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 data- Nominal
High School Class Ranking	Discrete data- Nominal
Celsius Temperature	Continuous- Interval
Weight	Continuous- ratio
Hair Color	Discrete-Ratio
Socioeconomic Status	Continuous- Interval
Fahrenheit Temperature	Continuous- ratio
Height	Continuous- ratio
Type of living accommodation	Discreate-Ordinal
Level of Agreement	Discreate-Interval
IQ(Intelligence Scale)	Discreate-Interval

Sales Figures	Discreate-Interval
Blood Group	Discreate-Ratio
Time Of Day	Continues-Interval
Time on a Clock with Hands	Continues-Interval
Number of Children	Discrete-Interval
Religious Preference	Discreate-Ratio
Barometer Pressure	Discrete-Interval
SAT Scores	Discreate-Ratio
Years of Education	Discreate-Nominal

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

Ans:- When three coins are tossed the total number of possible combinations are 2^3 = 8.

These combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

The number of combinations which have two heads and one tail are:
HHT, HTH, TTH which makes them 3 in number

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 :- (b) less than or equal to 4

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 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

=7C2

=(7×6)/(2×1)

=21

Let E = Event of drawing 2 balls, none of which is blue.

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

=5C2

=(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

Ans:- Child A

 –

 probability of having 1 candy = 0.015.Child B

 –

 probability of having 4 candies = 0.20

Sol: Expected number of candies for randomly selected child =
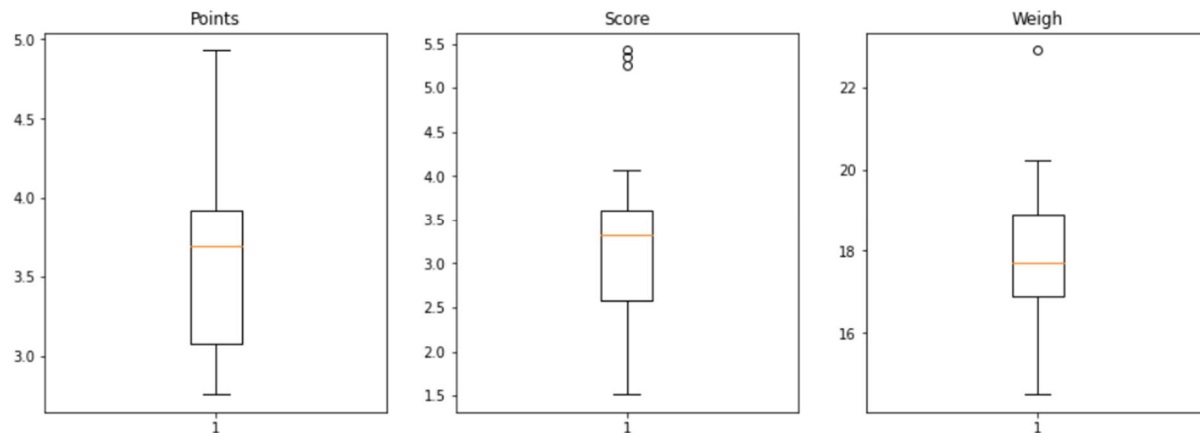1*0.015+ 4*0.20+3*0.65+ 5*0.005+ 6*0.01 +2*0.120

=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.**

Use Q7.csv file

Ans:-

Points
: Mean =3.596563, Median= 3.695, Mode= "numeric",

Variance= 0.2858814, Standard deviation= 0.5346787.

Score:
Mean= 3.21725, Median= 3.325, Mode= "numeric",

Variance= 0.957379, Standard deviation= 0.9784574

Note: Mean value are closer for both 'Point' and 'Score'.

Weight:
Mean= 17.84875, Median= 17.71, Mode= "numeric",

Variance= 3.193166, Standard deviation= 1.786943

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 :-

```
In [1]:   1  import numpy as np

In [3]:   1  x = np.array([108, 110, 123, 134, 135, 145, 167, 187, 199])

In [4]:   1  x
Out[4]:  array([108, 110, 123, 134, 135, 145, 167, 187, 199])

In [5]:   1  x.mean()
Out[5]:  145.33333333333334
```

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 of the SP dataset is: 1.611kurtosis of the given SP dataset is 2.98skewness of the WT dataset is: -0.614kurtosis of the given WT dataset is 0.95

A) As the skewness for SP dataset is greater than 1, we can conclude the data ishighly skewed towards right and as the kurtosis is positive and high value, we cansay data tends to have heavy tails and outliers.

From both skewness and kurtosisfor given SP dataset we can assume outliers are existing on right side.

B) As the skewness for WT dataset is in b/w [-1,-0.5], we can conclude the data ismoderately skewed towards left and as the kurtosis is positive , we can concludedata tends to have slightly heavy tails and outliers, from both the skewness andkurtosis values we can conclude that most of the outliers tend to exist on left side.

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight
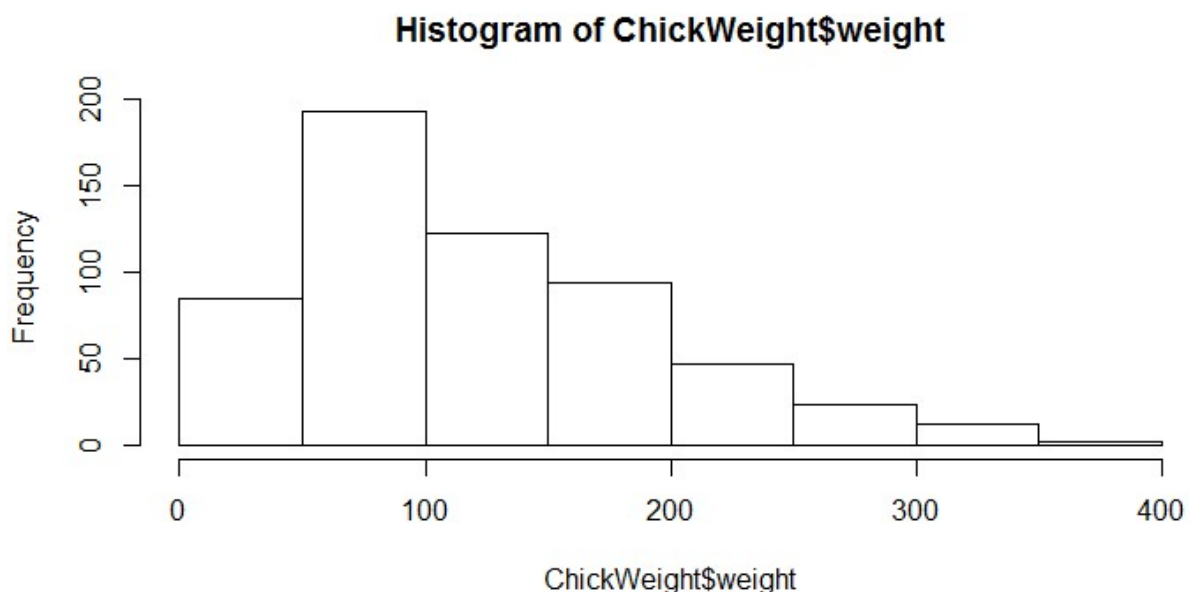
The most of the data points are concerated in the range 50-100 with frequency 200.And least range of weight is 400 somewere around 0-10.So the expected value the above distribution is 75.Skewness- we can notice a long tail towards right so it is heavily right skewed.
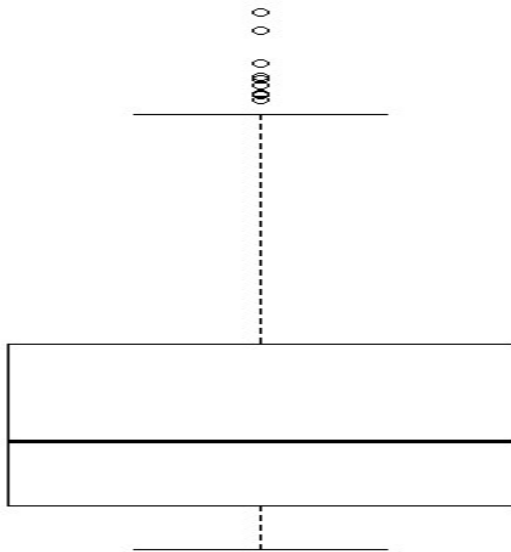
Medican is less than mean right skewed and we have outlier on the upperside of box plot and there is less data points between Q1 and bottom point.

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 :- X+/-(Z1-α.σ/sqrt(n)

Degrees of freedom= 2000-1= 1999
Confidence interval= 94%(1-σ/2)= 1-0.03) =0.97

for confidene interval for 94% is 1.882
Confidence interval for 98%= 2.33
Confidence interval for 96% = 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?

Ans :- 1) Mean= 41, Median= 40, variance= 24.111, Standard deviation= 4.910

2)All student passed in exam

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

ANS :- Symetrical

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

ANS:-Right skewed

Q15) What is the nature of skewness when median > mean?
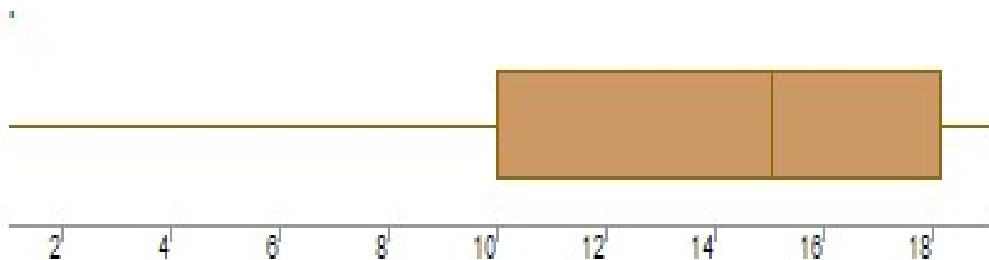
ANS:- Left Skewed

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

ANS:- The data is notmally 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 a flatter peaks than the normal distribution.

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

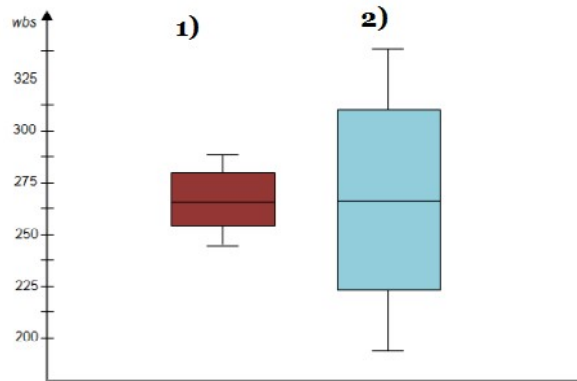What can we say about the distribution of the data?

What is nature of skewness of the data?

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

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

ANS:- Approximately= -8

Q19) Comment on the below Boxplot visualizations?

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

ANS:- By observing both the plots whisker's level is high in boxplot 2, 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)

ANS:- By using filter command
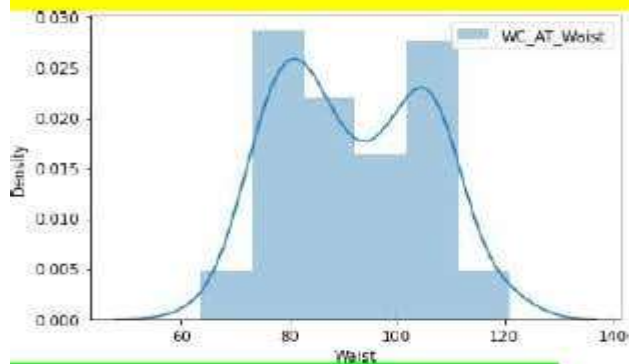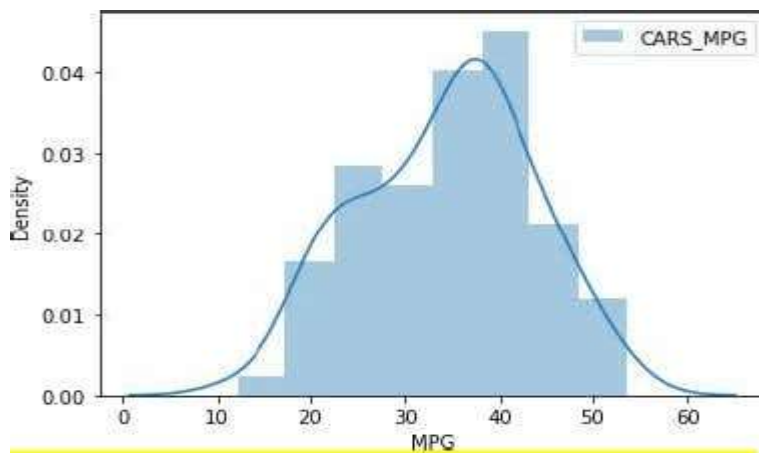
and installing the dplyr package into the 'R'.
a) There are 33 observations in MPG which are greater than 38
b)61 observations in MPG which are lesser than 40

Q 21) Check whether the data follows normal distribution

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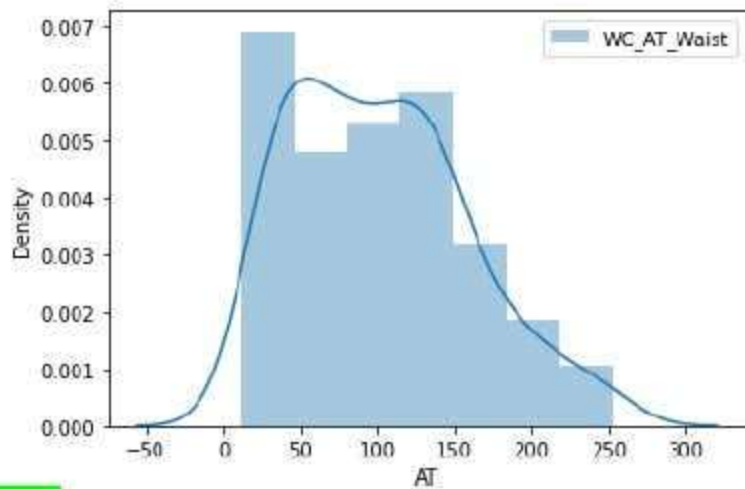Dataset: Cars.csv

ANS:- For Waist Circumference(Waist)

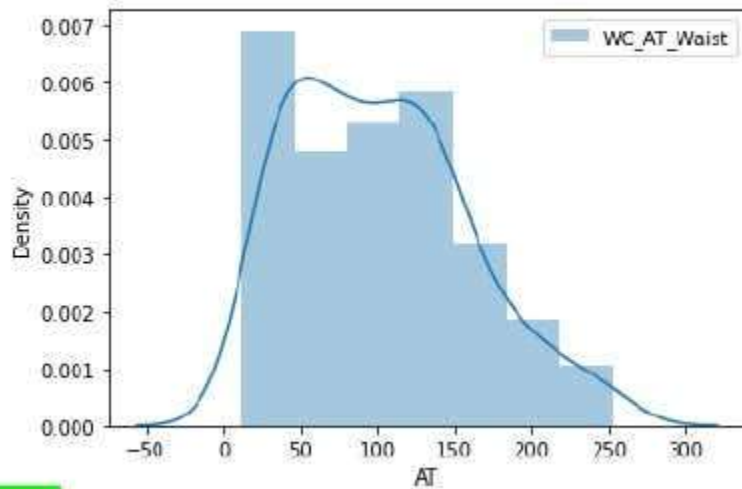mean is :91.90183486238533median is :90.8mode is:0 94.5

1 106.02 108.5skewness is:0.1340560824786468
As mean, median, mode values are nearly same and skewness value isless than 1 we can consider this dataset tends to be normal

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:-

Z score of 90% confidence interval is 1.65

Z score of 94% confidence interval is 1.55
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

ANS:-For 95%= 1.96

For 96%= 2.5
For 99% = 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:

rcode → pt(tscore,df)

df → degrees of freedom

Ans:Given: μ = population mean = 270 x bar= mean of sample bulbs = 260s = standard deviation of the sample = 90n = no of randomly selected bulbs for sample = 18Formula:

As the population standard deviation is not specified we do t test to find the probability = $(x - \mu)/(s/sqrt(n))$ t = $(260 - 270)/(90/sqrt(18))$ = -0.47 Probablity using t value in python stats.t.cdf(t-value,df = n-1) stats.t.cdf(-0.47,17) = 0.32216

The probability of bulb avg life less than 260: 0.32216394448907903