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

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

Answer: S = { HHH , HHT , HTH , THH , TTT , TTH , THT , HTT }

n(S) = 8

P = { HHT , HTH , THH }

n(P) = 3

Probability of two heads and one tail is = n(P) / n(S)

= 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

Answer:

1. Probability=0
2. Probability=.1

since there are only 6 possible outcomes and total no of outcomes is 36, so 6/36.

1. Probability = 0.1

since there are only 6 possible outcomes and total no of outcomes is 36, so 6/36

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?

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

Let S be the sample space.  
Then, n(S) = Number of ways of drawing 2 balls out of 7  
= 7C2 = (7×6) / (2×1) = 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​ = (5×4) / (2×1)​ = 10

∴ P(E) = n(E) / n(S) = 10/21​

The probability of getting no blue balls 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

Answer:

Expected number of candies for a randomly selected child = 3.09

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

= 3.09

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

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?

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



Answer: The histogram and boxplot figure are positively skewed on right side. i.e., Mean and median of the data is greater than mode. 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?

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?

Answer: If the mean is equal to the median as well as the mode, hence the skewness is zero. If the distribution is symmetric, the mean equals to median, and the skewness of the distribution is zero.

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

Answer: : If the mean is greater than the median, then distribution is positively skewed.

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

Answer: : If the mean is less than the median, the distribution is negatively skewed.

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

Answer: Positive kurtosis means the curve is more peaked

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

Answer: The peak is flatter and broader also the tails are heavily distributed

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



What can we say about the distribution of the data?

Answer: The distribution is positively distributed by data

What is nature of skewness of the data?

Answer: The Mean of negatively skewed data will be less than the Median.

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

Answer: The IQR describes the middle 50% of values when ordered from lowest to highest.

IQR= Q3 upper quartile - Q1 lower quartile

=18-10 IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

Answer: According to the plots shown, there are no outliers,the median of both the plot are similar as shown ,that is the range from 250-275 ,both the data are normally distributed with zero to no skewness at the minimum or maximum whisker range.

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