|  |  |
| --- | --- |
| 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 | Ratio |
| 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: Given that,

Three coins are tossed, therefore the sample space (S) =2^3 = 8

Sample space(S) = {HHH, HHT, HTH, THH, TTT, TTH, THT, HTT}

No. of events that two heads and one tail are obtained are = n(e) = 3

Therefore, the probability of getting two heads and one tail = P = n(e)/S = 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: Given that,

Two dice are rolled, then the sample space = S = 6^2 = 36

The sample space (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)}

a): Equal to 1:

The no. of events that sum is equal to 1 is 0, as the minimum sum is equal to 2 i.e., (1,1).

Therefore, the probability that sum is equal to 1 is 0.

b): Less than or equal to 4:

The no. of events that sum is less than or equal to 4 =n(e) = 6 i.e. {(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)}.

Therefore, the probability of getting sum less than or equal to 4 is P = n(e)/S= 6/36 = 1/6

c): Sum is divisible by 2 and 3:

The no. of events that sum is divisible by 2 and 3 = n(e)= 6 i.e. {(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}.

Therefore, the probability of getting sum is divisible by 2 and 3 is P = n(e)/S = 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: Here, given that a bag contains 2 red, 3 green and 2 blue balls and two balls are drawn at random.

Therefore, the sample space = S = 7c2

The number of events that none of the balls drawn is blue = 7-2 = 5

n(e) = 5c2.

The probability that none of the balls drawn is blue = n(e)/S = 5c2/7c2 = 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 = (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**

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: Given that,

The weights(X) of patients are 108, 110, 123, 134, 135, 145, 167, 187, 199.

And given that one of the patient is chosen at random therefore, the probability is P = 1/9.

The Expected value of the weight of that patient = 1/9\*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(108+110+123+134+135+145+167+187+199)

=1/9(1308)

Therefore, the expected of that patient = 145.333.

**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: It is given that,

🡪In histogram, the distribution of data is asymmetric.

And also it is right skewed and we can say it is a positively skewed distribution of data.

🡪In boxplot, the distance from the median to maximum length is higher than the

distance from the median to minimum length.

And we can say it as a positively skewed distribution of data.

**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: The data is normally distributed and it is symmetric in nature.

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

Ans: The data is right skewed and it is positively skewed in nature.

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

Ans: The data is left skewed and it is negatively skewed in nature.

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

Ans: Positive kurtosis value indicates that there is more no. of outliers in the data. And also there are heavier tails than the normal distribution.

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

Ans: Negative kurtosis value indicates that there is less no. of outliers or infrequent outliers present in the data. And also there are lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

Ans: The distribution of the data is asymmetric.

What is nature of skewness of the data?

Ans: The nature of skewness of the data is negatively skewed.

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

Ans: Q3 = 18,

Q1 = 10

Inter Quartile Range (IQR) = Q3 – Q1

= 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: Above boxplot visualizations present over there shows, normally distributed data. There are no outliers present in the data. They are symmetric in nature where median is 260.

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