

AVALIAÇÃO DE IMPACTO: Métodos e conceitos

O principal objetivo do presente boletim é apresentar, de maneira mais conceitual e menos teórica, duas metodologias para mensurar o impacto de um programa ou política pública. Para exemplificar essas metodologias, foi criado um banco de dados fictício, a partir do qual iremos aplicar essas metodologias, com o objetivo de averiguar a assertividade dos resultados.

A PRIMEIRA PERGUNTA QUE SURGE É, POR QUE SE PREOCUPAR EM FAZER UMA AVALIAÇÃO DE IMPACTO?

A resposta é simples: gestão! Ela nos trará indicativos de sucesso, deficiências, melhor alocação de recursos e retorno do valor investido. Não é razoável que um programa seja lançado e que os gestores simplesmente “cruzem os dedos” para que ele dê certo.

NÃO CONFUNDA AVALIAÇÃO DE IMPACTO COM ANÁLISE DE MONITORAMENTO DE INDICADORES!

São análises completamente diferentes. Uma avaliação de impacto mensura o efeito causal exclusivo da intervenção do programa, eliminando todos os possíveis efeitos externos. Já o monitoramento de indicadores, é importante, mas não pode ser usado para informar o impacto. Ele trata apenas da evolução (ou não) do indicador mensurado.

QUANTAS METODOLOGIAS DE AVALIAÇÃO DE IMPACTO EXISTEM?

Existem muitos métodos disponíveis e a escolha de cada um depende da especificidade do programa, da estrutura dos dados, bem como da qualidade e quantidade de informação. Os métodos podem ser divididos em dois grupos: experimental e quase-experimental. Os métodos experimentais são ensaios aleatórios controlados. Já os métodos quase-experimentais, tentam imitar ensaios aleatórios. Falaremos, nesse boletim, sobre a metodologia propensity score matching.

QUAL RESULTADO TERIA ACONTECIDO COM O PARTICIPANTE CASO ELE NÃO TIVESSE PARTICIPADO DO PROGRAMA?

Isso é o que se chama análise de contrafactual. O propósito é comparar o participante em dois cenários distintos: um em que ele recebe a intervenção e outro em que ele não recebe. Ao mesmo tempo, mesmo lugar e sob as mesmas circunstâncias. Logicamente, isso não é possível. Para contornar essa impossibilidade, utiliza-se um grupo de tratamento (que recebeu atendimento Sebrae, por exemplo) e um grupo de controle (que não recebeu atendimento). Uma avaliação de impacto nada mais é do que a comparação desses dois grupos. A questão é: como identificar um grupo de controle confiável?

Métodos experimentais são caracterizados pelo sorteio aleatório dos participantes para decidir quem fará parte do grupo tratado e quem fará parte do “grupo controle”. A aleatorização procura manter características como idade, sexo, perfil socioeconômico e outras variáveis não observáveis distribuídas de forma homogênea entre os grupos. Essa é a metodologia mais precisa para se gerar um “grupo controle” confiável.

NO QUE SE BASEIA UM EXPERIMENTO ALEATÓRIO?

- Grupos controle e tratado são definidos previamente, de forma aleatória;
- Ambos os grupos são acompanhados ao longo do tempo;
- A estimativa dos resultados é metodologicamente simples;
- Implementação mais complicada e com maior custo.

NO QUE SE BASEIA O PROPENSITY SCORE MATCHING?

- O método diminuiu o viés de seleção em estudos observacionais, pois cria um “grupo controle” comparável ao “grupo tratamento”.
- O “Grupo controle”, em geral, só é definido após o “grupo tratado” já ter recebido o tratamento;
- A estimativa de resultados é metodologicamente mais complexa;
- É fundamental ter acesso a diversos bancos de dados identificados;
- Às vezes, as principais variáveis de interesse não estão disponíveis nos bancos de dados existentes;
- Baixo custo de implementação.

DADOS FICTÍCIOS

Utilizando o software estatístico R, foi criado um banco de dados com 19.734 observações, sendo que cada observação indica um indivíduo. Para ficar mais didático, vamos supor que esses indivíduos são MEIs do setor de artesanato de uma localidade específica, em um determinado mês.

Esse banco é composto de seis variáveis: (i) identificador, simula o CNPJ; (ii) uma variável discreta, assimétrica à esquerda, que indica o número de produtos do tipo X vendidos pelo artesão no mês em questão; (iii) uma variável contínua, assimétrica à direita, que indica o peso do produto do tipo Y vendido pelo artesão no mês em questão; (iv) uma variável dicotômica, que indica se o artesão vende seus produtos na feira da cidade ou não; (v) uma variável discreta, que varia entre 0 e 10, indicando o número de outros produtos que são vendidos pelo artesão; e, finalmente, (vi) o rendimento bruto do artesão (combinação linear das variáveis (ii) a (v), acrescido de um erro aleatório, normalmente distribuído com média zero e variância 150) no mês em questão.

Todas as variáveis foram geradas de forma aleatória. Tomamos o cuidado para que elas não fossem correlacionadas entre si, para evitar problemas de multicolinearidade na modelagem estatística. Também criamos variáveis de diferentes tipos: simétricas, assimétricas, numéricas, categóricas, contínuas e discretas.

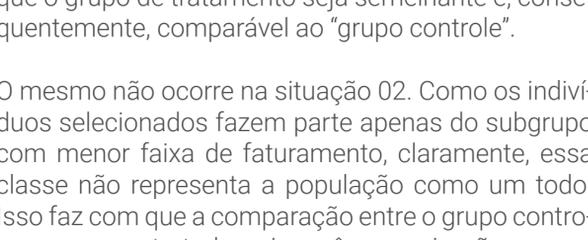
PROGRAMA QUE SERÁ AVALIADO

Suponhamos que o Sebrae crie um programa com a finalidade de aumentar o faturamento desses MEIs artesãos. E que, no mês em questão, participaram desse programa exatos 143 MEIs. Suponhamos ainda que o atendimento do Sebrae aumente o faturamento de cada um dos MEIs atendidos em 35%.

O próximo passo é definir as duas situações. Na primeira, a seleção desses 143 indivíduos será completamente aleatória. Na segunda, adicionaremos um viés de seleção, onde, por motivo desconhecido, todos os 143 indivíduos selecionados para participar do programa fazem parte do grupo dos 25% de artesãos com menores rendimentos mensais brutos.

Note que a primeira situação nada mais é do que um experimento aleatório e a segunda situação se constitui de um estudo observacional, onde não ocorreu qualquer cuidado na seleção dos indivíduos que participarão do programa. Esse descuido acarretou, ocasionalmente, na seleção de indivíduos pertencentes ao subgrupo com menores médias de faturamento - fato não conhecido pelo pesquisador.

Adicionamos esse viés com o objetivo de que, após a aplicação da metodologia propensity score matching, consigamos estimar o efeito do aumento do faturamento próximo aos 35% definidos anteriormente.



Os gráficos acima mostram a robustez e a facilidade de se obter o efeito do tratamento em um experimento aleatório (Situação 01). O faturamento mediano dos MEIs atendidos foi R\$ 4.053,51 contra um faturamento mediano dos não atendidos de R\$ 2.990,13. Fazendo a razão entre esses dois valores, encontramos 1,3556. Ou seja: o faturamento dos MEIs atendidos foi 35,56% superior aos não atendidos.

O mesmo não ocorre na Situação 02. Nessa situação, o faturamento mediano dos MEIs atendidos foi R\$ 3.050,41 contra um faturamento mediano dos não atendidos de R\$ 2.995,87. Fazendo a razão entre esses dois valores, encontramos 1,0182. Ou seja, se realizássemos apenas um monitoramento do indicador faturamento, sem o devido cuidado com a seleção aleatória dos participantes, inferiríamos que o aumento do faturamento nos MEIs atendidos foi de apenas 1,82%.

PORQUE AS DIFERENÇAS ENTRE OS RESULTADOS?

A diferença entre esses resultados mostra a fragilidade de uma análise feita sem que seja utilizada a metodologia adequada. É importante observar, inclusive, que se essa situação fosse real, o respectivo programa, que apresenta excelentes resultados (aumento do faturamento em 35%), poderia ser cancelado, pois a análise sem metodologia adequada indicou um crescimento de apenas 1,8%.

A diferença ocorreu, pois pelo princípio da aleatoriedade, os indivíduos tratados selecionados na situação 01, equivalem a uma amostra representativa de todo universo de MEIs estudado; englobando indivíduos de todos os subgrupos possíveis. Isso assegura que o grupo de tratamento seja semelhante e, consequentemente, comparável ao “grupo controle”.

O mesmo não ocorre na situação 02. Como os indivíduos selecionados fazem parte apenas do subgrupo com menor faixa de faturamento, claramente, essa classe não representa a população como um todo. Isso faz com que a comparação entre o grupo controle e o grupo tratado seja errônea, pois são grupos completamente diferentes entre si.

O erro cometido ao se estimar um aumento de 1,82% não decorre do fato de, na hora do atendimento, terem aparecido apenas MEIs presentes na faixa de faturamento mais baixa. O erro ocorreu quando comparamos esses MEIs atendidos com o universo total estudado, que são grupos com características distintas.

NA SITUAÇÃO 02, COMO INFERIR O IMPACTO DO PROGRAMA DE MANEIRA CERTA?

A metodologia Propensity Score Matching seleciona uma amostra do grupo que não recebeu tratamento; de forma que essa amostra seja a mais semelhante possível ao grupo que recebeu o tratamento. Ela está implementada no software estatístico R, na função matchit, dentro do pacote MatchIt.

O modelo selecionou um total de 8.512 MEIs não atendidos, que são semelhantes aos MEIs atendidos, e descartou os outros 10.936. Desse universo de 8.512, utilizando o método de vizinhos mais próximos, o modelo selecionou 143 indivíduos que mais se assemelham aos indivíduos tratados. Essa ação gerou dois grupos com 143 indivíduos cada: grupo tratado e grupo controle que foi selecionado pelo modelo estatístico.

Analisando o faturamento desses dois novos grupos, obtemos o faturamento mediano de R\$ 3.050,41 no grupo tratado, contra um faturamento mediano de R\$ 2.258,23 no grupo controle. Fazendo a razão entre esses dois valores, encontramos 1,3507. Ou seja, o faturamento dos MEIs atendidos foi 35,07% superior aos do grupo controle.

Importante notar que o faturamento do grupo tratado aqui é o mesmo faturamento do grupo tratado na situação 02. Isso porque, em ambas situações, estamos analisando os mesmos 143 indivíduos. O que mudou foi o faturamento do grupo controle. Na situação 02, o grupo controle era composto por todos indivíduos que não receberam atendimento. Aqui, o grupo controle é composto pelos 143 indivíduos mais semelhantes ao grupo tratado.

Por fim, concluímos que, além da definição da metodologia adequada, é fundamental atentar para a qualidade dos registros de atendimento e para o cadastro do público de interesse, objeto de um estudo de impacto. Só assim teremos capacidade de selecionar as variáveis corretas e definir um grupo controle com as mesmas características do grupo tratado.