

**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, and Ratio.**

Data	Data Type
Gender	Nominal
High School Class Ranking	Ordinal
Celsius Temperature	Interval
Weight	Ratio
Hair Color	Nominal
Socioeconomic Status	Interval
Fahrenheit Temperature	Interval
Height	Ratio
Type of living accommodation	Nominal
Level of Agreement	Ordinal
IQ(Intelligence Scale)	Ordinal
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	Interval
Years of Education	Ordinal

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

**Ans:**

Population Set=[HHH,HHT,HTH,THH,TTH,THT,HTT,TTT] = 8

Two heads and one tail = [HHT,HTH,THH] = 3

**So probability is = 3/8**

**Q4) Two Dice are rolled, find the probability that the sum is**

- a) Equal to 1
- b) Less than or equal to 4
- c) Sum is divisible by 2 and 3

**Ans:** Population Set=[ (1,1),(1,2),(1,3).....(1,6)  
(2,1),(2,2).....(2,6)  
(3,1).....(3,6)  
(4,1).....(4,6)  
(5,1).....(5,6)  
(6,1).....(6,6) ] = 36

**a)Equal to 1=[Null] = 0**

**Probability=0/36=0**

**b)Less than or equal to 4 = [(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)] = 6**

**Probability=6/36= 1/6**

**c)Sum is divisible by 2 and 3 = [(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)] = 6**

**Probability=6/36 = 1/6**

**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 no. of balls=7

Two balls drawn at random =  ${}^7C_2 = \frac{7*6}{2*1} = 21$

None balls are blue=  ${}^5C_2 = \frac{5*4}{2*1} = 10$

**The probability that none of the balls drawn is blue =  $\frac{10}{21}$**

**Q6) Calculate the Expected number of candies for a randomly selected child**

**Below are the probabilities of the 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 – the probability of having 1 candy = 0.015.**

**Child B – probability of having 4 candies = 0.20**

**Ans:**

The expected number of candies for a randomly selected child is =

$$(1*0.015+4*0.20+3*0.65+5*0.005+6*0.01+2*0.120)$$

$$= 3.09$$

**Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset**

- **For Points, Score, and Weigh.**  
**Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.**

Use Q7.csv file

**Ans:**

	<b>Points</b>	<b>Score</b>	<b>Weigh</b>
<b>Mean</b>	3.596	3.217	17.84
<b>Median</b>	3.69	3.325	17.71
<b>Mode</b>	3.07/3.92	3.44	17.02/18.9
<b>Variance</b>	0.285	0.957	3.193
<b>Standard Deviation</b>	0.534	0.978	1.786
<b>Range</b>	2.17	3.911	8.4

Comments: 1) As the mean is approximately equal to the median. thus we can say that the data is normally distributed.

2)As the variance is far from the mean we can say that data is widely spread.

**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:**

Total no. of patients= 9

Probability of a patient having weight is =  $1/9$

Expected Value of the Weight of that patient =

$$\begin{aligned} & [(1/9)*108+(1/9)*110+(1/9)*123+(1/9)*134+(1/9)*135+(1/9)*145+(1/9)*167+(1/9)*187+(1/9)*199] \\ & = \underline{\underline{145.318}} \end{aligned}$$

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

	<b>Speed</b>	<b>Distance</b>
<b>Skewness</b>	-0.117510	0.806895
<b>Kurtosis</b>	-0.508994	0.405053

Inference: 1) As the Skewness for Speed is negative, the plot is left-tailed.

2) As the Skewness for Distance is positive, the plot is right-tailed.

3) The Kurtosis of Speed is negative, the distribution is wider, thin-tailed.

4) The Kurtosis of Distance is positive, and the distribution is having a peak.

## SP and Weight(WT)

Use Q9\_b.csv

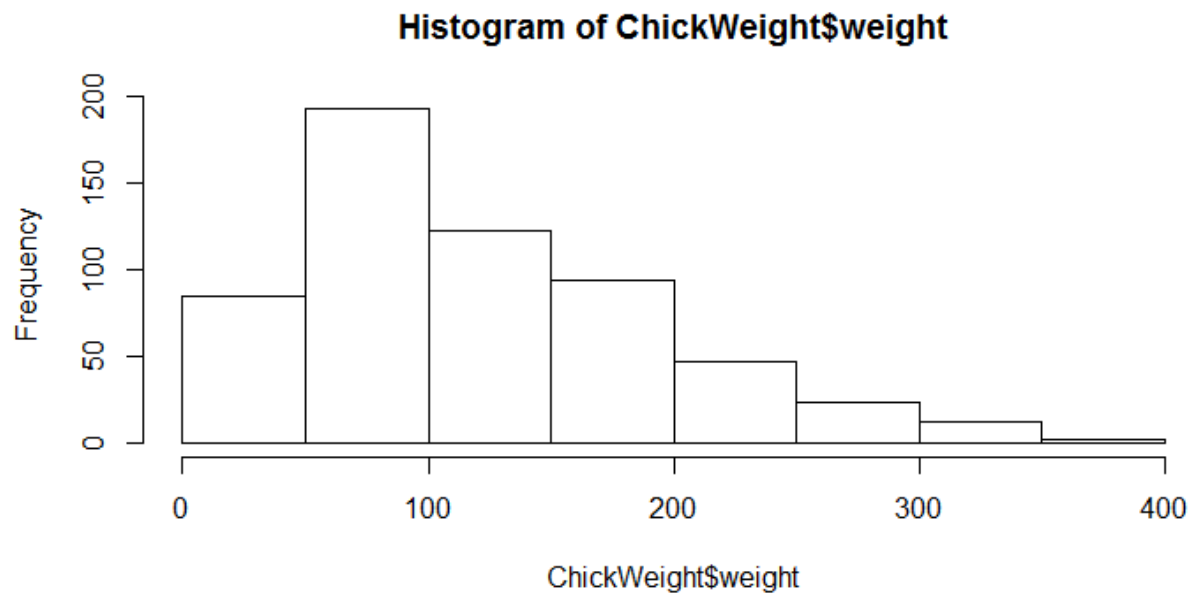
Ans:

	SP	WT
Skewness	1.637477	-0.623600
Kurtosis	3.132348	1.048415

- Inference: 1) As the Skewness for SP is positive, the plot is right-tailed.  
2) As the Skewness for WT is negative, the plot is left-tailed.  
3) The Kurtosis of SP is positive, the distribution is having peak  
4) The Kurtosis of WT is positive, and the distribution is having a peak.

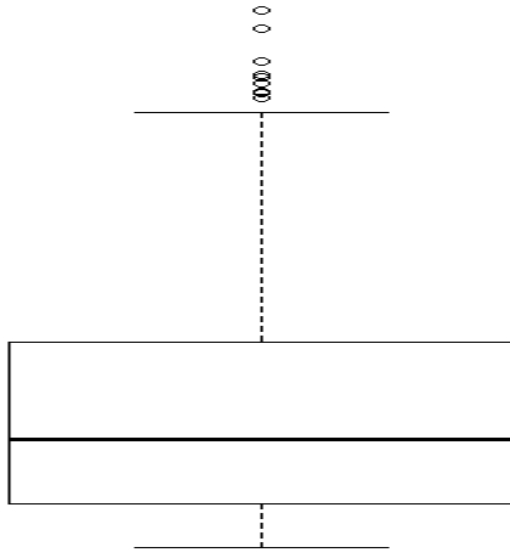
**Q10) Draw inferences about the following boxplot & histogram**

---



Inferences:

- 1) Interval 50-100 has a maximum frequency.
- 2) The histogram is positively skewed i.e. left tailed.
- 3) Interval 350-400 has a minimum frequency.



Inferences:

- 1) The boxplot has approximately 7 outliers that needed to be treated.
- 2) The outliers are on the maximum side.

**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%, and 96% confidence intervals.**

**Ans:**

Confidence Intervals		
94%	198.738325292158	201.261674707842
96%	198.62230334813333	201.37769665186667
98%	198.43943840429978	201.56056159570022

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

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

- 1) Find mean, median, variance, and standard deviation.**
- 2) What can we say about the student marks?**

**Ans:**

Mean=41

Median=40.5

Variance=25.529

Standard deviation = 5.052

- 1) The Mean is greater than median so the data is right-tailed.
- 2) Maximum students have marks between 39 to 41.

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

**Ans:** When mean and median are equal the data is symmetric in nature i.e. the data follows normal distribution.

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

**Ans:** When mean > median the data is right-tailed which means data is concentrated more on the left side.

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

**Ans:** When median > mean the data is left-tailed which means data is concentrated more on the right side.

**Q16) What does a positive kurtosis value indicate for data?**

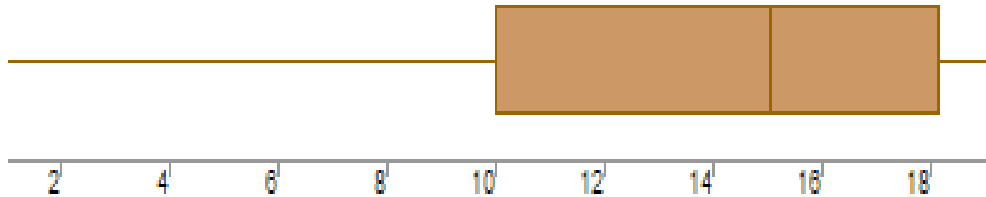
**Ans:** Positive kurtosis indicates peakedness in the data.



**Q17) What does a negative kurtosis value indicate for data?**

**Ans:** Negative kurtosis indicates data has wider peaks and thin tails.

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



**a)What can we say about the distribution of the data?**

**Ans:** The data is asymmetrically distributed towards the maximum side.

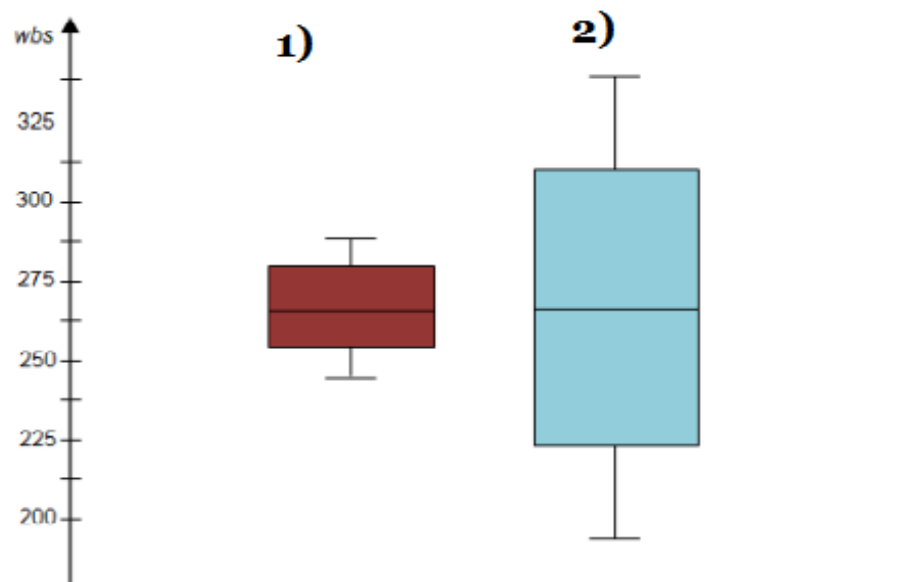
**b)What is the nature of the skewness of the data?**

**Ans:** The data is negatively skewed.

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

**Ans:** IQR = 8.

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



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

- Ans:** 1) Both plots have no outliers.  
2) Both plots have approximately the same median.  
3) Both plots have no skewness from both sides.

**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.  $P(\text{MPG} > 38)$**

**Ans:** 0.34759394041453007

**b.  $P(\text{MPG} < 40)$**

**Ans:** 0.7293498604157946

**c.  $P(20 < \text{MPG} < 50)$**

**Ans:** 0.89886890

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

**a) Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

**Ans:** Mean= 34.4220, Median=35.1527

As we can see the mean and median are approximately the same we can say the MPG data is normally disturbed.

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

**Dataset: wc-at.csv**

**Ans:**

a) Mean= 91.901835, Median=90.8000

As we can see the mean and median are approximately the same we can say the **Waist Circumference(Waist)** data is normally distributed.

b)a) Mean= 101.8940, Median=96.5400

As we can see the mean and median are slightly different so we can say **Adipose Tissue (AT)** the data is slightly skewed, and the right tail.

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

Ans: Z scores of

a) 90% confidence interval= 1.644853

b) 94% confidence interval= 1.880793

c) 60% confidence interval= 0.841621

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

Ans: t scores of

a) 95% confidence interval= 2.0638

b) 96% confidence interval= 2.17154

c) 99% confidence interval= 2.7969

**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:**

code □ `pt(tscore, df)`

df □ degrees of freedom

Ans: The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.32167.