|  |  |
| --- | --- |
| 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) | Continuous |

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 | Ratio |
| Weight | Ratio |
| Hair Color | Ordinal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Ordinal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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?

Ans = If three coins are tossed,

Total number of possible combinations = 23 = 8

Sample Space = { 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 =

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

= 3/8 = 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

Ans =

If two dices were rolled, then total possible cases =36

a) Total Favourable cases (Having sum =1) = 0

As minimum sum is 2 for outcome (1,1).

Hence, probability is 0.

b) Total Favourable cases (Having sum <=4) =6

P (Sum<=4) = 6/36

= 1/6 = 0.16

1. Having sum is divisible by 2 and 3 i.e. Sum should be 6,12

Total Favourable cases (Having sum is divisible by 2 and 3) = 6

P = 6/36 =1/6 = 0.16

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?

Ans = Total no. of balls = 7

Let S be the sample space,  
Then, n(S) = Number of ways of drawing 2 balls out of 7  
 = 7C2​  
 = (7\*6)​/ (2\*3)  
 = 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

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

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**



**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?

Ans = The distribution will be approximately Zero Skew.

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

Ans = The distribution will be Positively Skew.

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

Ans = The distribution will be Negatively Skew.

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

Ans = Positive kurtosis value indicates that the distribution is peaked and possesses thick tails.

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

Ans = Negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

Ans = The 50% of the data are distributed in between 10 to 18. The 25% data lies below 10 and 25% data lies above 18.

What is nature of skewness of the data?

Ans = The distribution is negatively skew.

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

Ans = IQR = Q3-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.

Ans = Median of both box plots are same. Most of the datapoints of Boxplot 2 are the outliers for the boxplot 1. The datapoints of boxplot 2 are more distributed than boxplot 1. All the datapoints of Boxplot 1 are lies in between the IQR of 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

Ans = At 90% Confidence Interval z score = 1.645,

At 94% Confidence Interval z score = 1.555,

At 60% Confidence Interval z score = 50.253.

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