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	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	Interval
Time on a Clock with Hands	Interval

Number of Children	Ordinal
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?

Sol: When three coins are tossed the total number of possible combinations are $2^3 = 8$

: P(H H T) + P(H T H) +P(T H H)

=1/8 + 1/8 + 1/8

3/8=0.37

==

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**

Sol: The set of possible outcomes when we roll a die are {1,2,3,4,5,6}
So, when we roll two dice there are 6*6=36 possibilities.

A) There is no outcomes equal to **1. 0/36 (Probability is 0).**

B) Less than 4 : (1,1)(1,2)(2,1)=3 Outcomes, 3/36 i.e 1/12
   Equal to 4: (1,3)(2,2)(3,1) = 3 Outcomes, 3/36 i.e 1/12

C) Sum is divisible by 2 and 3 (6,12)
Favorable outcomes=(1,5) (5,1) (3,3) (2,4) (4,2) (6,6)

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?

Sol: Total number of balls=(2+3+2) =7

Total sample space =7C2

        = 21

Two ball are drawn at random which are not blue ==5C2

            =10

  P(E)=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

Sol: 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**

==

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?

Sol: Mean = 108+110+123+134+135+145+167+187+199/9

= 1308/9

Expected Value of the Weight of that patient = 145.33

==

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

Cars speed and distance

Use Q9_a.csv

Car speed

Sol: Skewness on data car speed : -0.117510

skewness value is negative so it is left skewed. Since magnitude is slightly greater than 0 it is slightly left skewed

Kurtosis on data car speed: -0.508994

Car Distance

Skewness on data car distance: 0.806895 Right skewed (Positive) slight magnitude to right.

Kurtosis on data car distance : 0.405053

SP and Weight(WT)
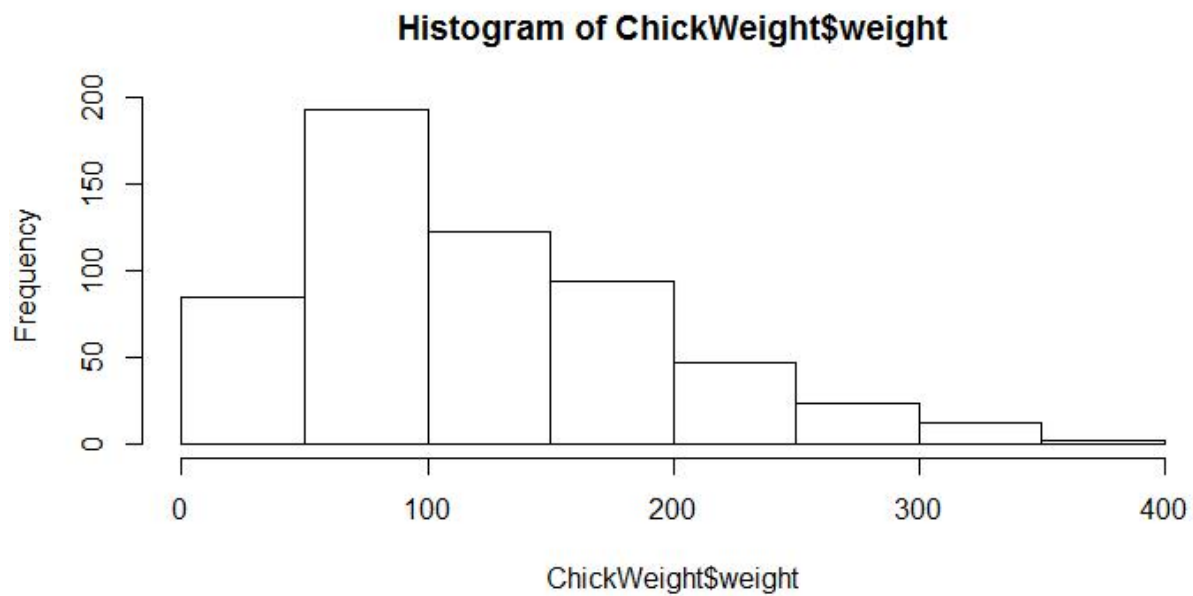
Use Q9_b.csv

Skewness on data car SP : 1.611450

Kurtosis on data car Speed: 2.977329

Skewness on data car Weight : -0.614753
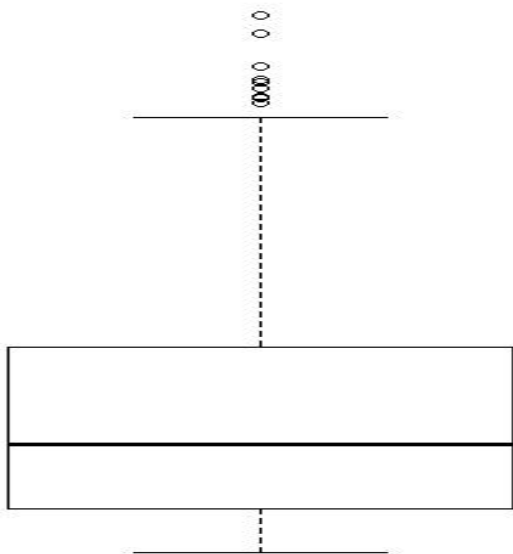
Kurtosis on data car Weight : 0.950291

===

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight

Sol: The most of the Chick weight is between 50 to 100 with frequency 200. And least range of weight is 400. More than 50% chick weight is between 50-150.

Chick weight data is right skewed or positively skewed.

Sol: The data is right skewed.

There are outliers at upper 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%,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?**
   Sol:
 From above data 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 41-42, there are two outlier 49,56.
Mean>Median,hence the data is positively Skewed.

==

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

Sol: Mean=Median=Mode

If the mean is equal to the median as well as the mode, hence the skewness is zero. If the distribution is symmetric, the mean equals the median, and the skewness of the distribution is zero

Null Skewed.

==

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

 Sol: Right Skewnee

==

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

Sol: Left Skewness

==

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

Sol: The data is notmally distributed and kurtosis value is 0.(Thin Peaks)
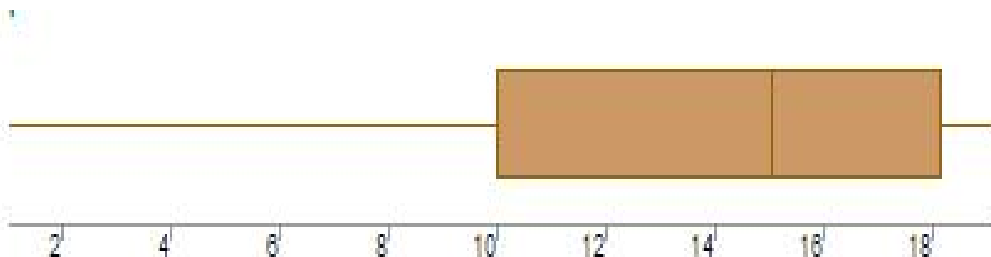
Q17) What does negative kurtosis value indicates for a data ?

Sol: The distribution of the data has lighter tails and a flatter peaks than the normal distribution.

Wider Peaks

===

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.

50% of the people are above 10 y

rs old and remainig are less.

And students who's age is above 15 are approx 40%.

What can we say about the distribution of the data?

Sol:Normal Distribution
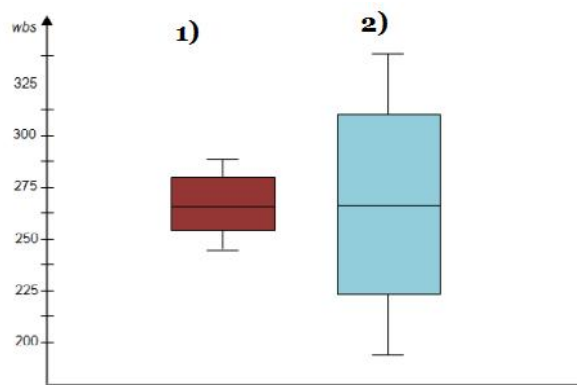
What is nature of skewness of the data?

Negatively Skewed

What will be the IQR of the data (approximately)?
Approximately= 8

==

Q19) Comment on the below Boxplot visualizations?

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

Sol: By observing both the plots whisker's level is high in boxplot 2, mean and

median are equal hence distribution is symetrical

==

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)

Sol: Calculate the probability of MPG of cars for the below cases.

===

Q 21) Check whether the data follows normal distribution
   a) **Check whether the MPG of Cars follows Normal Distribution
      Dataset:Cars.csv**
   From above plot and values we can say that data is fairly symmetrical, i.e
   fairly normally distributed.

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

   mean> median, both the whisker are of same length, median is slightly
   shifted towards left. Data is fairly symetrically display.

===

   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