PS3 R Solutions

Matthew Alampay Davis

October 18, 2021

Question 1

Non-empirical. See official solutions.

Question 2

```
gpa4 <- read.dta13("GPA4.dta")
gpa.mod1 <- lm_robust(colGPA ~ hsGPA + skipped, data = gpa4,
    se_type = "stata")
gpa.mod2 <- lm_robust(colGPA ~ hsGPA + skipped + PC, data = gpa4,
    se_type = "stata")
gpa.mod3 <- lm_robust(colGPA ~ hsGPA + skipped + PC + bgfriend +
    campus, data = gpa4, se_type = "stata")
```

Part a

```
summary(gpa.mod1)
```

```
##
## Call:
## lm_robust(formula = colGPA ~ hsGPA + skipped, data = gpa4, se_type = "stata")
##
## Standard error type:  HC1
##
## Coefficients:
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)  1.57917    0.32525   4.855 3.212e-06   0.9360  2.22228 138
## hsGPA        0.45880    0.09415   4.873 2.974e-06   0.2726  0.64497 138
## skipped     -0.07743    0.02539  -3.050 2.741e-03  -0.1276 -0.02724 138
##
## Multiple R-squared:  0.2227 ,    Adjusted R-squared:  0.2115
## F-statistic:  20.9 on 2 and 138 DF,  p-value: 1.181e-08
```

$$\underset{(0.325)}{\widehat{\text{colGPA}} = 1.579} \underset{(0.094)}{+0.458\text{hsGPA}} \underset{(0.025)}{- 0.077\text{skipped}}$$

Part b

The coefficient on hsGPA is 0.458. This means that a one point increase in hsGPA will lead to an increase of 0.458 points in colGPA (college GPA). Students with higher GPAs in high school tend to have higher college GPAs.

Part c

The t-statistic for H0: skipped=0 vs H1: skipped!= 0 is -3.05. Since |-3.05|>1.96 we can reject H0 and conclude skipped is statistically different than 0. We are essentially testing if on average skipping classes would affect students' college GPA (in layman words).

Part d

```
gpa.mod1$statistic["skipped"]
```

```
##   skipped
## -3.050356
```

```
gpa.mod2$statistic["skipped"]
```

```
##   skipped
## -2.528452
```

```
gpa.mod3$statistic["skipped"]
```

```
##   skipped
## -2.736817
```

The critical value for a two-sided test at the 1% significance level is 2.58, so we reject H_0 (that the coefficient on skipped is equal to zero) in Regressions 1 and 3. We cannot reject H_0 in regression 2. (We may also compare the p-values with 0.01 as instructed in the problem)

Part e

```
gpa.mod3$coefficients["campus"]
```

```
##     campus
## -0.1243919
```

The coefficient on campus is -0.124. This means living on campus reduces colGPA by 0.124 points. The negative sign on the coefficient might be because students who live on campus have more distractions that they would if they lived at home. However, students who live on campus are also more able to study with one another, so it is not clear whether the sign of the coefficient should be positive or negative; it would depend on which effect was stronger. The size of the coefficient is about one-tenth of a point, which is small. Note that the coefficient is not statistically different than 0, so there is not too much we can say.

Part f

```
gpa.mod3$coefficients["bgfriend"]
```

```
##   bgfriend
## 0.08586053
```

The coefficient on bgfriend is 0.085. This means having a boyfriend or girlfriend increases GPA by 0.085. One would expect the coefficient to be positive or negative, such as: the coefficient could be positive if students who are more intelligent are more likely to be dating (here we have a correlation), alternatively, the sign could be negative if dating distracts a student from studying. However, here the magnitude of the coefficient is very small also note that the coefficient is not statistically significant. Hence, we can conclude that there is no significant effect on GPA of dating.

Question 3

Non-empirical. See official solutions.