|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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

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

Answer: **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?

Answer: 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

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**

**Answer:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | SD | Range |
| Points | 3.6 | 3.7 | 3.07, 3.92 | 0.28 | 0.53 | 2.76 – 4.93 |
| Score | 3.22 | 3.32 | 3.44 | 0.96 | 0.98 | 1.51 - 5.42 |
| Weigh | 17.85 | 17.71 | 17.02, 18.9 | 3.17 | 1.78 | 14.5 - 22.9 |

1. No variable`s data is normally distributed

2. In the Points column, there is no outlier. Whereas 3 exist in Score and one in

Weigh column.

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?

**Answer: 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | Kurtosis | Skewness |
| Speed | **-0.58** | **-0.11** |
| Distance | **0.25** | **0.78** |

Speed: data is negatively screwed but fairly symmetric

Distance: data in positively moderately skewed

For both the columns kurtosis is close to zero which means data is fairly normally distributed.

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | Kurtosis | Skewness |
| SP | **2.72** | **1.58** |
| WT | **0.81** | **-0.6** |

SP: data is positively screwed and asymmetric

Distance: data in negatively screwed and moderately symmetric

SP: data has positive kurtosis which indicated "heavy-tailed" distribution. Distribution is too peaked and it is called leptokurtic distribution.

WT: data has positive kurtosis but it is less than 1, so we may say data is fairly normally distributed.

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



Data is positively screwed which doesn’t indicate normal distribution.

Dataset has got some outliers too.

Mean>Median>Mode

**Q11) S**uppose 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?

Answer: 94% = (198.74, 201.26)

96% = (198.62, 201.38)

98% = (198.43, 201.56)

**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

Mode = 41

1. What can we say about the student marks?

Since mean, median and mode are almost similar, we may say marks are distributed uniformly.

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

The distribution has zero skewness

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

The distribution is positively skewed.

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

The distribution is negatively skewed.

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

It indicated distribution is peaked and possesses heavy tailed distribution.

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

It indicated distribution is flat and possesses light tailed distribution.

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



What can we say about the distribution of the data?

Data is not distributed normally. Mean<Median<Mode

What is nature of skewness of the data?

Negatively skewness.

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

10-18

Q19) Comment on the below Boxplot visualizations?



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

Median for both the datasets is similar.

Boxplot 2 has flatter tail than boxplot 1.

Boxplot 1 has Leptokurtic curve whereas boxplot 2 has Platykurtic Curve.

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) = 0.35
  2. P(MPG<40) = 0.73

c. P (20<MPG<50) =0.9

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

It is fairly normally distributed.

1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Dataset doesn’t follow normal distribution.

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

90% = 1.64

94% = 1.88

60% = 0.84

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

95% = 1.711

96% = 1.828

99% = 2.492

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

**Answer: 0.3218**