Continuous statistical variables

Continuous variable

A continuous statistical variable is every quantitative (numerical) variable whose data may assume any value in an interval.

Example:

Classification of students according to their score on a math test over 20:

“resist” corresponds to a score in [0;10[

“Pass” corresponds to a score in [10;12[

“fair” corresponds to a score in [12;14[

“good” corresponds to a score in [14;1[

“very good” corresponds to a score in [16;18[

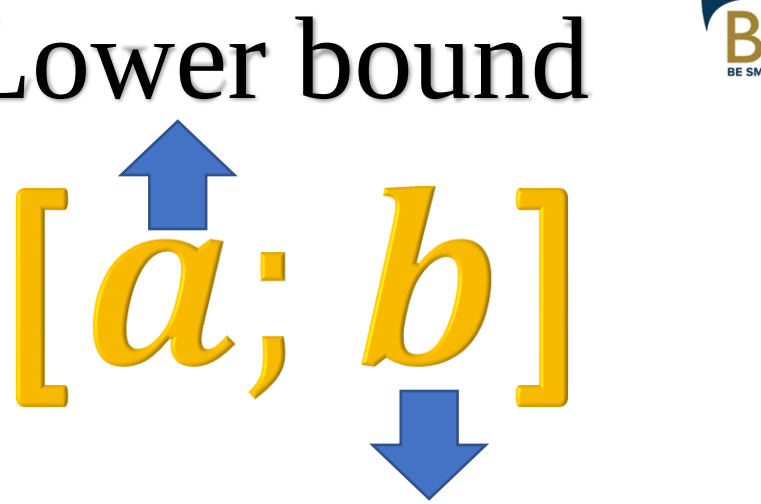
“Excellent” corresponds to a score in [18;20]

Grouping data in classes

Recall (Interval)

Center (midpoint): $c = \frac{a+b}{2}$

width: $l = b - a$



How to organize the values into classes?

Grouping data in classes

The marks of students on a math test over 20 are as following:

14 – 19 – 10 – 11 – 15 – 19 – 6 – 12 – 17 – 15

3 – 20 – 9 – 8 – 13 – 19 – 15 – 17 – 7 – 14

To group the data set into classes we need to choose k the number of classes needed, the width of classes is $l = \frac{\text{range}}{k}$.

In this case, all the classes have the same width that are called regular.

It is better that k divides the range so that the classes fit all the data set.

Grouping data in classes

The marks of students on a math test over 20 are as following:
14 – 19 – 10 – 11 – 15 – 19 – 6 – 12 – 17 – 15
3 – 19 – 9 – 8 – 13 – 19 – 15 – 17 – 7 – 14

range=19-3=16
k can be 2 of width 8, 4 of width 4, 8 of width 2.
Suppose that k=4, the classes are: [3;7[- [7;11[- [11;15[- [15;19]

Class	[3;7[	[7;11[	[11;15[	[15;19]	N
n_i	2	4	5	9	20

Grouping data in classes

The marks of students on a math test over 20 are as following:

14 – 19 – 10 – 11 – 15 – 19 – 6 – 12 – 17 – 15

3 – 19 – 9 – 8 – 13 – 19 – 15 – 17 – 7 – 14

Remark:

Suppose that we choose a number of classes that doesn't divide the range.

Example k=3 that doesn't divide 16

$\frac{16}{3} = 5.3333$

It is better that the width l must be natural number.

Suppose l=5, the classes becomes: [3;8[- [8;12[- [12;17] in this case, the number of classes didn't fit all the data set, so increase the width: ex: l=6 [3;9[- [9;14[- [14;20] or increase the number of classes with same width: [3;8[- [8;12[- [12;17[- [17;22].

Grouping data in classes

The marks of students on a math test over 20 are as following:

14 – 19 – 10 – 11 – 15 – 19 – 6 – 12 – 17 – 15

3 – 19 – 9 – 8 – 13 – 19 – 15 – 17 – 7 – 14

There is another method to group data in classes that is more precise:

- Instead of choosing the number of classes, determine the best number of classes: k is the first natural number verifying $2^k \geq N$ where N is the size of the population.

N = 20

$2^1 = 2 < 20$

$2^2 = 4 < 20$

$2^3 = 8 < 20$

$2^4 = 16 < 20$

$2^5 = 32 > 20$ so k=5

Grouping data in classes

The marks of students on a math test over 20 are as following:

14 – 19 – 10 – 11 – 15 – 19 – 6 – 12 – 17 – 15

3 – 19 – 9 – 8 – 13 – 19 – 15 – 17 – 7 – 14

➢ Calculate the width of the classes:

l is the first natural number verifying $l \geq \frac{\text{Range}}{k}$

$\frac{range}{k} = \frac{16}{5} = 3.2$ so l=4

Class	[3;7[	[7;11[	[11;15[	[15;19[	[19;23]	N
n_i	2	4	5	5	4	20

Relative frequency

Class	[3;7[	[7;11[	[11;15[	[15;19[	[19;23]	N
n_i	2	4	5	5	4	20
f_i	$\frac{2}{20} = 0.1$	$\frac{4}{20} = 0.2$	$\frac{5}{20} = 0.25$	**0.25**	**0.2**	**1**
$\% = f_i \times 100$	**10**	**20**	**25**	**25**	**20**	**100**

Increasing/Decreasing frequency

Class	[3;7[	[7;11[	[11;15[	[15;19[	[19;23]	N
n_i	2	4	5	5	4	20
ICF	2	6	11	16	20	
DCF	20	18	14	9	4	

Relative Increasing/Decreasing frequency

Class	[3;7[	[7;11[	[11;15[	[15;19[	[19;23]	N
n_i	2	4	5	5	4	20

First method

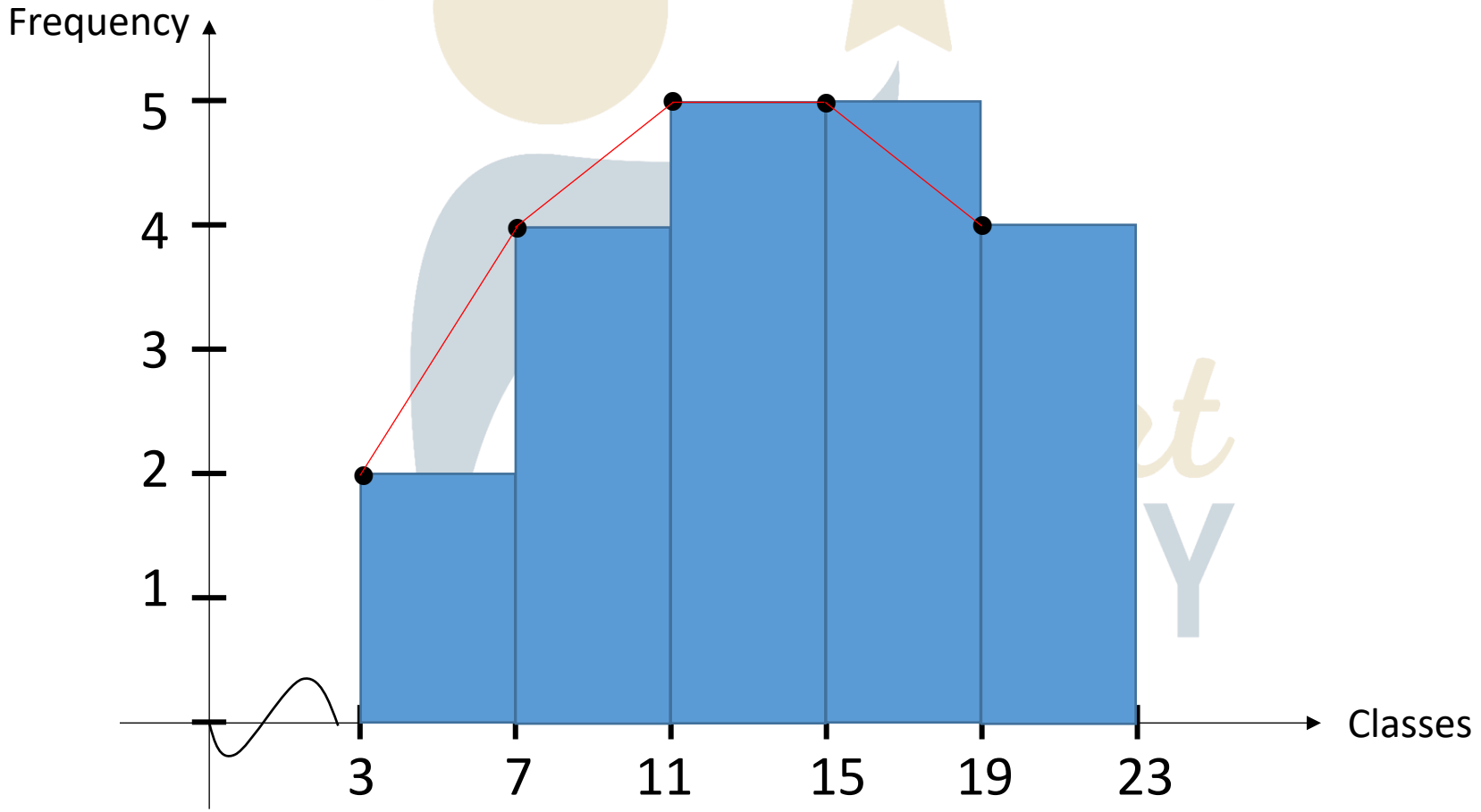
ICF	2	6	11	16	20	
ICRF	$\frac{2}{20} = 0.1$	$\frac{6}{20} = 0.3$	$\frac{11}{20} = 0.55$	$\frac{16}{20} = 0.8$	$\frac{20}{20} = 1$	

Second method

f_i	$\frac{2}{20} = 0.1$	$\frac{4}{20} = 0.2$	$\frac{5}{20} = 0.25$	0.25	0.2	1
ICRF	0.1	0.3	0.55	0.8	1	

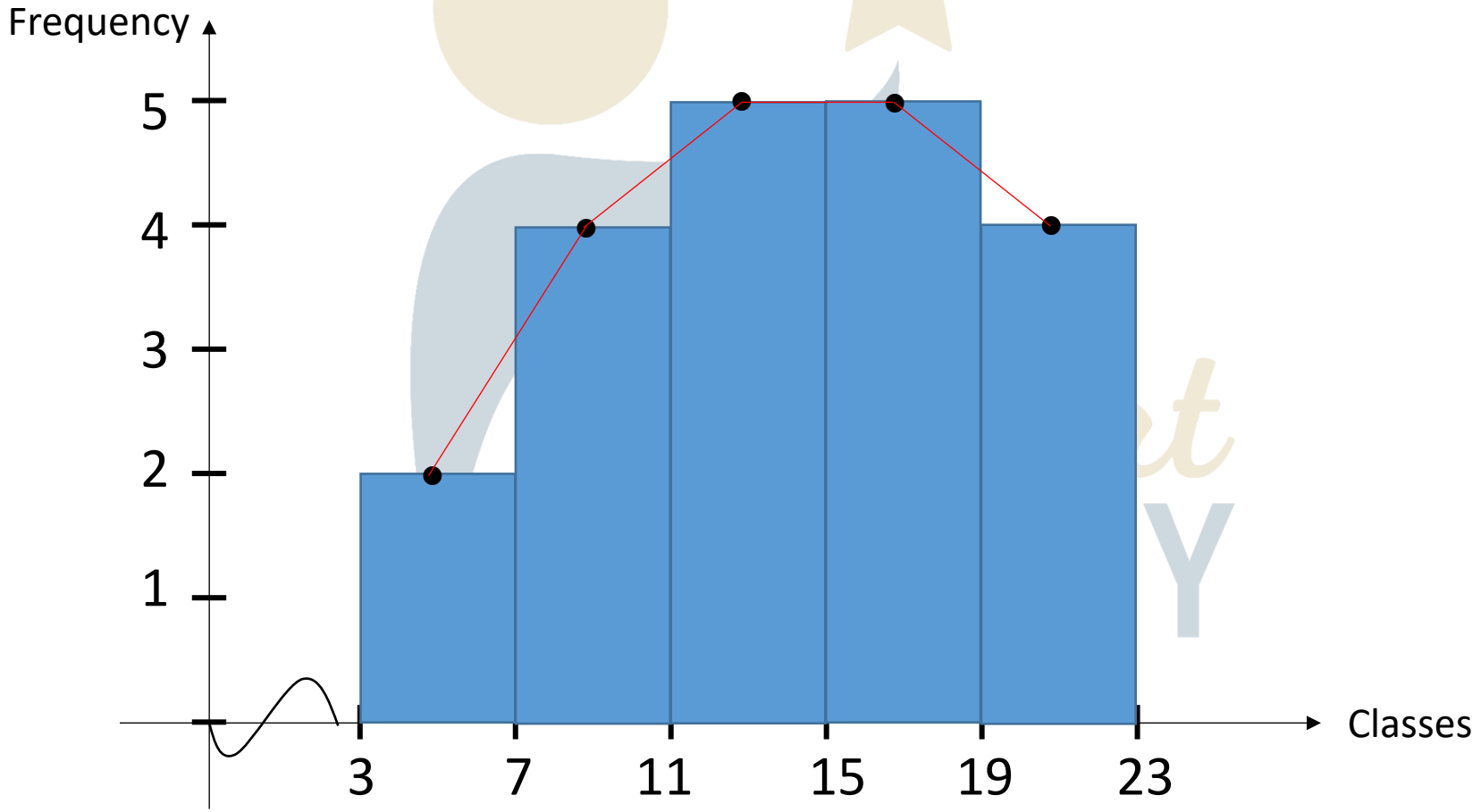
Graphical representation(Histogram+polygon)

Class	[3;7[	[7;11[	[11;15[	[15;19[	[19;23]	N
n_i	2	4	5	5	4	20



Graphical representation(Histogram+polygon)

Class	[3;7[	[7;11[	[11;15[	[15;19[	[19;23]	N
n_i	2	4	5	5	4	20



Graphical representation(Histogram+polygon)

Class	[3;7[	[7;11[	[11;15[	[15;19[	[19;23]	N
n_i	2	4	5	5	4	20