-se não observada para nós. O que deveríamos esperar do balanceamento dessas características, se os 2 mil potenciais participantes fossem aleatorizados em dois grupos com 1.000 indivíduos em cada? É intuitivo esperar que cada grupo fosse composto de aproximadamente 60% de homens, 50% de pessoas de cabelo preto e 20% com habilidades com números, ou seja, devemos esperar que os dois grupos na média sejam semelhantes entre si (e com o conjunto total de indivíduos).

Note que essa semelhança deve se dar tanto nas características observadas (sexo e cor do cabelo) quanto nas não observadas (habilidade com números, no nosso exemplo). Isso é importante, pois é a aleatorização que garante que, pelo menos em média, os dois grupos sejam similares em todas as dimensões. Na realidade, é essa equivalência entre os grupos que permite utilizar o grupo de controle como contrafactual para o grupo de tratamento. Esse é o principal motivo que faz com que o método de aleatorização seja considerado o “padrão-ouro” de estudos de efeitos causais.

Em muitas áreas, inclusive, a validade interna de uma avaliação de impacto só é aceita se for baseada na aleatorização dos que vão ou não receber o tratamento que se quer estudar. Por exemplo, na área de farmacêutica, os medicamentos só são autorizados para uso geral se os laboratórios conduzirem, com sucesso, um experimento em que o grupo que recebe a medicação e o que não recebe (ou recebe um placebo) são selecionados de forma aleatória.

3.1.2. Validade externa

Sem a validade interna de uma avaliação, pouca utilidade ela terá para revelar o verdadeiro efeito da intervenção. No entanto, mesmo quando a validade interna está assegurada, pode ocorrer de a avaliação ter sido realizada para um grupo de pessoas que não é representativo da população-alvo do programa. Por exemplo, se a avaliação de um programa nacional foi feita somente para uma localidade (bairro, cidade ou estado), não é possível generalizar os resultados da avaliação para o país como um todo. Em casos como esses, não se pode garantir a validade externa da avaliação, ou seja, os resultados não são extensíveis a outros grupos ou contextos.

Uma dimensão relevante para assegurar a validade externa é que a amostra de indivíduos utilizada na avaliação seja representativa da população de elegíveis a receber o programa. Por exemplo, a avaliação de impacto de um programa de qualificação profissional para jovens de ambos os sexos não terá validade externa se a avaliação for baseada numa amostra não representativa desses grupos demográficos. De fato, se a amostra contiver indivíduos adultos ou for formada somente por homens, as conclusões da avaliação não podem ser generalizadas para a toda a população-alvo do programa. No entanto, não basta que a amostra só contenha indivíduos da população-alvo; é necessário também que os indivíduos da amostra representem acuradamente essa população.

Na prática, a forma mais utilizada para garantir a representatividade de uma amostra é que ela seja uma amostra aleatória da população que ela pretende representar.² As mais importantes pesquisas

2. Para tanto, ela deverá ter sido escolhida com base em algum método de amostragem probabilística. Há uma diversidade de métodos para selecionar uma amostra aleatória de uma população. Entre os mais empregados estão as amostragens simples e estratificada. Foge do escopo deste livro um tratamento mais amplo dos métodos de amostragem. Uma apresentação desses métodos pode ser encontrada em Deaton (1997) ou no livro clássico de Cochran (1977).

socioeconómicas conduzidas pelo IBGE e outros institutos oficiais de estatística do mundo são baseadas em amostras aleatórias retiradas da população dos países. No caso de uma amostra para fins de avaliação de impacto de um programa, a representatividade da população de elegíveis está assegurada se a amostra a ser empregada na avaliação for escolhida de forma aleatória a partir dessa população.

Note que a aleatorização que permite assegurar a validade externa de uma avaliação não garante a validade interna dessa avaliação, e vice-versa. Se a avaliação utiliza uma amostra aleatória da população-alvo do programa, mas a participação no tratamento não é realizada pela aleatorização, então a amostra deve ser representativa da população de elegíveis, porém os resultados da avaliação podem não ter validade interna. Distintamente, caso a amostra não seja aleatória da população-alvo, mas a participação no programa seja aleatorizada entre os grupos de tratamento e controle, os impactos computados pela avaliação serão internamente válidos, porém não terão validade externa.

3.2. Custos da aleatorização

Durante as últimas décadas, o método experimental vem sendo cada vez mais utilizado em avaliações de impacto de programas sociais. Esse crescimento tem sido observado tanto em países ricos quanto pobres, e tem coberto intervenções nas áreas de educação, saúde, treinamento profissional, pobreza, entre outras.

Contudo, apesar desse crescimento, o método de aleatorização ainda está longe de ser o predominantemente utilizado em avaliações de impacto de programas sociais. Mas por que ele não é o mais utilizado, visto que é o único que garante com poucas hipóteses a validade interna das avaliações de impacto? A resposta a essa pergunta reside em três conjuntos de fatores: éticos, financeiros e técnicos.

Muitos argumentam que não é ético negar a participação em certo programa social com base num mecanismo de aleatorização. Afinal, dizem os que se opõem ao método, quando se trata de um programa social não é eticamente correto deixar pessoas fora do programa com base num simples “lançamento de moeda”. Sem dúvida, não se cogita utilizar aleatorização para casos de intervenções

de ampla escala (por exemplo, programas de vacinação), ou para públicos-alvo muito específicos (por exemplo, pessoas com algum tipo de deficiência). No entanto, há algumas situações em que esse problema ético não se coloca ou pelo menos pode ser contornado. A seguir, apresentamos as situações mais comuns e os correspondentes métodos de aleatorização mais empregados.

Excesso de demanda

Uma situação comum é quando existe escassez de vagas (normalmente devido à insuficiência de recursos financeiros, de pessoal ou operacionais) para atender toda a demanda ou público-alvo do programa. Em contextos como esse, que são muito comuns em países, estados e municípios pouco desenvolvidos, o próprio excesso de demanda pelo programa já requer que algum critério de alocação das vagas entre os demandantes seja utilizado. Tradicionalmente, o critério utilizado é o da ordem de chegada dos demandantes. Todavia, se pensarmos bem, o método de aleatorização pode ser um critério mais justo de alocação das vagas existentes. De fato, em situações de excesso de demanda, pode ser mais correto oferecer as vagas disponíveis garantindo a todos os elegíveis a mesma probabilidade de acesso a elas. Se, por exemplo, há escassez relativa de vagas e não se conhece bem onde os elegíveis moram, parece mais justo receber todas as inscrições e selecionar os participantes via um processo de aleatorização – por exemplo, retirando aleatoriamente os nomes de uma urna na presença dos inscritos – do que oferecer as vagas disponíveis para os primeiros que chegarem ao local de inscrição. A aleatorização das vagas, nesse sentido, é recomendada não somente para assegurar a validade interna da avaliação, mas também para garantir a igualdade de oportunidade de acesso aos serviços do programa. Ao final deste capítulo, são apresentadas duas avaliações de impacto que fizeram uso da aleatorização das vagas disponíveis devido à existência de excesso de demanda.

Ordem de entrada

Semelhante ao contexto de excesso de demanda, outra situação na qual é eticamente mais fácil utilizar o método de aleatorização é quando os recursos disponíveis para o programa são limitados e, apesar de se planejar atender toda a população-alvo ao final da implementação, só é possível incluir os elegíveis ao longo do tempo. Em casos como esse, ninguém será privado de participar do programa; o que se aleatoriza é somente a ordem com que os elegíveis entram no programa. Ao longo do processo gradual de entrada no programa, os que ainda não tiverem participado

servem como grupo de controle para os que já foram selecionados. Muitas vezes, a entrada gradual não é realizada no contexto do indivíduo ou da família, e sim em um contexto mais agregado, como escolas, comunidades ou municípios. Um exemplo bastante conhecido de aleatorização por ordem de entrada foi o programa mexicano de transferência condicional de renda, Progresa (atualmente chamado Oportunidades), que teve a primeira parcela dos municípios elegíveis selecionada de forma aleatória pelos gestores do programa.

Apesar de amplamente utilizada, a aleatorização por ordem de entrada embute algumas limitações. A primeira é que a comparação dos tratados com os controles pode ficar comprometida se esses últimos alterarem seu comportamento devido à expectativa de recebimento do programa no futuro. Por exemplo, no caso de um programa de transferência de renda, se as famílias das comunidades de controle souberem que receberão a transferência no futuro, é possível que muitas aumentem seus gastos antes de receberem a transferência. Outra limitação do método é que ele pode impossibilitar a identificação dos efeitos de mais longo prazo do programa. Isso se deve ao fato de que, em geral, a entrada do último grupo ocorre no máximo dois anos após a entrada do primeiro, limitando o horizonte de tempo de comparação entre os grupos já tratados e os de controle.

Intragrupos

Em alguns casos, não é possível, ou ético, aleatorizar nem as unidades que receberão o programa, nem a ordem de entrada nele. No entanto, dependendo do tipo de intervenção, pode ser viável aleatorizar subgrupos que receberão o programa em algumas unidades enquanto outros subgrupos o receberão em outras unidades. Esse foi o método de aleatorização utilizado por Banerjee et al. (2007) em que escolas de áreas pobres na Índia foram aleatorizadas para receber um tutor para ajudar as crianças com matemática e linguagem na 3^a série enquanto as demais escolas elegíveis foram aleatorizadas para receber um tutor na 4^a série. De forma semelhante, Glewee et al. (2009) avaliam um programa de concessão de livros didáticos de inglês e matemática no Quênia, no qual um grupo de escolas elegíveis foi aleatorizado para entrar no programa, e as crianças da 3^a à 7^a série dessas escolas receberam o livro de inglês, enquanto as da 3^a, 5^a e 7^a séries receberam o livro de matemática.

Encorajamento

Quando o acesso ao programa tem natureza universal, a questão ética pode ser contornada por meio do encorajamento aleatorizado de grupos de indivíduos elegíveis a participar efetivamente do

programa. Em outras palavras, o programa está disponível para todos, mas apenas certos grupos são aleatoriamente estimulados a participar do tratamento. Como esse método encoraja mas não garante que os indivíduos de fato entrem no programa, é de se esperar que a aleatorização afete somente a probabilidade de os indivíduos de participar. Na medida em que a decisão de participar efetivamente do programa pode depender de características individuais não observáveis relacionadas à variável de resultado, a aleatorização por encorajamento não permitirá identificar o efeito do programa como no caso em que há cumprimento total do que foi especificado pela aleatorização.³

Mas não é só por razões éticas que os experimentos sociais não são realizados com mais frequência. Outro motivo importante é que eles podem ser dispendiosos em relação ao orçamento global do programa. Tipicamente, a avaliação aleatorizada de um programa social requer a mobilização de vários recursos, tais como o tempo dos gestores do programa, a contratação de pessoal técnico e a realização de pesquisas de campo com as unidades tratadas e de controle. Este último componente requer o uso de questionários especialmente desenhados para a avaliação do programa em questão, que são geralmente aplicados em dois momentos do tempo: o primeiro, tipicamente no momento da inscrição no programa, e o segundo, alguns meses após o final da intervenção. Normalmente, é esse componente que requer o maior volume de recursos financeiros e, portanto, é o que merece mais atenção no planejamento da avaliação. Assim, apesar de todas as vantagens do método experimental, é importante contrapor os benefícios e custos financeiros de sua aplicação *vis-à-vis* a de outros métodos.

Há também razões de natureza técnica que dificultam o uso de experimentos aleatórios. Entre as mais importantes estão: o não comparecimento ao programa de uma parcela dos indivíduos do grupo de tratamento; a possibilidade de que pelo menos parte do grupo de controle acabe encontrando uma maneira de entrar no programa ou substituí-lo por um semelhante; e a perda de observações por atrito amostral, isto é, a perda das informações sobre unidades tratadas e de controle ao longo do tempo. Essas e outras dificuldades técnicas serão discutidas com mais detalhes na seção 3.4. Na seção a seguir, analisaremos a dimensão econométrica do método de aleatorização desconsiderando essas dificuldades técnicas.

3. Ver seção 3.4 para uma apresentação mais extensa desse ponto e o capítulo 6 para um tratamento detalhado dos problemas econométricos decorrentes desse tipo de cumprimento parcial da aleatorização.

3.3. Como a aleatorização resolve o problema de seleção

Tipicamente, estamos interessados em estimar o efeito médio do programa sobre os tratados (*EMPT*). Como vimos no capítulo 2 (seção 2.3), esse parâmetro é definido por:

$$(1) \quad EMPT = E[Y_i(1) - Y_i(0)|T_i=1] = E[Y_i(1)|T_i=1] - E[Y_i(0)|T_i=1],$$

onde a segunda igualdade decorre da propriedade de linearidade das médias (valores esperados). Esse parâmetro corresponde, portanto, à diferença entre a média da variável de resultado potencial para os tratados na situação de tratamento e a média do resultado potencial para os tratados, porém na situação contrafáutal de não tratamento.

Olhando para expressão (1), é possível perceber que a primeira média pode ser calculada diretamente das informações para a população de tratados. O grande problema é como obter a segunda média, ou seja, como encontrar a verdadeira média da variável de resultado caso os tratados não tivessem passado pela intervenção. É aqui que o método de aleatorização dá a sua maior contribuição, pois o grupo que foi aleatorizado para fora do programa é capaz de prover essa média. Especificamente, ao construir os grupos de tratamento e controle com base num mecanismo de aleatorização, o experimento social torna a variável que indica tratamento, T_i , independente dos resultados potenciais, $Y_i(0)$ e $Y_i(1)$. Essa independência se dá por força do próprio processo de aleatorização.⁴ Assim, enquanto um grupo qualquer de indivíduos não tratados fornece $E[Y_i(0)|T_i=0]$, um grupo de controle escolhido de forma aleatória assegura que $E[Y_i(0)|T_i=0] = E[Y_i(0)|T_i=1]$. Note que, ao garantir a validade desta última igualdade, o experimento aleatório consegue resolver o viés de seleção.

Podemos também utilizar equações lineares típicas dos modelos de regressão para entender como a aleatorização resolve o problema do viés de seleção. Assumindo que o efeito do programa é constante na população, a variável de resultado observada pode ser expressa por (ver seção 2.3 do capítulo 2):

4. Formalmente, temos: $Y_i(0), Y_i(1) \perp T_i$. Uma das consequências dessa independência é que $E[Y_i(0)|T_i] = E[Y_i(0)]$ e $E[Y_i(1)|T_i] = E[Y_i(1)]$, ou seja, as médias condicionais de ambos os resultados potenciais não dependem de o indivíduo ser aleatorizado para dentro ou para fora do programa. Assim, em experimentos aleatórios (com cumprimento total), o $EMPT = E[Y_i(1) - Y_i(0)|T_i=1]$ coincide com o $EMP = E[Y_i(1) - Y_i(0)]$.

$$(2) \quad Y_i = \alpha + \beta T_i + \varepsilon_i,$$

onde o termo ε_i é um componente não observável (pelo avaliador) que afeta Y_i .⁵ Nessa equação, o parâmetro β deve captar o efeito do programa, e o viés de seleção aparece pela potencial correlação entre T_i e ε_i . Naturalmente, essa correlação depende do tipo de programa e do contexto em que ele se insere. Por exemplo, no caso de um programa de treinamento, é provável que os indivíduos mais motivados tenham maior propensão de participar do programa.

A expressão para a média condicional da equação (2) é dada por:

$$(3) \quad E[Y_i|T_i] = \alpha + \beta T_i + E[\varepsilon_i|T_i].$$

A aleatorização garante que a participação no programa seja independente das características não observadas dos indivíduos. A geração dessa independência entre T_i e ε_i faz o método de aleatorização ser capaz de eliminar o viés de seleção. Em relação à equação (3), isso significa que a aleatorização faz com que:

$$(4) \quad E[\varepsilon_i|T_i] = 0,$$

ou seja, garante o atendimento de uma das hipóteses fundamentais (média condicional 0 do erro) para que modelo de Mínimo Quadrados Ordinário (MQO) gere um estimador não enviesado dos parâmetros da regressão. Devido ao resultado da expressão (4), podemos escrever que:

$$(5) \quad E[Y_i|T_i=1] - E[Y_i|T_i=0] = (\alpha + \beta) - (\alpha) = \beta,$$

o que comprova que o método de aleatorização isola corretamente o efeito de interesse. Note que a estimação do efeito pode ser feita com base numa simples regressão de MQO aplicada em (2) ou, como mostra a expressão (5), pelo simples cômputo da diferença de duas médias amostrais:

$$\bar{Y}_1 - \bar{Y}_0 = \frac{\sum_{i=1}^{N_1} Y_{1i}}{N_1} - \frac{\sum_{i=1}^{N_0} Y_{0i}}{N_0}, \text{ onde } Y_{1i} \text{ e } Y_{0i} \text{ representam a variável de resultado respectivamente}$$

5. Tal como no capítulo 2, não incluímos as características observadas dos indivíduos para facilitar o entendimento. Essas características serão introduzidas mais à frente no capítulo.

para os N_1 indivíduos tratados e N_0 indivíduos não tratados. Essa simplicidade de cálculo é uma das grandes vantagens do método de aleatorização.⁶

Embora não tenhamos incorporado as características observadas dos indivíduos na equação (2), é possível introduzi-las no modelo. Como a aleatorização torna a participação no programa, T_i , independente das características observadas (e não observadas) dos indivíduos, a inclusão do vetor de variáveis observadas X_i na equação (2) não deveria afetar, pelo menos em amostras grandes, a estimativa pontual do parâmetro β . Isso significa que, em princípio, a inclusão de X_i no modelo não enviesa (mas também não ajuda a estimar) o efeito causal de interesse. O principal benefício da introdução de X_i é que ela pode trazer ganhos de eficiência na estimativa do efeito de interesse, isto é, pode gerar estimativas mais precisas de β . Esse ganho de precisão, todavia, dependerá de as variáveis em X_i terem algum poder explicativo sobre a variável de resultado Y_i .⁷

Considere então a equação (2) modificada pela introdução do vetor de características observadas X_i :

$$(2') Y_i = X_i' \alpha + \beta T_i + \varepsilon_i$$

onde o primeiro elemento de X_i é igual a 1, e α é um vetor de parâmetros conforme X_i . Se bem realizado, o experimento aleatório assegura que:

$$(4') E[\varepsilon_i | X_i, T_i] = 0,$$

6. Com relação à inferência, o teste de o parâmetro β ser igual a 0 (o conhecido teste-t) é equivalente ao teste de diferenças de médias. Este último é baseado na estatística $d = \frac{\bar{Y}_1 - \bar{Y}_0}{\sigma_{\bar{Y}_1 - \bar{Y}_0}}$, onde $\sigma_{\bar{Y}_1 - \bar{Y}_0}$ representa o desvio-padrão da diferença das médias. Como as amostras dos grupos de tratamento e controle são independentes, no caso geral $\sigma_{\bar{Y}_1 - \bar{Y}_0} = \sqrt{\frac{\sigma_t^2}{N_1} + \frac{\sigma_o^2}{N_0}}$, em que σ_t^2 e σ_o^2 representam as variâncias de Y_{it} e Y_{ot} e N_t e N_o denotam o número de observações das amostras dos grupos de tratamento e controle, respectivamente.

7. Na realidade, somente as variáveis em X que têm poder preditivo sobre Y é que podem ajudar a estimar o parâmetro β com maior precisão; as que têm pouco ou nenhum poder explicativo tendem a reduzir a precisão. Isso abre espaço para que haja escolha das variáveis que entrarão no modelo após a aleatorização. Essa escolha, no entanto, é bastante questionada nos meios estatísticos, pois pode gerar vieses de pré-teste decorrentes da busca pela especificação (no nosso caso, quais as variáveis em X) que melhor se adeque a algum critério estabelecido pelo analista. Para evitar esse tipo de prática, recomenda-se que a escolha das características observadas que entrarão no modelo ocorra *antes* do experimento ser realizado.

o que implica que a diferença $E[Y_i|X_i, T_i=1] - E[Y_i|X_i, T_i=0] = (X_i'\alpha + \beta) - (X_i'\alpha) = \beta$ isola o parâmetro de interesse.

Com uma aleatorização bem realizada, as características observadas dos grupos de tratamento e controle devem ser balanceadas. Ou seja, se os dados contêm K diferentes características dos indivíduos (sexo, idade, escolaridade, renda familiar, entre outras), devemos esperar que a média de cada característica X_{ik} contida em $X_i = (X_{i1}, X_{i2}, \dots, X_{iK})$ para o grupo de tratamento seja igual à média da mesma variável para o grupo de controle. Formalmente, isso significa que $E[X_{ik}|T_i=1] = E[X_{ik}|T_i=0], \forall k \in \{1, \dots, K\}$, um resultado que pode ser testado com base nos dados coletados para os dois grupos experimentais.⁸ Na realidade, é sempre recomendável testar o balanceamento das características observadas entre os grupos, pois caso ocorra rejeição da hipótese de balanceamento para um conjunto significativo de variáveis em X , há indícios de que a aleatorização não foi bem realizada, o que pode comprometer a identificação do efeito do programa.

É comum que os programas sociais gerem efeitos distintos para diferentes grupos da população. Por exemplo, programas de treinamento de mão de obra tendem a ter impactos diferenciados entre mulheres e homens e entre jovens e adultos. Tipicamente, a incorporação desse tipo de heterogeneidade ao modelo é importante, já que pode revelar aos gestores para que grupos socioeconômicos o programa está gerando os efeitos desejados e para quais ele pode ser aperfeiçoado.⁹ Com um experimento bem desenhado, podemos também utilizar uma variante da equação (2) para computar os impactos heterogêneos do programa para subgrupos formados pelas características que aparecem no vetor X . Para operacionalizar essa estratégia, basta interagir as variáveis em X para as quais temos interesse em medir separadamente o impacto do programa com o indicador de tratamento T . Para verificar isso, suponha que se queira medir o efeito médio de uma

8. Há várias formas de se testar este resultado. Uma das mais empregadas é o teste para diferenças de médias de amostras independentes. Esse teste é bastante confiável quando a aleatorização dos grupos de tratamento e controle é feita com base no método de amostragem aleatória simples. Todavia, caso a aleatorização dos grupos seja realizada de forma estratificada (por exemplo, no nível de escolas, turmas, clínicas, bairros e municípios) e as amostras em cada estrato sejam pequenas, não necessariamente esse é o teste mais recomendado. Para mais detalhes, ver Firpo et al. (2011).

9. Para poder estimar adequadamente o impacto do programa para diferentes subgrupos num experimento aleatorizado, é tecnicamente recomendado que o desenho da amostra leve essa dimensão em consideração. Veja o Apêndice ao final do capítulo para uma apresentação sucinta de algumas das principais dimensões envolvidas no desenho da amostra de um experimento aleatorizado.

intervenção separadamente para homens e mulheres, os quais são identificados na base de dados pela variável binária $X_K=\{1,0\}$, com 1 para mulheres e 0 para homens. Denotemos por X^* todas as variáveis em X que não X_K . Se escrevermos:

$$Y_i = X_i' \alpha^* + X_{Ki} \alpha_K + \beta T_i + X_{Ki} T_i \gamma + \varepsilon_i,$$

o efeito do programa para as mulheres pode ser obtido por meio de:

$$\begin{aligned} E[Y_i | X_i^*, X_{Ki} = 1, T_i = 1] - E[Y_i | X_i^*, X_{Ki} = 1, T_i = 0] &= [X_i' \alpha^* + \alpha_K + \beta + \gamma] - \\ [X_i' \alpha^* + \alpha_K] &= \beta + \gamma, \end{aligned}$$

e para os homens via:

$$E[Y_i | X_i^*, X_{Ki} = 0, T_i = 1] - E[Y_i | X_i^*, X_{Ki} = 0, T_i = 0] = [X_i' \alpha^* + \beta] - [X_i' \alpha^*] = \beta.$$

Comparando as duas últimas expressões, vê-se que a potencial heterogeneidade de impacto entre os grupos de gênero é dada pelo parâmetro γ . A realização do teste t para γ permite inferir se ele é estatisticamente nulo. Da mesma forma que para X_K , outras variáveis presentes em X^* poderiam interagir com T para obtermos o impacto diferenciado do programa por essas características.

3.4. Dificuldades técnicas da aleatorização

Apesar de os experimentos sociais apresentarem importantes vantagens sobre outros métodos de avaliação de impacto, eles não estão livres de problemas. Na seção 3.2, abordamos os problemas de natureza ética e financeira que podem surgir para se implementar um experimento social. Nesta seção, discutiremos algumas dificuldades de natureza mais técnica que, caso não sejam adequadamente tratadas, podem comprometer bastante a acurácia dos resultados obtidos.

Quatro são as principais dificuldades técnicas que aparecem na realização dos experimentos aleatorizados. A primeira é o não comparecimento de uma parcela dos indivíduos selecionados para o grupo de tratamento para realizar o programa. A segunda é a substituição do programa

em avaliação por outros semelhantes por parte dos indivíduos do grupo de controle.¹⁰ A terceira dificuldade, chamada atrito amostral, aparece quando há coleta de dados em mais de um ponto do tempo e ocorre perda de observações de um conjunto de indivíduos da amostra entre períodos. A quarta complicação surge quando o próprio programa gera externalidades para os não tratados – incluindo os membros do grupo de controle –, o que pode distorcer os efeitos computados.

Não comparecimento

Um dos problemas mais comuns de um experimento social aleatório ocorre quando, após terem sido selecionados para participar do programa, alguns indivíduos do grupo de tratamento não aparecem de fato para realizar as atividades previstas na intervenção. O não comparecimento desse subgrupo altera o grupo de tratamento, o que pode comprometer a identificação dos impactos do programa, especialmente se a desistência estiver associada com características não observadas dos desistentes. Por exemplo, se os menos motivados entre os selecionados para o programa são os que não comparecem, a composição do grupo de tratamento deixa de ser parecida com a do grupo de controle nessa característica. Assim, se motivação for um determinante importante da variável de resultado (por exemplo, busca por emprego), a diferença de composição que aparece *a posteriori* entre os dois grupos irá comprometer o isolamento do impacto do programa.

Uma forma de olhar para esse problema é notar a distinção entre a oferta e a participação propriamente dita no programa. Essa distinção, que ocorre com vários programas sociais e não somente com experimentos aleatorizados, pode criar uma diferença entre o impacto que os gestores intencionavam que o programa tivesse para o público elegível em geral e o impacto que de fato ocorreu para o público efetivamente tratado. Na realidade, essa diferença de impactos tem sido amplamente discutida na literatura de avaliação e pode ser entendida dentro do arcabouço dos experimentos aleatórios.

Suponhamos que os gestores de um programa de treinamento tenham aleatorizado a oferta de vagas entre os indivíduos de uma comunidade. Seja Z_i uma variável binária que assume valor unitário

10. Tanto o não comparecimento quanto a substituição fazem parte do que se denomina cumprimento parcial da aleatorização. Este tema será retomado no capítulo 6, onde se discutirá mais formalmente como tratar as consequências do cumprimento parcial para a identificação do efeito do programa.

quando um indivíduo é aleatorizado para o grupo de tratamento e valor 0 quando a aleatorização é para o grupo de controle. Caso uma fração dos que foram alocados para o grupo de tratamento não aparecerem para participar do curso, haverá uma distinção entre a variável Z_i e a variável que identifica quem efetivamente compareceu, T_i . Assim, se a estimativa for realizada com base na diferença $E[Y_i|Z_i=1]-E[Y_i|Z_i=0]$, não estaremos identificando o efeito médio do programa (*EMPT* ou *EMP*), mas sim outro parâmetro, que recebe o nome de *efeito da intenção de tratar* (*EIT*).¹¹

A estimativa do *EIT* pode ser tão relevante para a política pública quanto para o efeito médio do programa. De fato, o *EIT* informa os gestores sobre como a disponibilidade do programa afeta em média a população de elegíveis. Isso é importante, já que, na prática, o máximo que os gestores normalmente conseguem fazer é oferecer o programa à população-alvo, não tendo como obrigar-a a frequentá-lo.

Mas o não recebimento dos serviços do programa por parte de uma parcela do grupo de tratamento não impede que o método de aleatorização seja capaz de identificar algum efeito do programa. Se olharmos com cuidado a situação, é possível perceber que a variável Z_i deve ser correlacionada com a variável T_i , porém não associada com qualquer determinante (observado ou não) da variável de resultado, Y_i . Isso permite utilizá-la como um instrumento no método de variáveis instrumentais, um tema que será tratado com detalhes no capítulo 6.

Substituição

O problema inverso também costuma ocorrer com os indivíduos que compõem o grupo de controle. Idealmente, num experimento social, nenhum membro desse grupo deveria receber os serviços do programa ou de um substituto próximo dele. Contudo, na prática, não é incomum que uma parcela do grupo acabe entrando no programa em avaliação ou em algum programa similar. Quanto mais substituível a intervenção sendo avaliada, maior a possibilidade de que os controles busquem um tratamento parecido em algum outro lugar.

Uma vez que a decisão do grupo de controle de participar do próprio programa ou de um similar possa estar relacionada a características não observáveis das pessoas (por exemplo, “excesso” de

11. Implicitamente, estamos assumindo que o efeito da intervenção é heterogêneo entre as pessoas. Naturalmente, na hipótese de que o efeito é igual para todos, não haveria diferença entre o *EIT* e o *EMPT* (ou *EMP*).

motivação ou percepção de que o ganho com o programa é elevado), a principal consequência desse comportamento é que o grupo de controle deixa de corresponder ao contrafactual desejado. Como decorrência, a média da variável de resultado para o grupo de controle, $E[Y_i|Z_i=0]$, deixa de fornecer a média que se buscava com o experimento aleatório, $E[Y_i(0)|T_i=1]$. Com isso, a simples diferença das médias calculadas para os grupos de tratamento e controle experimentais não isola mais o impacto do programa.¹²

Naturalmente, para se estimar o efeito causal da intervenção nessa situação de substituição do programa, há que se recorrer a estratégias de identificação alternativas. Como no caso do não comparecimento às atividades do programa por parte do grupo de tratamento, uma estratégia é fazer uso da variável Z_i como instrumento, quando há substituição ou entrada no próprio programa por parte dos controles. Essa estratégia de identificação será discutida no capítulo 6.

Atrito

Em quase todas as pesquisas que coletam dados de campo em mais de um momento do tempo, há perda de informações sobre alguns indivíduos que não são encontrados nos períodos subsequentes ao primeiro. Há diversos motivos que levam ao aparecimento desse atrito amostral, mas muitos deles são não aleatórios, ou seja, podem estar relacionados às características (não observadas) dos indivíduos ou ao impacto de interesse. Normalmente, quando o atrito não é aleatório, a estimação do efeito causal do programa por meio da comparação entre tratados e controles torna-se potencialmente enviesada.

A aparição desse tipo de viés é mais provável quando a taxa de atrito difere entre o grupo de tratamento e o grupo de controle. Na prática, essa é a situação mais comum, com a taxa do grupo de controle sendo tipicamente mais elevada que a do grupo de tratamento. Muitas vezes, a negação do programa aos controles gera reação de recusa por parte deles em fornecer informações aos entrevistadores do programa. Em alguns casos, os controles, quando se veem na situação de não encontrar um programa alternativo perto de onde moram, mudam-se para outras localidades, tornando-se mais difícil a coleta subsequente de informações.

12. Novamente, sob a hipótese de que o efeito do programa é homogêneo para todos, essa diferença identificará o impacto do programa sob algumas hipóteses.

Como já mencionado, é sempre boa prática comparar as características observadas dos grupos de tratamento e controle, antes e depois do tratamento. Essa comparação pode ser particularmente informativa quando a taxa de atrito é substancial, pois, se aparecerem diferenças marcantes entre os dois grupos nas características observadas, é provável que o atrito tenha alterado, também, a composição dos atributos não observados entre os grupos. Tipicamente, a resolução do problema do viés de atrito amostral requer o uso complementar de hipóteses comuns aos métodos não experimentais.¹³

Externalidades

Além das limitações acima apontadas, os experimentos aleatórios também podem estar associados a dois outros tipos de problemas. O primeiro é o efeito de transbordamento do programa sobre o grupo de controle, ou seja, uma externalidade (positiva ou negativa) que é gerada pelo próprio tratamento sobre a variável de resultado do grupo de controle. Por exemplo, se os grupos de tratamento e controle de um programa de combate à obesidade infantil são escolhidos aleatoriamente dentro das escolas, é provável que o tratamento acabe afetando indiretamente o grupo de controle, já que as crianças tratadas podem influenciar o comportamento alimentar das crianças do grupo de controle (dentro ou fora da escola). Em casos como esse, é mais recomendado realizar a aleatorização dos grupos experimentais entre escolas, preferencialmente guardando certa distância entre as selecionadas para cada grupo. Esse tipo de cuidado deve ser sempre tomado com intervenções relacionadas a saúde, educação, alimentos, entre outras, em que podem ocorrer externalidades devido ao contato físico ou à comunicação dos indivíduos tratados com os não tratados.¹⁴

13. Uma forma comumente utilizada é especificar a equação dos determinantes do atrito amostral e estimá-la conjuntamente com a equação da variável de resultado. Geralmente, esse procedimento é realizado utilizando-se variáveis instrumentais, um método que será apresentado no capítulo 6. O leitor mais interessado encontrará em Grasdal (2001) uma discussão sobre esse procedimento no contexto de avaliação de impacto. Duflo et al. (2008) provê uma apresentação do problema de atrito amostral no contexto de experimentos aleatórios.

14. Há algumas técnicas para averiguar a existência de efeitos de transbordamento para fora do grupo de tratados. Uma delas, utilizada num estudo sobre os efeitos de um programa de desparasitação em escolas do Quênia (Miguel e Kremer, 2004), faz uso de diferentes níveis de exposição ao tratamento para analisar as externalidades (positivas) do programa nas localidades próximas às escolas. Ver Duflo et al. (2008) para uma apresentação de outras técnicas para estudar a presença de externalidades de intervenções avaliadas pelo método de aleatorização.

Finalmente, o último problema que precisa ser levado em consideração é a ocorrência de efeitos de equilíbrio geral, isto é, efeitos que o programa pode ter sobre o funcionamento da economia do local onde ele ocorre. Esses efeitos são mais comuns quando a escala do programa é grande relativamente ao tamanho da economia. Por exemplo, o programa Bolsa Família atende uma proporção que pode chegar a quase 70% da população de vários municípios pequenos no Brasil. Em municipalidades como essas, as transferências de renda às famílias beneficiadas pelo programa pode gerar impacto significativo sobre a demanda por diversos bens e serviços dessas localidades, resultando em mudanças de preços e salários. Esse tipo de efeito agregado da intervenção tende a alterar o próprio estado do contrafactual de não tratamento, tipicamente concebido para uma situação em que não há mudanças na economia decorrentes do próprio programa.

Os efeitos de equilíbrio geral podem provocar uma reação dos agentes sobre suas decisões futuras em relação à variável de interesse, pelo menos em longo prazo. Por exemplo, um amplo programa de concessão de bolsas de estudos universitários pode ter efeito positivo sobre a entrada dos beneficiários na universidade. No entanto, devido a sua escala, o programa também tende a alterar o retorno à educação superior, o que pode afetar as decisões das pessoas de fazer curso superior no futuro.

A maior parte das avaliações de impacto não leva em conta esse tipo de efeitos indiretos do programa na economia e nas decisões dos agentes. Essa prática não é recomendada, especialmente nos casos de intervenções de larga escala. A forma típica de resolver esse problema é lançar mão de um modelo estrutural geral que incorpore de forma explícita o comportamento dos agentes. Após a estimativa dos parâmetros do modelo – inclusive o referente ao impacto do programa –, é possível simulá-lo para conhecer os efeitos diretos e indiretos da intervenção.

Box 3.1: Avaliação experimental do Planfor

Lançado em 1995 pelo governo federal, o Plano Nacional de Formação Profissional (Planfor) foi um programa de qualificação que ofereceu cursos de treinamento profissional para uma variedade de ocupações, tais como garçom, cabeleireiro, vendedor, eletricista, costureiro etc. Os principais objetivos do programa eram aumentar a empregabilidade e a capacidade de geração de renda dos trabalhadores da sua população-alvo, que era prioritariamente formada de desempregados e pessoas em situação de desvantagem social. A inscrição no programa era voluntária e os cursos costumavam durar entre 30 e 60 dias, em geral com uma carga horária de 60 horas por mês. Os serviços de treinamento eram providos por escolas especializadas, universidades, sindicatos e organizações não governamentais.

Em 1998, o Planfor passou por uma avaliação de impacto realizada com base no método de aleatorização. Especificamente, foram avaliados os cursos oferecidos no mês de setembro de 1998 nas regiões metropolitanas do Rio de Janeiro e de Fortaleza. Na inscrição para os cursos, os indivíduos responderam a um questionário – normalmente chamado de linha de base

– com uma série de perguntas sobre suas condições socioeconômicas e de trabalho. Como houve excesso de demanda, os grupos de tratamento e controle foram formados pela aleatorização dos inscritos nos diversos cursos oferecidos em cada região. Em novembro de 1999, um segundo questionário foi aplicado aos indivíduos de ambos os grupos – normalmente chamado de acompanhamento (ou *follow-up*) – que puderam ser encontrados cerca de um ano depois do início dos cursos. Novamente, perguntas sobre a situação de trabalho foram levantadas no segundo questionário. Apesar da ocorrência de atrito (isto é, perda de observações entre a linha de base e o *follow-up*), as amostras finais permaneceram grandes, com a do Rio de Janeiro contendo 2.256 indivíduos (1.201 tratados) e a de Fortaleza, 2.517 (1.210 tratados).

A Tabela 3.1 contém as estimativas do impacto do Planfor separadamente para cada região metropolitana (Ipea e Cedeplar, 2000). As variáveis de resultado são a proporção de ocupados e a taxa de desemprego. As estimativas para cada grupo experimental estão apresentadas em termos das diferenças nessas variáveis entre os períodos anterior e posterior aos

cursos. O impacto é calculado simplesmente pela subtração dessas diferenças entre os dois grupos, ou seja, pelo procedimento conhecido por diferenças em diferenças, a ser tratado em detalhes no capítulo 4.

Os resultados mostram que o programa possa ter um impacto positivo em termos de ocupação em ambas as regiões. Especificamente, vemos que a proporção de ocupados aumentou mais entre os tratados do que entre os controles tanto no Rio

de Janeiro quanto em Fortaleza, gerando impacto positivo de cerca de 2 pontos de percentagem (p.p.) na primeira região e de aproximadamente 3 p.p. na segunda. No caso da taxa de desemprego, as estimativas indicam que o programa só foi capaz de reduzir essa taxa no Rio de Janeiro. De fato, em Fortaleza, embora a estimativa pontual do impacto seja na direção de aumento da taxa de desemprego, do ponto de vista estatístico o efeito pode ser considerado nulo, pelo menos, com significância de 10%.

Tabela 3.1: Estimativas do Impacto do Planfor para Rio de Janeiro e Fortaleza

	Rio de Janeiro			Fortaleza		
	Tratamento	Controle	Impacto	Tratamento	Controle	Impacto
Ocupação	0,1064 (0,0098)	0,0876 (0,0099)	0,0188 [0,1000]	0,0914 (0,3990)	0,0607 (0,4224)	0,0307 [0,0621]
	0,0924 (0,0127)	0,1174 (0,0128)	-0,0250 [0,1000]	0,0859 (0,4134)	0,0174 (0,4409)	0,0685 [0,1285]
Desemprego						

Fonte: Ipea e Cedeplar (2000).

Nota: os números entre parênteses correspondem aos erros-padrão calculados para as amostras de cada grupo experimental e região metropolitana. Os número entre colchetes referem-se aos p-valores reportados para as estimativas do impacto do programa.

Box 3.2: Impacto das creches públicas no Rio de Janeiro sobre a participação das mulheres no mercado de trabalho

Na cidade do Rio de Janeiro, existe um conjunto amplo de creches que recebem algum financiamento do governo para funcionar. Espera-se que as creches tenham efeitos positivos tanto no desenvolvimento das crianças quanto na inserção das suas mães no mercado de trabalho. As evidências mostram que um desenvolvimento adequado na primeira infância tem efeitos cognitivos e não cognitivos positivos, com repercussões de longo prazo para as pessoas na fase juvenil e adulta. A liberação de tempo que as creches propiciam para maior engajamento das mães com o mercado de trabalho também é considerada importante, especialmente para as famílias de baixa renda.

Em 2007, utilizou-se um mecanismo de loteria na cidade do Rio de Janeiro para selecionar entre as 25 mil famílias candidatas às 10 mil vagas a serem oferecidas nas creches municipais com financiamento público no ano de 2008. Explorando essa aleatorização, uma pesquisa foi realizada com cerca de 3.800 famílias entre junho e outubro de 2008 para levantar uma série de informações sobre a situação de trabalho das mulheres. Metade dessas famílias tinha sido selecionada aleatoriamente para fazer parte do grupo de

tratamento, e a outra metade, para o grupo de controle. Na prática, no entanto, ocorreram os problemas de não comparecimento por parte das famílias do grupo de tratamento e de substituição do programa pelas famílias do grupo de controle. Especificamente, 94% das crianças do grupo de tratamento entraram de fato nas creches cobertas pelo programa, ao passo que 51% das crianças do grupo de controle acabaram conseguindo vagas nesse mesmo conjunto de creches.

Como mencionado no texto, a existência desses problemas de cumprimento parcial requer o uso de métodos complementares ao proposto pelo método de aleatorização. Em Barros et al. (2011), empregou-se o procedimento de variáveis instrumentais – a ser discutido no capítulo 6 – para se estimar um parâmetro conhecido como efeito médio local do tratamento (*EMLT*). Sendo Y_i a variável de resultado, Z_i uma variável (instrumento) binária que indica se a mulher foi aleatorizada para o grupo de tratamento ou controle (1 ou 0, respectivamente) e T_i uma variável também binária que marca quais mulheres efetivamente tiveram ou não seus filhos matriculados nas creches (1 ou 0, respectivamente), o estudo citado estimou o seguinte parâmetro:

$$EMLT = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{P[T_i = 1|Z_i = 1] - P[T_i = 1|Z_i = 0]}$$

Em linhas gerais, esse parâmetro pode ser interpretado como efeito do programa que seria estimado pelo método de aleatorização (isto é, o efeito da intenção de tratar, *EIT*, apresentado na seção 3.4) corrigido para o fato de que nem todos os tratados e controles “cumpriram” o que a seleção aleatória determinou. Note que essa correção é dada pela diferença entre os 94% das mulheres do grupo de tratamento que tiveram seus filhos efetivamente tratados (isto é, o termo $P[T_i=1|Z_i=1]$ do denominador) e os 51% das mulheres do grupo de controle

que descumpriam o determinado pela loteria e conseguiram colocar seus filhos nas creches do programa (o termo $P[T_i=1|Z_i=0]$ do denominador).

As estimativas do impacto do programa indicam aumentos relativos de 8% na taxa de participação das mulheres no mercado de trabalho (de 74% para 79%) e de 27% na proporção de ocupadas (de 36% para 46%). A intervenção parece também ter reduzido a taxa de desemprego das mulheres em 16% (de 50% para 42%).

3.5. Apêndice: tópicos sobre o desenho de experimentos aleatorizados

A realização de um experimento aleatorizado requer o planejamento cuidadoso das suas etapas. Entre os vários aspectos a serem planejados, a construção da amostra dos grupos de tratamento e controle é particularmente importante, já que a capacidade do experimento de identificar e estimar o efeito do programa corretamente depende da qualidade com que a amostra foi construída.

Uma das dimensões mais importantes na construção da amostra é a definição do seu tamanho. Como vimos no texto, a estimativa do efeito do programa em experimentos aleatorizados corresponde à diferença das médias amostrais da variável de interesse entre os grupos de tratamento e controle. A precisão com que essas médias – e, portanto, a diferença entre elas – serão estimadas aumenta com o tamanho das amostras de cada um dos grupos. Se não houvesse custos para colher informações, o tamanho das amostras poderia ser muito amplo. No entanto, como normalmente é custoso levantar informações, é necessário planejar o tamanho da amostra com base em algum método que permita estimar o efeito do programa com o nível de precisão desejado.

O método mais amplamente empregado no planejamento do tamanho da amostra de um experimento é denominado *cálculo do poder* (estatístico) que a estimativa a ser obtida terá para detectar o verdadeiro efeito do programa.¹⁵ Num contexto de experimentos aleatorizados, o poder estatístico corresponde à probabilidade de detectar que a estimativa do efeito do programa é diferente de 0 quando de fato ela é não nula.¹⁶ Para se evitar concluir que o programa não teve efeito quando ele realmente teve, as médias dos grupos de tratamento e controle devem ser estimadas de forma precisa. Assim, o poder do experimento será tão mais elevado quanto maior

15. Como esse tema é bastante amplo e técnico, o material a seguir apresenta de forma sucinta os principais aspectos envolvidos no cálculo do poder de um experimento. Um exposição mais detalhada pode ser encontrada em Bloom (1995) e Duflo et al. (2008).

16. Em termos estatísticos, o poder do experimento é inversamente relacionado ao risco de cometer um erro do tipo II (especificamente: 1 – probabilidade do erro do tipo II). O erro do tipo II ocorre quando se conclui a partir de um teste de hipótese convencional que a estimativa não é diferente de 0 quando de fato ela é. Distintamente, o erro do tipo I ocorre quando se conclui que a estimativa é diferente de 0 quando de fato ela não é.

for o tamanho da amostra (mantendo-se tudo o mais constante). Experimentos com baixo poder estatístico correm o risco de concluir equivocadamente que o programa não está gerando efeitos, quando ele de fato está.

Para se realizar o cálculo de poder, é preciso levar em conta uma série de fatores. Em primeiro lugar, é preciso saber se o experimento será realizado em âmbito pessoal ou mais agregado (*clusters*), tais como escolas, clínicas ou comunidades.¹⁷ No primeiro caso, a amostra pode ser construída por meio de amostragem simples de toda a população-alvo do programa. Já no segundo, a precisão da estimativa do efeito será diretamente relacionada com o número de *clusters* a serem aleatorizados. Em amostragem com *clusters* (por exemplo, comunidades) é preciso levar em consideração a existência de correlação na variável de resultado entre as unidades que compõem os *clusters* (por exemplo, as famílias das comunidades). Se a correlação for alta, as unidades dentro dos *clusters* tenderão a ser homogêneas, o que implica que amostrar mais indivíduos dentro dos *clusters* não deve aumentar muito a precisão das estimativas das médias amostrais. Isso indica que, em geral, quanto maior a correlação *intraccluster* na variável de resultado, maior a necessidade de amostrar um número maior de *clusters* (em vez de indivíduos dentro dos *clusters*) para manter o poder estatístico do experimento.¹⁸

O segundo fator que deve ser levado em conta é o interesse em estimar o impacto do programa para subgrupos da população (por exemplo, mulheres e homens, jovens e adultos, diferentes grupos de renda). Embora o interesse em estimar o efeito do programa para o conjunto de subgrupos permaneça, cada subgrupo pode ser pensado como um “miniexperimento” isolado e, nesse sentido, o tamanho da amostra deve ser elevado o suficiente para permitir detectar com precisão o efeito do programa para cada subgrupo (e para todos eles conjuntamente). Esse tipo de situação tem correspondência com o que se denomina amostragem estratificada, onde os estratos (também chamados blocos) são formados pela combinação das características (por exemplo, sexo,

17. Razões de ordem política, ética ou operacional podem impedir que a aleatorização seja realizada no âmbito individual. Nesses casos, a aleatorização acaba ocorrendo para *clusters* de pessoas.

18. Em muitos experimentos aleatorizados, utiliza-se um método de aleatorização em que pares de *clusters* com características próximas são pareados, e um deles é sorteado para o grupo de tratamento, e o outro, para o grupo de controle. Esse procedimento gera uma série de benefícios estatísticos para a estimativa do efeito do programa. Os interessados podem encontrar os detalhes técnicos desse método em Imai et al. (2009).

faixa etária e grupos de renda) da população de interesse. Uma das vantagens da estratificação é que os tratados e controles em cada bloco são, por construção, semelhantes nas suas características observáveis (e não observáveis). Outra vantagem é que a estratificação permite que a precisão da estimativa do efeito global do programa seja mais elevada do que se o método de amostragem fosse simples.¹⁹

O terceiro fator a ser levado em consideração no cálculo do poder é o tamanho do efeito que se espera com o programa. Se a expectativa é que a intervenção tenha impacto pequeno, deve-se aumentar o tamanho da amostra de forma a permitir testar com precisão se a diferença de médias entre tratados e controles será capaz de detectar o (pequeno) efeito esperado. Inversamente, quando o efeito esperado é alto, a amostra pode ser menor. Todavia, não é recomendável estipular um efeito mínimo esperado muito elevado, já que isso compromete a capacidade do experimento de detectar efeitos menores que o mínimo estipulado.

Outro fator importante é a dispersão da variável de resultado para a população-alvo do programa. Quanto mais heterogênea for essa variável na população, mais difícil é estimar sua média com precisão, exigindo então amostras maiores. Em geral, a dispersão é medida por meio de informações de outras pesquisas que contêm informações sobre a população-alvo do programa. Por exemplo, se a variável de interesse é a renda das famílias, uma pesquisa domiciliar realizada recentemente na região onde o programa ocorrerá pode fornecer uma estimativa razoável da variância da renda familiar.

O quarto fator para determinar o tamanho da amostra é determinar o nível de poder desejado para o experimento, de tal forma que a probabilidade de conclusões equivocadas sobre o “não efeito” do programa seja baixa. Como mencionado anteriormente, quanto maior o nível de poder estatístico, menor o risco de concluir que o programa não teve efeito quando ele de fato teve impacto. Mas estipular um nível de poder elevado exige maior precisão das estimativas e, portanto, amostras (e custos) maiores. Em geral, utiliza-se um nível de poder de 80%, que significa que, se fosse possível replicar o experimento inúmeras vezes, em 80% dos casos ele seria capaz de detectar que houve impacto quando de fato ele ocorreu.

19. Vale observar que, em geral, a estratificação nas características é superior em termos de precisão do que incluir *ex post* essas características na regressão. Ver Imbens et al.(2009) para uma análise dos benefícios estatísticos da estratificação em experimentos aleatorizados.

É preciso, também, determinar qual a proporção de tratados e controles na amostra. Se o custo do programa (incluindo o do levantamento das informações) por unidade tratada e de controle for similar, recomenda-se que a proporção seja a mesma entre os grupos (isto é, 50%). No entanto, se os custos unitários diferirem consideravelmente (por exemplo, porque o custo da intervenção propriamente dita é elevado), o tamanho da amostra dos controles deve ser maior que a dos tratados.

A possibilidade de que as unidades dos grupos de tratamento e controle não cumpram o que foi determinado pela aleatorização também deve ser prevista no cálculo do tamanho da amostra. Quando há cumprimento parcial da aleatorização, o experimento identifica o impacto do programa apenas para um subgrupo da população (denominado *compliers*, ver capítulo 6). Esses casos são muito comuns e, para manter o mesmo poder estatístico do experimento, há que se ajustar o tamanho da amostra para cima.

3.6. Exercícios:

1. Um governo estadual está atualmente planejando realizar programa de redução do analfabetismo e quer que uma avaliação seja realizada para medir os impactos do programa. Assinale se as assertivas abaixo são verdadeiras ou falsas e explique sua resposta.
 - a. Se esse governo estadual utilizar apenas os dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) – que é uma pesquisa com amostra probabilística realizada pelo IBGE em todo o território nacional – para avaliar o programa, os resultados assegurarão a validade interna do estudo.
 - b. A validade externa da avaliação estará garantida se houver a aleatorização dos serviços do programa (por exemplo, entre os municípios do estado).
2. Uma entidade que recebe verba pública presta serviços de treinamento de mão de obra para trabalhadores de baixa renda em certo município. Para selecionar os beneficiados pelo programa, a entidade utiliza o critério do “primeiro a chegar” ao local da inscrição. Um dia, um gestor decidiu tabular as informações das fichas de inscrição para o programa nessa entidade e descobriu que cerca de 80% dos beneficiados moravam no mesmo bairro do local da inscrição. O gestor

também observou que as características socioeconômicas dos inscritos não diferiam muito entre si. Baseado apenas na ideia de igualdade de oportunidade para participar do programa, você proporia alguma mudança de critério de seleção dos beneficiados? Pensando agora somente em termos de avaliação de impacto do programa, você proporia utilizar um critério baseado na seleção aleatória dos inscritos? Desenvolva suas respostas.

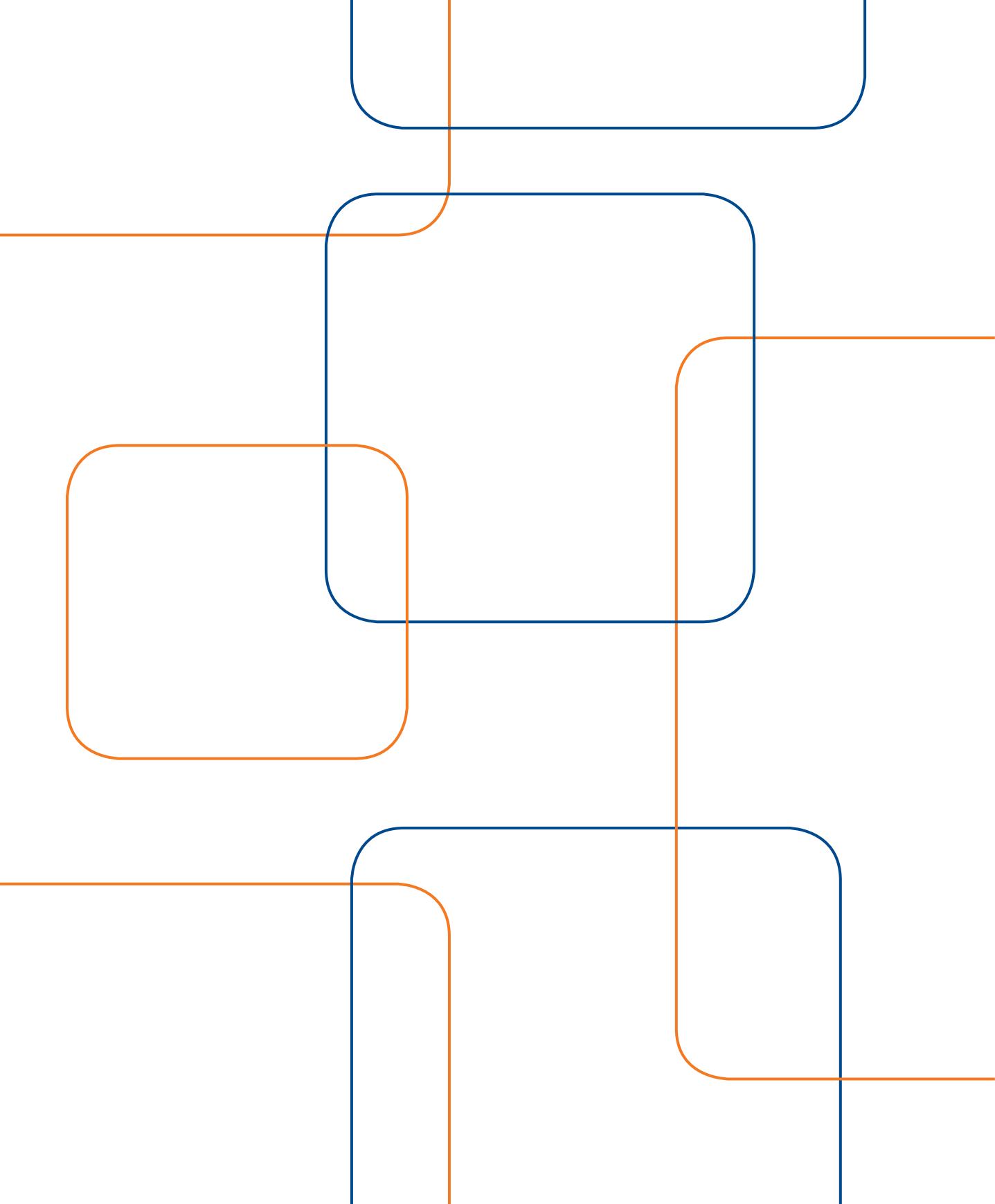
3. Em laboratórios de física, biologia e áreas científicas afins, é muito comum a realização de experimentos controlados, ou seja, experimentos nos quais o pesquisador tem ampla capacidade de controlar as variáveis que afetam o objeto de estudo. Nas ciências sociais, entretanto, não é tão fácil realizar experimentos desse tipo. Discuta pelo menos duas razões (técnicas ou não) que dificultam a realização de experimentos aleatórios na área de avaliação de programas sociais.
4. Em uma avaliação de impacto aleatorizada de um programa de reforço escolar, descobriu-se que 10% dos alunos do grupo de tratamento desistiram do programa e 10% dos indivíduos do grupo de controle encontraram uma forma de entrar no programa. Ao analisar as informações disponíveis, os avaliadores observaram que os alunos do grupo de tratamento que não compareceram tinham históricos escolares piores e que os do grupo de controle que acabaram entrando tinham históricos melhores.
 - a. Você trataria essa informação como evidência de que há características não observadas dos alunos que influenciaram a participação no programa?
 - b. As diferenças entre os grupos que foram aleatorizados para participar e não participar e os que efetivamente participaram impedem o uso do experimento para estimar o impacto do programa? Qual o efeito causal que poderia ser estimado, usando-se apenas a informação de quem foi aleatorizado para dentro e para fora do programa? Formalize sua resposta.
5. A equação de regressão abaixo apresenta as estimativas obtidas com base nas informações de um experimento aleatorizado de um programa de treinamento. A variável de resultado é o rendimento do trabalho, e os números entre parênteses são os erros-padrão estimados dos coeficientes da regressão. Assuma que houve cumprimento total do que foi determinado pela aleatorização.

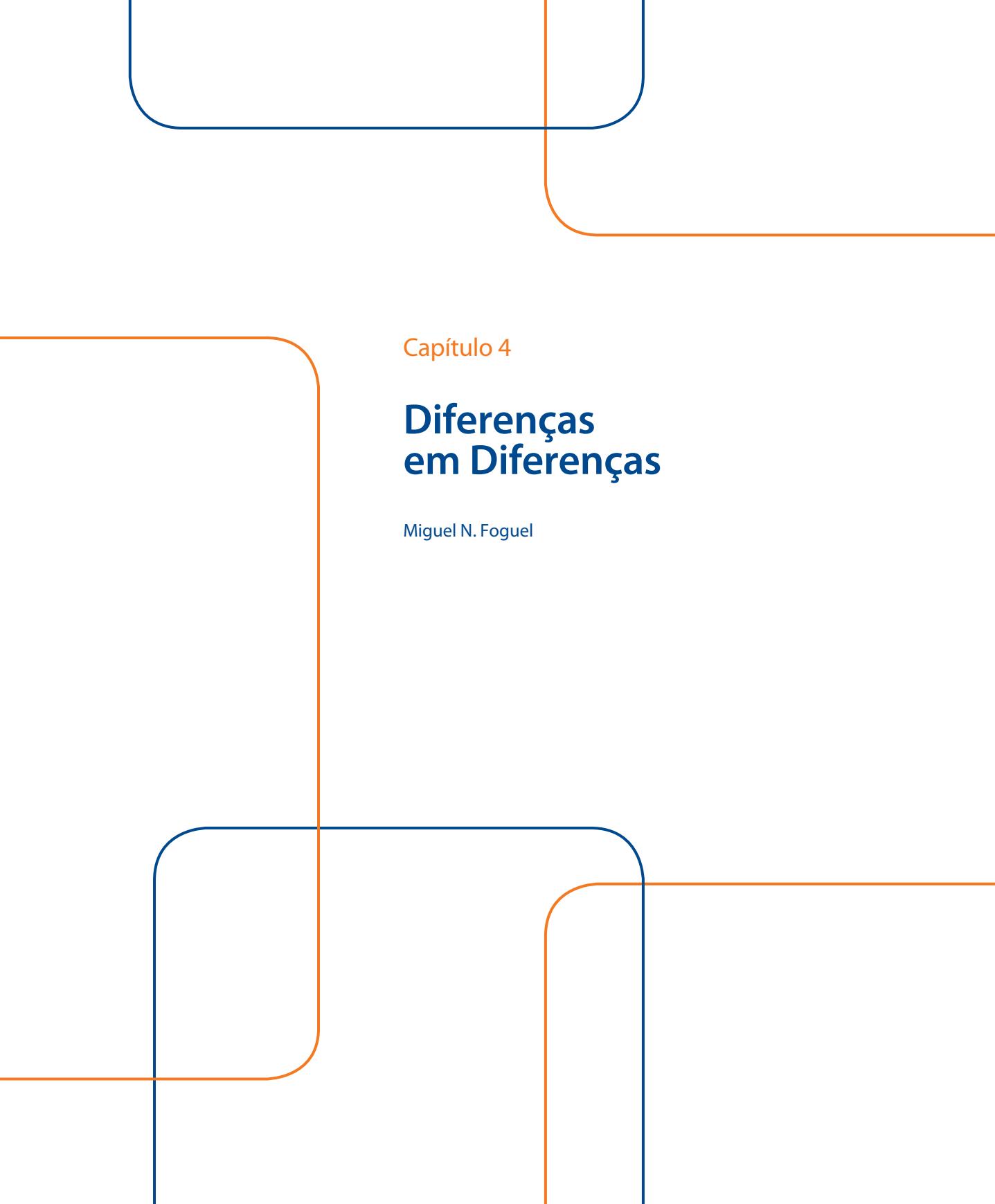
$$Y_i = 912,23 + 91,15 T_i$$
$$(115,54) \quad (12,25)$$

- a.** Qual o rendimento do trabalho dos não tratados?
- b.** Qual a magnitude do impacto do programa? Ele é significativo do ponto de vista estatístico?
- c.** Como você calcularia o impacto do programa em termos relativos?
- d.** Mostre como o impacto poderia ser estimado sem usar a regressão.
- e.** Mostre como sua resposta do item anterior se relaciona com os parâmetros da regressão.
- 6.** Dois estudos de avaliação de um programa de transferência de renda estimaram o efeito do programa sobre o consumo de alimentos. Uma delas utilizou dados de uma pesquisa domiciliar já existente (avaliação A) e a outra empregou dados de um experimento aleatorizado (avaliação B). A primeira avaliação estimou um impacto (em R\$) de $\hat{\beta}_A = 90$ e a segunda de $\hat{\beta}_A = 118$. Supondo que os dados se referem ao mesmo período e que não tenha havido cumprimento parcial da aleatorização, calcule a magnitude absoluta e relativa do viés de seleção.
- 7.** Os dados da tabela abaixo mostram estimativas para algumas variáveis referentes aos grupos de tratamento e controle de experimento aleatorizado. Além das médias de cada grupo, a tabela apresenta a diferença dessas médias, o seu erro-padrão e o P-valor associado ao teste de hipótese de as médias serem iguais. A variável de resultado é o salário, e as demais variáveis correspondem às características observadas dos grupos.

	Média de Tratados	Média de Controles	Diferença	Erro-padrão	P-valor
Salário	1.049,99	1.000,01	49,98	0,0201	0,0000
Mulher	0,315	0,308	0,007	0,0093	0,4530
Jovem	0,397	0,502	-0,104	0,0099	0,0000
Escolaridade	12,4	11,5	0,9	0,0573	0,0000

- a.** Qual a magnitude do impacto do programa? Ele é estatisticamente igual a 0?
- b.** As características observáveis são todas iguais do ponto de vista estatístico entre os grupos de tratamento e controle?
- c.** Com base na sua resposta ao item anterior, você diria que o efeito do programa está bem identificado? Explique.





Capítulo 4

Diferenças em Diferenças

Miguel N. Foguel

Como vimos nos capítulos anteriores, os métodos de avaliação de impacto são geralmente divididos em dois grupos: experimental e não experimental. O primeiro, apresentado no capítulo anterior, é baseado na seleção aleatória dos participantes e não participantes do programa e é considerado o método de referência na área de avaliação. O segundo grupo é composto por uma variedade de métodos distintos que, embora façam uso de hipóteses específicas para identificar o efeito causal do programa, procuram mimetizar o método experimental. Os próximos capítulos serão dedicados aos principais métodos não experimentais utilizados na área de avaliação de impacto. Neste capítulo, discutiremos um dos mais empregados dentre eles, o chamado método das diferenças em diferenças (**DD**).

Como sugere o seu nome, o método de **DD** é baseado no cálculo de uma dupla subtração: a primeira se refere à diferença das médias da variável de resultado entre os períodos anterior e posterior ao programa para o grupo de tratamento e para o de controle, e a segunda se refere à diferença da primeira diferença calculada entre esses dois grupos. Naturalmente, a operacionalização do método requer a existência de informações para ambos os grupos para, pelo menos, um período de tempo antes e um período depois do programa. Embora essa necessidade de dados torne a aplicação do procedimento mais demandante, o método possui uma série de vantagens em comparação a outros métodos não experimentais. Em particular, como veremos, ele é capaz de lidar com o viés de seleção associado a certo tipo de características não observáveis dos indivíduos, especificamente àquelas que são invariantes no tempo.

A aplicabilidade do método de **DD** é ampla, e ele tem sido utilizado isoladamente ou em combinação com outros métodos em diversas situações.¹ Muitas vezes, ele é empregado no contexto denominado “experimento natural” (ou “quase-experimento”), uma situação na qual a ocorrência de um evento fortuito – ou em larga medida imprevisto – permite formar grupos de tratamento e controle parecidos em diversos aspectos. Esses eventos podem ter origem em mudanças que ocorrem na própria natureza (por exemplo, terremotos e furacões), alterações institucionais (modificações inesperadas em leis ou programas em algum município ou estado) ou até mesmo na seleção dos participantes, que ocorre devido a circunstâncias que envolvem

1. Por exemplo, no capítulo 5, apresentaremos como o método de pareamento é combinado com o de **DD**.

um alto grau de aleatoriedade.² Qualquer que seja o caso, a ideia é que a fortuidade do evento permita criar uma seleção dos indivíduos ou unidades de observação que comporão os grupos de tratamento e controle próxima daquela do método experimental.³ No entanto, mesmo que o evento gerador pareça fortuito, não há garantias de que o grupo de controle seja uma boa representação contrafactual do grupo tratado. O método de **DD** procura resolver isso levando em consideração as diferenças de características preexistentes entre tratados e controles. Essa característica torna o método de **DD** empregável não só em casos de experimentos naturais, mas em diversos outros contextos onde haja disponibilidade de informações pré e pós-programa para os dois grupos.

A principal hipótese do método de **DD** é que a trajetória temporal da variável de resultado para o grupo de controle represente o que ocorreria com o grupo tratado caso não houvesse a intervenção. Naturalmente, essa hipótese não pode ser testada diretamente nos dados, mas uma indicação de sua validade aparece quando as trajetórias dos dois grupos são parecidas com pré-programa. A ideia básica é que, se as trajetórias se assemelham durante o período antes do programa, então parece razoável supor que a evolução do grupo de controle após o programa represente com fidedignidade o que ocorreria com o grupo de tratados na situação de não tratamento. Assim, quando existem dados para vários períodos de tempo pré-programa, tipicamente testa-se se as séries da variável de resultado, para os dois grupos, possuem a mesma tendência temporal. Embora não seja um teste cabal da hipótese, ele é bastante intuitivo e pode ser realizado com facilidade.

O procedimento de **DD** pode ser utilizado com dados de indivíduos/famílias, mas também com dados em níveis mais agregados, por exemplo, setores de atividade, ocupações, municípios ou estados. Por exemplo, se certa UF implementou um programa de educação ou saúde

2. Um exemplo interessante dessa última situação é o estudo de Braido et al. (2012), que avaliou os efeitos do programa brasileiro Bolsa Alimentação sobre o padrão de consumo dos domicílios. Por conta de uma incompatibilidade temporária dos softwares utilizados nos sistemas de recrutamento e seleção de beneficiários do programa, um conjunto de domicílios acabou não sendo selecionado devido à presença de letras nos nomes de membros dos domicílios que tivessem acentos, til ou cedilha. Dada a abrangência com que essas letras aparecem nos nomes das pessoas no Brasil, esse equívoco, que foi não intencional, constitui um mecanismo fortuito de seleção ao programa. Esse experimento “natural” permitiu aos autores construir grupos de tratamento e controle muito próximos aos de experimentos aleatórios.

3. Vale observar que o método de **DD** também é utilizado com experimentos puramente aleatórios, sempre que levantam informações para os períodos pré e pós-intervenção.

num determinado ano, é possível empregar o método para avaliar o impacto desse programa comparando os dados agregados (por exemplo, frequência escolar ou incidência de uma doença) da UF tratada com os dados de um conjunto de UFs não tratadas. Sua aplicabilidade com dados agregados é uma importante vantagem do método de **DD**, pois permite que o impacto de uma série de intervenções possam ser avaliadas apenas com base em informações rotineiramente coletadas por governos e institutos de pesquisa e estatística.

Outra importante vantagem do método de **DD** é que ele permite controlar para características não observáveis dos indivíduos que sejam invariantes no tempo (por exemplo, habilidades inatas). Como vimos em capítulos anteriores, os atributos não observados dos indivíduos são fonte potencial de geração de viés de seleção. Embora o método de **DD** só consiga levar em conta esse tipo de atributo que é fixo no tempo, esse controle é capaz de resolver o problema de seleção para um conjunto amplo de aplicações. Essa capacidade confere ao **DD** uma vantagem relativa a outros métodos não experimentais, por exemplo, o método do emparelhamento, que não consegue controlar para vieses de seleção decorrentes de atributos não observáveis, fixos ou não no tempo.

4.1. Comparando com os métodos “ingênuos”

Em uma avaliação de impacto sempre se busca encontrar o contrafactual para o grupo tratado, ou seja, o que teria ocorrido a esse grupo na ausência da intervenção. No capítulo 2 (seções 2.1 e 2.2), vimos dois métodos “ingênuos” de identificar o impacto do programa, a saber: (i) quando apenas se compara o grupo de tratados antes e depois da intervenção e (ii) quando se compara o grupo de tratados com um grupo qualquer de não tratados após a intervenção. A primeira forma de identificação não possui grupo de controle explícito e só consegue isolar o impacto do programa sob a hipótese de que o único fator que afeta a variável de resultado do grupo de tratados é a própria intervenção. Obviamente, trata-se de uma hipótese forte, já que sempre pode haver um conjunto de fatores que podem influenciar a trajetória dessa variável além da própria intervenção. A segunda forma de identificação utiliza um grupo de controle explicitamente, mas só consegue isolar o efeito do programa sob a hipótese de que esse grupo representa corretamente o que teria ocorrido com o grupo tratado na ausência do programa. Novamente, trata-se de uma suposição forte, uma vez que parte da diferença na variável de resultado entre os dois grupos pode se dever

à existência de diferentes características observáveis e não observáveis (ou medidas pelo avaliador) entre os grupos.⁴

O método de diferenças em diferenças (*DD*) busca contornar os problemas dessas duas formas de identificar o impacto de um programa. Em relação à primeira, ao invés de simplesmente contrastar o grupo de tratamento com ele mesmo antes e depois do programa, o método de *DD* vai buscar um grupo de comparação que se pareça ao máximo com o grupo tratado. A ideia é que esse grupo esteja sujeito às mesmas influências dos fatores que afetam a variável de resultado dos tratados, ou seja, que esse grupo funcione como um grupo de controle semelhante ao que se constrói a partir de experimentos aleatórios. Obviamente, sem aleatorização, não há garantias de que a única diferença entre os grupos seja a participação no programa, mas o método de *DD*, assim como os demais métodos não experimentais, busca mimetizar a situação de aleatorização. Uma característica do método de *DD* que ajuda nesse processo é que ele é capaz de levar em conta uma das principais fontes de diferença entre tratados e controles: as características não observáveis dos indivíduos que são invariantes no tempo. Nesse sentido, ele é um avanço em relação ao método que simplesmente contrasta o grupo de tratamento com o grupo de indivíduos não tratados.

4.2. Hipóteses básicas

Para identificar o efeito causal de uma intervenção, o método de *DD* lança mão de um conjunto de hipóteses. Idealmente, o método quer encontrar um grupo de indivíduos (ou unidades mais agregadas de observação, como municípios ou estados) para os quais a evolução da variável de resultado corresponda à trajetória dessa variável para o grupo tratado na ausência do programa. Naturalmente, essa é uma condição não diretamente testável pelos dados: a despeito da quantidade ou qualidade das informações disponíveis, não é possível saber se a evolução da variável de resultados para o grupo de controle representa bem o contrafactual dessa variável para o grupo tratado após o programa.

4. Os capítulos seguintes tratarão de situações em que esse tipo de grupo de controle pode ser utilizado para identificar o efeito de interesse sob as hipóteses de seleção em observáveis (capítulo 5) e não observáveis (capítulos 6 e 7).

Mas o fato de essa hipótese não ser diretamente passível de confirmação empírica não significa que os dados não possam ser utilizados para revelar algo sobre sua validade. De fato, sempre que possível, a verificação da hipótese é feita de forma indireta por meio de um teste que requer que a tendência temporal da variável de resultado dos dois grupos seja a mesma antes do programa. A ideia é que uma trajetória temporal semelhante indica que ambos os grupos vinham reagindo de forma similar a todo e qualquer fator que afeta a variável de resultado antes da intervenção. Supõe-se então que, na ausência da intervenção, essa trajetória continuaria ao longo do tempo para o grupo tratado após o programa tal como ela efetivamente segue para o grupo de controle. Note que a condição é que os grupos tenham a mesma tendência temporal, mas não é necessário que eles partam exatamente do mesmo ponto antes do programa: o método de *DD* é capaz de lidar com diferenças existentes entre os grupos pré-programa. Se a condição de mesma tendência é atendida, então desvios na trajetória da variável de resultado entre os grupos após o programa são atribuídos aos efeitos causais da intervenção.

Caso haja disponibilidade de dados para vários pontos no tempo antes da intervenção, o teste de mesma trajetória temporal para os dois grupos pode ser facilmente realizado. Naturalmente, esse teste só pode ser feito com base nos dados do período pré-programa, já que as trajetórias dos grupos podem diferir por causa dos efeitos da intervenção logo após o início dela.⁵ Vale ressaltar que, embora a existência de mesma tendência temporal pré-programa não garanta que o grupo de controle seja um bom contrafactual para o grupo tratado, é recomendável que o teste seja realizado sempre que possível. De fato, caso a hipótese de mesma tendência não seja rejeitada, o teste confere maior credibilidade ao grupo de controle como contrafactual adequado para o grupo tratado.

O método de *DD* pode ser utilizado tanto com dados de painel quanto com dados seccionais (*cross-section*) repetidos no tempo. Os primeiros tipicamente referem-se a informações de indivíduos, famílias ou empresas que são seguidos ao longo do tempo, enquanto os segundos são dados temporais calculados para unidades mais agregadas, tais como escolas, hospitais, setores de

5. Em alguns casos, podem ocorrer efeitos antecipatórios do programa, isto é, mudanças no comportamento do grupo de tratamento ou controle antes mesmo do começo da intervenção. Por exemplo, quando as notícias da introdução de uma política ou programa vêm a público com alguma antecedência, certas decisões dos indivíduos tratados ou de controle podem se alterar, modificando as trajetórias da variável de resultado pré-programa. Em casos como esses, a verificação da existência de tendências semelhantes no período pré-programa pode ser realizada, retirando-se os dados mais próximos do início da intervenção.

atividade, bairros, municípios ou estados. A agregação dos dados normalmente é feita em termos de proporções ou médias das variáveis de interesse para a avaliação. Já os momentos no tempo geralmente são meses ou anos.

A despeito do tipo de dados utilizados, outra hipótese do método de **DD** é que a composição dos grupos de tratamento ou controle não se altere de forma significativa entre os períodos anterior e posterior à intervenção. Por exemplo, quando as informações são de painel, é possível que tenha ocorrido atrito amostral não aleatório com o grupo de tratamento ou controle (ou ambos). Nesse caso, o impacto do programa pode ser mal identificado pelo método de **DD**, já que as diferenças na variável de resultado entre períodos de tempo e entre grupos podem refletir não apenas o efeito do tratamento, mas também mudanças nas características não observadas dos indivíduos. De forma semelhante, quando os dados são de *cross-sections* repetidas, as unidades de observação deveriam manter, pelo menos aproximadamente, a mesma composição de características das subunidades (indivíduos, firmas etc.) que as compõem. Por exemplo, quando se está comparando uma região tratada (digamos, certas áreas de uma cidade) com regiões não tratadas (outras áreas da cidade), se estiver ocorrendo esvaziamento de pessoas de uma das áreas, é provável que a distribuição das características observáveis e não observáveis estejam mudando em um ou ambos os grupos. Esse problema torna-se mais acentuado se as mudanças na distribuição dessas características, principalmente a das não observadas, decorrem da própria intervenção. Parte então do que o método atribuirá ao efeito do tratamento sobre a variável de resultado estará, na realidade, associada às mudanças de composição dos grupos nas áreas tratadas e não tratadas.

Outra condição requerida pelo método de **DD** é que os grupos de tratamento e controle não sejam afetados de forma específica por mudanças de qualquer natureza que ocorram após o programa. Por exemplo, se houver uma mudança idiossincrática com o grupo de controle após a intervenção, a trajetória da variável de resultado desse grupo pode se alterar de forma a que ela deixe de representar adequadamente o contrafactual do grupo tratado. Uma mudança idiossincrática (independente do programa) também pode ocorrer com o grupo de tratados. Em ambos os casos, o método de **DD** captará os efeitos dessas mudanças, os quais serão erroneamente atribuídos aos efeitos do programa.

4.3. O modelo de diferenças em diferenças

A forma mais simples de expressar o estimador do método de *DD* é calculando uma dupla diferença de médias da variável de resultado. Se denotarmos por $T=\{1,0\}$ a participação ou não no programa e por $t=\{1,0\}$ os períodos posterior e anterior à intervenção, respectivamente, o estimador de *DD* é dado por:

$$(1) \quad \beta_{DD} = \{E[Y_i|T_i=1,t=1] - E[Y_i|T_i=1,t=0]\} - \{E[Y_i|T_i=0,t=1] - E[Y_i|T_i=0,t=0]\},$$

ou seja, pela diferença temporal pós e pré-programa da variável de resultado para o grupo tratamento subtraída da diferença correspondente calculada para o grupo de controle. Embutida nesse estimador está a hipótese de que a variação temporal na variável de resultado para o grupo de controle representa a variação contrafactual do grupo tratado, isto é, a variação que seria experimentada pelo grupo de tratamento na ausência do programa. Para ver isso, podemos usar a noção de resultados potenciais apresentada no capítulo 2, onde $Y_i(1)$ e $Y_i(0)$ correspondem à variável de resultado do indivíduo i respectivamente nas situações de tratamento e não tratamento e os valores efetivamente observados são expressos por $Y_i=T_i Y_i(1)+(1-T_i)Y_i(0)$. Num contexto com mais de um período de tempo, os resultados potenciais dependem se já ocorreu ou não o tratamento. Assim, as médias observadas que compõem a expressão (1) correspondem a:

$$E[Y_i|T_i=1,t=1] = E[Y_i(1)|T_i=1,t=1]$$

$$E[Y_i|T_i=1,t=0] = E[Y_i(0)|T_i=1,t=0]$$

$$E[Y_i|T_i=0,t=1] = E[Y_i(0)|T_i=0,t=1]$$

$$E[Y_i|T_i=0,t=0] = E[Y_i(0)|T_i=0,t=0],$$

o que permite escrever o estimador de *DD* como:

$$(2) \quad \beta_{DD} = \{E[Y_i(1)|T_i=1,t=1] - E[Y_i(0)|T_i=1,t=0]\} - \{E[Y_i(0)|T_i=0,t=1] - E[Y_i(0)|T_i=0,t=0]\}.$$

O efeito médio do tratamento sobre os tratados (*EMTT*) no contexto com mais de um período de tempo pode ser expresso por:

$$EMTT = E[Y_i(1)|T_i=1, t=1] - E[Y_i(0)|T_i=1, t=1],$$

ou seja, pela diferença na média dos resultados potenciais para os tratados após o programa. Naturalmente, não é possível observar a segunda média dessa expressão, mas se considerarmos a hipótese de que variação temporal na média da variável de resultado potencial para o grupo de tratados na situação de não tratamento é igual à variação correspondente para o grupo de não tratados, isto é, se considerarmos a hipótese de que:

$$E[Y_i(0)|T_i=1, t=1] - E[Y_i(0)|T_i=1, t=0] = E[Y_i(0)|T_i=0, t=1] - E[Y_i(0)|T_i=0, t=0], \quad (H1)$$

é possível escrever a média contrafactual buscada como:

$$E[Y_i(0)|T_i=1, t=1] = E[Y_i(0)|T_i=1, t=0] + \{E[Y_i(0)|T_i=0, t=1] - E[Y_i(0)|T_i=0, t=0]\}.$$

Inserindo esse resultado na expressão anterior do *EMTT*, temos que:

$$EMTT = \{E[Y_i(1)|T_i=1, t=1] - E[Y_i(0)|T_i=1, t=0]\} - \{E[Y_i(0)|T_i=0, t=1] - E[Y_i(0)|T_i=0, t=0]\} = \beta_{DD},$$

ou seja, com a hipótese *H1*, o estimador de *DD* identifica o *EMTT*. Essa hipótese, portanto, permite usar a diferença entre a variação efetivamente observada para o grupo de tratamento e a variação fornecida pelo grupo de controle para capturar o efeito causal da intervenção sobre os tratados.

Como mencionado anteriormente, a média da variável de resultado para o grupo de controle no período anterior ao programa não precisa coincidir com a média correspondente para o grupo de tratamento, ou seja, os grupos podem ou não partir de um mesmo ponto. Na realidade, na maior parte das aplicações do método de *DD*, os dois grupos têm essas médias distintas, um fato que tipicamente reflete as diferentes influências dos atributos observáveis e não observáveis dos indivíduos sobre a variável de resultado. O que o método de fato requer é que a variação temporal do que ocorre com o grupo de controle antes e depois do programa reflita corretamente a variação temporal do grupo de tratados na situação contrafactual de não tratamento.

Uma forma equivalente de expressar o estimador de **DD** é pela dupla diferença:

$$(3) \quad \beta_{DD} = \{E[Y_i|T_i=1,t=1] - E[Y_i|T_i=0,t=1]\} - \{E[Y_i|T_i=1,t=0] - E[Y_i|T_i=0,t=0]\},$$

que é apenas um rearranjo da expressão (1). A expressão (3) mostra que o estimador de **DD** também pode ser visto como o contraste das diferenças de médias existentes entre os dois grupos no período anterior e posterior ao programa. Aqui, torna-se ainda mais claro que o método permite a existência de diferenças de médias entre os dois grupos no período pré-programa, ou seja, o método não requer que o segundo termo entre chaves na expressão (3) seja igual a 0.

Para ilustrar como o método funciona, retornaremos ao exemplo hipotético que utilizamos no capítulo 2, no qual imaginamos um curso de treinamento profissional oferecido a trabalhadores de baixa escolaridade em certo município. O interesse recaía sobre o impacto do programa sobre o salário médio dos tratados, que era de R\$ 1.000 um pouco antes do início do curso e de R\$ 1.100 alguns meses após o programa. Argumentamos naquele capítulo que nem a comparação do grupo tratado antes e depois da intervenção, nem a comparação dos tratados com um grupo de não tratados somente após a intervenção isolam corretamente o impacto do programa. Suponha agora que fomos capazes de levantar informações sobre o grupo de não tratados no período anterior ao programa. A Tabela 4.1 contém os valores dos salários médios dos grupos de tratamento e controle – (T_1, T_0) , respectivamente – para os períodos anterior e posterior ao programa – (t_0, t_1) , respectivamente. A notação \bar{Y} representa a média amostral do salário, e Δ significa o operador de diferença.

Distintamente dos outros dois métodos, o procedimento de **DD** computa uma dupla diferença de médias. Considerando as expressões (1) e (3), há duas maneiras de se calcular essa dupla diferença. Seguindo a expressão (1), a primeira é obtida fazendo-se inicialmente a diferença das médias entre os períodos antes e depois do programa para cada grupo – isto é, $\Delta\bar{Y}_{T_1,t_1-t_0} = 100$ e $\Delta\bar{Y}_{T_0,t_1-t_0} = 50$ – e, posteriormente, calculando-se a segunda diferença entre os grupos: $\Delta\bar{Y}_{T_1,t_1-t_0} - \Delta\bar{Y}_{T_0,t_1-t_0} = 100 - 50 = 50$. A segunda maneira corresponde ao análogo amostral da expressão (3), sendo obtida pela subtração das médias salariais entre os grupos de tratamento e controle para cada período de tempo – isto é, $\Delta\bar{Y}_{T_1-T_0,t_0} = 150$ e $\Delta\bar{Y}_{T_1-T_0,t_1} = 200$ – e, a seguir, pela diferença dessa primeira subtração entre os períodos anterior e posterior ao programa: $\Delta\bar{Y}_{T_1-T_0,t_1} - \Delta\bar{Y}_{T_1-T_0,t_0} = 200 - 150 = 50$. O método de **DD** mostra então que o impacto

do curso foi de R\$ 50, o que representa um aumento salarial de 5% em média para os tratados relativamente à situação deles antes do início do programa.

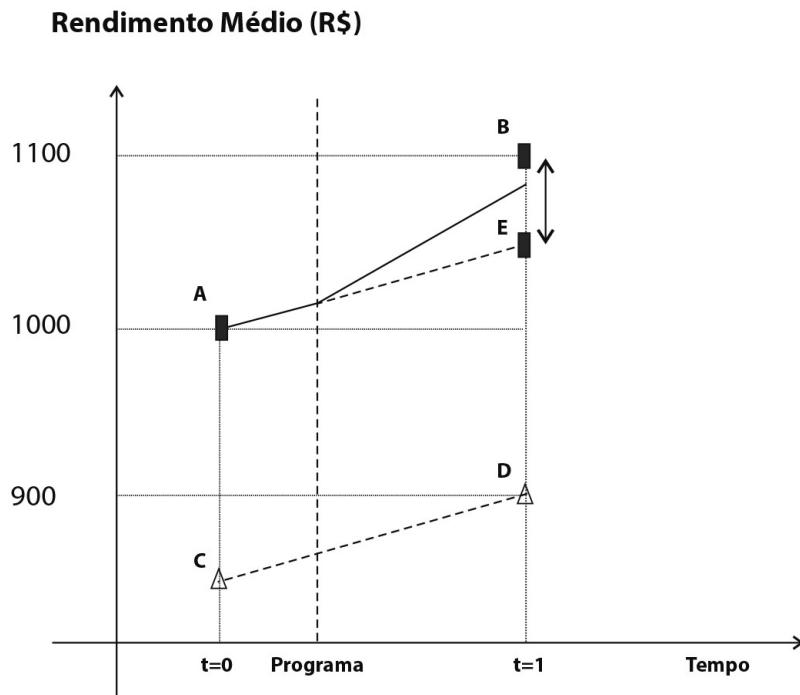
Tabela 4.1: Média salarial por grupo e período de tempo – R\$

	Tratados	Controle	Diferença
Antes	$\bar{Y}_{T_1,t_0} = 1000$	$\bar{Y}_{T_0,t_0} = 850$	$\Delta\bar{Y}_{T_1-T_0,t_0} = 150$
Depois	$\bar{Y}_{T_1,t_1} = 1100$	$\bar{Y}_{T_0,t_1} = 900$	$\Delta\bar{Y}_{T_1-T_0,t_1} = 200$
Diferença	$\Delta\bar{Y}_{T_1,t_1-t_0} = 100$	$\Delta\bar{Y}_{T_0,t_1-t_0} = 50$	$\begin{aligned} & \Delta\bar{Y}_{T_1,t_1-t_0} - \Delta\bar{Y}_{T_0,t_1-t_0} \\ &= \Delta\bar{Y}_{T_1-T_0,t_1} - \Delta\bar{Y}_{T_1-T_0,t_0} \\ &= 50 \end{aligned}$

O funcionamento do método de **DD** também pode ser visualizado pela Figura 4.1. Nela, as médias salariais observadas dos grupos de tratamento e controle estão representadas pelos quadrados e triângulos preenchidos, respectivamente. O principal aspecto a se destacar nessa Figura é o ponto *E*, que representa o contrafactual do grupo de tratamento após o programa. Note que a posição dele é obtida pela projeção da média salarial dos tratados, que é paralela à que ocorreu com o grupo de controle depois do programa. Essa projeção, ilustrada pela linha pontilhada que se liga ao ponto *E* na Figura 4.1, é baseada na hipótese de que a trajetória do salário médio do grupo de tratamento na ausência do programa é igual à trajetória dessa variável para o grupo de controle após a intervenção. Essa é a principal hipótese do método de **DD**.

A magnitude do impacto do programa pode ser obtida na Figura 4.1 a partir da diferença de pontos: $(B-A)-(D-C)=(1100-1000)-(900-850)=50$, ou seja, pela primeira expressão de dupla diferença descrita anteriormente. Note que o impacto não é calculado pela simples variação salarial antes e depois do programa para o grupo tratado (i.e., **B-A**), mas sim pelo contraste dessa variação com a experimentada pelo grupo de controle no mesmo intervalo (i.e., **D-C**). Utilizando a segunda expressão de dupla diferença apresentada anteriormente, obtemos $(B-D)-(A-C)=(1100-900)-(1000-850)=50$, o resultado mostra que o impacto não é dado apenas pela diferença entre tratados e controles pós-tratamento (i.e., **B-D**), mas sim pela comparação dessa diferença com a sua correspondente antes do programa (i.e., **A-C**).

Figura 4.1: Rendimento médio dos tratados e não tratados, antes e depois do programa



Podemos também apresentar o modelo de *DD* por meio de regressões lineares. Suponhamos inicialmente que só possuímos observações para dois períodos de tempo, um anterior e o outro posterior ao programa. A equação básica do modelo pode ser especificada, como:

$$(4) \quad Y_{it} = \alpha + \gamma T_i + \beta (T_i \cdot dt_t) + \varepsilon_{it},$$

onde o subscrito *t* foi acrescentado à notação dos capítulos anteriores para denotar o período de tempo no qual o indivíduo (ou unidade de observação) *i* se encontra. Como antes, T_i é uma variável binária que assume valor unitário se o indivíduo é tratado e valor nulo caso contrário. A variável dt_t também é binária, assumindo valor 1 no período pós-programa e valor 0 caso contrário. O termo ε_{it} representa distúrbio aleatório com média nula. Note que, na equação (4), as variáveis T_i e dt_t aparecem tanto

isoladamente quanto interagidas. Suas presenças isoladas captam as diferenças da média de Y_{it} entre o grupo de tratamento e controle e entre o período anterior e posterior ao programa, respectivamente. Já o termo de interação capta o que ocorreu especificamente com o grupo de tratamento no período pós-programa, ou seja, se a média da variável de resultado para esse grupo tornou-se diferente após a intervenção. É o parâmetro β , portanto, que mede o efeito do programa na equação de regressão (4).

A equação (4) também pode ser obtida a partir do arcabouço de resultados potenciais. Na ausência da intervenção, a média da variável de resultado é expressa por:

$$(5) \quad Y_{it}(0) = \alpha + \varrho dt_t + \varepsilon_{it},$$

ao passo que média correspondente para a situação de tratamento é dada por:

$$(6) \quad Y_{it}(1) = \alpha + \gamma + \varrho dt_t + \beta dt_t + \varepsilon_{it} = Y_{it}(0) + \gamma + \beta dt_t.$$

Introduzindo as equações (5) e (6) na expressão para variável de resultado observada $Y_{it} = T_i Y_{it}(1) + (1 - T_i) Y_{it}(0)$, obtemos:

$$(7) \quad Y_{it} = T_i [Y_{it}(0) + \gamma + \beta dt_t] + (1 - T_i) Y_{it}(0) = Y_{it}(0) + \gamma T_i + \beta T_i dt_t = \alpha + \gamma T_i + \varrho dt_t + \beta (T_i dt_t) + \varepsilon_{it},$$

que é uma expressão igual à equação de regressão especificada em (4).

Essa relação entre as equações de regressão e de resultados potenciais permite visualizar a hipótese básica (H1) do método de **DD**. Note que a expressão (5) especifica a variável de resultado na situação de não tratamento tanto para os controles quanto para os tratados ao longo do tempo. Em particular, note que o parâmetro ϱ impõe que a variável de resultado na ausência da intervenção varia entre os períodos de tempo da mesma forma para os não tratados e para os tratados. Em outras palavras, a expressão (5) garante que a média da variável de resultado contrafáctual para os tratados evolui exatamente como a média para os controles entre os períodos de tempo.

Outro ponto a se notar é que as equações (5) e (6) têm formato aditivo (i.e., os termos de cada expressão entram de forma separada por somas). Embora comum, esse formato aditivo é particularmente importante para o método de **DD**, pois é ele que permite o isolamento do efeito

de interesse pela operacionalização da dupla diferença que caracteriza o método. Para verificar isso, iniciemos por escrever as médias da variável de resultado para os seguintes casos (retirando os subscritos para facilitar a leitura):

(A) Tratados, pré-programa:

$$E[Y|T=1, t=0] = \alpha + \gamma + E[\varepsilon|T=1, t=0] = \alpha + \gamma + E[\varepsilon_{10}]$$

(B) Tratados, pós-programa:

$$E[Y|T=1, t=1] = \alpha + \gamma + \varrho + E[\varepsilon|T=1, t=1] = \alpha + \gamma + \varrho + E[\varepsilon_{11}]$$

(C) Controles, pré-programa:

$$E[Y|T=0, t=0] = \alpha + E[\varepsilon|T=0, t=0] = \alpha + E[\varepsilon_{00}]$$

(D) Controles, pós-programa:

$$E[Y|T=0, t=1] = \alpha + \varrho + E[\varepsilon|T=0, t=1] = \alpha + \varrho + E[\varepsilon_{01}],$$

onde $E[\varepsilon|T=m, t=k] = E[\varepsilon_{mk}]$ corresponde à média do termo de erro para o grupo $T=m=\{1,0\}$ e período de tempo $t=k=\{0,1\}$. Calculando agora a dupla diferença $\{(B)-(A)\} - \{(D)-(C)\}$, temos que:

$$(8) \quad \{(B)-(A)\} - \{(D)-(C)\} = \{(\alpha + \gamma + \varrho + \beta + E[\varepsilon_{11}]) - (\alpha + \gamma + E[\varepsilon_{10}])\} - \{(\alpha + \varrho + E[\varepsilon_{01}]) - (\alpha + E[\varepsilon_{00}])\} = \{\varrho + \beta + E[\varepsilon_{11}] - E[\varepsilon_{10}]\} - \{\varrho + E[\varepsilon_{01}] - E[\varepsilon_{00}]\} = \beta + \{(E[\varepsilon_{11}] - E[\varepsilon_{10}]) - (E[\varepsilon_{01}] - E[\varepsilon_{00}])\} = \beta,$$

onde a última igualdade decorre da hipótese (H1).⁶ Portanto, como se pode ver pela expressão (8), o formato aditivo (combinado com H1) permite que vários termos se anulem na operacionalização da dupla diferença, possibilitando assim isolar o parâmetro de interesse.

6. Usando a expressão (5), a hipótese H1 pode ser escrita como: $E[Y_i(0)|T_i=1, t=1] - E[Y_i(0)|T_i=1, t=0] = E[Y_i(0)|T_i=0, t=1] - E[Y_i(0)|T_i=0, t=0]$. $(\alpha + \gamma + \varrho + E[\varepsilon_{11}]) - (\alpha + \gamma + E[\varepsilon_{10}]) = (\alpha + \varrho + E[\varepsilon_{01}]) - (\alpha + E[\varepsilon_{00}])$. Isso mostra que o termo entre colchetes na expressão antes da última igualdade de (8) é igual a 0.

Uma das utilidades do modelo de regressão para expressar o método de **DD** é que ele facilita a inclusão de controles para características observáveis. A introdução dessas características é muitas vezes importante, já que elas podem absorver parte do efeito que, na ausência delas, seria equivocadamente atribuído à intervenção. Ademais, quando a variável de resultado é afetada pelas características observáveis, a introdução delas no modelo também pode permitir estimativas mais precisas do efeito de interesse, já que elas tendem a absorver parte da variabilidade da variável de resultado.

Denotando por $X_{it} = (X_{it1}, \dots, X_{itK})$ o vetor com K características observáveis do indivíduo i no período t , a equação de regressão (4) passa a ser expressa por:

$$(9) \quad Y_{it} = X'_{it}\alpha + \gamma T_i + \beta(T_i d_{it}) + \varepsilon_{it}$$

onde α agora é um vetor de parâmetros conforme o número de variáveis em X_{it} e que contém um intercepto (ou seja, um dos elementos do vetor X_{it} é igual a 1). Embora esteja indexado com o subscrito t , o vetor X_{it} que aparece em (9) pode conter características observáveis dos indivíduos tanto invariantes quanto variantes no tempo. Exemplos típicos do primeiro caso são sexo e cor, e do segundo, idade e renda.

Uma das implicações da inclusão das características observáveis no modelo é que a análise passa a ser condicionada a X_{it} . Por exemplo, as médias da variável de resultado passam a ser expressas por $E[Y_{it}|X_{it}, T_i, t]$, ou seja, as médias passam agora a depender não somente do tipo de grupo (T_i) e do período de tempo (t) mas também das características observadas presentes em X_{it} .⁷ O parâmetro β da equação (9) continua a fornecer o efeito de interesse, porém agora levando-se em consideração que ele é obtido com a inclusão na regressão das variáveis de controles para as características em X_{it} .

A equação (9) pode ser expandida para o caso em que há disponibilidade de informações para vários períodos de tempo (por exemplo, meses ou anos). Seja então $t=1, 2, \dots, \tau$, onde $\tau \geq 3$ representa o número máximo de períodos para o qual temos informações para os grupos de tratados e controles. Digamos que a intervenção ocorra em certo mês (ou ano) t_0 ao longo do período de

⁷ Em particular, a hipótese H1 passa a ser expressa por: $E[Y_{it}(0)|X_{it}, T_i=1, t=1] - E[Y_{it}(0)|X_{it}, T_i=1, t=0] = E[Y_{it}(0)|X_{it}, T_i=0, t=1] - E[Y_{it}(0)|X_{it}, T_i=0, t=0]$.

observação – isto é, $1 < t_0 < \tau$. Considere a equação a seguir, que é uma modificação da equação (9) para múltiplos períodos:

$$(10) \quad Y_{it} = X_{it}\alpha + \gamma T_i + \varrho_t + \beta D_{it} + \varepsilon_{it},$$

onde $T_i = \{1, 0\}$ permanece uma variável binária que assume valor 1 se o indivíduo i é tratado e valor 0 caso contrário. O componente ϱ_t capta efeitos agregados de tempo (por exemplo, inflação) que afetam a variável de resultado de todos os indivíduos, tanto no grupo de tratamento quanto no de controle. Note que, com $t=1, \dots, \tau$, $\tau \geq 3$, haverá múltiplos ϱ_t a serem estimados (e não apenas um como no contexto com dois períodos), o que é operacionalizado incluindo *dummies* de tempo para cada período t na regressão.⁸ A variável $D_{it} = \{1, 0\}$ também é binária, mas só assume o valor unitário se o indivíduo i é tratado no período t .⁹ Nesse contexto de múltiplos períodos, essa variável cumpre o papel do termo de interação que aparece na equação (9) para dois períodos (i.e., $T_i d_{it}$). Portanto, o efeito causal do programa também será capturado pelo parâmetro β da equação (10).

Uma das principais vantagens do método de **DD** é que ele é capaz de controlar para as influências sobre a variável de resultado das características não observáveis dos indivíduos que sejam fixas no tempo. Essa é uma vantagem importante do método pois, muitas vezes, esse tipo de características não observáveis influencia não apenas a variável de resultado, mas, também, a decisão de participar do programa. Por exemplo, as habilidades inatas tendem a influenciar não somente os salários, mas também a decisão das pessoas de participar em cursos de treinamento. A motivação ou certos traços de comportamento (inibição, autoestima, entre outros) são outros exemplos de características não observáveis que podem afetar tanto salários quanto a decisão de participar do programa. Quaisquer que sejam as características não observáveis em questão, o importante é que o método de **DD** é capaz de levar em conta a associação entre a variável de resultado, a participação no programa e as características não observáveis dos indivíduos que sejam invariantes no tempo (pelo menos durante o período de observação). Assim, como o problema do viés de seleção pode surgir da associação entre essas dimensões, o método de **DD** oferece uma ferramenta poderosa para contornar esse problema.

8. Em geral, exclui-se uma dessas *dummies* para evitar o problema de multicolinearidade perfeita com o vetor unitário presente em X_{it} .

9. Ou seja, ela será igual a 1 somente para os indivíduos tratados nos períodos após o início da intervenção (isto é, $t \geq t_0$).

Para verificar como o método de **DD** controla para a influência das características não observáveis que não se alteram no tempo, podemos utilizar uma especificação da regressão linear tipicamente empregada em modelos de dados de painel. Nesse tipo de modelo, a expressão para a variável de resultado inclui o chamado efeito fixo individual, que representa um elemento que absorve a influência de qualquer característica (observável ou não) fixa no tempo do indivíduo sobre a variável de resultado. Assumindo a existência de múltiplos períodos de tempo $t=1,2,\dots,\tau$ e uma amostra com $i=1,2,\dots,N$ indivíduos, a equação de regressão passa a ser escrita como:

$$(11) \quad Y_{it} = X_{it}\alpha + \varrho_t + \beta D_{it} + \mu_i + \varepsilon_{it}$$

onde μ_i representa o componente específico do indivíduo i que não varia no tempo. A inclusão desses efeitos fixos específicos dos indivíduos permite controlar para heterogeneidades existentes entre as pessoas em características que sejam fixas no tempo.¹⁰ Uma propriedade dos efeitos fixos é que eles podem ser correlacionados com qualquer característica observada dos indivíduos, incluindo a que determina a participação no programa, que está embutida na variável D_{it} . Ao permitir esse tipo de correlação, o método de **DD** é capaz de contornar o problema de viés de seleção associado às características fixas (observáveis ou não) dos indivíduos no tempo.

Para que o parâmetro β possa ser estimado de forma consistente pela equação (11) é importante que o termo ε_{it} não seja correlacionado com as variáveis do modelo, em particular com D_{it} , que está associada com a participação no programa. Mais especificamente, condicionando-se à presença do efeito fixo, μ_i , e das variáveis em X_{it} , requer-se que qualquer fator não observado que afete a variável de resultado ao longo do tempo não influencie a decisão de participar do programa.¹¹

Uma forma típica de estimar a equação (11) é calcular a diferença entre dois períodos de tempo para todas as variáveis da equação. Esse procedimento, conhecido como “primeiras diferenças”, elimina μ_i

10. Estrito senso, a presença do efeito fixo individual na equação capta a influência sobre Y_{it} de qualquer característica, observável ou não, que não varie com o tempo. Isso implica que os efeitos de variáveis fixas no tempo não são identificados conjuntamente com o efeito fixo individual. Por essa razão, o termo T_i não aparece na equação (11). De forma semelhante, eventuais variáveis tempo-invariantes contidas no vetor X_{it} (por exemplo, sexo e cor) também não apareceriam na equação.

11. Formalmente, uma hipótese que garante essa situação é a de exogeneidade estrita, que é expressa por: $E[\varepsilon_{is}|X_{it},\mu_i,D_{it}] = 0, \forall i \text{ e } s \leq t$.

da equação, já que este componente é fixo no tempo.¹² Especificamente, a equação a ser estimada torna-se:

$$(12) \Delta Y_{it} = Y_{it} - Y_{it-1} = (X'_{it} - X'_{it-1})\alpha + (\varrho_t - \varrho_{t-1}) + \beta(D_{it} - D_{it-1}) + (\mu_i - \mu_i) + (\varepsilon_{it} - \varepsilon_{it-1}) = \Delta X'_{it}\alpha + \lambda_t + \beta\Delta D_{it} + \Delta\varepsilon_{it},$$

onde $\lambda_t = \varrho_t - \varrho_{t-1}$ é o novo componente que capta efeitos agregados de tempo e Δ é o operador da primeira diferença.¹³ Vale notar que, neste procedimento, não só o efeito fixo é removido da equação, mas também qualquer variável presente em X_{it} que não varie no tempo (por exemplo, o sexo e a cor do indivíduo).

Embora estejamos nos referindo a indivíduos ao apresentar as equações acima, *mutatis mutandis*, elas também podem ser aplicadas para o caso em que as unidades de observação estão num nível mais agregado, como escolas, setores de atividade, municípios ou estados. Esse tipo de aplicação do método **DD** é útil por pelo menos duas razões. Em primeiro lugar, porque muitas vezes não há informações para todas as variáveis relevantes do modelo no âmbito individual, mas há disponibilidade de dados nas dimensões agregadas de interesse. Por exemplo, por motivo de confidencialidade das informações de indivíduos ou empresas, é comum que sejam disponibilizados somente dados agregados de escolas, setores de atividade ou mesmo municípios inteiros.¹⁴ Em segundo lugar, porque muitas intervenções são introduzidas para as unidades agregadas como um todo, o que faz com que, pelo menos potencialmente, todos os indivíduos ou empresas dessas

12. Outra forma de estimação comumente utilizada com dados de painel é o denominado método “intragrupos”. Nesse caso, cada variável é subtraída do valor de sua média temporal para cada indivíduo, por exemplo, subtrai-se Y_{it} de $\bar{Y}_i = \frac{1}{T} \sum_{t=1}^T Y_{it}$. Como o efeito fixo não varia temporalmente, ele é removido da equação por esse tipo de subtração.

13. Embora fora do escopo do livro, é importante observar que, se o termo $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$ apresentar correlação serial – isto é, se a correlação entre ε_{it} e ε_{it-1} for diferente de 0 –, há que se fazer alguma correção para os erros-padrão estimados pelo procedimento de **DD**. Por exemplo, é comum que choques passados sobre o salário e o estado ocupacional das pessoas (emprego, desemprego, inatividade) sejam correlacionados. Ignorar esse problema tipicamente subestima o erro-padrão de β , levando a conclusões equivocadas sobre a significância estatística do efeito do programa. Mais detalhes sobre o problema de correlação serial assim como algumas soluções para ele no contexto de **DD** podem ser encontrados em Bertrand et al. (2004).

14. Vale assinalar que o uso de dados agregados tem a vantagem de que eles tendem a atenuar problemas de erro de medida que tipicamente aparecem em informações coletadas no âmbito do indivíduo ou da família. Por exemplo, o uso de médias (ou proporções) no nível do estado ou município tende a mitigar os erros de medida em diversas variáveis comumente utilizadas em avaliações, tais como salários, status ocupacional, escolaridade e participação no programa.

unidades sejam afetados pela intervenção que se quer avaliar. Exemplos típicos são os programas para todo o sistema escolar ou de saúde de um estado ou município ou mudanças institucionais para grupos inteiros de setores de atividade (e.g., desonerações fiscais).

A equação de regressão é expressa de formas distintas dependendo do tipo de dados disponíveis (individual ou agregado). Por exemplo, assumindo que a variável de resultado está disponível no nível individual e que as demais variáveis somente no nível de agregação, a equação é expressa como:

$$Y_{igt} = X'_{gt}\alpha + \beta Q_t + \mu_g + \varepsilon_{igt} \quad (13)$$

onde o subscrito $g=1, \dots, G$ representa o nível de agregação no qual está o indivíduo i (por exemplo, escola, município ou estado). O termo μ_g pode ser visto como um efeito fixo para a população de indivíduos que pertence ao grupo g , e o termo ε_{igt} denota um erro aleatório não correlacionado com as demais covariáveis do modelo.¹⁵ A variável de tratamento D_{gt} representa uma *dummy* que assume valor 1 para o grupo tratado no período t e valor 0 caso contrário. Note que vários grupos não tratados vão desempenhar o papel de grupo de controle para o grupo que passou pela intervenção. Como mencionado anteriormente, é importante testar se as tendências temporais do grupo de tratamento e do grupo de controle são semelhantes antes da intervenção.¹⁶

Para que o método de *DD* identifique corretamente o efeito causal de uma intervenção nesse contexto de dados agregados, é importante que tanto indivíduos tratados quanto os de controle permaneçam similares nas suas características não observáveis ao longo do tempo, especialmente após a intervenção. Caso ocorram mudanças expressivas, a composição dos indivíduos que formam os grupos $g=1, \dots, G$ se alterará ao longo do tempo, o que complica a comparabilidade temporal

15. Em casos como o da equação (13), onde há informações individuais e agregadas, é provável que haja correlação dos termos de erros entre os indivíduos de um mesmo grupo (também chamados de *clusters* na literatura econométrica) – ou seja, formalmente, $\text{corr}(\varepsilon_{igt}, \varepsilon_{jgt}) \neq 0, i \neq j$. Por exemplo, é provável que os alunos de uma mesma escola ou município estejam sujeitos a mudanças não observadas comuns entre eles. Nesses casos, o erro-padrão do parâmetro de interesse β pode ser seriamente subestimado. Há, portanto, que se operacionalizar correções na estimativa dos erros-padrão do modelo. Mais detalhes sobre essas correções podem ser encontrados nos livros de estatística ou econometria que tratam de regressão com a presença de *clusters*. Outra referência é Cameron e Miller (2015).

16. Proposto por Abadie et al. (2010) para contextos em que a unidade de tratamento é agregada, o método de controle sintético procura encontrar de forma explícita um grupo de controle que se assemelhe ao grupo tratado antes da intervenção. Uma descrição sucinta desse método é apresentada no Box 4.1 ao final deste capítulo.

dos grupos. Por exemplo, a migração de pessoas entre municípios ou estados pode tornar os grupos muito diferentes antes e depois de uma intervenção em características não observáveis (por exemplo, habilidades valorizadas no mercado de trabalho). A perda de comparabilidade dos grupos ao longo do tempo compromete a acurácia do método de **DD**, já que o impacto estimado pode estar misturado com efeitos das mudanças de características não medidas dos grupos. Uma boa prática em avaliações que utilizam dados de *cross-section* repetidos é verificar se as características observáveis dos grupos tratados e não tratados permanecem relativamente estáveis no tempo. Esse tipo de problema também pode ocorrer com dados longitudinais de indivíduos quando ocorre atrito amostral.

4.4. Desvantagens

Embora possua uma série de vantagens, o método de **DD** tem dificuldade de lidar com casos em que alguma mudança temporária num fator não observável dos indivíduos afeta a decisão de participar no programa. Por exemplo, já se detectou que muitos trabalhadores que fazem cursos de treinamento experimentaram choques negativos e temporários de renda um pouco antes do começo do programa (Ashenfelter, 1978, e Heckman e Smith, 1999). Se essa mudança temporária não ocorrer com o grupo de controle, o procedimento de **DD** estimará o impacto do programa de forma incorreta. De forma geral, o estimador de **DD** não identificará o efeito de interesse quando houver algum tipo de característica não observável que varie no tempo e afete simultaneamente a variável de resultado e a participação no programa.¹⁷

17. Este tipo de situação leva ao uso de métodos de painel em que a variável de resultado aparece de forma defasada na regressão. Está fora do escopo deste livro tratar desse tipo de regressão. Os interessados podem encontrar material sobre o tema em Wooldridge (2002) e Angrist e Pischke (2009).

Box 4.1: Método do controle sintético

Desenhado para contextos em que a intervenção ocorre em uma unidade agregada (por exemplo, um estado de um país), Abadie et al. (2010) propõe um método que utiliza um conjunto de unidades não tratadas para “sintetizar” um grupo de controle que se pareça com a unidade tratada antes da intervenção. Distantemente do método de diferenças em diferenças convencional (*DD*), em que tipicamente há alguma incerteza sobre quais as unidades não tratadas são as mais apropriadas para servirem como grupo de controle, o método do controle sintético constrói explícita e intencionalmente o grupo de controle com o objetivo de torná-lo semelhante ao grupo tratado, em particular no que diz respeito à tendência temporal da variável de resultado antes do programa. Para tanto, ele emprega um procedimento no qual as unidades não tratadas (por exemplo, os estados do país não afetados pela intervenção) recebem pesos que são obtidos com base num conjunto de preditores da variável de resultado antes do programa. Um aspecto importante do método é que esses preditores podem incluir tanto as características observadas quanto as próprias trajetórias da variável de resultado das unidades antes do programa.

Formalmente, o método considera a existência de $j=1, \dots, J+1$ unidades agregadas, com $j=1$ a unidade tratada e $j=2, \dots, J+1$ as unidades não tratadas (cujo conjunto é chamado de *donor pool*) e $t=1, \dots, T$ os períodos de observação, sendo $t=1, \dots, T_0$ o intervalo pré-intervenção ($T_0 < T$). Usando a notação de resultados potenciais, se $Y_{1t} \equiv Y_{1t}(1)$ denota a variável de resultado para a unidade tratada e $Y_{1t}^N \equiv Y_{1t}(0)$ a sua correspondente contrafactual no período t , o efeito do programa é dado por $\beta_{1t} = Y_{1t} - Y_{1t}^N$. Como só observamos Y_{1t} , é preciso encontrar Y_{1t}^N .

Essa tarefa é operacionalizada com base num vetor de pesos $W=(w_2, \dots, w_{J+1})$ – cujos elementos são positivos e cuja soma é igual a um – que vai servir para ponderar a variável de resultado das unidades não tratadas (i.e., $\sum_{j=2}^{J+1} w_j Y_{jt}$) de forma a construir um grupo de controle sintético. Cada valor de W corresponde a um grupo de controle sintético potencial, e o método vai buscar um vetor W^* que minimize a distância da variável de resultado antes do programa e de outros preditores relevantes dessa variável entre a unidade tratada e as unidades da *donor pool*. Sob certas condições, Abadie et al. (2010) mostram

Box 4.1: Método do controle sintético (cont.)

que o vetor W^* torna a média ponderada da variável de resultado das unidades da *donor pool* muito próxima de Y_{1t}^N , que é variável contrafactual buscada. O impacto do programa pode então ser estimado por $\hat{\beta}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$.

Na aplicação empírica apresentada em Abadie et al. (2010), os autores avaliam o impacto sobre o consumo de tabaco de uma lei antitabagista que entrou em vigor no estado da Califórnia (EUA) em 1988. Eles utilizam 38 outros estados americanos que

não foram afetados pela lei para formarem a *donor pool* e obtêm dados sobre o consumo de cigarros e outros preditores dessa variável (p.e., percentagem de jovens e consumo de cerveja) para todos os estados durante o período entre 1970 e 2000. A aplicação do método permitiu construir um grupo de controle sintético muito parecido com o da Califórnia tanto nos preditores quanto na trajetória de consumo de cigarros pré-introdução da lei. O impacto dessa intervenção reduziu o consumo de cigarros ao longo do tempo.

Box 4.2: O impacto do Programa Saúde da Família

O Programa de Saúde da Família (PSF) procura prover atenção básica de saúde às famílias de comunidades carentes no Brasil. O funcionamento do programa baseia-se no uso de equipes de profissionais de saúde (médicos, enfermeiras, auxiliares de enfermagem e agentes comunitários) cujos principais objetivos são a provisão de práticas básicas de saúde (e.g., estratégias para lidar com problemas simples de saúde) às famílias, o encaminhamento das pessoas para clínicas e hospitais em casos específicos e o acompanhamento da situação de saúde das famílias beneficiadas ao longo do tempo. Implementado com a participação das três esferas de governo (federal, estadual e municipal), o PSF iniciou-se na metade dos anos 90. O programa foi se expandindo

progressivamente ao longo dos anos e em meados da década de 2000 já atendia a quase totalidade dos municípios do país.

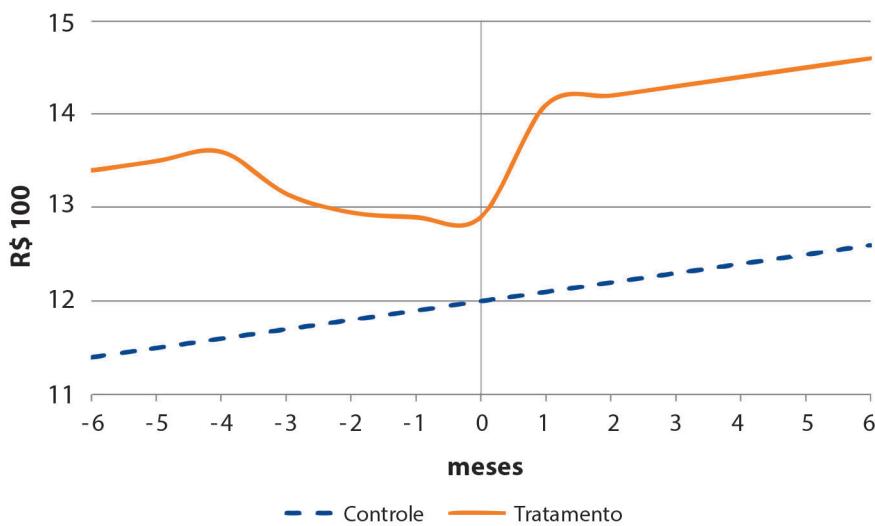
Utilizando informações no nível municipal, Rocha e Soares (2010) avaliam o impacto do PSF comparando os municípios tratados e não tratados pelo programa ao longo do tempo. Os autores utilizam o método das diferenças em diferenças explorando as diferentes datas de entrada dos municípios no programa. Os resultados mostram que o PSF foi capaz de reduzir a taxa de mortalidade das pessoas de várias faixas etárias, notadamente do grupo de crianças. Outros efeitos positivos foram estimados sobre a oferta de trabalho dos adultos e a frequência à escola das crianças das famílias das regiões NO e NE.

4.5. Exercícios:

- 1.** Distintamente dos casos de experimentos naturais (terremotos, tsunamis, erupções vulcânicas etc.), na área de avaliação de políticas ou programas sociais não se tem certeza de que a introdução ou mudança de política ocorreu de forma totalmente exógena. Crie ou busque exemplos de programas sociais que se aproximem da situação de experimento natural.
- 2.** Foi noticiado em dois jornais locais que a política atual de redução da criminalidade de seu estado era efetiva. Em um dos jornais, a comprovação dessa efetividade foi associada à informação de que a taxa de roubos de veículos havia caído 10% nos últimos dois anos. Já no outro jornal, a efetividade da política foi consubstanciada por meio da comparação da taxa de roubo de veículos no último ano do seu estado com a média da mesma taxa para os demais estados do país. Como um conhecedor da área de avaliação de impacto, ao ler os jornais você considera ambas as matérias ingênuas e decide tentar estimar o verdadeiro impacto do programa. Responda aos itens a seguir.

 - a.** Por que você acha a matéria do primeiro jornal ingênuas? E a segunda? Apresente as equações de regressão que você escreveria para representar cada uma das matérias dos jornais.
 - b.** Você decide estimar o efeito da política usando o método de diferenças em diferenças (*DD*). Suponha que existem dados disponíveis para vários estados (incluindo o seu) ao longo do tempo para os períodos anterior e posterior à introdução da política. Apresente a equação de regressão que você usaria para estimar o efeito da política pelo método *DD*. Em que ela difere das equações de regressão do item anterior?
 - c.** O que você deveria testar com dados anteriores ao início da política para averiguar se o grupo de controle parece adequado aos seus propósitos?
- 3.** Uma das grandes vantagens do método de diferenças em diferenças é que ele é capaz de lidar com a seleção decorrente de certo tipo de características não observadas dos indivíduos. Que tipo de características é esse, e como o método as incorpora na regressão? Mostre que variáveis de características observadas dos indivíduos invariantes no tempo (e.g. sexo) não são identificadas nessa regressão.

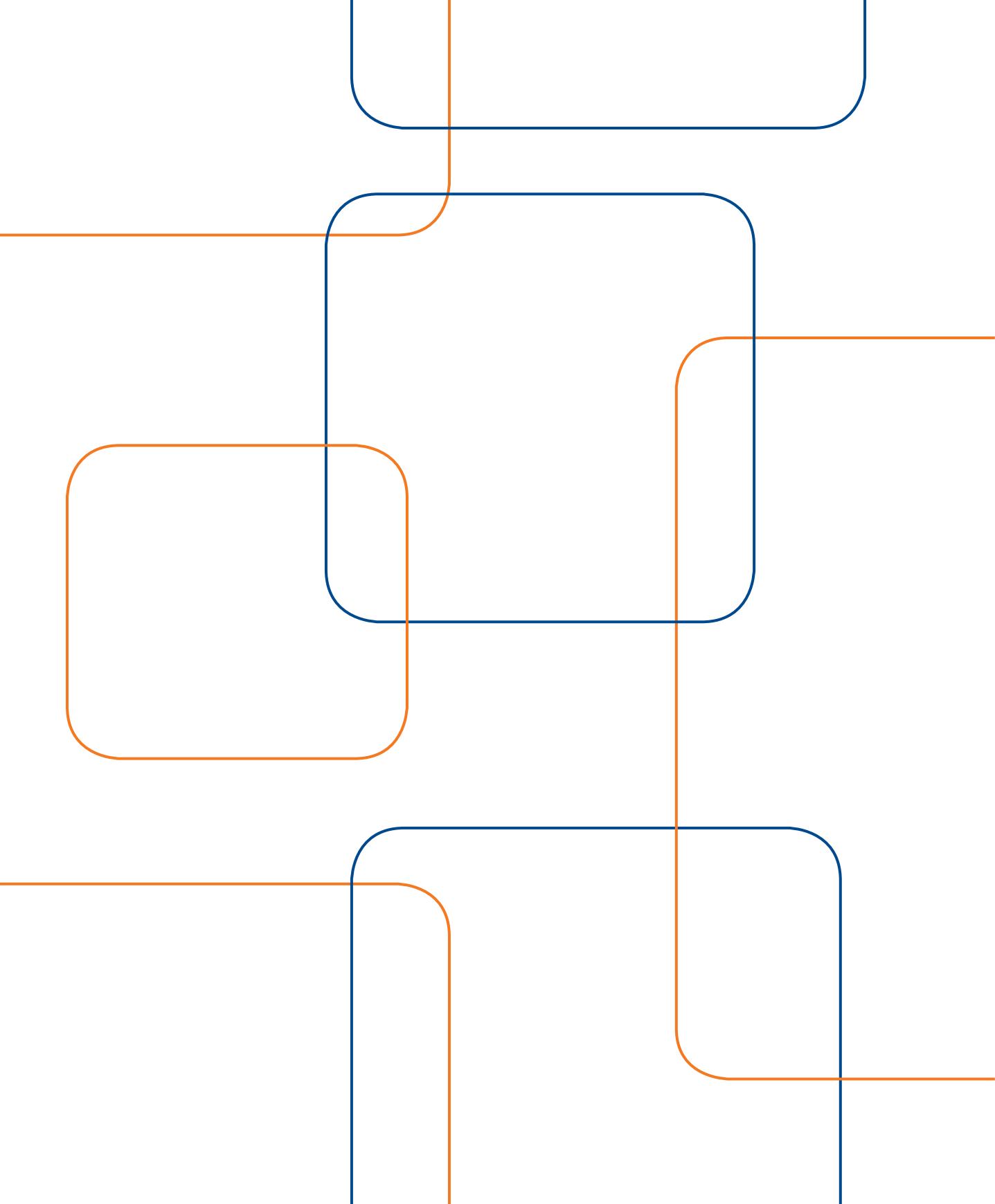
4. Certo município implementou um programa habitacional que começou a atrair um conjunto significativo de moradores de outros municípios para residirem lá. Suponha que um analista é contratado para avaliar o impacto do programa e decide utilizar o método de diferenças em diferenças com informações no nível municipal. Discuta os problemas que o avaliador precisa levar em consideração com os dados antes e após o programa para aplicar corretamente o método.
5. O gráfico 4.2 abaixo mostra as trajetórias do salário médio dos grupos de tratamento e controle de uma avaliação de impacto para um curso de treinamento hipotético que ocorreu no momento 0 no gráfico. É possível que o método das diferenças em diferenças estime de forma enviesada o impacto do programa? Por quê?

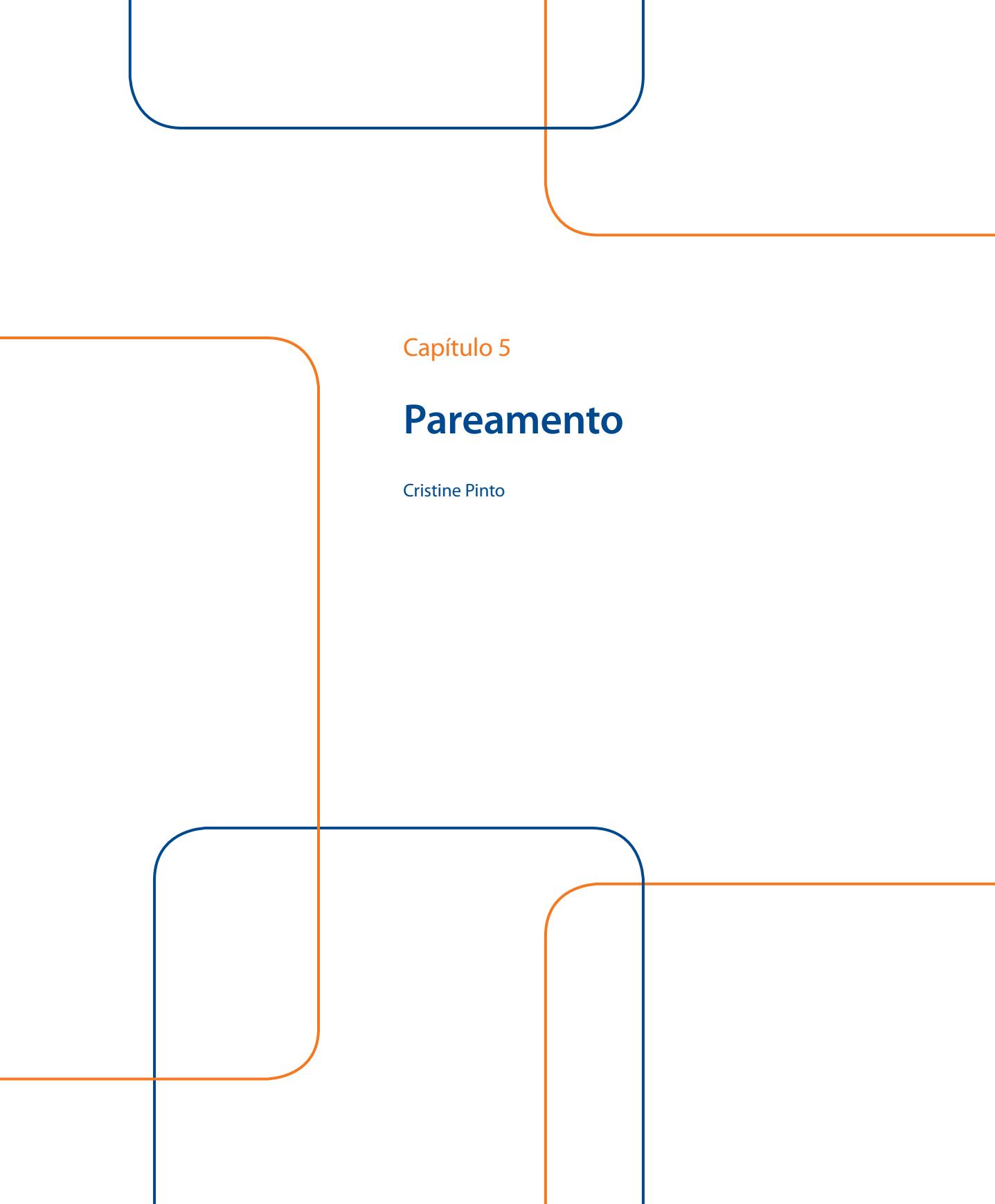


6. Com base na equação de regressão para dois períodos:

$$Y_{it} = \alpha + \gamma T_i + \varrho d_{it} + \beta(T_i * d_{it}) + \varepsilon_{it},$$

ofereça uma interpretação para cada parâmetro α, γ, ϱ e β utilizando as médias condicionais de Y_{it} para tratados e controle nos períodos anterior e posterior à intervenção. Plote esses parâmetros num gráfico hipotético com a variável de resultado no eixo vertical, e os períodos, no eixo horizontal.





Capítulo 5

Pareamento

Cristine Pinto

O método de pareamento busca construir um grupo de controle semelhante ao grupo de tratamento referente a determinadas características observáveis.¹ De acordo com as hipóteses desse método, cada membro do grupo de tratamento teria um par no grupo de controle, que representa o resultado que ele teria obtido caso não fosse tratado. De outra forma, as hipóteses do pareamento postulam que, ao comparar dois indivíduos, um no grupo de controle e outro no grupo de tratamento, com as mesmas características observáveis, o único fator que diferencia os resultados desses indivíduos é a participação ou não no programa.

A hipótese principal desse método, que tem como objetivo estimar o efeito médio do tratamento sobre os tratados (*EMPT*) e o efeito médio do tratamento (*EMP*), é que o vetor de variáveis observáveis X_i contém todas as informações sobre o resultado potencial na ausência do tratamento ($Y_i(0)$) e sobre o resultado potencial sob o tratamento ($Y_i(1)$) que o indivíduo possui ao tomar a decisão de participar ou não do tratamento.

Ao levar em consideração essas características observáveis, o avaliador controla por todas as variáveis que estão relacionadas aos resultados potenciais e que também afetam a decisão do indivíduo em participar ou não. Assim, ao controlar pelo vetor X_i , as variáveis $Y_i(0)$ e $Y_i(1)$ tornam-se independentes da variável binária que indica tratamento (T_i). Essa hipótese é conhecida como seleção nos observáveis, ou ignorabilidade, ou ainda não confundimento, e pode ser formalizada como:

$$Y_i(0) \perp T_i | X_i \quad (\text{H1})$$

para o caso em que estamos interessados no *EMPT*, e como:

$$(Y_i(0), Y_i(1)) \perp T_i | X_i \quad (\text{H1}')$$

para o caso em que estamos interessados no *EMP*.

A implicação dessa hipótese é que o resultado de um indivíduo no grupo de controle é um bom previsor do resultado potencial na ausência de tratamento de um indivíduo no grupo de tratamento

1. Existe extensa literatura empírica e teórica sobre o método de pareamento. Ver Rosenbaum (1995, 2002), Rubin (1973, 1979), Heckman, Ichimura e Todd (1998), Abadie e Imbens (2002), Lalonde (1986) e Dehejia e Wahba (1999).

que possui o mesmo vetor de variáveis observáveis (X_i), e que o resultado de um indivíduo no grupo de tratamento é um bom preditor do resultado potencial para um indivíduo controle com as mesmas características X_i caso ele fosse tratado.² Dessa forma, para estimar o *EMPT* via pareamento, precisamos que cada indivíduo no grupo de tratamento tenha um par no grupo de controle, cujo resultado reproduz o que seria o resultado desse indivíduo na ausência de tratamento. Dessa forma, precisamos que a região do vetor X_i que engloba as características dos indivíduos tratados também represente as características dos indivíduos que estão no grupo de não tratados. Assim, a segunda hipótese do método do pareamento é a hipótese de sobreposição e pode ser formalizada como:

$$Pr[T_i=1|X_i] < 1 \quad (\text{H2})$$

Da mesma forma, para estimar o *EMP* via pareamento, precisamos que, para cada indivíduo no grupo de tratamento, haja um par no grupo de controle, cujo resultado potencial seja um bom previsor do resultado desse indivíduo na ausência do tratamento. E também precisamos que, para cada indivíduo no grupo de controle, tenhamos um par no grupo de tratamento, cujo resultado potencial seja um bom previsor do resultado desse indivíduo, caso ele fosse tratado. Nesse caso, a hipótese de sobreposição deve ser expandida para:

$$0 < Pr[T_i=1|X_i] < 1 \quad (\text{H2}')$$

Essa hipótese garante que as características dos tratados sejam representadas no grupo de controle e vice-versa.

Uma das críticas aos métodos baseados nas duas hipóteses acima é que, ao tomar a decisão de participar ou não, os indivíduos olham não só para um vetor de características observáveis, mas também para características que não são observadas pelo avaliador. Se essas características não observáveis forem correlacionadas tanto com a decisão de participar do programa como com os

2. Essa hipótese de hipótese de independência é mais forte do que a hipótese mínima necessária para identificar o efeito médio do tratamento sobre os tratados. A hipótese necessária seria a independência na média, isto é, $E[Y_i(0)|T_i, X_i] = E[Y_i(0)|T_i=0, X_i] = E[Y_i(0)|X_i]$ no caso de H1.

resultados potenciais desse indivíduo, o método do pareamento não conseguirá eliminar o viés de seleção, e as estimações do *EMPT* e do *EMP* estarão sujeitas a um viés.

Sob as hipóteses H1 e H2, conseguimos identificar o efeito médio de tratamento sobre os tratados, e sob as hipóteses H1' e H2', o efeito médio de tratamento. O efeito médio do tratamento sobre os tratados para a subpopulação com características observáveis $X_i = x$ pode ser escrito como:

$$EMPT(x) = E[Y_i(1)|T_i=1, X_i=x] - E[Y_i(0)|T_i=1, X_i=x] \quad (5.1)$$

em que $E[Y_i(1)|T_i=1, X_i=x]$ é a média populacional de Y_i para os tratados com uma determinada combinação de características X_i , e $E[Y_i(0)|T_i=1, X_i=x]$ é a média de Y_i que os tratados com essas características teriam caso não tivessem recebido o tratamento.

De outra forma, o efeito médio do tratamento sobre os tratados para os indivíduos com certa combinação de características de X_i seria a diferença entre a média populacional do resultado para os indivíduos tratados e a média populacional do resultado de interesse para esses indivíduos tratados, caso eles não recebessem o tratamento. Na prática, podemos estimar a primeira média condicional em (5.1), usando a média amostral dos resultados obtidos pelos indivíduos tratados. No entanto, não conseguimos observar os resultados desses indivíduos tratados, caso eles não tivessem recebido tratamento, e não podemos simplesmente estimar a segunda média condicional usando uma média amostral.

Já o efeito médio do tratamento para a subpopulação com características observáveis $X_i = x$ pode ser escrito como:

$$EMP(x) = E[Y_i(1)|X_i=x] - E[Y_i(0)|X_i=x] \quad (5.2)$$

em que $E[Y_i(1)|X_i=x]$ é a média populacional do resultado potencial, caso fosse tratado para todos os indivíduos, incluindo tratados e controles, com uma determinada combinação de características X_i , e $E[Y_i(0)|X_i=x]$ é a média populacional do resultado potencial, caso não fosse tratado para todos os indivíduos com uma determinada combinação de X_i . Note que nesse caso, não podemos estimar a primeira média, usando somente os resultados dos tratados, e a segunda média, usando somente os resultados dos controles, pois queremos o efeito médio para todos os indivíduos tratados de

controles que possuem características e comportamentos diferentes. Para calcular as médias, precisamos estimar qual seria o resultado para os indivíduos tratados, caso eles não fossem tratados, e os resultados potenciais para os controles, caso eles fossem tratados.

No entanto, sob as hipóteses H1 e H2, conseguimos recuperar o segundo termo da equação (5.1). Sob a hipótese H1,

$$E[Y_i(0)|T_i=1, X_i=x] = E[Y_i(0)|X_i=x] = E[Y_i(0)|T_i=0, X_i=x] = E[Y_i|T_i=0, X_i=x] \quad (5.3)$$

Essa igualdade indica que a esperança dos resultados sem tratamento dos tratados que apresentem determinada combinação de características presentes em X_i é igual à esperança dos resultados sem tratamento dos não tratados (grupo de controle) com essas mesmas características. E essa esperança para os tratados pode ser estimada, usando os dados que são observáveis pelo pesquisador.

E, portanto, podemos escrever:

$$EMPT(x) = E[Y_i|T_i=1, X_i=x] - E[Y_i|T_i=0, X_i=x] \quad (5.4)$$

Esse resultado junto à hipótese H2 indica que podemos estimar o segundo termo da equação (5.1), usando a média dos resultados de interesse para os indivíduos no grupo de controle, que possuem a mesma combinação de características X_i que os indivíduos no grupo de tratamento.

Assim, podemos calcular o efeito do tratamento para os indivíduos tratados, com determinada combinação de características X_i , comparando a média observada de Y_i para os tratados com essas características com a média observada de Y_i para os indivíduos do grupo de controle com essas mesmas características.

Para obter o efeito médio do tratamento sobre os tratados, precisamos calcular o valor esperado (expectativa) dessa diferença de médias na distribuição de X_i , condicional a $T_i=1$. Por exemplo, no caso em que X_i inclui apenas uma variável explicativa discreta:

$$EMPT = \sum_x EMPT(x) \cdot \Pr[X_i = x | T_i = 1]$$

Nesse caso, $EMPT$ é a média ponderada dos efeitos médios para os indivíduos com determinada combinação de características X_i ($EMPT(x)$), na qual os pesos são dados pela porcentagem de indivíduos tratados que possuem essa combinação de características X_i .

Sob as hipóteses H1' e H2', podemos identificar o EMP da mesma forma que fizemos com o $EMPT$. Nesse caso, vamos escrever o nosso parâmetro como função das esperanças dos resultados potenciais para os grupos de tratados e de controle. Para isso, vamos definir $p=Pr[T_i=1]$, isto é, p é igual a probabilidade de ser tratado. Como T_i é uma variável aleatória binária, $1-p=Pr[T_i=0]$. Usando essa definição, podemos escrever o nosso parâmetro $EMP(x)$ como:

$$EMP(x) = E[Y_i(1)|T_i=1, X_i=x].p + E[Y_i(1)|T_i=0, X_i=x].(1-p) - \\ E[Y_i(0)|T_i=1, X_i=x].p - E[Y_i(0)|T_i=0, X_i=x].(1-p)$$

Note que conseguimos estimar o primeiro e o último termo dessa equação usando o resultado observado para os tratados e para os controles, respectivamente. No entanto, para recuperarmos os termos do meio, teremos que usar a hipótese H1', usando uma derivação análoga a equação 5.3 para o grupo de tratamento:

$$E[Y_i(1)|T_i=0, X_i=x] = E[Y_i(1)|X_i=x] = E[Y_i(1)|T_i=1, X_i=x] = E[Y_i|T_i=1, X_i=x] \quad (5.5)$$

Da mesma forma que anteriormente, a esperança dos resultados potenciais com tratamento para os controles que apresentam determinada combinação de características observáveis X_i é igual à esperança dos resultados potenciais sob tratamento para os tratados com a mesma combinação de características.

Usando os resultados das igualdades (5.3) e (5.5), podemos escrever o efeito médio do tratamento como:

$$EMP = E[EMP(x)]$$

$$= E[E[Y_i|T_i=1, X_i=x]] - E[E[Y_i|T_i=0, X_i=x]]$$

Note que para obter o efeito médio do tratamento, precisamos calcular a expectativa de $EMP(x)$

sobre a distribuição de X_i . No caso em que X_i inclui apenas uma variável explicativa discreta,

$$EMP = \sum_x EMP(x) \cdot \Pr[X_i = x]$$

5.1. Pareamento simples

Para cada indivíduo no grupo de tratamento, o estimador de pareamento busca os indivíduos no grupo de controle mais próximos (em termos do seu vetor de variáveis observáveis) e usa os resultados desses indivíduos para obter o que seria o resultado do indivíduo no grupo de tratamento, caso ele não fosse tratado (contrafactual). Se considerarmos o efeito médio do tratamento, além de estimar o contrafactual para o grupo de tratamento, o método do pareamento também busca os indivíduos no grupo de tratados mais próximos em termos de variáveis observáveis a certo indivíduo no grupo de controle, e usa os resultados desses indivíduos como o que aconteceria com o controle caso ele fosse tratado.

As principais diferenças entre os vários estimadores de pareamento dizem respeito à métrica usada para definir os indivíduos mais “próximos” dos tratados (ou dos controles) em termos do vetor de variáveis X_i . Além disso, os métodos também diferem em relação a quantos indivíduos do grupo de não tratados (ou de tratados) serão relacionados a cada indivíduo no grupo de tratamento (ou de controle) para obter o seu contrafactual.

Vamos considerar a estimação do $EMPT$ e pensar em como estimar o contrafactual para os indivíduos tratados. Suponha um pareamento em que mais de um indivíduo no grupo de controle possa ser relacionado com o indivíduo no grupo de tratados, e que o mesmo indivíduo no grupo de controle possa ser relacionado a vários indivíduos no grupo de tratados. Usando uma medida de distância,³ que nos diz quais são os indivíduos no grupo de controle que possuem características

3. Um exemplo de medida de distância é a métrica definida em Imbens e Wooldridge (2007) para determinar quanto diferentes os indivíduos são em relação ao vetor X_i , $\sum_{j|T_j=0, T_i=1} 1\{\|X_j - X_i\| \leq \|X_l - X_i\|\} = M$ na qual 1 é uma função

indicadora que é igual a 1 se a expressão entre chaves é verdadeira, e 0 caso contrário. Nesta métrica, $l \equiv l_m(i)$ é o índice do indivíduo no grupo de não-tratados com a distância M mais próxima do indivíduo i no grupo de tratados em termos a métrica $\|\cdot\|$. Por exemplo, $l_1(i)$ representa o par a distância 1 do indivíduo 1, isto é, o indivíduo mais próximo do indivíduo i .

observáveis mais próximas do indivíduo tratado i , identificamos os M pares mais próximos desse indivíduo. De modo mais formal, definimos como L_M o conjunto de índices com M pares mais próximos do indivíduo i , $L_M(i) = \{l_1(i), \dots, l_M(i)\}$, e o análogo amostral dos resultados potenciais, como:

$$Y_i(1) = Y_i \text{ e } \hat{Y}_i(0) = \frac{1}{M} \sum_{j \in L_M(i)} Y_j \quad (5.1.1)$$

Ou seja, para um indivíduo i no grupo de tratamento, o resultado observado para ele (Y_i) representa o resultado sob o tratamento, e a média dos resultados potenciais dos seus M pares no grupo de controle representa o resultado para esse indivíduo i caso ele não tivesse sido tratado.

Usando esses resultados estimados e observados para o indivíduo i , o estimador simples de pareamento para o efeito médio do tratamento sobre os tratados considerando os M vizinhos mais próximos é:

$$\widehat{EMPT}_M = \frac{1}{N_1} \sum_{i=1}^N T_i (Y_i - \hat{Y}_i(0)) \quad (5.1.2)$$

no qual N_1 é o número de indivíduos no grupo de tratamento e a multiplicação por T_i indica que só estamos considerando o efeito sobre os tratados.

Se estivermos interessados no efeito médio do tratamento, teríamos que estimar o contrafactual para cada indivíduo no grupo de controle. Nesse caso, para cada indivíduo i no grupo de controle, escolhemos os M indivíduos no grupo de tratamento com as características mais próximas dele, e o análogo amostral dos resultados potenciais para o indivíduo i no grupo de controle é:

$$\hat{Y}_i(1) = \frac{1}{M} \sum_{j \in L_M(i)} Y_j \text{ e } Y_i(0) = Y_i \quad (5.1.3)$$

Ou seja, no caso do indivíduo i no grupo de controle a média dos resultados dos indivíduos tratados mais próximos a ele representa o que aconteceria com esse indivíduo caso ele tivesse recebido o tratamento, e o resultado do próprio indivíduo representa o que aconteceria com esse indivíduo sem o tratamento.

Usando os análogos amostrais em (5.1.1) e em (5.1.3), o estimador do efeito médio do tratamento usando o método dos M vizinhos mais próximos é:

$$\widehat{EMP}_M = \frac{1}{N} \sum_{i=1}^N \left(T_i \cdot Y_i - (1 - T_i) \cdot \hat{Y}_i(1) \right) - \frac{1}{N} \sum_{i=1}^N \left(T_i \cdot \hat{Y}_i(0) - (1 - T_i) \cdot Y_i \right) \quad (5.1.4)$$

no qual N é o número total de indivíduos na amostra, incluindo tratados e controles.

O estimador de pareamento é muito sensível à escolha de variáveis presentes em X_i . Se o vetor X_i incluir número grande de variáveis explicativas, a hipótese de sobreposição (H2 e H2') pode não ser verdadeira, pois a região de X_i a qual precisamos de sobreposição das características entre o grupo de controle e de tratamento é muito grande. Se incluirmos número limitado de variáveis no vetor X_i , a hipótese de seleção nos observáveis (H1 e H1') pode ser violada.

O vetor X_i deve incluir as variáveis observáveis que influenciaram o indivíduo no momento da decisão de participar e são simultaneamente correlacionadas com os resultados potenciais dos indivíduos. Esse conjunto de variáveis pode garantir que a hipótese de seleção nos observáveis se verifica, mas não necessariamente que teremos sobreposição. A sobreposição pode ocorrer somente em um pedaço da região de X_i . Nesse caso, o pareamento estará olhando para outro parâmetro, que seria o efeito médio do tratamento sobre os tratados naquela região de suporte comum.⁴

Como vimos no capítulo 3, se os indivíduos fossem alocados de forma aleatória entre os grupos de tratamento e de controle, como no experimento natural, na média eles deveriam ter características bem parecidas. O método do pareamento assume que ao controlar por todas as características que afetam simultaneamente a decisão de participar no programa e os resultados potenciais, compararmos indivíduos semelhantes, exceto pelo fato de que um recebeu o tratamento, enquanto o outro não sofreu influência do programa. Para verificar se o pareamento está de fato comparando indivíduos que possuem um vetor X_i de características semelhantes, podemos realizar um teste de comparação de médias. Se fizermos um teste de comparação de médias das variáveis observáveis entre os grupos de controle e de tratamento antes do pareamento, devemos notar que na média eles são diferentes em algumas características, pois eles não foram selecionados de forma aleatória

4. Ver Crump, Hotz, Imbens e Mitnik (2009, 2008).

entre os dois grupos. No entanto, após o pareamento, se fizermos o mesmo teste de comparação de médias, o grupo de controle e o grupo de tratamento usados para estimar o efeito médio do tratamento devem ser semelhantes nas características observáveis, pois escolhemos para cada indivíduo no grupo de tratamento os seus pares com características bastante parecidas.

5.2. Pareamento com escore de propensão

A implementação do estimador de pareamento torna-se cada vez mais difícil quanto maior for a dimensão do vetor X_i . Quando acrescentamos variáveis ao vetor X_i , a dimensão da região de X_i cresce exponencialmente, e fica cada vez mais difícil encontrar para cada indivíduo no grupo de tratamento sua contraparte no grupo de controle com todas as características observáveis semelhantes. Rosenbaum and Rubin (1983) sugeriram que, ao invés de parearmos os indivíduos com base em todo o vetor X_i , podemos parear os indivíduos usando uma função de X_i , que resume toda a informação contida nesse vetor. Essa função é a probabilidade de receber o tratamento, dado o conjunto de características X_i , e é denominada de escore de propensão. De maneira formal, o escore de propensão é definido como:

$$P(X_i) = \Pr[T_i=1|X_i]$$

Como o escore de propensão contém a mesma informação contida no vetor de características observáveis X_i , se a hipótese de seleção nos observáveis for válida, também será válida a independência entre o resultado potencial na ausência de tratamento e a decisão de participar ou não, quando condicionarmos no escore de propensão.⁵ De modo mais formal,

$$(Y_i(0), Y_i(1)) \perp\!\!\! \perp T_i | X_i \Rightarrow (Y_i(0), Y_i(1)) \perp\!\!\! \perp T_i | p(X_i)$$

Esse resultado nos diz que, se as hipóteses H1 e H2 (ou H1' e H2') se verificarem e se o escore de propensão for conhecido, poderemos estimar o efeito médio do tratamento sobre os tratados fazendo o pareamento entre indivíduos tratados e indivíduos não tratados com base somente no

5. Para uma prova formal desse resultado, ver Rosenbaum and Rubin (1983).

escore de propensão. Na prática, não conhecemos o escore de propensão e precisaremos estimá-lo. A forma mais comumente usada para estimar o escore de propensão é usar os procedimentos paramétricos para estimação de probabilidade, por exemplo, os modelos probit ou logit. No caso do modelo logit, assumimos que a probabilidade de participar ou não segue o seguinte modelo:

$$\Pr[T_i = 1 | X_i = x] = \frac{\exp(x\beta)}{1 + \exp(x\beta)}$$

Em que β é vetor de parâmetros que será estimado no primeiro estágio. Seja $\hat{\beta}$ o estimador do parâmetro β . Nesse caso, o escore de propensão é estimado como:

$$\hat{P}(x) = \frac{\exp(x\hat{\beta})}{1 + \exp(x\hat{\beta})}$$

Como no caso do pareamento baseado no X_i , o pareamento baseado no escore de propensão também irá depender de uma métrica predeterminada, que definirá a proximidade do escore de propensão dos indivíduos tratados em relação ao escore de propensão dos indivíduos não tratados. Um dos estimadores mais usados é o estimador de pareamento do vizinho mais próximo ("The Nearest Neighbor Matching") que usa os resultados dos M indivíduos no grupo de não tratados que possuem escores de propensão mais próximos ao escore de propensão do indivíduo tratado i para estimar qual seria o resultado do indivíduo i , caso ele não recebesse tratamento.⁶⁷

Formalmente, sendo H_M o conjunto das M observações com o menor valor de $|\hat{P}(X_j) - \hat{P}(X_i)|$, podemos construir o análogo amostral para o resultado potencial do indivíduo, caso ele não fosse tratado, como:

6. Outros estimadores de pareamento baseado no escore de propensão são pareamento baseado em Kernel e o pareamento baseado em polinômio local. Nesse caso, para construir o contrafactual para o indivíduo no grupo de tratamento, usa-se os indivíduos no grupo de não tratados que estão em uma vizinhança em torno do escore de propensão do indivíduo i . Além disso, cada indivíduo não tratado pode receber peso diferente ao calcular o contrafactual do indivíduo i .

7. Um exemplo é usar a mesma medida de distância que usamos no caso do pareamento baseado no vetor X ,

$\sum_{j|T_j=0, T_i=1} 1\{|\hat{P}(X_j) - \hat{P}(X_i)| \leq |\hat{P}(X_h) - \hat{P}(X_i)|\} = M$. Nesse caso, $h \equiv h_m(i)$ é o índice do indivíduo no grupo de não tratados, cujo escore de propensão está à distância m do escore de propensão do indivíduo tratado i .

$$\hat{Y}_i(0) = \frac{1}{M} \sum_{j \in H_M(i)} Y_j \quad (5.2.1)$$

e o efeito médio do tratamento sobre os tratados estimado como na equação (5.1.2), usando essa definição para $\hat{Y}_i(0)$. Para implementar esse estimador na prática, primeiro usamos a média de Y dos M indivíduos mais próximos de cada participante do tratamento (em termos do escore de propensão) para calcular o contrafactual de cada indivíduo tratado ($Y_i(0)$) e depois calculamos a diferença de médias entre os tratados e seus contrafatuais, como em (5.1.2).

Para estimar o efeito médio do tratamento, estimamos também o contrafactual para o grupo de controle baseado no escore de propensão e calculamos a média usando ambos os contrafatuais, como na equação (5.1.4).

É importante ressaltar que as propriedades do estimador de pareamento baseado no escore de propensão também dependem da escolha do vetor X_i . É necessário escolher o vetor X_i que afete simultaneamente a decisão de participar ou não do tratamento e os resultados potenciais. Além disso, o estimador de pareamento baseado no escore de propensão também é sensível à especificação do escore de propensão, isto é, a forma funcional que vai ser escolhida na estimação do escore de propensão. É necessário que essa forma funcional seja flexível o suficiente para que a hipótese de sobreposição (H2 ou H2') se verifique. Rosenbaum e Rubin (1995) propuseram uma forma simples de buscar evidência se estivermos violando a hipótese de sobreposição. Eles sugeriram uma comparação simples das médias de cada uma das variáveis observáveis no grupo de tratamento e no grupo de seus pares. Se a diferença dessas médias for muito grande, o teste de diferença de médias irá rejeitar a hipótese de que os grupos de tratamento e o grupo de pares são balanceados em relação a essas variáveis observáveis. Nesse caso, provavelmente o escore de propensão não está sendo flexível o suficiente para garantir que a distribuição dos observáveis seja a mesma no grupo de tratamento e no grupo de pares.

Existem diversos métodos que usam o escore de propensão para ajustar as diferenças entre os indivíduos nos grupos de controle e de tratamento, sempre com base nas hipóteses de seleção nos observáveis e de sobreposição. Na próxima seção, iremos apresentar dois métodos bastante usados: o método de imputação baseado em uma regressão linear e o método de responderação.

5.3. Regressão linear

O método de regressão linear também pode ser visto como um estimador de pareamento⁸ e também se baseia na hipótese de seleção nos observáveis. Usando as definições do capítulo 2 para o caso em que consideramos que o programa tem o mesmo efeito em todos os indivíduos, podemos escrever os resultados potenciais, como:

$$Y_i(1)=X_i'\alpha_1+\beta+\varepsilon_{1i} \quad (5.3.1)$$

$$Y_i(0)=X_i'\alpha_0+\varepsilon_{0i} \quad (5.3.2)$$

Uma maneira de pensar no método de regressão linear é como usar o modelo para o resultado potencial caso não tivesse sido tratado para estimar o contrafactual para os tratados, e como usar o modelo do resultado potencial caso tivesse sido tratado para estimar o contrafactual para os controles.

Sob a hipótese de seleção nos observáveis, as igualdades (5.3) e (5.5) são válidas:

$$E[Y_i(0)|T_i=1, X_i=x]=E[Y_i|T_i=0, X_i=x]$$

$$E[Y_i(1)|T_i=0, X_i=x]=E[Y_i|T_i=1, X_i=x]$$

E assumindo exogeneidade do vetor X_i , i.e., $E[\varepsilon_{0i}|X_i]=0$ e $E[\varepsilon_{1i}|X_i]=0$, temos os seguintes modelos para as esperanças dos resultados potenciais:

$$E[Y_i|T_i=1, X_i=x]=X_i'\alpha_1+\beta, \text{ e } E[Y_i|T_i=0, X_i=x]=X_i'\alpha_0 \quad (5.3.3)$$

E, nesse caso, podemos estimar o efeito médio do tratamento e o efeito médio do tratamento sobre os tratados estimando os modelos de regressão acima para as subamostras de controle e tratamento.

8. Rubin (1977) usa o método de regressão linear para estimar o efeito médio do tratamento.

Para estimar o contrafactual para os indivíduos tratados, iremos estimar uma regressão linear de Y_i em X_i para a amostra de indivíduos que não foram tratados. Os coeficientes estimados nessa regressão simples para a subamostra do grupo de controle são então usados para prever qual teria sido o resultado de cada indivíduo do grupo de tratamento caso ele não tivesse recebido o tratamento. Nesse caso, o estimador do efeito médio do tratamento sobre os tratados seria:

$$E\overline{MPT}_R = \frac{1}{N_1} \sum_{i=1}^N T_i(Y_i - \hat{\mu}_0(X_i)) \quad (5.3.3)$$

em que $\hat{\mu}_0(X_i)$ é o valor previsto para o grupo de indivíduos no grupo de tratamento usando os coeficientes estimadores pela regressão linear simples no grupo de controle.

Os estimadores baseados em regressão dependem fortemente da extração dos valores $Y_i(0)$ para valores fora do intervalo dos usados na regressão. Os coeficientes estimados nessa regressão são usados para encontrar os valores preditos do resultado para os indivíduos tratados, caso eles não recebessem tratamento. Para essa extração, necessitamos de sobreposição na região de X_i do grupo de controle e no grupo de tratamento, isto é, precisamos da nossa hipótese H2. Por exemplo, se estimássemos uma regressão linear simples com intercepto, o valor médio previsto para os indivíduos no grupo tratamento seria:

$$\hat{\mu}_0(X_i) = \bar{Y}_0 + (\bar{X}_1 - \bar{X}_0)' \hat{\alpha}_0$$

em que \bar{Y}_0 é o valor médio do resultado para o grupo de controle, \bar{X}_1 é o valor médio das variáveis explicativas no grupo de tratamento e \bar{X}_0 é o valor médio das variáveis explicativas no grupo de controle. Se \bar{X}_1 é muito diferente de \bar{X}_0 , os valores previstos usando a regressão linear serão muito sensíveis à especificação dessa regressão. Podemos usar sempre uma forma funcional flexível para a regressão, mas não teremos garantias de que esse modelo funciona bem nas regiões onde não há sobreposição da região de X_i no grupo de controle e no grupo de tratamento.

Para estimar o efeito médio do tratamento, precisamos também estimar a regressão linear simples para os indivíduos tratados e usar os resultados dessa regressão para prever qual seria o resultado para os indivíduos no grupo de controle, caso eles recebessem o tratamento. Nesse caso, o estimador do efeito médio do tratamento seria:

$$\widehat{EMP}_R = \frac{1}{N} \sum_{i=1}^N T_i(Y_i - \hat{\mu}_0(X_i)) + (1 - T_i)(Y_i - \hat{\mu}_1(X_i))$$

em que $\hat{\mu}_1(X_i)$ é o valor previsto para os indivíduos no grupo de não tratados usando os resultados da regressão estimada para os indivíduos que receberam o tratamento.

Um método alternativo para obter o efeito médio do tratamento sobre os tratados por meio de regressão linear é usar o escore de propensão como variável explicativa.⁹ Nesse caso, iremos estimar uma regressão linear simples de Y_i em $P(X_i)$ para a subamostra de indivíduos não tratados e usar os valores estimados nessa regressão para prever os valores dos resultados dos indivíduos tratados, caso eles não recebessem tratamento.

Pelos modelos representados em (5.3.1) e (5.3.2), o efeito médio do tratamento e o efeito médio do tratamento sobre os tratados são:

$$EMP = \beta + E[X_i]'(\alpha_1 - \alpha_0)$$

$$EMPT = \beta + E[X_i|T_i = 1]'(\alpha_1 - \alpha_0)$$

Além disso, podemos representar os modelos em (5.3.1) e (5.3.2) em formas alternativas,

$$Y_i = \gamma_0 + \tau T_i + X'_i \gamma_1 + (X_i - \bar{X})T_i \gamma_2 + \vartheta_i \quad (5.3.4)$$

em que τ representa o efeito médio do tratamento, \bar{X} é a média de X_i em toda amostra e $E[\vartheta_i|X_i] = 0$. E podemos estimar esse efeito pela regressão linear de Y_i em X_i , T_i e $(X_i - \bar{X})T_i$.¹⁰

O efeito médio do tratamento sobre os tratados pode ser obtido pelo seguinte modelo:

9. Esse método foi sugerido por Rosenbaum and Rubin (1983).

10. Note que, se $\alpha_1 = \alpha_0$, o efeito médio do tratamento é igual a β e pode ser estimado pela regressão de Y_i em X_i e T_i .

$$Y_i = \gamma_0 + \delta T_i + X'_i \gamma_1 + (X_i - \bar{X}_1) T_i \gamma_2 + \epsilon_i \quad (5.3.5)$$

em que δ representa o efeito médio do tratamento sobre os tratados, \bar{X}_1 é a média de X_i na subamostra de tratados e $E[\epsilon_i | X_i] = 0$.

É importante salientar que o método de regressão pode ser entendido como um método de pareamento, no qual todos os indivíduos no grupo de controle tem pesos iguais na construção do contrafactual para o indivíduo tratado, e da mesma forma todos os indivíduos no grupo de tratado têm pesos iguais na construção do resultado sob tratamento para o indivíduo no grupo de controle. Enquanto, nos métodos de pareamento apresentados anteriormente, os indivíduos mais próximos referentes de características observáveis (X_i) recebiam peso maior.

5.4. Métodos de reponderação

O escore de propensão também pode ser usado como um peso¹¹ que, ao balancear os indivíduos no grupo de controle, torna esse grupo semelhante na média ao grupo de indivíduos tratados. E ao balancear os indivíduos no grupo de tratamento os torna mais semelhantes aos indivíduos no grupo de controle. Nesse caso, ponderamos cada unidade no grupo de controle pelo inverso da probabilidade de não receber tratamento, e cada unidade no grupo de tratamento pelo inverso da probabilidade de receber tratamento. Quando maior for a probabilidade do indivíduo no grupo de controle de não receber o tratamento, menor será o seu peso ao balancearmos o grupo de controle. Da mesma forma, quanto maior for a probabilidade de receber tratamento, menor será o peso do indivíduo tratado ao balancearmos o grupo de tratamento.

Usando a Lei de Bayes, podemos definir uma relação entre a distribuição de X_i na população de tratados e a distribuição de X_i na população de não tratados. Pela Lei de Bayes, a distribuição de X_i condicional a $T_i=1$ pode ser escrita, como:

$$f(X_i | T_i = 1) = \frac{f(X_i, T_i = 1)}{\Pr(T_i = 1)} = \frac{\Pr(T_i = 1 | X_i) \cdot f(X_i)}{\Pr(T_i = 1)}$$

11. Para uma revisão do método de reponderação, ver Imbens (2004) e Imbens and Wooldridge (2007).

e, de forma análoga, podemos escrever a distribuição condicional de X_i em $T_i=0$, como:

$$f(X_i|T=0) = \frac{f(X_i, T_i=0)}{\Pr(T_i=0)} = \frac{\Pr(T_i=0|X_i) \cdot f(X_i)}{\Pr(T_i=0)}$$

e dividindo as duas distribuições condicionais, temos

$$\frac{f(X_i|T_i=1)}{f(X_i|T_i=0)} = \frac{\Pr(T_i=1|X_i) \cdot \Pr(T_i=0)}{\Pr(T_i=0|X_i) \cdot \Pr(T_i=1)}$$

Usando a definição do escore de propensão e da probabilidade de ser tratado $\Pr(T_i=1) \equiv Q_0$, podemos reescrever a relação entre as distribuições condicionais de X_i ,

$$\frac{f(X_i|T_i=1)}{f(X_i|T_i=0)} = \frac{P(X_i) \cdot (1-Q_0)}{(1-P(X_i)) \cdot Q_0} \quad (5.4.1)$$

Usando essa relação e a hipótese de seleção nos observáveis, podemos escrever os parâmetros de interesse (**EMPT** e **EMP**) usando o escore de propensão como peso:

$$EMPT = E[Y_i(1)|T_i=1] - E[Y_i(0)|T_i=1]$$

$$= E[T_i Y_i | T_i=1] - E[E[(1-T_i)Y_i | T_i=1, X_i]]$$

$$= E\left[\frac{T_i}{Q_0} Y_i\right] - E\left[\frac{1-T_i}{Q_0} \cdot \frac{P(X_i)}{1-P(X_i)} Y_i\right]$$

e

$$EMP = E[Y_i(1)] - E[Y_i(0)] = E[E[Y_i|T_i=1, X_i]] - E[E[Y_i|T_i=0, X_i]]$$

$$= E[E[T_i Y_i | T_i=1, X_i]] - E[E[(1-T_i)Y_i | T_i=0, X_i]]$$

$$= E\left[\frac{T_i}{P(X_i)} Y_i\right] - E\left[\frac{1-T_i}{1-P(X_i)} \cdot Y_i\right]$$

Nas derivações acima, usando primeiramente a hipótese de seleção nos observáveis e a lei das expectativas iteradas, e para chegar à última igualdade usamos as fórmulas derivadas acima usando a Lei de Bayes.

Como no caso do estimador de pareamento, estimamos o escore de propensão usando o modelo paramétrico, como o logit (ou o probit). Dado que T_i é uma variável binária, o estimador de Q_0 será simplesmente a proporção de tratados, isto é, $\hat{Q}_0 = \frac{N_1}{N}$. Usando o análogo amostral para a equação acima, o estimador de responderação de D é

$$\widehat{EMPT}_{RP} = \frac{1}{N_1} \sum_{i=1}^N T_i Y_i - \frac{1}{N_1} \sum_{i=1}^N (1 - T_i) \left(\frac{\hat{P}(X_i)}{1 - \hat{P}(X_i)} \right) Y_i \dots \quad (5.4.2)$$

e

$$\widehat{EMPR}_{RP} = \frac{1}{N} \sum_{i=1}^N \frac{T_i Y_i}{\hat{P}(X_i)} - \frac{1}{N} \sum_{i=1}^N \left(\frac{(1 - T_i)}{1 - \hat{P}(X_i)} \right) Y_i$$

Esses estimadores se baseiam somente na estimação do escore de propensão, e uma das limitações desses estimadores é serem bastante sensíveis à especificação do escore de propensão. Além disso, quando valor do escore de propensão for próximo de um, esses estimadores podem assumir valores muito altos.

5.5. Combinação de métodos

Alguns estimadores surgiram da combinação de dois métodos descritos nas seções anteriores. Esses métodos são bastante atrativos na prática, por terem comportamentos menos sensíveis a certas hipóteses, como a especificação correta do modelo do escore de propensão ou a especificação correta da função de regressão que relaciona $Y_i(0)$ e $Y_i(1)$ com X_i . Nessa seção, descrevemos três combinações: regressão e pareamento; pareamento e diferenças em diferenças; e regressão e responderação.

5.5.1. Pareamento e regressão linear

Abadie e Imbens (2002) mostram que à medida que se aumenta a dimensão do vetor X_i , podemos estar diminuindo a variância do estimador de pareamento baseado em X_i , mas aumentamos o viés. Para cada indivíduo no grupo de tratamento, podemos definir um resíduo que seria a diferença entre o efeito de tratamento para aquele indivíduo, $\widehat{EMPT}_{iM} = Y_i(1) - \widehat{Y}_i(0)$. A variância do

estimador de pareamento é uma função da soma do quadrado desses resíduos. Ao aumentarmos o número de variáveis incluídas em X_i , diminuímos esse resíduo e consequentemente a variância do estimador. O viés surge porque o vetor de variáveis explicativas para o indivíduo i é próximo do vetor de variáveis explicativas dos seus M vizinhos mais próximos usados no pareamento, mas não é exatamente igual. No caso do pareamento com base em poucas variáveis discretas em que fosse possível parear os indivíduos no grupo de tratamento com indivíduos no grupo de controle com valores exatamente iguais para o vetor X_i , o viés desapareceria. Uma das maneiras de corrigir o viés do estimador de pareamento é ajustar pela diferença nos valores das variáveis explicativas X_i , usando o modelo de regressão linear. Nesse caso, usamos o mesmo conjunto $L_M(i)$ que engloba os M indivíduos no grupo de controle, cujos vetores X_j estão mais próximos do vetor X do indivíduo i , como na seção (5.1) acima. Além disso, estimamos a regressão linear que relaciona o resultado Y_i com X_i somente para a subamostra de indivíduos que não receberam tratamento, como na seção (5.2) acima. Com os coeficientes estimados na regressão, calculamos o valor previsto de Y_i para os indivíduos no grupo de controle e também para os indivíduos no grupo de tratamento e usamos esses valores preditos para ajustar o estimador de pareamento. Nesse caso, os valores do resultado para os indivíduos tratados, caso eles não recebem tratamento, são:

$$\tilde{Y}_i(0) = \frac{1}{M} \sum_{j \in H_M(i)} (Y_j + \hat{\mu}_0(X_i) - \hat{\mu}_0(X_j)) \quad (5.5.1.1)$$

e o estimador do efeito médio do tratamento sobre os tratados que combina pareamento e regressão é:

$$\widehat{EMPT}_{MR} = \frac{1}{N_1} \sum_{i=1}^N T_i (\tilde{Y}_i(1) - \tilde{Y}_i(0)) \dots \quad (5.5.1.2)$$

Podemos usar esse mesmo método para estimar o efeito médio do tratamento, para isso precisamos construir $\tilde{Y}_i(1)$, isto é, o resultado para os indivíduos não tratados, caso eles recebessem o tratamento usando uma fórmula análoga a (5.5.12) e recuperar o efeito médio do tratamento pela fórmula do estimador de pareamento.

5.5.2. Pareamento e o método de diferenças em diferenças

No caso de dados em painel, o método do pareamento pode ser combinado com o método de diferenças em diferenças.¹² Essa combinação de métodos faz com que algumas das hipóteses usadas em cada um dos métodos possam ser substituídas por hipóteses mais fracas. Por exemplo, a hipótese de seleção nos observáveis estabelece que, condicional ao vetor de variáveis observáveis X_i , não podem existir nenhum fator não observável que influencie simultaneamente a decisão de participar ou não no tratamento e os resultados potenciais. Ao combinar o pareamento com o método de diferenças em diferenças, podemos permitir que fatores não observáveis, que sejam constantes ao longo do tempo, influenciem simultaneamente a decisão de participar e os resultados potenciais, mesmo controlando pelo vetor de variáveis observáveis X_i , como vimos no capítulo 4.

Para aplicar essa combinação de métodos, precisamos observar os indivíduos no grupo de controle e no grupo de tratamento em pelo menos dois períodos no tempo, um anterior ao recebimento do tratamento e um período após o tratamento. Para simplificar, vamos considerar apenas dois períodos: t_0 , antes do tratamento e t_1 , após o tratamento. Esse método usa basicamente a ideia do método de diferenças em diferenças, isto é, ele compara a diferença do resultado do grupo de tratamento entre os períodos t_0 e t_1 com a diferença de resultados do grupo de controle no mesmo período e atribui essa diferença da diferença ao efeito do tratamento. Formalmente, o método que combina pareamento e o método de diferenças em diferenças estimam o seguinte parâmetro populacional:

$$EMPT_{t_0, t_1} = E[Y_{it_1}^1 - Y_{it_0}^0 | T_i = 1, p(X_i)] - E[Y_{it_1}^0 - Y_{it_0}^0 | T_i = 0, p(X_i)]$$

Em que Y_{it}^d representa o resultado no tempo t para o indivíduo i quando o seu *status* de tratamento é d . Para que esse parâmetro corresponda ao efeito médio do tratamento sobre os tratados, precisamos da seguinte hipótese:

$$E[Y_{it_1}^0 - Y_{it_0}^0 | T_i = 1, p(X_i)] = E[Y_{it_1}^0 - Y_{it_0}^0 | T_i = 0, p(X_i)] \quad (\text{H3})$$

12. Esse método foi proposto por Heckman, Ichimura e Todd (1995, 1997).

Essa hipótese é mais fraca do que a hipótese de seleção nos observáveis, pois ela continua válida mesmo que existam fatores não observáveis comuns aos dois períodos que afetem simultaneamente a decisão de participar e os resultados potenciais. No entanto, essa hipótese exige informação em pelo menos um período anterior à intervenção. Se a tendência na evolução dos resultados entre os períodos pré e pós-programa é a mesma no grupo de controle e no grupo de tratamento, essa hipótese é válida. O vetor X_i tem que incluir toda a informação disponível para o indivíduo quando ele decidiu participar ou não no programa, isto é, são características observáveis em períodos anteriores ao programa e que estão relacionadas à decisão de participação.

A hipótese H3 não é suficiente para identificarmos o efeito médio do tratamento sobre os tratados. Como no estimador de pareamento, precisamos também da hipótese de sobreposição da região de X_i para o grupo de controle e o grupo de tratamento. Se tivermos suporte comum, podemos estimar o efeito médio do tratamento sobre os tratados fazendo um pareamento baseado no escore de propensão¹³ e calculando a diferença das diferenças. Essa hipótese de suporte comum deve valer nos dois períodos de tempo, antes e depois do tratamento. Usando o estimador de pareamento baseado no vizinho mais próximo, definimos o conjunto H_M como o conjunto das M observações que possuem o escore de propensão mais próximo do indivíduo i , isto é, os M indivíduos tais que $\sum_{j|T_j=0,T_i=1} 1\{|\hat{P}(X_j) - \hat{P}(X_i)| \leq |\hat{P}(X_h) - \hat{P}(X_i)|\} = M$. Usando o pareamento, construímos o análogo amostral para a diferença do resultado potencial do indivíduo entre os períodos t_0 e t_1 , caso ele não fosse tratado:

$$\widehat{Y_{it_1}^0 - Y_{it_0}^0} = \frac{1}{M} \sum_{j \in H_M(i)} (Y_{jt_1} - Y_{jt_0}) \quad (5.5.2.1)$$

e o estimador de pareamento no escore de propensão para dados longitudinais é:

$$\widehat{EMPT}_{t_0,t_1} = \frac{1}{N_1} \sum_{i=1}^N T_i \left((Y_{it_1} - Y_{it_0}) - \frac{1}{M} \sum_{j \in H_M(i)} (Y_{jt_1} - Y_{jt_0}) \right) \quad (5.5.2.2)$$

Para implementar esse estimador na prática, primeiro calculamos para cada indivíduo no grupo de tratamento qual seria a diferença de resultados entre os períodos t_0 e t_1 caso ele não fosse tratado. Para isso, estimamos o escore de propensão para todos os indivíduos e selecionamos

13. Podemos também fazer o pareamento baseado no vetor X . Mas, neste caso, a hipótese H3 deve ser válida condicional em X , e não em $P(X)$.

os M indivíduos não tratados mais próximos de cada indivíduo tratado em relação ao escore de propensão e calculamos a média da diferença de resultados entre os dois períodos para os M indivíduos. No segundo passo, para cada indivíduo no grupo de tratamento, calculamos a diferença entre a sua diferença de resultado entre os períodos t_0 e t_1 , e a média dos M indivíduos no grupo de controle mais próximos a ele. O estimador do efeito médio do tratamento sobre os tratados é a média dessas diferenças das diferenças calculadas para os indivíduos tratados. Para estimar o escore de propensão usando dados em painel, precisamos usar a amostra de controles e de tratados. A variável dependente T_i é igual a 1 se o indivíduo recebeu o tratamento, e 0, caso contrário, e o vetor X_i engloba as informações que os indivíduos tinham ao decidir participar ou não no programa, isto é, características nos períodos anteriores ao programa. Nesse vetor X_i , podemos incluir variáveis que influenciaram a decisão do indivíduo entre participar ou não do programa, e não devemos incluir variáveis que possam ser resultado do programa. Ao final do capítulo, apresentamos um exemplo de como esse método pode ser empregado.

5.5.3. Regressão linear e reponderação

Como vimos acima, além do pareamento, podemos usar regressão linear ou reponderação para estimar o efeito médio do tratamento. No primeiro método, estimamos uma regressão linear para o grupo de controle e para o grupo de tratamento usando o vetor X_i como variáveis explicativas. Nessa primeira abordagem, não precisamos estimar o escore de propensão. Já o método de reponderação se baseia somente na estimativa do escore de propensão. Nesse segundo método, não precisamos estimar a função de regressão que relaciona os resultados potenciais com X_i . Sob algumas hipóteses, ambos os métodos resultam em estimadores consistentes e eficientes para o efeito médio do tratamento e para o efeito médio do tratamento sobre os tratados. No entanto, enquanto o comportamento do estimador de regressão é sensível à especificação usada para o modelo que relaciona os resultados potenciais com X_i , o estimador de reponderação é sensível ao modelo usado para o escore de propensão. Robins and Ritov (1997) propuseram um estimador que combina o método de regressão com o método de reponderação. A motivação para combinar métodos é que o estimador se torna menos sensível a apenas uma das hipóteses. Esses estimadores para o efeito médio do tratamento e para o efeito médio do tratamento sobre os tratados têm a propriedade de ser duplamente robustos, isto é, se o modelo paramétrico para o escore de propensão ou se o modelo paramétrico para a regressão linear forem corretamente especificados, o

estimador é consistente para o efeito médio e para o efeito médio do tratamento sobre os tratados, respectivamente.

Como vimos na seção (5.3), uma maneira de estimar o efeito médio do tratamento sobre os tratados usando uma regressão linear é estimar o seguinte modelo linear:

$$Y_i = \alpha_0 + \tau T_i + Z'_i \alpha_1 + T_i (X_i - \bar{X}_1)' \alpha_2 + \varepsilon_i \quad (5.5.3.1)$$

no qual \bar{X}_1 é a média amostral de X_i na subamostra de indivíduos tratados. Nesse caso, τ representa o efeito médio do tratamento sobre os tratados.

Essa regressão linear pode ser combinada com o método de reponderação, o qual ponderamos as unidades no grupo de controle por $\left(\frac{\hat{P}(X_i)}{1-\hat{P}(X_i)}\right)$. Ao final, estimamos uma regressão ponderada com o peso igual a 1 para um indivíduo no grupo de tratamento e igual a $\left(\frac{\hat{P}(X_i)}{1-\hat{P}(X_i)}\right)$ para uma unidade no grupo de controle. De modo mais formal, estimamos o modelo linear acima com a seguinte função-peso:

$$w(t, x) = t + (1-t) \left(\frac{\hat{P}(X_i)}{1-\hat{P}(X_i)} \right) \quad (5.5.3.2)$$

em que $\hat{P}(X_i)$ representa o estimador paramétrico do escore de propensão, obtido por uma regressão logit ou probit.¹⁴

Para obter o efeito médio do tratamento usamos uma regressão ponderada com duas modificações em relação ao modelo anterior. Como na seção (5.3), a regressão é estimada excluindo a média para toda a amostra:

$$Y_i = \alpha_0 + \tau T_i + Z'_i \alpha_1 + T_i (X_i - \bar{X})' \alpha_2 + \varepsilon_i$$

em que \bar{X} é a média amostral de X_i na amostra total, incluindo os indivíduos no grupo de tratamento. Temos que não só reponderar os indivíduos no grupo de controle para melhor

14. Esse método é mais geral, permitindo que a regressão linear inclua apenas um subvetor de variáveis explicativas, bem como o modelo para o escore de propensão. De outra maneira, o vetor de variáveis observáveis que determinam a participação no tratamento pode ser diferente do vetor de variáveis explicativas que determinam os resultados.

representar os indivíduos tratados, mas também os indivíduos no grupo de tratamento para melhor representar os indivíduos não tratados. Nesse caso, o peso para os indivíduos tratados será o inverso da probabilidade de receber tratamento e para os indivíduos não tratados será o inverso da probabilidade de não receber tratamento. Nesse caso, usamos a seguinte função-peso:

$$w_{EM}(t, x) = \frac{t}{\hat{P}(X_i)} + \frac{1-t}{1-\hat{P}(X_i)}$$

Esse método é amplamente usado por ser menos sensível a especificações do escore de propensão e da reta de regressão.

5.6. Exemplo 1: pareamento, regressão linear e reponderação

Mattos, Maia e Marques (2010) usam o pareamento baseado no escore de propensão para estimar o efeito de programas redistributivos incondicionais no número de horas trabalhadas de homens e mulheres. O Programa Bolsa Família engloba políticas de transferências condicionais e incondicionais de renda. Esse programa abrange a população em extrema pobreza e é elegível às famílias com renda *per capita* inferior a R\$ 50. Essas famílias podem receber a transferência sem nenhuma contrapartida. Além disso, existe um valor adicional para cada criança da família que for para a escola. Esse programa difere do antigo Bolsa Escola, em que as transferências ocorriam somente se os filhos frequentassem a escola. O objetivo do artigo é comparar a oferta de trabalho dos beneficiários de programas incondicionais como o Renda Mínima em relação a programas condicionais como o Bolsa Escola. A ideia seria que os beneficiários dos programas de transferência incondicionais não teriam incentivos para mudar suas horas de trabalho, pois eles teriam um aumento de renda sem nenhuma contrapartida. Já nos programas condicionais, como o Bolsa Escola, as crianças precisam ficar na escola; isso poderia reduzir as horas trabalhadas pelas crianças, o que, por sua vez, poderia afetar a oferta de trabalho dos adultos. Além disso, ao trabalhar mais horas, os adultos poderiam perder o direito ao benefício. Os autores usam o método do pareamento um a um baseado no escore de propensão, isto é, cada indivíduo no grupo de tratados é pareado com apenas um indivíduo no grupo de controle para estimar o efeito médio dos programas de transferências de renda nas horas trabalhadas dos beneficiários. Nesse exercício, usamos dados semelhantes aos usados em Mattos, Maia e Marques (2010) para estimar o efeito do

programa de Renda Mínima e do programa Bolsa Escola sobre o número de horas remuneradas de homens e mulheres, beneficiários do programa, com idade entre 12 e 65 anos.¹⁵

A tabela (5.6.1) apresenta os resultados da estimação desse efeito, usando alguns dos métodos descritos nesta seção. A parte superior da tabela apresenta os resultados para os homens, enquanto a parte inferior apresenta os resultados para as mulheres. Para os homens, o efeito do programa Bolsa Escola sobre o número de horas trabalhadas remuneradas ao ano é negativo e estatisticamente significativo em todos os métodos, exceto no método que usa o pareamento com o método do vizinho mais próximo (coluna 2). É interessante notar que os resultados estimados pelo pareamento com método do vizinho mais próximo, pela regressão e pela regressão com reponderação são bem semelhantes. No entanto, no caso do Programa de Renda Mínima, o efeito do programa sobre o número de horas remuneradas não é estatisticamente significativo, exceto para o caso em que combinamos o método da regressão com o método da reponderação (coluna 4). Nesse último caso, encontramos um efeito positivo sobre o número de horas trabalhadas remuneradas. Esse resultado contradiz a teoria de que programas de transferências incondicionais não levariam os indivíduos a trabalhar mais, pois eles teriam um aumento da renda sem contrapartida. No caso das mulheres, quando usamos o método do pareamento, encontramos que os efeitos de ambos os programas são não significativos. Quando usamos o método da regressão, encontramos que o efeito do programa de Renda Mínima é negativo e significativo a 5%, enquanto que o efeito do programa Bolsa Escola é não significativo. Os resultados encontrados com o método que combina regressão com ponderação indicam que ambos os programas tem efeito negativo e significativo no número de horas remuneradas das mulheres beneficiárias, sendo esse efeito maior em valor absoluto no caso do programa de Renda Mínima. Esse resultado indica que, no caso das mulheres, o programa leva à diminuição da oferta de trabalho.

Uma das hipóteses usadas por todos os métodos baseados no escore de propensão é a de que existe sobreposição da distribuição do escore de propensão para tratados e da distribuição do escore de propensão para os controles na região de X_i . O gráfico (5.6.2) mostra a sobreposição do escore de propensão para homens e mulheres beneficiários e não beneficiários para cada um dos programas. As barras azuis mostram o histograma para os indivíduos que não receberam

15. No artigo em Exercício, os autores usam como variável dependente o número de horas trabalhadas ao ano. Nesse exemplo, usamos como variável dependente o número de horas de trabalho remuneradas ao ano.

tratamento, enquanto as barras laranjas mostram o histograma para os indivíduos no grupo de tratamento. Logo, podemos verificar qual a porcentagem de indivíduos no grupo de controle e no grupo de tratamento para cada intervalo do escore de propensão. Se há sobreposição, em cada intervalo que tiver indivíduo tratado, temos que ter também indivíduos não tratados. Os gráficos indicam que há sobreposição das distribuições do escore de propensão tanto para os homens quanto para as mulheres.

Tabela 5.6.1: Métodos usando o escore de propensão

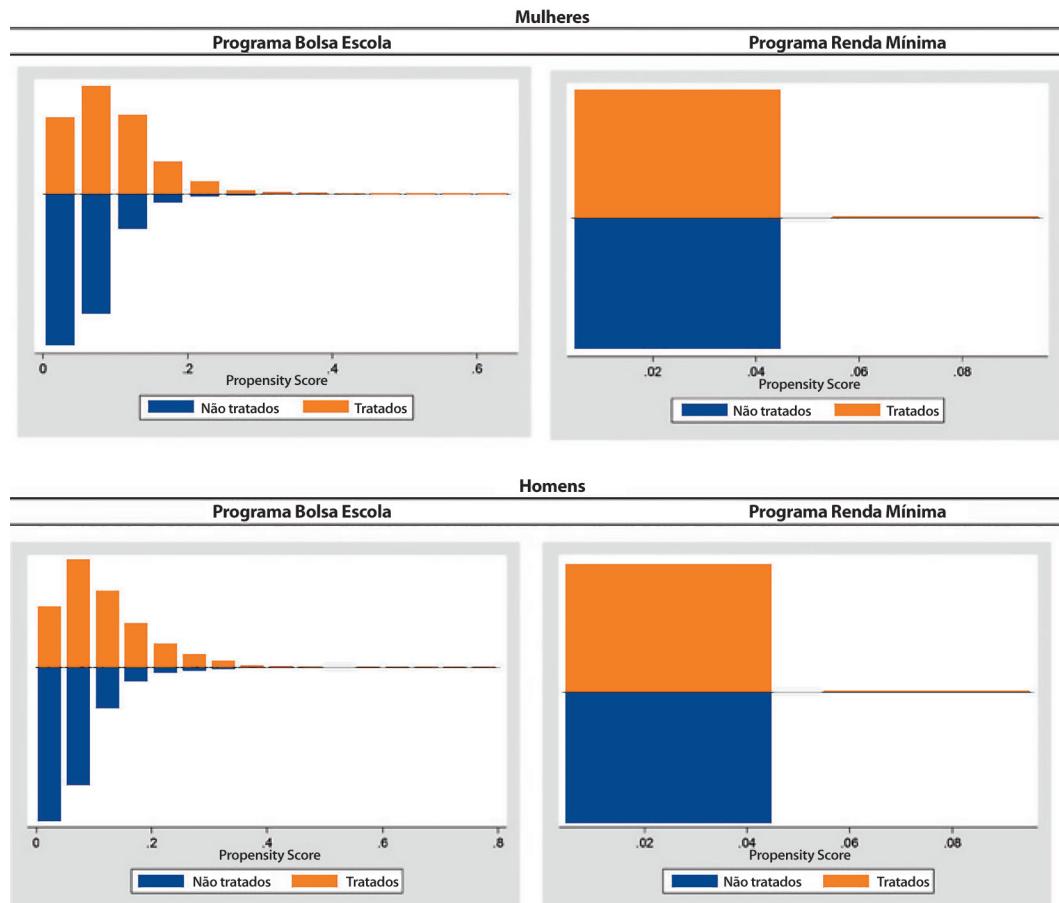
	Horas de Trabalho por Ano – Homens			
	Pareamento (um a um)	Pareamento (10 vizinhos)	Ressagem	Ressagem com Reponderação
Recebe Bolsa Escola	-0,044** (0,019)	-0,0232 (0,015)	-0,029** (0,012)	-0,027*** (0,007)
	0,054 (0,062)	0,067 (0,045)	0,043 (0,040)	0,070** (0,032)
Horas de Trabalho por Ano – Mulheres				
Recebe Bolsa Escola	Pareamento (um a um) -0,015 (0,029)	Pareamento (10 vizinhos) -0,028 (0,022)	Ressagem -0,029 (0,017)	Ressagem com Reponderação -0,029*** (0,010)
	-0,155 (0,13)	-0,136 (0,097)	-0,156** (0,065)	-0,083* (0,049)

Fonte: Elaboração própria dos dados da POF 2002-2003.

O vetor de variável X inclui escolaridade, participação no mercado de trabalho, ciclo de vida, composição familiar, raça, saúde, qualidade de moradia, controles estaduais.

O modelo é o mesmo usado em Mattos, Maia e Marques (2010).

Gráfico 5.6.2: Distribuição do escore de propensão para tratados e para não tratados



5.7. Exemplo 2: diferenças em diferenças com pareamento

Heckman, Ichimura e Todd (1997) combinaram uma base de dados secundários de pessoas que não participaram de um programa com uma base de dados de um grande experimento aleatório para avaliar a performance de vários estimadores de pareamento, incluindo o estimador que combina o método de diferenças em diferenças (*DD*) com pareamento. Eles compararam a estimação que usa o grupo de controle do experimento, e faz apenas uma comparação de médias com as estimativas que usam o grupo de controle vindo dos dados secundários e os diversos métodos de pareamento.

Eles olham para o efeito de um programa nacional de treinamento de mão de obra (JTPA) implementado nos Estados Unidos com o objetivo de prover treinamento para os trabalhadores menos qualificados. Esse programa consta com treinamentos *online* e presenciais e assistência na procura por emprego. As pessoas eram elegíveis para o programa se a renda familiar estava próxima ou abaixo da linha de pobreza nos seis meses anteriores ao programa ou se eles participavam de algum programa governamental de auxílio à pobreza.

Foi feito um experimento que coletou dados longitudinais de um grupo de controle e outro de tratamento que foram aleatorizados da população que se inscreveu no programa, e de um grupo de controle criado à parte da população elegível que não aplicou para participar do programa. Dois terços de todos os indivíduos que aplicaram para o programa foram alocados para o grupo de tratamento e um terço foi para o grupo de controle. Os dados foram coletados em três pontos no tempo: um período antes da implementação do programa (*baseline*) e dos períodos posteriores à implementação do programa (*follow-ups*).

O escore de propensão foi estimado por um modelo logit, usando as informações socioeconômicas e de mercado de trabalho dos indivíduos antes da implementação do programa. O primeiro teste que eles fazem no artigo é olhar se a hipótese de sobreposição do escore de propensão é válida. Para isso, eles fazem gráficos similares aos apresentados no exemplo 1 para cada um dos quatro grupos demográficos considerados na análise: homens adultos, mulheres adultas, homens jovens e mulheres jovens.

O quadro a seguir resume os resultados principais encontrados no artigo:

Tabela 5.7.1: Viés de métodos de pareamento

	Impacto Diferença das Médias	Viés	
		Pareamento	DID Pareamento
Homens Adultos (Renda)	-418	123	97
Mulheres Adultas (Renda)	-26	67	74
Homens Jovens (Renda)	-51	49	43
Mulheres Jovens (Renda)	6	-7	-7

A primeira coluna apresenta o impacto do programa sobre a renda utilizando apenas a diferença de médias, a segunda e terceira colunas representam a diferença entre o resultado empregando o método experimental e os resultados por meio do método do pareamento dos M vizinhos mais próximos, e o método do pareamento combinado ao método de diferenças em diferenças. A tabela indica que o resultado obtido pela combinação do pareamento com o método de diferenças em diferenças é sempre melhor ou igual que o obtido apenas pelo pareamento, isto é, o viés do método que combina *DD* com pareamento é sempre menor. Os autores argumentam que esse resultado se deve pelo fato de que o método que combina pareamento com *DD* controla pelo viés que advém de fatores não observáveis fixos no tempo poderem estar correlacionados com a decisão de participar ou não.

5.8. Exercícios

1. Estamos interessados em estimar o efeito médio do tratamento,

$$EM = E[Y_i(1)] - E[Y_i(0)]$$

Usamos a seguinte hipótese de seleção nos observáveis

$$(Y_i(0), Y_i(1)) \perp T_i | X_i$$

e a seguinte hipótese de sobreposição,

$$0 < \Pr[T_i = 1|X_i] < 1$$

- a. Mostre que se simplesmente tirarmos a diferença entre a expectativa dos resultados potenciais para os tratados e a expectativa dos resultados potenciais para o controle não necessariamente conseguimos o efeito médio do tratamento.

- b. Mostre que podemos usar o método da reponderação para obter o efeito médio do tratamento. Ou seja, mostre a seguinte igualdade:

$$EM = E \left[\frac{T_i}{P(X)} Y_i \right] - E \left[\frac{1 - T_i}{1 - P(X)} Y_i \right]$$

- c. Proponha um estimador para o efeito médio do tratamento baseado no resultado obtido na letra (b).

2. No exemplo deste capítulo, Mattos, Maia e Marques (2010) usam o método do pareamento para estimar o efeito médio do tratamento de programas de transferências de renda sobre o número de horas remuneradas dos beneficiários do programa.

- a. Quais hipóteses devem ser válidas para que o estimador de pareamento seja consistente para o efeito médio do tratamento sobre os tratados?

- b. Interprete essas hipóteses no contexto do exemplo e discuta a validade dessas hipóteses.

- c. Suponha que você obtenha uma base de dados que contém as características (X) dos beneficiários e não beneficiários do programa antes e depois da implementação. Além disso, você também observa o resultado de interesse nos períodos pré e pós-programa. Como você modificaria a análise realizada por Mattos, Maia e Marques (2010)? Discuta a validade das hipóteses desse novo método.

3. Para estimar o efeito de um programa de treinamento no mercado de trabalho sobre salário, obtém-se uma base de dados que tem o salário dos indivíduos tratados e não tratados e um

vetor com várias características do *background* sócio-econômico dos indivíduos e também da sua atividade no mercado de trabalho. Primeiro, decide-se usar essas informações para estimar o efeito médio do tratamento sobre os beneficiários usando um pareamento simples baseado no vetor de características dos indivíduos. Ao tentar o pareamento baseado nas características individuais e nas características dos postos de trabalho, o vetor X tem uma dimensionalidade muito grande e decide-se fazer o pareamento baseado no escore de propensão.

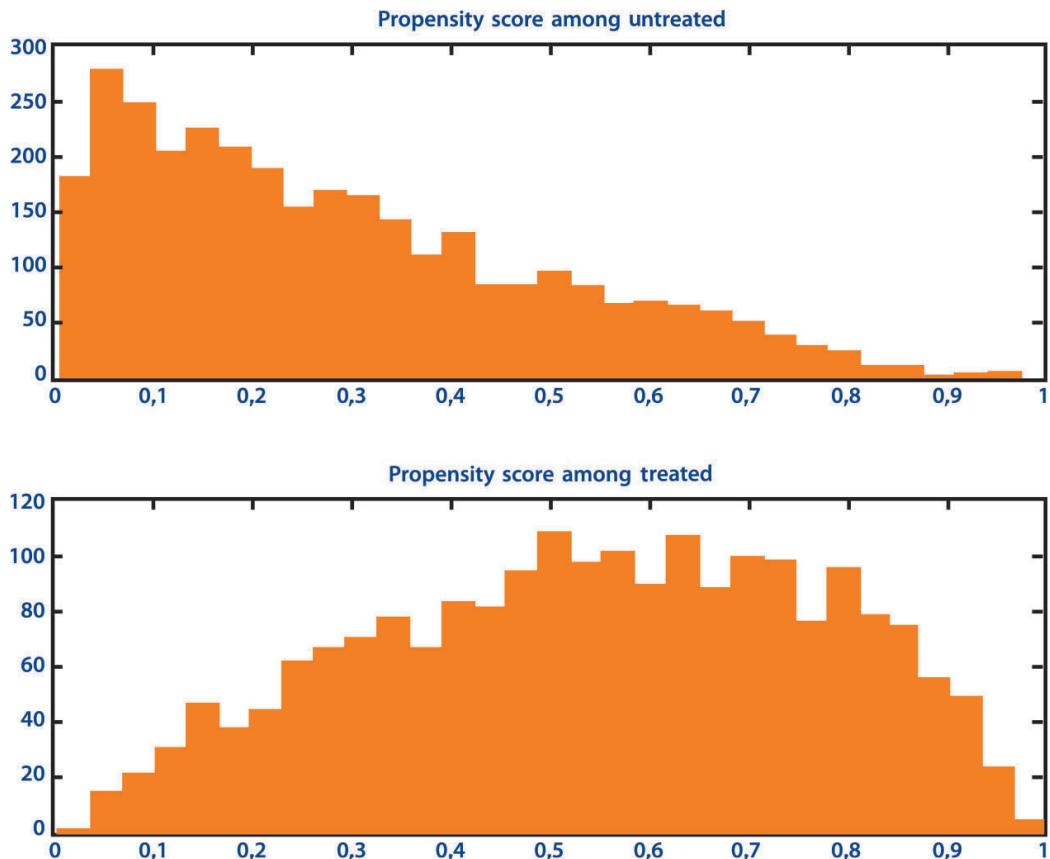
- a.** Quais hipóteses precisam ser satisfeitas para que esse estimador baseado no escore de propensão represente o efeito médio do tratamento sobre os beneficiários do programa?
 - b.** Poderia-se obter esse efeito médio do tratamento sobre os tratados por meio de regressão linear? Se sim, explique como. Se não, justifique a sua resposta.
 - c.** Um pesquisador lê o seu trabalho e argumenta que o seu resultado é muito sensível à especificação do escore de propensão. Ele acredita que a sua especificação do escore de propensão não é flexível o suficiente para que as hipóteses especificadas em (a) sejam satisfeitas. Como você responderia a essa crítica? Proponha um método que seja menos sensível à especificação do escore de propensão e interprete as suas hipóteses.
- 4.** Hirano e Imbens (2001) usam os métodos baseados no escore de propensão para estimar o efeito médio de um procedimento cardíaco, o cateterismo, sobre a probabilidade de sobrevivência dos indivíduos. Eles usam observações para 5.735 indivíduos, sendo 2.184 tratados e 3.551 não tratados. Para cada indivíduo, eles observam se um procedimento de cateterismo foi aplicado no período de 24 horas após a entrada no hospital com problemas cardíacos. Os indivíduos no grupo de tratamento são aqueles que sofreram a intervenção de um cateterismo, enquanto os indivíduos no grupo de controle são aqueles que não sofreram a intervenção. Além disso, eles observam um vetor com 72 características para cada um dos indivíduos, como sexo, raça, educação e renda, se possui plano de saúde, tipo de doença que possui, peso etc. Eles estimam o escore de propensão baseado em um modelo logístico.

- a. Antes de estimar o efeito médio do tratamento sobre os tratados, eles apresentam uma tabela que contém um teste de diferença de médias para as características observáveis. Eles obtêm resultados para esse teste antes e depois da ponderação pelo escore de propensão. Essa tabela encontra-se abaixo. Interprete os resultados.

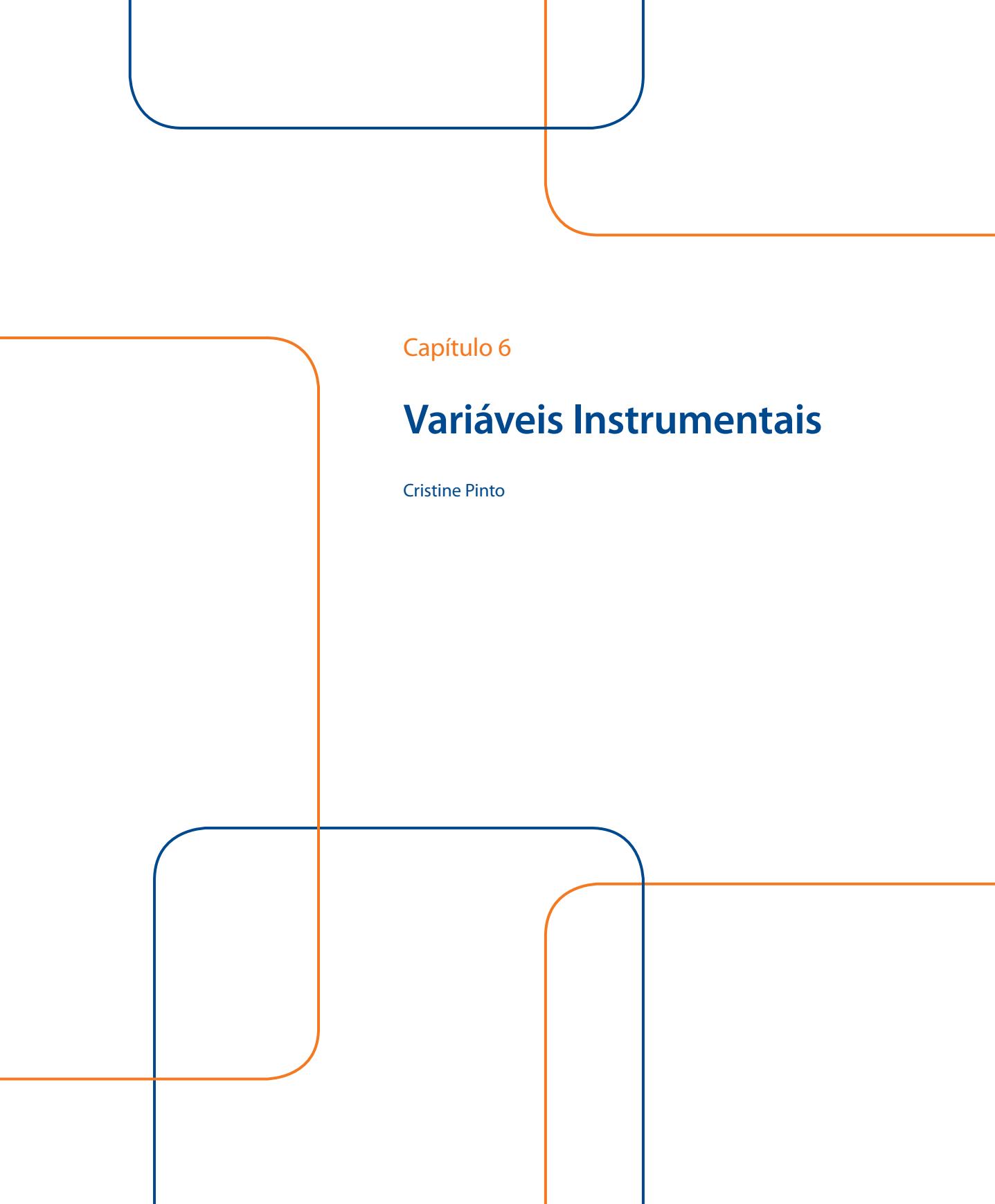
Variáveis				Pareando pelo Escore de Propensão		
	Não Tratado	Tratado	t-teste	Não Tratado	Tratado	t-teste
Idade	61,76	60,74	-2,28	61,25	61,15	-0,19
Sexo	0,46	0,41	-3,42	0,44	0,43	-0,85
Preto	0,16	0,15	-1,14	0,16	0,15	1,09
Educação	11,56	11,85	3,35	11,68	11,71	0,39
Renda 1	0,20	0,20	0,56	0,20	0,19	-1,19
Renda 2	0,14	0,17	3,88	0,14	0,16	1,05
Renda 3	0,07	0,08	2,19	0,07	0,07	0,12

Nesta tabela, Renda 1, Renda 2 e Renda 3 representam três faixas de renda baixa, média e alta, respectivamente.

- b. Antes de estimar os efeitos, eles também mostram os histogramas do escore de propensão para o grupo de tratados e para o grupo de não tratados. Interprete estes gráficos:



- c. Eles usam diferentes métodos para estimar o efeito médio do tratamento sobre a taxa de sobrevivência. Com um estimador de pareamento simples baseado nos vizinhos mais próximos, eles encontram um efeito de -0,081 (erro-padrão de 0,017). Ao combinar o método de pareamento com regressão simples, o efeito é de -0,063 (erro-padrão de 0,016). Estimando o efeito médio do tratamento por regressão linear, eles encontram um efeito médio de -0,065 (erro-padrão de 0,014), enquanto que, usando o método de reponderação, eles encontram um efeito médio de -0,060 (erro-padrão de 0,018). Combinado os métodos de regressão e reponderação, eles encontram um efeito médio de -0,062 (erro-padrão de 0,015). Interprete essas evidências, dando ênfase às hipóteses em que se baseiam cada estimador. Qual seria a conclusão sobre o efeito desse procedimento na taxa de sobrevivência?



Capítulo 6

Variáveis Instrumentais

Cristine Pinto

Ao contrário dos métodos apresentados no capítulo 5, o método de variável instrumental baseia-se no caso de seleção em variáveis não observáveis, isto é, variáveis que influenciaram a decisão de participar ou não do programa e que também afetam os resultados potenciais, mas não são observáveis pelo pesquisador. Neste capítulo, consideraremos o arcabouço de resultados potenciais apresentado no capítulo 2, mas lidaremos com a estimação do efeito médio do tratamento sob a hipótese de seleção em características não observáveis.

O nosso problema consiste em estimar o efeito médio do tratamento quando existem outros fatores, além das características observáveis presentes no vetor X_i , que afetam simultaneamente a decisão de participar no programa e os resultados potenciais, isto é, quando:

$$\Pr[T_i = 1|Y_i(1), Y_i(0), X_i] \neq \Pr[T_i = 1|X_i] = p(X_i)$$

Nesse caso, precisamos de uma variável exógena Z_i que afeta a decisão de participação e que não está correlacionada com nenhum fator não observável relacionado ao resultado potencial. No caso clássico de variável instrumental com efeito homogêneos do tratamento, estamos pensando no seguinte sistema de equações:

$$Y_i = \alpha + \beta T_i + \varepsilon_i$$

$$T_i = \begin{cases} 1 & \text{se } \gamma + \delta Z_i + \vartheta_i \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

em que T_i é igual a 1 se o indivíduo recebeu tratamento, e 0 se o indivíduo é não tratado. Nesse caso, assumimos que o indivíduo foi tratado dada certa condição que depende de variáveis observáveis e não observáveis. Além disso, assumimos que $\text{Cov}(Z_i, \varepsilon_i) = 0$ e $\text{Cov}(\vartheta_i, \varepsilon_i) \neq 0$, ou seja, que não há relação linear entre o instrumento e o termo aleatório da equação principal e que os termos aleatórios das duas equações são relacionados. Esse modelo pode ser estimado por mínimos quadrados em dois estágios. Nesse caso, em primeiro estágio, estimamos um modelo de probabilidade linear que relaciona T_i com Z_i e obtemos o valor predito:

$$\hat{T}_i = \hat{\gamma} + \hat{\delta} Z_i$$

Em segundo estágio, estimamos uma regressão linear que relaciona o resultado de interesse (Y_i) com este valor predito:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}\hat{T}_i$$

No sistema de equação acima, a variável Z_i não afeta diretamente Y_i . Ela só afeta o resultado de interesse por meio da sua relação com a participação ou não no tratamento. Além disso, assumimos que o efeito de tratamento é homogêneo, isto é, $\beta=\beta_i$ para todo indivíduo i . No caso de tratamento homogêneo, o resultado do indivíduo depende apenas da sua participação ou não no programa, e não está relacionada a como a participação no programa é afetada pelo instrumento Z_i . No caso de tratamento homogêneo, o efeito médio do tratamento (EMP) é igual ao efeito médio do tratamento sobre os tratados ($EMPT$), ou seja:

$$EMP = EMPT = \beta.$$

Se os indivíduos souberem que os ganhos de participação podem diferir para certos grupos, eles irão levar em consideração essa informação na hora de decidir se participam ou não no programa. Nesse caso, tanto seus ganhos individuais β_i como Z_i irão afetar a decisão de participar do indivíduo i , e variações em Z_i irão afetar a decisão de participar de forma diferenciada para cada indivíduo, dependendo do seu ganho com o tratamento, β_i . Nesse caso, a hipótese de homogeneidade do tratamento é violada, e o estimador de variável instrumental não identifica o efeito médio do tratamento ou o efeito médio do tratamento sobre os tratados.

Imbens e Angrist (1994) mostraram que quando os efeitos do tratamento são heterogêneos, o método de variável instrumental permite identificar um efeito médio de tratamento local, isto é, um efeito médio do tratamento para uma subpopulação específica. Nesse caso, β será o efeito médio do tratamento para aqueles indivíduos cuja variação em Z_i provoca uma variação no *status* de participação sem afetar os resultados potenciais. Para o efeito no resultado ocasionado exclusivamente pelo efeito do instrumento Z_i na taxa de participação é o efeito local do tratamento. Esse efeito representa um efeito médio do tratamento para a subpopulação de indivíduos induzidos

ao tratamento por variações no instrumento é chamado de efeito médio local do tratamento (*EML* ou *LATE*¹).

Iremos entender esse efeito dentro do arcabouço de resultados potenciais. Nesse arcabouço, definimos como *compliers* os indivíduos que são induzidos pelo instrumento a ter uma mudança de comportamento na variável endógena, no caso em questão, na participação ou não no programa. O *LATE* seria o efeito médio do tratamento para esta população de *compliers*. Como no capítulo 2, definimos $Y_i(0)$ como o resultado potencial do indivíduo i caso ele não tivesse participado no programa e $Y_i(1)$ como o resultado potencial do indivíduo i caso ele tivesse participado. Podemos definir o resultado observado, como:

$$Y_i = Y_i(T_i) = \begin{cases} Y_i(1) & \text{se } T_i = 1 \\ Y_i(0) & \text{se } T_i = 0 \end{cases} \quad (6.1)$$

Suponha que, além de observarmos (Y_i, T_i) para cada indivíduo, observamos também uma variável instrumental binária Z_i . Nesse caso, $T_i(0)$ seria o valor potencial da variável endógena (participação no tratamento) se a pessoa tivesse um valor de Z_i que a induzisse a não participar e $T_i(1)$ seria o valor potencial da variável endógena (participação no tratamento) se a pessoa tivesse um valor de Z_i que a induzisse a participar.² O valor realizado para T_i pode ser definido como uma função do instrumento:

$$T_i = T_i(Z_i) = \begin{cases} T_i(1) & \text{se } Z_i = 1 \\ T_i(0) & \text{se } Z_i = 0 \end{cases} \quad (6.2)$$

Nesse arcabouço de resultados e tratamentos potenciais, podemos pensar no resultado potencial como uma função de Z_i e T_i , $Y_i(Z_i, T_i)$ que seria o resultado potencial observado se o instrumento assumisse valor z e o tratamento valor t . Com um instrumento binário, temos quatro resultados potenciais, $Y_i(0,0)$, $Y_i(0,1)$, $Y_i(1,0)$, $Y_i(1,1)$. Por exemplo, $Y_i(0,0)$ seria o resultado potencial do indivíduo que obteve um valor de Z_i igual a 0 e decidiu não participar do programa. Um exemplo clássico da estimativa de um efeito médio de tratamento local é o artigo do Angrist (1990), estima

1. Essa sigla *LATE* vem do nome em inglês “Local Average Treatment Effect”.

2. $T_i(0)$ e $T_i(1)$ são os valores potenciais para a participação no tratamento, e seriam respectivamente os valores de T que eles teriam se não tivessem sido tratados ou tivessem sido tratados, independentemente se eles de fato receberam tratamento ou não.

o efeito de ser veterano de guerra na renda dos homens americanos que serviram na guerra do Vietnã. Para isso, ele usa uma loteria que foi usada para selecionar os homens que deveriam ir para guerra. Caso o homem tirasse um número alto nessa loteria, ele não precisava servir na guerra do Vietnã. Neste caso, T_i é igual a 1 se o indivíduo i serviu a guerra do Vietnã e 0, caso contrário. E Z_i é a variável instrumental baseada na loteria. Nesse caso, $Z_i=1$ se o indivíduo 1 tirou um número baixo na loteria; e é igual a 0 se o indivíduo tirou um número alto. Nesse caso, $T_i(1)$ representa o resultado da participação caso o indivíduo tirasse um número baixo na loteria, independentemente de o indivíduo ter tirado ou não um número baixo; e $T_i(0)$ representa o resultado da participação caso o indivíduo tirasse um número alto na loteria, independentemente de o indivíduo ter tirado um número alto ou baixo. Da mesma maneira que no capítulo 2, $Y_i(0)$ representa o resultado potencial do indivíduo i se ele não tivesse servido ao exército, independentemente de ele ter servido ou não; e $Y_i(1)$ representa o resultado potencial do indivíduo i se ele tivesse servido ao exército, independentemente de ele ter servido ou não.

Note que nesse instrumental, podemos definir quatro resultados potenciais, $Y_i(z,t)$. Nesse caso, $Y_i(0,0)$ representa o resultado potencial do indivíduo i se ele tivesse tirado um número alto na loteria e não participado na guerra, independentemente se ele tirou de fato um número alto ou baixo na loteria e se serviu ao exército ou não. De forma análoga, podemos definir outros resultados potenciais: $Y_i(0,1)$, $Y_i(1,0)$ e $Y_i(1,1)$.

Como no caso clássico de variável instrumental, precisamos que o instrumento Z_i seja não correlacionado com os termos não observáveis que afetam o nosso resultado de interesse. Para garantir que isso aconteça em um arcabouço com efeitos heterógenos, precisamos que Z_i seja independente não apenas dos resultados potenciais, mas também dos tratamentos potenciais. Precisamos que a variável instrumental seja tão boa como se ela fosse alocada de forma aleatória entre os indivíduos, isto é, precisamos que o instrumento seja independente dos quatro resultados potenciais e dos tratamentos potenciais.

Alocação Aleatória: Z_i é independente de $(Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1), T_i(1), T_i(0))$ (H1)

Além de Z_i ser alocado de forma aleatória, precisamos que ele só afete o resultado de interesse por meio da variação que provoca na taxa de participação, ou seja, precisamos de uma hipótese conhecida como restrição de exclusão. Essa hipótese garante que o instrumento só afete Y_i por

meio de um canal conhecido. De modo formal, essa hipótese garante que, ao controlarmos por t , $Y_i(z,t)$ não é uma função de z , mas somente de t :

Restrição de Exclusão: $Y_i(z,t) = Y_i(z^*,t)$ para todo t, z, z^* (H2)

A restrição de exclusão permite que os resultados potenciais sejam definidos somente com base no *status* do tratamento:

$$Y_{1i} \equiv Y_i(1,1) = Y_i(0,1) \quad (6.3)$$

$$Y_{0i} \equiv Y_i(1,0) = Y_i(0,0) \quad (6.4)$$

e o resultado observado pode ser escrito como:

$$Y_i = Y_i(0, Z_i) + (Y_i(1, Z_i) - Y_i(0, Z_i)) \cdot T_i \quad (6.5)$$

No caso de tratamentos heterogêneos, cada indivíduo responde a diferentes valores do instrumento, dependendo do resultado recebido pelo tratamento. Podemos pensar em quatro tipos de comportamento induzidos pelo instrumento. Suponha que os indivíduos que recebem valores de Z_i igual a 1 são convidados a participar do programa, enquanto os indivíduos com valores de Z_i igual a 0 não são convidados a participar. Podemos pensar, por exemplo, em um programa em que o aluno pode se matricular em um curso técnico dependendo de sua nota em um teste de proficiência e o seu histórico escolar. Se a nota do aluno no teste for superior a determinado valor c , ele é convidado a se matricular no curso. Caso ele receba uma nota inferior a c , ele não recebe o convite. Nesse caso, o instrumento seria uma variável binária que é igual a 1 se o aluno tirou uma nota no teste de proficiência acima de c , e 0, caso contrário. Podemos pensar em um primeiro grupo de alunos que gostariam de se matricular no programa independentemente da sua nota no teste de proficiência, isto é, independentemente do valor do instrumento. Esses alunos são chamados de *always-takers*, pois eles sempre aceitariam participar do programa independentemente do valor do instrumento. No exemplo do Angrist (1990), os *always-takers* são os homens que sempre serviriam na guerra do Vietnã independentemente do valor tirado na loteria. Existe outro grupo de indivíduos que não gostaria de se matricular no curso técnico, independentemente do valor do instrumento. Esse grupo de indivíduos são conhecidos como *never-takers* e são aqueles que nunca

aceitariam participar, independentemente da sua nota no teste de proficiência. No caso da guerra, esse grupo é composto pelos homens que nunca iriam servir na guerra, independentemente da loteria. Além disso, existem indivíduos que irão mudar o seu comportamento de acordo com o valor do instrumento. O terceiro grupo é composto pelos indivíduos que, se tiverem uma nota acima de c no teste de proficiência, irão realizar o curso técnico, mas se receberem uma nota abaixo de c no teste, eles não irão se matricular. Esses indivíduos são chamados de *compliers*. No caso do exemplo da guerra, os *compliers* são os homens que foram para a guerra, dado que tiraram um número baixo na loteria e não iriam para a guerra se tirassem um número alto. E também aqueles que não foram para a guerra porque tiraram um número alto na loteria, mas que iriam para a guerra se tirassem um número baixo. Ou seja, são os homens que tiveram os seus comportamentos induzidos pela loteria. Finalmente, teremos os indivíduos que terão comportamento oposto àquele induzido pelo instrumento. Nesse grupo, se os indivíduos recebessem notas abaixo de c , eles gostariam de se matricular no curso; mas se eles recebessem notas acima de c , eles não se matriculariam no curso. Esse último grupo é conhecido como *defiers*. No caso da guerra, são os homens que teriam um comportamento oposto a regra da loteria, isto é, se tirassem um número baixo não serviriam na guerra, e se tirassem um número alto, iriam para a guerra. Podemos representar esses quatro grupos em uma tabela³, de acordo com o comportamento de cada um, a variações do instrumento:

	$T_i(0) = 0$	$T_i(0) = 1$
$T_i(1) = 0$	<i>never-taker</i>	<i>defier</i>
$T_i(1) = 1$	<i>complier</i>	<i>always-taker</i>

Por exemplo, a primeira célula desta tabela representa os indivíduos que, independentemente da realização do instrumento (0 ou 1), seu tratamento potencial é sempre igual a 0, ou seja, eles nunca participam. Esses são os *never-takers*. Baseado nas variáveis que observamos para cada indivíduo (Y_i, T_i, Z_i), não podemos classificar esses indivíduos em cada um dos grupos acima. Para isso, precisaríamos saber como eles iriam reagir, caso eles recebessem outra realização para a variável instrumental. Com as informações que observamos, podemos montar a seguinte tabela:

3. Estas tabelas foram retiradas de Imbens e Wooldridge (2007).

	$Z_i = 0$	$Z_i = 1$
$T_i = 0$	complier/never-taker	never-taker/defier
$T_i = 1$	always-taker/defier	complier/always-taker

Esta tabela indica que, ao observamos para o indivíduo i o par $(z,t)=(0,1)$, só sabemos que esse indivíduo não participaria da guerra, se ele tivesse um número alto na loteria. Esse indivíduo pode estar no grupo de *compliers* ou de *never-takers*. Para saber de fato a qual grupo o indivíduo realmente pertence, precisaríamos saber qual seria o seu comportamento caso ele tirasse um número baixo na loteria. Para que o *LATE* identifique o efeito médio para alguma subpopulação, precisamos identificar o comportamento de cada um dos grupos e para isso assumimos que existem poucos indivíduos que farão o oposto do que foi sugerido pelo instrumento, isto é, assumimos monotonicidade dos tratamentos potenciais:

Monotonicidade: $T_i(1) \geq T_i(0)$ (H3)

Essa hipótese de monotonicidade garante que todas as pessoas são afetadas na mesma direção pelo instrumento, isto é, não podem existir os indivíduos que, se recebessem um número alto na loteria, iriam para a guerra e se recebessem um número baixo na loteria, não serviriam na guerra. Com a hipótese de monotonicidade, eliminamos os *defiers*, e a informação que temos nos dados observáveis aumenta e pode ser representada pela tabela abaixo,

	$Z_i = 0$	$Z_i = 1$
$T_i = 0$	complier/never-taker	never-taker
$T_i = 1$	always-taker	complier/always-taker

A hipótese de monotonicidade elimina a existência dos *defiers*. Usando a informação da tabela acima, podemos identificar o efeito médio do tratamento sobre a população de *compliers*. Podemos decompor a expectativa condicional de Y_i em Z_i no comportamento de cada um dos grupos,

$$E[Y_i|Z_i=1] = E[Y_i|Z_i=1, T_i=1] \cdot Pr[T_i=1|Z_i=1] + E[Y_i|Z_i=1, T_i=0] \cdot Pr[T_i=0|Z_i=1] \quad (6.6)$$

$$E[Y_i|Z_i=0] = E[Y_i|Z_i=0, T_i=1] \cdot Pr[T_i=1|Z_i=0] + E[Y_i|Z_i=0, T_i=0] \cdot Pr[T_i=0|Z_i=0] \quad (6.7)$$

Sob a hipótese de monotonicidade, $Pr[T_i=1|Z_i=0]$ representa a proporção de *always-takers*, $Pr[T_i=0|Z_i=1]$ representa a proporção de *never-takers* e $Pr[T_i=1|Z_i=1]-Pr[T_i=1|Z_i=0]=Pr[T_i=0|Z_i=0]-Pr[T_i=0|Z_i=1]$ representa a proporção de *compliers*. Podemos decompor a diferença dos resultados potenciais em quatro grupos:

$$\begin{aligned} & E[Y_i|Z_i=1]-E[Y_i|Z_i=0] \\ & = E[Y_i|Z_i=1, T_i=1] \cdot (Pr[T_i=1|Z_i=1]-Pr[T_i=1|Z_i=0])- \\ & E[Y_i|Z_i=0, T_i=0] \cdot (Pr[T_i=0|Z_i=0]-Pr[T_i=0|Z_i=1])+ \\ & (E[Y_i|Z_i=1, T_i=1]-E[Y_i|Z_i=0, T_i=1]) \cdot Pr[T_i=1|Z_i=0]+ \\ & (E[Y_i|Z_i=1, T_i=0]-E[Y_i|Z_i=0, T_i=0]) \cdot Pr[T_i=0|Z_i=1] \end{aligned}$$

Note que as duas primeiras linhas representam o grupo de *compliers*, a terceira linha se refere ao grupo de *always-takers* e a última linha representa o grupo de *never-takers*. Sob a hipótese de restrição de exclusão,

$$E[Y_i|Z_i=1, T_i=1]-E[Y_i|Z_i=0, T_i=1]=0$$

$$E[Y_i|Z_i=1, T_i=0]-E[Y_i|Z_i=0, T_i=0]=0$$

e podemos reescrever a expressão acima como:

$$\begin{aligned} & E[Y_i|Z_i=1]-E[Y_i|Z_i=0] \\ & =(E[Y_i|Z_i=1, T_i=1]-E[Y_i|Z_i=0, T_i=0]).(Pr[T_i=1|Z_i=1]-Pr[T_i=1|Z_i=0]) \end{aligned}$$

O efeito médio do tratamento para o grupo de *compliers* é simplesmente a seguinte razão:

$$(E[Y_i|Z_i = 1, T_i = 1] - E[Y_i|Z_i = 0, T_i = 0]) = \frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{\Pr[T_i=1|Z_i=1] - \Pr[T_i=1|Z_i=0]} \quad (6.8)$$

Essa razão é o parâmetro estimado pelo método de variável instrumental quando o instrumento é uma variável binária. Os dados observáveis só nos dão informação sobre o efeito médio do tratamento para a subpopulação de indivíduos que tiveram o comportamento induzido pelo instrumento, isto é, para o grupo de *compliers*. O estimador representado pela equação (6.8) é conhecido como estimador de Wald. No entanto, para facilitar o processo de inferência (testes de hipóteses e intervalos de confiança), geralmente, estimamos o *LATE* por meio do método de mínimos quadrados ordinários em dois estágios.

Suponha que observamos um vetor de características adicionais dos indivíduos (X_i) que podem ser usados nos modelos para o resultado de interesse e para a participação no tratamento. Nesse caso, esse vetor seria incluído nos dois modelos que iremos estimar por mínimos quadrados em dois estágios:

$$Y_i = \alpha + \beta T_i + \pi X_i + \varepsilon_i \quad (6.9)$$

$$T_i = \gamma + \delta Z_i + \varphi X_i + \vartheta_i \quad (6.10)$$

Consideramos que X_i são variáveis exógenas e possuem efeito direto no resultado de interesse. Se a variável instrumental Z_i for de fato alocada de forma aleatória entre os indivíduos, ela seria independente das demais variáveis explicativas do modelo (X_i), e a inclusão do vetor X_i no modelo só levaria ao aumento da precisão das estimativas. No entanto, como no caso tradicional de variáveis instrumentais, a razão principal de se incluir outras variáveis explicativas no modelo é que a hipótese de alocação aleatória de Z_i e a restrição de exclusão são válidas somente se condicionarmos em X_i , isto é, se compararmos indivíduos com as mesmas características observáveis. De maneira mais formal, nesse caso, as hipóteses de alocação aleatória e a restrição de exclusão seriam:

Alocação Aleatória*: Z_i é independente de $(Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1), T_i(1), T_i(0))$, condicional em X_i (H1*)

Essa hipótese nos diz que entre indivíduos com o mesmo vetor de características X_i , o instrumento Z_i foi alocado de forma aleatória.

Restrição de Exclusão*: $Y_i(z, t, x) = Y_i(z^*, t, x)$ para todo t, x, z, z^* (H2)

Nesse caso, somente para indivíduos com o mesmo vetor de variáveis explicativas, os resultados potenciais não irão depender da realização de z .

No caso em que incluímos um vetor adicional de variáveis explicativas no modelo, a estimação também pode ser realizada por mínimos quadrados em dois estágios. No primeiro estágio estimamos o modelo linear que relaciona a participação com todas as variáveis exógenas e calculamos o valor predito,

$$\hat{T}_i = \hat{\gamma} + \hat{\delta}Z_i + \hat{\varphi}X_i$$

e no segundo estágio, usamos esse valor predito como variável explicativa no modelo linear para Y_i ,

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}\hat{T}_i + \hat{\pi}X_i$$

6.1. Interpretação do parâmetro LATE e a sua relação com os outros métodos

Como já foi dito, apesar de muito semelhante ao estimador clássico de variável instrumental (se eliminarmos a hipótese de efeitos heterogêneos), o **LATE** não representa o efeito médio do tratamento ou o efeito médio do tratamento sobre os tratados, representa apenas o efeito médio do tratamento para uma subpopulação bem específica, a população de indivíduos que tem seu comportamento alterado em função de uma variação no instrumento. Em geral, esse grupo de indivíduos não é representativo da população como um todo. Dada essa limitação, podemos perguntar se esse efeito local identificado pelo **LATE** é interessante para uma análise política. A resposta a essa pergunta irá depender do instrumento que será usado na análise. Por exemplo, se a variável instrumental for alguma mudança de política que afeta a taxa de participação, o

LATE representa um parâmetro de interesse para essa política, pois representa o efeito médio do tratamento para a subpopulação que teve o seu comportamento alterado por tal política. Se, por outro lado, usarmos uma variável que simplesmente afeta a taxa de participação e não tem efeito direto no resultado, o *LATE* pode ser um parâmetro menos informativo.

É importante pensar também na relação entre o arcabouço do *LATE* e os demais métodos apresentados neste livro. Como já foi dito, diferentemente dos métodos de pareamento, o método de variável instrumental é usado quando a hipótese de seleção em características observáveis é violada, ou seja, o processo de seleção para o tratamento é baseado em características não observáveis. Na próxima seção, apresentaremos o método de regressão descontínua e veremos que o caso fuzzy desse novo método pode ser entendido como um caso especial de variável instrumental.

Além disso, podemos fazer uma analogia entre o *LATE* e um experimento aleatório. Suponha que o instrumento seja uma oferta de tratamento que foi alocado de forma aleatória entre os indivíduos. Por exemplo, suponha que se queira testar o efeito de um exame novo que foi criado para detectar câncer de mama e para isso utiliza-se uma amostra aleatória de mulheres de 35 a 50 anos para realizar o exame. As mulheres que recebem a carta-convite podem se dirigir à clínica e fazer o exame ou podem não aceitar o convite e não fazer o exame. Como são as mulheres que escolhem se irão fazer ou não o exame, as mulheres que decidem fazê-lo podem ser, na média, diferentes das mulheres que se recusaram. Por exemplo, as mulheres que aceitaram os convites são geralmente mais bem instruídas, preocupam-se mais em realizar os seus exames periódicos para detecção de doenças e têm hábitos mais saudáveis, tendo na média uma saúde melhor que as mulheres que rejeitaram o convite. Logo, para encontrar o efeito médio do tratamento sobre a probabilidade de detectar câncer de mama, não podemos comparar a proporção de mulheres que fizeram o exame e foram diagnosticadas com câncer de mama com a proporção de mulheres que não fizeram o exame e receberam o diagnóstico da doença, pois nesse caso, estariam misturando o efeito do programa com o efeito de uma vida mais saudável. Assim, estamos no arcabouço em que o instrumento vem de um experimento real e foi alocado de forma aleatória entre as mulheres. O instrumento seria uma variável binária que assume valor igual a 1 se a mulher foi convidada a fazer o exame, e 0, caso contrário. Dessa forma, o *LATE* é o efeito médio sobre as mulheres que de fato receberam o tratamento, isto é, o efeito médio do tratamento sobre os tratados. Essa é uma das situações comentadas no capítulo 3 em que houve o não comparecimento dos tratados na intervenção, de

modo que não podemos usar diretamente a comparação de médias entre tratados e controles para estimar o efeito médio do tratamento, como no caso ideal de experimento aleatório.

Bloom (1984) avalia o uso de variáveis instrumentais na presença de normas que impeçam as pessoas que não foram selecionadas para entrar no programa a participar dele. Esse é o caso em que $T_i(0)=0$, isto é, que as pessoas que de fato não foram alocadas ao programa não poderiam participar. Nesse caso, o *LATE* sempre identifica o efeito médio do tratamento sobre os tratados (*EMPT*). Isso é o que acontece no exemplo da mamografia, as mulheres que não receberam a carta não tinham direito de participar. Nesse caso, eliminamos o grupo de *always-takers* e os *compliers* serão de fato a população tratada.

6.2. O caso de múltiplos instrumentos

Imbens e Angrist (1994) e Angrist e Imbens (1995) mostram que o estimador que usa múltiplos instrumentos é uma média ponderada de cada um dos *LATEs* obtidos com cada instrumento específico. Nesta seção, iremos entender esse resultado dentro de um instrumental simples no qual consideramos um par de instrumentos binários Z_{1i} e Z_{2i} , que são mutuamente exclusivos. Suponha que a hipótese de monotonicidade é satisfeita para cada uma dessas variáveis binárias. Para estimar o efeito médio do tratamento para a população de *compliers*, usamos o estimador de MQO em dois estágios. No primeiro estágio, estimamos o modelo linear que relaciona o tratamento com as duas variáveis instrumentais:

$$\hat{T}_i = \hat{\pi}_{11}Z_{1i} + \hat{\pi}_{12}Z_{2i} \quad (6.2.1)$$

No segundo estágio, usamos esse valor predito como a variável explicativa no modelo linear para o resultado Y_i e estimamos o seguinte modelo linear simples:

$$Y_i = \alpha + \rho\hat{T}_i + \epsilon_i \quad (6.2.2)$$

O estimador de mínimos quadrados em dois estágios para esse caso é:

$$\hat{\rho} = \frac{Cov(Y_i, \hat{T}_i)}{Cov(T_i, \hat{T}_i)} \quad (6.2.3)$$

Substituindo a equação do primeiro estágio para \hat{T}_i :

$$\hat{\rho} = \hat{\pi}_{11} \frac{Cov(Y_i, Z_{1i})}{Cov(T_i, \hat{T}_i)} + \hat{\pi}_{12} \frac{Cov(Y_i, Z_{2i})}{Cov(T_i, \hat{T}_i)} \quad (6.2.4)$$

Multiplicando e dividindo o primeiro termo por $Cov(T_i, Z_{1i})$, o segundo termo por $Cov(T_i, Z_{2i})$, temos que:

$$\hat{\rho} = \frac{\hat{\pi}_{11} Cov(T_i, Z_{1i})}{Cov(T_i, \hat{T}_i)} \cdot \frac{Cov(Y_i, Z_{1i})}{Cov(T_i, Z_{1i})} + \frac{\hat{\pi}_{12} Cov(T_i, Z_{2i})}{Cov(T_i, \hat{T}_i)} \cdot \frac{Cov(Y_i, Z_{2i})}{Cov(T_i, Z_{2i})}$$

Note que $\frac{Cov(Y_i, Z_{1i})}{Cov(T_i, Z_{1i})}$ seria o *LATE* usando apenas o instrumento Z_1 , e $\frac{Cov(Y_i, Z_{2i})}{Cov(T_i, Z_{2i})}$ seria o *LATE* usando apenas o instrumento Z_2 . Logo, podemos escrever o estimador de MQO como uma média ponderada dos estimadores usando cada uma das variáveis instrumentais,

$$\hat{\rho} = \omega \cdot \frac{Cov(Y_i, Z_{1i})}{Cov(T_i, Z_{1i})} + (1 - \omega) \cdot \frac{Cov(Y_i, Z_{2i})}{Cov(T_i, Z_{2i})} \quad (6.2.5)$$

Os pesos de cada um dos *LATES* depende de quanto forte é a relação de cada instrumento com a participação no tratamento, isto é, será dado maior peso ao instrumento que apresenta uma relação mais forte com a participação no tratamento. Podemos mostrar que esse resultado se estende para o caso discreto de múltiplos instrumentos.

6.3. O caso de instrumentos contínuos

Nessa seção, consideraremos a estimativa do impacto do tratamento sobre toda a distribuição de um instrumento contínuo. Nesse caso, o instrumento é contínuo, e a probabilidade de participação condicional a Z_i também é uma função contínua, isto é, $p(z) = E[T_i | Z_i = z]$ é uma função contínua em z . Nesse caso, usamos mudanças infinitesimais na taxa de participação para medir o efeito médio local do tratamento (*LATE*). Heckman (1990), Vytlacil (2000), Heckman e Vytlacil (2001), Carneiro, Heckman e Vytlacil (2005) usam um modelo de seleção para interpretar esse efeito marginal.

A seleção é feita por uma função de Z_i e de um componente não observável, ϵ_i :

$$T_i = \begin{cases} 1 & \text{se } g(z_i) - \epsilon_i \geq 0 \\ 0 & \text{caso contrário} \end{cases} \quad (6.3.1)$$

Sob a hipótese de alocação aleatória de Z_i , e se a restrição de exclusão e a hipótese de monotonicidade são satisfeitas para todos os pares de z e z^* , o modelo latente implica que podemos ordenar os indivíduos de acordo com os seus componentes não observáveis ε_i . Se o fator não observável do indivíduo i é menor que o do indivíduo j ($\varepsilon_i < \varepsilon_j$), então $T_i(z) \geq T_j(z)$ para todo z . Dado esse ordenamento baseado no não observável, podemos definir o efeito marginal do tratamento condicional em ε_i :

$$\beta(\varepsilon) = E[Y_i(1) - Y_i(0) | \varepsilon_i = \varepsilon] \quad (6.3.2)$$

Esse efeito se relaciona diretamente com o limite do efeito médio local do tratamento definido para os valores de ε tal que existe um valor z que satisfaz $g(z) = \varepsilon$,

$$\beta(\varepsilon) = \beta(z) \text{ para } g(z) = \varepsilon \quad (6.3.3)$$

Para entender esse parâmetro como o limite do *LATE*, voltamos ao caso mais simples em que $g(Z_i)$ é uma função linear em Z_i . Nesse caso, o estimador do *LATE* para dois pontos na distribuição de z ,

$$\beta^{LATE}(z, z^*) = \frac{E[Y_i|Z_i=z] - E[Y_i|Z_i=z^*]}{\Pr[T_i=1|Z_i=z] - \Pr[T_i=1|Z_i=z^*]} \quad (6.3.4)$$

E podemos pensar no efeito marginal do tratamento (*EMT*) como o estimador de *LATE* quando z fica arbitrariamente próximo de z^* ,

$$\beta^{EMT}(z) = \frac{dE[Y_i|Z_i=z]}{d\Pr[T_i=1|Z_i=z]} \quad (6.3.5)$$

Esse parâmetro *EMT* utiliza um instrumento contínuo na tentativa de reconstruir a distribuição das probabilidades de participação e também essa distribuição para estimar o efeito na variação marginal de z .

6.4. Limitações do método

Como foi dito no capítulo 2 e na seção (6.1), o *LATE* é muito usado em experimentos aleatórios nos quais alguns indivíduos selecionados para participar do programa decidiram não participar. Essa situação é conhecida como experimento aleatório com *no-compliance*. Nesse caso, o *LATE* usa o resultado da aleatorização com instrumento e identifica o efeito médio do tratamento para aquelas pessoas que tiveram o comportamento modificado devido à aleatorização. No exemplo do Box 3.2 do capítulo 3, houve uma loteria que selecionou algumas crianças para frequentarem a creche, no entanto muitas famílias que conseguiram a vaga pela loteria não matricularam o seu filho na creche. O método de variáveis instrumentais pode ser usado para estimar o efeito médio local da creche no desenvolvimento das crianças do Rio de Janeiro. Nesse caso, os *compliers* seriam as crianças que foram matriculadas na creche, pois foram sorteadas na loteria, mas que não seriam matriculadas caso elas não fossem sorteadas.

Voltando à discussão introduzida no capítulo 3 sob validade externa e validade interna de um método, em que é dado um instrumento válido, isto é, que atende as hipóteses de restrição de exclusão e aleatoriedade, o método de variável instrumental apresenta uma grande validade interna, isto é, ele estima o efeito causal do tratamento para a população de *compliers*. No entanto, o grau de validade externa desse método pode ser baixo, pois ele só representa o efeito para essa subpopulação de *compliers*. Além disso, a interpretação do efeito de tratamento médio local depende do instrumento que estamos usando. Diferentes instrumentos mudam o comportamento de subpopulações distintas e irão estimar efeitos diferentes. É importante conhecer bem as características das subpopulações de *compliers* e também qual seria o tamanho desse grupo na população como um todo.⁴ Se essa subpopulação for bem parecida em observáveis com a população de interesse, podemos extrapolar esse efeito para a população como um todo, mas nem

4. Usando as fórmulas apresentadas na identificação do *LATE* (equação 6.8), é fácil ver como podemos calcular o tamanho do grupo de *compliers*. A proporção de tratados que são *compliers* é dada por:

$$\Pr[\text{compliers}|\text{tratado}] = \frac{\Pr[Z_i = 1](E[T_i|Z_i = 1] - E[T_i|Z_i = 0])}{\Pr[T_i = 1]}$$

Note que $E[T_i|Z_i=1]-E[T_i|Z_i=0]$ é o resultado do primeiro estágio do estimador de MQO em dois estágios. O tamanho do grupo de *compliers* na população de tratados é o resultado do primeiro estágio vezes a probabilidade do instrumento ser igual a 1 dividido pela probabilidade de ser tratado.

sempre é o que ocorre. No caso de a estimação do efeito ser de veterano de guerra sobre a renda, Angrist (1990) usando a loteria com instrumento estima o efeito para a subpopulação de homens que mudaram o seu comportamento devido à loteria. No entanto, na guerra do Vietnã, a maioria dos soldados era de voluntários, e o efeito encontrado no artigo não representa o efeito médio de servir à guerra para toda a população de homens que foram para a guerra. Para estimar o efeito sobre os voluntários, outro método de estimativa é necessário.

6.5. Exemplo

Ponczek e Souza (2011) usam o arcabouço de variável instrumental para estimar o efeito do tamanho da família em alguns resultados das crianças. Em uma amostra de famílias cuja mulher teve duas ou mais gestações, eles usam como instrumento uma variável binária que indica se na segunda gestação ocorreu o nascimento de gêmeos. Segundo os autores, esse acontecimento de gêmeos é claramente relacionado com o tamanho de família e condicional ao vetor de características da família, ele só afeta os resultados de interesse pelo efeito no tamanho da família. Nesse caso, a subpopulação de *compliers* é composta pelas famílias com pelo menos um filho nascido da primeira gestação que tiveram o tamanho da família aumentado em um número inesperado devido ao nascimento de gêmeos. Os resultados de interesse neste artigo que serão apresentados aqui são: se a criança participa da força de trabalho, se a criança vai à escola e se a criança não repetiu a série/ano. As tabelas (6.4.1) e (6.4.2) apresentam os resultados para as crianças com idade entre 10 e 15 anos que moram com a mãe e o marido da mãe. Os efeitos médios locais foram estimados por mínimos quadrados em dois estágios. A tabela (6.4.1) apresenta os resultados do primeiro estágio, enquanto a tabela (6.4.2) contém os resultados do segundo estágio. As variáveis explicativas usadas como controle em ambos os modelos são: variáveis binárias para anos e estados brasileiros, educação do chefe da família, gênero e sexo do chefe de família e educação e idade da mãe. Com os dados do Censo Brasileiro de 1991, os autores encontram que tamanho da família apresenta efeito significativo na progressão escolar das meninas, sendo esse efeito negativo. Os autores apresentam evidências de que quanto maior o tamanho da família, menor será a probabilidade da criança progredir na escola, mantendo todas as demais características da família constantes. Como salientado pelos autores, esse resultado é importante para pensarmos nas políticas públicas de transferência de renda que condicionam o montante recebido pelas famílias somente na sua renda e no número de filhos que elas possuem.

Nessas políticas, o efeito do benefício da renda pode ser cancelado pelos efeitos negativos do incentivo ao aumento do tamanho da família.

Tabela 6.4.1: Resultados do primeiro estágio

Variável dependente: número de crianças na família com 10 a 15 anos		
	Meninos	Meninas
Presença de gêmeos	0,852*** (0,028)	0,862*** (0,028)
Teste de F	1.070,76***	1.137,98***
N	139.214	148.106

Tabela 6.4.2: Resultados do segundo estágio

Participação na força de trabalho		
	Meninos	Meninas
Tamanho da família	-0,029 (0,021)	0,019 (0,027)
N	38.928	28.609

Frequência à escola		
	Meninos	Meninas
Tamanho da família	0,013 (0,021)	-0,016 (0,026)
N	38.928	28.609

Progressão na escola		
	Meninos	Meninas
Tamanho da família	-0,005 (0,010)	-0,044*** (0,011)
N	38.928	28.609

6.6. Exercícios

1. Duflo (2001) estima o efeito da construção de escolas na Indonésia no salário dos indivíduos.

Neste artigo, a autora usa um modelo linear para o salário w do indivíduo i nascido na região j e na coorte k que pode ser simplificado pela equação:

$$w_{ijk} = \beta + \delta_1 P_j + \delta_2 C_{ijk} + \delta_4 (P_j x C_{ijk}) + \varepsilon_{ijk}$$

em que P_j é uma variável binária que indica se na região j o programa foi muito intenso. Essa variável é igual a 1 se na região j duas ou mais escolas foram construídas, e é igual a 0 caso contrário. Além disso, C_{ijk} é igual a 1 se o indivíduo estudou em uma escola que foi construída pelo programa, e 0 caso contrário, e ε_{ijk} representa os fatores não observáveis. Esse modelo assume que o efeito médio do tratamento varia com a intensidade do programa.

- a.** Calcule $E[w_{ijk}|C_{ijk}=1, P_j=1]$ e $E[w_{ijk}|C_{ijk}=0, P_j=1]$. Qual seria o efeito médio do programa para os indivíduos que moravam em regiões nas quais o programa teve grande intensidade? Qual o efeito médio do programa para os indivíduos em regiões com baixa intensidade do programa?
- b.** Na tentativa de capturar efeitos heterogêneos, os autores propõem o modelo composto pelo seguinte sistema de equações:

$$w_{ijk} = \alpha + \gamma C_{ijk} + \vartheta_{ijk}$$

$$C_{ijk} = \theta + \pi P_j + \varphi_{ijk}$$

em que $\text{Cov}(\vartheta_{ijk}, \varphi_{ijk}) \neq 0$, $\text{Cov}(P_j, \varphi_{ijk}) = \text{Cov}(P_j, \vartheta_{ijk}) = 0$.

O que o parâmetro γ representa? Quem seriam os *compliers* nesse modelo?

- c.** Como podemos estimar o efeito médio do tratamento para a subpopulação de *compliers* usando o sistema acima?

2. Usamos o seguinte sistema de equações para estimar o retorno da educação sobre o salário,

$$Y_i = \alpha + \rho S_i + \varepsilon_i$$

$$S_i = \theta + \pi Z_i + \vartheta_i$$

Em que S_i representa o número de anos de estudo do indivíduo i , Z_i é uma variável binária que é igual a 1 se o indivíduo i nasceu no primeiro trimestre do ano, e 0 se o indivíduo nasceu no quarto trimestre. Assumimos que $Cov(\varepsilon_i, \vartheta_i) \neq 0$ e $Cov(Z_i, \varepsilon_i) = 0$. Suponha que $Pr[Z_i=1]=p$.

a. Interprete as duas hipóteses $Cov(\varepsilon_i, \vartheta_i) \neq 0$ e $Cov(Z_i, \varepsilon_i) = 0$ no contexto deste exercício.

b. Mostre que $Cov(Y_i, Z_i)=\{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]\}p(1-p)$

c. Mostre que $Cov(S_i, Z_i)=\{E[S_i|Z_i=1]-E[S_i|Z_i=0]\}p(1-p)$

d. Usando os resultados das letras (a) e (b), mostre que

$$\rho = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[S_i|Z_i = 1] - E[S_i|Z_i = 0]}$$

e. Baseado nos estimadores acima, proponha um estimador para p .

f. Angrist e Imbens (1995) usam uma base de dados para os Estados Unidos para estimar o efeito dos anos de escolaridade sobre salário, usando o trimestre de nascimento como instrumento para anos de estudo. A ideia é que as pessoas só podem entrar no ensino básico com 6 anos de idade. Nos Estados Unidos, o ano letivo começa em agosto. Assim, as crianças que nascem no quarto trimestre do ano, têm de esperar um ano a mais para entrar na escola que as crianças que nascem no primeiro trimestre. Essa variação levaria um grupo a permanecer mais tempo na escola do que o outro. Os autores separam os indivíduos em dois grupos. O primeiro grupo é composto pelos indivíduos nascidos no primeiro trimestre, e o segundo, por aqueles nascidos no quarto trimestre. Para cada grupo, eles obtêm as seguintes médias para os alunos de estudo e logaritmo do salário:

	Nascidos no	
	Primeiro trimestre	Quarto trimestre
log (salário)	5,892	5,905
Anos de estudo	12,688	12,839

Com o resultados da tabela acima e o estimador proposto em (e), calcule a estimativa para o efeito médio de anos de estudo no logaritmo do salário. Interprete esse coeficiente como o efeito médio de um tratamento para a subpopulação de *compliers*. Quem são os *compliers* nesse caso?

3. Considere o seguinte modelo com efeitos heterogêneos, em que T_i é uma variável binária que é igual a 1 se o indivíduo recebeu o tratamento, e Z_i é uma variável instrumental binária. O efeito do tratamento sobre o resultado de interesse é representado por:

$$Y_{1i} - Y_{0i} \equiv \rho_i$$

ou, de outra forma, podemos escrever esse modelo, usando o resultado observado para cada indivíduo,

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})T_i = \alpha_0 + \rho_i T_i + \vartheta_i$$

e também assumimos um modelo para o efeito do instrumento sobre o tratamento,

$$T_i = T_{0i} + (T_{1i} - T_{0i})Z_i = \pi_0 + \pi_1 Z_i + \varepsilon_i$$

- a. Mostre que, se tivermos efeitos heterogêneos do instrumento sobre o tratamento, mas tivermos efeito homogêneo do tratamento ρ , o estimador do *LATE* será igual a ρ mesmo que a hipótese de monotonicidade seja violada.
- b. Considere o caso heterogêneo, derive o viés do estimador do *LATE* caso a hipótese de monotonicidade seja violada. Sob quais circunstâncias, as violações da hipótese de monotonicidade não invalidam a análise de efeito médio local do tratamento.

- c. Suponha que as hipóteses do *LATE* sejam verdadeiras, mas

$$E[T_i|Z_i=0]=Pr[T_i=1|Z_i=0]=0$$

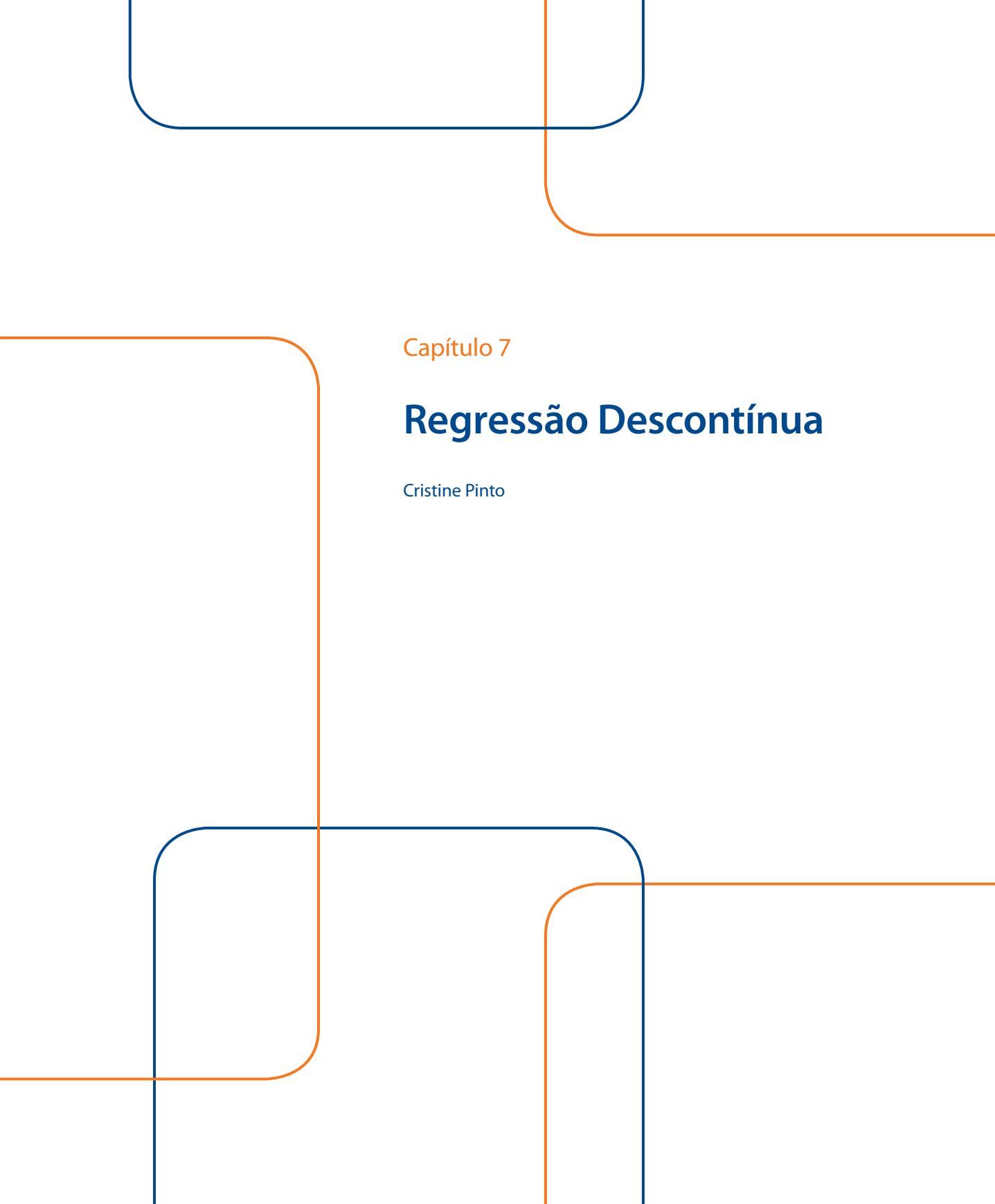
mostre que

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{\Pr [T_i = 1|Z_i = 1]} = E[Y_{1i} - Y_{0i}|T_i = 1]$$

Qual o significado dessa hipótese adicional? Interprete esse resultado e discuta como você iria estimar o efeito médio local do tratamento nesse contexto.

4. Um programa é criado pelo governo para promover o consumo de frutas e vegetais entre os alunos na 5^a série do ensino fundamental. Nas escolas que receberam o programa, os alunos foram convidados a participar de aulas que explicavam cada tipo de alimento e em quais eles podiam experimentar as frutas e vegetais. Além disso, esses alunos também faziam excursões para mercados e feiras na tentativa de conhecer e experimentar um número grande de frutas e verduras. As aulas e as excursões eram atividades extracurriculares, e os pais tinham que autorizar os alunos a participar das mesmas. Dentro do estado, metade das escolas foi selecionada para o grupo de controle, enquanto a outra metade foi selecionada para o grupo de tratamento. Essa seleção foi realizada de forma aleatória. O governo gostaria de avaliar o efeito do programa sobre a saúde das crianças. Para isso, eles possuem uma base de dados que contém algumas características socioeconômicas dos alunos, se eles estavam em uma escola tratada e se ele participou do programa. Além disso, ele possui informações sobre a saúde do aluno, como o número de vezes em que o aluno ficou gripado nos últimos 12 meses.

- a. Proponha um método para estimar o efeito do programa sobre a saúde dos alunos.
- b. Discuta as hipóteses necessárias para que o método proposto em (a) seja válido.
- c. Quais seriam os resultados esperados dessa avaliação? Você acha que os resultados obtidos para esse estado se manteriam se essa política fosse implementada em outros estados?



Capítulo 7

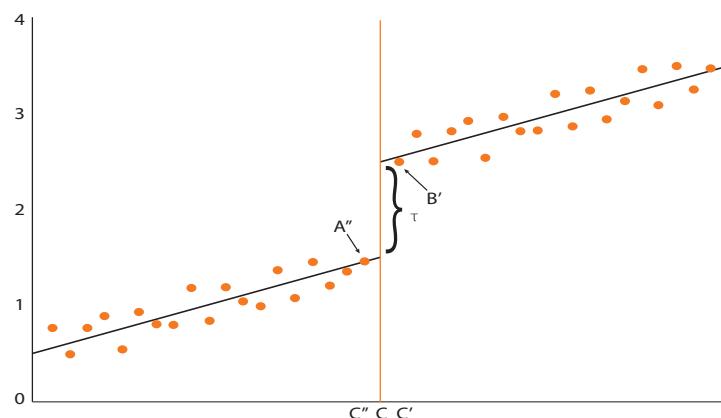
Régressão Descontínua

Cristine Pinto

O método de regressão descontínua (RD) pode ser usado quando a probabilidade de receber tratamento muda de forma descontínua com uma variável, que vamos chamar de Z_i . Um dos primeiros trabalhos que usaram regressão descontínua foi o de Thristlewaite and Cook (1960) que estimaram o impacto de um prêmio por mérito no desempenho acadêmico dos alunos. Nesse estudo, o prêmio por mérito era dado de acordo com as notas dos alunos. Os alunos que tinham notas acima de certo valor recebiam o prêmio, enquanto os alunos que tinham notas abaixo desse ponto de corte não podiam receber o prêmio. Os autores argumentam que os indivíduos que estão logo abaixo do ponto de corte são muito semelhantes aos indivíduos logo acima do ponto de corte, exceto pelo fato que não receberam o prêmio, e seriam um bom grupo de controle. Nesse estudo, a variável Z_i , cujo valor determina o recebimento do tratamento, é a nota do aluno.

Podemos representar o método de regressão descontínua de forma gráfica. Suponha que na figura abaixo o eixo vertical represente a variável de interesse, e no eixo horizontal tenhamos a variável Z . E que indivíduos com valores de Z_i acima de c receberam o programa, enquanto que indivíduos com valores abaixo desse valor não o receberam. A ideia do método é que os indivíduos que tiveram valores de Z_i muito próximos de c são muito parecidos. A única diferença é que alguns tiveram "sorte" e tiveram um Z_i um pouco acima de c , e outros tiveram "azar" e tiveram um Z_i um pouco abaixo de c . Se compararmos a média do Y_i para os indivíduos com Z_i acima do ponto c (ponto B no gráfico) com a média de Y_i para os indivíduos com o Z_i um pouco abaixo de c (ponto A), obtemos o efeito médio do tratamento em torno desse ponto c (no gráfico é representado por τ).

Figura 7.1 – Regressão Descontínua



Uma das vantagens do método de regressão descontínua é que ele requer hipóteses mais fracas do que as hipóteses usadas pelos métodos não experimentais apresentados até aqui. Em vez de assumirmos uma hipótese de independência, como no capítulo anterior, iremos assumir que existe uma relação contínua entre os determinantes do resultado de interesse e a variável Z , e uma relação descontínua entre a participação no tratamento e Z . Com essas hipóteses, a descontinuidade que ocorre na função que relaciona Y a Z pode ser atribuída ao salto que há na participação do tratamento em determinado valor de Z , e o tamanho desse salto pode ser interpretado como efeito médio local do tratamento. A desvantagem do método de regressão descontínua é que estimamos um efeito médio do tratamento, comparando apenas os indivíduos em torno desse ponto de corte (no caso do gráfico no ponto c). Se esses indivíduos forem muito diferentes do restante dos indivíduos na população de interesse, não podemos afirmar que ele seria o efeito médio do tratamento para a população de interesse. Assim, devemos ter cuidado ao tentar extrapolar o efeito médio local encontrado pelo método de regressão descontínua para o restante da população. Como acontecia com o método de variáveis instrumentais, é um método que apresenta uma validade interna alta, estimando o efeito médio local do tratamento, mas uma validade externa baixa.

Como nos capítulos anteriores, representamos o recebimento ou não do tratamento pelo indivíduo i por T_i . Mas, no instrumental de regressão descontínua, essa variável binária T_i é uma função de Z_i . Existem dois casos de regressão descontínua, depende se a descontinuidade na probabilidade de receber tratamento é *fuzzy* ou *sharp*. No caso *sharp*, a participação é uma função determinística de Z_i , isto é, $T_i=1$ se $Z_i \geq c$, e $T_i=0$ se $Z_i < c$. No caso *fuzzy*, ocorre um salto na probabilidade de participação no ponto em que Z_i é igual a c , mas não necessariamente de 0 para 1 como no caso *sharp*. Em ambos os casos, ter a descontinuidade na probabilidade de participação quando Z_i é igual ao ponto de corte c não é suficiente para identificarmos o efeito médio local do tratamento. Precisamos pensar quais hipóteses garantem que essa descontinuidade identifica o efeito médio local do tratamento.

Se os indivíduos têm controle sobre a variável Z_i e eles sabem os benefícios do tratamento, os indivíduos abaixo da descontinuidade podem ser sistematicamente diferentes dos indivíduos acima da descontinuidade. Por exemplo, no estudo de Thristlewaite and Cook (1960) se os estudantes puderem determinar a sua nota no teste de proficiência de forma perfeita por meio do esforço, os alunos mais esforçados e determinados irão escolher notas acima de c e irão ganhar o prêmio, e esses alunos podem ser sistematicamente diferentes dos demais, pelo menos no que se refere ao

esforço. Nesse caso, os indivíduos não são comparáveis em torno do ponto de corte, e não podemos usar os indivíduos abaixo de c como um grupo de controle para os indivíduos logo acima de c que receberam o tratamento. No entanto, suponha outro cenário onde os alunos, mesmos se esforçando mais, não têm garantia de que sua nota será maior que c . Nesse novo cenário, parece mais razoável pensar que os alunos que ficaram logo abaixo do ponto de corte são semelhantes aos que ficaram logo acima, exceto que os primeiros tiveram sorte e acertaram uma questão a mais na prova do que os outros. Mesmo sabendo dos benefícios do tratamento, os alunos não são capazes de manipular de forma perfeita a sua nota de modo que ela fique acima de c . Nesse caso, temos uma variação exógena no tratamento em torno da descontinuidade, e usamos essa variação para estimar o efeito médio do tratamento.

Se a variação no *status* do tratamento em torno do ponto de corte é aleatória, como descrito no caso acima, as características determinadas antes da realização de Z_i devem ter as mesmas distribuições no grupo de indivíduos com valores Z_i acima do ponto de corte e no grupo de indivíduos com valores abaixo do ponto de corte. Como no caso dos métodos de pareamento, podemos verificar se os indivíduos em torno do ponto de corte são semelhantes em suas características observáveis por meio de um teste de comparação de médias. Se tivermos evidências de que, na média, os indivíduos situados em torno do ponto de corte são diferentes, devemos desconfiar da validade da estrutura da regressão descontínua. Além disso, podemos representar a estrutura de uma regressão descontínua por uma análise gráfica. Por exemplo, uma representação gráfica da relação de Y_i e Z_i pode ser usada para nos dar uma ideia do tamanho da descontinuidade, da forma funcional adequada para modelar essa relação etc. Voltaremos à análise gráfica de uma regressão descontínua ao final do capítulo.

Antes de descrevermos os casos *fuzzy* e *sharp*, iremos definir o efeito médio local do tratamento que identificamos dentro da estrutura de uma regressão descontínua, usando o arcabouço de resultados potenciais. O efeito médio local no ponto de descontinuidade c pode ser definido como:

$$EML(c) = E[Y_i(1)|Z_i=c] - E[Y_i(0)|Z_i=c] \quad (7.1)$$

Mas não observamos os resultados potenciais $Y_i(1)$ e $Y_i(0)$ para o mesmo indivíduo. Para cada indivíduo i , observamos:

$$Y_i = Y_i(0) + (Y_i(1) - Y_i(0)) \cdot T_i$$

Suponha um $\varepsilon > 0$, para os indivíduos com valores de Z_i logo acima do ponto de corte podemos definir a seguinte:

$$E[Y_i|Z_i=c+\varepsilon] = E[Y_i(0)|Z_i=c+\varepsilon] + E[(Y_i(1) - Y_i(0)) \cdot T_i|Z_i=c+\varepsilon]$$

e de forma análoga para os indivíduos logo abaixo do ponto de corte:

$$E[Y_i|Z_i=c-\varepsilon] = E[Y_i(0)|Z_i=c-\varepsilon] + E[(Y_i(1) - Y_i(0)) \cdot T_i|Z_i=c-\varepsilon]$$

Definimos Y^+ como o valor que a esperança de Y_i aproxima quando Z_i chega próximo de c para os indivíduos acima do ponto de corte, e Y^- como o valor que a esperança de Y_i aproxima quando Z_i chega próximo de c para os indivíduos abaixo do ponto de corte.¹

A hipótese de continuidade exige que as funções de regressões condicionais dos resultados do tratamento e do controle em Z_i sejam contínuas.² De maneira mais formal,

$E[Y_i(1)|Z_i=z]$ e $E[Y_i(0)|Z_i=z]$ são funções contínuas em z ³ (H1)

1. De modo formal, temos que

$$Y^+ = \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c + \varepsilon]$$

$$= \lim_{\varepsilon \downarrow 0} E[Y_i(0)|Z_i = c + \varepsilon] + \lim_{\varepsilon \downarrow 0} E[(Y_i(1) - Y_i(0)) \cdot T_i|Z_i = c + \varepsilon]$$

$$Y^- = \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c - \varepsilon]$$

$$= \lim_{\varepsilon \downarrow 0} E[Y_i(0)|Z_i = c - \varepsilon] + \lim_{\varepsilon \downarrow 0} E[(Y_i(1) - Y_i(0)) \cdot T_i|Z_i = c - \varepsilon]$$

2. Sob a hipótese de que o resultado de interesse (Y) é uma função contínua de Z , temos que:

$$\lim_{\varepsilon \downarrow 0} E[Y_i(0)|Z_i = c + \varepsilon] = \lim_{\varepsilon \downarrow 0} E[Y_i(0)|Z_i = c - \varepsilon]$$

3. Uma hipótese mais forte que teria como consequência a continuidade das funções de regressão seria assumirmos que as funções de distribuições condicionais são contínuas, isto é, $F_{Y(0)|Z}(y|z)$ e $F_{Y(1)|Z}(y|z)$ em z para todo y .

Se em torno do ponto $Z_i=c$, o *status* do tratamento foi determinado de forma aleatória, temos ignorabilidade local, ou seja:

$$(Y_i(0), Y_i(1)) \perp T_i | Z_i = c \quad (\text{H2})$$

Sob essa hipótese de ignorabilidade⁴,

$$E[Y_i(1) \cdot T_i | Z_i = c] = E[Y_i(1) | Z_i = c] \cdot E[T_i | Z_i = c] \quad (7.2)$$

$$E[Y_i(0) \cdot T_i | Z_i = c] = E[Y_i(0) | Z_i = c] \cdot E[T_i | Z_i = c] \quad (7.3)$$

Sob as hipóteses de ignorabilidade local e continuidade podemos representar a diferença entre $Y_i^+ - Y_i^-$ como:

$$Y_i^+ - Y_i^- = E[Y_i(1) - Y_i(0) | Z_i = c](T^+ - T^-)$$

em que T^+ como o valor que a esperança de T_i aproxima quando Z_i chega próximo de c para os indivíduos acima do ponto de corte, e T^- como o valor que a esperança de T_i aproxima quando Z_i chega próximo de c para os indivíduos abaixo do ponto de corte.

E podemos escrever o efeito médio local de tratamento como

$$EML(c) = E[Y_i(1) | Z_i = c] - E[Y_i(0) | Z_i = c] = \frac{Y^+ - Y^-}{T^+ - T^-} \quad (7.4)$$

A hipótese básica para a identificação do efeito médio local do tratamento é que a probabilidade de participação no tratamento é uma função de Z_i , e no ponto c esta probabilidade sofre um salto. Essa variação no *status* de participação em torno de c pode ser interpretada como uma variação aleatória. Além disso, o resultado é uma função contínua de Z_i , de modo que qualquer descontinuidade em Y_i em torno de c pode ser atribuída somente ao tratamento.

4. Note que como no caso dos métodos de pareamento, a hipótese de ignorabilidade é mais forte do que a hipótese mínima necessária para identificar o efeito médio local do tratamento. As hipóteses mais fracas necessárias para identificação seriam as equações (7.2) e (7.3).

7.1. Caso sharp

Nesse caso, T_i é uma função determinística de Z_i . Todos os indivíduos com um valor de Z_i acima de c recebem o tratamento, e todos os indivíduos com valores abaixo desse valor não recebem. A probabilidade de receber o tratamento muda de 0 para 1 ao passar o ponto de corte. Nesse caso, sob as hipóteses de continuidade e ignorabilidade local, não existem outros fatores que expliquem a descontinuidade de Y_i em c a não ser o recebimento do tratamento, e interpretamos o salto de T_i no ponto c como o efeito causal do tratamento. Nesse caso, $T^+=1$ e $T^-=0$, e o efeito médio local do tratamento é:

$$EML^S(c) = Y^+ - Y^- \quad (7.1.1)$$

Nesse caso, para os indivíduos com o mesmo valor de Z_i não existe variação no tratamento. Os indivíduos com valores de Z_i iguais ou acima de c terão $T_i=1$; e os indivíduos com valores abaixo de c , sempre terão $T_i=0$.

O gráfico (7.1.1) ilustra a estrutura da regressão descontínua no caso *sharp*. A figura 1 representa a probabilidade de receber tratamento em função de Z_i e ilustra que esta probabilidade muda de 0 para 1 quando Z_i é igual ao valor do ponto de corte. A figura 2 representa as curvas dos resultados potenciais caso o indivíduo receba tratamento ($Y_i(1)$) e os resultados potenciais caso o indivíduo não receba tratamento ($Y_i(0)$). Só observamos os pontos da função de $Y_i(0)$ para indivíduos com valores de Z_i abaixo de c , e os pontos da função de $Y_i(1)$ para os indivíduos com valores de Z_i acima de c . Se essas funções são contínuas em Z_i , podemos atribuir o salto que acontece no ponto c como o efeito médio do tratamento em torno dessa descontinuidade.

Gráfico 7.1.1: Regressão descontínua sharp

Figura 1: Probabilidade condicional de receber tratamento

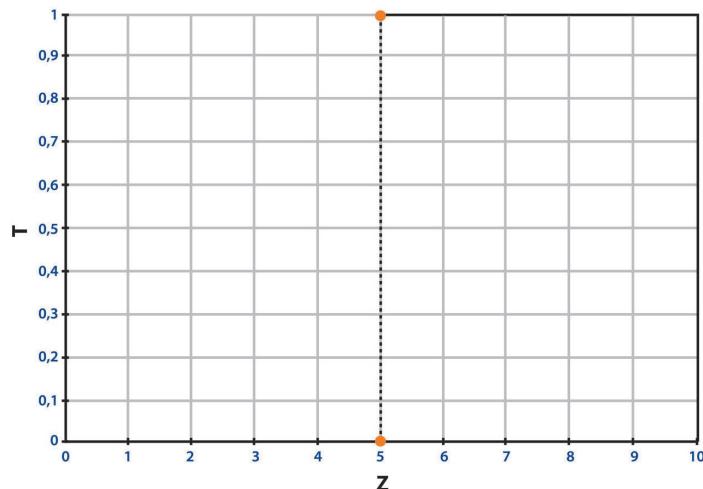
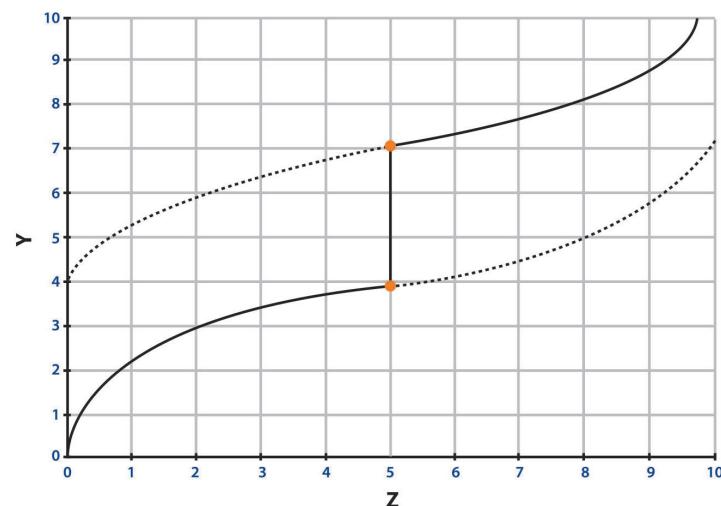


Figura 2: Expectativa condicional dos resultados



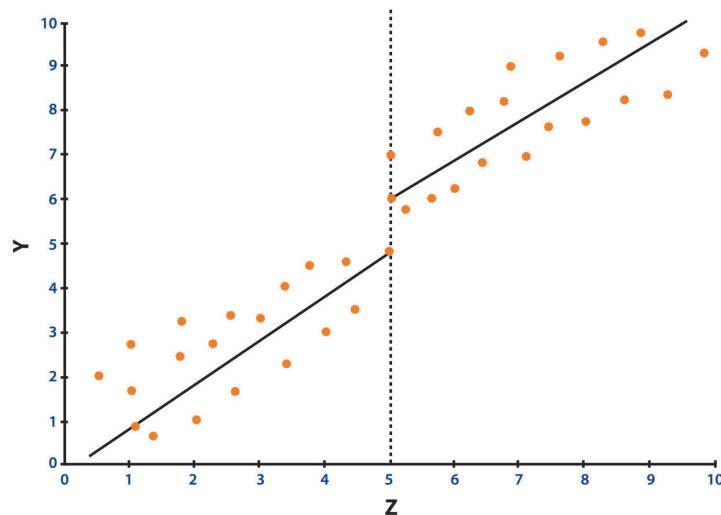
Suponha o caso simples em que a relação entre Y_i e Z_i é uma função linear, e queremos identificar o efeito médio do tratamento. Podemos representar o caso *sharp* de regressão descontínua pelo sistema de equações:

$$Y_i = \alpha + \tau T_i + \beta Z_i + \varepsilon_i$$

$$T_i = \begin{cases} 1 & \text{se } Z_i \geq c \\ 0 & \text{caso contrário} \end{cases}$$

Nesse caso, a variável de resultado é uma função linear de T_i , e T_i é igual a 1 se Z_i for maior que o ponto de corte c . O gráfico (7.1.2) ilustra esse problema.

Gráfico 7.1.2 - O caso linear



O salto na variável Y_i no ponto c pode ser interpretado como o efeito causal do tratamento se todos os outros fatores que afetam Y_i estão evoluindo de forma suave em torno de c^5 .

5. De maneira formal, esse salto irá representar τ se $\lim_{\varepsilon \downarrow 0} E[\varepsilon_i | Z_i = c + \varepsilon] = \lim_{\varepsilon \downarrow 0} E[\varepsilon_i | Z_i = c - \varepsilon]$

7.2. Caso fuzzy

Esse é o caso mais comum em avaliações de programas sociais, e pode ser entendido como um caso especial do *LATE* apresentado no capítulo 6. Nesse caso, o tratamento é determinado parcialmente pela descontinuidade em Z_i . A probabilidade de receber tratamento não muda de 0 para 1 no ponto de corte c , mas acontece apenas um salto na probabilidade de receber tratamento no ponto em que $Z_i=c$.

Como a probabilidade de receber tratamento tem um salto menor que 1, o salto da expectativa condicional de Y_i em Z_i , que acontece no ponto em que Z_i é igual a c , não pode ser interpretado como efeito médio do tratamento. Nesse caso, voltamos à equação (7.4) e identificamos o efeito médio do tratamento como a razão entre o salto que ocorre em Y_i e o salto que ocorre em T_i no ponto de corte,

$$EML^F(c) = \frac{Y^+ - Y^-}{T^+ - T^-} \quad (7.2.1)$$

Um exemplo desse caso seria a situação em que o recebimento do prêmio por mérito não depende somente das notas dos alunos no exame de proficiência, mas também de uma entrevista com uma banca examinadora. Nesse caso, uma nota acima de c aumenta a probabilidade de receber o prêmio, mas alunos que tiraram uma nota abaixo de c podem receber o prêmio desde que façam uma boa entrevista.

Para a identificação da razão das descontinuidades como efeito médio local do tratamento, precisamos de hipóteses semelhantes às hipóteses usadas no estimador de variável instrumental, *LATE*. Esse parâmetro identifica o efeito médio do tratamento para os indivíduos que mudaram o seu *status* de participação quando o valor do instrumento passa do ponto c . Como visto no capítulo anterior, para identificar o efeito médio local do tratamento representado pelo *LATE* precisávamos de hipóteses de monotonicidade, independência do instrumento e de restrição de exclusão. Vamos tentar entender o estimador de RD no caso fuzzy como um *LATE*. Para isso, vamos definir $T_i(z)$ como o *status* potencial de tratamento no ponto em que $Z_i=z$, para z em uma pequena vizinhança de c . $T_i(z)$ é igual a 1 se o indivíduo i recebe o tratamento no ponto em que $Z_i=z$. Usando esse tratamento potencial, podemos expressar a hipótese de monotonicidade como:

$T_i(z)$ é uma função não decrescente em z (H3)

Dentro do instrumental do **LATE**, definimos como *compliers* os indivíduos que têm o seu comportamento alterado ao ter valores de Z_i abaixo ou acima do ponto de corte, isto é, um indivíduo i é denominado de *complier* se $T^+=1$ e $T^-=0$. Os *compliers* são definidos como os indivíduos que recebem o tratamento se o valor de Z_i é igual ou acima de c , mas não recebem o tratamento se o valor de Z_i está abaixo de c . Considere o exemplo em que os indivíduos que recebem uma nota no teste de proficiência abaixo de c são encorajados a fazer aulas de reforço escolar no contraturno. O *complier* é o indivíduo que irá participar das aulas de reforço se a sua nota ficar abaixo de c , e não irá participar se a nota for igual ou maior que c . Como acontecia no caso do **LATE**, a hipótese de monotonicidade elimina o grupo de indivíduos que receberia o tratamento se o valor de Z_i fosse abaixo do ponto de corte e não receberia caso o valor de Z_i fosse acima do ponto de corte, isto é, essa hipótese elimina o que chamamos de *defiers*. Logo, o efeito médio local do tratamento em torno da descontinuidade pode ser decomposto no efeito para os *compliers*, no efeito para os indivíduos que sempre aceitam o tratamento, independentemente do valor de Z_i (*always-takers*) e no efeito para os indivíduos que sempre recusam o tratamento independentemente do valor de Z_i (*never-takers*).

No entanto, como acontecia no caso do **LATE** com a hipótese de restrição de exclusão, a hipótese de continuidade da esperança dos resultados potenciais irá garantir que em torno da continuidade, estamos captando o efeito do tratamento apenas para os *compliers*. Sob as hipóteses de independência, continuidade e monotonicidade, o estimador de regressão descontínua para o caso *fuzzy* identifica o efeito médio local do tratamento para os *compliers*.

Da mesma maneira que no caso clássico de variável instrumental, temos que verificar se existe uma relação forte entre o instrumento e a variável endógena; no caso da regressão descontínua, temos que verificar se existe uma descontinuidade na probabilidade de T_i condicional a Z_i . O gráfico (7.2.1) ilustra o instrumental de uma regressão descontínua *fuzzy*. Na primeira figura temos o salto que acontece na probabilidade de receber tratamento. Notamos que essa probabilidade não salta de 0 para 1 como acontecia no caso *sharp*. A segunda figura mostra o salto que acontece no resultado potencial, e é semelhante ao gráfico para o caso *sharp*. A diferença é que não podemos dizer que o tamanho desse salto é o efeito causal do tratamento. Nesse caso, o efeito médio local do tratamento é dado pela razão entre os saltos da figura 1 e da figura 2.

Gráfico 7.2.1: Regressão descontínua fuzzy

Figura 1: Probabilidade condicional de receber tratamento

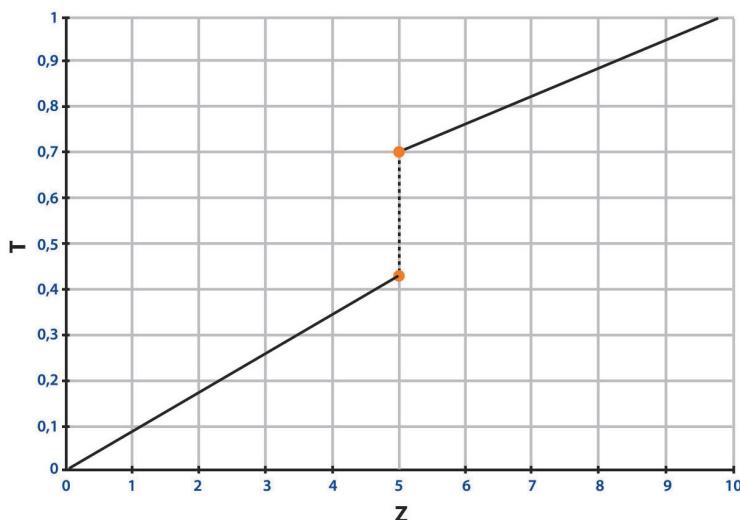
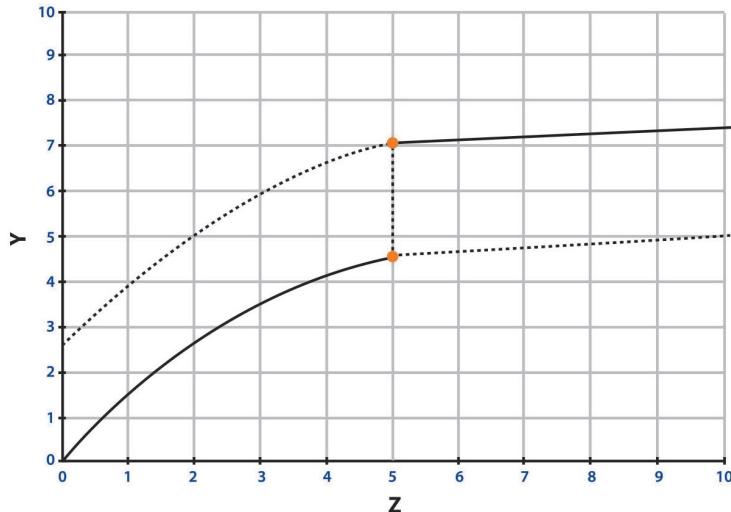


Figura 2: Expectativa condicional dos resultados



7.3. Estimação

Uma maneira simples de implementar o método de regressão descontínua é usar regressões lineares locais separadas para estimar as relações de um lado e de outro do ponto de corte⁶. No caso *sharp*, precisamos estimar apenas a relação de Y_i com Z_i dos dois lados em torno do ponto de corte c . O modelo de regressão do lado esquerdo do ponto de corte é:

$$Y_i = \alpha_l + \beta_l \cdot (Z_i - c) + \varepsilon_i \text{ se } c - h \leq Z_i < c \quad (7.3.1)$$

e o modelo de regressão do lado direito é:

6. Nesta seção, apresentamos o método mais comumente usado para estimar a regressão descontínua. Existem outros métodos não paramétricos que podem ser usados, como regressão por Kernel. Para a revisão de alguns métodos que podem ser implementados, ver Imbens e Lemieux (2007) e Lee e Lemieux (2009).

$$Y_i = \alpha_r + \beta_r \cdot (Z_i - c) + \varepsilon_i \text{ se } c \leq Z_i < c + h \quad (7.3.2)$$

Como as variáveis explicativas estão subtraídas do valor no ponto de corte, o valor esperado de Y de cada regressão no ponto de corte é dado por α_l e α_r . O efeito médio local do tratamento será a diferença entre o intercepto das duas regressões, $\tau = \alpha_r - \alpha_l$. Uma forma mais direta de estimar o efeito médio do tratamento seria usar um modelo de regressão que combina os modelos em ambos os lados do ponto de corte:

$$Y_i = \alpha_l + \tau \cdot T_i + \beta_l \cdot (Z_i - c) + \gamma \cdot T_i \cdot (Z_i - c) + \varepsilon_i, \text{ se } c - h \leq Z_i < c + h \quad (7.3.3)$$

no qual $\gamma = \beta_r - \beta_l$.

A vantagem desse modelo que combina os efeitos à direita e à esquerda do ponto de corte é que o efeito médio do tratamento e o seu erro-padrão são estimados de forma direta em uma única regressão.

Dada uma vizinhança h , é fácil estimar o modelo de regressão representado pela equação (7.3.3). Mas para estimar essa regressão, precisamos escolher o h . Se escolhermos um h muito grande, iremos usar um número maior de observações na estimação, o que aumenta a precisão das estimativas. No entanto, quanto maior o h , maior a chance de a especificação linear não ser a correta para estimar a relação entre Y_i e Z_i , e podemos aumentar o viés do efeito médio do tratamento. Se a expectativa condicional de Y_i não é uma função linear em Z_i , o modelo linear pode ser uma boa aproximação em uma região limitada de Z_i , mas à medida que aumentarmos a região de valores de Z_i , o modelo linear pode tornar uma aproximação muito ruim. Da mesma forma, quando menor o h , maior a chance de a especificação linear ser uma aproximação adequada para estimar a relação entre Y_i e Z_i naquela vizinhança, mas menor a precisão das estimativas. O h ideal é aquele que balanceia de forma ótima esse *trade off* entre viés e precisão. Existem várias regras usadas para escolher essa janela h . Se o número de observações aumentar, podemos diminuir h , pois teremos maior número de observações por intervalo.

Existem dois procedimentos mais comumente usados para escolher essa janela ótima⁷. O primeiro seria usar uma janela que leva em consideração alguns componentes da distribuição conjunta de (Y_i, Z_i) , como a curvatura do modelo de regressão. A regra de bolso usada para a escolha de h é:

$$h^* = 2,702 \cdot \left(\frac{\tilde{\sigma}^2 R}{\sum_{i=1}^N \{f''(z_i)\}^2} \right)^{1/5} \quad (7.3.4)$$

em que $\tilde{\sigma}$ é o erro-padrão estimado da regressão, $f''(\cdot)$ é a segunda derivada do modelo de regressão estimado para captar a relação em Y_i e Z_i , e R representa o intervalo da variável Z_i usada para estimar a regressão. Para usar essa regra de bolso, primeiramente precisamos estimar um modelo de regressão arbitrário em uma vizinhança em torno do ponto de corte. Por exemplo, suponha que nessa primeira etapa, estimamos uma regressão quadrática para estimar a relação entre Y_i e Z_i , $f(z_i) = \beta_0 + \beta_1 z_i + \beta_2 z_i^2 + \beta_3 z_i^3 + \beta_4 z_i^4$. Nesse caso, após a estimação da regressão, precisamos calcular $f''(z_i) = 2\hat{\beta}_2 + 6\hat{\beta}_3 z_i + 12\hat{\beta}_4 z_i^2$, e estimarmos $\tilde{\sigma}^2$ usando a soma do quadrado dos resíduos dividido pelo número de observações, subtraindo os graus de liberdade. Se usarmos as observações na vizinhança, $c-5 \leq Z_i < c+5$, R será igual a 10.

O segundo procedimento é conhecido como validação cruzada (*Cross-Validation*)⁸. Para cada observação i , estimamos uma regressão, a qual a observação i é deixada de fora e usamos os valores estimados nessa regressão para obter o valor predito de Y_i quando Z_i é igual a z_i . Como estamos dentro do instrumental de regressão descontínua, na estimação da regressão consideramos somente as observações em uma vizinhança h de z_i , isto é, $z_i - h \leq Z_i < z_i + h$. Fazemos esse exercício para cada observação, e obtemos um conjunto de valores preditos para Y_i . A janela ótima será aquela que minimizará a média dos quadrados das diferenças em cada Y_i predito e o Y_i observado. De maneira formal, para determinado valor de h , calculamos o valor de Y_i , $\hat{Y}^h(z_i)$, e calculamos o erro quadrático médio:

7. Em todos os procedimentos descritos nesta seção, estamos escolhendo uma janela única para ambos os lados do ponto de corte. A função de densidade de Z_i é provavelmente similar em ambos os lados do ponto de corte, e com uma amostra grande a janela ótima para o lado direito do ponto de corte será igual à janela do lado esquerdo do ponto de corte. No entanto, podemos aplicar os procedimentos descritos para as observações somente à direita de c e para as observações somente à esquerda de c , obtendo janelas ótimas diferentes para o lado esquerdo e para o lado direito.

8. Neste capítulo, iremos descrever o procedimento proposto por Imbens e Lemieux (2008).

$$CV_Y(h) = \frac{\sum_{i=1}^N (Y_i - \hat{Y}^h(z_i))^2}{N} \quad (7.3.5)$$

e o h ótimo será aquele que minimizará esse critério da validação cruzada:

$$h_{CV}^{Y*} = \operatorname{argmin}_h CV_Y(h) \quad (7.3.6)$$

Na prática, calculamos o erro quadrático médio do critério de validação cruzada para uma série de valores fixos de h e escolhemos o valor de h relacionado ao menor erro quadrático médio.

No caso da regressão descontínua *fuzzy*, temos que estimar também como a probabilidade de receber tratamento está relacionada com Z_i . Nesse caso, podemos usar um modelo de regressão linear que relaciona a variável T_i ao vetor Z_i . Da mesma maneira que fizemos anteriormente, podemos usar um modelo para as observações que estão à direita do ponto de corte e um modelo para as observações à esquerda do ponto de corte:

$$T_i = \gamma_l + \delta_l \cdot (Z_i - c) + u_i \text{ se } c - h \leq Z_i < c \quad (7.3.7)$$

$$T_i = \gamma_r + \delta_r \cdot (Z_i - c) + u_i \text{ se } c + h \leq Z_i < c \quad (7.3.8)$$

E nesse caso, o efeito médio do tratamento será dado pela razão entre a diferença dos interceptos: $\tau_{FD} = \frac{\alpha_r - \alpha_l}{\gamma_r - \gamma_l}$. Podemos usar o mesmo critério de validação cruzada para estimar a janela ótima no caso em que a variável dependente é T_i . Nesse caso, usaremos duas janelas diferentes na estimativa do efeito médio de tratamento na regressão descontínua *fuzzy*. A primeira janela seria aquela que minimiza $CV_Y(h)$, e a segunda seria aquela que minimiza $CV_T(h)$. Imbens e Lemieux (2008) argumentam que na prática podemos usar a mesma janela para o numerador e para o denominador. Para minimizar o viés assintótico, podemos usar a menor janela escolhida por validação cruzada:

$$h_{CV}^{*} = \min(\operatorname{argmin}_h CV_Y(h), \operatorname{argmin}_h CV_T(h)) \quad (7.3.9)$$

Em alguns casos, pode existir um vetor de outras variáveis explicativas além de Z_i que estão correlacionadas com a variável de interesse. A inclusão dessas outras variáveis no modelo não

pode afetar a estimativa da descontinuidade, se todas as hipóteses da estrutura de uma regressão descontínua forem válidas. Se os valores estimados mudarem com a inclusão dessas variáveis adicionais, podemos desconfiar que outros fatores também saltam em torno do ponto de corte, e que o salto que ocorre em Y_i em torno de c não pode ser atribuído somente ao efeito do tratamento. Além disso, se essas demais variáveis explicativas forem correlacionadas com Z_i , elas podem aumentar a precisão das estimativas ao diminuir o resíduo da regressão.

Além da estimativa, é necessário testar hipóteses para verificar se o efeito médio do tratamento é estaticamente significativo. No caso de uma regressão descontínua *sharp*, podemos estimar a equação (7.3.3) por mínimos quadrados ordinários e usar o erro-padrão estimado nessa regressão. No caso de uma regressão descontínua *fuzzy*, podemos estimar o erro-padrão de forma fácil, se entendermos esse modelo como um caso particular do estimador de mínimos quadrados ordinários em dois estágios. Usando a mesma janela para o modelo que relaciona Y_i com Z_i e para o modelo que relaciona T_i com Z_i , o sistema de equação que contém as duas equações que combinam os modelos para as observações à direita e à esquerda do ponto de corte é dado por:

$$Y_i = \alpha_l + \tau \cdot T_i + \beta_l \cdot (Z_i - c) + \gamma \cdot T_i \cdot (Z_i - c) + \varepsilon_i, \text{ se } c - h \leq Z_i < c + h \quad (7.3.10)$$

$$T_i = \gamma_l + \theta \cdot D_i + \delta_l \cdot (Z_i - c) + \pi \cdot D_i \cdot (Z_i - c) + u_i, \text{ se } c - h \leq Z_i < c + h \quad (7.3.11)$$

no qual D_i indica se Z_i excede o valor c , isto é, D_i é uma variável binária igual a 1 se Z_i é maior que c e 0, caso contrário.

O estimador de MQO em dois estágios para o sistema acima, usando D_i como instrumento para T_i é numericamente igual à razão que representa o efeito médio do tratamento para o caso *fuzzy* se usarmos a mesma janela para as duas equações do modelo. Nesse caso, para fazermos inferência podemos usar os erros-padrões estimados pelo método de mínimos quadrados em dois estágios⁹.

9. Imbens e Lemieux (2008) computam as variâncias assintóticas para os estimadores e propõem um método para estimá-las de forma consistente.

7.4. Análise gráfica

O método de regressão descontínua requer uma análise gráfica. Podemos verificar várias hipóteses de uma estrutura de regressão descontínua por meio desse tipo de análise. Em uma regressão descontínua, o efeito médio do tratamento está relacionado ao valor da descontinuidade no valor esperado de Y em certo ponto. O primeiro gráfico a ser feito é um histograma que relaciona o valor médio de Y com Z . Nesse gráfico, temos que observar se em torno do ponto de corte há evidência de uma descontinuidade na média condicional de Y . Além disso, devemos verificar se não existem outros saltos na expectativa condicional de Y em relação a Z . Se esses outros saltos existirem e forem grandes quando comparados ao salto que ocorre em torno do ponto de corte c , podemos desconfiar que outros fatores além da participação no tratamento estão influenciando a relação entre Y e Z .

O segundo gráfico seria um histograma que relaciona o valor de outras variáveis explicativas X que poderiam estar explicando a relação entre Y e Z . A ideia é que se a descontinuidade de Y em torno do ponto de corte se deve somente ao tratamento, essas demais variáveis X devem ter uma relação suave com Z , não apresentando saltos em torno desse ponto de corte c . No caso da regressão descontínua fuzzy, é necessário também fazer um gráfico dos valores médios de T em relação a Z para mostrar que de fato existe um salto na probabilidade de receber tratamento em torno do ponto de corte, c .

Além disso, no modelo de regressão descontínua é importante testar se as pessoas não se alocaram de forma endógena em torno do ponto de corte. Por exemplo, temos que testar se as pessoas, sabendo do tratamento, não manipularam Z_i de modo a ficar abaixo ou acima do ponto c . McCrary(2008) propôs um teste indireto para esse problema baseado em uma análise gráfica. Nesse teste, fazemos um gráfico em que dividimos o suporte de Z em intervalos e estimamos o número de observações em cada intervalo. Com esse gráfico, avaliamos se o número de indivíduos muda de forma descontínua em torno do ponto de corte. Se houver uma mudança descontínua do número de indivíduos em torno do ponto de corte, podemos desconfiar que os indivíduos manipularam o valor de Z , o que invalidaria a regressão descontínua. Por exemplo, no caso em que o prêmio por mérito é determinado pela nota em um teste de proficiência. Se os indivíduos conseguissem escolher a sua nota no teste, iríamos observar um número maior de indivíduos logo acima do ponto de corte do que logo abaixo dele.

7.5. Comparação da RD com os outros métodos

Podemos entender a regressão descontínua no caso *sharp* como uma aplicação muito particular da hipótese de seleção nos observáveis (H1), usada pelos métodos apresentados no capítulo 5. No caso da regressão descontínua *sharp*, a hipótese de seleção nos observáveis é trivialmente satisfeita, pois condicional em Z_i , o tratamento é completamente determinístico. Nesse caso, para todos os indivíduos com o valor de Z_i abaixo de c , T_i é sempre igual a 0, e para os indivíduos com valores de Z_i acima de c , T_i é sempre igual a 1. Ou seja, condicional em Z_i , não há variação no tratamento, e a seleção nos observáveis é sempre satisfeita. No entanto, a segunda hipótese, usada pelos métodos baseados em seleção nos observáveis, é a de sobreposição da região de valores de Z_i tanto no grupo de tratados como no grupo de controles, $0 < Pr[T_i=1|Z_i] < 1$. Essa hipótese é violada no método de regressão descontínua, pois os indivíduos tratados terão valores de Z_i acima de c enquanto os indivíduos não tratados terão valor de Z_i abaixo de c . Para todos os valores de z , a $Pr[T_i=1|Z_i]$ é sempre 1 ou 0, e nunca está entre esses valores. Ao invés de assumirmos a hipótese de sobreposição, assumimos a hipótese de continuidade das funções de expectativa condicional de Y_i em Z_i para o grupo de controle e tratamento. Essa hipótese de continuidade garante que podemos comparar essas duas funções ao longo de todos os valores de Z_i . A ideia da regressão descontínua no caso *sharp* é que os indivíduos que possuem valores de Z_i em torno do ponto de corte c são semelhantes na média, exceto que, por sorte, um indivíduo teve valor Z_i acima de c , e o outro teve valor de Z_i abaixo de c . Logo, podemos pensar na regressão descontínua como forma-limite de pareamento, que acontece em torno de um único ponto.

Outra forma de interpretar a regressão descontínua *sharp* é como um experimento aleatório local. A ideia da regressão descontínua é que em torno do ponto de corte, o *status* do tratamento é selecionado de forma aleatória. Dessa maneira, os indivíduos em torno do ponto de corte são semelhantes na média em todas as suas características, exceto que por sorte um grupo obteve valor de Z_i um pouco maior que c , e o outro grupo valor um pouco menor. Nesse sentido, dentre todos os métodos apresentados neste livro, o método de regressão descontínua é o que mais se aproxima de um experimento real (aleatório).

Podemos entender a regressão descontínua *fuzzy* dentro do ferramental de variável instrumental. Nesse caso, a probabilidade de receber tratamento salta no ponto de corte, mas não de 0 para 1

como no caso *sharp*. No caso *fuzzy*, usamos a variável binária que indica se Z_i é maior que o ponto de corte como instrumento para o tratamento, dado que estamos assumindo que a probabilidade muda em torno de c somente devido ao tratamento, e não aos demais fatores que podem estar afetando o resultado.

7.6. Exemplo

Ferraz e Finan (2008) usam o método de regressão descontínua para estimar o efeito de uma possível reeleição em corrupção. De acordo com os autores, os prefeitos que têm incentivo a se reelegerem são, em média, menos corruptos que os prefeitos que não tem a possibilidade de reeleição. Para estimar esse efeito, eles compararam municípios em que o prefeito incumbente ganhou as eleições de 2000 com uma margem muito pequena de votos com municípios nos quais o prefeito incumbente perdeu as eleições por uma quantidade pequena de votos. Ou seja, eles compararam prefeitos que tentaram a reeleição e tiveram a votação em torno de 50%. No entanto, um deles ficou com um percentual de votos um pouco abaixo de 50 e perdeu a eleição; enquanto o outro ficou um percentual um pouco acima de 50 e venceu a eleição. A ideia é de que esses prefeitos são, na média, muito parecidos em características que acarretam corrupção e capacidade de reeleição (como preferências ideológicas). A única diferença é que, por sorte, um obteve votação pouco maior que o ponto de corte, enquanto o outro obteve votação um pouco menor. Para estimar o efeito da reeleição para esses municípios, que tiveram prefeitos com percentual de votos muito próximo a 50%, os autores estimam o seguinte modelo:

$$R_i = \beta I_i + f(Z_i) + X_i \varphi + W_i \gamma + \varepsilon_i$$

em que R_i mede o nível de corrupção no município i , Z_i representa a diferença entre o percentual de votos do prefeito incumbente e o segundo colocado, I_i é uma variável binária que é igual a 1 se $Z_i \leq 0$, e 0, caso contrário, $f(Z_i)$ representa uma função contínua e suave da margem de votos, como um polinômio. Além disso, X_i é o vetor que contém as características do município, por exemplo, o coeficiente de Gini, população etc.; e o vetor W_i inclui características do município que afetam o nível de corrupção, como número de partidos políticos no legislativo, quantidade de recursos repassada ao município etc.

Usando os dados de auditorias realizadas pelo Tribunal de Contas da União (TCU) em 496 municípios brasileiros, os autores estabeleceram uma medida de corrupção por meio da classificação das irregularidades descritas nestes relatórios em três categorias: (i) fraudes na contratação ou compra de bens e serviços; (ii) superfaturamento de bens e serviços; (iii) desvios de fundos públicos para pessoas ou firmas privadas. Baseados nessas irregularidades classificadas, eles definiram uma medida de corrupção que seria o total de recursos relativo a essas atividades irregulares expresso como a fração do total de recursos auditados. Para obter as características dos municípios, os autores usaram os dados fornecidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE), Tribunal Superior Eleitoral (TSE) e Tesouro Nacional.

A tabela (7.6.1) apresenta o principal resultado da regressão descontínua *sharp* estimada pelos autores. A coluna (1) dessa tabela apresenta o caso para a especificação linear, isto é, $f(Z_i) = Z_i \alpha$. As colunas (2) e (3) usam um polinômio quadrático em Z_i e outro cúbico, respectivamente. Os resultados indicam que a reeleição tem efeito negativo na corrupção.

Tabela 7.6.1: O efeito da reeleição na corrupção

Variável dependente: % de recursos auditados envolvendo corrupção			
	Linear (1)	Quadrática (2)	Cúbica (3)
Prefeito no primeiro mandato	-0,031** [0,014]	-0,040** [0,019]	-0,038** [0,022]
Número de observações	328	328	328
R ²	0,27	0,27	0,27

Para verificar se, de fato, existe descontinuidade na medida de corrupção em torno do ponto de corte da margem de votos, os autores fazem o gráfico da função da expectativa condicional de R_i em Z_i . O gráfico (7.6.2) apresenta tal relação e comprova a descontinuidade da medida de corrupção em torno do ponto de corte. Usando o instrumental de regressão descontínua, os autores atribuem essa descontinuidade ao efeito da reeleição.

Gráfico 7.6.2: Corrupção vs. margem de votos

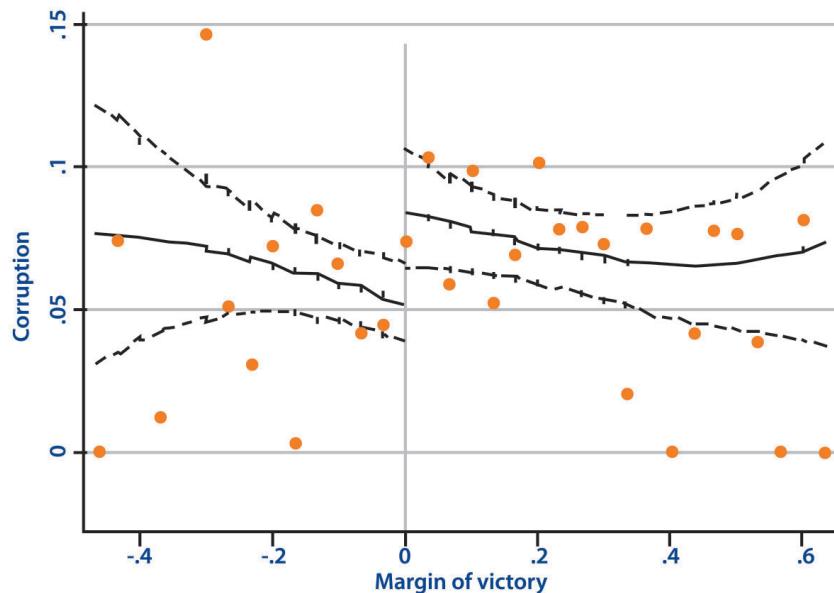


Figure 2: The Effects of Re-Election Incentives on Corruption

7.7. Exercícios

- Suponha que os estados brasileiros mudem a sua política educacional, e os alunos para progredirem do ensino fundamental para o ensino médio são obrigados a fazer uma prova estadual de matemática e português. Essa prova é aplicada a todos os alunos em todas as escolas públicas estaduais no último ano do ensino fundamental, e para progredir eles devem acertar no mínimo 50% das questões em cada uma das provas. Se o aluno não passar em um dos exames ou em ambos, ele deve cursar aulas de recuperação no período de férias e tentar o exame novamente após o período de recuperação. O objetivo é avaliar se o curso de recuperação está tendo impacto no desempenho dos alunos. Para isso, observamos uma base de dados que contém as informações de todos os alunos no último ano do ensino fundamental em 2004 e as

notas que eles obtiveram em cada um dos testes. Além disso, observamos o desempenho de todos os alunos em um teste padronizado no último ano do ensino médio.

Para avaliar o programa de recuperação, propomos o seguinte modelo

$$Y_i = \beta X_i + \eta T_i + \varepsilon_i$$

em que Y_i representa o desempenho do aluno no teste padronizado ao final do ensino médio, T_i é uma variável binária que é igual a 1 se o aluno participou do curso de recuperação, X_i é um vetor de características socioeconômicas do indivíduo i , e ε_i é o fator não observável do modelo.

- a.** Discuta o viés do estimador de mínimos quadrados ordinário no modelo acima.
 - b.** Como você usaria o desenho do programa para estimar o efeito médio da política de recuperação por uma regressão descontínua. Seria uma regressão descontínua *fuzzy* ou *sharp*?
 - c.** Discuta dentro do contexto da política em questão, as hipóteses necessárias para a validade do método de regressão descontínua proposto em (b).
- 2.** O governo gostaria de saber se uma política que expandisse o acesso a planos de saúde teria um efeito positivo na saúde dos indivíduos, pois os levaria a ir mais vezes ao médico. Para responder a tal pergunta, usamos como medida de acesso ao serviço de saúde o número de vezes que o indivíduo i foi ao médico nos últimos 12 meses. O modelo que gostaríamos de estimar é:

$$Y_i = \beta_0 + \beta_1 C_i + \gamma X_i + \varepsilon_i$$

em que Y_i é o número de vezes que o indivíduo i foi ao médico, C_i é uma variável binária que é igual a 1 se o indivíduo i possui plano de saúde, X_i é um vetor de características socioeconômicas do indivíduo i , e ε_i é o termo não observável da equação.

Nos Estados Unidos, ao completar 65 anos, todos os indivíduos têm acesso a um plano de saúde completo que é pago pelo governo. Card, Dobkin e Maesta (2008) usam a descontinuidade quando o indivíduo completa 65 anos para estimar o efeito do plano de saúde no número de consultas

médicas. A ideia é que os indivíduos com 64 anos são semelhantes aos indivíduos com 65 anos, exceto que o segundo grupo tem plano de saúde completo provido pelo governo.

- a.** Alguns indivíduos se aposentam aos 65 anos e têm queda da renda anual. Como isso poderia causar um viés no estimador de regressão descontínua?

Considere o seguinte modelo que relaciona idade com acesso a plano de saúde:

$$C_i = \alpha_0 + \delta X_i + \pi m_i + \omega_i$$

em que m_i é uma variável binária que assume valor igual a 1 se o indivíduo tem 65 anos ou mais, e 0 caso contrário.

- b.** Qual a interpretação de π ?

- c.** Suponha que $\beta_1 > 0$ e $\pi > 0$, faça um diagrama que ilustre a relação entre o resultado de interesse (Y_i) e a idade.

- d.** Suponha que $\beta_1 > 0$ e $\pi < 0$, faça um diagrama que ilustre a relação entre o resultado de interesse (C_i) e a idade.

- e.** As hipóteses do arcabouço de regressão descontínua são válidas nesse exemplo?
Discuta cada uma delas.

- 3.** Considere o seguinte modelo:

$$Y_i = \alpha_l + \tau \cdot T_i + \beta_l \cdot (Z_i - c) + \gamma \cdot T_i \cdot (Z_i - c) + \varepsilon_i, \text{ se } c - h \leq Z < c + h \quad (7.3.10)$$

$$T_i = \gamma_l + \theta \cdot D_i + \delta_l \cdot (Z_i - c) + \pi \cdot D_i \cdot (Z_i - c) + u_i, \text{ se } c - h \leq Z < c + h \quad (7.3.11)$$

em que Y é o resultado de interesse, T é uma variável indicadora de tratamento, Z seria um instrumento que influencia a participação ou não no programa e D é uma variável binária igual a

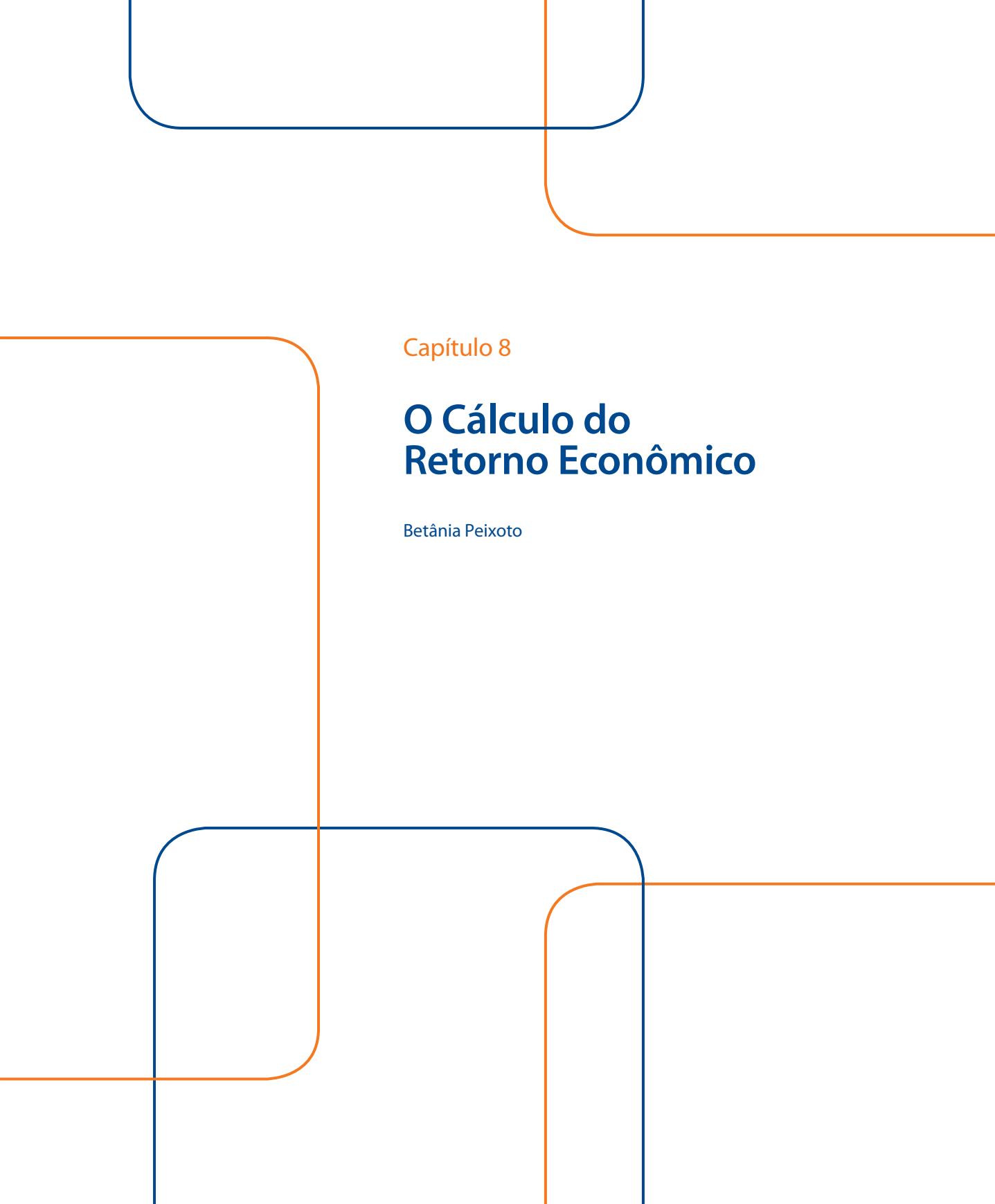
1 se Z é maior que c e 0, caso contrário. Supomos que todas as hipóteses no método de regressão descontínua são válidas.

- a.** Qual seria o parâmetro que captaria o efeito médio local do tratamento dentro do arcabouço de regressão descontínua?
 - b.** Suponha que ao analisar os dados, verificamos que o salto na probabilidade de receber tratamento em torno de Z igual a c é muito pequeno, sendo próximo de 0. De outra maneira, suponha que $\hat{\theta}$ é próximo de 0. O que aconteceria com o estimador do efeito médio local proposto na letra (a)?
 - c.** Você resolve analisar melhor os dados e percebe que existem inúmeras descontinuidades no gráfico que relaciona a probabilidade de receber tratamento com o instrumento Z . Como você interpreta essa evidência? Ela invalida o uso de regressão descontínua?
 - d.** Suponha que, ao analisar os dados, você percebe que existem inúmeros saltos no gráfico que relaciona o resultado de interesse Y a Z . Como você interpreta essa evidência? Ela invalida o uso do método de regressão descontínua?
- 4.** Devido a um aumento inesperado nos preços dos aluguéis, o governo cede à pressão popular e resolve criar uma política que facilite o acesso à casa própria ao trabalhadores brasileiros. Para isso, o governo estabelece reduções nos juros do crédito para casa própria. Essas reduções nos juros variam com a renda mensal do trabalho. Nos primeiros anos da política, os trabalhadores com renda até R\$ 700,00 têm redução de 2% ao ano, os com renda entre R\$ 701 e R\$ 1.500,00 têm redução de 1%, e os com salário entre R\$ 1.501,00 e R\$ 3.000,00 têm redução de 0,5% ao ano. Os demais trabalhadores não têm redução na taxa de juros do crédito para compra de imóveis.

Suponha que o governo tem uma base de dados com informações dos trabalhadores em anos que antecederam a política e depois da implementação dela. Nessa base de dados, temos informação sobre as características do trabalhador como idade, sexo, renda mensal do trabalho, região de domicílio e se o domicílio é próprio ou não. Com esses dados, vamos estimar o efeito dessa política de crédito sobre a probabilidade de adquirir um novo domicílio. Com essa base de dados, três pesquisadores propõem metodologias diferentes para estimar o efeito de tal política.

O primeiro pesquisador propõe o método de diferenças em diferenças, o segundo usa o método de variáveis instrumentais, e o último se baseia no método de regressão descontínua. Para cada um dos métodos:

- (i) Descreva qual o parâmetro de interesse;
- (ii) Discuta as hipóteses necessárias para identificar estes parâmetros e interprete estas hipóteses no contexto deste exemplo;
- (iii) Descreva o processo de estimação de cada um dos métodos;
- (iv) Compare os resultados que seriam obtidos com cada método.



Capítulo 8

O Cálculo do Retorno Econômico

Betânia Peixoto

A avaliação econômica de um projeto é composta de duas etapas: a avaliação de impacto e a avaliação do retorno econômico. Essas duas etapas são complementares, sendo que a avaliação de retorno econômico é realizada a partir dos resultados obtidos na avaliação de impacto.

Como vimos no capítulo introdutório, para valer a pena alocar mais recursos em um projeto não basta que o programa tenha impacto positivo. É necessário que os benefícios do programa sejam maiores que seus custos. Assim, a avaliação de retorno econômico de um projeto visa à mensuração da sua viabilidade econômica. Por exemplo, o retorno econômico de um investimento compara o seu custo com o benefício financeiro gerado, de forma a determinar se o investimento é economicamente viável. Da mesma forma, o cálculo do retorno econômico procura responder se o benefício gerado pelo projeto é maior do que seu custo.



Calcular o retorno econômico é importante porque, mesmo que um projeto tenha impacto positivo (estimado por meio da avaliação de impacto), há situações em que seu custo seja tão alto que inviabilize a sua replicação. Concluiríamos, nesse caso, que o projeto funciona, mas não compensa. Muitas vezes, existem projetos que exibem impacto de grande magnitude, mas quando comparados com seus custos, apresentam-se pouco viáveis, ou seja, têm custo muito alto em relação aos benefícios que geram.

Por analogia, imagine a decisão de comprar um carro. Se olharmos apenas o lado das benesses do carro – como a potência e o conforto –, poderíamos, por exemplo, escolher uma Ferrari. Entretanto, essa escolha não considera o custo para adquiri-lo. Como vivemos em um mundo com recursos escassos, esse elemento é fundamental nas tomadas de decisões. Um dos principais motivos de carros Ferrari não serem consumidos por todos é o custo. O mesmo deveria acontecer com o investimento nos programas sociais. Se analisarmos somente o impacto do programa – que equivale às benesses da Ferrari, muitas vezes estaremos investindo em programas com alto impacto, mas que apresentam custo tão elevado que não seriam viáveis economicamente.

A aplicação do conceito de viabilidade econômica em projetos sociais exige uma visão mais ampla do seu significado, pois não é trivial pensar em um projeto social da mesma forma que se pensa em um investimento econômico. Isso porque, o benefício advindo do projeto social não é auferido somente pelo financiador (a partir desse ponto, referido como investidor social). Grande parte dos benefícios é auferida pelas pessoas que foram impactadas pelo projeto, gerando benefícios privados aos participantes. Porém, parte dos benefícios do projeto incide sobre as pessoas que não participaram do mesmo, caracterizando-se como benefício público. O benefício público ocorre uma vez que o projeto, ao atingir seus objetivos e gerar benefícios para os participantes, também beneficia uma parte da sociedade, que ganha (indiretamente) com essas melhorias. Esse fenômeno, em que o projeto atinge indiretamente os que não participam do programa, é denominado de externalidade.

Vamos entender a aplicação dos conceitos mencionados anteriormente, no caso de avaliação de projetos sociais, por meio de um exemplo: o projeto “Olimpíada Brasileira de Matemática das Escolas Públicas – OBMEP”, avaliado por VASCONCELLOS et al. (2009). A OBMEP tem como um dos principais objetivos melhorar a qualidade da educação pública, especialmente no ensino de matemática. Para atingir esse fim, desempenha várias ações, ou seja, incorre em custos. Apesar dos custos, esse projeto não é um investimento financeiro no sentido clássico. Entretanto, para sua realização é necessária a aplicação de recursos, que pode ser interpretada como investimento social. O projeto não visa lucros para os financiadores, mas se ocorrer aumento da qualidade da educação, o ganho para o indivíduo e para a sociedade desse aumento pode ser interpretado como “retorno social” do projeto. O ganho para o participante (a criança) será melhor formação e maior escolaridade que levará por toda vida, seu benefício privado. Já as externalidades e os benefícios públicos advindos da maior escolarização dos participantes podem ser vários, como aumento da produtividade do trabalho, melhora na qualidade dos serviços públicos, diminuição da subnutrição infantil dos filhos desses tratados, entre outros.

Se os custos do projeto social são considerados como investimento, e os seus benefícios, como receita proveniente desse investimento, para calcularmos o retorno econômico é necessário dimensionarmos o valor monetário desses dois elementos.

8.1. Benefícios do projeto social em valores monetários

O benefício do projeto social ou política pública é o valor monetário do seu impacto. Portanto, necessariamente, a avaliação de impacto antecede a avaliação de retorno econômico e somente se realiza quando a avaliação de impacto evidencia que o projeto gerou resultados. O que fazemos é transformar o impacto em valor monetário, de forma a obter o benefício.

A transformação do impacto estimado em benefício é uma das maiores dificuldades da avaliação de retorno econômico. Ela exige a associação de um valor monetário a um impacto, muitas vezes, não monetário. A única exceção são os projetos para geração de renda, em que o indicador de impacto é a renda, ou seja, já é estimado em valor monetário, não exigindo nenhuma transformação. Para os demais projetos sociais, que não tenham como indicador de impacto a renda dos participantes, a transformação do impacto em benefício precisa ser realizada.

A pergunta que se faz é como efetuar essa transformação em projetos sociais com objetivos tão diversos, como a melhoria na qualidade da educação e da saúde; a redução da criminalidade, entre outros. A resposta não é simples. Para efetuarmos essas transformações, temos que fazer hipóteses que permitam atribuir valor monetário ao impacto. Essas hipóteses variam de projeto para projeto e exigem profundo conhecimento sobre a temática a qual se aplica. O avaliador precisa conhecer a literatura a respeito do indicador de impacto, saber quanto a mudança causada no indicador pelo programa economiza para a sociedade e quanto aumenta o bem-estar, devido à contribuição advinda das mudanças geradas nos beneficiários e nas circunstâncias que os envolvem.

Analisemos, por exemplo, um programa para o aumento da escolaridade, no qual o indicador de impacto é a taxa de jovens com ensino superior completo. A literatura mostra que escolaridade mais alta propicia salários maiores, menor probabilidade de acessar programas sociais, menor probabilidade de encarceramento, entre outros efeitos. Na transformação do impacto estimado em benefício (valor monetário do aumento da taxa de jovens com ensino superior completo), o avaliador deve considerar: os ganhos salariais devido ao aumento dos salários; a probabilidade daqueles jovens tratados acessarem programas sociais e a média dos custos em prover esses programas que seriam acessados caso o jovem não tivesse aumentado sua escolaridade – isso equivale a uma economia para a sociedade; a probabilidade de os jovens tratados serem presos, e o custo eventual dessas prisões se a escolaridade não tivesse sido aumentada – é o valor que

a sociedade deixou de gastar, também uma economia. Não seria possível estimar todas essas probabilidades e ganhos se não conhecermos bastante a literatura para podermos fazer hipóteses plausíveis sobre todos os parâmetros que serão empregados nessas estimativas.

No caso de projetos na área de saúde, por exemplo, para prevenir a incidência de diabetes, a transformação do impacto em benefício é também baseada em uma série de hipóteses. Digamos que o impacto do projeto seja mensurado com base na diminuição da taxa de incidência de diabetes na população. Menor taxa de incidência de diabetes leva a: diminuição dos gastos com o sistema de saúde para tratamento de doenças derivadas dela; melhora na vida das pessoas que adquiriram hábitos saudáveis por causa do projeto e que ficariam doentes caso ele não existisse; economia para sociedade com os dias de trabalho que seriam perdidos se as pessoas ficassem doentes e diminuição dos gastos com previdência social que teria de ser paga aos doentes e inválidos.

Se o projeto é para melhorar a empregabilidade dos jovens, por exemplo, o avaliador deve conhecer muito sobre o mercado de trabalho. Ele precisará determinar qual o valor para a sociedade do aumento do número de jovens empregados. O fato de mais jovens entrarem no mercado de trabalho gera ganhos para os próprios indivíduos (com relação a salário) e para toda a sociedade como: aumento da produção – porque agora há mais pessoas contribuindo para o produto; diminuição da probabilidade de o jovem que arrumou emprego se envolver com álcool, drogas e crime, o que, por sua vez, leva a menores gastos com tratamentos toxicômanos e prisões; melhora na qualidade do aprendizado dos filhos daqueles que foram empregados, pois terão melhor estrutura familiar; aumento do nível de atividade econômica à medida que boa parte dos salários é gasta no comércio, entre outros efeitos.

Dos exemplos apresentados, pode-se concluir que, para determinar o valor ganho ou economizado pela sociedade devido ao projeto, o avaliador terá de conhecer, profundamente, a literatura sobre o tema. Só assim, será capaz de estimar as várias facetas afetadas pelo projeto e adotar boas hipóteses para a transformação. Ao final desse capítulo, apresentamos três exemplos de avaliações de projetos no Brasil, nas áreas de educação, criminalidade e saúde.

Outra questão que o avaliador terá de considerar na estimativa do benefício do projeto é que o impacto, em geral, perdura mais do que um período de tempo. Algumas vezes o benefício perdura por toda a vida produtiva do indivíduo, como no caso dos projetos de qualificação profissional; ou

por séculos, como o caso de projetos de reflorestamento. Outras vezes, os benefícios são grandes logo após a implementação e reduzem-se ao longo do tempo, por exemplo, campanhas de prevenção a doenças epidemiológicas, como a dengue. Existem também projetos que só geram efeitos muitos anos após sua implementação, como aqueles para o aumento da escolaridade. O que todos esses casos têm em comum é que os benefícios ocorrem em momentos distintos no tempo.

Se os benefícios são recebidos em momentos diferentes, é preciso considerar o tempo no cálculo do benefício total. Não se pode, simplesmente, somar os benefícios em cada momento. Isso acontece porque, em geral, as pessoas atribuem mais valor ao dinheiro no presente do que no futuro. Afinal, o que você prefere: ganhar 500 reais hoje ou somente no ano que vem? A maioria das pessoas prefere hoje. Isso se chama “preferência intertemporal”.

Para calcularmos o benefício monetário total do projeto, precisamos considerar o efeito do tempo no seu recebimento. Para isso, montamos o fluxo de caixa do benefício e utilizamos a taxa de desconto intertemporal para deduzir do montante o fato de que o valor só será recebido no futuro. A taxa de desconto intertemporal indica o quanto o indivíduo valoriza o consumo presente em relação ao futuro, ou seja, é o percentual pelo qual ele está disposto a abrir mão do consumo no presente para só fazê-lo no futuro. Para efetuar o desconto intertemporal aplicamos, então, os conceitos de “valor presente” e “valor futuro” oriundos da matemática financeira. Esses conceitos são bastante intuitivos:

- i.** Valor presente (VP) é quanto determinado montante de dinheiro a ser recebido no futuro vale no tempo inicial do investimento;
- ii.** Valor futuro (VF) é quanto um montante de dinheiro hoje valerá em momento posterior.

Suponha que desejamos saber quanto vale hoje um montante de dinheiro a ser recebido em n períodos (meses ou anos). A fórmula que relaciona esses dois conceitos é:

$$VP = \frac{VF}{(1+i)^n} \quad (1)$$

Em que: n é o número de períodos

i é a taxa de desconto intertemporal

Suponha agora que desejamos saber o valor presente dos benefícios gerados ao longo do tempo por um projeto. Para obtermos o benefício total do projeto, aplicamos essa fórmula considerando o desconto intertemporal na soma dos benefícios ao longo do tempo:

$$VPTB_{t_0} = VB_{t_0} + \sum \frac{VFB_{t_n}}{(1+i)^n} \quad (2)$$

Onde o $VPTB_{t_0}$ é o valor presente do benefício total; VB_{t_0} é o valor do benefício no tempo inicial¹, se ocorrer do benefício ser imediato à implementação; e VFB_{tn} é o valor do benefício recebido em cada um dos n períodos t .

A taxa de desconto intertemporal adequada é uma decisão discricionária. Em geral, os economistas consideram a taxa de desconto intertemporal igual à taxa de juros. A intuição para isso é que a taxa de juros seria a recompensa financeira por deixar de consumir hoje, para consumir amanhã. Ela pode ser interpretada como a taxa que os indivíduos valorizam o presente em relação ao futuro. Quanto mais elevada é a taxa de desconto intertemporal, mais as pessoas valorizam o consumo presente. Por outro lado, quanto menor essa taxa, menos as pessoas valorizam o consumo presente em relação ao futuro.

A taxa de juros considerada como representativa da taxa de desconto intertemporal é a taxa de juros real. A taxa de juros real é igual à taxa de juros nominal menos a inflação/deflação. A taxa de juros nominal é a de face, ou seja, aquela que é paga quando fazemos um investimento. Por exemplo, se em determinado ano uma aplicação na poupança rendeu 7,5% (taxa de juros nominal) e a inflação acumulada foi de 6,5%, então a taxa de juros real foi de 1%. Utilizamos a taxa de juros real porque o cômputo do montante do benefício já está isento da inflação/deflação, pois quando o calculamos usamos as estimativas com os valores monetários referentes ao ano-base (via de regra, em t_0). Portanto, se usássemos a taxa de juros nominal, estaríamos descontando a variação do valor da moeda duas vezes.

Não necessariamente precisamos igualar a taxa de desconto intertemporal à taxa de juros. Muitas vezes, os avaliadores adotam taxas de desconto intertemporal menores que a taxa de juros corrente.

1. Desafio: por que não aplicamos a fórmula do valor presente no valor do benefício em t_0 ?

Resposta: porque o valor do benefício em t_0 já está, necessariamente, em valor presente desse período.

Isso acontece porque, como a maioria dos projetos tem impactos de longo prazo, os avaliadores utilizam uma taxa de juros de longo prazo. A taxa de juros de longo prazo, em geral, é menor que a taxa de juros corrente. Alternativamente, muitos estudos utilizam a taxa de juros dos Estados Unidos, por se tratar de uma economia mais estável que a brasileira. Como a economia brasileira passou nas últimas décadas por profundas mudanças, a sua taxa de juros de longo prazo é muito volátil e pode não ser uma boa estimativa para o futuro.

Para entender melhor os conceitos trabalhados, está disponível como último tópico do capítulo um exemplo fictício de avaliação de retorno econômico de um projeto para qualificação profissional de jovens.

8.2. Custo do projeto

Uma vez calculado o benefício total, a outra dimensão necessária para a avaliação de retorno econômico é o custo do projeto. O custo considerado nesse tipo de análise é denominado “custo econômico” e transcende o custo contábil por considerar o custo de oportunidade do projeto. Vamos entender cada um desses conceitos:

$$\text{Custo econômico} = \text{Custo Contábil} + \text{Custo de Oportunidade}$$

O custo contábil é todo dispêndio que, via de regra, aparece nos demonstrativos contábeis. É todo desembolso monetário necessário para o funcionamento do projeto, por exemplo, pagamento conta de luz, de pessoal, aquisição de material, etc. O custo de oportunidade difere do custo contábil por ser um custo implícito, ou seja, em geral não aparece nos demonstrativos contábeis.

O conceito do custo de oportunidade foi cunhado pela economia para designar o rendimento que se deixa de obter ao realizar determinada escolha. Em outras palavras, é o ganho que poderia advir da aplicação alternativa de um determinado recurso. Muitos projetos funcionam com um exíguo orçamento próprio, mas utilizam muitos recursos fornecidos pela sociedade. Por exemplo, um programa de fornecimento de sopa para pessoas carentes que utiliza mão de obra voluntária, que recebem alimentação de um doador, que utiliza a cozinha cedida de uma instituição benéfica, entre outros recursos. O custo contábil provavelmente será muito baixo, somente se

ele for computado no cálculo do retorno, ele estaria distorcido. O custo de oportunidade resolve o problema da distorção do retorno, uma vez que computa o quanto valeria: a aplicação da mão de obra em atividade produtiva; os alimentos doados se fossem aplicados de outra maneira, por exemplo, vendidos; o aluguel da cozinha para um restaurante. Nesse caso específico, o custo econômico do projeto tem com maior parcela o custo de oportunidade. Ao utilizá-lo para o cálculo do retorno econômico, esse não mais estaria distorcido.

Um exemplo clássico de custo de oportunidade no terceiro setor é o aluguel não pago. Muitos projetos sociais funcionam em imóveis cedidos por parceiros, pelos quais o aluguel não é pago. Como não pagam pelo recurso utilizado, ele não consta no custo contábil. Porém, para o projeto funcionar, ele precisa usufruir o imóvel. Nesse sentido, o projeto usa o recurso, mesmo não pagando por ele, e portanto esse custo precisa ser computado no cálculo do retorno. O custo de oportunidade faz esse cômputo. Se o imóvel fosse usado para outro fim, por exemplo, aluguel para ponto comercial, estaria gerando renda e remunerando o capital investido na sua aquisição. Portanto, o custo de oportunidade do uso do imóvel pelo projeto é o aluguel que ele deixou de render.

Para melhor compreensão vejamos outro exemplo. Considere o projeto de curso profissionalizante ministrado no contraturno escolar. O jovem que escolheu participar do curso deixou de fazer um estágio pelo qual receberia um salário. Mesmo que seja desejado pelos gestores do projeto, o fato de o jovem assistir às aulas em vez de ir ao estágio remunerado é um custo de oportunidade de participação no projeto e precisa ser computado na avaliação de retorno econômico. Ainda que esse jovem não fizesse o estágio, o tempo que ele permanece no curso poderia ser usado de forma alternativa para ajudar os pais nos afazeres domésticos ou para cuidar dos irmãos mais novos. Portanto, o custo dos afazeres domésticos e do cuidado com os irmãos é o custo de oportunidade de participar do projeto e precisa ser valorado.

Outro custo de oportunidade, muito comum em projetos sociais, é o dispêndio de horas de trabalho dos voluntários. Quando o projeto social funciona com a colaboração de voluntários, o custo de horas trabalhadas por eles não aparece no custo contábil, simplesmente porque não foram pagas. Entretanto, para o seu funcionamento, esse recurso (as horas de trabalho dos voluntários) foi utilizado, e o cômputo do seu custo deve considerá-lo. O uso alternativo desse recurso seria despender as horas em serviço remunerado. Assim, o custo de oportunidade do trabalho voluntário é o quanto esse profissionais estariam recebendo caso fossem remunerados.

Existem muitos outros exemplos de custo de oportunidade, específicos para cada projeto social ou política pública, e para descobri-los basta aplicar o conceito na análise de cada caso. Ao final do capítulo, apresentamos três exemplos de custos econômicos considerados em avaliações no Brasil.

Calculado o custo de oportunidade e o custo contábil, a soma de ambos resulta no custo econômico. Assim como acontece com o benefício, o custo econômico pode ocorrer em diferentes momentos no tempo. Novamente, vale relembrar que não podemos somar o valor gasto em cada momento para obtermos o custo total sem considerar a taxa de desconto intertemporal. Para isso, aplicamos os conceitos de valor presente e valor futuro da matemática financeira e obtemos a seguinte relação:

$$VPTC_{t_0} = VC_{t_0} + \sum \frac{VFC_{t_n}}{(1+i)^n} \quad (3)$$

Em que o **VPTC** é o valor presente do custo total; **VC_{t₀}** é o valor do custo no tempo inicial; e **VFC_{t_n}** é o valor do custo incorrido nos **n** períodos **t**.

Essa fórmula é análoga àquela aplicada ao cálculo do benefício total do programa. Assim, toda a discussão em relação à taxa de desconto intertemporal realizada na seção 8.1 continua válida. Além disso, não podemos esquecer que a taxa de desconto intertemporal tem que ser a mesma considerada no cálculo de valor presente do benefício total, ou seja, ao escolhermos uma taxa de desconto intertemporal, ela será a mesma para todos os cálculos do retorno econômico.

8.3. Medidas do retorno econômico

Após a estimação do custo e do benefício de um projeto social ou política pública, estamos aptos a calcular o seu retorno econômico. O retorno econômico visa à identificação da viabilidade econômica do projeto, por meio da comparação do seu custo com o seu benefício. A viabilidade econômica se refere ao fato de o projeto gerar “lucro social”, ou seja, seu benefício ser maior que seu custo para a sociedade.

O cálculo do retorno permite descobrir a viabilidade econômica de um projeto social ou política pública e, também, comparar alternativas de investimentos em projetos diferentes. A comparação

de dois projetos indica que aquele com maior retorno é o mais atrativo como investimento. Entretanto, é preciso ter cuidado com essa comparação.

A comparação de projetos só deve ser realizada se eles forem semelhantes em seus objetivos e público-alvo. Cada área de investimento social apresenta retornos diferenciados para a sociedade, mas todas precisam de investimento. A decisão de investimento não passa somente pelo retorno de cada área, mas também pela vocação de cada instituição financiadora. Por exemplo, projetos de educação básica costumam gerar retornos elevados para a sociedade pelas características inerentes ao seu público-alvo – o impacto perdura por toda a vida do participante. Projetos para idosos, em geral, apresentam retornos mais baixos, também por características inerentes ao público-alvo – o impacto perdura poucos anos. Não seria correto usar a comparação do retorno econômico de ambos para decidir em qual investir, pois essas duas áreas sociais têm características muito diferentes. O ideal seria comparar somente o retorno de projetos de educação básica com objetivos semelhantes e comparar somente o retorno de projetos para idosos e, dentre cada grupo, decidir em qual investir. Além disso, o número de atendidos e a capacidade de expansão do projeto também devem ser levados em consideração.

Existem várias formas de calcularmos o retorno, e cada uma delas constitui uma estatística de retorno econômico. A diferença entre as estatísticas é apenas a ótica da sua construção, pois todas medem a relação entre o custo e o benefício do projeto. Nesse livro, apresentamos quatro estatísticas de retorno econômico: Valor Presente Líquido (*VPL*), Taxa Interna de Retorno (*TIR*), Razão Custo-Benefício e Razão Custo-Efetividade.

O *VPL* e a *TIR* decorrem da aplicação direta de conceitos da matemática financeira. A Razão Custo-Benefício é um conceito aplicado em diversas áreas da economia. Essas estatísticas são muito comuns na análise de viabilidade econômica de investimentos do setor privado. Por fim, a Razão Custo-Efetividade é calculada quando não for possível transformar o impacto encontrado em benefício, ou seja, quando não for possível atribuir valor monetário ao impacto. Veremos detalhadamente cada uma dessas estatísticas.

8.3.1. Valor Presente Líquido –VPL

O valor presente líquido estima o retorno por meio da subtração do valor presente do custo total pelo valor presente do benefício total. A intuição é que se os benefícios são maiores que os custos, então o projeto é viável economicamente. Ou seja, o projeto gera para sociedade mais valor do que retira dela em termos do investimento necessário para sua execução. Assim, temos:

$$VPL = Benefício - Custo \quad (4)$$

Considerando que os benefícios e os custos acontecem em momentos diferentes no tempo, para realizarmos essa subtração, precisamos que ambos estejam a valor presente de um mesmo período. Considerando t_0 o primeiro ano do investimento, trazemos os valores do benefício e do custo de cada ano a valor presente por meio da fórmula:

$$VPL_{t_0} = \left(VB_{t_0} + \sum \frac{VFB_{t_n}}{(1+i)^n} \right) - \left(VC_{t_0} + \sum \frac{VFC_{t_n}}{(1+i)^n} \right) \quad (5)$$

A análise do resultado é:

Se $VPL > 0$ temos um retorno positivo, o projeto é viável economicamente, pois seu benefício supera os custos.

Se $VPL < 0$ temos um retorno negativo, o projeto não é viável economicamente, pois seu custo supera o benefício gerado.

Se $VPL = 0$ temos um retorno neutro, valor do benefício gerado é igual ao custo incorrido.

8.3.2. Taxa Interna de Retorno – TIR

A taxa interna de retorno é a taxa de oportunidade de investir o recurso no projeto. Ela representa o retorno implícito do investimento. Matematicamente, a **TIR** é a taxa de desconto intertemporal implícita que iguala o benefício do projeto ao seu custo, de forma que o valor presente líquido seja 0. Assim:

$$VPL_{t_0} = \left(VB_{t_0} + \sum \frac{VFB_{t_n}}{(1+TIR)^n} \right) - \left(VC_{t_0} + \sum \frac{VFC_{t_n}}{(1+TIR)^n} \right) = 0 \quad (6)$$

Ou, reagrupando os termos de (6), temos:

$$VPL_{t_0} = \left(VB_{t_0} - VC_{t_0} \right) + \left(\frac{VFB_{t_1}}{(1+TIR)^1} + \dots + \frac{VFB_{t_n}}{(1+TIR)^n} \right) - \left(\frac{VFC_{t_1}}{(1+TIR)^1} + \dots + \frac{VFC_{t_n}}{(1+TIR)^n} \right) = 0 \quad (6a)$$

Calculadoras financeiras ou softwares fazem essa estimativa.

A **TIR** é uma taxa anual, ou na unidade de tempo utilizada para construir o fluxo de caixa. Para interpretarmos o resultado da **TIR**, é necessária a comparação com a **TIR** de outro projeto ou com uma taxa mínima de atratividade (**TMA**). A taxa mínima de atratividade é a taxa de juros mínima que um investidor pretende obter com o seu investimento. Assim a **TMA** pode ser a taxa de juros básica do mercado, a taxa de remuneração da poupança ou outra taxa que se considerar a mínima para remunerar um investimento. Assim, são três os resultados possíveis:

Se a **TIR > TMA**, o projeto é viável economicamente, pois o retorno gerado é maior que a taxa mínima de atratividade, o que gera retorno para a sociedade. Ou seja, o projeto tem um retorno implícito maior do que o mínimo considerado pela sociedade como razoável.

Se a **TIR < TMA**, o projeto não é viável economicamente, porque seu retorno é menor que a mínima de atratividade, não gerando retorno para a sociedade. Ou seja, o projeto tem um retorno implícito menor do que o mínimo desejado.

Se a **TIR = TMA**, o projeto está em uma zona de indiferença de retorno, mas é economicamente viável. Ou seja, o projeto gera o mínimo desejado pela a sociedade.

8.3.3. Razão Custo-Benefício

A Razão Custo-Benefício é uma das estatísticas de retorno econômico mais utilizadas, devido à sua fácil compreensão. No dia a dia, costumamos fazer análises de custo-benefício em todas as nossas

decisões. Em geral, quando estamos diante de uma escolha, pensamos nos prós e nos contras que podem advir da decisão, para somente depois, balizarmos nossas ações. Por exemplo, com relação à compra de um automóvel: primeiramente analisamos a potência, o custo de manutenção, o conforto, o *design* e vários outros aspectos. Depois, verificamos o custo do automóvel. Por fim, relacionamos os benefícios de cada modelo e comparamos com o custo. Somente depois de todo esse processo é que decidimos pela compra. Quando se realizam essas comparações, o que estamos fazendo de forma intuitiva é uma análise custo-benefício.

No caso de projetos privados e sociais a análise não pode ser intuitiva. O que se faz é relacionar o custo e o benefício do projeto por meio da Razão Custo-Benefício. Essa razão fornece uma estatística do retorno de cada real investido no projeto para a sociedade. Para o seu cálculo basta dividir o benefício total pelo custo econômico total². Tanto o benefício quanto o custo devem estar a valor presente (como definidos nas seções anteriores):

$$\text{Razão Custo-Benefício} = VPTB_{t_0} / VPTC_{t_0} \quad (7)$$

A interpretação do resultado nos fornece o montante de benefício que será gerado por cada real investido. Por exemplo, se a Razão Custo-Benefício foi de 1,5, significa que a cada real investido o benefício gerado será de um real e cinquenta centavos. Portanto, o projeto é viável economicamente. Por outro lado, se a Razão Custo-Benefício for de 0,60, significa que a cada real investido no projeto, valerá no futuro apenas sessenta centavos. Ou seja, o dinheiro investido perdeu valor com o projeto, pois não gerou benefícios suficientes para compensar os custos.

Se quisermos saber o retorno para a sociedade em termos percentuais, aplicamos a seguinte fórmula:

$$\text{Retorno} = (\text{Razão Custo-Benefício} - 1) * 100 \quad (8)$$

Esse resultado nos fornece quanto o investimento no projeto gera no futuro em termos percentuais. Se a Razão Custo-Benefício é de 1,5, então o retorno será de 50%. Portanto, o projeto apresenta

2. Apesar da razão ser a divisão do benefício pelo custo – razão benefício-custo, adotamos a nomenclatura razão custo-benefício por ser a denominação difundida nacionalmente. Em inglês o termos é benefit-cost ratio.

retorno positivo para a sociedade, sendo viável economicamente. Com a Razão Custo-Benefício de 0,6, o retorno será de -40%, ou seja, retorno negativo, indicando que o programa não é viável economicamente. O retorno percentual é, somente, uma outra maneira de apresentar o mesmo resultado da razão benefício-custo. A decisão de como apresentar os resultados é escolha discricionária do avaliador e dependerá do seu conhecimento prévio do público para o qual se dirige, de modo a determinar qual a forma de mais fácil compreensão.

8.3.4. Razão Custo-Efetividade

A Razão Custo-Efetividade é uma alternativa para a comparação dos custos com os benefícios dos projetos, quando não se deseja, ou quando não é possível, valorar monetariamente o impacto. Essa estatística dimensiona o impacto em relação aos custos, sem transformá-lo em benefício. Assim, utilizamos a Razão Custo-Efetividade quando não é possível encontrar boas hipóteses para transformar o impacto estimado em valor monetário ou caso não seja interessante essa transformação.

Alguns projetos têm indicadores de impacto de difícil valoração. Por exemplo, projetos para a preservação das baleias. O impacto será medido em termos de número de baleias salvas. Para transformar o impacto em benefício, teríamos que saber quanto vale para a sociedade cada baleia salva. Isso exigiria hipóteses nada triviais. Outras vezes, por questões ideológicas, os avaliadores não querem atribuir valor a alguns indicadores de impacto. No caso de projetos para prevenção de doenças, em que o impacto é o número de vidas salvas. Para transformar esse impacto em benefício é preciso valorar a vida. Muitos autores consideram que o valor da vida é o quanto a pessoa vai produzir no mercado de trabalho, ou seja, quanto de salário receberá ao longo da sua vida laboral. Isso pode parecer lógico para alguns. Entretanto, para outros, pode parecer um absurdo, pois o valor da vida transcende o seu sentido econômico. É muito mais, em termos de contribuição à sociedade, educação dos filhos, etc. Em fim, é o papel social do homem. Nos dois exemplos, o avaliador poderia optar por apresentar, como estatística de retorno econômico, a Razão Custo-Efetividade.

A Razão Custo-Efetividade é calculada pela divisão do impacto estimado (na avaliação de impacto) pelo custo econômico total:

Razão Custo-Efetividade= Impacto Estimado / Custo Econômico_{tn} (9)

Razão Custo-Efetividade= Impacto Estimado / VPTC_{tn} (9a)

Essa estatística fornece o quanto de unidade de impacto se obtém por real gasto. Por exemplo, no caso do projeto para preservação das baleias, a Razão Custo-Efetividade fornece quantidade de vidas salvas por real investido. Se a Razão Custo-Efetividade é de 0,000003, significa que a cada real investido salva 0,000003 baleias. A fim de tornar o resultado mais intuitivo podemos multiplicar o resultado por 1.000.000. Assim, teríamos que para cada 1.000.000 de reais investido no programa três baleias são salvadas. No exemplo do projeto de prevenção às doenças, a Razão Custo-Efetividade indica quantas vidas foram salvadas para cada real despendido com o programa. Se a Razão Custo-Efetividade é de 0,00045 e multiplicarmos esse resultado por 100.000, teremos que cada cem mil reais investidos no projeto salvam 45 vidas.

Após calcular a Razão Custo-Efetividade, a pergunta é se o valor encontrado seria alto ou baixo. A resposta só poderá ser fornecida pelas outras estatísticas de retorno econômico. Isso indica que essa estatística, apesar de considerada uma estatística de mensuração de retorno econômico, não permite avaliar a viabilidade econômica do projeto. Por esse motivo, deve sempre que possível ser substituída pelas outras estatísticas apresentadas nesse capítulo. Mesmo com essa limitação, a estatística de custo-efetividade é muito usada para comparar projetos. Lembre-se de que a comparação deve se restringir a projetos com o mesmo indicador de impacto, o mesmo público-alvo e a mesma finalidade. Se o gestor precisar decidir entre dois projetos, ele escolherá o que tem menor custo-efetividade, pois o custo para se atingir o mesmo objetivo é menor. No exemplo da baleia, se outra metodologia de preservação das baleias fosse aplicada com Razão Custo-Efetividade de 0,000003, seria preferível àquela em que a Razão Custo-Efetividade é de 0,000001, pois com o mesmo montante é possível preservar mais baleias.

Alguém poderia argumentar que se o objetivo é comparar dois projetos, poderíamos analisar somente o impacto estimado. Entretanto, a comparação do impacto estimado poderia levar a conclusões errôneas, pois pode ocorrer a situação em que um projeto tenha o impacto maior do que o outro, mas por causa do seu custo muito mais elevado, tem menor Razão Custo-Efetividade.

8.4. Análise de sensibilidade

A análise de sensibilidade é a última etapa da avaliação econômica. Seu objetivo é testar a sensibilidade (variação) dos resultados da avaliação em relação aos parâmetros utilizados. Como a avaliação econômica é baseada em métodos estatísticos, que estão sujeitos a erros, e em parâmetros escolhidos de forma discricionária, faz necessário testar se os resultados encontrados sofreriam alteração, caso a amostra e os parâmetros da avaliação fossem diferentes. Assim, a análise de sensibilidade é uma técnica de verificação de robustez dos resultados da avaliação. O termo robustez em avaliação é empregado, no sentido figurado, como a “força” da avaliação, ou seja, sua confiabilidade.

Toda avaliação econômica envolve um conjunto de informações incertas. A incerteza se deve às variações da amostra e ao espectro de plausíveis valores para os parâmetros. Por exemplo, podemos calcular o retorno econômico considerando uma taxa de desconto intertemporal de 5%. Entretanto, seria plausível utilizar uma taxa de desconto de 6% ou 4%, entre outros valores. Nesse caso, na análise de sensibilidade, refazemos a avaliação de retorno econômico com essas possíveis variações da taxa de desconto intertemporal, para verificar se os resultados não sofrem alterações.

Então, na prática, a análise de sensibilidade implica em refazer a avaliação econômica alterando os parâmetros utilizados e verificar se as estatísticas de retorno econômico continuam com o mesmo resultado. Se o resultado é mantido dizemos que ele é robusto.

Existem várias técnicas para se realizar a análise de sensibilidade. Nesse capítulo abordaremos uma delas, conhecida como análise de sensibilidade univariada (*univariate sensitivity approach*). A análise é fácil de ser realizada e permite considerar os dois tipos de incertezas da avaliação econômica – a causada pelas variações da amostra e pelo espectro de plausíveis valores dos parâmetros.

Como vimos ao longo desse livro, a avaliação econômica de um projeto social envolve a estimativa de vários parâmetros para os quais utilizamos amostras e técnicas estatísticas. Isso acontece, por exemplo, na mensuração do impacto do projeto, quando os métodos de regressão são empregados. Ao utilizarmos uma amostra para estimar determinado parâmetro na população, incorremos em erros de estimativa devido a características da amostra analisada. Provavelmente, a seleção de uma amostra diferente levaria a valor diverso do parâmetro. A construção do intervalo

de confiança procura captar os possíveis valores que o parâmetro poderia assumir se amostras diferentes fossem utilizadas, com uma probabilidade de certeza.

Para considerar a incerteza proveniente de variações na amostra, em geral, se considera o menor e o maior valor possível do parâmetro, ou seja, os valores do limite inferior e superior do intervalo de confiança a 95% ou 90% de confiança. Para cada um desses valores, recalcula-se as estatísticas de retorno econômico e verifica-se a manutenção dos resultados. Se os resultados se mantiverem, então dizemos que o retorno é robusto. Esse procedimento é repetido para cada um dos parâmetros estimados com base em amostras.

O procedimento para o teste de sensibilidade em relação às incertezas proveniente dos valores dos parâmetros é similar ao proveniente da variação na amostra. A diferença é que, nesse caso, não temos os valores dos limites superiores e inferiores com uma probabilidade de certeza, porque os parâmetros foram escolhidos discricionariamente, com base em conhecimento prévio do avaliador. Assim, a determinação de quais valores esse parâmetro poderia assumir também será escolha discricionária e difícil. Uma forma muito utilizada pelos avaliadores para determinar os menores e maiores valores é a análise dos valores passados do parâmetro.

A taxa de desconto intertemporal é um exemplo de parâmetro escolhido discricionariamente pelo avaliador. Alguns economistas consideram essa taxa igual à taxa de juros real da economia. Entretanto, existem várias taxas de juros, por exemplo, a taxa básica chamada de SELIC, a taxa preferencial de juros, a taxa de remuneração da poupança, etc. Além disso, também existem vários índices de inflação deflação, como Índice de Preços ao Consumidor Amplo (IPCA) e o Índice Geral de Preços (IGP). Assim, surge uma incerteza quanto à precisão do retorno econômico calculado, devido à discricionariedade da escolha desse parâmetro.

O raciocínio anterior, para a taxa de desconto intertemporal, aplica-se a qualquer outro parâmetro considerado na avaliação econômica. Pode-se, por exemplo, utilizar parâmetros (incertos) para a estimativa dos custos de oportunidade ou até mesmo dos custos contábeis. Desse modo, faz-se a análise de sensibilidade para cada um dos parâmetros discricionariamente adotados. Portanto, existem vários parâmetros utilizados na avaliação econômica que podem e devem ser testados na análise de sensibilidade, a fim de que se possa confirmar a robustez do retorno.

Além de testar o efeito no retorno econômico de mudanças em cada parâmetro, mantendo os demais constantes, os avaliadores, também, testam a variação conjunta de dois ou mais parâmetros para obterem um espectro ainda maior de testes e verificar, de forma mais precisa, a robustez dos resultados. Cada novo cálculo do retorno, supondo valores diferentes dos parâmetros, é chamado comumente de cenários.

Considere x o número de parâmetros a ser analisado. Cada parâmetro terá o seu limite inferior e superior e seu valor original. Assim, teremos a combinação de três elementos x vezes, menos a combinação original da avaliação. O total de cenários é dado por:

$$\text{Total de cenários} = 3^x - 1 \quad (10)$$

Após montarmos todos os cenários, calculamos a proporção daqueles que tiveram resultados contrários ao da avaliação. Quanto menor o percentual de cenários com resultados contrários, mais robusto o seu resultado será. Quanto maior o percentual de cenários com resultados contrários ao da avaliação, mais frágeis os resultados estimados.

A fim de evitar que o relatório da avaliação fique enfadonho, os avaliadores não costumam reportar o resultado de cada cenário. Eles apenas descrevem os parâmetros testados, o número total de cenários e o percentual de estimativas contrário ao da avaliação.

8.5. Exemplo fictício

Neste ponto, finalizamos a teoria relativa à avaliação econômica de projetos sociais e, em especial, a avaliação de retorno econômico, tema específico deste capítulo. Para fixar melhor o conteúdo e servir de roteiro para futuras avaliações de retorno econômico, apresentamos a seguir a avaliação econômica de um projeto fictício, realçando todas as etapas e cálculos necessários à estimação do retorno do projeto.

Nome do projeto: Trabalha Brasil

Objetivo: inserir jovens no mercado de trabalho.

Público-alvo: jovens de 17 anos nos dois últimos anos do Ensino Médio.

Ações: curso profissionalizante com duração de dois anos, oferecido no contraturno escolar.

Número de beneficiários: 150 indivíduos.

Indicador de impacto: empregabilidade dos jovens.

Resultado da estimativa de impacto um ano após o fim do projeto: em média, 10% dos tratados conseguiram emprego por causa do projeto.

8.5.1. Cálculo do benefício

Transformação do impacto em benefício:

Dos 150 participantes, 10% conseguiram emprego por causa do projeto. Então, o projeto conseguiu inserir 15 jovens no mercado de trabalho, que não o acessariam se não tivessem participado do curso profissionalizante.

Para transformar o impacto em benefício, considere as seguintes hipóteses:

- i. O benefício de colocar o jovem no mercado de trabalho é o salário que ele recebe, equivalente ao que ele produz para a sociedade;
- ii. Na média, os beneficiários empregados recebem R\$ 1.000,00. Esse valor foi calculado com base nas informações de salário indicadas pelos jovens.

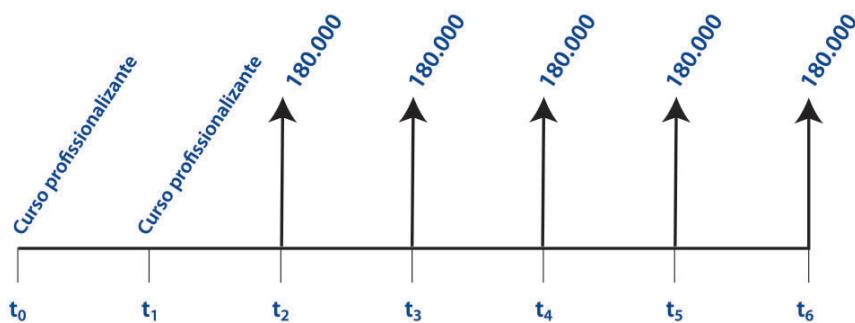
Para calcular o benefício do projeto por um ano, multiplicamos o número de jovens que estão empregados por causa do projeto (15) pelo salário médio que recebem por mês (R\$ 1.000,00) durante um ano – perfazendo um total de R\$ 180.000,00 por ano.

Fluxo de caixa e cálculo do benefício total:

Para elaborar o fluxo de caixa, considere que:

- iii.** Se não fosse o projeto, os jovens beneficiários entrariam no mercado de trabalho, de qualquer forma, cinco anos após o fim do projeto. Então, o benefício perdura durante cinco anos.
- iv.** A taxa de desconto intertemporal é de 5%.

A figura a seguir ilustra o fluxo de caixa.



Para somar os valores em momentos diferentes no tempo, aplicamos a taxa de desconto intertemporal e calculamos o Valor Presente Total do Benefício em t_0 :

$$VPTB_{t_0} = VB_{t_0} + \sum \frac{VFB_{t_n}}{(1+i)^n}$$

$$VPTB_{t_0} = VB_{t_0} + \frac{VFB_{t_1}}{(1+i)^1} + \frac{VFB_{t_2}}{(1+i)^2} + \frac{VFB_{t_3}}{(1+i)^3} + \frac{VFB_{t_4}}{(1+i)^4} + \frac{VFB_{t_5}}{(1+i)^5} + \frac{VFB_{t_6}}{(1+i)^6}$$

$$VPTB_{t_0} = 0 + 0 + \frac{180.000}{(1+0,05)^2} + \frac{180.000}{(1+0,05)^3} + \frac{180.000}{(1+0,05)^4} + \frac{180.000}{(1+0,05)^5} + \frac{180.000}{(1+0,05)^6} = 742.196,00$$

Portanto, o benefício total do projeto a valor presente em t_0 é de R\$ 742.196,00. O cálculo do impacto, consequentemente, do benefício só considerou o benefício individual. As externalidades advindas do salário dos jovens inseridos no mercado de trabalho por causa do projeto não foram computadas. Dessa forma, sabemos de antemão que esse benefício está subestimado. O benefício real deve ser maior do que o estimado pela avaliação, fazendo com que o seu retorno também seja maior.

8.5.2. Cálculo do custo econômico:

Para o funcionamento dos cursos, a gestão do projeto despende de recursos com salário dos professores, material de laboratório, apostilas e lanches fornecidos para os alunos (tabela 1). Além dos valores efetivamente pagos, o projeto utiliza o laboratório fornecido por uma empresa parceira que não cobra aluguel, luz ou água. Esse valor é computado no custo de oportunidade. Os alunos, ao decidirem alocar seu tempo nas aulas em vez de investi-lo em outras atividades, também incorrem no custo de oportunidade (tabela 2). Os valores das tabelas se referem aos custos de um ano de funcionamento.

Tabela 1: Custo Contábil

Custo Contábil	Valor (R\$)
Salário dos professores e coordenador	60.000,00
Material de laboratório	100.000,00
Apostilas	5.000,00
Lanches	72.000,00
Total	137.000,00

Tabela 2: Custo de Oportunidade

Custo de Oportunidade	Valor (R\$)
Aluguel do laboratório	10.000,00
Conta de luz	300,00
Conta de água	150,00
Tempo dos alunos	210.000,00
Total	220.450,00

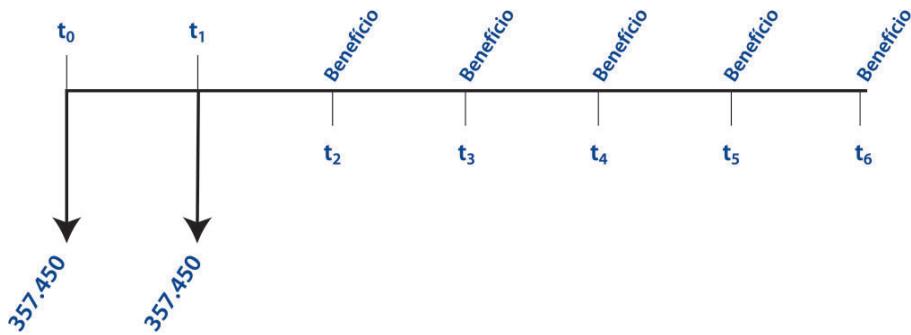
$$\text{Custo Econômico} = \text{Custo contábil} + \text{Custo de Oportunidade}$$

$$\text{Custo Econômico} = 137.000,00 + 220.450,00 = 357.450,00$$

Assim, o custo do projeto é de R\$ 357.450,00 por ano. Os jovens permanecem no projeto por dois anos.

Fluxo de caixa e cálculo do custo econômico total:

A figura a seguir ilustra o fluxo de caixa dos custos do projeto:



Para somar o custo do projeto nos diferentes momentos do tempo, aplicamos a taxa de desconto intertemporal de 5%, exatamente igual ao cálculo do benefício total. Note que estamos trazendo o custo do projeto a valor presente em t_0 . Portanto, o custo desse período já está em valor presente.

$$VPTC_{t_0} = VC_{t_0} + \sum \frac{VFC_{t_n}}{(1+i)^n}$$

$$VPTC_{t_0} = VC_{t_0} + \frac{VFC_{t_1}}{(1+i)^1}$$

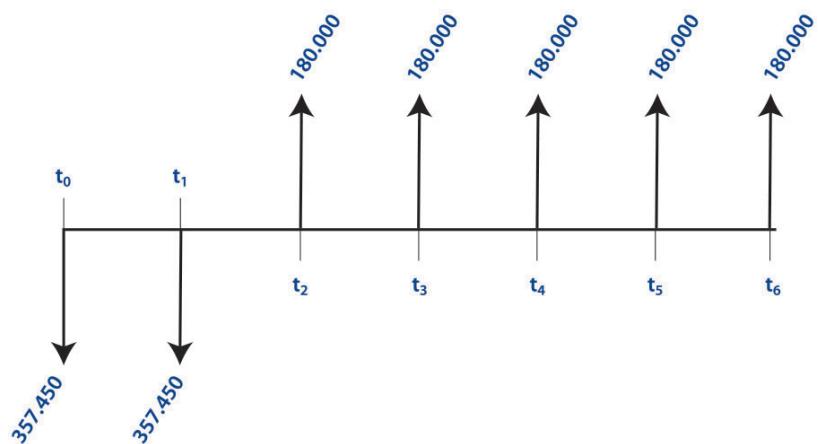
$$VPTC_{t_0} = 357.450 + \frac{357.450}{(1+0,05)^1} = 697.878,57$$

Assim, o custo econômico total do projeto em t_0 é de R\$ 697.878,57.

8.5.3. Retorno econômico

Cálculo do Valor Presente Líquido

O fluxo de caixa completo do projeto é:



Considere a taxa de desconto intertemporal de 5%. O Valor Presente Líquido é de:

$$\begin{aligned}
 VPL_{t_0} &= \left(VB_{t_0} + \sum \frac{VFB_{t_n}}{(1+i)^n} \right) - \left(VC_{t_0} + \sum \frac{VFC_{t_n}}{(1+i)^n} \right) \\
 VPL_{t_0} &= \left(\frac{180.000}{(1+0,05)^2} + \frac{180.000}{(1+0,05)^3} + \frac{180.000}{(1+0,05)^4} + \frac{180.000}{(1+0,05)^5} + \frac{180.000}{(1+0,05)^6} \right) - \left(357.450 + \frac{357.450}{(1+0,05)^1} \right) \\
 VPL_{t_0} &= 44.317,43
 \end{aligned}$$

Como já conhecíamos o valor presente do benefício e do custo, o VPL poderia ser calculado de forma alternativa:

VPTB_{t0}: 742.196,00

VPTC_{t0}: 697.878,57

$$VPL_{t0} = VPTB_{t0} - VPTC_{t0}$$

$$VPL_{t0} = 742.196,00 - 697.878,57 = 44.317,43$$

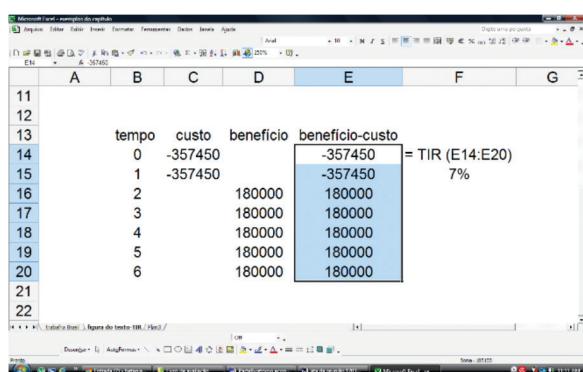
O resultado significa que os benefícios do projeto excedem os custos em R\$ 44.317,43. Então, podemos concluir que o projeto apresenta retorno econômico positivo para a sociedade, sendo viável economicamente.

Cálculo da Taxa Interna de Retorno

No exemplo do projeto "Trabalha Brasil" a *TIR* é de 7%. A fórmula para seu cálculo é:

$$(VB_{t_0} - VC_{t_0}) + \left(\frac{VFB_{t_n}}{(1+TIR)^1} + \dots + \frac{VFB_{t_n}}{(1+TIR)^n} \right) - \left(\frac{VFC_{t_n}}{(1+TIR)^1} + \dots + \frac{VFC_{t_n}}{(1+TIR)^n} \right) = 0$$

Para obtermos a *TIR* utilizamos o software Excel³, no qual inserimos o fluxo de caixa, selecionamos as células e inserirmos a função "*TIR*", como apresentado na figura a seguir:



3. É possível obter a *TIR* com calculadora financeira e outros pacotes computacionais. Optamos por apresentar o exemplo utilizando o Excel, por ser um software muito difundido.

A interpretação da **TIR** depende da taxa mínima de atratividade que considerarmos. Suponha que a taxa mínima de atratividade seja de 5%. Então, a **TIR** de 7% indica que o projeto é viável economicamente, pois tem um retorno implícito maior do que o mínimo que o investidor desejava.

Cálculo da Razão Custo-Benefício

Dos cálculos anteriores, temos que:

VPTB_{t0}: 742.196,00

VPTC_{t0}: 697.878,57

$$\text{Razão Custo-Benefício} = \frac{VPTB_{t_0}}{VPTC_{t_0}}$$

$$\text{Razão Custo-Benefício} = \frac{742196}{697878,57} = 1,06$$

O resultado indica que, para cada real gasto no projeto, o retorno para a sociedade será de R\$ 1,06. Lembre-se de que esse valor foi calculado, considerando uma taxa de desconto intertemporal de 5%.

No exemplo:

$$\text{Retorno} = (1,06 - 1) * 100 = 6,0\%$$

Ou seja, sob à taxa de desconto intertemporal de 5%, o retorno do projeto é de 6%.

8.6. Exemplo 1: projeto para melhoria na qualidade da educação (VASCONCELLOS et al. 2009)

O projeto “Olimpíada Brasileira de Matemática das Escolas Públicas - OBMEP” tem como objetivo a melhoria da qualidade da educação em matemática nas escolas públicas brasileiras. A unidade tratada são as escolas públicas, e o indicador de impacto é a nota média da escola na Prova Brasil dos alunos do 9º ano (antiga 8ª série). A avaliação de impacto utilizou metodologia de diferenças em diferenças com dados da Prova Brasil de 2005 e 2007. Os resultados mostram que o projeto melhora a nota média das escolas participantes (foram utilizados vários métodos no cálculo do impacto).

Para transformar o impacto em benefício, os autores consideraram as seguintes hipóteses:

- i. O impacto positivo estimado na nota média dos alunos do 8º ano vale em valores absolutos para todos os alunos matriculados nas escolas participantes do projeto.
- ii. A melhora no desempenho dos jovens na 8º série afetará os salários futuros dos jovens no mercado de trabalho - com elasticidade estimada de 0,3, segundo estudo de Curi e Menezes-Filho (2007, apud VASCONCELLOS et al., 2009)
- iii. Os retornos da educação no salário são constantes no tempo.

Com base nos dados da Pnad 2007, os autores projetaram qual seria o salário anual correspondente de um jovem com 18 anos de idade e oito anos de estudo. Para obter o benefício do projeto, multiplicaram o impacto na nota em relação aos salários futuros, considerando a elasticidade pressuposta. Depois, multiplicaram o impacto do projeto no aluno médio pelo número de alunos beneficiados, assim obtiveram o benefício do projeto.

No cálculo do custo, o trabalho considera dois cenários: apenas com o custo contábil por aluno, que é de R\$ 2,00; e com o custo econômico (contábil + oportunidade) calculado com base no custo econômico de outro projeto similar, a “Olimpíada de Língua Portuguesa Escrevendo o Futuro”, concurso semelhante à OBMEP realizado pela Fundação Itaú Social e Cenpec. Para os dois cenários, a taxa de desconto intertemporal considerada é de 5%.

Os autores calculam o *VPL* total do projeto, o *VPL* por aluno e a *TIR*, diferenciando as estatísticas de retorno para escolas e alunos que só participaram uma vez da OBEMP, que participaram duas e três vezes. Além disso, repetem os cálculos considerando os dois cenários de custo. Em seguida, calculam a média de cada estatística estimada. Os resultados apontam para um *VPL* total médio de 901 milhões, um *VPL* por aluno médio de R\$ 181,70 e uma *TIR* média de 45% ao ano. Assim, concluem que a OBMEP apresenta taxa de retorno elevada e gera benefícios salariais futuros aos jovens participantes.

8.7. Exemplo 2: projeto para prevenção da violência (PEIXOTO, 2010).

O projeto “Fica Vivo” foi implementado de forma piloto em uma das favelas mais violentas de Belo Horizonte, Minas Gerais, no ano de 2002. O objetivo do projeto é a redução do número de homicídios, por meio de ações que combinam policiamento ostensivo com ações sociais. A unidade tratada é a favela, e o indicador de impacto é a taxa de homicídios por cem mil habitantes. A avaliação de impacto utilizou a combinação da metodologia de pareamento por escore de propensão com a de diferenças em diferenças e os dados foram provenientes do Banco de Dados Georeferenciados da Polícia de Minas Gerais – de 2000 a 2006 – e o Censo Demográfico do IBGE – 2000. Os resultados mostram o impacto do projeto entre 2002 e 2006, que foi de redução de 15 homicídios.

Para transformar o impacto em benefício, a autora considerou as seguintes hipóteses:

- i. O benefício de um homicídio evitado é igual à perda social do homicídio. Assim, o benefício do homicídio evitado é a perda que a sociedade deixou de sofrer porque o homicídio foi evitado. Essa perda tem vários fatores, como anos de produção perdidos, custos judiciais, psicológicos, entre outros.
- ii. No caso do Brasil, Carvalho et al. (2007, apud PEIXOTO, 2010) calculam o valor dos anos de produção perdidos devido ao homicídio. A autora supõe que a proporção dos anos de produção perdidos devido ao homicídio em relação à perda total gerada para a sociedade é igual à do Reino Unido, onde o custo da perda gerada pelo homicídio foi calculado considerando vários fatores (DUBOURG E HAMED, 2005, apud PEIXOTO, 2010).

Para obter o benefício do projeto, a autora aplicou regra de três considerando o valor dos anos de produção perdidos pela proporção que esse fator representa na perda total gerada pelo homicídio no Reino Unido. Assim, calcularam que a perda total gerada pelo homicídio no Brasil é de R\$ 490 mil.

A autora considera o custo econômico do projeto, mas não o subdividiu em custo contábil e custo de oportunidade e sim por uma tipologia de custo própria. Além disso, estima o custo do projeto por três formas diferentes. Os valores de custo anuais médios por beneficiário variam de R\$ 43,17 a R\$ 52,43, e os custos totais anuais médios variam de R\$ 1,019 milhão a R\$ 1,238 milhão, de acordo com o método de estimativa. Para realizar o fluxo de caixa do projeto, considera-se a taxa de desconto intertemporal igual ao índice de inflação IPCA-IBGE variável mensalmente.

A autora calcula a razão custo-efetividade de um homicídio evitado pelo projeto e também a razão custo-benefício, pelas três formas de estimativa do custo. A razão custo-efetividade encontrada varia, de acordo com o método de cálculo do custo, de R\$ 201 mil a R\$ 244 mil e a razão custo-benefício varia de R\$ 1,99 a R\$ 2,42. Os resultados indicam que o projeto gera retorno para a sociedade.

Como foram utilizados vários parâmetros para a estimativa dos resultados, a autora realiza a análise de sensibilidade variando os diferentes valores dos anos de produção perdidos, do percentual que esse componente representa na perda total que o homicídio acarreta para a sociedade e do método de rateio. A análise de sensibilidade mostra que apenas dois casos, das 62 combinações de parâmetros utilizadas, não se mostraram economicamente viáveis. Para as demais combinações de parâmetros o projeto apresenta retorno para a sociedade. Essa análise indica robustez dos resultados estimados.

8.8. Exemplo 3: projeto de prevenção em saúde (CAMELO JUNIOR et al., 2011).

Avaliação econômica da triagem neonatal da galactosemia, no estado de São Paulo. A galactosemia é uma doença metabólica hereditária, que pode ter seus malefícios prevenidos, caso seja detectada precocemente. Se o diagnóstico for tardio a criança que apresenta a doença pode precisar de tratamento de urgência e intensivo, para infecções generalizadas

e insuficiência hepática. Além de a doença poder evoluir para cirrose, insuficiência hepática, cataratas, hipertensão intracraniana, edema cerebral, letargia, hipotonia, retardo mental e morte. Com o diagnóstico precoce, a criança necessita apenas de acompanhamento médico ambulatorial e, se a evolução do quadro clínico for favorável, adquire qualidade de vida e possibilidade real de produtividade social futura. Os autores verificaram o impacto da inclusão do diagnóstico dessa doença no conhecido “Teste do Pezinho”, por meio de um estudo-piloto no ano de 2006, em São Paulo. Os autores consideram como indicador de impacto a incidência da doença em crianças. A base de dados é proveniente do teste realizado em 59.953 recém-nascidos escolhidos aleatoriamente.

Os resultados indicaram que 1:19.984 nascidos vivos teriam o diagnóstico precoce caso o teste para galactosemia fosse incluído no exame do pezinho. Para transformar o impacto em benefício, os autores assumiram as seguintes hipóteses:

- i. O benefício do diagnóstico precoce é igual à diferença entre a perda social com diagnóstico tardio e a perda social causada pela doença quando o diagnóstico é precoce.
- ii. O diagnóstico tardio gera perda para a sociedade de despesa com terapia intensiva e enfermarias; procedimentos cirúrgicos; tratamento dietético; seguimento ambulatorial a longo prazo; e perda de produção por causa das complicações geradas pela doença.
- iii. Mesmo com o diagnóstico precoce, a doença gera perda de despesa com seguimento ambulatorial a longo prazo, sem complicações e com o tratamento dietético.
- iv. O valor da perda de produtividade foi considerado como a perda de meio dia de trabalho (R\$ 21,65), com base no rendimento médio mensal real das pessoas ocupadas no Brasil, em março, de 2009, R\$ 1.298,70.
- v. O paciente trabalharia entre 20 e 65 anos de idade.

Para obter o benefício do projeto, os autores subtraíram a perda gerada pelo diagnóstico tardio da gerada pelo diagnóstico precoce, calculando assim o benefício individual. Em seguida, multiplicaram o benefício pelo total de nascidos vivos com a doença (incidência da doença multiplicada pelo número de nascidos vivos em São Paulo no ano de 2009).

Os autores consideraram o custo econômico do projeto, mas não o subdividiram em custo contábil e custo de oportunidade. São computados os custos do projeto com os *kits* de detecção da

doença, o retorno das crianças detectadas para confirmação, bem como o custo do contato para que esse retorno aconteça, e a perda de produtividade do trabalho do acompanhante. Para realizar o fluxo de caixa do projeto consideraram a taxa de desconto intertemporal igual à SELIC de junho de 2009 – de 9,25%.

O retorno econômico foi mensurado com base na razão custo-benefício da triagem neonatal da galactosemia, considerando-se ainda mais dois níveis de incidência da doença dados pelos limites do intervalo de confiança a 95%. A razão custo-benefício estimada foi de 1,33, e para os intervalos de confiança foi de 0,44 e 3,54. Assim, os resultados mostram que o projeto gera retorno para a sociedade.

A análise de sensibilidade considerou o efeito da variação da taxa de juros, no intervalo entre 0 e 20%, na razão custo-benefício. Observa-se que somente com juros a 20% ao ano é que o projeto deixaria de apresentar retorno.

8.9. Exercícios de revisão

1. A avaliação econômica é composta de duas etapas, cite quais são e os principais objetivos.
2. Por que não podemos comparar dois programas apenas usando os resultados da avaliação de impacto?
3. O que é o benefício do projeto e por que transformar o impacto em benefício?
4. O que é a taxa de desconto intertemporal? E quando ela é utilizada na avaliação econômica?
5. De que é composto o Custo Econômico?
6. Se o custo de oportunidade é implícito, como é possível calculá-lo? Dê exemplos.
7. O que o avaliador pretende captar ao calcular o retorno econômico do projeto?

8. Cite as estatísticas de retorno econômico vistas neste capítulo.

9. Qual a diferença entre as estatísticas de retorno econômico?

10. O que é análise de sensibilidade e qual é a sua finalidade?

8.10. Exercícios práticos

1. O programa “Mulher do Futuro” – fictício – visa ao aumento da renda das mulheres chefes de família. Para isso, oferece cursos profissionalizantes em diversas áreas e ajuda na inserção das participantes no mercado de trabalho, por meio de convênios com empresas. Os cursos duram em média seis meses. Um total de 1.590 mulheres participou do treinamento. A avaliação de impacto do programa foi pensada desde sua concepção. Assim, realizou-se uma pesquisa com as participantes e não participantes antes da implementação. Dois anos após o início do programa, o grupo de avaliação voltou a campo e novamente coletou informações sobre as mulheres. A avaliação de impacto foi realizada pelo método de diferenças em diferenças com pareamento. O resultado da avaliação de impacto mostrou que o programa aumenta em R\$ 277,12, por mês, a renda das participantes.

O custo do projeto é expresso como segue:

Custo Contábil	Valores em R\$
Professores	954.000,00
Pessoal administrativo	254.720,00
Material de consumo	1.670.394,00
Apostilas	67.800,00
Total Contábil	2.946.914,00

Custo de Oportunidade	
Aluguel da sala de aula	460.000,00
Horas dispendidas pelas mulheres com o curso	7.461.642,86
Total Oportunidade	7.921.642,86

Custo do Programa	10.868.556,86
--------------------------	----------------------

- a.** Por que não precisamos transformar o impacto em benefício? Calcule o benefício anual do projeto.
- b.** Monte o fluxo de caixa do projeto supondo que:
- i.** A renda a mais gerada pelo programa perdurará durante toda a vida produtiva da mulher.
 - ii.** As mulheres entram no mercado de trabalho com 20 anos e trabalham por 35 anos.
- c.** Em t0, calcule o benefício total, custo total, *VPL*, a *TIR*, a Razão Custo-Benefício e a Razão Custo-Efetividade. Suponha uma taxa de desconto intertemporal de 5%.
- d.** Interprete os resultados da letra c. O que podemos concluir?
- e.** Realize a análise de sensibilidade considerando:
- i.** O impacto, que apresenta o intervalo de confiança a 95% variando de 248,45 a 305,79.
 - ii.** O retorno do projeto é recebido por apenas 5 anos.

REFERÊNCIAS BIBLIOGRÁFICAS

ABADIE, A.; DIAMOND, A.; HAINMUELLER, J. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*. v. 105, p. 493-505, 2010.

ABADIE, A.; IMBENS, G. Simple and Bias-Corrected Matching Estimator for Average Treatment Effect. NBER Working Paper. n. 283, 2002.

ANGRIST, J. D. Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records. *American Economic Review* 80(3), 313Y335, 1990.

ANGRIST, J. D.; IMBENS G. W. Two-stage least squares estimation of average causal effects using instrumental variables. *Journal of American Statistical Association*. v. 90, p. 430-442, 1995.

ANGRIST, J. D.; IMBENS, G. W.; RUBIN, D. B. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*. v. 91, p. 444-472, 1996.

ANGRIST, J. D.; LAVY, V. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*. v. 114, p. 533-575, 1999.

ANGRIST, J. Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records. *American Economic Review*. v. 80, p. 313-335, 1990.

ANGRIST, J.; KRUEGER, A. The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American Statistical Association*. v. 418, p. 328-336, 1992.

ANGRIST, J.; PISCHKE J. Mostly harmless econometrics: an empiricist's companion. Nova Jersey: Princeton University Press, 2008. 392 p.

ANGRIST, J.D. e PISCHKE, J.S. Mostly Harmless Econometrics: An Empiricists Companion, Princeton University Press, 2009.

ASHENFELTER, O. Estimating the effect of training programs on earnings. *Review of Economics and Statistics*. v. 60, p. 47-57, 1978.

ASSAF NETO, A. *Matemática Financeira e suas Aplicações*. 11^a edição. São Paulo: Atlas, 2009. 296 p.

BANEREE, A.; DUFLO, E. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. Nova York : Public Affairs, 2011. 320 p.

BARROS, R.P.; OLINTO, P.; LUNDE, T.; CARVALHO, M. The impact of access to free childcare on women's labor market outcomes: evidence from a randomized trial in low-income neighborhoods of Rio de Janeiro. 2011. Não publicado.

BERTRAND, M.; DUFLO, E.; MULLAINATHAN, S. How should we trust difference-in-differences estimates?. *Quarterly Journal of Economics*. v. 119, p. 249-275, 2004.

BIONDI, R. L.; VASCONCELLOS, L.; MENEZES FILHO, N. Evaluating the impact of the brazilian public school math olympics on the quality of education. *Economia*. Spring, v. 2, n. 2, p. 143-170, 2012.

BLOOM, H. Accounting for no-shows in experimental evaluation designs. *Evaluation Review*. v. 8, p. 225-246, 1984.

BLOOM, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review* 8(2), 225-246.

BLOOM, H.S. Minimal Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs, *Evaluation Review*, 19: 547-556, 1995.

BLUNDELL, R.; DIAS, M. C. Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*. v. 44, p. 565-640, 2009.

BLUNDELL, R.; DIAS, M. C. Evaluation methods for non-experimental data. *Fiscal Studies*, v. 21, p. 427-468, 2008.

BOARDMAN, A. E.; GREENBERG, D. H.; Vining, A. R.; WEIMER, D. L. Cost Benefit Analysis: Concepts and Practice. Prentice Hall, 2005, 560 pp.

BOURGUIGNON, F.; FERREIRA, F. Exante evaluation of policy reforms using behavioral models. In: BOURGUIGNON, F.; SILVA, L. P. The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools. Washington: World Bank Publications, 2003, p. 123-141.

BRAIDO, L.; OLINTO, P.; PERRONE, H. Gender bias in intrahousehold allocation: evidence from an unintentional experiment. Review of Economics and Statistics. v. 94, p. 552-565, 2012.

BURTLESS, G. The case for randomized field trials in economic and policy research. Journal of Economic Perspectives. v. 9, p. 63-84, 1995.

CAMELO JUNIOR, J. S.; FERNANDES, M. I. M.; JORGE, S. M.; MACIEL, L. M. Z.; SANTOS. J. L. F.; CAMARGO JR, A. S.; PASSADOR, C. S.; CAMELO, S. H. H. Avaliação econômica em saúde: triagem neonatal da galactosemia. Cadernos de Saúde Pública. v. 27, n. 4, p. 666-676, 2011.

CAMELO JUNIOR, J. S.; FERNANDES, M. I. M.; JORGE, S. M.; MACIEL, L. M. Z.; SANTOS. J. L. F. CAMARGO JR, A. S.; PASSADOR, C. S.; CAMELO, S. H. H. Avaliação Econômica em Saúde: Triagem Neonatal da Galactosemia. Cadernos de Saúde Pública, Rio de Janeiro, 27(4):666-676, abril, 2011.

CAMERON, C.A. e MILLER, D.L. A Practitioner's Guide to Cluster-Robust Inference, Journal of Human Resources, 50: 317-372, 2015.

CAMPBELL, H. F.; BROWN, R. P. C. Benefit-Cost Analysis: Financial and Economic Appraisal using Spreadsheets. Cambridge University Press, 2003, 360 pp.

CARD, D.; DOBKIN, C.; MAESTAS, N. The impact of nearly universal insurance coverage on health care utilization and health: evidence from medicare. American Economic Review. v. 98, n. 5, p. 2242-2258, 2008.

CARNEIRO, P.; HECKMAN, J.; VYTLACIL, E. Understanding what instrumental variables estimate: estimating the average and marginal return to schooling. 2005. Não publicado.

CHAY, K.; GREENSTONE, M. Does air quality matter; evidence from the housing market. *Journal of Political Economy*. v. 113, p. 376-424, 2005.

COCHRAN, W. *Sampling Techniques*. New York: John Wiley& Sons, 1977. 428 p.

COHEN, E.; FRANCO, R. *Avaliação de projetos sociais*. 9^a edição. Petrópolis: Vozes, 2011. 312 p.

CRESPO, A. A. *Matemática Financeira Fácil*. 14^a edição. São Paulo: Saraiva, 2009. 255 p.

CRUMP, R.; HOTZ, V. J.; IMBENS, G.; MITINIK, O. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*. v. 96, n. 3, p. 187-199, 2009.

CRUMP, R.; HOTZ, V. J.; IMBENS, G.; MITINIK, O. Nonparametric tests for treatment effect heterogeneity. *Review of Economics and Statistics*. v. 90, n. 3, p.389-405, 2008.

DEATON, A. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, World Bank, Washington, DC. 1997

DEHEJIA, R.; WAHBA, S. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*. v. 84, n. 1, p. 151-161, 2002.

DEHEJIA, R. Program evaluation as a decision problem. *Journal of Econometrics*. v. 125, p. 141-173, 2005.

DEHEJIA, R.; WAHBA, S. Casual effects of nonexperimental studies: reevaluating the evaluation of training programs. *Journal of American Statistical Association*. v. 94, p. 1053-1062, 1999.

DRUMMOND, M. F.; SCULPHER, M. J.; TORRANCE, G. W.; O'BRIEN, B. J.; STODDART, G. L. *Methods for the economic evaluation of health care programmes*. Nova York: Oxford University Press, 2005, 379 p.

DUFLO, E. Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment. *American Economic Review*. v. 91, n. 4, p. 795-813, 2001.

DUFLO, E.; GLENNISTER, R.; KREMER, M. Using randomization in development economics research: A Toolkit. In: T. SCHULTZ; J. STRAUSS. *Handbook of Development Economics*. Oxford: North Holland, 2008. v. 4, p. 3895-3962.

EUROPEAN COMISSION, Guide to Cost-Benefit Analysis of Investment Projects: Structural Funds, Cohesion Fund and Instrument for Pre-Accession, 2008. Disponível em: <http://ec.europa.eu/regional_policy/sources/docgener/guides/cost/guide2008_en.pdf>. Acesso em: 7 maio 2012.

FERMAN, B.; ASSUNÇÃO, J. Does affirmative action enhance or undercut investment incentives?: evidence from quotas in brazilian universities. 2009. Não publicado.

FERRAZ, C.; FINAN, F. Electoral accountability and corruption: evidence from the audit reports of local governments. *American Economic Review*. v.101, p. 1274-1311, 2011.

FIRPO, S.; FOGUEL, M.; JALES, H. Evaluating the impact of stratified randomized experiments with an application to a Brazilian public training program. 2011. Não publicado.

GERTLER, P.; MARTINEZ, S.; PREMAND, P.; RAWLINGS, L e VERMEERSH, C. *Impact Evaluation in Practice*, interactive book, The World Bank.

GRASDAL, A. The performance of sample selection estimators to control for attrition bias. *Health Economics*. v. 10, p. 385-398, 2001.

GUJARATI, D. N. *Econometria Básica*. 3ª edição. São Paulo: Makron Books, 2000. 860 p.

HAHN, J. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*. v. 66, n. 2, p. 315-331, 1998.

HAHN, J.; TODD, P.; VAN DER KLAUW W. Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica*. v. 69, p. 201-209, 2001.

HECKMAN, J. Varieties of selection bias. *American Economic Review*. v. 80, p.313-318, 1990.

HECKMAN, J.; HOTZ, J. Alternative methods for evaluating the impact of training programs. *Journal of the American Statistical Association*. v. 84, n. 804, p. 862-874, 1989.

HECKMAN, J.; ICHIMURA, H.; TODD, P. Matching as an econometric evaluation estimator. *Review of Economic Studies*. v. 65, p. 261-294, 1998.

HECKMAN, J.; LALONDE, R.; SMITH, J. The economics and econometrics of active labor market programs. In: ASHENFELTER, O.; CARD, D. *Handbook of Labor Economics*. Oxford: North Holland, 1999. v. 3, p. 1865-2097.

HECKMAN, J.; LOCHNER, L. e TABER, C. Explaining raising wage inequality: explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *NBER Working Paper*. v. 1, n. 6384, p. 1-58, 1998.

HECKMAN, J.; ROBB, R. Alternative methods for evaluating the impact of interventions. In: HECKMAN, J.; SINGER, B. *Longitudinal analysis of labor market data*. Nova York: Cambridge University Press, 1985.

HECKMAN, J.; SMITH, J. Assessing the case for social experiments. *Journal of Economic Perspectives*. v. 9, p. 85-110, 1995.

HECKMAN, J.; SMITH, J. The pre-program earnings dip and the determinants of participation in a social program: implications for simple program evaluation strategies. *Economic Journal*. v. 109, p. 313-348, 1999.

HECKMAN, J.; VYTLACIL, E. Local Instrumental Variables. In: HSIAO, C; MORIMUNE, K; POWELL, J. *Nonlinear Statistical Modeling: Essays in Honor of Takeshi Amemiya*. Nova York: Cambridge University Press, 2001.

HECKMAN, J.; VYTLACIL, E. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*. v. 73, n. 3, p. 669-738, 2006.

HECKMAN, J; ICHIMURA, H.; TODD, P. Matching as an econometric evaluation estimator: evidence from a job training program. *Review of Economic Studies*. v. 64, n 4, p. 605-54, 1997.

HIRANO, K.; IMBENS, G. Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. *Health Services & Outcomes Research Methodology*. v. 2, p. 259-278, 2001.

HIRANO, K.; IMBENS, G.; RIDDER, G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*. v. 71, n. 4, p. 1161-1189, 2003.

HIRANO, K.; IMBENS, G.; RUBIN, D.; Zhou, X. Identification and estimation of local average treatment effects. *Biostatistics*. v. 1, n. 1, p. 69-88, 2000.

HOFFMANN, R. *Estatística para Economistas*. 4^a edição. São Paulo: Cengage Learning, 2006. 446 p.

IMAI, K, KING, G. e NALL, C. The Essential Role of Pair Matching in Cluster Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation, *Statistical Science*, 24: 29-53. 2009.

IMBENS, G. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*. v. 86, n. 1, p. 1-29, 2004.

IMBENS, G. The role of the propensity score in estimating dose-response functions. *Biometrika*. v. 87, n 3, p. 706-710, 2000.

IMBENS, G., KING, G., MCKENZIE, D. e RIDDER, G. On the Benefits of Stratification on Randomized Experiments, mimeo, Departamento de Economia, Universidade de Harvard. 2009.

IMBENS, G.; ANGRIST, J. Identification and estimation of local average treatment effects. *Econometrica*. v. 61, n. 2, p. 467-476, 1994.

IMBENS, G.; LEMIEUX, T. Regression discontinuity designs: a guide to practice. *Journal of Econometrics*. v. 142, n. 2, p. 615-635, 2008.

IMBENS, G.; RUBIN, D. Rubin causal model. In: DURLAUF, S.; BLUME, L. *The New Palgrave Dictionary of Economics*. Hampshire: Palgrave Macmillan, 2008, 2^a edição. 7344 p.

IMBENS, G.; WOOLDRIDGE, J. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*. v. 47, n. 1, p. 5-86, 2009.

IPEA e CEDEPLAR. Uma proposta de avaliação do plano nacional de formação profissional (planfor). 2000. Relatório parcial. Não publicado.

LALONDE, R. J. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*. v. 76, p. 604-620, 1986.

LEE, D. S.; CARD, D. Regression discontinuity inference with specification error. *Journal of Econometrics*. v. 142, n. 2, p. 655-674, 2008.

LEE, D. S.; LEMIEUX, T. Regression Discontinuity Designs in Economics. *NBER Working Paper*. v. 48, n. 14723, p. 281-355, 2009.

LEE, D. S.; MORETTI, E.; BUTLER, M. Do voters affect or elect policies? Evidence from the U.S. house. *Quarterly Journal of Economic*. v. 119, p. 807-859, 2004.

MATTOS, E.; MAIA, S.; MARQUES F. Evidências da relação entre oferta de trabalho e programas de transferência de renda no Brasil: bolsa escola versus bolsa família. *Pesquisa e Planejamento Econômico* v. 40, p. 1-44, 2010.

MCEWAN, H. M.; LEVIN, P. J. Cost-effectiveness analysis: methods and applications. California: Sage Publications, 2001. 308 p.

MCRARY, J. Testing for manipulation of the running variable in the regression discontinuity design. *Journal of Econometrics*. v. 142, n. 2, p. 698-714, 2008.

MEYER, B. Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics*. v. 13, p. 151-161, 1995.

MIGUEL, E.; KREMER, M. Worms: identifying the impacts on education and health in the presence of treatment externalities. *Econometrica*. v. 72, p.159-217, 2004.

PEIXOTO, B. Avaliação econômica do projeto fica Vivo: o caso piloto. In: Ministério do Planejamento, Orçamento e Gestão, Secretaria de Orçamento Federal. (Org.). Orçamento Público: II Prêmio SOF de Monografias - Coletânea. Brasília, 2010.

PONCZEK, V.; SOUZA, A. New evidence of the causal effect of family size on child quality in a developing country. *Journal of Human Resources*. v. 47, p. 64-106, 2011.

PORTER, J. Estimation in the regression discontinuity model. Department of Economics. 2003. Não publicado.

ROBINS, J. M.; RITOY, Y. A curse of dimensionality appropriate (coda) asymptotic theory for semiparametric models. *Statistics in Medicine*. v. 16, p.285-319, 1997.

ROCHA, R.; SOARES, R. Evaluating the impact of community-based health interventions: evidence from brazil's family health program. *Health Economics*. v. 19, p. 126-158, 2010.

ROSENBAUM, P. Observational studies. Nova York: Springer Verlag, 2002. 375p.

ROSENBAUM, P. The role of a second control group in an observational study. *Statistical science*. v. 2, n. 3, p. 292–316, 1987.

ROSENBAUM, P.; RUBIN, D. The central role of the propensity score in observational studies for causal effects. *Biometrika*, v. 70, p. 41-55, 1983.

RUBIN, D. Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*. v. 2, n. 1, p. 1-26, 1977.

RUBIN, D. Estimating causal effects to treatments in randomized and nonrandomised studies. *Journal of Educational Psychology*. v. 66, p. 688-701, 1974.

RUBIN, D. Matching to remove bias in observational studies. *Biometrics*. v. 29, p. 159-183, 1973.

RUBIN, D. The use of matched sampling and regression adjustments to remove bias in observational

studies. *Biometrics*. v. 29, p. 185-203, 1973.

RUBIN, D. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of American Statistical Association*. v. 74, p. 318-328, 1979.

SMITH, J e TODD, P. Does matching overcome LaLonde's critique of nonexperimental estimators?, *Journal of Econometrics*, 125:305-353, 2005.

THISTLEWAITE, D.; CAMPBELL, D. Regression-discontinuity analysis: an alternative to the ex post facto experiment. *Journal of Educational Psychology*. v. 51, p.309-317, 1960.

TODD, P; WOLPIN, K. Ex ante Evaluation of Social Programs. PIER Working Paper Penn Institute for Economic Research. n. 06-122, 2006.

TRIOLA, M. F. *Introdução à Estatística*. 10^a edição. Rio de Janeiro: LTC, 2008. 686 p.

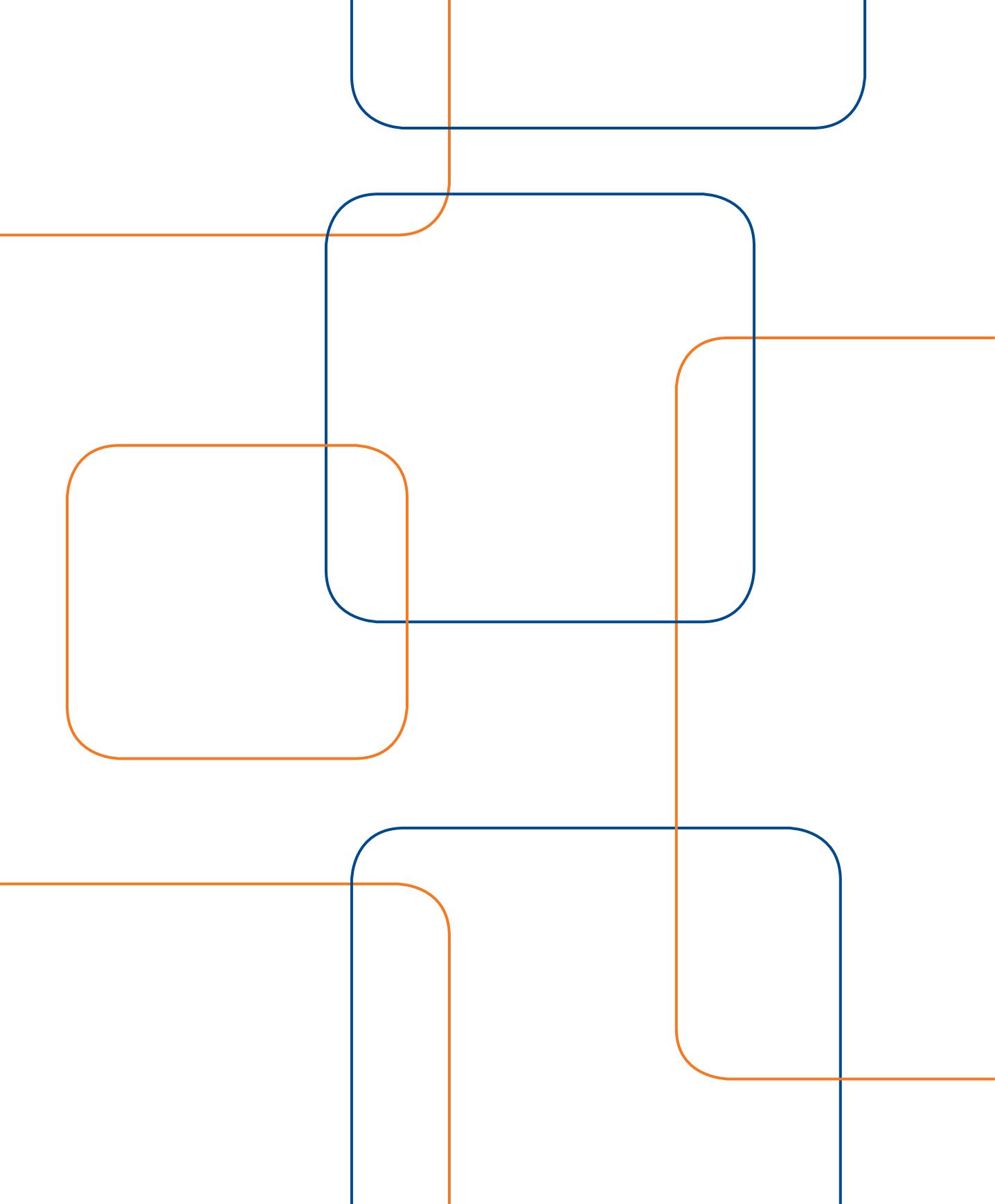
VAN DER KLAUW, W. Estimating the effect of financial aid offers on college enrollment: a regression-discontinuity approach. *International Economic Review*. v. 43, p. 1249-1287, 2002.

VASCONCELLOS, L.; BIONDI, R. L.; MENEZES FILHO, N. Avaliação Econômica da Olimpíada Brasileira de Matemática das Escolas Públicas (OBMEP), Relatório de Avaliação Econômica – Fundação Itaú Social, 2009. Disponível em: <<http://ww2.itau.com.br/itausocial2/pdf/obmep.pdf>>. Acesso em: 7 maio 2012.

VYTLACIL, E. Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*. v. 70, n. 1, p. 331-341, 2002.

WOOLDRIDGE, J. M. *Introdução à Econometria: Uma Abordagem Moderna*. São Paulo: Thomson, 2006. 725 p.

WOOLDRIDGE, J.M. *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Massachusetts, Estados Unidos. 2002.



Apêndice

EXERCÍCIOS PRÁTICOS

- 1.** O programa “Mulher do Futuro” – fictício – objetiva aumentar a renda de famílias chefiadas por mulheres. Para isso, oferece cursos profissionalizantes em diversas áreas e ajuda na inserção das participantes no mercado de trabalho, por meio de convênios com empresas. Os cursos duram em média seis meses. A avaliação de impacto do programa foi pensada desde sua concepção. Assim, realizou-se uma pesquisa com as participantes e não participantes antes da implementação. Dois anos após o início do programa, o grupo de avaliação voltou a campo e novamente coletou informações sobre as mulheres. O banco de dados “DDM” contém as informações em dois momentos no tempo – escolaridade, idade, se o domicílio possui luz elétrica, se a mulher vive com cônjuge e a renda familiar *per capita*.
- a.** Suponha que o experimento foi aleatório e que possuímos dados do programa em um momento do tempo - após o programa. Calcule o impacto na renda *per capita*.
- b.** Faça o teste de diferença de médias entre tratado e controle para as variáveis: escolaridade, idade, cônjuge, no período após o programa.
- c.** Como o teste de diferença de médias mostrou que os grupos tratado e de controle não são estatisticamente iguais em todas as características observadas, calcule o impacto do programa controlando para as variáveis observáveis, ainda supondo os dados em um momento do tempo - após o programa. Esse resultado mostrou alguma diferença para o obtido na letra b? Se sim, explique.
- d.** Ainda considerando os dados em apenas um momento do tempo, estime o impacto do programa pelo método de pareamento por escore de propensão – um para um, vizinho mais próximo, raio e Kernel.
- e.** Calcule o impacto do programa pelo método de diferenças em diferenças. Analise os resultados.
- f.** Faça uma análise descritiva e o teste de diferença de médias das características dos participantes e não participantes antes do início do programa. O que podemos concluir dos resultados do teste?

- g.** Calcule o impacto do programa combinando o método de pareamento com o de diferenças em diferenças. Suponha que o resultado da avaliação de impacto mostrou que o programa aumenta em R\$ 277,12, por mês, a renda das participantes. O custo do projeto é expresso como segue:

Custo Contábil	Valores em R\$
Professores	954.000,00
Pessoal administrativo	254.720,00
Material de consumo	1.670.394,00
Apostilas	67.800,00
Total Contábil	2.946.914,00

Custo de Oportunidade	
Aluguel da sala de aula	460.000,00
Horas dispendidas pelas mulheres com o curso	7.461.642,86
Total de Oportunidade	7.921.642,86

Custo do Programa	10.868.556,86
--------------------------	----------------------

- h.** Por que não precisamos transformar o impacto em benefício? Calcule o benefício anual do projeto.
- i.** Monte o fluxo de caixa do projeto supondo que a renda a mais gerada pelo programa perdurará durante toda a vida produtiva da mulher e que as mulheres entram no mercado de trabalho com 20 anos e trabalham por 35 anos.
- j.** Em t_0 , calcule o benefício total, custo total, VPL , a TIR , a Razão Custo-Benefício e a Razão Custo-Efetividade. Suponha uma taxa de desconto intertemporal de 5%.
- k.** Faça uma análise de sensibilidade considerando:
- i.** O impacto, que apresenta o intervalo de confiança a 95%, variando de R\$ 248,45 a R\$ 305,79.
 - ii.** O retorno do projeto é recebido por apenas 5 anos.

- 2.** O programa “Brasil Melhor” – fictício – visa a melhorar o desempenho escolar dos alunos do ensino fundamental. Esse programa foi desenvolvido com 245 alunos da Escola Estadual João dos Santos. A escola possui ao todo 745 alunos. O desenho do programa não abrangia a sua avaliação de impacto que foi implementada após um ano de seu funcionamento. Dessa forma, só foi possível obter informações em um momento no tempo – depois do programa implementado. O banco de dados “PSM1” apresenta informações sobre a nota no exame de proficiência realizado em toda a escola fundamental, a participação no programa, o sexo, a cor e os anos de estudos da mãe do aluno.
- a.** Calcule o impacto do programa supondo que o experimento foi aleatório. Analise os resultados.
- b.** Faça o teste de diferença de médias para as variáveis: sexo; cor; anos de estudos da mãe. O grupo tratado e controle apresentam diferenças estatisticamente significativas nas características médias? O que podemos concluir com esse teste?
- c.** Calcule o impacto do programa supondo que o experimento foi não aleatório pelo método de regressão MQO, usando todos os não tratados como controle. Analise os resultados.
- d.** Calcule o impacto do programa supondo que o experimento foi não aleatório pelo método de pareamento por escore de propensão – método um para um.
- e.** Faça o teste de diferença de médias para as variáveis: sexo; cor; anos de estudos da mãe. Utilize apenas o grupo de controle selecionado pelo pareamento. Compare o resultado desse teste de médias com o realizado na letra b. O que podemos concluir com essa comparação?
- f.** Faça o gráfico do escore de propensão para os tratados e controles selecionados pelo método de pareamento. O que podemos concluir com esse gráfico?
- g.** Repita os procedimentos de “e” a “g” utilizando as metodologias de pareamento por escore de propensão – método vizinho mais próximo, raio e método de Kernel.

PRÉ-REQUISITO

Conceitos de Estatística e Econometria

Estatística

1. Distribuições

- Conjuntas, Marginais e Condicionais
- Lei de Bayes

2. Momentos das Distribuições

- Esperança Condicional e Não Condisional
- Variância Condicional e Não Condisional
- Lei das Expectativas Iteradas

3. Grandes Amostras

- Conceitos de Convergência: Probabilidade e Distribuição
- Lei Fraca dos Grandes Números
- Teorema Central do Limite

4. Estimadores

- Propriedades em Pequena Amostra: Não Viés e Eficiência
- Propriedades em Grande Amostra: Consistência e Eficiência Assintótica

Econometria

5. Regressão Linear Múltipla

- Conceito de Exogeneidade
- Conceito de Homocedasticidade
- Estimador de Mínimos Quadrados Ordinários
- Propriedades dos Estimadores: Consistência e Não Viés
- Eficiência do Estimador de Mínimos Quadrados Ordinários
- Distribuição Assintótica do Estimador

6. Variável Instrumental

- Estimador de Variável Instrumental
- Estimador de Mínimos Quadrados em Dois Estágios

7. Modelos Logit e Probit

- Estimador de Máxima Verossimilhança
- Aplicação e Interpretação do Estimador

8. Dados em Painel

- Estimador de Efeito Fixo
- Estimador de Primeira Diferença

LEITURA COMPLEMENTAR

I. Estatística

TRIOLA, Mario F. Introdução à Estatística. 10^a ed. Rio de Janeiro: LTC, 2008.

Esse manual é bastante adotado em cursos de graduação nas ciências humanas. Apresenta uma linguagem fácil e compreensível para alunos com um conhecimento menos profundo de álgebra. Além disso, apresenta seções especiais com profissionais de vários campos que utilizam a estatística como ferramenta de trabalho, o que torna a leitura fluida e atraente. O diferencial é a quantidade de exercícios (mais de 1.500) em várias áreas das ciências humanas e seu foco mais centrado na interpretação do que nos cálculos. Muitos dos exercícios usam dados reais, o que os torna mais interessantes para os leitores. O livro aborda temas mais avançados como correlação, regressão e estatística não paramétrica.

HOFFMANN, Rodolfo. Estatística para Economistas. 4^a ed. São Paulo: Thompson, 2006.

Esse livro ensina estatística básica para iniciantes. Apesar de o título direcioná-lo para economistas, atende qualquer aluno das ciências humanas interessado em aplicar a estatística no seu campo de trabalho. Com uma linguagem simples e objetiva, sem perder o rigor formal, apresenta os aspectos conceituais e metodológicos da estatística. Mais ainda, o livro adentra o tópico de regressão linear, o que permite ao leitor ter uma introdução ao tema.

II. Econometria

WOOLDRIDGE, Jeffrey. M. "Econometric Analysis of Cross Section and Panel Data", 2010, MIT Press.

Esse é o principal livro-texto de microeconometria que existe atualmente, utilizado nos principais cursos de pós-graduação em economia do mundo. O texto combina rigor metodológico na abordagem econométrica com uma série de exemplos práticos. A parte inicial faz um ótima revisão dos modelos lineares e generalizados para estimativa de uma equação e também de sistemas de equações. Em seguida são expostos os modelos econométricos para dados em painel e também para variáveis dependentes discretas e censuradas. O capítulo 21 traz uma discussão importante sobre os modelos de resultados potenciais e estimativa de efeitos de tratamentos, um pouco mais

técnica mas bastante complementar à abordagem adotada nesse livro. Enfim, um livro essencial para quem quiser se aprofundar em microeconometria.

WOOLDRIDGE, Jeffrey. M. Introdução à Econometria: Uma Abordagem Moderna. São Paulo: Thomson, 2006.

Esse manual de introdução à econometria alia a matemática existente nesse método de pesquisa empírica econômica com uma ampla interpretação prática de problemas estudados em vários campos das ciências humanas. Assim, mostra que esse instrumental analítico, antes restrito à problemática econômica, pode ser aplicado em outras ciências para análise de questões empíricas. Com esse enfoque, o livro combina questões mais técnicas da econometria com suas aplicações práticas, facilitando seu entendimento. A estrutura dos tópicos apresentados é diferente da tradicional e segue a divisão por tipos de dados. Essa abordagem moderna é muito intuitiva e torna o conteúdo mais acessível ao estudante.

GUJARATI, Damodar N. Econometria Básica. 3^a Ed., Makron Books, 2000.

O manual de econometria básica adota uma abordagem tradicional do tema. Primeiramente é apresentado o modelo de regressão linear, em seguida são listadas as hipóteses necessárias e os problemas decorrentes da sua não observação. Por fim, tópicos especiais são discutidos detalhadamente, como regressão sobre variável binária, equações simultâneas e modelo de séries temporais. Para compreensão do conteúdo é requerido nível básico de cálculo, álgebra e estatística.

III. Matemática Financeira

CRESPO, Antônio A. Matemática Financeira Fácil. 14^a ed. São Paulo: Saraiva, 2009.

Esse livro é uma obra básica de matemática financeira. Escrito para quem nunca estudou o tema, traz os tópicos iniciais como juros simples e compostos, descontos, séries de pagamentos e amortizações. Com muitos exemplos resolvidos e exercícios práticos, é excelente para um primeiro contato com a matéria, principalmente para aqueles que não são familiarizados com a matemática. Entretanto, para quem deseja se aprofundar, a complementação com outros livros é necessária.

ASSAF NETO, Alexandre. Matemática Financeira e suas Aplicações. 11^a ed. São Paulo: Atlas, 2009.

Esse manual de matemática financeira é utilizado em cursos de graduação e pós-graduação. Tem uma abordagem ampla do tema, incluindo desde os princípios básicos da matemática financeira, como juros simples e compostos, até tópicos mais avançados, como análise de ativos financeiros. A sequência dos conteúdos é intuitiva e inclui os principais produtos financeiros existentes no Brasil. Dessa forma, o leitor aprende os fundamentos teóricos da matemática financeira e suas aplicações práticas com ativos financeiros no país.

CURRÍCULOS – AUTORES

Naercio Menezes Filho

Naercio Menezes Filho é professor titular (Cátedra IFB) e coordenador do Centro de Políticas Públicas do Insper Instituto de Ensino e Pesquisa. Membro Titular da Academia Brasileira de Ciências, Professor associado da FEA-USP, colunista do Valor Econômico e consultor do Itaú Social, Naercio é PhD em Economia pela University College London.

Cristine Campos de Xavier Pinto

PhD em Economia pela Universidade da Califórnia. Berkeley. Professora associada da Escola de Economia de São Paulo, FGV. Especialista nas áreas de Econometria da Educação e Econometria. Ministra cursos de avaliação de políticas para alunos de graduação e pós-graduação e também gestores.

Lycia Lima

Coordenadora do Centro de Aprendizagem em Avaliação e Resultados da Escola de Economia de São Paulo da FGV (FGV/EESP Clear) e pesquisadora associada do Centro de Microeconomia Aplicada (C-Micro) e Centro de Política e Economia do Setor Público (CEPESP) da FGV. É doutoranda em Administração Pública e Governo na Fundação Getulio Vargas, mestre em Economia do Desenvolvimento pela School of Oriental and African Studies da Universidade de Londres e graduada em Economia pela Universidade Federal de Minas Gerais (UFMG). Possui mais de 12 anos de experiência na área de políticas públicas e desenvolvimento e 8 anos de experiência com avaliação de programas nas áreas de educação, primeira infância, pobreza, governança e reformas institucionais. Trabalhou previamente no Banco Mundial em Washington, na ONU em Genebra, na Secretaria de Assuntos Estratégicos da Presidência da República e coordenou o núcleo de Avaliação de Políticas Públicas da Fundação João Pinheiro do Governo de Minas Gerais.

Betânia Peixoto

É Auditora Federal de Finanças e Controle e Diretora de Ensino a Distância da Escola de Administração Fazendária do Ministério da Fazenda. Tem doutorado em economia pelo CEDEPLAR/UFMG, 2008. Possui experiência na área de Economia, com ênfase em Avaliação Econômica de Programas Sociais e Econometria aplicada, principalmente nos seguintes temas: Avaliação Econômica de Políticas Públicas e Programas Sociais, Análise Econométrica Aplicada, Criminalidade e Segurança Pública.

Ricardo Paes de Barros

É graduado em engenharia eletrônica pelo Instituto Tecnológico da Aeronáutica (ITA), com mestrado em estatística pelo Instituto de Matemática Pura e Aplicada (IMPA) e doutorado em Economia pela Universidade de Chicago. Possui pós-doutorado pelo Centro de Pesquisa em Economia da Universidade de Chicago e pelo Centro de Crescimento Econômico da Universidade de Yale.

Integrou o Instituto de Pesquisa Econômica e Aplicada (IPEA) por mais de 30 anos, onde realizou inúmeras pesquisas focadas em questões relacionadas aos temas de desigualdade e pobreza, mercado de trabalho e educação no Brasil e na América Latina. Entre 1990 e 1996, Ricardo foi professor visitante da Universidade de Yale e, entre 1999 e 2002, diretor do Conselho de Estudos Sociais do IPEA. Entre 2011 e 2015, foi subsecretário de Ações Estratégicas da Secretaria de Assuntos Estratégicos da Presidência da República.

Como acadêmico, publicou diversos artigos e livros sobre seus temas de pesquisa, recebendo importantes prêmios em reconhecimento ao seu trabalho. Entre eles, cabe destacar o Prêmio Haralambos Simeonidis em 1995 e em 2000 e o Prêmio Mario Henrique Simonsen em 2000. Ricardo foi agraciado com a comenda da Ordem Nacional do Mérito Científico em 2005, eleito membro titular da Academia Brasileira de Ciências em 2010 e recebeu, em 2012, a primeira edição do Prêmio Celso Furtado em Estudos Sociais, oferecido pela Academia Mundial de Ciências (The World Academy of Sciences – TWAS).

No ano de 2015, Ricardo P. B. deixou o serviço público, assumindo a Cátedra Instituto Ayrton Senna no Insper, onde se dedica ao uso de evidência científica para identificação de grandes desafios nacionais e para a formulação e avaliação de políticas públicas, cobrindo os temas de produtividade do trabalho, educação, primeira infância, juventude, demografia, imigração, além dos tradicionalmente recorrentes em sua trajetória, desigualdade, pobreza e mercado de trabalho.

Miguel Nathan Foguel

É pesquisador do IPEA desde 1998 e professor auxiliar da PUC/RJ desde 2009. Tem mestrado pela PUC/RJ e doutorado pela UFF. Já publicou diversos estudos e artigos em livros e revistas científicas nas áreas de avaliação de políticas sociais, economia do trabalho, desigualdade de renda e pobreza. Recebeu o Prêmio Mario Henrique Simonsen, concedido pela Revista Brasileira de Economia, em 2000. É coordenador da rede internacional de pesquisa Network on Inequality and Poverty (NIP) no Brasil desde 2009.

RESPOSTAS DE EXERCÍCIOS SELECIONADOS

Capítulo 2 – Exercício 5:

- a. Calcule o efeito médio do programa (*EMP*) e o efeito médio do programa sobre os tratados (*EMPT*).

$$\begin{aligned}\widehat{EMP} &= \hat{E}[Y_i(1)] - \hat{E}[Y_i(0)] = \frac{1}{5} \sum_{i=1}^5 Y_i(1) - \frac{1}{5} \sum_{i=1}^5 Y_i(0) \\ &= \frac{1}{5} [15 + 16 + 14 + 17 + 14] - \frac{1}{5} [15 + 12 + 15 + 15 + 11] = \frac{76 - 68}{5} \\ &= \frac{8}{5} = 1,6\end{aligned}$$

$$\begin{aligned}\widehat{EMPT} &= \hat{E}[Y_i(1)|T_i = 1] - \hat{E}[Y_i(0)|T_i = 1] \\ &= \frac{1}{3} \sum_{i=1}^3 [Y_i(1)|T_i = 1] - \frac{1}{3} \sum_{i=1}^3 [Y_i(0)|T_i = 1] \\ &= \frac{1}{3} [16 + 14 + 14] - \frac{1}{3} [12 + 15 + 11] = \frac{44}{3} - \frac{38}{3} = \frac{6}{3} = 2,0\end{aligned}$$

- b. Calcule as médias dos resultados observados para quem passou e não passou pela intervenção.

$$\hat{E}[Y_i|T_i = 1] = \frac{1}{3} \sum_{i=1}^3 [Y_i|T_i = 1] = \frac{1}{3} [16 + 14 + 14] = \frac{44}{3} = 14,7$$

$$\hat{E}[Y_i|T_i = 0] = \frac{1}{2} \sum_{i=1}^2 [Y_i|T_i = 0] = \frac{1}{2} [15 + 15] = 15,0$$

- c. Qual a diferença das médias do item (b) para quem passou e não passou pela intervenção?

$$\hat{E}[Y_i|T_i = 1] - \hat{E}[Y_i|T_i = 0] = 14,7 - 15,0 = -0,3$$

- d. Compare a diferença de médias do item (c) com o *EMPT*. Por que eles são diferentes?

Notando que $\hat{E}[Y_i|T_i = 1] = \hat{E}[Y_i(1)|T_i = 1]$ e que $\hat{E}[Y_i|T_i = 0] = \hat{E}[Y_i(0)|T_i = 0]$, a diferença do item (c) pode ser reescrita como: $\hat{E}[Y_i(1)|T_i = 1] - \hat{E}[Y_i(0)|T_i = 0]$ (Note que por causa do

segundo termo, essa nova expressão não fornece o *EMPT*) Somando e subtraindo $\hat{E}[Y_i(0)|T_i = 1]$ dessa expressão:

$$\begin{aligned} & \{\hat{E}[Y_i(1)|T_i = 1] - \hat{E}[Y_i(0)|T_i = 1]\} + \{\hat{E}[Y_i(0)|T_i = 1] - \hat{E}[Y_i(0)|T_i = 0]\} \\ &= \left\{ \frac{44}{3} - \frac{38}{3} \right\} + \left\{ \frac{38}{3} - \frac{30}{2} \right\} = \{2,0\} + \{12,7 - 15\} = \{2,0\} + \{-2,3\} \\ &= \text{EMPT} + \text{Viés de Seleção} \end{aligned}$$

Ou seja, a simples diferença de médias do item (c) contém o *EMPT* somado ao viés de seleção. Note que ela forneceria um efeito negativo da intervenção, ao passo que o verdadeiro efeito é positivo.

Capítulo 2 – Exercício 6

Usando a equação (6), temos:

$$E[Y_i|T_i = 1] - E[Y_i|T_i = 0] = \{\alpha + E[\beta_i|T_i = 1] + E[\varepsilon_i|T_i = 1]\} - \{\alpha + E[\varepsilon_i|T_i = 0]\} = E[\beta_i|T_i = 1] + \{E[\varepsilon_i|T_i = 1] - E[\varepsilon_i|T_i = 0]\},$$

onde $E[\beta_i|T_i = 1]$ é o *EMPT* e o viés de seleção é representado por $\{E[\varepsilon_i|T_i = 1] - E[\varepsilon_i|T_i = 0]\}$

Capítulo 8 – Exercícios de Revisão

1. Resposta: a avaliação econômica é composta da avaliação de impacto econômico e avaliação de retorno econômico. A primeira visa a descobrir se o projeto alcançou os objetivos propostos, ou seja, seu impacto. A avaliação de retorno, por sua vez, visa a comparar os impactos alcançados com os custos incorridos para sua obtenção.

2. Resposta: a avaliação de impacto fornece o resultado obtido com o programa, mas não considera o custo incorrido para obtenção do resultado. Ocorrem situações em que um programa tem impacto maior do que o outro, porém o custo para atingir esse impacto é tão alto que o programa de menor impacto é mais viável economicamente. Portanto, comparar programas

apenas com o resultado da avaliação de impacto pode levar a conclusões erradas sobre qual é mais eficiente, pela ausência de um elemento importante que é o custo.

3. Resposta: o benefício do projeto é a expressão do seu impacto em termos monetários. Os avaliadores transformam o impacto em benefício, pois assim podem comparar com o custo. Tanto o benefício quanto o custo estão em uma mesma unidade de medida, a moeda. Se não fosse essa transformação, o impacto e o custo seriam expressos em unidades diferentes, o custo em moeda e o impacto na unidade do indicador de impacto.

4. Resposta: a taxa de desconto intertemporal indica quanto o indivíduo valoriza o consumo presente em relação ao futuro. É o percentual pelo qual ele está disposto a abrir mão do consumo no presente para só fazê-lo no futuro. Ela deve ser utilizada na avaliação todas as vezes que for necessário somar valores em momentos diferentes no tempo.

5. Resposta: o custo econômico é a soma do custo contábil com o custo de oportunidade. O custo contábil é todo dispêndio incorrido para o funcionamento do projeto que, via de regra, é registrado nos demonstrativos contábeis. O custo de oportunidade é o ganho que poderia advir de uma aplicação alternativa do recurso. É um custo implícito e não aparece nos registros contábeis.

6. Resposta: o cálculo do custo de oportunidade é realizado por estimação. O avaliador precisa investigar os ganhos advindos de aplicações alternativas. Suponha um programa para reflorestamento de uma área. Um custo de oportunidade de reflorestar é a safra agrícola que deixou de ser produzida no mesmo local. Para calculá-la o avaliador precisa multiplicar a área que foi reflorestada pela produção média da região e pelo preço do produto.

7. Resposta: ao calcular o retorno econômico do projeto, o avaliador pretende descobrir se ele é viável economicamente. Viabilidade econômica diz respeito ao projeto gerar um retorno para sociedade, ou seja, se o benefício gerado é maior que o recurso absolvido para sua realização.

8. Resposta: são quatro as estatísticas de retorno econômica vistas nesse capítulo, Valor Presente Líquido –VPL, Taxa Interna de Retorno – TIR, Razão Custo-Benefício, Razão Custo-Efetividade.

9. Resposta: todas as estatísticas de retorno econômico pretendem mensurar se o projeto é viável economicamente, ou seja, todas comparam o custo do projeto com as benesses que são geradas para a sociedade. O que as diferem é a forma pela qual essa comparação é estimada. O *VPL* subtrai os custos do benefício. Quando positivo, o benefício gerado é maior do que o custo, sendo o programa viável economicamente. A *TIR*, por sua vez, calcula a taxa implícita de retorno do projeto. Se a *TIR* é maior que uma taxa mínima de atratividade, então o projeto é viável economicamente, pois rende mais que o mínimo desejado pelo investidor. Ou contrário, se a *TIR* é menor que uma taxa mínima de atratividade, o projeto não é viável economicamente, pois gera menos retorno para sociedade do que inicialmente desejado. A Razão Custo-Benefício é calculada pela divisão do benefício pelo custo. Essa razão indica o quanto de benefício é gerado para cada real investido no projeto. Por fim, a Razão Custo-Efetividade é calculada quando não se quer, ou não se consegue, imputar valor ao impacto do programa. É computada pela divisão do custo pelo impacto alcançado. O resultado da divisão indica o custo incorrido em cada unidade de impacto gerada. Essa estatística não permite que se verifique a viabilidade econômica, exatamente porque não atribui valor monetário ao impacto.

10. Resposta: a análise de sensibilidade é uma técnica de verificação de robustez dos resultados da avaliação econômica. Tem por finalidade analisar até que ponto os resultados estimados se matêm, tendo em vista a variação dos parâmetros adotados na avaliação econômica.

Capítulo 8 – Exercícios Práticos

a. Resposta: nos casos em que o indicador de impacto é expresso em termos monetários, a transformação é desnecessária. Portanto, é possível sem nenhuma transformação compará-lo com o custo, porque ambos são expressos em termos monetários.

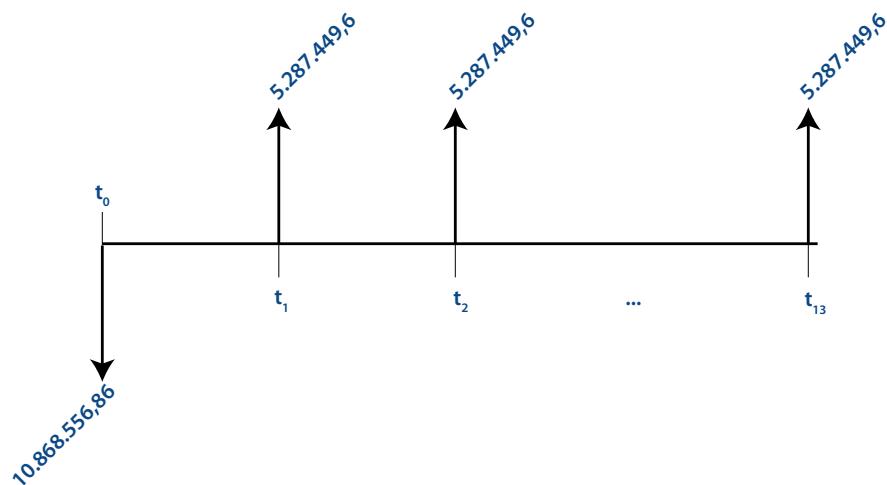
Cálculo do benefício anual:

Item	Valor	Fórmula de Cálculo
Benefício mensal individual	277,12	
Benefício anual individual	3.325,44	=277,12*12
Número de mulheres tratadas	1.590	
Benefício anual total do projeto	5.287.449,60	=1.590*3.325,44

b. Resposta: a média da idade das mulheres tratadas no ano seguinte ao tratamento é de 42 anos (média da variável de idade das tratadas na planilha *DDM*). Assim, elas, em média, já trabalharam 22 anos. Portanto, elas receberão o benefício por 13 anos.

Item	Valor	Fórmula de Cálculo
Idade média das tratadas	42	
Idade de início de trabalho	20	
Total de anos que se trabalha	35	
Total de anos que as mulheres receberão o benefício do projeto	13	=35-(42-20)

O custo do programa acontece apenas no tempo 0 e o benefício começa a ser recebido no tempo 1 e perdura por 13 anos.



c. Resposta:

Estatística	Resultado
VPTB	R\$ 49.668.043,73
VPTC	R\$ 10.868.556,86
VPL	R\$ 38.799.486,88
TIR	48%
C/B	R\$ 4,57

Nota: C/B é a Razão Custo-Benefício

d. Resposta: o benefício do programa superou os custos em cerca de 38,8 milhões, como indica o *VPL*.

A *TIR* de 48% precisa ser comparada com uma taxa mínima de atratividade. O retorno da *TIR* foi tão alto que é maior do que qualquer taxa de investimento, indicando que o programa gera um retorno muito alto.

A Razão Custo-Benefício de R\$ 4,57 mostra que o retorno de cada real investido no projeto é de R\$ 4,57.

O cálculo da Razão Custo-Efetividade nos casos em que o projeto tem impacto medido monetariamente não faz sentido.

Com base nesses resultados podemos concluir que o projeto gera retorno para a sociedade.

e. Resposta: foram estimados um total de 8 cenários. Para todos eles o programa apresentou retorno para a sociedade, indicando a robustez dos resultados da avaliação econômica.





Itaú Social