July 30, 2023

The results below are generated from an R script.

```
# Assignment: ASSIGNMENT 7
# Name: Reppeto, Brian
# Date: 2023-07-24

## Set the working directory to the root of your DSC 520 directory
setwd("~/DSC520/dsc520")

## Load the `data/r4ds/heights.csv` to
heights_df <- read.csv("data/r4ds/heights.csv")

# Fit a linear model
earn_lm <-  lm(earn ~ height + sex + ed + age + race, data=heights_df)

# View the summary of your model
summary(earn_lm)

##
## Call:
## lm(formula = earn ~ height + sex + ed + age + race, 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 ***
## height          202.5      185.6   1.091 0.275420
## sexmale       10325.6     1424.5   7.249 7.57e-13 ***
## ed             2768.4      209.9  13.190  < 2e-16 ***
## age             178.3       32.2   5.537 3.78e-08 ***
## 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
## ---
## 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, newdata=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(heights_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)
```

The R session information (including the OS info, R version and all packages used):

```
sessionInfo()

## R version 4.3.0 (2023-04-21)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Ventura 13.4.1
##
## Matrix products: default
## BLAS:   /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versio
## LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib;  LAPACK ve
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
```

```
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] knitr_1.43       lm.beta_1.7-2    corrplot_0.92    lmtest_0.9-40    zoo_1.8-12
##  [6] car_3.1-2        carData_3.0-5    conflicted_1.2.0 readxl_1.4.3     lubridate_1.9.2
## [11] forcats_1.0.0    stringr_1.5.0    dplyr_1.1.2      purrr_1.0.1      readr_2.1.4
## [16] tidyr_1.3.0      tibble_3.2.1     ggplot2_3.4.2    tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] utf8_1.2.3        generics_0.1.3    stringi_1.7.12    lattice_0.21-8
##  [5] hms_1.1.3         magrittr_2.0.3    evaluate_0.21     grid_4.3.0
##  [9] timechange_0.2.0  fastmap_1.1.1     cellranger_1.1.0  tinytex_0.45
## [13] fansi_1.0.4       scales_1.2.1      abind_1.4-5       cli_3.6.1
## [17] rlang_1.1.1       munsell_0.5.0     withr_2.5.0       cachem_1.0.8
## [21] tools_4.3.0       tzdb_0.4.0        memoise_2.0.1     colorspace_2.1-0
## [25] vctrs_0.6.3       R6_2.5.1          lifecycle_1.0.3   pkgconfig_2.0.3
## [29] pillar_1.9.0      gtable_0.3.3      glue_1.6.2        highr_0.10
## [33] xfun_0.39         tidyselect_1.2.0  rstudioapi_0.15.0 farver_2.1.1
## [37] xtable_1.8-4      labeling_0.4.2    compiler_4.3.0
```

```
Sys.time()
```

```
## [1] "2023-07-30 16:17:41 EDT"
```