|  |  |
| --- | --- |
| 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 | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| 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?

Ans : HHH, TTT, HHT, HTH, TTH, THT, HTT

P = 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 : HH, TT, HT, TH.

1. P=0
2. P=6/36 or 1/6
3. P=6/36 or 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?

Ans : 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

Ans : P = 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 : Refer Q7.csv file

From box plot we see that there are outliers in score and weigh.

For points mean is less than median then the distribution plot is negatively skewed data

For score mean is less than median then the distribution plot is negatively skewed data

For weigh mean is greater than median then the distribution plot is positively skewed data.

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

=145.3333

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

**Cars speed and distance**

**Use Q9.csv**

Ans : refer Q9.csv

For cars speed, skewness = -0.11395477 hence it is negatively skewed data

For cars distance, skewness = -0.7824352 hence indicate positively skewed data.

**SP and Weight(WT)**

**Use Q9\_b.csv**

Refer Q9.csv file

For SP, Skewness = 1.58145368 hence it is positively skewed data

For cweight(WT), Skewness = -0.60330993 hence it is negatively skewed data.

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



Ans : By seeing the distribution plot the data is positively skewed.

The boxplot shows a uneven distribution of data that leans to the right. There are also outliers different from the rest and make the unevenness more extreme.

**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 : Confidence interval for 94% = (134.8507, 265.1492)

Confidence interval for 96% = (130.1535, 269.8464)

Confidence interval for 98% = (122.6512, 277.3487)

**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?

Ans : the students marks are not normally distributed. The majority of students marks range is from 35 - 45 marks.

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

Ans : if mean and median of data are equal then it shows the normal distribution or the data is symmetrically distributed. And the skewness is equal to zero.

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

Ans : if mean > median then it shows that the data is asymmetrically distributed and mass of the distributed concentrated on left side. Which indicates the data is positively skewed or right skewed.

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

Ans : if median > mean then it shows that the data is asymmetrically distributed and mass of the distribution concentrated on right side. Which indicates the data is negatively skewed or left skewed.

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

Ans : Positively values of kurtosis indicate that distributed is peaked and possesses thick tails.

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

Ans : Negative vakues of kurtosis indicate that distribution is flatter (less peaked) and possesses lower tails.

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



What can we say about the distribution of the data?

Ans : data is assymmetrically distributed.

What is nature of skewness of the data?

Ans : negatively skewed or left skewed.

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

Ans : IQR = 18-10=8.

Q19) Comment on the below Boxplot visualizations?



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

Ans : data is normally distributed. No Outliers.

The IQR for boxplot -2 is greater than the IQR of boxplot -1.

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 : refer the attached Q20.csv file

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans : MPG is 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

Ans : Both AT and Waist doesn’t follow Normal Distribution

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

Ans : Z scores of 90% confidence interval = 1.64

Z scores of 94% confidence interval = 1.88

Z scores 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

Ans : t scores of 95% confidence interval = 2.06

t scores of 96% confidence interval = 2.17

t scores of 99% confidence interval = 2.79

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)

Ans : t-score = -0.4714

Degree of freedom = 17

P(t) = 0.3216725.