Educação Pública
Seção Divulgação Científica e Ensino de Ciências

Statistical analysis of socio-demographic factors as predictors of performance in a Brazilian science fair

Análise estatística de fatores sociodemográficos como preditores de desempenho em uma feira de ciências brasileira

Daniel Verdi do Amarante
ORCID: 0000-0002-2700-3135

Kritim Rijal
ORCID: 0000-0001-7947-6711

Resumo:

Feiras de ciências oferecem uma oportunidade para estudantes compartilharem seus projetos científicos e praticarem diversas habilidades. Além disso, o julgamento e a premiação são partes importantes do processo, fornecendo retorno e reconhecimento aos participantes. Para manter a imparcialidade na competição e evitar reforçar desigualdades sistêmicas, é crucial examinar o cenário atual das feiras de ciências. Este artigo analisa o impacto de fatores sociodemográficos no desempenho discente na Feira Brasileira de Jovens Cientistas. Focando em gênero, raça, nível de inglês, tipo de escola e localização, modelos de regressão linear e logística foram utilizados para prever as notas e os prêmios dos participantes. Os resultados sugerem que o tipo de escola e o nível de inglês, características associadas ao acesso a recursos, influenciam significativamente o sucesso na competição. Vale ressaltar que o único critério que não apresentou diferença significativa nas pontuações para essas variáveis foi o da ideia do projeto, indicando que pesquisas inovadoras, propostas por alunos menos privilegiados, podem ter sido prejudicadas pela falta de recursos para seu desenvolvimento. Ao lidar com essas disparidades, organizadores de feiras e formuladores de políticas públicas podem promover oportunidades mais equitativas e incentivar novos cientistas, independentemente de sua origem.

Palavras-chave: Equidade. Premiação. Acesso a recursos.

Abstract:

Science fairs allow students to showcase their scientific endeavors and practice important skills. Additionally, judging and awarding are essential parts of the process, providing feedback and recognition to participants. It is crucial to examine the current landscape of science fairs and identify potential areas for improvement to maintain fairness in the competition and avoid reinforcing systemic inequalities. We explore this issue by analyzing the impact of socio-demographic factors on students' performance in the Brazilian Fair of Young Scientists. Focusing on gender, race, English level, school type, and location, we use linear and logistic regression models to predict students' scores and awards. The results suggest that school type and English level, features associated with access to resources, significantly influence success in the competition. Notably, the only criterion that did not show a significant difference in scores for these features was the one about the project idea, indicating that innovative research questions from underserved students may have been hindered by a lack of resources for project development. By addressing these disparities, science fair organizers and policymakers can foster more equitable opportunities and encourage the next generation of scientists, regardless of their background.

Keywords: *Equity. Awarding. Access to resources.*

1. Introduction

Science fairs are an opportunity for students, primarily from middle and high school, to explore their scientific interests. The fairs allow students to reach beyond the realm of their school curriculum and participate in the unique experience of something new on their own (Abernathy; Vineyard, 2001). As part of the process, students develop a research project, which provides them with early exposure to the scientific method. This approach of conducting a research project and presenting it at a science fair helps students develop key skills like communication, problem-solving, and critical thinking (Da Silva Gallon *et al*., 2019; Schmidt; Kelter, 2017).

The primary goal of the science fair is to pique students' interests in research and learning. The freedom to choose their projects, enabling a hands-on learning experience, is a particularly attractive factor of science fairs. However, students do not only learn from their own projects. By meeting participants from various regions and backgrounds, depending on the scope of the science fair, students receive exposure to different perspectives. Along with that, the reaffirmation of a job well done through awards and prizes encourages and inspires students to continue in the field of science (Magalhães; Massarani; Rocha, 2022). However, such inspiration can be hard to achieve if the underlying foundation of the fair is biased towards certain social groups. Thus, maintaining fairness in the competition is crucially important for its success.

Considering the importance of the topic, only a handful of recent studies have been done to address this issue. Those are mostly focused on the judges' perspective (Adams, 2011; Kiddell, 1987; Levin; Levin, 1991) or limited to analysis of specific characteristics, primarily gender (Mcdonnell, 1996; Sonnert; Sadler; Michaels, 2013; Steegh *et al*., 2019). For instance, Adams (2011) attempts to find if there is a statistically significant disparity between the scores of judges of the same or different discipline as the project. On the other hand, Sonnert, Sadler, and Michaels (2013) and Steegh *et al*. (2019), limit their analysis to the disparity of scores due to gender and not other factors like race or quality of education. While these studies are important, they do not demonstrate the full scope of the issue. In this paper, we go in-depth into the problem by exploring evidence of biases in students' performance at science fairs based on their socio-demographic background.

A few studies intended to investigate the important factors determining success in science fairs, but all of them were done in the 1990s and might not reflect the current landscape. Bellipanni (1994), for instance, investigated whether receiving awards at the International Science and Engineering Fair (ISEF) could be predicted from variables including resources and facilities. Similarly, Gifford and Wiygul (1992) tried to understand how much use of outside facilities and resources impacts the science fair awarding results. Czerniak (1996) used different variables, such as demographics, anxiety, and parental influences, as predictors of success in a district science fair in the United States.

Expanding on this research, we hypothesize that such biases on science fairs stem from the historical marginalization of certain social groups - related to gender or race, for instance - or the dependencies that projects demand in order to excel. Enabling a science project for a student requires appropriate mentors proficient in the field and, for some research fields, specific materials and facilities such as laboratories and computers. However, access to these resources may not always be distributed equally among all the participants. In most science fairs, there is no factor that acknowledges this discrepancy, which could then result in an unequal distribution in performance of the students of unprivileged backgrounds. Therefore, we use data from a national Brazilian science fair to answer our research question: Do participants of certain demographics have a palpable advantage going into the competition?

2. Methodology

In this section, we discuss the source and nature of our data, their privacy protection policy, and the analysis tools used. Our research methodology constituted of the following steps:

- Retrieving data from the repository of the Brazilian Fair of Young Scientists;
- Collecting data about social vulnerability and income per capita from governmental databases;
- Performing exploratory data analysis (proportion and average calculations);
- Creating statistical models that describe relationships between success-measuring variables (award, scores) and participants' socio-demographic factors;
- Interpretation of results from the statistical models.

2.1. Data

The analyses in this paper used three different data sources, with the crucial information coming from a Brazilian science fair and supporting data from governmental databases. Each of them is presented in detail below.

2.1.1. Science Fair Data

The datasets contain data for the three first editions (2020, 2021, and 2022) of the Brazilian Fair of Young Scientists (*Feira Brasileira de Jovens Cientistas – FBJC* in Portuguese), an online national science fair in Brazil. All participants consented to share their information with the institution and third parties for research purposes.

- One of the datasets consists of information about the participants (n = 1941), stored anonymously. The data includes event year and participation status (selected, not selected, and finalist), along with features self-reported by the participants on the fair's website in specific and standardized fields that enforced data validation for the authenticity of the information:

- Gender: options "Male," "Female," "Non-binary," and "Other";
- Race: options following the Brazilian Institute of Geography and Statistics conventions (IBGE, 2019);
- English level: options "Basic," "Intermediate," and "Advanced/Fluent";
- School: previously registered schools were displayed as options, and participants had the option to add a new one if needed.
- City and state: options pre-populated from a government's database (respective cities only displayed after selecting the state);

While the first dataset includes data for all students who submitted a project to the science fair, the two other ones are only for finalists (n = 1264). The second dataset records scores reported from the judges' analyses of the projects. All judges had research experience, and most were graduate students or university professors (with Master's or Doctorate degrees), and were accepted to judge at the event after going through a selection process. Scores ranged from 0 (non-existent/very poor) to 5 (excellent) and were given based on the event's standards, with specific and detailed criteria and sub-criteria, which are available on the fair's website.

The last dataset contains data about different awards provided to the projects (n = 531) by the events' coordinators. We only analyzed the awards that are considered the most prestigious: the 1st, 2nd, and 3rd places in each research field. The awards were distributed to the top 30% of the projects, with projects in the top 5% of the scores getting first place, the next 10% second, and the next 15% third place.

To improve the quality of our analysis, we excluded data points of students that belong to groups with n < 25. This resulted in a reduction in our dataset to n = 1841 students, with n = 1221 being confirmed finalists. We understand that this filtering may result in the loss of some important information, particularly of underrepresented students, such as non-binary people. However, since conducting analysis from small sample groups can bring inaccurate results, we opted for a more robust data modeling process, leading to more reliable conclusions. We strongly encourage other researchers to explore the students who are not represented in this analysis, as the goal is to make science fairs more inclusive.

2.1.2. Supporting datasets

Two additional datasets were used to aid the location-based analyses. One of them was obtained from the Continuous National Household Sample Survey (PNAD-Contínua) conducted by IBGE (2021). The data consists of the nominal monthly household per capita income of the resident population in 2021 for each Federation Unit (state) in Brazil. The second dataset is the Atlas of Social Vulnerability from IPEA (2019). The most relevant feature of this dataset for our purposes is the Social Vulnerability Index (SVI), which is an arithmetic mean of the indices of three dimensions: income and work, urban infrastructure, and human capital. Both datasets are publicly available. We decided to work with these two datasets because per capita income and SVI have been used in other studies as indicators of access to resources (Carmichael *et al*., 2019; Fekete, 2009; Nassif-Pires; Carvalho; Lederman Rawet, 2021; Noorderhaven *et al*., 2004; Posel; Rogan, 2009). We incorporated both indicators in our analysis to regulate their individual benefits and limitations: while we have SVI data from 2015 for each city, the dataset of per capita income is for each state but is more recent (2021).

2.2. Data management and privacy

In order to allow other researchers or interested parties to better understand this work's research process, we share a repository with the code used for the analyses presented in this paper. However, to protect students' privacy and be in compliance with the Brazilian General Data Protection Law (LGPD, particularly considering Article 7, Clause IV), we are not sharing the data. Instead, we provide sample data files with fictional data points that make it possible to understand the data format and run the code. While we understand the importance of replicability, since the data used here contains sensitive information and a relatively high number of features, it could be possible to identify some data points through triangulation even after anonymization.

2.3. Analysis

We focused our analysis on 6 features of our students' data: gender, race, English level, school type, city SVI, and state income per capita. These were chosen because they relate to our goal (investigating socio-demographic factors) and have a large enough sample size for the groups within. All analyses were done using the R language.

The code used for the analyses presented in this paper is available on the project's GitHub repository.

2.3.1. Descriptive statistics

We performed four major calculations in the relevant features:

- Number of Students: group counts of the total number of students who submitted a project to the event;
- Acceptance Rate: percentage of students that submitted a project and were accepted to go to the next round of the competition;
- Average Score: unweighted mean score of all judging criteria for the student's project;

- Award %: proportion of participants that got an award.

It is relevant to note that while the two first items take into consideration all the students who submitted a project, the last two ones are only applicable to confirmed finalists.

2.3.2. Inferential statistics

To better understand whether certain groups have significant advantages in the judging and awarding process, we created linear and logistic regression models to predict students' performance based on the four variables that came from the science fair dataset (school type, English level, race, and gender).

For award prediction, we followed Sonnert, Sadler, and Michaels (2013) and ran a logistic regression with the binary outcome, of being awarded or not awarded. We then analyzed the models using ANOVAs and subsequently ran pairwise comparisons with Bonferroni correction to check the significance of the differences. To predict the score, we ran a linear regression model with the score as a continuous dependent variable and also tested an additional linear regression model with an interaction. A similar subsequent analysis was conducted, but instead of predicting the final average score, we created models to predict scores of each judging criterion (Presentation; Introduction; Methodology; Results and conclusions or Product; Writing; and Project Aspects). The goal was to investigate whether the pattern for the overall score would be the same for the individual criteria.

Finally, we used the per capita income and SVI datasets to conduct a location-based analysis, examining whether those indicators could be used as predictors of the project scores.

3.Results and Discussion

As demonstrated by Figure 1, some groups represented a larger share of the participation in the event. In particular, students from public federal schools, students with a Basic English level, White students, and female students had a considerably higher number of finalists compared to the other demographics in their respective groups. This may indicate facilitated access to this type of event to particular groups. The case of students from public federal schools is relevant, considering that these schools serve less than 3% of the Brazilian student population (Inep, 2022), but have the highest representation in our data.

We also noted some relevant patterns when analyzing the award distribution within each group for the different variables, as shown in Figure 2. For school type, Public Federal schools call attention again, with over 40% of the finalists in that group receiving an award, a much higher proportion than the other school types. Students with Advanced and Intermediate English levels also had higher proportions of awards than students with Basic knowledge of the English language, despite the higher participation in the absolute number of finalists of the former group.

For race, an important observation is that despite the low participation of Black students, this was the group with the highest proportion of awarded students compared to students of other races. Finally, the award proportion was very similar for both genders.

The results of the acceptance rates calculations did not bring many interesting insights, since the acceptance rates for all groups were very similar (around 60 to 70 %). We mention them in the next subsection as support for the other results, but the specific results are presented in Table 1 of the Appendix.

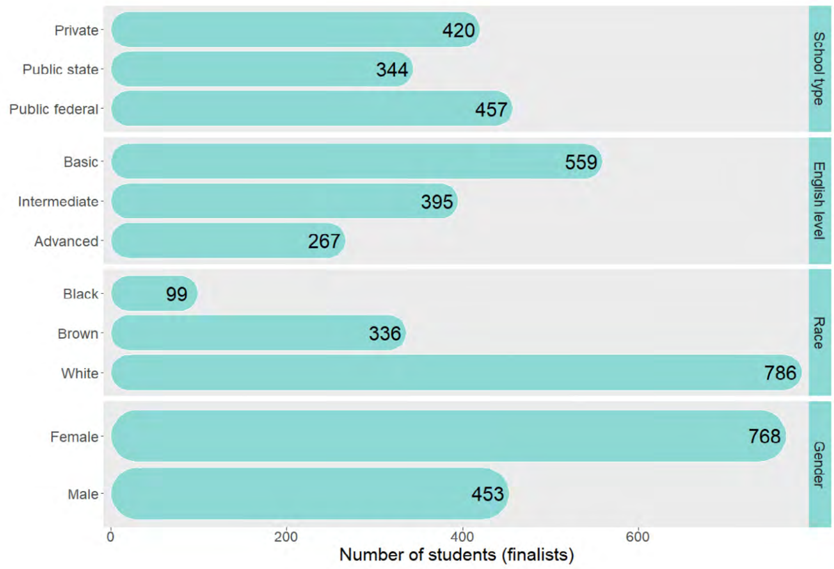

Figure 1: (left) - Number of Confirmed Finalists from each group.

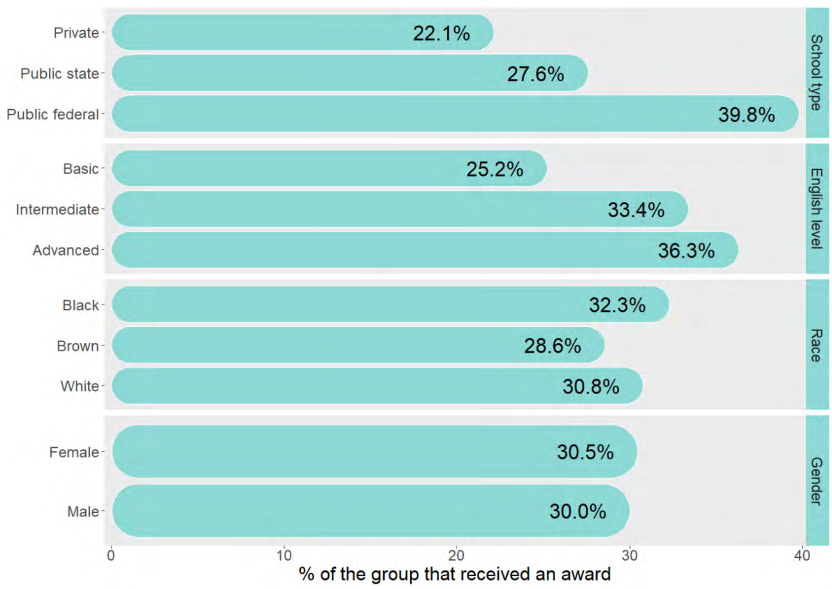

Figure 2: (right) - Proportion of awarded participants from each group.

Figure 3 presents the unweighted average scores for the projects in each group and for each variable (a), along with the coefficient estimates of the linear model we ran (b). These visualizations support the descriptions of results in subsections 3.1 through 3.4. The results from the ANOVAs of the models analyzing each judging criterion were mostly similar to the model using the overall score. Therefore, we generally only report the results that deviate from the overall results, and the specific values from the other analyses are reported in Table 2 of the Appendix.

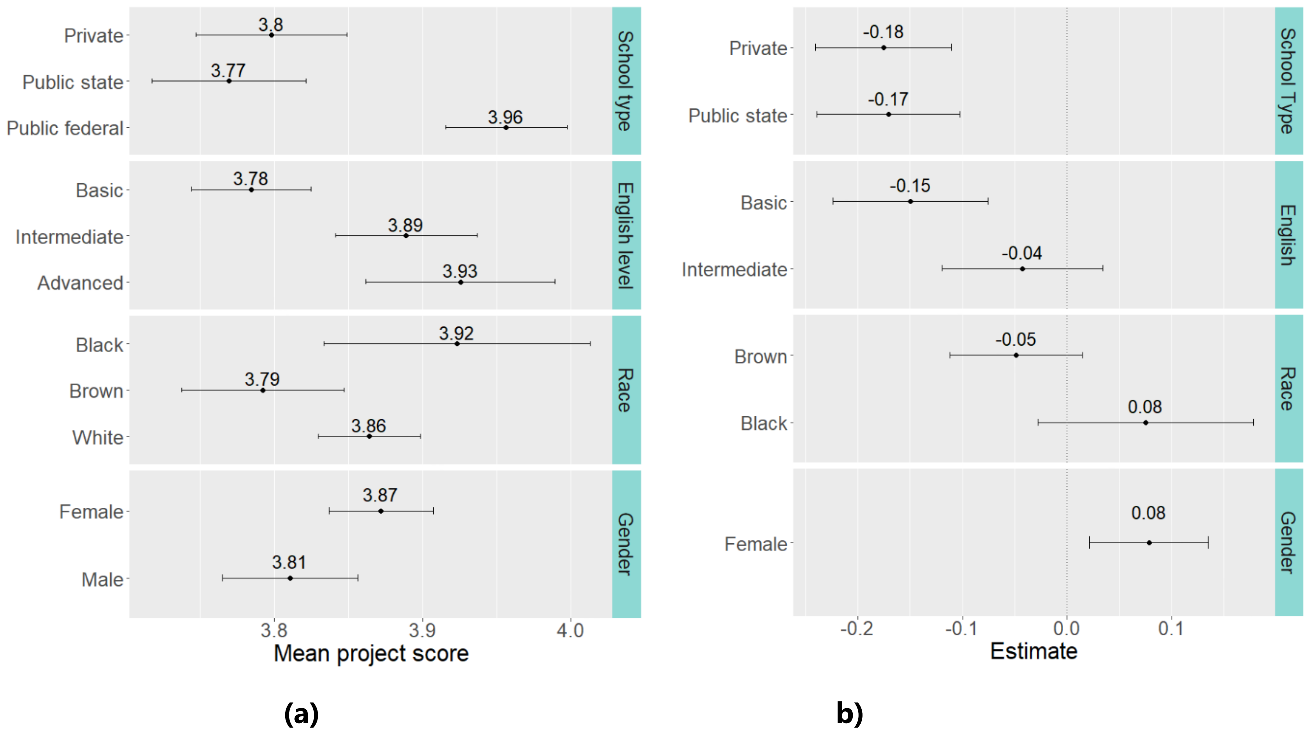

Figure 3: (a) Mean project scores by a group for each of the variables (the original scale of scores is from 0 to 5); **(b)** Coefficient plot of the Linear Regression model (the conditions not shown in the plot were the default values for the model, so the coefficient estimates are in comparison to them). For both graphs, the error bars indicate a 95% confidence interval.

3.1. School type

Federal Schools displayed the highest acceptance rate, average score, and award percentage among all school types. The ANOVA result indicated that the difference in the results was significant in the models for predicting scores ($F = 18.12$, $p < 0.001$) and awards ($p < 0.001$).

For the score model, the T-test pair-wise comparison with Bonferroni Correction showed that the Public Federal school type was significantly different from the Private (adjusted $p < 0.001$) and Public State (adjusted $p < 0.001$) school types, while the difference between Public State and Private was not statistically significant (adjusted $p = 1$).

The same pattern emerged for the award model, in which the pair-wise T-test comparison with Bonferroni Correction showed that the Public Federal school type was significantly different from the Private (adjusted $p < 0.001$) and Public State (adjusted $p < 0.001$) school types, while the difference between Public State and Private was not statistically significant (adjusted $p = 0.29$).

Finally, for the criteria analysis, there was only one criterion that was not statistically significant for school type: “Project Aspects” ($F = 1.04$, $p = 0.353$).

3.2. English level

In the English Level category, students with the Advanced level had the biggest acceptance rate, the highest average score, and also the most considerable award percentage. The ANOVA results for the score model showed a significant difference ($F = 9.87$, $p < 0.001$), and so did the ANOVA for the award model ($p = 0.001$).

The pair-wise T-test comparison between all the levels in the score model demonstrated a particularly significant difference in the other levels compared with Basic ($p < 0.001$ when compared with Advanced, and $p = 0.004$ compared with Intermediate). However, there was not a significant difference in the Intermediate and Advanced pair ($p = 1$).

Similarly, the pair-wise comparison of the award model illustrated a greater difference between the Basic level with the other two levels ($p = 0.003$ against Advanced and $p = 0.020$ against Intermediate) and not a significant difference between Intermediate and Advanced ($p = 1$).

The results of individual criteria models indicated significant differences for all criteria, showing the same pattern as the overall score, with the Basic level having statistically lower scores.

3.3. Race

For the Race variable, White participants have a bigger acceptance rate (81%), while Black participants had the greatest average score (3.92). The regression and ANOVA results were mixed, since the ANOVA for the score model ($F = 3.79$, $p = 0.023$) depicted a significant difference in scores among different races. However, the results from the linear regression and the ANOVA for the award model ($p = 0.686$) told a different story. For the pair-wise comparisons in the score model, none of the comparisons were significant ($p > 0.05$). To further explore this, we created an additional model adding the interaction between Race and School type, which showed the interaction is statistically significant. As illustrated by Figure 4, scores for Black students present a different pattern from the others, with the lowest scores for Public Federal schools, while White and Brown students from Public Federal schools have instead the highest scores compared to the other school types. These results are inconclusive and can be explored in further research.

The results indicating that White students did not had an advantage over students of other races, as we hypothesized, may be because the White race was overrepresented (as it consists of 64.4% of the finalists but 42.7% of the country’s population, according to IBGE, 2020) and the other races were underrepresented. With that, it could be the case that the Brown and Black students who had the opportunity to be finalists at the science fair were not “average students,” but instead high-achieving ones.

For the criteria analysis, the only criteria that differed from the pattern encountered for the overall score were Methodology (F = 2.585, p = 0.076) and Results (F = 2.398, p = 0.091), which did not show any significance.

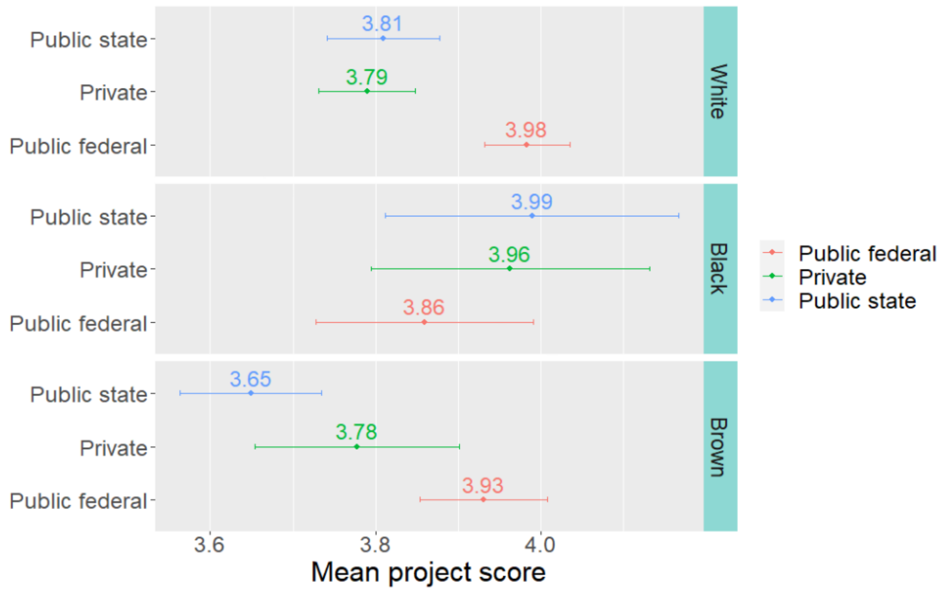

Figure 3: (a) Mean project scores by Race and School type (the original scale of scores is from 0 to 5).

3.4. Gender

Females had the highest acceptance rate and the highest average score in the gender category. Again, the regression and ANOVA results were mixed, with the score model (F = 4.564, p = 0.033) indicating a significant difference, but the award model (p = 0.870) not showing statistical significance. This difference in the models is likely due to the categories' division in the science fair, which can be further investigated in future research.

The finding that female students had similar performance as male students is in line with the literature, considering that most studies investigated award outcomes and not scores. Regarding connections between students' gender and performance at science fairs, Steegh *et al*. (2019) conducted a systematic review and found that male and female achievement in science fairs was similar.

In the same line, Sonnert, Sadler, and Michaels (2013) found that even with females making up more than 60 percent of the participants in a specific science fair, girls and boys were equally likely to be awarded. The researchers found the scientific field to be the most relevant predictor of success in the science fair, with projects in the life sciences having more chances of receiving an award. This is an interesting point to note because science fairs in Brazil are generally organized by scientific fields, with projects mostly competing only with other projects in the same field and each field receiving its respective awards.

The different criteria models revealed significance for Introduction ($F = 3.490$, $p < 0.062$), Methodology ($F = 1.518$, $p = 0.218$), Results ($F = 0.1425$, $p = 0.706$), and Project Aspects ($F = 1.6350$, $p = 0.201$), while the criteria of Presentation ($F = 10.01$, $p = 0.002$) and Writing ($F = 17.121$, $p < 0.001$).

3.5. Location-based analyses

When using the per capita income of the student's state as a predictor of the score, we found that it was not statistically significant according to the model ($t(1217) = 0.133$, $p = 0.894$). Similarly, when using SVI as a predictor for the score, the model indicated only a marginal statistical significance ($t(1208) = -1.874$, $p = 0.061$). Although per capita income and SVI can be considered related to access to resources, we suppose these were not significant at the 0.05 level because we didn't have enough data points for every city or state and also because city and state may not portray the students' realities specifically. The later explanation is supported by the result that SVI, which comes from city data and therefore is more specific than the state's income per capita, was marginally significant, while the income model was not significant at all.

4. Further discussion

The results indicate that school type and English level were important predictors of success in the competition. This pattern suggests a disparity among students from specific types of schools and with a better domain of the English language.

It is important to note that both school type and English level are related to having more access to resources. Federal schools in Brazil receive more funding (with more than twice the investment per student), have more qualified teachers (many holding Ph.D. degrees), more science laboratories and other scientific resources, and generally encourage more scientific research compared to other types of schools in the county (Inep, 2019; Inep, 2020; Spautz, 2019). Similarly, since the education of English as a second language in Brazilian schools is usually unsatisfactory, students with better English levels tend to be the ones who either can pay for a language course or are particularly motivated and can self-teach the language (Martiniano, 2021; Jareta, 2015).

Interestingly, the only criterion that did not present a statistically significant difference in scores for school type and English level was "Project aspects." That is a relevant point because this criterion is the only one that evaluates the idea more than the final project, with its sub-criteria being relevance, impact, creativity, and innovation. This supports the idea that students from state and private schools, as well as the ones with a basic English level, had promising ideas but did not have the necessary resources to develop their research projects.

Regards the relation between awards in science fairs and access to resources, results of studies such as the ones by Gifford and Wiygul (1992) and Bellipanni (1994) indicated access to resources and facilities as good predictors of success in the competitions.

5.Conclusion

One of the main innovations of this study was to amplify the scope of quantitative analysis and go beyond the science fairs in the United States and the International Science and Engineering Fair (ISEF), analyzing a science fair in Brazil. Additionally, it provides useful evidence for the influence of socio-demographic factors other than gender (for which more research has been done) on science fairs' performance. Finally, it brings more recent perspectives on the field since most of the research on the topic was done in the 1990s.

One of the main limitations of this project is that data was gathered directly from the science fair's registration website and not from a direct survey with the students. This limited the factors that could be used as elements in the analysis, consequently making interpretations of the results to be educated guesses considering the context of the events. Also, the data was collected from only one science fair, which limits the generalization of the results and might not reflect the relevance of the findings beyond the event analyzed here.

With the evidence provided by this paper, it is worth considering making changes either in the science fair judging and awarding process or in access to scientific education. With the former, one possible route would be to follow approaches by Science Olympiads, such as the Brazilian Mathematics Olympiad for Public Schools or the Brazilian Astronomy and Astronautics Olympiad, both of which have a limit or a quota on the number of awards given to some school types (IMPA, 2022; SAB, 2022). Another possible action that science fairs could take would be to include or expand programming that encourages students to be each other's allies during the event, giving space for them to collaborate, have conversations, and amplify marginalized voices, for instance, to promote social justice in scientific environments.

A more indirect approach to solve the problem would be creating programs specifically targeted at schools and students with limited resources. Thus, when the marginalized students do go to the competition, they wouldn't have a disadvantage against fellow participants.

Acknowledgements

We are grateful for the support received from the University of Richmond, particularly from Dr. Taylor Arnold, Dr. Lilla Orr, and Matthew Lowder.

References

ABERNATHY, T. V.; VINEYARD, R. N. Academic competitions in science: What are the rewards for students? **The Clearing House**, v. 74, n. 5, p. 269-276, 2001.

ADAMS, D. **Science Fair "Fairness"**. 2011. Thesis (Master of Statistics) - Department of Educational Psychology, University of Utah, Salt Lake City, 2011. Disponível em: https://mstat.utah.edu/_resources/documents/science-fair-fairness.pdf. Acesso em: 18 abr. 2023.

BELLIPANNI, L. J. The science fair experience: profile of science fair winners. **Education Resources Information Center (ERIC)**, Dec. 1994. Disponível em: https://eric.ed.gov/?id=ED395793. Acesso em: 27 abr. 2023.

CARMICHAEL, H.; MOORE, A.; STEWARD, L.; VELOPULOS, C. G. Using the social vulnerability index to examine local disparities in emergent and elective cholecystectomy. **Journal of Surgical Research**, v. 243, p. 160-164, 2019.

CZERNIAK, C. M. Predictors of success in a district science fair competition: an exploratory study. **School Science and Mathematics**, v. 96, n. 1, p. 21-27, 1996. Disponível em: https://doi.org/10.1111/j.1949-8594.1996.tb10208.x. Acesso em: 19 out. 2022.

DA SILVA GALLON, M.; ZOTTI DA SILVA, J.; SOUSA DO NASCIMENTO, S.; DA ROCHA FILHO, J. B. Feiras de ciências: uma possibilidade à divulgação e comunicação científica no contexto da educação básica. **Revista Insignare Scientia**, v. 2, n. 4, p. 180-197, 2019.

FEKETE, A. Validation of a social vulnerability index in context to river-floods in Germany. **Natural Hazards and Earth System Sciences**, v. 9, n. 2, p. 393-403, 2009. Disponível em: https://doi.org/10.5194/nhess-9-393-2009. Acesso em: 18 ago. 2020.

GIFFORD, V. D.; WIYGUL, S. M. The effect of the use of outside facilities and resources on success in secondary school science fairs. **School Science and Mathematics**, v. 92, n. 3, p. 116, 1992.

IBGE. **Características gerais dos domicílios e dos moradores**: 2019. Rio de Janeiro: IBGE, 2020. Disponível em: https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101707. Acesso em: 22 maio 2024.

IBGE. **IBGE divulga o rendimento domiciliar per capita 2021**. Rio de Janeiro: IBGE, 2021. Disponível em: https://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Renda_domiciliar_per_capita/Renda_domiciliar_per_capita_2021.pdf. Acesso em: 22 maio 2024.

IMPA. **Regulamento da 17ª Olimpíada Brasileira de Matemática das Escolas Públicas** – OBMEP. 2022. Disponível em: https://www.obmep.org.br/docs/2022/regulamento.pdf. Acesso em: 22 maio 2024.

INEP. **Notas estatísticas:** censo escolar 2018. Brasília, DF: Inep; MEC, 2019. Disponível em: https://download.inep.gov.br/educacao_basica/censo_escolar/notas_estatisticas/2018/notas_estatisticas_censo_escolar_2018.pdf. Acesso em: 22 maio 2024.

INEP. **Censo da educação básica 2019:** notas estatísticas. Brasília, DF: Inep; MEC, 2020. Disponível em: https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/notas_estatisticas_censo_da_educacao_basica_2019.pdf. Acesso em: 22 maio 2024.

INEP. **Censo da educação básica 2021:** notas estatísticas. Brasília, DF: Inep; MEC, 2022. Disponível em: https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/notas_estatisticas_censo_escolar_2021.pdf. Acesso em: 22 maio 2024.

IPEA. **Atlas da vulnerabilidade social**. On-line: Ipea, 2019. Disponível em: https://dados.gov.br/dados/conjuntos-dados/ivs. Acesso em: 22 maio 2024.

JARETA, G. Por que o ensino do inglês não decola no Brasil. **Educação**, on-line, 4 nov. 2015. Disponível em: https://revistaeducacao.com.br/2015/11/04/por-que-o-ensino-do-ingles-nao-decola-no-brasil/. Acesso em: 22 maio 2024.

KIDDELL, R. B. **Science fair project adjudication:** a study of 3 judges. 1987. Thesis (Master of Arts) - Department of Science Education, University of British Columbia, Vancouver, 1987.

LEVIN, K. N.; LEVIN, R. E. How to judge a science fair: use the consensus method. **The Science Teacher**, v. 58, p. 43-45, 1991.

MAGALHÃES, D. C.; MASSARANI, L.; ROCHA, J. N. Memórias das feiras de ciências: entrevistas com participantes das feiras estudantis de ciências das décadas de 1960 E 1970 em São Paulo. **Linguagens, Educação e Sociedade**, v. 26, n. 51, p. 234-267, 2022.

MARTINIANO, M. O ensino do Inglês no Brasil e o problema das escolas públicas. 2 Jul. 2021. **Blog Flexge**. Disponível em: https://blog.flexge.com/ensino-ingles-brasil-escolas/. Acesso em: 29 maio 2024.

MCDONNELL, A. M. **Gender and ethnic preferences in selection of science fair topics**. 1996. Thesis (Master of Science) - College of Arts and Sciences, Texas Woman's University, Denton, 1996.

NASSIF-PIRES, L.; CARVALHO, L. B. D.; LEDERMAN RAWET, E. Multi-dimensional inequality and Covid-19 in Brazil. **Investigación económica**, v. 80, p. 33-58, 2021.

NOORDERHAVEN, N.; THURIK, R.; WENNEKERS, S.; STEL, V. The role of dissatisfaction and per capita income in explaining self–employment across 15 European countries. **Entrepreneurship Theory and Practice**, v. 28, p. 447-466, 2004.

POSEL, D.; ROGAN, M. Women, income and poverty: gendered access to resources in post-apartheid South Africa. **Agenda**, v. 23, p. 25-34, 2009.

SCHMIDT, K. M.; KELTER, P. Science fairs: a qualitative study of their impact on student science inquiry learning and attitudes toward STEM. **Science Educator**, vol. 25, p. 126–132, 2017.

SAB (SOCIEDADE ASTRONÔMICA BRASILEIRA). Regulamento da 25ª Olimpíada Brasileira de Astronomia e Astronáutica - OBA. **oba.org.br**, 2022.

SONNERT, G.; SADLER, P.; MICHAELS, M. Gender aspects of participation, support, and success in a state science fair. **School Science and Mathematics**, v. 113, p. 135-143, 2013.

SPAUTZ, D. Escolas federais custam menos que as militares e têm desempenho superior no Enem. **NSC Total**, on-line, 8 jan. 2019. Disponível em: https://www.nsctotal.com.br/colunistas/dagmara-spautz/escolas-federais-custam-menos-que-as-militares-e-tem-desempenho-superior. Acesso em: 22 maio 2024.

STEEGH, A. M.; HÖFFLER, T. N.; KELLER, M. M.; PARCHMANN, I. Gender differences in mathematics and science competitions: a systematic review. **Journal of Research in Science Teaching**, v. 56, p. 1431-1460, 2019.

Sobre os autores

Daniel Verdi do Amarante

Daniel is a Cognitive Science undergrad at the University of Richmond. His research spans education data science, science and technology studies, and natural language processing. As a previous science fair participant and organizer, he now wants to contribute to making them even more equitable.

email: daniel.verdidoamarante@richmond.edu

Kritim Rijal

Kritim is a rising senior at the University of Richmond studying Mathematics and Computer Science. His previous research areas include adversarial neural networks and understanding human psychological behaviors. He is particularly interested in the fields of Data Science and Machine Learning.

email: kritim.rijal@richmond.edu