|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Integer |
| Results of rolling a dice | Integer |
| Weight of a person | Float |
| Weight of Gold | Float |
| Distance between two places | Floating- Number Point |
| Length of a leaf | Floating- Number Point |
| Dog's weight | Floating- Number Point |
| Blue Color | String |
| Number of kids | Integer |
| Number of tickets in Indian railways | Integer |
| Number of times married | Integer |
| Gender (Male or Female) | String |

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 | Interval |
| 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?

n = 3 (three-coin tosses)

k = 2 (two heads)

p = 1/2 (probability of getting a head in a fair coin toss)

Now, let's calculate the probability:

P (X = 2) = (3 choose 2) \* (1/2) ^2 \* (1 - 1/2) ^ (3 - 2)

P (X = 2) = (3! / (2! \* (3 - 2)!)) \* (1/4) \* (1/2)

P (X = 2) = (3 / 2) \* (1/4) \* (1/2)

P (X = 2) = (3/8)

So, the probability of getting two heads and one tail when three coins are tossed is 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

Total number of outcomes when rolling two dice = 6 (for the first die) \* 6 (for the second die) = 36 outcomes.

a) Probability that the sum is equal to 1:

There is only one way to get a sum of 1, which is when both dice show 1.

Number of outcomes with a sum of 1 = 1

Probability = (Number of favorable outcomes) / (Total number of outcomes) = 1 / 36

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

For a sum of less than or equal to 4, you can have the following combinations: (1, 1), (1, 2), (2, 1), (1, 3), (2, 2), and (3, 1).

Number of outcomes with a sum less than or equal to 4 = 6

Probability = (Number of favorable outcomes) / (Total number of outcomes) = 6 / 36 = 1 / 6

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

To have a sum divisible by both 2 and 3, the sum must be divisible by 6. The possible combinations are (1, 5), (2, 4), and (3, 3).

Number of outcomes with a sum divisible by 2 and 3 = 3

Probability = (Number of favorable outcomes) / (Total number of outcomes) = 3 / 36 = 1 / 12

So, the probabilities are as follows:

a) Probability that the sum is equal to 1: 1/36

b) Probability that the sum is less than or equal to 4: 1/6

c) Probability that the sum is divisible by 2 and 3: 1/12

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?

First Draw:

The probability of drawing a non-blue ball on the first draw is the sum of the probabilities of drawing a red ball and a green ball because these are the non-blue options.

Number of non-blue balls = 2 (red) + 3 (green) = 5

Total number of balls = 2 (red) + 3 (green) + 2 (blue) = 7

Probability of drawing a non-blue ball on the first draw = Number of non-blue balls / Total number of balls = 5/7

Second Draw:

After drawing a non-blue ball on the first draw, there are now 4 non-blue balls left in the bag (2 red and 2 green) and a total of 6 balls remaining in the bag.

Probability of drawing a non-blue ball on the second draw = Number of non-blue balls left / Total number of balls left = 4/6 = 2/3

To find the overall probability of both events happening (drawing a non-blue ball on the first draw and then another non-blue ball on the second draw), you multiply the probabilities:

Overall Probability = (Probability of first draw) \* (Probability of second draw)

Overall Probability = (5/7) \* (2/3)

Now, calculate the product:

Overall Probability = (5/7) \* (2/3) = 10/21

So, the probability that none of the balls drawn is blue is 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 = Σ (X \* P)

Where:

X is the value (candies count)

P is the probability associated with that value

Let's calculate it step by step:

Child A:

Candies count (X) = 1

Probability (P) = 0.015

Child B:

Candies count (X) = 4

Probability (P) = 0.20

Child C:

Candies count (X) = 3

Probability (P) = 0.65

Child D:

Candies count (X) = 5

Probability (P) = 0.005

Child E:

Candies count (X) = 6

Probability (P) = 0.01

Child F:

Candies count (X) = 2

Probability (P) = 0.120

Now, calculate the expected value:

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

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

Expected Value = 3.115

The expected number of candies for a randomly selected child is 3.115 candies.

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**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
|  | Maximum Value: 19.75 | | |  |  |  |  |
|  | Minimum Value: 11.06 | | |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  | Range = Maximum Value - Minimum Value = 19.75 - 11.06 ≈ 8.69 | | | | | |  |
|  |  |  |  |  |  |  |  |
|  | So, the range for your dataset is approximately 8.69. | | | | |  |  |

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?

Given the weights of patients:

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

Now, calculate the expected value:

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

E = 1273 / 9

E ≈ 141.44

**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**

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



**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?

**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?

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

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

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

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

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

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)?

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

* 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

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