

**MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE**  
**SIMON KUZNETS KHARKIV NATIONAL UNIVERSITY OF ECONOMICS**

# **STATISTICS**

**Guidelines**  
**to laboratory and independent work**  
**for Bachelor's (first) degree**  
**students of all specialities**

**Kharkiv**  
**S. Kuznets KhNUE**  
**2020**

UDC 311(07.034)

S81

**Compiled by:** I. Sierova

V. Derykhovska

V. Shlykova

O. Brovko

Затверджено на засіданні кафедри статистики і економічного прогнозування.

Протокол № 5 від 05.11.2019 р.

*Самостійне електронне текстове мережеве видання*

**Statistics** [Electronic resource] : guidelines to laboratory and independent work for Bachelor's (first) degree students of all specialities / compiled by I. Sierova, V. Derykhovska, V. Shlykova, O. Brovko. – Kharkiv : S. Kuznets KhNUE, 2020. – 97 p. (English)

The tasks for laboratory and independent work on the academic discipline and guidelines to them are given to help students gain theoretical and practical skills in the use of statistical observations, statistical analysis methods and forecasting the social and economic phenomena and processes.

For Bachelor's (first) degree students of all specialities.

**UDC 311(07.034)**

# Introduction

The right answer depends on  
the fact you are looking for.  
*The inferencial statistics gold rule*

The current trends in the world economy, the development of the "knowledge economy", where the main resource is the mobile and highly qualified "human capital", require the achievement of a new quality of the specialists' professional training.

In a market environment, education is directly connected with competitiveness.

In this sense, the role of innovative approaches, which focuses not only on the acquisition of a certain amount of knowledge and skills, but also on the formation of a set of competences, manifested in the ability to solve problems and carry out tasks in various spheres of economic activity, increases significantly.

The effectiveness of higher education modernization depends on the factors and conditions that are aimed at the identification of the additional opportunities of a competent specialist training.

The transition from the paradigm of learning to the paradigm of education assumes that the basic form of the educational process in the conditions of a rapid change of knowledge should be a combination of practical skills and independent work of a student.

In today's workplace, students can have an immediate competitive edge over both new graduates and experienced employees if they know how to apply statistical analysis skills to the real-world decision making.

The academic discipline "Statistics" is useful for professionals engaged in business analytics of economic activity and management at various levels. Modern business and administrative decision making require correct operations with primary data.

Statistics is useful in cases where you have to analyze processes being massively observed and demonstrating an obvious pattern.

The statistical regularity manifests itself more clearly in the case of the law of large numbers. This law reflects the laws inherent in random events of a mass nature.

With a large number of observations, the influence of random factors is mutually balanced and the main causes, which are reflected in a certain constant of the average number, come into effect.

In the theory of statistics, it is usual to distinguish between descriptive statistics and analytical (or inferential) one.

Descriptive statistics is mainly associated with the planning of the research, the collection of data and the presentation of the gained results in the form of statistical indicators. The most convenient form of presenting statistical information is tables and graphs.

The task of analytical statistics is to identify causal relationships, to assess the influence of the factors being studied and to draw the right conclusions.

A powerful tool for performing statistical research is the use of computers. In this regard, special packages of programs applied are widely used in the business sphere. They make it possible to ensure the speed of statistical calculations, high reliability and validity of the results, the possibility to easily present data in analytical, graphical or tabular forms.

The proposed tasks for laboratory work provide the opportunity to master the information technology of descriptive statistics based on the example of the popular office suite MS Excel.

The guidelines provide only a summary of the methods used.

Therefore, the proper use of the program requires an understanding of the relevant sections of statistics theory.

The tasks for laboratory work are performed using the Data Analysis module, as well as the built-in functions of MS Excel.

Using the Microsoft Excel program for statistical analysis, a student should be able to make the right choice of a method, but also to know the conditions of using it.

The **object** of the research is the capabilities of Excel.

The **subject** of the study is the use of MS Excel for statistical problem solving.

The **relevance** of the tasks is determined by the implementation of the capabilities of MS Excel for statistical problem solving.

The goal of the study is to generate sustainable knowledge about the capabilities of MS Excel for statistical analysis.

To achieve this goal, it was necessary to solve the following **problems**:

- to study MS Excel capabilities;
- to determine how to use the capabilities of MS Excel in statistics problem solving.

Each laboratory work is considered as an example of a specific problem solving and is provided with the basic comments and elements of graphical analysis.

Each laboratory work is accompanied by questions and tasks for students' independent work.

The evaluation of the laboratory work performance is carried out according to the results of the presentation of the work. Special attention is paid to the combination of theoretical and practical skills of students, the completeness of economic interpretation of the results obtained.

In the process of teaching the focus should be on the bachelor professional competences listed in Table 1.

Table 1.1

**The professional competences acquired by students  
when studying the academic discipline**

Competence name	Competence components
The ability to identify and combine the methods of statistical and economic analysis to assess the social and economic content of the problem	The ability to use common software packages to obtain the generalized characteristics of social and economic phenomena
	The ability to develop a statistical observation plan in accordance with a certain social and economic situation
	The ability to apply methods of data visualization based on the analytical material availability
	The ability to create a general system of statistical indicators to assess the social and economic problems at various levels
	The ability to detect regularity and to determine the forms of its functioning in the existing statistical populations
The ability to generate analytical reviews	The ability to apply scientific principles of information preprocessing according to a certain social and economic problem
	The ability to form representative research tools
	The ability to identify cause-and-effect relations between social and economic phenomena and to form an overall scheme of assessment of these relations

# **Content module 1. The methods of descriptive statistics**

## **Topic 1. The methodological principles of statistics**

### **Laboratory work 1. An overview of Excel**

*The purpose of the work* is to acquire the skills in operating the MS Excel spreadsheet.

*The task* is to learn the basics of using the spreadsheet: introduction and editing of data formatting, designing tables.

#### ***The guidelines***

Excel is a program that belongs to the category of spreadsheets and is a part of the Microsoft Office package. One of the main advantages of Excel is a wide range of use. There are several possible applications of Excel, among them being the following:

- solution of numeric problems that require time-consuming calculations, reporting, analysis of research results, and the use of different methods of financial analysis;

- creating charts. Excel contains tools to create different types of charts;

- organization of lists. Excel allows you to efficiently create and use structured tables with columns containing similar types of data;

- access to the data of other types and the ability to import data from many different sources;

- creating drawings and charts. Using Excel AutoShapes to create simple (and complex) schemes;

- automation of complex tasks by means of Excel macros to perform the same type of tasks with one click.

#### **The workbook and the worksheet**

All the actions performed in Excel, are saved in the workbook, which opens in a separate window. By default, the workbook files have the XLS extension.

Each workbook contains one or more worksheets, each consisting of individual cells. The cell can store a number, a formula or a text. To move from one worksheet to another, you need to click on the appropriate sheet tab, which is located at the bottom of the workbook. Furthermore, workbooks

can contain chart sheets, each of them inculcating one chart. One can go to the chart sheet by clicking on its label.

In Fig. 1.1, the most important elements of the Excel window are shown.

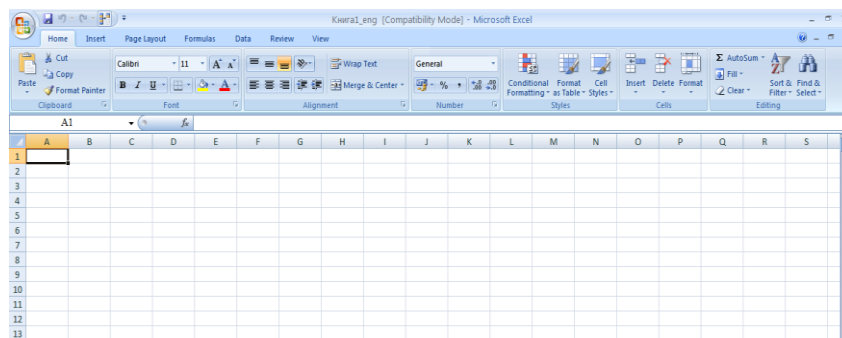


Fig. 1.1. The most important elements of the Excel window

### Moving within the worksheet

Each worksheet consists of rows (numbered from 1 to 1,048,576) and 16,384 columns (denoted by the letters A, B, ..., Z, AA, AB, ..., IT, IU, IV, etc.). At the intersection of a row and a column there is a single cell. At the moment, only one cell can be active. The active cell has dark contour (Fig. 1.2).

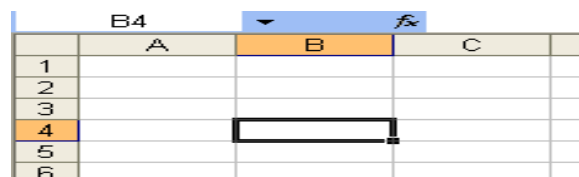


Fig. 1.2. The active cell

Its address, i.e. the column letter and the row number are specified in the *Name* field. Depending on the chosen method of moving within the workbook, the active cell can move or remain unchanged. The titles of the row and column at the intersection of which the active cell is, are marked by dark colors.

### Using the keyboard

To move within the worksheet, you can use the cursor keys. Pressing the "↓" key moves the table's cursor one row lower, pressing the "→" key – one column to the right, etc.

The <Num Lock> button controls the buttons on an additional digital keypad. When the *Num Lock* indicator is on, it is possible to use the keys on the numeric keypad to enter numbers in the Excel worksheet. This option is not selected by default.

## Using the mouse

Moving within the worksheet with the help of the mouse is also possible as it might have been expected. To change the active cell, one should click the desired cell, and it will become active. If the cell that you want to activate is not visible in the workbook, one can use the scrollbar to scroll the window in any direction.

When the scrollbars or scrolling through IntelliMouse is used, an active cell does not change, only the worksheet scrolls. To change the active cell, it is necessary to click on a new cell after scrolling.

## Using the menus and toolbars

If you have worked in any program before, then you should not have any problems when working in Excel. The user interface in Excel (that is, the menu and the toolbar) will look familiar, so far as the location and purpose of instructions in all the programs are the same.

## Using the menu

In Excel, like in all other Windows applications, there is a menu bar, located below the title bar (Fig. 1.3). Excel menu changes depending on the nature of your actions. For example, if you are working with a chart, the Excel menu adjusts so as to give you tools to work with charts.

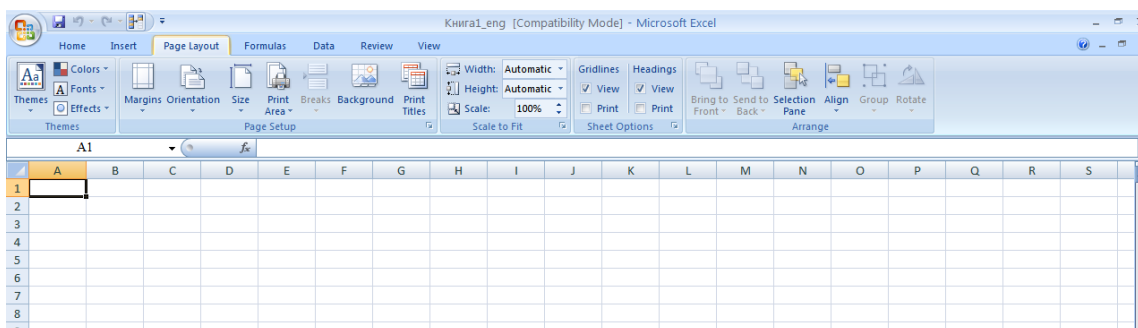


Fig. 1.3. The Excel menus and submenus

There is nothing easier than to get access to the menu with the help of the mouse. Click on the required menu, it will open and you will see a list of its elements (see Fig. 1.3). Some menu items have additional submenus. If you click on such menu item, the submenu will appear on the right side of it. Menu items that have a submenu are marked with a small triangle on the right side. For example, the *View* Toolbars submenu is shown in Fig. 1.3. Sometimes a menu item is highlighted with gray color. This is the way to indicate that this item is not designed for the operation you are performing at the moment.



Selecting a menu item, the name of which includes the ellipsis (three dots), always results in the dialog box opening. Menu commands that have no ellipses are executed immediately. For example, when you select *Insert > Cells...*, a dialog box appears because the program needs more information about the executed command. The *Insert > String* command does not need any additional information, so it is executed as soon as it has been chosen. Almost any command in Excel can be disabled with the *Corrections > Undo* command. Thus, you can undo the last 16 executed commands. Instead of choosing the *Corrections > Undo* command, you can use the *Undo* button, which is located on the standard toolbar. In addition, to undo the last action you can use the key combination <Ctrl + Z>.

### Using the context menus

In addition to the main menu, Excel provides many other context menus. These menus are context-sensitive, that is, their content depends on the nature of the operation that you are performing at the moment. There are only the commands that are most commonly used to work with the selected element at this point. The context menu can be called for almost any object in Excel. To do this, click on the right mouse button. In Fig. 1.4, the context menu that appears when you right-click on the cell is shown.

### The use of the key combinations

For some menu items the shortcuts are provided. Usually a key combination is shown next to the name of the menu item. For example, the keyboard equivalent of the *Corrections > Find* command is the combination <Ctrl + F>. The combinations of frequently used commands can be seen in the menu to the right of their names.

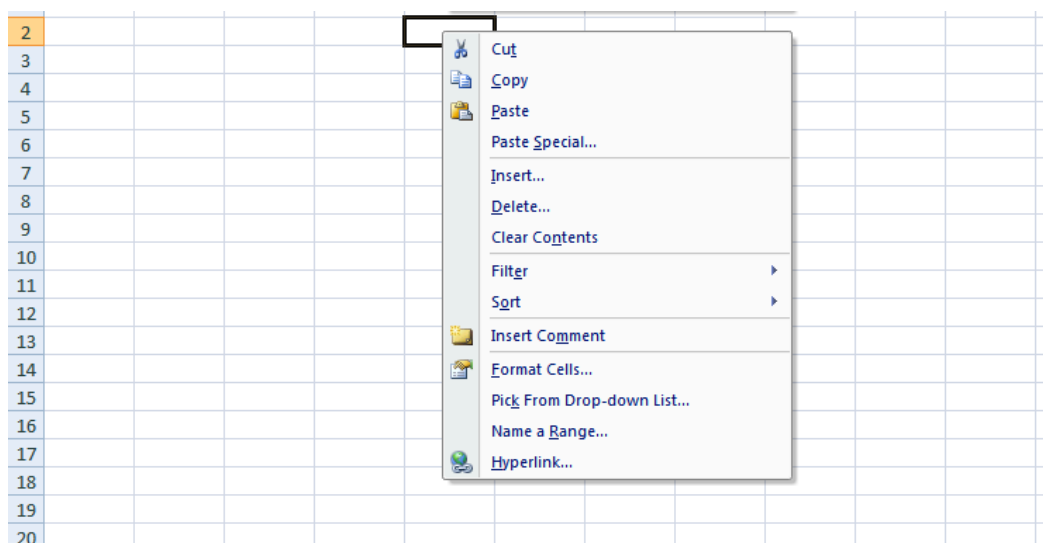


Fig. 1.4. The context menu

## The Excel toolbars

Excel, like almost all modern applications, includes a handy graphic toolbar. Most of the toolbar buttons replace some of the menu commands. For example, the *Copy* button replaces the *Corrections > Copy* command. However, some buttons do not have equivalents in the menu, such as the *AutoSum* button which automatically inserts the formula for calculating the sum of the values in a specified range of cells.

To find out what a particular toolbar button is meant for, move the mouse pointer to it (but do not click on it). Next to the cursor, a small yellow window appears with a tip that contains the name of the button. If a tip is not seen on the screen, select the *Tools > Setting up* command. In the window that appears, navigate to the tab *Options* and select the *Display button tool tips*. Table 1.2 lists the most commonly used built-in Excel toolbars.

### Hiding and displaying the toolbars

By default, Excel displays two toolbars – the standard and formatting ones. In addition, you can create your own toolbars, which will consist of the buttons that you find most useful.

Table 1.2

## The Excel toolbars

Name	Characteristics
Standard	Running the most frequently used commands
Formatting	Changing the appearance of a worksheet or chart
WordArt	Inserting or editing an artistic text
Web site	Access to the Internet with Excel
Border	Adding border (frame) around the selected cells
Charts	Working with Charts
Dependences	Identifying errors in the worksheet and displaying the relationships between formulas
Protection	Management of protected worksheets and books
Peer review	Operations with the notes to the cells
Drawing	Inserting and editing the pictures on a worksheet
Pivot tables	Working with pivot tables
Forms	Adding to the worksheet controls (buttons, lists, etc.)

To hide a toolbar or display it on the screen, choose *View > Toolbars* or right-click any toolbar or the menu bar. The result is a list of toolbars. The toolbar activated in the list is visible. To hide the toolbar, click on its name in the list. And vice versa, to show a hidden toolbar, you can also click on its name.

To access all the toolbars, choose *Service > Settings*. The *Toolbar* tab of the dialog box appears on the screen *Settings* for a list of available toolbars. Check the boxes next to each panel to display them.

### Constructing the first worksheet

At the first stage you need to run Excel and deploy a work window so that it fills the whole screen. Then, expand a clean workbook, which will be called *Book1*. If any book is open, in order to open a new one, click on the *New* button of the standard toolbar. First you need to enter the headers of the row and column in a worksheet *Sheet1*.

Perform the following steps.

1. Move the table cursor to the cell *A3* with the arrow keys. In the *Name* field the address of the cell will appear.

2. In the cell *A3*, enter the *North*. To do this, type the text and press the <Enter> key. Depending on your settings, Excel will either move the cursor to the cell *A4*, or leave it in the cell *A3*.

3. Move the table cursor to the cell *A4*, enter the *South* and press the <Enter> key.

4. In the cell *A5*, type *Total* and press the <Enter> key.

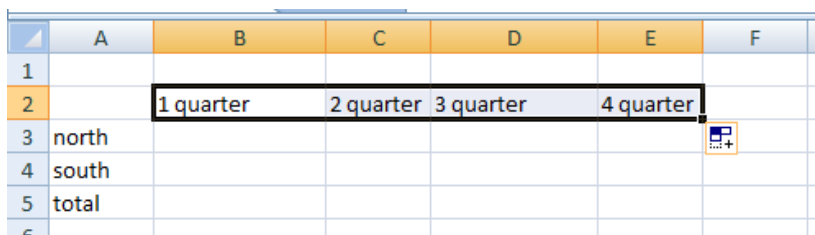
5. Move the table cursor to the cell *B2*, enter *1st quarter* and press the <Enter> key.

Three other titles could be entered manually, but let's make Excel work.

6. Find a small box in the lower right corner of the table cursor. It is called the fill handle. If you place the mouse pointer on the marker, it will take the shape of a black cross.

7. Place the mouse pointer over the fill handle so that the pointer turns in to a cross. Then click and drag the pointer to the right until you have the identified three cells on the right side (*C2*, *D2* and *E2*). Release the mouse button and you will see that the program automatically enters, instead of you, three headers. This is an example of using the *AutoComplete* function.

As a result, you should get a table such as shown in Fig. 1.5.



	A	B	C	D	E	F
1						
2		1 quarter	2 quarter	3 quarter	4 quarter	
3	north					
4	south					
5	total					
6						

Fig. 1.5. The worksheet after entering the headers

## Data entering

At this stage, you need to enter the sales volumes per quarter for each region.

1. Move the table cursor to the cell *B3*, type the number 123 and press the <Enter> key.
2. Move the table cursor to the other table cells and enter the following data.

The worksheet will look as shown in Fig. 1.6.

	A	B	C	D	E	F
1						
2		1 quarter	2 quarter	3 quarter	4 quarter	
3	north	123	411	456	357	
4	south	101	258	951	105	
5	total sum					
6						

Fig. 1.6. The worksheet with the entered data on the sales volumes

## Creating a formula

Let's, now, create a formula for calculating the total sum based on the regions. To do this, follow the instructions below.

1. Move the table cursor to the cell **B5**.
2. In the standard toolbar located below the menu bar, find the *AutoSum* button and click on it. This button depicts the Greek letter "sigma". In the cell *B5* the program has placed the following: = *SUM(B3:B4)*. This is the formula that calculates the sum of the values of the cells that are in the range *B3:B4*.
3. Press the <Enter> key to enter the formula. You will see that in the cell *B5* the sum of two numbers appears. You could repeat the same operation for the three quarters remaining, but it is much easier to copy the formula into the three cells, located on the right side.
4. Move the table cursor to the cell *B5*.
5. Place the mouse pointer over the fill handle. When it turns into a cross, click and drag the pointer three cells to the right. When you release the mouse button, you will see that Excel copied the formula into the selected cell. When the data changes, these formulas are recalculated automatically and the new results are displayed.

## Formatting the table

The table you have created can look better. At this stage one can take an opportunity of using the Excel automatic format. Perform the following steps.

1. Place the table cursor into any cell.
2. Select the command *Format > AutoFormat*. The result will be the following: Excel will determine the boundaries of the table and highlight all of it, and will open the *Auto Format* dialog box.
3. The *Auto Format* dialog box offers 16 ready formats to choose from. Click on the format you like.
4. Click on the button *OK*.

Your table should look like that in Fig. 1.7.

## Adding a title heading

At this stage, we introduce the table header, select it in bold and place in the center relative to the table columns. For this purpose do the following.

1. Place the table cursor into the cell **A1**.
2. Enter the *volume of sales based on the regions* and press the <Enter> key.
3. Move the cursor back to the table cell *A1*, if it was in a different cell, and click on the button *Bold* located on the formatting toolbar. This button depicts the capital large letter *B*.
4. Click on the button drop-down *Size*, which is located on the formatting toolbar. Choose from the list number 14.
5. Click on the cell *A1* and drag the mouse pointer to the right to select the cells *A1, B1, C1, D1* and *E1* (i.e. the range *A1:E1*). Just don't drag the fill handle. You need to select the cell, and not to copy the cell *A1*.
6. Click on the button *Merge&Center*, located on the formatting toolbar. As a result, the text that is in cell *A1*, will be centered relative to the selected cells.

Now your worksheet should look as that shown in Fig. 1.7.

	A	B	C	D	E	F
1		<b>Volume of sales by region</b>				
2		1 quarter	2 quarter	3 quarter	4 quarter	
3	north	123	411	456	357	
4	south	101	258	951	105	
5	total sum	224	669	1407	462	
6						

Fig. 1.7. The formatted table with the header

## **Saving the workbook**

All you have done so far has been stored in your computer RAM. And if suddenly power is lost, all will be lost. So it's time to save the results of the work done in the file. Name it, for example, *My first workbook*.

1. It depicts a diskette. Excel will open the dialog box *Saving document*.
2. Click on the *Save* button on the standard toolbar.
3. In the *File Name* enter *My first workbook*, and click on the *Save* button or press <Enter>.

The program will save the workbook in a file. Thus the workbook is left open, and you can continue using it.

## **The tasks for independent work**

1. Go to the official site of the State Statistics Committee of Ukraine ([www.ukrstat.gov.ua](http://www.ukrstat.gov.ua)) and read today's top analysis. Do the following:
  - give an example of a numerical variable found in the poll;
  - give an example of a categorical variable found in the poll.
2. Which types (quantitative or qualitative, discrete or continuous) include the following features:
  - a) the number of employees,
  - b) the relationships of family members,
  - c) the gender and age of a person,
  - d) the social position of depositors of commercial banks,
  - e) the number of floors of the houses,
  - f) the number of children in a family,
  - g) the retail turnover of shops?
3. Using the official site of the State Statistics Committee of Ukraine, provide a list of indicators which can characterize: a) population; b) a consumer market; c) an industry; d) a firm.
4. What are varying and not varying features inherent in a person, a firm? Present a comparative characteristic in a table form with MS Excel.

## **The list of questions for independent work**

1. Supply examples of social life sectors, which statistics studies.
2. Formulate a definition of statistics as a science and give it an appropriate justification.

3. Describe the main features of a statistics object.
4. Determine populations which can be identified at a university for the statistical study.
5. What quantitative and attributive (qualitative) characteristics can describe a population of high school students, a population of banks?
6. Name the main categories of statistics and explain their meaning.
7. Explain the distribution statistics for specific industries and why the study of statistics begins with the general theory of statistics.
8. Name the specific techniques (methods) of statistical research.
9. Describe the basic tasks of statistics in modern terms.
10. What are the main features which determine the variation of a student's achievement?

## **Topic 2. Statistical observation**

### **Laboratory work 2. Submission of statistical data: graphs**

*The purpose of the work* is to acquire the skills in the analysis of statistical data using the graphical method in the package MS Excel.

*The task* is to construct the graphs on the basis of statistical data, see the types of the graphs in MS Excel, get acquainted with the *Chart Wizard* possibilities.

#### ***The guidelines***

##### **Building a chart by pressing a keyboard single key**

The following sequence of actions gives the opportunity to create a chart on a separate sheet.

1. Enter the input data which will be used to create a chart (Fig. 2.1).

	A	B	C
1		<b>Enterprise 1</b>	<b>Enterprise 2</b>
2	January	1 843	982
3	February	2 283	1 092
4	March	2 183	1 143

**Fig. 2.1. The input data**

2. Select the range of data, including the row and column headers. For example, if the chart is based on the data presented in Fig. 2.1, you need to highlight the range *A1:C4*.

3. Press the button *<F11>*. Excel will build a chart basing on a selected range and put it on a new chart sheet with the name of chart 1. In Fig. 2.2 the result of this procedure is shown.

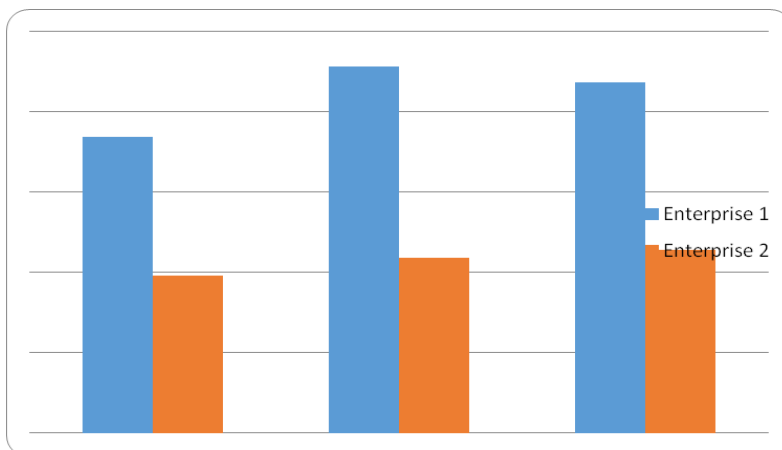


Fig. 2.2. The chart built by one click of the keyboard button

### Building a chart by pressing a single button mouse

The approach implies the following actions.

1. Make sure that the *Charts* toolbar is on the screen.
2. Select the data basing on which chart 1 will be created.
3. Press the *Chart type* button, which is located on the *Charts* toolbar, and select the type of chart (Fig. 2.3).

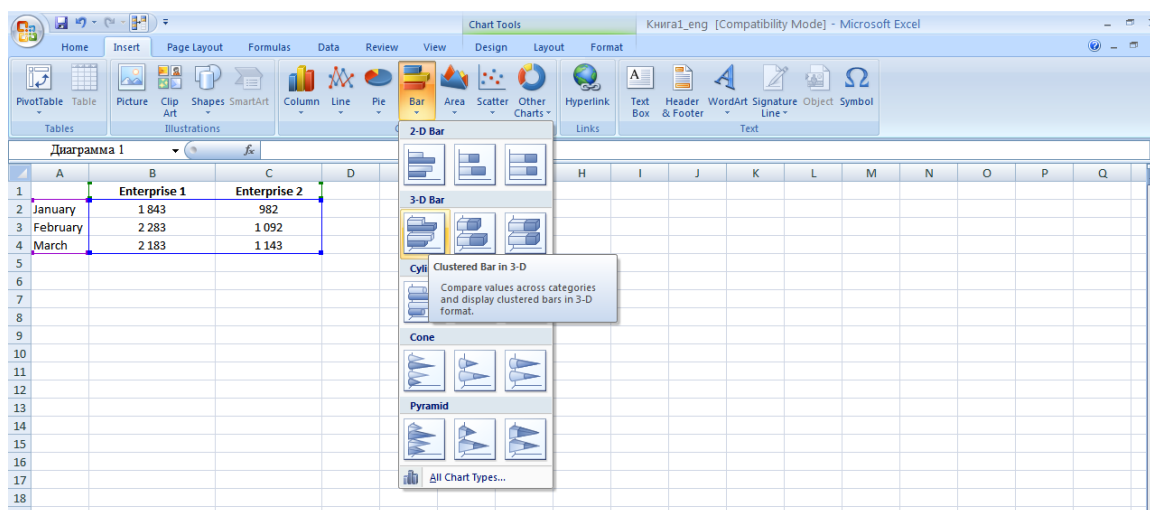


Fig. 2.3. Building a chart by clicking a single mouse button

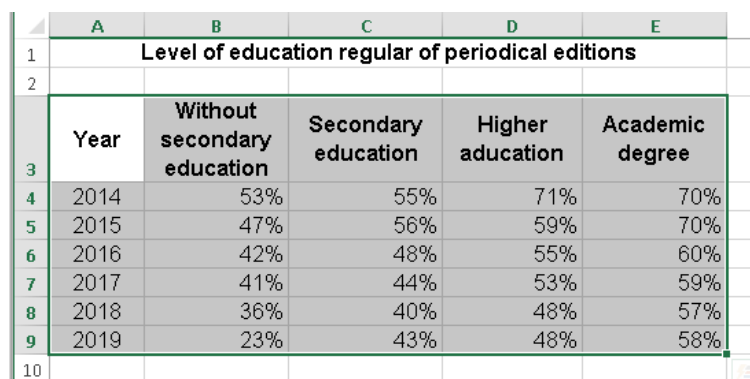


## Using the *Chart Wizard* function

In practice, it often happens that it is necessary to specify certain chart settings. In this case, you need to make changes to the settings manually after creating a chart. But for more control it is better to use the *Chart Wizard*, which provides for the following sequence of actions.

After entering the input data, highlight indicators for the construction of the chart to simplify the work (Fig. 2.4). Otherwise, you can choose them in the second *Wizard* dialog box.

In this example (Fig. 2.4), the range *A3:E9* was selected. This range contains the column headers, but it doesn't include the name of the chart, which is stored in the cell *A1*.



	A	B	C	D	E
1	Level of education regular of periodical editions				
2					
3	Year	Without secondary education	Secondary education	Higher education	Academic degree
4	2014	53%	55%	71%	70%
5	2015	47%	56%	59%	70%
6	2016	42%	48%	55%	60%
7	2017	41%	44%	53%	59%
8	2018	36%	40%	48%	57%
9	2019	23%	43%	48%	58%
10					

Fig. 2.4. The input data

Click on the *Chart Wizard* button, which is located on the toolbar, or choose *Insert* → *Chart*. You will see the first of the four *Chart Wizard* dialog boxes on the screen.

You can always go back to the previous stage of the work, pressing the *Back* button, and the next one – the *Next* button. To immediately complete the construction of the chart, click *Finish*.

**Chart Wizard: Step 1 of 4.** Fig. 2.5 presents the first *Wizard* window, where the type of chart is determined. The dialog box contains two tabs: *Standard* and *Custom*. The *Standard* tab includes 14 basic chart types with relevant subtypes, and the *Custom* tab contains some special types of charts, including the manufacturer's.

For this example, it is expedient to choose the chart type *Chart with markers*. From the list *Type* select the element *Graph*, then click the icon graph of a suitable type (subtype) in the list *View*.

Then determine the *Type* and *View* (subtype) charts, click on the *Next* button to proceed to the next step.

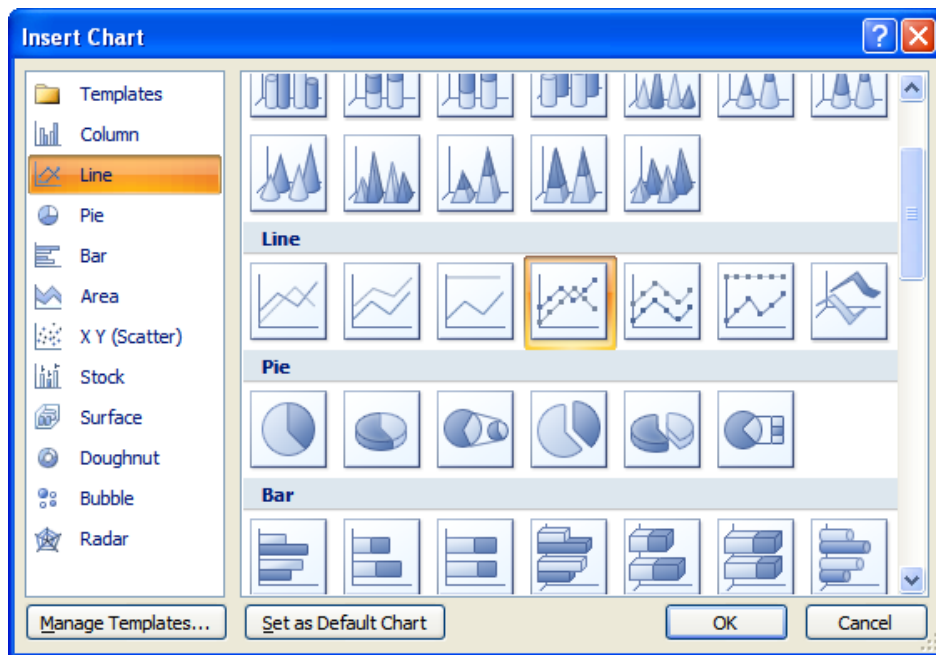


Fig. 2.5. The first *Chart Wizard* window: the **Chart Type**

**Chart Wizard: Step 2 of 4.** In the second step of the *Chart Wizard* function (Fig. 2.6) you can define the following parameters:

check out the range of the input data, if you want to change it;

determine the orientation of the data (rows or columns);

verify whether Excel divided the categories' source data and ranks correctly.

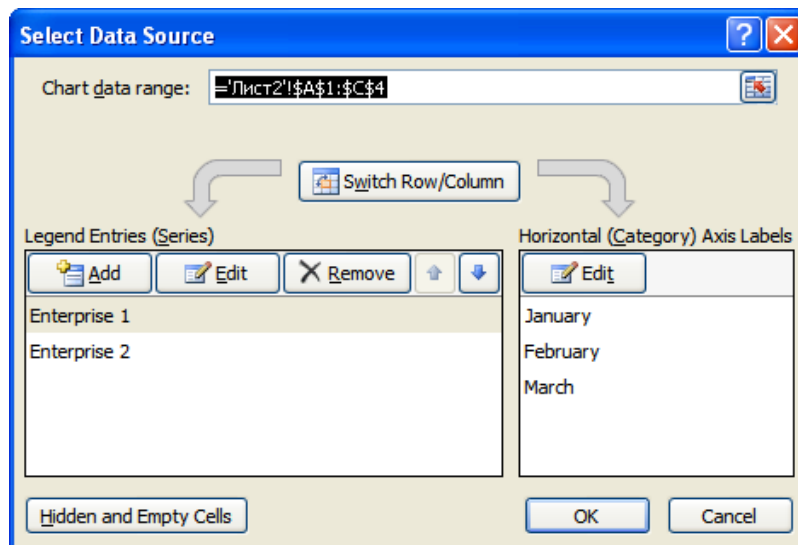


Fig. 2.6. The second *Chart Wizard* window: the **Data range** tab

The range field contains a reference to a range that has been allocated before starting the *Chart Wizard* tool. If you have selected only one cell, Excel will show a link to the foreseen range of the input data.

If the data range is not defined properly, it can be changed. To do this, press the *Range* field and highlight the data necessary in the worksheet.

The orientation of data cardinally affects the appearance of a chart. Excel itself performs this procedure. If the automatic execution of this Excel option is not implemented correctly, the orientation can be set manually by selecting the desired position for the *Series* switch.

If Excel mistakenly identified the years in column A as series of numeric value, but in fact, the number of the year is the name of the category and should be treated as text and does not appear in the chart as a numeric value, this must be set manually. To do this, go to the *Series* tab (Fig. 2.7).

The *Series* list contains the names of all the series data of the chart. In Fig. 2.7, the name of the *Year* series is in the cell A3, and the value – in the range of A4:A9.

As the year series is not really a series, and contains the name of the X-axis, it must be removed from this list. To do this, highlight the necessary series and press the *Delete* button. As a result, the appearance the chart will change and will not have a signature on the category axis.

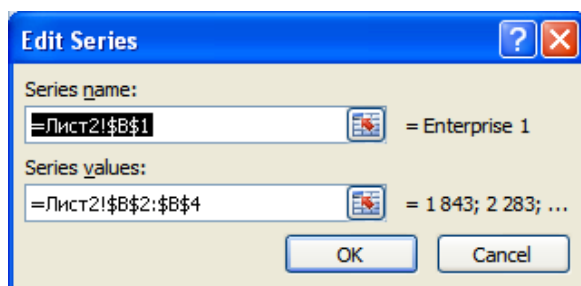


Fig. 2.7. The third *Chart Wizard* window: the *Series* tab

The *Signatures of the X-axis* field is empty, so you must specify a range that contains these signatures. To add a signature, click on the field and select the A4:A9 range. After that the years are displayed as categories on the horizontal axis, and the chart itself consists of four rows of data (Fig. 2.8).

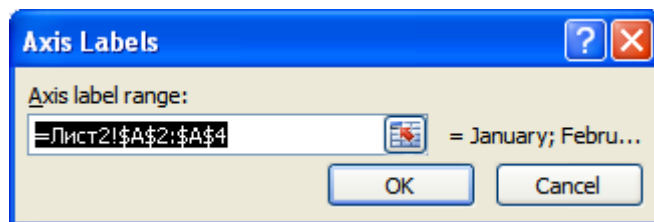


Fig. 2.8. The *Series* tab after the removal of one row and adding the X-axis labels

**Chart Wizard: Step 3 of 4.** In the fourth dialog box the *Chart Wizard* sets most of the options that define the appearance of the chart (Fig. 2.9). The options depend on the type of chart. The dialog box contains six tabs:

*Headlines* – adding chart titles;

*Axis* – displaying and deleting the axis, as well as delimiting their types;

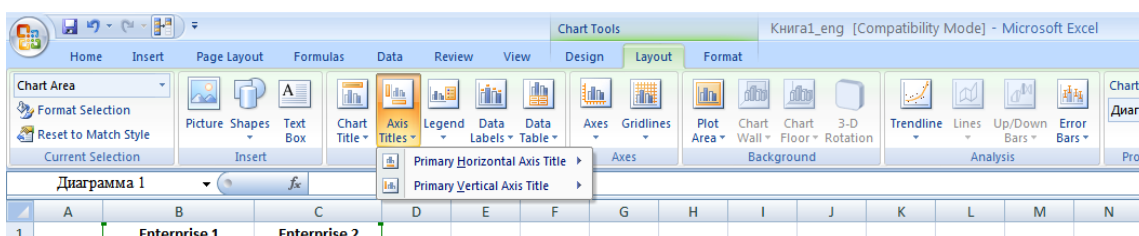
*Grid lines* – defining the grid settings if they are in the chart;

*Legend* – determining the presence and location of a legend;

*Data labels* – determining the presence and place of the data labels;

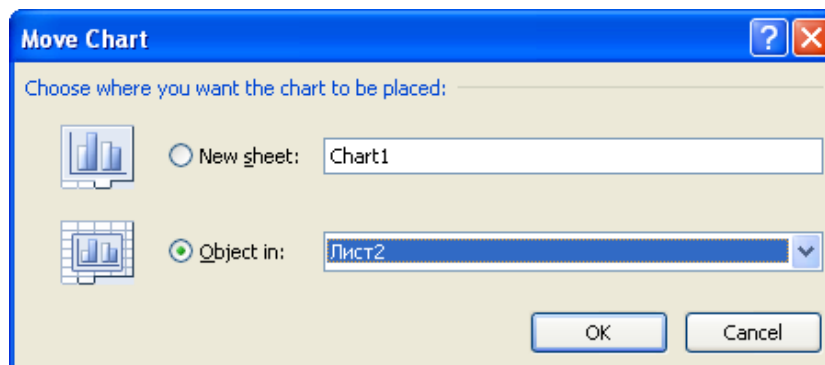
*Table of data* – determining the presence of the data table.

Fig. 2.9 displays the change of only two parameters in the *Titles* tab. In the *Chart* the level of education of regular readers of the periodicals, and on the X-axis (of categories) the years were entered.



**Fig. 2.9. The fourth Chart Wizard window:  
the options that determine the appearance of the chart**

**Chart Wizard: Step 4 of 4.** In the last dialog box, you should indicate where to place the chart. Click on *New sheet* and enter a name if you want to place the chart on a separate sheet, or click *Object in* if you want to place the chart on the active sheet (Fig. 2.10).



**Fig. 2.10. The fifth Chart Wizard window: the location of the chart specified**

If you created a nested chart, Excel will place it in the center of the active window (Fig. 2.11).

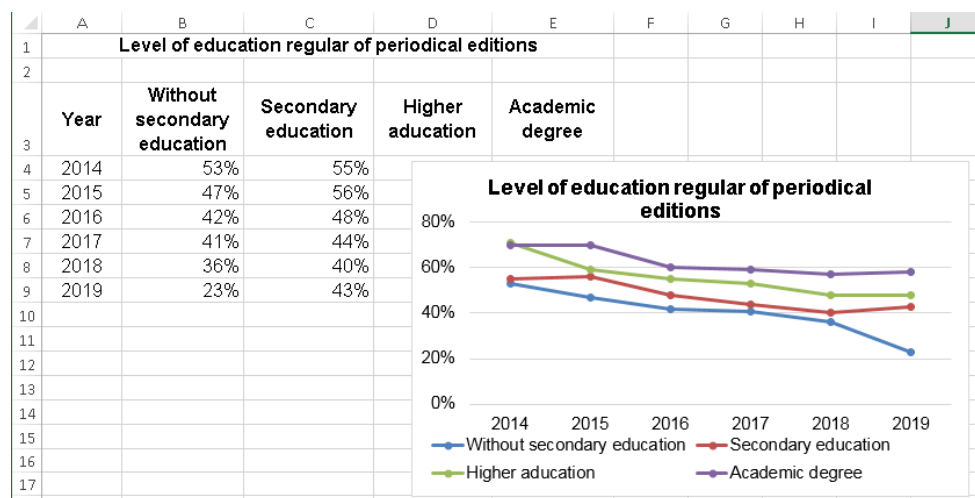


Fig. 2.11. The chart constructed by means of the *Chart Wizard* tool based on the specified parameters

### Changing the settings of the chart (elements)

The constructed chart can be modified any time. But before starting to modify the chart (any item), it is necessary to activate it by clicking on a chart (or element).

The settings of charts that are often used are as follows:

- moving and resizing a chart (nested), i.e. in order to move the chart one must click on its border, and then drag. To change the size of the chart you need to drag one of the eight size markers in the border of the chart;
- changing the type of a chart. The first method involves pressing the *Chart Type* button, which is located on the tool bar *Charts* (Fig. 2.3) and choosing one of the 18 main types of charts. The second way is to right-click on the chart and choose from the shortcut menu *Chart Type*;
- moving and deleting chart elements. Some elements of the chart can be moved (for example, titles, legend, or data table). To move a chart element, select it by clicking with the mouse, then drag to the right place of the chart. To delete a chart element, you need to select it and then press <Delete>;
- formatting the chart and its elements. Activation of a chart (chart element) allows you to change the parameters. The first method involves the use of commands from the context menu. The second one assumes the use of the *Charts* toolbar commands (Fig. 2.3), and the third method is the use of a set of commands that are on the other toolbars (such as *Fill color* data series, *Bold* chart legends, etc.).

### **The tasks for independent work**

1. Define an observation unit and a unit of population for carrying out a census of the industrial equipment, the inspection of workers' or employees' budgets; the account of availability of cash registers in commercial shops of a city.

2. Make a list of questions for an observation program:

a) interrogation of listeners of a business school concerning the orientation of the educational process to practical activities;

b) selective observation of the marketing campaign participants;

c) accounting for currency transactions of commercial banks.

3. Define the type of statistical observation concerning the coverage of the population units which are considered and the time of data recording:

a) recording the number of the registered crimes;

b) data recording of the tenders at universal stock exchanges;

c) registration of the price level of the agricultural products which are sold in the markets of large cities;

d) inspections of the budgets of unprotected layers of the population;

e) recording the number of newborns;

f) interrogation of young families of a region concerning the family planning.

4. Define the organizational form and the way of:

a) property inventory in public catering establishments;

b) the list of all the registered economic structures with the indication of their requisites (name, address, phone number, fax), the kind and the field of activity;

c) inspections of responses of the clients of service stations;

d) the monthly account of the manufacture volume and production realization of the enterprises;

e) accounting the balances of enterprises of separate branches of a national economy;

f) marketing research of the saturation of the audio, video equipment market.

5. By means of the logic control, find mistakes and ways to correct them in the following sheet. Specify, which of the answers are not consistent with each other.

Name and surname: Kirichenko Ivan.

Relation to the head of the family: the head.

Sex: female.

Age: 3.

Marital status: divorced.

Nationality: Ukrainian.

Native language: Russian.

Education: special.

Source of the means of existence: work in an institution.

### **The list of questions for independent work**

1. What is the nature of statistical observation? What is the difference between statistical observations and other kinds of observations?
2. The object and census of statistical observation. Their interconnection.
3. What is statistical observation, what is its essence and main tasks?
4. What are the main programmatic, methodological and organizational issues of statistical observation?
5. What is the object and unit of observation?
6. What is the critical moment and time of observation?
7. Define the surveillance program.
8. What are the requirements that are met when designing a statistical observation program?
9. What are statistical observation errors and what are their types?
10. How is the monitoring of results controlled? What are some ways to troubleshoot statistical observations?

## **Topic 3. Summarization and grouping of statistical data**

### **Laboratory work 3. Summarization and grouping of statistical data**

*The purpose of the work* is to acquire the skills in grouping the data in MS Excel.

*The task* is to group statistical data using the "Data Analysis" add-in of MS Excel.

#### ***The guidelines***

Grouping is partitioning (dividing) data sets into groups according to a certain attribute (feature, characteristics).

Let us consider the method of grouping based on the example:

We have some data on the foreign trade of Ukraine (export) with EU countries in 2019 (Table 3.1)

Table 3.1

### The input data

Country	Export, thousand USD	Country	Export, thousand USD
Austria	361 323.9	Malta	7 746.0
Belgium	251 541.9	The Netherlands	995 322.6
Bulgaria	418 193.3	Germany	1 423 735.2
The United Kingdom	317 792.1	Poland	2 200 010.1
Greece	159 123.2	Portugal	228 126.0
Denmark	155 453.9	Romania	716 981.4
Estonia	98 048.2	Slovakia	471 362.6
Ireland	45 483.9	Slovenia	16 758.2
Spain	1 004 547.4	Hungary	1 053 084.2
Italy	1 929 575.6	Finland	62 355.7
Cyprus	53 481.4	France	453 674.3
Latvia	138 155.3	Croatia	39 065.9
Lithuania	258 222.5	The Czech Republic	560 756.1
Luxemburg	5 087.0	Sweden	71 275.4

Grouping of statistical data should be performed in the *Data Analysis* add-in (Fig. 3.1) and in the *Histogram* mode (Fig. 3.2), in the package Excel.

The mode *Histogram* is used to calculate the frequency of the contact data in the specified boundaries of intervals, and to build the histogram of the interval variation distribution range.

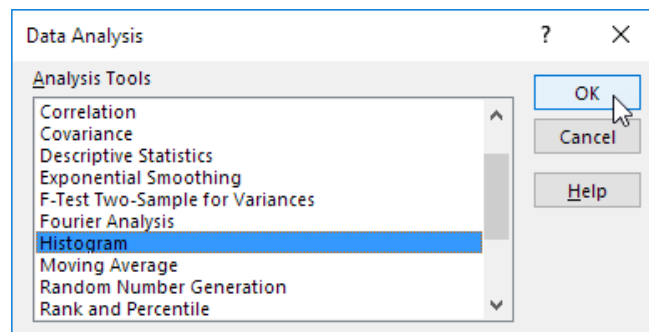


Fig. 3.1. The *Data Analysis* dialog box



The dialog box of this mode (Fig. 3.2) assumes the following parameters:

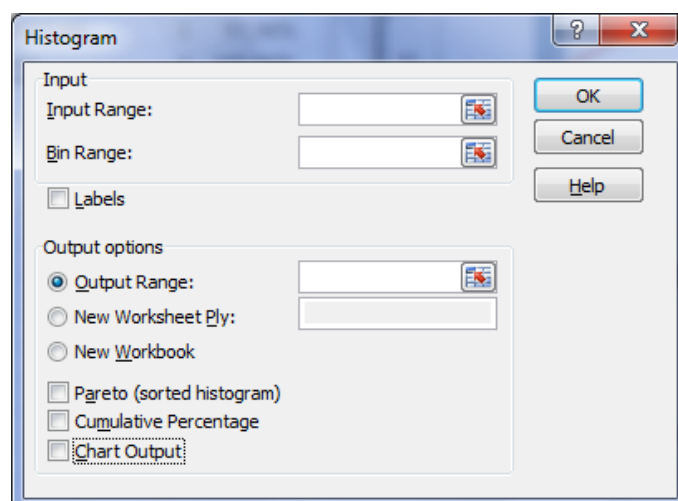


Fig. 3.2. The *Histogram* window options

1. The *Input Range*, i.e. the location of the input data on your worksheet (only the numeric data).

2. The *Bin Range* (an optional parameter), i.e. entering a link to the cells containing a set of limits that define the intervals (pocket). These values must be entered in ascending order. In MS Excel the number pointing the intervals by the data is calculated. The limits of intervals have a strict lower limit and a non-strict upper limit.

If the *Bin Range* was not entered, then a set of intervals, evenly distributed between the minimum and maximum values of the data will be generated automatically.

3. The *Labels*, which are set in an active state, if the first row (column) in the input range contains titles. If no titles, the checkbox should be deactivated. In this case, standard names for the output range will be automatically created.

4. The *Output Range*, *New Worksheet*, *New Workbook* activates the field where the link to the upper-left cell of the output range should be entered. The *Output Range* is automatically detected and a message appears in case of the possible imposition of the original range to the output range.

5. The *Pareto* (sorted histogram) is set in an active state to represent the data in descending order of the frequency. If it is unchecked, the data in the output range will be listed in order of intervals.

6. The *Cumulative Percentage* is set in an active state to calculate the percentage of accumulated frequencies and inclusion of the graph of accumulated frequencies in the histogram. Table 3.2 explains the calculation of the accumulated frequencies. Cumulative frequency is calculated based on the frequency (see Table 3.2, the Frequencies column). Each value of the cumulative frequency is divided by the maximum accumulated value. The result is relative frequency expressed in fractions of a unit. After the transformation to the percentage format the final result is obtained.

7. The *Chart Output* is set in the active state for automatic building of an embedded chart on a worksheet containing the input range.

You need to construct a histogram and accumulated frequencies in the histogram, also visualize the results of grouping. In order to do the task, the *Histogram* mode could be used. In the dialog box that opens (Fig. 3.2), set the following parameters:

in the *Input Range*, enter **\$B\$8:\$B\$35**;

in the *Output Range*, enter **\$G\$4\$**;

also select *Cumulative Percentage* and *Chart Output*.

The frequencies and cumulative frequencies are presented in Table 3.2.

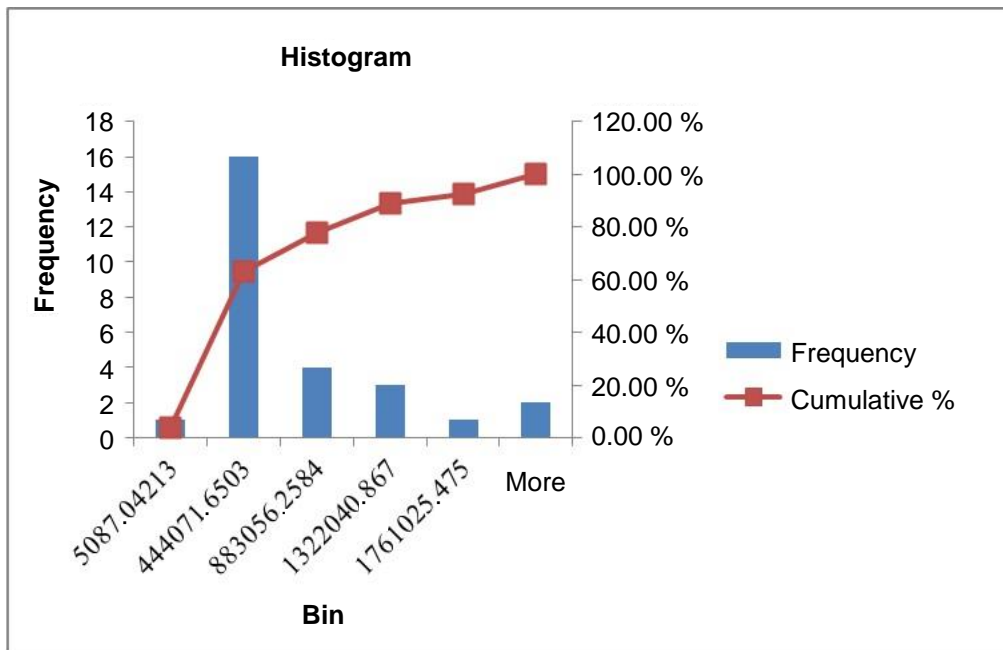
Table 3.2

### The frequencies and cumulative frequencies

Bin	Frequencies	Cumulative %
5 087.04213	1	3.70 %
444 071.6503	16	62.96 %
883 056.2584	4	77.78 %
1 322 040.867	3	88.89 %
1 761 025.475	1	92.59 %
More	2	100.00 %

So, we can see uneven distribution of countries based on the indicator "export" to 16 countries: Ukraine exported from 5 087.04 to 444 071.65 thousand dollar worth of products.

The histogram of the constructed and accumulated frequencies is shown in Fig. 3.3.



**Fig. 3.3. The histogram of distribution**

In conclusion, we can say that we got unequal (uneven) distribution. The countries of the European Union, to which Ukraine exported its products, can be divided into 6 groups. At the same time, the maximum volume of exports falls to Poland and Italy, and the minimum to Luxembourg. On the whole, exports are at a low level, which means that the Ukrainian government should reform the foreign economic policy and strengthen promotion (marketing), improve the quality and competitiveness of inventory products in the markets of Europe.

**The tasks for independent work**

1. We have the data about the distribution of 20 workers (Table 3.3) according to their skill level.

Table 3.3

**The input data**

5	6	4	2	3
3	2	6	5	4
4	5	3	4	3
4	5	4	5	1

Construct a discrete line of distribution of workers according to their skill level (categories), show it graphically. Draw conclusions.

2. We have the following data about the experience of work of employees of a credit department:

7 8 2 18 4 5 0 6 8 7 10 12 1 2 4 3 11 3 1 2 3 4 7  
 4 1 3 15 12 13 19 12 3 0 3 5 1 14 1 8 9 25 12 8 21 2 2  
 16 8 4 20 5 7 18 13 14 5 4 11 23 9

Construct an interval variation line of distribution. Specify the elements.

Show the distribution of workers on a histogram. Draw conclusions.

3. We have some data about the cost of the fixed capital and commodity output of industrial enterprises during a year (Table 3.4).

Table 3.4

**The distribution of enterprises according to the output and cost of the fixed capital**

The number of the enterprise	The cost of the fixed capital, thousand Euros	The commodity output, thousand Euros	
		Under the plan	Actual
1	480	560	590
2	690	1 435	1 470
3	1 750	3 535	3 742
4	3 420	7 150	8 210
5	1 146	1 124	1 217
6	1 205	1 784	1 955
7	1 308	2 966	2 994
8	364	408	409
9	2 120	3 151	3 276
10	1 332	2 150	2 166
11	1 000	2 100	2 300
12	4 073	10 967	12 099
13	1 014	2 976	2 138
14	586	978	992
15	905	1 296	1 404
16	1 746	4 010	4 157
17	3 140	6 572	6 206
18	529	1 088	1 057
19	2 861	6 068	6 412
20	407	618	724
21	1 990	3 912	4 075

Make a grouping of the enterprises according to the cost of the fixed capital, having selected the groups: up to 500 000 Euros, from 501 000 up to 1 000 000 Euros, from 1 001 000 up to 1 500 000 Euros, from 1 501 000 up to 2 500 000 Euros, above 2 501 000 Euros. In each group, count the number of the enterprises, the cost of the fixed capital, the commodity output under the plan and the actual one, the percent of the plan performance, the fund feedback under the plan and the actual one. Calculate the relative density of each group in the general actual output.

Tabulate the results of the calculation and draw an analytical conclusion.

4. We have some data about the sum of the capital and profit of 26 commercial banks (Table 3.5).

Table 3.5

**The distribution of banks according to the amount of the capital and profit**

The number of the bank	The capital, UAH m	The profit, UAH m
1	6.3	4.7
2	11.8	8.6
3	7.6	5.3
4	10.5	8.8
5	8.1	6.2
6	8.3	4.1
7	12.0	8.2
8	5.1	3.6
9	7.8	4.1
10	5.4	3.3
11	6.3	5.1
12	8.3	5.8
13	5.4	3.5
14	6.3	4.8
15	8.4	7.1
16	5.4	4.0
17	7.0	5.8
18	9.6	7.8
19	8.1	6.9
20	5.2	4.3
21	7.3	6.0
22	8.2	6.4
23	5.4	4.1
24	3.2	2.8
25	4.4	3.0
26	3.0	1.2

Construct:

1) the combinational distribution of banks according to these attributes, form three groups with equal intervals and draw a conclusion;

2) the analytical grouping which reflects the dependence of profit on the capital. Present the results of the grouping in a tabulated form.

5. We have some data about the percentage of the privatized property and the profitability of spending of 20 companies (Table 3.6).

Table 3.6

**The distribution of companies according to the profitability  
and percentage of the privatized property**

The number of the enterprise	The privatized property, %	The profitability of spending, %
1	95	31
2	81	22
3	68	24
4	66	15
5	79	28
6	81	29
7	90	30
8	100	37
9	56	20
10	75	27
11	86	29
12	55	18
13	81	30
14	58	21
15	94	27
16	77	23
17	48	19
18	78	25
19	100	36
20	70	29

Make a grouping of the companies according to the percentage of the privatized property based on the groups with equal intervals.

Construct a combinational distribution and an analytical grouping.

### **The list of questions for independent work**

1. What is the statistical summary, its tasks, the essence and types? Describe them briefly.

2. What is the essence of centralized and decentralized summarization? Give examples and define the terms of the use of summarization.

3. What is the role and tasks of statistical groupings?

4. What is statistical grouping? What are the types of groupings? Describe them briefly and give examples.

5. What is the advantage of combining groupings compared to simple groupings?

6. What are the possibilities of factor grouping?

7. How is the number of groups determined in attributive grouping?

8. To which grouping attributes – qualitative or quantitative – do the following belong: a) the age of a person; b) profession; c) ownership form; d) wages?

9. Which of the following groupings is typological:

a) grouping of population based on the nationality;

b) grouping of enterprises based on the level of profitability;

c) grouping of enterprises based on the form of ownership?

10. Describe the main principles of the methodology of statistical groupings.

## **Topic 4. Generalizing the statistical indicators**

### **Laboratory work 4. Analysis of data using the generalizing statistical indicators**

*The purpose of the work* is to acquire the skills in the calculation of relative and mean values.

*The task* is to use MS Excel to analyze the statistical data using relative and mean values and to provide an economic interpretation of results.

### ***The guidelines***

Absolute values characterize the size of socioeconomic phenomena. This refers to the aggregate amount or its certain parts (a number of elements) and the corresponding values of volume attributes.

Absolute statistical indicators are of indisputable statistical significance in management, but in-depth socioeconomic analysis of the facts requires various comparisons. The values of statistics are compared in time (in relation to one object) and in space (between the objects), different features of the same object are correlated.

The result of the comparison is the relative statistical indicator that characterizes the measure of quantitative correlation of dissimilar or similar indicators.

Each relative indicator is a fraction, the numerator of which is the comparable value, and the denominator is the base of comparison. A relative indicator shows how many times the comparable value exceeds the baseline or the proportion of the base of comparison.

The **relative indicator of dynamics (RID)** is calculated using the following formula:

$$\text{RID} = \frac{Y_1}{Y_0}, \quad (4.1)$$

where  $Y_0$  is the base level of the statistical indicator;

$Y_1$  is the actual level of the statistical indicator.

The **relative indicator of plan performance (RIPP)** is calculated as:

$$\text{RIPP} = \frac{Y_1}{Y_{pl}}, \quad (4.2)$$

where  $Y_{pl}$  is the planned level of the statistical indicator.

The **relative indicator of a planned target** is calculated as:

$$\text{RIPT} = \frac{Y_{pl}}{Y_0}. \quad (4.3)$$

If we have ungrouped data, we use simple means, if we have grouped data, we use weighted means.



The **simple arithmetic mean** ( $\bar{X}$ ) is calculated as:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}, \quad (4.4)$$

where  $x_i$  is the numerical value of the data set;  
 $n$  is the total number of items in the set.

The **weighted arithmetic mean** is calculated as:

$$\bar{X} = \frac{\sum_{i=1}^n x_i \times f_i}{\sum_{i=1}^m f_i}, \quad (4.5)$$

where  $f_i$  is the weight of the  $i$ -th group or the frequency of occurrence of the set item.

The **simple harmonic mean** is calculated as:

$$\bar{X} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}. \quad (4.6)$$

The **weighted harmonic mean** is calculated as:

$$\bar{X} = \frac{\sum_{i=1}^n M_i}{\sum_{j=1}^n \frac{1}{x_j} \times M_j}, \quad (4.7)$$

where  $M_j$  is the weight equal to  $x_j$  multiplied by  $f_j$ .

The **simple geometric mean** is calculated as:

$$\bar{X} = \sqrt[n]{x_1 \times x_2 \times x_3 \times \dots \times x_n} = \sqrt[n]{\prod_{i=1}^n x_i}. \quad (4.8)$$

The **weighted geometric mean** is calculated as:

$$\bar{X} = \sqrt[w]{\prod_{i=1}^m x_i^{w_i}}, \quad (4.9)$$

where  $w$  is the sum of the weights  $w_1, w_2, \dots, w_n$ .

Let's consider an example.

**Task 1.** We have some data about the sales of the industry products (Table 4.1).

Table 4.1

### Information about the sales

Enterprise number	Sales in the 1st quarter, UAH	Plan for the 2nd quarter, UAH	Sales in the 2nd quarter, UAH
1	8 500	8 650	8 720
2	7 540	7 650	7 700
3	9 900	10 400	10 670
4	5 900	6 000	5 950

Calculate the relative statistic indicators that characterize the planned target, the performance of the plan and the dynamics of sales for each enterprise and for the industries as a whole. Draw a conclusion.

The relative indicator of the planned target is calculated as the ratio of the planned feature in the current period to the actual characteristics of the previous period. The calculation formulas in the program Excel (Fig. 4.1) begin with the sign "=" in the cell *E2*. Select the value that must be in the numerator *C2*, put the division sign "/" and select the value that must be in the denominator *B2* and multiply the fraction by 100 ( $C2 / B2 \times 100$ ), press *<Enter>* in order to obtain the numerical result. The indicators for other enterprises are calculated in a similar way or you can select the result and apply it to the column.

	A	B	C	D	E
1	Enterprise number	Sales in the 1st quarter, UAH	Plan for the 2nd quarter, UAH	Sales in the 2nd quarter, UAH	Relative indicator of the planned target
2	1	8500,00	8650,00	8720,00	=C2/B2*100
3	2	7540,00	7650,00	7700,00	101,46
4	3	9900,00	10400,00	10670,00	105,05
5	4	5900,00	6000,00	5950,00	101,69

**Fig. 4.1. The calculation of the relative indicator of the planned target**

The calculation of the relative indicator of the performance of the plan and the relative indicator of the dynamics is similar to the calculation of the relative indicator of the planned target (Fig. 4.2 and 4.3). The numerator and denominator of these variables are relevant indicators which are used for the calculations.

	A	B	C	D	E	F
1	Enterprise number	Sales in the 1st quarter, UAH	Plan for the 2nd quarter, UAH	Sales in the 2nd quarter, UAH	Relative indicator of the planned target	Relative indicator of the plan performance
2	1	8500,00	8650,00	8720,00	101,76	=D2/C2*100
3	2	7540,00	7650,00	7700,00	101,46	100,65
4	3	9900,00	10400,00	10670,00	105,05	102,60
5	4	5900,00	6000,00	5950,00	101,69	99,17

**Fig. 4.2. The calculation of the relative indicator of the plan performance**

	A	B	C	D	E	F	G
1	Enterprise number	Sales in the 1st quarter, UAH	Plan for the 2nd quarter, UAH	Sales in the 2nd quarter, UAH	Relative indicator of the planned target	Relative indicator of the plan performance	Relative indicator of the dynamics
2	1	8500,00	8650,00	8720,00	101,76	100,81	=D2/B2*100
3	2	7540,00	7650,00	7700,00	101,46	100,65	102,12
4	3	9900,00	10400,00	10670,00	105,05	102,60	107,78
5	4	5900,00	6000,00	5950,00	101,69	99,17	100,85

**Fig. 4.3. The calculation of the relative indicator of the dynamics**

In order to determine these indicators for the industry as a whole we need to calculate the sum of the input data (in columns) using *AutoSum* "Σ" on the toolbar Excel or the formula "*=SUM()*" (Fig. 4.4). The calculation of relative indicators of the planned target, the performance of the plan and the dynamics for the industry are calculated similar to the calculation of indicators for each enterprise based on the ratio of two values (Fig. 4.5).

	A	B	C	D
1	Enterprise number	Sales in the 1st quarter, UAH	Plan for the 2nd quarter, UAH	Sales in the 2nd quarter, UAH
2	1	8500,00	8650,00	8720,00
3	2	7540,00	7650,00	7700,00
4	3	9900,00	10400,00	10670,00
5	4	5900,00	6000,00	5950,00
6	Total	=SUM(B2:B5)		33040,00

Fig. 4.4. The calculation of the sum sales for the industry

	A	B	C	D	E
1	Enterprise number	Sales in the 1st quarter, UAH	Plan for the 2nd quarter, UAH	Sales in the 2nd quarter, UAH	Relative indicator of the planned target
2	1	8500,00	8650,00	8720,00	101,76
3	2	7540,00	7650,00	7700,00	101,46
4	3	9900,00	10400,00	10670,00	105,05
5	4	5900,00	6000,00	5950,00	101,69
6	Total	31840,00	32700,00	33040,00	=C6/B6*100

Fig. 4.5. The calculation of the indicators for the industry

The relative indicator of the planned target characterizes the plan change in the current period compared to the actual value of the previous period. It can be concluded based on the first enterprise – in the current period the enterprise planned to realize 1.76 % more products than actually realized in the first quarter. For the industry it is planned to realize 2.7 % more products than in the previous period.

The relative indicator of the plan performance shows the percentage of the actual value changes compared to the plan. In the second quarter, the second enterprise sold 0.65 % more products than it was planned. The industrial enterprises sold 1.04 % more products than it was planned.

The relative indicator of the dynamics determines the dynamics of the changes of the socioeconomic phenomenon in time. It can be concluded that in the second quarter all the enterprises sold 3.77 % more products than in the first one.

**Task 2.** Consider the combination of using the relative indicators of the dynamics, the structure and coordination (Table 4.2).

The relative indicators of the structure are calculated as the ratio of the share to the whole, that is the row line "Raw materials" (B3) is divided by the

row line "Total" (B5) and multiplied by 100 % (Fig. 4.6). The sign "\$" in the denominator fixes the cell if then you apply the formula to the rest of the cells.

Table 4.2

**The material reserves of the company**

Material reserves	Reserves, thousand UAH, at the end of the quarter		The 4th quarter as % of the 3rd quarter	The structure of reserves, % by the end of the quarter		Structural shifts, per cent points
	3	4		3	4	
Raw materials	119	122.5	102.5	68	62	-6
Finished product	56	74	132.1	32	38	+6
Total	175	196	112.0	100	100	0

Structural changes are calculated as the difference between the structure of reserves in the fourth and third quarters.

	A	B	C	D	E	F	G
1	Material reserves	Reserves, thousand UAH, at the end of the quarter		The 4th quarter as % of the 3rd quarter	Structure of reserves, % by the end of the quarter		Structural shifts, per cent points
2		3	4		3	4	
3	Raw materials	119	122,5	102,5	=B3/\$B\$5*100		
4	Finished product	56	74	132,1			
5	Total	175	196	112			

Fig. 4.6. The calculation of the relative indicator of the structure

In the fourth quarter, the reserves generally increased 1.12 times, i.e. by 12 % (Fig. 4.7). As the reserves consist of two functionally distinct components, it is necessary to assess the dynamics of each of them. So, the raw materials increased only by 2.5 %, while the reserves of the finished product – by 32.1 %. Irregularity of dynamics of individual components has led to changes in the structure of the reserves. In the third quarter, the share of the raw materials was 68 %, in the fourth quarter it decreased to 62 %, that is by 6 percentage points. Due to the structural changes the proportions between the constituent parts have changed: in the third quarter, 1 UAH of the finished product reserves accounted for 2.125 UAH of the raw material reserves (119 : 56 = 2.125), in the fourth quarter this proportion decreased to 1.65 (122 : 74 = 1.65).

	A	B	C	D	E	F	G
1	Material reserves	Reserves, thousand UAH, at the end of the quarter		The 4th quarter as % of the 3rd quarter	Structure of reserves, % by the end of the quarter		Structural shifts, per cent points
2		3	4		3	4	
3	Raw materials	119	122,5	102,5	68	62,5	=F3-E3
4	Finished product	56	74	132,1	32	37,8	
5	Total	175	196	112	100	100	

Fig. 4.7. The calculation of the structural changes

Most socioeconomic phenomena are characterized by additivity of volumes (sugar, fuel, etc.), so the arithmetic mean is the most common one. It is calculated by dividing the total volume of attribute values by the total number of items in the set. The simple arithmetic mean is calculated for primary ungrouped data.

**Task 3.** An insurance company paid insurance benefits for five damaged objects, thousand UAH: 18, 27, 22, 30, 23.

In order to calculate in Excel the average payment of insurance, one should put the sign "=" into the cell B2 and select *Function* (f) on the formula bar. The possibilities of this module are presented in the window (Fig. 4.8).

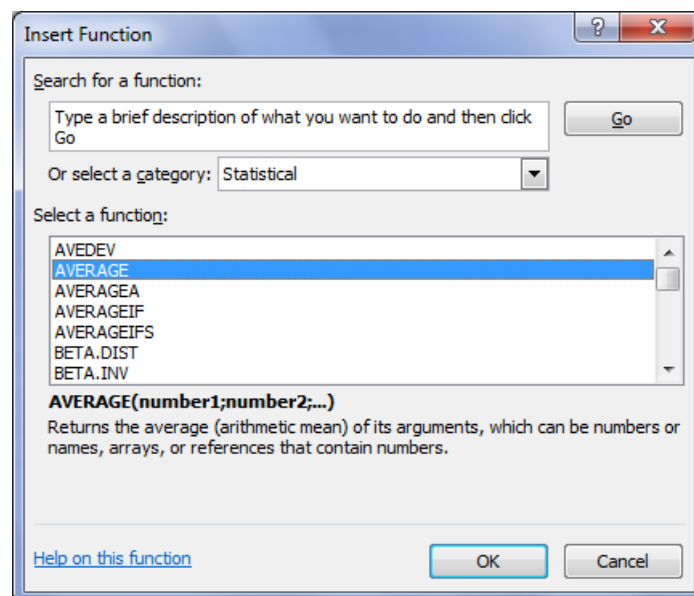


Fig. 4.8. The *Wizard Functions* dialog window

In order to calculate the mean value, one should select the *AVERAGE* function, and in the dialog window *Function Arguments* – (B3:F3) (Fig. 4.9), click "OK" (Fig. 4.10).

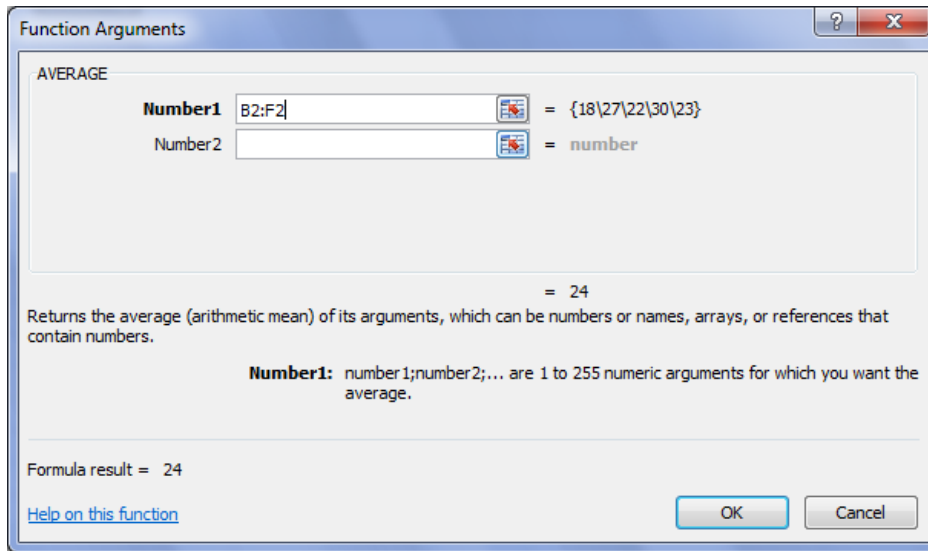


Fig. 4.9. The *Function Arguments* dialog window

	A	B	C	D	E	F
1	Damaged objects	1	2	3	4	5
2	Insurance benefits, ths. UAH	18	27	22	30	23
3	Average payment	24				

Fig. 4.10. The result of the calculation of the average insurance payment

On the average, the insurance company paid insurance benefits for five damaged objects in the amount of 24 thousand UAH.

**Task 4.** We have some data concerning two companies that produce the same products (Table 4.3).

Table 4.3

### The input data

Company	2018		2019	
	Time expenditure per unit, hours	Produced units	Time expenditure per unit, hours	Time expenditure on all products, hours
A	2.0	150	1.9	380
B	3.0	250	3.05	840

Determine the average time expenditure per unit by two companies in 2018 – 2019.

To determine the average time expenditure in 2018, use the formula of the arithmetic weighted average, because we have an attribute value (time expenditure per unit is "x") and the frequency of its appearance (produced units – "f"). Enter the formula " $=B3*C3$ " into the cell F3 and apply this formula to the whole column. In the cell F5, calculate the sum of the column: press the sign *AutoSum* ( $\Sigma$ ) on the toolbar or press "=" and choose the *SUM* formula in the formula bar. Calculate the sum in the cell C5 (Fig. 4.11).

	A	B	C	D	E	F
1		2018		2019		
2	Company	Time expenditure per unit, hours	Produced units	Time expenditure per unit, hours	Time expenditure on all products,	x*f
3	A	2	150	1,9	380	=B3*C3
4	B	3	250	3,05	840	=B4*C4

Fig. 4.11. The calculation of the arithmetic weighted average

To calculate the average value, enter the formula " $=F5/C5$ " into the cell G5 and press <Enter> and get the result: in 2018 the average time expenditure on production of companies A and B was 2.625 hours.

In order to determine the average time expenditure in 2019, we need to use the formula of the weighted harmonic mean value as we have the signs "x" (time spent per unit of product) and "M" general expenses. Enter the formula " $=E3/D3$ " in the cell H3 and apply this formula to the whole column. Calculate the sum of the column in the cell H5: press the *AutoSum* ( $\Sigma$ ) sign on the toolbar or press "=" and choose the *SUM* formula in the formula bar. Calculate the sum in the cell E5 (Fig. 4.12).

	A	B	C	D	E	F	G	H	I
1		2018		2019					
2	Company	Time expenditure per unit, hours	Produced units	Time expenditure per unit, hours	Time expenditure on all products, hours	x*f	The average time spent in 2018	M/x	The average time spent in 2019
3	A	2	150	1,9	380	=B3*C3		=E3/D3	
4	B	3	250	3,05	840	=B4*C4		=F4/D4	
5			=SUM(C3:C4)		=SUM(E3:E4)	=SUM(F3:F4)	=F5/C5	=SUM(H3:H4)	=E5/H5

Fig. 4.12. The calculation of the harmonic weighted mean



To calculate the average value, enter the formula " $=E5/H5$ " in the cell I5, press <Enter> and get the result: in 2018 the average time spent on production of companies A and B was 2.57 hours.

### The tasks for independent work

**Task 1.** We have some data about the population of a region which has secondary and higher education, thousand people (Table 4.4).

Table 4.4

#### The distribution of the population of a region according to the educational level

Indicator	Years				
	1999	2004	2009	2014	2019
People having secondary and higher education, total including:	58.7	95.0	139.1	164.3	177.3
higher complete	3.8	8.3	14.8	20.8	28.4
higher incomplete	1.7	2.6	3.2	3.5	7.5
special secondary	7.9	13.4	23.5	30.9	30.4
secondary	9.9	23.4	45.1	65.4	66.3
incomplete secondary school	35.4	47.3	52.5	43.7	44.7

Determine:

the structure of the population based on the education;

the number of people with higher education relative to 100 people with incomplete secondary education;

the dynamics of the population based on the received education (base = 1999).

Specify the type of the calculated relative indicators. Present the data in the form of a statistical table.

**Task 2.** The planned output growth for the current year in a branch should make 4.6 %. The actual output growth has made 106.5 % in this year. Determine the share of the plan performance as for the output.

**Task 3.** We have some data about the level of profitability of some enterprises of the food industry (Table 4.5).

Table 4.5

**The distribution of the enterprises according to the level of profitability**

Profitability of production, %	The number of enterprises
Less than 5	12
5 – 10	24
10 – 15	8
15 – 20	4
20 and more	2

Determine the average level of profitability for the enterprises of the food industry. Explain how you did the calculations and draw a conclusion.

**Task 4.** The discipline of businesses in different types of economic activity, concerning the payment of taxes is characterized by the following data (Table 4.6).

Table 4.6

**The distribution of the enterprises concerning the payment of taxes**

Type of activity	The number of enterprises which pay taxes		The sum of the tax which is paid by one enterprise, mln UAH
	Total, thousand people	% to all the registered businessmen	
Industry	8	60	1.7
Trade	19	70	1.2
Transport	37	53	4.3

Determine the average share of the enterprises, which pay taxes and the average size of the tax paid by one enterprise.

**Task 5.** We have some data about the time spent on production of the same product in different factories (Table 4.7).

Table 4.7

**The time spent on production in factories**

Factory	Time expenditure per unit of production, hours	Time spent on the production of all products, hours
1	45.0	2 250.0
2	52.0	3 200.0
3	60.0	1 800.0
4	58.0	6 380.0

Determine the average time expenditure per unit of production for all factories together. Explain the choice of the type of average.

**The list of questions for independent work**

1. What are absolute statistical values? What is their role and significance in statistics?
2. What are the relative values, their role and significance in statistics?
3. Name the types of relative values, describe them and give examples.
4. Give a definition of the average value, its essence and significance in the statistical data analysis.
5. What kind of average values do you know? Write their formulas.
6. When are simple and weighted means used?
7. What are the conditions for applying averages?

8. Name the main properties of the arithmetic mean. What is the essence of the simplified calculation of the arithmetic mean?
9. In which cases are the arithmetic and harmonic mean values applied?
10. In what cases is the geometric mean applied?

## **Topic 5. Analysis of the distribution series**

### **Laboratory work 5. Analysis of the statistical distribution series**

*The purpose of the work* is to acquire the skills in the calculation of the indicators of distribution series.

*The task* is to use MS Excel to analyze the statistical distribution series and to provide an economic interpretation of calculated indicators.

#### ***The guidelines***

The **distribution series** is ordered distribution of population units into groups according to a certain feature.

The **elements** of the distribution series are:

**variant** ( $x_i$ ) – the value of a grouping sign, which can vary;

**frequency** ( $f_i$ ) – which shows how often individual values of a data set are repeated.

The three major properties which describe a set of numerical data are: **central tendency, variation, shape.**

The **central tendency** is the extent to which all the values group around a typical or central value.

The **variation** is the **value** of dispersion, or scattering, of values away from a central value.

The **shape** is the pattern of the distribution of values from the lowest value to the highest value.

The **mean, mode** and **median** are all valid measures of the central tendency.

The **mode** is the most frequent score in our data set. When data are already grouped in a frequency distribution, we can assume that the mode is

located in the class with most items. In order to determine a single value for the **mode (Mo)** from this modal class, we use the formula:

$$Mo = x_{mo} + i_{mo} \times \frac{f_{mo} - f_{mo-1}}{(f_{mo} - f_{mo-1}) + (f_{mo} - f_{mo+1})}, \quad (5.1)$$

where  $x_{mo}$  is the lower class boundary of the modal group;

$i_{mo}$  is the modal group width;

$f_{mo}$  is the frequency of the modal group;

$f_{mo-1}$  is the frequency of the group before the modal group;

$f_{mo+1}$  is the frequency of the group after the modal group.

The **median** is the middle score for a set of data that has been arranged in order of magnitude.

If there is an odd number of values in the data set, the median is the middle-ranked value.

If there is an even number of values in the data set, the **median** is the average of the two middle-ranked values.

To determine **median** for grouped data in a frequency distribution, we can assume that the median is located in the first class with the value of cumulative frequency equal to at least  $n/2$  ( $n$  is the total frequency). In order to determine a single value for the **median (Me)** from this modal class, we use the formula:

$$Me = x_{me} + i_{me} \times \frac{\frac{\sum_{i=1}^n f_i}{2} - S_{me-1}}{f_{me}}, \quad (5.2)$$

where  $x_{me}$  is the lower class boundary of the median group;

$i_{me}$  is the median group width;

$f_i$  is the frequency of the  $i$ -th group of the distribution;

$S_{me-1}$  is the cumulative frequency of the group before the median group;

$f_{me}$  is the frequency of the median group.

The **quartiles (Q)** split a set of data into four equal parts.

The **number** of the first quartile (**Q<sub>1</sub>**) and the third quartile (**Q<sub>3</sub>**) for ungrouped data are calculated as follows:

$Q_1 = (n+1)/4$  ranked value;

$Q_3 = 3(n+1)/4$  ranked value.

Using the same method of calculation as with the median, we can get the **first quartile (Q<sub>1</sub>)** and the **third quartile (Q<sub>3</sub>)** for grouped data:

$$Q_1 = x_{Q_1} + i_{Q_1} \times \frac{\frac{\sum_{i=1}^n f_i}{4} - S_{Q_1-1}}{f_{Q_1}}; \quad (5.3)$$

$$Q_3 = x_{Q_3} + i_{Q_3} \times \frac{\frac{3}{4} \cdot \sum_{i=1}^n f_i - S_{Q_3-1}}{f_{Q_3}}. \quad (5.4)$$

The **decile (D)** divides the population into ten parts and the **first decile (D<sub>1</sub>)** for grouped data is:

$$D_1 = x_{D_1} + i_{D_1} \times \frac{\frac{\sum_{i=1}^n f_i}{10} - S_{D_1-1}}{f_{D_1}}. \quad (5.5)$$

The **standard deviation (σ)** is the square root of **variance (σ<sup>2</sup>)**. It is a measure of the extent to which data varies from the mean. For ungrouped data **standard deviation (σ)** is defined using the formula:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}. \quad (5.6)$$

For grouped data:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \times f_i}{\sum_{i=1}^n f_i}}. \quad (5.7)$$

The **coefficient of variation** ( $V_\sigma$ ) is a measure of relative variability. It is the ratio of the standard deviation to the mean (average):

$$V_\sigma = \frac{\sigma}{\bar{x}} \times 100\%. \quad (5.8)$$

A statistical distribution is considered homogeneous if the coefficient of variation does not exceed 33 %.

The **within group variance** ( $\sigma_i^2$ ) is calculated as:

$$\sigma_i^2 = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)^2 \times f_i}{\sum_{i=1}^n f_i}. \quad (5.9)$$

The **intergroup variance** ( $\delta^2$ ) is:

$$\delta^2 = \frac{\sum_{i=1}^n (\tilde{x}_i - \bar{x}_i)^2 \times f_i}{\sum_{i=1}^n f_i}. \quad (5.10)$$

The **average of the group variance** ( $\overline{\sigma^2}$ ) is:

$$\overline{\sigma^2} = \frac{\sum_{i=1}^n \sigma_i^2 \times f_i}{\sum_{i=1}^n f_i}. \quad (5.11)$$

The **total variance** is:

$$\sigma_{\text{total}}^2 = \overline{\sigma^2} + \delta^2. \quad (5.12)$$

The **coefficient of determination** is:

$$\eta^2 = \frac{\delta^2}{\sigma_{\text{total}}^2}. \quad (5.13)$$

The **correlation ratio** is:

$$\eta = \sqrt{\eta^2}. \quad (5.14)$$

Let's consider two examples.

### **Task 1. Analysis of ungrouped statistical data.**

The work starts with running Excel (similar to running the other applications – using the *Start* menu or by means of the shortcut). The indicators are not necessarily entered into the input spreadsheet when forming the file, you can move it from Microsoft Office documents through the clipboard. Then you can transform and visualize the data. We have some input data (Fig. 5.1).

	A	B
	Enterprise	Cost price of the product, UAH
1		
2	1	499,04
3	2	518,78
4	3	495
5	4	472,96
6	5	626,96
7	6	506,12
8	7	518,52
9	8	502,93
10	9	519,97

Fig. 5.1. **The input data**

The *Descriptive Statistics* mode is used to generate a one-dimensional statistical report on the basic indicators of position, scattering and asymmetry of the population which is analyzed. To change to this mode, you need to enter the menu item *Data – Data Analysis* and select this mode (Fig. 5.2).



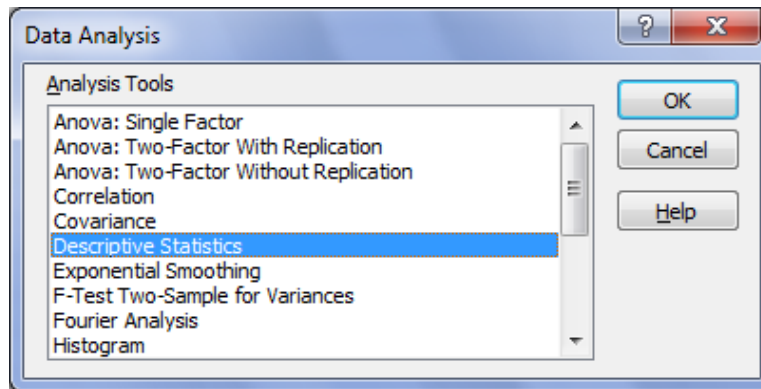


Fig. 5.2. **Selecting the *Descriptive Statistics* mode**

In the dialog box of this mode (Fig. 5.3) the following parameters are given:

1. The *Input Range*, i.e. entering the cells that contain statistical data.
2. *Grouping*, i.e. setting the position according to columns or according to rows, depending on the location of the data in the input range.
3. The *Labels in the first row* which are activated if the first row (column) in the input range contains column labels. If you have no column labels, this parameter should be deactivated. In this case standard names for the output range data will be automatically created.
4. The *Output Range, New Worksheet, New Workbook*.

In the *Output Range* position you need to enter a link to the upper-left cell of the output range in the box. The output range is automatically detected, and a message appears in the case of the possible imposition of the original range of the input data.

In the *New Worksheet* position a new worksheet opens. Excel will enter the results starting with the cell *A1*. If you want to specify the name of the new worksheet, enter the name in the box located in front of it.

In the *New Workbook* position a new workbook opens. Excel will enter the results in the cell *A1* on the first worksheet.

5. *Summary Statistics* which is activated if the output range gets one field for each Descriptive Statistics indicator.

6. The *Confidence Level for Mean* which is activated if the output table includes a row for a maximum sampling error at the prescribed level of reliability.

7. The *K-th Largest* which is activated if the row for the *k*-th largest (from a maximum *x<sub>max</sub>*) value of the sample should be included in the output table. If  $k = 1$ , the row will contain the maximum value of the sample.

8. The *K-th Smallest* which is activated if the output table includes the row for the *k*-th smallest (from *x<sub>min</sub>*) value of the sample. Enter a number *k* in the box. If  $k = 1$ , then the row will contain the minimum value of the sample.

The input parameters of the *Descriptive Statistics* mode are presented in Fig. 5.3, and the calculated parameters of this mode – in Fig. 5.4.

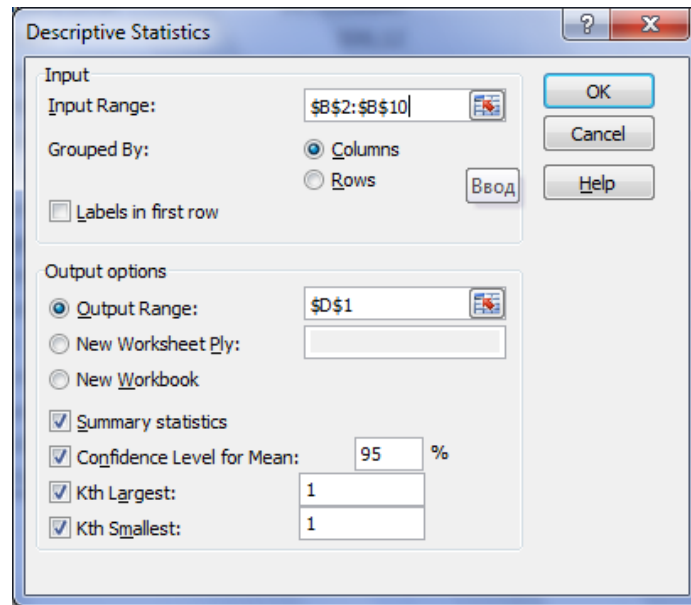


Fig. 5.3. The parameters of the *Descriptive Statistics* mode

	A	B	C	D	E
	Enterprise	Cost price of the product, UAH			
1	1	499,04			
3	2	518,78		Mean	517,8088889
4	3	495		Standard Error	14,51150649
5	4	472,96		Median	506,12
6	5	626,96		Mode	#N/D
7	6	506,12		Standard Deviation	43,53451948
8	7	518,52		Sample Variance	1895,254386
9	8	502,93		Kurtosis	6,349512563
10	9	519,97		Skewness	2,311803096
11				Range	154
12				Minimum	472,96
13				Maximum	626,96
14				Sum	4660,28
15				Count	9
16				Largest(1)	626,96
17				Smallest(1)	472,96
18				Confidence Level(95.0%)	33,46359398

Fig. 5.4. The calculated parameters of the *Descriptive Statistics* mode

According to the data, the **coefficient of oscillation** is calculated as follows:  $K_R = \frac{R}{\bar{X}} \times 100 \% = \frac{154.00}{517.81} \times 100 \% = 29.74 \%$ ;

and the **coefficient of variation** is calculated as:

$$K_\sigma = \frac{\sigma}{\bar{X}} \times 100 \% = \frac{43.53}{517.81} \times 100 \% = 8.40 \%$$

**Task 2. Analysis of the grouped statistical data.** After entering or moving the input data from Microsoft Office documents via the clipboard into a spreadsheet Excel, if necessary, you can transform and visualize the row data. In this case, you need to add a column with individual values of signs (the midpoint of the interval) for each group (Fig. 5.5).

The function for calculating the weighted arithmetic average is not available in Excel. But it can be obtained by combination of other functions.

The cell C9 contains the formula `"=SUMPRODUCT(C3:C8;B3:B8)/SUM(B3:B8)"`, by means of which the average current assets are calculated (Fig. 5.5).

	A	B	C
1	Input data		Output data
2	Current assets, m. UAH	The number of enterprises	Individual values of an average sign (middle of an interval)
3	less than 5	15	4
4	5 - 7	43	6
5	7 - 9	35	8
6	9 - 11	47	10
7	11 - 13	31	12
8	more than 13	27	14
9	The average current assets		9,18

Fig. 5.5. The calculation of the average current assets

In order to determine the mode (5.1) and the median (5.2), one should make some calculations (Fig. 5.6 and 5.7).

The calculation of the mode is presented in Fig. 5.6. The content of the cells is as follows:

the cell C10 contains the formula `"=MAX(B3:B8)"` with the help of which the modal number of companies is calculated;

the cell C11 contains the formula `"=MATCH(C10;B3:B8;0)"` by means of which the displacement of the modal value is calculated in an array of B3:B8;

the cell C12 contains the formula `"=INDEX(A3:A8;C11;1)"`, i.e. the modal interval of the amount of current assets in the A3:A8 array;

the cell C13 contains the formula `"=LEFT(C12;1)"` which shows the lower limit of the modal interval of the amount of current assets;

the cell C14 contains the formula `"=INDEX(B3:B8;C11-1;1)"`, i.e. the number of enterprises with the modal interval of current assets ( $f_{m0-1}$ ) in the B3:B8 array;

the cell *C15* contains the formula " $=INDEX(B3:B8;C11+1;1)$ ", i.e. the number of enterprises with a larger amount of current assets ( $f_{mo+1}$ ) in the *B3:B8* array;

the cell *C16* contains the formula " $=C12+2*((C9-C13)/((C9-C13)+(C9-C14)))$ " for calculating the mode of the amount of current assets.

	A	B	C
1		Input data	Output data
2	Current assets, m, UAH	The number of enterprises	Individual values of an average sign (middle of an interval)
3	less than 5	15	4
4	5 - 7	43	6
5	7 - 9	35	8
6	9 - 11	47	10
7	11 - 13	31	12
8	more than 13	27	14
9			
10		Modal number of companies	47
11		Displacement of modal value	4
12		Modal interval of the amount of current assets	9 - 11
13		The lower limit of the modal interval	9
14		The number of enterprises	35
15		The number of enterprises	31
16		The mode of the amount of current assets	9,86

**Fig. 5.6. The calculation of the mode**

Due to the fact that the median divides the population in half, it will be where the cumulative frequency is a half or more than half of the total amount of frequencies, and the previous cumulative frequency is less than a half of the population size (Fig. 5.7).

The content of cells (Fig. 5.7) is as follows:

in the cells *C3:C8* the cumulative frequency is calculated (for example, the cell *C5* contains the formula " $=C4+B5$ ");

the cell *B9* contains the formula " $=SUM(B3:B8)$ " by means of which the number of population (the number of enterprises) is calculated;

the cell *C10* contains the formula " $=B9/2$ " which determines a half of the number of population (50 % of enterprises);

the cell *C11* contains the formula " $=MATCH(C10;C3:C8;1)$ ", i.e. the number of the position of the largest value among the numbers less than it or equal to the middle of the interval is determined in the *C3:C8* array;

the cell *C12* contains the formula " $=INDEX(C3:C8;C11;1)$ ", i.e. a number that meets the search criteria formed in the cell *C11* and taken from the *C3:C8* array;

the cell *C13* contains the formula " $=IF(C10=C12;C11;C11+1)$ " by means of which the shift on the median interval is calculated;

the cell C14 contains the formula " $=INDEX(B3:B8;C13;1)$ " reflecting the frequency of the median interval;

the cell C15 contains the formula " $=INDEX(A3:A8;C13;1)$ ", i.e. the median interval found in the A3:A8 array;

the cell C16 contains the formula " $=LEFT(C15;1)$ " reflecting the lower limit of the median interval;

the cell C17 contains the formula " $=INDEX(C3:C8;C13-1;1)$ " calculating the value of the cumulative frequency before the median interval;

the cell C18 contains the formula " $=C16+2*((B9/2-C17)/C14)$ " calculating the median of the current assets amount.

	A	B	C
1		Input data	Output data
2	Current assets, m, UAH	The number of enterprises	Cumulative frequency
3	less than 5	15	15
4	5 - 7	43	58
5	7 - 9	35	93
6	9 - 11	47	140
7	11 - 13	31	171
8	more than 13	27	198
9	In total	198	
10		50 % of companies	99
11		The number of position of the largest value among the numbers less than it or equal to the middle of the interval	3
12		The number that meets the search criteria	93
13		Shift on the median interval	4
14		The frequency of the median interval	47
15		Median interval	9 - 11
16		Lower limit of median interval	9
17		The value of the cumulative frequency before the median interval	93
18		The median of the current assets	9.26

**Fig. 5.7. The calculation of the median**

The calculation of the mean square deviation makes it possible to determine the coefficient of variation (Fig. 5.8).

These parameters were determined with the help of the formulas in the following cells (Fig. 5.8):

the cell C10 which contains the formula " $=(SUMPRODUCT(POWER(C3:C8-C9;2);B3:B8))/SUM(B3:B8)$ " for calculating the variance;

the cell C11 which contains the formula " $=SQRT(C10)$ " for calculating the mean square (standard) deviation;

the cell C12 which contains the formula " $=(C11/C9)*100$ " for calculating the coefficient of variation.

	A	B	C
1	Input data		Output data
2	Current assets, m, UAH	The number of enterprises	Individual values of an average sign (middle of an interval)
3	less than 5	15	4
4	5 - 7	43	6
5	7 - 9	35	8
6	9 - 11	47	10
7	11 - 13	31	12
8	more than 13	27	14
9	The average current assets		9,18
10	Variance		9,05
11	Standard deviation		3,01
12	Coefficient of variation		32.76

**Fig. 5.8. The calculation of the variance, standard deviation and the coefficient of variation**

To calculate the quartiles, one should determine the upper and lower quartiles.

The calculation of the first quartile (Fig. 5.9) is similar to the calculation of the median, except for these cells:

the cell C10 containing the formula " $=B9*0,25$ ";

the cell C18 containing the formula " $=C16+2*((B9*0,25-C17)/C14)$ ".

	A	B	C
1	Input data		Output data
2	Current assets, m, UAH	The number of enterprises	Cumulative frequency
3	less than 5	15	15
4	5 - 7	43	58
5	7 - 9	35	93
6	9 - 11	47	140
7	11 - 13	31	171
8	more than 13	27	198
9	In total	198	
10		25 % of companies	49,5
11	The number of position of the largest value among the numbers less than it or equal to the middle of the interval		1
12	The number that meets the search criteria		15
13	Shift on the first quartile interval		2
14	The frequency of the first quartile interval		43
15	The first quartile interval		5 - 7
16	Lower limit of the first quartile interval		5
17	The value of the cumulative frequency before the first quartile interval		15
18	The first quartile of current assets		6.60

**Fig. 5.9. The calculation of the first quartile**

The third quartile is in the range 11 – 13 years and is equal to 11.5. To define it, use the cell B10 which contains the formula " $=B9*0,75$ ", and the cell B18 which contains the formula " $=C16+2*((B9*0,75-C17)/C14)$ ".

So, the quartile deviation is equals to:  $Q = \frac{Q_3 - Q_1}{2} = \frac{11.5 - 6.6}{2} = 4.9$ ,

and the quartile parameter of variation is:

$$K_Q = \frac{Q}{Me} \times 100\% = \frac{4.90}{9.26} \times 100\% = 52.91\%$$

The results of calculating the total variance and its components are presented in Fig. 5.10.

	A	B	C	D
1	Input data			Output data
2	Enterprises	The volume of production, m. UAH		
3		state	private	
4	1	420.00	3,980.00	
5	2	690.00	6,120.00	
6	3	790.00	6,030.00	
7	4	950.00	7,790.00	
8	5	580.00	5,050.00	
9	In total	3,430.00	28,970.00	
10	The average volume of production at two types of ownership enterprises			3,240.00
11	The average volume of production at the state enterprises			686.00
12	The average volume of the production at private enterprises			5,794.00
13	Intragroup variance (state property)			32,504.00
14	Intragroup variance (private property)			1,598,024.00
15	Average of the group variances			815,264.00
16	Intergroup variance			6,522,916.00
17	Total variance			7,338,180.00

Fig. 5.10. The calculation of the total variance and its components

The content of cells (Fig. 5.10) is as follows:

the cell *D10* contains the formula " $= (B9 + C9) / (5 + 5)$ " for calculating the average volume of production at two types of ownership enterprises;

the cell *D11* contains the formula " $= B9 / 5$ " for calculating the average volume of production at state enterprises;

the cell *D12* contains the formula " $= C9 / 5$ " for calculating the average volume of production at private enterprises;

the cell *D13* contains the formula " $= \text{VARP}(B4 : B8)$ " for calculating the intragroup variance (state enterprises);

the cell *D14* contains the formula " $= \text{VARP}(C4 : C8)$ " for calculating the intragroup variance (private enterprises);

the cell *D15* contains the formula " $= (D13 \times 5 + D14 \times 5) / 10$ " for calculating the average of the group variances;

the cell *D16* contains the formula " $= ((\text{POWER}(D11 - D10; 2)) \times 5 + (\text{POWER}(D12 - D10; 2)) \times 5) / 10$ " for calculating the intergroup variance;

the cell *D17* contains the formula " $= \text{SUM}(D15; D16)$ " for calculating the total variance.

The variance obtained and the mean square deviation of alternative signs is shown in Fig. 5.11.

	A	B	C	D	E
1	Input data				Output data
2	Batch	Finished products	among them		
3			suitable	defective	
4	1	1 200,00	800,00	400,00	
5	2	1 000,00	840,00	160,00	
6	3	1 100,00	1 000,00	100,00	
7	The average percentage of suitable products in three batches				0,80
8	The average percent of defective products				0,20
9	The variance of suitable products proportion				0,16
10	Mean square of alternative signs				0,40

Fig. 5.11. The calculation of the variance and the mean square deviation

These parameters (Fig. 5.11) were determined by means of the following formulas:

the cell *E7* contains the formula " $=\text{(SUM(C4:C6))/SUM(B4:B6)}$ " for calculating the average percentage of suitable products in three batches;

the cell *E8* contains the formula " $=1-E7$ " for calculating the average percent of defective products;

the cell *E9* contains the formula " $=E7 \times E8$ " for calculating the variance of suitable products proportion (alternative characteristics variance);

the cell *E10* contains the formula " $=\text{SQRT}(E9)$ " for calculating the mean square deviation of alternative signs.

The determination of the skewness and kurtosis coefficients involves the calculation of the moments of the third and fourth orders (Fig. 5.12).

	A	B	C
1	Input data		Output data
2	Current assets, m, UAH	The number of enterprises	Individual values of an average sign (middle of interval)
3	less than 5	15	4
4	5 - 7	43	6
5	7 - 9	35	8
6	9 - 11	47	10
7	11 - 13	31	12
8	more than 13	27	14
9	The average current assets		9,18
10	Standard deviation		3,01
11	The moment of the third order		1,06
12	Skewness coefficient		0,04
13	The moment of the fourth order		160,70
14	Kurtosis coefficient		-1,04

Fig. 5.12. The calculation of the skewness and kurtosis coefficients

The content of the cells (see Fig. 5.12) is the following:

the cell *C11* contains the formula " $=\text{(SUMPRODUCT(POWER(C3:C8-C9;3);B3:B8))/SUM(B3:B8)}$ " for calculating the moment of the third order;

the cell *C12* contains the formula " $=C11/\text{POWER}(C10;3)$ " for calculating the skewness coefficient;



the cell C13 contains the formula " $=\text{(SUMPRODUCT(POWER(C3:C8-C9;4);B3:B8))}/\text{SUM(B3:B8)}$ " for calculating the moment of the fourth order;  
the cell C14 contains the formula " $=\text{(C13}/\text{POWER(C10;4)}\text{)-3}$ " for calculating the kurtosis coefficient.

### The tasks for independent work

**Task 1.** We have some data about the distribution of workers based on the level of support payments (Table 5.1). Determine the mode and the median of material assistance of workers.

Table 5.1

### The distribution of workers depending on the level of support payments

The average size of financial assistance, thousand UAH	3.5	4	4.7	5	5.6	6.2	8	Total
The number of workers, people	10	15	22	34	45	26	18	170

**Task 2.** We have some data about the distribution of the population of Kharkiv region based on the income level in 2018 (Table 5.2).

Determine:

- the average monthly income of the residents of Kharkiv;
- the modal and median income;
- the lower (first) and upper (third) quartiles;
- the lower and upper deciles;
- the decile coefficient of differentiation of incomes.

Table 5.2

### The distribution of the population based on the income level

Income, UAH per month	The population of Kharkiv region, %
Up to 2000	1.9
2000 – 2500	4.4
2500 – 3000	6.2
3000 – 3500	14.5
3500 – 4000	13.8
4000 – 4500	11.7
4500 – 5000	17
5000 and more	30.5

**Task 3.** We have some data about the distribution of the data on the amount of loans issued according to the credit rates (Table 5.3).

Table 5.3

**The distribution of loans according to the credit rate size**

The credit rate, %	The sums of the given loans, million UAH
Up to 30	5
30 – 40	11
40 – 50	8
50 and more	6
Total	30

Determine the mean credit rate, the mode and the median. Determine the variance and the standard deviation, the coefficient of the variation. Estimate the homogeneity of data. Draw an analytical conclusion based on the results of calculations.

**Task 4.** Labor productivity at the enterprises of the coal industry is characterized by the following data (Table 5.4).

Table 5.4

**The distribution of the enterprises of the coal industry based on the labor productivity**

The enterprise	The relative share of workers, %	The monthly average coal mining per worker, thousand t	The dispersion of the coal mining
Mines	75	40	3400
Surface coal mines	25	200	2600
On the whole	100	80	–

Determine the intergroup, average of the group variance and the total variance of the coal mining per worker. Explain the sense of each variance.

**Task 5.** Consumption of natural gas in the residential sector per year is characterized by the data given in Table 5.5.

Determine the intergroup and average of the group variance of the consumption of gas if it is known that the total variance is equal to 4800.

Table 5.5

**The distribution of consumers  
according to the level of the natural gas consumption**

The consumer category	The number of consumers, thousand	The average annual gas consumption per consumer, m <sup>3</sup>
An apartment with a gas cooker	40	250
An apartment with a gas cooker and a water heater	10	400
Total	50	–

**The list of questions for independent work**

1. Explain the difference between the distribution series and the time series. Name the main components of a distribution series.
2. What statistical indicators can be used to characterize the distribution center?
3. Characterize the distribution series frequency and the use of the frequency in statistical study.
4. What are the main disadvantages of the average value?
5. Define the mode and the median. How can the mode and the median for an ungrouped series be found?
6. How can the modal and median values for a grouped series be defined?
7. For any data set, which measures of central tendency have only one value?
8. What are the properties of variation?
9. What types of variations are there and how do they differ?
10. When should quartiles be used to characterize the statistical distribution?

## Content module 2. The methods of inferential statistics

### Topic 6. Sampling and sampling distributions

#### Laboratory work 6. Analysis of statistical data with the use of sampling and sampling distributions

*The purpose of the work* is to acquire the skills in the sample observation.

*The task* is to use MS Excel to conduct a sample observation data set, calculating the standard and margin errors, construct the confidence interval.

#### *The guidelines*

**Population** and **sample** are two of the most important terms in statistics.

A **population** consists of the items or individuals about which you want to draw a conclusion.

A **sample** is the part of a population selected for analysis.

**Sampling with replacement** means that after you have selected an item, you return it to the frame, where it has the same probability of being selected again.

**Sampling without replacement** means that once you have selected an item, you cannot select it again.

A more scientific method of selection is a **table of random numbers**.

The system of conventional symbols is presented in Table 6.1.

Table 6.1

#### The symbols for indicators

Indicator	Conventional symbols	
	sample	frame (general population)
Size population	n	N
Mean	$\bar{x}$	$\bar{X}$
Proportion	w	P
Alternative proportion	(1-w)	Q
Dispersion (population variance)	$w \times (1-w)$	$p \times q$

The sampling error is the difference between the measure (a statistic) computed from a sample and a corresponding measure (a parameter) computed from the population.

The correspondence between the probability and the confidence level (*t*-value) is the most common one presented in Table 6.2.

Table 6.2

**The probability and the confidence level**

P (t)	0.683	0.950	0.954	0.990	0.997
t	1.00	1.96	2.00	2.58	3.00

The standard error is the value which measures the spread of the sample mean around the population mean.

The formula for a standard error for sampling with replacement is:

$$\mu = \sqrt{\frac{\sigma^2}{n}}. \tag{6.1}$$

The formula for a standard error for sampling without replacement is:

$$\mu = \sqrt{\frac{\sigma^2}{n} \times \left(1 - \frac{n}{N}\right)}. \tag{6.2}$$

The margin sampling error is:

$$\Delta_x = t \times \mu. \tag{6.3}$$

The margin sampling error for sampling with replacement is:

$$\Delta_x = t \times \sqrt{\frac{\sigma^2}{n}}. \tag{6.4}$$

The sampling error for sampling without replacement is:

$$\Delta_x = t \times \sqrt{\frac{\sigma^2}{n} \times \left(1 - \frac{n}{N}\right)}. \quad (6.5)$$

The confidence interval is an interval selected from sample values. This interval must include the true population parameter.

The confidence interval for the mean is:

$$\tilde{x} - \Delta_x \leq \bar{x} \leq \tilde{x} + \Delta_x. \quad (6.6)$$

Sometimes we must compute the sample error for a proportion as follows:

$$\Delta_p = t \times \mu. \quad (6.7)$$

The sampling error of a proportion for sampling with replacement is:

$$\Delta_p = t \times \sqrt{\frac{w \times (1-w)}{n}}, \quad (6.8)$$

where  $w = m/n$ .

The sampling error of a proportion for sampling without replacement is:

$$\Delta_p = t \times \sqrt{\frac{w \times (1-w)}{n} \times \left(1 - \frac{n}{N}\right)}. \quad (6.9)$$

The confidence interval for a proportion is:

$$P - \Delta_p \leq P_0 \leq P + \Delta_p. \quad (6.10)$$

Let's consider an example.

**Task 1.** An enterprise is required to conduct a 10 % sampling control of the components which are produced according to the light alloys molding technology. The initial data are shown in Table 6.3.

### The background data for the sample survey

Enter-prise No.	Weight of the compo-nents	Enter-prise No.	Weight of the compo-nents	Enter-prise No.	Weight of the compo-nents	Enter-prise No.	Weight of the compo-nents	Enter-prise No.	Weight of the compo-nents
1	16.048	11	15.989	21	16.732	31	16.154	41	16.881
2	14.442	12	16.064	22	16.289	32	16.229	42	16.955
3	14.747	13	16.138	23	14.659	33	16.305	43	17.030
4	16.194	14	16.212	24	14.968	34	14.365	44	17.104
5	15.544	15	16.287	25	16.437	35	13.485	45	17.178
6	15.618	16	16.361	26	15.777	36	15.489	46	17.252
7	15.692	17	16.435	27	15.852	37	15.570	47	17.327
8	15.766	18	16.509	28	15.927	38	16.132	48	17.401
9	15.841	19	16.584	29	16.003	39	16.694	49	16.166
10	15.915	20	16.658	30	16.078	40	16.807	50	16.243

The filled box *Sampling* of the *Data Analysis* package is shown in Fig. 6.1.

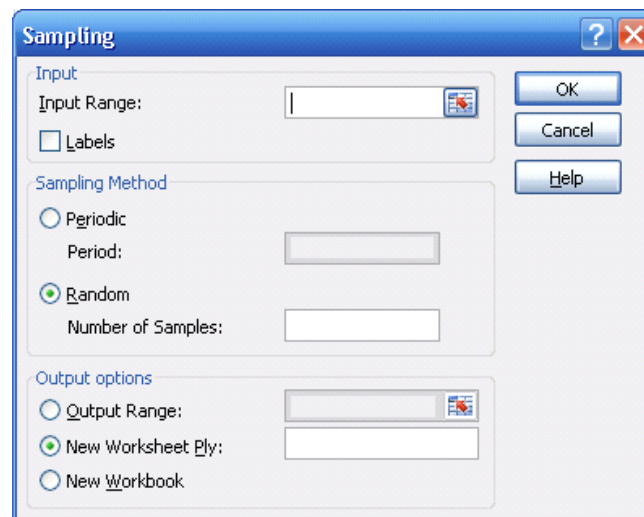


Fig. 6.1. The *Sampling* window filled with the data of the task

In the *Input Range* line, one should enter the references for the range of data that contains the population of values you want to sample. Microsoft Excel draws the samples from the first column, then the second column, and so on.

The *Labels* should be selected if the first row or column of your input range contains labels. It is clear, if your input range has no labels, Excel generates appropriate data labels for the output table.

The *Sampling Method* means that you should click *Periodic* or *Random* to indicate the sampling interval you want.

The *Period* assumes the necessity to enter the periodic interval at which you want sampling to take place.

In the *Number of Samples* line, you should enter the number of random values you want in the output column. Each value is drawn from a random position in the input range, and any number can be selected more than once.

In the *Output Range* line, enter the reference for the upper-left cell of the output table. Data is written in a single column below the cell. If you select *Periodic*, the number of values in the output table is equal to the number of values in the input range, divided by the sampling rate. If you select *Random*, the number of values in the output table is equal to the number of samples.

In the *New Worksheet Ply* line *Click* is used to insert a new worksheet in the current workbook and paste the results starting with the cell *A1* of the new worksheet. To name the new worksheet, type a name in the box.

In the *New Workbook* line, click *What* to create a new workbook and paste the results on a new worksheet in the new workbook.

After pressing the "OK" button on your work area, the result of 10 per cent sampling is displayed (Fig. 6.2).

Thus we receive a sampling frame, the characteristics of which can be calculated by means of MS Excel, with the help of *Descriptive statistics* of the *Data Analysis* add-in (see lab No. 5).

Besides the main characteristics of the distribution, the maximum sampling error has to be calculated. Due to the fact that Microsoft Excel implements random sampling with replacement, the margin error is calculated by the formula:

$$\Delta_x = t \times \sqrt{\frac{\sigma^2}{n}},$$

where  $n$  is the number of elements in a sample population;

$t$  is the confidence coefficient, which depends on the probability with which the margin sampling error is guaranteed.



	A	B
1	17,327	
2	17,178	
3	16,078	
4	15,618	
5	16,048	
6	16,732	
7	15,852	
8	15,57	
9	17,03	
10	14,659	
11	14,365	
12	16,694	
13	14,968	
14	16,078	
15	15,489	

Fig. 6.2. The result of the automatic random sampling

Fig. 6.3 displays the calculation of the sampling error limit with a 95 % probability.

D5		fx		=D4+SQRT(D2/D1)
	A	B	C	D
1	17,327		Average	=AVERAGE(A1:A15)
2	17,178		Variance	=VAR(A1:A15)
3	16,078		n	=COUNT(A1:A15)
4	15,618		t	2
5	16,048		$\Delta X$	=D4+SQRT(D2/D1)
6	16,732			
7	15,852			
8	15,57			
9	17,03			
10	14,659			
11	14,365			
12	16,694			
13	14,968			
14	16,078			
15	15,489			

Fig. 6.3. The calculation of the limit of the sampling error with a 95 % probability

In Fig. 6.4, one can see the results of the calculations.

Average	15.97907
Variance	0.813232
n	15
t	2
$\Delta_x$	2.225596

Fig. 6.4. **The results of calculating the margin sampling error**

Fig. 6.5 shows the calculation and the results of calculating the confidence interval.

Confidence interval	
lower limit	=D1-D5
upper limit	=D1+D5

Confidence interval	
lower limit	13.75347057
upper limit	18.20466276

Fig. 6.5. **The calculation and the results of calculating the confidence interval**

Thus, according to the results, one can draw a conclusion that the average for the total population with a 95 % probability belongs to the interval from 13.75 to 18.2.

### **The tasks for independent work**

**Task 1.** According to the sampling of 64 small business enterprises the average cost recovery makes 27 % at an average standard deviation 7.2 %. Determine the standard error of sample and the confidence intervals in which there is an average level of cost recovery with probability 0.997. Draw a conclusion.

**Task 2.** According to the sample, 45 % of respondents consider that advertising is the basic source of the information on the commodity market. The standard error of sample of this parameter is 2.6 %. With the probability

of 0.954, estimate the border confidence intervals of the users of advertising. Draw a conclusion.

**Task 3.** In a region, out of 200 commercial banks, 100 banks (a casual sample without replacement) were distributed based on the share of highly liquid assets (Table 6.4).

Table 6.4

**The distribution of banks based on the share of highly liquid assets**

Percent of highly liquid assets	Up to 10	10 – 15	15 – 20	20 – 25	More than 25
Number of banks	5	12	44	26	13

With the probability of 0.997, determine the borders of the average percent of highly liquid assets of 200 banks of the region.

With the probability of 0.954, determine the borders of the share of banks in which the percent of highly liquid assets is over 15 %.

Draw conclusions.

**Task 4.** At the enterprise, 100 out of 1000 workers were surveyed by casual sampling. They were distributed depending on the level of their daily productivity (Table 6.5).

Table 6.5

**The distribution of workers depending on the level of daily productivity**

Production, units	518 – 624	624 – 730	730 – 836	836 – 942	In total
Number of workers	10	30	40	20	100

Determine:

the average daily production per worker;

the average and standard error of sample with the probability of 0.954;

the probability that average daily production of the working enterprise will deviate by no more than 6 units;

the volume of sample such that with the probability 0.954 the standard error of sample for the definition of the average production would not exceed 50 units;

the share of workers of the enterprise with daily productivity of 730 units and more with probability of 0.954;

the probability that at the whole enterprise, for workers with daily productivity 730 units of product and more, the error for the selective share is no more than 2 %.

**Task 5.** According to a survey of 100 entrepreneurs (a 19 % sample), 20 of them estimate the economic conditions in the region as adverse. With the probability of 0.954, define the sampling error for the proportion of respondents who are not satisfied with the economic conditions in the region.

### **The list of questions for independent work**

1. Describe the main idea of the sample survey. Give some examples.
2. Define the general population and the sample.
3. What are the advantages and disadvantages of using selective surveillance?
4. List the principles that sample observation must meet.
5. What kind of selection methods are used?
6. What are the stages of sample observation?
7. What are the standard and margin error of sample?
8. How is the confidence interval constructed?
9. How is sample size determined?
10. What is the difference between point and interval estimates in statistics?

## **Topic 7. Time series analysis**

### **Laboratory work 7. Analysis of data using the time series methods**

*The purpose of the work* is to acquire the skills in the calculation of dynamics indicators.

*The task* is to use MS Excel to analyze the time series and to provide an economic interpretation of the calculated indicators.

#### ***The guidelines***

Time series apply to studying the change of the phenomena in time.

The analysis of speed and intensity of the phenomenon in time is carried out by means of statistical indicators received as a result of comparison of

levels. The compared level is called actual, and the level to which the comparison is made is basic. There are chain and basic indicators of the change of levels of a time series, and an average characteristic of the rows of dynamics.

If a base of comparison is the previous period, it is called a variable, and the received characteristics are called a chain. If a base of comparison is the initial level, it is called a constant, and the received characteristics are called basic.

**The absolute increase ( $\Delta$ )** shows how much the studied indicator changed in comparison with the previous (7.1) or basic period of time (7.2):

$$y_C = y_i - y_{i-1}, \quad (7.1)$$

$$y_B = y_i - y_0, \quad (7.2)$$

where  $\Delta_B$  is the base absolute increase;

$\Delta_C$  is the chain absolute increase;

$y_0$  is the base level of the time series;

$y_i$  is the actual level of the time series;

$y_{i-1}$  is the previous level of the row.

**The growth rate (Gr)** shows by how many times the studied indicator changed in comparison with the previous period of time (7.3) or with the basic period of time (7.4):

$$Gr_C = \frac{y_i}{y_0}, \quad (7.3)$$

$$Gr_B = \frac{y_i}{y_{i-1}}. \quad (7.4)$$

**The rate of increase (Ri)** shows the percentage of change of the studied indicator in comparison with the previous period of time or with the basic period of time (7.5).

$$Ri = Gr - 100. \quad (7.5)$$

The absolute value of one percent of an increase  $A\%$  shows how many absolute units are contained in 1 % of the increase (7.6):

$$A\% = \frac{y_{i-1}}{100}. \quad (7.6)$$

Let's consider some examples.

**Task 1.** We have some data on the amount of investment in the sector on a yearly basis (Table 7.1).

Table 7.1

**The volume of investments in the sector**

Years	2010	2013	2014	2015	2016	2017
The volume of investments in the sector, thousand UAH	365.00	374.00	381.00	396.00	405.00	380.00

It is necessary to calculate all the parameters of the dynamic series (the base of comparison is 2010), the average annual growth rate and the increase rate for the following periods: 1) 2010 – 2013; 2) 2013 – 2017 and 3) 2010 – 2017.

Let's create a file "Time series analysis". In column A and column B of sheet 1, the source data columns will be "Years" and "The volume of investments into the sector, thousand UAH" respectively:

Years	The volume of investments in the	Absolute increase		Growth rate (Tp),		Rate of increase		A 1 %, thousand
		Base	Chain	Base	Chain	Base	Chain	
2010	365							
2013	374							
2014	381							
2015	396							
2016	405							
2017	380							

Fig. 7.1. Creating the table with the input data

1. In order to define the absolute increase:

- enter the formula " $= B4 - \$B\$3$ " in the cell *C4*. Apply this formula to the entire column;
- enter the formula " $= B5 - B4$ " in the cell *D5* (for 2013 it is impossible to calculate the absolute increase in the chain manner, because there is no previous period – 2012) and also apply the formula to the entire column.

The results are presented in Fig. 7.2.

	A	B	C	D
1	Years	The volume of investments in the sector, thousand UAH	$(\Delta y)$ , thousand UAH	
2		UAH	Base	Chain
3	2010	365	-	-
4	2013	374	9	-
5	2014	381	16	7
6	2015	396	31	15
7	2016	405	40	9
8	2017	380	15	-25
9				

Fig. 7.2. The results of the absolute growth rate calculations

2. In order to define the growth rate:

- enter the formula " $= B4 / \$B\$3 \times 100$ " in the cell *E4*. Apply this formula to the entire column;
- enter the formula " $= B5 / B4 \times 100$ " in the cell *F5* (for 2013, the growth rate as an absolute increase, cannot be calculated by the chain method, because there is no previous period, i.e. 2012) and also apply the formula to the entire column.

3. To define the rate of increase:

- enter the formula " $= E4 - 100$ " in the cell *G4*. Apply this formula to the entire column;
- enter the formula " $= F5 - 100$ " in the cell *H5* and also apply the formula to the entire column.

4. To define the absolute value of 1 % of the increase:

- enter the formula " $= B4 / 100$ " in the cell *I5*. Apply this formula to the entire column.

The calculation results are given in Fig. 7.3.

	A	B	C	D	E	F	G	H	I
1	Years	The volume of investments in the sector, thousand UAH	$(\Delta y)$ , thousand UAH		Growth rate (Tp), %		Rate of increase (Tnp),%		A 1 %, thousand UAH
2			Base	Chain	Base	Chain	Base	Chain	
3	2010	365	-	-	-	-	-	-	-
4	2013	374	9	-	102,47	-	2,47	-	-
5	2014	381	16	7	104,38	101,87	4,38	1,87	3,74
6	2015	396	31	15	108,49	103,94	8,49	3,94	3,81
7	2016	405	40	9	110,96	102,27	10,96	2,27	3,96
8	2017	380	15	-25	104,11	93,83	4,11	-6,17	4,05
9									

Fig. 7.3. The results of the growth rate and increase rate calculations

5. Let's define the average dynamics parameters.

5.1. The average level of the series is determined in the following way. Since we have a number of missed levels, it is expedient to calculate the average level of series for the period of 2013 – 2017, i.e. to select the built-in aggregate function *AVERAGE* in the cell *B10*.

5.2. To calculate the average absolute growth rate, enter the formula  $= (B8-B3) / 8$  in the cell *B11*.

5.3. The average growth rate is calculated as follows:

1. For 2010 – 2013, one should enter the formula  $"= (B4/B3)^{(1/3)"$  in the cell *B14*.

2. For 2013 – 2017:

the basic method is to enter the formula  $"= (B8/B4)^{(1/4)"$  in the cell *B16*;  
the chain method is to enter the formula  $"= ((F5/100) \times (F6/100) \times (F7/100) \times (F8/100))^{(1/4)"$  in the cell *B17*.

3. For 2010 – 2017:

the basic method is to enter the formula  $"= (B8/B3)^{(1/7)"$  in the cell *B19*;  
the chain method of the weighted geometric mean formula is to enter the formula  $"= ((B14)^3 \times (B16)^4)^{(1/7)"$  in the cell *B20*.

5.4. The average increase rate is calculated in the following way:

1. For 2010 – 2013, one should enter the formula  $"= (B14-1) \times 100"$  into the cell *C14*.

2. For 2013 – 2017, one should enter the formula  $"= (B16-1) \times 100"$  into the cell *C16*.

3. For 2010 – 2017, one should enter the formula  $"= (B20-1) \times 100"$  into the cell *C20*.

The results of the calculations of the average values are shown in Fig. 7.4.



9			
10	The average level of the series	387,2	
11	The average absolute growth	1,875	
12		The average growth rate	The average rate of increase
13	1. 2010 - 2013		
14	basic method	1,0082	0,0082
15	2. 2013 - 2017		
16	basic method	1,0040	0,0040
17	chain method	1,4006	0,4006
18	3. 2010 - 2017		
19	basic method	1,0058	0,0058
20	chain method of weight geometric mean formula	1,0058	0,0058
21			

**Fig. 7.4. The results of determining the average dynamics**

Thus, during the period of 2010 – 2016, a steady growth in investment was observed, but in 2017 the volume of investments decreased in comparison with 2016 by 25 thousand UAH (compared to 2010 it increased by 15 thousand UAH). The growth rate in 2017, calculated by the basic method, was 104.11 %, i.e. the increase in the output was 4.11 %. The chain increase rate showed a decrease in the volume of production in comparison with 2016 by 6.17 %. A 1 % growth rate in 2017 was equal to 405 UAH. In the period of 2010 – 2017, the investment increased annually by an average of 1.875 UAH, i.e. 0.58 %. The average investment in the period of 2013 – 2017 was 387.2 thousand UAH. In the period from 2010 to 2013 the investment was growing on average by 0.82 %, from 2013 to 2017 by 0.4 %.

### **Task 2. Analysis of development trends (extrapolation).**

In theory and practice, in the process of forecasting financial indicators, quite often the extrapolation methodology is used, in which conclusions about the value of predictive indicators in future periods are drawn on the basis of studying their dynamics in the previous periods. A necessary element in this case is the construction and analysis of the so-called series of dynamics, which classifies the values of time indices in the context of individual periods and describes the dynamics of their development. We emphasize that the analysis of a number of dynamics of a separate indicator, for example, proceeds from the sale of products, is purely descriptive and does not explain the reasons for certain changes in the trend.

Extrapolation methods are used for the relatively stable development of the enterprise or in the presence of seasonal or cyclical fluctuations with a clearly

defined trend. A trend is a long-term tendency of changes in economic indicators in economic forecasting. If the development of indicators of financial and economic activity of the enterprise in previous periods is characterized by significant instability and significant fluctuations of financial indicators, their extrapolation for future periods will be impossible, and therefore, it is inappropriate to use such methods.

Perform extrapolation in Ms Excel: we have some data on the gross domestic product (GDP) of a particular country in 2006 – 2017 (Table 7.2). It is necessary to determine the forecast value of the GDP in 2018 by means of analytical alignment.

Table 7.2

**The GDP volume in 2006 – 2017**

Years	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
The GDP, billion UAH	190	210	200	220	240	230	220	240	260	260	270	280

On page 1 of the Excel workbook, columns A – B will be the columns of the input data (Fig. 7.5).

	A	B
	Years	The GDP, billion UAH
1		
2	2006	190
3	2007	210
4	2008	200
5	2009	220
6	2010	240
7	2011	230
8	2012	220
9	2013	240
10	2014	260
11	2015	260
12	2016	270
13	2017	280

Fig. 7.5. The input data in the Excel workbook

To construct a trend line, you need to build a graph of the input, select a time series and select the *Add Trend Line* command from the context menu (Fig. 7.6).

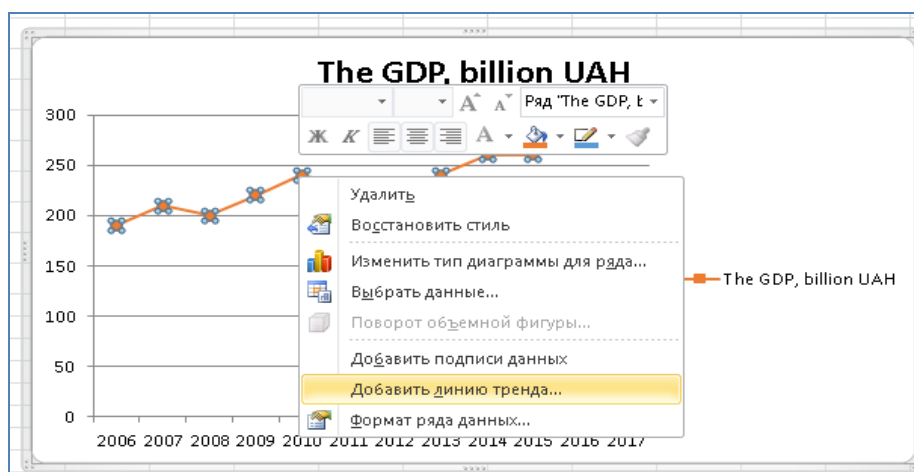


Fig. 7.6. Visualization of the input data and the command *Add Trend Line*

After choosing the command, a dialog will open, in which we choose the type and parameters of the trend construction (Fig. 7.7).

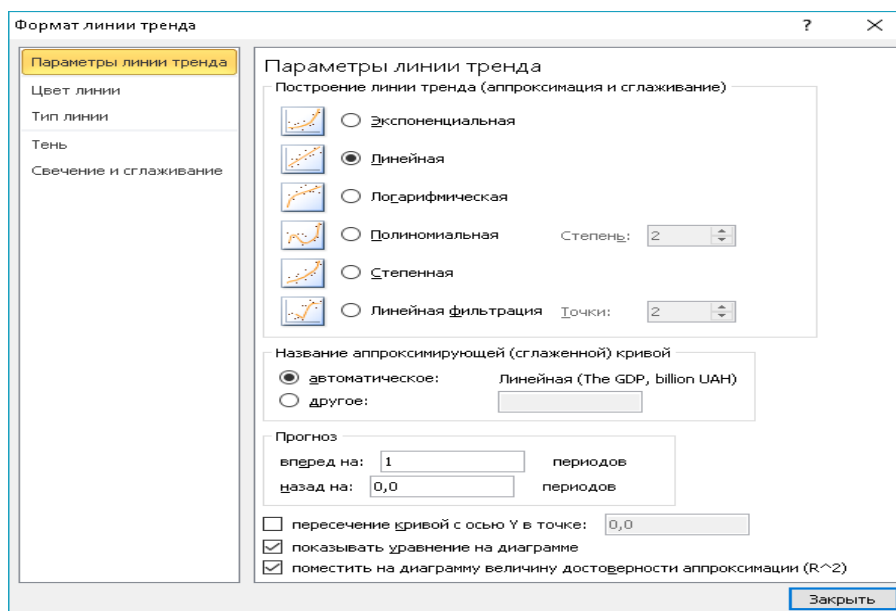


Fig. 7.7. Choosing the parameters of the *Trend Line*

The result of constructing a linear trend is shown in the Fig. 7.8.

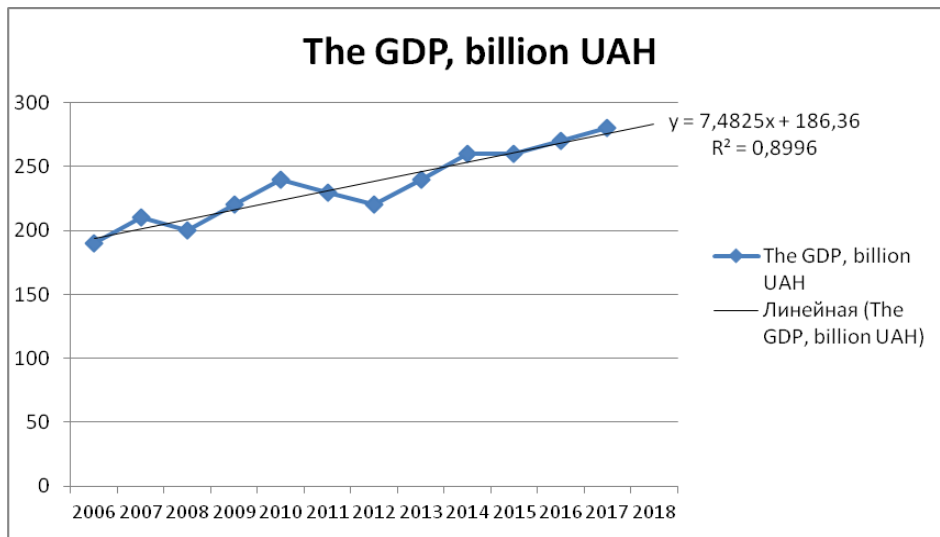


Fig. 7.8. The result of adding the *Trend Line*

Thus, the equation of the straight line has the form:

$$Y = 186.36 + 7.44825 \times t.$$

To calculate the GDP in 2017, you can:

1. Use the point forecast. Use the equation of the straight line, where  $t$  is a conditional indicator of the time at which the forecast will be constructed. That is, the volume of the GDP in 2017 is expected at the level:

$$Y = 186.36 + 7.44825 \times t;$$

$$t = 13;$$

$$Y = 283.63 \text{ billion UAH.}$$

2. Use the standard *ПРЕДСКАЗ (FORECAST)* function. After selecting this function, a dialog opens in which we set the parameters (Fig. 7.9, 7.10).

	A	B	C
	Years	The GDP, billion UAH	
1			
2	2006	190	
3	2007	210	
4	2008	200	
5	2009	220	
6	2010	240	
7	2011	230	
8	2012	220	
9	2013	240	
10	2014	260	
11	2015	260	
12	2016	270	
13	2017	280	
14	2018	=пр	
15			
16			
17			
18			
19			

Fig. 7.9. Choosing the function *ПРЕДСКАЗ*

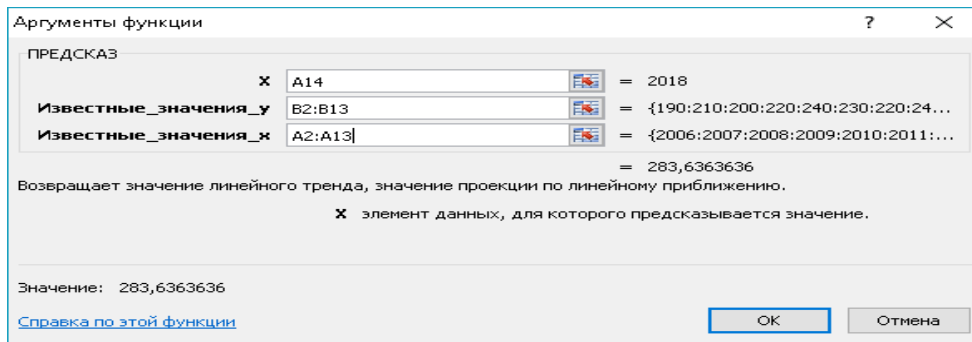


Fig. 7.10. Choosing the parameters of the function **ПРЕДСКАЗ**

The results of the predicted value are shown in Fig. 7.11.

B14		=ПРЕДСКАЗ(A14;B2:B13;A2:A13)				
	A	B	C	D	E	F
1	Years	The GDP, billion UAH				
2	2006	190				
3	2007	210				
4	2008	200				
5	2009	220				
6	2010	240				
7	2011	230				
8	2012	220				
9	2013	240				
10	2014	260				
11	2015	260				
12	2016	270				
13	2017	280				
14	2018	283,6363636				
15						

Fig. 7.11. The window of the results

3. Use the standard **ТЕНДЕНЦИЯ** (TREND) function. After selecting this function, a dialog opens, in which we set the parameters (Fig. 7.12, 7.13).

	A	B
	Years	The GDP, billion UAH
1		
2	2006	190
3	2007	210
4	2008	200
5	2009	220
6	2010	240
7	2011	230
8	2012	220
9	2013	240
10	2014	260
11	2015	260
12	2016	270
13	2017	280
14	2018	=ТЕ
15		ТЕКСТ
16		ТЕНДЕНЦИЯ
17		
18		

Fig. 7.12. Choosing the function **ТЕНДЕНЦИЯ**

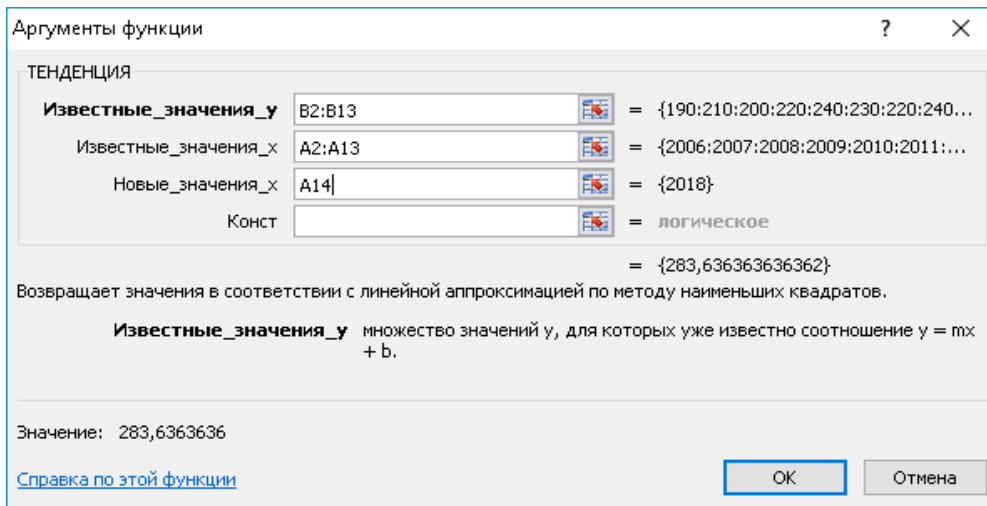


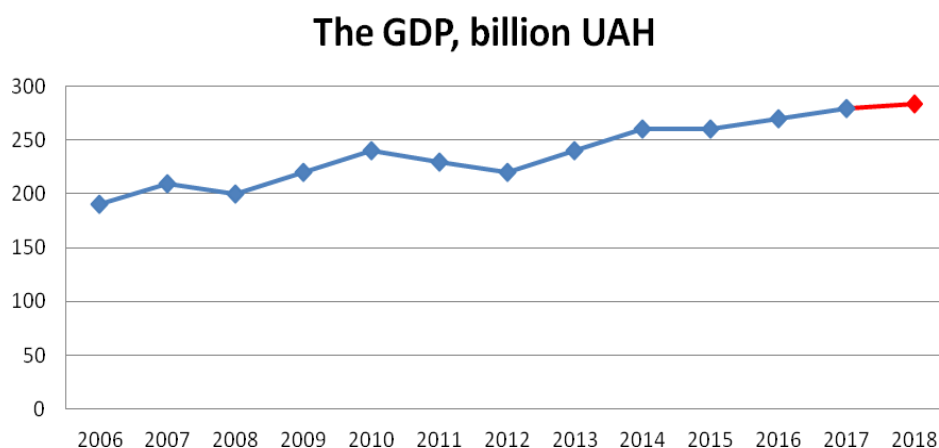
Fig. 7.13. Choosing the parameters of the function **ТЕНДЕНЦИЯ**

The results of the predicted value using the *TREND* function are shown in Fig. 7.14.

		=ТЕНДЕНЦИЯ(B2:B13;A2:A13;A14)				
	A	B	C	D	E	F
	Years	The GDP, billion UAH				
1						
2	2006	190				
3	2007	210				
4	2008	200				
5	2009	220				
6	2010	240				
7	2011	230				
8	2012	220				
9	2013	240				
10	2014	260				
11	2015	260				
12	2016	270				
13	2017	280				
14	2018	283,6363636				
15						

Fig. 7.14. The window of the results

As you can see, the use of three different methods produced the same results to be obtained graphically in Fig. 7.15.



**Fig. 7.15. Visualization of the analysis results**

In conclusion, it is expedient to note that the GDP of this country has a positive upward trend.

### The tasks for independent work

**Task 1.** According to the selected research topic, use the time series analysis methods to analyze the past, current and future situation.

The results of the assignments should be presented in the form of a report on the laboratory work.

**Task 2.** We have some information about placing the advertising posters in the region, thousand pieces (Table 7.3)

Table 7.3

### The number of advertising posters

Years	2012	2013	2014	2015	2016	2017	2018	2019	2020
Advertising posters, thousand pieces	11.9	12.3	14.5	16.3	15.2	16.3	17.2	18.2	19.8

For the analysis of profit dynamics, calculate:

1. For 2012 – 2020: the absolute increase, the growth rate and the rates of increase (using the chain and the base methods of calculation); the absolute

values of one percent of the increase. Present the received parameters in the form of a table.

2. The average placing of advertising posters in the region and the average absolute increase of placing the advertising posters in the region in the analyzed period.

3. The average annual growth rates and the rates of the increase of placing the advertising posters in:

a) 2000 – 2012 (provided that in 2000 the region placed 18.3 thousand pieces of advertising posters);

b) 2012 – 2020.

Draw an analytical conclusion based on the results of the calculations.

**Task 3.** The profit of a commercial bank is characterized by the following data, thousand UAH (Table 7.4).

Table 7.4

### The dynamics of the commercial bank profit

Year	Profit, thousand UAH
2014	253
2016	273
2017	303
2018	314
2019	294
2020	344

For the analysis of the profit dynamics, calculate:

1. For 2016 – 2020: the absolute increase, the rate of growth and the rate of increase (on the year-to-year basis and to the basic 2016); the absolute values of one percent of increase. Present the received parameters in the form of a table. Check up the interrelation between the chain and basic rates of growth.

2. The average annual growth rates and the rates of increase of profit in: 2014 – 2016; 2016 – 2020; 2014 – 2020.

3. The average level of profit in 2014 – 2020 and the average absolute increase of profit in this period.

Draw an analytical conclusion based on the results of the calculations.



**Task 4.** Using the interrelation of the characteristics of dynamics, determine the volumes of production of goods, the absolute and relative rate of reduction of volumes (Table 7.5).

Table 7.5

**The data for the characteristic of dynamics**

Year	Production of goods	Basic characteristics of dynamics		
		absolute increase, thousand UAH	rates of growth, %	rates of increase, %
1	600	–	–	–
2				-2
3		-28		
4			97	
5				-6

**Task 5.** The dynamics of production of goods of the light industry in the region is characterized by the following data (Table 7.6).

Table 7.6

**The dynamics of production of goods**

Kind of goods	Production in 2017	Annual absolute change of production	
		2018	2019
Fabrics, one million m <sup>2</sup>	240	28	20
Knitted products, million pieces	150	15	-2
Footwear, million pairs	80	12	8

For each kind of goods, determine:  
the mean absolute gain of production in 2017 – 2018;  
the level of production in 2018;  
the rate of the gain of production in 2019 in comparison with 2017;  
the mean rate of the gain for 2017 – 2019.

### **The list of questions for independent work**

1. What are the rows of dynamics and what is their role in statistical analysis?
2. What are the elements of a series?
3. What types of time series do you know? Give examples.
4. What conditions should be observed when constructing a series of dynamics?
5. What kind of averages are used in calculating the average level of moment and interval time series?
6. What are the basic and chain methods for calculating the dynamics?
7. Name the dynamics indicators and tell how they are calculated.
8. How is the average growth rate calculated in terms of dynamics?
9. What indicators of time series should be used to calculate the sales of the company for the period from 2010 to 2020?
10. What is the linear trend used for?

## **Topic 8. The index method**

### **Laboratory work 8. Analyzing the data by the index method using MS Excel**

*The purpose of the work* is to acquire the skills in statistical data analysis using the index method.

*The task* is to use *MS Excel* to analyze different types of indices and to provide an economic interpretation of the calculated indicators.

#### ***The guidelines***

**Index** is a relative value that, based on the systematic nature of the phenomenon, characterizes its change in time and space, or the degree of deviation from a certain standard (norm).

The scorecard used to build the indices has the following conventions:

**q** – the volume of production in natural units (physical volume of sales);

**p** – the price per unit;

**Q** or (**qp**) – the cost of production (turnover);

**z** – the unit cost;

**zq** – the total cost of production;

**w** – labor productivity;

**f** – average wage;

**T** – the number of employees (or hours worked);

**t** – the time spent per unit of output.

Each index includes data for two periods: **reporting (1)** – current (compared) and **base (0)** – used as the benchmark.

Table 8.1

### The formulas for calculating different types of indices

Name of indicator	Formula
Individual index of actual volume	$i_q = \frac{q_1}{q_0}$
Individual index of price	$i_p = \frac{p_1}{p_0}$
General cost index	$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$
Total index of actual volume	$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \text{ or } I_q = \frac{\sum i_q p_0 q_0}{\sum p_0 q_0}$
General price index	$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} \text{ or } I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1}{i_p} \times q_1}$
Index of variable composition	$I_{vc} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}$
Index of constant composition	$I_{cc} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1}$
Index of proportion shifts	$I_{ps} = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}$
The interconnection of indices	$I_{pq} = I_q \times I_p$ $I_{vc} = I_{cc} \times I_{ps}$

Let's consider some examples.

**Task 1.** We have some data on the sales of products (Table 8.2).

Table 8.2

**The data on the sales of products**

Production	The base period		The reporting period	
	The quantity of products, items	Price per unit, UAH	The quantity of products, items	Price per unit, UAH
A	350	200	420	225
B	170	120	110	150

On the basis of these data, it is necessary to determine: the total turnover index, the overall price index, the general index of the physical volume of trade, the absolute amount of savings or cost overruns on the price changes. Draw a conclusion.

To determine the overall index of turnover, enter the formula " $= D3 \times E3$ " into the cell *F3* and apply the formula to the whole column, calculate the amount of the column in the cell *F5*, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or by pressing the "=" button and selecting the *SUM* formula in the formula row, then enter the formula " $= B3 \times C3$ " into the cell *G3* and apply the formula to the whole column and calculate the amount in the cell *G5* (Fig. 8.1).

	A	B	C	D	E	F	G	H
1		The base period		The reporting period				
	Production	The quantity of products, items	Price per unit, UAH	The quantity of products, items	Price per unit, UAH	$q_1 \times p_1$	$q_0 \times p_0$	$I_{pq}$
2								
3	A	350	200	420	225	94500	70000	
4	B	170	120	110	150	16500	20400	
5						111000	90400	1,23

**Fig. 8.1. The calculation of the total turnover index**

To obtain the final result, enter the formula " $= F5/G5$ " into the cell *H5*. The overall turnover index is equal to 1.23, or 123 %, that is, the turnover in the current period increased by 23 % compared to the reference period due to the changes in product prices, and also due to the changes in the number of products sold.

To determine the general price index, enter the formula " $= D10 \times E10$ " into the cell *F10* and apply the formula to the whole column, calculate the amount of the column in the cell *F12*, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or by pressing the "=" button and selecting the *SUM* formula in the formula row, enter the formula " $= D10 \times C10$ " into the cell *G10* and apply the formula to the whole column, calculate the sum in the cell *G12*.

To get the value of the general price index, enter the formula " $= F12/G12$ " into the cell *H12* (Fig. 8.2).

	A	B	C	D	E	F	G	H
8		The base period		The reporting period				
	Production	The quantity of products, items	Price per unit, UAH	The quantity of products, items	Price per unit, UAH	$q_i \times p_i$	$q_i \times p_o$	$I_p$
9								
10	A	350	200	420	225	94500	84000	
11	B	170	120	110	150	16500	13200	
12						111000	97200	1,14

**Fig. 8.2. The calculation of the general price index**

The overall price index is equal to 1.14, or 114 %, that is, the turnover increased by 14 % in the current period compared to the turnover in the base period, due to the changes in the product prices at a constant output.

The absolute amount of the cost overrun or savings on the price changes is calculated as the difference between the numerator and denominator of the price index using the formula " $= F12-G12$ " in the cell *H12* (Fig. 8.3).

	A	B	C	D	E	F	G	H	I
8		The base period		The reporting period					
	Production	The quantity of products, items	Price per unit, UAH	The quantity of products, items	Price per unit, UAH	$q_i \times p_i$	$q_i \times p_o$	$I_p$	$\Delta p$
9									
10	A	350	200	420	225	94500	84000		
11	B	170	120	110	150	16500	13200		
12						111000	97200	1,14	13800

**Fig. 8.3. The calculation of the absolute amount of the cost overrun or savings due to the price changes**

The absolute amount of the overrun on the price changes in the reporting period compared with the base one was 13.800 UAH.

To determine the total volume index of the turnover, enter the formula " $= D17 \times C17$ " into the cell *F17* and apply the formula to the whole column, calculate the amount of the column in the cell *F19* either by clicking the sign

$\Sigma$  *AutoSum* button on the toolbar, or by pressing the "=" button and selecting the *SUM* formula in the formula bar, enter the formula " $= B17 \times C17$ " into the cell *G17* and apply the formula to the whole column, then, in the cell *G19*, calculate the sum. To get the value of the overall index of the physical volume of trade, enter the formula " $= F19/G19$ " into the cell *H19* (Fig. 8.4).

The overall index of the physical volume of trade is equal to 1.08, or 108 %, that is, the turnover increased by 8 % in the current period, compared with the turnover in the base period, due to the changes in the physical volume of sales at a fixed price.

	A	B	C	D	E	F	G	H
15		The base period		The reporting period				
	Production	The quantity of products, items	Price per unit, UAH	The quantity of products, items	Price per unit, UAH	$q_1 \times p_0$	$q_0 \times p_0$	$I_q$
16								
17	A	350	200	420	225	84000	70000	
18	B	170	120	110	150	13200	20400	
19						97200	90400	1,08

**Fig. 8.4. The calculation of the general index of the physical volume of the commodity turnover**

**Task 2.** We have the data on the cost of production and changes in the unit cost of product (Table 8.3).

Table 8.3

**The characteristics of the production in view of the product range**

Product name	The cost of production in the 2nd quarter, thousand UAH	Changing the unit cost of the product in the 2nd quarter compared to the 1st quarter, %
A	1850	+18
B	670	+11
C	755	No change

Identify the common indices of: 1) the prime cost; 2) the volume of production; and 3) the cost of production, provided that the cost of production in the 2nd quarter compared to the 1st quarter increased by 27 %.

In the cell *D24* the prime cost index is calculated individually using the formula " $= (C24 + 100)/100$ ", in the second column (the cost of production

in the 2nd quarter, thousand UAH) the amount is calculated by clicking the  $\Sigma$  *AutoSum* sign on the toolbar after selecting the entire column. Enter the formula " $= B24/D24$ " into the cell *E24*, calculate the amount of the column in the cell *E27*, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or by pressing the "=" button and selecting the *SUM* formula in the formula bar. To get the value of the total cost of the index, enter the formula " $= B27/E27$ " into the cell *F27* (Fig. 8.5).

	A	B	C	D	E	F
23	Product name	The cost of production in the 2nd quarter, thousand UAH	Changing the unit cost of the product in the 2nd quarter compared to the 1st quarter, %	$i_z$	$\frac{z_1 \times q_1}{i_z}$	$I_z$
24	A	1850	18	1,18	1567,80	
25	B	670	11	1,11	603,60	
26	C	755	No change	1	755,00	
27		3275			2926,40	1,12

**Fig. 8.5. The calculation of the total cost index**

The total prime cost index is equal to 1.12, or 112 %, i.e. the total cost of production increased by 12 % in the reporting period compared to the overall cost base, due to increased costs at a constant output.

Based on the conditions of the task, the overall cost index ( $I_{zq}$ ) is 1.27 or 127 %, as the cost of production in the second quarter compared with the first one increased by 27 % due to the changes in the cost and volume of output. For the calculation of the overall index of production using the relationship between the indexes, enter the formula " $= G27/F27$ " into the cell *H27* (Fig. 8.6).

	A	B	C	D	E	F	G	H
23	Product name	The cost of production in the 2nd quarter, thousand UAH	Changing the unit cost of the product in the 2nd quarter compared to the 1st quarter, %	$i_z$	$\frac{z_1 \times q_1}{i_z}$	$I_z$	$I_{zq}$	$I_q$
24	A	1850	18	1,18	1567,80			
25	B	670	11	1,11	603,60			
26	C	755	No change	1	755,00			
27		3275			2926,40	1,12	1,27	1,13

**Fig. 8.6. The calculation of the overall index of production**

The total production expenses increased by 13 % due to the changes in the volume of production at a fixed cost.

**Task 3.** We have some data on the manufacture of products of the same name "A" and its cost at two plants (Table 8.4).

Table 8.4

**The data on the production and the production prime cost**

Plant	Production, thousand units		The unit cost of production, UAH	
	years		years	
	2017	2018	2017	2018
1	90	100	210	205
2	180	110	180	240

Determine the prime cost indices: the constant composition, the variable composition, the proportion shift. Explain the difference between these indices.

To determine the cost index of constant composition, enter the formula " $= E34 \times C34$ " into the cell *F34* and apply the formula to the whole column, calculate the amount of the column in the cell *F36* either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or by pressing the "=" button and selecting the *SUM* formula in the formula bar, enter the formula " $= F36 / C36$ " into the cell *G36* to get the value of the first fraction, enter the formula " $= D34 \times C34$ " into the cell *H34* and apply the formula to the whole column, count the amount in the column in the cell *H36*, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or pressing the "=" button and selecting the *SUM* formula in the formula bar, enter the formula " $= H36 / C36$ " into the cell *I36* to get the value of the second fraction.

To get the prime cost index of constant composition, enter the formula " $= G36 / I36$ " into the cell *J36* (Fig. 8.7).

	A	B	C	D	E	F	G	H	I	J
31	Plant	Production, thousand units		The unit cost of production, UAH		$z_1 \times q_1$	$\frac{\sum z_1 \times q_1}{\sum q_1}$	$z_0 \times q_1$	$\frac{\sum z_0 \times q_1}{\sum q_1}$	$I_{cc}$
32		years		years						
33		2017	2018	2017	2018					
34	1	90	100	210	205	20500		21000		
35	2	180	110	180	240	26400		19800		
36		270	210			46900	223,33	40800	194,29	1,15

**Fig. 8.7. The calculation of the cost index of constant composition**



The prime cost index of constant composition was 1.15 (115 %), the average unit cost of production increased by 15 % due to the changes in the cost.

To determine the cost index of variable composition, enter the formula " $= E42 \times C42$ " into the cell *F42* and apply the formula to the whole column, calculate the amount of the column in the cell *F44*, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or by pressing the "=" button and selecting the *SUM* formula in the formula bar, enter the formula " $= F44 / C44$ " into the cell *G44* and get the value of the first fraction, enter the formula " $= B42 \times D42$ " into the cell *H42* and apply the formula to the whole column, calculate the amount in the column in the cell *H44*, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or pressing the "=" button and selecting the *SUM* formula in the formula bar, enter the formula " $= H44 / B44$ " into the cell *I44* and get the value of the second fraction.

To get the cost index of variable composition, enter the formula " $= G44 / I44$ " into the cell *J44* (Fig. 8.8).

	A	B	C	D	E	F	G	H	I	J
39	Plant	Production, thousand units		The unit cost of production, UAH		$Z_1 \times q_1$	$\frac{\sum Z_1 \times q_1}{\sum q_1}$	$Z_0 \times q_0$	$\frac{\sum Z_0 \times q_0}{\sum q_0}$	$I_{vc}$
40		years		years						
41		2017	2018	2017	2018					
42	1	90	100	210	205	20500		18900		
43	2	180	110	180	240	26400		32400		
44		270	210			46900	223,33	51300	190,00	1,18

Fig. 8.8. The calculation of the cost index of variable composition

The cost index of variable composition was 1.18 (118 %), that is, the average unit cost of production increased by 18 % due to the changes in the cost and the quantity of the output.

To determine the index of the cost of structural changes, enter the formula " $= D50 \times C50$ " into the cell *F50* and apply the formula to the whole column, calculate the amount of the column in the cell *F52*, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or by pressing the "=" button and selecting the *SUM* formula in the formula bar, enter the formula " $= F52 \times C52$ " into the cell *G52* to obtain the value of the first fraction, enter the formula " $= B50 \times D50$ " into the cell *H50* and apply the formula to the whole column, calculate the amount in the column in the cell *H52*, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or pressing the "=" button and selecting the *SUM* formula in the formula bar, enter the formula " $= H52 / B52$ " into the cell

/I52 and get the value of the second fraction. To get the index of the cost of structural changes, enter the formula = G52/I52 into the cell J52 (Fig. 8.9).

	A	B	C	D	E	F	G	H	I	J
47	Plant	Production, thousand units		The unit cost of production, UAH		$Z_0 \times q_1$	$\frac{\sum Z_0 \times q_1}{\sum q_1}$	$Z_0 \times q_0$	$\frac{\sum Z_0 \times q_0}{\sum q_0}$	$I_{ps}$
48		years		years						
49		2017	2018	2017	2018					
50	1	90	100	210	205	21000		18900		
51	2	180	110	180	240	19800		32400		
52		270	210			40800	194,29	51300	190,00	1,02

Fig. 8.9. The calculation of the cost index of the proportion shift

The prime cost index of the proportion shift was 1.02 (102 %), that is, the average cost of production increased by 2 % due to the changes in the number of output at constant unit costs.

### The tasks for independent work

**Task 1.** According to the selected statistical information, use all types of indices and make index analysis.

The results of the assignments should be presented in the form of a report on the laboratory work.

**Task 2.** We have the data about the sale of goods (Table 8.5).

Table 8.5

### The input information

Product A	Base period		Current period	
	Number of products, units	Price per unit, UAH	Number of products, units	Price per unit, UAH
Enterprise 1	200	120	185	150
Enterprise 2	150	80	120	100

Calculate: the *individual indices*: 1) for price and for volume; the *aggregative indices*: 2) the general cost index; 3) the general price index; 4) the total index of the actual volume; 5) determine the absolute amount of savings or cost overruns on the price change; the *average level dynamics*

*indices*: 6) the index of variable composition; 7) the index of constant composition; 8) the index of the shift of proportions. Draw conclusions.

**Task 3.** There are some data about the cost of the exported products and the change in the prices of the given products in the current period in comparison with the base period (Table 8.6).

Table 8.6

**The dynamics of the cost indices according to the kinds of the exported products**

Kind of product	Export of products in the current period, thousand US dollars	Change in the unit price of the exported products in the current year in comparison with the base year, %
A	450	-4.2
Б	487	-7.4
B	325	+3.1

Based on the index analysis, determine how the export has changed in relative and absolute expression due to changes in prices if it is known, that the export of products has decreased by 15 %. How has the export changed due to changes in the actual volume of the exported products? Interpret the received parameters.

**Task 4.** We have the following data about the labor productivity of workers in different trades (Table 8.7).

Table 8.7

**The distribution of workers based on the labor productivity level**

The trade	The volume of the given services in the period, thousand UAH		The general expenditure of time on granting of services in the period, hours	
	the basic period	the current period	the basic period	the current period
1. The hairdresser	15.1	15.5	1 069	1 095
2. The repairman of home appliances	4.8	5.0	340	350
3. The repairman of footwear	11.7	12.1	670	710

Based on the averages of the index analysis, determine, how the labor productivity in each trade and the average labor productivity in all trades has changed. Draw conclusions.

**Task 5.** The general costs of the enterprise production have increased in the current period in comparison with the basic period by 8 %, the volume of production has increased by 11.5 %. Determine, how the cost price of a production unit has changed in the current period in comparison with the base period.

### **The list of questions for independent work**

1. What is the index in statistics?
2. What is the role of the index method in the economic research?
3. What tasks are solved using the indices?
4. What types of indices are there?
5. What indices are called individual, and what are the general ones?

What do they characterize?

6. Name the types of indices: the degree of the coverage of the elements of the population, the basis of comparison, the method of calculation, the nature of the indexed quantities, the weights, and the co-factors.

7. Expand on the content, types and methods of calculating the indices of the physical volume, prices, labor productivity and the cost of production.

8. What is the relationship between the chain and base indices?

9. Provide some examples of interconnected indices.

10. What characterizes the indices of structural shifts and how are they calculated?

## References

### Main

1. Горкавий В. К. Статистика : навч. посіб. / В. К. Горкавий. – Київ : Алерта, 2012. – 608 с.
2. Двігун А. О. Статистика : навчальний посібник для студентів вищих навчальних закладів / А. О. Двігун, П. А. Борисенко, К. І. Дерев'янку. – Запоріжжя : Акцент Інвесттрейд, 2012. – 307 с.
3. Мармоза А. Т. Теорія статистики : підручник / А. Т. Мармоза. – Київ : Центр учбової літератури, 2013. – 592 с.
4. Сєрова І. А. Організація статистичних спостережень. Конспект лекцій для студентів спеціальності 8.050110 усіх форм навчання / І. А. Сєрова, І. В. Аксьонова. – Харків : Вид-во ХНЕУ, 2008. – 236 с.
5. Статистика (модульний варіант з програмованою формою контролю знань) : навчальний посібник / А. Т. Опря, Л. О. Дорогань-Писаренко, О. В. Єгорова, Ж. А. Кононенко. – (2-ге вид., перероб. і допов.). – Київ : Центр учбової літератури, 2017. – 536 с.
6. Статистика : навчальний посібник / за ред. О. В. Раєвської. – Харків : ВД "Інжек", 2011. – 504 с.

### Additional

7. Вуколов Э. А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL : учебное пособие. – Москва : Форум, 2010. – 464 с.
8. Макарова Н. В. Статистика в Excel : учебное пособие / Н. В. Макарова, В. Я. Трофимец. – Москва : Финансы и статистика, 2002. – 368 с.
9. Мармоза А. Т. Практикум з теорії статистики : навчальний посібник. – 3-е вид., випр. – Київ : Ельга, Ніка–Центр, 2007. – 348 с.
10. Сигел Э. Ф. Практическая бизнес-статистика / Э. Ф. Сигел : пер. с англ. – Москва : Издательский дом "Вильямс", 2008. – 1056 с.
11. Статистика: теоретичні засади і прикладні аспекти / Р. В. Фещур, А. Ф. Барвінський, В. П. Качур та ін. ; за заг. ред. Р. В. Фещура. – Львів : Інтеллект-Захід, 2003. – 576 с.

12. Ткач Є. І. Загальна теорія статистики : підручник [для студ. вищ. навч. закл.] / Є. І. Ткач, В. П. Сторожук. – [3-тє вид.] – Київ : Центр учбової літератури, 2017. – 442 с.

13. Тринько Р. І. Основи теоретичної і прикладної статистики : [навчальний посібник] / Р. І. Тринько, М. Є. Стадник. – Київ : Знання, 2011. – 400 с.

14. Few S. Show Me the Numbers: Designing Tables and Graphs to Enlighten / S. Few. – Oakland, California : Analytics Press, 2004. – 263 p.

15. Keller G. Statistics for Management and Economics / G. Keller, B. Warrack. – Fifth edition. – London : Duxbury, 2000. – 1000 p.

16. Leboucher L. Introduction à la statistique descriptive / L. Leboucher, M. Voisin. – Toulouse : Cépaduès, 2013. – 702 p.

### **Information resources**

17. The official site of the State Statistics Service of Ukraine. – Access mode : [www.ukrstat.gov.ua](http://www.ukrstat.gov.ua).

18. The only state web portal of open source of data. – Access mode : <https://data.gov.ua>.

### **Methodological support**

19. Sierova I. Statistics : summary of lectures for full-time students of training directions 6.140103 "Tourism", 6.030601 "Management" specialization "Business Administration" / I. Sierova. – Kharkiv : S. Kuznets KhNUE, 2014. – 100 p.

20. Statistics [Electronic resource] : syllabus for Bachelor's (first) degree students of all specialities / compiled by I. Sierova, V. Derykhovska. – Kharkiv : S. Kuznets KhNUE, 2018. – 13 p.

21. Tasks for self-instruction on the academic discipline "Statistics" for students of training directions 6.030601 "Management" of specialization "Business Administration", 6.140103 "Tourism" of all forms of study / compiled by I. Sierova, O. Zirko. – Kharkiv : S. Kuznets KhNUE, 2015. – 28 p. (English)

# Content

Introduction.....	3
Content module 1. The methods of descriptive statistics .....	6
Topic 1. The methodological principles of statistics .....	6
Laboratory work 1. An overview of Excel .....	6
The tasks for independent work .....	14
The list of questions for independent work .....	14
Topic 2. Statistical observation .....	15
Laboratory work 2. Submission of statistical data: graphs .....	15
The tasks for independent work .....	22
The list of questions for independent work .....	23
Topic 3. Summarization and grouping of statistical data .....	23
Laboratory work 3. Summarization and grouping of statistical data.....	23
The tasks for independent work .....	27
The list of questions for independent work .....	31
Topic 4. Generalizing the statistical indicators .....	31
Laboratory work 4. Analysis of data using the generalizing statistical indicators .....	31
The tasks for independent work .....	41
The list of questions for independent work .....	43
Topic 5. Analysis of the distribution series .....	44
Laboratory work 5. Analysis of the statistical distribution series ....	44
The tasks for independent work .....	57
The list of questions for independent work .....	59
Content module 2. The methods of inferential statistics .....	60
Topic 6. Sampling and sampling distributions .....	60
Laboratory work 6. Analysis of statistical data with the use of sampling and sampling distributions.....	60
The tasks for independent work .....	66
The list of questions for independent work .....	68
Topic 7. Time series analysis.....	68
Laboratory work 7. Analysis of data using the time series methods.....	68
The tasks for independent work .....	79
The list of questions for independent work .....	82

Topic 8. The index method .....	82
Laboratory work 8. Analyzing the data by the index method using MS Excel .....	82
The tasks for independent work .....	90
The list of questions for independent work .....	92
References .....	93
Main.....	93
Additional.....	93
Information resources .....	94
Methodological support.....	94



НАВЧАЛЬНЕ ВИДАННЯ

# СТАТИСТИКА

**Методичні рекомендації  
до лабораторних та самостійних робіт  
для студентів усіх спеціальностей  
першого (бакалаврського) рівня**

**(англ. мовою)**

*Самостійне електронне текстове мережеве видання*

Укладачі: **Сєрова** Ірина Анатоліївна  
**Дериховська** Вікторія Ігорівна  
**Шликова** Вікторія Олександрівна  
**Бровко** Ольга Іванівна

Відповідальний за видання *О. В. Раєвська*

Редактор *З. В. Зобова*

Коректор *З. В. Зобова*

Подано завдання для лабораторної та самостійної роботи з навчальної дисципліни та методичні рекомендації до їх виконання, що допоможуть студентам одержати теоретичні та практичні навички використання статистичних спостережень, методів статистичного аналізу та прогнозування соціальних і економічних явищ та процесів.

Рекомендовано для студентів усіх спеціальностей першого (бакалаврського) рівня.

План 2020 р. Поз. № 108 ЕВ. Обсяг 97 с.

---

Видавець і виготовлювач – ХНЕУ ім. С. Кузнеця, 61166, м. Харків, просп. Науки, 9-А

*Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру  
ДК № 4853 від 20.02.2015 р.*