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

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

Ans. 1/2 = 3/6

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. a) 0

b) 1/6

c) 6/36

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. 5/7

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.

**Use Q7.csv file**

1. Points
2. Mean = 3.5965625000000006
3. Median = 3.6950000000000003
4. Mode = 0 3.07
5. 3.92

Dtype : float64

1. Variance = 0.28588135080645166
2. Standard deviation = 0.5346787360709716
3. Range = 2.17
4. Score
5. Mean = 3.2172499999999995
6. Median = 3.325
7. Mode= 0 3.44

dtype: float64

1. Variance= 0.9573789677419354
2. Standard Deviation= 0.9784574429896966
3. Range= 3.9109999999999996
4. Weigh
5. Mean= 17.848750000000003
6. Median= 17.71
7. Mode= 0 17.02
8. 18.90

dtype: float64

1. Variance= 3.193166129032258
2. Standard deviation= 1.786943
3. Range= 8.399999999999999

* Merc 280 and Merc 280c is the most repeated car

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 SP and Weight (WT)**

**Use Q9\_b.csv**

**Ans) a) Skewness:** Index 0.000000

speed -0.117510

dist 0.806895

dtype: float64

1. **Kurtosis:** Index -1.200000

speed -0.508994

dist 0.405053

dtype: float64

* Distance is right skewed
* Speed is left skewed

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



* The histogram is right skewed



**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 at 94%,98%,96% confidence interval?

**Ans**:

* + - Population = N = 30,00,000
    - Sample size = n = 2000
    - Sample average = 200
    - Standard deviation = σ = 30
    - Confidence level at 94% = 201.2617
    - Confidence level at 98% = 201.5606
    - Confidence level at 96% = 201.3777

**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**: mean = 41

Variance = 25.52941

Standard deviation = 5.052664

There is more deviation from the mean leading to inconsistency in the performance of the students.

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

**Ans:** If the distribution is symmetric, then the mean is equal to the median, and the distribution has zero skewness

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

**Ans:** If the mean is greater than the median, the distribution is positively skewed

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

**Ans:** If the mean is less than the median, the distribution is negatively skewed

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

Ans: Positive values of kurtosis indicate that a distribution is peaked and possess thick tail

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

Negative values of kurtosis indicate that a distribution is flat and has thin tails

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



What can we say about the distribution of the data?

**Ans:**  It indicates the mean is towards the left of the median and not normal distribution

What is nature of skewness of the data?

Ans: It indicates left skewed or negatively skewed

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

8

Q19) Comment on the below Boxplot visualizations?



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

Ans :

* Since the median of the two-box plot overlap with each other or almost overlap. The data in box plot 1 are more associated values with less difference. The data in box plot 2 are widely spread and with more difference.
* The IQR is more in boxplot 2 when compared to 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) = 0.6524061
  2. P(MPG<40) = 0.7293499

c. P (20<MPG<50) = 0.8988689

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

P -value : 1.6574419104049988e-16 < 0.05

According to Shapiro test Null hypothesis is rejected

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

Dataset: wc-at.csv

P-value : 5.258175228561868e-09

According to Shapiro test Null hypothesis is rejected -> Not normal distrubution

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

* 90% - 1.644854
* 94% - 1.880794
* 60% - 0.8416212

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

* 95% - 2.0638985616280205
* 96% - 2.1715446760080677
* 99% - 2.796939504772804

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