UNIVERISTY OF PENNSYLVANIA

SCHOOL OF SOCIAL POLICY AND PRACTICE

MASTER OF SCIENCE IN SOCIAL POLICY AND DATA ANALYTICS

CASE STUDY REPORT ASSIGNMENT 1

THE IMPACT OF TEST PREPARATION ON ADMISSIONS PROSPECTS AND SCORES ON MULTI-RACIAL STUDENTS

Autor: Rose Barragan Barranco

Professor: Richard Hartwell

Course: MSSP 6070 Practical Programming For Data Science

October 15th, 2023

INDEX

1. EXECUTIVE SUMMARY

This report conducts a thorough analysis of the Student Performance Dataset by analyzing among ten different categories of variables to understand which are the most relevant factors in obtaining high scores in the admission exams, both in terms of the total score and the interdisciplinary scores that compose it.

The report also examines further fundamental aspects in terms of the student's academic and personal background, such as the highest level of education attained by the parents or the nutritional level of the school system. Fundamental aspects for the correct cognitive and social development of the student that have a strong relevance on the students' test performance.

Once all these aspects are taken into account, and how they are interrelated, the data will support our recommendations to the institutions, with the ultimate goal, that students can have a satisfactory academic and professional achievement once they start classes in the new admissions center.

2. INTRODUCTION

Prior to starting with the analysis, it is necessary to understand what kind of data we have in our dataset – "*Student Performance Data Set"* – collecting a sample of a total of 1000 students and a set of ten different variables:

- **Gender**: categorical variable, that distinguishes two genders: male and female.
- **Race/ethnicity**: categorical variable that classifies students into five racial/ethnic groups: Group A/B/C/C/D/D/E
- **Parental level of education**: categorical variable, which reflects the highest level of education attained by the students' parents: some high school, high school, some college, associate's degree, bachelor's degree, master's degree,
- **Lunch**: categorical variable reflecting the type of lunch students have: free/reduced or standard.
- **Math score**: numerical variable indicating the score on the math test, with the lowest score being 0 and the highest being 100
- **Reading score**: numerical variable indicating the score on the reading test, with the lowest score being 17 and the highest 100.
- **Writing score**: numerical variable indicating the score on the writing test, with the lowest score being 10 and the highest 100.
- **Total score:** total sum of the three interdisciplinary exams, with the lowest score being 27 and the highest 300.
- **Test preparation course**: categorical variable indicating whether the student has taken an admissions preparation course, where "none" means that he/she has not taken the course; and "completed" means that he/she has successfully completed the course.
- **Admission prospect**: categorical variable where the admissions prospect is collected and classified into three levels: low, medium, and high.

Considering these ten variables and the relationship between them, we will assess whether or not the exam preparation courses help obtain a better admissions score. Based on the insights, we will proceed with a series of recommendations regarding the exam preparation course requirements that students should take.

3. ANALYSIS AND METHODS USED

To proceed with the analysis of the Student Performance Dataset, we will work with Google Colab tools - programming in Python - and Microsoft 365 Toolkit for a thorough analysis of the data, resulting in a comprehensive storytelling that combines a seamless narrative with visualizations that illustrate and facilitate the understanding of the analysis.

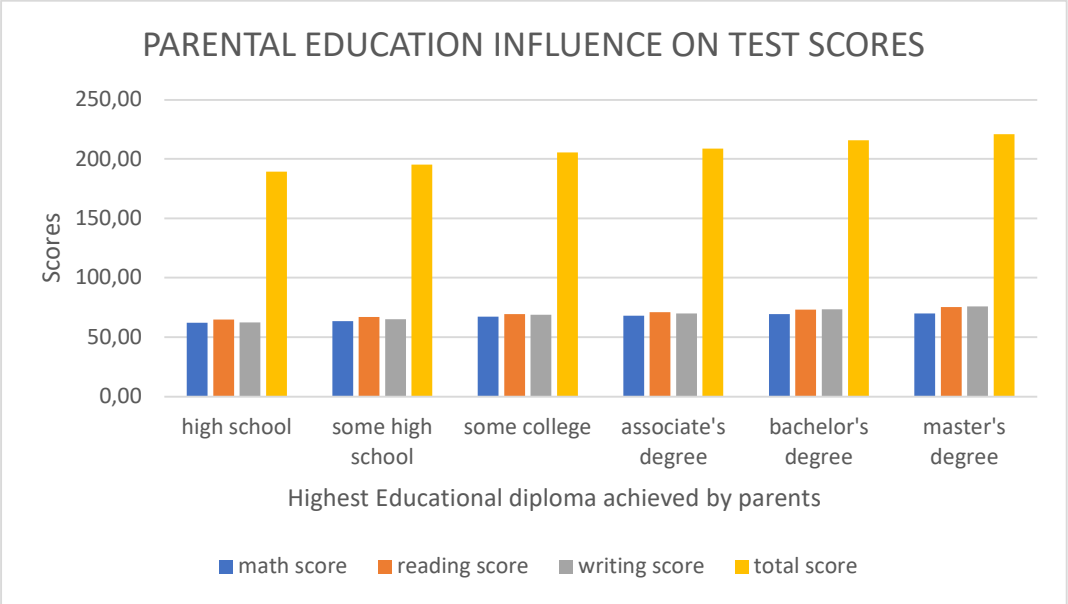
3.1. PARENTAL EDUCATION AND ITS INFLUENCE ON THE TEST SCORES

The level of degree that the students' parents have received is very relevant since those parents who have obtained the highest certification (master's degree) have children with the highest average total score, equal to 220.7966. This group is followed by parents with a "bachelor's degree," whose children have a mean total score of 215.771.

Among the intermediate groups, the students whose parents had obtained an "associate's degree" came in with an average total score of 208.707, and "some college degrees" with an average total score of 205.429.

Finally, the lowest ranking is held by those students whose parents have a high school degree ("some high school degree" with an average total score of 195.32; and in last place, "high school diploma" with an average total score of 189.29) [Table #1].

Table #1 - Parental education influence on test scores



Source: Rose Barragan Barranco (own graph).

Thus, the highest score corresponds to parents with a level of education equivalent to a master's degree, their children's total score being 220.7966, while the lowest score corresponds to parents with a level of education equivalent to a high school, their children's total score being 189.29.

Hence, the percentage difference between the highest score and the lowest score [((220.796610 - 189.290816) / 189.290816) * 100] is equal to 16.67%. That is, students whose parents have obtained a master's degree have 16.67% higher scores than students whose parents have obtained the lowest level of education (high school).

What is interesting to see is that this ranking is not only sustained at the "total score" level, but also in each of the three scores that make up the exam (math score, reading score, writing score).

Thus, the fact that parents with a higher level of education have a higher average score at both the total and interdisciplinary levels is accentuated.

This may be because in the corresponding households, parents with a higher level of education could invest more time and resources in their children's education, as they did with them previously.

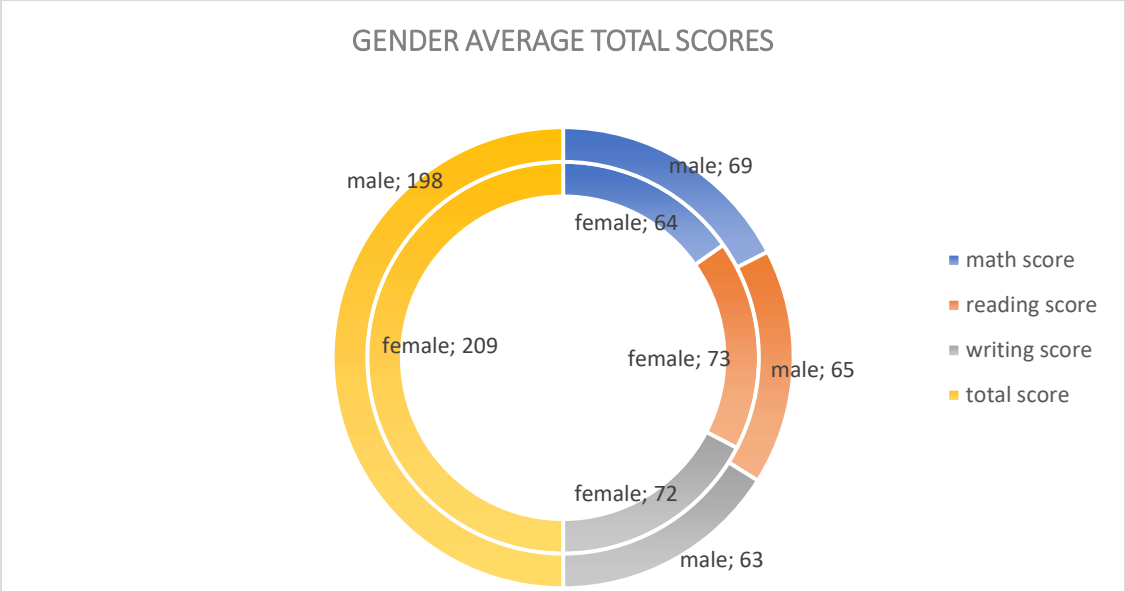
3.2. GENDER PERFORMANCE REGARDING TEST PREPARATION COURSE

The sample of 1,000 students includes 518 female students and 482 male students. Of these, the female group has a higher mean total score than the male group, 208.70 points compared to 197.51, respectively.

However, it is interesting to note that this difference in the total score is not repeated in the interdisciplinary scores. Here, we can observe that male students have a higher average score on the mathematics test (68.72) than girls (63.63).

But this is not sustained in the rest of the interdisciplinary tests, since females have obtained higher average scores in both the reading test and the written test (72.60 and 72.46 respectively) with respect to the score of male students (65.47 and 63.31 respectively) [Table #2].

Table #2 – Average total and interdiciplinary scores sorted by gender



Source: Rose Barragan Barranco (own graph).

Therefore, here we can observe a clear difference in the performance of the science and social studies subjects, where students have achieved very different results from each other.

This premise can be supported by the study called ***Social Psychology of Education*** [1], which tries to find out why women choose studies related to the humanities and social sciences while men prefer technological or scientific careers.

As a result, these scores reflect that male students have obtained a better average score in the mathematics test, while female students have obtained a better average score in the reading and writing test.

On the other hand, it is worth noting the effects of taking the test preparation courses in relation to the mean total and interdisciplinary scores for each of the student genders.

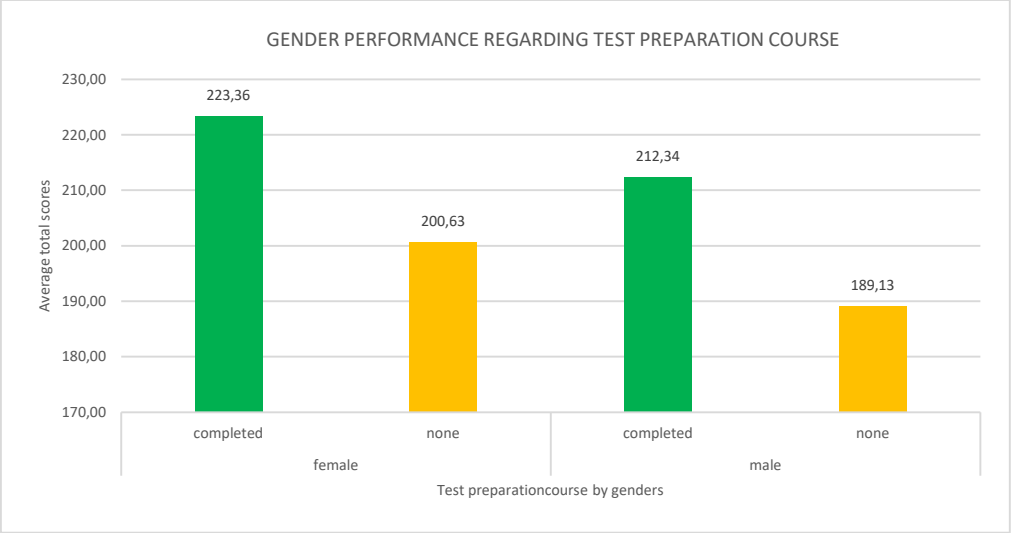
We observe that all those students (both male and female) who have participated and completed the exam preparation courses have obtained a much higher mean total score than those students who have not completed the exam preparation test.

The difference in score between girls who have completed (223.36) and those who have not completed (200.63) the preparation course is 11.34% [((223.36 - 200.63]/200.63) *100].

The same occurs for the male gender, i.e., the difference in the mean total scores of the boys who have completed the exam preparation course (212.34) to those who have not (189.122) is 12.26% [((212.34 - 189.133]/189-133)*100].

Thus, we see that the preparatory courses contribute positively to achieving a higher overall and interdisciplinary grade point average than those students who have not taken the course. Furthermore, we can see that for the male group, the percentage difference is even higher. So, this preparatory course has a very positive impact on both groups, but especially on the male group [Table #3].

Table #3 - Gender performance regarding test preparation course



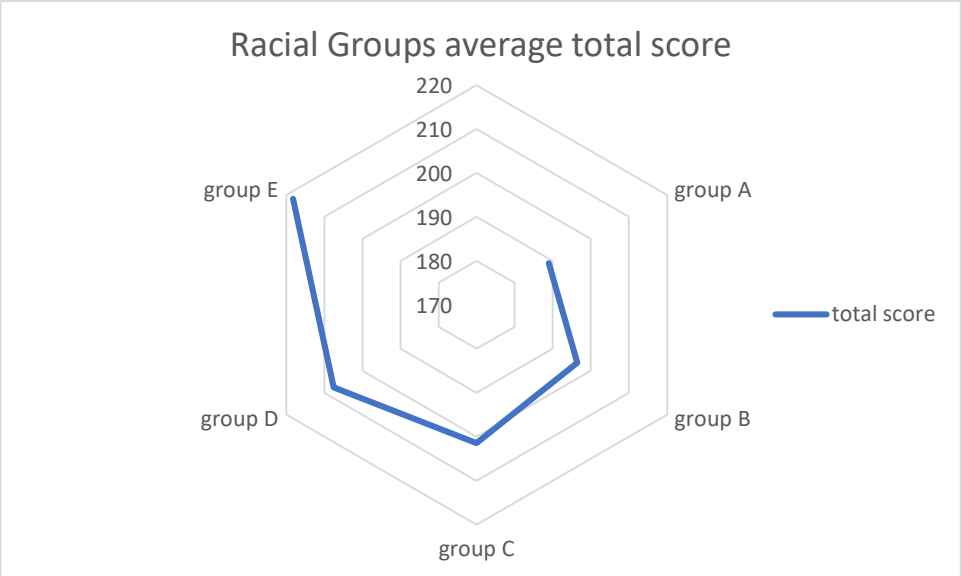
Source: Rose Barragan Barranco (own graph).

3.3. BEST PERFORMING GROUP FOR ADMISSIONS PROSPECT

According to our study, among the five participating racial groups (A/B/C/D/D/E), the group that obtained the best scores is group E.

On the one hand, this premise is observed in that group E obtained a higher average total score and higher average interdisciplinary score than the other four groups, scoring an average of 73.821 on the mathematics test, 73.0285 on the reading test, and 71.407 on the written test; with a total average score of 218.257 [Table #4].

Table #4 – Average total score by racial groups



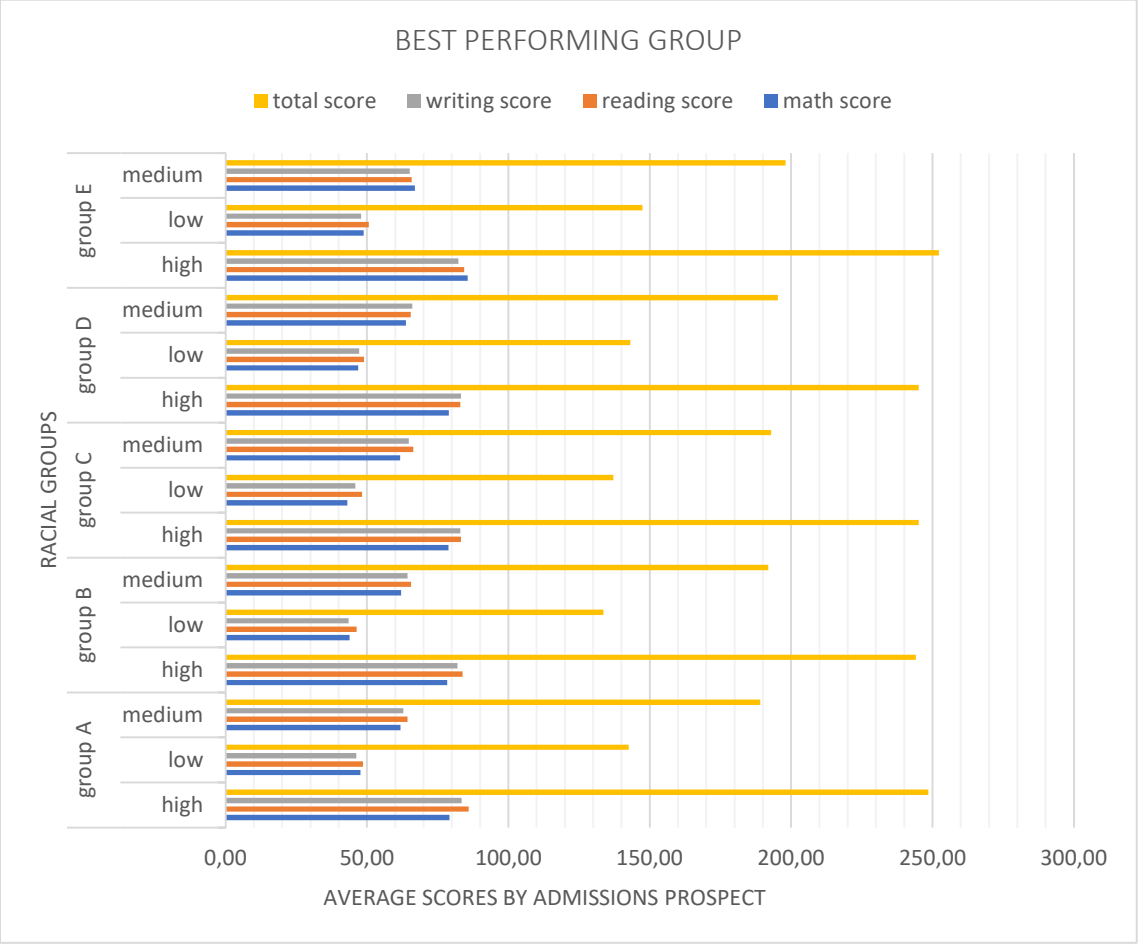
Source: Rose Barragan Barranco (own graph).

On the other hand, race group E not only has achieved the best interdisciplinary and total scores, but it is also the race group that has achieved the best admission prospects.

The admission prospect at the low level equals an average total score of 147.44, the medium level equals an average total score of 197.98, and the high level equals an average total score of 252.304.

We cannot ignore the fact that group E is the racial group by excellence in the admissions prospection. This is observed in that the other groups (A/B/C/D) have obtained an average high total score of around 245, an average medium total score of 191.75, and an average low total score of 139.10. Group E thus becomes the racial group by excellence in the admissions prospect [Table #5].

Table #5 – Best performing groups in relation to the admissions prospects.



Source: Rose Barragan Barranco (own graph).

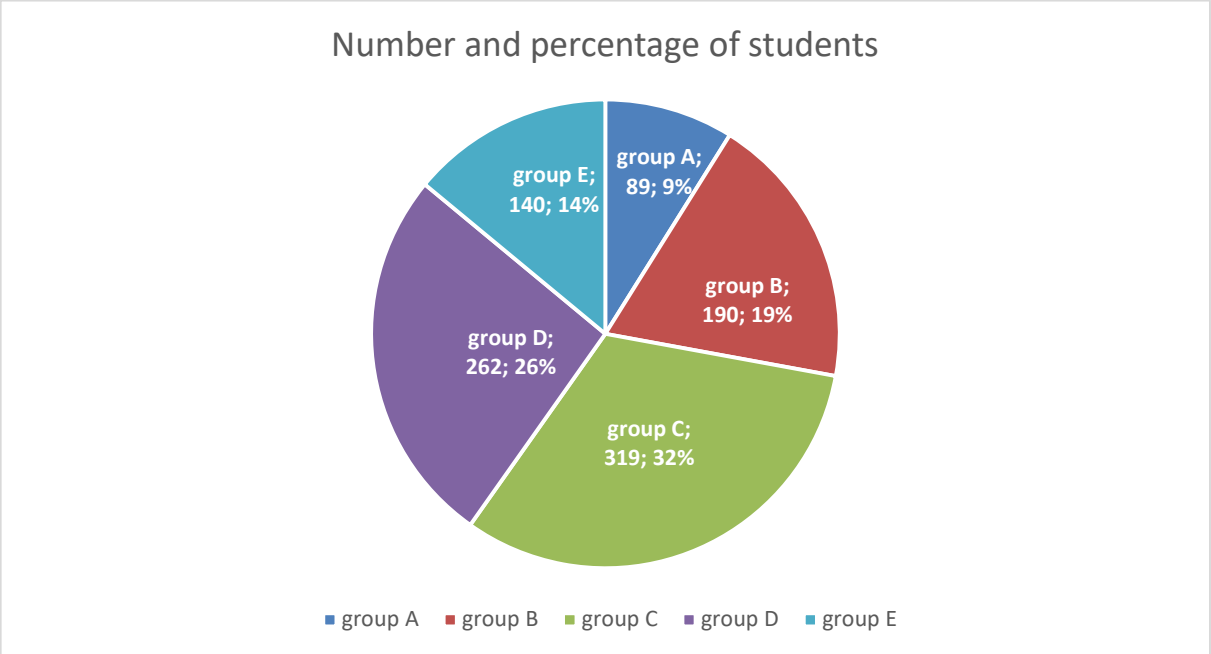
3.4. OTHER DIFFERENCES BETWEEN THE RACIAL GROUPS

Continuing with the analysis of the racial groups, we note an apparent disparity in the number of students belonging to each racial group since C is the majority group with 319 students, D has 262 students, B has 190 students, E has 140 students, and finally, group A has 89 students.

So, the difference between the majority group (C) and the minority Group (A) is 72.41%. In other words, Group C is approximately 3.6 times larger than Group A.

In the same manner, we make the comparison with the rest of the groups, where Group B is about 40% smaller than Group C; Group D is about 18% smaller than Group C; and finally, Group E is 56% smaller than Group C. Therefore, this shows that the sample collected is unbalanced [Table #6].

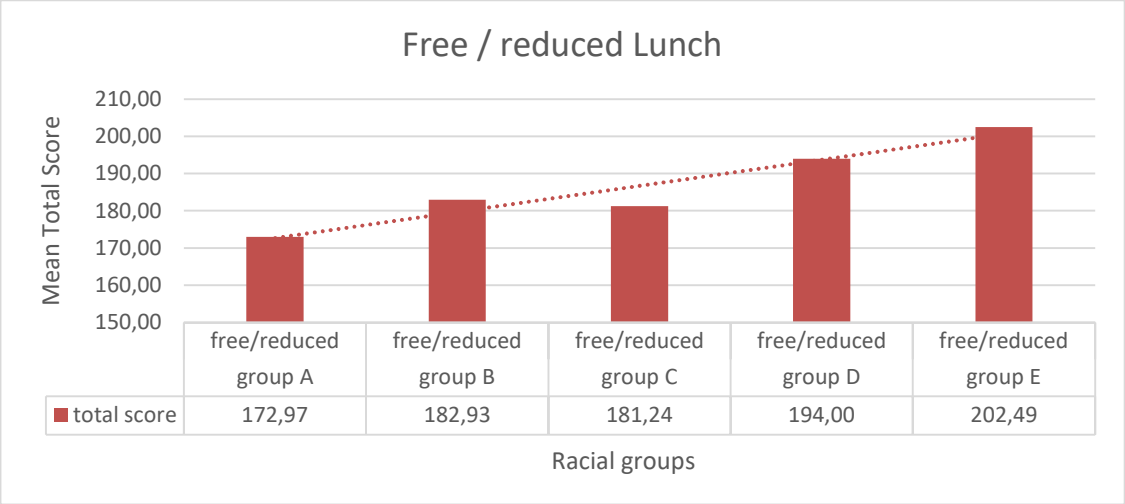
Table #6 – Number and Percentage of students in each racial group.



Source: Rose Barragan Barranco (own graph).

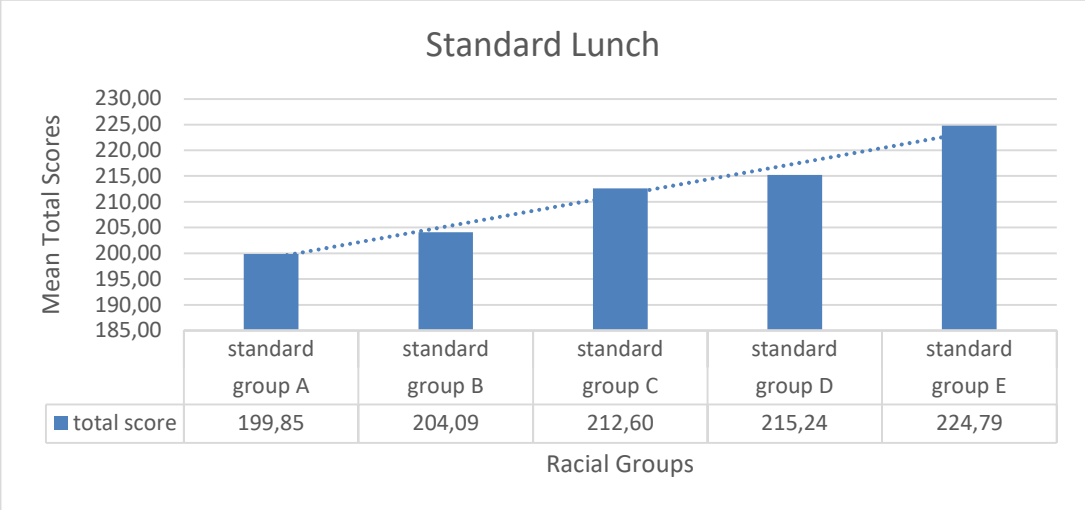
Another relevant difference is concerning the "lunch" category since all students in Groups A/B/C/D/E who receive the standard type of lunch have a much higher mean total and interdisciplinary score (199.85; 204.09; 212.60; 215.24, 224.79 respectively) than those students who receive a free/reduced lunch (172.97; 182.93; 181.24; 194; 202.49 respectively) [Table #7 & Table #8].

Table #7 – Free/reduced Lunch for each student group



Source: Rose Barragan Barranco (own graph).

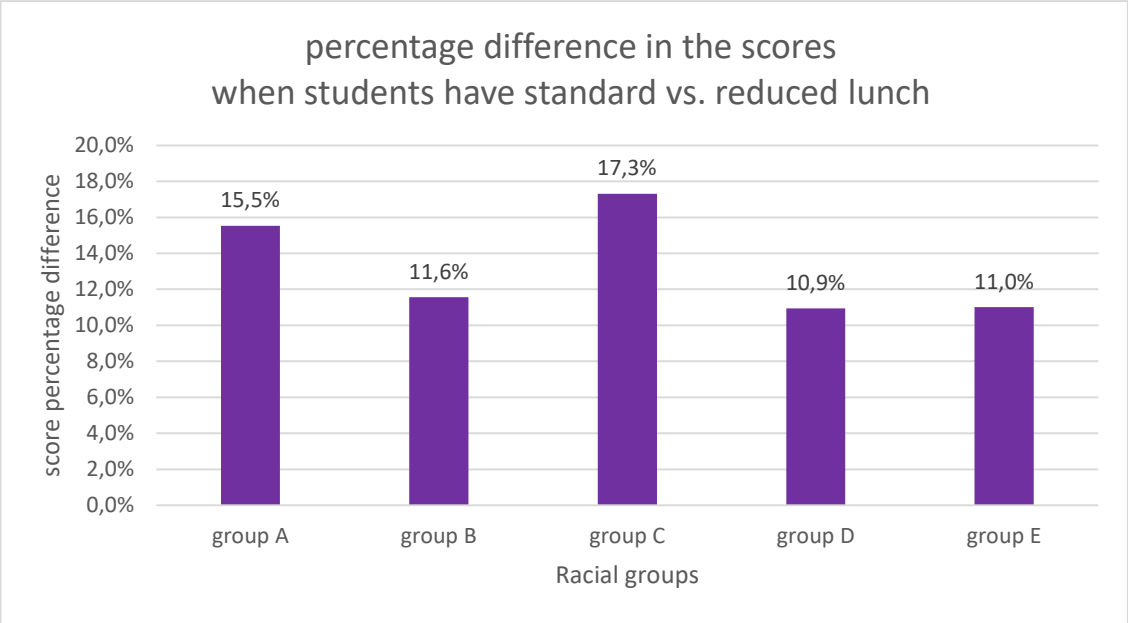
Table #8 – Standard Lunch for each student group



Source: Rose Barragan Barranco (own graph).

This is because child nutrition is an essential aspect for the students' cognitive and social proper development. Thus, students who receive the most deficient nutrition (free/reduced) tend to have a percentage difference of about -13.3% in the average total scores [Table #9].

Table #9 – percentage difference in the scores when students have standard vs. Free/ reduced lunch



Source: Rose Barragan Barranco (own graph).

3.5. BIASES IN THE SAMPLE

It is paramount to have an unbiased data sample whose representation of the data can be as close to reality as possible so that the results can be accurate, valid, equitable, and fair (non-discriminatory), especially when addressing issues related to gender, racial, or ethnic diversity.

Thus, it is essential that both the researchers, the dataset, and its analysis are free of bias. In our Research, we have analyzed biases from two different perspectives: racial/ethnic group and gender of the students.

On the one hand, the analysis of the racial group has been anticipated in section 3.4, where we conclude that the sample collected shows a disparity (unbalance) in the number of students composing each racial group, there being a tangible difference in the number of students composing each racial group.

Therefore, the data collection has not been equitable and may show discriminatory results that benefit the majority groups or harm the others, as well as perpetuate stereotypes and inequalities.

On the other hand, from the point of view of gender, as mentioned above, the sample includes 518 female students (51.8%) and 482 male students (48.2%).

In order for the results of this study to apply to society, the sample must be representative of that population. For this, we will reference the Research called "***Total population of the United States by gender 2010-2027,"*** performed by the Statista Research Department in 2023 [2]. The study confirms that since 2013, the gender distribution in the United States has remained stable, with women comprising about 51.1 percent of the population.

Therefore, we conclude that from a gender standpoint, the sample is representative of the U.S. population, with an accurate gender distribution.

4. RECOMMENDATIONS & CONCLUSION

After conducting this thorough study of the data sample, we can conclude that the students who obtained higher total and interdisciplinary scores are influenced by several factors: being part from racial group E; having an excellent dietary support system; having parents who have obtained a higher education equivalent to a master's degree; having successfully participated and completed the exam preparation courses. This last factor is a differential factor not only in the exam score but also for the admission prospects.

Therefore, our recommendation is that all study centers implement a mandatory requirement of satisfactory completion of the exam preparation course. This does not guarantee that the student will be admitted to the institution since there are other factors to consider (curriculum, motivations, concerns, fit with the university, personal trajectory, etc.). But, it does guarantee that these students will have higher total and interdisciplinary scores, which will result in a higher probability in the admissions prospection.

Moreover, the obligatory nature of this exam can offer more security to the students in order to face the courses with a more solid and foundational knowledge background than if they did not take the exam, achieving better academic and professional achievement, as well as good mental health and a lower degree of stress during the new academic year.

5. REFERENCES

[1] Trusz, S. (2020). Why do females choose to study humanities or social sciences, while males prefer technology or science? some intrapersonal and interpersonal predictors. *Social Psychology of Education*, *23*(3), 615–639. https://doi.org/10.1007/s11218-020-09551-5

[2] Statista Research Department. (2023, October 6). *U.S. population by gender 2027*. Statista. https://www.statista.com/statistics/737923/us-population-by-gender/#:~:text=Projection%20estimates%20calculated%20using%20the,US%20Census%20data%20for%202021