ON SOME DETERMINANTS OF SAT PERFORMANCE: A STATISTICAL ANALYSIS

(Proceedings of the Decision Sciences Institute Annual Meeting, Honolulu, Hawaii, November 1994)

RAJARAM GANA and ARTHUR E. HOERL
University of Delaware, Institutional Research, 325 Hullihen Hall, Newark, DE 19716. (302) 831-8946

ABSTRACT

Since the publication of the highly influential *Coleman Report* of 1966, researchers have tried to find inputs, such as school quality, that are associated with student achievement on standardized tests, but have found little [3]. This study finds an underlying association between Scholastic Aptitude Test scores and several measures of socio-ethnic inequality.

INTRODUCTION

This research examines state variations in Scholastic Aptitude Test (SAT) scores to address the empirical correctness of viewing economic inequality in an affluent nation, like America, as capability failure. The "capability approach" which originates in Aristotle's *Nicomachean Ethics* is discussed in [7].

HISTORICAL EVIDENCE

State-by-state breakdown of the percentage of potential military draftees who failed the mental test given by the United States's Selective Service System during 1966-70 was published by the U.S. Bureau of the Census. Most 18 year-old American males took this test. The aptitude test given draft candidates was an overly simplified one. An inability to pass the test would indicate a marginal expectation for success in the labor market. Walberg and Rasher [8] [9] examined state variations in the draft failure rates (DFR) to address the central problem of the *Coleman Report*: the relation of test results to educational, socio-economic, and demographic variables. Gana [2] reexamined the Walberg-Rasher study.

If state-by-state and year-by-year variations in DFR are modeled as a two-factor factorial design with a single replicate, then both factors have a significant effect on DFR. If this factorial design is used to model DFR over successively shortened spans of time (the last four, three, and two years, respectively), then both factors continue to have a significant effect on DFR. The factorial design diagnostics do not reveal any violations of model assumptions. These results indicate the heteroscedastic variability in DFR. Other members of the Box-Cox power transformation family, such as the square root, reciprocal, and logarithm of DFR, also display heteroscedastic variability. Furthermore, the estimated correlation between the mean and sample standard deviation of DFR over the last five, four, three, and two years by state are -0.71, -0.66, -0.46, and -0.38, respectively. The estimated correlation between the mean and sample standard deviation of the logarithm of DFR over the last five, four, three, and two years by state are 0.84, 0.85, 0.83, and 0.72, respectively.

The mean of DFR over the five years by state is the selected regressand. The sample variances of DFR over the five years by state are used to achieve homoscedasticity. If the five values of DFR in each state are considered as replicates, then the selected weights produce a weighted pure mean square error (MSE) of unity. This indicates the average estimated lower bound on the MSEs of generalized least squares estimated models explaining the regressand in high-dimensional space may be 1/5. A question of special interest to ask is how close to 0.2 can one bring MSEs of regression models explaining the regressand in order to control all influences from the demand for draftees.

The regressor selection procedure is not well-defined because it is difficult to select, *a priori*, a set of regressors explaining the regressand. Among several plausible sets of regressors competing to "explain" the regressand, the set which produces the smallest MSE is the one selected. For example, the regressors selected in [8] explain the regressand with MSE 1.6. Two regressors that are closely correlated may be exchangeable in the model, but one of them may decrease the MSE substantially more than the other and, hence, become a candidate for entry into the model. An entering regressor, selected by trial and error, may make some of the regressors selected at earlier stages insignificant. Multicollinearity may be a serious problem whenever a variance inflation factor (VIF) is greater than ten. Cook's distance and DFBETAS are used to assess model sensitivity.

Data published by the U.S. Bureau of the Census, U.S. Department of Housing and Urban Development, and U.S. Department of Education is used to model DFR. The selected regressors and the corresponding t-ratios (in parentheses) are shown below. For the resultant model, values of the MSE, F-ratio, and R^2 are 0.21, 114.73, and 0.974, respectively. No data values or variables are unreasonably influential. All VIFs are less than four. At this stage of the search for a DFR model, several other regressors are considered as potential entrants into the model. However, none of them are significant. Hence, the search is terminated.

1) Ratio of the number of all year-round housing units to resident population (-7.2).
2) Percent of occupied housing units with all plumbing facilities (-8.4).
3) Proportion of year-round housing units with air conditioning (-1.9).
4) Percent of one-family homes with average value of site per square foot $1.50 or more (2.9).
5) Average life expectancy of males (-6.2).
6) Ratio of Black male population 15 to 34 years old to resident male population (9.7).
7) Ratio of male population of Spanish origin to resident male population (3.1).
8) Public school pupils as percent of total average daily attendance of pupils transported at public expense (2.5).
9) Ratio of public elementary and secondary school enrollment in average daily attendance to resident population 5 to 17 years old (2.1).
10) Ratio of personal income per-capita to cost-adjusted public school expenditure (6.6).
11) Percent of males 18 to 24 years old in labor force (-2.9).
12) Annual average temperature based on the period 1941-70 (4.6).

High property values may reflect wealth concentration or maldistribution, or scarcity of affordable housing. Dropping this regressor from the model raises the MSE to 0.25. Economic deprivation among men, reflected by variations in male life expectancy [6], may influence their test performance. Ethnicity may influence test performance. Public schools and disparities in educational resources within states may influence student achievement on standardized tests. Members of the labor force might have had the basic skills necessary to pass the test. The correlation between the percent of homes with air conditioning and annual average temperature is 0.62. Dropping the temperature variable (which may reflect ecological factors) from the model makes the air conditioning variable insignificant, and raises the MSE to 0.32. These results have a point of contact with Christopher Jencks's work entitled *Inequality*.

NEW EVIDENCE

State-by-state breakdown of mean SAT scores during 1979-80, 1980-81, 1982-83, 1984-85, and 1985-86, was announced, in 1986, by the Educational Testing Service's College Entrance Examination Board (Princeton). The percent of graduates taking the SAT in 1982 by state was also announced. The sum of the mean verbal and mean mathematics SAT scores by state is the selected response. This response, and several of its transformations, display heteroscedastic variability. Hence, the mean of the response over the five years by state is the selected regressand. The sample variances of the response over the five years by state are used to achieve homoscedasticity. There is no correlation between the regressand and the sample variances of the response. The variability in the regressand is negatively correlated with the variability in the percentage of graduates taking the SAT, which has a value of 3 in South Dakota and a value of 69 in Connecticut. Hence, the percentage of graduates taking the SAT is considered a potential regressor in the model in order to control for the confounding effect the diversity of eligible students taking the SAT may have on state SAT averages [4].

The selected regressors and the corresponding t-ratios are shown below. For the resultant model, values of the MSE, F-ratio, and R^2 are 3.10, 84.22, and 0.965, respectively. No data values or variables are unreasonably influential. All VIFs are less than five. The regressors measure general population characteristics rather than test-taking population characteristics. Hence, the MSE may be relatively distant from 0.2. If the DFR and SAT models are reestimated using ordinary least squares, then no unreasonable changes occur in the levels of significance of the regressors. The resultant values of the R^2 are 0.97 and 0.95, respectively. Hence, the selected weights, estimated nonparametrically from the data based on low degrees of freedom, do not introduce unnecessary variability into the models.

1) Percentage of graduates taking the SAT (-17.80).
2) Ratio of families below poverty level to families earning $25,000 to $35,000 (2.67).
3) Ratio of White population 15 to 24 years old to resident population (6.90).
4) Ratio of American Indian, Eskimo, and Aleut population to resident population (4.64).
5) Ratio of Asian Indian population to resident population (5.29).
6) Ratio of population of Spanish origin 15 to 24 years old to resident population (-3.22).
7) Ratio of private high-school graduates to resident population (3.67).
8) Average life expectancy of females (6.89).
9) Pupils in national school lunch programs (-2.75).
10) Percent of year-round housing units with public or private water system (2.79).
11) Percent of occupied housing units lacking complete plumbing facilities for exclusive use (-3.15).
12) Inverse proportion of 16 to 19 year-old students in labor force (2.62).

As the number of families below poverty level increases relative to the number of families earning between $25,000 and $35,000, there may be less diversity in the pool of test-takers. Ethnicity may influence SAT performance. Private schooling may have a positive impact on SAT performance. Economic deprivation among women, reflected by variations in female life expectancy, may influence their SAT performance. Economic deprivation among students in school lunch programs may influence their SAT performance. Teenage students in the labor force may be unable to allocate sufficient time to assimilate material taught in school. These results have a point of contact with the work of Crouse and Trusheim [1].

The fitted studentized residuals for the model may measure inequalities in capabilities among states. State comparison of SAT scores may have important political and economic consequences. The role of the SAT in capability accounting is to take note of some fundamental diversity that might be present in the complex idea called "humanity". The model indicates raw SAT scores may measure capability inequality and economic inequality. Hence, comparing raw SAT scores may not be "justice as fairness" because unequal holdings of "primary goods" may have a confounding effect on the SAT as an identifier of unequal mental capabilities [5]. From the capability perspective it may be arguable that the SAT as a measure of inequalities in mental capabilities (assuming equal holdings of primary goods) should account for a person's difficulties in converting primary goods into SAT performance. This may become an important issue if SAT scores are viewed as a means for increasing an individual's freedom of choice of educational opportunities. Results indicate aptitude testing may conflict with certain long-championed aspects of the complex idea called "equality of opportunity". This study offers empirical evidence in support of the wisdom of focusing on economic inequality as lack of freedom for American teenagers to achieve.

REFERENCES

[1] Crouse, J.H. & Trusheim, D.W. *The case against the SAT*. Chicago, Illinois: The University of Chicago Press, 1988.

[2] Gana, R. *Improvement of the draft failure rate regression model of Walberg and Rasher*. Newark, DE: University of Delaware, 1989.

[3] Hanushek, E.A. The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, 1986, 24, 1141-1177.

[4] Powell, B. & Steelman, L.C. Variations in state SAT performance: meaningful or misleading?. *Harvard Educational Review*, 1984, 54, 389-412.

[5] Rawls, J. *A theory of justice*. Cambridge, Massachusetts: Harvard University Press, 1971.

[6] Sen, A.K. The economics of life and death. *Scientific American*, May 1993, 40-47.

[7] Sen, A.K. *Inequality reexamined.* Cambridge, Massachusetts: Harvard University Press, 1992.

[8] Walberg, H.J. & Rasher, S.P. Public school effectiveness and equality: new evidence and its implications. *Phi Delta Kappan*, 1974, 66, 3-9.

[9] Walberg, H.J. & Rasher, S.P. Improving regression models. *Journal of Educational Statistics*, 1976, 1, 253-277.