DATA HANDLING

Discrete Frequency Distribution:
Let the given data consist of n distinct values x_1, x_2,..., x_n , each occurring with corresponding frequencies f_1, f_2,...,f_n. This data can be represented in a tabular form with each distinct value listed alongside its corresponding frequency. The resulting table is known as a **Discrete Frequency Distribution**.

The table typically includes the following columns:
1.**Value (x_i):** The distinct data points x_1, x_2,..., x_n
2.**Frequency (f_i):** The number of occurrences f_1, f_2,...,f_n of each corresponding value.
3.**Cumulative Frequency:** The running total of frequencies up to the current value.
4.**Frequency × Value ($f_i \times x_i$):** The product of each frequency and its corresponding value.

Value (x_i)	Frequency (f_i)	Cumulative Frequency	Frequency × Value ($f_i x_i$)
x_1	f_1	f_1	$f_1 x_i$
x_2	f_2	$f_1 + f_2$	$f_2 x_2$
......	......	......	
x_n	f_n	$f_1 + f_2 {}_{+............+} f_n$	$f_n x_n$

$$\text{Mean} = \frac{\sum_{i=1}^{n}(f_i \times x_i)}{\sum_{i=1}^{n} f_i}$$

The mean, median, and mode are fundamental measures of central tendency in a dataset, each providing a different perspective on the data's distribution, allowing for more effective analysis and interpretation of the data's characteristics.

1. **Mean:** The mean of a dataset, often referred to as the average, is calculated by summing all the values in the dataset and dividing the total by the number of values. It is a key measure of central tendency that gives an overall idea of the data's general magnitude.

2. **Median:** The median is the middle value in a dataset when the numbers are arranged in ascending or descending order. If the dataset contains an odd number of values, the median is the middle number. If the dataset has an even number of values, the median is the average of the two middle numbers. The median is particularly useful for understanding the central point in a dataset, especially when dealing with skewed data or outliers.

3. **Mode:** The mode is the value that appears most frequently in a dataset. A dataset can have one mode (unimodal), more than one mode (bimodal or multimodal), or no mode at all if no value repeats. The mode helps to identify the most common value(s) in the dataset, making it useful in categorical data analysis.

Problem:

The following are the results of a monthly test in which 25 students out of a total of 30 marks. Make a discrete frequency distribution of the data. 15, 13, 16, 26, 25, 16, 17, 14, 15, 16, 26, 12, 19, 22, 22, 23, 27, 14, 25, 16, 23, 17, 12, 25, 16.

Calculate the average marks, median and mode.

Problem 2:

The frequency of grades obtained by students in a subject is given in the table.

Find out the average grade, median and mode and the range /grade where they lie if DD, CD,.......AA corresponds to the marks range 30-40, 40-50,.....90-100 respectively.

Display the Data in a suitable graphical representation

Grade	Frequency
DD	3
CD	5
CC	7
BC	16
BB	11
AB	8
AA	4

1. Calculate Average Marks

Data:

- **Grades**: DD, CD, CC, BC, BB, AB, AA

- **Ranges**: 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100

- **Frequencies**: 3, 5, 7, 16, 11, 8, 4

Midpoints of Ranges:

- DD (30-40): Midpoint = 35

- CD (40-50): Midpoint = 45

- CC (50-60): Midpoint = 55

- BC (60-70): Midpoint = 65

- BB (70-80): Midpoint = 75

- AB (80-90): Midpoint = 85

- AA (90-100): Midpoint = 95

Weighted Average Calculation:

$$\text{Average} = \frac{\sum(\text{Midpoint} \times \text{Frequency})}{\sum \text{Frequency}}$$

$$\text{Average} = \frac{(35 \times 3) + (45 \times 5) + (55 \times 7) + (65 \times 16) + (75 \times 11) + (85 \times 8) + (95 \times 4)}{54}$$

$$\text{Average} = \frac{105 + 225 + 385 + 1040 + 825 + 680 + 380}{54}$$

$$\text{Average} = \frac{4390}{54} \approx 81.85$$

2. Find the Median and Mode

Total Number of Students:

Total=3+5+7+16+11+8+4=64

Cumulative Frequencies:
- DD: 3
- CD: 3 + 5 = 8
- CC: 8 + 7 = 15
- BC: 15 + 16 = 31
- BB: 31 + 11 = 42
- AB: 42 + 8 = 50
- AA: 50 + 4 = 54

Median Calculation:

With 54 students (an even number), the median is the average of the ($n/2$ and $1+n/2$) 27th and 28th values :

Both fall into the "BC" range (60-70).

Mode Calculation:

The mode is the grade with the highest frequency, which is "BC" with a frequency of 16.

Results:

Average Marks: Approximately 81.85

Median Range: BC (60-70)

Mode Range: BC (60-70)

Grade	Range	Mid-point	Frequency	Frequency × Midpoint	Cumulative Frequency	Average/Median/Mode
DD	30-40	35	3	105	3	
CD	40-50	45	5	225	8	
CC	50-60	55	7	385	15	
BC	60-70	65	16	1040	31	Median Mode
BB	70-80	75	11	825	42	Average: 81.85 Mode
AB	80-90	85	8	680	50	Average: 81.85
AA	90-100	95	4	380	54	
Total			54	4390		Mode: BC (60-70)

Problem:

The following are the results of a monthly test in which 25 students out of a total of 30 marks. Make a discrete frequency distribution of the data.
15, 13, 16, 26, 25, 16, 17, 14, 15, 16, 26, 12, 19, 22, 22, 23, 27, 14, 25, 16, 23, 17, 12, 25, 16.

Calculate the average marks, median and mode.

Calculations

1. Average Marks (Mean):

To find the average marks:

$$\text{Average} = \frac{\sum(\text{Frequency} \times \text{Marks})}{\text{Total Frequency}}$$

$$\text{Sum of Frequency} \times \text{Marks} = 24 + 13 + 28 + 30 + 64 + 34 + 19 + 44 + 46 + 75 + 52$$

$$\text{Total Frequency} = 25$$

$$\text{Average} = \frac{470}{25} = 18.8$$

2. Median:
The median is the middle value when the data is arranged in ascending order.
For 25 students, the median will be the 13th value.
From the cumulative frequency column, the 13th value falls in the 17 marks category.
So, the **Median** is 17.

3. Mode:
The mode is the value that appears most frequently.
From the frequency column, 16 appears the most often (4 times).
So, the **Mode** is 16.

Marks	Frequency	Cumulative Frequency	Frequency × Marks	Results
12	2	2	24	
13	1	3	13	
14	2	5	28	
15	2	7	30	
16	4	11	64	Mode (16)
17	2	13	34	Median (17)
19	1	14	19	
22	2	16	44	
23	2	18	46	
25	3	21	75	
26	2	23	52	
27	1	24	27	
Total	25		470	Mean (18.8)