Continuous statistical series

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	

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	

100 - 200 ⇒ [100; 200)

200 - 300 ⇒ [200; 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	

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

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					

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

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					

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

$$h = 200 - 100 = 100$$

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					

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

$$h = 300 - 200 = 100$$

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				

$$x_2' = \frac{200+300}{2} = \frac{500}{2} = 250$$

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			

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

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		

$$x'_4 = \frac{400 + 500}{2} = \frac{900}{2} = 450$$

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	

$$x_5' = \frac{500+600}{2} = \frac{1100}{2} = 550$$

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						

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

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					

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

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				

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

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			

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

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		

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

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	

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

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**

Empirical distribution function

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

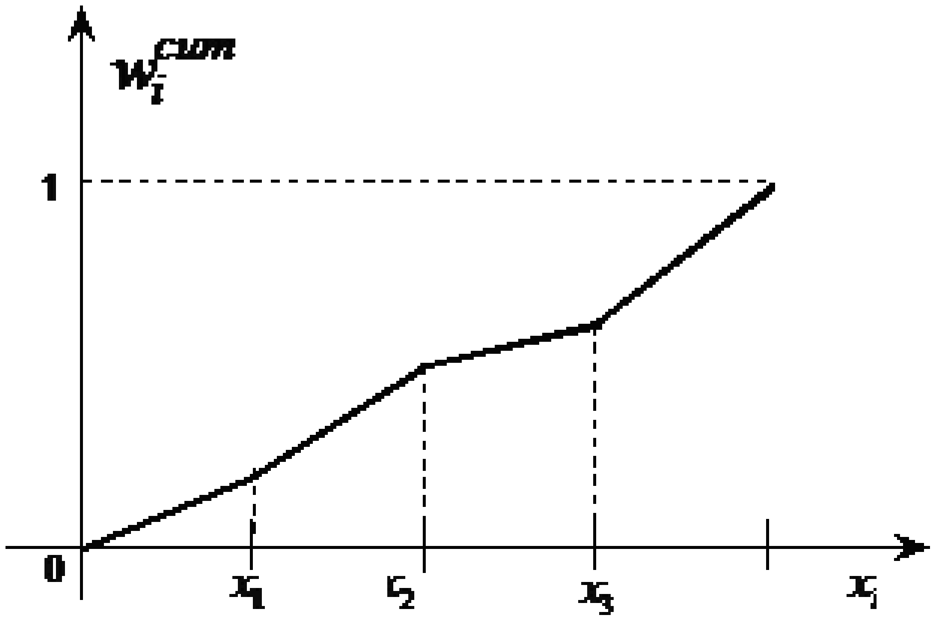
$$F_i = w_1 + w_2 + ... + w_i$$

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

A cumulative function

It is denoted and calculated as

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



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						

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						

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					

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				

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			

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		

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	

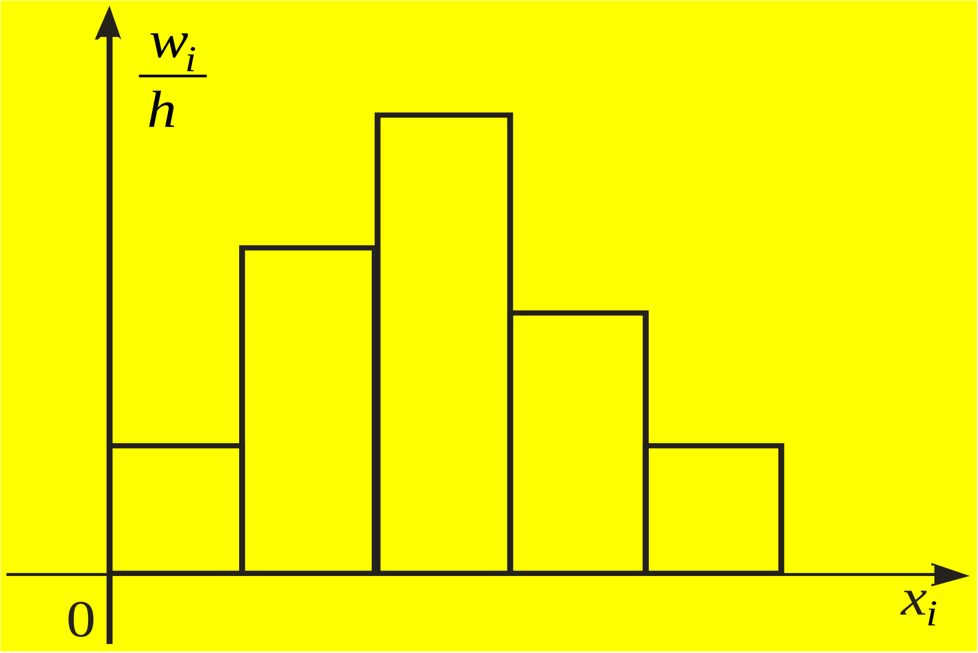
CONTINUOUS STATISTICAL SERIES

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 $[x_i, x_{i+1})$.

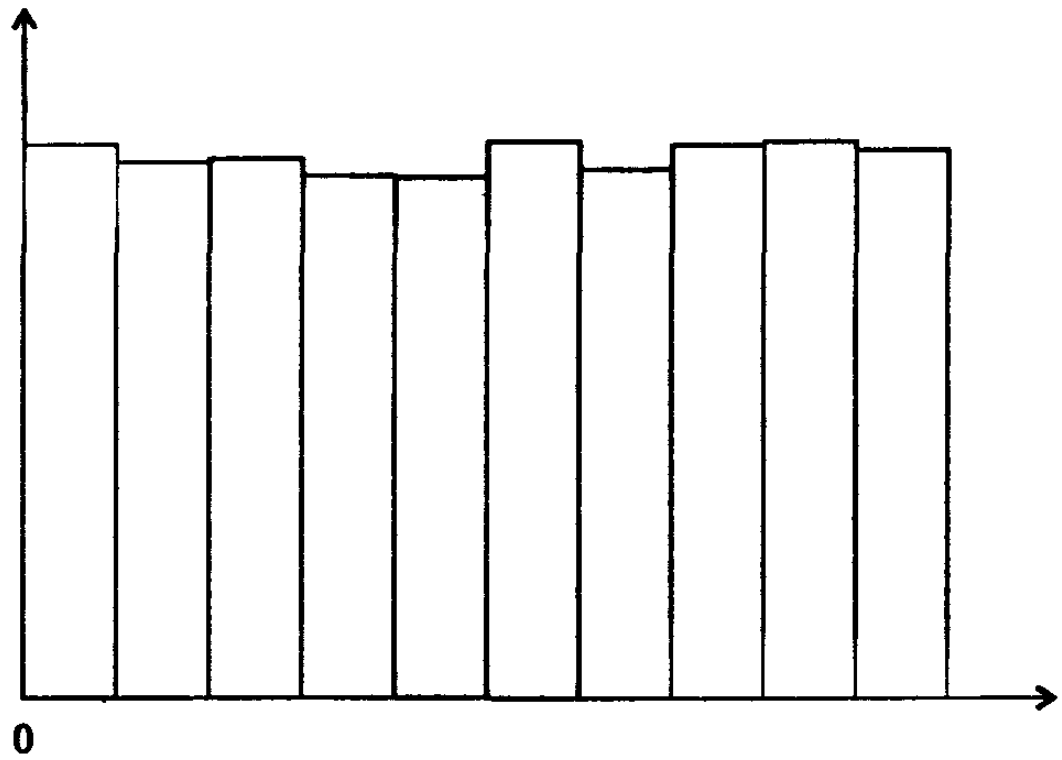
CONTINUOUS STATISTICAL SERIES ($n \geq 30$)

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



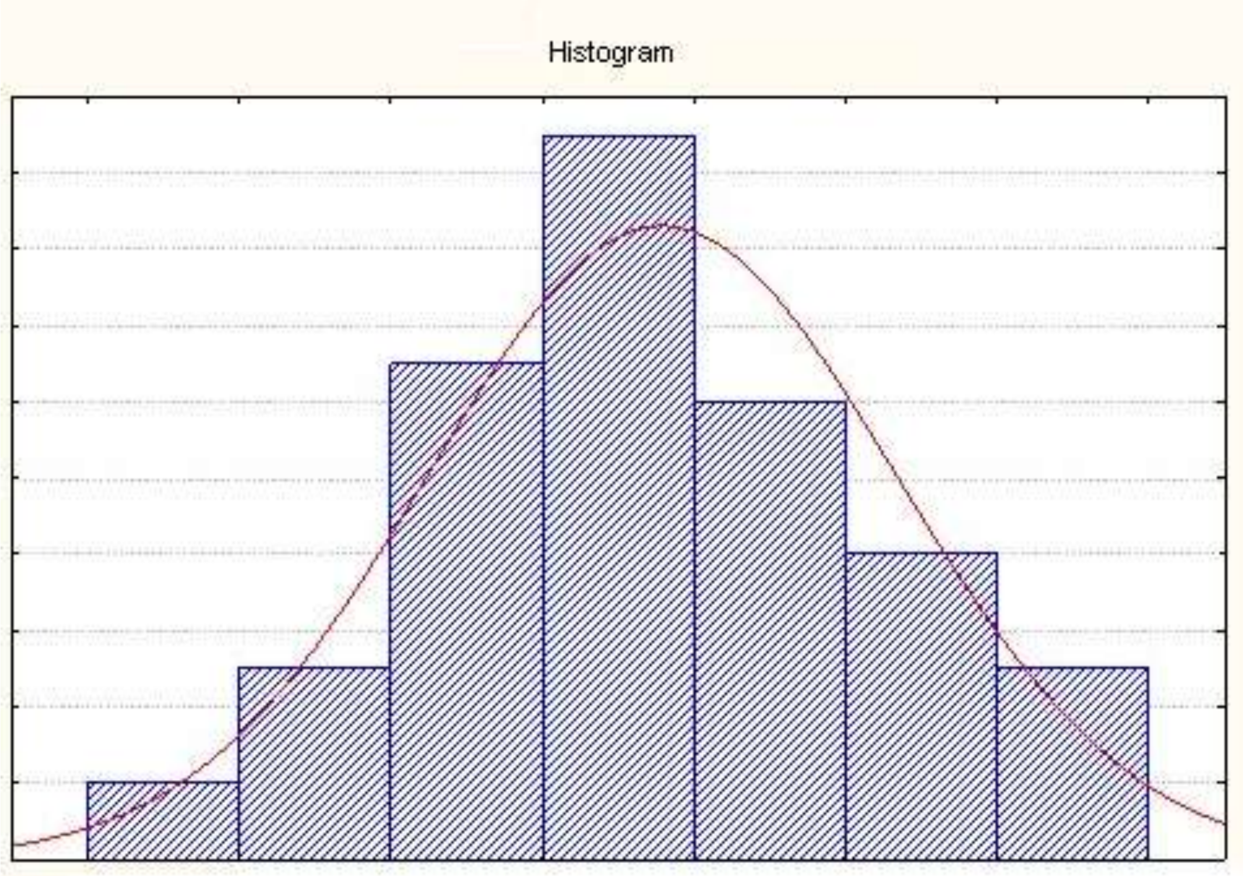
THE ASSUMPTION ABOUT THE DISTRIBUTION LAW (THE HYPOTHESIS)

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



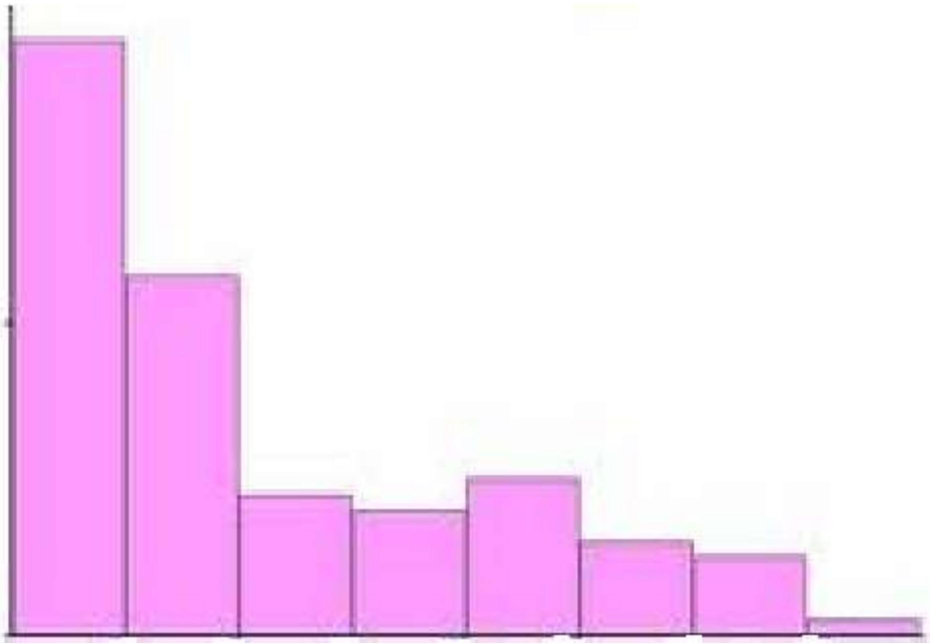
THE ASSUMPTION ABOUT THE DISTRIBUTION LAW (THE HYPOTHESIS)

- 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						

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

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				

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

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		

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

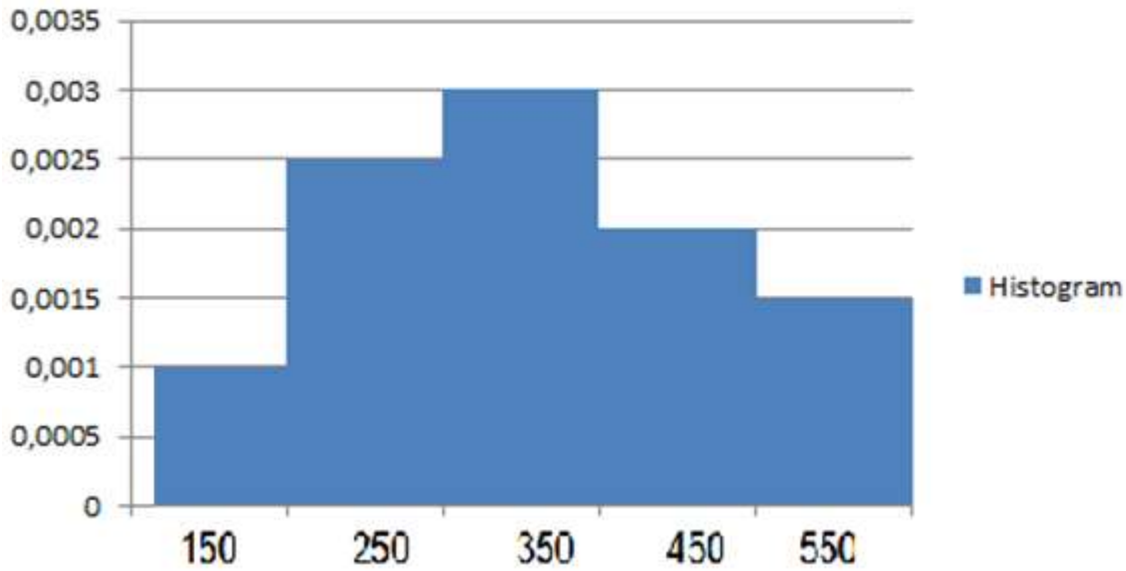
EXAMPLE (CSS)

$[x_i; x_{i+1})$	**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	

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'

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**
$x_i' \cdot w_i$						

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					

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

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					

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

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			

$$x'_4 \cdot w_4 = 450 \cdot 0{,}2 = 90$$

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		

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

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**

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** $= \overline{x_s}$

$$\overline{x_s} = 355$$

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				

$$(x'_2)^2 \cdot w_2 = 250 \cdot 62{,}5 = 15625$$

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			

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

56

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		

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

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	

$$(x_5')^2 \cdot w_5 = 550 \cdot 82{,}5 = 45375$$

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**	$= \overline{x_s^2}$

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

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** $= \overline{x_s^2}$

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

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** $=\overline{x_s^2}$

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

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** $= \overline{x_s^2}$

$$S_x^2 = 14475$$

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** $= \overline{x_s^2}$

$$S_x^2 = 14475$$

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

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** $= \overline{x_s^2}$

$$S_x^2 = 14475$$

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

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** $= \overline{x_s^2}$

$$S_x^2 = 14475$$

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

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***

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

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

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

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

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 $F_i > 0,5$

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 $F_i > 0{,}5$

$F_i = 0{,}65$

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 $F_i > 0{,}5$

$F_i = 0{,}65$ $x_i = 300$

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: $x_i = 300$

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 $F_i > 0,5$

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

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 $F_i > 0{,}5$

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

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	

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

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

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	

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

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

$$Mo = ?$$

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

$$m_{M_0} = 60$$

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

$$m_{M_0} = 60$$

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

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 estimations of the distribution parameters

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

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

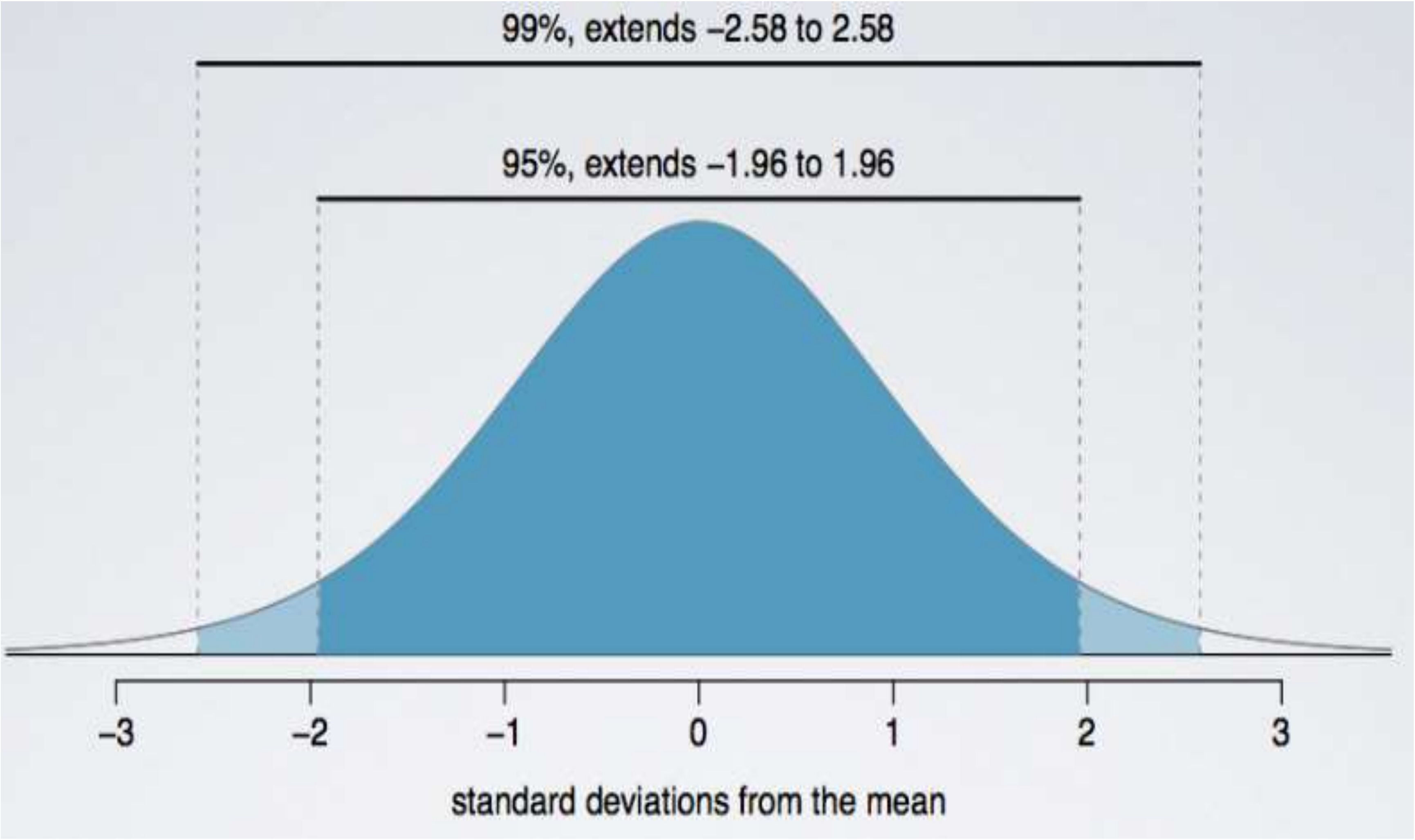
Confidence Level : $1 - \alpha$

The **probability** that the confidence interval actually contains the population parameter.

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

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

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

$\overline{x_s} = 355$ $S_{cor} = 120{,}61$ $n = 200$

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

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

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

$$355 - \frac{120{,}61 \cdot 1{,}96}{\sqrt{200}} < \overline{X}_{pop} < 355 + \frac{120{,}61 \cdot 1{,}96}{\sqrt{200}}$$

$t_\gamma = 1{,}96$

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

$$355 - 16{,}72 < \overline{X}_{pop} < 355 + 16{,}72$$

$$338{,}28 < \overline{X}_{pop} < 371{,}72$$

$\overline{X}_{pop}$ lies in the confidence interval $(338{,}28; 371{,}72)$ with the confidence probability 95%.

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

Table 3 of values $q = q(\gamma, n)$

n	γ 0,95	γ 0,99	γ 0,999	n	γ 0,95	γ 0,99	γ 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

$S_{cor} = 120{,}61$ $n = 200$

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

$$q = q(\gamma, n) = q(0{,}99,\ 200) = 0{,}136 < 1$$

n	γ			n	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

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

$$S_{cor} = 120{,}61$$

$$120{,}61\cdot(1-0{,}136) < S_{pop} < 120{,}61\cdot(1+0{,}136)$$

$$104{,}21 < S_{pop} < 137{,}01$$

with the confidence probability 99%

STATISTICAL PARAMETERS
(NUMERICAL CHARACTERISTICS)
OF A SAMPLE

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

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

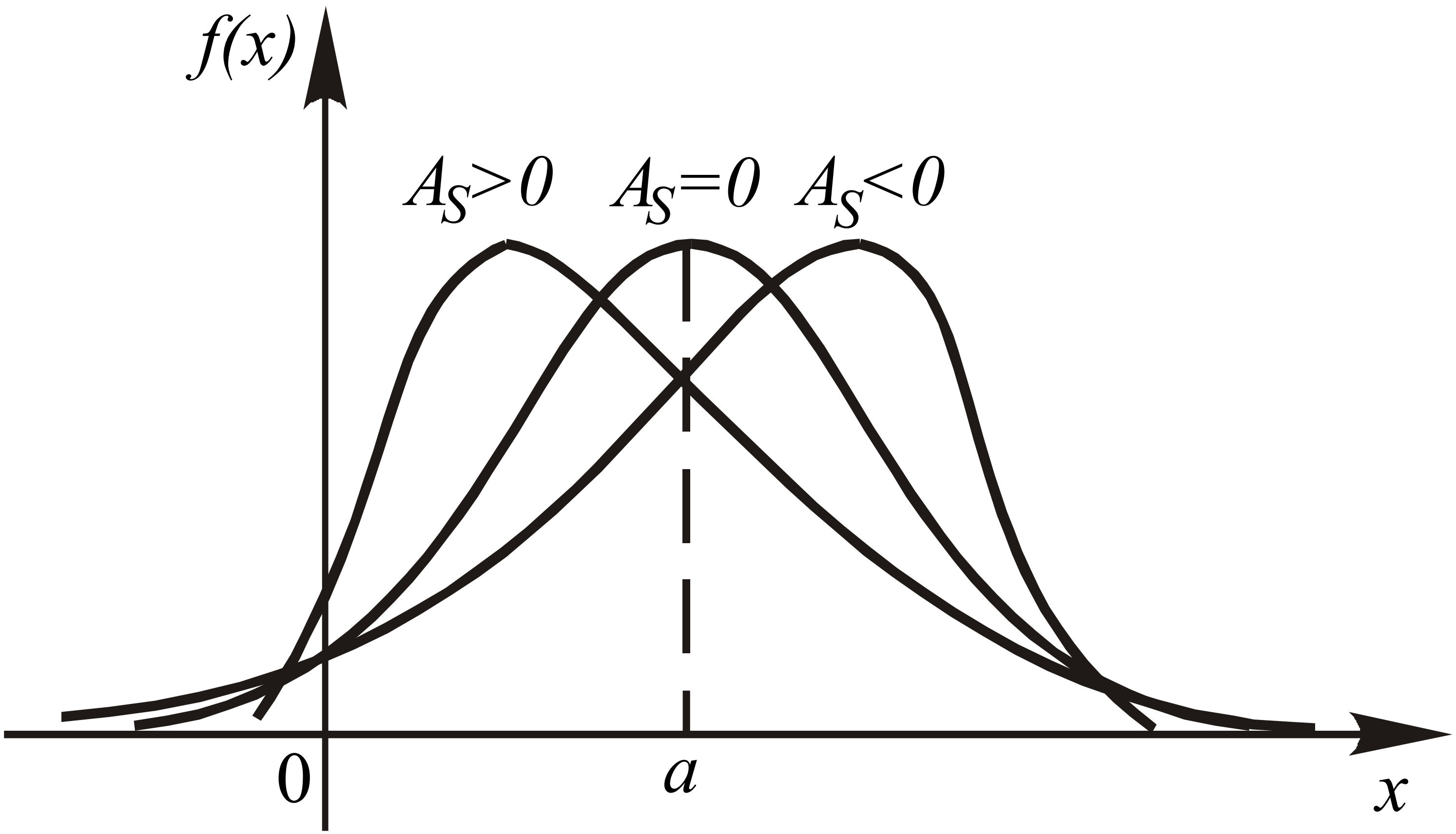
Asymmetry shows a deviation of a value from its central position on the left or on the right:

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

where

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

S_x – the root-mean square deviation.


f(x)
AS>0
AS=0
AS<0
0
a
x

STATISTICAL PARAMETERS

(NUMERICAL CHARACTERISTICS)

OF A SAMPLE

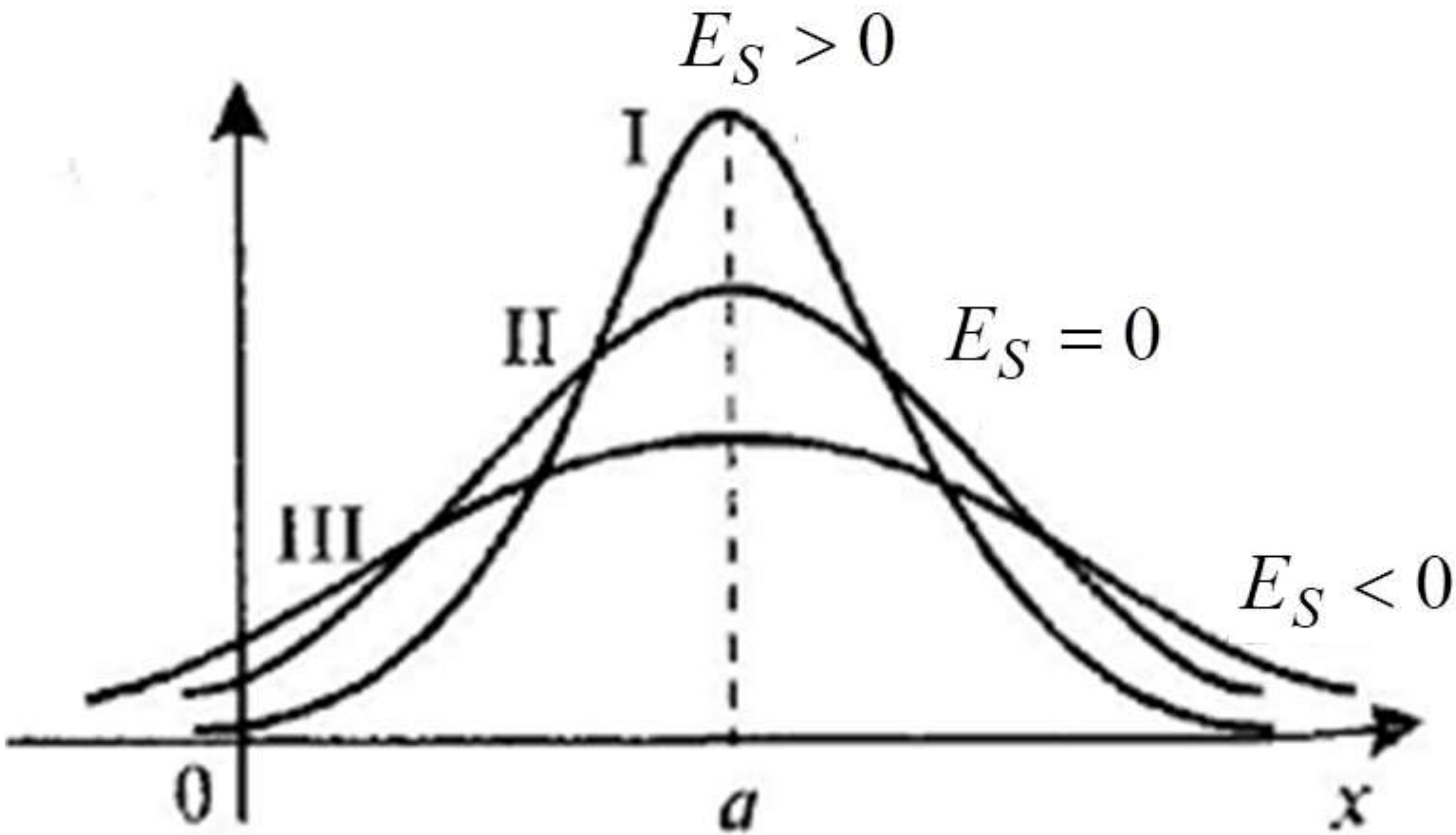
10. ***Excess*** characterizes a deviation of a value from its central position down or up:

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

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

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.

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