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

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	Discrete data- Nominal
High School Class Ranking	Discrete data- Nominal
Celsius Temperature	Continuous- Interval
Weight	Continuous- Ratio
Hair Color	Discrete data- Ratio
Socioeconomic Status	Continuous- Interval
Fahrenheit Temperature	Continuous – Ratio
Height	Continuous- Ratio
Type of living accommodation	Discrete- Ordinal
Level of Agreement	Discrete- Interval
IQ(Intelligence Scale)	Discrete- Interval
Sales Figures	Discrete- Interval
Blood Group	Discrete- Ratio
Time Of Day	Continuous – Interval
Time on a Clock with Hands	Continuous- Interval

Number of Children	Discrete- Interval
Religious Preference	Discrete- Ratio
Barometer Pressure	
SAT Scores	Continuous- Ratio
Years of Education	Discrete- Nominal

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

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

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

= 3/8

Or (Alternative approach)

Binomial distribution.

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: a) There is no outcomes which corresponds sum is equal to one. i.e. 0/36. Probability is 0.
b) (1,3) (2,2) (3,1) = 3 outcomes, 3/36 i.e. 1/12
c) 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: 10/21

P (2R, 3G, 2B)

P (5/7, 4/6) = 20/42 I.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: 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.

Points	Score	Weigh>
3.9	2.62	16.46
3.9	2.875	17.02
3.85	2.32	18.61
3.08	3.215	19.44
3.15	3.44	17.02
2.76	3.46	20.22
3.21	3.57	15.84
3.69	3.19	20
3.92	3.15	22.9
3.92	3.44	18.3
3.92	3.44	18.9
3.07	4.07	17.4
3.07	3.73	17.6
3.07	3.78	18
2.93	5.25	17.98
3	5.424	17.82
3.23	5.345	17.42
4.08	2.2	19.47
4.93	1.615	18.52
4.22	1.835	19.9
3.7	2.465	20.01
2.76	3.52	16.87
3.15	3.435	17.3
3.73	3.84	15.41
3.08	3.845	17.05
4.08	1.935	18.9
4.43	2.14	16.7
3.77	1.513	16.9
4.22	3.17	14.5
3.62	2.77	15.5
3.54	3.57	14.6
4.11	2.78	18.6

Sol: **Points**: Mean =3.596563, Median= 3.695, Mode= “numeric”,

Variance= 0.2858814, Standard deviation= 0.5346787.

Score: Mean= 3.21725, Median= 3.325, Mode= “numeric”,

Variance= 0.957379, Standard deviation= 0.9784574

Note: Mean value are closer for both ‘Point’ and ‘Score’.

Weight: Mean= 17.84875, Median= 17.71, Mode= “numeric”,

Variance= 3.193166, Standard deviation= 1.786943

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= 1308

` Mue= 163.5

Expected value= ∑[x.p(x)]

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

Cars speed and distance

speed	dist
4	2
4	10
7	4
7	22
8	16
9	10
10	18
10	26
10	34
11	17
11	28
12	14
12	20
12	24
12	28
13	26
13	34
13	34
13	46
14	26
14	36
14	60
14	80
15	20
15	26
15	54
16	32

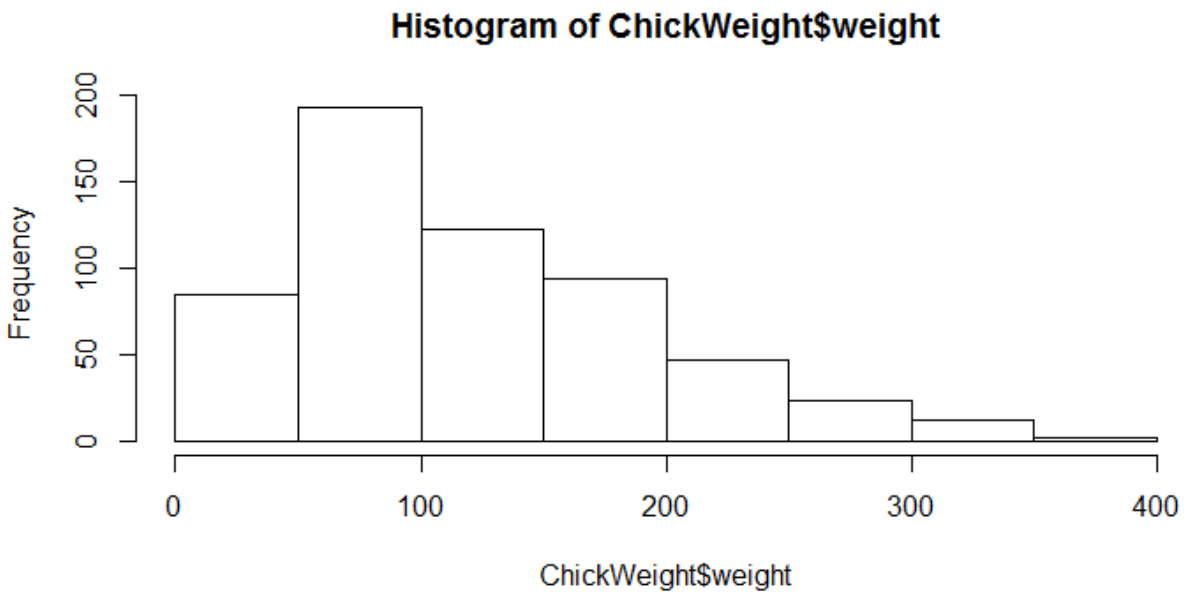
SP and Weight(WT)

SP	WT
104.1854	28.76206
105.4613	30.46683
105.4613	30.1936
113.4613	30.63211
104.4613	29.88915
113.1854	29.59177
105.4613	30.30848
102.5985	15.84776
102.5985	16.35948
115.6452	30.92015
111.1854	29.36334
117.5985	15.75353
122.1051	32.81359
111.1854	29.37844
108.1854	29.34728
111.1854	29.60453
114.3693	29.53578
117.5985	16.19412
114.3693	29.92939
118.4729	33.51697
119.1051	32.32465
110.8408	34.90821
120.289	32.67583
113.8291	31.83712
119.1854	28.78173
114.5985	16.04317
120.7605	38.06282
119.1051	32.83507
99.56491	34.48321
121.8408	35.54936
113.4846	37.04235
112.289	33.23436
119.9211	31.38004
121.3926	37.57329

Sol: Skewness for speed= -0.1139548, skewness value is negative so it is left skewed. Since magnitude is slightly greater than 0 it is slightly left skewed

And for distance= 0.7824835, right skewed (Positive) slight magnitude to right.

Q10) Draw inferences about the following boxplot & histogram

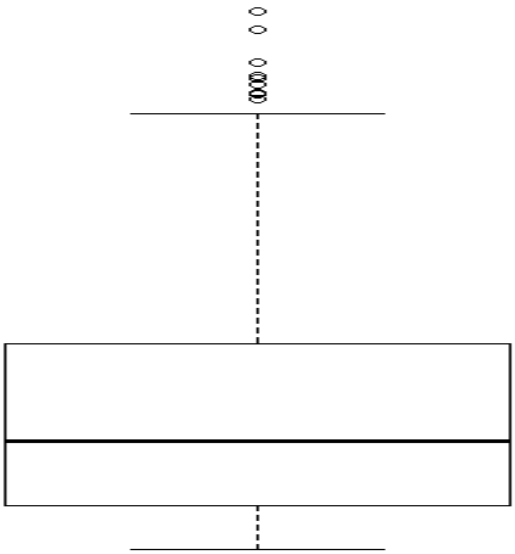


Sol: The most of the data points are concerated in the range 50-100 with frequency 200.

And least range of weight is 400 somewere around 0-10.

So the expected value the above distribution is 75.

Skewness- we can notice a long tail towards right so it is heavily right skewed.



Sol: Medican is less than mean right skewed and we have outlier on the upper side of 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 ?

Sol: X+/-(Z1- α. σ/sqrt(n)

Degrees of freedom= 2000-1= 1999

Confidence interval= 94%

(1- σ/2)= 1-0.03) =0.97

for confidene interval for 94% is 1.882

Confidence interval for 98%= 2.33

Confidence interval for 96% = 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.
2) What can we say about the student marks?

Mean= 41, Median= 40, variance= 24.111, Standard deviation= 4.910

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

Sol: Symetrical

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

Right skewed

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

Left Skewed

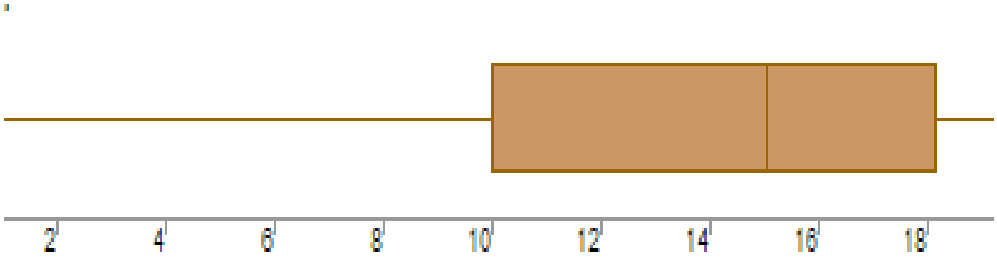
Q16) What does positive kurtosis value indicates for a data ?

The data is notmally distributed and kurtosis value is 0.

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

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

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



What can we say about the distribution of the data?

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

50% of the people are above 10 yrs old and remainig are less.

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

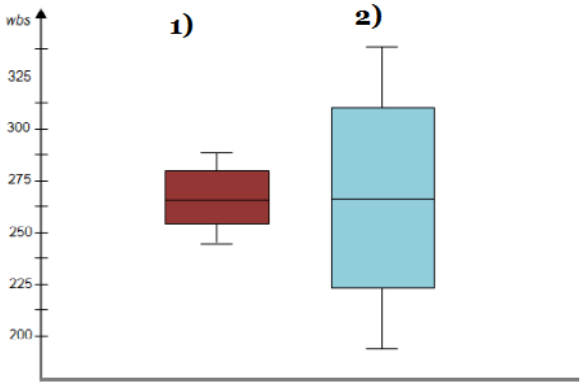
What is nature of skewness of the data?

Sol: Left skewed, median is greater than mean.

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 FILTERING

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: By using filter command and installing the dplyr package into the 'R'.

a) There are 33 observations in MPG which are greater than 38

b)61 observations in MPG which are lesser than 40

c)

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

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

For 95%= 1.96

For 96%= 2.5

For 99% = 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