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

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

A. If three coins are tossed,

Total number of possible combinations = 23 = 8

The combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

Number of combinations that have two heads and one tail = 3, i.e., HHT, HTH, TTH

The probability of two heads and one tail when three coins are tossed simultaneously are

P (Two heads and One tail) = Number of desired outcomes

= ⅜ or 0.375

**the probability of two heads and one tail is ⅜ or 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

A.To find the probability of different sums when two dice are rolled, we can use a probability table. Each die has 6 sides, numbered 1 through 6. There are a total of 6 x 6 = 36 possible outcomes when rolling two dice.

a) Probability that the sum is equal to 1:

There is only one way to get a sum of 1, which is rolling a 1 on the first die and a 1 on the second die. So, there is 1 favorable outcome.

P(sum = 1) = (Number of favorable outcomes) / (Total possible outcomes) = 1/36

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

The sums less than or equal to 4 can be achieved in the following ways:

- Sum = 2 (1+1)

- Sum = 3 (1+2, 2+1)

- Sum = 4 (1+3, 3+1, 2+2)

There are 10 favorable outcomes in this case.

P(sum ≤ 4) = (Number of favorable outcomes) / (Total possible outcomes) = 10/36 = 5/18

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

To have a sum divisible by both 2 and 3, we're looking for sums that are divisible by 6. The only way to get a sum of 6 on two dice is by rolling a 3 on each die.

P(sum is divisible by 2 and 3) = P(sum = 6) = 1/36

So, the probabilities are as follows:

a) P(sum = 1) = 1/36

b) P(sum ≤ 4) = 5/18

c) P(sum is divisible by 2 and 3) = 1/36

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?

A. To find the probability that none of the balls drawn is blue, you can use the concept of conditional probability. First, calculate the total number of ways to draw 2 balls from the bag, which is a combination of 7 balls taken 2 at a time:

Total ways to draw 2 balls = C(7, 2) = 7! / [2!(7-2)!] = 21 ways

Now, calculate the number of ways to draw 2 balls without getting any blue ball. You can draw both balls from the red and green balls. The number of ways to choose 2 balls from 2 red and 3 green balls is:

Ways to choose 2 non-blue balls = C(2, 2) \* C(3, 0) = 1 \* 1 = 1 way

So, the probability of drawing 2 non-blue balls is the number of favorable outcomes (1 way) divided by the total number of possible outcomes (21 ways):

Probability = 1/21

So, the probability that none of the balls drawn is blue is 1/21.

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 |

A. To calculate the expected number of candies for a randomly selected child, you can multiply the candy count for each child by their respective probabilities, and then sum up these values. The formula for calculating the expected value (E) is:

E = Σ (X \* P)

Where:

- E is the expected value (in this case, the expected number of candies).

- X is the value (candy count).

- P is the probability.

Using the provided data:

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

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

E = 3.115

So, the expected number of candies for a randomly selected child is 3.115.

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**

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?

A. To calculate the expected value of the weight of a randomly chosen patient, you need to find the mean (average) of the given weights.

The formula for the expected value (E) is:

E = (ΣX) / N

Where:

- E is the expected value.

- ΣX is the sum of all the values (weights in this case).

- N is the number of values (number of patients).

First, sum up the weights:

ΣX = 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199

ΣX = 1098

Now, calculate the expected value:

E = ΣX / N = 1098 / 9 = 122

So, the expected value of the weight of a randomly chosen patient is 122 pounds.

**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**



**A.The most of the data points are concerated in the range 50-100 with frequency 200.And least range of weight is 400 somewere around 0-10.So the expected value the above distribution is 75.Skewness- we can notice a long tail towards right so it is heavily right skewed.**



**Sol: Medican is less than mean right skewed and we have outlier on the upperside of box plot and there is less data points between Q1 and bottom point.**

**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?

A. To calculate confidence intervals for the average weight of adult males in Mexico, you can use the standard error of the sample mean and the Z-scores corresponding to the desired confidence levels.

The formula for the confidence interval for the population mean is:

Confidence Interval = X̄ ± Z \* (σ/√n)

Where:

- X̄ is the sample mean (200 pounds).

- Z is the Z-score corresponding to the desired confidence level (1.96 for 94%, 2.33 for 98%, and 2.05 for 96% confidence level). You can look up these values in a standard normal distribution table.

- σ is the population standard deviation (since you have a large population, it's assumed to be the same as the sample standard deviation, which is 30 pounds).

- n is the sample size (2,000 men).

Let's calculate the confidence intervals for the three desired confidence levels:

1. 94% Confidence Interval:

Z = 1.96 (for 94% confidence)

Confidence Interval = 200 ± 1.96 \* (30/√2000) = 200 ± 2.47

The 94% confidence interval is (197.53, 202.47).

2. 98% Confidence Interval:

Z = 2.33 (for 98% confidence)

Confidence Interval = 200 ± 2.33 \* (30/√2000) = 200 ± 2.73

The 98% confidence interval is (197.27, 202.73).

3. 96% Confidence Interval:

Z = 2.05 (for 96% confidence)

Confidence Interval = 200 ± 2.05 \* (30/√2000) = 200 ± 2.40

The 96% confidence interval is (197.60, 202.40).

So, the 94% confidence interval is (197.53, 202.47), the 98% confidence interval is (197.27, 202.73), and the 96% confidence interval is (197.60, 202.40) for the average weight of adult males in Mexico.

**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?

**A).Symetrical**

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

**Right skewed**

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

**Left Skewed**

Q16) What does positive kurtosis value indicates for adata ?

**The data is notmally distributed and kurtosis value is 0.**

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

**The distribution of the data has lighter tails and a flatter peaks than the normaldistribution.**

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



What can we say about the distribution of the data?

A.The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

A.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)?

A. 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.

 A.Here when we compare box plot 1 with box plot 2 we can say that the data in boxplot 1 is widely spread. Here the main inference is that since the data range varieshigh in box plot 2 it is hard to make a prediction in box plot 2. The median in the 2box 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 ofCars 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 scoresof 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 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