|  |  |
| --- | --- |
| 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 | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Soln:- Total outcomes= 8 (i.e TTT,TTH,THT,THH,HTT,HTH,HHT, HHH)

Sample =3(i.e THH,HTH,HHT)

Probability=3/8

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

Sol:- Total Outcomes =36

1. Required outcomes= 0

Probability =0/36=0

1. Required outcomes=6 (i.e (1,1),(1,2),(2,1),(1,3),(2,2),(3,1))

Thus, the probability is:

P(sum≤4)= =

1. The possible outcomes of rolling two dice and their sums:

Dice1: 1,2,3,4,5,6

Dice2: 1,2,3,4,5,6

Sum: 2,3,4,5,6,7,8,9,10,11,12

Now, we need to find the sums that are divisible by both 2 and 3. This means they must be divisible by 6.

So the possible outcomes is :- (1,5),(2,3), (3,3),(6,6)

Therefore, the probability is:

P( sum divisible by 2 and 3)= =

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?

Soln: Total number of ways to draw 2 balls out of 7 is given by

Total number of ways = =21

The number of ways to draw 2 balls such that none of them is blue.

There are 5 non-blue balls in the bag (2 red + 3 green).

Number of ways to choose 2 non-blue balls out of 5: = =10

Therefore, the probability that none of the balls drawn is blue is:

P(none blue)= =

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

Soln:- Let X as as the random variable representing the number of candies a child has. Xi as the number of candies for child .

P(X=xi ) as the probability of child i having xi candies.

The expected value E[X] is calculated as:

E[X]= ∑i xi ∙ P(X=xi )

E[X]= (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

=3.075

Therefore, the expected number of candies for a randomly selected child is 3.075.

7) 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**

Answer- answer in csv file-

Q8) Calculate Expected Value for the problem below

1. 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?

Soln:- sum of weight =108+ 110+ 123+ 134+ 135+ 145+ 167+ 187+ 199=1308

Total number of patients = 9

Expected value (E)= = = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

Answer- answer in csv file-

**Q10) Draw inferences about the following boxplot & histogram**



**Answer:-**

1. Most of chick weight data is between 50 to 100 has high frequency i.e 200.
2. Second largest chick weight is between 100 to 150 that is between frequency of 100 to 150.

As much as ChickWeight$Weight is increasing the frequency is gradually decreasing



**Answer:-** Data which is out of the the boxplot is called outliers. There are 6 data points in this boxplot which is outliers

**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?

Soln: As we know

Confidence interval = Sample Mean ± (Critical value \* )

Where sample mean=200 pounds(Given)

Standard Deviation= 30 pounds (given)

Sample Size=2000 men (given)

1. For 94% confidence level:

The critical z-value is z=1.88 (approx.)

1. For 98% confidence level:

The critical z-value is z=2.33 (approx.)

1. For 96% confidence level:

The critical z- value z= 2.05 (approx.)

1. For 94% confidence level:

Confidence interval= 200 ± (1.88 \* ) = 200 ±2.12

1. For 98% confidence level:

Confidence interval= 200 ± (2.33 \* ) = 200 ±2.73

1. For 96% confidence level:

Confidence interval= 200 ± (2.05 \* ) = 200 ±2.30

**12)** 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.

Solution- Mean= (34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56)/18

Mean= 41

Meadian=40.5

Variance= 24.11

Standard Variance= 4.91

1. What can we say about the student marks?

Answer- Most of the student performed average is 41 marks

Median indicates that half of the student scored below 40.5 marks and half of the student is scored above the 40.5 marks. And mode indicates that most of the students obtained 41 marks.

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

Ans- When the mean (average) and median of a dataset are equal, it indicates that the data distribution is symmetric. In other words, the distribution is balanced around the center point. Skewness refers to the measure of asymmetry of a distribution. When the mean and median are equal, the skewness of the distribution is typically close to zero or exactly zero, indicating that the distribution is symmetric. when the mean and median are equal, there's no significant skewness in the distribution, as it is symmetric.

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

Ans:- When the mean is greater than the median, it typically indicates that the distribution is positively skewed. In a positively skewed distribution, the tail on the right-hand side (the upper tail) is longer or fatter than the tail on the left-hand side (the lower tail). This means that there are more extreme values on the right side of the distribution, pulling the mean towards them and causing it to be greater than the median. Positive skewness often occurs when there are a few extremely high values in the dataset, which pull the mean upwards. However, the majority of the data points are clustered towards the lower end of the distribution, closer to the median. Therefore, when the mean is greater than the median the distribution is positively skewed.

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

Ans: When the median is greater than the mean, it typically indicates that the distribution is negatively skewed. In a negatively skewed distribution, the tail on the left-hand side (the lower tail) is longer or fatter than the tail on the right-hand side (the upper tail). This means that there are more extreme values on the left side of the distribution, pulling the mean downwards and causing it to be less than the median. Negative skewness often occurs when there are a few extremely low values in the dataset, which pull the mean downwards. However, the majority of the data points are clustered towards the higher end of the distribution, closer to the median. Therefore, when the median is greater than the mean then the distribution is negatively skewed.

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

Ans:

A positive kurtosis value indicates that the distribution of the data has heavier tails and a sharper peak compared to a normal distribution. In other words, it indicates that the data is more "peaked" and has more outliers (extreme values) than a normal distribution.Positive kurtosis is often associated with distributions that have heavier tails, meaning there is a higher probability of extreme values occurring. This can occur in distributions such as heavy-tailed distributions like the t-distribution, or in distributions with long tails on both sides like the Laplace distribution. Therefore a positive kurtosis value suggests that the data has a peakier distribution with heavier tails, indicating the presence of more extreme values compared to a normal distribution.

Top of Form

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

Ans : A negative kurtosis value indicates that the distribution of the data has lighter tails and a flatter peak compared to a normal distribution. In other words, it suggests that the data has fewer outliers (extreme values) and is more dispersed around the mean compared to a normal distribution. Negative kurtosis is often associated with distributions that have lighter tails, meaning there is a lower probability of extreme values occurring. This can occur in distributions such as the uniform distribution, which is evenly spread across its range and lacks significant peaks or tails. In summary, a negative kurtosis value suggests that the data has a flatter distribution with lighter tails, indicating a lower likelihood of extreme values compared to a normal distribution.

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



What can we say about the distribution of the data?

**Answer: -** The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

**Answer: -**The data is a skewed towards left. The whisker range of minimum value is greater than maximum

What will be the IQR of the data (approximately)?   
**Answer: -** The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 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.

**Answer:-**

1. First there are no outliers.
2. Both the boxplot share same median that is approx. in an range between 275 to 250.
3. They are normally distributed with zero to no skewness either at minimum or maximum whisker range.

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

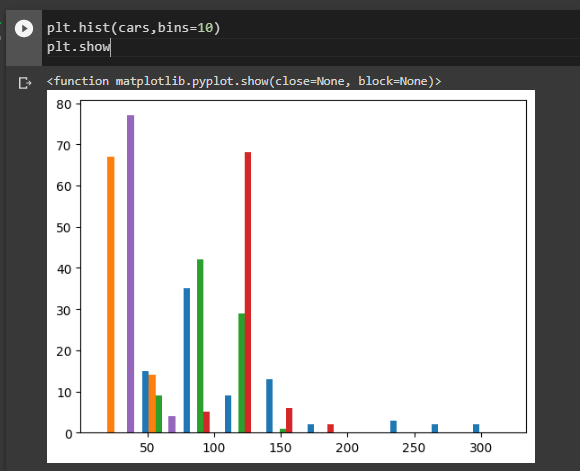
* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

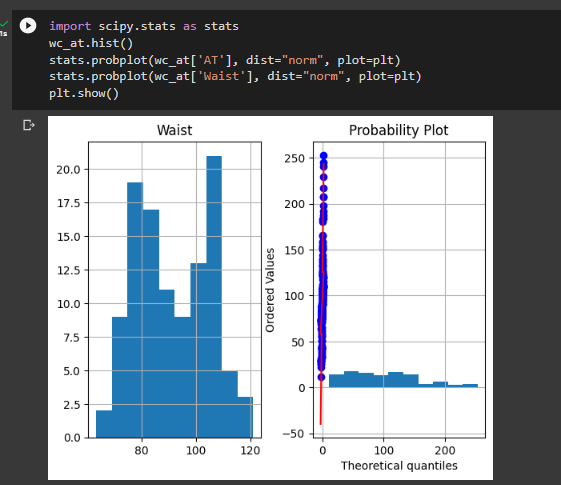
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

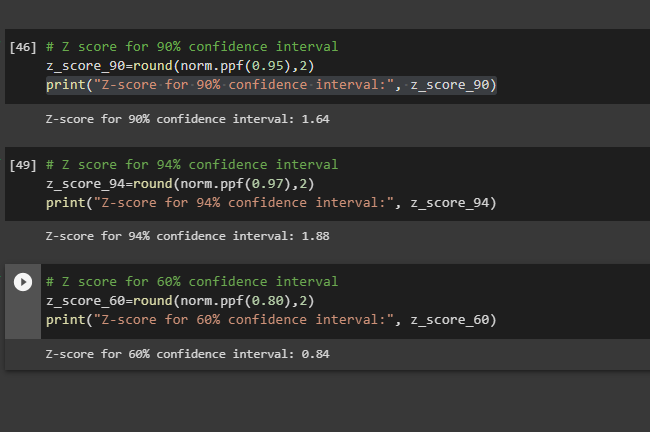


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

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

**A computer screen with text and numbers

Description automatically generated**

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

**A screenshot of a computer

Description automatically generated**