DOCUMENT RESUME

ED 120 230 TM 005 189

AUTHOR Brager, Gary L.
TITLE Outliers and Accountability, Fact or Fiction?
PUB DATE Apr 76
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS *Achievement Rating; Comparative Analysis; *Educational Accountability; Educational Assessment; *Educational Quality; Formative Evaluation; *Multiple Regression Analysis; Predictor Variables; *Schools; Validity
IDENTIFIERS *Statistical Outliers

ABSTRACT

The present paper sought to investigate the usefulness to accountability of identifying outlying schools through multiple regression. The outlier approach identifies schools which differ significantly for reasons other than variation in the predictors. The present study was not able to confirm the construct validity of the outlier approach for the author's school system. That is, educational process variables (observations, interviews, school size, staff ratios, ability scores, etc.) did not discriminate the positive from the negative outliers. Moreover, the study showed the outlier approach may be merely a tautology (i.e., 5% of the schools are outliers at the 5% level of significance.) (Author)

ED120230

OUTLIERS AND ACCOUNTABILITY
FACT OR FICTION?

by: Gary L. Brager, Ph.D.
Office of Educational Research
Baltimore County Public Schools
Baltimore County, Maryland
January 6, 1976

TM005 189

OUTLIERS AND ACCOUNTABILITY: FACT OR FICTION?

Recently, the use of outliers to identify exceptionally achieving schools has been proposed or noted in several accountability systems (e.g., Klitgaard, 1973; Maryland State Department of Education, 1975). Essentially, an outlier is a data element extremely distant from the central tendency of its distribution. "Extremely distant" is typically defined in much the same way as "statistical significance"; often the same percentiles are used as cutoff points. Detection of outliers in industrial quality control is commonplace (e.g., Zwickl, 1975); detection of outlying schools in educational accountability is emerging (Cooperative Accountability Project, 1975).

In the Maryland Accountability Program (MAP), outliers are located in the following manner. School mean achievement scores are regressed upon a set of predictors (e.g., school mean ability scores, median family income). Residuals are calculated for each school. Any school whose absolute residual is larger than a given value is identified as an outlier.

Approximately ± 1.96 standard deviations from the mean of the residual distribution (zero) are used as the cutoff points, yielding the $2\frac{1}{2}$ and $97\frac{1}{2}$ percentile points assuming normality, respectively.

Supposedly, a school in the upper tail of the residual distribution (termed "positive outliers") is exhibiting average achievement significantly greater than expected (or predicted), based upon its average ability, socioeconomic levels or other variables. Inversely, a school labeled as a "negative outlier" has significantly lower average achievement than expected.

The regression process works much as a blocking procedure, wherein schools are grouped by abilities (or socioeconomic levels). Extreme schools, if any exist within a block, are then identified as outliers.

Once schools are denoted as outliers, the next step in the Maryland Accountability Plan is a process analysis of observable educational variables for attribution of the test results. Though there are several statistical problems with the use of regression analysis (e.g., correlated predictors (Luecke and McGinn, 1975), reliability of the data (Stanley, 1971)), the present paper is concerned only with a logical analysis of regression and outliers in accountability.

It is the author's contention that the use of regression to identify outliers may constitute a tautology, not related at all to process variables of educational significance, or to improving instruction. (Please note that this paper is not a proven theorem; it consists of thoughts and concerns which hopefully will provoke reflection and consideration about accountability assessment).

Let us examine an example of test score data from an ESEA Title I class. Referring to Table 1, we have, for twenty-five pupils, their reading score (the predictor), math score (the criterion), predicted math score (the point on the regression line $(X,\hat{Y})$ and the residual (the algebraic difference between Y and $\hat{Y}$. Some common facts about regression lines are apparent from the table: the mean of the regression residuals is zero, since $\hat{Y} = E(Y|X)$; the mean of the predicted math scores is equal to the mean of the actual math scores; the standard deviation of the predicted math scores is smaller than the standard deviation of the actual math scores, since some of the error variance has been reduced by predicting; the standard deviation of the residuals is the well-known standard error of estimate.

According to the Maryland Accountability Program, any data element (in this case, pupils), whose absolute residual is larger than 1.96 standard errors of estimate is an outlier. Pupil 23 is the outlier in this example.

Figure 1 graphs the (X,Y) data points and the regression line $\hat{Y} = E(Y|X)$. Figure 2 shows a histogram of the residuals. Even for only 25 cases, the histogram is remarkably mound-shaped.

Now, where does the claim of a tautology surface? If our cutoff points encompass five percent of the residual distribution, we would expect to identify five percent of the data elements as outliers. In the above example, we would expect one outlier; we obtained one outlier (pupil 23).

The Baltimore County School System has 108 elementary schools. In the 1974 accountability assessment, 6 were outliers in reading (we would expect 5.4); in 1975, 5 schools were outliers (two schools were outliers both years).

Notice also, in the example, that ± 1.0 standard errors of estimate from zero enclose approximately two-thirds of the residuals, a result entirely predicted from normal curve theory.

Thus it seems plausible, that identifying outlying residuals from the data upon which the residuals are calculated may be no more than a classical Type I error. Is it the case that locating a school in the tails of the residual distribution is neither necessary nor sufficient for observing educationally significant processes? (See also Duncan, 1974)

Another difficulty in explaining the use of outliers in accountability is that a positive outlier school may have <u>lower</u> average achievement than a negative outlier school. Suppose we encounter a school whose average ability is 110, and average achievement in the fifth grade is 5.3. Another school has an average ability of 95 and average achievement of 4.5. The first school has significantly low achievement for its ability, the second school has significantly high achievement for its ability. However, would observations of the processes in the second school be expected to help the first school? How do we account for error in the tested ability? How do we transfer processes between two schools so unalike?

Certainly, if we had two schools of <u>equal</u> ability, one of which was a positive outlier and the other a negative outlier, we might expect to profit by a process evaluation. But the present regression method does not necessarily identify such pairs.

It would seem that the process evaluation is necessary to validate the statistical identification procedures, rather than to explain the results.

Although the Maryland Accountability Program process evaluation has not yet occurred, the present author has conducted three such similar evaluations on a much smaller scale in the Baltimore County schools. In each case, the process evaluation failed to confirm the construct validity of the statistical identification. Of course, it seems imperative that the process evaluators not know which schools are categorized as positive or negative outliers, or the results are obvious (this "blind" evaluation appears close to Scrivea's goal-free evaluation).

In the first study, instructional supervisors observed schools some of which had recorded outlying gains in achievement (positive and negative), some of which had not. The data were collected using narratives and interviews. No differences among the positive and negative outliers were observed. The second study involved Title I schools which were outliers. Project staff observed in these schools but were unable to discern which schools were the outliers. The third study also involved Title I schools, but the observers were not on the county staff and some used formal observational scales. Once again, the search for processes was not fruitful.

In addition, a glance at seven background variables for the three positive and three negative outliers identified in the 1974 Maryland accountability testing reveals no differences in their average enrollments, average abilities, pupil/staff, average years staff experience, percent of staff with advanced degrees, percent disadvantaged, and median family income (1970 dollars). If there are other educational variables which explain why three schools' achievements are significantly higher than their abilities predict and three are lower, they are not encumbered with the school resource or input variables usually encountered.

Of course, lack of power in the process evaluation to detect the causes of the statistical differences is a plausible explanation, but so is a Type I error. That is, the outliers are part of the error variance of the regression analysis and not of any educational significance.

Several future courses may be worthwhile to pursue. The present author welcomes comments, criticisms and suggestions.

(1) Consider schools for process evaluation only if they are outliers at least two years in a row. (e.g., Klitgaard, 1973)

(2) Consider schools for process evaluation only if they are outliers on most of the subtests of the test being used.

(3) Use a random sample of schools not statistical outliers in the process evaluation as well as those schools so identified as outliers.

(4) Try to develop a theoretical equation against which to test for outliers (possibly a given year's regression equation--sort of a norming process), rather than using the data to generate their own outliers.

(5) Keep the process evaluation a "blind" experiment relative to the statistical results.

(6) Don't count the number of outliers compared to an expected number, unless the distribution of residuals can be assumed normal (or it's parameters known)

(7) Consider the use of blocks or groups of schools alike in ability (or whatever covariate is used) rather than regression as a tool to locate unusually achieving schools. Blocking is far easier to explain to a school PTA than is regression analysis.

The present paper has discussed a concern about using regression analysis to identify outlying schools in pursuing accountability assessment.

GLB:w
1-7-76

REFERENCES

Duncan, A. J., "Some Recent Developments Related to MIL.-STD.-414". (paper presented at ASQC Technical Conference, Boston, May 20, 1974.)

Klitgaard, R. E., and Hall, G. R., A Statistical Search for Unusually Effective Schools. Santa Monica, California: Rand Corporation, 1973.

Luecke, D. and McGinn, N. F. "Regression Analyses and Education Production Functions: Can They Be Trusted". Harvard Educational Review. 45(3), 1975, 325-350.

Maryland State Department of Education, Maryland Accountability Program Report, School Year 1973-74. Baltimore, Maryland: Maryland State Department of Education, 1975.

Stanley, J. C., "Reliability" Chapter 13 of Educational Measurement, Ed. R. L. Thorndike. Washington, D. C.: American Council on Education, 1971.

Tukey, J. W. Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley Publishing Co., 1970.

Zwickl, R. D., "Acceptance Sampling with ESP". Lecture given at Sampling Seminar of the American Society for Quality Control, Baltimore, Maryland, March 21, 1975.

G. L. BRAGER

Table 1. LINEAR REGRESSION & RESIDUALS

EXAMPLE: PREDICTING MATH SCORE FROM READING SCORE

Pupil	READING SCORE	MATH SCORE	PREDICTED MATH SCORE	RESIDUAL
1	2.5	2.9	2.86	+.04
2	2.9	3.7	3.04	+.66
3	3.7	3.8	3.40	+.40
4	2.0	2.1	2.63	-.53
5	2.7	2.7	2.95	-.25
6	3.6	3.6	3.35	+.25
7	3.6	2.7	3.35	-.65
8	3.2	4.0	3.17	+.83
9	3.0	2.5	3.08	-.58
10	2.7	3.0	2.95	+.05
11	4.0	4.0	3.53	+.47
12	3.6	2.3	3.35	-1.05
13	1.8	2.9	2.54	+.36
14	2.3	2.9	2.77	+.13
15	2.9	3.7	3.04	+.66
16	3.2	2.9	3.17	-.27
17	2.5	2.7	2.86	-.16
18	3.5	3.8	3.31	+.49
19	2.7	3.9	2.95	+.95
20	3.8	3.5	3.44	+.06
21	2.9	2.6	3.04	-.44
22	3.0	2.4	3.08	-.68
23	3.3	2.0	3.22	-1.22
24	3.0	3.9	3.08	+.82
25	2.3	2.3	2.77	-.47
MEAN	2.99	3.07	3.07	0.00
S.D.	0.56	0.64	0.25	0.59

Table 2. $\boxed{\text{SUMMARY STATISTICS}}$

REGRESSION EQUATION:

PREDICTED MATH SCORE = 1.73 + (0.45)(READING SCORE)

CORREATIONS:

READING SCORE WITH MATH SCORE	.39
PREDICTED MATH SCORE WITH RESIDUALS	.00
READING SCORE WITH RESIDUALS	.00

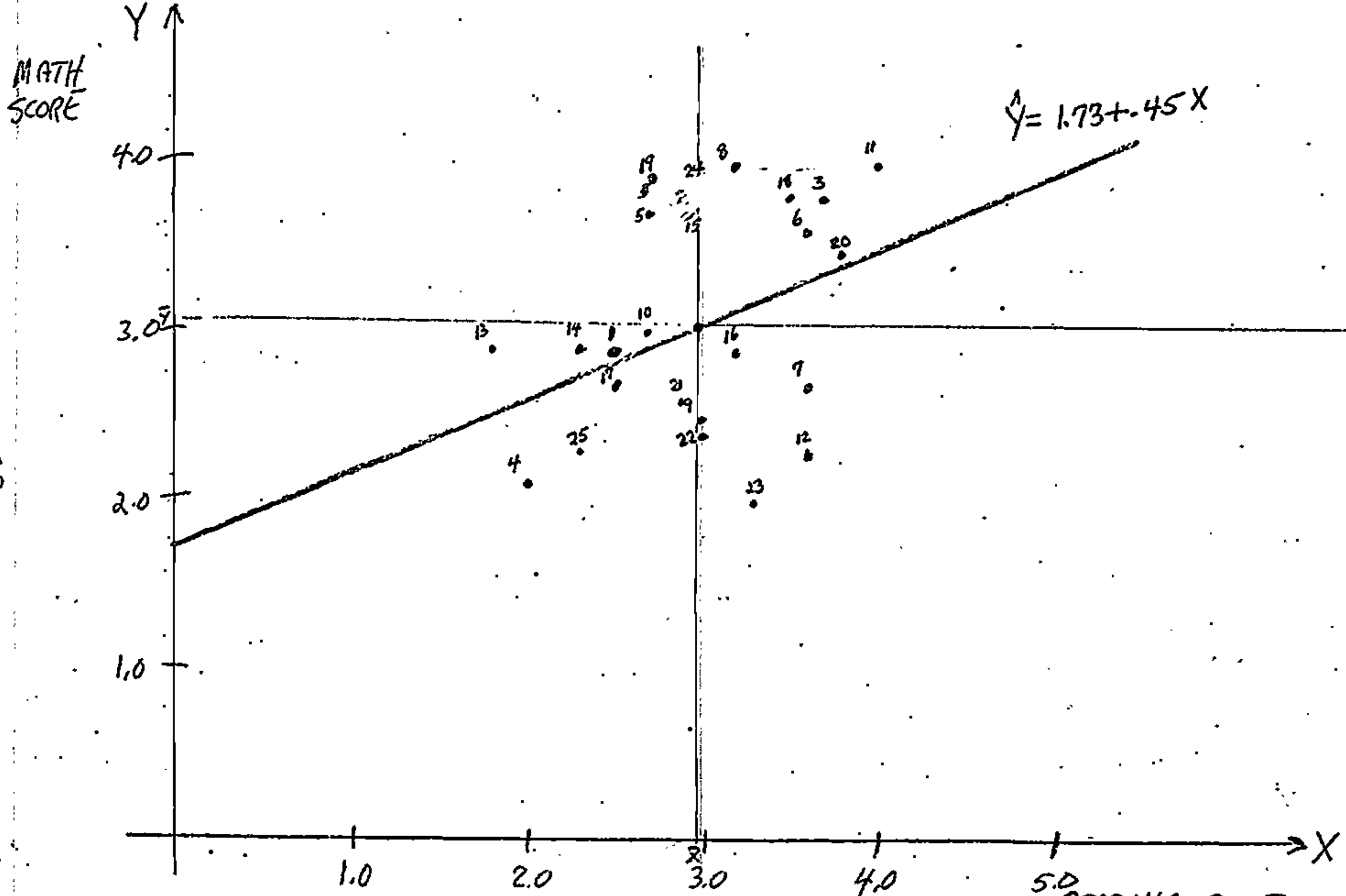
STANDARD ERROR OF ESTIMATE

$$S_{y \cdot x} = S_y\sqrt{1-r^2} = .64\sqrt{1-.39^2} = .59 = \text{S.D. of residuals.}$$

NOTE: INTERCEPT OF REGRESSION LINE, $\overset{a}{\cancel{b}}, = \bar{Y} - b\bar{X}$

$$= 3.07 - (.45)(2.99) = 1.73$$

CORRELATION, $r = b \cdot \frac{S_x}{S_y} = (.45)\left(\frac{.56}{.64}\right) = .39$



F I G 1. REGRESSION LINE FOR EXAMPLE

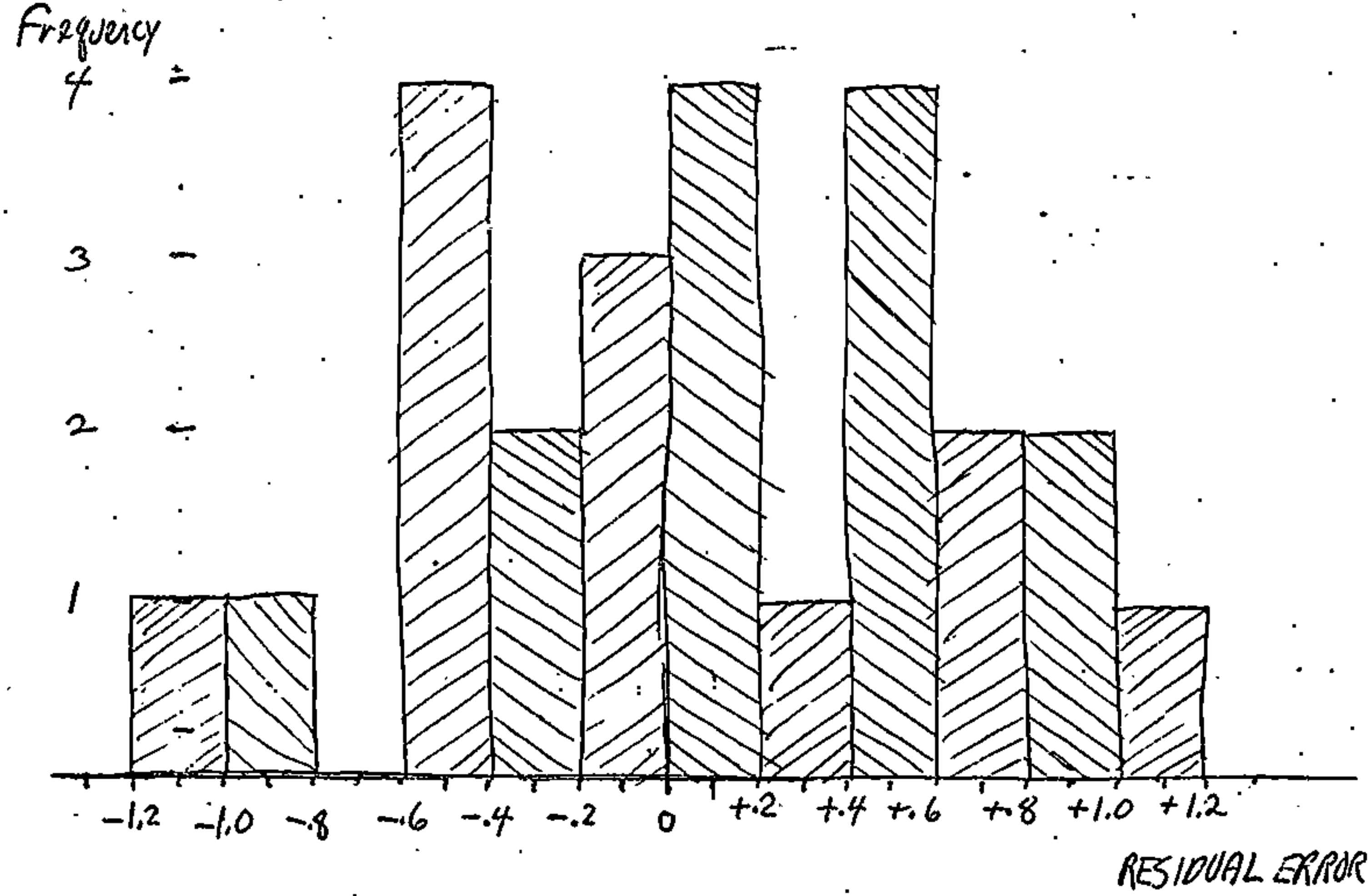


FIG 2. RESIDUAL DISTRIBUTION FOR EXAMPLE