Bayesian Analysis for Correlating Academic Performance with Student Statistics

Harika Malapaka, Emili Moan, and Pragya Vaishanav | Applied Bayesian Analysis | North Carolina State University

Motivation

Improving the teaching and learning effectiveness of a course will have a significant impact on student success and retention by virtue of the large number of students affected.

Introduction

This study is purposed to analyze the correlations between student performance or success in a particular course and the demographic, behavioral, personal and academic background of the student. The aim is to use Bayesian techniques to come up with a model which allows to predict the grade of a student; henceforth, guiding a student to choose the courses accordingly or an institution to select the students for a particular curriculum.

Data Description

- Released by LMS (Learning Management System) in 2016
- Categorical data for 460 students: these type of data can be divided into groups i.e. for our case race, sex, educational level etc. The ordering of variables remains undefined for categorical data.

Covariates

- Raised hand
- Visited resources
- Viewed announcements
- Discussion groups
- Gender (male = 0, female = 1)
- Semester (first = 0, second = 1)
- Parent responsible (father = 0, mom = 1)
- Parent answering survey (yes = 0, no = 1)
- Parent school satisfaction (yes = 0, no = 1)
- Student absence days (below 7 = 0, above 7 = 1)
- Nationality (reference = Kuwait)
- Place of birth [excluded]
- Educational stages (reference = lower)
- Grade levels [excluded]
- Section ID (reference = A)
- Topic (reference = IT)

Model

Our response variable is $Y_i = low, med, high.$ We assume an underlying latent variable: $Y_i = \begin{cases} low\ if\ \ Y_i^* \le \theta_1 \\ med\ if\ \theta_1 < \ Y_i^* \ \le \theta_2 \\ high\ if\ \theta_2 < \ Y_i^* \end{cases}.$

We will assume $Y_i^* = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon_i$ where $\epsilon_i \sim F(\epsilon_i)$. Then $\Pr(Y_i = j) = \Pr(\theta_{j-1} < Y_i^* \le \theta_j) = F(\theta_j - \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p) - F(\theta_{j-1} - \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p)$.

Model 1

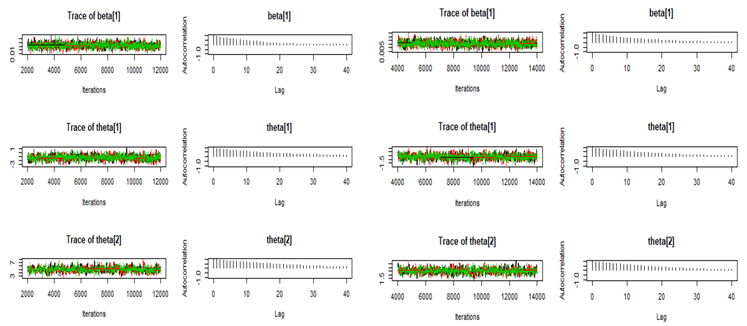
- Logit Model
- $F(z) = \frac{e^z}{1+e^z}$

Model 2

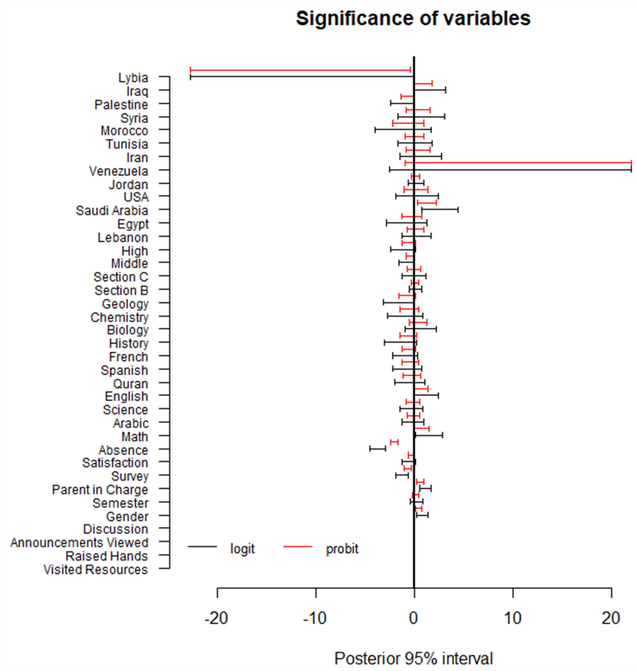
- Probit Model
- $F(z)$ ~ Normal CDF

Convergence

	Min ESS	Mean ESS	Max ESS	Min GS	Mean GS	Max GS
Model 1	544	5076	13785	1	1.001	1.007
Model 2	654.6	5100.6	14546.6	1	1.001	1.003



Significance of covariates



Results

Model 1

Variable	Mean	2.5%	97.5%
Visited Resources	0.03	0.02	0.04
Raised Hands	0.03	0.02	0.04
Gender	0.85	0.29	1.41
Parent	1.16	0.55	1.78
Survey	-1.18	-1.83	-0.54
Absence	-3.64	-4.45	-2.90
Math	1.48	0.10	2.87
Middle Level	-0.77	-1.54	-0.02
Saudi Arabia	2.55	0.76	4.43
Libya	-8.38	-22.71	-0.05
Theta1	-1.19	-2.38	0.03
Theta2	4.95	3.62	6.39

Model 2

Variable	Mean	2.5%	97.5%
Visited Resources	0.02	0.01	0.02
Raised Hands	0.01	0.01	0.02
Gender	0.45	0.14	0.75
Parent	0.62	0.29	0.95
Survey	-0.62	-0.96	-0.27
Absence	-1.97	-2.37	-1.59
Math	0.79	0.07	1.52
English	0.70	0.03	1.39
Middle Level	-0.42	-0.83	-0.01
Saudi Arabia	1.32	0.40	2.26
Iraq	0.91	0.06	1.80
Libya	-8.21	-22.69	-0.34
Theta1	-0.64	-1.27	-0.03
Theta2	2.56	1.87	3.26

Five-Fold Cross Validation

Model	Misclassification Rate
1	0.3055
2	0.3285

General Conclusion

- Female students make higher grades than male students
- Higher absences lead to lower grades
- Math students make higher grades than IT students

Future Work

- Hierarchical models
- Predictions

References

- Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. International Journal of Database Theory and Application, 9(8), 119-136.
- Amrieh, E. A., Hamtini, T., & Aljarah, I. (2015, November). Preprocessing and analyzing educational data set using X-API for improving student's performance. In Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference on (pp. 1-5). IEEE.
- Wingfeet. “Ordinal Data.” *R-Bloggers*, R-Bloggers, 25 Mar. 2013, www.r-bloggers.com/ordinal-data/.