An Analysis of the Department of Education Quality Survey and Its Efficacy

Andrew Bowen[1], Glen Dale Davis[1], Josh Forster[1], Shoshana Farber[1], & Charles Ugiagbe[1]

[1] City University of New York

Abstract

Abstract coming soon!

Keywords: Educational Outcomes, School Quality, Education

An Analysis of the Department of Education Quality Survey and Its Efficacy

Introduction

The NYC School Survey seeks to collect data to provide an overview of New York City Schools. Beginning in 2005, the survey looks to collect demographic and achievement data for New York City Public Schools, and provide a standardized rating of various elements of school quality.

The survey has changed over the years. This change has come from recommendations of public policy analysts in order to more accurately define the quality of schools *New York City Schools (2018)*. The 2020-21 academic year report provides a robust dataset defined at the school level with academic and socioeconomic data provided.

Research Question: This study aims to determine whether the school ratings within the NYC School Quality Survey accurately reflect educational outcomes, or if other variables related to certain schools can be used as a better proxy.

Literature Review

One of the main predictors of academic performance is the socioeconomic background of a student. Students from low-income families are nearly four times more likely to drop out of high school than students from wealthy families *Education Statistics (2008)*.

Attempts to use more sophisticated modeling techniques and different sources datasets come from several prior studies. *Bernacki, Chavez, and Uesbeck (2020)* based their modeling off trying to predict based on student digital behavior, rather than social factors. The model in this study reached an accuracy of 75%, and was able to flag early interventions. While this modeling technique attempts to predict the same variable (educational achievement, albeit a different metric where we are predicting college attainment), the base dataset used to train the model and input variables are different.

Similarly, *Musso, Cascallar, Bostani, and Crawford (2020)* attempted to train an artificial neural network (ANN) to identify variable relationships to educational performance data. They modeled educational performance of Vietnamese students in grade 5. They included individual characteristics as well as information related to daily routines in their training data. This method uses a more sophisticated model, and resulted in accuracy in prediction of $95 - 100$.

Yağcı (2022) predicted final grade exams for Turkish students as well via machine learning models. Their input variables were prior exam grades. These can be a good "vacuum" comparison to compare one set of academci performance to another. However, there is a concern that good exam grades (even in one subject) do not correspond to a higher rate of career success later in life *Afarian and Kleiner (2003)*. Additionally, a parent study also found a correlation of up to 0.3 between academic grades and later job performance *Roth, BeVier, Switzer III, and Schippmann (1996)*.

Measuring the input variables that impact educational outcomes is a difficult task. With so many confounding variables, it can be difficult to determine direct causal relationships that have an outsized impact

Data Sourcing

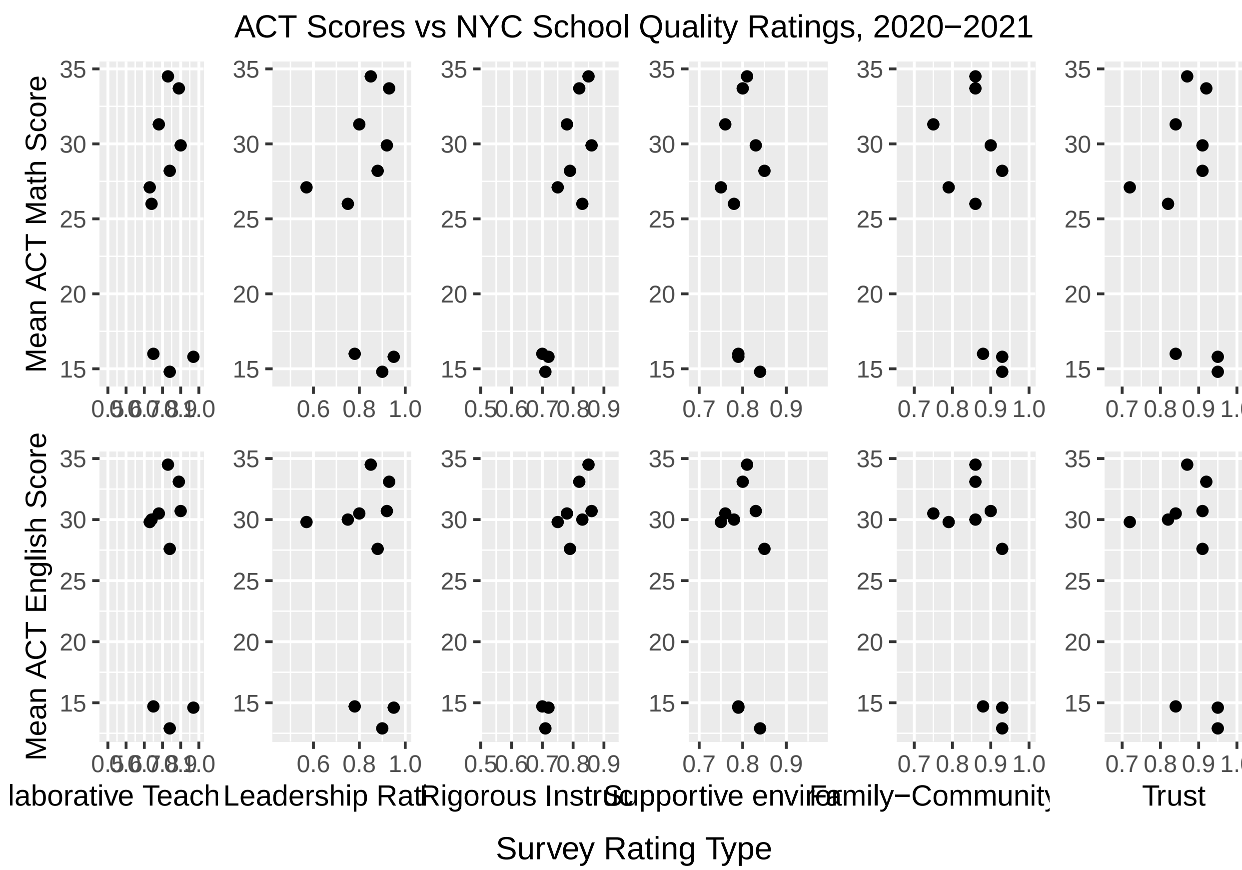
The dataset used in this study is published from the NYC School Quality Report for the Academic Year 2020 - 2021. It consists of data from 487 New York City public schools, and 391 variables (in the form of columns). This dataset is defined at the school level, indexed by a school's *district borough number* (DBN).

In addition to the school quality ratings provided from survey responses in the data, there is average and raw academic performance data included. In addition to thesea academic indicators, there are socioeconomic variables included as well, such as the percentage of students at a given school in temporary housing services.

Methodology

We create a 20% holdout set of data to be used later on in order to evaluate the efficacy of our model's predictive capability. The remaining 80% of the data is to be used for model training and exploratory data analysis (EDA).

The below plot shows the raw relationship between each survey rating (*Collaborative Teaching*, *Trust*, etc) and the response variables of interest: *Average English/Math SAT scores* per school.



Experimentation and Results

First, we construct a basic linear model to predict both English and Math ACT average scores for a given school.

As we see from summary stats below $Rating \rightarrow English/Math$ models perform decently well at predicting ACT English and Math scores, respectively. We see adjusted R^2 values for each academic subject below:

- *English:* 0.76
- *Math:* 0.493

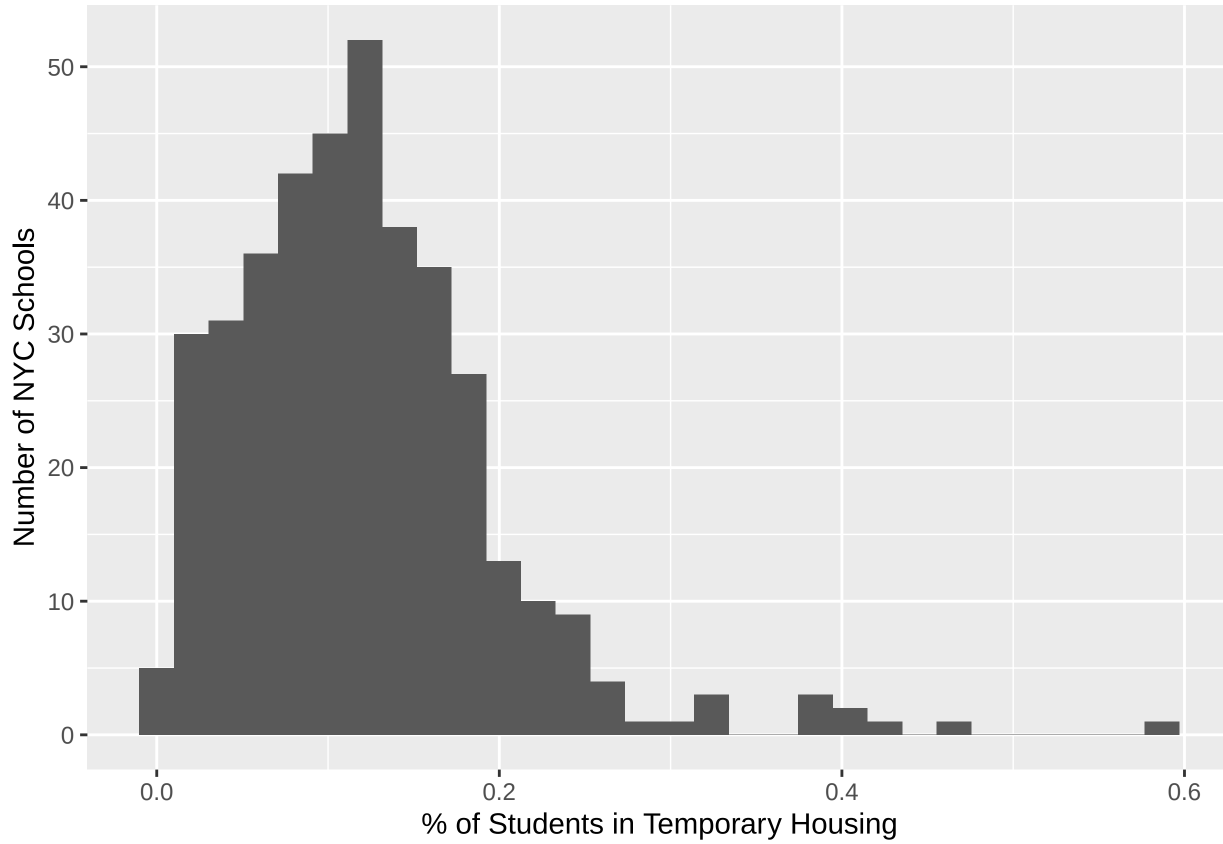
```
##
## Call:
## lm(formula = english_formula, data = train)
##
## Residuals:
##      39      90     128     132     147     193     257     259
##  2.1072  0.1175 -0.1037 -0.7313 -0.6176  0.1158  0.7941 -1.6820
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -154.76      88.19  -1.755    0.330
## survey_pp_RI   118.97      32.77   3.630    0.171
## survey_pp_CT    43.29      40.36   1.073    0.478
## survey_pp_ES  -223.10     144.88  -1.540    0.367
## survey_pp_SE   -23.08     130.37  -0.177    0.888
## survey_pp_SF   -73.85      49.68  -1.486    0.377
## survey_pp_TR   370.05     271.36   1.364    0.403
##
## Residual standard error: 2.976 on 1 degrees of freedom
##   (382 observations deleted due to missingness)
## Multiple R-squared:  0.966,  Adjusted R-squared:  0.7618
## F-statistic:  4.73 on 6 and 1 DF,  p-value: 0.3381
```

```
##
## Call:
## lm(formula = math_formula, data = train)
##
## Residuals:
##      39      90     128     132     147     193     257     259
##  2.9350  0.1636 -0.1444 -1.0186 -0.8602  0.1613  1.1060 -2.3428
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -134.18     122.84  -1.092    0.472
## survey_pp_RI    81.91      45.65   1.794    0.324
## survey_pp_CT    46.38      56.22   0.825    0.561
## survey_pp_ES  -171.25     201.80  -0.849    0.552
## survey_pp_SE    40.36     181.58   0.222    0.861
## survey_pp_SF  -101.32      69.20  -1.464    0.381
## survey_pp_TR   296.15     377.97   0.784    0.577
##
## Residual standard error: 4.145 on 1 degrees of freedom
##   (382 observations deleted due to missingness)
## Multiple R-squared:  0.9276, Adjusted R-squared:  0.493
## F-statistic: 2.135 on 6 and 1 DF,  p-value: 0.4808
```

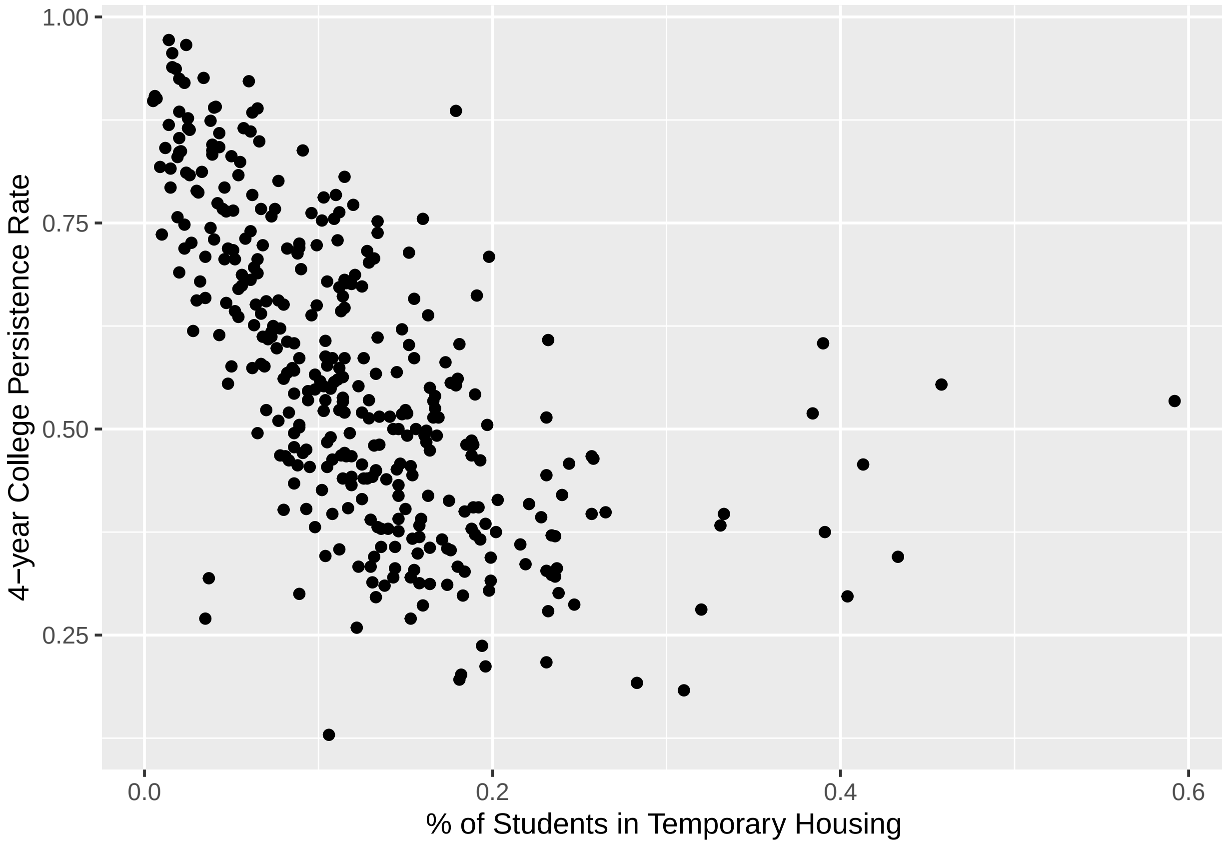
We can use two variables as a proxy for the school's survey rating in predicting college persistence:

- Percent in Temp Housing (`temp_housing_pct`) - percentage of students at a given school living in NYC temporary housing

- Economic Need Index (`eni_hs_pct_912`) - this is a measure of the percent of students facing economic hardship at a school

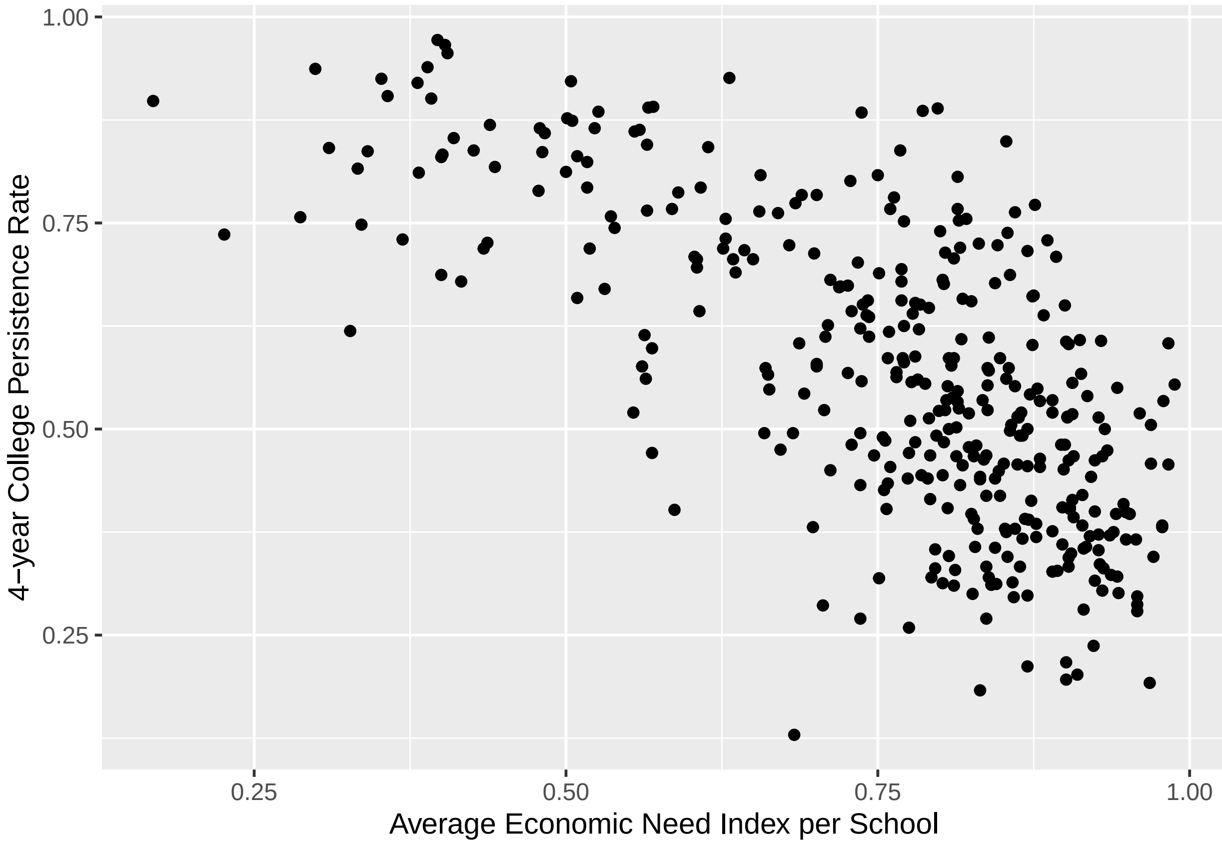


First, we should check an assumption of linearity between our predictor and response variables.

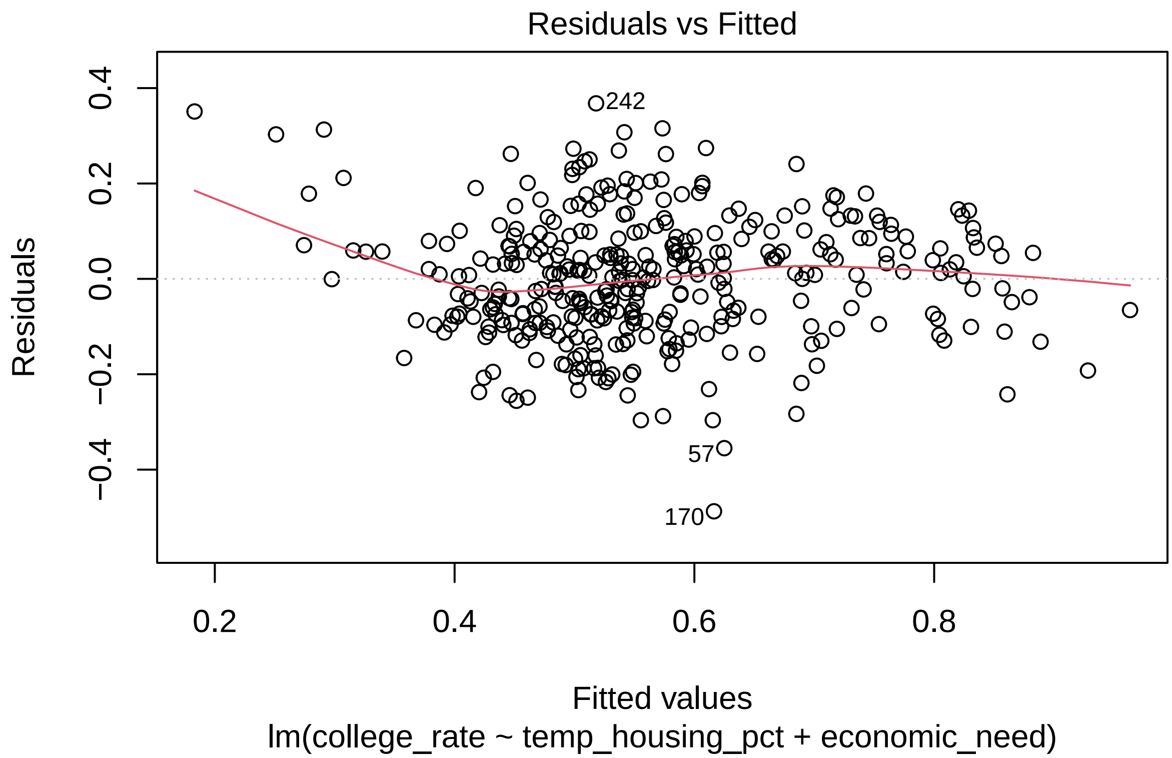


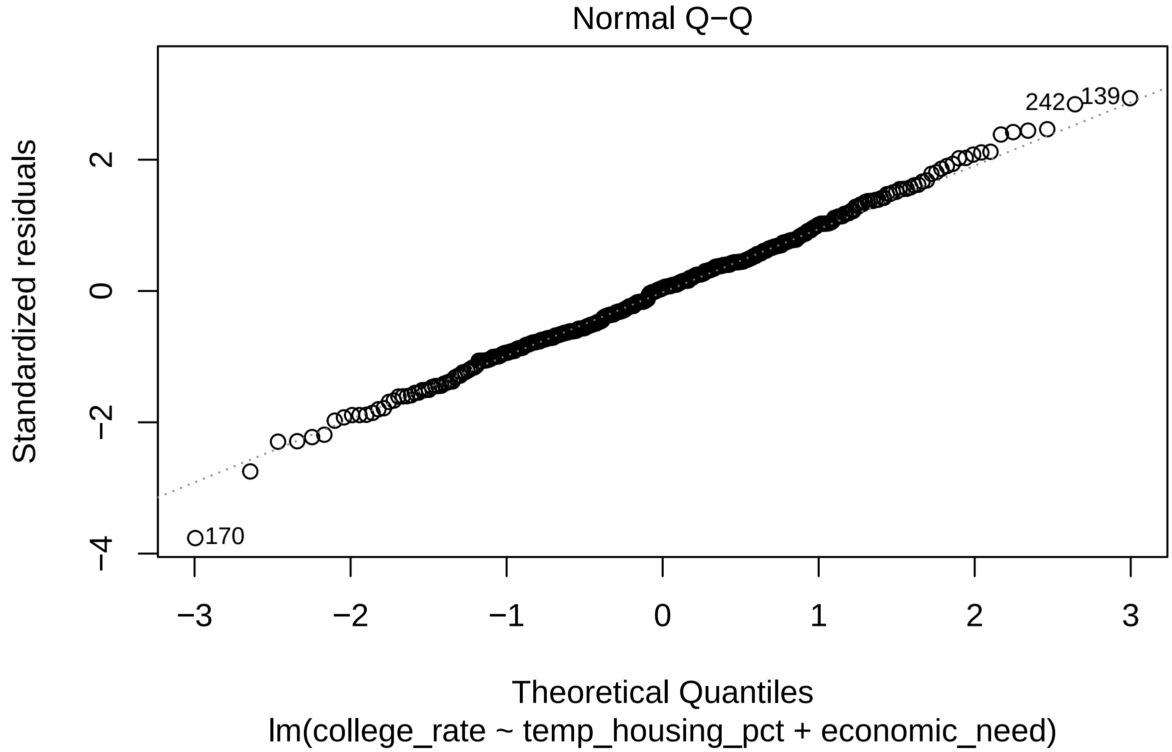
We see a general linear relationship for schools with lower rates of students in temp housing. However, this linear relationship does **not** visually hold for schools with hisgher rates of temp housing use.

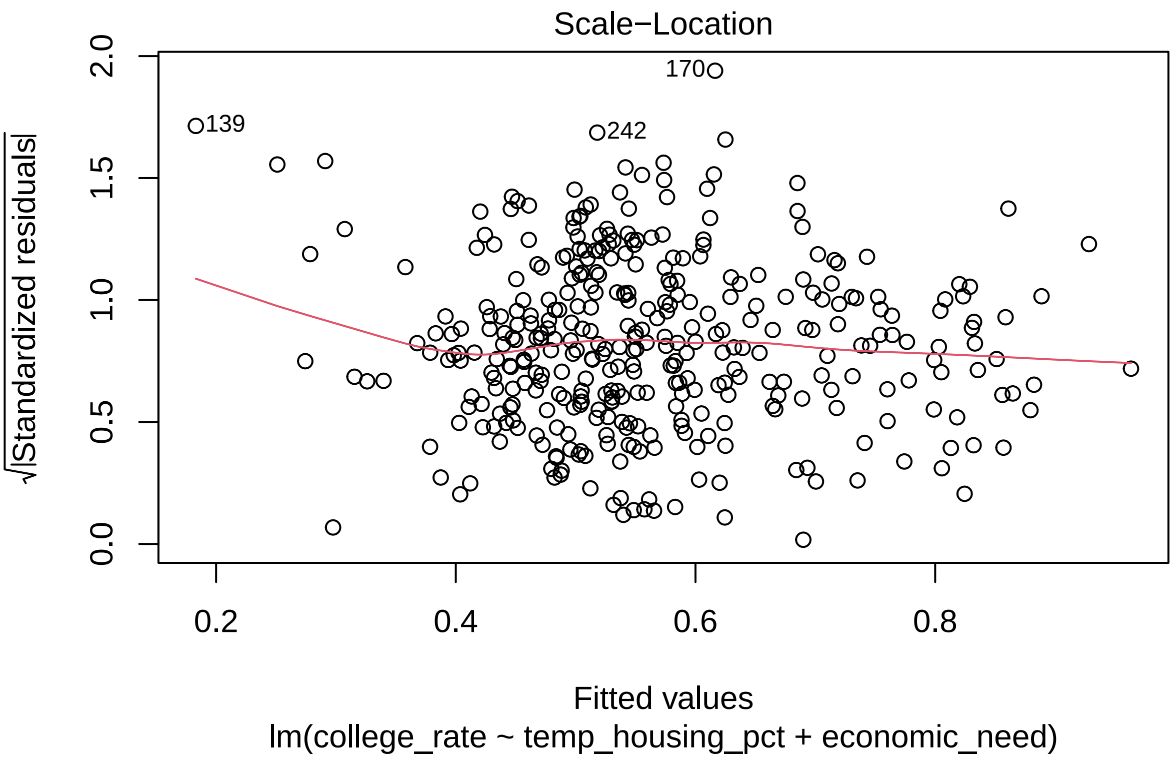
Plotting the relationship below between a school's economic need index

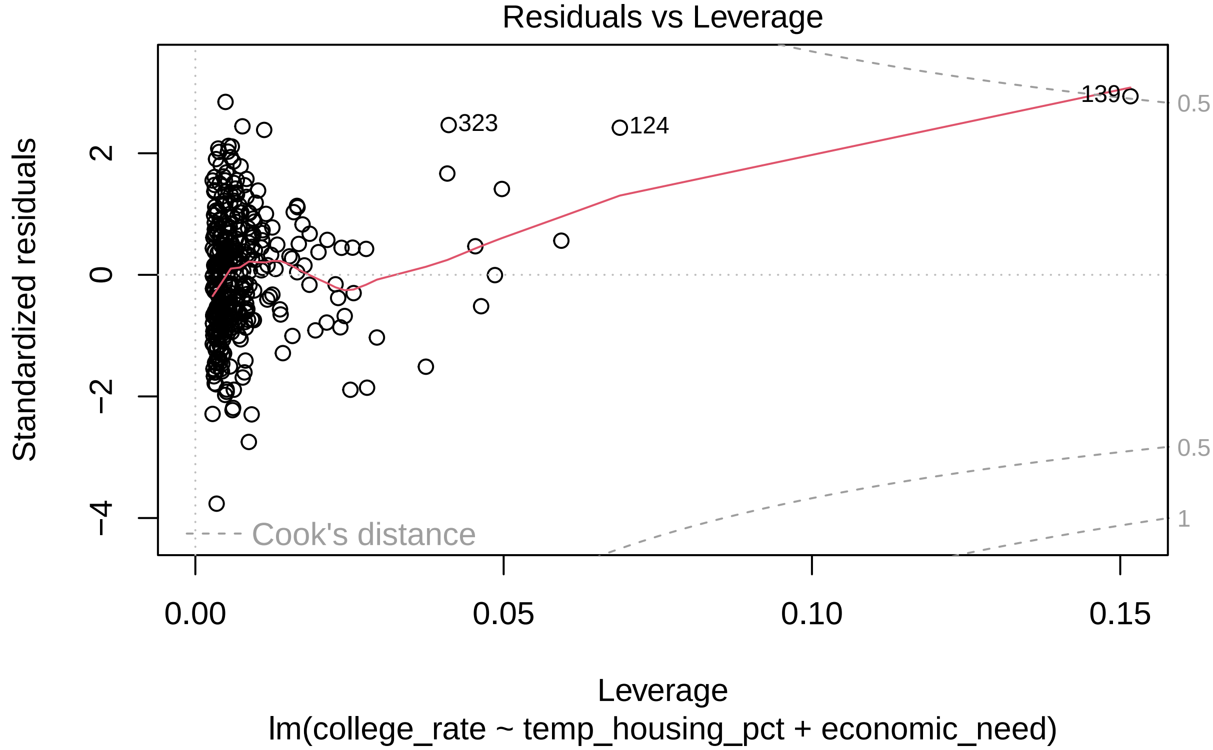


Again, we see a non-linear relationship between our predictor (*Economic Need Index*) and Outcome Variable (*College Persistence Rate*)


Residuals vs Fitted
Residuals
0.4
0.2
0.0
−0.2
−0.4
242
57
170
0.2
0.4
0.6
0.8
Fitted values
lm(college_rate ~ temp_housing_pct + economic_need)


Normal Q−Q
Standardized residuals
2
0
−2
−4
242
139
170
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(college_rate ~ temp_housing_pct + economic_need)


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
170
139
242
0.2
0.4
0.6
0.8
Fitted values
lm(college_rate ~ temp_housing_pct + economic_need)


Residuals vs Leverage
Standardized residuals
−4
−2
0
2
323
124
139
0.5
0.5
1
Cook's distance
0.00
0.05
0.10
0.15
Leverage
lm(college_rate ~ temp_housing_pct + economic_need)

Conclusion

TODO

- Merge/Join in ACT/SAT information by DBN
- Model Selection

References

Afarian, R., & Kleiner, B. (2003). The relationship between grades and career success. *Management Research News*, *26*, 42–51. https://doi.org/10.1108/01409170310783781

Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting achievement and providing support before STEM majors begin to fail. *Computers & Education*, *158*, 103999. https://doi.org/https://doi.org/10.1016/j.compedu.2020.103999

Education Statistics, N. C. for. (2008). *Percentage of high school dropouts among persons 16 through 24 years old*. Retrieved from https://nces.ed.gov/programs/digest/d08/tables/dt08_110.asp

Musso, M. F., Cascallar, E. C., Bostani, N., & Crawford, M. (2020). Identifying reliable predictors of educational outcomes through machine-learning predictive modeling. *Frontiers in Education*, *5*. https://doi.org/10.3389/feduc.2020.00104

New York City Schools, T. R. A. for. (2018). *Redesigning the Annual NYC School Survey: Lessons from a Research-Practice Partnership*. https://steinhardt.nyu.edu/sites/default/files/2021-01/Lessons_from_a_Research-Practice_Partnership.pdf.

Roth, P. L., BeVier, C. A., Switzer III, F. S., & Schippmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology*, *81*(5), 548–556. https://doi.org/10.1037/0021-9010.81.5.548

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11. https://doi.org/10.1186/s40561-022-00192-z

Appendices

Below is the code used to generate this report. It's also available on GitHub here

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)
library(tidyverse)
```

```r
library(gridExtra)
library(glue)
# library(autoReg)
library("papaja")
r_refs("r-references.bib")
# Read in our dataset from GitHub
# https://www.opendatanetwork.com/dataset/data.cityofnewyork.us/bm9v-cvch
df <- read.csv("https://data.cityofnewyork.us/api/views/26je-vkp6/rows.csv?date=2023110
label_cols <- c("dbn", "school_name", "school_type")
# Convert needed columns to numeric typing
df <- cbind(df[, label_cols], as.data.frame(lapply(df[,!names(df) %in% label_cols], as.

df$college_rate <- df$val_persist3_4yr_all
df$economic_need <- df$eni_hs_pct_912
set.seed(42)

# Adding a 20% holdout of our input data for model evaluation later
train <- subset(df[sample(1:nrow(df)), ]) %>% sample_frac(0.8)
test  <- dplyr::anti_join(df, train, by = 'dbn')

p1 <- ggplot(df, aes(x=survey_pp_CT, y=val_mean_score_act_math_all)) + geom_point() + la
p2 <- ggplot(df, aes(x=survey_pp_ES, y=val_mean_score_act_math_all)) + geom_point() + la
p3 <- ggplot(df, aes(x=survey_pp_RI, y=val_mean_score_act_math_all)) + geom_point() + la
p4 <- ggplot(df, aes(x=survey_pp_SE, y=val_mean_score_act_math_all)) + geom_point() + la
p5 <- ggplot(df, aes(x=survey_pp_SF, y=val_mean_score_act_math_all)) + geom_point() + la
p6 <- ggplot(df, aes(x=survey_pp_TR, y=val_mean_score_act_math_all)) + geom_point() + la
```

```r
# Plot english scores
p7 <- ggplot(df, aes(x=survey_pp_CT, y=val_mean_score_act_engl_all)) + geom_point() + la
p8 <- ggplot(df, aes(x=survey_pp_ES, y=val_mean_score_act_engl_all)) + geom_point() + la
p9 <- ggplot(df, aes(x=survey_pp_RI, y=val_mean_score_act_engl_all)) + geom_point() + la
p10 <- ggplot(df, aes(x=survey_pp_SE, y=val_mean_score_act_engl_all)) + geom_point() + l
p11 <- ggplot(df, aes(x=survey_pp_SF, y=val_mean_score_act_engl_all)) + geom_point() + l
p12 <- ggplot(df, aes(x=survey_pp_TR, y=val_mean_score_act_engl_all)) + geom_point() + l

# Panel plot
grid.arrange(
  p1, p2,
  p3, p4,
  p5, p6,
  p7, p8,
  p9, p10,
  p11, p12,
  nrow=2,
  ncol=6,
  top = "ACT Scores vs NYC School Quality Ratings, 2020-2021",
  bottom="Survey Rating Type"
)

english_formula <- val_mean_score_act_engl_all ~ survey_pp_RI + survey_pp_CT + survey_pp

math_formula <-  val_mean_score_act_math_all ~ survey_pp_RI + survey_pp_CT + survey_pp_E

# Create lineaer model to predict english and math scores based on sruvey ratings
```

```
lm_english <- lm(english_formula, data=train)

lm_math <- lm(math_formula, data=train)

summary(lm_english)

summary(lm_math)

hist(train$college_rate)

ggplot(train, aes(x=temp_housing_pct)) + geom_histogram() + labs(x="% of Students in Ter

ggplot(train, aes(x=temp_housing_pct, y=college_rate)) + geom_point() + labs(x="% of Stu

ggplot(train, aes(x=economic_need, y=college_rate)) + geom_point() +
  labs(x="Average Economic Need Index per School", y="4-year College Persistence Rate")

proxy_lm <- lm(college_rate ~ temp_housing_pct + economic_need, train)

plot(proxy_lm)
```