assignment\_07\_RiceMichelle.R

mrice

2021-01-30

# Assignment: ASSIGNMENT 7  
# Name: Rice, Michelle  
# Date: 2021-01-27  
  
## Set the working directory to the root of your DSC 520 directory  
setwd("/Users/mrice/DSC520\_Rice/dsc520")  
  
## Load the `data/r4ds/heights.csv` to  
heights\_df <- read.csv("data/r4ds/heights.csv")  
  
# Fit a linear model  
earn\_lm <- lm(earn ~ ed + race + height + age + sex, data=heights\_df)  
  
# View the summary of your model  
summary(earn\_lm)

##   
## Call:  
## lm(formula = earn ~ ed + race + height + age + sex, data = heights\_df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -39423 -9827 -2208 6157 158723   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -41478.4 12409.4 -3.342 0.000856 \*\*\*  
## ed 2768.4 209.9 13.190 < 2e-16 \*\*\*  
## racehispanic -1414.3 2685.2 -0.527 0.598507   
## raceother 371.0 3837.0 0.097 0.922983   
## racewhite 2432.5 1723.9 1.411 0.158489   
## height 202.5 185.6 1.091 0.275420   
## age 178.3 32.2 5.537 3.78e-08 \*\*\*  
## sexmale 10325.6 1424.5 7.249 7.57e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 17250 on 1184 degrees of freedom  
## Multiple R-squared: 0.2199, Adjusted R-squared: 0.2153   
## F-statistic: 47.68 on 7 and 1184 DF, p-value: < 2.2e-16

predicted\_df <- data.frame(  
 earn = predict(earn\_lm, heights\_df),  
 ed=heights\_df$ed, race=heights\_df$race, height=heights\_df$height,  
 age=heights\_df$age, sex=heights\_df$sex  
 )  
  
## Compute deviation (i.e. residuals)  
mean\_earn <- mean(heights\_df$earn)  
## Corrected Sum of Squares Total  
sst <- sum((mean\_earn - heights\_df$earn)^2)  
## Corrected Sum of Squares for Model  
ssm <- sum((mean\_earn - predicted\_df$earn)^2)  
## Residuals  
residuals <- heights\_df$earn - predicted\_df$earn  
## Sum of Squares for Error  
sse <- sum(residuals^2)  
## R Squared  
r\_squared <- ssm/sst  
  
## Number of observations  
n <- nrow(predicted\_df)  
## Number of regression paramaters  
p <- 8  
## Corrected Degrees of Freedom for Model  
dfm <- p-1  
## Degrees of Freedom for Error  
dfe <- n-p  
## Corrected Degrees of Freedom Total: DFT = n - 1  
dft <- n-1  
  
## Mean of Squares for Model: MSM = SSM / DFM  
msm <- ssm/dfm  
## Mean of Squares for Error: MSE = SSE / DFE  
mse <- sse/dfe  
## Mean of Squares Total: MST = SST / DFT  
mst <- sst/dft  
## F Statistic  
f\_score <- msm/mse  
  
## Adjusted R Squared R2 = 1 - (1 - R2)(n - 1) / (n - p)  
adjusted\_r\_squared <- 1 - (1 - r\_squared) \* (n - 1) / (n - p)