

- **Statistics science:** it is a science that employs the use of a scientific method of data collection methods, compiled, summarized, displayed and analyzed in order to reach them on the conclusions and the occasion decisions .

It can be divided the statistics into two main sections :

- 1.Descriptive statistics : It includes statistical methods that used to describe a group set of data, including statistical methods on the collection of data in form of numerical measurements methods and tabulation or organizing , summarizing , presenting and account of some different statistical measures .
- 2.Statistical inference: includes statistical methods that designed to work conclusions or inferences about the source of the data were collected from this section .There are two major branches :
 - (A).Estimate is interested to find a discretionary values to deduce them on the real values of the source of the data collection and the estimated values are either any specific at a certain point in point estimation or recognition or interval estimation .
 - (B).Test of hypotheses: includes hypotheses that set an explanation of the initial phenomenon to be studied to get them on the decision of acceptance or refusal to test .

-**Variable** : It is a phenomenon where appeared differences between the vocabulary show and symbolized by the symbol (y) or any symbol such as (x or z.). For example is length of the students in universities are considered variable symbolized by the symbol (y) and the length of any student is symbolized by (y_i) .These are called observation. The variables are divided into the following :

- 1.Qualitative variables: It is namely those phenomena or qualities that can not be measured directly in numerical numbers such as eye color (Blue, Black, Brown) and social status (poor, middle status, rich) and gender (female, male). . . etc .
- 2.Quantitative variables: It is namely those phenomena or traits that can be measured directly numerical numbers such as recipe height, weight, age and the amount of the crop. . . etc . The quantitative variables are divided into two parts:
 - (A).Continuous variables: It is continuous variable that takes viewing in numeric value in a certain extent if we assume that length students at the University that are between 130 cm and 170 cm,the range of length recipe is $(130.5) \geq y \geq (.170.0)$. Examples of continuous variables are weight, age and the amount of the crop, temperature and time .
 - (B).Discrete variables : The observation that takes a spaced valuable or intermittent or separate , if we assume that the members number in four families is 2, 3, 4.5 respectively, $y = 2, 3, 4, 5$. Examples of discrete variables is constant is the number of fruits in the plants and the number of production units in the factory and the number of students in the early stages of a university.

- **Functions of statistics :**

- 1.Description data .
- 2.Statistical inference .
- 3.Forecasting .

-Samples can be divided according into two types :

1.Probability samples: The samples are selected vocabulary according to the possibilities rules, It means are chosen vocabulary of the population study random method in order to avoid bias that resulting from the choice of samples. The most important types of probability samples are :

- a. Simple random sample .
- b. stratified random sample .
- c. Random sample of regular systematic sample .
- d. Cluster sampling .

2.Non-probability sample: that is selected vocabulary in a non-random sample such as the farms that produce dates from the type of diabetes .

- **Community**: It is values that can be taken lengths of students at the University of the community in this case are the lengths of all the students at that university .There are two types of community which are finite population of any possible infinite number of such vocabulary students lengths in Mosul University or the number of production units in the factory on a particular day and infinite population is difficult to count the number of vocabulary, such as the type of fish in the Tigris River and account of bacteria in the field .

-**Sample**: It is groups of observation were chosen in some way from the community and the sample are part of the community they represent the properties of the original community that took him to this sample .

- Statistical symbols:

Mathematical mean (\bar{Y})	lower or equal (\geq)	total (Σ)
arithmetic mean of the community (μ)	greater or equal (\leq)	absolute value ()
Geometric mean (\bar{G})	percentage (%)	median (\bar{M}_e)
harmonic mean (\bar{H})	repetition (f_i)	mode (\bar{M}_o)
quadratic mean (\bar{Q})	level of significance (α)	Coefficient of Variation (CV)
standard deviation of the sample (SD)	Sum of Squares (SS or S^2)	standard deviation of the community (σ)
Mean of squares. (MS)	standard error (SE)	standard deviation (S)

- Statistical terms :

1. Ungrouped data: It is an initial or original data (raw data) was collected and not classified.
2. Grouped data: It is data that organized in the frequency distribution table .
3. The classes: It aggregates that have divided them variable values and every category take a certain range of variable values.
4. The class limits: There are found for each category two limits which are lower class limit and upper class limit.

5. The class boundaries or real class limit : There are found for each category two limits boundaries which are lower class boundary and upper class boundary.
6. The class length: It is amount of range between two category and preferably that groups lengths are equal to facilitate the operation calculations and symbolizes the category length is symbol (c).
7. The category center or class mid-point : There is a found for each category is center and has the symbol (y_i) and is a mid-range between category boundaries .
8. The class frequency : It is a number of values that lied in that category range and symbolized by the symbol (f_i) and the sum of the duplicates must always be equal to the total number of the values of the phenomenon.

- Solved examples

Example (1): the value of variables (X_i) , (Y_i) are in following :

$$X_i = 1, 2, 3 . \quad Y_i = 2, 4, 6 .$$

find the following :

$$(a)-\sum_{i=1}^n y_i , (b)-\sum_{i=2}^3 y_i , (c)-\sum y_i^2 , (d)-(\sum y_i)^2 , (e)-\sum x_i y_i , (f)-(\sum x_i) (\sum y_i) .$$

Solution :

$$(a) \sum_{i=1}^n y_i = y_1 + y_2 + y_3 = 2 + 4 + 6 = 12 .$$

$$(b)-\sum_{i=2}^n y_i = y_2 + y_3 = 4 + 6 = 10 .$$

$$(c)-\sum y_i^2 = y_1^2 + y_2^2 + y_3^2 = 2^2 + 4^2 + 6^2 = 4 + 16 + 36 = 56 .$$

$$(d)-(\sum y_i)^2 = (y_1 + y_2 + y_3)^2 = (2 + 4 + 6)^2 = (12)^2 = 144 .$$

$$(e)-\sum x_i y_i = (x_1 y_1) + (x_2 y_2) + (x_3 y_3) = (1 \times 2) + (2 \times 4) + (3 \times 6) = (2) + (8) + (18) = 28 .$$

$$(f)-(\sum x_i) (\sum y_i) = (x_1 + x_2 + x_3) (y_1 + y_2 + y_3) = (1 + 2 + 3) (2 + 4 + 6) = (6) (12) = 72 .$$

Example (2): the value of variables (X_i) , (Y_i) are in following :

$$X_i = 1, 2, 3 . \quad Y_i = 2, 4, 6 . \quad \text{find the following :}$$

$$(a)-\sum (y_i - x_i)^2 , (b)-\sum (x_i - 3)(y_i - 5) , (c)-\sum x_i y_i^2 , (d)-\sum (y_i - 3) , (e)-\sum y_i - 3 , (f)-\sum (x_i + 2)/y_i .$$

$$(g)-\sum (x_i + 2)/\sum y_i , (h)-\sum y_i^2 - (\sum y_i)^2/n , (i)-\sum x_i y_i - \sum (x_i)(y_i)/n .$$

Solution :

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	2	2	1	4
2	4	8	4	16
3	6	18	9	36
Sum = 6	12	28	14	56

$$(a)-\sum (y_i - x_i)^2 = (y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2 = (2 - 1)^2 + (4 - 2)^2 + (6 - 3)^2 = (1)^2 + (2)^2 + (3)^2 = 1 + 4 + 9 = 14 .$$

$$\text{or : } \sum (y_i - x_i)^2 = \sum (y_i^2 - 2x_i y_i + x_i^2) = \sum y_i^2 - 2\sum x_i y_i + \sum x_i^2 = 56 - 2(28) + 14 = 56 - 56 + 14 = 14$$

$$(b) = \sum (x_i - 3)(y_i - 5) = (x_1 - 3)(y_1 - 5) + (x_2 - 3)(y_2 - 5) + (x_3 - 3)(y_3 - 5) = (1 - 3)(2 - 5) + (2 - 3)(4 - 5) + (3 - 3)(6 - 5) = (-2)(-3) + (-1)(-1) + (0)(1) = 6 + 1 + 0 = 7.$$

$$\text{or } \sum (x_i - 3)(y_i - 5) = \sum (x_i y_i - 5x_i - 3y_i + 15) = \sum x_i y_i - \sum 5x_i - \sum 3y_i + \sum (15) = 28 - 5(6) - 3(12) + (n)15 = 28 - 30 - 36 + (3)15 = 28 - 30 - 36 + 45 = 7.$$

$$(c) - \sum x_i y_i^2 = x_1 y_1^2 + x_2 y_2^2 + x_3 y_3^2 = (1)(2)^2 + (2)(4)^2 + (3)(6)^2 = (1)(4) + (2)(16) + (3)(36) = 4 + 32 + 108 = 144.$$

$$(d) - \sum (y_i - 3) = \sum y_i - \sum 3 = \sum y_i - (n)3 = \sum y_i - (3)3 = 12 - 9 = 3.$$

$$(e) - \sum y_i - 3 = 12 - 3 = 3.$$

$$(f) - \sum (x_i + 2)/y_i = (x_1 + 2)/y_1 + (x_2 + 2)/y_2 + (x_3 + 2)/y_3 = (1 + 2)/2 + (2 + 2)/4 + (3 + 2)/6 = 3/2 + 4/4 + 5/6 = 1.50 + 1 + 0.83 = 3.33.$$

$$(g) - \sum (x_i + 2)/\sum y_i = \sum x_i + \sum 2 / \sum y_i = \sum x_i + (n)(2) / \sum y_i = 6 + (3)(2)/12 = 12/12 = 1.$$

$$(h) - \sum y_i^2 - (\sum y_i)^2/n = 56 - (12)^2/3 = 56 - 144/3 = 56 - 48 = 8.$$

$$(i) - \sum x_i y_i - \sum (x_i)(y_i)/n = 28 - (6)(12)/3 = 28 - 72/3 = 28 - 24 = 4.$$

- The tabular presentation :

There are two main types of statistical tables, namely:

1. The simple tables, It is a table on which the data is distributed according to one recipe usually consists of two columns, first is represented the category or phenomenon divisions to categories or groups, and the second are shows the number of fixed vocabulary for each class or group .

Table 1. Distribution of the students number according to their weight .

categories weight (kg)	number of students
60 - 62	5
63 - 65	15
66 - 68	45
69 - 71	27
71 - 74	8
total	100

2. Composite table: It is a table which the data is distributed according to the two traits or phenomena or more at the same time usually consists of rows and columns.

Rows represent are classes or groups of one traits .

Columns represent are categories or groups of other traits .

As for squares that correspond to the rows and columns were contained several vocabulary or occurrences common into categories for both traits in table (2) .

Table 2: Distribution of the number of students according to the length and weight .

Length (cm)	Weight (kg)			Total
	51 - 60	61 - 70	71 - 80	
121 - 140	20	6	4	30
141 - 160	2	40	10	52
161 - 180	2	6	10	18
Total	24	52	24	100

- Frequency distribution table, It is a simple table that consists of two columns, the first which contain the variable values are divided into sections or groups called classes and the second shows the vocabulary of each category is called frequency.

Table 3: Frequency distribution table for the 80 lengths of cotton plant (cm).

length categories (cm)	redundancy (number of plants)
31- 40	1
41-50	2
51- 60	5
61-70	15
71- 80	25
81- 90	20
91- 100	12
TOTAL	80

Table 4: Recurring distribution table for lengths of cotton plant stating the real limits and categories centers .

Respectively	Classes	Real class boundaries	Class centers (y_i)	Recurring (f_i)
1	31 - 40	30.50 – 40.50	35.50	1
2	41 - 50	40.50 – 50.50	45.50	2
3	51 - 60	50.50 – 60.50	50.50	5
4	61 - 70	60.50 – 70.50	65.50	15
5	71 - 80	70.50 – 80.50	75.50	25
6	81 - 90	80.50 – 90.50	85.50	20
7	91 – 100	90.50 – 100.50	95.50	12
Total				80

For example, fourth category = (61-70). Minimum of fourth class = 61 The upper limit of fourth class = 70.