BASIC STATISTICS -1

Q1) Identify the Data type for the Following:

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)	Continuous

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

BASIC STATISTICS -1

SAT Scores	Interval
Years of Education	Ratio

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

Ans: 3/8=0.375

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

a) Equal to 1
   Ans: 0
b) Less than or equal to 4
   Ans: 6/36 = 0.17
c) Sum is divisible by 2 and 3
   Ans: 6/36 = 0.17

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?
   Ans: (5C2/7C2)=10/21=0.476

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

Ans: 3.09

BASIC STATISTICS -1

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

Points - Right Skew
Score- Right Skew
Weigh- Right Skew

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?

Ans: The Expected Value of the Weight of that patient is 145.33

Q9) Calculate Skewness, Kurtosis & draw inferences on the following data
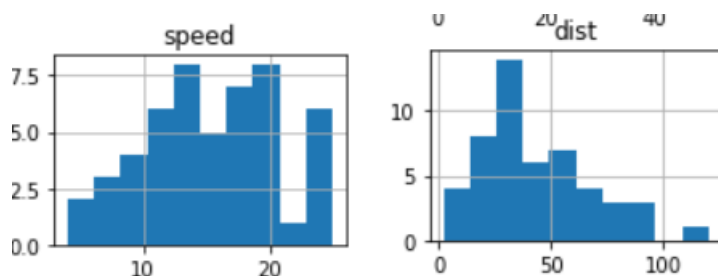
Cars speed and distance

Use Q9_a.csv

Skewness of Cars speed: -0.117510

Skewness of Distance: 0.806895

Kurtosis of cars: -0.508994

Kurtosis of Distance: 0.405053
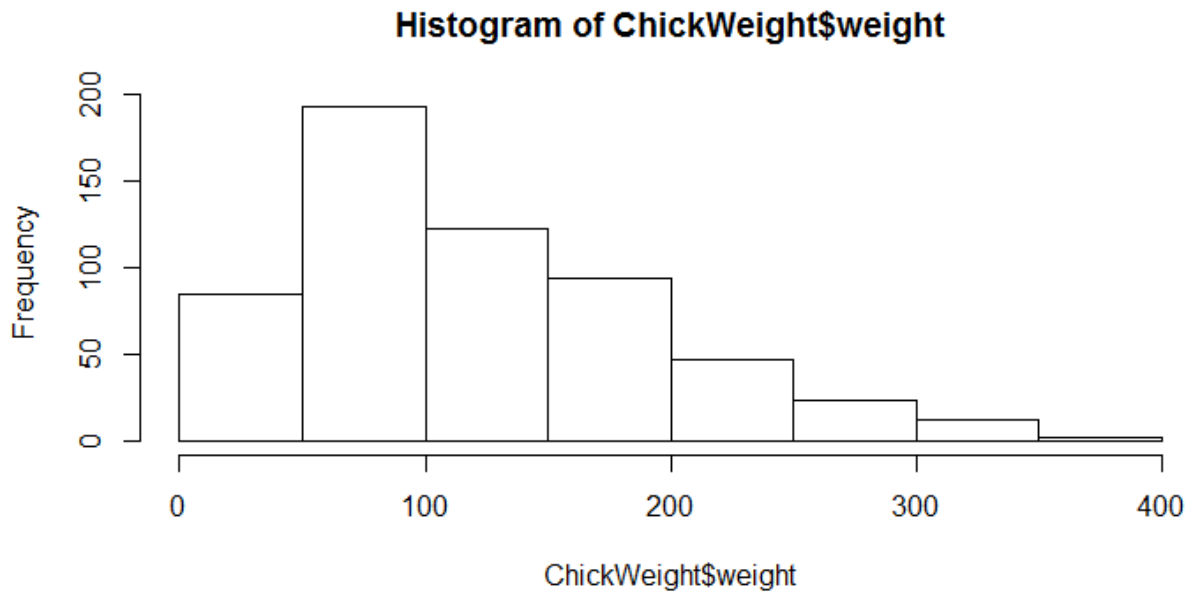
BASIC STATISTICS -1

SP and Weight(WT)

Use Q9_b.csv

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight

ChickWeight$weight

Ans: Positive skew

BASIC STATISTICS -1
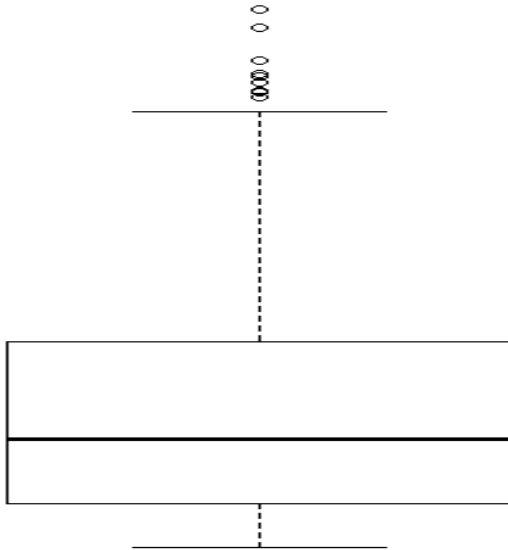
Q11) Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Ans: Avg. weight of Adult in Mexico with 94% CI is (198.738, 201.261)

Avg. weight of Adult in Mexico with 96% CI is (198.622, 201.377)

Avg. weight of Adult in Mexico with 98% CI is (198.439, 201.560)

Q12) Below are the scores obtained by a student in tests

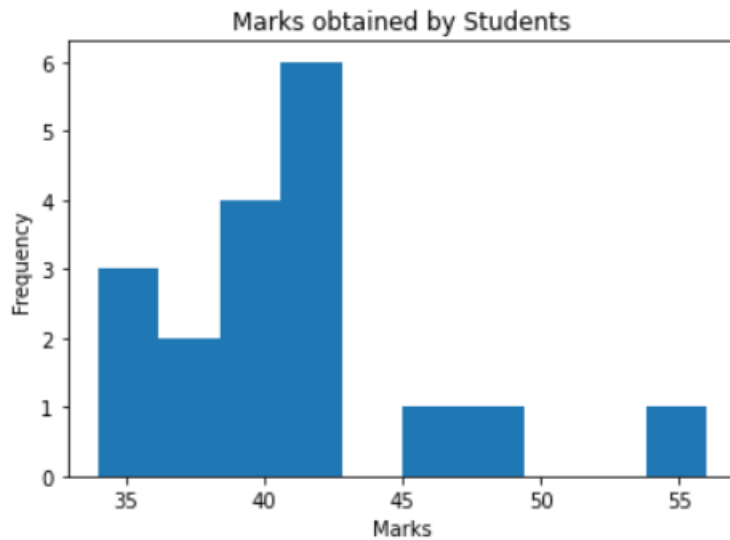**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1) Find mean, median, variance, standard deviation.
   Ans: Mean= 41, Median= 40, variance= 24.111, Standard deviation= 4.910

BASIC STATISTICS -1

2) What can we say about the student marks?

Ans: From the below histogram, we can able to see that most of students got marks between 35 to 41.

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

Ans: No skewness

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

Ans: Right skewness

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

Ans: Left skewness

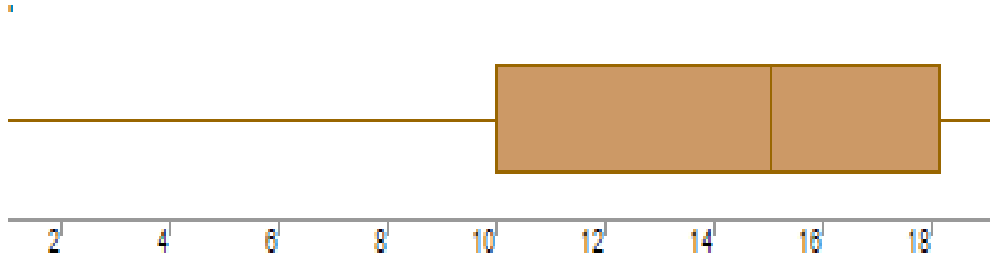Q16) What does positive kurtosis value indicates for a data ?

Ans: Sharp peak and it has more outliers

BASIC STATISTICS -1

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

Ans: Flat peak and it has lack of outliers

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

What can we say about the distribution of the data?

Ans: Normal distribution and it has low kurtosis because it has light tails and lack of outliers
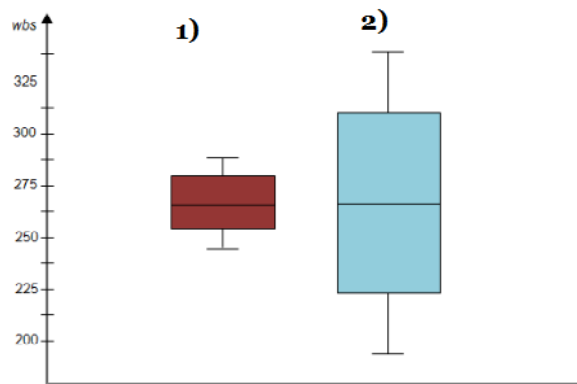
What is nature of skewness of the data?

Ans: Left skewed, median is greater than mean

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

Ans: Approximately= 8

BASIC STATISTICS -1

Q19) Comment on the below Boxplot visualizations?

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

Ans: From these above both the boxplots are following normal distribution and having same median values but it have different ranges

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) = 0.3476

b. P(MPG<40) = 0.7293

c. P (20<MPG<50) = 0.8988

BASIC STATISTICS -1

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

Ans: It follows normal distribution because

- mean is approximately equal to median.
- It is symmetrical

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

Ans: Waist Circumference(Waist) follows normal distribution because

- mean is approximately equal to median.
- It is symmetrical

Ans: Adipose tissue(AT)does not follows normal distribution because

- mean is not equal to median.
- It is not symmetrical

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval
Ans: 90% CI - 1.2815
       94% CI - 1.5547
       60% CI - 0.2533

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

Ans: 95% CI - 2.0638
       96% CI – 2.1715
       99% CI - 2.7969

BASIC STATISTICS -1

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

Ans: 0.322