Experimentos e Inferência Causal

Manoel Galdino

2024-04-16

Section 1

Experimentos

Introdução

- Um experimento é o desenho de pesquisa no qual a pesquisadora controla o mecanismo de atribuição do tratamento e controle
- Seja $p_i = P(T_i = 1)$. Então p_i é conhecido e controlado pela pesquisadora.
- Em contraposição, um estudo observacional é quando a pesquisadora não controla o mecanismo (natureza ou realidade social)
- Quando uma quantidade potencial (estimando) pode ser descrita em função da distribuição de dados observáveis, dizemos que o estimando é identificável. De outro modo, não identificado.
- Veremos porque experimentos produzem desenhos críveis de identificação causal
- Vamos supor experimentos ideais (sem attrition ou non-compliance)

Diferença de média observada e Viés de seleção

- Supondo SUTVA, Diferença Simples de Média pode ser decomposta em ATE + viés de seleção
- $\underbrace{\mathbb{E}[Y_i|T_i = 1] - \mathbb{E}[Y_i|T_i = 0]}_{\text{Simple Difference in Outcomes (SDO)}} = \mathbb{E}[Y_i^1|T_i = 1] - \mathbb{E}[Y_i^0|T_i = 0]$
- Podemos adicionar e subtrair os resultados contrafactuais para os tratados
- $= \mathbb{E}[Y_i^1|T_i = 1] - \mathbb{E}[Y_i^0|T_i = 1] + \mathbb{E}[Y_i^0|T_i = 1] - \mathbb{E}[Y_i^0|T_i = 0]$
- $= \underbrace{\mathbb{E}[Y_i^1 - Y_i^0|T_i = 1]}_{\text{ATT}} + \underbrace{\mathbb{E}[Y_i^0|T_i = 1] - \mathbb{E}[Y_i^0|T_i = 0]}_{\text{Viés de Seleção}}$

Experimentos aleatórios

- Mecanismo de atribuição de tratamento é probabilístico (Positividade): $0 < p_i < 1$.
- Unconfoundedness ou Permutabilidade (ou *assignment mechanism–ignorability*): $P(T_i = 1|y^1, y^0) = P(T_i)$.

O que é Permutabilidade (uncounfoudedness)?

- A distribuição dos resultados potenciais é independente do tratamento.
- $\mathbb{E}[Y^1|T = 1] = \mathbb{E}[Y^1|T = 0]$
- $\mathbb{E}[Y^0|T = 1] = \mathbb{E}[Y^0|T = 0]$
- Resultados potenciais são independentes do tratamento, dadas as covariáveis.
- Se a condição de tratamento fosse hipoteticamente trocada, os resultados esperados permaneceriam os mesmos.
- Isso significa que em um experimento com permutabilidade, não temos viés de seleção (Por quê?).

Distinção Entre Resultados Potenciais e Realizados

- **Permutabilidade de Resultados Potenciais**:
  - Estamos discutindo $Y^1 \perp\!\!\!\perp T$, não $Y \perp\!\!\!\perp T$.
  - Importante distinguir entre resultados potenciais (hipotéticos) e resultados observados (realizados).

Experimento remove o viés de seleção

- Independência entre tratamento e resultados potenciais implica que $\mathbb{E}[Y_i^0|T_i = 1] = \mathbb{E}[Y_i^0|T_i = 0] = \mathbb{E}[Y_i^0]$
- Portanto, o viés de seleção, dado por $\mathbb{E}[Y_i^0|T_i = 1] - \mathbb{E}[Y_i^0|T_i = 0]$, fica:
- $\mathbb{E}[Y_i^0|T_i = 1] - \mathbb{E}[Y_i^0|T_i = 0] = \mathbb{E}[Y_i^0] - \mathbb{E}[Y_i^0] = 0$
- Ou seja, SDO estima o ATE (via ATT).
- $\underbrace{\mathbb{E}[Y_i|T_i = 1] - \mathbb{E}[Y_i|T_i = 0]}_{\text{Simple Difference in Outcomes (SDO)}} = \underbrace{\mathbb{E}[Y_i^1 - Y_i^0|T_i = 1]}_{\text{ATT}} = ATE$

Restrição de Exclusão

- Formalmente, podemos separar a alocação do tratamento e o tratamento efetivamente recebido. Seja Z_i a alocação do tratamento e T_i o tratamento recebido.
- Então, a restrição de exclusão quer dizer que o que importa é o tratamento efetivamente recebido T_i, e não a variável que aloca o tratamento Z_i.
- Formalmente, isso quer dizer que: $Y_i^{1,z=1,T} = Y_i^{1,z=0,T} = Y_i^{1,T}$ e similarmente, $Y_i^{0,z=0,T} = Y_i^{0,z=0,T} = Y_i^{0,T}$
- Quando não ocorre isso? Se o mecanismo de atribuição do tratamento dispara outras causas
- Suponha que um experimento é sobre efeito de transferência de dinheiro em bem-estar
- Se ongs, sabendo do experimento, forem ajudar quem não tiver sido alocado para receber dinheiro

Restrição de Exclusão - erro de mensuração

- Se tiver erro de mensuração assimétrico
- Pesquisadores distintos entrevistam recipientes e não-recipientes da transferência de dinheiro, com habilidades distintas
- Ou questionários diferentes. Erro de mensuração assimétrico

Erro de mensuração

- Nova *switching equation*.
- Seja e_{i1} o erro de mensuração cometido se uma observação é atribuída para o tratamento, e, analogamente, e_{i0} o erro para o controle.
- De $Y_i = T_i Y_i^1 + (1 - T_i) Y_i^0$ para $Y_i = T_i(Y_i^1 + e_{i1}) + (1 - T_i))(Y_i^0 + e_{i0})$.
- Novo SDO: $\mathbb{E}[Y_i | T_i = 1] - \mathbb{E}[Y_i | T_i = 0] = \mathbb{E}[Y_i^1 + e_{i1} | T_i = 1] - \mathbb{E}[Y_i^0 + e_{i0} | T_i = 0] = \mathbb{E}[Y_i^1 | T_i = 1] + \mathbb{E}[e_{i1} | T_i = 1] - \mathbb{E}[Y_i^0 | T_i = 0] - \mathbb{E}[e_{i0} | T_i = 0]$
- Novo SDO pode se rearranjado: $\underbrace{\mathbb{E}[Y_i^1 | T_i = 1] - \mathbb{E}[Y_i^0 | T_i = 0]}_{\text{antigo SDO}} + \underbrace{\mathbb{E}[e_{i1} | T_i = 1] - \mathbb{E}[e_{i0} | T_i = 0]}_{\text{Dif média no erro de mensuração}}$
- Se $\mathbb{E}[e_{i1} | T_i = 1] \neq \mathbb{E}[e_{i0} | T_i = 0]$, então SDO será viesado.

Garantindo a Restrição de Exclusão

- Double blindness (duplo cego)
- Paralelismo na administração do experimento (mesmo questionário e mesmos entrevistadores)
- Na pior das hipóteses, aleatorização dos entrevistadores.

Section 2

Tipos de experimentos

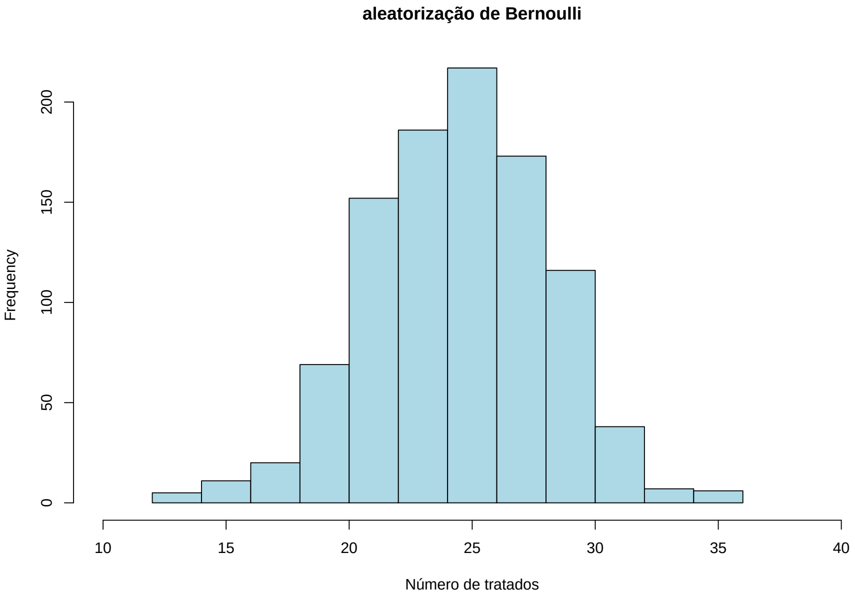
Aleatorização de Bernoulli

- É o experimento com aleatorização simples (basicamente, lançamento de moeda)
- Matematicamente, $p_i(T_i = 1) = p$.
- Problema: Possível “má aleratorização” (todo mundo no controle ou tratamento)
- ps.: toda aleatorização realizada é matematicamente equivalente.
- Possui 2^n configurações possíveis de alocação entre tratamento e controle

Aleatorização de Bernoulli - Sim no R - código

```r
set.seed(10)
n <- 50
hist(replicate(1000, sum(rbinom(n, 1, 0.5))),
     main = "aleatorização de Bernoulli",
     xlab = "Número de tratados",
     col = "lightblue") + xlim(0,50)
```

Aleatorização de Bernoulli - Sim no R - Histograma



Aleatorização Completa

- Seleciono aleatoriamente um número fixo de pessoas para tratamento e controle
- Ex.: 25 para tratamento e 25 para controle
- Basta numerar cada unidade (de 1 a 50) e amostrar 25 aleatoriamente para tratamento (e restante para controle)
- Vantagem: garanto número de obs em cada condição
- Possui $\binom{N}{\frac{n}{2}}$ configurações possíveis de alocação entre tratamento e controle.
- Intuição: estamos jogando fora as aleatorizações "indesejáveis".
- Cálculo da variância é mais complexo

Aleatorização Condicional (Block Random Assigment)

- **Definição**: Experimento é condicionalmente aleatório se a aleatorização depende de variáveis pré-tratamento X.
- **Exemplo Binário**: Duas moedas, uma para $X = 1$ e outra para $X = 0$.
- **Aleatorização Marginal vs. Condicional**:
  - Marginal: Aleatorização uniforme para todos os indivíduos.
  - Condicional: Aleatorização depende de X, gerando permutabilidade condicional a X.
- **Permutabilidade Condicional**: $(Y^1, Y^0 | X = x) \perp\!\!\!\perp T$.

Implicações da Aleatorização Condicional

- Não gera permutabilidade (não-condicional).
- Permutabilidade condicional a X é crucial para inferência em contextos com variáveis pré-tratamento.

Pensando aleatorização em bloco

- Ex.: digamos que em um amostra de 100 pessoas, queremos 25 homens e 25 mulheres no tratamento e controle
- Sorteio 25 homens para tramento e depois 25 mulheres.
- Cada bloco possui tamanho 25, neste exemplo.
- Blocos de tamanho 2 são chamados de *pair-matched design*.
- Em geral, estudos com *matching* em muitas variáveis
- Útil para amostras pequenas

ATE com Aleatorização Condicional (Bloco)

- Estratitificação
- Efeito heterogêneo por estrato?
- Podemos calcular o ATE por estrato, já que é aleatório no interior de cada estrato.
- Efeito geral na população.
- Podemos calcular ponderando os ATEs.
- Seja J o número de estratos, indixados por j. Seja N o número de unidades e N_j o número de unidades no bloco j. Então:
- $ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$

Aleatorização em bloco

– Pela Lei dos Grandes números, tende a gerar balanceamento entre blocos – Balanceamento quer dizer que blocos são similares – Em variáveis observadas e não-observadas – Probabilidade de tratamento pode variar por bloco. – Chamada de propensity score.

Precisão da aleatorização em bloco

- Em geral a precisão aumenta (erro padrão diminui) com aleatorização em bloco.
- Intuição é que removemos parte da variância (amostras possíveis), condicionando nos estratos
- Vamos checar uma simulação no R para ver um exemplo do ganho na precisão
- Lembrem-se que se X e Y são independentes, então $Var(aX + bY) = a^2 Var(x) + b^2 Var(Y)$.

Precisão da aleatorização em bloco - R sim p. 1

```r
# Set up Potential outcomes and units and blocks
n1 <- 10
n2 <- 16
N <- n1+n2
J <- 2
index_block <- c(rep(2, n2), rep(1, n1))
set.seed(12)
# potential outcome control
y0 <- c(rnorm(n1, 2, 1),rnorm(n2, 6, 1))
y1 <- y0 + 1.5 # potential outcome treatment
```

Precisão da aleatorização em bloco - R sim p. 3

```r
# block assignment
t_bloco1 <- sample(1:n1, n1/2)
c_bloco1 <- (1:n1)[!(1:n1 %in% t_bloco1)]

t_bloco2 <- sample((n1+1):(n1+n2), n2/2)
c_bloco2 <- ((n1+1):(n1+n2))[!((n1+1):(n1+n2) %in% t_bloco2)]

y1_obs_bloco1 <- y1[t_bloco1]
y1_obs_bloco2 <- y1[t_bloco2]
y0_obs_bloco1 <- y0[c_bloco1]
y0_obs_bloco2 <- y0[c_bloco2]
```

Precisão da aleatorização em bloco - R sim p. 2

```r
# random assignment
units_simple_treatment <- c(t_bloco1, t_bloco2)
units_simple_control <- c(c_bloco1, c_bloco2)
y1_obs <- y1[units_simple_treatment]
y0_obs <- y0[units_simple_control]

# erro padrão
erro_pad_simple <- t.test(y1_obs, y0_obs)$stderr
simple_p_value <- t.test(y1_obs, y0_obs)$p.value

my_t <- mean(y1_obs - y0_obs)/erro_pad_simple
```

Precisão da aleatorização em bloco - R sim p. 4

```
erro_pad1 <- t.test(y1_obs_bloco1, y0_obs_bloco1)$stderr
erro_pad2 <- t.test(y1_obs_bloco2, y0_obs_bloco2)$stderr

erro_padrao_geral <- sqrt(erro_pad1^2*(n1/N)^2 + erro_pad2^2*(
ate1 <- mean(y1_obs_bloco1 - y0_obs_bloco1)*(n1/N)
ate2 <- mean(y1_obs_bloco2 - y0_obs_bloco2)*(n2/N)
ate <- ate1 + ate2
my_t <- ate/erro_padrao_geral
p_value <- 2*(1 - pt(abs(my_t), df = 23.76567))
```

Comparação de SEs

Table 1: Comparação de Erros padrão

Method	Standard_Error
Simple Randomization	0.9552452
Block 1	0.6465630
Block 2	0.4502482
General Block Randomization	0.3723061

Cluster randomization

- Quando aleatorizo o cluster, em vez das unidades.
- Ex.: Se não for possível aleatorizar um tratamento entre estudantes, aleatorizo escolas
- No interior de cada escola, todo mundo é tratado ou não-tratado. Não há variação *within* escolas, apenas entre (between) escolas.
- Grande perda de variabilidade nos dados, reduzindo precisão (aumento no erro padrão)
- Às vezes é a única aleatorização possível.

Comparação

Table 2: Comparação de valoes p

Method	p_value
Simple Randomization	0.1413670
Block Randomization	0.0006818

Tabelas em artigos

SOCIAL ESTEEM AND PARTICIPATION IN CONTENTIOUS POLITICS

TABLE 1 Differences in Rates of Participation

	Intent to Participate	Actual Participation
Effect of newsletter treatment		
Newsletter	6.08%	3.04%
	N = 1217	N = 1217
Information-only	3.53%	1.72%
	N = 1217	N = 1217
Difference	**2.55**	**1.32**
(S.E.)	**(0.867)**	**(0.618)**
p-value (two-sided Welch)	**0.003**	**0.034**

Figure 1: Tabela de resultados.

Estimador ATE

- estimativa: $\frac{\sum_{i=1}^{n} Y_i T_i}{n_1} - \frac{\sum_{i=1}^{n} Y_i(1-T_i)}{n_0} = .0608 - .0353 = 0.0255$
- Erro padrão: $\sqrt{\frac{\hat{\sigma_1^2}}{n_1} + \frac{\hat{\sigma_0^2}}{n_0}} = \sqrt{\frac{.0608 \cdot (1-.0608)}{1217} + \frac{.0353 \cdot (1-.0353)}{1217}} = 0.00865$
- Pequena diferença com os coeficientes da tabela
- Typo? Alguma informação não reproduz exatamente? Fizemos algo errado?

Estimador ATE - R

```r
treatment <- c(rep(1, 74), rep(0, 1217 - 74))
control <- c(rep(1, 43), rep(0, 1217 - 43))
var_treat <- var(treatment)
var_control <- var(control)
erro_padrao <- sqrt(var_treat/1217 + var_control/1217)
round(erro_padrao, 5)

t.test(treatment, control)
```

Estimador ATE - R print

```
## [1] 0.00866

##
##  Welch Two Sample t-test
##
## data:  treatment and control
## t = 2.9414, df = 2286.3, p-value = 0.0033
## alternative hypothesis: true difference in means is not equ
## 95 percent confidence interval:
##  0.008490408 0.042454539
## sample estimates:
##  mean of x  mean of y
## 0.06080526 0.03533279
```

Key Takeaways

- Experimento (sob SUTVA) elimina o viés de seleção
- Depende de restrição de exclusão e simetria
- Vários tipos de experimentos: block aumenta precisão
- Com N grande, diferença diminui
- Sempre supomos condições ideais (sem attrition, compliance perfeito etc.)