# P8122-Homework3-yz4184

Yunlin Zhou

2022-10-09

```
# data import
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)
```

## Question a

p: The baseline covariate C (obesity) follows Bernoulli distribution. p is the probability of C taking 1 (with obesity).

theta0: The probability of assigning the units to treatment group (light) if the units are not with obesity (C = 0).

theta1: The probability of assigning the units with obesity to treatment group (light) is 1/5 times lower than the units without obesity on average.

beta0: The mean of baseline glucose when the mice are non-obese and unexposed to light.

beta1: Intervening to increase C (obesity) by one unit will, on average, increase the outcome Y (glucose) by 20 units, holding other conditions to be the same.

beta2: Intervening to increase A (light) by one unit will, on average, decrease the outcome Y (glucose) by 5 units, holding other conditions to be the same.

## Question b

$Pr(C=1) = \frac{1}{5}$

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

Marginal PACE:

$$
\begin{aligned}
E[Y_1] - E[Y_0] &= \sum_c E[Y|A=1, C=c] \cdot Pr(C=c) - \sum_c E[Y|A=0, C=c] \cdot Pr(C=c) \\
&= E[Y|A=1, C=1] \cdot Pr(C=1) + E[Y|A=1, C=0] \cdot (1 - Pr(C=1)) - \\
&\quad [\, E[Y|A=0, C=1] \cdot Pr(C=1) + E[Y|A=0, C=0] \cdot (1 - Pr(C=1)) \,] \\
&= (\beta_0 + \beta_1 + \beta_2) \cdot Pr(C=1) + (\beta_0 + \beta_2)[1 - Pr(C=1)] - [(\beta_0 + \beta_1) \cdot Pr(C=1) + \beta_0[1 - Pr(C=1)]] \\
&= \beta_2 \\
&= -5
\end{aligned}
$$

Conditional PACE

$$
\begin{aligned}
E[Y_1 | C=c] - E[Y_2 | C=c] &= E[Y|A=1, C=c] - E[Y|A=0, C=c] \\
&= (\beta_0 + \beta_1 \cdot C + \beta_2) - (\beta_0 + \beta_1 \cdot C) \\
&= \beta_2 \\
&= -5
\end{aligned}
$$

**Marginal PACE**

We can identify marginal PACE when there is no backdoor path (randomized experiments).

The marginal PACE is counterfactual. We assume the expectation if the whole population is assigned to treatment or control group.

**Conditional PACE**

We can identify conditional PACE when there is no unobserved confounding (for observational experiment).

The conditional PACE is conditional on specific level of covariate. It's not across the population.

## Question c

For randomized study

$$E(Y_a) = \sum_c E(Y \mid A=a, C=c) \cdot Pr(C=c)$$
$$= E(Y \mid A=a)$$
$$E(Y_1) = E(Y \mid A=1)$$
$$E(Y_0) = E(Y \mid A=0)$$

For observational study

$$E(Y_a) = \sum_c E(Y \mid A=a, C=c) Pr(C=c)$$
$$E(Y_1) = \sum_c E(Y \mid A=1, C=c) Pr(C=c)$$
$$= E(Y \mid A=1, C=1) Pr(C=1) + E(Y \mid A=1, C=0) \cdot (1 - Pr(C=1))$$
$$E(Y_0) = \sum_c E(Y \mid A=0, C=c) Pr(C=c)$$
$$= E(Y \mid A=0, C=1) Pr(C=1) + E(Y \mid A=0, C=0) \cdot (1 - Pr(C=1))$$

For the *randomized study*, we assume that there are no confounders thus we can conclude the g-formula as above (without covariate C).

For the *observational study*, we assume that there is a confounder which will affect the assignment, thus we need to calculate the expectation conditional on the covariate C.

## Question d

```
mice_df = cbind(Y, A, C)%>%
  as.data.frame()
```

```
EYA1_1=mean(mice_df[which(mice_df$A == 1 & C == 1), ]$Y)
EYA1_1
```

```
## [1] NaN
```

```
EYA1_0=mean(mice_df[which(mice_df$A == 1 & C == 0), ]$Y)
EYA1_0
```

```
## [1] 114.5484
```

```
EYA0_1=mean(mice_df[which(mice_df$A == 0 & C == 1), ]$Y)
EYA0_1
```

```
## [1] 129.4415
```

```
EYA0_0=mean(mice_df[which(mice_df$A == 0 & C == 0), ]$Y)
EYA0_0
```

```
## [1] 109.7194
```

d.

$$E[Y|A=1] = \sum_c E(Y|A=1, C=c)\, Pr(C=c|A=1)$$
$$= E[Y|A=1, C=1)\, Pr(C=1|A=1) + E[Y|A=1, C=0)\, Pr(C=0|A=1)$$
$$= 0 + 114.5484 \quad = 114.5484$$

$$E[Y|A=0] = \sum_c E(Y|A=0, C=c)\, Pr(C=c|A=0)$$
$$= E[Y|A=0, C=1)\, Pr(C=1|A=0) + E[Y|A=0, C=0)\, Pr(C=0|A=0)$$
$$= 129.4445 \times \tfrac{3}{8} \quad + 109.7194 \times \tfrac{5}{8}$$
$$= 117.1151875$$

$$E(Y|A=1] - E(Y|A=0) = 114.5484 - 117.1151875 = -2.5667$$

e.

$$E[Y_1] - E(Y_2) = E(Y|A=1, C=1)\, Pr(C=1) + E(Y|A=1, C=0)\cdot(1-Pr(C=1))$$
$$- E(Y|A=0, C=1)\, Pr(C=1) + E(Y|A=0, C=0)\cdot(1-Pr(C=1))$$
$$= 114.5484 \times \tfrac{13}{16} - 109.7194 \times \tfrac{13}{16} - 129.4415 \times \tfrac{3}{16}$$
$$= -20.34672$$

In this question, we consider about the crude mean. This formula shows the association other than causal effect.

On average, the the treatment group will have a value of outcome Y which is 2.5667 units lower.

## Question e

d.

$E[Y|A=1] = \Sigma_c E(Y|A=1, C=c) \Pr(C=c|A=1)$

$= E[Y|A=1, C=1) \Pr(C=1|A=1) + E[Y|A=1, C=0) \Pr(C=0 \ |A=1)$

$= 0 + 114.5484 \quad = 114.5484$

$E[Y|A=0] = \Sigma_c E(Y|A=0, C=c) \Pr(C=c|A=0)$

$= E[Y|A=0, C=1) \Pr(C=1|A=0) + E[Y|A=0, C=0) \Pr(C=0 \ |A=0)$

$= 129.445 \times \frac{3}{8} \quad + 109.1194 \times \frac{5}{8}$

$= 117.1151875$

$E(Y|A=1] - E(Y|A=0) = 114.5484 - 117.1151875 = -2.5667$

e.

$E[Y_1] - E(Y_2) = E(Y|A=1, C=1) \Pr(C=1) + E(Y|A=1, C=0) \cdot (1-\Pr(C=1))$

$- E(Y|A=0, C=1) \Pr(C=1) + E(Y|A=0, C=0) \cdot (1-\Pr(C=1))$

$= 114.5484 \times \frac{13}{16} - 109.1194 \times \frac{13}{16} - 129.445 \times \frac{3}{16}$

$= -20.34672$

In this question, we calculate the marginal ACE using g-formula. In this situation, covariate C is included and this formula shows the causal effect.

If we assign the population to the treatment group (A=1) and the control group (A=0) conditional on the covariate C, then, on average, the the treatment group will have a value of outcome Y which is 20.34672 units lower.

**The differences between the inferences obtained in (d) and (e):**

- In question d, the covariate C is not included and we assume a randomized assignment. But in question e, the covariate C is included and we assume an assignment based on covariate.
- The formula in question d tells us about the association, but the formula in question e tells us about the casual effect.

# Question f

- The linear regression to be correctly specified.
- All confounders are included in the new covariates.