|  |  |
| --- | --- |
| 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 Data |
| High School Class Ranking | Ordinal Data |
| 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 | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

Answer : Three coins are tossed, therefore there are 8 possibilities as follows

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

Probability of getting two heads and one tail = 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

Answer : Two dice are rolled therefore the possible outcome will be as follows

S =

(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)

N(S)=36

1. Equal to 1

Ans.: Two dice are rolled therefore minimum sum we will get is 2.

Probability of getting sum equal to 1 = 0/36 = 0

1. Less than or equal to 4

Ans.: Probability of getting sum equal to or less than 4 = 6/36 = 0.1667

1. Sum is divisible by 2 and 3

Ans.: Probability of getting sum divisible by 2 and 3 = 6/36 = 0.1667

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?

Total Number of Balls = { 2 red, 3 green, 2 blue }= 7

n(S) = 7

Probability of First ball not being blue is = 5/7

Probability of Second ball not being blue is = 4/6

Therefore,

Probability of none of the balls drawn is blue = (5/7)\*(4/6)

= 20/42

= 10/21

Probability of none of the balls drawn is blue = 0.47

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

Answer :

Expected number of candies for a randomly selected child =

1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120 = 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 : Q.7 Answer Github link given below**

## **https://github.com/khuspure/DataScience\_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.7%20-%20Assignment%201.ipynb**

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. :** n=9

Therefore, probability of selecting one of them = 1/9

Expected value of weight = 1/9(sum of all weights)

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

=1/9(1308)

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans : Q.9\_ a) Answer Github link given below**

## **https://github.com/khuspure/DataScience\_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.9a%20-%20Assignment%201.ipynb**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans : Q.9\_ b) Answer Github link given below**

## [**https://github.com/khuspure/DataScience\_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.9b%20-%20Assignment%201.ipynb**](https://github.com/khuspure/DataScience_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.9b%20-%20Assignment%201.ipynb)

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



**Ans. :**

**Histogram**

Given histogram is about distribution of chick weight data and as we see the plot

1. It is positively skewed.
2. More than 50% data is in between 50 to 150
3. Most of the chick weight is in between 50 to 100



**Ans. :**

As we see the plot,

1. Given boxplot is positively skewed.
2. There are many upper outliers in the boxplot.

**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 : Q.11 Answer Github link given below**

## [**https://github.com/khuspure/DataScience\_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.11%20-%20Assignment%201.ipynb**](https://github.com/khuspure/DataScience_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.11%20-%20Assignment%201.ipynb)

**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 : Q.12 Answer Github link given below**

## [**https://github.com/khuspure/DataScience\_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.12%20-%20Assignment%201.ipynb**](https://github.com/khuspure/DataScience_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.12%20-%20Assignment%201.ipynb)

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

Answer : **Symmetrical**

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

Answer : **Positively Skewed or Right Skewed**

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

Answer : **Negatively Skewed or Left Skewed**

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

Answer : **The given data consists Normal Distribution and Kurtosis value is 0.**

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

Answer : **The distribution of the data has lighter tails and a flatter peaks than the normal distribution.**

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



What can we say about the distribution of the data?

**Ans.: The mean has higher values and most of data is concentrated between q1 and lower end.**

What is nature of skewness of the data?

**Ans.: Negative Skewness**

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

**Ans.:18\*10=180**  
  
  
Q19) Comment on the below Boxplot visualizations?



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

**Ans.: Symmetrical**

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)

**Ans : Q.20 Answer Github link given below**

## [**https://github.com/khuspure/DataScience\_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.20%20-%20Assignment%201.ipynb**](https://github.com/khuspure/DataScience_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.20%20-%20Assignment%201.ipynb)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

**Ans : Q.21 Answer Github link given below**

## **https://github.com/khuspure/DataScience\_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.21%20-%20Assignment%201.ipynb**

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

**Ans : Q.22 Answer Github link given below**

## [**https://github.com/khuspure/DataScience\_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.22%20-%20Assignment%201.ipynb**](https://github.com/khuspure/DataScience_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.22%20-%20Assignment%201.ipynb)

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

**Ans : Q.23 Answer Github link given below**

## [**https://github.com/khuspure/DataScience\_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.23%20-%20Assignment%201.ipynb**](https://github.com/khuspure/DataScience_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.23%20-%20Assignment%201.ipynb)

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 : Q.24 Answer Github link given below**

## **https://github.com/khuspure/DataScience\_Assignments/blob/main/Assignment%201-%20Basic%20Statistics%201/Q.24%20-%20Assignment%201.ipynb**