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	Nominal
Level of Agreement	Ordinal
IQ(Intelligence Scale)	Interval
Sales Figures	Ratio
Blood Group	Nominal
Time Of Day	Ordinal
Time on a Clock with Hands	Interval
Number of Children	Ratio
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. 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. a) probability that sum is equal to 1

P(1) = 0

b) When two dice are rolled then possibility of getting number 4 is (1,1) ,(1,2),(1,3),(2,1), (2,2), and (3,1) so,

The no. of favorable outcomes = 6

Total No. of possibilities = 36

Probability = 6/36 = 1/6

Thus,1/12 is the probability of rolling two dice and getting a sum of 4.

C) Probabilities of sum is divisible by 2 and 3
Total no. of possible outcomes = 36
Favorable outcomes = (1,5) , (3,3) , (4,2) ,(2,4), (5,1), (6,6) = 6
Probability = No. of favorable outcome/No. of possible outcomes
= 6/36
= 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 No. of balls = (2+3+2) = 7

n(S) = (7*6)/ (2*1) = 21

n(E) = (5*4)/ (2*1) = 10

Probability of none of the balls draw is blue = p(E)= n(E)/n(S)

= 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 no. of candies for a randomly selected child = ((1*0.015) + (4*0.20) + (3*0.65) + (5*0.005) + (6*0.01) + (2*0.120))

= 3.090

= 3.09

Expected no. of candies for randomly selected child is 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. Name	Point	Score	Weigh
Mean	3.62	3.22	17.85
Median	3.79	3.32	17.71
Mode	4.08	3.44	18.9
Variance	0.286	0.9574	3.19
Standard Deviation	0.535	0.9784	1.78
Range	2.17 , 4.93	1.513 , 5.424	14.5 , 22.9

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. There are 9 patients at a clinic

Probability of selecting each patient = 1/9

Expected Value = (1/9) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) (1308)

= 145.33

Expected Value is 145.33

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

Ans. Skewness and Kurtosis of data Car speed and distance and SP and Weight are given below :

Car speed and distance data :

Skewness = -0.114 (car speed) , 0.782(distance)

Kurtosis = -0.51(car speed) , 0.41(distance)

SP and Weight(WT) data

Skewness = 1.61 (SP) , -0.615(weight)

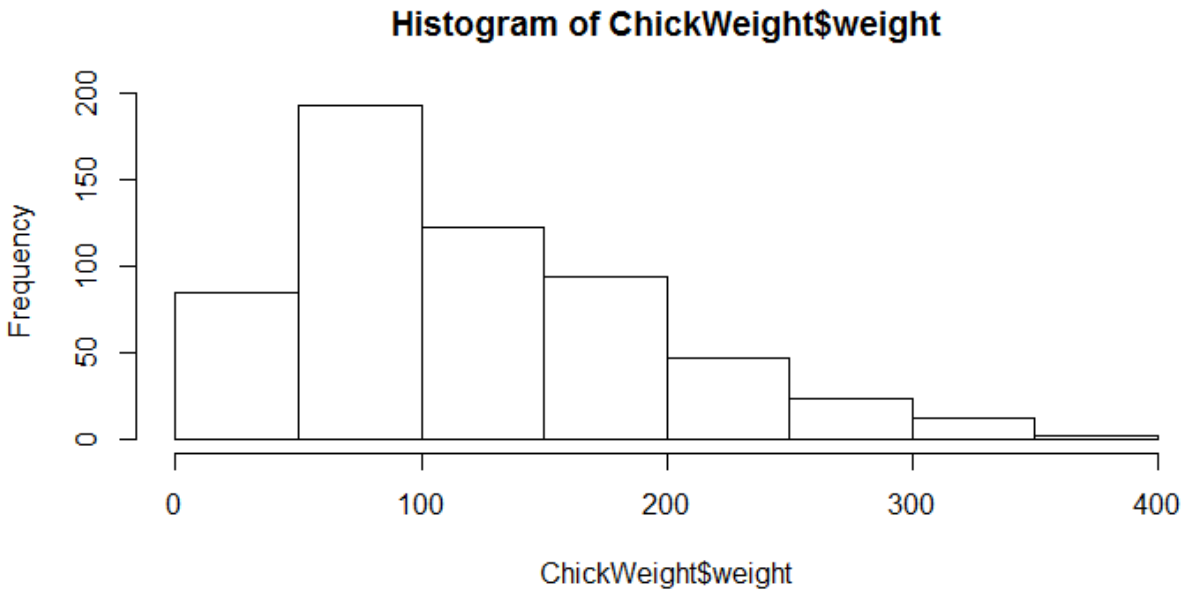
Kurtosis = 2.98 (SP), 0.95(weight)

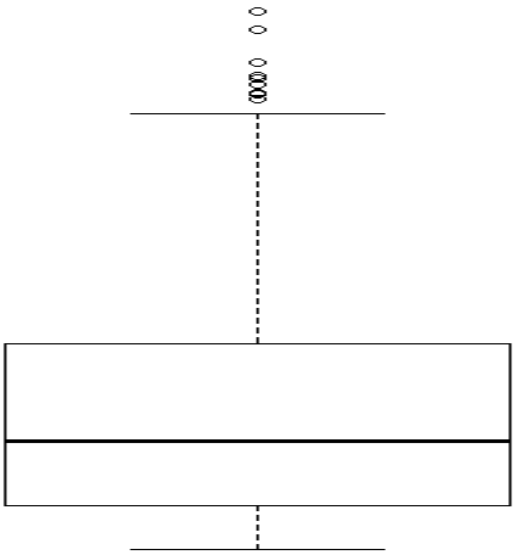
1. In given data of car speed and distance the skewness value of car speed is negative so, the data are negatively skewed or skewed left, meaning that the left tail is longe, and In kurtosis case car speed is negative so, simply means that more data values are located near the mean and less data values are located on the tails. negative kurtosis is the uniform distribution, which has no peak at all and is a completely flat distribution.

In Distance data skewness value is positive so,the mean and median is greater than mode.In kurtosis case of distance data is positive so, simply means that

fewer data values are located near the mean and more data values are located on the tails.

Q10) Dra inferences about the following boxplot & histogram





Ans. In above figure histogram of ChickWeight$weight is shown. In this figure x axis shown chickweight$weight and in yaxis shown frequency.In above histogram figurer and boxplot figure is positively skewed on right side.That means Mean and median of the data is greater than mode.

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. Using the t-distribution it is found that :

Sample Mean = 200

Sample standard deviation = 30

Sample size = 2000

1. The 94% confidence interval is (198.73 , 201.39) (t= 1.8916)
2. The 96% confidence interval is (198.61 , 201.39) (t= 2.0623)
3. The 98% confidence interval is (198.43 , 201.57) (t= 2.3452)

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. From above given of student score obtained in test :

Mean = 41, Median = 40.5 , Variance = 25.529 , Standard Deviation = 5.05

2.Mass of students marks between 38-42. Skewness(1.52) is positive because mass of marks in left side of plot.

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

Ans. If the mean is equal to the median as well as the mode, hence the skewness is zero. If the distribution is symmetric, the mean equals the median, and the skewness of the distribution is zero.

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

Ans. If the mean is greater than the median, the distribution is positively skewed. If the mean is less than the median, the distribution is negatively skewed.

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

Ans. If the mean is greater than the median, the distribution is positively skewed. if the mean is less than the median, the distribution is negatively skewed.

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

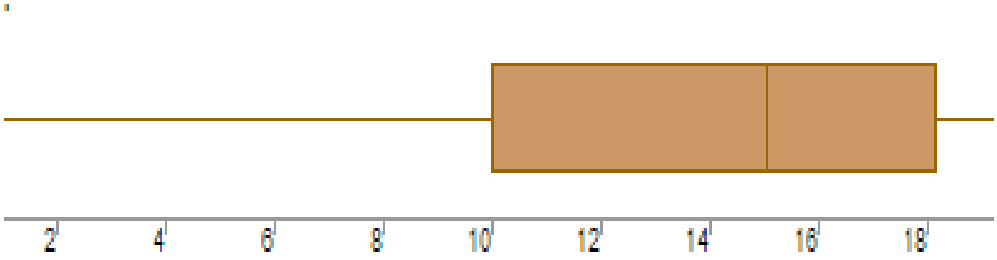
Ans. Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. An extreme positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

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

Ans. If a distribution has negative kurtosis, it is said to be platykurtic**,** which means that it has a flatter peak and thinner tails compared to a normal distribution. This simply means that more data values are located near the mean

and less data values are located on the tails. negative kurtosis is the uniform distribution, which has no peak at all and is a completely flat distribution.

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



What can we say about the distribution of the data?

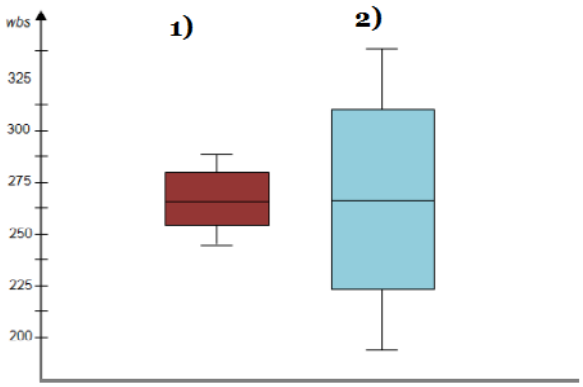
What is nature of skewness of the data?

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

Ans. In above boxplot visualization

1. A negatively skewed distribution is a type of distribution in which more values are concentrated on the right side (tail) of the distribution graph while the left tail of the distribution graph is longer.
2. The nature of skewness of the data is the Mean of negatively skewed data will be less than the Median.
3. The IQR describes the middle 50% of values when ordered from lowest to highest. In above example of data the IOR = (-10-18)

Q19) Comment on the below Boxplot visualizations?



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

Ans. In above figure of boxplot. Boxplot 1 is positively skewed that's means Mean and Median is greater than Mode. And In boxplot 2 is shows normal distribution that means the skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero.

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. From Given Data of cars

MPG <-CarsMPG

a. P(MPG>38)-
1- pnorm(38,34.422,9.13144) = 0.3475908
b. P(MPG<40)-
p-norm(40,34.422,9.13144)= 0.7293527
c. P (20<MPG<50)-
p-norm (50,34.422,9.13144) - (1-pnorm(20,34.422,9.13144))= 0.01311818

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution
Dataset: Cars.csv

Ans. In above data MPG of Cars is highly negative skewed.

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

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

Ans. Confidence interval Z scores

Z score of 60%	=	0.8416212
Z score of 90%	=	1.644854
Z score of 94%	=	1.880794

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

Ans. Confidence Interval t score

95%	2.063899
96%	2.171545
99%	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. 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

`Solution= 0.3217`