hmwrk-12 - Beimnet Taye

2023-04-19

Worked with Joan Shim and Lucas Yoshida

P1.

1.

- The coefficient would probably decrease in magnitude, since age would soak up some of the effect that previously appeared attributable to chemo. and chemo doesn't matter.
- The std error would decrease, since adjusting for age would make the effect of chemo more clearly visible.

2.

- The coefficient would probably stay the same, since SES doesn't do much to help explain survival relative to chemo.
- The std error would probably stay the same, since the correlation between SES and chemo doesn't matter.

P2.

1.

- Biostatistician A is correct since it is an RCT so the probability of being exposed is independent from all other predictors. This means that the effect of the drug cannot be "transferred" over to other predictors in the model due to the independence. Thus the STE of the ATE can only decrease when adding predictors. Additionally it is unlikely for the direction of the coeffecient to reverse when adding predicators in an RCT.

2.

```
mu <- \(x,a) 2*x^2 + 3*a
f <- \(x) 3*x^3

DGP <- \(n){
  tibble(
    X = runif(n),
    A = rbern(n,.5),
    Y1 = mu(X,1) + rnorm(n,0,1),
    Y0 = mu(X,0) + rnorm(n,0,1),
    Y = ifelse(A,Y1,Y0)
```

```
  )
}

DGP(100)
```

```
## # A tibble: 100 x 5
##        X     A    Y1     Y0      Y
##    <dbl> <int> <dbl>  <dbl>  <dbl>
##  1 0.546     0  4.25  1.54   1.54
##  2 0.698     0  4.10 -0.972 -0.972
##  3 0.383     1  4.27  0.999  4.27
##  4 0.521     1  2.51  0.113  2.51
##  5 0.636     0  4.86 -0.124 -0.124
##  6 0.270     0  2.90  1.96   1.96
##  7 0.544     1  4.01  0.186  4.01
##  8 0.943     1  5.24  2.13   5.24
##  9 0.143     1  1.67  0.367  1.67
## 10 0.499     0  3.61  0.601  0.601
## # ... with 90 more rows
```

3.

```
c_ATE <- DGP(10000) %$% {
  mean(Y1) - mean(Y0)
}
c_ATE
```

```
## [1] 3.000519
```

4.

```
lin_reg_est <- function(data) {
  fit <- lm(Y~A+X, data)
  mu_hat <- function(a,x) {predict(fit,tibble(A=a,X=x))}
  ATE_est <- data %>%
    mutate(
      Y1_hat = mu_hat(1,X),
      Y0_hat = mu_hat(0,X)
    ) %$% {
      mean(Y1_hat-Y0_hat)
    }
  return(ATE_est)
}

obsv <- DGP(1000) %>%
  select(-Y1,-Y0) %>%
  lin_reg_est()

lin_reg_est_m <- function(data) {
```

```
  mdata <- data %>%
    mutate(M = f(X))
  fit <- lm(Y~A+X+M, mdata)
  mu_hat <- function(a,x,m) {predict(fit,tibble(A=a,X=x, M=m))}
  ATE_est <- mdata %>%
    mutate(
      Y1_hat = mu_hat(1,X,M),
      Y0_hat = mu_hat(0,X,M)
    ) %$% {
      mean(Y1_hat-Y0_hat)
    }
  return(ATE_est)
}

obsvm <- DGP(1000) %>%
  select(-Y1,-Y0) %>%
  lin_reg_est_m()
```

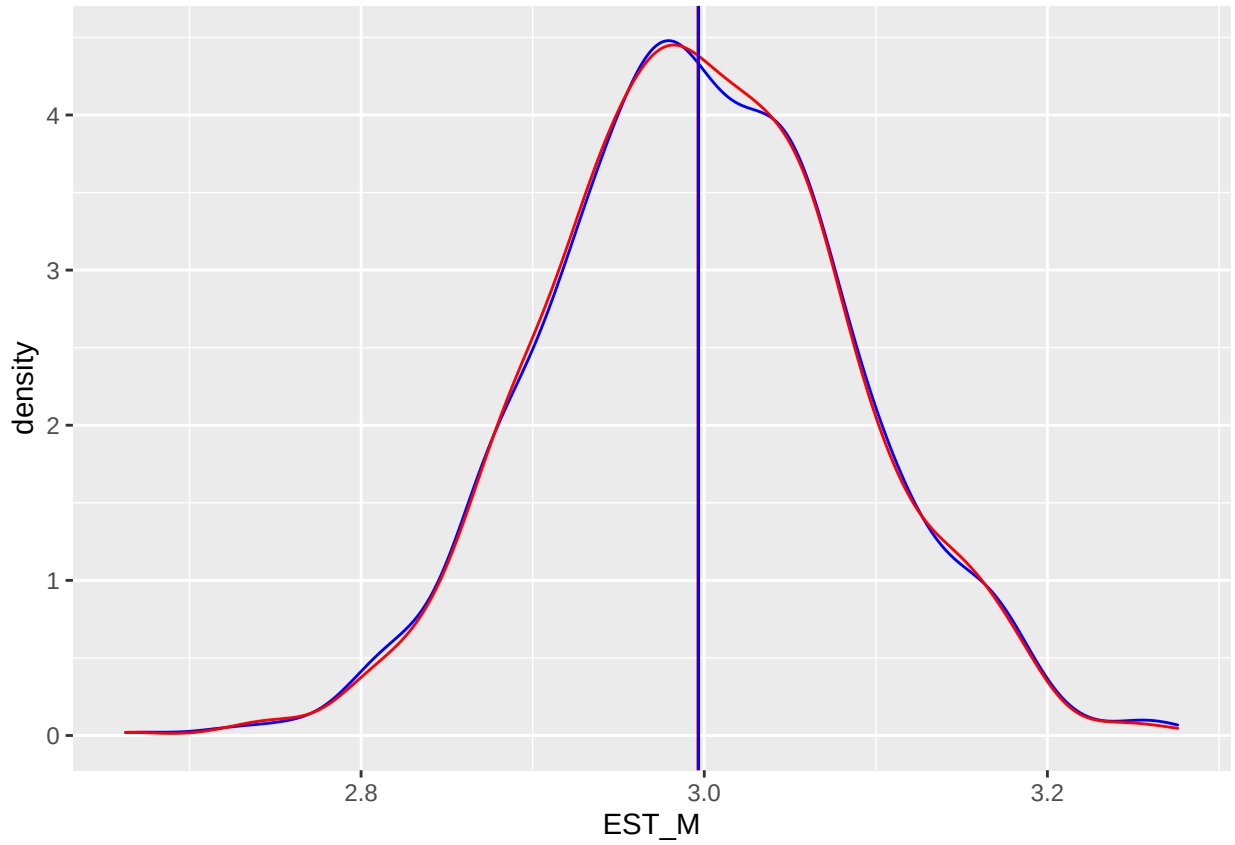
5.

```
samp <- map_df(1:1000,function(.x) {
 data <- DGP(500) %>%
  select(-Y1,-Y0)
 return(tibble(EST = lin_reg_est(data),
               EST_M = lin_reg_est_m(data)
               )
        )
  }
)

distb <- samp %>%
  ggplot() +
  geom_density(aes(x = EST_M),color = "blue") +
  geom_density(aes(x = EST),color = "red") +
  geom_vline(aes(xintercept = mean(EST)), color = "red") +
   geom_vline(aes(xintercept = mean(EST_M)), color = "blue")
  geom_vline(xintercept = c_ATE, color = "green")
```

```
## mapping: xintercept = ~xintercept
## geom_vline: na.rm = FALSE
## stat_identity: na.rm = FALSE
## position_identity
```

```
distb
```



6.

- It proves that bias is not introduced when adding the model as a predictor.

P3.

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)
)
```

```r
fitun <- study_data %>%
  lm(total_intubation_time ~ Randomization,.)

Y1_hat3 <- study_data %>%
  mutate(Randomization = 1) %>%
  predict(fitun,.) %>%
  mean()

Y0_hat3 <- study_data %>%
  mutate(Randomization = 0) %>%
  predict(fitun,.) %>%
  mean()

ATE_un <- Y1_hat3-Y0_hat3

ATE_un
```

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

```r
STE_un <- sqrt(vcovHC(fitun)[2,2])

tibble(Estimate = ATE_un, upper = ATE_un + 1.96 * STE_un, lower = ATE_un - 1.96 * STE_un )
```

```
## # A tibble: 1 x 3
##   Estimate upper lower
##      <dbl> <dbl> <dbl>
## 1     15.7  23.4  7.88
```

2.

```r
fit_r <- study_data %>%
  lm(total_intubation_time ~ Randomization + age + gender + asa + BMI + Mallampati,.)

Y1_hat32 <- study_data %>%
  mutate(Randomization = 1) %>%
  predict(fit_r,.) %>%
  mean()

Y0_hat32 <- study_data %>%
  mutate(Randomization = 0) %>%
  predict(fit_r,.) %>%
  mean()

ATE_r <- Y1_hat32-Y0_hat32

ATE_r
```

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

```
STE_r <- sqrt(vcovHC(fit_r)[2,2])
tibble(Estimate = ATE_r, upper = ATE_r + 1.96 * STE_r, lower = ATE_r - 1.96 * STE_r)
```

```
## # A tibble: 1 x 3
##   Estimate upper lower
##      <dbl> <dbl> <dbl>
## 1     15.4  22.7  8.10
```

3.

```
fit_c <- study_data %>%
  lm(total_intubation_time ~ Randomization + age + gender + asa + BMI + Mallampati,.)

Y1_hat33 <- study_data %>%
  mutate(Randomization = 1) %>%
  predict(fit_c,.) %>%
  mean()

Y0_hat33 <- study_data %>%
  mutate(Randomization = 0) %>%
  predict(fit_c,.) %>%
  mean()

ATE_c <- Y1_hat33-Y0_hat33

ATE_c
```

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

```
STE_c <- sqrt(vcov(fit_c)[2,2])

tibble(Estimate = ATE_c, upper = ATE_c + 1.96 * STE_c, lower = ATE_c - 1.96 * STE_c )
```

```
## # A tibble: 1 x 3
##   Estimate upper lower
##      <dbl> <dbl> <dbl>
## 1     15.4  22.8  8.05
```

4.

- The two adjusted SE methods are mostly the same since the study was an RCT so the predictor function behaves linearly enough to use the classical standard error. The classical method assumes a linear conditional mean function. With this being an RCT our predictor A is independent of any other potential covariate thus or predictor function will behave linearly. As such the linearity assumption is met to use the classical method and is why it is equal to the robust method in this case.Additionally the noise of the outcome is not related to our outcome and is homoscedastic. Additionally the noise of the outcome is largely normal. The unadjusted estimator has a larger standard error however than the other two adjusted methods but this makes sense since this is an RCT and as highlighted in the previous problem in an RCT the additional predictors in the model can only decrease the variance of the ATE.

5.

- linear regression with robust SE.