STATISTICS

During the study of natural and social events, we are not always able to follow them by using an experiment. Therefore, we are forced to make on them many observations, measurements and tests. Then, the characteristic data, for a given event, are collected and conclusions are being drawn based on them, which can be formulated as a theory or law. The examination in this way is called **statistical study** and the methods are called **mathematical (statistical) methods**.

The science that applies these methods to the study is called **statistics**.

The statistical tests to an event are referred to as a set of things or objects that are qualitatively equal, but they differ in their quantitative values. They are called statistical sets or **population**. Their elements are the **individuals** on whom observations can be performed.

For example, when census, the population is the set of all the residents and the individual is every resident.

When comparing different elements (individuals) under the observation of an event, we can perceive that they are different because of their own **features**. They are classified in separate subsets, **classes** or groups with equal characteristics.

Example: The following characteristics (features): height, weight, performance at school, sex, nationality, eye color, etc., can be examined on the set of all children in Skopje at the age of ten (population).

Example: The following characteristics (features): weight, percentage of sugar, alcohol per cent, etc., can be tested on the grape which was purchased by a wine cellar (population).

The insignia can be qualitative and quantitative. **The quantitative characteristics** are: height, weight, performance (of the students), percentage of sugar, weight (of grape), etc. **The qualitative characteristics** are: sex, nationality, color of the eyes, etc.

The values of the quantitative characteristics are real numbers.

When performing of statistical tests, most often, it is impossible to examine all elements of the population. Sometimes, it happens the tests to last longer so a part of the population is elected (a specimen). All examinations are performed on this specimen, and all the conclusions which will be drawn should be valid for the whole population with a certain probability.

The number of elements within the sample is called sample volume.

Formation of Statistical Tables

When preparing the data for examination, it is primarily, prior to not changing of the essence, some kind of an arrangement to be introduced between them and to give them a suitable form for rapid assessment. One of the stages in the preparation of the data for testing is forming of the tables that facilitate the statistical analysis.

Let be given a population of N individuals with a common feature X. Each feature has its own value.

The first step in resolving of the values lies in their sorting by size, i.e. representation in the following order:

$$\mathbf{x_1 \leq x_2 \leq \ldots\ldots \leq x_N}$$

The difference between the largest and the smallest value of X is called a distance variation h (shifting) of X.

Example 1: The following values were obtained by the measuring of the feature X: 15, 22, 55, 43, 36, 4, 46, 9, 55, 32, 22, 4, 9, and 22.

The subsequent order is obtained after collating their largest: 4, 4, 9, 9, 15, 22, 22, 22, 32, 36, 43, 46, 55, and 55. The difference between the largest and the smallest value is h= 55 – 4 = 51.

We can notice that some values of X are repeated.

This repetition is called **frequency** and it is denoted by **f**.

The sum of all frequencies is equal to N – the number of all individuals in the population.

The values x_1, x_2, … x_N together with the corresponding frequencies f_1, f_2, … f_n situated in growing order form a **statistical table**.

Values of X	x_1	x_2	. . . .	x_N	
Frequencies f	f_1	f_2	. . . .	f_n	Total N

The frequencies f_1, f_2, … f_n are allocated to the values x_1, x_2, … x_N of the feature X by using the statistical table.

Let's consider a case when **the feature X is interrupted** (when the feature receives final number of values).

Example 2: The number of members who are capable of work, in 20 farming families, was as follows: 2, 3, 4, 2, 3, 5, 2, 2, 1, 4, 5, 1, 3, 3, 4, 1, 1, 3, 2, and 2. Distribute the frequencies and create a statistical table.

Solution: We obtain the following order after trimming the values in size: 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 5, and 5. It can be noticed that the values are repeated, i.e. $f_1=4$, $f_2=6$, ……. Thus, the statistical table will be:

x	1	2	3	4	5	
f	4	6	5	3	2	20

Now, let's consider a case when **the feature X is not interrupted, continuous** (when the feature X accepts values from an interval [a, b]).

This interval is divided to n subintervals $[a_o, a_1]$, $[a_1, a_2]$, …… , $[a_{n-1}, b]$ which resulted in **grouping** of the values of the feature X according to the stated subintervals which are called group intervals or **classes**. The number of elements in the statistical set whose values of the feature X belong to the interval $[a_o, a_1]$ and it is denoted by f_1. The number of elements in the class $[a_1, a_2]$, it is denoted by f_2 and so on, f_n are the elements whose values of X belong to the interval $[a_{n-1}, b]$. The numbers f_1, f_2, … f_n represent the frequencies of the classes to the values of X.

The frequencies are allocated on the group intervals (classes) by using the statistical table. The group intervals are of the same length and their number should not be less than 5 and no more than 20.

Example 3: The annual consumption of fruit per kg was assessed in 20 families. The following values were obtained: 10,3; 17,2; 18,3; 4,3; 16,5; 19,8; 22,5; 12,4; 13,4; 11,5; 14,1; 4,1; 11,4; 7,6; 9,2; 28,0; 23,4; 8,7; 15,7; 15,8; 15,5 and 20,6. Group these values in group intervals with length 4 and determine the appropriate frequencies.

Solution: The 22 families represent a statistical set with N=22 elements to which the continuous feature X is measured: the annual consumption of fruit per kg. The following values were obtained after the arranging the values in size: 4,1; 4,3; 7,6; 8,2; 9,2; 10,3; 11,4; 11,5; 12,4; 13,4; 14,1; 15,5; 15,7; 15,8; 16,5; 17,2; 18,3; 19,8; 20,6; 22,5; 23,4; 28,0.

The upper values shall be grouped in 6 group intervals because h= 28 - 4,1= 23,9 and 23,9 : 6 ≈ 5,975 = 4.

The allocations of frequencies to characterize X values grouped by group intervals is presented in the following table:

The annual consumption in intervals	4,0 - 8,0	8,0 - 12,0	12,0 - 16,0	16,0 - 20,0	20,0 - 24,0	24,0 - 28,0
Number of families	3	5	6	4	3	1

Cumulative Frequency

The cumulative frequency with the "less than" landmark is the total number of elements (individuals) of the population that have less or the same characteristics of a given feature.

The cumulative frequency with the "greater than" landmark is the total number of elements (individuals) of the population that have greater or the same characteristics of a given feature.

The cumulative frequency is obtained when all frequencies whose features are less (greater) than or the same with the given characteristic are added up.

Example 4: The allocation of the frequencies to the feature X is presented in the following table.

x	4	5	8	10	14	17	20	
f	3	4	1	5	3	6	2	24

Determine the cumulative frequencies of the values of X, and then specify the number of elements (individuals) in the statistical set (population) for which the values of X are less than or equal to 14.

Solution: The cumulative frequencies are determined as follows:

x	4	5	8	10	**14**	17	20	
f	*3*	*4*	*1*	*5*	*3*	*6*	*2*	*24*
Cumulative frequency	*3*	*7*	*8*	*13*	***16***	*22*	*24*	

3 = 3   3 + 4 = 7 7 + 1 = 8 8+ 5 = 13 13+3 = 16 16+6 = 22 22+2=24

From this table, it can be noticed that there are 16 elements of the population whose values of X are smaller than or equal to 14.

Example 5: The allocation of frequencies to the feature X, whose values are grouped in group intervals (classes) is given in the following table:

Classes	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequencies	11	26	63	81	35	21	13

Determine the cumulative frequencies to the values of X, and then specify the number of elements (individuals) of the statistical set (population) whose values are smaller than or equal to 80.

Solution: The cumulative frequencies are determined in the same way as in the example 5:

Classes	30-40	40-50	50-60	60-70	70-***80***	80-90	90-100
Frequencies	*11*	*26*	*63*	*81*	*35*	*21*	*13*
Cumulative frequency	*11*	*37*	*100*	*181*	***216***	*237*	*250*

11=11 11+26=37 37+63=100 100+81=181 181+35=216 216+21=237 237+13=250

From this table, it can be seen that there are 216 elements in the population whose values of X are smaller than or equal to 80.

Tasks: (from population …)

1) 30 students did a written work in Mathematics. They obtained the following grades: 2, 1, 1, 3, 2, 2, 2, 4, 1, 3, 5, 5, 5, 4, 4, 2, 1, 4, 3, 1, 3, 2, 4, 2, 5, 4, 5, 5, 5, and 5.
Allocate the frequencies and create a statistical table.

2) Among 25 families, the monthly consumption of oil per liters is: 4, 2, 2, 2, 3, 4, 4, 5, 2, 2, 6, 6, 3, 2, 3, 4, 4, 4, 3, 2, 2, 3, 4, 5, and 5.
Allocate the frequencies and create a statistical table.

3) Five coins are tossed 1.000 times and the appearance of arms is observed during each throwing. It is determined that the coat of arms did not appear in 38 throws, one coat of arms appeared in 144 throws, two coats of arms in 342 throws, three coat of arms in 286 throws, four coats of arms in 164 throws and five in 25 throws.
Create a table with the allocation of frequencies and determine the number of throws in which the number of arms is less than or equal to 3.

4) The height of 40 students in centimeters is: 158, 184 170, 152, 164, 145, 169, 177, 166, 178, 160, 167, 156, 168, 172, 164, 188, 146, 158, 196, 183, 139, 174, 185, 166, 193, 162, 167, 155, 173, 160, 155, 181, 165, 155, 162, 170, 176, 165, and 148.
Perform grouping of these values in group intervals (classes) with a length of 8 cm. and allocate the corresponding frequencies.

5) The height of 40 students in centimeters is: 138, 164, 150, 132, 144, 125, 149, 157, 146, 158, 140, 147, 136, 148, 152, 144, 168, 126, 138, 176, 163, 119, 154, 165, 146, 173, 142, 147, 135, 153, 140, 135, 161, 145, 135, 142, 150, 156, 145, and 128.
Perform grouping of these values in 7 group intervals with a length of 9 cm. and determine the number of students whose height is less than or equal to 153 cm.

6) The length of illumination per hour is measured of 400 lights. It is determined that 14 lights shone from 300 to 399 hours, 46 lights shone from 400 to 499 hours, 58 lights shone from 500 to 599 hours, 76 lights shone from 600 to 699 hours, 68 lights shone from 700 to 799 hours, 62 lights shone from 800 to 899, 48 lights shone from 900 to 999, 22 lights shone from 1.000 to 1.099 and 6 lights shone from 1.100 to 1.199.
Create a table with the allocation of the frequencies and determine the number of lights that shone no more than 999 hours.

Arithmetic average of the feature X

- The arithmetic average of the feature X with ungrouped values $x_1, x_2, \dots, x_n$ is the number $\overline{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$

- The arithmetic average of the feature X with grouped values $x_1, x_2, \dots, x_n$ with the corresponding frequencies$f_1, f_2, \dots f_n$ is the number $\overline{X} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{n}$

- The arithmetic average of the continuous feature X with allocation of the frequencies $f_1, f_2, \dots f_n$, in group intervals is the number $\overline{X} = \frac{x_1' f_1 + x_2' f_2 + \dots + x_н' f_n}{n}$

  where x' is the middle of the interval (a, b) and it is calculated by the equation $x' = \frac{a+b}{2}$

Задачи (од аритметичка средина)

1) A student obtained the following grades: Macedonian – 4; Mathematics – 5; History – 5; Geography – 3; Biology – 5; Chemistry – 4 and Physics – 3.

Calculate the middle (average) success of the student.

2) The following success was achieved for the written work in Mathematics: 8 students received the grade 5, 7 students received the grade 4, 7 students received the grade 3 and 8 students received the grade 2.
Calculate the average success of the students that they achieved it in the written work in Mathematics.

3) The distribution of 120 students by the number of points that they achieved at the competition in Mathematics is given in the following table:

Points	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100
Number of students	2	5	10	21	43	30	9

Calculate the average of the achieved points per student.