Q1) Identify the Data type for the Following:

Activity	Data Type
Number of beatings from Wife	Discreate
Results of rolling a dice	Discreate
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	Discreate
Number of kids	Discreate
Number of tickets in Indian railways	Discreate
Number of times married	Discreate
Gender (Male or Female)	Discreate

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

Nominal, Ordinal, Interval, Ratio.

Data	Data Type
Gender	Nominal
High School Class Ranking	Ratio
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?

Heads=H   Tails= T

The possible outcomes:
(H,H,H),(H,H,T),(H,T,H),(T,H,H),(T,T,T),(T,H,T),(T,T,H),(T,H,H)

Therefore number of outcomes=8

So that the number of possible outcomes that gives two heads and one tail is:
(H,H,T),(H,T,H),(T,H,H)

Therefore number of favorable outcomes=3

Thus , the probability of getting two heads and one tail is 3/8

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

If two dice are rolled than the sample space is

(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)

If two dice are tossed then the possibility of outcomes is 36

The minimum sum of possible for two dice is (1,1)= sum of (2)

The maximum sum of possible for two dice is (6,6) = sum of(12)

A)

Sum = 1

The minimum possible sum is(1,1) = 2

Therefore p(1) = 0/36

= 0

b)

sum is less than or equal to 4

The sum is < or = 4 is (1,1),(1,2),(1,3)

(2,1),(2,2)

Hence the number of elements = 5

Hence the probability = 5/36

= 0.13888

c)

sum is divisible by 2 and 3

The sum is divisible by 2 and 3 is (1,5),(3,3),(4,2),(5,1),(6,6)

=5

Probability=5/36

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?

The total number of balls =2+3+2

=7

Two balls are selected at random is 7c2=21

We want to find the number of ways in which no blue ball is selected

2 balls can be selected from red and green balls set is 5c2 = 10

Therefore the probability of none of the balls drawn blue is =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

Expected umber of candies for a randomly selected child

=1*0.015+4*0.20+3*65+5*0.005+6*0.01+2*0.120

=0.015=0.8+1.95+0.025+0.06+0.24

=3.09

Therefore expected number of candies for a randomly selected child = 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 inference

For points:

   Mean=sum of values/no.of values

    =3.5871875

   Median=3.695

   Mode=3.92,3.07 (bimode)

   Range=maximum value-minimum value

    =4.93-2.76

    =2.17

  Variance=422.7907-413.928378125/32

    =8.862321875/32

    =0.276947

  Standard deviation=square of variance

     =0.5262575415

For score:

   Mean=3.21725

   Median =3.325

   Mode=3.44

   Range =3.911

   Variance =0.957379

   Standard deviation=0.9784574

For weight :

Mean=17.84875

Median =17.71

Mode=17.02

Range =8.4

Variance =3.193166

Standard deviation=1.78943

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?

Expected value=sum(probability*value)

Probability of selecting each patient=1/9

Expected value=   1/9(108+110+123+134+135+145+167+187+199))

=145.33

Expected value of weight of patient is =145.33

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

Cars speed and distance

Skewness for speed and distance

speed   -0.117510 it is negative values so that it follows   negatively skewed.
dist   0.806895 It is positive skewness.

kurtosis for speed and distance

speed   -0.508994
dist   0.405053
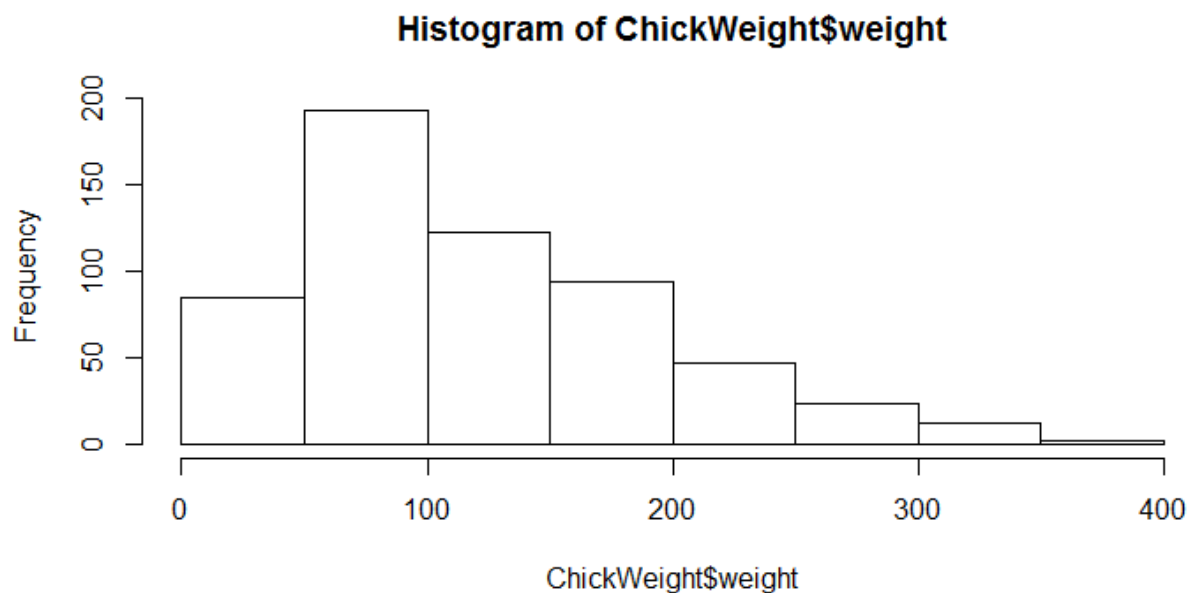
Skewness for sp and weight

SP   1.611450 positively skewed
WT   -0.614753 negatively skewed

kurtosis for sp and weight
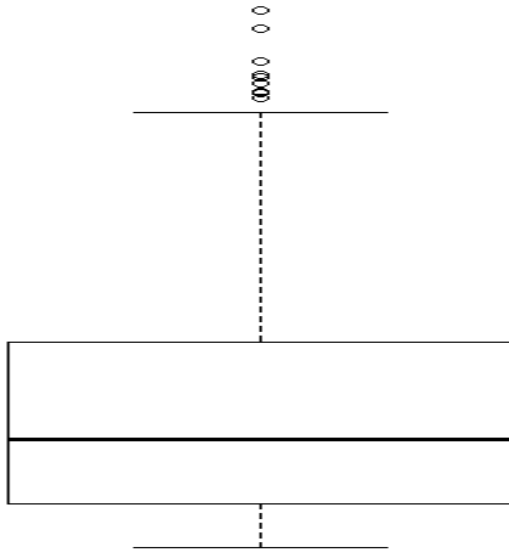
SP   2.977329
WT   0.950291

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight

In this histogram most of the points lies in the range of 50-100 with frequency of 200 and the least points lies in the range of 350-400 with the frequency of less than 10.

We notice long tail towards the right side so it a right skewed.

By the above box plots we notice that outlier of the upper side of boxplot and also we have less points between the Q1 and bottom(minimum) compared to the Q3 and upper(maximum).

It is also clear that the median of the box plot is less than the mean so it is right skewed.

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?

Stadard error:

30/(square root of 2000)=0. 670820393249937

Degrees of freedom=2000-1

=1999

Confidence interval=94%

(1-sigma/2)=(1-0.03)

=0.97

Confidence interval for 94% is 1.882

Confidence interval for 98% is 2.33

Confidence interval for 96% is 2.05

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.

Mean =340+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18

=41

Median =40.5

Variance=25.5294117764705884

Standard deviation=5.05266382858645

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

Symmetric distribution

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

Positively skewness

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

Negatively skewed

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

It indicates that distribution is peaked and possess thick tail. Platykurtic distribution have positive kurtosis value

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

It indicates that distribution is flat and has a thin tail. Leptokutric distribution have negative kurtosis value

Q18) Answer the below questions using the below boxplot visualization.
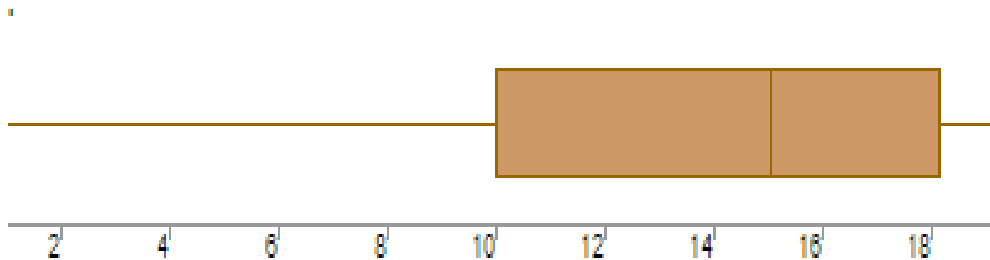
Here we get Q1=10

Q3=18

IQR=Q3-Q1

=18-10

=8

What can we say about the distribution of the data?

 Let us assume above box plot is about the salary of the employees. 50% of the employees are above 10k and remining are less and employee's salary is above 15k are approx 40%
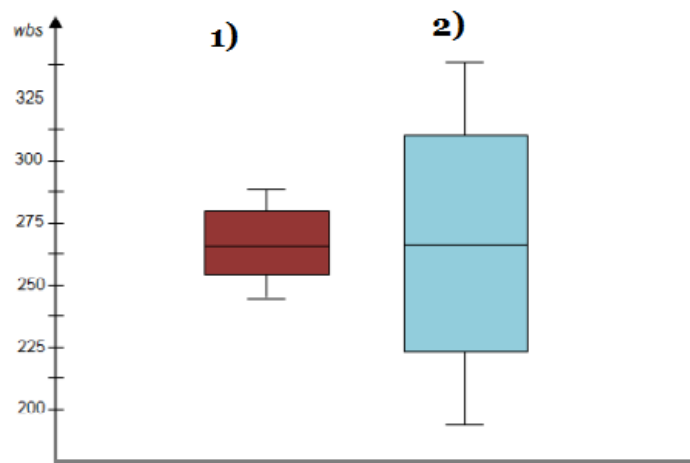
What is nature of skewness of the data?

It is a left skewed

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

IQR of the date is approximately 8

Q19) Comment on the below Boxplot visualizations?

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

1) Maximum value=287.5
   Minimum value=237.5

Q3=275

Q1=250

Median=262.5

2) Maximum value=287.5

Minimum value=200

Q3=312.5

Q1=225

Median =262.5

By observing the above data it is clear that mean and median is same so that it is a symmetric distribution.

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)

a. There are 33 observations in mpg which are greater than 38

b. 61 observations in MPG which are lesser than 40

c. 70 observations in MPG which are in between 20 and 50

Q 21) Check whether the data follows normal distribution
a) Check whether the MPG of Cars follows Normal Distribution
Dataset: Cars.csv

stat=0.889, p=0.000

Data does not follow Normal Distribution

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

stat=0.926, p=0.000
Data does not follow Normal Distribution

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

Z score of 90% confidence interval is 1.65
Z score of 94% confidence interval is 1.55
Z score of 60% confidence interval is 0.85

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

T score of 95% confidence interval is 1.96
T score of 96% confidence interval is 2.5
T score of 99% confidence interval is 2.47

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

    x=mean of the sample of bulbs=260

    u=population mean =270

    s=standard deviation of the sample=90

    n=number of items in the sample=18

    $t = x-u/s_{/root\ n}$

    $t = 260-270/90_{/root2}$

    t=-0.471

 Degrees of freedom d_f = n-1

            =18-1

        =17

      Failed test t-0.471 and d_f =17

      For p value = 0.3372(from normal distribution table)

    Therefore required probability = 0.3372

The probability that t<-0.471 with 17 degrees of freedom

The probability of the bulbs lasting less than 260days on average of 0.3218 assuming the mean of the bulbs is 300days.