

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

ILLUSTRATION OF A TEST WHICH COMPARES TWO PARALLEL REGRESSION
LINES WHEN THE VARIANCES ARE UNEQUAL

by

Richard F. Potthoff

April 1962

Contract No. AF 49(638)-213

In a situation where two regression lines are known to be parallel, it may be desired to test the hypothesis that the two lines are identical without assuming that the variances of the two sets of error terms are necessarily equal. This paper presents a relatively non-technical discussion of a test which can be used for this problem. The test statistic is analogous to the well-known Wilcoxon statistic. The obtaining of the test statistic involves a large number of routine calculations, and (in general) a computer is needed for this. This paper is intended for the practitioner rather than for the theoretician; the more technical aspects of the test are covered in a separate paper.

This research was partially supported by the Air Force Office of Scientific Research and partially supported by Educational Testing Service.

Institute of Statistics
Mimeo Series No. 323

ILLUSTRATION OF A TEST WHICH COMPARES TWO PARALLEL REGRESSION LINES WHEN THE
VARIANCES ARE UNEQUAL¹

by Richard F. Potthoff

1. Introduction. It is sometimes desired to make certain comparisons of two regression lines, when it is assumed that the two sets of error terms are normally distributed but with (possibly) unequal variances. An earlier paper [1] presented a method of testing the hypothesis that the two regression lines are parallel, under this condition of unequal variances. In the present paper it is assumed that the two regression lines are parallel (with the two variances being unequal), and a method is presented for testing the hypothesis that the two regression lines are identical. The test statistic to be used in this paper, like the test statistic presented in [1], is analogous to the Wilcoxon statistic.

We suppose that we have m pairs $(Y_1, X_1), (Y_2, X_2), \dots, (Y_m, X_m)$, such that, for each i , Y_i observes the relation

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

where α_Y and β are unknown parameters (regression coefficients), the X_i 's are specified constants, and the e_i 's are normal and independent with mean 0 and unknown variance σ_e^2 . We suppose also that we have n pairs $(Z_1, W_1), (Z_2, W_2), \dots, (Z_n, W_n)$, such that, for each j , Z_j observes the relation

$$(1.2) \quad Z_j = \alpha_Z + \beta W_j + f_j \quad ,$$

where α_Z is an unknown parameter, β is the same thing as in (1.1) (this is where the assumption that the two regression lines are parallel enters in), the W_j 's are specified constants, and the f_j 's are normal and independent with mean 0 and unknown variance σ_f^2 . We will present a test of the hypothesis that the two regression lines associated with (1.1) and (1.2) are identical: that is, we will

¹ This research was supported in part by Educational Testing Service, and was supported in part by the Air Force Office of Scientific Research.

test the hypothesis that $\alpha_Y = \alpha_Z$. We of course must have a test which will be valid regardless of what the values of σ_e^2 and σ_f^2 are.

This paper will give a numerical example, which will illustrate not only how the test of the hypothesis $\alpha_Y = \alpha_Z$ works, but will also illustrate how to obtain confidence bounds on $(\alpha_Z - \alpha_Y)$. Except when m and n are very small, the computations required to obtain our test statistic, although quite routine, will be prohibitively lengthy to be performed on a desk calculator; consequently, the numerical example given in this paper is for $m = 5$ and $n = 5$. In general, it appears that it will be necessary to utilize a high-speed computer in order to calculate the test statistic we are presenting here. Section 5 discusses briefly the problem of performing the lengthy calculations on a computer, and suggests some possible techniques for circumventing some of these calculations.

An experiment comparing two curriculums provides an illustration of a practical situation in which the problem treated by this paper could arise. This situation is discussed in some detail in [1]. We have m classes which receive curriculum # 1 and n classes which receive curriculum # 2. Y_i and X_i represent respectively the achievement measure (obtained after the course) and ability measure (obtained before the course) for the i -th class receiving curriculum # 1. Similarly, Z_j and W_j represent respectively the achievement measure and ability measure for the j -th class receiving curriculum # 2. The problem discussed in [1] was that of testing the hypothesis that the regression lines associated with the two curriculums are parallel (when the variances may be unequal). In the present paper, however, it is assumed that the two regression lines are parallel (1.1 - 1.2), and the problem is to test whether they are identical. (If they are identical, then there is no difference in the effects of the two curriculums.)

In this paper the discussion will be relatively non-technical. A separate paper covers the technical aspects of the topic of the present paper.

2. The test. This section presents the formula for the test statistic. There is a certain similarity to the test statistic of [1], but the present statistic is much more difficult to compute.

For every quadruple (i, I, j, J) such that

$$(2.1) \quad X_i < X_I \quad , \quad W_j < W_J$$

and

$$(2.2) \quad X_i \leq W_J \quad , \quad W_j \leq X_I \quad ,$$

let us define

$$(2.3) \quad V_{iIjJ} = \frac{(X_I - W_j)(Z_J - Y_i) - (W_J - X_i)(Y_I - Z_j)}{X_I + W_J - X_i - W_j} .$$

Each V_{iIjJ} will be normally distributed with expectation $(\alpha_Z - \alpha_Y)$, and hence will have median $(\alpha_Z - \alpha_Y)$. Thus we would expect (on the average) that half of the V_{iIjJ} 's would be positive and half negative, if and only if the null hypothesis $\alpha_Y = \alpha_Z$ is true.

Let S be the number of V_{iIjJ} 's which are positive, and let w be the proportion of them which are positive. Then

$$(2.4) \quad w = \frac{S}{T} \quad ,$$

where T is the total number of quadruples (i, I, j, J) which satisfy (2.1 - 2.2). If the null hypothesis $\alpha_Y = \alpha_Z$ is true, then w will have expected value $\frac{1}{2}$ and will be approximately normally distributed. The variance of w when the hypothesis $\alpha_Y = \alpha_Z$ is true will depend on σ_e^2 and σ_f^2 , but a number Q can be obtained such that

$$\text{var}(w) \leq Q$$

no matter what the values of σ_e^2 and σ_f^2 are. This number Q , which is the least upper bound of $\text{var}(w)$, is calculated by a formula to be given below.

The test of the hypothesis $\alpha_Y = \alpha_Z$ is as follows: if we wish (e.g.) to use a two-tailed test at the 5 o/o level, we reject if

$$(2.5) \quad \frac{|w - \frac{1}{2}|}{\sqrt{Q}} > 1.96$$

and accept otherwise. This test (2.5) will be conservative in the sense that the probability of rejecting the null hypothesis when it is true will generally be somewhat less than 5 o/o rather than exactly equal to 5 o/o; the reason for this is that the actual value of $\text{var}(w)$ (which is unknown) will generally be less than Q .

Although w (2.4) is relatively simple to calculate, the computation of Q is very lengthy. The formula for Q is

$$(2.6) \quad Q = \max(Q_1, Q_2)$$

(i.e., Q is the larger of the two quantities Q_1 and Q_2), where

$$(2.7) \quad Q_1 = \frac{1}{4T} + \frac{1}{\pi T^2} \sum_{\substack{iIjJ > \\ i'I'j'J'}} \sin^{-1} \rho_{iIjJ, i'I'j'J'}^{(1)}$$

and

$$(2.8) \quad Q_2 = \frac{1}{4T} + \frac{1}{\pi T^2} \sum_{\substack{iIjJ > \\ i'I'j'J'}} \sin^{-1} \rho_{iIjJ, i'I'j'J'}^{(2)}$$

the ρ 's being defined by

$$(2.9) \quad \rho_{iIjJ, i'I'j'J'}^{(1)} = \frac{\delta_{ii'}(X_I - W_j)(X_{I'} - W_{j'}) + \delta_{iI'}(X_I - W_j)(W_{J'} - X_{I'}) + \delta_{Ii'}(W_J - X_i)(X_{I'} - W_{j'}) + \delta_{II'}(W_J - X_i)(W_{J'} - X_{I'})}{\sqrt{(X_I - W_j)^2 + (W_J - X_i)^2} \sqrt{(X_{I'} - W_{j'})^2 + (W_{J'} - X_{I'})^2}}$$

and

$$(2.10) \quad \rho_{iIjJ,i'I'j'J'}^{(2)} = \frac{\delta_{jj'}(W_J - X_I)(W_{J'} - X_{I'}) + \delta_{jJ'}(W_J - X_I)(X_{I'} - W_{j'}) + \delta_{JJ'}(X_I - W_j)(W_{J'} - X_{I'}) + \delta_{JJ'}(X_I - W_j)(X_{I'} - W_{j'})}{\sqrt{(X_I - W_j)^2 + (W_J - X_I)^2} \sqrt{(X_{I'} - W_{j'})^2 + (W_{J'} - X_{I'})^2}}$$

The δ symbols in (2.9) and (2.10) are Kronecker deltas; for example, $\delta_{ii'}$ is equal to 0 if $i \neq i'$ and is equal to 1 if $i = i'$. The summations in (2.7) and (2.8) extend over all pairs of quadruples (i, I, j, J) and (i', I', j', J') such that both quadruples satisfy (2.1 - 2.2) and such that

$$(2.11) \quad (i, I, j, J) > (i', I', j', J')$$

[We assume that the T quadruples satisfying (2.1 - 2.2) are arranged arbitrarily in some kind of order; it is in this sense that the relation (2.11) is to be interpreted.]

Note that the formula for Q does not depend on the Y_i 's or the Z_j 's. In other words, Q can generally be calculated before the experiment is completed, if desired.

The reason for assuming normality of the e_i 's and f_j 's in using the test (2.5) is that this normality assumption was used in proving that Q is the least upper bound of $\text{var}(w)$. It is not known how the test (2.5) would be affected by non-normality, but it is possible that there might be certain non-normal symmetrical distributions of the e_i 's and f_j 's for which the test would still be valid.

Potentially, there are $\frac{1}{2} m(m-1) \times \frac{1}{2} n(n-1)$ quadruples (i, I, j, J) which will satisfy (2.1). However, there will be fewer quadruples than this satisfying (2.1) if any of the X_i 's are the same, or if any of the W_j 's are tied. If desired, it would be permissible to assign (by using random numbers) an arbitrary ranking to any set of tied X_i 's or of tied W_j 's and thereby bring the number of quadruples satisfying (2.1) up to its full potential number. [However, this will not be done in the numerical illustration to be given in Section 3 (in which there are

tied X_i 's and tied W_j 's), due to the fact that we are attempting to hold down the length of this illustration in order for it to be readily understandable.]

The question might be raised as to why the statistic w is not based on all quadruples which satisfy (2.1), rather than just on those which satisfy (2.2) as well as (2.1). One reason is that the proof which establishes the least upper bound for $\text{var}(w)$ would not work if quadruples satisfying (2.1) but not (2.2) were to enter in to the determination of w and so a new technique for proving an upper bound for $\text{var}(w)$ would somehow have to be worked out]. In addition to this, however, a couple of other possible advantages might result from excluding quadruples which do not satisfy (2.2).

(i) One effect of the requirement (2.2) is to exclude all V_{iIjJ} 's whose variances exceed a certain value. Since the test statistic is thus based on the V_{iIjJ} 's with the smallest variances, it is conjectured that the exclusion of some of the V_{iIjJ} 's via (2.2) does not necessarily result in a test with lower power, and that a test based on all $\binom{m}{2}\binom{n}{2}$ potentially possible V_{iIjJ} 's (if such a test could be devised) would have worse rather than better power in some circumstances. [On the other hand, the exclusion of some V_{iIjJ} 's via (2.2) can unquestionably be a hindrance if such exclusion results in too large a proportion of the V_{iIjJ} 's being eliminated.] (ii) An obvious effect of the exclusions based on (2.2) is to make the computations a bit less lengthy. (iii) Another effect of the restriction (2.2) is to exclude those quadruples for which the intervals (X_i, X_I) and (W_j, W_J) have no points in common; however, it is not known that this effect, as such, results in any advantage.

It thus appears that, although a test based on all V_{iIjJ} 's satisfying (2.1) (if it could be devised) would be a more desirable test in some situations than the one utilizing the restriction (2.2), it could be less desirable in other situations.

3. Numerical example. In this section we give a numerical example to illustrate the computations for the test (2.5). Because the computations are quite lengthy, the sample sizes in the example will be very small: $m = 5$ and $n = 5$. For purposes of the illustration we will go ahead and use the normal approximation for the distribution of w , even though actually the sample sizes are too small and $\text{var}(w)$ and Q are too large for the normal approximation to be suitable. We will counteract this difficulty to some extent by testing at the 10 o/o rather than the 5 o/o level.

Suppose that our two samples are

$$Y_1 = 4.42, \quad Y_2 = 27.59, \quad Y_3 = 30.78, \quad Y_4 = 32.65, \quad Y_5 = 69.36$$

$$X_1 = 0, \quad X_2 = 4, \quad X_3 = 4, \quad X_4 = 4, \quad X_5 = 9$$

and

$$Z_1 = 9.04, \quad Z_2 = 35.97, \quad Z_3 = 38.42, \quad Z_4 = 38.81, \quad Z_5 = 64.42$$

$$W_1 = 1, \quad W_2 = 5, \quad W_3 = 5, \quad W_4 = 5, \quad W_5 = 9.$$

Merely for the sake of orderliness, the first sample is arranged so that it is in order of increasing magnitude of the X_i 's, and the second sample is in order of increasing magnitude of the W_j 's.

Let us define v_{iIjJ} to be the numerator of the formula for V_{iIjJ} (2.3):

$$(3.1) \quad v_{iIjJ} = (X_i - W_j)(Z_j - Y_i) - (W_j - X_i)(Y_i - Z_j).$$

In all cases v_{iIjJ} will have the same sign as V_{iIjJ} . Hence we can determine S in (2.4) by counting the number of v_{iIjJ} 's which are positive, rather than by calculating all the V_{iIjJ} 's and counting the number of them which are positive.

Thus, if it is desired only to test the hypothesis $\alpha_Y = \alpha_Z$, it will not be necessary to calculate the V_{iIjJ} 's, but rather it will suffice to calculate only the v_{iIjJ} 's; On the other hand, if confidence bounds on $(\alpha_Z - \alpha_Y)$ are desired, then it will be necessary to have the V_{iIjJ} 's. Section 4 will illustrate how to obtain confidence bounds.

For the samples above with $m = 5$ and $n = 5$, there are potentially $\frac{1}{2} m(m-1) \times \frac{1}{2} n(n-1) = \frac{1}{2} \cdot 5(5-1) \times \frac{1}{2} \cdot 5(5-1) = 100$ different quadruples (i, I, j, J) which satisfy (2.1). Of these 100 quadruples, some satisfy both (2.1) and (2.2) but some do not. For example, consider the quadruple $(1, 4, 2, 5)$. We have

$$X_i = X_1 = 0, \quad X_I = X_4 = 4, \quad W_j = W_2 = 5, \quad W_J = W_5 = 9$$

Since $W_j \leq X_I$ does not hold, this quadruple fails to satisfy (2.2). On the other hand, the quadruple $(3, 5, 1, 3)$ does satisfy (2.1) and (2.2), since

$$X_i = X_3 = 4, \quad X_I = X_5 = 9, \quad W_j = W_1 = 1, \quad W_J = W_3 = 5$$

and so $X_i \leq W_J$ and $W_j \leq X_I$ are both satisfied [Note: As indicated earlier, we are excluding some quadruples because tied X's or tied W's prevent (2.1) from being satisfied. Some quadruples, such as $(2, 4, 1, 5)$, will satisfy (2.2) but will be excluded because a tie prevents (2.1) from being satisfied.]

Altogether it turns out that 40 of the 100 quadruples satisfy both (2.1) and (2.2), while 60 do not. Thus $T = 40$. [In general, one would suspect that, the higher T is (for a given m and n), the more powerful the test will be; in this example, the power of the test is hampered by the relatively small value of T .]

We now indicate what the 40 quadruples are which satisfy both (2.1) and (2.2) and we present the value of v_{iIjJ} for each quadruple (the 40 quadruples are the subscripts on the 40 v_{iIjJ} 's):

$$v_{1212} = 1.90, \quad v_{1213} = 9.25, \quad v_{1214} = 10.42, \quad v_{1215} = 13.05,$$

$$v_{1312} = -14.05, \quad v_{1313} = -6.70, \quad v_{1314} = -5.53, \quad v_{1315} = -15.66,$$

$$v_{1412} = -23.40, \quad v_{1413} = -16.05, \quad v_{1414} = -14.88, \quad v_{1415} = -32.49,$$

$$v_{1512} = -49.20, \quad v_{1513} = -29.60, \quad v_{1514} = -26.48$$

$$v_{1525} = -60.51, \quad v_{1535} = -38.46, \quad v_{1545} = -34.95, \quad v_{1515} = -62.88,$$

$$v_{2512} = 6.72, \quad v_{2513} = 26.32, \quad v_{2514} = 29.44,$$

$$v_{2525} = -19.63, \quad v_{2535} = -7.38, \quad v_{2545} = -5.43, \quad v_{2515} = -6.96,$$

$$\begin{aligned}
v_{3512} &= -18.80, & v_{3513} &= .80, & v_{3514} &= 3.92, \\
v_{3525} &= -32.39, & v_{3535} &= -20.14, & v_{3545} &= -18.19, & v_{3515} &= -32.48, \\
v_{4512} &= -33.76, & v_{4513} &= -14.16, & v_{4514} &= -11.04, \\
v_{4525} &= -39.87, & v_{4535} &= -27.62, & v_{4545} &= -25.67, & v_{4515} &= -47.44.
\end{aligned}$$

A couple of examples will suffice to demonstrate the calculation of the v_{iIjJ} 's:

$$\begin{aligned}
v_{2513} &= (X_5 - W_1)(Z_3 - Y_2) - (W_3 - X_2)(Y_5 - Z_1) \\
&= (9 - 1)(38.42 - 27.59) - (5 - 4)(69.36 - 9.04) \\
&= 8(10.83) - 1(60.32) & & = 26.32 \quad ,
\end{aligned}$$

and

$$\begin{aligned}
v_{1312} &= (X_3 - W_1)(Z_2 - Y_1) - (W_2 - X_1)(Y_3 - Z_1) \\
&= (4 - 1)(35.97 - 4.42) - (5 - 0)(30.78 - 9.04) = -14.05 \quad .
\end{aligned}$$

We find that, of the 40 v_{iIjJ} 's, 9 are positive and 31 negative. Thus

$S = 9$, and, by (2.4),

$$(3.2) \quad w = \frac{S}{T} = \frac{9}{40} = .225 \quad .$$

(Note: If one or more of the v_{iIjJ} 's had been exactly 0, we could have handled such a situation by counting $\frac{1}{2}$ for each such v_{iIjJ} .)

Computing Q is much less simple than computing w . In obtaining Q_1 and Q_2 , it is convenient to utilize the well-known trigonometric formula

$$(3.3) \quad 2 \sin^{-1} \rho = \cos^{-1}(1 - 2\rho^2)$$

in order to obviate the need for obtaining square roots. Table I. contains the value of

$$2 \sin^{-1} \rho_{iIjJ, i'I'j'J'}^{(1)}$$

to two decimal places, as computed by the formula (3.3), for every octuple $(i, I, j, J; i', I', j', J')$ embraced by the summation appearing on the right-hand side of (2.7). Similarly, Table II. contains the value of

$$2 \sin^{-1} \rho_{iIjJ, i'I'j'J'}^{(2)}$$

for every octuple needed for (2.8).

A few examples will demonstrate how the formulas (2.7), (2.8), and (3.3) are utilized to obtain the elements of Tables I. and II.:

$$\begin{aligned}
 2\sin^{-1} \rho_{4512,3525}^{(1)} &= \cos^{-1} \left[1 - 2 \frac{(W_2 - X_4)^2 (W_5 - X_3)^2}{\{(X_5 - W_1)^2 + (W_2 - X_4)^2\} \{(X_5 - W_2)^2 + (W_5 - X_3)^2\}} \right] \\
 &= \cos^{-1} \left[1 - 2 \times \frac{(5-4)^2 (9-4)^2}{\{(9-1)^2 + (5-4)^2\} \{(9-5)^2 + (9-4)^2\}} \right] \\
 &= \cos^{-1} \left[1 - \frac{50}{2665} \right] = \cos^{-1} .9812 = .1942
 \end{aligned}$$

$$\begin{aligned}
 2\sin^{-1} \rho_{4512,3525}^{(2)} &= \cos^{-1} \left[1 - 2 \frac{(X_5 - W_1)^2 (W_5 - X_3)^2}{\{(X_5 - W_1)^2 + (W_2 - X_4)^2\} \{(X_5 - W_2)^2 + (W_5 - X_3)^2\}} \right] \\
 &= \cos^{-1} \left[1 - \frac{3200}{2665} \right] = \cos^{-1} (-.2008) = 1.7730
 \end{aligned}$$

$$2\sin^{-1} \rho_{3545,1412}^{(1)} = 2\sin^{-1} 0 = 0 \quad \text{since } 3 \neq 1, 3 \neq 4, 5 \neq 1, 5 \neq 4$$

$$2\sin^{-1} \rho_{1412,3545}^{(2)} = 2\sin^{-1} 0 = 0 \quad \text{since } 1 \neq 4, 1 \neq 5, 2 \neq 4, 2 \neq 5$$

$$\begin{aligned}
 2\sin^{-1} \rho_{2535,2514}^{(1)} &= \cos^{-1} \left[1 - 2 \times \frac{\{(X_5 - W_3)(X_5 - W_1) + (W_5 - X_2)(W_4 - X_2)\}^2}{\{(X_5 - W_3)^2 + (W_5 - X_2)^2\} \{(X_5 - W_1)^2 + (W_4 - X_2)^2\}} \right] \\
 &= \cos^{-1} \left[1 - \frac{2738}{2665} \right] = \cos^{-1} (-.0274) = 1.5982
 \end{aligned}$$

$$2\sin^{-1} \rho_{2514,2535}^{(2)} = 2\sin^{-1} 0 = 0 \quad \text{since } 1 \neq 3, 1 \neq 5, 4 \neq 3, 4 \neq 5.$$

(Note: The only reason that Tables I. and II. are arranged in different orders is for convenience in grouping like elements. We might also mention that the elements of both tables were all computed to four decimal places, but only two decimal places are shown in the tables in order to save space.)

The sum of all the elements in Table I. is

$$(3.4) \quad 2 \sum_{\substack{iIjJ \\ i'I'j'J'}} \sin^{-1} \rho_{iIjJ, i'I'j'J'}^{(1)} = 628.732$$

The sum of all the elements in Table II. is

$$(3.5) \quad 2 \sum_{\substack{iIjJ \\ i'I'j'J'}} \sin^{-1} \rho_{iIjJ, i'I'j'J'}^{(2)} = 638.516$$

∫ The original figures, correct to four decimal places, were used in calculating both of these sums (3.4) and (3.5). ∫ We substitute (3.4) into (2.7) to obtain

$$\begin{aligned} Q_1 &= \frac{1}{4T} + \frac{1}{2\pi T^2} (628.732) \\ &= \frac{1}{160} + \frac{628.732}{3200\pi} = .068791 \end{aligned}$$

Similarly, we get

$$Q_2 = \frac{1}{4T} + \frac{1}{2\pi T^2} (638.516) = .069764$$

from (3.5) and (2.8). Hence, by (2.6) ,

$$Q = .069764$$

Since we already obtained w in (3.2), we have

$$\frac{w - \frac{1}{2}}{\sqrt{Q}} = \frac{-.275}{.2641} = -1.041$$

Inasmuch as the absolute value of -1.041 does not exceed 1.645 , we cannot reject the hypothesis $\alpha_Y = \alpha_Z$ at the $10^0/o$ level. ∫ We emphasize again that the sample sizes are not large enough and $\text{var}(w)$ and Q are not small enough for

the normal approximation to be very suitable, but for purposes of the illustration we are going ahead and applying the normal approximation anyhow. Although the normal approximation would appear to be suitable for just slightly larger values of m and n , it was felt that an illustration any lengthier than the one given here would consume an excessive amount of space.]

4. Confidence bounds. If we wish to obtain confidence bounds on $(\alpha_Z - \alpha_Y)$, we use roughly the same principle that was described in [1] and [2]. We need to find that value of $\delta = \alpha_Z - \alpha_Y$ which, when subtracted from every V_{iIjJ} , will cause (the resulting new) w to be on the threshold of significance. For example, suppose we want a 90% two-sided confidence interval for $(\alpha_Z - \alpha_Y)$. Then w will be on the threshold of being significantly large if 37 of the 40 V_{iIjJ} 's are positive and 1 of them is zero, since

$$(1.645 \times .2641 + \frac{1}{2}) \times 40 = .934 \times 40 = 37.4 \quad ;$$

and w will be on the threshold of being significantly small if 2 of the 40 V_{iIjJ} 's are positive and 1 is zero, since

$$(-1.645 \times .2641 + \frac{1}{2}) \times 40 = 2.6 \quad .$$

In Section 3 we obtained the v_{iIjJ} 's but not the V_{iIjJ} 's. We now must use the v_{iIjJ} 's to calculate the V_{iIjJ} 's by the formula

$$V_{iIjJ} = \frac{v_{iIjJ}}{X_I + W_J - X_i - W_j}$$

For example, we get

$$V_{3514} = \frac{v_{3514}}{x_5 + w_4 - x_3 - w_1} = \frac{3.92}{9+5-4-1} = \frac{3.92}{9} = .44 .$$

The values of all 40 V_{iIjJ} 's are as follows:

$$\begin{array}{llll} V_{1212} = .24, & V_{1213} = 1.16, & V_{1214} = 1.30, & V_{1215} = 1.09, \\ V_{1312} = -1.76, & V_{1313} = -.84, & V_{1314} = -.69, & V_{1315} = -1.31, \\ V_{1412} = -2.93, & V_{1413} = -2.01, & V_{1414} = -1.86, & V_{1415} = -2.71, \\ V_{1512} = -3.78, & V_{1513} = -2.28, & V_{1514} = -2.04, & \\ V_{1525} = -4.65, & V_{1535} = -2.96, & V_{1545} = -2.69, & V_{1515} = -3.70, \\ V_{2512} = .75, & V_{2513} = 2.92, & V_{2514} = 3.27, & \\ V_{2525} = -2.18, & V_{2535} = -.82, & V_{2545} = -.60, & V_{2515} = -.54, \\ V_{3512} = -2.09, & V_{3513} = .09, & V_{3514} = .44, & \\ V_{3525} = -3.60, & V_{3535} = -2.24, & V_{3545} = -2.02, & V_{3515} = -2.50, \\ V_{4512} = -3.75, & V_{4513} = -1.57, & V_{4514} = -1.23, & \\ V_{4525} = -4.43, & V_{4535} = -3.07, & V_{4545} = -2.85, & V_{4515} = -3.65 . \end{array}$$

We notice that 37 of the 40 V_{iIjJ} 's exceed -3.78 , while one V_{iIjJ} is equal to -3.78 . Hence -3.78 is the lower end of our confidence interval. Similarly, we find that 2 of the 40 V_{iIjJ} 's exceed 1.30 (with one V_{iIjJ} being equal to 1.30), so that 1.30 is the upper end of our confidence interval. Thus we can state that

$$-3.78 \leq (\alpha_Z - \alpha_Y) \leq 1.30$$

with confidence coefficient $\geq 90\%$.

In most practical situations the true values of the parameters would never be known. However, the numerical illustration in this paper was constructed artificially, and so all the parameters are known. The values used were $\beta = 7$, $\alpha_Y = 4$, $\alpha_Z = 2$, $\sigma_e^2 = 25$, and $\sigma_f^2 = 1$. Thus $(\alpha_Z - \alpha_Y) = -2$.

5. The problem of evaluating Q_1 and Q_2 . Obtaining Q_1 (2.7) and Q_2 (2.8) is by far the biggest problem associated with the calculation of the test statistic which we have presented. In general, the number of terms which are included in the summations in formulas (2.7) and (2.8) increases rapidly as m and n increase. It is therefore appropriate for us to consider possible techniques for reducing the computational burden involved in the evaluation of Q_1 and Q_2 .

It would be helpful if, by theoretical means, simpler but equivalent formulas for Q_1 and Q_2 could somehow be found. There is no indication, however, that any success in this direction could be achieved. Alternatively, it might be possible mathematically to obtain upper bounds on Q_1 and Q_2 which would be relatively easy to compute and which would be close to the actual values of Q_1 and Q_2 ; but at present no progress has been made in this direction.

Except when T is relatively very small, the direct calculation of Q_1 and Q_2 by means other than a high-speed computer is out of the question. Unfortunately, however, even a high-speed computer is not fast enough to calculate Q_1 and Q_2 directly when m and n are of moderate size. It has been estimated that the calculation of a typical non-zero element of the summation on the right-hand side of (2.7) or (2.8) would require roughly 15 milliseconds on the UNIVAC at Chapel Hill.² Thus, even if there were only a million of these ele-

²The author is indebted to Thomas G. Donnelly for furnishing this approximate estimate.

ments to be calculated (as there very well could be if m and n were larger than 10), the calculations would require (roughly) over four hours of computer time.

Hence it is necessary to find some kind of short-cut for determining Q_1 and Q_2 . Perhaps the most obvious approach is to take a sample of the elements which enter into the summations in (2.7) and (2.8), rather than calculating all $\frac{1}{2} T(T-1)$ elements in each summation. One possibility would be to draw random samples of sizes $s^{(1)}$ and $s^{(2)}$ (say) from all the $\frac{1}{2} T(T-1)$ elements (including both zero and non-zero elements) in the summations in (2.7) and (2.8) respectively, and then use the sample means as estimates of the respective population means [where each population consists of $\frac{1}{2} T(T-1)$ elements]. For example, if we want to estimate Q_1 , we could draw a sample of (and calculate) $s^{(1)}$ elements each of the form

$$(5.1) \quad \sin^{-1} \rho_{iIjJ, i'I'j'J'}^{(1)}$$

Let the sum of these $s^{(1)}$ elements (5.1) be $a^{(1)}$ (say). Then we could use the expression

$$(5.2) \quad Q_1 \doteq \frac{1}{4T} + \frac{1}{\pi T^2} \times \frac{1}{2} T(T-1) \left[\frac{a^{(1)}}{s^{(1)}} \right] = \frac{1}{4T} + \frac{T-1}{2\pi T} \left(\frac{a^{(1)}}{s^{(1)}} \right)$$

to estimate Q_1 . (We use the symbol \doteq to mean "is estimated by".) Similarly, we could write

$$(5.3) \quad Q_2 \doteq \frac{1}{4T} + \frac{T-1}{2\pi T} \left(\frac{a^{(2)}}{s^{(2)}} \right)$$

Perhaps a better technique of sampling, however, would be to determine exactly the number of elements in the population which are zero by virtue of satisfying

(5.4a) $i \neq i', i \neq I', I \neq i', I \neq I'$ in the case of elements for (2.7)

or

(5.4b) $j \neq j', j \neq J', J \neq j', J \neq J'$ in the case of elements for (2.8),

and then to draw the sample from that sub-population which consists only of the elements not satisfying (5.4). For example, suppose we are estimating Q_1 .

Suppose we determine that exactly $t_o^{(1)}$ of the $\frac{1}{2} T(T-1)$ elements (5.1) are zero by virtue of satisfying (5.4a). Then the remaining

$$t_p^{(1)} = \frac{1}{2} T(T-1) - t_o^{(1)}$$

elements do not satisfy (5.4a). We may draw a random sample of size $s_p^{(1)}$ (say) from this sub-population of size $t_p^{(1)}$. Suppose we calculate these $s_p^{(1)}$ elements in the sample and find that their sum is $a_p^{(1)}$ (say). Then we can use the expression

$$(5.5) \quad Q_1 \doteq \frac{1}{4T} + \frac{1}{\pi T^2} \left\langle t_p^{(1)} \frac{a_p^{(1)}}{s_p^{(1)}} \right\rangle$$

to estimate Q_1 . Similarly, we can write the expression

$$(5.6) \quad Q_2 \doteq \frac{1}{4T} + \frac{1}{\pi T^2} \left\langle t_p^{(2)} \frac{a_p^{(2)}}{s_p^{(2)}} \right\rangle$$

Still more sophisticated sampling techniques are available through the use of stratified sampling. In particular (for example), we can divide the elements not satisfying (5.4) into two groups or strata: those elements which satisfy

(5.7a) $i = i', I = I'$ or $i = I', I = i'$ in the case of elements for (2.7)

or

(5.7b) $j = j', J = J'$ or $j = J', J = j'$ in the case of elements for (2.8),

and those elements which do not. Samples can then be taken from each stratum.

[The reason for stratifying is that those elements which satisfy (5.7) will tend to have a much higher value than those elements which do not.] Suppose, for example, that we are estimating Q_1 . Of the $t_p^{(1)}$ elements (5.1) not satisfying (5.4a), suppose that $t_2^{(1)}$ satisfy (5.7a) while $t_1^{(1)} = t_p^{(1)} - t_2^{(1)}$ do not satisfy (5.7a). Suppose we take samples of sizes $s_2^{(1)}$ and $s_1^{(1)}$ from the two strata satisfying (5.7a) and not satisfying (5.7a) respectively. Let the sums of the elements in these samples be $a_2^{(1)}$ and $a_1^{(1)}$ respectively. Then we may estimate Q_1 by using the expression

$$(5.8) \quad Q_1 \doteq \frac{1}{4T} + \frac{1}{\pi T^2} \left\langle t_1^{(1)} \frac{a_1^{(1)}}{s_1^{(1)}} + t_2^{(1)} \frac{a_2^{(1)}}{s_2^{(1)}} \right\rangle .$$

Similarly, we can write

$$(5.9) \quad Q_2 \doteq \frac{1}{4T} + \frac{1}{\pi T^2} \left\langle t_1^{(2)} \frac{a_1^{(2)}}{s_1^{(2)}} + t_2^{(2)} \frac{a_2^{(2)}}{s_2^{(2)}} \right\rangle .$$

In addition to the three possible techniques just presented for estimating Q_1 and Q_2 , certain other techniques (involving stratified sampling) might be devised. The choice of what technique to use will be governed largely by the comparative costs of the different techniques (in terms mainly of computer time needed for obtaining a specified precision for the estimates of Q_1 and Q_2).

The discussions of the three sampling techniques just presented did not state explicitly whether sampling without replacement or sampling with replace-

ment should be used. Sampling without replacement is, of course, more efficient than sampling with replacement, assuming that the cost (i.e., computer time) per element sampled is the same under both schemes. However, there is no reason why sampling with replacement could not be used if it was justified by cost considerations.

If a sampling scheme is utilized to estimate Q_1 and Q_2 , the question might be raised as to what the sizes of the samples should be. For example, if we wish to use formulas (5.5) and (5.6), what should $s_p^{(1)}$ and $s_p^{(2)}$ be? It is estimated very roughly that, if we take the sample sizes $(s_p^{(1)}$ and $s_p^{(2)})$ each to be 10,000, then the chances will be about 99 out of 100 that the estimate of \sqrt{Q} will be within $\pm 1\%$ of the true value of \sqrt{Q} ; that, if both sample sizes are 2,500, then the chances will be about 99 out of 100 that \sqrt{Q} will be estimated to within $\pm 2\%$ of its true value; and that, if both sample sizes are 40,000, then \sqrt{Q} will be estimated to within $\pm \frac{1}{2}\%$ of its true value about 99% of the time. These statements will be roughly correct regardless of the values of m , n , and T , so long as $t_p^{(1)}$ and $t_p^{(2)}$ are considerably larger than $s_p^{(1)}$ and $s_p^{(2)}$. (For relatively small $t_p^{(1)}$ and $t_p^{(2)}$, the statements will still be roughly correct if sampling with replacement is used, but will be conservative if sampling without replacement is used.)

If the stratified sampling scheme associated with the formulas (5.8) and (5.9) [instead of the sampling scheme associated with (5.5 - 5.6)] is employed for estimating Q , then we should achieve greater precision with the same total sample size $s_p^{(1)} = s_1^{(1)} + s_2^{(1)}$ or $s_p^{(2)} = s_1^{(2)} + s_2^{(2)}$, and so the statements made in the preceding paragraph would be conservative. On the other hand, if the sampling scheme associated with (5.2 - 5.3) is used, then for given sample sizes we will have considerably less precision than for the scheme associated with (5.5 - 5.6); furthermore, the tendency will be that this

loss in precision will be relatively worse the larger m and n are [since, the larger m and n are, the larger will tend to be the expected proportion of zero elements sampled for the scheme associated with (5.2 - 5.3)] .

REFERENCES

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

[2] Richard F. Potthoff, "Comparing the Medians of Two Symmetrical Populations," Institute of Statistics Mimeograph Series No. 316, University of North Carolina, 1962.

