|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| 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 | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

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

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

Ans- Total occurance (S)-{HHH,HHT,HTH,THH,TTT,TTH,THT,HTT}

Occurance of an Event-{ HHT,HTH,THH}

Probablity of getting two head and one tail =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 2 and 3

Ans- Total Occurance –36

1. Probablity of getting sum equal to 1=0.
2. Number of getting less or equal to 4 =6. {11,12,13,21,22,31}

Probablity of getting less or equal to 4= 6/36.

c)Number of getting Sum is divisible by 2 and 3 =6. {15,24,33,42,51,66}

Probablity of getting Sum is divisible by 2 and 3=6/36.

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?

Ans- Total no of events: ncr = 7c2=21

None of the balls is blue=5c2=10

Probability of none of the balls drawn is blue=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

Ans: Expected number of candies for a randomly selected child =∑xP(x)

=1\*0.015+4\*0.20+3\*0.65+5\*0.05+6\*0.01+2\*0.120

=3.09

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**

Ans:

|  |  |  |  |
| --- | --- | --- | --- |
|  | point | score | weigh |
| Mean | 3.60 | 3.22 | 17.85 |
| Median | 3.70 | 3.33 | 17.71 |
| Mode | 3.07 | 3.44 | 17.02 |
| Variance | 0.28 | 0.93 | 3.09 |
| St. dev | 0.53 | 0.96 | 1.76 |
| Range | 2.76 – 4.93 | 1.513 – 5.424 | 14.5 – 22.9 |

Inference Drawn:

The mean is the most common measure of central tendency.

The median is a very useful statistics data as it is not impacted by extreme cases, but the mean is.

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?

Ans: The Expected Value of the Weight of that patient = ∑Xi/N

=108+110+123+134+135+145+167+187+199/9=1308/9=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Ans: For speed column:

>library(moments)

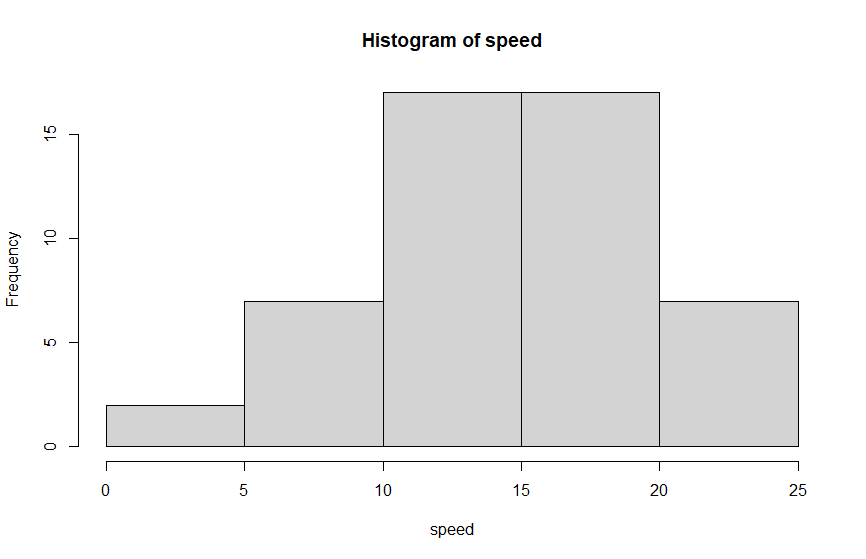
>attach(Q9\_a)

>skewness(speed)

-0.1139548 (left skewness)

>kurtosis(speed)

2.422853 (negative kurtosis)



Skewness is negative i.e the distribution is skewed towards left. Mean of distribution is less than the Median. Kurtosis Value says the distribution has broad peak and thin tails.

For dist column

>library(moments)

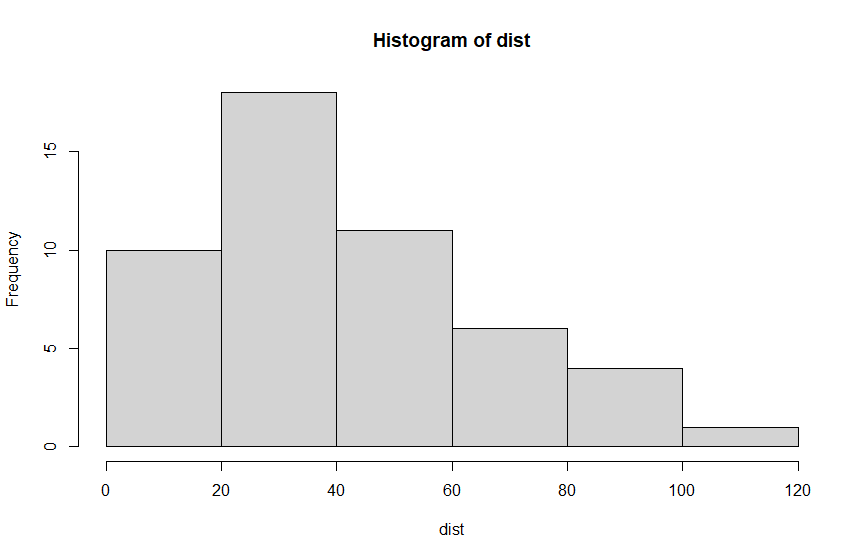
>attach(Q9\_a)

>skewness(dist)

-0.1139548 (right skewness)

>kurtosis(dist)

2.422853 (positive kurtosis)



Skewness is positive i.e the distribution is skewed towards right. Mean of distribution is more than the Median. Kurtosis Value says the distribution has sharp peak and wide tails.

**SP and Weight(WT)**

**Use Q9\_b.csv**

Ans:For SP:

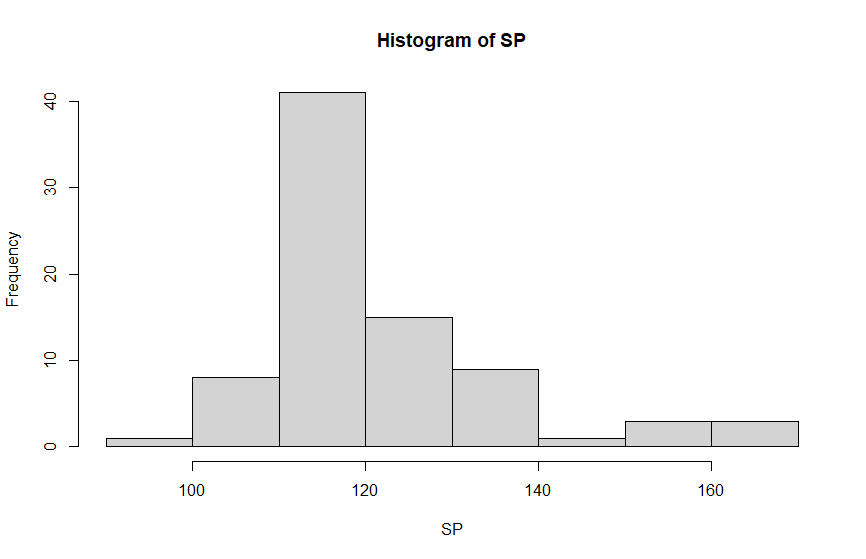
>attach(Q9\_b)

>skewness(SP)

1.581454 (

>kurtosis(SP)

5.723521

Skewness is positive i.e the distribution is skewed towards right. Mean of distribution is more than the Median. Kurtosis Value says the distribution has sharp peak and wide tails.

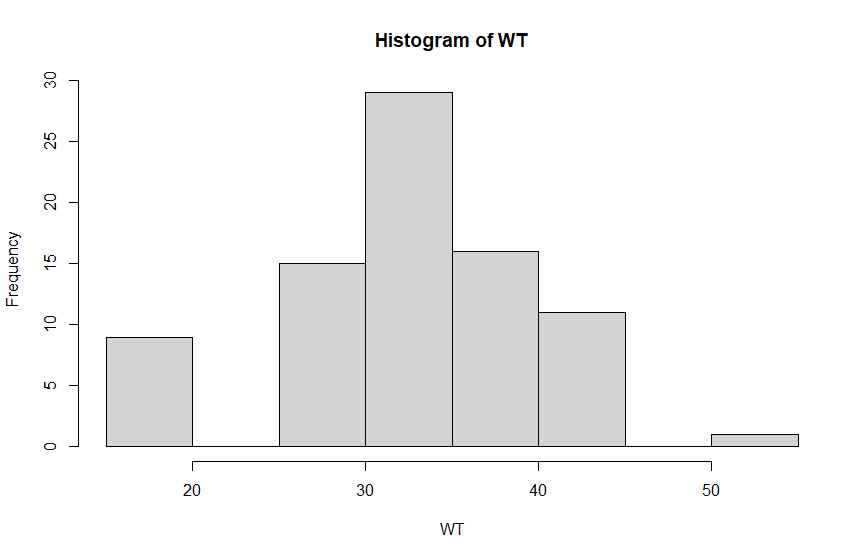
For WT:

>skewness(WT)

-0.6033099

>kurtosis(WT)

3.819466



Skewness is negative i.e the distribution is skewed towards left. Mean of distribution is less than the Median. Kurtosis Value says the distribution has sharp peak and wide tails.

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



The distribution is Right skewed.



The distribution has lots of outliers towards upper extreme**.**

**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?

Ans: Population = 3000000

Sample=2000

Sample mean = 200

Sample standard deviation = 30

CI = X(bar)± t(1-alpha,n-1)\*s/**√n**

**When confidence interval is 96% :**

**=200**±2.05\*30/**√2000=198.62-201.38**

**When confidence interval is 98% :**

**=200**±2.33\*30/**√2000=198.44-201.56**

**When confidence interval is 94% :**

**=200**±1.88\*30/**√2000=198.74-201.26**

**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.

Ans: Mean=41

Median=40.5

Variance=25.53

Standard deviation=5.05

1. What can we say about the student marks?

Ans: Mean is greater than Median so the distribution is skewed towards right and no outliers are present.

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

Ans: Symmetric in nature,skewness=0.

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

Ans: Right skewed

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

Ans: left skewed

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

Ans: Sharp peak and Thick tails

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

Ans: Broad peak and Thin tails

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



What can we say about the distribution of the data?

Ans: Not a Normal Distribution

What is nature of skewness of the data?

Ans: Left Skewed

What will be the IQR of the data (approximately)?

Ans: 18-10=8

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Ans: Both the distributions are Normally Distributed. No outliers are present.

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

Ans: >mean(MPG)

34.4

> sd(MPG)

9.1

* 1. P(MPG>38)

>pnorm(38,34.4,9.1)

0.65

>1-0.65

0.35

* 1. P(MPG<40)

>pnorm(40,34.4,9.1)

0.73

c. P (20<MPG<50)

>pnorm(50,34.4,9.1)-pnorm(20,34.4,9.1)

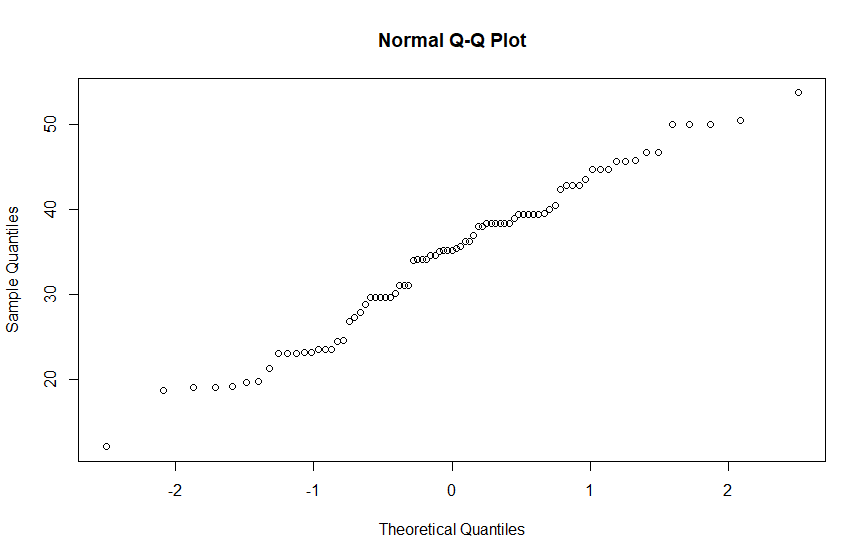
0.89

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:

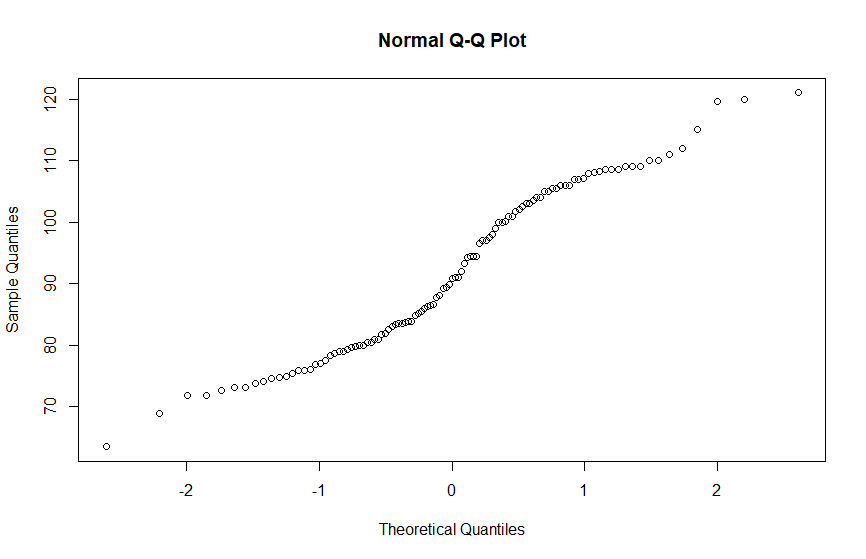


Follows normal distribution as shown in above plot.

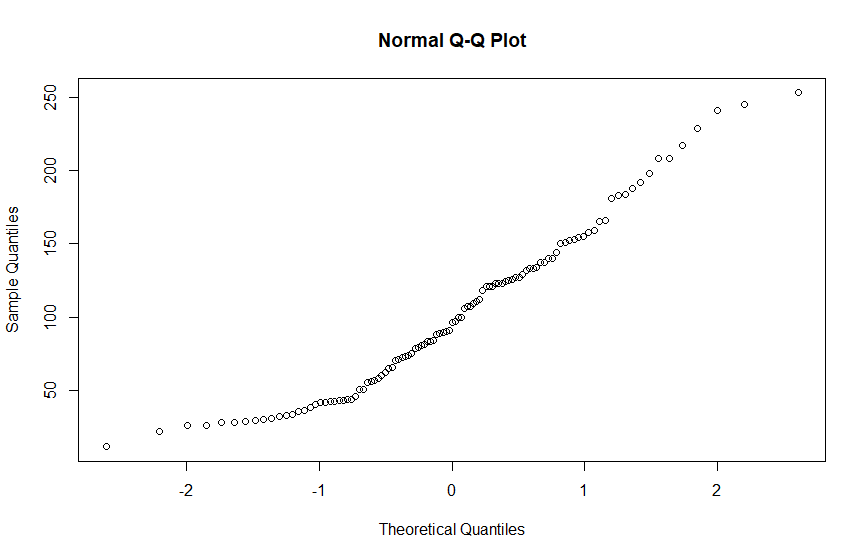
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans: Waist Circumference(Waist) follows normal distribution.



Adipose Tissue (AT) also follows normal distribution.



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

Ans: 90% confidence interval

>qnorm(0.95)

1.6445

94% confidence interval

>qnorm(0.97)

1.8807

60% confidence interval

>qnorm(0.80)

0.8416

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25?

Ans: 95% confidence interval

>qt(0.975,df=24)

2.063899

96% confidence interval

>qt(0.98,df=24)

2.171545

99% confidence interval

>qt(0.995,df=24)

2.79694

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

Ans: µ = 270, n= 18, x bar = 260, s = 90

t score= (x bar - µ)/(s/sqrt(n))

=(260-270)/(90/sqrt(18))

=-10/21.23

=-0.47

>qt(-0.47,df=17)

0.3221639

Required probability is 0.32.