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

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

Sol: S = [ HHH, TTT, HHT, HTH, HTT, THH, TTH, THT ] = 8

Two heads and One tail = [ HHT, HTH, THH ] = 3

Probability that two heads and one tail are = 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

Sol:

Two dice rolled : s = 6^2 =36

1. Probability that Sum = 1 : [(1,1)] = 1/36
2. Probability that Sum <= 4 : [(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)] = 6/36 = 1/6
3. Probability that sum is divisible by 2&3 : [(1,5),(5,1),(2,4),(4,2),(3,3),(6,6)] = 6/36=1/6

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?

Sol:

Total balls S= 2+3+2 = 7

Probability of 2 balls drawn at random = (7\*6)/(2\*1) = 21

Probability that none of the balls drawn is blue = [(5\*4)/(2\*1)]/21 = 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

Sol:

Expected no. of candies for a randomly selected child = (1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120) = 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?

Sol:

Total no. of patients = 9

Probability of one of the patients chosen at random = 1/9

Expected value = (1/9)108+(1/9)110+(1/9)123+(1/9)134+(1/9)135+(1/9)145+(1/9)167+(1/9)187+(1/9)199 = (1/9)(1308) = 145.33

Expected value of the weight of the patient is 145.33

**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**



Sol:

Based on this histogram, we can observe that the graph is positively skewed and max frequency is observed in the range of weight(50-100)



Based on this Box plot, we can observe that it has some outliers.

**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?

Sol: When mean and median of data is equal, the data is normally distributed . so, there is no skewness in the data.

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

Sol: When mean > median , the data is positively skewed or right skewed data. It means the right side of the data has some outliers.

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

Sol: When mean< median, the data is negatively skewed or left skewed data. It means the left side of the data has some outliers.

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

Sol: If the data has positive kurtosis, that means it has a sharper peak and heavier tails compared to normal distribution.

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

Sol: If the data has negative kurtosis, that means it has a flatter peak and thinner tails compared to normal distribution.

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



What can we say about the distribution of the data?

Sol: Based on the visualization, we can observe that the data is negatively skewed.

What is nature of skewness of the data?

Sol: The data is negatively skewed.

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

Sol: IQR = Q3-Q1 = 18 – 10 = 8 (approximately)

Q19) Comment on the below Boxplot visualizations?



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

Sol: Based on the visualization, the median of both boxplot 1 and boxplot 2 is same.

Both are normally distributed and there are zero outliers in both boxplot 1 and boxplot 2.

IQR of boxplot 1 = 275-250 = 25 (approximately)

IQR of boxplot 2 = 300-225 = 75 (approximately)

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