Chapter 3 - The Multiple Regression Analysis - Estimation

Thalles Quinaglia Liduares

21/02/2022

Exercise 3.4

Upload packages

```
library(wooldridge)
library(lmreg)
library(dplyr)
```

Upload database

```
data<-wooldridge::attend

attach(data)
```

Use the data in ATTEND.RAW for this exercise. (i) Obtain the minimum, maximum, and average values for the variables atndrte, priGPA, and ACT.

The `min`, `max` and `mean` for the variable `atndrte` are given, respectively by

```
atn_stat<-data %>%
          summarise(minimum = min(atndrte),
                    maximum = max(atndrte),
                    mean_value = round(mean(atndrte, na.rm=TRUE),2))
atn_stat
```

```
##   minimum maximum mean_value
## 1    6.25     100      81.71
```

The `min`, `max` and `mean` for the variable `priGPA` are given, respectively by

```
priGPA_stat<-data %>%
          summarise(minimum = min(priGPA),
                    maximum = max(priGPA),
                    mean_value = round(mean(priGPA, na.rm=TRUE),2))

priGPA_stat
```

```
##   minimum maximum mean_value
## 1   0.857    3.93       2.59
```

Finally, the `min`, `max` and `mean` for the variable `ACT` are given, respectively by

```
ACT_stat<-data %>%
          summarise(minimum = min(ACT),
                    maximum = max(ACT),
                    mean_value = round(mean(ACT, na.rm=TRUE),2))

ACT_stat
```

```
##   minimum maximum mean_value
## 1      13      32      22.51
```

(ii) Estimate the model

$$atndrte = \beta_0 + \beta_1 priGPA + \beta_2 ACT + u$$

and write the results in equation form. Interpret the intercept. Does it have a useful meaning?

```
lm1<-lm(atndrte~priGPA+ACT)

summary(lm1)
```

```
##
## Call:
## lm(formula = atndrte ~ priGPA + ACT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -65.373  -6.765   2.125   9.635  29.615
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   75.700      3.884   19.49   <2e-16 ***
## priGPA        17.261      1.083   15.94   <2e-16 ***
## ACT           -1.717      0.169  -10.16   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.38 on 677 degrees of freedom
## Multiple R-squared:  0.2906, Adjusted R-squared:  0.2885
## F-statistic: 138.7 on 2 and 677 DF,  p-value: < 2.2e-16
```

The estimated equation is given by

$$\widehat{atndrte} = 75.7 + 17.2 priGPA - 1.71 ACT$$

The intercept represents, the value of `atndrte` if `priGPA` and `ACT` are equal to zero, which not make pratical sense, given that students spent at least some time in studies.

Discuss the estimated slope coefficients. Are there any surprises?

Given the estimated coefficients, for a marginal variation in `priGPA` the percentual increase in `atndrte` is 17.2%, which makes sense from the theoretical point of view, given that students who frequent more classes, tends to perform better in exams. Otherwise, the coefficient associated to ACT, is -1.71, which not presents the expected sign, given that students with higher `ACT` tends to frequent more classes.

(iv) What is the predicted atndrte if priGPA = 3.65 and ACT = 20? What do you make of this result? Are there any students in the sample with these values of the explanatory variables?

substituting theses values in estimated equation, we obtain

$$\widehat{atndre} = 75.7 + 17.2 \times (3.65) - 1.71 \times (20) = 104.28$$

This value doesnt make sense, as `atndre` is the percentual value for attended classes by some student.

(v) If Student A has priGPA = 3.1 and ACT = 21 and Student B has priGPA = 2.1 and ACT = 26, what is the predicted difference in their attendance rates?

For student A: $\widehat{atndre}_A = 75.7 + 17.2 \times (3.1) - 1.71 \times (21) = 93.1$

For Student B: $\widehat{atndre}_B = 75.7 + 17.2 \times (2.1) - 1.71 \times (26) = 67.4$

Hence, the predicted difference is $\widehat{atndre}_A - \widehat{atndre}_B = 93.1 - 67.4 = 25.7$