**Steps to run mall sales regression analysis:**

Step 1: You need to copy mallsales data file from eLearning and copy it in your computer

Step 2: Set the directory of your R studio using the command below to set your working directory . Please note that you have to keep your data file in Working folder.

> setwd("C:/your\_folder")

Step3 Import Data file into Rstudio

> mallsales2 <- read.delim("C:/your\_folder")

Step 4: Run the following commands to see the results of regression

> View(mallsales2)

////Linear Regression lm(Y ~ X1 + X2, datasetname)

> m1 <- lm(Sales ~ Comp, mallsales2)

> summary(m1)

> m2 <- lm(Sales ~ Comp + Income, mallsales2)

> summary(m2)

**BMI Example**

Step1: Import Data file into Rstudio

Step 2 : Create dummy variable for education

df$Dlow <- ifelse(df$education2 == 0, 1,0)

df$Dmed <- ifelse(df$education2 == 1, 1,0)

df$Dhigh <-ifelse(df$education2 == 2, 1,0)

Step 2 : Create dummy variable for male

df$male <- 1-df$female

Q1: Is BMI higher for males than females? Does BMI increases with age ?

m1<-lm(BMI~female +age ,df)

summary(m1)

Q1a : what is average BMI for males and females?

What happens if we run following model?

m2<-lm(BMI~male ,df)

summary(m2)

Q2 As we go from (a) low to med and (b) low to higher education level , does BMI increases?

m3<-lm(BMI~female + Dmed + Dhigh ,df)

summary(m3)

Q3: As we go from medium to higher does BMI increase?

**Approach 1**: M3<-lm(BMI~female + Dmed + Dhigh ,df)

install.packages("car")

library("car")

linearHypothesis(m3," Dmed= Dhigh")

**Approach 2**:

m4<-lm(BMI~female + Dmed + Dlow ,df)

summary(m4)

**Question: We want to study whether the effect of going from lower education level to higher education level is higher for females than males. How can I achieve this?**

**We need interaction terms.**

Run the following regression

**BMI= a + b female + c Dmed + d Dhigh + e female\*Dmed + f Dfemale \* Dhigh**

**Interpretation**

**BMIlow,male = a**

**BMImed,male = a + c**

**BMIhigh,male = a + d**

**BMIlow,female = a + b**

**BMImed,female = a + b+ c + e**

**BMIhigh,female = a + b+ d + f**

**a = BMIlow,male**

**b= BMIlow,male - BMIlow,male**

**c= BMImed,male – BMIlow,male**

**d= BMIhigh,male – BMIlow,male**

**e= (BMImed,female – BMIlow,female ) –  (BMImed,male – BMIlow,male )**

**f= (BMIhigh,female – BMIlow,female ) –  (BMIhigh,male – BMIlow,male )**

df$Dhf <- df$Dhigh\*df$female

df$Dmf <- df$Dmed\*df$female

m5<-lm(BMI~female + Dmed + Dhigh + Dmf +Dhf,df)

summary(m5)

**Based on the results of above model, answer following questions?**

**Question: As you go from medium to high, what is the change in BMI for females? Is the change significant?**

**Question: As you go from medium to high, is the change in BMI higher for females than males? Is the difference significant?**

**What is the effect of age and gender on BMI?**

**BMI= a + b female + c age**

m6<-lm(BMI~female + age ,df)

summary(m6)

**What is the interpretation of c?**

**How does the effect of age changes with gender?**

**BMI= a + b female + c age + d age\* female**

df$fage <- df$age\*df$female

m7<-lm(BMI~female + age + fage ,df)

summary(m7)

**Please do following questions from BMI exercise:  1, 2, 12, 13, 14 and 16 in class?**