

Section 6. Sociology

<https://doi.org/10.29013/YSJ-22-5.6-38-42>

Qianran Bi,
Grier School, PA, USA

PREDICTIVE MODEL FOR STUDENTS' MATH PERFORMANCE

Abstract

Aim. This study aimed to build a predictive model for the final grade of math students.

Method. A public data was used in this study. All the records were randomly assigned into 2 groups: training sample (50%) and testing sample (50%). A linear regression model was built to predict the final grade of math students using the training sample and then was applied in the testing sample for performance assessment.

Results. The random sample size is 325 in the testing sample and 324 in training sample, a total of 649 records. The average score was 11.91 for the full sample, 11.76 in the training sample, and 12.05 in the testing sample. According to the linear regression, father education, study time, previous failures, school sup, paid tuition, higher, internet, daytime alcohol assumption can predict the final grade among math students. Multiple R-squared was 0.3154 and the adjusted R-squared was 0.258. The average mean squared error for the linear model in the testing sample was 8.5. The correlation between the predicted and the observed was 0.494. The min-max accuracy was 0.8355. Mean absolute percentage deviation is 0.99.

Conclusions. In this study, we identified important of predictors for the final grade of math students, for example father's education, daytime alcohol assumption.

Keywords: Math performance, model, predictors, linear regression.

1. Introduction

The ability to predict one's academic performance is a great asset for both the students and the institution administrators. For the students, they can adjust workload, career direction, etc. If they are aware of their capability. For the administrators and instructors, early warnings would facilitate intervention thus enabling a more successful academic environment. In addition, institutional resources can be utilized in an optimal way thus gaining operation efficiency [1].

This study aimed to build a predictive model for the days for the final grade among math students.

2. Data and Methods:

Data

This dataset is public available for research.

All the records were randomly assigned into 2 groups: training sample (50%) and testing sample (50%).

R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression.

The definition of R-squared is fairly straight-forward; it is the percentage of the response variable variation that is explained by a linear model. Or:

$$R\text{-squared} = \text{Explained variation} / \text{Total variation}$$

R-squared is always between 0 and 100%: 0% indicates that the model explains none of the variability of the response data around its mean. 100% indicates that the model explains all the variability of the response data around its mean.

Mean squared errors (MSE) were calculated and compared between both models. Min-Max Accuracy is defined as $\text{mean}(\min(\text{actual}, \text{predicted}) / \max(\text{actual}, \text{predicted}))$. The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of prediction accuracy of a forecasting method in statistics, for

example in trend estimation, also used as a loss function for regression problems in machine learning.

3 Results:

The random sample size is 325 in the testing sample and 324 in training sample, a total of 649 records. The average score was 11.91 for the full sample, 11.76 in the training sample, and 12.05 in the testing sample.

Table 1.

	min	Q1	Median	Mean	Q3	Max
Full sample	0	10	12	11.91	14	19
Training sample	0	10	11	11.76	14	18
Testing sample	0	10	12	12.05	14	19

final scores of students at the end of a math programs

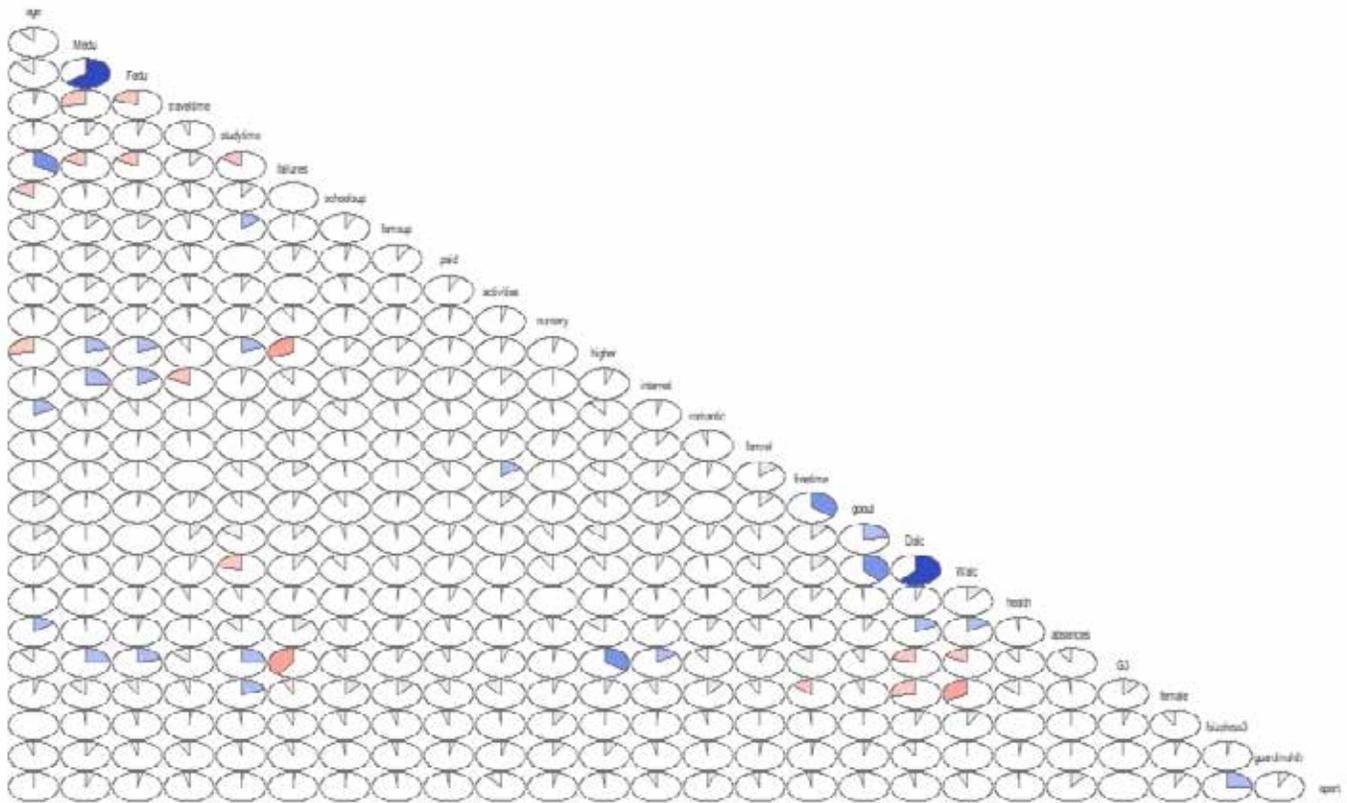


Figure 1. Pearson correlation coefficient across all the variables

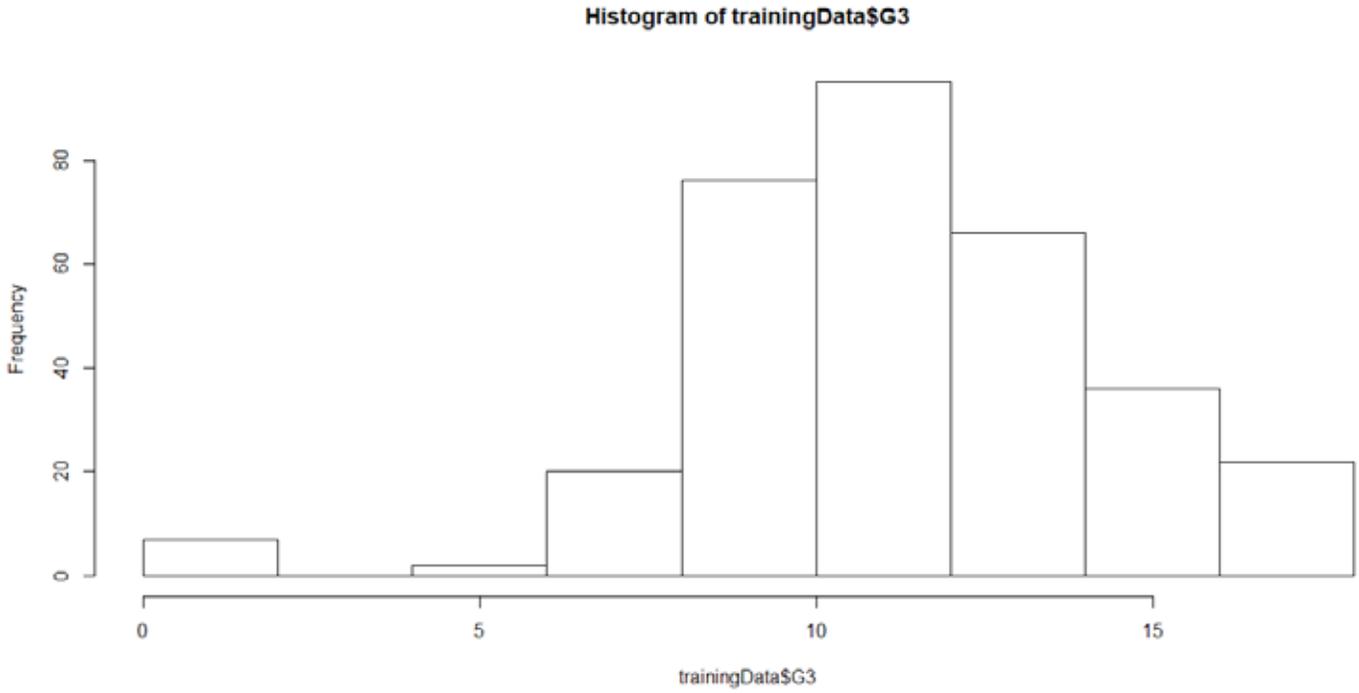


Figure 2. Distribution of Final Grade among Math Students in Training Sample

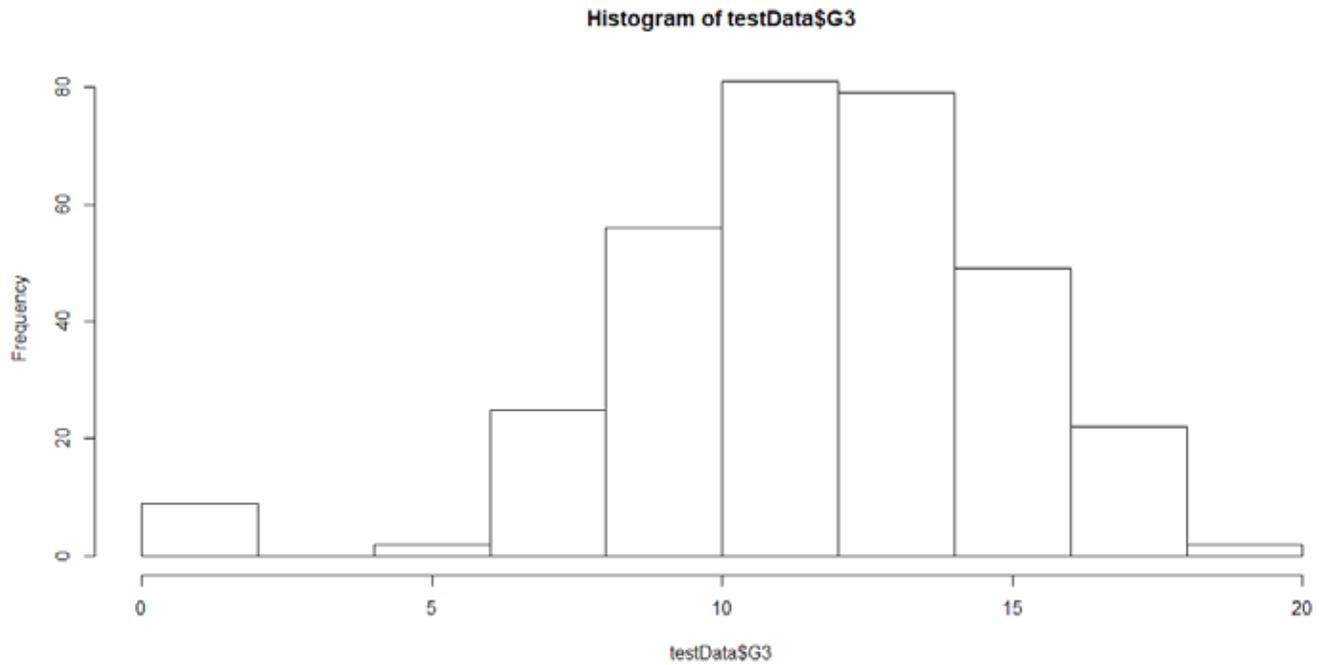


Figure 3. Distribution of Final Grade among Math Students in Test Sample

According to the linear regression, father education, study time, previous failures, school sup, paid

tuition, higher, internet, daytime alcohol assumption can predict the final grade among math students.

Table 2. – Linear Regression to Predict Final Grade among Math Students

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	10.97	2.65	4.15	0.00	***
age	0.01	0.14	0.08	0.93	
Medu	0.02	0.20	0.09	0.93	
Fedu	0.40	0.19	2.09	0.04	*
traveltime	0.14	0.22	0.63	0.53	
studytime	0.45	0.19	2.36	0.02	*
failures	-1.19	0.28	-4.28	0.00	***
schoolsup	-1.09	0.53	-2.07	0.04	*
famsup	-0.38	0.32	-1.19	0.23	
paid	-1.49	0.59	-2.51	0.01	*
activities	0.19	0.32	0.59	0.56	
nursery	-0.22	0.40	-0.55	0.58	
higher	1.59	0.55	2.89	0.00	**
internet	0.94	0.37	2.52	0.01	*
romantic	-0.41	0.32	-1.29	0.20	
famrel	0.01	0.17	0.08	0.93	
freetime	-0.23	0.16	-1.43	0.15	
goout	-0.11	0.15	-0.70	0.48	
Dalc	-0.59	0.22	-2.71	0.01	**
Walc	-0.02	0.17	-0.09	0.93	
health	-0.20	0.11	-1.78	0.08	
absences	-0.01	0.04	-0.14	0.89	
female	0.19	0.36	0.53	0.60	
fsizeless3	0.38	0.36	1.05	0.30	
guardmohth	-0.06	0.35	-0.18	0.86	
apart	-0.49	0.48	-1.02	0.31	

Multiple R-squared was 0.3154 and the adjusted R-squared was 0.258. The average mean squared error for the linear model in the testing sample was 8.5. The correlation between the predicted and the observed was 0.494. The min-max accuracy was 0.8355. Mean absolute percentage deviation is 0.99.

4. Discussion

In another study, in the multiple regression model 5 factors explained 24.6% of the variance: socio economic status (9.3% of the variance), parental involvement with the school (6.4%), academic goals (4.9%),

student–teacher interaction (leadership, rigor and orientation towards mastery in triennial courses) (2.6%), and academic related skills (1.1%) [2].

The random sample size is 325 in the testing sample and 324 in training sample, a total of 649 records. The average score was 11.91 for the full sample, 11.76 in the training sample, and 12.05 in the testing sample.

According to the linear regression, father education, study time, previous failures, school sup, paid tuition, higher, internet, daytime alcohol assumption can predict the final grade among math students.

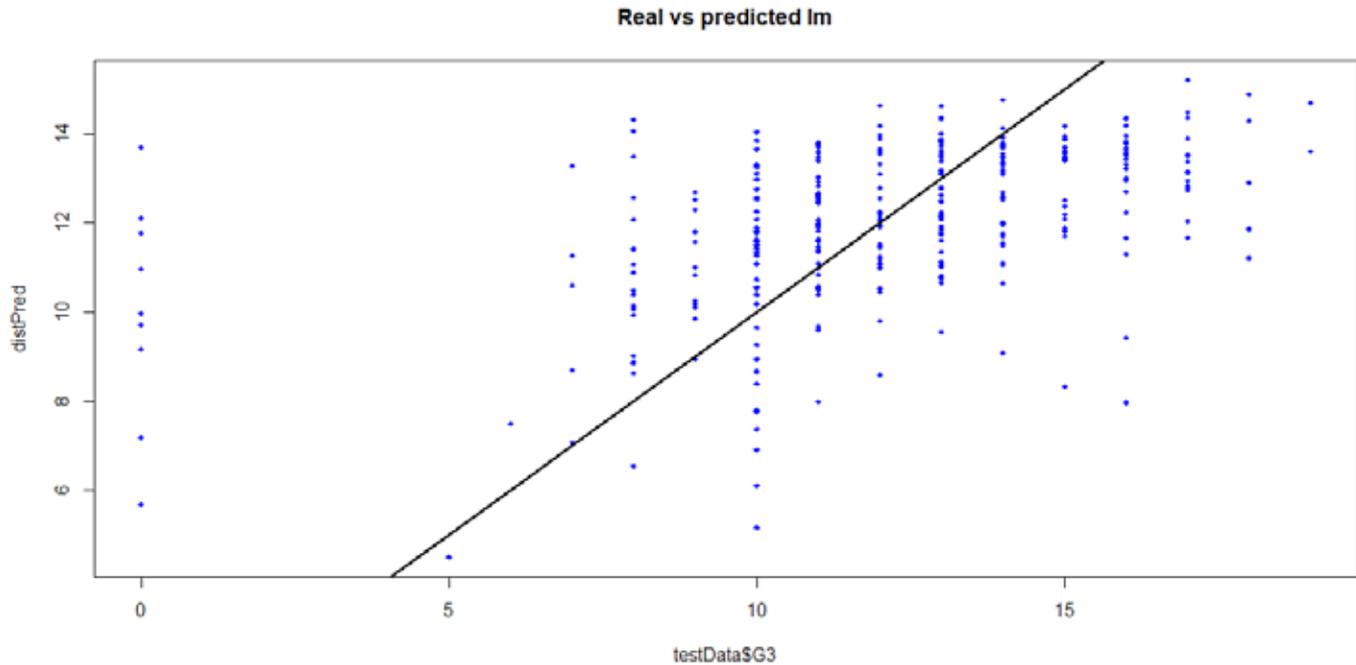


Figure 4: Predicted versus Observed Final Grade

Multiple R-squared was 0.3154 and the adjusted R-squared was 0.258. The average mean squared error for the linear model in the testing sample was 8.5. The correlation between the predicted and the observed was 0.494. The min-max

accuracy was 0.8355. Mean absolute percentage deviation is 0.99.

In this study, we identified important of predictors for the final grade of math students, for example father's education, daytime alcohol assumption.

References:

1. Academic Performance Predictors <https://ieeexplore.ieee.org/document/7096239>
2. Predictors of academic performance and school engagement — Integrating persistence, motivation and study skills perspectives using person-centered and variable-centered approaches. <https://www.sciencedirect.com/science/article/pii/S1041608012001677>
3. Peng, C. J., Lee, K. L., Ingersoll, G. M. An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, 96(1), 3–14.
4. Tabachnick, B., and Fidell, L. *Using Multivariate Statistics* (4th Ed.). Needham Heights, MA: Allyn & Bacon, 2001.
5. StatSoft, *Electronic Statistics Textbook*, <http://www.statsoft.com/textbook/stathome.html>.
6. Stokes, M., Davis, C. S. *Categorical Data Analysis Using the SAS System*, SAS Institute Inc., 1995.