hmrwk 11 - Beimnet Taye

2023-04-15

P1

- DGP A: 1,2,3,4
- DGP B: 1,3
- DGP C: 2,4
- DGP D: 3
- DGP E: None

P2

1.

```
library(medicaldata)
```

```
## Warning: package 'medicaldata' was built under R version 4.2.3
```

```
mean_impute = function(col) ifelse(is.na(col), mean(col, na.rm = T), col)
study_data = medicaldata::laryngoscope %>%
mutate(
BMI = mean_impute(BMI),
Mallampati = mean_impute(Mallampati)
)

un_ATE <- study_data %$% {
  mean(total_intubation_time[Randomization == 1]) - mean(total_intubation_time[Randomization == 0])
}
un_ATE
```

```
## [1] 15.65857
```

2.

```
fit <- lm(total_intubation_time~age+gender+asa+BMI+Mallampati+Randomization, study_data)
fit
```

```
##
## Call:
## lm(formula = total_intubation_time ~ age + gender + asa + BMI +
##      Mallampati + Randomization, data = study_data)
```

```
##
## Coefficients:
##    (Intercept)            age          gender            asa             BMI
##        17.7427         0.1237         17.0499         4.7095         -0.1485
##     Mallampati  Randomization
##        -2.5239        15.4073
```

```
Y1_hat <- study_data %>%
  mutate(Randomization = 1) %>%
  predict(fit,.) %>%
mean()

Y0_hat <- study_data %>%
  mutate(Randomization = 0) %>%
  predict(fit,.) %>%
mean()

ad_ATE <- mean(Y1_hat - Y0_hat)

ad_ATE
```

```
## [1] 15.40726
```

3.

- They are mostly the same. This makes sense since this is an RCT where exposure was randomized thus confounders are evenly balanced between the two groups and do not affect the final ATE.

4.

- No since this is an RCT so the predictor function will be linear.

5.

- Yes if "missingness" affected one group more than the other or the distribution of the absence of data is unequal between the two groups then imputing this data could bias our results by us systematically artificially affecting the covariate measurements of one group more than the other.

P3.

1.

- Our plug in estimator is as follows:
- let $\mu(X, d) = E[Y|X, D = d]$
- $E[\mu(X, d)]$
- $\frac{1}{n} \sum_i \mu(X, d)$
- In other words the expectation of the outcome given whether they responded or not is the conditional mean of the outcome given the covariates (X) and whether or not they responded to the survey (D). We thus need a function that predicts the outcome Y given the covariates X and their response sate D. We use linear regression to estimate this predictor function which we can then use, by plugging in our data, to predict the outcomes of each individual if they had and hadn't responded to the survey. We then take the difference in the means of the predicted outcomes to get the estimated ATE.

2.

- We have to assume that the underlying data truly has a linear relationship and that we have measured all possible confounders (i.e. no unmeasured confounders).