|  |  |
| --- | --- |
| 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 | **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) | **Ratio** |
| Sales Figures | **Ratio** |
| Blood Group | **Nominal** |
| Time Of Day | **Interval** |
| Time on a Clock with Hands | **Interval** |
| Number of Children | **Nominal** |
| 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. 3:8**

**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.) 0**

**B.) 1/6**

**C.) 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 = 0.476**

**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

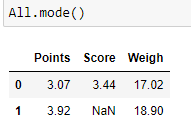
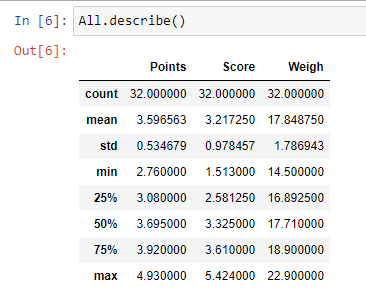
**Ans. 3.09**

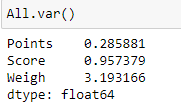
**Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset**

* 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.**



**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

**Ans. x̄ = 145.333**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans. Skewness of Speed and Distance is: -0.12, 0.81**

**Kurtosis of Speed and Distance is: 0.51 ,0.41**

**Skewness of both the features are different as speed is negatively skewed whereas distance is positively skewed**

**Kurtosis of both the features are positive**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness of Speed and Weight is: 1.61, -0.61**

**Kurtosis of Speed and Weight is: 2.98 ,0.95**

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



**Ans. A.) The data is positively skewed**

**B.) The frequency in range between 0-50,50-100,100-150,150-200**



**Ans. The above boxplot has too many outliers and is 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.** **A.) Average weight of adult in Mexico at 94% confidence interval [134.897 265.103]**

**B.) Average weight of adult in Mexico at 98% confidence interval [122.725 277.275]**

**C.) Average weight of adult in Mexico at 96% confidence interval [130.21 269.79]**

**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?

**Ans. A.) Mean is 41.0**

**Median is 40.5**

**Standard Deviation is 5.06**

**Variance is 25.53**

**B.) Data is not normally distributed, maximum marks are between 35-40**

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

**Ans. 0**

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

**Ans. Positive**

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

**Ans. Negative**

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

**Ans. Sharp peak less variation**

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

**Ans. Less peak high variation**

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



What can we say about the distribution of the data?

**Ans. It is not normally distributed**

What is nature of skewness of the data?

**Ans. Left skewed**

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

**Ans. 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.

**Ans. A.) There are no outliers.**

**B.) Data are normally distributed in both the boxplots.**

**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. Probability that 'MPG' > 38 = 0.348**

* 1. P(MPG<40)

**Ans. Probability that 'MPG' < 40 = 0.729**

* 1. P (20<MPG<50)

**Ans. Probability that 'MPG' < 40 = 0.84134**

**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. 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. A.) Z scores at 90% confidence interval is 1.64**

**B.) Z scores at 94% confidence interval is 1.88**

**C.) Z scores at 60% confidence interval is 0.84**

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

**Ans. A.) t scores at 95% confidence interval is 2.06**

**B.) t scores at 96% confidence interval is 2.17**

**C.) t scores at 99% confidence interval is 2.8**

**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. critical value = -0.47**

**probability for average life of no more than 260 days is 0.32**