|  |  |
| --- | --- |
| 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 probability of getting two heads and one tail when 3 coins are tossed are,

2^3 = 8 possible outcomes are available

HHT,THH,HTH are the outcomes that satisfy the condition.

Probability of getting 2 heads and 1 tail = 3/8=0.375

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

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

Ans: a) The probability that the sum is equal to 1 = 0

b) The probability that the sum is less than or equal to 2 and 3 ,

The possible outcomes are = (1,1),(1,2),(1,3),(2,1),(2,2),(3,1)

Probability = 6/36

= 1/6

c) The probability that the sum is divisible by both 2 and 3 are,

The possible outcomes are =(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)

There are 6 favourable outcomes out of 36 possible outcomes.

So,the probability of getting a sum is divisible by 2 and 3 = 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?

Ans: The total number of balls in a bag = 7 balls

The probability of drawing 2 balls out of 7 = 7C2 = 21

The probability of drawing 2 balls that are not blue = 5C2

= 10

Thus, the probability that none of the balls drawn is blue = 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,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

Ans: Mean of Points, Score,Weigh = (3.596,3.217,17.848)

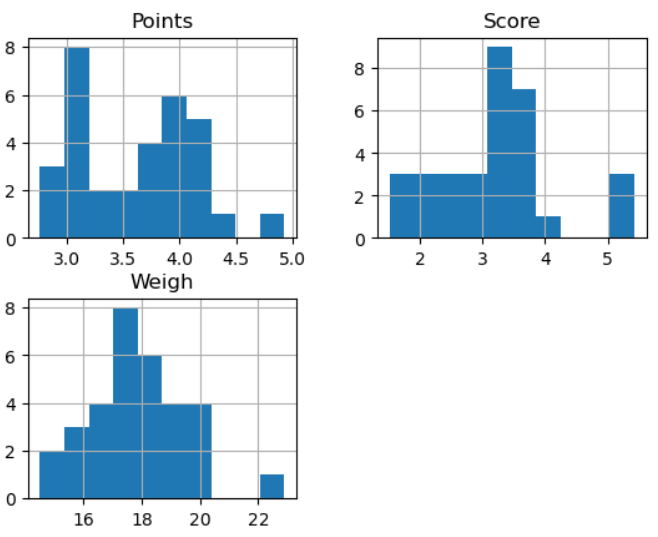
Median of Points, Score,Weigh = (3.695,3.325,17.710)

Mode of Points, Score,Weigh = (3.920,3.610,18.90)

Standard deviation of Points, Score,Weigh = (0.534,0.978,18.90)

Variance of Points, Score,Weigh =(0.731,0.988,4.34)

Range of Points, Score,Weigh = (2.17,3.911,8.4)



Q8) Calculate Expected Value for the problem below

1. 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.

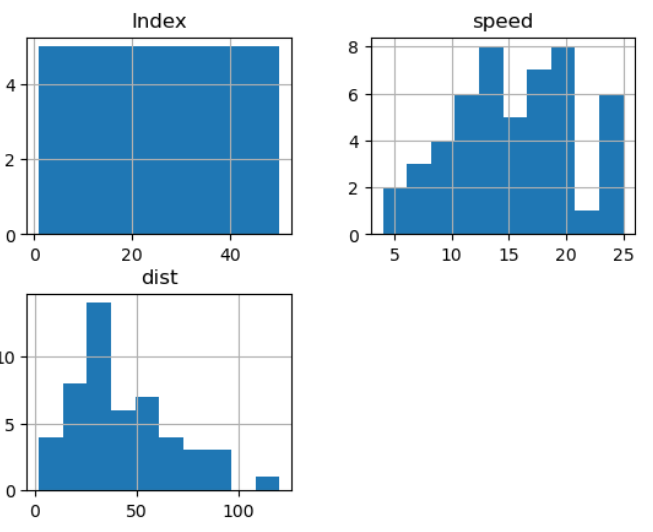
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**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]

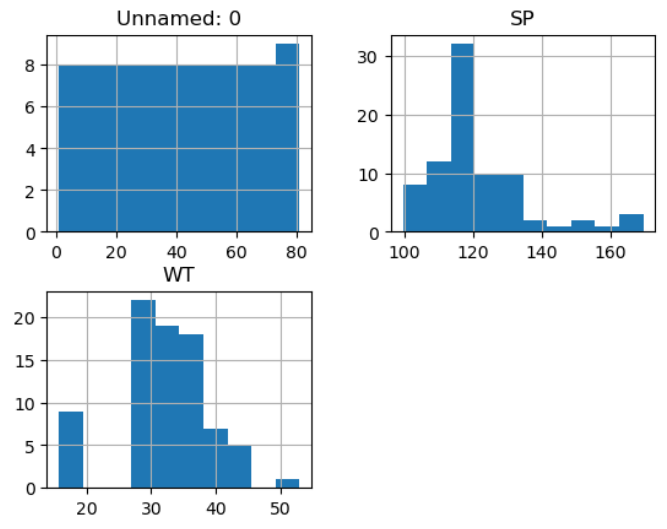


**SP and Weight(WT)**

**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.

Ans: Mean = 37.61

Median= 40

Variance = 57.12

Standard deviation = 7.55

1. What can we say about the student marks?

Ana: We can say that the student’s performance appears to be fairly consistent, with scores centered around the mean of 37.61 and there are some degrees of variability in the test of scores,indicated by variance and standard deviation which results that the student’s performance can change from test to test.

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

Ans: The nature of skewness = 0 when mean ,median of data are equal.

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

Ans: The nature of skewness is positively skewed when mean>median.

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

Ans: The nature of skewness is negatively skewed.

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

Ans: A positive kurtosis value indicates that a dataset has a highest peak value,

Kurtosis > 0

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

Ans: A negative kurtosis value indicates that a dataset has a lowest peak value,

Kurtosis < 0

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

* 1. P(MPG>38)

Ans: 0.330

* 1. P(MPG<40)

Ans: 0.729

* 1. P (20<MPG<50)

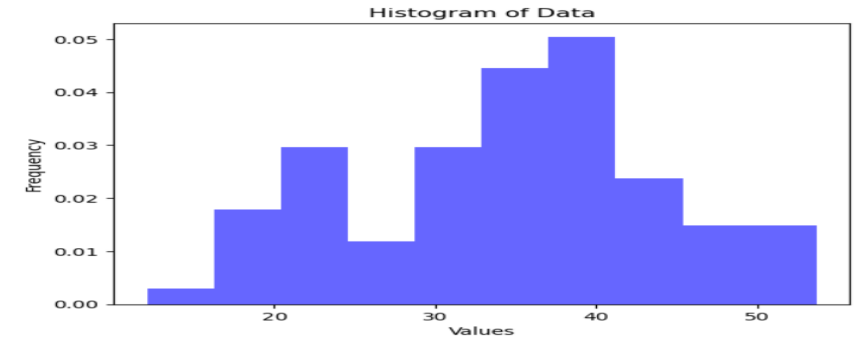
Ans: 0.898

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

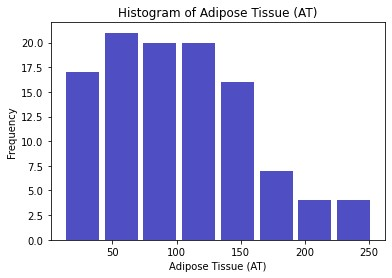
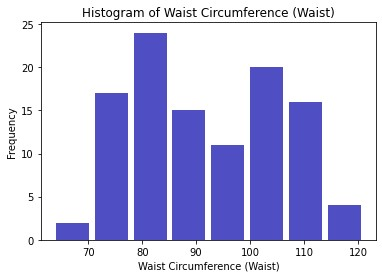
Ans: The ‘MPG’ from the dataset follows the normal distribution.



1. 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%.

​