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	Categorical
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	Nominal
Time on a Clock with Hands	Ordinal
Number of Children	Ratio
Religious Preference	Nominal

Barometer Pressure	Interval
SAT Scores	Ratio
Years of Education	Ratio

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

Ans) 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) 0
   b) 6/36
   c) 6/36

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) 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) $(1*0.015) + (4*0.2) + (3*0.65) + (5*0.005) + (6*0.01) + (2*0.120) = 3.085$

Therefore, the expected no. of candies for a randomly selected child is 3

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) Mean of Points = 3.5965

Mean of Score = 3.2172

Mean of Weigh = 17.8488

Median of Points = 3.6950

Median of Score = 3.325

Median of Weigh = 17.71

Mode of Points = 3.07, 3.92

Mode of Score = 3.44

Mode of Weigh = 17.02, 18.9

Variance of Points = 0.2859

Variance of Score = 0.9574

Variance of Weigh = 3.1931

Standard Deviation of Points = 0.5347

Standard Deviation of Score = 0.9785

Standard Deviation of Weigh = 1.7869

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
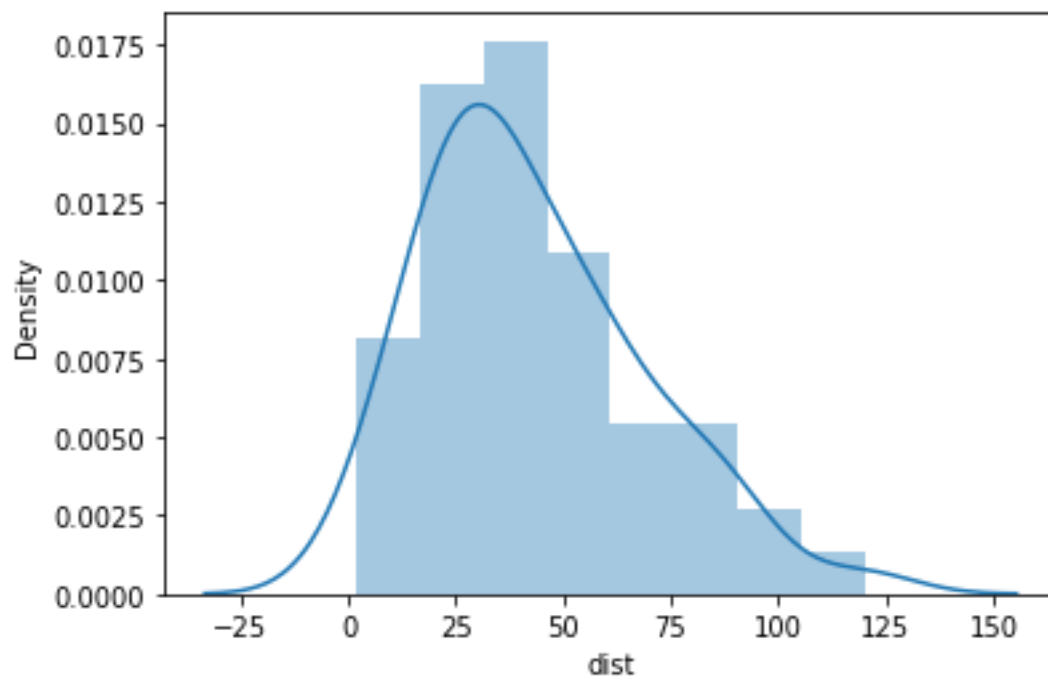Value of the Weight of that patient?

Ans) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)/9 = 145

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

Cars speed and distance

Use Q9_a.csv

Ans)

Skewness of distance column = 0.8068

Kurtosis of distance column = 0.405

The distplot of 'dist' column shows that the data is moderately right skewed and have more outliers. Because this is right skewed data, mode<median<mean.
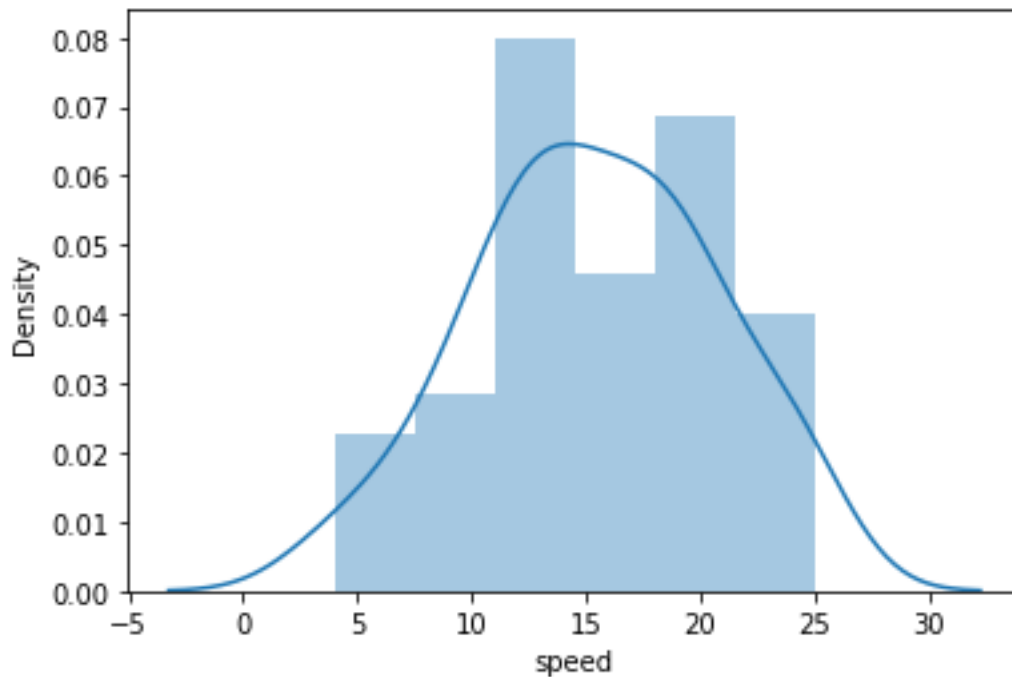
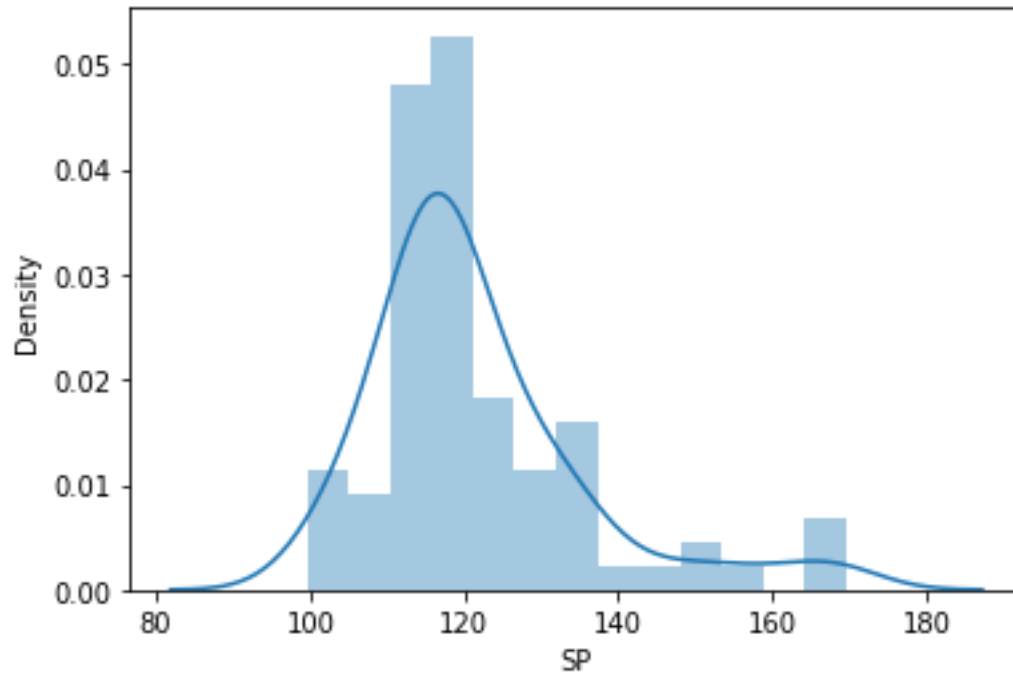

Skewness of speed column = -0. 1175

Kurtosis of speed column = -0.5089

The distplot plot of 'speed' column follows a normal distribution and have no outliers.
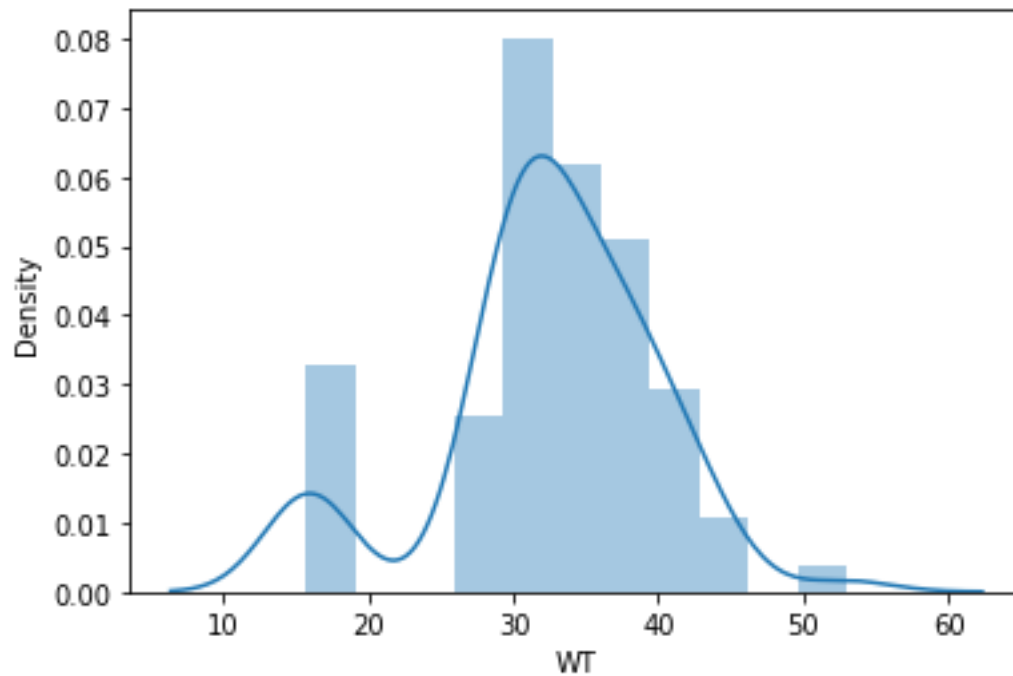

SP and Weight(WT)

Use Q9_b.csv

Ans)

Skewness of SP = 1.61145

Kurtosis of SP = 2.977329

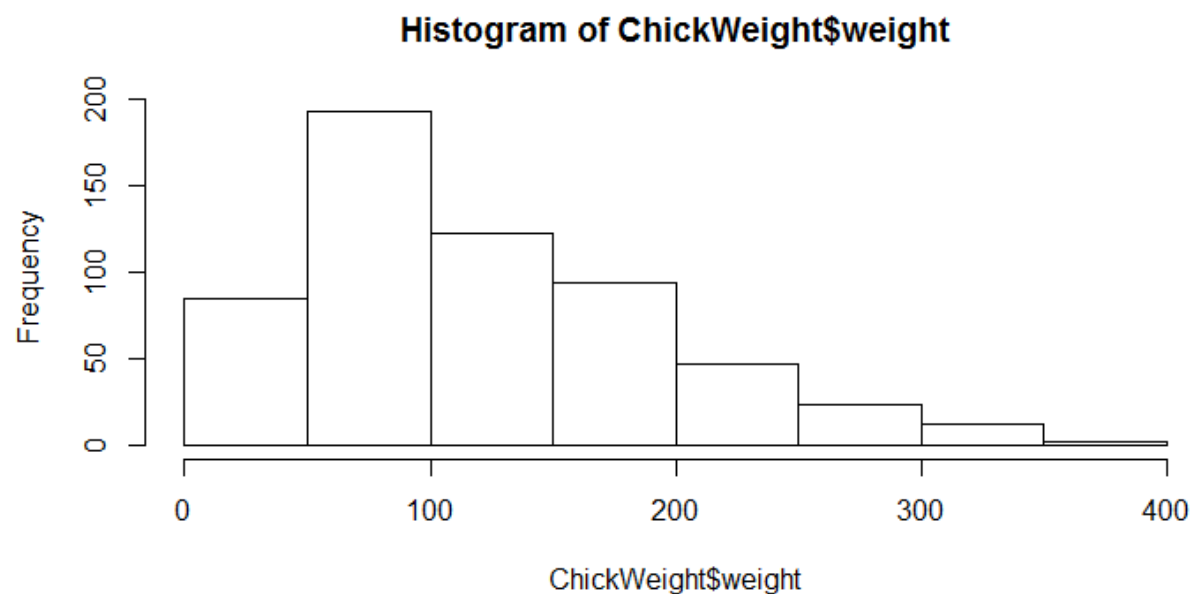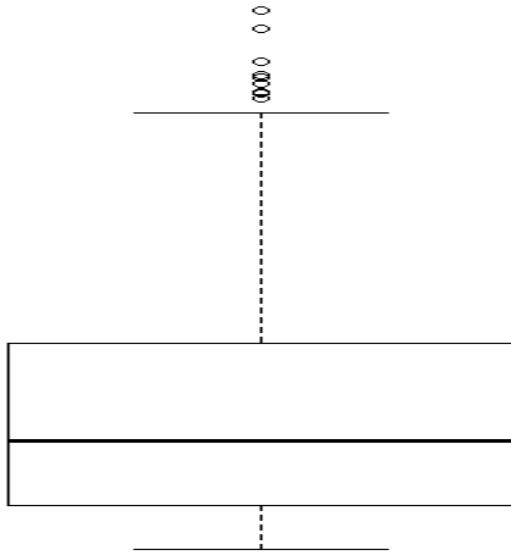The graph shows that the SP column is highly right skewed and therefore, mode<median<mean

Skewness of WT = -0.614753

Kurtosis of WT = 0.950291

WT column is moderately left skewed. So, the mean<median<mode

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight

Ans) The data is right skewed. Outliers exist. More than 75% of data lie in the left side of the median line. Since this is right skewed, (Q3-Q2) > (Q2-Q1).Also, the mode<median<mean.

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) For 94% CL, Confidence interval = (198.87161006007665, 201.12838993992335)

For 98% CL, Confidence interval = (199.62394888040726, 200.37605111959274)

For 96% CL, Confidence interval = (199.24783863179837, 200.75216136820163)

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

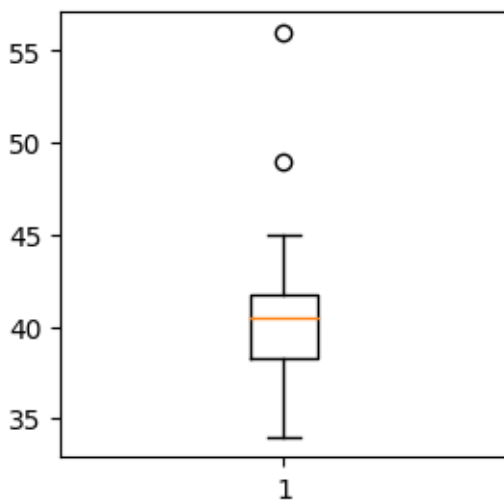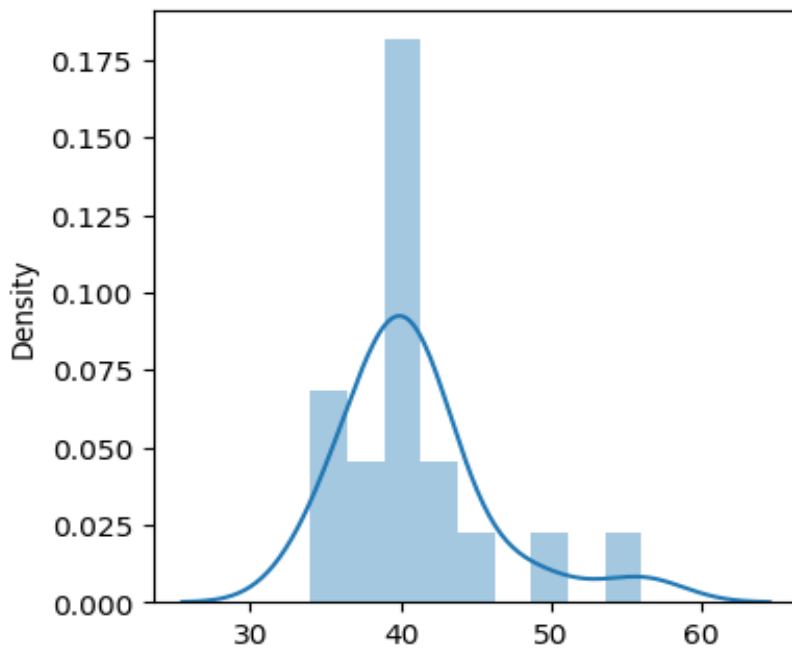1) Find mean, median, variance, standard deviation.

Ans) Mean = 41

Median = 40.5

Variance = 24.11111111

Standard deviation = 4.910306620885412

2) What can we say about the student marks?

Ans)   The range of the scores is 22. which means that the highest score is more than 3 standard deviations away from the mean. Overall, the student's marks seem to be fairly consistent, with only two outliers.

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

Ans) Skewness is 0.

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

Ans) Skewness will be positive (right skewed)

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

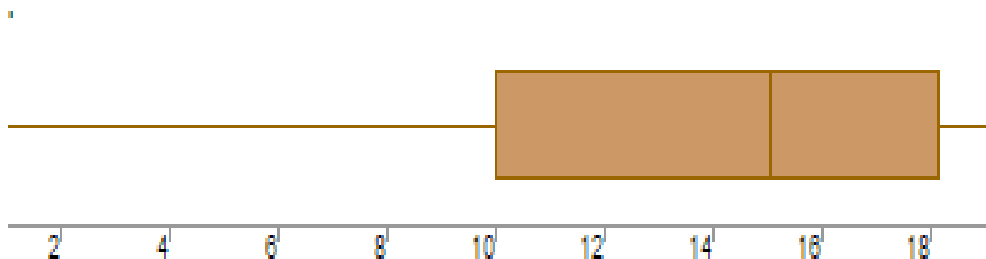Ans) Skewness will be negative (right skewed)

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

Ans) Outliers will be more.

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

Ans) Outliers will be less.

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

What can we say about the distribution of the data?
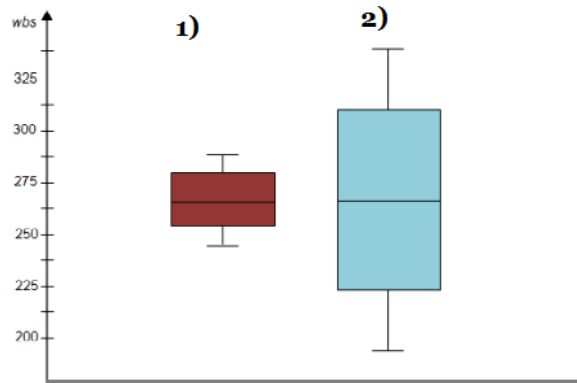
Ans) Negatively skewed data.

What is nature of skewness of the data?

Ans) Left skewed

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

Ans) 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.

Ans)   '1)' is slightly right skewed. Has thicker tails and have more outliers.

'2)' is a Normal distribution. Has thinner tails and don't have any outliers.

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)
Ans) 34.7%
b. P(MPG<40)
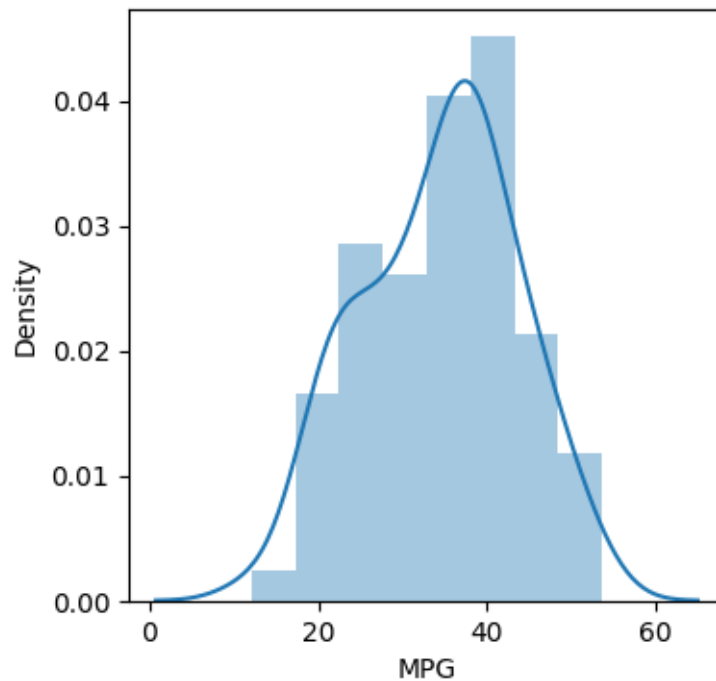Ans) 72.95%
c. P (20<MPG<50)

Ans) 89.9%

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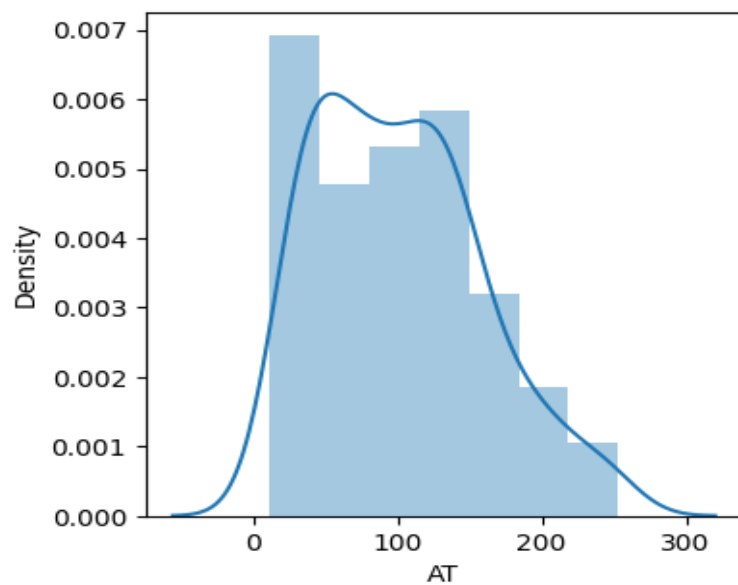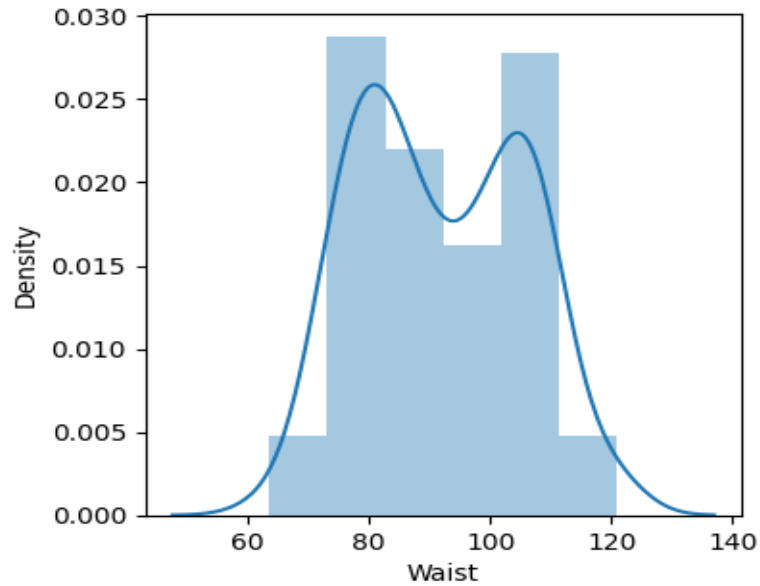
Ans)

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

Ans) Both AT and Waist does not follow a normal distribution.

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

Ans)   For 95% confidence level, Z score = 1.96

For 94% confidence level, Z score = 1.88

For 60% confidence level, Z score = 0.84

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

Ans)   For 95% confidence level, t score = 2.064

For 96% confidence level, t score = 2.171

For 99% confidence level, t score = 2.796

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) 0.3218