|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Descrete |
| Results of rolling a dice | Continuous |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Nominal |
| Number of kids | Descrete |
| Number of tickets in Indian railways | Descrete |
| Number of times married | Descrete |
| Gender (Male or Female) | Ordinal |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Ordinal |
| Weight | Nominal |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ordinal |
| Height | Nominal |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ordinal |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

**Answer:** Total number of coins n = 3

Total outcomes for three tossed coins = 8

Total Outcomes:

HHH, TTT, HTH, THT, HHT, TTH, THH, HTT

Probability for getting 2 heads and 1tail = 3/8

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2and 3

**Answer:** Total out comes for two rolled dice = 36

Total Outcome details:

(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)

(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)

(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)

(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)

(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)

(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)

1. Probability for two rolled dice when Sum is 1 = 0/36

= 0

1. Probability for two rolled dice when Sum is Less than or equal to 4

= 6/36

= 1/36

1. Probability for two rolled dice when Sum is divisible by 2 and 3

= 24/36

= 2/3

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

**Answer:** A bag contains total 7 balls which details are following:

Red, Red, Green, Green, Green, Blue, Blue

When two ball is randomly dawn from total balls then probabilty

= 7c5

=7\*6\*5\*4\*3\*2\*1 / 5\*4\*3\*2\*1\* 2\*1

= 21

The Probability of second ball where second ball not being blue

= 5c2

= 5\*4\*3\*2\*1/ 3\*2\*1\* 2\*1

= 10

When none of the balls drawn is blue then probability

= 10/21

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children(ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

**Answer:** Expected number of candies for randomly selected child:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CHILD | Candies count | Probability | Expected values | Expected candies for randomly selected child |
| A | 1 | 0.015 | 1\*0.015 = 0.015 | **3.09** |
| B | 4 | 0.20 | 4\*0.20 = 0.80 |
| C | 3 | 0.65 | 3\*0.65 = 1.95 |
| D | 5 | 0.005 | 5\* 0.005= 0.025 |
| E | 6 | 0.01 | 6\*0.01 = 0.06 |
| F | 2 | 0.120 | 2\*0.120 = 0.240 |

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points, Score, Weigh

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Answer:** As per R computation we find mean, median, variance, standard

deviation and range for Points, Score & Weight:

**For Points-**

**Mean =** 3.596563

**Median =** 3.695

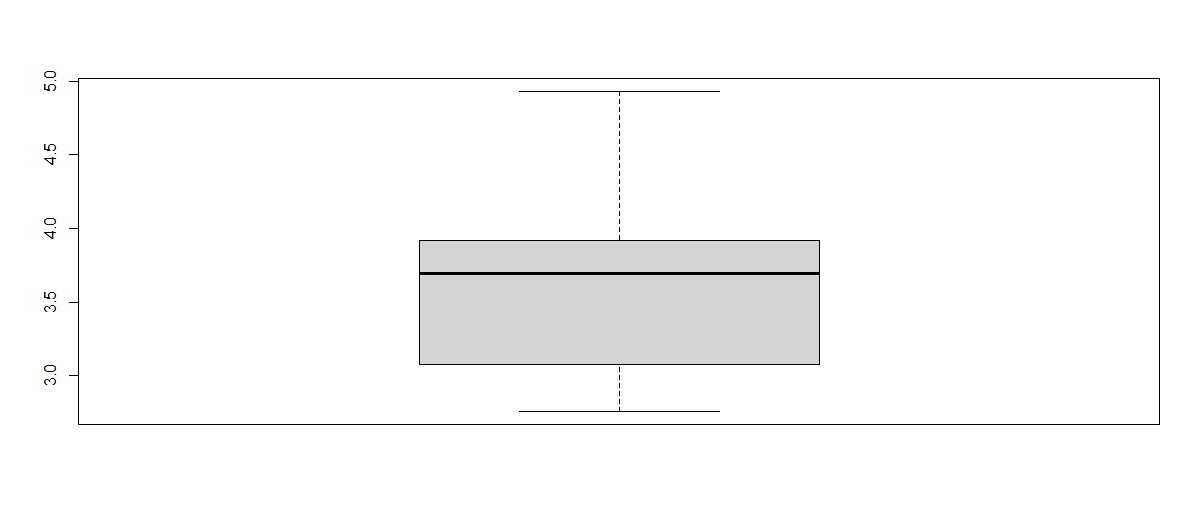
**Mode=** 3.90, 3.08, 3.07, 3.15, 4.08, 2.76, 3.92, 4.22

**Variance =** 0.2858814

**Standard deviation =** 0.5346787

**Range =** 2.17

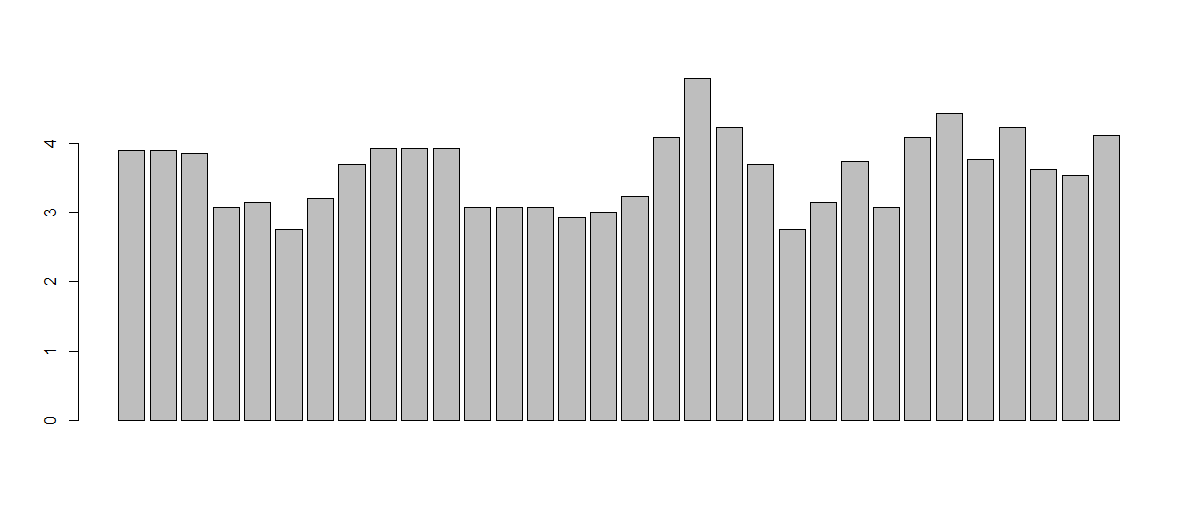
**Box Plot:**

****

**As per above box plot there is no out lairs it means that for points distribution is**

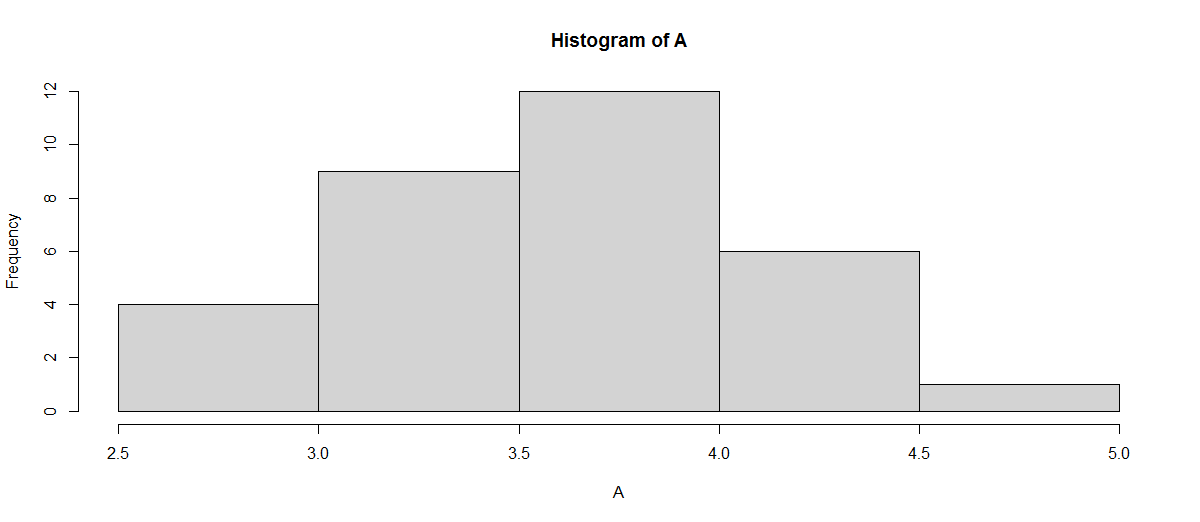
**good distribution.**

**Bar Plot:**

****

**As per above bar plot for points it’s a normal distribution.**

**Histogram:**

****

**As per above histogram we found that frequencies between points 3.0 & 4.0**

**is very high.**

**For Scores-**

**Mean =** 3.21725

**Median =** 3.325

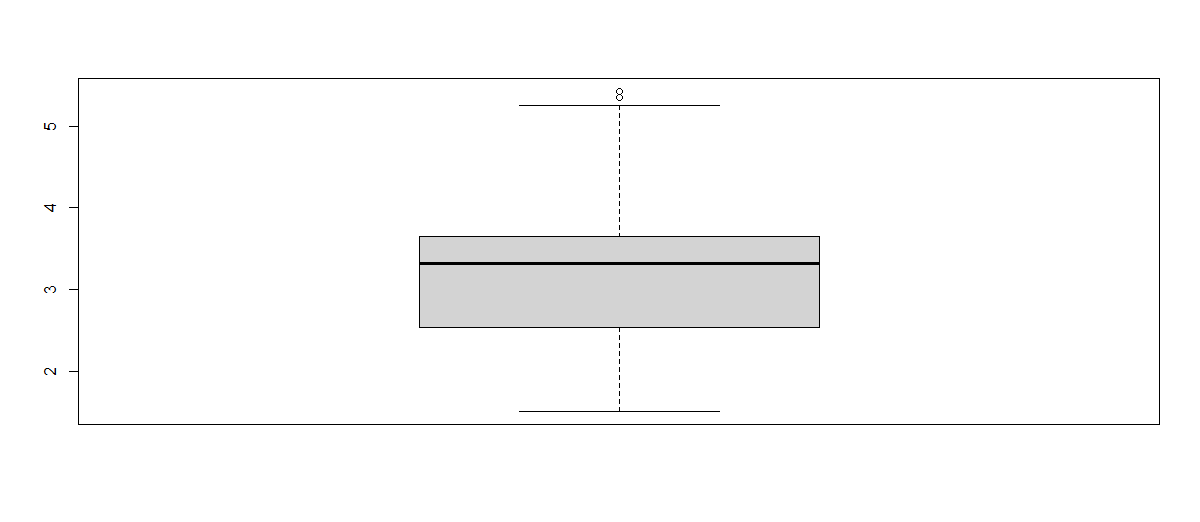
**Mode=** 3.440, 3.570

**Variance =** 0.957379

**Standard deviation =** 0.9784574

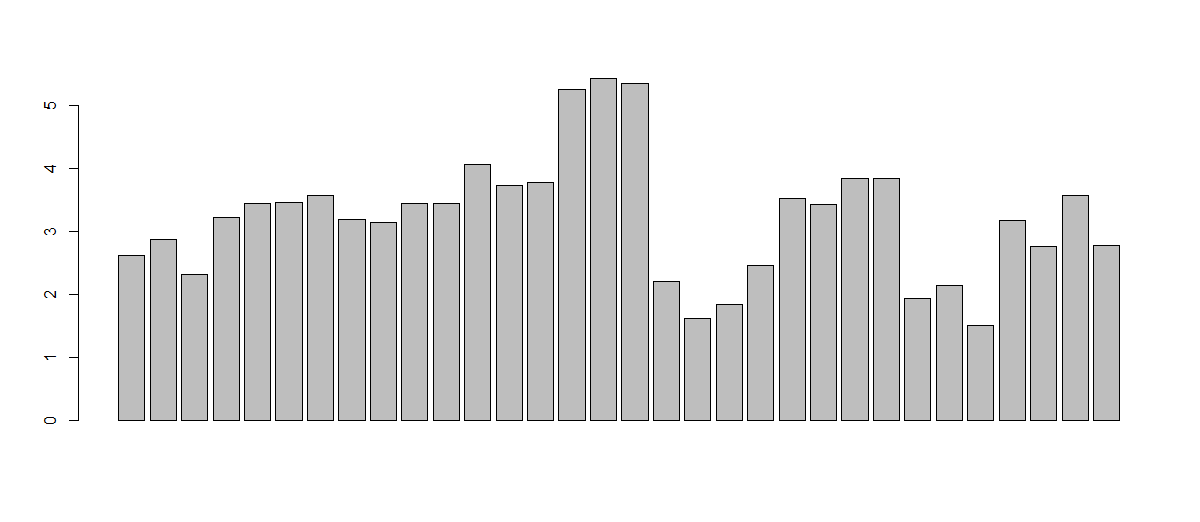
**Range =** 3.911

**Box Plot:**

****

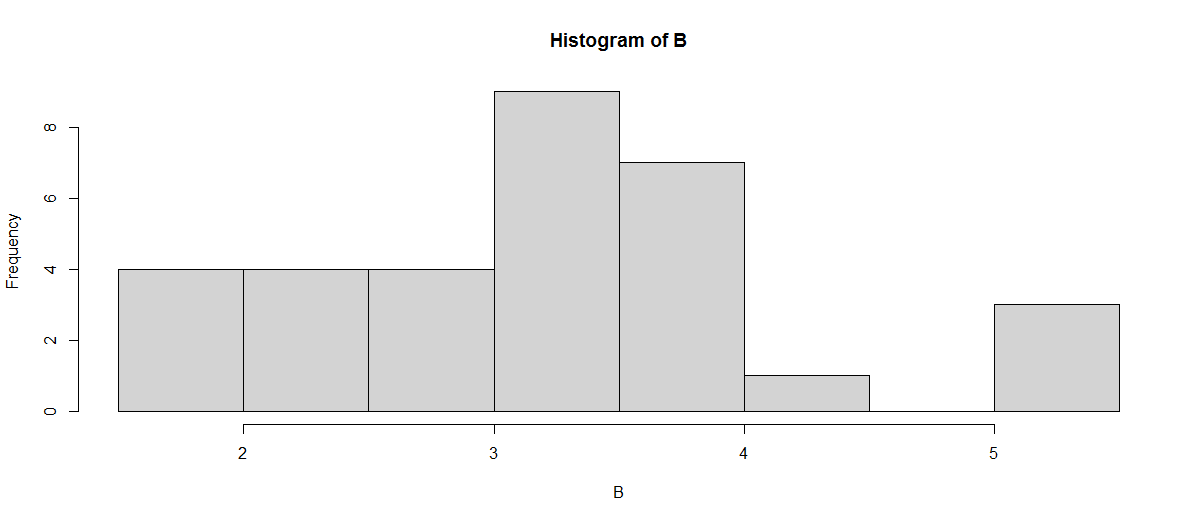
**As per above box plot there is two out lairs value 5.424 & 5.345 which needs to be removed for good distribution.**

**Bar Plot:**

****

**As per above bar plot for points it’s normal distribution.**

**Histogram:**

****

**As per above histogram we found that frequencies between scores 3 & 4**

**is very high.**

**For Weigh-**

**Mean =** 17.84875

**Median =** 17.71

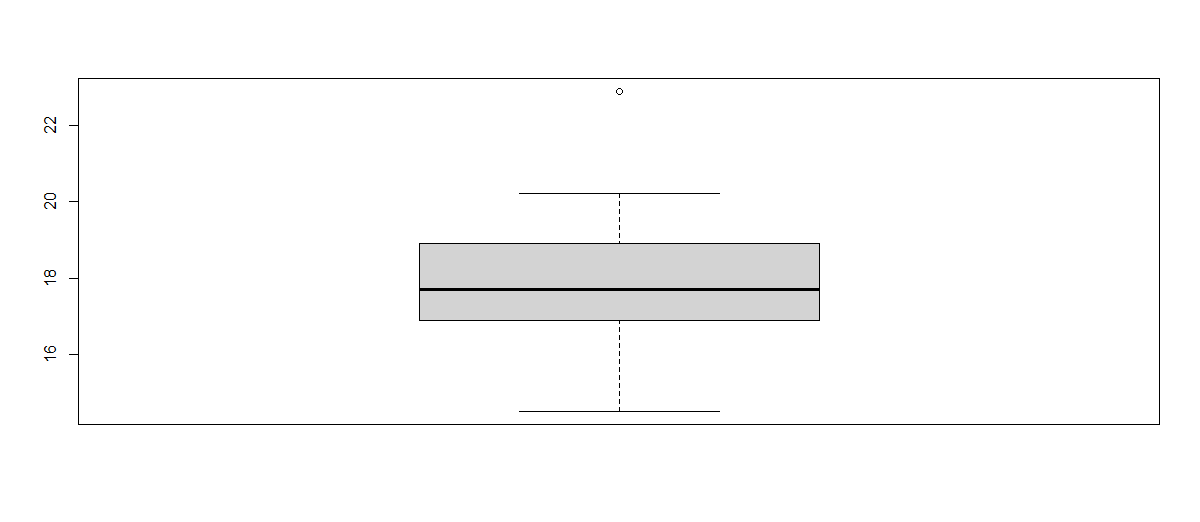
**Mode=** 17.02, 18.90

**Variance =** 3.193166

**Standard deviation =** 1.786943

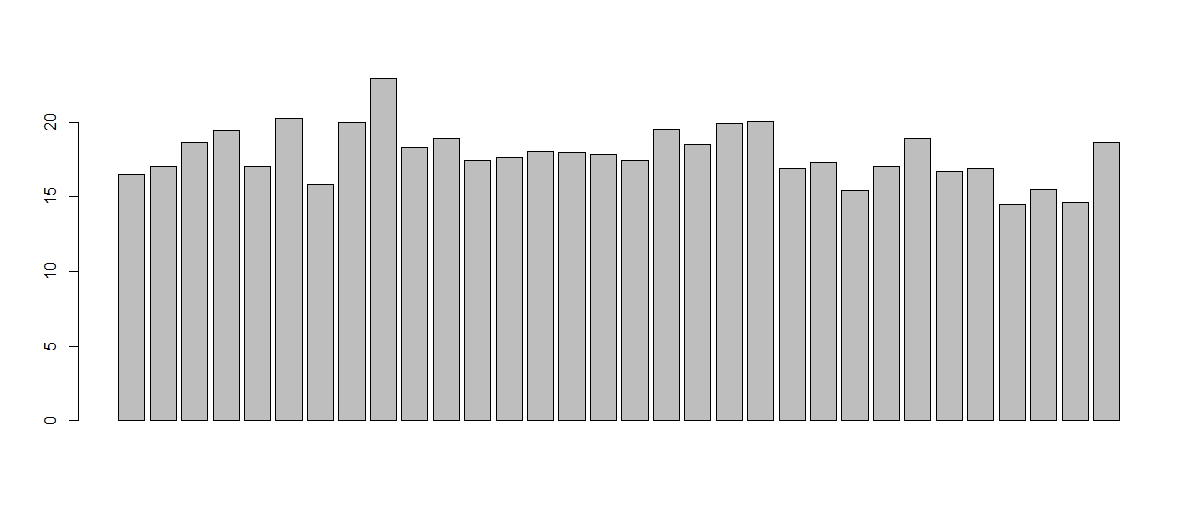
**Range =** 8.4

**Box Plot:**

****

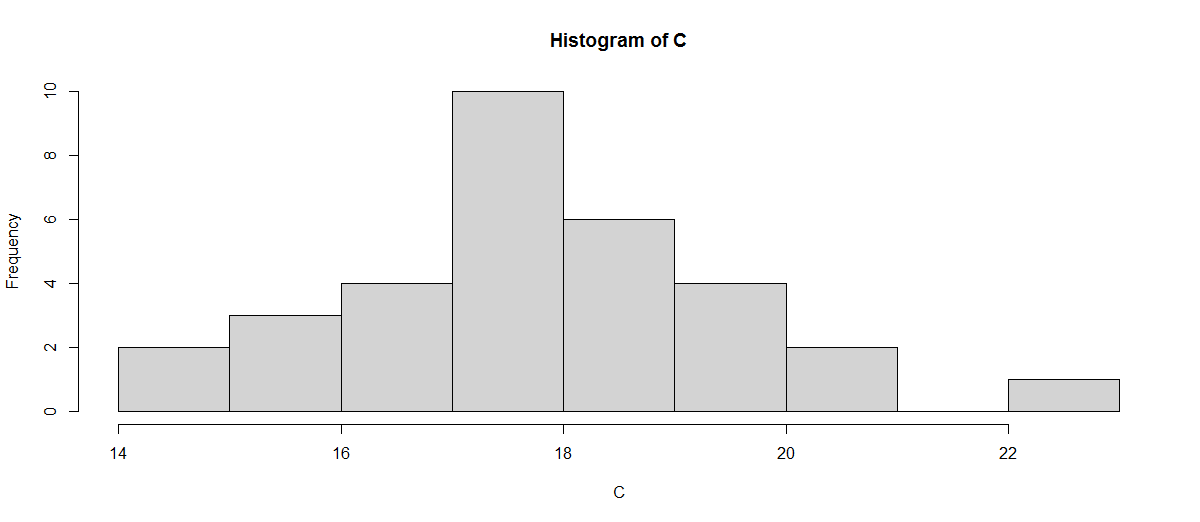
**As per above box plot there is one out lair value 22.90 which needs to be removed for good distribution.**

**Bar Plot:**

****

**As per above bar plot for points it’s normal distribution.**

**Histogram:**

****

**As per above histogram we found that frequencies between weigh 16.5 & 18.5**

**is very high.**

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

**Answer:** Expected value of the weight of one patient = 145.3333

Kindly find the R code file for above expected value solution.

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer:** Kindly find the R code file computation for Skewness, Kurtosis values of

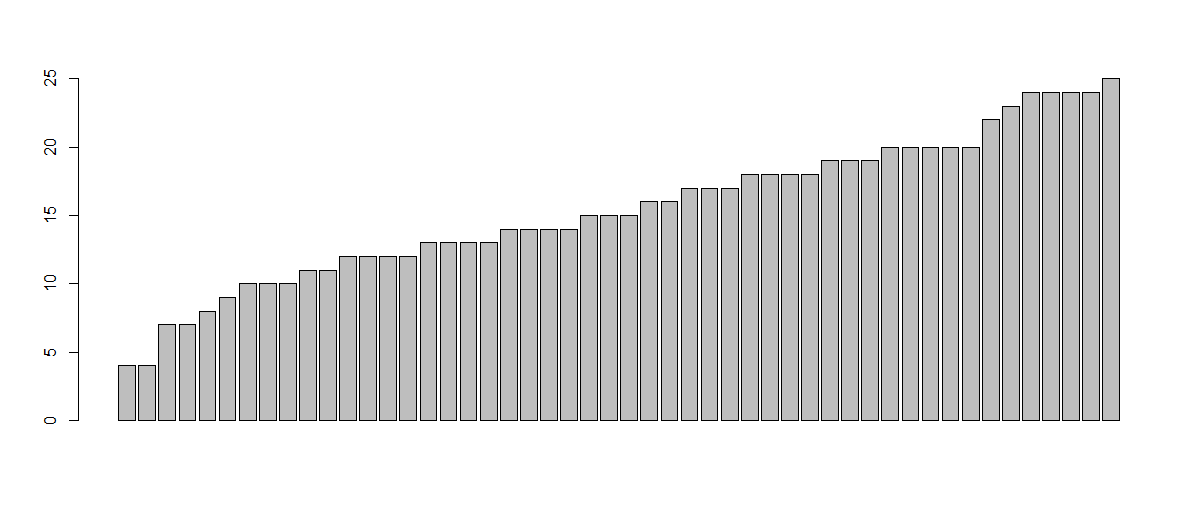
Cars speed, distance, SP & Weight which details are following:

**For Car Speed:**

Skewness = -0.1105533

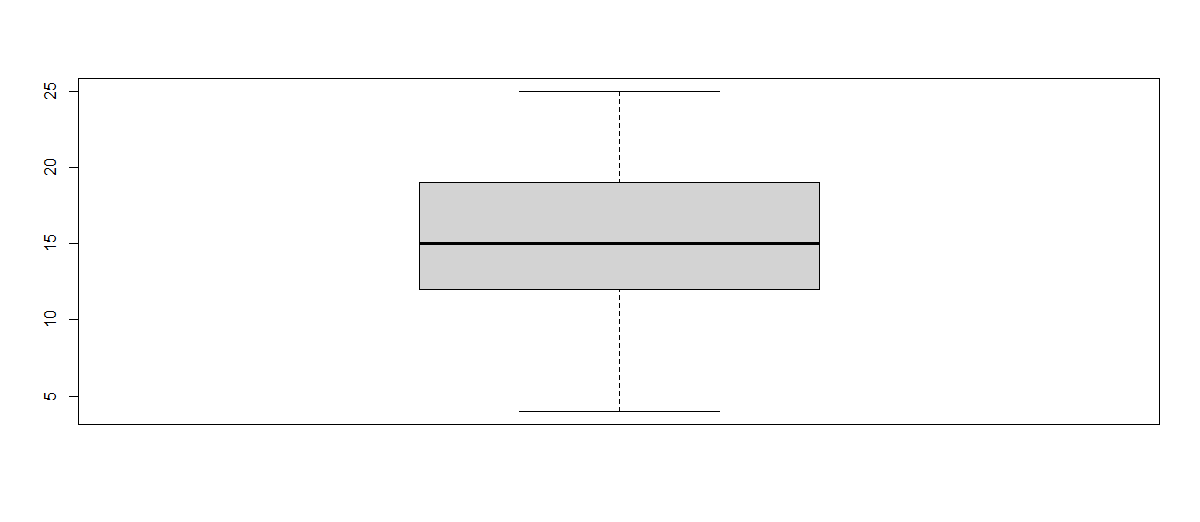
Kurtosis = -0.6730924

**Bar Plot:**



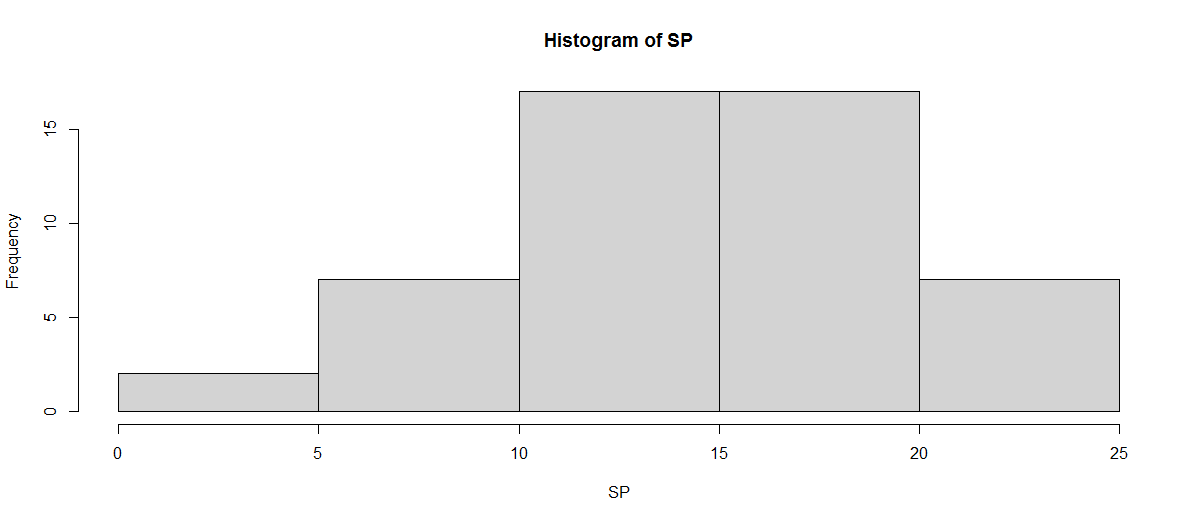
As per above bar plot for car speed it is normal distribution.

**Box Plot:**



As per above box plot for car speed there is no out lairs.

**Histogram:**



**As per above histogram we found that frequencies between weigh 10 & 20**

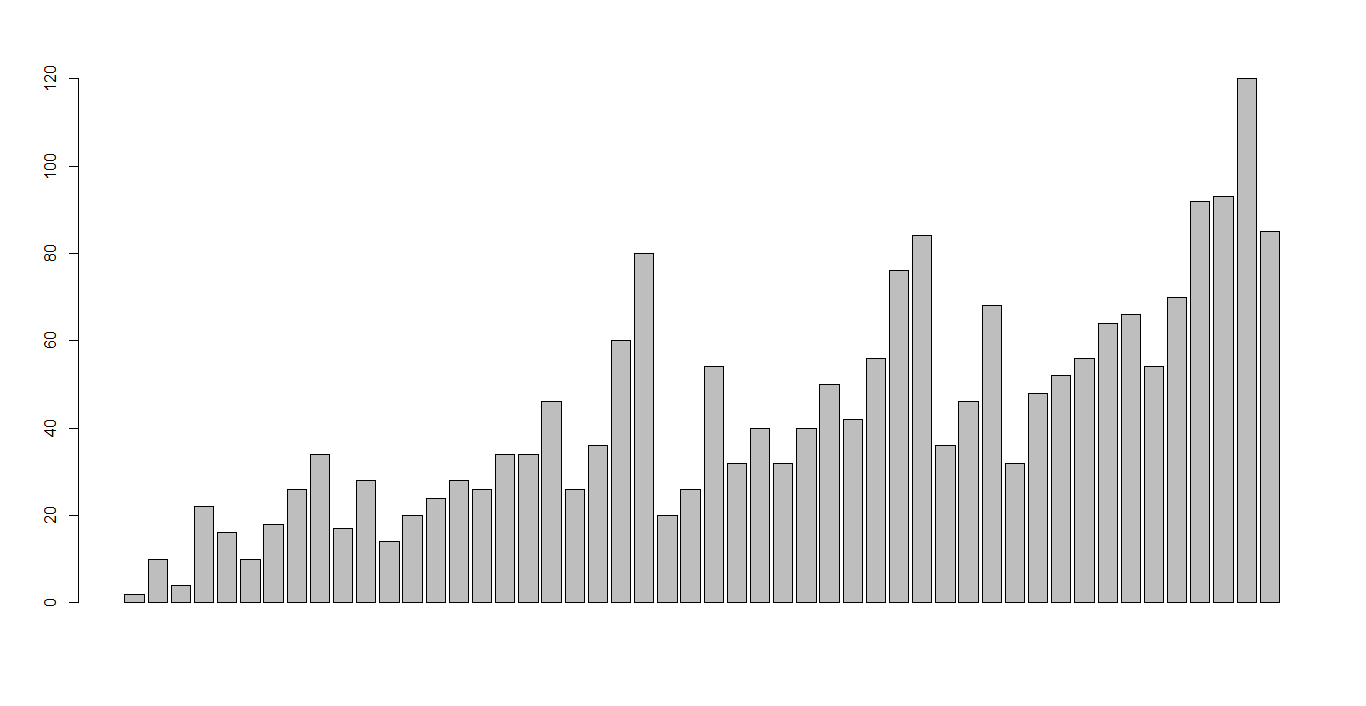
**is very high & there is a negative skewness & kurtosis.**

**For Car Distance:**

Skewness = 0.7591268

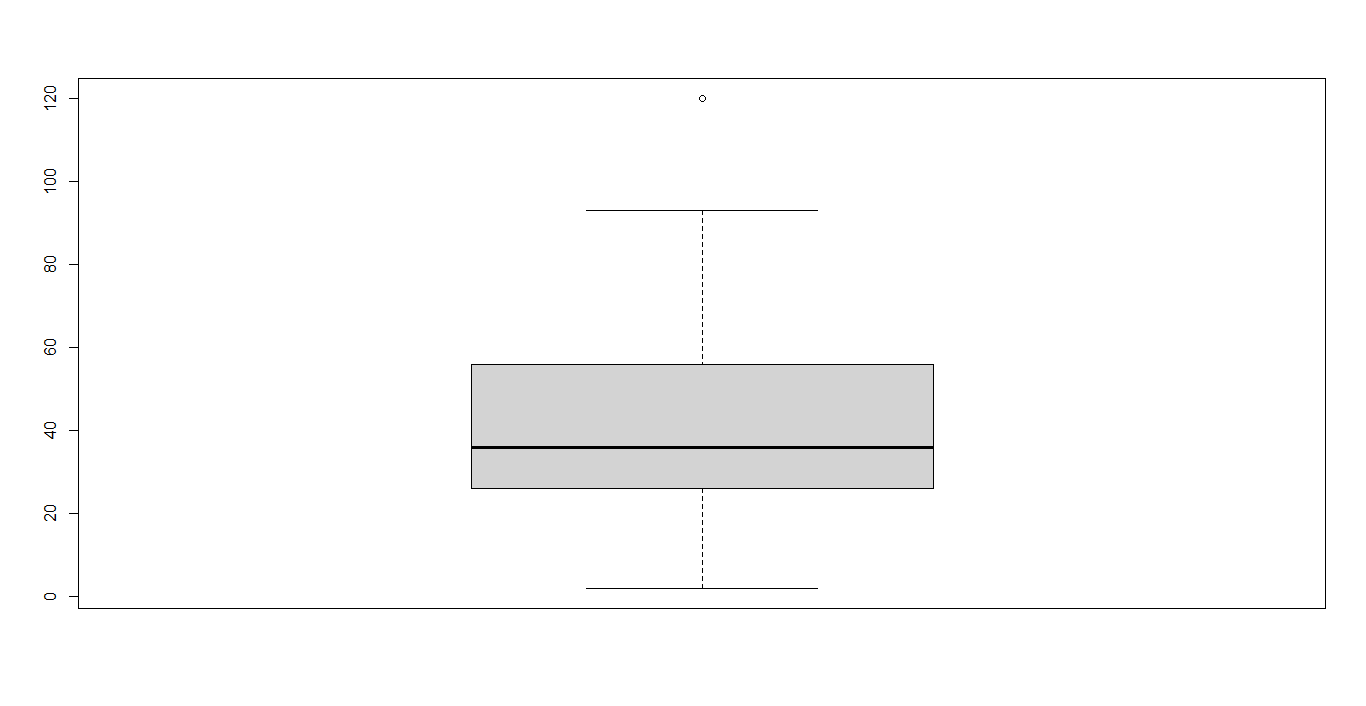
Kurtosis = 0.1193971

**Bar Plot:**



As per above par plot it’s a normal distribution.

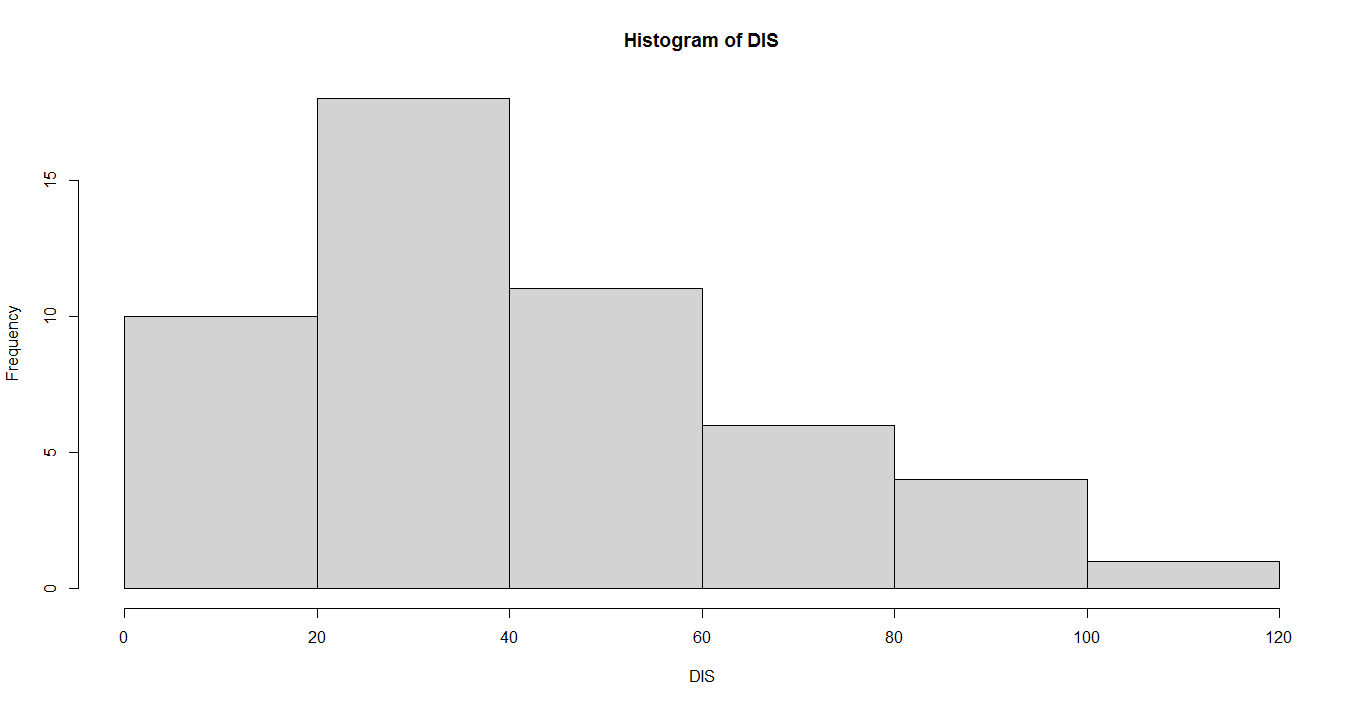
**Box Plot:**

****

As per box plot there is a for distance there is one out lair value which value

is 120 & needs to be removed for good distribution.

**Histogram:**

****

**As per above histogram we found that frequencies between weigh 20 & 60**

**is very high & there is a positive skewness and kurtosis.**

**For Car SP:**

Skewness = 1.552258

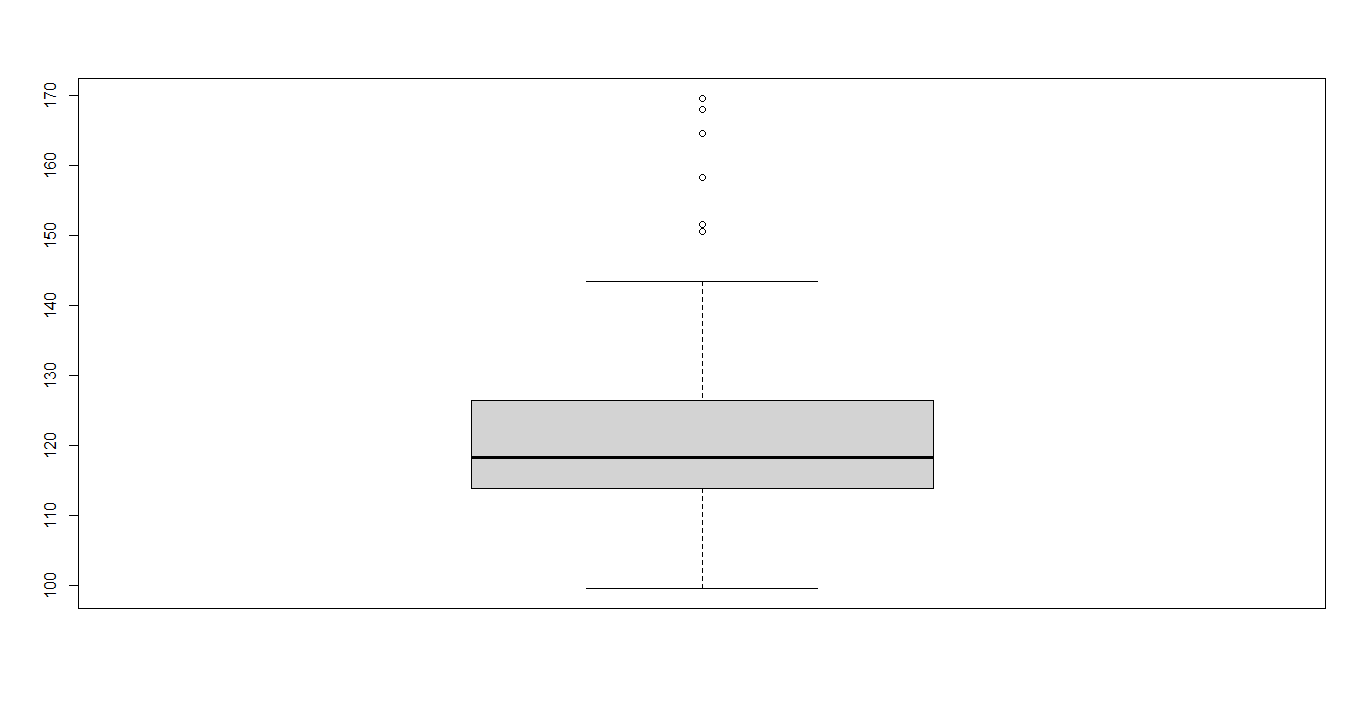
Kurtosis = 2.583072

**Bar Plot:**

****

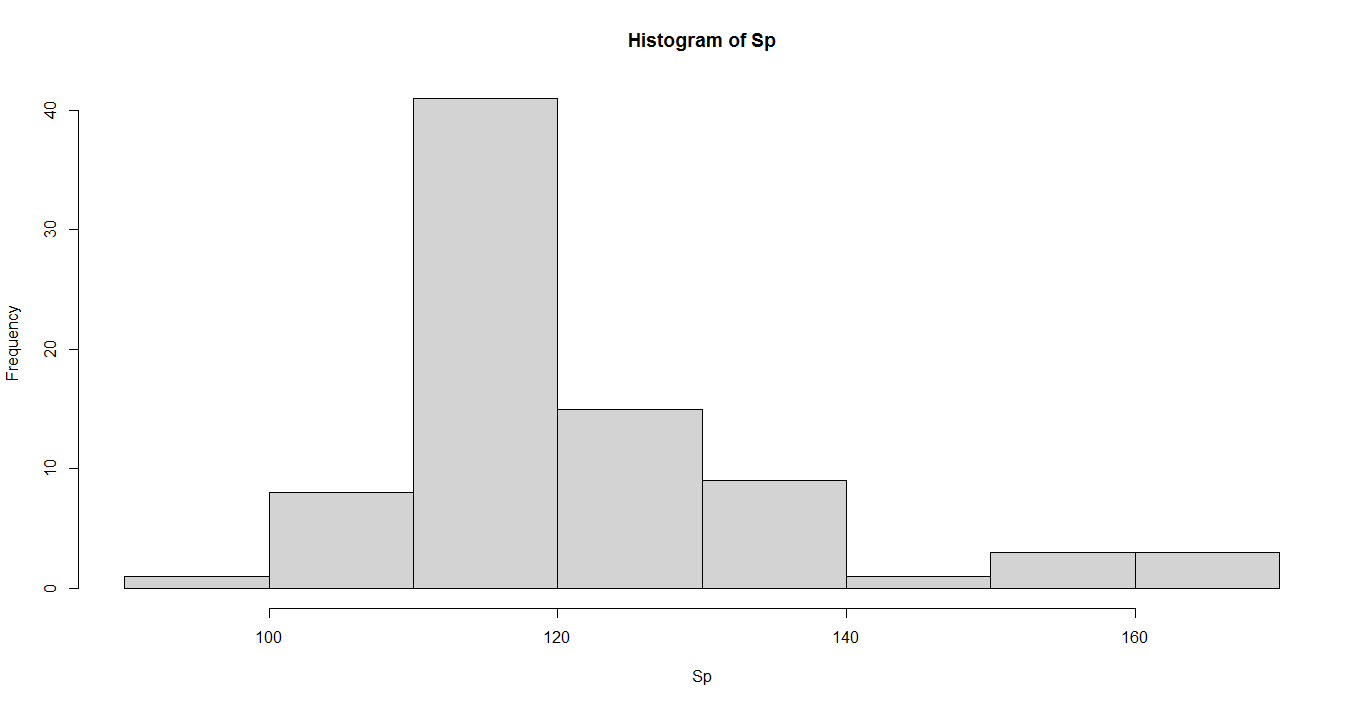
As per above par plot it’s a normal distribution.

**Box Plot:**

****

As per box plot there is a for distance there is 6 out lairs which values are 169.59851, 167.94446, 164.59851, 158.30067, 151.59851, 150.57658 & needs to be removed for good distribution.

**Histogram:**

****

As per above histogram we found that frequencies for SP between 110 & 130

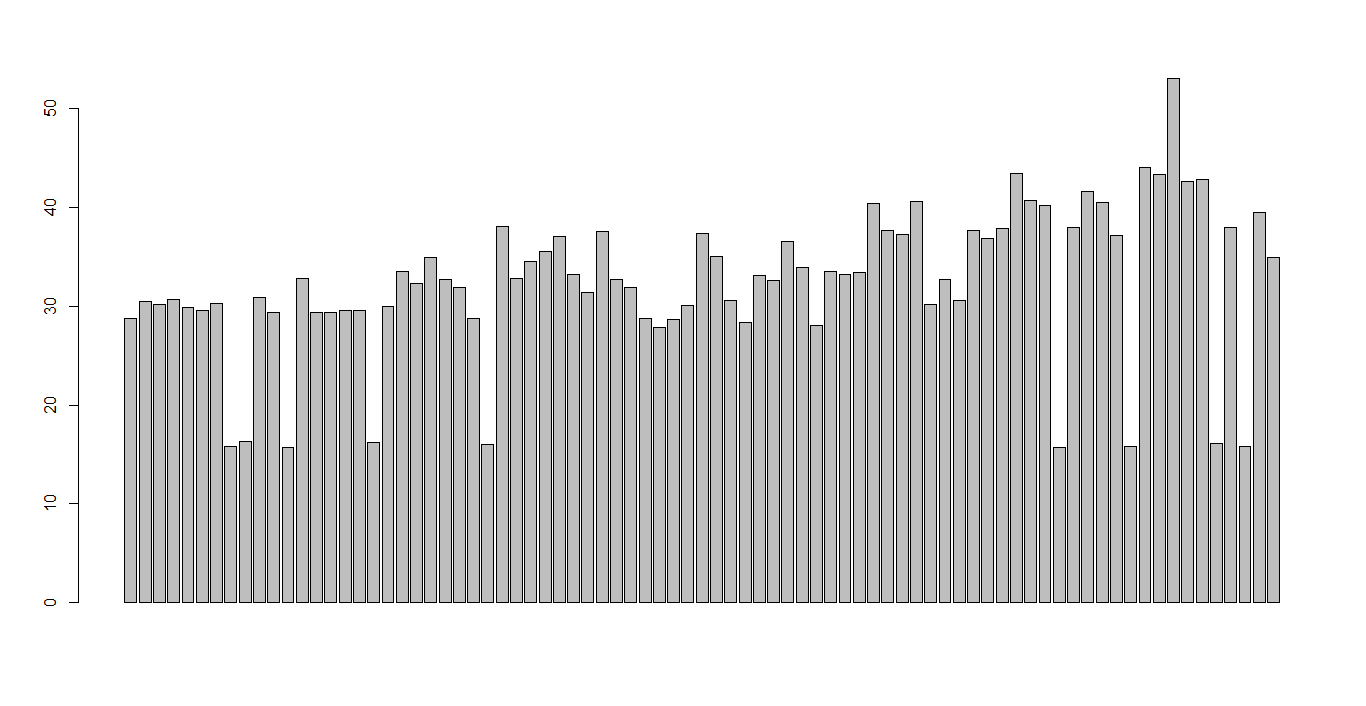
is very high & there is a positive skewness and kurtosis along with negative skewness & kurtosis**.**

**For Car Weight**

Skewness = -0.5921721

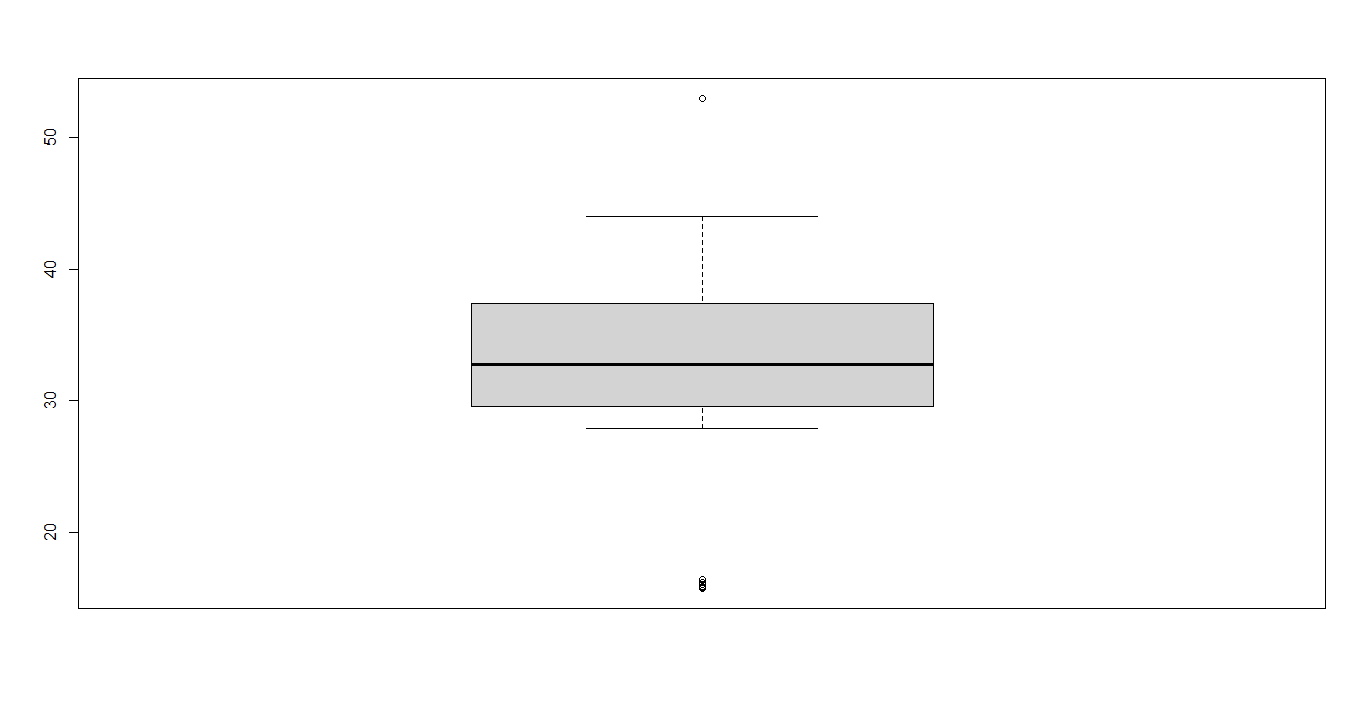
Kurtosis = 0.7257402

**Bar Plot:**

****

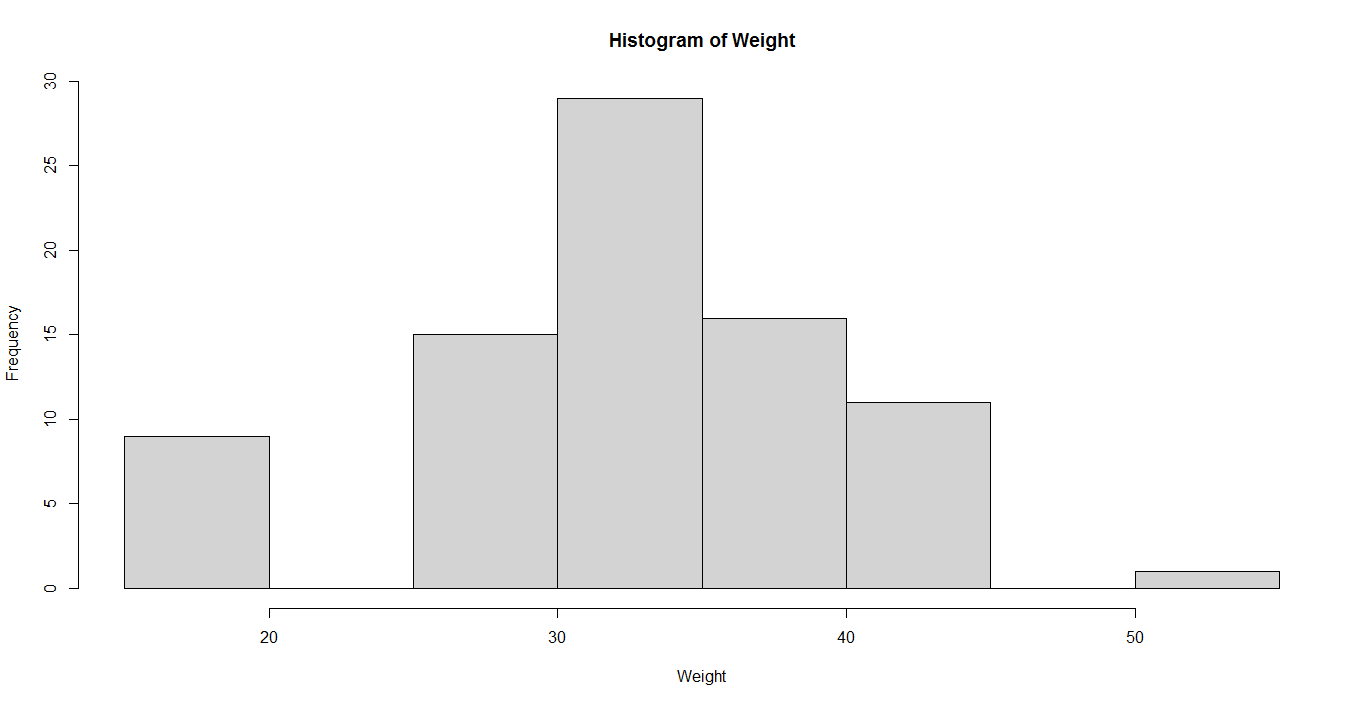
As per above par plot it’s a normal distribution.

**Box Plot:**

****

As per box plot there is a for distance there is 1upper & 9lower out lairs which values are 52.99775 & 15.71286, 15.75353, 15.76963, 15.82306, 15.84776, 16.04317, 16.13295, 16.19412 , 16.35948 and needs to be removed for good distribution.

**Histogram:**

****

As per above histogram we found that frequencies for weight between 25 & 40

is very high & there is a positive skewness and kurtosis along with negative skewness & kurtosis**.**

**Q10) Draw inferences about the following boxplot & histogram**



**Answer:** As per above box plot we can say that chick weight has 7 values

which are out lairs & needs to be removed for good distribution.

As per above histogram we can say that this is positive one sample one tail test for ChickWeight weight distribution & there is a positive skewness & kurtosis.

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

**Answer:** Number of men or Sample n = 2000

Total Population of men = 30,00000

Sample men average weight Xbar = 200pounds

Sample men standard deviation sigma = 30pounds

Confidence Intervals:

For 94% C.I. :

= Xbar + t(1-Alfa, n-1)\*sample standard deviation/root n

= 200 + 1.88 \* 30/root2000

= 200 + 1.88 \* 0.670

= 200 + 1.2596

= 201.26

Xbar - Z(1-Alfa,n-1)\* sample standard deviation /root n

= 200 – 1.88\*30/root2000

= 200 – 1.3596

= 198.64

For 98% C.I. :

= Xbar + t(1-Alfa, n-1)\*sample standard deviation/root n

= 200 + t(0.99,1999)\*30/root2000

= 200 + 2.33\*0.670

= 200 + 1.56

= 201.56

= Xbar - t (1-Alfa, n-1)\*sample standard deviation/root n

= 200 - t (0.99, 1999)\*30/root2000

= 200 - 2.33\*0.670

= 200 - 1.56

= 198.44

For 96% C.I.:

= Xbar + t(1-Alfa, n-1)\*sample standard deviation/root n

= 200 + t(0.98, 1999)\*30/root2000

= 200 + 2.05509\*0.670

= 200 + 1.38

= 201.38

= Xbar - t(1-Alfa, n-1)\*sample standard deviation/root n

= 200 - t(0.98, 1999)\*30/root2000

= 200 - 2.05509\*0.670

= 200 - 1.38

= 198.62

**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean,median,variance,standard deviation.
2. What can we say about the student marks?

**Answer:** As per computation of R we find the mean, median, variance, standard

deviation of student in class are following:

mean = 41

median = 40.5

variance = 25.52941

standard deviation = 5.052664

As per student marks we can say that they are getting good marks in class.

Q13) What is the nature of skewness when mean, median of data are equal?

**Answer:** When mean & median of data is equal then skewness would be zero.

Q14) What is the nature of skewness when mean >median?

**Answer:** If mean>median then distribution is positively skewed.

Q15) What is the nature of skewness when median > mean?

Answer: If median>mean then distribution is negatively skewed.

Q16) What does positive kurtosis value indicates for a data?

**Answer:** A distribution with a positive kurtosis value indicates that the distribution

has heavier tails and a sharper peak than the normal distribution.

Q17) What does negative kurtosis value indicates for a data?

**Answer:** A distribution with a negative kurtosis value indicates that the

distribution has lighter tails than the normal distribution.

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

**Answer:** As per above box-plot visualization data distribution is asymmetric.

What is nature of skewness of the data?

**Answer:** Nature of skewness of data is positive.

What will be the IQR of the data (approximately)?   
**Answer:** The IQR of data is 8’.

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot2.

**Answer:** Following Inferences from the distribution of data for Box-plot 1 with respect Boxplot2:

a] Boxplot1 is Boxplot2 are symmetric.

b] Boxplot1 doesn’t cover all data points and Boxplot2 cover all the data points.

c] Boxplot1 data & Boxplot2 data are equally distributed or

both are neither positively not negative skewed.

d] Boxplot1 have 4no lower (200, 212.5, 225,239) and 3no upper outliers (300,

312.5, 325) in dataset. But Boxplot2 don’t have outliers.

e] Boxplot1 and Boxplot2 have few values which are following:

**For Boxplot1:**

Minimum = 237.5

Maximum = 287.5

Median = 250+275/2

= 262.5

Boxplot1 IQR = 275-250

= 25

Quartile1 = 250

Quartile3 = 275

Five Number summery = 237.5, 250, 262.5, 275,287.5

**For Boxplot2:**

Minimum = 200

Maximum = 325

Median = 225+300/2

= 262.5

Boxplot2 IQR = 300-225

= 75

Quartile1 = 225

Quartile3 = 300

Five Number summery = 200,225,262.5,300,325

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

**Answer:** When P (MPG>38) = 49/81

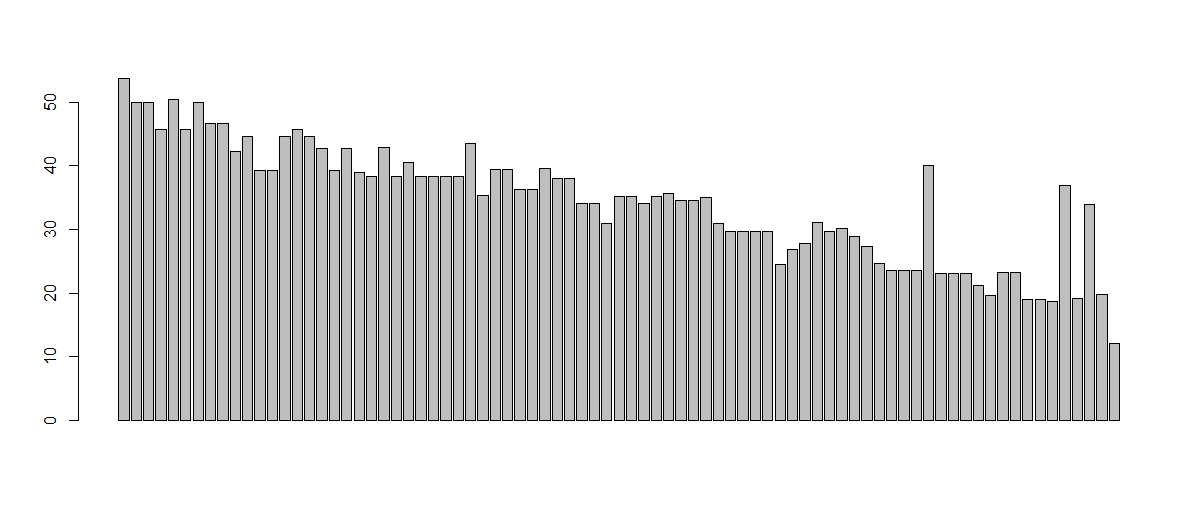
When P (MPG<40) = 63/81

When P (20<MPG<50) = 69/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

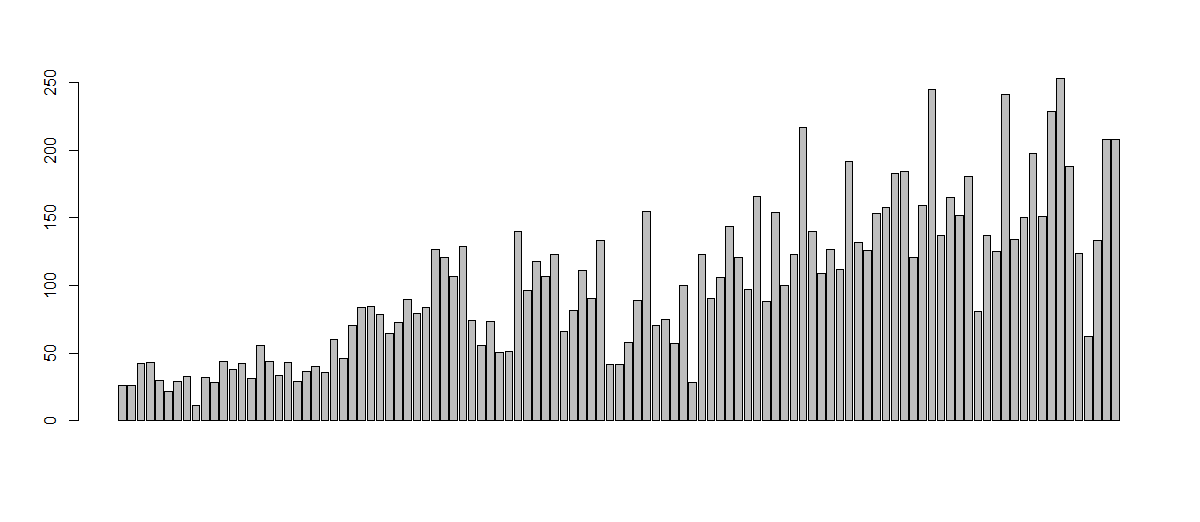


**Answer:** As per above bar plot and attached R file for MPG for cars It

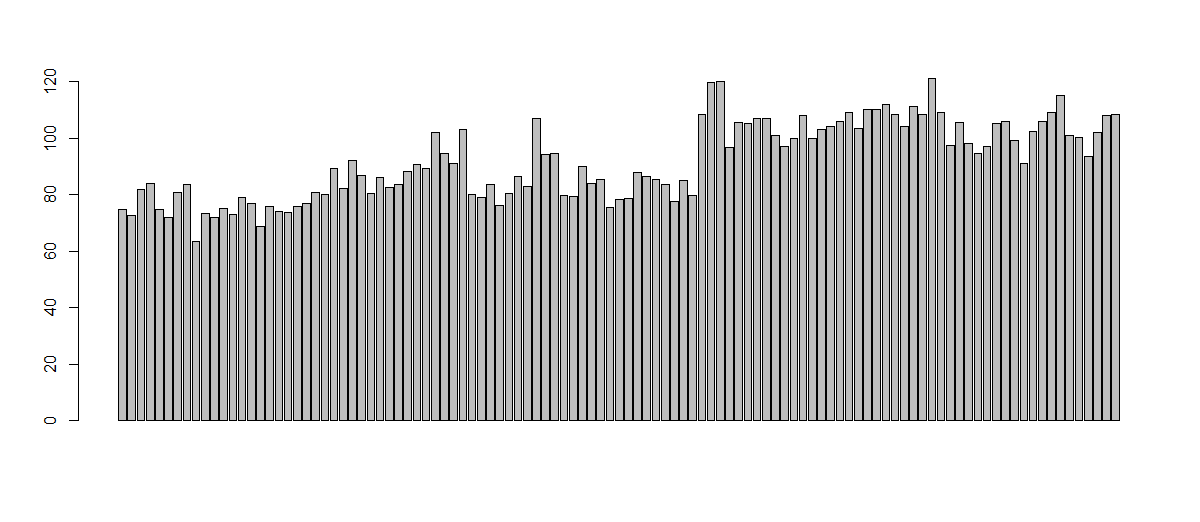
doesn’t follows the normal distribution.

1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv



Adipost Tissue Bar Plot



Waist Circumference Bar Plot

**Answer:** As per above both bar plots and attached R file the Adipose Tissue

(AT) and Waist Circumference (Waist) from wc- at data set follows

Normal Distribution.

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Answer:** **Z score of 90%, 94% & 60% confidence interval:**

**For 90% C.I.:**

= C.I. % + alpha/2

= 0.90+ 0.10/2

= 0.90 + 0.05

= 0.95

Compute qnorm(0.95) in R then get the Z score 1.645

as per attached R file.

**For 94% C.I.:**

= C.I. % + alpha/2

= 0.94 + 0.06/2

= 0.94 + 0.03

= 0.97

Compute qnorm(0.97) in R then get the Z score 1.88

as per attached R file.

**For 60% C.I.:**

= C.I. % + alpha/2

= 0.60 +0.40/2

= 0.60 + 0.2

= 0.80

Compute qnorm(0.80) in R then get the Z score 0.842

as per attached R file.

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence

interval, 99% confidence interval for sample size of 25.

**Answer:**

No. of sample n = 25

t scores of confidence intervals:

For 95% confidence interval:

= C.I.% + alpha/2

= 0.95 + 0.05/2

= 0.95 + 0.025

= 0.98

Compute qt(0.98, 24) in R then get the t score is 2.17 as per attached R file code.

For 96% confidence interval:

= C.I.% + alpha/2

= 0.96 + 0.04/2

= 0.95 + 0.02

= 0.97

Compute qt(0.97, 24) in R then get the t score is 1.974 as per attached R file code.

For 99% confidence interval:

= C.I. % + alpha/2

= 0.99 + 0.01/2

= 0.99 + 0.005

=0.995

Compute qt(0.995, 24) in R then get the t score is 2.80 as per attached R file code.

Q 24**)**A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode🡪pt(tscore,df)

df 🡪 degrees of freedom

**Answer:** R-code: p <- pt((260-270)/(90/sqrt(18)), 17)

t-Test formula applied for tscore:

t= xbar – mu\ S/root(n)

Probability = 0.3216725