

## Statistics

## الإحصاء

الإحصاء هو احد فروع الرياضيات التطبيقية الذي يختص بجمع المعلومات والحقائق، ويعرف بمجموع الطرق والوسائل والقوانين المستخدمة في التحليل ويعتبر وسيلة لقياس وتحليل الظواهر والحقائق واستخلاص النتائج ووضعها بصورة مناسبة لتوضيح العلاقة بينها، وهو وسيلة وليس غرضاً يساعد الباحثين في العلوم الاخرى للوصول الى النتائج بأقصر الطرق وأقلها كلفة.

مثال على ذلك التقييم الاحصائي للنتائج الكيميائية: إن النتائج للتحليل الكيميائي عامة تكون عديمة الفائدة ما لم تكن مقيمة احصائياً، وعند قياس اي خاصية فيزيائية لابد أن يكون هناك خطأ محتمل في قياسها ويمكن تقليل هذا الخطأ الى حد مقبول ولكن لا يمكن تلافيه تماماً.

### **Definition: (Statistics)**

It is a branch of mathematics dealing with the collecting, summarizing, presenting and analyzing data which measure numerically, this analysis may lead to conclusions and subsequent decisions.

**There are two types of statistics:**

Statistics → { Descriptive Statistics (الاحصاء الوصفي)  
Inferential Statistics (الاحصاء الاستدلالي)

### **(1) الاحصاء الوصفي (Descriptive Statistics):**

ويتم الاعتماد على هذا النوع لوصف مجموعة من البيانات (Data) على شكل عينة (Sample) من مجتمع (Population)، وذلك عن طريق حساب قيم خاصة مثل الوسط الحسابي والوسيط والانحراف المعياري، وايجاد هذه المعلومات والتوصل اليها ينتج استيعاب بيئة العينة التي تم اجراء الدراسة عليها.

### **(2) الاحصاء الاستدلالي (Inferential Statistics):**

يحفز هذا النوع الباحث للوصول الى المعلومات الاحصائية وذلك عن طريق الاستدلال والاستفسار عن خصائص العينة والتوزيع الاحصائي لبيانات العينة، ويتوافق تطبيق هذا الاحصاء اذا كانت البيانات المستقطبة يراد استخدامها كفرضية.



### Definition: (Data)

The numbers or values which represent any phenomenon (ظاهرة) in life are called data.

The data may be divided into two parts:

1. Ungrouped Data (بيانات غير مبوبة): which means, the values or numbers which is getting it for any phenomenon forms resources, population and institutes.
2. Grouped Data (بيانات مبوبة): is the data putting in table and the table contains two columns (classes, frequency).

التكرار      الفئات

### Example:

- 1) The density (كثافة) of steal (الفولاذ): 30, 47, 51, 82, 23, 35, 62

This kind of data for a various phenomena is ungrouped data.

- 2) Tabular presentation (frequency distribution table) is grouped data

العرض الجدولي (جدول التوزيع التكراري):

Classes	Frequency ( $f_i$ )	Center Classes ( $x_i$ )
$L_1$ 10 – 19 $u_1$	10	14.5
$L_2$ 20 – 29 $u_2$	17	24.5
$L_3$ 30 – 39 $u_3$	12	34.5
$L_4$ 40 – 49 $u_4$	7	44.5
$L_5$ 50 – 59 $u_5$	8	54.5
	$\sum_{i=1}^5 f_i = 54$	

To find the center classes, apply the following formula:

$$\text{Center classes} = \frac{\text{Lower limits} + \text{Upper limits}}{2}, \quad x_i = \frac{L_i + u_i}{2}$$

$$x_1 = \frac{L_1 + u_1}{2} = \frac{10 + 19}{2} = \frac{29}{2} = 14.5$$

## Measures of Central Tendency

## مقاييس النزعة المركزية

### 1. Mean (Average): الوسط الحسابي (المعدل)

Let  $x_1, x_2, \dots, x_n$  be a set of data (results) represent a finite sample of size (n), the sample mean which denoted by ( $\bar{x}$ ) is:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \text{ for ungrouped data}$$

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}, \text{ for grouped data}$$

Where  $x_i$ : is the center classes

$f_i$ : is a frequency

**Example:** Find the mean for the following measurements

7, 13, 16, 20, 23, 27, 29

**Solution:**

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^n x_i}{n} = \frac{7+13+16+20+23+27+29}{7} \\ &= \frac{135}{7} \\ &= 19.285\end{aligned}$$

**Example:** If you have the following data

1, 4, 31, 32, 33, 34, 36, 40

Find the average.

**Solution:**

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^n x_i}{n} \\ &= \frac{1+4+31+32+33+34+36+40}{8} = \frac{211}{8} = 26.375\end{aligned}$$



**Example:** If you have the frequency distribution table which represents the temperatures of the weather at 20 days in Baghdad.

Classes: 20-24 25-29 30-34 35-39 40-44

Frequency: 1 2 6 5 6

Find the mean.

**Solution:**

Classes	Frequency ( $f_i$ )	$x_i$	$f_i x_i$
20 – 24	1	22	22
25 – 29	2	27	54
30 – 34	6	32	192
35 – 39	5	37	185
40 – 44	6	42	252
	$\sum_{i=1}^5 f_i = 20$		$\sum_{i=1}^5 f_i x_i = 705$

$$x_i = \frac{L_i + u_i}{2}, \text{ for } x_1 = \frac{L_1 + u_1}{2} = \frac{20 + 24}{2} = \frac{44}{2} = 22.$$

$$\bar{x} = \frac{\sum_{i=1}^5 f_i x_i}{\sum_{i=1}^5 f_i} = \frac{705}{20} = 35.25, \text{ The mean (average) of temperature in Baghdad at 20 days.}$$

## 2. Median: الوسيط

Suppose that  $x_1, x_2, \dots, x_n$  is a finite sample of size (n), then the median which denoted by (Me) is:

- 1) For ungrouped data, arrange  $x_1, x_2, \dots, x_n$  in increasing or decreasing order of magnitude, say  $y_1, y_2, \dots, y_n$ . نرتب البيانات تصاعدياً

Then, if (n) is odd (فردى)

$$Me = y_{\frac{n+1}{2}}$$

If (n) is even (زوجي), then

$$Me = \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2}$$

2) For grouped data, applying the following express:

$$Me = L_i + \left[ \frac{(\sum_{i=1}^n f_i)/2 + F_i}{f_i} \right] \cdot c$$

Where,

$L_i$ : is the lower limit for the median class.

$F_i$ : is the cumulative frequency at the start of median class.

$f_i$ : is the median class frequency.

$c$ : is the median class length.

### **Example:**

(1) Find the median for the following data:

13, 20, 7, 16, 23, 29, 27

### **Solution:**

7, 13, 16, 20, 23, 27, 29      نرتب البيانات تصاعدياً

$n = 7$  (Odd)

$$Me = y_{\frac{n+1}{2}} = y_{\frac{7+1}{2}} = y_{\frac{8}{2}} = y_4 = 20$$

(2) The speed for many computers are:

317, 469, 375, 444, 216, 392, 285, 469

Find the median for this data.

### **Solution:**



216, 285, 317, 375, 392, 444, 469, 496

نرتب البيانات تصاعدياً

$n = 8$  (Even)

$$Me = \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2}$$

$$= \frac{y_{\frac{8}{2}} + y_{\frac{8}{2}+1}}{2}$$

$$= \frac{y_4 + y_5}{2}$$

$$= \frac{375 + 392}{2}$$

$= 383.5$  , The median.

★ (3) Find the median for the following results:

Classes	$f_i$
40 – 49	5
50 – 59	15
60 – 69	22
70 – 79	28
80 – 89	10

**Solution:**

Classes	$f_i$	C.F	$F_i$
40 – 49	5	Less than 49	5
50 – 59	15	Less than 59	20
60 – 69	22	Less than 69	42
70 – 79	28	Less than 79	70
80 – 89	10	Less than 89	80
	$\sum_{i=1}^5 f_i = 80$		

$$\text{Median order (رتبة الوسيط)} = \frac{\sum_{i=1}^5 f_i}{2} = \frac{80}{2} = 40$$

$$F_i = 20, \quad f_i = 15, \quad c = 10, \quad L_i = 50$$

$$Me = L_i + \left[ \frac{(\sum_{i=1}^n f_i)/2 + F_i}{f_i} \right] \cdot c$$

$$= 50 + \left[ \frac{40-20}{15} \right] \cdot 10$$

$$= 50 + \frac{20}{15} \cdot 10$$

$$= 50 + \frac{40}{3}$$

$$= 63.3333, \text{ The median}$$

### 3. Mode: المنوال

Assume that  $x_1, x_2, \dots, x_n$  be a finite set of sample of size (n), then the mode which denoted by (Mo) is:

- 1) For ungrouped data, the mode is the value repeated more than the other values. The data may contain more than one mode, or the data may not contain mode in it.
- 2) For grouped data, using the following express:

$$Mo = L_i + \left[ \frac{f_i - f_{i-1}}{2f_i - f_{i-1} - f_{i+1}} \right] \cdot c$$

Where,

$L_i$ : is the lower limit for mode class.

$f_i$ : is the frequency for mode class (greatest frequency).

$c$ : is the length for mode class.



**Example:**

(1) Find the mode for the following samples:

A.  $x_i: -2, 11, 4, -3, 7, 13$

B.  $x_i: 9, 10, 5, 9, 7, 8, 6, 9, 10, 4$

C.  $x_i: 2, 0, 3, 1, 2, 4, 2, 5, 4, 1, 0, 4$

**Solution:**

For (A) there no mode exists.

For (B)  $Mo = 9$ .

For (C)  $Mo = 2, 4$ .

(2) Consider the following frequency distribution table which represents particles (جسيمات) in five levels (مستويات) of energy (الطاقة).

الفئة التي تقابل اكبر تكرار

Classes: 40-49 50-59 60-69 70-79 80-89

Frequency: 5 15 20 30 10

Find the mode. اكبر تكرار

**Solution:** The greatest frequency is 30, then

$$f_i = 30, \quad f_{i-1} = 20, \quad f_{i+1} = 10$$

The mode class is 70 - 79,  $L_i = 70$ ,  $c = 10$ .

$$Mo = L_i + \left[ \frac{f_i - f_{i-1}}{2f_i - f_{i-1} - f_{i+1}} \right] \cdot c$$

$$= 70 + \left[ \frac{30-20}{2(30)-20-10} \right] \cdot 10$$

$= 73.333$ , is the mode for the five levels of energy.



### H.w

- 1) If you have the following ungrouped data which represents a calibration to measure sodium ratio:

25, 28, 7, 13, 16, 20, 12, 7

Find the mean, median and mode.

- 2) This measures ratio was calibrated and the results were found as follows:

6.87, 6.84, 6.88, 6.85, 6.83, 6.88

Find the average, median and mode.

- 3) Calibration of potassium nitrate concentration Mg/L was performed by using uv-vis and the data were found as follows:

Classes: 1.5-1.9 2-2.4 2.5-2.9 3-3.4 3.5-3.9 4-4.4 4.5-4.9

Frequency: 2 1 4 15 10 5 3

Find the mean, median and mode.

- 4) The following results represent 40 measurements of Na:

Classes: 7-10 11-14 15-18 19-22 23-26 27-30

Frequency: 4 6 8 12 9 5

Find the mean, median and mode.