Psychological Bulletin
1980, Vol. 88, No. 2, 307-321

Comparing Nonparallel Regression Lines

David Rogosa
Department of Education, University of Chicago

Statistical comparisons of experimental groups frequently are based on the within-group regressions of an outcome variable on a concomitant variable. I present a comprehensive strategy for the statistical comparison of within-group regressions that is suitable for both parallel and nonparallel regression lines. New results are obtained and new interpretations are formulated for standard statistical procedures such as analysis of covariance. New procedures for comparing regressions are proposed and illustrated.

The setting for this investigation is an experiment in which two groups are given alternative treatments. The groups are formed by random assignment of individual cases to the groups. One of these groups may be considered a control group, as in a study to evaluate the effectiveness of a curricular innovation relative to the effectiveness of a standard curriculum. The purpose of the experiment is to assess the treatment effect—the differential effectiveness of the two treatments.

A number of initial characteristics (usually called covariates or predictor variables) are recorded for each case before the initiation of the treatment; outcome variables are measured at the end of the treatment period. I limit consideration to one outcome variable (Y) and one predictor variable (X). The raw data from this experiment are measures of X and Y that are obtained from the members of the two experimental groups.

The raw data are summarized through estimation of the within-group regression lines, the average outcome for a given initial characteristic value within each group. The summary of the raw data consists of (a) the separate sample within-group regression lines and (b) estimates of the sampling variances and covariances associated with these sample regressions.

Requests for reprints should be sent to David Rogosa, who is now at the School of Education, Stanford University, Stanford, California 94305.

This article presents and evaluates statistical procedures for comparing the within-group regression lines. I view the treatment effect as a function of X and define the treatment effect as the difference between the population regression lines. The usual dichotomy between parallel and nonparallel regressions is shunned. The difference between the sample regression lines is used to estimate this treatment effect, and this estimate will depend on X (to some degree) whenever the sample within-group regressions are not parallel. Although this approach seems natural for problems of comparing regressions, the development and exposition of traditional statistical methods has proceeded along other lines.

Two types of assessments of the treatment effect are sought. First, an overall treatment effect, in which dependence on X of the treatment effect is ignored, can be estimated by evaluating the difference between the sample regression lines at a prespecified value of X. These procedures are called pick-a-point procedures. Second, an assessment of the difference between the regressions over the entire range of X can be used to evaluate the treatment effect as a function of X. The Johnson-Neyman technique as extended by Potthoff (1964) is one procedure in which the regressions are compared over the range of X. Nonsimultaneous inference procedures are associated with the first type of assessment, and simultaneous inference procedures are associated with the second type of assessment.

Copyright 1980 by the American Psychological Association, Inc. 0033-2909/80/8802-0307$00.75

First, the regression model and the basic notation are presented. Then, analysis of covariance (ANCOVA) is examined, and it is found that the ANCOVA estimate of the treatment effect is identical to an evaluation of the difference between the sample regressions at a particular value of X in the middle of the data. This result is used to form an interpretation of the ANCOVA estimate that does not depend on the assumption of parallel population regressions. Also, a modified version of ANCOVA is proposed. Next, an alternative overall assessment of the treatment effect is formed from a measure of the average distance between the within-group regressions, and a statistical procedure based on this measure is compared with ANCOVA. Then, assessment of the treatment effect as a function of X through the use of the Johnson-Neyman technique is described, and comparisons with pick-a-point procedures are developed. Also, analytic results developed in this investigation are used to expose serious logical and technical flaws in previous simulation studies on the robustness of ANCOVA to unequal regression slopes. Finally, the methods and computational procedures discussed in this article are illustrated with a small data set.

Regression Model and Parameter Estimation

The model for the population within-group regression lines allows both the slopes and intercepts to differ between the two groups. The within-group regression model for Groups A and B is

$$Y_j=\alpha_A+\beta_A X_j+\epsilon_j, \text{ for } j=1, \ldots, n_A;$$
$$Y_j=\alpha_B+\beta_B X_j+\epsilon_j, \text{ for } j=n_A+1, \ldots, N. \quad (1)$$

Group A contains n_A cases, and Group B contains n_B cases; $n_A + n_B = N$. The difference between the population regression lines at a specific value of X is $\Delta(X) = \alpha_A - \alpha_B + (\beta_A - \beta_B)X$.

This regression model can be rewritten using a dummy variable T, defined such that

$$T_j = 1, \text{ for } j = 1, \ldots, n_A;$$
$$T_j = 0, \text{ for } j = n_A + 1, \ldots, N.$$

Then an equivalent expression for Equation 1 is

$$Y_j = \beta_1 + \beta_2 T_j + \beta_3 X_j + \beta_4 T_j X_j + \epsilon_j \quad \text{for } j = 1, \ldots, N. \quad (2)$$

The difference between the within-group population slopes is $\beta_4 = \beta_A - \beta_B$, and the difference between the population regression lines evaluated at $X = 0$ is $\beta_2 = \alpha_A - \alpha_B$. Thus $\Delta(X) = \beta_2 + \beta_4 X$.

The statistical methods considered in this article condition on the values of X that are observed, as is conventional in regression analysis. The data can be thought of as consisting of N X,Y pairs that are a random sample from the population bivariate distribution of X and Y. The values of X are not chosen or fixed in advance. However, inferences from these data are restricted to subpopulations having the same values or configuration of X because inferences from the linear model are conditional on the observed values of X. All distributional statements in this article are conditional on X unless otherwise specified. (See Kendall & Stuart, 1967, and Sampson, 1974, for further discussion of conditional inference in regression models.)

The distributional assumptions used in the regression analyses are that in both Equation 1 and Equation 2 the ϵ_j are independently and identically distributed, with a normal distribution having mean 0 and within-group variance σ^2: $\epsilon \sim N(0, \sigma^2)$. Another way to express the distributional assumptions is that the conditional distribution of Y, given X and T, is

$$N(\beta_1 + \beta_2 T + \beta_3 X + \beta_4 T X, \sigma^2).$$

The ordinary least squares estimates of the parameters in Equation 1 and Equation 2 are denoted by $\hat{\alpha}$ and $\hat{\beta}$. The sample within-group regression lines are $Y = \hat{\alpha}_A + \hat{\beta}_A X$ and $Y = \hat{\alpha}_B + \hat{\beta}_B X$. The difference between these sample regression lines at any value of X is written as $D(X)$:

$$D(X) = \hat{\alpha}_A - \hat{\alpha}_B + (\hat{\beta}_A - \hat{\beta}_B)X = \hat{\beta}_2 + \hat{\beta}_4 X.$$

In the statistical comparison of within-group regression lines, $D(X)$ is the key summary of the data. All statistical inference procedures investigated in subsequent sections are based on $D(X)$.

Before investigating $D(X)$ further, I introduce some additional estimates and notation. An unbiased estimate of the common residual variance σ^2, denoted by s^2, is proportional to the sum of the residual sums of squares for the two within-group regressions:

$$s^2 = \frac{1}{N-4}\left[\sum_{j=1}^{n_A} (Y_j - \hat{\alpha}_A - \hat{\beta}_A X_j)^2 + \sum_{j=n_A+1}^{N} (Y_j - \hat{\alpha}_B - \hat{\beta}_B X_j)^2\right].$$

The distribution of $(N-4)s^2/\sigma^2$ is that of a central chi-square variable with $N-4$ degrees of freedom, written as $\chi^2(N-4)$.

The covariances of the estimates of the regression parameters in Equation 2 are written

$$\sigma_{ii'} = \text{cov}(\hat{\beta}_i, \hat{\beta}_{i'}), \quad i, i' = 1, \ldots, 4.$$

Estimates of these covariances are written

$$s_{ii'} = \widehat{\text{cov}}(\hat{\beta}_i, \hat{\beta}_{i'}), \quad i, i' = 1, \ldots, 4.$$

The sample estimates are obtained by substituting s^2 for σ^2 in the $\sigma_{ii'}$. Important quantities in later sections are the elements of the covariance matrix of the joint distribution of $\hat{\beta}_2$ and $\hat{\beta}_4$:

$$\sigma_{22} = \sigma^2\left(\frac{1}{n_A} + \frac{1}{n_B} + \frac{\bar{X}_A^2}{SSX_A} + \frac{\bar{X}_B^2}{SSX_B}\right),$$
$$\sigma_{44} = \sigma^2\left(\frac{1}{SSX_A} + \frac{1}{SSX_B}\right),$$
$$\sigma_{24} = -\sigma^2\left(\frac{\bar{X}_A}{SSX_A} + \frac{\bar{X}_B}{SSX_B}\right),$$

where SSX_A and SSX_B are the sums of squares for X in Groups A and B:

$$SSX_A = \sum_{j=1}^{n_A} (X_j - \bar{X}_A)^2,$$
$$SSX_B = \sum_{j=n_A+1}^{N} (X_j - \bar{X}_B)^2,$$

and $\bar{X}_A$ and $\bar{X}_B$ are the within-group sample means.

Under the regression model, $D(X)$ is the best linear unbiased estimate of $\Delta(X)$. The sampling distribution of $D(X)$ is $N[\Delta(X), \sigma^2_{D(X)}]$. The variance of $D(X)$ may be written

$$\sigma^2_{D(X)} = \sigma^2\left[\frac{1}{n_A} + \frac{1}{n_B} + \frac{(X-\bar{X}_A)^2}{SSX_A} + \frac{(X-\bar{X}_B)^2}{SSX_B}\right].$$

An unbiased estimate of $\sigma^2_{D(X)}$, denoted by $s^2_{D(X)}$, is formed by substituting s^2 for σ^2 in the previous expression. The quantity $(N-4)\cdot[s^2_{D(X)}/\sigma^2_{D(X)}]$ is distributed as $\chi^2(N-4)$.

A useful decomposition of $\sigma^2_{D(X)}$ is

$$\sigma^2_{D(X)} = \sigma^2_{D(C_a)} + \sigma_{44}(X - C_a)^2, \quad (3)$$

where

$$C_a = \frac{\bar{X}_A SSX_B + \bar{X}_B SSX_A}{SSX_A + SSX_B} = -\sigma_{24}/\sigma_{44}. \quad (4)$$

The quantity in Equation 4 is called C_a because of its role in the Johnson-Neyman technique as the center of accuracy. It can be seen from Equation 3 that C_a has the property that $D(C_a)$ is the $D(X)$ with the smallest $\sigma^2_{D(X)}$. This minimum value of $\sigma^2_{D(X)}$ can be written

$$\sigma^2_{D(C_a)} = \sigma_{22} + \sigma_{24}C_a = \sigma^2\left[\frac{1}{n_A} + \frac{1}{n_B} + \frac{(\bar{X}_A - \bar{X}_B)^2}{SSX_A + SSX_B}\right].$$

Another important quantity in the study of nonparallel regression lines is the point of intersection of the lines. The abscissa of the point of intersection of the population within-group regressions is denoted by X^o; $X^o = -\beta_2/\beta_4$. Properties of X^o are that $\Delta(X^o) = 0$, and that $\Delta(X) = \beta_4(X - X^o)$. The maximum likelihood estimator of X^o is the abscissa of the point of intersection of the sample within-group regression lines. This quantity is denoted by $\widehat{X^o}$; $\widehat{X^o} = -\hat{\beta}_2/\hat{\beta}_4$. Properties of $\widehat{X^o}$ are that $D(\widehat{X^o}) = 0$ and that $D(X) = \hat{\beta}_4(X - \widehat{X^o})$.

The statistical procedures considered in the following sections are based on two kinds of inferences about $\Delta(X)$: nonsimultaneous inferences and simultaneous inferences. Nonsimultaneous procedures are used to make inferences about $\Delta(X)$ at a single, prespecified value of X. Simultaneous procedures are used to make inferences about $\Delta(X)$ for a range of values of X.

A general expression for interval estimates of $\Delta(X)$ is given by

$$\Delta(X) \in [D(X) \pm \sqrt{K^\alpha s^2_{D(X)}}]. \quad (5)$$

Nonsimultaneous inferences for $\Delta(X)$ at a specified value of X are obtained from Expression 5 by setting the constant $K^\alpha = F^\alpha(1, N-4)$. Simultaneous inferences for $\Delta(X)$ are obtained from Expression 5 by setting $K^\alpha = 2F^\alpha(2, N-4)$. Additional discussion of

simultaneous versus nonsimultaneous inferences about $\Delta(X)$ is found in the section on the Johnson-Neyman technique.

ANCOVA With Nonparallel Regressions

ANCOVA tests the null hypothesis that $\beta_2 = 0$, given that $\beta_4 = 0$. Because the assumption in ANCOVA is that β_4 is identically zero, it is likely that this assumption is violated to some extent in many research settings. The results of this article speak to both large and small values of β_4. In this section the ANCOVA estimate of the treatment effect is expressed in terms of $D(X)$. This result gives rise to a novel interpretation of the ANCOVA estimate that is not tied to the assumption of parallel regressions. Also, statistical inference problems in ANCOVA are investigated, and a modified version of ANCOVA is proposed.

Preliminary Tests

Many textbook presentations of ANCOVA advise that a determination of the validity of the assumption of equal population within-group slopes be attempted before an ANCOVA is performed. A statistical test of the assumption of equal slopes is carried out by testing the null hypothesis that $\beta_4 = 0$ against the alternative that $\beta_4 \neq 0$. The test statistic is $\hat{\beta}_4^2/s_{44}$, which is distributed under the null hypothesis as $F(1, N - 4)$. Thus the null hypothesis of equal slopes is rejected (at level α) when $\hat{\beta}_4^2/s_{44} > F^{\alpha}(1, N - 4)$. Because this test assumes equal population within-group residual variances, inspection of the assumption of a common σ^2 should precede the test of equal slopes.

The often recommended decision rule to use ANCOVA when a significant difference in the slopes is not detected and to shun ANCOVA whenever a significant difference is detected has both statistical and logical drawbacks. Important differences in the slopes may go undetected because of a lack of power in the preliminary test. Conversely, with large enough samples the null hypothesis of equal slopes may be rejected when β_4 is essentially zero. Moreover, the strict dichotomy between equal and unequal slopes is an oversimplification of the problem of comparing regression lines. It seems plausible that small values of β_4 do not seriously affect the validity of conclusions based on an ANCOVA, whereas large values may have serious consequences. The following sections explicate these effects and describe some alternatives to ANCOVA.

ANCOVA Estimate of the Treatment Effect

When the assumption that $\beta_4 = 0$ is satisfied exactly, the vertical distance between the population within-group regressions is constant over the range of X and equals β_2. The estimate of the treatment effect obtained from ANCOVA, the adjusted mean difference, is an estimate of this constant difference of the population regressions. The adjusted mean difference calculated in ANCOVA is

$$\bar{Y}_A - \bar{Y}_B - \hat{\beta}_p(\bar{X}_A - \bar{X}_B), \qquad (6)$$

where $\hat{\beta}_p$ is a pooled estimate of the (assumed) common within-group regression slope.

$$\hat{\beta}_p = \frac{\hat{\beta}_A SSX_A + \hat{\beta}_B SSX_B}{SSX_A + SSX_B}.$$

When the within-group regressions are not parallel, the difference between the regressions is a function of X. A key question in the investigation of the properties of ANCOVA when $\beta_4 \neq 0$ is, For what value of X does $D(X)$ equal the adjusted mean difference from Expression 6? This question may be answered by solving for X in the equation $\hat{\beta}_2 + \hat{\beta}_4 X = \bar{Y}_A - \bar{Y}_B - \hat{\beta}_p(\bar{X}_A - \bar{X}_B)$. The value of X that satisfies this equation is C_a. A restatement of this result is that

$$D(C_a) = \bar{Y}_A - \bar{Y}_B - \hat{\beta}_p(\bar{X}_A - \bar{X}_B). \quad (7)$$

Whether or not $\beta_4 = 0$, the estimated treatment effect in ANCOVA is identical to $D(C_a)$. The ANCOVA estimate of the treatment effect can be obtained through the fitting of separate within-group regressions without reference to a pooled within-groups regression slope.

The adjusted mean difference of ANCOVA is the most precise evaluation of the difference between the within-group regressions. That is, in ANCOVA the value of X at which to evaluate $D(X)$ is chosen to minimize $\sigma^2_{D(X)}$ and thus maximize the precision of the comparison of the regressions. Under the ANCOVA assumption

of $\beta_4 = 0$, this precision is maintained for all values of X.

The conventional use of ANCOVA is to investigate whether the population regression lines coincide, given that the lines are parallel. The adjusted mean difference in Expression 6 is an estimate of the (assumed) constant vertical distance between the population regressions. The result that $D(C_a)$ is equal to this estimate gives rise to a novel interpretation of ANCOVA that is not tied to the assumption that the regression lines are parallel. Because C_a is a weighted average of $\bar{X}_A$ and $\bar{X}_B$, $D(C_a)$ is the vertical distance between the sample regression lines at a point in the middle part of the data. Consequently, regardless of whether the population regressions are parallel, ANCOVA has an interpretation as a procedure for evaluating an "average" treatment effect to the extent that $\Delta(C_a)$ can be thought of as a treatment effect for an "average" individual. Other procedures based on measures of an average treatment effect are considered in the next section.

Statistical Inference

Equation 7 shows that $D(C_a)$ is the ANCOVA estimate of β_2, assuming that $\beta_4 = 0$. Statistical inference procedures in ANCOVA further rely on the assumption that $\beta_4 = 0$ to obtain a more precise estimate of σ^2. I describe below the statistical inference procedures used in ANCOVA and note difficulties with the additional use of the assumption $\beta_4 = 0$.

Sampling variances. The estimated variance of $D(C_a)$ that is used in ANCOVA may be written

$$s_p^2\left[\frac{1}{n_A} + \frac{1}{n_B} + \frac{(\bar{X}_A - \bar{X}_B)^2}{SSX_A + SSX_B}\right].$$

The quantity $(N - 3)s_p^2$ is the residual sum of squares obtained by replacing $\hat{\beta}_A$ and $\hat{\beta}_B$ by $\hat{\beta}_p$ in fitting the sample within-group regressions. This residual sum of squares can be partitioned into two terms: one proportional to s^2 and one proportional to $\hat{\beta}_4^2$.

$$(N-3)s_p^2 = (N-4)s^2 + \frac{\hat{\beta}_4^2}{\frac{1}{SSX_A} + \frac{1}{SSX_B}}. \quad (8)$$

The estimated sampling variance of $D(C_a)$ used in ANCOVA is equal to $(s_p^2/s^2)s^2_{D(C_a)}$. From Equation 8

$$s_p^2 = \frac{s^2}{N-3}(N - 4 + \hat{\beta}_4^2/s_{44}). \qquad (9)$$

And from Equation 9, whenever $\hat{\beta}_4^2/s_{44} > 1$, $s_p^2/s^2 > 1$.

From the decomposition in Equation 8, it can be seen that s_p^2 is a pooled estimate of σ^2, in which both s^2 and the quantity

$$\frac{\hat{\beta}_4^2}{\frac{1}{SSX_A} + \frac{1}{SSX_B}}$$

are used as estimate of σ^2. This additional estimate is an unbiased estimate of σ^2 only when $\beta_4 = 0$;

$$E\left(\frac{\hat{\beta}_4^2}{\frac{1}{SSX_A} + \frac{1}{SSX_B}}\right) = \sigma^2\left(1 + \frac{\beta_4^2}{\sigma_{44}}\right).$$

The decomposition in Equation 8 shows that $(N - 3)s_p^2/\sigma^2$ is the sum of a chi-square variate with $N - 4$ degrees of freedom and a chi-square variate with 1 degree of freedom, the latter being a noncentral chi-square variate whenever $\beta_4 \neq 0$. Thus when $\beta_4 = 0$, $(N - 3)s_p^2/\sigma^2$ is distributed as $\chi^2(N - 3)$. When $\beta_4 \neq 0$, $(N - 3)s_p^2/\sigma^2$ has a noncentral chi-square distribution with $N - 3$ degrees of freedom and noncentrality parameter β_4^2/σ_{44}. Consequently, the expected value of s_p^2 may be written:

$$E(s_p^2) = \sigma^2\left[1 + \frac{\beta_4^2}{(N-3)\sigma_{44}}\right]. \qquad (10)$$

Hypothesis tests. The ANCOVA null hypothesis that $\beta_2 = 0$, given that $\beta_4 = 0$, is tested by use of the test statistic

$$\frac{[\bar{Y}_A - \bar{Y}_B - \hat{\beta}_p(\bar{X}_A - \bar{X}_B)]^2}{s_p^2\left[\frac{1}{n_A} + \frac{1}{n_B} + \frac{(\bar{X}_A - \bar{X}_B)^2}{SSX_A + SSX_B}\right]} = \frac{[D(C_a)]^2}{s^2_{D(C_a)}}\left(\frac{s^2}{s_p^2}\right). \quad (11)$$

Under the null hypothesis, the test statistic is distributed as $F(1, N - 3)$. This null hypothesis is rejected when the test statistic is greater than $F^{\alpha}(1, N - 3)$.

When $\beta_4 \neq 0$, the standard distribution

theory and probability statements for this test statistic are no longer valid. The test statistic in Equation 11 has a doubly noncentral F distribution with 1 and $(N-3)$ degrees of freedom and noncentrality parameters $[\Delta(C_a)]^2$, β_4^2/σ_{44}. Even when $\beta_2 = 0$, the distribution of the test statistic is a doubly noncentral F.

A modified version of ANCOVA can be formulated by using $s^2_{D(C_a)}$ as the denominator in the test statistic in Equation 11; that is, s^2, not s^2_p, is used to estimate σ^2. This modified procedure is appropriate for testing the null hypothesis $\beta_2 = 0 | \beta_4 = 0$. What is different is that the assumption that $\beta_4 = 0$ is not used in constructing the estimate of σ^2. Although this modified procedure is not optimal when the assumption $\beta_4 = 0$ holds exactly, the procedure may be thought of as a safer ANCOVA in that the procedure retains its statistical properties when the assumption is violated. For small values of β_4, the conditional null hypothesis is still of interest, because $\Delta(X)$ is only a weak function of X.

The test statistic for the null hypothesis that $\Delta(C_a) = 0$ in this procedure is

$$\frac{[D(C_a)]^2}{s^2_{D(C_a)}}. \tag{12}$$

Under the null hypothesis the test statistic of Expression 12 is distributed as $F(1, N-4)$, regardless of the value of β_4. The null hypothesis is rejected at level α when Expression 12 is greater than $F^\alpha(1, N-4)$.

Interval estimates. The population treatment effect identified in ANCOVA is the constant difference of the population within-group regression lines. An interval estimate of this treatment effect with confidence coefficient $1-\alpha$ is of the form of Expression 5 and is the interval bounded by the end points

$$D(C_a) \pm \sqrt{F^\alpha(1, N-3) s^2_{D(C_a)} \left(\frac{s^2_p}{s^2}\right)}. \tag{13}$$

The probability statement for this interval estimate is exact only when $\beta_4 = 0$.

Analogous to the safer version of ANCOVA, an interval estimate for $\Delta(C_a)$ with confidence coefficient $1-\alpha$ can be constructed that does not depend on the assumption $\beta_4 = 0$. This interval estimate is bounded by the end points

$$D(C_a) \pm \sqrt{F^\alpha(1, N-4) s^2_{D(C_a)}}. \tag{14}$$

Assessment of an Average Distance

Overall comparisons of the regression lines can be useful even when the treatment effect is a function of X. In this section a measure of the average distance between the within-group regressions is presented. Statistical procedures based on this measure are formulated, and comparisons with ANCOVA are obtained.

Measure of Average Distance

An obvious measure for an overall comparison of the regressions is the average difference of the within-group regressions. Following Rubin (1977), this average difference is defined in the population as the sum of the vertical difference of the population within-group regressions weighted by the population distribution of X. (See Rubin, 1977, Figure 1 and Equation 1.) Restricting the population to the subpopulation containing the observed values of X, the average distance can be written

$$\frac{1}{N}\sum_{j=1}^{N} (\beta_2 + \beta_4 X_j) = \Delta(\bar{X}_G), \tag{15}$$

where

$$\bar{X}_G = \frac{n_A \bar{X}_A + n_B \bar{X}_B}{N}$$

is the grand mean of X. One interpretation for $\Delta(\bar{X}_G)$ is as the difference of the expected group outcomes for a student having the average initial characteristic, that is, the treatment effect for the average student.

Estimation and Inference

In the sample $D(\bar{X}_G)$, the best linear unbiased estimate of $\Delta(\bar{X}_G)$ is the average vertical distance between the sample within-group regression lines:

$$\frac{1}{N}\sum_{j=1}^{N} (\hat{\beta}_2 + \hat{\beta}_4 X_j) = D(\bar{X}_G).$$

The variance of $D(\bar{X}_G)$ is

$$\sigma^2_{D(\bar{X}_G)} = \sigma^2\left[\frac{1}{n_A} + \frac{1}{n_B} + \frac{(\bar{X}_A - \bar{X}_B)^2}{N^2}\left(\frac{n_B^2}{SSX_A} + \frac{n_A^2}{SSX_B}\right)\right].$$

Because

$$\sigma^2_{D(\bar{X}_G)} = \sigma^2_{D(C_a)} + \sigma_{44}(\bar{X}_G - C_a)^2,$$

$$\sigma^2_{D(\bar{X}_G)} > \sigma^2_{D(C_a)}$$

whenever $\bar{X}_G \neq C_a$. An unbiased estimate of $\sigma^2_{D(\bar{X}_G)}$, denoted by $s^2_{D(\bar{X}_G)}$, is formed by substituting s^2 for σ^2 in the variance of $D(\bar{X}_G)$. [Atiqullah, 1964, briefly discussed $D(\bar{X}_G)$ as an alternative to the ANCOVA estimate of the treatment effect, and he derived $\sigma^2_{D(\bar{X}_G)}$ under the assumption that $n_A = n_B$.]

The test statistic for the null hypothesis that $\Delta(\bar{X}_G) = 0$ is

$$\frac{[D(\bar{X}_G)]^2}{s^2_{D(\bar{X}_G)}}. \tag{16}$$

Under the null hypothesis the test statistic of Expression 16 is distributed as $F(1, N-4)$, and the null hypothesis is rejected at level α when this test statistic is greater than $F^\alpha(1, N-4)$. An interval estimate for $\Delta(\bar{X}_G)$ with confidence coefficient $1-\alpha$ is the interval bounded by the end points

$$D(\bar{X}_G) \pm \sqrt{F^\alpha(1, N-4) s^2_{D(\bar{X}_G)}}. \tag{17}$$

Comparison With ANCOVA

The treatment effects identified in ANCOVA and through this measure of average distance differ because $\bar{X}_G$ differs, in general, from C_a.

$$E[D(C_a) - D(\bar{X}_G)] = \Delta(C_a) - \Delta(\bar{X}_G) = \beta_4(C_a - \bar{X}_G), \tag{18}$$

$$C_a - \bar{X}_G = (\bar{X}_A - \bar{X}_B)\left[\frac{n_B SSX_B - n_A SSX_A}{N(SSX_A + SSX_B)}\right]. \tag{19}$$

From Equation 19 $C_a = \bar{X}_G$ when either (a) $\bar{X}_A = \bar{X}_B$ or (b) $n_B SSX_B = n_A SSX_A$. A sufficient condition for (b) to be satisfied is that $n_A = n_B$ and the sample variances of X be identical in both groups. When n_A and n_B are large, the means and variances rarely will differ appreciably between groups.

The justification for the average distance measure makes $\Delta(\bar{X}_G)$ more attractive than $\Delta(C_a)$ for an overall comparison of nonparallel regression lines. Because $D(C_a)$ is the most precise comparison of the regressions, it is worthwhile to consider whether $D(C_a)$ is competitive with $D(\bar{X}_G)$ for estimating the average distance measure $\Delta(\bar{X}_G)$. Comparing the mean square errors (mse) of $D(C_a)$ and $D(\bar{X}_G)$ for estimating $\Delta(\bar{X}_G)$, we find

$$\text{mse}[D(C_a)] - \text{mse}[D(\bar{X}_G)] = \sigma_{44}(C_a - \bar{X}_G)^2\left(\frac{\beta_4^2}{\sigma_{44}} - 1\right), \tag{20}$$

where $\text{mse}[D(X)] = E[D(X) - \Delta(\bar{X}_G)]^2$. From Equation 20 $D(C_a)$ has a smaller mse than $D(\bar{X}_G)$ whenever $C_a \neq \bar{X}_G$ and $\beta_4^2/\sigma_{44} < 1$, and $D(\bar{X}_G)$ has the smaller mse whenever $C_a \neq \bar{X}_G$ and $\beta_4^2/\sigma_{44} > 1$.

Use of the Johnson-Neyman Technique

The Johnson-Neyman technique (J-N) is intended specifically for the analysis of data with nonparallel within-group regressions. J-N determines a region of significance—values of X for which the expected group outcomes differ. The assumptions of J-N and ANCOVA differ only in that the ANCOVA assumption that $\beta_4 = 0$ is not made in J-N. Consequently, J-N is considered an alternative to ANCOVA when this assumption is of doubtful validity (e.g., Kerlinger & Pedhazur, 1973).

J-N has a different objective than the procedures in previous sections. J-N produces an assessment of $\Delta(X)$ as a function of X, whereas the pick-a-point procedures assess only an overall treatment effect. When $\beta_4 = 0$, $\Delta(X)$ reduces to an overall effect, and thus when this assumption is satisfied exactly, ANCOVA yields an assessment of the treatment effect for the range of X. When $\beta_4 \neq 0$, an overall treatment effect can be thought of as an approximation to $\Delta(X)$ in which the dependence on X is suppressed. For slight dependence (β_4 small) this approximation can be useful. For large β_4 the approximation sacrifices considerable information and can be misleading, as when the regressions intersect in the middle of the data.

Because J-N does not depend on the assumption that $\beta_4 \neq 0$, J-N is useful for large, small, or zero values of β_4. In this section the use of J-N for assessing $\Delta(X)$ as a function of X and the additional use of J-N for assessing an overall treatment effect are investigated. The complementary roles of J-N and pick-a-point procedures guide the exposition.

Johnson-Neyman Technique

The region of significance determined by the Johnson-Neyman technique provides an assessment of $\Delta(X)$ as a function of X. Simultaneous regions of significance (R') and nonsimultaneous regions of significance (R) can be constructed. Simultaneous regions are appropriate for making statements about the difference between the regressions over the range of X. A nonsimultaneous region can be validly used when a statement about the difference between the regressions at a single X value is desired (Potthoff, 1964).

Simultaneous region of significance. A simultaneous region of significance is defined by Potthoff (1964) as

> a region such that, with confidence $\geq$ 95 per cent (for $\alpha = .05$) we can state that the two groups are different simultaneously for all points contained in it. In other words, in the long run, not more than 5 per cent of such regions which are calculated will contain any points at all for which the two groups are equal in expected criterion score. (p. 244)

R' is composed of two portions of the X-axis: one in which it can be said that $\Delta(X) > 0$ for all X simultaneously and the other in which it can be said that $\Delta(X) < 0$ for all X simultaneously, both with confidence $100(1 - \alpha)\%$. One or both of these regions may be null.

One derivation of the simultaneous region is to construct a $100(1 - \alpha)\%$ simultaneous confidence band for the line $Y = \Delta(X)$. The simultaneous confidence band used in Potthoff (1964) is the Working-Hotelling band (see Miller, 1966, p. 111) that consists of the region in the X,Y plane enclosed by the upper and lower confidence functions,

$$D(X) \pm \sqrt{2F^{\alpha}(2, N - 4)s^2_{D(X)}}. \quad (21)$$

The simultaneous region of significance is the portion of the X-axis that is composed of values of X that lie outside the intersection of this confidence band with the X-axis. (Other simultaneous regions could be based on the many simultaneous confidence bands that serve as alternatives to the Working-Hotelling band. Applications of some of these bands to J-N are found in Aitkin, 1973, and Rogosa, 1978.)

Any particular X lies in R' if and only if a Y value of zero is not included in the vertical range of the confidence band at that X. Thus the condition for X to lie in R' is that

$$[D(X)]^2 - 2F^{\alpha}(2, N - 4)s^2_{D(X)} > 0. \quad (22)$$

The set of X values that satisfy Expression 22 informs about values of the initial characteristic X for which the treatment is effective. Clearly, this information goes well beyond that provided by an assessment of an overall treatment effect.

The region of significance does not lose its validity when the regressions are parallel. When $\beta_4 = 0$, R' will include the range of X if the treatment effect is large and will be null if the treatment effect is negligible. If it were known that $\beta_4 = 0$, assessment of an overall treatment effect would be more efficient than the simultaneous region, but generally both procedures will indicate similar conclusions about the treatment effect.

Nonsimultaneous region of significance. The original formulation of the Johnson-Neyman (1936) technique and most subsequent technical work and substantive applications are limited to a nonsimultaneous region of significance. Potthoff (1964) argued that the probability statement for the nonsimultaneous region of significance is usually incorrectly interpreted: Statements about the difference of the regressions based on the nonsimultaneous region are valid only for a single specific value of X and not for all X values in R.

The point X lies in R if and only if $D(X)$ significantly differs from zero. Thus R is composed of all X values that satisfy the inequality,

$$\frac{[D(X)]^2}{s^2_{D(X)}} > F^{\alpha}(1, N - 4). \quad (23)$$

The condition in Expression 23 is identical to those used in previous sections in statistical tests for overall treatment effects. Specifically, for $X = \bar{X}_G$, Expression 23 is the test for the average distance meausre from Expression 16, and for $X = C_a$, Expression 23 is the test for the safer ANCOVA from Expression 12.

A statement equivalent to Expression 23 is that X lies in R if the confidence interval for the point $\Delta(X)$ does not include zero. This confidence interval is bounded by the end points,

$$D(X) \pm \sqrt{F^{\alpha}(1, N - 4)s^2_{D(X)}}.$$

The $100(1 - \alpha)\%$ nonsimultaneous confidence band for the line $Y = \Delta(X)$ is simply the concatenation of the confidence intervals for each point on the line and is bounded by the confidence functions,

$$D(X) \pm \sqrt{F^{\alpha}(1, N - 4)s^2_{D(X)}}.$$

A second interpretation of R is that values of X in R lie outside the intersection of the X-axis and this nonsimultaneous confidence band.

A third interpretation of the nonsimultaneous region of significance may be obtained from a confidence interval for the abscissa of the point of intersection of the within-group regressions, X°. Values of X that lie outside this confidence interval make up the region of significance. Because the confidence interval for X° is composed of the X values that do not satisfy Expression 23 (see Fisher, 1946, section 26.2; Kastenbaum, 1959), this interpretation leads to the same R as resulted previously.

Because $2F^{\alpha}(2, N - 4) > F^{\alpha}(1, N - 4)$, R' is a subset of R. When $D(X)$ has marginal statistical significance over parts of the range of X, R' will be substantially smaller than R, because all X for which $F^{\alpha}(1, N - 4) < [D(X)]^2/s^2_{D(X)} \leq 2F^{\alpha}(2, N - 4)$ lie in R but not in R'.

Overall Treatment Effects and the Region of Significance

For the values of X in R', it can be said (at the specified α level) that the within-group regressions differ simultaneously for these points. This statement is useful in identifying a range of initial characteristics for which there is a sizable treatment effect.

The major problem with the use of R' in assessments of the effects of a treatment is that R' may be very small or may only include points outside the plausible range of X. In these situations the data cannot support the strong statement of the simultaneous region about differential treatment effects over the range of X. In general, R' is likely to be most informative when large differential treatment effects exist. The region of significance and the overall treatment effect address different but complementary questions about the effectiveness of the treatment. To clarify some differences and similarities between the two approaches, relations between R' and overall treatment effects are presented below.

If an overall treatment effect is of sole interest in an investigation, then that effect should be evaluated directly through a nonsimultaneous pick-a-point procedure. Use of the simultaneous region to evaluate an overall effect is conservative. Whenever $\bar{X}_G$ is in R', $D(\bar{X}_G)$ differs significantly from zero (using the same α level for both determinations). This condition holds for any particular X value (e.g., $X = C_a$). The converse does not hold. (The converse does hold for the nonsimultaneous region). Another illustration of the difference between the nonsimultaneous and simultaneous assessments of $\Delta(X)$ is that the interval estimate for the point $\Delta(X)$ is narrower than the simultaneous confidence band for $Y = \Delta(X)$ at the X value by a factor of

$$\sqrt{2F^{\alpha}(2, N - 4)/F^{\alpha}(1, N - 4)}.$$

The most extreme example of a divergence in the results from the simultaneous and nonsimultaneous procedures is that when R' does not exist (no X values lie in R'), it is possible (although not likely) that a statistically significant overall treatment effect is present. R' does not exist when

$$\frac{[D(C_a)]^2}{s^2_{D(C_a)}} + \frac{\hat{\beta}^2_4}{s_{44}} \leq 2F^{\alpha}(2, N - 4).$$

(See Rogosa, 1978.) Consequently, $D(C_a)$ is statistically significant even when R' does not exist, whenever

$$F^{\alpha}(1, N - 4) < \frac{[D(C_a)]^2}{s^2_{D(C_a)}} \leq [2F^{\alpha}(2, N - 4) - \hat{\beta}^2_4/s_{44}].$$

On the other hand, in some situations an assessment of an overall treatment effect may be misleading. The most extreme example is when $\hat{\beta}_4$ is large and the sample regressions intersect in the middle of the data (i.e., $\hat{X^{\circ}} = \bar{X}_G$). Then $D(\bar{X}_G) = 0$, although the sample data indicate important differential treatment effects. The research questions and the variables measured will often provide guidance as to the appropriate statistical procedures. The intent here is to inform about properties of and tradeoffs among the procedures for comparing regressions.

Simulation Studies of the Robustness of ANCOVA

The robustness of ANCOVA to violations of the assumption of equal population regression slopes has been investigated in a number of Monte Carlo studies. The conclusion of Hamilton (1977) that "ANCOVA appeared robust to the violation of the assumption of homogeneity of regression when group sizes were equal; the technique appeared not to be robust for unequal group sizes" (p. 701) is representative of the results of the various simulations. My comments are directed specifically toward two published simulation studies: Hamilton (1976, 1977) and Peckham (1969), which is summarized in Glass, Peckham, and Sanders (1972). Serious logical and statistical flaws mar these simulations and render their conclusions misleading.

The results of the simulations can be explained and extended using the analytic results of previous sections. Equation 19 is used to show why the robustness of ANCOVA appears to depend on equal group sizes. Implicit in these simulations is the definition of the treatment effect as the difference between the nonparallel regressions at $X = \bar{X}_G$. Consequently, if the simulated data are generated so that $\bar{X}_G = C_a$, ANCOVA may appear not to be appreciably affected by the unequal regression slopes.

In the simulations, data are generated for two groups having nonparallel population within-group regressions. For each of a variety of within-group slope combinations, an ANCOVA F statistic is computed from the data generated at each replication. Then, for each of these slope combinations, the proportion of F statistics that exceed $F^{\alpha}(1, N - 3)$ is tabulated. The empirical alpha level or the empirical power is thus obtained.

In general, the ANCOVA F statistic computed in these simulations will be distributed as a doubly noncentral F with 1 and $(N - 3)$ degrees of freedom and noncentrality parameters $[\Delta(C_a)]^2$, β_4^2/σ_{44}. The percentage points of the doubly noncentral F distribution provide the actual Type I error rate and power. An approximation to the doubly noncentral F by a multiple of a central F (e.g., N. L. Johnson & Kotz, 1970, section 30.6) could be used to evaluate analytically the percentage points as a function of the parameters specified in the simulations. This analytic approach is preferable to the simulations.

The crucial role of equal group sizes is seen from Equation 19. In any of these simulations, the observed within-group variances of X will be nearly equal on the average because the groups are formed by random assignment. Under the condition $SSX_A/n_A = SSX_B/n_B$, Equations 18 and 19 show that $D(C_a) - D(\bar{X}_G)$ is proportional to $n_B^2 - n_A^2$. Thus when group sizes are equal, nonzero values of β_4 will have little effect on the numerator of the ANCOVA F statistic. Hamilton's (1977) conclusion quoted previously in an obvious consequence of the dependence of $D(C_a) - D(\bar{X}_G)$ on the relation of group sizes.

In Peckham's (1969) simulation the group sizes are equal, and the population regressions are specified to intersect at $\bar{X}_G$. Thus the null treatment effect is defined to be the difference of the population regressions at $\bar{X}_G$. In the first phase of the simulation, $\bar{X}_A = \bar{X}_B$, and thus $C_a = \bar{X}_G$. Consequently, any effects of the unequal slopes on the ANCOVA F statistic are transmitted through the denominator of the F statistic. As β_4 increases, s_p^2 increasingly overestimates σ^2, and as β_4^2/σ_{44} increases, the upper percentage points of the noncentral F decrease. As would be expected, Peckham detected a small effect of unequal slopes on ANCOVA; ANCOVA was found to be increasingly conservative with respect to making a Type I error as β_4 increases (see Glass et al., 1972, Table 17).

In the second phase of the simulation, nonzero values of $\bar{X}_A - \bar{X}_B$ were specified. But again $\bar{X}_G = C_a$ because the conditions of equal group sizes and equal sample variances of X in the two groups were retained. Consequently, differing values of $\bar{X}_A - \bar{X}_B$ can have no effect on the measured robustness of ANCOVA to unequal regression slopes, and the results of this second phase were found to closely duplicate those of the first phase (see Glass et al., 1972, Table 18).

In his investigation of Type I error rates, Hamilton (1976, 1977) used various combinations of group sizes crossed with a number of unequal slope combinations. Although the interpretation of Hamilton's results is complicated slightly by his use of a random covariate, some general conclusions can be drawn that are completely predictable from my analytic results. When group sizes were not equal, large effects of unequal slopes on the ANCOVA F statistic were detected. For the power study Hamilton used only equal group sizes; consequently, he found little effect of unequal slopes on the power of ANCOVA.

Regrettably, both simulations incorporated the unintended violation of the assumption of equal residual variances. The variances of X and Y were both standardized; consequently, whenever $\beta_4 \neq 0$, the residual variances were not equal.

Furthermore, the simulations possess logical flaws that render their conclusions misleading. The key logical flaw in the simulations is the ignorance of the need for an explicit definition of a treatment effect to be estimated. The simulations merely verify (a) that $D(C_a)$ is a good estimate of $\Delta(C_a)$ and (b) that $D(C_a)$ is a poor estimate of $\Delta(X)$ whenever $|C_a - X|$ and β_4 are large. The conclusions about the effects of group sizes resulting from the simulations are an inadequate guide for statistical practice. In no sense is ANCOVA always unaffected by unequal regression slopes when groups sizes are equal. Nor is ANCOVA necessarily shattered by unequal regression slopes when group sizes are unequal. Whether $D(C_a)$ is an adequate measure of a treatment effect depends crucially on the definition of that treatment effect. Whenever the structure of a simulation is such that $\Delta(C_a)$ is the treatment effect to be estimated, ANCOVA will prove to be adequate.

Table 1
Data and Summary Statistics

Group A		Group B	
X	Y	X	Y
.28	2.23	2.36	3.25
.97	4.99	2.11	5.30
1.25	3.37	.45	1.39
2.46	8.54	1.76	4.69
2.51	8.40	2.09	6.56
1.17	3.70	1.50	3.00
1.78	7.93	1.25	5.85
1.21	2.43	.72	1.90
1.63	5.40	.42	3.85
1.98	8.44	1.53	2.95
$\bar{X}_A = 1.52$	$\bar{Y}_A = 5.54$	$\bar{X}_B = 1.42$	$\bar{Y}_B = 3.87$
$s_X = .690$	$s_Y = 2.59$	$s_X = .699$	$s_Y = 1.70$
$r_{XY} = .882$		$r_{XY} = .542$	
$s_{Y \cdot X} = 1.29$		$s_{Y \cdot X} = 1.51$	

Computational Procedures and an Example

The statistical procedures for comparing regression lines are easy to carry out using a multiple-regression program and a hand calculator. The exposition of the computational procedures and the data analysis example make these steps explicit and serve as a review of the recommended statistical procedures for comparing regressions.

The statistical analysis proceeds from least squares estimation of the parameters in Equation 2, the regression of Y on T, X, and TX for all N cases.[1] The estimates of the $\beta_i (i = 1, \ldots, 4)$ yield $D(X) = \hat{\beta}_2 + \hat{\beta}_4 X$. Mean square for error in this regression is s^2. The estimated covariance matrix of the $\hat{\beta}_i$ (which, e.g., can be obtained through the COVB command of the SYSREG routine in SAS) yields the $s_{ii'}$.

The assessment of overall treatment effects through pick-a-point procedures requires evaluation of $D(X)$ and $s_{D(X)}^2$ for specified values of X. To determine $D(\bar{X}_G)$, obtain the mean of all N X values and substitute into $D(X)$. To determine $D(C_a)$, compute $C_a = -s_{24}/s_{44}$ and substitute into $D(X)$. The sampling variance is found from the relations $s_{D(X)}^2 = s_{D(C_a)}^2 + s_{44} \cdot (X - C_a)^2$, where $s_{D(C_a)}^2 = s_{22} + s_{24} C_a$. Interval estimates and tests for the average distance measures are constructed by substituting into Expressions 17 and 16. Interval estimates and tests for the safer ANCOVA are constructed by substituting into Expressions 14 and 12.

The Johnson-Neyman region of significance can be obtained through some additional cal-

[1] It is recommended that descriptive within-group analyses to examine possible outliers, nonlinearities, appropriate transformations, residual variances, and other descriptive measures precede the estimation of Equation 2. However, some within-group information can be recovered from the estimation of Equation 2. The within-group regression lines can be obtained because

$$\hat{\alpha}_A = \hat{\beta}_1 + \hat{\beta}_3; \hat{\beta}_A = \hat{\beta}_3 + \hat{\beta}_4; \hat{\alpha}_B = \hat{\beta}_1; \hat{\beta}_B = \hat{\beta}_3.$$

And the residuals from the within-group regressions are identical to the residuals from the fit to Equation 2.

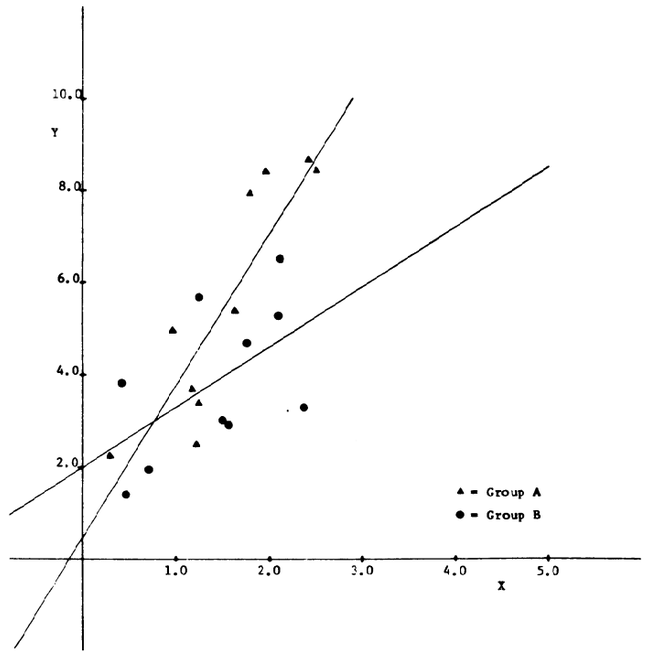

Figure 1. Plot of the data in Table 1 and the sample within-group regression lines.

culations. The end points of R' are the roots of the equation $[D(X)]^2 - 2F^{\alpha}(2, N-4)s^2_{D(X)} = 0$. This equation can be expanded as a quadratic equation in X:

$$[\hat{\beta}_4^2 - 2F^{\alpha}(2, N-4)s_{44}]X^2 + 2[\hat{\beta}_2\hat{\beta}_4 - 2F^{\alpha}(2, N-4)s_{24}]X + [\hat{\beta}_2^2 - 2F^{\alpha}(2, N-4)s_{22}] = 0. \quad (24)$$

Substitution of numerical values into Equation 24 yields an equation of the form $AX^2 + 2BX + C = 0$. If R' exists, the two distinct real roots of this equation are

$$\frac{-B \pm \sqrt{B^2 - AC}}{A}.$$

Denote these roots as X_+ for the larger root and X_- for the smaller root. When $\hat{\beta}_4^2/s_{44} > 2F^{\alpha}(2, N-4)$, R' is composed of the X values: $X > X_+$, $X < X_-$. When $\hat{\beta}_4^2/s_{44} < 2F^{\alpha}(2, N-4)$, R' is composed of the X values: $X_- < X < X_+$ (Rogosa, in press). Alternatively, a graphical procedure for R' is to plot the confidence functions in Expression 21 and find the X values outside the intersection of the X-axis and the simultaneous confidence band for $Y = \Delta(X)$.

The data in Table 1 illustrate these procedures. Summary statistics for each group are also presented in Table 1. The data and the within-group regression lines are plotted in Figure 1. The sample within-group regressions for Groups A and B, respectively, are

$$Y = .497 + 3.31X,$$
$$Y = 2.01 + 1.31X.$$

The sample regressions intersect at the point (.756, 3.00); $\widehat{X^{\circ}}(= -\hat{\beta}_2/\hat{\beta}_4)$ is larger than 4 of the 20 X values.

The difference of the sample regressions is

$$D(X) = -1.51 + 2.00X.$$

In these data $\bar{X}_G = C_a = 1.47$, and thus $D(\bar{X}_G) = D(C_a) = 1.43$. The estimated variance of $D(X)$ is $s^2_{D(X)} = .398 + .910(X - 1.47)^2$. A 95% confidence interval for $\Delta(\bar{X}_G)$ is the interval (.093, 2.77); $D(\bar{X}_G)/s_{D(\bar{X}_G)} = 2.27$.

The simultaneous region of significance (at $\alpha = .05$) is composed of X values that satisfy: $2.63X^2 - 2(6.71)X + 14.93 < 0$. The end points of R' are $X_+ = 3.47$ and $X_- = 1.64$. Because $\hat{\beta}_4^2/s_{44} < 2F^{.05}(2, 16)$, R' is composed of the X values: $1.64 < X < 3.47$. (Of the 20 X values, 7 lie in R'.) In Figure 2 $D(X)$ and the 95% simultaneous confidence band for $Y = \Delta(X)$ are plotted. Figures 1 and 2 constitute a good graphical summary for the comparison of regression lines.

It is useful as a separate matter to consider an analysis of these data using ANCOVA. The within-group slopes do not significantly differ at the .05 level ($\hat{\beta}_4^2/s_{44} = 4.37 < F^{.05}(1, 16) = 4.49$), and thus textbook advice would be to use ANCOVA. The ANCOVA test statistic in Equation 11 equals 4.28, which is not significant at the .05 level: $4.28 < F^{.05}(1, 17) = 4.45$. Recall that $D(C_a)$ differs significantly from zero at the .05 level; that is, the safer ANCOVA (at $\alpha = .05$) rejects the null hypothesis of no overall treatment effect. An equivalent way of contrasting standard ANCOVA and the procedures described in this article is to consider interval estimates of the treatment effect. With confidence coefficient .95, ANCOVA yields an interval estimate (−.028, 2.89), whereas the interval estimate for $\Delta(C_a)$ from Expression 14 of (.093, 2.77) is shorter than the interval from ANCOVA.

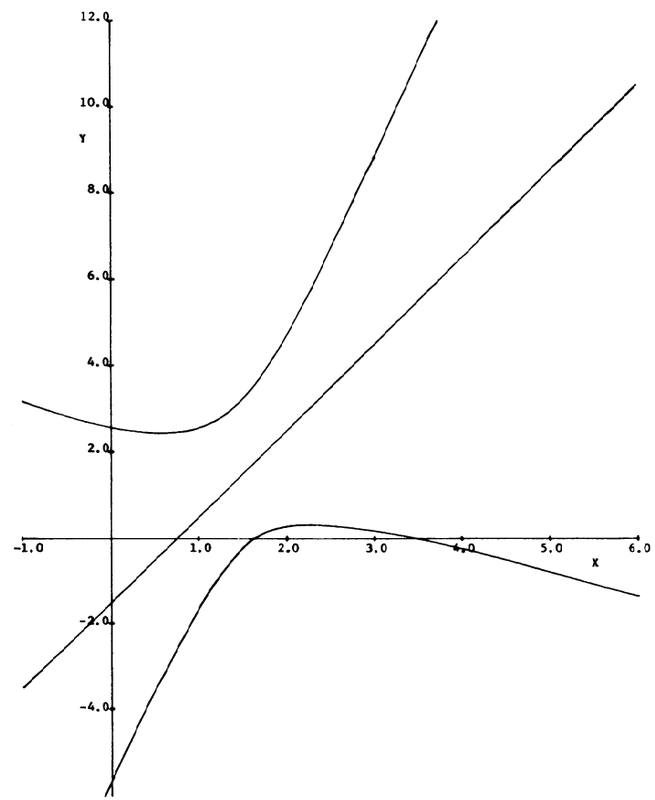

Figure 2. Plot of the line $Y = D(X)$ and the 95% simultaneous confidence band for the line $Y = \Delta(X)$.

Summary and Discussion

This article presents a framework for the assessment of the effects of a treatment through statistical comparisons of within-group regression lines. The effect of the treatment was formulated as a linear function of an individual initial characteristic X. Two types of assessments of the relative effectiveness of the treatment were investigated: assessment of an overall treatment effect through pick-a-point procedures and assessment of the treatment effect as a function of X through the region of significance. The use of both types of assessments was illustrated by the analysis of a small data set.

Measures of an overall treatment effect are most useful when the within-group regression slopes exhibit small to moderate differences. Overall treatment effects would be used, for example, in settings in which differential assignment of cases to alternative programs is not feasible or when individual differences on X have little substantive import. The measure of average distance between the regressions appears to be the most useful measure in general, but specific applications could well dictate a different choice of X for the pick-a-point procedure.

Assessments of the treatment effect as a function of X are crucial when the within-group regressions intersect in the range of X. The Johnson-Neyman technique identifies a range of X values for which the population within-group regressions differ. Undoubtedly, the widespread use of J-N has been hindered by the dearth of computer routines to perform the calculations. Even the specialized J-N computer programs described in Borich, Godbout, and Wunderlich (1976) offer only nonsimultaneous J-N procedures. However, as is shown in this article, the simultaneous (or nonsimultaneous) region of significance can be computed through a few manipulations of the output from a standard multiple-regression program.

The difference of the sample within-group regression lines, $D(X)$, is the key summary of the data for comparing regression lines. In the pick-a-point procedures, the estimate of the treatment effect is $D(X)$ evaluated at a particular X value. In particular, the average distance measure is estimated by $D(\bar{X}_G)$, and the ANCOVA estimate of the treatment effect is $D(C_a)$. Also, the Johnson-Neyman region of significance is formed from a confidence band for the line $Y = \Delta(X)$ that is constructed about the line $Y = D(X)$.

This article clarifies textbook advice on the use of ANCOVA when the within-group regressions are not parallel. Many current textbooks stress the importance of the assumption of equal population slopes. For example, Cohen and Cohen (1975) warned repeatedly that unequal slopes render ANCOVA "invalid" or "preclude its meaningful application." In this view the statistical significance of the test statistic for equal slopes entirely determines whether ANCOVA is meaningful. Although it is proper to question the logic of estimating an assumed constant distance between the regression lines when that distance is a function of X, the strict dichotomy implied by this advice is misleading. The effects of nonparallel regressions depend crucially on the degree of heterogeneity of the slopes. ANCOVA provides certain information about the treatment effect even when significant differences in the slopes are detected. With nonparallel population regressions, ANCOVA is not completely meaningless but it provides limited information and is subject to serious problems. This article argues that an appropriate pick-a-point procedure is preferable to ANCOVA for assessments of overall treatment effects.

The simulation studies of the robustness of ANCOVA to nonparallel population regressions reach different conclusions about the use of ANCOVA. These studies regard the nonadditive treatment effect not as a logical problem but as a violation of a statistical assumption. Many of these studies, because of flaws and limitations in the design of the simulation, conclude that the effects of unequal slopes on ANCOVA are often benign. However, I demonstrate that the logic and conclusions of these simulations are askew.

The results of this article blur the sharp distinction that is usually made between ANCOVA as a method for comparing parallel within-group regressions and J-N as a method for comparing nonparallel regressions. Overall treatment effects are useful measures when the heterogeneity of the regression slopes is not extreme. J-N is still valid when the population regressions are parallel; an assessment of the constant difference of the regressions is obtained with a loss of some power.

Moreover, the results of this article question the need for special procedures for the statistical comparison of parallel within-group regressions. The methods presented for the comparison of nonparallel regressions also accommodate parallel regressions. The risks in the use of special procedures for parallel regressions appear to outweigh the small increase in precision of estimation that can be obtained when the assumption of parallel regressions is satisfied exactly.

References

Aitkin, M. A. Fixed-width confidence intervals in linear regression with applications to the Johnson-Neyman technique. *British Journal of Mathematical and Statistical Psychology*, 1973, *26*, 261–269.

Atiqullah, M. The robustness of covariance analysis of a one-way classification. *Biometrika*, 1964, *51*, 365–372.

Borich, G. D., Godbout, R. D., & Wunderlich, K. W. *The analysis of aptitude-treatment interactions: Computer programs and calculations*. Austin, Tex.: Oasis Press, 1976.

Cohen, J., & Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: Halstead Press, 1975.

Fisher, R. A. *Statistical methods for research workers* (10th ed.). London: Oliver & Boyd, 1946.

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

Hamilton, B. L. A Monte-Carlo test of the robustness of parametric and nonparametric analysis of covariance against unequal regression slopes. *Journal of the American Statistical Association*, 1976, *71*, 864–869.

Hamilton, B. L. An empirical investigation of the effects of heterogeneous regression slopes in analysis of covariance. *Educational and Psychological Measurement*, 1977, *37*, 701–712.

Johnson, N. L., & Kotz, S. *Continuous univariate distributions* (Vol. 2). Boston: Houghton Mifflin, 1970.

Johnson, P. O., & Neyman, J. Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, 1936, *1*, 57–93.

Kastenbaum, M. A. A confidence interval on the abscissa of the point of intersection of two fitted regressions. *Biometrics*, 1959, *15*, 323–324.

Kendall, M. G., & Stuart, A. *The advanced theory of statistics* (Vol. 2). London: Griffin, 1967.

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

Miller, R. G. *Simultaneous statistical inference*. New York: McGraw-Hill, 1966.

Peckham, P. D. An investigation of the effects of nonhomogeneity of regression slopes upon the *F*-test of analysis of covariance (Doctoral dissertation, University of Colorado, 1968). *Dissertation Abstracts International*, 1969, *29*, 3531-B. (University Microfilms No. 69-4345)

Potthoff, R. F. On the Johnson-Neyman technique and some extensions thereof. *Psychometrika*, 1964, *29*, 241–256.

Rogosa, D. R. Some results for the Johnson-Neyman technique (Doctoral dissertation, Stanford University, 1977). *Dissertation Abstracts International*, 1978, *38*, 5366-A. (University Microfilms No. 78-02,225)

Rogosa, D. R. On the relationship between the Johnson-Neyman region of significance and statistical tests of parallel within-group regressions. *Educational and Psychological Measurement*, in press.

Rubin, D. B. Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 1977, *2*, 1–26.

Sampson, A. R. A tale of two regressions. *Journal of the American Statistical Association*, 1974, *69*, 682–689.

Received July 27, 1979 ■