UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

ILLUSTRATIONS OF SOME SCHEFFÉ-TYPE TESTS FOR SOME BEHRENS-FISHER-TYPE REGRESSION PROBLEMS

by

Richard F. Potthoff

October 1963

Contract No. AF AFOSR-62-196

It is sometimes required to compare two simple regression lines without assuming that the variances of the two sets of error terms are necessarily equal: one may wish to (A) test whether two parallel regression lines are identical, or (B) test whether two regression lines are parallel. In this paper, a t-test for each of these problems is explained and illustrated numerically. These t-tests bear a certain resemblance to Scheffé's test for the Behrens-Fisher problem, but (unlike the latter) are not randomized tests. This paper is on a practical rather than a theoretical level; the more technical aspects of the tests are covered in a separate paper (Mimeo Series No. 374).

This research was supported in part by Educational Testing Service; it was supported in part by the Air Force Office of Scientific Research; and it was supported in part by a grant from the NATO Research Grants Programme under a joint arrangement between the Institut de Statistique of the University of Paris, the Istituto di Calcolo delle Probabilità of the University of Rome, and the Department of Statistics of the University of North Carolina.

Institute of Statistics
Mimeo Series No. 377

ILLUSTRATIONS OF SOME SCHEFFÉ-TYPE TESTS FOR SOME BEHRENS-FISHER-TYPE REGRESSION PROBLEMS

by

Richard F. Potthoff[1]
University of North Carolina

1. Introduction. In educational and psychological applications, it is sometimes necessary to make certain comparisons between two regression lines when the two variances are different. Such problems arise particularly in connection with experiments comparing two curriculums or two teaching methods (which often have different variances), as is explained more fully in [1] and [2]. The two principal types of problems are (A) testing whether two parallel regression lines are the same, and (B) testing whether two regression lines are parallel. These will be referred to as Problem A and Problem B respectively.

By generalizing an idea which Scheffé [4] used to devise a test for the Behrens-Fisher problem, another paper [3] gives a theoretical development of Scheffé-type tests for Problems A and B. The purpose of the present paper is to provide numerical illustrations of some of the techniques presented in [3]. The illustrations of this paper are all for the case of simple regression, even though Scheffé-type tests are available in [3] both for simple regression and for the more involved case of multiple regression. Scheffé's test itself is a randomized test, but the tests in [3] include both randomized and non-randomized

[1]This research was supported in part by Educational Testing Service; it was supported in part by the Air Force Office of Scientific Research; and it was supported in part by a grant from the NATO Research Grants Programme under a joint arrangement between the Institut de Statistique of the University of Paris, the Istituto di Calcolo delle Probabilità of the University of Rome, and the Department of Statistics of the University of North Carolina.

tests. The tests illustrated in this paper, however, are all non-randomized (under the provision that the independent variates all have distinct values), for it is suggested in [3] that, for the case of simple regression, the non-randomized tests would ordinarily be used in preference to the randomized tests.

An entirely different test for Problem A is described in [2], and for Problem B in [1]; other possible tests for Problem B which are in the literature are mentioned in [3]. All these tests, however, are inexact (i.e., the actual level of significance is only approximately equal to the stated level of significance) and tend to be less exact the smaller the sample. The Scheffé-type tests, on the other hand, are exact tests based on the t-distribution. Thus these Scheffé-type t-tests are particularly useful for small sample sizes.

The mathematical models for Problems A and B may be expressed respectively as follows:

Problem A. We suppose that we have one group consisting of M pairs (Y_1, X_1), (Y_2, X_2), ..., (Y_M, X_M), and another group consisting of N pairs (Z_1, W_1), (Z_2, W_2), ... (Z_N, W_N). For each i, the observed value Y_i follows the relation

(1.1a) $$Y_i = \alpha_Y + \beta X_i + e_i \quad ;$$

and for each j, the observed value Z_j follows the relation

(1.1b) $$Z_j = \alpha_Z + \beta W_j + f_j \quad .$$

The α's and β are unknown parameters (regression coefficients), the X_i's and W_j's are specified constants, and the e_i's and f_j's are error terms which are independent and normal with mean 0 and unknown variances σ_e^2 and σ_f^2 respectively.

The hypothesis to be tested is that $\alpha_Y = \alpha_Z$, i.e., that the two parallel regression lines associated with (1.1a) and (1.1b) are identical. We will also be able to obtain confidence bounds on $(\alpha_Z - \alpha_Y)$. Standard regression methods cannot be used for this problem, of course, since two different variances σ_e^2 and σ_f^2 are involved.

Problem B. The model is the same as for Problem A, except that in place of (1.1) we have

$$(1.2a) \qquad Y_i = \alpha_Y + \beta_Y X_i + e_i$$

and

$$(1.2b) \qquad Z_j = \alpha_Z + \beta_Z W_j + f_j$$

The hypothesis to be tested is that $\beta_Y = \beta_Z$, i.e., that the two regression lines associated with (1.2a) and (1.2b) are parallel. Confidence bounds on $(\beta_Z - \beta_Y)$ can also be obtained.

For definiteness, we will always assume (with no loss of generality) that $M \leq N$.

The reader should refer to [1] and [2] for a further discussion of the relation of Problems A and B to the experimental situation of comparing two curriculums or two teaching methods.

2. The test for Problem A. This section tells how to calculate the test statistic for Problem A, and then Section 3 will present a numerical example.

Since this test, like that of Scheffé [4] for the Behrens-Fisher problem, is based on pairing of the two samples, it is first necessary to effect this pairing. To start, let the first group be arranged in order of increasing X and the second group in order of increasing W, so that $X_1 < X_2 < \ldots < X_M$ and

$W_1 < W_2 < \ldots < W_N$. (In case there are tied values of the X's or of the W's, it will be necessary to break the tie in some way, either by randomizing or else by some other objective but less arbitrary method. An alternative to randomizing, e.g., would be to break the tie on the basis of some second variable similar to X and W which otherwise would not even be utilized at all.)

Paired to each of the M members of the first group will be a member of the second group. Thus M of the W_j's will be paired, but the remaining (N-M) will be unpaired. Let $W_{(i)}$ denote that W_j (of the second group) which is paired to X_i. The pairing is to be done according to the formula

$$(2.1) \qquad \begin{aligned} W_{(i)} &= W_{N+1-i} , && 1 \leq i \leq \nu \\ &= W_{M+1-i} , && \nu+1 \leq i \leq M , \end{aligned}$$

where the integer ν is determined in a way which will be described shortly. Note, however, that in the special case M=N, no ν is necessary, and the pairing (2.1) simply reduces to matching the X's and W's in opposite order.

The optimal way to choose ν (for the general case) is to set it equal to the smallest integer ν such that

$$(2.2) \qquad \delta_\nu = \frac{W_{N-\nu} + W_{M-\nu}}{2} - \frac{\sum_{j=1}^{M-\nu-1} W_j + \sum_{j=N-\nu+1}^{N} W_j}{M-1} - \frac{\sqrt{MN}}{M-1} (X_{\nu+1} - \bar{X})$$

is < 0. [The function δ_ν (2.2) decreases as ν increases, and will turn from positive to negative somewhere between $\nu = 0$ and $\nu = M - 1$.]

Actually, <u>any</u> objective method of choosing ν may be used and the probability of Type I. error will be unaltered; however, the power of the test will be affected. If the experimenter prefers not to use the optimal method for choosing ν because he feels it may be too lengthy computationally, the following simple but somewhat sub-optimal method is recommended instead: choose $\nu = M/2$ if M is even, and choose $\nu = (M-1)/2$ [or $(M+1)/2$] if M is odd.

After completing the pairing, we proceed as follows. Let

$$(2.3)\quad \overline{X} = \frac{1}{M}\sum_{i=1}^{M} X_i,\quad \overline{Y} = \frac{1}{M}\sum_{i=1}^{M} Y_i,\quad \overline{W} = \frac{1}{N}\sum_{j=1}^{N} W_j,\quad \overline{Z} = \frac{1}{N}\sum_{j=1}^{N} Z_j,$$

$$\overline{W}^o = \frac{1}{M}\sum_{i=1}^{M} W_{(i)},\qquad \text{and}\qquad \overline{Z}^o = \frac{1}{M}\sum_{i=1}^{M} Z_{(i)},$$

where $Z_{(i)}$ denotes that Z_j which is associated with $W_{(i)}$. Thus $\overline{W}^o$ and $\overline{Z}^o$ are the means of the <u>paired</u> W_j's and Z_j's respectively. Now we calculate

$$(2.4)\qquad t = \frac{(\overline{Z}-\overline{Y}) - \frac{u'T_A}{u'u}(\overline{W}-\overline{X})}{\sqrt{\frac{1}{M} + \frac{(\overline{W}-\overline{X})^2}{u'u}}\ \sqrt{\frac{S_e^2}{M-2}}},$$

where

$$(2.5)\qquad u'u = \left(\sum_{i=1}^{M} X_i^2 - M\overline{X}^2\right) - 2\sqrt{\frac{M}{N}}\left(\sum_{i=1}^{M} X_i W_{(i)} - M\overline{X}\,\overline{W}^o\right) + \frac{M}{N}\left(\sum_{i=1}^{M} W_{(i)}^2 - M\overline{W}^{o^2}\right),$$

$$(2.6)\qquad u'T_A = \left(\sum_{i=1}^{M} X_iY_i - M\overline{X}\overline{Y}\right) - \sqrt{\frac{M}{N}}\left(\sum_{i=1}^{M} X_i Z_{(i)} - M\overline{X}\,\overline{Z}^o\right)$$

$$- \sqrt{\frac{M}{N}}\left(\sum_{i=1}^{M} Y_i W_{(i)} - M\overline{Y}\,\overline{W}^o\right) + \frac{M}{N}\left(\sum_{i=1}^{M} W_{(i)}Z_{(i)} - M\overline{W}^o\,\overline{Z}^o\right),$$

and

$$(2.7)\qquad S_e^2 = \left(\sum_{i=1}^{M} Y_i^2 - M\overline{Y}^2\right) - 2\sqrt{\frac{M}{N}}\left(\sum_{i=1}^{M} Y_i Z_{(i)} - M\overline{Y}\,\overline{Z}^o\right)$$

$$+ \frac{M}{N}\left(\sum_{i=1}^{M} Z_{(i)}^2 - M\overline{Z}^{o^2}\right) - \frac{(u'T_A)^2}{u'u}.$$

The statistic t (2.4) is used to test the null hypothesis $\alpha_Y = \alpha_Z$. It follows the t-distribution with $(M-2)$ degress of freedom if the null hypothesis is true.

3. Numerical example of the test for Problem A. Suppose there are $M=4$ observations in the first group,

Y	19.1	5.3	22.8	20.7
X	7	0	9	8

and $N=6$ observations in the second group,

Z	13.7	6.9	8.7	17.2	2.5	5.5
W	6	3	4	8	1	2

Arranging the first group in order of increasing X and the second group in order of increasing W, we have

$$Y_1 = 5.3 \qquad Y_2 = 19.1 \qquad Y_3 = 20.7 \qquad Y_4 = 22.8$$
$$X_1 = 0 \qquad X_2 = 7 \qquad X_3 = 8 \qquad X_4 = 9$$

and

$$Z_1 = 2.5 \quad Z_2 = 5.5 \quad Z_3 = 6.9 \quad Z_4 = 8.7 \quad Z_5 = 13.7 \quad Z_6 = 17.2$$
$$W_1 = 1 \quad W_2 = 2 \quad W_3 = 3 \quad W_4 = 4 \quad W_5 = 6 \quad W_6 = 8 \quad .$$

The next step is to find ν. We do not have to calculate δ_ν for all values of ν, but rather we need only find the place where δ_ν goes from positive to negative. Let us start with an intermediate value of ν: we evaluate δ_ν for $\nu = 2$. Using (2.2), and noting that $\bar{X} = 6$, we get

$$\delta_2 = \frac{W_4 + W_2}{2} - \frac{(W_1) + (W_5 + W_6)}{3} - \frac{\sqrt{24}}{3}\,(X_3 - \overline{X})$$

$$= \frac{6}{2} - \frac{15}{3} - \frac{\sqrt{24}}{3}\,(8 - 6) = -\frac{1}{3}\,(6 + 4\sqrt{6}) < 0 .$$

Hence the first negative δ_ν could occur at $\nu = 2$ or at $\nu < 2$. We next calculate

$$\delta_1 = \frac{6 + 3}{2} - \frac{(1 + 2) + (8)}{3} - \frac{\sqrt{24}}{3}\,(7 - 6)$$

$$= \frac{1}{6}\,(5 - 4\sqrt{6}) < 0 .$$

Now, evaluating δ_ν for the next lower value of ν, we find

$$\delta_0 = 4 + 4\sqrt{6} > 0 ,$$

which establishes that δ_1 is the first δ_ν which is < 0. Hence we set $\nu = 1$ in (2.1), and (2.1) then gives us

$$Z_{(1)} = Z_6 = 17.2 \qquad Z_{(2)} = Z_3 = 6.9 \qquad Z_{(3)} = Z_2 = 5.5 \qquad Z_{(4)} = Z_1 = 2.5$$

$$W_{(1)} = W_6 = 8 \qquad W_{(2)} = W_3 = 3 \qquad W_{(3)} = W_2 = 2 \qquad W_{(4)} = W_1 = 1$$

[Note: If we had simply chosen $\nu = M/2$, then ν would be 2, not 1.]

With respect to the quantities in (2.3), we may write

$$M\,\overline{X} = 24, \quad M\overline{Y} = 67.9, \quad N\overline{W} = 24, \quad N\overline{Z} = 54.5, \quad M\overline{W}^{o} = 14, \quad M\overline{Z}^{o} = 32.1.$$

Hence $\overline{Z} - \overline{Y} = -7.89$ and $\overline{W} - \overline{X} = -2$. In order to get t (2.4), we must obtain (2.5 - 2.7), for which we must first calculate

$\Sigma\ Y_i^2 = 1341.23, \quad \Sigma\ X_i^2 = 194, \quad \Sigma\ Z_{(i)}^2 = 379.95, \quad \Sigma\ W_{(i)}^2 = 78\ ,$

$\Sigma\ X_iY_i = 504.5, \quad \Sigma\ Y_iZ_{(i)} = 393.80, \quad \Sigma\ Y_iW_{(i)} = 163.9\ ,$

$\Sigma\ X_iZ_{(i)} = 114.8, \quad \Sigma\ X_iW_{(i)} = 46\ , \quad \Sigma\ W_{(i)}Z_{(i)} = 171.8\ .$

Now we use (2.5 - 2.7) to write

$$u'u = \left(194 - \frac{24^2}{4}\right) - 2\sqrt{\frac{4}{6}}\left(46 - \frac{24 \times 14}{4}\right) + \frac{4}{6}\left(78 - \frac{14^2}{4}\right) = 131.387\ ,$$

$$u'T_A = 97.1 - (-63.524) - (-60.217) + 39.633 = 260.474\ ,$$

and

$$s_e^2 = 516.935 - \frac{(260.474)^2}{131.387} = 0.547\ .$$

Finally, (2.4) becomes

$$t = \frac{-7.89 - \frac{260.474}{131.387}(-2)}{\sqrt{\frac{1}{4} + \frac{(-2)^2}{131.387}}\ \sqrt{\frac{.547}{4-2}}} = \frac{-3.93}{.277} = -14.19\ .$$

The number of degrees of freedom for t is $M-2=2$. Hence, for a two-sided test at the $\alpha = .05$ level, the critical region is $|t| > 4.30$. Since $|-14.19|$ is > 4.30, we reject the null hypothesis $\alpha_Y = \alpha_Z$.

Confidence bounds on $(\alpha_Z - \alpha_Y)$ are easily obtained. For example, a $95^o/o$ confidence interval for $(\alpha_Z - \alpha_Y)$ is given by

$$-3.93 \pm 4.30\ (.277)\ ,$$

i.e., we state that

$$-2.74 \leq (\alpha_Z - \alpha_Y) \leq -5.12$$

with $95^o/o$ confidence.

4. <u>The test for Problem B</u>. This section tells how to calculate the test statistic for Problem B, and then Section 5 will present two numerical examples.

As before, the first thing to do is to pair the two samples. We start again by assuming $X_1 < X_2 < \ldots < X_M$ and $W_1 < W_2 < \ldots < W_N$. Let ν be the smallest integer such that

$$(4.1) \qquad \gamma_\nu = \frac{1}{2}(W_{N-\nu} + W_{M-\nu}) - \frac{1}{M-1}\left(\sum_{j=1}^{M-\nu-1} W_j + \sum_{j=N-\nu+1}^{N} W_j\right)$$

is < 0. [This function γ_ν (4.1), like δ_ν (2.2), decreases as ν increases.] Now define the two pairings

$$(4.2a) \qquad \begin{aligned} W^+_{(i)} &= W_i \,, & 1 \le i \le M-\nu \\ &= W_{N-M+i} \,, & M-\nu+1 \le i \le M \end{aligned}$$

and

$$(4.2b) \qquad \begin{aligned} W^-_{(i)} &= W_{N+1-i} \,, & 1 \le i \le \nu \\ &= W_{M+1-i} \,, & \nu+1 \le i \le M \end{aligned} \quad .$$

[As in the case of the previous test, it is permissible to choose ν by an objective method different from the one just recommended. No ν is needed if $M=N$.]

Only one of the two pairings (4.2) will be used. To find out which one it is to be, we compute the correlation coefficients between the X_i's and $W^+_{(i)}$'s and between the X_i's and $W^-_{(i)}$'s. We select whichever pairing leads to the correlation coefficient with larger absolute value. (Note that the two correlation coefficients have the same denominator.) We use $W_{(i)}$ to denote the pairing which is chosen (i.e., either $W^+_{(i)}$ or $W^-_{(i)}$, whichever it turns out to be), and the correlation coefficient is then

$$(4.3) \qquad \rho = \frac{\sum_{i=1}^{M} X_i W_{(i)} - M \bar{X} \bar{W}^{o}}{\sqrt{\sum_{i=1}^{M} X_i^2 - M \bar{X}^2} \sqrt{\sum_{i=1}^{M} W_{(i)}^2 - M \bar{W}^{o^2}}}$$

[In this section, we define all means the same as in (2.3).]

Some future formulas will contain $(\mp)$ and $(\pm)$ signs. The upper sign is to be chosen in all places if pairing (4.2a) was selected, and the lower sign if pairing (4.2b) was selected.

We define

$$(4.4) \qquad M \sigma_X^2 = \sum_{i=1}^{M} X_i^2 - M \bar{X}^2 \quad ,$$

$$N \sigma_W^2 = \sum_{j=1}^{N} W_j^2 - N \bar{W}^2 \quad ,$$

and

$$N \sigma_{W^o}^2 = \sum_{i=1}^{M} W_{(i)}^2 - M \bar{W}^{o^2} \quad ,$$

and calculate

$$(4.5) \qquad R = \frac{\sigma_W}{\sigma_{W^o}} = \sqrt{\frac{N \sigma_W^2}{N \sigma_{W^o}^2}} \quad .$$

At this stage, we specify two cases: Case I. if $|\rho| \leq 1/R$, and Case II. if $|\rho| > 1/R$.

<u>Case I.</u>: $|\rho| \leq 1/R$. We calculate

$$(4.6) \qquad t = \frac{\hat{\beta}_Z - \hat{\beta}_Y}{\sqrt{\psi(R,\rho)} \sqrt{\frac{S_e^2}{M-3}}} \quad ,$$

where

(4.7) $$\psi(R,\rho) = \frac{1 + R^2 \mp 2\rho R}{1-\rho^2} = \frac{(R-1)^2}{2(1 \mp \rho)} + \frac{(R+1)^2}{2(1 \pm \rho)},$$

(4.8) $$\hat{\beta}_Y = \frac{(u_Y' T_B) \pm \rho R(u_Z' T_B)}{1-\rho^2},$$

(4.9) $$\hat{\beta}_Z = \frac{\pm \rho R(u_Y' T_B) + R^2(u_Z' T_B)}{1-\rho^2},$$

(4.10) $$S_e^2 = \frac{1}{M\sigma_X^2}\left(\sum_{i=1}^{M} Y_i^2 - M\,\overline{Y}^2\right) \mp \frac{2}{\sqrt{M\sigma_X^2}\sqrt{N\sigma_W^2}}\left(\sum_{i=1}^{M} Y_i Z_{(i)} - M\,\overline{Y}\,\overline{Z}^{o}\right)$$

$$+ \frac{1}{N\sigma_W^2}\left(\sum_{i=1}^{M} Z_{(i)}^2 - M\,\overline{Z}^{o^2}\right) - \hat{\beta}_Y(u_Y' T_B) - \hat{\beta}_Z(u_Z' T_B),$$

(4.11) $$u_Y' T_B = \frac{1}{M\sigma_X^2}\left(\sum_{i=1}^{M} X_i Y_i - M\,\overline{X}\,\overline{Y}\right) \mp \frac{1}{\sqrt{M\sigma_X^2}\sqrt{N\sigma_W^2}}\left(\sum_{i=1}^{M} X_i Z_{(i)} - M\,\overline{X}\,\overline{Z}^{o}\right),$$

and

(4.12) $$u_Z' T_B = \mp \frac{1}{\sqrt{M\sigma_X^2}\sqrt{N\sigma_W^2}}\left(\sum_{i=1}^{M} Y_i W_{(i)} - M\,\overline{Y}\,\overline{W}^{o}\right) + \frac{1}{N\sigma_W^2}\left(\sum_{i=1}^{M} W_{(i)} Z_{(i)} - M\,\overline{W}^{o}\overline{Z}^{o}\right).$$

<u>Case II.</u> $|\rho| > 1/R$. We calculate

(4.13) $$t = \frac{\hat{\beta}_Z - \hat{\beta}_Y}{\frac{1}{|\rho|}\sqrt{\frac{S_e^2}{M-3}}},$$

where

$$(4.14) \qquad \hat{\beta}_Y = \frac{(u_Y' T_B) + (u_Z' T_B)}{1 - \rho^2} ,$$

$$(4.15) \qquad \hat{\beta}_Z = \frac{(u_Y' T_B) + (1/\rho^2)(u_Z' T_B)}{1 - \rho^2} ,$$

$$(4.16) \qquad s_e^2 = \frac{1}{M\,\sigma_X^2}\left(\sum_{i=1}^{M} Y_i^2 - M\,\bar{Y}^2\right) - \frac{2\rho}{\sqrt{M\,\sigma_X^2}\sqrt{N\,\sigma_{W^o}^2}}\left(\sum_{i=1}^{M} Y_i Z_{(i)} - M\,\bar{Y}\,\bar{Z}^o\right)$$

$$+ \frac{\rho^2}{N\,\sigma_{W^o}^2}\left(\sum_{i=1}^{M} Z_{(i)}^2 - M\,\bar{Z}^{o^2}\right) - \hat{\beta}_Y(u_Y' T_B) - \hat{\beta}_Z(u_Z' T_B) ,$$

$$(4.17) \qquad u_Y' T_B \doteq \frac{1}{M\,\sigma_X^2}\left(\sum_{i=1}^{M} X_i Y_i - M\,\bar{X}\,\bar{Y}\right) - \frac{\rho}{\sqrt{M\,\sigma_X^2}\sqrt{N\,\sigma_{W^o}^2}}\left(\sum_{i=1}^{M} X_i\, Z_{(i)} - M\,\bar{X}\,\bar{Z}^o\right),$$

and

$$(4.18) \qquad u_Z' T_B = -\frac{\rho}{\sqrt{M\,\sigma_X^2}\sqrt{N\,\sigma_{W^o}^2}}\left(\sum_{i=1}^{M} Y_i W_{(i)} - M\,\bar{Y}\,\bar{W}^o\right) + \frac{\rho^2}{N\,\sigma_{W^o}^2}\left(\sum_{i=1}^{M} W_{(i)} Z_{(i)} - M\bar{W}^o\bar{Z}^o\right).$$

Note also that (4.14) and (4.15) imply the formula

$$(4.19) \qquad \hat{\beta}_Z - \hat{\beta}_Y = (1/\rho^2)\,(u_Z' T_B) .$$

To test the hypothesis $\beta_Y = \beta_Z$, the statistic t (4.6 or 4.13) is referred to tables of the t-distribution with $(M-3)$ degrees of freedom.

5. Numerical examples of the test for Problem B. We present two examples, one for Case I, and the other for Case II. For the first example, suppose the first group contains the $M=6$ observations

$$Y_1 = 0.7 \quad Y_2 = 2.4 \quad Y_3 = 1.9 \quad Y_4 = 2.4 \quad Y_5 = 4.2 \quad Y_6 = 4.5$$
$$X_1 = 0 \quad X_2 = 2 \quad X_3 = 4 \quad X_4 = 6 \quad X_5 = 13 \quad X_6 = 17 \quad ,$$

and suppose the second group contains the $N=7$ observations

$$Z_1 = 3.2 \quad Z_2 = 5.0 \quad Z_3 = 8.5 \quad Z_4 = 10.6 \quad Z_5 = 15.7 \quad Z_6 = 20.6 \quad Z_7 = 25.5$$
$$W_1 = 0 \quad W_2 = 1 \quad W_3 = 2 \quad W_4 = 3 \quad W_5 = 5 \quad W_6 = 7 \quad W_7 = 9 \quad .$$

Using (4.1), we find that

$$\gamma_2 = 4 - 3.8 \quad > \quad 0$$

and

$$\gamma_3 = 2.5 - 4.4 < 0 \quad ,$$

so that $\nu = 3$ is the proper ν-value to use in (4.2). Thus we obtain

$$W^+_{(1)} = 0 \quad W^+_{(2)} = 1 \quad W^+_{(3)} = 2 \quad W^+_{(4)} = 5 \quad W^+_{(5)} = 7 \quad W^+_{(6)} = 9$$

and

$$W^-_{(1)} = 9 \quad W^-_{(2)} = 7 \quad W^-_{(3)} = 5 \quad W^-_{(4)} = 2 \quad W^-_{(5)} = 1 \quad W^-_{(6)} = 0$$

from (4.2).

As for the means (2.3), we will need the values

$$M\,\overline{X} = 42, \quad M\,\overline{Y} = 16.1, \quad N\,\overline{W} = 27, \quad M\,\overline{W}^o = 24, \quad M\,\overline{Z}^o = 78.5 \quad .$$

Now we find

$$\Sigma\, X_i\, W^+_{(i)} - M\,\overline{X}\,\overline{W}^o = 284 - 168 = 116$$

and

$$\Sigma\, X_i\, W^-_{(i)} - M\,\overline{X}\,\overline{W}^o = 59 - 168 = -109 \quad .$$

Since $|116| > |-109|$, we take $W_{(i)} = W^+_{(i)}$, so that

$$Z_{(1)} = 3.2 \quad Z_{(2)} = 5.0 \quad Z_{(3)} = 8.5 \quad Z_{(4)} = 15.7 \quad Z_{(5)} = 20.6 \quad Z_{(6)} = 25.5$$

$$W_{(1)} = 0 \quad W_{(2)} = 1 \quad W_{(3)} = 2 \quad W_{(4)} = 5 \quad W_{(5)} = 7 \quad W_{(6)} = 9 \ .$$

The sums of squares and cross-products which we will need are

$$\Sigma Y_i^2 = 53.51 \ , \quad \Sigma X_i^2 = 514 \ , \quad \Sigma Z_{(i)}^2 = 1428.59 \ , \quad \Sigma W_{(i)}^2 = 160 \ ,$$

$$\Sigma X_i Y_i = 157.9 \ , \ \Sigma Y_i Z_{(i)} = 269.34, \quad \Sigma Y_i W_{(i)} = 88.1 \ ,$$

$$\Sigma X_i Z_{(i)} = 839.5, \quad \Sigma X_i W_{(i)} \doteq 284, \quad \Sigma W_{(i)} Z_{(i)} = 474.2 \ ,$$

and $$\sum_{j=1}^{N} W_j^2 = 169 \ .$$

Thus we obtain

$$M\,\sigma_X^2 = 220 \ , \quad N\,\sigma_W^2 = 64.857 \ , \quad N\,\sigma_{W^o}^2 = 64 \ ;$$

$$\rho = \frac{116}{\sqrt{220 \times 64}} = \sqrt{.955682} = .977590 \ ;$$

and

$$R = \sqrt{\frac{64.857}{64}} = \sqrt{1.013393} = 1.006674$$

from (4.4), (4.3), and (4.5). Since $R|\rho| = .984114 < 1$, this example falls within Case I.

Wherever there are two signs in (4.7 - 4.12), we choose the upper sign (since the upper signs are for $W^+_{(i)}$ and the lower signs for $W^-_{(i)}$). Thus (4.11 - 4.12) become

$$u_Y' T_B = \frac{1}{220}(45.2) - \frac{1}{\sqrt{220 \times 64.857}}(290.0) = -2.222317$$

and

$$u_Z' T_B = -\frac{1}{\sqrt{220 \times 64.857}}(23.7) + \frac{1}{64.857}(160.2) = 2.271636 \ ,$$

and so (4.8 - 4.10) are

$$\hat{\beta}_Y = \frac{(-2.222317) + .984114\ (2.271636)}{1 - .955682} = .29856 \quad ,$$

$$\hat{\beta}_Z = \frac{.984114\ (-2.222317) + 1.013393\ (2.271636)}{1 - .955682} = 2.5959 \quad ,$$

and

$$s_e^2 = \frac{1}{220}\,(10.3083) - \frac{2}{\sqrt{220 \times 64.857}}\,(58.6983)$$

$$+ \frac{1}{64.857}\,(401.5483) - .29856\ (-2.222317) - 2.5959\ (2.271636)$$

$$= .021884 \quad .$$

Finally, (4.7) is

$$\psi(R,\rho) = \frac{(.006674)^2}{2(1-.97759)} + \frac{(2.006674)^2}{2(1+.97759)} = 1.0191 \quad ,$$

and thus the test statistic (4.6) is

$$t = \frac{2.5959 - .2986}{\sqrt{1.0191}\,\sqrt{\frac{.021884}{6-3}}} = \frac{2.297}{.08622} = 26.64 \quad .$$

This t has $M-3=3$ degrees of freedom. Hence the critical region for a two-sided test at the $\alpha = .05$ level is $|t| > 3.18$. Since $|26.64| > 3.18$, we reject the null hypothesis $\beta_Y = \beta_Z$. A 95°/o confidence interval on $(\beta_Z - \beta_Y)$ is given by

$$2.30 \pm 3.18\ (.0862) \quad ,$$

i.e.,

$$2.03 \leq (\beta_Z - \beta_Y) \leq 2.57 \quad .$$

For our second example, we will use exactly the same figures as for the first example, except that we will suppose that $X_4 = 9$ instead of 6. Since the W_j's are the same as before, ν will also be the same as before, and therefore the $W^+_{(i)}$'s and $W^-_{(i)}$'s will likewise be as before. The means will stay the same except that $M\,\overline{X} = 45$. We find

$$\Sigma X_i W^+_{(i)} - M\,\overline{X}\,\overline{W}^o = 299 - 180 = 119$$

and

$$\Sigma X_i W^-_{(i)} - M\,\overline{X}\,\overline{W}^o = 65 - 180 = -115 \quad .$$

Thus we take $W_{(i)} = W^+_{(i)}$ (since $|119| > |-115|$); this means that the $W_{(i)}$'s and $Z_{(i)}$'s are exactly as before.

The sums of squares and cross-products will be the same as before except for

$$\Sigma X_i^2 = 559,\ \Sigma X_iY_i = 165.1,\ \ \Sigma X_iZ_{(i)} = 886.6,\ \ \Sigma X_iW_{(i)} = 299 \quad .$$

Hence $M\,\sigma^2_X = 221.5$ now, but $N\,\sigma^2_W$, $N\,\sigma^2_{W^o}$, and R are unaltered. We obtain

$$\rho = \frac{119}{\sqrt{221.5 \times 64}} = \sqrt{.998942} = .999471 \quad .$$

This time, $R|\rho| = 1.006141 > 1$, which means that we are thrown into Case II.

It remains to calculate (4.17 - 4.18), (4.14 - 4.16), and (4.13). We find

$$u'_Y\,T_B = \frac{1}{221.5}(44.35) - \frac{.999471}{\sqrt{221.5 \times 64}}(297.85) = -2.300067 \quad ,$$

$$u'_Z\,T_B = -\frac{.999471}{\sqrt{221.5 \times 64}}(23.7) + \frac{.998942}{64}(160.2) = 2.301528 \quad ,$$

$$\hat{\beta}_Y = \frac{(-2.300067) + (2.301528)}{1 - .998942} = 1.3809 \quad ,$$

$$\hat{\beta}_Z = \frac{(-2.300067) + (1/.998942)\,(2.301528)}{1 - .998942} = 3.6853 \quad ,$$

$$S_e^2 = \frac{1}{221.5}(10.3083) - \frac{2 \times .999471}{\sqrt{221.5 \times 64}}(58.6983)$$

$$+ \frac{.998942}{64}(401.5483) - 1.3809(-2.300067) - 3.6853(2.301528)$$

$$= .022953 \quad ,$$

and finally ,

$$t = \frac{3.6853 - 1.3809}{\frac{1}{.999471}\sqrt{\frac{.022953}{6-3}}} = \frac{2.304}{.08752} = 26.33 \quad .$$

Since $|26.33| > 3.18$, we reject the null hypothesis $\beta_Y = \beta_Z$. A $95^o/o$ confidence interval for $(\beta_Z - \beta_Y)$ is given by

$$2.30 \pm 3.18\,(.0875) \quad ,$$

i.e.,

$$2.03 \leq (\beta_Z - \beta_Y) \leq 2.58 \quad .$$

REFERENCES

[1] Potthoff, Richard F., "Illustration of a Technique Which Tests Whether Two Regression Lines Are Parallel When the Variances Are Unequal", Mimeo Series No. 320, Department of Statistics, University of North Carolina, 1962.

[2] Potthoff, Richard F., "Illustration of A Test Which Compares Two Parallel Regression Lines When the Variances Are Unequal", Mimeo Series No. 323, Department of Statistics, University of North Carolina, 1962.

[3] Potthoff, Richard F., "Some Scheffé -Type Tests for Some Behrens-Fisher-Type Regression Problems", Mimeo Series No. 374, Department of Statistics, University of North Carolina, 1963.

[4] Scheffé, Henry, "On Solutions of the Behrens-Fisher Problem, Based on the t-Distribution", Annals of Mathematical Statistics, volume 14 (1943), pp. 35-44.