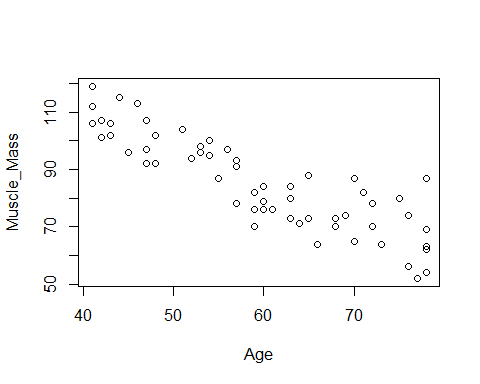
Chad Huntebrinker’s Homework 9

Chad Huntebrinker

2024-11-11

Problem 8.4 and 8.5

#Chad Huntebrinker  
  
#Load the library and data in to R  
library(readxl)  
excel\_data <- read\_excel("Muscle\_Mass\_Data.xlsx")  
excel\_data$Age.c <- excel\_data$Age-mean(excel\_data$Age)  
  
#Muscle\_Mass ~ Age  
#Fit the model  
#Problem 8.4a  
model\_1 <- lm(Muscle\_Mass~Age,data=excel\_data)  
sum\_of\_model\_1 <- summary(model\_1)  
plot(Muscle\_Mass~Age,data=excel\_data)



model\_2 <- lm(Muscle\_Mass~Age.c + I(Age.c^2),data=excel\_data)  
sum\_of\_model\_2 <- summary(model\_2)  
sum\_of\_model\_2

##   
## Call:  
## lm(formula = Muscle\_Mass ~ Age.c + I(Age.c^2), data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.086 -6.154 -1.088 6.220 20.578   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 82.935749 1.543146 53.745 <2e-16 \*\*\*  
## Age.c -1.183958 0.088633 -13.358 <2e-16 \*\*\*  
## I(Age.c^2) 0.014840 0.008357 1.776 0.0811 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.026 on 57 degrees of freedom  
## Multiple R-squared: 0.7632, Adjusted R-squared: 0.7549   
## F-statistic: 91.84 on 2 and 57 DF, p-value: < 2.2e-16

#The quadratic function seems like a good fit.  
  
#Problem 8.4b  
summary(model\_2)

##   
## Call:  
## lm(formula = Muscle\_Mass ~ Age.c + I(Age.c^2), data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.086 -6.154 -1.088 6.220 20.578   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 82.935749 1.543146 53.745 <2e-16 \*\*\*  
## Age.c -1.183958 0.088633 -13.358 <2e-16 \*\*\*  
## I(Age.c^2) 0.014840 0.008357 1.776 0.0811 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.026 on 57 degrees of freedom  
## Multiple R-squared: 0.7632, Adjusted R-squared: 0.7549   
## F-statistic: 91.84 on 2 and 57 DF, p-value: < 2.2e-16

qf(0.95, 2, 57)

## [1] 3.158843

#H0: B1 = B11 = 0, if F\_score <= 3.158843  
#Ha: not B1 and B11 = 0, if F\_score > 3.158843  
#F-statistic: 91.84 on 2 and 57 DF  
91.84 > 3.158843

## [1] TRUE

#Prove Ha is true.  
  
#Problem 8.4c  
predict(model\_2, data.frame(Age.c = 48 - mean(excel\_data$Age)),  
 interval="confidence", se.fit = TRUE, level = 0.95)

## $fit  
## fit lwr upr  
## 1 99.25461 96.28436 102.2249  
##   
## $se.fit  
## [1] 1.483295  
##   
## $df  
## [1] 57  
##   
## $residual.scale  
## [1] 8.025521

#Problem 8.4d  
predict(model\_2, data.frame(Age.c = 48 - mean(excel\_data$Age)),  
 interval="prediction", se.fit = TRUE, level = 0.95)

## $fit  
## fit lwr upr  
## 1 99.25461 82.9116 115.5976  
##   
## $se.fit  
## [1] 1.483295  
##   
## $df  
## [1] 57  
##   
## $residual.scale  
## [1] 8.025521

#Problem 8.4e  
sum\_of\_model\_2

##   
## Call:  
## lm(formula = Muscle\_Mass ~ Age.c + I(Age.c^2), data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.086 -6.154 -1.088 6.220 20.578   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 82.935749 1.543146 53.745 <2e-16 \*\*\*  
## Age.c -1.183958 0.088633 -13.358 <2e-16 \*\*\*  
## I(Age.c^2) 0.014840 0.008357 1.776 0.0811 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.026 on 57 degrees of freedom  
## Multiple R-squared: 0.7632, Adjusted R-squared: 0.7549   
## F-statistic: 91.84 on 2 and 57 DF, p-value: < 2.2e-16

qt(0.975, 57)

## [1] 2.002465

#t\* = 1.776  
#H0: B11 = 0 if t\* <= 2.002465  
#Ha: B11 != 0, if t\* > 2.002465  
1.776 > qt(0.975, 57)

## [1] FALSE

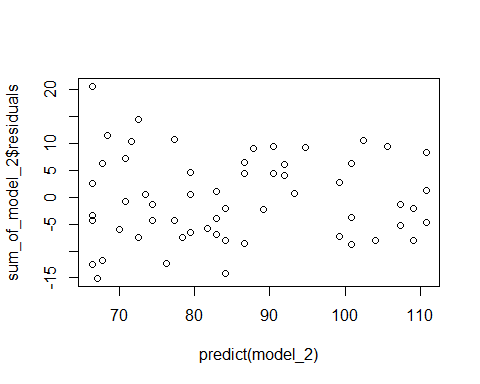
#We can conclude H0 is true  
  
#Problem 8.4f  
#Y = 207.350 − 2.96432X + .0148405X^2  
  
#Problem 8.4g  
cor(excel\_data$Age, excel\_data$Age^2)

## [1] 0.9960939

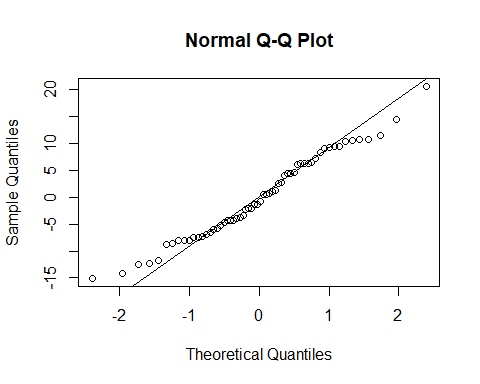
cor(excel\_data$Age.c, excel\_data$Age.c^2)

## [1] -0.03835694

#Yes, the use of a centered variable is helpful here.  
  
#Problem 8.5a  
plot(sum\_of\_model\_2$residuals~predict(model\_2))



qqnorm(residuals(model\_2))  
qqline(residuals(model\_2))



#Both of these plots seem pretty standard and no need for concern.  
  
#Problem 8.5b  
#H0: E{Y} = B0 + B1x + B11x^2, F\_score ≤ 1.875188  
#HA: E{Y} != B0 + B1x + B11x^2, F\_score > 1.875188  
reduced\_model <- lm(Muscle\_Mass~Age.c, data = excel\_data)  
model\_2\_residuals <- residuals(model\_2)  
reduced\_model\_residuals <- residuals(reduced\_model)  
  
#Calculate MSLF  
SST <- sum((excel\_data$Muscle\_Mass - mean(excel\_data$Muscle\_Mass))^2)  
  
SSR\_reduced <- sum((predict(reduced\_model) - mean(excel\_data$Muscle\_Mass))^2)  
  
SSLOF <- SST - SSR\_reduced  
  
dfLOF <- length(excel\_data$Muscle\_Mass) - length(coef(reduced\_model)) - 1  
  
MSLF <- SSLOF / dfLOF  
  
#Calculate MSPE  
  
MSPE <- mean((excel\_data$Muscle\_Mass - predict(model\_2))^2)  
  
F\_score <- MSLF / MSPE  
  
F\_score > qf(0.95, 29, 28)

## [1] FALSE

#Conclude H0  
  
#Problem 8.5c  
model\_3 <- lm(Muscle\_Mass~Age.c + I(Age.c^2) + I(Age.c^3), data = excel\_data)  
summary(model\_3)

##   
## Call:  
## lm(formula = Muscle\_Mass ~ Age.c + I(Age.c^2) + I(Age.c^3), data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.3671 -5.8483 -0.6755 6.1376 20.0637   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 82.9273444 1.5552264 53.322 < 2e-16 \*\*\*  
## Age.c -1.2678894 0.2489231 -5.093 4.28e-06 \*\*\*  
## I(Age.c^2) 0.0150390 0.0084390 1.782 0.0802 .   
## I(Age.c^3) 0.0003369 0.0009327 0.361 0.7193   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.087 on 56 degrees of freedom  
## Multiple R-squared: 0.7637, Adjusted R-squared: 0.7511   
## F-statistic: 60.34 on 3 and 56 DF, p-value: < 2.2e-16

#Y = = 82.9273 − 1.26789x + .01504x2 + .000337x3  
#t\_score = 0.361  
#H0: B111 = 0, t\_score <= 2.003241  
#Ha: B111 != 0, t\_score > 2.003241  
qt(0.975, 56)

## [1] 2.003241

0.361 <= qt(0.975, 56)

## [1] TRUE

#Thus, we can conclude H0.

Problem 8.15 and 8.19

#Chad Huntebrinker  
  
#Load the library and data in to R  
library(readxl)  
excel\_data <- read\_excel("Copier\_Maintenance\_Data.xlsx")  
  
#Problem 8.15a & b  
model\_1 <- lm(Total\_Minutes~Number\_Of\_Copiers + Type\_Of\_Copier,data=excel\_data)  
summary(model\_1)

##   
## Call:  
## lm(formula = Total\_Minutes ~ Number\_Of\_Copiers + Type\_Of\_Copier,   
## data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22.5390 -4.2515 0.5995 6.5995 14.9330   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.9225 3.0997 -0.298 0.767   
## Number\_Of\_Copiers 15.0461 0.4900 30.706 <2e-16 \*\*\*  
## Type\_Of\_Copier 0.7587 2.7799 0.273 0.786   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.011 on 42 degrees of freedom  
## Multiple R-squared: 0.9576, Adjusted R-squared: 0.9556   
## F-statistic: 473.9 on 2 and 42 DF, p-value: < 2.2e-16

#B0 is the intercept  
#B1 is one of the slopes, specifically related with the number of copiers  
#B2 is the other slopes, specifically related to the copier size.  
#Y = -0.9225 + 15.0461X1 + 0.7587X2  
  
#Problem 8.15c  
mean(excel\_data$Number\_Of\_Copiers)

## [1] 5.111111

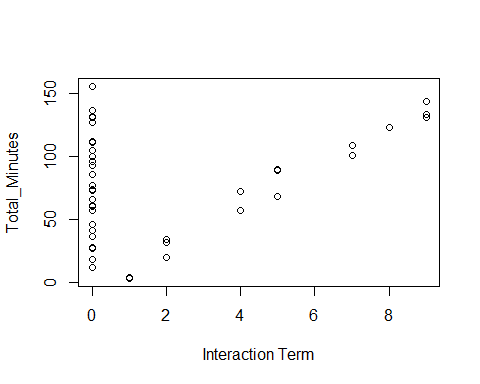
predict(model\_1, data.frame(Number\_Of\_Copiers = 5.111111, Type\_Of\_Copier = 0),  
 interval="confidence", se.fit = TRUE, level = 0.95)

## $fit  
## fit lwr upr  
## 1 75.98004 72.53907 79.421  
##   
## $se.fit  
## [1] 1.705069  
##   
## $df  
## [1] 42  
##   
## $residual.scale  
## [1] 9.011009

predict(model\_1, data.frame(Number\_Of\_Copiers = 5.111111, Type\_Of\_Copier = 1),  
 interval="confidence", se.fit = TRUE, level = 0.95)

## $fit  
## fit lwr upr  
## 1 76.73876 72.3191 81.15842  
##   
## $se.fit  
## [1] 2.190029  
##   
## $df  
## [1] 42  
##   
## $residual.scale  
## [1] 9.011009

#Problem 8.15d  
#We would be interested in X1 because the service time is still impacted by the  
#number of copiers even if we are interested in the effect of the type of copier (2 smaller  
#copiers might have a similiar amount of service time as 1 larger one, for example).  
  
#Problem 8.15e  
temp\_values <- excel\_data$Number\_Of\_Copiers \* excel\_data$Type\_Of\_Copier  
plot(Total\_Minutes~temp\_values, xlab = "Interaction Term", data = excel\_data)



#Yes, there seems to be a pattern so a interaction term would be helpful.  
  
#Problem 8.19a  
model\_2 <- lm(Total\_Minutes~Number\_Of\_Copiers + Type\_Of\_Copier +  
 Number\_Of\_Copiers\*Type\_Of\_Copier, data = excel\_data)  
summary(model\_2)

##   
## Call:  
## lm(formula = Total\_Minutes ~ Number\_Of\_Copiers + Type\_Of\_Copier +   
## Number\_Of\_Copiers \* Type\_Of\_Copier, data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.2072 -6.7887 -0.1708 7.1504 14.7441   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.8131 3.6468 0.771 0.4449   
## Number\_Of\_Copiers 14.3394 0.6146 23.333 <2e-16 \*\*\*  
## Type\_Of\_Copier -8.1412 5.5801 -1.459 0.1522   
## Number\_Of\_Copiers:Type\_Of\_Copier 1.7774 0.9746 1.824 0.0755 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.771 on 41 degrees of freedom  
## Multiple R-squared: 0.9608, Adjusted R-squared: 0.9579   
## F-statistic: 334.6 on 3 and 41 DF, p-value: < 2.2e-16

#Y = 2.8131 + 14.3394X1 - 8.1412X2 + 1.7774X1X2  
  
#Problem 8.19b  
#H0: B3 = 0, t\_value <= qt(0.95, 41)  
#HA: B3 != 0, t\_value > qt(0.95, 41)  
summary(model\_2)

##   
## Call:  
## lm(formula = Total\_Minutes ~ Number\_Of\_Copiers + Type\_Of\_Copier +   
## Number\_Of\_Copiers \* Type\_Of\_Copier, data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -19.2072 -6.7887 -0.1708 7.1504 14.7441   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.8131 3.6468 0.771 0.4449   
## Number\_Of\_Copiers 14.3394 0.6146 23.333 <2e-16 \*\*\*  
## Type\_Of\_Copier -8.1412 5.5801 -1.459 0.1522   
## Number\_Of\_Copiers:Type\_Of\_Copier 1.7774 0.9746 1.824 0.0755 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.771 on 41 degrees of freedom  
## Multiple R-squared: 0.9608, Adjusted R-squared: 0.9579   
## F-statistic: 334.6 on 3 and 41 DF, p-value: < 2.2e-16

#t\_value = 1.824  
#p\_value = 0.0755  
  
qt(0.95, 41)

## [1] 1.682878

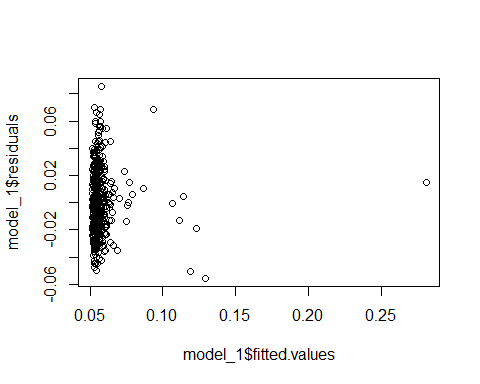
#Test H0  
1.824 <= qt(0.95, 41)

## [1] FALSE

#Reject H0, B3 != 0, keep it in the model.

Problem 8.37

#Chad Huntebrinker  
  
#Load the library and data in to R  
library(readxl)  
excel\_data <- read\_excel("CDI\_Data.xlsx")  
  
#Problem 8.37  
excel\_data$Pop\_Density <- excel\_data$Total\_Pop / excel\_data$Land\_Area  
Crime\_Rate <- excel\_data$Serious\_Crimes / excel\_data$Total\_Pop  
Pop\_Density.c <- excel\_data$Pop\_Density - mean(excel\_data$Pop\_Density)  
Unemployment.c <- excel\_data$Unemployment - mean(excel\_data$Unemployment)  
  
model\_1 <- lm(Crime\_Rate~Pop\_Density.c + Unemployment.c + I(Pop\_Density.c^2) +  
 I(Unemployment.c^2) + Pop\_Density.c\*Unemployment.c)  
sum\_of\_model\_1 <- summary(model\_1)  
  
plot(model\_1$residuals~model\_1$fitted.values)



sum\_of\_model\_1

##   
## Call:  
## lm(formula = Crime\_Rate ~ Pop\_Density.c + Unemployment.c + I(Pop\_Density.c^2) +   
## I(Unemployment.c^2) + Pop\_Density.c \* Unemployment.c)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.055642 -0.016851 -0.002889 0.014810 0.085485   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.629e-02 1.260e-03 44.662 < 2e-16 \*\*\*  
## Pop\_Density.c 4.585e-06 9.841e-07 4.659 4.23e-06 \*\*\*  
## Unemployment.c -8.800e-05 6.276e-04 -0.140 0.8886   
## I(Pop\_Density.c^2) 2.698e-12 5.932e-11 0.045 0.9637   
## I(Unemployment.c^2) 1.629e-04 9.541e-05 1.708 0.0884 .   
## Pop\_Density.c:Unemployment.c 8.334e-07 4.091e-07 2.037 0.0423 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.02383 on 434 degrees of freedom  
## Multiple R-squared: 0.2485, Adjusted R-squared: 0.2398   
## F-statistic: 28.7 on 5 and 434 DF, p-value: < 2.2e-16

#Multiple R-squared: 0.2485, Adjusted R-squared: 0.2398  
#The second-order model appears to fit the data pretty well, although there are some outliers  
#in the regression function  
  
#Problem 8.37b  
#H0: p\_value > 0.01, drop the quadratic and interaction terms  
#Ha: p\_value < 0.01, keep the quadratic and interaction terms  
reduced\_model <- lm(Crime\_Rate~Pop\_Density.c + Unemployment.c)  
anova(reduced\_model, model\_1)

## Analysis of Variance Table  
##   
## Model 1: Crime\_Rate ~ Pop\_Density.c + Unemployment.c  
## Model 2: Crime\_Rate ~ Pop\_Density.c + Unemployment.c + I(Pop\_Density.c^2) +   
## I(Unemployment.c^2) + Pop\_Density.c \* Unemployment.c  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 437 0.25186   
## 2 434 0.24638 3 0.005477 3.2159 0.02278 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#p\_value = 0.02278 > 0.01  
0.02278 > 0.01

## [1] TRUE

#H0 is true, so drop those terms  
  
#Problem 8.37c  
#Problem doesn't specify whether these variables need to be centered or not.  
#So, I will do both.  
model\_2 <- lm(Crime\_Rate~Total\_Pop + Land\_Area + Unemployment + I(Total\_Pop^2),  
 data = excel\_data)  
summary(model\_2)

##   
## Call:  
## lm(formula = Crime\_Rate ~ Total\_Pop + Land\_Area + Unemployment +   
## I(Total\_Pop^2), data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.05967 -0.01704 -0.00303 0.01410 0.19106   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.250e-02 3.889e-03 10.927 < 2e-16 \*\*\*  
## Total\_Pop 3.206e-08 3.941e-09 8.133 4.41e-15 \*\*\*  
## Land\_Area -5.576e-07 8.123e-07 -0.687 0.493   
## Unemployment 6.824e-04 5.302e-04 1.287 0.199   
## I(Total\_Pop^2) -3.356e-15 5.878e-16 -5.710 2.10e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.02539 on 435 degrees of freedom  
## Multiple R-squared: 0.1444, Adjusted R-squared: 0.1365   
## F-statistic: 18.35 on 4 and 435 DF, p-value: 6.022e-14

#Multiple R-squared: 0.1444, Adjusted R-squared: 0.1365  
  
Total\_Pop.c <- excel\_data$Total\_Pop - mean(excel\_data$Total\_Pop)  
Land\_Area.c <- excel\_data$Land\_Area - mean(excel\_data$Land\_Area)  
  
model\_3 <- lm(Crime\_Rate~Total\_Pop.c + Land\_Area.c + Unemployment.c + I(Total\_Pop.c^2))  
summary(model\_3)

##   
## Call:  
## lm(formula = Crime\_Rate ~ Total\_Pop.c + Land\_Area.c + Unemployment.c +   
## I(Total\_Pop.c^2))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.05967 -0.01704 -0.00303 0.01410 0.19106   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.850e-02 1.229e-03 47.595 < 2e-16 \*\*\*  
## Total\_Pop.c 2.942e-08 3.555e-09 8.276 1.57e-15 \*\*\*  
## Land\_Area.c -5.576e-07 8.123e-07 -0.687 0.493   
## Unemployment.c 6.824e-04 5.302e-04 1.287 0.199   
## I(Total\_Pop.c^2) -3.356e-15 5.878e-16 -5.710 2.10e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.02539 on 435 degrees of freedom  
## Multiple R-squared: 0.1444, Adjusted R-squared: 0.1365   
## F-statistic: 18.35 on 4 and 435 DF, p-value: 6.022e-14

#Multiple R-squared: 0.1444, Adjusted R-squared: 0.1365  
  
#While the coefficients are different, they both are on the lower end of the spectrum.