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

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

Solution:-when the three coins are tossed,the number of outcomes will be eight
{(H,H,H),(H,H,T),(H,T,H),(T,H,H),(T,T,T),(T,T,H),(T,H,T),(H,T,T)}

so, p(two heads and one tail) = 3/8
= 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

Solution:-When the two dice are rolled,the number of outcomes will be 36

a) Equal to 1 :-

p(Equal to 1) = 0/36
= 0

b) Less than or equal to 4 :-

p(Less than or equal to 4) = 6/36
= 1/6
= 0.166

c) Sum is divisible by 2 and 3 :-

p(Sum is divisible by 2 and 3) = 6/36
= 1/6
= 0.166

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?

Solution:- Total number of balls are 7

Total number of ways to draw 2 balls randomly = 7 C 2 = 21

Now, number of ways to draw 2 balls from 5 non-blue balls randomly =5 C 2 = 10

p(Non-blue balls) = 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

Solution :- Expected number of candies for a randomly selected child are

= 1*0.015 + 4*0.20 + 3*0.65 + 5*0.005 + 6*0.01 + 2*0.120

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 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

Solution :-

Mean of Points = 3.596

Median of Points = 3.695

Mode of Points = 3.92 & 3.07

Varience of Points = 0.285

Standard Deviation of Points = 0.534
Range of Points = 2.17

Mean of Score = 3.217

Median of Score = 3.325

Mode of Score = 3.44

Varience of Score = 0.957

Standard Deviation of Score = 0.978

Range of Score = 3.91

Mean of Weigh = 17.848

Median of Weigh = 17.71

Mode of Weigh = 17.02 & 18.9

Varience of Weigh = 3.193

Standard Deviation of Weigh = 1.786

Range of Weigh = 8.4

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?

Solution :- Expected value of the weight of randomly selected patients can be calculated by taking the mean of the weights of the patients

= Mean(108,110,123,134,135,145,167,187,199)
=145.333

Q9) Calculate Skewness, Kurtosis & draw inferences on the following data
Cars speed and distance
Use Q9_a.csv

Solution :- **Skewness of Cars speed = `-0.117`**

Skewness of Distance = `0.806`

Kurtosis of Cars speed = `-0.508`

Kurtosis of Diostance = `0.405`

SP and Weight(WT)
Use Q9_b.csv

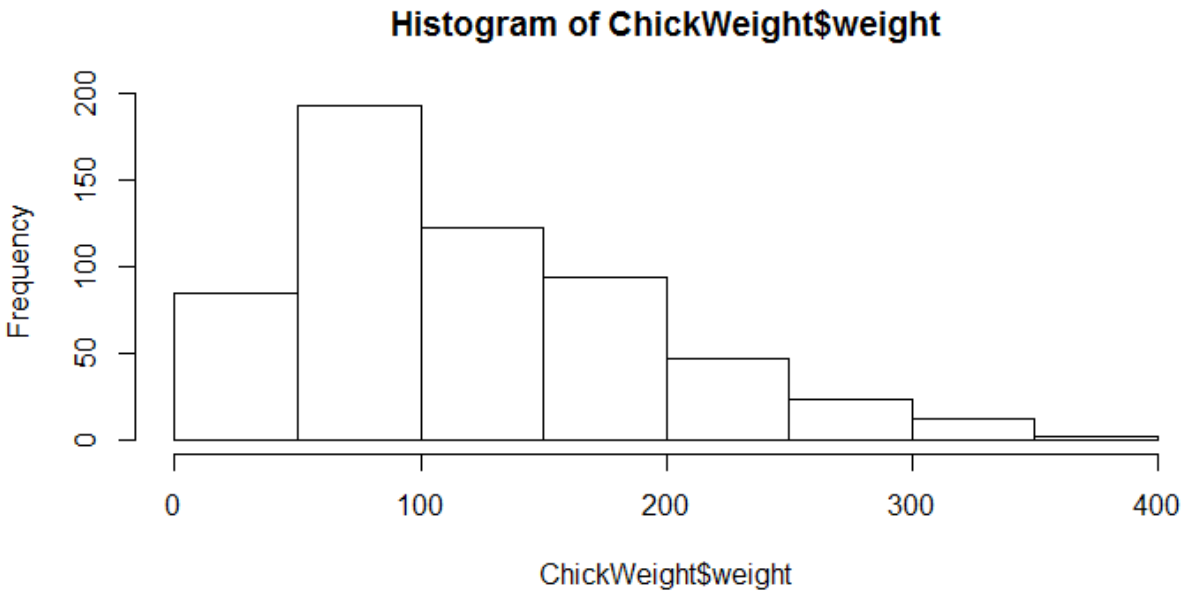
Solution :- Skewness of sp = **`1.611`**

Skewness of wt = **`-0.614`**

Kurtosis of sp = **`2.977`**

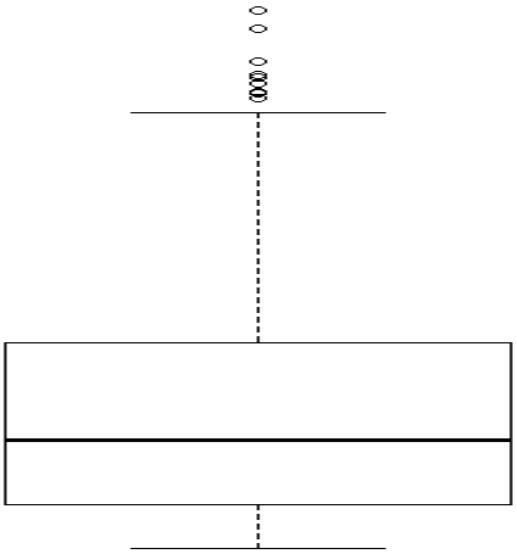
Kurtosis of wt = **`0.950`**

Q10) Draw inferences about the following boxplot & histogram



Solution :- 200 chickens have weight = 75

Therefore,the mode of the given histogram is 75.



Solution :- Box plot is mainly used to identify the outliers present in the data.

By using box plot ,we can easily find out the median of the data.

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?

Solution :- Mean of the sample = 200

Standard deviation of the sample = 30

Sample size = 2000

Here, Standard deviation/squareroot(sample size) = 0.67

For 94% confidence interval,the area under curve will be 0.97

Therefore,the Z value for 0.97 = Mean(1.88,1.89)

= 1.885

Interval Estimate for the above data = [200 – 1.885*0.67,200 +1.885*0.67]

= [198.737,201.263]

- For 98% confidence interval,the area under curve will be 0.99

Therefore,the Z value for 0.99 = Mean(2.32,2.33)

= 2.325

Interval Estimate for the above data = [200-2.325*0.67,200+2.325*0.67]

= [198.442,201.557]

- For 96% confidence interval,the area under curve will be 0.98

Therefore,the Z value for 0.98 = Mean(2.05,2.06)

= 2.055

Interval Estimate for above data = [200-2.055*0.67,200+2.055*0.67]

= [198.623,201.377]

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.

Solution :- 1) Mean = 41

Median = 40.5

Varience = 25.52

Standard deviation = 5.053

2)What can we say about the student marks?

Solution :- Out of 18 students,4 students have 41 marks which is equal to mean.

Also,students marks are not following normality and the datatype of students marks following is Discrete.

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

Solution : The Skewness of distribution is known as the lack of symmetry. Skewness can be positive or negative .When the mean,median of data are equal to each other,then the data is symmetrically distributed. There will 0 skewness in that case.

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

Solution :- If the mean is greater than medain,then the distribution is positively skewed.

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

Solution :- If the median is greater than mean,then the distribution is negatively skewed.

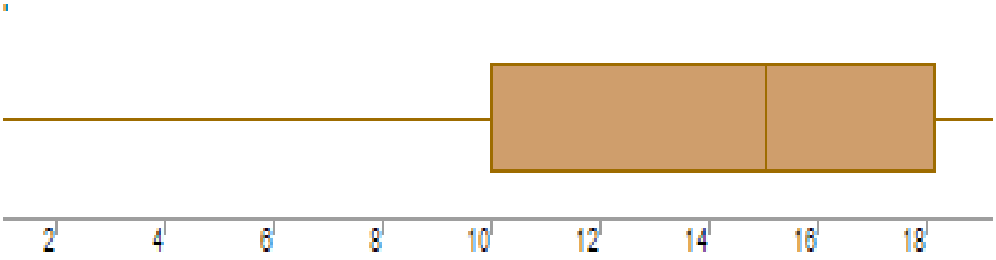
Q16) What does positive kurtosis value indicates for a data ?

Solution :- A distribution with a positive kurtosis value indicates that the distribution has heavier tails and a sharper peak than the normal distribution.

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

Solution :- A distribution with an negative kurtosis value indicates that the distribution has lighter tails and a flatter peak than the normal distribution.

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



What can we say about the distribution of the data?

Solution :- The distribution of the data is not following normality.

What is nature of skewness of the data?

Solution :- Right Skewed or negative

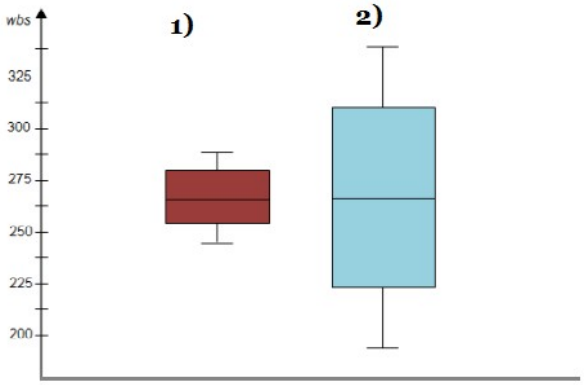
What will be the IQR of the data (approximately)?

Solution :- IQR(Inter-Quartile Range) = Upper quartile – Lower quartile

= 18 – 10

= 8(approx).

Q19) Comment on the below Boxplot visualizations?



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

Solution :- Both the plots are following normally distrubuted.

Median of both the plot is same which is 262.5

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)

Solution :- P(MPG>38) = 33/81 = 0.41

b. P(MPG<40)

Solution :- P(MPG<40) = 61/81 = 0.76

c. P (20<MPG<50)

Solution :- P(20<MPG<50) = 69/81 = 0.85

Q 21) Check whether the data follows normal distribution

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

Solution :- MPG of cars almost follow normal distribution.

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

Solution :-Waist Circumference followed normal distribution while Adipose Tissue(AT) does not followed normal distribution.

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

Solution :- Z scores of 90% confidence interval = Mean(1.64,1.65) = 1.645

Z scores of 94% confidence interval = Mean(1.88,1.89) = 1.885

Z scores of 60% confidence interval = Mean(0.84,0.85) = 0.845

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

Solution :- **Sample size = 25**

Degree of freedom = 24

t scores at 95% confidence interval for 24(n-1) = 2.064

t scores at 96% confidence interval for 24(n-1) = Mean(2.064,2.492) = 2.278

t scores at 99% confidence interval for 24(n-1) = 2.797

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

Solution :-Mean of sample(x) = 260

Mean of population(u) = 270

Standard deviation of sample(s) = 90

Sample size(n) = 18

Degree of freedom = n-1

t = (x-u)/(s/sqr_root(n))

t = -0.471

For negative value = 1-0.471 = 0.529

Therefore, P(avg life <=) = 0.7