



## Department of biology



## Department of biology

((Biostatistics))

1<sup>st</sup> stage

## chapter\_2

By

Mm. Ali Alawadi



# Department of biology



## Chapter Two

### Statistical Metrics

Statistical measures are divided into two parts:

1. Measures of central tendency
2. Measures of dispersion and variation

First: Measures of Central Tendency

*Arithmetic mean: Average or Mean*

$$\bar{X} = \frac{\Sigma x}{n}$$

It is one of the most important measures of central tendency and the most used in statistics and in practical life.

Way of calculating:

1. From unclassified data:

**Example 1:** Find the arithmetic mean of the following observations representing the level of hemoglobin in the blood to 6 men in mg/dL

$$X_i = 11, 12, 13, 12, 13, 11$$

Sol: -

$$\bar{X} = \frac{11 + 12 + 13 + 12 + 13 + 11}{6}$$

$$\bar{X} = \frac{72}{6} = 12$$



## Department of biology



**Example 2:** If the median follicle growth stimulating hormone is 18 mg/dL, where the level of follicle growth stimulating hormone in the first female rabbit is 18, the second 19, the third 17, and the fourth 19, find the hormone level in the fifth female rabbit:

Sol: -

$$\bar{X} = \frac{\sum x}{n}$$

$$18 = \frac{18 + 19 + 17 + 19 + \bar{X}5}{5}$$

$$18 = \frac{73 + \bar{X}5}{5}$$

$$90 = 73 + \bar{X}5$$

$$\bar{X}5 = 90 - 73 = 17$$

### Exercises:-

- ❖ If you know that the blood sugar level of eight men is as follows:  
10, 21, 15, 8, 24, 17, 5, 42 Calculate the arithmetic mean?
- ❖ If the mean of the marks of 3 students is 9, where the mark of the first student = 6 and the mark of the second student = 7, find the third student?



## Department of biology



### Example 3:-

If  $y_1, y_2, \dots, y_n$  represent the centers of the categories in the frequency distribution table with their frequencies  $f_1, f_2 \dots$ . Unwind. Respectively, the arithmetic mean is:

$$\bar{Y} = \frac{\sum^k f_i y_i}{\sum^k f_i}$$

تكرار × مركز الفئات ( $f_i \times y_i$ )	مراكز الفئات $y_i$	التكرارات $F_i$	الفئات $F_i$	تسلسل الفئات
42	14	3	13 – 15	1
85	17	5	16 – 18	2
240	20	12	19 – 21	3
230	23	10	22 – 24	4
104	26	4	25 – 27	5
29	29	1	28 – 30	6
$\sum y_i f_i = 730$		$\sum f_i = 35$		

To find the mean of classified data, we follow these steps:

1. Specifies the category center  $y_i$
2. Multiply the center of each class by the amount of its frequency (  $f_i, y_i$  )
3. Dividing the sum (**product of the center of each category × its repetition**) by the sum of the iterations.

$$\bar{Y} = \frac{\sum f_i y_i}{\sum f_i} = \frac{730}{35} = 20.86$$



## Department of biology



### Features of the arithmetic mean:

1. Scale is easy and easily subjected to algebraic operations.
2. It takes into account all the values under study.
3. It is the most understood metric in statistics.

### Disadvantages of the arithmetic mean:

1. It is influenced by outliers (which are very large and very small values).
2. It is difficult to calculate in the case of metadata.
3. It is difficult to calculate in the case of open frequency tables as it requires knowledge of the center of each category.

### *Arithmetic median*

When the data is arranged ascending or descending, the median is the value that falls before 50% of the views in the ranking and 50% of the views after it in the order, that is, it is the value that divides the data into two equal parts (the view that falls in the middle).

### Finding the median:

The values are arranged ascending or descending if the number of values (n) is odd,  $\frac{n+1}{2}$  the median is the value that is arranged, and if the number of values is even, the median is the arithmetic mean of the two values that are arranged  $\frac{n}{2}$  or  $\frac{n}{2+1}$

**Example 1:** Find the median for the data representing the level of hemoglobin in the blood of 7 men mg/dL

$$X_i = 11, 12, 13, 12, 13, 11, 14$$



## Department of biology



Sol:- We arrange the data in ascending order 11,11,12,12,13,13,14

Finding the median: Since the number of values (n) is odd,

$$\text{Me} = \frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$$

$$\text{Me} = 12$$

**Example 2:** Find the median for the data representing the level of hemoglobin in the blood (8) men mg/dL.

$$X_i = 11, 12, 13, 12, 13, 11, 14, 10$$

Sol: Sort data in ascending order 10, 11, 11, 12, 12, 13, 13, 14

Since the number of values (n) is even, the median is the arithmetic mean of the values that are arranged  $\frac{n}{2}$

$$X_i \frac{8}{2} = 4$$

$$\text{Me} = 12$$

**Example 3:** If you know that the following data represents the pH of the blood of people with viral hepatitis.

6.5 , 6.0 , 7.2 , 6.0 , 7.4 , 6.8 , 6.7 , 7.0 , 7.3

We arrange the data in ascending order as follows:

6.0 , 6.0 , 6.5 , 6.7 , 6.8 , 7.0 , 7.2 , 7.3 , 7.4

$$X_i \frac{n+1}{2}$$

$$X_i \frac{9+1}{2} = 5$$

that the median value is (6.8)



## Department of biology



### Exercise:-

Find the median from the following (7 views):

16 , 12 , 9 , 17 , 12 , 8 , 15

### Median features:

1. It is not affected by outliers.
2. It can be calculated in the case of metadata that can be arranged.
3. It can be calculated in the case of open frequency tables of quantitative data.

### Disadvantages of the median:

1. It does not take all values into account when calculating it.
2. It is not easy to deal with in statistical and mathematical analysis.

### *Mode*

The most common or frequent value or values among the dataset are called Mode or Common

**Example1:** Find the pattern in the following sample:

6, 3, 7, 5, 7, 8, 6, 7

Answer: The value (7) is the most frequent in the sample

$$Mo=7$$

**Example 2:** Find the following mode:

4, 5, 8, 1, 2, 5, 6, 2, 9

Answer: the two values (2 and 5).

$$Mo = 5 \text{ \& } 2$$

From this it is clear that it may be one mode or have two modes (two values) and may have more than two modes as there may be no mode of observations.



## Department of biology

### Features of the mode:

1. It is a scale that is easy to calculate.
2. It can be found for quantitative data, descriptive data, and open frequency distributions.

### Mode disadvantages:

1. Does not take all values into account when calculating
2. Some data may have more than one mode, so it is not possible to specify a single value for the mode.
3. Some data doesn't have a pattern