|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete -- Countable |
| 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-- Countable |
| Number of tickets in Indian railways | Discrete-- Countable |
| Number of times married | Discrete-- Countable |
| Gender (Male or Female) | Discrete – Classification – Nominal |
|  |  |

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

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

Since three coins are tossed, hence total combinations are 23

23 = 8.

The possible combinations of getting two heads out of total 8 combinations are HHT, HTH, TTH.

Therefore, the probability that two heads and one tail is 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

Total possible outcomes = 6\*6 = 36

1. For sum equal to 1 when two dice are rolled, one has to be ‘1’ and the other has to be ‘0’, which is not possible.

Hence, probability for having sum equal to 1 when two dice are rolled is zero.

Probability of getting sum of 1 = 0/36 = 0

1. There are total 6 ways in which the sum is less than or equal to 4.

(1,1), (1,2), (1,3), (2,1), (2,2), (3,1). So, the probability of getting sum less than equal to 4 is;6/36 = 1/6.

1. For having sum divisible by 2 or 3, we have 5 favorable outcomes;

(1,5) (3,3) (4,2) (5,1) (6,6)

Therefore, probability for getting sum divisible by 2 or 3 is;

5/36.

**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?

Total number of balls = 2Red + 3Green + 2Blue = 7 Balls.

Now, Total number of ways of drqawing 2 balls out of 7 is 7C2.

7C2 = (7\*6)/(2\*1) = 21.

Now, the Probability that none of the balls drawn is blue. i.e., either Red or Green .

Red 🡪 2Balls and Green 🡪 3Balls.

Therefore, the total number of ways of drawing 2 balls out other than Blue is 5C2.

5C2 = (5\*4)/(2\*1) = 10.

Therefore, the probability that none of the balls drawn is blue is:

[5C2 = (5\*4)/(2\*1) = 10] / [7C2 = (7\*6)/(2\*1) = 21] = 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

Expected number of candies = Sum of (Candies count x Probability)

* 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?

**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 data is right skewed or positively skewed
* More than 50% Chick Weight is between 50 to 150.
* Most of the chick weight is between 50 to 100.

**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?

If Mean = Median = Mode, then the distribution is both symmetric and unimodal.

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

If the mean is greater than median, then the nature of skewness is positive, or the data will be right skewed. Most of the data will be lying on the left side of the plot. Mean always tries to go or say tends to go towards the most skewed part since skewness influences the mean.

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

If the mean is less than median, then the nature of skewness is negative, or the data will be left skewed. Most of the data will be lying on the right side of the plot. Mean always tends to go towards the most skewed part since skewness influence the mean.

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

Kurtosis is the measure of the Peakness of the distribution. And positive value of Kurtosis indicates that the distribution is peaked and possess thick tails than the normal distribution.

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

Negative kurtosis value indicates that the distribution is flat and possess thin 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?

Median of the data is approx. 14.7

25% of the data lies between 0-10

50% of the data lies between 10-18

And the rest 25% of the data lies after 18(or to 20)

What is nature of skewness of the data?

As the whisker length on the upper quadrant is higher than the lower quadrant, the data is Negative Skewed, or say Left skewed.

What will be the IQR of the data (approximately)?   
 Inter Quartile Range(IQR) :

Q1 = 10

Q2 = 14.7 (approx.)

Q3 = 18

IQR = Q3 – Q1 = 18-10 = 8 (approx.)  
  
Q19) Comment on the below Boxplot visualizations?



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

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