PSET 4 Q.7

Allen Zhou

2023-11-15

```
A <- lm(birthweight ~ smoker, data = ds)
summary(A)
```

```
##
## Call:
## lm(formula = birthweight ~ smoker, data = ds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3007.06  -313.06    26.94   366.94  2322.94
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3432.06      11.87 289.115   <2e-16 ***
## smoker       -253.23      26.95  -9.396   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 583.7 on 2998 degrees of freedom
## Multiple R-squared:  0.0286, Adjusted R-squared:  0.02828
## F-statistic: 88.28 on 1 and 2998 DF,  p-value: < 2.2e-16
```

(a) difference is 253.23 grams for smoker vs non-smoker

(b)

```
B <- lm(birthweight ~ smoker + alcohol + nprevist, data = ds)
summary(B)
```

```
##
## Call:
## lm(formula = birthweight ~ smoker + alcohol + nprevist, data = ds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2733.53  -307.57    21.42   358.09  2192.70
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3051.249     34.016  89.701  < 2e-16 ***
## smoker      -217.580     26.680  -8.155 5.07e-16 ***
```

```
## alcohol      -30.491     76.234  -0.400    0.689
## nprevist      34.070      2.855  11.933  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 570.5 on 2996 degrees of freedom
## Multiple R-squared:  0.07285,    Adjusted R-squared:  0.07192
## F-statistic: 78.47 on 3 and 2996 DF,  p-value: < 2.2e-16
```

i) both alcohol and nprevist may be correlated and relevant variables to birthweight and smoking, so these variables are confounding variables that may cause part (a) to have an overestimated effect that smoking has on birthweight

ii) we see that the estimated difference fell from 253.23 to 217.58, so its possible that there may have been an omitted variable bias in part (a)

iii) 3051.249 - 217.580(1) - 30.491(0) +34.070(8) = 3106.229 grams

iv) R^2 is 0.07285 and R^2 adjusted is 0.7192 and they are so similar as the sample size is large

v) the coefficient does not measure a causal effect of prenatal visits on birthweight as nprevist is a control variable. It is used to elimiate the effects of other factors besides alcohol and smoking on birthweight.

(c) we would first regress Birthweight on Alcohol and Nprevist and store the residuals. Then we would regress Smoker on Alcohol and Nprevist and store the residuals. Finally, we would regress the residuals from step 1 with the residuals from step 2, which would result in the same estimator that we found for Smoker in part (b)

```
step1 <- lm(birthweight ~ alcohol + nprevist, data = ds)
step2 <- lm(smoker ~ alcohol + nprevist, data = ds)
step3 <- lm(step1$residuals ~ step2$residuals)
summary(step3)
```

```
##
## Call:
## lm(formula = step1$residuals ~ step2$residuals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2733.53  -307.57    21.42   358.09  2192.70
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -1.584e-14  1.041e+01   0.000        1
## step2$residuals -2.176e+02  2.667e+01  -8.158 4.95e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 570.3 on 2998 degrees of freedom
## Multiple R-squared:  0.02172,    Adjusted R-squared:  0.02139
## F-statistic: 66.55 on 1 and 2998 DF,  p-value: 4.955e-16
```