**Statistical and Predictive Modeling I (DATA 1204 - 02)**

**Assignment– Final Project**

**Name – Maharshi Mehta**

**Student Number-100799054**

**Professor: Ritwick Dutta.**

R code:

install.packages("psych")

library(psych)

library(ggplot2)

View(MultiRegDataset)

data\_new <- MultiRegDataset

head(data\_new)

str(data\_new)

#General statistics

describe(data\_new)

#histogram of expenses

x = data\_new$expenses

h<-hist(x, breaks=10, col="red", xlab="Expenses",

main="Household Studies")

xfit<-seq(min(x),max(x),length=40)

yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))

yfit <- yfit\*diff(h$mids[1:2])\*length(x)

lines(xfit, yfit, col="blue", lwd=2)

#The T-Test

#How the T-Test performing

t.test(x=data\_new$expenses, mu = 15000, alternative = "two.sided")

#Linear Model

modelData <- lm(smoker~expenses, data = data\_new)

linearModel<-modelData

summary(modelData)

#test for multivariate linear regression model

#Create Dataset

input <- data\_new[,c("expenses","age","sex","bmi","children","smoker","region")]

print(head(input))

#create the relationship model

model<-lm(expenses~age+sex+bmi+children+smoker+region, data = input)

#summary

summary(model)

**Thank you**