ACADEMIA HUMANITÁRIA · HUMANITAIRES JOBS

Mini MBA Programas e MEAL: conceber, pilotar e provar o impacto

Material do curso · Semana 10 : Medir o impacto: contrafactual, métodos e contribuição

Programa de exatamente 3 meses (12 semanas, 144 horas) para conceber, pilotar e avaliar programas humanitários: avaliação das necessidades e severidade, cadeia de resultados e indicadores, focalização e transferências monetárias, planeamento e orçamento, plano MEAL e qualidade dos dados, amostragem, análise, prestação de contas, avaliação, medição do impacto, otimização dos recursos e gestão adaptativa. Exercícios numéricos corrigidos, três exames mensais com estudos de caso e um diploma verificável da Academia Humanitária.

https://humanitairejobs.org/pt/academia/mini-mba-programmes-meal/semaine-10/

Documento gratuito da Academia Humanitária. Pode ser impresso e partilhado livremente para fins de formação, sem modificações.

Semana 10 : Medir o impacto: contrafactual, métodos e contribuição

Objetivos

- Compreender o problema do contrafactual e porque é que uma comparação antes-depois não mede um efeito.
- Escolher e aplicar um método quase experimental ou experimental adaptado ao contexto humanitário, com as suas condições de validade.
- Utilizar as abordagens de contribuição quando a atribuição é impossível, e decidir o que é razoável provar.

O problema fundamental: o que teria acontecido sem o programa?

O **impacto** de um programa é a diferença entre o que aconteceu com o programa e o que teria acontecido sem ele: o **contrafactual**. Nunca se pode observar diretamente o mesmo agregado familiar com e sem programa; todos os métodos de avaliação de impacto são formas de construir uma estimativa credível desse contrafactual.

Duas comparações ingénuas são enganadoras:

- **Antes-depois** nos beneficiários: a mudança observada mistura o efeito do programa com tudo o que mudou ao mesmo tempo (colheita, preços, segurança, outras ajudas). Se a colheita foi boa, a segurança alimentar melhora mesmo sem programa; se foi má, um programa eficaz pode parecer sem efeito.
- **Beneficiários contra não beneficiários** após o programa: os dois grupos eram diferentes à partida, precisamente porque a focalização escolheu os mais vulneráveis. A comparação subestima então o efeito, ou sobrestima-o se foram os mais motivados que se inscreveram (enviesamento de seleção).

> Pergunta a fazer perante qualquer afirmação de impacto: «em comparação com quê?». Se a resposta for «em comparação com antes», não é uma medição de impacto.

O ensaio aleatorizado controlado e a sua utilização no humanitário

Num **ensaio aleatorizado controlado** (EAC), sorteiam-se, entre as pessoas elegíveis, as que recebem o programa (grupo de tratamento) e as que não o recebem ou o recebem mais tarde (grupo de controlo). O sorteio torna os dois grupos comparáveis em média em todas as características, observadas ou não. O efeito é simplesmente a diferença dos resultados médios no fim.

No humanitário, recusar uma ajuda vital a pessoas elegíveis para constituir um grupo de controlo é **contrário à ética**. Mas os EAC continuam possíveis e produziram conhecimentos importantes (sobre o cash, a nutrição, a saúde mental) em situações precisas:

- **Recursos insuficientes** de qualquer forma para toda a gente: o sorteio é então uma forma equitativa de repartir, e o grupo não servido serve de controlo.
- **Implementação progressiva**: toda a gente receberá o programa, mas a ordem é sorteada (desenho em escada, «stepped wedge»).
- **Comparação de duas variantes** de um programa, ambas consideradas aceitáveis: cash mensal ou de uma só vez, com ou sem acompanhamento.

Os EAC são caros, demoram tempo e respondem a perguntas precisas; justificam-se quando a resposta servirá numerosos programas. Para a maioria dos projetos, os métodos quase experimentais bastam.

A diferença nas diferenças

A **diferença nas diferenças** compara a evolução de um grupo ajudado com a evolução de um grupo de comparação não ajudado: efeito = (depois − antes) no grupo ajudado − (depois − antes) no grupo de comparação. Elimina as diferenças iniciais estáveis entre os grupos e as tendências comuns (colheita, preços).

Exemplo. Proporção de agregados com um consumo aceitável. Grupo ajudado: 28 % antes, 52 % depois (+24 pontos). Grupo de comparação: 35 % antes, 44 % depois (+9 pontos). Efeito estimado: 24 − 9 = 15 pontos. Uma comparação antes-depois teria atribuído 24 pontos ao programa; uma comparação apenas dos resultados finais ter-lhe-ia atribuído 8.

Condição de validade: o pressuposto das **tendências paralelas**. Sem o programa, os dois grupos teriam evoluído da mesma forma. Verifica-se se possível com várias medições antes do programa, e assegura-se que o grupo de comparação não recebe outra ajuda importante durante o período (contaminação) e não sofre um choque específico.

No humanitário, o grupo de comparação é muitas vezes uma zona vizinha não coberta por falta de financiamento, ou agregados elegíveis inscritos numa lista de espera. Nunca se cria um grupo de comparação recusando a ajuda para as necessidades do estudo.

Emparelhamento e regressão descontínua

O **emparelhamento** constrói um grupo de comparação associando a cada agregado ajudado um ou vários agregados não ajudados que se lhe assemelham nas características medidas antes do programa (dimensão, bens, rendimentos, estatuto, zona). O método do **score de propensão** resume essas características numa probabilidade de ser ajudado, e comparam-se agregados com scores próximos. O seu limite: só corrige as diferenças observadas. Se a motivação ou as redes sociais influenciaram a inscrição e não forem medidas, o enviesamento permanece.

A **regressão descontínua** aplica-se quando a elegibilidade depende de um limiar numa pontuação: por exemplo, agregados com uma pontuação de vulnerabilidade de 60 ou mais ajudados, abaixo não. Os agregados imediatamente acima e imediatamente abaixo do limiar

(57-59 e 60-62) são quase idênticos; a diferença de resultados entre eles estima o efeito do programa em torno do limiar. Condições: o limiar é aplicado estritamente, e os agregados não podem manipular a sua pontuação para passar para cima.

Exemplo. Agregados com pontuação 57-59: 38 % com um consumo aceitável. Agregados com pontuação 60-62 (ajudados): 51 %. Efeito estimado em torno do limiar: cerca de 13 pontos. Este resultado vale para os agregados próximos do limiar, não necessariamente para os mais vulneráveis.

A **série temporal interrompida** compara a tendência de um indicador antes e depois do início de um programa, em numerosos pontos de medição (por exemplo, as admissões semanais em nutrição). É útil quando não existe nenhum grupo de comparação, desde que haja pontos suficientes antes da intervenção.

Abordagens de contribuição: quando a atribuição é impossível

Muitos efeitos importantes não se prestam aos métodos estatísticos: uma advocacia que muda uma lei, um reforço de capacidades de organizações locais, uma coordenação melhorada. Não se podem comparar com um contrafactual quantificado. Procura-se então estabelecer de forma credível que o programa **contribuiu** para a mudança.

- **Análise de contribuição** (Mayne): expõe-se a teoria da mudança, reúnem-se as evidências para cada ligação, examinam-se as explicações alternativas, e constrói-se uma «narrativa de contribuição» que se submete à crítica das partes interessadas.
- **Rastreamento de processos** (process tracing): testam-se hipóteses causais procurando evidências cuja presença ou ausência permite confirmar ou afastar uma explicação (por exemplo, a reprodução palavra por palavra de uma nota de advocacia num decreto).
- **Colheita de resultados** (outcome harvesting): identificam-se primeiro as mudanças observadas nos atores, e depois recua-se para determinar se e como o programa contribuiu para elas.
- **Mudança mais significativa**: recolhem-se histórias de mudança junto das pessoas, e painéis selecionam as mais significativas explicando porquê. Útil para descobrir efeitos não previstos.

> Rigor não quer dizer quantitativo. Uma análise de contribuição bem conduzida, que examina seriamente as explicações alternativas, é mais credível do que uma comparação antes-depois quantificada.

Escolher o método e decidir o que é preciso provar

A escolha de um método depende de quatro perguntas:

- **Que decisão?** Alargar um modelo-piloto à escala nacional justifica uma evidência forte; ajustar um programa em curso contenta-se com uma evidência suficiente.
- **As evidências já existem?** O efeito do cash no consumo alimentar está muito documentado: é inútil voltar a demonstrá-lo com um EAC em cada projeto. Basta muitas

vezes verificar que o programa é fiel ao modelo e que os efeitos intermédios estão presentes.

- **O que é exequível e ético?** Existe um grupo de comparação sem recusar a ajuda? Um limiar de elegibilidade claro? Uma implementação progressiva?
- **Que orçamento e que prazo?** Uma avaliação de impacto séria exige uma base antes do arranque: concebe-se com o programa, não no fim.

Quanto à **dimensão da amostra**, uma avaliação de impacto deve ter uma potência suficiente para detetar o efeito esperado. Detetar um efeito de 15 pontos exige muito menos agregados do que um efeito de 5 pontos. Se o efeito esperado for fraco e a amostra pequena, o estudo concluirá quase certamente «nenhum efeito significativo», o que será mal interpretado como «nenhum efeito».

Último ponto: apresentar honestamente o **nível de evidência**. «Os agregados ajudados melhoraram o seu consumo em 24 pontos; em comparação com uma zona de comparação, o efeito atribuível é estimado em 15 pontos, sob o pressuposto de tendências paralelas que os dados de duas campanhas anteriores sustentam» é uma frase que inspira confiança.

Exercício: estimar um efeito e criticar o método

Situação. Um programa de meios de subsistência apoiou 1 200 agregados. Rendimento mensal mediano: agregados ajudados 42 USD antes, 71 USD depois. Zona vizinha não coberta (por falta de financiamento): 47 USD antes, 58 USD depois. Durante o programa, outra ONG distribuiu sementes em metade das aldeias da zona vizinha. O relatório final anuncia: «o programa aumentou os rendimentos em 69 %».

Trabalho pedido. 1) De onde vem o número de 69 % e porque é enganador? 2) Calcule o efeito pela diferença nas diferenças. 3) Qual é o efeito provável da distribuição de sementes na sua estimativa, e o que propõe?

Correção. 69 %: (71 − 42) / 42, comparação antes-depois que atribui ao programa todo o aumento, incluindo o devido ao contexto. Diferença nas diferenças: (71 − 42) − (58 − 47) = 29 − 11 = 18 USD por mês, ou seja, cerca de 43 % do rendimento inicial dos agregados ajudados. As sementes provavelmente aumentaram os rendimentos da zona de comparação, o que inflaciona a sua tendência e subestima o efeito do programa (a contaminação viola o pressuposto das tendências paralelas). Proposta: analisar separadamente as aldeias de comparação sem sementes, ou controlar esta variável; apresentar a estimativa como um limite inferior e indicar claramente a limitação.

A reter

- O impacto mede-se em relação a um contrafactual; antes-depois não é uma medição de impacto.
- Diferença nas diferenças = evolução do grupo ajudado − evolução do grupo de comparação, sob o pressuposto de tendências paralelas.
- Um EAC no humanitário só é ético se não privar ninguém de uma ajuda disponível: recursos limitados, implementação progressiva, ou comparação de variantes.
- Regressão descontínua: comparar os agregados imediatamente acima e imediatamente abaixo de um limiar estrito de elegibilidade.
- Quando a atribuição é impossível: análise de contribuição, rastreamento de processos, colheita de resultados, mudança mais significativa.

Quiz

1. Grupo ajudado: 30 % antes, 50 % depois. Comparação: 32 % antes, 40 % depois. Efeito pela diferença nas diferenças:
   A. 20 pontos
   B. 12 pontos
   C. 10 pontos
   D. 8 pontos

2. Que pressuposto é essencial para a diferença nas diferenças?
   A. Os grupos são idênticos à partida
   B. Sem o programa, os dois grupos teriam evoluído da mesma forma
   C. A amostra ultrapassa 1 000 agregados
   D. O programa durou mais de um ano

3. Em que situação um sorteio é eticamente aceitável no humanitário?
   A. Para recusar uma ajuda vital disponível a metade dos elegíveis
   B. Quando os recursos de qualquer forma não chegam para todos os elegíveis, ou para a ordem de uma implementação progressiva
   C. Nunca
   D. Apenas se o doador o impuser

4. O principal limite do emparelhamento por score de propensão é:
   A. Só funciona com EAC
   B. Só corrige as diferenças medidas, não as diferenças não observadas como a motivação
   C. Exige um limiar de elegibilidade
   D. É proibido pela OCDE

5. Para demonstrar que uma ação de advocacia contribuiu para a adoção de uma lei, o método mais adequado é:
   A. Um ensaio aleatorizado
   B. Uma diferença nas diferenças
   C. Uma análise de contribuição ou um rastreamento de processos
   D. Um inquérito aos agregados familiares

Respostas do quiz : 1. B · 2. B · 3. B · 4. B · 5. C