123-72

INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

SELECTION BIAS IN EVALUATING TREATMENT EFFECTS: SOME FORMAL ILLUSTRATIONS

Arthur S. Goldberger

UNIVERSITY OF WISCONSIN ~ MADISON

SELECTION BIAS IN EVALUATING TREATMENT EFFECTS: SOME FORMAL ILLUSTRATIONS

Arthur S. Goldberger

The research reported here was supported in part by funds granted to the Institute for Research on Poverty at the University of Wisconsin by the Office of Economic Opportunity pursuant to the provisions of the Economic Opportunity Act of 1964. The conclusions are the sole responsibility of the author. The research draws heavily on the work of Burt Barnow and Glen Cain. The author thanks Irv Garfinkel and Edward Gramlich for helpful comments.

April 1972

ABSTRACT

Regression analyses of compensatory educational programs have been criticized on the grounds that the pupils were not randomly selected for the program. Specifically, it has been argued that a spurious, deleterious effect of the treatment will be observed when the selection procedure systematically puts lower-ability subjects into the treatment group and higher-ability students into the control group.

In this paper, we evaluate that argument in terms of a simple test-score model. Pretest score and posttest score are assumed to be fallible measures of underlying true ability, the true treatment effect being zero. Posttest score is regressed on pretest score and a treatment dummy variable. We find that the spurious effect arises when subjects are selected for treatment explicitly on the basis of true ability. However, when subjects are selected for treatment explicitly on the basis of pretest score, the spurious effect vanishes. Thus the criticism mentioned above is seriously misleading.

Our analysis is a formal one, regressions being expressed in terms of population parameters. We also evaluate the statistical efficiency of selection-on-the-basis-of-pretest relative to random selection. The appendices contain some general results on conditional distributions, moments, and regressions for various partitions of bivariate normal distributions.

SELECTION BIAS IN EVALUATING TREATMENT EFFECTS: SOME FORMAL ILLUSTRATIONS

Arthur S. Goldberger

1. Introduction

When subjects are not assigned randomly to treatment and control groups, the possibility arises that a spurious treatment effect may be observed. This possibility has been emphasized in a recent critique of the evaluation of compensatory educational programs. Campbell and Erlebacher (1970) assert:

> The compensatory program is made available to the most needy, and the 'control' group then sought from among the untreated children in the same community. Often this untreated population is on the average more able than the 'experimental' group. In such a situation, the usual procedures of selection, adjustment, and analysis produce systematic biases in the direction of making the compensatory program look deleterious.

But this critique may be misleading. The mere fact that the control group is more able than the treatment group does not suffice to produce bias in the evaluation of the treatment effect. We propose to demonstrate this point in terms of a highly idealized setting, that is in terms of a formal model.

2. The Basic Model

We suppose that true ability x^* is normally distributed with expectation zero and variance Q:

$$x^* \sim N(0, Q).$$

Further, we suppose that pretest score x and posttest score y are erroneous measures of true ability, more precisely, that

$$x = x^* + u \quad , \qquad\qquad y = x^* + v$$

where u and v are normally distributed with expectation zero and common variance. The common variance of the measurement errors can be written as $(1 - P)/P$ times the variance of x^* for some $0 < P < 1$; thus

$$u \sim N(0, (1 - P)Q/P) \ , \qquad\qquad v \sim N(0, (1 - P)Q/P) \ .$$

(The motive for parameterizing in terms of Q and P will soon become clear). Further, we suppose that x^*, u, and v are independent.

Consequently, the test scores have expectations

$$E(x) = E(x^* + u) = E(x^*) + E(u) = 0 + 0 = 0 = E(y),$$

variances

$$\begin{aligned} V(x) = V(x^* + u) = V(x^*) + V(u) + 2\ C(x^*, u) &= Q + (1\text{-}P)Q/P \\ &= Q/P = V(y), \end{aligned}$$

and covariance

$$C(x,y) = C(x^* + u, x^* + v) = V(x^*) = Q \ ,$$

and are joint-normally distributed. In a joint-normal distribution any regression function is linear and the slope(s) are readily calculated from the variances and covariances. Specifically, the regression of y on x is

(1) $$E(y|x) = (C(x,y)/V(x))\ x = (Q/(Q/P))\ x = P\ x.$$

Note that

$$P = V(x^*)/V(x) = V(x^*)/(V(x^*) + V(u)) = V(x^*)/V(y).$$

This is the variance ratio which plays a key role in the subsequent analysis.

As a matter of fact, x*, u, v, x, y are joint-normally distributed with zero expectations and variance-covariance matrix

	x*	u	v	x	y
x*	Q	0	0	Q	Q
u		(1-P)Q/P	0	(1-P)Q/P	0
v			(1-P)Q/P	0	(1-P)Q/P
x				Q/P	Q
y					Q/P

Thus, for example, the regression of posttest on true ability is

$$E(y|x^*) = (C(x^*,y)/V(x^*))\ x^* = (Q/Q)\ x^* = 1\ x^*\ .$$

Comparing this with (1) we see the familiar result that measurement error attenuates slopes -- here P < 1.

We now suppose that the population is split into two groups -- one which receives the treatment, the other which does not. Let z be a binary variable which indicates whether or not an individual receives the treatment:

$$z = \begin{cases} 1 & \text{if received treatment (i.e. selected for experimental group)} \\ 0 & \text{if did not receive treatment (i.e. selected for control group).} \end{cases}$$

Further, we suppose that <u>the true effect of the treatment is nil</u>. In terms of the model, this amounts to saying that

$$(2) \qquad C(z, v) = 0 \ ,$$

or in other words, that a multiple regression of y on x* and z would yield a zero coefficient on z. (This particular choice of a baseline is for convenience only and involves no essential loss of generality.)

In practice, the multiple linear regression of y on x and z,

$$(3) \qquad E(y|x, z) = \alpha_o + \alpha_1 x + \alpha_2 z \ ,$$

will be run to assess the effect of the treatment. Since x is an erroneous measures of x*, this procedure may be biased -- α_2 may be nonzero.

Clearly, what is relevant is the selection procedure -- the basis on which individuals were assigned to the treatment and control groups. If the assignment had been random with respect to true ability, so that

$$C(z,x^*) = 0 \ ,$$

<u>and</u> random with respect to the error component of pretest, so that

$$C(z, u) = 0,$$

then no bias would result. For in this situation,

$$C(z,y) = C(z, x^* + v) = C(z, x^*) + C(z, v) = 0 ,$$

$$C(z,x) = C(z, x^* + u) = C(z, x^*) + C(z, u) = 0 .$$

Thus, the normal equations determining the regression slopes, namely

$$\begin{aligned} V(x) \quad \alpha_1 + C(x,z)\ \alpha_2 &= C(x, y) \\ C(x,z)\ \alpha_1 + V(z) \quad \alpha_2 &= C(z, y) , \end{aligned} \tag{4}$$

would specialize to

$$\begin{aligned} (Q/P) \quad \alpha_1 + \quad 0 \qquad \alpha_2 &= Q \\ 0 \qquad \alpha_1 + V(z) \quad \alpha_2 &= 0 , \end{aligned}$$

the solution to which is $\alpha_1 = P$, $\alpha_2 = 0$.

As Campbell and Erlebacher indicate, such randomization is unlikely to occur in non-laboratory situations. Our main objective in this paper is to evaluate the bias -- the discrepancy of α_2 from zero -- in two idealized cases. The two cases are:

Case (i). Selection on basis of true ability. All individuals whose true ability is below the average are assigned to the experimental group; all those whose true ability is above the average are assigned to the control group. In terms of our model

$$z = \begin{cases} 1 & \text{if } x^* \leq 0 \\ 0 & \text{if } x^* > 0 \ . \end{cases}$$

Case (ii). Selection on basis of pretest score. All individuals whose pretest score is below the average are assigned to the experimental group; all those whose pretest score is above the average are assigned to the control group. In terms of our model

$$z = \begin{cases} 1 & \text{if } x \leq 0 \\ 0 & \text{if } x > 0 \ . \end{cases}$$

Case (i) is a variation on the model used by Campbell and Erlebacher in their critique of Head Start evaluations. Our variation lies in splitting a single normal population rather than using two distinct normal distributions. Case (ii) seems very similar, but has strikingly different implications, as shown by Barnow (1972). Neither case corresponds literally to reality. For example, when true ability is unobserved, it can't really provide the basis for selection. But these two polar cases should suffice to clarify the issues.

For future reference, note that either selection procedure splits the population into two equal-sized groups, so that in both cases the marginal distribution of z is given by

$$p_o = \text{Prob } \{z = 0\} = 1/2 \ , \quad p_1 = \text{Prob } \{z = 1\} = 1/2 \ ,$$

whence

$$E(z) = p_o\,0 + p_1\,1 = 1/2$$

(5)

$$V(z) = p_o\,(0 - 1/2)^2 + p_1\,(1 - 1/2)^2 = 1/4\ .$$

We shall see that the two cases differ with respect to the covariances of z with the other variables.

3. Technical Digression

In general, the covariance of any variable z with another variable w can be computed by taking the covariance of z with the conditional expectations of w given z, that is

$$C(z,w) = C(z,\ E(w|z)) = E((z - E(z))E(w|z)).$$

If z is a binary variable taking on the values 0 and 1 with probabilities p_o and p_1 respectively we have

$$\begin{aligned} C(z,\ w) &= p_o\ (0 - E(z))\ E(w|0) + p_1\ (1 - E(z))\ E(w|1) \\ &= -\ p_o\ p_1\ [E(w|0) - E(w|1)]; \end{aligned}$$

where $E(w|0) \equiv E(w|z = 0)$ and $E(w|1) \equiv E(w|z = 1)$. In the present setting, $p_o = p_1 = 1/2$, so that

(6) $$C(z,\ w) = -(1/4)\ [E(w|0) - E(w|1)].$$

In words, the covariance of the treatment dummy with any variable is one-fourth of the difference between the mean of the variable in the experimental group and the mean of the variable in the control group.

To compute these group means and related measures, we draw on the following theorem:

Let $w \sim N(\mu, \sigma^2)$, that is, let the density function of w be

$$f(w) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp \{- \tfrac{1}{2} \frac{(w - \mu)^2}{\sigma^2}\} .$$

Then, given that $a < w < b$, the conditional density of w is

$$\text{(7)} \qquad p(w|a < w < b) = \begin{cases} 0 & \text{for } w \leq a \\ f(w)/(F(b) - F(a)) & \text{for } a < w < b \\ 0 & \text{for } b \leq w \text{ ;} \end{cases}$$

the conditional expectation of w is

$$\text{(8)} \qquad E(w|a < w < b) = \mu + \sigma^2 \frac{f(a) - f(b)}{F(b) - F(a)} \text{ ;}$$

and the conditional variance of w is

$$\text{(9)} \qquad V(w|a < w < b) = \sigma^2 \left\{ 1 + \frac{(a-\mu)f(a) - (b-\mu)f(b)}{F(b) - F(a)} - \sigma^2 \left[\frac{f(a) - f(b)}{F(b) - F(a)} \right]^2 \right\} .$$

Here F(t) denotes the cumulative normal distribution, i.e. $F(t) = \int_{-\infty}^{t} f(w)dw$.

The theorem is proven in Appendix A. A numerical tabulation which illustrates the formulas is provided in Appendix B.

For the purposes of this paper, we apply the theorem directly to the upper half of a normal distribution. Setting $a = \mu$ and $b = \infty$, we find the conditional density function given that w is above μ:

$$p(w|\mu < w) = p(w|\mu < w < \infty) = \begin{cases} 0 & \text{for } w \leq \mu \\ 2f(w) & \text{for } \mu < w \end{cases} ,$$

using $F(\mu) = 1/2$, $F(\infty) = 1$; the conditional expectation given that w is above μ:

$$E(w|\mu < w) = \mu + \sigma\sqrt{2/\pi}\ , \tag{10}$$

using also $f(\mu) = (2\pi\sigma^2)^{-\frac{1}{2}}$ and $f(\infty) = 0$; and the conditional variance given that w is above μ:

$$V(w|\mu < w) = \sigma^2(\pi - 2)/\pi \tag{11}$$

By symmetry, for the lower half of the normal distribution, we have:

$$p(w|w < \mu) = \begin{cases} 2\ f(w) & \text{for } w \leq \mu \\ 0 & \text{for } \mu < w \end{cases}\ ,$$

$$E(w|w < \mu) = \mu - \sigma\sqrt{2/\pi}\ ,$$

and

$$V(w|w < \mu) = \sigma^2(\pi - 2)/\pi\ .$$

Introducing the binary variable

$$z = \begin{cases} 0 & \text{if } \mu < w \\ 1 & \text{if } w < \mu \end{cases}$$

we write the conditional expectations and variances compactly as

$$E(w|z) = \mu + (1 - 2\ z)\sqrt{2/\pi}\ \sigma \tag{12}$$

$$V(w|z) = \sigma^2(\pi - 2)/\pi\ . \tag{13}$$

In conjunction with (6), the conditional expectation formula implies

$$C(z,w) = (-1/4)\ 2\sqrt{2/\pi}\ \sigma = -\sigma/\sqrt{2\pi} \tag{14}$$

which means that the correlation between z and w is

$$\rho_{zw} = C(z,w)/\sqrt{V(z)V(w)} = (-\sigma/\sqrt{2\pi})/\sqrt{(1/4)\sigma^2} = -\sqrt{2/\pi}\ .$$

As is to be expected, the value of this correlation is entirely independent of the values of μ and σ^2.

To summarize the results of our digression: We have split a normal population into two groups, those above and those below the mean; found the within-group means and variances; and also found the between-group variance (expressed in terms of the covariance of the normal variable with a binary variable depicting the split). In what follows, x^* and x will alternately take on the role of w.

4. Selection on Basis of True Ability

In case (i) the individuals are assigned to the control group or to the experimental group according as their true ability is above or below the mean true ability in the population. Thus x^* plays the role that w did in Section 3. Recalling that $V(x^*) = Q$, we set σ in (14) equal to $\sqrt{Q}$ and find

$$C(z,x^*) = -\sqrt{Q/(2\pi)}\ . \tag{15}$$

Furthermore, since z is determined exactly (i.e. nonstochastically) by x^*, it must be independent of u and v, which, as will be recalled, are independent of x^*. Thus

$$C(z,y) = C(z,\ x^* + v) = C(z,x^*) + C(z,v) = C(z,x^*) = -\sqrt{Q/(2\pi)},$$

$$C(z,x) = C(z,\ x^* + u) = C(z,x^*) + C(z,u) = C(z,x^*) = -\sqrt{Q/(2\pi)}.$$

This gives us the moments we need to compute the slopes in the multiple linear regression of y on x and z. Specifically, the normal equations (4) specialize to

$$\begin{aligned} (Q/P)\quad \alpha_1 &+ (-\sqrt{Q/(2\pi)})\ \alpha_2 = Q \\ (-\sqrt{Q/(2\pi)})\quad \alpha_1 &+ \quad (1/4)\quad \alpha_2 = -\sqrt{Q/(2\pi)}\ . \end{aligned}$$

The solution to these normal equations is

$$\alpha_1 = P(\pi - 2)/(\pi - 2P)\ ,\quad \alpha_2 = -(1 - P)\ \sqrt{8\pi Q}/(\pi - 2P).$$

The intercept can then be obtained as $\alpha_o = E(y) - \alpha_1\ E(x) - \alpha_2\ E(z) = -\alpha_2/2$ using $E(y) = E(x) = E(x^*) = 0$ and $E(z) = -1/2$. Note that with $0 < P < 1$. we have $\alpha_1 < P$, $\alpha_2 < 0$, and $0 < \alpha_0$ for $\pi = 3.14... > 2 > 2P$.

The solution value for α_2 shows that the selection procedure makes the coefficient of z a biased measure of the true effect of the treatment (which is zero). As the variance ratio P falls from 1 to 0, the value of α_2 falls monotonically from 0 to $-\sqrt{8Q/\pi} \doteq -1.6\sqrt{Q}$, that is the magnitude of the bias ($|\alpha_2|$) rises. The regression spuriously attributes to the treatment deleterious effects on posttest when in fact it had no effect whatsoever. The source of this bias lies in the selection procedure which assigned low-ability individuals to the experimental group and high-ability individuals to the control group. Those differences in true ability were manifested in differences in posttest scores. The treatment variable z gets a negative coefficient because it is proxying (inversely) for true ability. This is the essence of the Campbell-Erlebacher argument, and we see that it holds up in our case (i).

To round out the discussion of case (i), we consider what happens when linear regressions of posttest on pretest are run separately in each of the two groups. To obtain those regressions we require the within-group moments -- the conditional exceptions, variances, and covariance -- of posttest and pretest. First, applying (12)-(13) with $\mu = E(x^*) = 0$ and $\sigma^2 = V(x^*) = Q$, we obtain

$$E(x^*|z) = (1 - 2z)\sqrt{2Q/\pi} \tag{16}$$

$$V(x^*|z) = (\pi - 2)Q/\pi \tag{17}$$

Next, recalling that u and v are independent of x* and thus of z (which is an exact function of x*), we deduce

$$\begin{aligned} E(x|z) &= E((x^* + u)|z) = E(x^*|z) + E(u|z) = E(x^*|z) + E(u) \\ &= E(x^*|z) = (1 - 2z)\sqrt{2Q/\pi} = E(y|z) , \\ V(x|z) &= V((x^* + u)|z) = V(x^*|z) + V(u|z) = V(x^*|z) + V(u) \\ &= (\pi - 2)Q/\pi + (1 - P)Q/P = (\pi - 2P)Q/(\pi P) = V(y|z), \\ C(x,y|z) &= C((x^* + u, x^* + v)|z) = V(x^*|z) = (\pi - 2)Q/\pi. \end{aligned}$$

Thus, within either group, the slope of the linear regression of posttest on pretest will be

$$C(x,y|z)/V(x|z) = P(\pi - 2)/(\pi - 2P) ,$$

which is exactly α_1 , the coefficient of x in the overall multiple regression. The intercepts of course will differ, that in the control group being

$$E(y|0) - \alpha_1 E(x|0) = (1 - \alpha_1) \sqrt{2Q/\pi}$$

and that in the experimental group being

$$E(y|1) - \alpha_1 E(x|1) = -(1 - \alpha_1) \sqrt{2Q/\pi}.$$

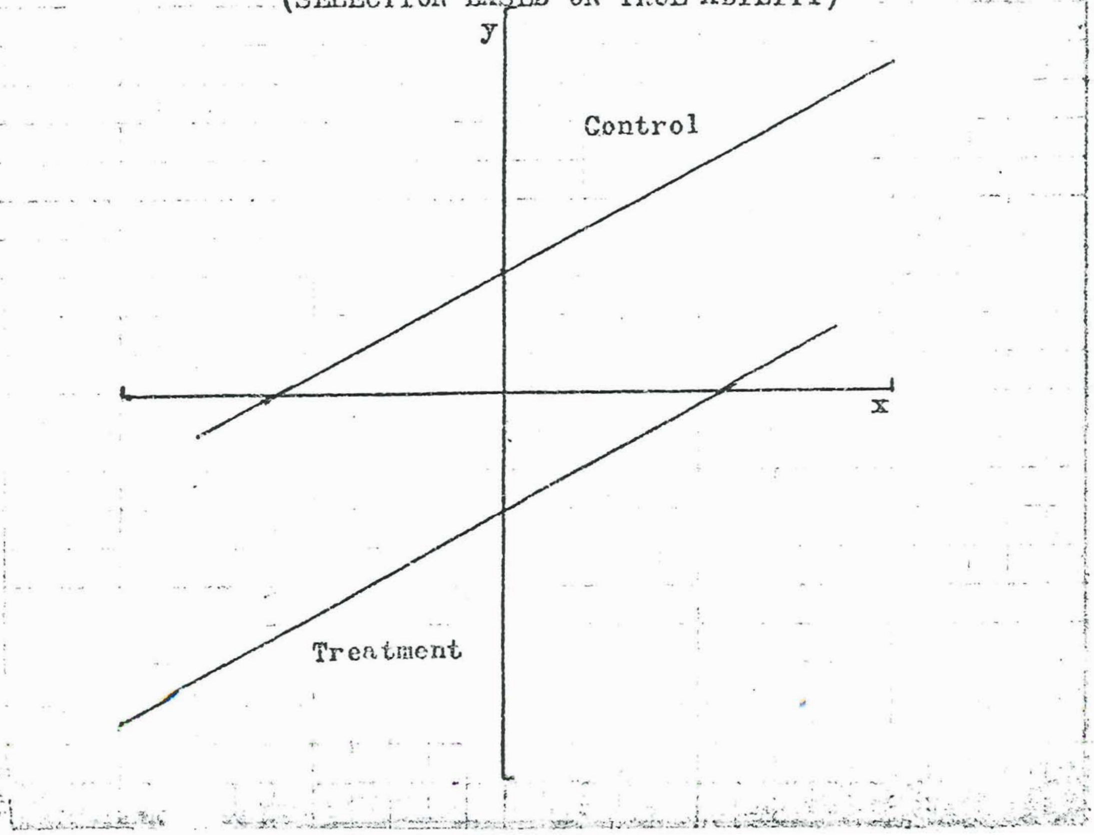
The difference between these two intercepts coincides with the coefficient of z in the overall multiple regression:

$$(1 - \alpha_1) (- \sqrt{2Q/\pi} - \sqrt{2Q/\pi}) = - 2 \frac{\pi(1 - P)}{\pi - 2 P} \sqrt{2Q/\pi} = - (1 - P) \sqrt{8\pi Q}/(\pi - 2P)$$

$$= \alpha_2 .$$

Thus the spurious effect of the treatment turns up again as a difference in the level of the two within-group linear regressions, as the diagram indicates.

WITHIN-GROUP REGRESSIONS OF POSTTEST ON PRETEST
(SELECTION BASED ON TRUE ABILITY)



We have relied on linear regression for this analysis, despite the fact that the true regression of y on x is nonlinear in the present situation, that is, $E(y|x, z)$ is nonlinear in x. This complication is examined in Appendix C and may be studied more closely in future work. But for the present, there is no reason to believe that it would change our qualitative conclusions. In any event, the results given above are still valid for the best linear approximation to the true conditional expectation function, which is presumably what is fitted in applied studies.

5. Selection on Basis of Pretest Score

In case (ii) the individuals are assigned to the control group or to the experimental group according as their pretest score (not their true ability) is above or below the mean pretest score in the population. Thus x (not x^*) plays the role that w did in Section 3. Recalling that $E(x) = 0$ and $V(x) = Q/P$, we set $\mu = 0$ and $\sigma = \sqrt{Q/P}$ in (12) and (14) to find

$$(18) \qquad E(x|z) = (1 - 2z)\sqrt{2Q/(\pi P)}$$

and

$$(19) \qquad C(z,x) = -\sqrt{Q/(2\pi P)}.$$

In the present case, z is determined exactly by x, hence it will depend on both x^* and u, but will remain independent of v.

To obtain $C(z,y)$, we proceed as follows. For the population at large, we know that

$$(20) \qquad E(u|x) = (1 - P)\, x\ ,$$

since $C(u,x) = V(u) = (1 - P)Q/P$ and $V(x) = Q/P$ together imply that in the regression of these joint-normal variables, the slope is $C(u,x)/V(x) = (1 - P)$, while $E(y) = 0 = E(x)$ implies that the intercept is zero. Since z is an exact function of x, it follows that

$$E(u|z) = E((E(u|x))|z) = E(((1 - P)x)|z) = (1 - P)\ E(x|z).$$

Consequently,

$$(21) \qquad \begin{aligned} E(x^*|z) &= E((x - u)|z) = E(x|z) - E(u|z) = P\ E(x|z) \\ &= (1 - 2z)\ \sqrt{2PQ/\pi}\ , \end{aligned}$$

and, since v is independent of z,

$$(22) \qquad \begin{aligned} E(y|z) &= E((x^* + v)|z) = E(x^*|z) + E(v|z) = E(x^*|z) + E(v) \\ &= E(x^*|z) = (1 - 2\ z)\ \sqrt{2PQ/\pi}\ . \end{aligned}$$

In conjunction with (6) this means that

$$(23) \qquad C(z,y) = (-1/4)\ 2\ \sqrt{2PQ/\pi} = -\ \sqrt{PQ/(2\pi)}\ .$$

Incidentally, (21) shows that the selection on the basis of pretest scores has made the two groups different in mean true ability, but a comparison of (21) with (15) shows that this difference is less (by the factor $\sqrt{P}$) than it was when the selection was strictly on the basis of true ability. In case (ii) the control group does not come entirely from the high-ability half of the population; it also includes low-ability individuals who happened to score unusually high on the pretest.

We now have the moments we need to compute the slopes in the multiple linear regression of y on x and z. The normal equations (4) specialize to

$$(Q/P)\ \alpha_1 + (-\sqrt{Q/(2\pi P)})\ \alpha_2 = Q$$

$$(-\sqrt{Q/(2\pi P)})\ \alpha_1 + \quad (1/4) \quad \alpha_2 = -\sqrt{PQ/(2\pi)}.$$

The solution to these normal equations is simply

$$\alpha_1 = P \quad , \qquad \alpha_2 = 0 \ ;$$

the intercept is $\alpha_o = E(y) - \alpha_1 E(x) - \alpha_2 E(z) = 0$.

The solution values show that the selection procedure <u>does</u> <u>not</u> make the coefficient of z a biased measure of the true effect of the treatment (which is zero). In striking contrast to case (i), the regression correctly attributes no effect to the treatment. Despite the fact that the selection procedure tended to assign low-ability individuals to the experimental group and high-ability individuals to the control group, no spurious effect arises. Despite the fact that z by itself is an (inverse) proxy for true ability, it fails to pick up, in the multiple regression, any credit for the effect of true ability on posttest.

The explanation for this serendipitous result is not hard to locate. Recall that z is completely determined by pretest score x. It cannot contain any information about x* that is not contained in x. Consequently, when we control on x as in the multiple regression, z has no explanatory power with respect to y. More formally, the partial correlation of y and z controlling on x vanishes although the simple correlation of y and z is nonzero.

To round out the discussion of case (ii), we consider what happens when separate regressions of y on x are run for the experimental and control groups. We already have the within-group means; it remains to find the within-group variances and covariance. For pretest, we can apply (13) directly, setting $\sigma^2 = V(x) = Q/P$ to find

$$V(x|z) = (\pi - 2)Q/(\pi P) \quad . \tag{24}$$

For posttest, the route is more roundabout; we start with the decomposition of the marginal variance of x* into its between- and within-group components:

$$V(x^*) = \underset{z}{V}\, E(x^*|z) + \underset{z}{E}\, V(x^*|z) \quad .$$

Using (21) we compute the between-group component:

$$\underset{z}{V}\, E(x^*|z) = p_o\, (E(x^*|0) - E(x^*))^2 + p_1\, (E(x^*|1) - E(x^*))^2$$
$$= (1/2)\ 2\ (2PQ/\pi) = (2PQ/\pi) \quad .$$

Since $V(x^*) = Q$, it follows that the expected conditional variance is

$$\underset{z}{E}\, V(x^*|z) = Q - (2\ PQ/\pi) = Q\ (1 - \frac{2P}{\pi}) = (\pi - 2P)Q/\pi.$$

By symmetry, this means that

$$V(x^*|z) = (\pi - 2P)Q/\pi.$$

Then, since v is independent of z and x*, we conclude that

$$V(y|z) = V((x^* + v)|z) = V(x^*|z) + V(v) = V(x^*|z) + (1 - P)Q/P$$
$$= Q\ (\frac{\pi-2P}{\pi} + \frac{1-P}{P}) = (\pi - 2P^2)Q/(\pi P).$$

Further, the marginal covariance of x* and x decomposes into

$$C(x^*,x) = \underset{z}{C}(E(x^*|z),\ E(x|z)) + \underset{z}{E}\ C(x^*,\ x|z).$$

Using (18) and (21) we compute the between-group component:

$$\begin{aligned}\underset{z}{C}(E(x^*|z),\ E(x|z)) &= p_o(E(x^*|0) - E(x^*))(E(x|0) - E(x)) \\ &\quad + p_1(E(x^*|1) - E(x^*))(E(x|1) - E(x)) \\ &= (1/2)2\ \sqrt{2PQ/\pi}\ \sqrt{2Q/(\pi P)} = 2Q/\pi.\end{aligned}$$

Since $C(x^*,x) = Q$, it follows by symmetry that

$$C(x^*,x|z) = Q - 2Q/\pi = Q\ (1 - 2/\pi) = (\pi - 2)Q/\pi.$$

Since v is independent of x and hence of z, we finally have

$$C(x,y|z) = C((x,\ x^* + v)|z) = C(x,x^*|z) = (\pi - 2)Q/\pi. \tag{25}$$

Taking together (24) and (25), we see that within either group, the slope of the linear regression of posttest on pretest is

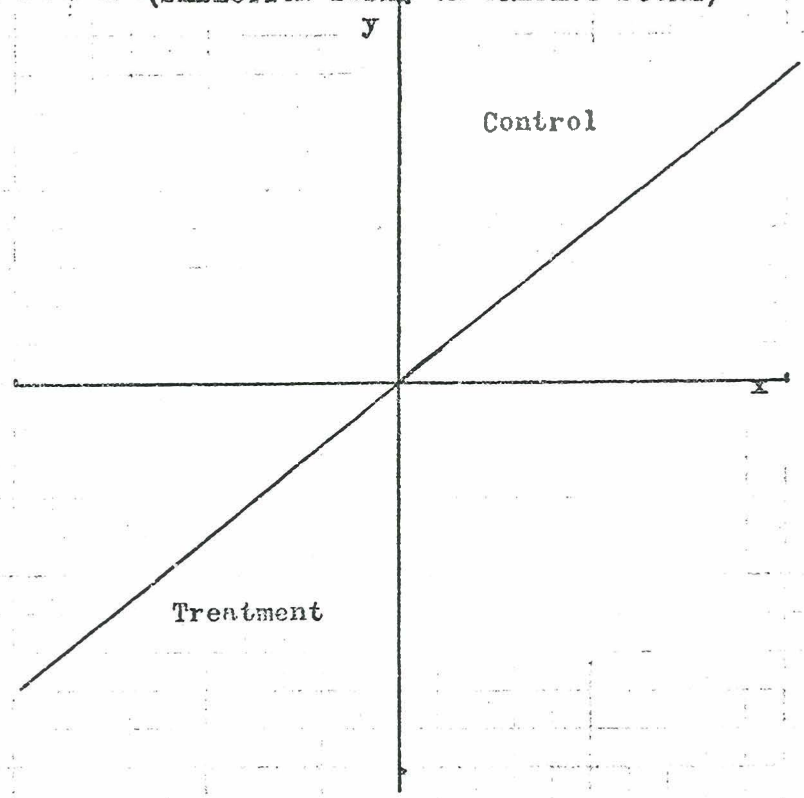
$$C(x,y|z)/V(x|z) = P\ ,$$

which coincides with the value for the coefficient of x in the overall multiple regression. The intercepts will also be the same, namely zero:

$$\begin{aligned}E(y|z) - P\ E(x|z) &= (1 - 2z)\ \sqrt{2PQ/\pi} - P\ (1 - 2z)\ \sqrt{2Q/(\pi P)} \\ &= (1 - 2z)0 = 0.\end{aligned}$$

As the diagram indicates, the two within-group regressions coincide with the overall regression, confirming the absence of a treatment effect.

WITHIN-GROUP REGRESSIONS OF POSTTEST ON PRETEST
(SELECTION BASED ON PRETEST SCORE)



The discussion in Lord and Novick (1968, pp. 141-147) provides a very simple derivation of the fact that no spurious treatment effect can arise in case (ii). Recall from (1) that for the population at large, the regression of posttest on pretest is linear. The same linear function holds over the entire range of x, and will be observed no matter what subrange of x we choose to observe, as long as we do not tamper with the conditional distribution of y given x. (Nothing in the usual regression model requires that the distribution of the explanatory variable be representative of its distribution over the entire population. The only requirement is that the conditional distribution of the dependent variable given the explanatory variable, be preserved.) For the within-group regressions, in case (ii) we have simply selected a range of x, we have not tampered with the distribution of y given x. Therefore within each group we must get the same regression as we get overall. This argument, incidentally, demonstrates that the true regression of y on x and z is linear (in contrast to case (i)). Note that this is true even within groups, where the distributions of y and x are clearly nonnormal.

Lord and Novick (1968, pp. 143-144) call attention to the fact that correlation coefficients, unlike regression coefficients, are sensitive to selection on the independent variable. In the present case, the overall correlation between x and y is

$$\rho_{xy} = C(x,y)/\sqrt{V(x)V(y)} = Q/\sqrt{(Q/P)(Q/P)} = P ,$$

while their within-group (i.e. partial given z) correlation is

$$\rho_{xy.z} = C(x,y|z)/\sqrt{V(x|z)V(y|z)} = P\sqrt{(\pi - 2)/(\pi - 2P^2)}.$$

Since $0 < P < 1$, we see that $\rho_{xy.z} < \rho_{xy}$, as might be expected.

6. Efficiency

The basic results for cases (i) and (ii) may be brought together in the following table, along with those for the random selection procedure discussed at the end of Section 3 and identified here as case (o):

Variances and covariances

	y	x	(o) z	(i) z	(ii) z
y	Q/P	Q	0	$-\sqrt{Q/(2\pi)}$	$-\sqrt{PQ/(2\pi)}$
x		Q/P	0	$-\sqrt{Q/(2\pi)}$	$\sqrt{Q/(2\pi P)}$
z			1/4	1/4	1/4
Regression coefficients					
α_1			P	$P(\pi-2)/(\pi-2P)$	P
α_2			0	$-(1-P)\sqrt{8\pi Q}/(\pi-2P)$	0

We are reminded that case (ii) -- selection on basis of pretest-- produces the same unbiased regression results as case (o) -- purely random selection. But we should not conclude that the two procedures are equally desirable. Recall that our analysis has been couched in terms of population parameters, so that sampling variability has been ignored. For finite samples, the case (ii) regression will remain unbiased, but as we now show, is subject to more sampling variability than the case (o) regression. That is, random selection provides a more efficient experimental design.

Inspecting the table we see that in case (o) the explanatory variables are uncorrelated, that is, $C(z,x) = 0$, while in case (ii) they are correlated, $C(z,x) \neq 0$. This suggests that the standard errors of the regression coefficient estimates are larger in the latter case. Indeed, the usual formula for regression on two explanatory variables shows that the variance of regression coefficient estimators is multiplied by a factor $1/(1 - \rho^2)$ in moving from an uncorrelated to a correlated design, where ρ^2 is the squared correlation of the explanatory variables; cf. Kmenta (1971, p. 388). In our case (ii), the relevant ρ^2 is

$$\rho_{xz}^2 = C^2(x,z)/(V(x)V(z)) = (Q/(2\pi P))/((Q/P)(1/4)) = 2/\pi,$$

implying that the sampling variances of the regression coefficients in case (ii) will be

$$1/(1 - (2/\pi)) = \pi/(\pi - 2) \doteq 2.75$$

times as large as they are in case (o). (It is interesting that this numerical conclusion is entirely independent of the values of the parameters of the model, P and Q.) Thus the efficiency of the random selection procedure is confirmed -- a random sample of size 100 being as good as a selected-on-pretest sample of size 275.

In the preceding calculation we relied implicitly on the assumption that the disturbance variance did not change with the change in experimental design. This assumption is justified since in both cases $V(y) = Q/P$ and $E(y|x, z) = E(y|x) = Px$, which implies that

$$\begin{aligned} V(y|x,z) = V(y|x) &= V(y) - \underset{x}{V}\,E(y|x) = Q/P - P^2\,V(x) \\ &= Q/P - P^2(Q/P) = (1 - P^2)(Q/P) \end{aligned}$$

in both cases.

Appendix A. Conditional Density and Moments

To evaluate the density, expectation, and variance conditional on a normally distributed variable lying within a specified interval, we begin the case of standard normal distribution.

Let $s \sim N(0, 1)$, that is let the density function of s be:

$$f^*(s) = (2\pi)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2} s^2\}.$$

The probability that s lies in the interval between a and b is $F^*(b) - F^*(a)$, where

$$F^*(s) = \int_{-\infty}^{s} f^*(r)\,dr = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{s} \exp\{-\tfrac{1}{2} r^2\}\,dr$$

denotes the cumulative standard normal distribution. Therefore the conditional distribution of s given that $a < s < b$ is given by the density function

$$\text{(A1)} \qquad p^*(s|.) \equiv p^*(s|a < s < b) = \begin{cases} 0 & \text{for } s \leq a \\ f^*(s)/(F^*(b) - F^*(a)) & \text{for } a < s < b \\ 0 & \text{for } \quad b \leq s\,. \end{cases}$$

The moment-generating function for this distribution is

$$\text{(A2)} \qquad m(t) \equiv E(e^{ts}) = \int_{-\infty}^{\infty} e^{ts}\, p^*(s|.)\,ds$$

$$= \frac{1}{F^*(b) - F^*(a)} (2\pi)^{-\frac{1}{2}} \int_{a}^{b} \exp\{ts\} \exp -\{\tfrac{1}{2} s^2\} ds.$$

Completing the square in the exponent via

$$ts - \tfrac{1}{2}s^2 = -\tfrac{1}{2}(s - t)^2 + \tfrac{1}{2}t^2$$

we rewrite (A1) as

$$(F^*(b) - F^*(a))\ m(t) = \exp\{\tfrac{1}{2}t^2\}\ [(2\pi)^{-\frac{1}{2}} \int_a^b \exp\{-\tfrac{1}{2}(s - t)^2\} ds].$$

The term in square brackets will be recognized as the probability that a $N(t, 1)$-variable lies in the interval between a and b, which is equal to the probability that a $N(0, 1)$-variable lies in the interval between $a - t$ and $b - t$, namely $F^*(b - t) - F^*(a - t)$.

Thus

$$(F^*(b) - F^*(a))\ m(t) = \exp\{\tfrac{1}{2}t^2\}\ (F^*(b - t) - F^*(a - t)).$$

Differentiating with respect to t gives

$$\begin{aligned}(F^*(b) - F^*(a))\ m'(t) &= t \exp\{\tfrac{1}{2}t^2\}\ (F^*(b - t) - F^*(a - t)) \\ &\quad + \exp\{\tfrac{1}{2}\ t^2\}\ (-f^*(b - t) + f^*(a - t))\end{aligned}$$

using $F^{*\prime}(s) = f^*(s)$. Setting $t = 0$ to generate the first moment we find

$$(F^*(b) - F^*(a))\ m'(0) = (-f^*(b) + f^*(a));$$

that is

$$\text{(A2)} \qquad E(s|a < s < b) = m'(0) = \frac{f^*(a) - f^*(b)}{F^*(b) - F^*(a)}.$$

Differentiating a second time with respect to t gives

$$
\begin{aligned}
(F^*(b) - F^*(a))\, m''(t) = {} & t \exp\{\tfrac{1}{2} t^2\} (-f^*(b - t) + f^*(a - t)) \\
& + t^2 \exp\{\tfrac{1}{2} t^2\} (F^*(b - t) - F^*(a - t)) \\
& + \exp\{\tfrac{1}{2} t^2\} (F^*(b - t) - F^*(a - t)) \\
& + \exp\{\tfrac{1}{2} t^2\} (-(b-t)f^*(b-t) + (a-t)f^*(a-t)) \\
& + t \exp\{\tfrac{1}{2} t^2\} (- f^*(b - t) + f^*(a - t))
\end{aligned}
$$

using $f^{*\prime}(s) = - s\, f^*(s)$. Setting t equal to zero to generate the second moment, we find

$$(F^*(b) - F^*(a))\, m''(0) = (F^*(b) - F^*(a)) + (a\, f^*(a) - b\, f^*(b));$$

that is,

$$\text{(A3)} \qquad E(s^2 | a < s < b) = m''(0) = 1 + \frac{a\, f^*(a) - b\, f^*(b)}{F^*(b) - F^*(a)}$$

The variance could now be computed as

$$V(s|a < s < b) = E(s^2|a < s < b) - E^2(s|a < s < b) = m''(0) - (m'(0))^2.$$

Proceeding to the general case, let $w \sim N(\mu, \sigma^2)$, that is the density function of w is

$$f(w) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{ -\tfrac{1}{2} \left(\frac{w - \mu}{\sigma}\right)^2 \right\},$$

and the cumulative distribution of w is

$$F(w) = \int_{-\infty}^{w} f(r)\, dr = (2\pi\sigma^2)^{-\frac{1}{2}} \int_{-\infty}^{w} \exp\left\{ -\tfrac{1}{2} \left(\frac{r - \mu}{\sigma}\right)^2 \right\} dr.$$

Note that

$$\text{(A4)} \qquad f(w) = (1/\sigma)\, f^*((w - \mu)/\sigma), \quad F(w) = F^*((w - \mu)/\sigma),$$

where $f^*(.)$ and $F^*(.)$ are the standardized functions defined above. We introduce the standardized variable $s = (w - \mu)/\sigma$. The event "$a < w < b$" is identical with the event "$a^* < s < b^*$", where

$$a^* = (a - \mu)/\sigma \,, \qquad b^* = (b - \mu)/\sigma.$$

Therefore the probability that $a < w < b$ is identical with the probability that $a^* < s < b^*$, namely $F^*(b^*) - F^*(a^*)$. From (A1) and (A4) it follows that the conditional probability distribution of w given that $a < w < b$ is given by the density function

$$p(w \mid a < w < b) = \begin{cases} 0 & \text{for } w \leq a \\ f(w)/(F(b) - F(a)) & \text{for } a < w < b \\ 0 & \text{for } b \leq w \,. \end{cases}$$

This is equation (7) in the text. Further, with $w = \mu + \sigma s$ everywhere, it must be true that for any event

$$E(w|.) = \mu + \sigma\, E(s|.).$$

Specifically

$$E(w|a < w < b) = \mu + \sigma\, E(s|a^* < s < b^*) = \mu + \sigma \frac{f^*(a^*) - f^*(b^*)}{F^*(b^*) - F^*(a^*)}$$

$$= \mu + \sigma \frac{\sigma f(a) - \sigma f(b)}{F(b) - F(a)}$$

$$= \mu + \sigma^2 \frac{f(a) - f(b)}{F(b) - F(a)} \,, \tag{A5}$$

using (A2) and (A4). This is equation (8) in the text. Similarly, it must be true that for any event,

$$E(w^2|.) = \sigma^2 E(s^2|.) + 2\,\mu\,\sigma\, E(s|.) + \mu^2 ,$$

so

$$V(w|.) = E(w^2|.) - E^2(w|.) = \sigma^2 \left[E(s^2|.) - E^2(s|.)\right].$$

Specifically,

$$V(w|a < w < b) = \sigma^2 \left[E(s^2|a^* < s < b^*) - E^2(s|a^* < s < b^*)\right]$$

$$= \sigma^2 \left\{ 1 + \frac{a^*f^*(a^*) - b^*f^*(b^*)}{F^*(b^*) - F^*(a^*)} - \sigma \left[\frac{f^*(a^*) - f^*(b^*)}{F^*(b^*) - F^*(a)} \right]^2 \right\}$$

(A6) $$= \sigma^2 \left\{ 1 + \frac{(a - \mu)f(a) - (b - \mu)f(b)}{F(b) - F(a)} - \sigma^2 \left[\frac{f(a) - f(b)}{F(b) - F(a)} \right]^2 \right\}$$

using (A2), (A3), and (A4). This is equation (9) in the text.

Appendix B. Illustration of Conditional Moments

The following tabulation may serve to illustrate the consequences of selecting a subpopulation from a normal distribution. Constructed for a standard normal distribution $s \sim N(0, 1)$, the table indicates for various values of a, the probability that a random drawing exceeds a, namely $1 - F^*(a)$; the conditional expectation given that it exceeds a, namely

$$E(s|a < s) = f^*(a)/(1 - F^*(a));$$

and the conditional variance given that it exceeds a, namely

$$V(s|a < s) = 1 + \frac{a\, f^*(a)}{1 - F^*(a)} - \left[\frac{f^*(a)}{1 - F^*(a)}\right]^2$$

$$= 1 + a\, E(s|a < s) - E^2(s|a < s)$$

$$= 1 - E(s|a < s)\, (E(s|a < s) - a)).$$

The formulas here are obtained by taking $b = \infty$ in (A2)-(A3).

Cutoff point a	Probability of Selection	Conditional Expectation	Conditional Variance
$-\infty$	1.000	0	1.00
-2	.977	.06	.88
-1	.841	.29	.62
$-.5$	.692	.51	.49
0	.500	.80	.36
.5	.308	1.14	.27
1	.159	1.52	.21
2	.023	2.38	.10

Our table may be compared with that in Lord and Novick (1968, p. 141), which uses different cutoff points, reports conditional standard deviations rather than variances, and does not report conditional expectations.

Appendix C. Exact Regressions When Selection is Based on True Ability

To develop the exact (nonlinear) regression functions of y on x when selection is based on true ability we proceed as follows. For typographical convenience in this appendix we denote x* by w and V(u) by

$$R = (1 - P)Q/P.$$

For the population at large we have $w \sim N(0, Q)$ and $x|w \sim N(w, R)$. The respective densities are

$$f(w) = (2\pi Q)^{-\frac{1}{2}} \exp\left\{ -\tfrac{1}{2} \frac{w^2}{Q} \right\},$$

$$p(x|w) = (2\pi R)^{-\frac{1}{2}} \exp\left\{ -\tfrac{1}{2} \frac{(x - w)^2}{R} \right\}.$$

For the lower half of the true ability distribution (i.e. the treatment group, for whom z = 1), the density of w is

$$p(w|1) = \begin{cases} 2\ f(w) & \text{for } w < 0 \\ 0 & \text{for } 0 \leq w , \end{cases}$$

while the density of x|w is still p(x|w). Thus the joint density of x and w is

$$\text{(C1)} \qquad p(x,w|1) = p(x|w)\ p(w|1) = \begin{cases} 2\ p(x|w)\ f(w) & \text{for } w < 0. \\ 0 & \text{for } 0 \leq w . \end{cases}$$

Now

$$p(x|w)\ f(w) = (2\pi R)^{-\frac{1}{2}}(2\pi Q)^{-\frac{1}{2}} \exp\left\{ -\tfrac{1}{2}\left[\frac{(x-w)^2}{R} + \frac{w^2}{Q}\right]\right\}$$

$$= (2\pi Q/P)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\ \frac{Px^2}{Q}\right\} (2\pi(1-P)Q)^{-\frac{1}{2}} \exp\left\{-\ \tfrac{1}{2}\frac{(w-Px)^2}{(1-P)Q}\right\} \qquad \text{(C2)}$$

where we have completed the square in the exponent via

$$\frac{(x-w)^2}{R} + \frac{w^2}{Q} = \frac{(Q+R)}{QR}\ \left(w - \frac{Qx}{Q+R}\right)^2 + \frac{1}{R}\ \left(1 - \frac{Q}{Q+R}\right)\ x^2$$

$$= \frac{(w-P\ x)^2}{(1-P)Q} + \frac{Px^2}{Q}\ ,$$

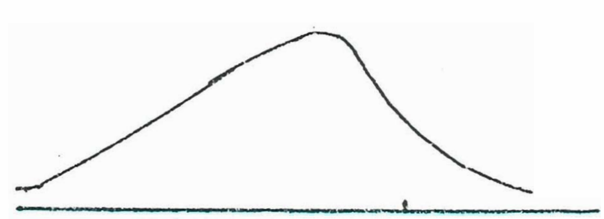
using the definition of R. Thus $p(x,w|1)$, the joint density of x and w in the treatment group, is zero for $0 \leq w$, and is twice the expression in (C2) for $w < 0$.

The density of x in the treatment group can now be obtained as

$$p(x|1) = \int_{-\infty}^{\infty} p(x,\ w|1)\ dw = \int_{-\infty}^{0} 2p(x|w)\ f(w)\ dw$$

$$= 2\ (2\pi Q/P)^{-\frac{1}{2}} \exp\left\{-\ \tfrac{1}{2}\frac{Px^2}{Q}\right\}\ (2\pi(1-P)Q)^{-\frac{1}{2}} \int_{-\infty}^{0} \exp\left\{-\tfrac{1}{2}\frac{(w-Px)^2}{(1-P)Q}\right\}dw \qquad \text{(C3)}$$

This distribution of pretest scores in the treatment group, sketched in the diagram below, is clearly nonnormal. **(High pretest scores are a rare** phenomenon in the treatment group, which by construction, has no high-ability individuals.)



From (C1) - (C3), the conditional density of w given x in the treatment group now follows:

$$p(w|x,1) = p(x,w|1)/p(x|1) = \begin{cases} f_x(w)/F_x(0) & \text{for } w < 0 \\ 0 & \text{for } 0 \leq w \end{cases}$$

where

$$f_x(w) = (2\pi(1 - P)Q)^{-\frac{1}{2}} \exp\{- \tfrac{1}{2} \frac{(w - Px)^2}{(1 - P)Q}\}$$

$$F_x(w) = \int_{-\infty}^{w} f_x(r)\, dr .$$

We recognize $f_x(w)$ as the density of a $N(Px, (1 - P)Q)$- variable, which means that $f_x(w)/F_x(0)$ is the conditional density of such a variable given that the variable is less than zero. Consequently,

$$E(w|x, 1) = \int_{-\infty}^{\infty} w\, p(w|x, 1)\, dw$$

$$= \int_{-\infty}^{0} (w\, f_x(w)/F_x(0))\, dw = (1/F_x(0)) \int_{-\infty}^{0} w\, f_x(w)\, dw$$

is the expected value of a $N(Px, (1 - P)Q)$ variable given that the variable is less than zero. Applying (A5) we find

$$E(w|x,\ 1) = P\ x + (1 - P)Q\ \frac{f_x(-\infty) - f_x(0)}{F_x(0) - F_x(-\infty)}$$

$$= P\ x - (1 - P)Q\ f_x(0)/F_x(0)\ . \tag{C4}$$

As in Appendix A, let $f^*(.)$ and $F^*(.)$ denote the standard normal density and cumulative functions respectively. Then using (A4), write

$$f_x(w) = (1/\sqrt{(1 - P)Q})\ f^*((w - Px)/\sqrt{(1 - P)Q}\ ),$$

and

$$F_x(w) = F^*((w - Px)/\sqrt{(1 - P)Q}\ ).$$

Introduce the transformation

$$s = P\ x/\sqrt{(1 - P)Q}\ , \tag{C5}$$

and write

$$f_x(0)/F_x(0) = (1/\sqrt{(1 - P)Q}\ )\ f^*(-s)/F^*(-s)$$

$$= (1/\sqrt{(1 - P)Q}\ )\ f^*(s)/(1 - F^*(s))\ , \tag{C6}$$

using $f^*(-s) = f^*(s)$ and $F^*(-s) = 1 - F^*(s)$. Finally, inserting (C6) into (C4) we have

$$E(w|x,1) = P\ x - \sqrt{(1 - P)Q}\ f^*(s)/(1 - F^*(s))\ ,$$

as the regression function of true ability on pretest in the treatment goup. By symmetry, the regression function of true ability on pretest in the <u>control</u> group (i.e. for $z = 0$) must be

$$E(w|x,0) = P\ x + \sqrt{(1 - P)Q}\ \ f^*(s)/F^*(s).$$

That these last two equations are also the regression of posttest on pretest for the two groups follows from the fact that $y = x^* + v$ with v independent of x^* (and hence of z). The shape of these curves is not hard to characterize. Consider the treatment group. As $x \to -\infty$, $s \to -\infty$, so $f^*(s)/(1 - F^*(s)) \to f^*(-\infty)/(1 - F^*(-\infty)) = 0/(1 - 0) = 0$, which means that $E(y|x,1) \to Px$. On the other hand, using L'Hospital's rule: $\lim f^*(s)/(1 - F^*(s)) = \lim f^{*\prime}(s)/(- F^{*\prime}(s)) = \lim (-s\ f^*(s))/(- f^*(s)) = \lim (-s)$. So as $x \to \infty$, $s \to \infty$, and $f^*(s)/(1 - F^*(s)) \to -\infty$, which means that $E(y\ x, |1)$ is asymptotic to the horizontal axis. For the control group, we have the mirror image, as sketched in the diagram on the next page.

The fact that $E(y|x,z)$ is always negative for $z = 1$ and always positive for $z = 0$ is an automatic consequence of the fact that the selection procedure kept true ability always negative for the treatment group and always positive for the control group.

We see that the exact regressions are nonlinear in x and that the spurious treatment effect shows up in a non-additive manner. What can be said about the spurious treatment effect in the exact (nonlinear) regressions as compared with the spurious treatment effect in the approximate (linear) regressions? In Section 4 we saw that the control group line was parallel to the treatment group line and lay above it by the <u>constant</u> amount

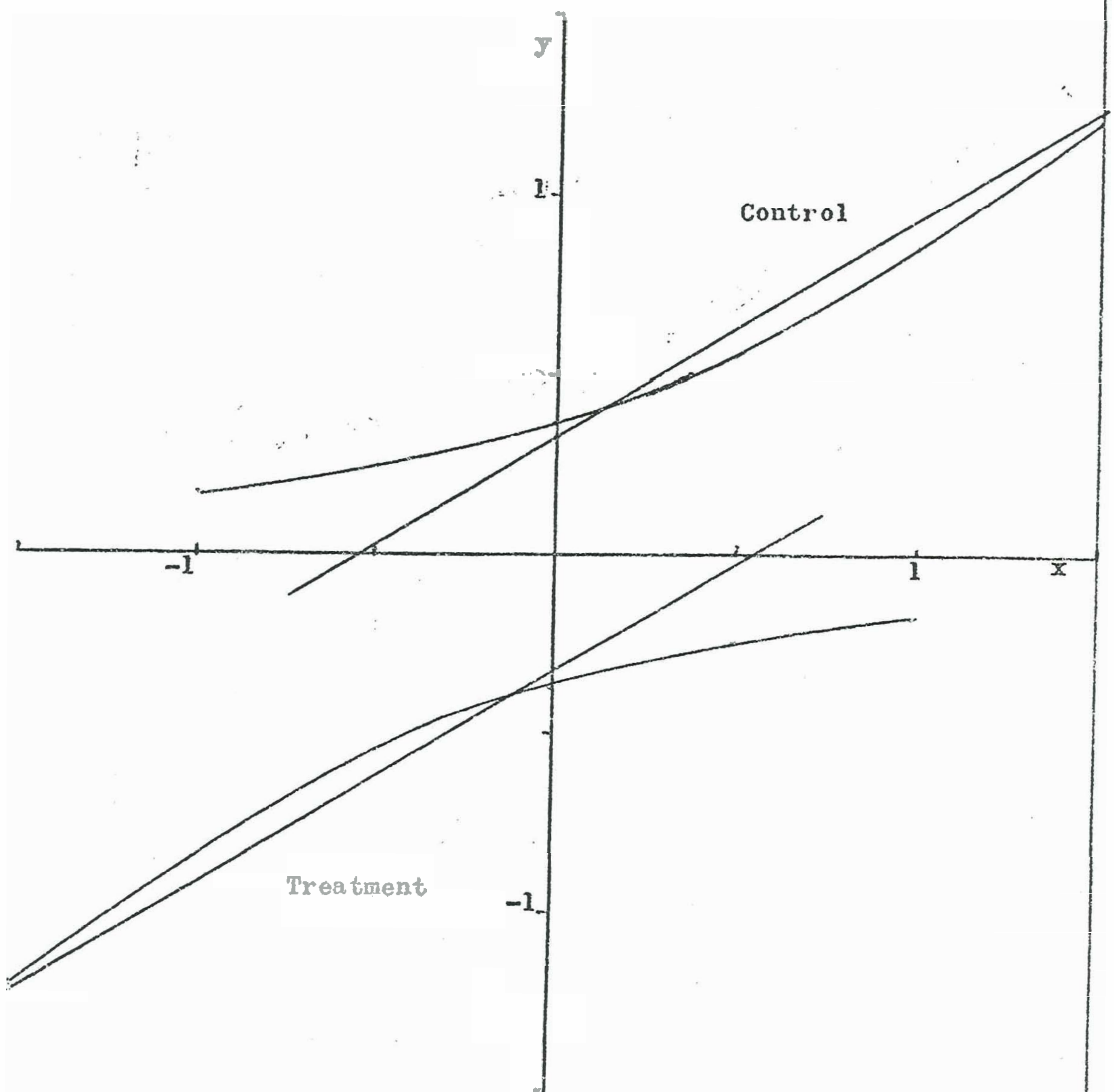
$$- \alpha_2 = (1 - P)\ \sqrt{8\pi Q}/(\pi - 2P) = \bar{h}\ ,$$

say. Now we see that the control group curve lies above the treatment group curve by the <u>variable</u> amount

EXACT AND APPROXIMATE WITHIN-GROUP REGRESSIONS OF POSTTEST ON PRETEST

(SELECTION BASED ON TRUE ABILITY)

$Q = 1$ $P = .8$



$$\sqrt{(1-P)Q}\ f^*(s)\ \left(\frac{1}{F^*(s)}+\frac{1}{1-F^*(s)}\right) = \sqrt{(1-P)Q}\ f^*(s)/(F^*(s)(1-F^*(s)))$$

$$= h(x)\ ,$$

say. In particular, at $x = 0$, the distance between the curves is

$$h(0) = \sqrt{(1-P)Q}\ (2\pi)^{-\frac{1}{2}}/((\tfrac{1}{2})(\tfrac{1}{2})) = \sqrt{8Q(1-P)/\pi}$$

$$= \bar{h}\ (\pi - 2P)/(\pi\sqrt{(1-P)} = \bar{h}\ g(P),$$

say. For $0 < P < .67$, we find that $.96 < g(P) < 1$, so that $h(0)$ is slightly less than, but virtually indistinguishable from, $\bar{h}$. For $.67 \leq P < .80$, we find $1 < g(P) < 1.10$, so that $h(0)$ is slightly greater than $\bar{h}$. Then for $.80 \leq P < 1$, $g(P)$ continues to rise and $h(0)$ becomes substantially greater than $\bar{h}$. The picture is somewhat mixed, but on balance it seems that the linear regressions may show less of a spurious treatment effect than the curvilinear regressions. Admittedly, it is not clear that $x = 0$ is a sensible point at which to compare the treatment and control curves.

As noted at the end of Section 5, the complication of curvilinearity does not arise when selection is on the basis of pretest scores rather than true ability.

REFERENCES

Barnow, B. S., "Conditions for the Presence or Absence of a Bias in Treatment Effect: Some Statistical Models for Head Start Evaluation," Institute Discussion Paper 122-72.

Campbell, D. T., and A. Erlebacher, "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful," in J. Hellmuth, editor, Compensatory Education: A National Debate Vol. III of The Disadvantaged Child, New York: Brunner Mazel, 1970.

Kmenta, J., Elements of Econometrics, New York: McGraw-Hill, 1971.

Lord, F. M., and M. R. Novick, Statistical Theories of Mental Test Scores, Reading, Pennsylvania: Addison-Wesley, 1968.