On an indicator of teaching effectiveness based on national assessments results

Luísa Canto e Castro Loura
Maria Isabel Hormigo
Ministry of Education and Science
Portugal

Second Lisbon Research Workshop on Economics, Statistics and Econometrics of Education – ISEG, Lisbon

18-19 January 2013

Motivation

- This study came in the sequence of the publication of the Regulatory Decree 13A/2012, June 2012, which laid down the rules for the new school year, giving more pedagogical and organizational autonomy in what concerns the work distribution for teachers and the organization of teaching schedules.

- Among many other aspects, the main aim was to establish a reward, translated into credit hours, that could promote the reinforcement of school autonomy, ensuring the appropriate mechanisms to its establishment.

Regulatory Decree 13A/2012 - a few indicators

- In each school, the reward (in credit hours) was dependent on several indicators, from which we refer:
  - **the capacity in managing human resources**;
  - **the "teaching effectiveness"** - appreciation of the school investment made in educational attainment, by analysing:
    - the results obtained in the National examinations;
    - the differences between student scores in the National examinations and the ones obtained as the average grades whitin school.
    - The anual variation in school results, by comparing the variation in the examinations scores in each school with the variation in the overall National scores.

How was this rewards assigned to schools that have shown good results of the students in the national examinations, or that showed an evident improvement from 2011 to 2012?

Regulatory Decree 13A/2012 – example: Table 3

TABELA N.º 3

Comparação da variação anual das classificações de exame de cada escola ou agrupamento com a variação anual nacional

Escola ou agrupamento com exames nos ensinos básico e secundário

Condições a verificar	IndSuc3
$CE_{sec\ n} - CE_{sec\ n-1} \geq A_1$ e $CE_{bas\ n} - CE_{bas\ n-1} \geq B_1$	+ 30 h
$CE_{sec\ n} - CE_{sec\ n-1} \geq A_1$ ou $CE_{bas\ n} - CE_{bas\ n-1} \geq B_1$	+ 20 h
$A_2 \leq CE_{sec\ n} - CE_{sec\ n-1} < A_1$ e $B_2 \leq CE_{bas\ n} - CE_{bas\ n-1} < B_1$ ...	+ 20 h
$A_3 \leq CE_{sec\ n} - CE_{sec\ n-1} < A_2$ e $B_3 \leq CE_{bas\ n} - CE_{bas\ n-1} < B_2$	+ 10 h
Nas restantes situações.	+ 0 h

The thresholds A_i and B_i should be high percentiles to be determined from the results of the exams.

CE_{bas} is the average of all the marks obtained by the students of each school in the national examinations of grades 6 and 9 (basic instruction – primary and lower secondary)

CE_{sec} is the average of all the marks obtained by the students of the school in the national examinations of grades 11 and 12 (upper secondary)

How to measure it?

Credit hours should be given to schools for which the differences

$$CE_{bas\ n} - CE_{bas\ n-1},\ CE_{sec\ n} - CE_{sec\ n-1},$$

or both, are high enough.

But how can we decide on what is "high enough" when there is such a big diversity of exams being performed in the schools and when the mean and standard deviation vary so much from discipline to discipline?

The choice went to a statistical significance indicator that measures the difference between the global averages of the marks obtained in each school and their expected values, in analogy with the standard approach in hypothesis testing for the difference of means.

More precisely, the indicator will be no more than the value of the test statistic for the case of known population standard deviations

$$\boxed{\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}}$$

The bigger the value of this quotient the bigger the evidence that the improvement of school results from year n-1 to year n was not due to chance but to the overall school effort.

Which must be the mean values and standard deviations?

Let us look again to the expression

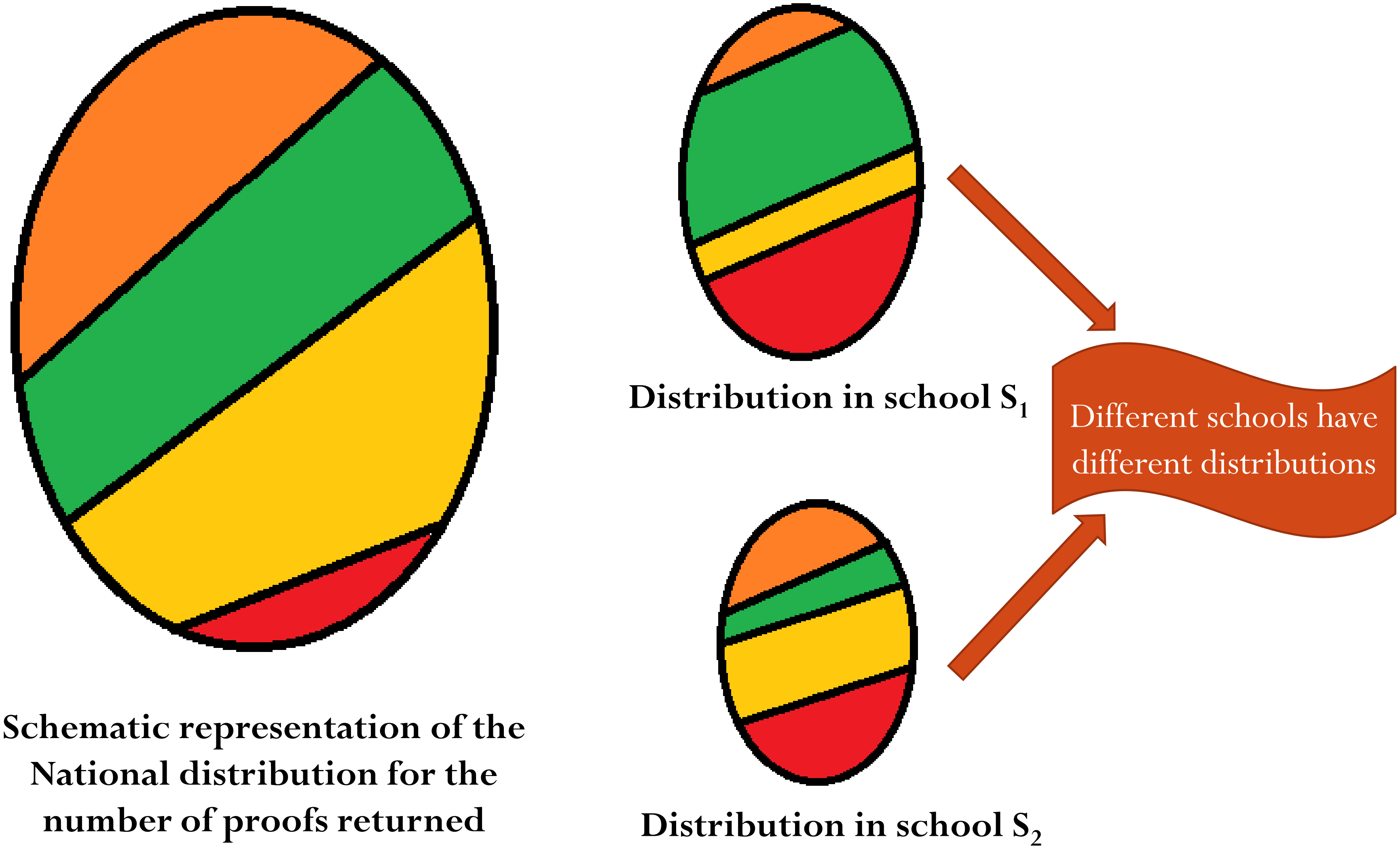
$$\boxed{\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}}$$

μ_1 is the population mean for year n and μ_2 the population mean for year n-1. It is clear that they must be determined from the global National results, but …

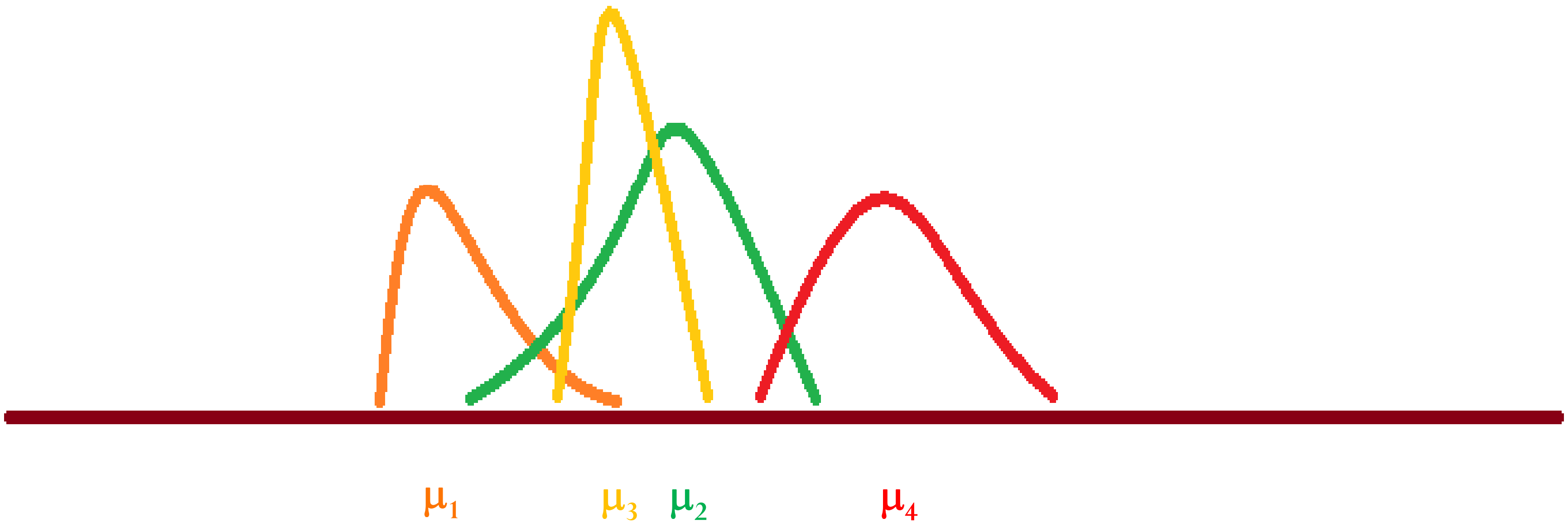
The question is
Shall they be equal for all schools, more precisely, shall we take the global National average of all the marks obtained in all the exams or will it be more reasonable to use some kind of weighting?

Naturally, the same question can be posed in what concerns the standard deviations.

Diversity in the number of proofs returned by the students

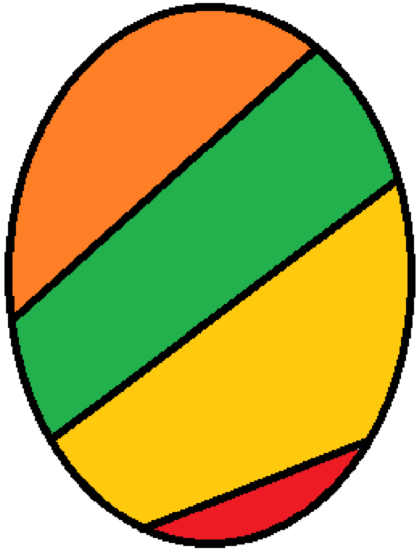


Diversity in location and scale parameters



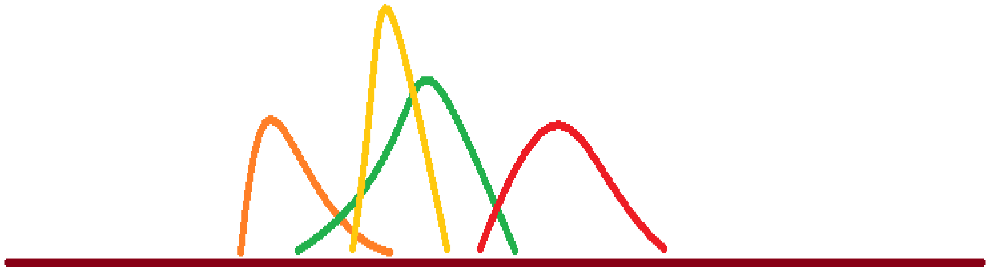
The marks obtained by the students in the exams are differently distributed when changing from one discipline to another, namely in what respects to location and dispersion.

Remark: in Portugal, in upper secondary education, we access to students' achievements in 25 disciplines using national exams.



Diversity in the number of proofs returned

+



Diversity in the distributions

Use of mixture models

Introducing the notation

Let us represent by S_i the i-th school and by E_j the j-th type of exam.
Also, represent by

μ_j – the national average for E_j
σ_j – the standard deviation of all the marks obtained in E_j at a national level
$N_j^{(i)}$- the number of students that performed E_j in the school S_i
$N^{(i)} = \sum_j N_j^{(i)}$ – the total number of proofs returned by the students in S_i
$\omega_j^{(i)} = \frac{N_J^{(i)}}{N^{(i)}}$ – the proportion of E_j proofs returned by the students in S_i.

We can now define the weighted mean or expected mean for the school S_i as being

$$\boxed{\mu^{(i)} = \sum_j \omega_j^{(i)} \mu_j \ .}$$

This is the reference threshold with which we will confront the average of all the scores obtained in the assessments performed in S_i, here represented by $\bar{x}^{(i)}$.

Calculus of the expected mean value

Example (Fictitious)

Disciplines under examination

	Math	Physics	Biology
National mean - μ_j	*102*	*98*	*120*
National s.d. - σ_j	*20*	*17*	*32*

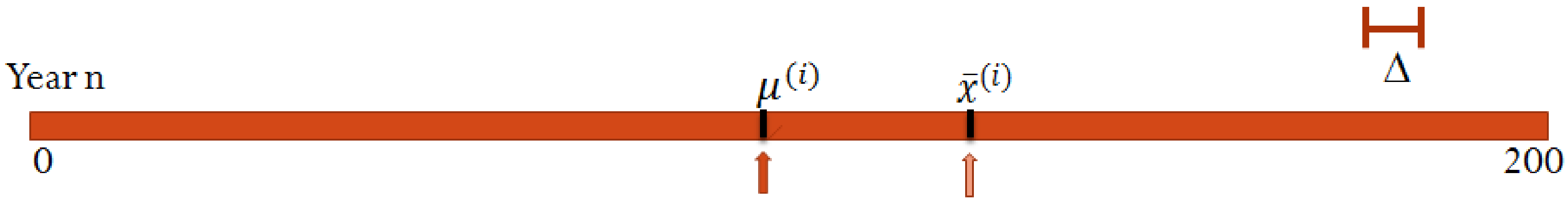
	Number of proofs returned in each school - $N_j^{(i)}$			Total $N^{(i)}$	$\mu^{(i)}$
School S_1	240	54	68	362	104,8
$\omega_j^{(1)}$	*0,66*	*0,15*	*0,19*		
School S_2	16	106	135	257	109,8
$\omega_j^{(2)}$	*0,06*	*0,41*	*0,53*		

In school S_1 the majority of the students performed the Math's examination whilst in school S_2 the main attendance was in Biology. As a consequence, the mean value to be used as a reference in school S_2 is higher than in school S_1.

The simplest case of a single type of examination

If a single type of examination (in a single disciplinine), let us say E_1, is performed in S_i, then it is well known that the standard deviation of the sample average $\bar{x}^{(i)}$ is given by

$$\boxed{\varDelta = \frac{\sigma_1}{\sqrt{N_1^{(i)}}}}$$



A measure of the statistical significance of the difference between $\bar{x}^{(i)}$ and $\mu^{(i)}$ is, in this case, given by the number of times that difference is bigger than $\varDelta$, i. e., it is given by the quocient

$$\boxed{\frac{\bar{x}^{(i)} - \mu^{(i)}}{\varDelta}}$$

The general case

Since we are interested in summing the information contained in the scores of all exams, the alternative is to replace σ_1 by the standard deviation of a mixture model and then standardize it by dividing by the square root of $N^{(i)}$, the total number of proofs returned in S_i .

Now, the expected variability for the marks in all the test assessments performed in S_i , assuming the population parameters μ_j and σ_j , is given by

$$\boxed{\left[\sigma^{(i)}\right]^2 = \sum_j \omega_j^{(i)} \left(\left(\mu_j - \mu^{(i)}\right)^2 + \sigma_j^2\right)}$$

$\sigma^{(i)}$, is no more than a weighted deviation that takes into account, not only the range of disciplines under examination, but also the proportion of students who performed each type of examination.

We then have that, regardless of the number of test types carried out in S_i, the standard deviation of the sample average $\bar{x}^{(i)}$will be given by

$$\boxed{\Delta = \frac{\sigma^{(i)}}{\sqrt{N^{(i)}}}}$$

Calculus of the weighted variance

Example (Fictitious) - continued

$$\left[\sigma^{(i)}\right]^2 = \sum_j \omega_j^{(i)} \left(\left(\mu_j - \mu^{(i)}\right)^2 + \sigma_j^2\right)$$

Disciplines under examination

	Math	Physics	Biology
National mean - μ_j	102	98	120
National s.d. - σ_j	20	17	32

Number of proofs returned in each school - $N_j^{(i)}$

	Math	Physics	Biology	Total $N^{(i)}$	$\mu^{(i)}$
School S_1	240	54	68	362	104,8
$\omega_j^{(1)}$	0,66	0,15	0,19		
School S_2	16	106	135	257	109,8
$\omega_j^{(2)}$	0,06	0,41	0,53		

Disciplines under examination

	Math	Physics	Biology
School S_1 $\mu_j - \mu^{(1)}$	-2,8	-6,8	15,2
School S_2 $\mu_j - \mu^{(2)}$	-7,8	-11,8	10,2

$0,66 \text{x}((2,8)^2+20^2)$

				$\sigma^{(i)}$ square	$\sigma^{(i)}$
$\sigma^{(1)}$ square comp.	270,3	50,0	235,8	556,2	23,6
$\sigma^{(2)}$ square comp.	28,7	176,7	592,5	797,9	28,2

The theoretical weighted deviation for school S_2 is more than four points higher than the one for school S_1.

Measuring the statistical significance

Example (Fictitious) - continued

Assuming for each school the average scores in the three disciplines as listed below, we can now derive the statistical significance of the difference between these average scores and the respective expected mean values.

	Average Scores - $\bar{x}^{(i)}$	S.d. of the average	$\bar{x}^{(i)} - \mu^{(i)}$	Stat. Sig. of $\bar{x}^{(i)} - \mu^{(i)}$
School S_1	116,3	1,24	11,5	9,3
School S_2	125,1	1,76	15,3	8,7

So, we can conclude that school S_1's results are more significantly higher than its mean reference value comparing with the school S_2's results despite of this one's higher average and higher difference to its mean reference. This is due, not only to school S_1's smaller expected variability but also to its larger size in what respects the number of students.

Comparison of the results in two different school years

We can now turn back to our main objective which is the comparison of the results of the school S_i in two different school years (year 1 and year 2). In fact, we can rewrite the test statistic for the difference of two means in a more precise form

$$\boxed{\frac{\left(\bar{x}_2^{(i)} - \bar{x}_1^{(i)}\right) - \left(\mu_2^{(i)} - \mu_1^{(i)}\right)}{\sqrt{\dfrac{\left[\sigma_1^{(i)}\right]^2}{N_1^{(i)}} + \dfrac{\left[\sigma_2^{(i)}\right]^2}{N_2^{(i)}}}}}$$

Where $\mu_1^{(i)}$ and $\mu_2^{(i)}$ are the means of appropriate mixture models and $\sigma_1^{(i)}$ and $\sigma_2^{(i)}$ are the corresponding standard deviations.

National examinations' global results for secondary level

(File excerpt)

School/cluster of schools code	School / cluster of schools denomination	School's average scores in the national assessments (A)	Total Number of tests performed (N)	National mean value, weighted by the distribution of tests performed in the school (M)	Standard deviation of the average (D =B/sqrt(N))	Statistical signifacance indicator (A-M)/D
...	Escola	103,63	404	95,12	2,11	4,033
...	Escola	119,79	647	98,47	1,75	12,183
...	Escola	97,32	301	98,12	2,42	-0,331
...	Escola	114,6	412	98,00	2,10	7,905
...	Escola	107,23	337	95,21	2,34	5,137
...	Escola	98,63	91	95,33	4,26	0,775
...	Escola	100,93	486	95,87	1,91	2,649
...	Escola	93,93	393	99,78	2,12	-2,759
...	Escola	97,92	458	101,07	2,00	-1,575
...	Escola	97,69	617	97,33	1,71	0,211
...	Escola	116,36	544	96,44	1,88	10,596

Comparative analisys: 2012 results versus 2011 results

		2012				2011			
School/ aggregated schools code	School/ aggregated schools denomination	National mean value, weighted by the distribution of tests performed in the school (A)	Total Number of tests performed	School's average score in the national assessments (B)	Standard deviation of the average score	National mean value, weighted by the distribution of tests performed in the school (C)	Total Number of tests performed	School's average score in the national assessments (D)	Standard deviation of the average score
---	Escola ...	93,35	1109	108,66	1,28	103,08	1565	100,65	1,02
---	Escola ...	91,51	493	108,92	1,88	102,97	714	101,46	1,53
---	Escola ...	90,58	82	86,46	4,45	100,64	142	87,78	3,42
---	Escola ...	98,26	474	106,94	2,02	103,41	865	104,45	1,41
---	Escola ...	100,26	112	83,11	3,92	104,68	228	94,48	2,67
---	Escola ...	94,97	345	104,52	2,38	105,64	649	106,32	1,55

2012 versus 2011				
Difference between the two reference means (E=A-C)	Standard deviation of the difference between the two average scores	Difference between the two average scores (F=B-D)	Mean improve (F-E)	Statistical signifacance indicator
-9,73	1,64	8,01	17,74	10,82
-11,46	2,42	7,46	18,93	7,81
-10,05	5,61	-1,32	8,73	1,56
-5,15	2,46	2,48	7,63	3,1
-4,42	4,74	-11,37	-6,95	-1,46
-10,67	2,84	-1,8	8,87	3,12

Thank you for your attention!