|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | ordinal |
| Time on a Clock with Hands | Interval |
| 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?

Answer:

When three coins are tossed, there are 2^3 = 8 possible outcomes because each coin can land either heads (H) or tails (T), and there are three coins.

The possible outcomes are: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT.

Out of these, the outcomes with two heads and one tail are: HHT, HTH, THH.

So, there are 3 favorable outcomes.

The probability (P) of getting two heads and one tail is given by:

P = Number of favourable outcomes / Total number of outcomes

Therefore, the probability of getting two heads and one tail when three coins are tossed is 3/8 = 0.375

The decimal representation of (3/8) is (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 2and 3

Answer=

When two dice are rolled, each die has numbers 1 through 6. The possible outcomes for the sum of two dice can be determined by considering all the combinations of the individual outcomes.

a) Probability that the sum is equal to 1:

There is only one way to get a sum of 1, which is if both dice show 1. The total number of outcomes is (6 \* 6 = 36) (as there are 6 possibilities for each die). Therefore, the probability is 1/36.

b) Probability that the sum is less than or equal to 4:

The possible combinations are (1,1), (1,2), (1,3), (2,1), (2,2), (3,1). There are 6 favorable outcomes. The total number of outcomes is still (6\*6 = 36\).

Therefore, the probability is 6/36 = 1/6.

c) Probability that the sum is divisible by 2 and 3:

The sums that are divisible by both 2 and 3 are 6 and 12. The combinations are (1,5), (2,4), (3,3), (4,2), and (5,1) for the sum of 6, and (6,6) for the sum of 12. There are 6 favorable outcomes. The total number of outcomes is (6 \* 6 = 36). Therefore, the probability is 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?

Answer:

Total number of balls = (2 + 3 + 2) = 7

Number of ways of drawing 2 balls out of 7 Total events =7C2 = 21

Number of ways of drawing 2 balls out of non-blue balls Interested events =5C2 = 10

Probability = (Interested Events) / Total Events = 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

Expected value for whole date set = 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24 = 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=attached

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?

**Solution:**

The expected value (or mean) of a set of values can be calculated by summing up all the values and then dividing by the number of values.

Given the weights of patients:

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

The expected value E(X) is calculated as follows:

E(X) = 1/n ∑ Xi

Where:

n is the number of values in the set.

(X\_i) is each individual value in the set.

Let's calculate it:

E(X) = 1/9 (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

E(X) = (1/9) 1318

E(X) = 146.44

Therefore, the expected value of the weight of a randomly selected patient is 146.44 pounds.

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

**Cars speed and distance**

**Use Q9\_a.csv**

Ans=attached

**SP and Weight(WT)**

**Use Q9\_b.csv**

Ans=attached

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

A graph of a person with a number of bars

Description automatically generated with medium confidence

Ans= Ans: The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side

Histogram: - Chick weight data is right skewed or positively skewed.---- Yes - More than 50% Chick Weight is between 50 to 150. ---- Yes - Most of the chick weight is between 50 to 100. --- Yes



Ans: The boxplot has outliers on the maximum side.

- The data is right skewed.

- There are outliers at upper side

**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?

Answer: Attached in file

**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?

Answer: Attached

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

Ans:

When mean = median then the skewness will be equal to 0. i.e. Normal Distribution

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

Ans:

When the median is greater than the mean, it suggests that the distribution of the data is negatively skewed, also known as left-skewed.

This skewness indicates that there are relatively more lower values in the dataset, pulling the mean to the left. The median, being less affected by extreme values, tends to be higher than the mean in a left-skewed distribution.

In summary:

- Median > Mean: Negatively skewed (left-skewed) distribution.

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

Ans:

Median > Mean: Negatively skewed (left-skewed) distribution.

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

Ans:

A positive kurtosis can be a sign of a distribution that has heavy tails, outliers, or is more peaked than a normal distribution. It suggests that there is more variability in the tails of the distribution.

In summary, positive kurtosis indicates a distribution with a more pronounced peak and heavier tails than a normal distribution.

Top of Form

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

Ans:

A negative kurtosis suggests that the data is less extreme or has less variability in the tails than a normal distribution. It implies a flatter, more spread-out distribution of values.

In summary, negative kurtosis indicates a distribution with lighter tails and a flatter peak than a normal distribution, suggesting less extreme values and less variability in the tails.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

What will be the IQR of the data (approximately)?   
  
1. Ans: As on the y-axis it plots from 2 to 18, it shows 25% of the data 1st quartile is Around 10,

where in 75% of the data is 18 and median which is 2nd quartile is around 14.8.

2. Ans: Data is negatively skewed as median is close to upper whisker of the box.

3. Ans: Q3 = 18.1, Q1 = 10 Hence, Q3 – Q1 = 18.1-10 = 8.1 = IQR.

Q19) Comment on the below Boxplot visualizations?



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

Ans: Box plot 1 and 2 have different IQR but Normally distributed as median for both is in center of both whiskers.

BP1’s min value is higher than the BP2 min value and also BP1’s max value is lower than the BP2’s max value.

BP1 covers lowest IQR wherein, BP2 covering Higher IQR. However median for both is equal.

Based on the observation, even if both Box plots have equal median and normally distributed. But, since Box plot 2 covering higher range of data is most accurate.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

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

c. P (20<MPG<50)

Ans=attached

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans=attached

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

Dataset: wc-at.csv

Ans=attached

Q 22) Calculate the Z scoresof 90% confidence interval,94% confidence interval, 60% confidence interval

Ans=attached

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

Ans=attached

Q 24**)** A Government companyclaims 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=attached