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

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

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

***Answer:***

To find: find the probability that two heads and one tail are obtained

On tossing 3 coins, the possible outcomes are {HHH, HHT, HTT, HTH,TTT,TTH,THH,THT}

Therefore, the total number of outcomes is 8

Getting two heads and one tail includes {HHT, HTH,THH}

Therefore, the number of favorable outcomes is 3

Hence, using the coin toss probability formula,

the probability of getting two heads and one tail is 3/8 = 0.375

**Answer: The probability of getting two heads and one tail 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 2 and 3

***Answer:***

**Assumptions**

1. The dice are “ fair “, that is, not biassed in any manner.
2. The dice are both six-sided dice, that is both have 6 faces, with each face on each dice, showing one of the numbers, 1 to 6, with no number repeated on the same dice.

Analysis:

With two dice, there are ( 6 ) \* ( 6 ) = ( 36 ) possible combinations of numbers.

The minimum sum possible for the two dice thrown is (1, 1) = a sum of (2 )

The maximum sum possible for the two dice thrown is (6, 6) = a sum of (12)

When 2 dice are rolled total possible outcomes are 36. They are :-

(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)

(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)

(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)

(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)

(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)

(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)

1. Sum Equal to 1

The minimum possible sum is (1, 1) = ( 2 ).

Therefore P( 1 ) = ( 0 )/( 36 ) = 0

***Thus ,there is zero probability of getting the sum=1***

1. Sum Less than or equal to 4

Total favourable outcomes to get a sum of less or equal to 4, when 2 dice are rolled simultaneously = 6

{i.e(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)}

Probability = favourable outcomes /total outcomes

P = 6/36

P = 1/6=0.1667=16.67%

***Thus ,there is 16.67% probability of getting the sum Less than or equal to 4***

1. Sum is divisible by 2 and 3

Total favourable outcomes to get a sum which is divisible by 2 and 3 ,when 2 dice are rolled simultaneously = 6

{i.e(1,5) ,(2,4), (3,3), (4,2), (5,1) ,(6,6)}

Probability = favourable outcomes /total outcomes

P = 6/36

P = 1/6=0.1667=16.67%

***Thus ,there is 16.67% probability of getting the sum is divisible by 2 and***

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.

n(S)= Number of ways of drawing 2 balls out of 7 = 7C2 =21

Let E = Event of drawing 2 balls, none of which is blue.

n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls. = 5C2=10

so, probability=10/21 = 0.4762= 47.62 %

***so , the probability that none of the balls drawn is blue , is =10/21=0.4762=47.62%***

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:***

**Step-by-step explanation:**

Expected number of candies for a randomly selected child

= 1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

=   3.090

= 3.09

***So,Expected number of candies for a randomly selected child  = 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**

**Answer:**

***Solution of this question is in the jupyternotebook 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?

Answer:

***Solution of this problem is solved in jupyternotebook file.***

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

Cars speed and distance

Use Q9\_a.csv

**Answer:**

***Solution of this problem is solved in jupyternotebook file.***

SP and Weight(WT)

Use Q9\_b.csv

**Answer:**

***Solution of this problem is solved in jupyternotebook file.***

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



***Solution:***

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



**Solution:**

Boxplot:

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?

Solution:

***Solution of this problem is solved in jupyternotebook 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?

Solution:

***Solution of this problem is solved in jupyternotebook file.***

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

**Answer:**

* skewness of data is zero when mean and median are equal.
* the left-hand side, and right-hand side of mean contain the same number of observations.
* Data is normally distributed.

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

**Anwer:**

* Data has positively skewed distribution
* In positively skewed, the mean of the data is greater than the median (a large number of data-pushed on the right-hand side).
* In other words, the results are bent towards the lower side.
* The mean will be more than the median as the median is the middle value and mode is always the highest value

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

**Answer:**

* Distribution is Negative skewed or left-skewed
* In negatively skewed, the mean of the data is less than the median (a large number of data-pushed on the left-hand side).
* Negatively Skewed Distribution is a type of distribution where the mean, median, and mode of the distribution are negative rather than positive or zero.
* a large number of data-pushed on the left-hand side

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

**Answer:**

* Kurtosis refers to the degree of presence of outliers in the distribution.
* Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. which means there are more chances of outliers
* A large kurtosis is associated with a high level of risk for an investment because it indicates that there are high probabilities of extremely large and extremely small returns
* An extreme positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

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

**Answer:**

* having a lower tail and stretched around center tails
* most of the data points are present in high proximity with mean.
* distribution is flatter (less peaked) when compared with the normal distribution.

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



Q.What can we say about the distribution of the data?

**Answer:**

* the median of data is approximately 15
* data is left skewed
* the mean of the data is less than the median
* a large number of data-pushed on the left-hand side.

Q.What is nature of skewness of the data?

**Answer:**

* left-skewed distribution or Negative skewed

Q. What will be the IQR of the data (approximately)?

**Answer:**

IQR = 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:**

* Median value of both the boxplot is same
* Range of distribution of value in case of the 1 st box plot is less as compared to 2nd boxplot.
* Both have the normally distributed data i.e zero skewness

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:***

***Solution of this problem is solved in jupyternotebook file.***

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

***Answer:***

***Solution of this problem is solved in jupyternotebook file.***

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

Dataset: wc-at.csv

***Answer:***

***Solution of this problem is solved in jupyternotebook file.***

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

***Answer:***

***Solution of this problem is solved in jupyternotebook file.***

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

***Answer:***

***Solution of this problem is solved in jupyternotebook file.***

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:***

***Solution of this problem is solved in jupyternotebook file.***