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	Nominal
Level of Agreement	Interval
IQ(Intelligence Scale)	Interval
Sales Figures	Ratio
Blood Group	Nominal
Time Of Day	Interval
Time on a Clock with Hands	Interval

Number of Children	Ratio
Religious Preference	Nominal
Barometer Pressure	Ratio
SAT Scores	Interval
Years of Education	Ordinal

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

ANS-The outcomes are HHT, THH, and HTH hence 3, the total outcomes are 8 hence the probability 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

ANS-the total outcomes are 36 hence

a) No possible outcomes

b) Number of outcomes are (1,1),(1,2),(1,3),(2,2),(3,1),(2,1) hence we get 6, probability is 6/36=1/6

c) Number of outcomes are (1,5),(2,4),(5,1),(4,2),(3,3),(6,6) hence we get 6,probability is 6/3=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?

ANS-Total balls are 2+3+2=7 , the probability of getting non blue balls are 5 balls i.e., green and red balls, so probability is 5/7 and if two balls are drawn means 2*5/7=10/14

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-The total probability is
1*0.015+4*0.20+3*0.65+5*0.005+6*0.01+2*0.120=3.09

Hence expected number is 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

ANS-Points

Mean-3.596563

Median-3.695

Mode-3.92

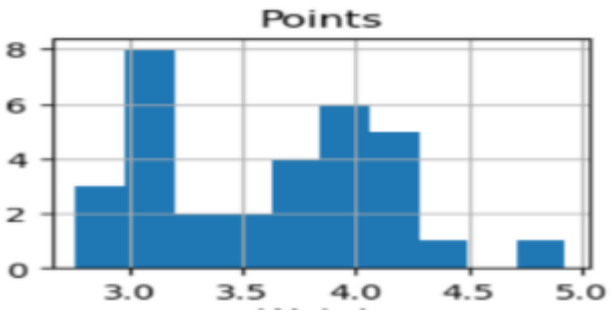
Variance-0.285881

Standard Deviation-0.534679

Range-max-min=4.93-2.76=2.17

Comments- it has a negative skew cause average value is less than median value

Drawing-



-Score

Mean-102.952

Median-3.325

Mode-3.44

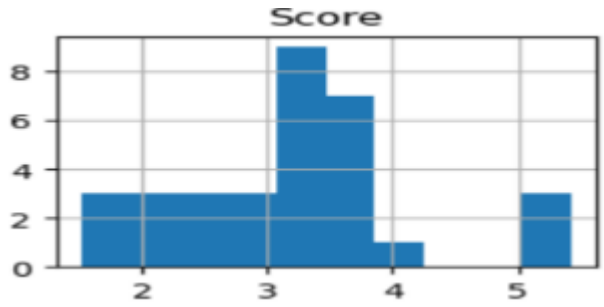
Variance-0.957379

Standard Deviation-0.978457

Range-max-min=-5.424-1.513=3.911

Comments- it has a negative skew cause average value is less than median value

Drawing-



-Weight

Mean-17.84875

Median-17.71

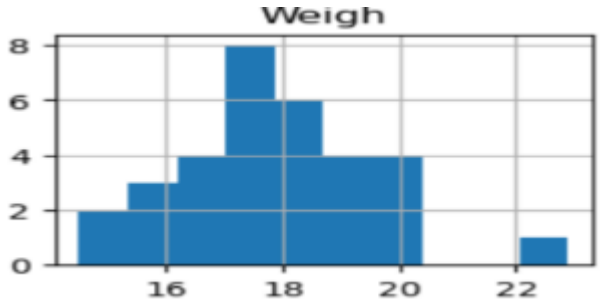
Mode-17.01

Variance-3.193166

Standard Deviation-1.786943

Range-max-min=22.9-14.5=8.4

Comments- it has a positive skew cause average value is more than median value



References-

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-108+110+123+134+135+145+167+187+199=1308

the expected weight can be calculated as average of the total weight =1308/9=145.3

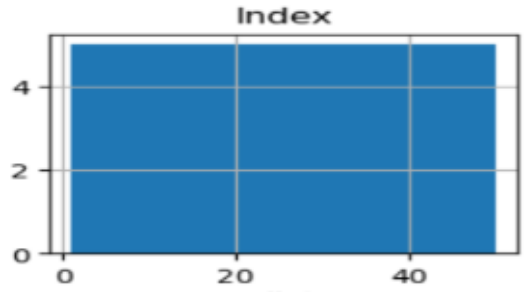
Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

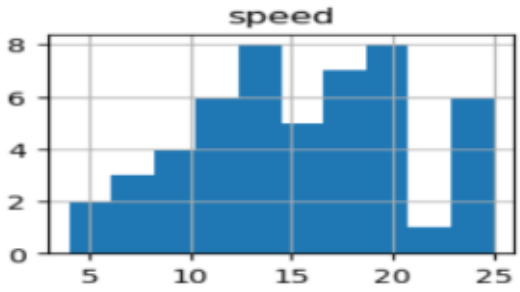
Cars speed and distance

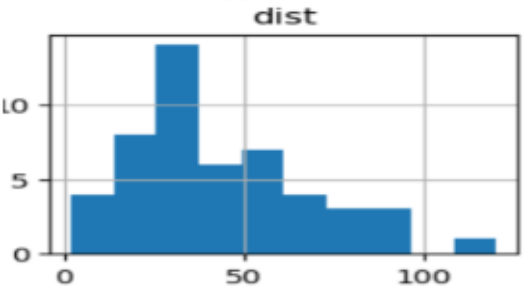
Use Q9_a.csv

ANS-skewness=[0.11395477-0.78248352]

kurtosis=[-1.20096038-0.24801866]





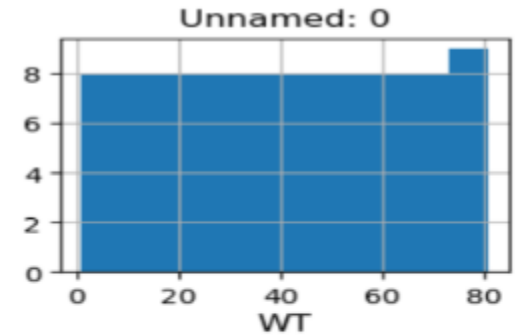


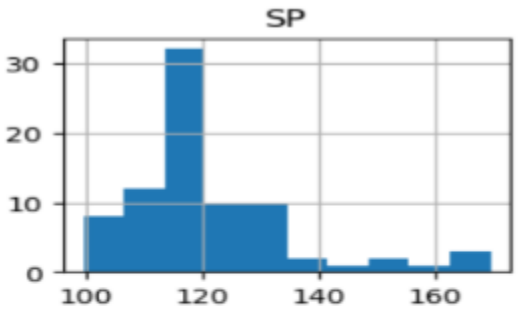
SP and Weight(WT)

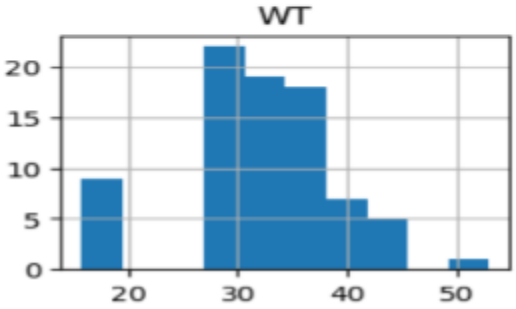
Use Q9_b.csv

ANS-sknewness=[1.58145363- -0.60330993]

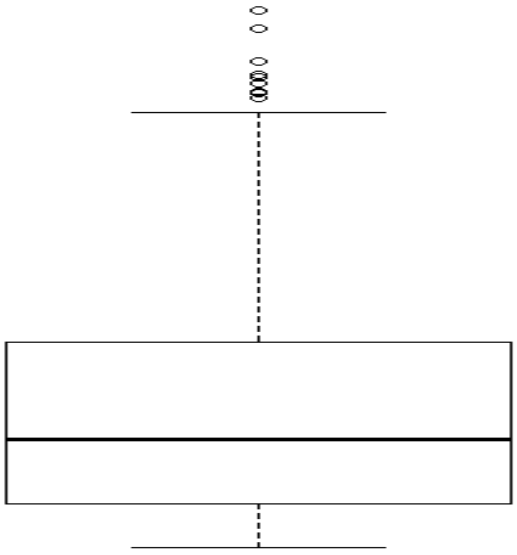
kurtosis=[-1.20036585-0.81946588]







Q10) Draw inferences about the following boxplot & histogram



ANS- The first plot is right-skewed, with most values on the left side. In the 300-400 range, values drop. This suggests a focus on lower values, potentially with outliers on the right.

The second plot is right-skewed, evident as the median is above the box plot. This indicates a focus on lower values with a tail extending to higher values.

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-The confidence interval of 94% is (197.58,202.42) pounds

-The confidence interval of 98% is (197.22,202.78) pounds

-The confidence interval of 96% is (197.66,202.34) pounds

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?

ANS-1)mean=41

median=40.5

Variance=24.11

Standard deviation=4.910306620885412

2)The performance of the students are average but compared to highest score 56 and lowest as 34, but others performed average of 41 which is fine.

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

ANS-If the mean median and mode are equal in a dataset the skewness is likely to be zero indicating a symmetrical distribution . This means the data is evenly distributed around the central point , there is no tail on either side.

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

ANS-The nature of skewness is right skewed as mean>median cause in such case the tail of mean is longer compared to left cause mean indicates the larger value compared to median pulling larger values to mean.

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

ANS-The nature of skewness is left skewed as median>mean cause in such case the tail of median is shorter compared to right cause median probably has less value compared to mean pulling larger values to mean.

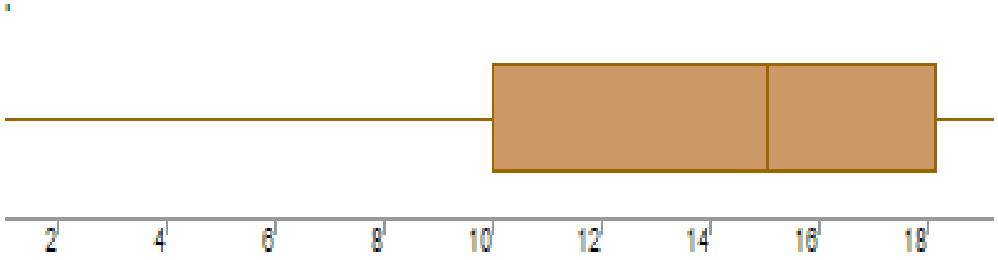
Q16) What does positive kurtosis value indicates for a data ?

ANS-A positive kurtosis value indicates that a dataset has heavier tails and sharper peaks than normal distribution. In other words the distribution has fatter tails and more extreme values than normal distribution.

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

ANS-A negative kurtosis value indicates that a dataset has lighter tails and non-sharp peaks than normal distribution. In other words the distribution has thinner tails and no more extreme values than normal distribution.

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



What can we say about the distribution of the data?

ANS-The median is towards the left side of the line , towards the higher side hence its not normal distribution.

What is nature of skewness of the data?

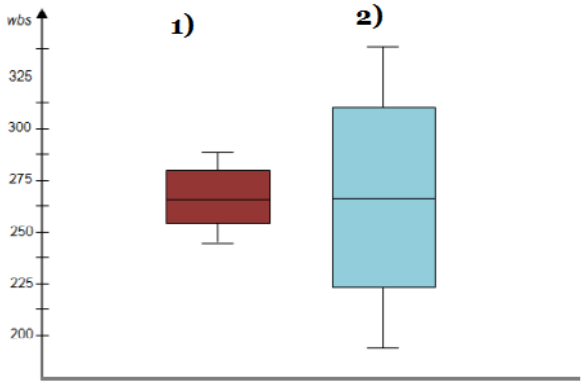
ANS-It is negatively skewed

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

=Q3-Q1

=18-10=8

Q19) Comment on the below Boxplot visualizations?



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

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