|  |  |
| --- | --- |
| 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 | CATEGORICAL |
| Number of kids | DISCRETE |
| Number of tickets in Indian railways | DISCRETE |
| Number of times married | DISCRETE |
| Gender (Male or Female) | CATEGORICAL |

**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) | RATIO |
| Sales Figures | RATIO |
| Blood Group | NOMINAL |
| Time Of Day | NOMINAL |
| Time on a Clock with Hands | INTERVAL |
| Number of Children | RATIO |
| Religious Preference | NOMINAL |
| Barometer Pressure | RATIO |
| SAT Scores | ORDINAL |
| Years of Education | ORDINAL |

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

**Solution**:

Let S be the sample space

S= {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} = 8.

Two heads and One tail Outcome= {HHT, HTH, THH} =3.

Probability= 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**

**Solution:**

1. Total outcomes = 6\*6 = 36.

Sum of two dice can never be 1, So probability = 0.

1. Possible outcomes = {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)} = 6

So, probability = 6/36 = 0.167.

1. Possible outcomes = 2,3,4,6,8,9,10,12,14,15,16,18,20,21,22,24,26,27,28,30,32,33,34,36.

So, probability = 24/36 = 0.67.

**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?**

**Solution:**

Total No. of balls = (2+3+2) = 7.

But we know that 2 balls are drawn random So, the sample space =

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

Let E the probability that none of the balls drawn is blue.

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

P(E) = n(E)/n(S) = 10/21 = 0.48.

**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**

**Solution:**

Expected no. of candies for randomly selected child

Hence, we know that

Expected Value = sum of (Candies count \* Probability)

= (1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.001 + 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?**

**Solution:** Expected weight = sum of(weights\*probabilities)

There are nine patients, so the probability of selecting each patient

E(x) = 108,110,123,134,135,145,167,187,199

P(x) = 1/9,1/9,1/9,1/9,1/9,1/9,1/9,1/9,1/9.

Expected weight= 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)

Expected weight = 1/9(1308) = 145.33~ 145 pounds.

**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**



**Solution: -** 1) The mode of the data is between 50-100(as we seen the highest peak between 50-100)

2) The median of the data is approx. 200.

3) As mode is less than median(mode<median) so the above distribution Is positively skewed.



**Solution:**  The Boxplot is positively skewed (right skewed) because the upper quartile is greater than lower quartile and the upper whisker is long tailed as compared to lower whisker.

**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?**

**Solutions:** 1) Average marks obtained was 41 with the spread around 5.

2) The median is 40.5(50%) of students scored below 40.5 and rest 50% scored above 40.5.

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

**Solution:** The skewness is 0.

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

**Solution:** The Distribution is positively skewed.

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

**Solution:** The distribution is negatively skewed.

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

**Solution:** A Distribution with a positive kurtosis value indicates that the distribution has heavier tails than the normal distribution.

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

**Solution:** A Distribution with a negative kurtosis value indicates that the distribution has heavier 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?**

**Solution:** The data has a range from approx. 1 to 19. There are no outliers present in the data.

**What is nature of skewness of the data?**

**Solution:** The nature of skewness of respective boxplot is left skewed.

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

**Solution:** q1(app.) =10, q3(app.) =18, so IQR=q3-q1=8(app.)

Q19) Comment on the below Boxplot visualizations?



**Solution:** 1) Data 1 and Data 2 both have the same median values (~262.5)

1. Both datasets are symmetrically distributed as evident from the nature of their boxplots.
2. The variation in dataset 1 is less compared to dataset 2. Dataset 1 is spread over smaller range compared to dataset 2 whose range is wider.

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