RB-76-9

RESEARCH BULLETIN

ASSIGNMENT TO TREATMENT GROUP ON THE BASIS OF A COVARIATE

Donald B. Rubin

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
March 1976

Assignment to Treatment Group on the Basis of a Covariate

Abstract

When assignment to treatment group is made solely on the basis of the value of a covariate, X, effort should be concentrated on estimating the conditional expectations of the dependent variable Y given X in the treatment and control groups. One then averages the difference between these conditional expectations over the distribution of X in the relevant population. There is no need for concern about "other" sources of bias, e.g., unreliability of X, unmeasured background variables. If the conditional expectations are parallel and linear, the proper regression adjustment is the simple covariance adjustment. However, since the resulting estimates may be somewhat sensitive to the adequacy of the underlying model, it is wise to search for non-parallelism and non-linearity in these conditional expectations. Blocking on the values of X is also appropriate, although the resulting estimates may be somewhat sensitive to the coarseness of the blocking employed.

Assignment to Treatment Group on the Basis of a Covariate

1. Introduction and summary

It sometimes happens in studies that the units are divided into two treatment groups solely on the basis of a covariate, X, supplemented, perhaps, with some randomization. For example, those who did poorly on X (a reading test) receive an experimental treatment (a compensatory reading program); others receive the standard control treatment; perhaps those with intermediate scores on X are randomly assigned. The critical point is that the probability that a unit is exposed to Treatment 1 rather than Treatment 2 is a function only of the values of X in the sample. Both groups are at a later time given test Y.

The central question is: what is the average effect on Y of Treatment 1 vs. Treatment 2 for the "relevant" population? For simplicity of discussion, we will usually assume the relevant population is the one from which all the units being studied are considered a random sample, say P. The associated effect is called τ.

Some researchers might wonder whether to use gain scores, simple posttest scores, covariance adjusted scores (possibly "adjusted" for reliability), or some other device to estimate τ. These issues are discussed in the general case of the non-randomized study in many places, e.g., Campbell and Stanley (1965), Campbell and Erlebacher (1970), Cochran and Rubin (1973), Lord (1967), and Rubin (1973a,b, 1974, 1975, 1976a,b).

The answer is straightforward when assignment is solely on the basis of X: the appropriate estimate is the average difference between the estimated regressions of Y on X in the two treatment groups, the average being taken over all units in the study if the relevant population is P. The regressions can be estimated using least squares, robust techniques, blocking, or matching methods. In the special case of linear and parallel regressions of Y on X in the treatment groups, this reduces to the simple covariance adjusted estimate. Neither gain scores nor scores adjusted for the reliability of X are generally appropriate (no matter how unreliable X).

In the following development we use unbiasedness as the criterion indicating the appropriateness of estimators. We do so only to indicate that the estimator tends to estimate the correct quantity without further adjustment. Clearly, we would prefer to use a consistent estimator whose bias and variance in small samples are small rather than an unbiased estimator whose variance is large.

2. Notation and preliminary results

Let $\mu_1(x)$ be the conditional expectation of Y for a random unit from the population P with score x on X given exposure to Treatment 1. And let $\mu_2(x)$ be the conditional expectation of Y for a random unit from the population P with score x on X and given exposure to Treatment 2. For example, under the usual linear models $\mu_1(x)$ and $\mu_2(x)$ are assumed to be linear functions of x. However, in this discussion we do not necessarily assume these conditional expectations to be linear. The effect of Treatment 1 vs. Treatment 2 at X = x in population P is given by $\mu_1(x) - \mu_2(x)$. (See Figure 1.) If $\mu_1(x) - \mu_2(x)$ = constant for all x, $\mu_1(x)$ and $\mu_2(x)$ are said to be parallel. The $\mu_i(x)$ are sometimes called the "response functions of Y given X" or the "regressions of Y on X."

The average effect in population P of Treatment 1 vs. Treatment 2, τ, is thus $\mu_1(x) - \mu_2(x)$ averaged over the distribution of X in the population:

$$(1) \qquad \tau = \underset{x \in P}{\text{ave}} \left[\mu_1(x) - \mu_2(x)\right] ,$$

where $\underset{x \in P}{\text{ave}} [\cdot]$ means the average (mean) value of the quantity in the brackets over the distribution of x in the population P. Notice that no matter which variable we choose to be X, the parameter to be estimated, τ, is the same.

Throughout the rest of the paper we will assume the following sampling situation. A random sample of size n_1+n_2 from the population

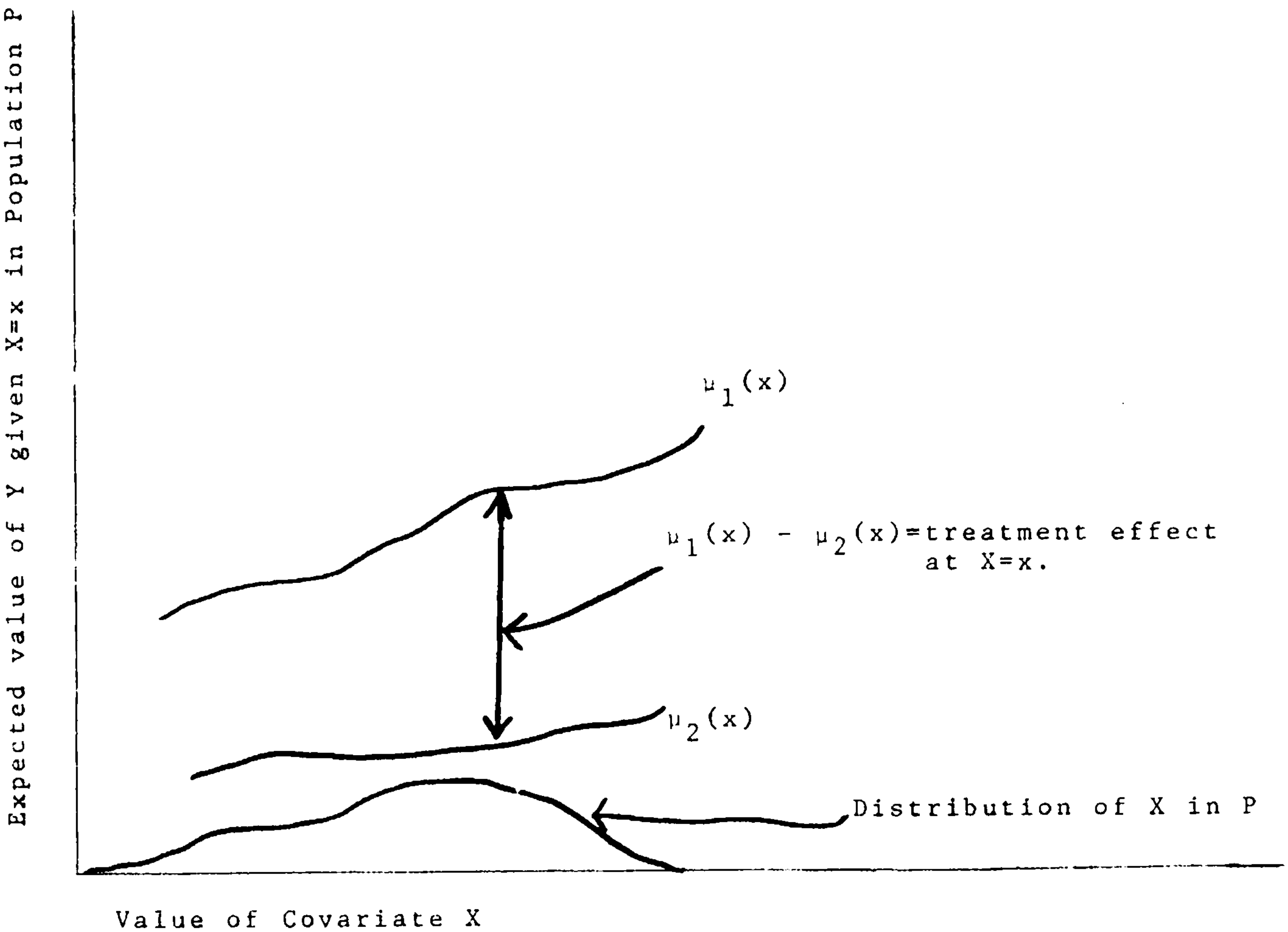


Figure 1

The Treatment Effect in Population P:

$$\tau = \operatorname*{Ave}_{x \varepsilon P} \left[\mu_1(x) - \mu_2(x)\right]$$

P is divided into two samples of sizes n_1 and n_2 solely on the basis of the values of X. The first group is exposed to Treatment 1 and the second group is exposed to Treatment 2. Let x_{ij}, y_{ij} $i=1,2$ $j=1,\ldots,n_i$ be the values of X and Y in the two samples.

Since the x_{ij} are a random sample from P, Result 1 is immediate from the definition of τ.

Result 1: The quantity

$$(2) \qquad \frac{1}{n_1+n_2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left[\mu_1(x_{ij}) - \mu_2(x_{ij})\right]$$

is unbiased for τ.

If we had conditionally unbiased estimates of the values $\mu_1(x_{ij})$ and $\mu_2(x_{ij})$, $i=1,2$ $j=1,\ldots,n_i$, we could substitute them into expression (2) to obtain an unbiased estimate of τ. By conditionally unbiased we mean unbiased conditionally given the values x_{ij} that occur in the sample. Notice, then, that only the values of x_{ij} in the sample and the conditional expectations of Y given X under the two treatments are needed. Nothing more. No matter how "unreliable" X or what $\mu_1(x)$ and $\mu_2(x)$ are like, no reliability correction is relevant.

Result 2 is the key to obtaining unbiased estimators of τ.

Result 2: The value y_{1j} is a conditionally unbiased estimate of $\mu_1(x_{1j})$ $j=1,\ldots,n_1$, and the value y_{2j} is a conditionally unbiased estimate of $\mu_2(x_{2j})$ $j=1,\ldots,n_2$.

Result 2 follows because the conditional expectation of y_{ij} given both (a) the actual values of the X data and (b) the values of the X data satisfy some selection criterion that determined exposure to Treatment i, equals the conditional expectation of y_{ij} given simply condition (a), and this conditional expectation equals $\mu_i(x_{ij})$ by definition.

By result 2, we now have conditionally unbiased estimates of $\mu_1(x_{1j})$ $j=1,\ldots,n_1$ and $\mu_2(x_{2j})$ $j=1,\ldots,n_2$. However, we still lack conditionally unbiased estimates of $\mu_1(x_{2j})$ $j=1,\ldots,n_2$ and $\mu_2(x_{1j})$ $j=1,\ldots,n_1$ which we need in order to use Result 1 to obtain an unbiased estimate of τ. There are two general methods for obtaining conditionally unbiased estimates for these quantities that we will discuss here: (a) fitting a model to the data to obtain estimates of the functions $\mu_1(x)$ and $\mu_2(x)$, and (b) grouping together Treatment 1 and Treatment 2 units with similar values of X to obtain estimates of the difference $\mu_1(x) - \mu_2(x)$ at particular values that are representative of the distribution of X in P.

3. Estimating $\mu_1(x_{ij})$ and $\mu_2(x_{ij})$ by model fitting

One method for estimating the values of $\mu_1(x_{2j})$ and $\mu_2(x_{1j})$ is via a model for these curves. This is most appropriate when X takes on many values (e.g., age, height). Obviously the accuracy of the resulting estimates will be somewhat dependent on the appropriateness of the model chosen.

The estimate of $\mu_1(x)$ and $\mu_2(x)$ will be illustrated in the simple case when we assume both are linear in x. The usual least squares estimates are:

$$(3) \qquad \hat{\mu}_1(x) = \bar{y}_1 + \hat{\beta}_1(x-\bar{x}_1)$$

$$(4) \qquad \hat{\mu}_2(x) = \bar{y}_2 + \hat{\beta}_2(x-\bar{x}_2)$$

where

$$(5) \qquad \hat{\beta}_i = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)(x_{ij} - \bar{x}_i) / \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2, \quad i=1,2.$$

Result 3: If both $\mu_1(x)$ and $\mu_2(x)$ are linear in x, the estimator

$$(6) \qquad \bar{y}_1 - \bar{y}_2 + (\bar{x}_1 - \bar{x}_2) \frac{n_1\hat{\beta}_2 + n_2\hat{\beta}_1}{n_1 + n_2}$$

is unbiased for τ.

In order to prove Result 3, first note that expression (6) equals

$$(7) \qquad \frac{1}{n_1+n_2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left[\hat{\mu}_1(x_{ij}) - \hat{\mu}_2(x_{ij})\right] \quad ,$$

where $\hat{\mu}_1(x_{ij})$ and $\hat{\mu}_2(x_{ij})$ are given by equations (3), (4) and (5). Now let $\mu_i(x) = \alpha_i + \beta_i x$ in accordance with the assumptions. Next note that by Result 2, the conditional expectation of $\hat{\mu}_i(x)$ is $\mu_i(x)$: (a) the conditional expectation of $\bar{y}_i$ is $\alpha_i + \beta_i \bar{x}_i$, and (b) the conditional expectation of $\hat{\beta}_i$ is $\sum_{j=1}^{n_i} (\mu_i(x_{ij}) - \overline{\mu_i(x_{ij})})(x_{ij}-\bar{x}_i)/\sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2 = \beta_i$. Consequently, $\hat{\mu}_1(x_{ij}) - \hat{\mu}_2(x_{ij})$ is a conditionally unbiased estimate of $\mu_1(x_{ij}) - \mu_2(x_{ij})$ (for $j=1, \ldots, n_i$; $i=1,2$), and thus by Result 1, expression (7) (and thus (6)) is unbiased for τ.

<u>Result 4</u>: If $\mu_1(x)$ and $\mu_2(x)$ are parallel and linear in x, then the simple analysis of covariance estimator

$$(8) \qquad \bar{y}_1 - \bar{y}_2 = (\bar{x}_1 - \bar{x}_2)\,\hat{\beta}$$

where $$\hat{\beta} = \sum_{i=1}^{2} \sum_{i=1}^{n_i} (y_{ij} - \bar{y}_i)(x_{ij} - \bar{x}_i) / \sum_{i=1}^{2} \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2$$

is unbiased for τ.

The proof of Result 4 is essentially the same as the proof of Result 3 with the change that now $\beta_1 = \beta_2 = \beta$ and both $\hat{\beta}_1$ and $\hat{\beta}_2$ are replaced by $\hat{\beta}$ which is a conditionally unbiased estimate of β.

Results analogous to Results 3 and 4 follow when $\mu_1(x)$ and $\mu_2(x)$ are polynomial in x or any linear combination of specified functions of x (e.g., e^x). The only change is that we must perform a multiple least squares regression.

Of course we never know that $\mu_1(x)$ and $\mu_2(x)$ really are linear in x (or linear in some specified functions of x). Hence, the estimators given by (6) and (8) are somewhat dependent upon the accuracy of the model. Rubin (1973b) presents evidence relevant to the estimators given by (6) and (8) indicating that in some cases the linear approximation is adequate to remove most of the bias present in the simple estimator $\bar{y}_1 - \bar{y}_2$ but in other cases it is inadequate even when $\mu_1(x)$ and $\mu_2(x)$ are smooth monotone functions. In addition, we are of course not limited to the least squares estimates of the regression functions $\mu_1(x)$ and $\mu_2(x)$. A more robust method may often be appropriate when trying to estimate these conditional expectations (c.f. Beaton and Tukey, 1975).

4. Estimating $\mu_1(x_{ij})$ and $\mu_2(x_{ij})$ by blocking on X

When the selection into treatment group allows the distribution of X in the two treatment groups to overlap substantially, it may be possible to obtain conditioning unbiased (or nearly so) estimates of $\mu_1(x_{2j})$ and $\mu_2(x_{1j})$ without fitting a model. The obvious but crucial point is that if $x_{1j} = x_{2j}$, then y_{1j} is unbiased for $\mu_1(x_{2j}) = \mu_1(x_{1j})$ and y_{2j} is unbiased for $\mu_2(x_{1j}) = \mu_2(x_{2j})$.

Suppose that in the samples there are only K distinct values of X , say $x_1,\ldots,x_k$, where n_{1k} Treatment 1 units and n_{2k} Treatment 2 units have X values equal to x_k , $k=1,\ldots,K$. It follows since assignment to the groups was solely on the basis of X , that the units with value x_k were probabilistically divided into the two groups, e.g., the study is randomized within K blocks. Let $\bar{y}_{1k}$ be the average Y value for the n_{1k} Treatment 1 units whose X value equals x_k ; similarly let $\bar{y}_{2k}$ be the average Y value for the n_{2k} Treatment 2 units whose X value equals x_k. If $n_{ik} = 0$ for some i and k, then the corresponding $\bar{y}_{ik}$ is not defined.

Result 5: If $n_{1k} > 0$ and $n_{2k} > 0$ for all $k = 1,\ldots,K$, then the estimator

$$(9) \qquad \frac{1}{n_1+n_2}\left[\sum_{k=1}^{K}(n_{1k}+n_{2k})(\bar{y}_{1k}-\bar{y}_{2k})\right]$$

$$= \frac{1}{n_1+n_2}\left[n_1\bar{y}_1 - n_2\bar{y}_2 + \sum_{k=1}^{K} n_{2k}\bar{y}_{1k} - \sum_{k=1}^{K} n_{1k}\bar{y}_{2k}\right]$$

is unbiased for τ.

Result 5 follows because by Result 2 $\bar{y}_{1k}$ is an unbiased estimate of $\mu_1(x_k)$, $\bar{y}_{2k}$ is an unbiased estimate of $\mu_2(x_k)$, and so $\bar{y}_{1k} - \bar{y}_{2k}$ is an unbiased estimate of $\mu_1(x_k) - \mu_2(x_k)$, $k=1,\ldots,K$. That is, the difference of the Y mean for those Treatment 1 units whose X value is x_k and the Y mean for those Treatment 2 units whose X value is x_k is an unbiased estimate of the Treatment 1 - Treatment 2 effect at x_k. Hence, from Result 1 we have Result 5.

The advantage of the estimator given by (9) is that it does not depend on the accuracy of some underlying model for its unbiasedness. The disadvantage of the estimator is that if X takes on many values, some n_{ik} may be zero and then the estimator is not defined. This occurrence is not unusual in practice.

A common practical method is to define a new variable that is a discretized version of X such that on this new variable the conditions of Result 5 are satisfied. However, since the assignment process was on the basis of X (not the discretized X), the estimator given by (9) based on the discretized X is no longer necessarily unbiased for τ. If the discretized X has many values, we would think, however, that the bias would be small. Cochran (1968) investigated a related situation and concluded that in many cases a discretized version of X with 5 or 6 values was adequate to remove over 90% of the bias present in the simple estimator $\bar{y}_1 - \bar{y}_2$.

Of particular interest is the case in which X is discretized as finely as possible (i.e., K is maximized subject to the constraint that each $n_{ik} > 0$). It would be of practical importance to investigate the bias of the estimate (9) based on this discretized

X under various (a) underlying distributions of X in P, (b) assignment processes based on X, and (c) response functions $\mu_1(x)$ and $\mu_2(x)$.

Another method for handling cases of some $n_{ik} = 0$ is to discard units. Result 6 is immediate from Result 1.

Result 6: If $\mu_1(x)$ and $\mu_2(x)$ are parallel, then

$$(10) \qquad \sum_{k=1}^{k} \delta_k(\bar{y}_{1k} - \bar{y}_{2k}) / \sum_{k=1}^{k} \delta_k$$

is unbiased for τ for any $\delta_k \geq 0$ such that $\delta_k = 0$ if either n_{1k} or $n_{2k} = 0$ (e.g., $\delta_k = (n_{1k} + n_{2k})$ if both n_{1k} and $n_{2k} > 0$).

Notice that the estimator given by (10) essentially discards those units whose X values are not the same as the X value of some unit who took the other treatment. This procedure is commonly known as matching on the values of X, and in some cases makes a lot of sense. For example, suppose X has been recorded and there is an additional cost in recording Y even though the treatments have already been given to all of the units. For example, the regular and compensatory reading programs have been given, background variables have been recorded, but there is an additional expense in giving and recording a battery of detailed posttests to each student. In these situations it is appropriate to ask how to choose the units on which to record Y. In practice, however, it may not be desirable to assume the regressions are parallel, so that the estimator given by (10) may not be useful in some cases. Matching has wider applicability when a subpopulation of P is of primary interest.

5. Generalizing to a subpopulation of P

At times, the relevant population to which we want to generalize the results of the study will not be the whole population P, but rather a subpopulation of P, say P_x defined (probabilistically, perhaps) by values on the covariate X. For example, the Treatment 1 units may be considered to be a random sample from the relevant population -- those in need of extra treatment because of low values of X.

In such cases, all the results presented here generalize to estimating $\tau_x = \underset{x \varepsilon P_x}{\text{ave}} \left[\mu_1(x) - \mu_2(x)\right]$. The quantity τ_x is the treatment effect in the population P_x because $\mu_1(x) - \mu_2(x)$ is the treatment effect in P_x at X=x as well as in P as X=x. That is, the conditional expectation of Y given (a) Treatment i, (b) X=x, and (c) X satisfied some criterion that defined membership in P_x is simply the conditional expectation of Y given (a) Treatment i and (b) X=x.

Hence, Result 1 generalizes to estimating τ_x if the average over all X values in the sample is replaced by the average over X values that are representative of the population P_x. Result 2 is true as stated for P_x. In Result (3), the corresponding estimator of τ_x is now given by expression (7) with the averaging over all units replaced by averaging over units representative of the population P_x. For example, if the units exposed to Treatment 1 are considered a random sample from P_x, this averaging of expression (7) leads to the unbiased estimator of τ_x given by $(\bar{y}_1-\bar{y}_2) - (\bar{x}_1-\bar{x}_2)\beta_2$. This estimator is discussed by Belsen (1956) and Cochran (1970) in some detail.

If $\mu_1(x)$ and $\mu_2(x)$ are parallel, $\tau = \tau_x$ so that Result 4 as well as Result 6 apply for obtaining unbiased estimates of the treatment effect for any subpopulation P_x.

The extension of Result 5 to the subpopulation P_x is somewhat more interesting, although equally straightforward. For example, again suppose the Treatment 1 units are a random sample from P_x; then if $n_{2k}>0$ whenever $n_{1k}>0$, the estimator

$$(11) \qquad \frac{1}{n_1} \sum_{k=1}^{K} n_{1k} (\bar{y}_{1k} - \bar{y}_{2k})$$

is unbiased for τ_x. This estimate discards those Treatment 2 units whose X values are not found among the Treatment 1 units. Finding for each Treatment 1 unit a Treatment 2 unit with the same X value and forming the estimate (11) is commonly known as matched sampling (Rubin, 1973a). As discussed at the end of Section 4, such estimates that discard data are really most appropriate when deciding which units should be retained to record Y.

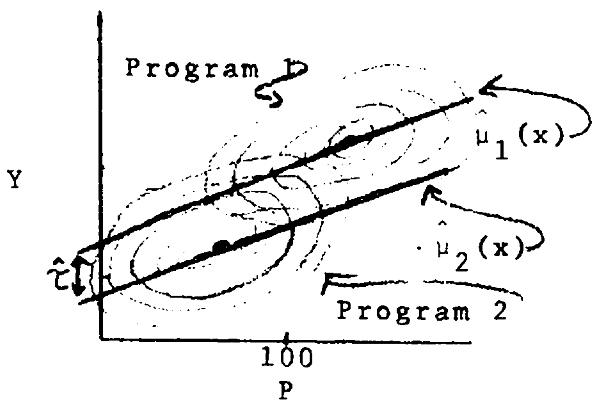
6. A simple example

In order to demonstrate some of the ideas presented here, assume we have a group of children and two reading programs (treatments) under study. First suppose that we assign the children to treatment groups on the basis of pretest score (P) as follows: the probability of being exposed to Program 2 is $\frac{100}{100+P}$ so that the children who did better on the pretest have a greater chance of being given Program 1. Suppose the resultant distributions of posttest score (Y) and P in the two treatment groups are given in Figure 2a, where the response functions $\hat{\mu}_1(x)$ and $\hat{\mu}_2(x)$ are essentially linear and parallel. Since assignment is on the basis of P, the covariate X is P, and the simple covariance adjusted estimate is an unbiased estimator of the effect of Program 1 - Program 2: $\hat{\tau}$ = the difference between $\hat{\mu}_1(x)$ and $\hat{\mu}_2(x)$.

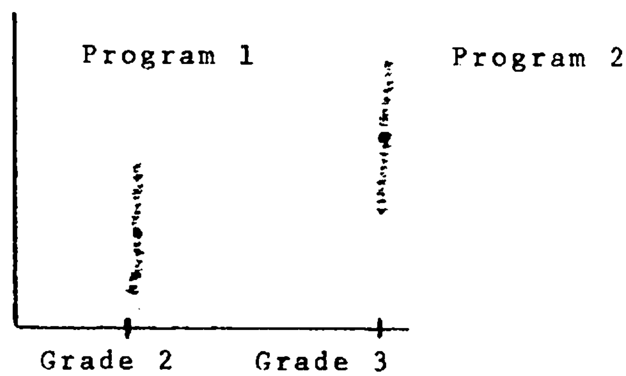
Now suppose that we assigned third graders to Program 1 and second graders to Program 2 with the same resultant plot as in Figure 2a. Is $\hat{\tau}$ still an unbiased estimator of the effect of Program 1 - Program 2? The answer is no, without further assumptions, because the covariate X that was used to assign treatments was not P but grade, and the plot of Y vs. grade looks like Figure 2b. Because each treatment group has only one value of X, using the methods discussed in Sections 3,4,5, we now cannot estimate the response functions $\mu_1(x)$ or $\mu_2(x)$. Nor can we block Program 1 and Program 2 children with similar values of X. Using these methods with pretest as the covariate makes the implicit assumption that in each treatment group, the expected value of Y given pretest and grade is the same as the expected value of Y given just pretest.

The crucial point is that knowledge of the assignment process is critical to drawing inferences about the effect of the programs. One cannot simply look at the plot of posttest on pretest and properly estimate treatment effects. An example similar to the above one is discussed by Lord (1967).

Figure 2: Posttest given covariate



2a: Covariate = pretest



2b: Covariate = grade

7. Discussion of needed investigations

In this paper we have stated the fact that if assignment to treatment group is on the basis of the value of a covariate, X, one must concentrate effort on the essential problem of estimating the conditional expectation of Y given X in each treatment. One then averages the difference between these conditional expectations over the values of X that are representative of the population of interest.

Two general methods for estimating these expectations were discussed: least-squares model fitting and blocking on the values of X. Little relevant work has been done on how well these techniques are likely to do in practice alone or in combination. A relevant statistical study would specify and then vary:

(a) the sample size, $n_1 + n_2$

(b) the distribution of X in P

(c) the assignment mechanism

(d) the conditional expectations, $\mu_1(x)$ and $\mu_2(x)$.

One would then find the distribution of estimators resulting from using the model fitting and blocking methods discussed here.

Of course the really interesting case is that of multivariate X. This follows because in natural settings we may not know the assignment mechanism, but may feel that it essentially is based on no more than a particular collection of p variables. For example, teachers when deciding which students should receive compensatory reading treatments presumably use only observable background characteristics of the children and test scores (not "true scores") in some way. Thus the assignment mechanism may be adequately approximated by some function (possibly probabilistic) of only those variables.

All the results presented here for univariate X generalize immediately (conceptually at least) to multivariate X (e.g., $\hat{\beta}$ is now a vector of regression coefficients). Some work on multivariate matching methods is given in Althauser and Rubin (1970), Cochran and Rubin (1973) and Rubin (1976a,b), but has received little attention otherwise.

Certainly a serious effort on both the univariate case and the multivariate case is worthwhile, not only in order to improve the analysis of existing non-randomized studies but also in order to study the possibility of finding designs that are tolerable given social constraints, not randomized in the usual sense, but still allow useful inferences for causality.

REFERENCES

Althauser, R. P., & Rubin, D. B. (1970): "The Computerized Construction of a Matched Sample", The American Journal of Sociology, 76, 325-346.

Beaton, A.E., & Tukey, J.W. (1975): "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data", Technometrics, 16, 2.

Belsen, W.A. (1956): "A Technique for Studying the Effects of a Television Broadcast", Applied Statistics, 5, 195-202.

Campbell, D.T., & Erlebacher, A. (1970): "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful", In J. Hellmuth (Ed.), The Disadvantaged Child. 3. Compensatory Education: A National Debate. New York: Brunner/Mazel.

Campbell, D.T., & Stanley, J.C. (1963): "Experimental and Quasi-Experimental Designs for Research on Teaching", Handbook of Research on Teaching. Chicago: Rand McNally, 171-246.

Cochran, W.G. (1968): "The Effectiveness of Adjustment by Sub-Classification in Removing Bias in Observational Studies", Biometrics, 24, 295-313.

Cochran, W.G. (1970): "Some Effects of Errors of Measurement on Linear Regression", In Proceedings of the 6th Berkeley Symposium, I, 527-539.

Cochran, W.G.,& Rubin, D.B. (1973): "Controlling Bias in Observational Studies", A review, Mahalanobis Memorial Volume Sankhyā-A, 417-446.

Lord, F.M. (1967): "A Paradox in the Interpretation of Group Comparisons", Psychological Bulletin, 68, 5, 304-305.

Rubin, D.B. (1973a): "Matching to Remove Bias in Observation Studies", Biometrics, 29, 1, 159-183.

Rubin, D. B. (1973b): "The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies", Biometrics, 29, 1, 184-203.

Rubin, D.B. (1974): "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies", Journal of Educational Psychology, 66, 688-701.

Rubin, D.B. (1975): "Bayesian Inference for Causality: The Importance of Randomization". In The Proceedings of the Social Statistics Section of the American Statistical Association.

Rubin, D. B. (1976a): "Multivariate Matching Methods That Are Equal Percent Bias Reducing, I: Some Examples". Biometrics, in press.

Rubin, D. B. (1976b): "Multivariate Matching Methods That Are Equal Percent Bias Reducing II: Maximums on Bias Reduction for Fixed Sample Sizes". Biometrics, in press.