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

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

Ans.: 0.375

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

Ans.: 0

1. Less than or equal to 4

Ans.: 0.166666

1. Sum is divisible by 2 and 3

Ans.: 0.166666

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.: 0.714285

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.: 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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.59656 | 3.217 | 17.85 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92, 3.07 | 3.44 | 17.02, 18.9 |
| Variance | 0.28588 | 0.957 | 3.193 |
| Std Dev | 0.53468 | 0.978 | 1.787 |
| Range | 2.17 | 3.911 | 8.4 |

A7)

From the results we can see that the average points a car would get is around **3.6** and manufacturers need to manufacture cars weighing in the range of **16.06 to 19.64** to get an average score of **3.217** and points ranging from **3.07** **to 4.13**

**Use Q7.csv file**

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.: 145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans.:**

|  |  |  |
| --- | --- | --- |
|  | Speed | Distance |
| Skewness | -0.11751 | 0.806895 |
| Kurtosis | -0.50899 | 0.405053 |

**Speed -** From the above table we can see the skewness for speed is in the acceptable range i.e., between **-0.5 to 0.5** and kurtosis is also in the acceptable range i.e., between **-2 to 2.**

**Distance –** From the above we can see that skewness for distance is outside the acceptable range hence it is **positively/right** **skewed** but the kurtosis lies in the acceptable range.

**SP and Weight (WT)**

**Use Q9\_b.csv**

**Ans.:**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.611450196 | -0.61475333 |
| Kurtosis | 2.977328944 | 0.950291491 |

**SP -** From the above table we can see that skewness for SP is outside the acceptable range i.e., of **-0.5 - 0.5** hence it is positively/right skewed and kurtosis is also outside the acceptable range i.e., of **-2 to 2**

**WT –** For **WT** we can see that skewness is slightly outside the acceptable range and is **negatively/left skewed** but the kurtosis is in the acceptable range

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



**Ans.:** From the histogram we can see that the average of the weight lies approximately around **75** and is not located at the center and we can also see that it is positively/right skewed, also there seem to be some outliers in the dataset.

**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?

**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.: Mean = 41.0, Median = 40.5, s2 = 25.529, s = 5.052

1. What can we say about the student marks?

**Ans.:** We can see that there is an outlier in the marks, the average marks a student would get is **41** and the variation of marks is from **36 to 46**

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

Ans.: The nature of skewness is symmetric when mean and median are equal.

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

Ans.: When mean > median the nature is right skewed.

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

Ans.: When mean < median the nature is left skewed.

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

Ans.: It means that it has sharper peak and fewer data values are located near the

mean and more data values are located on the tails.

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

Ans.: It mean that it has flatter peak and more data values are located near the

mean and less data values are located on the tails

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



What can we say about the distribution of the data?

**Ans.: By looking at the boxplot we can say that the lower quartile is around 10,**

**upper quartile is around 18 and median lies around 15.**

What is nature of skewness of the data?

**Ans.: the data is left skewed**

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

**Ans.: IQR = 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)

Ans.: 0.3475939

* 1. P(MPG<40)

Ans.: 0.729349

* 1. P (20<MPG<50)

Ans.: 0.8988689

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans.: MPG of cars does follow a 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.: Adipose Tissue and Waist Circumference do not follow a Normal Distribution**

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

**Ans.: 90% - 1.644, 94% = 1.88, 60% = 0.841**

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

**Ans.: 95% = 2.063, 96% = 2.171, 99% = 2.796**

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

**Ans.: 0.3218 ( t = (x – u) / (s / sqrt(n)))**

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom