```r
# Set Working directory
setwd('/home/noble_mannu/Documents/PhD/First/
STAT_2131_Applied_Statistical_Methods_I/HW6')

### Exercise 4 6.10 ###

## Part (a) ##

# Read the table with the information of the data
data <- read.delim('CH06PR09.txt', sep = '', dec = '.', header = FALSE)
names(data) <- c('Labor_Hours','Cases_Shipped','Indirect_Costs','Holiday')

# Make the multilinear regression model
linearMod = lm(formula = Labor_Hours~Cases_Shipped+Indirect_Costs+Holiday, data =
data)
# Display summary of our model
summary(linearMod)

# Obtain the regression coefficients
beta_0 = coef(linearMod)[1]
beta_1 = coef(linearMod)[2]
beta_2 = coef(linearMod)[3]
beta_3 = coef(linearMod)[4]

## Part (c) ##

# Obtain the residuals of the linear model
linearMod_res <- resid(linearMod) # Or just linearMod$residuals

# Plot the residuals against the fitted values
plot(linearMod$fitted.values, resid(linearMod), xlab = "Fitted_Values", ylab =
"Residuals",
     main = 'Residual plot against Y_Hat', col = 'blue')
abline(a=0, b=0)

# Here I separated the 2 clusters and computed the variance of each one
cluster1 = linearMod$fitted.values[which(linearMod$fitted.values < 4600)]
var(cluster1)
cluster2 = linearMod$fitted.values[which(linearMod$fitted.values > 4600)]
var(cluster2)

# PLot the residuals against X1
plot(data$Cases_Shipped, resid(linearMod), xlab = "Number of cases shipped", ylab =
"Residuals",
     main = 'Residual plot against X1', col = 'blue')
abline(a=0, b=0)

# PLot the residuals against X2
plot(data$Indirect_Costs, resid(linearMod), xlab = "Indirect cost", ylab =
"Residuals",
     main = 'Residual plot against X2', col = 'blue')
abline(a=0, b=0)

# PLot the residuals against X3
plot(data$Holiday, resid(linearMod), xlab = "Week with holiday", ylab =
"Residuals",
     main = 'Residual plot against X3', col = 'blue')
abline(a=0, b=0)

# PLot the residuals against X1X2
```

```
    ylab = "Estimated residual quantiles", col = 'red')
qqline(linearMod$residuals)

### Exercise 4 7.4 ###

## Part (b) ##

# Make the multilinear regression model (full and reduced models)
mod1 <- lm(formula = Labor_Hours ~ Cases_Shipped+Indirect_Costs+Holiday, data =
data)
summary(mod1)
mod1_red <- lm(formula = Labor_Hours ~ Cases_Shipped+Holiday, data = data)
summary(mod1_red)

# Using an ANOVA table
anova(mod1,mod1_red)

# For the  p-value
qf(.95, df1 = 1, df2 = length(data$Labor_Hours)-4)
1 - pf(0.3251, df1 = 1, df2 = length(data$Labor_Hours)-4)

### Exercise 4 6.16 ###

## Part (a) ##

# Read the data
data2 <- read.delim('CH06PR15.txt', sep = '', dec = '.', header = FALSE)
names(data2) <- c('Satisfaction','Age','Severity_Illness','Anxiety_Level')

# Regressing Satisfaction (Y) onto Age (X1), Severity of Illness (X2) and Anxiety
Level (FULL MODEL)
m1 <- lm(Satisfaction ~ Age+Severity_Illness+Anxiety_Level, data = data2)
summary(m1)

# Making the reduced model with coefficients equal to 0, except beta_0 (REDUCED
MODEL)
m1_reduced <- lm(Satisfaction ~ 1, data = data2)
summary(m1_reduced)

# Conducting the F-test I did this in two ways

# First way, doing all the computations separately
x <- m1$fitted.values - mean(data2$Satisfaction)
SSR <- x%*%x
MSR <- SSR/(4-1)
y <- m1$fitted.values - data2$Satisfaction
SSE <- y%*%y
MSE <- SSE/(length(data2$Satisfaction)-4)
f <- MSR/MSE

# Second way, using an ANOVA table
anova(m1,m1_reduced)

# For the p-value
qf(.90, df1 = 4-1, df2 = length(data2$Satisfaction)-4)
1 - pf(f, df1 = 4-1, df2 = length(data2$Satisfaction)-4)

## Part (b) ##

b_1 <- coefficients(m1)[2]
```

```
sd_b_2 <- 0.4920
sd_b_3 <- 7.0997
n <- length(data2$Satisfaction)
p <- 4
g <- 3
alpha <- 0.1
t <- qt(1 - alpha/(2*g), df=n-p)
# For beta1
b_1 - t*sd_b_1
b_1 + t*sd_b_1
# For beta2
b_2 - t*sd_b_2
b_2 + t*sd_b_2
# For beta3
b_3 - t*sd_b_3
b_3 + t*sd_b_3

## Part (c) ##

# We computed SSR and SSE in part (a) and we know SSTO = SSR+SSE
SSTO <- SSR + SSE
# We could also use the definition of SSTO
z <- data2$Satisfaction - mean(data2$Satisfaction)
SSTO1 <- z%*%z
# Computing the Coefficient of multiple determination
R_squared <- SSR/SSTO
R_squared1 <- 1 - SSE/SSTO
```