|  |  |
| --- | --- |
| 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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| 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 : S : {HHH ,HHT ,HTH, THH ,HTT ,THT, TTH , TTT }

Probability of getting two heads and one tail are : 3/8 = 0.375

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

Ans : 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, }

Probability of Sum is Equal to zero is : 0

1. Less than or equal to 4

ANS : Probability of Less than or equal to 4 : {1:1,1:2,1:3,2:1,2:2,3:1}

So 6/36 =1/6 = 0.16

c)Sum is divisible by 2 and 3

Ans:The probability of getting the sum which is divisible by 2&3 is 5/36

favourable outcomes =(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)

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 no of balls = 2+3+2 = 7

Number of ways of drawing 2 balls out of 7 = (7\*6)/ (2\*1)=21

Number of balls other than blue is 5 .

Hence number of ways of drawing 2 balls out of 5 = (5\*4)/(2\*1)= 10

So ,probability that none of the balls drawn blue is =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 : Expected number of candies for a randomly selected child

=  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=       3.090

=  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 : This is Right Skewed Distribution Histogram. It’s unimodal histogram.We have concentration of data to the left side and long tail to the right .



Ans: The histogram and boxplot in Fig is positively skewed on right side.

i.e mean and median of the data is greater than mode.

**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?

Ans : Then the skewness will be 0.

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

Ans: Then the data will be right skewed . And positive skewness will be there.

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

Ans : Then the data will be left skewed . And negative skewness will be there.

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

Ans : Positive kurtosis shows the sharp peek and fat tails.

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

Ans: Negative kurtosis shows flat peek and thin tails .

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



What can we say about the distribution of the data?

Ans: The distribution in which more values are concentrated on the right side (tail) of the graph is called Negatively Skewed Distribution, while the left tail of the distribution graph is longer.

What is nature of skewness of the data?

Ans: The mean of negative skewed data will be less than the median.

What will be the IQR of the data (approximately)?   
Ans: The IQR describes the middle 50% of value when ordered from lower to highest. The interquartile range (IQR)=Q(3)- Q(1) .In above example of data the IQR =(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: Ans: The median of the two boxplots are same approximately 262.

Both follows normal distribution.

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

Ans : From probability plot or Q-Q plot ,it follows 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: From probability plot or Q-Q plot , Adipose Tissue (AT) and Waist Circumference(Waist follows Normal Distribution

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

Ans : probability that 18 randomly selected bulbs would have an average life of no more than 260 days

= 0.32

=32%