Statistical methods for presenting and displaying data

It is the next step after collecting data in descriptive statistics, and there are two ways to share data:
1. Display data tabularly.
2. Display data graphically.
Data can be displayed in the form of a table, and the form of this table varies depending on the type and number of data. Below is the display of various data (descriptive or quantitative) in the form of a simple table.

If we are studying a phenomenon that contains one descriptive variable, its data can be presented in the form of a simple frequency table, which is a table consisting of two columns, one of which contains the categories (groups) of the variable, and the second contains the number of items (frequencies) for each class (group).
The following example shows us how raw descriptive data can be tabulated in the form of a frequency table.

Below is a group of grades for students in one of the secondary school classes, as follows

12	10	5	8	15	5	2	8	10	5
10	12	12	2	5	2	8	10	5	10

1. What is the type of variable?
2. Display the data in the form of a frequency table.
3. Create the percentage frequency distribution.
4. Comment on the results.

The solution
1- Type of variable is interval (quantitative).
2- Display the data in the form of a frequency table:
To reduce the time of finding information about the unclassified data mentioned above, we can find important statistics for this data by creating a frequency table of the data. When f_i represents the frequency of observation i, the frequency table will be as follows:

Marks (x_i)	Number of students (f_i)
2	4
5	5
8	3
10	5
12	2
15	1
	$\Sigma f_i = 20$

3 - percentage frequency distribution:

percentage frequency is calculated by dividing the group frequency by the sum of the frequencies.

Example 2

Below are data on the educational level of a sample of 50 individuals.

Elementary	**Middle school**	**Post graduate**	**secondary school**	**Middle school**	**secondary school**	**read and write**	**Middle school**
Middle school	**Elementary**	**secondary school**	**Middle school**	**secondary school**	**secondary school**	**Middle school**	**read and write**
secondary school	**read and write**	**Elementary**	**secondary school**	**university**	**read and write**	**secondary school**	**Elementary**
Middle school	**university**	**Middle school**	**Elementary**	**secondary school**	**Middle school**	**Elementary**	**Middle school**
Elementary	**secondary school**	**Elementary**	**read and write**	**secondary school**	**Elementary**	**Middle school**	**secondary school**
secondary school	**secondary school**	**Post graduate**	**university**	**Elementary**	**university**	**secondary school**	**university**
						read and write	**Middle school**

What is required: Display the data in the form of a frequency table.

The solution

1- Display the data in the form of a frequency table:

Educational classes (reads and writes - Elementary - Middle school - secondary school - university - Post graduate) is a descriptive ordinal variable, and the above data can be presented in the form of a frequency table by following:

• Configure the frequency table:

Frequency distribution of a sample size of 50 individuals according to educational level

Educational Class	*Number of individuals (frequencies)* f_i	Percentage frequency distribution
read and write	6	12%
Elementary	10	20%
Middle school	12	24%
secondary school	15	30%
university	5	10%
Post graduate	2	04%
Sum	50	100%

Example 3

Below are data on the scores of 70 students on the final exam for the statistics course.

56	65	70	65	55	60	66	70	75	56
60	70	61	67	61	71	67	62	71	66
68	72	57	68	72	69	57	71	69	75
72	62	67	73	58	63	66	73	63	65

58	73	74	76	74	80	81	60	74	58
76	82	77	83	77	85	91	78	94	72
79	64	57	79	55	87	64	88	78	62

1 - Create the frequency distribution of students' grades.

2- Create the percentage frequency distribution.

The solution

1- Configure the frequency distribution:

The student's score on the test is a continuous quantitative variable, and in order to tabulate the data in the form of a frequency table, the following is followed:

• Range calculation

Range = Maximum – Minimum

R = 94 - 55 = 39

• Determine the number of classes:

The number of categories is determined according to considerations including: the researcher's opinion, the goal of the research, and the size of the data. Many researchers believe that the best number of categories should range between 5 to 15, assuming that the number of categories is 8, meaning: (C=8).

• Calculate the length of the class:

• Define categories:

The class begins with a value called the lower limit, and ends with a value called the upper limit, and then we find that:

The minimum for the first class is the lowest reading (score), meaning that the minimum for the first class = 55

The upper limit of the first class = the lower limit + the length of the class = 55 + L = 60 = 55 + 5

So the first class is: "55 to less than 60" and it reads "from 55 to less than 60"

The minimum limit for the second class = the maximum limit for the first class = 60

The upper limit of the second class = the lower limit of the class + the length of the class = 65 = 60 + 5

So the second class is: "60 to less than 65" and it reads "from 60 to less than 65"

In the same way, the boundaries of other categories are formed, which are:

The third class: 65 to less than 70.

The fourth class: 70 to less than 75

Fifth class: 75 to less than 80.
Sixth class: 80 to less than 85
Seventh class: 85 to less than 90.
Eighth class: 90 to less than 95

• Configure the frequency table:

Frequency distribution of 70 students according to their scores in the statistics course test

Classes	*Number of individuals (frequencies)* f_i	Percentage frequency distribution
55 – 60	10	14.3
60 – 65	12	17.1
65 – 70	13	18.6
70 – 75	16	22.9
75 – 80	10	14.3
80 – 85	4	05.7
85 – 90	3	04.3
90 – 95	2	02.8
Sum	70	1.00

Graphical presentation of quantitative data

Graphical display of data is one of the methods that can be used to describe data, in terms of the form of distribution and the extent of data concentration. In many applied aspects, graphical display is easier and faster in describing the phenomenon under study. Methods of displaying data graphically differ according to the type of data classified in the form of a table. Iteratively, the following is a presentation of the different graphical shapes.

Histogram

The histogram is the graphical representation of a simple frequency table for continuous quantitative data. It consists of adjacent graph columns, where the frequencies are represented on the vertical axis, while the values of the variable (category boundaries) are represented on the horizontal axis. Each category is represented by a column, the height of which is the frequency of the category. The length of its base is the length of the category.

Below is the frequency distribution of the weights in grams of a sample of poultry, size 100, selected from a farm after 45 days.

weights	600-	620-	640-	660-	680-	700-720	Sum
Number of chickens	10	15	20	25	20	10	100

1 - How long is the category?

2- Draw the histogram.

3- Draw the relative histogram, then comment on the drawing.

The solution

1- Category length (L)

If class length = 20

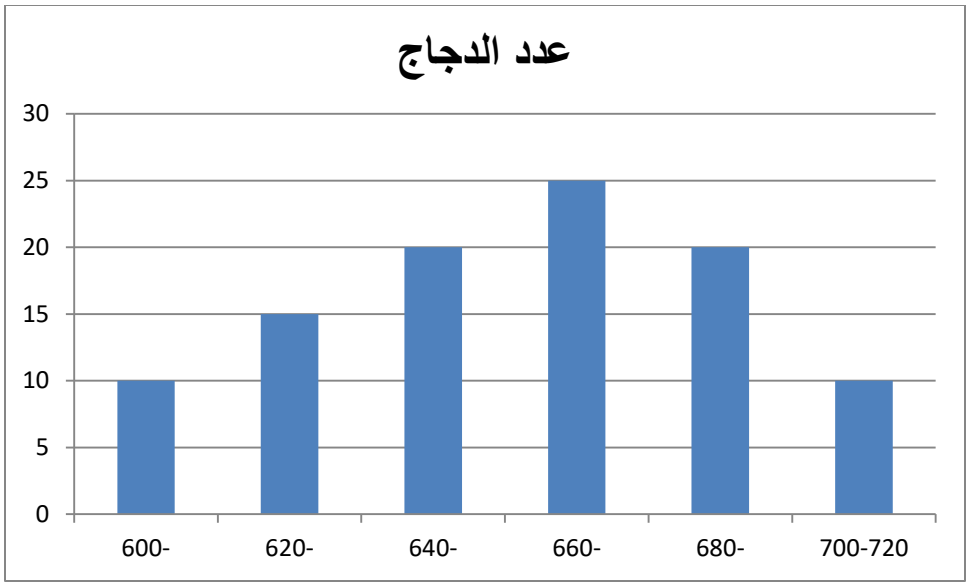
2- Draw the histogram.

To draw the histogram, the following steps are followed:

· Draw two perpendicular axes, the vertical represents frequencies, the horizontal represents weights.

· Each category is represented by a column whose height is the frequency of the category, and whose base length is the length of the category.

· Each column begins where the previous category column ended.



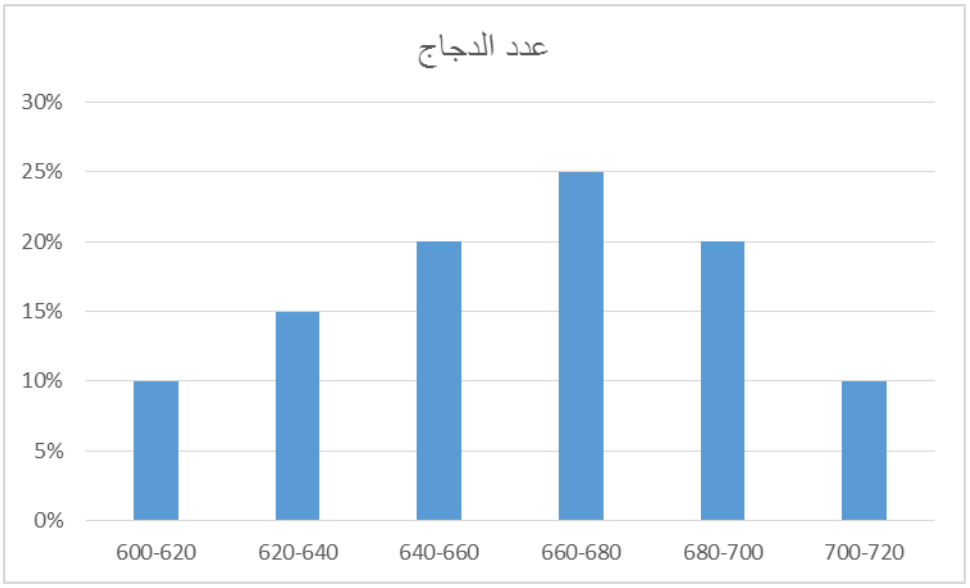
3- Drawing the relative histogram: To draw the relative histogram, do the following:

· Calculating relative frequencies.

weights	600-	620-	640-	660-	680-	700-720	Sum
Number of chickens	10	15	20	25	20	10	100
Percentage frequency	10	15	20	25	20	10	100

· By following the same previous steps when drawing the histogram, the relative histogram is drawn, replacing the relative frequencies with the absolute frequencies on the vertical axis, as shown in the following figure:

Relative histogram of weights for a sample of 100 chickens



From the figure above it is noted that:

· 25% of chickens weigh between 660 and 680 grams, which is the largest percentage.

· The shape is skewed to the left, which indicates that the distribution of chicken weights is negatively skewed.

Notes on the histogram format

A- The area under the histogram is equal to the sum of the frequencies (n).

B- As for the area under the relative frequency histogram, it expresses the sum of the relative frequencies, and it is equal to the correct one.

T- It is possible to estimate the common values, which are the values that correspond to the largest height. In the previous two figures, we find that the common weight falls in the category (660-680) and is called the mode.

Frequent polygon

It is also a graphical representation of a simple frequency table, where the frequencies are represented on the vertical axis, the centers of the categories on the horizontal axis, then the coordinates are connected with broken lines, and then the two ends of the polygon are connected to the horizontal axis.

The category center is the value that falls in the middle of the category, and is calculated by applying the following equation:

Since the actual frequency values for each category are not known, the category center is considered the appropriate estimate of the value of each item in the category.

Example 5

Use the frequency table data in Example (2-4) to draw the frequency polygon.

The solution

To draw the iterative polygon, do the following:

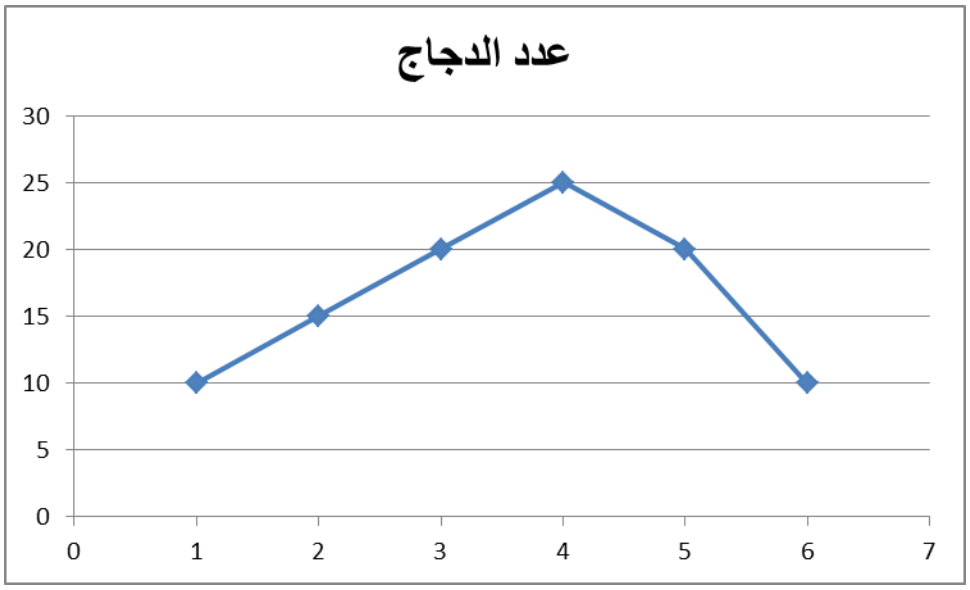
weights	Number of chickens	Middle of class x
600-	10	(600+620)/2= 610
620-	15	(620+640)/2=630
640-	20	650
660-	25	670
680-	20	690
700-720	10	(700+720)/710
Sum	100	

.

Middle of class x	590	610	630	650	670	690	710	730
Number of chickens y	0	10	15	20	25	20	10	0

· Graphical representation of coordinate points and connecting them with straight lines, as shown in following figure.

Frequency polygon of the weights of a sample of chickens with a size of 100 chickens



Frequency curve

By following the same previous steps in drawing a polygon, the iterative curve can be drawn, but the broken lines are smoothed into a curve so that it passes through the largest number of points. In the previous example, the iterative curve can be drawn, and following figure shows this figure.

Frequency curve of the weights of a sample of 100 chickens