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

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 | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |

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 = 0
2. Less than or equal to 4 : 1/6
3. Sum is divisible by 2 and 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?

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. 1\*0.015 + 4\*.20 + 3\*.65 + 5\*.005 + 6\*.01 + 2\*.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**

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?

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



Ans:

From the above Histogram and Boxplot following insights can be drawn:

1. The data ranges from 0 to 400. Most of the data occurs between 50 to 150.
2. Looking at the distribution of the data, we can see that it does not follow normal distribution. Its more spread out towards right then it is in left side.
3. Boxplot shows presence of outliers in the right side of the data i.e in upper digits.
4. Outlies causes mean to be greater that mode. So, testing visually and drawing inference about the data we can say that it is right skewed and as it has more outliers there are chances that it can be leptokutic.

**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.
2. What can we say about the student marks?

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

Ans. Mean=median; then normal distributed.

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

Ans. Mean > median; then right skewed.

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

Ans. mean < median; then left skewed.

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

Ans. Kurtosis is the measure of “tailedness”, which tells us how much data resides in tail. Positive kurtosis indicates heavy-tailed meaning showing the presence of more extreme values in the data. Also known as leptokurtic.

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

Ans. Negative Kurtosis means thin-tailedness meaning there are less extreme values in the data. In general it is known as Platykurtic.

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



What can we say about the distribution of the data?

Ans. From the above boxplot following insights can be drawn:

1. Data ranges from 1 to 20. Most of the data can be found in between 10 to 19.
2. Lowest value in the data is 2, while highest value is somewhere around 20. Lowest value is very less in comparison to highest value.
3. Looking at the distribution of the data it can be said that left side of the data has outliers.
4. Mean can be affected by the presence of extreme lowest value so mean <median.
5. We can say that data is left skewed.

What is nature of skewness of the data?

Ans. Left Skewed

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

Ans. Approximately (18 - 10 = 8) i.e 8.

Q19) Comment on the below Boxplot visualizations?



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

Ans. Following insights can be drawn from the figure:

1. Looking at the figure we can say that both are normally distributed.
2. Interval in boxplot 2 is more in comparison to boxplot1 .
3. No outliers present in 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)

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

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

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