Q1) Identify the Data type for the Following:

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

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)	Interval
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	Interval
SAT Scores	Interval
Years of Education	Ratio

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

Ans: Total outcome = 8

The outcomes are (HHH), (HTH), (HHT), (HTT), (THH), (TTH), (TTT), (THT)

Two heads and one tail are obtained in 3 cases, (HTH), (HHT), (THH)

So, probability that two heads and one tail are obtained = 3/8

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

a) Equal to 1
b) Less than or equal to 4
c) Sum is divisible by 2 and 3

Ans: Total possible outcome = 36

Total Events

{(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)} = 36

a) 0/36 i.e., zero probability, because when two dice are rolled then minimum sum we can get is 2 and the maximum sum will be 12. Therefore, the probability of getting a sum equal to 1 is zero.

b) The possible combination for sum less than or equal to 4 are (1,1), (1,2), (2,1), (1,3), (3,1) and (2,2). So the probability is 6/36=1/6

c) Probability for sum divisible by 2 and 3 is, 6/36 =1/6 because, the sum divisible by 2 and 3 are 6 and 12.The combinations for each pf these sum are (1,5), (2,4), (3,3), (4,2), (5,1), (6,6) Therefore required probability is 1/6.

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 balls = 7

The total number of ways to draw 2 balls from 7 balls = 7!/2!*(7-2)!

=21

The total number of ways to draw 2 balls, but none of them is blue

=5!/5!*(5-2)!

=10

Probability that 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:(1*0.015) + (4*0.20) + (3*0.65) + (5*0.005) + (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:

	Mean	Median	Mode	Variance	Standard deviation	Range
Points	3.596525	3.695	3.92	0.285881	0.53467873	2.17
Score	3.21725	3.325	3.44	0.957378	0.97845736	3.911
Weigh	17.84875	17.71	17.8	3.193166	0.97845744	8.3999

Q8) Calculate Expected Value for the problem below

a) 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: Expected Value of the Weight of the random patient = 145.3

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

Cars speed and distance

Use Q9_a.csv

Ans: skewness

Speed = -0.113955

distance = 0.782484

Kurtosis

Speed = -0.577147

Distance = 0.4050

Conclusion:

- Speed is Negatively skewed
- Distance is positively skewed
- Speed is Leptokurtic
- Distance is Leptokurtic

SP and Weight(WT)

Use Q9_b.csv

Ans: Skewness

SP = 1.581454

Weight = -0.603310

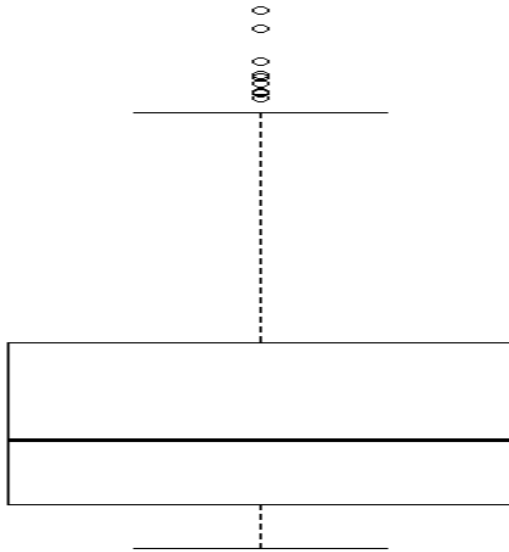**Kurtosis**

SP = 2.723521

Weight = 0.819466

Conclusion:

- SP is Positively skewed
- WT is Nedatively skewed
- SP is Leptokurtic
- Distance is Leptokurtic

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight
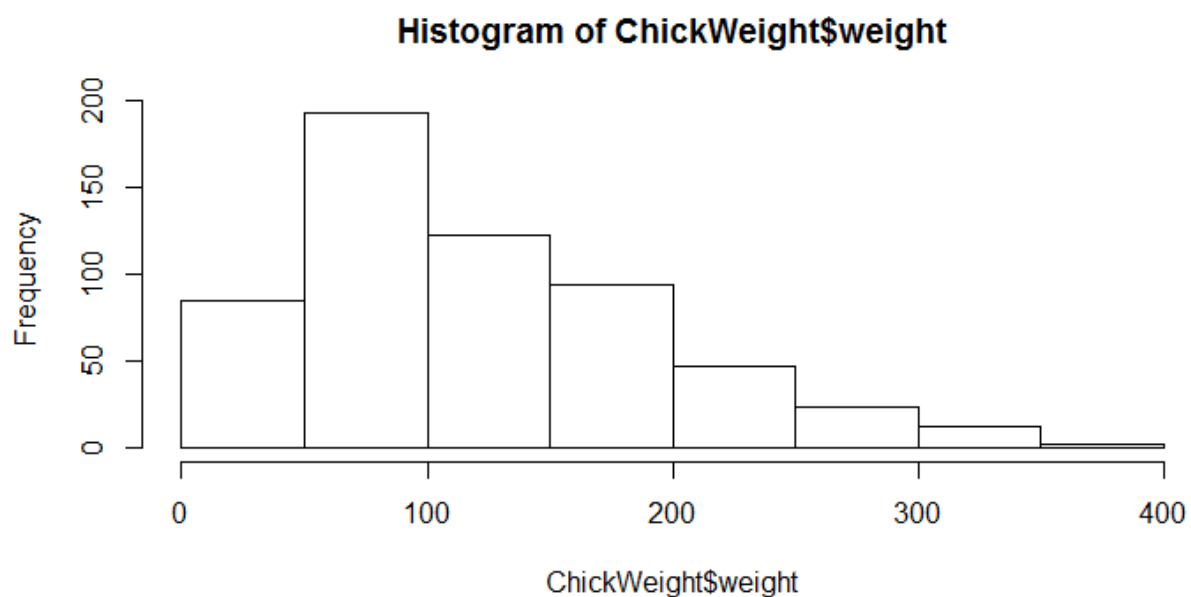
Ans: The histogram shows the greatest frequency in wight occurs between 150-200 and the weight that will least likely occur is between 350-400

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 : 94% Confidence Interval: (198.738325292158, 201.261674707842)

98% Confidence Interval: (198.43943840429978, 201.56056159570022)

96% Confidence Interval: (198.62230334813333, 201.37769665186667)

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?

Ans:1) Mean = 41

Median = 40.5

Variance = 25.52

Standard deviation = 5.052

2) 1. Average mark obtained by student is 41

2. There is no outlier is preset

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

Ans: skewness is zero

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

Ans: Right skewed

Q15) What is the nature of skewness when median > mean?
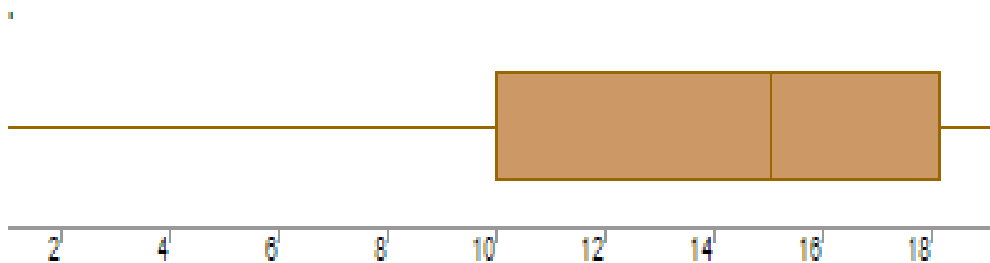
Ans: Left skewed

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

Ans: High peak at the center of data

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

Ans: Wider peak at the center of the data and thinner tails

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

What can we say about the distribution of the data?
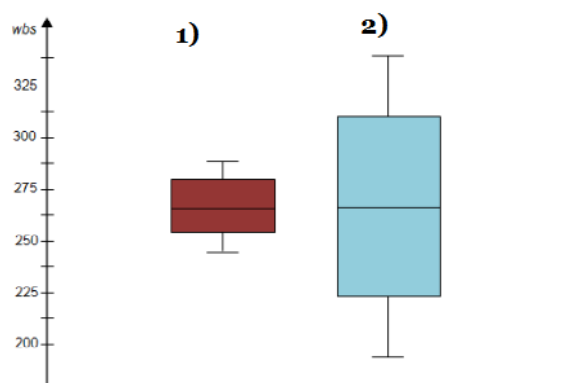
Ans: It is not normally distributed

What is nature of skewness of the data?

Ans: Its left skewed

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

Ans: IQR is 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: a) Both boxplots have same median approximate 262

     b) Both boxplots don't have outliers

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

      a. P(MPG>38)
      b. P(MPG<40)
      c. P (20<MPG<50)

Ans: a)0.347891
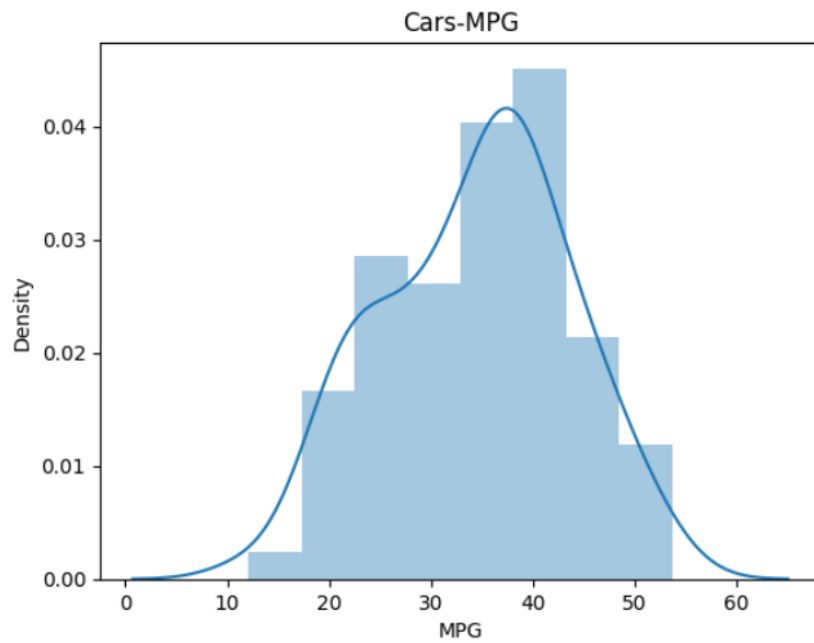
      b)0.729094

      c)0.898941

Q 21) Check whether the data follows normal distribution
     a) Check whether the MPG of Cars follows Normal Distribution
       Dataset: Cars.csv

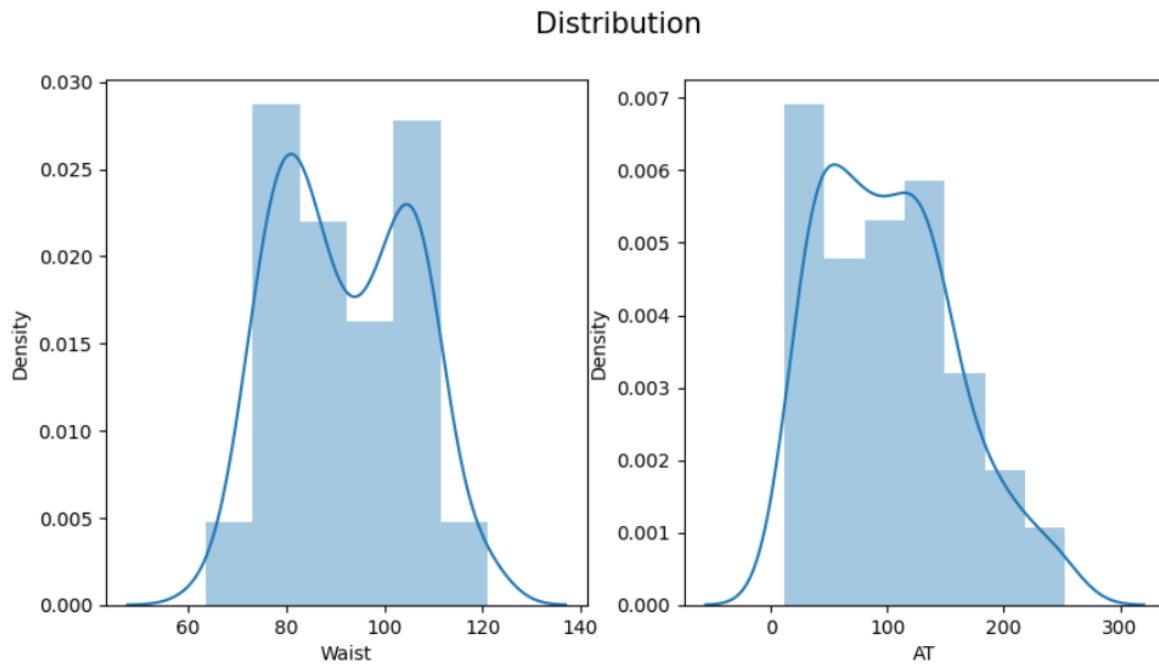**Ans:**

Cars-MPG

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

Ans:

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

Ans: Z score of 90% confidence interval is 1.6449

Z score of 94% confidence interval is 1.8808

Z score of 60% confidence interval is 0.8416

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

Ans: 95.0% Confidence Interval T-Score: 2.0639

96.0% Confidence Interval T-Score: 2.1715

99.0% Confidence Interval T-Score: 2.7969

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 , x = 260 , n = 18, s = 90

t score = (x – μ)/(s/sqrt(n))

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

=-10/21.23

=-0.47

Required probability = 0.319