

Trabalho Dados Reais - Introdução à Inferência Causal

José Luiz Padilha da Silva

Junho de 2025

Introdução

O objetivo deste trabalho é explorar o problema de não aderência ao tratamento em estudos experimentais. Por simplicidade, considere que o tratamento é binário. Em um estudo aleatorizado, as características relacionados com o desfecho são balanceadas entre os grupos de tratamento e controle. Assim, o efeito causal do tratamento pode ser estimado pela diferença entre as médias dos grupos. Contudo, é possível que não haja aderência ao tratamento, ou seja, que indivíduos do grupo controle recebam o tratamento e vice-versa. Nesse caso, a diferença entre as médias dos grupos não reflete o efeito causal do tratamento.

Seja T a variável indicadora do tratamento e Y o desfecho, ou variável resposta. Seja A a variável indicadora de aderência ao tratamento e H um vetor de variáveis que afetam o desfecho e podem estar relacionadas com a aderência ao tratamento. Seja $Y(t, a)$ a resposta potencial para Y quando fazemos $T = t$ e $A = a$. É assumido que $Y(t, a) = Y(a)$. Seja ainda $A(t)$ a resposta potencial para A quando fazemos $T = t$.

As seguintes estratégias de análise são usuais para estimação do efeito do tratamento:

1. *As-treated (AS)*, que compara $E(Y|A = 1)$ com $E(Y|A = 0)$, ou seja, compara os indivíduos que receberam o tratamento com os que não receberam, independentemente do grupo de tratamento ao qual foram alocados.
2. *Per-protocol (PP)*, que compara $E(Y|T = 1, Z = 1)$ com $E(Y|T = 0, Z = 1)$, em que $Z = 1$ se $A = T$. Ou seja, compara os indivíduos que receberam o tratamento e estavam alocados ao grupo de tratamento com os que não receberam o tratamento e estavam alocados ao grupo controle. Note que Z é uma variável observada pós aleatorização.
3. *Intention-to-treat (ITT)*, que compara $E(Y|T = 1)$ com $E(Y|T = 0)$. Ou seja, compara os indivíduos de acordo com o grupo a que foram aleatorizados, independentemente de terem aderido ou não ao tratamento. O ITT mede o efeito da intenção de tratar com $A = 1$ versus $A = 0$, não o efeito do tratamento de fato recebido. Devido à aleatorização e *consistência*, temos que ele é igual ao efeito causal $ITT = E[Y(1, A(1))] - E[Y(0, A(0))]$.
4. *Complier average causal effect (CACE)*. Seja $C = 1$ a indicadora de o indivíduo ser um *complier*, i.e., quando $A(t) = t$. O CACE é definido como o efeito médio do tratamento nos *compliers*, isto é,

$$\begin{aligned} CACE &= E[Y(1)|C = 1] - E[Y(0)|C = 1] \\ &= E[Y|T = 1, C = 1] - E[Y|T = 0, C = 1]. \end{aligned}$$

O CACE pode ser estimado via

$$\widehat{CACE} = \frac{\hat{E}(Y|T = 1) - \hat{E}(Y|T = 0)}{\hat{P}(A = 1|T = 1) - \hat{P}(A = 1|T = 0)},$$

com distribuição amostral estimada via bootstrap.

Questão 1

O arquivo de dados `dados.RData` contém dados simulados de um estudo aleatorizado com tratamento binário, em que a variável `T` indica o grupo de tratamento (1) ou controle (0), a variável `A` indica se o indivíduo aderiu ao tratamento (1) ou não (0), e a variável `Y` é o desfecho. As variáveis `escore` e `idade` são covariáveis que afetam o desfecho, e são balanceadas entre os grupos devido à aleatorização. A seguir, o cabeçalho dos dados.

```
library(tidyverse)
load("dados.Rdata")
head(dados)
```

```
##   T A      Y      escore idade
## 1 1 1 79.31438 0.7929343    31
## 2 1 1 97.08148 2.2774292    35
## 3 1 1 98.58202 3.0844412    35
## 4 1 1 92.69684 -0.3456977    40
## 5 1 0 72.04336 2.4291247    28
## 6 1 1 91.22757 2.5060559    31
```

```
dados %>% group_by(T) %>% summarise(n = n(), mean_e = mean(escore), mean_age = mean(idade))
```

```
## # A tibble: 2 x 4
##       T      n mean_e mean_age
##   <dbl> <int> <dbl>   <dbl>
## 1     0  1000    2.01    35.0
## 2     1  1000    1.97    35.1
```

Responda às seguintes questões.

1. Analise os dados usando a abordagem AS. Apresente o intervalo de confiança de 95%. O que você conclui sobre o efeito do tratamento?
2. Analise os dados usando a abordagem PP. Apresente o intervalo de confiança de 95%. O que você conclui sobre o efeito do tratamento?
3. Analise os dados usando a abordagem ITT. Apresente o intervalo de confiança de 95%. O que você conclui sobre o efeito do tratamento?
4. Apresente as estimativas do CACE e o intervalo de confiança de 95%. O que você conclui sobre o efeito do tratamento?
5. Estime o ATE via pareamento e apresente o intervalo de confiança de 95%. O que você conclui sobre o efeito do tratamento?
6. Por fim, compare as estimativas obtidas nos itens anteriores. Qual das análises você considera mais adequada para estimar o efeito causal do tratamento? Justifique sua resposta.

Avalie e justifique a necessidade de ajustar pelas covariáveis `escore` e `idade` nas análises.

O trabalho pode ser feito em duplas e deverá ser entregue pelo e-mail jlpadilha@ufpr.br até o dia 02/07/25.