Copy the Comments below into an R script. . Run your Script and copy your code solution as illustrate as directed below in item A-4. Upload you project in the M5 Project assignment in Module 6.

**#REFERENCES**

#Ref: https://www.statmethods.net/advgraphs/ggplot2.html

#Ref: https://www.datanovia.com/en/blog/ggplot-point-shapes-best-tips/

#Ref: http://environmentalcomputing.net/plotting-with-ggplot-colours-and-symbols/

#Ref: Andy Fields: Discovering Statistics Using R

#Ref: R Kabacoff: R in Action

**#DIRECTIONS**

#After each numbered item, Copy the comment and then your solution onto THIS Project Specification Sheet. #See Item A-4 as an example.

ls()

rm(list = ls())

#library

#install.packages("quantreg")

library(ggplot2)

library(quantreg)

##library end

#A-1 LOAD DATA SET: <. Album Sales 2.dat > & display the first 6 records

dataA2 <-read.table("AlbumSales2.dat", header = TRUE)

head(dataA2)

> head(dataA2)

adverts sales airplay attract

1 10.256 330 43 10

2 985.685 120 28 7

3 1445.563 360 35 7

4 1188.193 270 33 7

5 574.513 220 44 5

6 568.954 170 19 5

#A-2 EXPLORE the DATA SET: List the names of the 4 variables then display the dimensions of the dataset. Finally display the basics quartile statistics of the 4 variables,

colnames(dataA2)

>colnames(dataA2)

[1] "adverts" "sales" "airplay" "attract"

summary(dataA2)

> summary(dataA2)

adverts sales airplay attract

Min. : 9.104 Min. : 10.0 Min. : 0.00 Min. : 1.00

1st Qu.: 215.918 1st Qu.:137.5 1st Qu.:19.75 1st Qu.: 6.00

Median : 531.916 Median :200.0 Median :28.00 Median : 7.00

Mean : 614.412 Mean :193.2 Mean :27.50 Mean : 6.77

3rd Qu.: 911.226 3rd Qu.:250.0 3rd Qu.:36.00 3rd Qu.: 8.00

Max. :2271.860 Max. :360.0 Max. :63.00 Max. :10.00

#A-3 CREATE the LINEAR REGRESSION MODEL of Sales vs Advertisements. Save the model as an R object named < albumSales.3ad >

albumSales.3ad <- ggplot(data = dataA2[2:1], aes(x = adverts, y = sales)) + geom\_point() + geom\_quantile(formula = y~x ,quantiles = 0.5)

albumSales.3ad

**#**A-4 What is the CORRELATION COEFFICIENT between Advertisements and Album sales. Save the value as an object named < r >

**r <- cor(sales, adverts)**

**> r**

**[1] 0.5784877**

#A-5 Display the 3 most basic DESCRIPTIVE STATISTICSD for Sales and Advertisements

sapply(dataA2[2:1], mean)

sapply(dataA2[2:1], sd)

sapply(dataA2[2:1], var)

> sapply(dataA2[2:1], mean)

sales adverts

193.2000 614.4123

> sapply(dataA2[2:1], sd)

sales adverts

80.69896 485.65521

> sapply(dataA2[2:1], var)

sales adverts

6512.322 235860.981

#A-6 What is the COEFFICIENT OF DETERMINATION of Advertisements vs Sales

numad <- dataA2[["adverts"]]

numsa <- dataA2[["sales"]]

summary(lm(numsa ~ numad))$r.squared

> summary(lm(numsa ~ numad))$r.squared

[1] 0.3346481

#A-7 Compute Total Sum of Squares from a mathematics basic formula

tsssa <- sum((numsa-mean(numsa))^2)

> tsssa

[1] 1295952

#A-8 Compute the Residual Sum of Squares from a mathematical formula

coeff <- lm(formula = numsa ~ numad)$coefficients

predd <- numad \* coeff[2] + coeff[1]

rsssa <- sum((numsa-predd)^2)

> rsssa

[1] 862264.2

#A-9 Display the intercept and slope of the regression line of Sales vs Advertisements

coeff <- lm(formula = numsa ~ numad)$coefficients

coeff[1]

coeff[2]

> coeff[1]

(Intercept)

134.1399

> coeff[2]

numad

0.09612449

#A10 Create the regression line equation (eg y = a + bx) and display the y coordinate at x = 500

ggplot(data.frame(numad, predd),aes(x = numad))+

geom\_line(aes(y=predd))+

geom\_point(aes(x= c(500),y=c(500 \* coeff[2] + coeff[1])),color= "red")

#A-11 Create a plot identical to that which is shown: REPLACE THIS IMAGE WITH YOUR PLOT ANNOTATED WITH **YOUR NAME** AND CHANGE THE REGRESSION LINE TO **RED** AND THE ORANGE TRIANGLE TO **BLUE** Replace the Professor's last name with your own.

m <- ggplot(data.frame(numad, predd,numsa),aes(x = numad))+

geom\_point(aes(y=numsa), fill = "lightgray")+

geom\_smooth(aes(y = numsa),formula = y ~ x, method = lm, color = "red")+

geom\_point(aes(x= c(500),y=c(500 \* coeff[2] + coeff[1])),color= "blue",size = 3)

m <- m + labs(x = "Advertising Expense($000)",

y = "Album Sales(000 units)",

title = "Scatter Plot",

subtitle = "Sales vs Advertising",

caption = "Wanqin Su") +

theme(plot.title = element\_text(hjust = 0.5),

plot.subtitle = element\_text(hjust = 0.5),

plot.caption = element\_text(color = "red",face = "italic"))

m

