|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Nominal |
| SAT Scores | Ordinal |
| Years of Education | Interval |

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

**Ans : 3/8**

Q4) Two Dice are rolled, find the probability that sum is

1. **Equal to 1 Ans : 0**
2. **Less than or equal to 4 Ans : 1/6**
3. **Sum is divisible by 2 and 3 Ans : 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 : 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: E(X) = Candies count \* Probability = 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.

**ANS :**

* **Points –**

**Mean : 3.596563 (Average points given to the cars is 3.596563)**

**Median : 3.695**

**Mode: 3.92**

**Variance : 0.276948**

**Standard Deviation : 0.526258**

**Range : 2.17**

* **Score –**

**Mean : 3.21725**

**Median : 3.325**

**Mode : 3.44**

**Variance : 0.927461**

**Standard Deviation : 0.963048**

**Range : 3.91**

* **Weight –**

**Mean : 17.84875**

**Median : 17.71**

**Mode : 17.02**

**Variance : 3.09338**

**Standard Deviation : 1.758801**

**Range : 8.40**

**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 : 145.3333 pounds**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans :**

**Speed :**

**Skewness : -0.11395**

**Kurtosis : -0.50899**

**The skewness for speed is negative. Therefore we can conclude that the data is negatively skewed. Kurtosis value is less than 3. This implies that the given data of speed is Platykurtic in nature (Flatter than the Normal Curve)**

**Distance :**

**Skewness : 0.782484**

**Kurtosis : 0.405053**

**Skewness for distance is positive. Therefore, we can conclude that the data is positively skewed. Kurtosis value is less than 3. This implies that the given data of speed is Platykurtic in nature (Flatter than the Normal Curve)**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans :**

**SP :**

**Skewness : 1.581454**

**Kurtosis : 2.977329**

**The skewness value is positive, implies that the data for SP is positively skewed. Kurtosis value is just less than 3. The data is Platykurtic in nature. (Very close to Normal curve, slightly flatter than the normal curve.)**

**Weight :**

**Skewness : -0.60331**

**Kurtosis : 0.950291**

**The given data of weights is negatively skewed. Kurtosis < 3, implies that the curve is Platykurtic in nature. ( Very flatter than the normal curve.)**

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



**Ans :**

**The above Histogram of ChickWeight$weight is Positively Skewed. The data is not equally spread on the both sides of Median. Also we can observe peak of data between weights 50 units and 100 units. The Histogram is peaked (Leptokurtic in nature).**



**ANS :**

**Form above boxplot, we can observe that the data is asymmetric in nature. The upper whisker is bigger than the lover whisker. Outliers can be observed beyond the upper extreme. The data seems to be 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 : 40.5**

**Mode : 41**

**Variance : 64.86111**

**S.D : 8.053639612**

1. What can we say about the student marks?

**Ans :**

**The data is Positively Skewed, hence we can say that most of the students secured very less marks. The peak of the data between range 36 and 42. Most of the students scored between range 36 and 42.**

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

**Ans: The data is symmetric. Normal distribution curve.**

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

**Ans: Positively Skewed**

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

**Ans: Negatively Skewed**

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

**Ans : The data is Platykurtic in nature if Kurtosis is < 3. Flatter than normal distribution or equal to normal distribution if kurtosis = 3. Leptokurtic in nature if kurtosis > 3.**

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

**Ans : The data is highly Platykurtic. The curve is more flatter.**

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



What can we say about the distribution of the data?

**Ans: The data is asymmetric. Most part of the data is distributed on the left side of the median.**

What is nature of skewness of the data?

**Ans : The data is negatively skewed.**

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

**Ans : 8 units**

Q19) Comment on the below Boxplot visualizations?



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

**Ans: Both the datasets are normally distributed. Boxplot 1 has lesser inter-quartile range as compared to Boxplot 2. This implies that, the data for Boxplot 1 is not widely spread. The frequency of the data points is maximum between range 250-275. There are no outliers.**

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) = 0.4658**
  2. **P(MPG<40) = 0.5532**

**c. P (20<MPG<50) = 0.2809**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans: The data seems to be 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

**Ans: The data seems to be Normally Distributed.**

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

**Ans: 90% = 1.645**

**94% = 1.555**

**60% = 0.254**

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

**Ans: 95% = 1.711**

**96% = 1.828**

**99% = 2.492**

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 : -0.4814

p(t) : 0.3216725

DF : 17