# MATH3091 Coursework

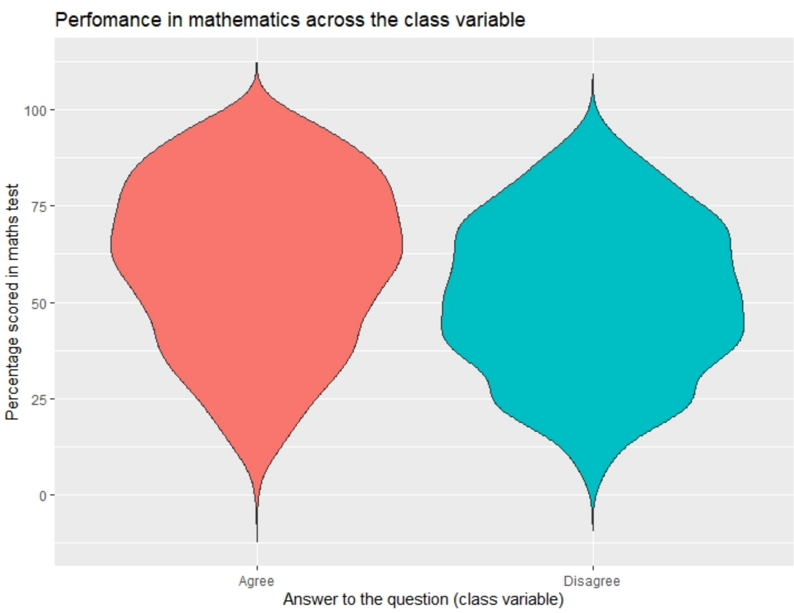
ID: 29557461

March 2019

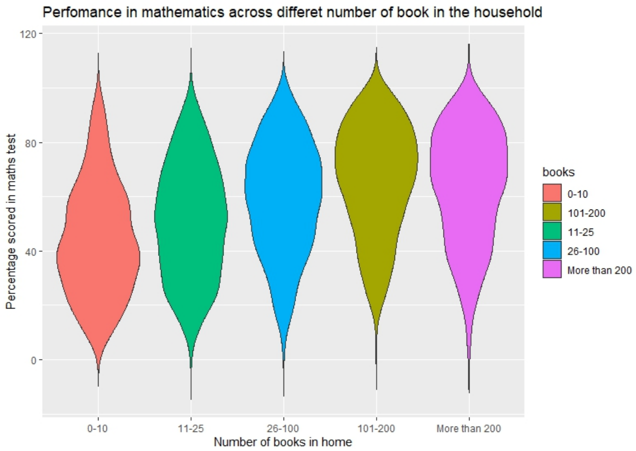
The purpose of this work is to investigate the responses to the question - "Are you good at Maths" - from 3813 pupils aged 9-10. The study was conducted in 2015 Trends in International Mathematics and Science Study. Children were also asked to complete an exam that would test their mathematical abilities. A study will also investigate the potential influence of the number of books stored in the household on pupil's answer to the statement.

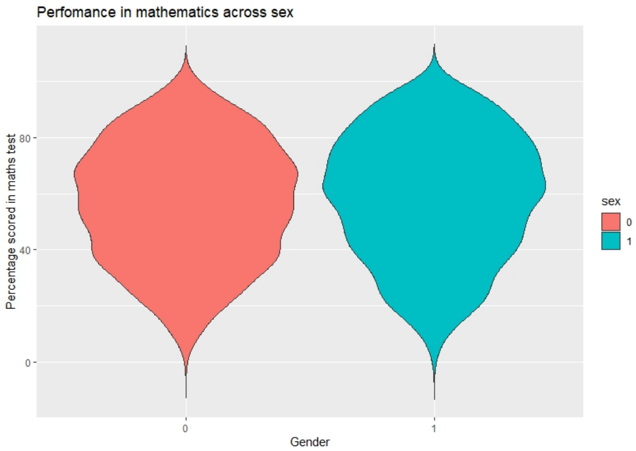
| | |
|---|---|
| good_at_maths | class variable, 1 if student didn't agree with the statement. |
| place_of_birth | binary variable, 1 if student was born in the UK. |
| sex | binary variable, 1 if student is male. |
| score | score from 0-100 on a Maths test. |
| books | number of books in the household, variable has five levels. |

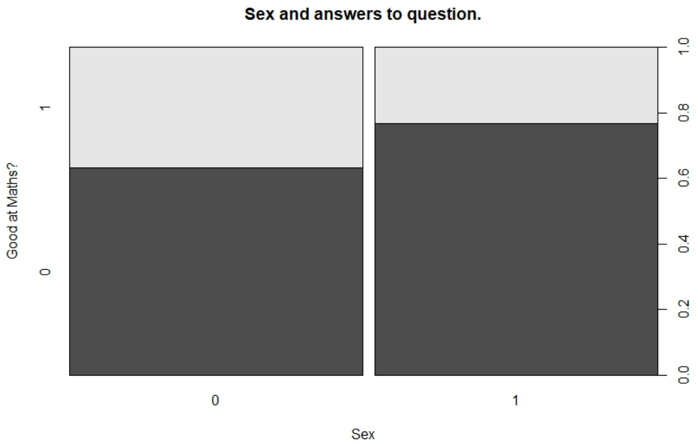
Table 1: Description of variables.




Firstly, let's investigate the distribution of variables. We can see that distribution of score is skewed and most pupils score rather high with a mean of 57.5% and a median of 59%. Looking at plots, we see that the distribution of score is significantly different for the two values of the class variable. Numerically, the mean score for pupils that agreed with the statement is equal to 60% in comparison to 51.3% for those that did not, similarly, medians are given as 62% and 52% respectively.




Another feature worth investigating is a number of books kept in the household. Following plot shows a clear dependence of score on books. However, we can also notice diminishing marginal returns - clearly, there is a huge difference between 0-10 and 26-100 books but difference dims as a number of books grow. This does not mean that number of books directly influence being confident of one's mathematical abilities, but plot shows some influence of a number of books on test scores.

We can also take a look at influence of sex on other features. The difference between test scores for boys and girls is negligible with means equal to 56.62% and 58.49% for girls and boys respectively. We have seen that gender has an influence on a variable (score) that is highly correlated with the class variable. We can take a look now at the direct influence of sex on the class variable. It is worth noting that the data set contains roughly 60% of boys' responses and 40% girls so our conclusions are meaningful.

The plot shows that around 77% of boys claim they are good at Maths in comparison to 63% of girls. A slight difference could be explainable by a slight difference in scores between genders but such a big difference of about 15% shows that boys are more confident about their skills.

We can now look at models. Starting from the most complicated model, including all of the variables and interactions we see that most of the interactions are not statistically significant. We conduct a likelihood ratio test and after setting significance level to 0.05 we are left with: score, sex, score:sex, and score:books being statistically significant. However, due to the marginality principle, we should not include books in our model and the final model becomes:

$$\log \frac{p_i}{1-p_i}) = \beta_0 + x_{1i}\beta_{SCR} + x_{2i}\beta_{SEX} + x_{1i} * x_{2i}\beta_{SCR,SEX}$$
$$\log(\frac{p_i}{1-p_i}) = 0.17 - x_{1i}0.012 + x_{2i}0.09 - x_{1i} * x_{2i}0.013$$

where $x_{1i}$ and $x_{2i}$ mean respectively score on the test (0-100) and gender (binary, 1 if male), $Y_i$ measures probability of disagreeing with the statement (1 if disagree, 0 if agree)

Interpreting this result we can see that students judge their mathematical abilities correctly and that boys are more confident about their skills. Moreover, interactions of *sex* and *score* also increases the chances of agreeing with the statement 'I am good at mathematics' as gender influence test scores. We excluded interaction score:books from the model as even though more books in household results in a better score, it does not influence confidence in judging one's abilities in mathematics.

If we take a look at higher order interactions of *score* we see that such model is still statistically significant. We can also observe that *score*'s coefficient is positive but that is weighted by the negative second interaction coefficient so that all in all this variable still increases the chances of agreeing with the statement.

$$\log \frac{p_i}{1-p_i}) = \beta_0 + x_{1i}\beta_{SCR} + x_{2i}\beta_{SEX} + x_{1i} * x_{2i}\beta_{SCR,SEX} + {x_{1i}}^2\beta_{SCR2}$$
$$\log(\frac{p_i}{1-p_i}) = -0.8387 + x_{1i}0.03119 + x_{2i}0.1961 - x_{1i} * x_{2i}0.01563 - {x_{1i}}^2 0.0004$$

Here are plots of two models where class variable depends on *score* and the second model where it depends also on second order of *score*. Note that it ranges from roughly 0.6 to 0. Even among worst students (<10% on the test), around 58% of them think they are good at the subject.

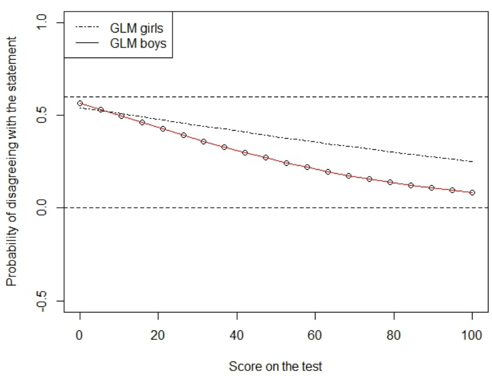



Figure 1: Simpler model.

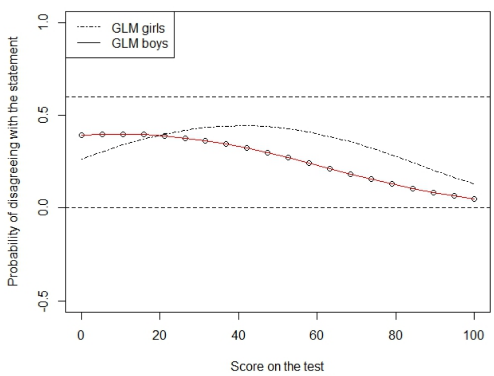



Figure 2: Model with second order of *score*