title: "assignment\_06\_RamirezKyle"

author: "Kyle Ramirez"

date: "2/13/2022"

output: pdf\_document

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

setwd("/Users/Kyle/Documents/GitHub/KR/Ramirez\_Kyle\_DSC510/dsc520")

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

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

## Load the ggplot2 library

library(ggplot2)

## Fit a linear model using the `age` variable as the predictor and `earn` as the outcome

age\_lm <- lm(age ~ earn, data = heights\_df)

## View the summary of your model using `summary()`

summary(age\_lm)

## Creating predictions using `predict()`

age\_predict\_df <- data.frame(earn = predict(50000, 70000), age=45)

## Plot the predictions against the original data

ggplot(data = age\_predict\_df, aes(y = earn, x = age)) +

geom\_point(color='blue') +

geom\_line(color='red',data = age\_predict\_df, aes(y=earn, x=age))

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 - age\_predict\_df$earn)^2)

## Residuals

residuals <- heights\_df$earn - age\_predict\_df$earn

## Sum of Squares for Error

sse <- sum(residuals^2)

## R Squared R^2 = SSM\SST

r\_squared <- ssm/sst

## Number of observations

n <- 3

## Number of regression parameters

p <- 2

## Corrected Degrees of Freedom for Model (p-1)

dfm <- p - 1

## Degrees of Freedom for Error (n-p)

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 = MSM/MSE

f\_score <- msm / mse

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

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

## Calculate the p-value from the F distribution

p\_value <- pf(f\_score, dfm, dft, lower.tail=F)