|  |  |
| --- | --- |
| 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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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?

**Ans**. Three coins are tossed.

Sample Space(S)= (HHH, HHT, HTH, THH, TTT, TTH, THT, HTT)

Two Heads and one tail (A)= (HHT, HTH, THH)

P(A) = No. of elements in event A/No. of elements in sample space.

P(A) =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.** The result of 2 dice can be as follows

S= {(1,1)(2,1)(3,1)(4,1)(5,1)(6,1)​(1,2)(2,2)(3,2)(4,2)(5,2)(6,2)​(1,3)(2,3)(3,3)(4,3)(5,3)(6,3)​(1,4)(2,4)(3,4)(4,4)(5,4)(6,4)​(1,5)(2,5)(3,5)(4,5)(5,5)(6,5)​(1,6)(2,6)(3,6)(4,6)(5,6)(6,6)}

N(s) =36​}

A: sum is equals to 1

A: ()

N(A) = 0

P(A) = 0

B: sum Less than or equal to 4

B: {(1,1)(1,2)(3,1)(1,2)(2,2)(1,3)}

N(B) = 6

P(B) = 6/36 = 1/6

C: Sum is divisible by 2 and 3

C: {(5,1)(4,2)(3,3)(2,4)(1,5)(6,6)}

N(C) = 6

P(B) = 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:** S: sample space

N(s) = Number of ways of drawing 2 balls out of 7 balls = 7 C2 = 21

A: Event of drawing two balls, none of which is blue.

n(A): Number of ways of drawing 2 balls out of 5 (2red, 3green) balls = 5C2 = 10

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

**Ans: =** 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.

**Use Q7.csv file**

**Ans**. [Tap to view](https://colab.research.google.com/drive/1NKvYTrVuXONKKWbU-qYZpUBmW8in6III?usp=sharing#scrollTo=Px5NNrbYjHVW)

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**. Expected Value =   Σ(Probability \* Value)

There are 9 points

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+ 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.** [Tap to View](https://colab.research.google.com/drive/1NKvYTrVuXONKKWbU-qYZpUBmW8in6III?usp=sharing#scrollTo=Px5NNrbYjHVW)

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans.** [Tap to view](https://colab.research.google.com/drive/1NKvYTrVuXONKKWbU-qYZpUBmW8in6III?usp=sharing#scrollTo=Px5NNrbYjHVW)

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



**Ans.**

**Histogram**

* Given histogram is about distribution of chick weight data, which is positively skewed.
* More than 50% data is in between 50 to 150
* Most of the chick weight is in between 50 to 100



**Boxplot**

Given boxplot is positively skewed.

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.** [Tap to view](https://colab.research.google.com/drive/1jt79X1qlpEhDDQLkMxA8hECuumDJB2Zf#scrollTo=6s3ECOzVrRQq)

**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.** [Tap to view](https://colab.research.google.com/drive/1SVHuW8O2rLCkDgxBehqkMM8LBbrMdnVC#scrollTo=nq3Kwn_AssI6)

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

*Ans- Symmetrical*

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

*Ans-* *Positive Skewness*

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

*Ans-* *Negative Skewness*

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

*Ans- 0 (Zero)*

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

*Ans- The distribution of the data has lighter tails and a flatten 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 min values*

What is nature of skewness of the data?

*Ans- Negative Skewness*

What will be the IQR of the data (approximately)?   
  
*Ans- 18\*10=8*

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**. [Tap to view](https://colab.research.google.com/drive/1g3XSNaGyTWP4U62ffPIXW4f0otBRI9cQ?usp=sharing)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans**. [Tap to View](https://colab.research.google.com/drive/1g3XSNaGyTWP4U62ffPIXW4f0otBRI9cQ?usp=sharing)

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

Dataset: wc-at.csv

**Ans.** [Tap to View](https://colab.research.google.com/drive/1g3XSNaGyTWP4U62ffPIXW4f0otBRI9cQ?usp=sharing)

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

**Ans**. [Tap to View](https://colab.research.google.com/drive/1AwTPlgsjXj3coydQ-uRCx70xdALW9Mn2?usp=sharing)

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

**Ans**. [Tap to View](https://colab.research.google.com/drive/1AwTPlgsjXj3coydQ-uRCx70xdALW9Mn2?usp=sharing)

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**. [Tap to View](https://colab.research.google.com/drive/1AwTPlgsjXj3coydQ-uRCx70xdALW9Mn2?usp=sharing)