Analysis of Factors Affecting Math Test Scores

Written by Liman Wei in Nov. 2020 and Posted on https://github.com/weiliman/Course-Assignments

Introduction

Some people believed that the math test scores are determined by the individual differences of students, while other people think good schools and good teachers are the keys for high math test scores. In this report, we are discussing what are main factors influencing math test scores of students, and how these factors affect the math test scores. The data we are using contains more than 3000 observations of over 1000 students from different junior schools.

Abstract

Using bayesian inference, we conducted a generalized linear mixed model to analyze the major factors influencing students' math test scores on a sample of over 3000 observations from more than 1000 students over 3 years. The result shows individual level difference is the most important factor, class level difference will somewhat influence the test scores, while school level difference has almost no effects on test scores. Therefore, we conclude the best way to enhance students' math test scores is identifying weak students and provide them with extra training, rather than offering extra help to poorly performing schools and classes.

Methods

The sample data includes 3236 records on over 1000 junior school students measured over 3 school years. It includes the following variables: gender, social class, school, class, grade, student ID, and math test score of each student.

The model we are using is a linear mixed model based on bayesian inference. From Figure 1 we can see the number of math test questions students get wrong (our response variable) seems to follow a poisson distribution, and hence we build our model as:

$$Y_{ijk} \sim Poisson(\lambda_{ijk}) \qquad log(\lambda_{ijk}) = \mu + X_{ijk}\beta + U_i + V_{ij} + W_{ijk}$$

with prior distributions:
(fixed effects) $\mu \sim N(0, 0.2), \beta \sim N(0, 10)$;
(random effects) $U_i \sim N(0, \sigma_1), V_{ij} \sim N(0, \sigma_2), W_{ijk} \sim N(0, \sigma_3)$ and $P(\sigma_i > 0.2) = 0.5$ for i = 1, 2, 3

where Y_{ijk} represents the number of wrong math test answers of student k at school i in class j, X_{ijk} is the covariates containing variables gender, social class and grade, U_i represents the school level difference, V_{ij} represents the class level difference and W_{ijk} represents the individual student level difference.

We used tables to show the posterior quantiles of fixed effects (Table 1) and standard deviations of random effects (Table 2). We also plotted both prior and posterior distributions for these variables (Figure 2 and Figure 3).

Results

From Table 1 and Table 2, we can see the individual level variation is the largest effect on the number of questions students get wrong. The second largest effect is the grade difference of students, students from higher grade would get more questions correct. The class level difference have nearly the same magnitude of influence on test scores as social class difference. Student from lower social classes are more likely to get more questions wrong. The school level difference and gender difference are the least important factors and have almost zero influence on math test scores.

Discussion

It turns out the individual effect is far larger than class level effect and school level effect. Therefore, to enhance the math test scores of students, the best strategy would be providing extra help to individual weak students. The second best strategy is, identifying students from lower social classes, or identifying poorly performing classes and providing extra training/funding to these students. Since school effect and gender effect is too small, offering extra help to some schools or a specific gender would not increase the math test scores of students.

Appendix: Plots and Tables

Distribution of y

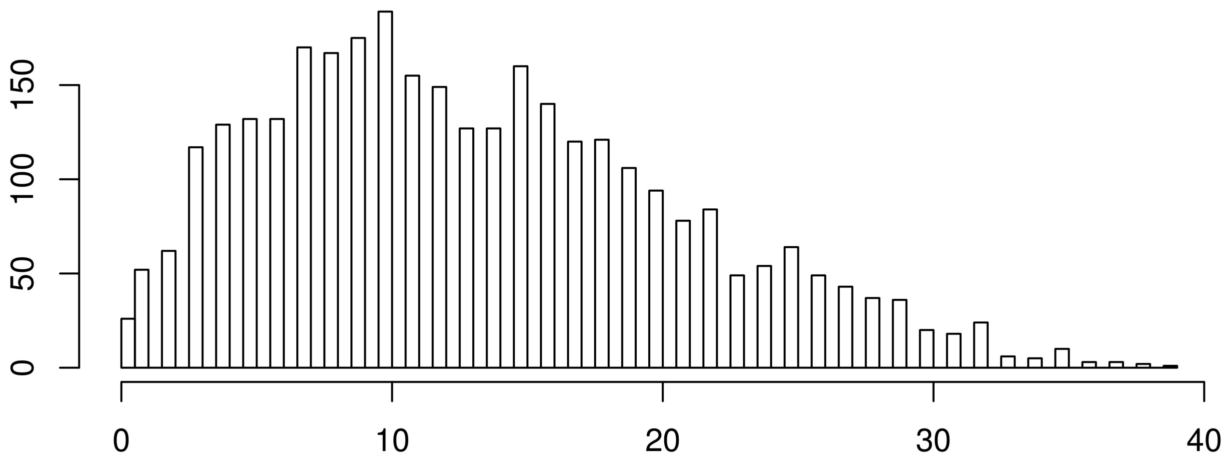

Figure 1: Histogram of the number of questions students get wrong

Tables of Posterior Quantitles

Table 1: Posterior quantiles of the fixed effects

	0.5quant	0.025quant	0.975quant
(Intercept)	2.502	2.377	2.626
genderm	-0.003	-0.060	0.055
socialClassII	-0.174	-0.308	-0.039
socialClassIIIn	0.013	-0.131	0.157
socialClassIIIm	0.137	0.017	0.258
socialClassIV	0.096	-0.047	0.238
socialClassV	0.215	0.067	0.363
socialClasslongUnemp	0.162	0.006	0.318
socialClasscurrUnemp	0.172	-0.042	0.385
socialClassabsent	0.166	0.035	0.297
grade1	-0.002	-0.023	0.020
grade2	-0.420	-0.447	-0.394

Table 2: Posterior quantiles of the random effects

	0.5quant	0.025quant	0.975quant
SD for school	0.041	0.008	0.105
SD for classUnique	0.181	0.138	0.238
SD for studentUnique	0.457	0.434	0.481

Plots of posterior distributions comparing to prior distributions

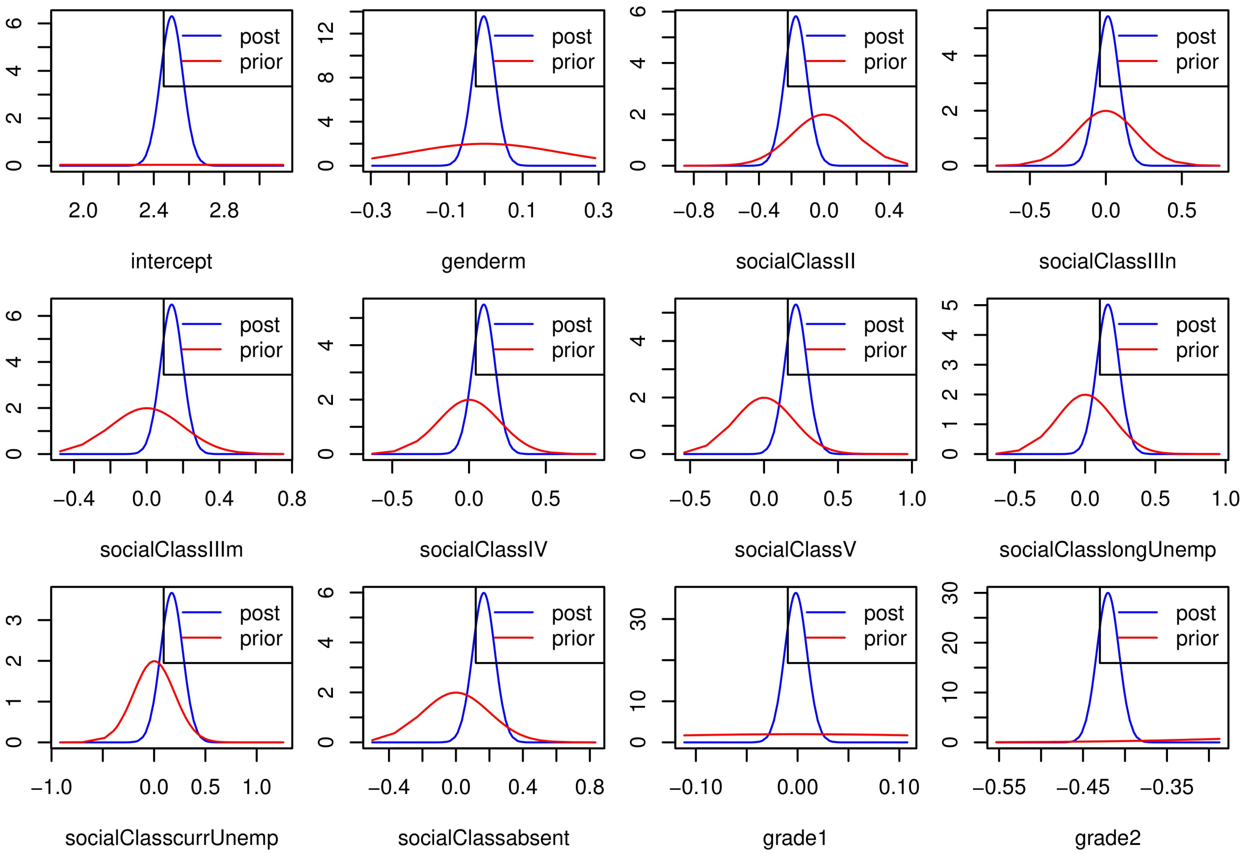

Figure 2: Priors and Posteriors for fixed effects

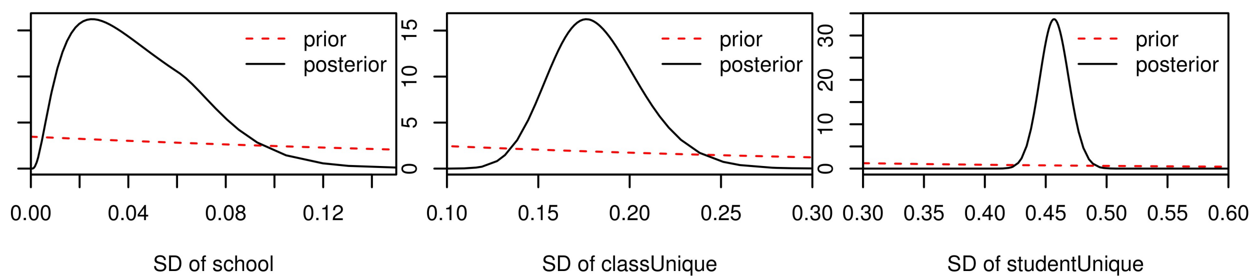

Figure 3: Priors and Posteriors for SD of random effects