ANOVA of Residuals and ANCOVA: Correcting an Illusion by Using Model Comparisons and Graphs
Author(s): Scott E. Maxwell, Harold D. Delaney and Jerry M. Manheimer
Source: *Journal of Educational Statistics*, Vol. 10, No. 3, Special Issue: Teaching Statistics (Autumn, 1985), pp. 197-209
Published by: American Educational Research Association and American Statistical Association
Stable URL: https://www.jstor.org/stable/1164792
Accessed: 17-09-2018 15:22 UTC

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/1164792?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

American Educational Research Association, American Statistical Association are collaborating with JSTOR to digitize, preserve and extend access to *Journal of Educational Statistics*

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

Journal of Educational Statistics
Fall 1985, Volume 10, Number 3, pp. 197–209

ANOVA OF RESIDUALS AND ANCOVA: CORRECTING AN ILLUSION BY USING MODEL COMPARISONS AND GRAPHS

SCOTT E. MAXWELL
University of Notre Dame
HAROLD D. DELANEY and JERRY M. MANHEIMER
University of New Mexico

KEY WORDS. *Analysis of covariance, teaching statistics, residuals, model comparisons, least squares, linear models.*

ABSTRACT. Analysis of covariance is often conceptualized as an analysis of variance of a single set of residual scores that are obtained by regressing the dependent variable on the covariate. Although this conceptualization of an equivalence between the two procedures may be intuitively appealing, it is mathematically incorrect. If residuals are obtained from the pooled within-groups regression coefficient (b_W), an analysis of variance on the residuals results in an inflated α-level. If the regression coefficient for the total sample combined into one group (b_T) is used, ANOVA on the residuals yields an inappropriately conservative test. In either case, analysis of variance of residuals fails to provide a correct test, because the significance test in analysis of covariance requires consideration of both b_W and b_T, unlike analysis of residuals. It is recommended that the significance test of treatment effects in analysis of covariance be conceptualized, not as an analysis of residuals, but as a comparison of models whose parameters are estimated by the principle of least squares. Focusing on model comparisons and their associated graphs can be used effectively here as in other cases to teach simply and correctly the logic of the statistical test.

One of the goals of most higher level statistics courses is to have the student perceive certain fundamental relationships between various statistical techniques such as the one between analysis of variance (ANOVA) and regression. ANOVA and regression meet, of course, in the analysis of covariance (ANCOVA), a topic that many students have difficulty mastering. It is easy for the student to miss the logic of the procedure by an overemphasis on computational formulas. To counter this difficulty, some instructors and authors, in explaining ANCOVA, make use of the notion of analyzing residual scores. Residual or error scores can be used effectively in explanations of ANCOVA and other statistical techniques; unfortunately, an equivalence that some think exists between ANCOVA and the analysis of variance of a set of residuals actually is illusory and needs to be eliminated from pedagogy concerning ANCOVA.

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

ANCOVA frequently has been erroneously presented as an ANOVA of a set of residual scores that are obtained when the dependent variable is regressed on the covariate. To cite one instance, in discussing an example in which IQ is used as a covariate and the scores students earn on a posttest are the values of the dependent variable, Marascuilo (1971) states,

> Covariance adjustment is equivalent to projecting the earned score in a direction parallel to the regression line to the IQ score defined at $\bar{X}$ [where $\bar{X}$ is the grand mean for the covariate]. This parallel projection is performed for all pairs of observations and an analysis of variance is then performed on the adjusted scores (p. 499).

Although the correct computational formulas for ANCOVA are invariably presented, a misleading reliance on the notion of analyzing a set of residuals is a common pedagogical device for introducing ANCOVA. Among authors who take this approach, some leave open the possibility that the analogy is not exact; others explicitly maintain that the two procedures are identical. Snedecor and Cochran (1967) illustrate the first approach when they state that, "The analysis of covariance is *essentially* an analysis of variance of the quantity $(Y - bX)$" (p. 424) [emphasis added]. Cohen and Cohen (1983) illustrate the more explicit approach in their text when they say,

> The ACV [ANCOVA] involves the analysis of (the residuals of) Y when one or more other variables (the covariates) have been partialed out. . . . In ACV [ANCOVA], the residual that is analyzed is $Y - \hat{Y}_A$ for each subject . . . in *exactly* the same way as Y itself is analyzed in AV [ANOVA] (p. 318) [emphasis added].

Similar statements can also be found in such sources as Dixon and Massey (1969, p. 223), Pedhazur (1982, pp. 496–497), and Lindquist (1953, p. 318). Of course, these authors are aware that ANCOVA and an analysis of variance conducted on the residuals (hereafter called ANORES) will be slightly different because the ANORES will have one additional but artificial degree of freedom. With an adjustment of the ANORES degrees of freedom, students are almost certain to get the impression that ANCOVA and ANORES should be identical. This seems to be what at least some authors intend.

That they are not identical thus needs to be emphasized. It is the purpose of this paper to develop the way in which ANORES and ANCOVA differ.

It also should be noted, though, that there clearly is evidence for a second, distinct approach to training students regarding the relationship between ANCOVA and ANORES. This is illustrated by the tack taken in writings by Werts and Linn (1970) and Corder-Bolz (1978). These writers are explicit about the fact that the two procedures differ, but suggest that ANORES is used by many as a viable alternative to the analysis of data that could be analyzed by ANCOVA. For example, Werts and Linn suggest that the "traditional procedure" for analyzing pre-post designs is "to remove the effect of

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

initial status by regressing the final score on the initial score, yielding the deviation of the final score from its predicted value. This deviation score is then used to find the correlates of change" (p. 17). Thus, students exposed to this approach to teaching about ANCOVA and ANORES might not conclude that one method is right and the other wrong, but may come to the conclusion instead that the procedures represent two viable analyses, either of which could be legitimately carried out on a given set of data. We believe that would be an erroneous conclusion. In fact, it is a second purpose of this paper to argue that when ANCOVA could be validly applied to particular data, ANORES would not be appropriate.

The problem will be approached by taking a perspective that works effectively in many other cases as well, namely by focusing on the model comparison implicit in the statistical test and using graphs to picture the models' fit to the data. Residuals are used, but one must consider the two models involved in the computation of *two* sets of residual scores. A focus either on the implied model comparison or on a geometrical representation of the alternative analyses makes clear the ways in which ANORES differs from ANCOVA. A numerical example illustrating the difference between the two procedures will be presented at the end of the paper.

ANCOVA

The analysis of covariance test of group differences can be conceptualized as the comparison of the following full (F) and restricted (R) models:

$$\text{Full:} \quad Y_{ij} = \mu + \alpha_j + \beta X_{ij} + \epsilon_{ij} \tag{1}$$

$$\text{Restricted:} \quad Y_{ij} = \mu + \beta X_{ij} + \epsilon_{ij} \tag{2}$$

where Y_{ij} is the score on the dependent variable of the ith subject in the jth group, μ is a "grand mean" parameter (appropriately thought of as the intercept in Equation 2 or the mean of the intercepts in Equation 1), α_j is a parameter indicating the effect of the jth treatment, β is a population regression coefficient, X_{ij} is the score on the covariate for the ith subject in the jth group, and ϵ_{ij} is an error term for the ith subject in the jth group. The models are compared by using least squares to estimate each parameter in each model and then comparing the error sum of squares (SSE) for the two models. The SSE's are based on the individual errors determined uniquely for that model, that is, the errors used are the deviations of the observations from the predictions computed using the numerical estimates of the parameters in that model. With the usual ANCOVA assumptions (see, e.g., Elashoff, 1969; Glass, Peckham, & Sanders, 1972), the following expression has a central F distribution with $k - 1$ and $N - k - 1$ degrees of freedom if the null hypothe-

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

sis of no treatment effects is true (k is the number of treatment groups and N is the total sample size):

$$F = \frac{[SSE(R) - SSE(F)]/(k - 1)}{SSE(F)/(N - k - 1)}. \tag{3}$$

The model comparisons approach makes clear that the F test in ANCOVA is a function of the extent to which scores on the dependent variable can be more accurately predicted if group membership is known than if it is not, where prediction is performed in both models by using least squares. Of crucial importance later in our argument is the estimation of β in the full and restricted models.

A common misconception among students when first presented with these models seems to be that because both models contain a β parameter, the least squares estimate of β in the full model will be identical to the least squares estimate of β in the restricted model. However, in the full model the estimator is b_W, the pooled within-groups regression coefficient for Y regressed on X, whereas in the restricted model the estimator is b_T, the regression coefficient for Y on X for the total sample of observations combined into one group. In the absence of group membership parameters, optimal prediction is obtained by using b_T as the slope coefficient, whereas when different intercepts for each group are allowed for by the introduction of group membership parameters but a common slope of Y on X is assumed, optimal prediction is obtained by using b_W. It should be noted that rarely will $b_W = b_T$. Even the corresponding β parameters being estimated (i.e., β in the full and restricted models, respectively) will themselves differ unless the null hypothesis is exactly correct or the mean on the covariate is exactly the same in each group. Of course, even if the β parameters were the same, b_W and b_T would almost certainly differ because of sampling variability.

The distinction between b_W and b_T, as well as the introduction of b_B, the between groups regression of $\bar{Y}_j$ on $\bar{X}_j$, often is difficult to explain to students when relying on extensions of the ANOVA approach, and hence contributes to the confusion surrounding ANCOVA. In contrast, the model comparisons approach (cf. Namboodiri, Carter, & Blalock, 1975; Searle, 1971) making the utilization of least squares explicit shows why it is necessary to define both b_W and b_T to discuss ANCOVA. However, b_B is not a least-squares estimator of either the β in the full model or the β in the restricted model, and thus does not necessarily need to be introduced in teaching ANCOVA. In fact, b_B seems to be of limited value in general, except when multilevels of unit of analysis are considered. For example, analyses might be conducted both at the level of students and classrooms, in which case b_B might be of interest. For further discussion, see Burstein, Linn, and Capell (1978).

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

ANORES

ANORES can also be conceptualized in terms of model comparisons. One virtue of this approach is that it necessitates explicit consideration of how the residuals are to be obtained, because either b_W or b_T could be used to define a residual score. Cohen and Cohen (1983, p. 318) argue that b_W should be used. This also seems to be the intent of Pedhazur (1982, especially equations 13.1 and 13.2, pp. 496–497). However, because none of the other previously referenced sources have stated which coefficient should be used, we will examine both, beginning with b_W. The residual score for the ith subject in the jth group can then be written as $Y_{ij} - b_W X_{ij}$. The full and restricted models compared by ANORES using b_W, which will be denoted models *FW* and *RW*, respectively, may then be written as follows:

$$FW:\quad Y_{ij} - b_W X_{ij} = \mu + \alpha_j + \epsilon_{ij} \tag{4}$$

$$RW:\quad Y_{ij} - b_W X_{ij} = \mu + \epsilon_{ij}. \tag{5}$$

The significance test is again obtained by comparing the error sum of squares for the two models as follows:

$$F = \frac{[SSE(RW) - SSE(FW)]/(k-1)}{SSE(FW)/(N-k-1)}. \tag{6}$$

The term $N - k - 1$ appears as the denominator degrees of freedom because β has been estimated to obtain the residual scores. The relevant question at this point is how this F test relates to that in Equation 3 from ANCOVA, which translates to how the errors associated with the models compare. First, consider the relationship between the ANCOVA full model and the full model for ANORES using b_W. It can be shown that $SSE(F)$ equals SS_W for ANCOVA, that is, the adjusted within-group sum of squares, which in turn equals (see Kirk, 1982, p. 723):

$$SS_W = \sum_j \sum_i (Y_{ij} - \bar{Y}_j)^2 - b_W^2 \sum_j \sum_i (X_{ij} - \bar{X}_j)^2. \tag{7}$$

The error sum of squares for model *FW*, on the other hand, is the within-group sum of squares for the dependent variable $Y_{ij} - b_W X_{ij}$. Since

$$SSE(FW) = \sum\sum [Y_{ij} - b_W X_{ij} - (\bar{Y}_j - b_W \bar{X}_j)]^2 \tag{8}$$

and

$$b_W = \frac{\sum_i \sum_j (X_{ij} - \bar{X}_j)(Y_{ij} - \bar{Y}_j)}{\sum_j \sum_i (X_{ij} - \bar{X}_j)^2}, \tag{9}$$

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

algebraic manipulation leads to

$$SSE(FW) = \sum_j \sum_i (Y_{ij} - \overline{Y}_j)^2 - b_W^2 \sum_j \sum_i (X_{ij} - \overline{X}_j)^2 = SSE(F). \qquad (10)$$

Consider next the relationship between the restricted models for ANCOVA and ANORES $- b_W$, which we have denoted models R and RW, respectively. In model R, estimates for μ and β are arrived at so as to minimize the error sum of squares for such a two-parameter model. In model RW, however, only μ is estimated through least squares. This estimate is given by

$$\hat{\mu} = \overline{Y} - b_W \overline{X}. \qquad (11)$$

It must be the case that $SSE(RW) \geq SSE(R)$, because least squares for model R could always "choose" $\hat{\mu}$ to be as in (11) and $\hat{\beta} = b_W$, duplicating the estimates of model RW; otherwise, the estimates of model R will differ and provide a better fit to the data, yielding a smaller error sum of squares, for this is precisely the goal of the least squares procedure. Reference to the formulas for the F test in ANCOVA (Equation 3) and ANORES (Equation 6) reveals that the observed F ratio for ANORES must be at least as large as the F for ANCOVA, because $SSE(F) = SSE(FW)$ but $SSE(RW) \geq SSE(R)$. The extent of the discrepancy will depend on the extent to which b_W differs from b_T; it can be shown that

$$SSE(RW) = SSE(R) + (b_W - b_T)^2 \sum_j \sum_i (X_{ij} - \overline{X})^2. \qquad (12)$$

This relationship shows that the test given in Equation 6 is not a legitimate F test, because the sampling distribution of the statistic differs systematically from the sampling distribution of the proper test statistic. The reason that Equation 6 is inappropriate is that the numerator expression is not distributed as a chi-squared random variable with $k - 1$ degrees of freedom. This can be seen clearly by considering the distribution of the numerator and denominator of the correct ANCOVA test. We know that under the ANCOVA assumptions, $SSE(F)$ for ANCOVA (see Equation 3), and hence $SSE(FW)$ for ANORES with b_W (see Equations 6 and 10), when divided by σ^2, that is, by the variance of the residual scores ϵ_{ij} in the ANCOVA full model, is distributed as a chi-squared random variable with $N - k - 1$ degrees of freedom. Similarly, under the null hypothesis, the ANCOVA expression $[SSE(R) - SSE(F)]$, when divided by the same σ^2, is distributed as a chi-squared random variable with $k - 1$ degrees of freedom. However, the ANORES expression, $[SSE(RW) - SSE(FW)]$, is systematically larger than that in ANCOVA, and so when divided by the same σ^2 cannot also have a chi-squared distribution with $k - 1$ degrees of freedom: Its expected value, for example, will be greater than $k - 1$. Hence, ANORES using b_W to obtain residual scores will lead to

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

an inflated α-level, and is certainly not equivalent to ANCOVA.

Although the use of b_W to obtain residuals does not reproduce ANCOVA, it is also possible to perform ANORES using b_T to obtain residuals. Perhaps it is this form of ANORES that certain authors have had in mind when they have written of the equivalence between ANCOVA and ANORES. It should be noted that this approach is often referred to as a residual gain analysis (e.g., Corder-Bolz, 1978). With this method of forming residual scores, the dependent variable for the ith subject in the jth group is $Y_{ij} - b_T X_{ij}$. The full and restricted models to be compared, which we will denote FT and RT, respectively, are then:

$$FT: \quad Y_{ij} - b_T X_{ij} = \mu + \alpha_j + \epsilon_{ij} \tag{13}$$

$$RT: \quad Y_{ij} - b_T X_{ij} = \mu + \epsilon_{ij}. \tag{14}$$

Once again, we might perform a significance test by comparing the error sum of squares for the two models as

$$F = \frac{[SSE(RT) - SSE(FT)]/(k-1)}{SSE(FT)/(N-k-1)}. \tag{15}$$

As before, $N - k - 1$ appears in the denominator because β has been estimated to obtain the residual scores. Now, models F, R, FT, and RT must be compared. Models R and RT are identical because both include b_T as the slope value and $\hat{\mu} = \overline{Y} - b_T \overline{X}$ in both cases. Hence,

$$SSE(RT) = SSE(R). \tag{16}$$

Consider next the relationship between models F and FT. In model F, estimates for μ, α_j and β are chosen so as to minimize the sum of squared errors, by the definition of least squares. In model FT, least squares estimates are obtained for μ and α_j subject to the constraint that $\hat{\beta} = b_T$. However, b_W is the least squares estimate and therefore must lead to a minimal sum of squared errors. Hence,

$$SSE(FT) \geq SSE(F). \tag{17}$$

Again, the extent of the difference in error sum of squares is related to the difference between b_W and b_T. In particular,

$$SSE(FT) = SSE(F) + (b_W - b_T)^2 \sum\sum (X_{ij} - \overline{X}_j)^2. \tag{18}$$

Referring to Equations 3 and 15 shows that the "F value" obtained by using this approach to ANORES must result in a value that is less than or equal to the F obtained with ANCOVA, with equality holding only if $b_W = b_T$. Thus, this form of ANORES also fails to be equivalent to ANCOVA and fails to provide a valid F test, for much the same reason as did ANORES with b_W.

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

Under the ANCOVA null hypothesis and assumptions, the ANCOVA F will be distributed as a central F with $k - 1$ and $N - k - 1$ degrees of freedom. A random variable such as that yielded by Equation 15, which has systematically smaller values, cannot also be said then to have a central F distribution with $k - 1$ and $N - k - 1$ degrees of freedom.

Thus, claims that ANORES and ANCOVA are equivalent are false, whichever approach to ANORES is employed. The fact that models F and FW are equivalent and that models R and RT are also equivalent suggests that it is possible to duplicate the ANCOVA test by examining residual scores. Specifically, the following test is equivalent to the ANCOVA test:

$$F = \frac{[SSE(RT) - SSE(FW)]/(k - 1)}{SSE(FW)/(N - k - 1)}. \tag{19}$$

The crucial fact is that ANCOVA depends on both b_W and b_T and consequently so must an equivalent analysis of residuals. It is insufficient to attend to only one regression coefficient. As stated previously, we believe that this is one of the points concerning ANCOVA that students most frequently misunderstand. Only by careful specification of the models and the least squares principle does the logic underlying the varying estimates of the regression coefficient parameter become clear.

Graphs

The difference between the approaches can be communicated quite easily and effectively using graphs. The three approaches we have discussed are represented schematically in Figure 1. To simplify comparisons across methods, all plots are in terms of the original X and Y variables. (To do this, the $b_W X_{ij}$ term in models RW and FW, and the $b_T X_{ij}$ term in models RT and FT are effectively shifted from the left hand side to the right hand side of the prediction equations (see panels (b) and (c), respectively). The regression lines, however, accurately represent the model comparisons we have considered. Thus, the deviation of any individual data point from the line equals the error score for that individual in the model represented by the regression line. The ANCOVA analysis is represented in panel (a). Clearly, for these data the full model's within-group regression lines are considerably less steep than the slope for the single regression line corresponding to the restricted model. But each slope is optimal for its model in a least squares sense—the full model's lines fit the data "better" than any other two parallel lines could, and the restricted model's line fits the data for the entire sample "better" than any other single line could.

The situation in panel (b) is quite different. This represents an analysis of variance of the residuals that would be computed using b_W as the single regression coefficient. The full model's fit to the data is the same as that in

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

panel (a), and hence the denominators of the F tests corresponding to panels (a) and (b) would be identical. However, the fit of the restricted model is now much worse. Constraining it to adopt the slope appropriate for the full model means its predictions will scarcely overlap with the data in either group. Thus the difference in goodness of fit between the models is greatly exaggerated, and the evidence for a treatment effect would thereby be made to appear much stronger than it really is.

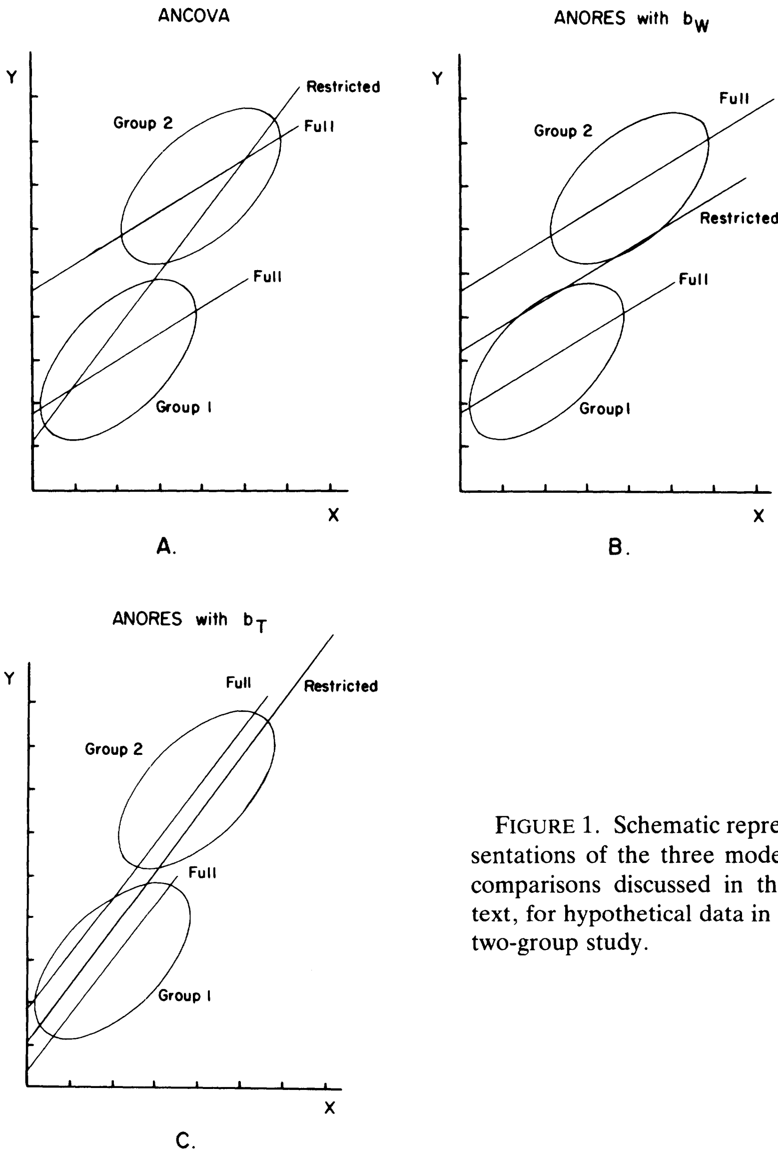

FIGURE 1. Schematic representations of the three model comparisons discussed in the text, for hypothetical data in a two-group study.

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

Panel (c) illustrates how much smaller the treatment effect would seem to be if ANORES with b_T were used. When the steep slope appropriate for the restricted model's single regression line is used in the full model, the parallel lines of the full model are much closer together. In addition, since the inflated lack of fit of this full model would be used as the yardstick for assessing the magnitude of the treatment effect, it will appear even smaller. Although the impact of using the wrong slope for a model is perhaps not as striking as it was for the restricted model in panel (b), the full model in panel (c) fits the data considerably less well than the correct parallel lines shown in panel (a).

Numerical Example

A numerical example will demonstrate the theoretical arguments concerning the relationship between ANCOVA and ANORES. Consider the hypothetical data given in Table I. The error sum of squares for the six models previously outlined are presented in Table II for these data, and Table III presents analysis of variance tables for ANCOVA and the two forms of ANORES. Results here verify that the "F" obtained with ANORES using b_W to form residuals is too large and in fact would be declared significant at $\alpha = .05$. The "F" obtained when b_T is used, on the other hand, is too small. This example illustrates that even when the two groups being compared have

TABLE I
Hypothetical Data to Illustrate ANCOVA—ANORES Relationship

Group 1		Group 2	
X	Y	X	Y
100	100	100	105
95	98	90	109
105	102	95	104
110	106	105	112
105	101	95	100
90	103	100	104

TABLE II
ANCOVA and ANORES Models and Associated Error Sums of Squares

Model	SSE
I. $Y_{ij} = \mu + \alpha_j + \beta X_{ij} + \epsilon_{ij}$	110.3
II. $Y_{ij} = \mu + \beta X_{ij} + \epsilon_{ij}$	170.7
III. $Y_{ij} - b_W X_{ij} = \mu + \alpha_j + \epsilon_{ij}$	110.3
IV. $Y_{ij} - b_W X_{ij} = \mu + \epsilon_{ij}$	175.7
V. $Y_{ij} - b_T X_{ij} = \mu + \alpha_j + \epsilon_{ij}$	114.9
VI. $Y_{ij} - b_T X_{ij} = \mu + \epsilon_{ij}$	170.7

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

TABLE III
Analysis of Variance Tables

	Source	*SS*	*df*	*MS*	*F*	*p*
ANCOVA	Between	60.4	1	60.4	4.9	.054
	Within	110.3	9	12.3		
	Total	170.7	10			
ANORES	Between	65.3	1	65.3	5.3	.046
with b_W [a]	Within	110.3	9[b]	12.3		
	Total	175.7	10			
ANORES	Between	55.8	1	55.8	4.4	.066
with b_T [c]	Within	114.9	9[b]	12.8		
	Total	170.7	10			

[a] $b_W = 0.20$ for these data
[b] $df_W = 9$ because of the estimation of β in forming the residual
[c] $b_T = 0.09$ for these data

similar distributions on the covariate, the use of b_W alone can lead to a conclusion of statistical significance when the correct *ANCOVA* test is in fact nonsignificant. With different data, the use of b_T alone might result in a failure to recognize an appropriate significant result. For example, by simply revising the data from the first example by subtracting 4 from each X score in group 2, the results of the analysis would be as shown in Table IV. Note that while the appropriate ANCOVA is now significant with $p = .048$, and the ANORES using b_W is significant with $p = .024$, an ANORES using b_T would not approach significance, $p > .10$. In fact, the F for ANORES $- b_T$ is less than half that for ANORES $- b_W$.

In addition, it is possible to duplicate the ANOVA results by employing both b_T and b_W to form residual scores and then applying Equation 19. The reason this procedure works can be seen in Tables III and IV. ANORES with b_W yields the correct adjusted SS_W, and ANORES with b_T yields the correct adjusted SS_T. Thus, Equation 19 provides a ratio of the Adjusted Mean Square Between divided by the Adjusted Mean Square Within, as is desired.

Discussion

Some authors of experimental design texts (see, e.g., Kirk, 1982, p. 727) explain why b_W should not be used to adjust both the numerator and the denominator of the F ratio by noting that this would violate the independence condition necessary for an F ratio. Although the numerator and the denominator must indeed be independent, the current argument shows that this use of b_W to obtain an adjusted sum of squares between groups also fails to provide even a numerator quantity that is distributed as a chi-square. Cochran (1957) has provided an alternative explanation of how dependence causes the use of

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

TABLE IV
Analysis of Variance Tables for Revised Example

	Source	*SS*	*df*	*MS*	*F*	*p*
ANCOVA	Between	64.3	1	64.3	5.2	.048
	Within	110.3	9	12.3		
	Total	174.7	10			
ANORES	Between	89.7	1	89.7	7.3	.024
with b_W[a]	Within	110.3	9[b]	12.3		
	Total	200.0	10			
ANORES	Between	46.1	1	46.1	3.2	.106
with b_T[c]	Within	128.5	9[b]	14.3		
	Total	174.7	10			

[a] $b_W = 0.20$ for these data
[b] $df_W = 9$ because of the estimation of β in forming the residual
[c] $b_T = -0.01$ for these data

b_W to provide an inappropriate numerator term.

In addition to explaining why the use of b_W alone results in an inflated and inappropriately distributed between-group sum of squares, the current approach also makes clear the different roles of b_W and b_T. This distinction between b_W and b_T has been presented in ways that are clearly misleading to students. Cohen and Cohen (1983), for example, err as we have shown by recommending use of b_W alone, saying that to use b_T would result in "removing from *Y*, in part, exactly what we mean to study" (p. 318). Although use of b_T does result in a lower adjusted SS_B than if b_W alone were used, use of b_T in the restricted model should not be viewed as removing part of "what we mean to study." Rather, it gives the restricted model a fair chance in that it allows the estimate of the regression parameter to be an optimal, least squares estimate, as b_W is in the full model.

In sum, although the concept of a residual score can be a useful pedagogical tool for explaining the logic of ANCOVA, it has typically not been utilized accurately. A correct SS_B can be calculated by using residuals, but only by considering both b_T and b_W, and hence at least implicitly considering *two* sets of residuals. In terms of the residual score models,

$$\text{Adjusted } SS_B = SSE(RT) - SSE(FW). \tag{20}$$

Further, approaches that present ANORES as an alternate data analysis strategy that could be used in situations where ANCOVA is legitimate are shown to be wrong, because under the ANCOVA assumptions the test statistic in ANORES does not follow an *F* distribution. Instead of relying on ANORES to explain ANCOVA, an approach utilizing model comparisons and least squares clarifies the ANCOVA procedure and its underlying rationale.

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms

Acknowledgments

Much of the work by Harold Delaney was done while he was a Post Doctoral Fellow at the University of Edinburgh, Scotland. The support of the Department of Psychology at the University of Edinburgh is gratefully acknowledged. Jerry Manheimer is now with Lockheed Corp.

References

Burstein, L., Linn, R. L., & Capell, F. J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics, 3,* 347–383.

Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics, 13,* 261–281.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Corder-Bolz, C. R. (1978). The evaluation of change: New evidence. *Educational and Psychological Measurement, 38,* 959–976.

Dixon, W. J., & Massey, F. J. (1969). *Introduction to statistical analysis.* New York: McGraw-Hill.

Elashoff, J. D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal, 6,* 383–401.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the analysis of variance and covariance. *Review of Educational Research, 42,* 237–288.

Kirk, R. E. (1982). *Experimental design.* Monterey, CA: Brooks/Cole.

Lindquist, E. F. (1953). *Design and analysis of experiments in psychology and education.* Boston: Houghton-Mifflin.

Marascuilo, L. A. (1971). *Statistical methods for behavioral science research.* New York: McGraw-Hill.

Namboodiri, N. K., Carter, L. F., & Blalock, H. M., Jr. (1975). *Applied multivariate analysis and experimental designs.* New York: McGraw-Hill.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research.* New York: Holt, Rinehart, and Winston.

Searle, S. R. (1971). *Linear models.* New York: Wiley.

Snedecor, G. W., & Cochran, W. G. (1967). *Statistical methods.* Ames: Iowa State University.

Werts, C. E., & Linn, R. L. (1970). A general linear model for studying growth. *Psychological Bulletin, 73,* 17–22.

Authors

SCOTT E. MAXWELL, Associate Professor, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556. *Specializations:* Behavioral statistics, quantitative psychology.

HAROLD D. DELANEY, Associate Professor, Department of Psychology, University of New Mexico, Albuquerque, NM 87131. *Specializations:* Quantitative psychology, human learning and memory.

JERRY M. MANHEIMER, Human Factors Engineer—Senior, Lockheed, Austin Division, 2124 E. St., Elmo Rd., Austin, TX 78744. *Specializations:* Cognitive models, multivariate techniques.

This content downloaded from 193.63.252.9 on Mon, 17 Sep 2018 15:22:41 UTC
All use subject to https://about.jstor.org/terms