DOCUMENT RESUME

ED 174 672	TM 009 634
AUTHOR	Raffeld, Paul; And Others
TITLE	A Simulation Study of the Effectiveness of Two Estimates of Regression in the Title I Model A Procedure.
PUB DATE	Apr 79
NOTE	16p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Achievement Gains; Critical Path Method; Elementary Education; *Evaluation Methods; Expectation; *Mathematical Models; Norm Referenced Tests; *Post Testing; Program Effectiveness; Research Design; Simulation; Standard Error of Measurement; *Statistical Bias
IDENTIFIERS	Elementary Secondary Education Act Title I; *Normal Curve Equivalent Scores; *RMC Models

ABSTRACT

The RMC Model A (norm-referenced) for evaluation of Title I programs is based upon the equipercentile assumption--that students maintain their percentile rank over a one-year period, provided that no special instrucional intervention is introduced. The control group, essentially the sample used to standardize the achievement test, represents the no-treatment expectation. The treatment effect is defined as the average posttest performance minus the average pretest performance. Title I participants are usually selected based upon low pretest scores, and may bias the results of Model A evaluation because low ability students generally regress toward the mean without any intervention. Two models which could estimate the no-treatment expectation and which might solve these problems were investigated: (1) Murray's recursive path analytic paradigm, encompassing three achievement measures, a group factor, an individual differences within group factor, and an error component; and (2) an extension of the one independent variable regression situation. Results indicated that for local districts with fewer than 500 students per grade, the no-treatment expectation would probably be overestimated. These two methods would be acceptable for use in districts with 2,000 or more students per grade. The use of local correlations, based on non-Title I participants, was also suggested. (GDC)

ED174672

A Simulation Study of the Effectiveness of Two Estimates of Regression in the Title I Model A Procedure

Paul Raffeld
David Stamman
George Powell
Resource Development Institute, Inc.
Title I Evaluation Technical Assistance Center

April 1979

Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California April 8-13, 1979.

TM009 634

As a result of the 1974 Congressional amendment of the law governing ESEA Title I programs, RCM Research Corporation developed a set of alternative evaluation models intended to provide a standardized framework from which to judge the effects of the Title I effort at the federal level. Each of the models was presented to the United States Office of Education (USOE) as a means of obtaining an unbiased estimate of the no-treatment expectation, provided that the assumptions of each model could be adequately satisfied. The three models developed by RMC Research Corporation are the Control Group Model (Model B), the Special Regression Model (Model C) and the Norm Referenced Model (Model A).

The intent for all three models is to provide an unbiased estimate of the Title I treatment effect which RMC has defined as

Treatment effect = Obtained Posttest performance - Expected Posttest performance.

The models differ primarily in the way in which the "expected posttest performance" is obtained. RMC defines this "expected posttest performance" or "no-treatment expectation" as the performance of Title I students who had not received any special instructional intervention.

In this paper we focus on Model A, the Norm Referenced Model because it is the most popular model among LEAs at the present time, and because it appears to be the model most prone to biased estimates of the no-treatment expectation. The major assumption underlying Model A is the "equipercentile" assumption. The basic concept is that students will maintain their percentile rank over a one year instructional period provided that no special instructional intervention is introduced.

The control group of Model A is essentially the norm group used for a particular standardized achievement test, since the national percentiles are obtained from this source. Furthermore, since the pretest percentiles are, by definition, the no-treatment expectation under the equipercentile assumption, the Title I treatment effect for Model A is simply defined as the average posttest performance minus the average pretest performance.

One of the major disadvantages of Model A is that Title I participants are generally selected on the basis of poor academic performance. And we know that low scoring students are likely to regress towards the population mean on a second measurement even with no special instructional intervention. In an effort to avoid this problem, Tallmadge (1976) recommended that selection of Title I participants be based upon some measure other than the pretest. The assumption that Tallmadge (1976) made is that all of the regression that is going to occur would do so between the selection measure and the pretest. Therefore, an unbiased estimate of the no-treatment expectation could be obtained.

Glass (1978) and Murray (1978) clearly demonstrate, however, that the use of a separate selection measure does not guarantee the elimination of all regression between the pretest and the posttest, although they do note that a reduction in the magnitude of the regression is to be expected. Murray (1978) further indicates that the only time that all regression would be eliminated when a selection test is used, is when the selection-pretest correlation is equal to the selection-posttest correlation.

The problem then, is to develop an unbiased estimate of the no-treatment expectation given that the selection procedure into Title I is fully understood. Murray (1978) has proposed one such estimation procedure which may lead to less biased estimates. In addition, there is the traditional multiple linear regression approach to this estimation task assuming some degree of multivariate normality and linearity.

The purpose of this paper is to investigate, through a limited simulation, the sampling distributions of these two estimation models and to determine the magnitude and direction of bias, if any, as well as the size of the standard errors under the assumptions of multivariate normality and linearity. The concern of greatest importance is that an estimation model will lead to large standard errors and consequently lead to frequent underestimates of the actual treatment effect.

The following notation will be used throughout this paper:

X_0 = Selection test given at time ∅.

X_1 = Pretest given at time 1.

X_2 = Posttest given at time 2.

r = Will represent population correlations.

Z = Unit normal standard scores.

$\hat{\bar{Z}}$ = Estimated mean.

$\bar{Z}$ = Obtained mean.

Estimation Procedures

The approach adopted by Murray (1978) for deriving the no-treatment expectation under a selection, pretest-posttest design, was to use a set of structural equations in a recursive path analytic paradigm. In addition to the three achievement measures, the paradigm included a group factor, an individual differences within group factor and an error component. The result of this path analytic approach indicates that an unbiased estimate of the no-treatment expectation can be obtained using the ratio of two correlations; the correlation of selection with posttest ($r_{x_0x_2}$) over the correlation of selection with pretest ($r_{x_0x_1}$) or $r_{x_0x_2}/r_{x_0x_1}$, times the estimated pretest mean. Murray points out that this estimate takes into account the distance of the treatment group mean from the population mean as well as the temporal erosion expressed by the ratio. Using standard score form, Murray's (1978) estimate of the no-treatment expectation is

(1) $$\hat{\bar{Z}}_2 = (r_{x_0x_2}/r_{x_0x_1})\,\hat{\bar{Z}}_1$$

where $\hat{\bar{Z}}_2$ is the estimated posttest mean,

$\hat{\bar{Z}}_1$ is the estimated pretest mean,

and $\hat{\bar{Z}}_1$ is obtained using $\hat{\bar{Z}}_1 = r_{x_0x_1}\,\bar{Z}_0$.

It becomes apparent when examining (1) that the regression effect from pretest to posttest will equal zero only when $r_{x_0x_1} = r_{x_0x_2}$.

As with all estimates of regression, this model is paramet(. dependent in that population correlations, means and variances (or very good estimates of them) are required. In a sense Murray's procedure is a stagewise approach in that $\hat{\bar{Z}}$, must be estimated from $\bar{Z}_0$ prior to estimating $\hat{\bar{Z}}_2$. That is, we first determine the regression from selection to pretest and then we compute the remaining regression from pretest to posttest using the ratio in (1). Actually, this stagewise process is not necessary since the same result can be obtained directly from

(2) $$\hat{\bar{Z}}_2 = r_{x_0x_2}\,\bar{Z}_0.$$

In fact this linear form appears in Murray's derivation one step prior to the ratio form. Through a simple substitution we have

$$\hat{\bar{Z}}_2 = (r_{x_0x_2}/r_{x_0x_1})\,\hat{\bar{Z}},$$

but $$\hat{\bar{Z}}_1 = r_{x_0x_1}\,\bar{Z}_0,$$

so $$\hat{\bar{Z}} = r_{x_0x_2}/r_{x_0x_1}\,(r_{x_0x_1}\,\bar{Z}_0)$$
$$= r_{x_0x_2}\,\bar{Z}_0.$$

This linear form, however, is less informative when a selection test is in use. For example, it would not be apparent from (2) that pretest to posttest regression is eliminated only when $r_{x_0x_1} = r_{x_0x_2}$.

The second estimation method examined in this paper is simply a direct extension of the one independent variable regression situation. When students are selected only on the pretest (with no prior preselection) then pretest to posttest regression is estimated directly from

$$\hat{\bar{Z}}_2 = r_{x_1x_2}\,\bar{Z}_1.$$

When a selection test is added to the picture, the logical extension is a multiple regression of the form

(3) $$\hat{\bar{Z}}_2 = b_1 \bar{Z}_0 + b_2 \bar{Z}_1 ,$$

where $$b_1 = (r_{x_0x_2} - r_{x_1x_2} r_{x_0x_1}) / 1 - r^2_{x_0x_1}$$

and $$b_2 = (r_{x_1x_2} - r_{x_0x_2} r_{x_0x_1}) / 1 - r^2_{x_0x_1} .$$

As in the case of (1), this form is parameter dependent in the correlations. However, in this case there is no need to estimate $\hat{\bar{Z}}_1$. Both the ratio form and the multiple linear form require the assumption of linearity and homogeneity of regression.

The questions of interest in this paper are 1) are the standard errors of the sampling distributions of $(\bar{Z}_2 - \hat{\bar{Z}}_2)$ sufficiently small to eliminate any major risk of underestimating the treatment effect, and 2) does varying the correlations significantly affect these standard errors?

Simulation Procedure

In order to simulate the regression effect of interest, it was necessary to produce three normally distributed random variables with specified intercorrelations. These variables could then serve as selection, pretest and posttest scores. The method used in this study is described in detail by Kaiser and Dickman (1962). The procedure uses component analysis for generating a sample score matrix $(\hat{Z})$ and a correlation matrix $(\hat{R})$ when given a specified population correlation matrix (R). The fundamental postulate of component analysis is

$$Z = FX,$$

where F is any factoring of R and X is a population score matrix. The components in F must be uncorrelated such as principal components or square root components. The sample score matrix is then produced using

$$\hat{Z} = F\hat{X}$$

where $\hat{X}$ is a matrix containing sample values from a normally distributed random variable. The variables in $\hat{X}$ are uncorrelated but the variables in $\hat{Z}$ are correlated according to the population correlation matrix (R).

The computer routine used in this study is GGNRM and it is found in the International Mathematics and Statistical Library (IMSL) package. It is an accurate, but slow and costly routine. The primary inputs to GGNRM are the population correlations, the number of cases to be generated and a seed to start the random number generator.

In this study, 3 x 3 correlation matricies were used and a 3 by 500 score matrix was generated. Descriptive statistics were then accumulated including the reproduced (sample) correlation matrix. This procedure was executed for 300 iterations in order to obtain a reasonable estimate of the sampling distributions of interest. This study is limited to a sample size of 500 due to the excessive time required.

The procedure involves using the first of the three score vectors as the selection test and selecting cases on the basis of a cutoff score. In this case, we are dealing with a multivariate normal distribution and therefore a Z of -.52 was used to select approximately the lower scoring 30% of the cases. After each iteration, we computed descriptive statistics on the total 500 cases as well as on the lower scoring 30% or about 150 cases that are of special interest.

Since a multivariate normal generator was used, it was a relatively simple matter to derive the expected values of all three variables. The first step involved determining the population mean of the lower scoring 30 percent. This is accomplished using

$$\overline{Z} = \psi(Z)/\phi(Z),$$

where $\psi(Z)$ is the height of the ordinate at a Z of -.52 and $\phi(Z)$ is the probability associated with a Z of -.52. Therefore we expect that the mean of the selection test variable will be

$$\bar{Z} = \psi(-.52) \ / \ \psi(-.52)$$
$$= .3485/.3015$$
$$= -1.1559.$$

Or, in terms of the NCE metric with mean of 50 and a σ of 21.06 we have

$$\bar{Z} = 21.06\ (-1.1559) + 50 = 25.6570.$$

In order to determine the expected pretest and posttest values, it is necessary to specify the correlations. For example in the run that used .9, .8, .9 as selection - pre, selection - post, and pre - post correlations respectively, the expected pretest value is

$$\hat{\bar{Z}}_1 = r_{x_0 x_1}\ \bar{Z}_0$$
$$= .9\ (-1.1559)$$
$$= -1.04,$$

or in NCEs $\hat{\bar{Z}} = 28.09$. The expected posttest value can then be determined using either estimation formula (1) or (3) since all input values are the same. Using (1) in this example we have

$$\hat{\bar{Z}}_2 = r_{x_0 x_2} \ / \ r_{x_0 x_1}\ (\hat{\bar{Z}}_1)$$
$$= .8/.9\ (-1.04)$$
$$= -.92,$$

or in NCEs $\hat{\bar{Z}}_2 = 30.53.$

The values of particular interest in this study, however, are the differences between obtained and estimated posttest means and the corresponding standard errors for these differences.

Results and Discussion

All computations were performed using unit normal standard scores (Z scores). For reporting purposes, these Z scores were then transformed to the NCE metric using

$$NCE = 21.06Z + 50.$$

The four correlation pattern runs were based upon a simulated district size of 500 cases per grade. Selection was based upon a Z of −.52 corresponding to the 30th percentile. Therefore, there were approximately 150 cases in each selected group. The means and standard errors reported in this paper were based upon a sampling distribution of 300 weighted averages. The range of sample sizes ran from 120 to 180.

Table I shows the population and reproduced sample correlations for each of the four runs. These correlations were selected in order to examine the effects of high, moderate, and low patterns. Run 2 was selected to demonstrate the case in which $r_{x_0x_2} = r_{x_0x_1}$.

Table II shows the obtained and estimated regression effects and corresponding standard errors, for selection to pretest, pretest to posttest and selection to posttest by run. In addition, the table shows the differences between obtained and estimated posttest means and corresponding standard errors. As would be anticipated, the magnitude of the correlations is inversely related to the amount of regression and to the size of the standard errors.

Run 1 represents the lowest set of correlations and consequently the greatest amount of regression. The pretest to posttest regression was just a little under 5 [illegible]'s. The regression estimates were all quite close to the obtained values.

TABLE I

Population and Reproduced Sample Correlations by Run

		Sel-Pre	Pre-Post	Sel-Post
Run 1	Pop.	.7	.6	.5
	Sample	.7017	.5982	.5006
Run 2	Pop.	.6	.7	.6
	Sample	.6016	.7017	.5981
Run 3	Pop.	.8	.8	.6
	Sample	.8001	.8000	.5997
Run 4	Pop.	.9	.9	.8
	Sample	.9003	.9002	.8001

TABLE II

OBTAINED AND ESTIMATED REGRESSION WITH STANDARD ERRORS BY RUN

(VALUES ARE EXPRESSED IN NCE UNITS)

Run 1

	Obtained Regression		Estimated Regression	
	Amount	S.E.	Amount	S.E.
Sel-Pre	7.177	1.342	7.221	1.308
Pre-Post	4.830	1.575	4.969*	1.320
Sel-Post	12.006	1.660	12.146	.476

$(\bar{Z}_2 - \hat{\bar{Z}}_2)$

Ratio		Mult. Linear	
Amount	S.E.	Amount	S.E.
-.140	1.553	-.085	1.407

Run 2

	Obtained Regression		Estimated Regression	
	Amount	S.E.	Amount	S.E.
Sel-Pre	9.611	1.306	9.728	.385
Pre-Post	.105	1.260	.117*	1.292
Sel-Post	9.715	1.396	9.728	.385

$(\bar{Z}_2 - \hat{\bar{Z}}_2)$

Ratio		Mult. Linear	
Amount	S.E.	Amount	S.E.
-.013	1.339	.05	1.126

TABLE II (continued)

Run 3

	Obtained Regression		Estimated Regression	
	Amount	S.E.	Amount	S.E.
Sel-Pre	4.942	1.047	4.877	.177
Pre-Post	4.867	1.063	4.812*	1.042
Sel-Post	9.790	1.475	9.753	.355

$(\bar{Z}_2 - \dot{\bar{Z}}_2)$

Ratio		Mult. Linear	
Amount	S.E.	Amount	S.E.
.037	1.426	-.021	1.047

Run 4

	Obtained Regression		Estimated Regression	
	Amount	S.E.	Amount	S.E.
Sel-Pre	2.485	.754	2.438	.089
Pre-Post	2.421	.769	2.391*	.752
Sel-Post	4.906	1.086	4.877	.177

$(\bar{Z}_2 - \dot{\bar{Z}}_2)$

Ratio		Mult. Linear	
Amount	S.E.	Amount	S.E.
.03	1.069	-.015	.764

*This value represents the ratio estimate.

For both the ratio and multiple linear estimators, there was a slight tendency to overestimate the no-treatment expectation when comparing obtained and expected posttest means. The magnitude of this overestimation is so small that it is of little pragmatic importance. On the other hand, the standard errors of both estimators appear to be too large to risk their use in estimating no-treatment expectations. That is, it would be quite possible for an LEA with 500 students or less per grade to underestimate Title I treatment effects by as much as two or three NCE's Using either estimator.

Run 2 also reflects relatively low correlations, however, in this case the population correlations were selected such that $r_{x_0x_2} = r_{x_0x_1}$. As can be seen, the amount of statistical regression from pretest to posttest is only about one-tenth of an NCE. Again the two estimators are quite accurate in their forecast. Although this result does demonstrate that pretest to posttest regression is virtually eliminated when $r_{x_0x_2} = r_{x_0x_1}$ regardless of the size of $r_{x_1x_2}$, the standard errors associated with $(\bar{Z}_2 - \hat{\bar{Z}}_2)$ are again too large to recommend that either estimator be used. In this case, under-estimates of treatment effect as great as two NCE's are still quite possible.

In Run 3 the correlations are moderately high with the exception of $r_{x_0x_2}$. And since this correlation is critical to the pretest to posttest regression, it is not surprising to find a pretest to posttest regression of almost 5 NCE's. However, the standard errors are all slightly smaller due to the increased magnitude of the other correlations.

It is also interesting to note that in this case both estimators produced sampling mean differences between obtained and estimated posttest means of less than 4/100 of an NCE. Although there is a slight drop in the standard errors associated with $(\bar{Z}_2 - \hat{\bar{Z}}_2)$ in this run, they are still too large to risk using either estimator.

Finally, in Run 4 the correlations are relatively high with $r_{x_0x_2}$ set at .8. In this case the pretest to posttest regression has dropped to 2.4 NCE's. The standard errors associated with $(\overline{Z}_2 - \hat{\overline{Z}}_2)$ are also smaller than those in runs one through three. Unfortunately, even these standard errors could result in underestimating the treatment effects by as much as 1.5 to 2 NCE's depending on the estimator used.

The results of this limited simulation indicate that, at least for districts with 500 or fewer students per grade and correlations of .9 or lower, the ratio and multiple linear estimators do not provide a viable option for obtaining an unbiased estimate of the no-treatment expectation. The chances of over estimating the no-treatment expectation are too great and consequently too many LEAs could end up seriously underestimating their treatment effects. On the other hand, it is possible that, for districts with 2,000 or more students per grade, either estimator would produce estimates with sufficiently small standard errors.

Another, and perhaps more viable option, is to use local correlations rather than requiring population correlation. In this case, we would be defining the population mean toward which low scoring students would regress as the district mean per grade. This might be a particularly attractive approach when the district means are close to the national means. One drawback to this approach would be that correlations would have to be based on non-Title I participants. As a result, there would be some attenuation due to restriction in range. However, the estimates are still likely to be more accurate and less variable than when depending upon national correlation data.

REFERENCES

Glass, G.. An unpublished note to the Technical Assistance Centers regarding the regression effect and selection testing, 1978.

Tallmadge, G. K.. The Regression Effect. ESEA Title I Evaluation and Reporting System, Technical Paper No. 3, 1976.

Kaiser, H. F. and Dickman, K.. Sample and population score matrices and sample correlation matrix. Psychometrica, Vol. 27, No. 2, 1962.