|  |  |
| --- | --- |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Countable |
| Blood Group | nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |

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

Ans: When three coins are tossed there are 8 outcomes.

(HHH),(HHT),(HTH),(THH),(TTT),(TTH),(THT),(HTT),

P(HHT)= 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

Ans: a) P(sum=1) = 0

b) P(sum<=4) = 6/36= 1/6

c) 6, 12 is divisible by 2 and 3 when two dice are rolled together.

(1,5),(5,1),(3,3),(2,4),(4,2),(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?

Ans: 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 number of candies for a randomly selected child is 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: 1). Points: Mean= 3.6, Median= 3.69, Mode= 3.92, Variance= 0.27, SD= 0.52, Range= 2.17

>> It is a discrete data and there are no outliers in the data.

>> Also, the mean, mode and median are almost same but the data doesn’t follow normal distribution as it’s a discrete data.

2). Score: Mean= 3.21, Median=3.33, Mode=3.44, Variance=0.92, SD=0.96, Range= 3.911

>> It is a discrete data with few outliers.

3). Weigh: Mean= 17.84, Median=17.71, Mode=17.02, Variance=3.09, SD=1.75, Range= 8.4

>> It is a continuous data.

>> Also, it is almost symmetrical with one outlier.

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 of the person would be the average of the weights of the patients that is 145.33. Therefore, the expected value would be 145 pounds.

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

**Cars speed and distance**

**Use Q9\_a.csv**

Ans: Skewness of Speed: -0.11750986144663393

Kurtosis of Speed: -0.5089944204057617

>> Speed and distance share a positive relationship.

>> Speed is almost symmetrical with a skewness of -0.11 and with zero outliers.

Skewness of Distance: 0.8068949601674215

Kurtosis of Distance: 0.4050525816795765

>> Distance is also almost symmetrical with a skewness of 0.8 with 1 outlier.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:** Skewness of SP: 1.6114501961773586

Kurtosis of SP: 2.9773289437871835

>> SP and WT share a slightly positive relationship.

Skewness of WT: -0.6147533255357768

Kurtosis of WT: 0.9502914910300326

>> SP is positively skewed with a few outliers whereas WT is slightly skewed with a few outliers.

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



Ans: >> The data in the above Histogram is positively skewed whereas the above Boxplot as a few 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?

Ans: 1). For 94% CI, Upper bound = 201.26, Lower bound= 198.73

2). For 98% CI, Upper bound = 201.56, Lower bound= 198.43

3). For 96% CI, Upper bound = 201.37, Lower bound= 198.62

**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?

Ans: 1). Mean= 41, Median= 40.5, Variance= 24.11, SD=4.91

2). Student scored an average of 41 in the tests and the marks are varied around 24 with a standard deviation of around 4.91 which means the marks scored are +4.91 or -4.91 from the mean. Also, the mean and median is almost same which means the data is almost symmetrical.

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

Ans: When the mean, median of data are equal then the data follows normal distribution which means the skewness would be zero.

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

Ans: If the mean is greater than the median then the data is positively skewed.

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

Ans: If the median is greater than the mean then the data is negatively skewed.

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

Ans: Positive Kurtosis value indicates that the data is symmetrical but the data is highly peaked.

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

Ans: Negative Kurtosis value indicates that the data is symmetrical but the data is negatively peaked.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

What will be the IQR of the data (approximately)?   
  
Ans: Data is not normally distributed.

Data is negatively skewed.

IQR is 8 approx.

Q19) Comment on the below Boxplot visualizations?



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

Ans: From the above boxplot visualization we can say that the data in boxplot 1 is not well spread as the minimum value is around 250 and the maximum value is around 280 whereas the data of boxplot 2 is well spread compared to boxplot 1 with a minimum value of around 200 and maximum value of around 330.

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) 0.40
  2. P(MPG<40)
  3. P (20<MPG<50)

Ans: 1). P(MPG>38) = 0.40

2). P(MPG<40) = 0.75

3). P (20<MPG<50) = 0.85

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: It almost follows a normal distribution with a skewness of -0.17.

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 AT and WC almost follows normal distribution with a skewness of 0.58 and 0.13 respectively.

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

Ans: 1). Z score of 90% is 1.644.

2). Z score of 94% is 1.88.

3). Z score of 60% is 0.84.

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

Ans: 1). T score of 95% is 2.063.

2). T score of 96% is 2.17.

3). T score of 99% is 2.79.

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 of randomly selected bulbs would have an average life of no more than 260 days is 0.321.