Assignment 07

Felipe Rodriguez

2023-02-12

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

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

```
setwd("/Users/feliperodriguez/Library/CloudStorage/OneDrive-BellevueUniversity/Github/dsc520")
heights_df <- read.csv("data/r4ds/heights.csv")
```

Fit a linear model

```
earn_lm <-  lm(age ~ earn + height + sex + ed + race, data=heights_df)
```

View the summary of your model

```
summary(earn_lm)
```

```
##
## Call:
## lm(formula = age ~ earn + height + sex + ed + race, data = heights_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -29.19 -11.42  -3.54   8.63  51.06
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.000e+02  1.072e+01   9.329  < 2e-16 ***
## earn          1.416e-04  2.557e-05   5.537 3.78e-08 ***
## height       -7.098e-01  1.641e-01  -4.325 1.66e-05 ***
## sexmale       3.706e-01  1.297e+00   0.286    0.775
## ed           -1.174e+00  1.974e-01  -5.948 3.57e-09 ***
## racehispanic -3.657e+00  2.391e+00  -1.530    0.126
## raceother     3.741e+00  3.417e+00   1.095    0.274
## racewhite     1.672e+00  1.537e+00   1.088    0.277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 15.37 on 1184 degrees of freedom
## Multiple R-squared:  0.06722,    Adjusted R-squared:  0.0617
## F-statistic: 12.19 on 7 and 1184 DF,  p-value: 4.142e-15
```

```
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(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)
```