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

Q1) Identify the Data type for the Following:

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	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	Ordinal
Time Of Day	Ordinal
Time on a Clock with Hands	Ratio

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?

Sol: P(H H T)+P(H T H)+P(T H H)

=1/8+1/8+1/8

=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
   Sol: a) There is no such outcomes
      i.e . 0/36 . Probability is 0.
      b) (1,3) (2,2) (3,1) = 3 outcomes,3/36 i.e.1/12
   c) 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?
Sol : Total number of balls

= 2 + 3 + 2 = 7

Let S be the sample space

Then, n(S) = Number of ways of drawing 2 balls out of 7

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

=21

Let E = event of 2 balls ,none of which is blue

n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls

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

   = 10

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

Sol : Child A- probability of having 1 candy = 0.015

   Child B - probability of having 4 candies = 0.20

   Expected number of canides for randomly selected child

   = 1*0.015 + 4*0.2+3*0.65+ 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.

Sol : For points ,

Mean=3.5965 , median = 3.695 , variance = 0.2858
standard deviation = 0.5346 range = 2.76 --4.93
For score ,
Mean = 3.2172 ,median = 3.325 ,variance = 0.9573 ,
standard deviation = 0.9784 range = 1.513 -- 5.424
For weigh ,
Mean = 17.84 , median = 17.71 , variance = 3.19 ,
standard deviation = 1.78 , range = 14.5 --22.9 .
Mean value are closer for both Point and Scores .
Mean and median of weigh are nearly equal .

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?

Sol :The mean is = 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199 / 9

= 145.33

Expected value = 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

a) Sol : Skewness for speed is (-1139548)

Skewness value is negative so it is left skewed.

Skewness for diatance is (0.7424835) ,right skewed .

Kurtosis for speed is (2.4228)

Kurtosis for distance is (3.2480)

b) skewness for SP is (1.5814) right skewed

Skewness for weight is (- 0.6033) ,left skewed

Kurtosis for SP is (5.7235)

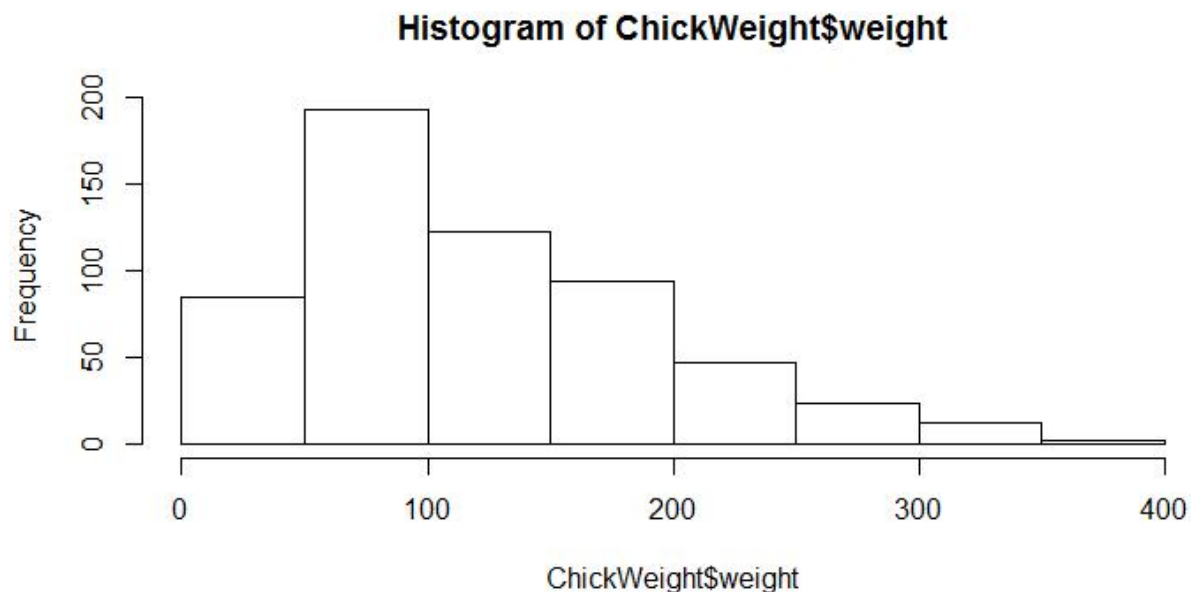**Kurtosis for weight is (3.8194)**

About kurtosis ,

When kurtosis > 3--Distribution is peaked (distance ,sp, weight)

Kurtosis = 3 --kurtosis is near to 0, same as normal distribution .
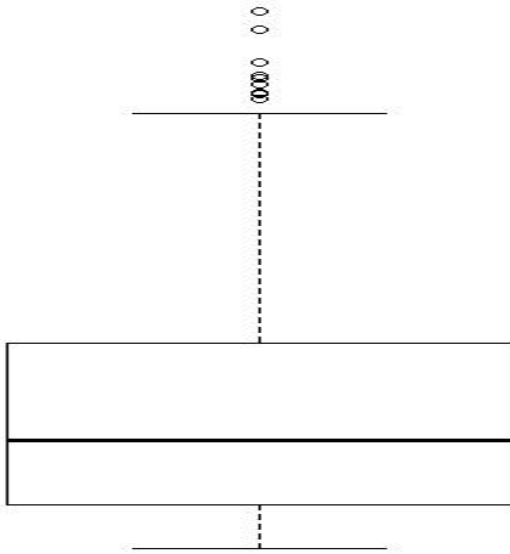
Kurtosis < 3 --less peaked .(speed)

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight

ChickWeight$weight

Sol : The most of the data points are concentrated in the range 50 -100 with frequency 200 . & least range of weigh is 400 around 0 - 10 .

So the expected value of above distribution is 75 .

Skewness : It is heavily right skewed .

Sol : Median is less than mean ,right skewed and outliers on the upper side of the box plot .

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?

Sol : Degrees of freedom = 2000 - 1= 1999

Confidence interval for 94 % is 1.88

Confidence interval for 98 % is 2.32

Confidence interval for 96 % is 2.05 .

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?

Sol : 1) Mean = 41 ,

Median = 40

Variance = 24.11

Standard Deviation = 4.9

2) scores obtained by students is in range between 34 to 56 .

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

Sol : Symmetrical

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

Sol : Right skewed

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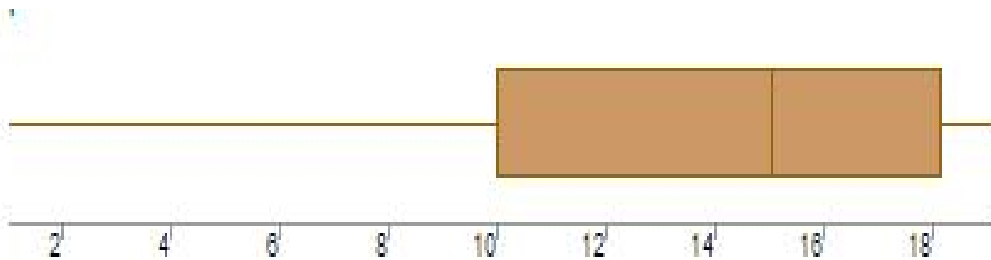
Sol : Left skewed

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

Sol : The data is normally distributed and kurtosis value is 0 .

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

Sol : The distribution of data is lighter tail and flatter peaks than the normal distribution .

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

What can we say about the distribution of the data?

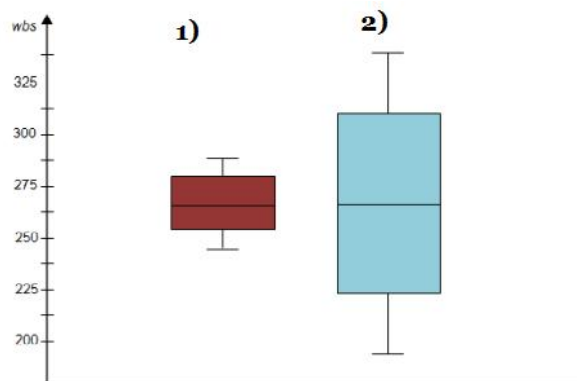Sol : 50 % of data distributed upto the 10 points .

What is nature of skewness of the data?

Sol : Left skewed ,median is greater than mean .

What will be the IQR of the data (approximately)?
Sol : 8

Q19) Comment on the below Boxplot visualizations?

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

Sol : Mean and median are equal hence distribution is symetrical .

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)

a. P(MPG>38)=0.3475939251582705

b. p(MPG<40)=0.7293498762151616

c. P(20<MPG<50)1.2430968797327613e−05

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

Sol : Mean = 34.42

    Median = 35.15

    Mode = 29.62

    Skew = −1779

    Kurtosis = −0.611

From above values we can say that data is fairly normally
Distributed .

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

Sol :   Mean   Median
AT   101.89   96.54
Waist   91.90   90.80
Data is positively skewed .

Data is fairly normally distributed .

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

   Z score of 90% confidence interval is 1.645
   Z score of 94% confidence interval is 1.89
   Z score of 60% 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

    t score of 95% confidence interval is = 2.06
    t score of 96% confidence interval is = 2.17
    t score of 99% confidence interval is = 2.79 .

 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

        df → degrees of freedom

To find:

If the ceo's claim were true

 Ans: Assume H0 = An average life of bulb>=260 days

 X=260

Using python code

T=(260-270)/(90/18**0.5)

t-0.4714045207910317

P value=0.32167411684460556

P0.005 so reject h0

So,an average life of bulb< 260 days.