

TRABALHO 5 - Inferência Causal

Vitor Kroeff e Helen Lourenço

02/07/2025

1 Questão 1

A abordagem **AS** compara o efeito entre os indivíduos que de fato aderiram ao tratamento, independentemente da alocação inicial. No entanto, essa abordagem pode fornecer uma estimativa viesada, pois ignora a alocação original dos participantes e o processo de aleatorização. Neste caso, observamos uma estimativa do efeito do tratamento de 15,88, com um intervalo de confiança de (15,10; 16,72).

```
# Pacotes
pacman::p_load(dplyr, boot, Matching)

# Carregand os dados
load(
  "C:/Users/vitor/OneDrive/R/Estatistica-UFPR/Inferência Causal/Trabalho 5/dados.RData"
)

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
```

```
# Estimativa
mean(dados$Y[dados$A == 1]) - mean(dados$Y[dados$A == 0])
```

```
## [1] 15.88422
```

```
# Intervalo de Confiança
```

```
AS <- function(df, linha) {
  data <- df[linha, ]
  mean(data$Y[data$A == 1]) - mean(data$Y[data$A == 0])
}
```

```
set.seed(1903)
```

```

as_boot <- boot(dados, AS, R = 2000)

ic_as <- boot.ci(as_boot, conf = 0.95, type = 'perc')

# Estativa Bootstrap
as_boot$t %>% mean()

## [1] 15.88355

# Intervalo de confinaça
ic_as

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 2000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = as_boot, conf = 0.95, type = "perc")
##
## Intervals :
## Level      Percentile
## 95%      (15.10, 16.72 )
## Calculations and Intervals on Original Scale

```

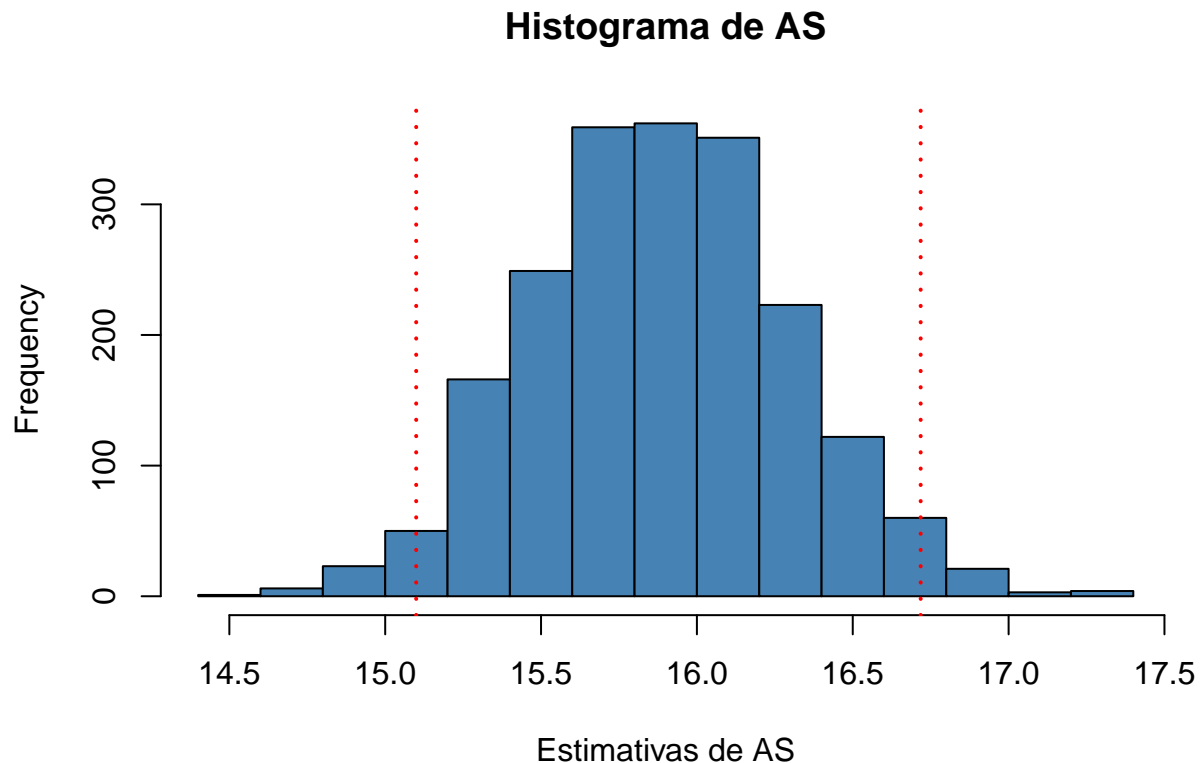
Podemos observar no histogram que as estimativas variam de 14,5 a 17,5.

```

# Visualização
hist(as_boot$t,
     col = 'steelblue',
     xlab = 'Estimativas de AS',
     main = 'Histograma de AS')

# Adiciona linhas do IC
abline(
  v = ic_as$perc[4:5],
  col = 'red',
  lwd = 2,
  lty = 3
) # limites inferior e superior

```



Questão 2

O **PP**, por outro lado, calcula o efeito do tratamento apenas entre os indivíduos que aderiram ao tratamento ao qual foram designados. Esse efeito também pode ser viesado, pois a adesão ao tratamento pode estar associada as outras variáveis, introduzindo um possível viés de seleção.

Ao observar as estimativas do efeito do tratamento, vemos um valor menor, de 14,38, com um intervalo de confiança de (13,34; 15,39). Esse intervalo contém a estimativa obtida pela abordagem AS, o que indica que os resultados são relativamente próximos.

```
# PP

# Estimativa

### Filtra A=T
pp_dados <- dados %>% filter(A==T)

# Estimativa
mean(pp_dados$Y[pp_dados$T == 1]) - mean(pp_dados$Y[pp_dados$T == 0])

## [1] 14.38197

# Intervalo de confiança
PP <- function(df, linha) {
  data <- df[linha, ]
  mean(data$Y[data$T == 1]) - mean(data$Y[data$T == 0])
}
```

```

}

set.seed(1903)
pp_boot <- boot(pp_dados, PP, R = 2000)
ic_pp <- boot.ci(pp_boot, conf = 0.95, type = 'perc')

# Estimativa Bootstrap
pp_boot$t %>% mean()

```

```
## [1] 14.37827
```

```

# Intervalo de Confiança
ic_pp

```

```

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 2000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = pp_boot, conf = 0.95, type = "perc")
##
## Intervals :
## Level      Percentile
## 95%      (13.34, 15.39 )
## Calculations and Intervals on Original Scale

```

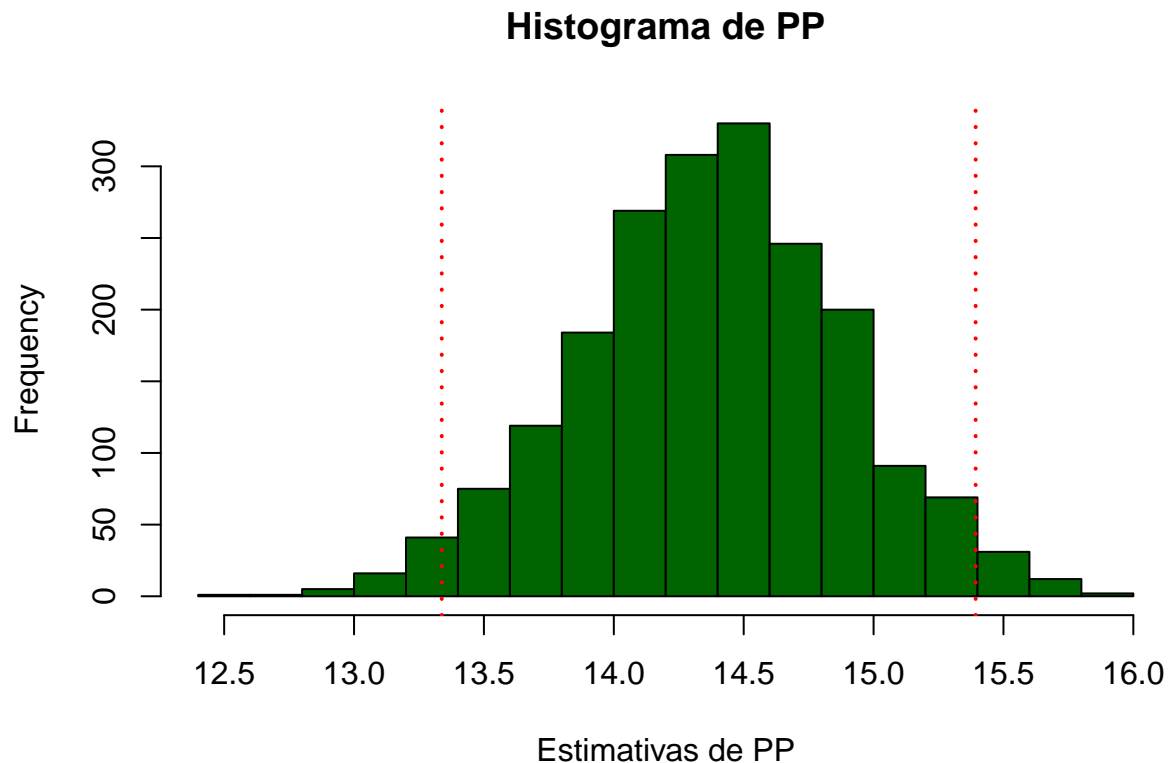
Observando o gráfico, vemos uma variação muito próxima a do método anterior.

```

# Visualização
hist(pp_boot$t,
      col = 'darkgreen',
      xlab = 'Estimativas de PP',
      main = 'Histograma de PP')

# Adiciona linhas do IC
abline(
  v = ic_pp$perc[4:5],
  col = 'red',
  lwd = 2,
  lty = 3
) # limites inferior e superior

```



Questão 3

O **ITT** compara os desfechos com base na alocação aleatória dos indivíduos. Devido à aleatorização do estudo, essa abordagem pode fornecer uma estimativa não viesada do efeito do tratamento.

Observamos que o efeito estimado por esse método é de 4,04, com um intervalo de confiança de (3,003; 5,079). Esse valor difere consideravelmente das estimativas anteriores, o que pode indicar um possível problema de adesão dos indivíduos ao tratamento proposto no estudo.

```
## ITT

# Estimativa
mean(dados$Y[dados$T == 1]) - mean(dados$Y[dados$T == 0])
```

```
## [1] 4.045822
```

```
# Intervalo de Confiança

ITT <- function(df, linha) {
  data <- df[linha, ]
  mean(data$Y[data$T == 1]) - mean(data$Y[data$T == 0])
}

set.seed(1903)
itt_boot <- boot(dados, ITT, R = 2000)

ic_itt <- boot.ci(itt_boot, conf = 0.95, type = 'perc')
```

```
# Estimativa Bootstrap
itt_boot$t %>% mean()
```

```
## [1] 4.049068
```

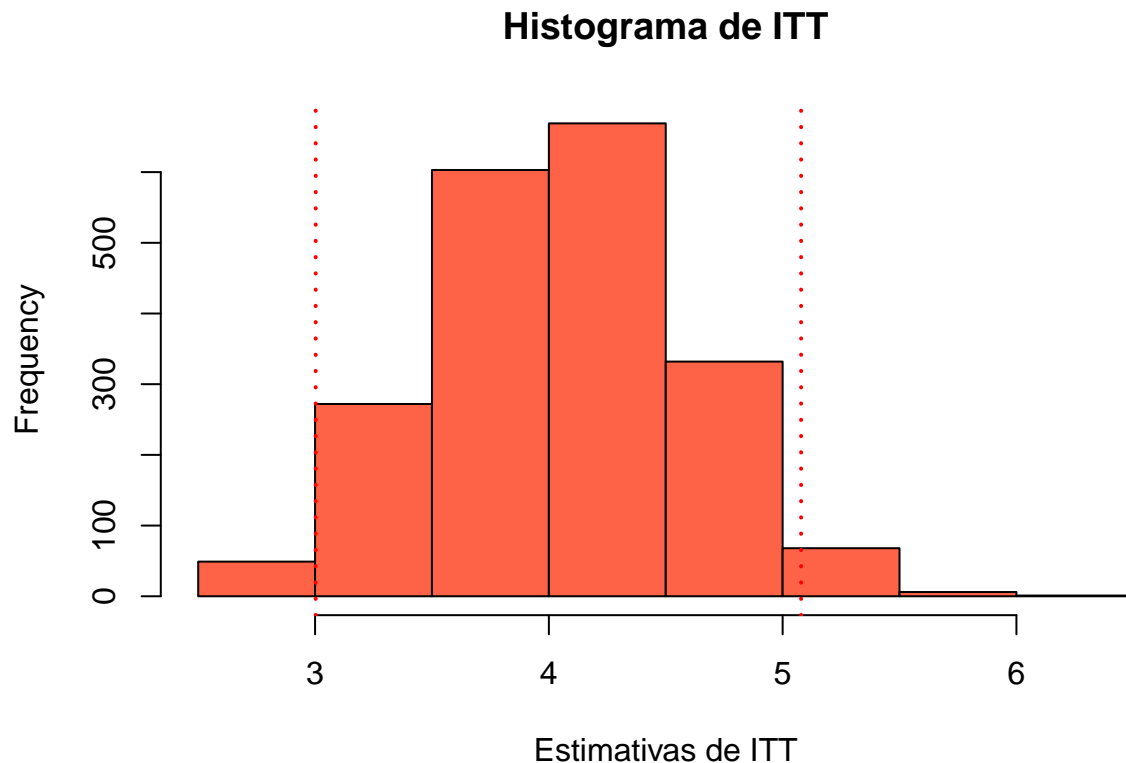
```
# Intervalo de Confiança
ic_itt
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 2000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = itt_boot, conf = 0.95, type = "perc")
##
## Intervals :
## Level      Percentile
## 95%      ( 3.003,  5.079 )
## Calculations and Intervals on Original Scale
```

Observando o gráfico, vemos que os valores estão menos dispersos e mais concentrados em torno da média dos dados.

```
# Visualização
hist(itt_boot$t,
     col = 'tomato',
     xlab = 'Estimativas de ITT',
     main = 'Histograma de ITT')

# Adiciona linhas do IC
abline(
  v = ic_itt$perc[4:5],
  col = 'red',
  lwd = 2,
  lty = 3
) # limites inferior e superior
```



Abaixo, podemos observar a distribuição da aderência ao tratamento em ambos os grupos.

```
tab <- round(prop.table(table(dados$A, dados$T), margin = 2), 2)

# Imprimir a tabela com nomes claros
colnames(tab) <- c("T = 0", "T = 1")
rownames(tab) <- c("A = 0", "A = 1")
tab
```

```
##
##      T = 0 T = 1
## A = 0  0.64 0.26
## A = 1  0.36 0.74
```

Questão 4

O **CACE** estima o efeito causal médio do tratamento especificamente na subpopulação de compliers, ou seja, aqueles indivíduos que adeririam ao tratamento se alocados ao grupo de tratamento e não o fariam se alocados ao grupo controle.

Observa-se que as estimativas são consideravelmente menores em relação às obtidas pelos dois primeiros métodos, mas maiores do que a estimativa do ITT, com um efeito de 10,56 e um intervalo de confiança associado de (8,32; 12,67).

```
# Estimativa

(mean(dados$Y[dados$T == 1]) - mean(dados$Y[dados$T == 0])) /
  (mean(dados$A[dados$T == 1] == 1) - mean(dados$A[dados$T == 0] == 1))
```

```
## [1] 10.5635
```

```
CACE <- function(df, linha) {  
  data <- df[linha, ]  
  (mean(data$Y[data$T == 1]) - mean(data$Y[data$T == 0])) / #  $E(Y/T = 1) - E(Y/T = 0)$   
  (mean(data$A[data$T == 1] == 1) - mean(data$A[data$T == 0] == 1)) #  $P(A = 1 / T = 1) - E(A = 1/T =$   
}  
  
set.seed(1903)  
cace_boot <- boot(dados, CACE, R = 2000)  
  
ic_cace <- boot.ci(cace_boot, conf = 0.95, type = 'perc')  
  
# Estimativa Bootstrap  
cace_boot$t %>% mean()
```

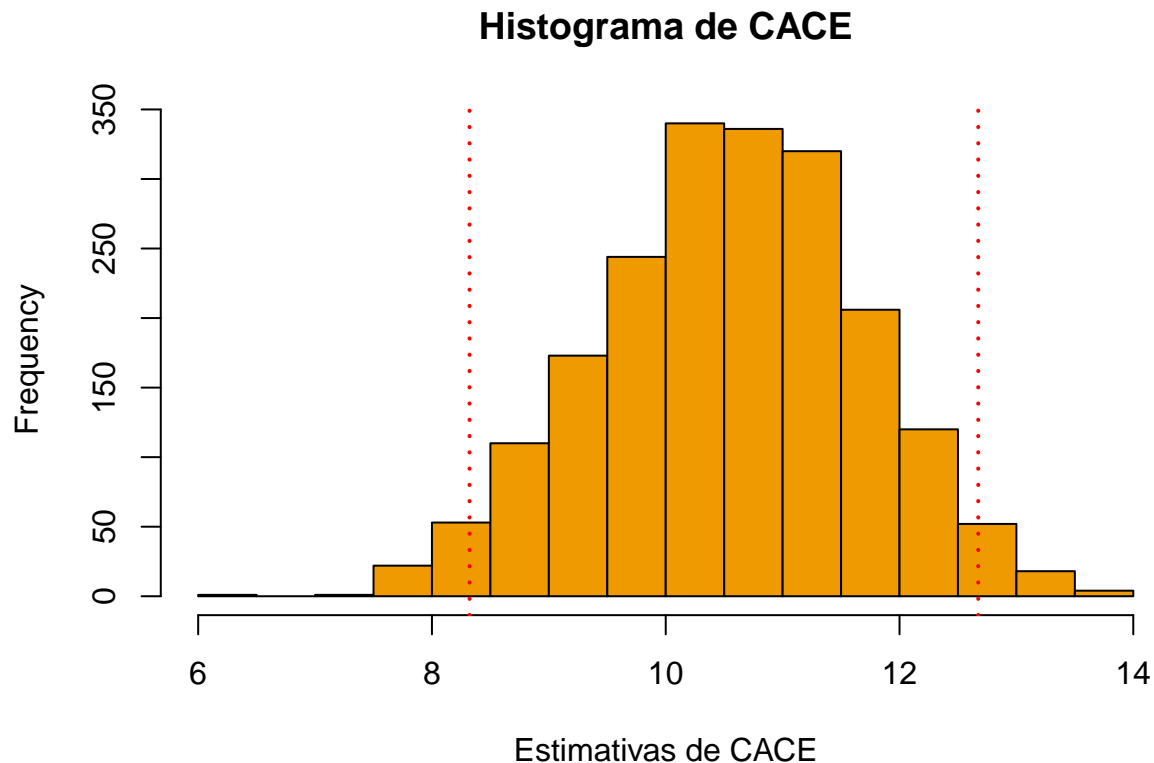
```
## [1] 10.55315
```

```
# Intervalo de Confiança  
ic_cace
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS  
## Based on 2000 bootstrap replicates  
##  
## CALL :  
## boot.ci(boot.out = cace_boot, conf = 0.95, type = "perc")  
##  
## Intervals :  
## Level      Percentile  
## 95%      ( 8.32, 12.67 )  
## Calculations and Intervals on Original Scale
```

No histograma, observamos um comportamento muito semelhante ao dos métodos anteriores.

```
# Visualização  
hist(cace_boot$t,  
     col = 'orange2',  
     xlab = 'Estimativas de CACE',  
     main = 'Histograma de CACE')  
  
# Adiciona linhas do IC  
abline(  
  v = ic_cace$perc[4:5],  
  col = 'red',  
  lwd = 2,  
  lty = 3  
) # limites inferior e superior
```

Questão 5

A estimação via **ATE** com pareamento busca incluir variáveis de controle na estimativa do efeito, equilibrando os grupos de tratamento e controle, a fim de estimar o efeito médio do tratamento. Devido à aleatorização do estudo, observamos um valor muito próximo ao estimado pelo ITT, o que reforça a indicação de que a não adesão ao tratamento pode ser um fator problemático na análise realizada pelos outros métodos.

```
Y <- dados$Y
Tr <- dados$T
X <- dados %>% dplyr::select(c(escore, idade))

set.seed(1903)

# utilizando pareamento
## Estimativa
rr <- Matching::Match(Y = Y, Tr = Tr, X = X, M = 1, estimand = "ATE")
rr$est

##           [,1]
## [1,] 3.853112

matching <- function(df, linha) {
  data <- df[linha, ]
  Y <- data$Y
  Tr <- data$T
  X <- data %>% dplyr::select(c(escore, idade))
  rr <- Matching::Match(Y = Y, Tr = Tr, X = X, M = 1, estimand = "ATE", ties = F)
```

```

    return(rr$est)
}

set.seed(1903)
matching_boot <- boot(dados, matching, R = 1000)

ic_matching <- boot.ci(cace_boot, conf = 0.95, type = 'perc')

# Estimativa Bootstrap
matching_boot$t %>% mean()

```

```
## [1] 3.834547
```

```

# Intervalo de Confiança
ic_matching

```

```

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 2000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = cace_boot, conf = 0.95, type = "perc")
##
## Intervals :
## Level      Percentile
## 95%      ( 8.32, 12.67 )
## Calculations and Intervals on Original Scale

```

Também observamos um histograma muito proximo ao do ATT.

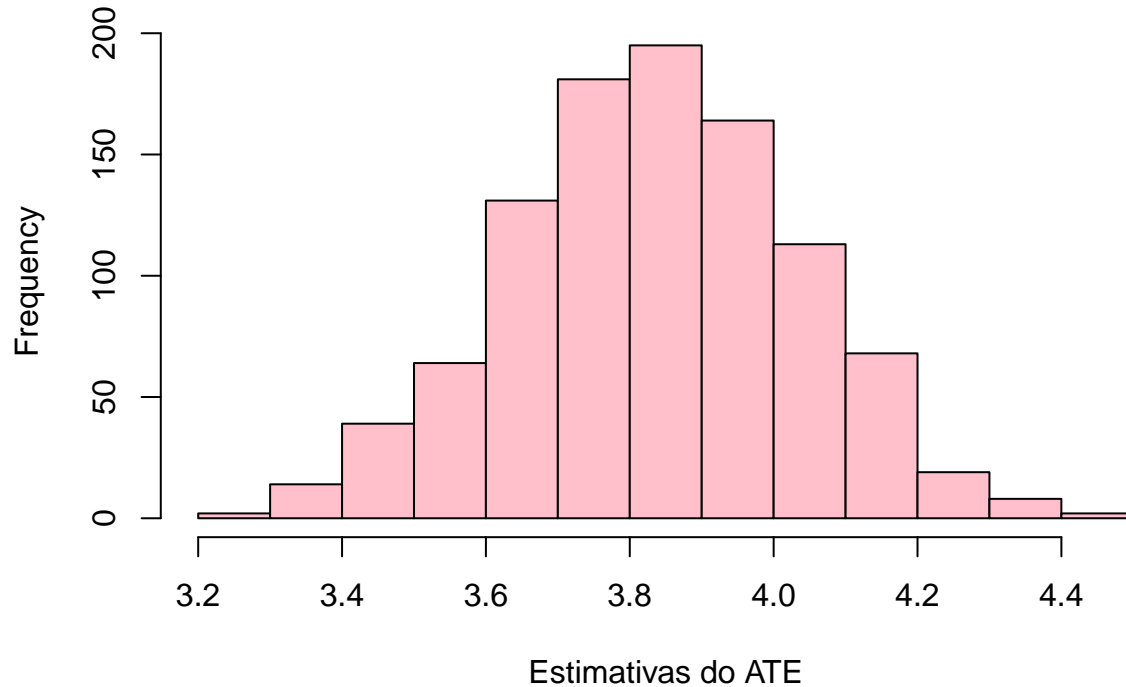
```

# Visualização
hist(matching_boot$t,
     col = 'pink',
     xlab = 'Estimativas do ATE',
     main = 'Histograma do ATE via Pareamento')

# Adiciona linhas do IC
abline(
  v = ic_matching$perc[4:5],
  col = 'red',
  lwd = 2,
  lty = 3
) # limites inferior e superior

```

Histograma do ATE via Pareamento



Questão 6

De forma geral, cada método tem um objetivo distinto ao estimar o efeito do tratamento e ao definir a população-alvo da análise. Como o estudo utiliza aleatorização dos participantes, o método **ITT** (Intention-to-Treat) é o mais indicado, pois preserva os benefícios da aleatorização e fornece uma estimativa não viesada do efeito causal do tratamento nos indivíduos. Embora o ATE com pareamento também possa ser utilizado, seu uso é mais comum em estudos observacionais, onde a aleatorização não está presente. Assim, neste caso, o ITT oferece uma estimativa mais apropriada do efeito causal.

```
res <- data.frame(Metodo = c('AS', 'PP', 'ITT', 'CACE', 'ATE (Pareamento)'),
  Estimativa_boot = c(itt_boot$t %>% mean(), pp_boot$t %>% mean(),
    itt_boot$t %>% mean(), cace_boot$t %>% mean(),
    matching_boot$t %>% mean()))
res$Estimativa_boot <- round(res$Estimativa_boot, 3)

knitr::kable(res, caption = 'Estimativas dos Métodos via Bootstrap')
```

Table 1: Estimativas dos Métodos via Bootstrap

Metodo	Estimativa_boot
AS	4.049
PP	14.378
ITT	4.049
CACE	10.553
ATE (Pareamento)	3.835

REFERÊNCIAS

McCoy CE. Understanding the Intention-to-treat Principle in Randomized Controlled Trials. *West J Emerg Med*. 2017 Oct;18(6):1075-1078. doi: 10.5811/westjem.2017.8.35985. Epub 2017 Sep 18. PMID: 29085540; PMCID: PMC5654877.

Smith VA, Coffman CJ, Hudgens MG. Interpreting the Results of Intention-to-Treat, Per-Protocol, and As-Treated Analyses of Clinical Trials. *JAMA*. 2021 Aug 3;326(5):433-434. doi: 10.1001/jama.2021.2825. PMID: 34342631; PMCID: PMC8985703.