|  |
| --- |
| Data Type |
| Number of beatings from Wife | Discrete Data |
| Results of rolling a dice | Discrete Data |
| Weight of a person | Continuous Data |
| Weight of Gold | Continuous Data |
| Distance between two places | Continuous Data |
| Length of a leaf | Continuous Data |
| Dog's weight | Continuous Data |
| Blue Color | Discrete Data |
| Number of kids | Discrete Data |
| Number of tickets in Indian railways | Discrete Data |
| Number of times married | Discrete Data |
| Gender (Male or Female) | Discrete Data |

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

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= Null
2. Less than or equal to 4= 1/6
3. Sum is divisible by 2 and 3= 1/12

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.

Expected number of candies for randomly selected child = 1\*0.15+ 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**

For Points, For Score, For Weigh,

Mean = 3.60 Mean= 3.21 Mean= 17.84

Median= 3.69 Median= 3.325 Median= 17.71

Mode= 3.92 Mode= 3.44 Mode= 17.02

Variance= 0.2 Variance= 0.95 Variance= 3.19

Standard Deviation= 0.53 Standard Deviation = 0.97 Standard Deviation= 8.4

Range= 2.17 Range= 3.91 Range= 8.4

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?

Expected value= ∑ P(x)\*E(x)

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

(187)+ (1/9) (199)

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Skewness Kurtosis**

Cars= 0.00 Cars= -1.200

Speed= -0.117510 Speed= -0.5089

Distance= 0.8068 Distance= 0.4050

**SP and Weight (WT)**

**Use Q9\_b.csv**

**Skewness for speed= -0.1139548, skewness value is negative so it is left skewed and for distance = 0.7824 it is right skewed**

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



Positively skewed distribution because the tail is to right side of distribution.



Median is less than mean and outliers are on the upper side of the box plot and there is less data points between Q1 and Q2

**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?

Adult in Mexico with 94% CI

198.7383, 201.26167

Adult in Mexico with 98% CI

198.4394, 201.5605

Adult in Mexico with 96% CI

198.6223, 201.3776

**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= 25.52941

Standard Deviation= 5.052

1. What can we say about the student marks?

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

Normalized skewness

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

Right Skewed

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

Left Skewed

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

It indicate the distribution has a higher peak.

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

It indicates flatter peak and it has thin tail.

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



What can we say about the distribution of the data?

The data possess a left skewed distribution

What is nature of skewness of the data?

The data is skewed to left

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

8

Q19) Comment on the below Boxplot visualizations?



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

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) =33
  2. P(MPG<40) =61

c. P (20<MPG<50) =81

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

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

Z score of 90% confidence interval = 1.64

Z score of 94% confidence interval = 1.88

Z score of 60% confidence interval = 0.84

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence

Interval, 99% confidence interval for sample size of 25.

For 95% 1.96

For 96% 2.5

For 99% 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 freedom

Ans :- 0.321