Take Home Exam 2

Joshua Oswari - A14751270

6/6/2019

Problem 1

Part A.

```
bootGOFdiff <- function(x,y, B=2000) {
m = length(x)
n = length(y)
D= abs(mean(x) - mean(y))
Z = c(x, y)
D.sim = numeric(B)
for (b in 1:B) {
    Zperm = sample(Z)
    D.sim[b] = mean(Zperm[1:m]) - mean(Zperm[(m+1):(m+n)])
    }
p.val = (sum(D.sim >= D)+1)/(B+1)
return (p.val)
}
```

Function helper to make Wage value to numeric ex. Messi Wage: €595K -> 595.000

cited from: https://towardsdatascience.com/exploratory-analysis-of-fifa-18-dataset-using-r-ba09aa4a2d3c

```
toNumberCurrency <- function(vector) {
    vector <- as.character(vector)
    vector <- gsub("(€|,)","", vector)
    result <- as.numeric(vector)

    k_positions <- grep("K", vector)
    result[k_positions] <- as.numeric(gsub("K","",vector[k_positions])) * 1000

    m_positions <- grep("M", vector)
    result[m_positions] <- as.numeric(gsub("M","",vector[m_positions])) * 1000000

    return(result)
}
```

Part B.

```
fifa = read.csv("~/Documents/Math185/fifa.csv")
fifa$Wage = toNumberCurrency(fifa$Wage)
```

```
## Warning in toNumberCurrency(fifa$Wage): NAs introduced by coercion
```

```
#first convert Wage to numerical value

under29 = fifa[fifa$Age <= 29,]
above30 = fifa[fifa$Age >= 30,]

bootGOFdiff(under29$Wage, above30$Wage, B=2000)
```

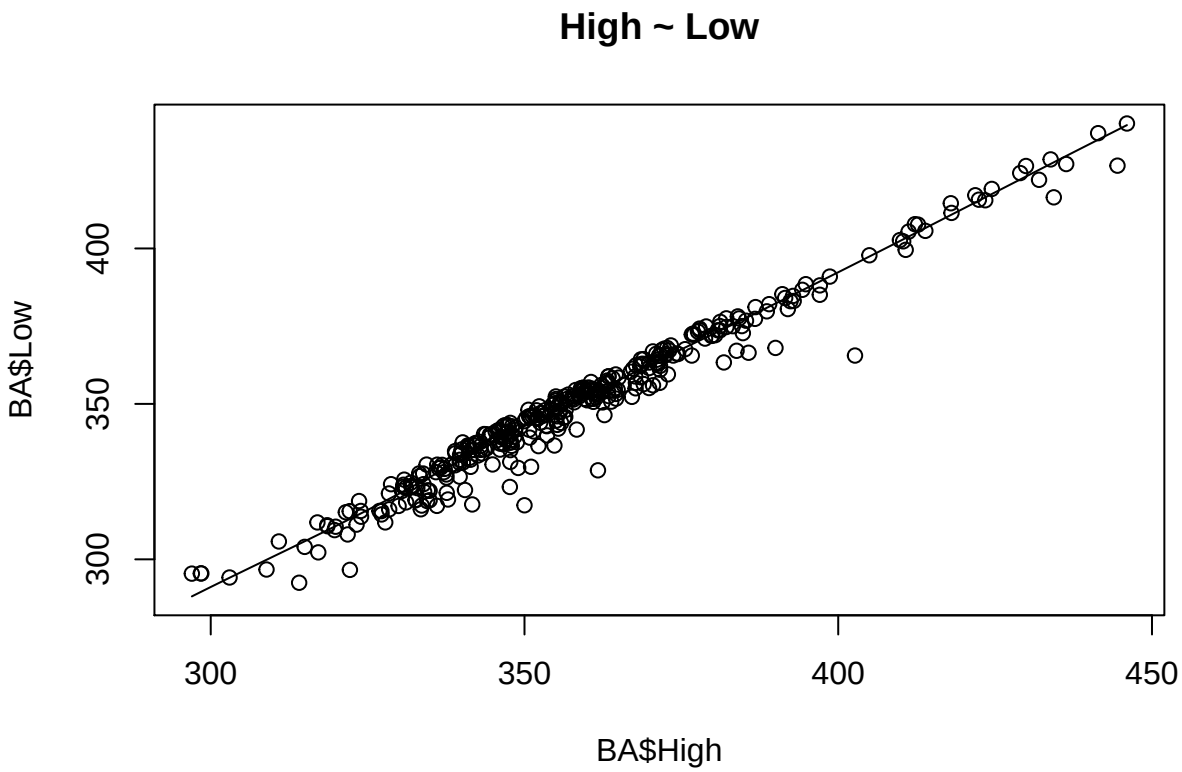
```
## [1] 0.0004997501
```

Problem 2.

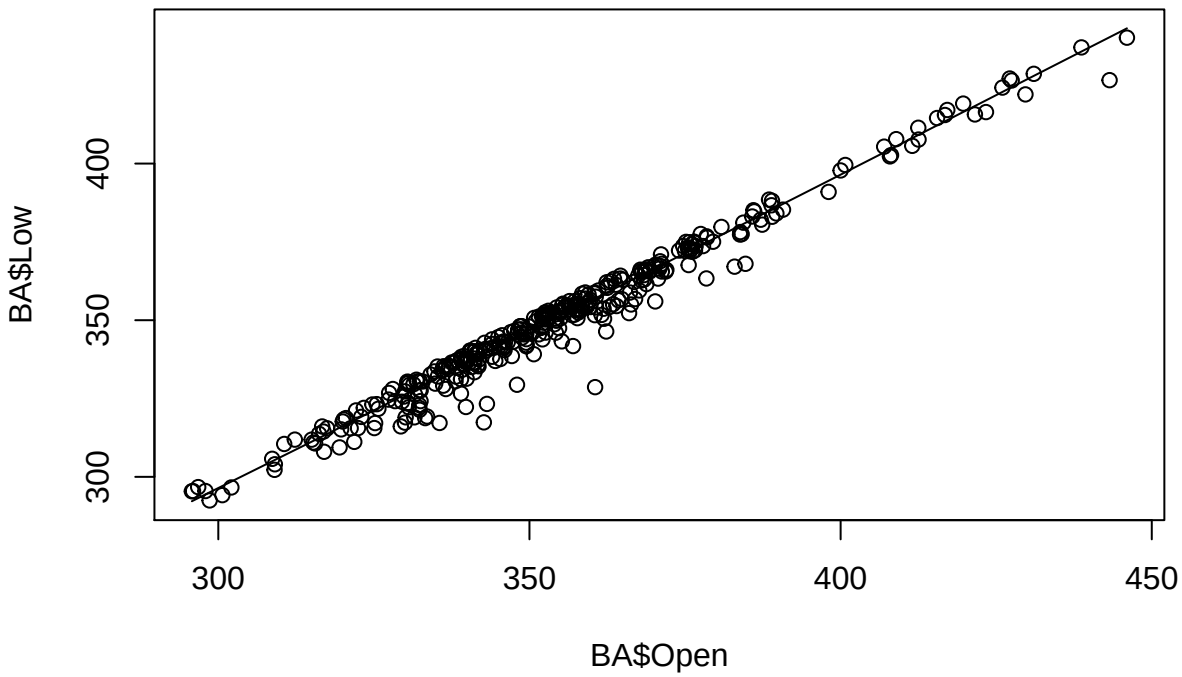
```
#kernSmooth & locfit
#localAbsLinearRegression(x, y, h, xnew = x) {

#}
```

Part B.

```
BA = read.csv("~/Documents/Math185/BA.csv")
#get rid of the date
BA = BA[,-1]

#a = localAbsLinearRegression(BA$High, BA$Low, 1, xnew = BA$High)

scatter.smooth(x=BA$High, y=BA$Low, main="High ~ Low")
```



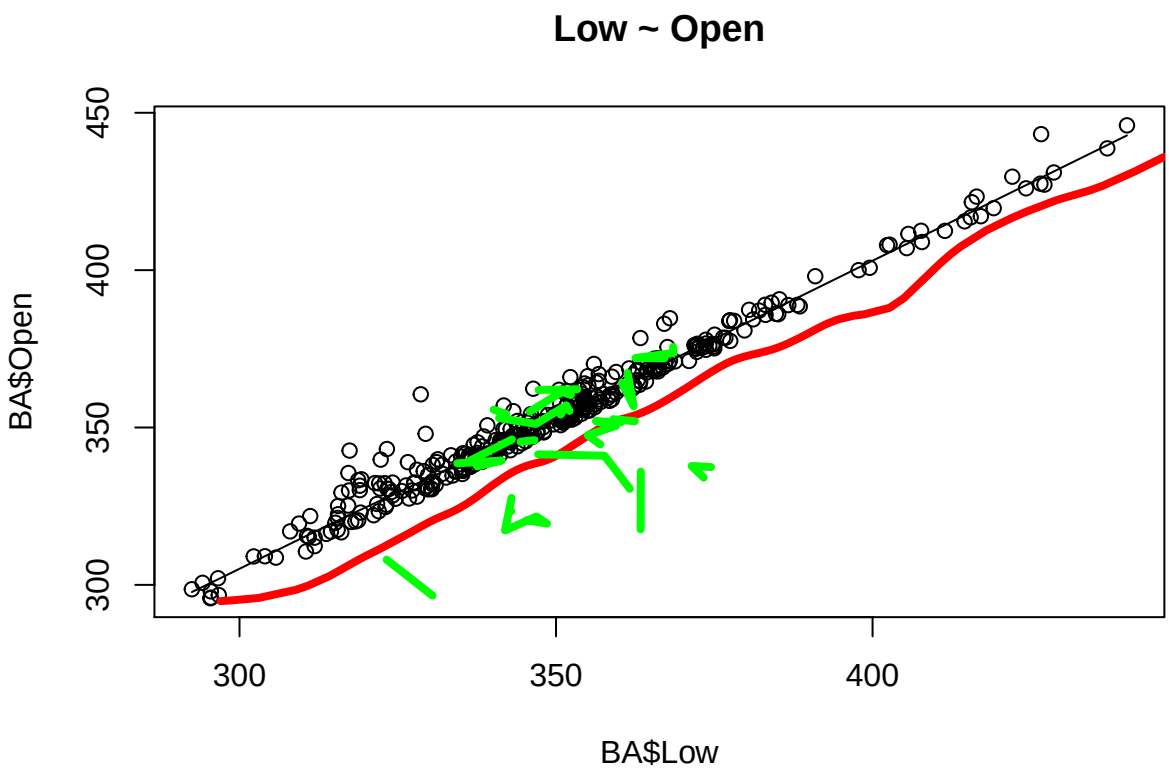

```
scatter.smooth(x=BA$Open, y=BA$Low, main="Open ~ Low")
```




```
scatter.smooth(x=BA$Low, y=BA$Open, main = "Low ~ Open")

fit = smooth.spline(BA$High, BA$Low)
lines(fit, col="red", lwd = 4)
fit$df  # degrees of freedom
```

```
## [1] 19.00843
```

```
# Kernel smoothing
h = 0.05
fit = ksmooth(BA$High, BA$Low, 'normal', bandwidth = h)
lines(BA$High, fit$y, col = 'green', lwd=4)
```



Part C.

```
#for(i in 1:10) {
#  a[i] = localAbsLinearRegression(BA$High, BA$Low, i, xnew = BA$Low)
#  a[i]
#}
```