DOCUMENT RESUME

ED 096 355 95 TM 003 971

AUTHOR	Hopkins, Kenneth D.
TITLE	Instructional Module on the Analysis of Covariance.
INSTITUTION	Colorado Univ., Boulder. Lab. of Educational Research.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
PUB DATE	Sep 73
NOTE	32p.; For related documents, see TM 003 967-973
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	*Achievement Tests; *Analysis of Covariance; *Autoinstructional Aids; Educational Researchers; Graduate Students; *Guides; *Problem Sets; Research Design; Validity

ABSTRACT

The purpose of this training workbook is to provide the user with an understanding of Analysis of Covariance (ANCOVA) sufficient to allow him to identify situations in which it can increase the credibility and the statistical power of the analysis. The module provides a conceptual, nonmathematical overview of the purposes of ANCOVA. The assumptions underlying the use of ANCOVA and the consequences of their violation are summarized. An illustrative ANCOVA problem is employed to graphically illustrate how ANCOVA removes bias and increases statistical power. A self-instructional problem set is included as illustration and reinforcement for the learner. The module concludes with a mastery test. The workbook is designed for students in intermediate statistics and experimental design courses and for research and evaluation personnel, especially those being trained on the job. The book requires familiarity with simple regression and one-factor analysis of variance. (Author/SE)

BEST COPY AVAILABLE

INSTRUCTIONAL MODULE ON THE ANALYSIS OF COVARIANCE

Kenneth D. Hopkins
Laboratory of Educational Research
University of Colorado

September, 1973

BEST COPY AVAILABLE

NCERD Reporting Form — Developmental Products

1. Name of Product	**2. Laboratory or Center**	**3. Report Preparation**
Instructional Module on the Analysis of Covariance (ANCOVA)	(LER)	*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 have weak internal validity because of non-comparable groups (selection bias).

Many research and evaluation studies fail to discover real differences because the analysis employed is inefficient and lacks power.

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

The training materials include a rewrite of a classic ANCOVA expository article by W.S. Cochran, adapting the illustration from agriculture to education.

The second part of the module includes self-instructional problem sets, followed by a mastery test.

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

10.71.A (D)

BEST COPY AVAILABLE

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:

The 26-page module provides a conceptual, non-mathematical overview of the purposes of ANCOVA. The assumptions and consequences of their violation is summarized. An illustrative ANCOVA problem is employed to graphically illustrate how ANCOVA removes bias and increases power. A self-instructional problem set is designed to illustrate to and reinforce the learner. The module concludes with a mastery test.

What it is Intended to do:

Provide the user with an understanding of ANCOVA sufficient to allow him to identify situations in which it can increase the credibility and power of the analysis.

Requirements for Use:

Familiarity with simple regression and one-factor analysis of variance.

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

The product is intended to be used by applied researchers in education and by students in intermediate courses in statistics or experimental design.

11. Product Outcomes: *The changes in user behavior, attitudes, effectiveness, 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 [illegible]*

An anonymous rating form was given to a group of twenty-five users who responded to the instructional value of the module. 35% of the users responded "very good," 50% responded "good," and only 15% rated the module as "fair." In addition, only 15% indicated that there were other sources that accomplished the same purposes that are as good or better.

The 86% indication of "good" or "very good" instructional value by users suggests learning value and efficiency for the module. The median reported error rate was 7.5%.

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. The use of analyses that will yield less equivocal results.

2. The use of more power analyses of research and evaluation studies.

13. Product Elements: *List the elements which constitute the product.*	**14. Origin:** *Circle the most appropriate letter.*
One self-contained and self-instructional module.	(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 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.*

Research and evaluation personnel, especially those bei g trained on the job.

Students in intermediate statistics and experimental design courses.

Instructional Module on the Analysis of Covariance[a]

This paper discusses the nature and principal uses of the analysis of covariance (ANCOVA). As Fisher (1934) has expressed it, the analysis of covariance "combine the advantages and reconciles the requirements of the two very widely applicable procedures known as regression and analysis of variance."

In experimental and quasi-experimental studies covariance can perform two distinct functions. One is to remove bias, that is, statistically equate groups on some confounding variables. In quasi-experimental studies coping with bias is typically its primary function. However, even if there are no real differences between the two groups on the covariable, hence no danger of bias, covariance may still be valuable for increasing the power of the analysis.

The Use of ANCOVA

To remove the effects of confounding variables in quasi-experimental studies.

In research endeavors in which randomized experiments are not feasible, two or more groups differing in some characteristic such as age, can be studied to discover whether there is a significant difference among groups on the dependent

[a]The ANCOVA overview is adapted from portions of W. S. Cochran's article in Biometrics (13:261-278), 1957.

variable when groups are statistically equated on the characteristic on which they differ (such as age, IQ, or pretest score). Examples where true experiments are not practicable or possible are studies constrasting cross-cultural studies, social class studies, urban vs. rural school districts, etc. In quasi-experimental studies it is widely realized that an observed association, even if statistically significant, may be due wholly or partly to other disturbing variables X_1, X_2 ... in which the groups differ, i.e., X_1 and X_2 are threats to the internal validity of the study. Where feasible, a common device is to match the groups for the disturbing variables thought to be most important. This matching often results in serious problems (cf. Hopkins, 1969). In the same way, the analysis of the X-variables can be treated as covariates and ANCOVA be employed to extricate the influence of X-variables, at least partially.

In a comparison of the heights of children from two different types of schools, Greenberg (1953) found that the two groups differed slightly, though not significantly, in mean age. A covariance adjustment for age resulted in a more sensitive comparison of the heights. Another study statistically equated mobile and non-mobile students on IQ when examining achievement consequences of mobility. School districts have been compared in pupil achievement after covarying on numerous socio-economic variables.

Unfortunately, quasi-experimental studies are subject to difficulties of interpretation from which true experiments are free. Although covariance has been skillfully applied, we can never be sure that bias may not be present from some disturbing variable that was overlooked. Indeed, unless the covariate is perfectly reliable, ANCOVA does not remove all of the bias due to X itself. In true experiments, the effects of all variables measured and unmeasured, real and illusory, are distributed among the groups by the randomization in a way that is taken into account in the standard tests of significance.

There is no such safeguard in the absence of randomization.

Secondly, when the X-variables show real differences among groups -- the case in which adjustment is needed most -- covariance adjustments involve a greater-or-less degree of extrapolation. To illustrate by an extreme case, suppose that we were adjusting for differences in parents' income in a comparison of private and public school children, and that the private school incomes ranged from \$10,000-\$12,000, while the public school incomes ranged from \$4,000-\$6,000. The covariance would adjust results so that they allegedly applied to a mean income of \$8,000 in each group, although neither group has any observations in which incomes are at or even near this level.

Two consequences of this extrapolation should be noted. Unless the statistical assumption of linear regression holds in the region in which observations are lacking, covariance will not remove all the bias, and in practice may remove only a small part of it. Secondly, even if the regression is valid in the "no man's land," the standard errors of the adjusted means become large, because the standard error formula in a covariance analysis takes account of the fact that extrapolation is being employed (although it does not allow for errors in the form of the regression equation). Consequently, the adjusted differences may become insignificant statistically merely because the adjusted comparisons are of low precision.

When groups differ widely on some confounding variable X, these difficulties imply that the interpretation of an adjusted analysis is speculative rather than definitive. While there is no sure way out of the difficulty, two precautions are worth observing.

1. Consider what internal evidence exists to indicate whether the regression is valid in the region of extrapolation. Sometimes the fitting of a more complex regression formula serves as a partial check.

2. Examine the standard erors of the adjusted group means, particularly when differences become non-significant after adjustment. Confidence limits for the difference in adjusted means will reveal how precise or imprecise the adjusted comparison is.

<u>The Use of ANCOVA</u>

<u>To increase power</u>.

The use of ANCOVA to increase power in true experiments is frequently overlooked. The covariate X is a measurement, taken or available on each experimental unit before the treatments are applied, which correlates with the dependent variable Y. This first illustration of the covariance method in the literature was of this type (Fisher, 1932). The variate X was the yield of tea per plot in a period preceding the start of the experiment, while Y was the tea yield at the end of a period of application of treatments. Adjustment of the responses Y for their regression on X removes the effects of variations in initial yields from the experimental errors, insofar as these effects are measured by the linear regression. In this example these effects might be due to either inherent differences in the tea bushes or to soil fertility differences that were permanent enough to persist during the course of the experiment.

With a linear regression equation, the gain in predision from the covariance adjustment depends primarily on the size of the correlation coefficient ρ between Y and X on experimental units that receive the same treatment. If σ_Y^2 is the error variance when no covariance is employed, ANCOVA reduces this error variance to a value which is about

$$\sigma_Y^2(1 - \rho^2)(1 + \frac{1}{f_e - 2})$$

where f_e is the degrees of freedom associated with the error term. The factor involving f_e is needed to take account of errors in the estimated regression coefficient. If, ρ, the correlation of covariate and the dependent variable, is less than 0.3 in absolute value, the reduction in variance is inconsequential (less than 9%), but as ρ increases sizeable increases in precision are obtained. In Fisher's example ρ was 0.928, reflecting a high degree of stability in relative yield of a plot from one period to another. The adjustment reduced the error variance roughly to a fraction $(1 - (0.928)^2)$, or about one-sixth, of its original value. Some of the most spectacular gains in precision from covariance have occurred in situations like this, in which the covariate represents an initial calibration of the responsiveness of the experimental units. In educational studies it is usually relatively easy to find pretreatment measures that correlate .6 or higher with posttest measures thereby reducing the error term by 36% or more -- approximately the same gain in power that would result from doubling the sample size.

In the use of ANCOVA to increase power, its function is the same as that of stratification and blocking. It removes the effects of an environmental source of variation that would otherwise inflate the experimental error and hence the error mean square. When the relation between Y and X is linear, covariance and blocking can be about equally effective. If, instead of using covariance, we can group the subjects into block such that the X values are equal within a block the error variance is reduced to $\sigma^2_{Y}(1 - \rho^2)$.

In a covariance analysis, the covariate X may be measured on a completely different scale from that of the dependent variable Y. Bartlett (1937) used a visual estimate of the degree of saltiness of the soil to adjust cotton yields. Federer and Scholottefeldt (1954) used the serial order (1, 2, ...7) of the plot within a replication as a basis for a quadratic regression adjustment of tobacco data, thereby removing the effects of an unexpected gradient in fertility within the replications. Similarly, the reading performances of children under different methods of instruction may be adjusted for variations in their initial IQ's. Note also that X need not be a direct causal agent of Y -- it may, for instance, merely reflect some characteristic of the environment that also influences Y.

When ANCOVA is used in this way, it is important to verify that the treatments have had no effect on X. This is obviously true when the X's were measured before treatments have been applied, as when plant number shortly before harvest is used to adjust crop yields for uneven growth, or as happened in the index of saltiness used by Bartlett. When the treatments do affect the X-values to some extent, the covariance adjustments take on a different meaning. They no longer merely remove a component of experimental error -- in addition, they distort the nature of the treatment effect that is being measured. If the higher performance by a superior reading treatment also improves IQ scores, a covariance adjustment (which attempts to measure what the means would have been if IQ means were equal for all treatments), may remove much of the real treatment effect.

ASSUMPTIONS REQUIRED FOR THE ANALYSIS OF COVARIANCE

The assumptions required for valid use of the analysis of covariance are the natural extension of those for an analysis of variance, namely,

(i) Treatment, block and regression effects must be additive as postulated by the model,

(ii) the residuals, e_{ij}, (differences between observed and predicted scores within each treatment group) must be normally and independently distributed with zero means and the same variance.

Much of the related work regarding the effects of violating statistical assumptions on the analysis of variance extends logically to ANCOVA --- for instance the practical unimportance of the additivity assumption (see Glass, Peckham, and Sanders, 1972, p. 241). Table 1 summarizes an abundance of research literature on the empirical consequence of violating assumptions in ANOVA.

Certain qualifications of the conclusions in Table 1 are regarded in the extension to ANCOVA. For example, non-normality in the dependent variable is inconsequential in ANCOVA only if the covariate is normally distributed (which in itself is not necessarily assumed in ANCOVA).

ANCOVA makes three assumptions that involve the regression term in covariance: (1) the regression lines for each group are assumed to be parallel, i.e., $\beta_1 = \beta_2 = \ldots = \beta_J$. If this is violated, the covariance adjustment may still improve the precision, but (i) the meanings of the adjusted treatment effects become cloudy, and (ii) if covariance is applied in a routine way, the investigator fails to discover the differential nature of the treatment effects -- a point that might be important for practical applications.

Peckham (see Glass, Peckham, and Sanders, 1972) found that violation of the parallel regression slopes to be inconsequential in a one-factor fixed-effects ANCOVA for a wide variety of conditions. The effects in more complex factorial design with mixed and random models appears not to have been studied.

(2) The covariance procedure assumes that the correct form of regression equation has been fitted. Perhaps the most common error to be anticipated is that linear regressions will be used when the true regression is curvilinear. In a randomized experiment, the randomization insures that the usual interpretations of standard errors and tests of significance are not seriously vitiated, although fitting the correct form of regression would presumably give a larger

Table 1

Summary of Consequences of Violation of Assumptions of the Fixed-effects ANOVA

ype of Violation	Equal n's: Effect on α	Equal n's: Effect on Power	Unequal n's: Effect on α	Unequal n's: Effect on Power
lon independence of errors	Non-independence of errors seriously affects both the level of significance and power of the *F*-test regardless whether *n*'s are equal or unequal.			
on-normality: Skewness	Skewed populations have very little effect on either the level of significance or the power of the fixed-effects model *F*-test; distortions of nominal significance levels of power values are rarely greater than a few hundredths. (However, skewed populations can seriously affect the level of significance and power of *directional*—or "one-tailed"—tests.)			
Kurtosis	Actual α is less than nominal α when populations are leptokurtic (i.e., $\beta_2>3$). Actual α exceeds nominal α for platykurtic populations. (Effects are slight.)	Actual power is less than nominal power when populations are platykurtic. Actual power exceeds nominal power when populations are leptokurtic. Effects can be substantial for small *n*.	Actual α is less than nominal α when populations are leptokurtic (i.e., $\beta_2>3$). Actual α exceeds nominal α for platykurtic populations. (Effects are slight.)	Actual power is less than nominal power when populations are platykurtic. Actual power exceeds nominal power when populations are leptokurtic. Effects can be substantial for small *n*'s.
eterogeneous Variances	Very slight effect on α, which is seldom distorted by more than a few hundredths. Actual α seems always to be slightly increased over the nominal α.	(No theoretical power value exists when variances are heterogeneous.)	α may be seriously affected. Actual α exceeds nominal α when smaller samples are drawn from more variable populations; actual α is less than nominal α when smaller samples are drawn from less variable populations.	(No theoretical power value exists when variances are heterogeneous.)[a]
ombined non-normality and heterogeneous ariances	Non-normality and heterogeneous variances appear to combine additively ("non-interactively") to affect either level of significance or power. (For example, the depressing effect on α of leptokurtosis could be expected to be counteracted by the elevating effect on α of having drawn smaller samples from the more variable, leptokurtic populations.)			

[a] From Glass, Peckham, and Sanders (1972).

increase in precision. The danger of misleading results is greater when there are real differences from treatment to treatment on the covariate. Fortunately, most cognitive and psychomotor variables are linearly related, and unless measurement procedures are faulty (e.g., a test that lacks ceiling), the linear regression model works well in most applications (see Li, 1964, for treatment of curvilinear ANCOVA). Frequently, curvilinear relationships can be made linear by mathematical transformations of either the dependent variable Y, or the covariate X, or both.

(3) An assumption of ANCOVA that is not widely recognized is that the covariate is fixed and measured without error. Lord (1960) has shown how large errors in the covariate can produce misleading results. The effects of the less-than-perfectly-reliable covariate are usually predictable so the nature of the bias in the adjustment can be considered in any interpretation. It should be emphasized, however, that, to the extent the covariate is unreliable, the statistically equating of the groups is incomplete.

Illustrative ANCOVA Problem

Suppose there are three intact groups (A, B, C), each was given a treatment. They were pretested (X) before the treatment and posttested following the treatment. The data are depicted graphically on the X and Y axes in Figure 1A.

Treatment					
A		B		C	
X	Y	X	Y	X	Y
2	5	14	7	20	20
4	8	16	8	18	22
5	7	15	10	23	26
8	9	19	13	25	28
6	11	11	12	24	24

Summary Data

	A		B		C		Totals (X)	Totals (Y)
ΣX	25		75		110		210	
ΣY		40		50		120		210
ΣX^2	145		1159		2454		3758	
ΣY^2		340		526		2920		3786
ΣXY	215		755		2670		3640	
Means	5	8	15	10	22	24	$\bar{X}_. = 14$	$\bar{Y}_. = 14$
E_{xx_j}		20		34		34	Within Treatments (E): $88 = E_{xx}$	
E_{xy_j}		15		5		30	$50 = E_{xy}$	
E_{yy_j}		20		26		40	$86 = E_{yy}$	

Total Data (S)	Between Treatments (T)
$S_{xx} = 818$	$T_{xx} = 730$
$S_{xy} = 700$	$T_{xy} = 650$
$S_{yy} = 846$	$T_{yy} = 760$

Let's ignore the pretest differences for the moment and perform a simple ANOVA on the Posttest (Y).

SV	SS	df	MS	F	p
Treatments	760	2	380	53.1	<.01
error	86	12	7.16		
Total	846	14			

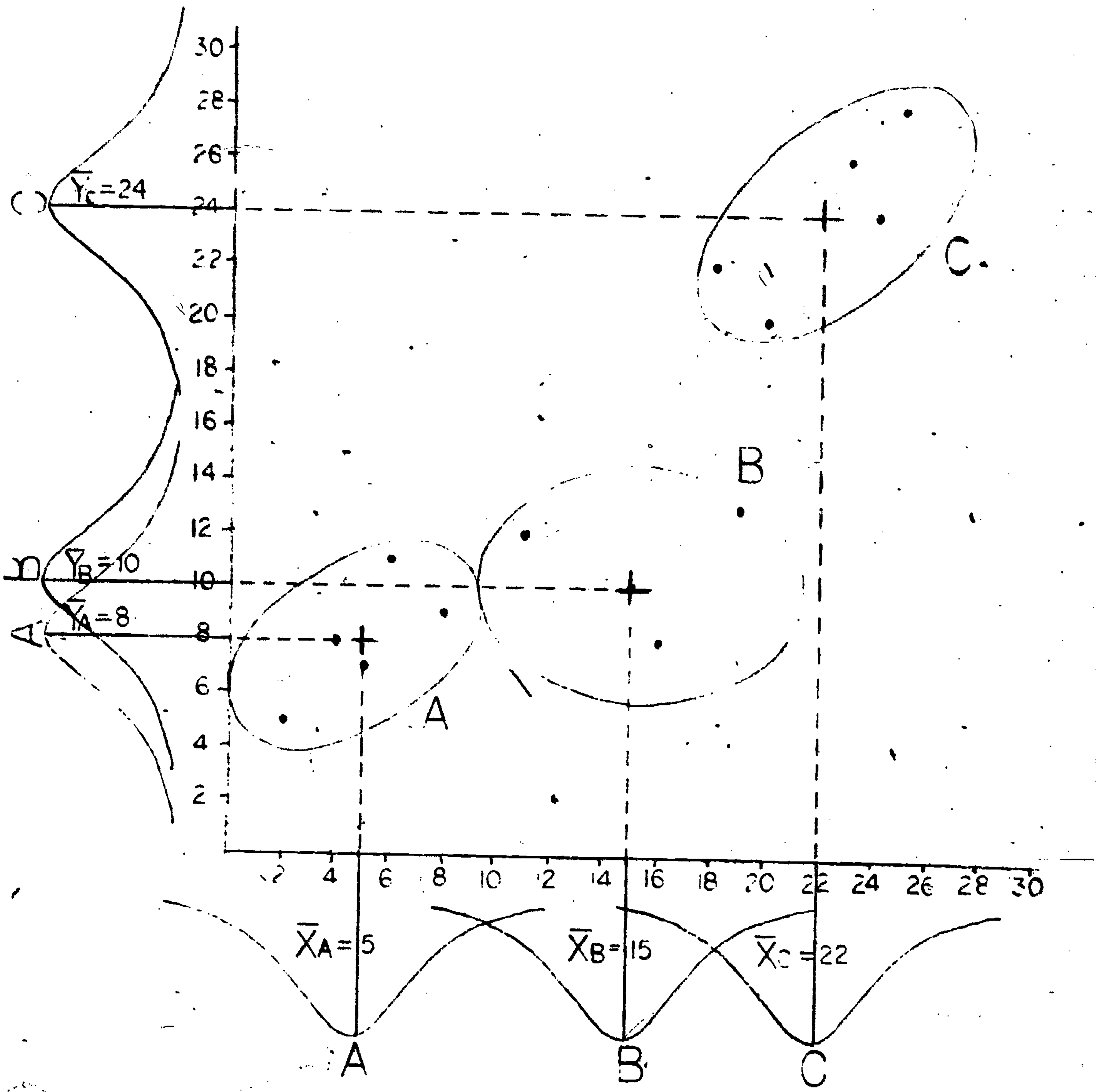


Figure 1A. Relationship between covariates and dependent variables for sample problem.

Obviously, this highly significant difference in posttest means is not very meaningful in light of the pretest differences. To confirm our suspicion that there were non-random, systematic differences between groups prior to the treatments, we run an ANOVA on pretest scores (X) and find that there were highly significant differences among groups prior to the treatments.

SV	SS	df	MS	F	p
Treatments	730	2	365	49.8	<.01
error	88	12	7.33		
Total	818	14			

Now, the crucial question is: when we statistically equate groups on the pretest, would there continue to be significant differences in posttest means. ANCOVA allows us to adjust the total sum of squares on the posttest (S_{yy}) to (1) remove predictable portion due to differences in pretest means (the "correcting" for bias function of ANCOVA) and (2) take advantage of predictability of posttest score from pretest score to reduce our error term (the power function of ANCOVA).

To adjust total sum of squares, S_{yy}:

$$S'_{yy} = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = 846 - \frac{(700)^2}{818} = 247$$

To adjust sum of squares error, E_{yy}:

$$E'_{yy} = E_{yy} - \frac{(E_{xy})^2}{E_{xx}} = 86 - \frac{(50)^2}{88} = 57.6$$

To adjust treatment sum of squares, T_{yy}:

$$T'_{yy} = S'_{yy} - E'_{yy} = 247 - 57.6 = 189.4$$

The summary ANCOVA table is shown below:

SV	SS'	df	MS'	F	p
Treatments	189.4	2	94.7	18.07	<.01
error	57.6	11	5.24		
Total	247.0	13			

(Note that one df is lost from error for each covariate)

We therefore conclude then that there are differences among the adjusted posttest means that are not explicable solely in terms of initial pretest differences.

For purposes of interpretation, we need to adjust the postttest means:

$$\bar{Y}'_j = \bar{Y}_j - b_w(\bar{X}_j - \bar{X})$$

$\bar{Y}'_j$ is the adjusted mean of the jth group. Except for b_w, all the information needed to adjust the means is given in the summary data. The regression coefficient, b_w, is the pooled estimate of β_w, the "average" slope within the treatment groups.

$$b_w = \frac{E_{xy}}{E_{xx}} = \frac{50}{88} = .57$$

The adjusted means of the treatment groups are then:

$$\bar{Y}'_A = 8 - .57(5-14) = 8 - (-5.1) = 13.1$$

$$\bar{Y}'_B = 10 - .57(15-14) = 10 - (.57) = 9.43$$

$$\bar{Y}'_C = 24 - .57(22-14) = 24 - (4.6) = 19.4$$

Figure 1B shows a regression line with slope b_w fitted to each of the three groups. The extension of this line to the point at which it intersects with the grand mean of the covariate, $\bar{X}.$, is the adjusted mean for the group.

Now is the assumption $\beta_A = \beta_B = \beta_C$, which legitimizes pooling, tenable?

To test H_0: $\beta_A = \beta_B = \beta_C$, we need to compare the sum of squares from the pooled regression line fitted for each group (E'_{yy}) with the sum of squares allowing each group to "find" its own best fitting individual regression line. Figure 1C gives the best fitting (least squares) regression line defined separately for each group together with the pooled regression line with slope b_w. Of course the regression line b_A will fit group A better than any other regression line including the one with slope b_w. Likewise b_B and b_C give least error for groups B and C. The real statistical concern is whether or not b_A, b_B, and b_C differ significantly, that is, is H_0: $\beta_A = \beta_B = \beta_C$ tenable? If H_0 is tenable then the use of the pooled regression coefficient b_w is legitimized.

We already have obtained the error sum of squares using the pooled regression coefficient b_w, i.e., $E'_{yy} = 57.6$. The error sum of squares for group A using b_A is:

$$E'_{yy_A} = E_{yy_A} - \frac{(E_{xy_A})^2}{E_{xx_A}} = 20 - \frac{(15)^2}{20} = 8.8$$

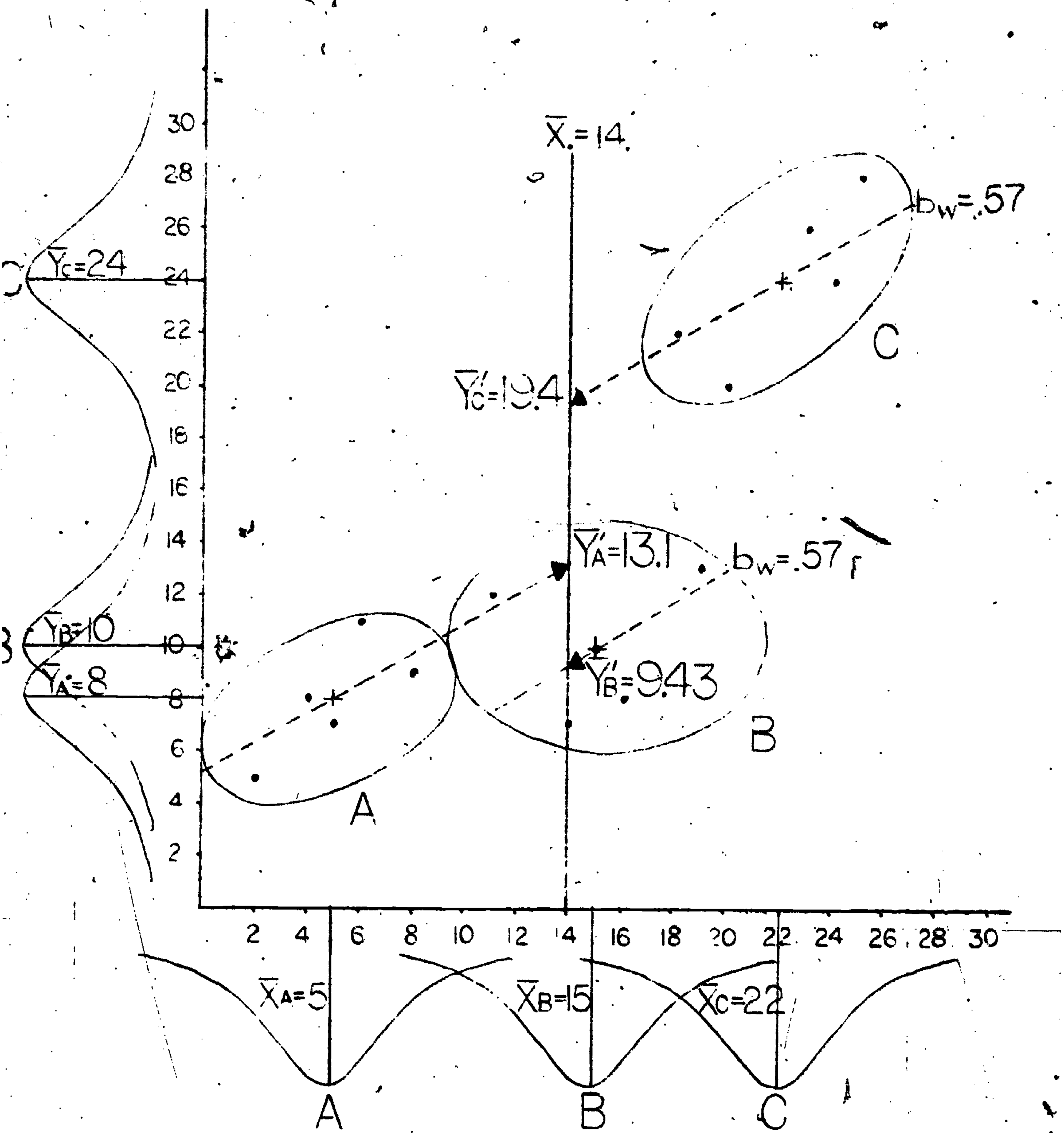


Figure 1B. An illustration of the process of adjusting means for pretreatment differences.

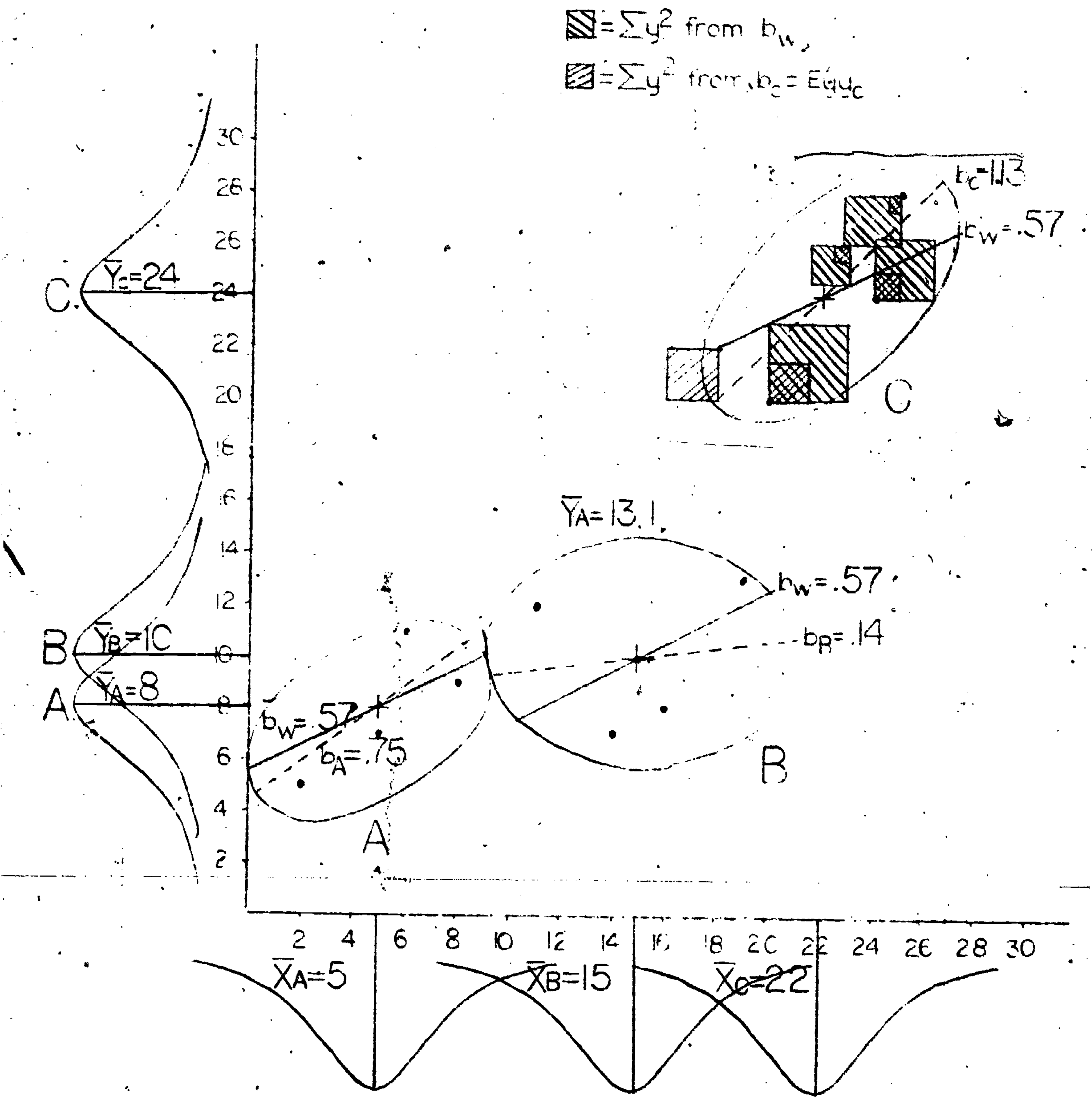


Figure 1C. The relationship between regression lines defined by separate groups with the regression line employing pooled data.

Similarly for groups B and C using b_B and b_C respectively:

$$E'_{yy_B} = 25.3, \quad E'_{yy_C} = 13.5$$

For convenience define $S_1 = \Sigma E'_{yy_j}$ = error sum of squares when each group defines its own best fitting regression line.

$$S_1 = 8.8 + 25.3 + 13.5 = 47.6$$

The reduction in sums of squares when best fitting individual regression lines are used (i.e., b_A, b_B, and b_C) in lieu of regression lines with the regression coefficient based on the pooled information (b_w), is defined as S_2:

$$S_2 = E'_{yy} - S_1 = 57.6 - 47.6 = 10.0$$

Obviously, if $b_A = b_B = b_C$, S_2 would be zero.

To test the significance of the non-parallelism in the individual regression lines:

$$F = \frac{S_2/(J - 1)}{S_1/J(n-2)} = \frac{10.0/2}{47.6/(3(3))} = \frac{5}{5.29} = .945$$ the F-ratio is below 1.0 -- obviously not significant.

In setting up confidence intervals about adjusted means and/or making multiple comparisons, MS'_e is not used, but MS''_e which is larger than MS'_e to the extent that the groups differed on the covariate, i.e., if $T_{xy} = 0$, $MS''_e = MS'_e$.

$$s_{\bar{Y}'} = \sqrt{\frac{MS''_e}{n}}; \quad MS''_e = MS'_e \left[1 + \frac{T_{xx}/(J-1)}{E_{xx}}\right]$$

ANCOVA Computational Problem Set

Fifteen subjects were administered a non-reactive pretest (X) and were randomly assigned to one of three treatments. The pretest and posttest data appear below (Winer notation; problem taken from Edwards).

Treatment Group

1		2		3	
X	Y	X	Y	X	Y
1	5	2	1	1	10
6	12	3	2	4	13
3	9	6	7	5	16
4	8	4	3	3	12
5	11	7	8	6	17

Summary Data	X	Y	X	Y	X	Y	Totals (X)	Totals (Y)
$\Sigma(\ )$	19	45	22	21	19	68	60	134
$\Sigma(\ )^2$	87	435	114	127	87	958	288	1520
ΣXY	191		118		270		579	
Means	3.8	9.0	4.4	4.2	3.8	13.6	4.0 ($\bar{X}.$)	8.93 ($\bar{Y}.$)

Within Groups Data (E)				
$\Sigma x_j^2 = E_{xx_j}$	14.8	17.2	14.8	$46.8 = E_{xx}$
$\Sigma xy_j = E_{xy_j}$	20.0	25.6	21.6	$67.2 = E_{xy}$
$\Sigma y_j^2 = E_{yy_j}$	30.0	38.8	33.2	$102.0 = E_{yy}$

Total Data (S)	Between Treatments (T)
$S_{xx} = 48.0$	$T_{xx} = 1.2$
$S_{xy} = 53.0$	$T_{xy} = -14.2$
$S_{yy} = 322.9$	$T_{yy} = 220.9$

1. Plot Y values against X values for each group. (Use different colors or marks for each group for visual separation.

   Following each exercise is a dotted line, below which provides the answers to the questions posed in the exercise. Attempt each question before consulting the answer.

2. Perform an analysis of variance of the posttest scores (Y) so that we may later compare the results with those from ANCOVA.

SV	SS	df	MS	F
Treatments		2		12.99
error	102.0		8.5	

$({}_{.99}F_{2,12} = 6.93)$

220.9/2 = 110.45; 102.0/12 = 8.5

3. Now perform ANCOVA, covarying on X

Adjusted total sum of squares, $S'_{yy} = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = (\quad) - \frac{(\quad)^2}{(\quad)} = 264.4$

$322.9 - \frac{(53)^2}{48}$

4. $E'_{yy} = (\quad) - \frac{(\quad)^2}{(\quad)} = 102 - \frac{(67.2)^2}{46.8} = 5.5$ = Adjusted error sums of squares

   ----

   $E_{yy} - (E_{xy})^2/E_{xx}$

5. $T'_{yy} = (\quad) - (\quad) = 258.9$ (Not $T_{yy} - (T_{xy})^2/T_{xx}$; this is affected by error in estimating β from b_j's.)

   $S'_{yy} - E'_{yy}$

6. Therefore:

SV	SS	df	MS	F
Treatments (T'_{yy})			129.45	258.4
Error (E'_{yy})			.50	

${}_{.99}F_{2,11} = 7.21$

258.9/2; 5.5/11

7. _____ degree(s) of freedom is (are) lost for each covariate employed (one in this example), which accounts for the slight ________ (increase or decrease) in the critical F-ratios.

----

ncrease

8. Why didn't T'_{yy} and T_{yy} differ greatly as they did in the earlier illustrative problem?

----

because the group means on the covariate differed minimally, hence the unadjusted means did not differ greatly from the adjusted means.

9a. Will T'_{yy} be larger than T_{yy} as a general rule in true experiments, i.e., when random assignment of subjects to treatments has been employed?

----

no, no consistent trend

9b. When will $T'_{yy} = T_{yy}$?

----

when $r_{xy} = 0$ (within cells), or when $\overline{X}_1 = \overline{X}_2 = \ldots = \overline{X}_j$

10. When will $E'_{yy} = E_{yy}$?

----

only when r_{xy} (within cells) = 0, hence $b_w = 0.0$

11. The relative advantage of ANCOVA over ANOVA can be seen best by comparing which one of these?

a. E'_{yy} with E_{yy}
b. T'_{yy} with T_{yy}
c. S'_{yy} with S_{yy}
d. computed F-values

----

a.

12. The gain in the power of ANCOVA over ANOVA is shown by the ratio of MS'_e to __________ or, in this example, .50/8.5.

MS_e

13. The gain in precision is a direct function of the correlation between the ____________ and the dependent variable (within cells, it is not r_{xy} for all observations combined).

----

covariate

14. $MS_e' \cong MS_e(1 - r^2)$, therefore in this problem: ($\cong$ means "approximately equal to")

$$r^2 \doteq 1 - \frac{(\quad)}{(\quad)} = 1 - \frac{(.5)}{(8.5)} = 1 - .059 = .941.$$

MS_e'/MS_e

15. More precisely, $r_j^2 = \frac{E_{xy_j}^2}{E_{xx_j} E_{yy_j}}$ for each group, or pooling our within groups information:

$$r_w^2 = \frac{E_{xy}^2}{E_{xx}E_{yy}} = \frac{(\quad)^2}{(\quad)(\quad)} = .946$$

$\frac{(67.2)^2}{46.8(102)}$ (This uncommonly high r is the reason MS_e' and MS_e differ so drastically.)

16. In order to adjust the $\overline{Y}_j$ values to $\overline{Y}_j'$ values we must find the pooled within-cell regression coefficient, b_w.

$$\overline{Y}_j' = \overline{Y}_j - b_w(\overline{X}_j - \overline{X}_.)$$

$$b_w = \frac{E_{xy}}{E_{xx}} \text{ or } \frac{(\quad)}{(\quad)} = 1.44$$

67.2/46.8

17. This value indicates that for every unit a score deviates from the grand mean of the covariate, $\overline{X}_.$, it will be expected to deviate ____ units from the grand mean of the dependent variable, $\overline{Y}_.$.

1.44

18. $\overline{Y}_1' = \overline{Y}_1 - b_w(\overline{X}_1 - \overline{X}.) = (\quad) - 1.44\ (\quad - \quad) = 9.0 + .29 = 9.29$

9.0 - 1.44 (3.8 - 4.0)

19. Since group A was below the grand mean $\overline{X}$, $\overline{Y}_A'$ would be __________ (smaller or larger) than $\overline{Y}_A$.

----

larger.

20. $\overline{Y}_2' = (\quad) - (\quad)(\overline{X}_2 - \overline{X}.) = 4.2 = 1.44(4.4 - 4.0) = 3.62$

$\overline{Y}_2 - b_w$

21. $\bar{Y}'_3 = 13.6 - 1.44(3.8 - 4.0) = 13.6 + .29 = 13.89$. The grand mean of the adjusted means, $\bar{Y}'_. = \frac{\Sigma \bar{Y}'_j}{j}$ (when n's are equal), is

$$\bar{Y}. = \frac{(\ ) + (\ ) + (\ )}{3} = \frac{26.80}{3} = 8.93.$$

$$\frac{(9.29) + (3.62) + (13.89)}{3}$$

22. Does $\bar{Y}'_. = \bar{Y}.$? Will this always be the case?

Yes. Yes.

Now let's turn to the question of evaluating our assumptions. (Ideally, one should do this prior to performing the analysis.)

23. An assumption in ANCOVA is that the within-group regression lines are parallel. In more symbolic form: $b_1, b_2 \ldots b_j$ differ only randomly from the parameter, _____; or equivalently; $\beta_1 = \beta_2 = \ldots = \beta_j$.

β

24. In order to test $H_0: \beta_1 = \beta_2 = \ldots = \beta_j$, one compares the pooled variation within each group about its own best-fitting regression line, with the pooled variation within groups about a regression line with the common "average" slope, b_w. We have already computed the latter, which carried the symbol: _________ = 5.5.

E'_{yy}

25. Now the E'_{yy} values (allowing each group to define its own least-squares regression line) are given by:

$$E'_{yy_j} = E_{yy_j} - \frac{(E_{xy_j})^2}{E_{xx_j}}; \text{ e.g., } E'_{yy_1} = 30 - \frac{(20)^2}{14.8} = 2.97$$

$$E'_{yy_2} = \_\_\_\_\_\_ - \frac{(\ )^2}{(\ )} = .70; \; E'_{yy_3} = 33.2 - \frac{(21.6)^2}{14.8} = 1.68$$

$$38.8 - \frac{(25.6)^2}{17.2}$$

26. Then the variation within each group about its own best-fitting regression, summed for all groups, S_1, is _____ + _____ + _____ = 5.35.
(Note that S_1 does not refer to group 1 but is total sum of squares when each group is allowed to define its best fitting regression line.)

2.97 + .70 + 1.68

27. Obviously, S_1 (can or cannot) exceed E'_{yy}

----

cannot

28. When would $S_1 = E'_{yy}$?

----

when all cells had precisely the same regression coefficient, i.e., $b_1=b_2=b_3=b_W$

29. S_1 and E'_{yy} should differ only randomly if H_0: __________ is true.

$\beta_1 = \beta_2 = \ldots = \beta_j$

30. The difference in unpredictable variance, allowing each group to use its own regression coefficient in predicting Y from X, from that in which all use the pooled value is then: S_2 = (   ) - (   ) = 5.5 - 5.35 = .15

$E'_{yy} - S_1$

31. By dividing S_2 and S_1 by their respective degrees of freedom, (J - 1) and J(n - 2), we have two unbiased estimates of population variance which will follow the central F-distribution when $H_0: \beta_1 = \beta_2 = \ldots = \beta_j$ is true.

$$F = \frac{S_2/(J-1)}{S_1/J(n-2)} = \frac{(\quad)/(\quad)}{(\quad)/(\quad)} = \frac{.075}{.594} = .126$$

$$\frac{(.15)/(2)}{(5.35)/(9)}$$

32. Is it necessary to reference the F-table? Why?

No, if $F < 1$, H_0 is never rejected in the typical (one-sided) F-test.

The test of linearity is considerably more involved, the basic rationale being, by allowing a quadratic, or cubic, etc. expression into the regression equation, to give the best-fitting curvilinear regression line, would the Σy^2 be significant less for the curved regression line than for a single straight line ? The researcher usually knows from previous study, the variables which are more likely to be related in a non-linear fashion, i.e., personal, social, affective variable. Curvilinearity may be removed by certain transformations or it may be built in an ANCOVA model (cf. Li, J. C. R., Statistical Inference, Vol. II, 1964). The procedures in a factorial design are the same, the cell being analagous to the group in the present example.

Comparing ANCOVA With Other Analysis Strategies.

It is interesting to compare the ANCOVA results with the probable results had a randomized blocks design been used, blocking on pretest scores.

SV	SS	df	MS	F
Treatments	220.9	2	110.45	235.02
Blocks	98.3	4	24.56	
Error	3.7	8	.47	
	322.9			

33. The MS_e' from ANCOVA is slightly _______ (larger, smaller) than the error MS from the analysis from the randomized blocks design.

    ----

    larger

34. However, the error term in the latter analysis is based on _____(fewer, more) (8 vs.___) degrees of freedom which requires a ______(larger, smaller) F-value in order to reject H_0. In this case for ANCOVA, $_{.95}F_{2,11} = 3.89$, and for the randomized blocks design, $_{.95}F_{2,8} = 4.46$.

    ----

    fewer; 8 vs. 11; larger

35. The randomized blocks analysis is more "robust" in that it is free from assumptions of parallel regression lines and _______ implicit in ANCOVA.

    ----

    linear regression

36. Edwards (1960) performed an ANOVA of the same data using gain scores (posttest-pretest) for each subject

SV	SS	df	MS	F
Treatments	250.5	2	125.3	104.4
Error	14.4	12	1.20	
	264.9			

It is evident in comparing error MS values, that the latter analysis is much ________(more, less) efficient than the ANCOVA and randomized blocks design.

----

less

Post Organizer

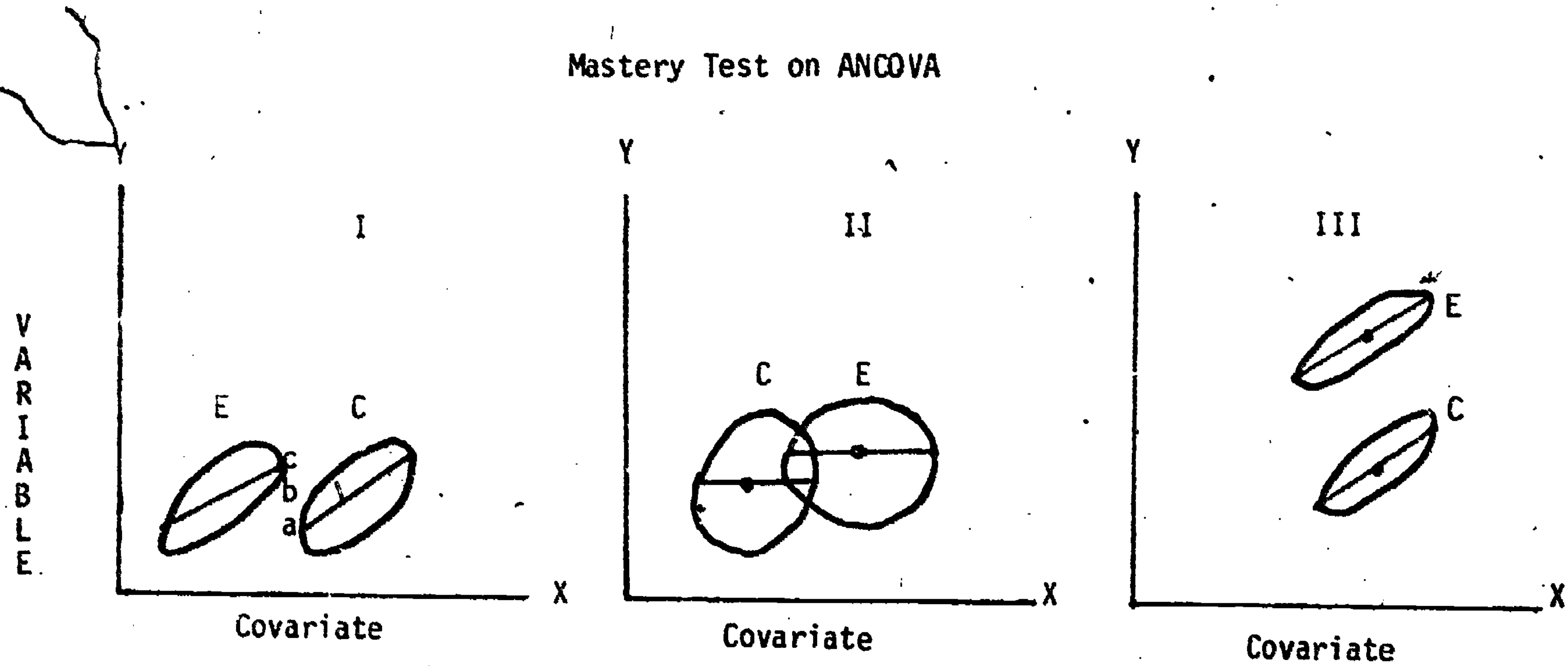
ANCOVA can be a useful statistical tool both for true and quasi-experiments. Its two potential advantages over ANOVA are (1) statistical compensation for pretreatment differences or bias, i.e., removing various "selection" threats to the internal validity of the study, and (2) increasing the power of the analysis.

With respect to the bias removing function it is important to be aware that pretreatment differences may exist on certain unmeasured variables, hence the adjustments are never complete and impeccable. In addition, the statistical compensation will be incomplete to the extent that the covariate is unreliable. ANCOVA cannot bring results from a quasi-experiment to the same level of credibility allowed by a true experiment.

Regarding the increase in power function, ANCOVA can make a substantial contribution to true experiments. If the covariate (or combination or covariates) correlate about .7 with the dependent variable within groups, the gain in power is approximately the same that would accompany a four-fold increase in N. There are other design and analysis strategies for capturing this gain in power, the most common of which is blocking or stratifying on the X-variable. These alternatives are generally preferable if the experimenter has complete control over the conditions of the study since the unique ANCOVA assumptions are of no concern and stratifying allows one to detect interaction effects between the treatments and the X-variable.

The basic ANCOVA rationale extends logically to multiple covariates where the covariates are the predictors in a multiple regression context.

Mastery Test on ANCOVA



1. By examining the situations depicted above, how does the adjusted $MS'_{treatments}$ from ANCOVA, compare to the $MS_{treatments}$ had the covariates been ignored and an ANOVA performed?

   $MS'_{treatments}$ would differ little in situations ________ and ________, and increase in situation ________.

2. In which situation will the adjusted error mean square, MS'_e, differ little from the unadjusted error mean square, MS_e? ____________

3. The gain in power from ANCOVA over ANOVA appears greatest in situation ________.

4. Do the data suggest any serious violation of ANCOVA assumptions? ________

5. Which situations appear to represent quasi-experiments? ________, ________

6. In which situation are the results from ANCOVA would be almost identical to those from ANOVA? ________

7. In figure I, the adjusted mean of the E group would be nearest of point a, b, or c?

8. An additional covariate appears to be needed least in situation I, II, or III?

9. Other things being equal, in which situation has the smallest adjusted error mean square, MS'_e?

10. b_w in group II is about ____________.

ANSWERS:

1. II and III, I	6. II
2. II	7. C
3. III	8. III
4. No	9. III
5. I and II	10. 0.0

REFERENCES

Cochran, W.G. Analysis of covariance: Its nature and uses. Biometrics, 1957, 13, 261-281.

Glass, G.V., Peckham, P.D., & Sanders, J.R. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Review of Educational Research, 1972, 42, No. 3, 237-284.

Hopkins, K.D. Regression and the matching fallacy in quasi-experimental research. The Journal of Special Education, 1969, 3, No. 4, 329-336.

Li, J.C.R. Statistical Inference II. Michigan: Edwards Brothers, Inc., 1964.

Lord, F.M. A paradox in the interpretation of group comparisons. Psychological Bull., 1967, 68, 304-305.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962; 2nd edition, 1971.