



# Avaliação Econômica <sup>de</sup> projetos sociais

# Ficha Técnica

ISBN 978-85-98149-04-2

## Iniciativa

Fundação Itaú Social

### Vice Presidente

Antonio Jacinto Matias

### Superintendente

Valéria Veiga Riccomini

### Gerente

Isabel Cristina Santana

### Coordenador do Programa

Antonio Bara Bresolin

## Coordenação Técnica

Itaú Unibanco

### Gerência de Avaliação de Projetos

Lígia Maria de Vasconcellos

Clarissa Gondim Teixeira

## Créditos da Publicação

### Organizador

Naercio Menezes Filho

### Autores

Betânia Peixoto

Cristine Campos de Xavier Pinto

Lycia Lima

Miguel Nathan Foguel

Ricardo Paes de Barros

### Projeto Gráfico e Diagramação

André Godoi Comunicação Completa

### Apoio

Carlos Eduardo Garrido

Evandro da Silva e Silva

Gabriela Gall Rosa

Luciana Rocha Vidal

### Revisão de textos

José Paulo Ferrer

## 1ª Edição - Brasil - 2012

Dinâmica Gráfica e Editora Ltda.

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

Avaliação econômica de projetos sociais / Betânia Peixoto...[et al.] ; organizador Naercio Menezes Filho. -- 1. ed. -- São Paulo : Dinâmica Gráfica e Editora, 2012.

Outros autores: Cristine Campos de Xavier Pinto, Lycia Lima, Miguel Nathan Foguel, Ricardo Paes de Barros  
Bibliografia

1. Avaliação de programas de ação social  
2. Avaliação econômica 3. Cálculos 4. Investimento social 5. Retorno de investimentos I. Peixoto, Betânia. II. Pinto, Cristine Campos de Xavier. III. Lima, Lycia. IV. Foguel, Miguel Nathan. V. Barros, Ricardo Paes de. VI. Menezes Filho, Naercio.

12-13022

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

1ª Edição - Brasil  
2012

# Avaliação Econômica de projetos sociais

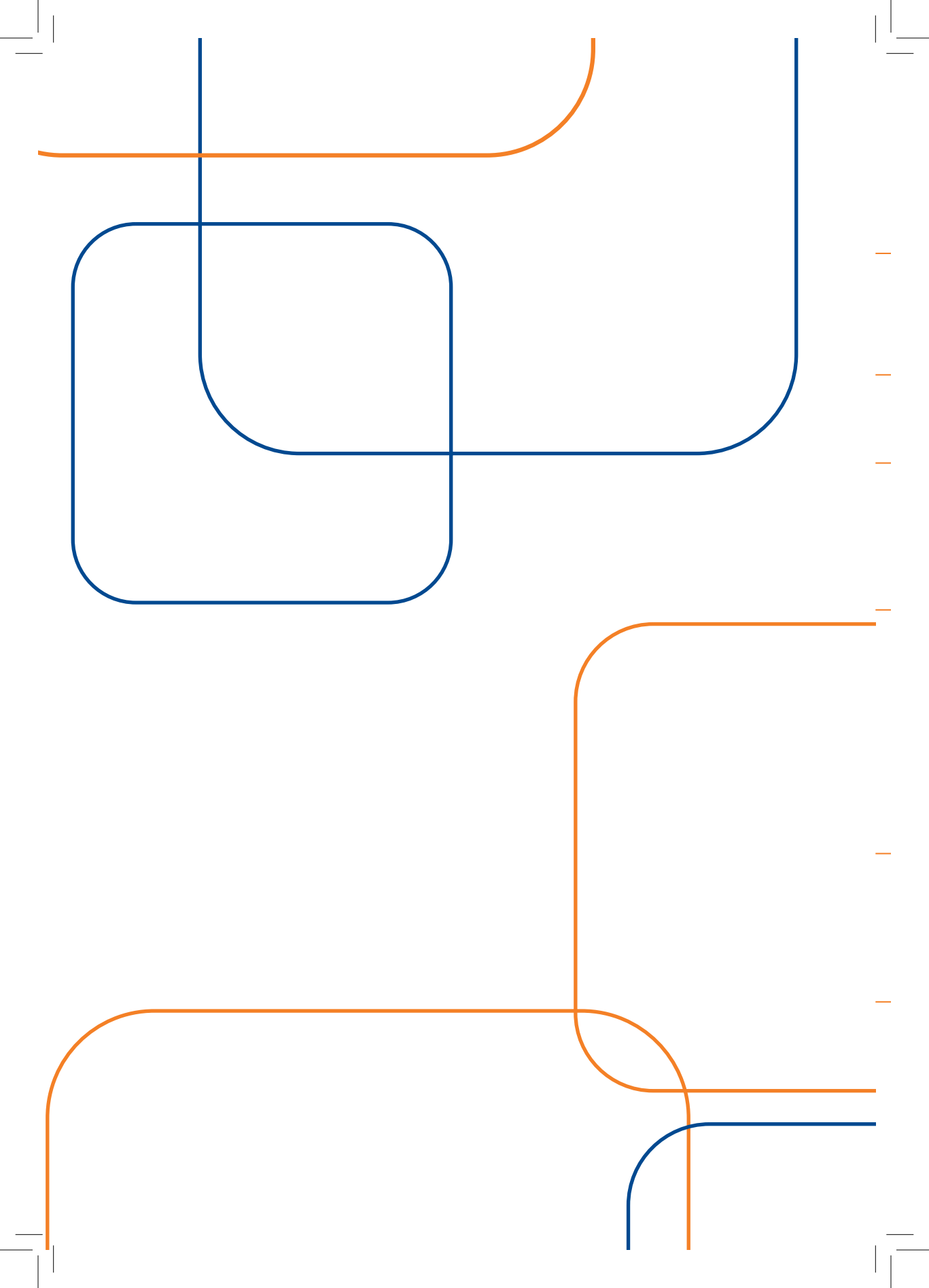
## ORGANIZADOR

Naercio Menezes Filho

## AUTORES

Betânia Peixoto  
Cristine Campos de Xavier Pinto  
Lycia Lima  
Miguel Nathan Foguel  
Ricardo Paes de Barros





# ÍNDICE

---

## Apresentação 07

Antonio Jacinto Matias  
Sérgio Ribeiro da Costa Werlang

---

## Prefácio 09

Naercio Menezes Filho

---

## Introdução 13

Ricardo Paes de Barros - Lycia Lima

1. Avaliação de Impacto de Programas Sociais:  
Por que, para que e quando fazer?

---

## Parte I - Métodos Básicos de Avaliação de Impacto 31

Miguel Nathan Foguel

2. Modelo de Resultados Potenciais \_\_\_\_\_ 35
3. Métodos de Aleatorização \_\_\_\_\_ 49
4. Diferenças em Diferenças \_\_\_\_\_ 69

Cristine Campos de Xavier Pinto

5. Pareamento \_\_\_\_\_ 85

---

## Parte II - Métodos Avançados de Avaliação de Impacto 107

Cristine Campos de Xavier Pinto

6. Variáveis Instrumentais \_\_\_\_\_ 107
7. Regressão Descontínua \_\_\_\_\_ 125

---

## Parte III - Análise de Retorno Econômico 147

Betânia Peixoto

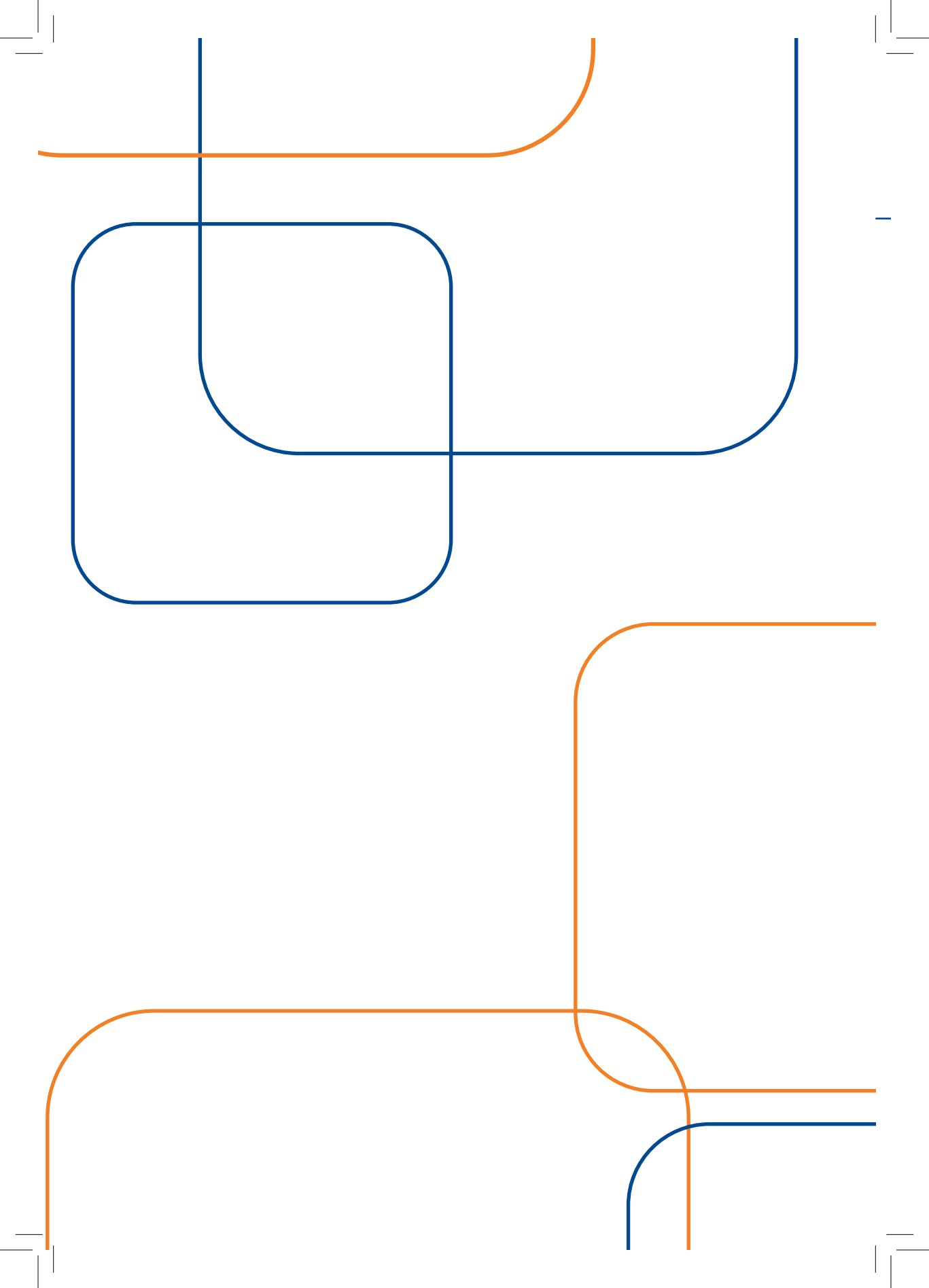
8. O Cálculo do Retorno Econômico \_\_\_\_\_ 149

---

Bibliografia \_\_\_\_\_ 175

Bibliografia Complementar \_\_\_\_\_ 183

Autores \_\_\_\_\_ 185



# APRESENTAÇÃO

O Itaú Unibanco, comprometido com o crescimento social e econômico do país, busca a sustentabilidade nos negócios, apostando em relações duradouras, fundamentadas nos seus valores de transparência, ética, respeito ao diálogo e capacidade de gestão de riscos. E, por entender que a educação é o principal caminho para o desenvolvimento sustentável do Brasil, este é o foco do investimento social do banco.

A Fundação Itaú Social, um dos braços do investimento social do Itaú Unibanco, tem como atividades centrais a formulação, a implantação e a disseminação de metodologias voltadas à melhoria de políticas públicas na área educacional. Sua atuação acontece em todo o território brasileiro, em parceria com governos, setor privado e organizações da sociedade civil.

Os valores estruturantes da cultura do Itaú Unibanco permeiam as ações da Fundação, compartilhando competências que garantem a execução das melhores práticas de gestão, na busca de resultados, nos ganhos de escala, no monitoramento e na avaliação das ações sociais apoiadas e desenvolvidas.

Neste contexto, a Fundação Itaú Social e o banco lançaram, em 2004, o Programa Avaliação Econômica de Projetos Sociais, colocando conhecimentos muito próprios da empresa a serviço da causa social. Com a convicção de que a avaliação é uma importante ferramenta de gestão, diversas ações foram desenvolvidas desde então, entre elas cursos e seminários ofertados em todo o país, com o objetivo de disseminar a cultura da avaliação.

Este livro complementa as estratégias de disseminação dos conceitos da avaliação econômica de projetos sociais ao tornar metodologias de impacto, hoje ainda ensinadas em poucos cursos de pós-graduação, acessíveis a alunos de graduação interessados no tema. Além disso, traz exemplos de avaliações de projetos brasileiros, mostrando que nos últimos anos a avaliação tem sido assumida, pouco a pouco, como um relevante instrumento de gestão e de prestação de contas à sociedade.

Colocamos foco também na análise de retorno econômico, pois, considerando-se que a escassez de recursos para investimentos voltados para o desenvolvimento social é uma realidade constante, é importantíssimo garantir que tais investimentos estejam alcançando seu melhor resultado.

Agradecemos ao coordenador e aos autores deste livro, em especial ao Ricardo Paes de Barros, por sua dedicação e empenho para a realização deste projeto, convencidos, como nós, da importância da disseminação deste conhecimento.

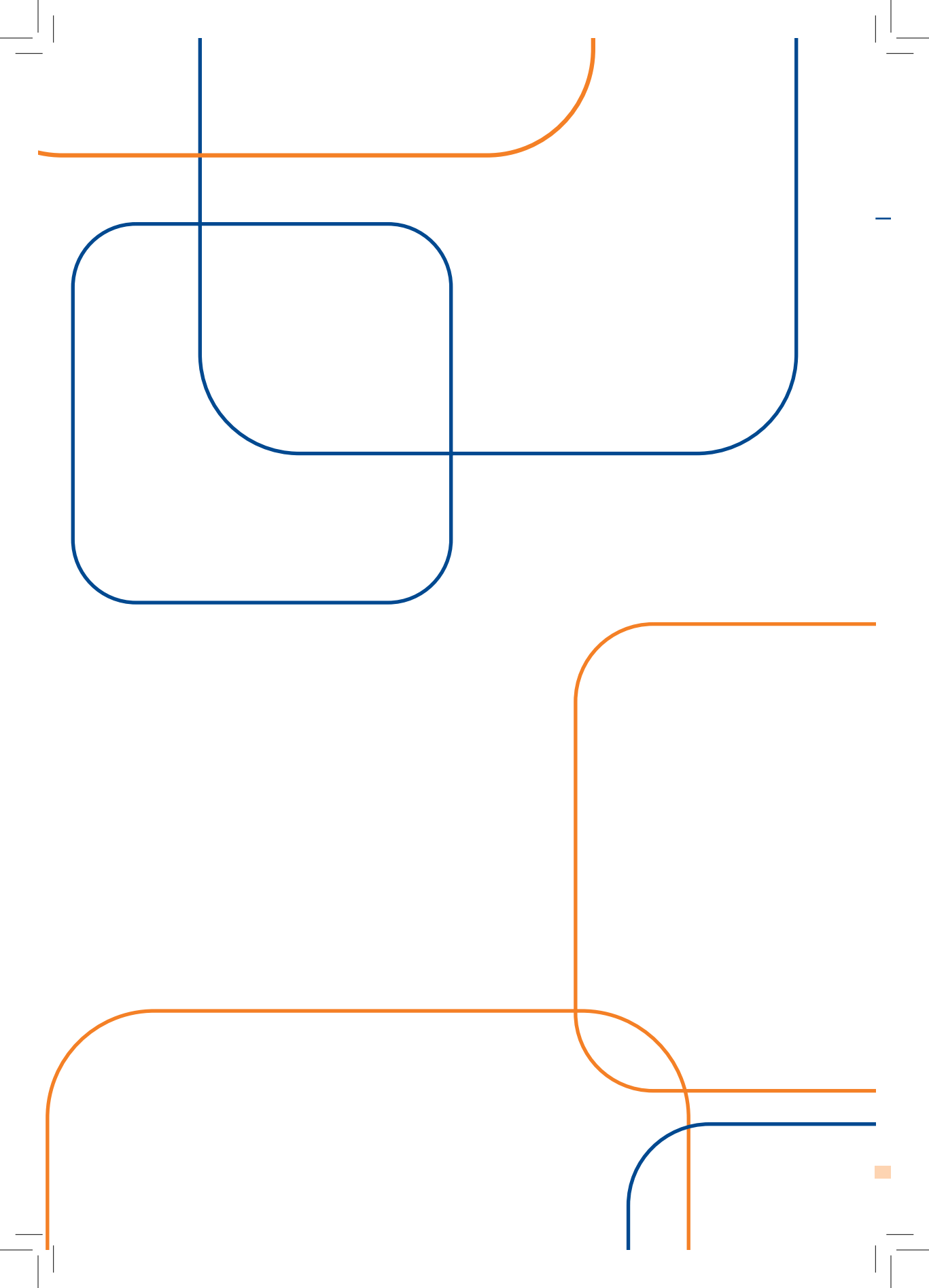
Esperamos, com mais este esforço de difundir a prática da avaliação econômica em projetos sociais, estar, de fato, contribuindo para a maior eficiência e alcance das ações sociais no nosso país.

**Antonio Jacinto Matias**

Vice-presidente da Fundação Itaú Social

**Sérgio Ribeiro da Costa Werlang**

Vice-presidente do Itaú Unibanco





## PREFÁCIO

---

Há meros 10 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 da Fundação Itaú Social na área de avaliação de impacto e retorno econômico iniciou-se em 2004, quando o programa “Raízes e Asas” foi avaliado economicamente.

Desde então, a Fundação Itaú Social oferece 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.

A preocupação com avaliação tem ganhado força na medida em 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 nas vidas das pessoas. Da mesma forma, o trabalho desenvolvido no terceiro setor, principalmente na área social, precisa prestar contas à sociedade e para os 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 das avaliações quantitativas existentes não chegam a essa parte, concluindo 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 desse livro estão entre os maiores especialistas em avaliação do país. No primeiro capítulo, à guisa de introdução, Ricardo Paes de Barros e Lycia Lima discutem por que a avaliação de impacto é tão necessária. Afinal, por que deveríamos gastar recursos avaliando um programa que, se não apresentasse resultados satisfatórios, não seria demandado por ninguém? 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 segundo capítulo, 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, através de um 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 quarto capítulo, Miguel explica a metodologia de diferenças em diferenças, que deve 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 termos de 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 de fatores específicos.

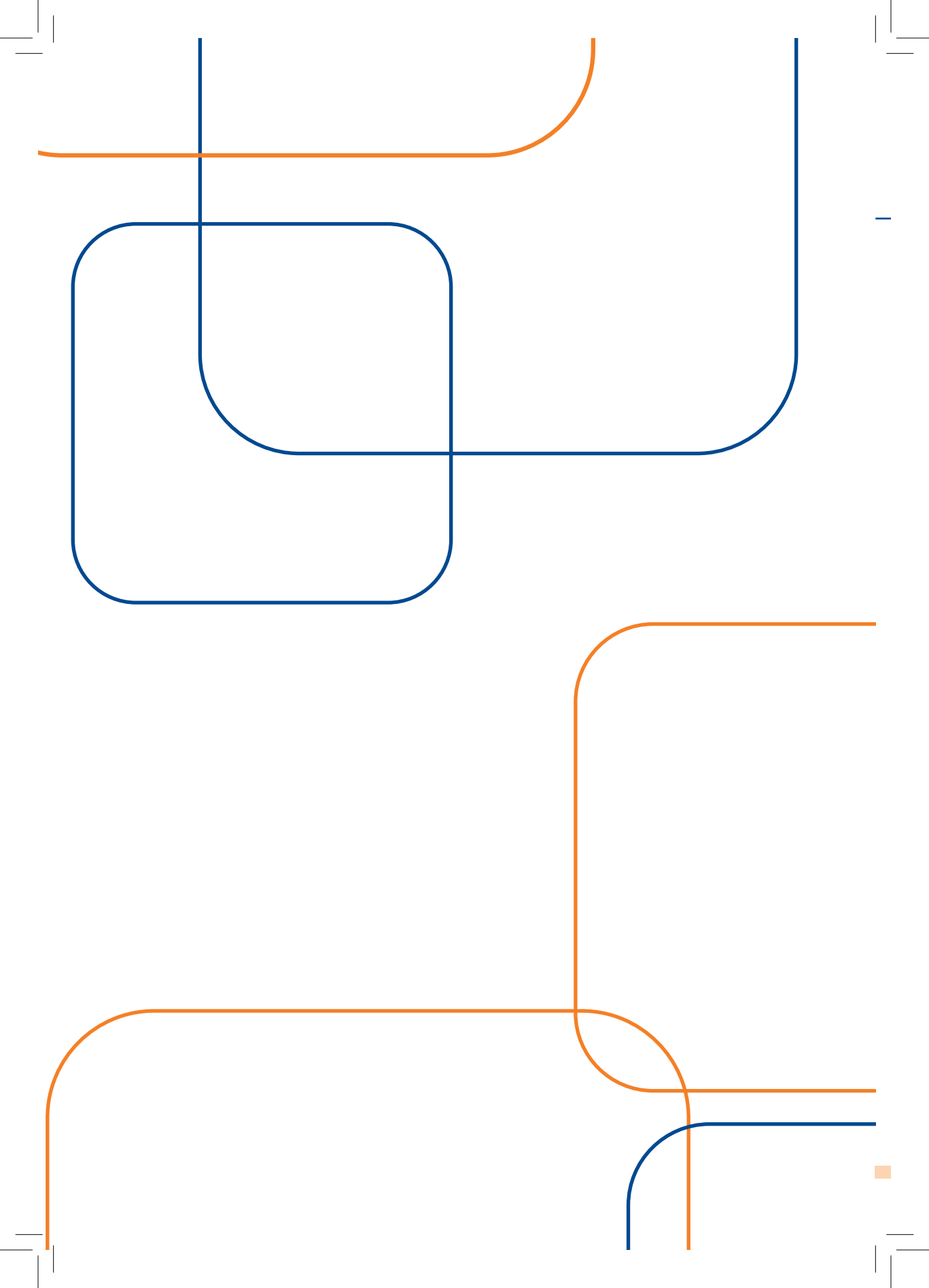
A terceira parte do livro lida com o cálculo do retorno econômico de um projeto social. No capítulo 8, Betânia Peixoto explica detalhadamente os conceitos e o instrumental necessários para o cálculo desse retorno, incluindo aplicações práticas e um exemplo para que os conceitos sejam sedimentados. Conceitos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

Este livro foi planejado tendo em vista um 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.

O apêndice do livro comenta uma bibliografia complementar nas áreas de estatística e matemática financeira que pode ser utilizada como auxílio para entendimento dos capítulos do livro. Além disso, cada capítulo traz exercícios específicos que servem para ajudar a fixação do conteúdo. Finalmente, a parte final traz um exercício prático para ser resolvido pelos leitores. As bases de dados e as respostas dos exercícios estão disponíveis na página da Fundação Itaú Social na internet.

Esse livro deve estimular ainda mais a realização de avaliações de impacto, para que possamos conhecer e replicar os projetos sociais e 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

## Introdução

A decisão dos formuladores de políticas públicas de investir em um programa social sempre parte do pressuposto de que aquela intervenção terá um impacto positivo sobre um conjunto de resultados de interesse coletivo ou individual que, em última instância, deve trazer benefícios para ao menos uma parcela da sociedade. Mas será que, na prática, as políticas públicas e os projetos sociais do terceiro setor atingem os objetivos para os quais foram desenhados?

O principal propósito de uma avaliação de impacto é verificar se, na realidade, um determinado programa está alcançando os objetivos, ou impactos, esperados. Entenderemos por impacto as diferenças entre a situação dos participantes do projeto após terem participado e a situação em que estariam, caso não tivessem tido acesso a ele. Dessa forma, o impacto do programa é definido como o contraste entre duas situações: uma real (a situação dos participantes após a participação no projeto) e outra hipotética (a situação em que estariam caso não tivessem tido a oportunidade de participar do programa). Esta definição de impacto, entretanto, é apenas um ponto de partida. Os demais capítulos deste livro se encarregarão de aprofundar essa discussão e apresentar os diferentes métodos existentes para a estimação do impacto.

Avaliações de impacto certamente consomem recursos que, alternativamente, poderiam ser investidos diretamente no próprio programa. A crescente expan-

são de estudos dessa natureza sinaliza que a importância de dedicar esforços e recursos para este fim é cada vez mais reconhecida. Mas por que avaliações de impacto são importantes?

Os resultados dessas avaliações têm uma variedade de utilidades potenciais. A mais óbvia delas é utilizar as informações produzidas para a melhoria do desenho do próprio programa em questão. Além disso, os resultados produzidos podem ser usados para auxiliar no desenho de outros programas subsequentes na mesma comunidade ou até mesmo para informar formuladores de políticas que pretendam desenhar programas com objetivos similares em outras comunidades.

Este livro busca descrever a metodologia necessária para a realização de avaliações de impacto. Entretanto, antes de iniciar a discussão metodológica, é fundamental entender por que, para que e quando avaliar programas sociais.

## 2

# Por que avaliar o impacto de um programa social?

Nossa sociedade investe uma grande quantidade de recursos, tanto humanos quanto financeiros, em programas sociais. Avaliações de impacto nos permitem verificar se estes recursos estão de fato sendo aplicados da melhor maneira possível. No entanto, avaliações de impacto também consomem recursos. Será que a realização da avaliação de impacto do programa é sempre o melhor uso daqueles recursos ou seria melhor revertê-los em um maior investimento no próprio programa? Afinal, os benefícios de uma avaliação de impacto superam seus custos?

## 2.1 Não é evidente que um programa amplamente utilizado tem impacto?

A justificativa da importância da aplicação de recursos para avaliar impacto não é tão óbvia. Uma contra-argumentação válida poderia ser fundamentada

na alegação de que é 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.

Em sua maioria, os programas sociais se baseiam em teorias que preveem um elo entre o acesso ao programa e impactos positivos sobre um conjunto pré-determinado de resultados. Portanto, os formuladores de políticas públicas, geralmente, contam com argumentos teóricos e, frequentemente, com evidência empírica, ambos sinalizando 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 pergunta.

Uma das motivações de uma avaliação de impacto é, precisamente, a validação das teorias que serviram de base ao desenho do programa. Por este motivo, é tão comum e importante a avaliação 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 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, podemos argumentar que o objetivo de uma avaliação de impacto vai muito além de simplesmente determinar a existência ou não do impacto. Mesmo que tenhamos convicção da existência de impacto de um dado programa, ainda assim será necessário conduzir uma avaliação para que possamos “entender” o impacto. Tanto para o gestor do programa, quanto para os 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 que canais permitem que este impacto se manifeste. Este conhecimento é indispensável para o aperfeiçoamento do desenho do programa e para a melhoria na sua adequação às necessidades de seus usuários.

## 2.2 Demanda pelos mais ricos é evidência de impacto?

Muitos dos programas sociais existentes são ofertados gratuitamente para a parcela mais pobre da sociedade. Nesse caso, adesão voluntária ao programa deve ser necessariamente interpretada como indício do impacto do programa? Pode-se argumentar que não. É possível que os pobres tenham conhecimento limitado sobre a eficácia daquele serviço ou programa. Além disso, como a oferta é gratuita, os beneficiários aderem porque não perdem nada participando.

Contudo, o que dizer do impacto de um serviço oferecido gratuitamente aos mais pobres, mas para o qual existe provisão privada para suprir a demanda dos mais ricos, que pagam preços significativos pelo acesso? 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 teria 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 a necessidade de realizar uma avaliação de impacto nesta situação?

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

Em segundo lugar, mesmo quando não há dúvidas sobre a existência do impacto do serviço, pode ser importante avaliar a sua magnitude sobre os mais pobres. Sabemos 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, a evidência da eficácia do programa entre os mais ricos pode não ser suficiente para justificar a implantação de um programa análogo para os mais pobres. Nesses casos, justifica-se, pois, a utilização da avaliação de impacto para produzir evidência direta do impacto do programa e do valor desses benefícios para o grupo dos mais pobres.



## 2.3 A relação entre impacto, ambiente socioeconômico e características da população beneficiada

Formuladores de políticas públicas, geralmente, desenham um programa social visando a uma população-alvo específica sendo beneficiada em um ambiente pré-estabelecido. Assim, é natural que existam poucas dúvidas sobre o impacto de um determinado programa quando este é implementado no local e momento corretos e beneficia a população para a qual foi originalmente desenhado. Porém, este mesmo programa implementado em um ambiente completamente distinto do planejado provavelmente não terá o mesmo resultado. É de se esperar que a mesma ação dirigida a outra população, outro local e outro momento do tempo não beneficie da mesma forma os que dela participam.

Uma das utilidades da avaliação de impacto é, precisamente, determinar em que medida a eficácia de um programa depende das características da população beneficiada, do momento no tempo e da natureza do ambiente socioeconômico em que ocorre.

## 2.4. Impacto potencial versus impacto efetivo

O impacto de um programa não depende apenas do seu desenho e da sua adequação ao perfil dos beneficiários e ao ambiente socioeconômico em que estes vivem. Igualmente determinante da magnitude do impacto é a forma como o programa é implementado. Mesmo sendo poucas as dúvidas sobre o impacto potencial de um programa bem desenhado e focalizado, podem persistir grandes incertezas relativas ao efetivo impacto desse programa quando implementado de maneira deficiente. Neste 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 maneira imperfeita como foi implementado.

Em suma, o impacto de um programa depende não apenas do seu impacto potencial, mas está também intrinsecamente ligado às condições da sua implementação. Ainda que um programa tenha comprovadamente impacto potencial, é fundamental avaliar o seu impacto real. Nesse caso, a avaliação de impacto é útil para verificar qual parcela do impacto potencial foi desperdiçada

devido a falhas no processo de implementação.

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

## 2.5 Avaliação de impacto como forma de monitoramento

Os impactos de um determinado programa social dificilmente são invariáveis ao longo do tempo. Assim, nada garante que um programa avaliado hoje continue tendo o mesmo impacto no próximo ano ou no ano subsequente.

Podemos apontar pelo menos quatro razões para justificar essa potencial inconstância temporal do impacto. Primeiramente, o impacto de um programa, normalmente, está relacionado com o 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 se for sensível a mudanças no ambiente socioeconômico. Por fim, se o perfil da população beneficiada pelo programa se modifica ao longo do tempo, é de se esperar que a magnitude do impacto também se altere.

Devido a cada um desses motivos, mesmo programas que têm impacto inicial 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 no caso em que o perfil da população beneficiada e o ambiente socioeconômico permanecem estáveis, avaliações de impacto periódicas podem ser de extre-

ma utilidade como um instrumento de gestão. Neste caso, flutuações na magnitude do impacto podem indicar uma melhoria ou deterioração na forma de funcionamento e, portanto, na gestão do programa.

Os argumentos apresentados sinalizam que o impacto de um programa tem caráter mutável ao longo do tempo. Portanto, uma avaliação de impacto deve ser considerada como uma ferramenta para a utilização contínua, já que as conclusões auferidas a partir de sua utilização pontual não são necessariamente válidas intertemporalmente.

## 2.6 A questão da relação custo-efetividade

A existência de impacto não é suficiente para justificar a alocação de recursos adicionais para um dado programa. Em uma sociedade, se existirem programas alternativos que perseguem os mesmos objetivos, a opção por um determinado programa dependerá de este ser aquele que produz o maior impacto por unidade de custo. Esse critério é comumente denominado custo-efetividade.

Assim, em um cenário onde os recursos são escassos e diferentes programas competem pelos mesmos recursos, a avaliação de impacto é essencial para mensurar a magnitude do impacto e calcular a relação custo-efetividade de cada programa. 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.

## 2.7 A questão da relação custo-benefício

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. No cálculo desta relação, deve-se considerar todos os custos e benefícios sociais, tanto aqueles sobre os agentes diretamente envolvidos como as externalidades sobre o restante da sociedade. Mesmo diante da evidência de que um programa tem impacto, é essencial ainda assim estimar a magnitude

deste impacto para que se possa verificar se o valor dos benefícios produzidos superam os custos demandados. Muitas vezes, diferentes ações com objetivos análogos não diferem substancialmente com relação ao seu custo. Nestes casos, são as diferenças na magnitude do impacto que irão determinar qual o programa que tem a melhor relação custo-benefício.

Em princípio, é possível argumentar que todos os programas com relação custo-benefício favorável deveriam ser implementados. Esta 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, então 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 custo-benefício são estimadas supondo que estes não foram nem serão implementados. Neste 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 custo-benefício são recalculadas para os programas não implantados, levando em consideração que aquele de melhor relação custo-benefício foi efetivamente implantado. Após esta reavaliação, caso ainda exista algum programa com relação custo-benefício favorável, o melhor deles deve ser implantado e o procedimento novamente conduzido. Note que esta análise sequencial requer a estimação do impacto de cada programa tanto na ausência quanto na presença dos demais.

## 2.8 Por que não investigar diretamente a propensão a pagar?

A seção anterior demonstrou que, em última instância, o objetivo de uma avaliação de impacto deve ser investigar a relação custo-benefício de um programa. 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 este pelo conjunto de beneficiários.

Nesse sentido, uma avaliação de impacto apresenta duas limitações. Primeiramente, um programa tem geralmente uma variedade de impactos, alguns intencionais e muitos outros colaterais. Neste 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 men-

suração. 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. De fato, 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 em termos monetários ou em termos do acesso a 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 um determinado programa, benefício ou serviço. O desafio neste caso é que frequentemente a resposta pode não ser fidedigna. Como a pergunta se refere a uma situação hipotética, as pessoas tendem a superestimar o valor que elas efetivamente estariam dispostas a pagar. Existem, entretanto, protocolos pré-estabelecidos que definem como investigações desta natureza podem ser conduzidas visando 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 efetivo comportamento 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 este 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 primeiramente decompor o impacto nos seus diversos componentes e

posteriormente valorar cada um destes individualmente. Sendo assim, dado que a avaliação de impacto se apresenta como um instrumento mais complexo e que via de regra exige maior esforço, qual seria então a justificativa para a sua utilização?

Existem essencialmente três justificativas que sustentam o uso da avaliação de impacto. 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 em questão 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 em questão tem impactos sobre não beneficiários. Neste 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.

Por fim, o uso da propensão a pagar tem a grande limitação de não permitir 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, na medida em que é impossível distinguir qual componente do programa está contribuindo mais ou menos para a satisfação do usuário.

Do ponto de vista dos beneficiários, é possível que seja suficiente saber o grau de satisfação dos beneficiários com o programa de um modo geral. Porém, para os gestores e aqueles que financiam o programa, essa informação pode ser insuficiente.

Existem duas razões que justificam a necessidade de 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 no alcance dos 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.

## 2.9 Heterogeneidade do programa e da população-alvo

A grande maioria das intervenções é heterogênea e oferecida a uma clientela também heterogênea. Na maioria dos casos, é possível ajustar os diversos parâmetros do programa, desde sua duração e intensidade até seu conteúdo e qualidade. É de se esperar, portanto, que a magnitude do impacto varie com estes parâmetros e, também, segundo as características do beneficiário e o contexto socioeconômico em que o programa se insere. Assim, a realização de uma avaliação de impacto é essencial para entender como a heterogeneidade na intervenção e no perfil dos usuários se reflete nos resultados alcançados.

Um mesmo programa pode ter muitos desenhos que variam em termos da sua eficiência. Uma avaliação de impacto possibilita não apenas identificar o desenho mais eficiente, mas também identificar os segmentos da população-alvo para os quais os benefícios são maiores. A identificação destes grupos é fundamental para o desenho do programa. No curto prazo, a atuação do programa deveria se concentrar nos segmentos da população-alvo que têm a maior probabilidade de se beneficiar destas ações. Por outro lado, a identificação dos segmentos que pouco se beneficiam das ações é igualmente importante na medida em que revela a necessidade do investimento em ações alternativas voltadas para estes grupos.

## 3. Para que avaliar o 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. 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.

### 3.1 Sobre o uso interno: o aperfeiçoamento do desenho, funcionamento e gestão do programa avaliado

Em seu uso interno, a avaliação de impacto serve como insumo para a tomada de decisão sobre o próprio programa. Neste sentido, a mesma pode ser utilizada com vários objetivos, dependendo 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 veredicto. Enquanto resultados positivos justificam a permanência ou ampliação de um programa, a ausência destes dá suporte à desativação progressiva ou imediata do programa. Em resumo, em primeiro lugar, uma avaliação de impacto serve para (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 com os parâmetros que definem a intervenção é a única maneira através da qual uma avaliação pode contribuir para a reformulação dos princípios que embasam o programa, possibilitando assim o aperfeiçoamento do seu desenho e a adequação do seu marco lógico.

O impacto de um programa, entretanto, não é determinado apenas por seu desenho. A natureza dos beneficiários e a forma e o grau com que estes utilizam as ações do programa também influenciam a magnitude do impacto. Assim, na medida em que a avaliação revela como a magnitude do impacto do programa varia com a forma e grau de utilização dos serviços e tipo de beneficiário, a mesma se torna um instrumento de extrema utilidade ao gestor responsável pela operação do programa. A partir dessas informações, o gestor é capaz de realizar ajustes finos no programa, podendo, potencialmente, readequar a natureza dos serviços e bens oferecidos às necessidades dos beneficiários e assim promover uma maior efetividade do programa em questão.

Por fim, deve-se ressaltar ainda que o impacto do programa também depende da forma como a gestão e operação do programa é conduzida. Assim, uma avaliação de impacto que contempla 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.



## 3.2 Sobre o uso externo: para o desenho e aperfeiçoamento de programas similares

O uso externo de uma avaliação de impacto está relacionado com a sua utilidade para auxiliar no desenho, na implementação ou na gestão de outros programas similares em contextos distintos. Neste 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 que são aquelas com as melhores relações custo-efetividade. A identificação destas 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, avaliações de impacto são fundamentais para o aprimoramento da eficácia dos programas sociais em vigor.

A avaliação dos impactos de um programa, quando associada à avaliação de seu custo, permite compará-lo com as alternativas disponíveis em termos de suas relações 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 estas investiguem não apenas o tamanho do impacto, mas também a interação deste com características do ambiente socioeconômico e cultural em que as avaliações se inserem. Este 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. Esta 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 um alerta para que a expansão do programa seja feita de forma cuidadosa, adequando o mesmo às características específicas do novo contexto. Nestes 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 esta 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 os ajustes necessários no seu desenho para que o seu impacto seja invariante ao contexto. Para que estes 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. Este 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.

## 4. Quando avaliar o impacto de um programa social? Avaliações ex-ante versus ex-post

Existem essencialmente três momentos na execução de um programa em que avaliações de impacto podem ocorrer: (i) antes do início do programa (são as ditas avaliações ex-ante), (ii) durante o período de execução do programa (são as ditas avaliações ex-post de percurso) e (iii) após a conclusão do programa (são as ditas avaliações ex-post de encerramento). Em cada caso, a avaliação é caracterizada por metodologias e objetivos distintos, condicionados à disponibilidade de informações relacionadas ao programa naquele momento do tempo.

### 4.1 Sobre a necessidade de avaliações ex-ante e ex-post

A decisão pela implementação de um dado programa é geralmente precedida pela apresentação de argumentos que justificam a racionalidade da opção por

aquela determinada intervenção. Em geral, justificativas deste tipo se baseiam em estimativas de custos e impactos esperados do programa em questão. Prever a magnitude dos impactos antes que a intervenção seja implantada é o objetivo das avaliações ex-ante de impacto. Embora o objetivo primordial de uma avaliação ex-ante seja justificar a adoção de um dado programa, é possível e recomendável que avaliações ex-ante sejam também utilizadas para decidir sobre o melhor desenho da intervenção.

Uma vez 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, são utilizadas para verificar a validade das previsões realizadas pelas avaliações ex-ante. Desta forma, esse tipo de avaliação mostra-se fundamental para traçar recomendações sobre o futuro do programa, seja no sentido de interromper, continuar ou aperfeiçoar as ações da intervenção em curso. Este tipo de avaliação é muito comum durante a fase piloto da implementação de um programa, embora seja importante que a sua aplicação não se limite a esta 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 quanto antes para que tenha sua influência sobre o desenho, operação, continuidade e expansão do programa maximizada. Por outro lado, dado que alguns impactos podem levar tempo para se manifestar, quanto mais cedo realizada a avaliação de impacto, maior a probabilidade de se subestimar ou até ignorar impactos importantes que não se manifestaram no curto prazo. Esse problema é particularmente grave já que, frequentemente, os impactos de maior importância podem ser os de longo prazo. Assim, toda avaliação ex-post de percurso precisa analisar cuidadosamente o trade-off entre fornecer respostas rápidas para auxiliar no aperfeiçoamento e expansão do programa e aguardar o tempo necessário para que parte substancial dos impactos do programa se manifeste. Portanto, avaliações ex-post de percurso são mais comuns em programas de longa duração ou mesmo de duração ilimitada, sendo difíceis de serem realizadas em programas de curta duração.

A realização de uma avaliação de impacto também é justificada 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 de uma avaliação ex-post das decisões realizadas no passado e da qualidade da gestão do programa. Não permite rever ou aprimorar ações, mas apenas verificar se os agentes envolvidos no processo decisório tomaram ou não decisões adequadas e se os responsáveis pelo programa fizeram uma boa gestão.

Por outro lado, uma utilidade extremamente importante das avaliações ex-post

em geral está relacionada à identificação de melhores práticas. Avaliações ex-post de programas com objetivos similares geralmente fornecem informações relevantes que podem auxiliar na tomada de decisão de gestores públicos empenhados em desenhar novos programas ou aperfeiçoar os programas existentes na área em questão.

Por fim, vale ressaltar que as avaliações ex-post de encerramento têm a grande vantagem de evitar o conflito entre o momento da avaliação e o prazo de maturação do impacto do programa. Neste caso, 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.

## 4.2 Diferenças informacionais e metodológicas

Em decorrência do momento em que são realizadas, as avaliações ex-ante e ex-post se baseiam cada qual em um conjunto distinto de informações, aquelas disponíveis naquele dado momento do tempo. Avaliações ex-post podem contar com informações sobre a situação de beneficiários e não beneficiários antes e em vários momentos após o início da intervenção. Por outro lado, avaliações ex-ante são conduzidas às vezes sem informações sobre quem serão os efetivos beneficiários do programa. Na medida em que informações naturalmente se acumulam ao longo do tempo, avaliações ex-post têm grande vantagem informacional sobre suas contrapartidas ex-ante.

Todo trabalho avaliativo é sempre o resultado da combinação de teoria com evidência empírica. No entanto, 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. Em uma avaliação ex-post cuidadosa, na qual uma grande quantidade de informações pertinentes e fidedignas foi coletada, o papel da teoria tenderá naturalmente a ser dominado pelo da evidência empírica. Já numa avaliação ex-ante, o papel da teoria passa a ser primordial, uma vez que as estimativas de impacto são obtidas a partir de simulações com dados na maioria das vezes secundários sobre o comportamento hipotético dos futuros beneficiários do programa a ser implantado. Neste caso, é com base na teoria sobre o comportamento destes beneficiários que são derivadas as hipóteses a partir das quais estas simulações de impacto são realizadas.

## 5. Considerações finais

Na nossa sociedade, formuladores de políticas continuamente propõem uma grande variedade de soluções visando atender às mais diversas necessidades

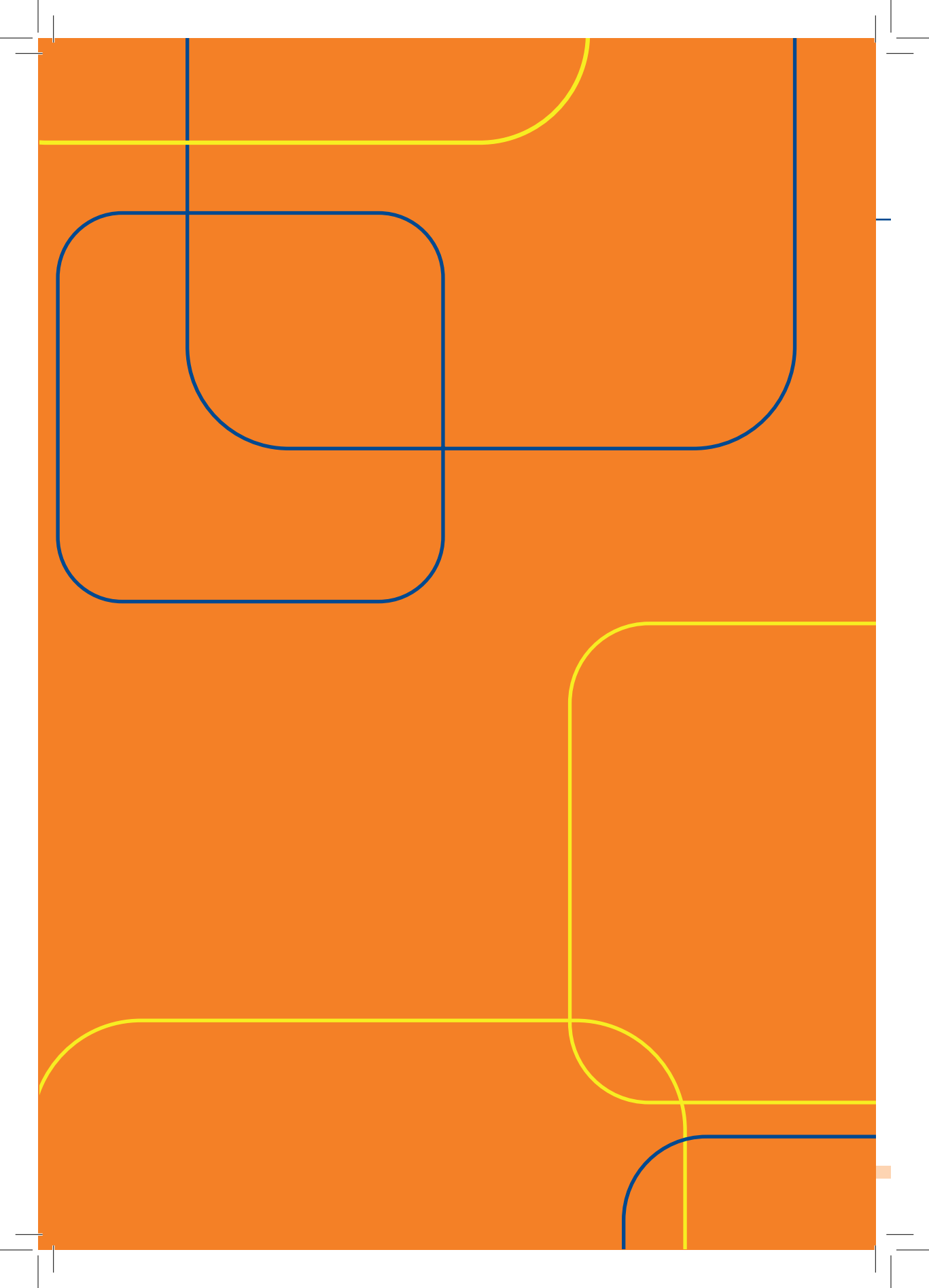
da população. Entretanto, dada a escassez dos recursos disponíveis, nem todas estas soluções podem ser efetivamente implementadas. É necessário, portanto, escolher bem. Quais das propostas apresentadas abrangem a maior parte das necessidades da população ao menor custo? Em outras palavras, quais destas propostas representam o melhor uso do orçamento disponível?

Um dos instrumentos fundamentais para se obter a resposta a estas questões é a avaliação de impacto, que permite isolar a contribuição de uma ação específica no 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.

## Exercícios

- 1) Quais são as possíveis motivações para uma avaliação de impacto?
- 2) O impacto de um programa não depende apenas do seu desenho e da sua adequação ao perfil dos beneficiários e ao ambiente socioeconômico em que estes vivem. Do que mais depende o impacto do programa?
- 3) Explique a importância da avaliação de impacto em um cenário onde os recursos são escassos e diferentes programas competem pelos mesmos recursos.
- 4) Existem essencialmente três momentos na execução de um programa em que avaliações de impacto podem ocorrer. Explique cada caso.
- 5) Explique os conceitos de uso interno e uso externo de uma avaliação.



## PARTE 1

# Métodos Básicos de Avaliação de Impacto

A parte I do livro é dedicada a apresentar um conjunto de métodos que são amplamente empregados para isolar e medir o impacto de programas sociais. Esses métodos são comumente divididos em duas categorias: o método experimental e os métodos não experimentais. O primeiro, a ser apresentado no capítulo 3, é baseado na seleção aleatória dos indivíduos que farão parte do grupo que receberá o programa (grupo de tratamento) e do grupo que não o receberá (grupo de controle). Como veremos, esse método faz com que a única diferença entre os grupos seja a participação no programa, uma vez que a aleatorização garante que eles sejam muito semelhantes tanto em termos das características observadas como das não observadas pelo analista. Pelo fato de conseguir fazer com que a intervenção seja a única diferença entre os grupos de tratamento e controle, esse método tem a denominação de “padrão-ouro” na área de avaliação de impacto.

Os demais capítulos apresentarão em detalhes os métodos não experimentais. Como não são baseados na seleção aleatória dos grupos de tratamento e controle, esses métodos não asseguram que os dois grupos difiram entre si apenas pela participação no programa. Na realidade, o que eles fazem é substituir a aleatorização do método experimental por certas condições e hipóteses que visam a tornar o grupo de controle parecido com o grupo de tratamento. Como veremos, cada método adota um conjunto específico de hipóteses para identificar o efeito causal do programa. Essas hipóteses não são diretamente testáveis, mas algumas podem ser confirmadas de modo indireto.

A escolha do método a ser empregado depende em larga medida do programa a ser avaliado, dos arranjos institucionais existentes, dos custos envolvidos na avaliação, além da disponibilidade de dados. O aumento da importância da avaliação de impacto e da análise de custo-benefício entre financiadores, gestores e fiscalizadores dos programas tem sido acompanhado pelo desenvolvimento e aperfeiçoamento de um conjunto de mé-

todos que formam hoje uma ampla “caixa de ferramentas” utilizada na área de avaliação de impacto. Ao apresentar uma introdução a alguns dos principais métodos disponíveis nessa caixa de ferramentas, esta parte do livro pretende auxiliar não somente no entendimento das metodologias, mas também no processo de escolha da metodologia mais adequada a ser adotada no contexto específico no qual se insere o programa. A esse propósito, vale assinalar que o foco do livro é sobre as metodologias ex-post. Em contraste com as avaliações ex-ante, que são baseadas em modelos sobre o comportamento dos agentes, as metodologias ex-post são baseadas em dados específicos colhidos para a avaliação do programa (ver Box 1 e seção 4 do capítulo 1).

Antes de entrarmos no detalhamento de cada método de nossa caixa básica de ferramentas, apresentaremos no capítulo 2 um arcabouço conceitual amplamente utilizado na área de avaliação de impacto. Além de estabelecer a notação que empregaremos em outros capítulos desta parte do livro, sua principal finalidade é apresentar um dos principais problemas da área de avaliação de impacto, a saber, o viés de seleção.



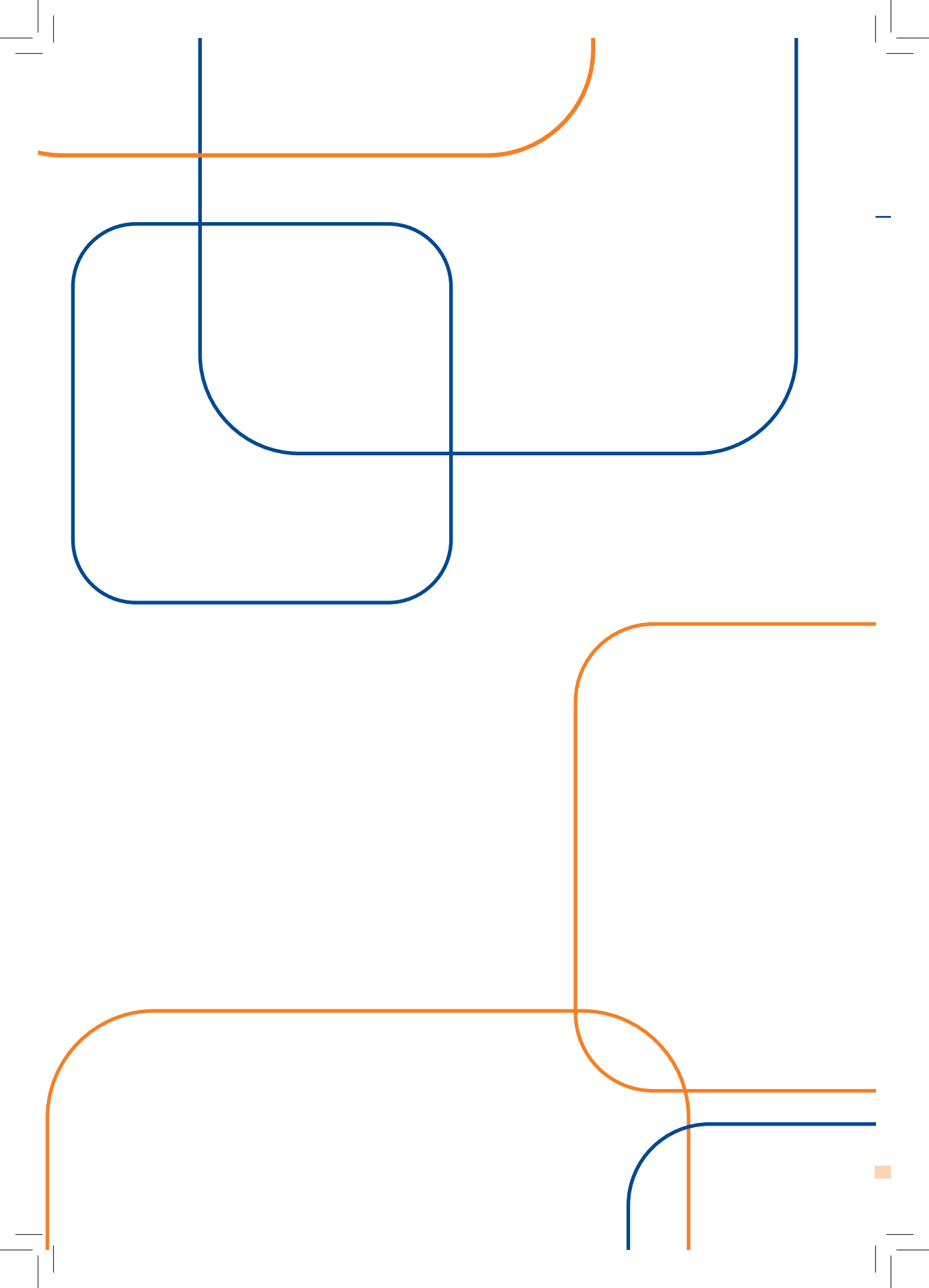
## Box 1: Avaliações Ex-ante e Ex-post

As avaliações ex-ante são elaboradas com base na construção de um modelo sobre o comportamento dos agentes e do ambiente econômico no qual eles e o programa em avaliação estão inseridos. Em alguns casos, o modelo proposto é estrutural, ou seja, baseado em hipóteses acerca das formas funcionais que representam o comportamento econômico dos indivíduos e do funcionamento da economia ao longo do tempo (por exemplo, Heckman et al., 1998). Em outros casos, o modelo proposto é baseado em formas reduzidas, isto é, em equações que procuram refletir as relações diretas entre as decisões dos agentes, os indicadores de resultado e demais variáveis consideradas relevantes na estimação do impacto do programa (ver, por exemplo, Bourguignon e Ferreira, 2003).

Tipicamente, as avaliações ex-ante utilizam dados de fontes secundárias ao programa para estimar os parâmetros do modelo proposto, incluindo o próprio impacto do programa. Após a estimação do modelo, é possível alterar os valores de alguns dos seus parâmetros e, por meio de simulações, avaliar como as variáveis de interesse mudam. Esse tipo de exercício é bastante útil para se obter um balizamento de como os impactos do programa podem variar para diferentes hipóteses sobre o comportamento dos agentes e o funcionamento da economia. Por outro lado, o método ex-ante tipicamente requer a adoção de um conjunto de hipóteses não testáveis, tanto sobre o comportamento dos indivíduos quanto sobre a identificação dos parâmetros do modelo. Distintamente, as avaliações de impacto ex-post não são baseadas em mode-

los teóricos, mas sim em métodos empíricos que procuram estimar diretamente dos dados os impactos do programa. Em contraste com o método ex-ante, que geralmente é empregado antes da implementação da intervenção, as avaliações ex-post só podem ocorrer durante ou após a execução do programa. Ademais, a maior parte das avaliações ex-post são “model-free”, isto é, preferem deixar os próprios dados “falarem” ao invés de proporem modelos teóricos para explicar os resultados encontrados. Embora extremamente úteis para se conhecer a efetividade das intervenções, as avaliações ex-post normalmente só estimam os impactos específicos dos programas. Nesse sentido, elas não permitem construir diferentes cenários para o programa e, portanto, não oferecem uma ideia das variações no impacto para diferentes desenhos do programa.

Mas as abordagens ex-ante e ex-post não são mutuamente excludentes. Recentemente, Todd e Wolpin (2006) combinaram os dois tipos de métodos para estudar o impacto do programa mexicano de transferência condicional de renda Progresa (atualmente chamado de Oportunidades). Basicamente, o exercício dos autores consistiu em prever o impacto do programa com base na abordagem ex-ante e posteriormente compará-lo com o impacto obtido pelo método ex-post de aleatorização (ver capítulo 3). A variável de resultado foi a frequência à escola de crianças entre 12 e 15. Os impactos obtidos pelas duas abordagens foram bastante semelhantes, o que permitiu a realização de cenários mais confiáveis a partir de variações nos parâmetros do modelo de base.



## CAPÍTULO 2

# Modelo de resultados potenciais

Miguel Nathan 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 está sempre em uma de duas situações distintas: ter sido ou não ter sido tratado pelo programa. Idealmente, o melhor grupo de comparação para os indivíduos tratados seria formado pelos mesmos indivíduos na situação em que eles não fossem tratados. 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, como não é possível observar essa situação contrafactual, 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 autosseleção. 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 um 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 um 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 de um curso ministrado por pessoal especializado 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.

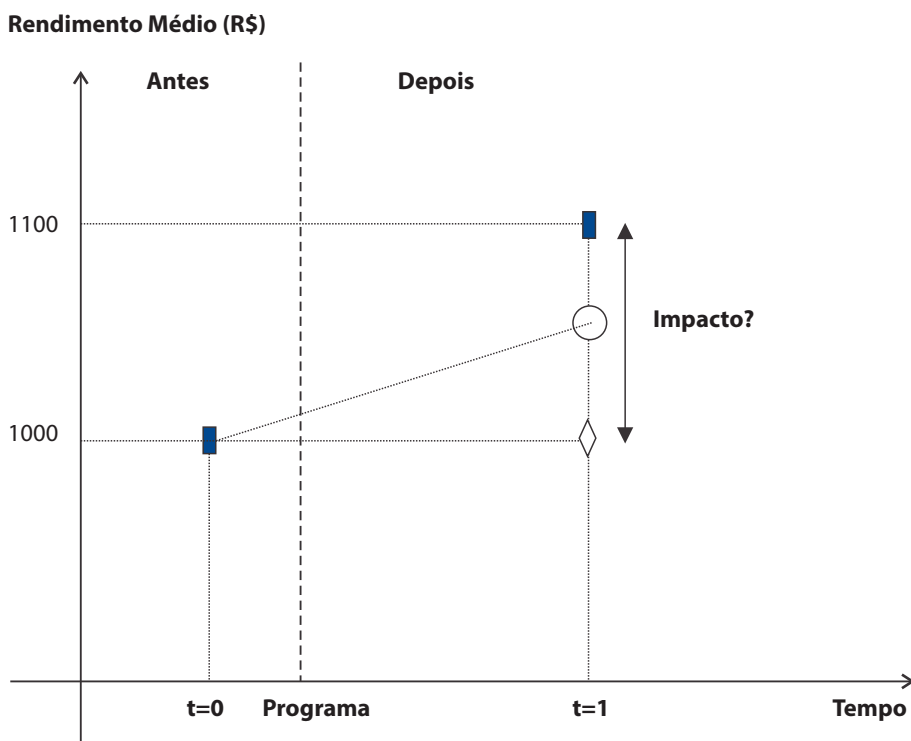
Um pouco antes do começo do programa (por exemplo, no dia da inscrição), todos os participantes responderam um questionário que levantou informações sobre suas características socioeconômicas, incluindo a situação de emprego e rendimento laboral naquele momento. Suponha que o rendimento médio do trabalho para todos os indivíduos que entraram no programa de treinamento tenha sido de R\$ 1000. 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\$ 1100.

A Figura 2.1 ilustra as informações obtidas para o rendimento médio para os períodos anterior e posterior ao programa<sup>1</sup>. Os retângulos estão associados aos valores observados. Alguns diriam que o efeito do programa foi de R\$ 100 (R\$ 1100 – R\$ 1000), 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?

---

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

**Figura 1: Rendimento médio dos participantes antes e depois do programa**



Para responder 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, em termos da 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 intervenção.<sup>2</sup>

## 2.2 Comparando os Tratados com os Não Tratados

Outra prática “ingênua” 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.

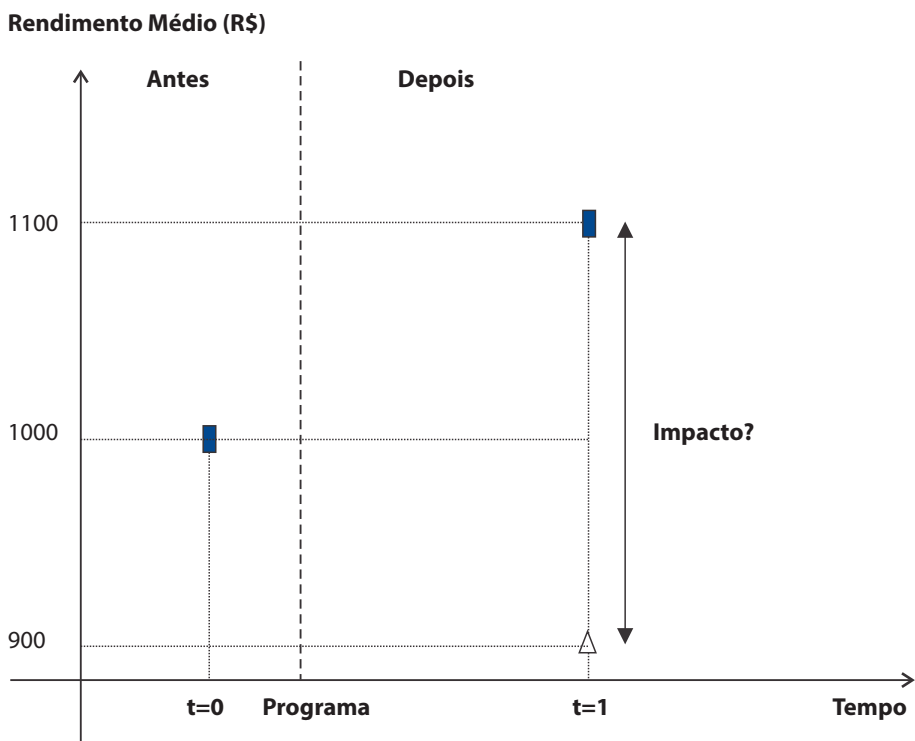
As principais razões para isso estão relacionadas com as diferenças nas características observáveis e não observáveis (pelo avaliador) entre os dois grupos. Por exemplo, é possível que, mesmo que os grupos sejam parecidos (em média) em atributos observáveis como gênero, idade, escolaridade, renda familiar 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 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 mo-

<sup>2</sup> 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.

tivação 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 1: Rendimento médio dos tratados e não tratados**



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.<sup>3</sup> 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

<sup>3</sup> 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.

que pode explicar porque 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 motivação. 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 contrafactual 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.

Cada indivíduo  $i$  possui um par de resultados potenciais  $(Y_i(1), Y_i(0))$  e características observáveis pelo avaliador, que denotaremos por meio do vetor  $X_i$ . Exemplos típicos de características observadas que são incluídas em avaliações são o gênero, a idade e a escolaridade da pessoa. Uma maneira amplamente utilizada para especificar o modelo de geração de valores dos resultados potenciais é o par de equações lineares:

$$(1) Y_i(1) = X_i' \alpha + \beta_i + \varepsilon_i$$

$$(2) Y_i(0) = X_i' \alpha + \varepsilon_i,$$

onde  $\varepsilon_i$  é o componente não observável (pelo avaliador) que influencia os resultados potenciais do indivíduo  $i$ .

Seja  $T_i$  uma variável discreta que assume valor igual a um se o indivíduo par-



ticipa do programa e valor zero caso ele não participe. Trata-se, portanto, do que chamamos de uma variável binária (ou *dummy*, em inglês), cujos valores unitário ou nulo podem depender da decisão do indivíduo de participar ou não da intervenção, respectivamente. O fato de que o valor dessa variável pode depender da decisão do indivíduo é importante, pois, como veremos, essa decisão pode estar correlacionada com outros elementos que determinam  $Y$ , gerando o viés de autosseleção.

Dado o apresentado até aqui, podemos expressar a equação de resultados potenciais como:

$$(3) Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0) = Y_i(0) + T_i (Y_i(1) - Y_i(0)).$$

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

Aplicando agora as equações (1) e (2) na equação (3), esta última se transforma em:

$$(4) Y_i = X_i' \alpha + \beta_i T_i + \varepsilon_i,$$

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 parâmetro que captaria o efeito do programa,  $\beta_i$ , varia entre os indivíduos. Em muitas avaliações, faz-se a hipótese de que o impacto do programa é igual para todos os indivíduos, ou seja,  $\beta_i = \beta$  para qualquer indivíduo da população. Nesse caso, a equação (4) se especializa em<sup>4</sup>:

$$(5) Y_i = X_i' \alpha + \beta T_i + \varepsilon_i.$$

O maior problema para estimar corretamente o impacto do programa por meio da equação (5) é que a variável  $T$  é potencialmente correlacionada com o componente não observável  $\varepsilon$ . 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  $\varepsilon$  da equação (5). Um exemplo é o que apresentamos na seção 2.2, no qual os indivíduos que participam de programas de treinamento

<sup>4</sup>Tratar o impacto da intervenção como homogêneo na população tende a ser uma hipótese forte. De fato, parece mais razoável esperar que o efeito de grande parte dos programas seja heterogêneo entre as pessoas. Contudo, a hipótese de homogeneidade facilita a exposição e, assim, a menos que explicitado no texto, estaremos supondo que ela se aplica.

são considerados mais motivados. Como motivação é uma dimensão do comportamento das pessoas que pode afetar  $Y$  (por exemplo, o rendimento do trabalho) e dificilmente é medida nos dados (ou seja, passa a ser incorporada em  $\varepsilon$ ), aparece um problema de correlação entre  $T$  e  $\varepsilon$  que precisa ser resolvido para que o parâmetro de interesse,  $\beta$ , seja corretamente estimado.

Esse problema de estimação incorreta de  $\beta$  devido à correlação entre a participação no programa ( $T$ ) e a parte não observável ( $\varepsilon$ ) geralmente recebe o nome de problema de autosseleção. A magnitude dessa incorreção é chamada de *viés de autosseleção*. Sua denominação decorre do fato de que são os próprios tratados que se selecionam para participar do programa. Se essa autosseleção estiver associada com 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 o isolamento do efeito causal do programa<sup>5</sup>.

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 autosseleção. A fim de apresentá-lo de forma mais geral, considere as seguintes médias populacionais da variável de resultado para os grupos de tratados e não tratados<sup>6</sup>:

$$D_{11} = E[Y_i(1)|T_i = 1]$$

$$D_{10} = E[Y_i(0)|T_i = 1]$$

$$D_{01} = E[Y_i(1)|T_i = 0]$$

$$D_{00} = E[Y_i(0)|T_i = 0].$$

Note que somente  $D_{11}$  e  $D_{00}$  são médias que podem ser calculadas diretamente dos dados. Já  $D_{10}$  e  $D_{01}$  representam médias contrafatuais, a primeira correspondendo à média para os tratados caso eles não tivessem sido tratados e a segunda a média para os não tratados caso eles tivessem sido tratados.

<sup>5</sup> Muitas vezes, é o governo ou entidade gestora do programa que escolhe o grupo de indivíduos ou famílias que receberão as ações da intervenção. Se essa escolha recair sobre parcelas “especiais” da população em características não observáveis, surge também uma correlação entre o termo  $\varepsilon$  e a variável de participação da equação (5). Nesse caso, muitas vezes diz-se que há um problema de viés de seleção, mas não é de autosseleção, pois não são os próprios indivíduos que se selecionam para participar do tratamento. A análise do problema permanece essencialmente a mesma, apesar das distintas denominações.

<sup>6</sup> Todas as expressões a seguir poderiam estar condicionadas ao vetor de características, ou seja, cada expressão poderia ser computada para diferentes combinações dos valores assumidos pelas variáveis contidas nesse vetor. Para não tornar a notação demasiadamente carregada, optamos por não o incluir nas expressões.

Em geral, na área de avaliação, estamos interessados em computar o que se costuma denominar de efeito médio do programa sobre os tratados (EMPT):

$$(6) D = E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 1] = D_{11} - D_{10},$$

ou seja, 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 de eles não terem passado pela intervenção. O EMPT é um parâmetro importante, uma vez que informa o impacto médio do programa para aquela parcela da população que foi efetivamente tratada<sup>7</sup>.

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 nosso exemplo hipotético do programa de treinamento da seção 2.1. Quando isso ocorre, geralmente calcula-se a diferença:

$$(7) D' = E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 0] = D_{11} - D_{00}$$

que é tomada como o impacto médio do programa. No entanto, isso tende a ser um procedimento incorreto, uma vez que o grupo de não tratados não necessariamente funciona como um bom contrafactual. Para ver isso formalmente, adicionaremos e subtrairemos da equação (7) a média contrafactual  $D_{10} = E[Y_i(0)|T_i = 1]$ :

$$\begin{aligned} (8) D' &= \{E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 0]\} - \{E[Y_i(0)|T_i = 1] - E[Y_i(0)|T_i = 1]\} \\ &= \{E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 1]\} + \{E[Y_i(0)|T_i = 1] - E[Y_i(0)|T_i = 0]\} \\ &= \{D_{11} - D_{10}\} + \{D_{10} - D_{00}\} = EMPT + V. \end{aligned}$$

Como se pode ver,  $D'$  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 vies de autosseleção. Note que ele corresponde à diferença de médias da variável de resultado na situação de não tratamento (isto é,  $Y(0)$ ), quando os indivíduos participam ( $T = 1$ ) ou não ( $T = 0$ ) do programa.

<sup>7</sup> Existem dois outros parâmetros que também recebem bastante atenção na área de avaliação. Um deles é chamado de efeito médio do programa sobre os não tratados (EMPNT):  $N = E[Y_i(1)|T_i = 0] - E[Y_i(0)|T_i = 0] = D_{01} - D_{00}$ , que mediria o impacto da intervenção sobre aquela parcela da população que não foi tratada. O outro parâmetro capta o impacto do programa sobre um indivíduo aleatório da população e recebe o nome de efeito médio do programa (EMP). Ele é definido como:  $P = E[Y_i(1)] - E[Y_i(0)]$ , onde  $E[Y_i(1)]$  e  $E[Y_i(0)]$  são respectivamente as médias para tratados no caso de eles participarem ou não participarem da intervenção.

Como mencionado anteriormente, o termo autosseleção está associado à decisão dos próprios indivíduos de participar ou não da intervenção.

Para que  $V$  seja igual a zero, a média observada para os não tratados,  $D_{00}$ , deve ser igual à média do contrafactual que buscamos,  $D_{10}$ . O problema é que não observamos  $D_{10}$  e, portanto, não temos como saber a magnitude do viés de autosseleção embutido em  $D'$ .

Nossa tarefa nos próximos capítulos será apresentar métodos que buscam solucionar o problema do viés de autosseleção. Cada método lança mão de certos procedimentos e hipóteses para tentar resolver esse problema. Embora cada metodologia vá ser discutida em detalhes em cada capítulo, faremos aqui um sumário das hipóteses básicas utilizadas.<sup>8</sup>

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, o fato de a variável  $T$  assumir valor um ou zero, respectivamente – é definida pelo procedimento de aleatorização garante que os grupos de tratamento e controle sejam parecidos tanto nas características observáveis quanto nas não observáveis. O viés de autosseleção fica assim contornado, permitindo que a comparação entre os grupos identifique o efeito causal do programa.

Outro método que discutiremos em detalhes é chamado *diferenças em diferenças* (DID). 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 mes-

<sup>8</sup> Assumindo a especificação linear da equação (5), se calcularmos  $D'' = E[Y_i|X, T_i = 1] - E[Y_i|X, T_i = 0] = \{X_i'\alpha + \beta + E[\varepsilon_i|X, T_i = 1]\} - \{X_i'\alpha + E[\varepsilon_i|X, T_i = 0]\} = \beta + \{E[\varepsilon_i|X, T_i = 1] - E[\varepsilon_i|X, T_i = 0]\}$ , obteremos o impacto médio do programa,  $\beta$ , somado ao viés de autosseleção, aqui representado por  $E[\varepsilon_i|X, T_i = 1] - E[\varepsilon_i|X, T_i = 0]$ .

ma 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 do grupo de tratamento na ausência da intervenção. Neste 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. A hipótese é que, ao levar em consideração essas características, a variável de resultado passa a independe da participação ou não dos indivíduos no programa – ou seja, as variáveis  $Y(0)$  e  $Y(1)$  tornam-se independentes de  $T$ , fazendo com que o viés de autosseleção 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 score de propensão (MEP). Neste método, assume-se que a probabilidade de um indivíduo participar do programa depende apenas de um conjunto de características observadas, que estão disponíveis nos dados a serem utilizados pelo avaliador. Estimada essa probabilidade, o pareamento dos grupos é realizado em termos das semelhanças dessa probabilidade estimada. A ideia por trás do MPE 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 autosseleção não é passível de ser testada diretamente nos dados.

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 Aprendizizes 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 autosseleção 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.

O último método a ser apresentado nesta parte do livro 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  $T$ , mas não com  $\varepsilon$  na equação (5). Como é capaz de afetar a participação no programa, mas não está correlacionada com as características não observadas, essa variável corrige o problema de autosseleção, 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.

## Exercícios

**1)** Um jornal de grande circulação apresentou numa 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. Essa afirmativa do jornal está 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. Seria o interior do estado um bom grupo de controle para avaliar o programa implantado na capital? Discuta.

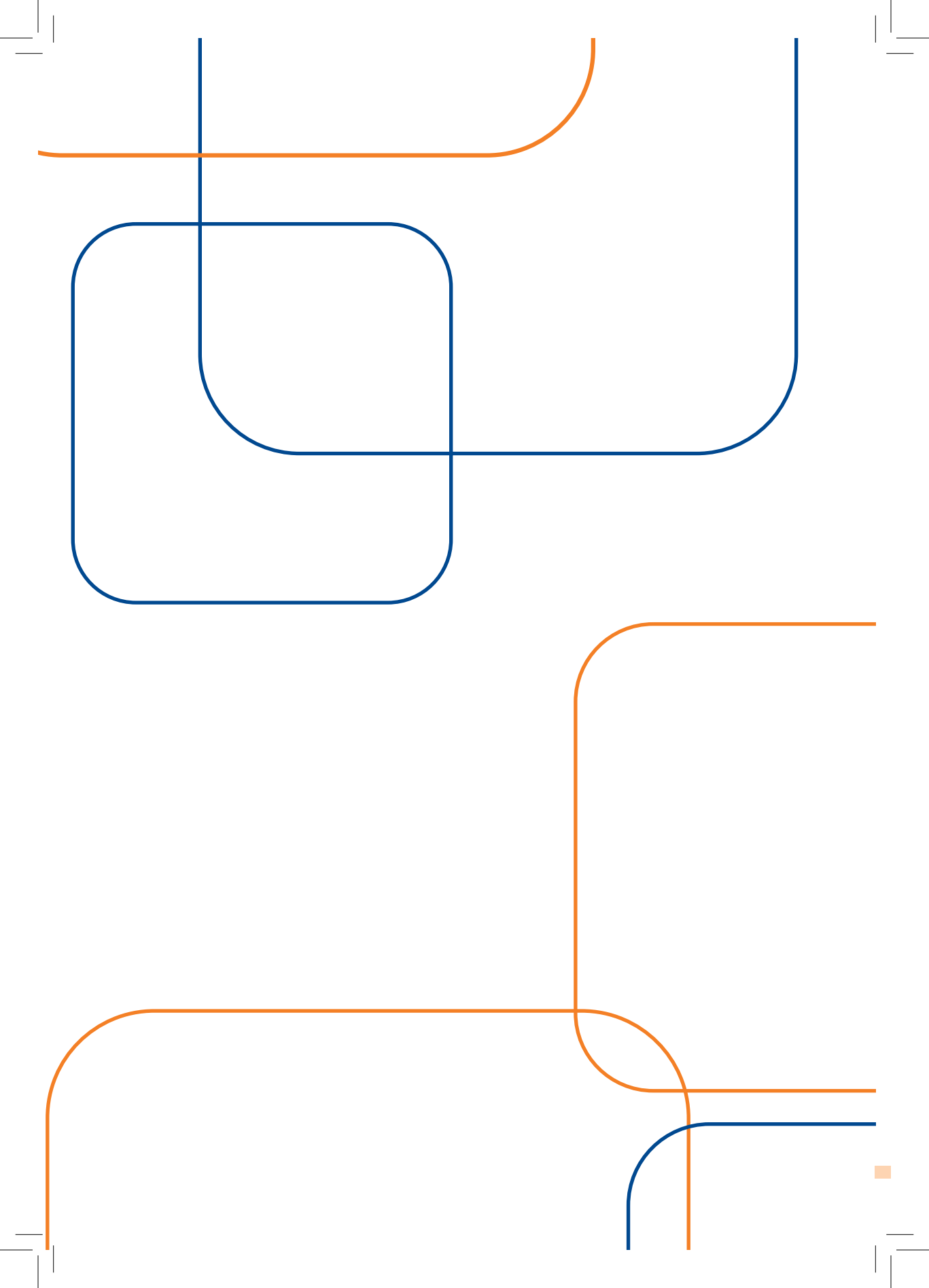
**3)** Qual é a mais importante função do grupo de controle em uma avaliação de impacto?

**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 os seguintes itens:

**a)** Seria o rendimento do trabalho de pessoas com ensino superior influenciado por características normalmente não observadas pelo avaliador? Cite algumas dessas características.

**b)** A escolha por entrar em certo curso universitário equivale à decisão de participar de certo programa. Seria essa decisão correlacionada com o rendimento futuro do trabalho dos formados?

**c)** Qual o principal tipo de problema que pode surgir se compararmos diretamente os rendimentos do trabalho dos estudantes formados em diferentes cursos?





## CAPÍTULO 3

# Método de Aleatorização

Miguel Nathan 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 bastante bem aproximada pelo uso do método de *aleatorização*. Este 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.<sup>9</sup> 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 autoseleção fica assim 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 efetividade (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

<sup>9</sup>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.

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 sobre os tratados. 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.

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 por parte de uma fração do grupo de controle. Em ambos os casos, pode haver uma 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 negligenciáveis na estimação 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 saber, 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 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 são 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. Uma forma de garantir 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 participantes potenciais tem a *mesma* probabilidade de ser selecionado a participar da intervenção, 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 2000 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, 1200 (60%) sejam homens, 1000 (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 2000 potenciais participantes fossem aleatorizados em dois grupos com 1000 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 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, 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 esta 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.

A validade externa é assegurada quando a amostra de indivíduos que é utilizada na avaliação é 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 terá validade externa se a avaliação for baseada numa amostra representativa desses grupos demográficos. Naturalmente, 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.<sup>10</sup> As mais importantes pesquisas socioeconômicas conduzidas pelo IBGE e outros institutos oficiais de estatística do mundo são baseadas em

<sup>10</sup> Para tanto, ela deverá ter sido escolhida com base em algum método de amostragem. 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 no livro clássico de Cochran (1977).

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 garante a validade externa de uma avaliação não assegura 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 através de 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 pelo menos duas situações importantes em que esse problema ético não se coloca.

A primeira, mais comum, ocorre quando existe escassez de vagas para atender toda a demanda ou público-alvo do programa. Nessa situação, pode-se inclusive justificar o uso da aleatorização como o critério mais justo de alocação

das vagas existentes entre os candidatos ao programa. De fato, se pensarmos bem, ao invés de utilizar o tradicional critério de atendimento por ordem de chegada, em muitas situações 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, não é recomendada para assegurar somente a validade interna da avaliação, mas também para garantir a igualdade de oportunidade de acesso aos serviços do programa.

A segunda situação na qual é mais fácil eticamente utilizar o método de aleatorização ocorre quando os recursos do programa (financeiros, de pessoal ou operacionais) 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, essa aleatorização gradual não é realizada no nível indivíduo ou família e sim em um nível mais agregado como bairros, comunidades ou municípios. Um exemplo bastante conhecido de aleatorização gradual foi o programa mexicano de transferência condicional de renda, Progresá (hoje em dia chamado Oportunidades), que teve a primeira parcela dos municípios considerados elegíveis selecionados de forma aleatória pelos gestores do programa.

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 este 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 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.

### 3.3 Como a Aleatorização Resolve o Problema de Autosseleção

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

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

ou seja, um parâmetro que mede a média da variável de resultado para a população de tratados quando ela é tratada subtraída da média da mesma variável para a mesma população, porém na situação contrafactual de não tratamento. Olhando para essa expressão, é possível perceber que a primeira média pode ser calculada diretamente a partir de informações para a população de tratados. No caso de um experimento social, essa média é fornecida pelos dados coletados para o grupo de tratamento, o qual foi aleatoriamente selecionado para participar do programa.

O grande problema é como obter a segunda média da expressão (1), ou seja, como encontrar a verdadeira média da variável de interesse caso os tratados não tivessem sido tratados. 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,  $T$ , independente dos resultados potenciais. Essa independência se dá por força do próprio processo de aleatorização. 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 autosseleção.

Podemos também utilizar equações lineares típicas dos modelos estimados por Mínimos Quadrados Ordinários (MQO) para entender como a aleatorização resolve o problema do viés de autosseleção. Considere a equação:

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

onde o termo  $\varepsilon_i$  aqui capta tanto as variáveis observadas quanto as não observadas do indivíduo  $i$ . Nessa equação, o parâmetro  $\beta$  deve captar o efeito do programa e o viés de autosseleção aparece pela potencial correlação entre  $T$  e  $\varepsilon$ . Essa correlação pode se dar porque a participação no programa é correlacionada tanto com algumas características observadas quanto não observadas das pessoas. Por exemplo, é possível que, quanto maior a motivação dos indivíduos, maior a chance de participação num programa de treinamento de mão de obra.

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

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

O que a aleatorização garante é que a participação no programa é independente das características observadas e não observadas dos indivíduos, exatamente como no exemplo da seção 2.1.1, em que os grupos de tratamento e controle têm todos os atributos semelhantes. É a geração dessa independência entre  $T$  e  $\varepsilon$  que faz o método de aleatorização ser capaz de anular o viés de autosseleção. Em termos da equação (3), isso significa que a aleatorização faz com que:

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

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

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

o que comprova que o método isola corretamente o efeito de interesse. Note que a estimação do efeito pode ser feita com base numa simples regressão de MQO ou, como mostra a expressão (5), através do simples cômputo da diferença de duas médias amostrais:  $\bar{Y}_1 - \bar{Y}_0 = \frac{\sum_{i=1}^{N_1} Y_{1i}}{N_1} - \frac{\sum_{i=1}^{N_0} Y_{0i}}{N_0}$ , onde  $Y_{1i}$  e  $Y_{0i}$  representam a variável de resultado respectivamente 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.<sup>11</sup>

<sup>11</sup> Em termos de inferência, o teste de se o parâmetro  $\beta$  é diferente de zero (o conhecido teste-t) é equivalente ao teste de diferenças de médias. Este último é baseado na estatística



Embora tenhamos incorporado as características observadas dos indivíduos ao termo  $\varepsilon$  na equação (2), é possível explicitá-las no modelo sem produzir grandes alterações nos resultados obtidos até agora. Considere então a equação (2) modificada pela introdução do vetor de características observadas  $X$ :

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

onde o primeiro elemento de  $X$  é igual a um e  $\alpha$  é um vetor de parâmetros conforme a  $X$ . Se bem realizado, o experimento aleatório assegura que:

$$(3') E[\varepsilon_i | X_i, T_i] = 0,$$

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.<sup>12</sup>

Um ponto adicional é que, em princípio, a aleatorização garante que tanto as características observadas como as não observadas são balanceadas entre os grupos de tratamento e controle. 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  $X_k$  contida em  $X = (X_1, X_2, \dots, X_K)$  para o grupo de tratamento seja igual à média da mesma variável para o grupo de controle. Formalmente, isso significa que  $E[X_k | T = 1] = E[X_k | T = 0]$ ,  $\forall k \in \{1, \dots, K\}$ , um resultado que pode ser testado com base nos dados coletados para os dois grupos experimentais.<sup>13</sup> Na realidade, é sempre recomendável testar o balanceamento das características observadas entre os grupos, pois caso ocorra rejeição da hipótese de balanceamento para muitas variáveis em  $X$ , há indicação de que o protocolo de aleatorização não foi bem realizado, o que pode comprometer a acurácia com que os efeitos do programa são estimados.

---

ca  $d = \frac{\bar{y}_1 - \bar{y}_0}{\sigma_{\bar{y}_1 - \bar{y}_0}}$ , onde  $\sigma_{\bar{y}_1 - \bar{y}_0}$  representa o desvio-padrão da diferença de médias. Como as amostras dos grupos de tratamento e controle são independentes,  $\sigma_{\bar{y}_1 - \bar{y}_0} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0}}$ , em que  $N_1$  e  $N_0$  denotam o número de observações das amostras dos dois grupos, respectivamente.

<sup>12</sup>Embora a prova da afirmação a seguir esteja fora do escopo deste livro, vale assinalar que o modelo expresso por (2') e (3') é pelo menos tão eficiente quanto o expresso por (2) e (3). Em outras palavras, o parâmetro pode ser estimado de forma mais precisa com a inclusão das características em no modelo (desde que os coeficientes em sejam diferentes de zero).

<sup>13</sup>Há várias formas de se testar este resultado. Um 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 – exemplos típicos de estratos são: escola, turma, clínica, bairro, etc. –, não necessariamente esse é o teste mais recomendado. Para maiores detalhes, ver Firpo et al. (2011).

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 mostrar impactos diferenciados entre mulheres e homens e entre jovens e adultos. Tipicamente, a incorporação desse tipo de heterogeneidade ao modelo é importante, já que revela aos gestores para que grupos socioeconômicos o programa pode ser aperfeiçoado.

Caso o tamanho da amostra de indivíduos nos grupos de tratamento e controle seja suficientemente grande, é possível estimar o impacto do programa particionando a amostra entre os estratos para os quais se quer conhecer o efeito. Por exemplo, se se quer conhecer o efeito de um programa de creche gratuita para diferentes áreas de um município, pode-se calcular a diferenças de médias da variável de interesse (e.g., participação no mercado de trabalho das mulheres) entre o grupo de tratamento e controle separadamente para cada área da cidade. Naturalmente, é importante que a amostra para cada área da cidade seja relativamente grande, caso contrário o impacto não será precisamente medido.

Podemos também utilizar uma variante da equação (2') para computar impactos heterogêneos de um programa para grupos 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 queiramos medir o efeito médio de uma 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 presentes 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:

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

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  $\gamma$ . A realização de um simples teste t para  $\gamma$  permite inferir se ele é nulo ou não. Da mesma forma que para  $X_K$ , outras variáveis presentes em  $X$  poderiam ser interagidas com  $T$  para obtermos o impacto diferenciado do programa segundo 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 os 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, muito frequente, é a substituição do programa em avaliação por outros semelhantes por parte dos indivíduos do grupo de controle. A terceira dificuldade, chamada atrito amostral, tipicamente 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 dificuldade 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 que é efetivamente o impacto do programa.

Uma forma de olhar para esse problema é que ocorre uma 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 de impacto 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  $Z$  uma variável binária que assume valor unitário quando um indivíduo é aleatorizado para o grupo de tratamento e valor zero 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  $Z$  e a variável que identifica quem efetivamente compareceu,  $T$ . Assim, se a estimação for realizada com base na diferença  $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$ , não estaremos identificando o EMPT, mas sim outro parâmetro, que recebe o nome de *efeito da intenção de tratar* (EIT).<sup>14</sup>

Dada a frequência com que o problema de desistência ocorre, a estimação do EIT pode ser tão relevante quanto a do EMPT. De fato, enquanto o EMPT fornece o impacto somente para os tratados, 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 obrigá-la 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 o efeito do programa. Se olharmos com cuidado a situação, é possível perceber que a variável  $Z$  deve ser correlacionada com a variável  $T$ , porém não associada com qualquer determinante (observado ou não) da variável de resultado,  $Y$ . 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.

<sup>14</sup> 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.

A principal consequência dessa substituição do programa é que o grupo de controle experimental deixa de corresponder ao contrafactual do estado não tratamento. 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.<sup>15</sup>

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  $Z$  como instrumento quando há substituição do programa por parte dos controles. Essa estratégia de identificação será discutida no capítulo 6.

Outra possibilidade é assumir que o efeito do programa substituto para os controles que dele participam é igual ao efeito do programa em avaliação para os efetivamente tratados (Heckman et al., 1999). Se  $S = \{1,0\}$  representa respectivamente as situações em que um indivíduo do grupo de controle recebe um tratamento alternativo e não recebe tratamento nenhum, essa hipótese pode ser expressa por:

$$E[Y_i(1) - Y_i(0)|D_i = 1, T_i = 1] = E[Y_i(2) - Y_i(0)|D_i = 1, S_i = 1],$$

onde  $Y_i(2)$  representaria a variável de resultado no caso do recebimento do tratamento alternativo ao em avaliação. Trata-se claramente de uma hipótese forte, cuja invocação requereria uma análise cuidadosa dos conteúdos das duas intervenções em questão.

## 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 viesada.

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

<sup>15</sup> Novamente, sob a hipótese de que o efeito do programa é homogêneo para todos, essa diferença identificará o impacto do programa.

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 uma 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, se mudam para outras localidades, tornando mais difícil a coleta subsequente de informações.

Como já mencionado, é sempre boa prática comparar as características observadas dos grupos de tratamento e controle, antes e após o 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.<sup>16</sup>

## Externalidades

Além das limitações acima apontadas, os experimentos aleatórios também podem estar associados a dois outros tipos de problemas. O primeiro é um 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 de saúde, educacionais, alimentares, entre outras, em que podem ocorrer externalidades devido ao contato físico ou de comunicação dos indivíduos tratados com os não tratados.<sup>17</sup>

<sup>16</sup> Uma forma comumente utilizada é especificar uma 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 Grisdal (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.

<sup>17</sup> 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

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 90% da população de vários municípios pequenos no Brasil. Em municipalidades como essas, as transferências de renda às famílias beneficiadas pelo programa pode gerar um 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 na qual 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 no longo prazo. Por exemplo, um amplo programa de concessão de bolsas de estudos universitárias 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 estimação dos parâmetros do modelo – inclusive o referente ao impacto do programa –, é possível simulá-lo para conhecer os efeitos diretos e indiretos da intervenção.

---

aplicação de drogas de combate a vermes 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.

## Box 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 basicamente 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 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 2256 indivíduos (1201 tratados) e a de Fortaleza, 2517 (1210 tratados).

A Tabela B.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 parece ter tido 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 um impacto positivo de cerca de 2 pontos de porcentagem (pp) na primeira região e de aproximadamente 3 pp 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 ao nível de significância de 10%.

**Tabela B.1: Estimativas do Impacto do Planfor para Rio de Janeiro e Fortaleza**

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

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 números entre colchetes referem-se aos p-valores reportados para as estimativas do impacto do programa.



## Box 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 um 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 3800 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 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  $Y$  a variável de resultado,  $T$  uma variável binária que indica se uma mulher pertence ao grupo de tratamento ou controle ( $T = 1$  ou  $0$ , respectivamente) e  $Z$  uma variável também binária que marca quais mulheres efetivamente tiveram ou não seus filhos matriculados nas creches ( $Z = 1$  ou  $0$ , respectivamente), o estudo citado estimou o seguinte parâmetro:

$$EMLT = \frac{E[Y|T=1] - E[Y|T=0]}{P[Z=1|T=1] - P[Z=1|T=0]}$$

Em linhas gerais, esse parâmetro pode ser interpretado como efeito do programa que seria estimado pelo método de aleatorização (o numerador) corrigido para o fato de que nem todos os tratados e controles “cumpriram” o que a seleção aleatória determinou (o denominador). Note que essa correção é dada pela diferença entre os 94% das mulheres do grupo de tratamento que tiveram seus filhos efetivamente tratados (o  $P[Z = 1|T = 1]$  do denominador) e os 51% das mulheres do grupo de controle que descumpriram o determinado pela loteria e conseguiram colocar seus filhos nas creches do programa (o  $P[Z = 1|T = 0]$  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%).

## 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 utilizar apenas os dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) – que é uma pesquisa com uma amostra aleatória 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).

**c)** Qual o principal tipo de problema que pode surgir se compararmos diretamente os rendimentos do trabalho dos estudantes formados em diferentes cursos?

**2)** Uma entidade que recebe verba pública presta serviços de treinamento de mão de obra para trabalhadores de baixa renda num 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 morava 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.

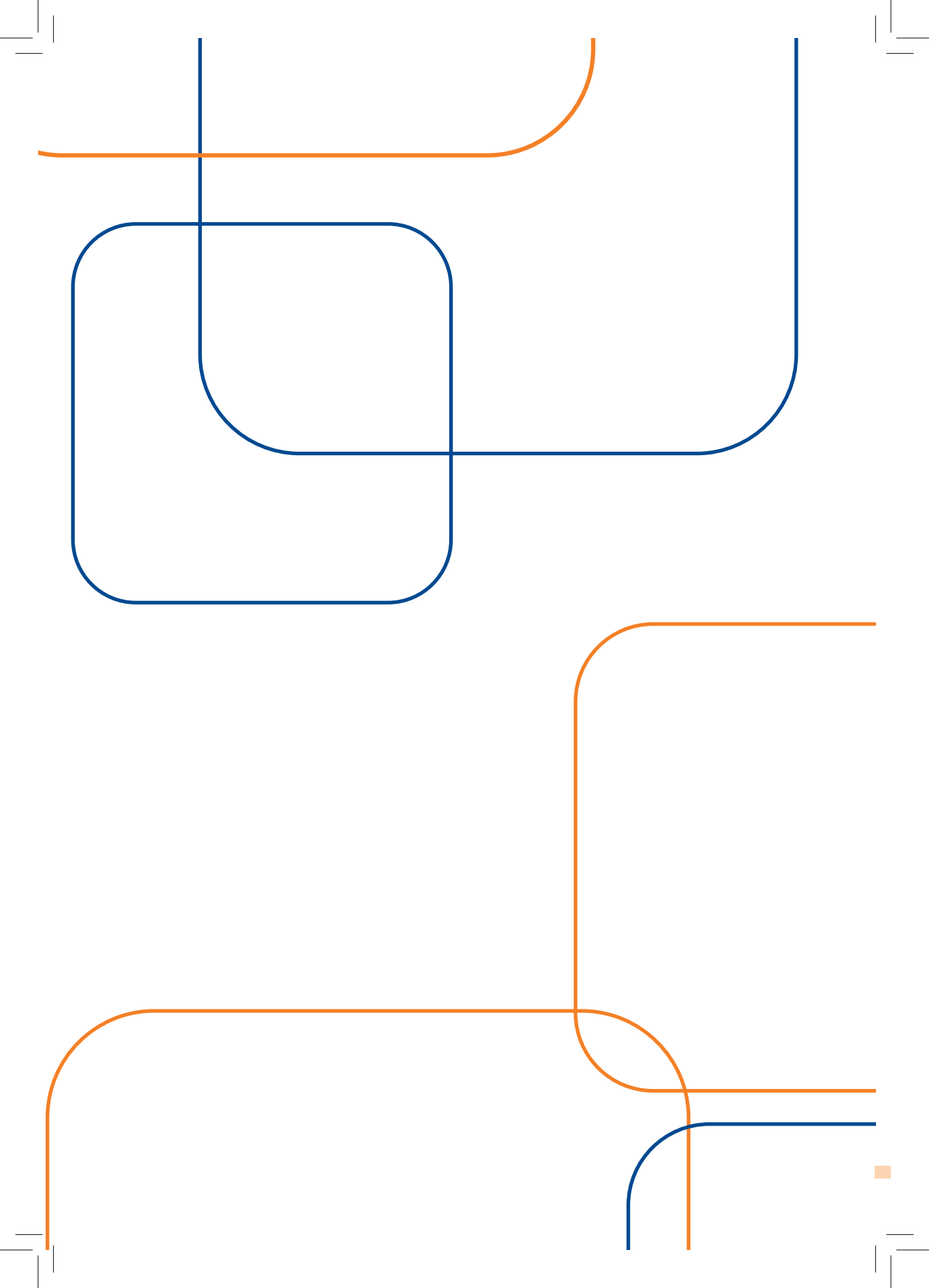
**4)** Em uma avaliação de impacto aleatorizada de um programa de reforço escolar, descobriu-se que 10% dos alunos do grupo de tratamento desistiram do programa. Ao analisar as informações disponíveis, os avaliadores observaram

que esses alunos que não compareceram tinham históricos escolares piores que os demais.

**a)** Você trataria essa informação como uma evidência de que há características não observadas dos alunos que influenciaram na desistência do programa?

**b)** A diferença entre o grupo que foi aleatorizado para participar e o que efetivamente participou não impede o uso do experimento para estimar o impacto do programa. Discuta o efeito causal que poderia ser estimado nessa situação.

**c)** Distintamente, se a situação fosse que 10% dos indivíduos do grupo de controle substituíram o programa por outros similares, o experimento aleatório estaria invalidado para se estimar o impacto do programa? Explique.



## CAPÍTULO 4

# Diferenças em Diferenças

Miguel Nathan 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 demanda por dados torne a aplicação do procedimento mais custosa, o método possui uma série de vantagens relativamente 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 procedimento de DD é ampla e ele tem sido utilizado isoladamente ou em combinação com outros métodos em diversas situações.<sup>18</sup> Muitas vezes, ele é empregado no contexto do que se chama *experimento natural* (ou *quase-experimento*) uma situação na qual a ocorrência de um evento fortuito – ou em larga medida imprevisto – permite formar grupos de trata-

<sup>18</sup> Por exemplo, no capítulo 5, apresentaremos como o método de pareamento é combinado com o método de DD.

mento 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 um alto grau de aleatoriedade.<sup>19</sup> 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.<sup>20</sup> 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 pré-existentes 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. Neste capítulo, a menos que explicitamente mencionado, não distinguiremos se o procedimento de DD está sendo aplicado num contexto de experimento aleatório, natural ou outra situação em que as condições demandadas pelo método são atendidas.

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 *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 este não seja um teste cabal da hipótese, ele é bastante intuitivo e pode ser realizado com facilidade.

<sup>19</sup> 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.

<sup>20</sup> Vale observar que o método de DD também é utilizado com experimentos puramente aleatórios, sempre que estes levantam informações para os períodos pré e pós-intervenção.

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 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 uma fonte potencial de geração de viés de autosseleção. 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 autosseleção 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 autosseleção 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 um 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 há 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 entre os grupos. Esse problema torna-se particularmente sério quando essas características não são observadas (ou medidas) pelo avaliador.

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 é 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 um 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.

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 através 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 inter-



venção, 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 do programa logo após o início deste.<sup>21</sup> Vale ressaltar que, embora a existência de uma mesma tendência temporal pré-programa não garanta que o grupo de controle é 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 um 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 ou famílias que são seguidos ao longo do tempo, enquanto os segundos são dados temporais calculados para unidades mais agregadas tais como escolas, hospitais, setores de atividade, bairros, municípios ou estados. A agregação dos dados normalmente é feita de termos de proporções ou médias das variáveis de interesse para a avaliação. Já os momentos no tempo geralmente são meses ou anos.<sup>22</sup>

A despeito do tipo de dados, outra hipótese do método de DD é que a compo-

---

<sup>21</sup> 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.

<sup>22</sup> Proposto por Abadie et al. (2010) para contextos em que a unidade de tratamento é agregada, há um método recente que é um pouco mais geral que o de DD. Nele, busca-se de forma flexível um conjunto de unidades agregadas não tratadas para o qual a trajetória da média da variável de resultado seja a mais próxima possível (segundo certo critério) da trajetória da unidade tratada antes do programa. Em linhas gerais, o método procura um vetor de pesos para unidades não tratadas que as aproxime ao máximo do que ocorreu com a unidade tratada pré-intervenção.

sição 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). O impacto do programa pode ser mal identificado pelo método de DD nesse caso, 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 um esvaziamento de pessoas de um dos dois tipos de áreas, é provável que a distribuição das características observáveis e não observáveis esteja mudando em um ou ambos os tipos de área. Parte então do que o método atribuirá ao efeito do tratamento na realidade pode se dever às mudanças de composição dos grupos.

Outra condição requerida pelo método de DD é que os grupos de tratamento e controle não sejam afetados de forma heterogênea 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.<sup>23</sup> 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 será dado por:

$$(1) \beta_{DD} = \{E[Y|T = 1, t = 1] - E[Y|T = 1, t = 0]\} - \{E[Y|T = 0, t = 1] - E[Y|T = 0, t = 0]\},$$

<sup>23</sup> É possível também que choques macroeconômicos tenham efeitos diferenciados sobre os grupos de tratamento e controle, ou seja, não se pode garantir nem mesmo que choques agregados sejam homogêneos sobre os grupos.

ou seja, pela diferença temporal do que ocorreu com o grupo tratamento subtraída da mesma diferença 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.<sup>24</sup> Em decorrência dessa hipótese, a diferença entre a variação efetivamente observada para o grupo de tratamento e a variação contrafactual fornecida pelo grupo de controle vai capturar o efeito causal da intervenção.

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 procedimento 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:

$$(2) \beta_{DD} = \{E[Y|T = 1, t = 1] - E[Y|T = 0, t = 1]\} - \{E[Y|T = 1, t = 0] - E[Y|T = 0, t = 0]\},$$

que é apenas um rearranjo da expressão (1). A expressão (2) 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 (2) seja igual a zero.

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\$ 1000 um pouco antes do início do curso e de R\$ 1100 alguns meses após o programa. Argumentamos naquele capítulo que nem a comparação do grupo tratado antes e depois da intervenção e nem a compa-

<sup>24</sup> Se utilizarmos o arcabouço de resultados potenciais apresentado no capítulo 2, a hipótese requerida pelo método de DD é:  $E[Y(0)|T = 1, t = 1] - E[Y(0)|T = 1, t = 0] = E[Y(0)|T = 0, t = 1] - E[Y(0)|T = 0, t = 0]$ , ou seja, que a variação temporal na média do contrafactual do grupo tratado seja igual à variação observada na média do grupo de controle.

ração 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 – respectivamente,  $(T_1, T_0)$  – para os períodos anterior e posterior ao programa –  $(t_0, t_1)$ , respectivamente. A notação  $\bar{Y}$  representa a média amostral do salário e  $\Delta$  significa o operador de diferença.

Distintamente dos outros dois métodos, o procedimento de DD computa uma dupla diferença de médias. Acompanhando as expressões (1) e (2), 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 (2), 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\$**

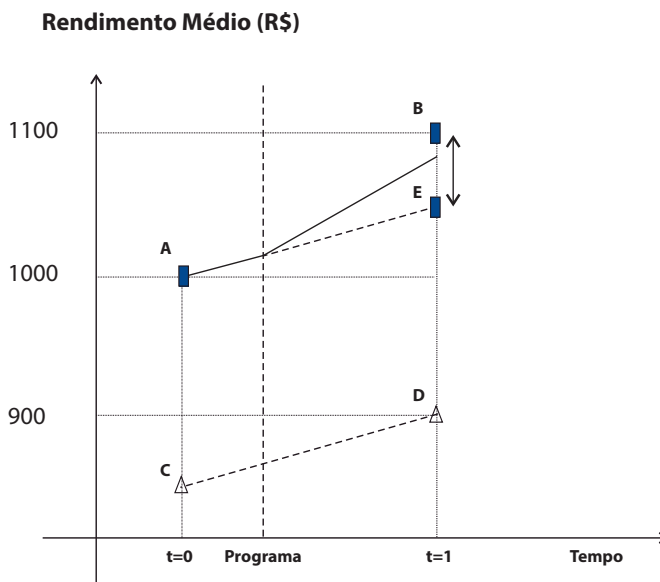
	Tratados	Controle	Diferença
<b>Antes</b>	$\bar{Y}_{T_1, t_0} = 1000$	$\bar{Y}_{T_0, t_0} = 850$	$\Delta\bar{Y}_{T_1 - T_0, t_0} = 150$
<b>Depois</b>	$\bar{Y}_{T_1, t_1} = 1100$	$\bar{Y}_{T_0, t_1} = 900$	$\Delta\bar{Y}_{T_1 - T_0, t_1} = 200$
<b>Diferença</b>	$\Delta\bar{Y}_{T_1, t_1 - t_0} = 100$	$\Delta\bar{Y}_{T_0, t_1 - t_0} = 50$	$\Delta\bar{Y}_{T_1, t_1 - t_0} - \Delta\bar{Y}_{T_0, t_1 - t_0}$ $= \Delta\bar{Y}_{T_1 - T_0, t_1} - \Delta\bar{Y}_{T_1 - T_0, t_0}$ $= 50$

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 quadrado listrado, que representa o contrafactual do grupo de tratamento após o programa. Note

que a posição dele é obtida através de uma projeção da média salarial dos tratados que é paralela à que ocorreu com o grupo de controle depois do programa. Essa projeção, que está 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 forma 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 forma de dupla diferença apresentada anteriormente, obtemos  $(B - D) - (A - C) = (1100 - 900) - (1000 - 850) = 50$ , um resultado que 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:

$$(3) Y_{it} = X'_{it}\alpha + \gamma T_{it} + \rho t_{it} + \beta(T_{it}t_{it}) + \varepsilon_{it},$$

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, o vetor  $X$  representa as características observadas do indivíduo e  $T$  é uma variável binária que assume valor unitário, se o indivíduo é tratado, e valor nulo, caso contrário. A variável  $t$  também é binária, assumindo valor um no período pós-programa e valor zero, caso contrário. O termo de erro é dado por  $\varepsilon$ . Sob a hipótese de que  $E[\varepsilon|X, T, t] = 0$ , o efeito causal do programa (condicional a  $X$ ) é medido pelo parâmetro  $\beta$ .<sup>25</sup>

Note que, na equação (3), as variáveis  $T$  e  $t$  aparecem tanto isoladamente quanto interagidas. Caso não houvesse o termo de interação entre as variáveis, suas presenças isoladas captariam respectivamente as diferenças da média de  $Y$  entre o grupo de tratamento e controle e entre o período anterior e posterior ao programa (condicional a  $X$ ). A introdução do termo de interação na equação procura captar 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 em particular tornou-se diferente após a intervenção. É fácil ver então que o papel do termo de interação entre as variáveis  $T$  e  $t$  é medir o impacto do programa. A magnitude desse impacto é, portanto, capturada pelo parâmetro  $\beta$ . Uma forma de verificar isso é escrevendo a média (condicional a  $X$ ) da variável de resultado para os seguintes casos (retirando os subscritos para facilitar a leitura):

Tratados, pré-programa:

$$E[Y|X, T = 1, t = 0] = X'\alpha + \gamma + E[\varepsilon|X, T = 1, t = 0] = X'\alpha + \gamma$$

Tratados, pós-programa:

$$E[Y|X, T = 1, t = 1] = X'\alpha + \gamma + \rho + \beta + E[\varepsilon|X, T = 1, t = 1] = X'\alpha + \gamma + \rho + \beta$$

Controles, pré-programa:

$$E[Y|X, T = 0, t = 0] = X'\alpha + E[\varepsilon|X, T = 0, t = 0] = X'\alpha$$

Controles, pós-programa:

$$E[Y|X, T = 0, t = 1] = X'\alpha + \rho + E[\varepsilon|X, T = 0, t = 1] = X'\alpha + \rho,$$

<sup>25</sup> Sem a presença do vetor de características  $X$ , o parâmetro  $\beta$  equivale à dupla diferença de médias do método de DD discutida anteriormente.

onde as últimas igualdades são obtidas sob a hipótese de média condicional nula:  $E[\varepsilon|X, T, t] = 0$ . Assim, calculando a dupla diferença:

$$\{(B) - (A)\} - \{(D) - (C)\} = \{(X'\alpha + \gamma + \rho + \beta) - (X'\alpha + \gamma)\} - \{(X'\alpha + \rho) - (X'\alpha)\} = \{\rho + \beta\} - \{\rho\} = \beta.^{26}$$

A equação (3) 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  $\tau$  representa o número máximo de períodos para o qual temos informações sobre os dois grupos. Digamos que a intervenção ocorre em certo mês (ou ano)  $k$  ao longo do período de observação – isto é,  $1 < k < \tau$ . Considere a equação a seguir, que é uma modificação da equação (3):

$$(4) Y_{it} = X'_{it}\alpha + \gamma T_i + \rho_2 dt_{i2} + \dots + \rho_{2\tau} dt_{i\tau} + \beta D_{it} + \varepsilon_{it},$$

onde  $T_i = \{1, 0\}$  é uma variável binária que assume valor um se o indivíduo  $i$  é tratado,  $D_{it} = \{1, 0\}$  também é binária mas só assume valor unitário se o indivíduo  $i$  é tratado e o período  $t$  se refere a qualquer mês após o início da intervenção (isto é,  $t \geq k$ ). Qualquer variável  $dt_{i\tau}$  do conjunto  $\{dt_{i2}, \dots, dt_{i\tau}\}$  assume valor unitário ou zero dependendo de se o mês da observação corresponde ao mês  $t$  ou não, respectivamente (por exemplo,  $dt_{i2}$  será igual a um se o mês for o segundo e zero, caso contrário).<sup>27</sup> A função de cada variável desse conjunto é controlar para efeitos agregados de tempo que possam afetar a variável de resultado de todos os indivíduos do grupo de tratamento e controle (por exemplo, inflação). Analogamente à equação (3), o efeito causal do programa também será dado pelo parâmetro  $\beta$  na equação (4).

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, essas características fixas não observáveis influenciam também a participação no 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) também seriam outros exemplos de características não observáveis que podem afetar tanto salários quanto a decisão de participar no programa. Qualquer que seja o caso, 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 obser-

<sup>26</sup> Note que o primeiro método “ingênuo” da diferença de médias antes e depois do programa para o grupo de tratamento fornece:  $\{(B) - (A)\} = \rho + \beta$ , o que revela um viés igual a  $\rho$ . Já o segundo método “ingênuo” compara dos tratados com os controles no período após o programa. Por este método, o impacto seria:  $\{(B) - (D)\} = \gamma + \beta$ , o que corresponde a um viés de magnitude  $\gamma$ .

<sup>27</sup> Note-se que excluímos da equação a variável binária  $dt_{i1}$  para evitar o problema de multicolinearidade perfeita com a (implícita) presença do intercepto no vetor de parâmetros  $\alpha$ .

váveis dos indivíduos que sejam invariantes no tempo (pelo menos durante o período de observação). Assim, como o problema do viés de autoseleção pode surgir da associação entre essas dimensões, o método de DD oferece uma ferramenta poderosa para contornar esse problema.

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 passa a poder incluir os chamados efeitos fixos individuais, os quais entram na equação como um conjunto de variáveis binárias que têm valor unitário ou nulo dependendo de com qual indivíduo elas estão associadas. 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 pode ser escrita como:

$$(5) Y_{it} = X'_{it}\alpha + \rho_2 dt_{i2} + \dots + \rho_{2\tau} dt_{i\tau} + \beta D_{it} + \mu_2 I_2 + \dots + \mu_N I_N + \varepsilon_{it},$$

onde  $I_i$  é igual a um para o indivíduo  $i$  e zero para todos os demais indivíduos da amostra (exemplo:  $I_2$  recebe valor um para o indivíduo  $i = 2$  e zero para o resto dos indivíduos)<sup>28</sup> A inclusão desses efeitos fixos específicos de cada indivíduo permite controlar para heterogeneidades existentes entre as pessoas em características que sejam fixas no tempo<sup>29</sup> 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,  $D_{it}$ . É ao permitir esse tipo de correlação que o método de DD controla para o viés de autoseleção.

Uma forma de estimar a equação (5) é aplicar o método do MQO à primeira diferença dessa equação:<sup>30</sup>

$$(6) Y_{it} - Y_{it-1} = \Delta Y_{it} = (X'_{it} - X'_{it-1})\alpha + \rho_2 dt_{i2} + \dots + \rho_{2\tau} dt_{i\tau} + \beta(D_{it} - D_{it-1}) + (\mu_2 I_2 - \mu_2 I_2) + \dots + (\mu_N I_N - \mu_N I_N) + (\varepsilon_{it} - \varepsilon_{it-1}) = \Delta X'_{it}\alpha + \rho_2 dt_{i2} + \dots + \rho_{2\tau} dt_{i\tau} + \beta \Delta D_{it} + \Delta \varepsilon_{it}.$$

Note que os efeitos fixos individuais são eliminados na equação (6) e isso só ocorre porque eles são invariantes no tempo. Note também que essa equação

<sup>28</sup> Novamente, excluímos a variável  $I_1$  para evitar o problema de multicolinearidade perfeita com o intercepto em  $X$ .

<sup>29</sup> Estrito senso, a presença do efeito fixo individual na equação capta a influência sobre  $Y$  de qualquer característica, observável ou não, do indivíduo 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  $T_1$  não aparece na equação (6). De forma semelhante, eventuais variáveis tempo-invariantes contidas no vetor  $X$  (por exemplo, sexo e cor) também não poderiam aparecer na equação.

<sup>30</sup> Outra forma de estimação comumente utilizada com dados de painel é a que se chama de intragrupos. Neste caso, cada variável é subtraída do valor de sua média temporal para cada indivíduo, por exemplo, subtrai-se  $Y_{it}$  de  $\bar{Y}_i = \tau^{-1} \sum_{t=1}^{\tau} Y_{it}$ .



está em linha com a aplicação do método de DD, pois corresponde ao cálculo de sua primeira diferença. A hipótese básica que permite estimar o impacto do programa pela equação (6) – o qual é dado pelo parâmetro  $\beta$  – é que o termo  $\Delta\varepsilon_{it}$  não seja correlacionado com as variáveis do modelo, em particular com  $D_{it}$ .<sup>31</sup>

Embora estejamos nos referindo a indivíduos ao apresentar as equações acima, *mutatis mutandis*, elas também podem ser utilizadas para o caso em que as unidades de observação estão num nível mais agregado, como escolas, setores de atividade, municípios ou estados. Uma das principais vantagens de usar dados agregados (e.g., médias, proporções) é que eles tendem a atenuar problemas de erros de medidas que tipicamente aparecem em informações coletadas no nível do indivíduo ou família. Por outro lado, para que o método de DD identifique corretamente o efeito causal de uma intervenção é importante que tanto o grupo de tratamento quanto o de controle permaneçam similares nas suas características ao longo do tempo. Em particular, é importante que não ocorram mudanças expressivas na composição dos grupos no que tange aos efeitos fixos dos indivíduos que compõem os agregados de cada grupo. Esse tipo de mudanças, ao tornar os grupos distintos entre períodos de tempo, enfraquece (ou impede) a comparabilidade temporal dos grupos. Por exemplo, a migração de pessoas entre municípios ou estados pode tornar os grupos muito diferentes antes e depois de uma intervenção. Ou a dinâmica própria do sistema educacional pode fazer com que a composição dos alunos das escolas se altere de ano para ano. 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 de cada grupo 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 não consegue lidar com casos em que alguma mudança temporária num fator não observável dos indivíduos afeta a decisão de participação 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). Além de poder ter influenciado os indivíduos a fazerem o curso, essa queda temporária nos rendimentos poderá fazer com que o grupo tratado experimente um crescimento maior nos seus rendimentos mesmo se ele não tivesse participado do programa. Se essa mudança temporária não ocorrer com o grupo de controle, o procedi-

<sup>31</sup> Embora fora do escopo do livro, vale observar que, se o termo  $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$  apresentar correlação serial – isto é, a correlação entre  $\varepsilon_{it}$  e  $\varepsilon_{it-1}$  ser diferente de zero –, há que se fazer alguma correção para os erros-padrão a serem estimados pelo procedimento de DD. Mais detalhes podem ser encontrados em Bertrand et al. (2004).

mento de DD tenderá a sobre-estimar o verdadeiro impacto do programa. De forma mais geral, vieses aparecerão no estimador de DD sempre que houver algum tipo de característica não observável que varie no tempo e afete simultaneamente a variável de resultado e a participação no programa.

#### Box 1: 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. Idealmente 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 pro-

gressivamente 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.

## Exercícios

1) Foi noticiado em dois jornais locais que a política atual de redução da criminalidade de seu estado era eficaz. Em um dos jornais, a comprovação dessa eficácia 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 (talvez porque você ou alguém próximo teve seu carro roubado recentemente!). Responda os itens a seguir.

- a) Por que você acha a matéria do primeiro jornal ingênuas? E a segunda?
- b) Você decide estimar o efeito da política usando o método das diferenças em diferenças. Quais os dados que você deve buscar?
- 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?

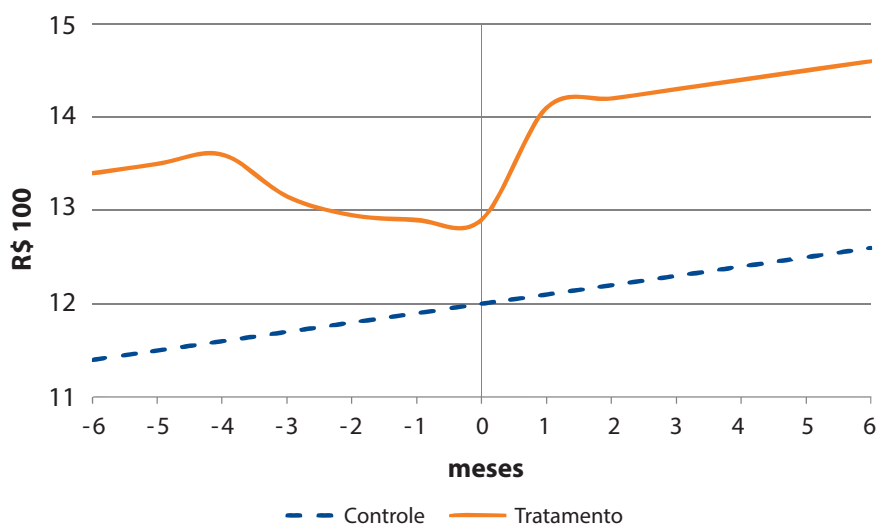
2) Um grupo de biólogos utilizou a erupção de um vulcão que ocorreu numa ilha para estudar o efeito de erupções vulcânicas sobre a flora e a fauna daquela região. Para tanto, eles compararam os dados da ilha que sofreu a erupção com os dados de outras ilhas antes e depois da erupção. Nem sempre na área de avaliação de políticas ou programas sociais tem-se a certeza de que a intervenção ocorreu de forma totalmente exógena como no caso da erupção do vulcão. Crie ou busque exemplos de programas sociais que se aproximem da situação de experimento natural.

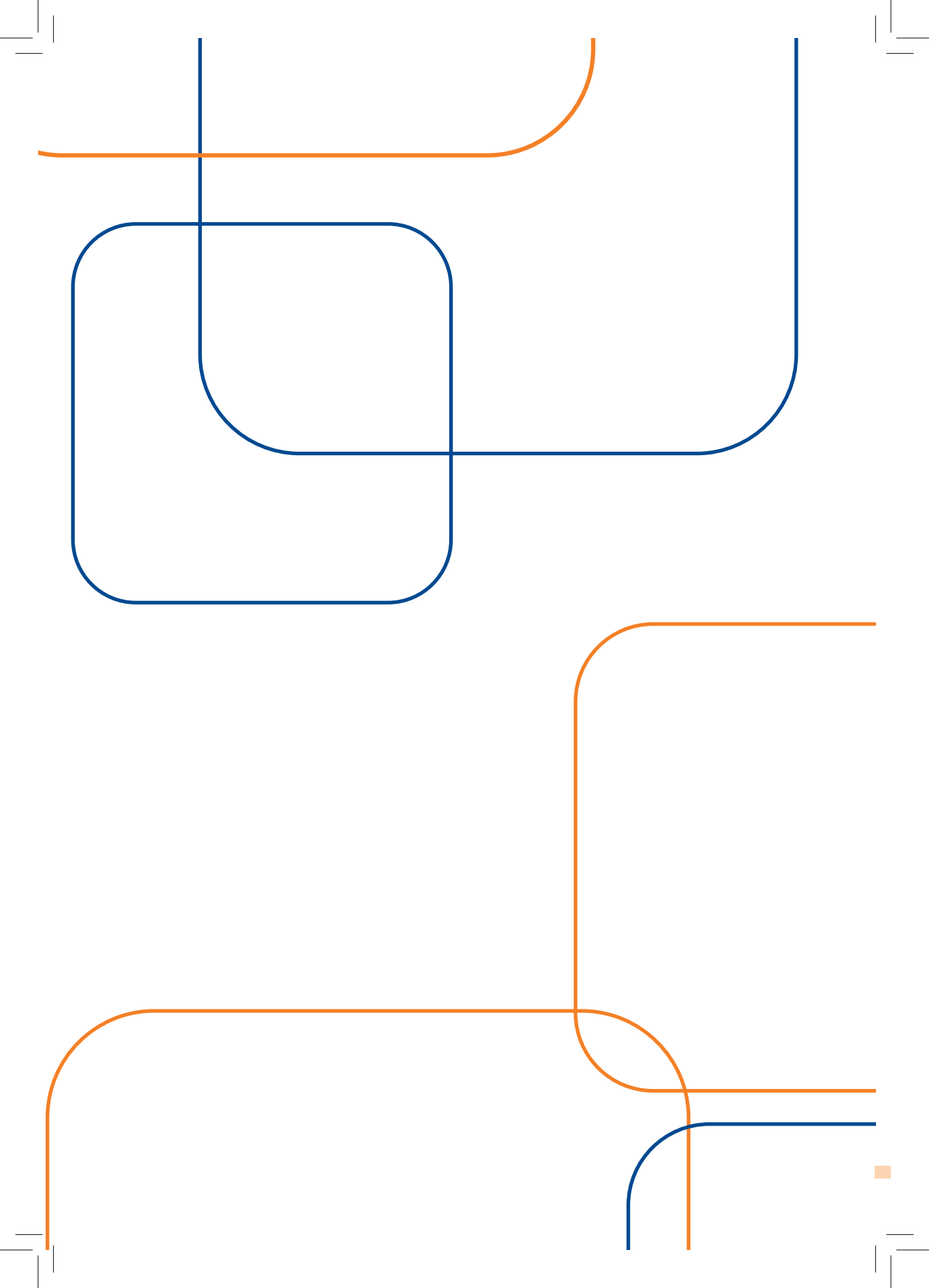
3) Uma das grandes vantagens do método de diferenças em diferenças é que ele é capaz de lidar com a autoseleção decorrente de certo tipo de características não observadas dos indivíduos. Que tipo de características é esse e como o método as incorpora no modelo?

4) Certo município começou a realizar 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 de município. Discuta os problemas que o avaliador precisa levar em consideração com os dados 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 zero no gráfico. É possível que o método das diferenças em diferenças estime de forma enviesada o impacto do programa? Por quê?

Gráfico 4.2: Trajetória do Salário Médio por Grupo





## CAPÍTULO 5

# Pareamento

Cristine Campos Xavier Pinto

O método do pareamento busca construir um grupo de controle semelhante ao grupo de tratamento em termos de determinadas características observáveis.<sup>32</sup> De acordo com as hipóteses deste 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 destes indivíduos é a participação ou não no programa.

— A hipótese principal deste método, que tem como objetivo estimar o efeito médio do tratamento sobre os tratados (EMPT), é que o vetor de variáveis observáveis  $X$  contém todas as informações sobre o resultado potencial na ausência do tratamento ( $Y(0)$ ) que o indivíduo possui ao tomar a decisão de participar ou não do tratamento. Ao levar em consideração estas características observáveis, o avaliador controla por todas as variáveis que estão relacionadas ao resultado potencial na ausência de tratamento e que também afetam a decisão do indivíduo em participar ou não. Assim, ao controlar pelo vetor  $X$ , a variável  $Y(0)$  torna-se independente de  $T$ . Esta hipótese é conhecida como seleção nos observáveis, ou ignorabilidade, ou ainda não-confundimento, e pode ser formalizada como:<sup>33</sup>

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

A implicação desta hipótese é que o resultado de um indivíduo no grupo de controle é um bom predictor do resultado potencial na ausência de tratamento de um indivíduo no grupo de tratamento que possui o mesmo vetor de variáveis observáveis ( $X$ ).<sup>34</sup> Desta forma, para estimar o EMPT via pareamento, pre-

<sup>32</sup> Existe uma 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).

<sup>33</sup> Para estudar o efeito médio do tratamento, precisamos da hipótese de independência nos dois resultados potenciais,  $(Y_i(1), Y_i(0)) \perp T_i | X_i$ .

<sup>34</sup> Esta 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

cisamos que cada indivíduo no grupo de tratamento tenha um par no grupo de controle, cujo resultado reproduz o que seria o resultado deste indivíduo na ausência de tratamento. Desta forma, precisamos que a região do vetor  $X$  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<sup>35</sup>

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

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 estas características não-observáveis forem correlacionadas tanto com a decisão de participar do programa como com o resultado potencial deste indivíduo, o método do pareamento não conseguirá eliminar o viés de seleção, e a estimação do EMPT estará sujeita a um viés.

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

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

em que  $E[Y_i(1)|T_i = 1, X = x]$  é a média populacional de  $Y$  para os tratados com uma determinada combinação de características  $X$  e  $E[Y_i(0)|T_i = 1, X = x]$  é a média de  $Y$  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  $X$  seria a diferença entre a média populacional do resultado para os indivíduos tratados e a média populacional do resultado de interesse para estes 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 destes

---

independência na média, isto é,  $E[Y_i(0)|T_i, X_i] = E[Y_i(0)|T_i = 0, X_i] = E[Y_i(0)|X_i]$ .

<sup>35</sup> Como no caso do EMP, precisamos que para cada indivíduo no grupo de tratamento tenhamos um par no grupo de controle cujo resultado potencial seja um bom predictor do resultado deste indivíduo na ausência do tratamento. Também precisamos que para cada indivíduo no grupo de controle tenhamos um par no grupo de tratamento cujo resultado potencial seja um bom predictor do resultado deste indivíduo caso ele fosse tratado. Nesse caso, a hipótese de sobreposição deve ser expandida para:  $0 < \Pr[T_i = 1|X_i] < 1$ . Esta hipótese garante que as características dos tratados sejam representadas no grupo de controle, e vice-versa.

indivíduos tratados caso eles não tivessem recebido tratamento, e não podemos simplesmente estimar a segunda média condicional usando uma média amostral. No entanto, a hipótese H1 nos permite escrever o segundo termo da equação 5.1 em função das variáveis que observamos.

Sob a hipótese H1,

$$\begin{aligned} E[Y_i(0)|T_i = 1, X = x] &= E[Y_i(0)|X = x] = E[Y_i(0)|T_i = 0, X = x] \\ &= E[Y_i|T_i = 0, X = x] \end{aligned}$$

Ou seja, em média, os resultados sem tratamento dos tratados que apresentem uma determinada combinação de características presentes em  $X$  são iguais aos resultados sem tratamento dos não tratados (grupo de controle) com essas mesmas características, que são observáveis para o avaliador.

Este resultado juntamente com a hipótese H2 indicam 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  $X$  que os indivíduos no grupo de tratamento.

E, portanto, podemos escrever:

$$D(x) = E[Y_i|T_i = 1, X = x] - E[Y_i|T_i = 0, X = x] \quad (5.2)$$

Assim, podemos calcular o efeito do tratamento para os indivíduos com uma determinada combinação de características  $X$  comparando a média observada de  $Y$  para os tratados com essas características com a média observada de  $Y$  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  $X$ , condicional a  $T=1$ :<sup>36</sup>

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

Por exemplo, no caso em que  $X_i$  inclui apenas uma variável explicativa discreta,

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

Neste caso,  $D$  é uma média ponderada dos efeitos médios para os indivíduos com determinada combinação de características  $X$  ( $D(x)$ ), na qual os pesos são

<sup>36</sup> Note que para obter o efeito médio do tratamento, precisamos calcular a expectativa de  $D(x)$  sobre a distribuição de  $X$ . No caso em que  $X_i$  inclui apenas uma variável explicativa discreta,

dados pela porcentagem de indivíduos tratados que possuem esta combinação de características  $X$ .

Para estimar o EMPT, precisamos estimar  $E[Y_i|T_i = 1, X = x]$  e  $E[Y_i|T_i = 0, X = x]$ . Se a hipótese H2 for violada, não será possível estimar ambas as médias, pois os valores de  $x$  serão válidos para o grupo de tratamento, mas não para o grupo de não-tratados.

## 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 destes indivíduos para obter o que seria o resultado do indivíduo no grupo de tratamento caso ele não fosse tratado (contrafactual). As principais diferenças entre os vários estimadores de pareamento dizem respeito à métrica usada para definir os indivíduos mais “próximos” dos tratados em termos do vetor de variáveis  $X$ . Além disso, os métodos também diferem em relação a quantos indivíduos do grupo de não-tratados serão relacionados a cada indivíduo no grupo de tratamento para obter o seu contrafactual.

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<sup>37</sup>, que nos diz quais são os indivíduos no grupo de controle que possuem um número de características observáveis mais próximas do indivíduo  $i$ , identificamos os  $M$  pares mais próximos deste indivíduo. De modo mais formal, definimos como  $L_M(i)$  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:

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

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

<sup>37</sup> Um exemplo de medida de distância é a métrica definida em Imbens e Wooldridge (2007) para determinar quão diferentes os indivíduos são em relação ao vetor  $X$ ,

$$\sum_{j|T_j=0, T_i=1} 1\{\|X_j - X_i\| \leq \|X_i - X_i\|\} = M$$

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,  $l_m(i)$  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  $i$ .



Usando este estimador do contrafactual do indivíduo  $i$ , o estimador simples de pareamento é:

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

no qual  $N_1$  é o número de indivíduos no grupo de tratamento.

O estimador de pareamento é muito sensível à escolha de variáveis presentes em  $X$ . Se o vetor  $X$  incluir um número grande de variáveis explicativas, a hipótese de sobreposição (H2) pode não ser verdadeira, pois a região de  $X$  no qual precisamos de sobreposição das características entre o grupo de controle e de tratamento é muito grande. Se incluirmos um número limitado de variáveis no vetor  $X$ , a hipótese de seleção nos observáveis (H1) pode ser violada. O vetor  $X$  deve incluir as variáveis observáveis que influenciaram o indivíduo no momento da decisão de participar, e que são simultaneamente correlacionadas com os resultados potenciais dos indivíduos. Este conjunto de variáveis garante 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  $X$ . Neste caso, o pareamento estará olhando para outro parâmetro, que seria o efeito médio do tratamento sobre os tratados naquela região de suporte comum<sup>38</sup>.

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, comparamos indivíduos semelhantes, exceto pelo fato de que um recebeu o tratamento, enquanto o outro não sofreu influência do programa. Para verificar se através do pareamento estamos de fato comparando indivíduos que possuem um vetor  $X$  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 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.

---

<sup>38</sup> Ver Crump, Hotz, Imbens e Mitnik (2009, 2008).

## 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  $X$ . Quando acrescentamos variáveis ao vetor  $X$ , a dimensão da região de  $X$  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  $X$ , podemos parear os indivíduos usando uma função de  $X$  que resume toda a informação contida neste vetor. Esta função é a probabilidade de receber o tratamento dado o conjunto de características  $X$ , e é denominada de escore de propensão. De forma formal, o escore de propensão é definido como:

$$P(X) = \Pr[T = 1|X]$$

Como o escore de propensão contém a mesma informação contida no vetor de características observáveis  $X$ , 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<sup>39</sup>. De modo mais formal,

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

Este resultado nos diz que, se as hipóteses 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 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:

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

Em que  $\beta$  é vetor de parâmetros que será estimado num primeiro estágio. Seja  $\hat{\beta}$  o estimador do parâmetro  $\beta$ . Neste caso, o escore de propensão é estimado como:

<sup>39</sup> Para uma prova formal deste resultado, ver Rosenbaum e Rubin (1983).

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

Como no caso do pareamento baseado no  $X$ , o pareamento baseado no escore de propensão também irá depender de uma métrica pré-determinada, 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  $i$  para estimar qual seria o resultado do indivíduo  $i$  caso ele não recebesse tratamento<sup>40 41</sup>.

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:

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

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 (5.2.1).

É importante ressaltar que as propriedades do estimador de pareamento baseado no escore de propensão também dependem da escolha do vetor  $X$ . É necessário escolher o vetor  $X$  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, à forma funcional que vai ser escolhida na

<sup>40</sup> Outros estimadores de pareamento baseado no escore de propensão são o pareamento baseado em Kernel e o pareamento baseado em polinômio local. Neste caso, para construir o contrafactual para o indivíduo no grupo de tratamento usam-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 um peso diferente ao calcular o contrafactual do indivíduo  $i$ .

<sup>41</sup> Um exemplo é usar a mesma medida de distância que usamos no caso do pareamento baseado no vetor  $X$ ,

$$\sum_{j|T_j=0, T_i=1} 1\{|\hat{P}(X_j) - \hat{P}(X_i)| \leq |\hat{P}(X_h) - \hat{P}(X_i)|\} = M$$

Neste caso,  $h \equiv h_m(i)$  é o índice do indivíduo no grupo de não-tratados cujo escore de propensão está a distância  $m$  do escore de propensão do indivíduo tratado  $i$ .

estimação do escore de propensão. É necessário que esta forma funcional seja flexível o suficiente para que a hipótese de sobreposição (H2) se verifique. Rosenbaum e Rubin (1983) propuseram uma forma simples de buscar evidência se estamos 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 destas 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 estas variáveis observáveis. Nesse caso, provavelmente o escore de propensão não está sendo flexível o suficiente para garantir que a distribuição dos observáveis é 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 reponderação.

### 5.3 Regressão Linear

O modelo de regressão linear também pode ser visto como um estimador de pareamento<sup>42</sup>. 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:

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

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

Para estimarmos o efeito médio do tratamento sobre os tratados, precisamos conhecer o modelo do resultado potencial para  $Y_i(0)$ . Sob as hipóteses de seleção nos observáveis e exogeneidade de  $X$ , i.e.,  $E[\varepsilon_{0i}|X_i] = 0$ , podemos estimar o efeito médio do tratamento sobre os tratados usando o estimador de mínimos quadrados ordinários da regressão de  $Y$  em  $X$ .

Neste caso, iremos estimar uma regressão linear de  $Y$  em  $X$  para a amostra de indivíduos que não foram tratados. Os coeficientes estimados nesta 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. Neste caso, o estimador do efeito médio do tratamento sobre os tratados seria:

---

<sup>42</sup> Rubin (1977) usa o método de regressão linear para estimar o efeito médio do tratamento.

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

em que  $\hat{\mu}_0(X_i)$  é o valor previsto para o grupo de indivíduos no grupo de tratamento usando os coeficientes estimador pela regressão linear simples no grupo de controle <sup>43</sup>.

Os estimadores baseados em regressão dependem fortemente da extrapolação dos valores  $Y_i(0)$  para valores fora do intervalo dos usados na regressão. Os coeficientes estimados nesta regressão são usados para encontrar os valores preditos do resultado para os indivíduos tratados caso eles não recebessem tratamento. Para esta extrapolação, necessitamos de sobreposição na região de  $X$  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:

$$\hat{\mu}_0(X_i) = \hat{\alpha}_{00} + \bar{X}_1' \hat{\alpha}_{10}$$

em que  $\hat{\alpha}_{00}$  é o intercepto da regressão usando somente os indivíduos do grupo de controle, e  $\hat{\alpha}_{10}$  é o coeficiente relacionado a  $X$  estimado na mesma regressão. Usando a fórmula do estimador do intercepto da regressão,

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

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 desta regressão. Podemos usar sempre uma forma funcional flexível para a regressão, mas não teremos garantias de que este modelo funciona bem nas regiões onde não há sobreposição da região de  $X$  no grupo de controle e no grupo de tratamento.

Um método alternativo para obter o efeito médio do tratamento sobre os tratados através de uma regressão linear é usar o escore de propensão como uma variável explicativa<sup>44</sup>. Neste caso, iremos estimar uma regressão linear simples de  $Y$  em  $P(X)$  para a sub-amostra de indivíduos não-tratados, e usar os valores

<sup>43</sup> Para estimar o efeito médio do tratamento, precisamos também estimar a regressão linear simples para os indivíduos tratados, e usar os resultados desta regressão para prever qual seria o resultado para os indivíduos no grupo de controle caso eles recebessem o tratamento. Neste caso, o estimador do efeito médio do tratamento seria:  $EMPR = \frac{1}{N} \sum_{i=1}^N T_i (Y_i - \hat{\mu}_0(X_i)) + (1 - T_i) (Y_i - \hat{\mu}_1(X_i))$  no qual  $\hat{\mu}_1(X_i)$  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.

<sup>44</sup> Este método foi sugerido por Rosenbaum e Rubin (1983).

estimados nesta regressão para prever os valores dos resultados dos indivíduos tratados caso eles não recebessem tratamento.

## 5.4 Reponderação

O escore de propensão também pode ser usado como um peso <sup>45</sup> que, ao balancear os indivíduos no grupo de controle, torna este grupo semelhante na média ao grupo de indivíduos tratados. Neste caso, ponderamos cada unidade no grupo de controle pela probabilidade de não 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. Usando a Lei de Bayes, podemos definir uma relação entre a distribuição de  $X$  na população de tratados e a distribuição de  $X$  na população de não-tratados. Pela Lei de Bayes, a distribuição de  $X$  condicional a  $D=1$  pode ser escrita como

$$f(X|D = 1) = \frac{f(X, D = 1)}{\Pr(D = 1)} = \frac{\Pr(D = 1|X) \cdot f(X)}{\Pr(D = 1)}$$

E, de forma análoga, podemos escrever a distribuição condicional de  $X$  em  $D=0$  como,

$$f(X|D = 0) = \frac{f(X, D = 0)}{\Pr(D = 0)} = \frac{\Pr(D = 0|X) \cdot f(X)}{\Pr(D = 0)}$$

e dividindo as duas distribuições condicionais, temos

$$\frac{f(X|D = 1)}{f(X|D = 0)} = \frac{\Pr(D = 1|X) \cdot \Pr(D = 0)}{\Pr(D = 0|X) \cdot \Pr(D = 1)}$$

Usando a definição do escore de propensão e definindo que  $\Pr(D = 1) \equiv Q_0$ , podemos reescrever a relação entre as distribuições condicionais de  $X$ ,

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

Usando esta relação e a hipótese de seleção nos observáveis, podemos escrever o parâmetro de interesse (EMPT) usando o escore de propensão como peso,

$$D = E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 1] = E\left[\frac{T_i}{Q_0} Y_i\right] - E\left[\frac{1 - T_i}{Q_0} \cdot \frac{P(X)}{1 - P(X)} Y_i\right]$$

Como no caso do estimador de pareamento, estimamos o escore de propen-

<sup>45</sup> Para uma revisão do método de reponderação, ver Imbens (2004) e Imbens e Wooldridge (2009).

são usando o modelo paramétrico, como o logit. Dado que  $T$  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 reponderação de  $D$  é

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

Este estimador se baseia somente na estimação do escore de propensão, e uma das limitações deste estimador é ser bastante sensível à especificação do escore de propensão. Quando o valor do escore de propensão for próximo de um, este estimador pode assumir valores muito altos.

## 5.5 Combinação de Métodos

Alguns estimadores surgiram da combinação de dois métodos descritos nas seções anteriores. Estes 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  $Y(0)$  com  $X$ . Nesta seção, descrevemos três combinações: regressão e pareamento, pareamento e diferenças em diferenças, e regressão e reponderação.

### 5.5.1 Pareamento e Regressão Linear

Abadie e Imbens (2002) mostram que, à medida que aumentamos a dimensão do vetor  $X$ , podemos estar diminuindo a variância do estimador de pareamento baseado em  $X$ , 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,  $\hat{D}_{iM} = Y_i(1) - \hat{Y}_i(0)$ , e o efeito médio do tratamento ( $\hat{D}_M$ ). A variância do estimador de pareamento é uma função da soma do quadrado dos resíduos. Ao aumentarmos o número de variáveis incluídas em  $X$ , diminuímos este resíduo e conseqüentemente 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  $X$ , o viés desapareceria. Uma das maneiras de corrigir o viés do estimador de pareamento é ajustar pela diferença

nos valores das variáveis explicativas  $X$ , usando o modelo de regressão linear. Neste caso, usamos o mesmo conjunto  $L_M(i)$  que engloba os  $M$  indivíduos no grupo de controle cujos vetores  $X$  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  $Y$  com  $X$  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  $Y$  para os indivíduos no grupo de controle e também para os indivíduos no grupo de tratamento, e usamos estes valores preditos para ajustar o estimador de pareamento. Neste 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:

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

## 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<sup>46</sup>. Esta 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  $X$ , não pode 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  $X$ , como vimos no capítulo 4.

Para aplicar esta 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. Este método usa basicamente a ideia do método de diferenças em diferenças, isto é, ele compara a diferença do re-

<sup>46</sup> Este método foi proposto por Heckman, Ichimura e Todd (1997,1998).



sultado 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 esta diferença da diferença ao efeito do tratamento. Formalmente, o método que combina pareamento e o método de diferenças em diferenças estima o seguinte parâmetro populacional,

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

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 este parâmetro corresponda ao efeito médio do tratamento sobre os tratados, precisamos da seguinte hipótese:

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

Esta 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, esta hipótese exige informação em pelo menos um período anterior à intervenção. Se a evolução dos resultados entre os períodos pré e pós-programa é a mesma no grupo de controle e no grupo de tratamento, esta hipótese é válida. O vetor  $X$  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  $X$  para o grupo de controle e o grupo de tratamento. No caso de dados longitudinais, precisamos da hipótese H2. Se tivermos suporte comum, podemos estimar o efeito médio do tratamento sobre os tratados fazendo um pareamento baseado no escore de propensão<sup>47</sup> e calculando a diferença das diferenças. 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:

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

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

<sup>47</sup> Podemos também fazer o pareamento baseado no vetor  $X$ .

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

Para implementar esse estimador na prática, primeiro calculamos para cada indivíduo no grupo de tratamento qual seria a diferença de resultados entre os períodos  $t_1$  e  $t_0$  caso ele não fosse tratado. Para isso, estimamos o escore de propensão para todos os indivíduos, e selecionamos 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. Em um segundo passo, para cada indivíduo no grupo de tratamento, calculamos a diferença entre a sua diferença de resultado entre os períodos  $t_1$  e  $t_0$  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 destas 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 tratamentos. A variável dependente  $T$  é igual a 1 se o indivíduo recebeu o tratamento, e 0 caso contrário, e o vetor  $X$  engloba as informações que os indivíduos tinham ao decidir participar ou não no programa, isto é, características nos períodos anteriores ao programa. Neste vetor  $X$ , 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.

### 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  $X$  como variáveis explicativas. Nesta primeira abordagem, não precisamos estimar o escore de propensão. Já o método de reponderação se baseia somente na estimação do escore de propensão. Neste segundo método, não precisamos estimar a função de regressão que relaciona os resultados potenciais com  $X$ . Sob algumas hipóteses, ambos os métodos resultam em estimadores consistentes e eficientes para o efeito médio do tratamento. No entanto, enquanto o comportamento do estimador de regressão é sensível à especificação usada para o modelo que relaciona os resultados potenciais com  $X$ , o estimador de reponderação é sensível ao modelo usado para o escore de propensão. Robins e 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. Este estimador para o efeito médio do tratamento tem a propriedade de ser duplamente robusto, isto é, se o modelo paramétrico para o escore de propensão for corretamente especificado ou se o modelo paramé-

trico para a regressão linear for corretamente especificado, o estimador é consistente para o efeito médio do tratamento sobre os tratados.

Esta combinação de métodos também pode ser usada para estimar o efeito médio do tratamento sobre os tratados. Uma maneira de estimar o efeito médio do tratamento sobre os tratados proposta por Hirano e Imbens (2001) 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  $X$  na subamostra de indivíduos tratados. Nesse caso,  $\tau$  representa o efeito médio do tratamento sobre os tratados.

Mas, essa regressão linear pode ser combinada com o método de reponderação, no 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:

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

no qual  $\hat{P}(X_i)$  representa o estimador paramétrico do escore de propensão, obtido por uma regressão logit ou probit<sup>48</sup>.

Para obter o efeito médio do tratamento usamos também uma regressão ponderada com duas modificações em relação ao modelo anterior. Primeiro, 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  $X$  na amostra total, incluindo os indivíduos no grupo de tratamento. Segundo, temos que não só reponderar os indivíduos no grupo de controle para melhor representar os indivíduos tratados, mas também os indivíduos no grupo de tratamento para melhor representar os indivíduos não-tratados. Neste 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. Neste caso, usamos a seguinte função peso:

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

<sup>48</sup> Este 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.

## 5.6 Exemplos

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. Este programa abrange a população em extrema pobreza e são elegíveis as famílias com renda per capita inferior a R\$ 50. Estas famílias podem receber a transferência sem nenhuma contrapartida. Além disso, existe um valor adicional para cada criança da família que for para a escola. Este programa difere do antigo Bolsa Escola, no qual 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, o que poderia reduzir as horas trabalhadas pelas crianças, o que, por sua vez, pode 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 tratado 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. Neste 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<sup>49</sup>.

A tabela 5.3.1 apresenta os resultados da estimação deste 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 ponderação 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

<sup>49</sup> No artigo em questão, os autores usam como variável dependente o número de horas trabalhadas ao ano. Neste exemplo, usamos como variável dependente o número de horas de trabalho remuneradas ao ano.

reponderação (coluna 4). Neste último caso, encontramos um efeito positivo sobre o número de horas trabalhadas remuneradas. Este resultado contradiz a teoria de que programas de transferências incondicionais não levariam os indivíduos a trabalharem 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 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 têm um efeito negativo e significativo no número de horas remuneradas das mulheres beneficiárias, sendo este efeito maior em valor absoluto para no caso do programa de Renda Mínima. Este resultado indica que, no caso das mulheres, o programa leva a uma 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. 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 tratamento, enquanto as barras vermelhas 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.

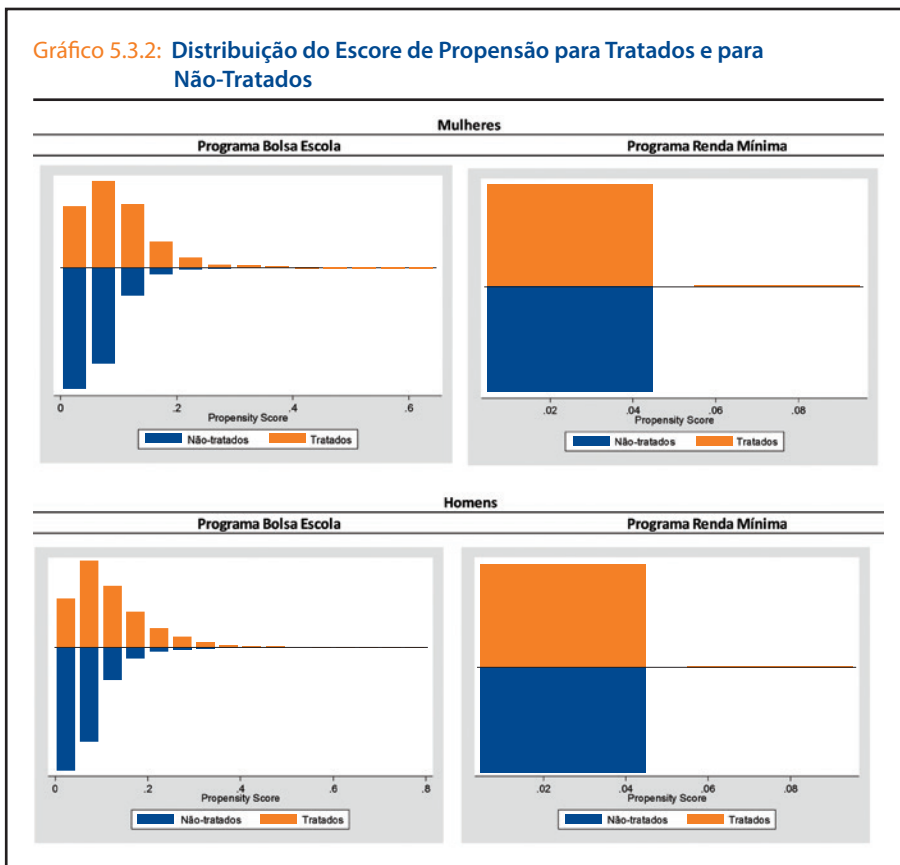
**Tabela 5.3.1: Métodos usando o Escore de Propensão**

<b>Horas de Trabalho por Ano- Homens</b>				
	Pareamento (um a um)	Pareamento (10 vizinhos)	Regressão	Regressão com Reponderação
Recebe o Bolsa Escola	-0,044** (0,019)	-0,0232 (0,015)	-0,029** (0,012)	-0,027*** (0,007)
Recebe a Renda Mínima	0,054 (0,062)	0,067 (0,045)	0,043 (0,040)	0,070** (0,032)
<b>Horas de Trabalho por Ano-Mulheres</b>				
	Pareamento (um a um)	Pareamento (10 vizinhos)	Regressão	Regressão com Reponderação
Recebe o Bolsa Escola	-0,015 (0,029)	-0,028 (0,022)	-0,029 (0,017)	-0,029*** (0,010)
Recebe a Renda Mínima	-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 5.3.2: Distribuição do Escore de Propensão para Tratados e para Não-Tratados



## 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.

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

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

**d)** 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 estas hipóteses no contexto do exemplo, e discuta a validade destas hipóteses.

**d)** 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 deste novo método.

**3)** Para estimar o efeito de um programa de treinamento no mercado de trabalho sobre salário, você obtém uma base de dados que tem o salário dos indivíduos tratados e não-tratados e um vetor com várias características do background socioeconômico dos indivíduos e também da sua atividade no mercado de trabalho. Em um primeiro momento, você decide usar estas 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 você decide fazer o pareamento baseado no escore de propensão.

**a)** Quais hipóteses precisam ser satisfeitas para que este estimador baseado no escore de propensão represente o efeito médio do tratamento sobre os beneficiários do programa?

**b)** Você poderia obter este efeito médio do tratamento sobre os tratados usando uma 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 esta crítica? Proponha um método que seja menos sensível a 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 por exemplo, sexo, raça, educação, renda, se possui plano de saúde, tipo de doença que possui, peso, etc. Eles estimam o escore de propensão baseado em um modelo logístico.

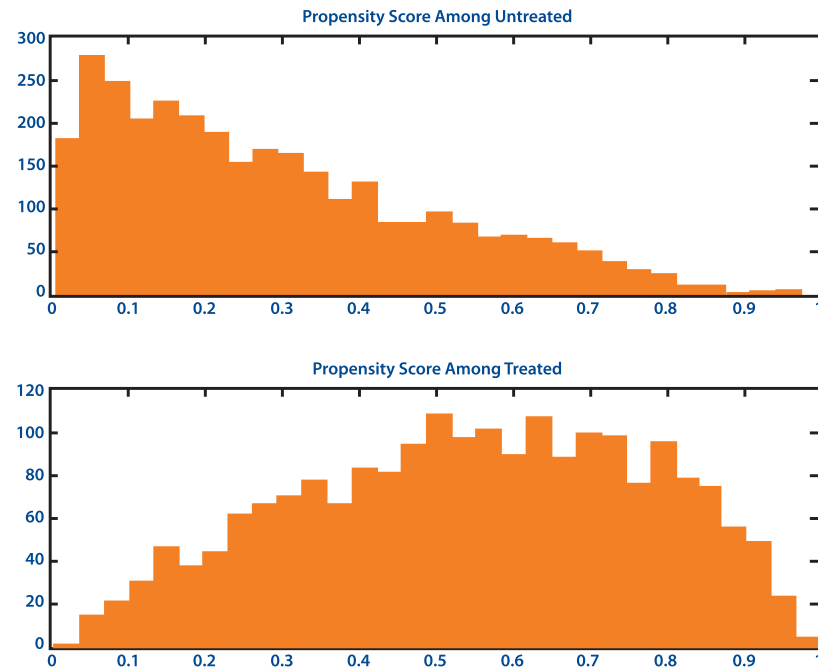
**a)** Antes de estimar o efeito médio do tratamento sobre os tratados, eles apresentam uma tabela que contém um teste de diferença de médias para as características observáveis. Eles obtêm resultados para este teste antes e depois da ponderação pelo escore de propensão. Esta tabela encontra-se abaixo.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

Variáveis	Não-Tratado	Tratado	t-teste	Pareando pelo Escore de Propensão		
				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

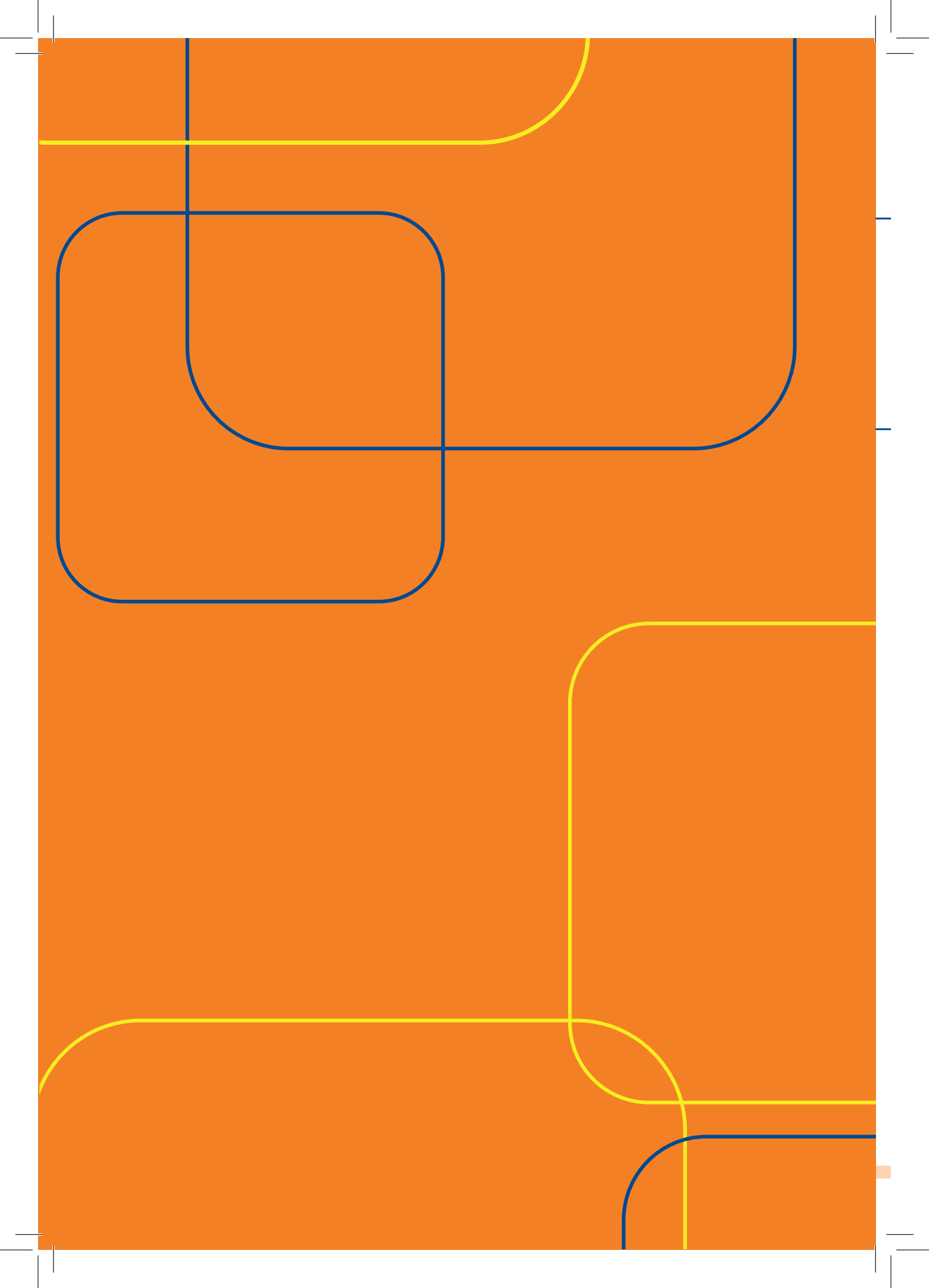
Nesta tabela, renda 1, renda 2 e renda 3 representam 3 baixas de renda faixa, média e alta, respectivamente.



**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. Usando 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 usando o método de reponderação, eles encontram um efeito médio de  $-0,060$  (erro-padrão de  $0,018$ ). Combinando os métodos de regressão e reponderação, eles encontram um efeito médio de  $-0,062$  (erro-padrão de  $0,015$ ). Interprete estas evidências, dando ênfase às hipóteses em que se baseia cada estimador. Qual seria a conclusão sobre o efeito deste procedimento na taxa de sobrevivência?



## PARTE 2

# Métodos Avançados de Avaliação de Impacto

## CAPÍTULO 6

# Variáveis Instrumentais

Cristine Campos Xavier Pinto

Ao contrário dos métodos apresentados no capítulo 5, o método de variável instrumental baseia-se no caso de seleção em variáveis não observáveis. 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  $X$ , que afetam simultaneamente a decisão de participar no programa e os resultados potenciais, isto é, quando:

$$\Pr[T = 1|Y(1), Y(0), X] \neq \Pr[T = 1|X] = p(X)$$

Neste caso, precisamos de uma variável exógena  $Z$  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 efeitos homogêneos do tratamento, estamos pensando no seguinte sistema de equações:

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

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

no qual  $T_i$  é igual a 1 se o indivíduo recebeu tratamento, e 0 se o indivíduo é não-tratado. Além disso, assumimos que  $Cov(Z_i, \varepsilon_i) = 0$  e  $Cov(\vartheta_i, \varepsilon_i) \neq 0$ , ou seja, que não há relação linear entre o instrumento e o termo aleatório da equação principal, mas os termos aleatórios não observáveis das duas equações são correlacionados. Neste sistema, os fatores não observáveis que afetam a decisão de participar no programa estão correlacionados com os fatores não observáveis que afetam o resultado de interesse, e precisamos de um instrumento  $Z$  que permita captar uma variação exógena na decisão de participar do programa, e que ao mesmo tempo não está relacionado de forma direta com o resultado potencial.

Este modelo pode ser estimado por mínimos quadrados em dois estágios. Neste caso, em um primeiro estágio, estimamos um modelo de probabilidade linear que relaciona  $T_i$  com  $Z_i$  e obtemos o valor predito:

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

Este valor predito representa uma variação exógena na decisão de participar ou não do programa, que não está relacionada a nenhum outro fator que possa influenciar o resultado de interesse.

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

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

No sistema de equação acima, a variável  $Z$  não afeta diretamente  $Y$ . Ela só afeta o resultado de interesse através 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 relacionado a como a participação no programa é afetada pelo instrumento  $Z$ . No caso de tratamento homogêneo, o efeito médio do tratamento (EMP) é igual ao efeito médio do tratamento sobre os tratados (EMPT),  $EMP = EMPT = \beta$ .

Se os indivíduos souberem que os ganhos de participação podem diferir para certos grupos, eles irão levar em consideração esta informação na hora de decidir se participam ou não do programa. Neste caso, tanto seus ganhos individuais  $\beta_i$  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 diferente para cada indivíduo, dependendo do seu ganho com o tratamento,  $\beta_i$ . Neste 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 tratamen-

to sobre os tratados.

Imbens e Angrist (1994) mostraram que quando os efeitos do tratamento são heterogêneos, o arcabouço de variável instrumental permite identificar um efeito médio de tratamento local, isto é, um efeito médio do tratamento para uma subpopulação específica. Neste caso,  $\beta$  será o efeito médio do tratamento para aqueles indivíduos cuja variação em  $Z$  provoca uma variação no status de participação sem afetar os resultados potenciais. Para estes indivíduos, a diferença na média dos resultados potenciais do grupo de tratados e do grupo de não-tratados ocasionada por uma variação em  $Z$  se dá exclusivamente através do efeito de  $Z$  na taxa de participação do programa.

Este efeito médio do tratamento para a subpopulação de indivíduos induzidos ao tratamento por variações no instrumento é chamado de efeito médio local do tratamento (LATE)<sup>50</sup> Iremos entender este efeito dentro do arcabouço de resultados potenciais. Neste arcabouço, definimos como “compliers” os indivíduos que são induzidos pelo instrumento a terem 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:

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

Suponha que, além de observarmos  $(Y_i, T_i)$  para cada indivíduo, observamos também uma variável instrumental binária  $Z_i$ . Neste 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<sup>51</sup>. O valor realizado para  $T_i$  pode ser definido como uma função do instrumento:

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

Neste arcabouço de resultados e tratamentos potenciais, podemos pensar no resultado potencial como uma função de  $Z$  e  $T$ ,  $Y_i(Z, T)$  que seria o resultado

<sup>50</sup> Esta sigla LATE vem do nome em inglês, “Local Average Treatment Effect”.

<sup>51</sup>  $T_i(0)$  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.

potencial que seria 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  $Z$  igual a 0 e decidiu não participar do programa.

Como no caso clássico de variável instrumental, precisamos que o instrumento  $Z$  só afete o resultado de interesse através do seu efeito na taxa de participação. Para garantir que isso aconteça em um arcabouço com efeitos heterôgenos, precisamos que  $Z$  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 através da variação que provoca na taxa de participação, ou seja, precisamos de uma hipótese conhecida como restrição de exclusão. Esta hipótese garante que o instrumento só afete  $Y$  através de um canal conhecido. De um modo mais formal, esta 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:

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

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

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 dependo do resultado recebido pelo tratamento. Podemos pensar em quatro tipos de comportamento induzidos pelo instrumento. Suponha que os indivíduos que recebem valores de  $Z$  igual a 1 são convidados a participar do programa, enquanto os indivíduos com valores de  $Z$  igual a 0 não são convidados a participar. Podemos pensar, por exemplo, em um progra-

ma 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 um valor determinado  $c$ , ele é convidado a se matricular no curso. Caso ele receba uma nota inferior a  $c$ , ele não recebe o convite. Neste 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 independente da sua nota no teste de proficiência, isto é, independente do valor do instrumento. Estes alunos são chamados de “always-takers”, pois eles sempre aceitariam participar do programa independentemente do valor do instrumento. Existe outro grupo de indivíduos que não gostaria de se matricular no curso técnico, independentemente do valor do instrumento. Este grupo de indivíduos é conhecido como “never-takers” e são aqueles que nunca aceitariam participar, independentemente da sua nota no teste de proficiência. 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, não irão se matricular. Estes indivíduos são chamados de “compliers”. Finalmente, teremos os indivíduos que terão comportamento oposto àquele induzido pelo instrumento. Neste 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. Este último grupo é conhecido como “defiers”. Podemos representar estes quatro grupos em uma tabela<sup>52</sup> de acordo com o comportamento de cada um a variações do instrumento,

	Nascidos no	
	Primeiro Trimestre	Quarto Trimestre
<b>log(salário)</b>	5,892	5,905
<b>Anos de Estudo</b>	12,688	12,839

Por exemplo, a primeira célula desta tabela representa os indivíduos cujo tratamento potencial, independentemente da realização do instrumento (0 ou 1), é sempre igual a 0, ou seja, eles nunca participam. Estes são os never-takers. Baseados nas variáveis que observamos para cada indivíduo ( $Y_i$ ,  $T_i$ ,  $Z_i$ ), não podemos classificar estes 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:

<sup>52</sup> Estas tabelas foram retiradas de Imbens e Wooldridge (2009)

	<b>Zi = 0</b>	<b>Zi = 1</b>
<b>Ti = 0</b>	complier/never-taker	never-taker/defier
<b>Ti = 1</b>	always-taker/defier	complier/always-taker

Esta tabela indica que, ao observarmos para o indivíduo  $i$  o par  $(Z, T) = (0, 1)$ , só sabemos que este indivíduo não participaria do curso técnico se a sua nota for abaixo do ponto de corte  $c$ . Este 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 uma nota superior a  $c$  no teste de proficiência. 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:**  $T_i(1) \geq T_i(0)$  para todo  $i$  (H3)

Esta 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 uma nota no teste de proficiência acima de  $c$ , não iriam se matricular no curso, e se recebessem uma nota no teste acima de  $c$  iriam se matricular no curso. Com a hipótese de monotonicidade, a informação que temos nos dados observáveis aumenta e pode ser representada pela tabela abaixo:

	<b>Zi = 0</b>	<b>Zi = 1</b>
<b>Ti = 0</b>	complier/never-taker	never-taker
<b>Ti = 1</b>	always-taker	complier/always-taker

Como pode ser visto pela tabela, 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  $Y$  em  $Z$  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:

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

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

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

Esta razão é o parâmetro estimado pelo método de variável instrumental quando o instrumento é uma variável binária e os efeitos são heterogêneos. 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”.

Suponha que observamos um vetor de características adicionais dos indivíduos ( $X$ ) que podem ser usados nos modelos para o resultado de interesse e para a participação no tratamento. Neste caso, este 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 que têm um efeito direto no resultado de interesse. Se a variável instrumental  $Z_i$  for de fato alocada de forma aleatória entre os indivíduos, ele seria independente das demais variáveis explicativas do modelo ( $X_i$ ) e a inclusão do vetor  $X$  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, neste 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\*)

Esta 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)

Neste 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. Em um primeiro momento 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 em um segundo estágio, usamos este valor predito como variável explicativa no modelo linear para  $Y$ ,

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

## 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. O LATE 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, este grupo de indivíduos não é representativo da população como um todo. Dada esta limitação, podemos perguntar se este efeito local identificado pelo LATE é interessante para uma análise de política. A resposta a esta 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 esta 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 contínua que simplesmente afeta a taxa de participação e não tem efeito direto no resultado, 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.

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 queremos testar o efeito de um exame novo que foi criado para detectar câncer de mama, e para isso convidamos 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 fazer o exame podem ser na média diferentes das mulheres que se recusaram a fazê-lo. Por exemplo, as mulheres que aceitaram os convites são geralmente mais bem instruídas, se preocupam 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 neste caso estaríamos misturando o efeito do programa com o efeito de uma vida mais saudável. Neste caso, 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. Neste caso, LATE é o efeito médio sobre as mulheres que de fato receberam o tratamento, isto é, o efeito médio do tratamento sobre os tratados.

## 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 este resultado dentro de um instrumental simples no qual consideremos 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 destas variáveis binárias. Para estimar o efeito médio do tratamento para a população de “compliers”, usamos o estimador de MQO em dois estágios. Em um primeiro estágio, estimamos o modelo linear que relaciona o tratamento com as duas variáveis instrumentais:

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

Em um segundo estágio, usamos este valor predito como a variável explicativa no modelo linear para o resultado  $Y$  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 este 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:

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

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

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

Os pesos de cada um dos LATES depende de quão 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 este resultado se estende para o caso de múltiplos instrumentos discretos.

### 6.3 O Caso de Instrumentos Contínuos

Nesta seção, consideremos a estimação do impacto do tratamento sobre toda a distribuição de um instrumento contínuo. Neste caso, o instrumento é contínuo e a probabilidade de participação condicional a  $Z$  também é uma função contínua, isto é,  $p(z) = E[T_i | Z_i = z]$  é uma função contínua em  $z$ . Neste caso, usamos mudanças infinitesimais na taxa de participação para medir o efeito médio local do tratamento (LATE). Heckman (1990), Vytlacil (2002), Heckman e Vytlacil (2001), Carneiro, Heckman e Vytlacil (2005) usam um modelo de seleção para interpretar este efeito marginal. A seleção é feita através de uma função de  $z$  e de um componente não-observável,  $\epsilon$ :

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

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  $\epsilon_i$ . Se o fator não-observável do indivíduo  $i$  é menor que o do indivíduo  $j$  ( $\epsilon_i < \epsilon_j$ ), então  $T_i(z) \geq T_j(z)$  para todo  $z$ . Dado este ordenamento baseado no não-observável, podemos definir o efeito marginal do tratamento condicional em  $\epsilon$ :

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

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

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

Para entender este parâmetro como o limite do LATE, voltamos ao caso mais simples em que  $g(z_i)$  é uma função linear em  $z_i$ . Neste 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  $z^*$ ,

$$\beta^{MTE}(z) = \frac{dE[Y|Z=z]}{d \Pr[T=1|Z=z]} \quad (6.3.5)$$

Este parâmetro EMT usa um instrumento contínuo na tentativa de reconstruir a distribuição das probabilidades de participação, e usar esta distribuição para estimar o efeito na variação marginal de  $z$ .

## 6.4 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, este 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 através do efeito no tamanho da família. Neste 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. 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. Usando os dados do Censo Brasileiro de 1991, os autores encontram que

o tamanho da família apresenta um efeito significativo na progressão escolar das meninas, sendo este efeito negativo. Os autores apresentam evidências de que, quanto maior o tamanho da família, menor será a probabilidade de a criança progredir na escola, mantendo todas as demais características da família constantes. Como salientado pelos autores, este 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. Nestas políticas o efeito 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**

<b>Variável Dependente: Número de crianças na família com 10 a 15 anos</b>		
	<b>Meninos</b>	<b>Meninas</b>
<b>Presença de gêmeos</b>	0,852***	0,862***
	(0,028)	(0,028)
<b>Teste de F</b>	1070,76***	1137,98***
<b>N</b>	139.214	148.106

Fonte: Ponczek e Souza (2011)

**Tabela 6.4.2: Resultados do Segundo Estágio**

	<b>Meninos</b>	<b>Meninas</b>
<b>Participação na Força de Trabalho</b>		
<b>Tamanho da Família</b>	-0,029	0,019
	(0,021)	(0,027)
<b>N</b>	38.928	28.609
<b>Frequência à Escola</b>		
<b>Tamanho da Família</b>	0,013	-0,016
	(0,021)	(0,026)
<b>N</b>	38.928	28.609
<b>Progressão na Escola</b>		
<b>Tamanho da Família</b>	-0,005	-0,044***
	(0,010)	(0,011)
<b>N</b>	38.928	28.609

Fonte: Ponczek e Souza (2011)

## 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:

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

em que  $P_j$  é uma variável binária que indica se na região  $j$  o programa foi muito intenso. Esta 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  $\varepsilon_{ijk}$  representa os fatores não observáveis. Este 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  $\gamma$  representa? Quem seriam os "compliers" neste 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  $\rho$ .

**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 que esperar um ano a mais para entrar na escola que as crianças que nascem no primeiro trimestre de nascimento. Esta 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, 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
<b>log(salário)</b>	5,892	5,905
<b>Anos de Estudo</b>	12,688	12,839

Usando os 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 este coeficiente como o efeito médio de um tratamento para a subpopulação de “compliers”. Quem são os “compliers” neste caso?

**3)** Considere o seguinte modelo com efeitos heterogêneos, no qual  $T_i$  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 este 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,

$$T_i = T_{0i} + (T_{1i} - T_{0i})Z_i = \pi_0 + \pi_{1i}Z_i + \varepsilon_i$$

**a)** Mostre que, se temos efeitos heterogêneos do instrumento sobre o tratamento, mas temos um efeito homogêneo do tratamento  $\rho$ , o estimador do LATE é igual a  $\rho$  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, violações da hipótese de monotonicidade não invalidam a análise de um 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 desta hipótese adicional? Interprete este resultado e discuta como você iria estimar o efeito médio local do tratamento neste contexto.

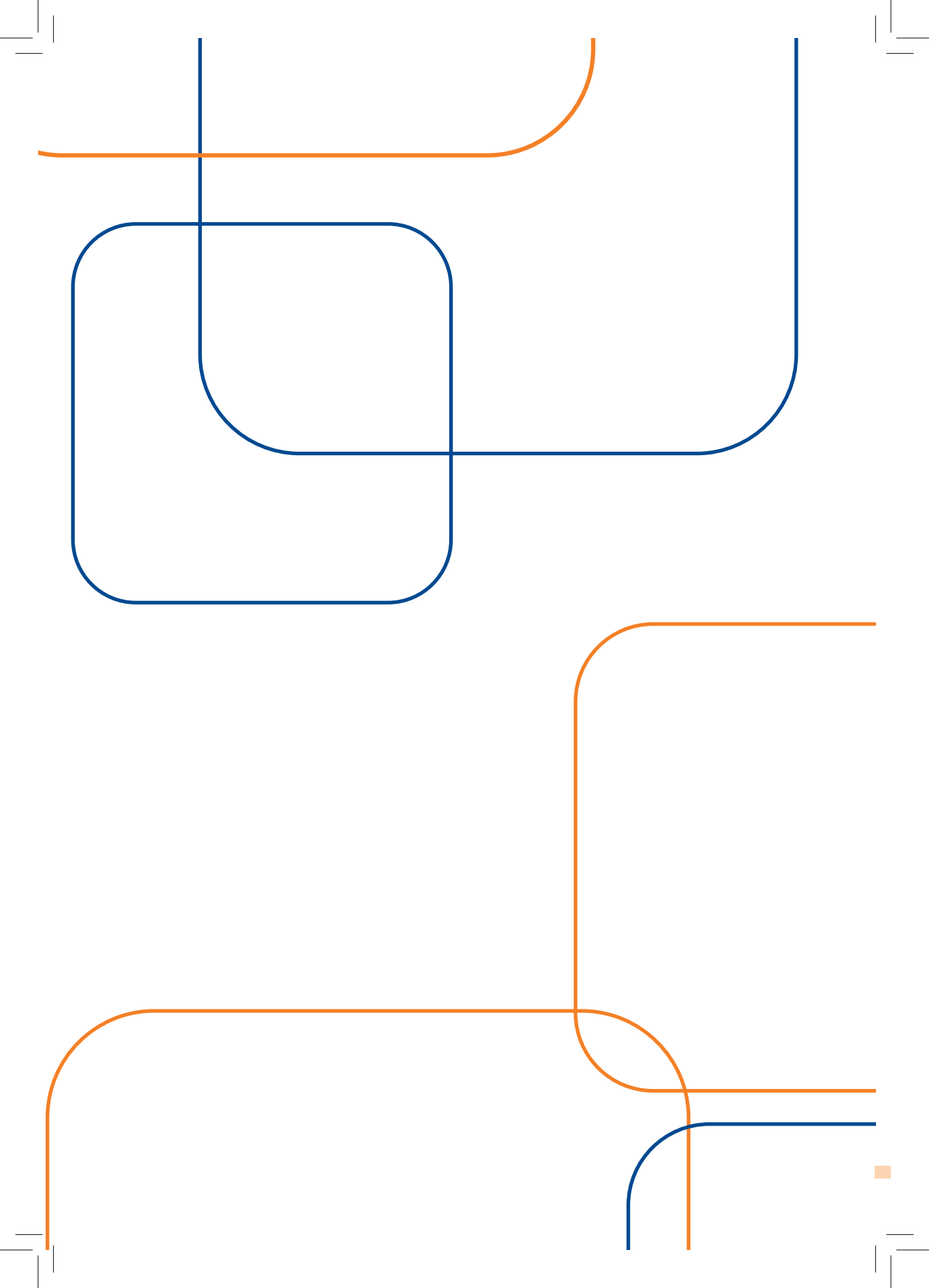
**4)** Um programa é criado pelo governo para promover o consumo de frutas e vegetais entre os alunos na 5ª série do ensino fundamental. Nas escolas que receberam o programa, os alunos foram convidados a participar de aulas que explicavam cada tipo de alimento e nas quais eles podiam experimentar as frutas e vegetais. Além disso, estes 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. Esta seleção foi realizada de forma aleatória. O governo gostaria de avaliar o efeito do programa sobre a saúde das crianças. Para isso, ele possui uma base de dados que contém algumas características socioeconômicas dos alunos, se eles estavam em uma escola tratada e se participaram do programa. Além disso, ele possui informações sobre a saúde do aluno, como o número de vezes que o aluno ficou gripada 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 desta avaliação? Você acha que os resultados obtidos para este estado se manteriam se esta política fosse implementada em outros estados?



## CAPÍTULO 7

# Regressão Descontínua

Cristine Campos Xavier Pinto

O método de regressão descontínua pode ser usado quando a probabilidade de receber tratamento muda de forma descontínua com uma variável, que vamos chamar de  $Z$ . Um dos primeiros trabalhos que usaram regressão descontínua foi Thristhlewaite e Cook (1960), que estimou o impacto de um prêmio por mérito no desempenho acadêmico dos alunos. Neste estudo, o prêmio por mérito era dado de acordo com as notas dos alunos. Os alunos que tinham notas acima de um certo valor recebiam o prêmio, enquanto os alunos que tinham notas abaixo deste 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 de que não receberam o prêmio, e seriam um bom grupo de controle. Neste estudo, a variável  $Z$ , cujo valor determina o recebimento do tratamento, é a nota do aluno.

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. Ao invés de assumirmos uma hipótese de independência, como no capítulo 5, 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 esta hipótese, qualquer descontinuidade que ocorra na função que relaciona  $Y$  a  $Z$  pode ser atribuída ao salto que ocorre na participação do tratamento em determinado valor de  $Z$ , e o tamanho deste salto pode ser interpretado como um efeito médio local do tratamento. A desvantagem do método de regressão descontínua é que ele estima um efeito médio do tratamento comparando apenas os indivíduos em torno deste ponto de corte. Se estes indivíduos forem muito diferentes do restante dos indivíduos na população de interesse, não podemos afirmar que este 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.

No instrumental de regressão descontínua, o recebimento do tratamento, denominado pela variável binária  $T$ , é uma função de  $Z$ . A descontinuidade pode ser *fuzzy* ou *sharp*. No caso *sharp*, a participação é uma função determinística de  $Z$ , isto é,  $T = 1$  se  $X \geq c$ , e  $T=0$  se  $X < c$ . No caso *fuzzy*, ocorre um salto na

probabilidade de participação no ponto em que  $Z$  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 em  $Z$  igual a  $c$  não é suficiente para identificarmos o efeito médio local do tratamento.

Se os indivíduos tem controle sobre a variável  $Z$  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 Thistlethwaite e Cook (1960), se os estudantes puderem determinar a sua nota no teste de proficiência de forma perfeita através do esforço, os alunos mais esforçados e determinados irão escolher notas acima de  $c$  e irão ganhar o prêmio, e estes alunos podem ser sistematicamente diferentes dos demais pelo menos no que se refere ao esforço. Neste 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 um outro cenário no qual os alunos, mesmos se esforçando mais, não têm garantia de que sua nota será maior que  $c$ . Neste novo cenário, parece mais razoável pensar que os alunos que ficaram logo abaixo do ponto de corte são semelhantes aos que ficaram 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  $c$ . Neste caso, temos uma variação exógena no tratamento em torno da descontinuidade, e usamos esta 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  $Z$  devem ter a mesma distribuição no grupo de indivíduos com valores  $Z$  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 através 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 através de uma análise gráfica. Por exemplo, uma representação gráfica da relação de  $Y$  e  $Z$  pode ser usada para nos dar uma ideia do tamanho da descontinuidade, da forma funcional adequada para modelar esta 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 é identificado dentro da estrutura de uma regressão descontínua, usando o arcabouço de resultados potenciais. O efeito médio lo-

cal no ponto de descontinuidade  $c$  pode ser definido como:

$$D(c) = E[Y_i(1)|Z = c] - E[Y_i(0)|Z = c] \quad (7.1)$$

Mas não observamos os resultados potenciais  $Y_i(1)$  e  $Y_i(0)$  para o mesmo indivíduo. Para cada indivíduo  $i$ , observamos:

$$Y_i = Y_i(0) + (Y_i(1) - Y_i(0)) \cdot T_i$$

Para os indivíduos com valores de  $Z$  logo acima do ponto de corte:

$$E[Y_i|Z_i = c + \varepsilon] = E[Y_i(0)|Z_i = c + \varepsilon] + E[(Y_i(1) - Y_i(0)) \cdot T_i|Z_i = c + \varepsilon]$$

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  $Y$  aproxima quando  $Z$  chega próximo de  $c$  para os indivíduos acima do ponto de corte, e  $Y^-$  como o valor que  $Y$  aproxima quando  $Z$  chega próximo de  $c$  para os indivíduos abaixo do ponto de corte. De modo formal, temos

$$Y^+ = \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c + \varepsilon] = \lim_{\varepsilon \downarrow 0} E[Y_i(0)|Z_i = c + \varepsilon] + \lim_{\varepsilon \downarrow 0} E[(Y_i(1) - Y_i(0)) \cdot T_i|Z_i = c + \varepsilon]$$

$$Y^- = \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c - \varepsilon] = \lim_{\varepsilon \downarrow 0} E[Y_i(0)|Z_i = c - \varepsilon] + \lim_{\varepsilon \downarrow 0} E[(Y_i(1) - Y_i(0)) \cdot T_i|Z_i = c - \varepsilon]$$

Sob a hipótese de que o resultado de interesse ( $Y$ ) é uma função contínua de  $Z$ ,

$$\lim_{\varepsilon \downarrow 0} E[Y_i(0)|Z_i = c + \varepsilon] = \lim_{\varepsilon \downarrow 0} E[Y_i(0)|Z_i = c - \varepsilon]$$

A hipótese de continuidade exige que as funções de regressões condicionais dos resultados do tratamento e do controle em  $Z$  sejam contínuas. De maneira mais formal,

H1:  $E[Y_i(1)|Z_i = z]$  e  $E[Y_i(0)|Z_i = z]$  são funções contínuas em  $z$ <sup>53</sup>

Se em torno do ponto  $Z=c$ , o status do tratamento foi determinado de forma aleatória, temos ignorabilidade local (H2), ou seja:

$$(Y_i(0), Y_i(1)) \perp T_i | Z_i = c$$

<sup>53</sup> 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  $F_{Y(1)|Z}(y|z)$  e em  $z$  para todo  $y$ .

Sob esta hipótese de ignorabilidade<sup>54</sup>,

$$E[Y_i(1) \cdot T_i | Z_i = c] = E[Y_i(1) | Z_i = c] \cdot E[T_i | Z_i = c] \quad (7.2)$$

$$E[Y_i(0) \cdot T_i | Z_i = c] = E[Y_i(0) | Z_i = c] \cdot E[T_i | Z_i = c] \quad (7.3)$$

Sob as hipóteses de ignorabilidade local e continuidade,

$$Y_i^+ - Y_i^- = \lim_{\varepsilon \downarrow 0} E[Y_i(1) | Z_i = c + \varepsilon] \lim_{\varepsilon \downarrow 0} E[T_i | Z_i = c + \varepsilon] \\ - \lim_{\varepsilon \downarrow 0} E[Y_i(1) | Z_i = c - \varepsilon] \lim_{\varepsilon \downarrow 0} E[T_i | Z_i = c - \varepsilon]$$

E podemos escrever o efeito médio local de tratamento como

$$D(c) = E[Y_i(1) | Z = c] - E[Y_i(0) | Z = c] = \frac{Y^+ - Y^-}{T^+ - T_i^-} \quad (7.4)$$

no qual  $T^+ = \lim_{\varepsilon \downarrow 0} E[T_i | Z_i = c + \varepsilon]$  e  $T^- = \lim_{\varepsilon \downarrow 0} E[T_i | Z_i = c - \varepsilon]$ .

A hipótese básica para a identificação do efeito médio local do tratamento é que a probabilidade de participação no tratamento é uma função de  $Z$ , e no ponto  $c$  esta probabilidade sofre um salto. Esta 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  $Z$ , de modo que qualquer descontinuidade em  $Y$  em torno de  $c$  pode ser atribuída somente ao tratamento.

## 7.1 Caso Sharp

Neste caso,  $T$  é uma função determinística de  $Z$ . Todos os indivíduos com um valor de  $Z$  acima de  $c$  recebem o tratamento, e todos os indivíduos com valores abaixo deste valor não recebem. A probabilidade de receber o tratamento muda de 0 para 1 ao passar o ponto de corte. Neste caso, sob as hipóteses de continuidade e ignorabilidade local, não existem outros fatores que expliquem a descontinuidade de  $Y$  em  $c$  a não ser o recebimento do tratamento, e interpretamos o salto de  $T$  no ponto  $c$  como o efeito causal do tratamento. Neste caso,  $T^+ = 1$  e  $T_i^- = 0$ , e o efeito médio local do tratamento é:

$$D^S(c) = Y^+ - Y^- = \lim_{\varepsilon \downarrow 0} E[Y_i | Z_i = c + \varepsilon] - \lim_{\varepsilon \downarrow 0} E[Y_i | Z_i = c - \varepsilon] \quad (7.1.1)$$

Neste caso, a hipótese de ignorabilidade é naturalmente satisfeita, pois para

<sup>54</sup> 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 (6.2) e (6.3).



os indivíduos com o mesmo valor de  $Z$  não existe variação no tratamento. Os indivíduos com valores  $Z$  iguais ou acima de  $c$  terão  $T=1$ ; e os indivíduos com valores abaixo de  $c$  sempre terão  $T=0$ .

O gráfico 7.1.1 ilustra a estrutura da regressão descontínua no caso *sharp*. A figura 1 representa a probabilidade de receber tratamento, ilustrando que esta probabilidade muda de 0 para 1 quando  $Z$  é igual ao valor do ponto de corte. A figura 2 representa as curvas dos resultados potenciais caso o indivíduo receba tratamento ( $Y(1)$ ) e os resultados potenciais caso o indivíduo não receba tratamento ( $Y(0)$ ). Só observamos os pontos da função de  $Y(0)$  para indivíduos com valores de  $Z$  abaixo de  $c$ , e os pontos da função de  $Y(1)$  para os indivíduos com valores de  $Z$  acima de  $c$ . Se estas funções são contínuas em  $Z$ , podemos atribuir o salto que acontece no ponto  $c$  como o efeito médio do tratamento em torno desta descontinuidade.

### Gráfico 7.1.1: Regressão Descontínua Sharp

Figura 1: Probabilidade Condicional de Receber Tratamento

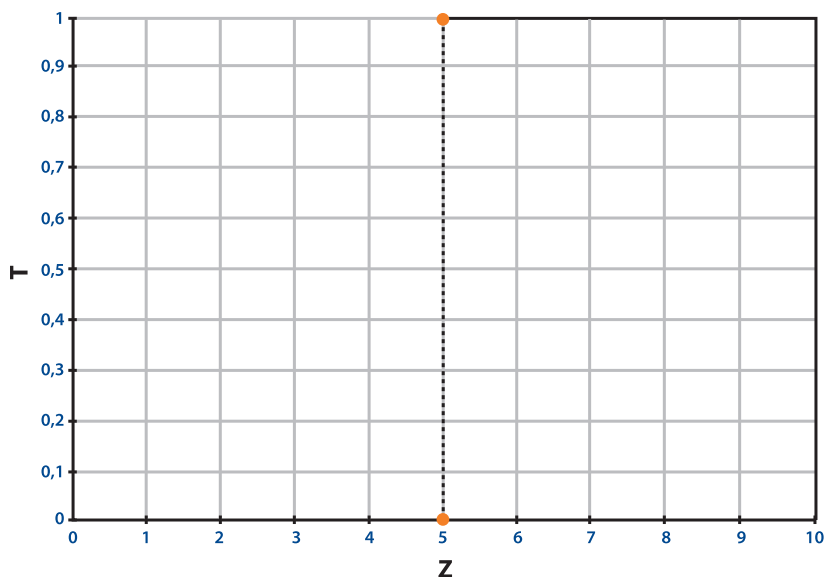
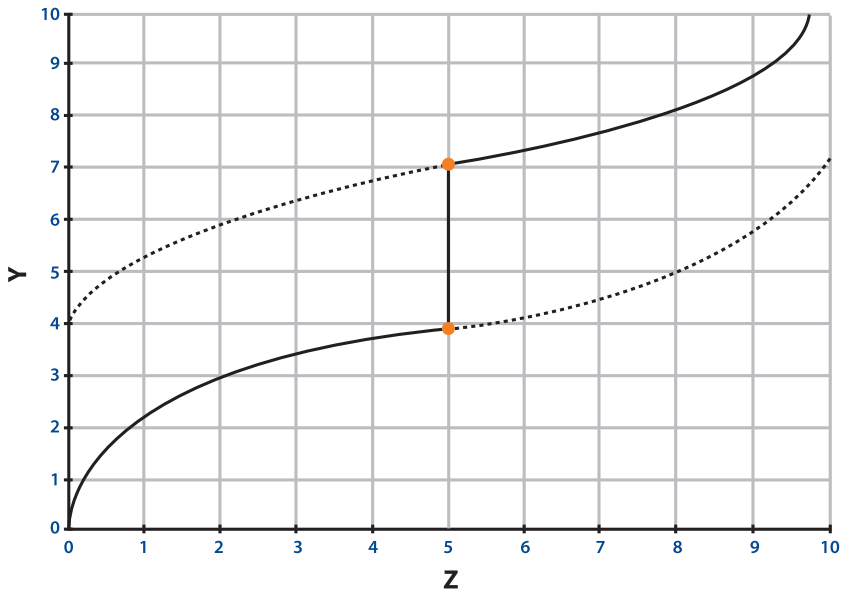


Figura 2: Expectativa Condicional dos Resultados



Suponha o caso simples em que a relação entre  $Y$  e  $Z$  é uma função linear, e queremos identificar o efeito médio do tratamento. Podemos representar o caso sharp de regressão descontínua através do sistema de equações:

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

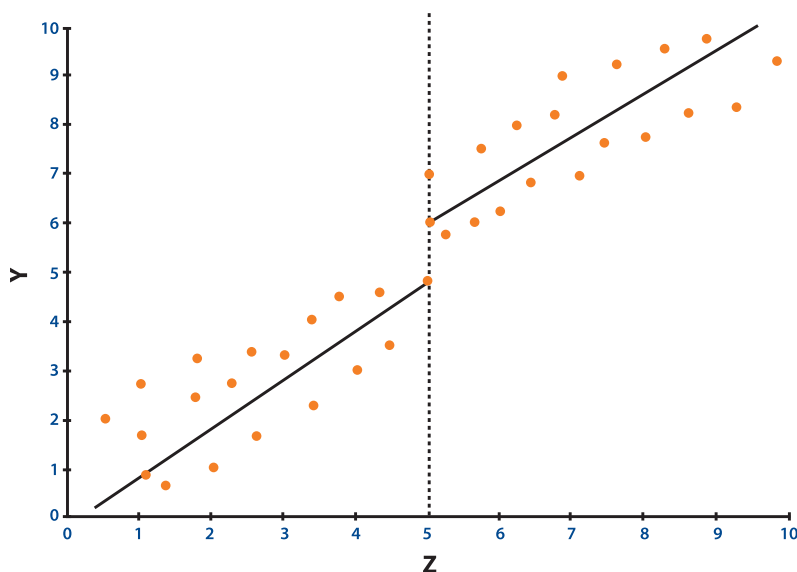
$$T_i = \mathbf{1}\{Z_i \geq c\}$$

no qual  $\mathbf{1}$  é uma função indicadora que assume valor 1, caso o indivíduo tenha o valor de  $Z$  acima de  $c$ ; e 0, caso contrário. O gráfico 7.1.2 ilustra este problema. O salto na variável  $Y$  no ponto  $c$  pode ser interpretado como o efeito causal do tratamento se todos os outros fatores que afetam  $Y$  estão evoluindo de forma suave em torno de  $c$ . De maneira formal, este salto irá representar  $\tau$  se

$$\lim_{\varepsilon \downarrow 0} E[\varepsilon_i | Z_i = c + \varepsilon] = \lim_{\varepsilon \downarrow 0} E[\varepsilon_i | Z_i = c - \varepsilon]$$

Esta hipótese pode ser entendida como a hipótese de exogeneidade que usamos em uma regressão linear tradicional, sendo válida somente em torno do ponto de corte.

Gráfico 7.1.2: O Caso Linear



## 7.2 Caso Fuzzy

Este é o caso mais comum nas aplicações em economia. Neste caso, o tratamento é determinado parcialmente pela descontinuidade em  $Z$ . 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  $Z=c$ , ou seja:

$$\lim_{\varepsilon \downarrow 0} Pr[T_i = 1 | Z_i = c + \varepsilon] \neq \lim_{\varepsilon \downarrow 0} Pr[T_i = 1 | Z_i = c - \varepsilon]$$

Como a probabilidade de receber tratamento tem um salto menor que um, o salto da expectativa condicional de  $Y$  em  $Z$ , que acontece no ponto em que  $Z$  é igual a  $c$ , não pode ser interpretado como efeito médio do tratamento. Neste caso, voltamos à equação (7.4) e identificamos o efeito médio do tratamento como a razão entre o salto que ocorre em  $Y$  e o salto que ocorre em  $Z$  no ponto de corte,

$$D^F(c) = \frac{\lim_{\varepsilon \downarrow 0} E[Y_i | Z_i = c + \varepsilon] - \lim_{\varepsilon \downarrow 0} E[Y_i | Z_i = c - \varepsilon]}{\lim_{\varepsilon \downarrow 0} E[T_i | Z_i = c + \varepsilon] - \lim_{\varepsilon \downarrow 0} E[T_i | Z_i = c - \varepsilon]} \quad (7.2.1)$$

Um exemplo deste caso seria a situação em que o recebimento do prêmio de mérito não depende somente das notas dos alunos no exame de proficiência, mas também de uma entrevista com uma banca examinadora. Neste caso, uma

nota acima de  $c$  aumenta a probabilidade de receber o prêmio de mérito, mas alunos que tiraram uma nota abaixo de  $c$  podem receber o prêmio deste que façam uma boa entrevista.

Para a identificação da razão das descontinuidades como um efeito médio local do tratamento, precisamos das hipóteses H1 e H2, estabelecidas no início deste capítulo. O estimador da regressão descontínua no caso *Fuzzy* pode ser relacionado com o estimador de variável instrumental conhecido como LATE, que foi apresentado no capítulo 6<sup>55</sup>. Como sabemos, o LATE identifica a efeito médio do tratamento para os indivíduos que mudaram o seu status de participação quando o valor do instrumento passa do ponto  $c$ . Para isso, ele precisa de uma hipótese de monotonicidade, além da hipótese de independência<sup>56</sup>. Defina  $T_i(z)$  como o status potencial de tratamento no ponto em que  $Z=z$ , 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  $Z=z$ . A hipótese de monotonicidade pode ser expressa usando este tratamento potencial:

H3:  $T_i(z)$  é uma função não-decrescente em  $z$ .

Dentro do instrumental do LATE, definimos como “compliers” os indivíduos que são de fato afetados pelo tratamento, isto é, um indivíduo  $i$  é denominado de “complier” se:

$$\lim_{\varepsilon \downarrow 0} Pr[T_i = 1 | Z_i = c + \varepsilon] = 1 \text{ e } \lim_{\varepsilon \downarrow 0} Pr[T_i = 1 | Z_i = c - \varepsilon] = 0$$

Os “compliers” são definidos como os indivíduos que recebem o tratamento se o valor de  $Z$  é igual ou acima de  $c$ , mas não recebem o tratamento se o valor de  $Z$  está abaixo de  $c$ . Considere o exemplo em que os indivíduos que recebem uma nota no teste de proficiência abaixo  $c$  são encorajados a fazer aulas de reforço escolar no contra-turno. 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  $c$ . A hipótese de monotonicidade elimina o grupo de indivíduos que receberia o tratamento se o valor de  $Z$  fosse abaixo do ponto de corte e não receberia caso o valor de  $Z$  fosse acima do ponto de corte. Como fizemos no capítulo anterior, usando a hipótese de monotonicidade, podemos decompor o efeito médio do tratamento em torno da descontinuidade no efeito para os “compliers”, no efeito para os indivíduos que sempre aceitam o tratamento independente do valor de  $Z$  e no efeito para os indivíduos que sempre recusam o tratamento independente do valor de  $Z$ .

<sup>55</sup> Para maiores detalhes sobre o LATE, ver Imbens e Angrist (1994) e Angrist, Imbens e Rubin (1996).

<sup>56</sup> Ver Imbens e Angrist (1994).

$$\begin{aligned} & \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c + \varepsilon] - \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c - \varepsilon] = \\ & \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c + \varepsilon, T_i = 1] \cdot (\lim_{\varepsilon \downarrow 0} Pr[T_i = 1|Z_i = c + \varepsilon] - \lim_{\varepsilon \downarrow 0} Pr[T_i = 1|Z_i = c - \varepsilon]) \quad (7.2.2a) \\ & - \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c - \varepsilon, T_i = 0] \cdot (\lim_{\varepsilon \downarrow 0} Pr[T_i = 0|Z_i = c - \varepsilon] - \lim_{\varepsilon \downarrow 0} Pr[T_i = 0|Z_i = c + \varepsilon]) \quad (7.2.2b) \\ & + (\lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c + \varepsilon, T_i = 1] - \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c - \varepsilon, T_i = 1]) \lim_{\varepsilon \downarrow 0} Pr[T_i = 1|Z_i = c - \varepsilon] \quad (7.2.2c) \\ & + (\lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c + \varepsilon, T_i = 0] - \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c - \varepsilon, T_i = 0]) \cdot \lim_{\varepsilon \downarrow 0} Pr[T_i = 0|Z_i = c + \varepsilon] \quad (7.2.2d) \end{aligned}$$

As duas primeiras equações representam o efeito médio local do tratamento para os “compliers”, a equação (7.2.2c) representa o efeito do tratamento para os indivíduos que sempre recebem o tratamento independente do valor de  $Z$  e a última equação representa o efeito médio do tratamento para as pessoas que sempre recusam o tratamento independente do valor de  $Z$ . Pela hipótese de continuidade,

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c + \varepsilon, T_i = 1] &= \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c - \varepsilon, T_i = 1] \\ \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c + \varepsilon, T_i = 0] &= \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c - \varepsilon, T_i = 0] \end{aligned}$$

e temos que:

$$\begin{aligned} & \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c + \varepsilon] - \lim_{\varepsilon \downarrow 0} E[Y_i|Z_i = c - \varepsilon] = \\ & = E[Y_i(1) - Y_i(0)|Z = c, i \text{ é complier}] \end{aligned}$$

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, apresentado no capítulo anterior, 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  $T$  condicional a  $Z$ . 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 esta 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 deste salto é o efeito causal do tratamento. Neste 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

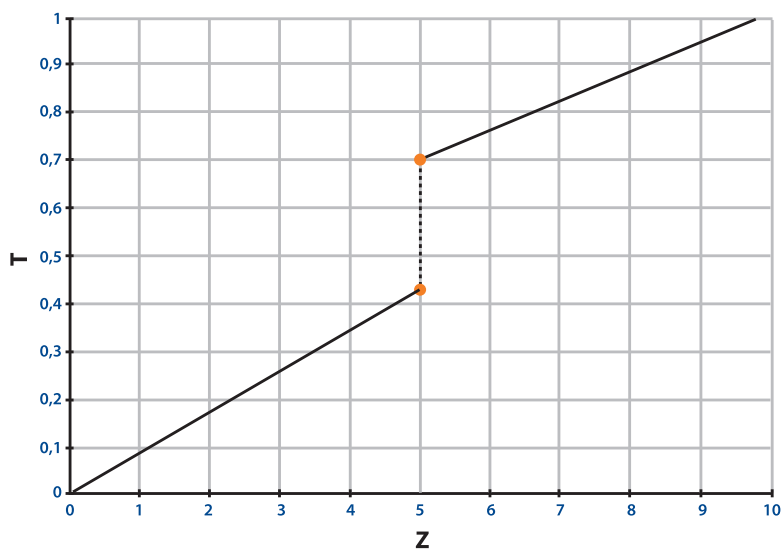
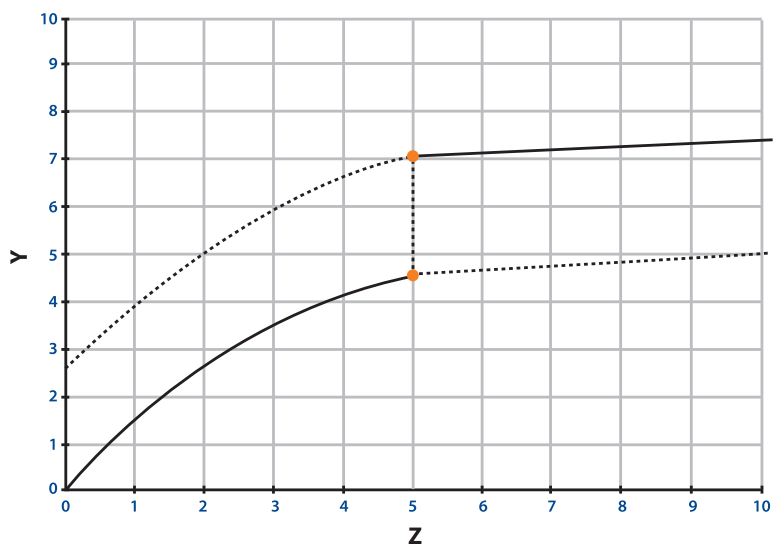


Figura 2: Expectativa Condicional dos Resultados



## 7.3 Estimação

Um 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 corte<sup>57</sup>. No caso *sharp*, precisamos estimar apenas a relação de Y com Z dos dois lados em torno do ponto de corte c. A modelo de regressão do lado esquerdo do ponto de corte é:

$$Y_i = \alpha_l + \beta_l \cdot (Z_i - c) + \varepsilon_i \text{ se } c - h \leq Z < c \quad (7.3.1)$$

e o modelo de regressão do lado direito é:

$$Y_i = \alpha_r + \beta_r \cdot (Z_i - c) + \varepsilon_i \text{ se } c \leq Z < c + h \quad (7.3.2)$$

Como as variáveis explicativas estão subtraídas do valor no ponto de corte, o valor esperado de cada regressão no ponto de corte é dado por  $\alpha_l$  e  $\alpha_r$ . 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:

$$Y_i = \alpha_l + \tau \cdot T_i + \beta_l \cdot (Z_i - c) + \gamma \cdot T_i \cdot (Z_i - c) + \varepsilon_i, \text{ se } c - h \leq Z < c + h \quad (7.3.3)$$

no qual  $\gamma = \beta_r - \beta_l$  e  $\tau = \alpha_r - \alpha_l$ .

A vantagem deste 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.

Dado uma vizinhança h, é fácil estimar o modelo de regressão representado pela equação (7.3.3). Mas para estimar esta 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, quando maior o h, maior a chance da especificação linear não ser a correta para estimar a relação entre Y e Z e podemos aumentar o viés do efeito médio do tratamento. Se a expectativa condicional de Y não é uma função linear em Z, o modelo linear pode ser uma boa aproximação em uma região limitada de Z, mas à medida que aumentarmos a região de valores de Z, o modelo linear pode se tornar uma aproximação muito ruim. Da mesma forma, quanto menor o h, maior a chance de a especificação linear ser uma aproximação adequada para estimar a relação entre Y e Z naquela vizinhança, mas menor a precisão

<sup>57</sup> Nesta seção, apresentamos o mesmo, mais comumente usado para estimar regressão descontínua. Existem outros métodos não-paramétricos que podem ser usados, como uma regressão por Kernel. Para uma revisão de alguns métodos que podem ser implementados, ver Imbens e Lemieux (2008) e Lee e Lemieux (2009).

das estimativas. O  $h$  ideal é aquele que balanceia de forma ótima este trade off entre viés e precisão. Existem várias regras usadas para escolher esta janela,  $h$ . Se o número de observações aumentar, podemos diminuir  $h$ , pois teremos um maior número de observações por intervalo.

Existem dois procedimentos mais comumente usados para escolher esta janela ótima<sup>58</sup>. O primeiro seria usar uma janela que leva em consideração alguns componentes da distribuição conjunta de  $(Y,Z)$ , como por exemplo, a curvatura do modelo de regressão. A regra de bolso usada para a escolha de  $h$  é:

$$h^* = 2,702 \cdot \left( \frac{\tilde{\sigma}^2_R}{\sum_{i=1}^N \{\check{f}''(z_i)\}^2} \right)^{1/5} \quad (7.3.4)$$

em que  $\tilde{\sigma}$  é o erro padrão estimado da regressão,  $\check{f}''(\cdot)$  é a derivada segunda do modelo de regressão estimado para captar a relação em  $Y$  e  $Z$ , e  $R$  representa o intervalo da variável  $Z$  usada para estimar a regressão. Para usar esta regra de bolso, primeiramente precisamos estimar um modelo de regressão arbitrário em uma vizinhança em torno do ponto de corte. Por exemplo, suponha que nesta primeira etapa, estimamos uma regressão quadrática para estimar a relação entre  $Y$  e  $X$ ,  $f(z_i) = \beta_0 + \beta_1 z_i + \beta_2 z_i^2 + \beta_3 z_i^3 + \beta_4 z_i^4$ . Neste caso, após a estimação da regressão, precisamos calcular  $\check{f}''(\cdot)(z_i) = 2\hat{\beta}_2 + 6\hat{\beta}_3 z_i + 12\hat{\beta}_4 z_i^2$ , e estimar  $\tilde{\sigma}^2$  usando a soma do quadrado dos resíduos dividido pelo número de observações, menos os graus de liberdade. Se usarmos as observações na vizinhança,  $c - 5 \leq Z < c + 5$ ,  $R$  será igual a 10.

O segundo procedimento é conhecido como validação cruzada (“Cross-Validation”)<sup>59</sup>. Para cada observação  $i$ , estimamos uma regressão no qual a observação  $i$  é deixada de fora e usamos os valores estimados nesta regressão para obter o valor predito de  $Y$  quando  $Z$  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,  $z_i - h \leq Z < z_i + h$ . Fazemos este exercício para cada observação, e obtemos um conjunto de valores preditos para  $Y$ . A janela ótima será aquela que minimiza a média dos quadrados das diferenças em cada  $Y$  predito e o  $Y$  observado. De maneira formal, para determinado valor de  $h$ , calculamos o valor de predito para  $\hat{Y}^h(z_i)$ , e calculamos o erro quadrático médio:

<sup>58</sup> 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  $Z$  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.

<sup>59</sup> Neste capítulo, iremos descrever o procedimento proposto por Imbens e Lemieux (2008).



$$CV_Y(h) = \frac{\sum_{i=1}^N (Y_i - \bar{Y}^h(z_i))^2}{N} \quad (7.3.5)$$

e o  $h$  ótimo será aquele que minimiza este critério da validação cruzada:

$$h_{CV}^* = \operatorname{argmin}_h CV_Y(h) \quad (7.3.6)$$

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  $Z$ . Neste caso, podemos usar um modelo de regressão linear que relaciona a variável  $T$  ao vetor  $Z$ . 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:

$$T_i = \gamma_l + \delta_l \cdot (Z_i - c) + u_i \text{ se } c - h \leq Z < c \quad (7.3.7)$$

$$T_i = \gamma_r + \delta_r \cdot (Z_i - c) + u_i \text{ se } c \leq Z < c + h \quad (7.3.8)$$

E neste 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  $T$ . Neste caso, usaremos duas janelas diferentes na estimação 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  $Z$  que estão correlacionadas com a variável de interesse. A inclusão destas outras variáveis no modelo não pode afetar a estimação da descontinuidade, se todas as hipóteses da estrutura de uma regressão descontínua são válidas. Se os valores estimados mudarem com a inclusão destas variáveis adicionais, podemos desconfiar que outros fatores também saltam em torno do ponto de corte e que o salto que ocorre em  $Y$  em torno de  $c$  não pode ser atribuído somente ao efeito do tratamento. Se estas demais variáveis explicativas forem correlacionadas com  $Z$ , elas podem aumentar a precisão das estimativas ao diminuir o resíduo da regressão.

Além da estimação, é necessário fazermos testes de 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 nesta regressão. No caso de uma regressão descontínua *fuzzy*, podemos estimar o erro padrão de forma fácil se entendermos este modelo como um caso particular do estimador de mínimos quadrados ordinários em dois estágios. Usando a mesma janela para o modelo que relaciona Y com Z e para o modelo que relaciona T com Z, 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:

$$Y_i = \alpha_l + \tau \cdot T_i + \beta_l \cdot (Z_i - c) + \gamma \cdot T_i \cdot (Z_i - c) + \varepsilon_i, \text{ se } c - h \leq Z < c + h \quad (7.3.10)$$

$$T_i = \gamma_l + \theta \cdot D_i + \delta_l \cdot (Z_i - c) + \pi \cdot D_i \cdot (Z_i - c) + u_i, \text{ se } c - h \leq Z < c + h \quad (7.3.11)$$

no qual D indica se Z excedeu o valor c, isto é, D é uma variável binária igual a 1 se Z é maior que c e 0, caso contrário.

O estimador de MQO em dois estágios para o sistema acima, usando D como instrumento para T 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. Neste caso, para fazermos inferência podemos usar os erros padrões estimados pelo método de mínimos quadrados em dois estágios<sup>60</sup>.

## 7.4 Análise Gráfica

O método de regressão descontínua requer uma análise gráfica. Podemos verificar várias hipóteses de uma estrutura de regressão descontínua através 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. Neste gráfico, temos que observar se em torno do ponto de c 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 estes 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, estas demais variáveis X devem ter uma relação suave com Z, não apre-

<sup>60</sup> Imbens e Lemieux (2008) computam as variâncias assintóticas para os estimadores e propõem um método para estimá-las de forma consistente.

sentando saltos em torno deste 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,  $c$ .

Uma última análise gráfica sugerida por McCrary (2008) seria investigar se houve seleção dos indivíduos em torno do ponto de corte olhando para o número de indivíduos abaixo e acima deste ponto. Com este propósito, usamos um gráfico em que dividimos o suporte de  $Z$  em intervalos e estimamos o número de observações em cada intervalo. Com este gráfico, podemos avaliar 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 de 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 acima do ponto de corte do que abaixo do mesmo.

## 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  $Z$ , o tratamento é completamente determinístico. Neste caso, para todos os indivíduos com o valor de  $Z$  abaixo de  $c$ ,  $T$  é sempre igual a 0, e para os indivíduos com valores de  $Z$  acima de  $c$ ,  $T$  é sempre igual a 1. Ou seja, condicional em  $Z$ , 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  $Z$  tanto no grupo de tratados como no grupo de controles,  $0 < \Pr[T_i = 1|Z_i] < 1$ . Esta hipótese é violada no método de regressão descontínua, pois os indivíduos tratados terão valores de  $Z$  acima de  $c$  enquanto os indivíduos não-tratados terão valor de  $Z$  abaixo de  $c$ . Para todos os valores de  $z$ , a  $\Pr[T_i = 1|Z_i]$  é sempre um ou zero, e nunca está entre 0 e 1. Ao invés de assumirmos a hipótese de sobreposição, impomos continuidade das funções de expectativa condicional de  $Y$  em  $Z$  para o grupo de controle e tratamento. Esta hipótese de continuidade garante que podemos comparar estas duas funções ao longo de todos os valores de  $Z$ . A ideia da regressão descontínua no caso *sharp* é que os indivíduos que possuem valores de  $Z$  em torno do ponto de corte  $c$  são semelhantes na média, exceto que, por sorte, um indivíduo teve um valor  $Z$  acima de  $c$  e o outro teve um valor de  $Z$  abaixo de  $c$ . Logo, podemos pensar a regressão descontínua como uma forma limite de pareamento, aquele 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. Desta 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 um valor de  $Z$  um pouco maior que  $c$ , e o outro grupo um valor um pouco menor. Neste sentido, dentre todos os métodos apresentados neste livro, o método de regressão descontínua é o que mais se aproxima de um experimento real (aleatório).

Na seção 7.3, mostramos que podemos entender a regressão descontínua *fuzzy* dentro do instrumental de variável instrumental. Neste 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  $Z$  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. Na regressão descontínua *fuzzy*, a variação na taxa de participação em torno do ponto de corte é usada para estimar o efeito médio do tratamento. Neste caso, a variação do resultado potencial em torno do ponto de corte é ocasionada somente pela variação da taxa de participação induzida por  $Z$  quando este passa de valores abaixo de  $c$  para valores logo acima de  $c$  e não está relacionada com a variação de nenhum outro fator observável ou não observável.

## 7.6 Exemplo

Ferraz e Finan (2011) 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 têm a possibilidade de reeleição. Para estimar este efeito, eles comparam municípios nos quais 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 comparam 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 com um percentual um pouco acima de 50 e venceu a eleições. A ideia é que estes prefeitos são na média muito parecidos em características que afetariam corrupção e a capacidade de reeleição (como preferências ideológicas). A única diferença é que por sorte um obteve um votação um pouco maior que o ponto de corte, enquanto o outro obteve uma votação um pouco menor. Para estimar o efeito da reeleição para estes 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 por exemplo, um polinômio. Além disso,  $X_i$  é o vetor que contém as características do município, como por exemplo, coeficiente de Gini, população, etc.; e o vetor  $W_i$  inclui características do município que afetam o nível de corrupção, como por exemplo, 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 constroem uma medida de corrupção através 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 nestas irregularidades classificadas, eles constroem uma medida de corrupção que seria o total de recursos relativo a essas atividades irregulares expresso como 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) desta tabela apresenta o caso para a especificação linear, isto,  $f(Z_i) = Z_i\alpha$ . As colunas (2) e (3) usam um polinômio quadrático em  $Z_i$  e outro cúbico. Os resultados indicam que a reeleição tem um efeito negativo na corrupção.

**Tabela 7.6.1: O Efeito da Reeleição na Corrupção**

**Variável Dependente: % de recursos auditados envolvendo corrupção**

<b>Variável Dependente: % de recursos auditados envolvendo corrupção</b>			
	<b>Linear (1)</b>	<b>Quadrática (2)</b>	<b>Cúbica (3)</b>
<b>Prefeito no primeiro mandato</b>	-0,031** [0,014]	-0,040** [0,019]	-0,038** [0,022]
<b>Número de observações</b>	328	328	328
<b>R<sup>2</sup></b>	0,27	0,27	0,27

Fonte: Ferraz e Finan (2011)

Para verificar se de fato existe uma 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 em Z. O gráfico 7.6.2 apresenta tal relação, e comprova a descontinuidade da medida da corrupção em torno do ponto de corte. Usando o instrumental de regressão descontínua, os autores atribuem esta descontinuidade ao efeito da reeleição.

Gráfico 7.6.2: O Efeito da Reeleição na Corrupção

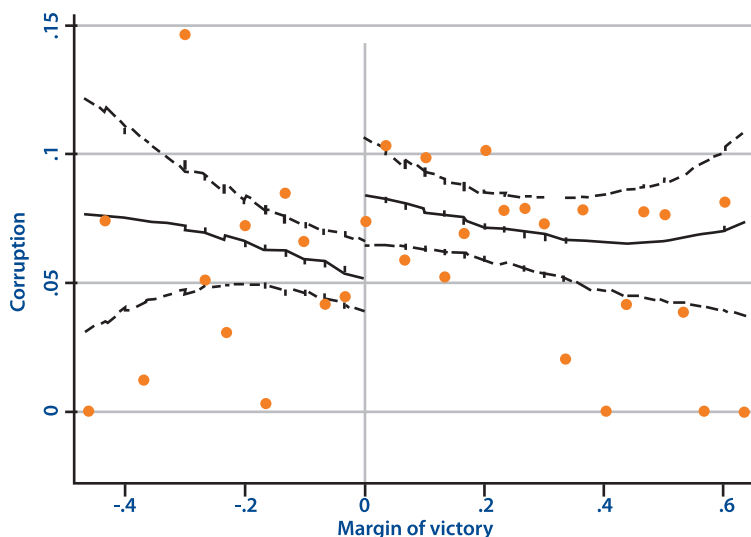


Figure 2: The Effects of Re-Election Incentives on Corruption

Fonte: Ferraz e Finan (2011)

## Exercícios

1) 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. Esta 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 um 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:

$$Y_i = \beta X_i + \rho T_i + \varepsilon_i$$

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  $\varepsilon_i$  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 um maior número de vezes ao médico. Para responder 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  $\varepsilon_i$  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 uma 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  $\pi$ ?

**c)** Suponha que  $\beta_1 > 0$  e  $\pi > 0$ , faça um diagrama que ilustre a relação entre o resultado de interesse ( $Y_i$ ) e 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 idade.

**e)** As hipóteses do arcabouço de regressão descontínua são válidas neste 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 zero. De outra maneira, suponha que  $\hat{\theta}$  é próximo de zero. 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 inú-



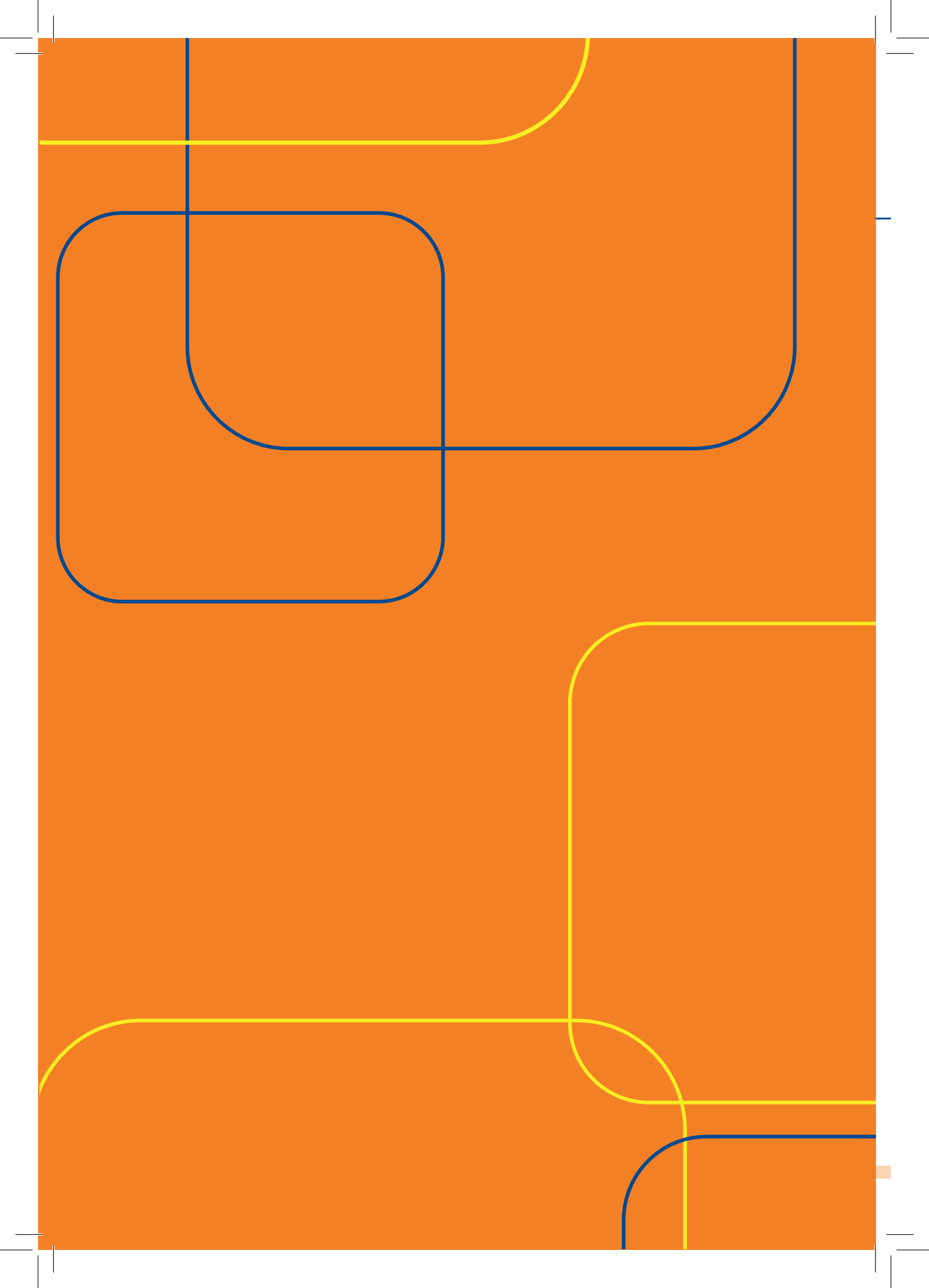
meras descontinuidades no gráfico que relaciona a probabilidade de receber tratamento com o instrumento Z. Como você interpreta esta evidência? Esta evidência invalida o uso de regressão descontínua?

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 esta evidência? Esta evidência 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. Estas 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\$ 701,00 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 de sua implementação. Nesta 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 estes dados, vamos estimar o efeito dessa política de crédito sobre a probabilidade de adquirir um novo domicílio. Com esta 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.



## PARTE 3

# Análise de Retorno Econômico

A parte III desse livro apresenta a análise do retorno econômico de projetos sociais e políticas públicas. A avaliação econômica de um projeto é composta de duas etapas, a avaliação de impacto e a avaliação do retorno econômico. Estas duas etapas são complementares, sendo que a avaliação de retorno econômico é realizada a partir dos resultados obtidos na avaliação de impacto.

A avaliação de retorno econômico visa à mensuração da viabilidade econômica de um projeto. Em outras palavras, 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. Assim, a avaliação de retorno econômico procura responder, principalmente, as seguintes questões: se o benefício gerado pelo projeto é maior do que seu custo; se o projeto é viável economicamente; e se existe algum projeto alternativo mais viável.



Calcular o retorno econômico é importante porque, mesmo que um projeto tenha impacto positivo (estimado por meio da avaliação de impacto), pode acontecer uma situação em que seu custo seja tão alto que inviabilize a replicação. Concluiríamos, nesse caso, que o projeto funciona, mas não compensa. Muitas vezes, existem projetos que exibem um impacto de grande magnitude, mas quando comparados com seu custo, 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 um elemento importante da análise que é o custo. Como vivemos em um mundo com recursos escassos, esse elemento é fundamental nas tomadas de decisões. E

um dos principais motivos de Ferrari não ser consumida 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 o projeto social da mesma forma que se pensa um investimento econômico. Isso porque o benefício advindo do projeto social não é auferido somente pelo financiador (a partir deste ponto, referido como investidor social). Grande parte dos benefícios é auferida pelas pessoas que foram impactadas pelo projeto. Nesse sentido, é um benefício privado. Parte dos benefícios é recebida ainda por pessoas que não participaram do projeto, caracterizando-se como um 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 também os que não participam, 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 BIONDI et al. (2012). 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. 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 uma melhor formação e mais escolaridade, que ela levará por toda a vida – benefício privado. As externalidades (benefícios públicos), advindas da maior escolarização dos participantes, podem ser várias, como aumento da produtividade do trabalho, melhora na qualidade dos serviços públicos, diminuição da subnutrição infantil na geração 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 destes dois elementos.

## CAPÍTULO 8

# O Cálculo do Retorno Econômico

Betânia Peixoto

### 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, por exemplo, a melhora na qualidade da educação, na saúde, a redução da criminalidade, entre outros. A resposta não é simples. Para efetuarmos essas transformações, temos que fazer hipóteses que permitam atribuir um valor monetário ao impacto. Essas hipóteses variam de projeto para projeto e exigem um profundo conhecimento sobre a temática à qual se aplica. O avaliador precisa conhecer a literatura a respeito do indicador de impacto. Ele precisa 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 tem que 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 dessas prisões que teriam acontecido 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 conhecemos bastante a literatura para podermos fazer hipóteses plausíveis sobre todos os parâmetros que serão empregados nessas estimações.

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 em termos da diminuição da taxa de incidência de diabetes na população. Uma menor taxa de incidência de diabetes leva a: uma diminuição dos gastos com o sistema de saúde para tratamento das doenças derivadas dessa; uma melhora na vida das pessoas que adquiriram hábitos saudáveis por causa do projeto e que ficariam doentes caso ele não existisse; economia para a sociedade com os dias de trabalho que seriam perdidos se as pessoas ficassem doentes; diminuição dos gastos com previdência social que teria que 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 (em termos de 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 de toxicômanos e prisões; melhora na qualidade do aprendizado dos filhos daqueles que foram empregados, pois terão uma 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 que 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 que 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 para obter o benefício total. Isto 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 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 determinando montante de dinheiro recebido no tempo inicial valerá em momentos posteriores.

A fórmula que relaciona estes dois conceitos é:

$$VP = \frac{VF}{(1+i)^n} \quad (1)$$

Onde:  $n$  é o número do período

$i$  é a taxa de desconto intertemporal

Para obter o benefício total do projeto, aplicamos diretamente essa fórmula, de forma a considerar 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 é o valor presente do benefício total;  $VB_{t_0}$  é o valor do benefício no tempo inicial, se ocorrer de o benefício ser imediato à implementação; e  $VFB_{t_n}$  é o valor do benefício recebido nos “n” períodos “t”.

A taxa de desconto intertemporal adequada é uma decisão discricionária. Em geral, os economistas consideram a taxa de desconto intertemporal com sendo 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 sobre quanto 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. 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 de 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 extras da sociedade. Por exemplo, um programa de fornecimento de sopa para pessoas carentes que utiliza mão de obra voluntária, que recebe os alimentos de um doador, que utiliza a cozinha cedida de uma instituição beneficente, entre outros recursos. O custo contábil provavelmente será muito baixo e se somente ele fosse computado no cálculo do retorno, esse estaria distorcido. O custo de oportunidade resolve o problema da distorção do retorno, uma vez que computa quanto valeria: a aplicação da mão de obra em outra 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 como 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 usado (o imóvel), esse não consta no custo contábil. Porém, para o projeto funcionar ele ocupa aquele imóvel. Nesse sentido, o projeto usa o recurso (o imóvel), 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 as aulas em vez de ir para o 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 das 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 quanto esses 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 lembrar 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:

$$VPTC_{t_0} = VC_{t_0} + \sum \frac{VFC_{t_n}}{(1+i)^n} \quad (3)$$

Onde o VPTC é o valor presente do custo total;  $VC_{t_0}$  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, essa 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 a 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 um “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, dentro 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. Neste 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 a sociedade mais valor do que retira dela em termos do investimento necessário para sua execução. Assim, temos:

$$\text{VPL} = \text{Benefício} - \text{Custo} \quad (4)$$

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 desse período, por meio da fórmula:

$$\text{VPL}_{t_0} = \left( \text{VB}_{t_0} + \sum \frac{\text{VFB}_{t_n}}{(1+i)^n} \right) - \left( \text{VC}_{t_0} + \sum \frac{\text{VFC}_{t_n}}{(1+i)^n} \right) \quad (5)$$

A análise do resultado é:

Se  $\text{VPL} > 0$  – retorno positivo, o projeto é viável economicamente, pois seu benefício supera o custo.

Se  $\text{VPL} < 0$  – retorno negativo, o projeto não é viável economicamente, pois seu custo supera o benefício gerado.

Se  $\text{VPL} = 0$  – 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 zero. 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:

$$VPL_{t_0} = (VB_{t_0} - VC_{t_0}) + \left( \frac{VFB_{t_1}}{(1 + TIR)^1} + \dots + \frac{VFB_{t_n}}{(1 + TIR)^n} \right) - \left( \frac{VFC_{t_1}}{(1 + TIR)^1} + \dots + \frac{VFC_{t_n}}{(1 + TIR)^n} \right) = 0 \quad (6a)$$

Calculadoras financeiras ou *softwares* fazem essa estimação.

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. 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. 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 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 tomadas. 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 de quanto é o retorno para a sociedade de cada real investido no projeto. Para o seu cálculo basta dividir o benefício total pelo custo econômico total<sup>61</sup>. Tanto o benefício quanto o custo têm que estar a valor presente de um mesmo período de tempo.

$$\text{Razão Custo-Benefício} = \text{VPTB}_{t_n} / \text{VPTC}_{t_n} \quad (7)$$

Onde  $t_n$  indica o tempo  $n$ , no qual o benefício e o custo foram calculados. 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 cada real investido no projeto valerá no futuro apenas sessenta centavos. Ou seja, o dinheiro investido perdeu valor com o projeto, pois esse não foi capaz de gerar 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 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.

<sup>61</sup> Apesar de a 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 termo é *benefit-cost ratio*.

### 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 a transformação daquele 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 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. Nos dois exemplos, o avaliador poderia optar por apresentar a Razão Custo-Efetividade.

A Razão Custo-Efetividade é calculada pela divisão do impacto estimado (na avaliação de impacto) pelo custo econômico total:

$$\text{Razão Custo-Efetividade} = \text{impacto estimado} / \text{VPTC}_{t_n} \quad (9)$$

Esta estatística fornece quanta unidade de impacto se obtém por real gasto. Por exemplo, no caso do projeto para preservação das baleias, a Razão Custo-Efetividade fornece quantas baleias foram salvas para cada real gasto. 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 investidos no programa, três baleias são salvas. No exemplo do projeto de prevenção às doenças, a Razão Custo-Efetividade indica quantas vidas foram salvas 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 que seja avaliada 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 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 tem que decidir entre dois projetos, ele escolherá o que tem maior custo-efetividade, pois com o mesmo custo mais impacto é atingido. No exemplo da baleia, se uma outra metodologia de preservação das baleias fosse aplicada com Razão Custo-Efetividade de 0,000001, esta não seria preferível àquela em que a Razão Custo-Efetividade é de 0,000003, pois com o mesmo montante menos baleias são preservadas.”

Alguém poderia argumentar que para 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 impacto maior do que o outro, mas por seu custo ser muito mais elevado, tenha 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 às vezes de forma discricionária, faz-se 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 de “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 alteração.

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 man-



tido, dizemos que ele é robusto. Isso significa que, mesmo que nossa estimação apresentasse um erro, o valor real fosse diferente do estimado, o resultado não se alteraria.

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 estimação 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 amostra para estimar determinado parâmetro na população, incorremos em erros de estimação 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, recalculamos as estatísticas de retorno econômico e verificamos a manutenção dos resultados. Se os resultados tiverem se mantido, 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. Como vimos, 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 e deflação, como Índice de Preços ao Consumi-

dor 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. Para fazer a análise de sensibilidade escolhemos o menor e o maior valor da taxa de juros real da economia, de acordo com a série histórica das várias taxas existentes.

O raciocínio anterior, para a taxa de desconto intertemporal, se aplica a qualquer outro parâmetro considerado na avaliação econômica. Pode-se, por exemplo, utilizar parâmetros (incertos) para a estimação 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 que utilizamos na avaliação econômica que podem e devem ser testados na análise de sensibilidade a fim de que possamos 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 a fim de 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 serem analisados. Cada parâmetro terá o seu limite inferior e limite superior e seu valor original. Assim, temos  $x$  vezes 3 elementos a serem combinados 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 opostos ao da avaliação, em que o valor presente líquido encontrado foi negativo ao invés de positivo, por exemplo. Quanto menor o percentual de cenários com resultados contrários ao da avaliação, 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 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 estimação 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 acersariam 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. Esse equivale 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 informadas 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) por 12 meses – perfazendo um total de R\$ 180.000,00 por ano.

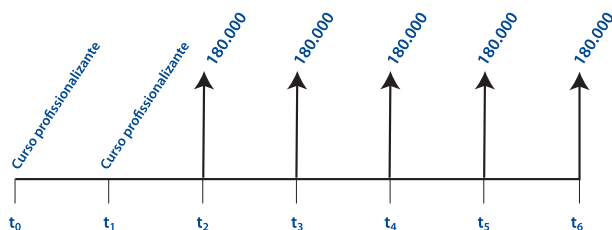
### Fluxo de caixa e cálculo do benefício total:

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.



t<sub>0</sub>

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<sub>0</sub>:

$$VPTB_{t_0} = VB_{t_0} + \sum \frac{VFB_{t_n}}{(1+i)^n}$$

$$VPTB_{t_0} = VB_{t_0} + \frac{VFB_{t_1}}{(1+i)^1} + \frac{VFB_{t_2}}{(1+i)^2} + \frac{VFB_{t_3}}{(1+i)^3} + \frac{VFB_{t_4}}{(1+i)^4} + \frac{VFB_{t_5}}{(1+i)^5} + \frac{VFB_{t_6}}{(1+i)^6}$$

$$VPTB_{t_0} = 0 + 0 + \frac{180.000}{(1+0,05)^2} + \frac{180.000}{(1+0,05)^3} + \frac{180.000}{(1+0,05)^4} + \frac{180.000}{(1+0,05)^5} + \frac{180.000}{(1+0,05)^6} = 742.196,00$$

Portanto, o benefício total do projeto a Valor Presente em t<sub>0</sub> é de 742.196 reais. O cálculo do impacto, conseqüentemente, 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 despense 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. Também os alunos, ao decidirem alocar seu tempo nas aulas em vez de investi-lo em outras atividades, incorrem no custo de oportunidade – tabela 2. Os valores das tabelas se referem aos custos de um ano de funcionamento.

Tabela 1: Custo Contábil		Tabela 2: Custo de Oportunidade	
Custo Contábil	Valor (R\$)	Custo de oportunidade	Valor (R\$)
Salário dos professores e coordenador	60.000,00	Aluguel do laboratório	10.000,00
Material de laboratório	100.000,00	Conta de luz	300,00
Apostilas	5.000,00	Conta de água	150,00
Lanches	72.000,00	Tempo dos alunos	210.000,00
<b>Total</b>	<b>137.000,00</b>	<b>Total</b>	<b>220.450,00</b>

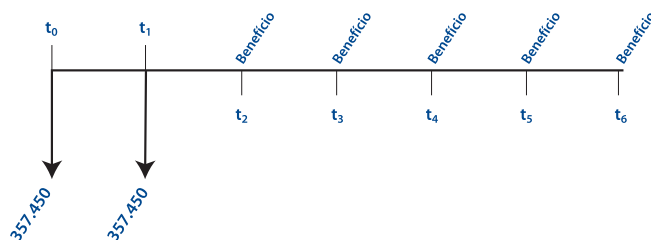
Custo econômico = custo contábil + custo de oportunidade

Custo econômico = 137.000,00 + 220.450,00 = 357.450,00

Assim, o custo do projeto é de 357.450,00 reais 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.

$$VPTC_{t_0} = VC_{t_0} + \sum \frac{VFC_{t_n}}{(1+i)^n}$$

$$VPTC_{t_0} = VC_{t_0} + \frac{VFC_{t_1}}{(1+i)^1}$$

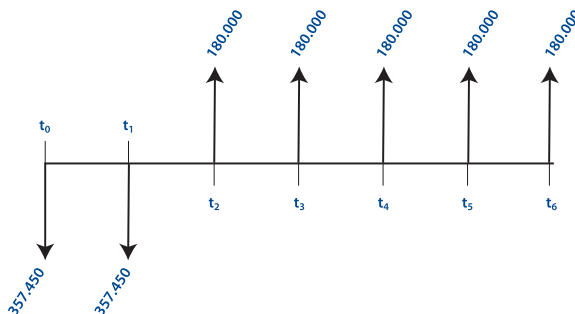
$$VPTC_{t_0} = 357.450 + \frac{357.450}{(1+0,05)^1} = 697.878,57$$

Assim, o custo econômico total do projeto em  $t_0$  é de 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:

$$VPL_{t_0} = \left( VB_{t_0} + \sum \frac{VFB_{t_n}}{(1+i)^n} \right) - \left( VC_{t_0} + \sum \frac{VFC_{t_n}}{(1+i)^n} \right)$$

$$VPL_{t_0} = \left( \frac{180.000}{(1+0,05)^2} + \frac{180.000}{(1+0,05)^3} + \frac{180.000}{(1+0,05)^4} + \frac{180.000}{(1+0,05)^5} + \frac{180.000}{(1+0,05)^6} \right) - \left( 357.450 + \frac{357.450}{(1+0,05)^1} \right)$$

$$VPL_{t_0} = 44.317,43$$

Como já conhecíamos o valor presente do benefício e do custo, o VPL poderia ser calculado de forma alternativa:

$$VPTB_{t_0} : 742.196,00$$

$$VPTC_{t_0} : 697.878,57$$

$$VPL_{t_0} = VPTB_{t_0} - VPTC_{t_0}$$

$$VPL_{t_0} = 742.196,00 - 697.878,57 = 44.317,43$$

O resultado significa que os benefícios do projeto excedem os custos em 44.317,43 reais. Então, podemos concluir que o projeto apresenta um 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:

$$(VB_{t_0} - VC_{t_0}) + \left( \frac{VFB_{t_1}}{(1 + TIR)^1} + \dots + \frac{VFB_{t_n}}{(1 + TIR)^n} \right) - \left( \frac{VFC_{t_1}}{(1 + TIR)^1} + \dots + \frac{VFC_{t_n}}{(1 + TIR)^n} \right) = 0$$

Para obtermos a TIR utilizamos o *software* Excel<sup>62</sup>, no qual inserimos o fluxo de caixa, selecionamos as células e inserimos a função “TIR”, como apresentado na figura a seguir:

tempo	custo	beneficio	beneficio-custo
0	-357450		-357450
1	-357450		-357450
2		180000	180000
3		180000	180000
4		180000	180000
5		180000	180000
6		180000	180000

Formula: =TIR(E14:E20) = 7%

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 desejado pelo investidor.

### Cálculo da Razão Custo-Benefício

Dos cálculos anteriores temos que:

$$VPTB_{t_0} : 742.196,00$$

$$VPTC_{t_0} : 697.878,57$$

$$\text{Razão Custo - Benefício} = \frac{VPTB_{t_0}}{VPTC_{t_0}}$$

$$\text{Razão Custo - Benefício} = \frac{742196}{697878,57} = 1,06$$

<sup>62</sup> 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.

O resultado indica que para cada real gasto no projeto, o retorno para a sociedade será de 1,06 reais. Lembre-se 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 uma taxa de desconto intertemporal de 5%, o retorno do projeto é de 6%.

### Exemplo 1: Projeto para melhora na qualidade da educação (BIONDI et al. 2012):

O projeto "Olimpíada Brasileira de Matemática das Escolas Públicas - OBMEP" tem como objetivo a melhora 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 da 8ª série vale em valores absolutos para todos os alunos matriculados nas escolas participantes do projeto.
- ii. A melhora no desempenho dos jovens na 8ª 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 BIONDI et al., 2012)
- 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 181,70 reais e uma TIR média de 45% ao ano. Assim, concluem que a OBMEP apresenta uma taxa de retorno elevada e gera benefícios salariais futuros aos jovens participantes.



## 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 dos números 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 – 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 au-

tora 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 490 mil reais.

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 43,17 a 52,43 reais e os custos totais anuais médios variam de 1,019 a 1,238 milhões de reais, de acordo com o método de estimação. Para realizar o fluxo de caixa do projeto considera a taxa de desconto intertemporal igual ao índice de inflação IPCA-IBGE variável mensalmente.

A razão custo-efetividade de um homicídio evitado pelo projeto e também a razão custo-benefício é calculada para as três formas de estimação do custo. A razão custo-efetividade encontrada varia, de acordo com o método de cálculo do custo, de 201 a 244 mil reais e a razão custo-benefício varia de 1,99 a 2,42. Os resultados indicam que o projeto gera retorno para a sociedade.

Como foram utilizados vários parâmetros para a estimação dos resultados, a autora realiza a análise de sensibilidade variando os diferentes valores dos anos de produção perdidos, do percentual que este 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.

### 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 de um 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 2011

v. R\$ 1.298,70.

vi. o paciente trabalharia entre 20 e 65 cinco 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 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.

## Exercícios

- 1)** A avaliação econômica é composta de duas etapas. Cite quais são essas etapas e os seus 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 devemos transformar o impacto em benefício?
- 4)** O que é a taxa de desconto intertemporal? E quando ela é utilizada na avaliação econômica?
- 5)** Quais são os componentes do 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 nesse capítulo.
- 9)** Qual a diferença entre as estatísticas de retorno econômico?
- 10)** O que é análise de sensibilidade e qual é a sua finalidade?

## 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, luz, 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 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 277,12 reais, por mês, a renda das participantes.

O custo do projeto é expresso como segue:

<b>Custo Contábil</b>	<b>Valores em R\$</b>
Professores	954.000,00
Pessoal administrativo	254.720,00
Material de consumo	1.670.394,00
Apostilas	67.800,00
<b>Total Contábil</b>	<b>2.946.914,00</b>
<b>Custo de Oportunidade</b>	
Aluguel da sala de aula	460.000,00
Horas dispendidas pelas mulheres com o curso	7.461.642,86
<b>Total Oportunidade</b>	<b>7.921.642,86</b>
<b>Custo o Programa</b>	<b>10.868.556,86</b>

**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 248,45 a 305,79.

**ii.** O retorno do projeto é recebido por apenas 5 anos.

**2)** O programa “Brasil Melhor” –fictício – visa 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 este 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 deste teste de médias com o realizado na letra b. O que podemos concluir com esta comparação?

**f)** Faça o gráfico do escore de propensão para os tratados e controles selecionados pelo método de pareamento. O que podemos concluir com este 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.

## BIBLIOGRAFIA

---

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. 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.; 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.

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. 11ª 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.

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*. Nova Jersey: Prentice Hall, 2005. 560 p.

BOURGUIGNON, F.; FERREIRA, F. Ex-Ante 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.

CAMPBELL, H. F.; BROWN, R. P. C. *Benefit-cost analysis: financial and economic appraisal using spreadsheets*. Cambridge: Cambridge University Press, 2003. 360 p.

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*. 9ª edição. Petrópolis: Vozes, 2011. 312 p.

CRESPO, A. A. *Matemática Financeira Fácil*. 14ª 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.

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.; GLENNERSTER, 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](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.

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.; 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. Varieties of selection bias. American Economic Review. v. 80, p.313-318, 1990.

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.; HOTZ, J. Alternative methods for evaluating the impact of trai-

ning programs. *Journal of the American Statistical Association*. v. 84, n. 804, p. 862-874, 1989.

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.; 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.

HECKMAN, J.; ICHIMURA, H.; TODD, P. Matching as an econometric evaluation estimator. *Review of Economic Studies*. v. 65, p. 261-294, 1998.

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.

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*. 4ª edição. São Paulo: Cengage Learning, 2006. 446 p.

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.; 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, 2ª 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.

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.

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.; 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.

LEE, D. S.; LEMIEUX, T. Regression Discontinuity Designs in Economics. NBER Working Paper. v. 48, n. 14723, p. 281-355, 2009.

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*. Califórnia: 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.; RITOV, 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. Estimating causal effects to treatments in randomized and nonrandomised studies. *Journal of Educational Psychology*. v. 66, p. 688-701, 1974.

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. 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.

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*. 10ª 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.

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.

## BIBLIOGRAFIA COMPLEMENTAR

### I. Estatística

TRIOLA, Mario F. **Introdução à Estatística**. 10ª ed. Rio de Janeiro: LTC, 2008.

Esse manual é bastante adotado em cursos de graduação nas ciências humanas. Apresenta uma linguagem fácil, 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. Um 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**. 4ª ed. São Paulo: Thomson, 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. **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**. 3ª 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**. 14ª 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 uma complementação com outros livros é necessária.

ASSAF NETO, Alexandre. **Matemática Financeira e suas Aplicações**. 11ª 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.



## AUTORES

---

### Betânia Peixoto

Graduada em Ciências Econômicas pela FACE/UFMG (2000), mestre em Economia pelo CEDEPLAR/UFMG (2003) e doutora em Economia pelo CEDEPLAR/UFMG (2008). Tem experiência na área de Economia, com ênfase em Economia dos Programas de Bem-Estar Social e Econometria aplicada, atuando principalmente nos seguintes temas: Avaliação Econômica de Políticas Públicas e Programas Sociais, Análise Econométrica Aplicada e Criminalidade.

### Cristine Campos de Xavier Pinto

Possui Ph. D. em Economia pela University of California, Berkeley. Atualmente é professora assistente da Escola de Economia de São Paulo, FGV. Suas áreas de pesquisa são: Econometria, Economia da Educação e Economia do Trabalho.

### Lycia Lima

Graduou-se em Economia pela UFMG em 2003. Concluiu o mestrado em Development Economics pela School of Oriental and African Studies da Universidade de Londres em 2007. Desde então, trabalhou como pesquisadora na área de monitoramento e avaliação de políticas públicas no Banco Mundial em Washington, na Fundação João Pinheiro do governo do Estado de Minas Gerais e na Secretaria de Assuntos Estratégicos da Presidência da República do Brasil, tendo conduzido principalmente estudos de avaliação de impacto nas áreas de saúde, educação, governança, pobreza rural e primeira infância.

### 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.

## 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. Professor associado da FEA-USP, colunista do Valor Econômico e consultor da Fundação Itaú Social, Naercio é PhD em Economia pela University College London.

## Ricardo Paes de Barros

Graduou-se em Engenharia Eletrônica no Instituto Tecnológico da Aeronáutica – ITA em 1977. Concluiu mestrado em Estatística pelo Instituto de Matemática Pura e Aplicada – IMPA em 1982 e doutorado em Economia pela Universidade de Chicago em 1987. Possui pós-doutorado pelo Centro de Pesquisa em Economia, Universidade de Chicago (1988), e pelo Centro de Crescimento Econômico, Universidade de Yale (1989). Desde 1979, tem trabalhado como pesquisador do Instituto de Pesquisa Econômica Aplicada (IPEA), conduzindo pesquisas nos campos de desigualdade social, educação, pobreza e mercado de trabalho no Brasil e na América Latina. Atualmente é subsecretário da Secretaria de Assuntos Estratégicos da Presidência da República – SAE/PR.