Does Parental Level of Education Affect Student's Performance? An Analysis of Kaggle Data

Group 47

Faaiza Khan, Muhammad Maaz Ahmad, Asif Nawaz, Ayesha Naweed, Eraj Zafar

Abstract, Context: For a student to perform well in their education, there are lots of factors taken into consideration, such as gender of the student, whether the student has completed a test preparation course.

Objective: To calculate the correlation between the parental level of education and the mean scores.

Method: The scores in the given Kaggle dataset of 1000 students were analysed to assess the average score of each student and its correlation with the parental level of education.

Results: Most students indicated their parent's level of education scatter around the lower level of education. The number of parents with a higher level of education were lower in ratio. However, the data shows that most students averagely scored in between 64% to 80%.

Conclusions: For our analysis, we hypothesise that the parental level of education will impact the students' grades. We believe that the higher the level of parental education, such as bachelor's or master's degree, the higher the students' scores.

1. Introduction

For a student to perform well in their education, there are lots of factors taken into consideration. In reference to the dataset(Kaggle.com, 2019) we analysed among others in Kaggle, the factors affecting student performance include the gender of the student, whether the school has reduced or standard lunch options, whether the student has completed a test preparation course. For this analysis, we answered the question "Does parental level of education affect student performance?" To answer the question, we will attempt to calculate the correlation between the parental level of education and the mean scores.

The scores in the given dataset(Kaggle.com, 2019) has been divided into writing score, reading score and writing score. These can be further analysed in analysing whether the writing score affects the math score and so on. For us to understand the performance in the various scores, we will first change the data so as to indicate the average scores of each student between the maths, reading and writing score. Currently, the dataset displays the maths, reading and writing scores separately.

It is hypothesised that student-parent dynamics are essential "to the success of a child's academic career". Currently, few studies specifically address the relationship between

parental education accomplishment and student academic performance. (Lib.dr.iastate.edu, 2019)

1.1. Research Question and Background

In this paper, our research question is “To what extent does parental level of education affect student performance?”

According to De Graaf et al., In the Netherlands, the reading habits of parents strongly impacted the student’s academic performance rather than the participation of parents. It is necessary to consider, however, that Jackson et al. (2007) and Erikson (2007) took social class as their indicator of social background and not parental level of education. Kloosterman, R., Ruiter, S., De Graaf, P.M. and Kraaykamp, G. (2009)

When we discuss the term home literacy environment, it refers to literacy-related conversations with parents, resources for students available at home and attitudes at home that affect the student’s performance. In the paper, Hyujoon Park concludes that even though there is “some correlation between parental education and home literacy environments”, there is still a notable amount of parents with low level of education who are often involved with student when it comes to “literacy activities, positive attitudes toward reading, and have a large number of books at home”. (Park, 2019)

For parents to increase literacy-related conversations with students, they can carry out the following activities:

- “Monitor their child’s academic progress
- Helping with homework and reading to and with the child
- Provide a structure for homework, including rewards and punishments
- Monitor their children’s homework (Barge, J.K. and Loges, W.E. (2003))

The above activities carried out by parents will help ensure a child's responsibility when it comes to homework submission deadlines and also to help teachers monitor student performance. (Barge, J.K. and Loges, W.E. (2003))

In this paper, Barge highlights parents who stress the “importance of involving their children in extracurricular activities and engaging in communal and shared parenting.” Barge et al. emphasises that “effective communication is about creating supportive relationships among parents, teachers, and community members to foster the academic and social development of” the child.

“Dornbusch, Ritter, Leiderman, Roberts, and Fraleigh (1987) purport that parental education level affects parenting style, which, on the other hand, affects children's academic success. Dombrusch et al. added that parents with higher educational levels are likely to be more permissive and less strict in parenting. According to Mullis and Jenkins (1990) and White (1982), parental education shows a strong correlation to students' academic achievement”. (Lib.dr.iastate.edu, 2019)

“In a study of nearly 25,000 8th graders, Lee and Peng (1994) found that students whose parents only attained high school diplomas were five times more likely than their peers to drop out of high school by the 10th grade”. (Lib.dr.iastate.edu, 2019)

The discussion so far convinced us to say that parents influence a child’s behaviour, decisions, performance and day to day routines. However, children’s performance is also influenced by their environment and genetically transferred habits. That is why the distinctive personal abilities and personalities of children can shape the parenting they receive Walker, J. and Smrekar, C. (2019). In his paper, longinus Chukwudi, O. (2017) justifies the importance of parental level of education by the point of view that, parents can act as second teachers to their children in order to guide them and help them perform better in education. On the other hand, research by Okpala, C., Okpala, A. and Smith, F. (2019) emphasises that economic status of a child is correlated with its academic performance.

However, parents' level of education should not be considered as a predictor of a child's academic performance. According to Walker, J. and Smrekar, C. (2019), parent's level of education is not directly related to a child's performance, and rather, it is one of the attributes as a part of psychological, physical and sociological variables influencing a child's achievements. Hence, it should not be considered a single source of influence upon a child’s academic outcomes.

2. Method and Data Analysis

We began the analysis by carrying out some content analysis of the Kaggle dataset (Barge, J.K. and Loges, W.E. (2003)). Kaggle dataset obtained marks secured by the students in high school from the United States.

The dataset was divided into eight parts: gender, race/ethnicity, and parental level of education, lunch, and test preparation course, math score, writing score and reading score.

We first adjusted the data by calculating the mean scores of each student. To take into consideration student 1 in the dataset who had 72, 72, 74 in math score, reading score and writing score, respectively, the mean score calculated is 73. This is the values we used to calculate the correlation with parental level of education.

The parental level of education is divided into six levels which have been described in the table below:

Some high school	Level - 1
High school	Level - 2
Some college	Level - 3
Associate’s degree	Level - 4
Bachelor’s degree	Level - 5
Master’s degree	Level - 6

According to the table above we adjusted the Kaggle dataset by changing the parental level of education in numeric levels to calculate the correlation. For example, we ranked some high schools as level 1.

Upon that, we carried out some descriptive analysis and plotted some graphs, such as a scatter plot and box plot. We also used the bar graph to give us a ratio of the number of parents at each level of education. Most students indicated their parent's level of education, which scatter around the lower level of education, as shown below.

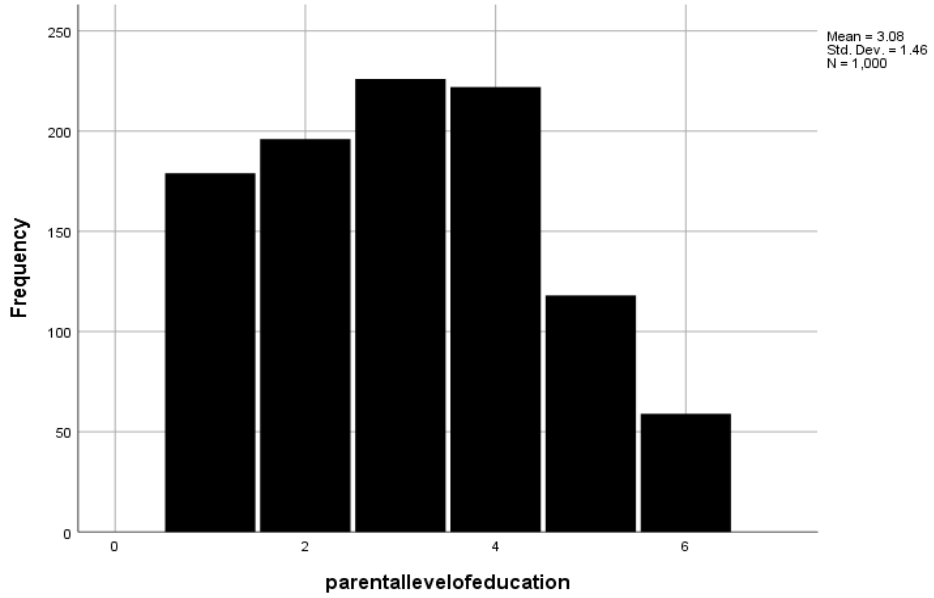

Figure 1: Histogram depicting the count of parents in each level of education

3. Results

The data shows that most students averagely scored in between 64% to 80%, as shown below.

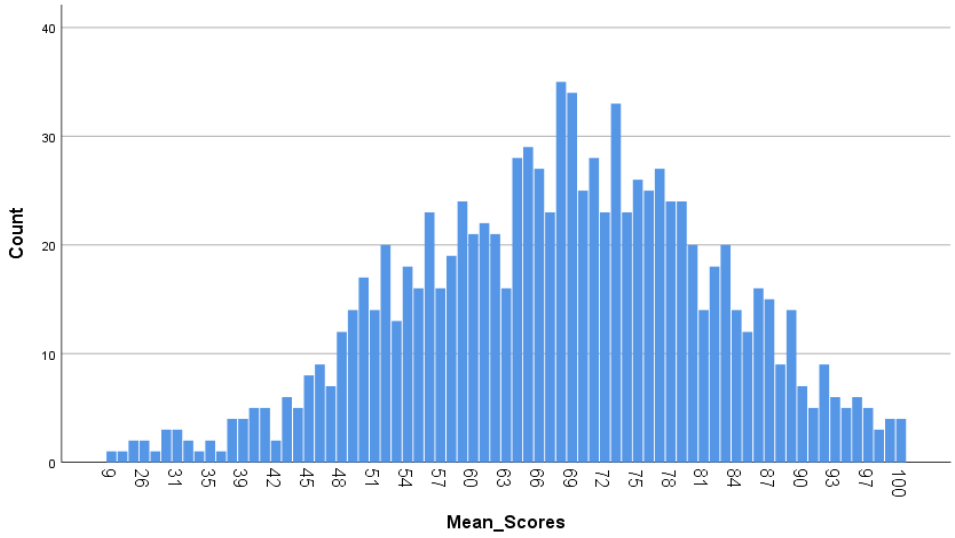

Figure 2: Bar graph representing the mean scores

In reference to the scatter plot generated using the parental level of education and the mean scores. The scatter plot indicates that most of the points cluster together tight after 40% average scores which shows there is a strong relationship between parental level of education and the mean scores. We realised that most students mean scores started at 40% with the exception of a few outliers which will be clearly described in the box plot.

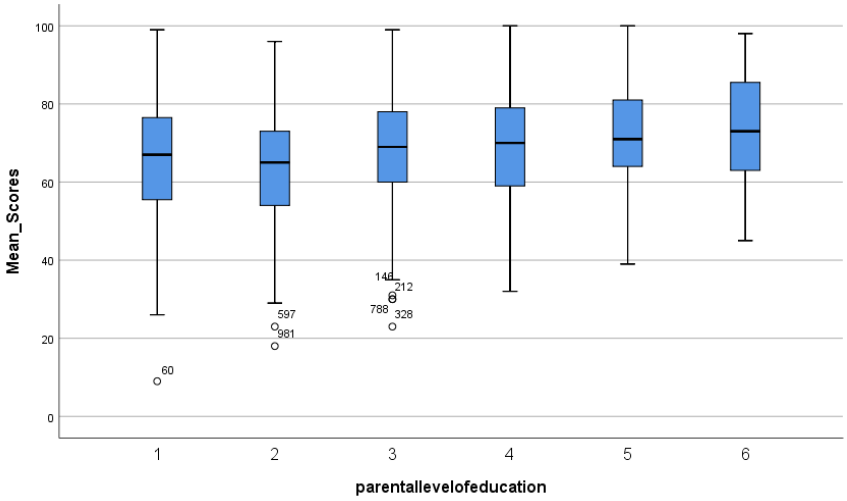

Figure 3: Box plot depicting the correlation between parental level of education and mean scores

The few outliers identified in the box plot are present in level 1, 2, 3, which is the lower level of education. Therefore, we can hypothesise that the students whose parents are highly educated on average performed better. Moreover, to argue our hypothesis, we carried out Pearson's R correlation and also Kendall's tau b and Spearman's rho correlation.

4. Discussions and Conclusion

To sum up our results, there are a few inferences we made. First being that most of the students in the Kaggle dataset did perform well on average. As stated above, most student performed around 64%- 80%. In the case processing summary table for the box plot, it states the number of cases (N) of the number of students in each parental level of education. Such that, for the level 6 which is master's degree the number of students who mentioned this level is the lowest at 59, whereas for the level 3 and level 4 which are some college and associates it has a higher occurrence which is 226 and 222 respectively. Hence this shows there are very few parents with master's level of education.

We will now discuss the Pearson's r correlation carried out where it calculated a correlation coefficient of 0.206, which is closer to 0. Therefore, we believe that there is a weak relationship between the parental level of education and the mean scores of students.

Secondly, when we make an inference from the scatter plot diagram, the line sloped upwards hence also depicting a positive correlation.

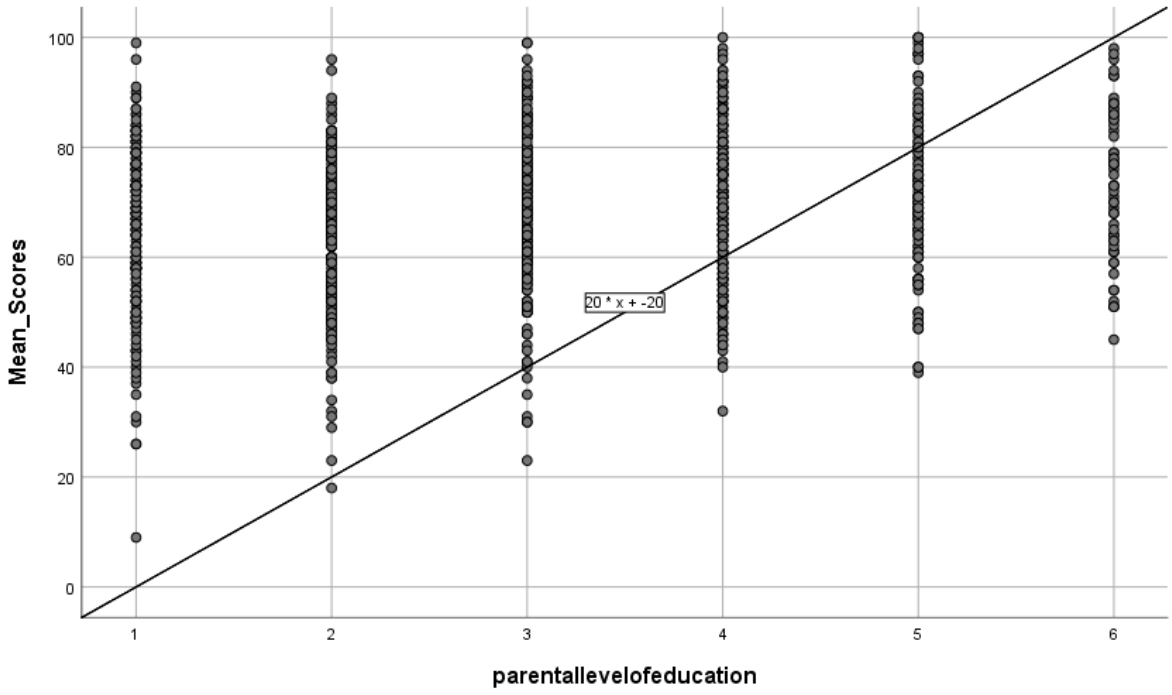

Figure 4: Scatter plot depicting the correlation between parental level of education and mean scores

The Kendall's tau b and Spearman's rho correlation calculated a 0.138 and 0.187 correlation, respectively. Hence, this further enforces that since the values are closer to 0, therefore there is a weak positive relationship between the parental level of education and mean scores.

The Kaggle dataset did not provide information as from which school they obtained the results from and also did not provide the grading system of the schools. Moreover, the dataset described the race/ethnicity as group A or B not defining clearly what the groups mean. Due to some of the limitations in the dataset, we cannot decisively prove that there is a correlation between the parental level of education and student performance.

As to make the analysis more enforced we would like to make some recommendations; one is the clear indication of the form of data acquisition and secondly indicating the school and state from which the information was obtained to be sure whether the correlation is the same for all school within the United States or only some particular states. Moreover, this will add significance to the analysis to realise how poverty, socioeconomic status, and family structure impact students' performance.

5. References

1. Barge, J.K. and Loges, W.E. (2003). Parent, Student, and Teacher Perceptions of Parental Involvement. Journal of Applied Communication Research, 31(2), pp.140–163.
2. Kaggle.com. (2019). Students Performance in Exams. [online] Available at: https://www.kaggle.com/spscientist/students-performance-in-exams/version/1#StudentsPerformance.csv [Accessed 19 Dec. 2019].
3. Kloosterman, R., Ruiter, S., De Graaf, P.M. and Kraaykamp, G. (2009). Parental education, children's performance and the transition to higher secondary education: trends in primary and secondary effects over five Dutch school cohorts (1965-99). The British Journal of Sociology, 60(2), pp.377–398.
4. Lib.dr.iastate.edu. (2019). [online] Available at: https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1428&context=rtd [Accessed 18 Dec. 2019].
5. Longinus Chukwudi, O. (2017). Influence of Parental Occupation and Level of Education on Academic Performance of Accounting Students in Nigeria.
6. Okpala, C., Okpala, A. and Smith, F. (2019). Parental Involvement, Instructional Expenditures, Family Socioeconomic Attributes, and Student Achievement: The Journal of Educational Research: Vol 95, No 2. [online] Tandfonline.com. Available at: https://www.tandfonline.com/doi/pdf/10.1080/00220670109596579?needAccess=true [Accessed 12 Dec. 2019].
7. Park, H. (2019). *Home literacy environments and children's reading performance: a comparative study of 25 countries: Educational Research and Evaluation: Vol 14, No 6*. [online] Tandfonline.com. Available at: https://www.tandfonline.com/doi/pdf/10.1080/13803610802576734?needAccess=true [Accessed 18 Dec. 2019].
8. Walker, J. and Smrekar, C. (2019). Parenting - Influence Of Parents' Level Of Education. [online] Education.stateuniversity.com. Available at: https://education.stateuniversity.com/pages/2311/Parenting-INFLUENCE-PARENTS-LEVEL-EDUCATION.html [Accessed 12 Dec. 2019].

6. Appendix

Correlations

		Mean_Scores	parentallevelofeducation
Mean_Scores	Pearson Correlation	1	.206**
	Sig. (2-tailed)		.000
	N	1000	1000
parentallevelofeducation	Pearson Correlation	.206**	1
	Sig. (2-tailed)	.000	
	N	1000	1000

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5: Pearson's r correlation

Correlations

			Mean_Scores	parentallevelofeducation
Kendall's tau_b	Mean_Scores	Correlation Coefficient	1.000	.138**
		Sig. (2-tailed)	.	.000
		N	1000	1000
	parentallevelofeducation	Correlation Coefficient	.138**	1.000
		Sig. (2-tailed)	.000	.
		N	1000	1000
Spearman's rho	Mean_Scores	Correlation Coefficient	1.000	.187**
		Sig. (2-tailed)	.	.000
		N	1000	1000
	parentallevelofeducation	Correlation Coefficient	.187**	1.000
		Sig. (2-tailed)	.000	.
		N	1000	1000

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 6: Kendall's tau_b and Spearman's rho correlation