ABSTRACT
        This self-contained and self-instructional unit is
intended for use by evaluation and development personnel and by
students in introductory research and evaluation courses. The unit
contains a discussion of the regression employing graphic
illustrations with actual data. The user is introduced to the
regression effect in the single group pretest-posttests design, after
which he responds to mastery test instructional exercises. The second
part illustrates how the regression effect confounds the matched-pair
type of design and this, too, is followed by mastery test
instructional exercises. The user should be familiar with the basic
statistical concepts of mean, standard deviation, correlation, and
z-scores. (Author/SE)

INSTRUCTIONAL MODULE ON
REGRESSION AND THE MATCHING FALLACY
.IN QUASI-EXPERIMENTAL RESEARCH

Kenneth D. Hopkins
Laboratory of Educational Research
University of Colorado

September, 1973

NCERD Reporting Form — Developmental Products

1. Name of Product	2. Laboratory or Center	3. Report Preparation
Instructional Module on Regression and the Matching Fallacy in Quasi.-Experimental Research.	Laboratory of Educational Research, University of Colorado	Date prepared 11/9/73 Reviewed by K.D. Hopkins, director

4. Problem: *Description of the educational problem this product designed to solve.*

Many research and evaluation studies yield misleading, erroneous, or misinterpreted finding due to the failure to recognize the regression and matched-groups fallacy. This module·is designed to develop the competencies needed to identify situations in which the regression effect confounds results.

5. Strategy: *The general strategy selected for the solution of the problem above.*

The training materials present a conceptual, non-mathematical treatment of the regression phenomenon using graphical·illustration with actual data. Self-instructional exercises are included that can also be used as a pretest.

6. Release Date: *Approximate date product was (or will be) ready for release to next agency.*	7. Level of Development: *Character-istic level (or projected level) of development of product at time of release. Check one.*	8. Next Agency: *Agency to whom product was (or will be) released for further development diffusion.*
12/1/73	___ *Ready for critical review and for preparation for Field Test (i.e. prototype materials)* X *Ready for Field Test* ___ *Ready for publisher modification* ___ *Ready for general dissemination/ diffusion*	NIE

10.71.A (D)

9. Product Description: *Describe the following; number each description.*

- 1. *Characteristics of the product.*
- 2. *How it works.*
- 3. *What it is intended to do.*
- 4. *Associated products, if any.*
- 5. *Special conditions, time, training, equipment and/or other requirements for its use.*

Characteristics of the Product:

An 18 page discussion of the regression and matched-groups fallacy employing graphic illustrations with actual data. The module is self-contained and self-instructional.

How it Works:

The user is introduced to the regression effect in the single group pretest-posttest design, after which he responds to mastery test instructional exercises.
The second part illustrates how the regression effect confounds the matched-pair type of design, followed by mastery test instructional exercises.

What it is Intended to do:

Provide the user with recognition of situations in which the regression effect confounds results.

Requirements for Use:

User should be familiar with the basic statistical concepts of mean, standard deviation, correlation, and z-scores.

10. Product Users: *Those individuals or groups expected to use the product.*

The product is to be used by evaluators and consumers of research and evaluation reports as well as students in related courses.

11. Product Outcomes: *The changes in user behavior, attitudes, efficiency, etc. resulting from product use, as supported by data. Please cite relevant support documents. If claims for the product are not yet supported by empirical evidence please so indicate.*

Twenty-eight users responding to questions pertaining to the instructional quality of the module, the error rate for the programmed learning, and whether or not the materials were superflous (duplicated other equally-good sources).
The results are given below:

Instructional value?: "Poor": 0%; "fair": 7%; "good": 46%; "very good": 46%.

Median error rate: 10%.

Materials superflous: "Yes": 15%; "no": 85%.

The rating of "good" or "very good" by 92% of the users suggests instructional value for this module.

12. Potential Educational Consequences: *Discuss not only the theoretical (i.e. conceivable) implications of your product but also the more probable implications of your product, especially over the next decade.*

1. Fewer research and evaluation studies vulnerable to the confounding effects of regression.

2. Recognition by consumers of research and evaluation reports of regression fallacy where it exists and, hence, fewer misinterpretations of findings.

-3-

13. Product Elements: List the elements which constitute the product.	14. Origin: Circle the most appropriate letter.
One self-contained module including narrative and graphic illustrations, with instructional exercises intersperced.	ⓓ M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D M A
	D= Developed M= Modified A= Adopted

15. Start-up Costs: _Total expected costs to procure, install and initiate use of the product._

Reproduction costs only.

16. Operating Costs: _Projected costs for continuing use of product after initial adoption and installation (i.e., fees, consumable supplies, special staff, training, etc.)._

Reproduction only.

17. Likely Market: _What is the likely market for this product? Consider the size and type of the user group; number of possible substitute (competitor) products on the market; and the likely availability of funds to purchase product by (for) the product user group._

Evaluation and development personnel, especially those who are being trained on the job at regional labs and at state departments of education.
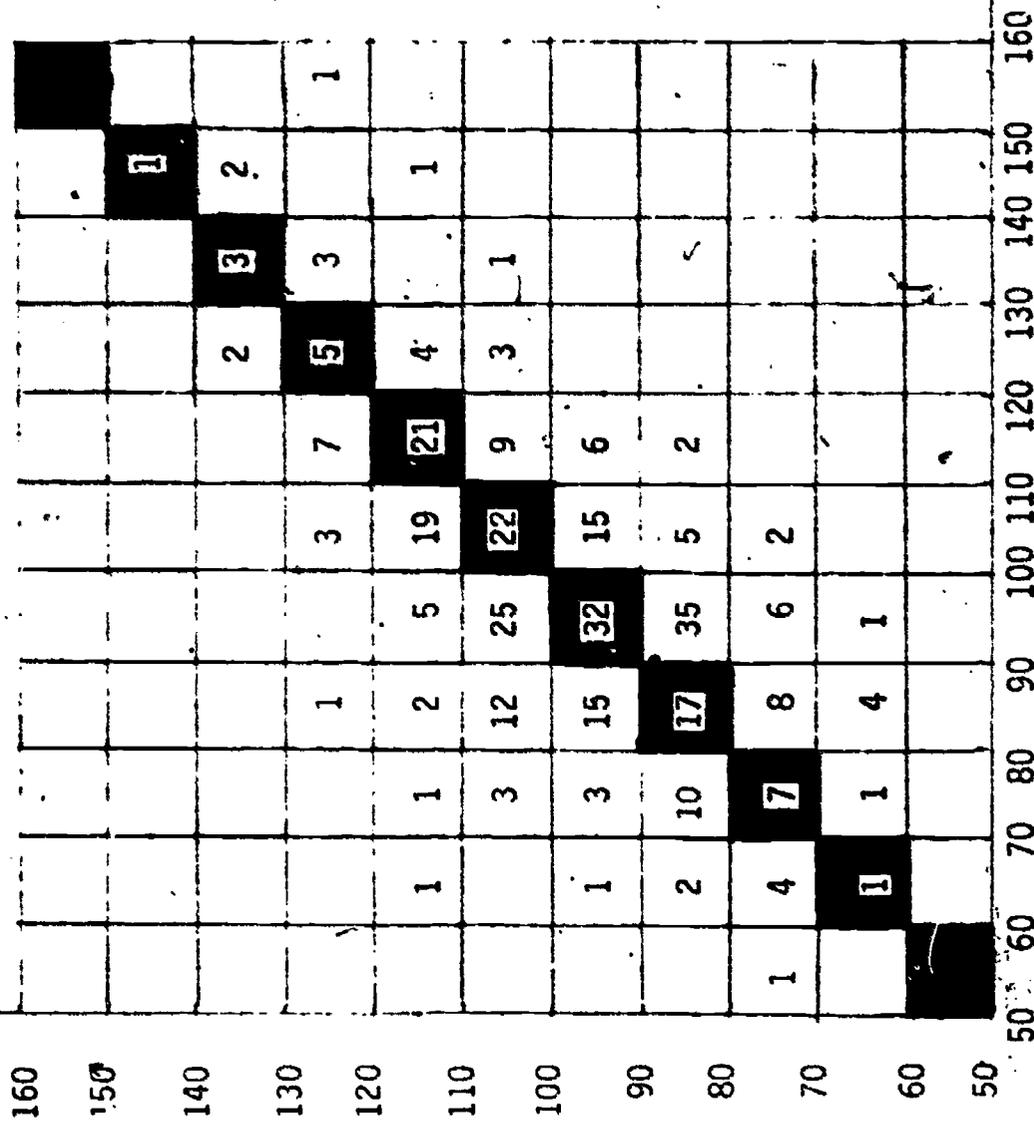
Students in introductory research and evaluation courses.

INSTRUCTIONAL MODULE ON
REGRESSION AND THE MATCHING FALLACY
IN QUASI-EXPERIMENTAL RESEARCH[1]

Perhaps the most subtle source of invalidity in behavioral research is the
elusive phenomenon of regression. Even seasoned researchers have frequently
failed to detect its presence; hence, it has spoiled many otherwise good
research efforts. Studies of atypical and special groups have probably been
the victimes of the regression phenomenon more often than those in any other
single area of inquiry. A simple statistical truism is that when subjects are
selected because they deviate from the mean on some variable, regression will
always occur.

Many studies on remediation and treatment of the handicapped and other
deviant groups follow this pattern: those in greatest "need" are selected, a
treatment is administered, and a reassessment then follows. For example,
suppose all children having IQ scores below 80 were given some special treatment
(e.g., glutamic acid) over a periou of a year and were then retested. Assume
that the time interval, etc., between testings was such that there was no prac-
tice or carryover effect. If the treatment had absolutely no effect, how would
the experimental group fare on the posttest? Figure 1 illustrates the regres-
sion effect using actual IQ score on 354 pupils tested in grade five and three
years later in grade eight. The regression line shows the average IQ score at
grade eight for any IQ score at grade five. For example, persons scoring 130 at
grade five obtained an average score at grade eight of approximately 120. In

---

Posttest (Grade 8) \ Pretest (Grade 5)	50	60	70	80	90	100	110	120	130	140	150	160
160												
150											1	1
140									2	3	2	
130					1		3	7	5	3		
120				1	2	5	19	21	4	1	1	1
110				3	12	25	22	9	3			
100			1	3	15	32	15	6	1			
90			3	10	17	35	5	2				
80			4	7	8	6	2					
70			11	1	4	1						
60	1		1		1							
50		1										

IQ Scores on Pretest (Grade 5)
$\bar{X} = 98$. $s_X = 15$

IQ Scores on Posttest (Grade 8)
$\bar{Y} = 98$ $s_Y = 15$

$r = .68$

Figure 1. The regression effect illustrated with data from 354 pupils tested at grades five and eight. (Data from Hopkins and Bibelheimer, 1971)

If there was no regression effect, for each column the number of scores above and below the shaded area would be about equal. Note that for scores below the mean, the average score is above the shaded area (i.e., regression toward the mean), for scores above the mean, the average scores are below the shaded area (i.e., regressing toward the mean).

other words, on the average, the grade five 130's regressed about 10 points to 120 at grade eight. Notice that there is a similar regression toward the mean for low-scorers at grade five. Notice also that the scores are just as variable at grade eight as they were at grade five -- the example was selected so that the regression effect would not be confounded with changes in means on the X and Y variable. There is a correlation of .6 between pretest and posttest IQ scores in this illustration. Figure 2 depicts a simplified illustrative situation. No treatment or practice effects are present for the treated group; the means and variances are identical in both distributions (as they are in most tests where standard scores are employed). Figure 2 illustrates that there is a definite and pronounced tendency for subjects to regress toward the posttest mean to the point where subjects tend to be, on the average., only six-tenths as far from the posttest mean as they were on the pretest; i.e., on the average, examinees tend to deviate only 60 percent as much from the posttest mean as they did from the pretest mean. Those examinees with pretest IQ scores of 80 would, on the average, be only 60 percent as far below the posttest mean -- they would be expected to have an average posttest score of 88, a substantial "gain" of 8 points. Those having IQ scores of 70 initially would appear to have gained 12 points, with a posttest mean of about 82.

The standard error of estimate ($s_{y.x} = s_y \sqrt{1-r^2}$) gives the standard deviation of posttest scores for persons having the same pretest score; in this example $s_{y.x} = 12$ IQ points. Using the $s_{y.x}$, we can accurately predict the proportion of those with a given pretest score who will fall above (or below) any other IQ score on the posttest (provided the common assumptions of linearity and homoscedasticity between the two variables are met). Those scoring 70 on the first test will have a mean of 82 on the second test, with a standard deviation of 12 IQ points. Using a normal curve table, it is readily apparent that about 84
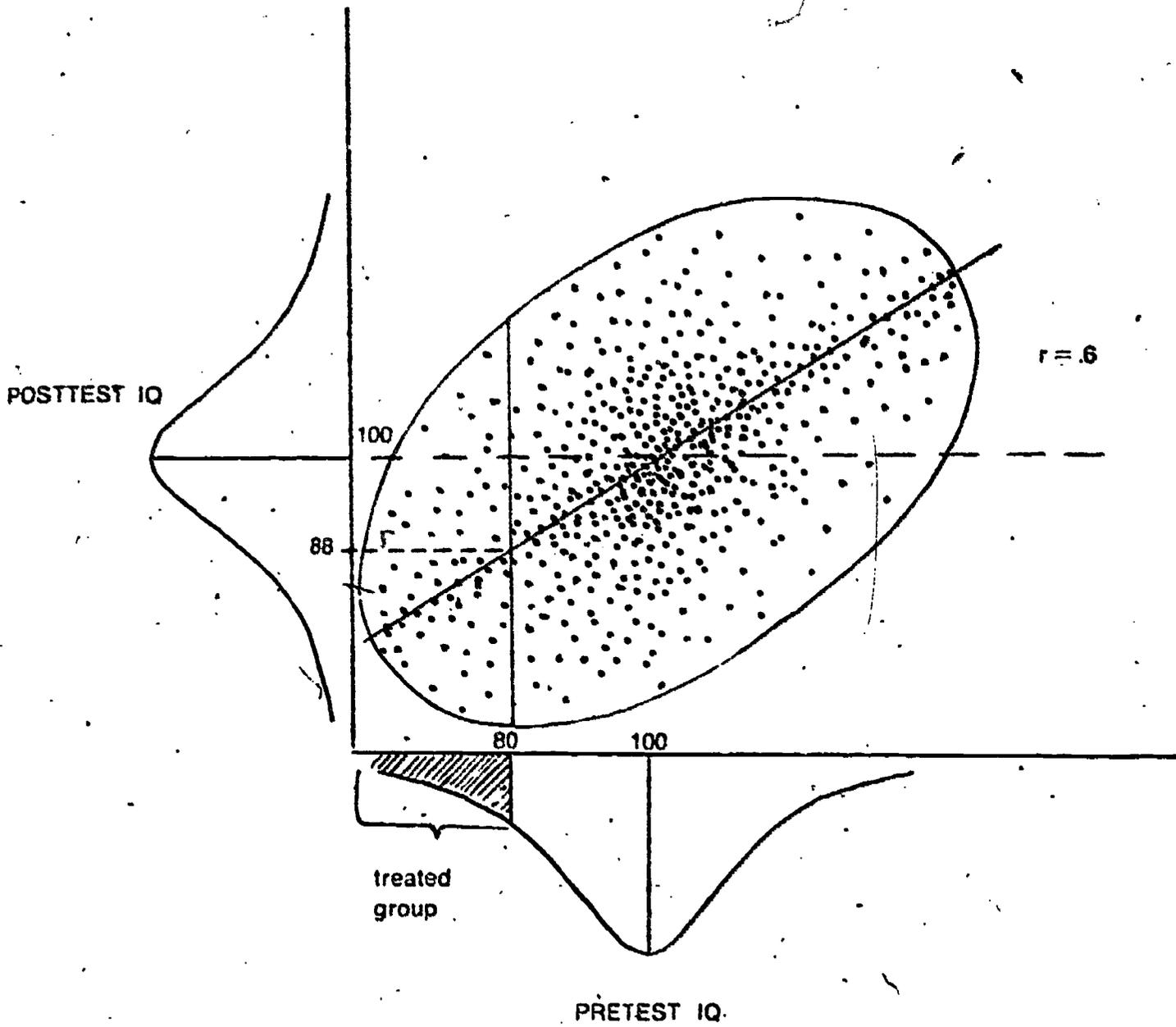
4

Figure 2. Graphic presentation of a hypothetical situation in which a deviate group is selected and administered an inefficacious treatment.

percent will regress and hence receive higher IQ scores on the posttest even without any practice effect. One-half will "gain" 12 or more IQ points; one-sixth will have IQs that "increased" by 24 or more points (i.e., obtain IQ scores of 96 or more). Further, about 10 percent of those with an initial IQ of 70 will obtain an IQ score of 100 or more on the second test, apart from any treatment or practice effect. Obviously, what may appear to an enthusiastic investigator to be striking improvements in a deviant population can result solely from the regression phenomenon. The following examples will serve to illustrate the problems:

Figure 3 is included to demonstrate that the regression effect is not simply a result of measurement error. Indeed, Galton first observed the phenomenon in stature of father and sons, as illustrated in Figure 3, and termed it the "law of filial regression." Note that tall fathers tend to have sons that are not as tall as they; short fathers tend to have sons that are not as short as they are -- that is, they regress toward the mean. Notice also that tall sons have fathers that are not as tall as they -- regression occurs going from X to Y or from Y to X.

INSTRUCTIONAL EXERCISES

Assuming no practice or testing effect in the situation depicted in Figure 2:

1. The expected or average score on the posttest for persons scoring 110 on the pretest is _____.
   106

2. The average score on the posttest for persons scoring 90 on the pretest is

   _____.
   94

3. If those scoring 90 on the pretest tend to score higher on the posttest, are they regressing? _____

   (Yes, statistical regression is movement toward the mean of group from which persons were selected.)

Height of Sons $\overline{Y} = 68$ $s_y = 2.4$

	62	63	64	65	66	67	68	69	70	71	72	73	$\overline{X}_i$
75											1			71.0
74											1		r	72.0
73								.1	1		1	1	1	70.6
72						1		1	2	2	2	.1	1	70.0
71					1	2	2	2	3	4	2	2	1	69.1
70				1	1	4	2	4	4	4	3	1	1	68.5
69			1	1	3	4	3	5	6	4	2	1		67.8
68			1	2	2	5	6	5	6	3	2	2		67.7
67		1	1	3	4	5	5	4	2	3	1			66.7
66			1	2	2	2	4	3	1					66.3
65			1	2	3	2	2	1	1					65.8
64		1	1	1	2	2	1							64.7
63					1	1								65.5
	62	63	64	65	66	67	68	69	70	71	72	73	
$\overline{Y}_i$	65.5	66.5	66.8	66.8	67.6	67.6	68.6	69.1	69.5	70.6	70.3	72.0	

Height of Fathers

$\overline{X} = 68$ $s_X = 2.5$

Figure 3 . Scatterplot showing the regression phenomenon ($r = .56$) in height of 192 fathers (X-variable) and sons (Y-variable). The average height of sons is given for fathers in each column ($\overline{Y}_i$); the average height of fathers is given for sons in each row ($\overline{X}_i$). (Note that in each instance (X to Y and Y to X) there is regression, although means and standard deviations are approximately equal for both X and Y.)

Data from McNemar (1962, p. 117).

4. Did both the 90 and 110 groups regress equally? _____
   Yes

5. The expected posttest score for persons scoring 120 on the pretest would be

   _____
   112

6. Did the "120s" regress more than the "110s?" _____

   (Yes, 8 points vs. 4 points.)

7. Was the ratio below the same for both the "110s" and "120s"?

   $$\frac{\text{(expected deviation from posttest mean)}}{\text{(actual deviation from pretest mean)}} \quad \text{or} \quad \frac{y'}{x} \quad \text{or} \quad \frac{Y' - \bar{Y}}{X - \bar{X}}$$

   (Yes, 6/10 = .6, and 12/20 = .6.)

8. The above ratio is the coefficient of correlation when the standard

   deviations for the pretest and posttests are equal, i.e., $\sigma_x = \sigma_y$. A more

   general expression is illustrated below:

   $$r = \frac{z_{y'}}{z_x}$$ where $z_{y'}$ is the expected standard z-score on the posttest, and

   z_x is the actual standard z-score on the pretest.

   Recall that z-scores are also called sigma scores because they express

   performance in standard deviation units. A z-score of +1.5 indicates the

   score was 1.5 standard deviations above the _____.
   mean

9. Suppose the x-variable in Figure 2 is unchanged, but the y-variable is a

   standardized grade-five reading test. Descriptive data for each are given

   below: $(r_{xy} = .6)$.

	IQ	Reading
Means	$\bar{X} = 100$	$\bar{Y} = 5.0$ (in grade equivalent)
S.D.	$\sigma_x = 16$	$\sigma_y = 1.5$

For persons with IQ scores of 132 on the IQ test (pretest), what is expected ("most" probable) reading score? Since $r = \frac{z'_y}{z_x}$, $z'_y = rz_x$. A score of 132 in z-score units is _____?

$(x = X - \bar{X} = 32; \; z_x = \frac{x}{\sigma_x} = 32/16 = 2.0.)$

10. Then, $z'_y = rz_x = (\;\;)(\;\;) = 1.2$.
$\qquad\qquad\qquad\quad (.6)\;\;(2.0)$

11. To convert z'_y to grade equivalent units, recall that the $z'_y = 1.2$ indicates a performance 1.2 standard deviations above the mean, hence

$y' = z'_y \; \sigma_y \;$ or $(\;\;)(\;\;) = 1.8$.
$\qquad\qquad\qquad (1.2)\;(1.5)$

12. Hence, the expected reading score for persons scoring 132 on the IQ test is 1.8 grade equivalents above the mean or $(\;\;) + (\;\;) = 6.8$.
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (5.0)\;\;\;(1.8)$

13. We have been illustrating the statistical basis for the regression effect. Conceptually we should understand that whenever subjects are selected because they are atypically low or high on some measure, on reassessment they will tend to _____ toward the mean.
$\qquad\qquad\qquad\qquad\quad$ (regress)

14. In the above example, was the percentile rank of the IQ score of 132 above the percentile rank of reading level?

(Yes, 98%ile vs. 88%ile.)

15. One study compared the IQ scores of retarded mothers with corresponding IQ scores of their offspring who had been given a cognitive enrichment program. Will the children have significantly higher scores even if the enrichment is without efficacy? _____
$\qquad\qquad\qquad\qquad\qquad$ (Yes)

Other actual examples follow.

'Webb's (1963) study reported that a group of Negro pupils in EMR classes had an average IQ of 68 on the WISC (on which they qualified for EMR classes) but obtained a mean IQ of 74 on the WAIS given two years later. The report concluded, "The most striking finding in this study is the significantly higher IQ's derived from the WAIS..." This reported increase easily falls within the range expected from regression alone.

Delacato (1959) reported large "gains" on a reading test for a group of pupils achieving at least 1.5 years below their "expectancy levels" who received Doman-Delacato therapy. Large apparent gains could have been predicted, since the regression phenomenon would have been operating strongly.

Another study (Scott & Brinkely, 1960) used the Minnesota Teacher Attitude Inventory and reposted that student teachers "...working with supervising teachers whose attitudes toward pupils were, in each instance, superior to their own, improved significantly, as a group, in their attitudes toward pupils during student teaching..." These results would be predicted from the regression phenomenon alone.

Some researchers have mistakenly assumed that if a pretest, different from that on which a group was selected, is administered before the treatment, that the regression taking place on the second pretest completely eradicates the problem of post-treatment test regression. They incorrectly assume that post-test means can be meaningfully compared with the second pretest mean to assess possible treatment effects. However, other things being equal, tests administered closely in time correlate more highly than those separated by a greater time interval. Hence, greater regression would be expected from the first pretest to posttest than from the first pretest to the second pretest. All of the regression thus is not eliminated by the use of a second pretest. Those scoring below the mean on the first pretest will score closer to the mean on the posttest

than they did on the second pretest in the absence of any treatment or practice effect. For example, going back to Figure 2, suppose the group selected on the pretest was administered another pretest prior to the treatment. If the two pretests correlated .8, then those with IQ scores of 80 on the first pretest would show an average IQ of about 84 on the second pretest, yet they would be expected to have a mean of 88 on the posttest without any treatment effect. Thus, in a pretest-posttest comparison, employment of a second pretest, following a pretest on which subjects are selected, does not eliminate all of the regression artifacts.

THE MATCHING FALLACY AND INTERNAL VALIDITY

The regression effect probably goes unnoticed most often in studies using the matched-pair design. Consider the example given in Figure 4 in which cerebral\palsied persons were "matched" on IQ with normal persons. Obviously, the intent was to have a CP vs. non-CP comparison on other variables, free from confounding resulting from intelligence or IQ differences. Typically, the subjects paired together have IQs which fall within a narrow range (e.g., five points). Unfortunately, this procedure almost always results in a real difference between the means of the groups, even on the variable on which they were "matched." In most pairs, the pair-member from the population with a higher mean will have a higher score than his matched-pair from the control population.

What if the investigator is aware of this problem and requires that the member of the cerebral palsied group have the higher score in one-half of the pairs of subjects? Regrettably, a real and important difference between the groups on the matching variable will continue to result. (It may not be "statistically significant," however, if the sample size is small, since power

would be low.) When the CP child has the higher score of the pair, the differ-
ence between the paired IQs would tend to be less than when the normal member
has the greater score. Figure 4 graphically illustrates this point: For
normals with IQ scores of, for example, 90, almost two-thirds of those CPs
having scores within five points of this value will be below 90.0. On the other
hand, however, for CPs with IQs of 90, most of those normals within the matching
range are above 90.

Now consider the situation in which the researcher is aware of the above
problems and requires identical scores on the matching variable, does he
eliminate the regression problem? Unfortunately not.

Suppose an investigator wanted to ascertain whether his creativity-inducing
treatment would be more effective with Negro pupils than with Anglo pupils.
Assume that he required his matched pairs to have identical pretest scores on
Form 1 of the ABC creativity test, which was the selection instrument. The
distribution of pretest scores for the total groups (from which the matched
pairs were to be selected) is shown on the horizontal axis in Figure 5.

For simplicity, assume the standard T-score means were 40 and 60 for the
Negro and Anglo groups, respectively. The investigator then found fifty matched-
pairs having equal scores on Form 1 of the ABC creativity test, who then became
the members of his experimental and control groups. What would happen if he
retested his sample with the parallel form of the ABC creativity test (a reli-
ability coefficient of .60 is typical of such tests and is assumed here)? The
results are given in Figure 5.

The illustration shows that on the retest, the Anglo pupils would, on the
average, be 8 T-score points (or .8 standard deviations) higher than their Negro
matched-pairs who had an identical score on the initial test. For example, of
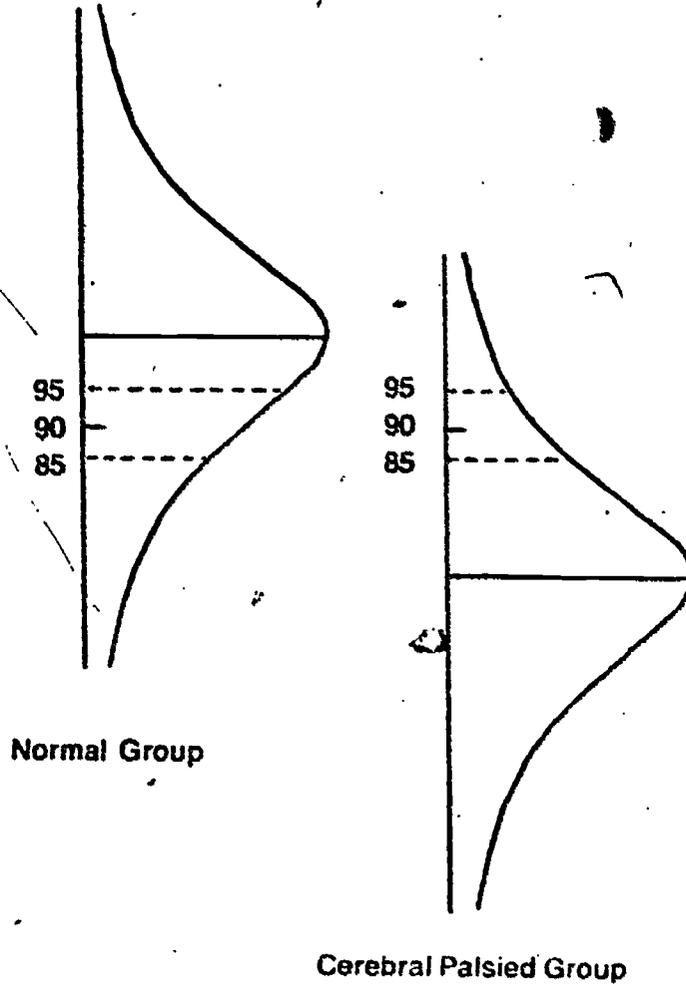the matched-pairs having a score of 40 on the first test, the Negro mean on the

95 - - - -
90 —
85 - - - -

95 - - - -
90 —
85 - - - -

Normal Group

Cerebral Palsied Group

Figure 4
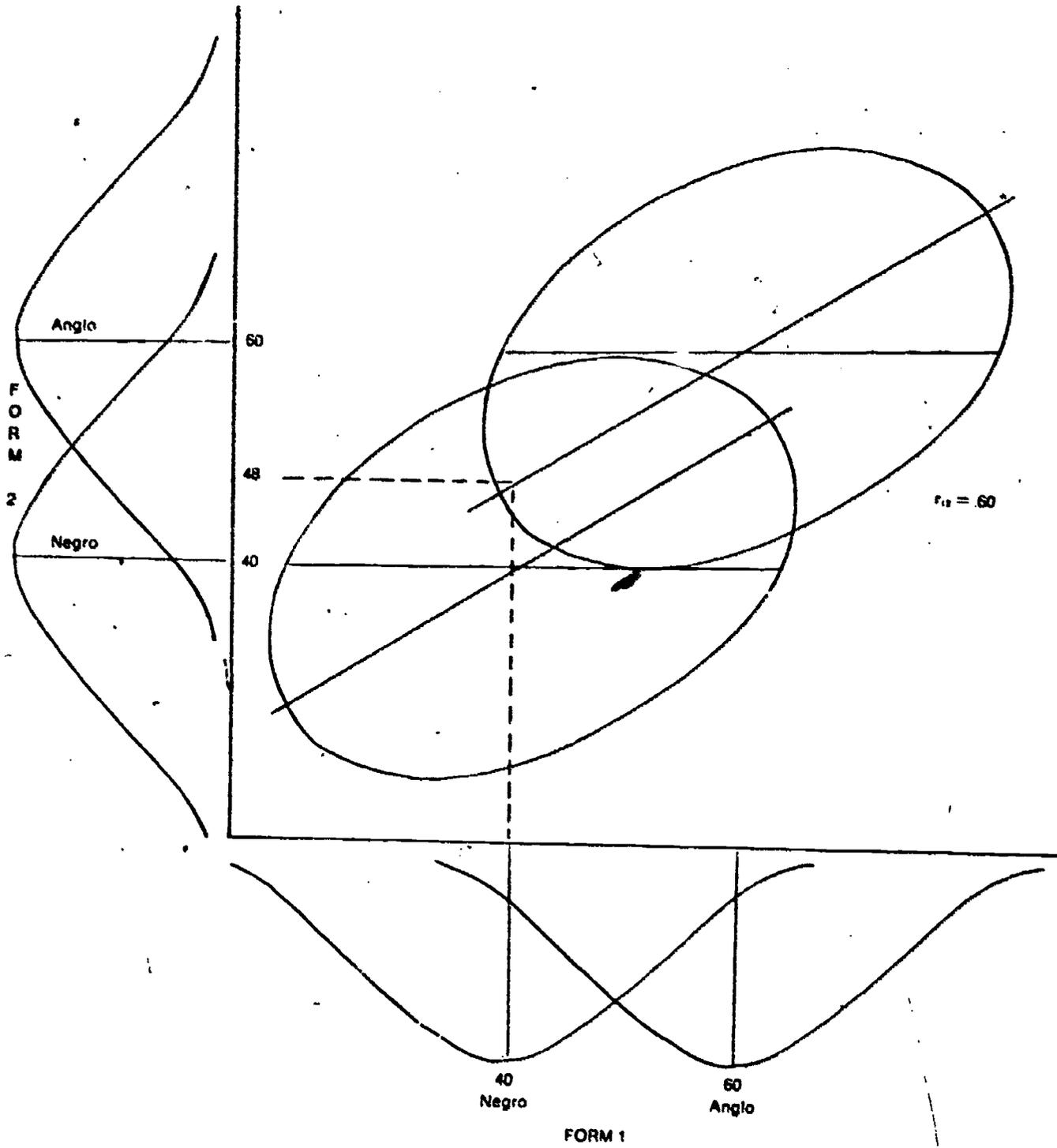Hypothetical IQ distributions of cerebral palsied and normal children

Figure 5
Illustration of a matching situation for Negro and Anglo pupils on a hypothetical creativity test

second test would be 40, whereas the Anglo mean would be about 48. In other words, the Anglo mean will be much greater (8 points) on the retest than the Negro mean, due simply to regression effects. The scores of each sample have "regressed" toward the mean of their respective total groups.

In many studies, of which the above example is typical, the investigator pretests, matches subjects, applies treatment, and then retests. He frequently concludes that the treatment was more effective for one group. This conclusion is based upon inadequate awareness of the regression that took place from test to test. We should note that in the example above we are observing only the regression phenomenon, not testing or maturation effects.

Instructional Exercises

16. Suppose high school male and female students were matched on height (within ½ inch) prior to being compared in some psychomotor skill. Has all of the effect of height been removed from the comparison? _____
(No)

17. Would the average height of the matched-pair males be higher than their matched-pair females? _____
(Yes)

18. Why?

(Since population means differ (see Figure 4), female pair-members would be more apt to be the shorter pair-member, i.e., of all males within ½ inch of a typical female's height, perhaps two-thirds would be taller and only one-third shorter.)

19. If the research design required that the female was to be the taller pair-member in 50% of the pairings, would the average height of the males and females be expected to be equal? _____

(No, the males would still have a higher mean.)

20. Return to Figure 4. Suppose you are to find matched-pairs (± 5 IQ points) of normal persons for a group of C.P.'s. If a C.P. had an IQ score of 90, what would the most probable or frequent qualifying IQ score (± 5 IQ points) that you would find among the normal group be? _____ (The height of the normal distribution curve indicates score frequency.)

    (95)

21. But, to turn the illustration around, suppose you are seeking from the C.P. group a matched-pair for a normal pupil having an IQ score of 90. The most probable qualifying pair member from the C.P. group would have an IQ score of _____.

    (85)

22. In other words, regardless of whether one first has scores from the C.P. group and then finds the matched-pair from the normal group or vice-versa, the most probable discrepancy is 5 points favoring the _____ group.

    (normal)

23. If identical observed scores are required, would the mean observed scores be equal? _____

    (Yes)

24. But would the mean true scores be equal? _____

    (No)

25. If the group matched-pairs with identical scores were retested using a parallel form, which would score higher? _____

    (normals)

26. Why?

    (The average of each group would tend to regress toward its population mean. Since most C.P.'s have IQ scores below 100, more than half of the normals among the matched pairs would have scored below the normals' mean. Upon retesting, the means of the normal group would have (increased or decreased)?

    (increased)

One recent study (Dobbs & Neville, 1967) matched 30 non-promoted pupils on race, sex, age, MA, reading, and SES and concluded, "The promoted were better after 2nd and 3rd years in both reading and arithmetic." These results could have been anticipated on the basis of the regression effect alone.

Would the use of gain scores avoid the difficulty? Unfortunately, the correlation of gain and initial scores presents some statistical difficulties (cf. McNemar, 1962) and is inefficient. If the pretest were used as a covariate in order to equate groups, would the problem be solved? No, the adjusted means would still differ without a treatment effect (eight points in the Anglo-Negro example) in spite of the fact that the original means of both groups remained unchanged. Lord (1967) has graphically illustrated this paradox.

Matching and External Validity

One can easily see from the example given in Figure 3 that the matched-pair approach also seriously restricts the external validity of the findings when the "matched" subjects are drawn from populations having different means. The majority of the members of the Negro matched-pair sample discussed above would have had scores higher than the Negro group mean, whereas the Anglo sample would have represented below-average subjects from the Anglo population.

Recommendations

Random assignment to treatment and non-treatment groups should be utilized whenever possible when working with non-organismic independent variables (e.g., variables to which subjects can be randomly assigned). However, if a researcher is comparing groups differing in organismic variables, e.g., factors such as sex, ethnic group and IQ, which do not lend themselves to random assignment, the dependent variable should be residual gain scores, i.e., the difference between predicted scores and obtained scores on the posttest. (This may be difficult to

establish since, in order to predict performance, data on a previous group is required.) Using this approach, in the present illustration, no differences would have been found between Anglo and Negro groups in residual gain scores. Additional technical discussions of this problem may be found in Harris (1963), Stanley (1967), and Thorndike (1942,1963).

1

REFERENCES

Delacato, C. The Treatment and Prevention of Reading Problems. Springfield, Ill: Charles C. Thomas, 1959.

Dobbs, Virginia & Neville, D. The Effect of Non-promotion on the Achievement of Groups Matched from Retained First Graders and Promoted Second Graders. J. educ. Res., 1967, 60, 472-475.

Harris, C.W. (Ed.) Problems in Measuring Change. Madison, Wisc.: Univ. Wisconsin Press, 1963.

Hopkins, K.D. & Bibelheimer, M.H. Five-year Stability of Intelligence Quotients from Language and Non-language Group Tests. Child Development, 1971, 42, 645-649.

Lord, F.M. A Paradox in the Interpretation of Group Comparisons. Psychol. Bull., 1967, 68, 304-305.

McNemar, Q. Psychological Statistics. (3rd ed.) New York: Wiley, 1962.

Scott, O. & Brinkley, S.G. Attitude Changes of Student Teachers and the Validity of the Minnesota Teacher Attitude Inventory. J. educ. Psychol., 1960, 51, 76-81.

Stanley, J.C. Problems in Equating Groups in Mental Retardation Research. J. spec. Educ., 1967, 1, 241-256.

Thorndike, R.L. Regression Fallacies in the Matched Groups Experiment. Psychometrika, 1942, 7, 85-102.

Thorndike, R.L. The Concepts of Over- and Under-achievement. New York: Teachers College, Columbia Univ., 1963.

Webb, A.P. A Longitudinal Comparison of the WISC and WAIS with Educable Mentally Retarded Negroes. J. clin. Psychol., 1963, 19, 101-102.