

Inferência Causal e Cenário Contrafactual

Começamos por examinar dois conceitos que se integram ao processo de condução de avaliações que sejam rigorosas e confiáveis: inferência causal e contrafatuais.

Inferência Causal

A questão básica da avaliação de impacto constitui, essencialmente, um problema de inferência causal. Avaliar o impacto de um programa em uma série de resultados equivale a avaliar o efeito causal do programa sobre estes resultados. A maioria das questões sobre políticas envolvem relação de causa e efeito: a formação do professor melhora os resultados das provas dos alunos? Os programas de transferência de renda geram resultados melhores na saúde das crianças? Os programas de capacitação profissional aumentam a renda dos indivíduos capacitados?

Embora as questões de causa e efeito sejam comuns, não é uma questão simples determinar se a relação é causal. No contexto de um programa de capacitação profissional, por exemplo, simplesmente observar que a renda de um indivíduo capacitado aumenta após ele ou ela ter concluído tal programa não é suficiente para estabelecer a causalidade. A renda do indivíduo

capacitado poderia ter aumentado mesmo se ele não tivesse participado de nenhum treinamento, devido a seus próprios esforços, a mudanças nas condições do mercado de trabalho, ou mesmo por causa de uma série de outros fatores que possam afetar a renda. As avaliações de impacto nos ajudam a superar o desafio de determinar a causalidade ao demonstrar empiricamente em que grau um determinado programa - e somente este programa - contribuiu para a mudança de um resultado. Para estabelecer causalidade entre um programa e um resultado, usamos métodos de avaliação de impacto para descartar a possibilidade de que quaisquer outros fatores, além do programa de interesse, expliquem o impacto observado.

A resposta à pergunta básica da avaliação de impacto - *qual é o impacto ou o efeito causal de um programa P sobre um resultado de interesse Y?* - é dada pela fórmula básica da avaliação de impacto:

$$\alpha = (Y | P = 1) - (Y | P = 0)$$

De acordo com esta fórmula, o impacto causal (α) de um programa (P) sobre um resultado (Y) corresponde à diferença entre o resultado (Y) na presença do programa (em outras palavras, quando $P = 1$) e o mesmo resultado (Y) na ausência do programa (isto é, quando $P = 0$).

Por exemplo, se P denota um programa de capacitação profissional e Y designa renda, então, o impacto causal do programa de treinamento vocacional (α) é expresso pela diferença entre a renda do indivíduo (Y) após participar do programa de treinamento vocacional (em outras palavras, quando $P = 1$) e a renda do mesmo indivíduo (Y), no mesmo momento no tempo, considerando que este não tenha participado do programa (ou seja, quando $P = 0$). Em outras palavras, gostaríamos de medir a renda no mesmo instante no tempo, para a mesma unidade de observação (um indivíduo, neste caso), mas em dois cenários diferentes. Se fosse possível fazer isso, estaríamos observando qual seria o valor da renda do mesmo indivíduo no mesmo ponto do tempo com e sem a participação no programa, de forma que a *única* explicação possível para qualquer diferença na renda do indivíduo seria o programa. Ao comparar o indivíduo com ele mesmo, no mesmo momento, seriam eliminados quaisquer fatores exógenos que poderiam igualmente ter explicado a diferença nos resultados. Poderíamos, então, dizer com segurança que a relação entre o programa de capacitação profissional e a renda é causal.

A fórmula básica da avaliação de impacto é válida para qualquer unidade que estiver sendo analisada: um indivíduo, uma família, uma comunidade, uma empresa, uma escola, um hospital, ou qualquer outra unidade de observação que possa receber ou ser impactada por um programa. A fórmula também é válida para qualquer resultado (Y) que esteja plausivelmente relacionado ao

programa em questão. Uma vez medidos os dois componentes-chave desta fórmula - o resultado (Y), tanto na presença quanto na ausência do programa - poderemos responder a qualquer pergunta sobre o impacto do programa.

O Contrafactual

Conforme discutido anteriormente, podemos pensar no impacto (α) de um programa como a diferença nos resultados (Y) para o mesmo indivíduo, com e sem a sua participação em um determinado programa. Contudo, sabemos que é impossível medir o mesmo indivíduo em duas situações diferentes ao mesmo tempo. Em um dado instante no tempo, um indivíduo participou ou não de um dado programa. O indivíduo não pode ser observado simultaneamente em dois estados diferentes (em outras palavras, participando e não participando de um programa). Isto se chama o “problema do contrafactual”: Como é possível medir o que teria acontecido se a outra circunstância tivesse prevalecido? Embora possamos observar e medir o resultado (Y) para os participantes do programa ($Y | P = 1$), não há informações para estabelecer quais teriam sido os resultados na ausência do programa ($Y | P = 0$). Na fórmula básica da avaliação de impacto, o termo ($Y | P = 0$) *representa o cenário contrafactual*. Podemos ver isto como o *que teria acontecido* se um indivíduo não tivesse participado do programa. Em outras palavras, o contrafactual é qual teria sido o resultado (Y) na ausência do programa (P).

Por exemplo, imagine que o “Sr. Desafortunado” tome um comprimido vermelho e morra cinco dias depois. Somente porque o Sr. Desafortunado morreu após ter tomado o comprimido vermelho não se pode concluir que o comprimido vermelho tenha *provocado* a sua morte. Talvez ele estivesse muito doente quando tomou o comprimido vermelho e tenha sido a doença, e não o comprimido vermelho, que provocou a sua morte. Inferir causalidade exige a exclusão de outros fatores potenciais que possam afetar os resultados em consideração. No simples exemplo de determinar se o fato do Sr. Desafortunado ter tomado o comprimido vermelho provocou a sua morte, um avaliador precisa estabelecer o que teria acontecido com o Sr. Desafortunado se ele *não* tivesse tomado o comprimido. Na medida em que o Sr. Desafortunado de fato tomou o comprimido vermelho, não é possível observar diretamente o que teria acontecido caso ele o não tivesse feito. O que teria acontecido com ele caso não tivesse tomado o comprimido vermelho é o cenário contrafactual e o principal desafio do avaliador é determinar como tal estado contrafactual da realidade se manifestaria (vide quadro 3.1).

Ao realizar uma avaliação de impacto, é relativamente simples obter-se o primeiro termo da fórmula básica ($Y | P = 1$), o resultado do tratamento.

Conceito-chave:

O contrafactual é uma estimativa do que seria o resultado (Y) para um participante do programa na ausência deste programa (P).

Simplesmente medimos o resultado de interesse para a população que participou do programa. No entanto, o segundo termo da fórmula ($Y | P = 0$) não pode ser observado diretamente pelos participantes do programa - por isso a necessidade de preencher esta lacuna de informação *estimando-se o contrafactual*. Para tanto, usamos, geralmente, *grupos de comparação*

Quadro 3.1: Estimativa do Contrafactual

A Srta. Única e o Programa de Transferência de Renda

A Srta. Única é uma garotinha recém-nascida, cuja mãe receberá uma transferência de renda mensal contanto que a Srta. Única faça exames médicos regulares no posto de saúde local, seja vacinada e tenha seu crescimento monitorado. O governo supõe que a transferência de renda motivará a mãe da Srta. Única a buscar os serviços de saúde requeridos pelo programa e ajudará a Srta. Única a crescer forte e com a altura adequada. Para a avaliação de impacto, o governo escolheu o crescimento como um indicador de resultado para saúde no longo prazo e medirá a altura da Srta. Única após três anos de duração do programa de transferência de renda.

Suponha que você possa medir a altura da Srta. Única aos três anos de idade. Idealmente, para avaliar o impacto do programa, você teria que medir a altura da Srta. Única aos três anos de idade, considerando que sua mãe tenha recebido a transferência de renda e, também, a altura da Srta. Única aos três anos de idade caso sua mãe não tivesse recebido a transferência de renda. Você compararia, então, as duas alturas. Se você pudesse comparar a altura da Srta. Única aos 3 anos de idade com e sem o programa, você saberia que qualquer diferença na altura teria sido causada somente por conta programa. Porque tudo mais a respeito da Srta. Única seria igual, não haveria outras características que pudessem explicar a diferença na altura.

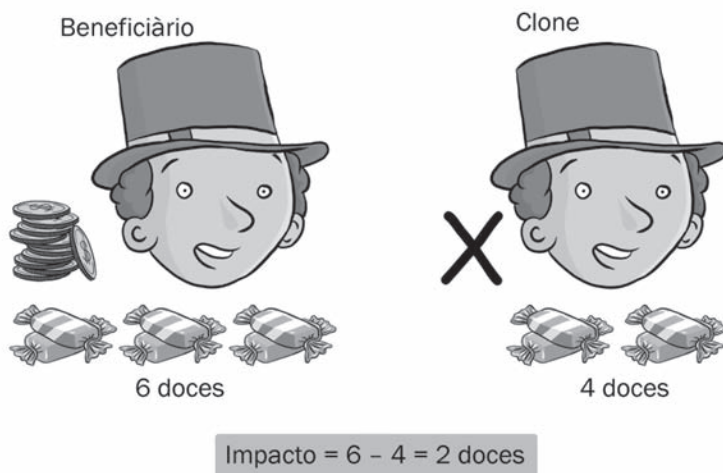
Infelizmente, no entanto, é impossível observar a Srta. Única com e sem o programa de transferência de renda: ou a família dela participa ou não participa do programa. Em outras palavras, não sabemos qual seria o cenário contrafactual. Uma vez que a mãe da Srta. Única de fato participou do programa de transferência de renda, nós não podemos saber qual teria sido a altura da Srta. Única caso sua mãe não tivesse recebido a transferência de renda. Encontrar uma comparação adequada para a Srta. Única será um desafio, precisamente porque a Srta. Única é única. Sua exata situação socioeconômica, características genéticas e pessoais não podem ser encontradas em mais ninguém. Caso tivéssemos simplesmente comparando a Srta. Única a criança que não se inscreveu no programa de transferência de renda, digamos, o Sr. Inigualável, a comparação poderia não ser adequada. A Srta. Única não é idêntica ao Sr. Inigualável. A Srta. Única e o Sr. Inigualável podem não se assemelhar, podem não viver no mesmo local, podem não ter os mesmos pais e podem não ter apresentado a mesma altura ao nascer. Desse modo, se observarmos que o Sr. Inigualável é mais baixo do que a Srta. Única aos 3 anos de idade, não poderemos saber se a diferença é devida ao programa de transferência de renda ou a uma das muitas outras diferenças entre essas duas crianças.

(às vezes denominados “grupos de controle”). O restante da segunda parte deste livro enfocará os diferentes métodos ou abordagens que podem ser usados para identificar grupos válidos de comparação que reproduzam exatamente ou simulem o cenário contrafactual. Identificar tais grupos de comparação é o ponto crucial de toda avaliação de impacto, independentemente do tipo de programa sendo avaliado. Em suma, sem uma estimativa válida do contrafactual, o impacto de um programa não pode ser estabelecido.

Estimando o Contrafactual

A fim de ilustrar melhor a estimação do contrafactual, passaremos a um exemplo hipotético que, embora não seja relevante enquanto política, nos ajudará a refletir um pouco mais sobre este conceito-chave. Em nível conceitual, resolver o problema do contrafactual requer que o avaliador identifique um “clone perfeito” para cada participante (figura 3.1). Por exemplo, digamos que o Sr. Fulano de Tal recebe um adicional de \$12 na sua mesada e que queremos medir o impacto deste tratamento em seu consumo de doces. Se fosse possível identificar um clone perfeito para o Sr. Fulano de Tal, a avaliação seria fácil: poderíamos simplesmente comparar o número

Figura 3.1 O Clone Perfeito



Fonte: Autores

de docos consumidos pelo Sr. Fulano de Tal (digamos que seis) ao número de docos consumidos por seu clone (digamos que quatro). Neste caso, o impacto da mesada adicional seria a diferença entre esses dois valores, ou seja, dois docos. Na prática, sabemos que é impossível identificar clones perfeitos: mesmo entre gêmeos geneticamente idênticos há diferenças importantes.

Embora não exista um clone perfeito para um único indivíduo, existem ferramentas estatísticas que podem ser usadas para gerar dois grupos de indivíduos que, se forem grandes o suficiente, serão estatisticamente indistinguíveis um do outro. Na prática, um objetivo-chave de uma avaliação de impacto é identificar um grupo de participantes do programa (ou grupo de tratamento) e um grupo de não-participantes (o grupo de comparação), que sejam estatisticamente idênticos na ausência do programa. Se os dois grupos forem idênticos, com a única exceção de que um grupo participa do programa e o outro não, então poderemos ter a certeza de que qualquer diferença nos resultados deve-se ao programa.

Conceito-chave:

Um grupo válido de comparação terá as mesmas características que o grupo de participantes do programa ("grupo de tratamento"), exceto pelo fato de que as unidades no grupo de comparação não se beneficiam do programa.

Conceito-chave:

Quando o grupo de comparação para uma avaliação for inválido, então a estimativa do impacto do programa também será inválida, não estimará o verdadeiro impacto do programa. Em termos estatísticos, será "enviesado".

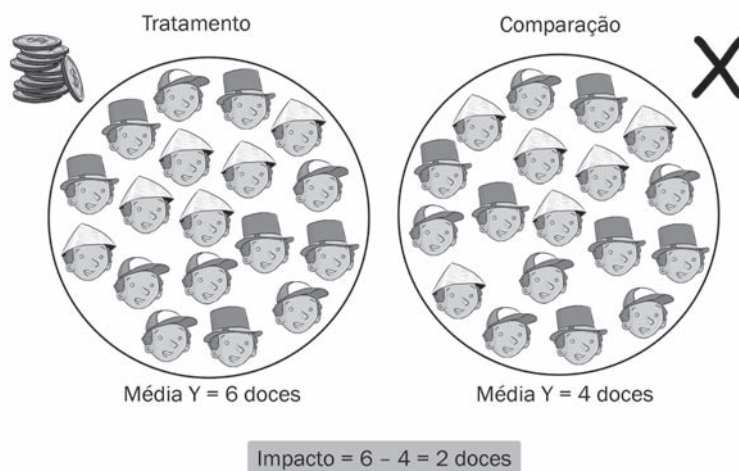
O principal desafio, assim, é identificar um grupo de comparação válido, que possua as mesmas características que o grupo de tratamento. Especificamente, os grupos de comparação e tratamento devem ser equivalentes em pelo menos três aspectos: Primeiro, o grupo de tratamento e o grupo de comparação devem ser idênticos na ausência do programa. Embora não seja necessário que cada observação dentro do grupo de tratamento seja idêntica a cada observação dentro do grupo de comparação, na média, as características de interesse dos grupos de tratamento e de comparação devem ser as mesmas. Por exemplo, a idade média do grupo de tratamento deve ser a igual à idade média do grupo de comparação. Em segundo lugar, o grupo de tratamento e o grupo de comparação devem apresentar a mesma reação ao programa. Por exemplo, as rendas do grupo de tratamento devem ter a mesma probabilidade de serem beneficiadas pelo treinamento que as rendas do grupo de comparação. Em terceiro lugar, os grupos de tratamento e de comparação não podem ser expostos de maneira diferente a outras intervenções durante o período de avaliação. Por exemplo, se isolarmos o impacto da mesada adicional sobre o consumo de docos, o grupo de tratamento não poderia ser privilegiado também com mais idas à confeitaria do que o grupo de comparação, já que isso poderia confundir os efeitos da mesada com o efeito do maior acesso aos docos.

Quando essas três condições são atendidas, então somente a existência do programa de interesse explicará quaisquer diferenças no resultado (Y) entre os dois grupos, contanto que o programa seja implementado. O motivo é que a única diferença entre os grupos de tratamento e de comparação é que os membros do grupo de tratamento receberão o programa, enquanto os

membros do grupo de comparação não o receberão. Quando as diferenças nos resultados puderem ser integralmente atribuídas ao programa, o impacto causal do programa terá sido identificado. Então, ao invés de olhar para o impacto da mesada adicional somente para o Sr. Fulano de Tal, poderíamos olhar para o impacto sobre um grupo de crianças (figura 3.2). Se for possível identificar outro grupo de crianças totalmente semelhante, exceto pelo fato de que elas não recebem o acréscimo de mesada, a estimativa do impacto do programa seria a diferença entre os dois grupos em relação à média de consumo de doces. Consequentemente, se o *grupo tratado* consome uma média de 6 doces por pessoa, enquanto o *grupo de comparação* consome uma média de 4, o impacto médio da mesada adicional sobre o consumo de doces seria 2.

Agora que definimos um grupo de comparação válido, é importante considerar o que aconteceria se decidíssemos continuar com a avaliação sem identificar tal grupo. Intuitivamente, isto agora deveria estar claro: um grupo de comparação inválido é aquele que se diferencia do grupo de tratamento por conta de algum outro aspecto que não somente a ausência do tratamento. Estas diferenças adicionais podem invalidar ou, em termos estatísticos, enviesar a estimativa: não será estimado o impacto real do programa. Mais do que isso, será estimado o efeito do programa combinado ao impacto destas outras diferenças.

Figura 3.2 Um Grupo de Comparação Válido



Fonte: Autores.

Dois Tipos de Estimativas de Impacto

Após estimar o impacto do programa, o avaliador precisa saber como interpretar os resultados. Uma avaliação sempre estima o impacto de um programa comparando os resultados do grupo de tratamento à estimativa do cenário contrafactual obtida de um grupo de comparação válido, usando a equação básica da avaliação de impacto. Dependendo do que o tratamento e o contrafactual de fato representem, a interpretação do impacto de um programa pode variar.

O impacto estimado α é chamado de estimativa da “intenção ao tratamento” (ITT, do inglês, “*intention-to-treat*”), quando a fórmula básica se aplica àquelas unidades às quais o programa foi oferecido, independentemente de terem se inscrito ou não no programa. O estimador ITT é importante para os casos nos quais tentamos determinar o impacto médio de um programa na população-alvo do programa. Por outro lado, o impacto estimado α é chamado de “efeito do tratamento sobre os tratados” (estimador TOT, do inglês, *treatment-on-the-treated*), quando a fórmula básica se aplica àquelas unidades às quais o programa foi oferecido e que realmente se inscreveram. Os estimadores ITT e TOT serão os mesmos quando houver um cumprimento perfeito, isto é, quando todas as unidades para as quais o programa foi oferecido de fato decidiram inscrever-se nele. Voltaremos a falar sobre a diferença entre os estimadores ITT e TOT em detalhes mais adiante. Começaremos com um exemplo.

Considere o Programa de Subsídio ao Seguro Saúde, ou HISP (*Health Insurance Subsidy Program*, na sigla em inglês), exemplo descrito na introdução da parte 2, no qual qualquer família em uma comunidade de tratamento pode inscrever-se em um subsídio de assistência médica. Embora todas as famílias nas comunidades de tratamento estejam qualificadas a se inscrever no programa, uma fração delas, digamos 10%, pode decidir não se inscrever (talvez porque já possuam um plano de assistência médica fornecido pela empresa onde trabalham ou porque sejam saudáveis e não antecipem a necessidade de cuidados médicos, ou devido a inúmeras outras razões). Neste cenário, 90% das famílias na comunidade de tratamento decidem se inscrever no programa e, de fato, receber os serviços que o programa oferece. O estimador ITT seria obtido computando-se a fórmula básica de avaliação de impacto para todas as famílias às quais o programa foi oferecido, isto é, para 100% das famílias das comunidades de tratamento. Por outro lado, o estimador TOT seria obtido calculando-se a fórmula básica da avaliação de impacto somente para o subconjunto de famílias que realmente decidiu se inscrever no programa, isto é, 90% das famílias inscritas nas comunidades de tratamento.

Duas Estimativas Falsas do Contrafactual

No restante da parte 2 deste livro, discutiremos os diversos métodos que podem ser usados para construir grupos de comparação válidos que permitirão estimar o cenário contrafactual. Antes disso, no entanto, é útil discutir dois métodos comuns, porém altamente arriscados, de construção de grupos de comparação, que podem levar a estimativas inadequadas do cenário contrafactual. Estas duas estimativas “falsas” dos cenários contrafatuais são: (1) comparações *antes-e-depois*, ou *pré-pós*, que comparam os resultados dos participantes antes e após a introdução do programa; e (2) comparações *com-e-sem*, entre unidades que escolheram se inscrever no programa e as que preferiram não fazê-lo.

Cenário Contrafactual Falso 1: Comparando o Antes e o Depois

A comparação *antes-e-depois* tenta determinar o impacto de um programa acompanhando as mudanças nos resultados para os participantes do programa ao longo do tempo. Voltando à fórmula básica da avaliação de impacto, o resultado para o grupo de tratamento ($Y | P = 1$) é simplesmente o resultado pós-intervenção. No entanto, o contrafactual ($Y | P = 0$) é estimado usando o resultado da pré-intervenção. Em essência, esta comparação assume que, se o programa nunca tivesse existido, o resultado (Y) para os participantes do programa teria sido exatamente o mesmo que a sua situação pré-programa. Infelizmente, na grande maioria dos casos, esta suposição simplesmente não é válida.

Podemos tomar como exemplo a avaliação de um programa de microfinanças para agricultores pobres na zona rural. Digamos que o programa ofereça aos fazendeiros microempréstimos para a compra de fertilizantes para aumentar sua produção de arroz. Nota-se que, um ano antes do lançamento do programa, os fazendeiros colhiam uma média de 1.000 quilogramas (kg) de arroz por hectare. O programa de microfinanças é lançado e, um ano depois, a produção de arroz aumenta para 1.100 kg por hectare. Se você estiver tentando avaliar o impacto usando uma comparação *antes-e-depois*, o resultado antes da intervenção seria usado como cenário contrafactual. Aplicando a fórmula básica de avaliação de impacto, você concluiria que o programa aumentou a produção de arroz em 100 kg por hectare.

No entanto, imagine que o índice pluviométrico tenha sido normal durante o ano anterior ao lançamento do programa, mas que uma seca ocorreu no ano em que o programa começou. Neste contexto, o resultado pré-intervenção não pode constituir um cenário contrafactual adequado.

Embora as comparações *antes-e-depois* possam ser inválidas na avaliação de impacto, isto não significa que não sejam válidas para outros fins. De fato, os sistemas de informação administrativos de muitos programas geralmente registram dados dos participantes ao longo do tempo. Por exemplo, um sistema de informações sobre gestão educacional pode rotineiramente coletar dados sobre as matrículas de alunos nas escolas onde um programa de merenda escolar esteja sendo executado. Estes dados permitem que os gestores do programa observem se o número de crianças matriculadas na escola está aumentando ao longo do tempo. Essas informações são importantes e

Avaliação de Impacto na Prática

valiosas para os gestores que planejam e desenvolvem relatórios sobre o sistema educacional. No entanto, estabelecer que o programa de merenda escolar tenha *provocado* a mudança observada na matrícula é algo muito mais desafiador, pois muitos fatores diferentes afetam a matrícula de alunos ao longo do tempo. Desse modo, embora monitorar mudanças nos resultados ao longo do tempo para um grupo de participantes seja extremamente valioso, geralmente não nos permite concluir se (ou quanto) um determinado programa contribuiu para uma melhoria, uma vez que podem existir outros fatores que variam no tempo que podem afetar tal resultado.

Nos exemplos do programa de microfinanciamento e da produção de arroz, vimos que muitos fatores podem afetar a produção ao longo do tempo. Do mesmo modo, muitos fatores podem afetar a maioria dos resultados de interesse para os programas de desenvolvimento. Por este motivo, o resultado pré-programa quase nunca é uma boa estimativa do contrafactual, e é por isso que o chamamos de “falso contrafactual”.

Fazendo uma Avaliação Antes-e-Depois do Programa Subsídio ao Seguro Saúde

Suponhamos que o HISP seja um novo programa no seu país, que oferece subsídio para a compra de plano de assistência médica às famílias rurais de baixa renda, e que este plano cubra despesas relacionadas ao atendimento básico de saúde e à aquisição de medicamentos. O objetivo do HISP é reduzir os gastos das famílias de baixa renda com despesas de saúde e, no final das contas, melhorar os resultados de saúde. Embora muitos indicadores de resultados possam ser considerados na avaliação do programa, o governo está particularmente interessado em analisar os efeitos do HISP no montante que as famílias de baixa renda gastam em cuidados básicos de saúde e medicamentos, medidos como despesas anuais per capita da família (doravante denominadas “despesas com a saúde”).

O HISP representará uma grande fatia do orçamento nacional se for expandido para todo o país - até 1,5% do produto interno bruto (PIB), conforme algumas estimativas. Além disso, o desenvolvimento de um programa dessa natureza envolve grandes complexidades logísticas e administrativas. Por estas razões, foi tomada a decisão, nos níveis mais altos do governo, de introduzir o HISP primeiramente como um programa piloto e, depois, dependendo dos resultados da primeira etapa, expandi-lo gradualmente ao longo do tempo. Com base nos resultados de análises financeiras e de custo-benefício, o Presidente e o seu gabinete anunciaram que, para viabilizar o HISP e estendê-lo nacionalmente, o programa deveria reduzir a média anual da despesa per capita com saúde das famílias rurais de baixa

renda em, pelo menos, \$9 em relação ao que gastariam na ausência do programa, e isso deveria acontecer dentro de um período de 2 anos.

O HISP será introduzido em 100 municípios rurais durante a fase inicial do programa piloto. Logo antes do início do programa, o governo do seu país contrata uma empresa de pesquisa para realizar um estudo de linha de base incluindo todas as 4.959 famílias destes municípios. O estudo coleta informações detalhadas sobre cada família, incluindo a composição demográfica, ativos, acesso a serviços médicos e despesas com saúde incorridas no ano anterior. Pouco tempo depois da realização do estudo de linha de base, o HISP é introduzido nos 100 municípios piloto acompanhado de grande divulgação, incluindo eventos comunitários e outras campanhas promocionais para motivar as famílias elegíveis a se inscrever.

Das 4.959 famílias na amostra da linha de base, um total de 2.907 se inscrevem no HISP durante os dois primeiros anos do programa. Durante estes dois anos, o HISP opera com sucesso, de acordo com a maioria das medidas. As taxas de cobertura são altas e as pesquisas mostram que a maior parte das famílias inscritas está satisfeita com o programa. Ao término do período de 2 anos do piloto, uma segunda rodada de dados de avaliação é coletada na mesma amostra de 4.959 famílias.¹

O Presidente e o Ministro da Saúde deixaram você responsável pela supervisão da avaliação de impacto do HISP e lhe incumbiram de informar se o programa deve ou não ser estendido nacionalmente. A sua pergunta de interesse na avaliação de impacto é: *em quanto o HISP reduziu as despesas com saúde das famílias rurais de baixa renda?* Lembre-se que o que está em jogo é muito importante. Se for concluído que o HISP reduziu as despesas com saúde em cerca de \$9 ou mais, ele será estendido nacionalmente. Se o programa não atingir a meta de \$9, você não recomendará a expansão.

O primeiro consultor “especialista” em avaliação que você contrata informa que, para estimar o impacto do HISP, é preciso calcular a mudança nas despesas com saúde ao longo do tempo para as famílias que se inscreveram. O consultor argumenta que, pelo fato de o HISP cobrir todos os custos relacionados aos cuidados básicos de saúde e medicamentos, qualquer redução nas despesas ao longo do tempo deve ser, em grande parte, atribuída ao efeito do HISP. Desse modo, usando somente o subconjunto de famílias inscritas, você estima a média das despesas com saúde antes da implementação do programa e dois anos mais tarde. Em outras palavras, você executa uma avaliação *antes-e-depois*. Os resultados são apresentados na tabela 3.1.

Você observa que as famílias que se inscreveram no HISP reduziram as despesas com saúde de \$14,40, antes da introdução do HISP, para \$7,80 dois anos depois - uma redução de \$6,60 (ou 45%) durante o período. Conforme

Tabela 3.1 Caso 1 — Impacto do HISP Usando a Comparação Antes-e-Depois (Comparação de Médias)

	Depois	Antes	Diferença	Teste t
Despesas da família com saúde	7,8	14,4	-6,6	-28,9

Fonte: Cálculos dos autores a partir de um conjunto de dados hipotéticos.

indicado pelo valor do teste t), a diferença entre as despesas com saúde antes e depois do programa é *estatisticamente significativa*, isto é, a probabilidade de que o efeito estimado seja estatisticamente igual a zero é muito baixa.

Embora a comparação *antes-e-depois* seja para o mesmo grupo de famílias, você se preocupa com o fato de que outros fatores possam ter se modificado ao longo do tempo e tenham afetado as despesas com saúde. Por exemplo, uma variedade de intervenções na saúde têm sido feitas simultaneamente nos municípios em questão. Alternativamente, algumas mudanças nas despesas familiares podem ser resultado da crise financeira pela qual seu país passou recentemente. Para responder a algumas destas questões, o seu consultor conduz uma *análise de regressão* que controlará os fatores externos. Os resultados são apresentados na tabela 3.2

Neste caso, a regressão linear é das despesas com saúde sobre uma variável binária (0-1) que pode assumir o valor de 0, quando a observação se refere à linha de base, ou 1, quando se refere ao momento pós-intervenção. A regressão linear multivariada também *controla* ou *mantém constantes* outras características observadas para as famílias de sua amostra, incluindo indicadores de riqueza (ativos), composição familiar e assim por diante. Note que a regressão linear simples é equivalente à diferença simples entre o antes e o depois nas despesas com a saúde (uma redução de \$6,59). Uma vez controlados outros fatores disponíveis em seus dados, será encontrado um resultado semelhante - uma redução de \$6,65.

Tabela 3.2 Caso 1 — Impacto do HISP Usando a Comparação Antes-e-Depois (Análise de Regressão)

	Regressão linear	Regressão linear multivariada
Impacto estimado nos gastos da família com saúde	-6,59** (0,22)	-6,65** (0,22)

Fonte: Autores.

Observação: Os erros padrão estão entre parênteses.

** 1% de significância estatística.

PERGUNTA 1

- A. Com base nos resultados do caso 1, o HISP deveria ser estendido nacionalmente?
- B. Pode-se dizer que esta análise controla todos os fatores que afetam as despesas com saúde ao longo do tempo?

Cenário Contrafactual Falso 2: Comparando os Inscritos aos Não Inscritos

A comparação das unidades que recebem um programa com as unidades que não o recebem (“com-e-sem”) constitui outro cenário contrafactual falso. Considere, por exemplo, um programa de capacitação profissional para jovens desempregados. Suponha que, dois anos após o lançamento do programa, uma avaliação tente estimar seu impacto sobre a renda comparando as rendas médias de um grupo de jovens que escolheram se inscrever no programa com aquelas de um grupo que decidiu não se inscrever. Suponha que os resultados demonstrem que os jovens que se inscreveram no programa ganham duas vezes mais do que aqueles que não se inscreveram.

Como esses resultados podem ser interpretados? Neste caso, o cenário contrafactual é estimado com base nas rendas de indivíduos que decidiram não se inscrever no programa. No entanto, os dois grupos de jovens tendem a ser fundamentalmente diferentes. Os indivíduos que escolheram participar podem estar altamente motivados a melhorar sua condição de vida e possivelmente têm grandes expectativas quanto ao retorno proporcionado pela capacitação. Por outro lado, os que optaram por não se inscrever podem ser jovens desmotivados que não esperam se beneficiar desse tipo de programa. É provável que estes dois tipos de jovens tenham desempenhos diferentes no mercado de trabalho e também tenham rendas diferentes, mesmo na ausência do programa de capacitação profissional.

Desse modo, o grupo que escolheu não se inscrever não oferece uma boa estimativa do cenário contrafactual. Se uma diferença de renda for observada entre os dois grupos, não poderemos determinar se ela é devida ao programa de capacitação ou às diferenças subjacentes, em termos de motivação e de outros fatores, existentes entre os dois grupos. O fato de indivíduos menos motivados decidirem não se inscrever no programa de treinamento leva, conseqüentemente, a um viés em nossa avaliação do impacto do programa.² Este viés é chamado de “viés de seleção”. Neste caso, se os jovens que se inscreveram tivessem obtido rendas mais altas mesmo na ausência do programa, o viés de seleção seria positivo; em outras palavras, nós estaríamos superestimando o impacto do programa de capacitação profissional sobre a renda.

Conceito-chave:

O viés de seleção ocorre quando as razões pelas quais um indivíduo participa de um programa estão correlacionadas com os resultados. Este viés normalmente ocorre quando o grupo de comparação não é elegível para o programa ou decide não participar dele.

Comparando Indivíduos que Escolheram se Inscrever no Programa de Subsídio ao Seguro Saúde a Aqueles que Decidiram não se Inscrever

Refletindo um pouco mais sobre a comparação *antes-e-depois* com a sua equipe de avaliação, você percebe que ainda há muitos fatores que variam no tempo que podem explicar parte da mudança nas despesas com saúde ao longo do tempo (em particular, o Ministro da Fazenda está preocupado que a recente crise financeira tenha afetado as despesas familiares com saúde e possa explicar a mudança observada). Outro consultor sugere que seria mais apropriado estimar o cenário contrafactual no período pós-intervenção, isto é, dois anos após o início do programa. O consultor observa corretamente que, das 4.959 famílias na amostra da linha de base, somente 2.907 realmente se inscreveram no programa e, então, aproximadamente 41% das famílias na amostra permanecem sem a cobertura do HISP. O consultor argumenta que as famílias da mesma localidade seriam expostas às mesmas intervenções de saúde do lado da oferta e às mesmas condições econômicas locais, de modo que os resultados, no período pós-intervenção, do grupo de famílias não inscritas ajudariam a controlar muitos dos fatores que afetam tanto as famílias inscritas quanto as não inscritas.

Assim, você decide calcular as despesas médias com saúde no período pós-intervenção, tanto para as famílias que se inscreveram no programa quanto para as famílias que decidiram não se inscrever, produzindo as observações apresentadas na tabela 3.3.

Usando as despesas médias com saúde das famílias não inscritas como estimativa do cenário contrafactual, você conclui que o programa reduziu as despesas médias com saúde em aproximadamente \$14. Ao discutir um pouco mais os resultados com o consultor, você pergunta se as famílias que decidiram não se inscrever no programa não poderiam ser sistematicamente diferentes daquelas que se inscreveram. Por exemplo, as famílias que se registraram para o HISP podiam ser aquelas que esperavam ter maiores despesas com saúde, ou mesmo pessoas mais bem informadas sobre o programa ou pessoas mais preocupadas com a saúde de suas famílias. Alternativamente, talvez as famílias que se inscreveram fossem, em média,

Tabela 3.3 Caso 2 — Impacto do HISP Usando Inscritos e Não Inscritos (Comparação de Médias)

	Inscritos	Não inscritos	Diferença	Teste t
Despesas da família com a saúde	7,8	21,8	-13,9	-39,5

Fonte: Autores.

Tabela 3.4 Caso 2 — Impacto do HISP Usando Inscritos e Não Inscritos (Análise de Regressão)

	Regressão linear	Regressão linear multivariada
Impacto estimado nos gastos da família com saúde	-13,9** (0,35)	-9,4** (0,32)

Fonte: Autores.

Observação: Os erros padrão estão entre parênteses.

** 1 % de significância estatística.

mais pobres do que as que não se inscreveram, visto que o HISP é voltado às famílias carentes. O consultor garante que a análise de regressão pode controlar potenciais diferenças entre os dois grupos. Ao controlar todas as características das famílias que estão disponíveis na base de dados, o consultor estima o impacto do programa, conforme apresentado na tabela 3.4.

Com uma regressão linear simples das despesas com saúde sobre uma variável categórica que indica se as famílias se inscreveram ou não no programa, você encontra um impacto estimado de menos \$13,90; em outras palavras, você estima que o programa causou um decréscimo nas despesas médias em saúde de \$13,90. No entanto, quando todas as outras características da população amostral são mantidas constantes, você estima que o programa reduziu as despesas das famílias inscritas em \$9,40 ao ano.

PERGUNTA 2

- A. Com base nos resultados do caso 2, o HISP deveria ser estendido nacionalmente?
- B. Pode-se dizer que esta análise controla todos os fatores que determinam as diferenças nas despesas com saúde entre os dois grupos?

Notas

1. Observe que estamos assumindo ausência de atrição na amostra durante o período de dois anos, ou seja, nenhuma família terá deixado a amostra. Esta não é uma hipótese realista para a maioria das pesquisas envolvendo famílias. Na prática, famílias que mudam de residência às vezes não são encontradas e algumas famílias se desfazem e deixam de existir como unidade familiar.
2. Como outro exemplo, se os jovens que esperam se beneficiar do programa de capacitação têm maior probabilidade de se inscrever (porque esperam, por exemplo, salários mais altos com a capacitação), então estaremos comparando um grupo de indivíduos que esperavam rendas mais altas com um grupo de indivíduos que não esperam rendas mais altas.