PhD Misiura Ie.Iu. (доцент Місюра Є.Ю.)

Theme: Problems of mathematical statistics. Primary processing of statistical data. Statistical estimates of distribution parameters. Statistical evaluation methods in international trade

PART 1. Primary processing of statistical data

MATHEMATICAL STATISTICS

THE OBJECTIVE OF THIS COURSE is

TO INTRODUCE

A MATHEMATICAL STATISTICS

TO STUDENTS

MATHEMATICAL STATISTICS

MS is the study of statistics with a mathematical standpoint using methods of probability theory.

MS deals with an information from data. In practice, data contain some randomness *(случайность)* and uncertainty *(неопределенность)*.

MATHEMATICAL STATISTICS

The basic objects of MS

are called

A POPULATION and A SAMPLE

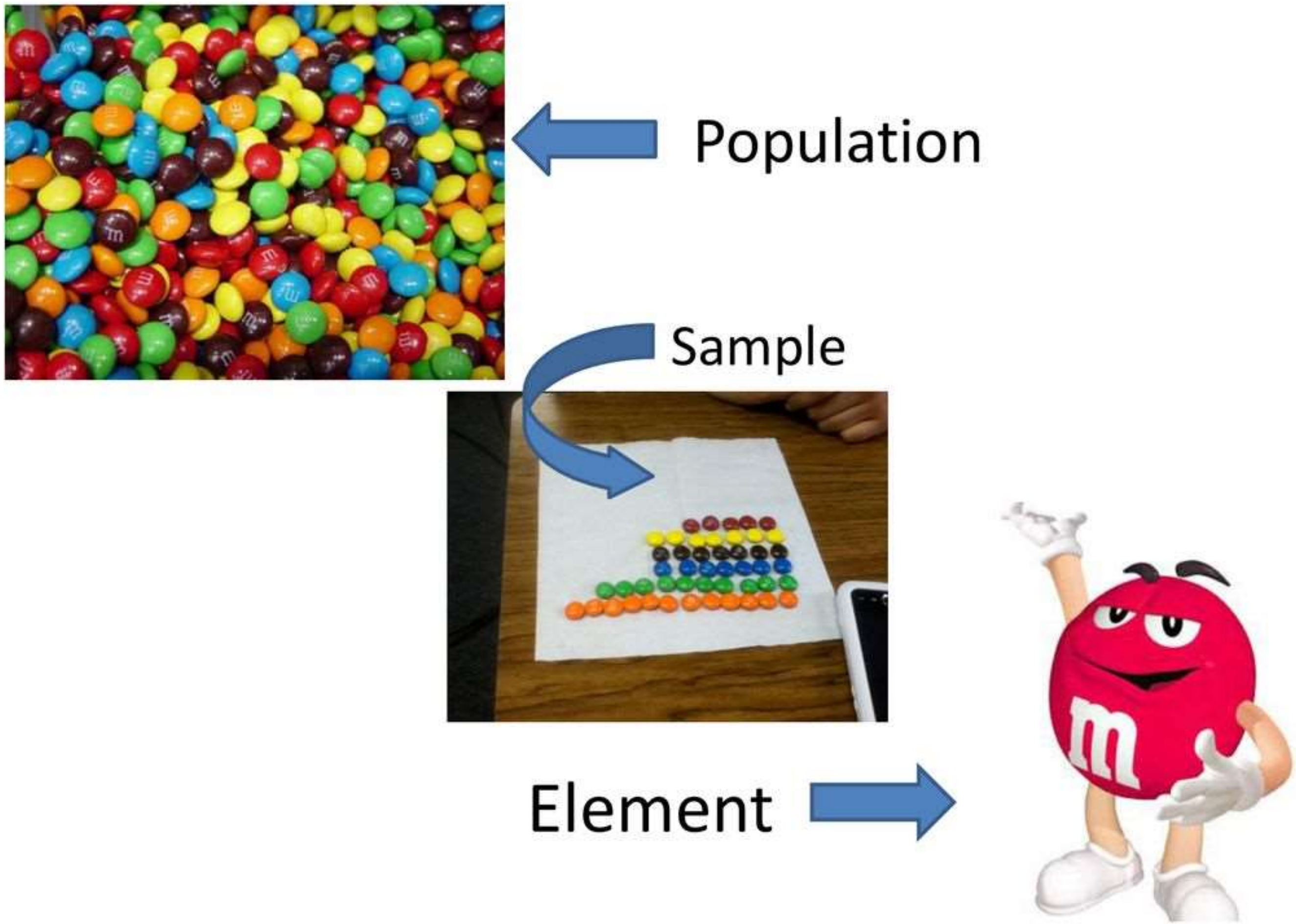
MATHEMATICAL STATISTICS

The basic objects of MS are called

A POPULATION and A SAMPLE

(генеральна сукупність) **(виборка)**

Example



POPULATION
(генеральна сукупність)

A ***population*** is a space of all elements (data) obtaining in a particular experiment.

A ***population size*** (***N)*** is a number of all elements in this population.

POPULATION

(генеральна сукупність)

A ***population*** is a space of all elements (data) obtaining in a particular experiment.

A ***population size*** (***N)*** is a number of all elements in this population.

We can consider **any set of things having the same property** in a certain sense, e.g., every article of a certain production process or all students of our university.

POPULATION

Sometimes a population size can be **very large**, **even practically infinite** and then we take its **subset** (**sample**) and work up data of this sample. After this we conclude the results for a population with the help of the interval estimations for the numerical characteristics.

SAMPLE (виборка)

In order not to check the total population about the considered property, **data are collected** only from a subset, from a so-called ***sample*** of size ***n*** ***(n < N).*** We talk about a random choice if every element of the population has the same chance of being chosen.

A *sample of size n* from a finite population of size *N* is **a sample** selected such that each possible sample of size ***n*** has the same probability of being selected.

For example, a sample is a group of students.

Descriptive Statistics

A random variable *X* can be characterized by its distribution function, by its parameters, where the distribution function itself is determined completely by the properties of the population. These are unknown at the beginning of a statistical investigation, so we want to collect as much information as possible with the help of *samples*.

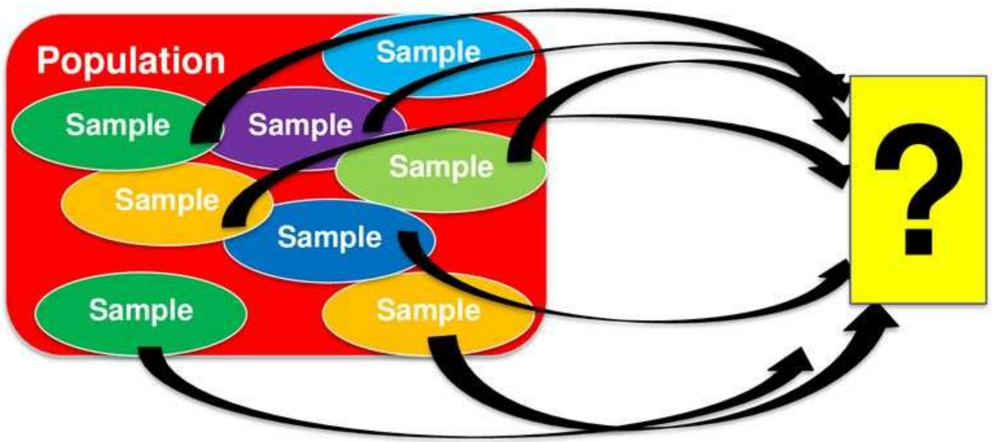
Usually we do not restrict our investigation to one sample but we apply more samples (with same size *n* if it is possible, for practical reasons).

The elements of a sample are chosen randomly, so the realizations take their values randomly, i.e., the first value of the first sample is usually different from the first value of the second sample. Consequently, the first value of a sample is a random variable itself denoted by X_1.

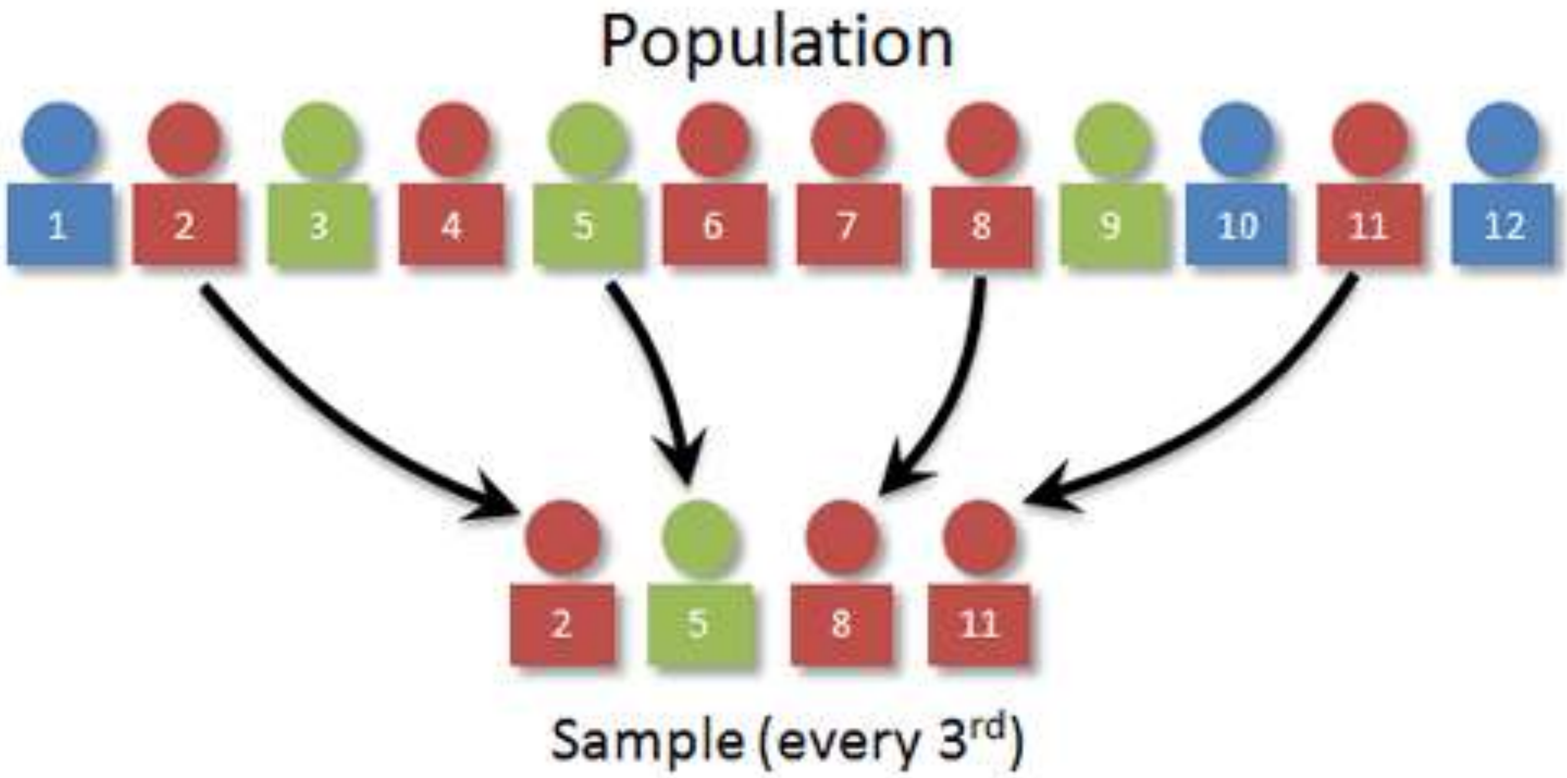
Analogously, we can introduce the random variables $X_2, X_3, \dots, X_n$ for the second, third,..., *n*-th sample values, and they are called *sample variables.* Together, they form the *random vector* $X = (X_1, X_2, X_3, \dots, X_n)$

Example

Sampling Variability



Example



Let ***X*** be a random variable, then

$X = (X_1, X_2, X_3, ..., X_N)$ is a population and

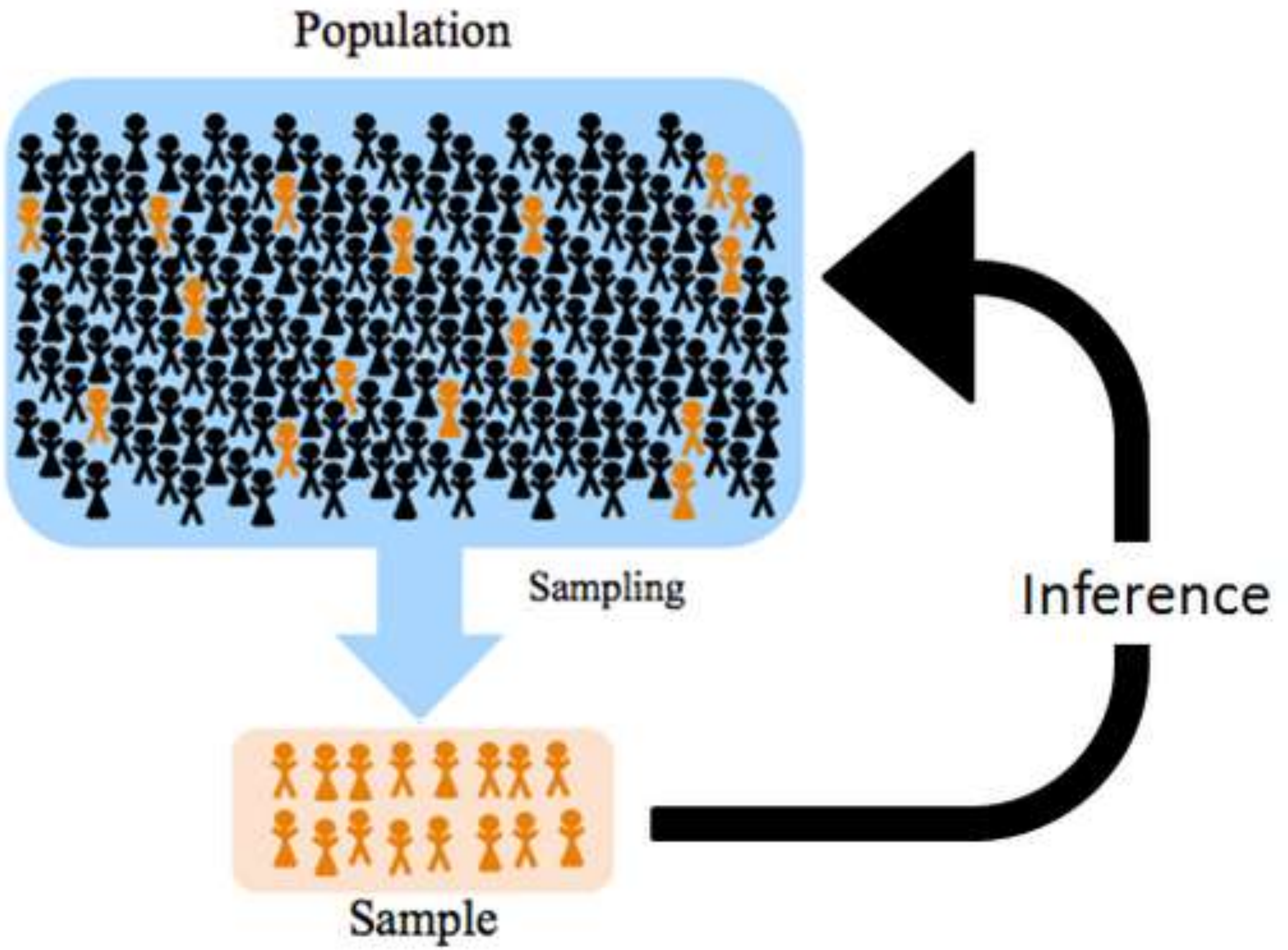
$x = (x_1, x_2, x_3, ..., x_n)$ is a sample.

To describe the properties of a population (if a population has a large size) we use a sample.

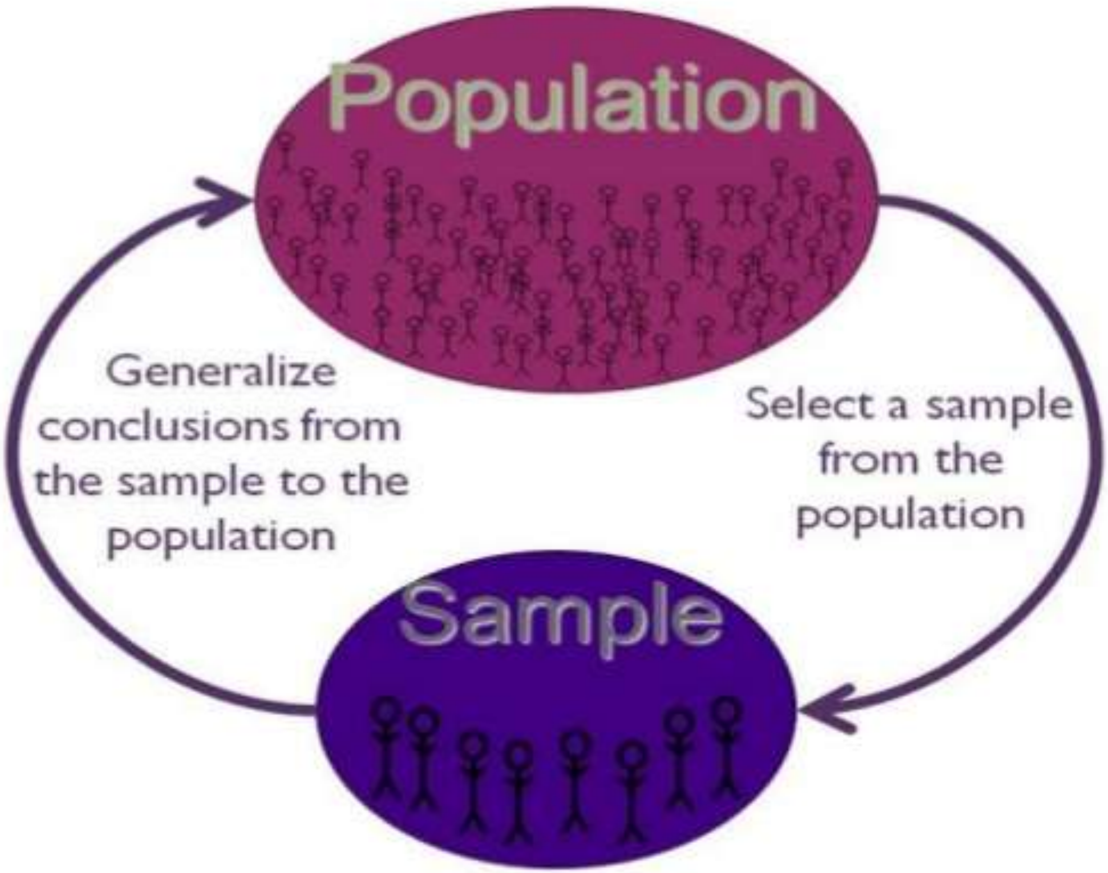
MATHEMATICAL STATISTICS

- **studies** data obtaining in a particular experiment,
- **works up** them,
- **gets** statistical parameters (or numerical characteristics),
- **defines** a type of a distribution law for data.

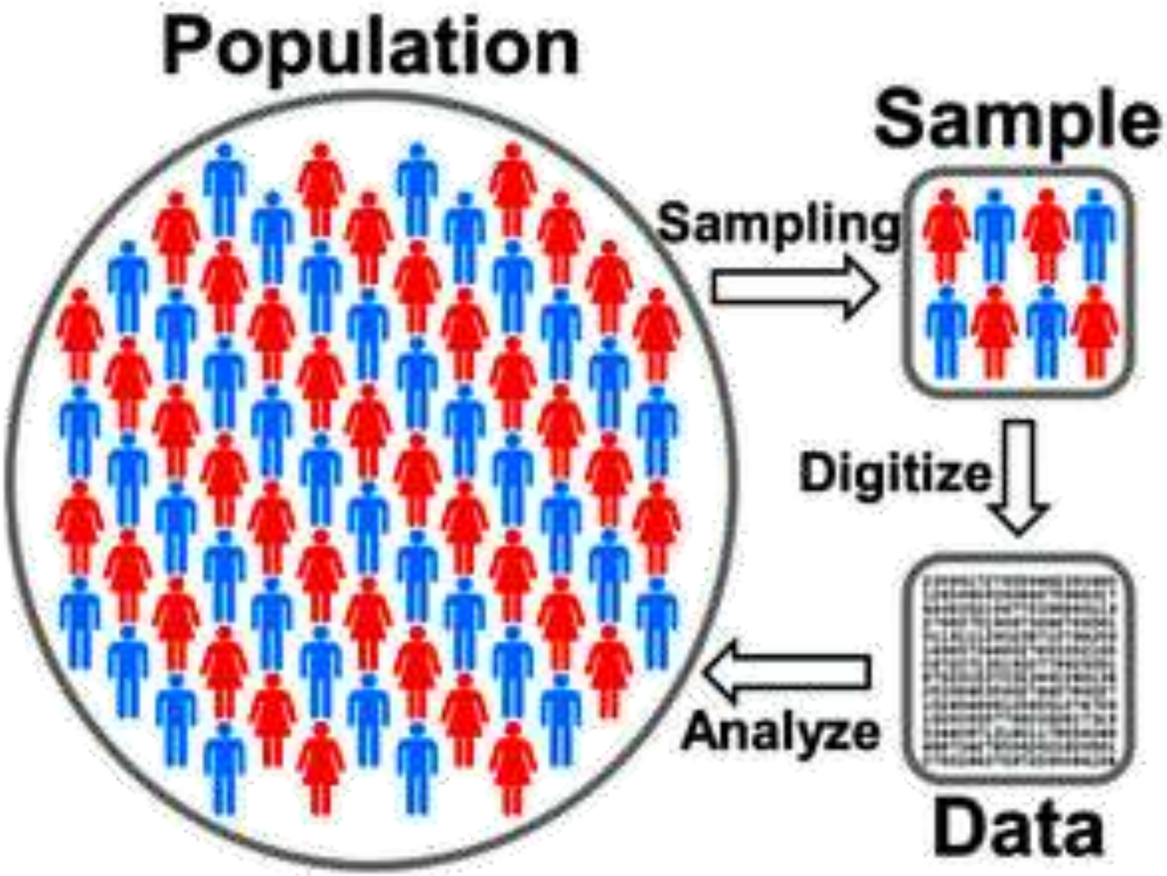
Example



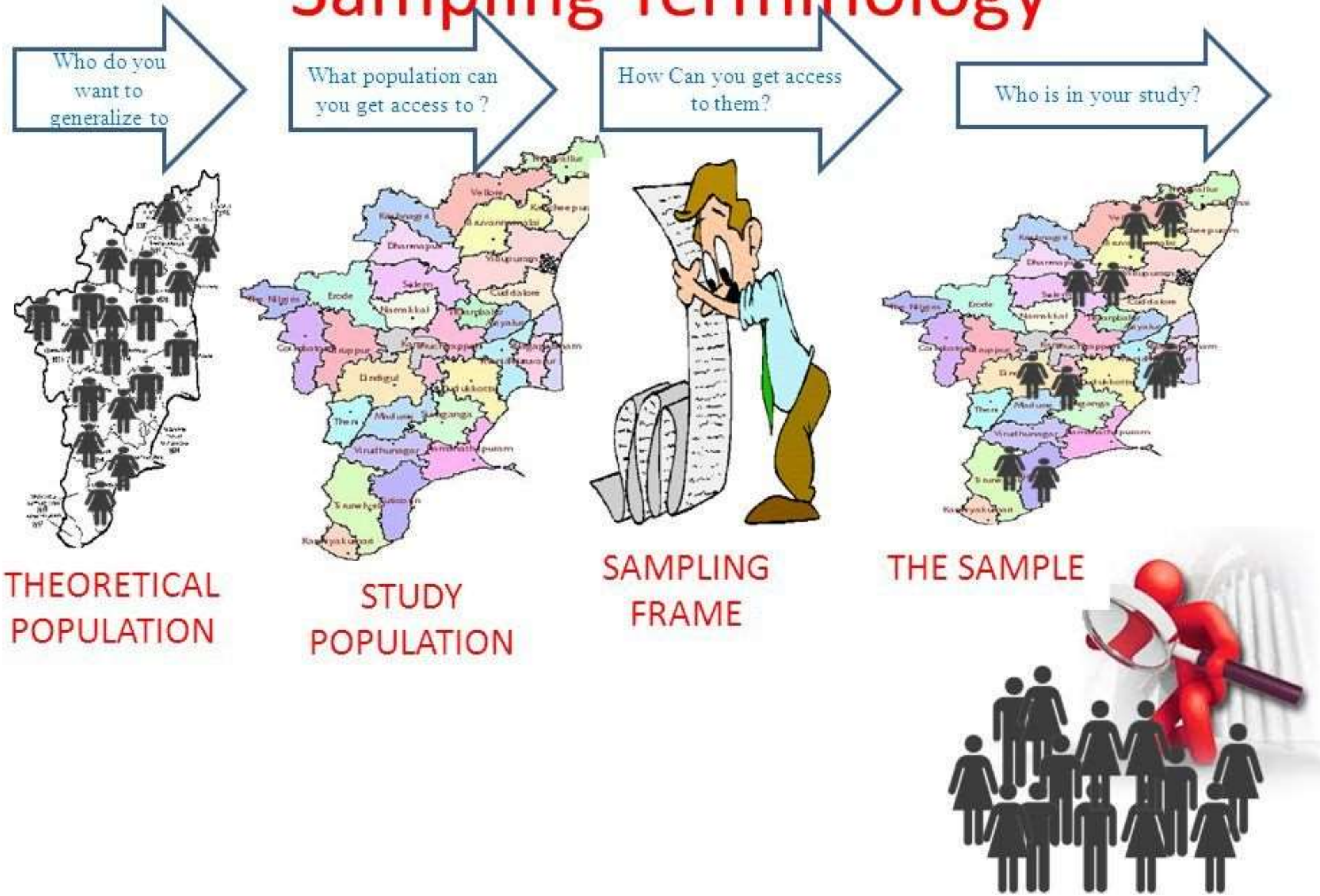
Example


Population
Generalize conclusions from the sample to the population
Select a sample from the population
Sample

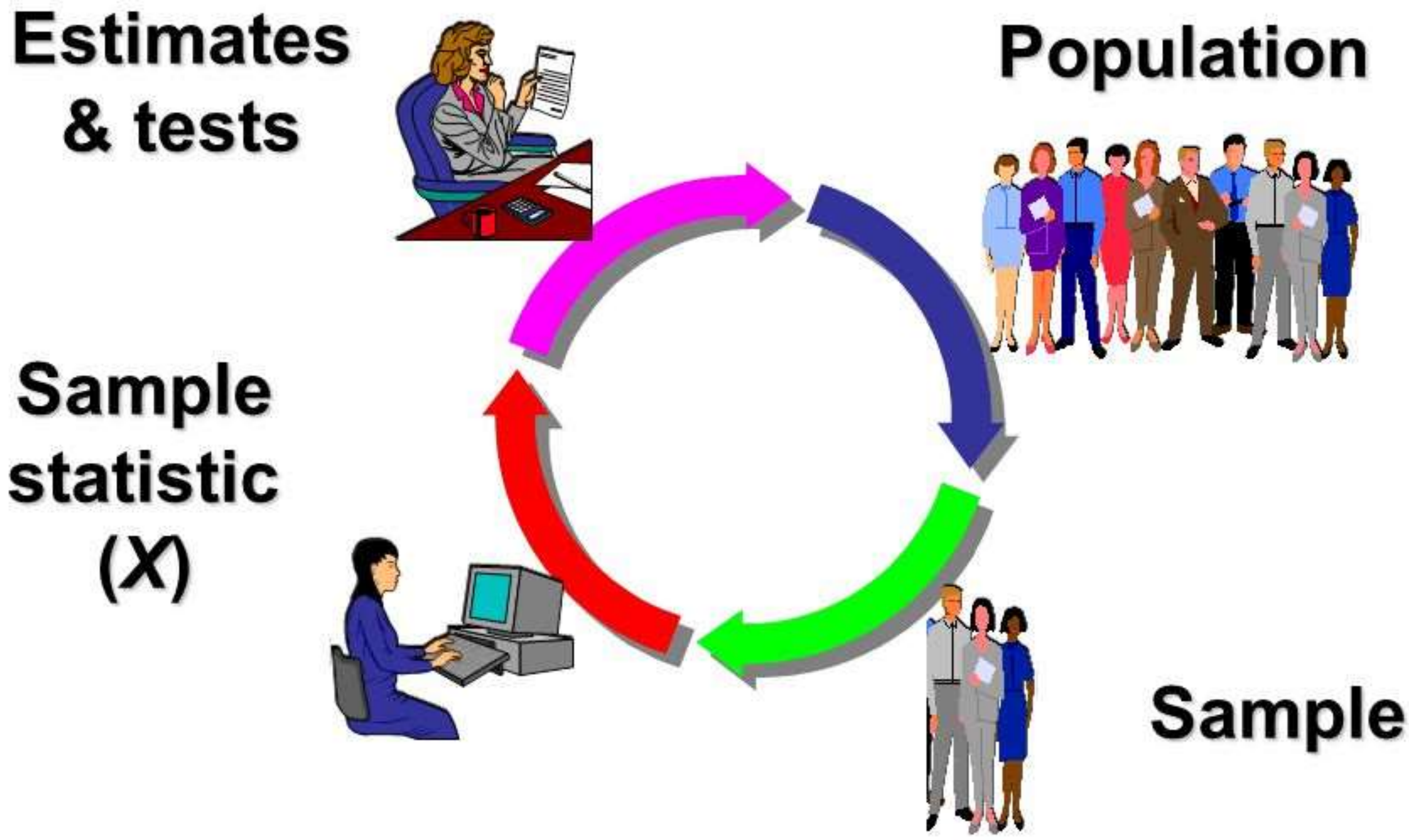
Example



Sampling Terminology



Inference Process



If $n < 30$ (a small value), we construct a **discrete statistical series** from the given sample (in this case we take separate values).

If ***n*** is a large value ($n \geq 30$), we construct **a continuous statistical series** (in this case we make intervals for values).

STATISTICAL SERIES

1. Discrete statistical series ($n < 30$)

2. Continuous statistical series ($n \geq 30$)

Discrete statistical series

DISCRETE STATISTICAL SERIES

($n < 30$)

Let we have a sample of the population

$x = (x_1, x_2, x_3, ..., x_n)$.

For example, 10, 5, 15, 7, 20.

DISCRETE STATISTIC SERIES

For example, 10, 5, 15, 7, 20.

If we write this sample in the form of increasing sequence or decreasing sequence, then we obtain ***a discrete statistic series***.

DISCRETE STATISTIC SERIES

For example, 10, 5, 15, 7, 20.

An increasing sequence is a sequence of elements from the smallest to the greatest.

An decreasing sequence is a sequence of elements from the greatest to the smallest.

DISCRETE STATISTIC SERIES

For example, **10, 5, 15, 7, 20**.

An increasing sequence is a sequence of elements from the smallest to the greatest.

For example, **5, 7, 10, 15, 20 (increasing sequence)**.

DISCRETE STATISTICAL SERIES

For example, **10, 5, 15, 7, 20**.

An decreasing sequence is a sequence of elements from the greatest to the smallest.

For example, **20, 15, 10, 7, 5 (decreasing sequence)**.

DISCRETE STATISTICAL SERIES

If some values of a random variable **are repeated**, then a statistic series is written in the form of the table, where we write the values of a random variable x_i and their frequencies m_i .

A frequency m_i is called a number of occurrences of particular values of a random variable.

DISCRETE STATISTICAL SERIES

If some values of a random variable **are repeated**, then a statistic series is written in the form of the table, where we write the values of a random variable x_i and their frequencies m_i .

x_i	x_1	x_2	$\cdots$	x_k
m_i	m_1	m_2	$\cdots$	m_k

DISCRETE STATISTICAL SERIES

$$n = \sum_{i=1}^{k} m_i$$

(***n*** is sample size)

x_i	x_1	x_2	$\cdots$	x_k
m_i	m_1	m_2	$\cdots$	m_k

DISCRETE STATISTICAL SERIES

For example, we have a sample 10,15,20,10, 5,7,10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

DISCRETE STATISTICAL SERIES ($n < 30$)

For example, we have a sample 10,15,20,10, 5,7,10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

$$x_1 = 5$$

DISCRETE STATISTICAL SERIES

For example, we have a sample 10,15,20,10,5,7, 10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

$$x_1 = 5$$

$$m_1 = 2$$

DISCRETE STATISTICAL SERIES

For example, we have a sample 10,15,20,10,5,7, 10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

$x_1 = 5$ $x_2 = 7$

$m_1 = 2$ m_2

DISCRETE STATISTICAL SERIES

For example, we have a sample 10,15,20,10,5,7, 10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

$x_1 = 5$ $x_2 = 7$

$m_1 = 2$ $m_2 = 4$

DISCRETE STATISTICAL SERIES

For example, we have a sample 10,15,20,10,5,7, 10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

$x_1 = 5$ $x_2 = 7$ $x_3 = 10$

$m_1 = 2$ $m_2 = 4$ m_3

DISCRETE STATISTICAL SERIES

For example, we have a sample 10,15,20,10,5,7, 10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

$x_1 = 5$ $x_2 = 7$ $x_3 = 10$

$m_1 = 2$ $m_2 = 4$ $m_3 = 7$

DISCRETE STATISTICAL SERIES

For example, we have a sample 10,15,20,10,5,7, 10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

$x_1 = 5$ $x_2 = 7$ $x_3 = 10$ $x_4 = 15$

$m_1 = 2$ $m_2 = 4$ $m_3 = 7$

DISCRETE STATISTICAL SERIES

For example, we have a sample 10,15,20,10,5,7, 10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

$x_1 = 5$ $x_2 = 7$ $x_3 = 10$ $x_4 = 15$

$m_1 = 2$ $m_2 = 4$ $m_3 = 7$ $m_4 = 5$

DISCRETE STATISTICAL SERIES

For example, we have a sample 10,15,20,10,5,7, 10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

$x_1 = 5$ $x_2 = 7$ $x_3 = 10$ $x_4 = 15$

$m_1 = 2$ $m_2 = 4$ $m_3 = 7$ $m_4 = 5$

$x_5 = 20$

DISCRETE STATISTICAL SERIES

For example, we have a sample 10,15,20,10,5,7, 10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

$x_1 = 5$ $x_2 = 7$ $x_3 = 10$ $x_4 = 15$

$m_1 = 2$ $m_2 = 4$ $m_3 = 7$ $m_4 = 5$

$x_5 = 20$ $m_5 = 2$

DISCRETE STATISTICAL SERIES

For example, we have a sample 10,15,20,10,5,7, 10,10,15,5,7,10,15,15,7,20,15,10,10,7.

Task. Make a statistic series of this sample.

x_i	**5**	**7**	**10**	**15**	**20**
m_i	2	4	7	5	2

DISCRETE STATISTICAL SERIES

Task. Make a statistic series of this sample.

x_i	**5**	**7**	**10**	**15**	**20**
m_i	2	4	7	5	2

The sample size is $n=\sum_{i=1}^{5} m_i = 2+4+7+5+2=20$

The ratio $w_i = \frac{m_i}{n}$, where the ratio of the frequency m_i and the sample size n is called ***relative frequency*** and

$$\sum_{i=1}^{k} w_i = 1$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	**20**
w_i						

$$n = \sum_{i=1}^{5} m_i = 2 + 4 + 7 + 5 + 2 = 20$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i						

$$w_1 = \frac{m_1}{n} = \frac{2}{20} = 0{,}1$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1					

$$w_1 = \frac{m_1}{n} = \frac{2}{20} = 0{,}1$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1					

$$w_2 = \frac{m_2}{n} = \frac{4}{20} = 0,2$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2				

$$w_2 = \frac{m_2}{n} = \frac{4}{20} = 0,2$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2				

$$w_3 = \frac{m_3}{n} = \frac{7}{20} = 0{,}35$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35			

$$w_4 = \frac{m_4}{n} = \frac{5}{20} = 0,25$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25		

$$w_5 = \frac{m_5}{n} = \frac{2}{20} = 0{,}1$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	

$$w_5 = \frac{m_5}{n} = \frac{2}{20} = 0{,}1$$

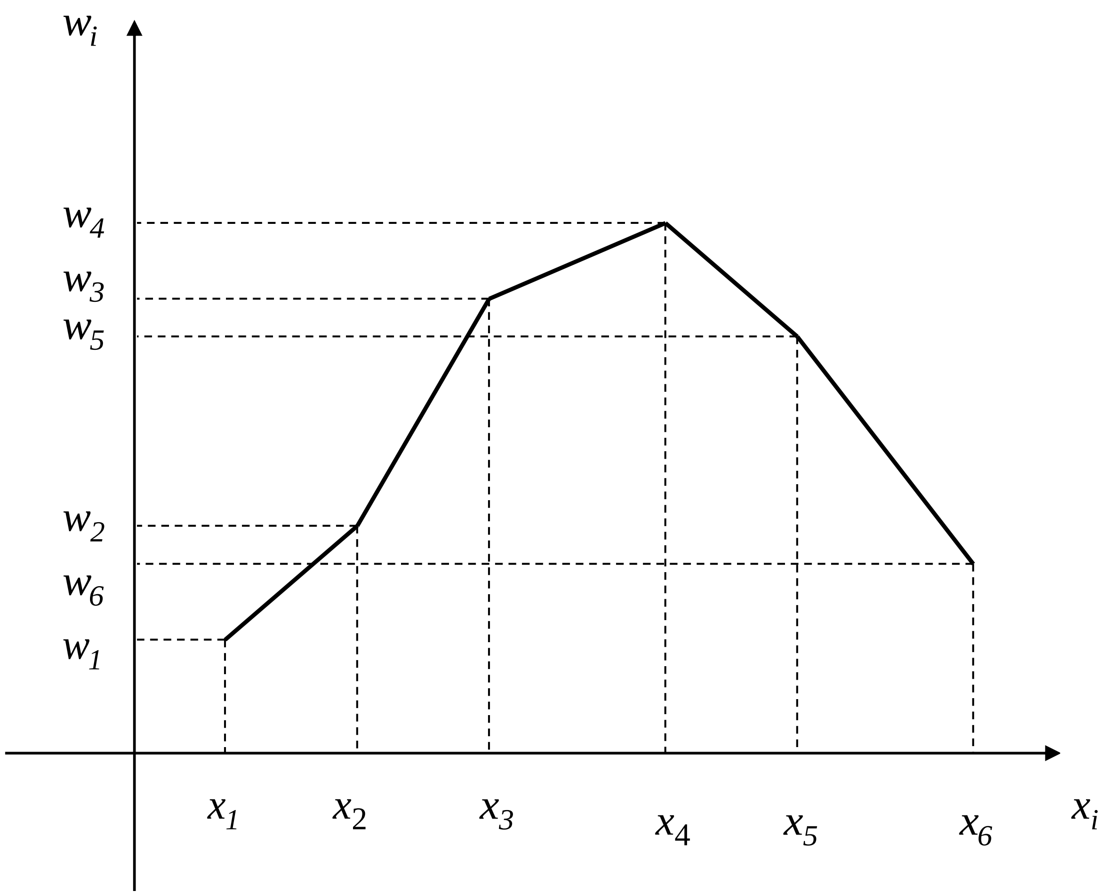
EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**

DISCRETE STATISTICAL SERIES

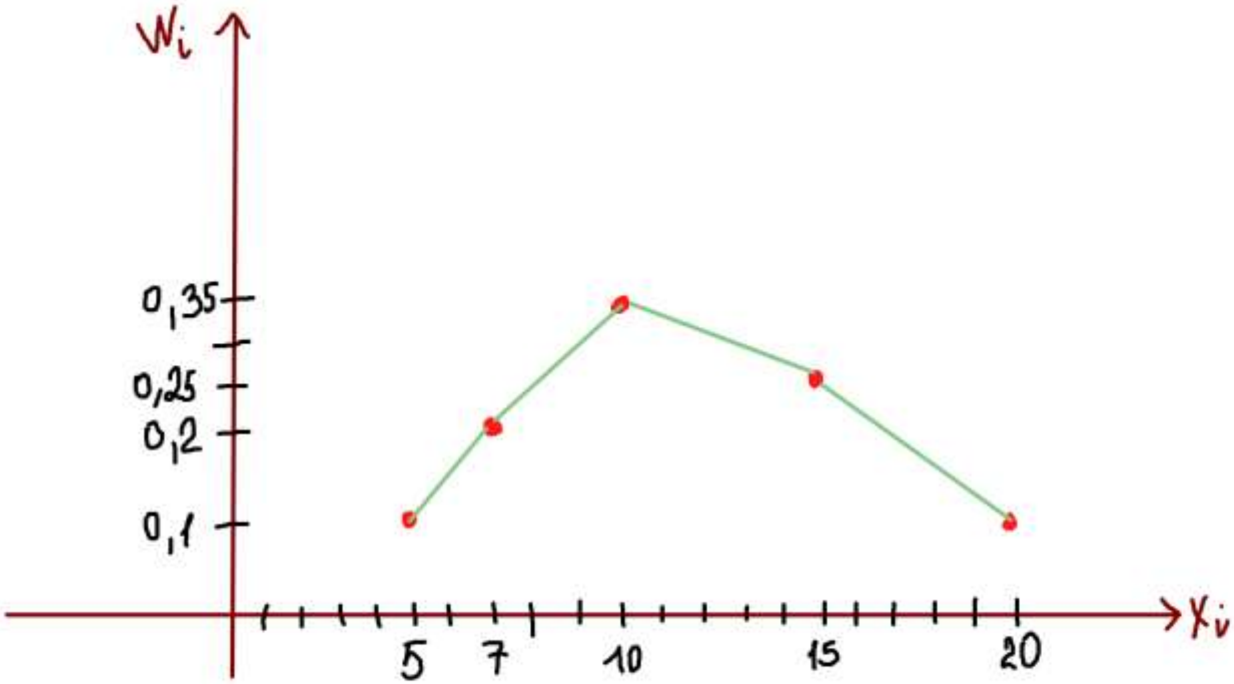
If the values x_i and w_i are represented as the coordinates of the points, then we get a graphical representation which is called ***a distribution polygon.***

A distribution polygon



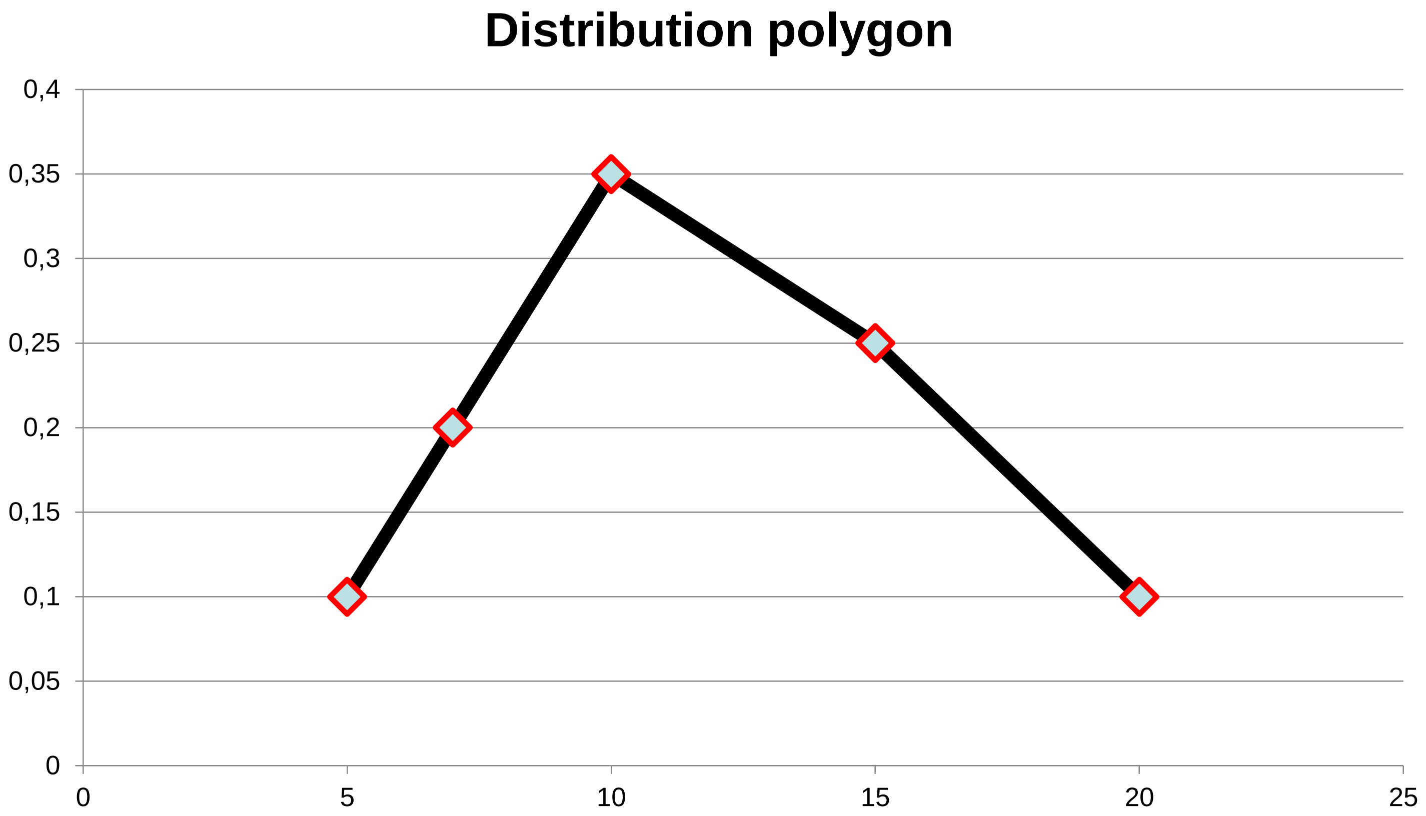
The distribution polygon

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**



The distribution polygon

w_i



x_i

Empirical distribution function

Adding the relative frequencies, we get the *cumulative frequency*

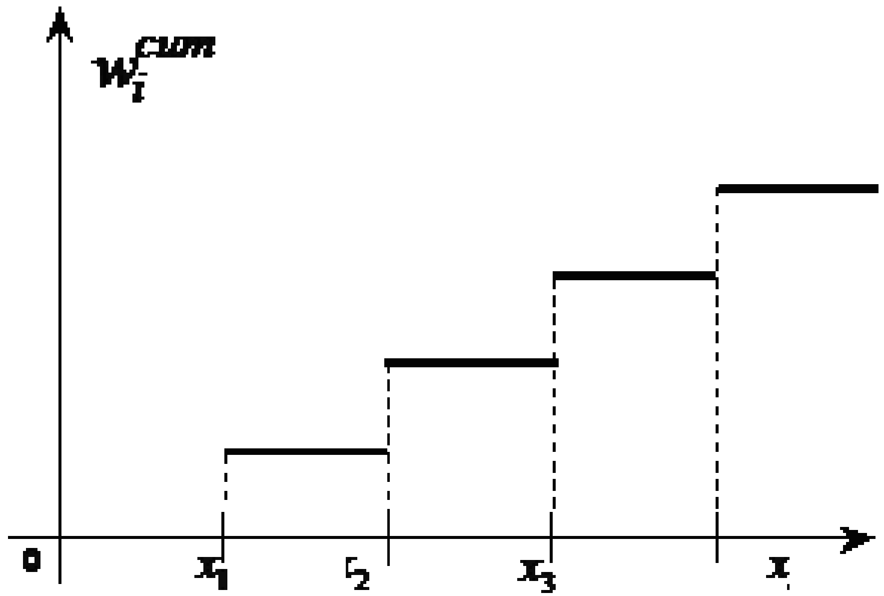
$$F_i = w_1 + w_2 + \dots + w_i$$

$(i = 1,2, \dots, k)$

A cumulative function

It is denoted and calculated as

$$F^{cum}\left(X < x_i\right) = \frac{m_i^{cum}}{n} = w_i^{cum}$$



EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i						

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0						

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1					

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3				

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65			

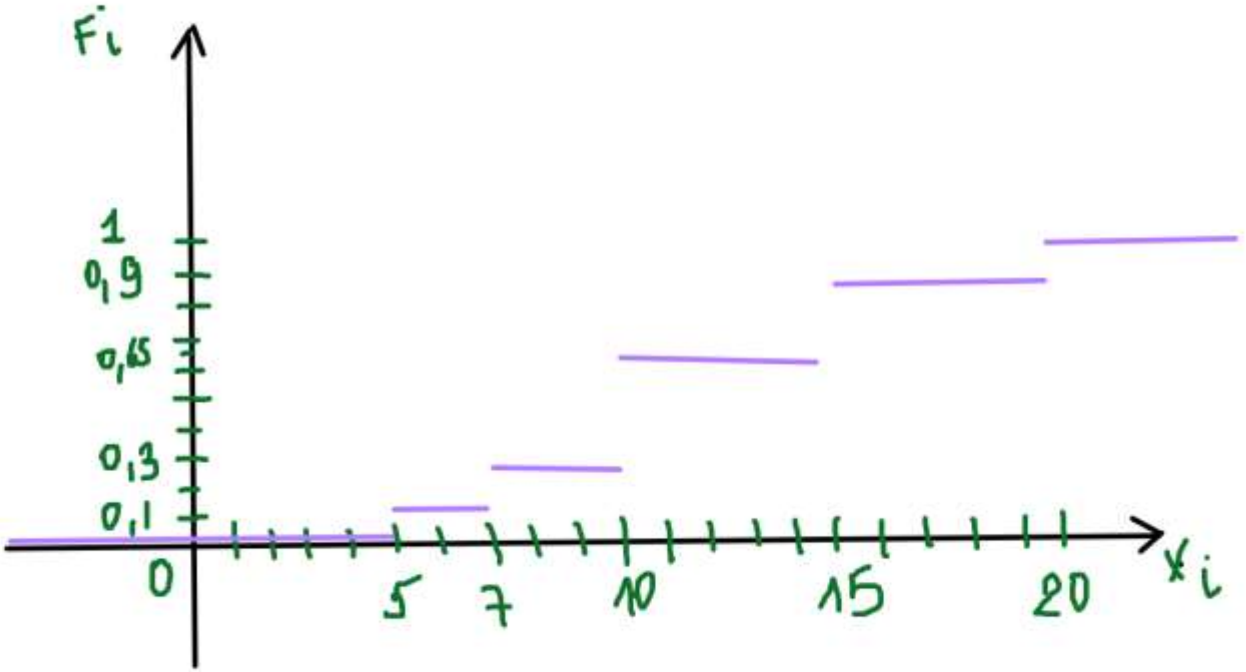
EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9		

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	

The cumulative function



STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

After summarizing and analyzing the data of the sample, we can calculate the following parameters (the numerical characteristics):

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

1. *The sample mean* of a discrete statistical series is

$$\overline{x_s} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i$$

or

$$\overline{x_s} = \sum_{i=1}^{k} x_i w_i$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$						

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5					

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4				

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5			

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75		

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**

$$\overline{x_s} = 11{,}15$$

79

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15** $= \overline{x_s}$

80

DISCRETE STATISTICAL SERIES ($n < 30$)

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

2. ***The sample variance*** of a discrete statistical series is

$$S_x^2 = \sum_{i=1}^{k} \left(x_i - \overline{x_s}\right)^2 \cdot w_i$$

or

$$S_x^2 = \sum_{i=1}^{k} (x_i)^2 \cdot w_i - \left(\overline{x_s}\right)^2 = \overline{x_s^2} - \left(\overline{x_s}\right)^2$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$= \overline{x_s}$
$x_i^2 w_i$							

Мисюра Е.Ю.)
82

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15** $= \overline{x_s}$
$x_i^2 w_i$	2,5					

83

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15** $= \overline{x_s}$
$x_i^{\,2} w_i$	2,5	9,8				

84

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15** $= \overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35			

Місюра Є.Ю.)
85

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15** $=\overline{x}_s$
$x_i{}^2 w_i$	2,5	9,8	35	56,25		

86

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$= \overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40		

87

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15** $= \overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**

88

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**	$=\overline{x_s^2}$

89

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15** $=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55** $=\overline{x_s^2}$

$$S_x^2 = \overline{x_s^2} - \left(\overline{x_s}\right)^2 =$$

90

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**	$=\overline{x_s^2}$

$$S_x^2 = \overline{x_s^2} - \left(\overline{x_s}\right)^2 = 143{,}55 - 11{,}15^2 =$$

91

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**	$=\overline{x_s^2}$

$$S_x^2 = 143{,}55 - 11{,}15^2 = 19{,}2275$$

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

3. ***The sample root-mean-square deviation*** of a discrete statistical series is

$$S_x = \sqrt{S_x^2}$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**	$=\overline{x_s^2}$

$$S_x^2 = 143{,}55 - 11{,}15^2 = 19{,}2275$$

$$S_x = \sqrt{S_x^2} = \sqrt{19{,}2275} = 4{,}4$$

94

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

3*. ***The corrected root-mean-square deviation*** of a discrete statistical series is

$$S_{cor} = \sqrt{\frac{n}{n-1} \cdot S_x^2}$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**	$=\overline{x_s^2}$

$$S_x = \sqrt{S_x^2} = \sqrt{19,2275} = 4,4$$

$$S_{cor} = \sqrt{\frac{n}{n-1} \cdot S_x^2} = \sqrt{\frac{20}{20-1} \cdot 19,2275} =$$

96

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**	$=\overline{x_s^2}$

$$S_x = \sqrt{S_x^2} = \sqrt{19,2275} = 4,4$$

$$S_{cor} = \sqrt{\frac{n}{n-1} \cdot S_x^2} = \sqrt{\frac{20}{20-1} \cdot 19,2275} = 4,5$$

STATISTICAL PARAMETERS (DSS)

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

4. ***The median*** Me may not be a uniquely determined point. The median of a sample is

$$M_e = \begin{cases} x_{m+1}, & \text{if } k = 2m+1 \\ \dfrac{x_{m+1} + x_m}{2}, & \text{if } k = 2m \end{cases}$$

STATISTICAL PARAMETERS (DSS)

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

For example, 1, 2, 3, 4, 5, 7, 8, 9, 10. $Me = ?$

For example, 2, 3, 4, 5, 7, 8, 9, 10. $M_e = ?$

STATISTICAL PARAMETERS (DSS)

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

For example, 1, 2, 3, 4, 5, 7, 8, 9, 10. $Me = 5$

For example, 2, 3, 4, 5, 7, 8, 9, 10. $M_e = \frac{5+7}{2} = 6$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15** $=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55** $=\overline{x_s^2}$

$$M_e = ?$$

$$M_e = \begin{cases} x_{m+1}, & \text{if } k = 2m+1 \\ \dfrac{x_{m+1} + x_m}{2}, & \text{if } k = 2m \end{cases}$$

Місюра Є.Ю.)
101

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15** $= \overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55** $= \overline{x_s^2}$

$$M_e = 10$$

5 5 7 7 7 7 10 10 10 10 | 10 10 10 15 15 15 15 15 20 20

$$Me = \frac{10+10}{2} = 10$$

n=20 even

102

STATISTICAL PARAMETERS
(NUMERICAL CHARACTERISTICS)
OF A SAMPLE

5. ***The mode*** is the data value that occurs with greatest frequency. It is denoted by M_o, i.e.

$$M_o = x_i \quad \left(\underset{1 \le i \le k}{m_i} = \max \right)$$

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

For example, 2,2,5,7,9,9,9,10,10,11,12,18.

$$M_o = ?$$

$$M_o = x_i \quad \left(\underset{1 \le i \le k}{m_i} = \max \right)$$

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

For example, 2,2,5,7,9,9,9,10,10,11,12,18.

This sample has one mode!

$$M_o = 9$$

$$M_o = x_i \quad \left(\underset{1 \le i \le k}{m_i} = \max \right)$$

STATISTICAL PARAMETERS
(NUMERICAL CHARACTERISTICS)
OF A SAMPLE

For example, 3,5,8,10,12,16.

$$M_o = ?$$

$$M_o = x_i \quad \left(\underset{1\le i\le k}{m_i} = \max \right)$$

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

For example, 3,5,8,10,12,16.

This sample has no mode!

$$M_o = x_i \quad \left(\underset{1 \le i \le k}{m_i} = \max \right)$$

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

For example, 2,3,4,4,4,5,5,7,7,7,9.

$$M_o = ?$$

$$M_o = x_i \quad \left(\underset{1 \le i \le k}{m_i} = \max \right)$$

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

For example, 2,3,4,4,4,5,5,7,7,7,9.

This sample has two modes!

$$M_{o(1)} = 4 \qquad M_{o(2)} = 7$$

$$M_o = x_i \quad \left(\max_{1 \le i \le k} m_i \right)$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**
F_i 0	0,1	0,3	0,65	0,9	1	
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15** $=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55** $=\overline{x_s^2}$

$$M_o = ?$$

$$M_o = x_i \quad \left(\underset{1 \le i \le k}{m_i} = \max \right)$$

110

EXAMPLE

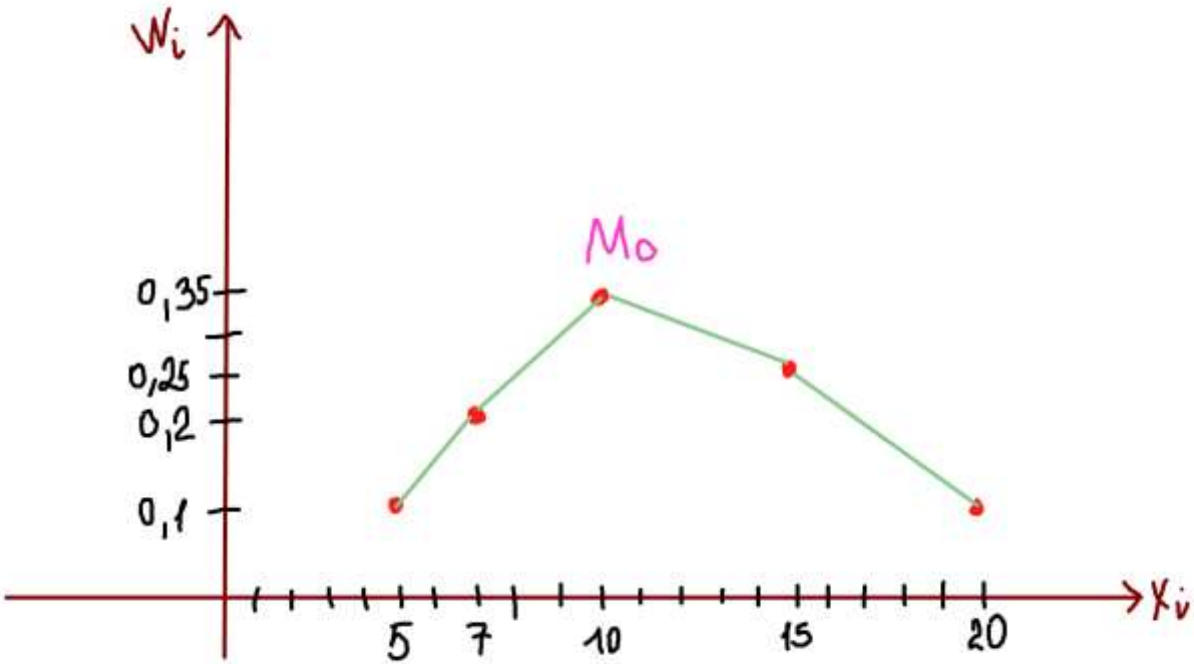
x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**	$=\overline{x_s^2}$

$$M_o = 10$$

111

The distribution polygon and the mode Mo

x_i	**5**	**7**	**10**	**15**	**20**	Sum
m_i	2	4	7	5	2	20
w_i	0,1	0,2	0,35	0,25	0,1	**1**



STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

6. Range

$$R = x_{\max} - x_{\min}$$

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**	$=\overline{x_s^2}$

$$R = x_{\max} - x_{\min} =$$

114

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**	$=\overline{x_s^2}$

$$R = x_{\max} - x_{\min} = 20 -$$

115

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**	$=\overline{x_s^2}$

$$R = x_{\max} - x_{\min} = 20 - 5 =$$

116

EXAMPLE

x_i	**5**	**7**	**10**	**15**	**20**	Sum	
m_i	2	4	7	5	2	20	
w_i	0,1	0,2	0,35	0,25	0,1	**1**	
F_i 0	0,1	0,3	0,65	0,9	1		
$x_i w_i$	0,5	1,4	3,5	3,75	2	**11,15**	$=\overline{x_s}$
$x_i^2 w_i$	2,5	9,8	35	56,25	40	**143,55**	$=\overline{x_s^2}$

$$R = x_{\max} - x_{\min} = 20 - 5 = 15$$

117

Continuous statistical series

CONTINUOUS STATISTICAL SERIES

It is not suitable to use a discrete statistic series if ***n*** is very large.

Let's consider an interval, which contains the *n* data x_i (*i* = 1,2,..., *n*) of the sample, and divide it into *k* subintervals, so-called ***classes*** or ***class intervals***.

CONTINUOUS STATISTICAL SERIES

A number of subintervals *k* is defined by ***Sturges' formula***:

$$\log_2 n + 1 = 1 + 3.322 \lg n$$

We round off ***k*** to ***an integer***.

CONTINUOUS STATISTICAL SERIES

A number of subintervals k is defined by ***Sturges' formula***:

$$k = 1 + 3.322 \cdot \lg n$$

We round off ***k*** to ***an integer***.

CONTINUOUS STATISTICAL SERIES ($n \geq 30$)

For example, if ***n*=100**, then the number of subintervals *k* is

$$k = 1 + 3,322 \cdot \lg 100 = 1 + 3,322 \cdot 2 = 7,644$$

CONTINUOUS STATISTICAL SERIES

For example, if ***n*=100**, then the number of subintervals *k* is

$$k = 1 + 3{,}322 \cdot \lg 100 = 1 + 3{,}322 \cdot 2 = 7{,}644$$

We round off ***k*** to ***the integer*** and get ***k*=8**.

CONTINUOUS STATISTICAL SERIES

Usually 5-10 classes are selected with equal length *h*, and their boundaries are called ***class boundaries.***

$$h = \frac{R}{k} = \frac{x_{\max} - x_{\min}}{k}$$

CONTINUOUS STATISTICAL SERIES

The endpoints of the total interval are not uniquely defined; in general, we choose them approximately symmetrically with respect to the smallest and largest value of the sample, and class boundaries should be different from any sample value.

CONTINUOUS STATISTICAL SERIES

$[x_i; x_{i+1})$	$[x_1; x_2)$	$[x_2; x_3)$	...	$[x_i; x_{i+1})$	...	$[x_{k-1}; x_k)$
m_i	m_1	m_2	...	m_i	...	m_k
			...		...	

CONTINUOUS STATISTICAL SERIES

$[x_i; x_{i+1})$	$[x_1; x_2)$	$[x_1; x_2)$	...	$[x_i; x_{i+1})$	...	$[x_{k-1}; x_k)$
m_i	m_1	m_2	...	m_i	...	m_k
			...		...	

$$\sum_{i=1}^{k} m_i = n$$

CONTINUOUS STATISTICAL SERIES

$[x_i; x_{i+1})$	$[x_1; x_2)$	$[x_1; x_2)$	...	$[x_i; x_{i+1})$	...	$[x_{k-1}; x_k)$
m_i	m_1	m_2	...	m_i	...	m_k
x_i'	x_1'	x_2'	...	x_i'	...	x_k'

CONTINUOUS STATISTICAL SERIES

$[x_i; x_{i+1})$	$[x_1; x_2)$	$[x_1; x_2)$	...	$[x_i; x_{i+1})$	...	$[x_{k-1}; x_k)$
m_i	m_1	m_2	...	m_i	...	m_k
x'_i	x'_1	x'_2	...	x'_i	...	x'_k

$$x'_i = \frac{x_i + x_{i+1}}{2}$$

$$h = x_{i+1} - x_i$$

CONTINUOUS STATISTICAL SERIES

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	

$$n = \sum_{i=1}^{5} m_i = 20 + 50 + 60 + 40 + 30 = 200$$

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	200

$$n = \sum_{i=1}^{5} m_i = 20 + 50 + 60 + 40 + 30 = 200$$

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	200
x_i'						

$$x_i' = \frac{x_i + x_{i+1}}{2}$$

EXAMPLE (CSS)

$[x_i;x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	200
x_i'	150					

$$x_1' = \frac{100+200}{2} = \frac{300}{2} = 150$$

134

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	200
x_i'	150					

$$x_1' = \frac{100+200}{2} = \frac{300}{2} = 150$$

$$h = 200 - 100 = 100$$

135

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	200
x_i'	150					

$$x_1' = \frac{100+200}{2} = \frac{300}{2} = 150$$

$$h = 300 - 200 = 100$$

136

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	200
x'_i	150	250				

$$x'_2 = \frac{200 + 300}{2} = \frac{500}{2} = 250$$

137

EXAMPLE (CSS)

$[x_i ; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	200
x_i'	150	250	350			

$$x_3' = \frac{300+400}{2} = \frac{700}{2} = 350$$

Мисюра Е.Ю.)
138

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	200
x_i'	150	250	350	450		

$$x_4' = \frac{400+500}{2} = \frac{900}{2} = 450$$

139

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	200
x'_i	150	250	350	450	550	

$$x'_5 = \frac{500 + 600}{2} = \frac{1100}{2} = 550$$

140

The ratio $w_i = \frac{m_i}{n}$, where the ratio of the frequency m_i and the sample size n is called ***relative frequency*** and

$$\sum_{i=1}^{k} w_i = 1$$

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	200
x_i'	150	250	350	450	550	
w_i						

$$w_1 = \frac{20}{200} = 0{,}1$$

Мисюра Е.Ю.)
142

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	200
x_i'	150	250	350	450	550	
w_i	0,1					

$$w_2 = \frac{50}{200} = 0,25$$

143

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	200
x_i'	150	250	350	450	550	
w_i	0,1	0,25				

$$w_3 = \frac{60}{200} = 0,3$$

Мисюра Е.Ю.)
144

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	200
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3			

$$w_4 = \frac{40}{200} = 0,2$$

145

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	200
x'_i	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2		

$$w_5 = \frac{30}{200} = 0{,}15$$

146

EXAMPLE (CSS)

$[x_i ; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	200
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	

$$w_5 = \frac{30}{200} = 0{,}15$$

Мисюра Е.Ю.)
147

EXAMPLE (CSS)

$[x_i ; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	200
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**

148

Empirical distribution function

Adding the relative frequencies, we get the *cumulative frequency*

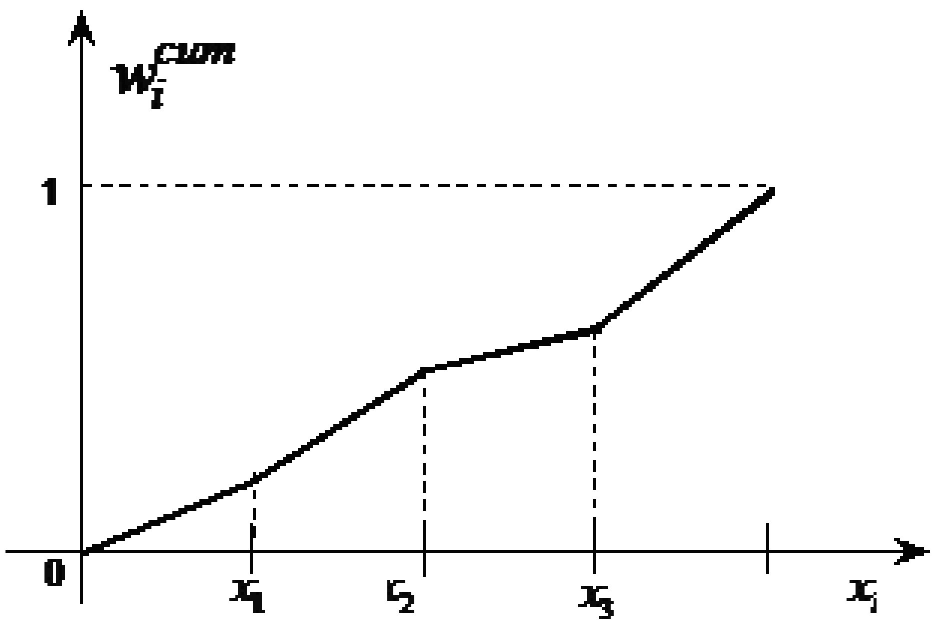
$$F_i = w_1 + w_2 + ... + w_i$$

$(i = 1,2, ...,k)$

A cumulative function

It is denoted and calculated as

$$F^{cum}\left(X < x_i\right) = \frac{m_i^{cum}}{n} = w_i^{cum}$$



EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i						

151

EXAMPLE (CSS)

$[x_i;x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0						

152

EXAMPLE (CSS)

$[x_i ; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1					

153

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35				

154

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65			

155

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x'_i	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85		

156

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	

157

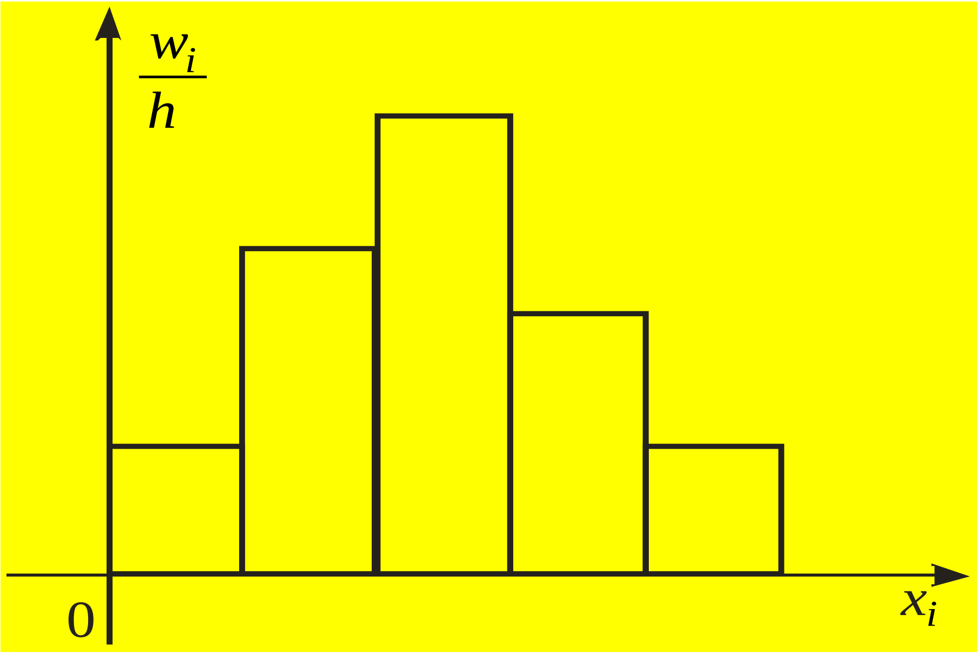
CONTINUOUS STATISTICAL SERIES

A graphical representation of a continuous statistical series is called a *histogram of relative frequencies*.

Here $\frac{w_i}{h}$ is the height of the corresponding rectangle on the interval $\left[x_i, x_{i+1}\right)$.

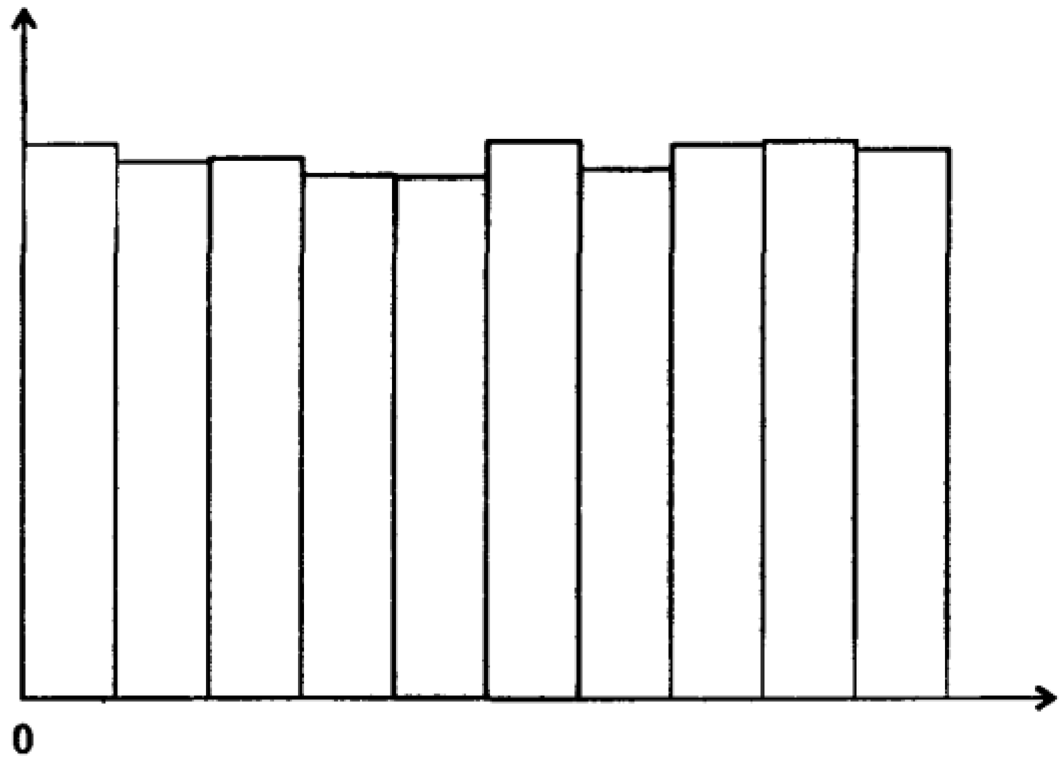
CONTINUOUS STATISTICAL SERIES ($n \geq 30$)

A graphical representation of a continuous statistical series is called a *histogram of relative frequencies*.



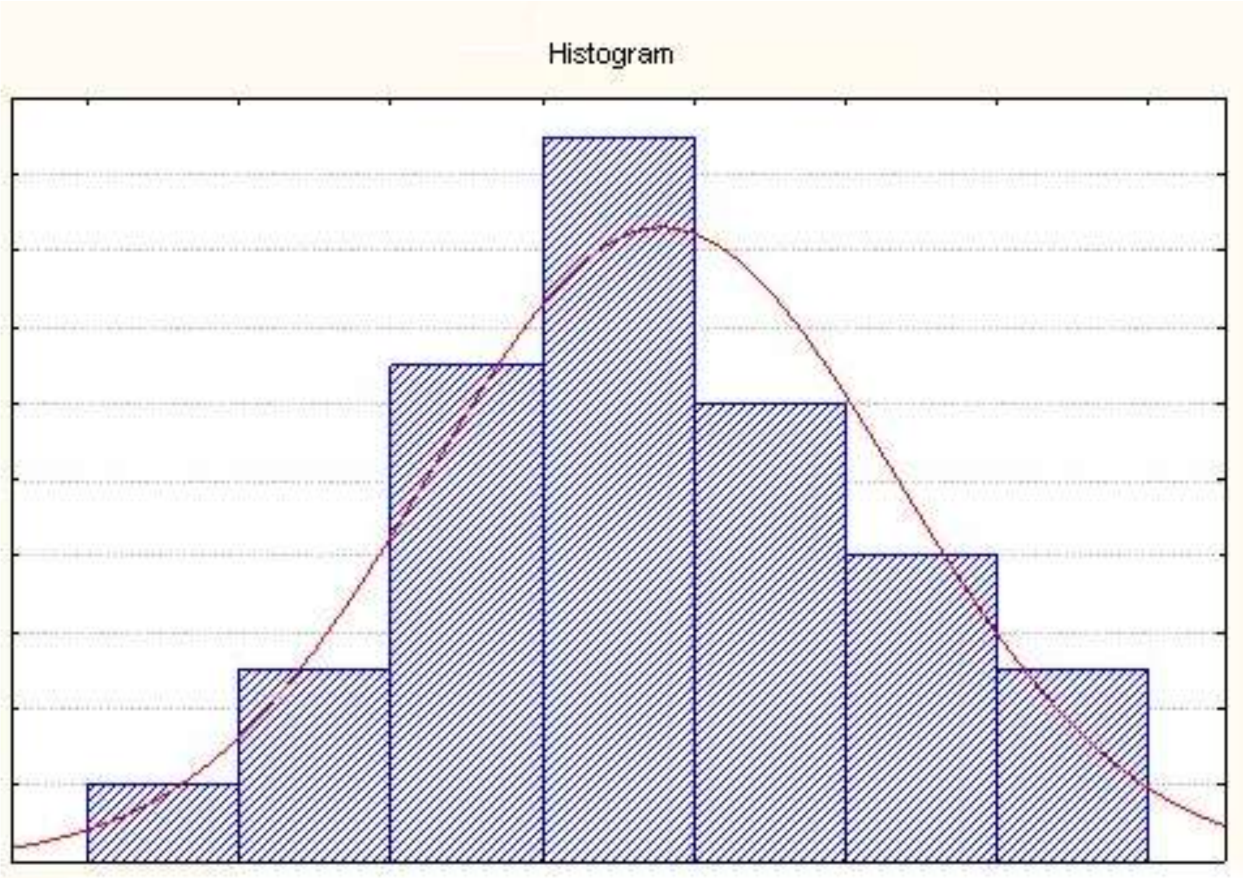
THE ASSUMPTION ABOUT THE DISTRIBUTION LAW (THE HYPOTHESIS)

- Let’s consider the basic distribution laws:
- a) the uniform distribution



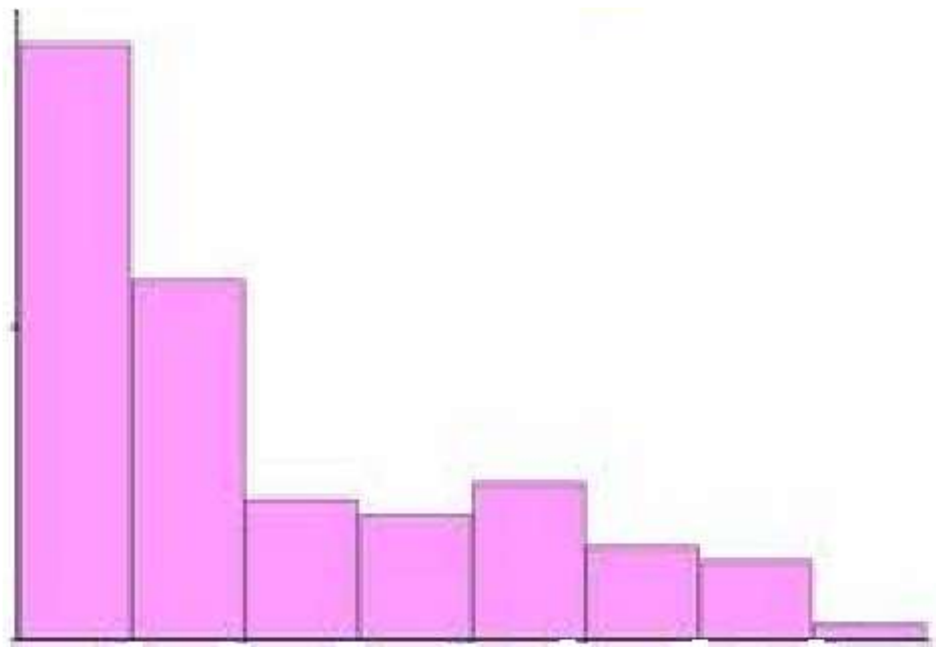
THE ASSUMPTION ABOUT THE DISTRIBUTION LAW (THE HYPOTHESIS)

- Let’s consider the basic distribution laws:
- b) the normal distribution



THE ASSUMPTION ABOUT THE DISTRIBUTION LAW (THE HYPOTHESIS)

- Let’s consider the basic distribution laws:
- c) the exponential distribution



EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	
w_i / h						

$$\frac{w_1}{h} = \frac{0,1}{200-100} = \frac{0,1}{100} = 0,001$$

163

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x'_i	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	
w_i / h	0,001					

$$\frac{w_2}{h} = \frac{0,25}{100} = 0,0025$$

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x'_i	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	
w_i / h	0,001	0,0025				

$$\frac{w_3}{h} = \frac{0,3}{100} = 0,003$$

165

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	
w_i / h	0,001	0,0025	0,003			

$$\frac{w_4}{h} = \frac{0,2}{100} = 0,002$$

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x'_i	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	
w_i / h	0,001	0,0025	0,003	0,002		

$$\frac{w_5}{h} = \frac{0{,}15}{100} = 0{,}0015$$

EXAMPLE (CSS)

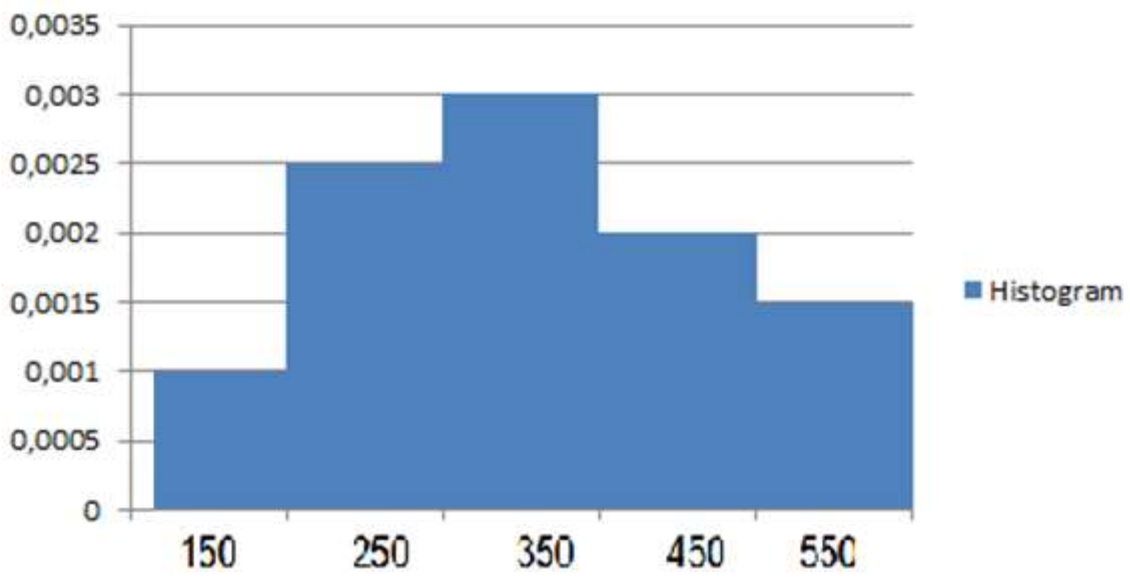
$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	
w_i / h	0,001	0,0025	0,003	0,002	0,0015	

Мисюра Е.Ю.)
168

EXAMPLE (CSS)

x_i'	150	250	350	450	550	
w_i / h	0,001	0,0025	0,003	0,002	0,0015	

w_i / h



x_i'

169

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

1. *The sample mean* of a continuous statistical series is

$$\overline{x_s} = \frac{1}{n}\sum_{i=1}^{k} x_i' \cdot m_i = \sum_{i=1}^{k} x_i' \cdot w_i$$

$$x_i' = \frac{x_i + x_{i+1}}{2}$$

EXAMPLE (CSS)

$[x_i;x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$						

171

EXAMPLE (CSS)

$[x_i ; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15					

$$x_1' \cdot w_1 = 150 \cdot 0{,}1 = 15$$

172

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15					

$$x_2' \cdot w_2 = 250 \cdot 0{,}25 = 62{,}5$$

173

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5				

$$x_3' \cdot w_3 = 350 \cdot 0{,}3 = 105$$

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105			

$$x_4' \cdot w_4 = 450 \cdot 0{,}2 = 90$$

Мисюра Е.Ю.)
175

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90		

$$x_5' \cdot w_5 = 550 \cdot 0{,}15 = 82{,}5$$

176

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x'_i	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x'_i \cdot w_i$	15	62,5	105	90	82,5	**355**

177

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$

$$\overline{x_s} = 355$$

178

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

2. *The sample variance* of a continuous statistical series is

$$S_x^2 = \sum_{i=1}^{k} \left(x_i' - \overline{x_s} \right)^2 \cdot w_i$$

or

$$S_x^2 = \sum_{i=1}^{k} \left(x_i' \right)^2 \cdot w_i - \left(\overline{x_s} \right)^2 = \overline{x_s^2} - \left(\overline{x_s} \right)^2$$

EXAMPLE (CSS)

$[x_i ; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x_i')^2 \cdot w_i$						

EXAMPLE (CSS)

$[x_i;x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $=\overline{x_s}$
$(x_i')^2 \cdot w_i$	2250					

$$(x_1')^2 \cdot w_1 = 150 \cdot 15 = 2250$$

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x_i')^2 \cdot w_i$	2250	15625				

$$(x_2')^2 \cdot w_2 = 250 \cdot 62{,}5 = 15625$$

182

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x_i')^2 \cdot w_i$	2250	15625	36750			

$$(x_3')^2 \cdot w_3 = 350 \cdot 105 = 36750$$

183

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x_i')^2 \cdot w_i$	2250	15625	36750	40500		

$$(x_4')^2 \cdot w_4 = 450 \cdot 90 = 40500$$

184

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x'_i	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x'_i \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x'_i)^2 \cdot w_i$	2250	15625	36750	40500	45375	

$$(x'_5)^2 \cdot w_5 = 550 \cdot 82,5 = 45375$$

185

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x_i')^2 \cdot w_i$	2250	15625	36750	40500	45375	**140500** $= \overline{x_s^2}$

$$\overline{x_s^2} = \sum_{i=1}^{5} (x_i')^2 \cdot w_i = 140500$$

186

EXAMPLE (CSS)

$[x_i ; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x_i')^2 \cdot w_i$	2250	15625	36750	40500	45375	**140500** $= \overline{x_s^2}$

$$S_x^2 = \overline{x_s^2} - \left(\overline{x_s}\right)^2 = 140500 - (355)^2 =$$

187

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x_i')^2 \cdot w_i$	2250	15625	36750	40500	45375	**140500** $= \overline{x_s^2}$

$$S_x^2 = 140500 - (355)^2 = 14475$$

188

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x_i')^2 \cdot w_i$	2250	15625	36750	40500	45375	**140500** $= \overline{x_s^2}$

$$S_x^2 = 14475$$

189

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

3. ***The sample root-mean-square deviation*** of a continuous statistical series is

$$S_x = \sqrt{S_x^2}$$

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x_i')^2 \cdot w_i$	2250	15625	36750	40500	45375	**140500** $= \overline{x_s^2}$

$$S_x^2 = 14475$$

$$S_x = \sqrt{14475} =$$

191

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x'_i	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x'_i \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x'_i)^2 \cdot w_i$	2250	15625	36750	40500	45375	**140500** $= \overline{x_s^2}$

$$S_x^2 = 14475$$

$$S_x = \sqrt{14475} = 120,31$$

192

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

3*. ***The corrected root-mean-square deviation*** of a continuous statistical series is

$$S_{cor} = \sqrt{\frac{n}{n-1} \cdot S_x^2}$$

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x_i')^2 \cdot w_i$	2250	15625	36750	40500	45375	**140500** $= \overline{x_s^2}$

$$S_x^2 = 14475$$

$$S_{cor} = \sqrt{\frac{n}{n-1} \cdot S_x^2} = \sqrt{\frac{200}{200-1} \cdot 14475} =$$

194

EXAMPLE (CSS)

$[x_i ; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
$x_i' \cdot w_i$	15	62,5	105	90	82,5	**355** $= \overline{x_s}$
$(x_i')^2 \cdot w_i$	2250	15625	36750	40500	45375	**140500** $= \overline{x_s^2}$

$$S_{cor} = \sqrt{\frac{n}{n-1} \cdot S_x^2} = \sqrt{\frac{200}{200-1} \cdot 14475} = 120,61$$

STATISTICAL PARAMETERS (CSS) (NUMERICAL CHARACTERISTICS) OF A SAMPLE

4. ***The median***

$$M_e = x_i + \frac{0.5 - F^{cum}(x_i)}{F^{cum}(x_{i+1}) - F^{cum}(x_i)} \cdot h$$

where $F^{cum}(x_{i+1}) > 0.5, F^{cum}(x_i) < 0.5$

$$h = x_{i+1} - x_i$$

EXAMPLE (CSS)

$[x_i;x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	60	40	30	**200**
x'_i	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	

Let's find the median interval

197

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x'_i	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	

Let's find the median interval. It's the interval with $F_{i+1} > 0,5$

198

EXAMPLE (CSS)

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	**60**	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	

Let's find the median interval. It's the interval with $F_{i+1} > 0,5$

$$F_{i+1} = 0,65$$

199

EXAMPLE (CSS)

$[x_i ; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x'_i	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	

Let's find the median interval. It's the interval with $F_{i+1} > 0,5$

$F_{i+1} = 0,65$ $x_i = 300$

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	200
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	1
F_i 0	0,1	0,35	0,65	0,85	1	

Let's find the lower limit of the median interval: $x_i = 300$

Мисюра Е.Ю.)
201

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	

Let's find the median interval. It's the interval with $F_{i+1} > 0{,}5$

$x_i = 300$ $F_{i+1} = 0{,}65$ $F_i = 0{,}35$

202

EXAMPLE (CSS)

$Me = ?$

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	

Let's find the median interval. It's the interval with $F_{i+1} > 0,5$

$x_i = 300$ $F_{i+1} = 0,65$ $F_i = 0,35$ $h = 200 - 100 = 100$

203

EXAMPLE (CSS)

$$Me = ?$$

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	60	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	

$x_i = 300$ $F_{i+1} = 0{,}65$ $F_i = 0{,}35$ $h = 200 - 100 = 100$

$$M_e = 300 + \frac{0.5 - 0.35}{0.65 - 0.35} \cdot 100 =$$

204

EXAMPLE (CSS)

$Me = ?$

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	**60**	40	30	**200**
x_i'	150	250	350	450	550	
w_i	0,1	0,25	0,3	0,2	0,15	**1**
F_i 0	0,1	0,35	0,65	0,85	1	

$x_i = 300$ $F_{i+1} = 0{,}65$ $F_i = 0{,}35$ $h = 200 - 100 = 100$

$$M_e = 300 + \frac{0.5 - 0.35}{0.65 - 0.35} \cdot 100 = 300 + \frac{0.15}{0.3} \cdot 100 = 300 + 50 = 350$$

$$Mo = ?$$

STATISTICAL PARAMETERS (CSS)
(NUMERICAL CHARACTERISTICS)
OF A SAMPLE

5. ***The mode*** is $$M_o = x_i + \frac{m_{M_0} - m_{M_0 -1}}{2m_{M_0} - m_{M_0 -1} - m_{M_0 +1}} \cdot h$$

where x_i is the lower limit of the modal class (interval); m_{M_0} is the frequency of the modal class; $m_{M_0 -1}$ is the frequency of the class preceding the modal interval; $m_{M_0 +1}$ is the frequency of the class following the modal interval; h is the length of the modal interval

EXAMPLE (CSS)

$$Mo = ?$$

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	**60**	40	30	**200**

Let's find the modal interval

EXAMPLE (CSS)

$$Mo = ?$$

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	**60**	40	30	**200**

The modal interval will be the interval 300-400

EXAMPLE (CSS)

$$Mo = ?$$

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	**60**	40	30	**200**

The modal interval will be the interval 300-400

$$m_{M_0} = 60$$

$$Mo = ?$$

EXAMPLE (CSS)

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	**60**	40	30	**200**

The modal interval will be the interval 300-400

$$m_{M_0} = 60$$

$$m_{M_0 - 1} = 50$$

EXAMPLE (CSS)

$$Mo = ?$$

$[x_i; x_{i+1})$	100–200	200-300	300-400	400-500	500-600	Sum
m_i	20	50	**60**	40	30	**200**

The modal interval will be the interval 300-400

$$m_{M_0} = 60$$

$$m_{M_0+1} = 40$$

$$m_{M_0-1} = 50$$

EXAMPLE (CSS)

$$Mo = ?$$

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	**60**	40	30	**200**

The modal interval will be the interval 300-400

$$m_{M_0} = 60$$

$$m_{M_0+1} = 40$$

$$m_{M_0-1} = 50$$

$$x_i = 300$$

EXAMPLE (CSS)

$$Mo = ?$$

$[x_i; x_{i+1})$	**100–200**	**200-300**	**300-400**	**400-500**	**500-600**	**Sum**
m_i	20	50	**60**	40	30	**200**

$$m_{M_0} = 60$$

$$m_{M_0+1} = 40$$

$$m_{M_0-1} = 50$$

$$x_i = 300$$

$$M_o = 300 + \frac{60-50}{2 \cdot 60 - 50 - 40} \cdot 100 = 300 + \frac{10}{30} \cdot 100 = 333$$

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

6. Range

$$R = x_{\max} - x_{\min}$$

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

7. The ***l-th initial moment*** satisfies the relation

$$\nu_l = \overline{X^l} = \frac{\sum_{i=1}^{k} x_i^l m_i}{n} = \sum_{i=1}^{k} x_i^l w_i$$

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

7. The ***l-th initial moment*** satisfies the relation

$$\nu_1 = \overline{x_s} \qquad \nu_2 = \overline{x_s^2} = \sum_{i=1}^{k} x_i^2 w_i$$

$$\nu_3 = \overline{x_s^3} = \sum_{i=1}^{k} x_i^3 w_i \qquad \nu_4 = \overline{x_s^4} = \sum_{i=1}^{k} x_i^4 w_i$$

STATISTICAL PARAMETERS
(NUMERICAL CHARACTERISTICS)
OF A SAMPLE

8. The ***l-th central moment*** satisfies the relation:

$$\mu_l = \overline{\left(X - \overline{X}\right)^l} = \frac{\sum_{i=1}^{k} \left(x_i - \overline{x}\right)^l m_i}{n} = \sum_{i=1}^{k} \left(x_i - \overline{x}\right)^l w_i$$

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

8. The ***l-th central moment*** satisfies the relation:

$$\mu_1 = 0 \qquad \mu_2 = \sum_{i=1}^{k} \left(x_i - \overline{x_s}\right)^2 w_i = S_x^2 = v_2 - v_1^2$$

$$\mu_3 = \sum_{i=1}^{k} \left(x_i - \overline{x_s}\right)^3 w_i = v_3 - 3v_2 v_1 + 2v_1^2$$

$$\mu_4 = \sum_{i=1}^{k} \left(x_i - \overline{x_s}\right)^4 w_i = v_4 - 4v_3 v_1 + 6v_2 v_1^2 - 3v_1^4$$

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

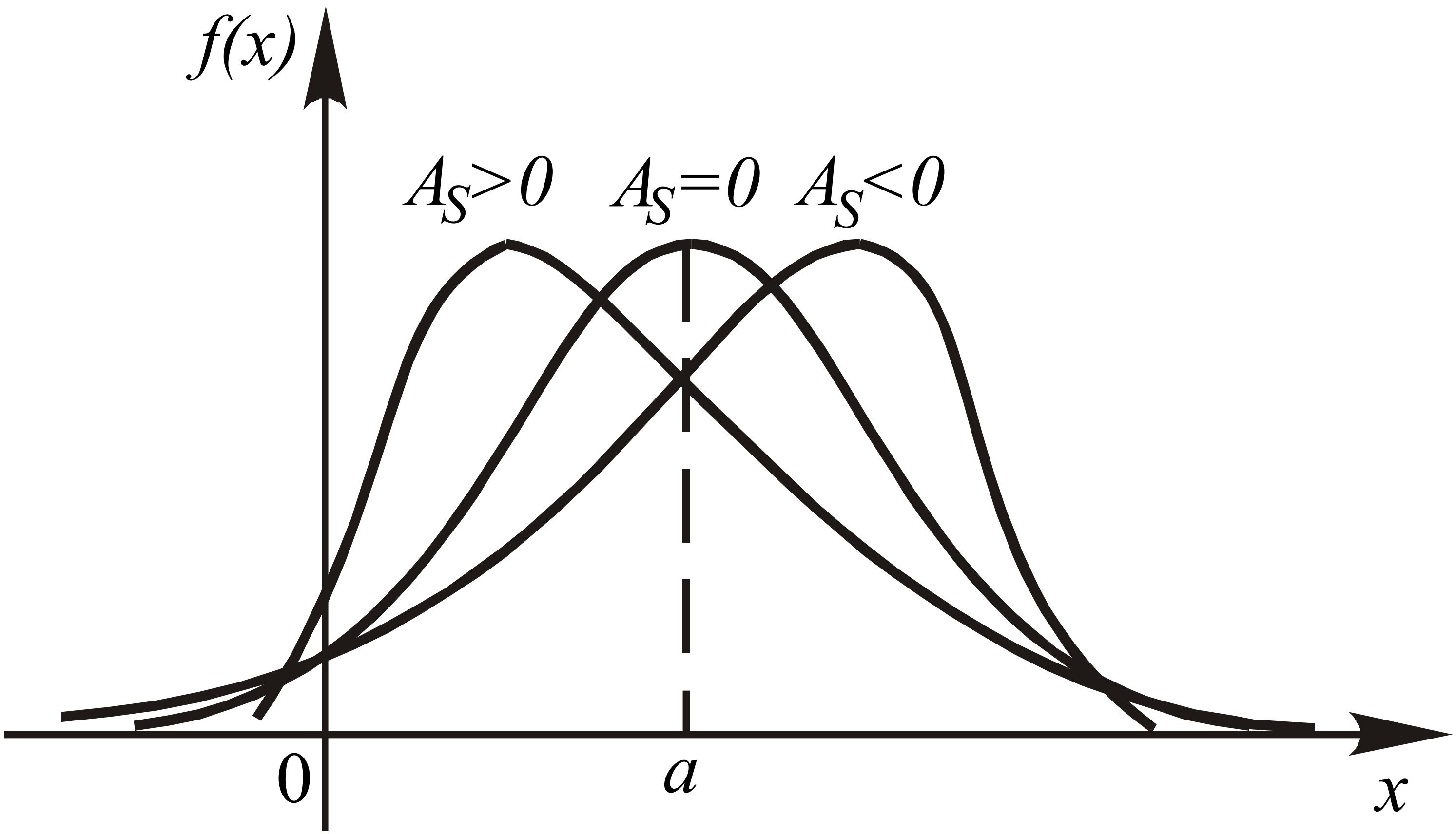
9. ***Asymmetry*** shows a deviation of a value from its central position on the left or on the right:

$$A_S = \frac{\mu_3}{S_x^3}$$

where

μ_3 – the central moment of the 3-rd order;

S_x – the root-mean square deviation.


f(x)
$A_S>0$
$A_S=0$
$A_S<0$
0
a
x

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

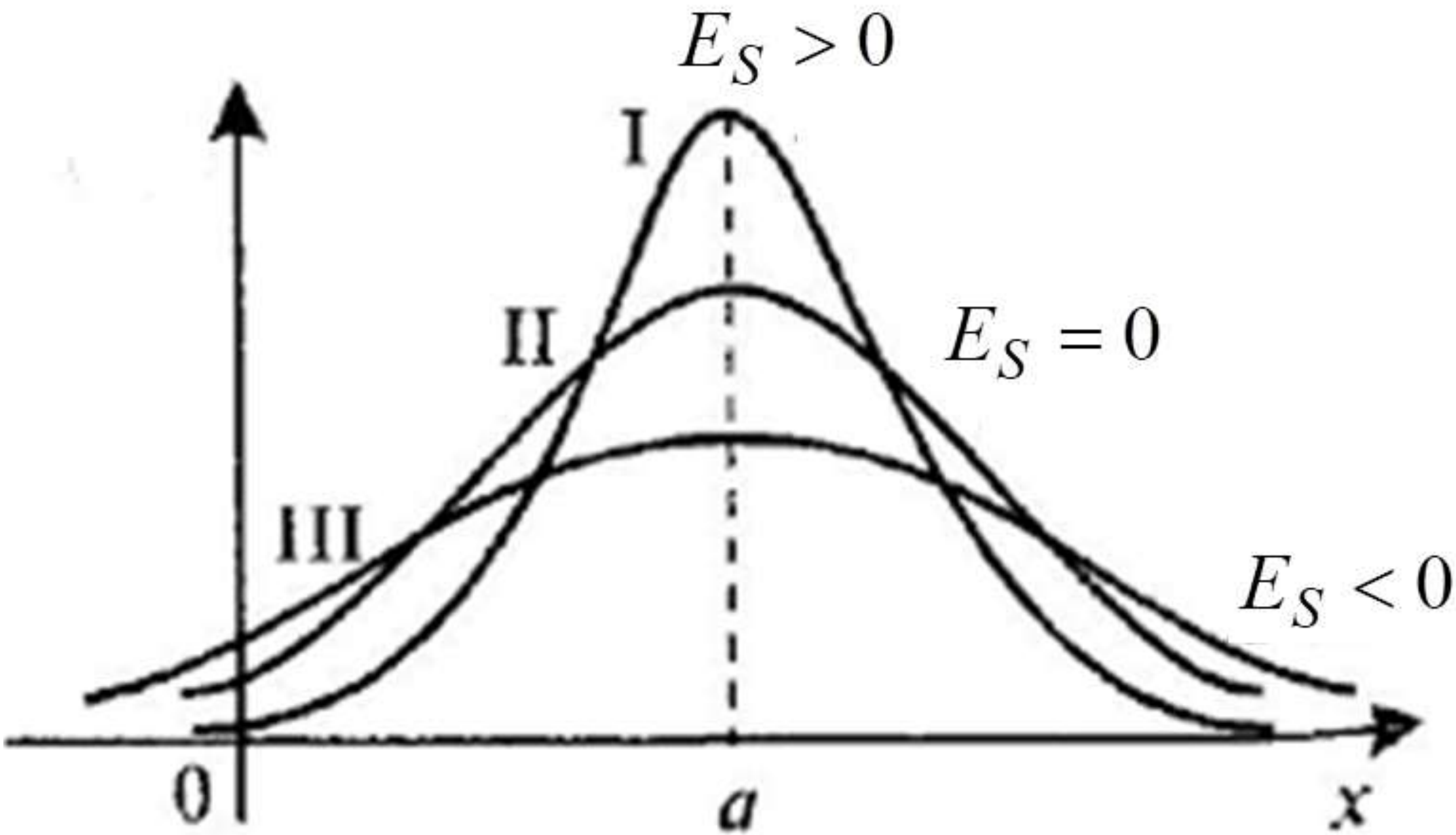
10. ***Excess*** characterizes a deviation of a value from its central position down or up:

$$E_S = \frac{\mu_4}{S_x^4} - 3$$

where

μ_4 – the central moment of the 4-th order;

S_x – the root-mean square deviation.



$E_S > 0$
I
II
$E_S = 0$
III
$E_S < 0$
0
a
x

11. ***Coefficient of variation*** is a ratio of a root-mean-square deviation and a mathematical expectation in percent:

$$\nu(X) = \frac{S_x}{\overline{\overline{x_s}}} \cdot 100\,\%$$

Coefficient of variation gives a possibility to compare a level of dispersion of values which have different character.

HOMEWORK

$[x_i ; x_{i+1})$	**10–15**	**15-20**	**20-25**	**25-30**	**30-35**
m_i	5	15	20	35	25

224

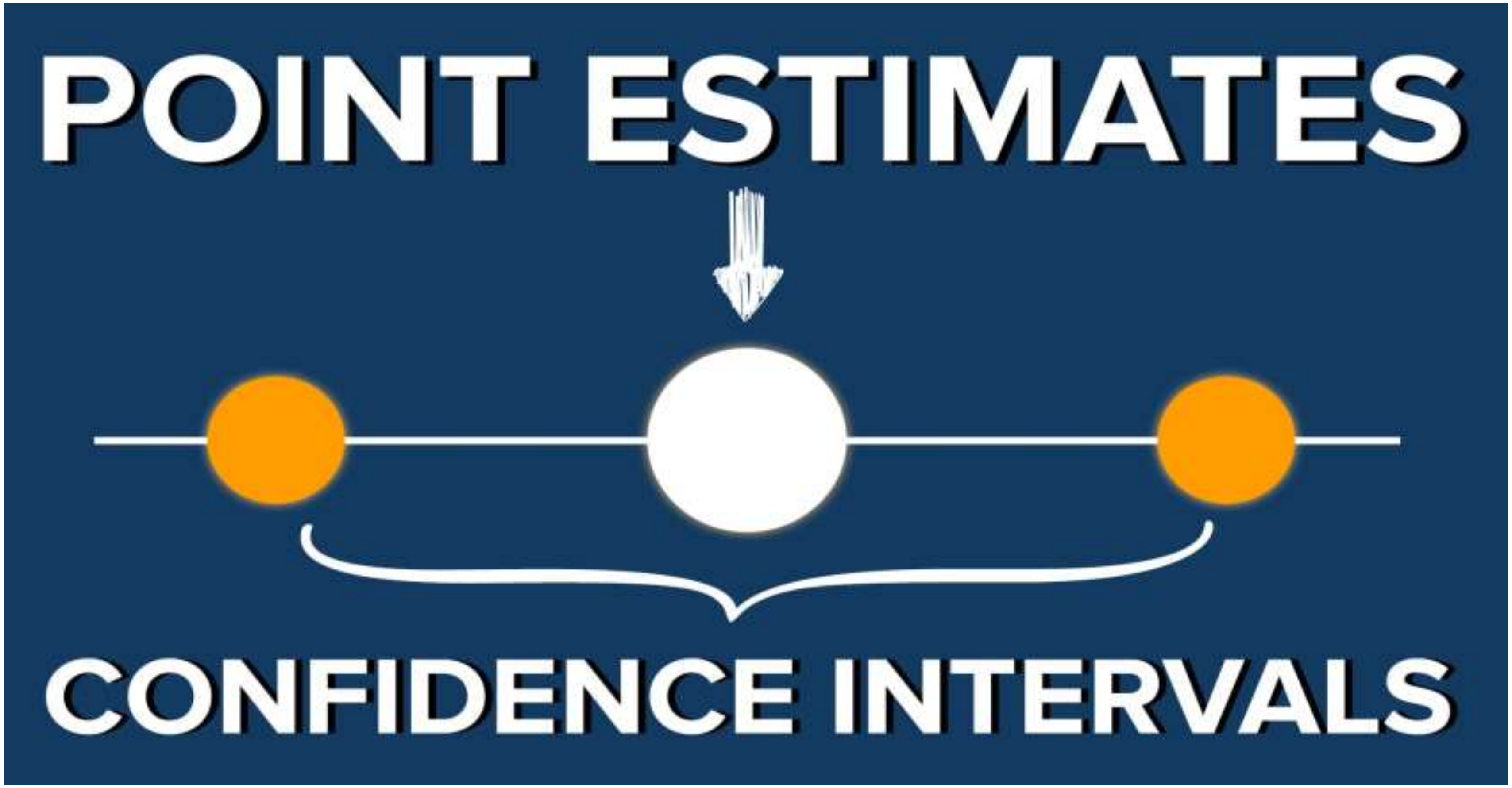
PART 2
Statistical estimations of the distribution parameters

ESTIMATIONS

Point and Interval Estimates

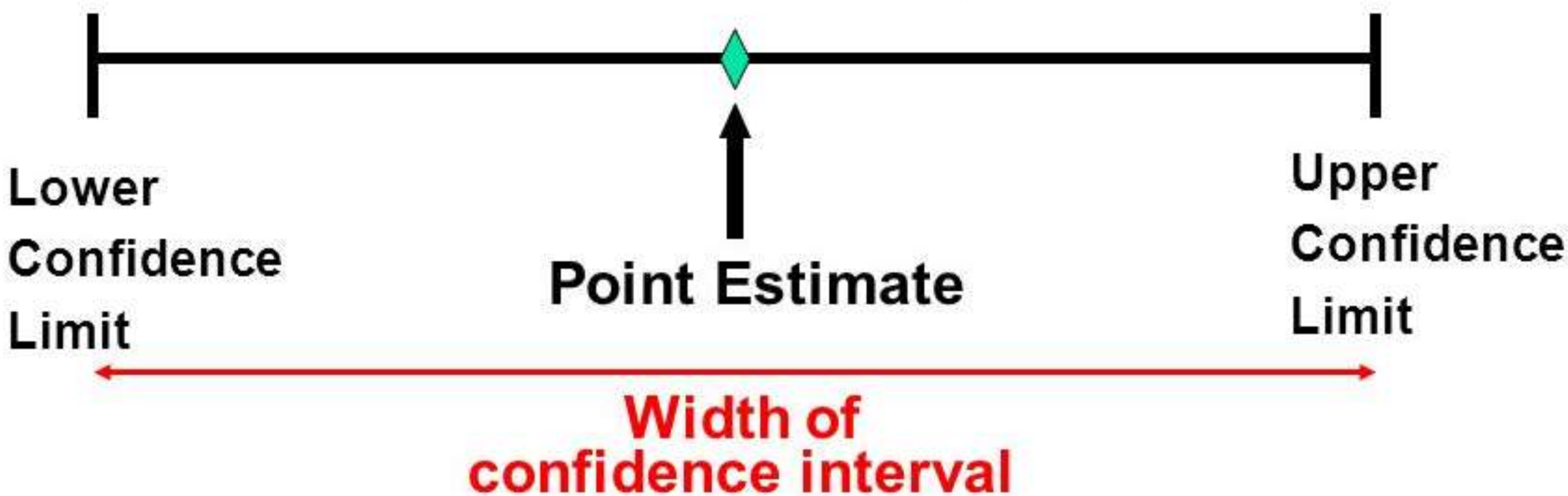
- A point estimate is a single value (statistic) used to estimate a population value (parameter).
- A confidence interval is a range of values within which the population parameter is expected to occur.

ESTIMATIONS



Point and Interval Estimates

- A point estimate is a single number,
- a confidence interval provides additional information about variability



ESTIMATIONS

Point estimate VS Interval estimate

Point estimate

$$\bar{x} \approx \hat{\mu}$$

$$s^2 \approx \hat{\sigma}^2$$

Interval estimate

$$\bar{x} - Error \leq \mu \leq \bar{x} + Error$$

Until now we didn't specify what is meant by error

INTERVAL ESTIMATIONS

If the numerical characteristics of a sample are known, then we can define the interval characteristics for a population.

A confidence interval is an interval of values bounded by confidence limits within which the true value of a population parameter is stated to lie with a specified probability.

INTERVAL ESTIMATIONS

A confidence interval is an interval of values bounded by confidence limits within which the true value of a population parameter is stated to lie with a specified probability.

INTERVAL ESTIMATIONS

If the numerical characteristics of a sample are known, then we can define the interval characteristics for a population.

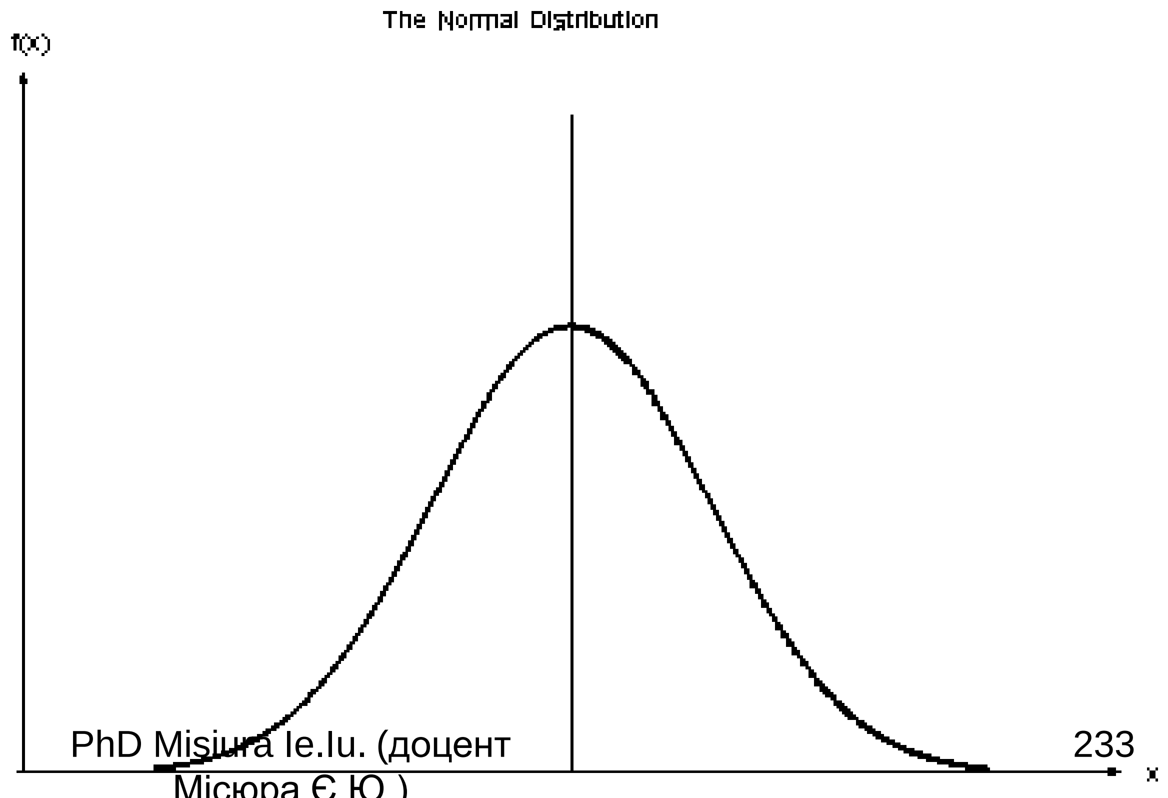
1. The <u>confidence interval</u> of the <u>population mean</u>

X has a normal distribution with parameters a and σ^2.

INTERVAL ESTIMATIONS

1. The confidence interval of the population mean

X has a normal distribution with parameters a and σ^2.



INTERVAL ESTIMATIONS

1. The confidence interval of the population mean

$$P\left(\left| X - a \right| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right)$$

INTERVAL ESTIMATIONS

1. The <u>confidence interval</u> of the <u>population mean</u>

$$P\left(\left| X - a \right| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right)$$

$$\varepsilon = \sigma t$$

INTERVAL ESTIMATIONS

1. The confidence interval of the population mean

$$P\left(\left| X - a \right| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right)$$

$$\varepsilon = \sigma t$$

$$P\left(\left| X - a \right| < \sigma t\right) = 2\Phi\left(t\right)$$

INTERVAL ESTIMATIONS

1. The <u>confidence interval</u> of the <u>population mean</u>

$$P\left(\left| X - a \right| < \sigma t\right) = 2\Phi\left(t\right)$$

$$a = \overline{X}_{pop}$$

INTERVAL ESTIMATIONS

1. The <u>confidence interval</u> of the <u>population mean</u>

$$P\left(\left| X - a \right| < \sigma t\right) = 2\Phi\left(t\right)$$

$$a = \overline{X}_{pop}$$

$$\sigma = \frac{S_{cor}}{\sqrt{n}}$$

INTERVAL ESTIMATIONS

1. The <u>confidence interval</u> of the <u>population mean</u>

$$\overline{x_s} - \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}} < \overline{X}_{pop} < \overline{x_s} + \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}}$$

Here $\overline{x_s}$ is the mean of a sample;

INTERVAL ESTIMATIONS

1. The **confidence interval** of the **population mean**

$$\overline{x_s} - \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}} < \overline{X}_{pop} < \overline{x_s} + \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}}$$

Here $\overline{x_s}$ is the mean of a sample;

$\overline{X}_{pop}$ is the mean of a population;

INTERVAL ESTIMATIONS

1. The <u>confidence interval</u> of the <u>population mean</u>

$$\overline{x_s} - \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}} < \overline{X}_{pop} < \overline{x_s} + \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}}$$

Here $\overline{x_s}$ is the mean of a sample;

$\overline{X}_{pop}$ is the mean of a population;

S_{cor} is the corrected root-mean-square deviation of a sample;

INTERVAL ESTIMATIONS

1. The **confidence interval** of the **population mean**

$$\overline{x_s} - \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}} < \overline{X}_{pop} < \overline{x_s} + \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}}$$

Here $\overline{x_s}$ is the mean of a sample;

$\overline{X}_{pop}$ is the mean of a population;

S_{cor} is the corrected root-mean-square deviation of a sample;

t_γ is a parameter that $\gamma = 2 \cdot \Phi(t_\gamma)$

INTERVAL ESTIMATIONS

Here S_{cor} is the **corrected root-mean-square deviation** of a sample.

INTERVAL ESTIMATIONS

Here S_{cor} is the **corrected root-mean-square deviation** of a sample.

We calculate the corrected root-mean-square deviation

$$S_{cor} = \sqrt{\frac{n}{n-1} \cdot S_x^2}$$

INTERVAL ESTIMATIONS

t_γ is a parameter that $\gamma = 2 \cdot \Phi(t_\gamma)$

Here γ is the **confidence probability (level)**;

$\alpha = 1 - \gamma$ is the **significance level**.

INTERVAL ESTIMATIONS

t_γ is a parameter that $\gamma = 2 \cdot \Phi(t_\gamma)$

Here γ is the **confidence probability**;

$\alpha = 1 - \gamma$ is the **confidence (significance) level**.

The **basic confidence probabilities** are **95%, 99%** and **99.9**%.

Definition

<u>Confidence Level</u> : $1 - \alpha$

The **probability** that the confidence interval actually <u>contains the population parameter</u>.

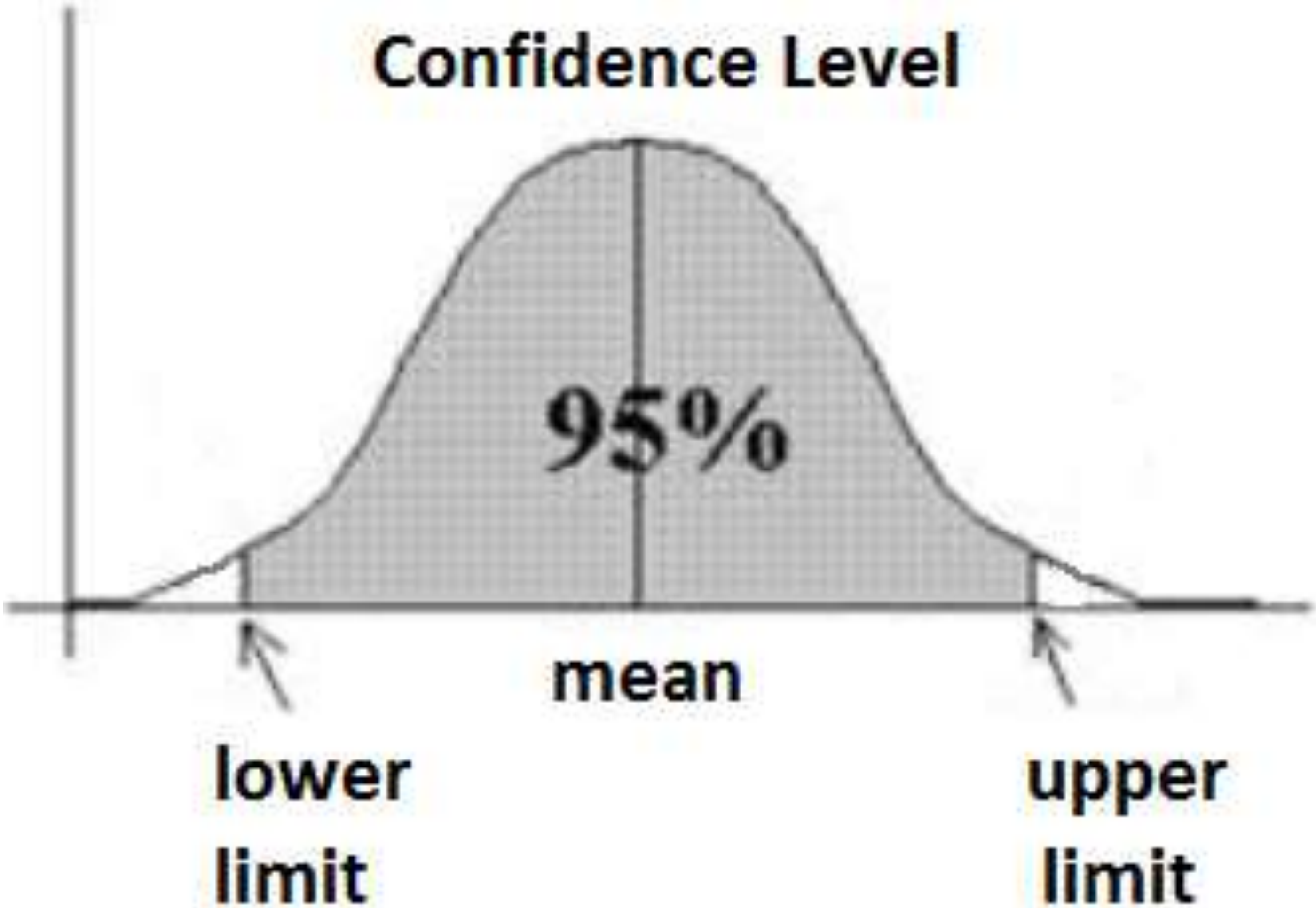
The most common confidence levels used are **95%**, **99%**

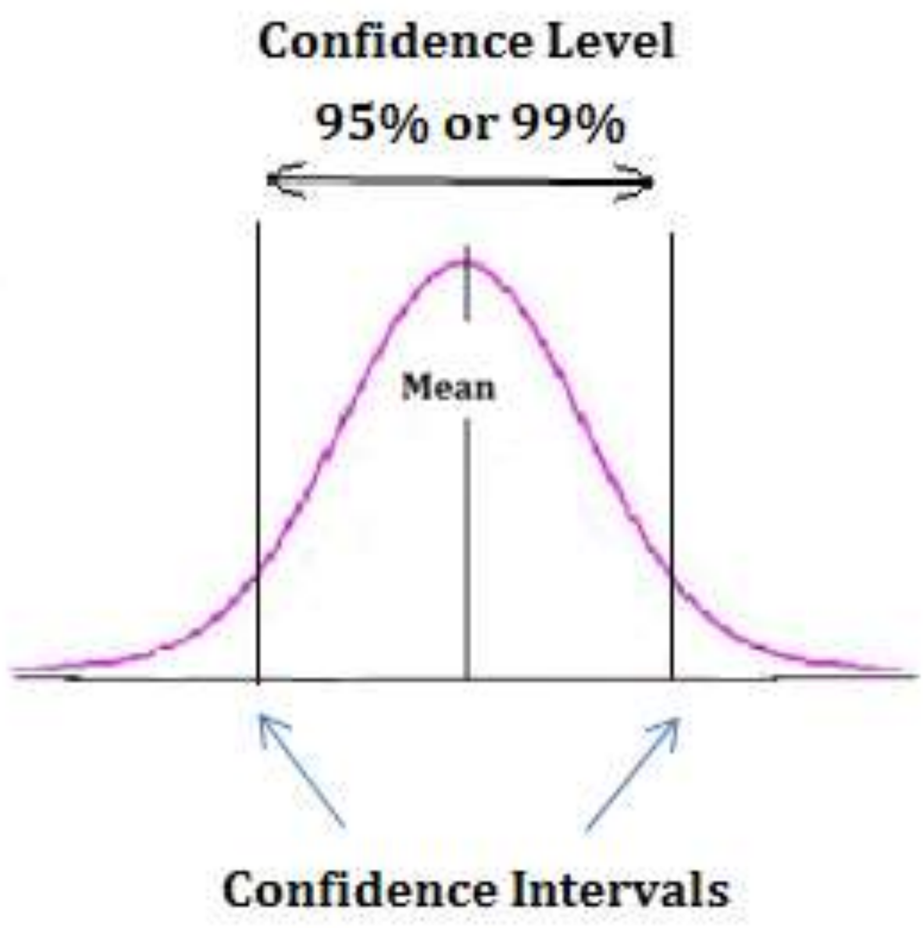
95% : $\alpha = 0.05$ 99% : $\alpha = 0.01$

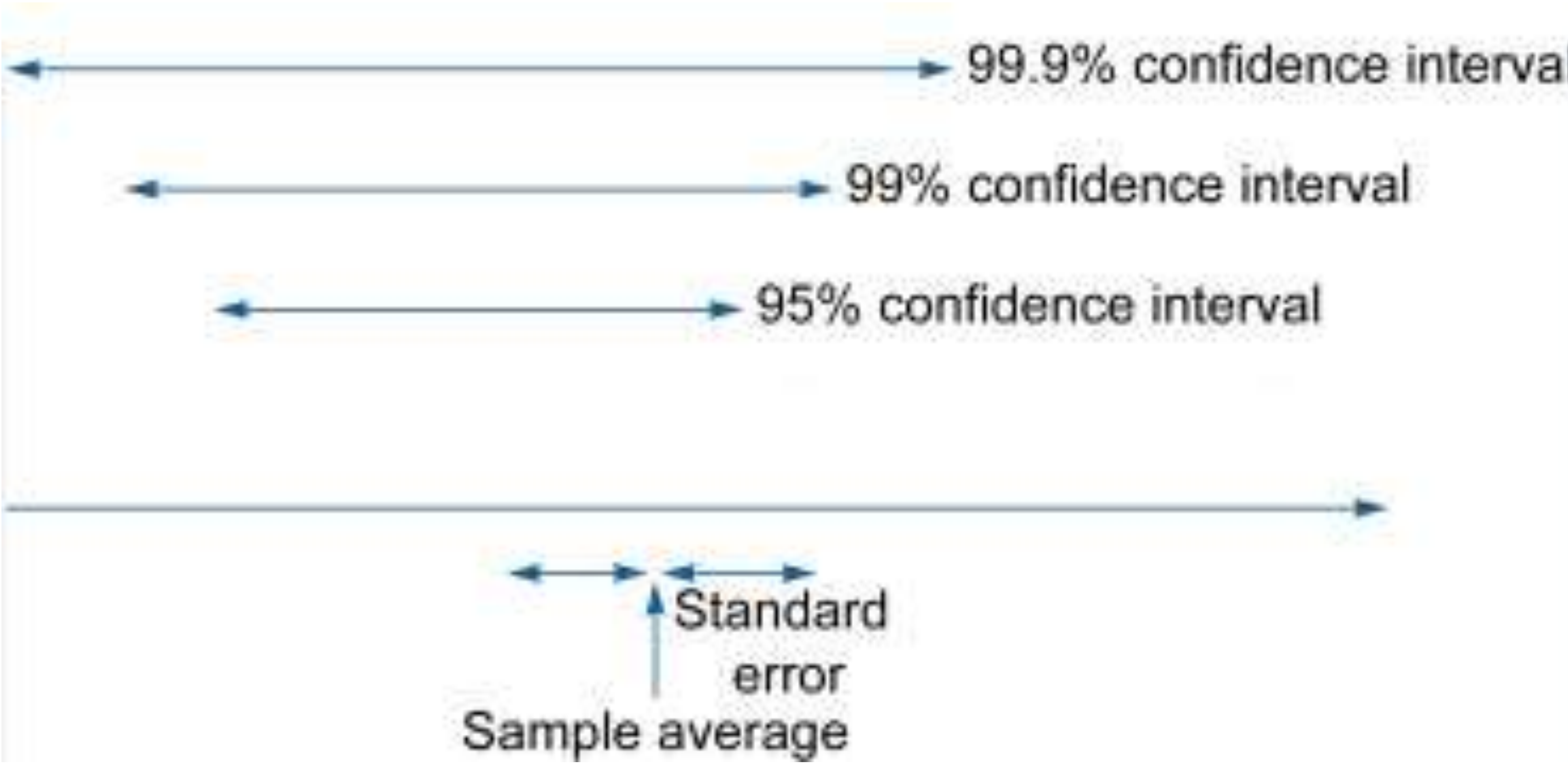
INTERVAL ESTIMATIONS

The **basic confidence levels** are **95%, 99%** and **99.9**%.

A confidence level γ statistics is a measure of the reliability of a result.


Confidence Level
95%
mean
lower
limit
upper
limit


Confidence Level
95% or 99%
Mean
Confidence Intervals


99.9% confidence interval
99% confidence interval
95% confidence interval
Standard
error
Sample average

INTERVAL ESTIMATIONS

The **basic confidence levels** are **95%, 99%** and **99.9**%.

A confidence level γ statistics is a measure of the reliability of a result.

A confidence level of 95 percent or 0.95 means that there is a probability of at least 95 percent that the result is reliable.

INTERVAL ESTIMATIONS

The **parameter** t_γ is defined by Table 2 (it's the table for **Laplace cumulative distribution function**)

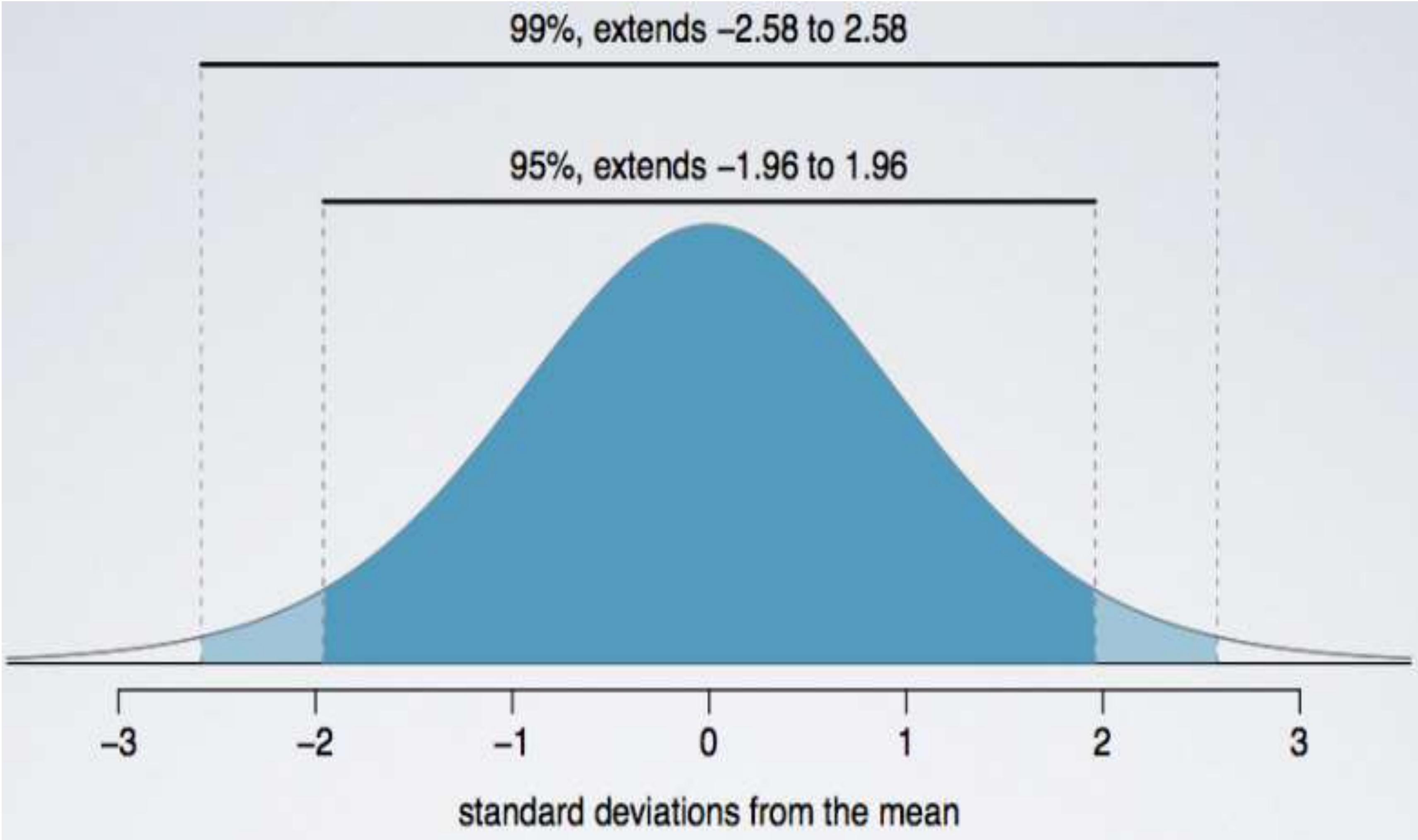
x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2356	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621

x	0	1	2	3	4	5	6	7	8	9
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3906	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,1049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4274	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4430	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4648	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4700	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4762	0,4757
2,0	0,4772	0,4778	0,4783	0,4788	0,4796	0,4798	0,4803	0,4808	0,48Ї2	0,4817

x	0	1	2	3	4	5	6	7	8	9
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4874	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4903	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4924	0,4927	0,4929	0,4930	0,1932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4958	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4973
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4980	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4986	0,4986	0,4987	0,4987	0,4988	0,4988	0,4988	0,4989	0,4989	0,4990

x	0	1	2	3	4	5	6	7	8	9
3,1	0,4990	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993
3,2	0,4993	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998	0,4998	0,4998	0,4998	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,5000	0,5000	0,5000	0,5000	0,5000
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Confidence level	value
95%	1.96
99%	2.58
99,9%	3.291


99%, extends –2.58 to 2.58
95%, extends –1.96 to 1.96
–3
–2
–1
0
1
2
3
standard deviations from the mean

INTERVAL ESTIMATIONS

Task 1. A random variable has

$\overline{x_s} = 12{,}57$ $S_x^2 = 5$ $n = 25$

Construct a confidence interval of the population mean with the confidence probability 95%.

INTERVAL ESTIMATIONS

Task 1. A random variable has

$\overline{x_s} = 12{,}57$ $S_x^2 = 5$ $n = 25$

Construct a confidence interval of the population mean with the confidence probability 95%.

Homework 1. Find confidence intervals with the confidence probability 99% and 99,9%.

INTERVAL ESTIMATIONS

$\overline{x_s} = 12{,}57$

$S_x^2 = 5$

$n = 25$

INTERVAL ESTIMATIONS

$\overline{x_s} = 12{,}57$

$S_x^2 = 5$

$n = 25$

$$S_{cor} = \sqrt{\frac{n}{n-1} \cdot S_x^2} = \sqrt{\frac{25}{25-1} \cdot 5} = \sqrt{5{,}2083} \approx 2{,}28$$

INTERVAL ESTIMATIONS

$\overline{x_s} = 12{,}57$

$S_x^2 = 5$

$n = 25$

$$S_{cor} = \sqrt{\frac{n}{n-1} \cdot S_x^2} = \sqrt{\frac{25}{25-1} \cdot 5} = \sqrt{5{,}2083} \approx 2{,}28$$

$$95\% = 2 \cdot \Phi\left(t_\gamma\right)$$

INTERVAL ESTIMATIONS

$\overline{x_s} = 12{,}57$

$S_x^2 = 5$

$n = 25$

$$S_{cor} = \sqrt{\frac{n}{n-1} \cdot S_x^2} = \sqrt{\frac{25}{25-1} \cdot 5} = \sqrt{5{,}2083} \approx 2{,}28$$

$$95\% = 2 \cdot \Phi(t_\gamma)$$

$$0{,}95 = 2 \cdot \Phi(t_\gamma)$$

INTERVAL ESTIMATIONS

$\overline{x_s} = 12{,}57$ $S_x^2 = 5$ $n = 25$

$$S_{cor} = \sqrt{\frac{n}{n-1} \cdot S_x^2} = \sqrt{\frac{25}{25-1} \cdot 5} = \sqrt{5{,}2083} \approx 2{,}28$$

$$95\% = 2 \cdot \Phi(t_\gamma)$$

$$0{,}95 = 2 \cdot \Phi(t_\gamma)$$

$$0{,}475 = \Phi(t_\gamma)$$

x	0	1	2	3	4	5	6	7	8	9
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3906	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,1049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4274	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4430	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4648	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4700	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4762	0,4757
2,0	0,4772	0,4778	0,4783	0,4788	0,4796	0,4798	0,4803	0,4808	0,48Ї2	0,4817

$$\Phi(1,96) = \Phi(t_\gamma)$$

$$t_\gamma = 1,96$$

$$\overline{x_s} - \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}} < \overline{X}_{pop} < \overline{x_s} + \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}}$$

$$\Phi(1{,}96) = \Phi(t_\gamma)$$

$$\overline{x_s} - \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}} < \overline{X}_{pop} < \overline{x_s} + \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}}$$

$$t_\gamma = 1{,}96$$

$$12{,}57 - \frac{2{,}28 \cdot 1{,}96}{\sqrt{25}} < \overline{X}_{pop} < 12{,}57 + \frac{2{,}28 \cdot 1{,}96}{\sqrt{25}}$$

$$\Phi(1{,}96) = \Phi(t_\gamma)$$

$$\overline{x_s} - \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}} < \overline{X}_{pop} < \overline{x_s} + \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}}$$

$$t_\gamma = 1{,}96$$

$$12{,}57 - \frac{2{,}28 \cdot 1{,}96}{\sqrt{25}} < \overline{X}_{pop} < 12{,}57 + \frac{2{,}28 \cdot 1{,}96}{\sqrt{25}}$$

$$11{,}67 < \overline{X}_{pop} < 13{,}46$$

$$\Phi(1{,}96) = \Phi(t_\gamma)$$

$$t_\gamma = 1{,}96$$

$$\overline{x_s} - \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}} < \overline{X}_{pop} < \overline{x_s} + \frac{S_{cor} \cdot t_\gamma}{\sqrt{n}}$$

$$12{,}57 - \frac{2{,}28 \cdot 1{,}96}{\sqrt{25}} < \overline{X}_{pop} < 12{,}57 + \frac{2{,}28 \cdot 1{,}96}{\sqrt{25}}$$

$$11{,}67 < \overline{X}_{pop} < 13{,}46$$

$\overline{X}_{pop}$ lies in the confidence interval $(11{,}67; 13{,}46)$ with the confidence probability 95%.

INTERVAL ESTIMATIONS

2. The <u>confidence interval</u> of the <u>population root-mean-square deviation</u>

$$P\left(\left| S_{cor} - S_{pop} \right| < \varepsilon\right) = \gamma$$

INTERVAL ESTIMATIONS

2. The <u>confidence interval</u> of the <u>population root-mean-square deviation</u>

$$P\left(\left|S_{cor} - S_{pop}\right| < \varepsilon\right) = \gamma$$

$$q = \frac{\varepsilon}{S_{cor}}$$

INTERVAL ESTIMATIONS

2. The <u>confidence interval</u> of the <u>population root-mean-square deviation</u>

$$P\left(\left| S_{cor} - S_{pop} \right| < \varepsilon\right) = \gamma$$

$$q = \frac{\varepsilon}{S_{cor}}$$

$$S_{cor} - \varepsilon < S_{pop} < S_{cor} + \varepsilon$$

INTERVAL ESTIMATIONS

2. The <u>confidence interval</u> of the <u>population root-mean-square deviation</u>

$$S_{cor} - \varepsilon < S_{pop} < S_{cor} + \varepsilon$$

INTERVAL ESTIMATIONS

2. The <u>confidence interval</u> of the <u>population root-mean-square deviation</u>

$$S_{cor} - \varepsilon < S_{pop} < S_{cor} + \varepsilon$$

$$S_{cor}\left(1 - \frac{\varepsilon}{S_{cor}}\right) < S_{pop} < S_{cor}\left(1 + \frac{\varepsilon}{S_{cor}}\right)$$

INTERVAL ESTIMATIONS

2. The <u>confidence interval</u> of the <u>population root-mean-square deviation</u>

$$S_{cor} - \varepsilon < S_{pop} < S_{cor} + \varepsilon$$

$$S_{cor}\left(1 - \frac{\varepsilon}{S_{cor}}\right) < S_{pop} < S_{cor}\left(1 + \frac{\varepsilon}{S_{cor}}\right)$$

$$q = \frac{\varepsilon}{S_{cor}}$$

INTERVAL ESTIMATIONS

2. The <u>confidence interval</u> of the <u>population root-mean-square deviation</u>

$$S_{cor} \cdot (1-q) < S_{pop} < S_{cor} \cdot (1+q)$$ if $q < 1$

INTERVAL ESTIMATIONS

2. The <u>confidence interval</u> of the <u>population root-mean-square deviation</u>

$$S_{cor} \cdot (1-q) < S_{pop} < S_{cor} \cdot (1+q)$$ if $q < 1$

$$0 < S_{pop} < S_{cor} \cdot (1+q)$$ if $q > 1$

INTERVAL ESTIMATIONS

2. The <u>confidence interval</u> of the <u>population root-mean-square deviation</u>

$S_{cor} \cdot (1-q) < S_{pop} < S_{cor} \cdot (1+q)$ if $q < 1$

$0 < S_{pop} < S_{cor} \cdot (1+q)$ if $q > 1$

The parameter $q = q(\gamma, n)$ is defined by the **table 3** with the confidence level γ (**95%, 99%** and **99.99%**)

INTERVAL ESTIMATIONS

Values $q = q(\gamma, n)$ **(TABLE 3)**

	γ				γ		
	0,95	**0,99**	**0,999**		**0,95**	**0,99**	**0,999**
5	1,37	2,67	5,64	**20**	0,37	0,58	0,88
6	1,09	2,01	3,88	**25**	0,32	0,49	0,73
7	0,92	1,62	2,98	**30**	0,28	0,43	0,63
8	0,80	1,38	2,42	**35**	0,26	0,38	0,56
9	0,71	1,20	2,06	**40**	0,24	0,35	0,50
10	0,65	1,08	1,80	**45**	0,22	0,32	0,46
11	0,59	0,98	1,60	**50**	0,21	0,30	0,43
12	0,55	0,90	1,45	**60**	0,188	0,269	0,38
13	0,52	0,83	1,33	**70**	0,174	0,245	0,34
14	0,48	0,78	1,23	**80**	0,161	0,226	0,31
15	0,46	0,73	1,15	**90**	0,151	0,211	0,29
16	0,44	0,70	1,07	**100**	0,143	0,198	0,27
17	0,42	0,66	1,01	**150**	0,115	0,160	0,211
18	0,40	0,63	0,96	**200**	0,099	0,136	0,185
19	0,39	0,60	0,92	**250**	0,089	0,120	0,162

INTERVAL ESTIMATIONS

Task 2. A random variable has

$$S_{cor} = 2{,}82 \qquad n = 50$$

Construct a confidence interval of the population root-mean-square deviation with the confidence probability 99%.

INTERVAL ESTIMATIONS

Task 2. A random variable has

$S_{cor} = 2{,}82$ $\quad\quad n = 50$

Construct a confidence interval of the population root-mean-square deviation with the confidence probability 99%.

Homework 2. **Find** confidence intervals with the confidence probability 95% and 99,9%.

$$q = q(\gamma, n) = q(0{,}99;\, 50) = 0{,}30$$

Values $q = q(\gamma, n)$ (**TABLE 3**)

	γ				γ		
	0,95	**0,99**	**0,999**		**0,95**	**0,99**	**0,999**
5	1,37	2,67	5,64	**20**	0,37	0,58	0,88
6	1,09	2,01	3,88	**25**	0,32	0,49	0,73
7	0,92	1,62	2,98	**30**	0,28	0,43	0,63
8	0,80	1,38	2,42	**35**	0,26	0,38	0,56
9	0,71	1,20	2,06	**40**	0,24	0,35	0,50
10	0,65	1,08	1,80	**45**	0,22	0,32	0,46
11	0,59	0,98	1,60	**50**	0,21	0,30	0,43
12	0,55	0,90	1,45	**60**	0,188	0,269	0,38
13	0,52	0,83	1,33	**70**	0,174	0,245	0,34
14	0,48	0,78	1,23	**80**	0,161	0,226	0,31
15	0,46	0,73	1,15	**90**	0,151	0,211	0,29
16	0,44	0,70	1,07	**100**	0,143	0,198	0,27
17	0,42	0,66	1,01	**150**	0,115	0,160	0,211
18	0,40	0,63	0,96	**200**	0,099	0,136	0,185
19	0,39	0,60	0,92	**250**	0,089	0,120	0,162

$$S_{cor} \cdot (1-q) < S_{pop} < S_{cor} \cdot (1+q)$$

$$S_{cor} \cdot (1 - q) < S_{pop} < S_{cor} \cdot (1 + q)$$

$$S_{cor} = 2{,}82$$

$$S_{cor} \cdot (1-q) < S_{pop} < S_{cor} \cdot (1+q)$$

$$S_{cor} = 2{,}82$$

$$2{,}82 \cdot (1-0{,}3) < S_{pop} < 2{,}82 \cdot (1+0{,}3)$$

$$S_{cor} \cdot (1-q) < S_{pop} < S_{cor} \cdot (1+q)$$

$$S_{cor} = 2{,}82$$

$$2{,}82 \cdot (1-0{,}3) < S_{pop} < 2{,}82 \cdot (1+0{,}3)$$

$$1{,}974 < S_{pop} < 3{,}666$$

with the confidence probability 99%

Thank You for your time and attention! Hope this presentation was educational and helpful to you ☺