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

Q1) Identify the Data type for the Following:

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

Nominal, Ordinal, Interval, Ratio.

Data	Data Type
Gender	Nominal
High School Class Ranking	Nominal
Celsius Temperature	Interval
Weight	Ratio
Hair Color	Ratio
Socioeconomic Status	Interval
Fahrenheit Temperature	Ratio
Height	Ratio

Type of living accommodation	Ordinal
Level of Agreement	Interval
IQ(Intelligence Scale)	Interval
Sales Figures	Interval
Blood Group	Ratio
Time Of Day	Interval
Time on a Clock with Hands	Interval
Number of Children	Interval
Religious Preference	Ratio
Barometer Pressure	Ratio
SAT Scores	Ratio
Years of Education	Nominal

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

-> P(H,H,T) + P(H,T,H) + P(T,H,H)

-> 1/8 + 1/8 + 1/8

= 3/8 = 0.375

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

   a) Equal to 1 = 0

   b) Less than or equal to 4 = (1,3) (2,2) (3,1), 3/24 -> 1/12

   c) Sum is divisible by 2 and 3 = 6/36 = 1/6

Q5)

-> P(2R, 3G, 2B)

P(5/7,4/6)= 20/42 = 10/21

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

-> 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

->

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?

    Mue = 163.5

    Expected value = 145.33

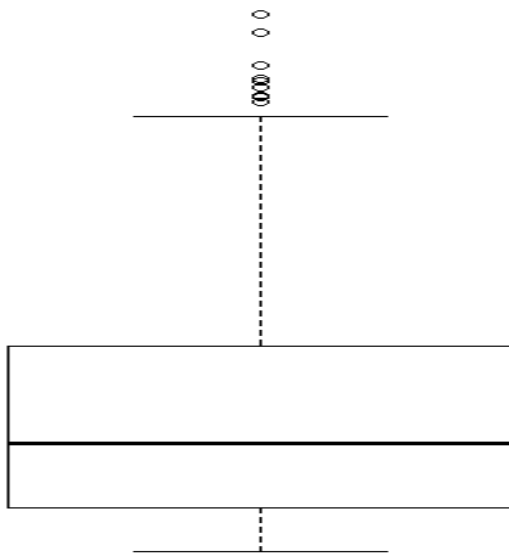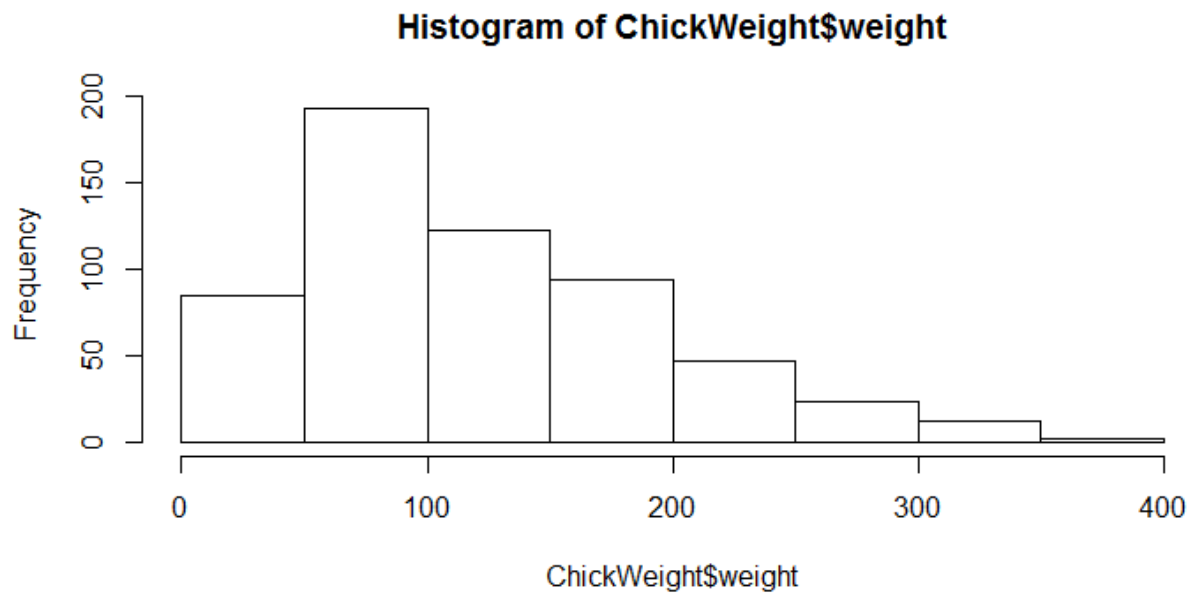**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

   Cars speed and distance

Use Q9_a.csv

SP and Weight(WT)

Use Q9_b.csv

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight

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.

  2) What can we say about the student marks?

Mean = 41 Median =40, Standard Deviation =4.910, Variance = 24.111

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

Symmetrical

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

Right Skewed

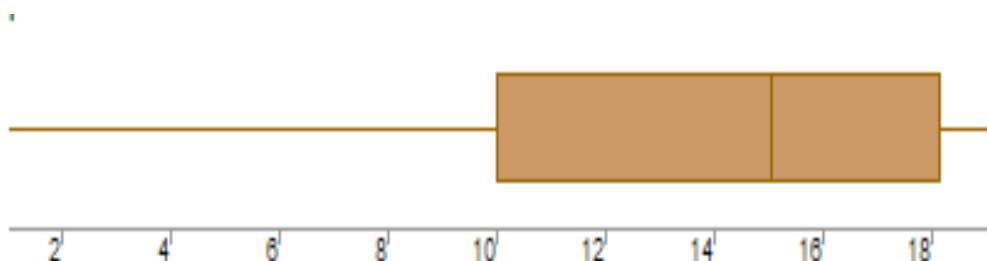Q15) What is the nature of skewness when median > mean?

Left Skewed

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

Data is Normally distributed.

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

Data has lighter tails and flatter peaks then normally distribution

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

What can we say about the distribution of the data?

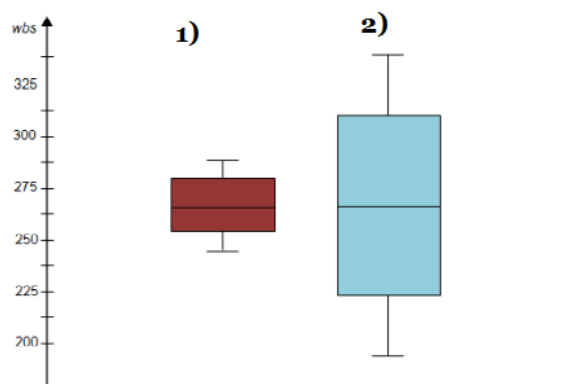-> its is related to age of students in school, 50% students are above 10 years old

And students who is in between 14-16 are approx 40% most probably 15 or somewhere up

What is nature of skewness of the data?

-> left skewed, Median is greater than mean

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

Q19) Comment on the below Boxplot visualizations?

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

->

Level of whisker in box plot 2 is higher, but mean and median are equal for both so distribution is symmetrical

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(MPG>38)
b. P(MPG<40)
c. P (20<MPG<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