|  |  |
| --- | --- |
| 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 | **Nominal** |
| 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 | **Ordinal** |

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

**Possible events:**

**HHH**

**HHT**

**HTH**

**HTT**

**THH**

**THT**

**TTH**

**TTT**

**Number with 2 heads and 1 tail: 3**

**Total number: 8**

**Probability = 3/8 = 0.375 = 37.5%**

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

**a) Total Outcomes=6\*6=36  
 equal to1=0  
 0/36=0**

**b) Less than or equal to 4=6  
 6/36=1/6**

**c)Sum is divisible by 2 and 3=6**

**6/36=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?

**Total balls = 7**

**2 balls out of 7=7c2=21**

**None is blue=5c2=10  
P=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

**Expected candies = 1\*0.015 + 4\*0.20 + 3\*0.65 + 5\* 0.005 + 6\* 0.01 + 2\* 0.120**

**= 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**

**Attached ipynb file.  
File Name= Assignments\_1**

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?

**Total no of patients=9  
P(N)=1/9  
P(S)= 108, 110, 123, 134, 135, 145, 167, 187, 199**

**Expected Value= ∑ (P(N)\*P(S))**

**(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**

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

Cars speed and distance

**Use Q9\_a.csv**

**Attached ipynb file.  
File Name= Assignments\_1**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Attached ipynb file.  
File Name= Assignments\_1**

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



**Histogram is positively skewed**

**Box plot has many outliers**

**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?

**Given= N=2000, SD=30, Avg=200**

**The interval is**

**x+t s ^n= 200 – 1.8916 30^2000 =198.73**

**x+ t s ^n= 200+ 1.816 30 ^2000=201.27**

**The 94% confidence interval is (198.73,201.27)**

**x-ts ^n=200 – 2.0673 30 ^2000=198.61**

**x+ t s^n=200 + 2.0673 30-^2000201.39**

**The 96% confidence interval Is (198.61,201,39).**

**x-ts^n =200 -2.3452 30 =198.43 30^2000**

**x+ ts^n 200 + 2.3452 30 = 201.57^2000**

**The 98% confidence interval is (198.43,201.57)**

**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.

**Mean= 41**

**Median=40.5**

**Variance=24.1**

**S.D=4.91**

1. What can we say about the student marks?

**Avg of class marks is 41**

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

**No skewness**

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

**Positive skewness**

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

**Negative skewness**

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

**Heavy tailed distribution, Thinner peak**

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

**Light tailed distribution, Flatter peak**

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



What can we say about the distribution of the data?

**Not Normally distributed**

What is nature of skewness of the data?

**Negative skewed**

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

**Q1=11, Q2=14, Q3=12  
 IQR=Q1-Q3=6**

Q19) Comment on the below Boxplot visualizations?



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

**Median is same so distribution is symmetrical**

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)
  2. P(MPG<40)
  3. P (20<MPG<50)

**Attached ipynb file.  
File Name= Assignments\_1**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Attached ipynb file.  
File Name= Assignments\_1**

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

Dataset: wc-at.csv

**Attached ipynb file.  
File Name= Assignments\_1**

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

**Attached ipynb file.  
File Name= Assignments\_1**

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

**Attached ipynb file.  
File Name= Assignments\_1**

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

**Attached ipynb file.  
File Name= Assignments\_1**

**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**