GUIDELINES FOR PRACTICAL LESSONS

Course:	Introduction to Scientific Research
Course code de	ISR 2212
Educational program:	6B10115 «Medicine»
Number of academic hours/credits:	180/6
Year/Term:	2/4
Practical lessons	48 hours

Shymkent, 2024 y.

Guidelines for practical lessons were developed in accordance with the working curriculum of the discipline (syllabus) "Introduction to Scientific Research" and discussed at a meeting of the departments:

Medical biophysics and information technology

Protocol no. 11 from "30" 05 2024 g.

Head of department, PhD, ass. prof, _____ M.B. Ivanova

Social health insurance and public health

Protocol no. 15 from "10" 06 202__ g.

Head of department, PhD, ass. prof. _____ G.Zh. Sarsenbayeva

Lesson №1

1. **The theme:** Introduction to Biostatistics
2. **The aim:** formation in students of a basic understanding of the discipline "Biostatistics", its subject, tasks and stages of development; familiarizing students with the types of medical data, as well as the types of measurement scales.
3. **The learning objectives:**
   - understand the role of biostatistics in medical education and the work of a practicing physician;
   - get acquainted with the main stages of the development of biostatistics;
   - learn to classify medical data;
   - learn to distinguish between measurement scales used in medical research.
4. **Key questions of the theme:**
   1. Subject and tasks of biostatistics.
   2. Stages of development of biostatistics.
   3. Classification of medical data.
   4. The main types of measurement scales used in biomedical research.
5. **Methods / technologies of teaching and learning:** Practice

❖ **Tasks:**

1. Learn the theory:

1.1. Subject and tasks of biostatistics

*Statistics*is a social science that studies the quantitative side of mass social phenomena.

It appeared as an independent science at the end of the 17th century.

Biostatistics- statistics that studies issues related to biology, medicine, pharmacy, hygiene and health care.

The science of bostatistics was formed in the middle of the 19th century.

The role of biostatistics in the practical and scientific work of a healthcare manager, doctor, epidemiologist, nurse, and pharmacist is great.

Biostatistics uses various methods: collecting data, summarizing it, analyzing and drawing conclusions based on the observations obtained.

Statistical analysis helps extract information from data and evaluate the quality of that information.

Objectives of biostatistics:

- quantitative representation of biological facts (measurement);
- generalized description of a set of facts (statistical assessment);
- search for patterns (testing statistical hypotheses).

1.2. The role of scientists F. Galton, K. Pearson, R. Fisher in the development of biometrics

The founder of biometrics is considered to be the English scientist Francis Galton (1822-1911) (Figure 1.1, a).

A contemporary of F. Galton, Russian scientist K.A. Timiryazev said the following about him: "He was one of the most original scientists, researchers and thinkers of modern England."

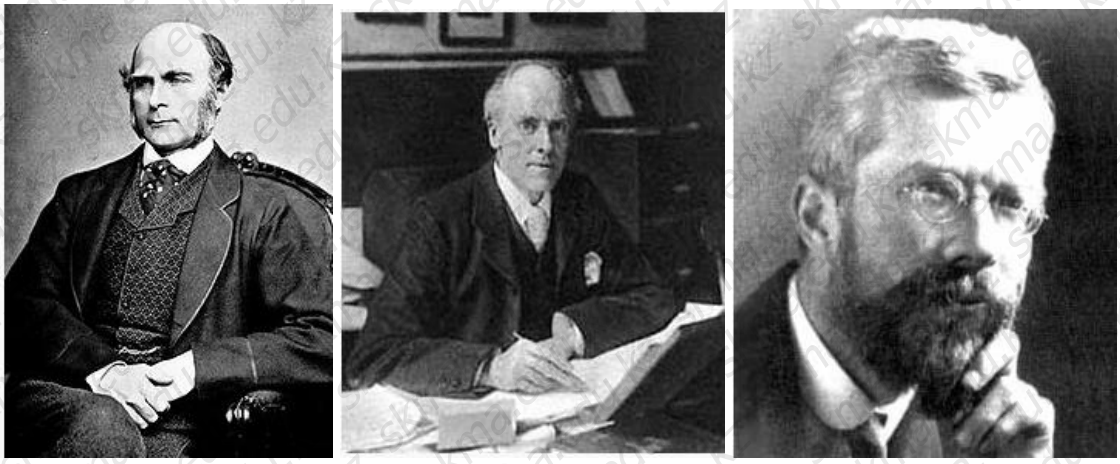
Indeed, F. Galton was a very versatile person; he was interested in natural science, anthropology, heredity, psychology, the theory of evolution, meteorology and statistics.

In one of his books on heredity, the term biometry was first introduced. F. Galton's important achievements include the development of the foundations of correlation analysis.

A follower of F. Galton is Karl Pearson (1857-1936) (Figure 1.1,b) - English mathematician, biologist and philosopher. Author of more than 400 scientific papers on mathematical statistics.

He introduced the concept of standard deviation and coefficient of variation, developed the mathematical apparatus of the theory of conjugation of features, nonlinear correlation and regression,

goodness-of-fit criteria, decision-making algorithms and parameter estimation.



a b c

Figure 1.1. a – F. Galton, b – K. Pearson, c – R. Fisher

Together with the English zoologist and biometrician W. Weldon, K. Pearson founded the journal Biometrics, dedicated to the application of mathematical and statistical methods in biology.

The successor and continuer of K. Pearson's work on biostatistics was Ronald Fisher (1890-1962) (Figure 1.1, c).

Danish statistician A. Hald described R. Fischer as "a genius who almost single-handedly laid the foundations of modern statistics," and English ethologist and biologist R. Dawkins called him "the greatest biologist like Darwin."

R. Fisher is the founder of the theory of sampling distributions, methods of variance and discriminant analysis, the theory of experimental design, the maximum likelihood method and much more, which forms the foundation of modern applied statistics and mathematical genetics.

1.3. Collection, classification and presentation of data

Regardless of what tasks are set when conducting scientific medical and biological research, it must be carried out step by step, in a certain sequence.

At the first stage, a research plan and program is drawn up, at the second,collection of statistical material, on the third - the analysis of the received data is carried out, on the fourth - the processing of the collected material takes place.

The stage of collecting statistical data is very important, because the correctness of all further research depends on the correctness of the collected data.

The process of obtaining information about the objects of the population under study and their properties is called collecting statistical data. This data is the subject of statistical processing and analysis.

At this stage, it is important to determine the type of data being considered.

Statistical data is divided into the following types: quantitative, qualitative and dates (Fig.1.2).

*Quantitative data*are divided into two categories: discrete and continuous.

Discrete data– quantitative data that is presented as integers. For example, the number of children in the family, the number of ambulance calls per hour, heart rate, etc.

Continuous Data– quantitative data obtained by measurement on a continuous scale. For example, body weight, height, blood pressure, etc.

Continuous data can be interval or relative.

Interval data– continuous data that is measured in absolute values that have a physical meaning.

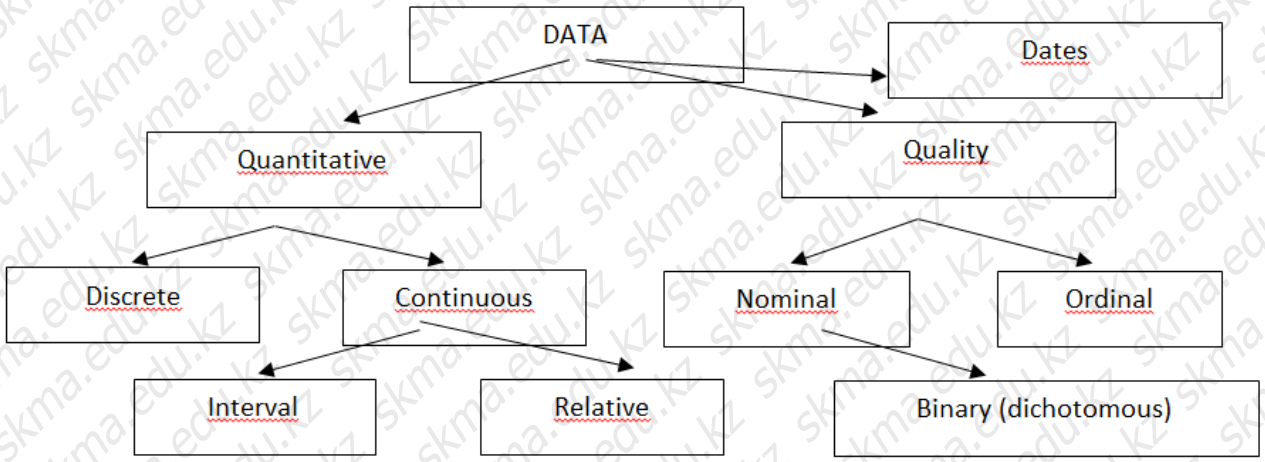


Figure 1.2. Types of statistics

Relative data– continuous data that reflects the proportion of change (increase or decrease) in the value of a characteristic in relation to the original (or some other) value of this characteristic. These data are dimensionless quantities or expressed as percentages.

*Qualitative data*are divided into nominal and ordinal.

Nominal data –qualitative data that reflects dummy codes for unmeasured categories. For example, diagnosis code, gender, marital status, nationality, etc.

Ordinal data- qualitative data that reflects the conditional degree of expression of any characteristic. For example, stages of cancer, degrees of heart failure.

Their main difference from discrete quantitative data is the absence of a proportional scale for measuring the severity of a trait.

Binary (or dichotomous) data -qualitative data that has only two possible meanings. For example, gender, the presence or absence of a disease, etc.

Dates- a special type of data, in some cases it may be necessary to perform some arithmetic operations with them, for example, to calculate the period of a patient's stay in the hospital.

1.4. Main types of measurement scales used in biomedical research

Measurement– this is a procedure for comparing objects according to certain indicators or characteristics (features).

Scale– a necessary, mandatory element of the measurement procedure.

The main types of measurement scales used in biomedical research include:

- *Nominal scale*- used to classify the properties of an object, assign them numerical, alphabetic and other symbolic characteristics (gender, nationality, diagnosis, etc.) (Figure 1.3 shows a fragment of the electronic international classifier of diseases, where each type of disease is assigned a special code);

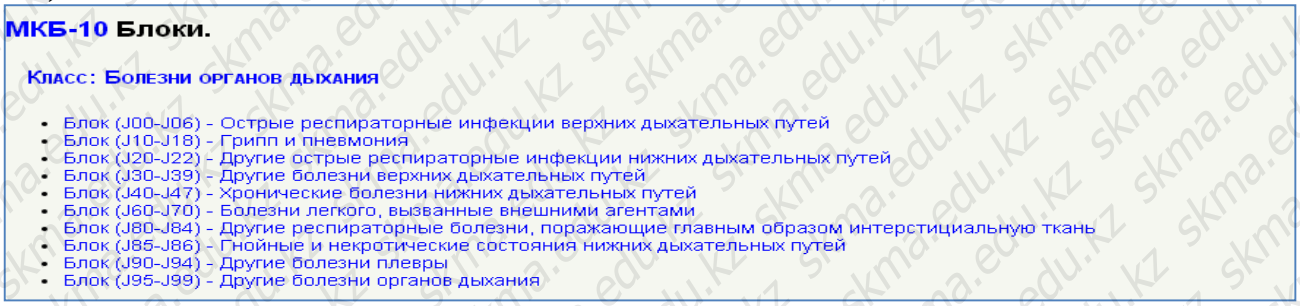


Figure 1.3. International Classification of Diseases - example of a naming scale

- *ordinal*or ranking scale – orders the values of a characteristic (scale of stages of hypertension according to Myasnikov, scale of degrees of heart failure according to Strazhesko-Vasilenko-Lang, scale of severity of coronary insufficiency according to Fogelson (Figure 1.4), etc.);

Stage	Signs
1	decreased myocardial contraction rate, decreased ejection fraction, shortness of breath, palpitations, fatigue during exercise
2	circulatory failure is expressed moderately or significantly. The signs of circulatory failure indicated for the initial stage are detected not only during physical activity, but also at rest.
3	significant disturbances in cardiac activity and hemodynamics at rest, as well as the development of significant dystrophic and structural changes in organs and tissues

Figure 1.4. Fogelson coronary insufficiency severity scale - an example of an ordinal scale

- *interval scale*– shows the "range" of individual measurements of a characteristic (time, temperature scale, test scores, etc.) (Figure 1.5);

Figure 1.5. Stopwatch, thermometer - examples of interval scale

- *ratio scale*– reveals the relationship between the measured values of a trait (height, weight, reaction time, number of completed test tasks, etc.) (Figure 1.6).

Figure 1.6. Scales - an example of a ratio scale

1.5. Reliability and validity of measurements in biostatistics

During the measurement process, the question arises: *reliability* And *reliability*.

Reliability measurements depends on:

- correctness (is the scale chosen correctly, are the readings recorded correctly, are systematic errors taken into account, etc.);
- stability (do the results coincidewith repeated measurements);
- validity (It is the specified property of the object that is measured, and not another that is similar to it).

Credibility measurement characterizes the accuracy of measurements of a quantity in relation to what exists in reality.

The main direction of checking the reliability of measurements is to obtain information from various sources.

2. Oral survey on the topic.
3. Take the test (completing test tasks using the "easyQuizzy" program)

6. **Assessment methods / technologies:** Oral questioning. Practical work. (assessment using a checklist).
7. **Literature:**

- Main:

1. Rosner Bernard Fundamentals of Biostatistics: Texbook/ B.Rosner - 8th ed.: GENGAGE learning, 2016.
2. Armitage P. Encyclopedia of Biostatistics. - Wiley, 2016. - 6100 p
3. Le C.T. Introductory biostatistics. - Wiley, 2013. - 536 p
4. Newman S.C. Biostatistical Methods in Epidemiology. - Wiley, 2017. - 388p
5. http://www.biometrica.tomsk.ru

8. **Control:**
1. What is "biostatistics"?
2. What is the role of scientists F. Galton, K. Pearson, R. Fisher in the development of biometrics?
3. What types of statistics do you know?
4. What types of measurement scales are used in biomedical research?

Lesson №2

1. **The theme:** Variation series.
1. **The aim:** developing skills in calculating the numerical characteristics of a variation series and their interpretation.familiarization with some of the capabilities of the STATISTICA program
2. **The learning objectives:**

- learn to calculate the numerical characteristics of a variation series;
- learn to interpret the numerical characteristics of a variation series
-learn to create, edit, save source data tables;
- learn to build simple graphs in the STATISTICA program.

4. **Key questions of the theme:**
1. Measures of the central tendency of a variation series.
2. Measures of the spread of the variation series.
3. Interpretation of the numerical characteristics of the variation series.
4. Stages of statistical analysis in the "STATISTICA" program.
5. Types of documents created in the STATISTICA program. Their extensions.
6. Elements of the working window of the "STATISTICA" program.
7. Operations performed on columns and rows of a spreadsheet.
8. Construction of the simplest graphs using the "STATISTICA" program.

5. **Methods / technologies of teaching and learning:** Computer-based work / Solving situational tasks.

Example 1. The study of a new highly effective antibiotic involved patients in the purulent surgery department who were prone to obesity. The following distribution of patients by weight was obtained:

Body weight, kg	90	100	120	130	140	Total
Number of patients, people	1	4	8	6	2	21

Calculate the indicators of the variation series: mean, mode, median, dispersion, standard deviation, coefficient of variation, range of variation, 25th and 75th percentiles. Does the presented variation series correspond to the law of normal distribution?

Solution.

Create a calculation table.

xi	*vi*	*xi·vi*	*xi-* $\bar{x}$	*(xi-* $\bar{x}$ *)2*	*(xi-* $\bar{x}$ *)2·vi*
90	1	90	-29.5	870.25	870.25
100	4	400	-19.5	380.25	1521
120	8	960	0.5	0.25	2
130	6	780	10.5	110.25	661.5
140	2	280	20.5	420.25	840.5
Total	21	2510			3895.25

1) Calculate the average:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i v_i}{\sum_{i=1}^{n} v_i} = \frac{2510}{21} \approx 119{,}5$$

those. average body weight of patients119.5 kg.

2) Determine the mode: Mo=120, i.e. the most common value found in observations is120 kg.

3) Determine the median: Me=120, i.e. meaning120 kgis in the middle of the variation series.

4) Calculate variance:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot v_i}{n-1} = \frac{3895{,}2}{20} \approx 194{,}76$$

5) Calculate standard deviation:

$$S = \sqrt{S^2} = \sqrt{194{,}76} \approx 13{,}96,$$

those. standard deviation of patient weight13.6 kg.

6) Calculate the coefficient of variation:

$$V = \frac{S}{\bar{x}} \cdot 100\% = \frac{13{,}96}{119{,}5} \cdot 100\% \approx 11{,}7\%$$

because V≤33%, then the sample is homogeneous.

7) Calculate the range of variation: $R = xmax - xmin = 140\text{-}90 = 50$

those. difference between the largest and smallest mass values50 kg.

8) Present the variation series in the form of an ordered sample: 90, 100, 100, 100, 100, 120, 120, 120, 120, 120, 120, 120, 120, 130, 130, 130, 130, 130, 130, 140, 140.

50th percentile number (median) $N_{P_{50}} = \frac{n+1}{2} = \frac{21+1}{2} = 11$ *Meh*=120.

25th percentile (lowest quartile) number $N_{P_{25}} = \frac{n+1}{4} = \frac{21+1}{4} = 5{,}5$ *P25*=110.

75th percentile (top quartile) number $N_{P_{75}} = \frac{3(n+1)}{4} = \frac{3(21+1)}{4} = 16{,}5$ *P75*=130.

Example 2. In the "Statistica" program, create a sample of 100 random numbers that have a

normal distribution with the parameters: μ=170, σ=7, which are the results of measuring the height of 1st year students at SKMA.

Calculate: mean, sum, median, geometric mean, harmonic mean, standard deviation, variance, standard error of the mean, confidence interval for the mean, skewness, kurtosis, largest and smallest sample value, lower and upper quartiles, range.

Construct a histogram and a normal probability plot, as well as a box-and-whisker plot. Explain the results.

1. *Create a spreadsheet.*

Open the “Statistica” program. A spreadsheet of size 10*10 will appear on the screen. If the table does not appear, then follow these steps: Home → File → New → Spreadsheet.

2. *Setting table sizes.*

The data requires 1 column and 100 rows. Remove extra 9 columns and add 90 rows.

3. *Table header design.*

Enter the title of the table “Growth of 1st year students at SKMA” in the white field under the line “Data: Spreadsheet1 (1v by 100c)”.

4. *Setting the variable name.*

Double-click the left mouse button on the variable name “Var1”. Open the variable specification window. In the “Name” field, write “Student height, cm”, click “OK”. If the name is partially visible, then stretch the column.

5. *Creation of a sample subject to the normal distribution law.*

While in the specification window for the variable “Var1”, in the “Long name” field, write the formula “=RndNormal(7)+170” (Figure 1), and click the “OK” button. The program will automatically fill the cells with numbers.

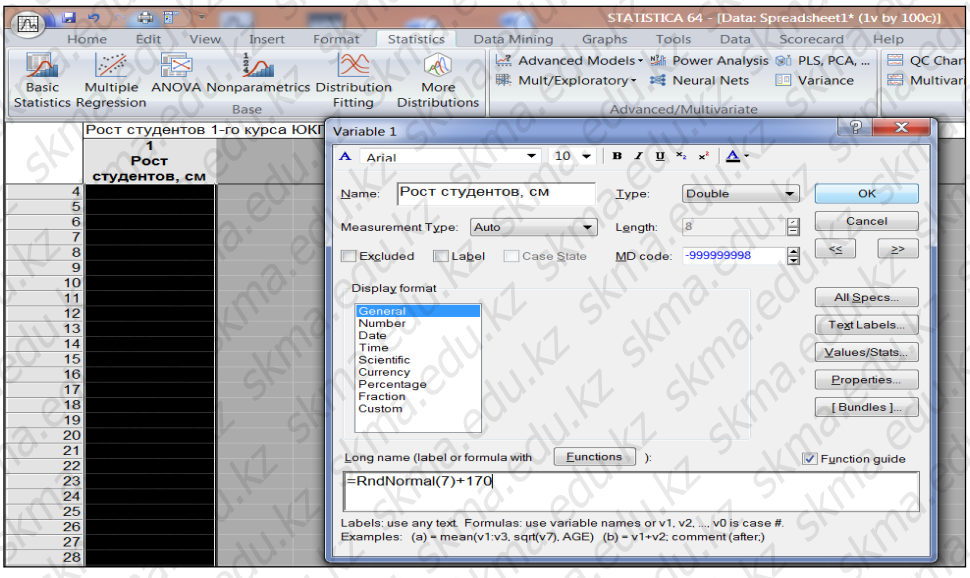


Figure 1. Creating a sample subject to the normal distribution law

6. *Changing the format of numeric data.*

In the specification window for the variable “Var1”, select “Number” in the “Display format” field, put “1” in the “Decimal places” field, and click the “OK” button.

7. *Calculation of numerical characteristics of the sample.*

Basic Statistics→ Descriptive statistics (Figure 2) → OK.

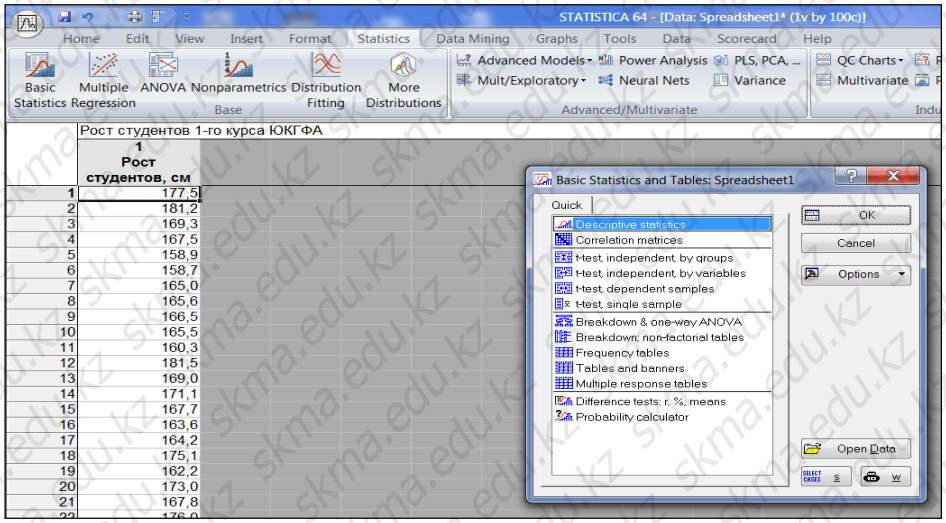


Figure 2. Selecting the Descriptive statistics procedure

Set the variable “Variables” (Figure 3).

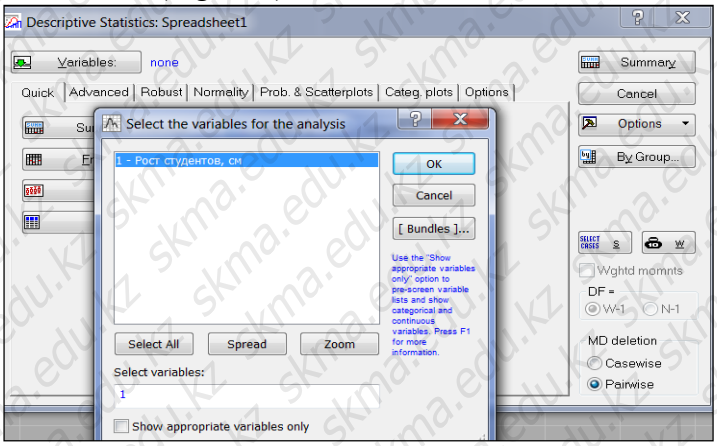


Figure 3. Setting a variable

Select the “Advanced” tab, mark the desired numerical characteristics:
Valid N - sample size;
Mean - average;
Sum - amount;
Median - median;
Geom. mean - geometric mean;
Harm. mean - harmonic mean;
Standard Deviation - standard deviation;
Variance - dispersion;
Std. err. of mean - standard error of the mean;
Conf. limits for means - confidence interval for the mean;
Skewness - asymmetry;
Kurtosis - excess;
Minimum & maximum - minimum and maximum;
Lower & upper quartiles - lower and upper quartiles;
Range - scope(Figure 4).
Click the “Summary” button.

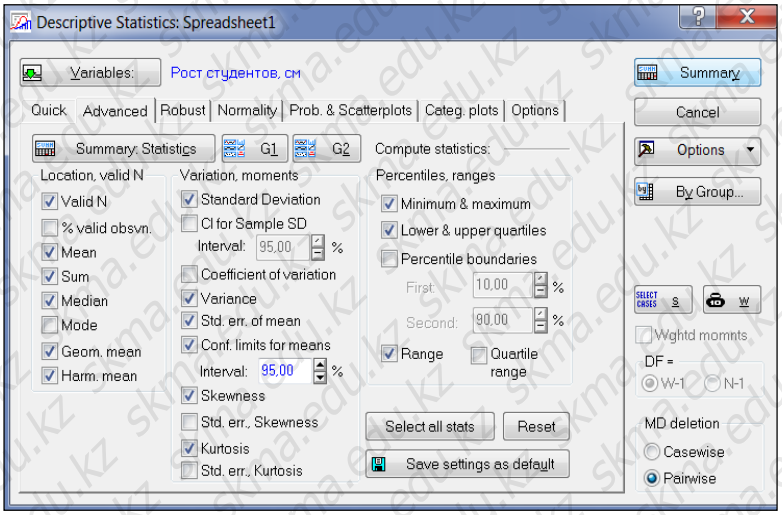


Figure 4. Selection of numerical characteristics

A summary table will appear on the screen (Figure 5).

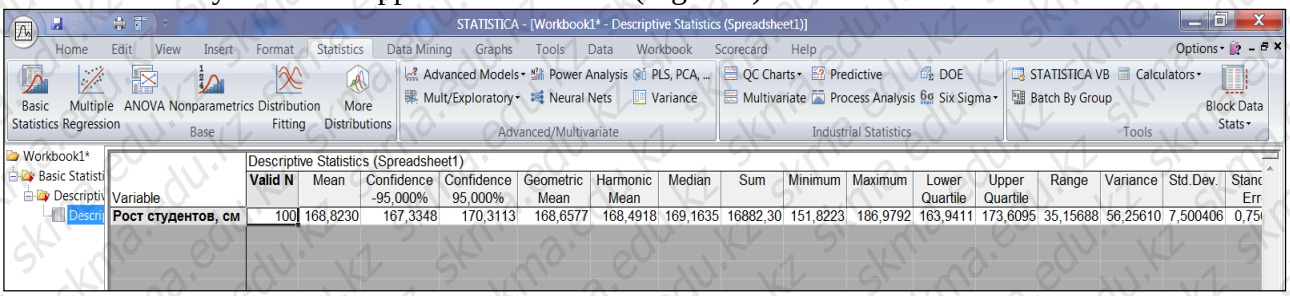


Figure 5. Final table with numerical characteristics of the sample

8. Constructing a histogram.

Return to the "Descriptive statistics" analysis window, select the "Normality" tab, click the "Histograms" button (Figure 6).

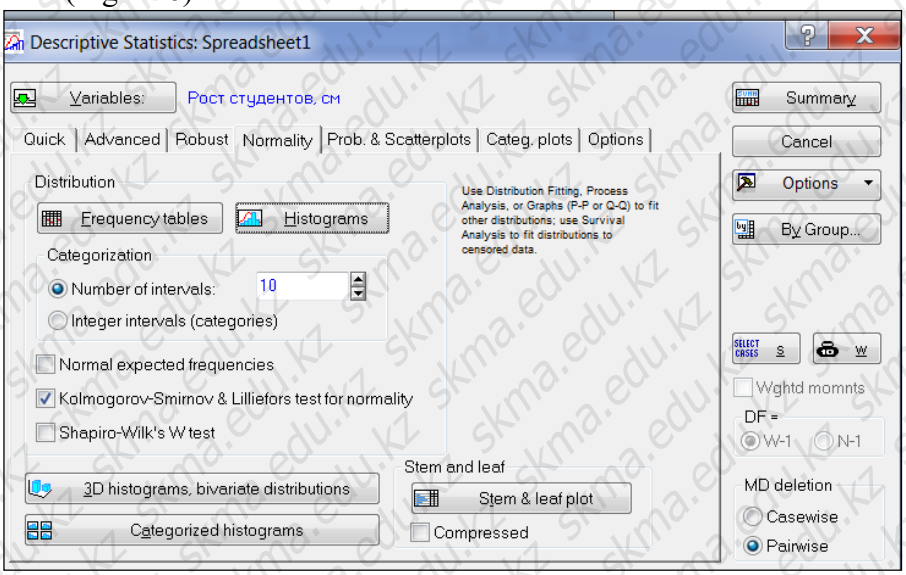


Figure 6. Plotting a histogram

A histogram will appear on the screen (Figure 7). The red line on the histogram is a graph of the density of a normal distribution.

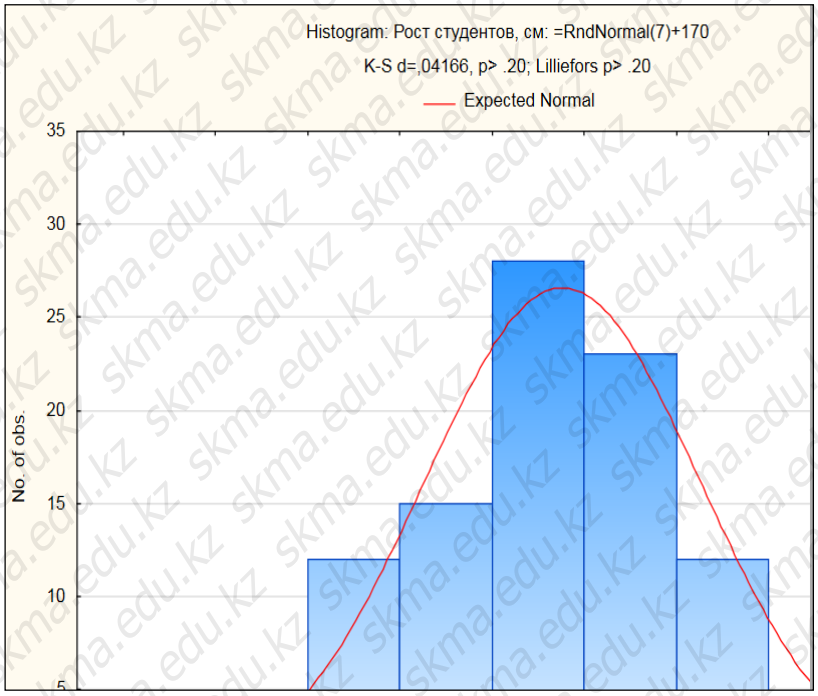


Figure 7. Histogram

9. Construction of a normal probability graph.

Return to the “Descriptive statistics” analysis window, select the “Prob. & Scatterplots” (Probability graphs and scatterplots), click the “Normal probability plot” button (Figure 8).

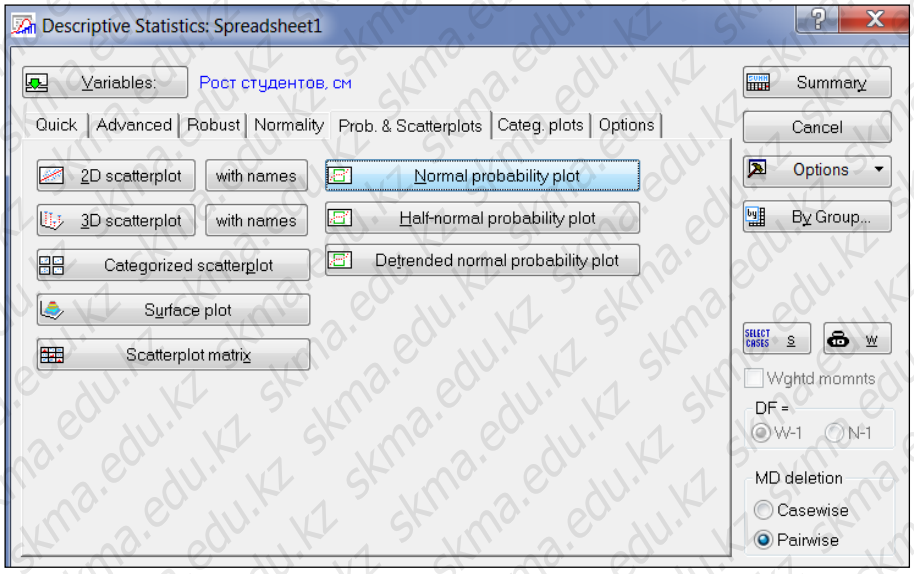


Figure 8. Construction of a normal probability graph

A normal probability graph will appear on the screen (Figure 9). The red line on the histogram is the density of the normal distribution, the blue dots are the sample observations. The closer the blue dots are to the red line, the more “normal” the distribution is.

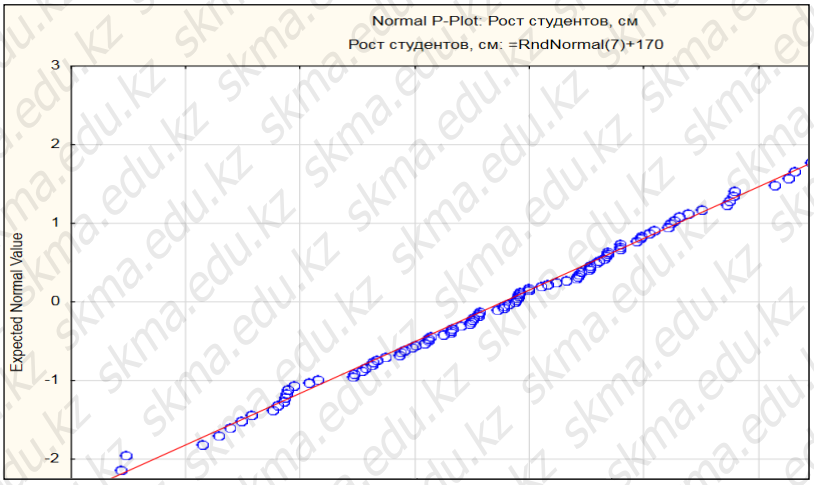


Figure 9. Normal probability plot

10. Constructing a box-and-whisker plot.

Return to the “Descriptive statistics” analysis window, select the “Quick” tab, click the “Box & whisker plot for all variables” button (Figure 10).

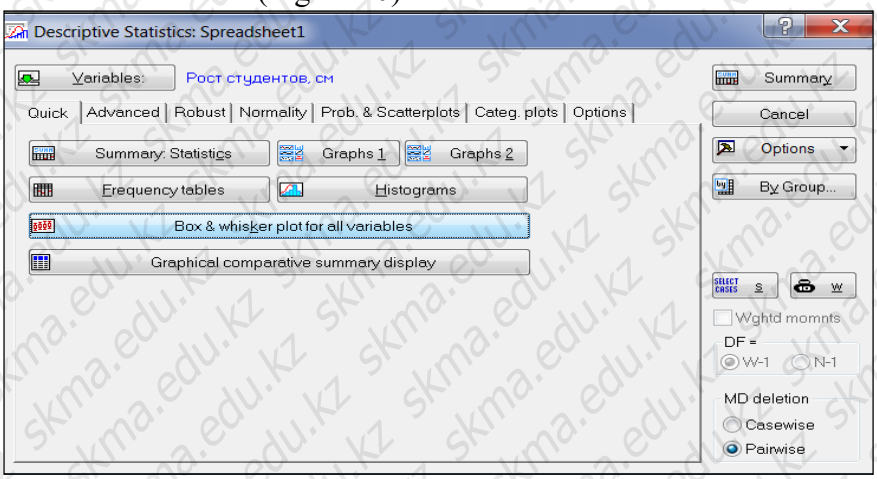


Figure 10. Plotting a box-and-whiskers plot

A box-and-whiskers plot will appear on the screen (Figure 11). The “legend” located at the bottom of the graph indicates: the mean and confidence intervals for the mean.

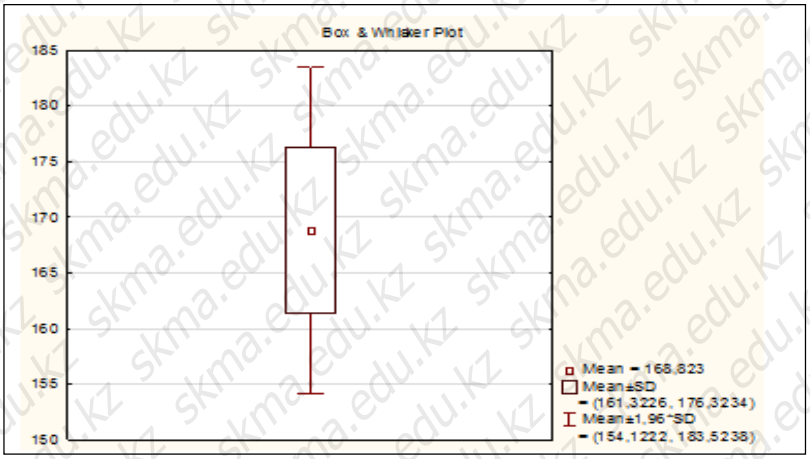


Figure 11. Box-and-whisker plot

If you want the "legend" of the graph to contain information about the median, lower and upper quartiles, as well as the range, then you need to select the "Options" tab and check "Median/Quartiles/Range" (Figure 12), click the "Sammary" button, and then plot the box-and-whiskers plot again (Figure 13).

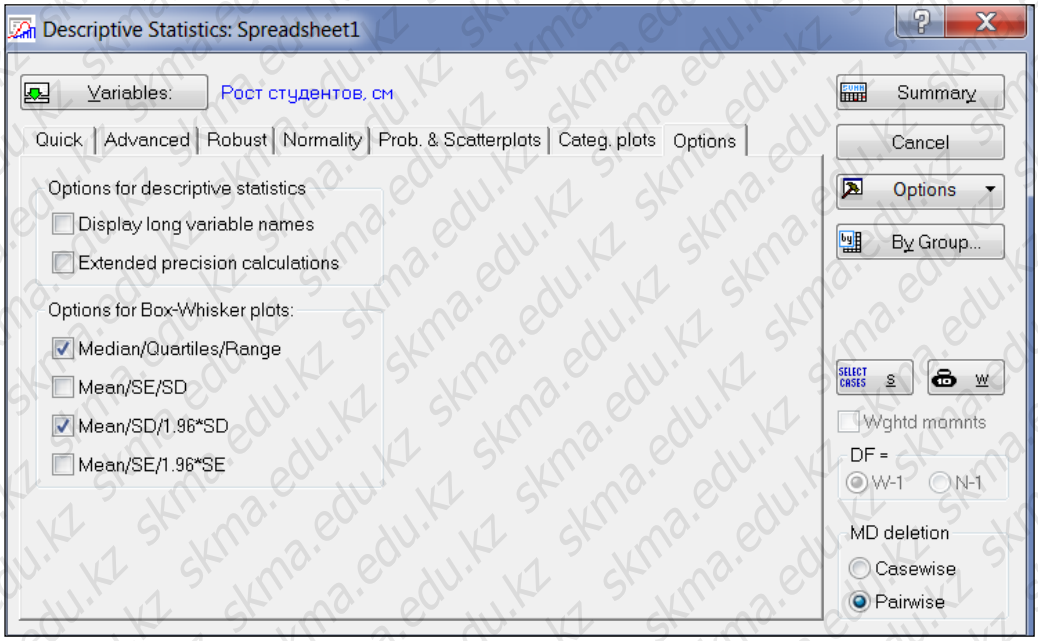


Figure 12. Selecting options for a box-and-whisker plot.

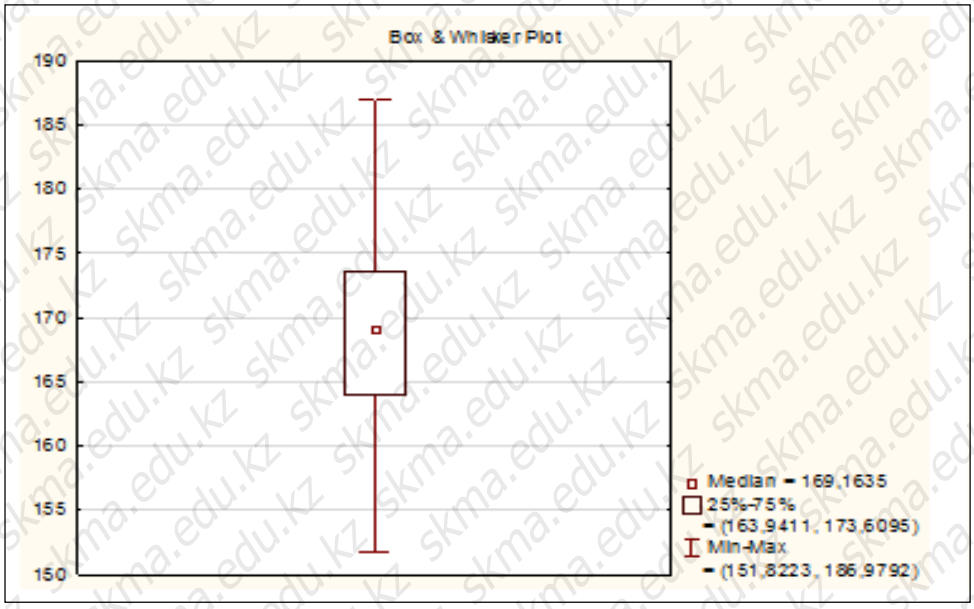


Figure 13. Box-and-whisker plot

11. Saving data.

Create a "Student" folder on your Desktop. In this folder, save the source data table under the name "Sample.sta": Home → File → Save → Save As

In the same folder, save the workbook under the name "Analysis results.stw": Home → File → Save → Save As.

Tasks:

1) There are data on clinical assessment of the severity of sickle cell anemia: 0; 0; 0; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 2; 2; 2; 2; 3; 3; 3; 3; 4; 4; 5; 5; 5; 5; 6; 7; 9; 10; 11. It is necessary to present the sample in the form of a variation series, calculate the sample mean, dispersion, standard deviation, coefficient of variation, median, mode, 25th and 75th percentiles, construct a polygon and a histogram. Can the sample be considered to be drawn from a population with a normal distribution? Check the results obtained in the Statistica program.

2) There are data on the duration (in seconds) of physical activity before the development of an angina attack in 12 people with coronary heart disease: 289, 203, 359, 243, 232, 210, 251, 246, 224, 239, 220, 211. Calculate the sample mean , variance, standard deviation, coefficient of variation, median, 25th and 75th percentiles. Can the sample be considered to be drawn from a population with a normal distribution? Check the results obtained in the Statistica program.

3) There are results for assessing the permeability of retinal vessels: 1.2; 1.4; 1.6; 1.7; 1.7; 1.8; 2.2; 2.3; 2.4; 6.4; 19.0; 23.6. Calculate sample mean, variance, standard deviation, coefficient of variation, median, 25th and 75th percentiles. Can the sample be considered to be drawn from a population with a normal distribution? Check the results obtained in the Statistica program.

4) For 25 days, the number of people who sought emergency medical help was recorded. The resulting sample is: 1, 0, 4, 2, 3, 5, 2, 4, 0, 1, 8, 5, 2, 4, 3, 3, 2, 5, 1, 3, 2, 5, 1 , 3, 2. It is necessary to present the sample in the form of a variation series, calculate the sample mean, dispersion, standard deviation, coefficient of variation, mode and median. Check the results obtained in the Statistica program.

6. Assessment methods / technologies: Oral questioning. Practical work. (assessment using a checklist).

7. Literature:

1. Biostatistics: electronic textbook / K. Zh. Kudabaev [et al.]. - Shymkent: SKFFA, 2015.
2. Rosner Bernard Fundamentals of Biostatistics: Texbook/ B.Rosner - 8th ed.: GENGAGE learning, 2016.
3. Armitage P. Encyclopedia of Biostatistics. - Wiley, 2016. - 6100 p
4. Le C.T. Introductory biostatistics. - Wiley, 2013. - 536 p
5. Newman S.C. Biostatistical Methods in Epidemiology. - Wiley, 2017. - 388p
6. http://www.biometrica.tomsk.ru

8. Control:

1. What indicators of the variation series do you know?
2. What indicators are indicators of central tendency?
3. What are the indicators of dispersion?
4. What uhDo you know the steps for conducting statistical analysis in the STATISTICA program?
5. What types of documents can be created in the STATISTICA program? What extensions do they have?
6. List the main elements of the working window of the STATISTICA program.
7. What operations can be performed on the columns and rows of a spreadsheet?

Lesson №3

1. The theme: Fundamentals of the theory of testing statistical hypotheses. Consent criteria.

2. The aim: Develop the ability to use consent signs to check the assumption of a uniform distribution of the set.

3. The learning objectives:

- in what cases is the Pearson agreement sign and the Kolmogorov-Smirnov agreement sign used;

- learn to form null and alternative forecasts;
- mastering the sign algorithm;
- learn to interpret the result;
- Implementation of the Pearson agreement and the Kolmogorov-Smirnov agreement in the program. STATISTICS.

4. Key questions of the theme:

1. Scheme for testing a statistical hypothesis.
2. Pearson test to test the hypothesis of normal distribution.
3. Kolmogorov-Smirnov test to test the hypothesis of normal distribution.
4. Testing the hypothesis of normal distribution in the STATISTICA program.

5. Methods / technologies of teaching and learning: Computer-based work / Solving situational tasks.

❖ **Tasks:**

Example 1. For statistical analysis of manufactured products, the fracture strength of the tablets was determined. The following results were obtained (in decinewtons):

Intervals	Frequencies, νi	Midpoints of intervals, xi
(373;421]	4	397
(421;445]	6	433
(445;469]	25	457
(469;493]	31	481
(493;517]	44	505
(517;541]	35	529
(541;565]	16	553
(565;589]	4	577

Test the hypothesis of normal distribution for tablet fracture using the χ2-Pearson goodness-of-fit test.

Solution.

1) Create a calculation table:

xi	*vi*	*xi·vi*	*xi- $\bar{x}$*	*(xi- $\bar{x}$)2*	*(xi- $\bar{x}$)2·vi*
397	4	1588	-103	10609	42436
433	6	2598	-67	4489	26934
457	25	11425	-43	1849	46225
481	31	14911	-19	361	11191
505	44	22220	5	25	1100
529	35	18515	29	841	29435
553	16	8848	53	2809	44944
577	4	2308	77	5929	23716
Total	165	82413			225981

2) Calculate the average: $\bar{x} = \frac{\sum_{i=1}^{n} x_i \nu_i}{\sum_{i=1}^{n} \nu_i} = \frac{82413}{165} \approx 500$.

3) Calculate variance:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \nu_i}{n-1} = \frac{225981}{164} \approx 1377,9$$

.

4) Calculate corrected sample variance:

$$s^2 = \frac{n}{n-1} S^2 = \frac{165}{165-1} \cdot 1377,9 \approx 1386.$$

5) Calculate the corrected standard deviation:

$$s = \sqrt{s^2} = \sqrt{1378} \approx 37.$$

Using the goodness-of-fit criterionχ2-Pearson test the hypothesis that the distribution is normal.

6) Determine the probability of a random variable falling into the intervals [xi, xi+1] using the formula:

$$p_i(x_i \le X \le x_{i+1}) = F\left(\frac{x_{i+1} - \bar{x}}{s}\right) - F\left(\frac{x_i - \bar{x}}{s}\right),$$

where F(x) is the distribution function of the normalized normal distribution(see table 1)

Table 1

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,500000	1,00	0,841345	2,00	0,977250
0,05	0,519939	1,05	0,853141	2,05	0,979818
0,10	0,539828	1,10	0,864334	2,10	0,982136
0,15	0,559618	1,15	0,874928	2,15	0,984222
0,20	0,579260	1,20	0,884930	2,20	0,986097
0,25	0,589706	1,25	0,894350	2,25	0,987776
0,30	0,617911	1,30	0,903200	2,30	0,989276
0,35	0,636831	1,35	0,911492	2,35	0,990613
0,40	0,655422	1,40	0,919243	2,40	0,991802
0,45	0,673645	1,45	0,926471	2,45	0,992857
0,50	0,691463	1,50	0,933193	2,50	0,993790
0,55	0,708840	1,55	0,939429	2,55	0,994614
0,60	0,725747	1,60	0,945201	2,60	0,995339
0,65	0,742154	1,65	0,950528	2,65	0,995975
0,70	0,758036	1,70	0,955434	2,70	0,996533
0,75	0,773373	1,75	0,959941	2,75	0,997020
0,80	0,788145	1,80	0,964070	2,80	0,997445
0,85	0,802338	1,85	0,967843	2,85	0,997814
0,90	0,815940	1,90	0,971283	2,90	0,998134
0,95	0,828944	1,95	0,974412	2,95	0,998411
				3,00	0,998650

$$p_1(373 < X \le 421) = \Phi\left(\frac{421 - 500}{37}\right) - \Phi\left(\frac{373 - 500}{37}\right) = 0,02,$$

$$p_2(421 < X \le 445) = \Phi\left(\frac{445-500}{37}\right) - \Phi\left(\frac{421-500}{37}\right) = 0{,}05,$$

$$p_3(445 < X \le 469) = \Phi\left(\frac{469-500}{37}\right) - \Phi\left(\frac{445-500}{37}\right) = 0{,}15,$$

$$p_4(469 < X \le 493) = \Phi\left(\frac{493-500}{37}\right) - \Phi\left(\frac{469-500}{37}\right) = 0{,}21,$$

$$p_5(493 < X \le 517) = \Phi\left(\frac{517-500}{37}\right) - \Phi\left(\frac{493-500}{37}\right) = 0{,}26,$$

$$p_6(517 < X \le 541) = \Phi\left(\frac{541-500}{37}\right) - \Phi\left(\frac{517-500}{37}\right) = 0{,}19,$$

$$p_7(541 < X \le 565) = \Phi\left(\frac{565-500}{37}\right) - \Phi\left(\frac{541-500}{37}\right) = 0{,}09,$$

$$p_8(565 < X \le 589) = \Phi\left(\frac{589-500}{37}\right) - \Phi\left(\frac{565-500}{37}\right) = 0{,}03.$$

7) Create a calculation table:

Interval	Frequencies νi	Probabilities *pi*	Theoretical frequencies $\nu_i^* = np_i$	$(\nu_i - \nu_i^*)^2$	$\frac{(\nu_i - \nu_i^*)^2}{\nu_i^*}$
(373;421]	4	0.02	3.3≈3	1	0.3
(421;445]	6	0.05	8.25≈8	4	0.5
(445;469]	25	0.15	24.75≈25	0	0
(469;493]	31	0.21	34.65≈35	16	0.46
(493;517]	44	0.26	42.9≈43	1	0.02
(517;541]	35	0.19	31.35≈31	16	0.52
(541;565]	16	0.09	14.85≈15	1	0.07
(565;589]	4	0.03	4.95≈5	1	0.2
Sum		1	165		2.07

8) Calculate:

$$\chi^2{}_{расч} = \sum_{i=1}^{k} \frac{(\nu_i - \nu_i^*)^2}{\nu_i^*} = 2{,}07.$$

9) Define:

$\chi 2_{table}(p; f)$, $\chi 2_{table}(0.05;5)=11.1$. (see table 2)

Table 2

Число степеней свободы	Уровень значимости *p*					
	0,01	0,05	0,1	0,90	0,95	0,99
1	6,6	3,8	2,71	0,02	0,004	0,0002
2	9,2	6,0	4,61	0,21	0,1	0,02
3	11,3	7,8	6,25	0,58	0,35	0,12
4	13,3	9,5	7,78	1,06	0,71	0,30
5	15,1	11,1	9,24	1,61	1,15	0,55

6	16,8	12,6	10,6	2,20	1,64	0,87
7	18,5	14,1	12,0	2,83	2,17	1,24
8	20,1	15,5	13,4	3,49	2,73	1,65
9	21,7	16,9	14,7	4,17	3,33	2,09
10	23,2	18,3	16,0	4,87	3,94	2,56
11	24,7	19,7	17,3	5,58	4,57	3,05
12	26,2	21,0	18,5	6,30	5,23	3,57
13	27,7	22,4	19,8	7,04	5,89	4,11
14	29,1	23,7	21,1	7,79	6,57	4,66
15	30,6	25,0	22,3	8,5	7,26	5,23
16	32,0	26,3	23,5	9,31	7,98	5,81
17	33,4	27,6	24,8	10,1	8,67	6,41
18	34,8	28,9	26,0	10,9	9,39	7,01
19	36,2	30,1	27,2	11,7	10,1	7,63
20	37,6	31,4	28,4	12,4	10,9	8,26
21	38,9	32,7	29,6	13,2	11,6	8,90
22	40,3	33,9	30,6	14,0	12,63	9,54
23	41,6	35,2	32,0	14,8	13,1	10,2
24	43,0	36,4	33,2	15,7	13,8	10,9
25	44,3	37,7	34,4	16,5	14,6	11,5
26	45,6	38,9	35,6	17,3	15,4	12,2
27	47,0	40,1	36,7	18,1	16,2	12,9
28	48,3	41,3	37,9	18,9	16,9	13,6
29	49,6	42,6	39,1	19,8	17,7	14,3
30	50,9	43,8	40,3	20,6	18,5	15,0

1) Compare $\chi 2_{calculation}$ And $\chi 2_{table}$: $\chi^2_{расч} < \chi^2_{табл}$, which means the hypothesis that the distribution is normal is accepted.

Example 2. Using the goodness-of-fit criterion Kolmogorov-Smirnov to test the hypothesis that the distribution is normal.

1) Determine the values of the theoretical distribution function using the formula:

$$F(x) = \frac{1}{2} + \Phi\left(\frac{x_{i+1} - \bar{x}}{s}\right),$$

where Φ(x) is the Laplace function (see table 3)

Table 3

x	0	1	2	3	4	5	6	7	8	9
	Сотые доли x									
0,0	0,0000	0040	0080	0112	0160	0199	0239	0279	0319	0359
0,1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0754
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2258	2291	2324	2357	2389	2422	2454	2486	2518	2549
0,7	2580	2612	2642	2673	2704	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2996	3023	3051	3079	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389

1,0	3413	3438	3461	3485	3508	3531	3553	3577	3599	3621
1,1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4430	4441
1,6	4452	4463	4474	4485	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4700	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4762	4767
	Десятые доли x									
2,	4773	4821	4861	4893	4918	4938	4953	4965	4974	4981
3,	4987	4990	4993	4995	4997	4998	4998	4999	4999	5000[8]

$$F_1(x)=\frac{1}{2}+\Phi\left(\frac{421-500}{37}\right)=0{,}02,\qquad F_2(x)=\frac{1}{2}+\Phi\left(\frac{445-500}{37}\right)=0{,}07,$$

$$F_3(x)=\frac{1}{2}+\Phi\left(\frac{469-500}{37}\right)=0{,}20,\qquad F_4(x)=\frac{1}{2}+\Phi\left(\frac{493-500}{37}\right)=0{,}43,$$

$$F_5(x)=\frac{1}{2}+\Phi\left(\frac{517-500}{37}\right)=0{,}68,\qquad F_6(x)=\frac{1}{2}+\Phi\left(\frac{541-500}{37}\right)=0{,}87,$$

$$F_7(x)=\frac{1}{2}+\Phi\left(\frac{565-500}{37}\right)=0{,}96,\qquad F_8(x)=\frac{1}{2}+\Phi\left(\frac{589-500}{37}\right)=0{,}99.$$

1) Create a calculation table:

Interval [xi, xi+1]	Frequencies νi	Cumulative frequencies νi, accumulated	$F_n(x)=\frac{\nu_{i,\ накопл}}{n}$	$F(x)$	$\lvert F_n(x)-F(x)\rvert$
(373;421]	4	4	4/165=0.02	0.02	0
(421;445]	6	4+6=10	10/165=0.06	0.07	0.01
(445;469]	25	10+25=35	0.21	0.20	0.01
(469;493]	31	35+31=66	0.4	0.43	0.03
(493;517]	44	66+44=110	0.67	0.68	0.01
(517;541]	35	110+35=145	0.88	0.87	0.01
(541;565]	16	145+16=161	0.98	0.96	0.02
(565;589]	4	161+4=165	1	0.99	0.01

$$d_{\max}=\max\left|F_n(x)-F(x)\right|=0{,}03.$$

2) Calculate: $\lambda_{расч}=d_{\max}\sqrt{n}=0{,}03\cdot\sqrt{165}\approx 0{,}39.$

3) $\lambda_{табл}=1{,}36$.

4) Compare $\lambda_{расч}$ And $\lambda_{табл}$: $\lambda_{расч}<\lambda_{табл}$, which means the hypothesis that the distribution is normal is accepted.

Tasks.

1) The sample is presented in the form of a statistical series (n=200):

xi	0.3	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9	2.1	2.3
v_i	6	9	26	25	30	26	21	24	20	8	5

Test the hypothesis about the normal distribution of the sample using the χ2-Pearson goodness-of-fit test.

6. Assessment methods / technologies: Oral questioning. Practical work. (assessment using a checklist).

7. Literature:

Main:

1. Biostatistics: electronic textbook / K. Zh. Kudabaev [et al.]. - Shymkent: SKFFA, 2015.
2. Rosner Bernard Fundamentals of Biostatistics: Texbook/ B.Rosner - 8th ed.: GENGAGE learning, 2016.
3. Armitage P. Encyclopedia of Biostatistics. - Wiley, 2016. - 6100 p
4. Le C.T. Introductory biostatistics. - Wiley, 2013. - 536 p
5. Newman S.C. Biostatistical Methods in Epidemiology. - Wiley, 2017. - 388p
6. http://www.biometrica.tomsk.ru

8. Control:

1. What is a statistical hypothesis? What types of statistical hypothesis do you know?
2. What is the general scheme for testing statistical hypotheses?
3. What are the consent criteria used for?
4. What is the scheme for applying the Pearson χ2 goodness-of-fit test?
5. What is the scheme for using the Kolmogorov-Smirnov goodness-of-fit test?
6. How are Pearson's χ2 goodness-of-fit tests implemented in the Statistica program?

Lesson №4

1. The theme: Parametric methods of comparative statistics.

2. The aim: studying the methodological foundations and conditions for using the two-sample and paired Student's t-test and introducing them into the "STATISTICA" program.

3. The learning objectives:

- understand in what cases two-sample and paired sampling is used*t*-Student's test;
- learn to formulate null and alternative hypotheses;
- master the algorithm of the criterion;
- learn to interpret the result;
-cdevelop skills in applying two-sample and paired Student's t-tests in the STATISTICA program.

4. Key questions of the theme:

1. Student's t-test for analyzing biomedical data.
2. Conditions for using the Student t-test.
3. Scheme for applying the two-sample Student t-test.
4. Scheme for applying the paired Student's t-test.
5. Interpretation of the result.
6. Checking the conditions for applying the Student's test (normal distribution of samples, equality of variances).
7. The procedure for implementing the Student's t-test in the "STATISTICA" program.

8. Interpretation of results.

5. Methods/technologies of teaching and learning: Computer-based work / Solving situational tasks.

❖ **Tasks:**

Example 1. If the cervix does not dilate for a long time during childbirth, the duration of labor increases and a caesarean section may be necessary. Scientists decided to find out whether a gel with prostaglandin E2 accelerates the dilatation of the cervix. The study included 2 groups of women in labor. Women in labor in the first group were injected into the cervix with prostaglandin E2 gel, while women in the second group were injected with a placebo gel. In both groups there were 21 women in labor, the age, height and duration of pregnancy were approximately the same. Labor in the group receiving prostaglandin E2 gel lasted on average 8.5 hours (standard deviation 4.7 hours), in the control group - 13.9 hours (standard deviation 4.1 hours). Is it possible to say that the gel with prostaglandin E2 shortened the duration of labor?

Solution.

1) $H_0: \bar{x}_1 = \bar{x}_2$.

$H_1: \bar{x}_1 \neq \bar{x}_2$.

2) α=0.05

3) $t_{расч} = \dfrac{13{,}9 - 8{,}5}{\sqrt{(21-1)\cdot 4{,}1^2 + (21-1)\cdot 4{,}7^2}} \cdot \sqrt{\dfrac{21\cdot 21}{21+21}(21+21-2)} \approx 4$.

4) $t_{табл}(0{,}05;40) = 2{,}02$.(See table 5)

Table 5

Erkindik darijeler sana	Mynyzdylyk money, r			
f	0.10	0.05	0.02	0.01
1	6.31	12.7	31.82	63.7
2	2.92	4.30	6.97	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.60
5	2.01	2.57	3.37	4.03
6	1.94	2.45	3.14	3.71
7	1.89	2.36	3.00	3.50
8	1.86	2.31	2.90	3.36
9	1.83	2.26	2.82	3.25
10	1.81	2.23	2.76	3.17
11	1.80	2.22	2.72	3.11
12	1.78	2.18	2.68	3.05
13	1.77	2.16	2.65	3.01
14	1.76	2.14	2.62	2.98
15	1.75	2.13	2.60	2.95
16	1.75	2.12	2.58	2.92
17	1.74	2.11	2.57	2.90
18	1.73	2.10	2.55	2.88
19	1.73	2.09	2.54	2.86
20	1.73	2.09	2.53	2.85

30	1.70	2.04	2.46	2.75
40	1.68	2.02	2.42	2.70
60	1.67	2.00	2.39	2.66
120	1.66	1.98	2.36	2.62
∞	1.64	1.96	2.33	2.58

5) Because $t_{расч} > t_{табл}$, then "H0" is rejected, i.e. gel with prostaglandin E2 shortened the duration of labor.

Example 2. To evaluate the effectiveness of the new hypoglycemic drug, blood glucose levels were measured in patients with diabetes mellitus before and after taking the drug:

Patient no.	Blood glucose level, mol/l	
	before taking the drug	after taking the drug
1	9.6	5.7
2	8.1	4.2
3	8.8	6.4
4	7.9	5.5
5	9.2	5.3
6	8.0	5.2
7	8.4	5.1
8	10.1	5.9
9	7.8	7.5
10	8.1	5.0
Average value	8.6	5.6

Can we assume that after taking the drug, the level of glucose in the blood of patients decreases?

Solution.

1) $H_0: \bar{x}_1 = \bar{x}_2$.

$H_1: \bar{x}_1 \neq \bar{x}_2$.

2) α=0.05- significance level.

Patient no.	Blood glucose level, mol/l		Difference of values $d = x_i - y_i$	$(d_i - \bar{d})^2$
	before taking the drug	after taking the drug		
1	9.6	5.7	3.9	0.77
2	8.1	4.2	3.9	0.77
3	8.8	6.4	2.4	0.38
4	7.9	5.5	2.4	0.38
5	9.2	5.3	3.9	0.77
6	8.0	5.2	2.8	0.048
7	8.4	5.1	3.3	0.078
8	10.1	5.9	4.2	1.39
9	7.8	7.5	0.3	7.4
10	8.1	5.0	3.1	0.006

Sum	30.2	12

1) $\bar{d} = \frac{30,2}{10} = 3,02$. $\bar{d} = \frac{30,2}{10} = 3,02$

2) $S_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$ $S_d = \sqrt{\frac{12}{9}} = 1,15$

3) $t_{расч} = \frac{\bar{d}}{S_d / \sqrt{n}}$ $t_{расч} = \frac{3,02}{1,15/\sqrt{10}} = 8,39$

4) $t_{табл}(0,05;9) = 2,26$. (See Table 5)

5) Because $t_{расч} > t_{табл}$, then "H0" is rejected, i.e. The blood glucose level has decreased after taking the drug, which means the new drug is effective.

Example 3. Medical students had their pulse examined before and after passing the exam. Pulse rate before the exam was 98.8±4.0, and after the exam 84.0±5.0. Can we assume that after the exam the heart rate decreases and approaches normal?

Solution.

1) $H_0: \bar{x}_1 = \bar{x}_2$.

$H_1: \bar{x}_1 \neq \bar{x}_2$.

2) α=0.05.

3) $t_{расч} = \frac{98,8 - 84}{\sqrt{4^2 + 5^2}} \approx 2,3.$

4) Because tcalc>2, then "H0" is rejected, i.e. After the exam, the heart rate decreases and approaches normal.

Example 4. Compare test results in two groups.

No.	Group No. 1 results (sec.)	Group No. 2 results (sec.)
1	thirty	46
2	45	49
3	41	52
4	38	55
5	34	56
6	36	40
7	31	46
8	30	51
9	49	58
10	50	46
11	51	46
12	46	56
13	41	53

14	37	57
15	36	44
16	34	42
17	33	40
18	49	58
19	32	54
20	46	53
21	41	51
22	44	57
23	38	56
24	50	44
25	37	42
26	39	49
27	40	50
28	46	55
29	42	43

1. Create a data table “Test Results” of size 2*58 in the Statistica program, enter the initial data.

2. Select Statistics → Basic Statistics → t-test independent by groups (t-test for independent samples) (Figure 1).

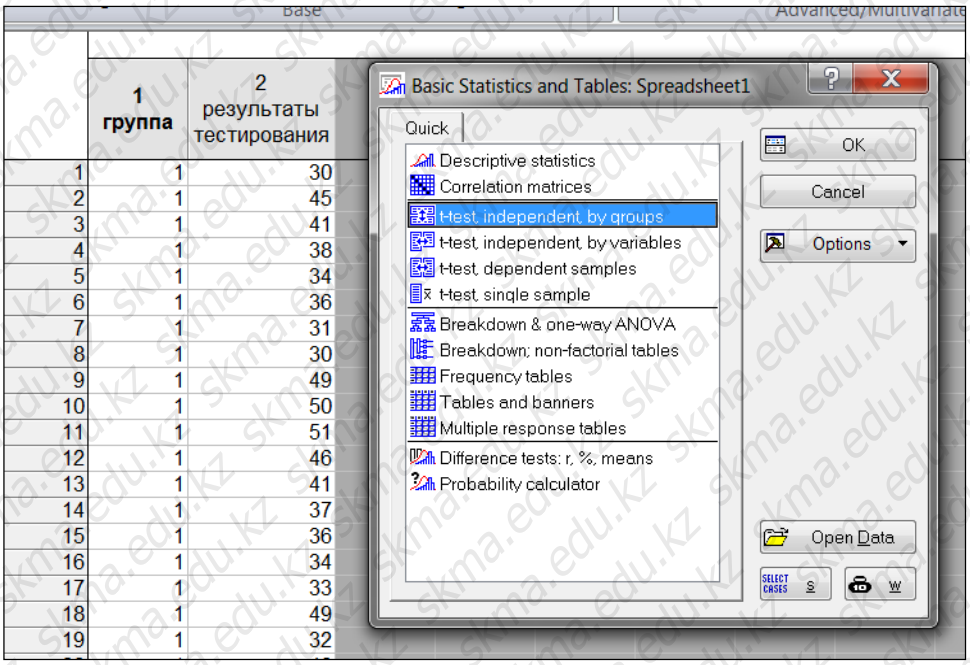


Figure 1. Selecting the “t-test independent by groups” procedure

3. In the dialog box, click the “Variables” button, indicate the grouping characteristic (column containing group codes) in the right part of the window, and the column containing the analyzed characteristic in the left part of the window (Figure 2), click the “OK” button.

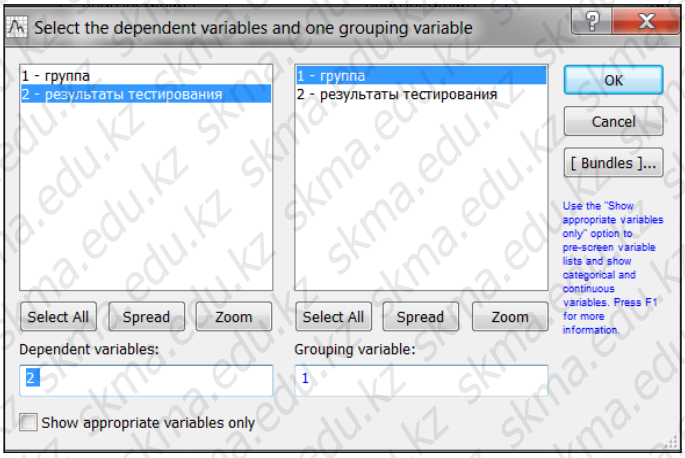


Figure 2. Setting variables

4. Check samples for normal distribution.

Select the “Advanced” tab, build “Categorized normal plots” and “Categorized histograms”. Draw a conclusion about the normality of the distribution (Figure 3).

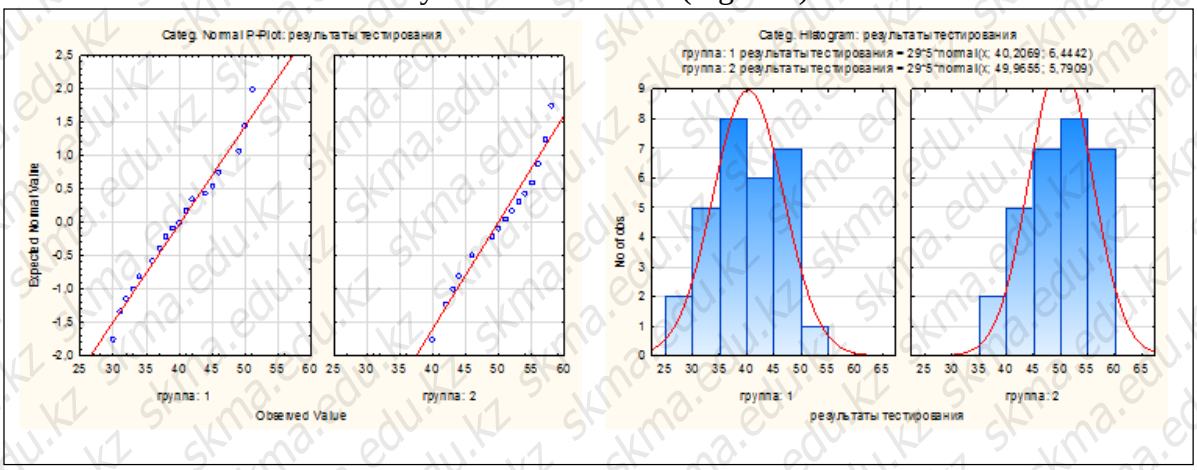


Figure 3. Testing samples for normal distribution

5. Check the equality of sample variances.

To check this condition, the program automatically uses Fisher's F-test, but you can also use Levene's test, to do this you need to select the “Options/Levene's test” option (Figure 4).

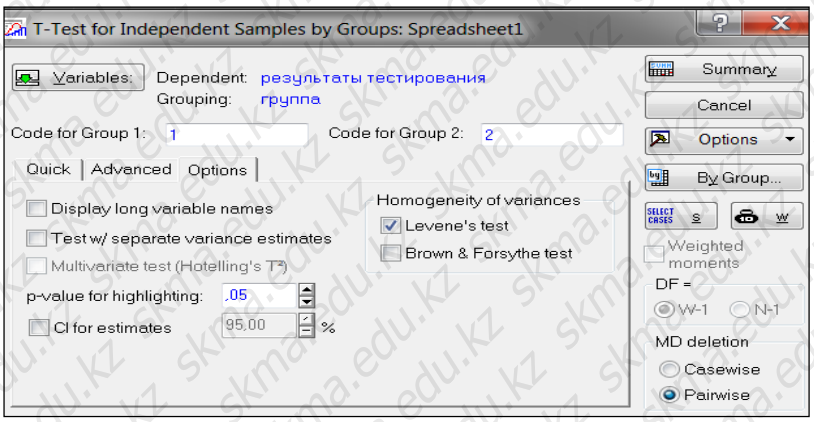


Figure 4. Selecting the “Levene's test” option

6. Click the "Summary" button, a final table will appear on the screen with the results of comparing two independent samples using the Student's t-test (Figure 5).

T-tests; Grouping: группа (Spreadsheet1)
Group 1: 1
Group 2: 2

Variable	Mean 1	Mean 2	t-value	df	p	Valid N 1	Valid N 2	Std.Dev. 1	Std.Dev. 2	F-ratio Variances	p Variances	Levene F(1,df)	df Levene	p Levene
результаты тестирования	40,20690	49,96552	-6,06567	56	0,000000	29	29	6,444152	5,790897	1,238340	0,575457	0,222792	56	0,638755

Figure 5. Final table with the results of comparison of two independent samples using Student's t-test

Names of the columns of the final table:
*Mean*1, Mean 2 – average values of variables;
t-value– t-test value;
df– number of degrees of freedom;
R– level of significance of the t-test;
*Valid*1, Valid 2 - number of observations in groups;
Std. Dev.– standard deviations of variable values;
F-ratio Variances– F-criterion value;
*p*Variances - level of significance of the F-test;
Levene F(1,df) –Lewin's criterion value;
df Levene- number of degrees of freedom of Lewin's criterion;
r Levene -level of significance of Lewin's criterion.

If Levine's test $p<0.05$, a conclusion should be drawn about the difference in variances in the compared groups.

If for Lewin's test $p>0.05$, one should conclude that the variances in the compared groups are equal.

Similarly for the F-test.

If the t-test value is $p>0.05$, then the null hypothesis of equality of means is accepted.

If the t-test has a p-value of <0.05 (such results are highlighted in red), then the null hypothesis of equality of means is rejected.

For this example:

- *p*Variances>0.05, which means the variances are equal (Fisher's F test);
- *r Levene* >0.05, which means the variances are equal (Levine's criterion);
- *p*<0.05, which means the null hypothesis of equality of means is rejected (Student's t-test).

Example 5.Compare the test results of students before and after training.

No.	Results before training (sec.)	Results after training (sec.)
1	thirty	46
2	45	49
3	41	52
4	38	55
5	34	56
6	36	40
7	31	46
8	30	51
9	49	58
10	50	46

eleven	51	46
12	46	56
13	41	53
14	37	57
15	36	44
16	34	42
17	33	40
18	49	58
19	32	54
20	46	53
21	41	51
22	44	57
23	38	56
24	50	44
25	37	42
26	39	49
27	40	50
28	46	55
29	42	43

1. Create a data table “Test results before and after training” of size 2*29 in the Statistica program, enter the initial data.

2. Check samples for normal distribution.

Statistics→ Basic statistics → Descriptive statistics → Normality. Normality testing can be done using the Kolmogorov-Smirnov & Lilliefors test of normality or the Shapiro-Wilk's W test. To do this, select the appropriate options and click the “Histograms” button (Figure 1).

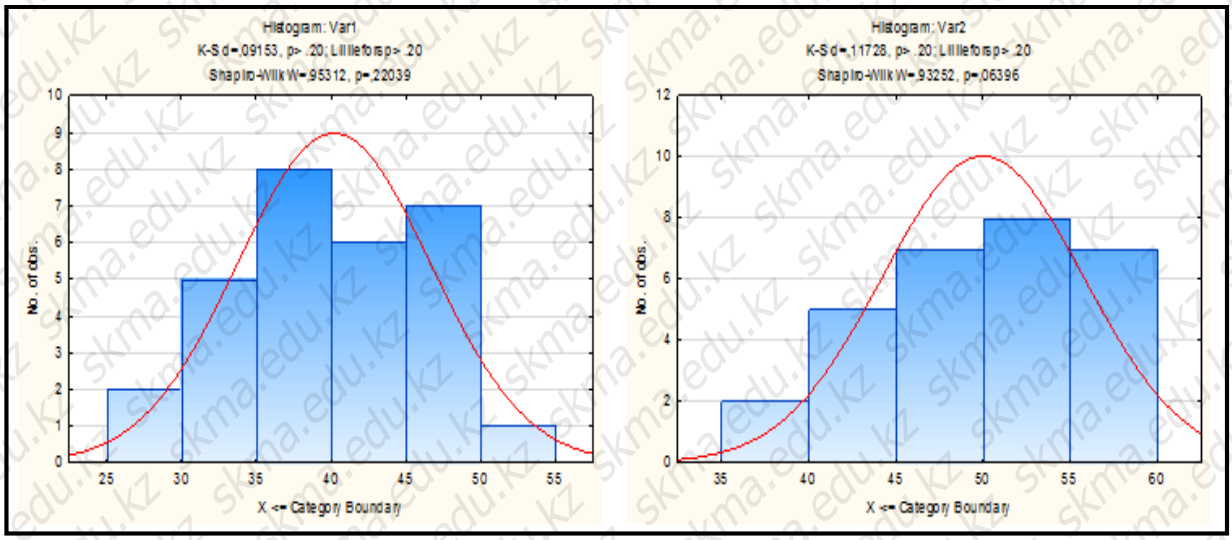
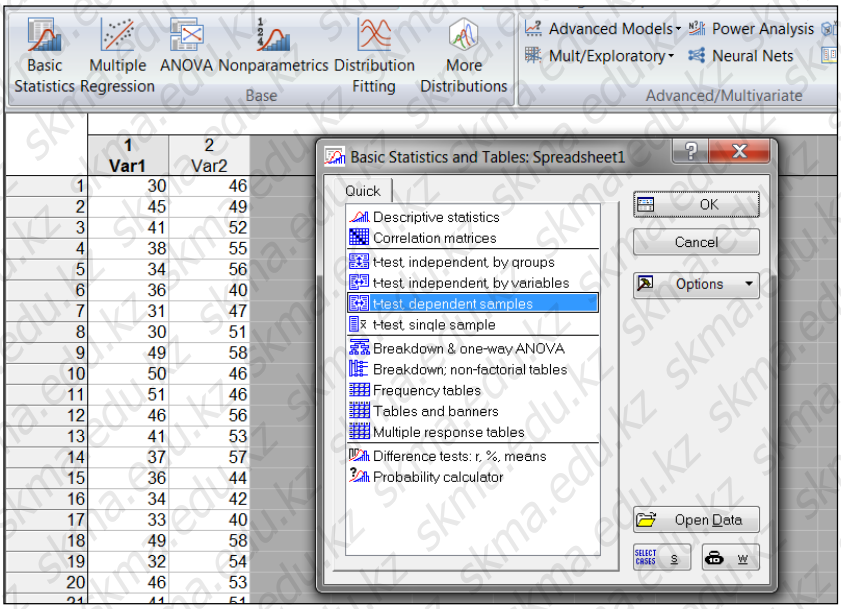


Figure 1. Testing samples for normal distributions

The histograms that appear (Figure 1) at the top show the results of tests for normality, which can be explained as follows:

- if in these tests p>0.05, then the hypothesis of normal distribution is accepted;
- if p<0.05, then the hypothesis of normal distribution is rejected.

3. To test the null hypothesis about the equality of means in dependent groups using the Student t-test, select: Statistics → Basic statistics → t-test dependent samples (t-test for dependent samples) (Figure 2).



	1 Var1	2 Var2
1	30	46
2	45	49
3	41	52
4	38	55
5	34	56
6	36	40
7	31	47
8	30	51
9	49	58
10	50	46
11	51	46
12	46	56
13	41	53
14	37	57
15	36	44
16	34	42
17	33	40
18	49	58
19	32	54
20	46	53

Figure 2. Selecting the t-test dependent samples procedure

4. Set variables "Variables".

5. Click the "Summary" button, a summary table will appear on the screen (Figure 3).

T-test for Dependent Samples (Spreadsheet1)
Marked differences are significant at p < ,05000

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Confidence +95,000%
Var1	40,20690	6,444152								
Var2	50,00000	5,769377	29	-9,79310	7,537427	-6,99675	28	0,000000	-12,6602	-6,92602

Figure 3.Final table with the results of comparison of two dependent samples using Student's t-test

The results obtained can be interpreted as follows:

- if p<0.05, then the null hypothesis is rejected (such results are highlighted in red font);
- if p>0.05, then the null hypothesis is accepted.

For this example*R*<0.05, which means the null hypothesis of equality of means is rejected.

Example 6. Compare test results in two groups. The samples are normally distributed and have equal variances. It is known that $n_1 = 29$, $n_2 = 29$, $\bar{x}_1 = 40$, $\bar{x}_2 = 50$, $s_1 = 6{,}4$, $s_2 = 5{,}8$.

1. Select Statistics → Basic Statistics → Difference tests: r, %, means (Other tests).

2. Substitute the appropriate parameters for each of the two groups into the dialog box and perform the calculation to obtain the "p" value (Figure 1).

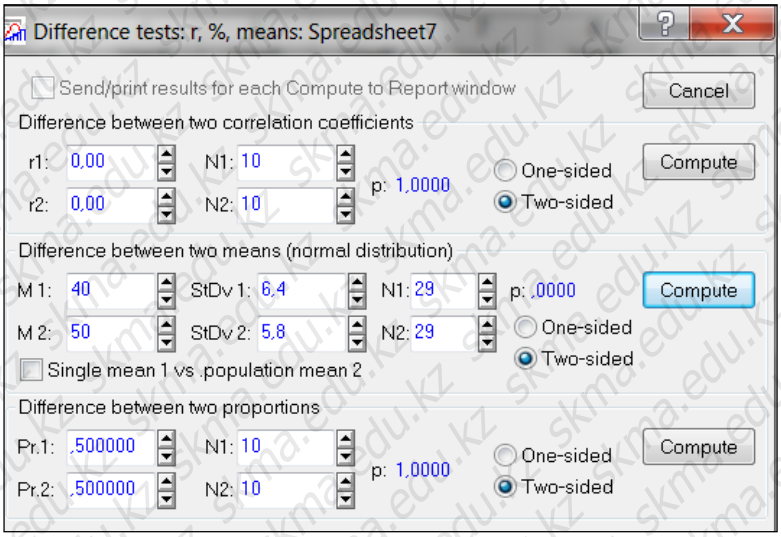


Figure 1. Dialog box of the procedure “Difference tests: r, %, means”

The results obtained can be interpreted as follows:

- if p>0.05, then the null hypothesis of no differences in means is accepted;
- if p<0.05, then the null hypothesis is rejected.

For this example, p < 0.05, which means the null hypothesis of equality of means is rejected.

Tasks

1. Two batches of tablets were produced at different compression pressures (80 and 100 MPa). The fracture strength of tablets of the first group turned out to be 50.4; 53.6; 54.4; 46.4; 44.0; 48.2; 49.4 (in Newtons), in the second group – 47.2; 62.4; 64.8; 62.4; 58.9; 55.4; 66.2; 49.5; 67.8; 68.9 (in Newtons). Using the Student's test at p=0.05, check the hypothesis about the equality of means (the alternative hypothesis is about their inequality).

2. The psychological test measured choice reaction time in two groups. The first group included athletes, the second group included people not involved in sports. In the first group the following results were obtained: 0.42, 0.52, 0.48, 0.46, 0.55, 0.62, 0.58, 0.64, 0.56 (seconds). In the second: 0.51, 0.67, 0.54, 0.52, 0.56, 0.66, 0.68 (seconds). Using the Student's test at p=0.05, check the hypothesis about the equality of means (the alternative hypothesis is about their inequality).

3. Using paired Student's t-test, compare the results of performing logical tasks before and after the training course. The initial data is presented in the table.

No.	Results of performing logical tasks before the course (sec.)	Results of performing logical tasks after the course (sec.)
1	25	22
2	23	25
3	28	23
4	29	22
5	35	30
6	31	27
7	24	20
8	24	19
9	38	32
10	26	25
11	20	20

4. Using a paired Student's t test, determine whether smoking affects platelet function. The initial data are shown in the table.

No.	Platelet aggregation	
	Before smoking	After smoking
1	25	27
2	25	29
3	27	37
4	44	56
5	30	46
6	67	82
7	53	57
8	53	80
9	52	61
10	60	59
11	28	43

5. Using a paired Student's t test, check whether the drug under study functions as a diuretic. The initial data are shown in the table.

No.	Daily diuresis, ml	
	Before taking the drug	After taking the drug
1	1490	1600
2	1300	1850
3	1400	1300
4	1410	1500
5	1350	1400
6	1000	1010

6. Using paired Student's t-test, check the effectiveness of a special diet to get rid of excess weight. The initial data is presented in the table.

No.	Weight (kg) before experiment	Weight (kg) after experiment
1	93.2	88.9
2	98.2	94.5
3	105.6	106.1
4	86.8	84.3
5	95.5	92.5

6. Assessment methods / technologies: Oral questioning. Practical work. (assessment using a checklist).

7. Literature:

Main:

1. Biostatistics: electronic textbook / K. Zh. Kudabaev [et al.]. - Shymkent: SKFFA, 2015.
2. Rosner Bernard Fundamentals of Biostatistics: Texbook/ B.Rosner - 8th ed.: GENGAGE learning, 2016.
3. Armitage P. Encyclopedia of Biostatistics. - Wiley, 2016. - 6100 p
4. Le C.T. Introductory biostatistics. - Wiley, 2013. - 536 p
5. Newman S.C. Biostatistical Methods in Epidemiology. - Wiley, 2017. - 388p
6. http://www.biometrica.tomsk.ru

8. Control:

1. Why is the Student's t-test very popular in statistical analysis of biomedical data?

2. What conditions must be met when using the Student's t-test?
3. How are the null and alternative hypotheses formulated for the Student's t-test?
4. In what ways can it be implemented*t*-Student's test in the program "STATISTICA"?
5. How to check the conditions for applying the Student's test in the STATISTICA program?
6. How is the information contained in the summary table interpreted?

Lesson №5

1. **The theme:** Nonparametric methods of comparative statistics.
2. **The aim:** studying the methodological foundations and conditions for using the Mann-Whitney U-test and the Wilcoxon W(T)-test, becoming familiar with some of the capabilities of the "STATISTICA" program
3. **The learning objectives:**
   - understand in what cases it is used*U*Mann-Whitney test and Wilcoxon W(T) test;
   - learn to formulate null and alternative hypotheses;
   - master the algorithm of the criterion;
   - learn to interpret the result.
4. **Key questions of the theme:**
   1. Mann-Whitney U test is a nonparametric analogue of the two-sample Student t test.
   2. Wilcoxon W (T) test is a nonparametric analogue of the paired Student t test.
   3. Conditions for using the Mann-Whitney U test and the Wilcoxon W (T) test.
   4. Scheme of application of the Mann-Whitney U-test and Wilcoxon W (T)-test.
   5. The procedure for implementing the Mann-Whitney criterion in the "STATISTICA" program.
   6. The procedure for implementing the Wilcoxon test in the "STATISTICA" program.
   7. Interpretation of results.
5. **Methods / technologies of teaching and learning:** Computer-based work / Solving situational tasks.

❖ **Tasks:**

Example 1. Researchers decided to find out the effectiveness of the drug, which allows obese patients to lose excess weight. In this case, a group of patients was prescribed a certain diet.

A month later, the amount of weight lost was recorded. 8 people were selected to conduct the experiment. 3 of them received the study drug (experimental group), and 5 received placebo (control group). The selection of 3 out of 8 subjects into the experimental group was carried out randomly. All participants in the experiment believed that they were taking the drug.

	Lost weight, kg				
Experimental group	6,2	3,0	3,9		
Control group	4,0	-0,5	3,3	1,5	3,0

Solution.

1) $H_0: \bar{x}_1 = \bar{x}_2$.

$H_1: \bar{x}_1 \neq \bar{x}_2$.

2) α=0.05 - significance level.

3) Let's make a single row.

Lost weight, kg	6.2	3.0	3.9	4.0	-0.5	3.3	1.5	3.0
Rank	8	3.5	6	7	1	5	2	3.5

Let us divide the single ranked series into two, consisting of units of the first and second samples.

Experimental group		Control group	
Lost weight, kg	Rank	Lost weight, kg	Rank
6.2	8	4.0	7
3.0	3.5	-0.5	1
3.9	6	3.3	5
		1.5	2
		3.0	3.5
	T1=17.5		*T2*=18.5

T1 and T2 – sums of ranks; T1<T2, which means T2=Tx, nx=n2=5.

$$U_{расч} = 3 \cdot 5 + \frac{5 \cdot (5+1)}{2} - 18{,}5 = 11{,}5.$$

4) *Utable* $(0{,}05; 3; 5)$=1 (see Appendix 5).

p=0,05

N1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N2																			
3	...	0																	
4	...	0	1																
5	0	1	2	4															
6	0	2	3	5	7														
7	0	2	4	6	8	11													
8	1	3	5	8	10	13	15												
9	1	4	6	9	12	15	18	21											
10	1	4	7	11	14	17	20	24	27										
11	1	5	8	12	16	19	23	27	31	34									
12	2	5	9	13	17	21	26	30	34	38	42								
13	2	6	10	15	19	24	28	33	37	42	47	51							
14	3	7	11	16	21	26	31	36	41	46	51	56	61						
15	3	7	12	18	23	28	33	39	44	50	55	61	66	72					
16	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83				
17	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96			
18	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109		
19	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	
20	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138

5) $U_{расч} > U_{табл}$, then "H0" is accepted, i.e. the drug is ineffective.

Example 2. To check whether there is a difference in the blood sugar level on an empty stomach before work and three hours after work in 12 workers working on ultrasonic units. The initial data are shown in the table.

No.	Sugar content before work	Sugar content after work
1	112	54
2	82	67
3	101	96
4	72	59
5	79	79
6	82	76
7	64	66
8	70	66

9	88	48
10	81	50
11	66	61
12	88	61

Solution.

1) $H_0: \bar{x}_1 = \bar{x}_2$.

$H_1: \bar{x}_1 \neq \bar{x}_2$.

2) $\alpha \approx 0.05$ - significance level.

No.	Blood sugar level											
Sugar content BEFORE work	112	82	101	72	79	82	64	70	88	81	66	88
Sugar content AFTER work	54	67	96	59	79	76	66	66	48	50	61	61
Sugar difference	58	15	5	13	0	6	-2	4	40	31	5	27
Ranking. row	0	-2	4	5	5	6	13	15	27	31	40	58
Ranks		1	2	3.5	3.5	5	6	7	8	9	10	11
T+	**65**											
T-	**1**											

3) *Trasch*=1

4) *Ttable (*0,052; 12 *)*=50 (see Appendix 6).

n	T	α	n	T	α
5	15	0,062	13	65	0,022
6	21	0,032		57	0,048
	19	0,062	14	73	0,020
7	28	0,016	15	63	0,050
	24	0,046		80	0,022
8	32	0,024	16	70	0,048
	28	0,054		88	0,022
9	39	0,020	17	76	0,050
	33	0,054		97	0,020
10	45	0,020	18	83	0,050
	39	0,048		105	0,020
11	52	0,018	19	91	0,048
	44	0,054		114	0,020
12	58	0,020	20	98	0,050
	50	0,052		124	0,020
				106	0,048

5) *Trasch*< Ttable, then "H0" is rejected, which means there is a difference in the blood sugar levels of workers before and after work.

Example 3. The effectiveness of a drug that allows obese patients to lose excess weight is being studied. In this case, a group of volunteers is prescribed a certain diet.

After a month, in order to check compliance with the diet and regular use of the drug, the

amount of weight lost (kg) is recorded. A group of 8 people was selected to conduct the experiment. 3 of them received the study drug (experimental group), and 5 received placebo (control group). The selection of 3 out of 8 subjects into the experimental group was carried out randomly. All participants in the experiment believed that they were taking the drug.

	Lost weight, kg				
Experimental group	6,2	3,0	3,9		
Control group	4,0	-0,5	3,3	1,5	3,0

1. Create a data table "Drug Efficacy" in the "Statistica" program with a size of 2*8 and enter the initial data.

2. Select Statistics→Nonparametrics (Figure 1)→Comparing two independent samples (groups) (Figure 2), click the "OK" button.

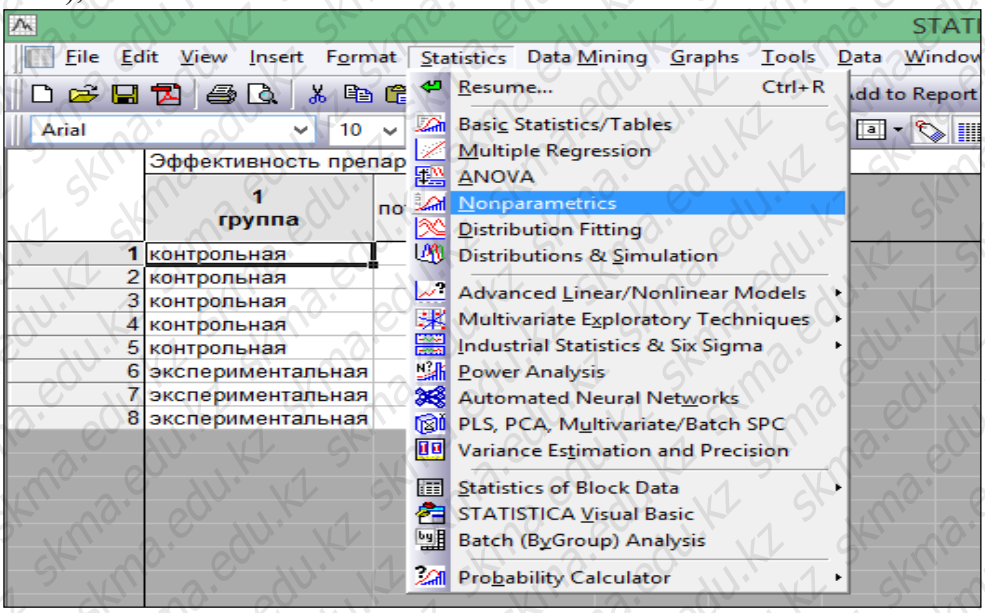


Figure 1. Selecting the "Nonparametrics" module

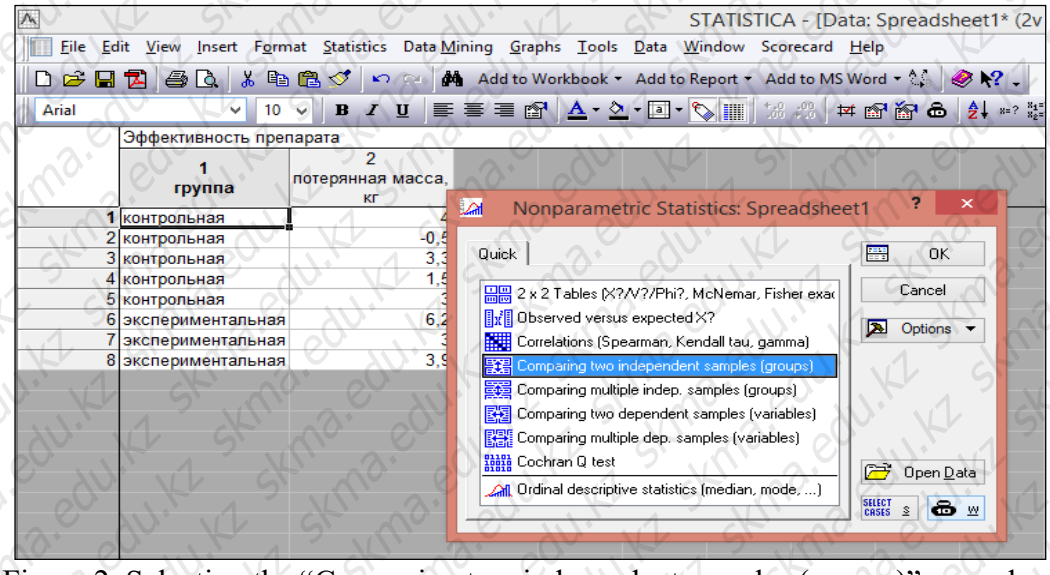


Figure 2. Selecting the "Comparing two independent samples (groups)" procedure

3. In the dialog box, click the "Variables" button, indicate the grouping characteristic (column containing group codes) on the right side of the window, and the column containing the analyzed characteristic on the left side of the window (Figure 3) and click the "OK" button.

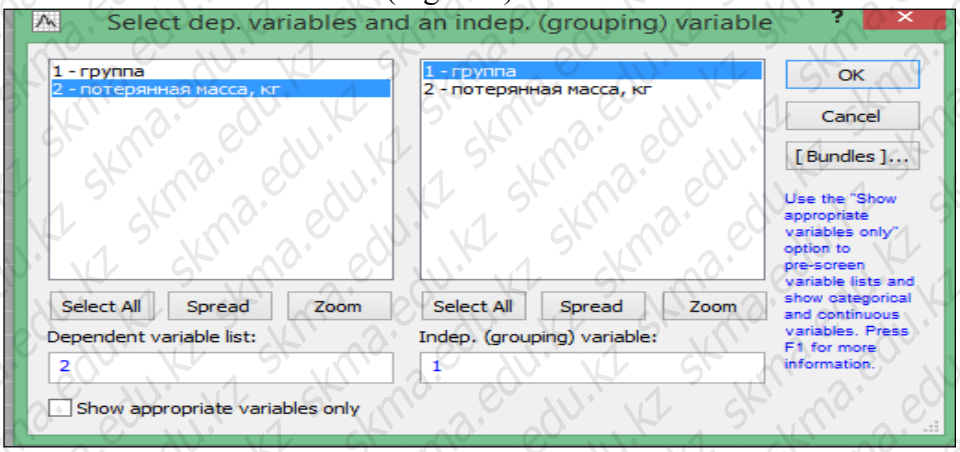


Figure 3. Setting variables

4. In the dialog box, click the "Mann-Whitney U test" button.

A summary table will appear on the screen (Figure 4).

Mann-Whitney U Test (Spreadsheet1)
By variable группа
Marked tests are significant at p <,05000

variable	Rank Sum контрольная	Rank Sum экспериментальная	U	Z	p-value	Z adjusted	p-value	Valid N контрольная	Valid N экспериментальная	2*1sided exact p
потерянная масса, кг	18,50000	17,50000	3,500000	-1,04350	0,296718	-1,04977	0,293827	5	3	0,250000

Figure 4. Final table with the results of comparing two independent samples using the Mann-Whitney test

The first and second columns indicate the Rank Sum.

The results obtained can be interpreted as follows:

- if $p<0.05$, then the null hypothesis is rejected (such results are highlighted in red font);
- if $p>0.05$, then the null hypothesis is accepted.

In this example, $p = 0.296718$ and $p = 0.293827$, which means the hypothesis of equality of means is accepted, i.e. the drug is ineffective.

*5.*For clarity, you can build a "box with whiskers" plot by clicking the "Box & whisker plots for all variables" button in the analysis window (Figure 5).

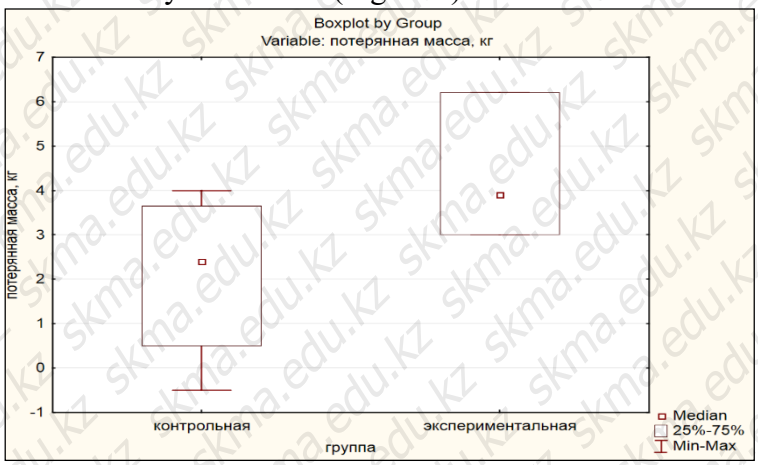


Figure 5. Box-and-whisker plot

Example 4. To check whether there is a difference in the blood sugar level on an empty stomach before work and three hours after work in 12 people working on ultrasound machines.

No.	Sugar content before work	Sugar content after work
1	112	54
2	82	67
3	101	96
4	72	59
5	79	79
6	82	76
7	64	66
8	70	66
9	88	48
10	81	50
11	66	61
12	88	61

1. Create a data table "Blood sugar content" in the Statistica program, size 2*12, and enter the initial data.

2. Select Statistics → Nonparametrics → Comparing two dependent samples (variables) (Figure 1) and click the "OK" button.

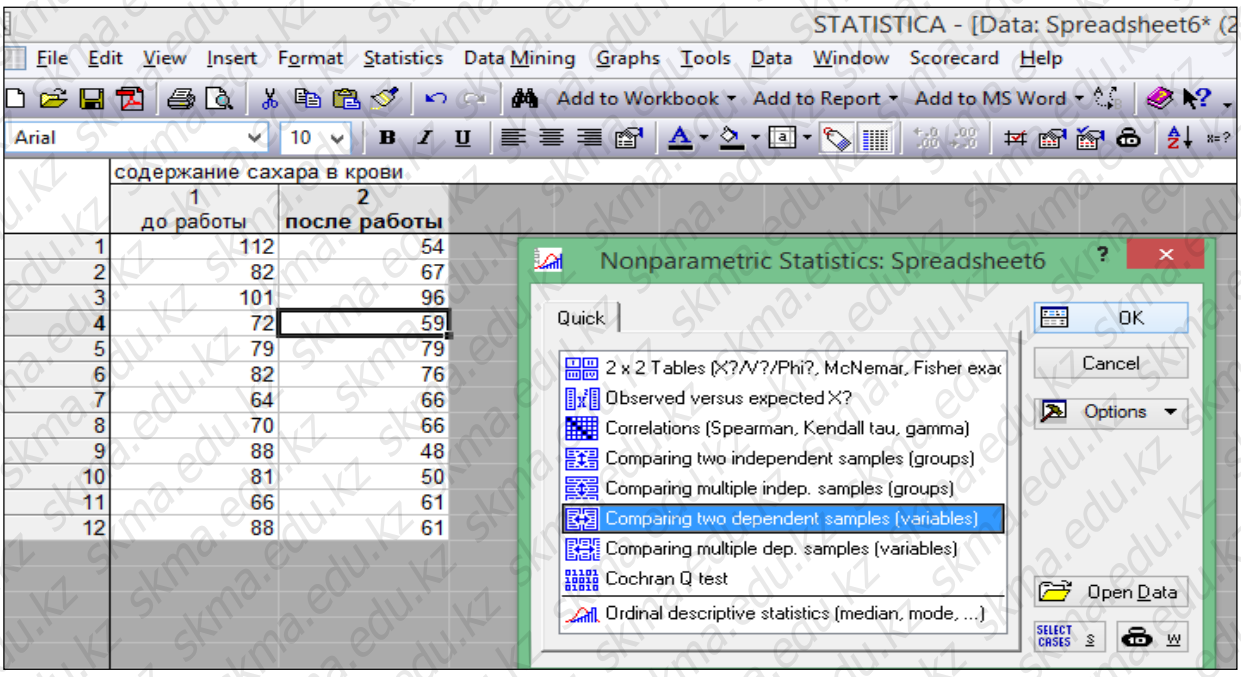


Figure 1. Selecting the procedure Comparing two dependent samples (variables)

3. In the dialog box, click the "Variables" button, indicate the first variable on the left side of the window, and the second variable on the right side of the window (Figure 2) and click the "OK" button.

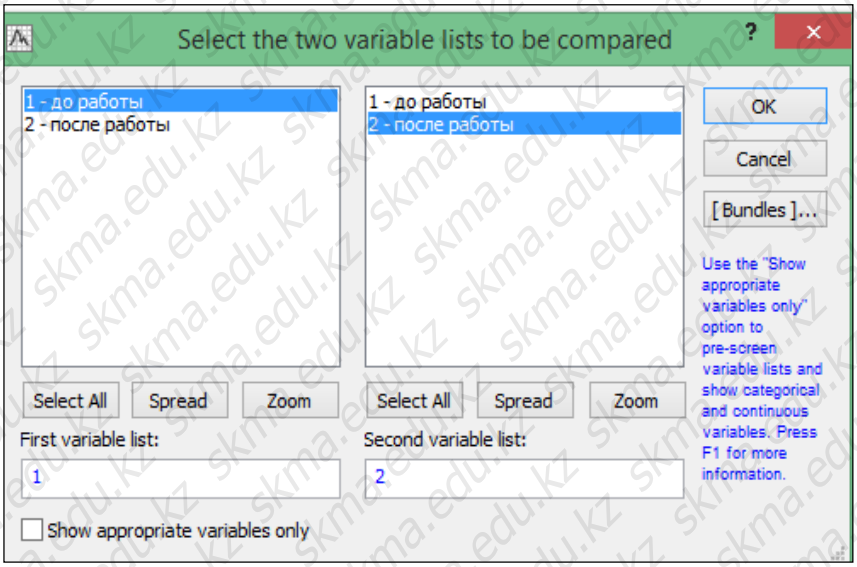


Figure 2. Setting variables

4. In the dialog box, click the “Wilcoxon matched pairs test” button.
A summary table will appear on the screen (Figure 3).

Wilcoxon Matched Pairs Test (Spreadsheet6)
Marked tests are significant at p <,05000

Pair of Variables	Valid N	T	Z	p-value
до работы & после работы	11	1,000000	2,845147	0,004439

Figure 3. Final table with the results of comparison of two dependent samples using the Wilcoxon test

The results obtained can be interpreted as follows:

- if p<0.05, then the null hypothesis is rejected (such results are highlighted in red font);
- if p>0.05, then the null hypothesis is accepted.

In this example, p = 0.004439, which means the null hypothesis of equality of means is rejected, i.e. There is a difference in the blood sugar levels of workers before and after work.

*5.*For clarity, you can build a “box with whiskers” plot by clicking the “Box & whisker plots for all variables” button in the analysis window (Figure 4).

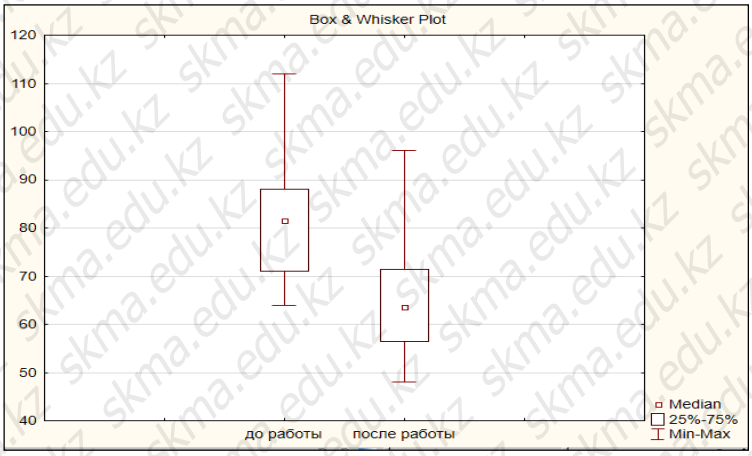


Figure 4. Box-and-whisker plot

Tasks:

1. Using the Mann-Whitney U test, compare the level of intelligence of students in two groups. The initial data are shown in the table.

	IQ scores									
1 group	112	105	109	90	130	117	117	125	134	109
2nd group	121	120	134	119	115	106	107	101	97	117

2. Using the Mann-Whitney U test, compare the duration of wakefulness in the first hour of life of infants born using the conventional and Leboyer methods. The initial data are shown in the table.

Childbirth according to the usual method	Childbirth using the Leboyer method
5.0	2.0
10.1	19.0
17.7	29.7
20.3	32.1
22.0	35.4
24.9	36.7
26.5	38.5
30.8	40.2
34.2	42.1
35.0	43.0
36.6	44.4
37.9	45.6
40.4	46.7
45.5	47.1
49.3	48.0
51.1	49.0
53.1	50.9
55.0	51.2
56.7	52.5
58.0	53.3

3. Using the Mann-Whitney U test, compare systolic blood pressure (SBP) (in mm hg) in two homogeneous groups of healthy men:

- group 1 - persons with many years of work experience in conditions of disturbed sleep and wakefulness;
- group 2 - persons without disturbances in the circadian rhythm of sleep and wakefulness.

The initial data are shown in the table.

	SBP										
1 group	90	95	100	105	120	135	135	135	140	140	145
2nd group	110	115	115	122	122	125	130	150			

4. Using the Wilcoxon W test, check the effectiveness of a special diet to get rid of excess weight. The initial data are shown in the table.

No.	Weight (kg) before experiment	Weight (kg) after experiment
1	93.2	88.9
2	98.2	94.5
3	105.6	106.1
4	86.8	84.3
5	95.5	92.5

5. Using the Wilcoxon W test, check whether smoking affects platelet function. The initial data are shown in the table.

No.	Platelet aggregation	
	Before smoking	After smoking
1	25	27
2	25	29
3	27	37
4	44	56
5	30	46
6	67	82
7	53	57
8	53	80
9	52	61
10	60	59
11	28	43

6. Using the Wilcoxon W test, check whether a certain drug under study is a diuretic. The initial data are shown in the table.

No.	Daily diuresis, ml	
	Before taking the drug	After taking the drug
1	1490	1600
2	1300	1850
3	1400	1300
4	1410	1500
5	1350	1400
6	1000	1010

6. Assessment methods / technologies: Oral questioning. Practical work. (assessment using a checklist).

7. Literature:

Main:

1. Biostatistics: electronic textbook / K. Zh. Kudabaev [et al.]. - Shymkent: SKFFA, 2015.
2. Rosner Bernard Fundamentals of Biostatistics: Texbook/ B.Rosner - 8th ed.: GENGAGE learning, 2016.
3. Armitage P. Encyclopedia of Biostatistics. - Wiley, 2016. - 6100 p
4. Le C.T. Introductory biostatistics. - Wiley, 2013. - 536 p
5. Newman S.C. Biostatistical Methods in Epidemiology. - Wiley, 2017. - 388p
6. http://www.biometrica.tomsk.ru

8. Control:

1. What is the difference between parametric and nonparametric statistical tests?
2. Why is the Mann-Whitney test called an analogue of the two-sample Student t-test?
3. What conditions must be met when using the Mann-Whitney test?
4. Why is the Wilcoxon test called an analogue of the paired Student t-test?
5. What conditions must be met when using the Wilcoxon test?

Lesson №6

1. The theme: Aanalysis of quality features.

2. The aim: formation of an idea of qualitative characteristics, contingency tables, familiarization with some of the capabilities of the STATISTICA program

2. The learning objectives:

- understand the difference between qualitative and quantitative characteristics;
- learn to present qualitative data using contingency tables;
- understand in what cases Pearson's $\chi 2$ test is used;
- learn to formulate null and alternative hypotheses;
- understand in what cases Pearson's $\chi 2$ test is used for tables of size 2x2;
- find out why the Yates amendment is needed;
- understand in what cases McNemar's $\chi 2$ test is used;
- master algorithms of criteria;
- learn to interpret the results;

4. Key questions of the theme:

1. Quantitative and qualitative characteristics.
2. Contingency tables.
3. Conditions for applying the Pearson $\chi 2$ test.
4. Scheme for applying the Pearson $\chi 2$ test.
5. Contingency tables of size 2x2.
6. Scheme for applying the Pearson $\chi 2$ test for tables of size 2x2.
7. Yates Amendment.
8. Scheme for applying the McNemar $\chi 2$ test.

5. Methods / technologies of teaching and learning: Computer-based work / Solving situational tasks.

❖ **Tasks:**

There are data on the number of observations and cases of mortality for four forms of acute purulent destruction of the lungs. Using Pearson's $\chi 2$ test, it is necessary to assess the significance of the difference between groups in the number of deaths.

№ group	Form diseases	Number of cases		Number of patients
		Lethal outcomes	recovery	
1	Purulent abscess	5	136	141
2	Gangrenous abscess	11	37	48
3	Gangrene of the lobes	7	8	15
4	Total gangrene	6	5	11

Solution.

1) H_0: there is no connection between the signs
H_1: there is a connection between the signs
2) p=0.05 – significance level

3) $\nu_{ij}^{*} = \nu_{.i} \cdot \frac{\nu_{.j}}{\nu_{..}}$

$$\frac{(\nu_{ij} - \nu_{ij}^{*})^2}{\nu_{ij}^{*}}$$

3.1) Calculate theoretical frequencies

	B_1	B_2	Total
A_1	5	136	141
A_2	11	37	48
A_3	7	8	15
A_4	6	5	11
Total	29	186	215

	B_1	B_2	Total
A_1	29*141/215=19	186*141/215=122	141
A_2	29*48/215=6,5	186*48/215==41,5	48
A_3	29*15/215=2	186*15/215=13	15
A_4	29*11/214=1,5	186*11/214=9,5	11

3.2) Let's calculate the quantities

	B_1	B_2
A1	(5-19)2/19=10.3	(136-122)2/122=1.6
A2	(11-6.5)2/6.5=3.1	(37-41.5)2/41.5=0.5
A3	(7-2)2/2=12.5	(8-13)2/13=1.9
A4	(6-1.5)2/1.5=13.5	(5-9.5)2/9.5=2.1

3.3) Let's calculate

$$\chi^2_{расч} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(\nu_{ij} - \nu_{ij}^{*})^2}{\nu_{ij}^{*}} = 10{,}3 + 3{,}1 + 12{,}5 + 13{,}5 + 1{,}6 + 0{,}5 + 2 + 2{,}1 = 45{,}5$$

4) $\chi^2_{табл}(p, f)$, where f =(r-1)(s-1) =(4-1)(2-1)=3 - number of degrees of freedom

$\chi^2_{табл}(0{,}05;3) = 7{,}8$ (see table 2)

table 2

Число степеней свободы	Уровень значимости p					
	0,01	0,05	0,1	0,90	0,95	0,99
1	6,6	3,8	2,71	0,02	0,004	0,0002
2	9,2	6,0	4,61	0,21	0,1	0,02
3	11,3	7,8	6,25	0,58	0,35	0,12
4	13,3	9,5	7,78	1,06	0,71	0,30
5	15,1	11,1	9,24	1,61	1,15	0,55
6	16,8	12,6	10,6	2,20	1,64	0,87
7	18,5	14,1	12,0	2,83	2,17	1,24
8	20,1	15,5	13,4	3,49	2,73	1,65
9	21,7	16,9	14,7	4,17	3,33	2,09
10	23,2	18,3	16,0	4,87	3,94	2,56
11	24,7	19,7	17,3	5,58	4,57	3,05
12	26,2	21,0	18,5	6,30	5,23	3,57
13	27,7	22,4	19,8	7,04	5,89	4,11
14	29,1	23,7	21,1	7,79	6,57	4,66
15	30,6	25,0	22,3	8,5	7,26	5,23

16	32,0	26,3	23,5	9,31	7,98	5,81
17	33,4	27,6	24,8	10,1	8,67	6,41
18	34,8	28,9	26,0	10,9	9,39	7,01
19	36,2	30,1	27,2	11,7	10,1	7,63
20	37,6	31,4	28,4	12,4	10,9	8,26
21	38,9	32,7	29,6	13,2	11,6	8,90
22	40,3	33,9	30,6	14,0	12,63	9,54
23	41,6	35,2	32,0	14,8	13,1	10,2
24	43,0	36,4	33,2	15,7	13,8	10,9
25	44,3	37,7	34,4	16,5	14,6	11,5
26	45,6	38,9	35,6	17,3	15,4	12,2
27	47,0	40,1	36,7	18,1	16,2	12,9
28	48,3	41,3	37,9	18,9	16,9	13,6
29	49,6	42,6	39,1	19,8	17,7	14,3
30	50,9	43,8	40,3	20,6	18,5	15,0

5) If $\chi^2_{расч} > \chi^2_{табл}$ then "H0" is rejected, which means the differences between groups in the number of deaths are statistically significant.

An example of applying the Pearson χ2 criterion (2x2)

The relationship between maternal use of contraceptive pills and jaundice in breastfed children is being investigated. The data for the study are presented in the table.

Mother taking pills	There is jaundice	No jaundice	Total
Took pills	33	24	57
Didn't take pills	14	45	59
Total	47	69	116

Solution.

1) H_0: jaundice in children does not depend on the mother taking contraceptive pills
   H_1: jaundice in children depends on the mother taking contraceptive pills
2) α=0.05 – significance level
3) $\chi^2_{расч} = \frac{(ad-bc)^2 n}{(a+b)(c+d)(a+c)(b+d)} = \frac{(33\cdot 45 - 24\cdot 14)^2 \cdot 116}{57\cdot 59\cdot 47\cdot 69} = 14{,}04$
4) $\chi^2_{табл}(0{,}05;1) = 3{,}8$ (see table 2)
5) Because $\chi^2_{расч} > \chi^2_{табл}$, then the hypothesis of independence between jaundice and taking contraceptive pills is rejected, i.e. there is a dependency.

Yates Amendment

The above formula for χ2 in the case of a 2x2 table gives overestimated values. In practice, this results in the null hypothesis being rejected too often. To compensate for this effect, the Yates correction is introduced into the formula:

$$\chi^2_{расч} = \frac{n\left(ad - bc - \frac{n}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

For the example considered above, the calculated value of the criterion with the Yates correction

$$\chi^2_{расч} = \frac{n\left(ad - bc - \frac{n}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{116\left(33 \cdot 45 - 24 \cdot 14 - \frac{116}{2}\right)^2}{57 \cdot 59 \cdot 47 \cdot 69} = 12{,}66$$

Example of application of McNemar's χ2 test

Example 3. The effectiveness of a metabolite-type probiotic in complex therapy for complicated mixed respiratory viral infection and its effect on intestinal microbiocenosis are being investigated. 32 patients took part in the study. The data for the study are presented in the table.

	After probiotic treatment	
Before probiotic treatment	No dysbacteriosis	There is dysbacteriosis
There is dysbacteriosis	9	5
No dysbacteriosis	18	0

Example of application of McNemar's χ2 test

Solution.

1) H_0: the incidence of dysbacteriosis after using the prebiotic did not change.
   H_1: the incidence of dysbacteriosis after using the prebiotic changed.
2) α=0.05 – significance level

3) $\chi^2_{расч} = \frac{(|a-d|-1)^2}{(a+d)} = \frac{(|9-0|-1)^2}{(9+0)} = 7{,}11$

4) $\chi^{2keste}(0.05;1)=3.8$ (see table 2)

5) Because $\chi^2_{расч} > \chi^2_{табл}$, then the hypothesis that the number of patients did not change after the use of the pribiotic is rejected.

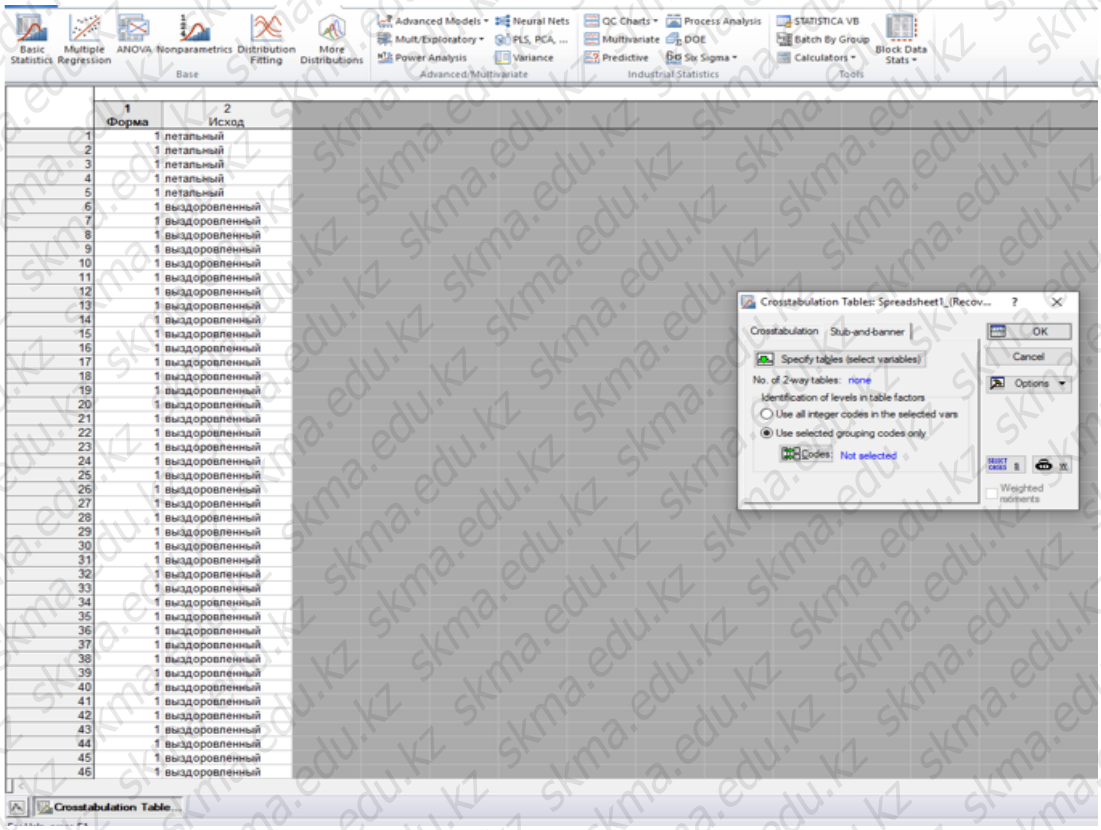
Example 4. There are data on the number of observations and cases of mortality for four forms of acute purulent destruction of the lungs. Using Pearson's χ2 test, it is necessary to assess the significance of the difference between groups in the number of deaths.

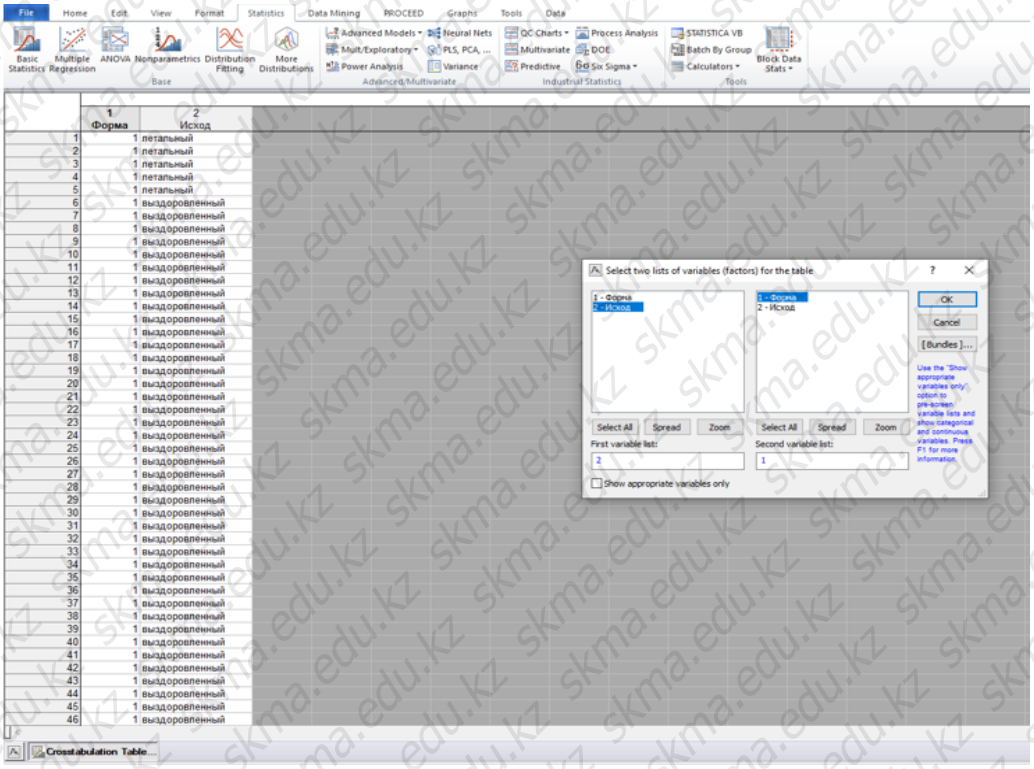
№ group	Form diseases	Number of cases		Number of patients
		Lethal outcomes	recovery	
1	Purulent abscess	5	136	141
2	Gangrenous abscess	11	37	48
3	Gangrene of the lobes	7	8	15
4	Total gangrene	6	5	11

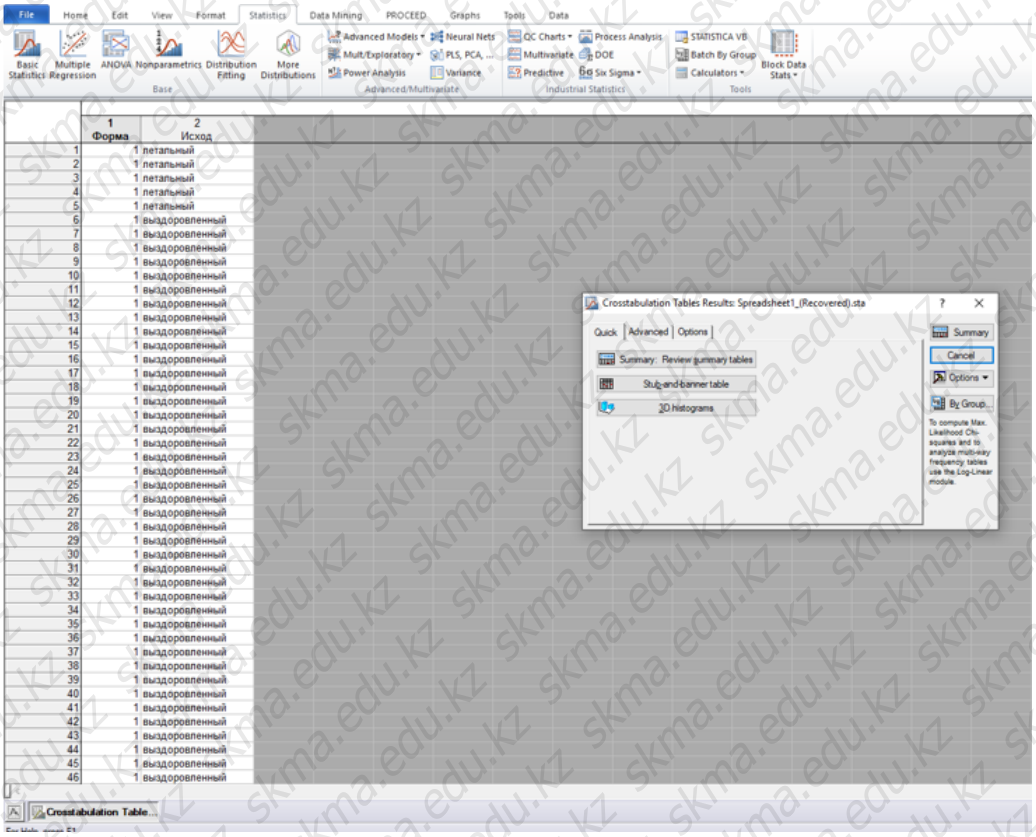
1. Create a table in the Statistica program with a size of 2*215 and enter the initial data.

	1 Форма	2 Исход
1	1	летальный
2	1	летальный
3	1	летальный
4	1	летальный
5	1	летальный
6	1	выздоровлени
7	1	выздоровлени
8	1	выздоровлени
9	1	выздоровлени
10	1	выздоровлени
11	1	выздоровлени
12	1	выздоровлени
13	1	выздоровлени
14	1	выздоровлени
15	1	выздоровлени
16	1	выздоровлени
17	1	выздоровлени
18	1	выздоровлени
19	1	выздоровлени
20	1	выздоровлени
21	1	выздоровлени

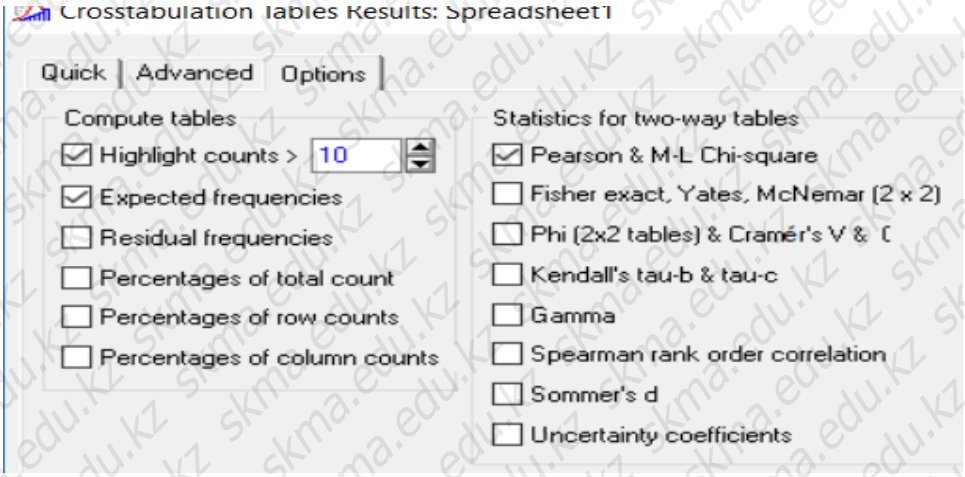
2. Open the “Basic statistics and tables” module, select the “Tables and banners” tab. Make a contingency table.







3. Options “Expected frequencies”, “Pearson & ML Chi-square”



Summary Frequency Table (Spreadsheet1)
Marked cells have counts > 10
(Marginal summaries are not marked)

Форма	Исход выздоровление	Исход летальный	Row Totals
1	136	5	141
2	37	11	48
3	8	7	15
4	5	6	11
All Grps	186	29	215

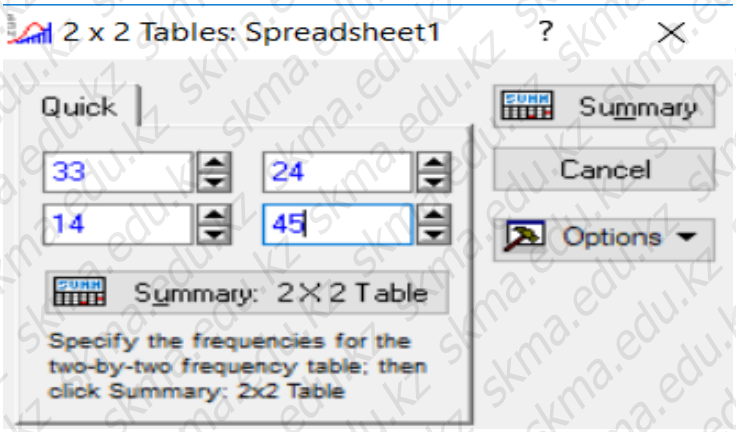
Summary Table: Expected Frequencies (Spreadshee
Marked cells have counts > 10
Pearson Chi-square: 45,6414, df=3, p=,000000

Форма	Исход выздоровление	Исход летальный	Row Totals
1	121,9814	19,01860	141,0000
2	41,5256	6,47442	48,0000
3	12,9767	2,02326	15,0000
4	9,5163	1,48372	11,0000
All Grps	186,0000	29,00000	215,0000

Conclusion: P <0.05, which means the hypothesis is rejected, that is, there is a relationship between the characteristics.

Example 2. Implementation of the calculated value of the characteristic with the Yates correction in the Statistica program for the above example.

1. Open the “Nonparametrics” module, select the “2x2 tables XI/VI/Phil, McNemar, Fisher exact” tab
2. Enter the data, click the “Summary” button



	2 x 2 Table (Spreadsheet1)		
	Column 1	Column 2	Row Totals
Frequencies, row 1	33	24	57
Percent of total	28,448%	20,690%	49,138%
Frequencies, row 2	14	45	59
Percent of total	12,069%	38,793%	50,862%
Column totals	47	69	116
Percent of total	40,517%	59,483%	
Chi-square (df=1)	14,04	p= ,0002	
V-square (df=1)	13,92	p= ,0002	
Yates corrected Chi-square	12,66	p= ,0004	
Phi-square	,12105		
Fisher exact p, one-tailed		p= ,0002	
two-tailed		p= ,0003	
McNemar Chi-square (A/D)	1,55	p= ,2130	
Chi-square (B/C)	2,13	p= ,1443	

Conclusion: $p<0.05$, which means the hypothesis “H0” is rejected.

Tasks

1. The influence of the learning process on the results of a certain psychological test is being studied. Tests carried out on 100 schoolchildren revealed the results presented in the table below. Using the $\chi 2$ criterion, it is required to investigate the presence of an influence of training on test results.

Age of schoolchildren	Test results			
	Low	Average	High	Total
Juniors	10	15	5	30
Average	6	16	8	30
Elder	7	13	20	40
Total	23	44	33	100

Note. **Check your results in the STATISTICA program.**

2. The table below provides information on newly diagnosed cases of primary and secondary syphilis, age and race rates per 100,000 population, USA,1989.Using the $\chi 2$ test, it is required to

examine the presence of an effect of race on morbidity.

Age group	Racial rate per 100,000			
(years)	White race	Black race	Other races	Total
< 14	0.0	2.4	0.8	0.4
15-19	2.4	131.5	51.0	24.3
20-24	5.8	323.0	139.2	55.9
25-29	5.4	270.9	117.9	44.1
30-34	4.7	256.6	83.2	38.8
35-44	2.9	135.0	47.8	19.0
45-54	1.7	76.7	29.6	10.5
>55	0.5	19.4	10.4	2.4
Total	2.2	115.8	45.8	17.7

Note. Check your results in the STATISTICA program**.**

3. Scientists compared the effectiveness of three antibiotics for recurrent urinary tract infections in girls 3-16 years old. After a short course of one of the antibacterial drugs (assigned at random), repeated urine cultures were taken over the course of a year. When bacteriuria was detected, a relapse was noted. The research results are presented in the table below. Using the $\chi 2$ criterion, it is necessary to study the effectiveness of drugs.

A drug	Relapse	
	Eat	No
Ampicillin	20	7
Trimethoprim/sulfamethoxazole	24	21
Cephalexin	14	5

Note. Check your results in the STATISTICA program**.**

4. The effectiveness of vaccination against typhus is being studied. The data is presented in the table below. Using the $\chi 2$ criterion, draw conclusions about the effectiveness of vaccinations.

Observed values	Number of cases	Number of non-sick people	Total
With vaccination	72	7988	8060
Without vaccination	303	9322	9625
Total	375	17310	17685

Note. Check your results in the STATISTICA program.

5. The table below shows the classification data1000 people based on color blindness.Using the $\chi 2$ test,check, is there a relationship between the presence of color blindness and a person's gender.

Observed values	Men	Women	Total
Colorblind	38	6	44
Not colorblind	442	514	956
Total	480	520	1000

Note. Check your results in the STATISTICA program.

6. It has been suggested that the skin reaction to dinitrochlorobenzene can serve as an assessment of the integrity of the patient's immunity. To test this assumption, we decided to compare the reaction to dinitrochlorobenzene with the reaction to croton oil, which has only a local irritant effect. To do this, a group of patients were applied to adjacent areas of clean skin with the agents being compared, and the reactions were recorded. The results are presented in the table below.

Reaction to croton oil	Reaction to dinitrochlorobenzene		Total
	Eat	No	
There is	5	48	53

No	23	10	33
Total	28	58	86

Note. Check your results in the STATISTICA program**.**

7. There are data on mortality among patients suffering from open pulmonary tuberculosis during the first year after the disease.

It is required to establish whether the difference in mortality between men and women is significant.

Observed values	Alive	Died
Men	53	8
Women	11	10

Note. Check your results in the STATISTICA program**.**

8. There is evidence of the sociability of soldiers recruited from cities and rural areas. Test the hypothesis that city dwellers are more sociable than rural dwellers.

Observed values	Sociable	Unsociable
City	10	14
Countryside	6	16

Note. Check your results in the STATISTICA program**.**

9. The effectiveness of high-frequency nerve stimulation as a pain reliever during tooth extraction is being studied. All patients were connected to the device, but in some cases it worked, in others it was turned off. Neither the dentist nor the patient knew whether the device was turned on. Do the following data suggest that high-frequency nerve stimulation is an effective analgesic agent?

Observed values	The device is turned on	The device is turned off
No pain	20	5
There is pain	16	12

Note. Check your results in the STATISTICA program**.**

6. Assessment methods / technologies: Oral questioning. Practical work. (assessment using a checklist).

7. Literature:

Main:

1. Biostatistics: electronic textbook / K. Zh. Kudabaev [et al.]. - Shymkent: SKFFA, 2015.
2. Rosner Bernard Fundamentals of Biostatistics: Texbook/ B.Rosner - 8th ed.: GENGAGE learning, 2016.
3. Armitage P. Encyclopedia of Biostatistics. - Wiley, 2016. - 6100 p
4. Le C.T. Introductory biostatistics. - Wiley, 2013. - 536 p
5. Newman S.C. Biostatistical Methods in Epidemiology. - Wiley, 2017. - 388p
6. http://www.biometrica.tomsk.ru

8. Control:

1. What is special about the analysis of qualitative features?
2. What is a contingency table of size mxn?
3. What conditions must be met when applying Pearson's $\chi 2$ test?
4. Why is the Yates correction calculated?
5. To what cases does McNemar's criterion apply?
6. In what ways can it be implementedcriterion$\chi 2$-Pearsonin a programme "STATISTICA"?
7. How is the information contained in the summary table interpreted?

Lesson №7

1. **The theme:** Correlation analysis
2. **The aim:** studying the principles of correlation analysisand the reliability of the obtained rank correlation coefficient when assessing risk factors for developing the disease
3. **The learning objectives:**find and study information on the following issues:
   - types of correlation dependence;
   - Pearson pair correlation coefficient;
   - assessment of the reliability of the correlation coefficient.
   - conducting correlation analysis;
   - calculate Spearman's rank correlation coefficient
4. **Key questions of the theme:**
   1. What is Correlation Analysis?
   2. How is it located?Pearson's pairwise correlation coefficient?
   3. How is the correlation coefficient determined?
   4. What is the reference correlation coefficient used to estimate the relationship?
   5. How p is calculatedSpearman's ang correlation coefficient?
   6. What data will be included in the final table of correlation and regression analysis in Statistica?
5. **Methods / technologies of teaching and learning:** Computer-based work / Solving situational tasks.

❖ **Tasks:**

Example 1. For the following data, calculate the linear Pearson correlation coefficient:

Incidence of acute respiratory infections per 1000 population, x	352	228	340	300	196	258	237
Incidence of pneumonia per 1000 population, in	64	60	52	48	46	41	32

***Solution*:**

1) Create a calculation table:

No.	X	U	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})\cdot(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
1	352	64	79	15	1185	6241	225
2	228	60	-45	11	-495	2025	121
3	340	52	67	3	201	4489	9
4	300	48	27	-1	-27	729	1
5	196	46	-77	-3	231	5929	9
6	258	41	-15	-8	120	225	64
7	237	32	-36	-17	612	1296	289
Sum	1911	343	0	0	1827	20934	718
Average	273	49					

2) Calculate the correlation coefficient:

$$r_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} = \frac{1827}{\sqrt{20934 \cdot 718}} = 0{,}47 .$$

3) Analyze the result obtained: the relationship between the characteristics under consideration is direct and moderate.

4) Calculate the average error of the correlation coefficient:

$$m_r = \pm\frac{1-r_{xy}^2}{\sqrt{n}} = \pm\frac{1-0{,}47^2}{\sqrt{7}} = 0{,}3,$$

The correlation coefficient is not reliable, because does not exceed its average error by three times.

Example 2. In a certain locality, a chronic epidemic of Flexner's dysentery has been reported. Laboratory studies have shown that in the drinking water of the water supply network there are frequent "slips" of non-standard samples according to bacteriological indicators (risk factor). It is necessary to check whether there is a connection between these two signs.

Month	Number of patients with dysentery (x)	Proportion of non-standard water samples (y)
January	10	0
February	9	0.5
March	2	1.1
April	7	2.0
May	6	1.8
June	11	2.9
July	26	6.7
August	32	4.5
September	46	8.7
October	38	7.1
November	8	3.2
December	5	0

***Solution*:**

1) Create a calculation table:

No.	X	at	r_x	r_y	$r_x - r_y$	$(r_x - r_y)^2$
1	2	0	7	1.5	5.5	30.25
2	9	0.5	6	3	3	9
3	2	1.1	1	4	-3	9
4	7	2.0	4	6	-2	4
5	6	1.8	3	5	-2	4
6	11	2.9	8	7	1	1
7	26	6.7	9	10	-1	1
8	32	4.5	10	9	1	1
9	46	8.7	11	12	-1	1
10	38	7.1	12	11	1	1
11	8	3.2	5	8	-3	9
12	5	0	2	1.5	0.5	0.25
Sum						70.5

2) Calculate the correlation coefficient:

$$\rho = 1 - \frac{6}{n^3 - n}\sum_{i=1}^{n}\left(r_{x_i} - r_{y_i}\right)^2 = 1 - \frac{6}{12^3 - 12} \cdot 70.5 \approx 0{,}75 .$$

3) Analyze the result obtained: the connection between the characteristics under consideration is direct and high.

4) Calculate the average error of the correlation coefficient:

$$m_r = \pm\frac{1 - r_{xy}^2}{\sqrt{n}} = \pm\frac{1 - 0{,}75^2}{\sqrt{12}} \approx 0{,}12 ,$$

The correlation coefficient is reliable because exceeds its average error by more than three times.

Example 3. In one locality, a chronic epidemic of Flexner's dysentery was registered. Preliminary analysis and laboratory studies have shown that in the drinking water of the water supply network there are frequent "slips" of non-standard samples according to bacteriological indicators (risk factor). Calculate Spearman's rank correlation coefficient. The initial data is presented in the table.

Month	Number of patients with dysentery (n)	Proportion of non-standard water samples (x)
January	10	0
February	9	0.5
March	2	1.1
April	7	2.0
May	6	1.8
June	11	2.9
July	26	6.7
August	32	4.5
September	46	8.7
October	38	7.1
November	8	3.2
December	5	0

1. Select Statistics → Nonparametrics → Correlations (Spearman, Kendall tau, gamma) (Figure 1), click the "OK" button.

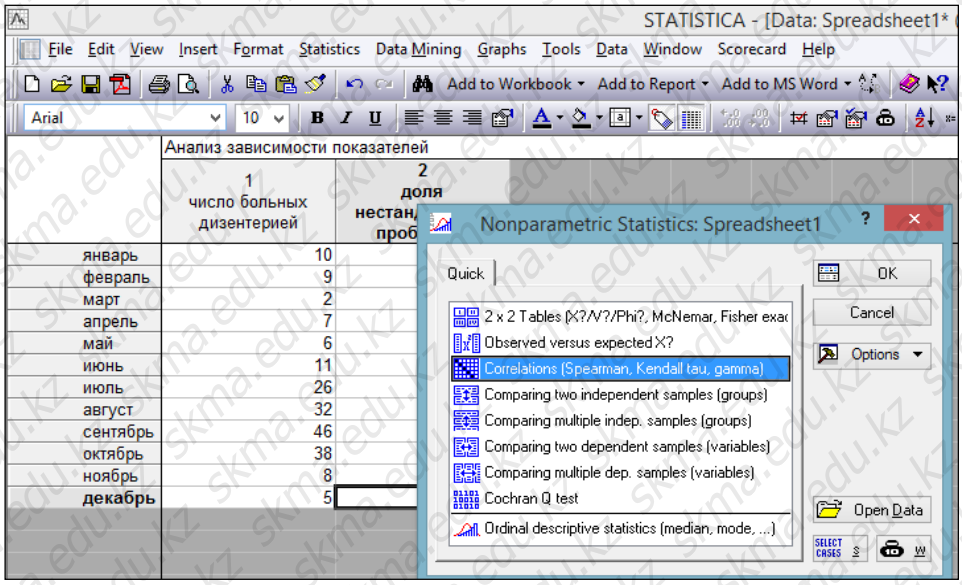


Figure 1. Selecting the "Correlations (Spearman, Kendall tau, gamma)" procedure

2. Set the variables "Variables" (Figure 2), click the "OK" button.

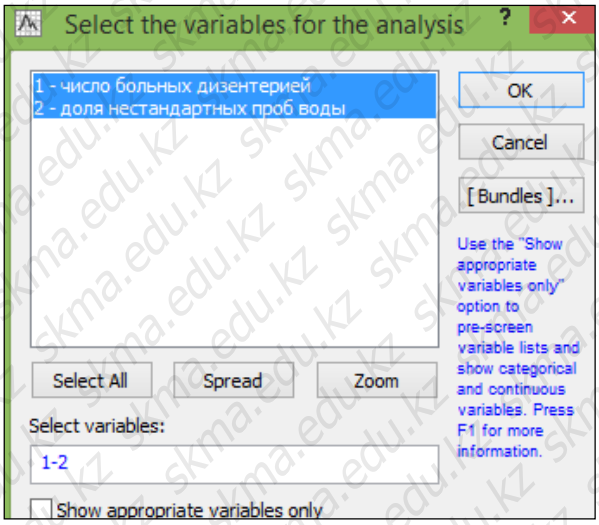


Figure 2. Setting variables

3. Press the "Spearman rank R" button.

The result of calculating the Spearman rank correlation coefficient is the following table (Figure 3).

The value of the rank correlation coefficient of 0.760071 indicates that the relationship between the characteristics is direct and high.

Spearman Rank Order Correlations (Spreadsheet1)
MD pairwise deleted
Marked correlations are significant at p <,05000

Variable	число больных дизентерией	доля нестандартных проб воды
число больных дизентерией	1,000000	0,760071
доля нестандартных проб воды	0,760071	1,000000

Figure 3. Calculation result of Spearman's rank correlation coefficient

Tasks

1. The relationship between the incidence of myocardial infarction by month of the year and the average monthly air temperature is studied. The initial data are presented in the table.

Month	Incidence of myocardial infarction by month (per 10,000 thousand inhabitants)	Average monthly air temperature
January	1.6	-7.1
February	1.23	-7.7
March	1.14	-5.8
April	1.13	-4.1
May	1.12	+13
June	1.02	+14.9
July	0.91	+18.8
August	0.82	+15.6
September	1.06	+9.0

October	1.22	+6.0
November	1.33	-1.0
December	1.4	-7.7

Calculate the Pearson correlation coefficient, determine the direction, strength and reliability of the relationship between indicators. Draw a conclusion.

2. The relationship between systolic blood pressure (SBP) and weight in women aged 20 to 30 years is studied. The initial data is presented in the table.

SBP, at	110	125	80	120	115	140	120	110	85
Weight, (kg)x	53	60	58	55	68	70	64	55	55

Calculate the Pearson correlation coefficient, determine the direction, strength and reliability of the relationship between indicators. Draw a conclusion.

3. The relationship between the duration of smoking and the number of respiratory tract diseases is studied. The initial data is presented in the table.

Smoking experience (years), x	2	4	5	1	3	2	4	5	6
Number of diseases	5	6	4	1	2	5	5	6	4

Calculate the Spearman correlation coefficient, determine the direction, strength and reliability of the relationship between indicators. Draw a conclusion.

4. The relationship between the thickness of the skin scar and the time of its freezing for cryodestruction is being studied. The initial data is presented in the table.

Skin scar thickness (mm), x	3	5	8	9	12	14	17	20
Freezing of the skin scar during cryodestruction (min.), y	0.6	1	1.6	1.5	1.7	1.6	2.4	3

Calculate the Spearman correlation coefficient, determine the direction, strength and reliability of the relationship between indicators. Draw a conclusion.

5. The relationship between the duration of contact of workers with industrial poisons and the incidence of toxic hepatitis is studied. The data is presented in the table.

Duration of work (years), x	Up to 1	2	3	4	5	6
Morbidity (‰), y	2	8	7	11	10	13

Calculate the Spearman correlation coefficient, determine the direction, strength and reliability of the relationship between indicators. Draw a conclusion.

6. In one of the districts of the West Kazakhstan region, cases of anicteric leptospirosis (water fever) appeared. A study was carried out on the relationship between the number of diseases and the amount of precipitation. The data is presented in the table.

Number of diseases, x	0	19	4	1	2	68	131	14	11	2
Precipitation amount (mm), y	54	101	185	85	30	128	143	74	28	132

Calculate Spearman's rank correlation coefficient, determine the direction, strength and reliability of

the relationship between indicators. Draw a conclusion.

7. The relationship between the thickness of the coal seam and the incidence of hypertension among miners is studied. The data is presented in the table.

Seam thickness (m), x	0.6	0.8	1.0	1.2	1.4	1.5	1.6
Incidence of hypertension (per 1000 miners), in	3.5	4.2	5.6	6.3	7.4	8.9	10

Calculate Spearman's rank correlation coefficient, determine the direction, strength and reliability of the relationship between indicators. Draw a conclusion.

6. Assessment methods / technologies: Oral questioning. Practical work. (assessment using a checklist).

7. Literature:

Main:

1. Biostatistics: electronic textbook / K. Zh. Kudabaev [et al.]. - Shymkent: SKFFA, 2015.
2. Rosner Bernard Fundamentals of Biostatistics: Texbook/ B.Rosner - 8th ed.: GENGAGE learning, 2016.
3. Armitage P. Encyclopedia of Biostatistics. - Wiley, 2016. - 6100 p
4. Le C.T. Introductory biostatistics. - Wiley, 2013. - 536 p
5. Newman S.C. Biostatistical Methods in Epidemiology. - Wiley, 2017. - 388p
6. http://www.biometrica.tomsk.ru

8. Control:

1. How is linear correlation coefficient calculated?
2. What types of correlation do you know?
3. How is the reliability of the correlation coefficient determined?
4. What is correlation analysis?
5. What is the reference correlation coefficient used to estimate the relationship?
6. How is p calculatedSpearman's ang correlation coefficient?
7. What data will be included in the final table of correlation and regression analysis in Statistica?

Lesson №8

1. The theme: Regression analysis

2. The aim: Formation of skills in drawing up a linear regression equation using the least squares method, checking the significance of regression coefficients and finding the coefficient of determination.

3. The learning objectives:

- learn to create a linear regression equation using the least squares method;
-check the significance of regression coefficients, check the significance of the regression equation;
-developing the ability to teach how to find the coefficient of determination.

4. Key questions of the theme:

1. What is the main meaning of regression analysis?
2. What is regression and what are its types?
3. What types of pairwise regression equations do you know?
4. What is the main meaning of least squares?
5. What formula is used to determine the coefficients of a paired regression equation?

5. Methods / technologies of teaching and learning: Computer-based work / Solving situational tasks.

❖ **Tasks:**

Example 1. Based on the following data, construct and analyze a regression equation:

Incidence of influenza per 1000 population, x	352	228	340	300	196	258	237
Incidence of pneumonia per 1000 population, in	64	60	52	48	46	41	32

Solution.

1) Calculate the correlation coefficient: $r_{xy} = 0{,}47$. The relationship between the signs is direct and moderate.

2) Construct a paired linear regression equation.

2.1) Create a calculation table.

No.	*X*	*U*	*xy*	*x2*	y_x	$(y_x - \bar{y})^2$	$(y-y_x)^2$	$(x - \bar{x})^2$
1	352	64	22528	123904	55.89	47.54	65.70	6241
2	228	60	13680	51984	45.07	15.42	222.83	2025
3	340	52	17680	115600	54.85	34.19	8.11	4489
4	300	48	14400	90000	51.36	5.55	11.27	729
5	196	46	9016	38416	42.28	45.16	13.84	5929
6	258	41	10578	66564	47.69	1.71	44.77	225
7	237	32	7584	56169	45.86	9.87	192.05	1296
Sum	1911	343	95466	542637	343	159.45	558.55	20934
Average	273	49	13638	77519.6	49	22.78	79.79	2990.6

2.2) Calculate regression coefficients:

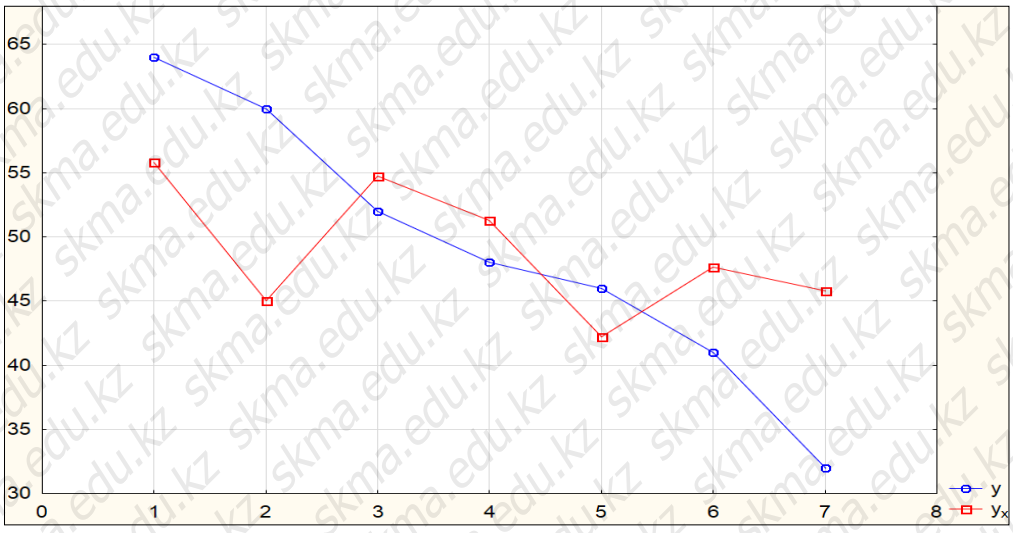
$$b = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{13638 - 49 \cdot 273}{77519{,}6 - 273^2} = 0{,}087,$$

$$a = \bar{y} - b\bar{x} = 49 - 0{,}087 \cdot 273 = 25{,}17.$$

Paired linear regression equation: yx=25.17+0.087x.

3) Find the theoretical values of "yx" by substituting the actual values of "x" into the regression equation.

4) Construct graphs of actual "y" and theoretical values of "yy" of the resultant characteristic:



5) Check the statistical significance of regression coefficients:

5.1) Calculate random errors:

$$m_b = \sqrt{\frac{\sum (y - y_x)^2}{n-2} \Big/ \sum (x - \bar{x})^2} = \sqrt{\frac{558{,}55}{7-2} \Big/ 20934} \approx 0{,}073 ;$$

$$m_a = \sqrt{\frac{\sum (y - y_x)^2}{n-2} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = \sqrt{\frac{79{,}8}{7-2} \cdot \frac{542637}{7 \cdot 2990{,}6}} \approx 20{,}34 .$$

5.2)

$$t_{b\ расч} = \frac{b}{m_b} = \frac{0{,}087}{0{,}073} \approx 1{,}19 , \qquad t_{a\ расч} = \frac{a}{m_a} = \frac{25{,}17}{20{,}34} \approx 1{,}24 .$$

5.3) t_{table} (0.05; 5) = 2.57 (see table 1).

Table 5.

the number of degrees of freedom	significance level, α			
f	0,10	0,05	0,02	0,01
1	6,31	12,7	31,82	63,7
2	2,92	4,30	6,97	9,92
3	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,60
5	2,01	2,57	3,37	4,03
6	1,94	2,45	3,14	3,71
7	1,89	2,36	3,00	3,50
8	1,86	2,31	2,90	3,36
9	1,83	2,26	2,82	3,25
10	1,81	2,23	2,76	3,17
11	1,80	2,22	2,72	3,11
12	1,78	2,18	2,68	3,05
13	1,77	2,16	2,65	3,01
14	1,76	2,14	2,62	2,98
15	1,75	2,13	2,60	2,95
16	1,75	2,12	2,58	2,92
17	1,74	2,11	2,57	2.90
18	1,73	2,10	2,55	2,88
19	1,73	2,09	2,54	2,86
20	1,73	2,09	2,53	2,85
30	1,70	2,04	2,46	2,75
40	1,68	2,02	2,42	2,70
60	1,67	2,00	2,39	2,66
120	1,66	1,98	2,36	2,62
∞	1,64	1,96	2,33	2,58

5.4) $t_{b\ calc} < t_{table}$, which means the coefficient "b" is insignificant,

$t_{a\ calc} < t_{table}$, which means the coefficient "a" is insignificant.

6) Check the statistical significance of the regression equation:

6.1) $F_{calc} = \frac{\frac{\sum(y_x - \bar{y})^2}{k}}{\frac{\sum(y - y_x)^2}{n-k-1}} = \frac{\frac{159{,}45}{1}}{\frac{558{,}55}{5}} = 1{,}43.$

6.2) F_{table}(p;k; nk-1)=(0.05; 1; 5)=6.61 (see table 2).

6.3) $F_{calc} < F_{table}$, then the regression equation is chosen incorrectly. This result can be explained by the low closeness of the relationship (r_{xy}=0.47) and the small number of observations.

7) Calculate the coefficient of determination: R_2=(0.47)2=0.22, the constructed equation is of poor quality.

Tasks:

1. The relationship between the incidence of myocardial infarction by month of the year and the average monthly air temperature is studied. The initial data are presented in the table.

Month	Incidence of myocardial infarction by month (per 10,000 thousand inhabitants)	Average monthly air temperature
January	1.6	-7.1
February	1.23	-7.7
March	1.14	-5.8
April	1.13	-4.1
May	1.12	+13
June	1.02	+14.9
July	0.91	+18.8
August	0.82	+15.6
September	1.06	+9.0
October	1.22	+6.0
November	1.33	-1.0
December	1.4	-7.7

Construct and estimate a paired linear regression equation.

2. The relationship between systolic blood pressure (SBP) and weight in women aged 20 to 30 years is studied. The initial data is presented in the table.

SBP, at	110	125	80	120	115	140	120	110	85
Weight, (kg)x	53	60	58	55	68	70	64	55	55

Construct and estimate a paired linear regression equation.

3. The relationship between the duration of smoking and the number of respiratory tract diseases is studied. The initial data is presented in the table.

Smoking experience (years), x	2	4	5	1	3	2	4	5	6
Number of diseases	5	6	4	1	2	5	5	6	4

Construct and estimate a paired linear regression equation.

Table 2

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,52
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

4. The relationship between the thickness of the skin scar and the time of its freezing for cryodestruction is being studied. The initial data is presented in the table.

Skin scar thickness (mm), x	3	5	8	9	12	14	17	20
Freezing of the skin scar during cryodestruction (min.), y	0.6	1	1.6	1.5	1.7	1.6	2.4	3

Construct and estimate a paired linear regression equation.

5. The relationship between the duration of contact of workers with industrial poisons and the incidence of toxic hepatitis is studied. The data is presented in the table.

Duration of work (years), x	Up to 1	2	3	4	5	6
Morbidity (‰), y	2	8	7	11	10	13

Construct and estimate a paired linear regression equation.

6. In one of the districts of the West Kazakhstan region, cases of anicteric leptospirosis (water fever) appeared. A study was carried out on the relationship between the number of diseases and the amount of precipitation. The data is presented in the table.

Number of diseases, x	0	19	4	1	2	68	131	14	11	2
Precipitation amount (mm), y	54	101	185	85	30	128	143	74	28	132

Construct and estimate a paired linear regression equation.

6. Assessment methods / technologies: Oral questioning. Practical work. (assessment using a checklist).

7. Literature:

Main:

1. Biostatistics: electronic textbook / K. Zh. Kudabaev [et al.]. - Shymkent: SKFFA, 2015.
2. Rosner Bernard Fundamentals of Biostatistics: Texbook/ B.Rosner - 8th ed.: GENGAGE learning, 2016.
3. Armitage P. Encyclopedia of Biostatistics. - Wiley, 2016. - 6100 p
4. Le C.T. Introductory biostatistics. - Wiley, 2013. - 536 p
5. Newman S.C. Biostatistical Methods in Epidemiology. - Wiley, 2017. - 388p
6. http://www.biometrica.tomsk.ru

8. Control:

1. What types of pairwise regression equations do you know?
2. What is the main significance of the least squares method?
3. What formula is used to determine the coefficients of a paired regression equation?

Lesson №9

1. The theme: Health systems in Kazakhstan. International cooperation in health care.
2. The aim: to familiarize students with the healthcare system in Kazakhstan.
3. The learning objectives:

The student must know and be able to:

•Basic principles and objectives of the state policy in the field of health care.
•The healthcare system in the Republic of Kazakhstan.
•International cooperation in the field of healthcare.

4. Key questions of the theme:

1. State and non-state health sector.
2. Pros and cons of the healthcare system.
3. Prospects for the development of healthcare.

5. Methods / technologies of teaching and learning: Training cases, question and answer
6. Assessment methods / technologies Assessment using a check list
7. Literature: see Appendix 1.
8. Control (questions):

1. What is included in the healthcare system?
2. What is the healthcare system in Kazakhstan?
3. At what level is medicine in Kazakhstan?
4. What problems exist in the healthcare system?
5. Name the priority directions of international cooperation in the field of health care.

Lesson №10

1. The theme: Methodology for calculating and analyzing medical and demographic indicators.
2. The aim: to familiarize students with the medical and social aspects of demography and talk about the calculation and analysis of demographic indicators in a comprehensive health assessment.
3. The learning objectives:
The student must know and be able to:

- Medico-social aspects of demography.
- The principle of calculation and analysis of demographic indicators.
- Comprehensive assessment of public health.

4. Key questions of the theme:

1) What are the medical and social aspects of demography?
2) What demographics shape public health?
3) How are demographics calculated?
4) What methods are used in the analysis of demographic indicators?

5. Methods / technologies of teaching and learning: Training cases, case-study
6. Assessment methods / technologies Assessment interview using a checklist
7. Literature: see Appendix 1.
8. Control (questions):

1. What current health questions can be answered by examining demographics?
2. What demographics shape public health?
3. What methods are used in the analysis of demographic indicators?
4. What indicators are needed to calculate a comprehensive assessment of public health?

Lesson №11

1. The theme: Modern medical and social problems, issues of strengthening the health of the population.
2. The aim: to familiarize students with the criteria for the incidence of the population.
3. The learning objectives:
The student must know and be able to:

- morbidity of the population.
- Classification of the incidence of the population.
- Quantitative criteria for morbidity according to WHO.
- Methods for calculating the criteria for morbidity (indicator of the frequency of primary morbidity, the prevalence of general morbidity, the incidence rate).

4. Key questions of the theme:

1) What do you understand by population morbidity?
2) On what grounds is the classification of the incidence of the population?
3) What morbidity criteria are quantitative according to WHO?
4) What calculation methods are intended for quantitative criteria of morbidity?
5) According to what principle is the incidence of the population divided into primary, secondary and tertiary?

5. Methods / technologies of teaching and learning: Educational cases, case-study
6. Assessment methods / technologies Assessment using a checklist
7. Literature: see Appendix 1.

8. Control (questions):

1. What morbidity rates do you know?
2. How does the ICD-10 classification affect the definition and calculation of incidence rates?
3. By what criteria does WHO classify incidence rates?
4. What methods of calculating incidence rates are most often used to determine the level of public health?

Lesson №12

1. The theme: Organization and conduct of medical and social expertise.
2. The aim: to familiarize students with the activities of MSE.
3. The learning objectives:
The student must know and be able to:

- Main functions of MSE.
- Criteria for determining disability groups.
- The procedure for appealing the conclusion of MSE.

4.Key questions of the theme:

- Grounds for conducting a medical and social examination.
- Inspection procedure.
- Criteria for establishing disability.

5. Methods / technologies of teaching and learning: Training cases, case-study
6. Assessment methods / technologies Assessment interview using a checklist
7. Literature: see Appendix 1.
8. Control (questions):

1. What is a medical and social expert commission?
2. Who gives direction to MSE?
3. Who is part of MSE?
4. How to challenge the conclusion of MSE?

Lesson №13

1. The theme: Medical care and its types.
2. The aim: to familiarize students with the types of medical care provided and the forms of its provision.
3. The learning objectives:
The student must know and be able to:

- Types of medical care.
- Forms of providing medical care.
- The principle of distribution of children and adolescents by health groups.
- The principle of distribution of adults by health groups.
- Distinctive features of the distribution of adults into 3 groups, and children and adolescents into 5 health groups.

4. Key questions of the theme:

1) What types of medical care are provided by health facilities in the Republic of Kazakhstan?
2) What forms of medical care are used in the Republic of Kazakhstan?
3) What health groups do you know?
4) What is the difference between the distribution of adults and children by health groups?

5. Methods / technologies of teaching and learning: Educational cases, case study

6. Assessment methods / technologies Assessment using a checklist
7. Literature: see Appendix 1.
8. Control (questions):

1. On what basis is medical care provided?
2. What are the general and distinctive features of the distribution of medical care into forms and types?
3. What is the reason for the distribution of health groups in the adult population into 3 groups, and in children - into 5 groups?
4. What are the key distribution criteria for each adult and child health group?

Lesson №14

1. The theme: Medical secrecy.
2. The aim: to acquaint students with medical secrecy, the procedure for providing information and the responsibility for disclosing it.
3. The learning objectives:
The student must know and be able to:
•Definition of the term "confidentiality".
•Hippocratic oath of medical secrecy.
•Legal basis of medical secrecy.
4. Key questions of the theme:

1. Medical secrecy: regulation.
2. Preservation of medical secrecy: medical documents.
3. Conditions for disclosure of medical confidentiality.
4. Punishment for disclosure of medical secrets.

5. Methods / technologies of teaching and learning: Training cases
6. Assessment methods / technologies Assessment using a checklist
7. Literature: see Appendix 1.
8. Control (questions):

1. What is included in the concept of medical secrecy?
2. When is disclosure of medical confidentiality allowed?
3. What is the responsibility for disclosing medical confidentiality?

Lesson №15

1. The theme: The universality of the ethical norm and the uniqueness of the moral choice.
2. The aim: explain to the students the universality of the ethical norm and the uniqueness of the moral choice.
3. The learning objectives:
The student must know and be able to:

- The principle of justice
- The concept of the universality of the ethical norm
- Moral choice and morality
- The uniqueness of moral choice in medicine

4.Key questions of the theme:

1. How are the principle of justice and the universality of the ethical norm combined?
2. What is a moral choice?
3. What is meant by the term "uniqueness"?

4. What is the uniqueness of moral choice in medicine?

5. Methods / technologies of teaching and learning: Educational cases, case study
6. Assessment methods / technologies Assessment using a checklist
7. Literature: see Appendix 1.
8. Control (questions):

1. What are the main criteria for the principle of justice?
2. How do you understand the terms "universality" and "uniqueness"? Give examples.
3. What are the main differences between morality and moral choice?
4. How is the interaction of the principle of justice and the uniqueness of the moral choice?

Lesson №16

1. The theme: Ethical regulations.
2. The aim: familiarize students withnormative-legal acts regulating medical activity.
3. The learning objectives:
The student must know and be able to:

- the Constitution of the Republic of Kazakhstan.
- Code of honor for medical workers.
- Rights and obligations of persons in the field of healthcare.
- Code of the Republic of Kazakhstan "On the health of the people and the healthcare system".

4.Key questions of the theme:

- The Constitution of the Republic of Kazakhstan.
- Ethical principles and requirements for medical workers.
- Principles of state policy in the field of healthcare.
- Criminal Code of the Republic of Kazakhstan.

5. Methods / technologies of teaching and learning: Educational cases, case study
6. Assessment methods / technologies Assessment using a checklist
7. Literature: see Appendix 1.
8. Control (questions):

1. What legal acts regulate medical activity?
2. What normative acts include provisions on the protection of the health of citizens?
3. What is the fundamental legal document for doctors?

Appendix 1

Literature:

Main:

1. Public health: textbook / A. A. Akanov [and others]. - ; Approved and rec. committee for control in the field of education and science. Ministry of Education and Science of the Republic of Kazakhstan. - M. : "Litterra", 2017. - 496 p.
2. Boleshov, M. A. Kogamdyk densaulyk zhane densaulykty saktau: okulyk / M. A. Boleshov. - Almaty: Evero, 2015. - 244 bet p.
3. Campbell, A. Medical Ethics / A. Campbell, G. Gillette, G. Jones; ed. Yu. M. Lopukhin. - M. : GEOTAR - Media, 2014. - 368 bet. With.

Additional:

1. Rymanov D.M., Medic V. A. Public health and health care: hands. to practical exercises. - M.: GEOTAR - Media, 2012. - 400 p.

Electronic resources:

1. Lisitsyn Yu. P. Public health and healthcare [Electronic resource]: textbook / Yu.P. Lisitsyn, G.E. Ulumbekova. - 3rd ed., revised. and additional - Electron. text data. (43.1Mb). - M. : GEOTAR - Media, 2017. - el. opt.
2. Medic V. A. Public health and healthcare [Electronic resource]: textbook / V.A. Medic, V. K. Yuryev. - Electron. text data. (47.6 Mb). - M. : GEOTAR - Media, 2013. - 608 p. email
3. Lisitsyn Yu. P. Public health and healthcare [Electronic resource]: textbook / Yu. P. Lisitsyn, G. E. Ulumbekova. - 3rd ed., revised. and additional - Electron. text data. (40.9 Mb). - M.: Ed. group "GEOTAR-Media", 2011. - 544 el.
4. Shchepin O. P. Public health and healthcare [Electronic resource]: textbook / O. P. Shchepin, V. A. Medic. - Electron. text data. (43.6 Mb). - M.: Ed. group "GEOTAR-Media", 2011. - 592 p. email opt. disc (CD-ROM).
5. Medic V. A. Public health and healthcare [Electronic resource]: a textbook for honey. Schools and colleges / V. A. Medik., V. K. Yuriev. - 3rd ed., revised. and additional - Electron. text data. (37.2 MB). - M.: Publishing group "GEOTAR-Media", 2011. - 288 p. email opt. disk.

Electronic database:

No.	Name	Links
1	Digital library	http://lib.ukma.kz
2	Digital catalogue - For internal users - For external users	http://10.10.202.52 http://89.218.155.74
3	Intermediate republican higher educational institutions electronic library	http://rmebrk.kz/
4	Electronic Library of the Medical School "Student Advisor"	http://www.studmedlib/ru
5	Section "Paragraph" information system "Medicine"	https://online.zakon.kz/Medicine
6	Electronic source of legal information "law"	https://zan.kz
7	Scientific electronic library	https://elibrary/ru/
8	"BooksMed" electronic library	http://www.booksmed.com

9	"Web of science" (Thomson Reuters)	http://apps.webofknowledge.com
10	"Science Direct" (Elsevier)	https://www.sciencedirect.com
11	"Scopus" (Elsevier)	www.scopus.com
12	PubMed	https://www.ncbi.nlm.nih.gov/pubmed