|  |  |
| --- | --- |
| 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 | 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 | 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?

**Ans:** Events May happen: {HHH, HTT, HHT, HTH, TTT, THH, TTH, THT} = 8

Two heads and one tail events: {HHT, HTH, THH} = 3

Probability of two heads and one tail is = 3/8

**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:** Total Events

{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)

(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)

(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)

(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)

(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)

(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)} = 36

1. Equal to 1: 0% Probability
2. Less than or equal to 4: 6/36 = 1/6
3. Sum is divisible by 2 and 3: 6/36 = 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:** 14/21

The total number of balls in the bag is 2 (red) + 3 (green) + 2 (blue) = 7.

Drawing one non-blue ball and one blue ball

The number of favorable outcomes for this scenario is the number of ways to choose 1 non-blue ball from the 2 red and 3 green balls, multiplied by the number of ways to choose 1 blue ball from the 2 blue balls.

Number of favorable outcomes = (C(2, 1) + C(3, 1)) \* C(2, 1) = (2 + 3) \* 2 = 1

Total number of possible outcomes = C(7, 2) = 21

Therefore, the total number of favorable outcomes is 4 + 10 = 14.

The probability of not drawing any blue balls is given by:

Probability = Number of Favorable Outcomes / Total Number of Possible Outcomes

Probability = 14 / 21

Simplifying the fraction, we get:

Probability = 2 / 3 = 0.66 = 66%

**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

Expected values when the probability is provided for a data point is given as

E(X) = ∑ X \* P(X)

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count | Probability | Expected Value |
| A | 1 | 0.015 | 1 \* 0.015 = 0.015 |
| B | 4 | 0.20 | 4 \* 0.20 = 0.80 |
| C | 3 | 0.65 | 3 \* 0.65 = 1.95 |
| D | 5 | 0.005 | 5 \* 0.005 = 0.025 |
| E | 6 | 0.01 | 6 \* 0.01 = 0.06 |
| F | 2 | 0.120 | 2 \* 0.120 = 0.240 |

To get the expected number of candies for a randomly selected child we sum up all the calculated expected values

0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.240 = 3.09

the Expected number of candies for a randomly selected child is 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.

**Ans:**

**Use Q7.csv file**

|  |  |  |  |
| --- | --- | --- | --- |
| Metrics | Points | Score | Weigh |
| Mean | 3.596 | 3.217 | 17.848 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 0 3.07,  1 3.92 | 0 3.44 | 0 17.02  1 18.90 |
| Standard Deviation | 0.534 | 0.978 | 1.786 |
| Variance | 0.285 | 0.957 | 3.193 |
| range | 2.17 | 3.911 | 8.399 |

Ans –we get the idea about skewness when compare mean and median of columns of data, here we can see that mean and media are equivalent to each other with verry slight difference and mean, median, & mode are also equivalent to each other so we conclude that data is normally distributed

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:** 145.33

Ans –

Expected values when the probability is provided for a data point is given as

E(X) =∑ (X \* P(X))

Count of weight of patient at a clinic = 9

Probability of patient selected randomly = 1/9

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

= 1308 \* 1/9

= 1308/9

= 145.33

the Expected Value of the Weight of that patient chosen randomly = 145.33

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

**Car’s speed and distance**

**SP and Weight (WT)**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | -0.11751 | -0.50899 |
| **Distance** | 0.806895 | 0.405053 |
| **SP** | 1.61145 | 2.977329 |
| **Weight** | -0.61475 | 0.950291 |
|  |  |  |

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



**Ans:**

1. Most of the chick wight lign in bwtween 50 - 100
2. The data is positively Skewed



**Ans:** There is lots of outliers is present and the data is 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?

**Ans:**

n=2000

x̅=200

s=30

SE=

CI = SE \* tvalue

|  |  |  |  |
| --- | --- | --- | --- |
|  | **94%** | **98%** | **96%** |
| **Lower** | 201.04 | 201.38 | 201.17 |
| **Upper** | 198.96 | 198.62 | 198.83 |

**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.

**Ans:**

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.5 |
| Deviation | 5.05 |

1. What can we say about the student marks?

**Ans:** 1. Average of students get 41 marks

2. There is no outlier is preset

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

**Ans:** Skewness is totally 0 and perfectly symmetric bell-shaped curve

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

**Ans:** Right Skewed

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

**Ans:** Left Skewed

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

**Ans:** High peak at center of data

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

**Ans:** Wide peak at center of data

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



What can we say about the distribution of the data?

**Ans:** It is not a Normal Distribution

What is nature of skewness of the data?

**Ans:** It is left skewed

What will be the IQR of the data (approximately)?   
**Ans:** IQR is 8

**Q19)** Comment on the below Boxplot visualizations?



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

**Ans:**

1. Both boxplots have same median approximate 262
2. Both boxplots don’t have outliers

**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)

**Ans:** 0.4074074

1. P(MPG<40)

**Ans:** 0.7530864

1. P (20<MPG<50)

**Ans:** 0.8518519

**Q 21)** Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:** MPG have 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

**Ans:** Adipose Tissue (AT) and Waist Circumference (Waist) follow Normal Distribution

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

Ans:

|  |  |
| --- | --- |
| **90%** | 2.06 |
| **94%** | 2.17 |
| **60%** | 2.79 |

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

**Ans:**

|  |  |
| --- | --- |
| **95%** | 2.06 |
| **96%** | 2.17 |
| **99%** | 2.79 |

**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

**Ans:** µ=270, =260, SD=90, n=18, df=17

t-score = =-0.4714  
Degree of freedom = 17   
P(t) = 0.3216725