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)	Discrete

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 probability og getting exactly two heads and one tail when three coins are tossed is 3/8, 0.375(37.5%)

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: a) When two dice are rolled, the minimum possible sum is 2 (rolling a 1 on both dice). So, it's impossible to get a sum of 1 with two dice. Therefore, the probability of getting a sum equal to 1 is 0

b) (1,1),(1,2),(1,3),(2,1),(2,2),(3,1) Therefore these are the only possibilities to get less than or equal to 4. i.e.,6 so the probability is

PROBABILITY=favourable outcomes/Total outcomes
=6/36=1/6=0.166

0.166 is the probabiliy to get the less than or equal to 4 when two dice are rolled.

c) The only numbers which divisible by both 2 and 3 is 6 and 12.
(1,5),(2,4),(3,3),(4,2),(5,1)--------> are the only outcomes for sum 6
(6,6) is the only outcome to get sum 12 so there are total 6 favourable outcomes here .

PROBABILITY= favourable outcomes/Total outcomes
=6/36=1/6=0.166

0.166 is the probabiliy to get the sum is divisible by 2 and 3 when two dice are rolled.

= 6/36

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?
Ans: 2 red+ 3 green+2 blue balls = Total 7 balls
so first let us find out the favourable outcomes for the blue balls
Therefore probability of getting 1 blue ball= 2/7
so the probability fo getting non blue balls =1-2/7 or 5/7
if non blue ball is drawn the remaining balls are then Total=6 balls
now again calculate the probability of getting the blue ball
Probability of getting 1 blue ball=2/6
so the probability fo getting non blue balls =1-2/6 or 4/6
Therefore, the probability of drawing 2 balls without drawing any blue ball is
(5/7) * (4/6) = 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

Ans: To calculate the expected number of candidates for a randomly selected child,then we need to calculate the expected value,

Expected value = sigma(X * P(X))

Expected value = (1*0.015)+(4*0.20)+(3*0.65)+(5*0.005)+(6*0.01)+(2*0.120)

= 0.015+0.80+1.95+0.025+0.06+0.24

= 3.095

So,the expected number of candies for a randomly selected child is 3.095 on average.

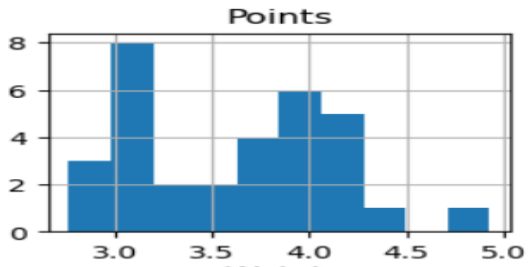
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given data set

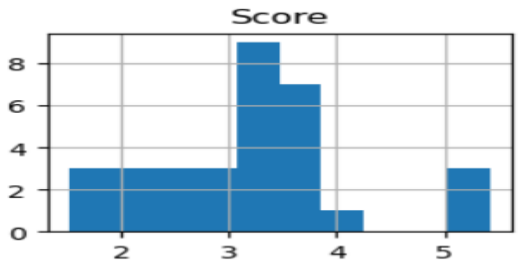
for points

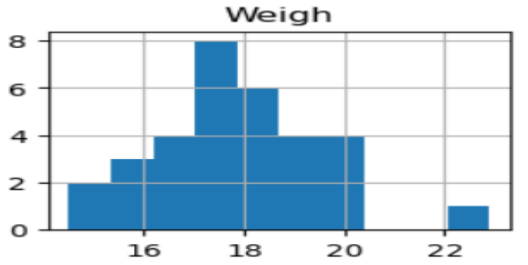
mean - 3.5965625

median - 3.6950000000000003
var - 0.28588135080645166
std - 0.5346787360709716
#range=max-min
range=max1-min
range= 12.17
The relatively small variance of 0.286 and the standard deviation of approximately 0.535 suggest that the "Points" data is not highly spread out from the mean.
The mean (average) "Points" value is approximately 3.60, while the median (middle value) is approximately 3.70. This suggests that the central tendency of the data is slightly lower than the median, indicating a possible slight negative skew in the distribution.
for score:
mean -3.2172500000000004
median -3.325
var -0.9573789677419356
std -0.9784574429896967
#range=max-min
range=max1-min1
range=3.9110000000000005
The mean "Score" value is approximately 3.22, while the median is approximately 3.33. This suggests that the central tendency of the data is slightly lower than the median, indicating a possible slight negative skew in the distribution
For Weigh:

mean -17.848750000000003
median -17.71
var - 3.193166129032258
std -1.7869432360968431
#range=max-min
range=max1-min1
range = 8.399999999999999
The mean "Weigh" value is approximately 17.85, while the median is approximately 17.71. This suggests that the central tendency of the data is slightly higher than the median, indicating a possible slight positive skew in the distribution.







Q8) Calculate Expected Value for the problem below

a) The weights (X) of pa-tients 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: To calculate the expected value we need to calculate the mean of the patients,where

X = {108,110,123,134,135,145,167,187,199}

E(X) = (108+110+123+134+135+145+167+187+199)\9

= 145.5

Thus, the expected value of the weight of a randomly chosen patient from the clinic is 145.5 pounds.

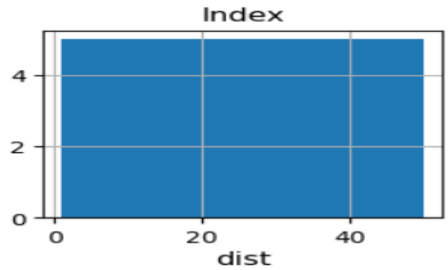
Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

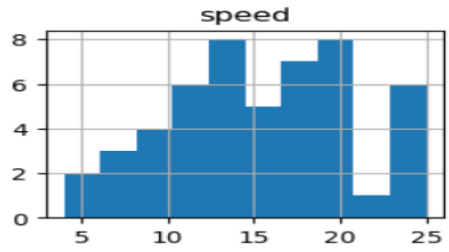
Cars speed and distance

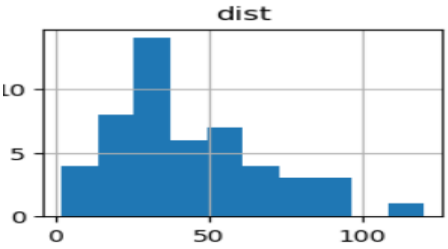
Use Q9_a.csv

Ans: Skewness of the data set = [-0.11395477 - 0.78248352]

Kurtosis of the data set = [-1.20096038 - 0.24801866]





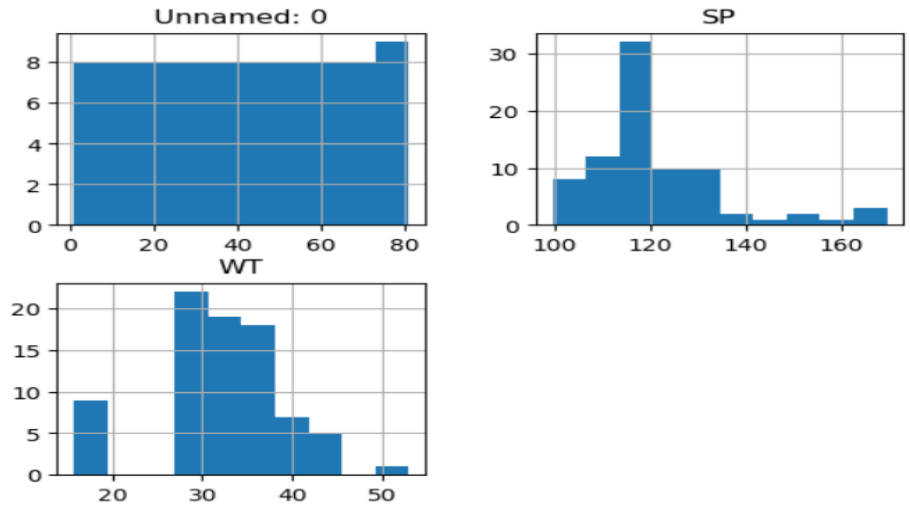


SP and Weight(WT)

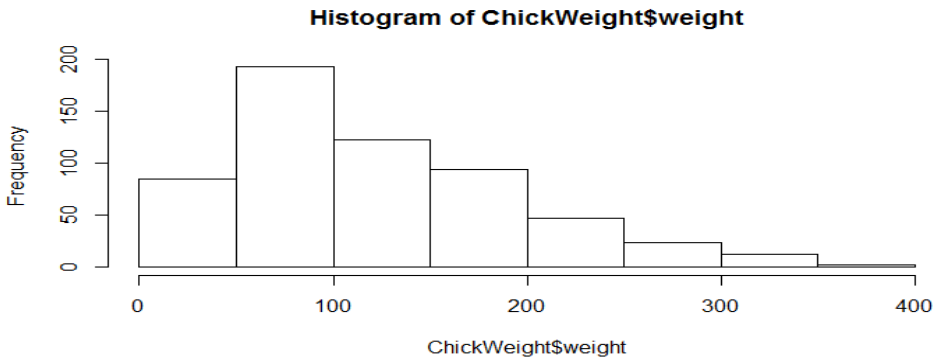
Use Q9_b.csv

Skewness of the data set =[1.58145368 - -0.60330993]

Kurtosis of the data set = [-1.20036585 - 0.81946588]



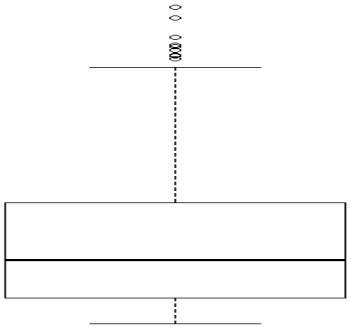
Q10) Draw inferences about the following boxplot & histogram



Ans: Here we can see that the major Chick weights fall in the category of 50-100g(measures in X) as the maximum which is 200.The minimum weights have a frequency if less than or equal to 5.

The plot is Right skewed which show that there is lasser concentration of chick weights in the 300-400grams category.

The Expected value should be above 46.65



Ans: Median is less than mean,right skewed and we have the Outlier on the upperside of box plot and there is less data points between Q1 and lower points.

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 94% confidence interval is (197.58,202.42) pounds,

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

The 96% confidence interval 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.

mean--->41.0

Median-->40.5

variance--> 24.11111111111111

standard deviation--->4.910306620885412

The mean score is approximately 41.5, and the median is 40.5. mean and median are close, suggesting that the scores are relatively symmetrically distributed around the center.
There are no extreme outliers in the dataset. All the scores fall within a reasonable range.The highest score is 56, and the lowest score is 34, which are not extremely distant from the central tendency.
2.What can we say about the student marks?
The variance is approximately 25.5, and the standard deviation is approximately 5.05.The relatively low variance and standard deviation indicate that there is not a significant amount of variability in the scores.The scores are relatively tightly grouped around the mean, suggesting that the student's performance is consistent across the tests.This suggests that the student's performance is consistently good, with occasional scores slightly above or below the central value.

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

Ans: When the mean and median of a dataset are equal, it indicates that the data is symmetrically distributed.The skewness in this case is zero because there is no skew in the data. The distribution is perfectly balanced on both sides of the mean.

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

Ans: When the mean is greater than the median, it suggests that the data is right-skewed (positively skewed).In a right-skewed distribution, the tail on the right side is longer.

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

Ans: When the median is greater than the mean, it suggests that the data is left-skewed (negatively skewed).In a left-skewed distribution, the tail on the left side is longer

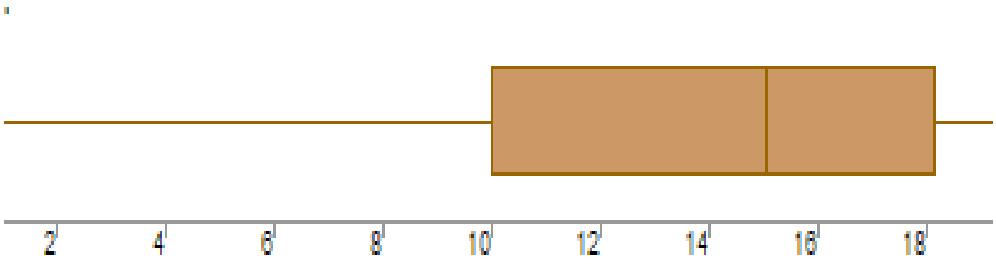
Q16) What does positive kurtosis value indicates for a data ?

Ans: :A positive kurtosis value indicates that the data has fatter tails and is more peaked at the center compared to a normal distribution.Such a distribution is called leptokurtic.

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

Ans: A negative kurtosis value indicates that the data has thinner tails and is flatter at the center compared to a normal distribution.Such a distribution is called platykurtic

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



What can we say about the distribution of the data?

Ans: The above Box plot is not normally distributed the median is towards the higher value.

What is nature of skewness of the data?

Ans: The nature of skewness is negative skewness.

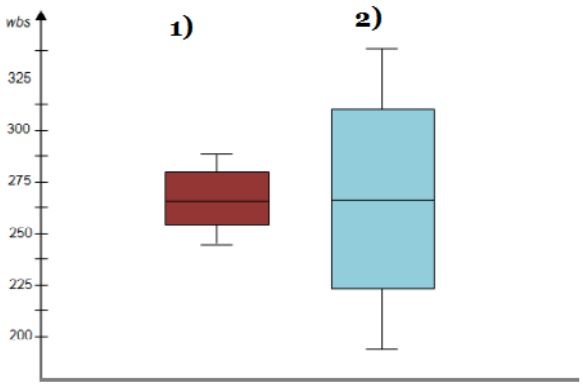
What will be the IQR of the data (approximately)?

Ans: The Inter Quantile Range = Q3 - Q1

= 18-10

= 8

Q19) Comment on the below Boxplot visualizations?



Ans: Here there is a representation of 2 box plots in which box plot 2 is highly distributed across the plane and box plot 1 is slightly less distributed.(variances)

Whiskers in these diagrams also show this 100% of the data is spread across values from 350 in 2 whereas its spread in range250-290 approx in 1.

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

Ans: Here when we compare box plot 1 with box plot 2 we can say that the data in box plot 1 is widely spread.Here the main inference is that since the data range varies high in box plot 2 it is hard to make a prediction in box plot 2.The median in the 2 box plots are equal and the data spread in both of them are symmetrical.

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)

Ans: 0.330

b. P(MPG<40)

Ans: 0.729

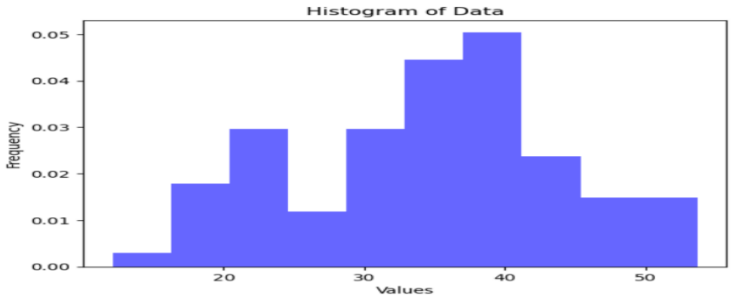
c. P (20<MPG<50)

Ans: 0.898

Q 21) Check whether the data follows normal distribution

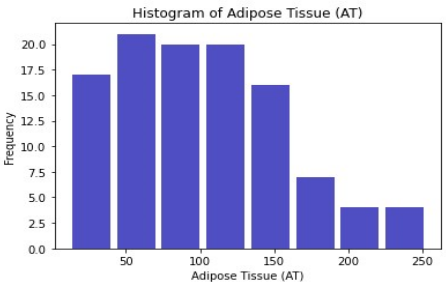
a) Check whether the MPG of Cars follows Normal Distribution
Dataset: Cars.csv

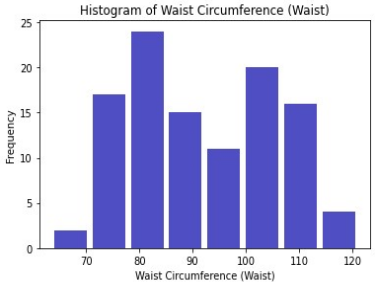
Ans: The 'MPG' from the dataset follows the normal distribution.



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

Ans: The Adipose Tissue and Waist from the data set does not follow the normal distribution.





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

Ans: For 90% of confidence interval Z score is approximately 1.645.
For 94% of confidence interval Z score is approximately 1.881.
For 60% of confidence interval Z score is approximately 0.842.

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

Ans: The t-scores for the given confidence intervals for a sample size of 25 is,

For 95% confidence interval = -2.064 to +2.064

For 96% confidence interval = -2.171 to +2.171

For 99% confidence interval = -2.797 to +2.797.

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: The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is approximately 32.17%.