KELVIN NJUKI
STA 651, MODULE 1, PROJECT
SEPTEMBER 23RD, 2020

DOES GENDER, EDUCATION LEVEL, AND PARTICIPATION IN DISTANCE LEARNING AFFECT NUMERACY SCORE?

ABSTRACT

"Good numeracy is the best protection against unemployment, low wages, and poor health." Andreas Schleicher – OECD. Having good numeracy skills can as well improve individual self-esteem. Numeracy can be affected by various factors such as age, gender, education level, race, distance learning, among others. Using PIAAC data collected between 2011 – 2018, this study aims at investigating whether numeracy scores are affected by an individual's gender, education level, and participation in distance learning. Also, the study finds out which combination of gender, education level, and distance learning levels, have higher numeracy score. Analysis results showed that gender, education level, and an interaction between education level and distance learning significantly affected numeracy scores. A male with at least college education and had participated in distance learning scored highest while a female with less than a college education, and had not participated in distance learning, scored the least.

INTRODUCTION

The Program for the International Assessment of Adult Competencies (PIAAC) is a program of assessment and analysis of adult skills. The adult skills survey is a major survey conducted by PIAAC. It measures adults' proficiency in key information-processing skills – literacy, numeracy, and problem-solving. The survey is conducted in over 40 countries/economies and measures the key cognitive and workplace skills needed for individuals to participate in society and for economies to prosper. The survey is administered every 10 years. The first cycle had three rounds of data collection between 2011 – 2018. In this cycle, adults were surveyed in 24

participating countries in 2011, 9 additional countries in 2014, and 6 more countries in 2017. Among the key variables reported in the data are numeracy, literacy and problem-solving scores, age, gender, education level, distance learning, race, income level, and employment status. This study focusses on data collected for the United States only. It aims at answering two statistical questions based on the data collected.

i) Does numeracy score gets affected by gender, education level, and distance learning?

ii) Which combination of gender, education level, and distance learning yield a higher numeracy score?

Numeracy is defined as the ability to use, apply, interpret, and communicate mathematical information and ideas. It is an essential skill in an age when individuals encounter an increasing amount and wide range of quantitative and mathematical information in their daily lives.

EXPLORATORY DATA ANALYSIS

Exploratory data analysis was essential to gain insight into the data. Visualizing the relationship between numeracy score, gender, education level, and distance learning was done using a comparative bar chart. The distribution of numeracy scores for different combinations of gender, education levels, and distance learning levels, were done using combined violins and box plots.

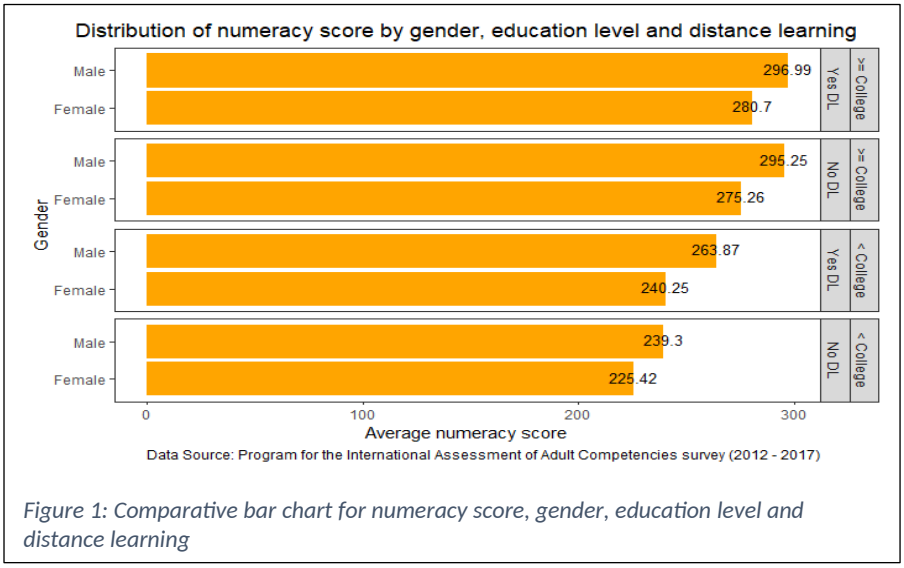

Figure 1: Comparative bar chart for numeracy score, gender, education level and distance learning

The chart shows a clear difference in average numeracy scores for different combinations of gender, education level and distance learning levels. Males who attended distance learning classes and had at least college education scored highest while females who did not attend distance leaning classes and had less than college education scored the least.

The plot exhibits a clear disparity in the median of numeracy score for different combinations of gender, education level and distance learning levels. Further, the numeracy scores are normally distributed across these combinations.

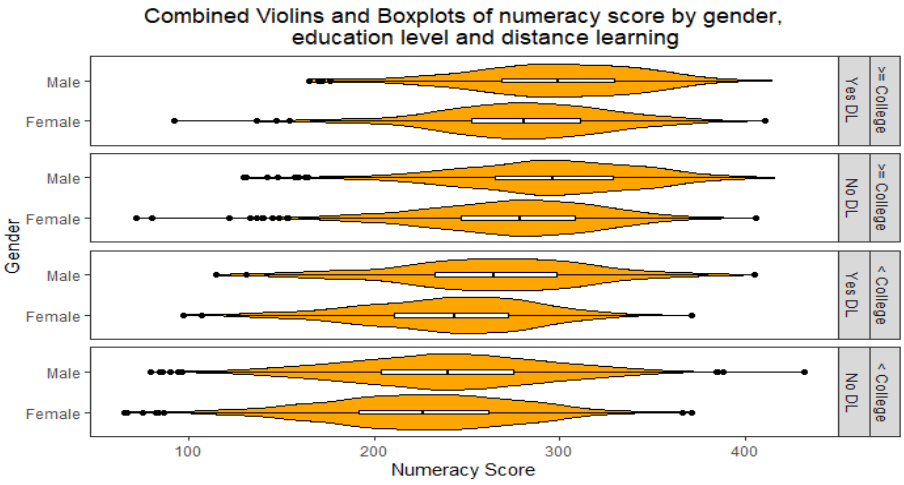

Figure 2: Combined Violins and Boxplots of numeracy score by gender, education level and distance learning

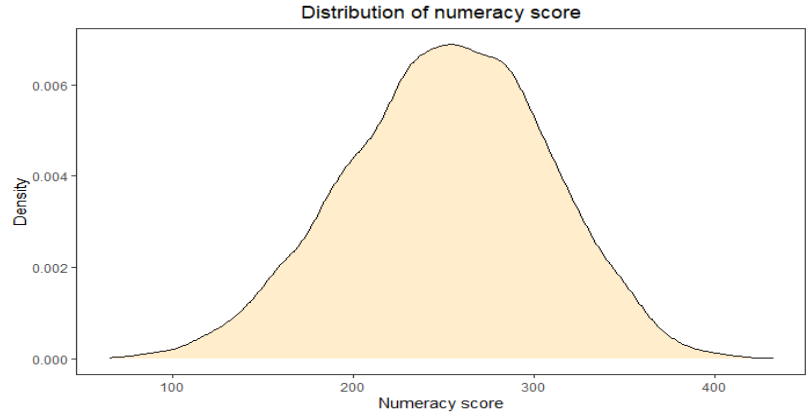

Figure 3: Distribution of numeracy score

Further, exploration of the distribution of numeracy score was done. Numeracy scores were found to be normally distributed as suggested by the plot. Therefore, a normal general linear regression model with identity link would be appropriate when modelling the data.

ANALYSIS METHODS AND SUMMARY OF FINDINGS

i) Analysis Methods

Based on exploratory data analysis, a normal generalized linear model was appropriate to model the numeracy scores as a function of gender, education level, and distance learning. That is.

$$Y = \beta_0 + \beta_1 G + \beta_2 EL + \beta_3 DL + \beta_4 G*EL + \beta_5 G*DL + \beta_6 EL*DL + \beta_7 G*EL*DL + \varepsilon \text{ where}$$

$$\mathbf{G = Gender} = \begin{cases} 1, \text{If gender is male} \\ 0, \text{If gender is female} \end{cases} \quad \mathbf{EL = Education\ Level} = \begin{cases} 1, \text{If} < \text{College} \\ 0, \text{If} >= \text{College} \end{cases} \quad \mathbf{DL = Distance\ Learning} = \begin{cases} 1, \text{If No} \\ 0, \text{If Yes} \end{cases}$$

A model selection and diagnostic were done to come up with a reduced validated model as discussed in the next section.

ii) Model Selection and Diagnostics

The model selection procedure was conducted on the above model using a forward selection procedure with the Bayesian Information Criterion (BIC). A model with the three factors and an interaction term was found to be the most appropriate, that is

$$\widehat{Num.score} = 280.668 + 16.372\,G - 37.433\,EL - 3.800\,DL - 15.199\,EL * DL$$

Model-checking/diagnostics were conducted to validate the model.

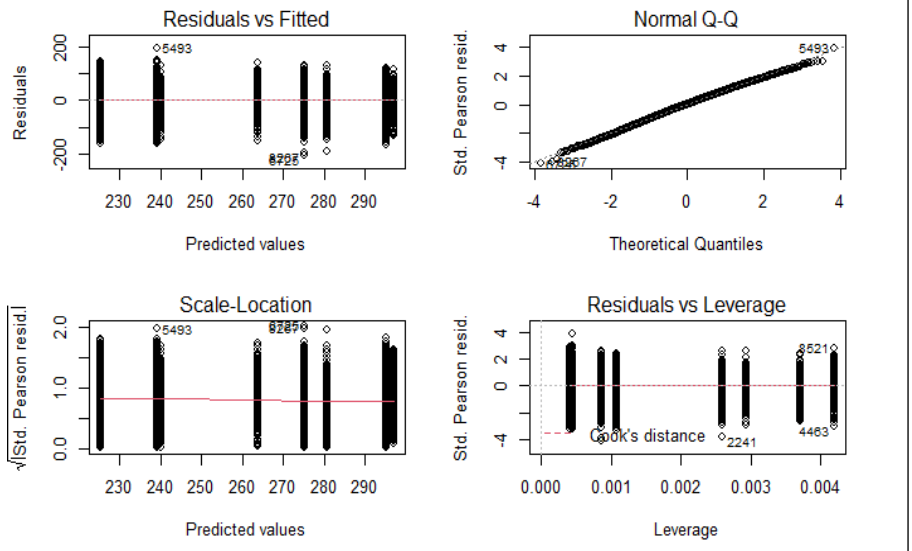

Figure 4: Model diagnostics

The top-left plot indicates homogeneity of variance for the residuals. Top-right plots indicates that the residuals are normally distributed. The bottom left emphasizes on homogeneity of variance using standardizes residuals.

To find out whether the three factors of interest significantly contributed to numeracy scores, the following hypotheses were tested using the Wald test.

H_0: Gender had no significant effect on numeracy score when education level, distance learning, and the interaction term were already in the model ($\beta_1 = 0$)

H_0: Education level had no significant effect on numeracy score when gender, distance learning, and the interaction term were already in the model ($\beta_2 = 0$)

H_0: Distance learning had no significant effect on numeracy score when gender education level and the interaction term were already in the model ($\beta_3=0$)

H_0: The interaction between education level and distance learning had no significant effect on numeracy score when gender and education level were already in the model ($\beta_4=0$)

iii) Summary of Findings

Testing the hypotheses using the Wald test

Model Parameter	Wald Test P-Value
β_1	0.0
β_2	0.0
β_3	0.085
β_4	8.9e-07

Table 1: Wald test p-values

The Wald test indicated that gender and education level were important when the other two factors and the interaction term were in the model. The interaction between education level and distance learning significantly contributed to the numeracy score when gender and distance learning were in the model. Distance learning did not have a significant effect on its own when gender and education level were in the model.

We further fitted models for the eight different combinations of gender, education level, and distance learning levels. The models report that males with at least college formal education and had participated in distance learning performed better in numeracy tests. Females with less than college formal education and did not participate in distance learning scored the least in the numeracy test.

Gender	Education Level	Attended Distance Learning	Expected Numeracy Score
Male	>=College	Yes	297.04
Male	>=College	No	293.24
Female	>=College	Yes	280.67
Female	>=College	No	276.87
Male	<College	Yes	259.61
Female	<College	Yes	243.24
Male	<College	No	240.61
Female	<College	No	224.24

Figure 4: Individual models for different combinations of levels for gender, education level, and distance learning

DISCUSSION

The analysis results indicate that gender, education level, and distance learning had a significant effect on numeracy scores. Male participants with at least a college formal education and had attended distance learning performed had the highest score. Females with less than a college education and had not participated in distance learning had the least score. Other factors such as age and race can be incorporated in the model in the future to find out their contribution.

CONCLUSION

The numeracy score is affected by the individual's level of education, gender, and participation in distance learning as suggested by the analysis. "Good numeracy is the best protection against unemployment, low wages, and poor health." Andreas Schleicher – OECD. This tells that the United States which was the focus in this analysis would need to invest much in education to cut unemployment rates and improve the country's health.

Reference

1. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
2. http://www.oecd.org/skills/piaac/
3. https://www.nationalnumeracy.org.uk/why-numeracy-important