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

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

Solution: When 3 coins are tossed, possible combinations are 2^3=8, they are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT where we have 3 such combinations we will get two heads & one tail i.e., HHT,HTH,THH   
So, probability 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 2and 3**

Solution:

If two dice are rolled, possible combinations are {(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)}

1. Equal to 1 = 0
2. Less than or equal to 4 = 6/36=1/6=0.16=16.76%
3. Sum is divisible by 2 and 3 = (1,5),(2,4),(3,3),(4,2),(5,1),(6,6)

Possible ways are 6 so, 6/36=1/6=0.16=16.76%

**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 balls = (2+3+2) =7

No, of ways of drawing 2 balls out of 7 = 7C2 = (7\*6)/(2\*1) = 21

When drawing 2 balls none is blue(2+3) = 5C2=(5\*4)/(2\*1)=10

Therefore, probability= 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**

Solution: 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

=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**

Solution:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | 3.597 | 3.217 | 17.849 |
| **Median** | 3.695 | 3.325 | 17.710 |
| **Mode** | 3.920 | 3.440 | 17.020 |
| **Variance** | 0.286 | 0.957 | 3.193 |
| **Std Dev** | 0.535 | 0.978 | 1.787 |
| **Max** | 4.930 | 5.424 | 22.900 |
| **Min** | 2.760 | 1.513 | 14.500 |
| **Range** | 2.170 | 3.911 | 8.400 |
|  |  |  |  |

**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 value = (108+110+123+134+135+145+167+187+199)/9

=145.3333

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | -0.11751 | -0.50899 |
| **Dist** | 0.806895 | 0.405053 |

**SP and Weight(WT) Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **SP** | 1.6114502 | 2.977329 |
| **WT** | -0.6147533 | 0.950291 |

**Q10) Draw inferences about the following boxplot & histogram**



Solution: The data has large side on the right & tail on the left side, this is called as Right Skewed Data/ Positively Skewed data. i.e., Mean>Median>Mode

Solution: This Boxplot has Outliers on the Maximum side



**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?**

Solution:

For 94% confidence interval Range is 198.73-201.26  
For 98% confidence interval Range is 198.43-201.56  
For 96% confidence interval Range is 198.62-201.37

**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?**

Solution: Mean = 41  
 Median = 40.5  
 Mode = 41  
 Variance = 25.52  
 Std Dev = 5.052

Mean> Median so data will slightly skewed towards right.

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

Solution: No skewness is present we have perfect symmetrical distribution

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

Solution: If the mean is greater than median, then distribution is positively skewed.

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

Solution: If the median is greater than mean, distribution is negative skewed.

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

Solution: Positive kurtosis values indicates that the curve is more peaked & it is leptokurtic

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

Solution: Negative kurtosis value indicates that the curve will be flatter & broader.

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



**What can we say about the distribution of the data?**

Answer: The boxplot is not normally distributed as the median is towards the higher value.

**What is nature of skewness of the data?**Answer: The data is skewed towards left as the whisker range of minimum value is greater than maximum.

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

Answer: The Inter Quartile Range = 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.**

Answer:

1)No outliers

2)The boxplot shares the same median which ranges between 250-275

3)They are normally distributed with 0 to no skewness neither at the minimum or maximum whisker range.

**Q 20) Calculate probability from the given dataset for the below cases**

**Data \_set: Cars.csv**

**Calculate the probability of MPG ofCars for the below cases.**

**MPG<- Cars$MPG**

* 1. **P(MPG>38)**
  2. **P(MPG<40)**

**c. P (20<MPG<50)**

Solution:

1. P(MPG>38)**= 0.348**
2. P(MPG<40)**= 0.729**
3. P(MPG>20) **= 0.943**
4. P(MPG<50) **= 0.956**
5. P (20<MPG<50)**= 0.013000000000000012**

**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**Answer: From plot & values we can say that data is fairly symmetrical, that means it is normally distributed.

1. **Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution**

**Dataset: wc-at.csv**

Answer: From plot & values we can say that data does not follow normal distribution.

**Q 22) Calculate the Z scoresof 90% confidence interval,94% confidence interval, 60% confidence interval**

|  |  |
| --- | --- |
| CONFIDENCE INTERVAL | Z-SCORES |
| 90% | 1.644 |
| 94% | 1.880 |
| 60% | 0.841 |

**Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25**

Solution:

|  |  |
| --- | --- |
| CONFIDENCE INTERVAL | T-SCORES |
| 95% | 2.063 |
| 96% | 2.171 |
| 99% | 2.796 |

**Q 24)A Government companyclaims 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**

**Solution:**

\*Population mean = 270 days  
\*Sample mean = 260 days  
\*Sample Std Dev = 90 days  
\*Sample(n) = 18 bulbs  
\*Degree of freedom = n-1 = 18-1 = 17  
\*t-score = -0.471  
\*0.32%