**STAT 43000/MA59800 Applied Statistics Spring 2021**

**Lab 1**

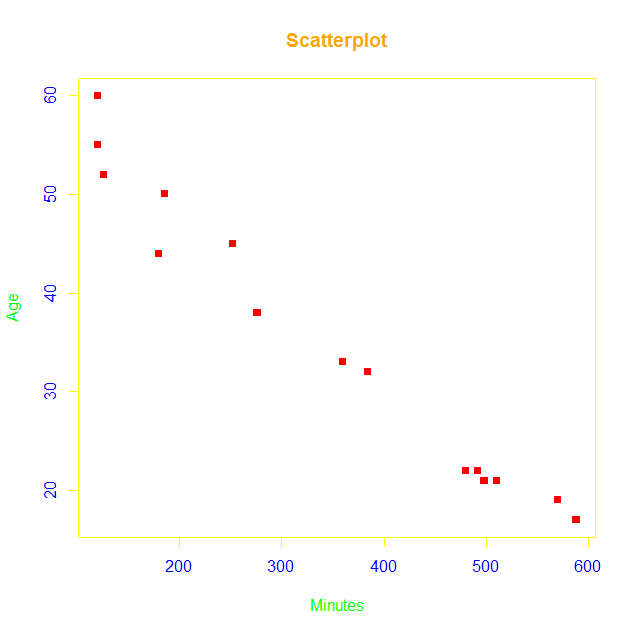
**Q.N. 1)** Several studies have found that there is a relationship between the age of an individual and the time spend on internet for online shopping. Below is the data showing the age of the respondents and their answer to the question “How many minutes do you browse online retailers per week?”

|  |  |
| --- | --- |
| Age | Minutes |
| 22 | 492 |
| 50 | 186 |
| 44 | 180 |
| 32 | 384 |
| 55 | 120 |
| 60 | 120 |
| 38 | 276 |
| 22 | 480 |
| 21 | 510 |
| 45 | 252 |
| 52 | 126 |
| 33 | 360 |
| 19 | 570 |
| 17 | 588 |
| 21 | 498 |

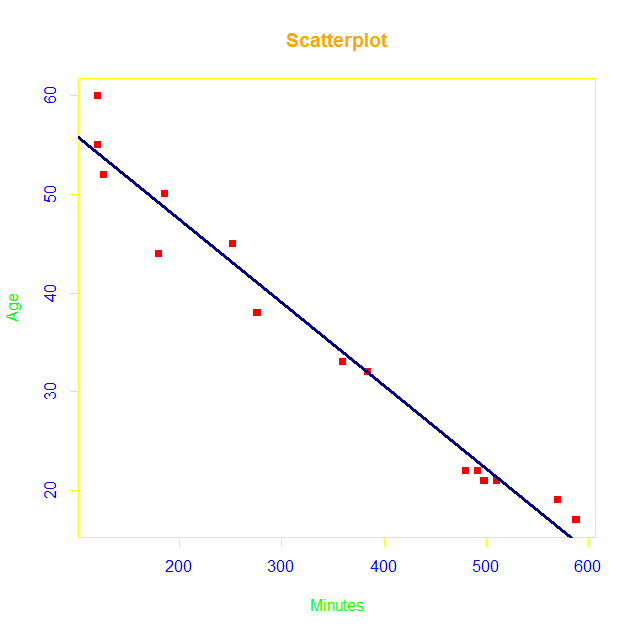
1. Make a scatterplot to display the data.

**R code for scatterplot:**

>plot(Minutes,Age,pch=15,col="red",main="Scatterplot",col.main="orange",col.axis="blue",col.lab="green",fg="yellow") #to plot a scatterplot first we need to scan the data and then we use this code to plot the graph.



1. Find the equation of the regression line. What does the slope parameter indicate?



**R code for linear model:**

>model=lm(Minutes~Age) #it constructs linear model

>model

>abline(model,col="dark blue",lwd=3) #line gets constructed on the graph

**Output**:

Call:

lm(formula = Minutes ~ Age)

Coefficients: #coefficients describes the equation of regression line

(Intercept) Age

750.0 -11.5

**Equation of regression line:**

Y= -11.5x+750.02

From the above equation we can say that, as the age increases by 1 year there is a decline of 11.5 minutes of browsing over internet for online shopping.

1. Construct a 90% confidence interval for the slope parameter.

**R code for confidence interval:**

>confint(model,level=0.9)

**Output**:

5 % 95 %

(Intercept) 61.37267559 67.11250222

Minutes -0.09166701 -0.07660947

**Alternative method:**

This all can also be done through UsingR library within one line, below is the code to construct scatterplot, line of regression, equation of regression line and confidence interval at 90% confidence:

**R code:**

>model=simple.lm(Age,Minutes,conf.level=0.9)

**Q.N. 2)** A marketing researcher studied annual sales of a product that had been introduced 10 years ago. The data are as follows, where x is the year coded and y is the sales in thousands of units:

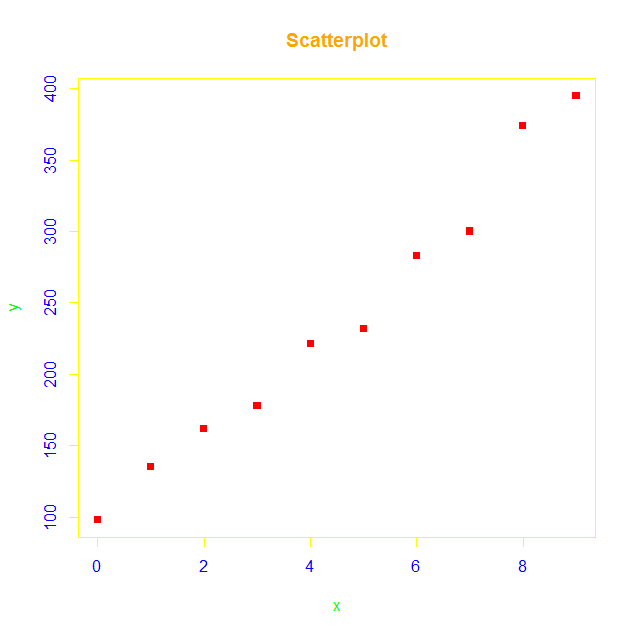
x: 0 1 2 3 4 5 6 7 8 9

y: 98 135 162 178 221 232 283 300 374 395

a) Prepare a scatter plot of the data

**R code for scatter plot:**

>plot(x,y,pch=15,col="red",main="Scatterplot",col.main="orange",col.axis="blue",col.lab="green",fg="yellow")



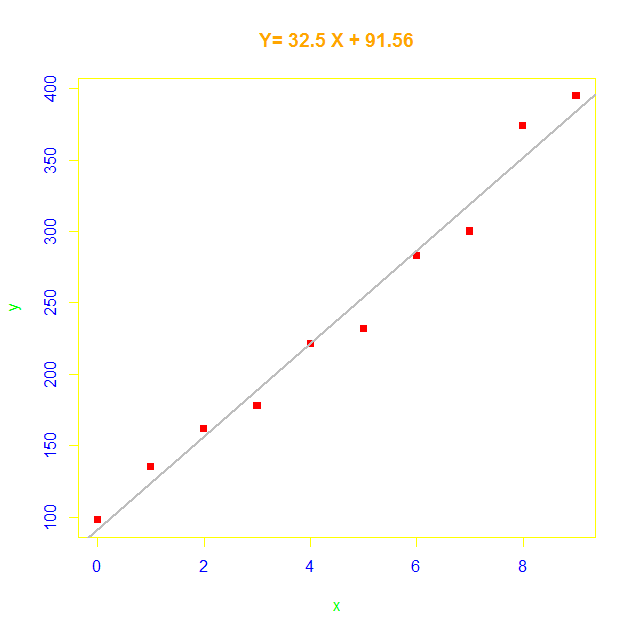
b) State the estimated regression line for the data and add it to the scatter plot.

**R code for line of regression:**

>model=lm(y~x) #linear model of the data

>plot(x,y,pch=15,col="red",main="Scatterplot",col.main="orange",col.axis="blue",col.lab="green",fg="yellow") #line code alone can’t be used to draw the line and it always comes after plotting the graph.

>abline(model,lwd=2,col="grey") # Estimated regression line.



c) Use the model to predict the sales in the 10th year (i.e. For x=10). Also provide the 95% confidence interval for the predicted value.

**R code for predicting the sales in the 10th year at 95% of confidence interval:**

>predict(model,data.frame(x=10),interval="conf")

**Output:**

fit lwr upr

416.5333 392.9089 440.1578

At 95% of confidence interval, 10th year sales predicted value is 416.5333 with upper limit as 440.1578 and lower limit as 392.9089.