



**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

# Measurement of study variables

Dr. R. Ramakrishnan  
MAE, PhD

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in

## Types of Data

- Qualitative
  - Nominal
    - Eg. Color of Eyes
  - Ordinal
    - Eg. Stages of disease condition
- Quantitative
  - Discrete
    - Eg. Family size
  - Continuous
    - Eg. Height / Weight

**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in



**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

## Describe – Central Value

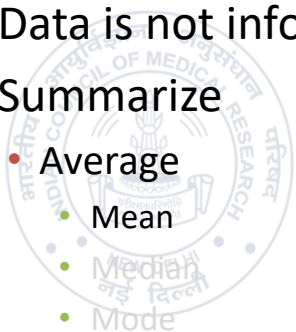
- Data is not information.
- Summarize

- Average

- Mean

- Median

- Mode



**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**

NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in

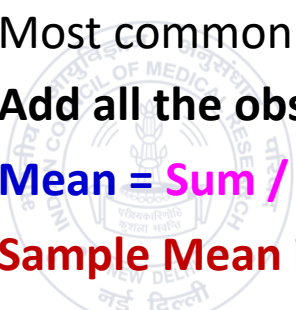


**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

## Arithmetic Mean (AM)

- Most commonly used; Simply called MEAN
- Add all the observed values ( $\text{Sum} = \sum X_i$ )
- $\text{Mean} = \text{Sum} / n$
- **Sample Mean** is denoted by  $\bar{x}$
- **Population Mean** is denoted by  $\mu$



**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**

NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in



**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

## Example

- Age of 10 Pregnant women

26, 31, 25, 21, 26, 26, 27, 25, 27, and 26

**Sum** =  $(26+31+25+21+26+26+27+25+27+26) = 260$

**n** = 10

**Mean** =  $\text{sum} / n = 260/10 = 26 \text{ years}$

## Describe – Central Value

- Data is not information.

- Summarize

- Average

- Mean

- Median

- Mode

## Median

**The Median** describes literally the middle value of the distribution

Divides the distribution **exactly into two halves** (i.e. 50% of the data will fall on either side)

Useful when there are extreme values

## Example

Duration (days) of hospital stay of 11 patients

1, 2, 3, 4, 5, 6, 7, 8, 8, 9, 77 (Arranged in ascending order)

Median is the middle value (6<sup>th</sup> value) = 6

(Mean = 11.8)

If n is even; then take average of middle two values.

## Describe – Central Value

- Data is not information.

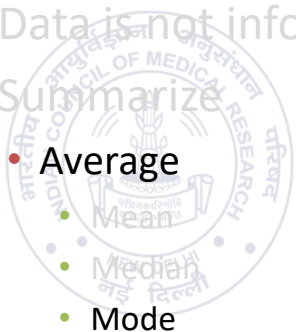
- Summarize

- Average

- Mean

- Median

- Mode



**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**

NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in



**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

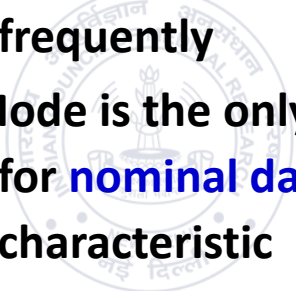
**NIE**  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

## Mode

**The Mode** is the value that occurs most frequently

Mode is the only location statistics to be used –  
for **nominal data** - **not measurable**  
characteristic

**Epidemiology – Describing an epidemic with  
respect to TIME**



**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**

NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in



**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

## Example

- Colour preference of people for their car

<u>Colour preference</u>	<u>No. of persons</u>
Green	354
Yellow	852
White	310
Red	474

**Mode = Yellow**

## Describe – Dispersion

- Is it enough to know the average?

- Example of swimming pool.

- Measures of variability

- Range

- Inter-quartile range

- Mean deviation from mean

- Variance / Standard deviation

# RANGE

## Definition:

The difference between the Minimum and the Maximum value of the observations

## Advantage:

A quick and easy indicator of dispersion.

## Disadvantage:

Influenced by extreme values; Uses only two data points

## Describe – Dispersion

- Is it enough to know the average?
  - Example of swimming pool.
- Measures of variability
  - Range
  - Inter-quartile range
  - Mean deviation from mean
  - Variance / Standard deviation

# INTER-QUARTILE RANGE

## Quartile Deviation

### Definition:

Defined as the interval between the value of the upper quartile (Q3) and the lower quartile (Q1)

$$\text{Inter Quartile Range} = Q_3 - Q_1$$

### Advantage:

Unaffected by the extreme values

### Disadvantage:

Covers only the middle 50% observations



icmr  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

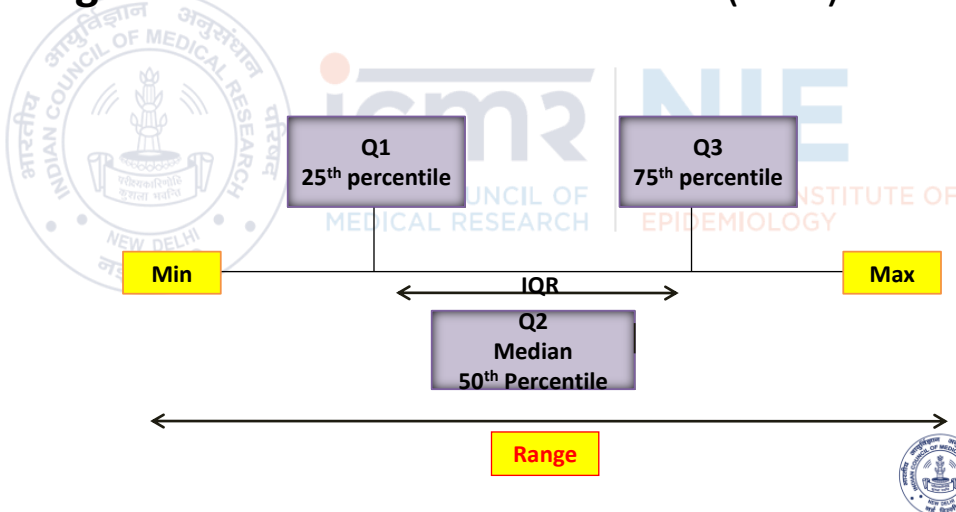
NIE  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

nie.gov.in

HEALTH RESEARCH FUNDAMENTALS

# INTER QUARTILE RANGE

The **range** of a data is divided into four (25%)



icmr  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

NIE  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

nie.gov.in

HEALTH RESEARCH FUNDAMENTALS



## Describe – Dispersion

- Is it enough to know the average?

- Example of swimming pool.

- Measures of variability

- Range
  - Inter-quartile range
  - Mean deviation from mean
  - Variance / Standard deviation



icmr  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

NIE  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

## MEAN DEVIATION

**Definition:** The mean deviation is the average of the absolute (ignoring the sign) deviations of the observations from the arithmetic mean.

**Advantage:** It is based on all the observations in the group. It is easy to grasp the meaning of the procedure.

**Disadvantage:** It ignores the sign of the difference of the value of the observation and arithmetic mean.

It is not widely used because of the availability of a more advantageous measure.



icmr  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

NIE  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

## Describe – Dispersion

- Is it enough to know the average?
  - Example of swimming pool.
- Measures of variability
  - Range
  - Inter-quartile range
  - Mean deviation from mean
  - Variance / Standard deviation

icmr  
INDIAN COUNCIL OF  
MEDICAL RESEARCHNIE  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

## STANDARD DEVIATION - SD ( $\sigma$ )

**Definition:** The SD is the square root of the average of the squared deviations of the observations from the arithmetic mean

The square of the SD is called variance

**Advantage:** The SD is the most important measure of distribution. While the variance is in unit squared, the SD is expressed in the same units of measurement as the observation. It is suitable for further analysis

The SD together with arithmetic mean is useful for description of the data

icmr  
INDIAN COUNCIL OF  
MEDICAL RESEARCHNIE  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

# Coefficient of Variation (CV)

Purpose: To compare the relative variability in different groups

Definition: The coefficient of variation is the SD expressed as a percentage of the arithmetic mean (AM).

$$CV = \left( \frac{SD}{AM} \right) \times 100$$



icmr  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

NIE  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

nie.gov.in

HEALTH RESEARCH FUNDAMENTALS

## Summary

- Choose appropriate central / dispersion value
  - Mean / SD – if no extreme values
  - Median / IQR – if there are extreme values
  - Mode / Range – for qualitative variables/ time distribution in epidemic curve
- Mean and SD are used the most.



icmr  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

NIE  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

nie.gov.in

HEALTH RESEARCH FUNDAMENTALS



**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

**icmr**  
INDIAN COUNCIL OF  
MEDICAL RESEARCH

**NIE**  
NATIONAL INSTITUTE OF  
EPIDEMIOLOGY

Thank you

---