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

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

**Sol**: P (H H T) + P(H T H) + P(T H H)

=1/8 + 1/8 + 1/8

= 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

**Sol**: (a) Sum is Equal to 1

= 0

(b) Sum is less than equal to 4

= 1/6

(c) Sum is Divisible by 2&3

=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?

**Sol**: 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

**Sol:** Expected number of candies for randomly selected child

=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**

**Sol:** PointMean=3.596563, Median= 3.695, Mode=NA

Variance=0.2858814, Standard Deviation=0.53467

Score Mean=3.21725, Median=3.325, Mode=NA

Variance=0.957379, Standard Deviation=0.97845

Weight Mean=17.84847, Median=17.71, Mode=NA

Variance=3.19316, Standard Deviation=1.786943

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?

**Sol:** 145.33

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

Car’s speed and distance

Use Q9\_a.csv

**Sol:** Skewness for speed = -0.11751

Skew value is negative, so it is left skewed

Kurtosis for speed = -0.50899

Skewness for Distance = 0.806895

Skew value is positive, so it is right skewed

Kurtosis for distance = 0.405053

SP and Weight (WT)

Use Q9\_b.csv

**Sol:** Skewness for SP =1.61145

Kurtosis for SP =2.977329

Skewness for Weight =-0.61475

Kurtosis for Weight =0.950291

Q10) Draw inferences about the following boxplot & histogram



**Sol:** The most data points are in range of 50-100 with frequency 200.

And at least range of weight is 400 somewhere in range of 0-10.

So expected value of distribution is 75.

We can notice long tail towards right, so it is heavily right 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?

**Sol:** Confidence interval for 94% =1.882

Confidence interval for 96% =2.33

Confidence interval for 98% =2.05

**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?

**Sol:(1)** Mean = 41, Median =40, Variance = 24.11, Standard Deviation = 4.910

**Sol:(2)**from above that we can say that mean of marks of student is 41 which is slightly greater than median.

Most of the students got marks in between 41-42, there are two outlier 49,56.

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

**Sol:** Symmetrical

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

**Sol:** Right Skewed

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

**Sol:** Left Skewed

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

**Sol:** Normally distributed

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

**Sol:** The distribution has lighter tail than normal distribution.

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



What can we say about the distribution of the data?

Sol: Assume that The box plot given above is about age of students in school

50% of student above 10 years old and remaining are less.

Student age above 15 is Approximate up to 40%

What is nature of skewness of the data?

**Sol:** Left skewed, Median is greater than mean.

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

**Sol:** -8

Q19) Comment on the below Boxplot visualizations?



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

**Sol:** Mean and median are equal hence 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)

Sol: 33/81

* 1. P(MPG<40)

Sol: 67/81

* 1. P (20<MPG<50)

Sol: 69/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Sol:** It nearly follows the 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

**Sol:** Mean is greater than median; hence the data is positively skewed, it does not follow the normal distribution.

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

**Sol:** Z score of 90% confidence Interval=1.65

Z score of 94% confidence Interval=1.55

Z score of 60% confidence Interval=0.85

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

**Sol:** T score for 95% confidence interval=1.96

T score for 96% confidence interval=2.5

T score for 99% confidence interval=2.47

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 free

Sol: Average life of bulb is 300 Days.