ASSIGNMENT 7

Amelia Farrell

October 18th 2021

## Multiple Linear Regression

Making predictions in earning potential

# Fit a linear model

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

Viewing the summary of your model

##   
## Call:  
## lm(formula = earn ~ age + sex + height + ed + 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 \*\*\*  
## age 178.3 32.2 5.537 3.78e-08 \*\*\*  
## sexmale 10325.6 1424.5 7.249 7.57e-13 \*\*\*  
## height 202.5 185.6 1.091 0.275420   
## 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   
## ---  
## 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

Compute deviation (i.e. residuals)

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

Corrected Sum of Squares Total

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  
r\_squared

## [1] 0.2198953

Number of observations

n <- nrow(heights\_df)  
n

## [1] 1192

Number of regression parameters

p <- 2  
p

## [1] 2

Corrected Degrees of Freedom for Model (p-1)

dfm <- p-1  
dfm

## [1] 1

Degrees of Freedom for Error (n-p)

dfe <- n-p  
dfe

## [1] 1190

Corrected Degrees of Freedom Total: DFT = n - 1

dft <- n-1  
dft

## [1] 1191

Mean of Squares for Model: MSM = SSM / DFM

msm <- ssm/dfm  
msm

## [1] 99302918656

Mean of Squares for Error: MSE = SSE / DFE

mse <- sse/dfe  
mse

## [1] 296041147

Mean of Squares Total: MST = SST / DFT

mst <- sst/dft  
mst

## [1] 379170348

F Statistic F = MSM/MSE

f\_score <- msm/mse  
f\_score

## [1] 335.4362

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

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

## [1] 0.2192397

## References

Field, A., J. Miles, and Z. Field. 2012. Discovering Statistics Using R. SAGE Publications. <https://books.google.com/books?id=wd2K2zC3swIC>.

Lander, J. P. 2014. R for Everyone: Advanced Analytics and Graphics. Addison-Wesley Data and Analytics Series. Addison-Wesley. <https://books.google.com/books?id=3eBVAgAAQBAJ>.