| Activity | Data Type |
| --- | --- |
| Number of beatings from Wife | **Continuous** |
| 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 | **Ratio** |
| Socioeconomic Status | **Interval** |
| Fahrenheit Temperature | **Interval** |
| Height | **Ratio** |
| Type of living accommodation | **Ordinal** |
| Level of Agreement | **Interval** |
| IQ(Intelligence Scale) | **Interval** |
| Sales Figures | **Ordinal** |
| Blood Group | **Ordinal** |
| Time Of Day | **Interval** |
| Time on a Clock with Hands | **Interval** |
| Number of Children | **Nominal** |
| Religious Preference | **Ratio** |
| Barometer Pressure | **Interval** |
| SAT Scores | **Ratio** |
| Years of Education | **Ordinal** |

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

**Ans:- Total no of coins tossed = 8**

**possibility of two heads & one tail is =3 {HHT,HTH,THH}**

**therefore it will be [ 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**

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:-** **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**

**Ans: (ipynb file is attached )**

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: **(108+110+123+134+135+145+167+187+199)/9=145.3333**

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

**Cars speed and distance**

**SP and Weight(WT)**

**Use Q9\_a.csv**

**Use Q9\_b.csv**

**Ans: (ipynb file is attached )**

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



**Ans: The histogram peak has a right skew and tail is on right. Mean > Median. We have outliers on the higher side.**

**Ans: The boxplot has outliers on the maximum side.**



**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: (ipynb file is attached )**

**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: (ipynb file is attached )**

1. What can we say about the student marks?

**Ans:- We don’t have outliers and the data is slightly skewed towards the right because the mean is greater than the median.**

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

**Ans:- There will be zero skewness**

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

**Ans:- It will be positively skewed distribution**

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

**Ans:- It will be negatively skewed distribution**

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

**Ans:- It indicated the heavier tails and more peaked distribution**

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

**Ans:- It indicated the lighted tails and flat distribution**

Q18) Answer the below questions using the below box plot visualization.



What can we say about the distribution of the data?

**Ans:- The distribution of data is asymmetrical.**

What is nature of skewness of the data?

**Ans: the distribution of data is more skewed on the left side , therefore mean < median .**

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

**Ans: IQR=Q3-Q1 =18-10=8**

**here Q3= 18 & Q1 = 10**

Q19) Comment on the below Box Plot visualizations?



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

**Ans:- Both are having median on the center**

**There are no outliers on both the boxplot .**

**There are no skewness on both the boxplot**

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)

c. P (20<MPG<50)

**Ans: (ipynb file is attached )**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

**Ans:** **(ipynb file is attached )**

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

**Ans: using z-table with C.I**

**At 90% confidence interval :- 1.645**

**At 94% confidence interval :- 1.88**

**At 60% confidence interval :- 0.841**

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

**Ans: using t-table with C.I**

**At 95% Confidence interval :- 2.064**

**At 96% Confidence interval :- 2.17**

**At 99% Confidence interval :- 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