Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 |

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 | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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, as each coin has two possible outcomes (heads or tails). The eight possible outcomes are:

{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

There are three outcomes that satisfy this condition (that two heads and one tail are obtained): HHT, THH, and HTH.

Probability of getting two heads and one tail is:

Probability = Number of favorable outcomes/Total number of outcomes

Probability = 3/8

Therefore, the probability of getting two heads and one tail when three coins are tossed is 3/8 = 0.375 = 37.5%

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

Answer:

When two dice are rolled, each die has 6 faces, and the possible outcomes for each die are 1, 2, 3, 4, 5, and 6. So, 36 outcomes are possible.

a) The sum equal to 1:

There is only one possibility to get a sum of 1, which is if both dice show 1. The possible outcome is (1, 1). So, the probability is:

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

= 1/36 = 0.027 = 2.7%

b) The sum less than or equal to 4:

The possible outcomes for sums less than or equal to 4 are (1, 1), (1, 2), (2, 1), and (1, 3), (3, 1), (2, 2). There are a total of 6 favorable outcomes.

So, the probability is:

Probability (Sum <= 4) = Number of favorable outcomes/Total number of possible outcomes

= 6/36 = 0.166 = 16.6%

c) The sum is divisible by 2 and 3:

The sums that are divisible by both 2 and 3 are 6 and 12. The possible outcomes for these sums are (3, 3) and (6, 6). There are a total of 2 favorable outcomes.

So, the probability is:

Probability (Sum is divisible by 2 and 3) = Number of favorable outcomes/Total number of possible outcomes

= 2/36 = 0.055 = 5.5%

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:

To find the probability that none of the balls drawn is blue, we can use the concept of combinations.

The total number of balls in the bag is 2 + 3 + 2 = 7.

The number of ways to draw 2 balls without any restriction is given by the combination formula:

Number of ways = \binom {n}{r} = n! / r! (n-r)!

where n! denotes the factorial of n.

For this problem, we want to find the number of ways to draw 2 balls out of the 5 non-blue balls. So,

Number of ways to draw 2 non-blue balls = \binom {5}{2} = 5! /2! (5-2)! = 10

Now, the total number of ways to draw 2 balls from the 7 balls is:

Total number of ways to draw 2 balls = \binom {7}{2} = 7! /2! (7-2)! = 21

The probability that none of the balls drawn is blue is the ratio of the number of ways to draw 2 non-blue balls to the total number of ways to draw 2 balls:

Probability (None are blue) = Number of ways to draw 2 non-blue balls/Total number of ways to draw 2 balls

Probability (None are blue) = 10/21

Therefore, the probability that none of the balls drawn is blue is 10/21

= 0.476 = 47.6%

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

Answer: To calculate the expected number of candies for a randomly selected child, you can use formula:

E(X) = sum\_{i} (xi \* Probability(xi))

Where:

-E(X) is the expected value,

-xi is the candies count for child i,

-Probability(xi) is the probability of child I having xi candies,

-the sum is taken over all possible values of xi.

Let’s calculate it for given data:

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

E(X) = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24

E(X) = 3.09

Therefore, the expected number of candies for a randomly selected child is 3.09 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**

Answer:

Answer is in python notebook.

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?

Answer:

The expected value (or mean) is calculated by summing up all the values and then dividing by the number of values. In this case, the weights (X) of patients are:

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

To find the expected value (E(X)), you sum up all the weights and divide by the total number of patients:

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

Calculating the sum:

E(X) = 1308 / 9

= 145.33

So, the expected value of the weight of a randomly chosen patient is approximately 145.33 pounds.

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

**Car’s speed and distance**

**Use Q9\_a.csv**

**SP and Weight (WT)**

**Use Q9\_b.csv**

Answer:

Answer is in python notebook.

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



Answer:

1. The histograms peak has positive(right) skewness and tail is on right.

Which means Mean > Median. We have outliers on the higher side.

The boxplot has outliers on the maximum 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:

94% Confidence interval of given problem is (198.7376, 201 .2624)

98% Confidence interval of given problem is (198.4382, 201 .5618)

96% Confidence interval of given problem is (198.6214, 201 .3786)

**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:

1. Mean for given data is 41, median is 40.5, variance is 25.5294, standard deviation is 5.0527.
2. We don’t have outliers and the data is slightly skewed towards right because mean is greater than median. No extreme value present in given marks.

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

Answer: No Skewness will be present, we have perfect symmetrical distribution.

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

Answer: Distribution will have positive skewness, right side tail is longer than left side tail.

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

Answer: Distribution will have negative skewness, left side tail is longer than right side tail.

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

Answer: Positive kurtosis means the curve is more peaked and it is Leptokurtic.

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

Answer: Negative Kurtosis means the curve will be flatter and broader

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 no outliers are present.

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 = Upper quartile(Q3) –Lower Quartile(Q1) = 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:

There are no outliers.

Both the box plot shares the same median that is approximately in a range between 275 to 250.

They are normally distributed with zero skewness neither at the minimum or maximum whisker range.

Range of both boxplots are different, therefore size of boxplots are different.

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)
  3. P (20<MPG<50)

Answer: Answer is in python notebook.

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: MPG of cars follows normal distribution

Answer is in python notebook.

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

Dataset: wc-at.csv

Ans: Adipose Tissue (AT) and Waist does not follow Normal Distribution

Answer is in python notebook.

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

Answer: Answer is in python notebook.

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

Answer: Answer is in python notebook.

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

Answer:

samp\_mean = 260, pop\_mean = 270, n = 18

To find t-score: t\_score = (260-270)/(90/np.sqrt(n))

Finally code to find probability with respect to t-score:

stats.t.cdf(t-score, df = 17)

Output: 0.3217

Therefore, the probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 32.17%.