

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 / CATEGORICAL
Number of kids	DISCRETE
Number of tickets in Indian railways	DISCRETE
Number of times married	DISCRETE
Gender (Male or Female)	DISCRETE / 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?

Ans =  $3/8$  , {HHT,HTH,TTH}

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

- a) Equal to 1      Ans = 0
- b) Less than or equal to 4      Ans =  $1/6$
- c) Sum is divisible by 2 and 3      Ans =  $6/36=1/6$

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?

$$\begin{aligned}
 \text{Ans} &= P(E) = n(E) / n(S) \\
 P(5/7, 4/6) &= 20/42 \text{ i.e. } 10/21 \\
 {}^7C_2 &= (7*6)/(2*1) = 21 \\
 {}^5C_2 &= (5*4)/(2*1) = 10 \\
 &= 10/21
 \end{aligned}$$

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 : -  $E(X) = np$

$$\begin{aligned}
 &\text{Expected number of candies for randomly selected child} \\
 &= 1*0.015 + 4*0.20 + 3*0.65 + 5*0.005 + 6*0.01 + 2*0.120 \\
 &= 3.09
 \end{aligned}$$

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**

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      $x=1308$

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?  
 $x.p(x)=145.3333$

Expected Value of the Weight of that patient is  $x.p(x)= 145.3333$

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

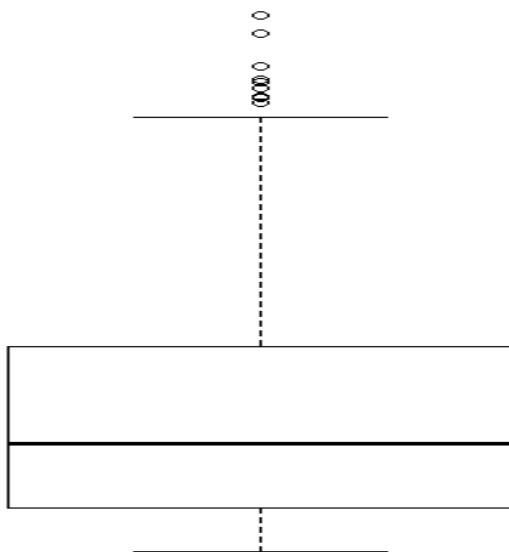
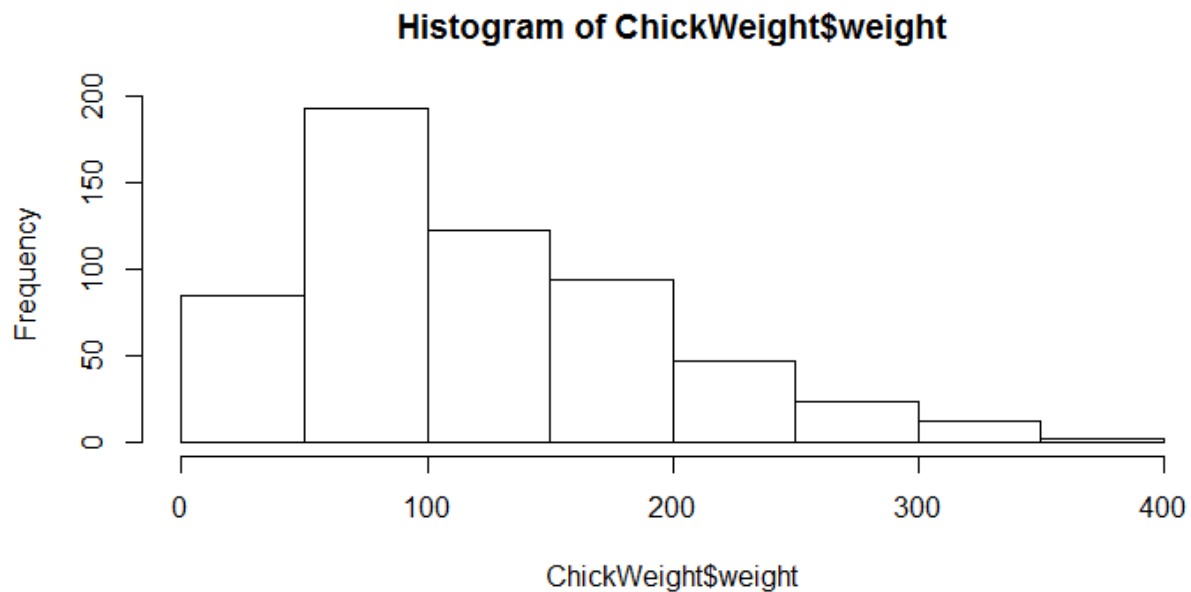
**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

Q10) Draw inferences about the following boxplot & histogram



Ans=the most of the data points are seen in the range 50-100 with frequency 200.  
 And least range of weight is 400 somewhere around 0-10.  
 So the expected value of the above distribution is 75.  
 Skewness- we can notice a long tail towards the right so it is strongly right skewed

Ans =Median is less than means right skewed and we have outlier on the upper side of the box plot and there is less data points between Q1 and bottom point.

**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?

**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 =  $738/18 = 41$   
median = 40  
std = 5.052664  
variance = 25.52941**
- 2) What can we say about the student marks?

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

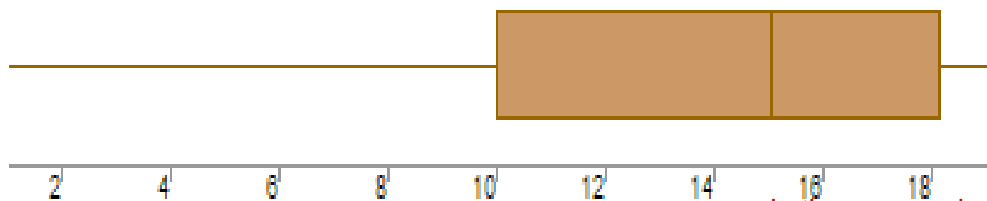
**Q14)** What is the nature of skewness when mean > median ? **Ans = Positive**

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

**Q16)** What does positive kurtosis value indicates for a data? **Ans = Distribution is Peaked and thicker tails, The data is normally distributed and kurtosis value is 0.**

**Q17)** What does negative kurtosis value indicates for a data? **Ans = Distribution is Flat and thinner tails.**

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



**Let's assume above box plot is about age's of the students in a school.**

**What can we say about the distribution of the data?** **50% of the people are above 10 yrs old and remaining are less and students who's age is above 15 are approx 40.**

**Ans = Low to high**

**What is nature of skewness of the data?** **Ans = Negative skewness, median is greater than mean.**

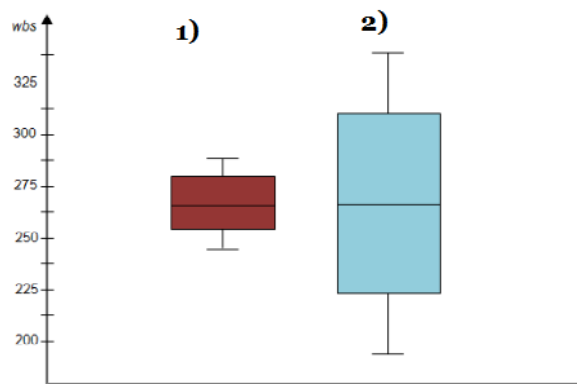
**What will be the IQR of the data (approximately)?**

**Ans =  $Q3 - Q1$**

**=  $18 - 10 = 8$**

**Here,  $Q0 = 2, Q1 = 10, Q2 = 15.5, Q3 = 18, Q4 = 20$   
approximately,  
**IQR = 8****

Q19) Comment on the below Boxplot visualizations?



Ans = By observing both the plots Whisker's level is high in boxplot 2, mean and median are equal hence distribution is symmetrical.

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

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. $P(\text{MPG} > 38)$      | a) There are 33 observations in MPG which are greater than 38.       |
| b. $P(\text{MPG} < 40)$      | b) 61 observations in MPG which are lesser than 40                   |
| c. $P(20 < \text{MPG} < 50)$ | c) 69 observations in MPG which are greater than 20 and less than 50 |

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

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

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

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