Analysis of student performance using statistical techniques

Pedro Henrique Alves Barros
State University of Rio de Janeiro – UERJ
E-mail: pedro.barros@pos.ime.uerj.br

Regina Serrão Lanzillotti
State University of Rio de Janeiro – UERJ
E-mail: reginalanzillotti@ime.uerj.br

ABSTRACT
In the Public School System, the Mathematics teacher from the 6th to the 9th grade is faced with the abyss between pedagogical planning and learning. Evaluations, plus bonuses for complementary activities, may be extracurricular excesses, as they generate unrealistic results in the evaluation of teaching. Descriptive statistical treatment was chosen according to histograms, position and dispersion measurements, asymmetry coefficients, short coefficients and Pearson's correlation coefficients. In inferential statistics, ANOVA was used to compare the means. The descriptive statistical evaluation indicated high variability in the results of the written evaluations, and the correlation coefficient between the written evaluations corresponded to 0.67, which indicates a small association. The comparison of these averages showed a significant difference between the Mathematics averages in the written and extra-class assessments, which shows the significant difference in the construction of the assessments of structured curricular knowledge in Mathematics and the extra-class assessment. It should be suggested that the personalized assessment would be more coherent with the student's profile due to the heterogeneity observed in the evaluations inherent to the concepts of Mathematics that will be the foundation in the daily life of the student body. To combat the discouragement generated by school failure, it should be thought that formative assessments are possible to provide students with greater feedback, directing them to structured knowledge that is appropriate to the demands of their daily lives.

Keywords: Position and Dispersion Measurements, Correlation, ANOVA.

1 INTRODUCTION

The teacher of Mathematics in the Public School System, when working in the 6th to 9th grade, is faced with the abyss between the current school pedagogical planning and structured learning, many times, there is a disparity between what is planned and what is carried out. Student evaluations, when added to bonuses for activities complementary to classroom activities, can be considered true extracurricular excesses, as they often come by masking knowledge gaps, generating unrealistic results for the education system.

In the Public School System, the Mathematics teacher from the 6th to the 9th grade is faced with the abyss between pedagogical planning and learning. Evaluations, plus bonuses for complementary activities, may be extracurricular excesses, as they generate unrealistic results in the evaluation of teaching. It is vital to help a teacher feel good about himself and that his performance is not minimized, as it is important to maintain emotional balance and psychological resilience to live with the stress generated by the current building and educational conditions, avoiding the psychosomatic effects. In this domain, it is essential that the pedagogical processes are in an adequate position to contribute to functional learning, that is, the skills (activities) that encompass self-care, hygiene habits, school attendance, commitment to tasks and interpersonal interaction.

It is noteworthy that the importance of teaching focused on the themes of the National Common Curricular Base (BRASIL, 2018) should be recognized, with regard to universality, plurality, and fluidity among disciplines. It is noteworthy that despite so many efforts, learning problems, especially in Mathematics, persist over the years. The experience in Youth and Adult Education as well as in High School in the basic operations of Fundamental Mathematics, allows us to perceive the extreme difficulty in the algorithm referring to division. It is worth reflecting on the importance of its use in financial planning, especially in the notions of family budgeting, home economics, in which the mastery of at least the four arithmetic operations are fundamental. The abyss between the political pedagogical project and curricular learning masks the problem, since extracurricular activities generate bonuses for basic curricular subjects, often without any connection in relation to the teaching of Mathematics, because the bonus is due only to the student participating in the activities, exemplified by the gang of the June festivals without any correlation with the concepts of fundamental measures referring to the time of high school of the performances during the rehearsals of the quadrille and the geometric shapes in the operational of the dance.

This reflection motivates the use of statistical modeling in the evaluations applied in the first quarter of 2023 referring to one of the 8th grade classes in a municipal school, which provides the confrontation of performance before and after bonuses in order to outline strategies that can improve student performance contributing to school management, mitigating possible deficiencies of students, This can support future innovative pedagogical and curricular strategies to improve the teaching and learning process. In this teaching segment, the evaluation takes place in a triple way under the criterion of three learning assessments, hereinafter AV1, AV2 and AV3, where each one has a score from zero to ten, but each of them has its own proposal and approach. AV1 is directed to extracurricular activities that directly or indirectly involve the concepts portrayed in the classroom, unlike AV2 and AV3, which are restricted to the contents taught in class. The student's final grade is the sum of the three assessments.

2 OBJECTIVE

Propose a method that seeks to analyze student performance to contribute to fair evaluation in the Elementary School cycle.

Within this scope, we seek to meet the following specific objectives:

a) to analyze the students' performance according to statistical modeling of AV2 and AV3 assessments.

b) To compare the results of the AV2 and AV3 evaluations with the result obtained in the AV1 evaluation, a simultaneous diagnostic tool for the teacher and the institution regarding the ongoing teaching and learning process.

3 METHODOLOGY

Initially, a literature review was carried out with articles published in the period from 2017 to 2021 in the electronic databases: *Scientific Electronic Library Online* - Scielo and Google Scholar, using the descriptors: Elementary Education and Legislation of the Department of Education of the Municipality of Maricá. Articles published in journals that presented *Digital Object Identifiers* that dealt with the subject, available online, were included.

4 DEVELOPMENT

The descriptive statistical treatment was performed through histograms, which allowed a visualization of the students' profile in relation to learning. Position measures such as mean, median, and mode were used as a way to evaluate the heterogeneity of the results, and in a symmetrical distribution there is equality of position measures. (CORREA, 2002). In the positive asymmetric distribution, the mean value is greater than the mode, ~~and~~ in the negative distribution, the mode is greater than the mean. One of the measures of dispersion corresponds to the coefficient of variation (CV), obtained by the ratio between the standard deviation and the arithmetic mean, and the standard deviation is the square root of the variance.

$$CV = \frac{S}{\bar{x}}, \qquad (1)$$

where S = standard deviation

$\bar{x}$ = arithmetic mean.

CV values close to zero are considered low dispersion, as one approaches, there is an expressiveness of the variation around the mean value.

The kurtosis percentile coefficient indicates the degree of flattening or elongation of the histogram frequency polygon, according to the ratio between the mean distance from the distance of the central percentiles, Percentiles 75 and 25 (Quartile 3 and Quartile 1), to the distance between the percentiles considered extreme, Percentile 90 and Percentile 10, and which indicates the concentration in the lower and upper tails" (NAGHETTINI and PINTO, 2007). The equation is expressed as follows:

$$K = \frac{(Q_3 - Q_1)/2}{(P_{90} - P_{10})}, \text{ where } (2)$$

$Q_3, 3^{º}\, quartil,\ Q_1, 1^{º}\, quartil, \qquad P_{90}, percentil\ 90, \qquad P_{10}, percentil\ 10.$

In the Normal Distribution, the kurtosis coefficient corresponds to 0.263, when it is classified as mesocurtic, if the value is higher, the distribution will be designated as platicuritic, if not, leptocurtic.

The sample stratification process is indicated when there is expressiveness of the coefficient of variation, asymmetry and kurtosis deviates from the mesocurtic classification. In this process, the guidelines of (CÂMARA, 1965) were followed, which suggests arranging the data in ascending order and dividing the total of the variable in haste, according to the number of strata to be established, but suggests that it should not exceed five strata. Subsequently, to ensure the adequacy of the stratification, obtain the weighted mean of the variances of the strata and divide it by the total variance, generating the T. Newmann coefficient. If this coefficient is less than 0.15, the stratification can be considered adequate, otherwise another stratum is created and the process is redone until a value close to this coefficient is reached.

The association between two variables in the ratio scale can be obtained through Pearson's correlation coefficient, which is a dimensionless index that also indicates the direction and intensity of the association, obtained according to the expression:

$$Pearson = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 (\sum_i (y_i - \bar{y})^2}} \qquad (3)$$

The comparison of the means of two or more groups can be treated according to Analysis of Variance - ANOVA (NEWBOLD *et al.*,2013), which compares the equality between them, according to the Variance Within and Variance Between and the Total Variance. In this study, single-factor ANOVA was used, as it was intended to evaluate the means for only one category and it was assumed that there would be independence between the samples. It is considered to be the overall mean of the sample, x$\bar{x}_{Ij}$ the j-th observation of group i, where i = 1 to k. Snedocor's F-test statistic corresponds to the ratio between the estimates of Variance Between (MQB) and Variance Within Variance (MQW), whose mathematical expressions are as follows.

$$SQB = \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2 \qquad (4)$$

$$SQW = \sum_{i=1}^{K} \sum_{j=1}^{n_1} (x_{ij} - \bar{x}_i)^2 \qquad (5)$$

$$SQT = \sum_{i=1}^{K} \sum_{j=1}^{n_1} (x_{ij} - \bar{x})^2 \qquad (6)$$

$$SQT = SQB + SQW \qquad (7)$$

These statistics are shown in the ANOVA Table. The F-test statistic is used under the following hypotheses:

Null hypothesis (H0, statistical equality between the means of the groups.

Alternative hypothesis (H1), at least from the population averages may be different. Use the Fisher-Snedocor F function, with a pre-established significance level (α) (e.g., α = 5%). H_0 is rejected if $_{F\ calculated}$ > tabulated $F_{,\ k-1,\ n-k,\ \alpha}$.

Table 1: ANOVA.

Source of variation	Quadratic Sum	Degrees of freedom	Variances (Mean Square)	Reason, F-Senedocor (calculated)
Between groups	SQB	k – 1	Variance between, MQB = $\frac{SQB}{k-1}$	$F_{(k-1,\ n-k)\ gl}$ $\frac{MQB}{MQW}$
Within the group	SQW	N – K	Variance within, MQW = $\frac{SQW}{n-k}$	
Total	SQT	n – 1		

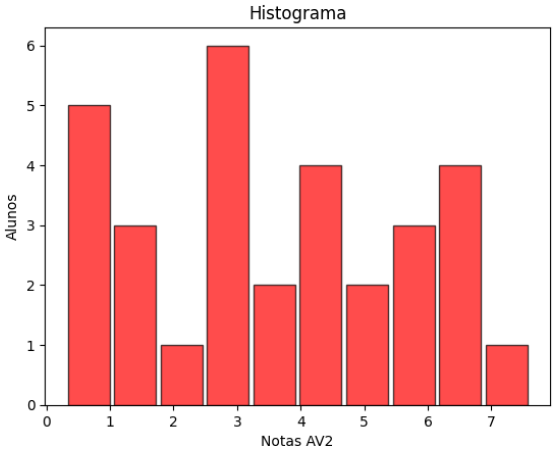
5 RESULTS AND DISCUSSIONS

In this study, the statistical treatment that investigates the performance of students in the AV2 and AV3 evaluations was observed, with scores on the ratio scale from 0 to 10. The AV1 evaluation score has a punctual attribution on the same scale, but in addition to measuring all the syllabus corresponding to the quarter, it includes extracurricular activities such as event organization, Festa Junina and interdisciplinary activities such as Cultural Exhibitions on various topics pertinent to Inclusion and Combating Bullying. It is noteworthy that the AV1 assessment includes a differentiated approach and it has been noticed that its values are differentiated in relation to AV2 and AV3, which often becomes a resource for the outcome of non-failure.

Chronologically, the AV2 evaluation is the first of the three, and was analyzed in a descriptive manner and by the graphical visualization of the histogram, Figure 1. It was considered as a sample analysis, since not all students were considered. The histogram shows the heterogeneity of the VA 2 scores, emphasized by the position measures that took over the statistics: mean 3.61, median 3.50, mode 3.00, the distribution is asymmetrical to the right, indicating concentration for the lower values. The coefficient of variation, a measure of dispersion obtained by dividing the standard deviation 2.13 in relation to the mean value, indicated a value of 0.59, dispersion of the scores in relation to the mean value. The percentile kurtosis

0.345 was higher than 0.262, a statistic referring to the Standard Normal Curve, which indicates a platycurtic configuration, emphasizing the dispersion of the notes.

Figure 1: Histogram of AV2 assessment.



Once the evaluation scores indicated significant dispersion, the stratification method suggested by Lourival Câmara2 was adopted to identify homogeneous groups of students in relation to the learning of Mathematics concepts. The values referring to the scores are ranked in descending order, initially adopting five strata, which did not meet Newman's stratification criterion, which should have a stratification coefficient of 14.00%, corresponding to the ratio of the weighted mean of the variances of the strata in relation to the total variance. Thus, it was decided to increase the number of groups until this relative threshold of 14.00% was reached. Similarly, this procedure was performed for the AV3 evaluation, Table 2.

Table 2: Validation of the stratification of AV2 and AV3 evaluations using the T. Newman criterion.

No. of strata	AV2	AV3
	Coefficient T. Newman	T. Newman coefficient
5	1,33	0,427
6	0,95	0,377
7	0,56	0,333
8	0,14	0,257
9	-	0,236
10	-	0,190
11	-	0,169
12	-	0,158

13	-	0,140

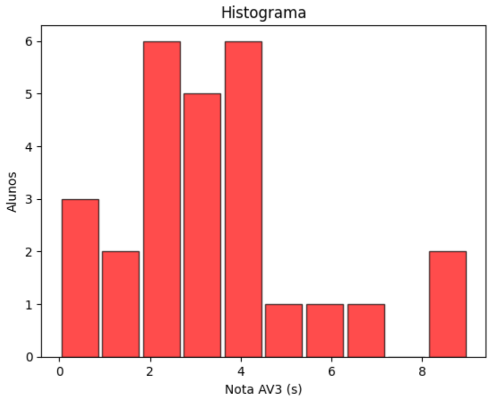
The distribution of the AV2 grades of the 31 students arranged in eight groups in descending order of knowledge measurement, Table 3 showed that most groups presented variability around the means considered low, since the coefficients of variability were between 0.12 and 0.04, aggregating students with similar learning profiles within each group.

Table 3: Descriptive statistics of the results of the AV2 evaluation, according to stratification.

Statistics	Strata (Groups)							
	1	2	3	4	5	6	7	8
Average	0,76	1,68	2,74	3,82	4,73	5,60	6,30	6,90
Variance	0,11	0,15	0,11	0,22	0,10	0,13	0,07	0,28
Standard deviation	0,34	0,39	0,34	0,46	0,32	0,36	0,26	0,53
Coefficient of Variation	0,44	0,23	0,12	0,12	0,07	0,06	0,04	0,08
Students	5	5	5	4	3	3	3	3

Similarly, in the AV3 evaluation, the objective is to evaluate the learning of polynomials, and the grades allowed the construction of the histogram, Figure 2, which allowed to obtain the measures of position, mean, median that assumed values, 3.24, 3.00, being discriminated two modes, 2.00 and 4.00, indicating positive asymmetry, that is, most students had difficulties in learning this syllabus. The dispersion measurements showed a standard deviation of 2.30, resulting in a coefficient of variation of 0.71. Percentile kurtosis of 0.159, which indicates a leptocurtic distribution, high concentration in low modal values.

Figure 2: Histogram of AV3 assessment.



Similarly to Table 3, Table 4 was constructed, which indicates the high heterogeneity of the results of AV3 within groups, but the group with lower acquisition of the concepts to be evaluated presented greater variability, with a coefficient of variation of 1.73.

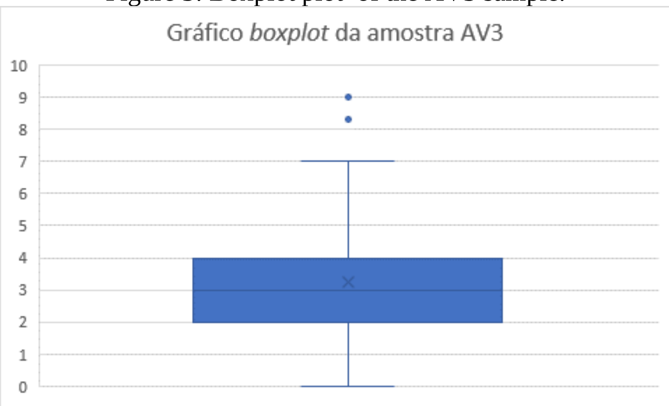
Table 4: Descriptive statistics of the results of the AV3 evaluation, according to stratification.

Statistics	Strata (Groups)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Average	0,00	0,33	1,83	2,3	2,5	3	3,05	3,65	3,90	4	4,5	6,5	8,65
Variance	0,00	0,33	0,08	0,02	0	0	0,01	0,05	0,02	0	0,5	0,5	0,25
Standard deviation	0,00	0,58	0,29	0,14	0	0	0,07	0,21	0,14	0	0,71	0,71	0,50
Coefficient of Variation	-	1,73	0,16	0,06	0	0	0,02	0,06	0,04	0	0,16	0,11	~~0,06~~
Sample size	2	3	3	2	2	2	2	2	2	2	2	2	2

Other The analysis consisted of obtaining the degree of correlation between AV2 and AV3 assessments, since there is an obligation to launch the concept, regardless of whether or not the student has made the 2nd call. Some students were absent in at least one of the evaluations mentioned, because they did not take the test, which resulted in a score of zero in the calculation of the average, and in the correlation, students who missed the 2nd call were not included. It is imagined that the lack of interest in replacing one of the lost evaluations, or because of the illusory performance in the last evaluation, AV1.

The Chart *Boxplot* of the AV3 evaluation, Figure 3, indicated two *outliers* that correspond to differentiated students, which stimulated the withdrawal of the AV2 and AV3 evaluations for the calculation of Pearson's correlation, which presented a value of 0.67, which indicates a moderate positive correlation. If AV2 assessment is increasing, VA3 is likely to be as well, although Figure 3 shows that learning is falling short.

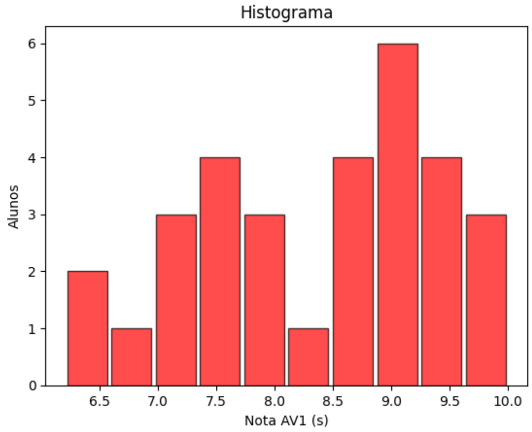
Figure 3: Boxplot plot of the AV3 sample.



Subsequently, ANOVA was applied in order to evaluate the Null Hypothesis of the statistical equality of the means of the AV2 and AV3 evaluations, values of 3.61 and 3.23, establishing a significance level of 5%. The F-test statistic$_{Calculated}$ was 1.34, which confronted with the critical value, $F_{critical}$, 4.03, validates the null hypothesis.

In AV1 there was no absence, but in this case the extracurricular activities related to folkloric cultural activities are aggregated and the position measures assumed values of 8.37, 8.50 and 9.00 for the mean, mode and median, which indicates negative asymmetry, which may explain the significant difference in relation to the performance in the activities of the curriculum that has been minimized in Mathematics. The coefficient of variation assumed the value of 0.13, indicating low variability around the mean, although kurtosis 0.33 is slightly different from the Standard Normal Probability Distribution, 0.262. In the histogram of this evaluation, figure 4, it is possible to identify three groups, scores lower than 6.5, from this value to 8.0 and above this limit.

Figure 4: Histogram of AV1 assessment.



Pedagogically, the AV3 evaluation presented a more extensive evaluative content than in the AV2, as it encompassed the subjects of the previous evaluation. The coefficient of variation indicated that AV3 has greater variability around the mean, which was confirmed in the stratification into groups of students. Due to these characteristics, AV3 was selected instead of AV2 to be compared with AV1 using ANOVA, adopting a significance level of 5%, and the means were 8.52 and 2.82 for AV1 and AV3, respectively. The objective of this study was to verify whether statistically these evaluations could be considered with equal means, and those who were absent from the AV3 evaluation were excluded. The ANOVA lower than the calculated test statistic $F_{Snedocor}$ at 194.1, which compared with the critical value $F_{critical}$, 4.03, rejects the Null Hypothesis of equality between the means.

6 FINAL CONSIDERATIONS

Students present discrepant evaluative effects that result from the inclusion of the AV1 evaluation in the average to define the progression of school performance, so it is necessary to review this criterion so that it can not hinder the learning process in High School, as there may be effects that compete with harmful impacts for both teachers and students.

Statistical analysis proved to be an excellent tool for analyzing school performance, providing important information about the first trimester of the eighth-grade class of the municipal public network. In general, many schools have management staff and pedagogues, but the quantitative analysis does not deserve attention, being restricted to the pedagogical processes in the scenario of Elementary School II. It has been noticed that the students have advanced in the serial curricular context, but maintaining the learning gaps inherent to the knowledge of previous years. On the other hand, many students use extracurricular assessments, designated as events, as a way to recover their learning deficiencies, masking both the school and the student for the learning deficit that accumulates. Aware of this problem, the school seeks pedagogical strategies to combat the learning deficits recorded in various assessments, especially in Mathematics, because it uses complementary activities to teaching, which in theory aim to contemplate and value forms of extracurricular learning, respecting the curricular framework pertinent to each stage of teaching.

The ANOVA indicates that the averages referring to the knowledge of the Mathematics concepts required in AV3 are in statistical inequality with AV1, since the means were 3.24 and 8.37, with coefficients of variation 0.71 and 0.13, in that order. These statistics indicate a significant difference in the construction of these assessments, in which it is perceived that the curricular knowledge in Mathematics is far below the complementary activities.

The written evaluations referring to the knowledge of the mathematical concepts required in AV2 and AV3 have statistical equality revealed by the ANOVA in which the mean values were 3.61 and 3.23, with coefficients of variation 0.59 and 0.71, respectively, and two students were considered differentiated from the others in AV3. It was also observed that the class profile indicates 13 groups when T. Newmann's stratification was used, with eleven strata with two students and two strata with three students. This scenario emphasizes the diversity of the student group, making it difficult to plan activities in class, which indicates that for a better evaluation, the student should be the center of the evaluation activity and not only the content addressed, as it suggests a tendency towards the elaboration of personalized evaluations. Low performance, when theoretical knowledge is required to the detriment of evaluations that include extracurricular activities, which led to high mean values with low variations.

It should be noted that the personalized assessment would be more coherent with the differences of each student, combating the discouragement generated by the so-called "school failure". Another way to combat discouragement would be the use of the so-called formative assessments, in which the student is continuously asked to do tasks of items of the mathematical knowledge approach, being possible to provide the student with a larger *feedback*, directing them towards the acquisition of structured knowledge. As Perrenoud (2003) mentions, it would be interesting for the school to promote strategies that will improve the students' attendance, preventing them from sometimes being absent from the assessments. At the same time, the school could rethink its educational mission, investing in the training of teachers in order to produce creative and diversified assessments, with classes focused on the daily needs of students, including paying attention to the regionality of the school's location and not only on the content.

REFERENCES

BRASIL. Ministério da Educação. Base Nacional Comum Curricular. Brasília, 2018

CORREA, Sonia Maria Barros Barbosa, "Probabilidade e estatística", 2ª ed., PUC Minas Virtual, Belo Horizonte, 2003, 116 p.

CÂMARA, Lourival. Tecnologia da amostragem, Rio de Janeiro, ENCE, 1968. Descrição física: 2v.

NAGHETTINI, M.; PINTO, E. J. A. Hidrologia Estatística. Belo Horizonte: CPRM, 2007. 552 p.

NEWBOLD, Paul; CARLSON, William; THORNE, Betty. "Statistics for Business and Economics", 8º ed, Pearson, London, 2013, 792p.

PERRENOUD, Philippe. "Os ciclos de aprendizagem: Um caminho para combater o fracasso escolar". Porto Alegre: Artmed, 2003[2002].