# Assignment: ASSIGNMENT 7

# Name: Sahu, Jyoti

# Date: 2021-28-01

**Github- https://github.com/bellevue-university/dsc520/pull/48**

# Assignment: ASSIGNMENT 7

# Name: Sahu, Jyoti

# Date: 2021-01-28

1## Set the working directory to the root of your DSC 520 directory

setwd("/Users/sahujyot/Documents/DSC520")

2## Load the `data/r4ds/heights.csv` to

heights\_df <- read.csv("data/r4ds/heights.csv")

3# Fit a linear model

earn\_lm <- lm(earn ~ height + sex + ed+ age +race, data=heights\_df)

Call:

lm(formula = earn ~ height + sex + ed + age + race, data = heights\_df)

Coefficients:

(Intercept) height sexmale ed age racehispanic raceother racewhite

-41478.5 202.5 10325.6 2768.4 178.3 -1414.3 371.0 2432.5

4# View the summary of your model

summary(earn\_lm)

earn height sex ed age race

Min. : 200 Min. :57.50 Length:1192 Min. : 3.0 Min. :18.00 Length:1192

1st Qu.: 10000 1st Qu.:64.01 Class :character 1st Qu.:12.0 1st Qu.:29.00 Class :character

Median : 20000 Median :66.45 Mode :character Median :13.0 Median :38.00 Mode :character

Mean : 23155 Mean :66.92 Mean :13.5 Mean :41.38

3rd Qu.: 30000 3rd Qu.:69.85 3rd Qu.:16.0 3rd Qu.:51.00

Max. :200000 Max. :77.05 Max. :18.0 Max. :91.00

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

)

5## Compute deviation (i.e. residuals)

mean\_earn <- mean(heights\_df$earn)

[1] 23154.77

6## Corrected Sum of Squares Total

sst <- sum((mean\_earn - heights\_df$earn)^2)

[1] 451591883937

7## Corrected Sum of Squares for Model

ssm <- sum((mean\_earn - predicted\_df$earn)^2)

[1] 99302918657

8## Residuals

residuals <- heights\_df$earn - predicted\_df$earn

9## Sum of Squares for Error

sse <- sum(residuals^2)

[1] 352288965280

10## R Squared

r\_squared <- ssm/sst

[1] 0.2198953

11## Number of observations

n <- nrow(predicted\_df)

1192

12## Number of regressions parameters

p <- 8

[1] 8

13## Corrected Degrees of Freedom for Model

dfm <- p-1

[1] 7

14## Degrees of Freedom for Error

dfe <- n-p

[1] 1184

15## Corrected Degrees of Freedom Total: DFT = n - 1

dft <- n-1

[1] 1191

16## Mean of Squares for Model: MSM = SSM / DFM

msm <- ssm/dfm

[1] 14186131237

17## Mean of Squares for Error: MSE = SSE / DFE

mse <- sse/dfe

[1] 297541356

18## Mean of Squares Total: MST = SST / DFT

mst <- sst/dft

[1] 379170348

## F Statistic

f\_score <- msm/mse

[1] 47.67785

## Adjusted R Squared R2 = 1 - (1 - R2)(n - 1) / (n - p)

adjusted\_r\_squared <- 1 - ((1 - r\_squared)\*(n - 1) / (n - p))

[1] 0.2152832