Chapter 14: Statistics

This chapter deals with collecting, organizing, presenting, and analyzing data. You'll learn different ways to represent data visually and calculate measures that describe the "center" or "average" of a dataset.

1. Basic Definitions

1. **Data**
   - A collection of facts, numbers, or information.
   - Example: Marks of students: 85, 90, 78, 92, 88.
2. **Frequency**

$$\text{Frequency} = \text{Number of times a value occurs}$$

   - **Usage**: Count how many times each data value appears in the dataset.
3. **Class Interval**

$$\text{Range of values grouped together, e.g., } 10-20, 20-30$$

   - **Lower Limit**: Smallest value in class (e.g., 10).
   - **Upper Limit**: Largest value in class (e.g., 20).
   - **Class Mark (x_i)**: Mid-point $= \frac{\text{Lower Limit}+\text{Upper Limit}}{2}$.

2. Measures of Central Tendency (Ungrouped Data)

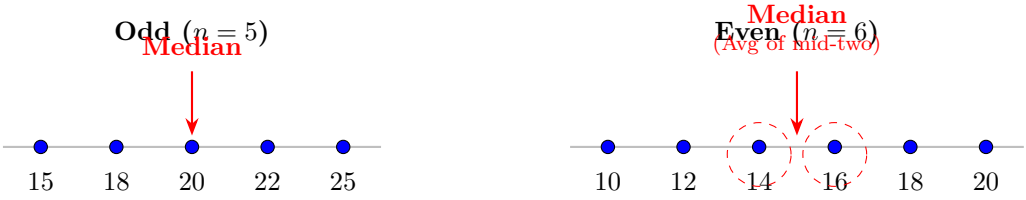
These are single values that represent the "center" or typical value of a dataset.

4. **Mean (Average)**

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

   - $\bar{x}$: Mean (read as "x-bar").
   - x_i: Individual data values.
   - n: Total number of observations.
5. **Median (Middle Value)**

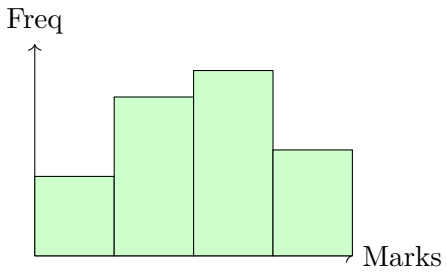
$$\text{Median} = \begin{cases} \text{Value at } \left(\frac{n+1}{2}\right)^{\text{th}} \text{ position} & \text{if } n \text{ is odd} \\ \text{Avg of } \left(\frac{n}{2}\right)^{\text{th}} \text{ and } \left(\frac{n}{2}+1\right)^{\text{th}} \text{ positions} & \text{if } n \text{ is even} \end{cases}$$



6. **Mode (Most Frequent Value)**
   - The value that occurs with the highest frequency.
   - Best for categorical data (e.g., most popular shoe size).

3. Graphical Representation of Data

7. **Bar Graph**: Used for discrete/categorical data with gaps between bars.
8. **Histogram**: Used for continuous data with **no gaps** between bars.



Histogram (Continuous)

9. **Frequency Polygon**
   - Formed by joining the mid-points (class marks) of the histogram bars with straight lines.

---

4. Mean for Grouped Data

10. **Direct Method**: $\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$
11. **Assumed Mean Method**: $\bar{x} = a + \frac{\sum f_i d_i}{\sum f_i}$ where $d_i = x_i - a$.
12. **Step Deviation Method**: $\bar{x} = a + \left(\frac{\sum f_i u_i}{\sum f_i}\right) \times h$ where $u_i = \frac{x_i - a}{h}$.

---

5. Mode for Grouped Data

13. **Mode Formula**

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$$

- l: lower limit of modal class.
- f_1: frequency of modal class.
- f_0: frequency of class preceding modal class.
- f_2: frequency of class succeeding modal class.

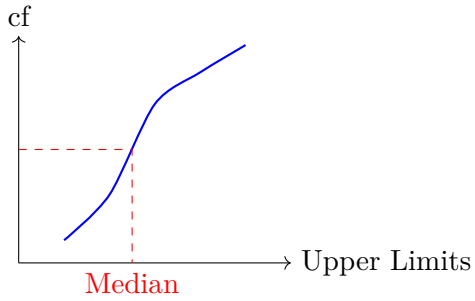
---

6. Median for Grouped Data

14. **Median Formula**

$$\text{Median} = l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times h$$

15. **Cumulative Frequency (cf)**: The "running total" of frequencies.



Ogive (Cumulative Frequency Curve)

---

7. Empirical Relationship

$$3 \times \text{Median} = \text{Mode} + 2 \times \text{Mean}$$

Quick Revision Summary

1. **Class Mark**: $\frac{\text{Upper Limit}+\text{Lower Limit}}{2}$
2. **Mean**: $\frac{\sum f_i x_i}{\sum f_i}$
3. **Mode**: Value with max frequency
4. **Median**: Middle value (positional average)
5. **Empirical**: $3\text{Med} = \text{Mode} + 2\text{Mean}$