|  |  |
| --- | --- |
| 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 | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | 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 | 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: The probability of getting two heads and one tail in the toss of three coins

Simultaneously is 3/8 or 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**

Ans: a) The probability of sum is equal to 1 is zero because they starts

with (1,1) .

b) The possible outcomes for sum less than or equal to 4 are

(1,3),(2,2),(3,1) therefore n(b) = 3/36 = 1/12.

c) The probability that sum is divisible by 2 and 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?**

Ans: The probability that none of the balls drawn is blue is 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 is 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**

Ans**:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | Std. Dev | Range |
| Points | 3.59 | 3.70 | 3.92 | 0.29 | 0.53 | 2.17 |
| Score | 3.22 | 3.33 | 3.44 | 0.96 | 0.98 | 3.91 |
| Weigh | 17.85 | 17.71 | 17.02 | 3.19 | 1.79 | 8.40 |

* No case has the variable Mean = Median = Mode
* Thus as seen in the graph dataset ‘score’ and ‘Weigh’ has outliers

**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: there are 9 patients,

Probability of selecting each patients is 1/9.

Expected value =(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).

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Ans**:**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| speed | -0.11 | 2.42 |
| dist | 0.76 | 3.25 |

* “dist” is positively skewed where as “speed” is negatively skewed
  + Thus dist has distribution of data concentrated on the left whereas speed has distribution on the right. As seen in the graph
* Both dist and speed has positive Kurtosis

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| SP | 1.55 | 5.72 |
| WT | -0.59 | 3.82 |

* “SP” is positively skewed where as “WT” is negatively skewed
  + Thus SP has distribution of data concentrated on the left whereas WT has distribution on the right. As seen in the graph
* Both WT and SP has positive Kurtosis

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



Ans:

1. Majority of the Chicks has weight in range 50 to 100.
2. The data is Right Skewed or positively Skewed.
3. 3 Categories can be define

\*Under weight >50.

\*Avg weight 51 to 150.

\* Over weight <150.



Ans. =

1. Data has outliers
2. Data is right skewed or positively 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?**

Ans:

|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| Upper | 201.04 | 201.38 | 201.17 |
| Lower | 198.96 | 198.62 | 198.83 |

**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.**

Ans:

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.50 |
| Variance | 25.53 |
| Std Dev | 5.05 |

1. **What can we say about the student marks?**

Ans:

1. Data is not normally distributed.
2. Majority of the students scored between 35 to 45 Marks

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

Ans: Skewness = 0.

Perfectly symmetric bell shaped curve.

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

Ans: Skewness = Positive.

Data is distributed more on left.

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

Ans: Skewness = Negative.

Data is distributed more on right.

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

Ans: High and narrow peak on central part of the data

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

Ans: wider peak on central part of the data

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



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

Ans: The data is not symmetric, Data is more concentrated towards right side.

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

Ans: Skewness is Negative.

**What will be the IQR of the data (approximately)?**Ans: IQR data is 8 (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.**

\* Data in the first plot is Normally Distributed.

No Outliers.

Center around 262.5.

First plot has less range compared to second.

\* Data in second plot is Normally Distributed.

No Outliers.

Center around 262.5.

Second graph has more range compared to first.

**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)**

Ans: 0.4074074

* 1. **P(MPG<40)**

Ans: 0.7530864

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

Ans: 0.8518519

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

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

**Dataset: Cars.csv**

Ans: MPG is Normal Distribution.

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

**Dataset: wc-at.csv**

Ans: Both AT and Waist doesn’t follow Normal Distribution.

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

Ans:

|  |  |
| --- | --- |
| 90% | ± 1.711 |
| 94% | ± 1.828 |
| 60% | ± 2.492 |

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

Ans:

|  |  |
| --- | --- |
| 95% | ± 2.060 |
| 96% | ± 2.167 |
| 99% | ± 2.787 |

**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**

Ans: t-score = -0.4717.

Degree of freedom = 17.

P(t) = 0.3216725.