DOCUMENT RESUME

ED 317 588 TM 014 646

AUTHOR Loftin, Lynn
TITLE The Extreme Dangers of Covariance Corrections.
PUB DATE Jan 90
NOTE 23p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 25-27, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Covariance; Mathematical Models; *Regression (Statistics); Research Methodology; *Research Problems

ABSTRACT

Although analysis of covariance (ANCOVA) is used fairly infrequently in published research, the method is used much more frequently in dissertations and in evaluation research. This paper reviews the assumptions that must be met for ANCOVA to yield useful results, and argues that ANCOVA will yield distorted and inaccurate results when these assumptions are violated. For ANCOVA to provide meaningful statistical control and to not obscure or mislead, it must be ascertained that the data set fulfills several requirements, especially those pertaining to the homogeneity of regression slopes. For ANCOVA to increase power against a Type II error, there must be a high correlation between the covariate and the dependent variable and no correlation between the covariate and the independent variable. ANCOVA practitioners should examine their data sets carefully to insure that ANCOVA is an appropriate analytic method rather than a method that generates more problems for a given data set than it actually solves. Numerous examples, five figures, and one table are provided. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED317588

ancova.wp 1/22/90

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
LYNN LOFTIN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Extreme Dangers of Covariance Corrections

Lynn Loftin

University of New Orleans 70148

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX, January 25, 1990.

TM014646

ERIC

Abstract

Although ANCOVA is used fairly infrequently in published research, the method is used much more frequently in dissertations and in evaluation research. The purpose of this paper is to review the assumptions that must be met for ANCOVA to yield useful results, and to argue that ANCOVA will yield distorted and inaccurate results when these assumptions are violated. Numerous examples and Venn diagrams are used to make the discussion concrete and more accessible to the nontechnical reader.

Analysis of covariance (ANCOVA) is useful in very limited situations and thus is not employed with great frequency within the behavioral sciences literature (Elmore & Woehlke, 1988; Gaither & Glorfeld, 1985; Goodwin & Goodwin, 1985). However, due to the promise of "control" and "power", some researchers, especially doctoral students (Thompson, 1988), erroneously believe that ANCOVA always provides "control" and more power against Type II error. This misconception accounts for some of the continued use of this method even in cases in which the method is not appropriate.

ANCOVA purports to be a method of "leveling" groups or statistically removing from dependent variable variance the effects of a continuous extraneous variable or variables so that treatment effects can be clarified and the probability of obtaining statistically significant results will be increased (Shavelson, 1981, p. 530; Wildt & Ahtola, 1978). Huitema (1980, p. 13) cites two advantages of ANCOVA: (a) ANCOVA generally has greater power against Type II error than analysis of variance (ANOVA); and (b) ANCOVA reduces the bias caused by variations between experimental groups existing before the treatment was administered. Cliff (1987, p. 272) portrays these advantages in a different light when he says that (a) ANCOVA "tries" to enhance the relationship between dependent and independent variables and (b) ANCOVA "attempts" to correct for extraneous differences and to rule out alternative explanations for findings.

These prospective promises and ANCOVA's favorable treatment in Handbook of Research on Teaching (Campbell & Stanley, 1963) have caused a generation of less informed researchers to regard this method as a statistical panacea that answers many research problems. Unfortunately, the limitations of ANCOVA are not recognized until these apparent advantages are explored thoughtfully and in detail. Explanation of ANCOVA's severe limitations does not appear until later in the chapter in which Shavelson (1981, p. 537) explains the method, in subsequent chapters six and seven of Huitema (1980, pp. 98-156), or in the case of Campbell and Stanley (1963), problems with the use of ANCOVA were not elaborated until a subsequent publication (Campbell & Erlebacher, 1975).

Because of the delay in pointing out problems, and in some cases due to seeming trivialization of the problems accomplished by presenting the assumptions of the methods almost as afterthoughts, ANCOVA procedures are still used inappropriately. The purpose of this paper is to argue that ANCOVA is rarely useful, for reasons that Campbell and Erlebacher (1975) forcefully argue in their subsequent recant. The importance of testing the data set to determine if ANCOVA procedures are appropriate is emphasized.

Conditions Required for Correct ANCOVA Usage

Instead of initially emphasizing the possible benefits derived in those cases in which ANCOVA is indeed appropriate, it might be best to first state the assumptions underlying the successful use of ANCOVA:

1. The covariate (or covariates) should be an independent variable highly correlated with the dependent variable.
2. The covariate should be uncorrelated with the independent variable or variables.
3. With respect to the dependent variable, (a) the residualized dependent variable (Y^*) is assumed to be normally distributed for each level of the independent variable, and (b) the variances of the residualized dependent variable (Y^*) for each level of the independent variable are assumed to be equal.
4. The covariate and the dependent variable must have a linear relationship, at least in conventional ANCOVA analyses.
5. The regression slopes between covariate and independent variable must be parallel for each independent variable group.

With these limitations (Huitema, 1980, Ch. 6) in mind, it seems that the procedure would have limited value, which is probably true. However, in the few cases in which the procedure can be correctly employed, the procedure does have the generally presumed benefits.

The ANCOVA procedure is actually a regression of the dependent variable with the covariate, and then an ANOVA is computed on the adjusted (residual or error) scores on the dependent variable, as illustrated by Thompson (1989). The "correction" of the dependent variable scores is seen by some as a device to adjust for all kinds

ERIC

of problems with random assignment, but very few data sets can meet the very specific requirements that make the adjustment appropriate. And, as pointed out by Campbell and Erlebacher (1975, p. 613), "The more one needs the 'controls' and 'adjustments' which these statistics seem to offer, the more biased are their outcomes." Thompson (1989) notes that the inappropriate use of ANCOVA almost resulted in the termination of major compensatory federal education funding and also stimulated the development of the qualitative research paradigm.

Condition 1

Condition 1 requires a "high" correlation between the covariate (X) and the dependent variable (Y). If the correlation is not high, the covariate will do little to reduce the error sum of squares, and this is the primary objective of ANCOVA. To illustrate this point, Table 1 presents three keyouts for results involving a single data set and one independent variable with three levels. Keyout (a) illustrates a regular ANOVA without a covariate, keyout (b) presents an ANCOVA in which the covariate (X) has a smaller correlation with the dependent variable (r^2 = 15/100 = .15 = 15%), and keyout (c) illustrates an ANCOVA involving a covariate with a larger correlation (r^2 = 30/100 = .30 = 30%) with the dependent variable. The treatment effects (SOS explained = 15) are constant across the analyses, because for this example the covariate is perfectly uncorrelated with the treatment conditions, thus the use of the covariate does not impinge at all on the effect size involving the treatment and the dependent variable (eta^2 = 15/100 = .15 = 15%).

INSERT TABLE 1 ABOUT HERE.

The change from not rejecting the null (Ho) for Table 1 keyouts (a) and (b) to rejecting the treatment effect Ho in keyout (c) is due to the reduction of the sum of squares error when the covariance adjustment is made when the covariate has the larger correlation (r^2 = 30%) with the dependent variable. However, notice that the use of the covariate also causes the loss of one degree of freedom (df) for each covariate, which in and of itself tends to lead to a larger mean square (MS) error, and thus larger calculated *F* values. The change in degrees of freedom also leads to a corresponding increase in the critical *F* used to test the intervention hypothesis, e.g., from 3.35 to 3.37.

In many cases, the decrease in error sum of squares more than compensates for the loss of degrees of freedom in error. But this is **not** always the case. For example, if the covariate is perfectly uncorrelated with the dependent variable, then (a) the sum of squares error will remain totally unchanged after the covariance adjustment, (b) degrees of freedom error will be decreased by one for each covariate, (c) mean square error will then be larger, and (d) *F* calculated will then become smaller, making it less likely to get statistical significance. Critical *F* values will change as well, as illustrated in the Table 1 example.

Figure 1 uses Venn diagrams to illustrate these dynamics. The circle in (a) represents the total sum of squares on the dependent variable. The diagram in (b) shows a covariate (X) with a fairly small correlation (r^2 = 15%) with the dependent variable (Y) and no correlation between X and the treatment variable. Diagram (c) shows a covariate (X) with a higher correlation (r^2 = 30%) with Y. The sum of squares residual after the covariance adjustment is further reduced in this analysis, thanks to the higher correlation between X and Y. The more "pie" (sum of squares total) that is consumed by the covariate regression, the smaller the sum of squares error becomes, as long as X and the treatment variable are perfectly uncorrelated (and thus do not overlap in the Venn diagram). This situation makes the mean square error smaller and helps yield significant treatment effects because the smaller mean square error then ultimately leads to larger calculated *F* values for treatment effects.

INSERT FIGURE 1 ABOUT HERE.

<u>Condition 2</u>

Condition 2 requires the covariates and the independent variables to be perfectly uncorrelated, as they were in the previous examples. The best way to illustrate this is with the Venn diagram in Figure 2. The desired situation is (a) where the covariate (X) and the independent variable are perfectly uncorrelated and thus explain different portions of the total variance of the dependent variable. In other words, they "eat

ERIC

different parts of the sum of squares (SOS) pie." In (b) the covariate (X) and the treatment variable are somewhat correlated. In this case, because the covariance adjustment is made first in ANCOVA, all of the sum of squares explained by both X (15) and the treatment variable (10) will be attributed **solely** to the covariate. In this case the estimated treatment effects will be reduced (15 - 10 = 5).

INSERT FIGURE 2 ABOUT HERE.

Part (c) of Figure 2 presents the worst possible case, since the covariate explains all the sum of squares total also predicted by the treatment variable. The effect of the treatment is obfuscated by the covariate causing confusion regarding the effects of the intervention. The researcher in this case would erroneously conclude that the treatment had no effect whatsoever (15 - 15 = 0).

Condition 3

The first aspect of Condition 3 assumes a normal distribution of the residualized dependent variable ($Y^* = Y - YHAT = Y - (a + bX)$) for each independent variable group. ANCOVA is apparently more sensitive to the violation of this distribution assumption than is ANOVA. However, if the covariate (X) is normally distributed, ANCOVA is reasonably robust to the violation of this assumption (Huitema, 1980, p. 117).

The second aspect of Condition 3 assumes that variances of the residualized dependent variable (Y^*) are assumed to be equal for each independent variable group. If sample sizes are equal and random, this requirement is usually met. This requirement is illustrated graphically in Figure 3.

INSERT FIGURE 3 ABOUT HERE.

Condition 4

The fourth requirement concerns the regression analysis used to "correct" or "adjust" the dependent variable (Y) scores using the covariate (X). A regression assumes a straight line, i.e., linear relationship between (Y) and (X) for each level of the dependent variable. The regression line is the best representation of the relationship between (Y) and (X) in the sense that the "best fit" regression line minimizes mean square error. The regression line can be used to predict (Y) from (X) using the straight line equation:

$$Y^*_i = a + b(X_i) \ ,$$

where, b = slope of the regression line,

X_i = the score on X of any given individual, and

a = the Y intercept of the regression line when X = 0.

A straight line relationship between dependent variable and covariate means that the relationship does not change at some point on the covariate, i.e., that the relationship is constant within the range of the two variables tested. Thus, a change in magnitude on X is presumed to cause a proportional change on Y at each level of X. This condition is extremely important if the ANCOVA adjustment is not to bias the data set. Figure 4 illustrates how a scatterplot for Y and X scores in each independent variable group can indicate or contraindicate linearity of the relationship between the dependent variable and the covariate.

INSERT FIGURE 4 ABOUT HERE.

Condition 5

The use of ANCOVA requires that the regression slopes (b) of the dependent variable Y versus the covariate (X) be equal for each level of the independent variable. This "homogeneity of regression" condition indicates that the relationship between the dependent variable Y and the covariate X is constant for all levels of the independent variable. Any adjustment in the covariate (X) will result in the same proportionate adjustment in the dependent variable (Y) for each level of the independent variable. If the regression slopes are equal, a single pooled regression slope may be used for all groups to calculate the solicited adjustments in the dependent variable (Y). For example, for a one-way three-level ANCOVA the assumption is:

$$b \text{ pooled} = b \text{ group } 1 = b \text{ group } 2 = b \text{ group } 3$$

The test for homogeneity of regression slopes is based on the concept that each regression line produces the least sum of squares error (Y) for each level of the independent variable. The pooled regression line totally ignores the level of the independent variable during the statistical adjustments invoked by the regression procedure. The slopes are homogeneous if the residual sums of squares from separate regression analyses equal the residual sum of squares produced by a pooled regression equation developed by ignoring group membership on other independent variables.

An *F* test can be computed to test the homogeneity of regression to determine if the difference in the slopes across groups is statistically significant and the slopes are not equal. If this proves to be the case, ANCOVA is not appropriate for the data set. But as vitally as important as the homogeneity of regression assumption is, Willson (1982, p. 6) found that "few [published] studies using straight ANCOVA tested for homogeneity." Thompson (1988) found the same thing with respect to doctoral dissertations. Figure 5 illustrates three examples of the pooled regression.

INSERT FIGURE 5 ABOUT HERE.

The ANCOVA Correction

If the regression lines for the dependent variable versus the covariate are parallel for each level of the independent variable, the correction can be determined from the graph from the mean of the pooled regression equation. The point where this mean intersects the regression line for each independent variable level will become the new adjusted dependent variable mean. All Y scores in each level will be accordingly adjusted into new Y^* scores which are then subjected to an ANOVA (Thompson, 1989).

This should make it clear why homogeneity of regression is important. The adjustment of Y to the mean of X is proportional for each regression line. Without that relationship, the adjusted Y (Y^*) would change with respect to the level of the independent variable at each level of X. The mathematical expression for calculating the adjusted scores, Y^*, assumes a straight line relationship between X and Y, and a common slope for the single, corrective, "pooled" regression equation across groups. **Since a single regression equation is used to correct all Y scores, regardless of independent variable groups, if the "pooled" or "averaged" single equation does not accurately describe the Y and X relationship in a given group, the corrections producing the residual Y scores, Y^* (Y - YHAT = Y - (a + bX)), will actually bias the data rather than "correct" them.**

The expression for formulating corrected group means (Huitema, 1980, p. 32) is:

$\text{Ymean}^*_j = \text{Ymean}_j - \text{pooled b } (\text{Xmean}_j - \text{X grand mean})$,

where

pooled b = pooled slope of the XY regression line developed by <u>ignoring</u> group membership on the independent variable(s), and

j = level or group on the independent variable.

Adjusting scores of data in which the homogeneity of regression requirement is not met changes the relationship of the levels to each other and completely muddles the data.

<u>Summary</u>

ANCOVA is perceived by some researchers to be a miracle method that magically produces correct statistical control and always yields greater power against Type II error. Unfortunately, these benefits are not always realized and ANCOVA must be used with considerable caution. In addition to normal statistical assumptions for ANOVA and regression procedures, ANCOVA data sets must meet other stringent requirements.

For ANCOVA to provide meaningful statistical control, and not obscure or mislead, extreme caution must be used to ascertain that the data set fulfills these several requirements, especially those pertaining to the homogeneity of regression slopes. For ANCOVA to increase power against a Type II error, there must be a high correlation between the covariate and the dependent variable and no correlation between the covariate and the independent variable. Thus, readers of ANCOVA results should carefully examine research studies for reports of evidence that these requirements were met. And ANCOVA practitioners should thoughtfully examine their data sets to insure that ANCOVA is an appropriate analytic method, rather than a method that for a given data set will generate more problems that the analysis actually solves.

References

Campbell, D. T., & Erlebacher, A. (1975). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In M. Guttentag & E. L. Struening (Eds.), Handbook of evaluation research (pp. 597-617, Vol. 1). Beverly Hills: SAGE.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching (pp. 171-246). Chicago: Rand McNally.

Cliff, N. (1987). Analyzing multivariate data. San Diego: Harcourt, Brace, Jovanovich.

Elmore, P. B., & Woehlke, P. L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Gaither, N., & Glorfeld, L. (1985). An evaluation of the use of tests of significance in organizational behavior research. Academy of Management Review, 10, 787-793.

Goodwin, L. D., & Goodwin, W. L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.

Huitema, B. E. (1980). The analysis of covariance and alternatives. New York: John Wiley & Sons.

Shavelson, R. J. (1981). Statistical reasoning for the behavioral sciences (2nd ed.). Boston: Allyn & Bacon.

Thompson, B. (1988, November). Common methodology mistakes in dissertations: Improving dissertation quality. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, KY. (ERIC Document Reproduction Service No. ED 301 595)

Thompson, B. (1989). The place of qualitative methods in contemporary social science. In B. Thompson (Ed.), Advances in social science methodology (pp. 1-49, Vol. 1). Greenwich, CT: JAI Press.

Wildt, A. R., & Ahtola, O. T. (1978). Analysis of covariance. Beverly Hills: SAGE.

Willson, V. L. (1982, February). Misuses of regression approaches to ANCOVA and ANCOVA. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 222 522)

Table 1

Three Keyouts for Three-level One-way Design (n=30)

(a) ANOVA--treatment Ho not rejected

Source	SOS	dF	MS	Fcalc	Fcrit	Effect size
Treatment	15	2	7.5	2.3825	3.35	.15
Error	85	27	3.1481			
Total (Y)	100	29				

(b) ANCOVA with smaller correlation between X and Y and zero correlation between X and treatment--treatment Ho not rejected

Source	SOS	dF	MS	Fcalc	Fcrit	Effect size
Covariate (X)	15	1	15	5.5714	4.22	.15
Treatment	15	2	7.5	2.7857	3.37	.15
Error	70	26	2.6923			
Total (Y)	100	29				

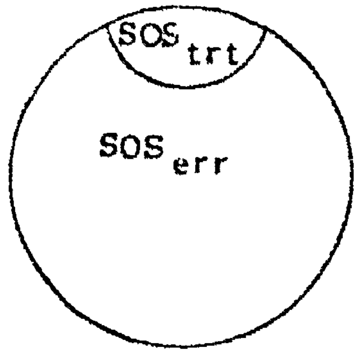
(c) ANCOVA with higher correlation between X and Y and zero correlation between X and treatment--treatment Ho rejected

Source	SOS	dF	MS	Fcalc	Fcrit	Effect size
Covariate (X)	30	1	30	14.1817	4.22	.30
Treatment	15	2	7.5	3.5455	3.37	.15
Error	55	26	2.1154			
Total (Y)	100	29				

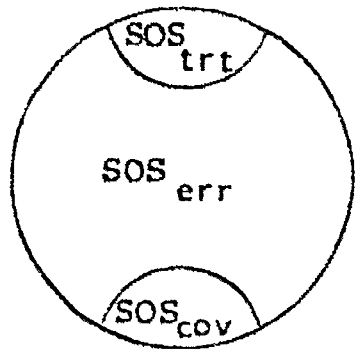
ERIC

Figure 1. Three Venn Diagrams Illustrating the Reduction of SOS_{err} with ANCOVA

(a) ANOVA without a covariate, SOS_{trt} explains .15 of SOS_y



(b) ANCOVA with covariate (X) that explains a unique .15 of SOS_y



(c) ANCOVA with covariate (X) that explains a unique .30 of SOS_y

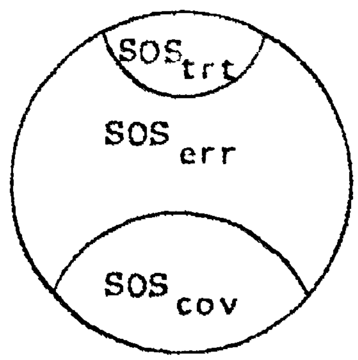
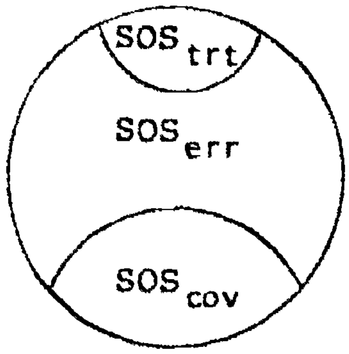
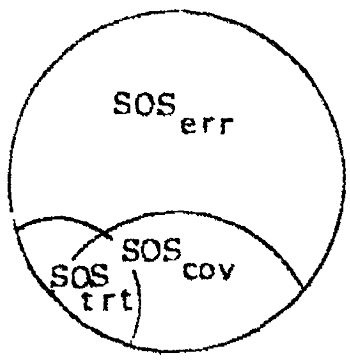


Figure 2. Venn Diagrams Illustrating the Necessity for the Covariate and the Treatment to be Uncorrelated



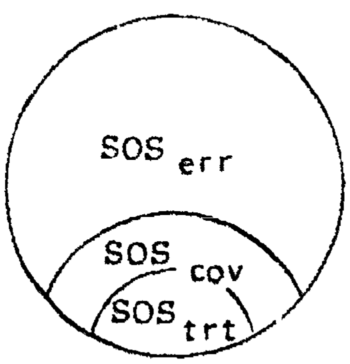
Source	SOS	df	MS	Fcalc	Fcrit	Eff Size
Covariate	30	1	30.00	14.18	4.22	.30
Treatment	15	2	7.50	3.54	3.37	.15
Error	55	26	2.12			
Total	100	29				

(a) Treatment and covariate (X) uncorrelated



Source	SOS	df	MS	Fcalc	Fcrit	Eff Size
Covariate	30	1	30.00	12.00	4.22	.30
Treatment	5	2	2.50	1.00	3.37	.05
Error	65	26	2.50			
Total	100	29				

(b) Treatment and covariate (X) partially correlated

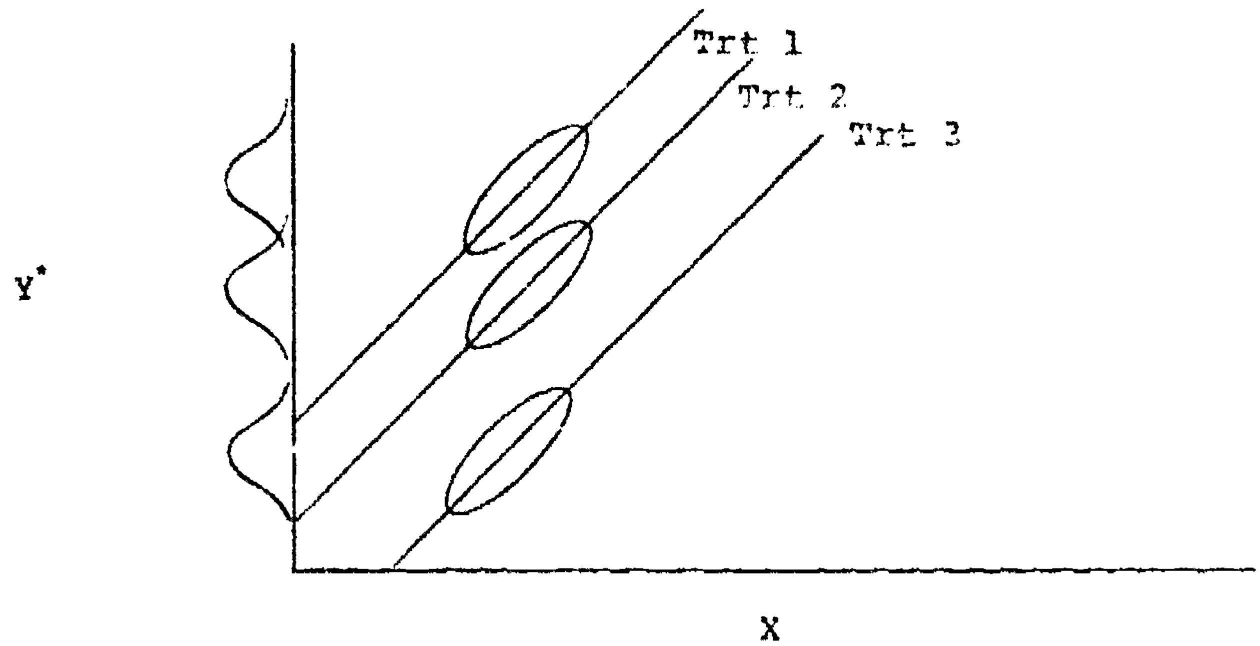


Source	SOS	df	MS	Fcalc	Fcrit	Eff Size
Covariate	30	1	30.00	11.14	4.22	.30
Treatment	0	2	0.00	----	3.37	.00
Error	70	26	2.69			
Total	100	29				

(c) Covariate subsumes Treatment SOS

Figure 3. Illustrations of Normal Distribution and
Homogeneity of Variance Assumptions

(a) The residualized dependent variable (Y^*) has a normal distribution within each treatment group



(b) The residualized dependent variable (Y^*) has equal variances for each treatment group at each level

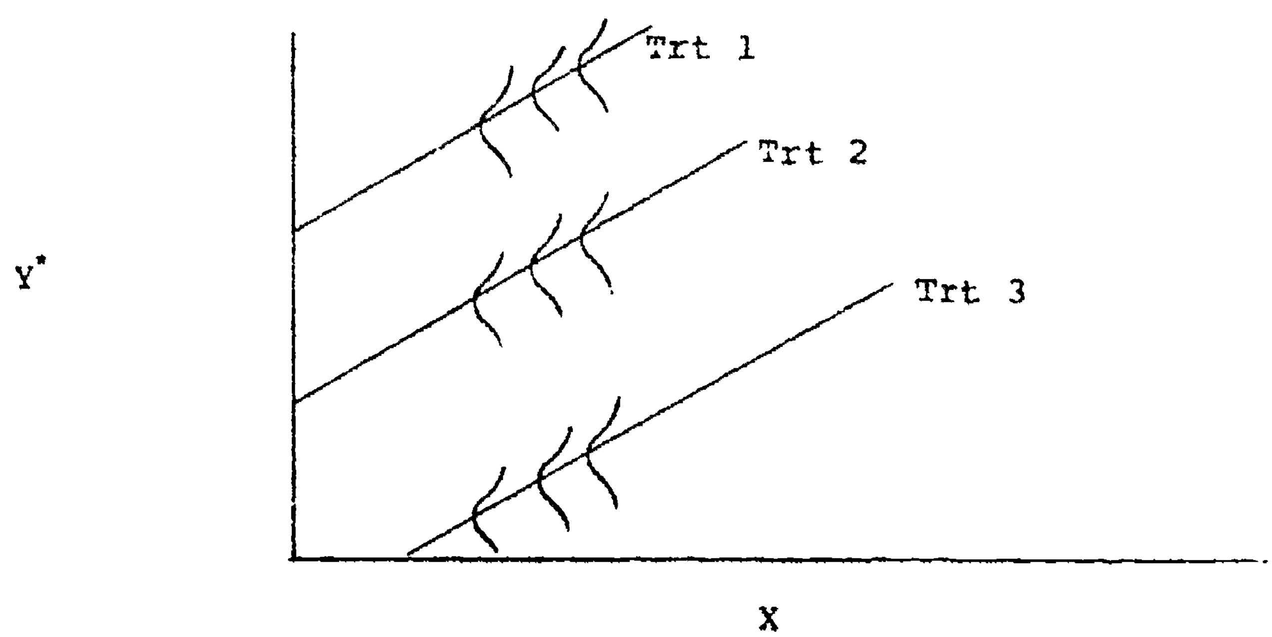
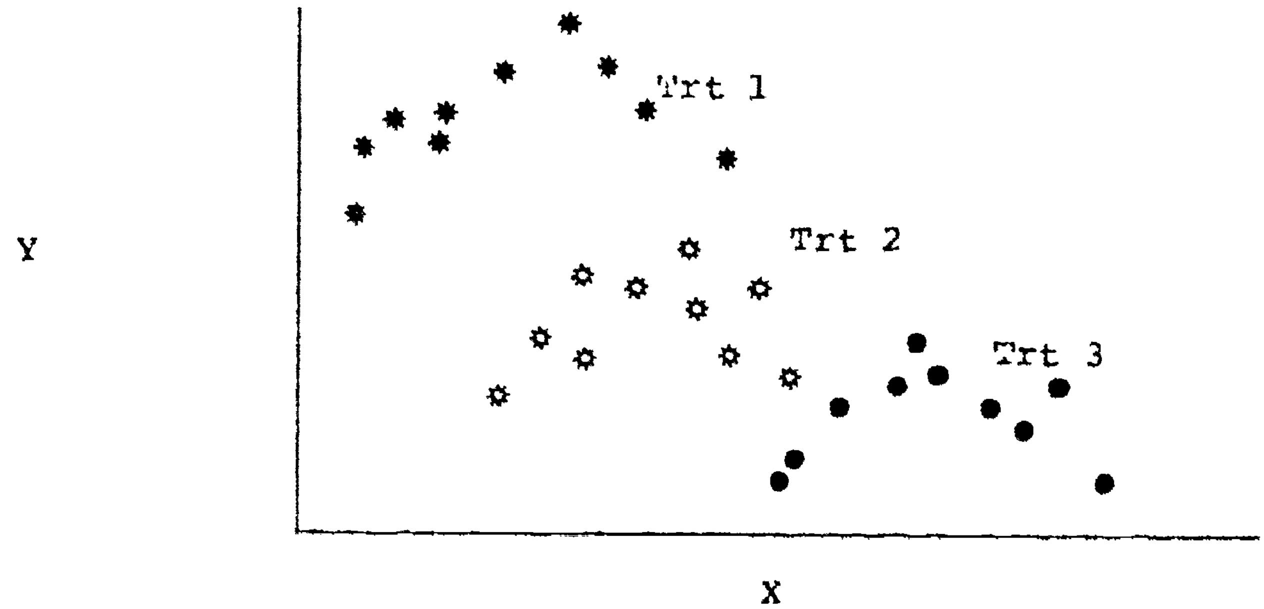


Figure 4. Illustrations of Requirement that Y and Covariate (X) be Linearly Related

(a) Data points probably not suited for ANCOVA



(b) Data points that define straight lines that would be suitable for this ANCOVA assumption

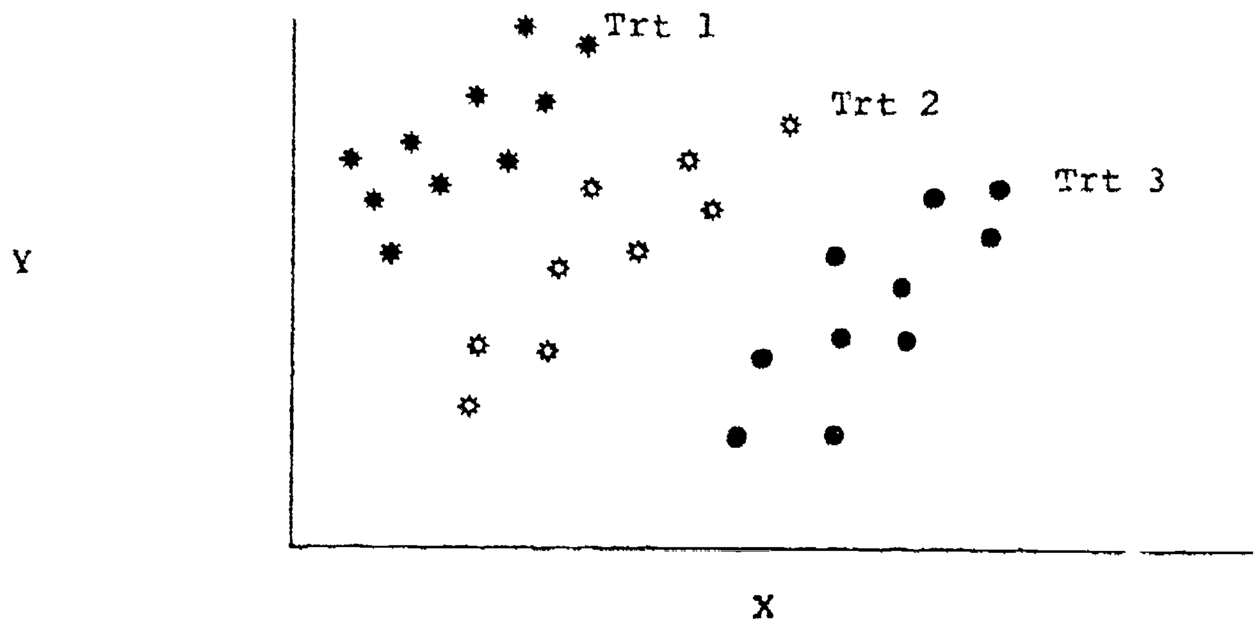
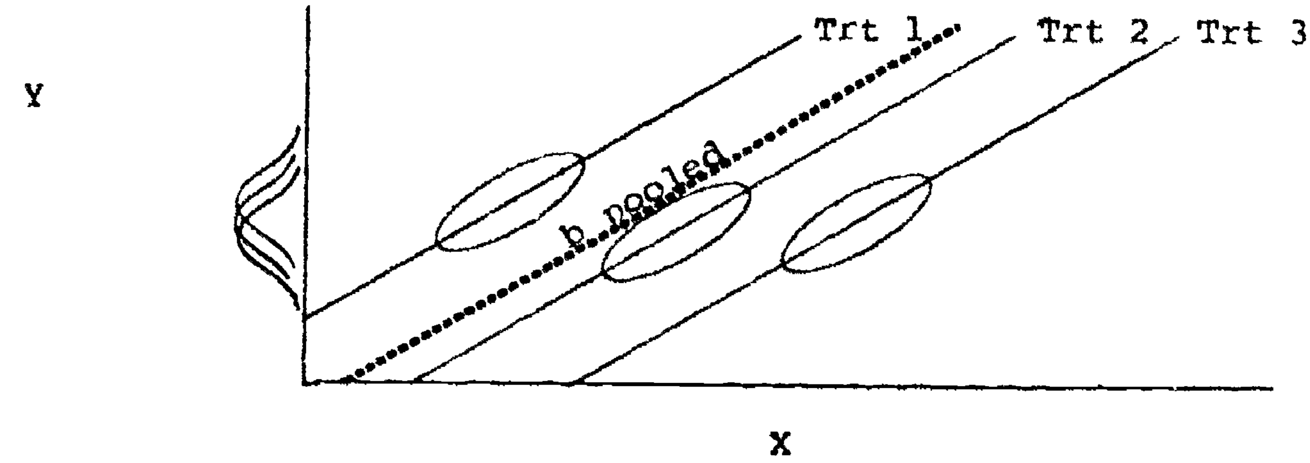
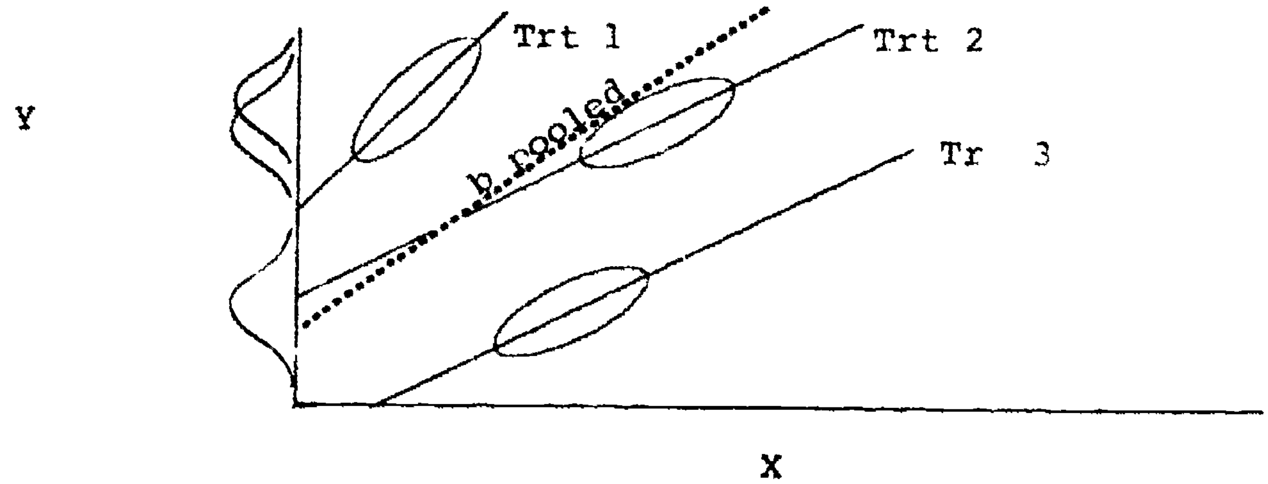


Figure 5. Illustration of Various Slopes Across Treatment Groups

(a) Slopes are equal, means of Y are very close



(b) Slopes are not equal, and any adjustment using X would bias the data



(c) Slopes are equal, means of Y are separated