yc4384_P8122_HW3

yc4384_Yangyang_Chen

2024-10-07

Preparing mice data

```
set.seed(124)
n = 16
p_C = 1/5
C = rbinom(n, 1, p_C)
theta0 = 1/2
theta1 = -1/5
p_A = theta0 + theta1 * C
A = rbinom(n, 1, p_A)
beta0 = 110
beta1 = 20
beta2 = 5
sigma_Y = 1
mu_Y = beta0 + beta1 * C + beta2 * A
Y = rnorm(n, mu_Y, sigma_Y)

mice_data = data.frame(
  mouse_id = 1:n,
  obesity = C,
  light_exposure = A,
  glucose = Y
)

head(mice_data)
```

```
##   mouse_id obesity light_exposure  glucose
## 1        1       0              1 114.8490
## 2        2       0              1 113.7769
## 3        3       0              1 114.1312
## 4        4       0              0 108.9575
## 5        5       0              1 113.8964
## 6        6       0              1 115.4442
```

Problem 1: Interpretation of parameters:

p: The baseline covariate C (obesity) follows Bernoulli distribution, where p represents the probability of $C = 1$ (with obesity).

θ_0: The probability of assigning the units to treatment group (light) if the units are not with obesity ($C = 0$).

θ_1: The probability of assigning the units with obesity to treatment group (light) is 1/5 times lower than the units without obesity on average.

β_0: The mean of baseline glucose when the mice are non-obese and unexposed to light.

β_1: Intervening to increase C (obesity) by one unit will, on average, increase the outcome Y (glucose) by 20 units, holding other coviariates constants.

β_2: Intervening to increase A (light) by one unit will, on average, decrease the outcome Y (glucose) by 5 units, holding other coviariates constants.

Problem 2: Marginal and Conditional PACE

Since $P(C = 1) = \frac{1}{5}$

and $E[Y \mid A, C] = \beta_0 + \beta_1 C + \beta_2 A = 110 + 20C - 5A$

The Marginal PACE is:

$E[Y_1] - E[Y_0]$

$= \sum E[Y \mid A = 1, C = c] \cdot P(C = c) - \sum E[Y \mid A = 0, C = c] \cdot P(C = c)$

$= E[Y \mid A = 1, C = 1] \cdot P(C = 1) + E[Y \mid A = 1, C = 0] \cdot (1 - P(C = 1))$

$-[E[Y \mid A = 0, C = 1] \cdot P(C = 1) + E[Y \mid A = 0, C = 0] \cdot (1 - P(C = 1))]$

$= (\beta_0 + \beta_1 + \beta_2) \cdot P(C = 1) + (\beta_0 + \beta_2) \cdot (1 - P(C = 1)) - [(\beta_0 + \beta_1) \cdot P(C = 1) + \beta_0 \cdot (1 - P(C = 1))]$

$= (\beta_0 + \beta_1 + \beta_2) \cdot P(C = 1) + (\beta_0 + \beta_2) \cdot (1 - P(C = 1)) - [(\beta_0 + \beta_1) \cdot P(C = 1) + \beta_0 \cdot (1 - P(C = 1))]$

$= \beta_2$

$= -5$

Similarly, the conditional PACE is:

$E[Y_1 \mid C = c] - E[Y_0 \mid C = c]$

$= E[Y \mid A = 1, C = c] - E[Y \mid A = 0, C = c]$

$= (\beta_0 + \beta_1 \cdot C + \beta_2) - (\beta_0 + \beta_1 \cdot C)$

$= \beta_2$

$= -5$

I. Under the following assumptions Marginal PACE can be identified:

1) Consistency:
$$Y = Y_1 \text{ if treated, and } Y = Y_0 \text{ if not treated.}$$
The observed outcome is equal to the potential outcome for the treatment received.

2) Exchangeability:
$$Y_1, Y_0 \perp A \mid X.$$
Conditional on covariates X, treatment assignment is independent of the potential outcomes.

3) Positivity:
$$0 < P(A = 1|X) < 1.$$
There must be a positive probability of receiving each treatment level for all values of covariates.

II. Under the following assumptions Conditional PACE can be identified:

1) Conditional Consistency:

$$Y = Y_1 \text{ if } A = 1, \quad Y = Y_0 \text{ if } A = 0, \quad \text{given } X.$$

Similar to marginal consistency, but conditional on X.

2) Conditional Exchangeability:

$$Y_1, Y_0 \perp A \mid X = x.$$

The treatment assignment is independent of the potential outcomes, conditional on covariates X.

3) Conditional Positivity:

$$0 < P(A = 1 \mid X = x) < 1.$$

Within each level of X, there must be a positive probability of receiving both the treatment and control.

Problem 3: G-formula for Randomized and Observational Studies

I. In a randomized study, the treatment assignment $A \perp C$. Therefore:

$E[Y_a]$

$= \sum E[Y \mid A = a, C = c] \cdot P(C = c)$

$= E[Y \mid A = a]$

Thus, $E[Y_1] = E[Y \mid A = 1]$

and $E[Y_0] = E[Y \mid A = 0]$

II. In an observational study, treatment assignment A is not independent of the covariates C. As a result, we need to adjust for the distribution of covariates:

$E[Y_a]$

$= \sum E[Y \mid A = a, C = c] \cdot P(C = c)$

where

$E[Y_1]$

$= \sum E[Y \mid A = 1, C = c] \cdot P(C = c)$

$= E[Y \mid A = 1, C = 1] \cdot P(C = 1) + E[Y \mid A = 1, C = 0] \cdot (1 - P(C = 1))$

and

$E[Y(0)]$

$= \sum E[Y \mid A = 0, C = c] \cdot P(C = c)$

$E[Y \mid A = 0, C = 1] \cdot P(C = 1) + E[Y \mid A = 0, C = 0] \cdot (1 - P(C = 1))$

Thus, the key difference between the two studies is how the treatment assignment interacts with the covariates, which affects the application of the g-formula.