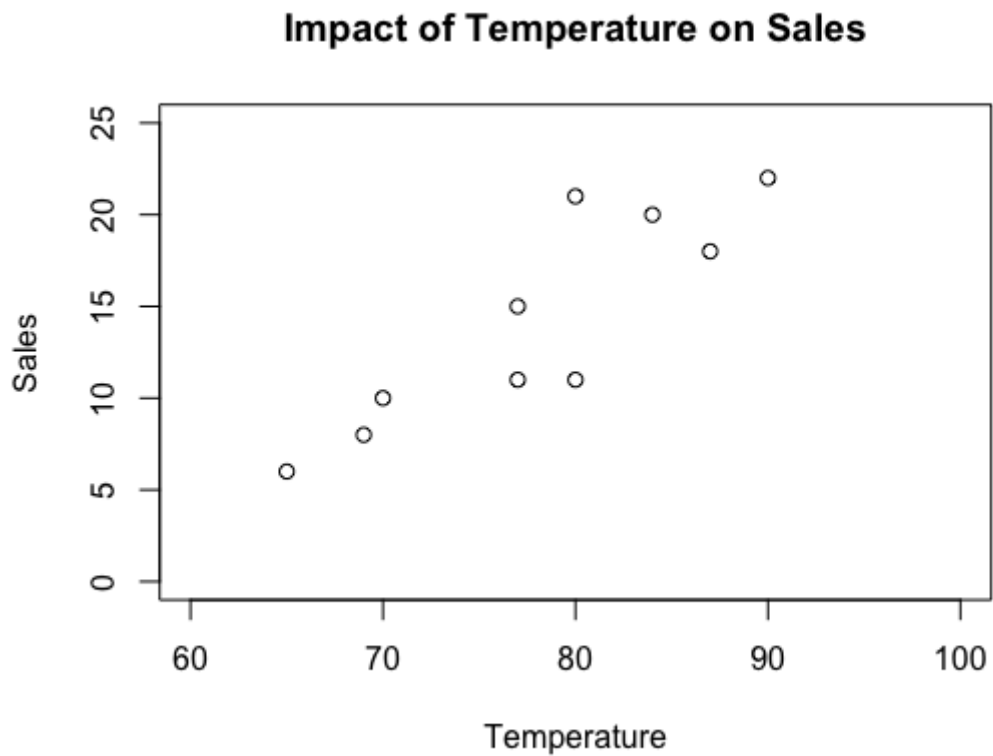


Introduction to Analytics
Module 1 - Project Assignment
Executive Summary Report 1
Devi Somalinga Bhuvanesh
17th January 2022

Key Findings

a. A scatter plot of the Sales ~ temp data



b. The mean temperature

```
Console | Terminal x | Background Jobs x
R 4.2.2 · ~/Documents/Devi/MPS Analytics/Introduction to Analytics/Module 1/
> #Q5 Mean temperature
> temperature <- c(69,80,77,84,80,77,87,70,65,90)
> mean(temperature)
[1] 77.9
```

c. Display the data after steps 6 and 7

```
Console Terminal x Background Jobs x
R 4.2.2 · ~/Documents/Devi/MPS Analytics/Introduction to Analytics/Module 1/
> #Q6 Delete 3rd element from sales vector
> sales <- c(8,11,15,20,21,11,18,10,6,22)
> sales1 <- sales[-3]
> sales1
[1] 8 11 20 21 11 18 10 6 22
>
> #Q7 Insert 16 as 3rd element into sales vector
> sales2 <- c(sales1[1:2],16,sales1[3:9])
> sales2
[1] 8 11 16 20 21 11 18 10 6 22
```

d. Display the names vector

```
Console Terminal x Background Jobs x
R 4.2.2 · ~/Documents/Devi/MPS Analytics/Introduction to Analytics/Module 1/
> #Q8 Create vector names with Tom, Dick, and Harry
> Names <- c("Tom","Dick","Harry")
> Names
[1] "Tom" "Dick" "Harry"
```

e. Display the matrix of 5 row by 2 column of 10 integers

```
Console Terminal x Background Jobs x
R 4.2.2 · ~/
> #Q9 Matrix of 10 integers with 5 rows and 2 columns
> y <- matrix(1:10, nrow=5, ncol=2)
> y
      [,1] [,2]
[1,]    1    6
[2,]    2    7
[3,]    3    8
[4,]    4    9
[5,]    5   10
```

f. Display the icSales data frame

```
Console Terminal x Background Jobs x
R 4.2.2 · ~/Documents/Devi/MPS Analytics/Introduction to Analytics/Module 1/
> #Q10 Dataframe - icSales
> sales <- c(8,11,15,20,21,11,18,10,6,22)
> temperature <- c(69,80,77,84,80,77,87,70,65,90)
> icSales <- data.frame(sales,temperature)
> icSales
  sales temperature
1     8           69
2    11           80
3    15           77
4    20           84
5    21           80
6    11           77
7    18           87
8    10           70
9     6           65
10   22           90
```

g. Display the summary of the icSales data frame

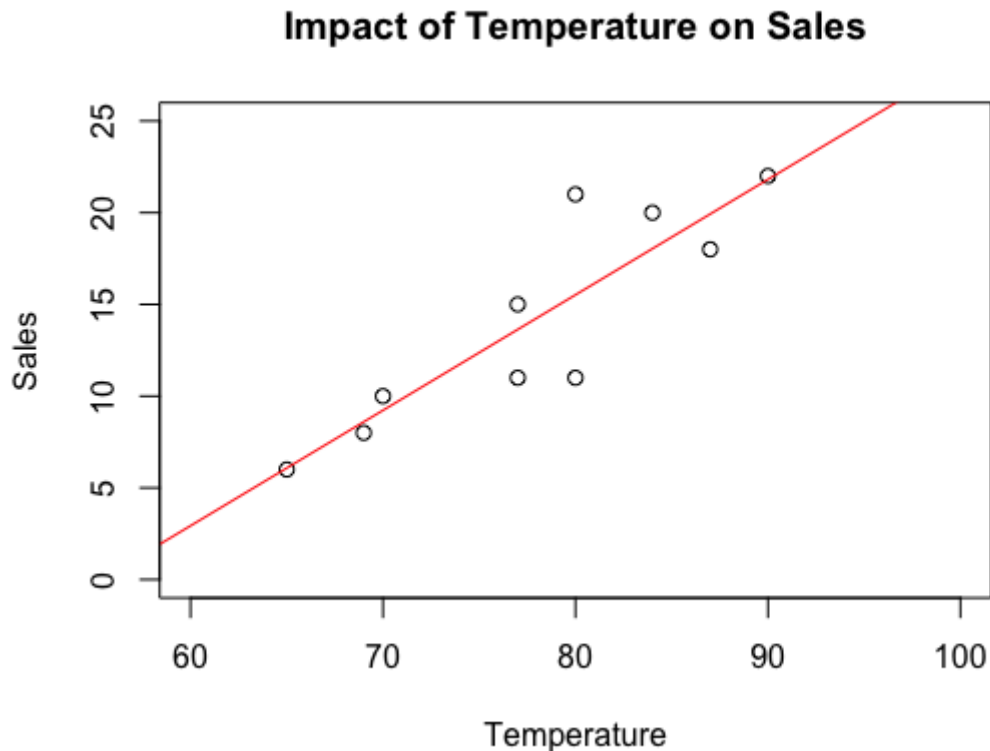
```
Console Terminal x Background Jobs x
R 4.2.2 · ~/Documents/Devi/MPS Analytics/Introduction to Analytics/Module 1/
> #Q12 Summary icSales
> sales <- c(8,11,15,20,21,11,18,10,6,22)
> temperature <- c(69,80,77,84,80,77,87,70,65,90)
> icSales <- data.frame(sales,temperature)
> summary(icSales)
  sales      temperature
Min.   : 6.00   Min.   :65.00
1st Qu.:10.25   1st Qu.:71.75
Median :13.00   Median :78.50
Mean    :14.20   Mean    :77.90
3rd Qu.:19.50   3rd Qu.:83.00
Max.    :22.00   Max.    :90.00
```

h. Display the variables only from the Student.csv data set.

```
Console Terminal x Background Jobs x
R 4.2.2 · ~/Documents/Devi/MPS Analytics/Introduction to Analytics/Module 1/
> #Q14 Display only variable name of student.csv
> studentdata <- read.csv("Student.csv",header=TRUE,sep=",")
Warning message:
In read.table(file = file, header = header, sep = sep, quote = quote, :
incomplete final line found by readTableHeader on 'Student.csv'
> studentdata
  StudentID First      Last Math Science Social.Studies
1       11  Bob      Smith   90    80         67
2       12  Jane    Weary   75    NA         80
3       10  Dan Thornton, III 65    75         70
4       40  Mary    O'Leary  90    95         92
> colnames(studentdata)
[1] "StudentID" "First"      "Last"       "Math"       "Science"    "Social.Studies"
```

i. A summary of the information you learned about the data sets based on the instructions you followed.

The key findings from the data sets and analysis are as follows:



- Overall, the above scatter-plot indicates that with the increase in temperature, the sales have gradually increased from minimum value of 6 to maximum 22 which is equally distributed on either side of the average temperature of 77.9 and average sales of 14.20.
- With the rise in temperature starting from 65 to 77, the sales consistently increased reaching to 11, i.e., slightly higher than the 1st quartile, and remained constant even when the temperature was little above the average (80).
- On the other hand, at a temperature of 77, two different instances at the same time can be noticed where the sales were at 11 and 15. Similarly, the sales were at 11 and 21 when the temperature hit 80.
- With further rise in temperature from 80 to 87, a steady decrease in sales was observed from 21 to 18 followed with a sudden peak in sales reaching to 22 at a maximum temperature of 90.
- The data approximately fits a linear model. At almost four instances, the sales value falls exactly at the estimated temperature in a linear fashion. In 3 situations, the sales value is higher than the estimation. However, the sales value fell below the expectation at 3 places.

Bibliography

- Kabacoff, R. (2015). *R in Action: Data Analysis and Graphics with R* (2nd ed.). Manning.
- GeeksforGeeks. (2021, February 17). How to Create a Scatterplot with a Regression Line in R? *GeeksforGeeks*.
<https://www.geeksforgeeks.org/how-to-create-a-scatterplot-with-a-regression-line-in-r/>
- R insert element how to insert element into vector. (2020, September 18). *Stack Overflow*.
<https://stackoverflow.com/questions/63960655/r-insert-element-how-to-insert-element-into-vector>
- Simplilearn. (2020, June 18). *R Tutorial For Beginners Part - 1 | R Programming For Beginners | R Language Tutorial | Simplilearn* [Video]. YouTube.
<https://www.youtube.com/watch?v=7NLPPFU0O3w>
- ALY 6000.20926.202325. Learning to use R - Instructor Videos Lesson 1-5.
https://northeastern.instructure.com/courses/131162/pages/learning-to-use-r-instructor-videos?module_item_id=8426786

Appendix

> #Q1 - Print name at top of Script

```
> print("Devi Somalinga Bhuvanesh")
```

```
[1] "Devi Somalinga Bhuvanesh"
```

> #Q2 To install VCD

```
> install.packages("vcd")
```

```
Error in install.packages : Updating loaded packages
```

Restarting R session...

```
> install.packages("vcd")
```

```
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/vcd_1.4-10.tgz'
```

```
Content type 'application/x-gzip' length 1286536 bytes (1.2 MB)
```

```
=====
```

```
downloaded 1.2 MB
```

The downloaded binary packages are in

```
/var/folders/1y/qwg6z9nj78nfkv1gf3qrqts40000gp/T//Rtmp7hL27B/downloaded_packages
```

> #Q3 To import VCD

```
> library("vcd")
```

```
Loading required package: grid
```

> #Q4 Scatter plot for Sales vs Temperature

```
> sales <- c(8,11,15,20,21,11,18,10,6,22)
```

```
> temperature <- c(69,80,77,84,80,77,87,70,65,90)
```

```
> plot(temperature,sales, main="Impact of Temperature on Sales", xlab="Temperature",  
ylab="Sales", xlim=c(60,100), ylim=c(0,25))
```

```
> salesvstemp <- data.frame(temperature,sales)
```

```
> salesvstemp
```

```
  temperature sales
```

```
1         69     8
```

```
2         80    11
```

```
3         77    15
```

```
4         84    20
```

```
5         80    21
```

```
6         77    11
```

```
7         87    18
```

```
8         70    10
```

```
9         65     6
```

```
10        90    22
```

```
> abline(lm(sales~temperature,data=salesvstemp),col='red') #Linear Model
> lm(sales~temperature,data=salesvstemp) #to know the slope and y-intercept
```

Call:

```
lm(formula = sales ~ temperature, data = salesvstemp)
```

Coefficients:

```
(Intercept) temperature
-34.8388      0.6295
```

> #Q5 Mean temperature

```
> temperature <- c(69,80,77,84,80,77,87,70,65,90)
> mean(temperature)
[1] 77.9
```

> #Q6 Delete 3rd element from sales vector

```
> sales <- c(8,11,15,20,21,11,18,10,6,22)
> sales1 <- sales[-3]
> sales1
[1] 8 11 20 21 11 18 10 6 22
```

> #Q7 Insert 16 as 3rd element into sales vector

```
> sales2 <- c(sales1[1:2],16,sales1[3:9])
> sales2
[1] 8 11 16 20 21 11 18 10 6 22
```

> #Q8 Create vector names with Tom, Dick, and Harry

```
> Names <- c("Tom","Dick","Harry")
> Names
[1] "Tom" "Dick" "Harry"
> class(Names) #to know the character
[1] "character"
```

> #Q9 Matrix of 10 integers with 5 rows and 2 columns

```
> y <- matrix(1:10, nrow=5, ncol=2)
> y
     [,1] [,2]
[1,]    1    6
[2,]    2    7
[3,]    3    8
[4,]    4    9
[5,]    5   10
```


> #Q10 Dataframe - icSales

```
> sales <- c(8,11,15,20,21,11,18,10,6,22)
> temperature <- c(69,80,77,84,80,77,87,70,65,90)
> icSales <- data.frame(sales,temperature)
> icSales
  sales temperature
1     8          69
2    11          80
3    15          77
4    20          84
5    21          80
6    11          77
7    18          87
8    10          70
9     6          65
10   22          90
```

> #Q11 Structure Dataframe

```
> sales <- c(8,11,15,20,21,11,18,10,6,22)
> temperature <- c(69,80,77,84,80,77,87,70,65,90)
> icSales <- data.frame(sales,temperature)
> icSales
  sales temperature
1     8          69
2    11          80
3    15          77
4    20          84
5    21          80
6    11          77
7    18          87
8    10          70
9     6          65
10   22          90
> str(icSales)
'data.frame':   10 obs. of  2 variables:
 $ sales      : num  8 11 15 20 21 11 18 10 6 22
 $ temperature: num  69 80 77 84 80 77 87 70 65 90
```

> #Q12 Summary icSales

```
> sales <- c(8,11,15,20,21,11,18,10,6,22)
> temperature <- c(69,80,77,84,80,77,87,70,65,90)
> icSales <- data.frame(sales,temperature)
> summary(icSales)
```

```

    sales    temperature
Min.   : 6.00  Min.   :65.00
1st Qu.:10.25 1st Qu.:71.75
Median :13.00 Median :78.50
Mean   :14.20 Mean   :77.90
3rd Qu.:19.50 3rd Qu.:83.00
Max.   :22.00 Max.   :90.00

```

> #Q13 To import dataset student.csv

```
> studentdata <- read.csv("Student.csv",header=TRUE,sep=",")
```

Warning message:

```
In read.table(file = file, header = header, sep = sep, quote = quote, :
incomplete final line found by readTableHeader on 'Student.csv'
```

```
> studentdata
```

	StudentID	First	Last	Math	Science	Social.Studies
1	11	Bob	Smith	90	80	67
2	12	Jane	Weary	75	NA	80
3	10	Dan	Thornton, III	65	75	70
4	40	Mary	O'Leary	90	95	92

> #Q14 Display only variable name of student.csv

```
> studentdata <- read.csv("Student.csv",header=TRUE,sep=",")
```

Warning message:

```
In read.table(file = file, header = header, sep = sep, quote = quote, :
incomplete final line found by readTableHeader on 'Student.csv'
```

```
> studentdata
```

	StudentID	First	Last	Math	Science	Social.Studies
1	11	Bob	Smith	90	80	67
2	12	Jane	Weary	75	NA	80
3	10	Dan	Thornton, III	65	75	70
4	40	Mary	O'Leary	90	95	92

```
> colnames(studentdata)
```

```
[1] "StudentID" "First" "Last" "Math" "Science" "Social.Studies"
```