|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete Data |
| Results of rolling a dice | Discrete Data |
| Weight of a person | Continuous Data |
| Weight of Gold | Continuous Data |
| Distance between two places | Continuous Data |
| Length of a leaf | Continuous Data |
| Dog's weight | Continuous Data |
| Blue Color | Discrete Data |
| Number of kids | Discrete Data |
| Number of tickets in Indian railways | Discrete Data |
| Number of times married | Discrete Data |
| Gender (Male or Female) | Discrete Data |

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 Data |
| High School Class Ranking | Ordinal Data |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal Data |
| Socioeconomic Status | Ordinal Data |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal Data |
| Level of Agreement | Ordinal Data |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal Data |
| Time Of Day | Ordinal Data |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal Data |
| Religious Preference | Nominal Data |
| 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:

There are 3 coins, which means the total possible combination is =(2)^3=8.

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

In that case, we are getting two heads and one tail 3 times- HHT, HTH, THH. The probability of getting 2 heads and 1 tail is = (3/8) [ans]

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:

If 2 dice are rolled total possible outcomes are =(6)^2=36.

1. The probability of that sum is equal to 1 is= 0

Because whenever we roll the 2 dice the minimum sum we will get is for (1,1)=2

1. If the sum is Less than or equal to 4 we will get 6 times.

(1,1), (1,2), (1,3), (2,1), (3,1), (2,2)

The probability is =(6/36)= (1/6)

C) The Sum is divisible by 2 and 3.

If 2 dice are rolled together we know that only 6 and 12 these two sums are divisible by 2 and 3.

6 we will get 5 times-(1,5), (5,1), (2,4), (4,2), (3,3)

12 we will get only 1 time -(6,6)

the probability that Sum is divisible by 2 and 3 is - (5+1)/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 :

Total no. Of balls=7

No. Of ways drawing 2 balls out of 7 = 7C2 = (7\*6)/(2\*1) = 21 Excluding blue there are 5 balls.

No. Of ways drawing 2 balls out of 5 = 5C2 =(5\*4)/(2\*1) = 10

The probability that none of the balls drawn is blue = 5C2 / 7C2 =10/21[ans]

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 no. Of candies for randomly selected child=

1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.120 =3.09 [ans]

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?

Ans :

The total weight of 9 patients are= 108 +110+ 123+ 134+ 135+ 145+ 167+ 187+ 199 = 1308

The Expected Value of the Weight of that patient is = 1308/9=145.33[ans]

**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**



* The histogram peak has the right skew and the tail is on the right. That means we have positive skewness.
* Here mean > median.
* More than 50% chick Weight is between 50 to 150.



Ans:

1. The box plot has outliers on the higher side.

2. The data is right skewed.

**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?

Ans:

1. The mean of this data is =41 The median of this data is =40.5

Variance =25.52

Standard deviation =5.05

2)

From this data, we can see that mean > median so the data is slightly right skewed. The range of this data is 12 which means we can say that we don’t have outliers. and most of the students get numbers between 40 to 42.

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

Ans: If n the data mean and median are equal then we can say that there is no skewness in the data. the distribution of the data is symmetric.

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

Ans: If in the data mean > median then we can say that there is positive skewness That means in the distribution of data the tail is on the right side.

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

Ans : If in the data median > mean then we can say that there is negative skewness That means in the distribution of data the tail is on the left side.

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

Ans: when we have a sharp or steep peak in the distribution of the data then it’s known as positive kurtosis.

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

Ans : when we have a flatter peak in the distribution of data then it’s known as negative kurtosis.

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:

* Looking at the box plot we can say that the median is towards the higher value like 15 That’s why this is not normally distributed. 50% of data lies between 10 to 18.
* Visualizing the box plot we can say that the data is skewed towards the left which means there is negative skewness. the whisker range of minimum value is greater than the maximum
* IQR of the data= upper quartile - 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 :

The first thing we can say after visualizing the box plot is that there are no outliers. In both box plots, we have the same median approximately within the range of 250 to 275.

And they are normally distributed we can say that the distribution of data is symmetric and there is no skewness neither in the upper whisker nor in the lower whisker for both box plots.

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