

## Avaliação de impacto na prática

### Atividade Final

**Instruções:** Este conjunto de problemas revisa a aplicação empírica de um conjunto básico de métodos de avaliação de impacto. Embora os dados sejam fictícios, eles são modelados a partir de exemplos do mundo real de avaliação de impacto.

**Contexto:** Este exercício analisa o impacto de um programa de subsídio de seguro saúde para os pobres, o HISP. Um dos principais objetivos do programa é reduzir a carga de despesas relacionadas à saúde para famílias de baixa renda. Você foi encarregado de realizar uma avaliação de impacto do efeito do HISP sobre os gastos diretos com saúde. Você usará o conjunto de dados denominado “evaluation.dta” para conduzir sua análise. O desenho da avaliação mudará a cada caso.

Antes de começar:

1. Crie um arquivo ipynb” para registrar sua implementação e textos, chame-o de “Solução\_setorPublicoII\_Seunome.ipynb”. Este arquivo deve conter suas análises e os resultados dos comandos de acordo com cada caso e respectivas questões.

### **Caso 1 de avaliação de impacto: antes e depois**

Nesse cenário, você estimará o efeito do programa comparando a mudança nos resultados ao longo do tempo para um grupo de famílias que se inscreveram no programa. Presuma conformidade total, o que significa que todas as famílias elegíveis para o programa se inscreveram nele.

1. Compare os gastos médios com saúde antes (round = 0) e depois do programa (round = 1) para o grupo de famílias elegíveis (eligible = 1) em comunidades de tratamento (treatment\_locality = 1). Apresente os testes de significância dessas médias.

## **Caso 2 de avaliação de impacto: Tratamentos auto selecionados**

Chamaremos isso de cenário de oferta aleatória. Nesse cenário, a inscrição no programa de seguro saúde é voluntária e todas as famílias nas comunidades piloto são elegíveis para se inscrever. Você tem à sua disposição uma pesquisa de linha de base (round = 0) que foi coletada antes do programa de seguro saúde ser implementado, e uma pesquisa de acompanhamento (round = 1) coletada 12 meses após o início do programa. Para este caso, analise apenas as comunidades onde o programa foi oferecido (manter se `treatment_locality = 1`). Ao executar regressões, certifique-se de agrupar os erros padronizados por localidade.

1. Descreva o número de observações, a média, o desvio padrão e os intervalos de confiança das variáveis no banco de dados de linha de base.
2. Para a linha de base, compare a média mensal per capita de gastos com saúde (`health_expenditures`) e outras covariáveis entre as famílias que se inscreveram no programa (`enrolled_ro = 1`) e aqueles que não se inscreveram. Os dois grupos são comparáveis?
3. Use a análise de regressão com Ordinary Least Squares (OLS) para estimar o efeito do programa sobre os beneficiários. Faça uma estimativa da correlação entre a participação no programa e os gastos diretos com saúde no período de tratamento (round = 1). Execute duas regressões:
  - a. Sem controles
  - b. Incluindo características do chefe da família e cônjuge
4. Qual é o “efeito” do programa sobre as despesas do bolso dos participantes?
5. Você acha que o coeficiente estimado da regressão **a** ou **b** é o verdadeiro impacto do programa? Com que tipo de viés você pode estar preocupado?

### **Caso 3 de avaliação de impacto: randomização**

Acontece que quando as comunidades estavam sendo selecionadas para inclusão no piloto de seguro saúde, havia muito mais comunidades elegíveis do que poderiam ser cobertas com o orçamento disponível. As autoridades provinciais decidiram fazer um sorteio para selecionar as comunidades que participariam do esquema de seguro no ano 1, dando assim a todas as comunidades uma chance justa de começar no programa primeiro. Seus dados contêm informações sobre as comunidades selecionadas aleatoriamente para participação no ano 1, bem como sobre as comunidades que só entrariam no programa nos anos subsequentes. A variável “treatment\_locality” indica comunidades de tratamento (treatment\_locality = 1) e comunidades de não tratamento ou controle (treatment\_locality = 0). Para este caso, use comunidades de tratamento e controle em sua análise. A amostra está estruturada da seguinte forma:

	Comunidades de tratamento	Comunidades de controle
Elegíveis	5.929	5.328
Inelegíveis	3.990	4.580

1. Compare os gastos de saúde da linha de base e outras covariáveis entre famílias elegíveis em comunidades de tratamento e controle. A amostra está balanceada em observáveis? É isso que você esperaria e por que ou por que não?
2. No período de tratamento (round = 1), compare a média dos gastos com saúde para a população elegível em tratamento e comunidades de controle. É este o impacto do programa de seguro saúde sobre os gastos diretos com saúde?
3. Agora use uma regressão OLS para estimar o efeito do programa sobre os gastos diretos com saúde:
  - a. Sem controles
  - b. Incluindo características do chefe da família e cônjuge
  - c. Incluindo covariáveis de linha de base
4. Qual é o impacto do programa sobre os gastos diretos com saúde? Qual é a redução percentual que pode ser atribuída ao programa?
5. Quais são alguns dos vieses potenciais com os quais você ainda pode estar preocupado com este método?

#### **Caso 4 de avaliação de impacto: Variáveis instrumentais**

No caso 3, você estimou o impacto do programa comparando as populações elegíveis nas populações de tratamento e controle. No entanto, você sabe que nem TODAS as famílias elegíveis em comunidades de tratamento decidiram se inscrever no programa. Você gostaria de saber qual foi o efeito do programa para as famílias elegíveis que realmente se inscreveram no programa e foram cobertas pelo seguro saúde. Para fazer isso, você precisará corrigir a seleção endógena para o tratamento usando um instrumento válido.

Para este caso, iremos alterar o desenho experimental. Suponha que o ministério da saúde tome uma decisão executiva de que o HISP seja disponibilizado a qualquer família que deseje se inscrever. No entanto, você sabe que, de forma realista, essa expansão nacional será incremental ao longo do tempo e, portanto, você chega a um acordo para acelerar a inscrição em um subconjunto aleatório de aldeias por meio de uma campanha de promoção. Você empreende um esforço intensivo de promoção em uma sub amostra aleatória de aldeias, incluindo campanhas de comunicação e marketing social destinadas a aumentar a conscientização sobre o HISP. A variável “promotion\_locality” indica comunidades de promoção (promotion\_locality = 1) e localidades de não promoção (promotion\_locality = 0). A variável “enrolled\_rp” indica se uma família se inscreveu ou não no programa neste cenário de promoção aleatória. Uma vez que o tratamento de promoção foi randomizado, você levanta a hipótese de que é um instrumento válido para corrigir a seleção para tratamento ao comparar aqueles que se inscreveram no programa com as comunidades de controle.

1. Estime uma regressão de “mínimos quadrados de dois estágios” (iv) do efeito do programa nos gastos diretos com saúde para famílias tratadas, usando promoção como sua variável instrumental.

- a. Sem controles
- b. Incluindo características do chefe da família e cônjuge
- c. Incluindo covariáveis de linha de base

2. Como você interpretaria esse novo coeficiente estimado?

### **Caso 5 de avaliação de impacto: Projeto de Descontinuidade de Regressão (RDD)**

Depois de fazer mais pesquisas sobre o desenho do programa de seguro saúde, você descobre que as autoridades decidiram direcionar o programa para famílias de baixa renda, portanto, apenas famílias abaixo da linha de pobreza existente seriam elegíveis para se inscrever. O índice de pobreza foi construído por meio de um teste de médias proxy que atribuiu a cada família na comunidade uma pontuação entre 0 e 100. Todas as famílias com uma pontuação menor ou igual a 58 foram classificadas como pobres e, portanto, elegíveis para se inscrever no programa de seguro saúde. Todas as famílias com uma pontuação acima de 58 foram consideradas não pobres e permaneceram inelegíveis para se inscrever. A variável “eligible” identifica domicílios abaixo da linha da pobreza (eligible = 1) e domicílios inelegíveis (eligible = 0).

1. Em primeiro lugar, normalize a pontuação da pobreza para que o limiar seja 0. Estime uma regressão OLS do índice de pobreza nas despesas com saúde do bolso no período de tratamento. Inclua os lados esquerdo e direito do limiar de pobreza na regressão.
2. Faça um gráfico do valor ajustado dos gastos diretos com saúde na pontuação do índice de pobreza
3. Estime o efeito do programa sobre os gastos diretos com saúde usando um modelo de RDD. Além de sua variável de tratamento / controle, certifique-se de controlar a elegibilidade do programa, incluindo os lados esquerdo e direito do limiar de pobreza em sua regressão.
4. Apresente em seu modelo de regressão os termos de pontuação de pobreza quadrática, cúbica e quártica.
5. Plote gráfico do valor ajustado dos gastos diretos com saúde estimados na etapa 4 em comparação com a pontuação da pobreza. Existe alguma evidência de não linearidade?
6. Realize um “experimento falso” usando o modelo RD no período da linha de base.
7. Plote o gráfico do valor ajustado dos gastos diretos com saúde estimados na etapa 6 em comparação com a pontuação da pobreza.

### **Caso 6 de avaliação de impacto: Diferenças em diferenças (dif em dif)**

Você não está muito satisfeito com os resultados de sua análise. Pensando mais a fundo, você percebe que, como tem dados para dois períodos para cada família na amostra, pode usar esses dados para resolver alguns dos desafios encontrados nos casos anteriores. Agora você vai comparar a mudança nos gastos diretos com saúde entre as famílias que se inscreveram no programa e as que não se inscreveram.

1. Gere uma nova variável com a diferença nos gastos diretos com saúde entre a linha de base (round = 0) e o acompanhamento (round = 1). Chame a nova variável de “delta\_he”.
2. Compare a mudança média nos gastos diretos com saúde entre as famílias que se inscreveram no programa e aquelas que não o fizeram nas localidades de tratamento. Suponha que estamos no cenário aleatório, portanto, certifique-se de usar a variável `enrolled_ro`.
3. Execute uma regressão OLS para estimar o efeito do programa sobre os gastos diretos com saúde usando a diferença nas diferenças:
  - a. Sem controles
  - b. Incluindo características do chefe da família e cônjuge
  - c. Incluindo covariáveis de linha de base
4. Qual é o “efeito” do programa sobre os gastos diretos com saúde?
5. Como este método melhora as estimativas obtidas no Caso 1? Existe algum viés potencial com o qual você ainda está preocupado?

### **Caso 7 de avaliação de impacto: Matching**

Desejando melhorar suas estimativas anteriores, você decide usar algumas técnicas de correspondência para selecionar um grupo de famílias participantes e não participantes que parecem semelhantes com base nas características observáveis.

1. Estime a probabilidade de participar do programa dadas as covariáveis observáveis das famílias participantes e não participantes (use uma regressão logit). Mais uma vez, suponha que estamos no cenário de oferta aleatória.
2. Use uma regressão linear para estimar o efeito do HISP nos gastos com saúde das famílias que decidiram se matricular.
  - a. Sem controles
  - b. Incluindo covariáveis de linha de base
5. Qual é o “efeito” do programa sobre os gastos diretos com saúde?
6. Como este método melhora as estimativas obtidas no Caso 2? Existe algum viés potencial com o qual você ainda está preocupado?