

Machine Learning e políticas públicas

Prof. Rafael B. Barbosa
Universidade Federal do Ceará
Fortaleza/2020

Sumário

- 1 O que é Machine Learning
 - Problemas de previsão em políticas públicas
 - Exemplo: Quem deve aguardar o julgamento em liberdade?
 - Problemas típicos de Big Data em economia
- 2 Introdução à avaliação de políticas públicas
- 3 Experimentos Randomizados Controlados (RCT)
 - Identificação por atribuição aleatória de tratamento
- 4 Seleção sobre os observáveis
 - Viés por omissão de variáveis (VOV)

O que é Machine Learning (ML)

Machine Learning (ML) são um conjunto de técnicas estatísticas/-computacionais desenvolvidas para realizar previsões e que permitem que seus **algoritmos** “aprendam” a partir dos dados.

Métodos de ML permitem que os algoritmos “aprendam” algo que não era inicialmente conhecido quando foram construídos.

Isso implica uma mudança radical na estrutura lógica dos algoritmos computacionais.

Abordagem Clássica de Programação

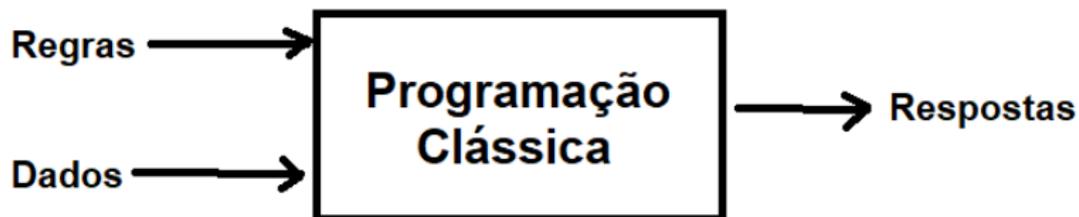
A abordagem clássica de programação utiliza construções lógicas para passar instruções do que um computador precisa fazer.

Exemplo:

Sejam os números $S = \{1, 2, 3, 4, 5\}$.

- > Se for digitado os números $\{1, 3, 5\} \implies$ escreva: “ímpar”
- > Se for digitado os número $\{2, 4\} \implies$ escreva: “par”
- > Se for digitado qualquer outro número \implies escreva: “Não consigo identificar”

Abordagem Clássica de Programação



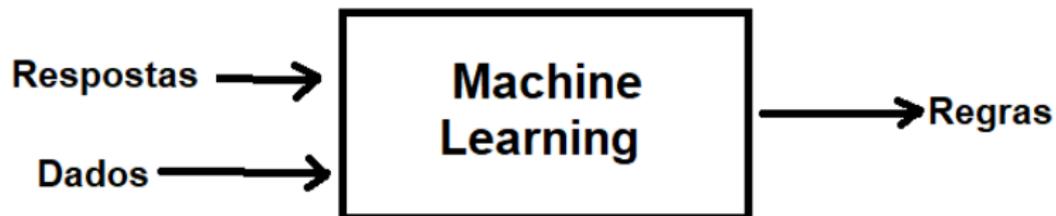
Abordagem Machine Learning

Na abordagem via machine learning não se sabe quais as regras que governam os dados. O algoritmo irá aprender sozinho analisando vários exemplos.

Exemplo:

- > Dados de entrada: {1 é ímpar}, {2 é par}, {3 é ímpar}, {4 é par}, {5 é ímpar}, {6 é par}
- > Personalização do algoritmo capaz de “aprender” o que é um número par ou ímpar.

Abordagem Machine Learning



Abordagem Machine Learning

A principal aplicação dos Métodos de ML é um campo da computação chamado de Inteligência Artificial (IA).

Ou seja, métodos de ML permitem que computadores “aprendam” regras sobre os dados que lhes são inputados e com isso possibilitam a tomada de decisão baseada nesse aprendizado.

Mas, que tipo de inteligência máquinas podem aprender?

ML permite que máquinas aprendam a tomar decisões baseadas em previsões sobre determinada situações.

Previsões são importantes no nosso dia a dia e são utilizadas por nós em diversas situações.

Exemplo simples de aplicação de ML e IA:

O google criou um algoritmo de ML para prever se um email era spam ou não era spam.

Problema: emails recebiam muitos spams e isso lotava a caixa de email dificultando o uso.

Saber antecipadamente que email é ou não é spam, permite tomar a decisão de colocá-lo em um lugar específico quando ele for enviado.

Exemplo simples de aplicação de ML e IA:

- 1 Inputar dados sobre emails que foram identificados a priori como sendo spams (*ground thuth*).
- 2 Criar vários modelos estatísticos para prever (classificar) determinados emails como sendo ou não spams
- 3 Escolher aquele modelo que tem o melhor desempenho
- 4 Incorporá-lo ao gmail para que possa realizar a ação de destinar emails classificados como spam para a caixa de spam.

Exemplo simples de aplicação de ML e IA:

- 1 Inputar dados sobre emails que foram identificados a priori como sendo spams (*ground thuth*).
- 2 Criar vários modelos estatísticos para prever (classificar) determinados emails como sendo ou não spams
- 3 Escolher aquele modelo que tem o melhor desempenho
- 4 Incorporá-lo ao gmail para que possa realizar a ação de destinar emails classificados como spam para a caixa de spam.

Exemplo simples de aplicação de ML e IA:

- 1 Inputar dados sobre emails que foram identificados a priori como sendo spams (*ground thuth*).
- 2 Criar vários modelos estatísticos para prever (classificar) determinados emails como sendo ou não spams
- 3 Escolher aquele modelo que tem o melhor desempenho
- 4 Incorporá-lo ao gmail para que possa realizar a ação de destinar emails classificados como spam para a caixa de spam.

Exemplo simples de aplicação de ML e IA:

- 1 Inputar dados sobre emails que foram identificados a priori como sendo spams (*ground thuth*).
- 2 Criar vários modelos estatísticos para prever (classificar) determinados emails como sendo ou não spams
- 3 Escolher aquele modelo que tem o melhor desempenho
- 4 Incorporá-lo ao gmail para que possa realizar a ação de destinar emails classificados como spam para a caixa de spam.

Exemplo simples de aplicação de ML e IA:

- 1 Inputar dados sobre emails que foram identificados a priori como sendo spams (*ground thuth*).
- 2 Criar vários modelos estatísticos para prever (classificar) determinados emails como sendo ou não spams
- 3 Escolher aquele modelo que tem o melhor desempenho
- 4 Incorporá-lo ao gmail para que possa realizar a ação de destinar emails classificados como spam para a caixa de spam.

Exemplo simples de aplicação de ML e IA:

Elementos importantes do procedimento

- 1 Sem dados não tem modelos de ML.
- 2 Algoritmos de ML são projetados para executar uma única tarefa.
- 3 Modelos de ML cometem erros!
- 4 Útil para a tomada de decisão que requer previsão (e que seja possível aplicar as técnicas estatísticas)

Exemplo simples de aplicação de ML e IA:

Elementos importantes do procedimento

- 1 Sem dados não tem modelos de ML.
- 2 Algoritmos de ML são projetados para executar uma única tarefa.
- 3 Modelos de ML cometem erros!
- 4 Útil para a tomada de decisão que requer previsão (e que seja possível aplicar as técnicas estatísticas)

Exemplo simples de aplicação de ML e IA:

Elementos importantes do procedimento

- 1 Sem dados não tem modelos de ML.
- 2 Algoritmos de ML são projetados para executar uma única tarefa.
- 3 Modelos de ML cometem erros!
- 4 Útil para a tomada de decisão que requer previsão (e que seja possível aplicar as técnicas estatísticas)

Exemplo simples de aplicação de ML e IA:

Elementos importantes do procedimento

- 1 Sem dados não tem modelos de ML.
- 2 Algoritmos de ML são projetados para executar uma única tarefa.
- 3 Modelos de ML cometem erros!
- 4 Útil para a tomada de decisão que requer previsão (e que seja possível aplicar as técnicas estatísticas)

Exemplo simples de aplicação de ML e IA:

Elementos importantes do procedimento

- 1 Sem dados não tem modelos de ML.
- 2 Algoritmos de ML são projetados para executar uma única tarefa.
- 3 Modelos de ML cometem erros!
- 4 Útil para a tomada de decisão que requer previsão (e que seja possível aplicar as técnicas estatísticas)

Exemplo simples de aplicação de ML e IA:

Elementos importantes do procedimento

- 1 Sem dados não tem modelos de ML.
- 2 Algoritmos de ML são projetados para executar uma única tarefa.
- 3 Modelos de ML cometem erros!
- 4 Útil para a tomada de decisão que requer previsão (e que seja possível aplicar as técnicas estatísticas ⇒ **BIG DATA**)

O que é Big Data?

- Grande número de observações
- Grande número de variáveis
- Dados com características não convencionais (dados não estruturados)

O que é Big Data?

- Grande número de observações
- Grande número de variáveis
- Dados com características não convencionais (dados não estruturados)

O que é Big Data?

- Grande número de observações
- Grande número de variáveis
- Dados com características não convencionais (dados não estruturados)

O que é Big Data?

- Grande número de observações
- Grande número de variáveis
- Dados com características não convencionais (dados não estruturados)

Classificação da IBM

Para ser chamado de Big Data a base de dados precisa ter ao menos uma das três características

1 Volume

- Grande número de observações e variáveis

2 Variedade

- Dados com diferentes estruturas.

3 Velocidade

- Dados com elevada frequência

Classificação da IBM

Para ser chamado de Big Data a base de dados precisa ter ao menos uma das três características

1 Volume

- Grande número de observações e variáveis

2 Variedade

- Dados com diferentes estruturas.

3 Velocidade

- Dados com elevada frequência

Classificação da IBM

Para ser chamado de Big Data a base de dados precisa ter ao menos uma das três características

1 Volume

- Grande número de observações e variáveis

2 Variedade

- Dados com diferentes estruturas.

3 Velocidade

- Dados com elevada frequência

Classificação da IBM

Para ser chamado de Big Data a base de dados precisa ter ao menos uma das três características

1 Volume

- Grande número de observações e variáveis

2 Variedade

- Dados com diferentes estruturas.
 - Dados estruturados: dados transversais, em painel ou séries temporais
 - Dados não estruturados: dados cujo significado é desconhecido à priori

3 Velocidade

- Dados com elevada frequência

Classificação da IBM

Para ser chamado de Big Data a base de dados precisa ter ao menos uma das três características

1 Volume

- Grande número de observações e variáveis

2 Variedade

- Dados com diferentes estruturas.
 - Dados estruturados: dados transversais, em painel ou séries temporais
 - Dados não estruturados: dados cujo significado é desconhecido à priori

3 Velocidade

- Dados com elevada frequência

Classificação da IBM

Para ser chamado de Big Data a base de dados precisa ter ao menos uma das três características

1 Volume

- Grande número de observações e variáveis

2 Variedade

- Dados com diferentes estruturas.
 - Dados estruturados: dados transversais, em painel ou séries temporais
 - Dados não estruturados: dados cujo significado é desconhecido à priori

3 Velocidade

- Dados com elevada frequência

Classificação da IBM

Para ser chamado de Big Data a base de dados precisa ter ao menos uma das três características

1 Volume

- Grande número de observações e variáveis

2 Variedade

- Dados com diferentes estruturas.
 - Dados estruturados: dados transversais, em painel ou séries temporais
 - Dados não estruturados: dados cujo significado é desconhecido à priori

3 Velocidade

- Dados com elevada frequência

Classificação da IBM

Para ser chamado de Big Data a base de dados precisa ter ao menos uma das três características

1 Volume

- Grande número de observações e variáveis

2 Variedade

- Dados com diferentes estruturas.
 - Dados estruturados: dados transversais, em painel ou séries temporais
 - Dados não estruturados: dados cujo significado é desconhecido à priori

3 Velocidade

- Dados com elevada frequência

Classificação da IBM

Para ser chamado de Big Data a base de dados precisa ter ao menos uma das três características

1 Volume

- Grande número de observações e variáveis

2 Variedade

- Dados com diferentes estruturas.
 - Dados estruturados: dados transversais, em painel ou séries temporais
 - Dados não estruturados: dados cujo significado é desconhecido à priori

3 Velocidade

- Dados com elevada frequência

Classificação da IBM

Exemplo:

Notícias de Jornal

1 Volume

- Cada texto possui inúmeras informações

2 Variedade

- Textos não são organizados na estrutura tradicional (Objeto -> Observação -> Variação)

3 Velocidade

- Todos os dias milhares de textos de jornal são publicados

Classificação da IBM

Exemplo:

Notícias de Jornal

1 Volume

- Cada texto possui inúmeras informações

2 Variedade

- Textos não são organizados na estrutura tradicional (Objeto -> Observação -> Variação)

3 Velocidade

- Todos os dias milhares de textos de jornal são publicados

Classificação da IBM

Exemplo:

Notícias de Jornal

1 Volume

- Cada texto possui inumeras informações

2 Variedade

- Textos não são organizados na estrutura tradicional (Objeto -> Observação -> Variação)

3 Velocidade

- Todos os dias milhares de textos de jornal são publicados

Classificação da IBM

Exemplo:

Notícias de Jornal

1 Volume

- Cada texto possui inumeras informações

2 Variedade

- Textos não são organizados na estrutura tradicional (Objeto -> Observação -> Variação)

3 Velocidade

- Todos os dias milhares de textos de jornal são publicados

Classificação da IBM

Exemplo:

Notícias de Jornal

1 Volume

- Cada texto possui inúmeras informações

2 Variedade

- Textos não são organizados na estrutura tradicional (Objeto -> Observação -> Variação)

3 Velocidade

- Todos os dias milhares de textos de jornal são publicados

Classificação da IBM

Exemplo:

Notícias de Jornal

1 Volume

- Cada texto possui inumeras informações

2 Variedade

- Textos não são organizados na estrutura tradicional (Objeto -> Observação -> Variação)

3 Velocidade

- Todos os dias milhares de textos de jornal são publicados

Classificação da IBM

Exemplo:

Notícias de Jornal

1 Volume

- Cada texto possui inumeras informações

2 Variedade

- Textos não são organizados na estrutura tradicional (Objeto -> Observação -> Variação)

3 Velocidade

- Todos os dias milhares de textos de jornal são publicados

Classificação da IBM

Exemplo:

Notícias de Jornal

1 Volume

- Cada texto possui inumeras informações

2 Variedade

- Textos não são organizados na estrutura tradicional (Objeto -> Observação -> Variação)

3 Velocidade

- Todos os dias milhares de textos de jornal são publicados

Classificação da IBM

Exemplo:

Notícias de Jornal

1 Volume

- Cada texto possui inúmeras informações

2 Variedade

- Textos não são organizados na estrutura tradicional (Objeto -> Observação -> Variação)

3 Velocidade

- Todos os dias milhares de textos de jornal são publicados

Aplicações em economia

Personalização de produtos

- Marketing Digital - Facebook
- Identificação e predição de características de clientes - Nubank
- Netflix

Personalização de produtos

- Marketing Digital - Facebook
- Identificação e predição de características de clientes - Nubank
- Netflix

Personalização de produtos

- Marketing Digital - Facebook
- Identificação e predição de características de clientes - Nubank
- Netflix

Personalização de produtos

- Marketing Digital - Facebook
- Identificação e predição de características de clientes - Nubank
- Netflix

Utilização de dados administrativos restritos

- Previsão e tomada de decisão governamental - Identificação de fraudes à Receita Federal
- Análise de documentos históricos

Utilização de dados administrativos restritos

- Previsão e tomada de decisão governamental - Identificação de fraudes à Receita Federal
- Análise de documentos históricos

Utilização de dados administrativos restritos

- Previsão e tomada de decisão governamental - Identificação de fraudes à Receita Federal
- Análise de documentos históricos

Prevendo eventos sociais

- Previsão e tomada de decisão
 - Resultados educacionais (Evasão, Alunos com elevado desempenho, Reprovação, etc)
 - Prevendo ocorrência de gripes
 - Prevendo decisões judiciais

Prevendo eventos sociais

- Previsão e tomada de decisão

- 1 Resultados educacionais (Evasão, Alunos com elevado desempenho, Reprovação, etc)
- 2 Prevendo ocorrência de gripes
- 3 Prevendo decisões judiciais

Prevendo eventos sociais

- Previsão e tomada de decisão
 - 1 Resultados educacionais (Evasão, Alunos com elevado desempenho, Reprovação, etc)
 - 2 Prevendo ocorrência de gripes
 - 3 Prevendo decisões judiciais

Prevendo eventos sociais

- Previsão e tomada de decisão
 - 1 Resultados educacionais (Evasão, Alunos com elevado desempenho, Reprovação, etc)
 - 2 Prevendo ocorrência de gripes
 - 3 Prevendo decisões judiciais

Prevendo eventos sociais

- Previsão e tomada de decisão
 - 1 Resultados educacionais (Evasão, Alunos com elevado desempenho, Reprovação, etc)
 - 2 Prevendo ocorrência de gripes
 - 3 Prevendo decisões judiciais

Novas formas de mensurar atividade econômica

- Construção de índices
 - Índice de polaridade política (Getzkow et al, 2014)
 - Índice de incerteza político-econômica (EPU) (Bloom e Baker, 2016)
- Medindo variáveis econômicas
 - Medindo a pobreza com dados de celular (Blumenstock et al, 2015) ou com luminosidade de imagens de satélite (Jean et al, 2016)
 - Medindo a riqueza por ruas usando Google Street View (Glaeser et al, 2016)

Novas formas de mensurar atividade econômica

- Construção de índices

- Índice de polaridade política (Getzkow et al, 2014)
- Índice de incerteza político-econômica (EPU) (Bloom e Baker, 2016)

- Medindo variáveis econômicas

- Medindo a pobreza com dados de celular (Blumenstock et al, 2015) ou com luminosidade de imagens de satélite (Jean et al, 2016)
- Medindo a riqueza por ruas usando Google Street View (Glaeser et al, 2016)

Novas formas de mensurar atividade econômica

- Construção de índices
 - Índice de polaridade política (Getzkow et al, 2014)
 - Índice de incerteza político-econômica (EPU) (Bloom e Baker, 2016)
- Medindo variáveis econômicas
 - Medindo a pobreza com dados de celular (Blumenstock et al, 2015) ou com luminosidade de imagens de satélite (Jean et al, 2016)
 - Medindo a riqueza por ruas usando Google Street View (Glaeser et al, 2016)

Novas formas de mensurar atividade econômica

- Construção de índices
 - Índice de polaridade política (Getzkow et al, 2014)
 - Índice de incerteza político-econômica (EPU) (Bloom e Baker, 2016)
- Medindo variáveis econômicas
 - Medindo a pobreza com dados de celular (Blumenstock et al, 2015) ou com luminosidade de imagens de satélite (Jean et al, 2016)
 - Medindo a riqueza por ruas usando Google Street View (Glaeser et al, 2016)

Novas formas de mensurar atividade econômica

- Construção de índices
 - Índice de polaridade política (Getzkow et al, 2014)
 - Índice de incerteza político-econômica (EPU) (Bloom e Baker, 2016)
- Medindo variáveis econômicas
 - Medindo a pobreza com dados de celular (Blumenstock et al, 2015) ou com luminosidade de imagens de satélite (Jean et al, 2016)
 - Medindo a riqueza por ruas usando Google Street View (Glaeser et al, 2016)

Novas formas de mensurar atividade econômica

- Construção de índices
 - Índice de polaridade política (Getzkow et al, 2014)
 - Índice de incerteza político-econômica (EPU) (Bloom e Baker, 2016)
- Medindo variáveis econômicas
 - Medindo a pobreza com dados de celular (Blumenstock et al, 2015) ou com luminosidade de imagens de satélite (Jean et al, 2016)
 - Medindo a riqueza por ruas usando Google Street View (Glaeser et al, 2016)

Novas formas de mensurar atividade econômica

- Construção de índices
 - Índice de polaridade política (Getzkow et al, 2014)
 - Índice de incerteza político-econômica (EPU) (Bloom e Baker, 2016)
- Medindo variáveis econômicas
 - Medindo a pobreza com dados de celular (Blumenstock et al, 2015) ou com luminosidade de imagens de satélite (Jean et al, 2016)
 - Medindo a riqueza por ruas usando Google Street View (Glaeser et al, 2016)

Problemas de previsão em políticas públicas

Kleinberg et al (2015) analisaram como previsões podem ser aplicadas nas políticas públicas.

Em geral, estamos preocupados com os determinantes de algum problema social, isto é, qual a causa fundamental dos problemas sociais.

Causalidade é diferente de previsibilidade.

Métodos de ML são especializados em realizar previsão e não são diretamente aplicados para análise de causalidade.

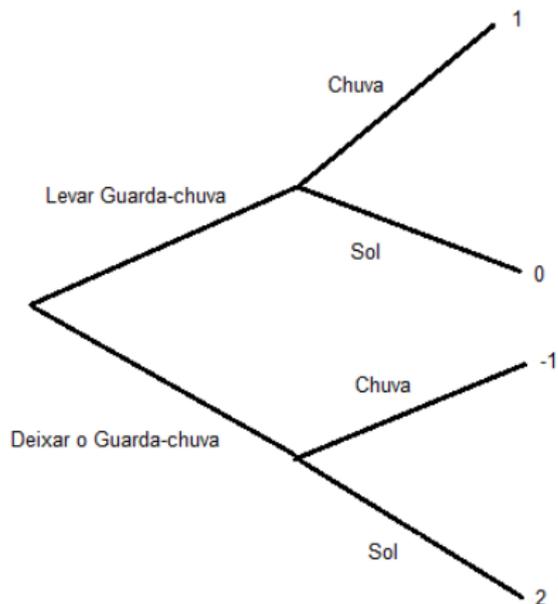
Mas previsão tem utilidade em políticas públicas?

A utilidade de ML em políticas públicas é a mesma da sua aplicação no setor privado.

A utilidade de ML em políticas públicas é a mesma da sua aplicação no setor privado.

Métodos de ML ajudam na tomada de decisões se suas previsões forem melhores que as previsões realizadas por agentes públicos.

Tomada de decisão em geral



O que podemos aprender deste problema simples?

- 1 Melhorar a capacidade de previsão do estado da natureza (chuva ou sol) ajuda na tomada de decisão com menor risco
- 2 Não precisamos de uma explicação das causas que determinam as chuvas. Simplesmente, uma previsão tem que funcionar!

O que podemos aprender deste problema simples?

- 1 Melhorar a capacidade de previsão do estado da natureza (chuva ou sol) ajuda na tomada de decisão com menor risco
- 2 Não precisamos de uma explicação das causas que determinam as chuvas. Simplesmente, uma previsão tem que funcionar!

O que podemos aprender deste problema simples?

- 1 Melhorar a capacidade de previsão do estado da natureza (chuva ou sol) ajuda na tomada de decisão com menor risco
- 2 Não precisamos de uma explicação das causas que determinam as chuvas. Simplesmente, uma previsão tem que funcionar!

Métodos de ML tem sido aplicados no setor privado em várias atividades onde uma melhor previsão ajuda a reduzir os riscos das decisões.

Exemplos

- Fintechs de crédito pessoal usam ML para saber para quem realizar empréstimos
- Amazon utiliza ML para recomendar compras associadas
- Walmart usa ML para determinar itens nos estoques em diferentes regiões

Métodos de ML tem sido aplicados no setor privado em várias atividades onde uma melhor previsão ajuda a reduzir os riscos das decisões.

Exemplos

- Fintechs de crédito pessoal usam ML para saber para quem realizar empréstimos
- Amazon utiliza ML para recomendar compras associadas
- Walmart usa ML para determinar itens nos estoques em diferentes regiões

Métodos de ML tem sido aplicados no setor privado em várias atividades onde uma melhor previsão ajuda a reduzir os riscos das decisões.

Exemplos

- Fintechs de crédito pessoal usam ML para saber para quem realizar empréstimos
- Amazon utiliza ML para recomendar compras associadas
- Walmart usa ML para determinar itens nos estoques em diferentes regiões

Métodos de ML tem sido aplicados no setor privado em várias atividades onde uma melhor previsão ajuda a reduzir os riscos das decisões.

Exemplos

- Fintechs de crédito pessoal usam ML para saber para quem realizar empréstimos
- Amazon utiliza ML para recomendar compras associadas
- Walmart usa ML para determinar itens nos estoques em diferentes regiões

Os mesmos princípios podem ser aplicados em políticas públicas

Exemplos

- Saber qual estudante irá evadir da escola permite intervir antes para tentar evitar a evasão
- Sinais de trânsito podem ser programados para serem mais rápidos em horários de grande movimento
- Comprar medicamentos em períodos de baixa demanda pode ajudar a economizar nos gastos.

Os mesmos princípios podem ser aplicados em políticas públicas

Exemplos

- Saber qual estudante irá evadir da escola permite intervir antes para tentar evitar a evasão
- Sinais de trânsito podem ser programados para serem mais rápidos em horários de grande movimento
- Comprar medicamentos em períodos de baixa demanda pode ajudar a economizar nos gastos.

Os mesmos princípios podem ser aplicados em políticas públicas

Exemplos

- Saber qual estudante irá evadir da escola permite intervir antes para tentar evitar a evasão
- Sinais de trânsito podem ser programados para serem mais rápidos em horários de grande movimento
- Comprar medicamentos em períodos de baixa demanda pode ajudar a economizar nos gastos.

Os mesmos princípios podem ser aplicados em políticas públicas

Exemplos

- Saber qual estudante irá evadir da escola permite intervir antes para tentar evitar a evasão
- Sinais de trânsito podem ser programados para serem mais rápidos em horários de grande movimento
- Comprar medicamentos em períodos de baixa demanda pode ajudar a economizar nos gastos.

Os mesmos princípios podem ser aplicados em políticas públicas

Exemplos

- Saber qual estudante irá evadir da escola permite intervir antes para tentar evitar a evasão (Previsão: Risco de evasão)
- Sinais de trânsito podem ser programados para serem mais rápidos em horários de grande movimento
- Comprar medicamentos em períodos de baixa demanda pode ajudar a economizar nos gastos.

Os mesmos princípios podem ser aplicados em políticas públicas

Exemplos

- Saber qual estudante irá evadir da escola permite intervir antes para tentar evitar a evasão (Previsão: Risco de evasão)
- Sinais de trânsito podem ser programados para serem mais rápidos em horários de grande movimento (Previsão: hora e locais com grande movimento na cidade)
- Comprar medicamentos em períodos de baixa demanda pode ajudar a economizar nos gastos.

Os mesmos princípios podem ser aplicados em políticas públicas

Exemplos

- Saber qual estudante irá evadir da escola permite intervir antes para tentar evitar a evasão (Previsão: Risco de evasão)
- Sinais de trânsito podem ser programados para serem mais rápidos em horários de grande movimento (Previsão: hora e locais com grande movimento na cidade)
- Comprar medicamentos em períodos de baixa demanda pode ajudar a economizar nos gastos (Previsão: Quando acontecerá um surto de virose)

Limitações

Se pudessemos ter previsto a ocorrência de uma pandemia em escala global, poderíamos ter comprado respiradores quando a demanda estava baixa.

Nem sempre os dados permitem fazer boas previsões!

Quem deve aguardar o julgamento em liberdade?

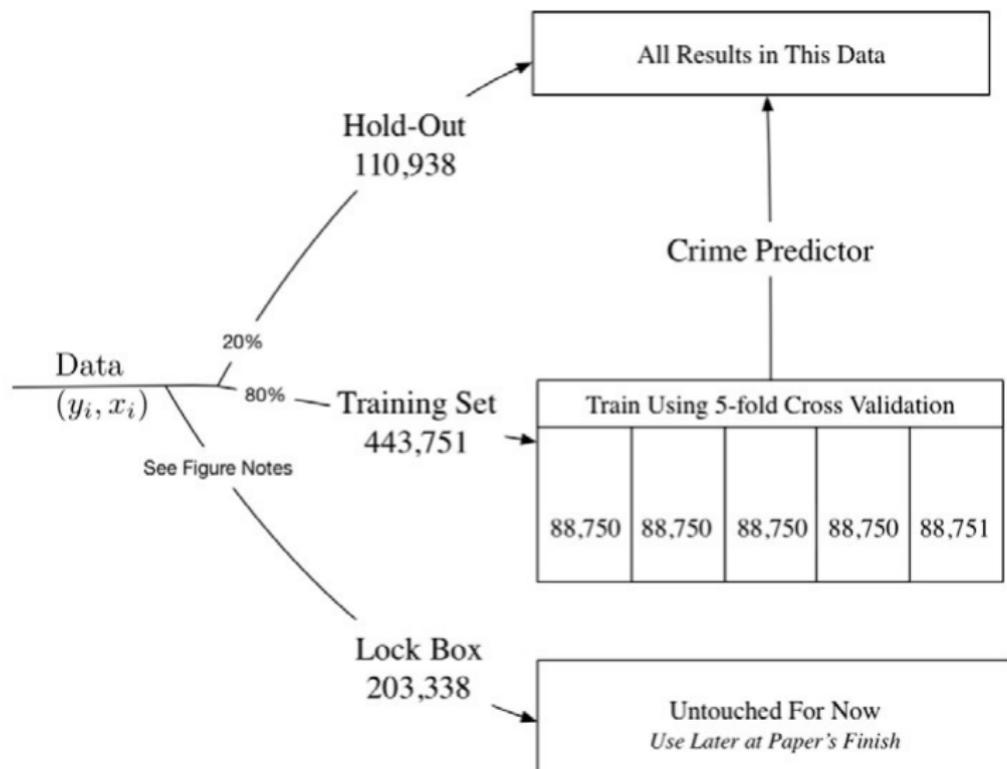
Todos os dias juizes precisam decidir se devem autorizar um preso a aguardar o julgamento por um crime em liberdade.

Apesar de existir algumas regras básicas, no geral, juízes precisam prever quais presos deve aguardar o julgamento em liberdade baseados em sua experiência. Espera-se que os presos liberados para julgamento em liberdade retornem para o julgamento ou não sejam novamente presos antes do julgamento.

Kleinberg et al (2015) questionaram se os métodos de ML podem ajudar os juízes a decidir quem deve ou não aguardar o julgamento em liberdade.

- Base de dados de todos os presos em Nova York durante 2008 a 2013 (1 460 462 casos)
- controles: histórico de condenações, variáveis individuais (gênero, cor e etnia, idade), tipo de crime (ex. usou arma de fogo ou não)

	Full Sample	Judge Releases	Judge Detains	P-value
Sample Size	554,689	408,283	146,406	
Release Rate	.7361	1.0000	0.00	
Outcomes				
Failure to Appear (FTA)	.1521	.1521		
Arrest (NCA)	.2581	.2581		
Violent Crime (NVCA)	.0372	.0372		
Murder, Rape, Robbery (NMRR)	.0187	.0187		
Defendant Characteristics				
Age	31.98	31.32	33.84	<.0001
Male	.8315	.8086	.8955	<.0001
White	.1273	.1407	.0897	<.0001
African American	.4884	.4578	.5737	<.0001
Hispanic	.3327	.3383	.3172	<.0001
Arrest County				
Brooklyn	.2901	.2889	.2937	.0006
Bronx	.2221	.2172	.2356	<.0001
Manhattan	.2507	.2398	.2813	<.0001
Queens	.1927	.2067	.1535	<.0001
Staten Island	.0440	.0471	.0356	<.0001
Arrest Charge				
<i>Violent Crime</i>				
Violent Felony	.1478	.1193	.2272	<.0001
Murder, Rape, Robbery	.0581	.0391	.1110	<.0001
Aggravated Assault	.0853	.0867	.0812	<.0001
Simple Assault	.2144	.2434	.1335	<.0001
<i>Property Crime</i>				
Burglary	.0206	.0125	.0433	<.0001
Larceny	.0738	.0659	.0959	<.0001
MV Theft	.0067	.0060	.0087	<.0001
Arson	.0006	.0003	.0014	<.0001
Fraud	.0696	.0763	.0507	<.0001
<i>Other Crime</i>				
Weapons	.0515	.0502	.0552	<.0001
Sex Offenses	.0089	.0086	.0096	<.0001
Prostitution	.0139	.0161	.0078	<.0001
DUI	.0475	.0615	.0084	<.0001
Other	.1375	.1433	.1216	<.0001
Gun Charge	.0335	.0213	.0674	<.0001
<i>Drug Crime</i>				
Drug Felony	.1411	.1175	.2067	<.0001
Drug Misdemeanor	.1142	.1156	.1105	<.0001



Metodologia:

1. Estimaram o modelo na amostra de treino e testaram na amostra de validação (*hold-out sample*)
2. Escolheram o melhor modelo de previsão
3. Comparam a previsão do modelo com a previsão dos juízes

Resultados

- Crimes podem ser reduzidos 24,8% sem mudanças na taxa de encarceramento
- A população carcerária pode ser reduzida em 48% sem aumentar a taxa de criminalidade
- O efeito parece ser heterogêneo para cada tipo de crime.

Resultados

- Crimes podem ser reduzidos 24,8% sem mudanças na taxa de encarceramento
- A população carcerária pode ser reduzida em 48% sem aumentar a taxa de criminalidade
- O efeito parece ser heterogêneo para cada tipo de crime.

Resultados

- Crimes podem ser reduzidos 24,8% sem mudanças na taxa de encarceramento
- A população carcerária pode ser reduzida em 48% sem aumentar a taxa de criminalidade
- O efeito parece ser heterogêneo para cada tipo de crime.

Resultados

- Crimes podem ser reduzidos 24,8% sem mudanças na taxa de encarceramento
- A população carcerária pode ser reduzida em 48% sem aumentar a taxa de criminalidade
- O efeito parece ser heterogêneo para cada tipo de crime.

Problemas típicos de Big Data em economia

Em economia, o problema mais comum de Big Data é conhecido como **Maldição da Dimensionalidade**

Este problema pode ocorrer tanto em aplicações empíricas macro quanto microeconômicas:

- Métodos tradicionais de estimação não são adequados quando o número de variáveis é maior que o número de observações
- Com muitas variáveis, quais são relevantes?

Em economia, o problema mais comum de Big Data é conhecido como **Maldição da Dimensionalidade**

Este problema pode ocorrer tanto em aplicações empíricas macro quanto microeconômicas:

- Métodos tradicionais de estimação não são adequados quando o número de variáveis é maior que o número de observações
- Com muitas variáveis, quais são relevantes?

Em economia, o problema mais comum de Big Data é conhecido como **Maldição da Dimensionalidade**

Este problema pode ocorrer tanto em aplicações empíricas macro quanto microeconômicas:

- Métodos tradicionais de estimação não são adequados quando o número de variáveis é maior que o número de observações
- Com muitas variáveis, quais são relevantes?

Exemplos:

- Macro: Muitas variáveis são utilizadas para realizar previsão em macroeconomia, muitas vezes o número de variáveis é maior que o número de observações temporais;
 - Stock e Watson (2002) utilizam 149 variáveis para prever 156 observações trimestrais
- Micro: Aplicações micro que possam ter efeito spillover entre suas unidades, requerem muitas variáveis para serem controladas.
 - Chodorow-Reich (2020) ressalta a dificuldade de controlar efeitos spillovers em análises regionais que mensuram o impacto de gasto público local.

Exemplos:

- Macro: Muitas variáveis são utilizadas para realizar previsão em macroeconomia, muitas vezes o número de variáveis é maior que o número de observações temporais;
 - Stock e Watson (2002) utilizam 149 variáveis para prever 156 observações trimestrais
- Micro: Aplicações micro que possam ter efeito spillover entre suas unidades, requerem muitas variáveis para serem controladas.
 - Chodorow-Reich (2020) ressalta a dificuldade de controlar efeitos spillovers em análises regionais que mensuram o impacto de gasto público local.

Exemplos:

- Macro: Muitas variáveis são utilizadas para realizar previsão em macroeconomia, muitas vezes o número de variáveis é maior que o número de observações temporais;
 - Stock e Watson (2002) utilizam 149 variáveis para prever 156 observações trimestrais
- Micro: Aplicações micro que possam ter efeito spillover entre suas unidades, requerem muitas variáveis para serem controladas.
 - Chodorow-Reich (2020) ressalta a dificuldade de controlar efeitos spillovers em análises regionais que mensuram o impacto de gasto público local.

Exemplos:

- Macro: Muitas variáveis são utilizadas para realizar previsão em macroeconomia, muitas vezes o número de variáveis é maior que o número de observações temporais;
 - Stock e Watson (2002) utilizam 149 variáveis para prever 156 observações trimestrais
- Micro: Aplicações micro que possam ter efeito spillover entre suas unidades, requerem muitas variáveis para serem controladas.
 - Chodorow-Reich (2020) ressalta a dificuldade de controlar efeitos spillovers em análises regionais que mensuram o impacto de gasto público local.

Exemplos:

- Macro: Muitas variáveis são utilizadas para realizar previsão em macroeconomia, muitas vezes o número de variáveis é maior que o número de observações temporais;
 - Stock e Watson (2002) utilizam 149 variáveis para prever 156 observações trimestrais
- Micro: Aplicações micro que possam ter efeito spillover entre suas unidades, requerem muitas variáveis para serem controladas.
 - Chodorow-Reich (2020) ressalta a dificuldade de controlar efeitos spillovers em análises regionais que mensuram o impacto de gasto público local.

Suponha o seguinte modelo linear geral:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i$$

$$i = 1, \dots, n$$

Uma estimação não viesada para os $\beta = (\beta_0, \dots, \beta_p)$ é obtida pelos MQO:

$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

Em que: $Y = (y_1 \dots y_n)'$, $X = (X_{i1} \dots X_{ip})$, tal que:
 $X_{ik} = (x_{1k} \dots x_{nk})'$.

Note que:

- Se $p, n \rightarrow \infty$ inverter a matriz $(X'X)$ é um desafio até para computadores avançados;
- Se $p > n$ então $\hat{\beta}$ possui muitas soluções.

Em que: $Y = (y_1 \dots y_n)'$, $X = (X_{i1} \dots X_{ip})$, tal que:
 $X_{ik} = (x_{1k} \dots x_{nk})'$.

Note que:

- Se $p, n \rightarrow \infty$ inverter a matriz $(X'X)$ é um desafio até para computadores avançados;
- Se $p > n$ então $\hat{\beta}$ possui muitas soluções.

Em que: $Y = (y_1 \dots y_n)'$, $X = (X_{i1} \dots X_{ip})$, tal que:
 $X_{ik} = (x_{1k} \dots x_{nk})'$.

Note que:

- Se $p, n \rightarrow \infty$ inverter a matriz $(X'X)$ é um desafio até para computadores avançados;
- Se $p > n$ então $\hat{\beta}$ possui muitas soluções.

Algumas soluções propostas para lidar com este problema são

- 1 Aplicar métodos que permitam selecionar previamente as variáveis mais relevantes;
- 2 Aplicar métodos que condensem as variáveis em um conjunto menor de variáveis latentes (redução da dimensão)

Algumas soluções propostas para lidar com este problema são

- 1 Aplicar métodos que permitam selecionar previamente as variáveis mais relevantes;
- 2 Aplicar métodos que condensem as variáveis em um conjunto menor de variáveis latentes (redução da dimensão)

Algumas soluções propostas para lidar com este problema são

- 1 Aplicar métodos que permitam selecionar previamente as variáveis mais relevantes;
- 2 Aplicar métodos que condensem as variáveis em um conjunto menor de variáveis latentes (redução da dimensão)

O problema de utilizar tais métodos é

- 1 Eles não geram inferência confiável (Leeb and Potsher, 2006)
- 2 São viesados
- 3 **Sua utilização em problemas de identificação causal requer cuidado!**

O problema de utilizar tais métodos é

- 1 Eles não geram inferência confiável (Leeb and Potosher, 2006)
- 2 São viesados
- 3 Sua utilização em problemas de identificação causal requer cuidado!

O problema de utilizar tais métodos é

- 1 Eles não geram inferência confiável (Leeb and Potosher, 2006)
- 2 São viesados
- 3 Sua utilização em problemas de identificação causal requer cuidado!

O problema de utilizar tais métodos é

- 1 Eles não geram inferência confiável (Leeb and Potosher, 2006)
- 2 São viesados
- 3 **Sua utilização em problemas de identificação causal requer cuidado!**

Introdução à avaliação de políticas públicas

Um dos mais importantes problemas em políticas públicas consiste em saber: Qual o efeito de algum programa (política) sobre determinada variável de interesse.

Exemplos

- 1 Qual o efeito de um programa de treinamento sobre o salário (medida de produtividade) de um indivíduo;
- 2 Qual o efeito da exposição a bons professores sobre os resultados educacionais e de mercado de trabalho dos alunos (Chetty et al (2014));
- 3 Qual o impacto de se viver em um bairro melhor sobre a educação dos filhos: (Hendren e Chetty (2018));
- 4 Qual o impacto da austeridade sobre a atividade econômica (Jordá e Taylor (2015)).

Um dos mais importantes problemas em políticas públicas consiste em saber: Qual o efeito de algum programa (política) sobre determinada variável de interesse.

Exemplos

- 1 Qual o efeito de um programa de treinamento sobre o salário (medida de produtividade) de um indivíduo;
- 2 Qual o efeito da exposição a bons professores sobre os resultados educacionais e de mercado de trabalho dos alunos (Chetty et al (2014));
- 3 Qual o impacto de se viver em um bairro melhor sobre a educação dos filhos: (Hendren e Chetty (2018));
- 4 Qual o impacto da austeridade sobre a atividade econômica (Jordá e Taylor (2015)).

Um dos mais importantes problemas em políticas públicas consiste em saber: Qual o efeito de algum programa (política) sobre determinada variável de interesse.

Exemplos

- 1 Qual o efeito de um programa de treinamento sobre o salário (medida de produtividade) de um indivíduo;
- 2 Qual o efeito da exposição a bons professores sobre os resultados educacionais e de mercado de trabalho dos alunos (Chetty et al (2014));
- 3 Qual o impacto de se viver em um bairro melhor sobre a educação dos filhos: (Hendren e Chetty (2018));
- 4 Qual o impacto da austeridade sobre a atividade econômica (Jordá e Taylor (2015)).

Um dos mais importantes problemas em políticas públicas consiste em saber: Qual o efeito de algum programa (política) sobre determinada variável de interesse.

Exemplos

- 1 Qual o efeito de um programa de treinamento sobre o salário (medida de produtividade) de um indivíduo;
- 2 Qual o efeito da exposição a bons professores sobre os resultados educacionais e de mercado de trabalho dos alunos (Chetty et al (2014));
- 3 Qual o impacto de se viver em um bairro melhor sobre a educação dos filhos: (Hendren e Chetty (2018));
- 4 Qual o impacto da austeridade sobre a atividade econômica (Jordá e Taylor (2015)).

Um dos mais importantes problemas em políticas públicas consiste em saber: Qual o efeito de algum programa (política) sobre determinada variável de interesse.

Exemplos

- 1 Qual o efeito de um programa de treinamento sobre o salário (medida de produtividade) de um indivíduo;
- 2 Qual o efeito da exposição a bons professores sobre os resultados educacionais e de mercado de trabalho dos alunos (Chetty et al (2014));
- 3 Qual o impacto de se viver em um bairro melhor sobre a educação dos filhos: (Hendren e Chetty (2018));
- 4 Qual o impacto da austeridade sobre a atividade econômica (Jordá e Taylor (2015)).

Métodos estatísticos que buscam gerar estratégias empíricas para responder tais perguntas são chamados de MÉTODOS DE ANÁLISE DE CAUSALIDADE.

Iremos revisar estes métodos por que muitas das aplicações de ML estão sendo realizadas para melhorar a análise de causalidade (Ver Athey e Imbens (2018) para uma revisão)

Análise de causalidade tem sido realizada em ciências médicas e biológicas desde o início do século passado (Fisher (1925), Neyman(1923))

O principal método utilizado é o experimento randomizado controlado (Randomized Control Trials (RCT)). Esses métodos permitem identificar adequadamente o efeito causal do tratamento sobre os resultados de interesse.

RCT's possuem inúmeras limitações quando aplicados a ciências sociais. Apesar de existir cada vez mais aplicações de RCT, seu uso como estratégia empírica não é factível em vários problemas de interesse.

Limitações da RCT

- 1 Expor alunos a bons e maus professores aleatoriamente não é eticamente aceitável. Especialmente por que o efeito de tal tratamento tem repercussões de longo prazo.
- 2 Muitos problemas decorrem de dados observacionais
- 3 Não é claro a validade externa de políticas testadas por RCT

Limitações da RCT

- 1 Expor alunos a bons e maus professores aleatoriamente não é eticamente aceitável. Especialmente por que o efeito de tal tratamento tem repercussões de longo prazo.
- 2 Muitos problemas decorrem de dados observacionais
- 3 Não é claro a validade externa de políticas testadas por RCT

Limitações da RCT

- 1 Expor alunos a bons e maus professores aleatoriamente não é eticamente aceitável. Especialmente por que o efeito de tal tratamento tem repercussões de longo prazo.
- 2 Muitos problemas decorrem de dados observacionais
- 3 Não é claro a validade externa de políticas testadas por RCT

Limitações da RCT

- 1 Expor alunos a bons e maus professores aleatoriamente não é eticamente aceitável. Especialmente por que o efeito de tal tratamento tem repercussões de longo prazo.
- 2 Muitos problemas decorrem de dados observacionais
- 3 Não é claro a validade externa de políticas testadas por RCT

Dessa forma, a análise de causalidade em ciências sociais tem se dedicado a resolver estes problemas.

Alguns métodos comuns para realizar avaliação de políticas públicas são:

- 1 Experimentos aleatórios controlados (RCT)
- 2 Seleção sobre observáveis
- 3 Variáveis Instrumentais
- 4 Regressão em discontinuidade
- 5 Métodos para dados em painel: Diferença em diferença (dif-dif) e efeitos fixos.

Dessa forma, a análise de causalidade em ciências sociais tem se dedicado a resolver estes problemas.

Alguns métodos comuns para realizar avaliação de políticas públicas são:

- 1 Experimentos aleatórios controlados (RCT)
- 2 Seleção sobre observáveis
- 3 Variáveis Instrumentais
- 4 Regressão em discontinuidade
- 5 Métodos para dados em painel: Diferença em diferença (dif-dif) e efeitos fixos.

Dessa forma, a análise de causalidade em ciências sociais tem se dedicado a resolver estes problemas.

Alguns métodos comuns para realizar avaliação de políticas públicas são:

- 1 Experimentos aleatórios controlados (RCT)
- 2 Seleção sobre observáveis
- 3 Variáveis Instrumentais
- 4 Regressão em discontinuidade
- 5 Métodos para dados em painel: Diferença em diferença (dif-dif) e efeitos fixos.

Dessa forma, a análise de causalidade em ciências sociais tem se dedicado a resolver estes problemas.

Alguns métodos comuns para realizar avaliação de políticas públicas são:

- 1 Experimentos aleatórios controlados (RCT)
- 2 Seleção sobre observáveis
- 3 Variáveis Instrumentais
- 4 Regressão em discontinuidade
- 5 Métodos para dados em painel: Diferença em diferença (dif-dif) e efeitos fixos.

Dessa forma, a análise de causalidade em ciências sociais tem se dedicado a resolver estes problemas.

Alguns métodos comuns para realizar avaliação de políticas públicas são:

- 1 Experimentos aleatórios controlados (RCT)
- 2 Seleção sobre observáveis
- 3 Variáveis Instrumentais
- 4 Regressão em discontinuidade
- 5 Métodos para dados em painel: Diferença em diferença (dif-dif) e efeitos fixos.

Dessa forma, a análise de causalidade em ciências sociais tem se dedicado a resolver estes problemas.

Alguns métodos comuns para realizar avaliação de políticas públicas são:

- 1 Experimentos aleatórios controlados (RCT)
- 2 Seleção sobre observáveis
- 3 Variáveis Instrumentais
- 4 Regressão em discontinuidade
- 5 Métodos para dados em painel: Diferença em diferença (dif-dif) e efeitos fixos.

Referências

- Papers

- Recent Developments in the Econometrics of Program Evaluation - Imbens and Wooldridge (2009)
- Econometric Methods for Program Evaluation - Abadie e Cattaneo (2018)

- Livros

- Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction - Rubin and Imbens (2015)
- Mostly Harmless Econometrics - Angrist and Pischke (2008)

Referências

● Papers

- 1 Recent Developments in the Econometrics of Program Evaluation - Imbens and Wooldridge (2009)
- 2 Econometric Methods for Program Evaluation - Abadie e Cattaneo (2018)

● Livros

- 1 Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction - Rubin and Imbens (2015)
- 2 Mostly Harmless Econometrics - Angrist and Pischke (2008)

Referências

• Papers

- 1 Recent Developments in the Econometrics of Program Evaluation - Imbens and Wooldridge (2009)
- 2 Econometric Methods for Program Evaluation - Abadie e Cattaneo (2018)

• Livros

- 1 Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction - Rubin and Imbens (2015)
- 2 Mostly Harmless Econometrics - Angrist and Pischke (2008)

Referências

• Papers

- 1 Recent Developments in the Econometrics of Program Evaluation - Imbens and Wooldridge (2009)
- 2 Econometric Methods for Program Evaluation - Abadie e Cattaneo (2018)

• Livros

- 1 Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction - Rubin and Imbens (2015)
- 2 Mostly Harmless Econometrics - Angrist and Pischke (2008)

Referências

• Papers

- 1 Recent Developments in the Econometrics of Program Evaluation - Imbens and Wooldridge (2009)
- 2 Econometric Methods for Program Evaluation - Abadie e Cattaneo (2018)

• Livros

- 1 Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction - Rubin and Imbens (2015)
- 2 Mostly Harmless Econometrics - Angrist and Pischke (2008)

Referências

• Papers

- 1 Recent Developments in the Econometrics of Program Evaluation - Imbens and Wooldridge (2009)
- 2 Econometric Methods for Program Evaluation - Abadie e Cattaneo (2018)

• Livros

- 1 Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction - Rubin and Imbens (2015)
- 2 Mostly Harmless Econometrics - Angrist and Pischke (2008)

Referências

• Papers

- 1 Recent Developments in the Econometrics of Program Evaluation - Imbens and Wooldridge (2009)
- 2 Econometric Methods for Program Evaluation - Abadie e Cattaneo (2018)

• Livros

- 1 Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction - Rubin and Imbens (2015)
- 2 Mostly Harmless Econometrics - Angrist and Pischke (2008)

Experimentos Randomizados Controlados (RCT)

Suponha que se deseje avaliar um programa (ou uma política ou um tratamento) cujo resultado seja dado por Y_i . Ou seja, Y_i refere-se ao resultado observado decorrente ou não de um tratamento qualquer.

Assuma que o tratamento é discreto e possa ser representado por:

$$D_i = \begin{cases} 1 & \text{se } i \text{ recebeu o tratamento} \\ 0 & \text{caso contrário} \end{cases}$$

Ou seja, $D_i = \{0, 1\}$.

Define-se como resultado potencial: Y_{1i} se o indivíduo i recebeu o tratamento ($D_i = 1$) e Y_{0i} se o indivíduo i não recebeu o tratamento ($D_i = 0$).

O resultado potencial refere-se ao possível resultado decorrente do indivíduo i ter recebido ou não o tratamento.

Ex: Qual o efeito de um programa de treinamento sobre o salário.

1. Y_i salário de um indivíduo qualquer;
2. D_i dummy indicando se o indivíduo i recebeu o treinamento;

Podemos escrever o resultado observado a partir dos resultado potencial:

$$Y_i = Y_{1i}(D_i) + Y_{0i}(1 - D_i) = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

O nosso objetivo é obter uma estimativa para $Y_{1i} - Y_{0i}$ que é o efeito causal de se receber determinado tratamento.

O problema fundamental em análise da causalidade (HOLLAND (1986)) decorre da observação de que não é possível calcular $Y_{1i} - Y_{0i}$ por que um mesmo indivíduo não pode ser observado, ao mesmo tempo, como tendo recebido o tratamento e como tendo não recebido o tratamento.

Uma solução é observar o efeito médio, isto é, comparar o resultado potencial médio de quem recebeu o tratamento $E[Y_i | D_i = 1]$ contra o resultado potencial médio de quem não recebeu o tratamento $E[Y_i | D_i = 0]$.

Nesse caso, temos:

$$\tau = E[Y_i | D_i = 1] - E[Y_i | D_i = 0]$$

Observe que podemos decompor o *ATE* da seguinte forma:

$$E[Y_i | D_i = 1] - E[Y_i | D_i = 0] = \quad (1)$$

$$E[Y_{1i} | D_i = 1] - E[Y_{0i} | D_i = 1] + E[Y_{0i} | D_i = 1] - E[Y_{0i} | D_i = 0]$$

O primeiro termo do lado direito chama-se Efeito Tratamento Médio sobre os Tratados (ATET):

$$ATET = E[Y_{1i} | D_i = 1] - E[Y_{0i} | D_i = 1] = E[Y_{1i} - Y_{0i} | D_i = 1]$$

O segundo termo é chamado de viés de seleção:

$$Viés = E[Y_{0i} | D_i = 1] - E[Y_{0i} | D_i = 0]$$

O termo de viés refere-se a diferença esperada daqueles que apresentam potencial tratamento Y_{0i} mas que podem ou não ter recebido o tratamento.

Este viés pode alterar o resultado do efeito tratamento médio.

O viés é causado por não conseguirmos contornar o problema fundamental da análise de causalidade e necessitarmos estimar o efeito médio.

A diferença média entre os resultados observados de quem recebeu e não recebeu o tratamento ($E[Y_i | D_i = 1] - E[Y_i | D_i = 0]$) refere-se a uma correlação. Porém, correlação não é o mesmo que causalidade, por isso que existe o termo de *Viés*.

Assim, para que $\tau = E[Y_i | D_i = 1] - E[Y_i | D_i = 0]$ represente efetivamente o efeito causal de receber ou não o tratamento é preciso que de alguma forma o termo de *Viés* seja eliminado.

Observe que o termo de viés ele depende de como o tratamento é atribuído aos indivíduos:

$$\text{Viés} = E[Y_{0i} | D_i = 1] - E[Y_{0i} | D_i = 0]$$

Se o tratamento for distribuído independente do resultado potencial, então:

$$E[Y_{0i} | D_i = 1] = E[Y_{0i} | D_i = 0] = E(Y_{0i})$$

O termo de viés desaparece.

Dessa forma, ao se assumir que a atribuição do tratamento é independente dos resultados potenciais ($Y_{1i}, Y_{0i} \perp D_i$), então a comparação entre os resultados potenciais de quem recebeu e quem não recebeu o tratamento será também causal.

Essa é a essência dos RCT em que o tratamento é distribuído aos indivíduos independentemente dos resultados potenciais.

A análise de causalidade neste caso é bastante simples. Considere as estimativas empíricas de τ :

$$\tau = \frac{1}{n_1} \sum_{i=1}^{n_1} y_{1i} - \frac{1}{n_0} \sum_{i=1}^{n_0} y_{0i}$$

Em que: n_1 refere-se a quantidade de indivíduos que receberam o tratamento e n_0 refere-se a quantidade de indivíduos que não receberam o tratamento. y_{1i} e y_{0i} são os resultados observados de quem recebeu e de quem não recebeu o tratamento. Logo:

$$\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$$

Identificação por Atribuição Aleatória de Tratamento

O caso acima pode ser generalizado para obtenção de identificação por atribuição aleatória de tratamento.

Chama-se de identificação a estratégia empírica adotada para eliminar o *Viés* decorrente da comparação entre os observáveis.

Essa caracterização é melhor explicada por meio da análise de regressão.

Considere a seguinte regressão:

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

Em que: Y_i é o resultado observado; $D_i = (0, 1)$ é a atribuição do tratamento; ε_i é o erro. Comparando a regressão acima com o que foi visto na seção anterior temos que: $Y_{0i} = \alpha$, $ATE = \beta$ e ε_i seria o viés.

Sob a hipótese de que o tratamento é independente dos potenciais resultados, pode-se garantir que: $E(\varepsilon_i | D_i) = 0$ e nesse caso, $\hat{\beta}$ é uma estimativa não viesada do efeito do tratamento sobre o resultado observado.

Sob condições normais, a significância de $\hat{\beta}$ pode ser calculada pelo teste t e esse indica se o tratamento é ou não significativo para alterar a variável de resultado. Ou seja:

$$H_0 : \hat{\beta} = 0 \implies H_{o'} : \bar{Y}_1 - \bar{Y}_0 = 0$$

Nesse caso, a identificação do efeito causal é obtida ao se garantir que a distribuição do tratamento é feita de forma independente dos resultados potenciais.

Esse tipo de identificação não necessariamente está associado a experimentos randomizados controlados. Pode-se obter tal identificação em experimentos quase-randomizados.

Um experimento quase-randomizado é aquele cujo tratamento pode ser hipoteticamente assumido como independente do resultado potencial, porém, tal experimento não foi controlado, mas aconteceu na natureza sem o controle do pesquisador.

EXEMPLO: Water Scarcity and Birth Outcomes in the Brazilian Semi-arid (Soares e Rocha (2015))

A pergunta que eles querem responder é:

- Períodos caracterizados por escassez de água durante a gravidez afetam a taxa de mortalidade?
- Escassez de água é definido como períodos com elevada temperatura e baixa precipitação chuvosa.
- Usando a base de dados georeferenciada, eles calcularam para cada município do Nordeste qual a taxa de precipitação e temperatura mensalmente (r_{mt}).
- Com isso puderam calcular o efeito de escassez de água nos 12 meses durante a gravidez para cada mulher no município para explicar a taxa de mortalidade (e outras variáveis de saúde).

EXEMPLO: Water Scarcity and Birth Outcomes in the Brazilian Semi-arid (Soares e Rocha (2015))

A pergunta que eles querem responder é:

- Períodos caracterizados por escassez de água durante a gravidez afetam a taxa de mortalidade?
- Escassez de água é definido como períodos com elevada temperatura e baixa precipitação chuvosa.
- Usando a base de dados georeferenciada, eles calcularam para cada município do Nordeste qual a taxa de precipitação e temperatura mensalmente (r_{mt}).
- Com isso puderam calcular o efeito de escassez de água nos 12 meses durante a gravidez para cada mulher no município para explicar a taxa de mortalidade (e outras variáveis de saúde).

EXEMPLO: Water Scarcity and Birth Outcomes in the Brazilian Semi-arid (Soares e Rocha (2015))

A pergunta que eles querem responder é:

- Períodos caracterizados por escassez de água durante a gravidez afetam a taxa de mortalidade?
- Escassez de água é definido como períodos com elevada temperatura e baixa precipitação chuvosa.
- Usando a base de dados georeferenciada, eles calcularam para cada município do Nordeste qual a taxa de precipitação e temperatura mensalmente (r_{mt}).
- Com isso puderam calcular o efeito de escassez de água nos 12 meses durante a gravidez para cada mulher no município para explicar a taxa de mortalidade (e outras variáveis de saúde).

EXEMPLO: Water Scarcity and Birth Outcomes in the Brazilian Semi-arid (Soares e Rocha (2015))

A pergunta que eles querem responder é:

- Períodos caracterizados por escassez de água durante a gravidez afetam a taxa de mortalidade?
- Escassez de água é definido como períodos com elevada temperatura e baixa precipitação chuvosa.
- Usando a base de dados georeferenciada, eles calcularam para cada município do Nordeste qual a taxa de precipitação e temperatura mensalmente (r_{mt}).
- Com isso puderam calcular o efeito de escassez de água nos 12 meses durante a gravidez para cada mulher no município para explicar a taxa de mortalidade (e outras variáveis de saúde).

EXEMPLO: Water Scarcity and Birth Outcomes in the Brazilian Semi-arid (Soares e Rocha (2015))

A pergunta que eles querem responder é:

- Períodos caracterizados por escassez de água durante a gravidez afetam a taxa de mortalidade?
- Escassez de água é definido como períodos com elevada temperatura e baixa precipitação chuvosa.
- Usando a base de dados georeferenciada, eles calcularam para cada município do Nordeste qual a taxa de precipitação e temperatura mensalmente (r_{mt}).
- Com isso puderam calcular o efeito de escassez de água nos 12 meses durante a gravidez para cada mulher no município para explicar a taxa de mortalidade (e outras variáveis de saúde).

A regressão que eles estimaram é:

$$H_{myt} = \alpha + \beta R_{myt} + \phi_{mt} + \lambda_y + \varphi Trend_{gyt} + \pi T_{myt} + \epsilon_{myt},$$

$$D_{m\tau} = 1 \text{ if } \sum_{t=-11}^{\tau} r_{mt} < (\bar{r}_m - r_m^{SD}), \text{ and } 0 \text{ otherwise,}$$

Onde: H_{myt} é um resultado de saúde para o município m no mês t no ano y , R_{myt} é a medida de choque de escassez de água; $Trend_{gyt}$ tendência do grid g no ano y e no ano t , T_{myt} é a temperatura média dos 12 últimos meses no município t .

Por que β representa uma relação causal?

Ou melhor:

Por que $E(\epsilon_{myt} | R_{myt}) = 0$?

Argumentos dos autores

- 1 Choques de escassez não são facilmente previstos pelos indivíduos;
- 2 Mesmo que haja uma certa previsibilidade; não existem evidências de que as mulheres se programam para engravidar em períodos mais amenos.

Argumentos dos autores

- 1 Choques de escassez não são facilmente previstos pelos indivíduos;
- 2 Mesmo que haja uma certa previsibilidade; não existem evidências de que as mulheres se programam para engravidar em períodos mais amenos.

Argumentos dos autores

- 1 Choques de escassez não são facilmente previstos pelos indivíduos;
- 2 Mesmo que haja uma certa previsibilidade; não existem evidências de que as mulheres se programam para engravidar em períodos mais amenos.

Drought before Birth	2.087 (1.118)*	2.265 (1.109)**	3.984 (1.299)***	3.761 (1.324)***	4.594 (1.620)***
Rainfall 12-24 Months before Birth					-0.771 (2.327)
Rainfall 1-12 Months after Birth					-2.224 (2.469)
Observations	157,403	155,688	155,688	155,688	143,319
Number of Municipalities	1,048	1,037	1,037	1,037	1,037
Municipality × Month of Birth Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year of Birth Fixed Effects	Yes	Yes	Yes	Yes	Yes
Grid Time Trend	Yes	Yes	Yes	Yes	Yes
Exclude Top 1% in Number of Births	No	Yes	Yes	Yes	Yes
Weighted (Average Number of Newborns)	No	No	Yes	Yes	Yes
Temperature before Birth	No	No	No	Yes	Yes

Retornando a análise da regressão, perceba que Y_i tenha sua distribuição de probabilidade dependente de outras variáveis de controle observáveis (X_i). Nesse caso, a condição de identificação é obtida ao se supor que:

$$Y_{1i}|X_i, Y_{0i}|X_i \perp D_i$$

Daí, a regressão fica:

$$Y_i = \alpha + \beta D_i + \gamma' X_i + \varepsilon_i$$

Existe uma implicação importante para esta hipótese muito mais realista:

O **efeito tratamento** pode ser **heterogêneo**, no sentido de que tanto a magnitude quanto a direção podem se alterar em diferentes sub-grupos. Em termos de políticas públicas essa possibilidade permite trabalhar o efeito de políticas em grupos específicos de indivíduos.

Seleção sobre os observáveis

Uma forma alternativa de obter a identificação refere-se a seleção sobre os observáveis. Neste caso, o tratamento é designado de acordo com algumas variáveis observáveis.

A identificação neste caso é feita de maneira semelhante ao caso anterior, porém, é preciso controlar para as características observáveis que fizeram que determinado indivíduo recebesse ou não o tratamento.

Exemplo: Curso de Capacitação de Professores

Imagine que se deseje avaliar um Programa de Formação Continuada em Matemática (PFCM). Um tipo de treinamento in-service para os professores da rede estadual.

Problema: Qual o efeito do PFCM sobre o desempenho dos estudantes?

Em termos de regressão, podemos descrever tal problema como:

$$y_{ijt} = \alpha + \beta D_{jt-1} + \delta' X_{it} + \epsilon_{ijt}$$

Nesse modelo, não posso garantir que a estimativa de β seja causal, pois não tenho como afirmar se o professor que foi tratado pelo PFCM é independente do resultado potencial.

Por ora, vamos supor que os professores são atribuídos aleatoriamente aos estudantes.

- 1) Um professor que esteja participando do PFCM pode ser mais aplicado que professores que não estão participando. Nesse caso, independente do PFCM a diferença entre a performance dos alunos destes professores pode ser significativa.
- 2) Um outro caso é o do professor que se inscreveu para participar do PFCM, porém, devido ao número de vagas limitadas, não foi selecionado para participar. Nesse caso, qual o critério de escolha do professor que irá participar do PFCM?

Suponha que o regulamento de seleção do PFCM seja:

- a. Professores efetivos tem preferência em relação a professores com contrato temporário de trabalho
- b. Professores que efetivamente lecionam matemática tem preferência frente a professores formados em matemática, mas que ensinam outras matérias.

Essas duas características são observáveis e podem ser utilizada para controlar a atribuição do tratamento

Uma vez que tais variáveis (W_{jt}) sejam incluídas na análise de regressão, podemos assumir que o tratamento foi designado independentemente das características dos professores.

$$y_{ijt} = \alpha + \beta D_{jt-1} + \delta' X_{it} + \beta_2 W_{jt} + \epsilon_{ijt}$$

$$y_{1i}, y_{0i} \perp D_{jt-1} | W_{jt}$$

A essa hipótese dá-se o nome de: Hipótese de Independência Condicional (HIC)¹

¹Em inglês: Conditional Independence Assumption (CIA) 

Ou seja, a inclusão das variáveis de controle que determinam a forma como o controle é designado faz com que o problema de identificação seja considerado aleatório.

OBS: Note que a inclusão das variáveis W_{jt} resolvem o problema de identificação em 2), mas não em 1). Pois a motivação para que um professor participe do PFCM não é observada e não pode ser controlada.

Isto implica que comparar professores que se candidataram a participar com aqueles que não se candidataram pode gerar viés decorrente do interesse de cada professor em participar do programa.

A Hipótese de independência condicional (HIC) faz parte de um caso mais geral de situação que será bastante útil para entender como identificar efeitos causais e como métodos de Machine Learning podem ajudar na identificação.

Seja o exemplo anterior novamente e considere a seguinte regressão:

$$y_{ijt} = \alpha + \beta D_{jt-1} + \delta' X_{it} + \epsilon_{ijt}$$

$$\epsilon_{ijt} = \gamma_1 W_{jt} + \gamma_2 Int_{jt} + u_{ijt}$$

Em que: Int_{jt-1} representa o interesse do professor j em se inscrever no tempo $t - 1$ no PFCM.

A variável Int_{jt-1} é bastante difícil de ser adequadamente mensurada, pois vários fatores podem determiná-la. Tal variável é correlacionada com o a atribuição do tratamento, isto é:
 $E(D_{jt-1}, Int_{jt-1}) \neq 0$.

Note que W_{jt} é observável, porém Int_{jt-1} não é diretamente observável.

Não incluir tais variáveis no modelo pode gerar o viés por omissão de variáveis (VOV). Todavia, no caso de W_{jt} é possível incluí-la, já no caso de Int_{jt-1} não existe essa possibilidade pois esta não é observável.