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	NOMINAL
Celsius Temperature	INTERVAL
Weight	RATIO
Hair Color	RATIO
Socioeconomic Status	INTERVAL
Fahrenheit Temperature	RATIO
Height	RATIO
Type of living accommodation	ORDINAL
Level of Agreement	INTERVAL
IQ(Intelligence Scale)	INTERVAL
Sales Figures	INTERVAL
Blood Group	RATIO
Time Of Day	INTERVAL
Time on a Clock with Hands	INTERVAL
Number of Children	NOMINAL
Religious Preference	RATIO

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:-

Probability if getting the H and T for each coin is $\frac{1}{2}$

$= P(H\,H\,T) + (H\,T\,H) + (T\,H\,H)$

$= \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2}\right) + \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2}\right) + \left(\frac{1}{2} + \frac{1}{2} + \frac{1}{2}\right)$

$= \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$

$= \frac{3}{8}$

So , The probability of getting Two heads and one tail is $\frac{3}{8}$

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 of getting the sum equal to 1 is Zero.**

b) $(1 , 1) (1 , 2) (1 , 3) (2 , 1) (2 , 2) (3 , 1)$

$$P = \frac{Number\ of\ favourable\ outcomes}{Total\ number\ of\ possible\ outcomes}$$

$$P = \frac{6}{36}$$

Probability of getting the sum is less than or equal to 4 is $\frac{1}{6}$

c) The no which is Divisible by 2 & 3 is 6 and 12
$(1 , 5) (2 , 4) (3, 3) (4 , 2) (5 ,1) (6 , 6)$

$$P = \frac{Number\ of\ favourable\ outcomes}{Total\ number\ of\ possible\ outcomes}$$

$$P = \frac{6}{36}$$

Probability of getting the sum which is divisible by 2 & 3 is $\frac{1}{6}$.

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:-

$$P = \frac{Number\ of\ favourable\ outcomes}{Total\ number\ of\ possible\ outcomes}$$

Number of favourable outcomes $= \frac{5*4*3}{2*3} = 10$

Total number of possible outcomes $= \frac{7*6*5}{2*5} = 21$

probability that none of the balls drawn is blue is $\frac{10}{21}$

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:-

No. of candies = 1*0.015 + 4*0.20 + 3*0.65 + 5*0.005 + 6*0.01 + 2*0.120 = 3.09

Expected number of candies for a randomly selected child are 3.09

Q7) Calculate <u>Mean, Median, Mode, Variance, Standard Deviation, Range</u> & comment about the values / draw inferences, for the given dataset
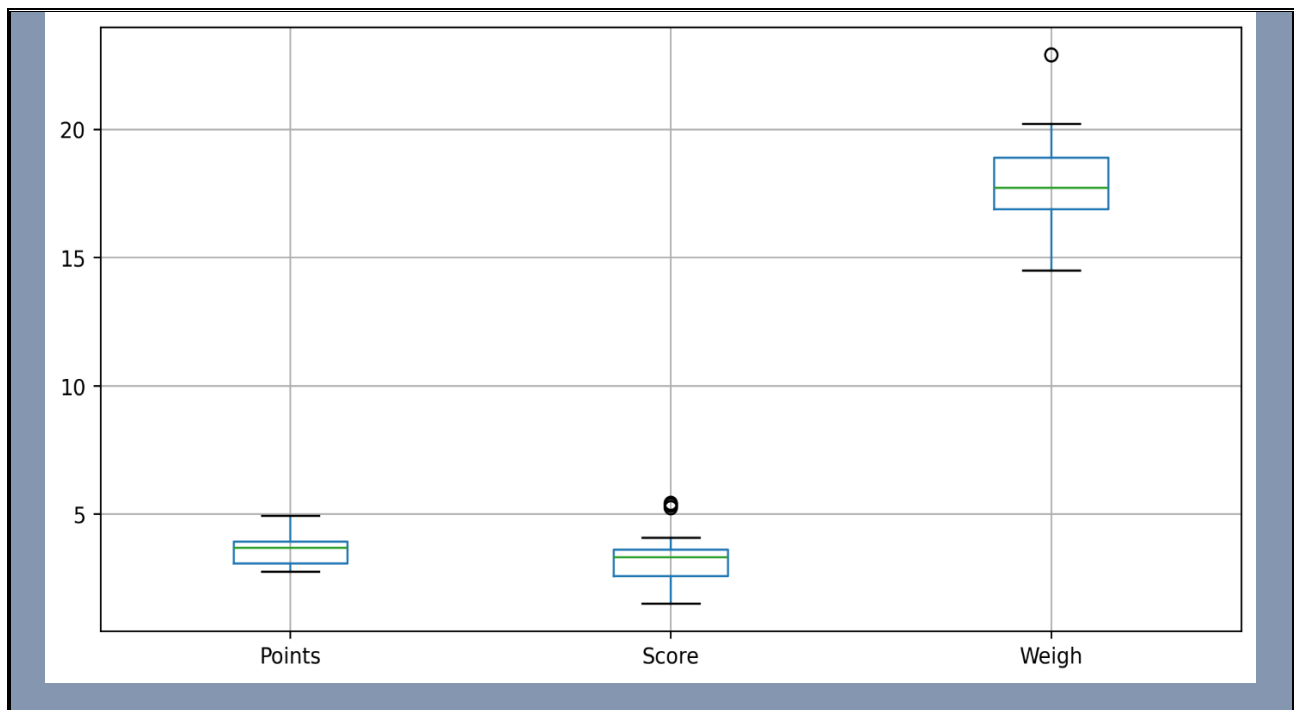
- For Points, Score, Weigh >
  Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

Use Q7.csv file

ANS:- <u>Click Here..</u>   **Please click the Ctrl + Click.**

Sr.No	Points	Score	Weigh
Mean	3.596	3.217	17.848
Mode	3.92	3.44	18.90
Median	3.695	3.325	17.71
Varience	0.276	0.927	3.093
Standard Deviation	0.526	0.963	1.758
Range	2.170	3.911	8.400

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:-

$$Mean = \frac{\sum(x)}{N}$$

$$Mean = \frac{108+110+123+134+135+145+167+187+199}{9}$$

$$Mean = 146.55 \cong 147$$

Expected Value of the weight of that patient is :- 147

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

Cars speed and distance   (Use Q9_a.csv)

ANS:-   Click here..   **Please click the Ctrl + Click.**

Sr.No	Cars Speed	Distance
Skewness	-0.117510	0.806895
Kurtosis	-0.508994	0.405053

SP and Weight(WT) **(Use Q9_b.csv)**

Click here... For ANS

Sr.No	Sp	Weight
Skewness	1.611450	−0.614753
Kurtosis	2.977329	0.950291

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight
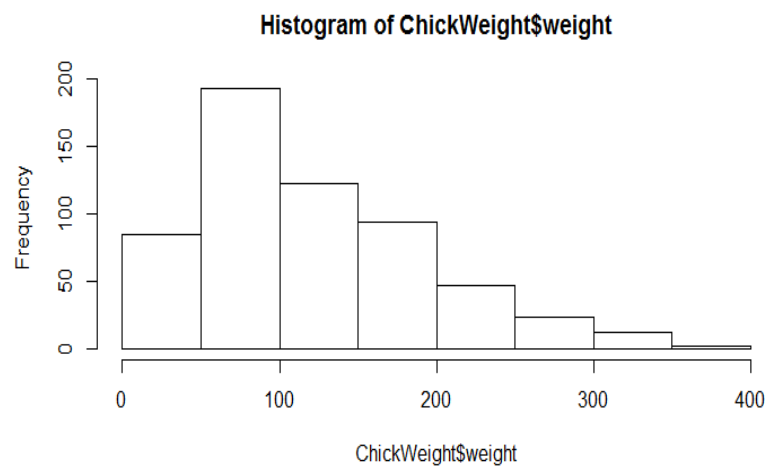
ANS:-

Here the most of the data points are concentrated in the range of 50 – 150 with the frequency of 200 & least range of data in the range of 300 – 400 with the frequency of 0 – 30.

We can clearly see that the right tail is longer than the left tail i.e the histogram is right skewed.

ANS:-
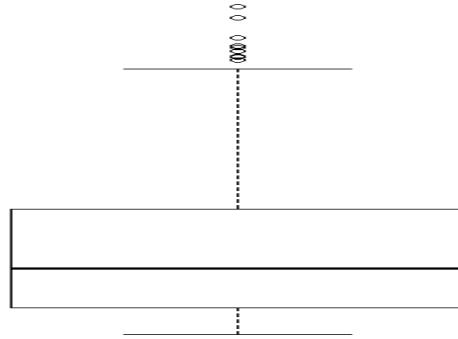
Medican is less than mean right skewed and we have outlier on the upper side of box plot and there is less data points between Q1 and bottom point.

So we can say that there is presence of the Outliers.

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:-

Click here.....For Ans

For 94% = (198.738325292158, 201.261674707842)

For 96% = (198.62230334813333, 201.37769665186667)

For 98% = (198.43943840429978, 201.56056159570022)

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:-

Click here...   For ANS

Mean	= 41
Median	= 40.5
Variance	= 24.11
Standard Deviation	= 4.91

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

ANS:-

Symmetrical

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

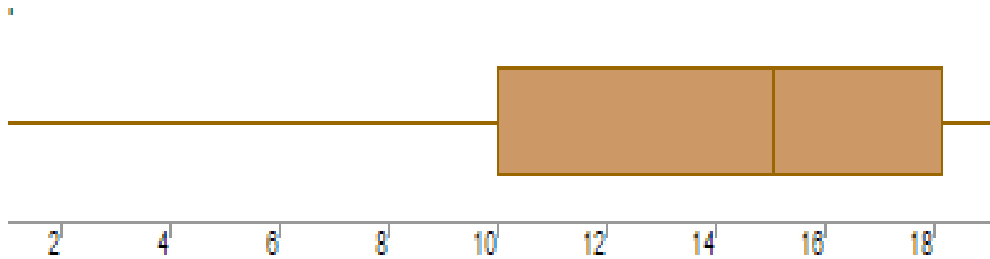Q16) What does positive kurtosis value indicates for a data ?

ANS:-

> *The data is normally distributed and kurtosis value is 0.*

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

ANS:-

> *The distribution of the data has lighter tails and flatter peaks than normal distribution.*

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

1. What can we say about the distribution of the data?

ANS:-   *We can say that here the data is not normally distributed.*
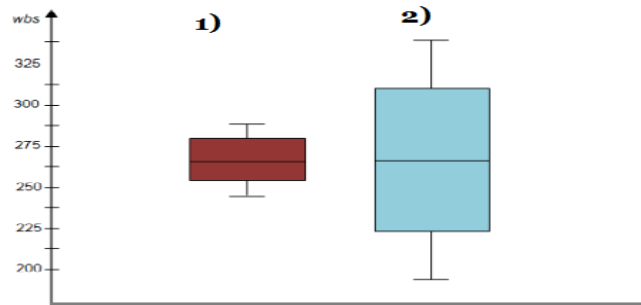
2. What is nature of skewness of the data?

ANS:-   *Left skewed , median is greater than mean.*

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

ANS:- $IQR = (Q_3 - Q_1) = (18 - 10) = (- 8)$

Q19) Comment on the below Boxplot visualizations?

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

ANS:-

The whiskers of the Boxplot No.2 is longer than the Boxplot No.1 and the median values of both the boxplots are exactly same which means the distribution of these two boxplots are Symmetrical

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:-   Click here.... For ANS

Following are the probabilities for the different conditions.

1) P(MPG > 38)   :- 0.347   =35%
2) P(MPG < 40)   :- 0.729   =72%
3)P(20 < MPG < 50) :- 0.898   =90%

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

ANS:-   Click here....For ANS

After plotting the Boxplot , Histogram and the q-q plot it is clear that the data is normally distributed and there is not presence of any outlier data points.

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

ANS:-

Click here....

After plotting the Boxplot , Histogram and the q-q plot it is clear that the data is not normally distributed . Even after the absence of the outliers the data is not normally distributed.

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

Z-Score of 90% Confidence interval is = 1.65

Z-Score of 90% Confidence interval is = 1.55

Z-Score of 90% Confidence interval is = 0.85

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

ANS:-

> *T-Score of 90% Confidence interval is = 1.96*
>
> *T-Score of 90% Confidence interval is = 2.5*
>
> *T-Score of 90% Confidence interval is = 2.47*

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:

ANS:-

Click here.....for ANS

> *Probability that 18 randomly selected bulbs would have an average life of no more than 260 days is : 32%*