

Causalidade e Estatística

Guilherme Jardim Duarte

Jota

guilherme.duarte@jota.info

November 24, 2018

- Duas revoluções:
 - Statistical Learning - Friedman et al. (2001); James et al. (2013)
 - Causalidade - Pearl (2009); Imbens Rubin (2015)
- Uma das revoluções tem como foco predição. Redução dos Erros
- a Outra tem como foco melhorar a predição quando realizamos uma intervenção.

- O que significa realizar uma intervenção?
- Exemplo 1: Suponha que desejamos saber que tipo de desconto dar para clientes de forma a maximizar o valor pago. Temos dados sobre descontos dados
- Exemplo 2: Queremos entender se hospitais melhoram a saúde dos pacientes. Temos dados de saúde sobre indivíduos que entraram e não entraram no hospital.
- Como resolver esse problema?
- $P(Y|x) \neq P(Y|do(x))$

- Problema fundamental da Inferência Causal: Nunca saberemos $y_1 - y_0$
- Solução: Buscar $E(y_1) - E(y_0)$
- Porém $E(y_1|D = 1) \neq E(y_1|D = 0)$ e $E(y_0|D = 1) \neq E(y_0|D = 0)$
- Geralmente temos $E(y_1|D = 1) - E(y_0|D = 0)$
- Que é equivalente a
$$E(y_1|D = 1) - E(y_0|D = 1) + E(y_0|D = 0) - E(y_0|D = 0)$$
- Como resolver?

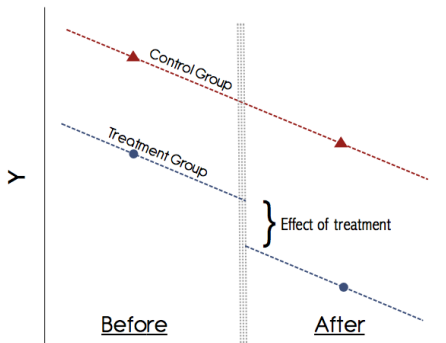
- Experimento Controlado.
- Métodos observacionais
 - Regressão
 - Matching
 - Diferenças em Diferenças
 - Regressão Descontínua
 - Que é equivalente a

$$E(y_1|D = 1) - E(y_0|D = 1) + E(y_0|D = 0) - E(y_0|D = 0)$$

- Suposição: Seleção em Observáveis
- Balanceamento de Data points a fim de que haja a mesma probabilidade de $P(D=1|X)$
- Técnicas
 - Nearest Neighbors
 - Exact
 - Optimal
- Problema: Variáveis não observadas

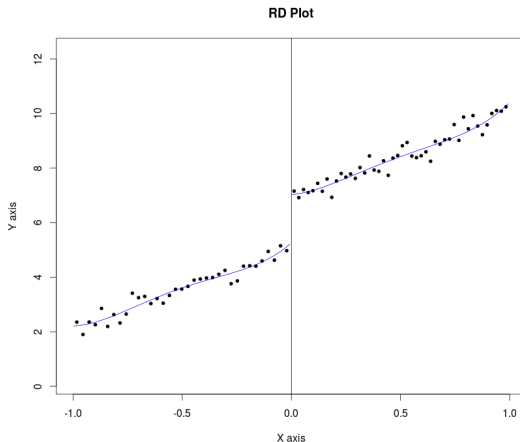
Diferenças em Diferenças

- Tempo ajuda a capturar efeito causal
- $(E(Y|D=1, t=1) - E(Y|D=1, t=0)) - (E(Y|D=0, t=1) - E(Y|D=0, t=0))$



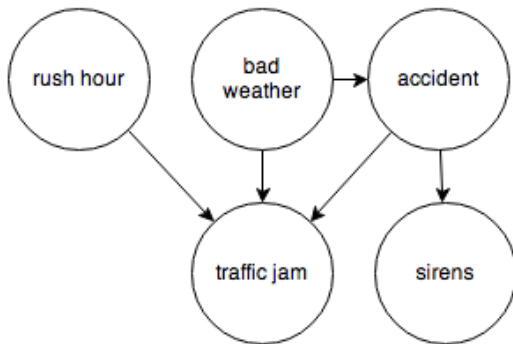
Regressão Descontínua

- Uso de uma descontinuidade pra captar o efeito causal
- Exemplo: Bolsa para estudantes que atingirem mais que 80% num teste
- Interrupção de time series é geralmente uma má ideia.



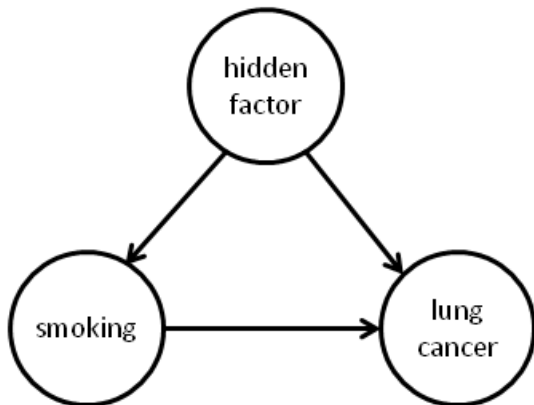
Contribuição de Pearl

- Judea Pearl desenvolveu um importante modelo de causalidade
- Estrutura física determina as relações de causalidade.
- Modelo de estrutura (SCM) precisa ser assumido antes da inferência.
- Representação de estrutura em forma de grafos.



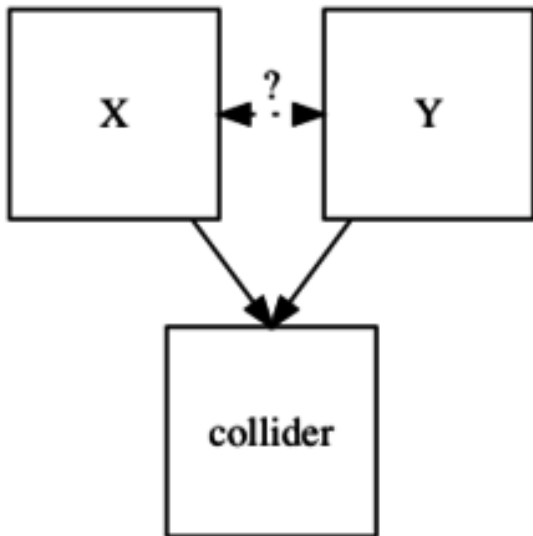
Tipos de Grafos - Confounding

Se X causa Y, Z é um confounder se Z causa X e Y ao mesmo tempo.



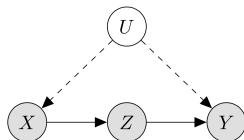
Tipos de Grafos - Collider

Se X causa Y, e ambos X e Y causam Z. Z é um collider.

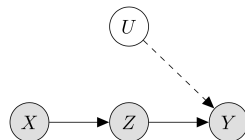


Intervenção

- Se queremos entender o efeito causal de X em Y , precisamos isolar outras variáveis que poderiam interferir na análise.
- Em outras palavras, estudar $P(Y|X)$ não é suficiente, precisamos de $P(Y|do(X))$.
- Experimentos controlados resolvem esse problema.
- Para dados observacionais, precisamos assumir um SCM e condicionar ou não em variáveis, respeitando alguns métodos:
 - Backdoor path
 - Front-door path



(a) Observational Model



(b) $do(X)$

É melhor condicionar em colliders. Isso melhora tanto a predição quanto a inferência causal.

```
z <- rnorm(1000)
```

```
x <- 0.6 + 1.3*z + rnorm(1000)
```

```
y <- 1.2 + 1.6 * x + 1.3*z + rnorm(1000)
```

```
df1 <- tibble(x = x, y = y, z = z)
```

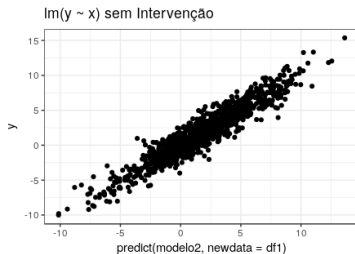
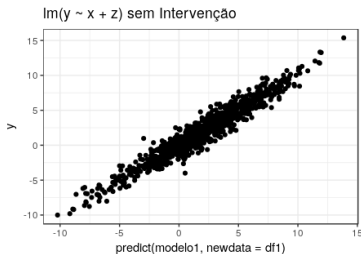
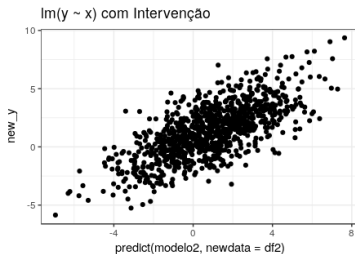
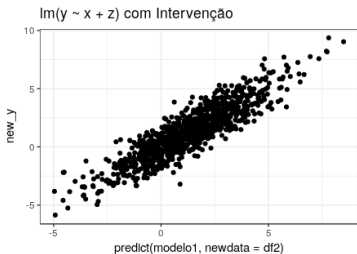
```
new_x <- rnorm(1000)
```

```
new_z <- rnorm(1000)
```

```
new_y <- 1.2 + 1.6 * new_x + 1.3*new_z + rnorm(1000)
```

```
df2 <- tibble(x = new_x, y = new_y, z = new_z)
```

Simulações - Predições x Causalidade - Confounding



É melhor condicionar sempre? M-Bias

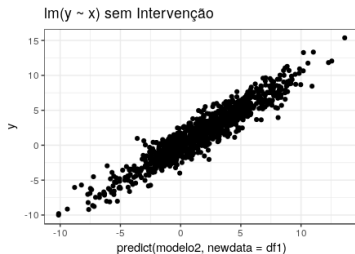
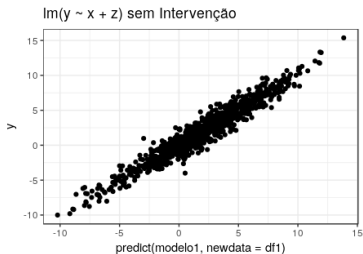
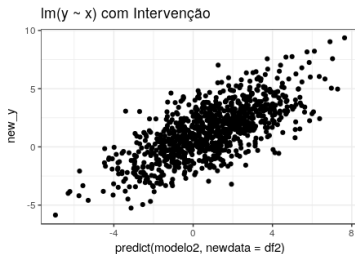
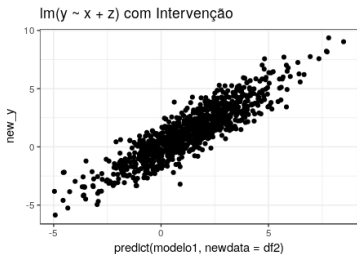
```
x1 <- rnorm(1000) y <- 1.2 + 1.6 * x1 + rnorm(1000) z <- 0.6 + 1.3*x1 + 1.5*y +  
rnorm(1000)
```

```
df1 <- tibble(x1 = x1, y = y, z = z)
```

```
new_x1 <- rnorm(1000) new_y <- 1.2 + 1.6 * new_x1 + rnorm(1000) new_z <-  
rnorm(1000)
```

```
df2 <- tibble(x1 = new_x1 , y = new_y, z = new_z)
```

Simulações - Predições x Causalidade - Colliders



- Vale a pena se preocupar com causalidade.
- É possível fazer inferência causal com dados observacionais, mas devemos assumir novos pressupostos.

Obrigado!

twitter: @guilhermejd1 email: guilherme.duarte@jota.info