AIED\_Prediction\_Assignment1

20201564 김성현

2023-05-22

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

infile <- outfile <- "C:/aied"  
setwd(infile)   
getwd()

## [1] "C:/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)  
summary(M1)

##   
## Call:  
## lm(formula = PV1MATH ~ EMOSUPS, data = PISA2018MS\_KOR)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -384.38 -65.52 4.73 68.24 325.60   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 526.144 1.208 435.67 <2e-16 \*\*\*  
## EMOSUPS 15.727 1.322 11.89 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 97.22 on 6563 degrees of freedom  
## (결측으로 인하여 85개의 관측치가 삭제되었습니다.)  
## Multiple R-squared: 0.0211, Adjusted R-squared: 0.02095   
## F-statistic: 141.5 on 1 and 6563 DF, p-value: < 2.2e-16

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

##   
## Call:  
## lm(formula = PV1MATH ~ as.factor(ST004D01T) + EMOSUPS, data = PISA2018MS\_KOR)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -380.64 -65.70 5.07 68.54 322.13   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 522.432 1.733 301.429 < 2e-16 \*\*\*  
## as.factor(ST004D01T)2 7.162 2.400 2.984 0.00285 \*\*   
## EMOSUPS 15.699 1.321 11.881 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 97.16 on 6562 degrees of freedom  
## (결측으로 인하여 85개의 관측치가 삭제되었습니다.)  
## Multiple R-squared: 0.02243, Adjusted R-squared: 0.02213   
## F-statistic: 75.28 on 2 and 6562 DF, p-value: < 2.2e-16

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

##   
## Call:  
## lm(formula = PV1MATH ~ as.factor(ST004D01T) + EMOSUPS + as.factor(ST004D01T):EMOSUPS,   
## data = PISA2018MS\_KOR)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -379.56 -65.87 5.11 68.60 322.92   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 522.544 1.738 300.583 < 2e-16 \*\*\*  
## as.factor(ST004D01T)2 6.937 2.416 2.872 0.00409 \*\*   
## EMOSUPS 14.548 1.919 7.581 3.9e-14 \*\*\*  
## as.factor(ST004D01T)2:EMOSUPS 2.190 2.646 0.827 0.40806   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 97.17 on 6561 degrees of freedom  
## (결측으로 인하여 85개의 관측치가 삭제되었습니다.)  
## Multiple R-squared: 0.02253, Adjusted R-squared: 0.02209   
## F-statistic: 50.41 on 3 and 6561 DF, p-value: < 2.2e-16

### Interpret the coefficients of M2 and M3

#### M2 Results

* intercept = 522.432:
* ST004D01T = 7.162:
* EMOSUPS = 15.699:

#### M3 Results

* intercept = 522.544:
* ST004D01T = 6.937:
* EMOSUPS = 14.548:
* as.factor(ST004D01T) = 2.190:
* Interaction term shows that One unit increase in EMOSUPS is associated with 14.548 higher mathematics score, on average, for girls, and 16.738 (14.548+2.190) higher mathematics score, on average, for boys.

# Check the assumptions of M2 using the plot function

plot(M2)

