Analysis of Student Performance

Sung Hee Hong
and Timothy Lee

Due 12/11/18

Introduction

Our data set is the collection of students circumstances, such as romantic status, internet access, family support, etc., and achievements in Portuguese language and math scores, in two different Portuguese secondary education schools. The goal of this report is to identify if any of the parameters gathered has a significant effect on the success rate of a student in mathematics. We attempt to estimate exactly how effective these circumstances are to a students success, and thereby draw conclusions on what other information should be gathered, or what else to test.

Questions of Interest

We consider the following research questions:

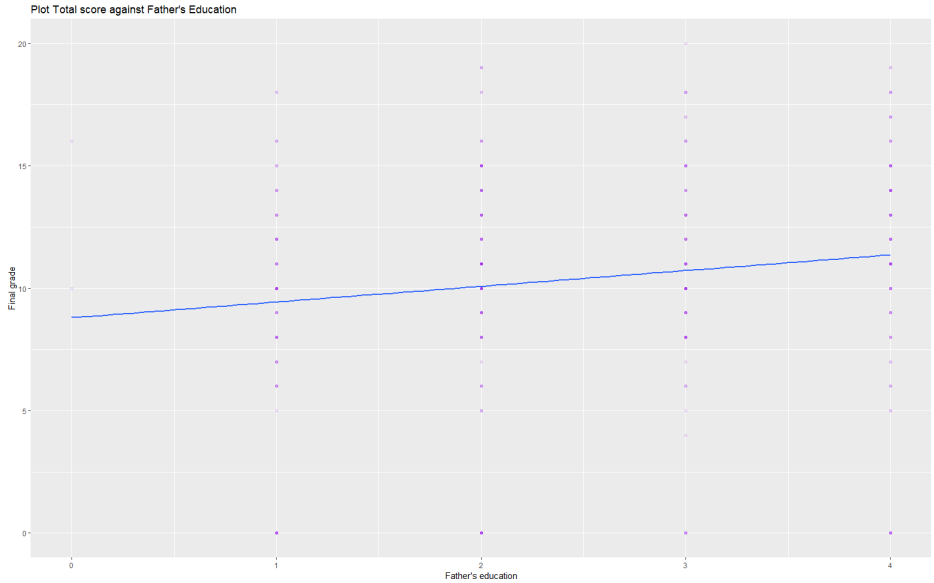
1. Does the mother's education or the fathers education have a bigger impact on their childs success?
2. Does the gender of the child play affect how much their mother or fathers education impacts their academic success?
    - Does the mothers education impact the female students academic success?
    - Does the mothers education impact the male students academic success?
    - Does the fathers education impact the male students academic success?
    - Does the fathers education impact the female students academic success?

Regression Method

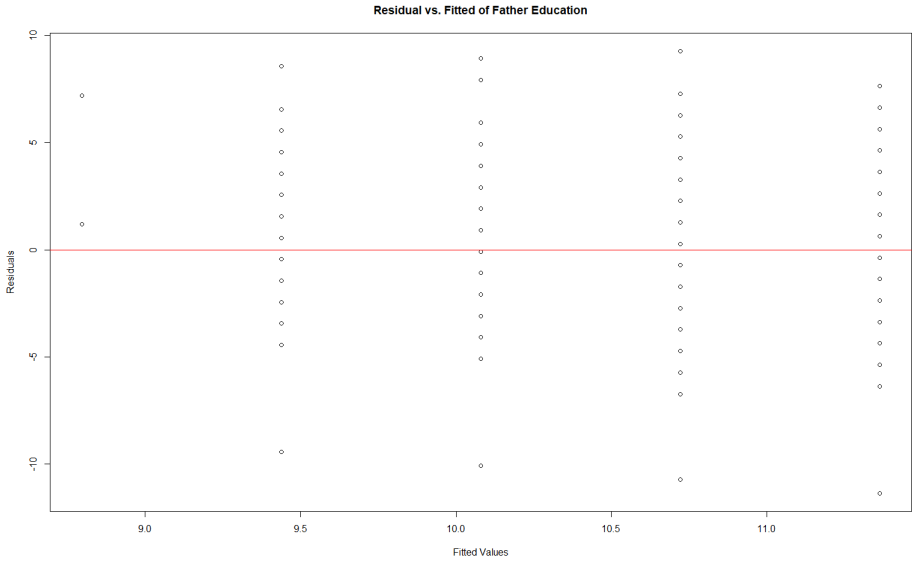
To answer our questions, we plot the final scores of each student as the response, and the respective parent's education level as the predictor, and use an information criterion to help us estimate the effect of a child's parents' education.

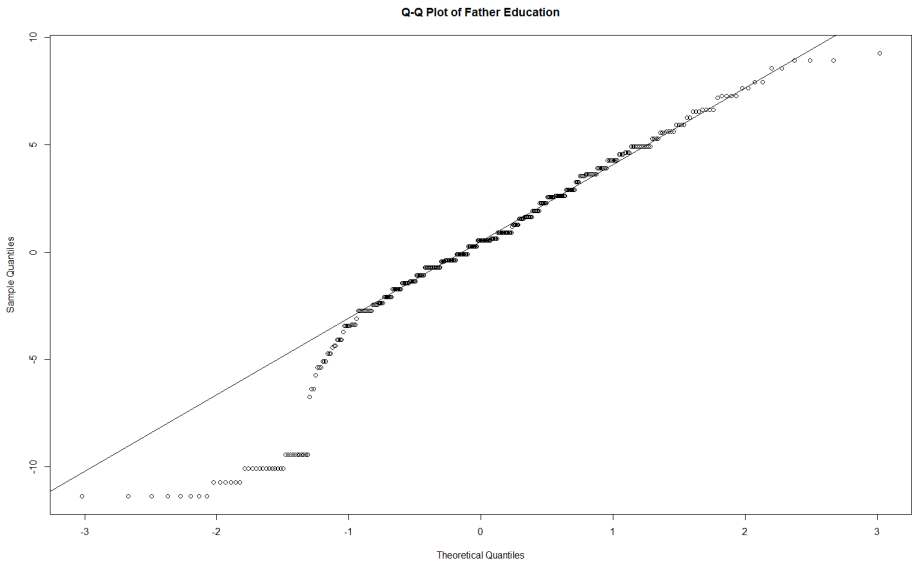
Regression Analysis: Father's Education

The predictor variable father's education level is measured from 0 to 4. Each number represents from the least education to higher education the father of the student has completed (0- no education, 1-primary education (4th grade), 2-5th to 9th grade, 3-secondary education, 4- higher education). We examine the final scores of the students, which is a score from 0 to 20. The scatterplot below shows total graph on fathers education.



To conduct linear regression analysis, the four LINE conditions must hold. By conducting the Residual cs. Fitted plot the graph succeeds to meet the criteria of linearity and equal variances. By testing the qqnorm, the plot is proven to be a normal distribution.





Research Question: Is higher education level of a child's father an indicator of greater academic success from said child?

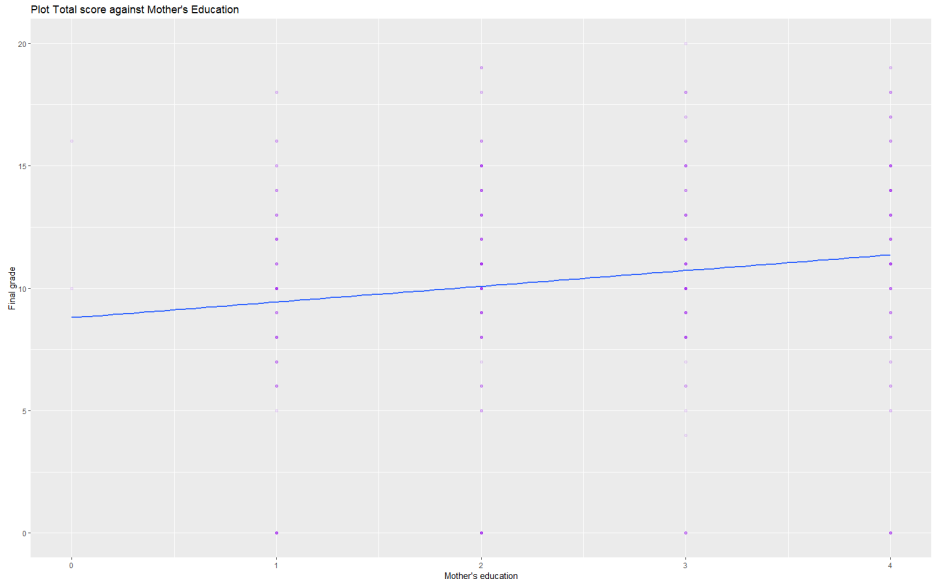
To answer this question, we conduct a t-test with the following hypotheses:

$$H_0 : \beta_{Fedu} = 0$$
$$H_a : \beta_{Fedu} \neq 0$$

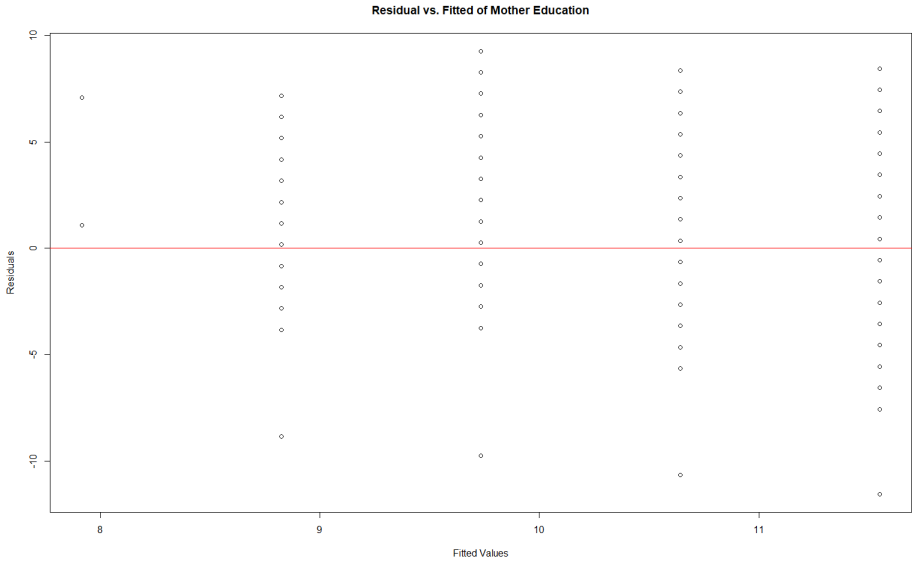
Executing this test, we obtain a t-value of 3.058, which corresponds to a p-value of 0.00238. Thus, this enables us to reject the null hypothesis of our this test at a 0.05 significance level, and we can then conclude that **a father's education level is an indicator of greater academic success of the child.**

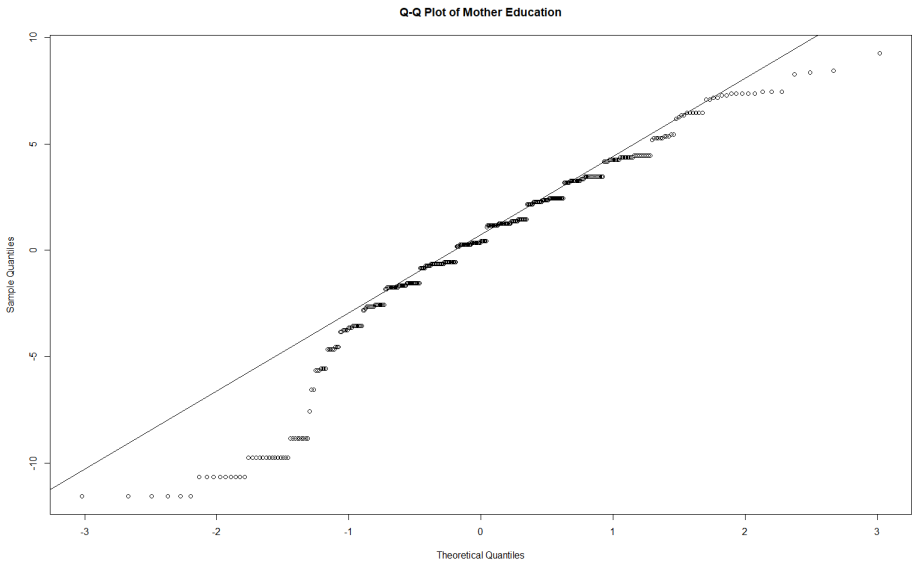
Regression Analysis: Mother's Education

For the mother's education, we do a similar method to the father's. Below is the scatterplot for the mother's education

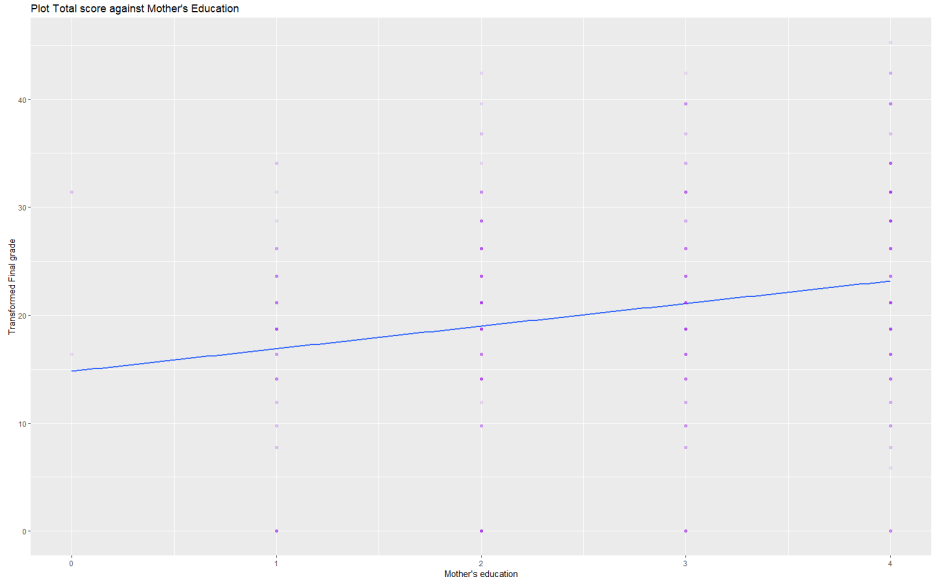


We test for the "LINE" conditions below.

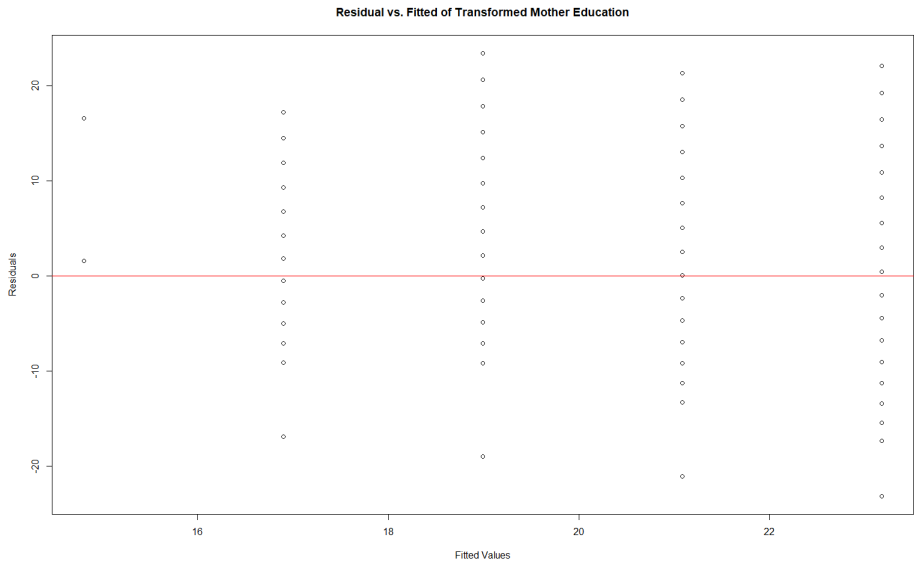


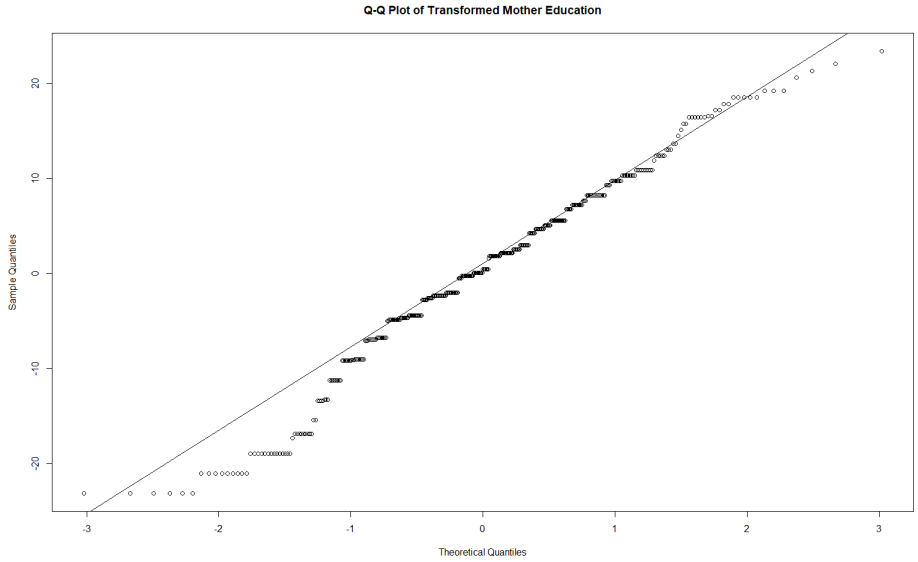


As we can see, it seems like the residuals are skewed upwards, and there seems to be a slight trend in the Q-Q plot. We solve these issues using a Box-Cox transformation, which gives us that the best power to use is 1.2727. We show the transformed scatterplot below.



We test for the "LINE" conditions of the transformed response below.





Observing these two plots, the Box-Cox transformation has successfully fulfilled the "LINE" conditions. Now, we move onto the research question.

Research Question: Is higher education level of a child's mother an indicator of greater academic success from said child?

To answer this question, we conduct a t-test with the following hypotheses:

$$H_0 : \beta_{Medu} = 0$$
$$H_a : \beta_{Medu} \neq 0$$

Executing this test, we obtain a t-value of 4.59, which corresponds to a p-value of $5.97 * 10^{-6}$. Thus, this enables us to reject the null hypothesis of our this test at a 0.05 significance level, and

we can then conclude that **a mother's education level is an indicator of greater academic success of the child.**

With this information, we can now answer the first research question we proposed.

Father's and Mother's Education Level Analysis

Does the mother's education or the fathers education have a bigger impact on their childs success?

To answer this, we simply look at the 95% confidence intervals for β_{Fedu} and β_{Medu}.

For β_{Fedu}, this is straight-forward: the interval is (0.2221,1.0617).

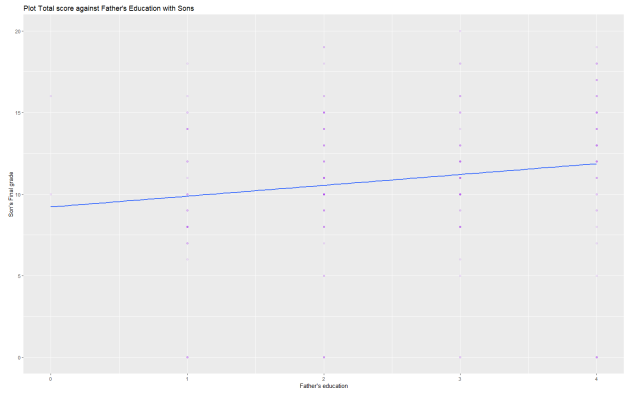
For β_{Medu}, this is a little trickier, because we used a Box-Cox transformation. While the general form of β_{Medu} is harder to show, we use a rough estimate of taking the $\frac{1}{\lambda}$ power of the transformed regression, where $\lambda = 1.2727$. With this estimate, we get the interval (1.1400,2.374431).

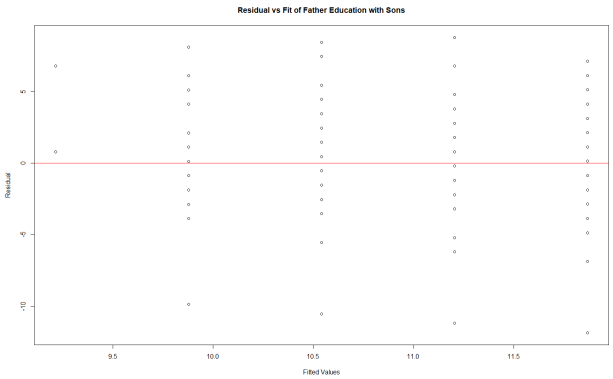
Confirming this seems to be outside of the scope of this class, but for now, we claim, with 0.05 significance level (or at least close to it), that **the mother's education level is more impactful to a child's academic success than the father's.**

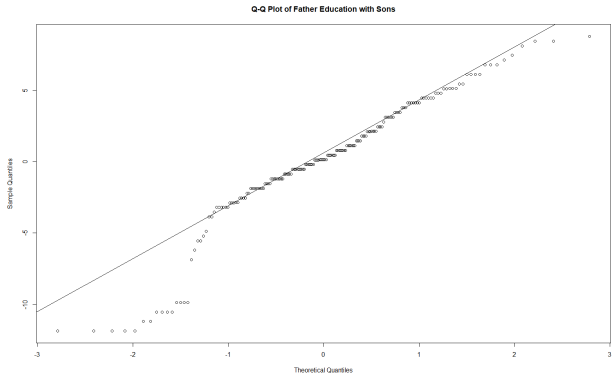
Regression Analysis: Father's Education with Sons and Daughters

Now, we want to look at how fathers' education levels impact their sons or daughters. In terms of procedure, this is all the same as before, so I will spare the details and simply put of some graphs and do the hypothesis testing towards the end.

Here are the graphs for father's education and son's performance.

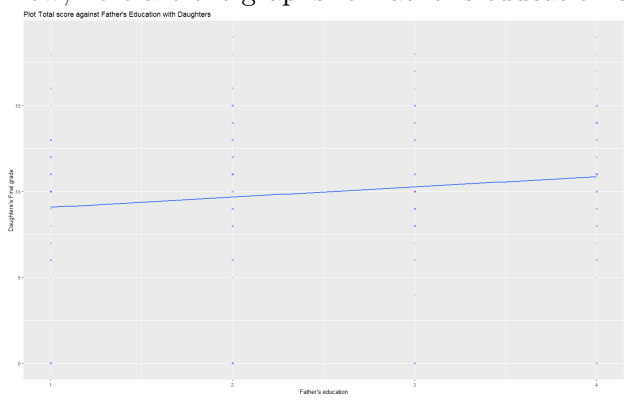


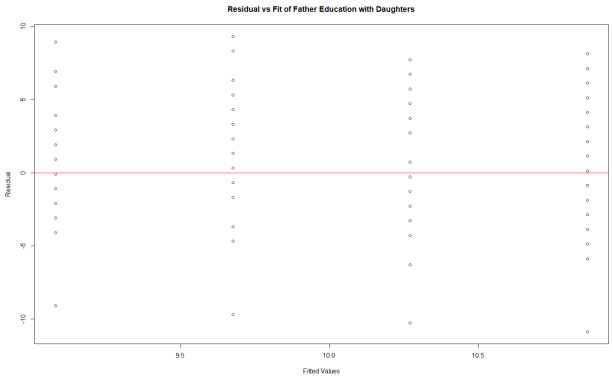


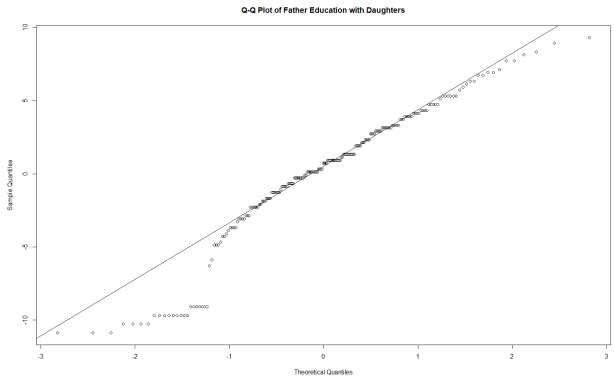


As we can see, this regression fulfills the "LINE" conditions.

Now, here are the graphs for father's education and daughter's performance.







Again, we see that the "LINE" conditions are met.

Research Question: How does a father's education level affect the academic performance of a son or daughter, specifically?

To answer this, we conduct some t-tests. Since the details are similar as before, I will simply state the results.

With a t-value of 2.212 (p-value 0.0282), we conclude at a 0.05 significance level that a father's education level does have a significant impact on a son's academic performance.

With a t-value of 2.035 (p-value 0.0432), we conclude at a 0.05 significance level that a father's education level does have a significant impact on a daughter's academic performance, as well.

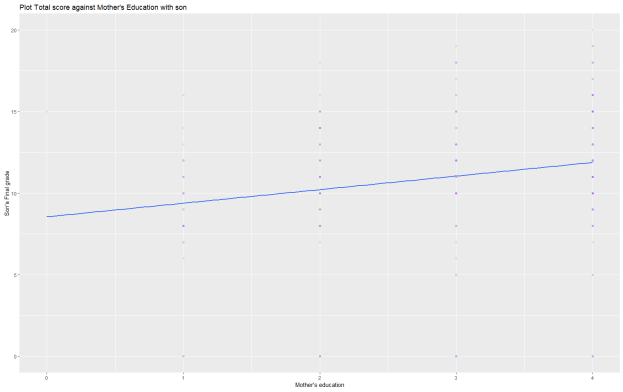
These results make sense since we concluded earlier that a father's education level has an impact on a child's grade, regardless of gender.

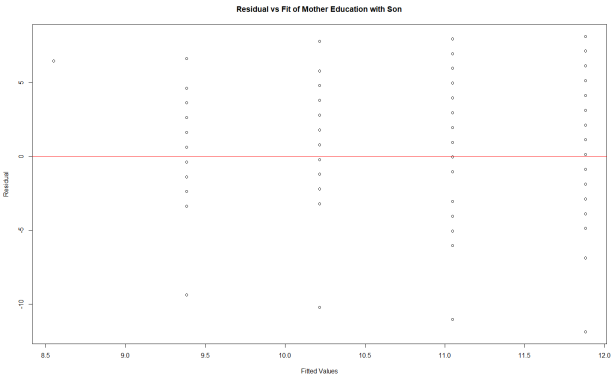
For sons, we get a 95% confidence interval of (0.0626,1.2622). For daughters, we get (0.0102,1.1802). Thus, we cannot conclude anything about the difference in the impact of a father's education depending on the gender of the child.

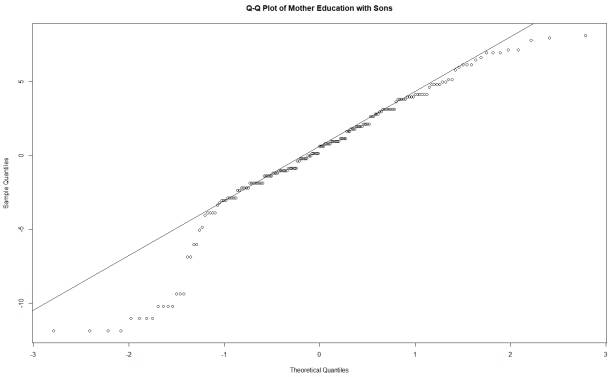
Regression Analysis: Mother's Education with Sons and Daughters

Now, we look at the mother's education level's impact on sons and daughters.

Here are the graphs for the mother's education and son's performance.

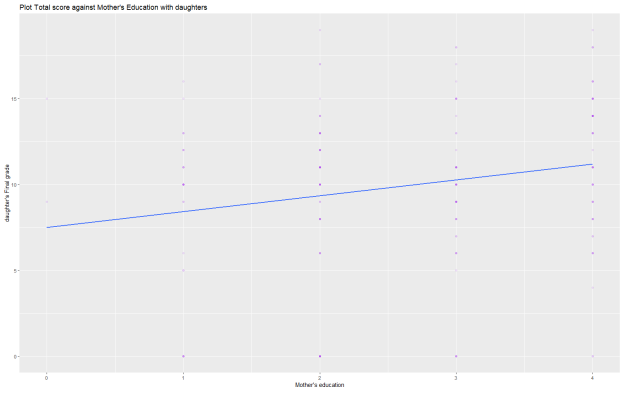


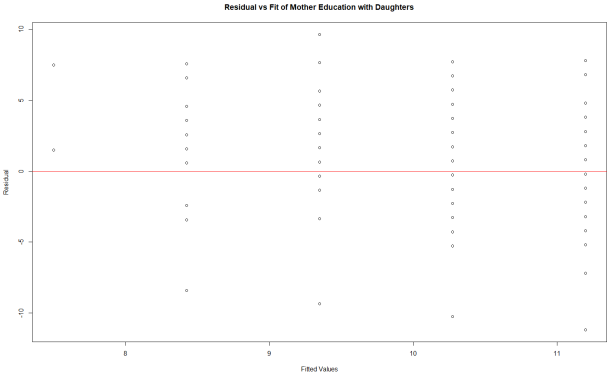


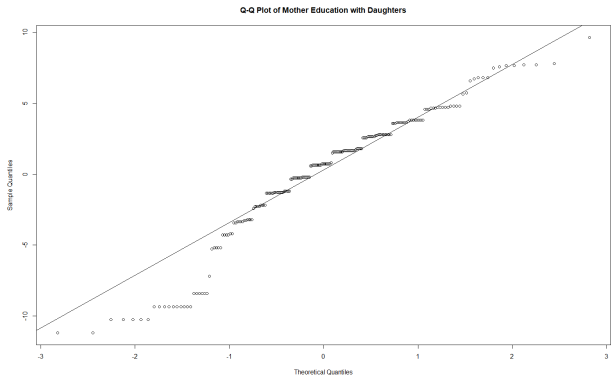


We see that the "LINE" conditions are met.

Here are the graphs for the mother's education and daughter's performance.







Again, we see that the "LINE" conditions are met.

Research Question: How does a mother's education level affect the academic performance of a son or daughter, specifically?

To answer this, like last time, we conduct some t-tests.

With a t-value of 2.830 (p-value 0.00516), we conclude at a 0.05 significance level that a mother's education level does have a significant impact on a son's academic performance.

With a t-value of 3.194 (p-value 0.00163), we conclude at a 0.05 significance level that a mother's education level does have a significant impact on a daughter's academic performance.

For sons, we get a 95% confidence interval of (0.2441, 1.4205), and for daughters, (0.3455,1.8035). Thus, we again cannot conclude anything about the difference in the impact of a mother's education depending on the gender of the child.

Fathers and Mothers Education Level Gender-Specific Analysis

Unfortunately, we cannot conclude anything significant about the differences in how a son's performance is impacted by either the mother or the father's education level, and likewise with a daughter's performance. The only thing we can conclude is that, perhaps by a small margin, a mother's education level is more impactful than the father's, which is in line with what we discovered earlier.

What is important to note, while obvious, is that for every case, we see that a parent's education level can only have a positive impact on a child's academic performance.

Conclusion

In summary, we tested and estimated the impact a parent's education level has on the academic performance of a child, considering cases of father with child, daughter with child, and the child's gender-specific case as well. We found that both the father and the mother's education level has a strictly positive impact on a child's academic performance, regardless of gender of the child or of the parent.

There are several things that could be improved upon in this analysis. For example, the intervals of a mother's impact on academic performance for gender-specific versus the gender-regardless is quite confusing, and would require some more testing to improve upon. We could have also looked at other predictors, such as primary guardian, parental status, family support, and family relationship. These predictors may have helped with understanding some of the variance we got. However, we were able to get a strong answer of mother's education level being more impactful to a child's performance than a father's, when disregarding gender, which is something very interesting to look into.