How well do PSAT scores predict SAT scores?

-Daniel Ellis Schwartz

Big Question

Which is a better predictor of SAT scores, PSAT scores or demographic data?

- How do they compare?
- What does this mean for the PSAT?

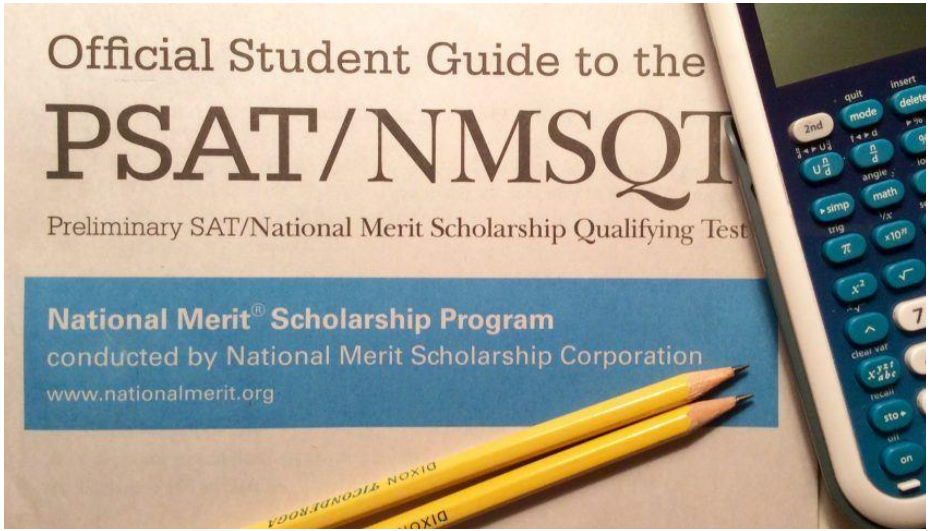
Results Summary

- PSAT is slightly better
- Demographic data is still pretty good
- Ambiguous implications

Background: What is the PSAT?

Preliminary SAT:

- Two parts
  - Math
  - English Reading & Writing (EBRW)
- High school sophomores
- Only affects scholarships
- Supposedly approximates future SAT scores



Background: Costs of Standardized Testing

- Financial
- Educational opportunity cost
- Health & wellbeing

Some tests in Colorado:

- CMAS
- MAPS
- SAT/PSAT
- ACCESS testing
- District specific testing

Can be 10+ test per year

Data Overview

- PSAT scores from 2017
- SAT scores from 2018
- Scores subdivided by
  - School
  - Gender
  - Ethnicity/Race
  - Free & Reduced Lunch eligible (FRL)
  - English Language Learner (ELL)
  - Individualized Education Plan status (IEP)
- Scores only reported if >16 in subgroup
- Median family income of school's zip code

COLORADO
Department of Education

SCHOOLview® State Assessment Data Lab

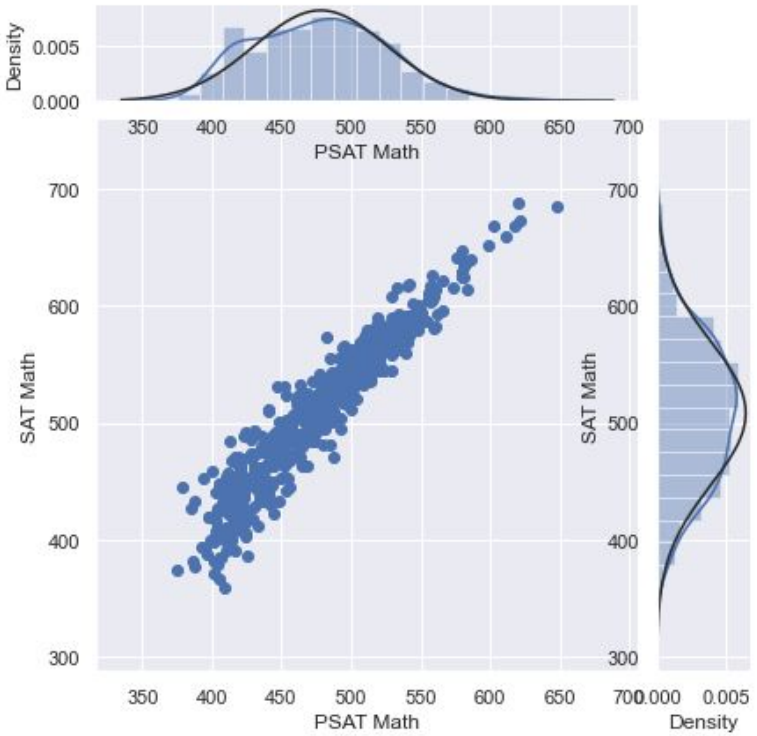
Scores & Scoring

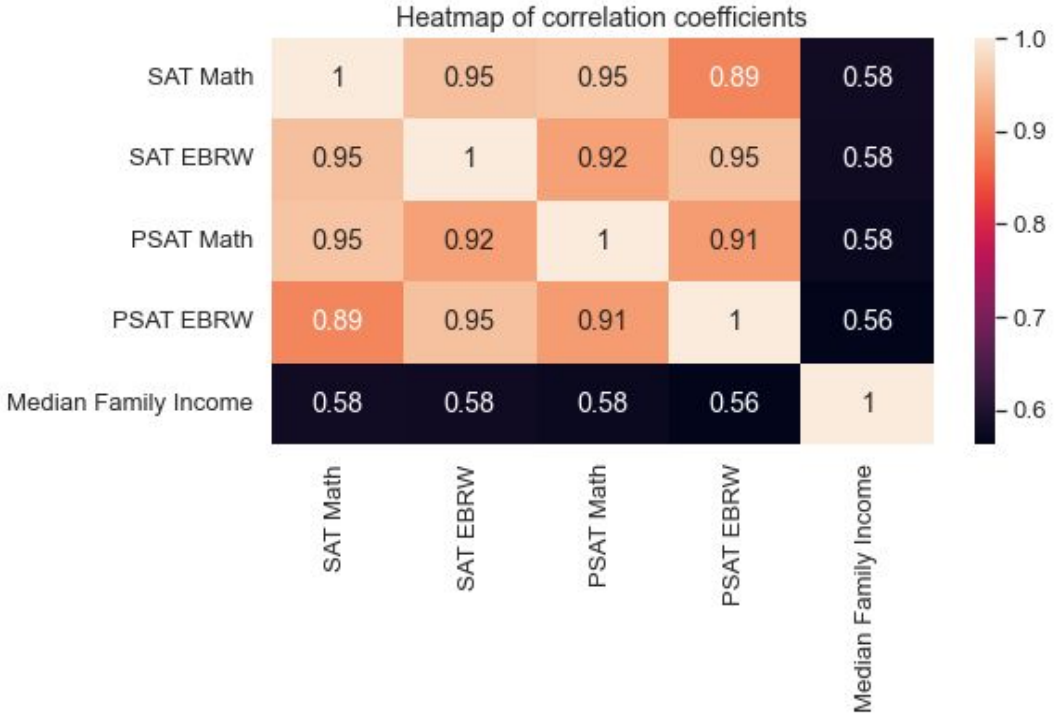
- Scores range from 200 to 800
- Math & EBRW scored separately
- No direct conversion from total correct
- Mean scores reported for each subgroup

RAW SCORE CONVERSION TABLE 1 SECTION AND TEST SCORES

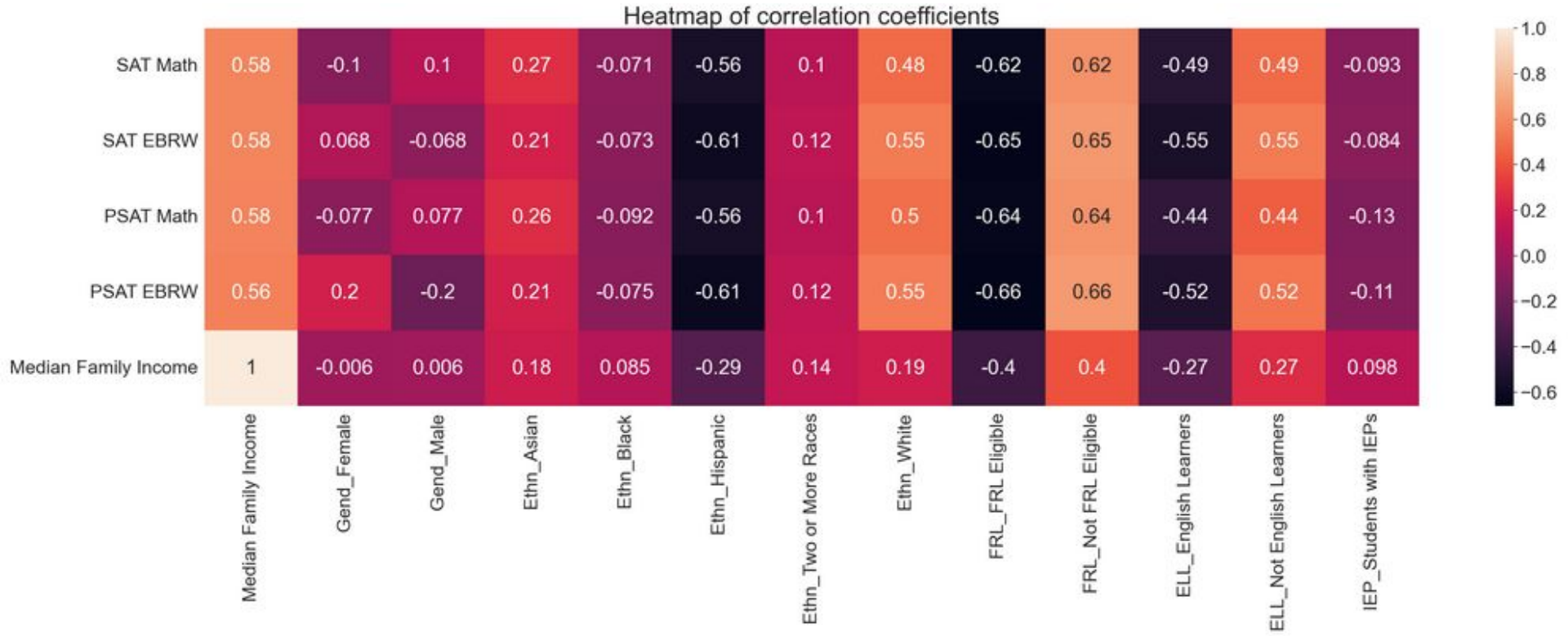
Raw Score (# of correct answers)	Math Section Score	Reading Test Score	Writing and Language Test Score
0	200	10	10
1	200	10	10
2	210	10	10
3	230	11	10
4	240	12	11
5	260	13	12
6	280	14	13
7	290	15	13
8	310	15	14
9	320	16	15
10	330	17	16
11	340	17	16
12	360	18	17
13	370	19	18
14	380	19	19
15	390	20	19
16	410	20	20
17	420	21	21
18	430	21	21
19	440	22	22
20	450	22	23
21	460	23	23
22	470	23	24
23	480	24	25
24	480	24	25
25	490	25	26
26	500	25	26
27	510	26	27
28	520	26	28
29	520	27	28
30	530	28	29
31	540	28	30
32	550	29	30
33	560	29	31
34	560	30	32
35	570	30	32
36	580	31	33
37	590	31	34
38	600	32	34
39	600	32	35
40	610	33	36
41	620	33	37
42	630	34	38
43	640	35	39
44	650	35	40
45	660	36	
46	670	37	
47	670	37	
48	680	38	
49	690	38	
50	700	39	
51	710	40	
52	730	40	
53	740		
54	750		
55	760		
56	780		
57	790		
58	800		

Initial Score Exploration





Initial Exploration



PSAT Based Model

Linear regression

Model metrics

Subject	MSE	RMSE	MAE	MAPE
Math	308.50	17.56	13.32	2.71%
EBRW	267.51	16.36	12.88	2.50%

MAE ~ one test question

Demographics Based Model

Features:

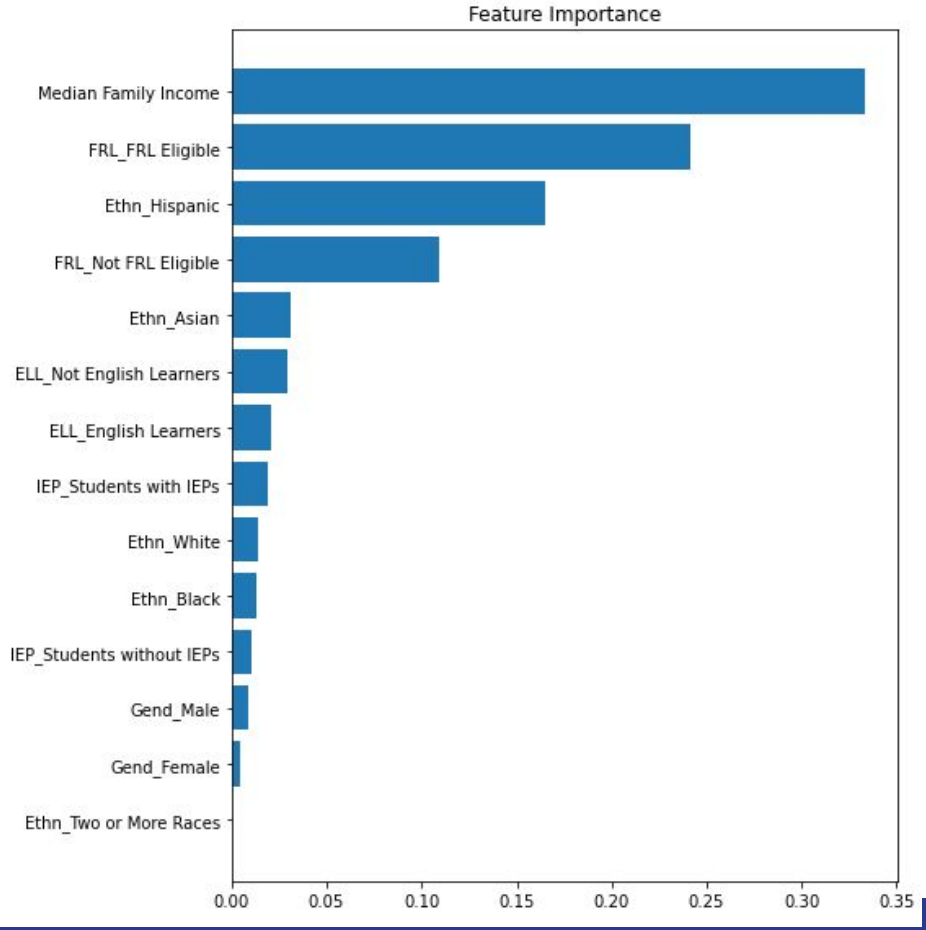
- Gender
- Ethnicity/Race
- Free & Reduced Lunch eligible (FRL)
- English Language Learner (ELL)
- Individualized Education Plan status (IEP)
- Median family income of school’s zip code

Models used:

- Linear regression
- Gradient boost regression (CV)
- Random forest regression (CV)

Best model: Gradient boost regression (CV)

Best Model Overview



Subject	MAE	MAPE
Math	21.62	4.21%
EBRW	19.69	3.73%

PSAT v Demographics Comparison

Method	RMSE	MAE	Raw score equivalent MAE	MAPE
PSAT	17.56	13.32	1.29	2.71%
Demographics	29.29	21.62	2.09	4.21%

- Each math test question ~10.34 points
- Difference in MAE ~0.8 questions
- Standard deviation of test scores: 62.61

Results

- PSAT scores are a better predictor than demographics but not by much
  - MAE for PSAT model: 13.32
  - MAE for demographic model: 21.62
  - Max possible score on test: 800
- Demographics model still makes good predictions
- Key features:
  - Median family income
  - Race/ethnicity
  - FRL eligibility

Conclusions & Recommendations

- If PSAT scores are unavailable then demographic data can work to predict SAT scores
- Neither SAT nor PSAT can be used to directly evaluate a school's academic performance
- PSAT should not be required for students