PRACTICAL LESSON No. 2
Topic: STATISTICAL ANALYSIS

OBJECTIVE
1. Learn the basics of primary data processing;
2. Learn to construct a histogram and frequency polygon;
3. Learn to calculate point and interval estimates of sample data.

The student must study the material on the topic and be ABLE to answer the following questions:

1. General population and sample. Examples. Sample size, representativeness.
2. Statistical distribution (variation series). Bar graph.
3. Characteristics of position (mode, median, sample mean) and dispersion (sample variance and sample standard deviation).
4. Estimation of the parameters of the general population based on its sample (point and interval).
5. Confidence interval, confidence probability, level of significance.

Mathematical statistics is a branch of mathematics that studies methods for finding distribution laws and numerical characteristics based on the results of an experiment.
In mathematical statistics two main areas of research are distinguished:
1. Estimation of the parameters of the general population.
2. Testing statistical hypotheses (some a priori assumptions).

General population is the set of all conceivable values of observations (objects) that are homogeneous with respect to some attribute.
The number of all observations (objects) that make up a general population is called its size *N*. Studying an entire population is laborious and expensive, and perhaps simply impossible. Therefore, data are collected from a sample of objects which are considered to be representatives of this population, allowing conclusions to be drawn about the population.
A sample is a collection of randomly selected observations (objects) for direct study from the general population. Sample size is denoted by *n*. The sample must necessarily satisfy the condition of representativeness, i.e. to give an informed

view of the general population. To form a representative sample the elements of a sample must be taken in a random way.

Each value x_i of a sample element property x is called a **variant**.

The number of observations of variants m_i is called the (absolute) **frequency** of occurrence.

A sequence of variants, written in ascending order, is called a **variation series**.

A **statistical distribution** is a set of variants x_i and the corresponding frequencies m_i.

Problem 2.1.

Sample size is 20. Statistical distribution is specified.

x_i	2	6	12
m_i	3	10	7

Find the distribution of relative frequencies.

Solution: to find relative frequencies the absolute frequencies must be divided by the sample size:

$$\frac{m_1}{n}=\frac{3}{20}=0{,}15\,; \quad \frac{m_2}{n}=\frac{10}{20}=0{,}5\,; \quad \frac{m_3}{n}=\frac{7}{20}=0{,}35\,.$$

The distribution of relative frequencies is as follows:

x_i	2	6	12
m_i	3	10	7
m_i/n	0,15	0,5	0,35

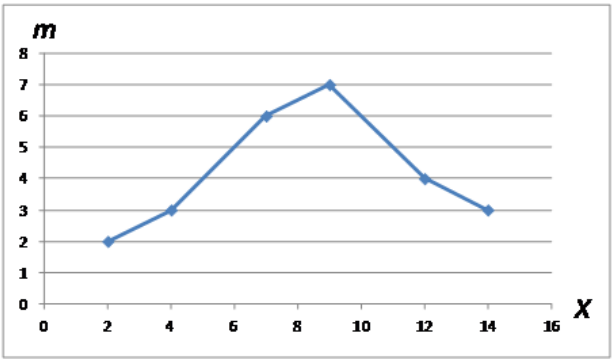
To visualize statistical distribution, a graphical representation of the variation series is used: a **polygon** and a **histogram**.

To draw a **polygon** the variants values x_i are plotted on the OX axis, and the values of frequencies m_i (or relative frequencies f_i) are plotted on the OY axis. The polyline, the segments of which connect the points (x_i; m_i) or (x_i; f_i) is called the polygon of frequencies (relative frequencies).

Problem 2.2.

The experimental results are given in the table. Draw the polygon of frequencies.

x	m	f
2	2	0,08
4	3	0,12
7	6	0,24
9	7	0,28
12	4	0,16
14	3	0,12



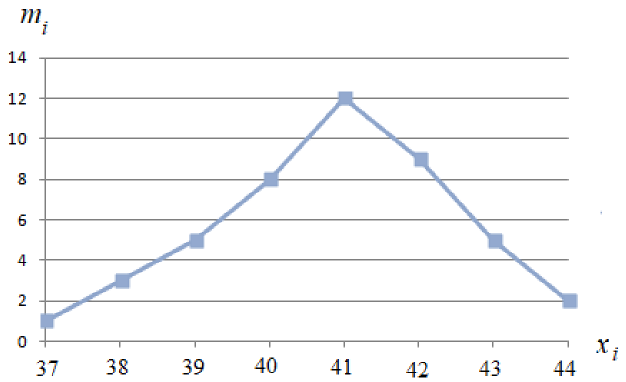
Problem 2.3.

Construct a discrete variation series and draw a polygon for the distribution of 45 applicants according to the number of points they received on entrance exams:
39 41 40 42 41 40 42 44 40 43 42 41 43 39 42 41 42 39 41 37 43 41 38 43 42 41 40 41 38 44 40 39 41 40 42 40 41 42 40 43 38 39 41 41 42.

Solution: To obtain a variation series the values of the attribute x must be arranged in ascending order and the corresponding frequencies of their appearance must be estimated.

x_i	37	38	39	40	41	42	43	44
m_i	1	3	5	8	12	9	5	2

The polygon of distribution:



A **histogram (bar chart)** is a stepwise figure, consisting of adjacent rectangles the bases of which are equal to the width h of an interval (bin, class) and the heights are equal to the frequencies (relative frequencies) of falling into a given interval.

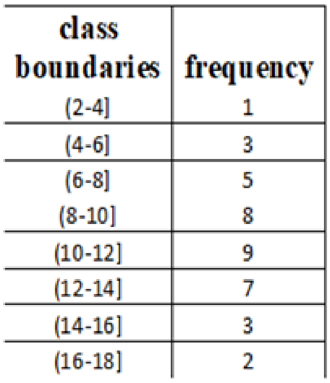
To draw a bar chart the whole interval of possible variant values is divided into equal partial intervals (classes, bins). The width *h* of a bin can be estimated according to Sturges’ formula:

$$h=\frac{x_{\max}-x_{\min}}{1+3{,}32\lg n},$$

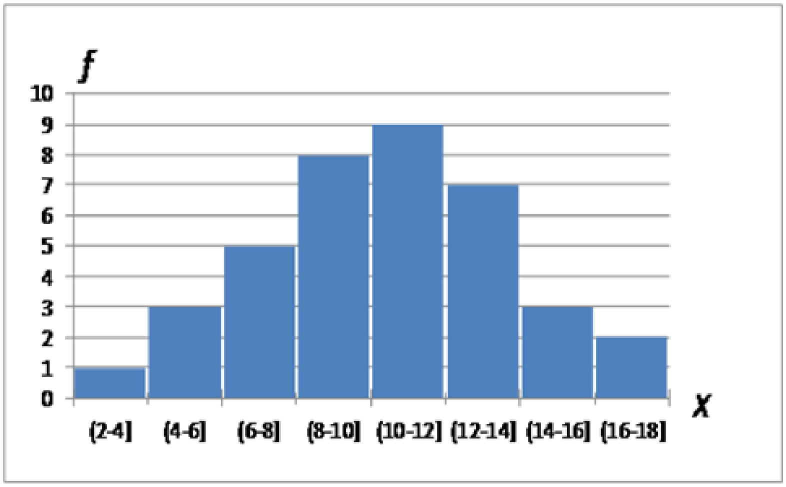
where $x_{\max}$ - maximum variant value, $x_{\min}$ – minimum variant value, *n* – sample size. (Usually the number of bins is 7 – 12).

Problem 2.4.

Plot the histogram for the data given in a table



class boundaries	frequency
(2-4]	1
(4-6]	3
(6-8]	5
(8-10]	8
(10-12]	9
(12-14]	7
(14-16]	3
(16-18]	2



Problem 2.5.

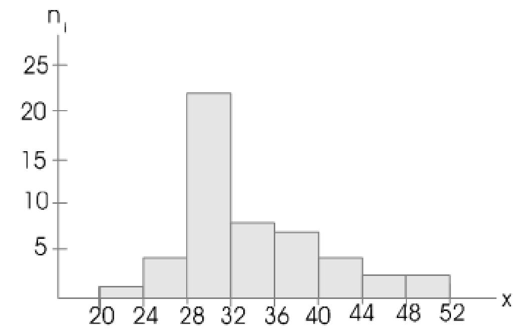
Observations of the number of particles entering the Geiger counter per minute gave the following results:

21 30 39 31 42 34 36 30 28 30 33 24 31 40 31 33 31 27 31 45 31 34 27 30 48 30 28 30 33 46 43 30 33 28 31 27 31 36 51 34 31 36 34 37 28 30 39 31 42 37

Construct an interval variation series with equal intervals (1-st interval 20-24; 2-nd interval 24-28, etc.) and draw a histogram.

Solution:

Interval	20-24	24-28	28-32	32-36	36-40	40-44	44-48	48-52
Frequency	1	4	22	8	7	4	2	2



STATISTICAL ESTIMATES OF DISTRIBUTION PARAMETERS. SAMPLE CHARACTERISTICS

CHARACTERISTICS OF POSITION

Sample mean value - that is arithmetic mean of the statistical series.

X	x_1	x_2	…	x_i	…	x_k
m	m_1	m_2	…	m_i	…	m_k

$$\bar{x}_s = \frac{m_1x_1 + m_2x_2 + ... + m_ix_i + ... + m_kx_k}{n} = \frac{\sum_{i=1}^{k} m_ix_i}{n}$$

Problem 2.6.

Hemoglobin concentration was measured in adult men. The results are given by a table below. Estimate the sample mean.

x, g/l	132	136	138	141	144	145	148
m	2	2	4	3	2	1	1

Solution:

Sample mean is:

$$\bar{x}_s = \frac{132 \cdot 2 + 136 \cdot 2 + 138 \cdot 4 + 141 \cdot 3 + 144 \cdot 2 + 145 \cdot 1 + 148 \cdot 1}{15} = 139{,}5$$

Problem 2.7.

In a sample of adult men *n* = 50, the blood hemoglobin content was determined. For n_1 = 30, it was equal to 70% on average. For another group of men n_2 = 20, this indicator was 50% on average. Find the arithmetic mean of these two means.

Solution: $\bar{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i = \frac{1}{50}(30 \cdot 70 + 20 \cdot 50) = 62\%$.

Mode of distribution (*Mo*) is such a variant that the preceding and the following variants have lower frequencies of occurrence.

For unimodal distributions, mode is the most often variant in a given population (the variant of greatest frequency).

For example, for the distribution given

x_i	16	17	18	20
f_i	3	8	11	6

the mode is equal to 18.

The calculation of the mode of classed data with equal bin widths (as in histograms) can be done according to the following equation:

$$Mo = x_{low} + h\frac{f_k - f_{k-1}}{2f_k - f_{k-1} - f_{k+1}},$$

with

k - the index of the bin containing the greatest number of objects (index of the modal class);

x_{low} - lower border of the bin containing the greatest number of objects;

h - bin width;

f_k - frequency of the k-th class (frequency of the modal class);

$f_{k\text{-}1}$, f_{k+1} - frequencies of the neighboring bins.

Problem 2.8.

Determine the mode of the distribution of calcium (mg%) in the blood serum of monkeys.

Bins	8,6-9,3	9,4-10.1	10,2-10,9	11,0-11,7	11,8-12,5	12,6-13,3	13,4-14,1	14,2-14,9
Frequency f_i	2	6	15	23	25	17	7	5

Solution: The frequency of the modal class n_2 = 25, its lower boundary 11,8. The frequency of the class preceding the modal is n_1 = 23; frequency of the class following the modal n_3 = 17; h = 0.8.

$$Mo = x_{low} + h\frac{f_k - f_{k-1}}{2f_k - f_{k-1} - f_{k+1}} = 11{,}8 + 0{,}8\frac{25-23}{2\cdot 25 - 23 - 17} = 11{,}96$$

Problem 2.9.

Determine the distribution mode.

Class intervals	2 - 4	4 - 6	6 - 8	8 - 10	10 - 12	12 - 14	14 - 16	16 - 18
m_i	1	3	5	8	9	7	3	2

Solution: $Mo = x_{low} + h\frac{f_k - f_{k-1}}{2f_k - f_{k-1} - f_{k+1}} = 10 + 2\frac{9-8}{2\cdot 9 - 8 - 7} = 10{,}7$.

Problem 2.10.

Find the mode of height distribution for 1000 adult males:

Height, cm	Number of men	Height, cm	Number of men
143-146	1	167-170	170
146-149	2	170-173	120
149-152	8	173-176	64

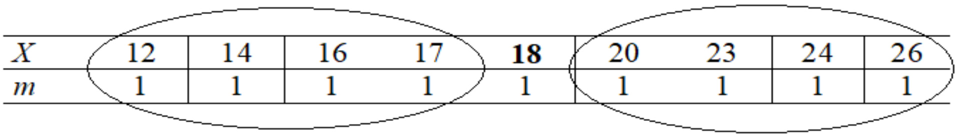
152-155	26	176-179	28
155-158	65	179-182	10
158-161	120	182-185	3
161-164	181	185-188	1
164-167	201		

Answer: 167,5 cm

Median, *Me* – the value of the variant, relatively to which the statistical series is divided into two parts equal by size.

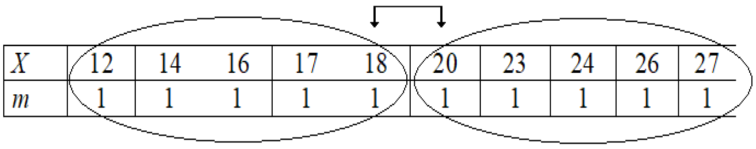
To find the median the set of data must be ordered (in an ascending way).

For the distribution shown below (odd number of measurements) *Me* = 18 (equal number of variants on both sides of 18).



X	12	14	16	17	**18**	20	23	24	26
m	1	1	1	1	1	1	1	1	1

For the series consisting of even number of variants the median is equal to half of central variants sum:



X	12	14	16	17	18	20	23	24	26	27
m	1	1	1	1	1	1	1	1	1	1

$$Me = \frac{18+20}{2} = 19 .$$

For the interval statistical series the median is calculated according to the formula:

$$Me = x_{\text{low}} + \frac{0{,}5n - n_{m-1}^{\text{acc}}}{n_m} h , \qquad \text{with}$$

n - statistical population size;

x_{low} - median interval lower boundary;

h - median interval length (bin size);

n_m - frequency of the median interval;

n_{m-1}^{acc} - accumulated frequency of the previous interval.

Problem 2.11.

Determine the median of distribution:

intervals	**frequencies,** n_i	**accumulated frequencies,** n_{m-1}^{acc}
20 – 24	10	10
24 – 28	20	30
28 – 32	50	80
32 – 36	15	95
36 – 40	5	100
	n = 100	

Solution: The third interval (28 – 32) is the median interval (the accumulated frequency for this interval is 80 (greater than $n/2 = 50$) and for the previous interval – 30, hence the median is in the third interval), so

$$Me = x_{\text{low}} + \frac{0{,}5n - n_{m-1}^{\text{acc}}}{n_m} h = 28 + \frac{0{,}5 \cdot 100 - 30}{50} 4 = 29{,}6 .$$

Problem 2.12.

Determine the median of distribution:

No. **of bin**	**intervals**	**frequencies,** n_i	**accumulated** **frequencies,** n_{m-1}^{acc}
1	20 – 25	8	
2	25 – 30	12	
3	30 – 35	27	
4	35 – 40	35	
5	40 – 45	18	
6	45 – 50	10	
		n = 110	

Answer: $Me = 36{,}1$

Sample mean value, mode, median are called the characteristics of position.

CHARACTERISTICS OF SCATTERING

Measures of statistical dispersion are the **variance, standard deviation and coefficient of variation.**

Sample **variance** is the arithmetic mean of the squares of the deviation of the variants from their mean value:

$$S_S^2 = \frac{1}{n}\sum_{i=1}^{\kappa}(x_i - \bar{x}_S)^2 \cdot n_i .$$

Standard deviation is the square root of the sample variance:

$$S_S = \sqrt{S_S^2} .$$

Coefficient of variation CV is the ratio of the standard deviation to the mean value of the attribute, expressed as a percentage:

$$CV = \frac{S_S}{\bar{x}_S} \cdot 100\% .$$

The coefficient of variation is a measure of the relative variability of a random variable that allows to compare variability of dissimilar values, for example, heart rate (beats/min), blood pressure (mm Hg) and temperature (° C) in relative units - percent.

The **variation range** is the difference between the highest and the lowest values of an attribute:

$$\Delta = x_{\max} - x_{\min} .$$

Problem 2.13.

The sample is given by the distribution table:

x_i	1	2	3	4
n_i	20	15	10	5

Find the sample variance.

Solution:

$$\bar{x}_S = \frac{20 \cdot 1 + 15 \cdot 2 + 10 \cdot 3 + 5 \cdot 4}{20 + 15 + 10 + 5} = 2;$$

$$S_S^2 = \frac{20(1-2)^2 + 15(2-2)^2 + 10(3-2)^2 + 5(4-2)^2}{20 + 15 + 10 + 5} = 1.$$

Problem 2.14.

Compare the 2 varying attributes.

One is characterized by mean value $\bar{x}_1 = 2{,}4\ \text{kg}$ and standard deviation $S_1 = 0{,}58\ \text{kg}$; the other is characterized by the values $\bar{x}_2 = 8{,}3\ \text{cm}$ and standard deviation $S_2 = 1{,}57\ \text{cm}$. Which attribute varies more strongly?

Solution:

$$CV_1 = \frac{S_1}{\overline{x}_1} \cdot 100\% = \frac{0.58}{2.4} \cdot 100\% = 24.2\%;$$

$$CV_2 = \frac{S_2}{\overline{x}_2} \cdot 100\% = \frac{1.57}{8.3} \cdot 100\% = 18.9\%.$$

Answer: $CV_1 > CV_2$, first attribute varies more strongly.

ESTIMATION OF THE GENERAL POPULATION PARAMETERS BASED ON A SAMPLE

The numerical values that characterize the general population are called parameters. One of the tasks of mathematical statistics is to determine the parameters of a large array by examining its part.

The point of statistical methods is that based on estimates obtained for a sample it is possible to make a reasoned conclusion about the parameters of the general population.

Statistical estimation (evaluation) can be done in two ways:

1) Point estimate - an estimate that is given by a certain number.

2) Interval estimate - according to the sample data, the interval is estimated in which the true value lies with a given probability.

POINT ESTIMATES OF THE GENERAL POPULATION PARAMETERS

A point estimate of a general parameter is an estimate given by one number, which is obtained according to a sample (it is determined by a sample and is a function of the sample results).

The quality of the estimate is established according to three properties: it must be consistent, effective, and unbiased.

A point estimate is called **consistent** if, with an increase in the sample size, the sample characteristic (estimate) tends to the corresponding characteristic of the general population (parameter).

A point estimate is called **unbiased** if its mathematical expectation is equal to the estimated parameter.

A point estimate is said to be **effective** if it has the smallest variance of all unbiased estimates of a given property.

Sample mean $\bar{x}_S = \frac{1}{n}\sum_{i=1}^{K} x_i n_i$ is an unbiased estimate of the general mean (mathematical expectation):

$$M(\bar{x}_S) = \mu .$$

Sample variance is not an unbiased estimate. This is a biased estimate of the general variance:

$$M\left(S_S^2\right) \neq \sigma_{gen}^2$$

$$M\left(S_S^2\right) = \frac{n-1}{n}\sigma_{gen}^2 .$$

So, a corrected variance $S_{corrected}^2$ is used for which

$$M\left(S_{corrected}^2\right) = \sigma_{gen}^2 .$$

Then $$S_{corrected}^2 = \frac{n}{n-1} S_S^2 ;$$

$$S_{corrected}^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x}_S)^2 \cdot n_i .$$

Corrected standard deviation (r. m. s. d.)

$$S = \sqrt{S_{corrected}^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x}_S)^2 \cdot n_i} .$$

And the accuracy of sample mean is given by sample mean standard deviation also called the **standard error of sample mean,** which is estimated as:

$$m_{\bar{x}_S} = \frac{S}{\sqrt{n}} .$$

Problem 2.15.

A sample is taken out of general population:

x_i	2	5	7	10
n_i	16	12	14	8

Calculate sample mean value, sample corrected variance, sample corrected standard deviation and sample mean standard error.

Solution: $\bar{x}_s = \frac{\sum_{i=1}^{k} m_i x_i}{n} = \frac{2\cdot16 + 5\cdot12 + 7\cdot14 + 10\cdot8}{50} = 5{,}4$

x_i	2	5	7	10
n_i	16	12	14	8
$(x_i - \bar{x}_S)$	2 – 5,4	5 – 5,4	7 – 5,4	10 – 5,4

$(x_i - \bar{x}_S)^2$	$(2 – 5,4)^2$	$(5 – 5,4)^2$	$(7 – 5,4)^2$	$(10 – 5,4)^2$

$$S^2_{corrected} = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x}_S)^2 \cdot n_i =$$

$$= \frac{16(2-5,4)^2 + 12(5-5,4)^2 + 14(7-5,4)^2 + 8(10-5,4)^2}{50-1} = 8$$

$$S = \sqrt{S^2_{corrected}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x}_S)^2 \cdot n_i} = \sqrt{8} = 2,8$$

$$m_{\bar{x}_S} = \frac{S}{\sqrt{n}} = \frac{2,8}{\sqrt{50}} = 0,4.$$

Problem 2.16.

Calculate the unbiased estimate of the general mean, population variance, and standard deviation over a sample of *n* = 12, describing the duration in seconds of physical activity before the onset of an angina attack. Calculate sample mean standard error.

289, 203, 359, 243, 232, 210, 251, 246, 224, 239, 220, 211.

Solution:

$$\bar{x}_s = \frac{289 + 203 + 359 + 243 + 232 + ... + 239 + 220 + 211}{13} = 244$$

$$S^2_{corrected} = \frac{(289-244)^2 + (203-244)^2 + (359-244)^2 + ... + (211-244)^2}{12-1} = 1849$$

$$S = \sqrt{S^2_{corrected}} = \sqrt{1849} = 43$$

$$m_{\bar{x}_S} = \frac{S}{\sqrt{n}} = \frac{43}{\sqrt{12}} = 12,4.$$

Problem 2.17.

A sample is taken out of general population, sample distribution given by the table:

x_i	2	5	10	12	15	16
n_i	5	8	14	8	3	1

Calculate sample mean value, sample corrected variance, sample corrected standard deviation and sample mean standard error.

INTERVAL EVALUATION. CONFIDENCE INTERVAL AND CONFIDENCE PROBABILITY.

In some cases, it is not a point estimate of the general parameter that is of interest, but the determination of a certain interval which, with a given probability, covers (includes) this parameter.

The confidence interval (a, b) for the general parameter Θ is an interval relative to which it is possible, with a preselected probability close to one, to assert that it contains an unknown parameter value.

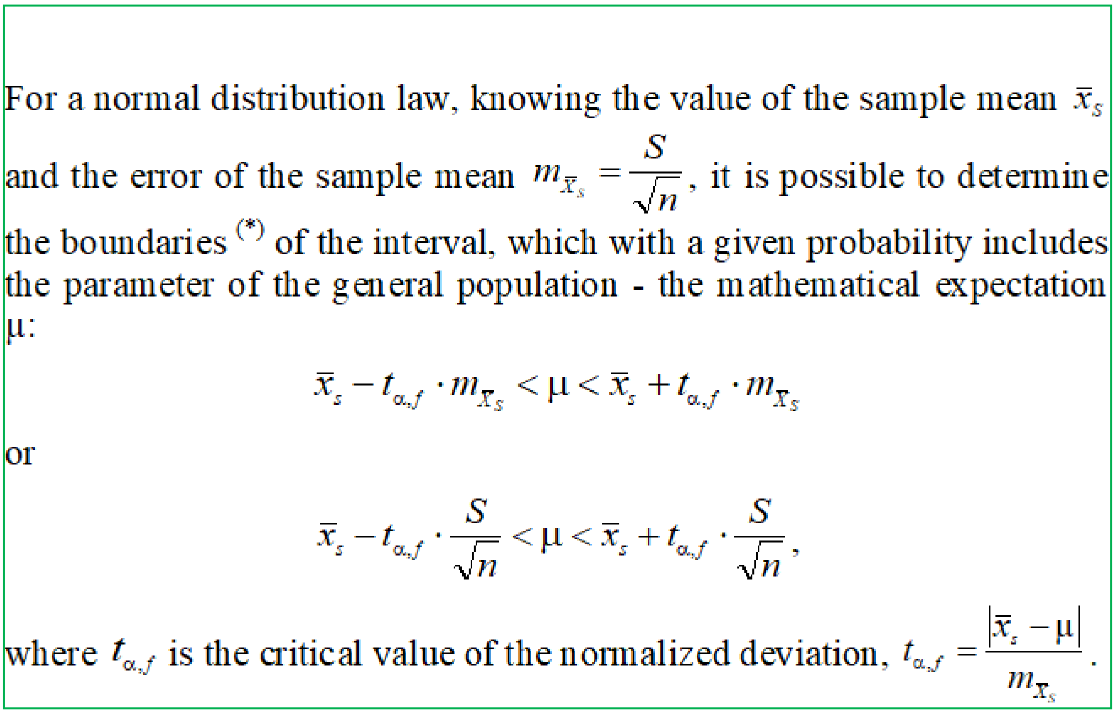
In statistics, a confidence interval (CI) is a type of estimate computed from the statistics of the observed data.

The probability with which the parameter of the general population is guaranteed to fall within the confidence interval is called the confidence probability.

More often probabilities $P_1 = 0.95$; $P_2 = 0.99$; $P_3 = 0.999$ are used as confidential.

So, for instance, the confidence probability $P = 0.95$ means that for 100 samples of the same size taken from a given general population, in 95 cases the confidence interval constructed from the sample will cover the estimated general parameter.

In some cases, it is not the confidence probability P that is indicated, but the probability of opposite event, when the confidence interval does not cover the general parameter. The probability of such unlikely events is called the level of significance α and is equal: $\alpha = 1 - P$.



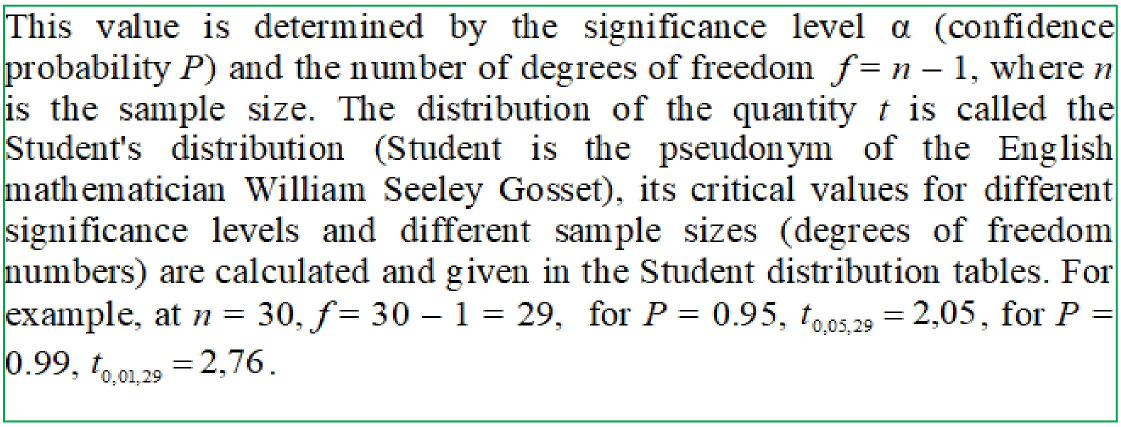
For a normal distribution law, knowing the value of the sample mean $\bar{x}_s$ and the error of the sample mean $m_{\bar{x}_s} = \dfrac{S}{\sqrt{n}}$, it is possible to determine the boundaries [(*)] of the interval, which with a given probability includes the parameter of the general population - the mathematical expectation μ:

$$\bar{x}_s - t_{\alpha,f} \cdot m_{\bar{x}_s} < \mu < \bar{x}_s + t_{\alpha,f} \cdot m_{\bar{x}_s}$$

or

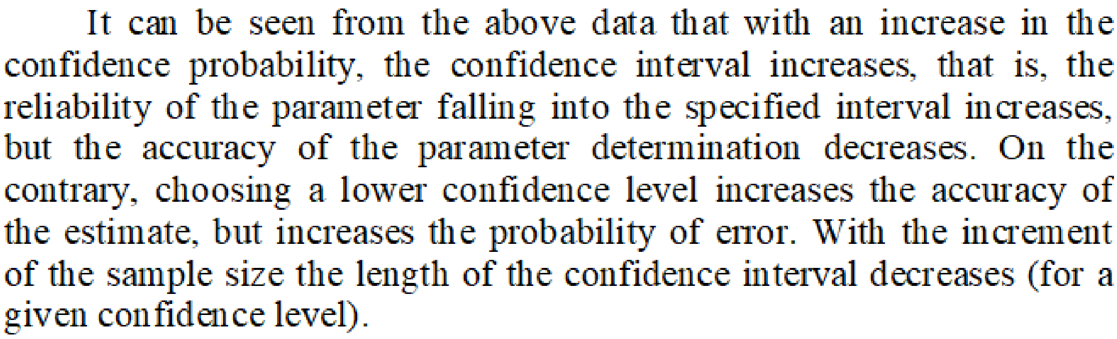
$$\bar{x}_s - t_{\alpha,f} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{x}_s + t_{\alpha,f} \cdot \frac{S}{\sqrt{n}},$$

where $t_{\alpha,f}$ is the critical value of the normalized deviation, $t_{\alpha,f} = \dfrac{|\bar{x}_s - \mu|}{m_{\bar{x}_s}}$.



This value is determined by the significance level α (confidence probability P) and the number of degrees of freedom $f = n - 1$, where n is the sample size. The distribution of the quantity t is called the Student's distribution (Student is the pseudonym of the English mathematician William Seeley Gosset), its critical values for different significance levels and different sample sizes (degrees of freedom numbers) are calculated and given in the Student distribution tables. For example, at $n = 30$, $f = 30 - 1 = 29$, for $P = 0.95$, $t_{0,05,29} = 2{,}05$, for $P = 0.99$, $t_{0,01,29} = 2{,}76$.

Despite the fact that we have *n* terms (sample size is *n*), this distribution will have (*n* – 1) degrees of freedom, since knowing the sample mean and (*n* – 1) sample elements, we can always accurately specify the last element.



It can be seen from the above data that with an increase in the confidence probability, the confidence interval increases, that is, the reliability of the parameter falling into the specified interval increases, but the accuracy of the parameter determination decreases. On the contrary, choosing a lower confidence level increases the accuracy of the estimate, but increases the probability of error. With the increment of the sample size the length of the confidence interval decreases (for a given confidence level).

The width of the confidence interval depends on *m* - the mean error of the sample mean, which in turn depends on the sample size (*n*) and on the variability of the data (*S*). If the sample is small, then the confidence interval is wider than in the case of a large sample. A wide confidence interval indicates an inaccurate estimate, and a narrow confidence interval indicates an accurate estimate.

Table 1. Critical points of two-tailed Student's *t*-criterion

Number of degrees of freedom *f*	Levels of significance α, % (two-sided test)			Number of degrees of freedom *f*	Levels of significance α, % (two-sided test)		
	5	1	0,1		5	1	0,1
1	12,71	63,66	64,60	18	2,10	2,88	3,92
2	4,30	9,92	31,60	19	2,09	2,86	3,88
3	3,18	5,84	12,92	20	2,09	2,85	3,85
4	2,78	4,60	8,61	21	2,08	2,83	3,82
5	2,57	4,03	6,87	22	2,07	2,82	3,79
6	2,45	3,71	5,96	23	2,07	2,81	3,77
7	2,37	3,50	5,41	24	2,06	2,80	3,75
8	2,31	3,36	5,04	25	2,06	2,79	3,73
9	2,26	3,25	4,78	26	2,06	2,78	3,71
10	2,23	3,17	4,59	27	2,05	2,77	3,69
11	2,20	3,11	4,44	28	2,05	2,76	3,67
12	2,18	3,05	4,32	29	2,05	2,76	3,66
13	2,16	3,01	4,22	30	2,04	2,75	3,65
14	2,14	2,98	4,14	40	2,02	2,70	3,55
15	2,13	2,95	4,07	60	2,0	2,66	3,46
16	2,12	2,92	4,02	120	1,98	2,62	3,37
17	2,11	2,90	3,97	∞	1,96	2,58	3,29

(Example: from the table it can be seen that

- for the sample size n = 20 and the significance level α = 5% the critical value of normalized t-value $t_{\alpha;f}$ = $t_{0,05;\,19}$ = 2,09;

- for the sample size n = 20 and the significance level α = 1% the critical value of normalized t-value $t_{\alpha;f}$ = $t_{0,01;\,19}$ = 2,86).

Problem 2.18.

The quantitative characteristic x of the general population is distributed normally. For a sample of size n = 16, the sample mean and standard deviation were found: $\bar{x}_S = 20{,}2; \quad S = 0{,}8$.

Determine the confidence interval for unknown mathematical expectation at confidence probability $p \geq 0{,}95$.

Solution:

$$\bar{x}_S - \frac{S}{\sqrt{n}} t < \mu < \bar{x}_S + \frac{S}{\sqrt{n}} t .$$

Let us find t from the Student distribution table at the significance level $\alpha \leq 0{,}05$ and the number of degrees of freedom f = n – 1 = 16 – 1 = 15:

$$t_{0,05;\,15} = 2{,}13.$$

Then, $20{,}2 - \frac{0{,}8}{\sqrt{16}} \cdot 2{,}13 < \mu < 20{,}2 + \frac{0{,}8}{\sqrt{16}} \cdot 2{,}13;$

$19{,}8 < \mu < 20{,}6$ (at $p \geq 0{,}95$).

Problem 2.19.

The results of measuring the systolic pressure (mm Hg) in 11 men (sample size n = 11) at the initial stage of shock are given:

x: 127, 124, 155, 129, 77, 147, 65, 109, 145, 141, 158.

Determine the sample mean value, corrected sample variance, standard error of sample mean. Construct the confidence interval for general mean (μ) at confidence probability $p \geq 0{,}95$.