Accuracy of API Index and School Base Report Elements
2003 Academic Performance Index

David Rogosa
Stanford University
December 2005
rag AT stat.stanford.edu

LIST OF FIGURES AND TABLES

Accuracy of API Index and School Base Report Elements, 2003 Academic Performance Index
David Rogosa, Stanford University, December 2005

Accuracy of API Index and School Base Report Elements: 2003 Academic Performance Index

David Rogosa
Stanford University
December 2005

This report for 2003 Base API updates and extends the previous accuracy reports on the 1999 API (Rogosa 2002a), year 2000 API (Rogosa 2002b), and year 2001 Base API (Rogosa 2002c). The 2001 Base API was the first to include elements for the California Standards Tests (CSTs), which by the 2003 Base API comprise 80% of the index. The four main sections of this accuracy report present: standard errors of School API 2003 Base, hit rate accuracy for the Statewide Decile Ranks, comparison of accuracy properties with prior year API, and a look at accuracy of subgroup API scores using the socioeconomically disadvantaged subgroup (SD). As in the prior reports, the purpose of studying the statistical properties of the quantities in the School Report is to address the question: How seriously can we regard these numerical values without (over)interpreting these numbers beyond the accuracy they can support?

From the CDE documentation for 2003 Base (CDE, 2004), the listing of changes and added features in the API:

> California Standards Tests (CSTs)
>
> • The California English-Language Arts Standards Test (ELA CST) was included for all grade levels assessed: grades two through eleven, including a writing assessment at grades four and seven.
>
> • The California Mathematics Standards Test (Mathematics CST) was included for all grade levels assessed: grades two through seven, and grades eight through eleven for course-specific tests:
>
> • The California History-Social Science Standards Test (History-Social Science CST) was included for grade ten (world history) and eleven (U.S. history).
>
> • NEW: The California Science Standards Test (Science CST) was included for grades nine through eleven for course-specific tests.
>
> NEW: The California Alternate Performance Assessment (CAPA) in English language arts and mathematics was included for grades two through eleven.
>
> Norm-referenced test (NRT)
>
> The California Achievement Test, Sixth Edition Survey (CAT/6 Survey), was included for all content areas and grade levels assessed: grades two through eleven. The content areas for grades two through eight included reading, language, spelling, and mathematics. The content areas for grades nine through eleven included reading, language, mathematics, and science.
>
> 2003 California High School Exit Examination (CAHSEE)
>
> The CAHSEE, administered in March 2003 (and May for make ups), was included for grade ten. The CAHSEE covers English-language arts and mathematics.
>
> [from CDE 2003 Base Information Guide]

Part I. Standard Errors of School API, Year 2003 Data

The first topic is the standard error (a measure of statistical uncertainty) of the school-level API index. Computation of the standard error for each school is through bootstrap resampling methods. The value for the standard error of the school score is simply the standard deviation for the (1000) API bootstrap replications.

Descriptive statistics for the standard error of the API, s.e.(API), are shown by School Type in Table 1. The upper frame shows s.e.(API) five-number summaries for the 5087 Elementary schools, 1173 Middle Schools, and 845 High Schools included in this analysis. Minimum school size is 100 students with test scores, and therefore schools with the "S" designation (small schools) are not included. (High Schools with ASAM status are considered separately in Table 3). The lower half of Table 1 shows the median standard error at each state decile. Table 2 expands that display by providing a five-number summary of s.e.(API) and median school size for the schools at each state decile. Regardless of school type, schools have a wide range of values for s.e.(API).

INSERT TABLE 1
INSERT TABLE 2

Further display of s.e.(API) is provided by the plots for Elementary, Middle and High Schools in Figure 1. A pair of plots is provided for each School Type: plots of s.e.(API) versus school API and s.e.(API) versus 1/sqrt(school size). (To calibrate those plots note that axis points .1, .05, .025., .02 correspond to n = 100, 400, 1600, 2500.)

INSERT FIGURE 1

A major feature of the s.e.(API) properties is the dependence on the number of students (denote by n) contributing to the school's API index. In California, Middle Schools have about twice the number of API students as Elementary Schools (median school sizes, students included, for 2003 Base API of 382 and 849), and High Schools have about three times the number as Elementary Schools (median school size 1266). Table 1 shows that the ratio of median standard errors for Elementary and Middle Schools is 1.43 (compare to sqrt(2) = 1.41). Furthermore, the plots in Figure 1 of s.e.(API) versus 1/Sqrt(n) for Elementary, Middle and High Schools show the strong dependence of the standard error on the number of students. As the school API score can be expressed as a mean of individual scores, the 1/sqrt(n) dependence of the standard error would be anticipated by any introductory statistics student.

In contrast, from Table 1 the 2003 High Schools appear not to follow the ratio indicated by relative school sizes (proportional to square root of relative sizes), as for example High School median and quartile values of s.e.(API) actually exceed

those for the Middle Schools. The inflation of s.e.(API) for High Schools is due to the inclusion of the High School Exit Exam (CAHSEE) in the 2003 Base. And Figure 1 shows the extreme effect for some High Schools with only a few student scores on the CAHSEE: large High Schools with school s.e.(API) exceeding values of small Elementary Schools (c.f. discussion of Table 3).

Although the dependence of s.e.(API), on the number of students is strong, the plots in Figure 1 also show notable differences for schools of the same size, mainly a result of the additional dependence of s.e.(API) on the school's API score (clearest for Elementary and Middle Schools). The plots of s.e.(API) versus API show a pattern of larger s.e.(API) for API scores in the middle of the distribution, a pattern readers with an introductory statistics course will recognize as characteristic of a proportion score. The display of the median s.e.(API) by state decile in Tables 1 and 2 similarly shows larger values for schools in the middle state deciles for each school type.

Even though different tests are employed for 2003 Base, patterns of s.e.(API) are quite similar to the prior years. Detailed comparisons with 1999-2001 API are shown in Part III of this report.

Additional issues for High schools: HSEE, ASAM
Table 3 provides additional detail on the distribution of s.e.(API) for the California High Schools. The leftmost column shows percentiles of the s.e.(API) distribution depicted in Figure 1 and Tables 1,2. The percentiles of the distribution show the magnitude of the upper tail (in fact the upper half of the distribution is notably inflated compared to past years data.) The two rightmost columns of Table 3 include results for the Alternative Schools Accountability Model (ASAM) schools, which receive APIs but do not receive decile ranks. API scores for school districts and ASAM schools are reported in order to comply with requirements of the federal No Child Left Behind Act of 2001 (NCLB) and are not required to be reported under the California law. The displays in Tables 1 and 2 and Figure 1 do not include the ASAM schools. The bottom part of Table 3 gives descriptive statistics for the 100 ASAM High Schools with at least 100 API student scores; the ASAM schools are smaller and with lower API scores (median is over 200 points below the non-ASAM High Schools). Inclusion of the smaller ASAM schools does make s.e.(API) for the full set of 945 High Schools larger than for the set of 845 PSAA High Schools.

INSERT TABLE 3

Statistical asides:
1. Because of the imputation procedure for missing individual test scores implicit in the calculation of the API (for school or subgroups), the naive standard error estimate, standard deviation of individual scores divided by square root of n, will understate (because of nonindependence) the uncertainty of the API. The bootstrap

resampling methods (or alternatively a sampling theory correction) will produce standard errors that are larger than the elementary estimate whenever some student scores are missing.

2. In forming the bootstrap estimates for standard error, the bootstrap samples could be constructed more carefully by stratifying on the subgroup classification (SD crossed with the 7 ethnic/racial categories) to guarantee that each bootstrap sample had the same subgroup memberships as the school dataset. Doing so does not reduce the bootstrap estimate of standard error noticeably.

3. An even more minor effect is the rounding to an integer in the reported API score. Whether the bootstrap standard error is computed using the replicates as the actual API value or a rounded value does not effect the standard error estimate.

4. Accountability Criteria and Award Programs. It is easy to see that the school-level API scores do contain enough uncertainty that properties of award programs (e.g., meeting growth targets) would be seen as unattractive if the criteria were based solely on the school API. Because of the inclusion of multiple subgroup improvement criteria into accountability criteria, intuitions formed by examining the standard error of the API score are not easily transformed into conclusions about the award programs. As the separate reports on accountability and awards explain, "it's not the size of the standard error, it's how you use it."

Table 1.
Standard Error of School API (bootstrap resampling)

Descriptive Statistics: s.e.(API)

	N	Median	Q1	Q3	Minimum	Maximum
Elem	5087	**9.680**	8.309	11.520	2.892	24.770
Mid	1173	**6.761**	5.825	8.019	3.119	21.730
High	845	**7.060**	5.969	9.789	2.532	39.340

Median s.e.(API) by CARank (state decile)

	Elem		**Middle**		**High**	
CARank	N	Median	N	Median	N	Median
1	507	9.454	116	6.335	84	7.054
2	507	9.694	113	6.534	83	7.794
3	502	9.975	116	7.014	83	7.329
4	517	10.190	122	7.170	87	7.478
5	496	10.510	116	7.428	81	8.012
6	518	10.570	120	7.200	87	6.965
7	508	10.110	116	7.146	81	7.651
8	504	9.719	116	6.548	85	6.500
9	506	8.988	119	6.454	88	6.441
10	522	7.294	119	5.668	86	5.717

Table 2.
Standard Error of 2003 Base API: Quantiles by State Decile Rank

Elementary Schools

CA Rank	Number Schools	Median Size	Quantiles s.e.(API) 0%	25%	50%	75%	100%
1	507	445.0	4.169	8.349	9.454	11.290	22.610
2	507	447.0	5.304	8.436	9.694	11.625	19.740
3	502	418.0	4.974	8.654	9.975	11.992	23.24
4	517	393.0	5.544	8.816	10.190	11.960	24.090
5	496	368.0	5.855	9.095	10.510	12.335	23.810
6	518	350.5	6.124	9.002	10.565	12.200	24.770
7	508	360.0	6.264	8.873	10.105	11.882	22.260
8	504	348.5	5.658	8.541	9.719	11.335	20.580
9	506	362.5	4.220	7.857	8.987	10.670	17.550
10	522	378.5	2.892	6.251	7.295	8.483	15.690

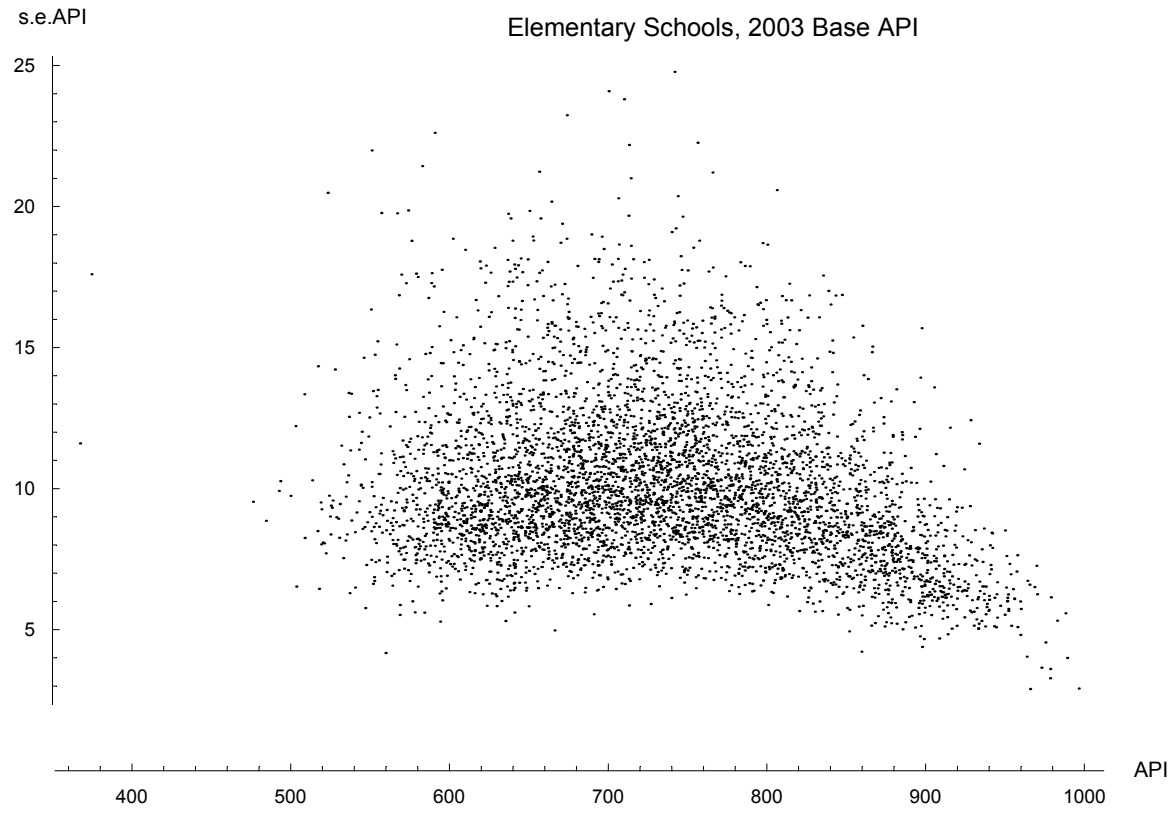
Middle Schools

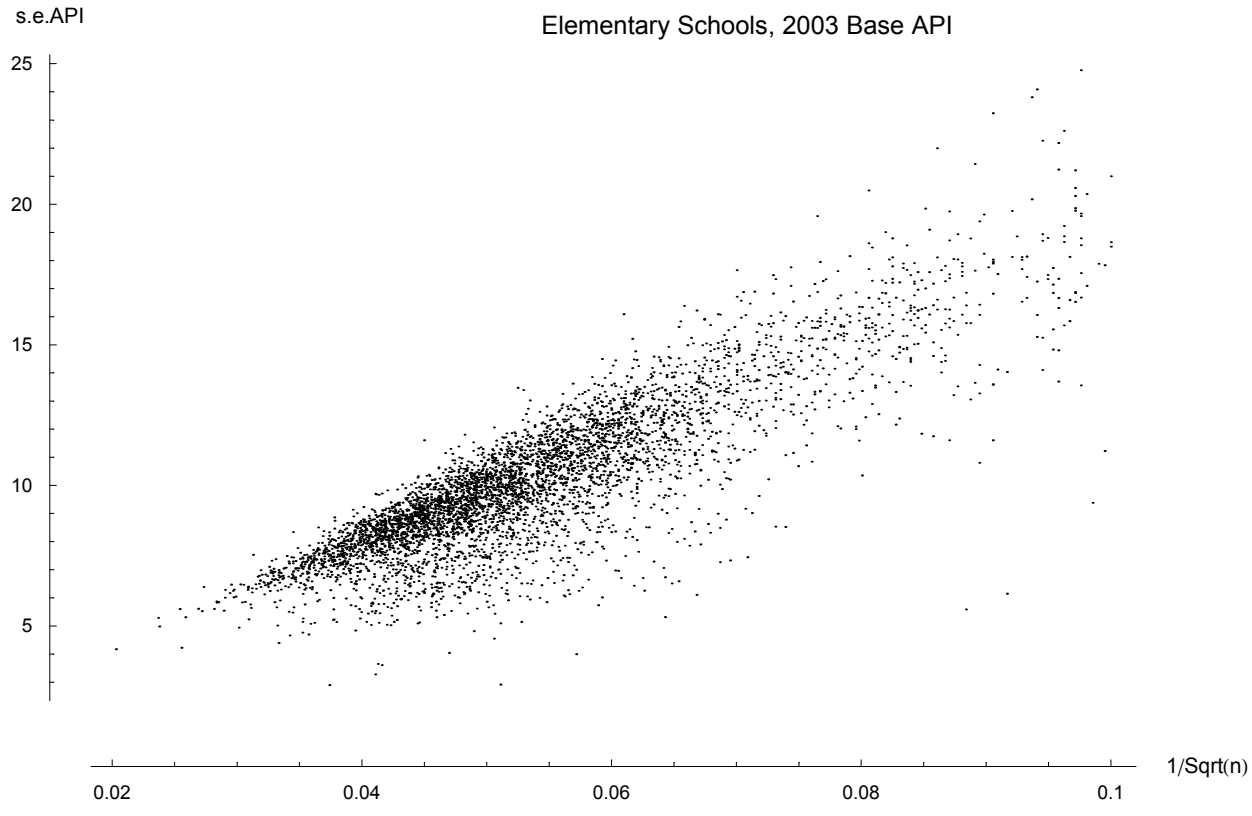
CA Rank	Number Schools	Median Size	Quantiles s.e.(API) 0%	25%	50%	75%	100%
1	116	965.0	3.119	4.488	6.335	7.843	17.740
2	113	980.0	3.534	5.775	6.534	7.613	17.240
3	116	857.0	4.563	6.046	7.013	7.950	18.510
4	122	854.0	5.023	6.159	7.170	8.209	15.800
5	116	788.5	4.999	6.526	7.428	8.690	21.730
6	120	808.0	4.548	6.330	7.200	8.507	20.750
7	116	799.0	4.393	6.124	7.146	8.914	16.880
8	116	888.5	4.123	6.002	6.548	8.009	15.030
9	119	828.0	4.169	5.714	6.454	7.257	17.260
10	119	825.0	3.345	4.882	5.668	6.392	12.340

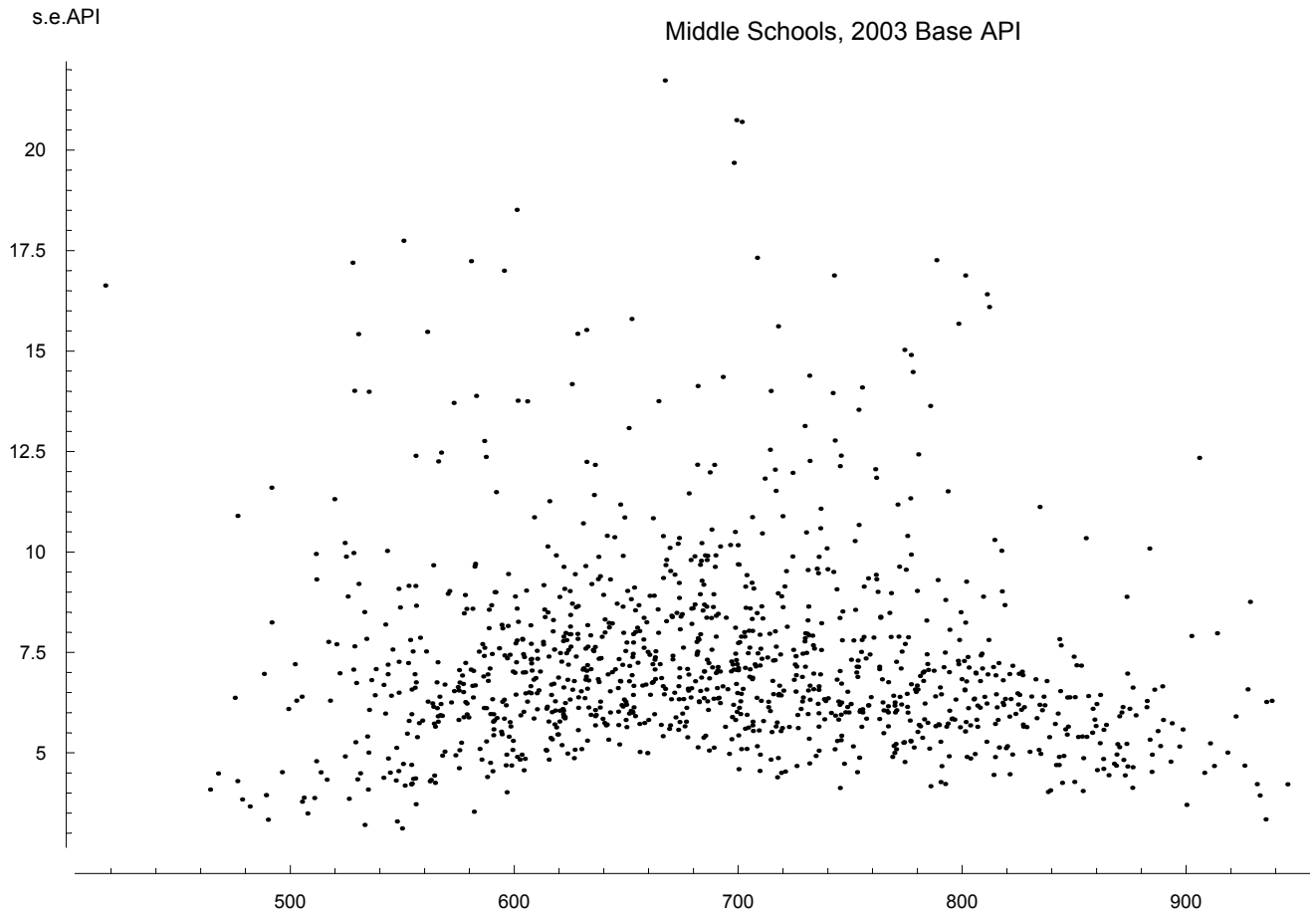
High Schools

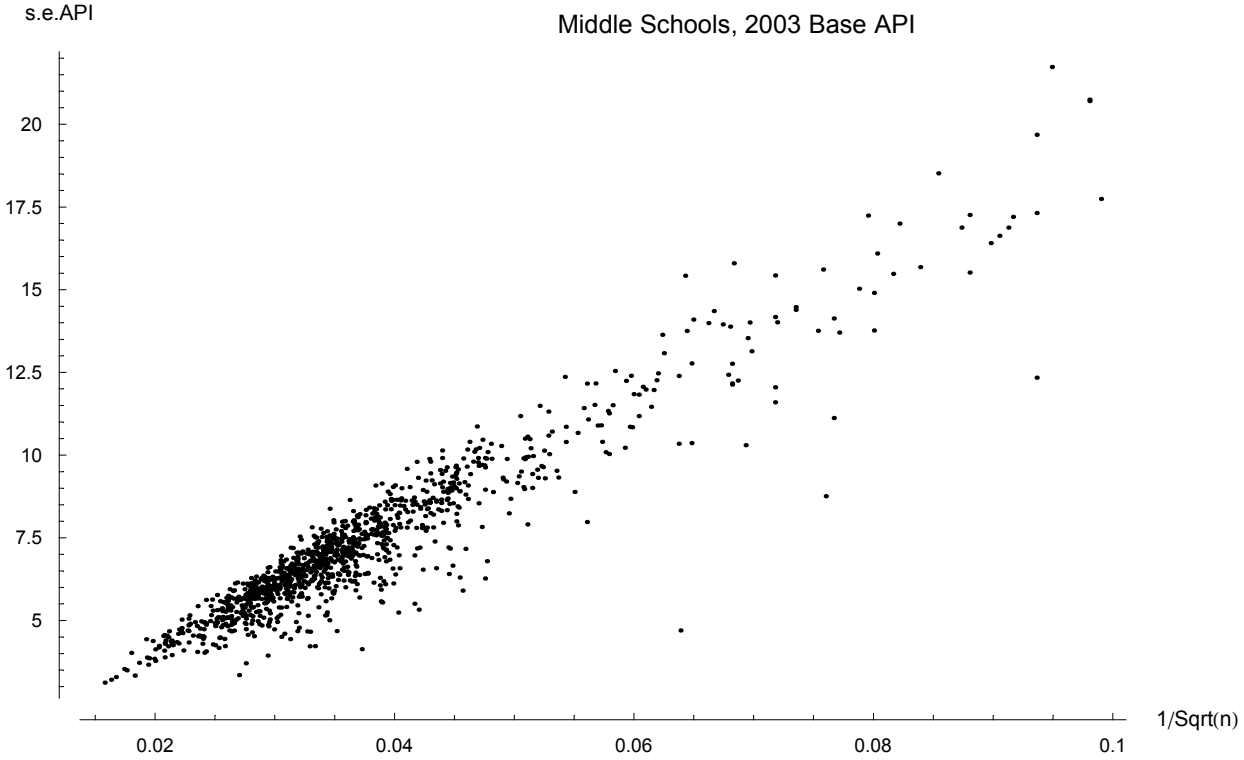
CA Rank	Number Schools	Median Size	Quantiles s.e.(API) 0%	25%	50%	75%	100%
1	84	1242	4.189	5.825	7.054	12.405	23.610
2	83	1214	4.415	6.092	7.794	10.480	35.990
3	83	1317	4.746	6.332	7.329	10.415	27.710
4	87	1293	4.006	6.498	7.478	9.845	31.980
5	81	1262	5.203	6.744	8.012	11.950	27.250
6	87	1358	4.592	6.101	6.965	9.360	33.080
7	81	1237	4.733	6.535	7.651	10.620	25.750
8	85	1270	4.539	5.812	6.500	9.434	39.340
9	88	1305	3.749	5.676	6.441	8.138	37.980
10	86	1196	2.532	5.024	5.717	6.792	16.750

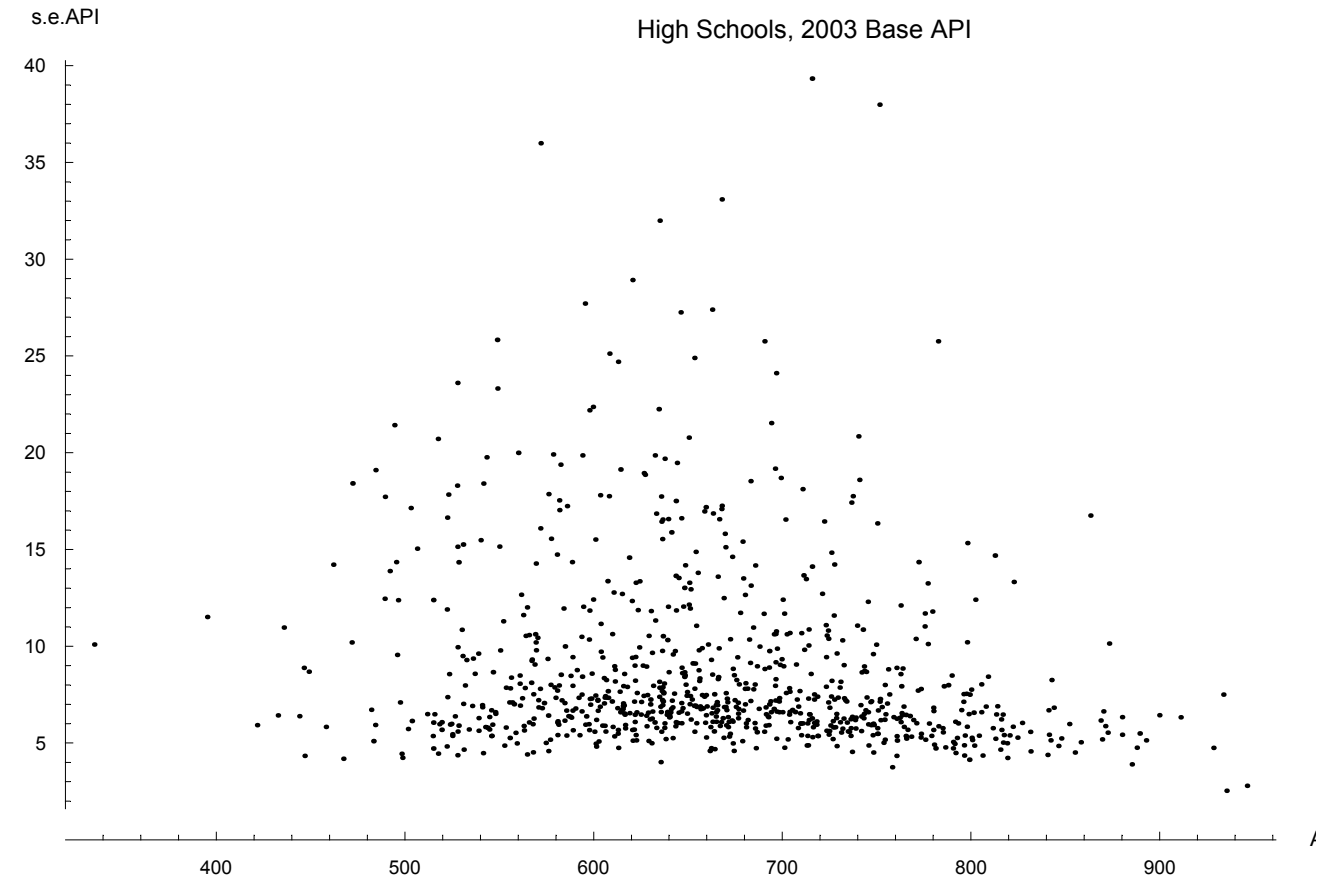
Figure 1 Plots for API Standard Errors






s.e.API
Middle Schools, 2003 Base API
20
17.5
15
12.5
10
7.5
5
500
600
700
800
900


s.e.API
Middle Schools, 2003 Base API
20
17.5
15
12.5
10
7.5
5
0.02
0.04
0.06
0.08
0.1
1/Sqrt(n)


s.e.API
High Schools, 2003 Base API
40
35
30
25
20
15
10
5
400
500
600
700
800
900

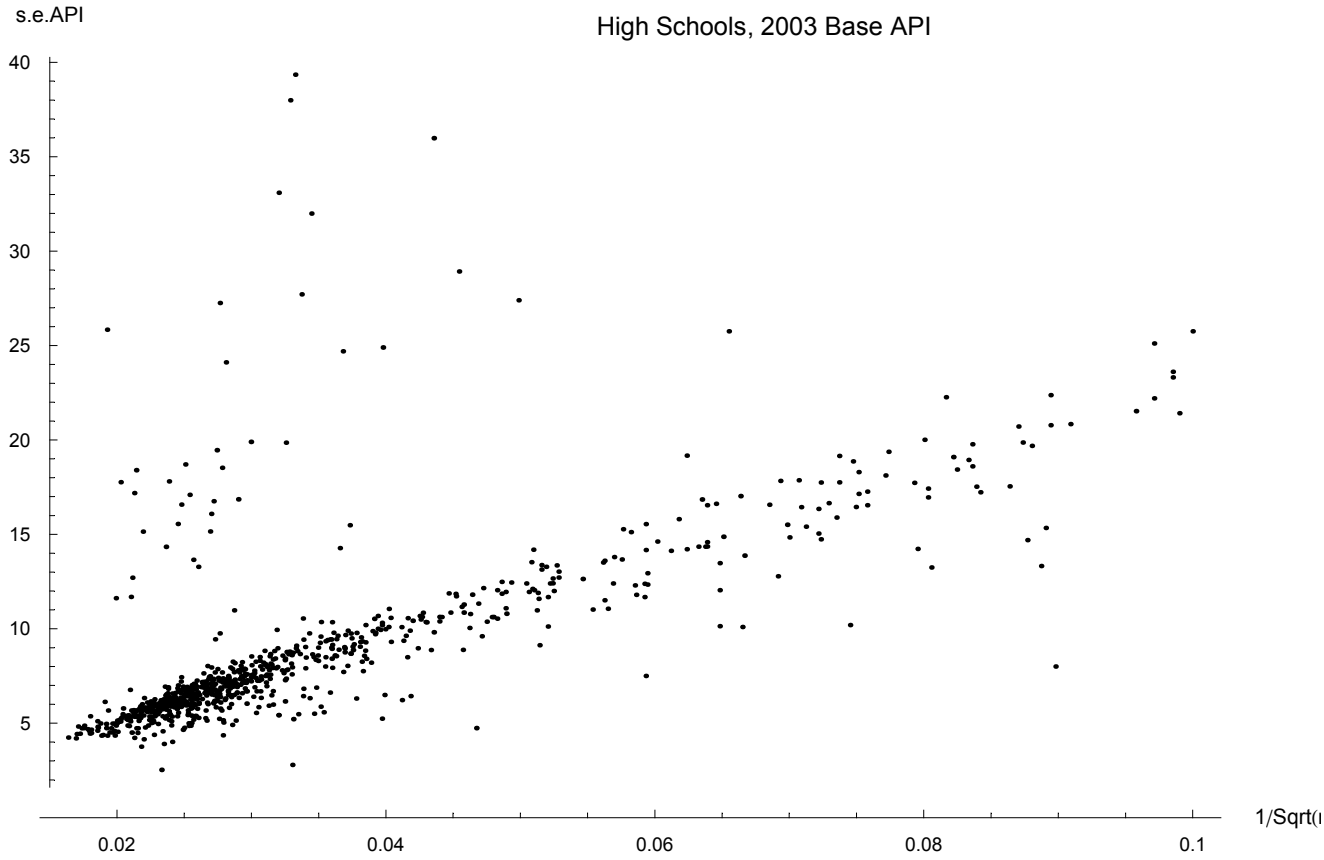

s.e.API
High Schools, 2003 Base API
40
35
30
25
20
15
10
5
0.02
0.04
0.06
0.08
0.1
1/Sqrt(

Table 3.
Distribution of s.e.(API) for High Schools; Inclusion/Exclusion ASAM

	845 non-ASAM HS	945 HS	100 ASAM HS
percentile	se(API)	se(API)	se(API)
0.05	4.830	4.869	8.805
0.10	5.254	5.336	11.420
0.15	5.515	5.600	12.340
0.20	5.780	5.876	13.310
0.25	5.969	6.070	13.730
0.30	6.161	6.336	13.840
0.35	6.391	6.554	14.710
0.40	6.591	6.794	15.290
0.45	6.822	7.066	15.850
0.50	7.060	7.348	16.450
0.55	7.318	7.853	16.810
0.60	7.729	8.489	17.680
0.65	8.236	9.129	18.330
0.70	8.884	10.200	18.650
0.75	9.789	11.820	19.280
0.80	10.850	13.480	20.010
0.85	12.550	15.350	20.610
0.90	15.220	17.530	21.290
0.95	18.420	19.990	22.740
1.00	39.340	39.340	29.410

ASAM HS

	quantiles				
	0%	25%	50%	75%	100%
Number Students	102.0	123.75	155.0	208.75	2002
School API	332.1	395.5	430.9	492.3	673.6

Part II. Hit Rate for Statewide Decile Ranks.

The second main item in the School API report is a state decile (aka statewide rank) for the school API score, computed separately for each school type and using the range 1 (lowest decile) to 10 (top decile). The accuracy of the use of the school API score to determine the reported statewide rank is quantified here by the hit-rate:

decile accuracy hit-rate = 1 – Prob{sampling variability in API score moves the school out of its assigned decile}.

The hit-rates are estimated from a bootstrap resampling (e.g. 1000 replications for the school API), by simply tabulating the proportion of the API bootstrap replications that fall into the reported decile. Other accuracy calculations for the decile rank could be presented (one alternative might be Bayesian-style Prob{'true' decile = k | reported decile = j}), but the hit-rate from resampling appears to be the most straightforward quantity.

The decile accuracy hit-rate quantifies an answer to the question: What is the effect of the statistical variability (wobble) in the school API score on the reported statewide rank? This hit-rate provides one useful summary for the accuracy of the decile rank for each school. The median hit-rate for Elementary schools is .775, with 70% of Elementary Schools having hit rates above 2/3. Quartiles and median values for the hit-rates are broken down by state decile in Table 4. High Schools have noticeably lower decile accuracy than the (smaller) Middle Schools because of the inflation of s.e.(API) from the CAHSEE inclusion.

INSERT TABLE 4

Of course, a school with an API score near a decile boundary will have a much larger probability of statistical variability moving the API score into a different decile. That motivates plotting hit-rate versus position in the decile, as shown in Figure 2 and Figure 3. Figure 2 shows the hit rate for 1999 schools in statewide deciles 2 through 9; Figure 3 gives separate plots for each state decile. Schools in the middle of a decile have high hit-rates, except for the smallest schools. The hit rate is dependent on the statistical uncertainty in the school API (one measure being the standard error of API) and the width of the decile. Therefore smaller schools near the middle of a decile will have smaller hit rates than larger schools. The hit rate results do indicate that schools smaller than the minimum size of 100 would have decile ranks with limited accuracy.

INSERT FIGURE 2
INSERT FIGURE 3

Table 4
Decile Accuracy Hit Rates by School type and Decile

decile accuracy hit-rate =
1 - Prob{sampling variability in API score moves the school score out of its assigned decile}.

Elementary Schools Hit-rate

Decile	Median	Lower Quartile	Upper quartile
1	0.994	0.888	1.
2	0.773	0.647	0.877
3	0.738	0.616	0.846
4	0.685	0.588	0.791
5	0.684	0.587	0.773
6	0.676	0.579	0.765
7	0.722	0.612	0.831
8	0.795	0.659	0.894
9	0.863	0.715	0.96
10	1.	0.933	1.

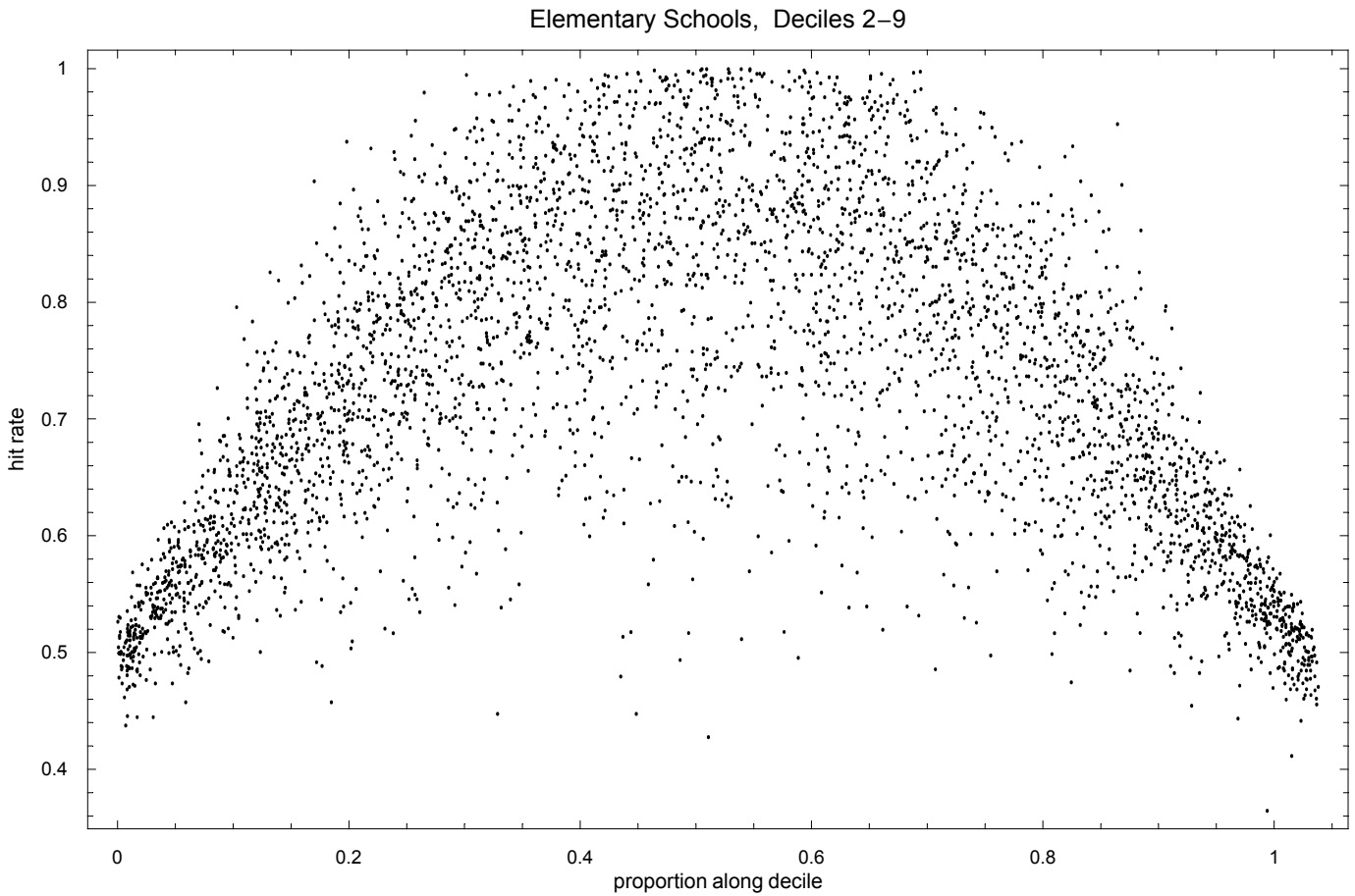
Middle Schools Hit-rate

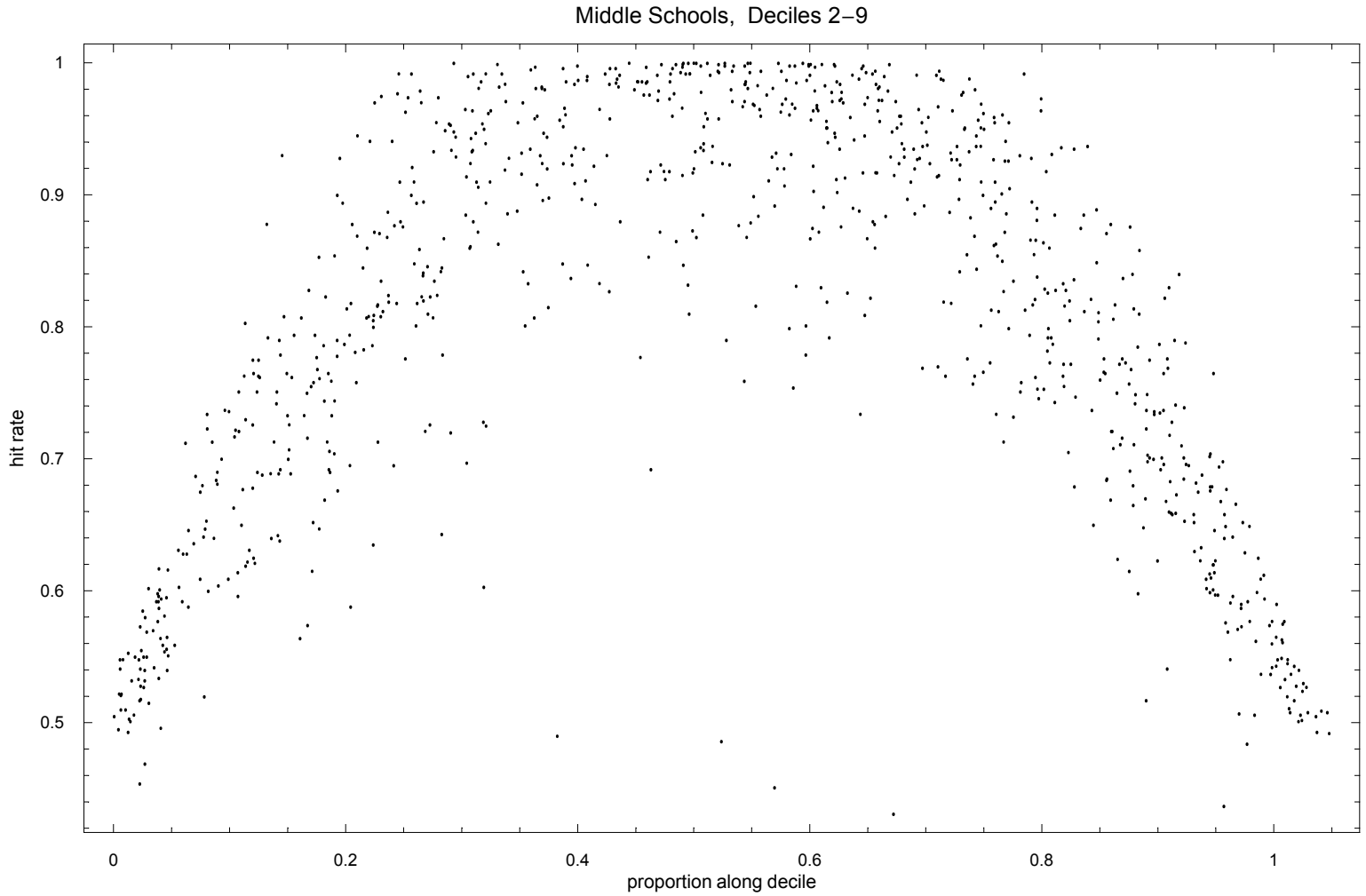
Decile	Median	Lower Quartile	Upper quartile
1	1.	0.968	1.
2	0.912	0.737	0.969
3	0.806	0.668	0.89
4	0.802	0.69	0.915
5	0.803	0.643	0.9
6	0.758	0.609	0.879
7	0.876	0.761	0.958
8	0.878	0.698	0.96
9	0.953	0.765	0.995
10	1.	0.999	1.

High Schools Hit-rate

Decile	Median	Lower Quartile	Upper quartile
1	1.	0.902	1.
2	0.91	0.662	0.968
3	0.772	0.616	0.933
4	0.634	0.541	0.842
5	0.626	0.545	0.778
6	0.758	0.577	0.867
7	0.803	0.671	0.916
8	0.817	0.623	0.927
9	0.911	0.744	0.989
10	1.	0.98	1.

Figure 2. Plots of API State Decile Hit Rates




Middle Schools, Deciles 2−9
hit rate
1
0.9
0.8
0.7
0.6
0.5
0
0.2
0.4
0.6
0.8
1
proportion along decile

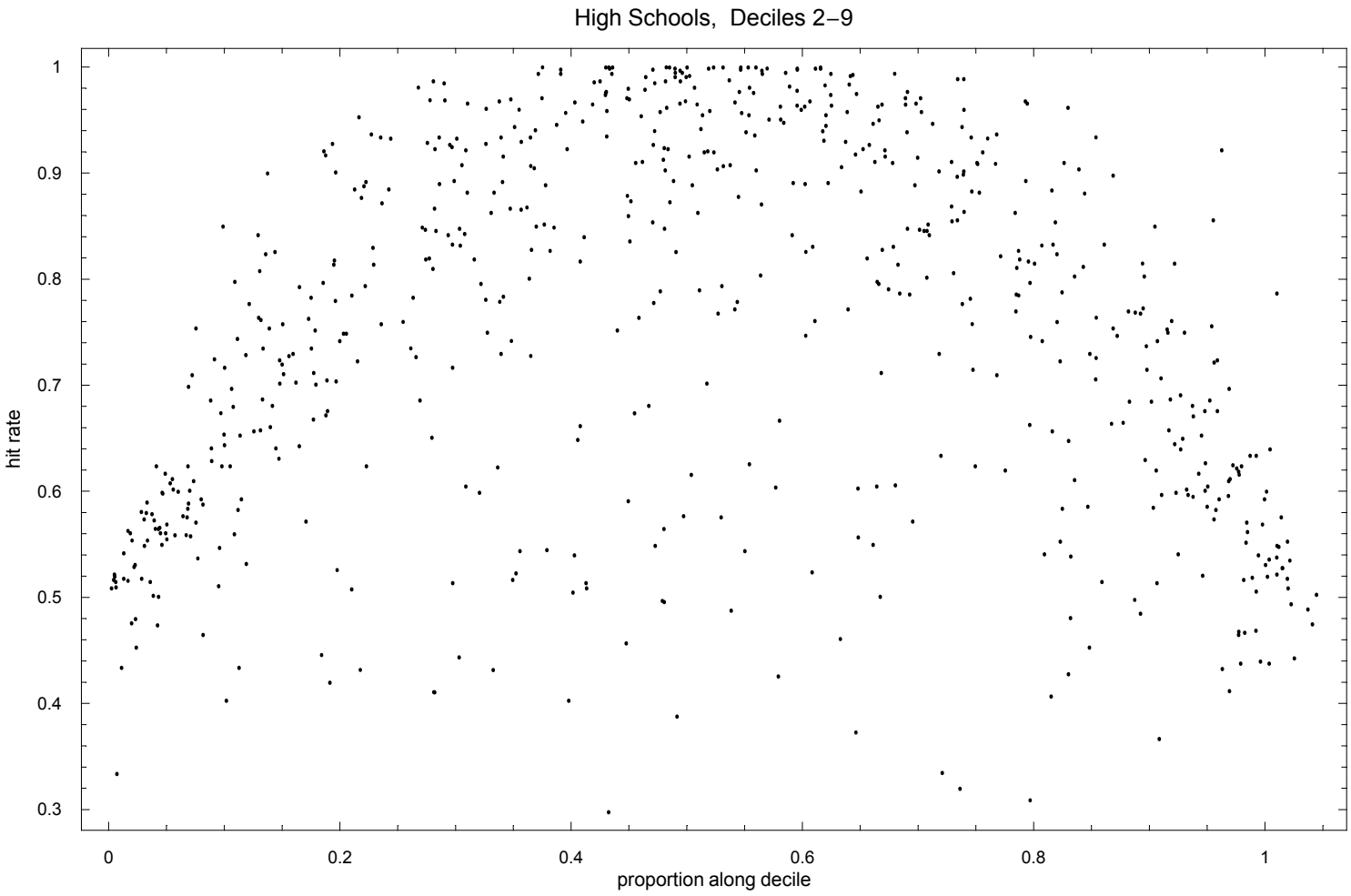
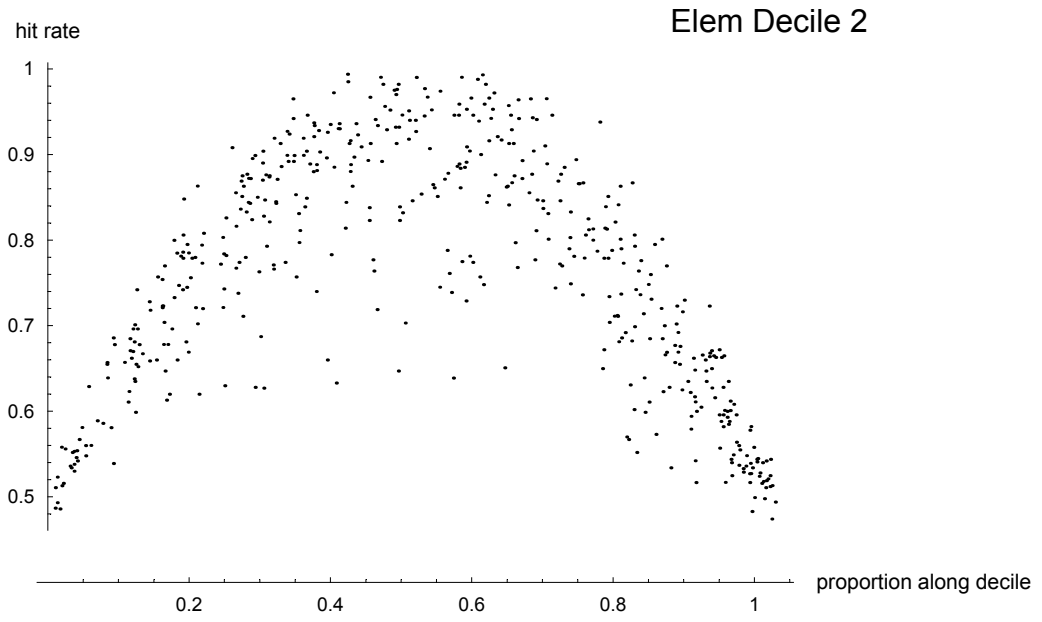
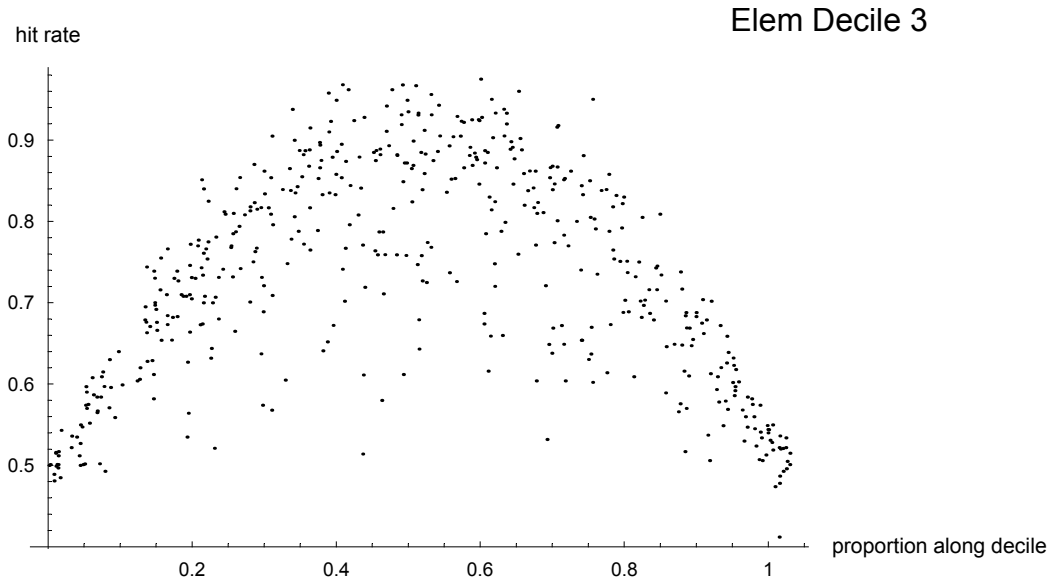
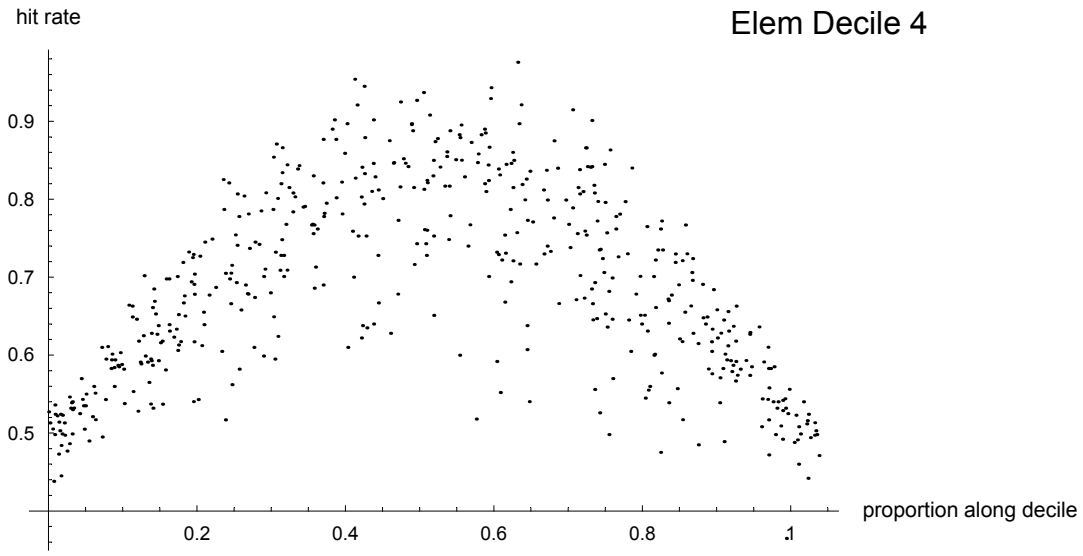
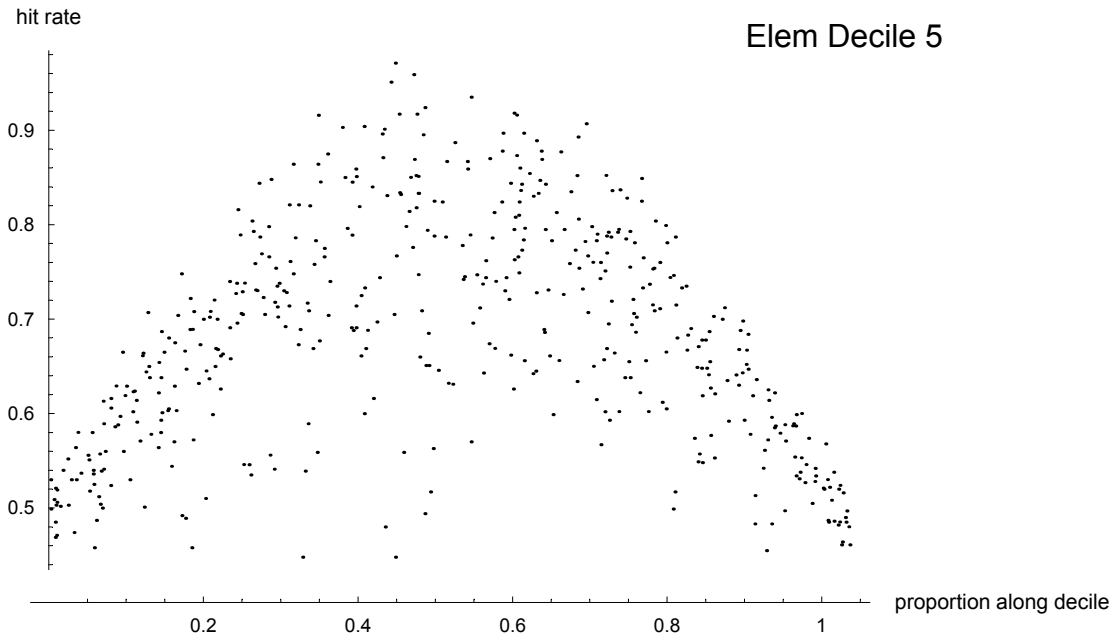

High Schools, Deciles 2−9
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
0.2
0.4
0.6
0.8
1
proportion along decile

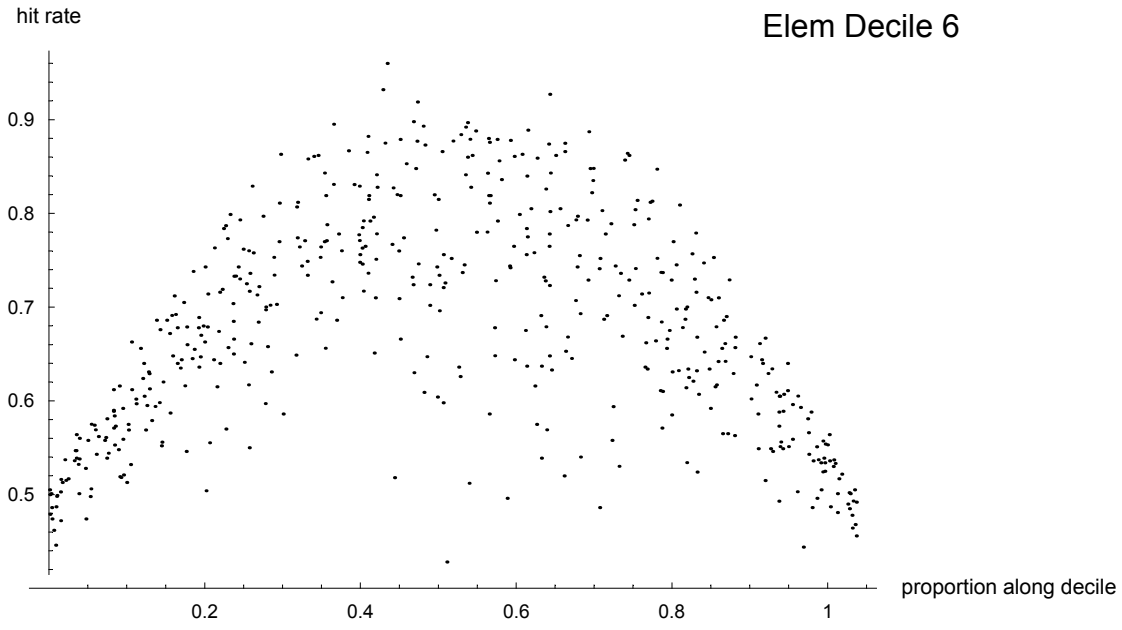
Figure 3 Plots of Decile Hit Rates by Decile

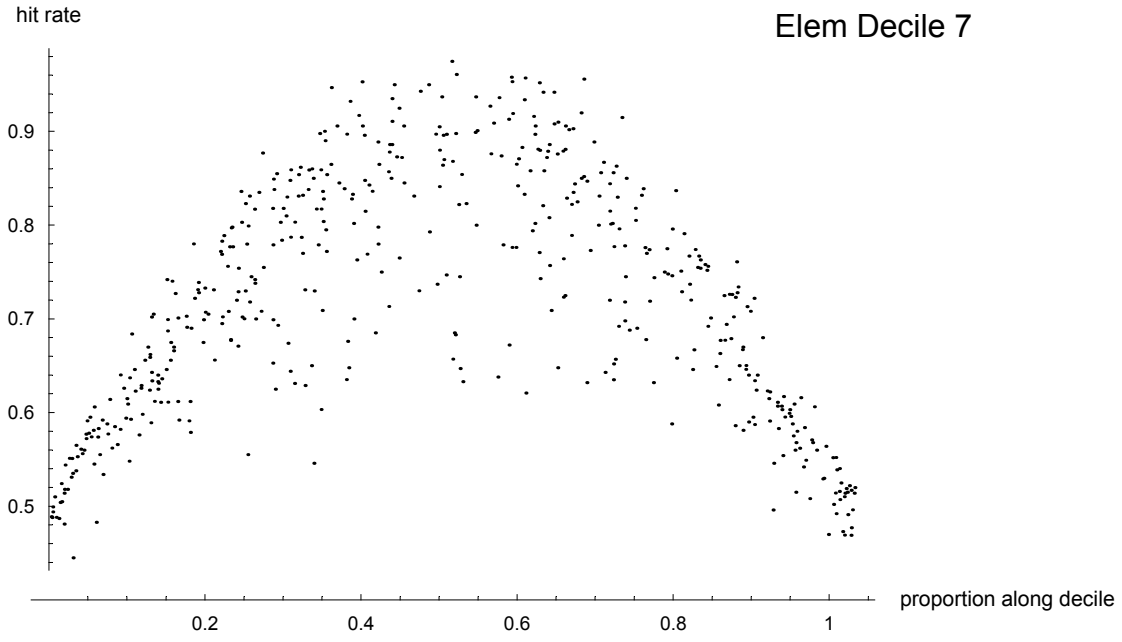


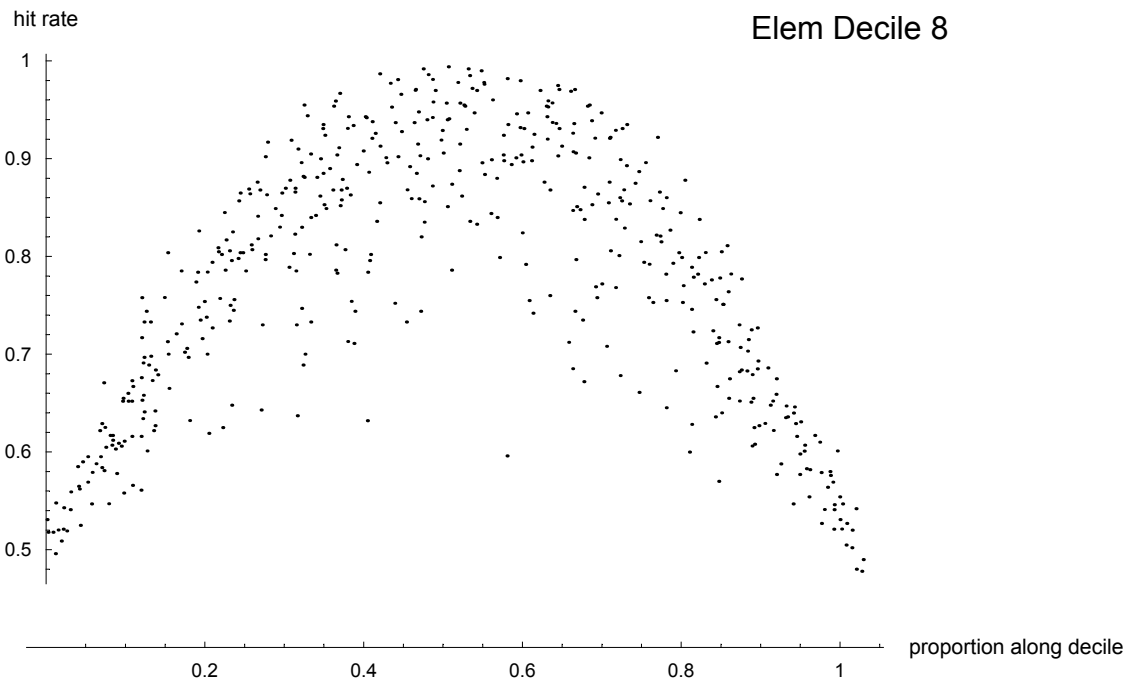


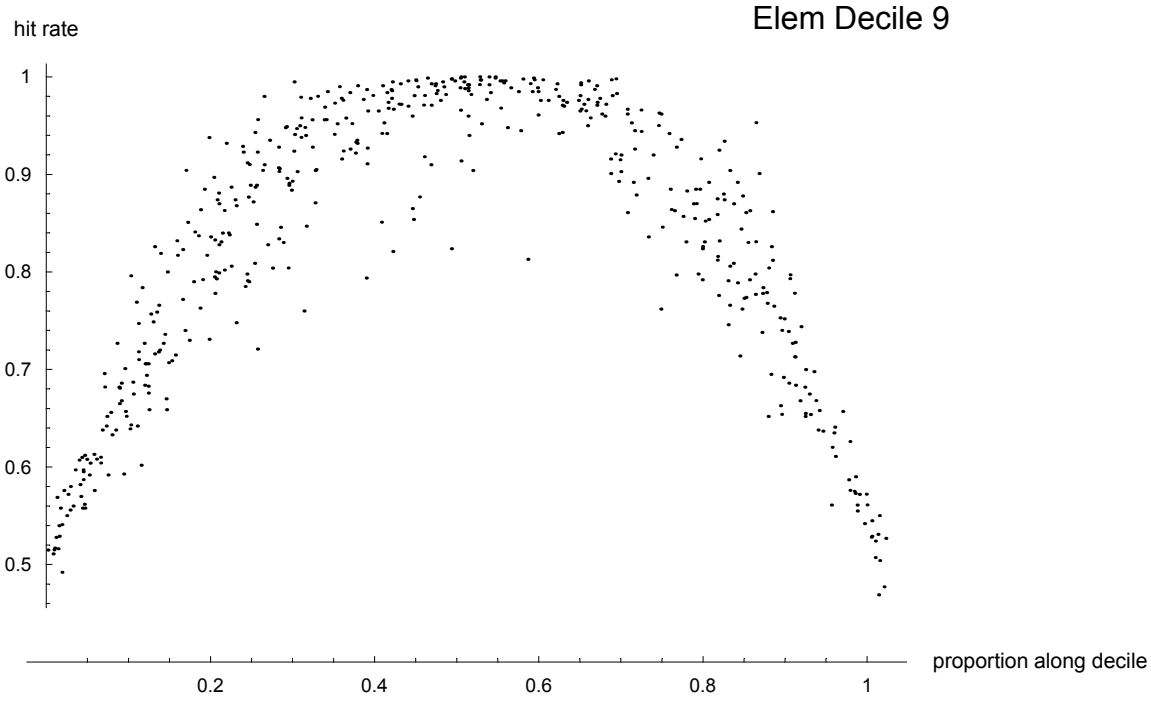

hit rate
Elem Decile 4
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

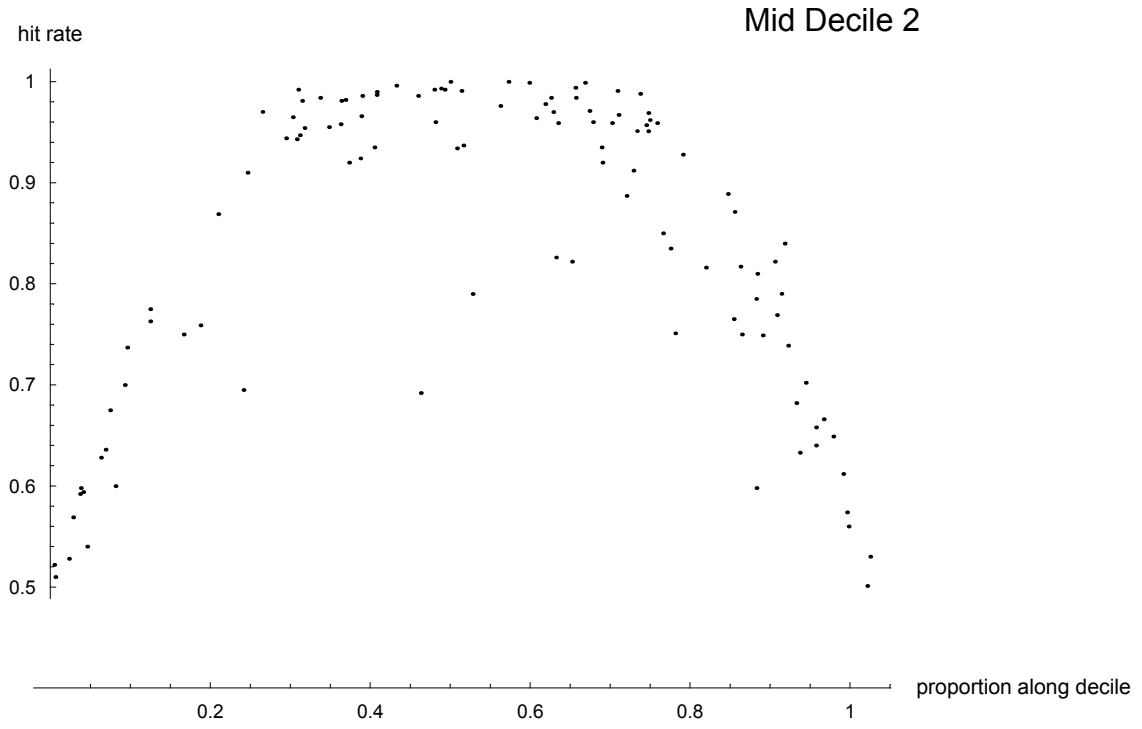

hit rate
Elem Decile 5
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

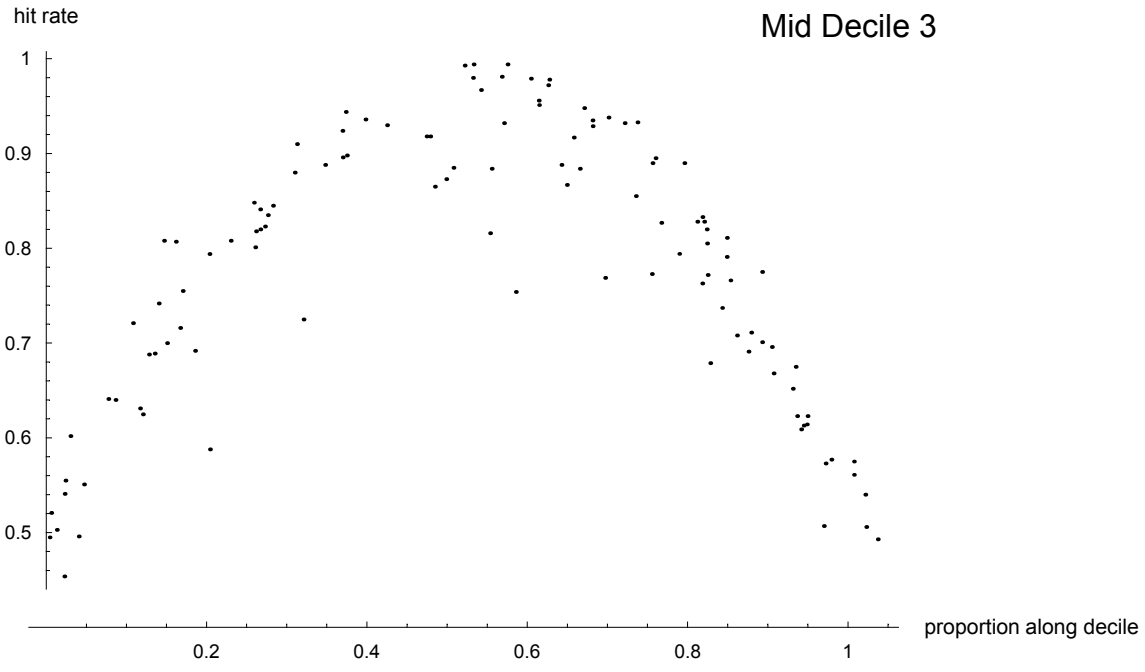

Elem Decile 6
hit rate
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

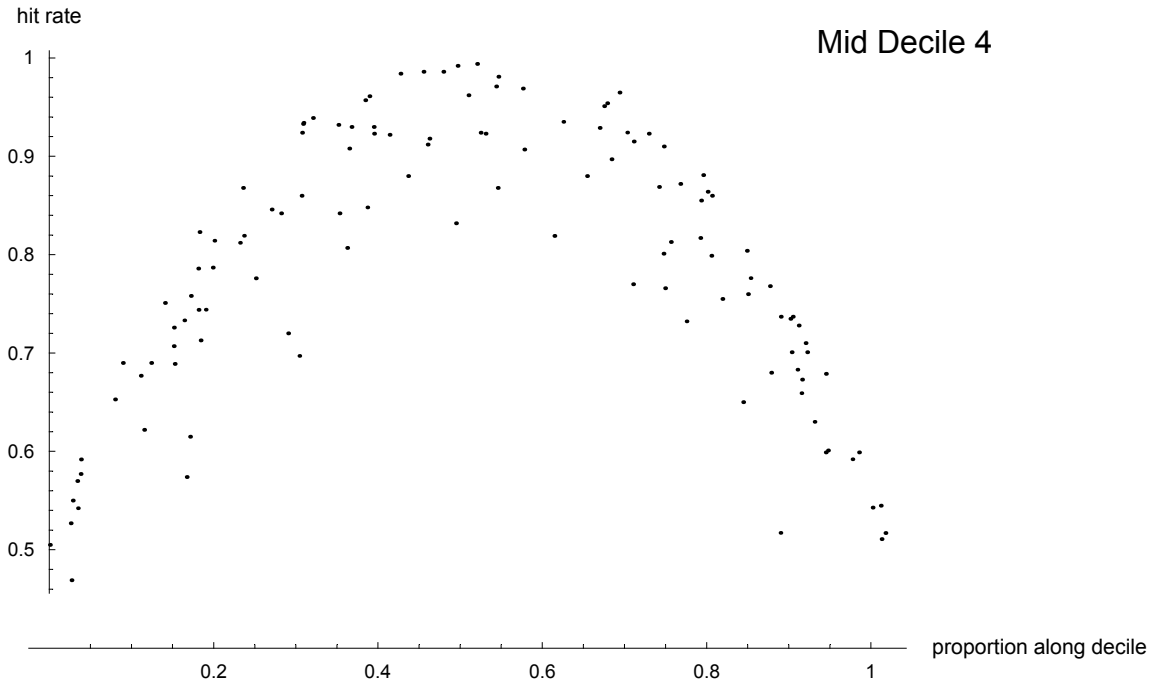

Elem Decile 7
hit rate
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

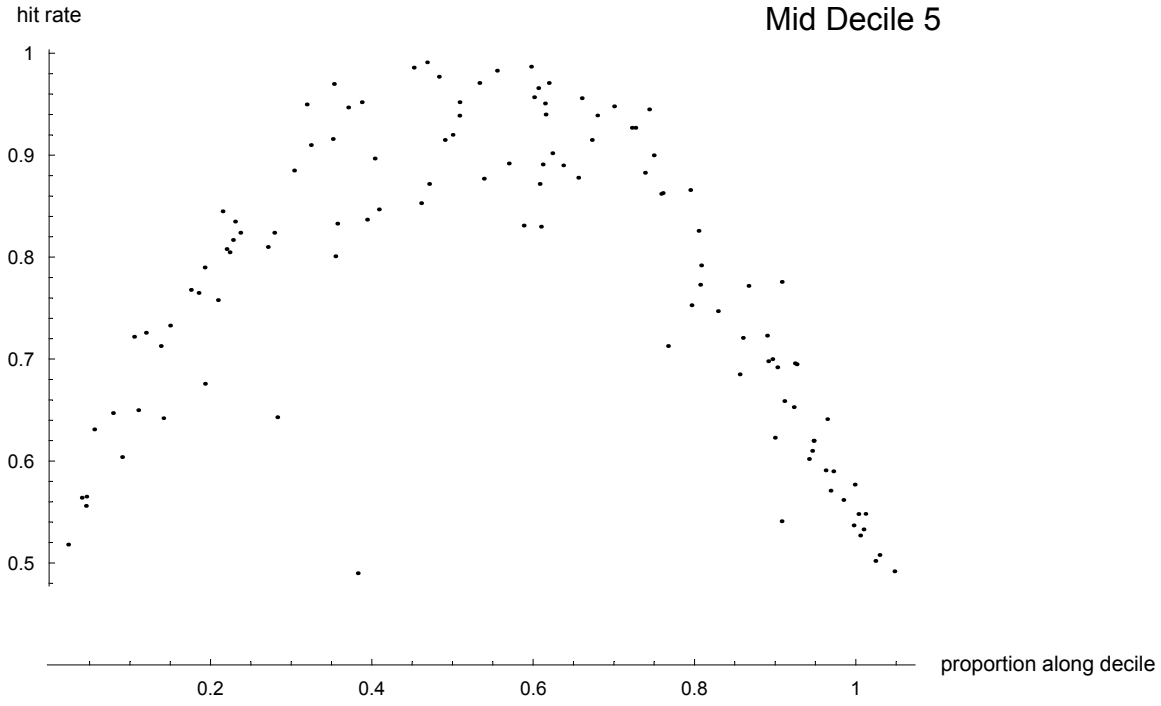

Elem Decile 8
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

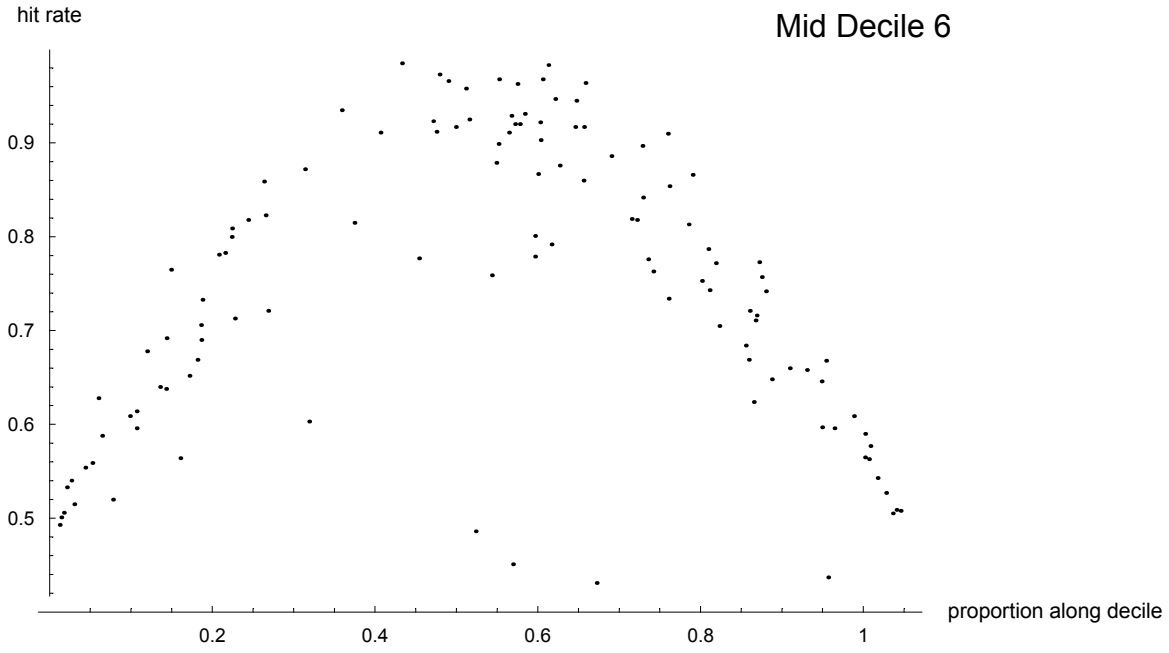

Elem Decile 9
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

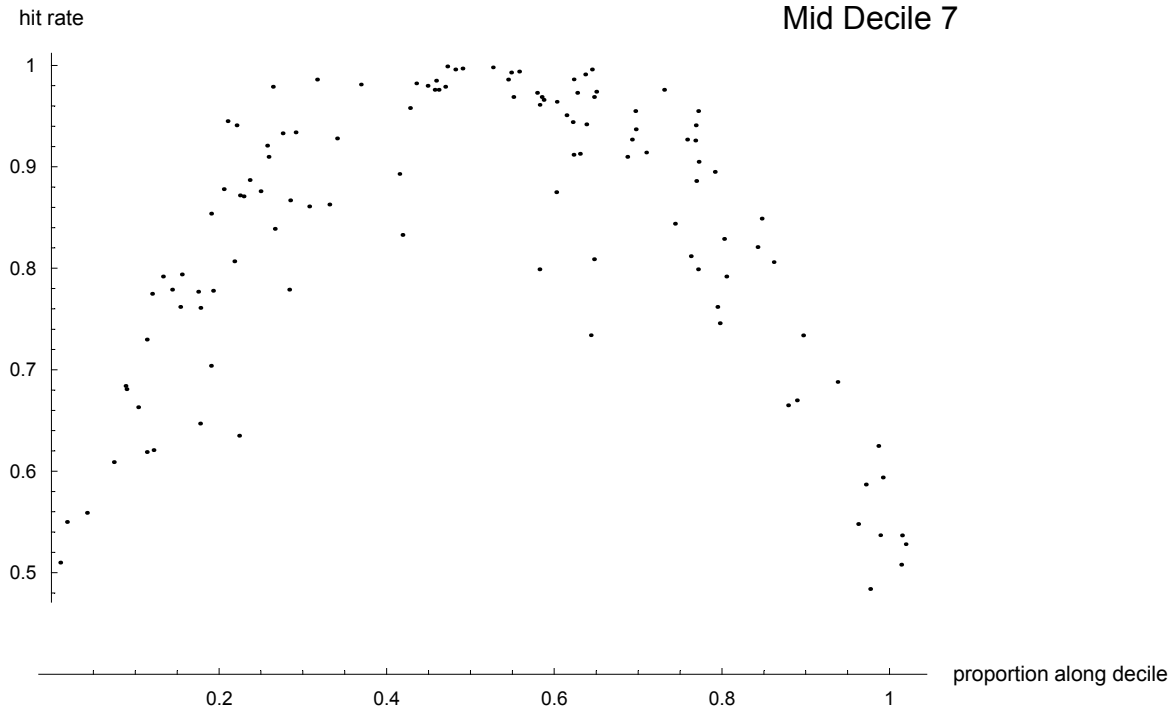

Mid Decile 2
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

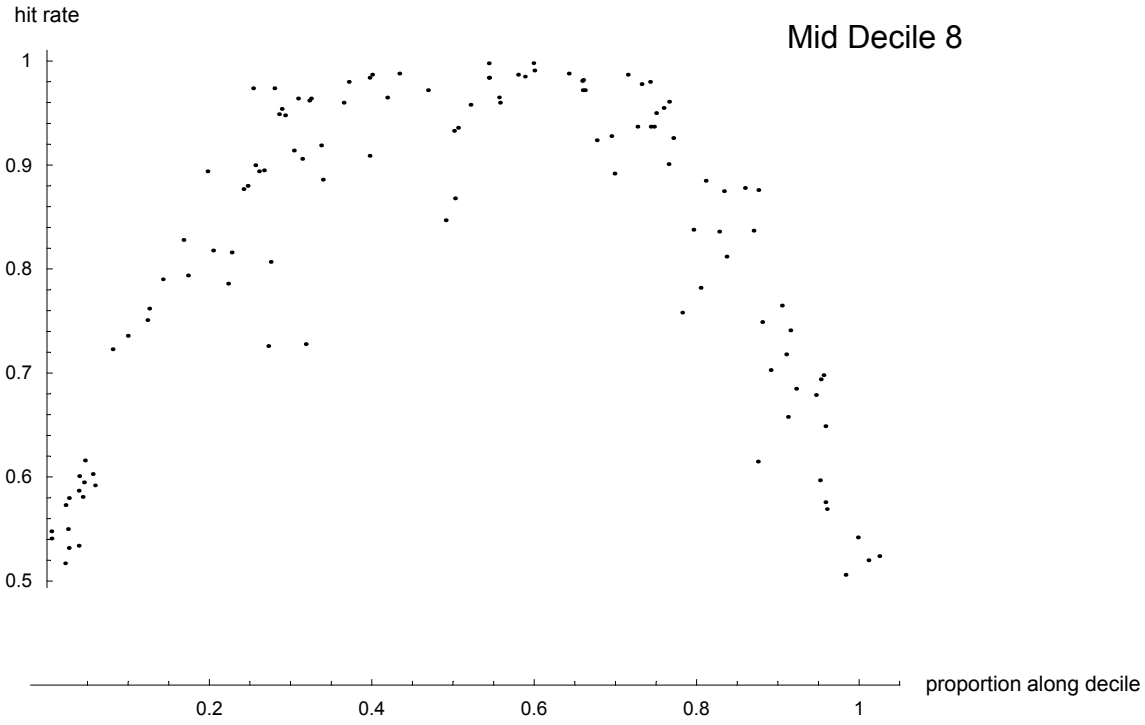

Mid Decile 3
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

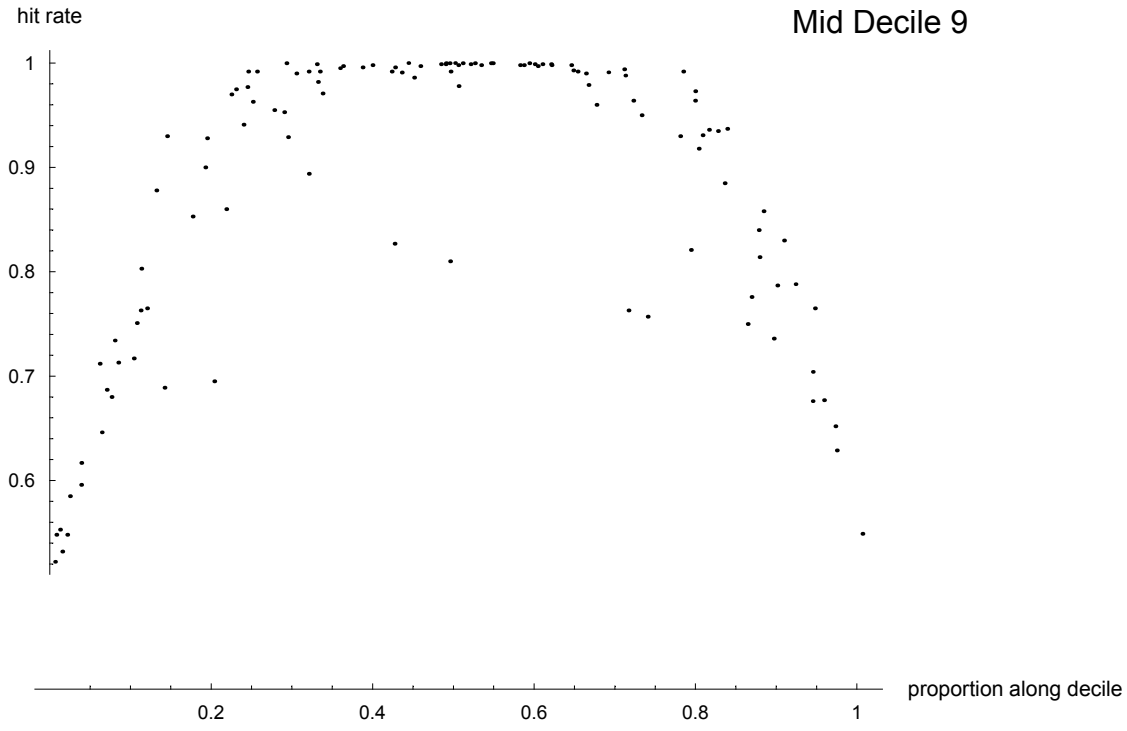

Mid Decile 4
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

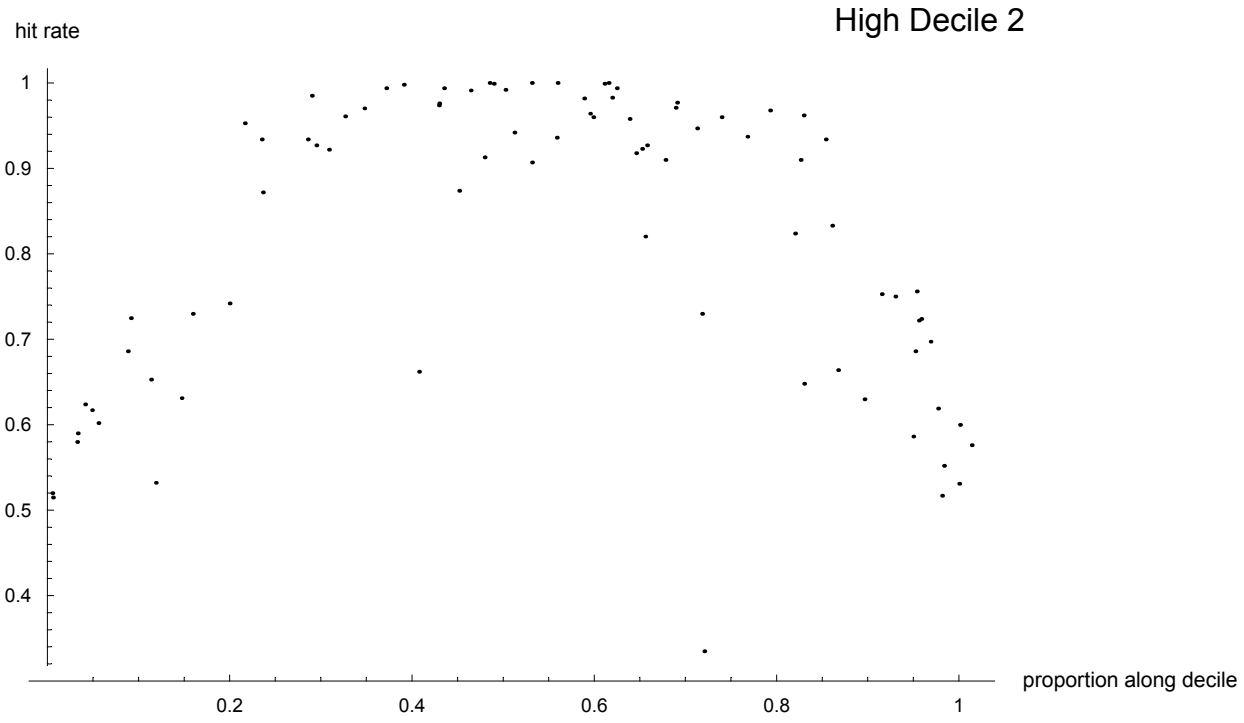

Mid Decile 5
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

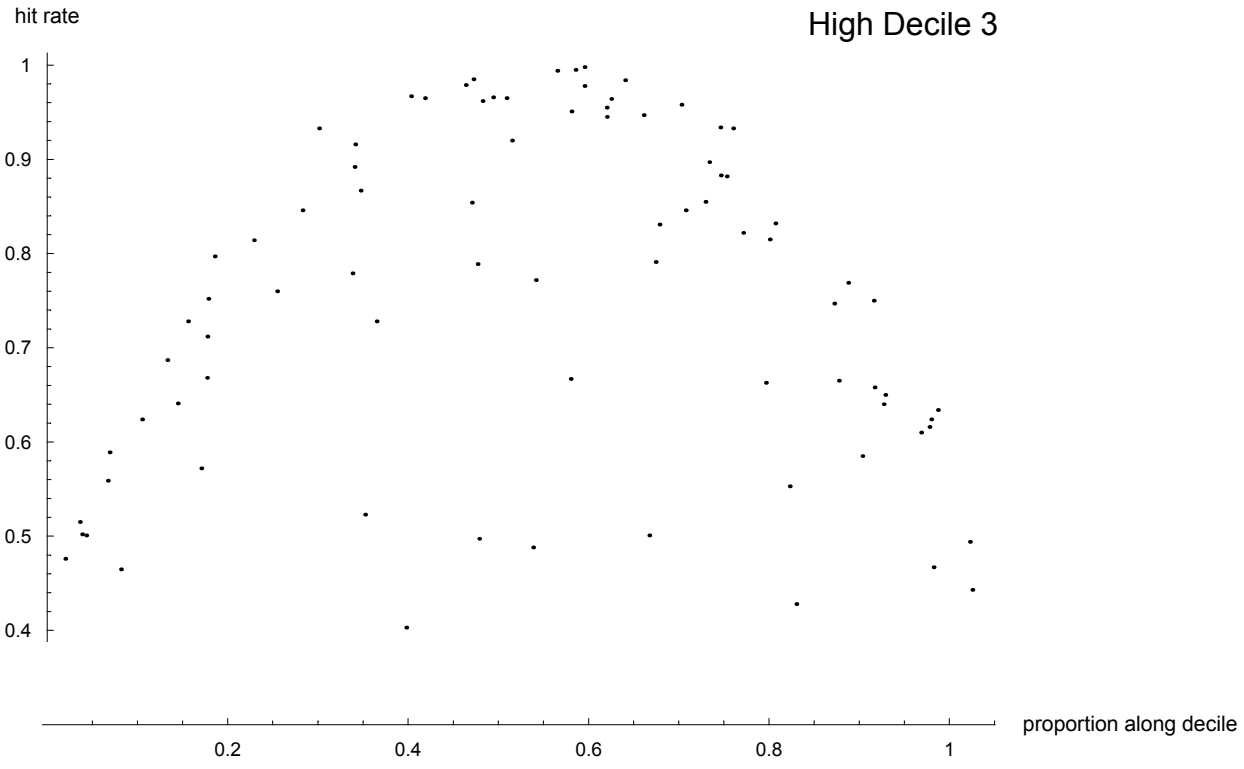

Mid Decile 6
hit rate
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

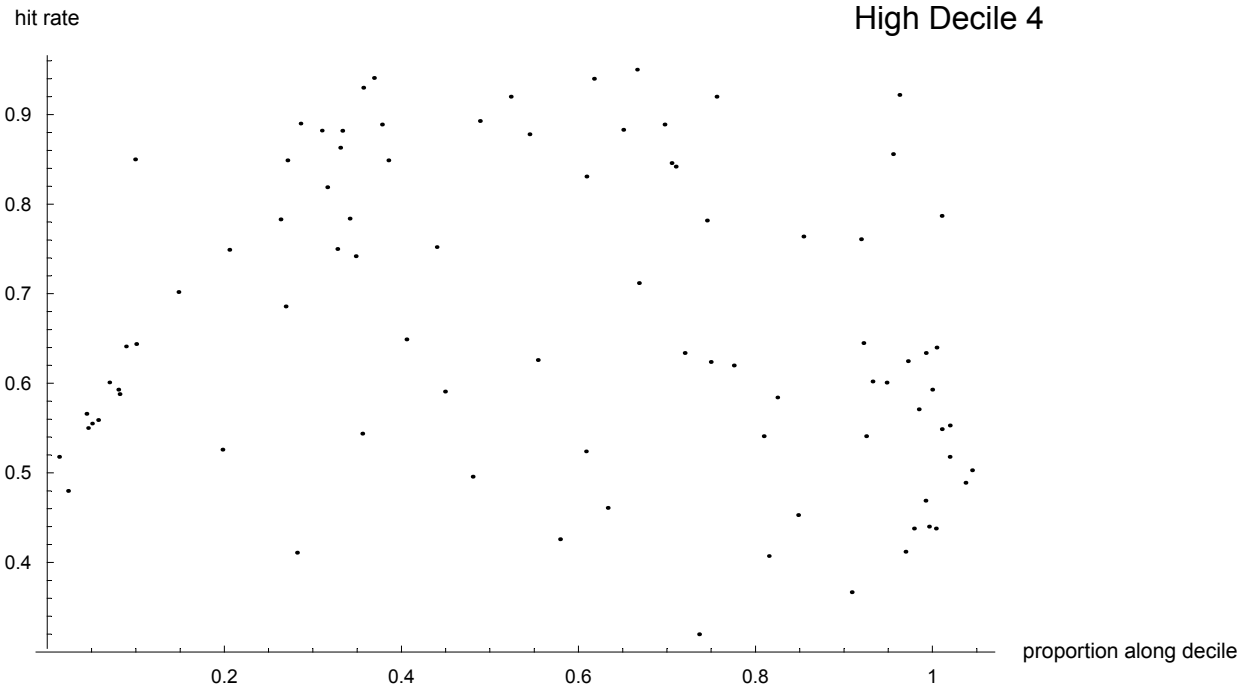

Mid Decile 7
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

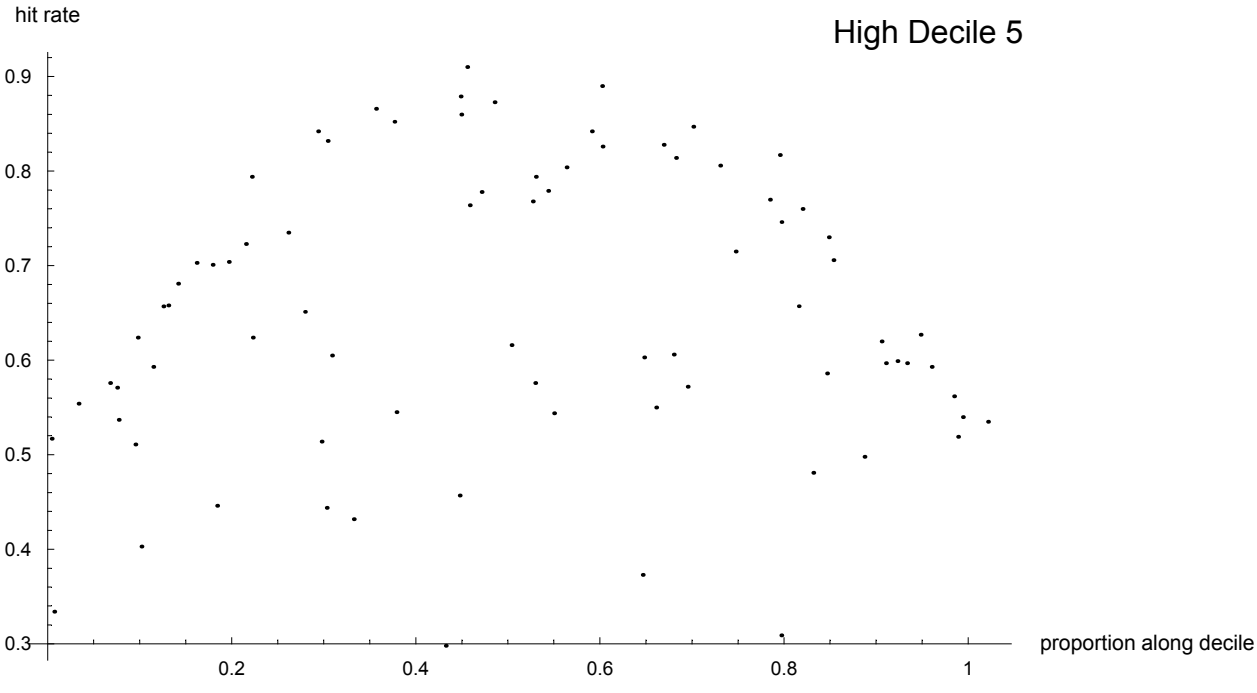

hit rate
Mid Decile 8
1
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

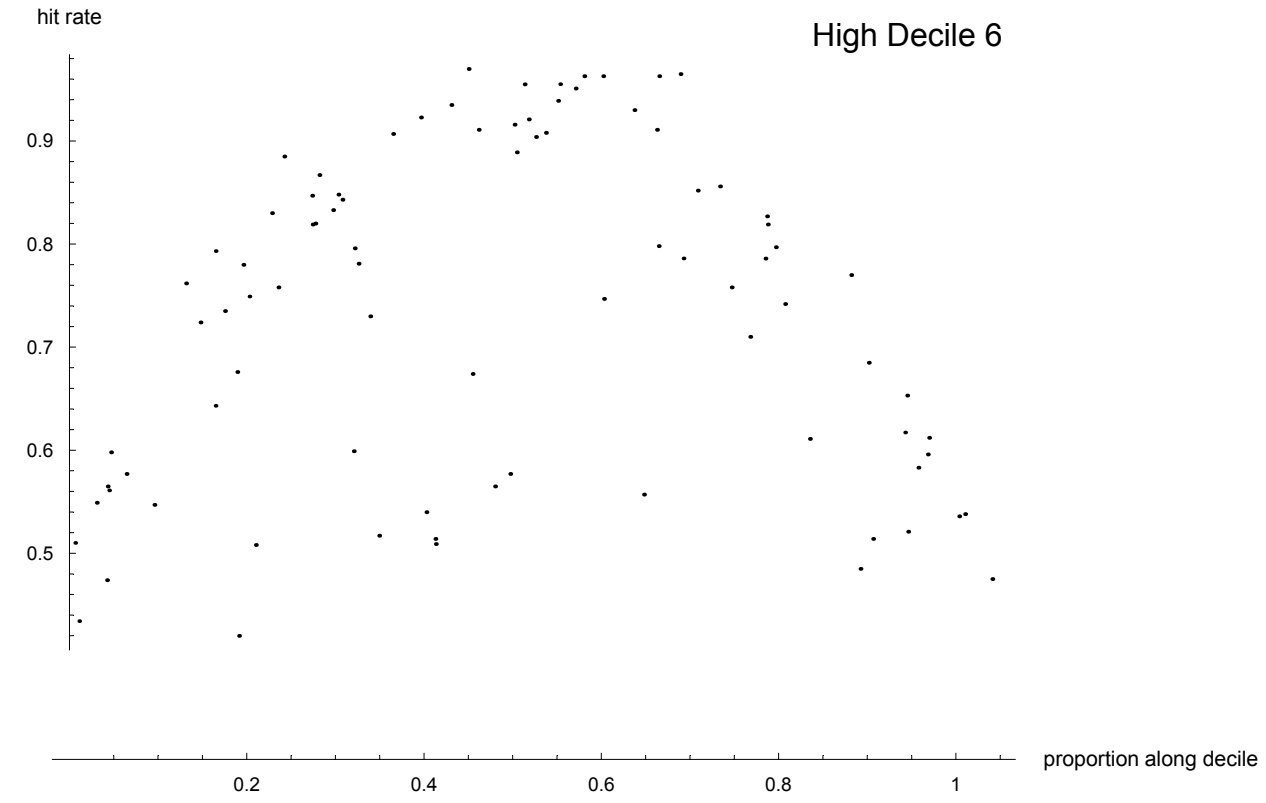

hit rate
Mid Decile 9
1
0.9
0.8
0.7
0.6
0.2
0.4
0.6
0.8
1
proportion along decile

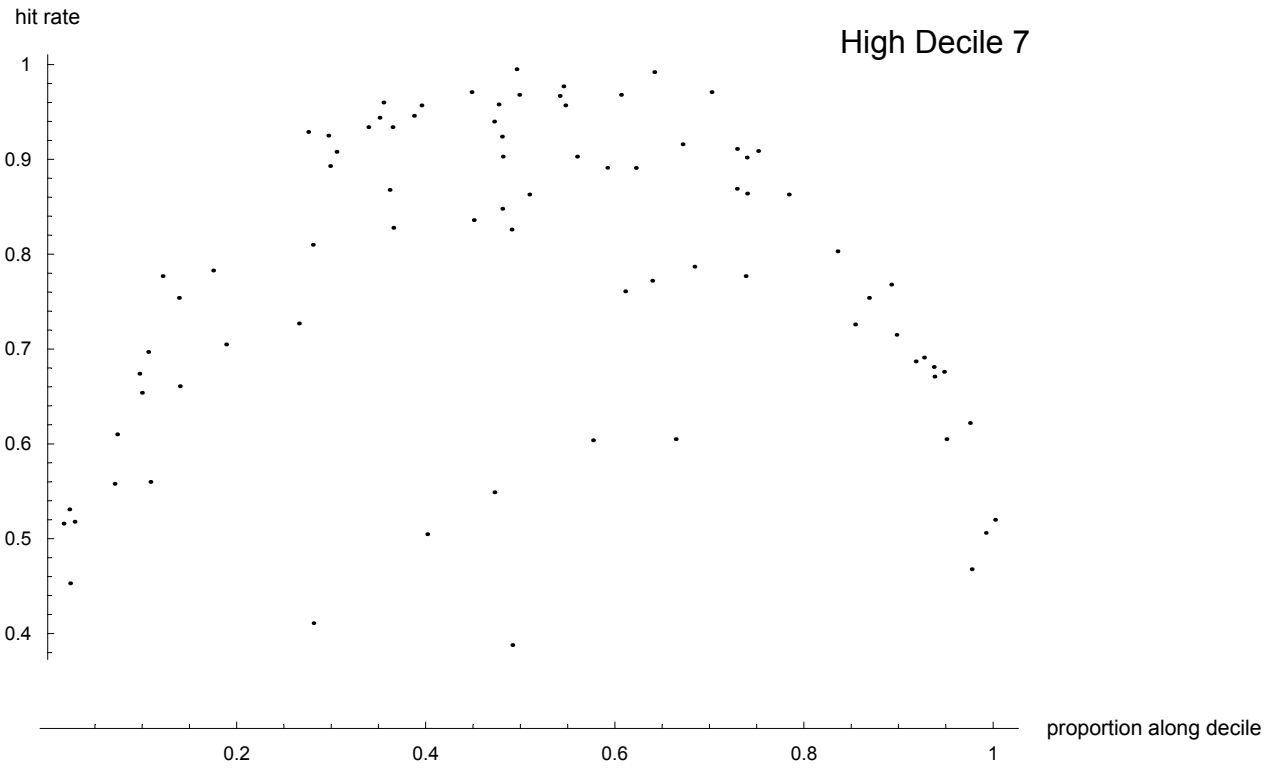

High Decile 2
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.2
0.4
0.6
0.8
1
proportion along decile

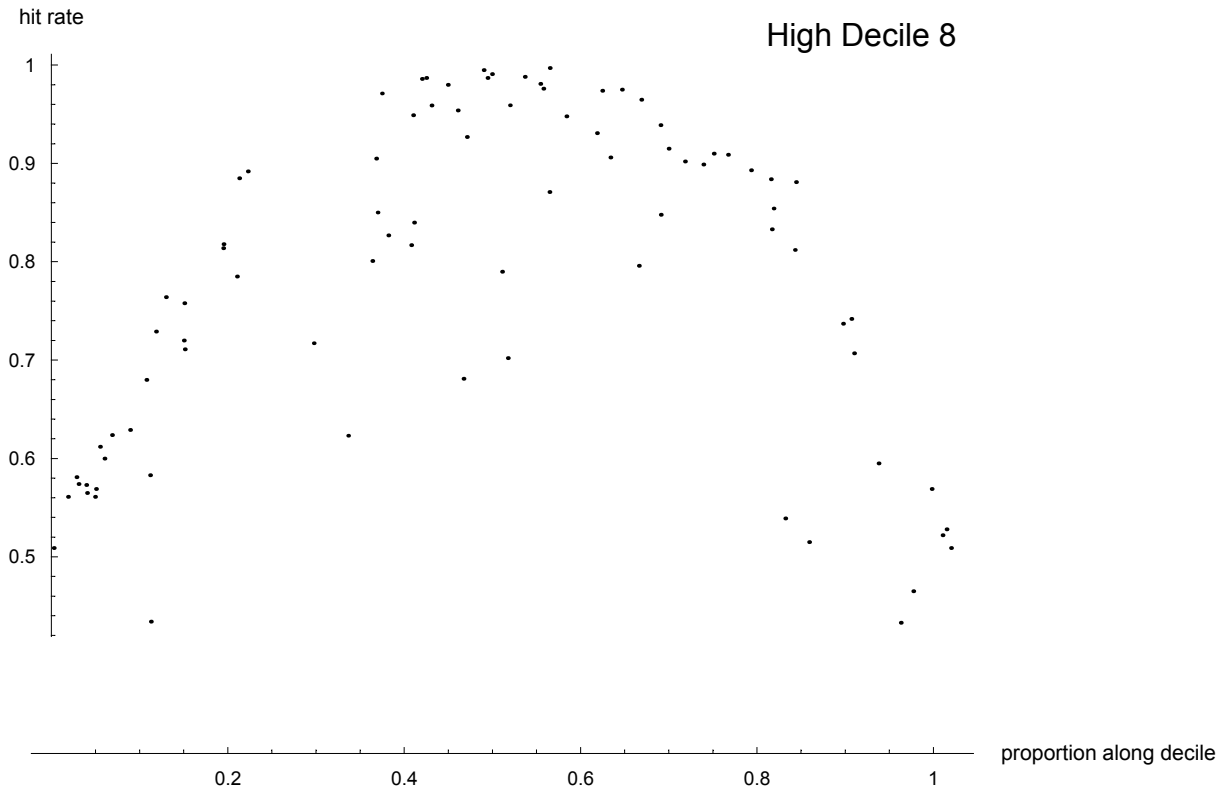

High Decile 3
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.2
0.4
0.6
0.8
1
proportion along decile

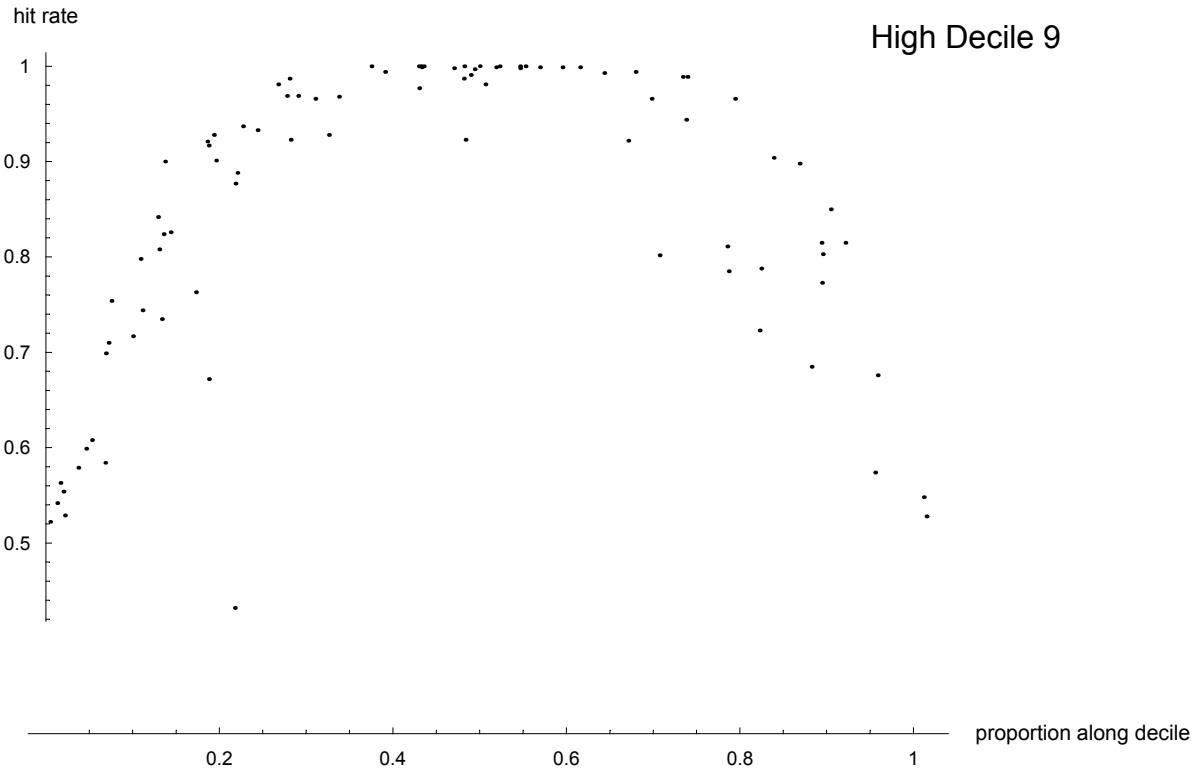

hit rate
High Decile 4
0.9
0.8
0.7
0.6
0.5
0.4
0.2
0.4
0.6
0.8
1
proportion along decile


hit rate
High Decile 5
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.4
0.6
0.8
1
proportion along decile


hit rate
High Decile 6
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile


hit rate
High Decile 7
1
0.9
0.8
0.7
0.6
0.5
0.4
0.2
0.4
0.6
0.8
1
proportion along decile


High Decile 8
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile


High Decile 9
hit rate
1
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
1
proportion along decile

Part III: Comparisons with Prior Years

As the composition of the API has changed from in 1999 being solely based on the Stanford 9 norm-referenced tests to in 2003 being predominately composed of the CST, comparison of 2003 statistical properties with prior years is of interest. In this section, reliability coefficients for 2003 are compared with those previously calculated in for the 1999 API, and school standard errors for 2003 Base are compared with those previously reported for the year 2000 and year 2001 Base API. Bottom line is: the more the index composition changes, the more the properties of the API stay the same.

A. API Reliability Coefficient Calculations

In the accuracy report for the 1999 API (Rogosa 2002a) the section "Derived Values Based on seAPI: Reliability Coefficients for California API Scores " describes the calculation of API (school score) reliability coefficient. Readers familiar with educational testing are conditioned to speak of accuracy or "quality of measurement" in terms of a reliability coefficient. The approach to the reliability calculations is a rough educational testing analogy where n (number of students) serves the role of test length and, as in IRT situations, the error variance in the score also depends on the score level. Shown in Figure 4 are fits to the plots of s.e.(API) from Figure 1 using a simple quadratic for the fit of standard error on API score and a straight-line for standard error on 1/Sqrt[n]. (More sophisticated fits using smoothers won't change the gist of the results).

INSERT FIGURE 4

These fits then allow calculation of standard errors and reliability coefficients for a population of schools of a specified size. For Elementary Schools, the sizes used in 1999 n = 150, 350, 500 represent the 4th, 42nd, and 75th percentiles of the school size (API students) distribution for 2003. For High Schools the sizes used in 1999, n = 500, 1000, 1500 represent the 17th, 35th, and 64th percentiles of the school size (API students) distribution in 2003.

The reliability coefficient can be expressed in a number of equivalent forms:

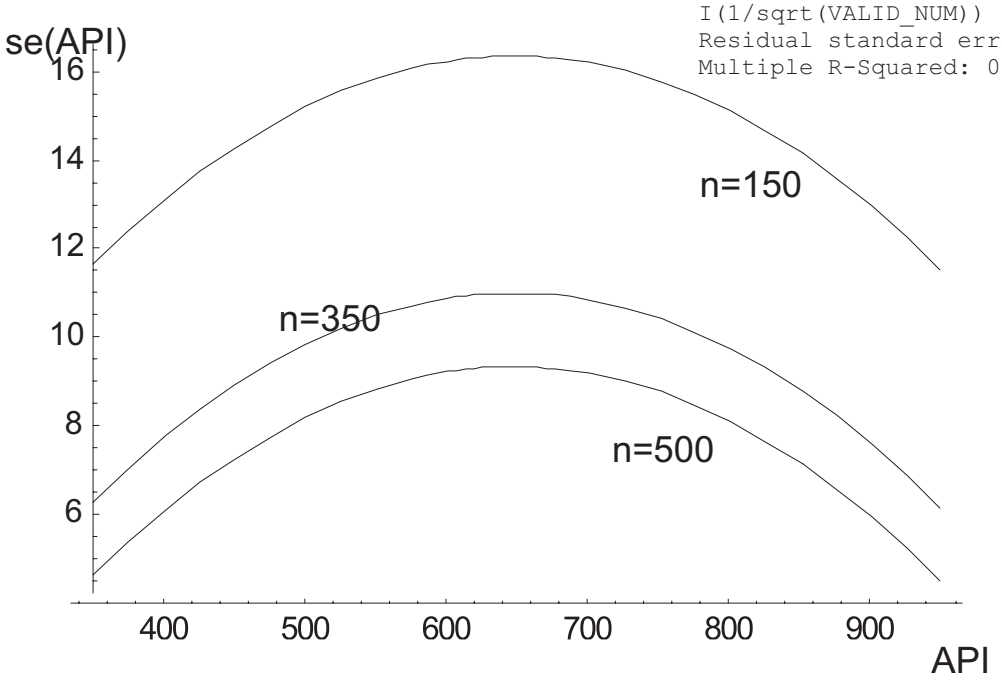
reliability =
(observed variance – average error variance)/observed variance

The average error variance for a specified n is computed by integrating (averaging) the error variance functions displayed in Figure 4 over a "true score" distribution taken as Gaussian with observed score mean, and variance computed as observed score variance minus overall average error variance. The observed variance is the

Figure 4. Error Variances for API 2003 Base Reliability Calculations

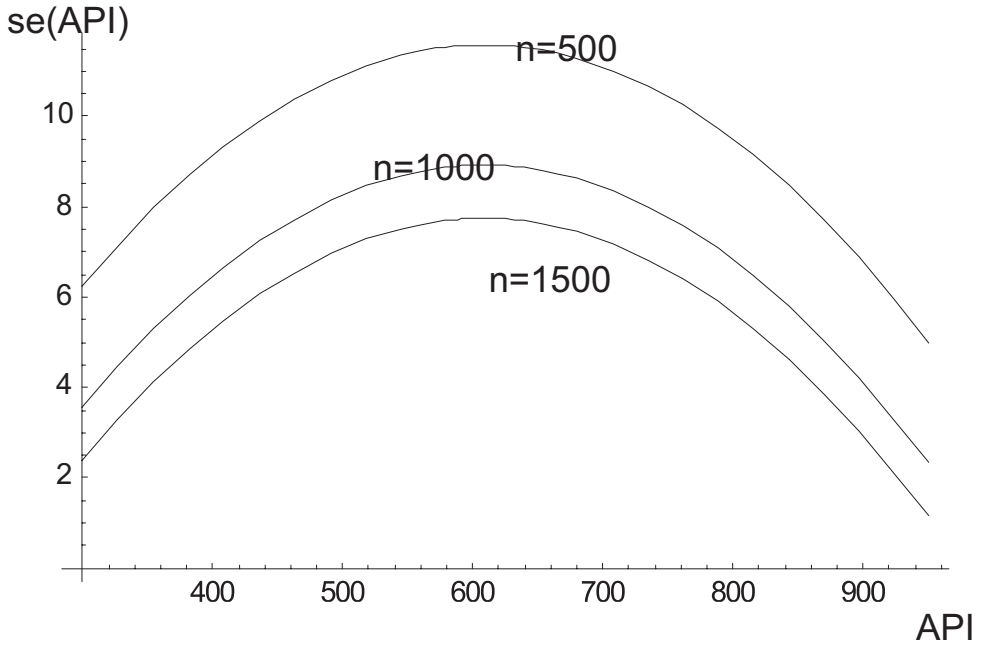
Elementary Schools fit,
5087 schools

```
Coefficients:
                          Estimate Std. Error t value
(Intercept)           -2.132e+01  5.899e-01  -36.14
API                    6.848e-02  1.614e-03   42.43
I(API^2)              -5.284e-05  1.088e-06  -48.58
I(1/sqrt(VALID_NUM))   1.895e+02  9.360e-01  202.50
Residual standard error: 0.8408 on 5083 df
Multiple R-Squared: 0.9041,Adjusted R-squared: 0.904
```



High School fit,
845 schools

```
Coefficients:
                          Estimate Std. Error
(Intercept)           -1.834e+01  4.216e+00
API                    6.854e-02  1.258e-02
I(API^2)              -5.635e-05  9.361e-06
I(1/sqrt(VALID_NUM))   2.029e+02  6.990e+00
Residual standard error: 3.285 on 841 df
Multiple R-Squared: 0.5212, Adjusted R-squared: 0.5194
```



sample variance for all included 2003 schools.

Table 5 presents reliability coefficients for Elementary School and High School scores for both the 2003 and 1999 API at the three chosen values of school size (API n). Reliability coefficients are slightly smaller for 2003 API, but very large in both years, even for the smaller schools. Why is reliability of school scores so high? Relative standing assessments give great weight not to the accuracy of the scores, but to the ability to distinguish between low-scoring and high-scoring schools, a distinction that even rather inaccurate school scores cannot obscure.

INSERT TABLE 5

Relation to Kane-Staiger "volatility"
In "Volatility in School Test Scores: Implications for Test-Based Accountability Systems" Kane and Staiger (2002) attempt to call into question the accuracy of API scores (and school assessment scores in general). Strangely, Kane and Staiger employ disguised forms of reliability coefficients to demonstrate "volatility" (see Rogosa, 2002d, 2005) The reliability values of .99 in Table 5 translate for Kane and Staiger as one percent of variability in school API scores due to (sampling or measurement) error. From these reliability results it would seem that California API scores should receive praise from Kane and Staiger for a lack of volatility.

B. Standard Error API Comparisons: 2003 Base with 2000, 2001

Table 6 displays five-number summaries for s.e.(API) for 2003, 2001, 2000 API by School Type (4675 Elementary Schools 1092 Middle Schools, and 766 High Schools with data all 3 years). For Elementary and Middle Schools, s.e.(API) decreases over years, but High Schools show an increase in 2003. Table 7 gives additional details by showing percentiles of the distributions of the ratio of 2003 s.e.(API) to the earlier years: ratio0301 = s.e.(APIB03)/s.e.(APIB01) and ratio032k = s.e.(APIB03)/s.e.(API2k). Almost identical results are seen for Elementary and Middle Schools. High Schools show, once again, the effects of the CAHSEE inclusion on the s.e.(API).

INSERT TABLE 6
INSERT TABLE 7

But as also seen in the lower half of Table 6, school size (number of students in the API) also increases 10-20% over the period. So a more refined comparison of s.e.(API) would adjust for the changes in school size. Table 8 displays comparisons using the indices comp0301 = s.e.(APIB03)/s.e.(APIB01)*sqrt(n03/n01) and comp032k = s.e.(APIB03)/s.e.(API2k)*sqrt(n03/n2k). The adjustment for school size reduces but does not erase the observed decrease in s.e.(API). For example, Table 8 shows that more than 80% of Elementary Schools smaller s.e.(API) (adjusted for

size by comp0301) in 2003 than 2001 and more than 90% are smaller in 2003 than 2000. One additional feature of s.e.(API), dependence on API score in the 200-1000 range, serves to make the 2003 standard error smaller. For example in 2000, 46% of Elementary Schools had API scores in the interval [500,700], whereas in 2003 the value is 38% of schools.

INSERT TABLE 8

Table 5.
API Reliability Coefficients

2003 Elementary Schools	
n	reliability
150	0.974
350	0.989
500	0.992

2003 High Schools	
n	reliability
500	0.986
1000	0.992
1500	0.994

1999 Elementary Schools	
n	reliability
150	0.982
350	0.992
500	0.994

1999 High Schools	
n	reliability
500	0.991
1000	0.996
1500	0.997

Table 6
API Standard Errors and School Size Comparisons: 2003Base, 2000, 2001Base

School s.e.(API) by School Type:
Elem (n=4675), Mid(n=1092), High(n=766)

Elem

	Min	Q1	Med	Q3	Max
2003	2.890	8.265	9.620	11.400	23.200
2001	3.050	9.285	11.000	12.800	23.100
2000	2.6	10.0	11.9	14.0	27.0

Mid

	Min	Q1	Med	Q3	Max
2003	3.120	5.820	6.695	7.890	20.700
2001	3.2800	6.5300	7.6900	9.1125	21.2000
2000	3.6700	7.0775	8.3300	9.9100	22.4000

High

	Min	Q1	Med	Q3	Max
2003	2.5300	5.9425	6.9100	9.3050	39.3000
2001	1.8900	5.3400	6.2500	7.7375	22.0000
2000	2.090	5.630	6.525	8.350	20.600

Comparisons of School Size by School Type

Elementary Schools

	Size '03	'03/'01	'03/2k
Min	100	0.324	0.298
Q1	287	1.000	1.020
Med	386	1.070	1.110
Q3	504	1.140	1.210
Max	2426	2.880	5.160

Middle Schools

	Size '03	'03/'01	'03/2k
Min	104	0.615	0.552
Q1	646	1.060	1.100
Med	861	1.130	1.190
Q3	1148	1.210	1.300
Max	4003	2.310	6.470

High Schools

	Size '03	'03/'01	'03/2k
Min	109	0.516	0.48
Q1	813	1.040	1.06
Med	1323	1.110	1.15
Q3	1725	1.170	1.27
Max	3478	2.230	4.66

Table 7
Ratio of s.e.(API) 2003 to prior years by School Type:
percentiles of school distributions

Elementary Schools (n=4675)

p	ratio0301	ratio032k
0.05	0.75	0.67
0.10	0.78	0.70
0.15	0.80	0.72
0.20	0.81	0.74
0.25	0.83	0.76
0.30	0.84	0.77
0.35	0.85	0.78
0.40	0.86	0.79
0.45	0.87	0.80
0.50	0.88	0.82
0.55	0.89	0.83
0.60	0.90	0.84
0.65	0.92	0.85
0.70	0.93	0.87
0.75	0.95	0.88
0.80	0.97	0.90
0.85	1.00	0.93
0.90	1.00	0.97
0.95	1.10	1.00
1.00	1.80	1.80

High Schools (n=766)

p	ratio0301	ratio032k
0.05	0.94	0.85
0.10	0.98	0.90
0.15	1.00	0.94
0.20	1.00	0.97
0.25	1.00	0.99
0.30	1.10	1.00
0.35	1.10	1.00
0.40	1.10	1.00
0.45	1.10	1.00
0.50	1.10	1.10
0.55	1.10	1.10
0.60	1.10	1.10
0.65	1.10	1.10
0.70	1.20	1.10
0.75	1.20	1.10
0.80	1.20	1.20
0.85	1.20	1.20
0.90	1.30	1.30
0.95	1.75	1.70
1.00	5.90	5.70

Middle Schools n=1092)

p	ratio0301	ratio032k
0.05	0.75	0.690
0.10	0.78	0.720
0.15	0.80	0.730
0.20	0.82	0.750
0.25	0.83	0.760
0.30	0.84	0.770
0.35	0.85	0.788
0.40	0.86	0.800
0.45	0.87	0.810
0.50	0.88	0.820
0.55	0.89	0.830
0.60	0.90	0.840
0.65	0.91	0.850
0.70	0.93	0.870
0.75	0.94	0.880
0.80	0.96	0.890
0.85	0.99	0.910
0.90	1.00	0.950
0.95	1.10	1.000
1.00	1.70	1.600

Table 8
Comparison Indices for s.e.(API) by School Type:
percentiles of school distributions

Elementary Schools (n=4675)

p	comp0301	comp032k
0.05	0.799	0.739
0.10	0.822	0.765
0.15	0.837	0.780
0.20	0.850	0.794
0.25	0.861	0.806
0.30	0.870	0.817
0.35	0.880	0.828
0.40	0.889	0.838
0.45	0.898	0.848
0.50	0.909	0.857
0.55	0.920	0.867
0.60	0.931	0.878
0.65	0.943	0.890
0.70	0.956	0.903
0.75	0.972	0.918
0.80	0.990	0.934
0.85	1.019	0.956
0.90	1.058	0.990
0.95	1.129	1.064
1.00	1.634	1.981

High Schools (n=766)

p	comp0301	comp032k
0.05	1.025	0.978
0.10	1.056	1.027
0.15	1.080	1.053
0.20	1.100	1.069
0.25	1.112	1.080
0.30	1.123	1.092
0.35	1.134	1.105
0.40	1.142	1.113
0.45	1.153	1.121
0.50	1.163	1.132
0.55	1.174	1.142
0.60	1.185	1.155
0.65	1.193	1.171
0.70	1.206	1.185
0.75	1.219	1.202
0.80	1.238	1.215
0.85	1.262	1.248
0.90	1.298	1.296
0.95	1.734	1.737
1.00	6.020	5.873

Middle Schools n=1092)

p	comp0301	comp032k
0.05	0.836	0.789
0.10	0.852	0.807
0.15	0.866	0.825
0.20	0.877	0.838
0.25	0.886	0.850
0.30	0.895	0.858
0.35	0.904	0.864
0.40	0.913	0.873
0.45	0.921	0.883
0.50	0.930	0.890
0.55	0.941	0.898
0.60	0.953	0.906
0.65	0.965	0.915
0.70	0.976	0.925
0.75	0.990	0.936
0.80	1.008	0.951
0.85	1.033	0.976
0.90	1.059	1.001
0.95	1.128	1.057
1.00	1.924	1.889

Part IV: Accuracy for Reported Subgroup scores: Socioeconomically Disadvantaged (SD) Subgroup

According to the definition adopted by the SBE, the "socioeconomically disadvantaged subgroup" consists of pupils who meet either one of two criteria:
1) Neither of the pupil's parents has received a high school diploma
OR
2) The pupil participates in the free-or reduced price lunch program.
The PSAA defines a "numerically significant ethnic or socioeconomically disadvantaged subgroup" as a subgroup "that constitutes at least fifteen percent of a school's total pupil population and consists of at least thirty pupils." Also, under the law, if a subgroup defined by ethnicity or socioeconomic disadvantage constitutes at least 100 pupils, i.e., at least 100 pupils with valid STAR scores, that subgroup is numerically significant. (CDE Explanatory Notes).

In the analyses of this section, results are presented for both SD subgroups of at least 30 students and the subset of those subgroups which are deemed "numerically significant". SD is the largest subgroup designation in the API reporting--over half of all 2003 California Elementary School and Middle School students are classified SD. Tables 9 and 10 provide some descriptive statistics on the number, sizes and API performance of the SD subgroups. Over 90% of California schools have at least 30 SD students and over 90% of those SD groups are deemed numerically significant by the CDE criteria. Significant SD subgroups will tend to be slightly larger. Of the 85% of Elementary Schools having significant subgroups, half are composed of at least 2/3 SD students. The size of SD subgroups tend to be smaller for higher scoring (high decile) schools (which tend to be slightly smaller). And from Table 10, API scores for SD subgroups tend to be 50 points lower than the full school. Figure 5 shows plots of SD API vs size of SD subgroup for each School Type.

INSERT TABLE 9
INSERT TABLE 10
INSERT FIGURE 5

Tables 11, 12 and 13 plus Figure 6 display results for the standard error of the SD subgroup API scores. Figure 6 replicates the displays in Figure 1 for both SD subgroups with at least 30 students and the subset of numerically significant SD subgroups (indicated with the SD_API > 0 in the plot axes labels reflecting a reported value in the CDE research files). The pair of plots for each School Type are s.e.(API) versus SD subgroup API and s.e.(API) versus 1/Sqrt(SD subgroup size). Tables 11 and 12 repeat the displays in Tables 1 and 2 and allow comparisons of magnitudes of s.e.(API) for SD subgroups with school s.e.(API). Table 13 gives further details (percentiles) on the distribution of s.e.(API) for the SD subgroups.

INSERT TABLE 11
INSERT TABLE 12
INSERT TABLE 13
INSERT FIGURE 6

Table 9
Number and Size of SD Subgroups

Number of SD subgroups by School Type

	Elem	Mid	High
SD at least 30 students	4566	1125	881
SD Significant Subgroups	4257	1044	806
Total schools (in database)	5053	1175	944

Size of SD Subgroups by School Type

	Min	Q1	Med	Q3	Max
SD at least 30 students	30	116	234	420	3718
Elem	30	107	205	353	2426
Mid	31	182	359	643	3718
High	30	117	290	649	3250
SD Significant Subgroups	30	137	256	441	3718
Elem	30	125	221	365	2426
Mid	31	220	398	671	3718
High	30	152	327	684	3250

Relative Size (school proportion SD)
of SD Significant Subgroups by School Type

	Min	Q1	Med	Q3	Max
Elem	0.0885	0.409	0.661	0.868	1.0
Mid	0.0628	0.326	0.539	0.752	1.0
High	0.0468	0.240	0.404	0.609	1.0

SD Subgroups and Size by School-Type and School State Decile Rank

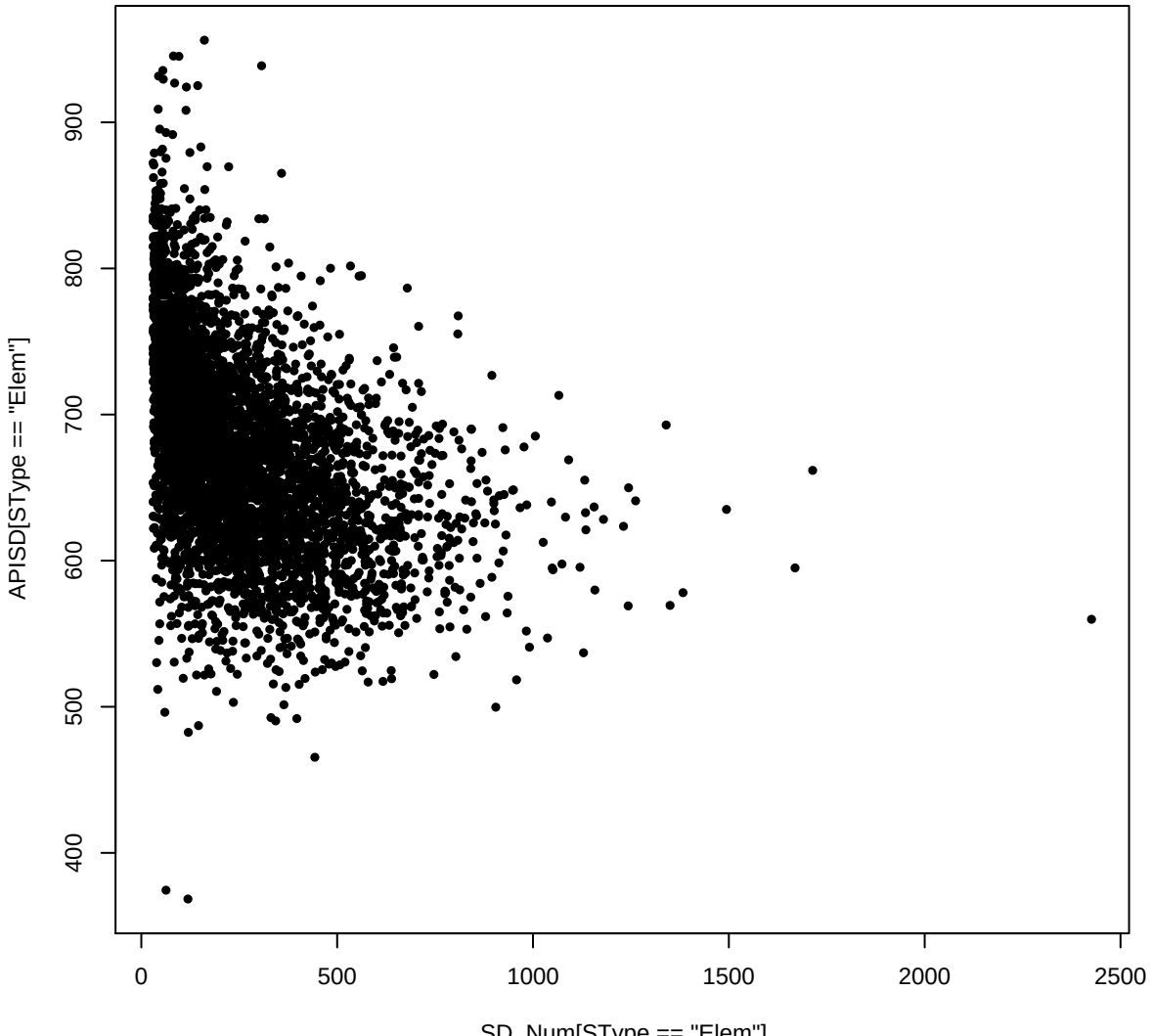
	State Decile Rank									
	1	2	3	4	5	6	7	8	9	10
SD Subgroups at least 30 students										
Elem										
number of subgroups	496	499	498	508	489	509	495	467	413	192
median size	411	387	328	279	224	170	136	103	67	49
Mid										
number of subgroups	116	113	115	122	116	119	116	115	115	76
median size	823	763	625	539	409	324	255	198	125	80
High										
number of subgroups	79	81	81	84	80	86	79	82	84	59
median size	868	775	632	550	373	362	235	215	136	94
SD Numerically Significant Subgroups										
Elem										
number of subgroups	496	498	498	507	489	504	486	431	281	67
median size	411	387	328	279	224	172	137	109	86	81
Mid										
number of subgroups	116	113	115	122	116	118	115	108	92	27
median size	823	763	625	539	409	324	257	205	155	143
High										
number of subgroups	79	81	80	82	78	84	75	75	59	30
median size	868	775	632	566	381	376	242	239	168	169

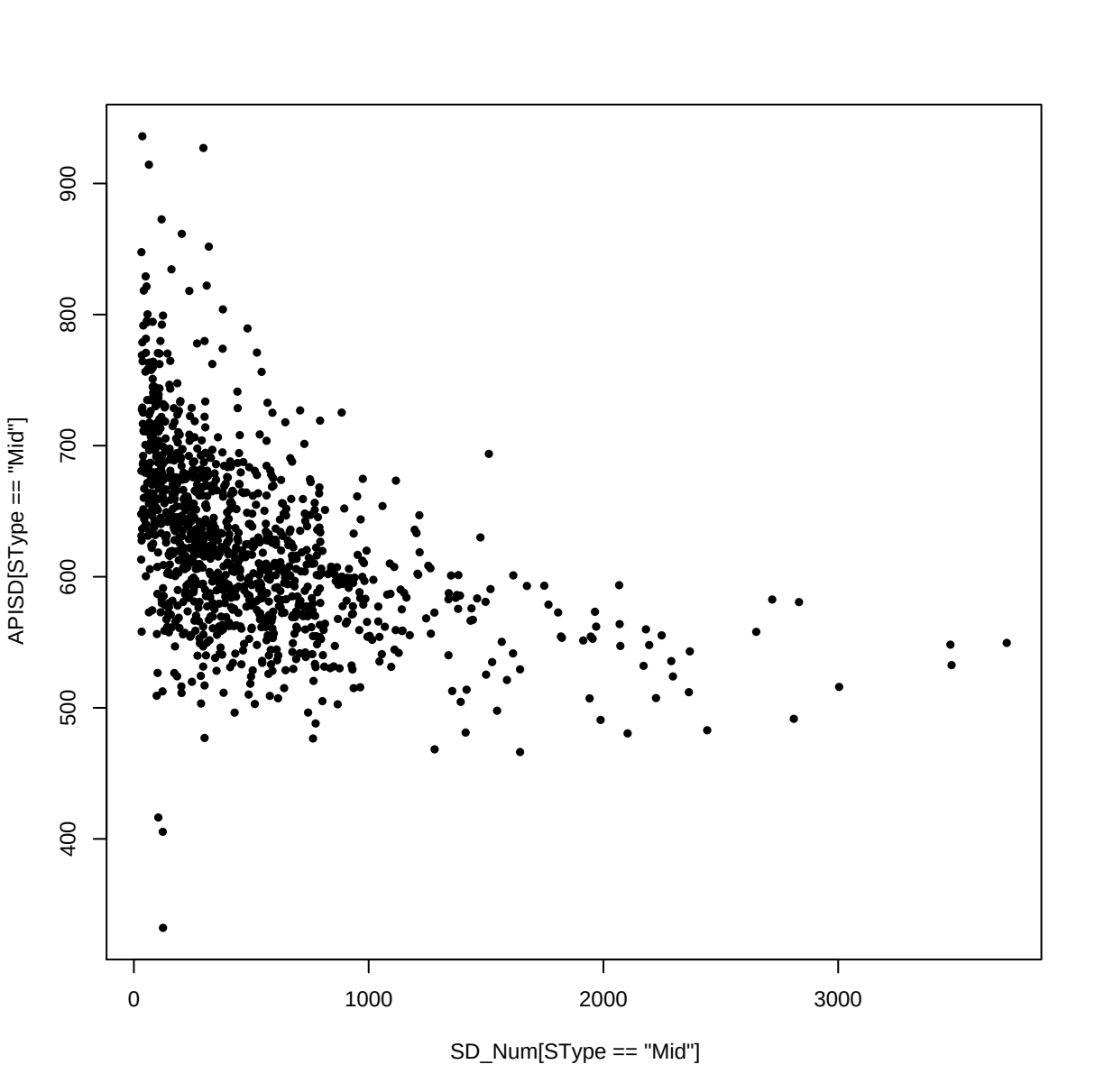
Table 10
API scores for SD Subgroups

API scores for SD Subgroups by School Type

	Min	Q1	Med	Q3	Max
SD at least 30 students	325.9	601.3	653.9	705.7	956.2
Elem	368.4	629.5	675.3	720.0	956.2
Mid	332.2	579.9	619.4	665.9	936.0
High	325.9	524.1	567.8	613.1	935.5
SD Significant Subgroups	325.9	599.2	649.2	698.8	956.2
Elem	368.4	626.4	670.8	712.8	956.2
Mid	332.2	577.2	614.0	659.4	927.0
High	325.9	522.4	564.0	608.7	935.5
School API 2003 Base					
Elem	367.4	659.1	728.9	804	996.4
Mid	339	617.2	686	763	957.5
High	332	576	645	714	946

Figure 5. Plots of SD API vs size of SD subgroup by School Type




APISD[SType == "Mid"]
400
500
600
700
800
900
0
1000
2000
3000
SD_Num[SType == "Mid"]

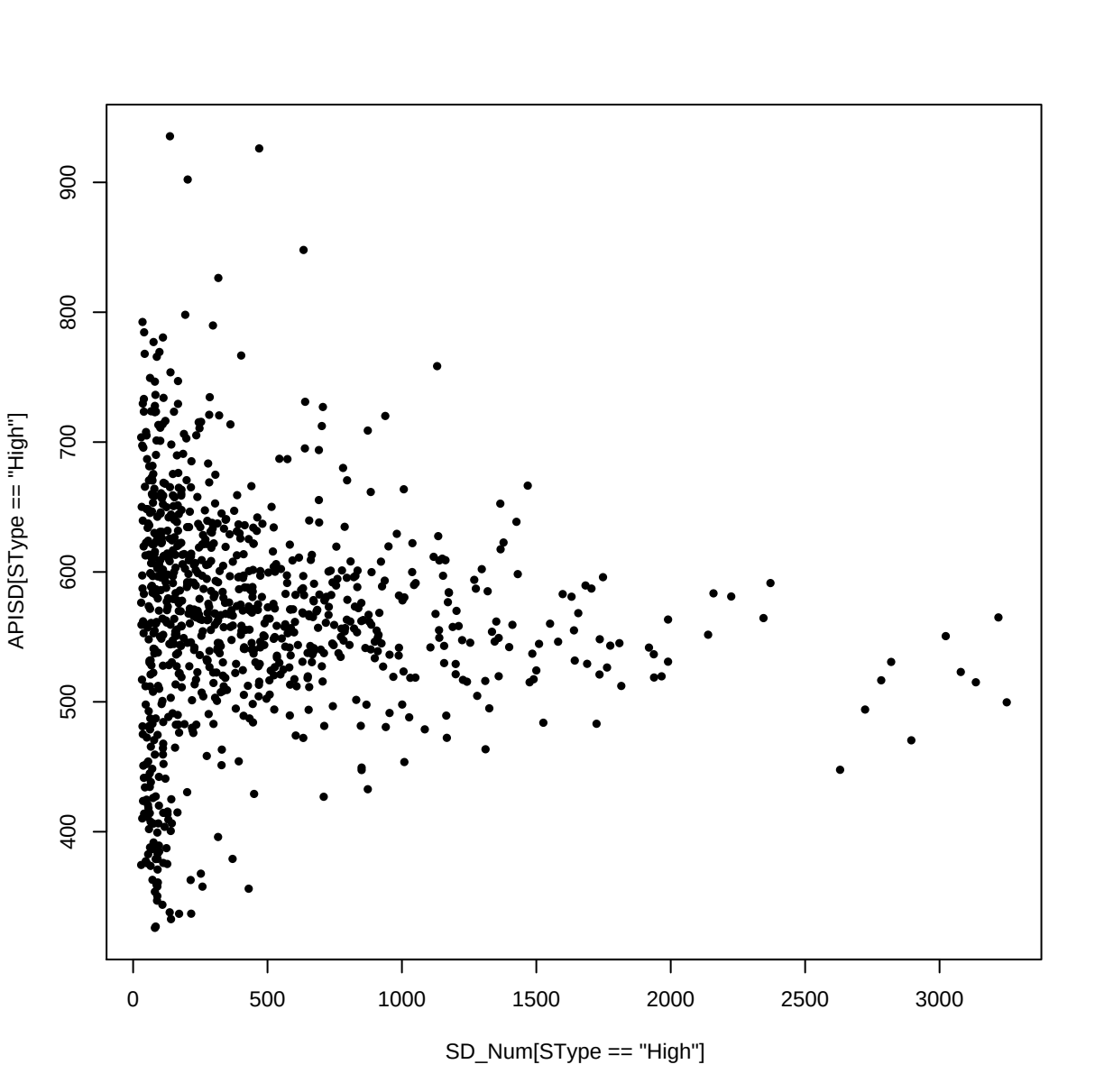

APISD[SType == "High"]
400
500
600
700
800
900
0
500
1000
1500
2000
2500
3000
SD_Num[SType == "High"]

Table 11.
Standard Error of SD Subgroup API (bootstrap resampling)

Descriptive Statistics: s.e.(API)

		Min	Q1	Med	Q3	Max
SD at least 30 students						
	Elem	4.36	10.50	13.70	18.70	52.20
	Mid	3.27	7.97	10.60	15.00	47.00
	High	4.03	9.87	14.70	21.50	65.00
SD Significant Subgroups						
	Elem	4.36	10.30	13.30	17.50	45.40
	Mid	3.27	7.72	10.10	13.80	47.00
	High	4.03	9.61	13.75	19.37	65.00

Table 12
Median se(API) and median size SD subgroup by School Type and School State Decile

Elementary Schools

SD at least 30 students

School State Decile	number SD subgroup	Median SD Size subgroup	Median se(API)
1	496	411.5	9.82
2	499	387.0	10.30
3	498	328.0	11.10
4	508	279.0	12.00
5	489	224.0	13.40
6	509	170.0	14.90
7	495	136.0	16.70
8	467	103.0	18.60
9	413	67.0	22.80
10	192	49.0	23.95

SD Numerically Significant Subgroups

School State Decile	number SD subgroup	Median SD Size subgroup	Median se(API)
1	496	411.5	9.82
2	498	387.5	10.30
3	498	328.0	11.10
4	507	279.0	12.00
5	489	224.0	13.40
6	504	172.5	14.90
7	486	137.0	16.65
8	431	109.0	18.00
9	281	86.0	20.20
10	67	81.0	18.40

Middle Schools

SD at least 30 students

School State Decile	number SD subgroup	Median SD Size subgroup	Median se(API)
1	116	823.5	6.895
2	113	763.0	7.080
3	115	625.0	8.060
4	122	539.0	8.900
5	116	409.0	10.200
6	119	324.0	11.200
7	116	255.0	12.800
8	115	198.0	14.600
9	115	125.0	17.400
10	76	80.0	21.000

SD Numerically Significant Subgroups

School State Decile	number SD subgroup	Median SD Size subgroup	Median se(API)
1	116	823.5	6.895
2	113	763.0	7.080
3	115	625.0	8.060
4	122	539.0	8.900
5	116	409.0	10.200
6	118	324.5	11.200
7	115	257.0	12.800
8	108	205.5	14.350
9	92	155.0	16.350
10	27	143.0	17.100

High Schools

SD at least 30 students

School State Decile	number SD subgroup	Median SD Size subgroup	Median se(API)
1	79	868.0	7.75
2	81	775.0	9.54
3	81	632.0	10.30
4	84	550.0	10.85
5	80	373.5	12.95
6	86	362.0	13.05
7	79	235.0	17.40
8	82	215.5	17.75
9	84	136.5	21.80
10	59	94.0	24.60

SD Numerically Significant Subgroups

School State Decile	number SD subgroup	Median SD Size subgroup	Median se(API)
1	79	868.0	7.75
2	81	775.0	9.54
3	80	632.5	10.20
4	82	566.5	10.65
5	78	381.0	12.60
6	84	376.0	12.80
7	75	242.0	17.30
8	75	239.0	17.30
9	59	168.0	19.20
10	30	169.5	17.70

Table 13
Percentiles of SD Subgroups s.e.(API) distribution by School Type

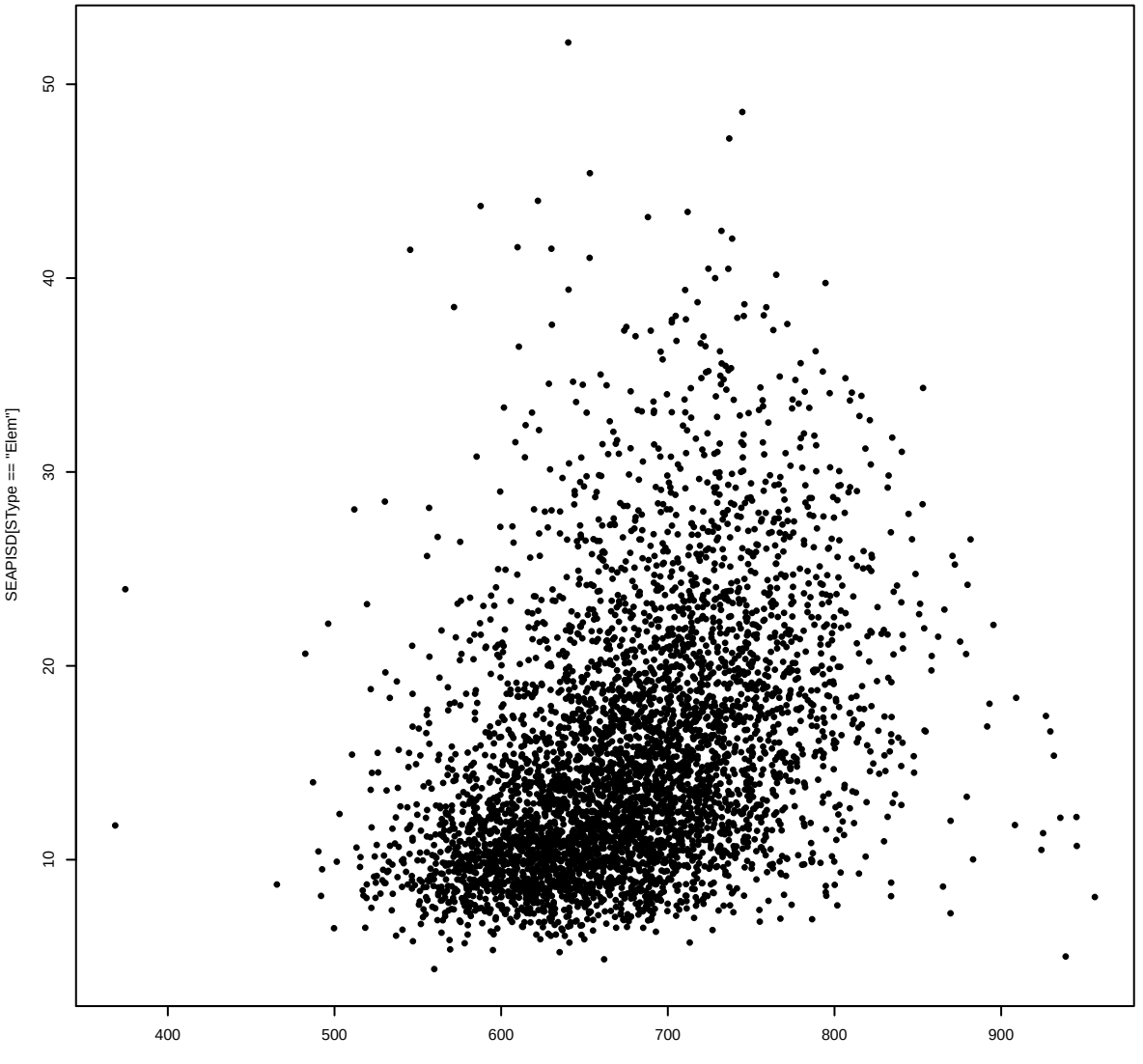
SD at least 30 students

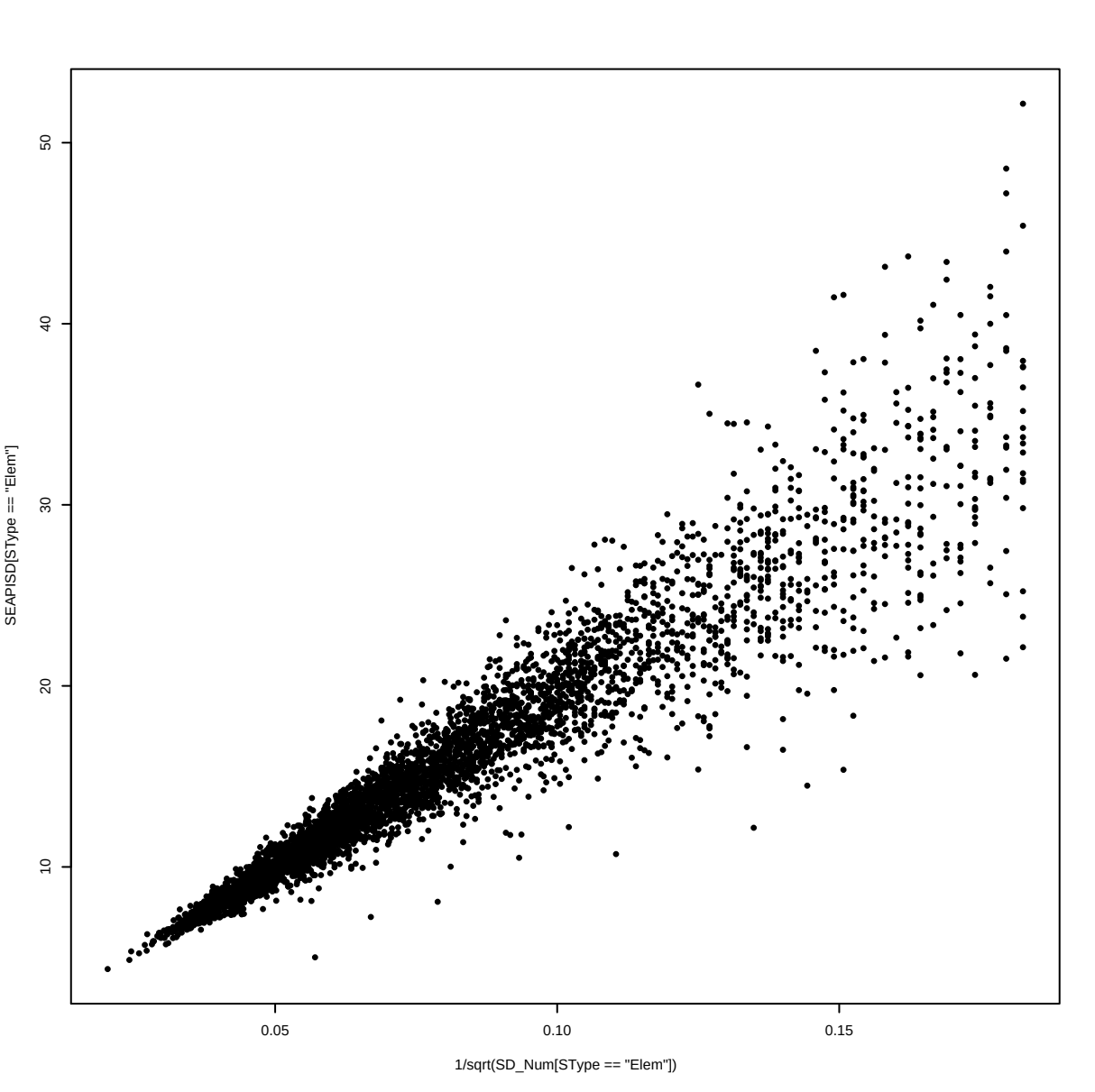
p	Elem	Mid	High
0.00	4.359936	3.272886	4.027012
0.05	7.881909	5.340971	6.346924
0.10	8.644623	6.300187	7.384903
0.15	9.402983	6.931016	8.269333
0.20	9.960144	7.405650	9.214317
0.25	10.498194	7.970459	9.869598
0.30	11.051823	8.415569	10.599937
0.35	11.661932	8.846384	11.494278
0.40	12.341652	9.332097	12.419495
0.45	12.973099	9.876840	13.521568
0.50	13.733894	10.584511	14.670982
0.55	14.554144	11.159293	16.041168
0.60	15.414604	11.801059	17.362000
0.65	16.490967	12.770102	18.481186
0.70	17.477295	13.806569	19.806091
0.75	18.706844	14.968575	21.480042
0.80	20.133347	16.480579	23.396591
0.85	21.953323	18.032407	25.254730
0.90	24.371231	21.053680	28.050095
0.95	28.333374	24.334021	32.932465
1.00	52.154266	47.041138	64.975830

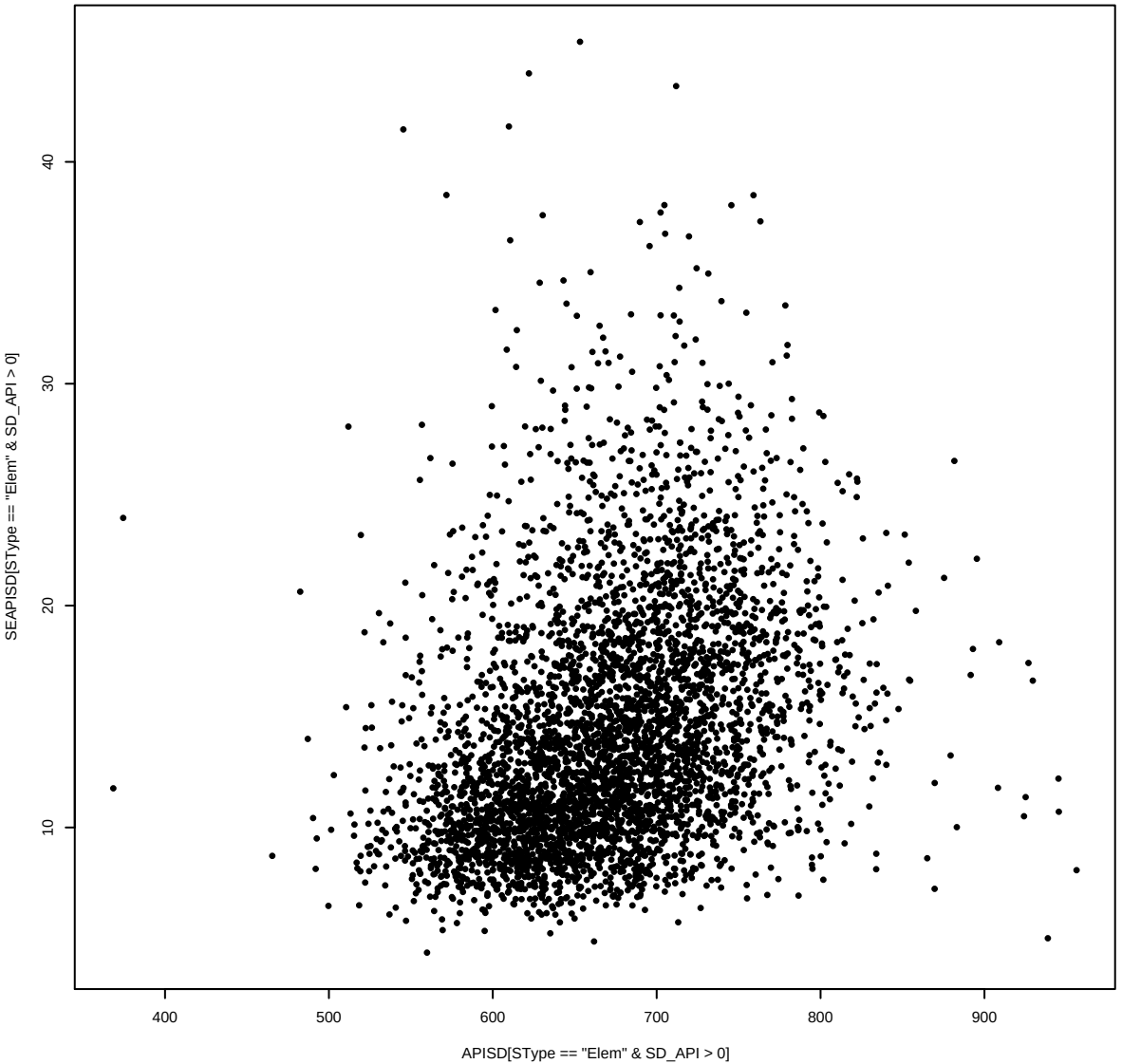
SD Numerically Significant Subgroups

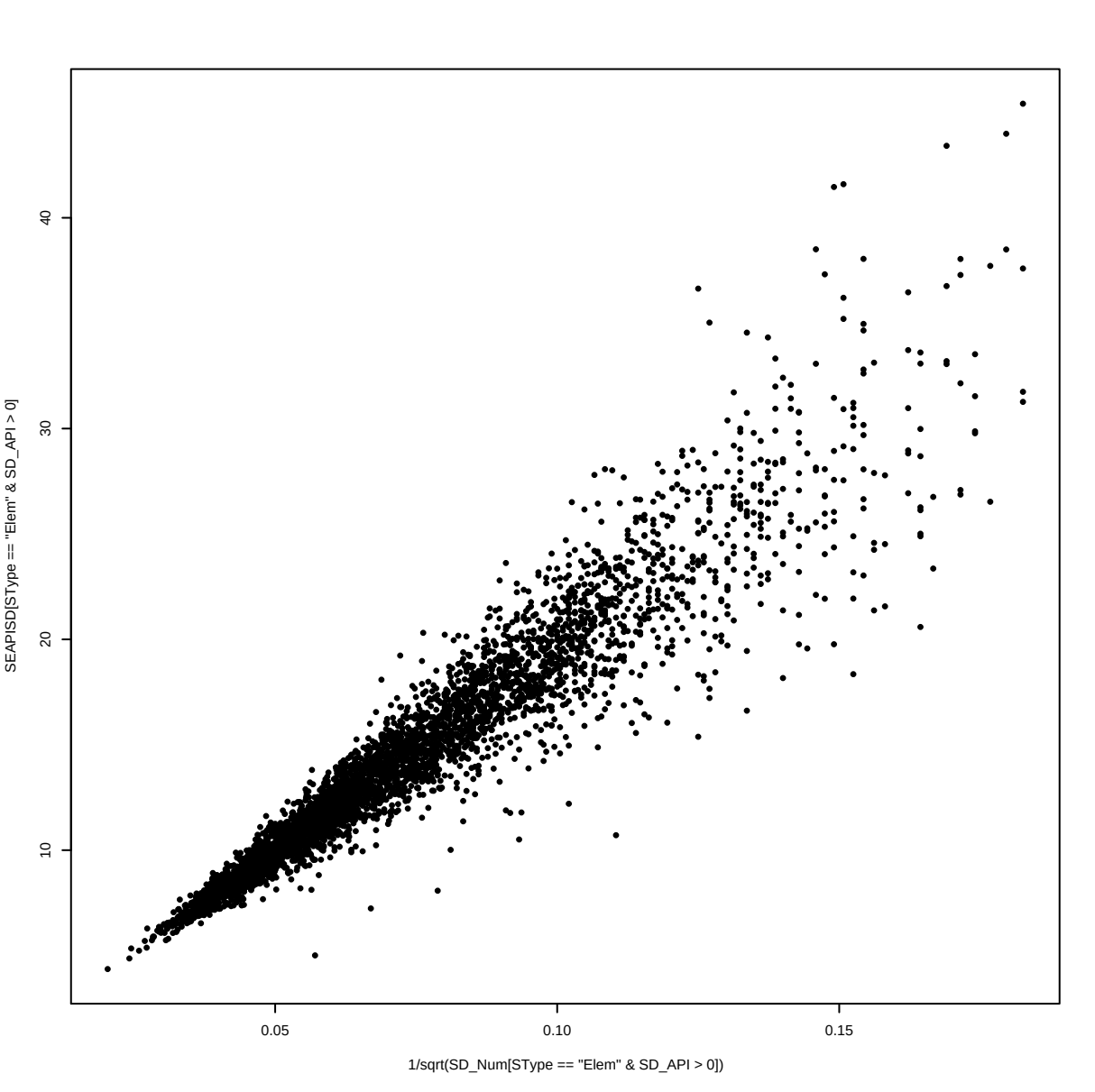
p	Elem	Mid	High
0.00	4.359936	3.272886	4.027012
0.05	7.812766	5.279841	6.168777
0.10	8.571751	6.211192	7.265350
0.15	9.240067	6.869435	8.061640
0.20	9.832622	7.277312	8.889580
0.25	10.286789	7.720880	9.609898
0.30	10.836005	8.210251	10.183163
0.35	11.374248	8.601574	10.942566
0.40	11.930025	9.027676	11.780045
0.45	12.583600	9.501501	12.567244
0.50	13.250534	10.084358	13.738010
0.55	13.913554	10.701131	14.758282
0.60	14.690503	11.304115	16.013474
0.65	15.495503	11.847312	17.225613
0.70	16.563565	12.771759	18.335281
0.75	17.486694	13.784510	19.371365
0.80	18.610437	14.811360	21.041443
0.85	19.959039	16.077129	22.730846
0.90	21.892661	17.635797	24.861618
0.95	25.034293	20.060606	28.027779
1.00	45.412659	47.041138	64.975830

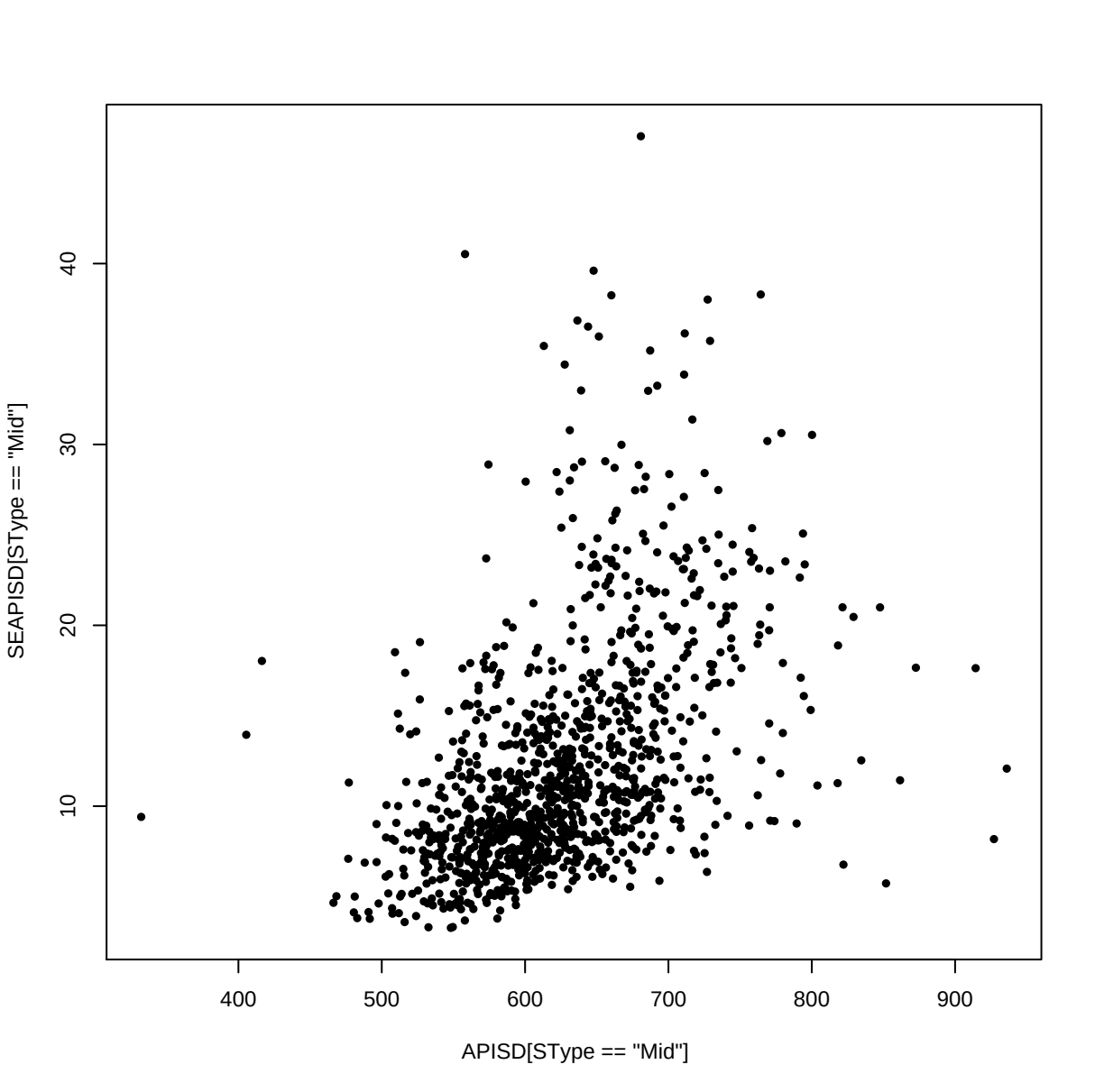
Figure 6. Plots of se(API) vs API for SD subgroups

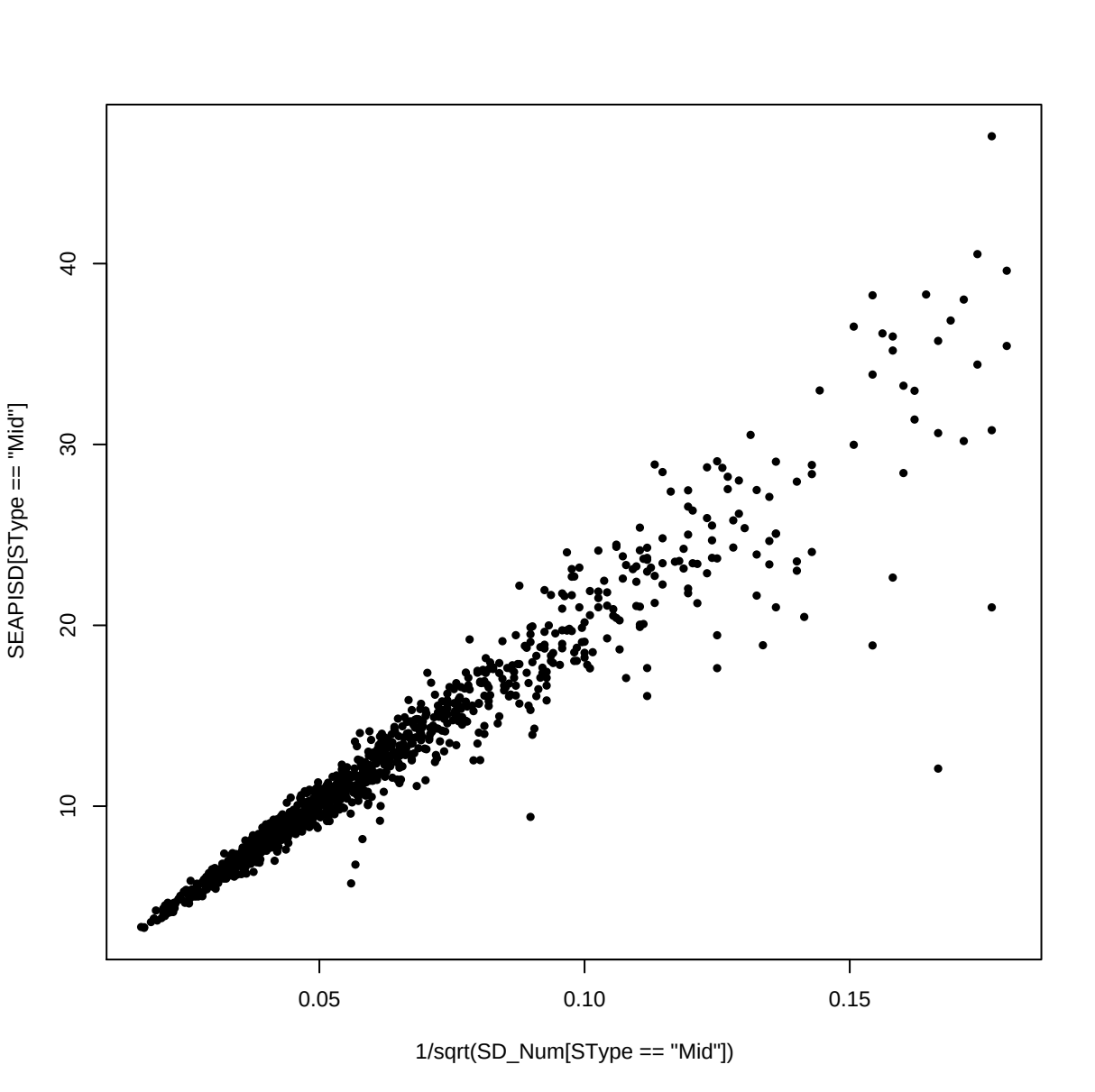


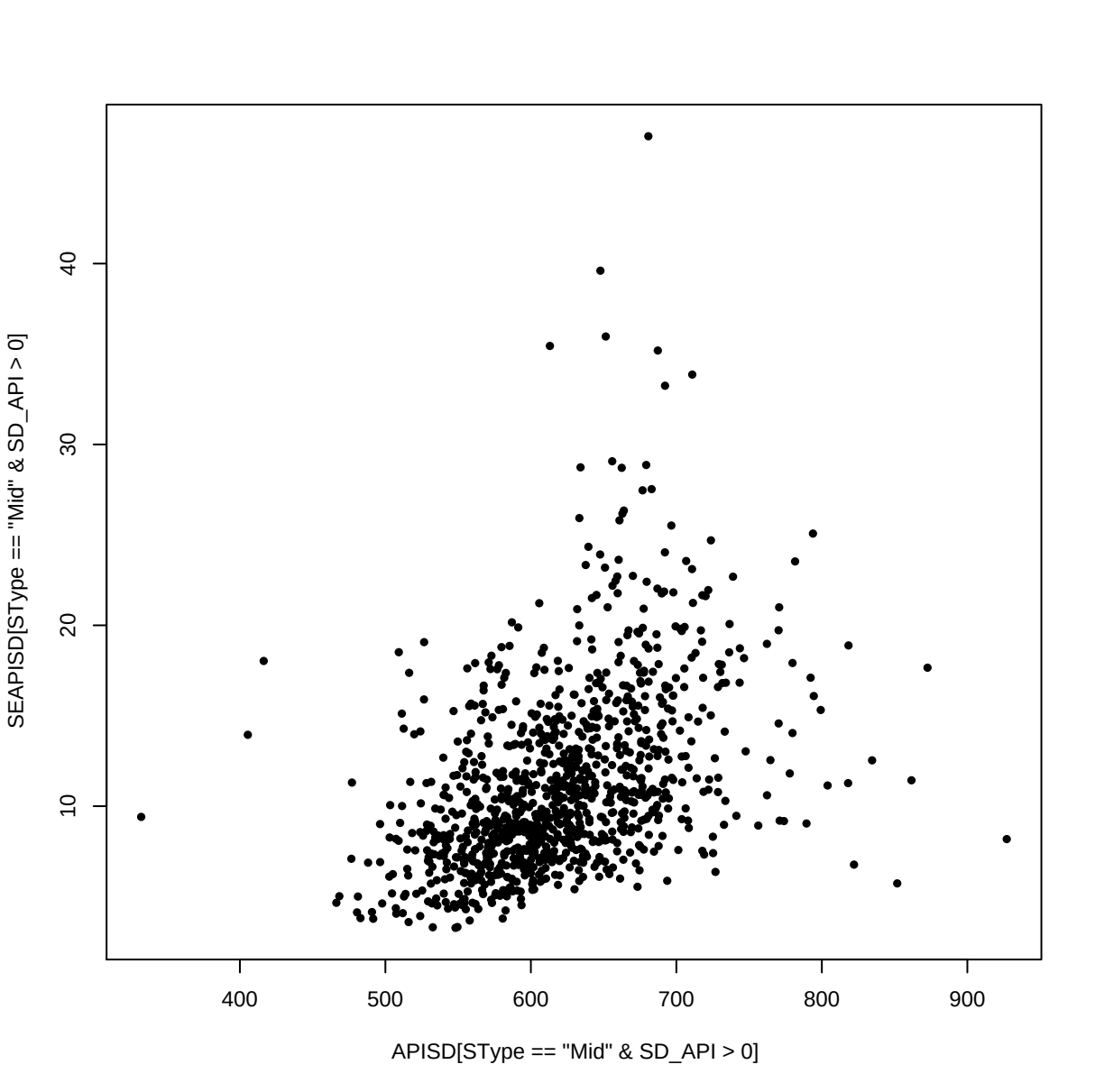

SEAPISD[SType == "Elem"]
10
20
30
40
50
0.05
0.10
0.15
1/sqrt(SD_Num[SType == "Elem"])

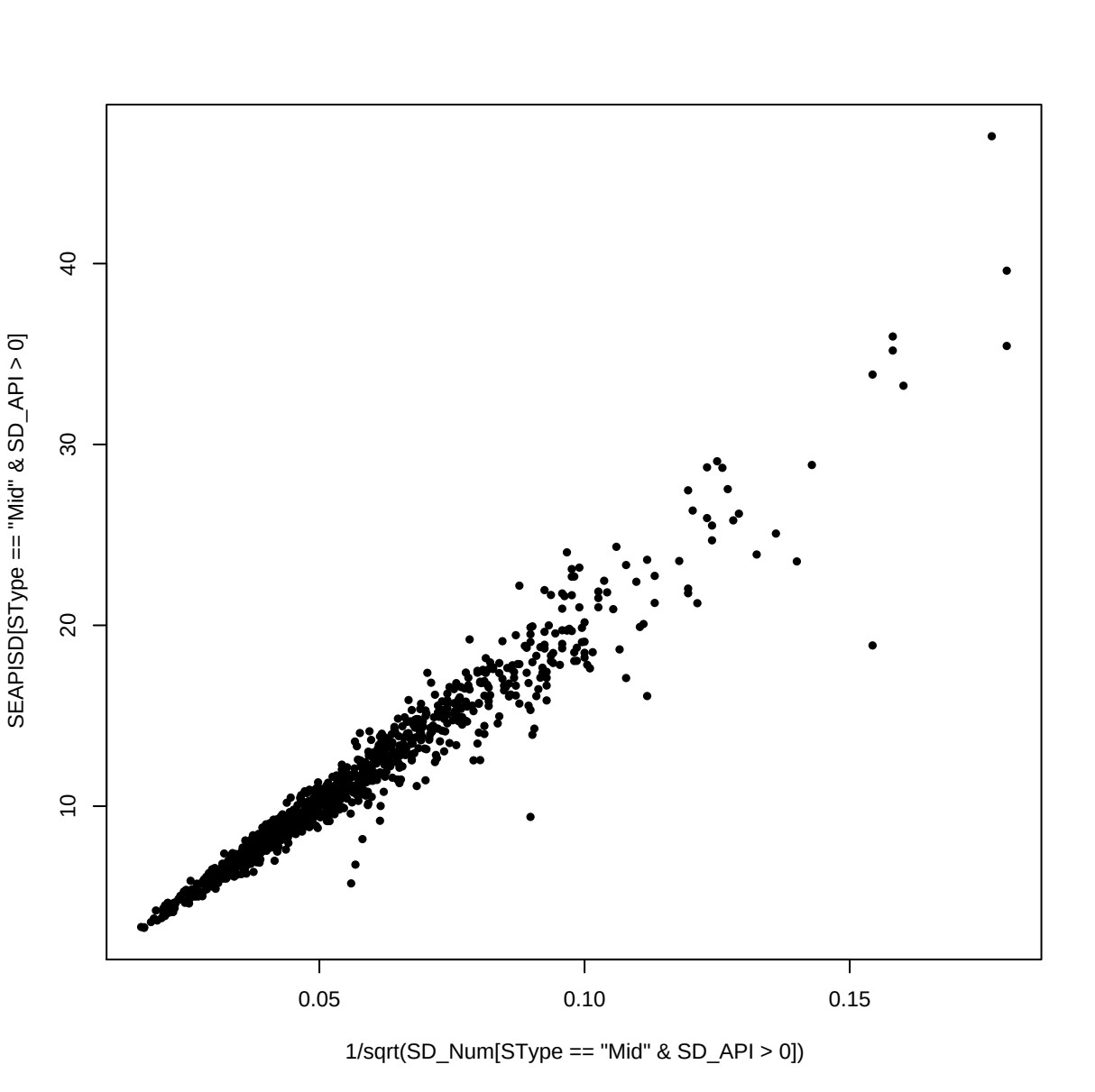

SEAPISD[SType == "Elem" & SD_API > 0]
10
20
30
40
APISD[SType == "Elem" & SD_API > 0]
400
500
600
700
800
900

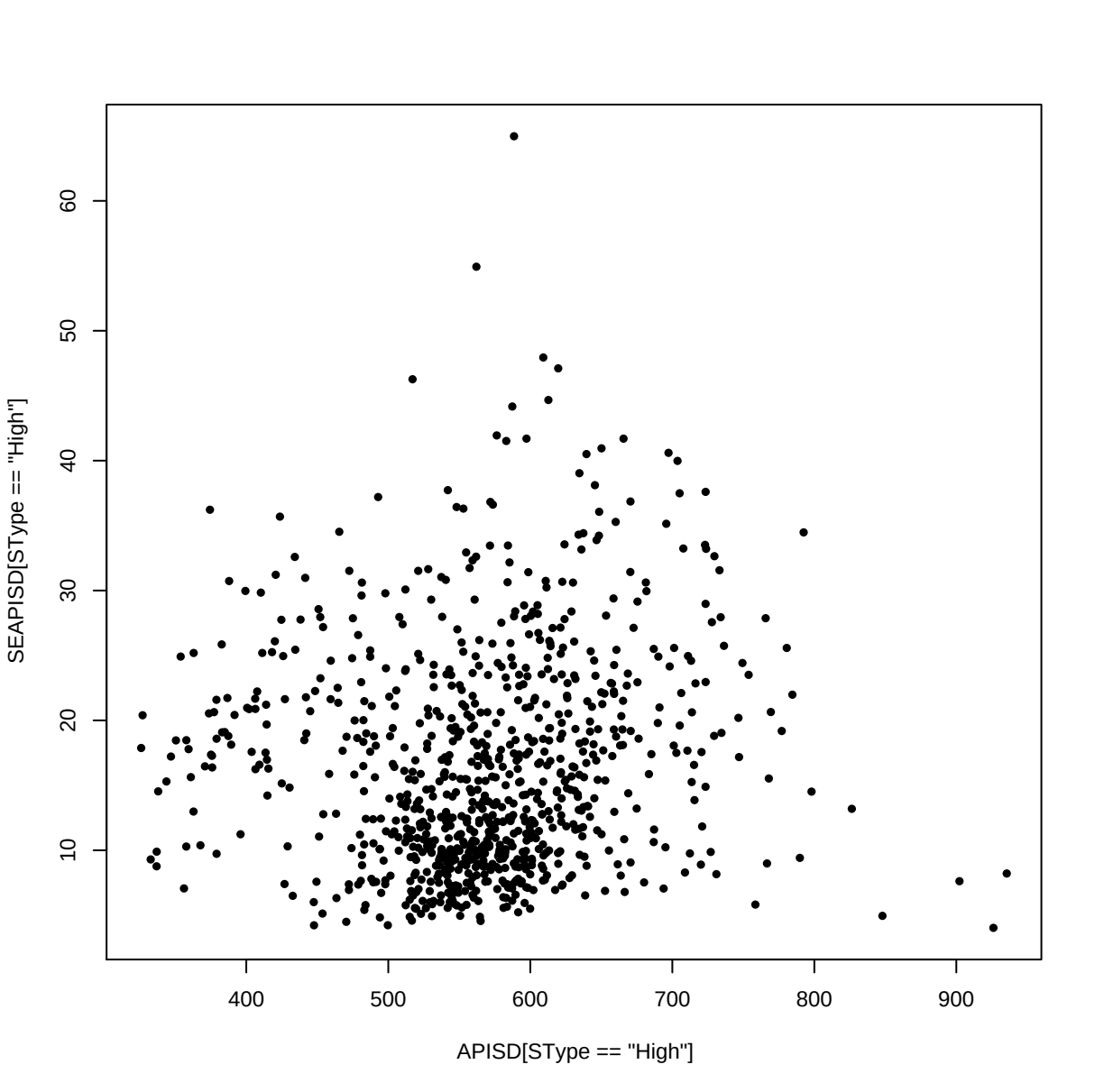

SEAPISD[SType == "Elem" & SD_API > 0]
10
20
30
40
0.05
0.10
0.15
1/sqrt(SD_Num[SType == "Elem" & SD_API > 0])

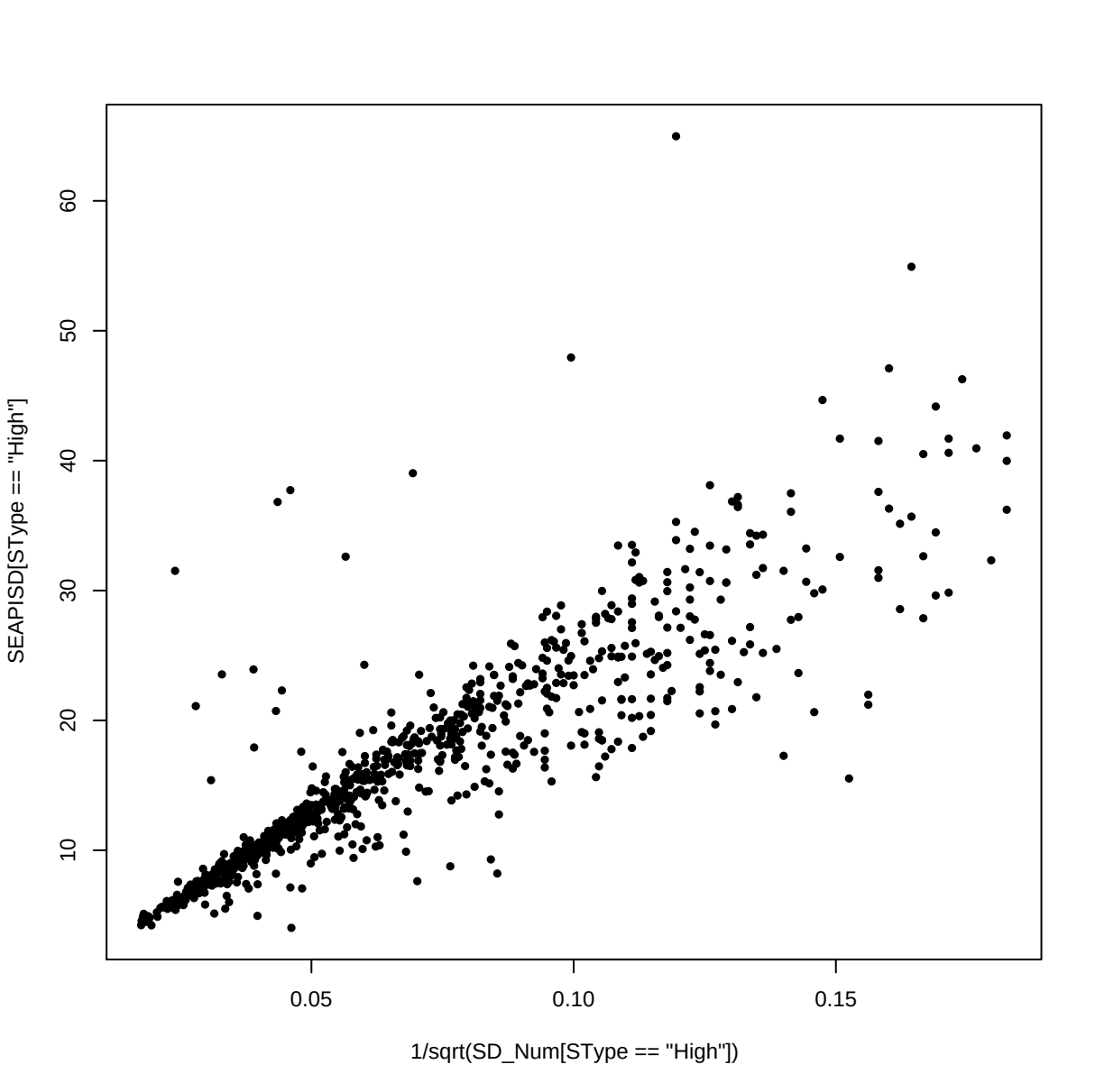

SEAPISD[SType == "Mid"]
10
20
30
40
400
500
600
700
800
900
APISD[SType == "Mid"]

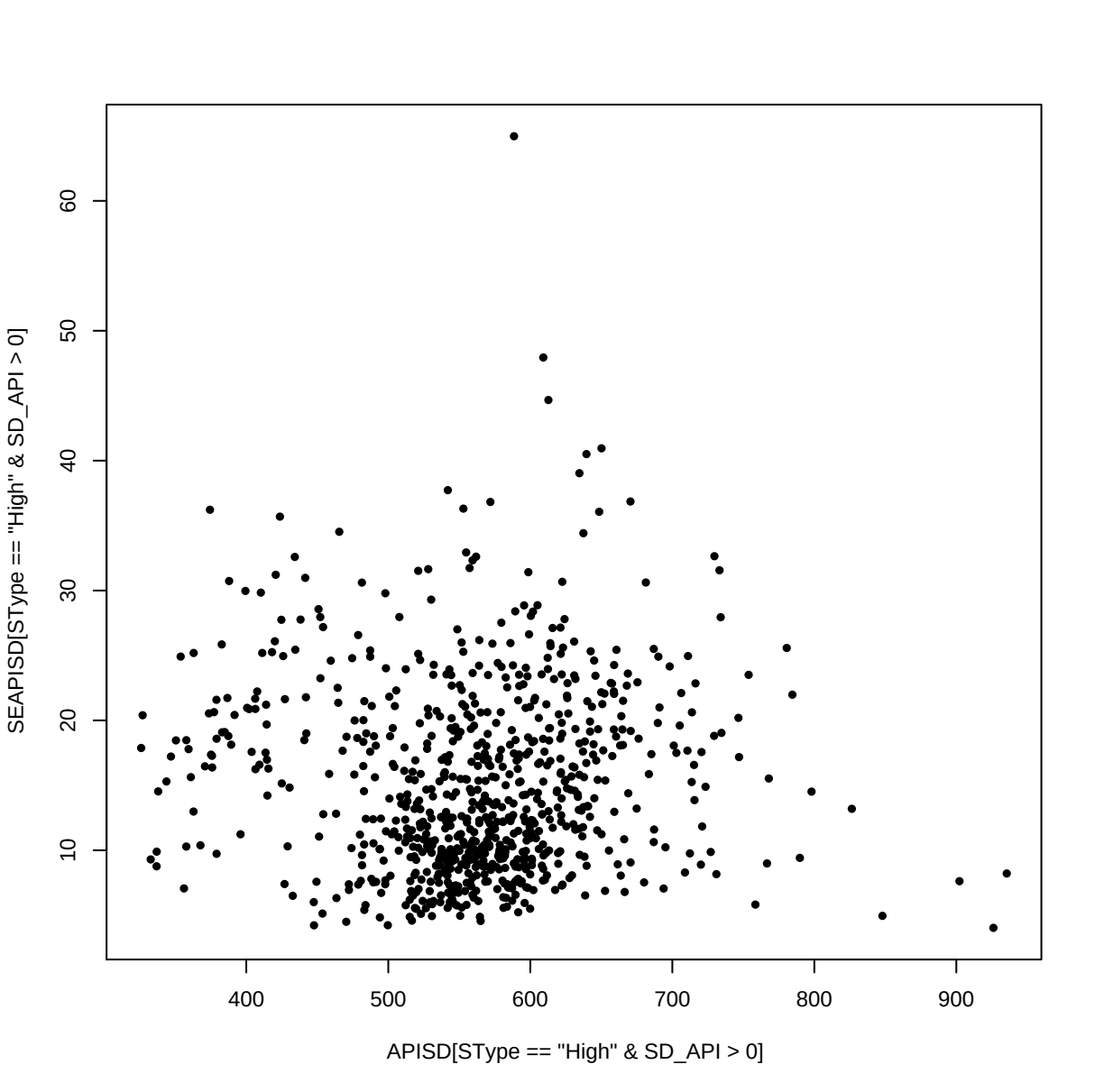

SEAPISD[SType == "Mid"]
10
20
30
40
0.05
0.10
0.15
1/sqrt(SD_Num[SType == "Mid"])

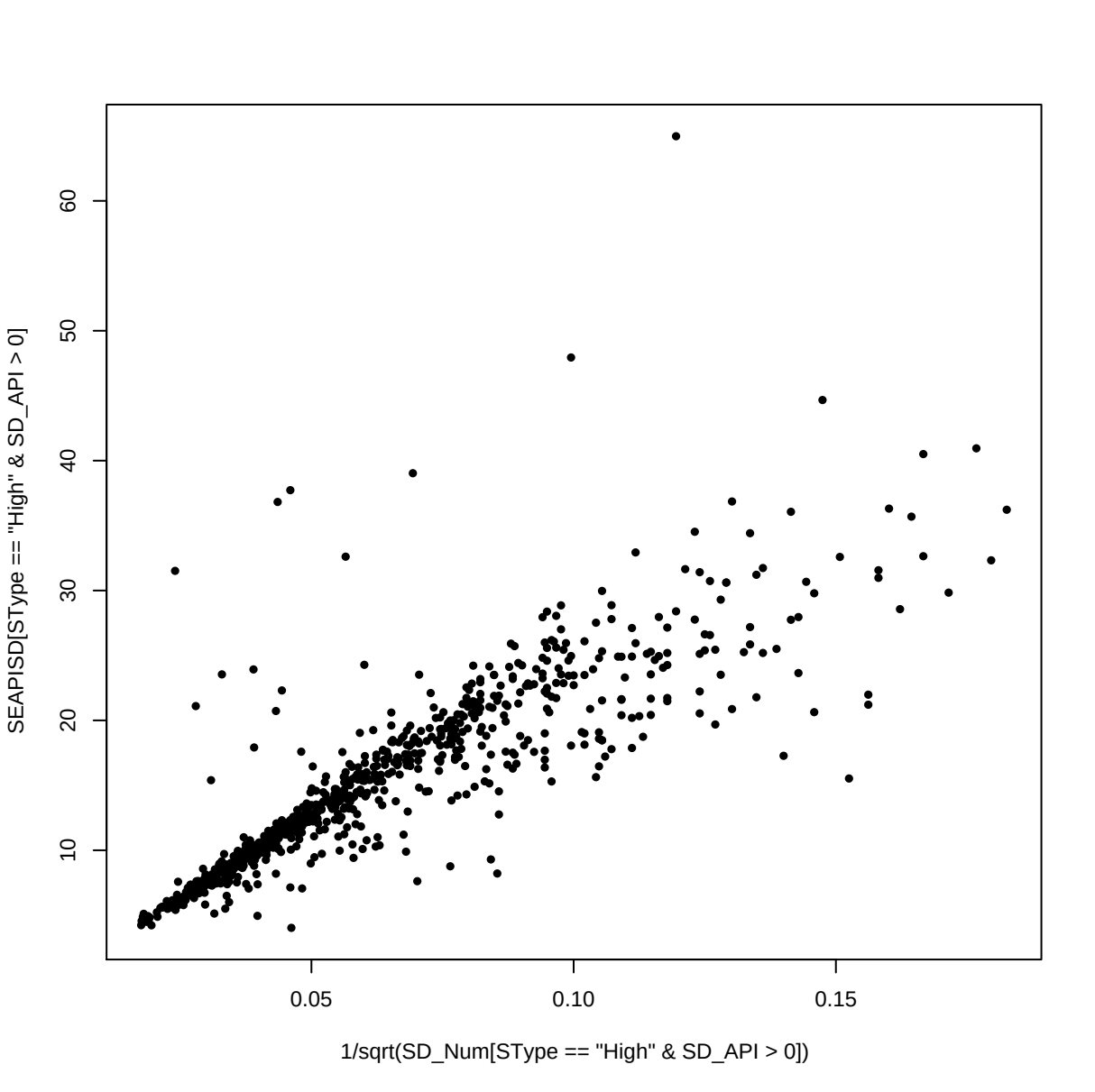

SEAPISD[SType == "Mid" & SD_API > 0]
10
20
30
40
400
500
600
700
800
900
APISD[SType == "Mid" & SD_API > 0]


SEAPISD[SType == "Mid" & SD_API > 0]
10
20
30
40
0.05
0.10
0.15
1/sqrt(SD_Num[SType == "Mid" & SD_API > 0])


SEAPISD[SType == "High"]
10
20
30
40
50
60
400
500
600
700
800
900
APISD[SType == "High"]


SEAPISD[SType == "High"]
10
20
30
40
50
60
0.05
0.10
0.15
1/sqrt(SD_Num[SType == "High"])


SEAPISD[SType == "High" & SD_API > 0]
10
20
30
40
50
60
400
500
600
700
800
900
APISD[SType == "High" & SD_API > 0]


SEAPISD[SType == "High" & SD_API > 0]
10
20
30
40
50
60
0.05
0.10
0.15
1/sqrt(SD_Num[SType == "High" & SD_API > 0])

References

California Department of Education, Policy and Evaluation Division. The 2003 Academic Performance Index Base Report: Information Guide (March 2004)
http://www.cde.ca.gov/ta/ac/ap/api0304.asp
http://www.cde.ca.gov/ta/ac/ap/documents/infoguide03b.pdf

California Department of Education, Policy and Evaluation Division. The 2003 Academic Performance Index (API) Base Report: Explanatory Notes
http://www.cde.ca.gov/ta/ac/ap/api0304.asp
http://www.cde.ca.gov/ta/ac/ap/expnotes03b.asp

Kane, T. J., and Staiger, D. O. (2002) "Volatility in School Test Scores: Implications for Test-Based Accountability Systems." Brookings Papers on Education Policy, 2002 (Washington, DC: Brookings Institution), 235-269.

Rogosa, D. R. Accuracy of API Index and School Base Report Elements. December 2002. California Department of Education website:
http://www.cde.ca.gov/psaa/apiresearch.htm

Rogosa, D. R. Year 2000 update: Accuracy of API Index and School Base Report Elements. December 2002. California Department of Education website:
http://www.cde.ca.gov/psaa/apiresearch.htm

Rogosa, D. R. Year 2001 update: Accuracy of API Index and School Base Report Elements. December 2002. California Department of Education website:
http://www.cde.ca.gov/psaa/apiresearch.htm

Rogosa, D. R. Irrelevance of Reliability Coefficients to Accountability Systems: Statistical Disconnect in Kane-Staiger "Volatility in School Test Scores". Stanford University, October 2002.
http://www-stat.stanford.edu/~rag/api/kscresst.pdf

Rogosa, D. R. (2005). Statistical Misunderstandings of the Properties of School Scores and School Accountability. In J. Herman and E. Haertel (Eds.) Uses and Misuses of Data in Accountability Testing. NSSE Yearbook, Vol. 104, Part 1. Chapter 7.