

Q1) Identify the Data type for the Following:

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

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

Data	Data Type
Gender	<b>Nominal</b>
High School Class Ranking	<b>Ordinal</b>
Celsius Temperature	<b>Interval</b>
Weight	<b>Ratio</b>
Hair Color	<b>Nominal</b>
Socioeconomic Status	<b>Ordinal</b>
Fahrenheit Temperature	<b>Interval</b>
Height	<b>Ratio</b>
Type of living accommodation	<b>Ordinal</b>
Level of Agreement	<b>Ordinal</b>
IQ(Intelligence Scale)	<b>Ratio</b>
Sales Figures	<b>Ratio</b>
Blood Group	<b>Nominal</b>
Time Of Day	<b>Ordinal</b>
Time on a Clock with Hands	<b>Interval</b>
Number of Children	<b>Ratio</b>
Religious Preference	<b>Nominal</b>

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?

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

**$P[X] = X[\text{HHH, HHT, HTH, THH, THT, TTH, HTT, TTT}]$**

**TOTAL 8 OUTCOMES**

**2HEADS, 1TAIL= [HHT, HTH, THH]**

**TOTAL 3 SUCH OUTCOMES**

**$P[2HEADS, 1TAIL] = 3/8 = .375$**

**HENCE 37% probability that two heads and one tail are obtained**

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

**2DICE RANDOMLY THROWN =**

**$P[X] = X[2, 3, \dots, 12]$**

**TOTAL = 36[6X6] OUTCOMES**

**a)  $P[X=1]=0$**

**b)  $P[X \leq 4] = [(1, 1), (1, 2), (1, 3), (2, 2), (2, 1), (3, 1)] / 36 = 6/36 = 1/6 = 0.166$   
16.6% PROBABILITY THAT SUM IS LESS THAN OR EQUAL TO 4**

**c)  $P[X/2 \text{ OR } 3] = [(2, 4), (4, 2)] / 36 = 2/36 = 1/18 = 0.0555$**

**5.5% PROBABILITY THAT SUM IS DIVISIBLE BY 2 AND 3**

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?

**TOTAL BALLS=7 COMBINATIONS OF 2 BALLS**

**HENCE TOTAL COMBINATIONS= ${}^7C_2 = \frac{7!}{(2! \cdot 5!)} = \frac{(7 \cdot 6)}{2} = 21$**

**5 BALLS ARE NOT BLUE**

**THUS 5 COMBINATIONS OF 2 BALLS= ${}^5C_2 = \frac{5!}{2!3!} = \frac{(5 \cdot 4)}{2} = 10$**

**PROBABILITY OF NOT BLUE BALL DRAWN=  $\frac{10}{21} = 0.4762$**

**47.62% IS the probability that none of the balls drawn is blue.**

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:

**Expected number of candies for a randomly selected child=**

**$(1 \cdot .015) + (4 \cdot .2) + (3 \cdot .65) + (5 \cdot .005) + (6 \cdot .01) + (2 \cdot .12) = 3.09$**

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: # python code

```
import pandas as pd  
df= pd.read_csv('Q7.csv')  
df.describe()
```

	Points	Score	Weigh
Mean	3.60	3.22	17.85
Median	3.70	3.32	17.71
Mode	3.07	3.44	17.02
Standard Deviation	0.53	0.97	1.78
Variance	0.29	0.96	3.19
Range [Min-Max]	2.76 – 4.93	1.51 – 5.42	14.50 – 22.9

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:

**Expected Value =  $\sum (\text{probability} * \text{Value})$**

**Probability of selecting each patient = 1/9**

**Expected value=  $\sum [x.p(x)]$**

**Expected value = Sum (X \* Probability of X)**

**= (1/9)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+  
(1/9)(187)+ (1/9)(199)**

**= 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**

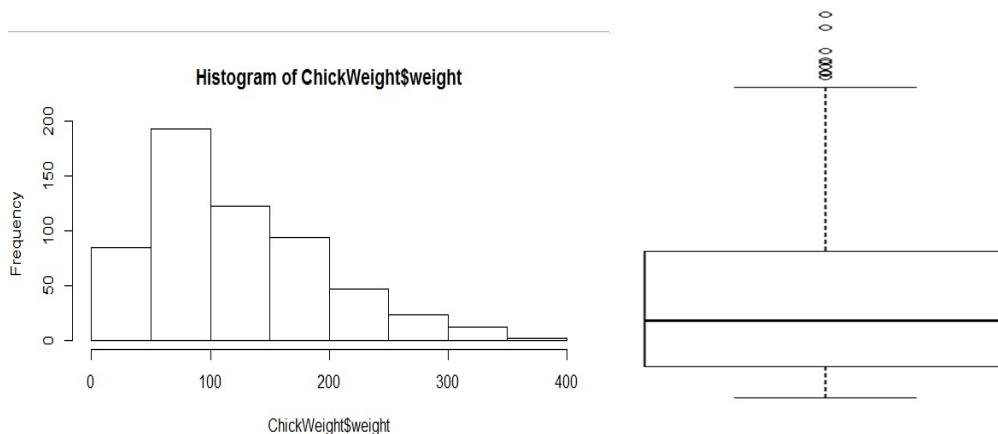
**Ans: python code IN jupyter notebook**

**Skewness for speed= -0.12, skewness value is negative so it is left**

**skewed. Since magnitude is greater than 0 it is slightly left skewed**

**distance= 0.78, right skewed (Positive) slight magnitude to right.**

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



**Ans: Most data concentrated at 50-100 range with frequency upto 200 and least between 350 and 400**

**The histogram's peak has right skew and tail is on right.**

**Mean > Median.**

Ans: The boxplot has outliers on the maximum side.

Here also Mean > Median

**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?

```
conf_94 = stats.t.interval(alpha = 0.94, df=1999, loc=200,
                           scale=30/np.sqrt(2000))
print(np.round(conf_94,0))
```

```
print(conf_94)
```

For 94% confidence interval Range is [ 198.73 – 201.26]

For 98% confidence interval range is [198.43 – 201.56]

For 96% confidence interval range is [198.62 – 201.37]

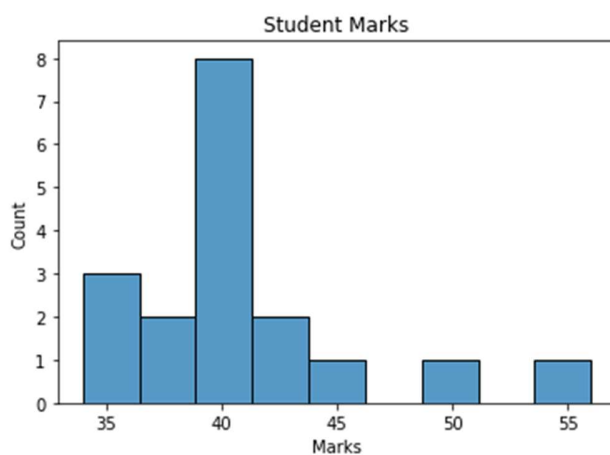
**Q12)** Below are the scores obtained by a student in tests

**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: Mean =41, Median =40.5, Variance =25.52 and Standard Deviation =5.05

we don't have outliers and the data is slightly skewed towards right because mean is greater than median.



From plot we can say that mean of marks of student is 41 which is slightly greater than median. Most of the students got marks in between 41-43.

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

**If the mean is equal to the median as well as the mode, hence the skewness is zero**

**No skewness is present we have a perfect symmetrical distribution**

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

**Ans: If the mean is greater than the median, then distribution is positively skewed. Skewness and tail is towards Right**

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

**Ans: : If the mean is less than the median, the distribution is negatively skewed. Skewness and tail is towards left**

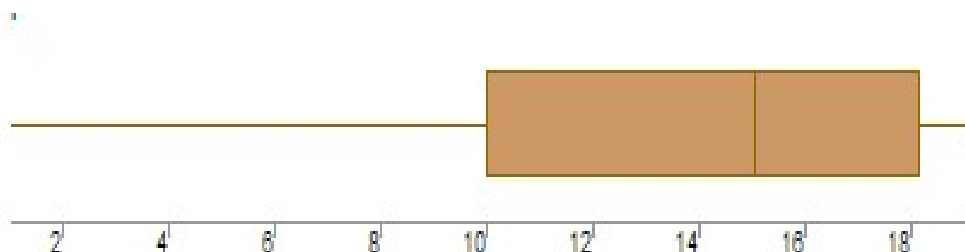
Q16) What does positive kurtosis value indicates for a data ?

**Ans: Positive kurtosis means the data distribution is more peaked and have thick tails.**

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

**Ans: Negative Kurtosis means the peak will be flatter and broader in the normal distribution. This simply means that more data values are located near the mean and less data values are located on the tails. Negative kurtosis is the uniform distribution, which has no peak at all and is a completely flat distribution.**

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



What can we say about the distribution of the data?

**Ans: The above Boxplot is not normally distributed and the median is towards the higher value(right side of the tail)**

What is nature of skewness of the data?

**Ans: data is a skewed towards left and median is greater than mean**

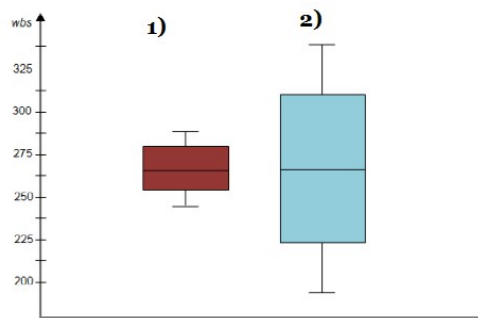
**The whisker range of minimum value is greater than maximum**



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

Ans: **The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 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:..

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.  $P(\text{MPG} > 38)$
- b.  $P(\text{MPG} < 40)$
- c.  $P(20 < \text{MPG} < 50)$

Ans:python code in jup notebook

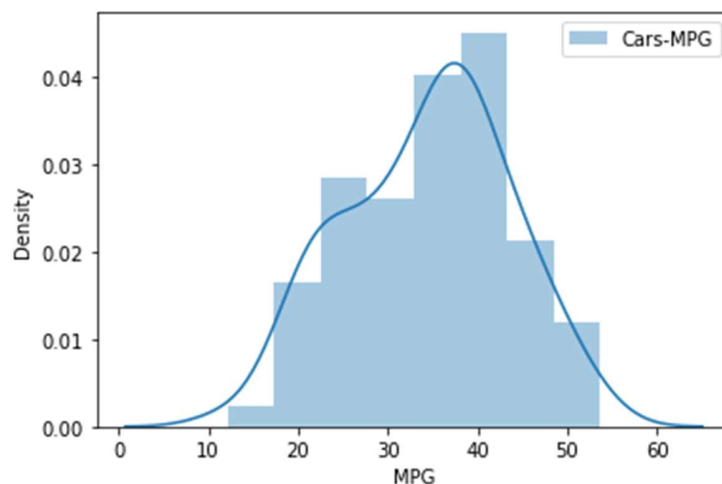
- a) **34% of the cars are greater than 38 MPG.**
- b) **72% of the cars are less than 40 MPG.**
- c) **89% of the cars are in between 20 to 50 MPG.**

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans: MPG of cars follows normal distribution, skewness=-0.177**

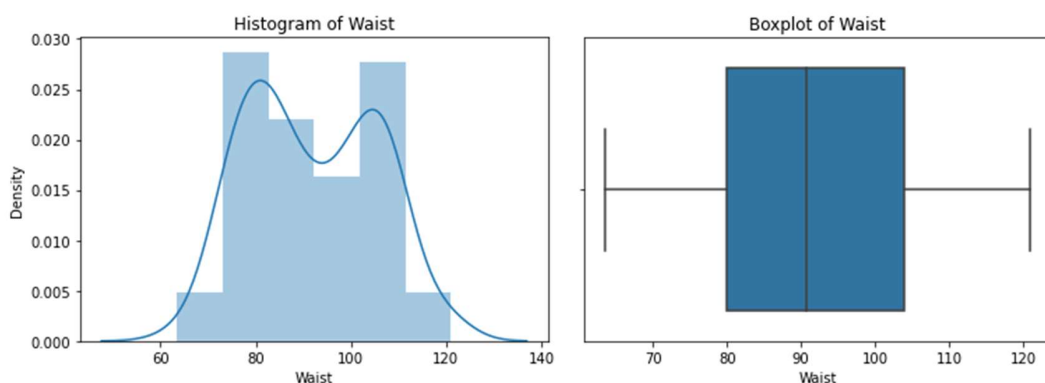


b) Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

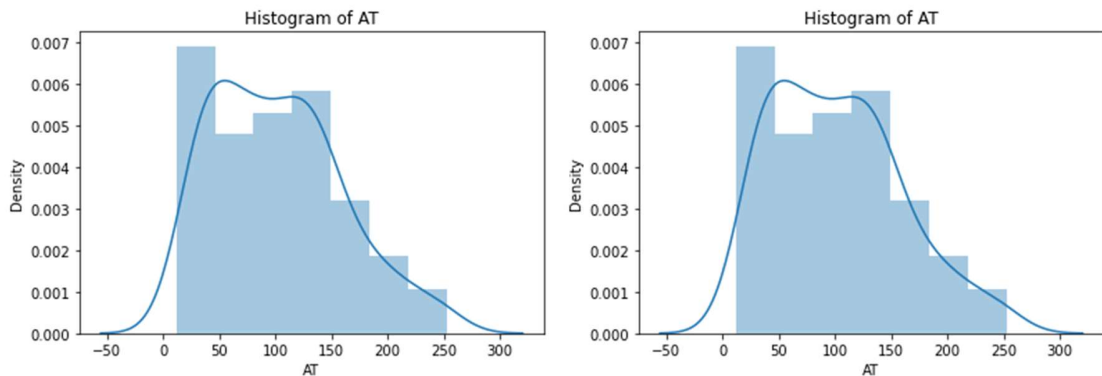
Dataset: wc-at.csv

**Ans: Adipose Tissue (AT) and Waist does not follow Normal Distribution**

**mean greater than median, both the whisker is of same length, median is slightly shifted towards left. Data is fairly symmetric**



**Mean greater than median, right whisker is larger than left whisker, data is positively skewed.**



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

Z score of 60% Confidence Interval = 0.841

Z score of 90% Confidence Interval = 1.644

Z score of 94% Confidence Interval = 1.880

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

Confidence Interval	T Score
95%	2.06
96%	2.17
99%	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

Ans: IN JUP NOTEBOOK

**Population mean = 270 days**

**Sample mean = 260 days**

**Sample SD = 90 days**

**Sample n = 18 bulbs**

**df = n-1 = 17**

$$t = \{(260-270) / (90/\sqrt{18})\}$$

$$t = (-1 * \sqrt{2}) / 3$$

$$t = - 0.471$$

-----X-----