Statistical Analysis (Lab Sheet 10.1)
Index No: - 18001149

1. Load the data set "cats" in the package "MASS". This data set contains Body weight (Bwt) and Heart weight (Hwt) of 144 domestic cats.

a) Check whether there is any relationship between Bwt and Hwt, by using a suitable graphical method. Comment on your plot.

b) Fit a simple linear regression to model the Heart weight of the cats using Body weight as the independent variable. Write the equation of the fitted model.

c) Test the goodness of fit of the fitted model and justify your answer.

d) Test the significance of the relationship between the variables Bwt and Hwt. (Mention all the steps clearly).

e) Predict the heart weight of a cat if its body weight is 2.52 kg.

```
beav1                   Body Temperature Series of Beaver 1
beav2                   Body Temperature Series of Beaver 2
biopsy                  Biopsy Data on Breast Cancer Patients
birthwt                 Risk Factors Associated with Low Infant Birth
                        Weight
cabbages                Data from a cabbage field trial
caith                   Colours of Eyes and Hair of People in
                        Caithness
cats                    Anatomical Data from Domestic Cats
cement                  Heat Evolved by Setting Cements
chem                    Copper in Wholemeal Flour
coop                    Co-operative Trial in Analytical Chemistry
cpus                    Performance of Computer CPUs
crabs                   Morphological Measurements on Leptograpsus
                        Crabs
deaths                  Monthly Deaths from Lung Diseases in the UK
drivers                 Deaths of Car Drivers in Great Britain
                        1969-84
eagles                  Foraging Ecology of Bald Eagles
epil                    Seizure Counts for Epileptics
farms                   Ecological Factors in Farm Management
fgl                     Measurements of Forensic Glass Fragments
forbes                  Forbes' Data on Boiling Points in the Alps
galaxies                Velocities for 82 Galaxies
gehan                   Remission Times of Leukaemia Patients
genotype                Rat Genotype Data
geyser                  Old Faithful Geyser Data
gilgais                 Line Transect of Soil in Gilgai Territory
hills                   Record Times in Scottish Hill Races
housing                 Frequency Table from a Copenhagen Housing
```

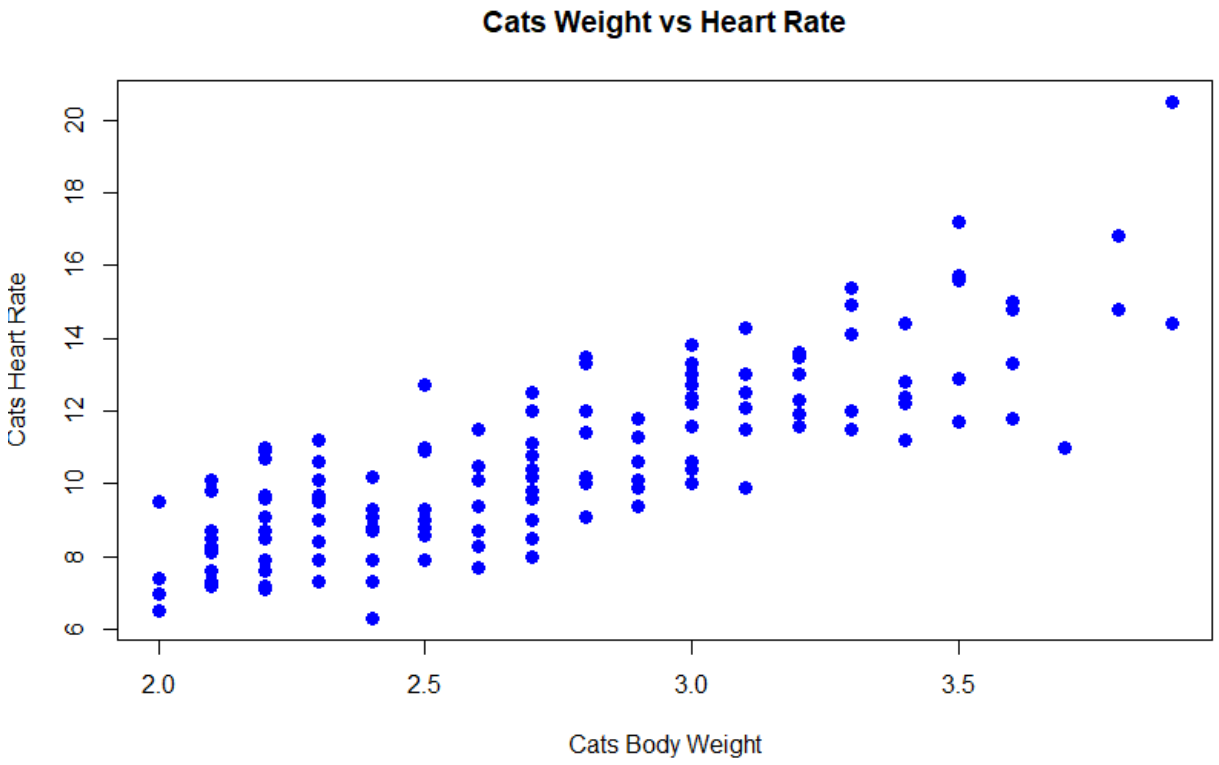
Console | Terminal | Jobs

C:/Local Disk D/Ucsc/Second Year/2nd sem/SCS2211 Laboratory II/tutorial/R tute5/

```
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> setwd("C:/Local Disk D/Ucsc/Second Year/2nd sem/SCS2211 Laboratory II/tutorial
R tute5")
> library('datasets')
> library(datasets)
> library(MASS)
warning message:
package 'MASS' was built under R version 4.0.4
> data()
```

a)

```
> bwt=cats[,2]
> hwt<-cats[,3]
> catone<-data.frame(bwt,hwt)
> plot(catone ,col="blue",pch=16,cex =1.3, main = "Cats Weight vs Heart Rate",xla
b = "Cats Body Weight",ylab = "Cats Heart Rate")
>
```



b)

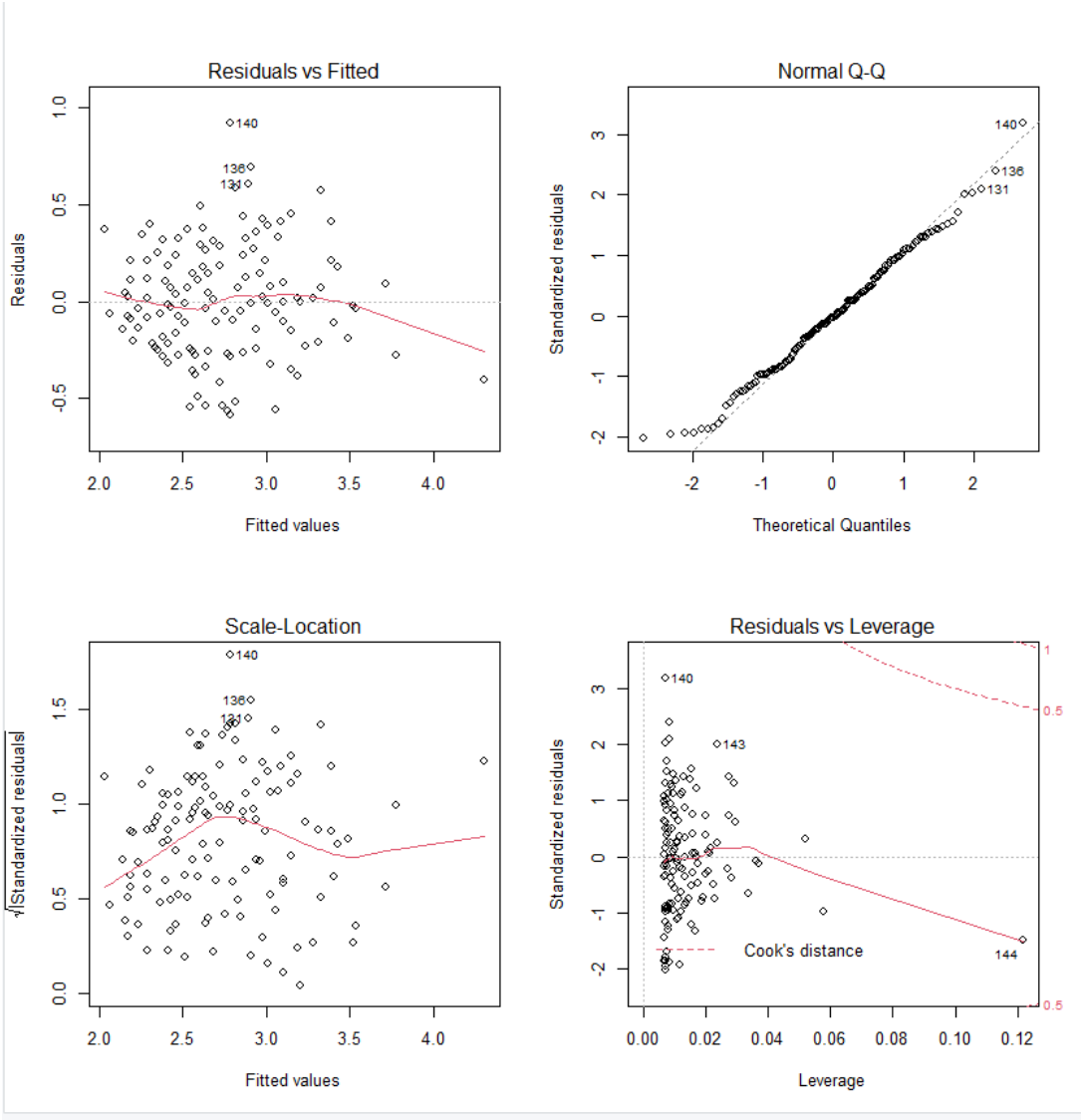
```
> lm(hwt~bwt,data=cats)

Call:
lm(formula = hwt ~ bwt, data = cats)

Coefficients:
(Intercept)          bwt
    -0.3567       4.0341
```

So, the Formula of This Plot is:

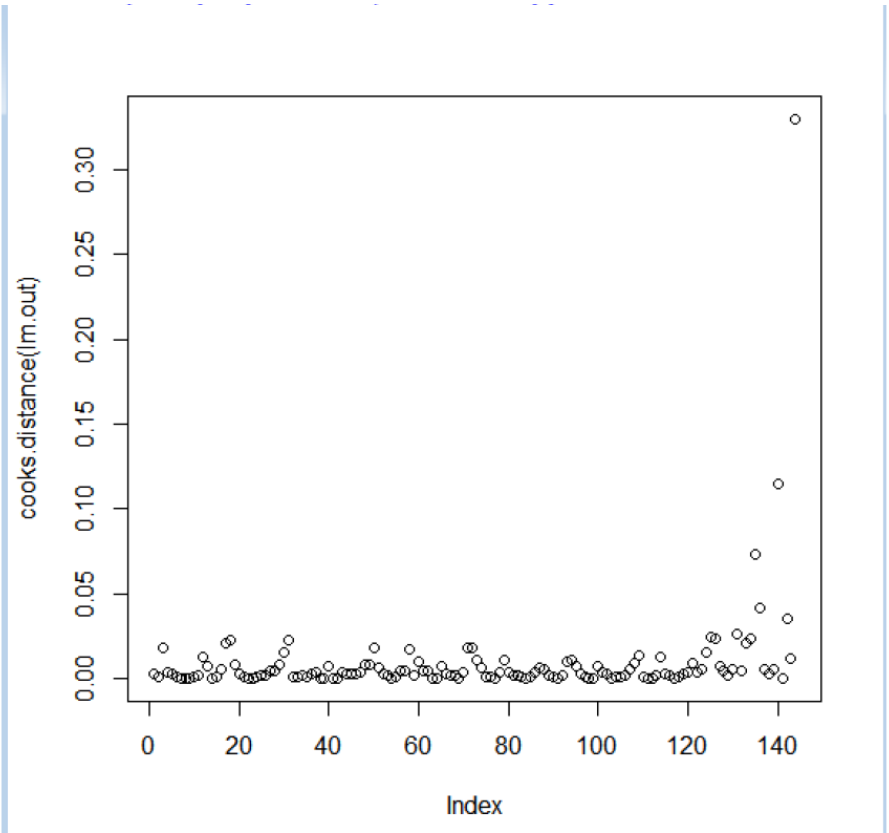
Hwt = 4.0341(Bwt) – 0.3567;

C)

```
> abline(lm(bwt~hwt,data=cats))
> par(mfrow=c(2,2))
> plot(lm(bwt~hwt,data=cats))
>
```



According to this diagram most fitted diagram can be identified as the **“cook’s Distance”.**



d)

e) Predict the heart weight of a cat if its body weight is 2.52 kg.

```
> predict.lm(lm.out,x0,interval = "prediction",cof.level=0.95)
        fit      lwr      upr
1  2.141668 1.562965 2.720372
```

2) Load the data set "iris" in the package "datasets". This data set gives the measurements in centimeters of the variable's sepal length and width and petal length and width and the species for 50 flowers from each of 3 species of iris.
Construct a matrix scatter plot for the data and interpret the relationships among variables.

```
> library(datasets)
> str(iris)
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

 # Correlation panel

```
> panel.cor<-function(x,y)
+ {
+   user<-par("usr");
+   on.exit(par(usr))
+   par(usr=c(0,1,0,1))
+
+   r<-round(cor(x,y),digits = 2)
+   txt<-paste0("R =",r)
+   cex.cor<-0.8/strwidth(txt)
+   text(0.5,0.5,txt,cex=cex.cor*r)
+ }
>
```

 # Customize upper panel

```
> upper.panel<-function(x,y)
+ {
+   points(x,y,pch=19,col=c("blue","green","red"))[iris$Species]
+   r<-round(cor(x,y),digits = 2)
+   txt<-paste0("R= ",r)
+   usr<-par("usr");
+   on.exit(par(usr))
+   par(usr=c(0,1,0,1))
+   text(0.5,0.9,txt)
+ }
```

 # Create the Plots

```
pairs(iris[,1:4],lower.panel = panel.cor,upper.panel = upper.panel)
```