



# University of Baghdad College of Medicine 2024-2025



**Title: Mathematical Presentation of Data -  
Part 1**

**Grade: Third**

**Module: Research Methodology**

**Speaker: Dr. Ghaith Sabri Mohammed**

**Date: 18/9/2024**

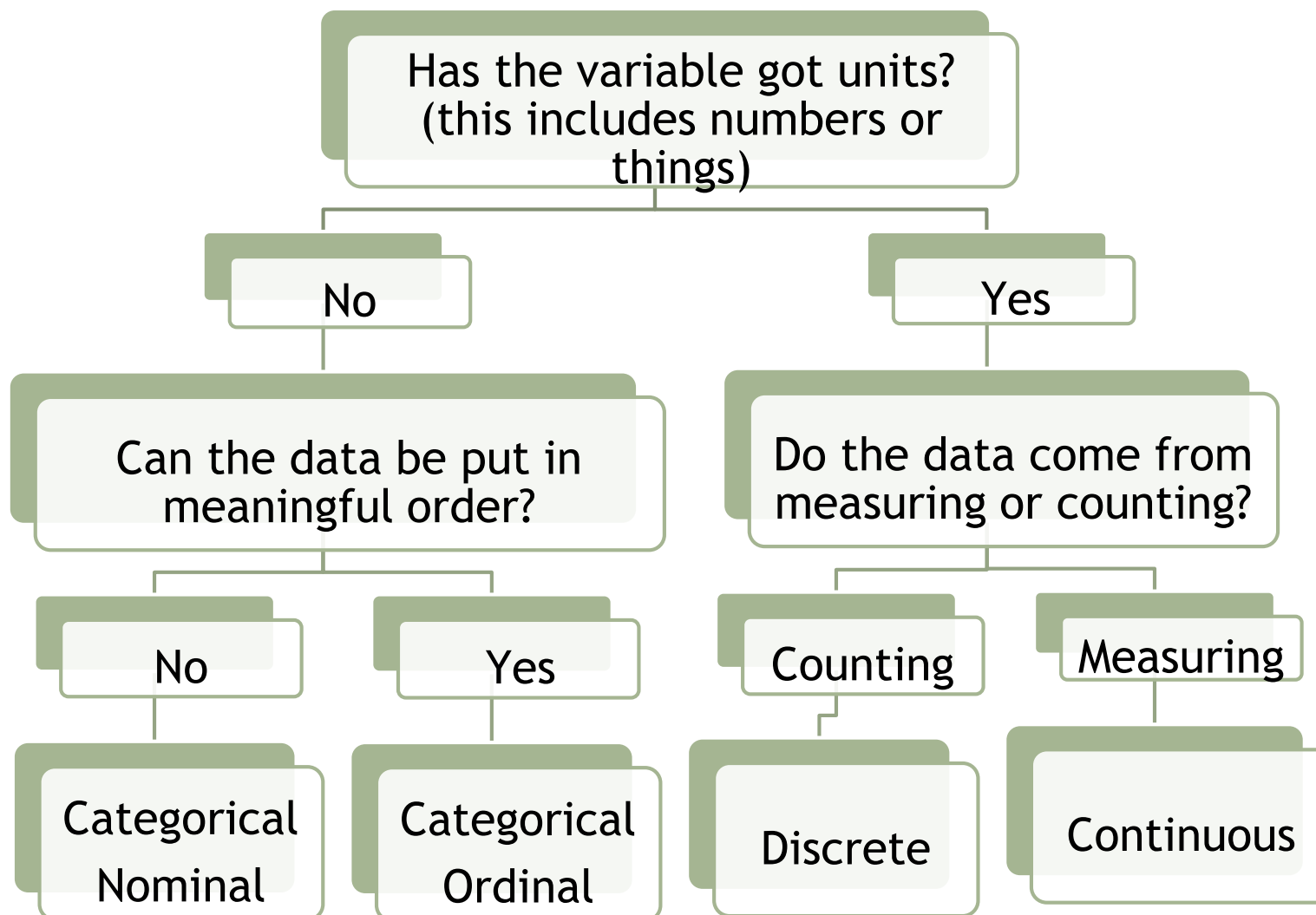
# Learning objectives



At the end of this lecture you should be able to:

- ◆ List the measures of central tendency, describe their characteristics and identify their uses
- ◆ Calculate the measures of central tendency
- ◆ List the measures of locations (Quantiles): centile, quintile, and quartile

# How can I tell what type of variable I am dealing with?





# Central Tendency

- Central tendency refers to our intuition that there is a center around which all these scores vary.

The three branches of central tendency are:

- The mean,
- The median, and
- The mode

# Measures of Central Tendency;

## The Mean

It is the average of the data divided the sum of all values of a set of observation by the number of these observations.

Calculated by this equation:

◆ Mean of *population*  $\mu = \frac{\sum X}{N}$

◆ Mean of *sample*  $\bar{X} = \frac{\sum X}{n}$



# Advantages of mean:



- ◆ Uniqueness; For a given set of data there is one and only one mean, it is single value.
- ◆ Simple to compute.
- ◆ All values are included.
- ◆ The mean is used when the data are interval or ratio scaled

## Disadvantage:

The main disadvantage of mean is the presence of extreme values, i.e. (very high or very low values).



# Measures of Central Tendency;

## The Weighted Mean:

The individual values in the set are *weighted* by their respective frequencies.

$$\bar{X}_w = \frac{\sum (n \cdot \bar{X})}{N}$$

# Weighted mean



- Let us imagine that four classes in Anatomy obtained the following mean scores on the final examination: 75, 78, 72, and 80.
  - Could you sum these four means together and divide by four, to obtain an overall mean for all four classes?
- ✓ This could be done only if  $n$  in each class is identical



# Weighted mean

- The individual values in the set are weighted by their respective frequencies.
- It can be expressed as the sum of **the mean of each group** multiplied by its respective weight (**the n in each group**) divided by the sum of the weights ( $\sum n = N$ ).

$$\bar{X}_w = \frac{\sum (n \cdot \bar{X})}{N}$$



# Weighted mean

- From the previous e.g. if , as a matter of fact, the mean of 75 is based on an  $n$  of 30, the second mean is based on 40 observations, the third on  $n = 25$ , and the fourth on  $n = 50$ .
- The total sum of scores may be obtained by multiplying each mean by its respective  $n$
- *Thus:*

$$\begin{aligned}\Sigma (n \cdot X) &= 30(75) + 40(78) + 25(72) + 50(80) \\ &= 11170\end{aligned}$$

*And by dividing it by the sum of weights (total  $N$ ), weighted mean will be obtained.*

$$= 11170 / 145 = 77.03$$



# Measures of Central Tendency;

## *The Median (50th percentile)*



### *After creating an ordered array*

- The median of a data set is the **value** that lies exactly in the *middle*.
- The **position** of the median depends on the *number* of observations
  - ❖ For **odd** number of observations:  $(n+1)/2$
  - ❖ For **even** number of observations:  
Two positions;  $(n/2)$  &  $((n/2) + 1)$
  - ❖ The value will be the mean of the two values

# Advantages of median :

- ✓ It is a single value,
- ✓ Simple, easy to compute & easy to understand,
- ✓ Unaffected by extreme values,
- ✓ The median is often used when the distribution of scores is either positively or negatively skewed
- ✓ Good with ordinal data.

## Disadvantages of median:

- ◆ It provides no information about all values (observations).
- ◆ It is less amenable than the mean to tests of statistical significance.



# Measures of Central Tendency;

## The Mode:

- It is the value which occurs most frequently.
- Data distribution with one mode is called unimodal
- If all values are different there is no mode or nonmodal.
- Sometimes, there are more than one mode.
- two modes is called bimodal; more than two is called multimodal distribution.



# Advantage of mode:

- ✓ Good for nominal data (It is the only measure of central tendency that is appropriate for nominally scaled data).
- ✓ Sometimes gives a clue about the *etiology* of the disease.

## Disadvantages of mode:

- ◆ With small number of observations, there may be no mode.
- ◆ It is less amenable to tests of statistical significance

### *Properties of the Mode:*

- Sometimes, it is not **unique**.
- It may be used for **describing qualitative data**.



Although the mean is often an excellent summary measure of a set of data, the data must be approximately normally distributed, because the mean is quite sensitive to extreme values that skew a distribution.

### *Example*

In an outbreak of hepatitis A, 6 persons became ill with clinical symptoms. The incubation periods for the affected persons ( $x_i$ ) were 29, 31, 24, 29, 30, and 25 days

$$\bar{X} = \frac{\sum X}{n} = \frac{168}{6} = 28 \text{ days}$$

If the largest value of the six listed incubation periods were 131 instead of 31, the mean would change from 28.0 to ?

$$(24+25+29+29+30+25+131)/6 = 44.7 \text{ days}$$



# Examples

## What about the Median & the Mode?

### Finding the Median

#### Position of the Median:

1. Arrange data in order (24, 25, 29,29,30,31)
2. Find position of the median; in even no.= $n/2$  &  $(n/2)+1$ , (observations no.3 &4)
3. The value of the median is the average of the **TWO VALUES (29)**

#### Finding the Mode:

- The most frequent observation
- **Mode = (29)**

- ▶ if the largest value of the six listed incubation periods were 131 instead of 31, what will happen to the **Median & the Mode?**

- **The Median will remain the same**
- **The Mode will remain the same in this example**







# Choosing a Measure of Central Tendency

IF variable is Nominal..

- ◆ Mode

IF variable is Ordinal...

- ◆ Mode or Median(or both)

IF variable is Interval-Ratio and distribution is Symmetrical...

- ◆ Mode, Median or Mean

IF variable is Interval-Ratio and distribution is Skewed...

- ◆ Mode or Median

# Measures of locations; Quantiles



- ◆ **Quantile** is a value below which a certain proportion of observations occurred in the ordered set of data values.
- ◆ **Quantile** is defined as equal-sized segments of a population.

# Measures of locations; Quantiles



- ▶ One of the most common metrics in statistical analysis, the median, is actually just the result of dividing a population into two quantiles.
- ▶ A population split into three equal parts is divided into tertiles

# Quintiles



- ◆ A quintile is a statistical value of a data set that represents 20% of a given population
- ◆ The first quintile represents the lowest fifth of the data (1-20%)
- ◆ The second quintile represents the second fifth (21% - 40%) and so on.



# Centiles

- ◆ Those values, in a series of observations arranged in ascending order of magnitude, which divide the distribution into 100 equal parts.
- ◆ 10th Percentile: it is the value below which 10% of the observations lie.
- ◆ We also frequently used 3rd, 97th, and the 50th (median) percentile.



# Quartiles

- ◆ These are the observations in an array that divide the distribution into four equal parts.
- ◆ **1st (lower Quartile):** the value below which 25% of observations lie in an ordered array

# Quartiles



- ◆ 2nd quartile = Median = 50th percentile
- ◆ Upper Quartile = 75th percentile
- ◆ Interquartile Range: is the middle 50% of all observations

## To summarize

### Measures of central tendency are:

- The mean, and weighted mean
- The median
- The mode

### Measures of locations :

- Centile
- Quintile
- Quartile







**THANK YOU**