

Técnicas Econométricas para Avaliação de Impacto

Avaliação experimental

Fabio Veras Soares
Centro Internacional de Pobreza (IPEA/PNUD)

Brasília, 16 de abril de 2008.

Introdução: como aleatorização resolve o problema do viés de seleção.

Como vimos na aula 1, o problema fundamental da avaliação de impacto é que a unidade de observação (indivíduo, família, firma, pais) não é observada simultaneamente em dois estados. Sinteticamente pode-se definir a variável de interesse, Y , como:

$$Y = dY_1 + (1 - d)Y_0,$$

onde $d=1$ indica que o indivíduo foi tratado e $d=0$ que não foi tratado. Neste caso, o impacto (ganho) em participar do programa é dado por:

$$\Delta = Y_1 - Y_0$$

A variável Y pode ser definida como uma função de variáveis observáveis e de um componente não observável:

$$Y_1 = g_1(X) + U_1$$

$$Y_0 = g_0(X) + U_0$$

Onde $E(U_1) = E(U_0) = 0$

Neste caso, o efeito médio do tratamento sobre os tratados (ATT) pode ser escrito como:

$$\begin{aligned} E(Y_1 - Y_0 | X, d = 1) &= E(\Delta | X, d = 1) \\ &= g_1(X) - g_0(X) + E(U_1 - U_0 | X, d = 1) \end{aligned}$$

e o efeito médio do tratamento (ATE):

$$\begin{aligned} E(Y_1 - Y_0 | X) &= E(\Delta | X) \\ &= g_1(X) - g_0(X) + E(U_1 - U_0 | X) \end{aligned}$$

O ganho (impacto) médio do programa poderia ser estimado como:

$$E(\Delta) = E(Y_1 | d = 1) - E(Y_0 | d = 0)$$

Adicionando e subtraindo, o valor esperado de um indivíduo tratado caso ele não tivesse recebido o tratamento,

$$E(Y_0 | d = 1) :$$

$$E(\Delta) = E(Y_1 - Y_0 | d = 1) - E(Y_0 | d = 0) + E(Y_0 | d = 1)$$

Onde o primeiro termo é o impacto do programa e o segundo termo é o viés de seleção.

Como $E(Y_0 | d = 1)$ não é observável, o objetivo essencial do trabalho empírico em avaliação de impacto é identificar situações em que se possa assumir que o viés de seleção não existe ou encontrar maneiras de corrigi-lo.

A aleatorização de indivíduos em grupos de tratamento e comparação é um dos casos onde o viés de seleção PODE ser inteiramente removido.

Motivo: Como o tratamento foi aleatoriamente distribuído entre os dois grupos, indivíduos tratados e não-tratados, em média, diferem apenas quanto a este status. Não tivesse o grupo de tratamento sido tratado ambos os grupos teriam em média o mesmo resultado:

$$E(Y_0 | d = 0) - E(Y_0 | d = 1) = 0$$

Hipóteses:

- Não há viés de aleatorização: a aleatorização não afeta o que seria o valor médio da variável de interesse na ausência do tratamento; e,
- O fato de uma unidade receber o tratamento não afeta o resultado potencial de uma unidade que não recebeu o tratamento (SUTVA):

$$\begin{aligned} E(\Delta) &= E(Y_1 | d = 1) - E(Y_0 | d = 0) \\ &= E(Y_1 - Y_0 | d = 1) \\ &= E(Y_1 - Y_0) \end{aligned}$$

Exclui a possibilidade efeitos de equilíbrio geral e externalidades.

- Isto implica que caso o experimento seja implementado de maneira adequada: ATT=ATE.

Aleatorização como variável instrumental

- IV tenta resolver o problema de endogeneidade de uma variável explicativa.
- A aleatorização (R) pode ser vista como uma variável instrumental no sentido em que ela determina a participação no programa, mas não é correlacionada com o resultado de interesse.
- Isto é particularmente importante quando os indivíduos decidem participar ou não em um programa baseados em suas expectativas de ganho que podem ser uma função tanto de variáveis observáveis (X) como de variáveis não observáveis (U).

▪ Nestas circunstâncias, ATT somente será igual a ATE se assumirmos que:

a) Não há heterogeneidade do impacto: modelo de efeito comum:

$$U_1 - U_0 = 0 \Rightarrow E(\Delta | X) = E(\Delta | d = 1, X) . O$$

impacto é o mesmo para pessoas com as mesmas características X . O componente não observável não afeta o ganho.

b) Há heterogeneidade, $U_1 \neq U_0$, mas a aleatorização é administrada somente para aqueles que participariam do programa ($d=1$). ATE=ATT.

c) Mas quando há heterogeneidade, em geral:

ATT \neq ATE

$$U_1 \neq U_0 \Rightarrow E(\Delta | X) \neq E(\Delta | d = 1, X)$$

- Heckman (1992) ressalta que as condições de identificação do impacto do programa quando se usa aleatorização não exigem que X seja independente ou ortogonal a U . Deste modo, os outros parâmetros estimados por um modelo experimental podem não ser identificados e, portanto, não fornecem informações sobre parâmetros estruturais que poderiam ser de interesse. O parâmetro do tratamento é identificado porque X é ortogonal a D por construção.

$$\begin{aligned} E(Y_1 - Y_0 | X) &= E(\Delta | X) \\ &= g_1(X) - g_0(X) + E(U_1 - U_0 | X) \end{aligned}$$

Desenhando uma Avaliação com Aleatorização

- Desenhando avaliações:
 - a) Parceiros: governos e ONGs
 - b) Pilotos: oportunidade de informar a expansão do programa
 - c) Objetivos: Testar o pacote, testar mecanismos e testar teoria
- Métodos de Aleatorização:
 - a) Super-inscrição;
 - b) Aleatorização da entrada no programa;
 - c) Aleatorização dentro do grupo;
 - d) Aleatorização com incentivo para participação;

- Tamanho da amostra, desenho e poder dos experimentos.

a) Princípios básicos:

- A amostra precisa ter um tamanho suficiente que permita identificar um impacto de tamanho x .
- O poder do teste é a probabilidade de rejeitar a hipótese nula quando ela é, de fato, falsa.
- O tamanho mínimo do efeito detectável (MDE) para um poder do teste κ , nível de significância α , tamanho da amostra N é dado por:

$$MDE = (t_{1-\kappa} + t_{\alpha}) * \sqrt{\frac{1}{P(1-P)}} * \sqrt{\frac{\sigma^2}{N}}$$

Esta equação implicitamente define o tamanho da amostra e explicita o dilema entre o poder e o tamanho do teste.

A função MDE é minimizada com $P=0.5$, mas se a coleta de dados for barata, a proporção ótima é ter um grupo de comparação maior. (Maior probabilidade de atrito)

b) Erros grupados:

Quando se aleatoriza grupos ao invés de indivíduos, é importante observar que o erro pode não ser independente entre indivíduos dentro do mesmo grupo:

$$MDE = \frac{(t_{1-\kappa} + t_{\alpha})}{P(1-P)J} * \sqrt{\rho \frac{1-\rho}{n}} * \sigma$$

Onde ρ é a correlação intra-cluster e J corresponde ao número de clusters.

A precisão das estimativas aumenta mais com um aumento do número de clusters (J) do que com um aumento no número de observações (n) nos clusters existentes.

c) Alocação imperfeita (*imperfect compliance*)

$$MDE = (t_{1-\kappa} + t_{\alpha}) * \sqrt{\frac{1}{P(1-P)}} * \sqrt{\frac{\sigma^2}{N}} * \frac{1}{c-s}$$

Onde c é a proporção de alocados no tratamento que realmente recebem o tratamento e s a proporção de alocados no grupo de comparação que recebem o tratamento.

d) Variáveis de Controle

- Controlar por variáveis que influenciam a variável de interesse não afeta o valor esperado do estimador β , mas pode reduzir sua variância, e, portanto, o tamanho da amostra.
- Estas variáveis, no entanto, devem ser capturadas no *baseline*, pois se elas forem influenciadas pelo tratamento, elas podem capturar parte do efeito do mesmo na variável de interesse.
- Controlando por variáveis que afetam pouco a variável de interesse pode aumentar a variância ao diminuir o número de graus de liberdade.

e) Estratificação

- A aleatorização é usada para garantir que para certas dimensões observáveis (X), os grupos de controle e tratamento tenham médias (esperanças) semelhantes.
- Estratificação é utilizada para garantir que isso se verifique na prática.
- A precisão será melhorada na medida em que os blocos forem formados por variáveis que afetam a variável de interesse.

- Além de reduzir a variância este procedimento é útil para analisar a heterogeneidade do impacto para diferentes grupos.

f) Cálculo do poder do teste na prática

- É necessário ter uma idéia da média e da variância da variável de interesse na ausência do experimento, depois de controlar por possíveis co-váriaveis e/ou estratificação.

- No caso de desenho grupado é preciso ter uma idéia da correlação da variável de interesse para membros do mesmo grupo.

- Decidir o tamanho do teste e o tamanho do efeito mínimo detectável.

- Decidir o poder do teste.
 - Existe um *freeware* chamado *Optimal design* para calcular o tamanho da amostra. O manual desse software também é um bom guia para o estudo de experimentos.

Implementando a aleatorização

- Nível de Aleatorização:

Em alguns casos é obvio se o nível de aleatorização deve ser o indivíduo ou algum grupo (escola, cidade, indústrias), mas em outros casos não é tão óbvio assim. Critérios para decisão.

- a) Quanto maior o número de grupos, maior o tamanho da amostra para o MDE.
- b) Efeito transbordamento (ou externalidade) pode enviesar a estimação dos efeitos do tratamento.

c) Impactos nos indivíduos não selecionados (ressentimento) e custos fixos.

- Desenhos fatoriais (*cross-cutting design*)

- Usados para testar varias intervenções e combinações destas em relação a um grupo de comparação e entre grupos de tratamento.

- Usados para testar interações de diferentes componentes de um programa.

- Pode ser usado para testar duas hipóteses ao invés de uma, sem um grande aumento de custos.

Levantamento de Dados

- Linha de base:

- a) reduz os requisitos do tamanho da amostra, por gerar variáveis de controle correlacionadas com a variável de interesse;
- b) interação entre valores iniciais e o impacto do programa;
- c) checar se a aleatorização foi bem feita;
- d) testar e refinar os procedimentos de coleta de dados.

- Uso de dados administrativos:

- a) pode reduzir atrito;

- b) metodologia de coleta quando se combina pesquisa de campo e dados administrativos;
- c) instrumento pode ser afetado pelo programa

Quando a aleatorização não é perfeita...

Ameaças à validade interna da avaliação:

- a) Probabilidade de seleção depende de variáveis observáveis: calcular médias ponderadas ou controlar por todos os X 's na regressão (incluindo interações);
- b) Alocação imperfeita:

Estimativa da intenção do tratamento:

$$ITT = E(Y_1 | Z = 1) - E(Y_0 | Z = 0)$$

Wald (LATE) Estimator:

$$\beta_w = \frac{E(Y_1 | Z = 1) - E(Y_0 | Z = 0)}{E[D | Z = 1] - E[D | Z = 0]}$$

Hipóteses:

- independência: Y e D são independentes de Z
- monotonicidade: o instrumento faz cada observação mais (ou menos) provável de participar no programa.
 - Independência não se sustenta caso exista externalidade: Z afeta Y .
 - Existência de externalidade positiva infla o estimador Wald.

Modelo de efeito comum: LATE=ATT=ATE

Modelo com heterogeneidade, mas instrumento não é correlacionado com o ganho do programa: $LATE=ATT$
Modelo com heterogeneidade e instrumento correlacionado com o ganho (auto-seleção): $LATE$.

c) Externalidades (SUTVA é violado)

- Físicas, mudanças de preços, efeito aprendizagem e imitação.
- Externalidades globais tornam a identificação do impacto do programa muito difícil.
- Externalidades locais (restritas a um grupo), permitem a identificação do impacto total do programa.
- experimentos podem ser desenhados para captar externalidades entre grupos.

d) Atrito na amostra

- identificar se é um problema: atrito aleatório apenas diminui o poder estatístico do experimento.

- aplicar técnicas de correção: paramétricas ou não paramétricas.

▪ Inferência

a) Dados grupados: Mínimos Quadrados Generalizados

b) Resultados Múltiplos: ajusta os p-valores; estimar modelos SUR e calcular o Efeito tratamento padronizado.

- c) Subgrupos: definir ex-ante e estratificar; ex-post tomar cuidado com tamanho da amostra e *data mining*.
- d) Covariáveis: apenas as que de fato afetam a variável de interesse.

Validade Externa e Generalização de Experimentos

- validade interna é necessária, mas não suficiente para a validade externa;
- incapacidade de captar efeitos de equilíbrio geral;

- *Hawthorne effect*: mudança de comportamento entre os tratados – *outperform*
- *John Henry effect*: mudança de comportamento no grupo de comparação – sabotagem ou competição. Ambos são efeitos de curto prazo. Dados de mais longo prazo podem testar o impacto destes vieses.
- Pilotos como “*gold plating*”
- Replicação para acumular conhecimento (viés de publicação dificulta esta prática).

- Modelos teóricos ajudam a fazer predição fora da amostra.

Debate:

Métodos experimentais versus quase-experimentais:

- a) quase-experimentos conseguem replicar os resultados de experimentos?
- Discussão sobre os NSW, JTPA e Progresas.
- b) Experimentos não estimam alguns estimadores de efeitos que podem ser relevantes e não dão tanta importância a modelos estruturais.
- c) Combinação de experimentos com métodos quase-experimentais

