Christopher Botts
DSC680
White Paper 2

Predicting Student Test Scores

Business Problem

Test scores are important for students as well as the school they attend. For students, these scores are a means for college admission. For administrators, test scores are used to evaluate the quality of educational instruction throughout k-12 education. The results of subject tests can indicate whether curriculum is effective or if students may need additional or alternative education.

Background/History

As of 2019, schools spent an average of $13,187 per student per year of k-12th grade education nationwide; states such as New York, spent upwards of $25,000 per student (Chen 2021). To maximize this investment, education expenditures should create a learning environment that is best suited for students to absorb and retain information. Test scores provide a way to evaluate the return on investment of public money allocated to the education system.

Data Explanation

The dataset analyzed in this project contains the profile of 1000 unique students. Features include demographic information such as gender; parental education status; potential score boosters such as participation in test preparation courses; and test scores across three subjects: math, writing and reading.

Data Dictionary	
gender	binary values of male and female
race_ethnicity	contains 5 values of anonymized classifications
parental_level_of_education	contains 6 values: some high school, high school, some college, associate degree, bachelor’s degree, master’s degree
lunch	binary values of standard or free/reduced
test_preparation_course	binary values of completed or none
math_score	math test scores ranging from 0-100
writing_score	writing test scores ranging from 0-100
reading_score	reading test scores ranging from 0-100

Table 1. Data dictionary for student dataset

The target feature for this analysis was ‘math_score’. The three subject test scores, ‘math_score’, ‘reading_score’, and ‘writing_score’ were not entirely redundant, but were similar in distribution and value, so math was chosen to represent student scores as an indicator of educational success.

‘math_score’ had a normal distribution, no duplicates, or outliers. It was ready for use in predictive modeling.

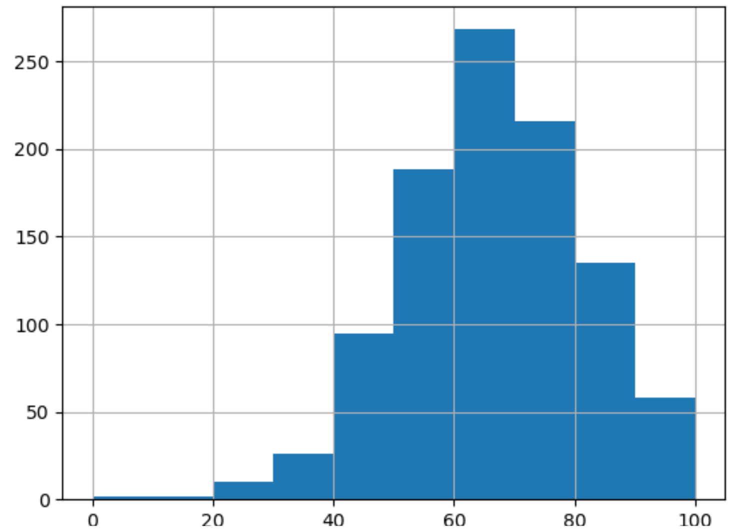


Chart 1. Distribution of math_score

The dataset was composed mostly of categorical variables, so the dataframe was one hot encoded to perform correlation analysis. A correlation matrix is presented below.

	math_score
math_score	1.000000
reading_score	0.817580
writing_score	0.802642
gender_female	-0.167982
gender_male	0.167982
race_ethnicity_group A	-0.091977
race_ethnicity_group B	-0.084250
race_ethnicity_group C	-0.073387
race_ethnicity_group D	0.050071
race_ethnicity_group E	0.205855
parental_level_of_education_associate's degree	0.063228
parental_level_of_education_bachelor's degree	0.079664
parental_level_of_education_high school	-0.128725
parental_level_of_education_master's degree	0.060417
parental_level_of_education_some college	0.037056
parental_level_of_education_some high school	-0.079852
lunch_free/reduced	-0.350877
lunch_standard	0.350877
test_preparation_course_completed	0.177702
test_preparation_course_none	-0.177702

Table 2. math_score correlation matrix

All features showed no or weak correlation with ‘math_score’ except for test scores of other subjects both of which showed high correlation. Because of the weak correlation across all variables, it was determined to build a multiple regression model.

Recursive feature elimination was used to narrow down the features for model building based on cross validation using mean squared error. Recursive elimination suggested that the most effective model includes all 19 features. The graph below shows the results. Including more than 9 features significantly improves the model. Improvement slows after 13 features but continues until all features are included.

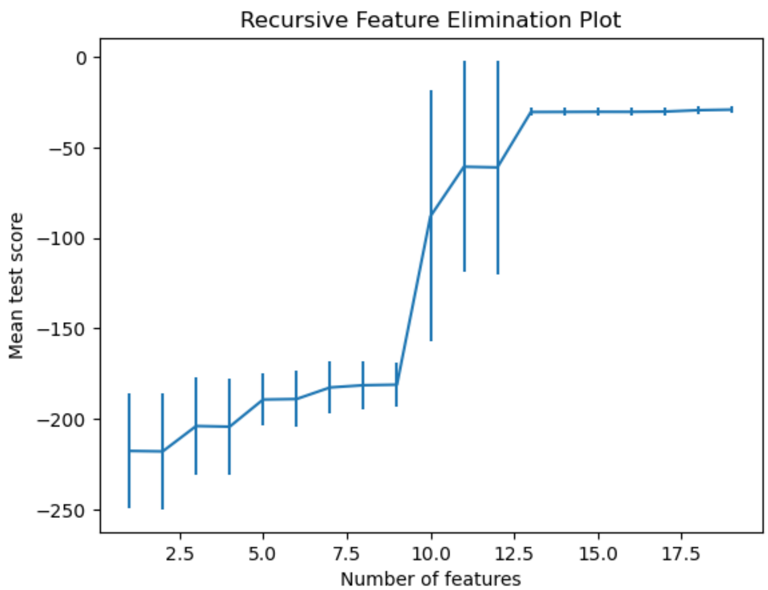


Chart 2. Model accuracy based on the number of features included

<u>Methods</u>

A multiple regression analysis was performed to predict test scores. Three models were built to test linear trend and variability. These models were evaluated using the r squared value and root mean square error.

Model 1.

The first model was a multiple linear regression using all 19 features from the dataset. The features were unmodified.

Model 2.

The second model was built to fit a nonlinear model. It evaluated both a quadratic and cubic transformation.

Model 3.

The third model was a ridge regression selected to reduce variance. All 19 features were standardized and used in the creation of this model.

Analysis

The ridge regression model had the highest r squared and root mean square error and the multiple linear regression had similar results. Nonlinear regression showed the worst fit.

	R2	RMSE
Linear Regression	0.866	5.494
Polynomial Regression	0.811	6.569
Ridge Regression	0.879	5.288

Table 3. Model Results

The lower score of the polynomial regression suggests that there is a linear relationship between the features and target variable. The positive impact of the ridge regression compared to model 1 indicates that the ordinary least squares linear regression may have been slightly overfit.

The high R squared value shows that the model has a good fit. Features inputted into the ridge regression model can explain 88% of variation in math scores. The low root mean square error value tells us that the model is also accurate.

The coefficients of the model describe the impact each feature has on the math score. The most impactful two features are reading and writing score. The next two largest coeffecients are demographic features; the third highest coeffecient is gender and the fourth is one of five race_ethnicity groups. All four of these features show statistically significant f-scores.

```
writing_score 10.5
reading_score 3.96
gender_male 3.29
race_ethnicity_group E 1.39
```

Conclusion

The results of this model indicate that education outside of math lessons are positively associated with math scores. Although there is not a determined cause effect relationship, the strong correlation and model results support a positive linear relationship between math scores and the scores in other subjects. Investing in other courses, or the educational experience as a whole, may consequently improve the student’s performance in math.

Although reading and writing scores contributed the most to predicting a student's math score, all features improved the model. Furthermore, all features showed a statistically significant difference from a model without these features with a p value $< .001$. This suggests that a person's understanding of math is highly multifaceted. A student's home life, culture, background, economic status, and any extra resources provided to a student all play a role in their educational success as measured by standardized exams.

Assumptions

One assumption with linear models is that the relationship between the features and target is linear. To test this assumption, a quadratic and cubic model were also evaluated. The result of this test was that the features have a linear relationship with the target variables.

Another assumption in the business question is that test scores are useful indicators of the effectiveness of teaching methods. However, test score may not be an accurate metric for measuring successful teaching. This model predicts test scores and extrapolates conclusions from the predicted test score results.

Limitations

One limitation of this model is that it predicts best with all features, but some demographic information may not be available for future instances. Furthermore, there is little detail about teaching methods, styles, study time. The exclusion of these details limits the explanatory power of the model. The dataset also does not include datetime information. This enables us to build a model to predict a student's math score given their demographic information but prevents us from predicting how math scores for the school will change over time.

Challenges

This is a clean and easy to work with dataset. The features are self-explanatory and new instances can be easily classified into the existing feature categories for use in the model. The main challenge with this project is that the dataset is limited in descriptive features.

Future Uses and Additional Applications

It is interesting to note that the top two highest predicting factors for math_score after test scores in other subjects was gender and race/ethnicity. In this model, each additional male identifying student was associated with a 3.29 point increase in math score. This suggests that there may be a gender inequality in education. This may be unsurprising; mathematical careers historically and presently are male dominated. This may lead to bias in education that invests more time and energy into teaching males math than other genders. Furthermore, women often face discrimination in math courses (Steele et al 2002). It is likely that the students from this dataset faced a similar experience.

There is room to explore these results further. Provided more demographic information, especially information pertaining to gender and race/ethnicity could enable a deeper dive into the equality of education for minorities and other historically prejudiced persons in our education system.

Recommendations

There is room for improvement in the model. Future models should try to capture additional information to predict how feature outside of school affect test scores. It would also be valuable to include information from the schools including the amount of funding for the math department, extracurriculars like math club, field trips, the time students begin and end the school day.

Implementation Plan

This model should be used for students in the elementary school systems on a quarterly basis. It can predict their scores for the end of the year test and provide feedback for individual students on their understanding of course material. If their score predictions improve, then it shows that the teaching methods are effective. If the students' score declines, then the teaching style is not working for that student. If this is seen as a broad trend in all students, then the curriculum should be reevaluated. However, if a score decline is observed in a small subset of students, the teacher should use an opportunity to provide more personalized training with those students.

Ethical Assessment

This model shows less accurate results lower without sensitive demographic information such as race / ethnicity and gender. By necessitating these features, this model risks perpetuating discrimination against student based on the ethnicity of gender identification. It also limits the classification of several features, such as gender which offers only male and female values. This may not be representative of student identification.

References

Chen, Erika (2021, May). K-12 School Spending UP 4.7% in 2019 From Previous Year. *United States Census Bureau.* Retrieved from: https://www.census.gov/library/stories/2021/05/united-statesspending-on-public schools-in-2019-highest-since-2008.html

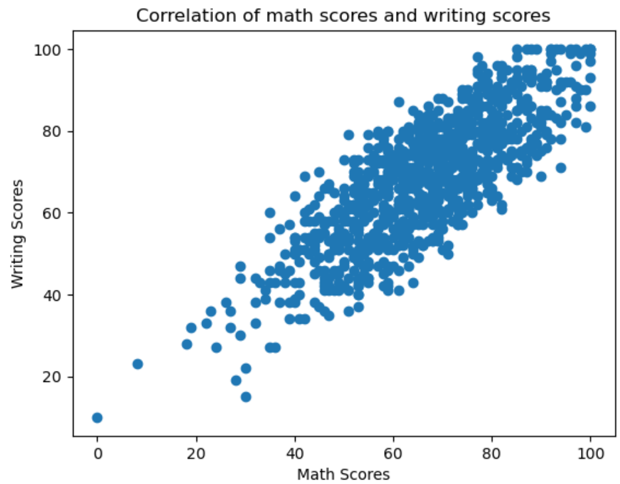
Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological science in the public interest*, *9*(3), 105-119.

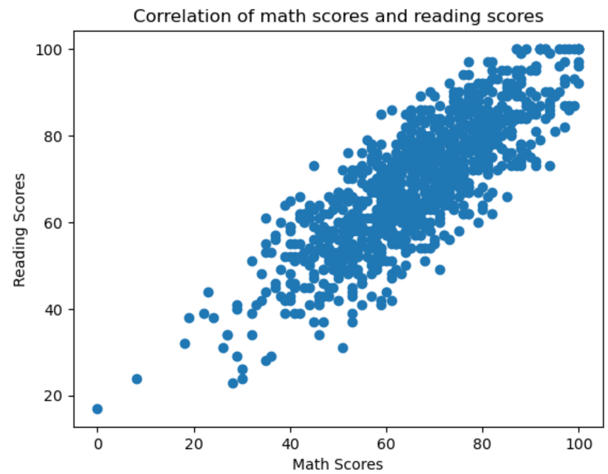
Steele, J., James, J. B., & Barnett, R. C. (2002). Learning in a Man's World: Examining the Perceptions of Undergraduate Women in Male-Dominated Academic Areas. *Psychology of Women Quarterly*, *26*(1), 46-50. https://doi.org/10.1111/1471 6402.00042

USC Rossier Online Teaching Degree. (n.d.) *U.S. Education Spending and Performance vs. The World [INFORGRAPHIC].* Retrieved from: https://rossieronline.usc.edu/blog/us-education-versus-the-world-infographic/

Appendix 1.

Correlation of math_score and test scores in other subjects





Results of Ridge Regression

Displaying feature name, coefficient, f-statistic, p-value

```
reading_score 3.9663358729263836 2011.981555312104 1.7877531098927278e-241
writing_score 10.509068766967616 1807.2166657826942 3.3760270425458192e-226
gender_female -3.298572849032812 28.97933609503111 9.12018554933112e-08
gender_male 3.298572849032812 28.97933609503111 9.12018554933112e-08
race_ethnicity_group A -0.29809267145667917 8.514901585538928 0.0036015171249518794
race_ethnicity_group B -0.08355333225486561 7.134517789068466 0.007684428392683 14
race_ethnicity_group C -0.40373155710704367 5.4039656242276095 0.020290932900978286
race_ethnicity_group D -0.407369410449121 2.508350825622873 0.11356119810525196
race_ethnicity_group E 1.3976386340622327 44.16280405596806 4.962294009781 4446e-11
parental_level_of_education_associate's degree -0.013411508857672061 4.005820152003389 0.04561378068806519
parental_level_of_education_bachelor's degree -0.34282404234632224 6.3740432446353505 0.011734201427985803
parental_level_of_education_high school 0.20630306813575672 16.815692092046685 4.4543256457807136e-05
parental_level_of_education_master's degree -0.4411124541489784 3.6562366770091312 0.056145748479768276
parental_level_of_education_some college 0.15441096532991594 1.372308741117535 0.24169520551730445
parental_level_of_education_some high school 0.19207391211897118 6.404406079151087 0.011536477504795651
lunch_free/reduced -0.772821806534914 140.1188415483533 2.4131955993124396e-30
lunch_standard 0.772821806534914 140.1188415483533 2.4131955993124396e-30
test_preparation_course_completed -0.8302950126547017 32.54264846908929 1.535913460715173e-08
test_preparation_course_none 0.8302950126547017 32.54264846908929 1.535913460715173e-08
```

U.S. Education Spending and Test Results Compared to other Nations

USC Rossier's Teaching Degree