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	Continuous
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	Interval
Time on a Clock with Hands	Interval
Number of Children	Ratio
Religious Preference	Categorical
Barometer Pressure	Ratio
SAT Scores	Interval
Years of Education	Ordinal

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

Q3_Solution:

Given,

Three Coins are tossed

Therefore, Total Number of outcomes $= 2^3$

$= 8$

$$\left[(HHH)\ (HHT)\ (HTH)\ (THH)\ (TTT)\ (TTH)\ (THT)\ (HTT) \right] - \text{Total outcomes}$$

Probability of obtaining Two heads and one Tail $= ?$

Let $p(E) = ?$

$\therefore\ p(E) = \dfrac{No.\ of\ favourable\ outcomes}{Total\ No.\ of\ outcomes}$

$p(E) = 3/8$

$\therefore p(E) = 0.375$

Therefore, probability of obtaining Two heads and one tail when Three coins are tossed is $\boxed{0.375}$

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

Q_4 - Solution :-

Given, Two dices are rolled.

    Therefore, Total Number of outcomes $= 6^2 = 36$.

a) Sum is equal to 1

$\therefore P(E) = 0/36$

$\therefore P(E) = 0$

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

b) Less than (or) Equal to 4

$P(E) = 6/36$

$P(E) = 1/6$

Sum of outcomes.

c) divisible by 2 & 3

$P(E) = \dfrac{\text{No. of favourable outcomes}}{\text{Total No. of possible outcomes}}$

$= 6/36$

$= 1/36$

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

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?

Q5 – Solution :

Given, A bag contains 2 red balls, 3 green balls and 2 blue balls.

Therefore, Total Number of Balls = 7

Sample Space = S

$\therefore n(S) = {}^nC_r$

$\quad n$ – total No. of Samples.

$\quad r$ – Samples taken at a time.

$= {}^7C_2$

$= \dfrac{(7 \times 6)}{(2 \times 1)} = \dfrac{\cancel{42}^{21}}{\cancel{2}} = 21$

Total ways of drawing two balls from 7 balls

$n(E)$ = Except blue balls

Therefore $n(E) = {}^5C_2$

$= \dfrac{(5 \times 4)}{(2 \times 1)} = \dfrac{\cancel{20}^{10}}{\cancel{2}} = 10$

$\therefore P(E)$ be the probability of drawing 2 balls from bag which are not Blue :-

$\therefore P(E) = \dfrac{n(E)}{n(S)} = \dfrac{10}{21}$

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

Q6 – Solution:

Given, Probability of candies count by 6 children

Child A – 1 candy $= 1 * 0.015$

Child B – 4 candies $= 4 * 0.20$

Child C – 3 candies $= 3 * 0.65$

Child D – 5 candies $= 5 * 0.005$

Child E – 6 candies $= 6 * 0.01$

Child f – 2 candies $= 2 * 0.120$

$P(E) = $ Randomly selected child candies count

$\{ P(S) * P(E)$

probability is already given

$\therefore P(E) = 1 * 0.015 + 4 * 0.20 + 3 * 0.65 + 5 * 0.005 + 6 * 0.01 + 2 * 0.120$

$= 3.09.$

$\therefore$ The count of candies from a randomly selected child will be

3.09 (a) 3 candies.

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

Q7_Solution:

The solution for the above question is attached in the .ipynb format as mentioned by the assignment team.

The file q7_Solution.ipynb contains the solution of the above question.

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?

Q8 - Solution :

Given, weights of patients in pounds

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

    Total 9 patients.

    Therefore $p(x) = \frac{1}{9}$.

    Expected value = ?

$$E_V = \{ \, p(x) \times \text{weights of patients} \,$$

$$E_V = \left(\frac{1}{9}\right) \times (108) + \left(\frac{1}{9}\right)(110) + \left(\frac{1}{9}\right)(123) + \left(\frac{1}{9}\right)(134) +$$

$$\left(\frac{1}{9}\right) \times (135) + \left(\frac{1}{9}\right)(145) + \left(\frac{1}{9}\right)(167) + \left(\frac{1}{9}\right)(187)$$

$$+ \left(\frac{1}{9}\right)(199)$$

$$= \left(\frac{1}{9}\right)(1308) = \frac{1308}{9} = 145.33$$

(or)

$$E_V = \frac{\text{Sum of observations}}{\text{No. of observations.}}$$

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

$$= \frac{1308}{9} = 145.33$$

∴ if one person is chosen at random, then weight = 145.

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

Car's speed and distance

Use Q9_a.csv

Q9_a_Solution:

       The solution for the above question is attached in the .ipynb format as mentioned by the assignment team.

       The file q9_a_Solution.ipynb contains the solution of the above question.

SP and Weight (WT)

Use Q9_b.csv

Q9_b_Solution:

       The solution for the above question is attached in the .ipynb format as mentioned by the assignment team.

       The file q9_b_Solution.ipynb contains the solution of the above question.

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight
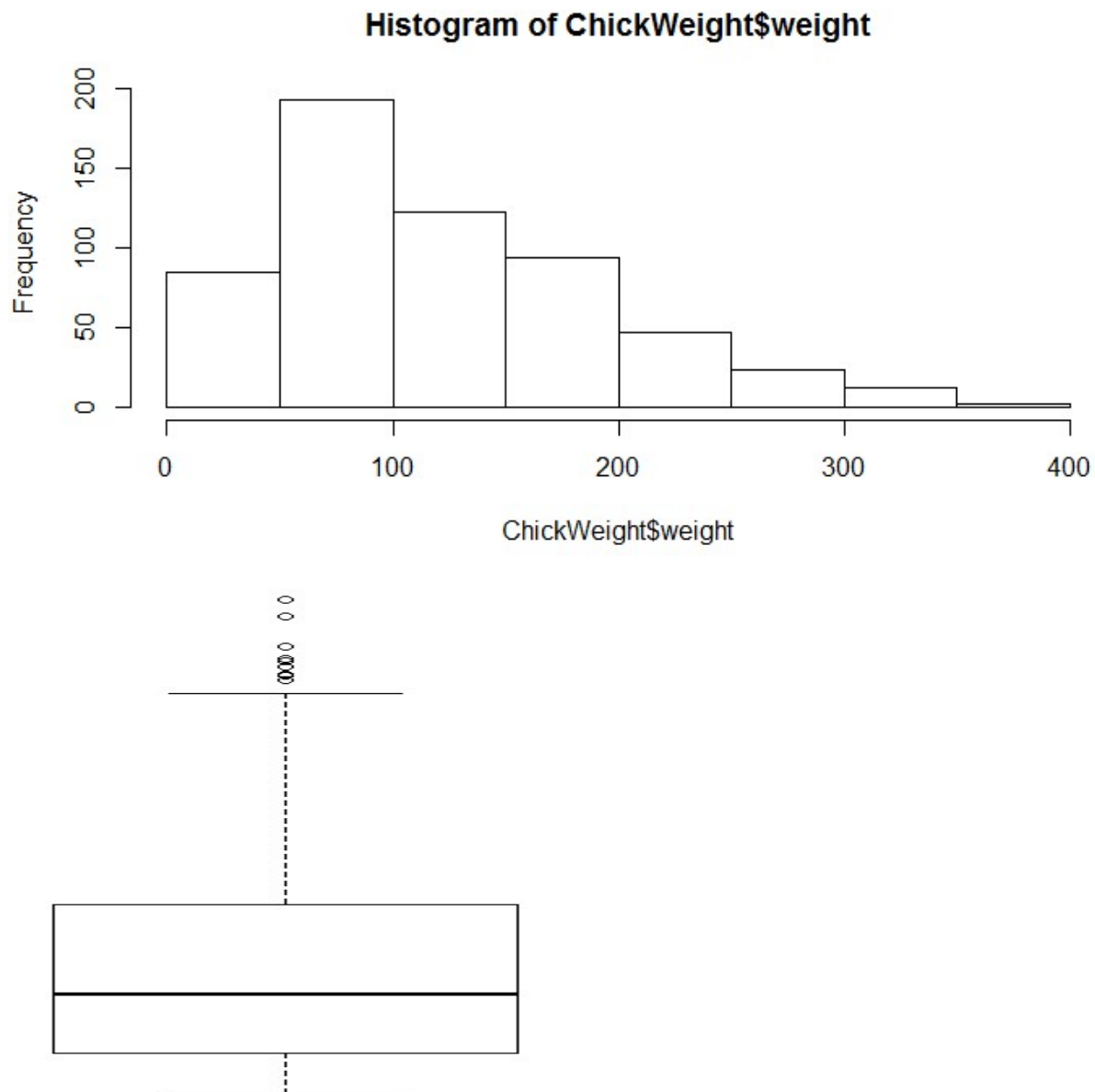
Q10_Solution:

ChickWeight$weight Histogram:

1. Histogram describes the shape of the data, Therefore we can see that the shape of the data is not symmetrical.

2. The histogram is positively skewed, Therefore the skewness>0
3. Positive skew means mode<median<mean.

ChickWeight$weight Boxplot:

1. Boxplot is used to determine the outliers in the data
2. Extreme values which are far from mean are known as outliers
3. We can see there are outliers present in the above boxplot
4. These outliers can be a reason behind the shape of the data going positively skewed instead of symmetrical
5. The outliers can be removed or recode with the mean value

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?

Q11_Solution:

The solution for the above question is attached in the .ipynb format as mentioned by the assignment team.

The file q11_Solution.ipynb contains the solution of the above question.

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?

Q12_Solution:

The solution for the above question is attached in the .ipynb format as mentioned by the assignment team.

The file q12_Solution.ipynb contains the solution of the above question.

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

Q13 solution:

Skewness=0, When the mean = median the data is in the shape of a bell

When we plot a histogram of the data and the data is ready for performing prediction or further analysis.

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

Q14 solution:

Skewness>0, When the mean > median

When we plot a histogram of the data we can see that the distribution has a long tail on its right side.

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

Q15 solution:

Skewness<0, When the median > mean

When we plot a histogram of the data we can see that the distribution has a long tail on its left side.

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

Q16 solution:

Kurtosis value indicates us about the measure of tailedness and peakedness of the data. Positive kurtosis value indicates that the distribution of data is not normal/symmetrical and has highest peakedness with thick tails.

Positive kurtosis is also known as leptokurtic.

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

Q17 solution:

Kurtosis value indicates us about the measure of tailedness and peakedness of the data. Negative kurtosis value indicates that the distribution of data is not normal/symmetrical i.e flatter with thinner tails.

Positive kurtosis is also known as Platykurtic.

Q18) Answer the below questions using the below boxplot visualization.
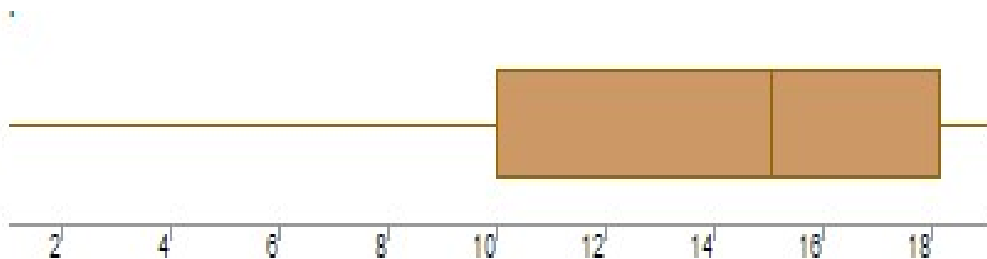
a)What can we say about the distribution of the data?

Solution:

We can say that the distribution of the data is not symmetrical as the box plot is plotted at the extreme end showing that the most of the data lies in the right side

b)What is the nature of skewness of the data?

Solution:

As we can see that in the boxplot the median line is dividing the upper and lower whisker lengths. The Lower Whisker is actually longer than the upper whisker in the boxplot so the data is negatively skewed, Skewness < 0 and mean < median.

c)What will be the IQR of the data (approximately)?
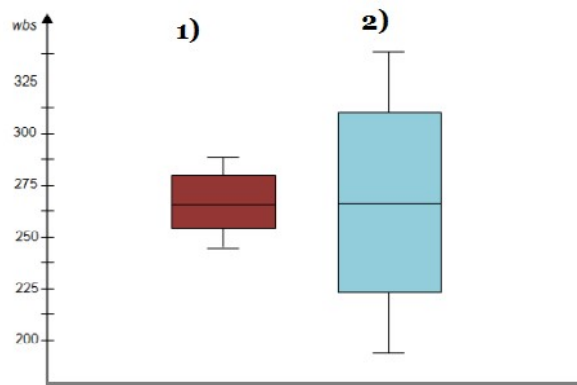Solution:

The formulae for calculating IQR is Q3-Q1

Therefore Q1=10,Q3=18

IQR=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.

1. We can see that the boxplot 2 has more data than boxplot 1
2. Boxplot 1 and boxplot 2 both have same median value
3. Boxplot 2 has wide spread of data when compared to boxplot 1
4. Both the boxplot have symmetrical distribution of data as the median divides the box into equal whisker length's.

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)

Q20_Solution:

      The solution for the above question is attached in the .ipynb format as mentioned by the assignment team.

      The file q20_Solution.ipynb contains the solution of the above question.

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

      Q21_a_Solution:

        The solution for the above question is attached in the .ipynb format mentioned by the assignment team.

      The file q21_a_Solution.ipynb contains the solution of the above question.

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

      Q21_b_Solution:

        The solution for the above question is attached in the .ipynb format as mentioned by the assignment team.

      The file q21_b_Solution.ipynb contains the solution of the above question.

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

Q22_Solution:

The solution for the above question is attached in the .ipynb format as mentioned by the assignment team.

The file q22_Solution.ipynb contains the solution of the above question.

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

Q23_Solution:

The solution for the above question is attached in the .ipynb format as mentioned by the assignment team.

The file q23_Solution.ipynb contains the solution of the above question.

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

Q24_Solution:

The solution for the above question is attached in the .ipynb format as mentioned by the assignment team.

The file q24_Solution.ipynb contains the solution of the above question.