

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

# **THE FUNDAMENTALS OF DATA ANALYSIS**

**Guidelines for laboratory work  
for Master's (second) degree students  
of speciality 122 "Computer Science"**

**Kharkiv  
S. Kuznets KhNUE  
2020**

UDC 004.6(07.034)

F97

**Compiled by:** O. Rayevnyeva

O. Brovko

Затверджено на засіданні кафедри статистики і економічного прогнозування.

Протокол № 6 від 05.12.2019 р.

*Самостійне електронне текстове мережеве видання*

**The Fundamentals** of Data Analysis [Electronic resource] :  
F97 guidelines for laboratory work for Master's (second) degree students  
of speciality 122 "Computer Science" / compiled by O. Rayevnyeva,  
O. Brovko. – Kharkiv : S. Kuznets KhNUE, 2020. – 55 p.

The tasks for laboratory work and guidelines to them are presented to consolidate and deepen the knowledge of the theoretical and practical material, to acquire the skills in the analysis of economic phenomena and processes with the help of MS Excel.

For Master's (second) degree students of speciality 122 "Computer Science".

**UDC 004.6(07.034)**

© Simon Kuznets Kharkiv National  
University of Economics, 2020

## Introduction

Data analysis is a science that studies important information and arrays of a large amount of data. It combines statistical methods for analyzing data using a number of IT technologies. The academic discipline offers a broad approach to data analysis and data searching methods and application of these methods to practice. It combines cutting-edge research and practice in related fields and provides students with the knowledge and ability to initiate and conduct data analysis that can turn company data into commercially valuable information.

The academic discipline "The Fundamentals of Data Analysis" belongs to the cycle of selective disciplines of the academic degree "Master" for students of speciality 122 "Computer Science" of the educational and professional program "Business Analytics and Information Systems in Entrepreneurship" and aims to form student's basic competences to analyze, objectively evaluate phenomena and processes, establish the dependence of the effective indicator on a set of factors, identify the trends of changes in the phenomena under study and predict and interpret the survey results and more.

Within the framework of laboratory workshops students are offered modern methods and technologies that are used to analyze data in various fields of activity.

Each laboratory work contains the purpose and tasks, as well as the guidelines to help students perform the work.

In order to defend the laboratory work, students have to draw up an individual report, which should contain the statement of the task, the printed results of building a model with a complete description of each stage of modeling and conclusions.

The table summarizes the professional competences that students develop when studying the academic discipline "The Fundamentals of Data Analysis".

Table

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

<b>Competence</b>	<b>Learning outcomes</b>
1	2
The ability to identify, interpret and analyze the needs of stakeholders of economic relations	The ability to determine the object, the subject of the study of socio-economic phenomena and processes

Table (the end)

1	2
	The ability to form the information base of the study
The ability to apply systemic thinking to understand complex systemic behavior, including interaction between the constituent and other systems (social, cultural, legislative, environmental, business, etc.)	The ability to calculate the average and marginal error of selective observation of socio-economic systems under different selection schemes. The ability to determine the sample size based on the use of application packages
The ability to establish the regularity of the behavior of economic objects on the basis of the formation of homogeneous groups	The ability to conduct primary and secondary grouping of statistical information. The ability to develop layouts of statistical tables. The ability to group statistics using the MS Excel add-in "Data Analysis"
The ability to identify quantitative and qualitative characteristics of the course of economic phenomena	The ability to perform calculations of average and relative indicators in the study of socio-economic systems. The ability to analyze the intensity of dynamics, identify the presence of trends in dynamic ranks and predict socio-economic phenomena based on the method of extrapolation. The ability to distinguish seasonal characteristics in the studied series of dynamics
The ability to form innovative management decisions regarding the development of economic processes and objects	The ability to construct various forms of expression of indices and assess their interconnection

# Laboratory work 1

## Formation of a representative sample

*The purpose* of the work is to master the skills in the work with the MS Excel spreadsheet in the process of forming the information space of research.

*The task* is, on the basis of the initial data presented on various official sites, to study the basic methods of work with spreadsheets: the input and editing of data, formatting, designing of tables.

### Guidelines

In the process of formation of information space of research in Excel, quite often it is necessary to carry out sampling based on a certain criterion or on several criteria. You can do this in a variety of ways using a wide range of tools, choosing from a number of options.

Formation of information space consists in the procedure of selecting, from the total array, of data that satisfy the given conditions, and then displaying them on a sheet on a separate list or in the output range.

There are several approaches to forming a research space:

Approach 1: applying an extended auto filter.

Approach 2: applying an array formula.

Approach 3: multiple criteria sampling using a formula.

Approach 4: random sampling.

Let's elaborate on each approach in more detail.

**Approach 1:** applying an extended auto filter.

The simplest way to make a selection is to use a special auto filter. Consider how to do it in a specific example.

Highlight the area on the worksheet to be sampled. In the *Home* tab (Fig. 1.1), click on the *Sort and Filter* button. It is placed in the *Edit Settings* block. In the list that opens, click on the *Filter* button.

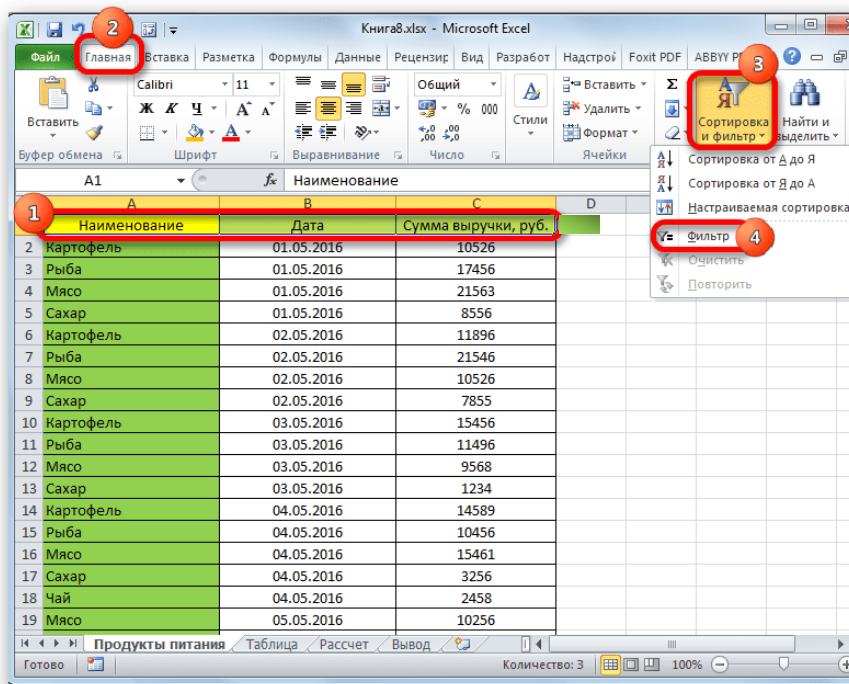


Fig. 1.1. The **Filter Options** window

There is another way to customize *Filter* settings. To do this, after highlighting the area on the worksheet, we move to the tab *Data* (Fig. 1.2). Click on the *Filter* button, which is placed on the *Ribbon* in the *Sort and Filter* group.

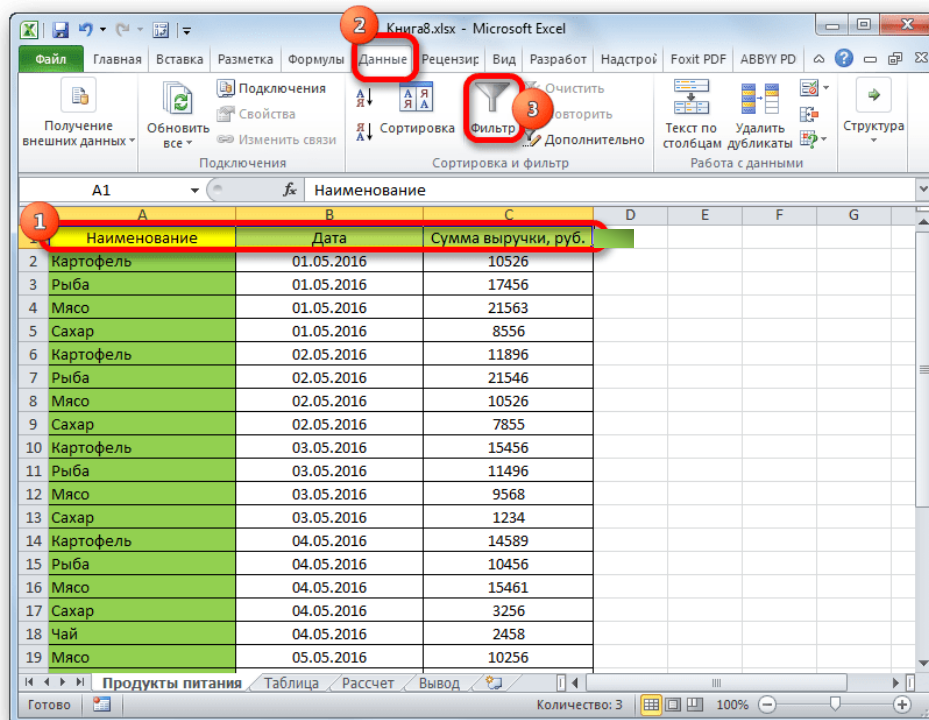


Fig. 1.2. **Data Insert**

The next step is to see the icons in the table header to start filtering in the form of inverted edges down the small triangles on the right edge of the cell. Click on this icon in the header of the column that we want to sample. In the menu that opens, go to *Text Filters*. Next, select *Custom Filter ...* (Fig. 1.3).

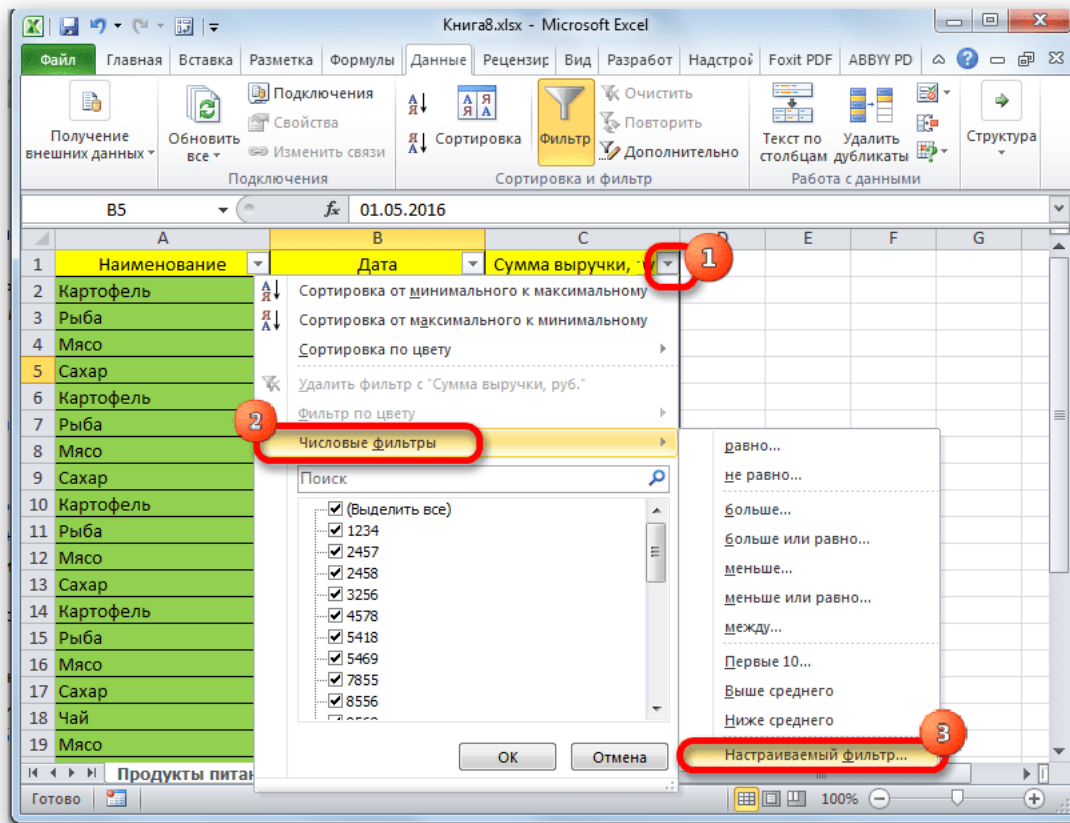


Fig. 1.3. The *Custom Filter* settings

After you have made the required settings, the *Custom Filter* window is activated. It allows you to specify a limit on which the selection will be made. The drop-down list for a column contains numeric format cells which we use as an example. You can choose from five types of conditions:

- is equal to;
- not equal;
- greater than or equal to;
- less.

For example, we set the condition to select only values for which the amount of revenue exceeds \$10 000. Set the switch to the *Greater* position. Enter the value "10 000" in the right field. Click the *OK* button to complete the action (Fig. 1.4).

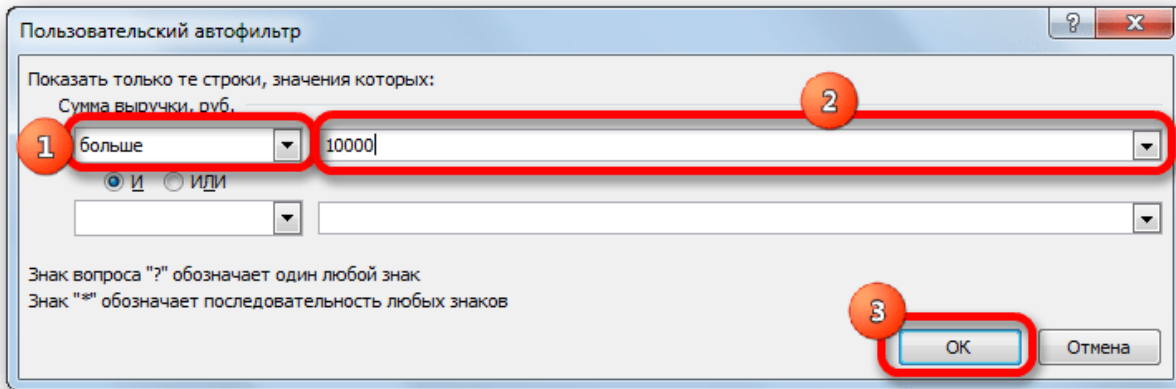


Fig. 1.4. The **Custom AutoFilter** dialog box

After filtering, there are only lines in which the revenue exceeds \$10 000. The result of data filtering is shown in Fig. 1.5.

	А	В	С	Д	Е	F	G
1	Наименование	Дата	Сумма выручки, руб.				
2	Картофель	01.05.2016	10526				
3	Рыба	01.05.2016	17456				
4	Мясо	01.05.2016	21563				
6	Картофель	02.05.2016	11896				
7	Рыба	02.05.2016	21546				
8	Мясо	02.05.2016	10526				
10	Картофель	03.05.2016	15456				
11	Рыба	03.05.2016	11496				
14	Картофель	04.05.2016	14589				
15	Рыба	04.05.2016	10456				
16	Мясо	04.05.2016	15461				
19	Мясо	05.05.2016	10256				
22	Картофель	06.05.2016	12546				
23	Рыба	06.05.2016	11784				
24	Мясо	06.05.2016	13485				
27	Картофель	07.05.2016	14256				
28	Рыба	07.05.2016	13858				
29	Мясо	07.05.2016	13978				

Fig. 1.5. The result of data filtering

But in the same column, you can add a second condition. To do this, we again return to the *Custom Filter* window. At the bottom of it there is another condition switch and the corresponding input field. It is now necessary to set



an upper selection limit of \$15 000 (Fig. 1.6). To do this, put the switch in the position "Less", and in the field to the right, enter the value "15 000".

In addition, there is also a condition switch. It has two positions – "AND" and "OR". By default, it is set in the first position. This means that only rows that satisfy both constraints will remain in the sample. If it is set to "OR", then there will remain values that fit either of the two conditions. In our case it is necessary to set the parameter "AND", that is, to leave this setting by default. After all the values have been entered, click on the OK button.

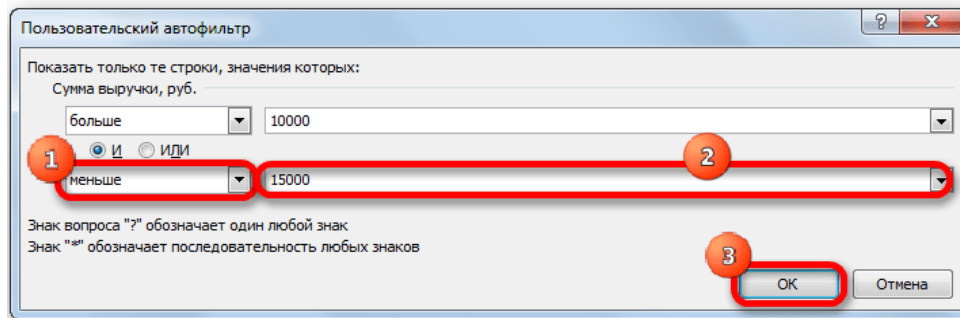


Fig. 1.6. The **Custom AutoFilter** dialog box

Now, in the table, there are only rows in which the amount of revenue is not less than 10 000 rubles, but does not exceed 15 000 rubles. The result of data filtering is shown in Fig. 1.7.

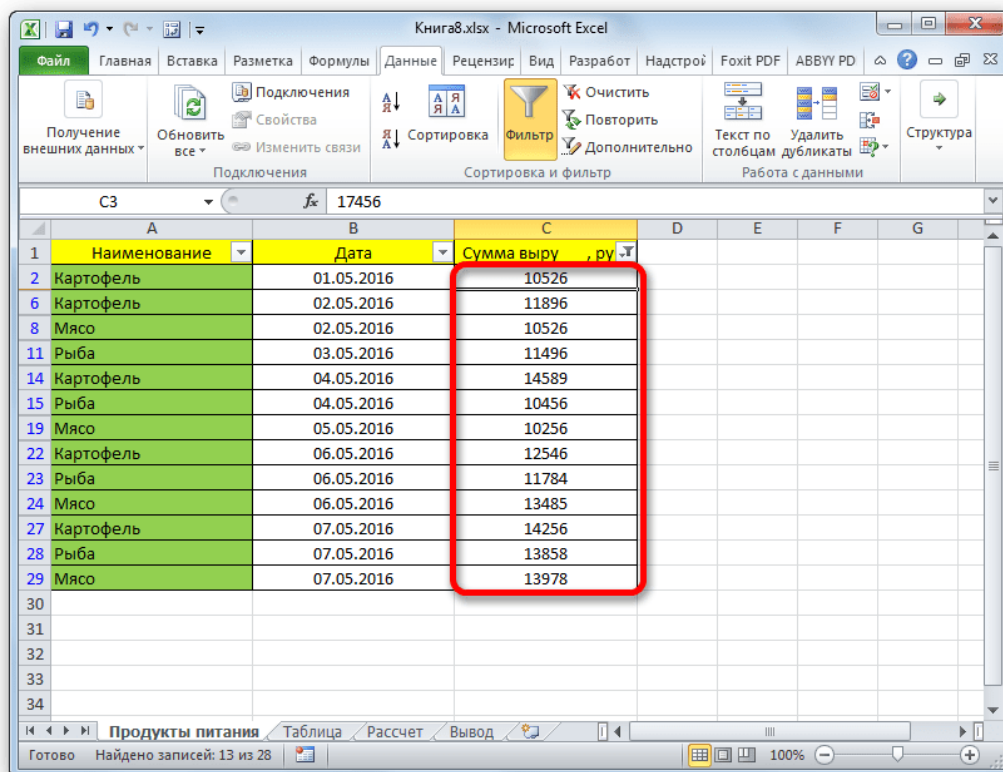


Fig. 1.7. The result of data filtering

You can also customize filters in other columns. It is also possible to save the filtering and preconditions that were specified in the columns.

Let's select the cell filter in the date format. Click on the filter icon in the corresponding column. There are multiple items in the *Filter by date* and *Custom Filter* lists (Fig. 1.8).

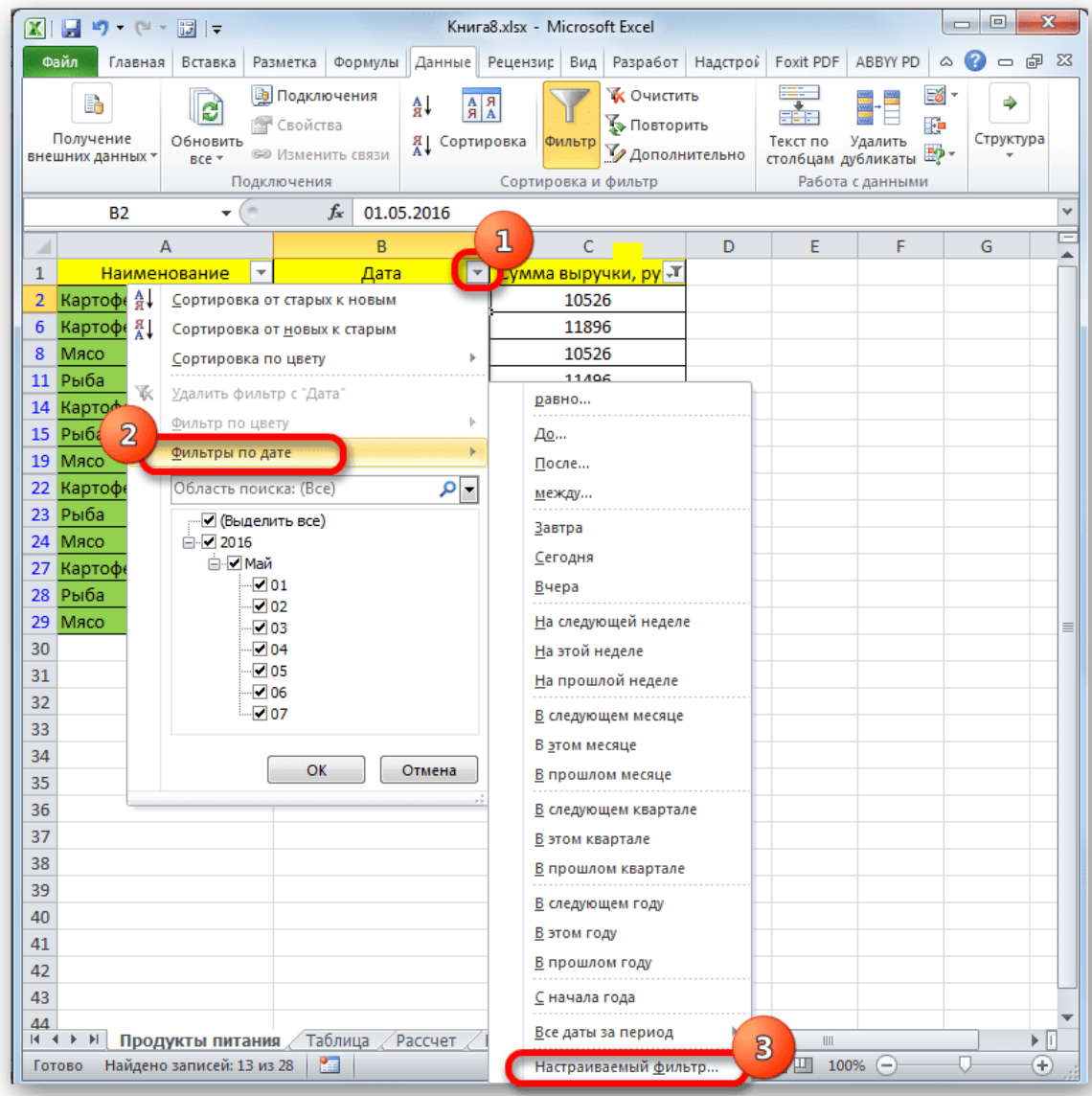


Fig. 1.8. The filter options selection window

The *Custom AutoFilter* window starts again. We will select the results in the table from May 4 to May 6, 2016 inclusive. There are even more options in the choice of conditions switch than for the numeric format. We select the position *After or equal*. In the box to the right we set the date "04.05.2016". In the lower block, set the switch to the "OR" position. In the right field enter

the date "06.05.2016" (Fig. 1.9). The condition compatibility switch is left in the default position – "AND". Click the *OK* button to filter the action.

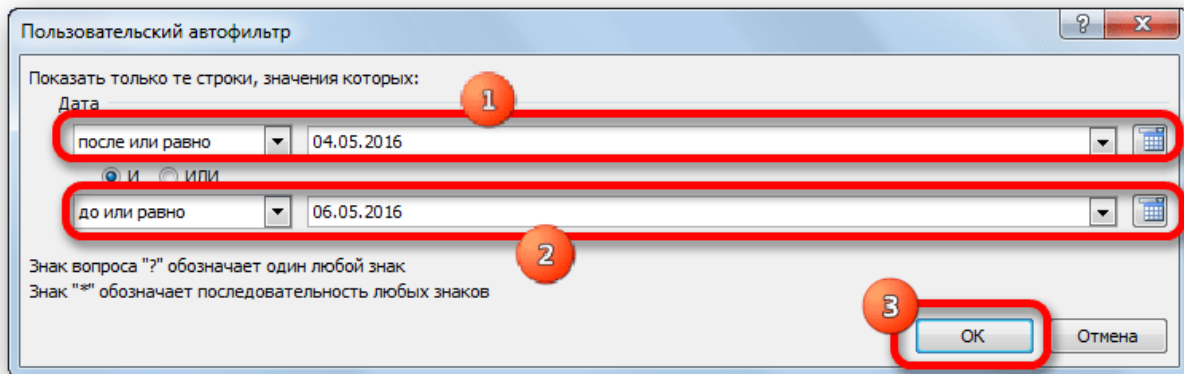


Fig. 1.9. The **Custom AutoFilter** dialog box

As a result, the list is further reduced. Now it leaves only lines in which the amount of revenue varies from 10 000 to 15 000 USD for the period from 04.05.2016 to 06.05.2016 inclusive. The result of filtering based on the given parameters is presented in Fig. 1.10.

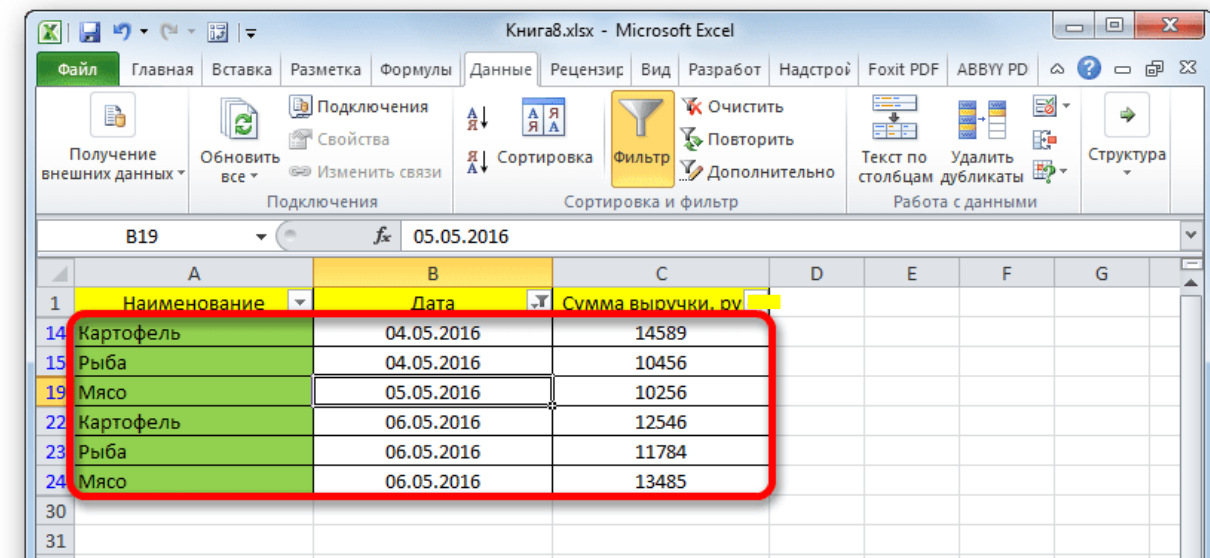


Fig. 1.10. The result of filtering based on the given parameters

When sampling, you can reset the filtering in one of the columns. For example, let's do this for revenue values. Click on the autofilter icon in the corresponding column. In the drop-down list, click on *Delete filter* (Fig. 1.11).

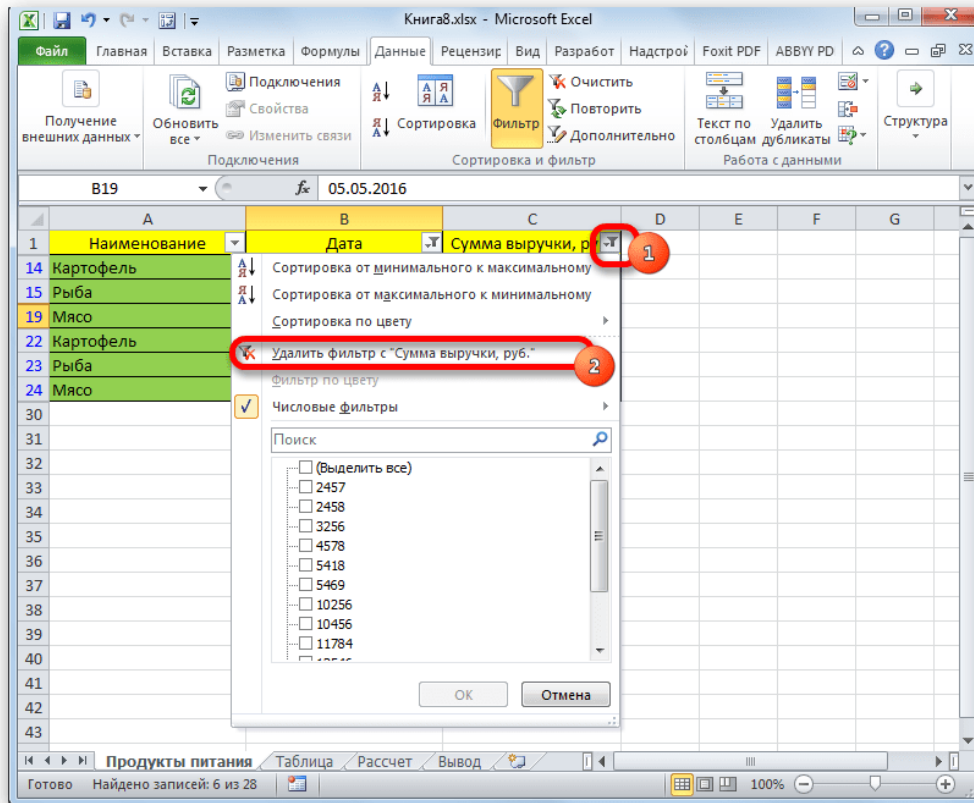


Fig. 1.11. The *Options* dialog box

After these actions, the sample based on the amount of revenue will be disabled, and only the selection in the date format will remain (from 04.05.2016 to 06.05.2016). The result of sample formation is shown in Fig. 1.12.

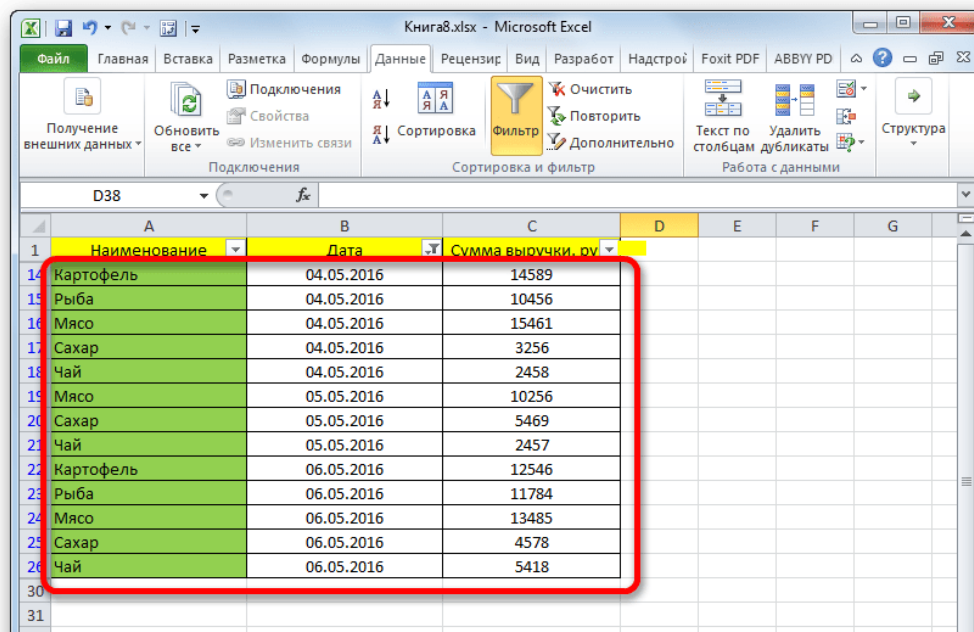


Fig. 1.12. The result of filtering based on the given parameters

This table has another column – "Name". It contains data in the text format. Let's look at how to sample by filtering based on this value.

Click on the column name filter icon. Go through the names of the list *Text Filters* and *Custom Filter ...* (Fig. 1.13).

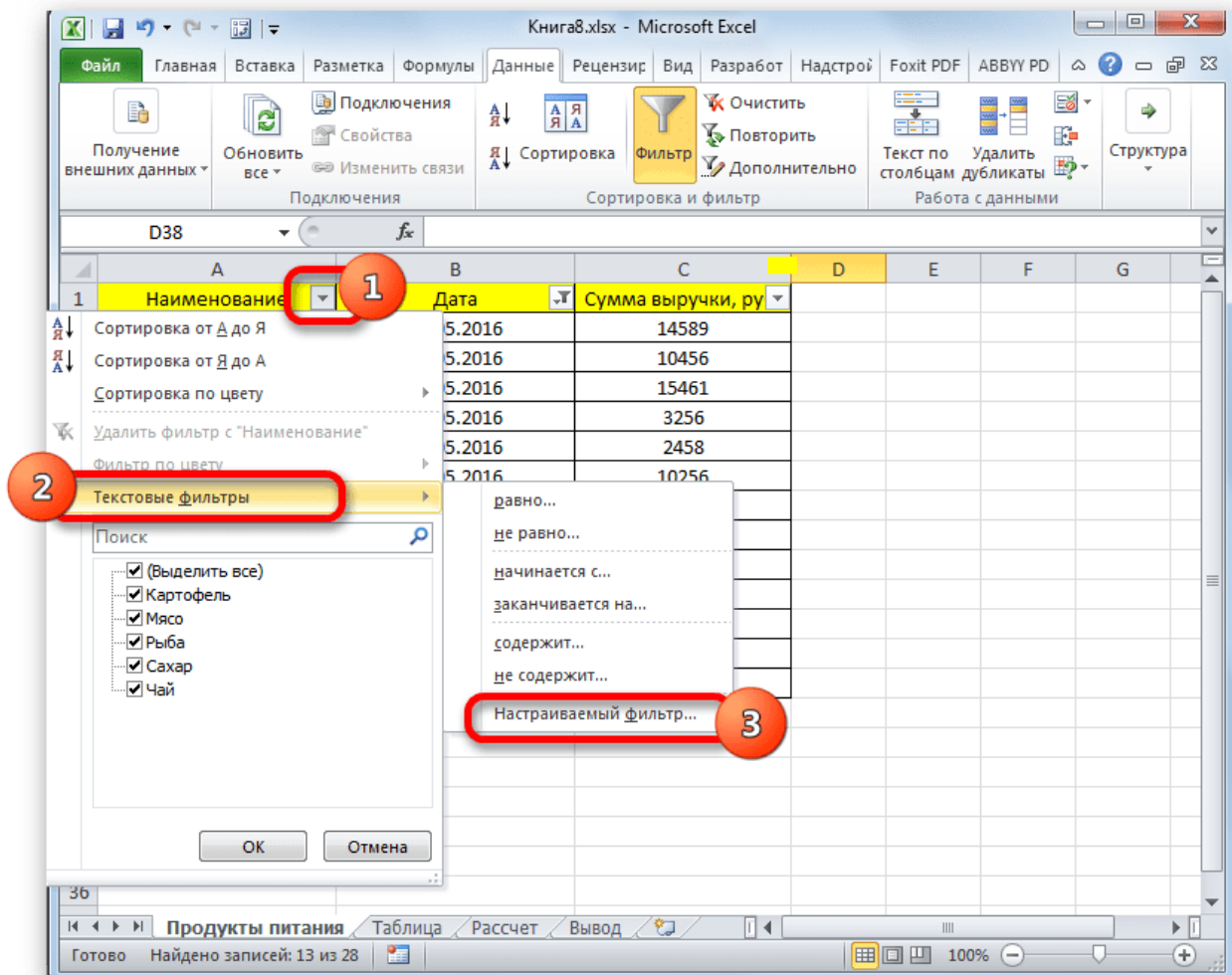


Fig. 1.13. The *Custom Filter* selection window

After selecting this filter, the *Custom Autofilter* window opens, which requires you to specify the appropriate name parameters. For example, make a sample using the names "Potatoes" and "Meat". In the first block, the condition switch is set to "". In the box to the right, enter the word "Potatoes". The lower unit switch is also set to "". In the box opposite make an entry – "Meat" (Fig. 1.14).

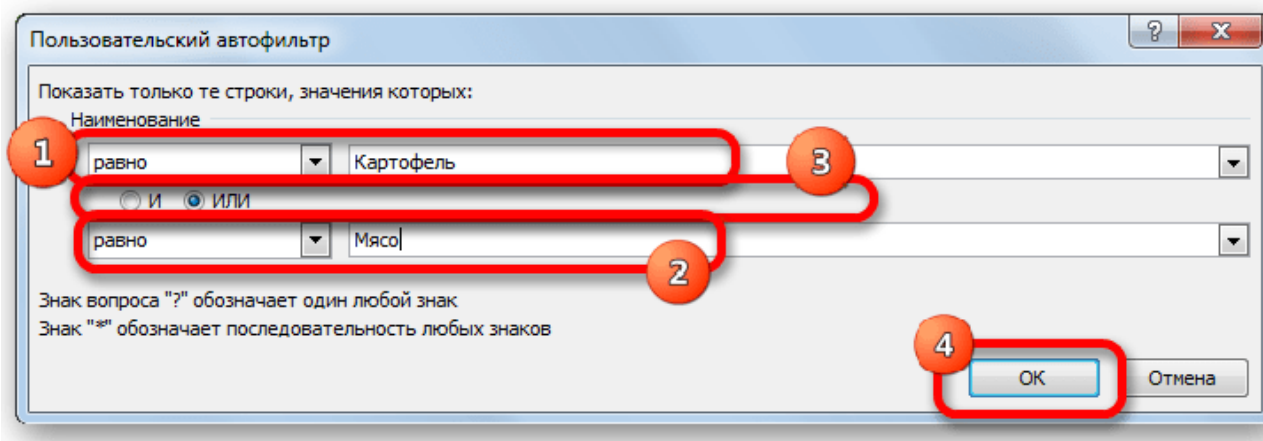


Fig. 1.14. The *Custom AutoFilter* dialog box

Then set the compatibility switch to the "OR" position. The line containing any of these conditions will now be displayed. Click on the *OK* button. The result of the generated sample is presented in Fig. 1.15.

	A	B	C
1	Наименование	Дата	Сумма выручки, руб.
14	Картофель	04.05.2016	14589
16	Мясо	04.05.2016	15461
19	Мясо	05.05.2016	10256
22	Картофель	06.05.2016	12546
24	Мясо	06.05.2016	13485
30			
31			

Fig. 1.15. The result of filtering based on the given parameters

As we can see from Fig. 1.15, in the new sample, there are restrictions on the date (from 04.05.2016 to 06.05.2016) and the name (potatoes and meat). There is no restriction on the amount of revenue.

You can completely remove the filter in the same ways that you used to install it. No matter what method was used. To reset the filter, while in the *Data* tab, click on the *Filter* button, which is located in the *Sort and Filter* section.

The second option is to go to the *Home* tab. There, click on the ribbon on the *Sort and Filter* button in the *Edit* block. In the activated list, click on the *Filter* button. Using either of the two methods above, the filtering will be removed and the results of the sampling will be cleared. That is, the table will show the entire array of data it owns.

**Approach 2:** applying an array formula.

You can also make a selection by applying a complex array formula. Unlike the previous variant, this method involves displaying the result in a separate table.

On the same sheet, we create an empty table with the same column names in the header as the original (Fig. 1.16).

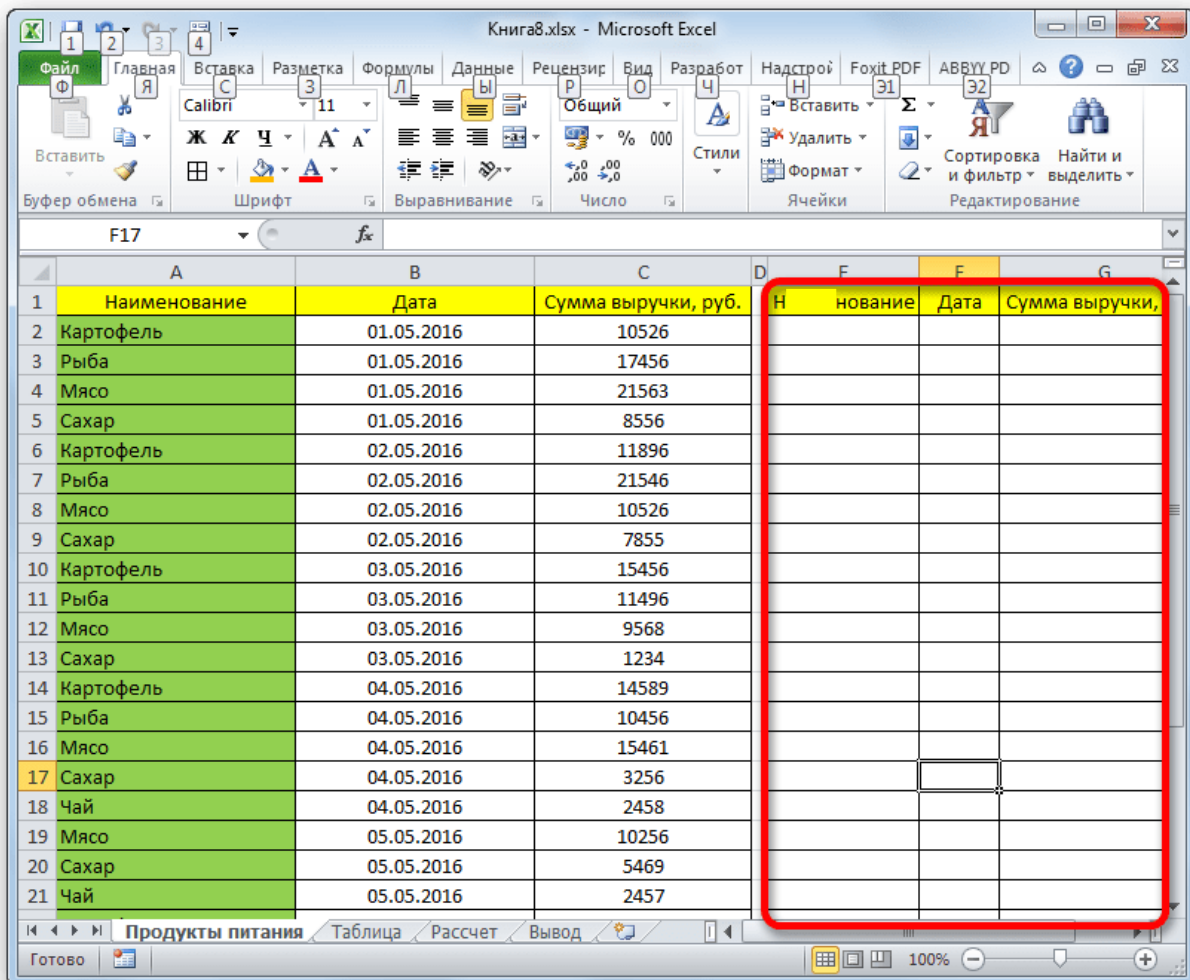


Fig. 1.16. The incoming data

Select all the empty cells in the first column of the table. We place the cursor in the formula row. This will include a sampling formula based on the

specified criteria. We will select the rows with a revenue exceeding \$15 000. In our specific example, the introduced formula will look like this (Fig. 1.17):

= INDEX(A2:A29;LESS(IF(15000<= C2:C29;LINE(C2:C29);"");LINE()-LINE(\$C\$1))-LINE(\$C\$1)).

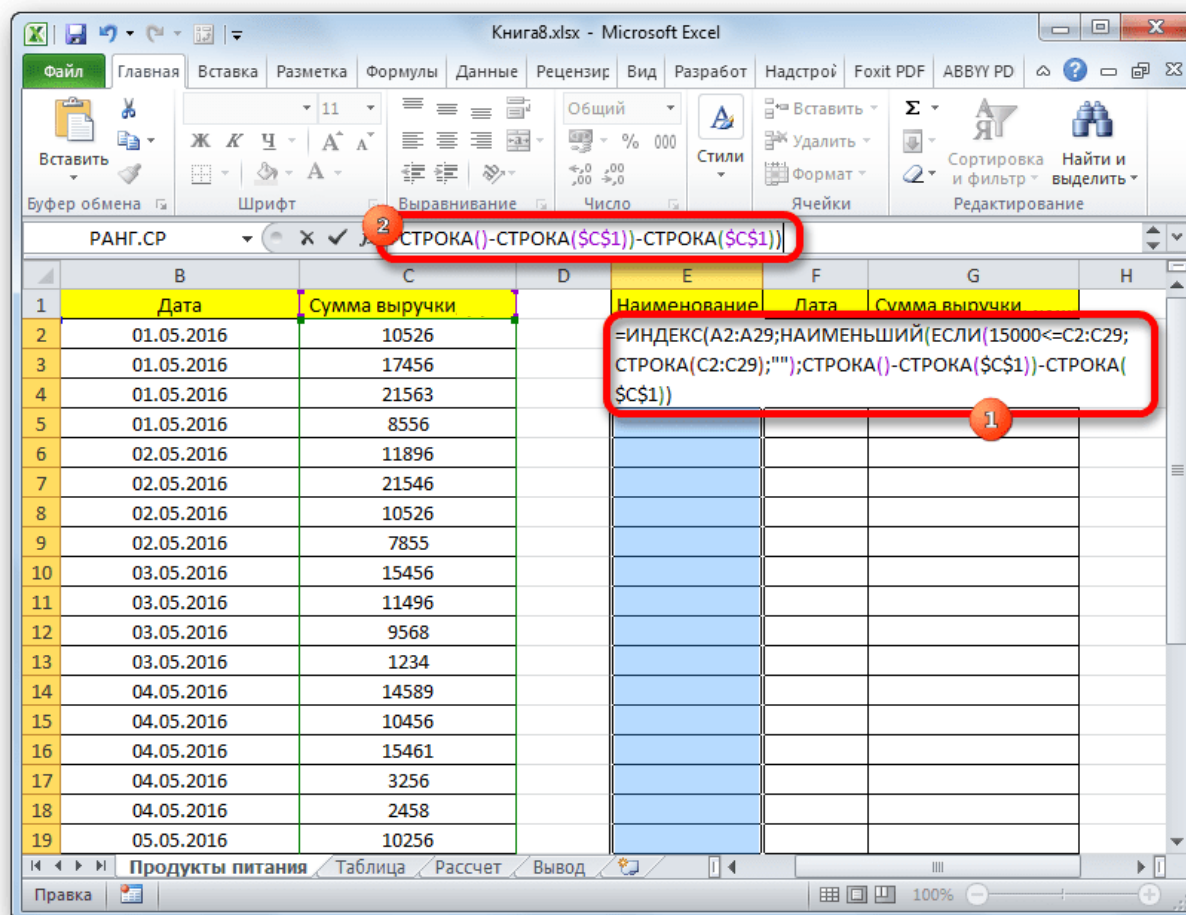


Fig. 1.17. The data filtering options

In each case, the cell and range addresses will be different. In this example, you can map the formula to the coordinates in the illustration and adjust it to your needs.

Since this is an array formula, in order to apply it, you must press the *Enter* button, and the keyboard shortcut *Ctrl + Shift + Enter*.

Selecting the second column with dates and placing the cursor in the formula row, enter the following expression:

= INDEX(B2:B29;LOWER(IF(15000<=C2:C29;LINE(C2:C29);"");LINE()-LINE(\$C\$1))-LINE(\$C\$1)).

We press *Ctrl + Shift + Enter*.



Similarly, enter the following formula in the revenue column:  
`=INDEX(C2:C29;LOWER(IF(15000<=C2:C29;LINE(C2:C29);""));LINE()-LINE($C$1))-LINE($C$1)).`

Again, press *Ctrl + Shift + Enter*.

In all three cases, only the first coordinate value changes, and for the rest, the formulas are completely identical. The result of filtering based on these parameters is presented in Fig. 1.18.

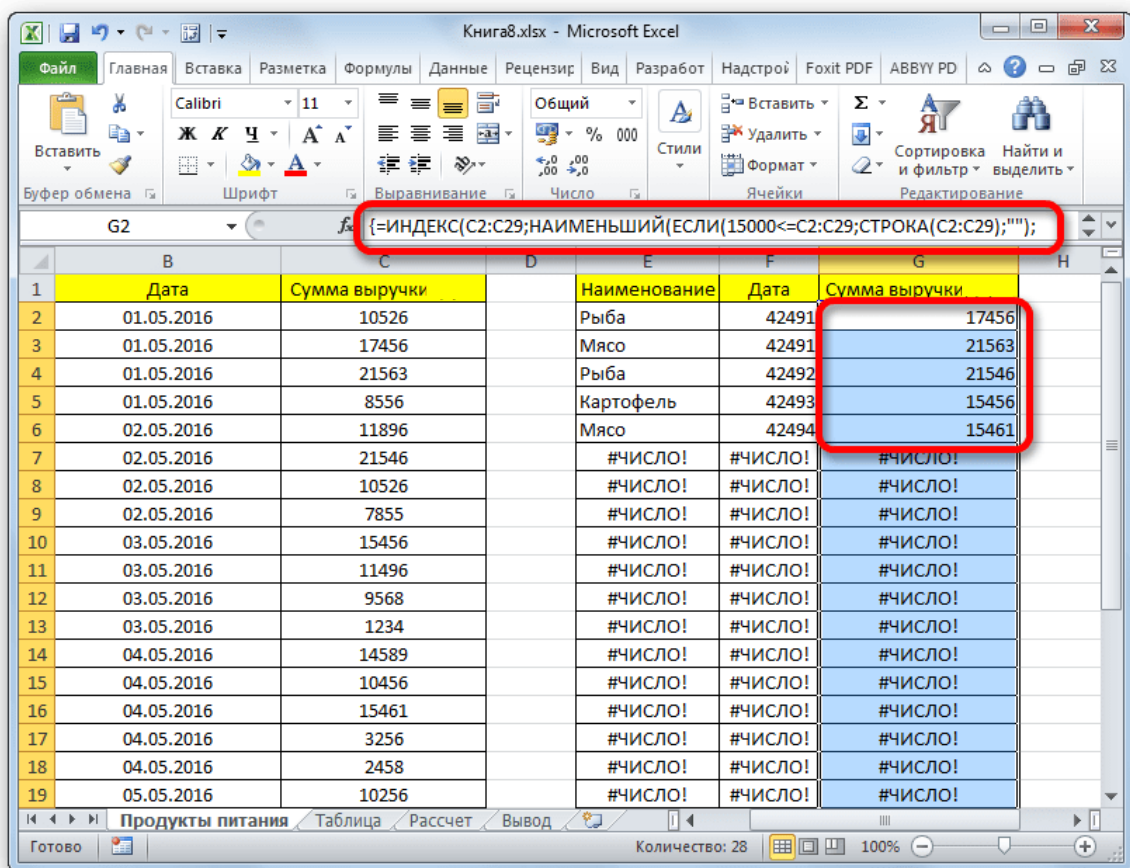


Fig. 1.18. The result of filtering based on the given parameters

As you can see, the table is filled with data, but its appearance is not quite attractive, in addition, the date values are filled in it incorrectly. We need to correct these shortcomings. Date incorrectness is due to the fact that the column cell format is general, and we need to set the date format. Highlight the entire column, including the cells with errors, and right-click. In the list, go to the item "Format cell ...".

In the *Formatting* window, open the *Number* tab. In the block *Numeric formats*, we highlight the value "Date" (Fig. 1.19). In the right part of the window you can select the desired date display type. Once the settings are made, click on the *OK* button.

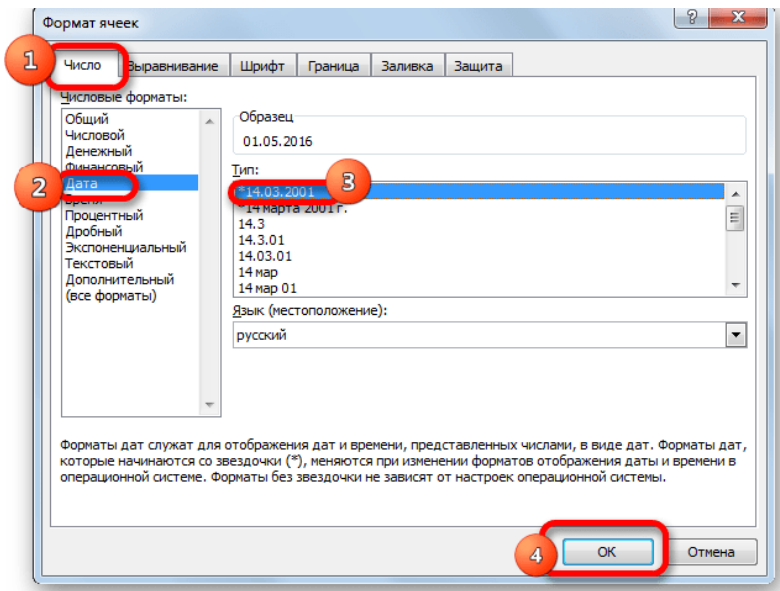


Fig. 1.19. The dialog window for setting data selection options

The date is now displayed correctly. But, as you can see, the entire bottom of the table is filled with cells that contain the false value "# NUMBER!" In fact, these are the cells that did not have enough sampling data. It would be more attractive if they appeared blank at all. For these purposes, we use conditional formatting. Select all cells in the table except the header. While in the *Home* tab, click the *Conditional Formatting* button in the *Styles* toolbar. In the list, select "Create rule ...".

In the window, select the type of rule "Format only cells that contain". In the first box, under the format "Format only cells for which the following condition is true", select *Errors*. Next, click the *Format ...* button (Fig. 1.20).

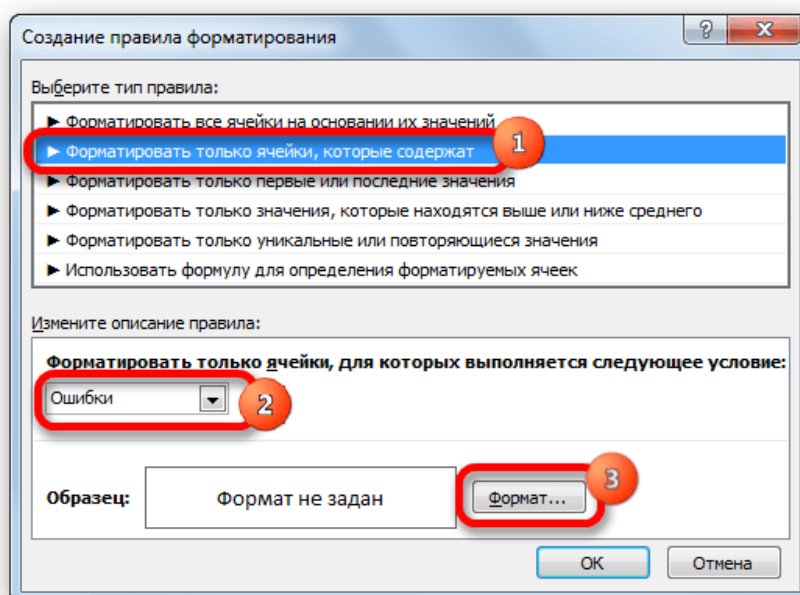


Fig. 1.20. The *Data Formatting Options* dialog box

In the formatting window that opens, go to the *Font* tab and select white in the appropriate box. After these actions, click on the *OK* button. Click the button with the same name after returning to the *Create* window.

After selecting all the necessary parameters, as a result, we will receive the specified restrictions in a separate, properly designed table. The result of such a sample is presented in Fig. 1.21.

The screenshot shows an Excel spreadsheet with two tables. The first table (columns B-C) has columns 'Дата' and 'Сумма выручки'. The second table (columns E-G) has columns 'Наименование', 'Дата', and 'Сумма выручки'. A red box highlights the second table's data rows from row 2 to row 19.

1	Дата	Сумма выручки		Наименование	Дата	Сумма выручки
2	01.05.2016	10526		Рыба	01.05.2016	17456
3	01.05.2016	17456		Мясо	01.05.2016	21563
4	01.05.2016	21563		Рыба	02.05.2016	21546
5	01.05.2016	8556		Картофель	03.05.2016	15456
6	02.05.2016	11896		Мясо	04.05.2016	15461
7	02.05.2016	21546				
8	02.05.2016	10526				
9	02.05.2016	7855				
10	03.05.2016	15456				
11	03.05.2016	11496				
12	03.05.2016	9568				
13	03.05.2016	1234				
14	04.05.2016	14589				
15	04.05.2016	10456				
16	04.05.2016	15461				
17	04.05.2016	3256				
18	04.05.2016	2458				
19	05.05.2016	10256				

Fig. 1.21. The result of the generated sample with the necessary restrictions

**Approach 3:** multiple criteria sampling using a formula.

Just like using a filter, you can use a formula to sample multiple criteria. For example, take the entire source table, as well as an empty table where the results will be displayed, with numeric and conditional formatting already completed. First, set the lower limit on revenue selection at \$15 000, and the second condition – the upper limit at \$20 000.

When forming a sample under this approach, it is necessary to enter the boundary conditions for the sample in a separate column.

As in the previous method, we in turn select the empty columns of the table and write the corresponding three formulas in them. In the first column, enter the following expression:

= INDEX (A2:A29;LOWER(IF((\$D\$2=C2:C29);LINE(C2:C29);""));LINE (C2:C29) -LINE(\$C\$1))-LINE(\$C\$1)).

In the following columns, enter exactly the same formulas, only changing the coordinates immediately after the name of the INDEX operator on the appropriate columns we need, by analogy with the previous method.

Every time you type, make sure you type *Ctrl + Shift + Enter*. The result is presented in Fig. 1.22.

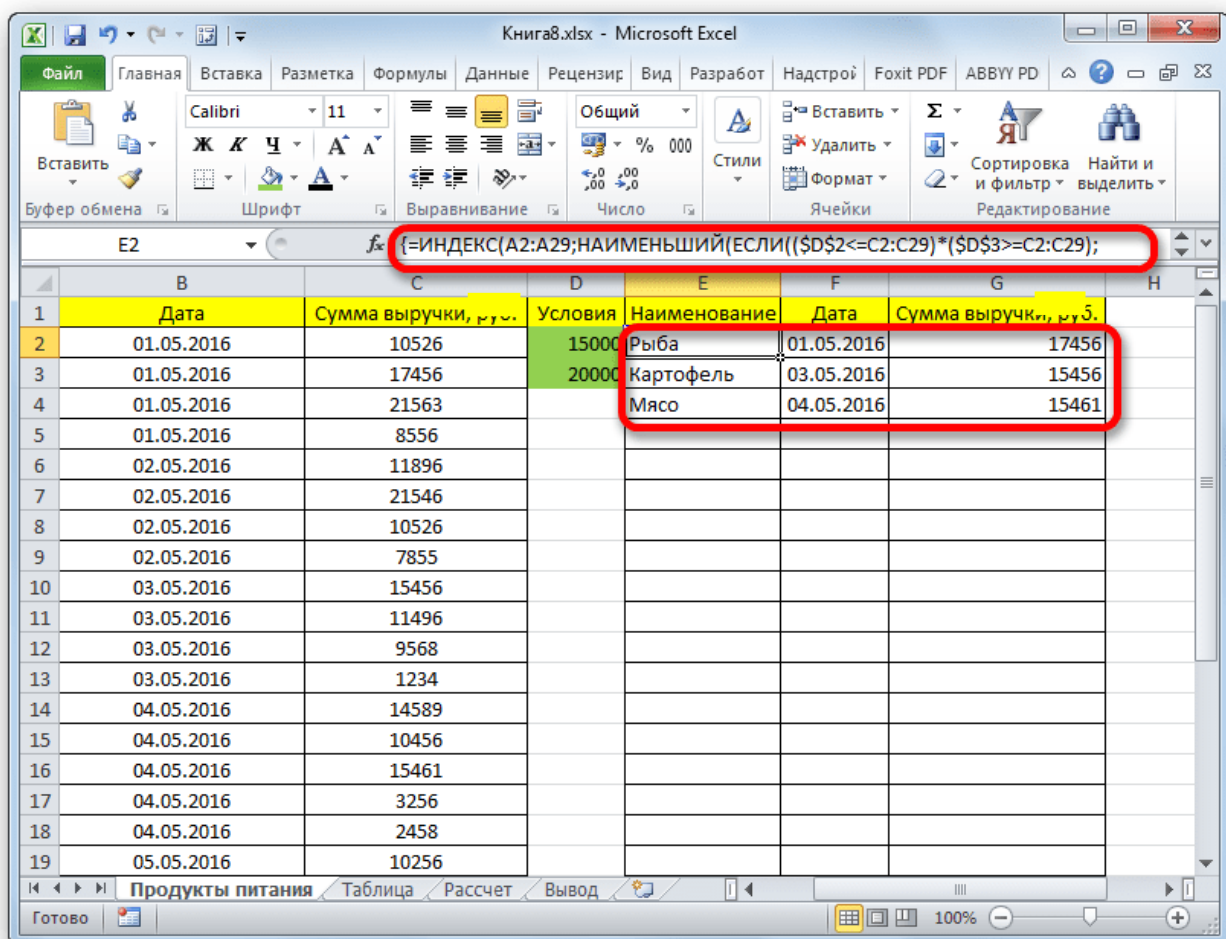


Fig. 1.22. The result of the generated sample

The advantage of this approach over the previous one is that if you want to change the boundaries of the sample, you do not need to change

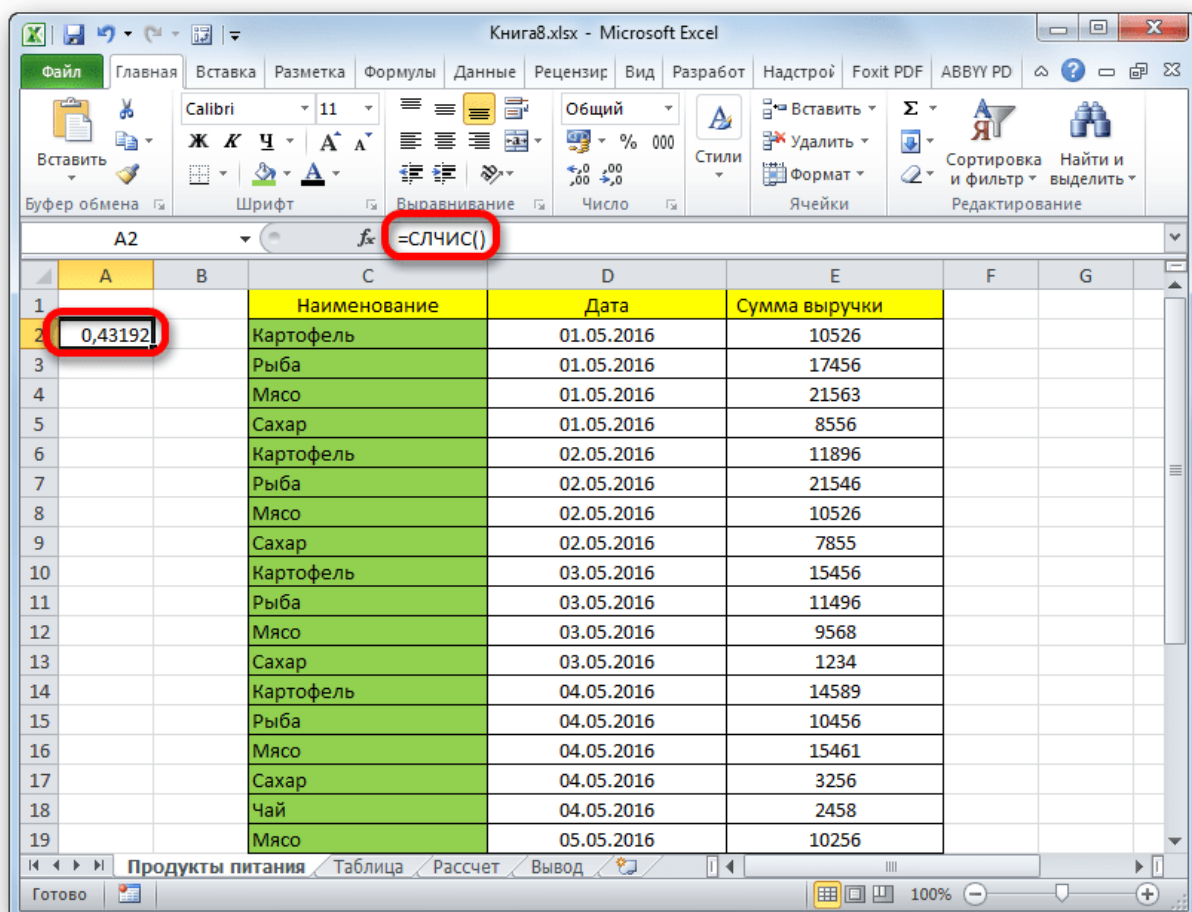
the formula of the array, which in itself is quite problematic. It is enough in the column of conditions on the sheet to change the limit numbers to those that the user needs. The results of the selection will automatically change immediately.

#### **Approach 4:** random sampling.

In Excel, you can also use random selection using the special RAND formula. It needs to be done in some cases when dealing with a large amount of data, when you want to present the big picture without comprehensive analysis of all data in the array.

We skip one column to the left of the table. In the cell of the next column opposite the first cell with the data of the table, enter the formula: = RAND ().

This function displays a random number. To activate it, press the *ENTER* button. The result is presented in Fig. 1.23.



The screenshot shows the Microsoft Excel interface with a table of food products. The formula bar at the top shows the formula =СЛЧИС() in cell A2, which is circled in red. The table has columns for Name, Date, and Revenue. The value 0,43192 is displayed in cell A2, also circled in red.

	A	B	C	D	E	F	G
1			Наименование	Дата	Сумма выручки		
2	0,43192		Картофель	01.05.2016	10526		
3			Рыба	01.05.2016	17456		
4			Мясо	01.05.2016	21563		
5			Сахар	01.05.2016	8556		
6			Картофель	02.05.2016	11896		
7			Рыба	02.05.2016	21546		
8			Мясо	02.05.2016	10526		
9			Сахар	02.05.2016	7855		
10			Картофель	03.05.2016	15456		
11			Рыба	03.05.2016	11496		
12			Мясо	03.05.2016	9568		
13			Сахар	03.05.2016	1234		
14			Картофель	04.05.2016	14589		
15			Рыба	04.05.2016	10456		
16			Мясо	04.05.2016	15461		
17			Сахар	04.05.2016	3256		
18			Чай	04.05.2016	2458		
19			Мясо	05.05.2016	10256		

**Fig. 1.23. The result of sampling**

To make an entire column of random numbers, place the cursor in the bottom right corner of the cell that already contains the formula. A fill marker appears. We drag it down with the left mouse button pressed parallel to the data table up to the end.

The result is a range of cells filled with random numbers. However, it contains the RAND formula. To work with values, you must copy to the empty column on the right. We distinguish a range of cells with random numbers. From the *Home* tab, click the *Copy* icon on the ribbon. Then select an empty column and right-click on the context menu. In the *Insert Options* toolbar, select the *Value* shown as a number icon.

After that, while in the *Home* tab, click on the *Sort and Filter* icon. In the drop-down list, select *Custom Sorting*.

The *Sorting Options* window is activated (Fig. 1.24). Be sure to check the box next to "My data contains headers" if there is a cap and there is no tick. In the *Sort by* field, specify the name of the column that contains the copied values of random numbers. In the *Sort* field we leave the default settings. In the *Order* field, you can select the parameter *Ascending* and *Descending*. For random sampling, this does not matter. Once the settings are done, click on the *OK* button.

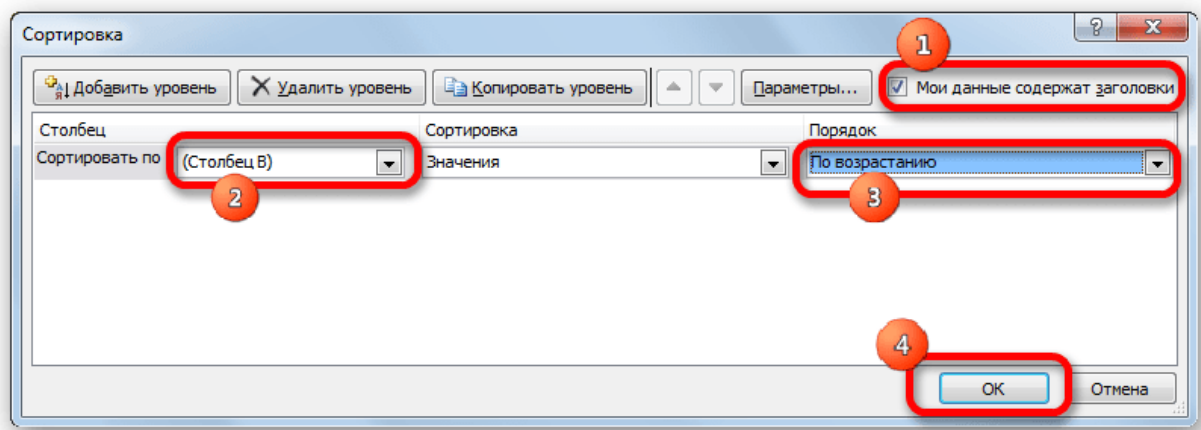


Fig. 1.24. The *Options* dialog window

After that, all values of the table are constructed in order of increasing or decreasing random numbers. You can take any number of the first rows from tables (5, 10, 12, 15, etc.) and it can be considered as a random sample (Fig. 1.25).

Книга8.xlsx - Microsoft Excel

	A	B	C	D	E	F	G
1			Наименование	Дата	Сумма выручки, р		
2	0,435323	0,000397	Картофель	03.05.2016	15456		
3	0,808892	0,024653	Рыба	06.05.2016	11784		
4	0,734098	0,058724	Рыба	02.05.2016	21546		
5	0,000684	0,116991	Мясо	05.05.2016	10256		
6	0,664038	0,167573	Сахар	06.05.2016	4578		
7	0,824356	0,271273	Картофель	07.05.2016	14256		
8	0,746063	0,27737	Картофель	01.05.2016	10526		
9	0,149064	0,315085	Сахар	04.05.2016	3256		
10	0,727885	0,328215	Рыба	01.05.2016	17456		
11	0,292055	0,421079	Картофель	02.05.2016	11896		
12	0,044867	0,480894	Мясо	04.05.2016	15461		
13	0,405097	0,507331	Сахар	02.05.2016	7855		
14	0,390004	0,507508	Рыба	03.05.2016	11496		
15	0,273046	0,509845	Мясо	06.05.2016	13485		
16	0,073342	0,581274	Мясо	01.05.2016	21563		
17	0,003232	0,599017	Сахар	05.05.2016	5469		
18	0,464047	0,649743	Рыба	07.05.2016	13858		
19	0,226712	0,660384	Чай	04.05.2016	2458		

Fig. 1.25. The result of the generated representative sample

In this way, the Excel spreadsheet can be sampled either by *AutoFilter* or by applying special formulas. In the first case, the result will be output to the source table, and in the second one – to a separate area. It is possible to make a selection both on one and several conditions. You can also randomly sample using the RAND function.

## Laboratory work 2

### Tools for visualization of statistical data

*The purpose* of the work is to learn how to visualize data with modern application packages.

*The task* is to acquire data interpretation skills using anomalous software.

#### Guidelines

Research, experiments, analysis are always accompanied by data visualization. Not only do diagrams and charts graphically represent values,

they also have a meaningful load on the topic of the study. One can see at once the general picture of the results of the study, a large amount of data on a limited interval. In addition, attention is paid to some elements of the data series, data fragments are compared, the trend in the change of properties under study is demonstrated, and so on.

Infographics (from the Latin *informatio* – informing, explaining, and the Greek *uraficos* – writing) is a visual display of information, statistics for simple and visual demonstration of trends, relationships, as well as interest in the subject of research.

It is important to choose the right type for the content load of a graph or chart. Consider the different types of diagrams in the examples.

Histograms are used to illustrate the dynamics of time-varying data or the distribution of data in the form of rectangular regions, the size of which corresponds to the values obtained in the statistical survey (Fig. 2.1).



Fig. 2.1. An example of a histogram

The normalized flat bar graphs show the percentage of each value in the total for the different categories.

Each year, the Ukrainian Center for Educational Quality Assessment publishes the results of a statistical analysis of higher education institutions. In Fig. 2.2 a line diagram of the compilation of external independent assessment in the Ukrainian language of the students of Kropyvnytskyi is presented. The results are presented for the educational institutions, so each band is responsible for the results of the type of educational institutions, and the color indicates the range of points.



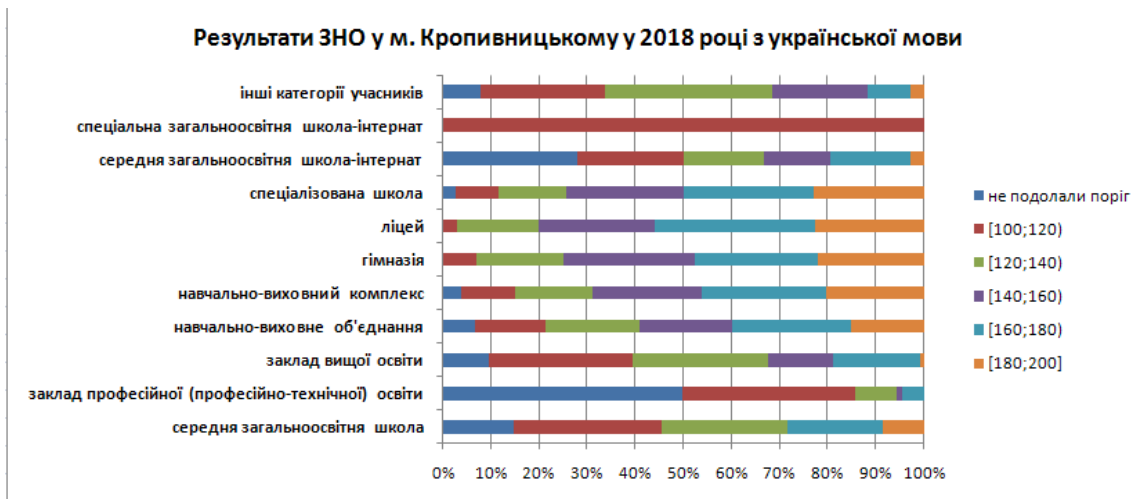


Fig. 2.2. An example of a stacked histogram

Pie charts are used to illustrate the relationship between the parts of a sample using circular diagrams. Fig. 2.3 shows the share of each speciality in the total number of students enrolled in 2018.

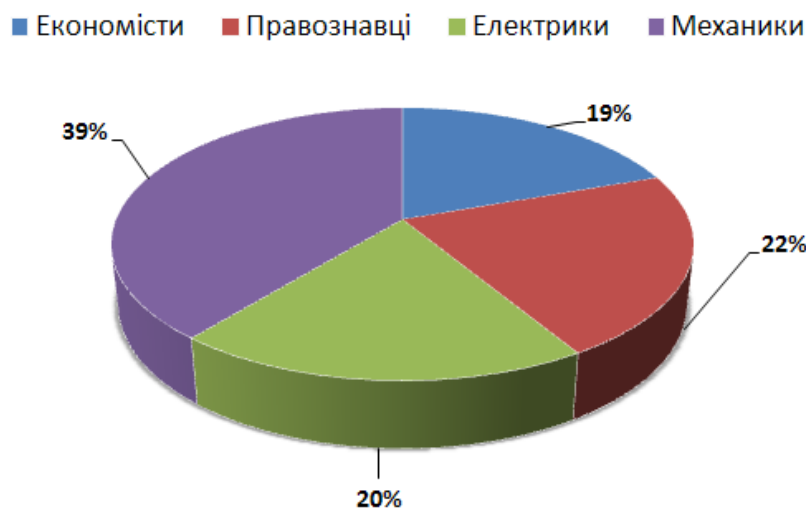


Fig. 2.3. A pie chart

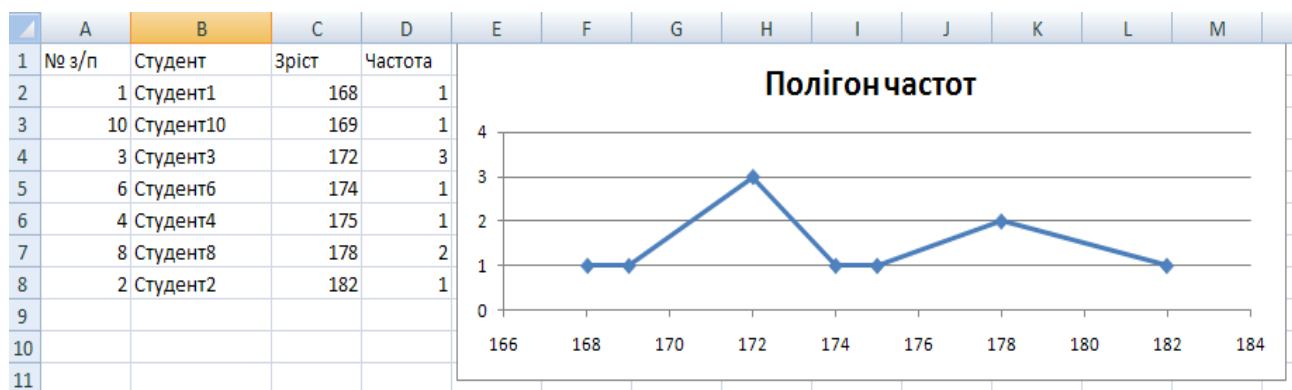
The polygon of frequencies is called a polynomial with vertices at points  $(z_i, n_i)$ ,  $i = 1, 2, \dots, k$ . Here,  $z_i$  is the value of the  $i$ -th variant, and  $p_i$  is the frequency corresponding to this variant.

To construct a polygon of frequencies, the variants  $z$  are plotted on the abscissa axis, and corresponding frequencies are plotted on the ordinate axis. The points  $(z_i, n_i)$  are connected by segments of straight lines and make a polygon of frequencies.

In fact, the frequency range is a graphical representation of the interval row.

The example that follows, shows how to use MS Excel to construct a frequency range. Suppose the frequency table is already built. There are two columns in it: row values (they will be plotted along the Ox axis) and relative frequencies (they will be presented by Oy). Now the Chart Wizard. It will ask us to do the following:

1. Determine the type of graph (the point diagram on which the values are connected by segments).
  2. Specify the source data (they are already selected: X is value, Y is frequency);
  3. Specify additional design options (cannot be highlighted);
  4. Specify the location for the chart (by default – on the same sheet).
- The frequency range is ready (Fig. 2.4).



**Fig. 2.4. An example of plotting a frequency range**

Mapping is designing and generating geological and topographic images and other graphs on display devices (mainly on a display screen) based on digital output, rules and algorithms for converting them.

Modern journalism and research are actively using infographic tools; in addition, all world organizations and international economic, political and social foundations use cartography as one of the interactive tools for visualization of data in reports.

In support of this fact, a revision of 7 most interesting visual maps prepared by various sources are presented.

In 2013, the World Economic Forum presented the results of a survey of residents in 140 countries. Respondents expressed their attitude to foreign visitors – so the organization was able to draw a map of the most and least friendly states. The most hospitable were Iceland, New Zealand and Morocco. However, look at your country (Fig. 2.5).

## The friendlies countries in the world

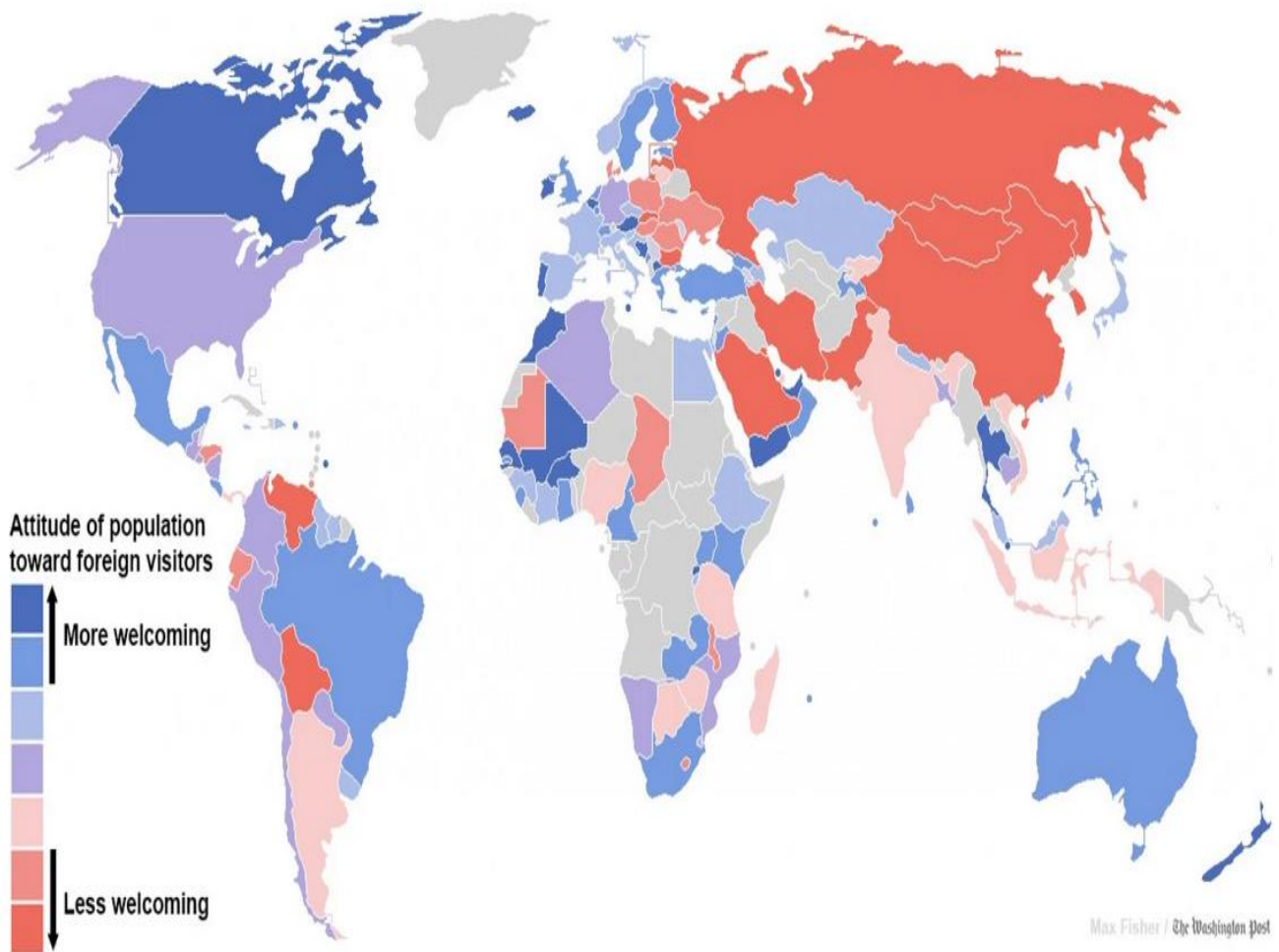


Fig. 2.5. The geographic map

For graphical representation, Excel provides the user with a significant set of features. You can make charts on either a single worksheet with a spreadsheet or on a separate worksheet called a chart sheet. A chart created on a single worksheet with a spreadsheet is called an embedded chart.

### Laboratory work 3 Grouping of data and research on groups

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

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

## Guidelines

Summarization and grouping of statistical data should be performed in the mode *Histogram*, *Data Analysis* add-in.

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

The dialog box of this mode (Fig. 3.1) assumes the following parameters:

1. The input range, i.e. the location of the input data on your worksheet.
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 Microsoft 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.

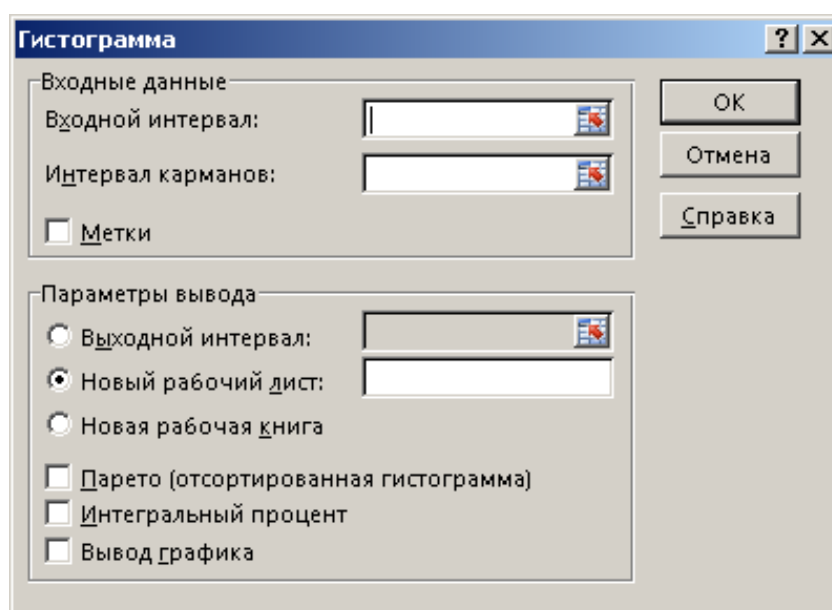


Fig. 3.1. The *Histogram* window

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. 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 on the output range.

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

6. 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.

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

*Example.* The GDP volume per capita (current USD) in 2011 is shown in Table 3.1.

Table 3.1

### The input data

Country name	2011	Country name	2011
1	2	3	4
Albania	4,029.73	Latvia	12,726.4
Armenia	3,305.49	Lithuania	13,339.2
Austria	49,707.1	Luxembourg	115,038
Azerbaijan	6,915.76	Macedonia, FYR	4,925.34
Belarus	5,820.35	Moldova	1,966.93
Belgium	46,469.2	Montenegro	7,197.13
Bosnia and Herzegovina	4,820.67	Netherlands	50,087.3
Bulgaria	7,158.12	Norway	98,102.5
Croatia	14,488.3	Poland	13,462.9
Cyprus	30,670.3	Portugal	22,329.8
Czech Republic	20,407.3	Romania	8,405.49
Denmark	59,683.8	Russian Federation	13,089.3
Estonia	16,555.8	Serbia	6,203.47
Finland	49,391.3	Slovak Republic	17,646
France	42,377.4	Slovenia	24,141.9
Georgia	3,202.53	Spain	32,244.2
Germany	43,689.3	Sweden	56,927
Greece	26,427.2	Switzerland	80,390.8
Hungary	14,043.7	Tajikistan	934.82
Iceland	44,072.3	Turkey	10,498.3

Table 3.1 (the end)

1	2	3	4
Ireland	48,423.2	Turkmenistan	4,721.96
Italy	36,115.7	Ukraine	3,615.38
Kazakhstan	11,244.9	United Kingdom	38,817.8
Kosovo	3,592.6	Uzbekistan	1,545.93
Kirgiz Republic	1,074.74		

Enter the input data into the cells A1:B50.

You need to construct a histogram and accumulated frequencies in the histogram. In order to solve the task, the *Histogram* mode could be used. The parameters in the dialog *Histogram* are shown in Fig. 3.2.

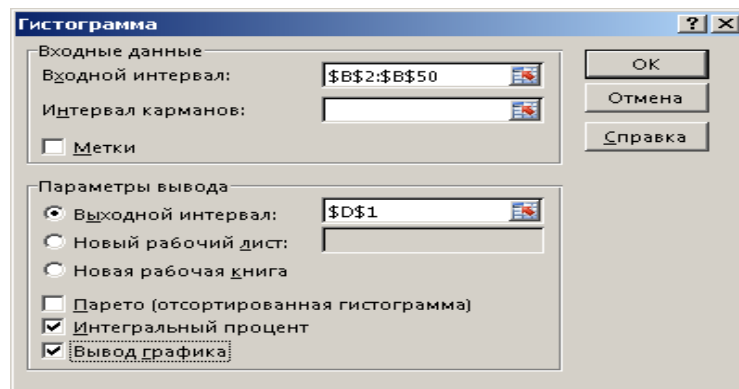


Fig. 3.2. Filling the *Histogram* window

The frequencies and cumulative frequencies are presented in Table 3.2.

Table 3.2

### The frequencies and cumulative frequencies

Bin	Frequency	Cumulative, %
934.8201	1	2.04
17,235.26	26	55.10
33,535.69	7	69.39
49,836.13	9	87.76
66,136.57	3	93.88
82,437	1	95.92
98,737.44	1	97.96
More	1	100.00

So we see uneven distribution of countries based on the indicator "the GDP per capita": 26 countries of 49 have a value of the GDP per capita from 934.82 dollars to 17,235.26 dollars.

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

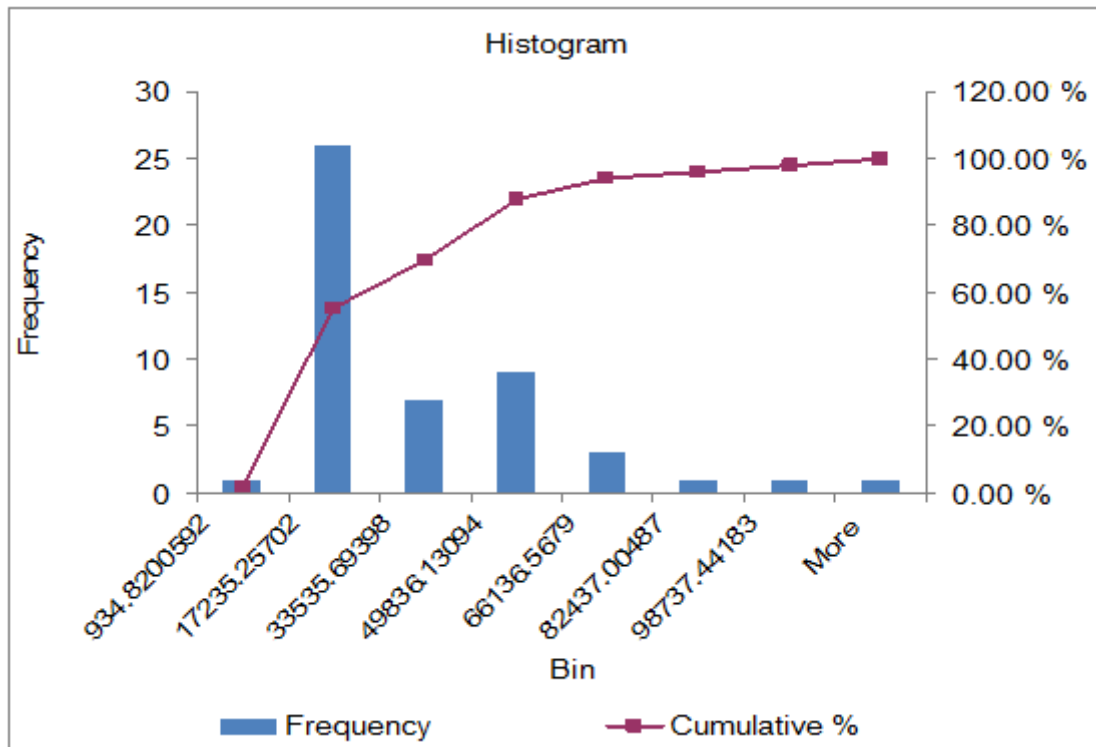


Fig. 3.3. The histogram of distribution

In Table 3.2, the column "Cumulative, %" explains the calculation of the accumulated frequencies. Cumulative frequency is calculated based on the frequency. Each value of cumulative frequency is divided by the maximum accumulated value. The result is relative frequency expressed in fractions of a unit. After the transformation to a percentage format the final result is obtained.

There is another way to aggregate and group statistics. At the beginning of grouping it is necessary to construct a summary table. The summary table is constructed as follows:

- 1) call *Data* → *Summary Table*;
- 2) on request for a data source, select the option *in the list or in the database Microsoft Office Excel*;
- 3) in response to the proposal to specify the range with the data, select on the worksheet *DB* columns from "A1" to "F8200" inclusive. There should be no empty row in the database!

4) click on the *Next* button. Excel will display a window asking you to define the dimensions of a pivot table or to build a layout. We click on the *Layout* button;

5) in the row area we put the field *Price*, in the data area we transfer the field *Sum*;

6) double-click on the field *Sum by field Sum*. The *PivotTable Field Calculation* window appears with a list of available operations. Select from the list the operation *Amount*;

7) in the field *Name*: enter the name "Amount, \$". In the *Calculate PivotTable Field* window, click *OK*;

8) in the window of creation of a layout we press on *OK*, then get to the main window of the *Master* of the summary tables;

9) here we click on the *Options* button, the *Summary table options* window will open;

10) in this window, we disable the checkboxes *Total amount by columns*, *Total sum by rows*, *Auto format*;

11) click on *OK* to return to the main window of the *Summary table wizard*;

12) Click on *Finish*.

As a result, we received a summary report, a form. Based on the resulting video table, you can group data.

When working with pivot tables, the database may contain empty rows. This is often used in practice. For example, instead of an exact block of values with the input data, only the column indices are entered. This is convenient because you do not need to rebuild the table when replenishing the database with new records. If you work with interval groups by numbers (or dates), this approach is not applicable! The grouping database should not contain empty records. Therefore, when creating a pivot table, you must explicitly specify the block of cells with the original data. Of course, this has some inconvenience, but it will have to be reconciled.

For example, it is necessary to determine which goods (cheap, expensive or mid-range) receive the bulk of the turnover. To do this, break down the entire set of interval prices and then recalculate the summary report values for each interval separately. In our example, the cost of printers ranges from 455 UAH (Printer Canon pixma ip1900 model) up to 3585 UAH (Printer Epson Stylus Photo 1410 A3). We split this range into eight intervals in 500 UAH increments. Follow these steps:

1) put a pointer on any value of the *Price* field of the summary report, for example on the cell "A5";



2) right-click. A context menu will appear;

3) from the context menu, select *Group and Structure*, then – *Group ...* .

A window will appear asking you to set the start and end intervals and the change step.

Note the checkboxes to the left of the *starting with:* and *by:* fields. If enabled, Excel selects the price grouping interval on its own, automatically. In our example, it will set a minimum price value (455 UAH) in the field *starting with:*, and a maximum value (3585 UAH) in the field *by:*. In most cases, this is wrong. Therefore, grouping is best done by yourself. Fix the values in the *Grouping* window;

4) left-click inside the *starting with:* field. Enter the value "0";

5) left-click on the field *with step:* and enter here "500";

6) click on *OK*. Excel groups the products and calculates the total revenue at each interval. That is, it will generate a sample with the golden parameters.

To unselect and return to the original pivot table form, you must:

1) put the pointer of the active cell on any value of the field *Price*;

2) right-click;

3) from the context menu select *Group and Structure* → *Unbundle*.

After that, the table will look original.

## **Laboratory work 4**

### **Statistical tools for studying the dynamics of phenomena and processes**

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

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

#### **Guidelines**

Let's analyze the dynamic series with the help of analytical indicators. The initial series of speakers is located in the cells (B3:B8) (Fig. 4.1). To calculate the chain, the basic and average indicators of the dynamics of sales volume, we use the formulas, which are presented in the lecture material.

	A	B	C	D	E	F	G	H	I	J
1	Років	Обсяг продукції, тис. грн. (y)	Абсолютний приріст, тис. грн.		Темп зростання, %		Темп приросту, %		Абсолютне значення 1% приросту, тис. грн.(A)	
2			базисної	лампровий	базисної	лампровий	базисної	лампровий		
3	2005	20	–	–	–	–	–	–	–	
4	2006	23	=B4-\$B\$3	=B4-B3	=B4/\$B\$3*100	=B4/B3*100	=E4-100	=F4-100	=0,01*B3	
5	2007	27	=B5-\$B\$3	=B5-B4	=B5/\$B\$3*100	=B5/B4*100	=E5-100	=F5-100	=0,01*B4	
6	2008	28	=B6-\$B\$3	=B6-B5	=B6/\$B\$3*100	=B6/B5*100	=E6-100	=F6-100	=0,01*B5	
7	2009	30	=B7-\$B\$3	=B7-B6	=B7/\$B\$3*100	=B7/B6*100	=E7-100	=F7-100	=0,01*B6	
8	2010	33	=B8-\$B\$3	=B8-B7	=B8/\$B\$3*100	=B8/B7*100	=E8-100	=F8-100	=0,01*B7	
9										
10										
11	Середньорічний обсяг продукції, тис. грн.				=CP3НАЧ(B3:B8)					
12	Середньорічний приріст продукції, тис. грн.				=CP3НАЧ(D4:D8)					
13	Середньорічний темп зростання, %				=CPГЕОМ(F4:F8)					
14	Середньорічний темп приросту, %				=E13-100					
15										
16										
17										

**Fig. 4.1. The initial data for the calculation of the analytical indicators of the dynamics intensity**

We obtain the following results of calculations of analytical indicators of the intensity of the volume of sales for six years (Fig. 4.2).

	A	B	C	D	E	F	G	H	I	J
1	Років	Обсяг продукції, тис. грн. (y)	Абсолютний приріст, тис. грн.		Темп зростання, %		Темп приросту, %		Абсолютне значення 1% приросту, тис. грн.(A)	
2			базисної	лампровий	базисної	лампровий	базисної	лампровий		
3	2005	20	–	–	–	–	–	–	–	
4	2006	23	3	3	115,0	115,0	15,0	15,0	0,20	
5	2007	27	7	4	135,0	117,4	35,0	17,4	0,23	
6	2008	28	8	1	140,0	103,7	40,0	3,7	0,27	
7	2009	30	10	2	150,0	107,1	50,0	7,1	0,28	
8	2010	33	13	3	165,0	110,0	65,0	10,0	0,30	
9										
10										
11	Середньорічний обсяг продукції, тис. грн.				26,8					
12	Середньорічний приріст продукції, тис. грн.				2,6					
13	Середньорічний темп зростання, %				110,5					
14	Середньорічний темп приросту, %				10,5					
15										
16										

**Fig. 4.2. The results of the calculation of analytical indicators of the dynamics intensity**

Thus, the average annual sales volume is 26.8 thousand UAH. During 2005 – 2010 the volume of sales increased by 13 thousand UAH or 65 %. On the average, over the year the sales increased by 2.6 thousand UAH or 10.5 %.

For a visual idea of the dynamics of sales, we present a series of dynamics in the form of a linear graph and a bar chart. To draw a line graph, use the *Chart Wizard* → *Type* → *Graph* → *View1* (Fig. 4.3).

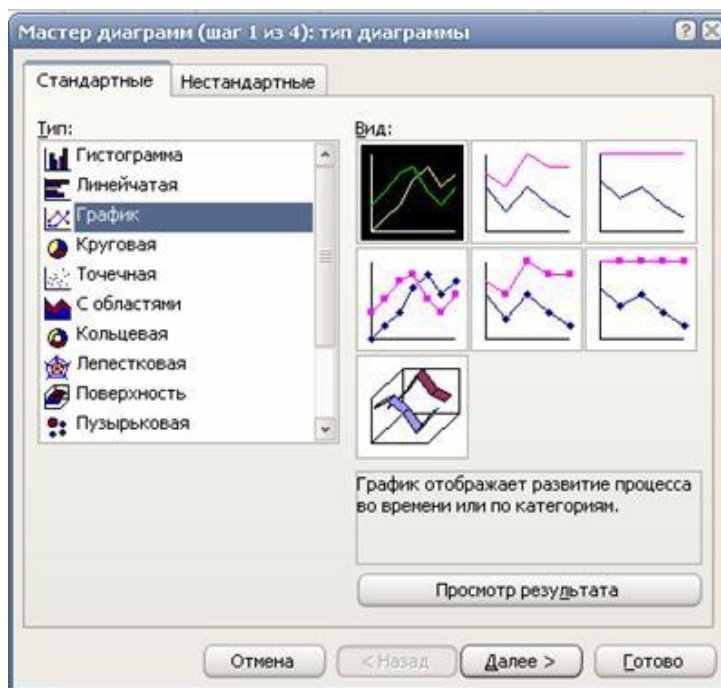


Fig. 4.3. The **Graphics** dialog box

In order to arrange the chart properly, it is advisable to click *Next* → *Chart Options* → *Headings* and enter the name of the chart, the names of indicators in the appropriate fields (Fig. 4.4).

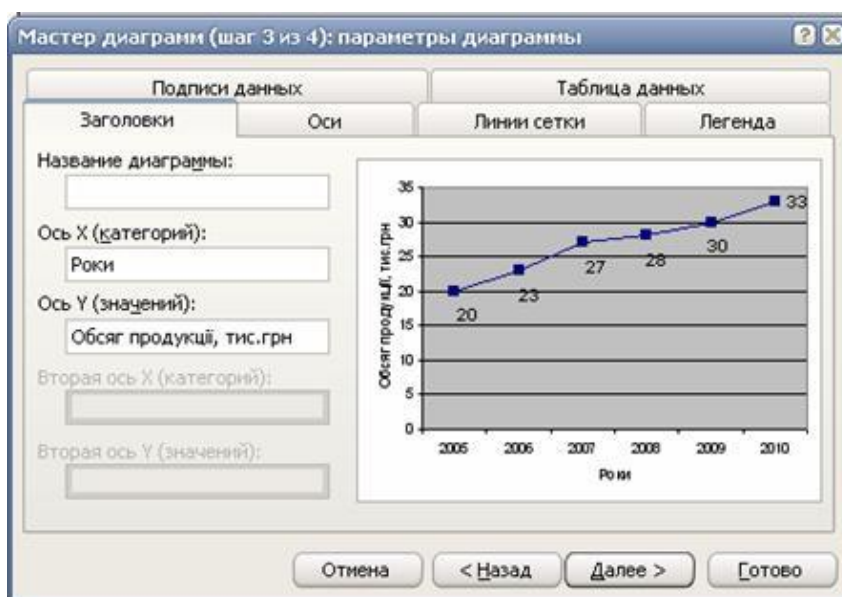


Fig. 4.4. The window for selecting the parameters of the graph

Similarly, a bar chart is built, only the type of chart is selected.

Using the *Chart Wizard* we get a linear graph and a bar chart (Fig. 4.5 and 4.6).

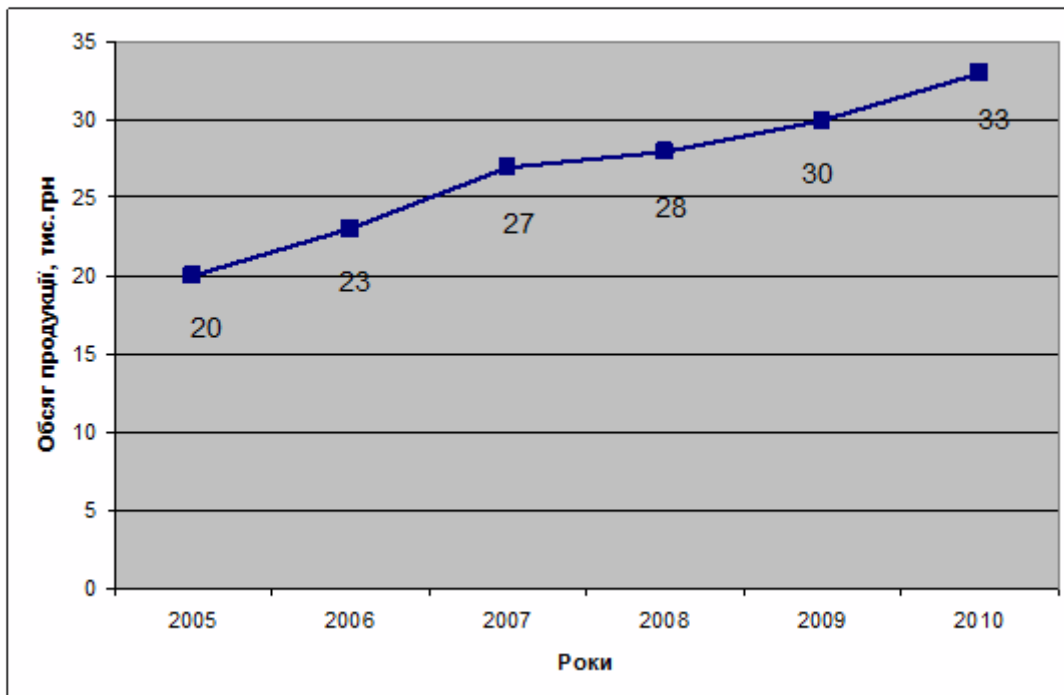


Fig. 4.5. The graph of the production volume dynamics for 2005 – 2010

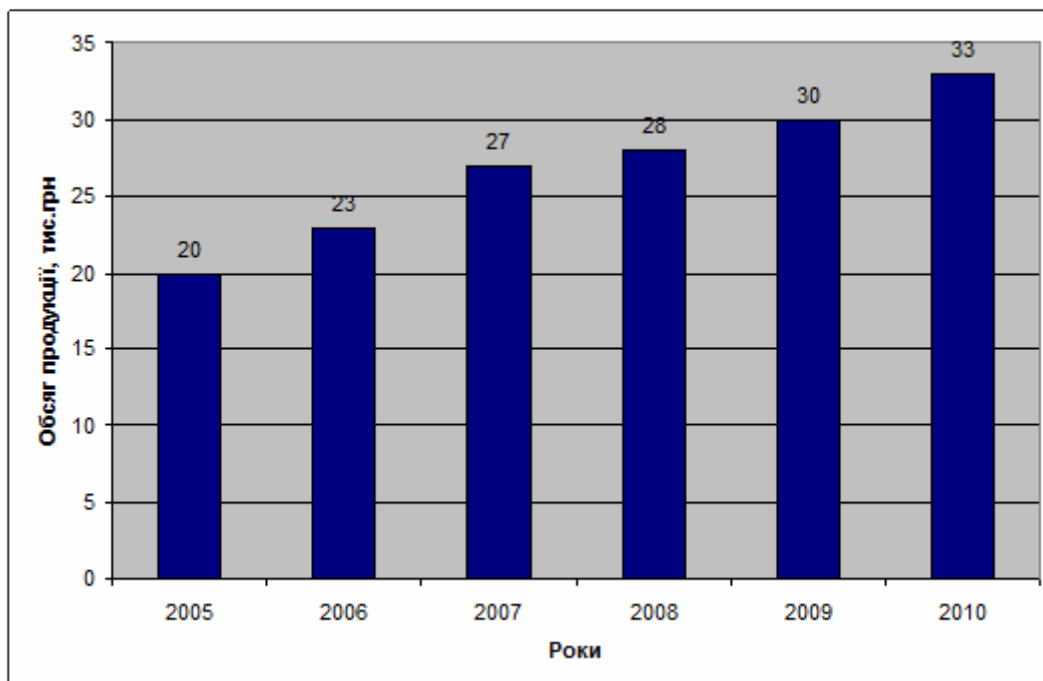


Fig. 4.6. The chart of dynamics of the volume of production for 2005 – 2010

To implement the method of a moving average you can use the package *Data analysis – Analysis tools – Moving average*.

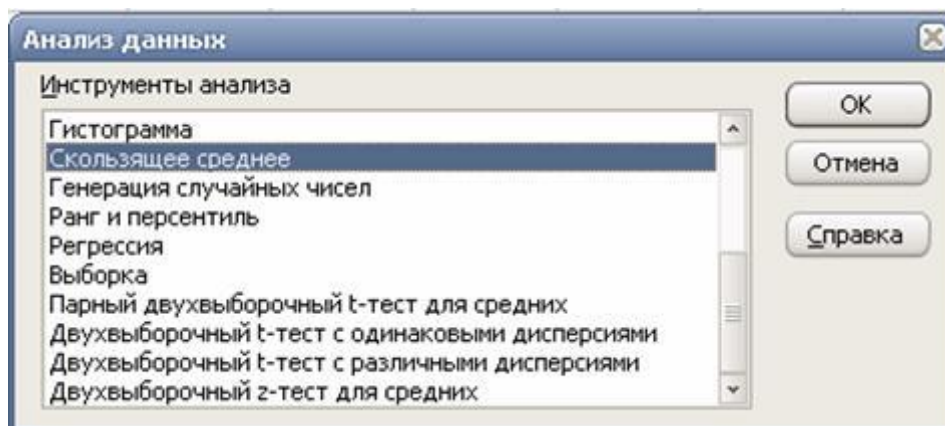


Fig. 4.7. The data analysis dialog

In the field *Input interval*, enter the initial series of dynamics (Fig. 4.8); in the field *Interval*, set the interval of alignment 3, make a mark for plotting.

	A	B	C	D	E	F	G	H	I	J	K
	Роки	Випуск продукції, тис. грн	Ковзна середня								
1											
2	1996	15,6									
3	1997	16									
4	1998	18,9									
5	1999	15,7									
6	2000	20									
7	2001	19,6									
8	2002	19,8									
9	2003	21,5									
10	2004	20									
11	2005	27,3									
12	2006	24,4									
13	2007	28,2									
14	2008	27,9									
15	2009	33,1									
16	2010	32,7									
17											
18											

**Скользящее среднее**

Входные данные  
 Входной интервал:

Метки в первой строке

Интервал:

Параметры вывода  
 Выходной интервал:

Новый рабочий лист:  
 Новая рабочая книга:

Вывод графика       Стандартные погрешности

ОК      Отмена      Справка

Fig. 4.8. The initial data on the implementation of the moving average method

The main disadvantage of this method is that the aligned row becomes shorter than the original one due to the loss of levels at the beginning and end of the row. After executing the program, we obtain the following results (Fig. 4.9).

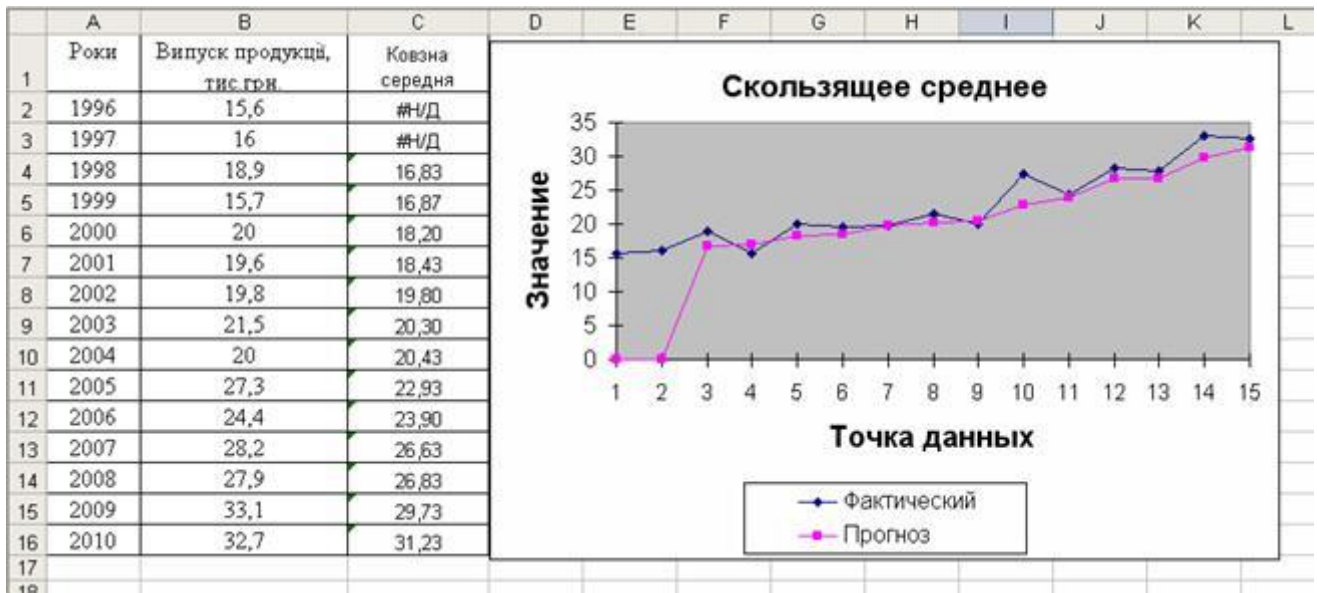


Fig. 4.9. The results of the method of a moving average

Determine the identification of trends in the phenomenon of analytical alignment. The original series of years is located in cells A1:B16.

To obtain the five types of trend equations predicted in Excel, use the *Chart Wizard* → *Type* → *Graph* (Fig. 4.10).

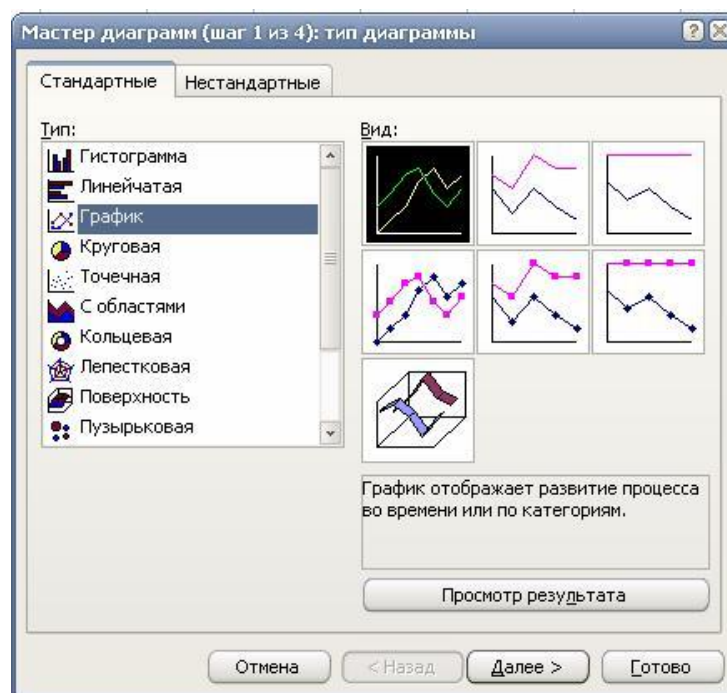


Fig. 4.10. Construction of a moving average graph

Once a linear graph is constructed, five trend equations can be obtained: linear, logarithmic, polynomial, power, and exponential. To select the best trend equation, it is necessary to construct all five types of graphs,

display the respective equation and the magnitude of the R2 error in the field of each graph. To do this, place the cursor on the graph line and right-click, in the window that appears, select the command *Add trend line*, select the desired type of equation (Fig. 4.11).

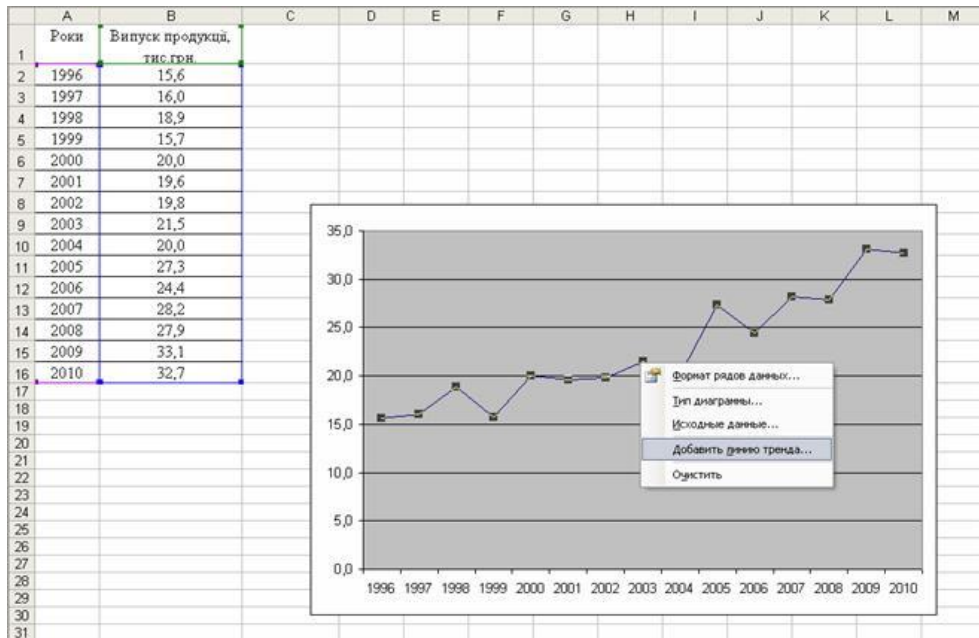


Fig. 4.11. Choosing a trend equation

Then you need to specify two commands in the *Parameters* submenu – *Show the equation on the chart* and *Plot the value of the approximation reliability* (Fig. 4.12).

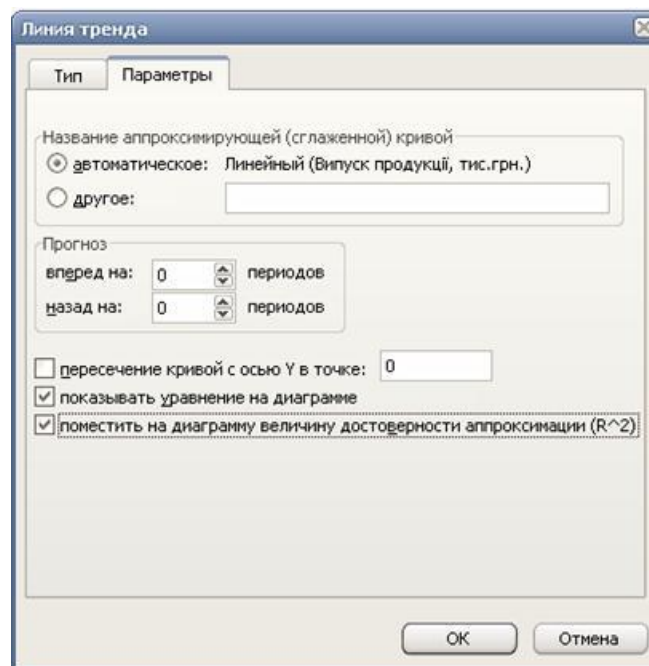


Fig. 4.12. The window for selecting the trend parameters

For example, the result of the selection of linear and polynomial (2nd degree polynomial) trend equations according to the data given in cells A2:B16 is presented in Fig. 4.13 and 4.14.

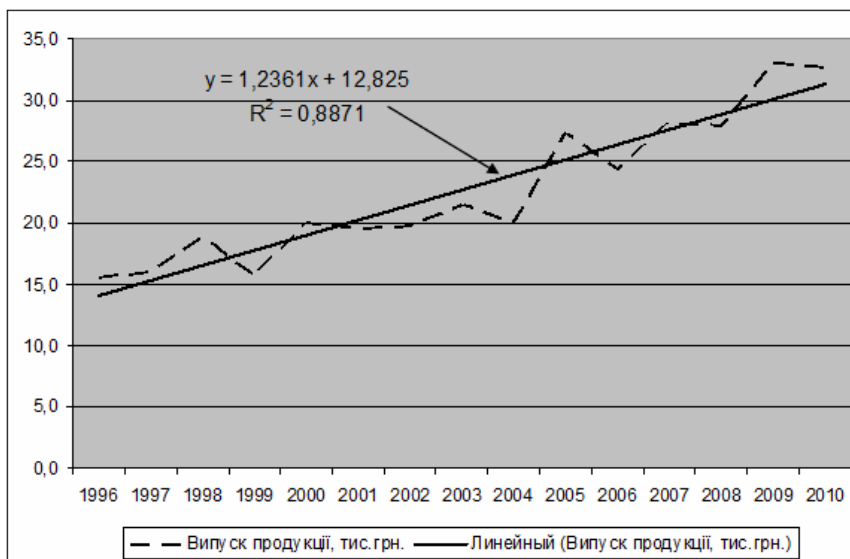


Fig. 4.13. The result of the selection of the linear trend equation

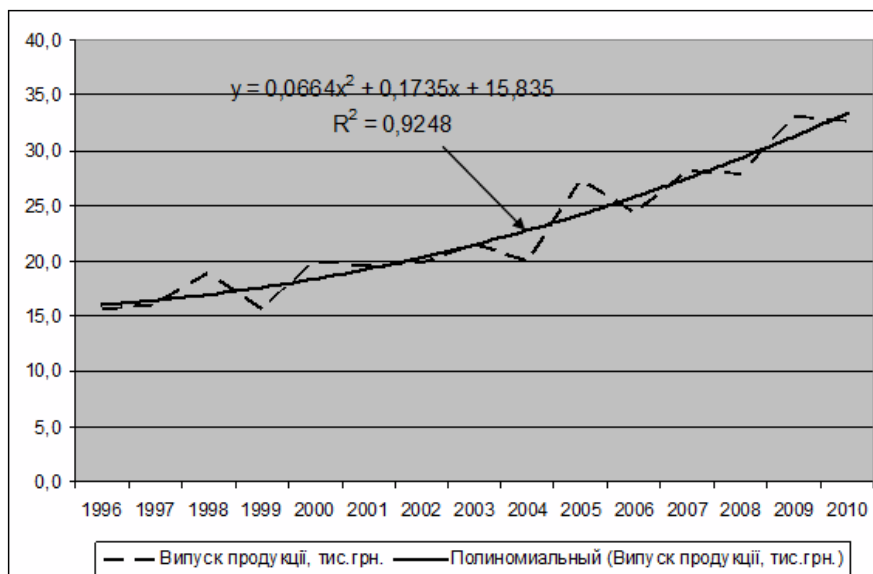


Fig. 4.14. The result of the selection of the polynomial (2nd degree polynomial) trend equation

By the criterion R2, the polynomial provides a higher level of approximation of the original series of dynamics, so this model has an advantage over the linear one. Similarly, you can compare the results of calculations by other types of equations, using the proposed list (Fig. 4.15).



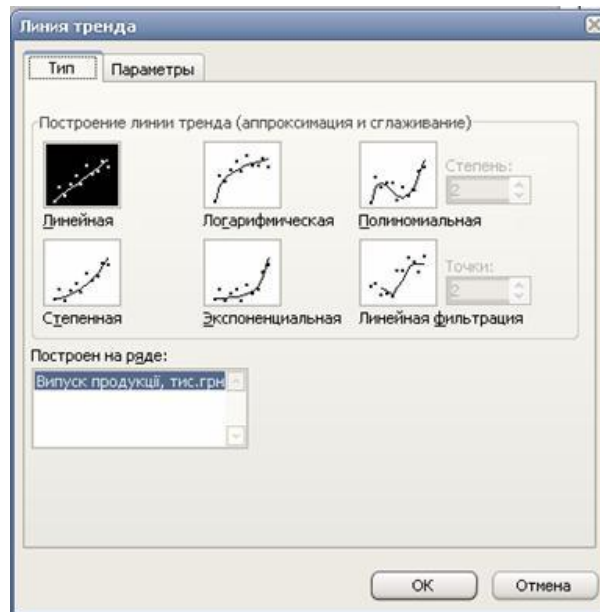


Fig. 4.15. The trend line selection window

Calculate the seasonality index according to the data entered in cells B3:E6 in Fig. 4.11.

To make a hypothesis about the presence and nature of seasonal variations, we plot a diagram using the *Wizard of diagrams*, but first we will form a single series of dynamics in four years (cells G2:G17) (Fig. 4.16).

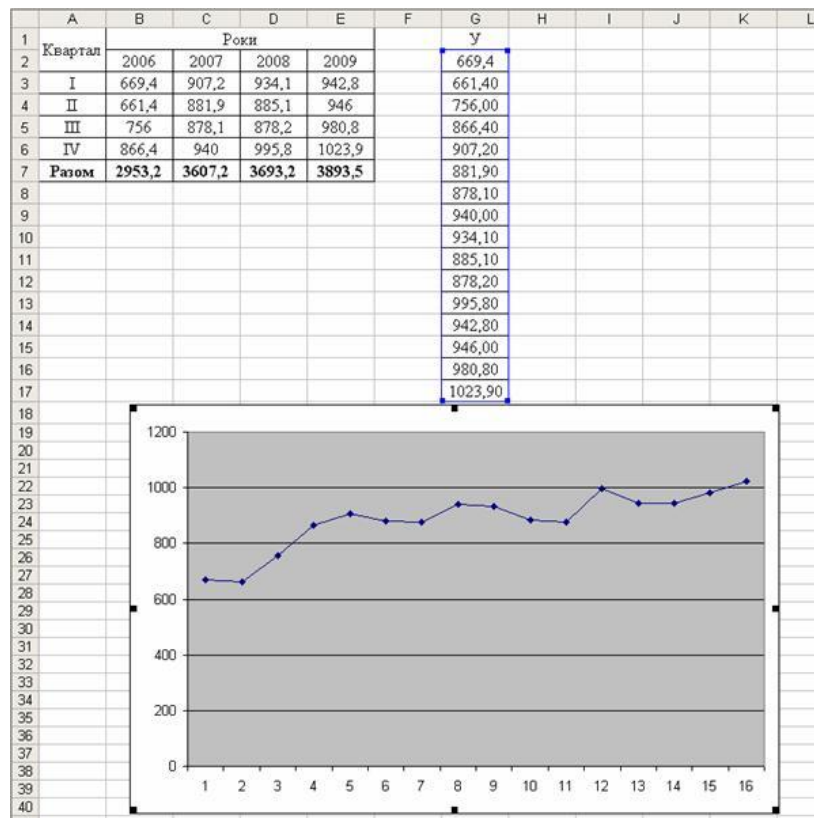


Fig. 4.16. Detection of seasonal variations by the graphical method

The obtained graph gives grounds for the conclusion that the values of the indicator do not fluctuate, that is, they fluctuate around the average level, so we use the formula to estimate seasonal fluctuations:

$$I_s = \frac{\bar{y}_i}{\bar{y}} \times 100.$$

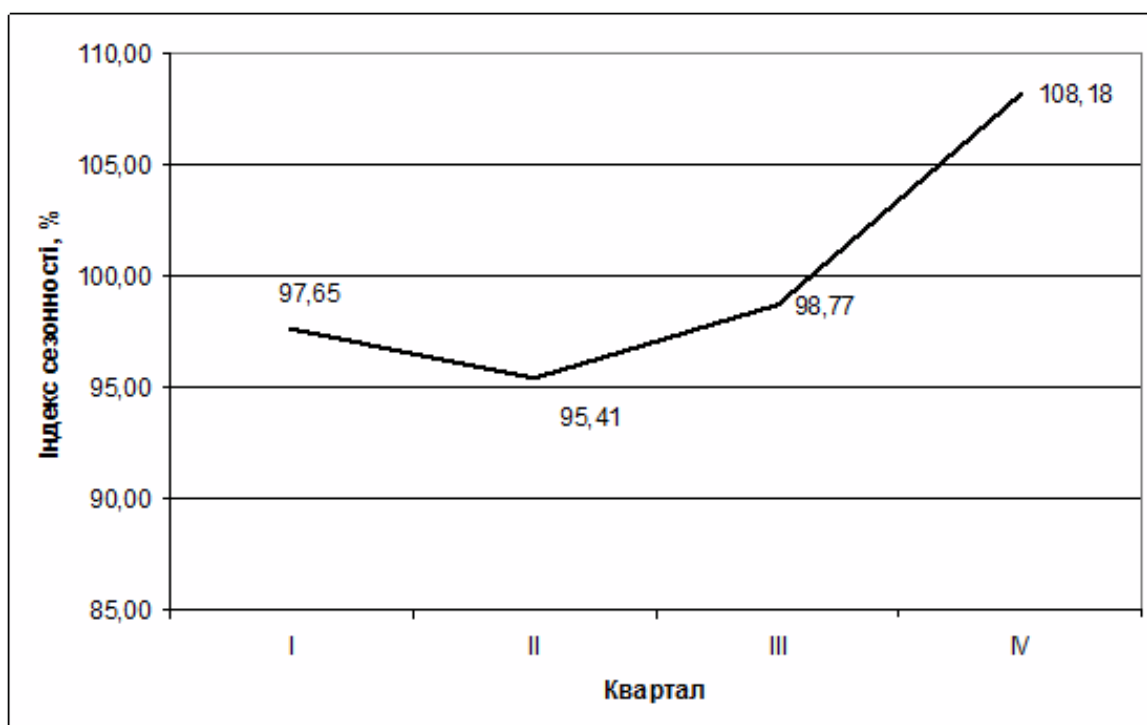
To calculate seasonality indices, we calculate the average values for each quarter as well as the overall average, after which we find the relation (Fig. 4.17).

	A	B	C	D	E	F	G	H	I	J	
1		Роки				Квартальні	Індекс				
2	Квартал	2006	2007	2008	2009	середні	сезонності				
3	I	669,4	907,2	934,1	942,8	863,4	97,65	=F3/\$F\$7*100			
4	II	661,4	881,9	885,1	946	843,6	95,41	=F4/\$F\$7*100			
5	III	756	878,1	878,2	980,8	873,3	98,77	=F5/\$F\$7*100			
6	IV	866,4	940	995,8	1023,9	956,5	108,18	=F6/\$F\$7*100			
7	Середні за роками	738,3	901,8	923,3	973,4	884,2					
8											
9		=СРЗНАЧ(B3:B6)				=СРЗНАЧ(F3:F6)		=СРЗНАЧ(B3:E3)			
10											
11											
12											
13											
14											

Fig. 4.17. The results of the calculation of seasonality indices

Thus, according to the results of the calculations, we obtain the following seasonality indexes of revenues from the provision of urban telephone services: 1st quarter – 97.65 %; 2nd quarter – 95.41 %; 3rd quarter – 98.77 %; 4th quarter – 108.18 %.

The seasonal wave graph constructed on the basis of the calculations confirms the above conclusions (Fig. 4.18).



**Fig. 4.18. The seasonal wave of the revenue from the provision of urban telephone services, 2006 – 2009**

It can be concluded that seasonal variations are quite strong. To visualize seasonal fluctuations, use a graph of the seasonal wave, in which the abscissa shows the time (period) and the ordinates show the seasonality indices.

## **Laboratory work 5**

### **The indicators of the study of world trends**

*The purpose* of the work is to acquire the skills in statistical data analysis by means of the index method in the MS Excel package.

*The task* is to analyze the data by the index method using MS Excel.

### **Guidelines**

**Task 1.** We have the data on the manufacture of products in the base and reporting years (Table 5.1), which will be placed in cells A1:E5.

Table 5.1

## Data for analysis

Company number	Products A			
	The base year		The reporting year	
	Price, UAH	Volume, items	Price, UAH	Volume, items
1	10.5	220	10.0	260
2	6.4	730	6.5	800
3	6.6	340	7.4	820

Then, using the formulas above, we calculate the individual price indices (F3:F5 cells), the volume (physical volume) of the manufactured products (G3:G5 cells) and the cost (H3:H5 cells).

1	Номер підприємства	Базисний період		Звітний період		Індивідуальні індекси			2010	2011	2011
		ціна, грн.	кількість, шт.	ціна, грн.	кількість, шт.	ціни	кількості	вартості			
3	1	10,5	220	10,0	260	0,952	1,182	1,126	2310	2730	2600
4	2	6,4	730	6,5	800	1,016	1,096	1,113	4672	5120	5200
5	3	6,6	340	7,4	820	1,121	2,412	2,704	2244	5412	6068
6	Разом	x	x	x	x	x	x	x	9226	13262	13868

Formulas shown in the image:

- $=D5/B5$  (Price Index)
- $=E5/C5$  (Volume Index)
- $=(D5*E5)/(B5*C5)$  (Cost Index)
- $=B5*C5$  (Base Cost)
- $=D5*E5$  (Reporting Cost)
- $=СУММ(K3:K5)$  (Sum of values in column K)
- $=СУММ(J3:J5)$  (Sum of values in column J)
- $=СУММ(I3:I5)$  (Sum of values in column I)

Fig. 5.1. Calculation of individual indices

The following conclusions can be drawn from the calculation of individual indices. In the reporting period, the volume of products increased by 18.2 % in company 1, by 9.6 % in company 2, and by 141.2 % in company 3. Compared to the base period, unit price decreased by 4.8 % in company 1, and increased by 1.6 % in company 2, and by 12.1 % in company 3. The cost of products in the reporting period increased by 12.6 % in company 1, by 11.3 % in company 2, and by 170.4 % in company 3.

**Task 2.** We have the data on the sales of products (Table 5.2).

Table 5.2

**The data on the sales of products**

Production	The base period		The reporting period	
	The volume of products, items	Price per unit, UAH	The volume 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 price changes. Draw a conclusion.

To determine the overall index of the turnover, enter the formula = D3 × 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 "=" and selecting the *SUM* formula in the formula row, then enter the formula = B3 × C3 into the cell G3 and apply the formula to the whole column and calculate the volume in the cell G5 (Fig. 5.2).

	A	B	C	D	E	F	G	H
1		The base period		The reporting period				
2	Production	The quantity of products, items	Price per unit, UAH	The quantity of products, items	Price per unit, UAH	q1p1	q0p0	lpq
3	A	350	200	420	225	94,500	70,000	
4	B	170	120	110	150	16,500	20,400	
5						111,000	90,400	1.23

**Fig. 5.2. 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 the product prices, and also due to the changes in the number of products sold.

To determine the general price index, enter the formula = D3 × 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 "=" and selecting the SUM formula in the formula row, enter the formula = D3 × C3 into the cell G3 and apply the formula to the whole column, calculate the sum in the cell G5.

To get the value of the general price index, enter the formula = F5/G5 into the cell H5 (Fig. 5.3).

	A	B	C	D	E	F	G	H
9	Production	The quantity of products, items	Price per unit, UAH	The quantity of products, items	Price per unit, UAH	q1p1	q1p0	Ip
10	A	350	200	420	225	94,500	84,000	
11	B	170	120	110	150	16,500	13,200	
12						111,000	97,200	1.14

**Fig. 5.3. 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 cost overruns or savings on the price changes are calculated as the difference between the numerator and denominator of the price index in the formula = F5 – G5 in the cell H5 (Fig. 5.4).

	A	B	C	D	E	F	G	H	I
9	Production	The quantity of products, items	Price per unit, UAH	The quantity of products, items	Price per unit, UAH	q1p1	q1p0	Ip	Δp
10	A	350	200	420	225	94,500	84,000		
11	B	170	120	110	150	16,500	13,200		
12						111,000	97,200	1.14	13,800

**Fig. 5.4. The calculation of the absolute amount of cost overruns or savings on the price changes**

The absolute amount of overruns 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 = D3 × C3 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 "=" and selecting the *SUM* formula in the formula bar, in the cell G3, enter the formula = B3 × C3 and apply the formula to the whole column, then, in the cell G5, calculate

the sum. To get the value of the overall index of the physical volume of trade, enter the formula = F5/G5 into the cell H5 (Fig. 5.5).

	A	B	C	D	E	F	G	H
15	Production	The base period		The reporting period		q1p0	q0p0	Iq
16		The quantity of products, items	Price per unit, UAH	The quantity of products, items	Price per unit, UAH			
17	A	350	200	420	225	84,000	70,000	
18	B	170	120	110	150	13,200	20,400	
19						97,200	90,400	1.08

**Fig. 5.5. The calculation of the general index of the physical volume of the commodity turnover**

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.

**Task 3.** We have the data on the cost of production and changes in the cost of the enterprise (presented in Table 5.3).

Table 5.3

**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 %. Draw conclusions.

In the cell D2, the prime cost index is calculated individually using the formula = (C2 100/100), in the second column (the cost of production in the 2nd quarter, thousand UAH), the amount is calculated by clicking the *AutoSum*  $\Sigma$  sign on the toolbar after selecting the entire column. Enter the formula = B2/D2 into the cell E3, calculate the amount of the column

in the cell E5, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or by pressing the "=" and selecting the *SUM* formula in the formula bar. To get the value of the total cost of the index, enter the formula = B5/E5 into the cell F5 (Fig. 5.6).

	A	B	C	D	E	F
22	<b>Task 2</b>					
	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, %	$i_z$	$\frac{z_1 * q_1}{i_z}$	$I_z$
23						
24	A	1,850	18	1.18	1,567.80	
25	B	670	11	1.11	603.60	
26	C	755	No change	1	755.00	
27		3,275			2,926.40	1.12

Fig. 5.6. 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, by increasing the cost and at the constant output.

Based on the conditions of the task, the overall cost index ( $I_{zd}$ ) 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 the volume of output. For the calculation of the overall index of production using the relationship between the indexes, enter the formula = G5/F5 into the cell H5 (Fig. 5.7).

	A	B	C	D	E	F	G	H
22	<b>Task 2</b>							
	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, %	$i_z$	$\frac{z_1 * q_1}{i_z}$	$I_z$	$I_{zq}$	$I_q$
23								
24	A	1,850	18	1.18	1,567.80			
25	B	670	11	1.11	603.60			
26	C	755	No change	1	755.00			
27		3,275			2,926.40	1.12	1.27	1.13

Fig. 5.7. The calculation of the overall index of production

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



**Task 4.** We have the data on the manufacture of products of the same name "A" and its cost at two plants (Table 5.4).

Table 5.4

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

Plant	Production, thousand units		The unit cost of production, UAH	
	year		year	
	2007	2008	2007	2008
1	90	100	210	205
2	180	110	180	240

Determine the cost indices: variable composition, constant composition, structural changes. Explain the difference between these indices.

To determine the index of the cost of variable composition, enter the formula = E5 × C5 into the cell F5 and apply the formula to the whole column, calculate the amount of the column in the cell F7, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or by pressing the "=" and selecting the *SUM* formula in the formula bar, enter the formula = F7/S7 into the cell G7 get the value of the first fraction, enter the formula = D5 × C5 into the cell H5 and apply the formula to the whole column, count the amount of the column in the cell H7, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or pressing the "=" and selecting the *SUM* formula in the formula bar, enter the formula = H7/S7 into the cell I7 to get the value of the second fraction.

To get the index value of the variable cost, enter the formula = G7/I7 into the cell J7 (Fig. 5.8).

	A	B	C	D	E	F	G	H	I	J
30	<b>Task 3</b>									
31	Plant	Production, thousand units		The unit cost of production, UAH		z1q1	$\frac{\sum z_1 * q_1}{\sum q_1}$	z0q1	$\frac{\sum z_0 * q_1}{\sum q_1}$	variable
32		year		year						
33		2007	2008	2007	2008					
34	1	90	100	210	205	20,500		21,000		
35	2	180	110	180	240	26,400		19,800		
36		270	210			46,900	223.33	40,800	194.29	1.15

**Fig. 5.8. The calculation of the index of the prime cost of variable composition**

To get the index of the prime cost of variable composition, enter the formula = G7/I7 into the cell J7 (Fig. 5.8).

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

To determine the index of the prime cost of the fixed structure, enter the formula = E4 × C4 into the cell F4 and apply the formula to the whole column, calculate the amount of the column in the cell F6, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or by pressing the "=" and selecting the *SUM* formula in the formula bar, enter the formula = F6/S6 into the cell G6 and get the value of the first fraction, enter the formula = B4 × D4 into the cell H4 and apply the formula to the whole column, calculate the amount of the column into the cell H6, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or pressing the "=" and selecting the *SUM* formula in the formula bar, enter the formula = H6/B6 into the cell I6 and get the value of the second fraction.

To get the index of the cost of fixed composition, enter the formula = G6/I6 into the cell J6 (Fig. 5.9).

	A	B	C	D	E	F	G	H	I	J
39	Plant	Production, thousand units		The unit cost of production, UAH		z1q1	$\frac{\sum z_1 * q_1}{\sum q_1}$	z0q0	$\frac{\sum z_0 * q_0}{\sum q_0}$	ifix
40		year		year						
41		2007	2008	2007	2008					
42	1	90	100	210	205	20,500		18,900		
43	2	180	110	180	240	26,400		32,400		
44		270	210			46,900	223.33	51,300	190.00	1.18

**Fig. 5.9. The calculation of the index of the prime cost of the fixed structure**

The index of the cost of fixed composition was 1.18 (118 %), that is, the average cost of production increased by 18 % due to the changes in unit costs at the constant amount of the output.

To determine the index of the cost of structural changes, enter the formula = D4 × C4 into the cell F4 and apply the formula to the whole column, calculate the amount of the column in the cell F6, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or by pressing the "=" and selecting the *SUM* formula in the formula bar, enter the formula = F6 × C6 into the cell G6

to obtain the value of the first fraction, enter the formula = B4 × D4 into the cell H4 and apply the formula to the whole column, calculate the amount of the column in the cell H6, either by clicking the sign  $\Sigma$  *AutoSum* button on the toolbar, or pressing the "=" and selecting the *SUM* formula in the formula bar, enter the formula = H6/B6 into the cell I6 and get the value of the second fraction.

To get the index of the cost of structural changes, enter the formula = G6/I6 into the cell J6 (Fig. 5.10).

	A	B	C	D	E	F	G	H	I	J
47	Plant	Production, thousand units		The unit cost of production, UAH		z0q1	$\frac{\sum z_0 * q_1}{\sum q_1}$	z0q0	$\frac{\sum z_0 * q_0}{\sum q_0}$	lstr. change
48		year		year						
49		2007	2008	2007	2008					
50	1	90	100	210	205	21,000		18,900		
51	2	180	110	180	240	19,800		32,400		
52		270	210			40,800	194.29	51,300	190.00	1.02

**Fig. 5.10. The calculation of the index of the cost of structural changes**

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

## Bibliography

### Main

1. Андрієнко В. Ю. Статистичні індекси в економічних дослідженнях / В. Ю. Андрієнко. – Київ : 2004. – 536 с.
2. Ефимова М. Р. Общая теория статистики : учебник / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцева. – 2-е изд., испр. и доп. – Москва : ИНФРА-М, 2007. – 417 с.
3. Лугін О. Є. Статистика : підручник. – 2-е видання, перероблене та доповнене / О. Є. Лугін. – Київ : Центр учбової літератури, 2007. – 608 с.
4. Сизова Т. М. Статистика : учебное пособие / Т. М. Сизова. – Санкт-Петербург : Изд-во ИТМО, 2005. – 190 с.
5. Статистика : підручник / С. С. Герасименко, А. В. Головач, А. М. Єріна та ін. ; за наук. ред. д-ра екон. наук С. С. Герасименка. – 2-ге вид., перероб. і доп. – Київ : КНЕУ, 2000. – 467 с.
6. Статистика : навчальний посібник / під ред. д-ра. екон. наук професора О. В. Раєвневої. – Харків : Вид-во ХНЕУ, 2010. – 520 с.
7. Статистическое моделирование и прогнозирование : учебное пособие / под. ред. А. Г. Гранберга. – Москва : Финансы и статистика, 2000. – 384 с.

### Additional

8. Дегтярев Д. А. Количественные методы в международных исследованиях / Д. А. Дегтярев // Международные процессы. – 2015. – Том 13, № 2. – С. 35–54.
9. Єріна А. М. Статистичне моделювання та прогнозування : навч. посіб. / А. М. Єріна. – Київ : КНЕУ, 2001. – 170 с.
10. Халафян А. А. STATISTICA 6. Статистический анализ данных / А. А. Халафян. – Москва : ООО "Бином-Пресс", 2008. – 512 с.

### Information resources

11. Офіційний сайт Всесвітнього Банку [Електронний ресурс]. – Режим доступу : <http://www.worldbank.org>.

12. Офіційний сайт департаменту статистики Організації Об'єднаних Націй [Електронний ресурс]. – Режим доступу : <http://unstats.un.org/unsd/default.htm>.
13. Офіційний сайт Європейського центрального банку [Електронний ресурс]. – Режим доступу : <https://www.ecb.europa.eu/stats/html/index.en.html>.
14. Офіційний сайт Міждержавного статистичного комітету об'єднання незалежних держав [Електронний ресурс]. – Режим доступу : <http://www.cisstat.com>.
15. Fernandes Marcelo. Statistics for Business and Economics [Electronic resource] / M. Fernandes. – 2009. – 150 p. – Access mode : [http://russell-davidson.arts.mcgill.ca/e257/Marcelo\\_Fernandes.pdf](http://russell-davidson.arts.mcgill.ca/e257/Marcelo_Fernandes.pdf).
16. Owen G. Scott. History of Visualization [Electronic resource] / G. S. Owen. – 2010. – 97 p. – Access mode : [https://www.researchgate.net/publication/40801120\\_Visualizing\\_Historical\\_Narratives\\_Geographically-Integrated\\_History\\_and\\_Dynamics\\_GIS](https://www.researchgate.net/publication/40801120_Visualizing_Historical_Narratives_Geographically-Integrated_History_and_Dynamics_GIS).

## Content

Introduction.....	3
Laboratory work 1. Formation of a representative sample .....	5
Laboratory work 2. Tools for visualization of statistical data.....	23
Laboratory work 3. Grouping of data and research on groups .....	27
Laboratory work 4. Statistical tools for studying the dynamics of phenomena and processes.....	33
Laboratory work 5. The indicators of the study of world trends .....	43

НАВЧАЛЬНЕ ВИДАННЯ

# ОСНОВИ АНАЛІЗУ ДАНИХ

**Методичні рекомендації  
до лабораторних робіт  
для студентів спеціальності  
122 "Комп'ютерні науки"  
другого (магістерського) рівня**

*Самостійне електронне текстове мережеве видання*

Укладачі: **Раєвська** Олена Валентинівна  
**Бровко** Ольга Іванівна

Відповідальний за видання *О. В. Раєвська*

Редактор *З. В. Зобова*

Коректор *З. В. Зобова*

Подано завдання для лабораторних робіт і методичні рекомендації до їх виконання, метою яких є закріплення й поглиблення знань теоретичного та практичного матеріалу, набуття навичок аналізу економічних явищ та процесів за допомогою MS Excel.

Рекомендовано для студентів спеціальності 122 "Комп'ютерні науки" другого (магістерського) рівня.

План 2020 р. Поз. № 113 ЕВ. Обсяг 55 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.*