Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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** |

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 | **Ordinal** |
| Religious Preference | **Nominal** |
| Barometer Pressure | **Ratio** |
| 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 that two heads and one tail -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

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**

**5C2 / 7C2**

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: 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**

**= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24**

**= 3.090**

**= 3.09**

**Expected number of candies for a randomly selected child = 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**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr No** | **Datasets**  **Values** | **Points** | **Score** | **Weigh** |
| **1** | **Mean** | **3.596563** | **3.21725** | **17.84875** |
| **2** | **Median** | **3.695** | **3.325** | **17.71** |
| **3** | **Mode** | **numeric** | **numeric** | **numeric** |
| **4** | **Variance** | **0.2858814** | **0.957379** | **3.193166** |
| **5** | **Standard Deviation** | **0.5346787** | **0.9784574** | **1.786943** |
| **6** | **Range** | **2.76 - 4.93** | **1.513 - 5.424** | **14.5 - 22.9** |

**Inferences: a) For Points data set:**

**1) The data is concentrated around Median**

**2) There are no outliers**

**3) The distribution is Right skewed**

**b) For Score data set:**

**1) The data is concentrated around Median**

**2) There are 3 Outliers: 5.250, 5.424, 5.345**

**3) The distribution is Left skewed**

**c) For Weigh data set:**

**1) The data is concentrated around Median**

**2) There is 1 Outlier: 22.90**

**3) The distribution is Left skewed**

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: Expected Value  =  ∑ ( probability  \* Value )

∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

Ex 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 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

= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) ( 1308)

= 145.33

Expected Value of the Weight of that patient = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans :Skewness of speed -0.1139548**

**Kurtosis of Speed 2.422853**

**Skewness of distance 0.7824835**

**Kurtosis of distance 3.248019**

**Cars speed is negative skewed & distance is positive skewed.**

**Both Speed and distance are not having outliers.**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:****Skewness of speed 1.581454**

**Kurtosis of Speed 5.723521**

**Skewness of weight -0.6033099**

**Kurtosis of weight 3.819466**

**Cars speed is positive skewed & distance is negative skewed.**

**Speed is having outliers at upper extreme**

**weight is having outliers at both lower & upper extrem**

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





**Ans: This is positively skewed chicks weight distribution & having positive whisker & outlaiers.**

**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: import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

**from** **scipy** **import** stats

**from** **scipy.stats** **import** norm

**# Avg. weight of Adult in Mexico with 94% CI**

stats.norm.interval(0.94,200,30/(2000\*\*0.5))

**(198.738325292158, 201.261674707842)**

**# Avg. weight of Adult in Mexico with 98% CI**

stats.norm.interval(0.98,200,30/(2000\*\*0.5))

**(198.43943840429978, 201.56056159570022)**

**# Avg. weight of Adult in Mexico with 96% CI**

stats.norm.interval(0.96,200,30/(2000\*\*0.5))

**(198.62230334813333, 201.37769665186667)**

**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 :**

|  |  |  |
| --- | --- | --- |
| **Sr No** | **Parameter** | **Values of Dataset** |
| **1** | **mean** | **41** |
| **2** | **median** | **40.5** |
| **3** | **variance** | **25.52941** |
| **4** | **Standard deviation** | **5.052664** |

1. What can we say about the student marks?

**Ans: 1 Maximum students have scored 35 to 40 marks in the test**

**2 No students have scored 50 to 55 marks in the test.**

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

**Ans: In perfectly symmetrical distribution mean & median of data are equal &skewness is zero**

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

**Ans: Distribution of data is skewed to the right side (Positively skewed)**

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

**Ans: Distribution of data is skewed to the left side (negatively Skewed)**

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

**Ans: Positive value of Kurtosis indicate that the distribution is peaked and possess thick tail**

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

**Ans: Negative value of Kurtosis indicate that the distribution is flat and has thin tail.**

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



What can we say about the distribution of the data?

**Ans:1 Maximum data is distributed Between IQR**

**2 No data in outlaiers**

What is nature of skewness of the data?

**Ans:** **Data is skewed to the left**

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

**Ans:** **IQR(Inter Quartile Range) = 5**Q19) Comment on the below Boxplot visualizations?



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

**Ans: 1 Both the Data is equally distributed**

**2 Both the dataset is having same Median**

**3 Both are having symmetric distribution.**

**4 Skewness is 0**

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

**Ans:** P(MPG>38) = **0.34749**

P(MPG<40)=**0.72946**

P (20<MPG<50) =**0.01316**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans : No,this is not normal distribution.the data is negative skewed.**

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

Dataset: wc-at.csv

**Ans : No, the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set does not follows Normal Distribution both are positively skewed**

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

**Ans:**

|  |  |  |
| --- | --- | --- |
| **Sr No** | **Confidence Interval** | **Z Score** |
| **1** | **90%** | **1.645** |
| **2** | **94%** | **1.89** |
| **3** | **60%** | **0.85** |

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

**Ans:**

|  |  |  |
| --- | --- | --- |
| **Sr No** | **Conference Interval** | **T Score** |
| **1** | **95%** | **-2.063899** |
| **2** | **96%** | **-2.171545** |
| **3** | **99%** | **-2.79694** |

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 - statistics for the data is given as follows:**

**x = mean of the sample of bulbs = 260**

**μ = population mean = 270**

**s = standard deviation of the sample = 90**

**n = number of items in the sample = 18**









**t = - 0.471**

**For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.**

**The probability that t < - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.**