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

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

Ans : Total Outcomes = 8

P(getting two heads) = 3/8

P(getting one tail) = 3/8

P(getting two heads and one tail) = 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: The total number of possible scenarios, if two dice were rolled, is 36

1. All of the favorable outcomes (with a sum of 1) = 0

= 0

1. All of the favorable outcomes (less than equal to 4) = 6
2. All of the favorable outcomes (sum is divisible by 2 & 3) = 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?

Ans: Total case = 2+ 3+ 2 = 7

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: anticipated amount of candy for a child chosen at random

= 1\*0.015+ 4\*0.20 + 3\*0.65+ 5\*0.005+ 6\*0.01 + 2\*0.120

= 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**

Ans**:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | point | Score | Weight |
| Mean | 3.596 | 3.217 | 17.848 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.891 | 3.54 | 17.43 |
| Variance | 0.285 | 0.957 | 3.19 |
| Standard Deviation | 0.534 | 0.978 | 1.786 |
| Range | 2.76,4.93 | 1.513,5.424 | 14.5,22.9 |

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: Total Number of patients = 9

Probability of each patience is 1/9

The expected value of the weight of that patients

= 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

= 145.33

Expected value of the weight of 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**



Ans: With a frequency of 200, the majority of the data points are concentrated in the 50–100 range. The lowest weight range is 400, which is roughly 0-10.For the distribution shown above, the expected value is 75.Skewedness: There is a noticeable long tail to the right, indicating a significant right skew.



Ans: There is an outlier on the upper side of the box plot, Medican is less than mean and right skewed, and there are fewer data points between Q1 and the bottom point.

**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: X+/-(Z1- α).σ /sqrt(n)

Degree of freedom = 2000-1

=1999

Confidence interval = 94%

(1-σ/2) = 1-0.03) =0.97

For confidence interval for 94% is 1.882

confidence interval for 96% is 2.05

confidence interval for 98% is 2.33

**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: Standard Deviation = 5.05,

Variance = 25.52,

Mean = 41,

Median = 40.5

1. What can we say about the student marks?

Ans: The data is slightly skewed to the right because the mean is higher than the median, and there are no outliers.

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

Ans: There is no skewness; the distribution is perfectly symmetrical.

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

Ans: Rightward skewness and tail direction.

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

Ans: The tail and skewness point left.

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

Ans: A positive kurtosis indicates a leptokurtic curve with more peaks.

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

Ans: A curve with negative kurtosis will be flatter and broader.

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



What can we say about the distribution of the data?

Ans: The median in the above Boxplot is pointing toward the higher value; it is not normally distributed.

What is nature of skewness of the data?

Ans: There is a leftward bias in the data. The minimum value's whisker range exceeds the maximum value's.

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

Ans: The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 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: Given that the mean and median of boxplot 2 are equal and the whisker level is high, the distribution is symmetrical.

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