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

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

1. When 3 coins are tossed the total no. of possible outcomes =8

The no. of outcomes which have 2 heads and 1 tail = 3

P(2 heads and 1 tail ) = 3/8

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

1. Equal to 1

If two dices were rolled, then total possible cases =36

Total Favourable cases (Having sum =1) = 0

As minimum sum is 2 for outcome (1,1).

Hence, probability is 0.

1. Less than or equal to 4
   1. Thus number of event n(S) = 36

Let “A” be the event of getting a total sum of **4** or less.

Now, A = {(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)}

n(A) = 6

Therefore, probability of getting a sum of 4 or less

P(A) = n(A) / n(S)

= 6/36

= 1/6

c. Sum is divisible by 2 and 3

**a. Total number of possible outcomes = 36**

Favorable outcomes = sum is divisible by 2 and 3

Sum should be divisible by both 2 and 3

Favorable outcomes = (1 , 5) , (2,4),(3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)

Therefore,

Number of favorable outcomes = 6

Probability = 6/36 = 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?

1. **The probability that none of the balls drawn is blue is 0.476**

Sum of number of balls =2 +3 +2 = 7

Assume that, S be the sample space and 'n(S)' be number of ways of drawing 2 balls out of 7 , which implies that;

n(S) = ⁷C₂ = (7\*6)/ (2\*1)  =21

Consider 'A' be event of drawing 2 balls, none of which is blue.

n(A) = Number of ways of drawing 2 balls out of (2+3) balls.

n(A) = ⁵C₂ = (5\*4)/(2\*1) = 10

P(A) = n(A)/ n(S) = 10/ 21 = 0.476

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | S/D | Range |
| Points | 3.596563 | 3.695 | 3.92 | 0.276948 | 0.526258 | 2.17 |
| Score | 3.21725 | 3.325 | 3.44 | 0.927461 | 0.963048 | 3.911 |
| Weigh | 17.84875 | 17.71 | 17.02 | 3.09338 | 1.758801 | 8.4 |

**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?

1. Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

Ex  =108, 110, 123, 134, 135, 145, 167, 187, 199

P(x)  = 1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

Expected Value  =  (1/9)(108) + (1/9)110  + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9)  (  1308)

= 145.33

Expected Value of the Weight of that patient = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**a.**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| speed | -0.11395 | -0.50899 |
| Distance | -0.78248 | -0.40505 |

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.61145 | -0.60331 |
| Kurtosis | 2.977329 | 0.950291 |

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

1. Max number of values (weight) lie in the range of 50 to 100





A . Histogram:

The data is skewed on the right side ie data is positively skewed.

There are no outliers in the data.

Boxplot:

Data is distributed on the right and positively 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 94%,98%,96% confidence interval?

For 94% - (198.73 or 201.27).

For 96% - (198.61or 201.39).

For 98% - (198.43 or 201.57).

**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.
   1. Mean – 40.94117647
   2. Median – 40
   3. Variance - 25.46712803
   4. Std Dev - 5.046496609
2. What can we say about the student marks?
   1. Density of students with 41 marks are highest.

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

1. Skewness is zero.

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

1. Skewness is positive

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

1. Skewness is negetive

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

1. indicate that distribution is peaked and possesses thick tails

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

1. Indicates most of the data points are present in high proximity with mean.

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



What can we say about the distribution of the data?

1. Median is between 14 and 16, Q1(25 percentile) – 10 and Q3 (75 percentile) - 18

What is nature of skewness of the data?

1. Negative skewness

What will be the IQR of the data (approximately)?   
a. 10 - 18

Q19) Comment on the below Boxplot visualizations?



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

Ans: First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)
     1. 33/81
  2. P(MPG<40)
     1. 61/81
  3. P (20<MPG<50)
     1. 69/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol. Yes the above data does show 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

Sol. Only Adipose Tissue (AT) follow normal distribution.

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

1. 90% - 1.645
2. 94% - 1.880746
3. 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% - 2.0638985616280205

96% - 2.1715446760080677

99% - 2.1715446760080677

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

1. The probability that t < - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.