

Comments from the peer reviewer

We received several suggestions from the peer reviews.

From the first peer, the peer advised that our paper should explain about the background of the model more deeply. We will partially accept this suggestion. Our intended readers already know what the Rasch model is, so we do not have to explain the basic backgrounds of the models, how `bayesglm()` and `glm()` functions are different. However, we shall distinguish the difference of the models more clearly in the paper. The peer claimed we should demonstrate on how to calculate AIC. We would not accept this because our intended audience have basic knowledge of statistics and regression modeling.

From the second peer, the peer suggested to move the description of the data to the method section. We will accept this advice. The peer also suggested to change the position of the table. We will accept this advice. The peer argued that we should add background knowledge about what MLE is and which MLE is the best. We would not accept this comment because, again, our intended audience are trained in basic statistics.

Introduction

Security and strict disciplinary codes in high schools have been done to prevent, or discourage, adolescent students from his/her misconduct. Although adolescence is an important period greatly influenced by the environment and regulations are useful to preserve orderliness and safety, excessive regulations sometimes negatively affect students emotionally, or physically. Undoubtedly, that we need to study more about the school security for growing students and to analyze how school security affects the student, it is necessary to quantify the ‘security measures’ for each school. The goal of this paper is to make relative quantified security measurement scores for each school.

In 2014 Finn and Servos studied security and student discipline practice. The research examined the relationship among student misbehavior, suspensions, and security measures in a nationwide sample of high schools. From the data, Finn and Servos tried to identify the characteristics of schools that implemented the most invasive security measures, and examine the conditions related to racial/ ethnic and gender inequities in suspensions (Finn and Servos, p. 1). In this paper, we first reconstructed the model similar to the one used by Finn and Servos, and the variations of the model that might improve the model. Next, we compared those models and selected the recommended model. Finally, we created school indices from the original Finn and Servos model and the recommended model and analyzed how two models showed different results.

Method

The data Finn and Servos used is the Education Logistic Study of 2002 (ELS:2002) conducted by the National Center for Education Statistics. From twenty security-related data elements from the responses to the ELS:2002, Finn and Servos selected seven questions to construct their data; Q38c) Require students to pass through metal detectors each day; Q38d) Perform one or more random metal detector checks on students; Q38f) Use one or more random dog sniffs to check for drugs; Q38g) Perform one or more random sweeps for contraband (e.g., drugs or weapons) but not including dog sniffs; Q38h) Require drug testing for any students (e.g., athletes); Q38n) Use one or more security cameras to monitor the school; Q39a) Have a formal process to obtain parent input on policies related to school crime and discipline. In this paper, we first used the same data to replicate Finn and Servos's result, and selected the other 6 questions, overlapping with Finn and Servos's questions, to find a better dataset to analyze. In particular, we removed Q38d and Q39a from Finn and Servos's questions and add one another question; Q38j) Enforce a strict dress code. From now, we will call those two data sets 'Finn and Servos's data set' and 're-selected data set'.

We first modified the original data to construct Finn and Servos data set and our re-selected data set. We calculated the sum of the index of schools from the Finn and Servos's data and we fit the Rasch model by using the maximum likelihood and a penalized variant of maximum likelihood for the Finn and Servos data to replicate the Finn and Servos's result. We used Bayes Method to construct a penalized variant of maximum likelihood, a model fits a logistic model combining the same logistic likelihood as generalized linear model, but with a multivariate Cauchy prior distribution on the coefficients. Then we fit the same models, maximum likelihood, and penalized variant of maximum likelihood, for our re-selected data

described above. To compare those four models, two for Finn and Servos and two for re-selected data, we compared the AIC and Deviance of the models and determined the best model among the four models.

To achieve our initial purpose, the quantification of the 'security measures', we scored each school's security for Finn and Servos's model and our determined best model. We indexed the schools in order of the security measurement scores and analyzed how the indices differ for the Finn and Servos's models and determined model. To deal with the large indices for the analysis, we constructed a visual graph that compares the indices of two models and calculated the correlation of the two models.

We also tried to replace the missing values in the data. We derived two models, one which replaced the missing values with the average of that variable for all data, and a model replaced the missing values by an average only from data with the same value for the determined independent variable. We used 'BYURBAN, a variable representing school urbanicity, as an independent variable since we determined this variable will be the most efficient variable to state the MAR.

Result

We first removed the rows that contained any missing data and calculated the sum of the index level of each school for Finn and Servos's data set by simply add the scores of each question selected for each school. As a result, we removed 44 school data out of 656 schools that contained the missing data and constructed a table of the sum of indices for each school. The result is described in Table 1.

Table 1. Table of sum of indices for each school, from Finn and Servos's data

Sum of indices	0	1	2	3	4	5	6	7
Number of schools	81	146	184	119	64	16	2	0

The AIC for model_0, maximum likelihood method model for the Finn and Servos's data set (which we also called as Finn and Servos's model), was 4471.554 and the Deviance for model_0 was 3235.554. The AIC for model_1, penalized variant for maximum likelihood method model for the Finn and Servos's data set, was 4846.643 and the Deviance for model_1 was 570.5367. The AIC for model_2, maximum likelihood method model for the re-selected data set, was 4245.643 and the Deviance for model_2 was 3011.735. Finally, the AIC for model_3, penalized variant for maximum likelihood method model for the re-selected data set, was 4628.027 and the Deviance was 528.6911. Overall, the model used penalized variant for maximum likelihood method showed higher AIC and lower Deviance than the model used maximum likelihood method. Also, the model from the re-selected data set showed lower AIC

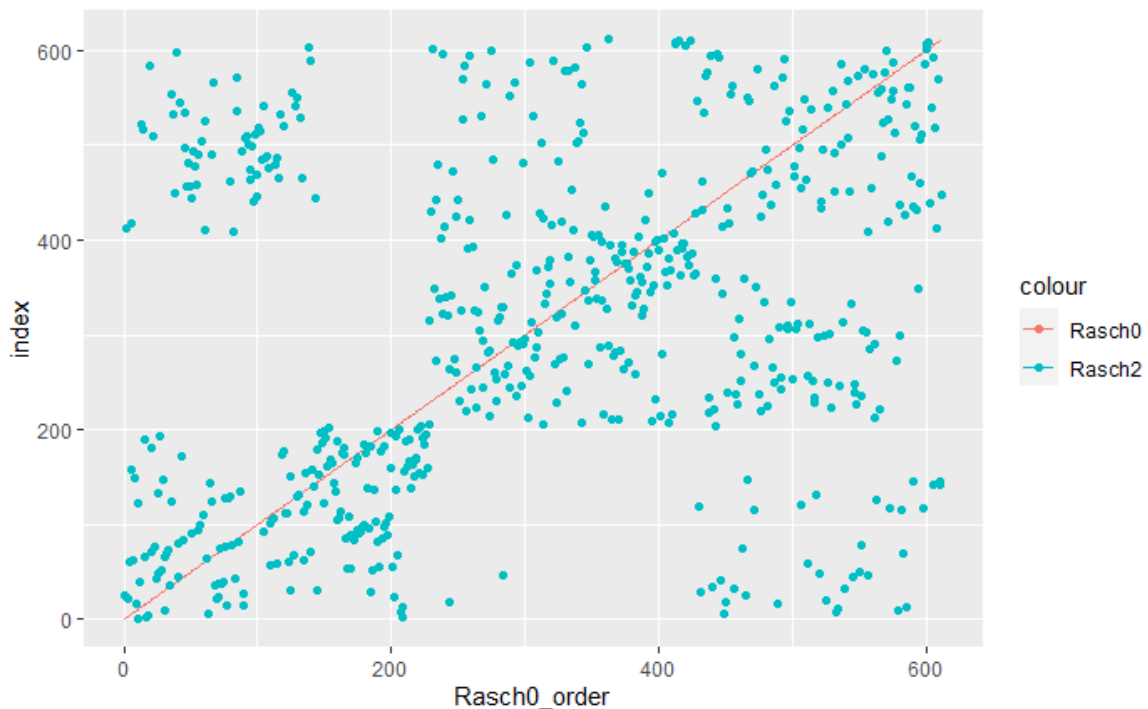


Figure 1. Rank of the model_0 verses rank of the model_2. Data are ordered in the rank of model_0 and the rank of the model_2 were described as blue dots. The blue dots showed four groups.

and lower Deviance than the model from the Finn and Servos's data set. Since we valued AIC more than Deviance for the model selection, we selected the model_2, maximum likelihood method model for the re-selected data set, as our recommended model.

We ranked the schools by the scores we provided from our recommended model and Finn and Servos's model, model_0. For the analysis of the ranks, Intuitive analysis was difficult due to a large number of schools. So, we visualized the data by rearranging them in the order of the rank of model_0 and comparing them with the rank of model_2.

For better visualization, we placed the rank of the model_0 as a red line and the rank of the model_2 as blue dots shown in the Figure 1. Unusually, the result of model_2 seemed to be divided into four groups. For the numerical analysis, the correlation of the rank of model_0 and model_2 was 0.3130.

We tried to replace the missing values. In our re-selected data set, there were 24 missing values. The first model replaced the missing values into the mean value of each questions. So, if the response of q38c is missed for the school "id3572", we replaced its value to the average of total q38c responses. The second model replaced the missing values into mean value of each questions calculated only from the data that has same 'BYURBAN' value, a value representing school urbanicity, with the school that has missing variables. For instance, if the 'BYURBAN' variable for the school "id3572" was 1, we replaced its missing value, the response of q38c, to the average of q38c responses from the data which have BYURBAN variable equals to 1.

Two model's AIC and deviance were almost equal. Their AIC were 4579.35 and 4579.40, and deviance were 3228.91 and 3228.98 Since we only replaced the missing variables and only 24 schools out of 656 schools had missing variables, AIC and deviance should not change much.

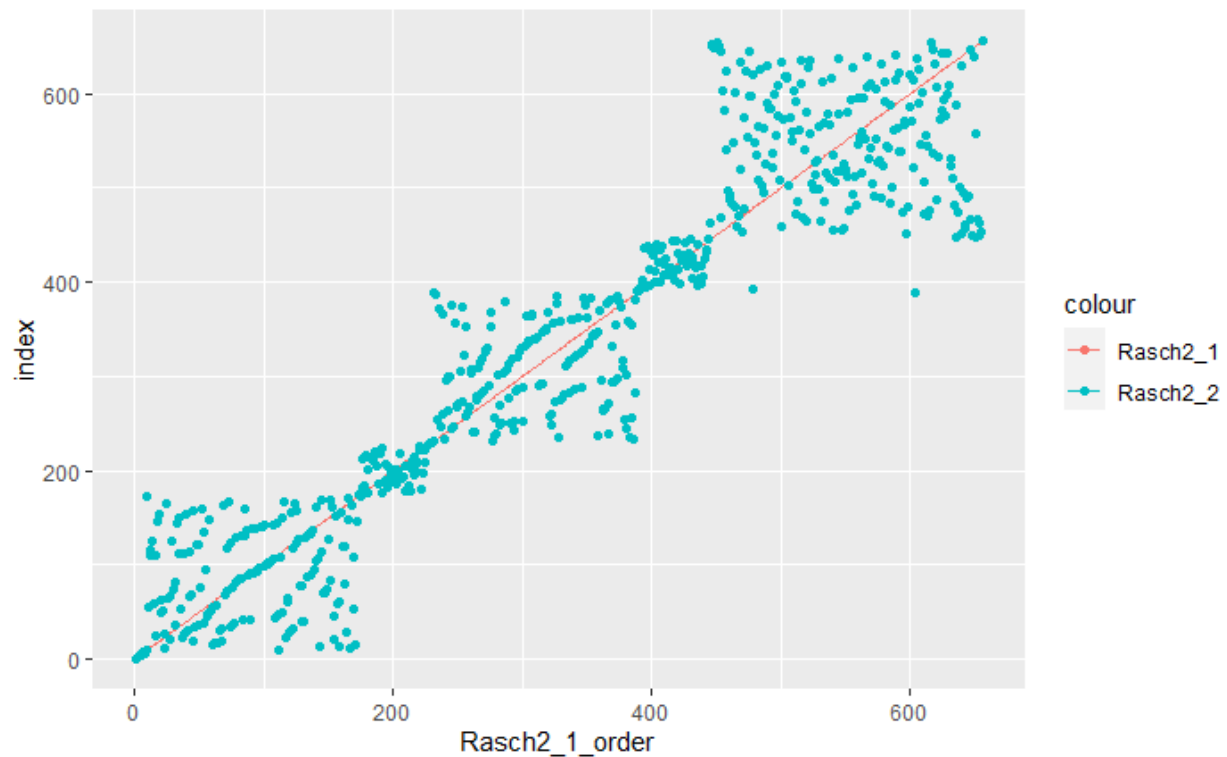


Figure 2. Rank of the Rasch2_1 model verses rank of the Rasch2_2. Data are ordered in the rank of Rasch 2_1 and the rank of the Rasch2_2 were described as blue dots.

Figure 2 shows the relationship between Rasch2_1 model, the model replaced the missing value from all data, and Rasch2_2 model, the model replaced the missing value from the data set that have same BYURBAN value. Two ranks seem highly related in the graph. We can also draw three groups for the blue dots. For numerical analysis, the correlation between the rank of two models were 0.9349.

Discussion

About the sum of indices model, this model is a very simple way to quantify the security measurement, but it had two problems. First, the importance of each question was not taken into account. It is hard to say that two schools with the same score really need to get the same score

because the scores for all questions were weighted as 1. Second, there are too many schools which got the same score which makes this model meaningless. To create a more convincing and effective model, we used the Rasch model by using maximum likelihood and a penalized variant for maximum likelihood mentioned above. Since we used two data sets, we produced four models as a result.

Clearly, we stated that the re-selected data set is better than Finn and Servos's data set because it showed lower AIC and Deviance. The reason for this result is 1) Q38c and Q38d both ask about how much schools consider the material objects owned by students. Because there were only 16 schools that answered the question 'yes, it makes sense to remove Q38c from the data set. 2) Q38f and Q38g both ask about how much schools consider drugs. So, it makes sense to remove Q38f from the data set. 3) The question added in the re-selected data set, Q38j) Enforce a strict dress code, working positively at determining the score.

Among the four models, we selected model_2 because it showed the lowest AIC. Interestingly, the rank of the model_2 showed four groups compare to the model_0, the original model performed by Finn and Servos. Moreover, the model_0 data passed through the centers of the two groups that contained most of the model_2 data. One might say that this is a significant positive relationship, but we should remind that the 5 questions out of 6 or 7 were overlapped for the two data sets. Furthermore, model_0 and model_2 used the same analyzing method, the maximum likelihood method. From these points of view, we might interpret that the correction value of 0.3130, which we obtained, is rather low.

We also replaced the missing values. The two models, separated by reference to the BYURBAN variable or not, were highly related with a correlation of 0.9349. We cannot decide which model is more 'right' from the data because it is just our choice how to rank the schools,

but this high correlation value was predicted that the two models would not be much different because there are only 24 of the 656 schools including missing values in the re-selected data set. We can also distinguish the rank of the second model into three groups, ordered by the rank of the first model.

For the further questions, we can analyze how important each question played in determining the score in each model. This will allow us to quantitatively analyze how the questions we removed and added from Finn and Servos's data played a major role in lowering AIC and lowering Deviance. In advance, we can expect that the added or removed questions had a significant effect, in good ways or bad ways, because we got a correction value that is lower than we expected from the number of overlapping questions.

We might also think about the reason why the rank data from model_2 was divided into four groups. Although the major of the data follows the positive relationship, some data groups showed low ranks in model_0 and high ranks in model_2, and low ranks in model_2, and high ranks in model_0. It would be good to analyze what differences these schools have and whether there are differences in overall other questions.

References

- [1] Finn, Jeremy D. and Servoss, Timothy J. (2014) "Misbehavior, Suspensions, and Security Measures in High School: Racial/Ethnic and Gender Differences," *Journal of Applied Research on Children: Informing Policy for Children at Risk*: Vol. 5 : Iss. 2 , Article 11.