An Analysis of High School Test Score for College in Northeastern U.S. Metropolitan Area

Introduction/Study Purpose

Today most good jobs require more than a high school diploma. Businesses want to hire people who know how to think and solve problems. Besides, most college graduates earn a lot more money during their working years than people who stop their education at high school.

The purpose of the study is to find out what factors would affect high school students getting better test score for college. The focus of this study is also to examine the relationships between college test score (for example, SAT Math score, SAT Verbal score, 10^{th} grade MCAS English score, 10^{th} grade MCAS Math score), and certain existing school recourses, such as school enrollment, school cost/expense, teacher salary, student/teacher ratio, student/counselor ratio, school dropout rate, SAT participation rate, and percentage of students going to college.

Data Preparation

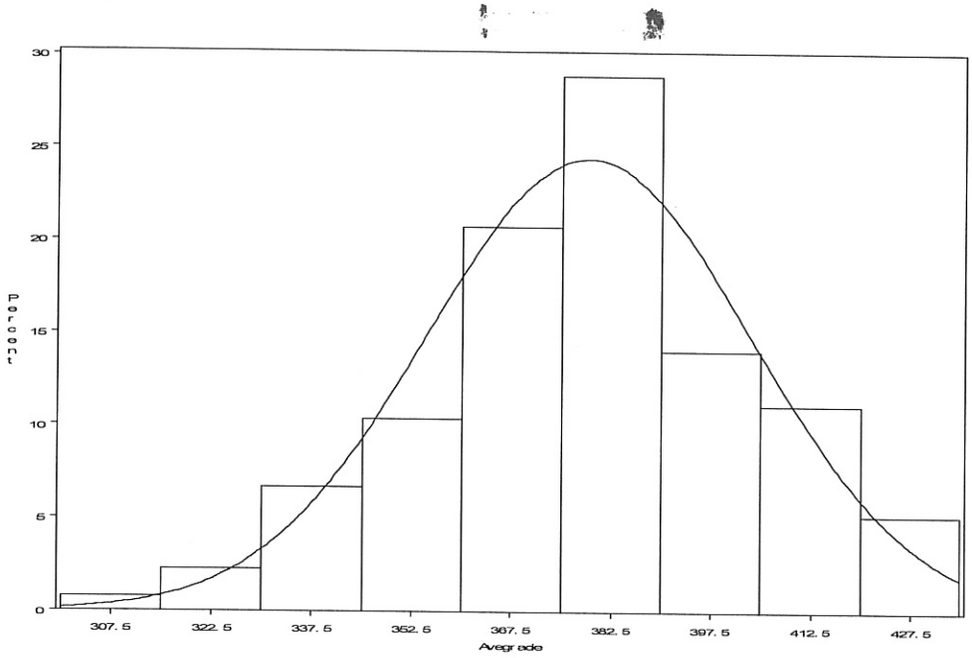
The data are collected from 135 public high schools in Northeastern U.S. metropolitan area. Since 4 major test scores (SAT Math score, SAT Verbal score, 10^{th} grade MCAS English score, 10^{th} grade MCAS Math score) are highly correlated in this case, they are re-calculated and transformed into one average test core.

Spearman Correlation Coefficients, N = 135

Prob > |r| under H0: Rho=0

	10GMCASMth	10GMCASEng	SATM	SATV
10GMCASMth	1.000	0.94524	0.89241	0.88579
10GMCASEng	0.94524	1.000	0.85404	0.87979
SATM	0.89241	0.85404	1.000	0.93956
SATV	0.88579	0.87979	0.93956	1.000

By using General Liner Regression Model to predict the average test score (or called Avegrade in the following histogram), and it shows that the average test score is normally distributed.



First we can test if data is following the normal distribution.

Ho: data is normally distributed.

Ha: data is not normally distributed.

The p-values we got are all greater than the critical value (0.05), so it fails to reject Ho. We conclude the data is normally distributed, and now we are able to conduct hypothesis testing.

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.99012	Pr < W	0.4557
Kolmogorov-Smirnov	D	0.046128	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.050418	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.327082	Pr > A-Sq	>0.2500

From the result of Analysis of Parameter Estimates, it shows that only Student/Teacher Ratio, Student/Counselor Ratio, SAT Participation Rate, School Dropout Rate are statistically significant.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-2.3957	0.1889	-2.7660	-2.0256	160.89	<.0001
S_TRatio	1	0.0405	0.0040	0.0327	0.0484	103.01	<.0001
S_CounselRatio	1	-0.0000	0.0001	-0.0003	0.0003	0.01	0.9359
SATPartRate	1	0.0246	0.0009	0.0229	0.0263	818.20	<.0001
DropoutRate	1	0.0143	0.0033	0.0078	0.0208	18.53	<.0001
AveTeach_	1	-0.0017	0.0014	-0.0045	0.0010	1.60	0.2053
Cost_Pupil	1	0.0069	0.0047	-0.0024	0.0161	2.11	0.1459
Scale	0	1.0000	0.0000	1.0000	1.0000		

Since we know the rest of variables are not significant in our model, eliminating them from our model would give us a better result.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	277.9284	16.2897	246.0011	309.8557	291.10	<.0001
S_TRatio	1	0.0405	0.0040	0.0327	0.0484	103.01	<.0001
SATPartRate	1	1.1017	0.1092	0.8877	1.3157	101.83	<.0001
S_CounselRatio	1	-0.0414	0.0215	-0.0834	0.0007	3.71	0.0341
DropoutRate	1	-3.0043	0.6474	-4.2732	-1.7355	21.54	<.0001

A better and higher high school test score is a free ticket to any colleges. What we have learned can be useful information for parents or students that a better Student/Teacher Ratio and Student/Counselor Ratio would directly affect student to achieve higher test score in school as well as a higher SAT participation rate and lower school dropout rate could be important factors. Before choosing which high school they want to attend, these four major factors may help them to make their final decision.

Appendix 1

The CORR Procedure

Spearman Correlation Coefficients, N = 135
Prob > |r| under H0: Rho=0

	_0GMCASMth	_0GMCASEng	S_TRatio	S_ Counsel Ratio	SATM
_0GMCASMth	1.00000	0.94524	-0.25325	-0.23560	0.89241
10GMCASMth		<.0001	0.0030	0.0059	<.0001
_0GMCASEng	**0.94524**	1.00000	-0.22322	-0.25113	0.85404
10GMCASEng	<.0001		0.0093	0.0033	<.0001
S_TRatio	-0.25325	-0.22322	1.00000	0.32370	-0.18654
S/TRatio	0.0030	0.0093		0.0001	0.0303
S_CounselRatio	-0.23560	-0.25113	0.32370	1.00000	-0.27164
S/CounselRatio	0.0059	0.0033	0.0001		0.0014
SATM	**0.89241**	0.85404	-0.18654	-0.27164	1.00000
SATM	<.0001	<.0001	0.0303	0.0014	
SATV	**0.88579**	0.87979	-0.21046	-0.23805	0.93956
SATV	<.0001	<.0001	0.0143	0.0054	<.0001
SATPartRate	0.81320	0.78581	-0.26678	-0.18476	0.77466
SATPartRate	<.0001	<.0001	0.0018	0.0319	<.0001
DropoutRate	-0.68165	-0.70428	0.23922	0.13779	-0.59695
DropoutRate	<.0001	<.0001	0.0052	0.1110	<.0001
AveTeach_	0.32077	0.29955	-0.03010	-0.09905	0.29624
AveTeach$	0.0001	0.0004	0.7289	0.2531	0.0005
Cost_Pupil	0.31166	0.30412	-0.51066	-0.36051	0.28805
Cost/Pupil	0.0002	0.0003	<.0001	<.0001	0.0007

Spearman Correlation Coefficients, N = 135
Prob > |r| under H0: Rho=0

	SATV	SATPart Rate	Dropout Rate	Ave Teach_	Cost_ Pupil
_0GMCASMth	0.88579	0.81320	-0.68165	0.32077	0.31166
10GMCASMth	<.0001	<.0001	<.0001	0.0001	0.0002
_0GMCASEng	0.87979	0.78581	-0.70428	0.29955	0.30412
10GMCASEng	<.0001	<.0001	<.0001	0.0004	0.0003

Spearman Correlation Coefficients, N = 135
Prob > |r| under H0: Rho=0

	SATV	SATPart Rate	Dropout Rate	Ave Teach_	Cost_ Pupil
S_TRatio	-0.21046	-0.26678	0.23922	-0.03010	-0.51066
S/TRatio	0.0143	0.0018	0.0052	0.7289	<.0001
S_CounselRatio	-0.23805	-0.18476	0.13779	-0.09905	-0.36051
S/CounselRatio	0.0054	0.0319	0.1110	0.2531	<.0001
SATM	0.93956	0.77466	-0.59695	0.29624	0.28805
SATM	<.0001	<.0001	<.0001	0.0005	0.0007

SATV	1.00000	0.78206	-0.64034	0.27212	0.25217
SATV		<.0001	<.0001	0.0014	0.0032
SATPartRate	0.78206	1.00000	-0.66479	0.25439	0.23629
SATPartRate	<.0001		<.0001	0.0029	0.0058
DropoutRate	-0.64034	-0.66479	1.00000	-0.30334	-0.27641
DropoutRate	<.0001	<.0001		0.0003	0.0012
AveTeach_	0.27212	0.25439	-0.30334	1.00000	0.28919
AveTeach$	0.0014	0.0029	0.0003		0.0007
Cost_Pupil	0.25217	0.23629	-0.27641	0.28919	1.00000
Cost/Pupil	0.0032	0.0058	0.0012	0.0007	

Appendix 2

The UNIVARIATE Procedure
Variable: Avegrade

Moments

N	135	Sum Weights	135
Mean	379.035185	Sum Observations	51169.75
Std Deviation	24.5040958	Variance	600.450712
Skewness	-0.1843463	Kurtosis	0.03808041
Uncorrected SS	19475596.1	Corrected SS	80460.3954
Coeff Variation	6.46486046	Std Error Mean	2.10897678

Basic Statistical Measures

Location		Variability	
Mean	379.0352	Std Deviation	24.50410
Median	379.7500	Variance	600.45071
Mode	367.0000	Range	130.00000
		Interquartile Range	32.00000

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	179.7247	Pr > |t|	<.0001
Sign	M	67.5	Pr >= |M|	<.0001
Signed Rank	S	4590	Pr >= |S|	<.0001

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.99012	Pr < W	0.4557
Kolmogorov-Smirnov	D	0.046128	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.050418	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.327082	Pr > A-Sq	>0.2500

Quantiles (Definition 5)

Quantile	Estimate
100% Max	433.00
99%	425.00
95%	420.50
90%	413.00
75% Q3	394.75

The UNIVARIATE Procedure
Variable: Avegrade

Quantiles (Definition 5)

Quantile	Estimate
50% Median	379.75
25% Q1	362.75
10%	345.00
5%	338.50
1%	320.00
0% Min	303.00

Extreme Observations

----Lowest----		-----Highest----	
Value	Obs	Value	Obs
303	53	424.25	83
320	26	424.75	2
327	111	424.75	124
327	85	425.00	84
336	14	433.00	128

The UNIVARIATE Procedure
Fitted Distribution for Avegrade

Parameters for Normal Distribution

Parameter	Symbol	Estimate
Mean	Mu	379.0352
Std Dev	Sigma	24.5041

Goodness-of-Fit Tests for Normal Distribution

Test	---Statistic----		-----p Value-----	
Kolmogorov-Smirnov	D	0.04612752	Pr > D	>0.150
Cramer-von Mises	W-Sq	0.05041797	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.32708199	Pr > A-Sq	>0.250

Quantiles for Normal Distribution

	------Quantile------	
Percent	Observed	Estimated
1.0	320.000	322.030
5.0	338.500	338.730
10.0	345.000	347.632
25.0	362.750	362.507
50.0	379.750	379.035
75.0	394.750	395.563
90.0	413.000	410.438
95.0	420.500	419.341
99.0	425.000	436.040

Appendix 3

The GENMOD Procedure

Model Information

Data Set	WORK.A
Distribution	Normal
Link Function	Identity
Dependent Variable	Avegrade

Number of Observations Read	135
Number of Observations Used	135

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	126	22047.5925	174.9809
Scaled Deviance	126	135.0000	1.0714
Pearson Chi-Square	126	22047.5925	174.9809
Scaled Pearson X2	126	135.0000	1.0714
Log Likelihood		-535.5154	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-2.3957	0.1889	-2.7660	-2.0256	160.89	<.0001
S_TRatio	1	0.0405	0.0040	0.0327	0.0484	103.01	<.0001
S_CounselRatio	1	-0.0000	0.0001	-0.0003	0.0003	0.01	0.9359
SATPartRate	1	0.0246	0.0009	0.0229	0.0263	818.20	<.0001
DropoutRate	1	0.0143	0.0033	0.0078	0.0208	18.53	<.0001
AveTeach_	1	-0.0017	0.0014	-0.0045	0.0010	1.60	0.2053
Cost_Pupil	1	0.0069	0.0047	-0.0024	0.0161	2.11	0.1459

Appendix 4

The GENMOD Procedure

Model Information

Data Set	WORK.A
Distribution	Normal
Link Function	Identity
Dependent Variable	Avegrade

Number of Observations Read	135
Number of Observations Used	135

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	130	22510.7470	173.1596
Scaled Deviance	130	135.0000	1.0385
Pearson Chi-Square	130	22510.7470	173.1596
Scaled Pearson X2	130	135.0000	1.0385
Log Likelihood		-536.9187	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	277.9284	16.2897	246.0011	309.8557	291.10	<.0001
S_TRatio	1	0.0405	0.0040	0.0327	0.0484	103.01	<.0001
SATPartRate	1	1.1017	0.1092	0.8877	1.3157	101.83	<.0001
S_CounselRatio	1	-0.0414	0.0215	-0.0834	0.0007	3.71	0.0341
DropoutRate	1	-3.0043	0.6474	-4.2732	-1.7355	21.54	<.0001

NOTE: The scale parameter was estimated by maximum likelihood.