THE ANALYSIS OF COVARIANCE: STANDARD AND NON-STANDARD PROCEDURES

Ronald J. Visco/Florida State University

Frequently, the researcher wishes to compare two or more treatment groups on some experimental or dependent variable. However, in making the comparison, the researcher may wish to utilize information which is available on some set of independent variables (covariates), correlated with the dependent variable. The analysis of covariance is a generic term for a family of procedures applicable in this context.

Basically, the analysis of covariance, through use of regression techniques, permits a post-hoc statistical control of the set of independent variables, "removing their influence from the comparison of groups on the main experimental variable" (Hays, 1973, p. 655). This statistical control is applied to independent variables which are "inadvisable, inconvenient, or impossible to control directly in the design of the experiment" (Johnson and Jackson, 1953, p. 410). Thus, the analysis of covariance is regarded as a combination of regression analysis and the analysis of variance (Hays, 1973, p. 654; Winer, 1971, p. 760).

Although analysis of covariance procedures are frequently applied, they are commonly misunderstood. The discussion here is centered on a geometric comceptualization of the underlying models, with an outline of assumptions, procedures, and appropriate statistics, as well as reference to available computer programs.

Attention is limited to linear models with two treatment groups (1 and 2), one dependent variable (Y) and one independent variable (X). Extensions of these techniques may be found in the references.

Standard Analysis of Covariance

The standard analysis of covariance procedure is the method most commonly presented, discussed, and applied.

Assumptions

Each X is a constant with no bias and no variability (Schluck, Note 1, Note 2). Within each treatment, each X defines a population of Ys. The mean for each Y population lies on a regression line representing the particular treatment. Furthermore, the Ys in each population are independently and normally distributed with a common variance among populations:

$$Y_{kj} \overset{d}{=} \text{NID}\ [M_{kj},\ {}^{2}]\ ;\ K = 1,2;\ j = 1,2,\ \ldots,\ n_k \qquad (1)$$

where K indicates the treatment and n_k is the sample size for treatment group K.

Also, the regression lines are parallel; that is, the regression coefficients are equal across treatments. The regression equations are given by:

$$\mu_{kj} = \alpha_k + \beta X_{kj}\ ;\ K = 1,2,;\ J = 1,2,\ \ldots,\ n_k. \qquad (2)$$

This assumption implies that there is no interaction between the treatment and the independent variable.

In addition, the value of X is unaffected by the treatment (Myers, 1966, chap. 3; Winer, 1971, chap. 10).

Procedure

Tests of assumptions. The researcher may wish to test the given assumptions. If the assumptions are unreasonable in terms of the data, another model should be adopted.

Some necessary statistics are defined by

$$\bar{X}_k = \frac{\sum_{i=1}^{n_k} X_{ki}}{n_k},\ K = 1,2,;\ \text{and} \qquad (3)$$

$$\bar{Y}_k = \frac{\sum_{i=1}^{n_k} Y_{ki}}{n_k},\ K = 1,2, \qquad (4)$$

the usual sample means; also let

$$SS_{x:k} = \sum_{i=1}^{n_k} (X_{ki} - \bar{X}_k)^2,\ K = 1,2, \qquad (5)$$

$$SS_{y:k} = \sum_{i=1}^{n_k} (Y_{ki} - \bar{Y}_{ki})^2,\ K = 1,2, \qquad (6)$$

the usual sums of squares. Finally, define

$$SP_{xy:k} = \sum_{i=1}^{n_k} (X_{ki} - \overline{X}_k)(Y_{ki} - \overline{Y}_k), K = 1,2, \quad (7)$$

the usual sums of cross products.

Let b_k be the estimate of the regression coefficient calculated independently for group K. Then.

$$b_k = \frac{SP_{xy:k}}{SS_{x:k}}, K = 1,2. \quad (8)$$

To test the assumption of equal variances, compute the ratio

$$F = \frac{S_1^2}{S_2^2}, \quad (9)$$

where

$$S_k^2 = \frac{SS_{y:k} - b_k SP_{xy:k}}{n_k - 2}$$

$$= \frac{SS_{y:k} - b_k^2 SS_{x:k}}{n_k - 2}; K = 1,2, \quad (10)$$

(Crow, Davis, and Maxfield, 1960, pp. 74, 161; Johnson and Jackson, 1953, pp. 417-8; Schluck, Note 1). The F-test suggested by Equation 9 is a two-tailed test with df = $(1, n_1, + n_2 - 4)$.

Often, the assumption of equal variances is not formally tested: the analysis of covariance is not sensitive to moderate departures from this assumption, although large departures may invalidate the results.

The assumption of linearity is frequently made without careful inspection of the data. In some instances, quadratic or even cubic regression equations may be more suitable. For example, the researcher may find that the model

$$\mu_{kj} = \alpha_k + \beta_1 X_{kj} + \beta_2 X_{kj}^2 ; K = 1,2,; j = 1,2, ..., n_k \quad (11)$$

is more appropriate for the data. Formal tests, if desired, are available (e.g., see Winer, 1971).

The most critical assumption of the standard analysis of covariance model is the equality of regression coefficients across treatments. Hays (1973, p. 657) calls this assumption "the biggest hitch" in applying the analysis. Although some researchers fail to test for equal coefficients before conducting a standard analysis of covariance, such an oversight may lead to serious errors of interpretation.

To test the hypothesis that $\beta_1 = \beta_2 = \beta$ (a common regression coefficient assumed by Equation 2), one may form the ratio

$$F = \frac{(b_1 - b_2)^2}{S^2_{RES:I} \left(\frac{1}{SS_{x:1}} + \frac{1}{SS_{x:2}} \right)},$$

where

$$S^2_{RES:I} = \frac{SS_{y:1} + SS_{y:2} - b_1 SP_{xy:1} - b_2 SP_{xy:2}}{n_1 + n_2 - 4} \quad (13)$$

(Crow, Davis, and Maxfield, 1960, p. 161; Hays, 1973, pp. 657-8; Johnson and Jackson, 1959, pp. 418-9; Myers, 1966, p. 310; Walker and Lev, 1953, pp. 390-3; Schluck, Note 2). The obtained F-ratio is compared to the tabled F with df = $(1, n_1 + n_2 - 4)$.

If the hypothesis of a common regression coefficient is rejected, the application of a standard analysis of covariance procedure is extremely questionable: the researcher should look for another model. It is not implied that a failure to reject this hypothesis justifies the employment of the standard model.

Common regression coefficient. Since the standard analysis of covariance model assumes that the regression coefficients are equal across treatments, differences in the coefficient estimates obtained independently for each group are attributed to sampling error.

Therefore, the data are pooled to obtain an estimate of the common regression coefficient. The pooled estimate is

$$b_c = \frac{b_1 SS_{x:1} + b_2 SS_{x:2}}{SS_{x:1} + SS_{x:2}}$$

$$= \frac{SP_{xy:1} + SP_{xy:2}}{SS_{x:1} + SS_{x:2}} \quad (14)$$

(Johnson and Jackson, 1959, p. 418; Walker and Lev, 1953, p. 390; Schluck, Note 2).

Note that any test of the assumption of equal regression coefficients across treatments must be conducted before employing the standard analysis of covariance procedure: the standard model imposes a pooled coefficient estimate without testing that assumption.

Figure 1 illustrates a hypothetical situation. Regression equations calculated independently for each group are represented by dashed lines. These

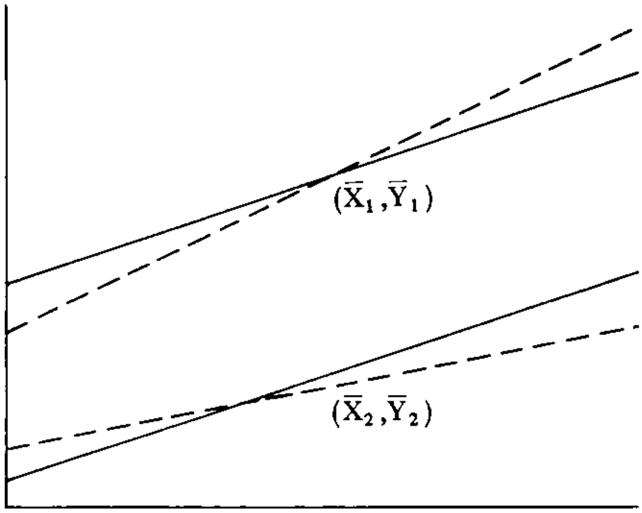

Figure 1. A standard analysis of covariance model. Dashed lines are regression lines calculated independently for each treatment group; solid lines are imposed by the model using a pooled estimate of an assumed common regression coefficient. ($\overline{X}_1$ and $\overline{X}_2$ are sample X means; $\overline{Y}_1$ and $\overline{Y}_2$ are sample Y means.)

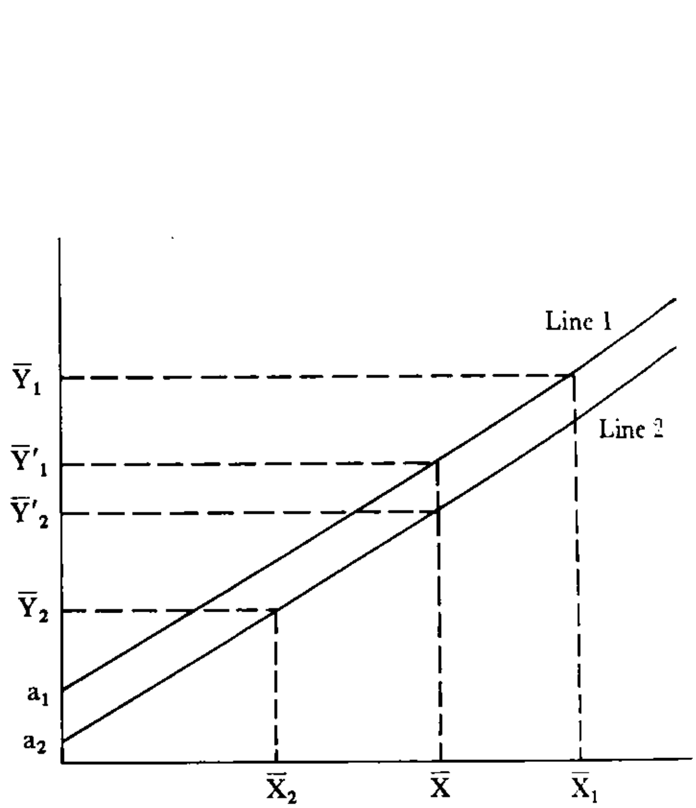

Figure 2. A standard analysis of covariance model. (With $\overline{X}_1$, $\overline{X}_2$ the sample X means; $\overline{Y}_1$, $\overline{Y}_2$ the sample Y means; $\overline{Y}'_1$, $\overline{Y}'_2$, the adjusted Y means; a_1, a_2 the intercepts of lines 1 and 2, respectively. $\overline{X}$ is the overall X mean.)

lines are not, in general, parallel. However, the standard analysis of covariance model assumes a common regression coefficient for all treatments and therefore imposes parallel lines on the data. In Figure 1, the imposed regression lines are solid.

For each group, both regression lines pass through the point $(\overline{X}_k, \overline{Y}_k)$ of sample means; this will always be the case. Roughly speaking, the pooled regression coefficient imposed on the data has the effect of rotating each independently estimated regression line about the point $(\overline{X}_k, \overline{Y}_k)$ until the lines are parallel.

Test of equal intercepts. In Figure 2, a pooled coefficient estimate has already been imposed on the data. The regression lines yield predicted Y means for given X values. The standard analysis of covariance procedure compares the predicted Y means for a particular X value, usually $X = \overline{X}$. These predicted Y means are frequently called adjusted Y means.

In Figure 2, the difference between the adjusted Y means is equal to the vertical distance between the regression lines at $X = \overline{X}$. However, since the lines are parallel, this distance remains constant for all values of X. At $X = 0$, the vertical distance between the lines is the difference $(a_1 - a_2)$ between the intercept estimates, where

$$a_k = \overline{Y}_k - b_c\overline{X}_k. \tag{15}$$

Therefore, in testing the hypothesis of equal treatment means for a given X value ($\mu_{1j} = \mu_{2j}$ at X_j), the X value chosen is of no consequence; either the difference between adjusted Y means or the difference between intercept estimates may be used for convenience. The ratio

$$F = \frac{(a_1 - a_2)^2}{S^2_{RES:II}\left[\frac{n_1 + n_2}{n_1 n_2} + \frac{(\overline{X}_1 - \overline{X}_2)^2}{SS_{x:1} + SS_{x:2}}\right]}, \tag{16}$$

where

$$S_{RES:II} = \frac{SS_{y:1} + SS_{y:2} - b_c(SP_{xy:1} + SP_{xy:2})}{n_1 + n_2 - 3}, \tag{17}$$

is compared to the critical F with $df = (1, n_1 + n_2 - 3)$.

If the F-ratio obtained in Equation 16 is significant, the researcher might conclude that, on the average, there is a difference between treatments: that one treatment is "more effective" than the other.

Non-Standard Analysis of Covariance

Removing the simplifying assumption that the treatment regression coefficients are equal allows the application of an alternative procedure: the non-standard analysis of covariance (Schluck, Note 2).

Here, the researcher may ask if there is a significant difference between regression lines (treatments) at a particular X value of interest. In general, the answer will depend on what X value is considered.

Clearly, this non-standard analysis may be used in situations where the standard analysis does not apply (Walker and Lev, 1953, pp. 398-9).

Assumptions

As with the standard model, each X is a constant unaffected by the treatment, and Equation 1 holds.

The means of the Y populations lie on a line defined by

$$\mu_{kj} = \alpha_k + \beta_k X_{kj}; K = 1,2,; j = 1,2, ..., n_k. \quad (18)$$

In general, $\beta_1 \neq \beta_2$.

Procedure

Tests of assumptions. The researcher may test the assumption of equal variances with the F-ratio in Equation 9. Again, the analysis is insensitive to violations of this assumption.

The researcher should examine the data for linearity, and may conduct a formal test.

Differences between predicted means. A regression line is now determined independently for each group. Since the lines are no longer forced by assumption to be parallel, the vertical distance between the lines will depend, in general, on the value of X. Therefore, the difference in predicted means is a function of X and is given by

$$D = (a_1 + b_1 X) - (a_2 + b_2 X)$$

$$= (a_1 - a_2) + (b_1 - b_2) X. \quad (19)$$

The difference D may assume negative values. The absolute value of the difference, $|D|$, is the distance between the lines.

Figure 3 illustrates a hypothetical situation. The regression equations are calculated independently for each group, and the lines are not parallel. The distance between the lines at $X = X'$ is less than the distance at $X = X''$.

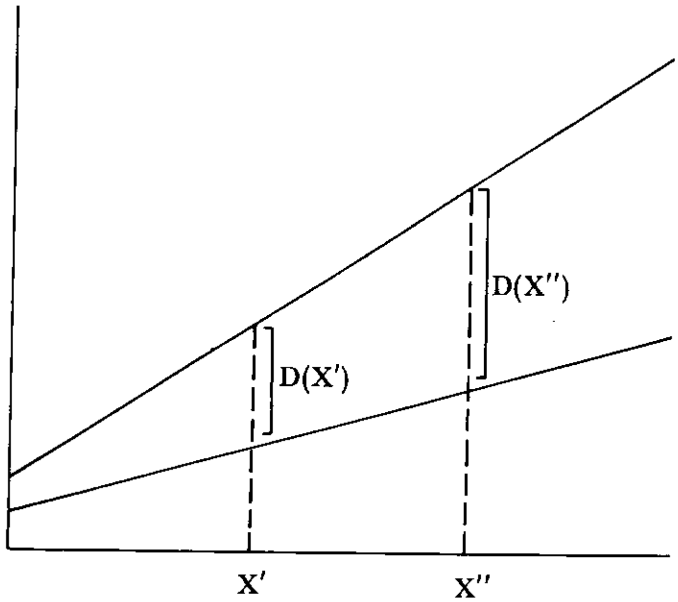

Figure 3. A non-standard analysis of covariance model. The difference D between regression lines is a function of X. Here, $D(X') < D(X'')$.

Test of equal treatment means. The researcher may now test the hypothesis of equal treatment means for a given X value of interest ($\mu_{1j} = \mu_{2j}$ at X_j) by testing the significance of D at that X value.

First, calculate S_D^2, the variance of D. The value of S_D^2 at $X = X_0$ is given by

$$S_D^2 = S^2_{RES:I} \left[\frac{n_1 + n_2}{n_1 n_2} + \frac{(X_0 - \bar{X}_1)^2}{SS_{x:1}} + \frac{(X_0 - \bar{X}_2)^2}{SS_{x:2}} \right] \quad (20)$$

and the $S^2_{RES:I}$ is given in Equation 13. The ratio D/S_D has the t-distribution for each fixed value of X (Walker and Lev, 1953, p. 400), and D^2/S_D^2 has the F-distribution (Schluck, Note 2).

To test the significance of D for a given X value, one may calculate the ratio

$$F = \frac{D^2}{S_D{}^2} \quad (21)$$

and compare it to the critical F with $df = (a, n_1 + n_2 - 4)$.

Regions of significance. The researcher may not have particular X values of interest for which he or she wishes to examine the difference between treatments. Suppose the researcher calculates the regression equations for the two treatments and obtains the results illustrated in Figure 4. The researcher may reasonably want to know the values of X for which Treatment 1 is (significantly) more effective than Treatment 2, as well as the X values for which Treatment 2 is more effective. Methods for this approach were largely developed by Johnson and Neyman (see Johnson and Fay, 1950; Abelson, 1953).

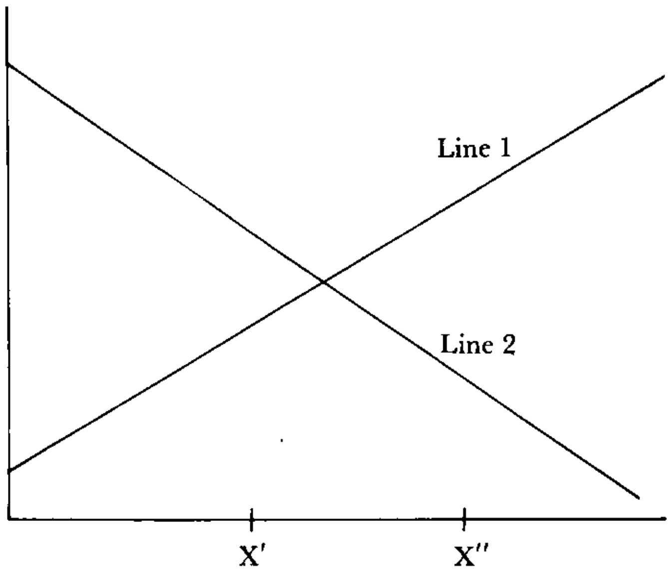

Figure 4. Regions of significance. Values of X between X′ and X″ constitute a region of nonsignificance for the difference between regression lines 1 and 2.

From Equation 21, it is seen that the set of X values which yield a significant F are those which satisfy the inequality

$$\frac{D^2}{S_D^2} > F_{critical}; \qquad (22)$$

That is, the set of X values which make

$$D^2 - F_{1-\alpha;1,n_1+n_2-4} S_D^2 > 0 \qquad (23)$$

for the given level of significance α. The set of X values satisfying Equations 22 and 23 are said to constitute the "region of significance." To determine the boundary values for this region, solve the equality

$$D^2 - F_{1-\alpha;1,n_1+n_2-4} S_D^2 = 0 \qquad (24)$$

which is a quadratic equation in X. The solutions are presented by Walker and Lev (1953, p. 401) and adapted by Kerlinger (1973, p. 256), and are given by a modified quadratic formula:

$$X = \frac{-B \pm \sqrt{B^2 - AC}}{A}, \qquad (25)$$

where

$$A = -F_{1-\alpha;1,n_1+n_2-4} S^2_{RES:I} \left[\frac{1}{SS_{x:1}} + \frac{1}{SS_{x:2}}\right] + (b_1 - b_2)^2; \qquad (26)$$

$$B = F_{1-\alpha;1,n_1+n_2-4} S^2_{RES:I} \left[\frac{\overline{X}_1}{SS_{x:1}} + \frac{\overline{X}_2}{SS_{x:2}}\right] + (a_1 - a_2)(b_1 - b_2); \qquad (27)$$

$$C = -F_{1-\alpha;1,n_1+n_2-4} S^2_{RES:I} \left[\frac{n_1+n_2}{n_1 n_2} + \frac{\overline{X}_1^2}{SS_{x:1}} + \frac{\overline{X}_2^2}{SS_{x:2}}\right] + (a_1 - a_2)^2. \qquad (28)$$

If $B^2 - AC < 0$, there is no real solution to Equation 25, and no boundary values exist. In that event, there is no region of significance (Walker and Lev, 1953, p. 402) unless the coefficient estimates are equal, when the entire X line may be either a region of significance (Borich, Godbout, and Wunderlich, Note 3) or one of nonsignificance. If $B^2 - AC = 0$, there is a single X value constituting the region of significance, a highly unlikely occurrence.

When Equation 25 yields two X values, say X′ and X″, each value separates a region of significance from a region of nonsignificance.

In Figure 4, hypothetical values of X′ and X″ are indicated. The set of all X values between X′ and X″ constitutes a region of nonsignificance. Therefore, the set of X values greater than or equal to X″ constitutes a region of significance where, loosely stated, Treatment 1 is (significantly) more effective than Treatment 2. For X values less than or equal to X′, Treatment 2 is more effective.

Figure 5 illustrates geometrically how the region of significance is determined. The solid line represents the difference function D for the two lines in Figure 4. Therefore, the X value at which the lines intersect in Figure 4 is the value at which D = 0 (no difference) in Figure 5. The curved

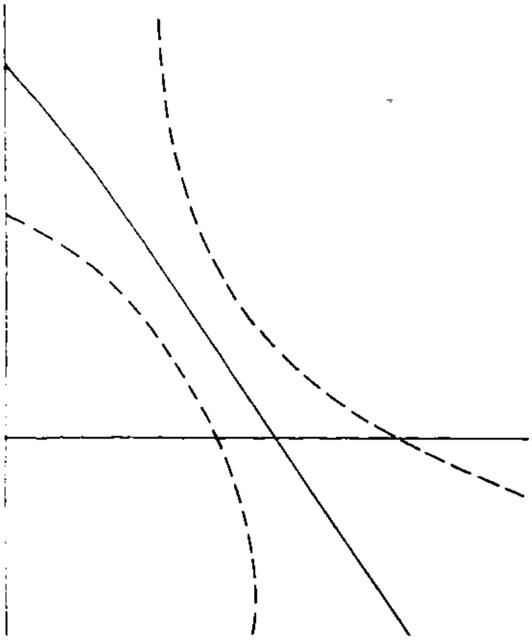

Figure 5. Confidence (dashed) lines about a difference (solid line) between two regression lines (not shown).

(dashed) lines in Figure 5 indicate a confidence interval about the D line. The two points at which the X-axis intersects the curved lines are the X′ and X″ values indicated in Figure 4. Roughly stated, for X values such that $X \leq X'$ or $X \geq X''$, one is confident that the difference between treatment means is not equal to zero.

Unfortunately, when Equation 25 has two solutions, the values of X between these solutions do not always constitute the region of nonsignificance. Sometimes they may constitute the region of significance. To determine which situation holds, one calculates the value of X for which $D(X) = 0$: this is, the X value for which the regression lines intersect. This X value will always lie in a region of nonsignificance. For the situation illustrated in Figures 4 and 5, this X value lies in the region between X′ and X″; therefore, this region is one of nonsignificance.

Computer Programs

Computer packages are available for performing the analyses. The standard analysis of covariance may easily be done using, for example, one of the Biomedical (BMD) Computer Programs: BMD04V, "The analysis of covariance." Furthermore, BMD05V (General Linear Hypothesis) can assist in performing a non-standard analysis but the data must be entered in a peculiar manner.

A more versatile package is MULGEN (Multivariate General Linear Model), although it is not widely available. MULGEN can routinely perform the standard analysis of covariance. In addition, when used in conjunction with the program COVARY, MULGEN performs non-standard analysis for any number of X values of interest. Analysis with non-linear regression lines may also be done.

If the researcher wishes to compute regions of significance, programs for this purpose are reported in the literature. For example, Carroll and Wilson (1970) and Borich (1971) report such programs. Borich and Wunderlich (1973) extended the computer analysis to the case of two covariates. A manual by Borich, Godbout, and Wunderlich (Note 3) discusses a set of programs which compute regions of significance with either one or two covariates, and also allow for curvilinear regression lines.

Summary and Conclusions

The researcher must choose the model with care. The data cannot be expected to conform to an unrealistic model. Perhaps theory, as well as the data, will guide the researcher.

Certainly, the researcher will wish to consider the assumptions of any model which may potentially be adopted. In some cases, the researcher may conduct formal tests of the assumptions; however, when employing multiple tests (or confidence intervals), care should be taken to adjust the level of significance for a single test (see Miller, 1966).

It is seldom the case in statistical analysis that clearly defined rules can be established to guide a researcher. For example, suppose the researcher tests an assumption of equal variances, and obtains an F-ratio which just exceeds the tabled F at the chosen level of significance. In deciding whether to proceed with an analysis of covariance, the researcher should consider the insensitivity of the analysis to that assumption. Of course, the obtained F might have failed to exceed the tabled F had another level of significance been chosen.

A critical difference between the standard and non-standard analysis of covariance models arises from the latter model's assumption of parallel regression lines. Some researchers will choose to test that assumption for their data. If the hypothesis of parallel lines is rejected, the researcher might be likely to employ a non-standard procedure. However, researchers often view a failure to reject the hypothesis as a demonstration that the standard model is appropriate for their data. The researcher is cautioned that such an attitude can frequently lead to oversimplifications and errors of interpretation.

There is no substitute for good judgment exercised by the researcher. Experienced researchers will choose their models and analyze their data with insight and care.

Reference Notes

1. Schluck, G. Personal communication, March, 1977.
2. Schluck, G. Seminar presented at Florida State University, Spring, 1975.
3. Borich, G. D., Godbout, R. C., and Wunderlich, K. W. *The analysis of aptitude-treatment interactions: computer programs and calculations* (Contract No. NIE-C-74-0088). Austin: The University of Texas at Austin, Research and Development Center for Teacher Education, March, 1976.

References

Abelson, R. P. A note on the Neyman-Johnson Technique. *Psychometrika*, 1953, *18*(3), 213-218.

Borich, G. D. Interactions among group regressions: Testing homogeneity of group regressions and plotting regions of significance. *Educational and Psychological Measurement*, 1971, *31*, 251-253.

Borich, G. D., and Wunderlich, K. W. Johnson-Neyman revisited: Determining interactions among group regressions and plotting regions of significance in the case of two groups, two predictors, and one criterion. *Educational and Psychological Measurement*, 1973, *33*, 155-159.

Carroll, John B., and Wilson, G. F. An interactive-computer program for the Johnson-Neyman Technique in the case of two groups, two predictor variables, and one criterion variable. *Educational and Psychological Measurement*, 1970, *30*, 121-132.

Crow, E. L., Davis, F. A., and Maxfield, M. W. *Statistics Manual.* New York: Dover Publications, Inc., 1960.

Hays, W. L. *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart & Winston, Inc., 1973.

Johnson, P. O., and Fay, L. C. The Johnson-Neyman Technique, its theory and application. *Psychometrika*, 1950, *15*, 349-367.

Johnson, P. O., and Jackson, R. W. B. *Modern statistical methods: Descriptive and inductive.* Chicago: Rand McNally & Company, 1959.

Kerlinger, F. N., and Pedhayur, E. J. *Multiple regression in behavioral research.* New York: Holt, Rinehart & Winston, 1973.

Miller, R. J., Jr. *Simultaneous statistical inference.* McGraw-Hill Book Co., New York, 1966.

Myers, J. L. *Fundamentals of experimental design.* Boston: Allyn & Bacon, 1966.

Walker, Helen M., and Lev, J. *Statistical inference.* New York: Henry Holt and Company, 1953.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: 1962.