

**МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ**  
**УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ**  
**«ГОМЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ»**

**Кафедра общественного здоровья и здравоохранения**

# **ОСНОВЫ МЕДИЦИНСКОЙ СТАТИСТИКИ**

**Учебно-методическое пособие  
для студентов 4 и 6 курсов  
факультета по подготовке специалистов для зарубежных стран,  
обучающихся по специальности «Лечебное дело»,  
медицинских вузов**

## **BASES OF MEDICAL STATISTICS**

**Manual for 4, 6<sup>st</sup> year students  
of higher medical educational institutions  
of the Faculty on preparation of experts for foreign countries,  
Specialty of «General medicine»**

**Гомель  
ГомГМУ  
2013**

УДК 61:31 (072=111)  
ББК 51.1 (2Англ)я73  
О-75

**Авторы:**

*Л. Г. Соболева, Н. П. Петрова, Т. М. Шаршакова, В. М. Дорофеев*

**Рецензенты:**

кандидат медицинских наук,  
доцент кафедры общей гигиены, экологии и радиационной медицины  
Гомельского государственного медицинского университета

***В. Н. Бортновский;***

кандидат медицинских наук,  
доцент кафедры общественного здоровья и здравоохранения,  
заведующий научно-исследовательским сектором  
Гомельского государственного медицинского университета

***И. А. Чешик***

**Основы медицинской статистики:** учеб.-метод. пособие для студентов  
О-75 4 и 6 курсов факультета по подготовке специалистов для зарубежных  
стран, обучающихся по специальности «Лечебное дело» медицинских  
вузов = Bases of medical statistics: manual for 4, 6<sup>st</sup> students of higher  
medical educational institutions of the Faculty on preparation of experts for  
foreign countries, specialty of «General medicine» / Л. Г. Соболева [и др.]. —  
Гомель: ГомГМУ, 2013.— 36 с.  
ISBN 978-985-506-473-3

Может использоваться для проведения практических занятий на кафедре общественного здоровья и здравоохранения с целью получения знаний, умений и навыков по теме «Основы медицинской статистики». Содержит современные материалы по основам медицинской статистики. Соответствует учебному плану и типовой учебной программе по дисциплине «Общественное здоровье и здравоохранение», утвержденной Министерством здравоохранения Республики Беларусь.

Предназначено для студентов 4 и 6 курсов факультета по подготовке специалистов для зарубежных стран, обучающихся по специальности «Лечебное дело», медицинских вузов.

Утверждено и рекомендовано к изданию Центральным учебным научно-методическим советом учреждения образования «Гомельский государственный медицинский университет» 28 сентября 2012 г., протокол № 6.

**УДК 61:31 (072=111)  
ББК 51.1(2Англ)я73**

**ISBN 978-985-506-473-3**

© Учреждение образования  
«Гомельский государственный  
медицинский университет», 2013

## CONTENTS

Introduction .....	4
1. Organization of statistic research. Graphic images in statistics.....	5
2. Statistic values. Their use in medicine.....	10
3. Dynamic rows. Ways of smoothing and analysis .....	15
4. Reliability estimation of the results of statistic investigation.....	17
5. Study of relation between phenomena .....	20
6. Tasks for students' self work .....	24
6.1. Situational tasks .....	24
6.1.1 Topic: «Topic: «Relative values. Mean values».....	24
6.1.2. Topic «Dynamic rows. Ways of smoothing and analysis».....	24
6.1.3. Topic: «Parametric methods of reliability estimation of statistic research. Correlative analysis» .....	25
7.2. Test control.....	26
7.2.1. Topic: «Organizing the statistic research. Graphic picture in statistics» .....	26
7.2.2. Topic: «Relative values. Mean values» .....	29
7.2.3. Topic: «Parametric methods of reliability estimation of the statistic research. Correlative analysis» .....	32
The literature .....	34

## INTRODUCTION

Every topic is studied for 5 hours and consists of 3 parts.

During the first part of the class on the topic «Organization of statistic research. Graphic image in statistics» we determine the initial level of students' knowledge, analyze and discuss the main questions of the topic. In the second part under the teacher's guidance students determine statistic research stages, make up a plan and program of statistic research according to an individual task. In the third part we test-control the students and sum up the classes.

In the first part of the class on the topic «Statistic values and their use in medicine. Dynamic rows, ways smoothing and analysis» we study and discuss the main questions of the topic. In the second part the students solve situational problems on calculation of relative indices, computation of indices and analysis of dynamic rows. We examine methods of making up variable rows, correct computation of average values, criteria of variety. The third part of the class is given up to students' self work with specific statistic documents on making up of variable rows, computation of average values and other characteristics of the row. We control the received knowledge and practical skills, sum up.

In the first part of the class on the topic «Parametric methods of statistic research reliability assessment. Correlation analysis» we go over theoretical aspects of the given topic, analyze the methods of correlation coefficient computation, methods of statistic research results reliability estimation. The second part is given to students' self work. The students solve situational problems. In the third part we estimate the received knowledge by means of computer testing.

## 1. ORGANIZATION OF STATISTIC RESEARCH. GRAPHIC IMAGES IN STATISTICS

*The term «statistics»* (from Latin word ‘status’ means ‘state’).

*Statistics* is the science studying quantitative side of mass social phenomena and processes indissoluble connection with their qualitative side in concrete conditions of place and time.

*Medical (sanitary) statistics* is the science studying social health and health care which contributes to working out of measures on population health improving.

***Purposes of medical statistics:***

— to study the most important problems of medicine, hygiene, health services, in particular:

— to determine population health indices;

— to estimate the influence of social-biological factors on people’s health;

— to analyze the data about a network, staff and activity of medical-prophylactic organizations;

— to determine effectiveness of medico-prophylactic measures;

— to use statistic methods in experimental, clinical-biological, social-hygienic researches.

### STATISTIC RESEARCH STAGES

I stage. Making up of the statistic research program and plan.

II stage. Organizing and collecting of statistic material.

III stage. Grouping, working out and summing of statistic material.

IV stage. Analysis of the received data, conclusions and suggestions.

Before the beginning of statistic research it is necessary to determine research aim and purposes, to formulate the topic.

**Research aim** must be topical for medical science and practice of health services; it must determine ways of solving the given problem.

**Research purposes** are concretized, broadened and specified determining of the aim. Usually the number of the tasks may be from 3 to 6.

**I STAGE OF STATISTIC RESEARCH** is making up of statistic research program and plan.

***Statistic research program*** includes the following issues:

1. To determine an observation unit and to make up the program of material collecting.

2. To make up the program of material working out.

3. To make up the program of the collected material analysis.

**1. Determining of an observation unit and the program of material collecting.**

***Observation unit (computation unit)*** is every primary element of the statistic totality having the signs of similarity and difference.

***Statistic totality*** is the group consisting of relatively similar elements taken together in the known terms of time and space according to the given aim.

We distinguish two kinds of totality: general and selective.

**General totality** is the totality of all possible units which may be related to it.

**Selective totality** is part of the general totality chosen by a special selective method and intended to characterize the whole general totality.

Selective totality must be representative, that is, it must include sufficient number of observations (quantitative representative ness) and must be typical for the whole general totality from which this selection is made (qualitative representative ness).

Signs by which the elements of statistic totality are distinguished and subjected to computation are called **registering signs**.

**By their character registering signs** may be:

- attributive (descriptive) signs — expressed by words;
- quantitative signs — expressed by a number.

**By the role in the totality:**

- factor signs influencing on the studied phenomenon;
- resulting signs changing under the influence of factor signs.

**Program of material collecting** is the working out of a primary statistic document pattern. *Registration document* may be represented as a list, journal, statistic blank or a card, questionnaire or a protocol. *It must have:* a passport part clearly formulated and put in a certain order questions of the program with variants of the prepared answers and the date of filling in the document. **Variants of the prepared answers are called «grouping».**

Grouping of the answers on attributive signs is called typological, and if the sign is quantitative it is called variable.

1. **The program of processing the received data envisages** making up statistic tables models.

**Vertical columns of the table are called graphs, and horizontal ones are lines.**

We distinguish a statistic subject and a statistic predicate in the table.

1. Statistic subject is the main registering sign analyzed in the table.
2. Statistic predicate (one or several) are statistic signs which supplement and reveal the subject.

All the tables are divided into simple and complex, and complex tables may be group and combinative.

A simple table is a statistic table allowing to analyze the received data having only one sign — a subject (Table 1).

Table 1 — Distribution of students-smokers by departments (in abs. numbers)

Department	Total number of students
General Medicine	
General Medicine and Diagnostics	
Preventive Medicine	
Total	

***In the group table*** the subject and predicate are interrelated; there is no relation between registering signs, and, thus every registering sign characterizes the subject independently (Table 2).

Table 2 — Distribution of different departments students according to sex, age at which they smoked their first cigarette (in abs. numbers)

Department	sex		Age at which first cigarette was smoked			Total
	m	f	under15	15–18	older than18	
General Medicine						
Preventive Medicine						
General Medicine and Diagnostics						
Total						

***In a combinative table*** there are two or some predicates which are related not only with the subject but between each other (Table 3).

Table 3 — Distribution of different department students by sex and average number of cigarettes smoked per day (in abs. numbers)

Department	Average number of cigarettes smoked by students per day									Total		
	10 and less			11–20			More than 20			m	f	both
	m	f	both	m	f	both	m	f	both			
General Medicine												
Preventive Medicine												
General Medicine and Diagnostics												
Total												

***Analysis program*** gives a list of statistic methods necessary to reveal patterns of the studied phenomenon.

**PLAN OF STATISTIC RESEARCH INCLUDES THE FOLLOWING ELEMENTS:**

1. Choice of the research object.
2. Determining of the statistic totality volume.
3. Fixing the terms of the research realization, its kinds and ways of observing and material collecting.
4. Determining of the personnel with the help of which the research will be carried out.
5. Determining of persons which will accomplish organizing and methodical supervision of the researches.
6. Characteristics of the technical equipment and required financial means.

***Statistic research object*** is the statistic totality (of phenomena, objects, persons) about which information will be collected.

**Statistic research volume** is the number of observations included into the research.

**Characteristics of the technical equipment and required financial means:**

- laboratory equipment;
- stationary;
- financial means.

**By time** we distinguish 2 kinds of statistic observation: one-time and current observation.

**One-time observation** is the observation at a certain moment of time. **Examples:** census of population, one-time computation of medico-prophylactic organizations, doctors, beds, etc.

**Current observation** is continuous observation carried out during a certain period of time (month, year, and so on).

**Examples:** registration of morbidity, mortality, birth rate, etc.

**By the degree of scope** we distinguish 2 kinds of statistic observation: overall and selective.

**Overall observation** is the registration of all the totality objects which interest the researcher (general totality).

**Selective observation** is the registration of not all cases but only their part (selective totality).

**Methods of the studied phenomena choice (selection):**

**Random selection** is the selection carried out by a lot or tables of random numbers. Random will be choice of patients surnames of which begin with a certain letter.

**Mechanical selection** is the selection when from the whole totality we take for studying mechanically chosen every fifth (20 %) or every tenth (10 %) observation unit, that is in a certain interval.

**Cluster (serial) selection** is the selection when from the whole general totality we chose not separate units but a series.

**Method of general massif** is used in studying of those objects in which the majority of the studied phenomena is concentrated. The main part characterizing the whole statistic totality is chosen from all the observation units forming a part of the given object.

**Direct selection** is the selection when from general totality only those observation units are chosen which allow to reveal influence of unknown factors in eliminating the influence of the known ones.

**Typological selection** is the selection of units from previously grouped qualitatively similar groups.

**Ways of receiving primary information:**

**1. Direct observation:** direct check-up or instrumental examination of a sick or healthy person and registration of the received data.

**2. Method of data extracting from primary medical documentation.** It is made on special cards of choice or statistic blanks.



**3. Anamnestic method of material collecting.** This method consists of registering the information received from the patient or his relatives.

**4. Reporting method** consists of information collecting with the help of the system of reporting-registration documentation.

**II STAGE OF STATISTIC RESEARCH** includes organizing and carrying out of statistic material collecting.

Material collecting is carried out according to previously made up research program and plan.

**III STAGE OF STATISTIC RESEARCH** includes grouping, processing and summing of statistic material.

It includes the following successively made actions:

**1. Control** is carried out with the aim of choosing registration documents having defects.

**2. Ciphering** is conventional symbol of every selected grouping of the studied sign by a certain cipher (code).

**3. Grouping** is the distribution of the collected material according to attributive or quantitative sign (typological or variable).

**4. Material summary** is the summarization of isolated cases received by observation into certain groups. The results of statistic summary and materials grouping are shown as tables. In the tables all the received data are registered firstly in absolute values.

**5. Computation of statistic indices.** The program of analysis is realized.

**6. Graphic images.**

**Graphs are divided into** diagrams, cartograms and cartodiagrams.

**Kinds of diagrams** are linear, radial, columnar, sector, figure.

**Linear diagrams** are used to show phenomenon dynamics. With the help of linear diagrams it is expedient to show indices dynamics of population movement, morbidity, network of medical establishments, etc.

**Radial diagram** is based on the system of polar coordinates in showing phenomenon dynamics within a closed period of time (24 hours, a week, a year). Seasonal variations of infectious morbidity, daily variations of calls to the First Aid Station, fluctuations of discharged patients number and the admitted ones to hospitals on particular days of the week and so on.

**Columnar diagram** is used to show dynamics of the phenomenon statistics as well as the phenomenon structure.

**Sector diagram** is used to show the phenomenon structure (structure of morbidity or structure of population mortality causes, etc.).

**Statistic cards — cartograms** — show intensity of statistic indices in geographic, administrative regions. For example, indices of birth rate, mortality, morbidity or other data are drawn onto a contour card by means of various coloring, shading or with the help of isolines showing regions with similar indices.

**Cartograms.** Diagrams of various kinds are drawn onto a contour card.

Columnar, linear, radial diagrams, cartograms, cartodiagrams are used for graphic image of *intensive indices*.

*Extensive indices* may be graphically shown as sector diagrams as well as cartograms.

*Indices of visibility* are shown by a columnar diagram, cartogram and cartodiagram.

#### **IV STAGE — ANALYSIS OF THE RECEIVED DATA, CONCLUSIONS AND SUGGESTIONS.**

Analyzing the material we study all interacted factors deeper, reveal main influencing causes, exclude accidental phenomena and so on. We think over, compare, discuss, make conclusions, work out measures to be used in practice.

## **2. STATISTIC VALUES. THEIR USE IN MEDICINE**

**Absolute value** is the value which characterizes the scope or singleness of a phenomenon. For example, size of the population, number of medico-prophylactic establishments, number of doctors, number of infectious diseases, patients with AIDS or virus-carriers, etc.

Comparing the sizes of the phenomena or studying phenomena changing within time it is necessary to reduce to the same denominator absolute numbers, more often it is the size of the population.

#### **Relative values, their kinds, methods of computation, use in health services.**

##### **Kinds of relative values:**

1. Extensive index.
2. Intensive index.
3. Index of correlation.
4. Index of visibility.

**Extensive index** is the index of the phenomenon structure.

Method of computing: (phenomenon part / whole phenomenon) x 100 %.

Example: in 2010 the number of all diseases of children in a kindergarten amounted to 102 cases, including 50 cases of enteritis and 52 cases of dysentery. If we accept all the cases as 100 %, then morbidity with enteritis amounted to —  $50 / 102 \times 100 \% = 49 \%$ , and morbidity with dysentery —  $52 / 102 \times 100 \% = 51 \%$ .

**Intensive index** is the index of phenomenon frequency in the medium.

Method of computing: (phenomenon / size of population) x base.

Base: 100, 1000, 10000, 100000.

Index of lethal outcome = (number of the dead / number of the sick) x 100 %.

Index of temporary disablement days frequency = (number of temporary disablement days / number of workers) x 100.

Birth rate = (number of the born within a year / average annual size of population) x 1000.

Index of primary invalidity = (number of people which were given invalidity group for the first time in life / average annual size of population) x 10000.

Maternal mortality = (number of women died during pregnancy, delivery and first 42 days of postnatal period / number of babies born alive) x 100000.

Index of correlation is characterized by relation of two statistic totalities which are not related between each other, and correlated only by their content.

Method of computing: (phenomenon / size of population) x 10000.

Example: index of population supply with doctors, nursing staff, beds.

Index of visuality is the index which is used to characterize phenomenon changing in dynamics (Table 4).

Table 4 — Dynamics of birth rate in the region within 2000–2002 years.

Year	Birth rate	Index of visuality (%)
2000	13,1	100
2001	14,2	108,4
2002	15,5	118,3

Method of computing:  $14,2 / 13,1 \times 100 = 108,4$ ;

$15,5 / 13,1 \times 100 = 118,3$ .

Conclusion: index of birth rate within 2002 year increased to 18,3 % compared to 2000 year.

### **Mean values.**

**Mean values** are summarizing indices characterizing the size of this or that sign which vary in separate units of the qualitatively similar totality.

**Variation** is changing of the varying sign meaning in separate units of the totality.

**Row of distribution** is ordered distribution of the totality units according to the meaning of varying sign. If the varying sign has qualitative measure, then such a variation is called **qualitative**, and the row of distribution — **attributive** (distribution of the sick according to nozologic forms of the disease, sex, profession and so on). If the varying sign has quantitative expression, such variation is called **quantitative**, and the row of distribution — variable.

**Variable row** is a statistic row showing distribution of the studied phenomenon according to quantitative sign. For example, the sick according to age, terms of treatment; newborns according to their weight, etc.

**Variance** is separate meanings of the sign by which grouping is made (marked *V*).

**Frequency** is a number showing how often this or that variance occurs (marked *P*). The sum of all frequencies shows **total number** of observations. **Total number** of observations may be marked with letters *N, n* or  $\sum P$ .

### **We distinguish the following variable rows:**

#### **1. Discontinuous and continuous.**

The row is considered continuous if the grouping sign is expressed by fractional quantities (weight, height and the like), and discontinuous, if the grouping sign is expressed by a whole number (days of disablement, pulse rate, and so on).

## 2. Simple and weighted.

Simple variable row is the row in which quantitative value of the sign is met one time. In the weighted variable row quantitative values of the sign reoccur with a certain frequency.

## 3. Grouped and non-grouped.

A grouped row has variances, combined into groups by quantity in the limits of a certain interval. In a non-grouped row certain frequency corresponds to every separate variance.

## 4. Even and odd.

In even variable rows the sum of frequencies or total number of observations is expressed by an even number and in the odd rows — by an odd number.

## 5. Symmetrical and asymmetrical.

In a symmetrical variable row all kinds of mean values coincide or they are very close (mode, median, mean arithmetical). In asymmetrical ones they do not coincide and they are not close.

In sanitary statistics the following kinds of mean values are used:

- mode, median;
- arithmetical mean;
- harmonic mean;
- geometrical mean;
- progressive mean.

**Mode ( $M_o$ )** is the variance which corresponds to the highest frequency.

**Median ( $M_e$ )** divides the variable row (ordered, that is variances values are placed in order of increase or decrease) into two equal halves.

**Arithmetical mean** is the most prevalent value. Arithmetical mean is marked with the letter **M**.

There are distinguished arithmetical mean simple and weighed.

**Arithmetical mean simple** is calculated:

$$M = \frac{\sum V}{n},$$

where  $V$  — individual values of the sign;

$n$  — number of individual values;

$\sum$  — symbol of the sum.

**Example:** to determine average duration of bed staying of 10 patients with pneumonia:

16 days — 1 patient; 17–1; 18–1; 19–1; 20–1; 21–1; 22–1; 23–1; 26–1; 31–1.

$$M = \frac{16 + 17 + 18 + 19 + 20 + 21 + 22 + 23 + 26 + 31}{10} = 21,3 \text{ bed-days.}$$

**Arithmetical mean** weighed is computed in those cases when individual values of the sign reoccur. It can be computed by the following formula:

$$M = \frac{\sum VP}{\sum P},$$

where P is frequency (number of cases) of each variance observations;  
 $\sum P$  is total number of observations.

**Example:** to determine average height of 8 year-old boys (Table 5).

Table 5 — Initial data

Height in cm	Number of boys (P)	Central variance (V)	VP
115–116	2	116	232
117–118	7	118	826
119–120	21	120	2520
121–122	33	122	4026
123–124	21	124	2604
125–126	12	126	1512
127–128	3	128	384
129–130	1	130	130
	n = 100		12234

Central variance — interval middle — is determined as half-sum of the initial values of two adjoining groups:

$$\frac{115+117}{2} = 116; \frac{117+119}{2} = 118 \text{ and, so on.}$$

Product VP is obtained by multiplying central variances to frequencies:  $116 \cdot 2 = 232$ ;  $118 \cdot 7 = 826$  and, so on. Then the obtained products are summed and we obtain  $\sum VP = 12234$ .

$$M = \frac{12234}{100} = 122,34 \text{ cm.}$$

Average height of 8 year-old boys = 122,34 cm.

While studying the varying sign one must not confine oneself only to computing mean values. It is necessary to compute indices characterizing the degree of the studied signs variety.

Characteristics of the variable row is mean quadratic deviation ( $\sigma$ ), which shows scattering of the studied signs concerning mean arithmetical. It can be determined by the formula:

$$\sigma = \pm \sqrt{\frac{\sum (V - M)^2 P}{\sum P}}$$

**Example of calculation:** to determine average number of sick-lists given at a polyclinic within a day (Table 6).

Table 6 — Initial data

Number of sick-lists given by a doctor within a day (V)	Number of doctors (P)	VP	d=V-M	(V-M) <sup>2</sup> (d <sup>2</sup> )	(V-M) <sup>2</sup> P (d <sup>2</sup> )P
4	12	48	-2	4	48
5	4	20	-1	1	4
6	19	114	0	0	0
7	3	21	1	1	3
8	2	16	2	4	8
n=40		219			63

$$M = \frac{219}{40} = 5,5; \quad \sigma = \pm \sqrt{\frac{63}{40}} = \pm \sqrt{1,575} = \pm 1,3$$

Conclusion: average number of sick-lists given by a doctor within a day is from 4 to 6 ( $5,5 \pm 1,3$ ).

According to a normal law of errors distribution (discovered by K. Gauss and P. Laplas) individual values of the sign are within limits of  $M \pm 3\sigma$ , which encompasses 99,73 % of all totality units.

If  $M \pm 2\sigma$ , then in the limits of the obtained values there are 95,45 % of all members of the variable row, and finally if  $M \pm 1\sigma$ , then in the limits of the obtained values there will be 68,27 % of all members of the variable row.

**In medicine the concept of norm is associated with  $M \pm 1\sigma$ .**

In sanitary statistics the rule of three sigmas is used in studying physical development, estimation of health care establishments activity, estimation of population health.

Thus, mean quadratic deviation serves for characteristics of signs variety degree which are determined by coefficient of variation:

$$C = \frac{\sigma}{M} \cdot 100 \%$$

Coefficient of variation less than 10 % means weak variety of signs, from 10 % to 20 % — middle, and more than 20 % — strong variety. Coefficient of variation to a certain extent is a criterion of arithmetical mean reliability.

### 3. DYNAMIC ROWS. WAYS OF THEIR SMOOTHING AND ANALYSIS

**Dynamic row** is the row of similar values which shows changing of the studied phenomenon within time. Numbers forming a dynamic row are called levels of the row (Table 7). **Levels of the row may be presented by absolute numbers, relative and mean values.**

**Interval row** consists of successive row of numbers characterizing the phenomenon changing for a certain period (by time).

**Moment row** consists of values determining the sizes of the phenomenon at a certain date — moment.

**Indices of a dynamic row:**

1. **Absolute increase** is the difference between every succeeding level and a preceding one. Absolute increase may be positive and negative (Table 7).

2. **Rate of growth** is the ratio of every succeeding level to the preceding one expressed in percentage (Table 7).

3. **Rate of increase** is the ratio of absolute increase to the preceding level accepted as 100 % (Table 7).

4. **Absolute value** of 1 % increase is a quotient from division of absolute increase within a certain period of time into rate of increase in percentage within the same period (Table 7).

**Example:** It is necessary to give the analysis of birth rate dynamics in a certain region.

Table 7 – Birth rate dynamics in the region within 2000–2004 years

Year	Birth rate, %	Absolute increase	Rate of increase, %	Rate of growth, %	Absolute value of 1% increase
2000	9,4	—	—	—	—
2001	8,9	-0,5	-5,3	94,7	0,09
2002	9,2	0,3	3,4	103,4	0,09
2003	9,3	0,1	1,1	101,1	0,09
2004	9,4	0,1	1,1	101,1	0,09

Order of computing:

1. We determine absolute increase:  $8,9 - 9,4 = -0,5$ ;  $9,2 - 8,9 = 0,3$  and, so on.
2. We compute the rate of increase:  $-0,5 \times 100 / 9,4 = -5,3$  and, so on.
3. We find out the rate of growth:  $8,9 \times 100 / 9,4 = 94,7$  and, so on.
4. We get absolute value of 1% increase:  $-0,5 / -5,3 = 0,09$

Dynamic row does not always consist of levels successively changing to the side of decrease or increase. Often the levels of the dynamic row greatly vary. In such cases we make smoothing of the dynamic row. There exist several ways of dynamic row smoothing: enlargement of the interval, leveling down by computing the sliding average, smoothing in a straight line, etc.

**1. Smoothing in a straight line** (Table 8). For such smoothing it is necessary to know the following formulae:

$Y_t$  (theoretical levels) =  $a_0 + a_1t$ , where  $t$  is conventional indication of time,  $a_0$  и  $a_1$  are parameters of the sought-for straight line.

$$a_0 = \Sigma y / n$$

$$a_1 = \Sigma yt / \Sigma t^2$$

Let us examine the following example.

Table 8 – Birth rate smoothing within 2003–2008 years

Year	Birth rate, (y)	Conventional indication of time, t	$y \times t$	$t \times t$	Smoothing in a straight line
2003	9,4	-5	-47	25	11
2004	8,9	-3	-26,7	9	10,1
2005	9,2	-1	-9,2	1	9,3
2006	8,3	1	8,3	1	8,5
2007	9,4	3	18,8	9	7,7
2008	8.4	5	25,2	25	6,9

$n = 6$

$\Sigma y = 53,5$

$\Sigma yt = -30,6$

$\Sigma t \times t = 70$

If the row is even, in finding out  $t$  the countdown begins with 1 (middle of the row), then successively odd numbers 3, 5, 7 and so on (upward with - ; downward with +). If the row is odd, the countdown of the conventional indication of time begins with 0 (middle of the row), then - 1, 2, 3 and, so on to the both sides.

Order of computing:

$$Y_t \text{ (theoretical levels)} = a_0 + a_1t$$

$$a_0 = \Sigma y / n; a_1 = \Sigma yt / \Sigma t^2$$

$$a_0 = 8,9; a_1 = -0,4;$$

$$8,9 + (-0,4) \times (-5) = 11;$$

$$8,9 + (-0,4) \times (-3) = 10,1; \text{ и т.д.}$$

## 2. Smoothing by calculating the sliding average (Table 9).

Table 9 – Birth rate smoothing within 2003–2008 years

Year	Birth rate, (y)	Three-year moving average
2003	9,4	–
2004	8,9	9,2
2005	9,2	8,8
2006	8,3	8,9
2007	9,4	8,7
2008	8.4	–



Order of computing:

For 2004 year  $(9,4 + 8,9 + 9,2) / 3 = 9,2$ .

For 2005 year  $(8,9 + 9,2 + 8,3) / 3 = 8,8$  and so on.

### 3. Enlargement of the interval (Table 10).

Enlargement of the interval is made by summing the data up within a row of adjacent periods.

Table 10 – Birth rate smoothing within 2003–2008 years.

Years	2003	2004	2005	2006	2007	2008
Birth rate	9,4	8,9	9,2	8,3	9,4	8,4

Within 2003–2005 birth rate amounts to  $9,4 + 8,9 + 9,2 = 27,5$ .

Within 2006–2008 birth rate amounts to  $8,3 + 9,4 + 8,4 = 26,1$ .

## 4. RELIABILITY ESTIMATION OF THE RESULTS OF STATISTIC INVESTIGATION

Precision and reliability measure of selective statistic values is the average mistakes of representativeness which depend on the number of choice and variety degree of selective totality on the examined sign.

Therefore, determining the reliability degree of statistic data requires computing the appropriate average mistake for every relevant and average value.

1. The average mistake of the intensive index  $m_p$  is computered according to the formula:

$$m_p = \pm \sqrt{\frac{P \cdot q}{n}}$$

If the observation number is less than 30  $m_p = \pm \sqrt{\frac{P \cdot q}{n-1}}$ ,

where P — index value in percents, promills, etc.;

q — completion of this index up to 100, if it's in percents, up to 1000, if ‰ etc. (i.t.  $q = 100 - P$ ,  $1000 - P$  etc.).

For example: there were 224 people ill with dysentery during the year in the particular region. The population size — 33000.

$$\text{The morbidity index of dysentery for 10000} = \frac{244 \cdot 10000}{33000} = 74.$$

$$\text{The average mistake of this index } m_p = \pm \sqrt{\frac{74 \cdot (10000 - 74)}{33000}} = \pm 4,79.$$

To determine the reliability degree of the index it's necessary to determine the confidential coefficient (t), which is equal to the ratio of the index to its average mistake, i.t.

$$t = \frac{P}{m_p}$$

In this example  $t = \frac{74}{4,79} = 15,4$

The higher t is, the higher the reliability degree is.

**If t = 1, the probability of the reliability index is 68,3 %, if t = 2–95,5 %, if t = 3–99,7 %.** So, in our example the morbidity index is reliable.

If the number of observation is less than 30, the value of the criterion is determined according to Student's table. If the received value is higher or equal to the tabular one, the index is reliable. If it's lower, the index is not reliable.

2. If it's necessary to compare two intensive indices the reliability of their difference is determined according to the formula:

$$t = \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}} \text{ (the lesser number is subtracted from the greater number),}$$

where  $P_1 - P_2$  — the difference between two compared indices;

$\sqrt{m_1^2 + m_2^2}$  — the average mistake of the difference of two indices.

For example: there were 208 people ill with dysentery during the year in region A. The population size — 36000. There were 265 people ill with dysentery during the year in region B. The population size — 40000. So, dysentery morbidity index:

$$P_1 = \frac{208 \cdot 10000}{36000} = 58$$

$$m_{p_1} = \pm \sqrt{\frac{58 \cdot (10000 - 58)}{36000}} = \pm 4$$

$$P_2 = \frac{265 \cdot 10000}{40000} = 66$$

$$m_{p_2} = \pm \sqrt{\frac{66 \cdot (10000 - 66)}{40000}} = \pm 4$$

As you see, morbidity in region A is lower than in region B. Then we determine the reliability of the difference between two indices according to the formula:

$$t = \frac{66 - 58}{\sqrt{4^2 + 4^2}} = \frac{8}{5,6} = 1,4$$

The probability of the reliability of the index difference is 68,3 %. In our example the morbidity in region B is reliably higher than in region A.

It is necessary to determine the average mistake for the arithmetical mean as well as for relative values.

The average mistake of the arithmetical mean  $m_x$  is determined according to the formula:

$$m_x = \pm \frac{\sigma}{\sqrt{n}} \text{ or } m_x = \pm \frac{\sigma}{\sqrt{n-1}} \text{ (if the number of observations is less than 30),}$$

where  $\sigma$  — the average quadratic deviation;

$n$  — the number of observations.

3. Reliability estimation of the arithmetical mean:

$t = \frac{M}{m}$  The result is reliable if the value of the confidential coefficient ( $t \geq 2$ ).

For example, 15 workers of the workshop with high air temperature were examined to see the pulse rate. Results of inspection table 11.

Table 11 — Results of inspection

Number of pulse beats X V	Number of workers (P)	VP	V-M	(V-M) <sup>2</sup>	(V-M) <sup>2</sup> P
76	3	228	-5,3	28,09	84,3
78	5	390	-3,3	10,89	55,0
82	3	246	0,7	0,49	1,5
88	2	176	6,7	44,89	90,0
90	2	180	8,7	75,69	151,4
n=15		1220			382,2

Determine the average pulse rate and its reliability.

1. Determine the arithmetical mean:

$$M = \frac{\sum VP}{n} = \frac{1220}{15} = 81,3 \text{ beats / min.}$$

2. Determine the average quadratic deviation:

$$\sigma = \pm \sqrt{\frac{\sum (V - M)^2 P}{\sum P}} = \pm \sqrt{\frac{382,2}{15 - 1}} = \pm \sqrt{27,3} = \pm 5,2$$

3. Compute the average mistake of the arithmetical mean:

$$m = \pm \frac{\sigma}{\sqrt{n-1}} = \pm \frac{5,2}{\sqrt{14}} = \pm \frac{5,2}{3,74} = \pm 1,4$$

4. Estimate the reliability of the arithmetical mean:

$$t = \frac{M}{m} = \frac{81,3}{1,4} = 58,1$$

The value is reliable.

5. Determine the reliability of the difference between two arithmetical means:

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

The results are reliable if the value of the confidential coefficient (t) is equal or more than 2.

For example, while examining 25 workers of the workshop with normal air temperature (18–20°) the average pulse rate turned out to be 72,4 beats / min. The average quadratic deviation — 4,8 beats / min., the average mistake — 0,96. It's necessary to determine the reliability of the difference in the pulse rate between the workers of both workshops. So, we need to put the data into the formula

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} = \frac{81,3 - 72,4}{\sqrt{1,4^2 + 0,96^2}} = \frac{8,9}{\sqrt{1,96 + 0,9216}} = \frac{8,9}{1,7} = 5,2$$

The difference in the pulse rate of the workers of the compared workshops is statistically reliable.

## 5. STUDY OF RELATION BETWEEN THE PHENOMENA

All the phenomena in the nature and society are interconnected. By the character of the phenomenon dependence there are:

- functional (complete);
- correlative (incomplete) connections.

The functional connection is often presented by the formulae:

$$(body\ volume = \frac{mass(m)}{densityt(p)}; distance = speed \times time, etc.)$$

In case of correlative connection the same value of one sign corresponds to different values of another sign. For example, there is a correlative connection between height and weight, morbidity with malignant neoplasms and age etc.

**By direction** there are direct and inverse correlative connections (Table 9). In case of direct connection increasing of one sign results in increasing another one; in case of inverse connection increasing of one sign results in decreasing another one.

**By strength** connection may be strong, moderate and weak (Table 12).

By means of statistic analysis it's possible to determine the connections, its direction and measure its strength. One of the means of measuring the connection between the phenomena is computing **the correlation coefficient marked as  $r_{xy}$** . It can be computed by different means. The most precise is the method of squares (Pirson's) in which the correlation coefficient is determined according to the formula:

$$r_{xy} = \frac{\sum d_x \cdot d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}}$$

where  $r_{xy}$  — correlation coefficient between the statistic rows X and Y;  
 $d_x$  — deviation of each figure of the statistic row X from its arithmetical mean;  
 $d_y$  — deviation of each figure of the statistic row Y from its arithmetical mean.

Table 12 — The scheme of strength estimation of the correlative connection by the correlation coefficient

Strength of connection	Value of correlation coefficient in case of	
	direct connection (+)	inverse connection (-)
No connection	0	0
Weak connection	from 0 to +0,29	from 0 to -0,29
Moderate connection	from +0,3 to +0,69	from -0,3 to -0,69
Strong connection	from +0,7 to +0,99	from -0,7 to -0,99
Complete (functional) connection	+1,0	-1,0

To compute the correlation coefficient according to the method of squares it's necessary to make a seven column table.

**Determining the strength and character of connection between the iodine content in water and the goiter affection (Table 13).**

Table 13 — Results of inspection

Iodine content (mg) in water ( $V_x$ )	Goiter affection in % ( $V_y$ )	$d_x = V_x - M_x$	$d_y = V_y - M_y$	$d_x d_y$	$d_x^2$	$d_y^2$
201	0,2	63	-3,6	-226,8	3969	0,04
178	0,6	40	-3,2	-128,0	1600	0,36
155	1,1	17	-2,7	-45,9	289	1,21
154	0,8	16	-3,0	-48,0	256	0,64
126	2,5	-12	-1,3	-15,6	144	6,25
81	4,4	-57	0,6	-34,2	3249	19,36
71	16,9	-67	13,1	-877,7	4489	285,61
				$\Sigma 1345,0$	$\Sigma 13996$	$\Sigma 313,47$

n=7

1. Determine the average iodine content in water (in mg / l).

$$M_x = \frac{\sum V_x}{n} = \frac{996}{7} = 138 \text{ mg / l}$$

2. Determine the average goiter affection in %.

$$M_y = \frac{\sum V_y}{n} = \frac{26,5}{7} = 3,8\%$$

3. Determine the deviation of each  $V_x$  from  $M_x$ , i.t.  $d_x$ .

$$201 - 138 = 63; 178 - 138 = 40 \text{ etc.}$$

4. In the same way determine the deviation of each  $V_y$  from  $M_y$ , i.t.  $d_y$ .

$$0,2 - 3,8 = -3,6; 0,6 - 3,8 = -3,2 \text{ etc.}$$

5. Determine the products of deviations. The received product is summed up.

$$\sum d_x d_y = -1345,0$$

6.  $d_x$  is squared and the results are summed up.

$$\sum d_x^2 = 13966$$

7. In the same way  $d_y$  is squared, the results are summed up.

$$\sum d_y^2 = 313,47$$

8. As a result, all the received sums are put into the formula:

$$r_{xy} = \frac{-1345}{\sqrt{13966 \cdot 313,47}} = \frac{-1345}{2094,59} = -0,64$$

As for the reliability of the correlation coefficient, the average mistake is determined according to the formula:

$$m_r = \frac{1 - r_{xy}^2}{\sqrt{n}}$$

(If the number of observations is less than 30, there is  $n-1$  in the denominator).

In our example

$$m_r = \frac{1 - 0,64^2}{\sqrt{7-1}} = \frac{1 - 0,4096}{2,45} = 0,24$$

The value of the correlation coefficient is considered reliable if it is three times as much as its average mistake.

In our example  $\frac{0,64}{0,24} = 2,7$

Therefore, the correlation coefficient is not reliable, which requires the increase in number of observations.

**Person fitting criterion x2 (chi-square)**

Convenient to prove the reliability of differences between samples is Person fitting criterion is x2 (chi-square), which is based on the suggestion that there is no difference between the actual values (P) and theoretical values (expected p'). Fitting criterion x2 has the null hypothesis which is rejected during calculation.

For example: There are 4178 people with CAD on the follow-up list, 1160 from them have hypertension disease, and 3018 people from them are without hypertension. During 3 years of follow-up from 1160 people, 12 died, 1148 are alive, from 3018 people, 55 from them died, 2963 are alive. Define if there is a difference in the probability of survival with ischemic heart disease, depending on the presence or absence of hypertension (Table 14.)

Fitting criterion index is calculated by the formula

$$\chi^2 = \sum \frac{(P - P')^2}{P'}$$

where P and p' – are actual expected numbers.

If P= p', then x2=0, i.e. null hypothesis would be confirmed. The less P differs from p', the closer is x2 to 0. To find out the significance of the difference, the received value x2 is assessed by the application (supplement 1).

n1 – is the number of degrees equals (S-1)(r-1), where S – is the number of columns; r – is the number of lines.

Table 14 — The calculation of fitting criterion index (x2)

	Actual numbers P		«Expected numbers p'»		$\frac{(P - P')}{P'}$	
	Alive	Dead	Alive	Dead	Alive	Dead
CAD with hypertension disease	1148	12	1141	19	$\frac{(1148 - 1141)^2}{1141} = 0,043$	$\frac{(12 - 19)^2}{19} = 2,58$
CAD without hypertension disease	2963	55	2970	48	$\frac{(2963 - 2970)^2}{2970} = 0,02$ $\Sigma = 0,06$	$\frac{(55 - 48)^2}{48} = 1,02$ $\Sigma = 3,6$
TOTAL	4111	67			x2 = 0,06 + 3,6 = 3,66	
	98,4%	1,6%				

Supposing that the probability of survival is the same – 98,4%, we find the number of alive from the first group and the second group accordingly

$\frac{1160 \times 98,4}{100} = 1141$  ;  $\frac{3018 \times 98,4}{100} = 2970$  . «Expected» numbers of dead in the first group  $\frac{1160 \times 1,6}{100} = 19$  or  $1160 - 1141 = 19$ ; In the second group  $\frac{3018 \times 1,6}{100} = 48$  or  $3018 - 2970 = 48$ . Comparing the actual numbers with expected ones by the given formula, we can define  $x^2$ . As the table critical value is 3,84, and ours is 3,66, which is less than table's value (supplement 1), so the null hypothesis is accepted, i.e. the difference is unreliable.

## 6. TASKS FOR STUDENT'S SELF-WORK

### 6.1. SITUATIONAL TASKS

#### 6.1.1. Topic: «Relative values. Mean values».

1. Population in town N. in 2002—47 020. In 2002 there were registered 550 cases of blood circulation diseases (BCD), including: hypertension — 305 cases, ischemic heart disease (IHD) — 115 cases, cerebrovascular diseases — 25 cases, others — 105 cases.

Compute the intensive and extensive indices of morbidity. Make conclusions.

2. Population in town N. in 2006 — 50 030 people. In 2006 there were registered 770 cases of BCD, including hypertension — 360 cases, IHD — 160 cases, cerebrovascular diseases — 30 cases, others — 220 cases.

Compute the intensive and extensive indices of morbidity. Make conclusions.

3. In 2002 in the Republic of Belarus there were 40 449 doctors (not including dentists). Indices of the provision of the population with doctors: in 2000—44,4 ‰, in 2001—40,7 ‰. Population size — 9899,2 thousand people.

Determine the index of provision with doctors in 2002, compute the index of visibility taken for 3 years and present it graphically.

4. The duration of temporary disablement was studied in patients with acute respiratory diseases. It was determined that during 5 days 5 patients were ill; 6 days — 5 patients; 7 days — 5 patients; 8 days — 2 patients; 9 days — 5 patients.

Make up a variable row and compute its main characteristics.

#### 6.1.2. Topic «Dynamic rows. Ways of smoothing and analysis».

1. Compute the main indices of the dynamic row (Table 15).

Table 15 — Data for calculation

Year	Mortality (%)	Absolute increase	Increase rate (%)	Growth rat (%)	Index of visibility (%)
2000	6,5				
2001	7,0				
2002	8,5				
2003	13,0				



2. Compute the main indices of the dynamic row (Table 16).

Table 16 — Data for calculation

Year	Mortality (%)	Absolute increase	Increase rate (%)	Growth rat (%)	Absolute value of 1% increase
2000	8,5				
2001	11,5				
2002	14,9				
2003	16,7				

3. Infantile mortality in the Republic of Belarus was: in 1999–11,4 ‰, in 2000–9,3 ‰, in 2001–9,1 ‰, in 2002–7,9 ‰.

Compute the indices of the dynamic row and to analyze them. Present the data graphically.

**6.1.3. Topic: «Parametric methods of reliability estimation of statistic research. Correlative analysis».**

1. While studying the average treatment duration of patients at the surgical and therapeutic departments of the city hospitals the following data were received: the average treatment duration at the surgical department was 12,1 ( $m_1 = \pm 0,05$ ) days, but at the therapeutic department — 23,2 ( $m_2 = \pm 0,09$ ) days. Determine the difference reliability of the average treatment duration in the compared groups.

2. In the region B in the year 270 people fell ill with dysentery, in the A - 130. A population of the region - 45 000, population region B - 75 000. To reliability estimation of the difference between the obtained parameters (t difference).

3. Determine the correlative connection between the age and the number of the hospitalized patients with heart failure (Table 17).

Table 17 — Initial data

Age in years (X)	Number of the hospitalized (Y)
19	12
29	24
39	30
49	23
59	26
69	30

4. Determine the correlative connection between the age and the number of the hospitalized patients with heart failure (Table 18).

Table 18 — Initial data

Age in years (X)	Number of the hospitalized (Y)
39	11
41	26
59	35
62	26
69	37
79	42

5. Determining the strength and character of connection between the iodine content in water and the goiter affection (Table 19).

Table 19 — Initial data

Iodine content (mg) in water ( $V_x$ )	Goiter affection in % ( $V_y$ )
221	0,6
199	0,8
176	1,3
153	0,9
134	2,7
97	4,9
74	17,9

## 7.2. TEST CONTROL

### 7.2.1. Topic: «Organizing the statistic research. Graphic picture in statistics».

#### 1. Medical statistics is:

*Answer variants:*

- branch of statistics studying health of the population;
- complex of statistic methods necessary for analyzing the activity of public health services;
- branch of statistics studying the issues concerning medicine, hygiene, sanitation and health care.

#### 2. Stages of statistic research:

*Answer variants:*

- gathering statistic material;
- carrying out standardization;
- analyzing statistic material;
- working out, grouping, summing statistic material.

#### 3. Statistic totality is:

*Answer variants:*

- the primary element of the statistic observation, the bearer of the signs to be registered;

- b) the group consisting of relatively similar elements taken together in the known terms of time and space according to the given aim;
- c) a list of elements determining the totality of observation.

**4. The property of the representativeness is typical for the statistic totality:**

*Answer variants:*

- a) general;
- b) selective.

**5. Grouping, working out, summing up the material are the stage of the statistic research:**

*Answer variants:*

- a) the first;
- b) the second;
- c) the third.

**6. By the time statistic observation may be:**

*Answer variants:*

- a) current (constant);
- b) continuous;
- c) selective.

**7. The program of collecting the material includes:**

*Answer variants:*

- a) choosing the object of research;
- b) extracting the data;
- c) controlling the quality of registration;
- d) working out the pattern of the statistic document.

**8. Statistic subject is:**

*Answer variants:*

- a) the studied totality in its main phenomena;
- b) the composite quality characteristics of the statistic totality by lines;
- c) the main registration sign analyzed in the table.

**9. Statistic predicate is:**

*Answer variants:*

- a) the sum of the indices in the graphs of the table;
- b) indices of the horizontal row (by lines);
- c) the studied groups of the statistic totality;
- d) signs used for characterizing the subject and located in the graphs of the table.

**10. One-time observations are:**

*Answer variants:*

- a) birth rate;
- b) preventive examination of population;
- c) listing the in-patients.

**11. The plan of the statistic research includes:**

*Answer variants:*

- a) determining the kinds of observation;
- b) choosing the aim and purposes of the research;
- c) organizing the research, financing, etc.;
- d) working out the statistic document.

**12. The current observation is:**

*Answer variants:*

- a) listing the population;
- b) registrating the new-born;
- c) registrating mortality.

**13. The fourth stage of the statistic research includes:**

*Answer variants:*

- a) analysis of the received data;
- b) graphic pictures (images).

**14. By the scope of the observation units the statistic totality may be:**

*Answer variants:*

- a) general;
- b) selective;
- c) one-time.

**15. By character the registration signs of the statistic totality may be:**

*Answer variants:*

- a) quantitative;
- b) attributive;
- c) resultive.

**16. By role in the totality, the registration signs of the statistic totality may be:**

*Answer variants:*

- a) factor;
- b) resultive;
- c) selective.

**17. Making the plan and program of the research is:**

*Answer variants:*

- a) the first stage of the statistic research;
- b) the second stage of the statistic research.

**18. Ways of getting the primary information:**

*Answer variants:*

- a) immediate observation;
- b) extracting the data from the primary medical documents;
- c) making a family history (anamnesis);
- d) at random.

**19. Tables may be:**

*Answer variants:*

- a) simple;
- b) group;
- c) combined;
- d) resulting.

**7.2.2. Topic: «Relative values. Mean values»**

**1. While computing the intensive coefficients it's necessary to know the statistic totalities:**

*Answer variants:*

- a) one;
- b) two.

**2. While computing the extensive coefficients it's necessary to know the statistic totalities:**

*Answer variants:*

- a) one;
- b) two.

**3. The structure of the studied phenomenon is characterized by the coefficient:**

*Answer variants:*

- a) intensive;
- b) extensive;
- c) visuality;
- d) correlation.

**4. Frequency of the phenomenon in the particular medium is characterized by the coefficient:**

*Answer variants:*

- a) intensive;
- b) extensive;

- c) visuality;
- d) correlation.

**5. Index of correlation characterizes:**

*Answer variants:*

- a) changing of the phenomenon within time;
- b) relation of two independent totalities;
- c) distribution of the whole into parts.

**6. Coefficient of correlation according to the computing method is close to the coefficient:**

*Answer variants:*

- a) visuality;
- b) extensive;
- c) intensive;
- d) relative intensity.

**7. Indicate the extensive coefficient:**

*Answer variants:*

- a) the average life-span;
- b) mortality of population;
- c) number of girls amongst the new-born;
- d) provision of population with hospital beds;
- e) the rate of growth;
- f) dynamics of birth rate within 10 years.

**8. Indicate the intensive coefficient:**

*Answer variants:*

- a) the rate of growth;
- b) provision of population with doctors;
- c) percent rate of blood circulation diseases as a cause of death;
- d) infantile mortality.

**9. Indicate the coefficient of correlation:**

*Answer variants:*

- a) correlation of doctors and nursing staff;
- b) number of disablement days for 100 workers;
- c) dynamics of birth rate of population in Gomel region;
- d) the average duration of staying in-patients;
- e) the number of 1-year-old children at the pediatric district;
- f) the rate of growth.

**10. The average means are used for characterizing the signs:**

*Answer variants:*

- a) qualitative;
- b) quantitative.

**11. The necessary condition for receiving the average means is:**

*Answer variants:*

- a) having the homogenous group of data;
- b) knowing the formulae of computing;
- c) making the variable row.

**12. The average means are:**

*Answer variants:*

- a) Mode, median;
- b) progressive mean, regressive mean;
- c) harmonic mean, geometric mean.

**13. The variable row is:**

*Answer variants:*

- a) the statistic row showing the distribution of the studied phenomenon by the quantitative sign;
- b) the statistic row consisting of two main elements: variances (V), frequencies (P), H;
- c) the row of value distribution by some qualitative sign, located in increasing or decreasing order.

**14. Formulae for computing the arithmetical mean are:**

*Answer variants:*

- a)  $M = \frac{\sum v \cdot p}{\sum P}$ ;
- b)  $M = \frac{\sum viP}{N}$ ;
- c)  $M = A + \frac{\sum d \cdot P}{N}$ .

**15. Definition of the mode:**

*Answer variants:*

- a) the average value expressing the most frequently occurring typical values of the signs;
- b) the variance corresponding to the highest frequency.

**16. Definition of the median:**

*Answer variants:*

- a) dividing the variation row into two equal halves;
- b) the average value not depending on the values of the marginal variances.

**17. The average values are computed at:**

*Answer variants:*

- a) the first stage of the statistic research;
- b) the third stage of the statistic research.

**7.2.3. Topic: «Parametric methods of reliability estimation of the statistic research. Correlative analysis».**

**1. The average quadratic deviation is calculated to determine:**

*Answer variants:*

- a) representativeness mistakes;
- b) degree of variation in the variable row;
- c) the reliability of the difference between the mean values.

**2. The data necessary to determine the mean mistake (m) the relative value (P):**

*Answer variants:*

- a) intensive index;
- b) mode;
- c) the index of visuality;
- d) total number of observations.

**3. The data necessary to determine the mean mistake (m) the average value (M):**

*Answer variants:*

- a) the arithmetic mean;
- b) the average quadratic deviation;
- c) number of observations.

**4. What is confines possible random fluctuations of the mean value (M) with a probability of 95.5%:**

*Answer variants:*

- a)  $M \pm m$ ;
- b)  $M \pm 2m$ ;
- c)  $M \pm 3m$ .

**5. The functional connection is expressed by:**

*Answer variants:*

- a) tables;
- b) formulae.

**6. The correlative connection is defined as the connection:**

*Answer variants:*

- a) in which the value of each quantity of one sign corresponds to several values of another interconnected sign;



b) in which any value of one of the signs corresponds to the definite value of another interconnected sign.

**7. Indicate the formula for computing the correlation coefficient by Pirson:**

*Answer variants:*

a)  $1 - \frac{6\sum d^2}{n(n^2 - 1)}$ ;

b)  $\frac{\sum d_x \cdot d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}}$ ;

c)  $\frac{u_{xy}}{m_q}$ .

**8. The correlation coefficient is considered reliable only if its value is several times its average mistake:**

*Answer variants:*

a) 3 and more times;

b) twice.

**9. If the correlation coefficient is equal to 1, it indicates to:**

*Answer variants:*

a) the functional connection between the phenomena;

b) the correlative connection between the phenomena.

**10. If the correlation coefficient is equal to 0, it indicates to:**

*Answer variants:*

a) the absence of connection between the phenomena;

b) the functional connection between the phenomena.

**11. The correlative connection may be direct and:**

*Answer variants:*

a) inverse;

b) conjugate.

**12. The value of the correlation coefficient  $r = 0,6$  indicates to the strength of connection between the phenomena:**

*Answer variants:*

a) middle;

b) weak.

## LITERATURE

### THE BASIC LITERATURE

1. Valchuk, E. A. Bases of organizational-methodical service and statistic analysis in health care service / E. A. Valchuk, N. I. Gulitskaya, F. P. Tsaruk. — Minsk, 2003. — 381 p.
2. Glushanko, V. S. Social health and health care /Lecture course / V. S. Glushanko. — Vitebsk, 2001. — 359 p.
3. Lisitsin, Y. P. Social health and health care / Manual for higher schools / Y. P. Lisitsin. — Moscow, 2002. — 520 p.
4. Social health and health care / Edited by V. A. Minyaeva, N. I. Vishnyakova. — M., 2003. — 520 p.
5. Sharshakova, T. M. Course of lectures in social health and health care: Lecture course / T. M. Sharshakova. — Gomel, 2004. — 154 p.

### ADDITIONAL LITERATURE

1. Alehnovich, M. V. Methods and organizing the statistic research / M. V. Alehnovich. — Grodno, 2001. — 24 p.
2. Merkov, A. M. Sanitary statistics/ A. M. Merkov, L. E. Polyakov. — L.: Medicine, 1977. — 384 p.
3. Applying the methods of statistic analysis for studying social health and health care. Manual. / Edited by V. Z. Kucharenko. — Moscow, 2005. — 188 p.
4. Teaching medical statistics. Twenty lecture notes and seminars / Edited by S. K. Lvanga and Chzho-Ek Tya. — WHO, Zheneva: Medicine, 1989. — 215 p.

## SUPPLEMENT 1

The number of degrees equals n	$\chi^2$ 0,95	$\chi^2$ 0,99	The number of degrees equals n	$\chi^2$ 0,95	$\chi^2$ 0,99
1	3,84	6,64	20	31,41	37,57
2	5,99	9,21	21	32,67	38,93
3	7,82	11,34	22	33,92	40,29
4	9,49	13,28	23	35,17	41,64
5	11,07	15,09	24	36,42	42,98
6	12,69	16,81	25	37,65	44,31
7	14,07	18,48	26	38,89	45,64
8	15,51	20,09	27	40,11	46,96
9	16,92	21,67	28	41,34	48,28
10	18,31	23,21	29	42,56	49,59
11	19,68	24,72	30	43,77	50,89
12	21,03	26,22	40	55,76	63,69
13	22,36	27,69	50	67,50	76,15
14	23,68	29,14	60	79,08	88,38
15	25,00	30,58	70	90,53	100,42
16	26,30	32,00	80	101,88	112,33
17	27,59	33,41	90	113,14	124,12
18	28,87	34,81	100	124,34	185,81
19	30,14	36,19	-	-	-

Учебное издание

**Соболева Людмила Григорьевна**  
**Петрова Надежда Петровна**  
**Шаршакова Тамара Михайловна**  
**Дорофеев Виталий Михайлович**

**ОСНОВЫ**  
**МЕДИЦИНСКОЙ СТАТИСТИКИ**  
(на английском языке)

**Учебно-методическое пособие**  
**для студентов 4 и 6 курсов**  
**факультета по подготовке специалистов для зарубежных стран,**  
**обучающихся по специальности «Лечебное дело»,**  
**медицинских вузов**

**Редактор *Т. Ф. Рулинская***  
**Компьютерная верстка *С. Н. Козлович***

Подписано в печать 22.01.2013.  
Формат 60×84<sup>1</sup>/<sub>16</sub>. Бумага офсетная 65 г/м<sup>2</sup>. Гарнитура «Таймс».  
Усл. печ. л. 2,09. Уч.-изд. л. 2,29. Тираж 70 экз. Заказ 24.

Издатель и полиграфическое исполнение  
Учреждение образования  
«Гомельский государственный медицинский университет»  
ЛИ № 02330/0549419 от 08.04.2009.  
Ул. Ланге, 5, 246000, Гомель.