

# GUIA DE AVALIAÇÃO DE IMPACTO

| CABO VERDE |



Ministério das Finanças  
e do Fomento Empresarial  
Direção Nacional do Planeamento



CLEAR  
África Lusófona e Brasil



FGV EESP



# GUIA DE AVALIAÇÃO DE IMPACTO

| CABO VERDE |



Ministério das Finanças  
e do Fomento Empresarial  
Direção Nacional do Planeamento



# ÍNDICE

## 07. ■ PREFÁCIO

## 08. ■ INTRODUÇÃO

O que são Políticas Públicas baseadas em Evidências? 9

O Ciclo da Política Pública, o Seguimento e a Avaliação 9

## 12. ■ CAPÍTULO 1 - O PAPEL DA AVALIAÇÃO DE IMPACTO

Por que realizar avaliação de impacto 13

Como medir o impacto 13

Grupo de tratamento e grupo de comparação 15

Parâmetros de interesse 17

## 18. ■ CAPÍTULO 2 - METODOLOGIAS DE AVALIAÇÃO DE IMPACTO

Aleatorização 19

Condições para que o método funcione 19

Como implementar o método 20

Pareamento 23

Condições para que o método funcione 25

Como implementar o método 25

<b>Diferença em diferenças</b>	<b>26</b>
Condições para que o método funcione	28
Como implementar o método	28
<b>Regressão descontínua</b>	<b>29</b>
Condições para que o método funcione	30
Como implementar o método	31
<b>Comparação entre métodos</b>	<b>32</b>
Validade interna e validade externa	35

## **36.** **CAPÍTULO 3 - ANÁLISES DE RETORNO ECONÓMICO E SOCIAL**

Análise de custo-benefício	37
Análise de custo-efetividade	38

## **39.** **CONCLUSÃO** **38**

<b>Figura 1.</b> O Ciclo da Política Pública	10
<b>Figura 2.</b> O S&A no Ciclo da Política Pública	11
<b>Figura 3.</b> Estimação do impacto com o método experimental	21
<b>Figura 4.</b> Ilustração do processo de pareamento	23
<b>Figura 5.</b> Método de Diferença em diferenças	26
<b>Figura 6.</b> Método de Regressão descontínua	29
<b>Quadro 1.</b> Comparativo dos métodos	33



# FICHA TÉCNICA

## Direção Nacional de Planeamento

Serviço de Planeamento Estratégico, Monitoramento e Avaliação

### Director Nacional

Gilson Pina

### Técnicos

Arcioliondo Pinheiro

Jessica Sancha

Vera Almeida

Viviane Andrade

.....

## FGV EESP Clear

### Diretoria

André Portela

### Vice-Diretoria

Lycia Lima

### Gerência Executiva

Gabriela Lacerda

### Elaboração

Priscilla Bacalhau

### Revisão técnica

Alei Fernandes Santos

@2023

O Planeamento, em qualquer circunstância, é a função administrativa que determina, de forma antecipada, o que fazer e os objetivos a alcançar. Hoje, mais do que nunca, sobretudo neste contexto de crises e de incertezas, torna-se imprescindível o recurso a este modelo teórico para a implementação de políticas públicas. Isto fundamenta-se enquanto um processo que, num contexto de economia de esforços e recursos, de domínio do tempo, de antecipação do futuro, de equacionamento de boas práticas, busca uma maior aproximação dos objetivos traçados, através da implementação de projetos e ações refletidos nos diferentes instrumentos de gestão.

Cabo Verde possui uma tradição e uma experiência de planeamento ao nível global e setorial, com um Sistema que se estriba em uma doutrina que integra a visão nacional de desenvolvimento de longo prazo, contida no Programa do Governo, com respostas de curto e médio prazos, ditadas pela conjugação das capacidades, necessidades, constrangimentos e problemas, e arbitradas por um processo, cada vez mais participativo, de definição de prioridades e de um sistema de Seguimento e Avaliação (S&A).

O Governo, na procura por melhores resultados para a sociedade, como uma questão da sua agenda e, na ótica de uma mudança de paradigma na gestão dos recursos públicos, decidiu implementar um modelo de gestão baseado em resultados transformadores para a sociedade, cuja mudança decorre da necessidade de se repensar o Governo e seu papel, tendo sempre em vista a melhoria dos serviços públicos.

A Lei de Bases do Sistema Nacional de Planeamento em vigor e cuja revisão em curso decorre dessa necessidade, define a Direção Nacional do Planeamento (DNP) como o órgão do Sistema Nacional de Planeamento (SNP), responsável pela programação, elaboração e coordenação dos instrumentos de planeamento de longo prazo, competindo-lhe, entre outras importantes atribuições, “acompanhar e avaliar, através do Serviço de Planeamento, Seguimento e Avaliação (SPSA), a implementação dos instrumentos de planeamento de longo, médio e curto prazos”.

A verificação da efetividade de qualquer política passa, decidida e necessariamente, pelo seu monitoramento e sua avaliação, partindo da premissa de que as avaliações qualificadas que produzem resultados confiáveis e credíveis possibilitam o aprimoramento e o ajuste de políticas e justificam investimentos ou economia de recursos, uma vez que evidenciam se os resultados esperados estão sendo alcançados, se os recursos estão sendo utilizados de modo eficiente e tendo o impacto previsto junto dos beneficiários, desempenhando, por isso, no contexto das políticas públicas, um papel essencial na determinação e no alcance dos objetivos e das prioridades do governo.

A Avaliação de Impacto, um tipo específico de Avaliação Ex Post e tema deste Guia, consiste numa metodologia quantitativa utilizada para verificar se a política pública analisada teve sucesso na promoção das mudanças e se teve o impacto pretendido, identificando a existência e a magnitude do impacto de uma política sobre um indicador de interesse ou de impacto.

Por ser o único tipo de avaliação capaz de estabelecer uma relação de causalidade entre a política e os resultados de interesse, a avaliação de impacto tem um papel

crucial no ciclo da política pública. Ao longo da implementação de uma política, muitos fatores podem afetar os resultados obtidos, inclusive fatores externos e, esse tipo de evidência proveniente desta avaliação de impacto pode, e deve ser utilizado para tomar decisões em relação à continuidade, expansão, exclusão ou adequação de uma política. Note-se que o foco da avaliação de impacto é a identificação de relações causais entre a política e as mudanças observadas no indicador de impacto, ou seja, aquela mudança que, na ausência do programa, não seria observada, devendo os seus resultados serem interpretados em conjunto com as evidências fornecidas pelo sistema de Seguimento e Avaliação, subsidiando a tomada de decisão.

Este guia, fruto do esforço e da experiência de equipas técnicas dos diferentes setores, comprometidas com a construção de políticas públicas consistentes e a melhoria do gasto público, detalha os procedimentos para a realização deste tipo de avaliação do ponto de vista prático para os gestores públicos e pode ser referência para todos os níveis de governo.

De caráter orientativo e com o fim de simplificar os procedimentos, o guia servirá, sobretudo, à DNP, aos Serviços Centrais com responsabilidade no planeamento, seguimento e avaliação, podendo influenciar, também e positivamente, as demais esferas governamentais e instituições da sociedade civil. O documento é composto por 3 Capítulos, aos quais antecede uma Parte Introdutória e precede a apresentação das Conclusões. No Capítulo 1, apresentam-se as características, usos e conceitos. O Capítulo 2 detalha os principais métodos utilizados para medir o impacto de uma política e no Capítulo 3, são apresentados os métodos de análise de retorno económico e social que completam os resultados de uma avaliação de impacto.

Torna-se, assim, mais do que oportuna a publicação deste Guia prático de Avaliação de Impacto que inspirado nas melhores práticas internacionais e tendo como base as experiências acumuladas e os exemplos do País, tem o intuito de fornecer referências às avaliações a serem implementadas no âmbito nacional, de disseminar as abordagens e os conceitos de boas práticas de avaliação aos gestores das várias esferas governamentais, permitindo melhor padronização para as rotinas de seguimento e avaliação das políticas públicas. Afinal, o objetivo geral da implantação de políticas públicas baseadas em evidências deve ser o aperfeiçoamento dos serviços prestados à sociedade para a solução de problemas públicos.

Espera-se que este Guia prático de Avaliação de Impacto venha contribuir expressiva e oportunamente para as discussões sobre a importância desta avaliação, a racionalização dos gastos públicos e seu impacto na sociedade. A qualidade dos gastos é um compromisso do governo e esforços têm sido enveredados para que, a médio e longo prazos, Cabo Verde tenha um processo bem definido e efetivo de análise de suas políticas.

# INTRODUÇÃO AO GUIA



Neste guia são apresentadas, de forma simplificada, metodologias para a realização da Avaliação de Impacto de Políticas Públicas, que utilizam técnicas quantitativas para medir o efeito causal da política na população de interesse, no contexto da gestão de políticas baseadas em evidências. Antes de entender os objetivos deste tipo de avaliação e como executá-la apresentamos, inicialmente, o contexto em que deve ser planejada, o contexto de políticas públicas baseadas em evidências.

## O QUE SÃO POLÍTICAS PÚBLICAS BASEADAS EM EVIDÊNCIAS?

A promoção de políticas públicas com base em evidências é a prática de usar um conjunto de informações existentes no processo de tomada de decisão. Pelo fato de os governos disporem de recursos limitados, e, naturalmente, terem de fazer escolhas sobre como aplicá-los, faz-se primordial a análise acerca de quais ações se mostram mais efetivas na solução de problemas públicos. Estas escolhas públicas tornam-se, ainda mais difíceis, quanto mais urgentes forem os problemas sociais enfrentados, e mais escassos forem os contributos *inputs* para resolvê-los.

Neste sentido, práticas de Seguimento e Avaliação (S&A) são importantes instrumentos para a tomada de decisão de gestores públicos, pois são meios através dos quais se geram evidências sobre o funcionamento e os resultados das políticas públicas. Estas práticas podem ser utilizadas para fundamentar a tomada de decisão, provendo informações confiáveis, uma vez que produzidas por métodos verificáveis. Cabe notar que evidência não é o mesmo que opinião. Pelo contrário, o uso de evidências minimiza o caráter subjetivo na escolha pública, eliminando vieses e fazendo com que as decisões se baseiem em informações verificáveis e objetivas.

Uma vez que essas práticas tendem a elevar a eficiência, a transparência e a prestação de contas (*accountability*) dos gastos públicos, existe uma pressão crescente por parte de diversos atores da sociedade para que sejam adotadas. Prova disso é que organismos internacionais, como a Organização das Nações Unidas (ONU) e o Banco Mundial, além, por exemplo, de Governos de países como Benin, Uganda, África do Sul, Chile, México, Colômbia e Estados Unidos, defendem o uso de evidências na formulação e condução das políticas públicas, bem como o uso de práticas de Seguimento e Avaliação no contexto em que atuam.

## O CICLO DA POLÍTICA PÚBLICA, O SEGUIMENTO E A AVALIAÇÃO

O Ciclo da Política Pública é uma representação de todas as etapas que compõem a ação para a solução de um problema público ou para um objetivo que se busca alcançar. No caso em que se envolve todo o processo de tomada de decisão referente a uma determinada política, o Ciclo pode ser representado conforme a Figura 1, que, embora represente uma simplificação da realidade, auxilia na sua compreensão.

O Ciclo tem início com a **identificação do problema**, que deve ser realizada com base em critérios objetivos, fazendo uso de dados quantitativos e qualitativos, sendo que sua relevância deve ser observada por meio de recortes (por exemplo, se a população de uma determinada comunidade é mais afetada do que outra), e dimensionada tendo em consideração as causas e consequências. Uma vez que o problema tenha sido bem identificado, parte-se para a **formulação da política**, e pela definição de seu funcionamento lógico e prático. O funcionamento lógico corresponde a toda a teoria que está por trás da política, e o prático, ao seu funcionamento em si<sup>1</sup>. A **implementação ou realização**, etapa seguinte do ciclo, compreende todos os processos que ocorrem enquanto decorre a política. A etapa de **avaliação**, corresponde à análise do conjunto de evidências obtidas pelo processo de seguimento e de avaliações, tendo como objetivo a etapa seguinte, de **tomada de decisão**, podendo ter casos de recomendações quanto à reformulação, expansão, continuidade ou mesmo interrupção da política.

1 Para mais informações sobre as etapas de identificação do problema e formulação da política, ver Guia Ferramentas para Desenho de Programa e Construção de Quadros Lógicos.

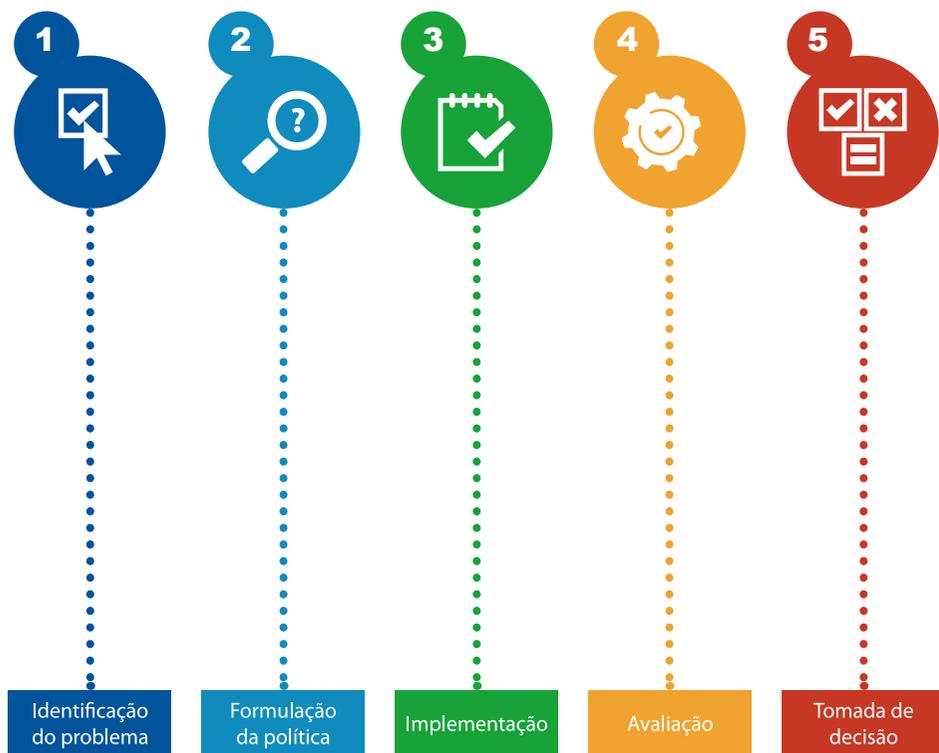


Figura 1. O Ciclo da Política Pública

Fontes: elaborado com base em Secchi (2014) e Kraft & Furlong (2020)

O Seguimento e a Avaliação formam um conjunto de métodos e ferramentas essenciais para o planejamento e para a gestão de políticas públicas, baseadas em evidências. Estas funções constituem processos que ajudam a medir e a analisar o desempenho da ação pública.

O **seguimento** ocorre durante toda a etapa de implementação da política. Conforme definido pela Organização para a Cooperação e Desenvolvimento Econômico - OCDE (2002), o seguimento corresponde a um “processo contínuo que usa a coleta sistemática de dados sobre os indicadores específicos para fornecer, à gestão e às principais partes interessadas de uma intervenção em curso, indicações sobre o progresso e o cumprimento dos objetivos e, ainda, sobre a execução dos fundos alocados” (páginas 27 e 28, tradução nossa).

Para cumprir o seu papel, o seguimento de uma intervenção é precedido da elaboração de um plano ou quadro lógico, que passa pela definição de indicadores relativos aos contributos/*inputs*, às atividades, aos produtos, resultados e impactos da intervenção. A importância da coleta sistemática desses indicadores reside na identificação de desvios, de padrões e de tendências nos dados que se mostrem divergentes do que foi planejado aquando da formulação da política. De posse dessas informações, o gestor pode ponderar a realização de ajustes na intervenção, de modo a favorecer o êxito da mesma. Além disso, o levantamento dos dados de seguimento pode servir de contributo / *input* para a condução de uma avaliação posterior.

Quanto à **avaliação**, de acordo com a OCDE (2002), pode ser vista como o “processo de determinação de valor ou importância de uma atividade, política ou programa, de forma sistemática e objetiva, esteja ela planeada, em andamento, ou concluída”. Seu objetivo é “determinar a relevância e o cumprimento dos objetivos, da eficiência de desenvolvimento, da eficácia, do impacto e da sustentabilidade. Deve fornecer informações que sejam críveis e úteis, permitindo a incorporação das lições aprendidas no processo de tomada de decisão, tanto dos recetores, quanto dos doadores” (páginas 21 e 22, tradução nossa).

As avaliações devem, sempre, partir de uma pergunta, e ser objetivas quanto às informações que pretendem levantar. Diferentes questões de interesse podem surgir, dependendo do estágio de desenvolvimento da política (se em fase de formulação ou implementação, ou após a implementação), sendo que as perguntas as quais se deseja responder deverão definir o tipo, o escopo e o método de avaliação a ser aplicado.

Pode-se classificar os tipos de avaliação em duas grandes categorias, dependendo do estágio em que a política se encontra: avaliação *ex ante* ou avaliação *ex post*. Entende-se por avaliação *ex ante* aquela que ocorre nas etapas anteriores à implementação da política. Conforme representado na Figura 2, o momento *ex ante* corresponde, no ciclo de política pública, às etapas de identificação do problema e formulação da política; isto é, este tipo de avaliação pode ser realizado desde o momento em que se identificam as causas e as consequências da existência de um problema que afeta determinada população, até o momento de planejar e de desenhar a intervenção que vai resolver a situação negativa. Uma avaliação nessa fase busca responder, primordialmente, a questões sobre se o diagnóstico do problema foi bem realizado ou não, se a política formulada apresenta hipóteses lógicas que levem à solução do problema identificado, e se o desenho e os objetivos da política estão bem delimitados e claros.

Por sua vez, a avaliação *ex post* corresponde àquela que ocorre em etapas concomitantes e/ou posteriores à implementação da política. As perguntas que podem nortear avaliações deste tipo estão relacionadas, substancialmente, com questões de desempenho, sendo que, para cada pergunta, pode ser aplicado um ou mais tipos de avaliação diferentes. Por exemplo, pode-se avaliar se o projeto vem sendo implementado conforme planejado por meio de análise da utilização de seus *inputs*, da realização de suas atividades e da entrega de seus produtos imediatos (avaliação de processos); de quais indivíduos o projeto beneficiou (avaliação de focalização); da manutenção da consistência lógica e das hipóteses formuladas na Teoria do Programa (avaliação *ex post* de desenho); do alcance das metas (avaliação de resultados); do impacto sobre as variáveis de interesse (avaliação de impacto); ou dos benefícios resultantes do projeto em relação aos recursos alocados (avaliação de custo-benefício).



Figura 2. O S&A no Ciclo da Política Pública

Fontes: elaborado com base em Secchi (2014) e Kraft & Furlong (2020)

### Evidência não é o mesmo que a opinião

Evidência	Opinião
Produzida por métodos (muitas vezes científicos), visa reduzir o viés e maximizar a objetividade	Está também presente nas políticas Declarações e afirmações que não atendem aos padrões de evidência
Dados que etendem aos padrões de validade interna e validade externa	Qualidades-chaves: <b>subjetiva, parcial (seletiva), de difícil verificação</b>
Qualidades-chave: <b>independência, objetiva, verificável</b>	

Fontes: Adaptado de Clear AA (2000)

A **avaliação, portanto, atua de modo complementar ao seguimento**: quando no processo de seguimento detecta-se que os indicadores estão seguindo uma trajetória diferente da planejada, a realização de uma avaliação pode ajudar a investigar e a chegar a conclusões sobre o porquê de determinada tendência e, que possam ser importantes para possíveis ajustes na política. Isso porque, enquanto o seguimento é contínuo e frequente, e ocorre durante todo o curso de implementação da política pública, a realização de uma avaliação dá-se em momentos-chave, motivada por questões específicas.

Cabe notar que o seguimento e a avaliação podem ser realizados para projetos, programas ou políticas. Embora estes três termos, muitas vezes, se confundem, há uma distinção entre eles. Para Imas e Rist (2009, página 14), **projeto** define-se por “uma única intervenção em um local ou um único projeto implementado em vários locais”; **programa**, “uma intervenção que inclui várias atividades ou projetos que se destinam a contribuir para um objetivo comum”; e **política**, são «padrões, diretrizes ou regras estabelecidas por uma organização para regular as decisões de desenvolvimento».

Este Guia, apenas para fins de simplificação, fará sempre referência ao termo política pública. Porém, métodos ou protocolos apresentados podem também ser adotados em programas ou em projetos.

Nos capítulos do Guia será apresentado um tipo específico de avaliação *ex post*, a Avaliação de Impacto. No Capítulo 1, apresentamos as características, usos e conceitos. O Capítulo 2 detalha os principais métodos utilizados para medir o impacto de uma política. São discutidas no Capítulo 3, métodos de análise de retorno econômico e social, que completam os resultados de uma avaliação de impacto. Por fim, são apresentadas as conclusões.

# *CAPÍTULO 1*

**O PAPEL DA AVALIAÇÃO DE IMPACTO  
NO CICLO DA POLÍTICA PÚBLICA**



## POR QUE REALIZAR AVALIAÇÃO DE IMPACTO

Como saber se os resultados observados foram causados por uma política específica? É possível identificar o quanto das mudanças ocorridas (durante ou após a implementação de uma política) foram geradas pela política e não por outros fatores? Entre as diferentes formas de implementação de uma política, qual é a mais efetiva? Esses são exemplos de perguntas que uma avaliação de impacto pode responder, estabelecendo **relações causais** entre a política avaliada e os resultados encontrados, o que é diferente de verificar se o resultado esperado foi alcançado. Por mais plausível que possa parecer, nem sempre os resultados positivos em indicadores de interesse estão associados à política em questão, pois vários fatores atuam de forma concomitante e, talvez, os resultados tivessem sido obtidos mesmo na ausência da política.

As evidências geradas pela avaliação de impacto permitem identificar se a política promove as mudanças pretendidas para a população de interesse e quantificar a magnitude deste impacto. Em outras palavras, o **impacto** medido é o efeito que foi causado pela política e não por outros fatores.

### DEFINIÇÃO: Avaliação de impacto

Metodologia quantitativa utilizada para verificar se a política pública analisada teve sucesso na promoção das mudanças, se teve o impacto pretendido. Identifica a existência e a magnitude do impacto de uma política sobre um indicador de interesse ou de impacto.

Por ser o único tipo de avaliação capaz de estabelecer uma relação de causalidade entre a política e os resultados de interesse, a avaliação de impacto tem um papel crucial no ciclo da política pública. Ao longo da implementação de uma política, muitos fatores podem afetar os resultados obtidos, inclusive fatores externos. Esse tipo de evidência proveniente da avaliação de impacto pode e deve ser utilizado para tomar decisões em relação a continuidade, expansão, exclusão ou adequação de uma política.

Os resultados encontrados por uma avaliação de impacto devem ser interpretados em conjunto com as evidências fornecidas pelo sistema de Seguimento e Avaliação, subsidiando a tomada de decisão. Metodologias qualitativas podem contribuir com a interpretação dos resultados quantitativos da avaliação de impacto e dos canais de transmissão dos efeitos encontrados. Por exemplo, a ausência de impacto positivo de uma política sobre o resultado de interesse pode ser explicada por falhas na teoria da mudança ou na sua implementação. Neste caso, a avaliação executiva da política já poderia apontar para a não necessidade de realização da avaliação de impacto, devido a problemas na implementação, pois uma política mal implementada dificilmente irá gerar os efeitos causais pretendidos.

## COMO MEDIR O IMPACTO

O foco da avaliação de impacto é a identificação de relações causais entre a política e as mudanças observadas no indicador de impacto, ou seja, aquela mudança que, na ausência do programa, não seria observada. Para isso, é necessário o uso de metodologias científicas que permitem comparar dois cenários:

- i) o resultado observado com a implementação da política pública;
- ii) o resultado que seria observado se a política não tivesse sido implementada.

O cenário (ii) é, por definição, uma situação hipotética e não pode ser observado, ou seja: um indivíduo que é contemplado pela política, jamais saberíamos o que ocorreria se ele não recebesse este tratamento, e vice-versa. Para se ter a certeza de que as mudanças observadas nos **indicadores de impacto** são fruto da participação na política, gostaríamos de poder observar os participantes no mesmo momento, mas em dois estados diferentes: tendo participado e não tendo participado<sup>2</sup>. Esse cenário hipotético é denominado de **contrafactual** (o que teria ocorrido na ausência da política) e, por ser não-observável, deve ser estimado. Para estimá-lo, é preciso observar, na realidade, uma aproximação à situação contrafactual dos indivíduos contemplados pela política. A estimação do contrafactual é realizada por meio da seleção de um **grupo de comparação**.

2 Ver Rubin (1974).

# MODELO DE RESULTADOS POTENCIAIS

Seja  $T$  uma variável indicadora de tratamento, tal que:

- $T_i = 1$  : se indivíduo  $i$  é tratado, ou seja, beneficiado pela política;
- $T_i = 0$  : se indivíduo  $i$  não é tratado, ou seja, não é beneficiado pela política.

Definem-se os resultados potenciais como:

- $Y_i^1$  é o resultado potencial para o indivíduo  $i$  quando ele é tratado ( $T_i = 1$ );
- $Y_i^0$  é o resultado potencial para o indivíduo  $i$  quando ele não é tratado ( $T_i = 0$ ).

Idealmente, gostaríamos de poder observar  $Y_i^1$  e  $Y_i^0$ , ou seja, para cada indivíduo  $i$ , seu resultado potencial em dois estados diferentes (tendo participado e não tendo participado). Dessa forma, o impacto causal da participação no programa seria definido como:

$$\beta_i = Y_i^1 - Y_i^0$$

No entanto, sabemos que é impossível observar o mesmo indivíduo em dois estados diferentes. Observamos apenas uma realização do resultado potencial.

A outra realização é a situação contrafactual, hipotética. Portanto, o resultado observado é dado por:

$$Y_i = Y_i^1 T_i + Y_i^0 (1 - T_i)$$

## GRUPO DE TRATAMENTO E GRUPO DE COMPARAÇÃO

A estimação do contrafactual implica a identificação de um grupo de não-beneficiários pela política que seja muito semelhante ao grupo de beneficiários, de forma que os resultados dos não-beneficiários representem o que teria ocorrido com os beneficiários na ausência da política. Conceitualmente, o grupo de beneficiários é denominado **grupo de tratamento**, e o de não-beneficiários utilizado para estimar o contrafactual é denominado **grupo de comparação**.

### DEFINIÇÃO: Grupo de tratamento

Indivíduos ou unidades beneficiárias pela política avaliada.

### DEFINIÇÃO: Grupo de comparação

Indivíduos ou unidades não-beneficiárias pela política avaliada, mas que apresentam características similares às do grupo de tratamento.

Os membros do grupo de comparação não devem ter sido afetados pela política nem direta nem indiretamente. Desse modo, e por serem semelhantes ao grupo de tratamento, o grupo de comparação representa o que teria acontecido com o grupo de tratamento se ele não tivesse tido acesso à política, isto é, a situação contrafactual.

Portanto, a escolha do grupo de comparação é uma etapa crucial na estimação do impacto de uma política<sup>3</sup>. No momento anterior à implementação da política, o grupo de tratamento e o grupo de comparação devem ser idênticos em média, diferindo-se apenas em relação à futura participação na política<sup>4</sup>. Além disso, o grupo de comparação reagiria à política de forma semelhante e não deve ser exposto a contexto diferente ao longo da implementação da política.

Quando essas condições são satisfeitas, qualquer diferença observada nos indicadores de impacto entre o grupo de tratamento e o de comparação, após a implementação da política, trata-se de seu impacto. O impacto será calculado sempre pela comparação entre esses grupos.

Na prática, não é factível imaginar ser possível encontrar um indivíduo idêntico a cada indivíduo do grupo de tratamento para compor o grupo de comparação. Contudo, quanto mais parecidos os grupos forem na média, mais garantia se tem de que a comparação entre eles após a implementação da política representa o impacto causal de ter sido beneficiado pela política.

Se o grupo de comparação não satisfizer as condições, ou seja, não for idêntico ao grupo de tratamento antes da política, a estimativa de impacto apresentará um viés. Nesses casos, as diferenças entre os grupos correspondem ao impacto do programa, acrescidas outras diferenças existentes entre os grupos.

A escolha do grupo de comparação adequado leva em consideração os critérios de seleção para a beneficiação da política. Esses critérios podem ser sorteio, ordem de inscrição, características como renda, idade, motivação, esforço, etc. Dependendo dos critérios de elegibilidade e de priorização utilizados, pode haver diferentes opções de grupos de comparação em potencial e a avaliação de impacto poderá ter que envolver métodos mais ou menos complexos. Em geral, seleciona-se a melhor metodologia, dados os critérios de seleção já existentes para a política. Em todo caso, busca-se igualar as características dos grupos para que sejam de fato comparáveis.

Se, por exemplo, a política exige a inscrição voluntária para definir quem será beneficiado, este critério pode ser fonte de um **viés de seleção** dos participantes, que é uma distorção no resultado do impacto devido à existência de diferenças sistemáticas entre beneficiários e não-beneficiários, anterior à política. Os indivíduos que não se inscrevem para se beneficiar da política podem ser diferentes dos participantes em termos de motivação ou características não observadas. Logo, selecionar indivíduos que não se inscreveram voluntariamente na política para compor o grupo de comparação pode não ser uma estratégia adequada para se medir o impacto.

<sup>3</sup> Ver Gertler et al (2018).

<sup>4</sup> Em alguns métodos de avaliação de impacto, como o de diferença em diferenças, como será visto adiante, os grupos de tratamento e de comparação não precisam ser idênticos na ausência da política, mas outras condições precisam ser atendidas.

# VIÉS DE SELEÇÃO

Supondo que, antes da política, a média do grupo de tratamento fosse igual à média do grupo de comparação, a diferença do indicador de impacto  $Y_i$  entre o grupo de tratamento ( $T_i = 1$ ) e o grupo de comparação ( $T_i = 0$ ) é dada por:

$$\begin{aligned} \text{Diferença} &= E[Y_i|T_i = 1] - E[Y_i|T_i = 0] \\ &= E[Y_i^1|T_i = 1] - E[Y_i^0|T_i = 0] \end{aligned}$$

Em que  $E[Y_i|T_i = 1]$  é o valor esperado do indicador para o grupo de tratamento,  $E[Y_i|T_i = 0]$  é o equivalente para o grupo de comparação. Somando e subtraindo o termo  $E[Y_i^0|T_i = 1]$ , que representa a situação contrafactual, é possível identificar o viés de seleção:

$$\begin{aligned} \text{Diferença} &= E[Y_i^1|T_i = 1] - E[Y_i^0|T_i = 0] \\ &+ E[Y_i^0|T_i = 1] - E[Y_i^0|T_i = 1] \quad \text{Impacto} \\ &= E[Y_i^1|T_i = 1] - E[Y_i^0|T_i = 1] \quad \text{Viés de seleção} \\ &+ E[Y_i^0|T_i = 1] - E[Y_i^0|T_i = 0] \end{aligned}$$

De forma análoga, a comparação antes e depois dos resultados do mesmo grupo também não se constitui necessariamente como uma estimativa válida para o impacto. Nesse caso, a situação do próprio grupo de tratamento antes da implementação da política está sendo utilizada como grupo de comparação para estimar o contrafactual. Para que esta estratégia seja válida, é necessário supor que, se a política nunca tivesse existido, o resultado para os participantes seria exatamente o mesmo que a sua situação anterior à política, o que, na prática, não costuma ser verdade para políticas implementadas ao longo do tempo.

Vale lembrar que a avaliação de impacto é um tipo de avaliação ex post, ou seja, é realizada após a implementação da política. Contudo, a avaliação de impacto pode ser desenhada de forma concomitante ao desenho do programa. Quando planeamos a avaliação de impacto desde a concepção da política, pode-se optar por critérios de seleção que sejam adequados ao contexto e à condução da avaliação. Isso é especialmente importante no método experimental, como será visto adiante.

## PARÂMETROS DE INTERESSE

Uma avaliação de impacto pode procurar medir diferentes parâmetros de interesse, dependendo do contexto da política avaliada e da disponibilidade de informações. Tendo em vista que é pouco provável que seja possível encontrar “clones” perfeitos das unidades de tratamento para compôr o grupo de comparação, não é possível calcular o impacto para cada beneficiário contemplado pela política de forma individual. Ainda assim, a comparação entre o grupo de tratamento e um grupo de comparação, feita de forma adequada, torna possível calcular outras medidas relevantes para a gestão de políticas baseadas em evidências. Exemplo é o efeito médio da política ou o efeito médio em subgrupos específicos, decorrente da eficácia dessas medidas. Os principais parâmetros de interesse são:

**1. Efeito médio do tratamento** (do inglês, *ATE – Average Treatment Effect*). Diz respeito ao efeito médio da política para toda a população. Trata-se do efeito médio que seria esperado caso uma pessoa da população-alvo, com um determinado conjunto de características, fosse sorteada para ser contemplada.

**2. Efeito médio do tratamento sobre os participantes** (do inglês, *ATT – Average Treatment Effect on Treated*). Diz respeito ao efeito médio da política sobre seus beneficiários.

**3. Efeito médio da intenção de tratar** (do inglês, *ITT – Intention to Treat*). Diz respeito ao efeito médio da política entre os indivíduos para os quais foi oferecida a possibilidade de participar, considerando tanto os participantes como os não participantes.

**4. Efeito médio local do tratamento** (do inglês, *LATE – Local Average Treatment Effect*). Diz respeito ao efeito médio da política na região ao redor de um ponto de corte de elegibilidade, ou apenas para um subconjunto específico da população. Esse efeito não pode ser generalizado para toda a população, pois é uma estimativa local.



# *CAPÍTULO 2*

**METODOLOGIAS DE AVALIAÇÃO DE  
IMPACTO**



O grupo de comparação deve sempre ser o mais parecido possível ao grupo de tratamento nas características relevantes e em média. Contudo, as metodologias de avaliação de impacto diferem entre si na forma em que selecionam o grupo de comparação, além de se basearem em hipóteses distintas<sup>5</sup>. As metodologias podem ser classificadas entre o **método experimental** e os **métodos não experimentais**. No método experimental, a seleção de beneficiários da política é realizada por meio de um processo aleatório, ou de um sorteio.

A ausência dessa aleatoriedade não impossibilita a realização de uma avaliação de impacto, mas a definição do grupo de comparação não é feita de forma aleatória. Quando os critérios de seleção das unidades beneficiárias por uma política não se baseiam em sorteio aleatório, e sim em características observáveis ou não observáveis, pode-se utilizar **métodos não experimentais** para estimar o impacto da política. Novamente, a escolha do grupo de comparação dependerá dos critérios de seleção para a participação.

Neste capítulo, serão apresentados o método experimental e alguns métodos não experimentais muito utilizados na literatura de avaliação de impacto. São eles: **pareamento**, **diferença em diferenças** e **regressão descontínua**.

## ALEATORIZAÇÃO

É comum haver excesso de demanda no contexto de programas sociais e políticas públicas em que se precisa realizar a alocação de recursos escassos. Nessas situações, gestores públicos precisam selecionar beneficiários entre os interessados que fazem parte do público-alvo que se pretende atender. Partindo da premissa de que os critérios de seleção para a participação em um programa social devem ser justos e claros, uma opção é a **aleatorização**<sup>6</sup>, também chamado de método experimental. A aleatorização corresponde à realização de um sorteio entre os interessados que fazem parte do público-alvo da política, dividindo-os entre beneficiários e não-beneficiários.

O método experimental de avaliação de impacto utiliza como grupo de tratamento os indivíduos ou unidades elegíveis que foram sorteados para serem beneficiados pela política, e, como grupo de comparação, os indivíduos ou unidades que, apesar de elegíveis, foram sorteados para não serem beneficiados. Quando o grupo de comparação é selecionado de forma aleatória, ele é também chamado de **grupo de controle**.

Uma aplicação comum da avaliação experimental ocorre na área de saúde, com a testagem de medicamentos ou vacinas, e foi nesta área que a nomenclatura de tratamento e controle surgiu. Na testagem de uma nova vacina, por exemplo, sempre haverá um grupo de tratamento composto por pessoas que receberão a vacina que está sendo testada, e um grupo de comparação composto por pessoas muito parecidas, na média, que receberão apenas um placebo. A vacina será considerada eficaz se as pessoas do grupo de tratamento se contaminarem menos, ou reagirem melhor quando contaminadas, quando comparadas às pessoas do grupo de comparação.

### *Condições para que o método funcione*

Na avaliação de políticas públicas ou programas sociais, a realização de uma aleatorização é capaz de garantir que os grupos de tratamento e de controle sejam semelhantes nas características relevantes. Desse modo, quaisquer diferenças nos indicadores de impacto após a implementação da política podem ser atribuídas à participação, sendo, por isso, a avaliação experimental considerada o **padrão-ouro da avaliação de impacto**, o método mais robusto para estimar o impacto de uma política.

Contudo, para que a aleatorização funcione, é preciso haver um número de participantes em potencial suficientemente grande, o que garantirá que os dois grupos tenham alta probabilidade de serem iguais, do ponto de vista estatístico. Isso é verdade tanto para características observáveis, como idade e escolaridade, quanto para características não observáveis ou de difícil medição, como motivação e esforço dos indivíduos. Ou seja, havendo uma amostra suficientemente grande de unidades elegíveis, os grupos aleatórios serão semelhantes nas características observáveis (exemplo: idade média) e nas não-observáveis (exemplo: motivação média). O **balanceamento** das características observáveis pode ser testado se estiverem disponíveis dados de um momento anterior ao início do programa. Já o balanceamento das características não-observáveis não pode ser testado.

Se a aleatoriedade for realizada considerando apenas a população elegível e com uma amostra suficientemente grande, diferenças encontradas entre o grupo de tratamento e o grupo de comparação, após a implementação da política, podem ser interpretadas como sendo o seu impacto, uma vez que, por construção, os grupos são semelhantes e estão expostos aos mesmos fatores externos ao longo do tempo.

<sup>5</sup> Para análises descritivas e comparativas dos métodos de avaliação de impacto, ver: Gertler et al (2018), Menezes Filho e Pinto (2017), Instituto Jones dos Santos Neves (2018), entre outros.

<sup>6</sup> O termo "aleatorização" é utilizado na literatura científica para designar processos aleatórios de seleção de beneficiários, ou seja, há uma aleatoriedade na definição de quem será beneficiário da política e quem não será beneficiário.

Por outro lado, nem sempre esse tipo de seleção de beneficiários é factível do ponto de vista ético ou legal. Por exemplo, sortear vagas para a participação em um programa de reforço escolar dentro da mesma turma pode ter implicações éticas para com os estudantes e suas famílias. Além disso, o método experimental só é viável quando planeado de forma conjunta com a política, uma vez que a seleção dos participantes deve ser definida de forma aleatória antes da implementação.

## POSSIBILIDADE DE AVALIAÇÃO DE IMPACTO EM CABO VERDE: Programa Casa para Todos<sup>1</sup>

O Programa Casa para Todos consistiu na construção de habitações sociais, viabilizado por uma parceria entre Cabo Verde e Portugal. O programa, implementado pelo Ministério da Descentralização, Habitação e Ordenamento do Território (MDHOT) entre 2008 e 2013, promoveu a construção de habitações de interesse social para a população de baixa e média renda. A parceria deveria construir 8 mil habitações e reabilitar 18 mil casas, por meio de uma linha de crédito no valor de 200 milhões de euros.

Os resultados esperados do programa são reduzir o déficit habitacional e garantir uma habitação digna a famílias com baixa e média renda. Deste modo, uma avaliação de impacto pode buscar responder se o programa contribuiu para que famílias de renda baixa tivessem acesso a uma habitação digna e para a melhoria no acesso ao saneamento básico (água potável, casa de banho, serviço de recolha de lixo, etc.), e qual a magnitude desse efeito.

A escolha das famílias que foram beneficiadas pelo programa ocorreu por meio de um processo aleatório, com base nas informações disponíveis no Cadastro Único de Beneficiários. A atribuição aleatória das moradias viabiliza o uso do método experimental para avaliar o impacto do programa:

*Grupo de tratamento:* composto pelas famílias inscritas que foram sorteadas e beneficiadas.

*Grupo de controle:* composto pelas famílias inscritas, mas que não foram sorteadas para serem beneficiárias.

## Como implementar o método

Alguns passos básicos devem ser seguidos ao implementar uma avaliação experimental:

**1.** *Definir a lista de participantes em potencial.* O primeiro passo na implementação de uma avaliação experimental é listar todos os participantes ou beneficiários em potencial da política, que sejam unidades elegíveis. Para tanto, é preciso que os critérios de elegibilidade estejam bem definidos. Esta etapa também pode incluir a inscrição dos possíveis participantes, havendo, portanto, uma autoseleção. É importante que, entre todos os participantes em potencial, sejam mantidos apenas aqueles que fazem parte do público-alvo da política. A unidade de análise da avaliação pode ser tanto o indivíduo diretamente beneficiado pela política, ou unidades maiores, como grupos de indivíduos, escolas, regiões, municípios ou ilhas.

**2.** *Definir o desenho da aleatorização.* A seleção aleatória dos participantes ou beneficiários por uma política pode ocorrer naturalmente, como uma regra no processo de alocação de recursos. Muitas vezes, quando há excesso de demanda pelos benefícios que a política irá oferecer, os participantes são selecionados de forma aleatória, por ser uma forma justa e imparcial. O desenho dessa aleatoriedade também pode envolver uma entrada escalonada, comum em cenários de escassez de recursos: parte da população-alvo será atendida primeiro, e outra parte em um momento futuro. Em outros casos, é relevante realizar a aleatorização dentro de grupos ou regiões, de modo a garantir que todos os grupos estejam representados na amostra da avaliação. Vale ressaltar que a amostra da avaliação não precisa necessariamente incluir toda a população de unidades elegíveis, desde que seja suficientemente grande para a estimação do impacto e representativa da população. O tamanho da amostra será determinado por cálculos estatísticos.

**3.** *Sortear de forma aleatória participantes e não participantes.* Assim como o desenho da aleatorização, a seleção aleatória dos participantes deve ser planeada e executada antes do início da implementação da política. O processo de seleção aleatória define a condição de tratamento para cada unidade da amostra da avaliação: selecionado ou não selecionado. Há várias formas de realizar a seleção aleatória, por exemplo, atribuindo-se um número aleatório a cada unidade, ordenando-se esses números e selecionando os menores valores. É importante que o critério de seleção do sorteio seja definido antes de sua realização, para que a seleção seja justa e transparente. Com o sorteio, define-se, portanto, os integrantes do grupo de tratamento, que são aqueles selecionados para participar da política, e os integrantes do grupo de comparação, que não foram selecionados.

**4.** *Verificar se os grupos sorteados são semelhantes.* Com informações do momento anterior à implementação da política, deve-se verificar se a seleção aleatória de beneficiários funcionou. Em outras palavras, espera-se que a seleção aleatória de participantes gere grupos de tratamento e de comparação muito semelhantes, e, portanto, haja um balanceamento das características observáveis entre os grupos na **linha de base** (antes da implementação da política). Havendo similitude entre os grupos, elimina-se o viés de seleção<sup>7</sup>. Caso o balanceamento não seja constatado, o método experimental deverá ser complementado por outros métodos para que sejam controlados os possíveis efeitos das características que são diferentes entre os grupos.

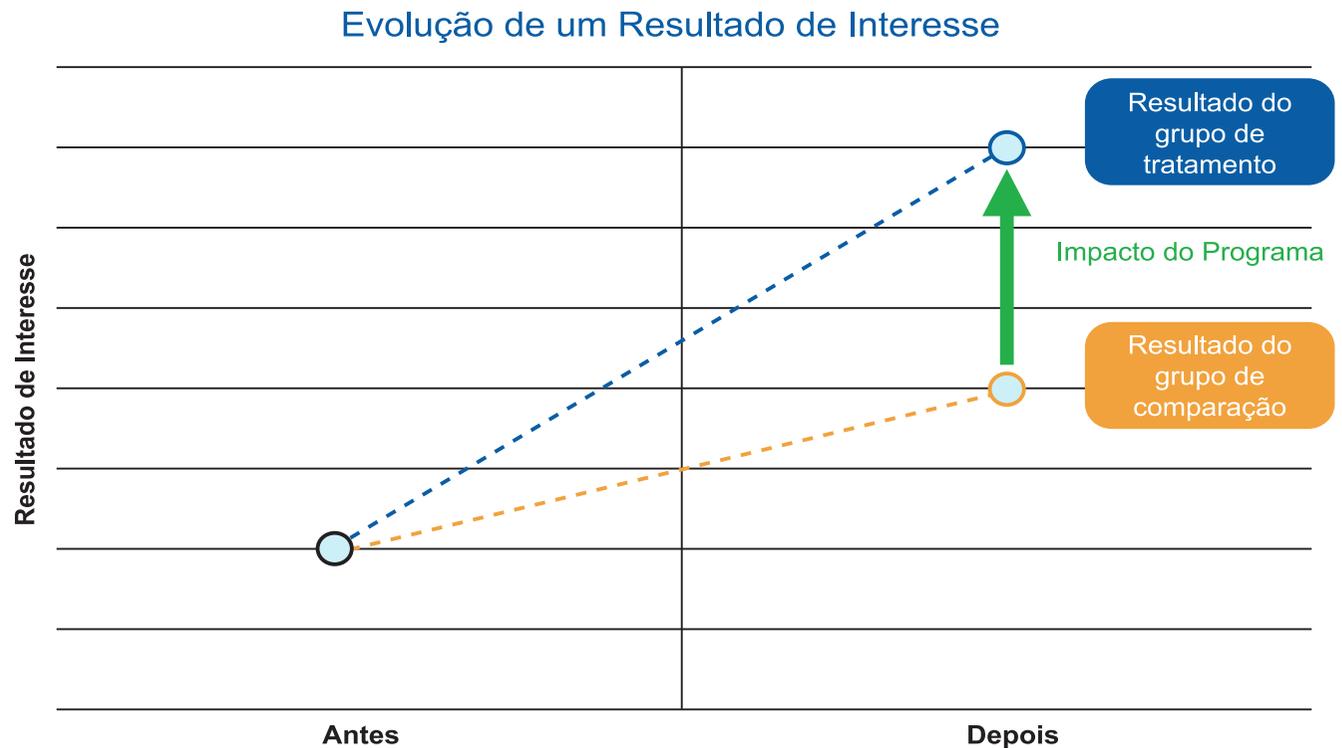
**5.** *Seguir a implementação da política.* O seguimento contínuo da implementação da política é relevante para assegurar a integridade do experimento. Alguns problemas comuns que afetam a integridade do experimento são: **não cumprimento** (quando unidades sorteadas para o grupo de tratamento não participam da política); **contaminação ou transbordamento** (quando unidades selecionadas para o grupo de comparação são beneficiadas pela política); ou **atrito** (quando não é possível encontrar mais os integrantes da avaliação após a realização do sorteio). A ocorrência de alguma dessas situações pode gerar viés para a estimativa de impacto.

**6.** *Estimar o impacto do programa a partir da comparação entre os grupos.* Por meio da coleta de informações após a implementação da política, estima-se o impacto a partir da comparação de médias entre os grupos: resultado obtido pelo grupo de tratamento subtraído do resultado obtido pelo grupo de comparação, como ilustrado na Figura 2. O impacto estimado deve ser submetido a testes estatísticos para verificar se é, de fato, diferente de zero.

<sup>7</sup> Ver Angrist e Pischke (2008, Capítulo 2).

Figura 3: Estimação do impacto com o método experimental

Fonte: Elaboração própria.



Podemos estimar o impacto a partir da comparação de médias entre os grupos, pois a aleatoriedade implica que a alocação seja independente dos resultados potenciais:

$$E[Y^0|T = 1] = E[Y^0|T = 0] = E[Y^0]$$

Como o viés de seleção é zero, o impacto é igual à diferença entre os resultados médios dos grupos:

$$\text{Impacto} = Y^1 - Y^0$$

Contudo, para afirmar que há impacto, é preciso verificar se o impacto estimado é estatisticamente diferente de zero. Uma vez que se trata de estimativas, há sempre um erro envolvido, ou seja, o valor pontual encontrado na estimativa pode não ser exatamente igual ao valor verdadeiro do parâmetro que está sendo estimado.

Para isso, é preciso calcular o intervalo de confiança associado ao impacto. Se o valor 0 (zero) estiver contido nesse intervalo, há indícios de que o programa não tem impacto (não se pode dizer que o valor encontrado é estatisticamente diferente de zero).

Embora seja possível calcular o impacto de um programa a partir da diferença de médias entre grupo de tratamento e grupo de comparação, é conveniente fazê-lo a partir de uma regressão pois:

1. A análise da regressão é uma forma fácil de testar a significância estatística dos efeitos estimados do tratamento;
2. Os métodos de regressão permitem considerar outras características que afetam o indicador de impacto, incluindo-as como variáveis explicativas na regressão, o que pode aumentar a precisão das estimativas.

Quando se está em um caso em que houve aleatoriedade da participação no programa, é possível estimar o impacto da participação no programa sobre o indicador de impacto a partir de uma regressão simples:

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

em que  $Y_i$  é o indicador de impacto para o indivíduo  $i$ ;  $T_i$  é a variável binária indicativa da participação no programa (assume o valor 1 para indivíduos do grupo de tratamento e o valor 0 para indivíduos do grupo de comparação);  $\varepsilon$  é o termo de erro.  $\alpha$  e  $\beta$  são os parâmetros a serem estimados, sendo que  $\alpha$  corresponde à média do indicador de impacto para o grupo de comparação e  $\beta$  corresponde à média do diferencial do indicador de impacto entre grupo de tratamento e grupo de comparação, ou seja, corresponde ao impacto da participação no programa (benefício médio dos indivíduos do grupo beneficiário face às condições do grupo de comparação).

Neste caso, o parâmetro de interesse é  $\beta$ . Uma vez estimado, precisamos verificar sua significância estatística, ou seja, verificar se o valor encontrado pode ser considerado estatisticamente diferente de 0 (zero), de forma a poder afirmar se o programa tem impacto ou não.

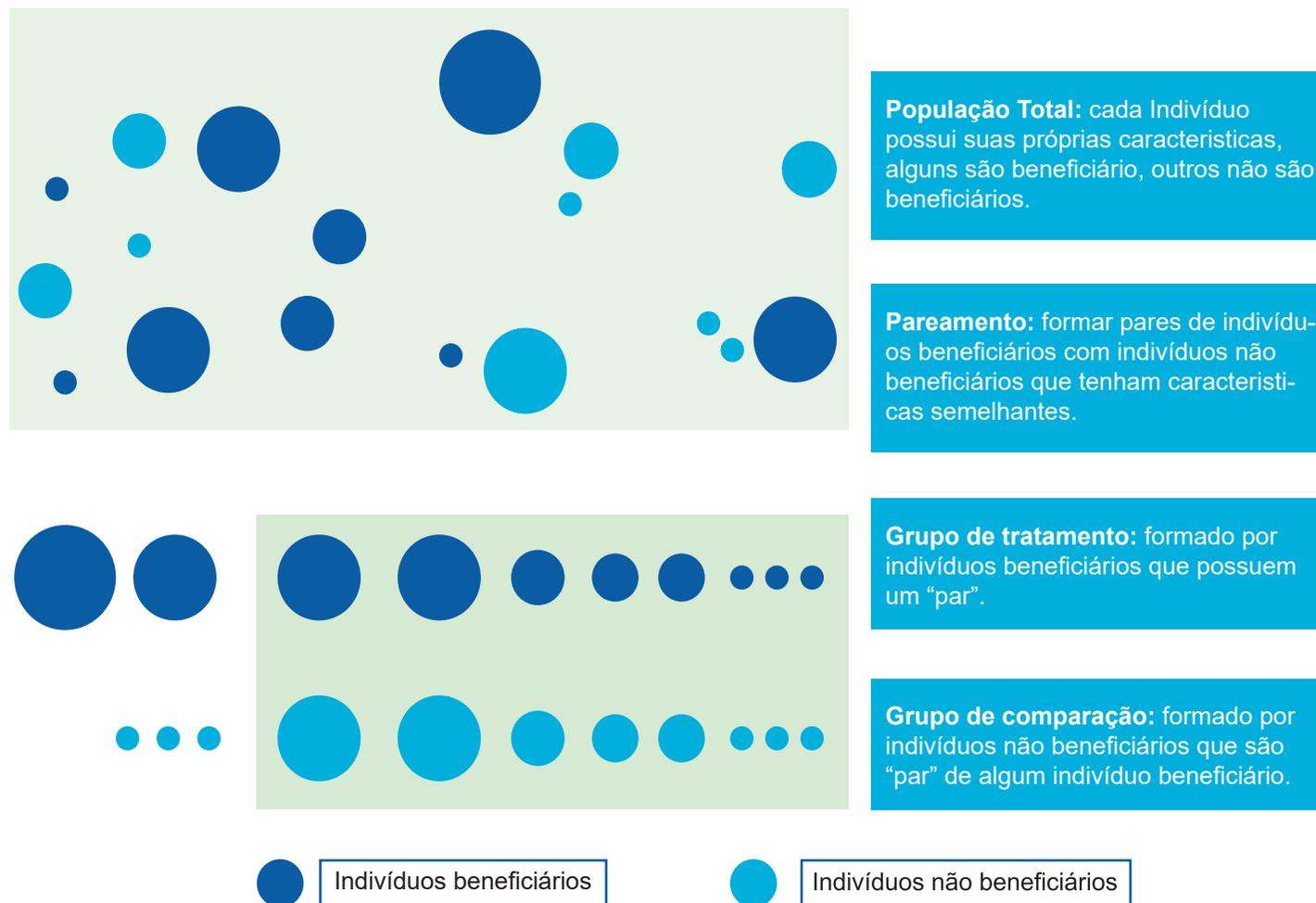
Para verificar a significância estatística, o primeiro passo é definir o nível de confiança desejado. Tipicamente, será utilizado o valor de 95% de confiança. A regressão permite verificar se o parâmetro em questão é estatisticamente significativo, a partir de seu intervalo de confiança associado ou de um teste de hipóteses.

## PAREAMENTO

O método do **pareamento**<sup>8</sup> (do inglês, *matching*) cria, por meio de técnicas estatísticas, um grupo de comparação artificial que se assemelha ao grupo de tratamento, baseado em características observáveis. A principal hipótese em que esse método se baseia é que, condicional a um conjunto de características observáveis relevantes, a participação no programa pode ser considerada aleatória e, em particular, aleatória em relação aos resultados de interesse (**hipótese de seleção em observáveis**). Esse método pode ser utilizado mesmo quando os critérios de seleção não forem tão claros ou bem definidos, desde que a seleção seja baseada em características observáveis.

O pareamento funciona da seguinte forma: para cada unidade do grupo de tratamento, deve-se encontrar uma ou mais unidades não tratadas que apresentem características observáveis mais semelhantes possíveis para servir como seu "par". Essas unidades não tratadas compõem o grupo de comparação. A Figura 3 ilustra o processo de pareamento.

Figura 4: Ilustração do processo de pareamento



Fonte: Elaboração própria.

<sup>8</sup> O termo "pareamento" (do inglês, *matching*) é utilizado na literatura de avaliação de impacto para designar o processo de formar pares, ou seja, encontrar indivíduos que sejam parecidos e possam ser "pareados", formar um par.

As características observáveis consideradas no processo de pareamento devem ser aquelas que sejam relevantes para determinar a participação no programa ou política avaliada e as expectativas para o resultado de interesse. O sucesso do método depende da escolha dessas características relevantes, às quais devem ser anteriores à política e, portanto, não afetadas por ela. Assim, não devem ser incluídas características que possam ter sido afetadas pela expectativa de ser beneficiado pela política. Uma vez escolhidas as variáveis observáveis relevantes para definir a participação na política, os grupos de tratamento e de comparação cruzados serão muito parecidos no momento anterior à implementação política, e a única diferença entre eles será que um será beneficiado e o outro não. Assim, quaisquer diferenças nos indicadores de impacto após o período de implementação podem ser atribuídas à política e não a outros fatores.

## POSSIBILIDADE DE AVALIAÇÃO DE IMPACTO EM CABO VERDE: Plataforma do Turismo<sup>2</sup>

O Programa Cabo Verde Plataforma do Turismo (CVPT) tem como principal objetivo reforçar o papel do turismo como motor de crescimento inclusivo, através da criação de emprego decente, diversificação da economia e sustentabilidade social e ambiental. O Ministério do Turismo deu início ao programa em 2017, com a implementação de um sistema de fiscalização e licenciamentos e oferecendo capacitações para mulheres e chefes de família do setor de turismo e profissionais de 300 estabelecimentos de alojamentos de todas as ilhas.

Espera-se que com esse programa os estabelecimentos hoteleiros aumentem a ocupação e, conseqüentemente, os resultados econômicos do setor de turismo.

Levando em consideração que houve um excesso de demanda no início da implementação do programa, o método do pareamento pode ser utilizado para encontrar um grupo de comparação válido. Características observáveis relevantes que podem ser utilizadas para comparar indivíduos e encontrar semelhantes, incluem: classificação do hotel (número de estrelas), capacidades de alojamento, pessoal ao serviço, localização/ilha, entre outras que afetem a probabilidade de participar do programa.

*Grupo de tratamento:* estabelecimentos de alojamento que se inscreveram para participar no programa e foram beneficiados.

*Grupo de comparação:* estabelecimentos de alojamento que se inscreveram para participar no programa, mas que não foram sorteados para participar, e apresentam características similares aos do grupo de tratamento.

Entretanto, encontrar um par para cada unidade tratada pode não ser possível, sobretudo se a lista de características que desejamos levar em consideração for muito extensa, incluir variáveis contínuas, ou quando os grupos são relativamente pequenos. Para tentar contornar esse problema, conhecido como **problema da dimensionalidade**, pode-se optar pelo pareamento baseado no **escore de propensão**<sup>9</sup> (do inglês, *propensity score matching*). Ao invés de buscar um par para cada unidade tratada, calcula-se a probabilidade de cada unidade tratada e de cada unidade não tratada participarem do programa. Essa probabilidade é denominada escore de propensão, medida que assume valores de 0 a 1 e sintetiza todas as características observáveis consideradas que influenciam a probabilidade de participar ou de se inscrever para participar.

Uma vez calculados os escores de propensão, é possível identificar unidades tratadas parecidas com unidades não tratadas que apresentem escores de propensão, o mais próximo possível, formando, assim, o grupo de comparação. Utilizar o escore de propensão resolve o problema da dimensionalidade, pois apenas uma variável será comparada (o escore de propensão), ao invés de precisar comparar todas as variáveis consideradas relevantes.

Há diversos critérios que podem ser utilizados para realizar o pareamento de indivíduos semelhantes. Por exemplo, para cada unidade do grupo de tratamento, podemos selecionar para formar o grupo de comparação apenas uma unidade que tenha um escore de propensão exatamente igual ao dela ou diversas unidades, cujo escore de propensão esteja próximo o suficiente, tendo definido um intervalo de referência. Os principais critérios são<sup>10</sup>:

- **VIZINHO MAIS PRÓXIMO:** para cada unidade tratada, busca-se uma ou mais unidades não tratadas que mais se assemelham a ela em termos da distância absoluta entre os escores de propensão para compor o grupo de comparação.
- **CALIPER:** variação do critério do vizinho mais próximo, em que é estabelecida uma distância máxima permitida para a distância absoluta entre os escores de propensão, de forma a evitar pares muito diferentes entre si em termos do escore de propensão.
- **RAIO:** todas as unidades não tratadas que se encontram dentro do caliper (nível de tolerância em termos da distância dos escores de propensão) da unidade tratada são utilizadas para compor o grupo de comparação.
- **KERNEL:** todas as unidades não tratadas são incluídas no grupo de comparação, mas com pesos diferentes dependendo da distância entre os escores de propensão.

<sup>9</sup> Ver Rosenbaum e Rubin (1983).

<sup>10</sup> Ver Caliendo e Kopeinig (2008).

A escolha da quantidade de unidades não tratadas usadas para compor o grupo de comparação deve ser feita de forma criteriosa, pois pode gerar viés e afetar a variância do estimador de impacto. É desejável minimizar tanto o viés quanto a variância do estimador, mas nem sempre é possível minimizar os dois simultaneamente. Usando apenas uma unidade não tratada para compor o grupo de comparação, por exemplo, a vantagem é que este será o mais parecido possível com o grupo de tratamento, reduzindo o viés. Entretanto, usando várias unidades não tratadas, reduz-se a variância do estimador (mais informação é usada para construir o contrafactual de cada unidade tratada).

### **Condições para que o método funcione**

Para que o método seja válido, é preciso assumir duas hipóteses.

**Hipótese de seleção em observáveis, ou de independência condicional:** a participação no programa pode ser considerada aleatória em relação aos resultados de interesse, condicionada a um conjunto de características relevantes. Essa hipótese não pode ser testada na prática e, em muitos contextos, pode ser considerada forte, pois, se houver qualquer característica não-observável que afete a probabilidade de um indivíduo se beneficiar da política e o resultado de interesse, o impacto estimado será enviesado.

**Hipótese de suporte comum:** para que o método do pareamento nos dê uma estimativa válida do impacto do programa, é preciso que cada unidade tratada tenha par com, pelo menos, uma unidade não tratada<sup>11</sup>.

Logo, o método de pareamento é recomendado apenas quando a seleção para participação se baseia, exclusivamente, em características observáveis. Outra desvantagem é a necessidade de um grande volume de informações disponíveis, pré-implementação da política, além de amostras grandes o suficiente, para a realização de testes estatísticos. Apesar disso, o método é bastante versátil, podendo ser utilizado em situações de diversos contextos.

### **Como implementar o método**

A implementação do método do pareamento passa pelas seguintes etapas:

- 1. Definir as características relevantes.** Devem ser consideradas características relevantes para a escolha de participar do programa (como ganhos futuros e custos associados) e para formar expectativas sobre o futuro. Vale ressaltar que são utilizadas as informações de características observáveis no momento anterior à política, ou seja, os grupos são semelhantes antes do grupo de tratamento ser beneficiado.
- 2. Definir o método: pareamento simples ou baseado no escore de propensão.** No pareamento simples, busca-se entre as unidades não participantes, aqueles que apresentem características observáveis semelhantes às observadas entre os indivíduos participantes. Na prática, estabelecer pares baseando-se no escore de propensão é mais factível, por resumir esse cruzamento em uma única medida ao invés de múltiplas variáveis.
- 3. Selecionar o grupo de comparação.** A escolha de unidades não tratadas que sejam parecidas às unidades tratadas envolve uma decisão sobre como fazer essa comparação entre elas, ou seja, sobre o critério de criar pares a ser realizado (vizinho mais próximo, por exemplo). Deve-se definir quantas unidades não tratadas devem ser selecionadas para compor o grupo de comparação para cada unidade tratada, e quão parecidas elas devem ser em termos da distância absoluta entre os escores de propensão. Além disso, deve-se considerar apenas unidades no suporte comum.
- 4. Verificar se os grupos estão balanceados.** Essa etapa é ainda mais relevante do que no método experimental, uma vez que as hipóteses são mais fortes. Espera-se que o cruzamento forneça grupos de tratamento e de comparação cujas características pré-implementação da política estejam balanceadas.
- 5. Calcular o impacto da política.** O impacto do programa será calculado considerando apenas indivíduos do grupo de tratamento que estão no suporte comum (formaram um par com um indivíduo não-beneficiário). O impacto é calculado a partir da comparação da média do indicador de impacto para o grupo de tratamento com a média do indicador de impacto para o grupo de comparação. Como no método experimental, pode-se utilizar regressões lineares, permitindo-se a inclusão de variáveis de controles e a utilização de funções dos escores de propensão como peso. O impacto estimado restringe-se à região do suporte comum.

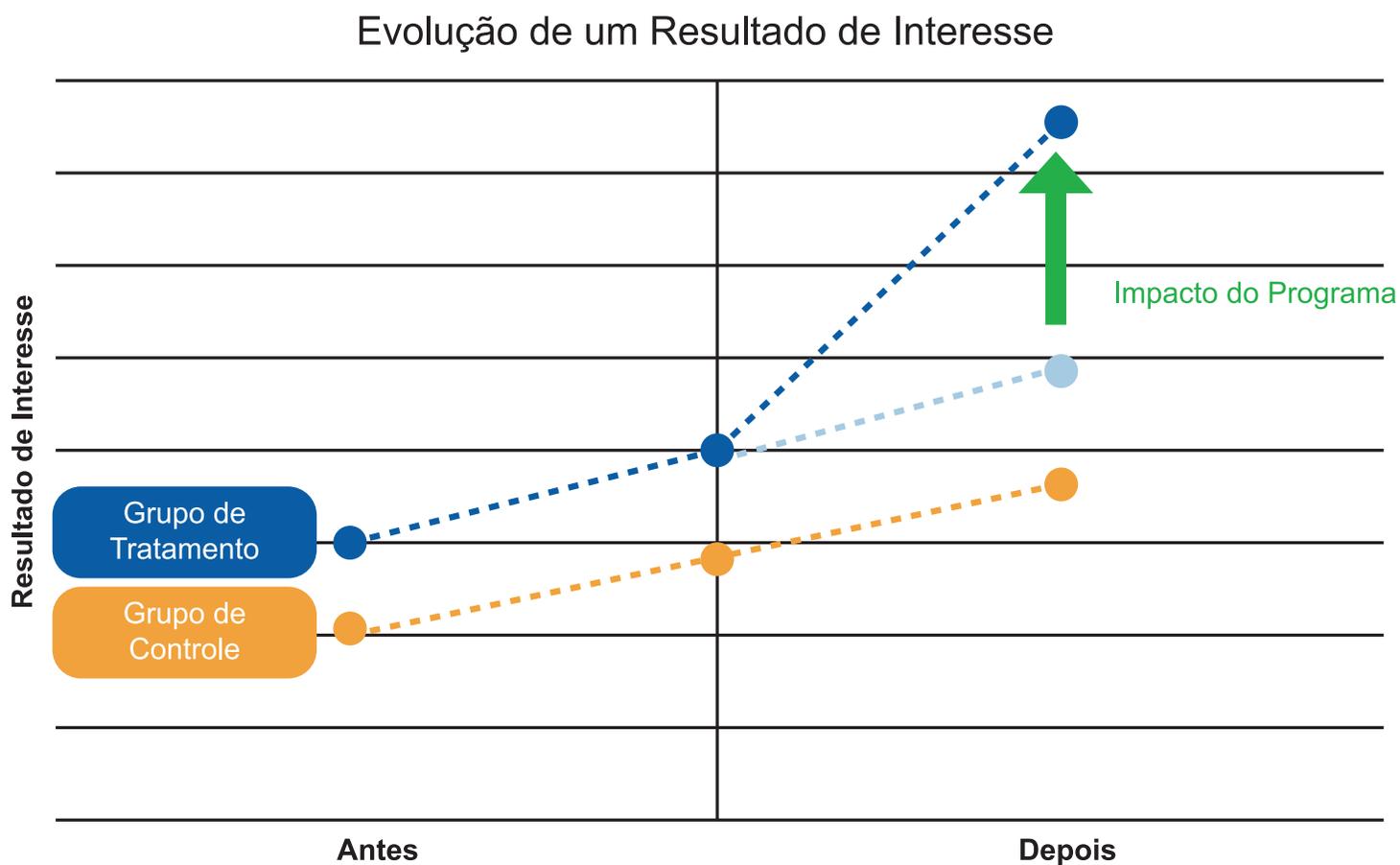
11 Ver Menezes Filho e Pinto (2017).

## DIFERENÇA EM DIFERENÇAS

O método de diferença em diferenças (ou dupla diferença) baseia-se na evolução, ao longo do tempo, do resultado de interesse do grupo de tratamento, comparado à evolução do grupo de comparação. Segundo esse método, há impacto positivo da política quando o resultado do grupo de tratamento melhora mais do que o resultado do grupo de comparação no período de implementação da política, conforme ilustrado na Figura 4.

Esse método pode ser utilizado mesmo quando os critérios de seleção da política não forem tão claros ou bem definidos, mas requer hipóteses mais fortes que outros métodos (como aleatorização e regressão descontínua) e também que estejam disponíveis dados da linha de base (momento anterior ao início do programa).

Figura 5: Método de Diferença em diferenças



Fonte: Elaboração própria.

A intuição do impacto via método de Diferença em diferenças pode ser resumida em duas etapas. Primeiro, calcula-se a primeira diferença (evolução do resultado de interesse ao longo do tempo, para os dois grupos):

$$\text{Grupo de tratamento: } Y_{GT}^1 - Y_{GT}^0$$

$$\text{Grupo de comparação: } Y_{GC}^1 - Y_{GC}^0$$

Em seguida, calcula-se a segunda diferença, que é diferença das diferenças da primeira etapa:

$$\text{Impacto} = (Y_{GT}^1 - Y_{GT}^0) - (Y_{GC}^1 - Y_{GC}^0)$$

Na prática, pode-se utilizar a seguinte regressão, que estimará o mesmo indicador e poderá ser realizado os testes de hipóteses para confirmar se o impacto estimado é diferente de zero:

$$Y_{it} = \alpha + \beta \text{Tratamento}_i + \gamma \text{Tempo}_i + \delta (\text{Tratamento}_i \times \text{Tempo}_i) + \varepsilon_{it}$$

Em que:  $Y_{it}$  é o indicador de impacto para o indivíduo  $i$  no momento de tempo  $t$ ,  $\text{Tratamento}_i$  é a variável binária indicativa de tratamento para  $i$ ,  $\text{Tempo}_i$  é uma variável binária indicativa de tempo (assume valor 1 para o momento depois e valor 0 para o momento antes da política);  $\varepsilon_{it}$  é o termo de erro. O impacto da política será dado pelo parâmetro estimado  $\delta$ .

Assim como nos demais métodos, outras variáveis de controle também podem ser acrescentadas à regressão. Essas variáveis podem afetar o indicador de impacto, mas não devem ser correlacionadas ao tratamento.

## Condições para que o método funcione

Uma importante vantagem do método é que os grupos de tratamento e comparação não precisam, necessariamente, apresentar as mesmas características antes do tratamento (podem partir de pontos diferentes). Ao tirarmos a primeira diferença (ao longo do tempo) para cada grupo, estamos eliminando o efeito de características tanto observáveis quanto não observáveis que são invariantes no tempo. No entanto, diferenças entre os grupos que variam ao longo do tempo não são eliminadas.

Para que o método forneça estimativas críveis de impacto, é preciso que a evolução, ao longo do tempo, do grupo de comparação represente as mudanças que teriam ocorrido com o grupo de comparação na ausência do tratamento (**hipótese de tendências paralelas**). Isso implica que, na ausência do tratamento, o grupo de tratamento e o grupo de comparação apresentariam mudanças que ocorreriam à mesma taxa no indicador de impacto, as trajetórias seriam semelhantes. Essa hipótese não pode ser testada, mas, se houver dados disponíveis, para mais de um momento, antes do início do programa, pode-se verificar se os grupos apresentavam tendências paralelas nesse período, o que dá indícios sobre a credibilidade dessa hipótese.

## Como implementar o método

- 1.** *Definir o grupo de comparação.* O grupo de comparação deve atender a hipótese de tendências paralelas, ou seja, a evolução do indicador de impacto ao longo do tempo representa a evolução que o grupo de tratamento teria na ausência da política. Além disso, os grupos de tratamento e de comparação não podem ser afetados de forma distinta por outros fatores concomitantes à política.
- 2.** *Verificar o balanceamento das variáveis no momento anterior à política.* Apesar de o método não requerer que haja balanceamento prévio, conhecer as variáveis que não estão balanceadas pode ser importante para calcular o impacto.
- 3.** *Calcular o impacto da política.* O impacto pode ser calculado através da estimativa das duas diferenças (antes e depois, e tratamento e comparação), ou via regressão linear, viabilizando a inclusão de outras variáveis de controle que possam ter afetado o resultado de interesse.

## POSSIBILIDADE DE AVALIAÇÃO DE IMPACTO EM CABO VERDE: Programa Transformação Agricultura<sup>3</sup>

O programa Transformação Agricultura consiste na implementação de um sistema de irrigação gota a gota, fornecimento de assistência técnica e financiamento. Os beneficiários serão 500 pequenos agricultores da região de Santiago Norte, região com alto potencial agrícola e elevado nível de pobreza. O programa tem um orçamento inicial de 100.000.000 ECV, sendo o custo de implementação do sistema de irrigação de 100.000 ECV por agricultor e um pacote de financiamento de 50.000.000 ECV.

Espera-se que, com esse programa, os agricultores aumentem a produtividade e, conseqüentemente, os rendimentos e os empregos formais fornecidos no campo. Os dados para estimativa do impacto serão coletados através do fornecimento do sistema de irrigação ao fim de dois anos.

O método de diferença em diferenças pode ser utilizado para avaliar o impacto do programa, considerando a hipótese de tendências paralelas:

*Grupo de tratamento:* pequenos agricultores beneficiados.

*Grupo de comparação:* pequenos agricultores não-beneficiados.

Para encontrar o grupo de comparação, o método de diferença em diferenças pode ser combinado com outros métodos, como a aleatorização ou o pareamento.

## REGRESSÃO DESCONTÍNUA

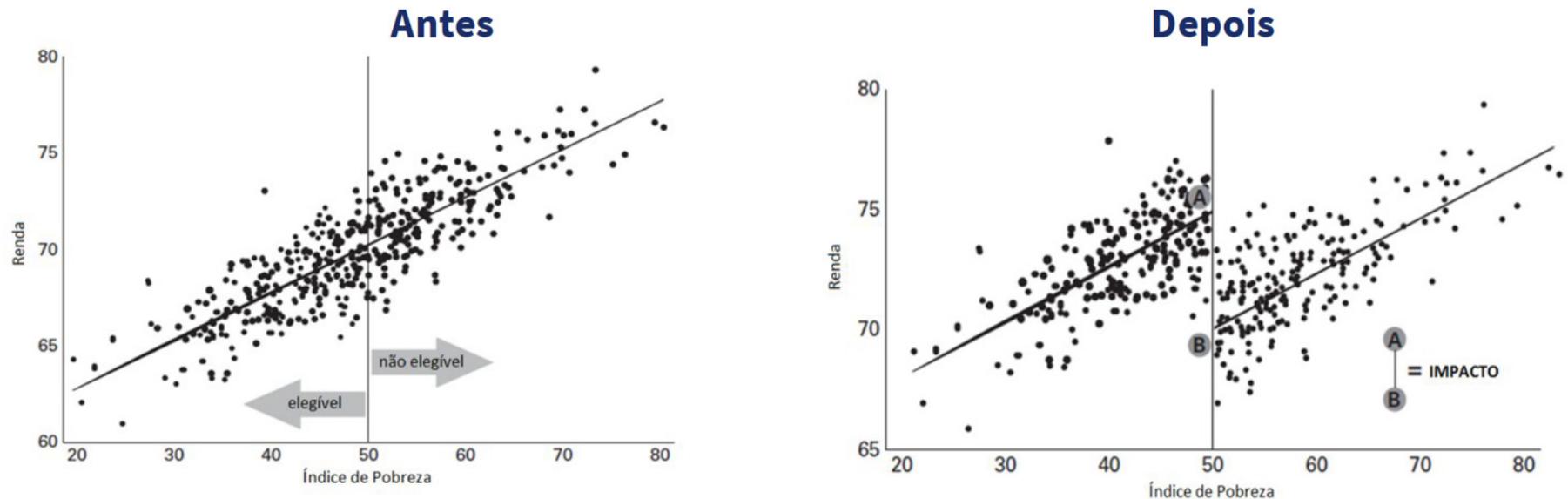
Na definição de critérios de elegibilidade para a participação em uma política pública, é usual estabelecer um ponto de corte, ou ponto de referência, em algum índice, como um indicador de pobreza, renda, ou nota em um exame. Quando há um ponto de corte bem definido em relação a um índice contínuo, de forma que todos os indivíduos ou unidades que estejam acima/abaixo deste ponto sejam elegíveis ou não elegíveis para participar da política, o método da regressão descontínua pode ser utilizado para estimar o impacto. Em especial, é estimado o efeito médio local do tratamento, pois explora-se a descontinuidade presente ao redor do ponto de corte para identificar o grupo de comparação e estimar o contrafactual.

A intuição do método da regressão descontínua refere-se à similaridade de indivíduos ou unidades, cujos valores do índice sejam muito próximos: indivíduos logo abaixo e logo acima do ponto de corte tendem a ser muito similares. Logo, pode-se escolher como grupo de comparação indivíduos que não são elegíveis, mas que estão próximos o suficiente do ponto de corte. Imagine, por exemplo, uma política de transferência de renda para populações em situação de vulnerabilidade social que considere como critério de elegibilidade para ser beneficiado, famílias cuja renda per capita esteja abaixo de 50 em um índice de pobreza pré-definido. Assim, uma família com índice igual a 49 é elegível para receber a transferência, mas outra família com 51 não é elegível. É razoável pressupor que essas duas famílias são muito parecidas em termos de pobreza e vulnerabilidade, exceto pela possibilidade de participar do programa. Dessa forma, famílias elegíveis que estejam próximas do ponto de corte (com índice um pouco abaixo de 50) são definidas como grupo de tratamento para a avaliação com regressão descontínua, e famílias não elegíveis também próximas do ponto de corte (com índice um pouco acima de 50) compõem o grupo de comparação. Ao redor do ponto de corte, é como se a ordem tivesse sido determinada de forma aleatória, por isso este método é considerado quase-experimental.

Para estimar o impacto, é preciso, ainda, definir uma janela ao redor do ponto de corte, para determinar quais os indivíduos que serão incluídos nas estimativas – o grupo de tratamento e o grupo de comparação. O tamanho da janela está relacionado com o balanceamento de outras características observáveis relevantes entre os grupos. Isto é, o tamanho da janela deve ser tão grande quanto o necessário para haver um número de observações suficiente para realizar as estimativas, mas não a ponto de os grupos de tratamento e de comparação não serem mais semelhantes. No caso da política de transferência de renda, enquanto a família com índice 51 deve ser muito similar àquela com índice 49, não se espera o mesmo entre uma família com índice 70 e outra com índice 30, por exemplo.

O impacto é calculado a partir da medição das diferenças no indicador de impacto após a política entre as unidades que estão próximas do ponto de corte. A Figura 5 ilustra a descontinuidade em um índice contínuo de elegibilidade, anterior à política, e respectivo impacto observado no momento depois. O salto observado ao redor da descontinuidade após a implementação da política pode ser interpretado como impacto. Os resultados de impacto podem ser sensíveis ao tamanho da janela escolhida.

Figura 6: Método de Regressão descontínua



Fonte: Gertler et al (2018).

Há dois tipos de regressão descontínua:

**Sharp:** Neste caso, a condição de participante/elegível ou não para a política é definida de forma determinística pelo índice contínuo. A variável  $T_i$  de participação na política (tratamento) depende de forma determinística do resultado do valor do índice contínuo  $Z_i$  em relação ao ponto de corte  $c$ . Ou seja,  $T_i = 1$  sempre que  $Z_i \leq c$ , e  $T_i = 0$  sempre que  $Z_i > c$ , por exemplo. A probabilidade de ser uma unidade tratada (beneficiada pela política) passa de 0 a 1 ao redor do ponto de corte.

**Fuzzy:** Quando a descontinuidade é deste tipo, a probabilidade de participar aumenta ou diminui dependendo do desempenho no índice contínuo  $Z_i$ , mas não se trata de um salto entre 0 e 1 como no caso *sharp*. A variável  $T_i$  de participação na política (tratamento) é uma função do índice  $Z_i$  e é descontínua no ponto de corte  $c$ . Por exemplo, são elegíveis para receber a transferência de renda famílias cujo índice de pobreza esteja abaixo de 50 e tenham crianças.

Em ambos os casos, estima-se o impacto da participação no programa sobre o indicador de impacto de interesse a partir da comparação entre as unidades ao redor do ponto que define a descontinuidade, por meio da comparação entre o grupo de tratamento e o grupo de comparação.

### Condições para que o método funcione

O método de regressão descontínua pode ser utilizado na avaliação de políticas que atendam às seguintes condições no que se refere à seleção dos participantes:

- Há um índice contínuo que pode ser utilizado para classificar a população de interesse, que são os participantes em potencial (hipótese da continuidade).
- Há um ponto de corte bem definido que corresponde a um ponto nesse índice em relação ao qual todos os indivíduos acima ou abaixo dele são classificados como sendo elegíveis ou não para participar do programa, o tratamento pode ser interpretado como sendo definido de forma aleatória (hipótese da ignorabilidade social).

Além disso, o método só deve ser utilizado quando não há possibilidade de manipulação do índice pelos participantes em potencial. Se houver manipulação, as estimativas de impacto apresentariam um viés.

### POSSIBILIDADE DE AVALIAÇÃO DE IMPACTO EM CABO VERDE: Rendimento Social de Inclusão <sup>4</sup>

Iniciativa do Ministério da Família e Inclusão Social, o programa Rendimento Social de Inclusão (RSI) prevê uma transferência mensal de 5.500 escudos cabo-verdianos para as famílias das classes mais baixas. Para que a família seja elegível, ela deve reunir as seguintes condições: estar no grupo 1 do escore no índice de pobreza, estar inscrito no Cadastro Social Único (com inscrição válida) e contar com uma criança de até 15 anos.

O programa oferece três componentes: um componente monetário, que consiste em uma prestação mensal de 5.500\$00 de carácter temporário de 2 anos, para satisfação das necessidades básicas; acompanhamento familiar, que consiste em encontros individuais ou em grupo de informação e formação para acesso a bens e serviços básicos; e inclusão produtiva, que consiste em acções na área de emprego e formação profissional.

Uma vez que o programa usa um índice contínuo não-manipulável como critério de elegibilidade – o escore de pobreza –, o método da regressão descontínua pode ser utilizado para estimar o impacto local na redução da pobreza, ou em outros indicadores de interesse. Para tanto, define-se a amostra da avaliação:

**Grupo de tratamento:** famílias com inscrição válida no Cadastro Social Único e com criança até 15 anos, logo abaixo do ponto de corte no índice de pobreza e, portanto, receberam o benefício.

**Grupo de comparação:** famílias com inscrição válida no Cadastro Social Único e com criança até 15 anos, logo acima do ponto de corte no índice de pobreza e, portanto, não receberam o benefício.

O tamanho da janela no índice de pobreza que definirá o tamanho dos grupos de tratamento e comparação dependerá de até que ponto os grupos continuam muito parecidos.

## Como implementar o método

De forma simplificada, os passos para implementar o método de regressão descontínua podem ser descritos como:

**1.** Identificar a descontinuidade no ponto de corte do critério de seleção. Em geral, a descontinuidade pode ser observada de forma gráfica, de acordo com os critérios definidos pela política, conforme ilustrado na Figura 5. É importante definir também se se trata de um caso *sharp* ou *fuzzy*.

**2.** Selecionar o tamanho da janela ao redor do ponto de corte para a definição dos grupos de tratamento e comparação da avaliação. Tendo em mente que os grupos de tratamento e comparação precisam ser muito semelhantes nas características relevantes, o tamanho da janela será aquele que garanta o balanceamento nessas características entre os grupos.

**3.** Estimar o impacto da política ao redor do ponto de corte. A estimativa do impacto envolve a estimativa de regressões locais separadas para cada lado do ponto de corte, e o impacto é dado pela diferença dos parâmetros estimados entre elas. Vale ressaltar que este impacto, que é o efeito médio local do tratamento, é válido apenas para o intervalo definido pelo tamanho da janela ao redor do ponto de corte, não podendo ser generalizado para o restante da distribuição.

## CÁLCULO DO IMPACTO – REGRESSÃO DESCONTÍNUA

De forma simplificada, a implementação do método se dá pela estimativa de regressões lineares locais para cada lado do ponto de corte  $c$ . Considerando o caso *sharp* citado anteriormente, em que  $T_i = 1$  sempre que  $Z_i \leq c$ , e  $T_i = 0$  sempre que  $Z_i > c$ , uma janela ao redor do ponto de corte de tamanho  $h$ :

$$Y_i = \alpha_{GT} + \beta_{GT}(Z_i - c) + \varepsilon_i, \quad \text{se } c - h < Z < c$$

$$Y_i = \alpha_{GC} + \beta_{GC}(Z_i - c) + \varepsilon_i, \quad \text{se } c \leq Z < c + h$$

O impacto, medido como o efeito médio local do tratamento, é dado pela diferença entre os interceptos das regressões:

$$\text{Impacto} = \alpha_{GT} - \alpha_{GC}$$

No caso de descontinuidade *fuzzy*, é preciso estimar as probabilidades de tratamento em cada lado do ponto de corte.

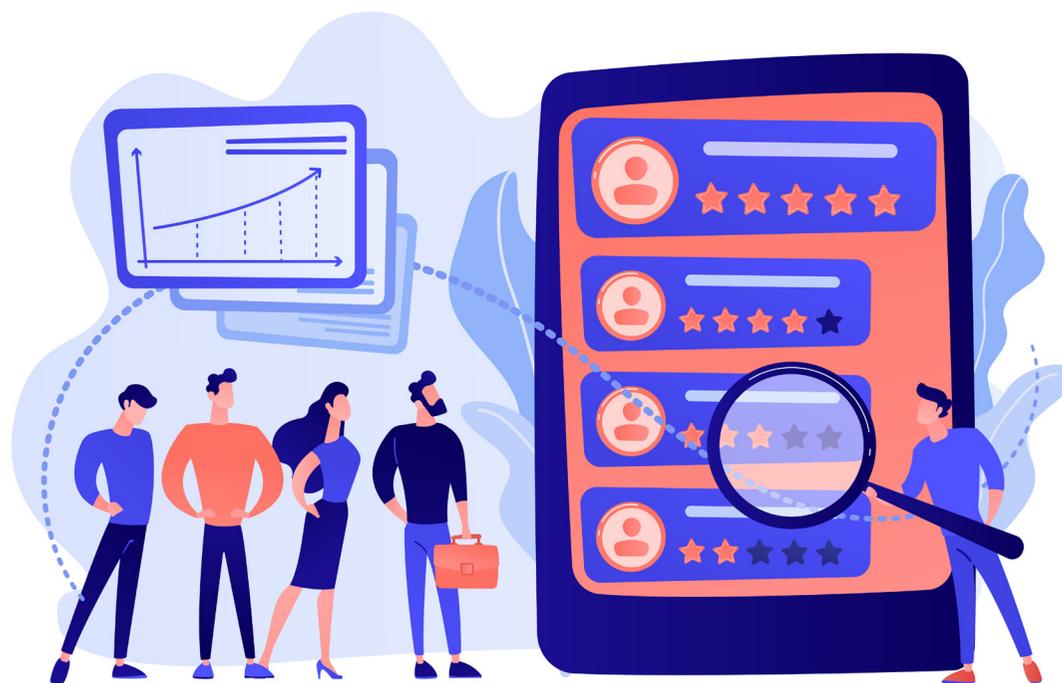
## COMPARAÇÃO ENTRE MÉTODOS

Como analisado ao longo deste capítulo, há diversas possibilidades de metodologias para executar uma avaliação de impacto de uma política pública. Todas elas implicam a escolha de um grupo de comparação semelhante a um grupo de tratamento, participante da política, de forma a estimar o contrafactual, qual teria sido o resultado do grupo de tratamento na ausência da política. Com isso, é possível estabelecer o impacto, o efeito causal da política no resultado de interesse.

Contudo, dependendo dos critérios de seleção para participação na política e das hipóteses que podem ser assumidas, alguns métodos podem ser preferíveis a outros em cada situação. De uma forma geral, o método experimental, que utiliza aleatoriedade para alocar as unidades entre os grupos de tratamento e controle, é tido como o padrão-ouro da avaliação de impacto, pois garante a eliminação do viés de seleção, quando há um número suficientemente grande de observações no sorteio. Esse método é apropriado em situações de excesso de demanda, já que se constitui como uma forma justa de alocar participantes. Por outro lado, pode haver impedimentos éticos ou legais para sua execução, além de precisar ser desenhado concomitantemente à formulação da política.

Quando não é possível realizar a aleatorização, ou quando se deseja avaliar políticas já em curso ou finalizadas, os métodos não experimentais fornecem alternativas para a estimativa do impacto causado por uma política. Em todo caso, é preciso analisar os critérios de seleção da política, a disponibilidade de dados e as condições para que cada método funcione. O Quadro 2 sintetiza, de forma comparativa, as informações sobre cada método, suas hipóteses, vantagens e desvantagens.

Vale ressaltar que, muitas vezes, é possível utilizar uma combinação de métodos na avaliação. Se, por exemplo, a aleatorização não envolve um número suficientemente grande de observações, de forma que os grupos não estejam balanceados no momento anterior à política, a combinação com o método de diferença em diferenças pode anular diferenças preexistentes entre os grupos, contanto que estas sejam invariantes no tempo. A combinação do método do pareamento com diferença em diferenças também é útil para minimizar as limitações desse cruzamento com base exclusivamente em características observáveis.



Quadro 1: Comparativo dos métodos

Método	Grupo de comparação	Hipóteses/Condições para que funcione	Vantagens	Desvantagens
Aleatorização	Indivíduos que não foram sorteados para participar.	Os dois grupos são estatisticamente idênticos em relação a características observáveis e não observáveis (isto é, a aleatorização funciona, o que pode ser garantido desde que os grupos sejam grandes o suficiente).	<ul style="list-style-type: none"> <li>• Forma justa e transparente de alocar recursos escassos entre indivíduos igualmente merecedores, uma vez que todos os indivíduos elegíveis têm a mesma oportunidade de ser beneficiado;</li> <li>• Fácil comunicação ao público e fácil execução;</li> <li>• Quando ocorre, é possível avaliar o programa da forma mais robusta possível, uma vez que se elimina o viés de seleção;</li> <li>• A aleatorização garante que os grupos de tratamento e de controle sejam similares antes de se implementar o programa, tanto nas características observáveis quanto nas não-observáveis.</li> </ul>	<ul style="list-style-type: none"> <li>• Pode não fazer sentido se não houver excesso de demanda e/ou escassez de recursos;</li> <li>• Pode não ser viável do ponto de vista ético;</li> <li>• Pode ocorrer mudanças no comportamento dos grupos ao longo do tempo (ex 1: indivíduos sorteados para participar podem desistir; ex 2: indivíduos não sorteados para participar podem procurar substitutos para o programa; ex 3: transbordamento do tratamento);</li> <li>• A semelhança entre grupos só é garantida quando há um número suficientemente grande de participantes em potencial, o que nem sempre ocorre.</li> </ul>
Pareamento	Não participantes que apresentam probabilidade parecida de participar em relação à observada entre os participantes (a partir de características observáveis).	Não devem existir variáveis omitidas que enviesem o resultado (ex: característica diferente entre os dois grupos ou característica correlacionada ao resultado).	<ul style="list-style-type: none"> <li>• Método versátil, que pode ser aplicado em vários contextos e independentemente dos critérios de seleção;</li> <li>• Pode ser combinado a vários métodos.</li> </ul>	<ul style="list-style-type: none"> <li>• São necessárias muitas informações sobre os indivíduos;</li> <li>• Só é aplicável em casos em que há um número relativamente grande de indivíduos envolvidos;</li> <li>• A hipótese de suporte comum pode não ser válida;</li> <li>• Não é possível incorporar características não observáveis, de forma que é preciso considerar se o grupo de tratamento e o grupo de comparação potencial não apresentam diferenças sistemáticas em relação a elas – hipótese forte que não pode ser testada;</li> <li>• Método menos robusto.</li> </ul>

<p>Diferença em diferenças</p>	<p>Não participantes para os quais se tem informação em dois momentos de tempo (antes e depois do programa).</p>	<p>Na ausência do programa, ambos os grupos apresentariam evoluções semelhantes ao longo do tempo (tendências paralelas).</p>	<ul style="list-style-type: none"> <li>• Método versátil, que pode ser aplicado em vários contextos e independentemente dos critérios de seleção;</li> <li>• grupo de tratamento e grupo de comparação não precisam necessariamente partir do mesmo ponto antes do início do programa;</li> <li>• Capaz de eliminar efeitos tanto de características observáveis quanto de não-observáveis que são invariantes no tempo.</li> </ul>	<ul style="list-style-type: none"> <li>• Método menos robusto e com hipóteses mais fortes que os métodos experimentais e regressão descontínua;</li> <li>• Não é possível testar a hipótese de tendências paralelas, que é essencial para que a estimativa de impacto seja não enviesada;</li> <li>• Se houver outros fatores simultâneos à realização do programa que afetem e/ou causem diferenças nas tendências, as estimações serão enviesadas.</li> </ul>
<p>Regressão descontínua</p>	<p>Comparar indivíduos ao redor do ponto de corte usado para definir a participação no programa.</p>	<p>Deve haver um ponto de corte bem definido e os candidatos não devem poder manipular seu desempenho em relação a essa característica.</p>	<ul style="list-style-type: none"> <li>• À medida que nos aproximamos do ponto de corte, os indivíduos à esquerda e à direita dele ficam cada vez mais similares, de forma a que o grupo de comparação seja considerado aleatório (método quase-experimental);</li> <li>• Se a pergunta que se pretende responder com a avaliação de impacto é do tipo “devemos reduzir ou expandir o programa na margem?”, esse método produz exatamente a estimativa local do tratamento para ser utilizada nessa decisão.</li> </ul>	<ul style="list-style-type: none"> <li>• Esse método só pode ser utilizado quando os participantes em potencial não conseguem manipular seu desempenho no índice em questão;</li> <li>• O impacto estimado é válido apenas na vizinhança do ponto de corte;</li> <li>• Se os indivíduos nessa vizinhança forem muito diferentes do restante do público-alvo, não podemos garantir que o impacto estimado seja igual para esse público como um todo.</li> </ul>

## Validade interna e validade externa

Ao planejar e executar a avaliação de impacto de uma política, os conceitos de validade interna e validade externa podem contribuir no processo de escolha de um método, ou combinação de métodos.

O conceito de **validade interna** refere-se à identificação de relações causais, ou seja, se a avaliação realizada, de fato, consegue estabelecer uma relação de causa e efeito entre a política avaliada e as mudanças observadas no resultado de interesse. Ou seja, é importante refletir se, de fato, foi identificada uma relação causal entre a política e os resultados de interesse.

Quando há validade interna, é possível garantir que as transformações observadas no(s) resultado(s) de interesse sejam fruto exclusivo da política. A validade interna, por exemplo, é garantida quando há aleatoriedade do tratamento. As avaliações com o método de regressão descontínua também possuem alta validade interna. Possíveis fontes de riscos para a validade interna são:

- Variação nos intervalos de coleta de informações;
- Atrito (quando integrantes do grupo de tratamento ou do grupo de controle não são mais encontrados e não é possível medir os resultados de interesse);
- Diferenças sistemáticas entre grupos de tratamento e de comparação;
- Ocorrência de outros programas/projetos ou fatos concomitantes.

Por outro lado, a validade interna não garante que seja possível generalizar os resultados encontrados. O conceito de **validade externa** refere-se à possibilidade de generalizar o efeito causal para outras populações, localidades e/ou períodos de tempo. Ou seja, é a possibilidade de afirmar que o efeito causal estimado pela avaliação seria o mesmo se a política fosse realizada em outro contexto. Para que haja validade externa, é preciso que a amostra usada na avaliação seja representativa do público elegível. Possíveis fontes de riscos para a validade externa são:

- Particularidades específicas do contexto da análise;
- Pequena variação disponível para análise em termos de intensidade e duração do tratamento;
- Diferenças sistemáticas entre o grupo que foi analisado e outras populações elegíveis.

Em muitos casos, há um dilema entre validade interna e validade externa: buscando-se garantir um, ocorre o comprometimento do outro. Experimentos controlados, por exemplo, costumam ter alta validade interna, mas nem sempre a amostra utilizada na avaliação é representativa da população no geral.

“ *O conceito de validade interna refere-se à identificação de relações causais, ou seja, se a avaliação realizada, de fato, consegue estabelecer uma relação de causa e efeito entre a política avaliada e as mudanças observadas no resultado de interesse.* ”

# *CAPÍTULO 3*

**ANÁLISES DE RETORNO  
ECONÓMICO E SOCIAL**





As **análises de retorno econômico e social** podem complementar as evidências geradas por uma avaliação de impacto, por fornecerem informações relevantes à tomada de decisão de políticas públicas. Essas análises ajudam a responder como os custos da política se comparam aos benefícios gerados para a sociedade, ou se o impacto causado compensa os custos de implementação. Além disso, podem ser úteis para comparar políticas ou projetos com objetivos semelhantes, apontando quais alternativas são mais eficientes.

Assim, a análise de retorno econômico tem como objetivos qualificar os resultados da avaliação de impacto, dar contexto e significado para o impacto estimado, e situar a política e seus resultados em um contexto econômico relevante. Dois métodos de realizar esta análise, abordados neste capítulo, são custo-benefício e custo-efetividade.

12 Ver Menezes e Filho (2017).

13 Ver J-PAL (2016).

## ANÁLISE DE CUSTO-BENEFÍCIO

Por **análise de custo-benefício entende-se** a comparação direta dos custos e benefícios gerados pela política, com o objetivo de identificar a sua viabilidade econômica:

$$\textit{Benefício Líquido} = \textit{Benefício Total} - \textit{Custo Total}$$

A política é considerada economicamente viável se os benefícios totais compensam os custos, ou seja, o benefício líquido é maior do que zero.

Contudo, muitas vezes os custos e benefícios associados a políticas públicas não estão diretamente expressos em unidades monetárias. A **monetização** é o processo de traduzir algo em valores monetários: para comparar os custos e benefícios da política pública, é preciso que estejam expressos na mesma unidade de medida (dinheiro). Todos os valores precisam ser expressos em valor presente<sup>12</sup>.

Se o indicador de impacto for não-pecuniário, será preciso realizar hipóteses para monetizá-lo, baseando-se na literatura sobre o tema. Para isso, deve-se realizar suposições sobre o quanto o impacto (não-monetário) causado pela política, proporciona ganhos monetários para a sociedade, ou o quanto é economizado do ponto de vista social. Por exemplo, se uma política gera como impacto uma maior escolaridade da população, é razoável supor que isso implicará em ganhos salariais futuros, que podem ser estimados em termos monetários. Porém, nem sempre é possível monetizar todos os benefícios da política. No processo de monetização, também deve-se determinar qual a duração do benefício: curto, médio ou longo prazo.

Em relação aos custos, todos os tipos de custos associados à política precisam ser considerados, inclusive aqueles menos evidentes e não-monetários<sup>13</sup>. Os custos contábeis da política, que se referem às despesas para implementação, devem constar em dados administrativos. Todos os custos contábeis, operacionais da política, devem ser considerados, independentemente da fonte de financiamento (setor público, financiador externo, doações, etc.).

Além dos custos contábeis, também são incluídos os custos de oportunidade, que são o valor dos usos alternativos dos recursos destinados à política. O custo de oportunidade é um custo implícito, em que se responde como aquele recurso poderia ser utilizado se não estivesse destinado à política.

A análise de retorno da política envolve incertezas em relação à magnitude dos custos e benefícios e à duração deles. A estimativa pode variar de acordo com acontecimentos futuros. Na estimativa dos custos e benefícios, é apropriado imaginar três cenários: otimista, o melhor resultado possível; realista, o resultado intermediário; e pessimista, o menor resultado possível.

## ANÁLISE DE CUSTO-EFETIVIDADE

A **análise de custo-efetividade** também busca averiguar se os benefícios gerados pela política compensam seus custos, mas não há necessidade de esses benefícios estarem expressos em unidades monetárias<sup>14</sup>. Este método é útil para relacionar o impacto gerado pela política e seus custos quando não é possível (ou pertinente) calcular o benefício monetário do projeto.

Seu objetivo é comparar a eficiência de diferentes programas/projetos que tenham objetivos similares e o mesmo indicador de impacto. A razão custo-efetividade identifica o investimento necessário para obter cada unidade de impacto não-monetário:

*Razão Custo – Efetividade = Custo Total / Impacto*

Logo, essa análise aponta o quanto é preciso investir para alcançar cada unidade de impacto causal.

O ideal é que este impacto seja estimado através das metodologias de avaliação, analisadas no capítulo anterior. Contudo, se não houver avaliação de impacto, uma opção é considerar o custo por unidade de produto ou resultado, por exemplo, custo por beneficiário que recebeu a política, custo por refeição entregue, etc.

A análise de custo-efetividade é útil para realizar comparação do valor relativo das alternativas de políticas. É um bom método para comparar políticas com o mesmo objetivo, apontando aquelas que são mais custo-efetiva, ou seja, menos custosas por unidade de impacto. Ademais, uma vez realizada uma avaliação de impacto robusta e estando disponíveis informações sobre os custos da política, a execução deste tipo de análise gera evidências úteis para tomada de decisões e eficiência dos gastos públicos.

Por outro lado, diferente da análise de custo-benefício, a avaliação de custo-efetividade não permite avaliar a viabilidade econômica da política.



*...ideal é que este impacto seja estimado através das metodologias de avaliação, analisadas no capítulo anterior. Contudo, se não houver avaliação de impacto, uma opção é considerar o custo por unidade de produto...*

# CONCLUSÃO



No contexto de políticas públicas baseadas em evidências, a avaliação de impacto apresenta-se como um tipo de avaliação ex post quantitativa, que, por meio de técnicas econométricas, permite identificar o efeito causal de uma política sobre resultados de interesse. Estabelecer relações de causalidade, tal qual realizado por avaliações de impacto robustas, é uma etapa imprescindível para tomar decisões a respeito de políticas. Por outro lado, os resultados de uma avaliação de impacto só poderão ser analisados, de forma completa, se acompanhados de dados de outras etapas do ciclo da política, como avaliação de processo e dados de monitoramento. Se uma política gera impactos positivos, é preciso analisar os canais de transmissão; se não gera impactos, é preciso identificar as falhas de desenho ou implementação da política.

Neste guia, foram discutidas as principais metodologias de avaliação de impacto, e também de análises de retorno econômico e social que complementam a avaliação. Todas as metodologias buscam isolar o efeito da política do efeito de outros fatores. A escolha da metodologia a ser utilizada dependerá dos critérios de seleção para participação na política e na disponibilidade de informações.

Vale ressaltar que, quando a avaliação de impacto é planejada antes da implementação da política, conseguimos utilizar métodos mais robustos que nos permitem estimar o impacto de forma mais acurada. O método experimental, que é o padrão-ouro da avaliação de impacto, só poderá ser implementado se houver aleatoriedade na definição de quem serão os participantes/beneficiários. Caso contrário, são viáveis apenas opções de métodos menos robustos, que exigem hipóteses mais fortes, o que prejudica a qualidade das respostas obtidas. Logo, sempre que possível, o desenho da avaliação de impacto – e da análise de retorno econômico e social – deve ser feito durante a fase de planejamento da política. Ademais, é preciso haver uma gestão da implementação da avaliação, de maneira a minimizar os riscos que podem vir a comprometer os resultados, como contaminação e atrito.

Finalmente, uma vez realizada a avaliação de impacto de uma política, com a utilização de técnicas científicas robustas e fontes de dados confiáveis, é preciso disseminar e, efetivamente, incorporar as evidências geradas à gestão. Os resultados serão informações relevantes para tomar decisões relativas à eventual continuidade de uma política, sua expansão, modificação ou mesmo conclusão. A disseminação dos resultados deve ser transparente e adequada aos diferentes públicos, que incluem pesquisadores, gestores e a população em geral. Afinal, o objetivo geral da implantação de políticas públicas baseadas em evidências deve ser o aperfeiçoamento dos serviços prestados à sociedade para a solução de problemas públicos.

# **REFERÊNCIAS** **BIBLIOGRÁFICAS**



ANGRIST, J.D.; PISCHKE, J.S. Mostly harmless econometrics: An empiricist's companion. Princeton University Press, 2008.

CALIENDO, M.; KOPEINIG, S. Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys, 22(1), pp.31-72, 2008.

DHALIWAL, I., DUFLO, E., GLENNERSTER, R., TULLOCH, C. Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education. Abdul Latif Jameel Poverty Action Lab (J-PAL), MIT, 2012. Disponível em: <<https://economics.mit.edu/files/15014>>

GERTLER, P.J., MARTINEZ, S., PREMAND, P.; RAWLINGS, L.B. Avaliação de Impacto na Prática, 2a Edição. World Bank Publications, 2018. Disponível em: <<https://publications.iadb.org/publications/portuguese/document/Avalia%C3%A7%C3%A3o-de-impacto-na-pr%C3%A1tica-Segunda-edi%C3%A7%C3%A3o.pdf>>

IMAS, L. G.; RIST, R. The road to results: Designing and conducting effective development evaluations. The World Bank, 2009.

INSTITUTO JONES DOS SANTOS NEVES. Guia para Avaliar Políticas Públicas | volume 4. E quando a política está em andamento? Avaliação ex post! Vitória, 2018. Disponível em <<http://www.ijns.es.gov.br/component/attachments/download/6421>>

J-PAL - ABDUL LATIF JAMEEL POVERTY ACTION LAB. J-PAL Costing Guidelines. J-PAL. 2016. Disponível em: <<https://www.povertyactionlab.org/sites/default/files/resources/costing-guidelines.pdf>>.

MENEZES FILHO, N.; PINTO, C.C.X. (organizadores). Avaliação Econômica de Projetos Sociais, 3a Edição. Fundação Itaú Social, 2017. Disponível em: <[https://www.itausocial.org.br/wp-content/uploads/2018/05/avaliacao-economica-3a-ed\\_1513188151.pdf](https://www.itausocial.org.br/wp-content/uploads/2018/05/avaliacao-economica-3a-ed_1513188151.pdf)>

OCDE. Organisation for Economic Co-operation and Development. Development Assistance Committee. Working Party on Aid Evaluation. Glossary of key terms in evaluation and results-based management, 2002. Disponível em: <<https://www.oecd.org/dac/evaluation/2754804.pdf>>

ROSENBAUM, P.R.; RUBIN, D.B. The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), pp.41-55, 1983.

RUBIN, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5), 688, 1974.

# FOOTNOTES



- 1.** As informações sobre o programa foram obtidas durante o Curso de Introdução à Avaliação de Impacto para Cabo Verde, realizado em janeiro de 2021 pelo FGV EESP Clear. A proposta de avaliação de impacto aqui apresentada é baseada nessas informações e no que foi desenvolvido por trabalho em grupo no curso.
- 2.** As informações sobre o programa foram obtidas durante o Curso de Introdução à Avaliação de Impacto para Cabo Verde, realizado em janeiro de 2021 pelo FGV EESP Clear. A proposta de avaliação de impacto aqui apresentada é baseada nessas informações e no que foi desenvolvido por trabalho em grupo no curso.
- 3.** As informações sobre o programa foram obtidas durante o Curso de Introdução à Avaliação de Impacto para Cabo Verde, realizado em janeiro de 2021 pelo FGV EESP Clear. A proposta de avaliação de impacto aqui apresentada é baseada nessas informações e no que foi desenvolvido por trabalho em grupo no curso.
- 4.** As informações sobre o programa foram obtidas durante o Curso de Introdução à Avaliação de Impacto para Cabo Verde, realizado em janeiro de 2021 pelo FGV EESP Clear. A proposta de avaliação de impacto aqui apresentada é baseada nessas informações e no que foi desenvolvido por trabalho em grupo no curso.







**Ministério das Finanças  
e do Fomento Empresarial**  
Direção Nacional do Planeamento

