|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Count data |
| Results of rolling a dice | Count 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 | Categorical data |
| Number of kids | Count data |
| Number of tickets in Indian railways | Count data |
| Number of times married | Count data |
| Gender (Male or Female) | Categorical 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 data |
| High School Class Ranking | Ordinal data |
| Celsius Temperature | Interval data |
| Weight | Ratio data |
| Hair Color |  |
| Socioeconomic Status | Ordinal data |
| Fahrenheit Temperature | Interval data |
| Height | Ordinal data |
| Type of living accommodation | Interval data |
| Level of Agreement | Ordinal data |
| IQ(Intelligence Scale) | Ratio data |
| Sales Figures | Ordinal data |
| Blood Group | Nominal data |
| Time Of Day | Ratio data |
| Time on a Clock with Hands | Ratio data |
| Number of Children | Ratio data |
| Religious Preference | Interval data |
| Barometer Pressure | Ratio data |
| SAT Scores | Ratio data |
| Years of Education | Ratio data |

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

HHH,HHT,HTT,TTT,THH,TTH,HTH,THT

HHT+HTH+THH=3/8=0.375

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

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?

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

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

Mean:

Unnamed: 0 0.000000

Points 3.596563

Score 3.217250

Weigh 17.848750

dtype: float64

Median

Points 3.695

Score 3.325

Weigh 17.710

dtype: float64

Mode = ?

STD

Points 0.534679

Score 0.978457

Weigh 1.786943

dtype: float64

Range = ?

* 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**

**skewness**

Index 0.000000

speed -0.117510

dist 0.806895

**kurtosis**

Index -1.200000

speed -0.508994

dist 0.405053

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



**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?

When mean, median is equal then skewness is zero.

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

Then it is positive skewness

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

Then it is negative skewness

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

Positive kurtosis means thin at start distribution is peak and thick as tail.

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

Negative kurtosis means thick at start and distribution is peak and thin at tail.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

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)
  2. P(MPG<40)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

Pair-plot and Q-Q plot is used to show or check the normal distribution of given data

Yes- it follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

yes

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