Attend Exercise

## Exercise ( attend )

## Setup

setwd("C:/Users/22700/Desktop")

## Load Packages

library(GGally)

## Loading required package: ggplot2

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

library(ggplot2)  
library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(data.table)

## Load Data

load("attend.RData")  
dt.attend <- data.table(data)  
rm(data)  
colnames(dt.attend)

## [1] "attend" "termGPA" "priGPA" "ACT" "final" "atndrte" "hwrte"   
## [8] "frosh" "soph" "missed" "stndfnl"

## i) Obtain the minimum, maximum, and average values for the variables atndrte, priGPA, and ACT.

dt.attend[, max(atndrte)]

## **[1] 100**

dt.attend[, min(atndrte)]

## **[1] 6.25**

dt.attend[, mean(atndrte)]

## **[1] 81.70956**

dt.attend[, max(priGPA)]

## **[1] 3.93**

dt.attend[, min(priGPA)]

## **[1] 0.857**

dt.attend[, mean(priGPA)]

## **[1] 2.586775**

dt.attend[, max(ACT)]

## **[1] 32**

dt.attend[, min(ACT)]

## **[1] 13**

dt.attend[, mean(ACT)]

## **[1] 22.51029**

## ii) How many students are in their freshman year?

dt.attend[, sum(frosh)]

## **[1] 158**

## iii) How many of the students in freshmen year have missed more than 10 classes?

dt.attend[missed>10, sum(frosh)]

## **[1] 29**

## iv) Is it true that, on average, students who miss less than 10 classes have a higher GPA? State the null and alternative hypothesis and conduct the appropriate test.

Remain: Error in match.arg(alternative) : 'arg' should be one of “two.sided”, “less”, “greater”

dt.attend[, t.test(termGPA ~ (missed>=10), alternative = c("greater"))]

##   
## Welch Two Sample t-test  
##   
## data: termGPA by missed >= 10  
## t = 12.237, df = 150.04, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.8018051 Inf  
## sample estimates:  
## mean in group FALSE mean in group TRUE   
## 2.761899 1.834686

**The p-value is way smaller than 0.05 at 95% confidence interval, therefore we can reject the null. It has a different on average GPA between those two groups.**

## v) Estimate the model equation, and write the results in equation form.Interpret the intercept. Does it have a useful meaning?

lm.attend <- lm(atndrte ~ priGPA + ACT, data=dt.attend)  
stargazer(lm.attend, type = "text")

##   
## ===============================================  
## Dependent variable:   
## ---------------------------  
## atndrte   
## -----------------------------------------------  
## priGPA 17.261\*\*\*   
## (1.083)   
##   
## ACT -1.717\*\*\*   
## (0.169)   
##   
## Constant 75.700\*\*\*   
## (3.884)   
##   
## -----------------------------------------------  
## Observations 680   
## R2 0.291   
## Adjusted R2 0.288   
## Residual Std. Error 14.379 (df = 677)   
## F Statistic 138.651\*\*\* (df = 2; 677)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Atndrte=75.7+17.261priGPA-1.717ACT, intercept indicates that, for a cumulative GPA and ACT score of zero will be equal to zero, the attendance should be 75.7%.**

## vi) Discuss the estimated slope coefficients. Are there any surprises?

## vii) 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?

new.student <- data.table( priGPA = 3.65  
 , ACT = 20)  
pred.new.student <- predict(lm.attend, newdata = new.student)  
pred.new.student

## 1   
## 104.3705

(dt.attend[priGPA==3.65, sum(ACT=20)])

## [1] 20

**There is a student could have both.**

## viii) 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?

#predict A  
studentA <- data.table( priGPA = 3.61  
 , ACT = 21)  
pred.studentA <- predict(lm.attend, newdata = studentA)  
pred.studentA

## 1   
## 101.9635

#predict B  
studentB <- data.table( priGPA = 2.1  
 , ACT = 26)  
pred.studentB <- predict(lm.attend, newdata = studentB)  
pred.studentB

## 1   
## 67.31727

#Difference  
pred.studentA - pred.studentB

## 1   
## 34.64626

## ix) Should we include attend along with alcohol as explanatory variables in a multiple regression model? (Think about how you would interpret alcohol.)

**Alcohol coefficient can explain the level of GPA to a certain extent, because there may be a certain correlation between alcohol consumption and attendance.**

## x) Should SAT and hsGPA be included as explanatory variables? Explain.

It might be needed, so one of these variables should be needed.