

# Predicting Math Success from Self-Reported Attitude

---

COMP 4441 Final Project

Luke Sonnanburg

# Research Question

---

Do student attitudes towards math predict measurable success in learning?

Can we identify a mindset or set of attitudes that predicts success in math learning?

- Q1: "I like math."
- Q2: "I'm good at math."
- Q3: "I'm capable of difficult math."
- Q4: "Math is my favorite subject."
- Q5: "I understand most math I'm taught."
- Q6: "Math is easy."
- Q7: "I do my best in math."
- Q8: "Math is useful for everyday life."
- Q9: "Math is useful in other subjects."
- Q10: "Math is useful for what I want to study."
- Q11: "Math is useful for the job I want."



# Why Does This Matter?

---

- Learning outcomes
  - Mathematical literacy
  - Skilled workforce
- Quality of life
  - Mathphobia
  - Student stress
- Allocation of resources, messaging

# Dataset

---

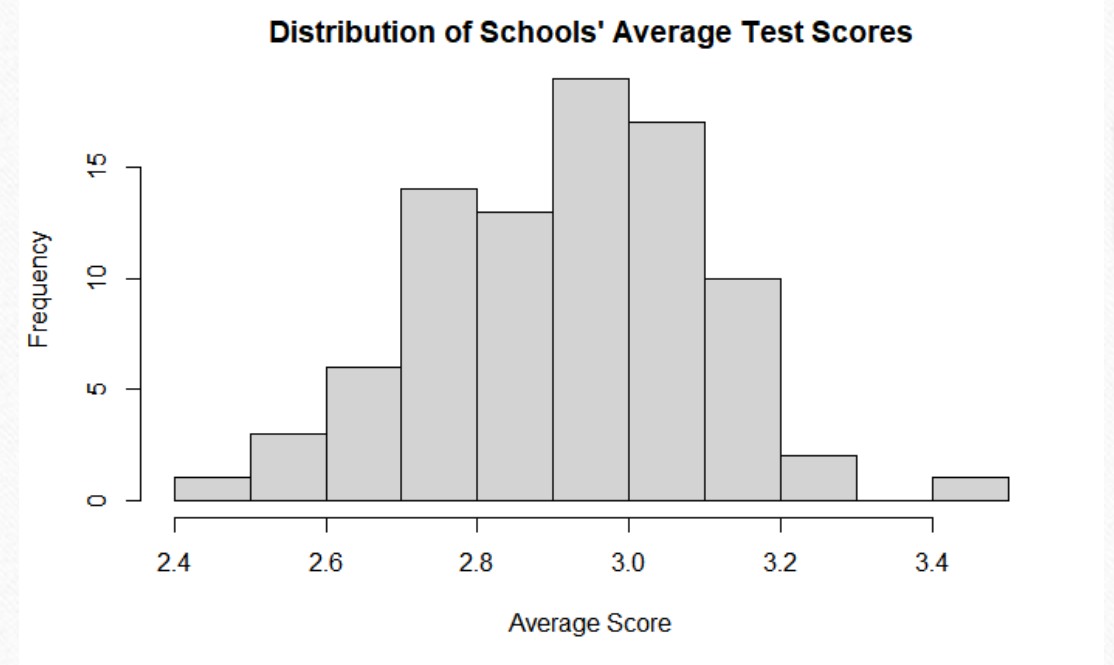
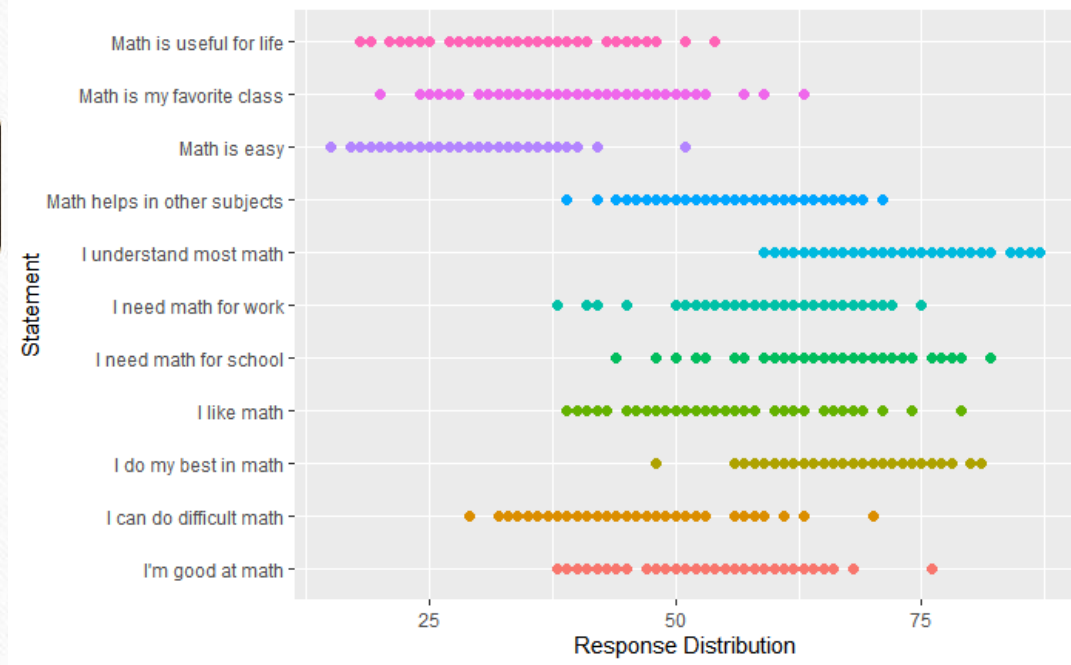
Subjects: 9th Grade Ontario math students

Input variables: Q1, Q2, ... Q11, represent the proportion of students in a given school who responded that they "agree" or "strongly agree" to the 11 queries

Output variable: Average score of students 9th grade standardized math test (individual scores range from 1 to 4)

All data is on the scale of entire school populations

# Dataset

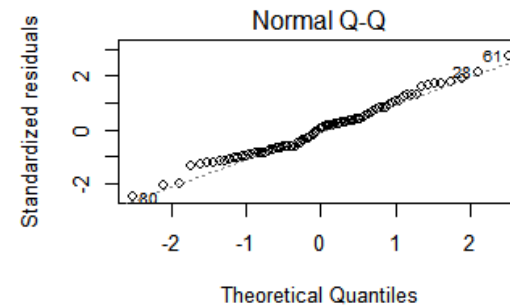
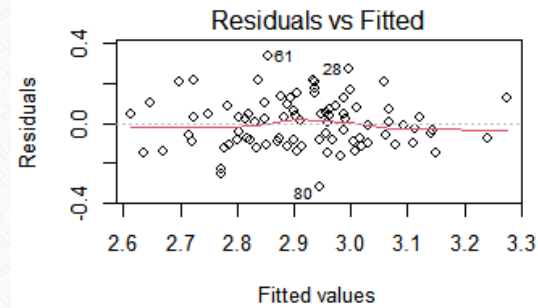


Survey response and standardized test data come from EQAO, Ontario's government agency for public schools (2013-2014)



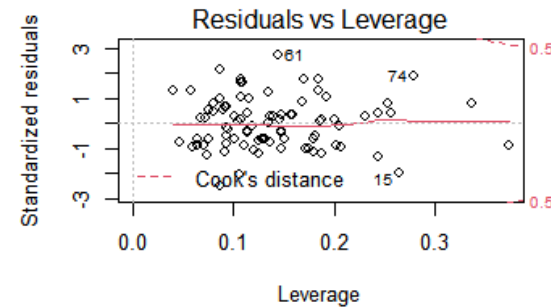
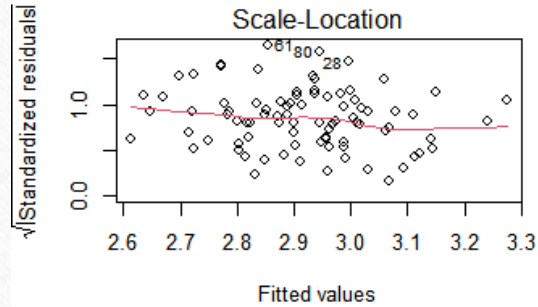
# Data Prep: Assumption Checking

Collinearity



Normally distributed errors

Very minor heteroscedasticity  
(attempting to fix it made it worse)



No extreme outliers

# Data Prep: Feature Selection

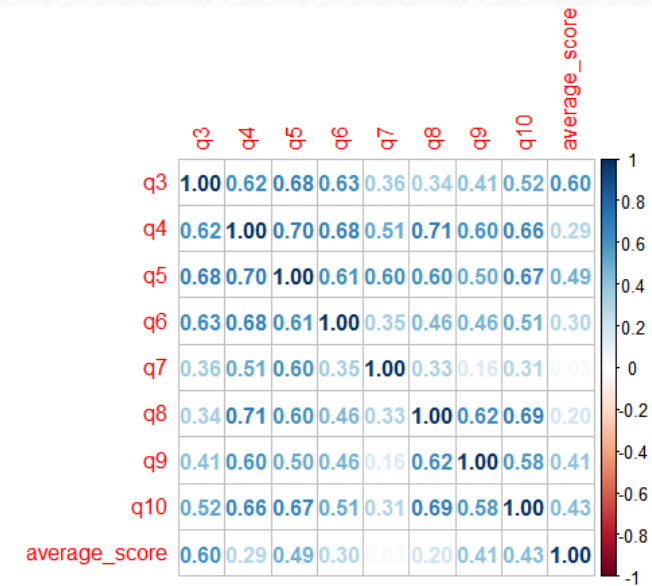
```
Call:
lm(formula = average_score ~ ., data = main_model[2:13])

Residuals:
    Min       1Q   Median       3Q      Max
-0.31787 -0.09233  0.00498  0.07845  0.33532

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.835992   0.236649   7.758 3.71e-11 ***
q1           0.001735   0.004760    0.365  0.71646
q2           0.001635   0.004680    0.349  0.72781
q3           0.011600   0.003919    2.960  0.00413 **
q4          -0.004719   0.004434   -1.064  0.29060
q5           0.011596   0.004512    2.570  0.01218 *
q6          -0.004817   0.003243   -1.486  0.14164
q7          -0.008722   0.003386   -2.576  0.01199 *
q8          -0.004599   0.003236   -1.421  0.15947
q9           0.005872   0.002894    2.029  0.04603 *
q10          0.006632   0.005133    1.292  0.20032
q11          -0.002593   0.004798   -0.540  0.59052
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1329 on 74 degrees of freedom
Multiple R-squared:  0.5403,    Adjusted R-squared:  0.4719
F-statistic: 7.906 on 11 and 74 DF,  p-value: 6.779e-09
```

Drop very large p-values: q1, q2, q11



Drop values with weak correlation to score but strong correlation to other values: q4, q7, q8

# Data Visualization



|               | q2   | q3   | q5   | q6   | q9   | average_score |
|---------------|------|------|------|------|------|---------------|
| q2            | 1.00 | 0.84 | 0.71 | 0.73 | 0.36 | 0.47          |
| q3            | 0.84 | 1.00 | 0.68 | 0.63 | 0.41 | 0.60          |
| q5            | 0.71 | 0.68 | 1.00 | 0.61 | 0.50 | 0.49          |
| q6            | 0.73 | 0.63 | 0.61 | 1.00 | 0.46 | 0.30          |
| q9            | 0.36 | 0.41 | 0.50 | 0.46 | 1.00 | 0.41          |
| average_score | 0.47 | 0.60 | 0.49 | 0.30 | 0.41 | 1.00          |



# Methodology

---

- Tested 3 related types of linear regression
  - LASSO, Ridge, Elastic Net
  - Each uses "shrinkage" to reduce low-impact coefficients
  - Shrinks data values closer to the mean
- 80/20 train/test split, model selected based on accuracy/simplicity

# Comparison

---

|             | RMSE      | R-squared | Number of Variables |
|-------------|-----------|-----------|---------------------|
| Ridge       | 0.1459139 | 0.3556914 | 5                   |
| Lasso       | 0.1510877 | 0.3428642 | 3                   |
| Elastic Net | 0.1500432 | 0.3440466 | 4                   |

LASSO regression results in only 3 variables

It's chosen for simplicity, as it is only very slightly less accurate in its predictions than Elastic Net

# Lasso Regression

---

- Forms of linear regression with "shrinkage"
- L1 regularization; shrinks coefficients based on absolute value of their magnitude and some constant lambda
- Goal is to minimize  $\sum_{i=1}^n (y_i - \sum_j x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$  ; this shrinks model closer to mean
- Implemented via R's glmnet library, including automatic lambda selection of 0.010212

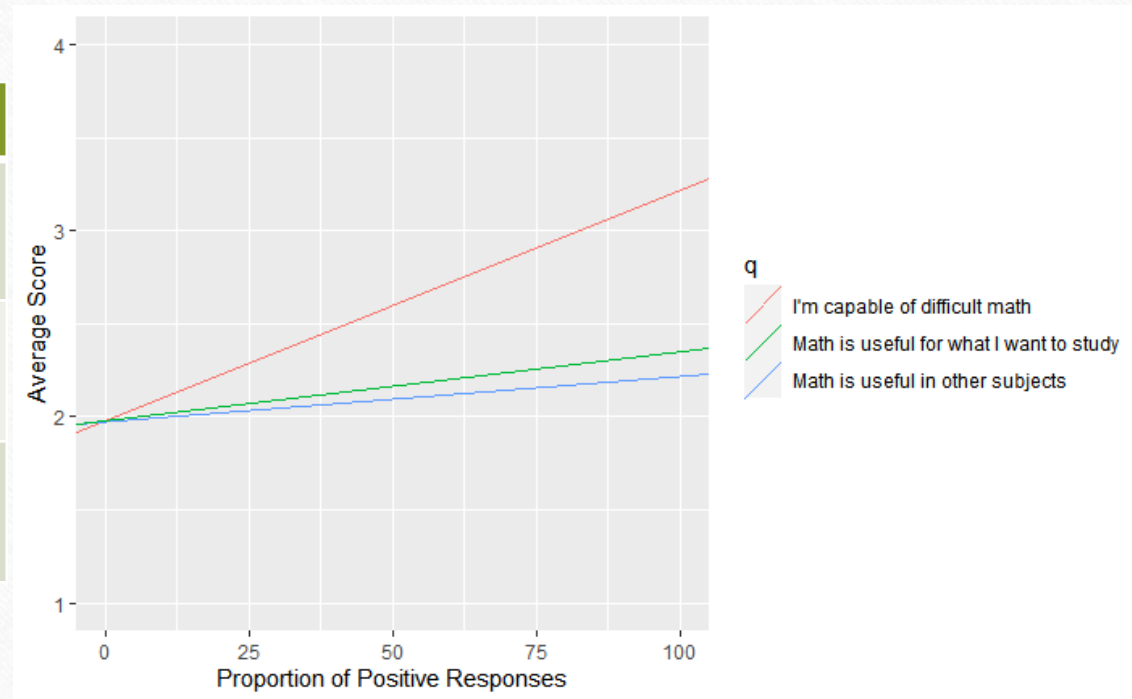


# Major Data Analysis & Modeling

Results:

| Variable                                  | Weight |
|---|--------|
| "I'm capable of difficult math"           | .0124  |
| "Math is useful in other subjects"        | .0024  |
| "Math is useful for what I want to study" | .0037  |

R-squared of .343; about 1/3 of variance can be explained by these variables



# Can we say that attitude predicts student success in mathematics?

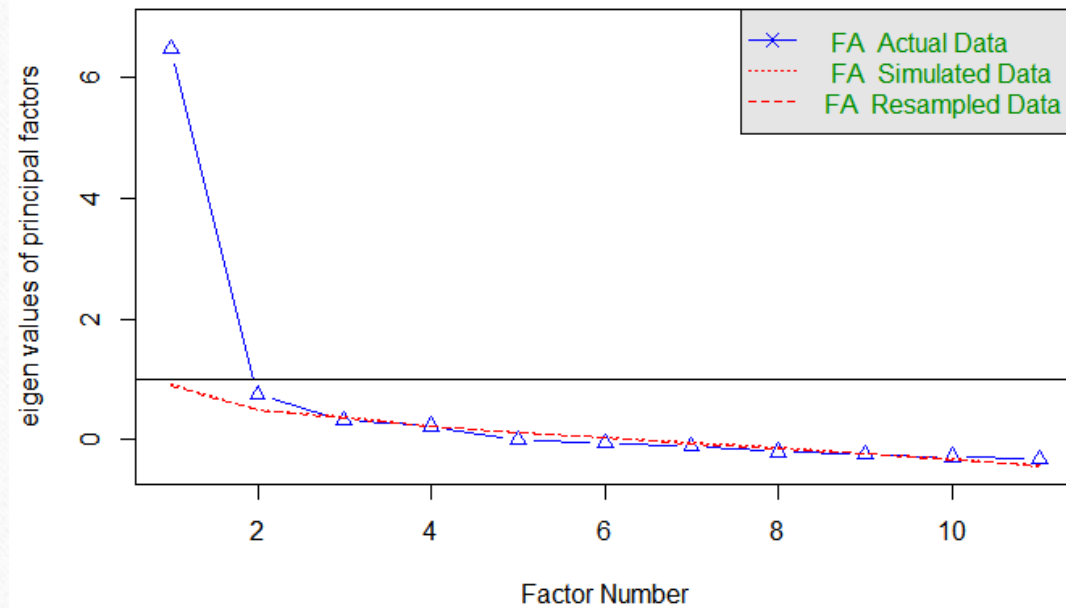
---

- 1/3 of variance explained by student attitude is higher than expected
- The greatest weight being given to "I can do difficult math"... might skew impressions
- Other prominent variables could be described as appreciation for math's applications; perhaps students would perform better with more focus on applied math problems

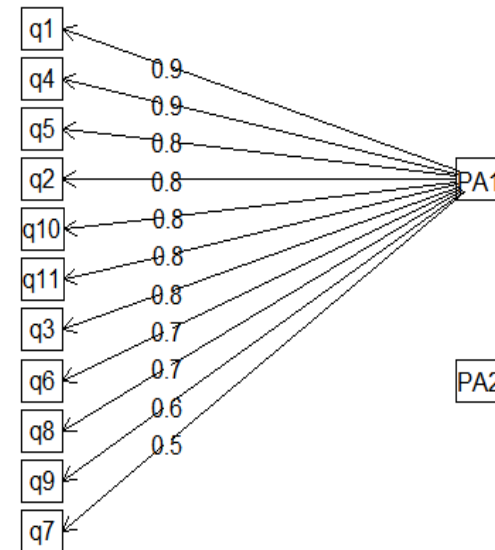
So to summarize: Technically yes, but I would suggest we could get more interesting results with different survey prompts

## Afterthoughts: Devising Alternate Questions With Factor Analysis

Parallel Analysis Scree Plots



Factor Analysis With No Rotation and Two Factors





## Afterthoughts: Devising Alternate Questions With Factor Analysis

| PA | Statement   |
|----|---|
| 1  | Q4: Math is one of my favorite subjects                   |
|    | Q1: I like mathematics                                    |
|    | Q8: Math is useful in everyday life                       |
|    | Q9: Math helps in other subjects                          |
| 2  | Q2: I am good at math                                     |
|    | Q3: I am able to answer difficult math questions          |
|    | Q6: Math is easy  |
|    | Q5: I understand math that I'm taught                     |
|    | Q7: I do my best in math                                  |
| 3  | Q10: I need to do well in math to study what I want later |
|    | Q11: I need to keep taking math for the job I want later  |

Factor Analysis With Varimax Rotation and Three Factors

