Natan Alper 2/3/2020

Business Intelligence & Consumer Insights- Professor Kovtun

HW #1

**1 (a)**

FullName <- c()

for (i in (1:nrow(titanic\_data))) {

name <- as.character(titanic\_data[i,colnames(titanic\_data)=="Name"])

# retrieve name of ith row

comma\_pos <- regexpr(",",name) # position of comma

period\_pos <- regexpr("\\.",name) # position of period

total <- nchar(name) # total number of characters in string "name"

FullName <- c(FullName, paste(substr(name, start=period\_pos+2, stop=total), substr(name, start=1, stop=comma\_pos-1)))

}

FullName

**1 (b)**

survivorFullName <- c()

for (i in (1:nrow(titanic\_data))) {

if (titanic\_data$Survived[i]==1) {

name <- as.character(titanic\_data[i,colnames(titanic\_data)=="Name"])

# retrieve name of ith row

comma\_pos <- regexpr(",",name) # position of comma

period\_pos <- regexpr("\\.",name) # position of period

total <- nchar(name) # total number of characters in string "name"

survivorFullName <- c(survivorFullName, paste(substr(name, start=period\_pos+2, stop=total), substr(name, start=1, stop=comma\_pos-1)))

}

}

survivorFullName

**2 (a)**

Yi = a + b1x1 + b2x2 + b3x3 + ei ei = epsilon

Value of NFL Team = a + b1\*(Debt to value)i + b2\*(Revenue)i + b3\*(Operating Income)i + ei

**2 (b)**

Call:

lm(formula = Current\_Value ~ Debt\_to\_Value + Revenue + Operating\_Income,

data = nflTeams)

**Coefficients**:

(Intercept) Debt\_to\_Value Revenue Operating\_Income

55.678 1553.736 4.951 2.226

**2 (c)**

Value of NFL Team = 55.678 + 1553.736\*(Debt to Value)i + 4.951\*(Revenue)i + 2.226\*(Operating Income)i + ei

**2 (d)**

nflTeams <- NFLTEAMVALUES2018

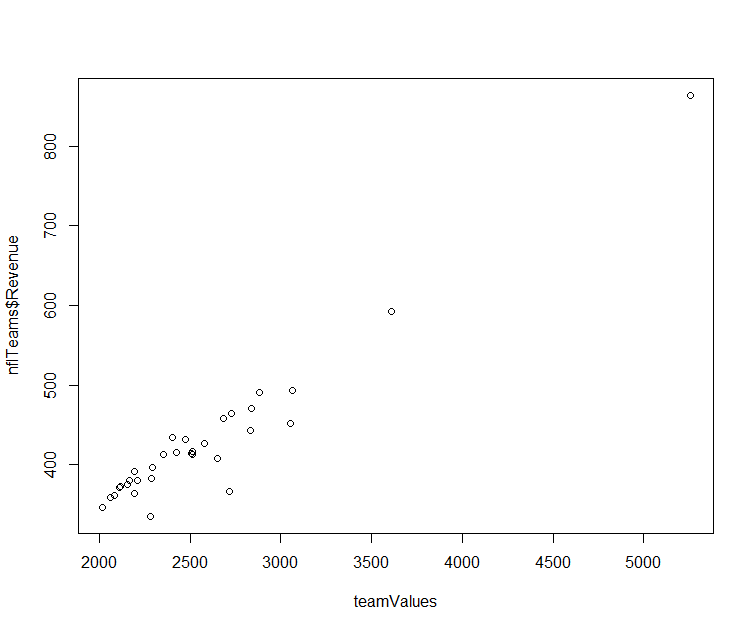
teamValues <- 55.678 + 1553.736\*nflTeams$Debt\_to\_Value + 4.951\*nflTeams$Revenue + 2.226\*nflTeams$Operating\_Income

teamValues

**2 (e)**

|2850 – 2833.561| = 16.439

**2 (f)**



**3 (a)**

x1 = sample(1:50, 30)

x2 = sample(-25:25, 30)

eps = rnorm(30, mean=0, sd=2)

y = 10 + 3\*x1 + 5\*x2 + eps # OLR Model

**3 (b)**

y = 10 + 3\*x1 + 5\*x2 + eps

**3 (c)**

Coefficients:

(Intercept) olrData$x1 olrData$x2

11.887 2.960 4.924

**3 (d)**

Coefficients:

(Intercept) olrData\_1$x\_1 olrData\_1$x\_2

60.6617 0.5735 -0.6947

The model in part (c), the model with the smaller variance, is does a better job estimating the coefficients.