

**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)	Categorical

**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	Ordinal
Time on a Clock with Hands	Interval
Number of Children	Ratio
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?

**Solution** :- When three coins are tossed the total number of possible combinations are  $2^3 = 8$ . The number of combinations which have two heads and one tail are: HHT, HTH, THH. Number of desired outcome is 3.

The probability of getting two heads and one tail in the toss of three coins is  $\frac{3}{8} = 0.375$ .

**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** :-  $n(S) = 36$

a) If two dice are rolled , probability that the sum is equal to 1 is zero.

b) Less than or equal to 4

$$B = \{(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)\}$$

$$n(B) = 06$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{6}{36} = \frac{1}{6} = 0.1666$$

c) Sum is divisible by 2 and 3

$$A = \{(1,5), (2,4), (3,3), (4,2), (5,1), (6,6)\}$$

$$n(A) = 06$$

$$P(B) = \frac{n(A)}{n(S)} = \frac{6}{36} = \frac{1}{6} = 0.1666$$

**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 =  $2+3+2 = 07$

Number of ways of drawing two balls out of 7 =  $n(S) = {}^7C_2 = \frac{7 \times 6}{2 \times 1} = 21$

Number of drawing two balls out of (2+3) balls none of blue =

$$n(A) = {}^5C_2 = \frac{5 \times 4}{2 \times 1} = 10$$

$$\therefore P(A) = \frac{n(A)}{n(S)} = \frac{10}{21} = 0.4762$$

**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

**Solution:-**

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Expected number of candies for a randomly selected child are

$$1*0.015+4*0.20+3*0.65+5*0.005+6*0.01+2*0.120 = 3.09$$

**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 :-** import pandas as pd

import numpy as np

df=pd.read\_csv("Q7.csv")

# Mean, Median, mode, standard deviation , variance, range for **Points** .

df["Points"].mean() = 3.5965625000000006

df["Points"].median() = 3.6950000000000003

df["Points"].mode() = 3.07 & 3.92

```
df["Points"].var() = 0.28588135080645166
df["Points"].std() = 0.5346787360709716
range=df["Points"].max()-df["Points"].min() = 2.17
```

#Mean, Median, mode, standard deviation , variance, range for **Score** .

```
df["Score"].mean() = 3.2172499999999995
df["Score"].median() = 3.325
df["Score"].mode() = 3.44
df["Score"].var() = 0.9573789677419356
df["Score"].std() = 0.9784574429896967
range=df["Score"].max()-df["Score"].min() = 3.9110000000000005
```

#Mean, Median, mode, standard deviation , variance, range for **Weigh** .

```
df["Weigh"].mean() = 17.848750000000003
df["Weigh"].median() = 17.71
df["Weigh"].mode() = 17.02 & 18.90
df["Weigh"].var() = 3.19316629032258
df["Weigh"].std() = 1.7869432360968431
range=df["Weigh"].max()-df["Weigh"].min() = 8.399999999999999
```

Points and Weigh is bimodal and score is unimodal. Range is very high in weigh compare to score and points. Variance and standard deviation are almost same in Score.

**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 :-** Probability of weights is  $\frac{1}{9}$

Expected Value of the Weight of that patient are

$$108 * \frac{1}{9} + 110 * \frac{1}{9} + 123 * \frac{1}{9} + 134 * \frac{1}{9} + 135 * \frac{1}{9} + 145 * \frac{1}{9} + 167 * \frac{1}{9} + 187 * \frac{1}{9} + 199 * \frac{1}{9}$$

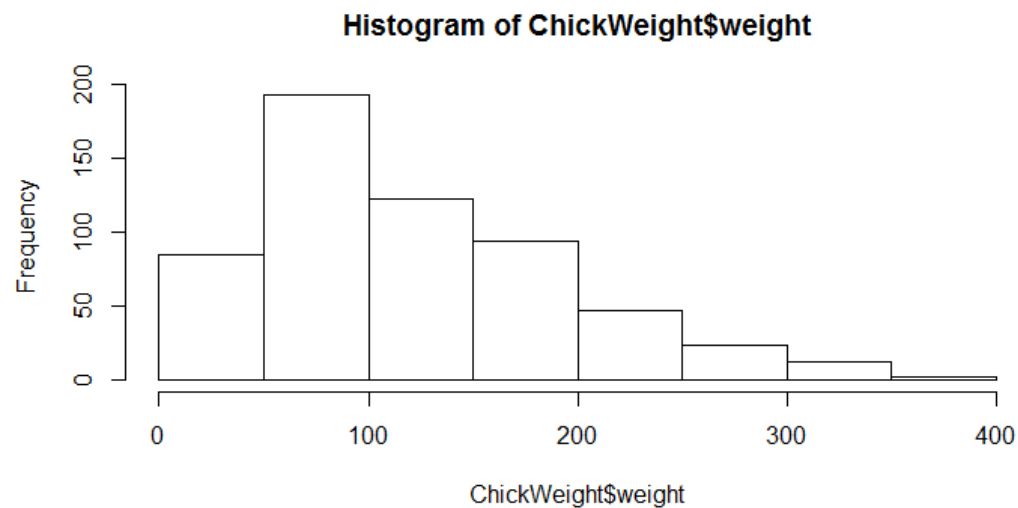
$$= 322.22222222222223$$

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

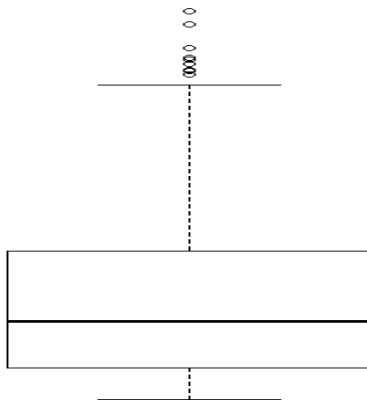
**Solution:-** `df["SP"].skew() = 1.6114501961773586`  
`df["WT"].skew() = -0.6147533255357768`  
`df["SP"].kurt() = 2.9773289437871835`  
`df["WT"].kurt() = 0.9502914910300326`

Speed is positively skewed where as mean>median>mode and less positive kurtosis. Weight is negative skewed where as mode>median>mean and less positive kurtosis .

### Q10) Draw inferences about the following boxplot & histogram



Histogram is used for continuous data set . The given graph is positively skewed and has its peak to the left , gradual tapering to the right. It has outliers to the right hand side.



**Solution:-**Box plot is applied for continuous data set. It has 1<sup>st</sup> quartile, 2<sup>nd</sup> quartile and 3<sup>rd</sup> quartile and also having minimum and maximum point. It has outliers above the maximum point. Median is nearer to 1<sup>st</sup> quartile. Distance between q3 and q2 is more compare to q2 and q1. Inter quartile range (IQR) = q3-q1 .

**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:-** from scipy import stats

```
import scipy
```

```
stats.norm.interval( 0.94,200,30) = (143.57619175546247, 256.42380824453755)
```

```
stats.norm.interval( 0.98,200,30) = (130.2095637787748, 269.7904362212252)
```

```
stats.norm.interval( 0.96,200,30) = (138.38753268104531, 261.61246731895466)
```

**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:-**

```
x=pd.Series([34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56])
```

```
x.mean() = 41.0
```

```
x.median() = 40.5
```

```
x.var() = 25.529411764705884
```

```
x.std() = 5.052663828586
```

2) What can we say about the student marks?

**Solution:-** Mean of the student marks is 41. Meadian is 40.5. Variance is 25.5294. Standard deviation is 5.0527.

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

**Solution:-** Non skewed

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

**Solution:-** Positively skewed

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

**Solution:-** Negatively skewed

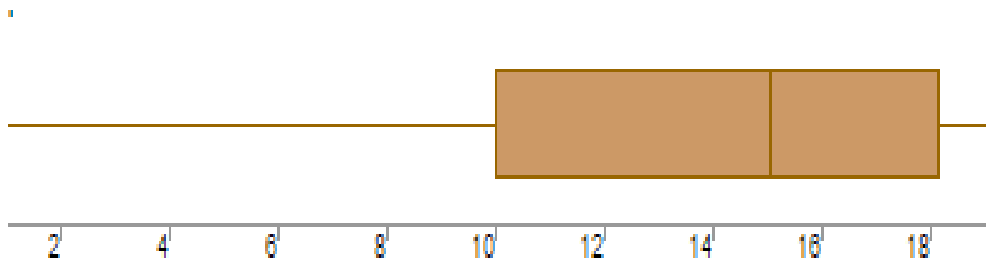
Q16) What does positive kurtosis value indicates for adata ?

**Solution:-** Peakendness is very high.

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

**Solution:-** Wider peak and thinner tails.

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



What can we say about the distribution of the data?

**Solution:-** Maximum point is 20 . Mean is less than median .

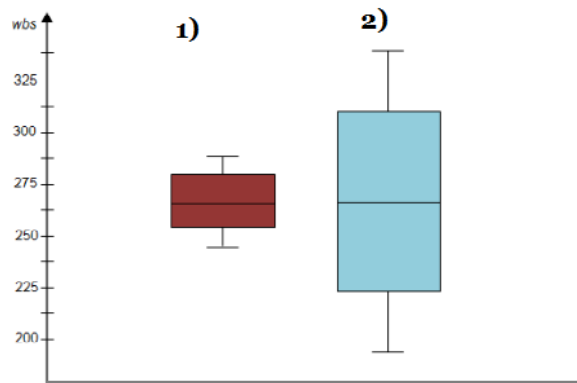
What is nature of skewness of the data?

**Solution:-** The nature of skewness of the data is negative skewness.

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

**Solution:-**  $q_3 - q_1 = 8$

**Q19)** Comment on the below Boxplot visualizations?



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

**Solution:-** Median for both boxplot 1 and 2 is 262.5. Median > mean > mode in both the plots. Therefore both are positively skewed. Data is larger in plot 2 compare to plot 1. Both has maximum and minimum points.

**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.

**Solution:-**

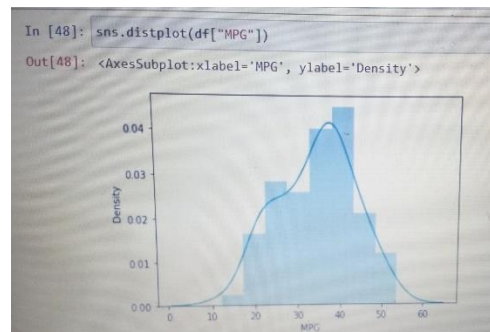
- a.  $P(\text{MPG} > 38) = 65\%$
- b.  $P(\text{MPG} < 40) = 80\%$
- c.  $P(20 < \text{MPG} < 50) = 89.88\%$

**Q 21)** Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution



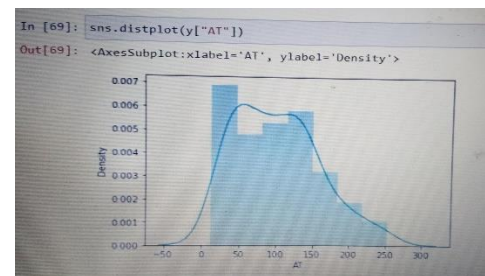
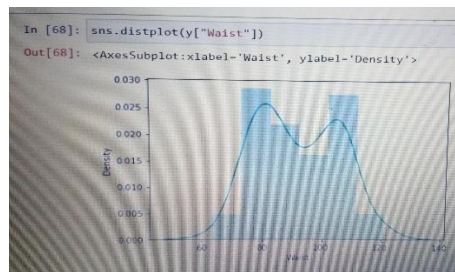
Dataset: Cars.csv



In this graph mean median mode is not equal so it is not 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



**Solution:-**In this data also mean median mode is not equal . Graph is not in bell shape. Therefore it is not normally distributed.

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

**Solution:-** Z-scores of 90% confidence interval = 1.645

Z-scores of 94% confidence interval = 1.88

Z-scores of 60% confidence interval = 0.842

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

**Solution:-** t-scores of 95% confidence interval = 2.060

t-scores of 96% confidence interval = 2.284

t-scores of 99% confidence interval = 2.787

**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

**Solution:-** Given, Mean = 270

Sample mean = 260

Sample size = 18

Standard deviation = 90

T statistic =  $(270-260)/(90/\text{np.sqrt}(18)) = 0.47$

$2*\text{stats.t.cdf}(0.47, \text{df}=17) = 1.35665$