Q1) Identify the Data type for the Following:

Activity	Data Type
Number of beatings from Wife	Numeric Data Discrete
Results of rolling a dice	Numeric Data Discrete
Weight of a person	Numeric Data Continuous
Weight of Gold	Numeric Data Continuous
Distance between two places	Numeric Data Continuous
Length of a leaf	Numeric Data Continuous
Dog's weight	Numeric Data Continuous
Blue Color	Categorical Data
Number of kids	Numeric Data Discrete
Number of tickets in Indian railways	Numeric Data Discrete
Number of times married	Numeric Data Discrete
Gender (Male or Female)	Categorical Data

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	Nominal

Barometer Pressure	Interval
SAT Scores	Ratio
Years of Education	Interval

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

Answer – 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

Answer – a) 0 b) 1/6 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?

Answer – 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

Answer –

CHILD	Expected Values
A	0.015
B	0.8
C	1.95
D	0.025
E	0.06
F	0.24

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

Answer – Using Pandas in Python

Points-

```
Mean - 3.5965625000000006
Median - 3.6950000000000003
Mode - 3.07
Variance - 0.28588135080645166
Std - 0.5346787360709716
Range - 2.17
```

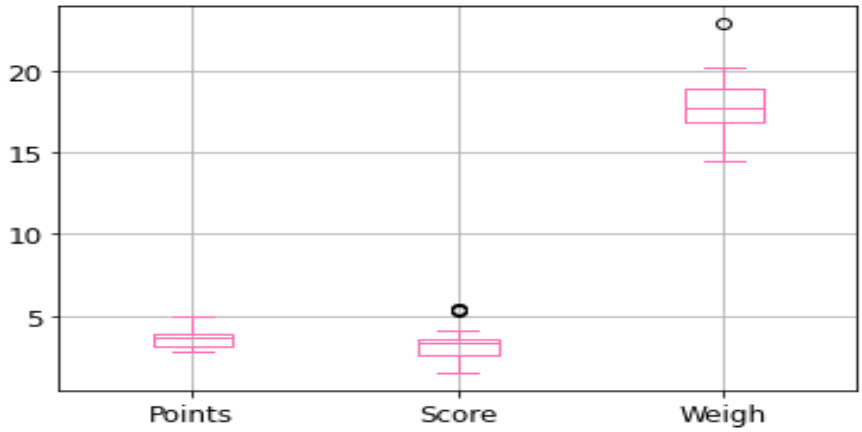
Score-

```
Mean - 3.2172499999999995
Median - 3.325
Mode - 3.44
Variance - 0.9573789677419354
```

Std - 0.9784574429896966
Range - 3.9109999999999996

Weigh-

Mean - 17.848750000000003
Median - 17.71
Mode - 17.02
Std - 1.7869432360968431
Range - 8.399999999999999
Variance - 3.193166129032258



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?

Answer – 145.334

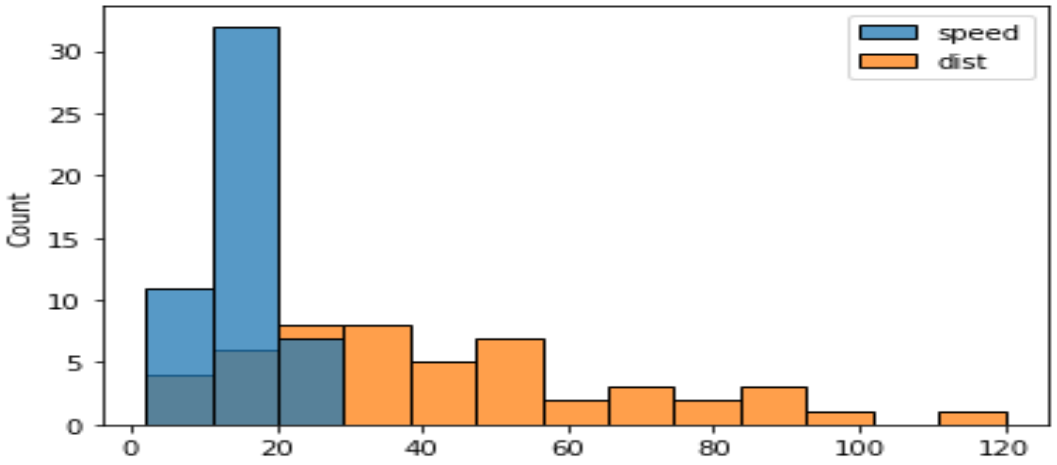
Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

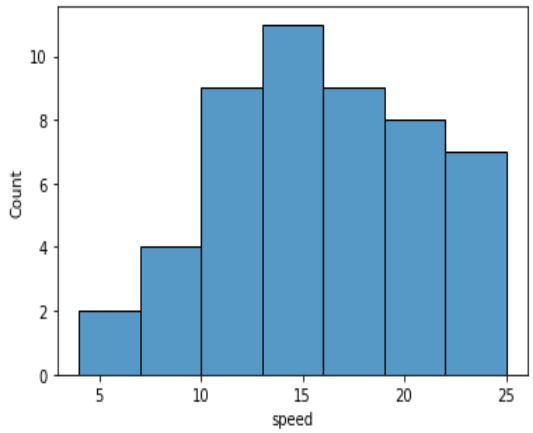
Cars speed and distance

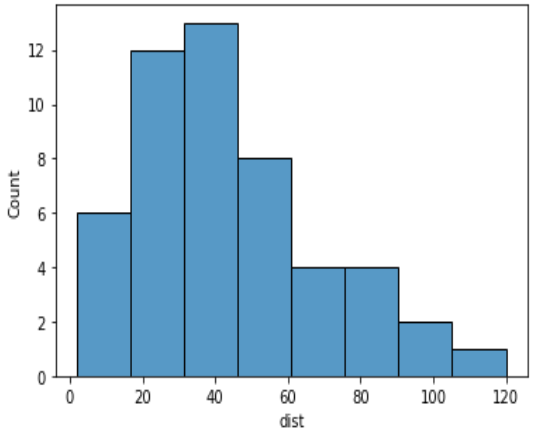
Use Q9_a.csv

Answer -

Cars parameters	skewness	kurtosis
speed	-0.11395477	-0.57714742
dist	0.78248352	0.24801866







Inference –

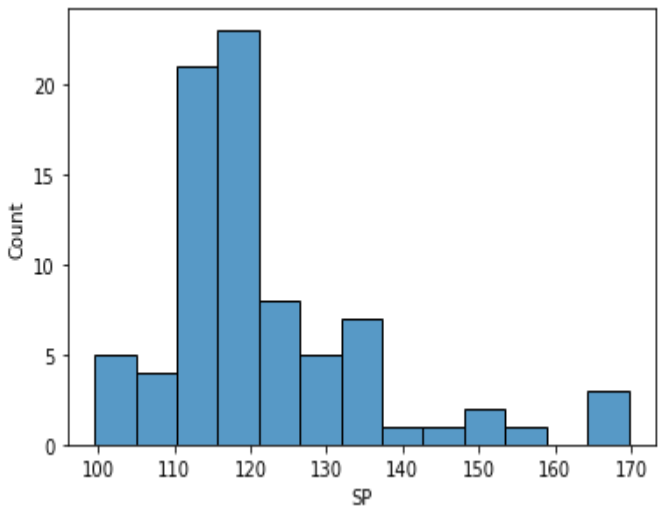
Skewness and kurtosis of speed is negative as we see the plot, Speed is left skewed most of the values are on the left side. On the other side skewness and kurtosis of distance is positive, distance is right skewed most of the values are on the right side. There will be extreme values in distance as kurtosis is positive.

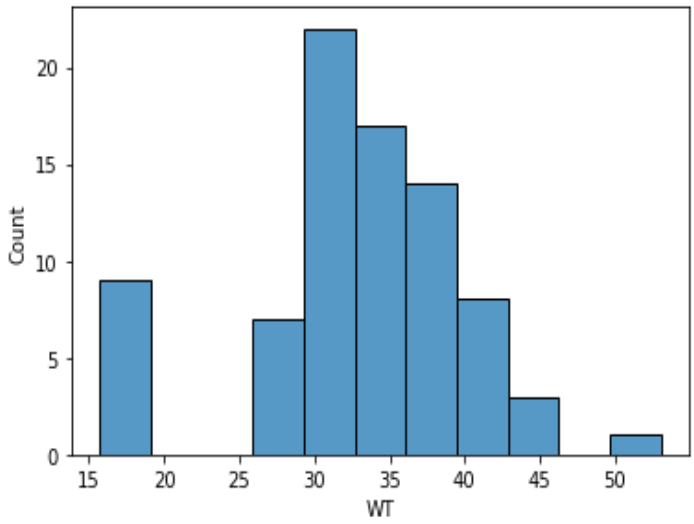
SP and Weight(WT)

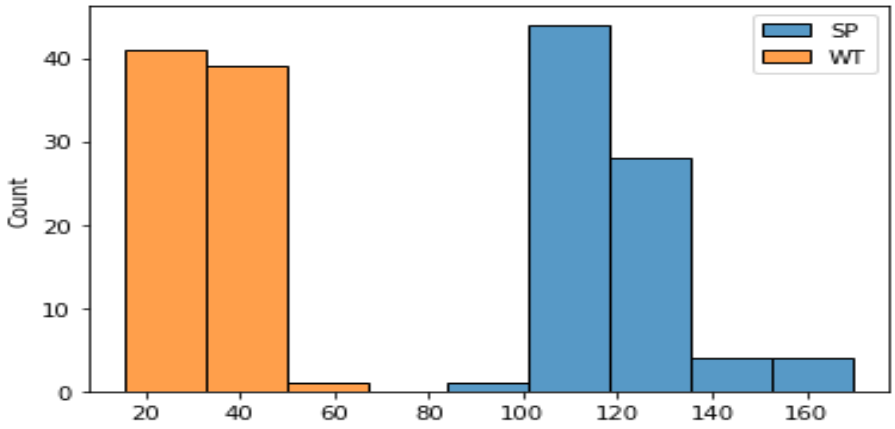
Use Q9_b.csv

Answer -

Cars parameters	Skewness	Kurtosis
SP	1.58145368	2.72352149
WT	-0.60330993	0.81946588



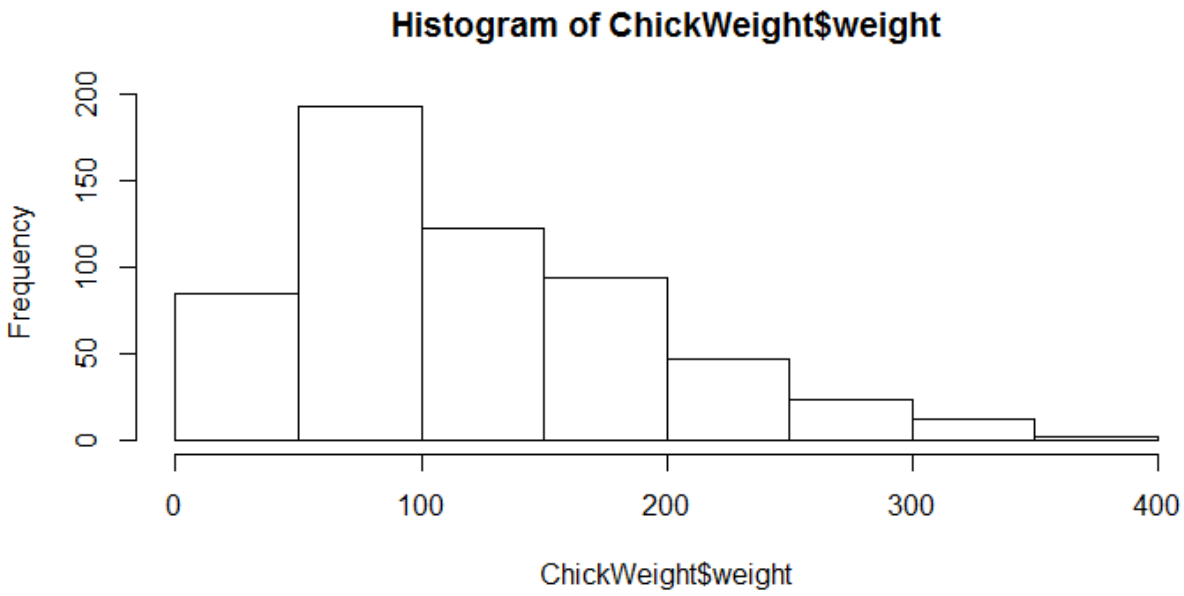


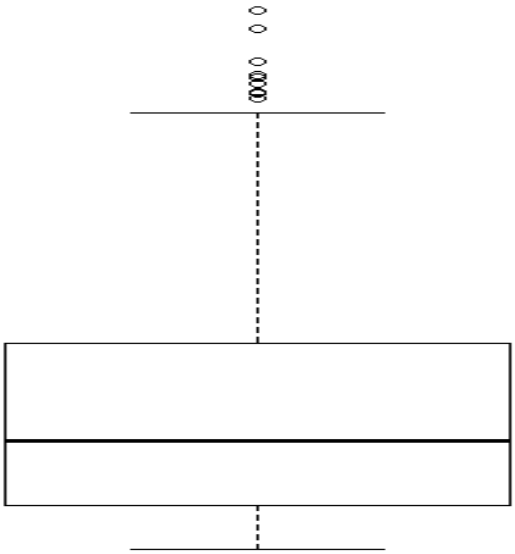


Inference –

SP has positive values for both skewness and kurtosis it shows that it is right skewed. WT has negative value for skewness and positive for kurtosis which means that it is left skewed and extreme values in the distribution ranged from 25 to 40 as we see from the above histogram of WT.

Q10) Draw inferences about the following boxplot & histogram





Answer –

The distribution's histogram shows that most chickweights lies in range 50-100 and decreases for high chick.

The boxplot is toward the lower limit. There are outliers towards the upper limit(1st quartile range) . the median (2rd quartile) is towards the lower limit.

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?

Answer –

Confidence level	Confidence Interval
94%	(143.57619175546247, 256.42380824453755)
96%	(138.38753268104531, 261.6124673189 5466)
98%	(130.2095637787748, 269.79043622122 52)

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?

Answer –

1) Mean = 41.00, Median= 40.5, Variance = 25.53, Standard Deviation = 5.052

2) Histogram plot of student mark

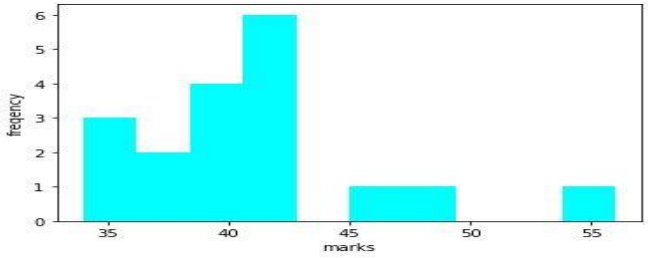
```
In [13]: marks = [34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56]

In [4]: import pandas as pd
        import matplotlib.pyplot as plt
        import numpy as np

In [14]: q = pd.DataFrame(marks)

In [25]: plt.hist(q, color = 'cyan')
         plt.xlabel('marks')
         plt.ylabel('freqency')

Out[25]: Text(0, 0.5, 'freqency')
```



2 answer – the graph represents that most of the students scored between 40 to 45 marks. Maximum marks obtained was 56.

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

Answer – it shows that data is normalized and perfectly symmetric. Skewness is 0.

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

Answer – negative skewed (left tailed distribution)

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

Answer – positive skewed(right tailed distribution)

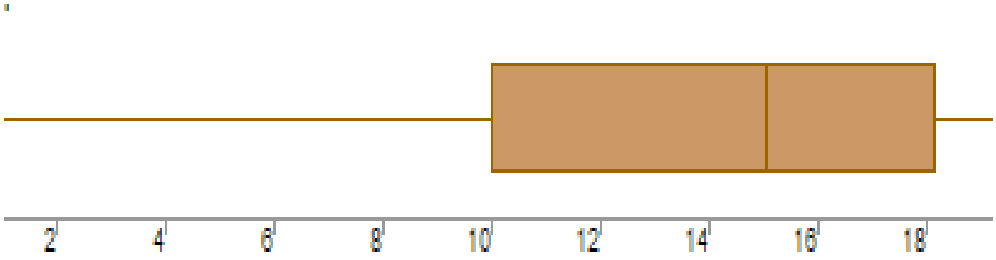
Q16) What does positive kurtosis value indicates for a data ?

Answer – wide tail and and peak.

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

Answer – thin tail and flattened peak.

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



What can we say about the distribution of the data?

Answer – uneven data distribution

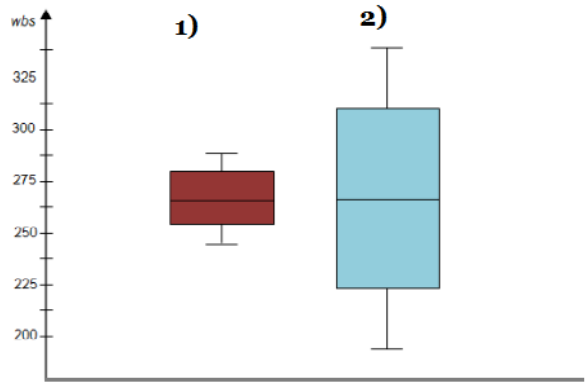
What is nature of skewness of the data?

Answer – negatively skewed (left tailed distribution)

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

Answer – IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

Answer – the data of boxplot 1 is less spread ranging between 240 and 280 with 40 IQR(approx.) . The data of boxplot 2 is more spread ranging between 200 and 330 with 130 IQR(approx.). the median is same for both the boxplots.

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)

Answer – 0.348

b. P(MPG<40)

Answer – 0.729

c. P (20<MPG<50)

Answer -0.013

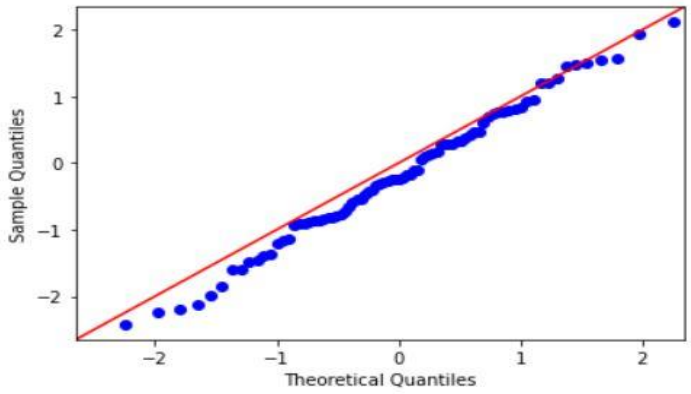
Q 21) Check whether the data follows normal distribution

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

Answer – yes MPG follows normal dist.

```
In [28]: sm.qqplot(cars.MPG, line = '45')
```

Out[28]:

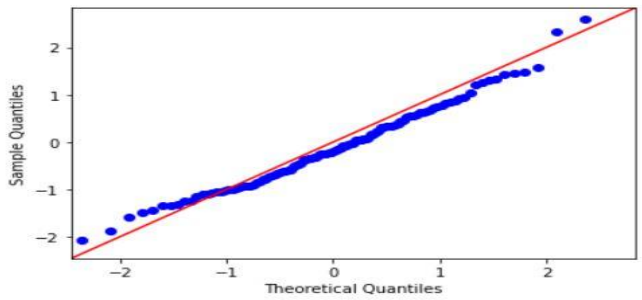


b) Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution
Dataset: wc-at.csv

Answer – yes AT follows normal dist.

```
In [40]: q21.AT = stats.norm.rvs(size = 109)
         sm.qqplot(q21.AT, line = '45')
```

Out[40]:



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

Answer -

CI	zscore
0.6	0.8416212335729143
0.94	1.8807936081512509
0.9	1.6448536269514722

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

Answer -

Confidence intervals	T scores
0.99	2.796939504772804
0.96	2.1715446760080677
0.95	2.0638985616280205

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

answer – prob of n bulbs with life <260

pop mean = 270

sample mean = 260

sample st. dv. = 90

n = 18

to calculate t scores

t score = -0.4714045207910317

probability = 0.32167253567098364 (32%)