

# International Journal of Human Sciences Research

## A STATISTICAL STUDY ABOUT FACTORS THAT CAN INFLUENCE THE RATE OF CORRECT ANSWERS IN THE ENEM

---

*Francisco Moisés Cândido de Medeiros*

PhD in Statistics from the  
"Universidade de São Paulo" – USP

*Ivone da Silva Salsa*

PhD in Education from the "Universidade  
Federal do Rio Grande do Norte" - UFRN

*Ivonaldo Silvestre da Silva Junior*

Graduating in Statistics from the  
"Universidade Federal do Rio  
Grande do Norte" – UFRN

All content in this magazine is licensed under a Creative Commons Attribution License. Attribution-Non-Commercial-Non-Derivatives 4.0 International (CC BY-NC-ND 4.0).



**Abstract:** The National Secondary Education Examination – ENEM was created in 1998 with the purpose of evaluating the performance of graduates from the last stage of basic education in Brazil. However, today, this exam has taken on a relevant position as it has become the main gateway to undergraduate courses for the overwhelming majority of Brazilian universities. The initial hypothesis – seed of this debate – is that some statistical variables can influence the rate of correct answers in the ENEM, especially the variables linked to socioeconomic conditions. In this research, we seek to identify these variables, considering only the percentages of correct answers in two areas of knowledge: Languages, Codes and their Technologies, and Mathematics and their Technologies. Furthermore, it is important to point out that the participants of this investigation are those candidates who managed to enter the “Universidade Federal do Rio Grande do Norte” - UFRN in 2018 through ENEM. The discussions woven here were supported by concepts inherent to Descriptive Statistics, measures of Central Tendency and the graphical resources of the Histogram and the Box Diagram. Among the variables studied, it was found that “father’s education”; “mother’s education”; “Internet availability” and “family income” influenced the ENEM correct answers.

**Keywords:** ENEM; Percentage of correct answers; Variables.

## INTRODUCTION

The data exposed in this study were, originally, presented at the VII National Congress of Education – CONEDU; now, after new reflections have been added to the debate around these data, they now come back to the fore to deepen ideas about the performance of candidates for the ENEM, when this performance is scrutinized from the perspective of certain existing conditions

that can affect them.

ENEM and Secondary Education are two actors with a relevant role in the Brazilian educational scenario, because, for the immense legion of those who complete this level of education, ENEM constitutes what represents the path that gives access to the transcendent dream called UNIVERSITY! However, unfortunately, there are countless Brazilians – especially those who live in precarious socioeconomic conditions – who are unable to fulfill this dream! Society is structured in such a way that it condemns thousands who aspire to access Higher Education to exclusion (Bourdieu, 1999).

On the other hand, the quality teaching that paves the way for success in the ENEM does exist, in a select group of private schools whose monthly fees are unfeasible for students from the most disadvantaged social classes; that is, a privilege intended for students from wealthy families. This brings to light the social inequality that punishes thousands of candidates participating in the ENEM, depriving them of the chance of becoming a university student, as warned by Pierre Bourdieu (1999): the selection for Higher Education precedes the selective examination for this level of education. In reality, this selection takes place throughout the schooling course, based on the unequal opportunities that are configured mainly for subjects from the most disadvantaged social classes. “Now, opportunities for access to higher education are seen as the result of a direct or indirect selection that, throughout schooling, weighs with unequal rigor on subjects from different social classes” (BOURDIEU, 1999, p.41).

It is in this scenario that ENEM operates, presenting, at the end of each annual event, the results associated with the performance of each candidate; they are the bearers of the news that define the future of student life for thousands of high school graduates, as the

need to have a higher education diploma seems to be, today, a requirement through which greater chances are achieved in the market of work.

From this perspective, the school plays a crucial role in the student's life! However, in current times, is the teacher being prepared/trained to provide his student with an education compatible with the demands of society in the 21st century? How is the school situated when it is challenged to build skills, in these times, whose scenario seems so impregnated with thousands and thousands of information that flow on the internet, uninterruptedly, 24 hours a day?

It is true that, for many sectors of society, the central role of school education is given to teachers. As we recognize, teachers play a vital role in the success of pedagogical projects, but they alone cannot produce the expected transformations in the system. There are problems that originate outside the school, which need to be worked on by everyone (family, community, State, unions, etc.). Without this perspective, teachers can do little. (RAMALHO; BELTRÁN NUÑEZ; GAUTHIER, 2003, p. 91).

In such a situation, the university emerges as a way to improve the professional qualification of those who seek knowledge, understanding it as a currency of undeniable value. Perhaps for this very reason, with rare exceptions, graduates of basic education try to enter a university; which, in a way, explains the large turnout to ENEM, given that this exam holds the entry passport for undergraduate courses in almost all Brazilian universities (and even some in Portugal). This exam annually submits thousands of candidates – especially young people – and, by extension, their families, to distressing expectations in the face of a situation that, due to the absolute lack of places in higher education institutions, can bring, depending on certain courses – medicine is a them – frustrations for many

candidates. This is likely to happen because many candidates, despite being qualified, are not successful in undertaking the said exam.

## **METHODOLOGICAL PATH**

According to the thought of Cecília Minayo (1999) the methodology adopted in a scientific research must consider, among other aspects, the creativity of the researcher; this, in its essence, presupposes the experience, personal capacity and sensitivity inherent to those who invest in scientific knowledge. The methodology encompasses much more than a set of techniques used to improve the understanding of an investigated reality. It exposes, during the course of the research, how the articulation theory versus reality of the facts/data studied was woven.

This research aims to identify variables that may influence the rate of correct answers in the ENEM, more specifically, in the area of Languages, Codes and their Technologies, as well as in Mathematics and their Technologies, considering only the candidates who managed to enter the UFRN in 2018 through this exam. Therefore, this means that the data that underpinned the analyzes presented here were generated by the candidates who took the ENEM tests, in the previous year, that is, at the end of 2017. In such conditions, there were precisely 29,610 freshmen at UFRN; these constitute the target population of the present study. The database used in this research is formed by the set of microdata provided by the Observatório da Vida do Estudante Universitário (OVEU), these coming from the Núcleo Permanente de Concursos (COMPERVE) which provides, among other information, data on the profile of freshmen at UFRN. As of 2013, this database is the result of the union of microdata made available by the National Institute of Educational Studies and Research Anísio Teixeira (INEP), by the Superintendency of Informatics at UFRN

(SINFO) and by the SISUGestão system.

At first, important highlights of the profile of the freshmen involved in this work were presented. Next, a screening was carried out, considering the various variables that were selected according to the possible degree of influence they could have in relation to the rate of correct answers in the area of Languages, Codes and their Technologies, as well as in Mathematics, and its Technologies. This screening was the way to identify which variables must be considered to be researched in order to meet the objectives of this study. The consecutive stage – supported by the methodology of Descriptive Statistics – consisted of the organization, presentation and descriptive analysis of the data associated with these variables, seeking precisely to discover possible influences/relations or patterns of the same with regard to the index of correct answers of the candidate. The R software (R CORE TEAM, 2021) was used for reading, organizing the data and building the graphs and tables of this work.

## PROFILING CANDIDATES

In the scenario of variables associated with the profile of freshmen, the first variable to be addressed is Gender; the data representing this variable are shown in the bar chart shown in figure 1 below.

Observing the graph in this figure, it is easy to see that the majority (54%) of the entrants were female, against 46% male.

The second variable explored is the “age” of freshmen. Data from this variable were organized in a statistical graph called histogram, which is shown in figure 2. Such data clearly show the existence of a huge concentration of freshmen up to 20 years old; then, as age increases, inversely, it is noted that the number of newcomers decreases, although, along this decrease, there are some increase peaks (see figure 2).

Still with regard to the variable “entry age”, other relevant results can be obtained by examining Table 1, which brings up important statistical measures, such as, for example, the average age of these entrants: 22.18 years. In addition to this information, there are some order statistics (also called separators): the 1st quartile being equal to 18 years old, reveals that 25% of new entrants were aged between 15 (the lowest observed) and 18 years old; in addition, the median is equal to 20 years, which means that half of these new entrants were between 15 and 20 years old. It is also possible to state that 50% of the entrants were between 18 and 24 years old (those who are between the first quartile (18 years old) and the third quartile (24 years old)). Another piece of information from Table 1 is that the highest age observed was Age 73. Here are the values that support these findings:

Next, an important variable will be analyzed: the education level of both the father and the mother. With regard to the father’s education, it was found that, of the 29,610 freshmen participating in this study, 1,839 of them had no knowledge about this reality.

(Probably this portrays a social situation that happens, as a rule, in less favored social classes: children raised without the knowledge or without the presence of the father in the family nucleus). In view of this specificity, the remainder was considered, that is, the 27,771 freshmen who were aware of their father’s education.

Given this specificity, inherent to this variable, the associated data produced the results that are shown in the statistical graph in columns, shown in figure 3-A.

In Graph A of Figure 3, it is clearly seen that, for the variable “father’s education”, the most prominent category was “completed high school or incomplete higher education” (33.86%); after it, the category “incomplete primary education” with 17.21% was the

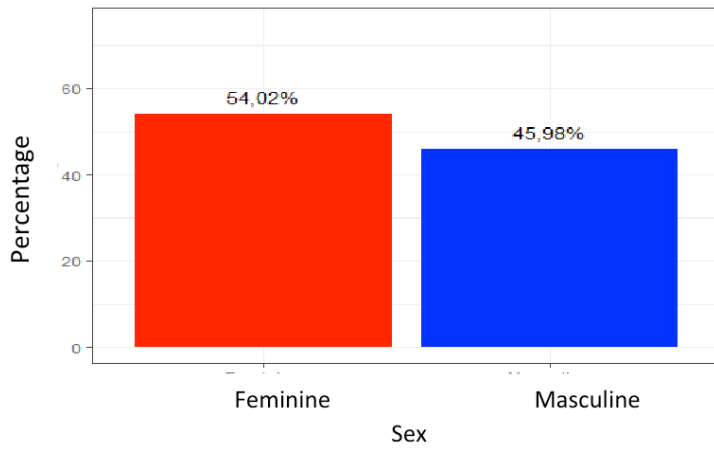


Figure 1: Bar graph of the percentage of freshmen according to gender.

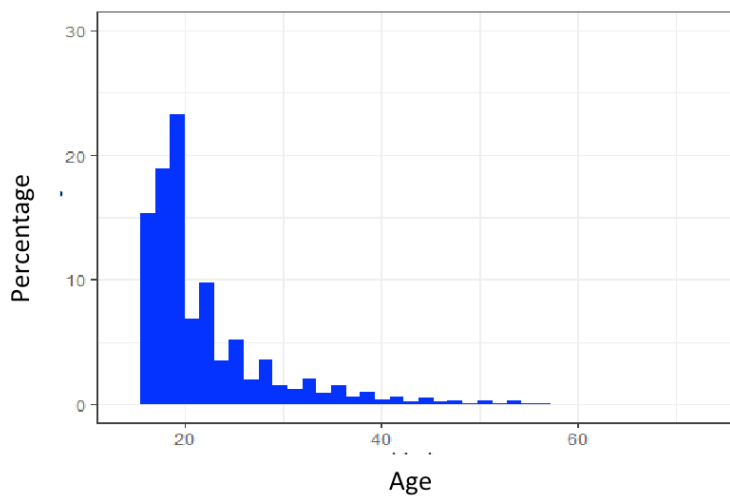


Figure 2: Histogram associated with the age of the freshman at UFRN.

Minimum	First quartile	Median	Third quartile	Maximum	Average	Standard deviation
15	18	20	24	73	22,18	6,57

Chart 1: Order, mean and standard deviation statistics for freshmen ages.

second most frequent. It can also be extracted that 53.85% of the freshmen, therefore, more than half, answered that the father has, at least, completed high school. (The value 53.85% is the result of the sum of the percentages 33.86%; 12.92% and 7.07%).

As for “mother’s education”, the number of freshmen who did not know about this issue was much smaller, compared to that on the father’s education; in the case of the mother, only 441 of the freshmen did not know the answer to this question. Similar to the specificity found in the previous question, subtracted from the total of 29,610 freshmen those 441 who were not aware of their mother’s education, leaving 29,169 freshmen; these generated the data in percentages shown in the statistical chart in columns shown in chart B of figure 3.

An analytical look at graph B makes it possible to perceive that the behavior of the variable “mother’s education” presents certain similarities in relation to the father’s education, previously discussed. For example, as in relation to the father, the greatest emphasis in the case of the mother’s education is associated with those who have completed high school or incomplete higher education (35.96%); in addition, more than half (62.92%) of the freshmen stated that their mother had completed at least high school. This last result shows a difference of almost 10% when compared with the result associated with the father. Another category to be highlighted is “never studied”; for the latter, when it comes to the father, the result is 4.45%; while, in the case of the mother, the percentage for this category was equal to 2.28%, therefore, almost half of that associated with the father. Another relevant result obtained from these data: the category “incomplete primary education” occupies second place in the occurrences associated with the father’s education, while, for the mother, in second place is “complete

higher education”.

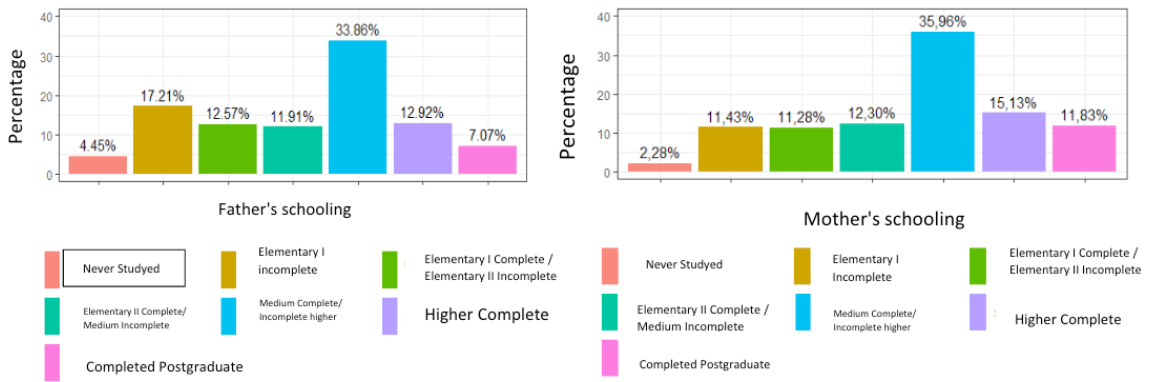
Another fact to be emphasized refers to the percentage associated with the “graduate” category: for the mother, the result was equal to 11.83%, which is markedly higher than that found for the father (7.07%) for this education level. In conclusion, the data displayed in the statistical graphs, A and B, of figure 3 bring to light an important observation regarding the behavior of the variable “education”: in general, the level of education of the newcomer’s mother was higher than that of the father.

In modern times, the internet is a very useful tool. For those who study and/or are always researching, undoubtedly, it is of great value. In this research, the variable “Internet availability at home” was contemplated, and the corresponding results are recorded in the column chart shown in Figure 4 shown below.

The message of the data is clearly perceptible just by looking at figure 4; he announces that a significant portion of newcomers, 82.33%, have internet access at home. This is undoubtedly an important tool, as it is capable of providing access to a vast collection of information which can be transformed into knowledge; this importance is especially important for those who are going to submit to any exam/competition, including the ENEM.

## **ANALYZING THE INFLUENCE OF CERTAIN VARIABLES ON THE BEHAVIOR OF THE RATE OF CORRECT ANSWERS IN THE ENEM**

As previously explained, the ENEM success rate, specifically in the area of Languages, Codes and its Technologies, as well as in the area of Mathematics and its Technologies is the focus that guides the steps in this research. However, it seems to be important to offer, at the outset, an overview regarding this rate of



A - Father's education level

B - Mother's education level

Figure 3: Level of education of parents of freshmen.

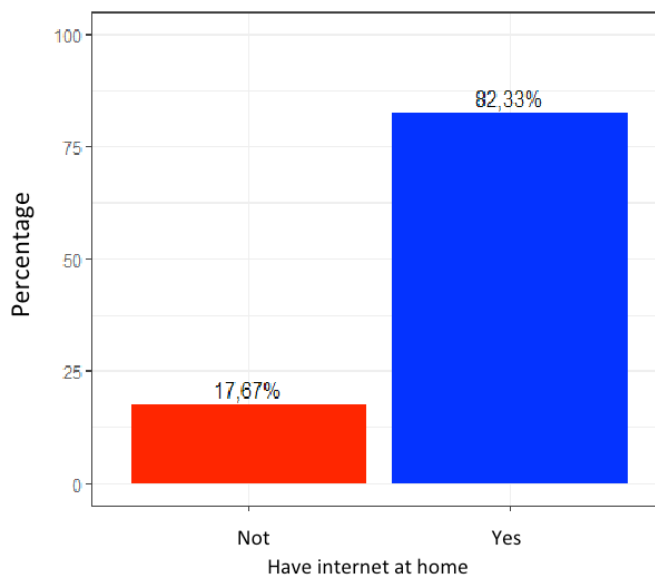


Figure 4: Percentage of new entrants in terms of internet availability at home.

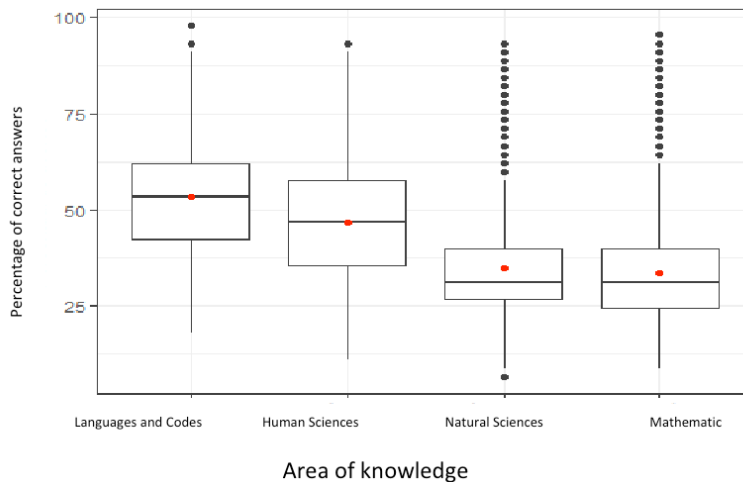


Figure 5: Box diagram of the percentage of correct answers in each area of knowledge

correct answers, also considering, distinctly, each of the other areas of knowledge explored in ENEM. This overview is presented through a statistical resource – the box diagram – shown in figure 5, presented below.

In this diagram, each box corresponds to an area of knowledge established in ENEM (there are four areas) and the average percentage of correct answers corresponding to each one of them is represented by the red dot inside the respective box.

Without requiring great effort, through just a more careful look at this statistical tool, it is possible to see how the data behaves; they show that, comparing all the areas, Languages and Codes has the highest average of correct answers, then, with a slightly lower average value, is the area of Human Sciences. On the other hand, when it comes to the areas of Natural Sciences and Mathematics, it can be clearly seen that these averages were much lower than those of the two areas of knowledge mentioned above, highlighting here that the lowest average among all of them are associated with Mathematics.

Another aspect worthy of note is that, precisely in the areas with the lowest success rates, namely: Natural Sciences and Mathematics, many occurrences of success rates well above the average of the respective data set emerged! This is the meaning of the black dots that are positioned, one after the other, at the top of the vertical line drawn from the middle, in each box; such points are called “outliers”: they precisely indicate scores that stand out in relation to the mean of their respective set, in this case, the highlight is for values above the mean because they are above it. The concepts that underlie the interpretation of these graphs can be apprehended in Bussab and Morettin (2006).

With regard to the results linked only to the success rates of two areas in the ENEM: Languages, Codes and their Technologies, as

well as Mathematics and their Technologies, which constitute the object of study of this work, Figure 6 shows two statistical graphs: the A-histogram and the B-histogram; they summarize these results separately. These graphs give news of the behavior of the data related to the referred indices. With a diligent look at these histograms, one can compare them, and, with this comparison, it is possible to notice that there is a significant difference in their shape (outline). This brings out the behavior of the percentage of correct answers, thus facilitating its analysis; From this one can obtain valuable findings, such as those commented below.

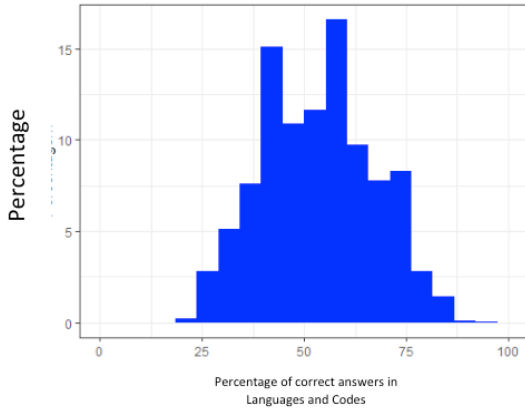
The highest concentration of hits for Languages and Codes orbits around 60%; in addition, extending to the left and right of this 60%, in a slightly symmetrical way, the other percentages of correctness extend approximately in the interval between 25% and 75%.

The percentage of correct answers referring to Mathematics is clearly lower than the percentage of correct answers for Languages and Codes: in the case of Mathematics, the highest concentration of correct answers is around 30%. Furthermore, the percentages of correct answers in Mathematics are distributed asymmetrically, spreading more to the right than to the left, reaching a little over 75%.

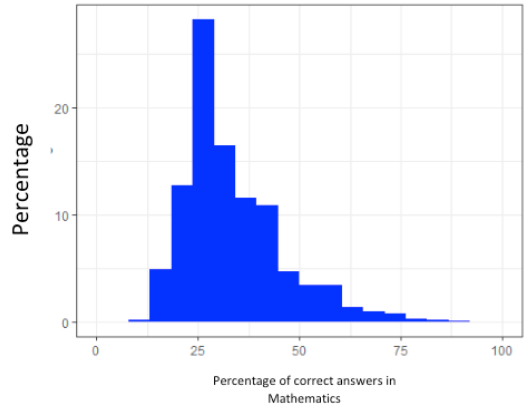
Next, four box diagrams will be presented; these show the behavior of the percentages of correct answers in Languages and Codes (graphs A), as well as those of correct answers in Mathematics (graphs B), in a scenario in which the variable “sex”, concomitantly, is considered. Said box diagrams are shown in figure 7 stamped below:

Examining the two box diagrams associated with the Languages and Codes area (diagrams A) shown in figure 7, it can be seen that the difference between male and female freshmen



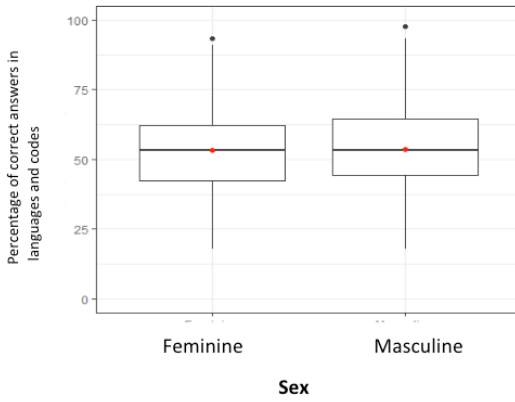


A - Languages and codes

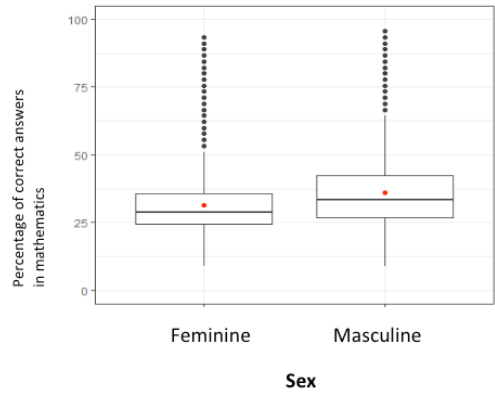


B - Mathematics

Figure 6 - Histograms of the percentage of correct answers by area of knowledge.

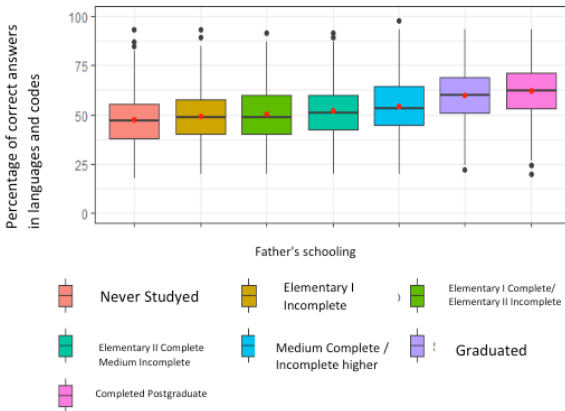


A - Language and codes

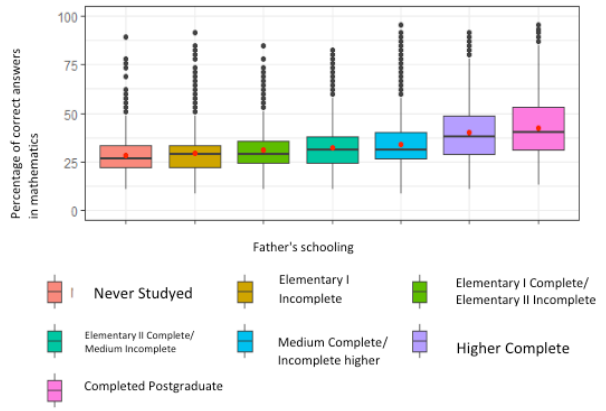


B - Mathematics

Figure 7: Box diagrams of the percentage of correct answers, by gender.

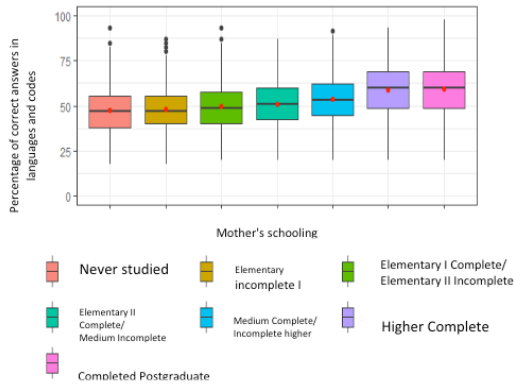


A - Language and Codes

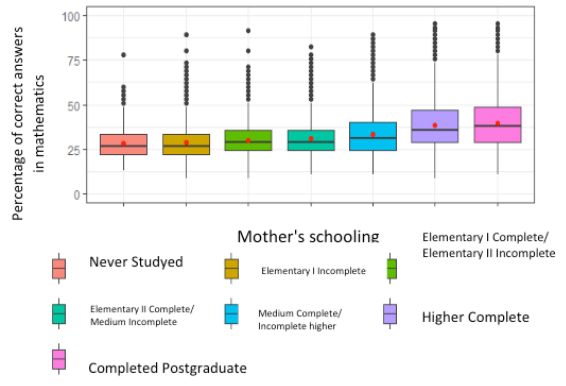


B - Mathematic

Figure 8: Box diagrams of the percentage of correct answers by father's education.

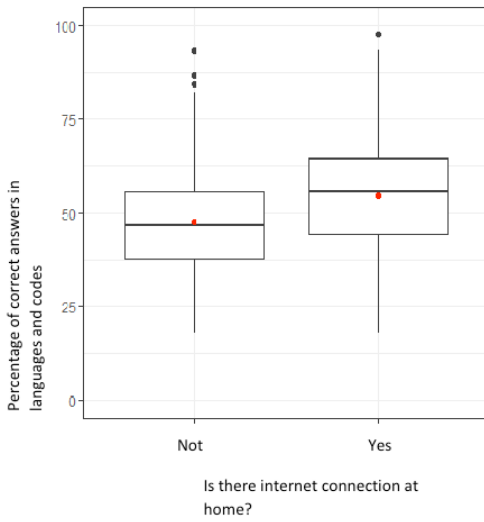


A - Language and Codes

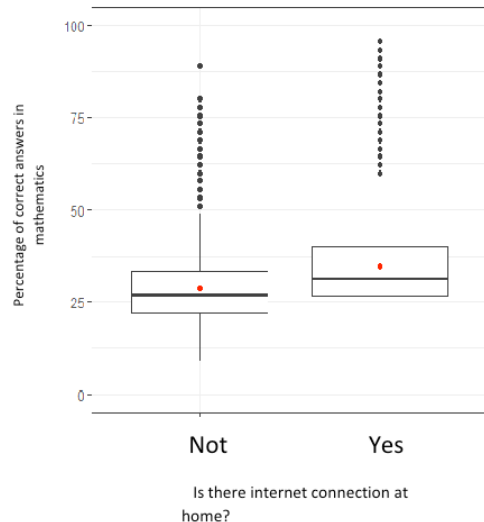


B - Mathematic

Figure 9: Box diagrams of the percentage of correct answers by mother's education.

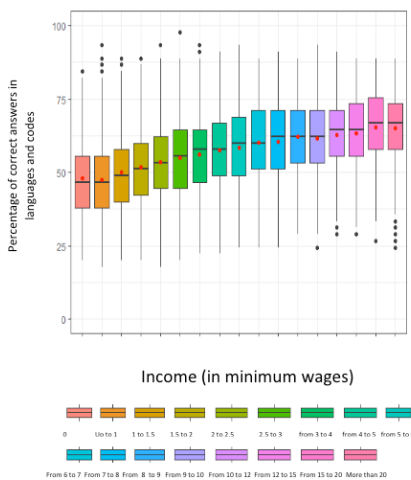


A - Language and Codes

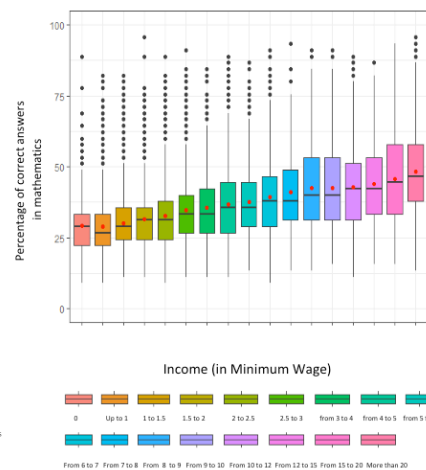


B - Mathematic

Figure 10 - Box diagrams of the percentage of correct answers considering having, or not, internet at home.



a) Language and codes



b) Mathematic

Figure 11: Box diagrams of the percentage of correct answers according to monthly income.

is really small, with their respective means (represented by the red dot) values that appear to match. This result suggests that the variable “gender” did not interfere in the percentage of correct answers in Languages and Codes.

With regard to the percentage of correct answers in Mathematics, shown in the two B diagrams shown on the right in Figure 7, a slight difference can be seen in the two distributions of these percentages: the “male” category obtained an average slightly higher than the “Female” category, (as can be attested by the position of the red dot). However, on the other hand, the “female” category had the highest number of correct answers that stood out in relation to the average (this is the reading based on the “outliers” that are above the average, represented by the black dots, located at the top, after the vertical line, above the mean). This fact is clearly perceptible when examining in figure 7, the two B diagrams, associated with correct answers in Mathematics.

Conclusively, the analysis of the rate of correct answers concomitantly with the variable “gender” suggests that this variable does not interfere significantly in the percentage of correct answers neither for Languages and Codes, nor for Mathematics.

There is consensus on the idea that quality basic education paves the way for success in the ENEM; however, not infrequently, the student coming from more disadvantaged social classes suffers from the indifference of the rulers regarding public policies for education, especially in the case of Basic Education. For this student, in addition to limited options in schools of excellence, his “Cultural Capital” from the perspective of Bourdieu (1999) also lacks a lot of knowledge that could help him in this marathon called ENEM.

Next, statistical data on the percentage of correct answers will be presented, using box diagrams to represent them. Such data

were analyzed by crossing them with certain relevant variables somehow linked to this concept of “Cultural Capital” constructed by Bourdieu.

First, an attempt was made to verify the existence of any influence of the variables “level of education of the father” and “level of education of the mother” on the percentage of correct answers of the freshman, for each of the areas: Language and its Codes, as well as in Mathematics, separately. The data associated with this crossing, “parents’ level of education” versus “percentage of correct answers” are represented in the graphs in figures 8 A and B (for “father’s level of education”), and, in figures 9 A and B (for “mother’s level of education”). Here they are:

Analyzing these box diagrams, it is perceptibly noted that there is a growing trend in the average percentage of correct answers, as the educational level of both the father and the mother grows. It can also be noted that this growth intensifies a little more, when the two highest levels of education are taken as a reference, namely: Higher Education and Postgraduate.

From the above, it can be concluded that the behavior of these data shows the existence of a trend, suggesting that this variable – the level of education – of both the father and the mother, interferes with the performance of newcomers in Language and its Codes, as well as, in math. It is worth highlighting the fact that the hit rate in Mathematics was always lower than in Language and its Codes, this occurred in the presence of the variable “father’s education”, and also, when analyzed concomitantly with “mother’s education”.

Now, this research tries to verify the influence of internet use in the percentage of correct answers in the ENEM. The internet is a tool that can be used on a large scale and, undeniably, a source of countless possibilities for accessing information. Although

information in itself does not translate into knowledge, it can pave the way to knowledge. The data analyzed here deal with the percentage of correct answers in a scenario of having or not having internet at home. Regarding this circumstance, figure 10 shows the behavior of the data for each of the two areas of knowledge in ENEM, discussed here, namely: Language and codes (Figure 10 A and Mathematics (figure 10 B). to verify whether the percentage of correct answers was affected by the variable associated with the internet, which took only two forms: either the entrant has internet at home, or he does not have this resource.

Without requiring much effort, examining the four diagrams shown in figure 10, it can be noted that the freshmen who had internet at home obtained better results than those who did not have this technological tool. This conclusion is visually perceptible and attested by comparing the averages (the red dot in each respective situation) in the two diagrams in figure 10 A that are associated with the area of Languages and their Codes, as well as in the two diagrams in figure 10 B whose data refer to Mathematics.

Therefore, having the internet resource at home is a factor that seems to influence results in the percentage of correct answers in the ENEM, in the two areas of knowledge: Languages and their Codes, and Mathematics. Probably, the non-availability of the internet is due to a lack of financial resources in the family; this type of situation highlights the possible difficulties that students face when they belong to an economically disadvantaged social class.

In this study, the last variable analyzed in the scenario of performances in ENEM tests, in the areas of Language and its Codes, and Mathematics was family income. Figure 11 shows the Box Diagrams constructed based on the percentage of correct answers

by income, considering these two areas of knowledge separately. By carefully observing these diagrams, one can understand that, as income grows, the means (the red dots) and medians (the horizontal line inside the “box”) also grow. Studies have shown that a per capita household income of more than 1.5 minimum wages, associated, *pari passu*, with the father’s higher education level are factors that strongly influence entry into higher education. (Carvalho and Waltenberg, 2015).

The data analyzed here suggest that the more economically favored classes have a greater chance of succeeding in the ENEM. Perhaps this is due to the fact that, as a rule, financially more affluent social classes tend to invest in situations that provide the components of the family nucleus with culture and education (investments in the cultural capital of the newcomer) through trips, various courses, school education in private schools that strive to offer better quality education, etc. All this can work as a lucrative savings whose income happens in a currency called knowledge! It is this savings, arising from cultural capital, which undoubtedly paves the way towards better conditions for success in assessments along the lines of ENEM.

These results suggest that, in a scenario of exams the size of the ENEM, the “gateway” seems more “open” to privileged candidates, coming from a more affluent social class whose accumulation of acquired knowledge provides them with greater cultural capital, than for those who did not have access to various knowledge opportunities. Families pass on to their children, throughout life, a certain cultural capital. (Bourdieu, 1999).

Today, the possibility of accessing a higher education course, through ENEM, is greater for students from social classes coming from municipal/state public schools who live in precarious financial situations, thanks to the inclusion policies adopted in

public universities and federal institutes of education, inaugurated in the government of Fernando Henrique and expanded in the Lula government. It is these inclusive policies that allow those who live in such adverse social conditions to feed and fulfill the dream of one day obtaining a university degree!

## FINAL CONSIDERATIONS

The ENEM success rates of UFRN freshmen in 2018, through this exam, constituted the raw material that embodied this research. More precisely, this study focused on the ENEM success rates, considering two areas of knowledge: Languages, Codes and their Technologies, and also Mathematics and their Technologies. The objective of the aforementioned research was to identify which

statistical variables, among those studied, could have an influence on the behavior of these indices. The initial suspicion was that the rate of correct answers on the ENEM could be influenced by certain variables. This suspicion was confirmed, because, among the variables investigated, it was found that “father’s education”; “mother’s education”; “Internet availability” and “household income” were variables that showed to have influence on the behavior of the ENEM test scores in the mentioned areas of knowledge.

The construction of such conclusions was woven under the methodological light of concepts inherent to Descriptive Statistics, highlighting here, measures of central tendency and graphic tools: histogram and box diagram.

## REFERENCES

BOURDIEU, P. **Escritos de Educação**. Organização M. A. Nogueira e A. Catani. Petrópolis, RJ: Vozes, 1999.

BUSSAB, Wilton de O.; MORETTIN, Pedro A. **Estatística Básica**. 5.ed. São Paulo: Saraiva, 2006.

CARVALHO, Márcia Marques de; WALTENBERG, Fábio D. Desigualdade de oportunidades no acesso ao ensino superior no Brasil: uma comparação entre 2003 e 2013. **Economia Aplicada**, [S.l.], v. 19, n. 2, p. 369-396, 2015. Disponível em: <https://doi.org/10.1590/1413-8050/ea124777>. Acesso em: 23jul. 2021.

MINAYO, Maria Cecília de Souza. **O desafio do conhecimento: Pesquisa Qualitativa em Saúde**. 9. ed. São Paulo: Hucitec, 2006.

R CORE TEAM. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. 2021.

RAMALHO, Betânia Leite; BELTRÁN NUÑEZ, Isauro; GAUTHIER, Clermont. **Formar o professor – profissionalizar o ensino: perspectivas e desafios**. Porto Alegre: Sulina, 2003.