|  |  |
| --- | --- |
| 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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| 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 | Ratio |

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

Ans: 3 coins tossed , sample space S is

S= {HHH, HHT, HTT, TTT, TTH, THH, HTH, THT}, A is event that have 2 heads and 1 tail, A = {HHT, THH, HTH}, n(A)= 3 then

P(A)=n(A)/n(S) = 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: Two diced are tossed then sample space S is ,

S={(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)}, n(S) = 36

1. A is event that sum is 1 i.e. A={}, n(A)=0 , P(A)= n(A)/n(S) = 0/36 = 0
2. B is event that sum is less than or equal to 4 i.e. B={(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)} n(B) = 6 , P(A) = n(A)/n(S) = 6/36
3. C is divisible by 2 and 3 i.e. C={(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)} , n(C) = 6

P(C) = n(C )/n(S) = 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?

Ans : Total no. of balls = n(S) = 7 , If 2 balls are drawn at random then Probability that none of the ball blue is ,

Prob = 5C2/7C2 = 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

Ans : The expected no. of candies for the children are 0.14714

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**

**Ans : For Points**

|  |  |
| --- | --- |
| **Mean** | **3.5966** |
| **Mode** | **3.695** |
| **Median** | **3.07 , 3.92** |
| **Std Dev** | **0.5347** |
| **Variance** | **0.2859** |
| **Range** | **2.17** |

**Conclusion : The distribution is bimodal.**

**For Score**

|  |  |
| --- | --- |
| **Mean** | **3.2173** |
| **Mode** | **3.44** |
| **Median** | **3.325** |
| **Std Dev** | **0.9785** |
| **Variance** | **0.9573** |
| **Range** | **3.911** |

**Conclusion :** Here mode>Median>mean , the distribution is left or negatively skewed.

**For Weigh**

|  |  |
| --- | --- |
| **Mean** | **17.8487** |
| **Mode** | **17.02, 18.90** |
| **Median** | **17.71** |
| **Std Dev** | **1.7869** |
| **Variance** | **3.1932** |
| **Range** | **8.3999** |

**Conclusion:** Here mean is greater than median and difference between Q3 and maximum is large, so this data is positively skewed i.e. right skewed.

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?

Ans : The expected value is 145.3333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans: For Speed**

**Skewness for speed is -0.117509**

**Kurtosis for speed is -0.50899**

**Inference: By skewness and kurtosis the distribution is approximately normal.**

**For Distance,**

**Skewness for distance is 0.8069**

**Kurtosis for distance is 0.4050**

**Inference: By skewness the distribution is moderately positive skewed and kurtosis the distribution is mesokurtic i.e. normal .**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans :**

**For SP**

**Skewness for SP is 1.6114**

**Kurtosis for SP is 2.9773**

**Inference: By skewness the distribution is positively skewed and by kurtosis it is leptokurtic.**

**For Weight**

**Skewness for weight is -0.6147**

**Kurtosis for weights is 0.9502**

**Inference: By skewness and kurtosis the distribution is approximately normal.**

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



**Ans: The distribution has larger right tail; hence it is positively skewed and it has outliers at the upper extreme.**

**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?

**Ans: Confidence interval for**

1. **94% is (198.9567,201.0433)**
2. **96% is (198.825,201.175)**
3. **98% 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.

**Ans: Mean = 41**

**Median =40.5**

**Variance = 25.5294**

**Standard deviation =5.0526**

1. What can we say about the student marks?

**Ans: The mean and median value are near close to each other and also variance is low , so data is approximately normal.**

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

**Ans: No skewness , data is normally distributed .**

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

**Ans: Data is positively distributed.**

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

**Ans: Data is negatively distributed.**

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

**Ans : Positive kurtosis value indicates that distribution has larger peak than usual and it is leptokurtic.**

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

**Ans: Negative kurtosis value indicates that data has flat peak than usual and distribution is platykurtic.**

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



What can we say about the distribution of the data?

**Ans: The distribution has larger left tail.**

What is nature of skewness of the data?

**Ans: The distribution is negatively skewed.**

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

**Ans: Approximately 8**

Q19) Comment on the below Boxplot visualizations?



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

**Ans: The boxplot 1 spread from about 240 to 286 and the boxplot 2 from about 185 to 340 , the spread of the boxplot 1 is less wider than 2 also range is smaller than 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)
  3. P (20<MPG<50)

**Ans:**

1. **0.3473**
2. **0.7294**
3. **0.8989**

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 follow Normal Distribution.**

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

Dataset: wc-at.csv

**Ans: Both the AT and Waist follows normal distribution.**

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

**Ans: For 90% Z value is 1.2815**

**For 94% Z value is 1.5547**

**for 60 % is 0.2533**

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

**Ans:**

1. **For 95 % CI t value is 1.71**
2. **For 96% CI t value is 1.8280**
3. **For 99% CI t value is 2.4921**

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

**Ans: The probability is 0.3216 .**