|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discreate |
| Results of rolling a dice | Discreate |
| 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 | Discreate |
| Number of kids | Discreate |
| Number of tickets in Indian railways | Discreate |
| Number of times married | Discreate |
| Gender (Male or Female) | Discreate |

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 | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Ans:- When three coins are tossed the total number of possible combinations are 2^3=8.

These combinations are HHH,HHT,HTH,THH,TTH,THT,HTT,TTT.

The number of combinations which have two heads and one tail are:

HHT,HTH,TTH which makes them 3 in number.

Therefore the probability of getting two heads and one tails in the toss of 3 coins

Simultaneously is defined as:- P(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:- a) Probability = 0

Because when two dies are rolled then least possible sum is 2.

Least possible outcomes from dice 1=1

Least possible outcomes from dice 2=1

b) There are only 6 possible outcomes and total no. of outcomes is 36.

Probability = 6/36 = 1/6

c) There are only 6 Possible outcomes and total no. of outcomes is 36.

Probability = 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:- There are 7 balls in total so the probability of first ball not being blue is 5 out of 7 and remains with 6 balls out which two are blue.

Probability of none of the balls drawn = 0.4

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:- As the probability of each candy count is given so multiplication of probability to the candies count and summation gives us the results.

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

Expected number of candies for a randomly selected child = 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?

**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:-

**HISTOGRAM**

1. The following histogram has right tail, to make the data normally distributed we have to remove the outlliers in the right.

2. The mean is greater than the median

3. The outlier will be >0(ChickWeight$weight)

4. The max weight = 200, Min weight = 10

5. The distribution is fairly symmetric

**BOX PLOT**

1. There is a positive correlation between the two variables. This means that as the value of one variable increases, the value of the other variable tends to increase as well.

2. The line of best fit is not a straight line. This suggests that the relationship between the two variables is not linear.

**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.

Ans:- Mean: 41, Median:40, Variance:25.52, Standard Deviation:5.05

1. What can we say about the student marks?

Ans:- The maximum score by any student is 56 and minimum score by any student is 34, while the average score by students is 41. We don’t have any outliners, since the mean is greater then median we don’t have skewed data.

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

Ans:- There is no skewness present we have a perfect symmetrical distribution.

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

Ans:- The nature of skewness when mean is greater than median is towards right.

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

Ans:- The nature of the skewness when median is greater then mean is towards left.

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

Ans:- The top curve is more peaked or leptokurtic.

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

Ans:- The peak is flatter and broader also the tails are heavily distributed.

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



What can we say about the distribution of the data?

Ans:- The distribution is positively distributed.

What is nature of skewness of the data?

Ans:- The data is sewed towards left, the whisker range of minimum value is greater than maximum value.

What will be the IQR of the data (approximately)?   
Ans:- IQR=Q3 uppeer quartile-Q1 lower quartile = 18-10

IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

Ans:- There are no outlines, the median of both the plot are similar as shown, that is the range from 250-275,both the data are normally distributed with zero to no skewness at the minimum or maximum whisker range.

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