**Summary of Liner Model Data:**

> 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 |

> summary(predicted\_df)

|  |
| --- |
| earn ed race height age sex  Min. :-10695 Min. : 3.0 Length:1192 Min. :57.50 Min. :18.00 Length:1192  1st Qu.: 16151 1st Qu.:12.0 Class :character 1st Qu.:64.01 1st Qu.:29.00 Class :character  Median : 23148 Median :13.0 Mode :character Median :66.45 Median :38.00 Mode :character  Mean : 23155 Mean :13.5 Mean :66.92 Mean :41.38  3rd Qu.: 29051 3rd Qu.:16.0 3rd Qu.:69.85 3rd Qu.:51.00  Max. : 47740 Max. :18.0 Max. :77.05 Max. :91.00 |

**R Console output from R Studio**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | |  | | --- | | Assignment: ASSIGNMENT 7 - Exercise 11: Multiple Regression with R (Github)  > # Name: Gunasekaran, Ragunath  > # Date: 2020-10-12  >  > ## Set the working directory to the root of your DSC 520 directory  > setwd("C:/Users/ragun/Documents/GitHub/dsc520-master/DSC520-new")  >  > ## 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$sex  + )  >  > # View the summary of your model  > summary(predicted\_df)  earn ed race height age sex  Min. :-10695 Min. : 3.0 Length:1192 Min. :57.50 Min. :18.00 Length:1192  1st Qu.: 16151 1st Qu.:12.0 Class :character 1st Qu.:64.01 1st Qu.:29.00 Class :character  Median : 23148 Median :13.0 Mode :character Median :66.45 Median :38.00 Mode :character  Mean : 23155 Mean :13.5 Mean :66.92 Mean :41.38  3rd Qu.: 29051 3rd Qu.:16.0 3rd Qu.:69.85 3rd Qu.:51.00  Max. : 47740 Max. :18.0 Max. :77.05 Max. :91.00  >  > ## Compute deviation (i.e. residuals)  >  > mean\_earn <- mean(heights\_df$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^2 = SSM\SST  > r\_squared <- ssm/sst  >  > ## Number of observations  > n <- length(coefficients(earn\_lm))  >  > ## 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  > f\_score  [1] 1.691275  >  > ## Adjusted R Squared R2 = 1 - (1 - R2)(n - 1) / (n - p)  > adjusted\_r\_squared <- 1 - ((1 - r\_squared)\*dft / dfe)  > adjusted\_r\_squared  [1] 0.08987781  >  > ## Calculate the p-value from the F distribution  > p\_value <- pf(f\_score, dfm, dft, lower.tail=F)  > p\_value  [1] 0.2346089 | |  | | |  | | --- | | > | | | | |  | | --- | |  | | |