AIED\_Prediction\_Assignment1

Hyo Jeong Shin

2023-06-02

## Set the working directory and read in the dataset (PISA2018MS\_KOR\_BQ.rdata)

infile <- outfile <- "/Users/hyojeong/Library/Mobile Documents/com~apple~CloudDocs/Teaching/2023\_1/UnderGrad\_AIED"  
setwd(infile)  
getwd()

## [1] "/Users/hyojeong/Library/Mobile Documents/com~apple~CloudDocs/Teaching/2023\_1/UnderGrad\_AIED"

load("PISA2018MS\_KOR\_BQ.Rdata")  
class(PISA2018MS\_KOR)

## [1] "data.frame"

## Prediction task using the linear regression

### Check the dimension of the dataset

dim(PISA2018MS\_KOR)

## [1] 6650 863

### Compute the mean and sd of PV1MATH variable (outcome variable)

mean(PISA2018MS\_KOR$PV1MATH)

## [1] 526.5421

sd(PISA2018MS\_KOR$PV1MATH)

## [1] 99.19173

### Fit the following three prediction models and return the summary of each model:

* M1: PV1MATH by EMOSUPS
* M2: PV1MATH by explanatory variable: EMOSUPS, ST004D01T (gender)
* M3: PV1MATH by explanatory variables: EMOSUPS, ST004D01T (gender), interaction between EMOSUPS and ST004D01T (gender)

M1 <- lm(PV1MATH ~ EMOSUPS, data = PISA2018MS\_KOR)  
M2 <- lm(PV1MATH ~ EMOSUPS + as.factor(ST004D01T), data = PISA2018MS\_KOR)  
M3 <- lm(PV1MATH ~ EMOSUPS + as.factor(ST004D01T) + as.factor(ST004D01T):EMOSUPS, data = PISA2018MS\_KOR)

### Interpret the coefficients of M2 and M3

#### M2 Results

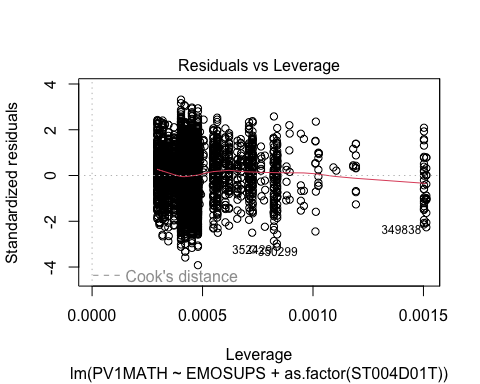
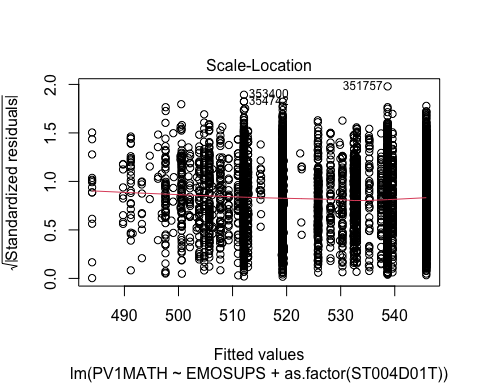
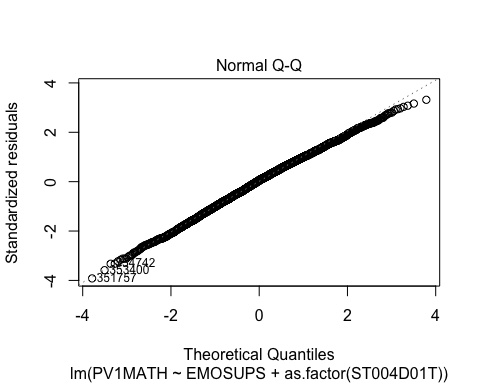
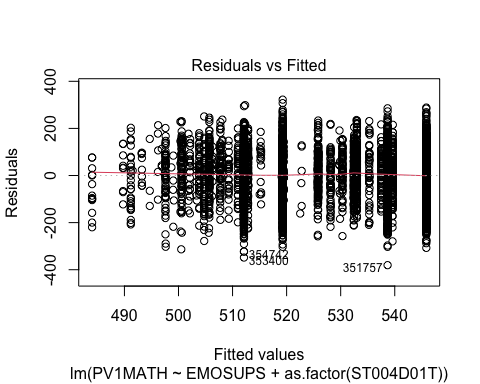
* intercept = 522.432 : 522.4 is the expected MATH score for girls when EMOSUPS is zero
* ST004D01T = 7.162 : MATH score is 7.16 higher for boys, on average, controlling for the EMOSUPS
* EMOSUPS = 15.699 : one unit increase in EMOSUPS is associated with 15.699 higher MATH score, on average, for the given gender (i.e., same EMOSUPS effect for boys and girls)

#### M3 Results

* intercept = 522.544: 522.5 is the expected MATH score for girls when EMOSUPS is zero
* ST004D01T = 6.937 : MATH score is 6.94 higher for boys, on average, when EMOSUPS is zero
* EMOSUPS = 14.548: one unit increase in EMOSUPS is associated with 14.55 higher MATH score, on average, for girls
* as.factor(ST004D01T):EMOSUPS = 2.190 : one unit increase in EMOSUPS is associated with 16.738 (14.548 + 2.190) higher MATH score, on average, for boys
* Interaction term that is not statistically significant indicates that the effect of EMOSUPS is not statistically different for boys and girls

# Check the assumptions of M2 using the plot function

plot(M2)



Normal Q-Q plot and the residual vs. fitted plot show that the errors follow the normal distribution and that the errors appear constant.