DOCUMENT RESUME

ED 160 632 TM 007 701

AUTHOR Hick, Thomas L.; Irvine, David J.
TITLE An Analysis of the Historical Regression Method of Predicting Posttest Grade Equivalents for Categorically-Aided Programs.
PUB DATE Mar 78
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Gains; Elementary Education; *Grade Equivalent Scores; *Mathematical Models; Maturation; *Post Testing; *Predictive Measurement; Pretesting; *Program Length; Statistical Analysis; Weighted Scores
IDENTIFIERS Least Squares Analysis

ABSTRACT

To eliminate maturation as a factor in the pretest-posttest design, pretest scores can be converted to anticipate posttest scores using grade equivalent scores from standardized tests. This conversion, known as historical regression, assumes that without specific intervention, growth will continue at the rate (grade equivalents per year of schooling) obtained at the time of pretest. Data were taken from reports of 213 Title I compensatory education programs in New York State to examine the predictive ability of the historical regression model. The approach was to: (1) express historical regression in algebraic terms; (2) produce a linear model with assigned weights from the algebraic formula; (3) produce a least squares historical regression model whose weights best fit the data; (4) compare the historical regression model with the least squares model; and (5) develop an alternative model. When compared with program-level data, historical regression underestimated final achievement for short programs with older children. It overestimated for younger children in long programs. An alternative method was developed which eliminated the bias, removed half of the error, and eliminated much computation since an expected achievement level for each child was not required. (Author/CP)

AN ANALYSIS OF THE HISTORICAL REGRESSION METHOD OF PREDICTING POSTTEST GRADE EQUIVALENTS FOR CATEGORICALLY-AIDED PROGRAMS

Thomas L. Hick
Assistant in Education Research
David J. Irvine
Coordinator
Prekindergarten Evaluation Unit
New York State Education Department
Albany, New York 12234

A paper prepared for presentation
at the Annual Meeting of the
American Educational Research Association,
Toronto, Canada, March 27-31, 1978.

Printed in U.S.A.

Abstract

An Analysis of the Historical Regression Method of Predicting Posttest Grade Equivalents for Categorically-Aided Programs

THOMAS L. HICK and DAVID J. IRVINE

New York State Education Department

Historical Regression follows directly from the assumption that, without specific intervention, growth will continue at the rate (grade equivalents per year of schooling) obtained at the time of the pretest. When compared with program-level data (n = 213) it was found that Historical Regression underestimated final achievement for short programs with older children. It overestimated for younger children in long programs. An alternative method was developed which eliminated the bias, removed half of the error, and eliminated much computation since an expected achievement level for each child was not required.

An Analysis of the Historical Regression Method of Predicting Posttest Grade Equivalents for Categorically-Aided Programs

T. L. HICK and D. J. IRVINE
New York State Education Department

T. L. Hick
Prekindergarten Evaluation Unit
State Education Department
Room 562 EBA
Albany, New York 12234

Objectives

To eliminate maturation as a factor in the pretest-posttest design at least one State (New York) recommends a conversion of the pretest to anticipate posttest scores when data are in grade equivalents from standardized tests. This conversion is known as Historical Regression, and follows directly from the assumption that, without specific intervention, growth will continue at the rate (grade equivalents per year of schooling) obtained at the time of pretest.

The District Evaluator's Handbook describes the procedure for obtaining anticipated scores by the following steps:

- Step 1. Obtain each pupil's pretest grade equivalent.
- Step 2. Subtract 1 (since most standardized tests start at 1.0).
- Step 3. Divide the figure obtained in step 2 by the number of months the pupil has been in school to obtain a hypothetical (historical regression) rate of growth per month. (Ignore kindergarten months. One school year = 10 months.)
- Step 4. Multiply the number of months of Title I treatment by the historical rate of growth per month.
- Step 5. Add the figure obtained in step 4 to the pupil's pretest grade equivalent (step 1).*

This paper examines the Historical Regression method for obtaining predicted posttest grade equivalents to determine its adequacy as a predictive model and to develop an alternative predictive model.

Method

The approach was to (1) express Historical Regression in algebraic terms, (2) from the algebraic formula, produce a linear model for

*Bureau of Urban and Community Programs Evaluation. District Evaluator's Handbook of Selected Evaluation Procedures for Categorically Aided Programs Serving Disadvantaged Learners. Albany: New York State Education Department, Spring 1972.

Historical Regression (this model has assumed weights), (3) produce a Least Squares Historical Regression model that has weights which best fit the data, (4) compare the Historical Regression model with the Least Squares Historical Regression model to determine similarities and differences, (5) devel an alternative model.

Data Source

The data were taken from ne reports of Title I Compensatory Education programs filed at the New York State Education Department for the 1972-73 school year. This file contained the following information for 213 programs:

Y - Program mean reading grade equivalent on posttest.
B - Program mean reading grade equivalent on pretest if over grade equivalent of 2.
D - Duration of program in years.
T - Mean previous time spent in school in years.

Characteristics of these variables are reported in Table 1.

Table 1
Means, Standard Deviations and Limits

Variable	M	SD	Minimum	Maximum
Y	4.67	1.34	2.10	9.2
B	3.83	1.23	2.03	7.7
D	.90	.32	.20	1.6
T	6.69	5.72	1.00	11.0

Is the Historical Regression algorithm adequate?

In terms of the defined variables the algebraic expression of the Historical Regression was found to be:

$$Y = B + (B - 1) \times (D/T)$$

In the form of a linear model this Historical Regression Model became:

Model 1. $Y = \text{Zero } U + 1\,(B) + 1\,(B \times D/T) - 1\,(D/T) + E_1$

The weights on the variables were dictated by the Historical Regression procedure. A model with the same variables used in the Historical Regression model but with the weights left for a least squares best fit was expressed as a Least Squares Historical Regression Model:

Model 2. $Y = a_0 U + a_1 (B) + a_2 (B \times D/T) + a_3 (D/T) + E_2$

The solution to the Least Squares Historical Regression model yielded: $a_0 = .68$, $a_1 = .99$, $a_2 = .59$ and $a_3 = -1.02$. The Historical Regression Model assumes that these weights are: $a_0 = 0$, $a_1 = 1$, $a_2 = 1$, $a_3 = -1$. The error sum of squares for the Least Squares model is 43. Upon substitution of the historical regression values, the error sum of squares for the restricted model is 85. With 5 degrees of freedom for the full model and none for the restricted model, the difference between the error sum of squares is tested with F at 5/208 degrees of freedom. The F of 38 is associated with a near zero probability of a chance difference.

Therefore, the Historical Regression Model (1) does a poorer job of predicting posttest grade equivalents than the Least Squares Historical Regression Model (2).

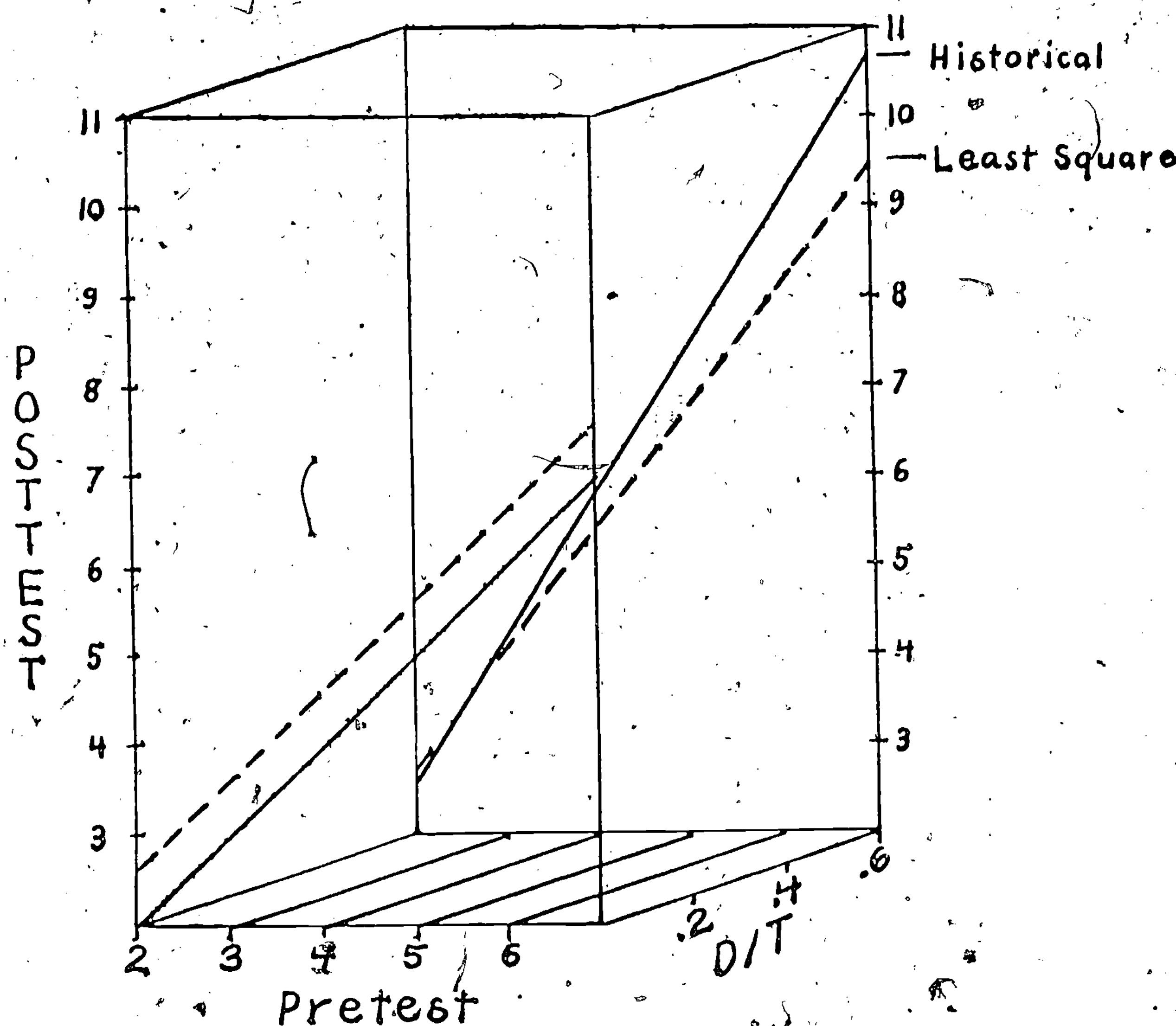
Since the Historical Regression deviates from a Least Squares best fit, Historical Regression was plotted as shown in Figure 1. It may be observed that initial grade equivalents are along the horizontal axis while the ratio of the program duration to the average time the students had previously spent in school is given on the right side of the cube. The historical regression plane twists so that the relationship of pretest to posttest is 1 to 1 as the ratio D/T approaches zero. This would be the situation of a very short program (e.g., 2 months) for children that had spent a long time in school (e.g., 50 months) yielding a ratio of D/T of .04, a value very close to zero. In this situation, according to historical regression, the children will have the same score at the end of program as at the beginning.

At the other end of the D/T ratio, the situation is shown where the program has lasted as long as the child has previously been in school. This would be the case for young children. At an initial grade equivalent of 2, the final achievement would be approximately 2.4 grade eq[illegible] At an initial grade equivalent of 3, the final achievement would be approximately 4.0 grade equivalents.

This historical model does fit the data quite well covering 88 percent of the variability in posttest scores, provided it is centered and twisted properly. As previously shown, however, the historical regression model does not match the least squares fit to the data.

Figure 1

TWO TECHNIQUES FOR ESTIMATING POSTTEST SCORES



A Detailed Examination of the Historical Regression Algorithm.

Each of the Historical Regression weights was tested in turn with either the hypothesized weight or the least squares weight retained in accord with the result of testing the historical regression weight. First, the $a_0 = 0$ hypothesis was tested. This hypothesis is that the regression plane passes through zero on the posttest when the pretest is zero and the D/T ration is zero. This turned out to not be the case. The least squares regression plane does not pass through zero at a very high level of probability ($F = 476$ at 1/209 degrees of freedom). The least squares value elevates the plane .68 grade equivalents above zero. The second hypothesis was that $a_2 = 1$. This weight describes the amount of twist in the plane. The weight for twist was not one ($P = 0$). A weight of .59 fits the data much better so that the twist is not as large as assumed by historical regression. The third hypothesis was that the tilt of the plane was 45 degrees when D/T was zero. This was the hypothesis that a_1 was equal to one. Test of this hypothesis failed to reject it, so that a value of one for a_1 was an acceptable fit by least squares ($P = .28$). The last hypothesis that a_3 is -1, also was not rejected ($P = .37$).

The final model is:

$$Y = .68 + B + \frac{((.59B) - L) \times D}{T}$$

The differences between this equation and the algorithm of historical regression are the addition of .68 to all projected scores and the multiplication of the pretest score by .59 before adjusting it for the beginning at grade 1.

Examination of the difference between the Historical Regression plane in Figure 1 and the least squares best fit plane reveals that the final achievement of older students in short duration programs has been underestimated by as much as 1/2 a year. Younger students in long programs did not fare as well. The effect was to overestimate their achievement by as much as one year, giving the appearance of poor performance.

Development of an Alternative Model

It would seem desirable to simplify the formulation of the expected achievement without losing the predictive power of historical regression. An attempt at this goal was undertaken beginning with this possible model:

$$Y = a_0U + a_1B + a_2D + a_3B^2 + a_4D^2 + a_5B^3 + a_6D^3 +$$
$$a_7B \times D + a_8B^2D + a_9BD^2 + a_{10}B^2D^2 + E_3$$

The R^2 for the above model was .896 and is obviously significant. The hypothesis of no curvilinear interaction, was rejected using Model 4:

$$Y = a_0U + a_1B + a_3B^2 + a_4D^2 + a_5B^3 + a_6D^3 + a_7BxD + E_4$$

which yielded an R^2 of .890. The difference between R^2's was significant ($F = 4.12$, df 3/202, $P = .008$); therefore curvilinear interaction was considered as present. The task of locating the interaction remained. The hypothesis that the interaction was very complex was tested with model 5 which was the same as model 3 except for the last term B^2D^2. Complex interaction was considered not present ($F = 4.57$, df = 1/202, $P = .03$) as all relationships are tested at the .01 level of significance. Therefore, model 5 became the full model for a test of the hypothesis that the interaction is curvilinear on program duration. Model 6 expressed this hypothesis:

$$Y = a_0U + a_1B + a_2D + a_3B^2 + a_4D^2 + a_5B^3 + a_6D^3 + a_7BxD + a_8B^2D + E_6$$

This model turned out an R^2 of .889 which was significantly different from model 5 at the .007 level of significance ($F = 7.51$, df = 1/203). Therefore, the term BD^2 was considered significant and included in the next model (7) that was used to test the effect of B^2D. This term was not considered significant ($R^2 = .892$, $F = 2.64$, df 1/203, $P = .10$).

To regain our bearings, the full model is now (#7) which includes third degree polynomial forms:

$$Y = a_0U + a_1B + a_2D + a_3B^2 + a_4D^2 + a_5B^3 + a_6D^3 + a_7BxD + a_9BD^2 + E_7$$

Comparison of this model with a model that lacked the third degree polynomials showed that B^3 and D^3 were not significant ($F = .80$, df 2/204, $P = .45$). The only remaining term not involved in the established interaction was B^2. It was found to be nonsignificant ($F = 1.38$, df = 1/206) and dropped from the final equation.

All the remaining terms were involved in the expression of the interaction and were therefore considered necessary for the expression of the relationship of beginning scores and program duration on posttest scores. The final acceptable model is:

$$Y = -2.14 + 1.56B + 7.26D - 4.04D^2 - 1.39BxD + .80BD^2 + E_8$$

This is a highly complex plane as it describes a different effect for program duration at various levels of the pretest scores. It should be noted that program duration, although highly significant, has an only very subtle relationship with posttest scores after initial skill level is included. This plane, described by the final acceptable model, is shown in Figure 2.

Conclusion

A method for eliminating maturation as a factor in pretest-posttest designs for Title I programs was examined in relation to data and an alternative predictive model was developed. The alternative method: (1) makes no assumptions regarding the relationship between pretest level and program duration, (2) does not require the computation of time span in school for each child, (3) does not require an expected achievement level for each child, (4) does not bias against some programs, (5) cuts the error in half, (6) requires only the beginning mean achievement level.

FIGURE 2

LEAST SQUARES MODEL FOR USING PRETEST LEVEL AND PROGRAM DURATION TO ESTIMATE FINAL ACHIEVEMENT FOR THE PROGRAM