

**1) Identify the Data type for the Following:**

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

**2) Identify the Data types, which were among the following  
Nominal, Ordinal, Interval, Ratio.**

<b>Data</b>	<b>Data Type</b>
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	Ordinal
Level of Agreement	Ordinal
IQ(Intelligence Scale)	Ratio
Sales Figures	Ratio
Blood Group	Nominal
Time Of Day	Nominal
Time on a Clock with Hands	Nominal
Number of Children	Ordinal
Religious Preference	Nominal

Barometer Pressure	Interval
SAT Scores	Ordinal
Years of Education	Ordinal

- 3) **Three Coins are tossed, find the probability that two heads and one tail are obtained.**

Sample space =  $2^3$

{(HHH), (HHT), (HTH), (THH), (TTT), (TTH), (THT), (HTT)}

Two head and one tail = {(H, H, T), (H, T, H), (T, H, H)}

Probability =  $3/8$

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

Sample space of two dice:

{(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)

(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)

(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)

(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)

(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)

(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)}

$n(S) = 36$

- a) **Equal to 1:**

**Soln.:** Let

A: There is no outcomes, which corresponds, sum equal to 1 = { }

$n(A) = 0$

$P(A) = n(A)/n(S) = 0/36$

- b) **Less than or equal to 4:**

**Soln.:** Let,

B: Sum is less than or equal to 4

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

$n(B) = 6$

$P(B) = 6/36 = 1/6$

- c) **Sum is divisible by 2 and 3:**

**Soln.:** Let,

C: Sum is divisible by 2 or 3

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

$n(C) = 24$

$P(C) = 24/36 = 2/3$

- 5) 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?

**Soln.:**

Total Balls in Bag = 2Red+3Green+2Blue = 7

Two balls are drawn at random =  ${}^7C_2 = 21$

None of balls drawn is blue =  ${}^5C_2 / {}^7C_2 = 10/21$

- 6) 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 – the probability of having 4 candies = 0.20

**Soln.:**

Expected number of candies = Sum of ( $X * p(x)$ )

= Sum of (Candies count \* Probability)

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

= 3.09

- 7) Calculate Mean, Median, Mode, Variance, Standard Deviation, and Range & comment on 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.

	Point	Score	Weigh
Mean	3.596563	3.21725	17.84875
Median	3.695	3.325	17.71
Mode	3.92	3.44	17.02
Variance	0.285881	0.957379	3.193166
Standard Deviation	0.534678	0.978457	1.786943
Range	2.17	3.91	8.39

**8) Calculate the 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?**

**Soln:**  $E(\text{weight of patient chosen at random}) = E(X)$   
 $= (\text{sum of weights of all patients}) / (\text{total number of patients})$   
 $= (108+110+123+134+135+145+167+187+199)/9$   
 $= 145.33 \text{ pounds}$

**9) Calculate Skewness, Kurtosis & draw inferences from the following data Cars' speed and distance.**

- a) Check Solution.pdf - Q9\_a for calculations.**

For speed column its skewness is between -0.5 to +0.5:

Distribution is normal. i.e., its mode, median, mean are almost equal.

In case of dist column its skewness is  $> 0.5$ ,

Distribution is right skewed. i.e. Dist column have higher values in its data.

Also its mode  $<$  median  $<$  mean

Both speed and dist columns have kurtosis close to 0,

Implies distributions are mesokurtic distributions and are Normal distribution.

- b) Speed (SP) and Weight(WT)**

**Check Solution.pdf - Q9\_b for calculations.**

For SP column its skewness is  $> 0.5$ :

Distribution is right skewed i.e., its mode  $<$  median  $<$  mean.

In case of WT column its skewness is  $< -0.5$ :

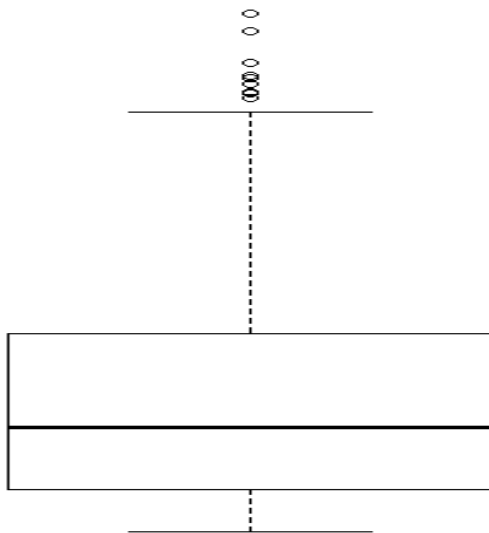
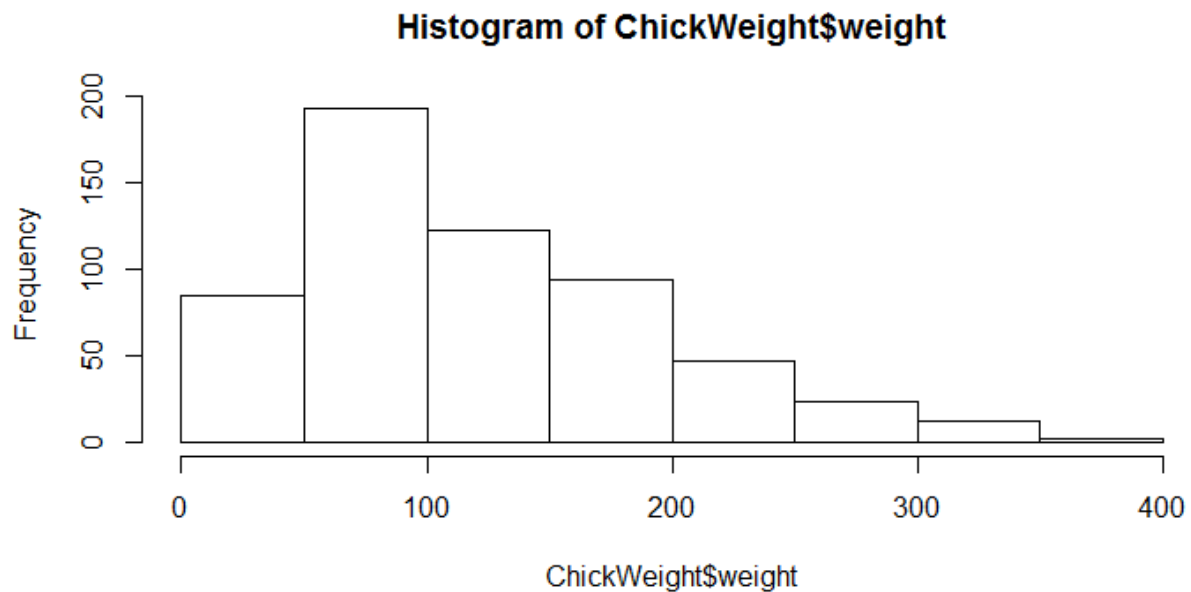
Distribution is left skewed. i.e., its mode  $>$  median  $>$  mean.

For SP column its kurtosis  $< 3$  means it have 0 excess kurtosis:

Distribution is mesokurtic and have normal distribution.

In case of WT its kurtosis  $< 3$  means it have 0 excess kurtosis, Implies distribution is mesokurtic and have normal distribution.

**10) Draw inferences about the following Boxplot & Histogram.**



1 > distribution have mean<median<mode, implies it is right skewed.

2 > lower outliers are present in distribution

3 > IQR is less suggesting more data points are distributed closer to mean.

**11) 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?**

**Soln:**

**Check Solution.pdf – Q11 for calculations.**

94% = (198.738325292158, 201.261674707842)

98% = (198.43943840429978, 201.56056159570022)

96% = (198.62230334813333, 201.37769665186667)

**12) Below are the scores obtained by a student in a test.**

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

**Check Solution.pdf – Q12 for calculations.**

**a) Find mean, median, variance, and standard deviation.**

**Soln:**

Mean = 41

Median = 40.5

Variance = 25.52941

Standard Deviation = 5.052664

**b) What can we say about the student marks?**

**Soln:**

Expected score (mean) for any student selected randomly is 41

From Standard deviation we can infer that

68% students have obtained scores between [36, 46]

95% students have obtained scores between [31, 51]

**13) What is the nature of skewness when the Mean and Median of data are equal?**

**Soln:**

In case when mean is equal to median, the data appears normally distributed. It appears nearly perfectly symmetrical and not absolute symmetrical. It is also called as '**Zero Skewness**'. The left and right side of Bell curve appears as mirror image.

**14) What is the nature of skewness when Mean > Median?**

**Soln:**

In case when mean is greater than median, the tail is inclined towards right side. It is also called as '**Positive Skewness**' where value of skewness is greater than zero. Mean always tends to go towards the most skewed part since skewness influences the mean.

**15) What is the nature of skewness when Median > Mean?**

**Soln:**

In case when mean is greater than median, the tail is inclined towards left side. It is also called as '**Negative Skewness**' where value of skewness is less than zero. Mean always tends to go towards the most skewed part since skewness influences the mean.

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

**Soln:**

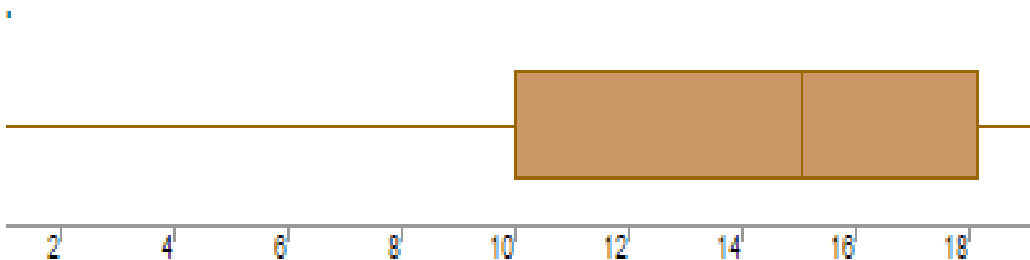
A distribution with a positive kurtosis value indicates that the distribution has heavier tails than the normal distribution. For example, data that follow a t distribution have a positive kurtosis value.

**17) What does a negative kurtosis value indicates for a data?**

**Soln:**

A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



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

**Soln:**

The data on whiskers plot appears asymmetrical across the plane. There might be an outliers influencing the data.

Median of the data is 14.7(approx)

25% of the data lies between 0-10

50% of the data lies between 10-18

25% of the data lies after 18-20 approx.

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

**Soln:**

As the data is inclined towards left, the whiskers plot shows it is negatively skewed distribution as left whisker is greater than right whisker. The mean is less than median in such case

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

**Soln:**

IQR is the Inter Quartile Range.

For left skew =  $Q2 - Q3 < Q2 - Q1$

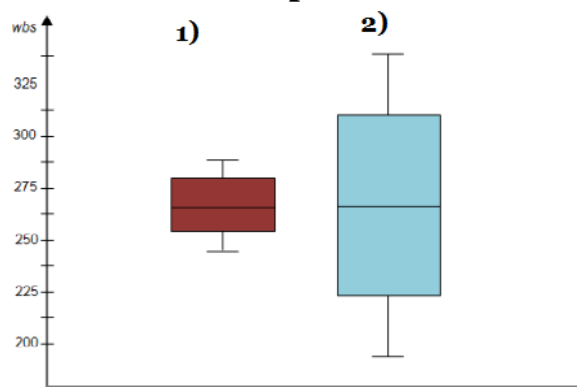
Here  $Q1 = 10$

$Q2 = 14.7$

$Q3 = 18$

$IQR = Q3 - Q1 = 8(\text{approx})$

19) Comment on the below Boxplot visualizations?



**Draw an Inference from the distribution of data for Boxplot 1 w.r.t Boxplot 2:**

**Soln:**

In the above plot comparison, we found Boxplot 1(B1) with Box plot 2(B2), the data is widely spread. Here the main inference is that since the data range varies high in B2 it is hard to make a prediction in B2. The median in the two box plots is equal to 267.5 wbs around. The data spread in both the plots is symmetrical.

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**

**Soln:**

**Check Solution.pdf – Q20 for calculations.**

a.  $P(MPG > 38) = 0.3475939251582705$

b.  $P(MPG < 40) = 0.7293498762151616$



c.  $P(20 < \text{MPG} < 50) = 1.2430968797327613e-05$

**21) Check whether the data follow the Normal Distribution.**

**Soln:**

**Check Solution.pdf – Q21 for calculations.**

**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**

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

**Check Solution.pdf – Q22 for calculations.**

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

**Check Solution.pdf – Q23 for calculations.**

**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**

**Soln:**

**Check Solution.pdf – Q24 for calculations.**

**Sample size = 18 = n**

**Sample mean = 260 days = x**

**Sample standard deviation = s = 90days**

**t-score = -0.4714**

**$P(X \Rightarrow 260) = 0.32167$**