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

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

by
Richard F. Potthoff

October 1963

In psychological and other applications, it may be necessary to make certain comparisons of two regression lines when the variances are unequal. Such problems arise rather frequently, for example, in studies comparing two alternative curriculums or two different teaching methods. By generalizing an idea which Scheffé used to obtain a test for the Behrens-Fisher problem, this paper develops some tests for comparing two regression lines when the two sets of error terms are normally distributed but with two different variances. Scheffé's test itself is a randomized test, but in this paper we present both randomized and non-randomized tests. Both simple and multiple regression are considered, but the simple regression tests are computationally easier than the multiple regression tests. The basic test statistic which is used is the ordinary t-statistic. Essentially two types of problems are dealt with: (A) determining whether the two regression lines are identical when they are known to be parallel; and (B) determining whether the two regression lines are parallel. Confidence bounds as well as tests of hypotheses are available. This paper is on a theoretical level; a more practically-oriented discussion, with numerical examples, is given in a separate paper.

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

Institute of Statistics
Mimeo Series No. 374

L'UNIVERSITÉ DE LA CAROLINE DU NORD

Départment de Statistique

Chapel Hill, Caroline du Nord

QUELQUES TESTS DU TYPE DE SCHEFFÉ POUR QUELQUES PROBLÈMES DE LA RÉGRESSION DU TYPE DE BEHRENS-FISHER

par

Richard F. Potthoff

octobre 1963

Dans les applications psychologiques et les autres applications, il peut être nécessaire de faire des certaines comparaisons de deux lignes de régression quand les variances sont inégales. De tels problèmes se présentent assez frequemment, par exemple, dans les études pour comparer deux curriculums alternatifs ou deux méthodes différentes d'enseignement. En généralisant une idée que Scheffé a utilisée pour obtenir un test pour le problème de Behrens-Fisher, ce papier développe quelques tests pour comparer deux lignes de régression quand les deux ensembles d'erreurs résiduelles suivent la distribution normale mais avec deux variances différentes. Le test de Scheffé lui-même est un test randomisé, mais dans ce papier on présente à la fois des tests randomisés et des tests non randomisés. On considére à la fois la régression simple et la régression multiple, mais les tests de la régression simple sont plus faciles à calculer que les tests de la régression multiple. La statistique fondamentale des tests c'est tout simplement le t de Student. On considére deux sortes de problèmes: (A) déterminer si deux lignes de régression sont identiques quand on sait qu'elles sont parallèles; et (B) déterminer si les deux lignes de régression sont parallèles. Les limites de confiance ainsi que les tests des hypothèses sont disponibles. Ce papier est plutôt théorique; dans un autre papier on présente une discussion plus pratique, avec des exemples numériques.

Cette recherche a été rendue possible partiellement grâce à des fonds du Programme d'Aide à la Recherche de l'OTAN sous un arrangement conjoint entre l'Institut de Statistique de l'Université de Paris, l'Istituto di Calcolo delle Probabilità de l'Université de Rome, et le Départment de Statistique de l'Université de la Caroline du Nord; c'était soutenue partiellement par l'Educational Testing Service; et c'était soutenue partiellement par l'Air Force Office of Scientific Research.

Institut de Statistique
Série Mimeographique, Numéro 374

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

by

Richard F. Potthoff[1]
University of Paris and University of North Carolina

1. INTRODUCTION

A solution to the Behrens-Fisher problem involving randomized pairing of the two samples was presented by Scheffé [7]. The present paper attacks certain Behrens-Fisher-type regression problems by developing some tests somewhat similar to the test of [7].

Two basic Behrens-Fisher-type regression problems will be considered in this paper. The first, which we will call Problem A, consists of testing whether two regression lines are identical when they are assumed to be parallel, under the condition that the two variances may be unequal. More specifically, we suppose that we have mutually independent observations $(Y_1, Y_2, \ldots, Y_M)$, $(Z_1, Z_2, \ldots, Z_N)$ such that

(1a) $$Y_i = \alpha_Y + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_r X_{ir} + e_i \qquad (i = 1, 2, \ldots, M)$$

and

(1b) $$Z_j = \alpha_Z + \beta_1 W_{j1} + \beta_2 W_{j2} + \ldots + \beta_r W_{jr} + f_j \qquad (j = 1, 2, \ldots, N) \quad ,$$

where the e_i's are $N(0, \sigma_e^2)$, the f_j's are $N(0, \sigma_f^2)$, the α's, β's, and σ^2's are unknown parameters, the $X_{i\ell}$'s and $W_{j\ell}$'s are (known) fixed constants, and

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

M is assumed (without loss of generality) to be $\leq N$. We desire to test

(2) $$H_{OA}: \alpha_Y = \alpha_Z \quad .$$

The second problem, to be called Problem B, consists of testing whether two regression lines are parallel when the variances may be unequal. More specifically, the set-up for Problem B is the same as for Problem A, except that instead of (1) our model is

(3a) $$Y_i = \alpha_Y + \beta_{Y1}X_{i1} + \beta_{Y2}X_{i2} + \cdots + \beta_{Yr}X_{ir} + e_i \qquad (i = 1, 2, \ldots, M)$$

and

(3b) $$Z_j = \alpha_Z + \beta_{Z1}W_{j1} + \beta_{Z2}W_{j2} + \cdots + \beta_{Zr}W_{jr} + f_j \qquad (j = 1, 2, \ldots, N) ,$$

and the hypothesis to be tested is

(4) $$H_{OB}: \beta_{Y1} = \beta_{Z1}, \ \beta_{Y2} = \beta_{Z2}, \ \ldots, \ \beta_{Yr} = \beta_{Zr} \quad .$$

Scheffé-type tests will be presented for Problems A and B. All our discussions, in addition to considering the multiple regression situation with general r, will also pay particular attention to the special case $r = 1$, for which the tests are simpler. Section 2 is concerned with randomized tests, while Section 3 develops non-randomized tests closely akin to the randomized tests of Section 2. For $r = 1$, the non-randomized tests are considerably easier to calculate than the corresponding randomized tests and at the same time appear to be almost as powerful, so that probably they would usually be preferred; for $r > 1$, however, the picture is less clear. Many users of statistics object to randomized tests on philosophical grounds which have been adequately expounded elsewhere but which have perhaps failed to convince many mathematical statisticians. Even if the randomized tests of Section 2 are not used, however, they still provide both a standard of comparison for and a basis of constructing the non-randomized tests of Section 3.

An entirely different approach to Problems A and B, based on test statistics analogous to the Wilcoxon statistic, was presented respectively in [6] and [5] for the case $r=1$; the test for Problem A [6] is rather complicated, but the one for Problem B [5] is much simpler. Welch [8] has proposed a technique, based on an approximating t-distribution, which is applicable to Problem B for $r=1$. Recent results of Hájek concerning a generalized t-distribution [3] can be utilized to obtain a conservative test for Problem B when $r=1$. Thus it seems to be chiefly for the case of Problem B that competing tests are available.

No detailed comparison of alternative tests will be attempted here. However, we may simply remark that the tests of [6], [5], [8], and [3] are all inexact (i.e., the actual level of significance is only approximately equal to the stated level of significance) but become less inexact with increasing M, whereas all the t-tests of the present paper are exact. Consequently, it appears that the tests of this paper will be particularly useful when M is small.

The technique of Scheffé [7], which provided the inspiration for the tests developed in the present paper, has also been generalized with respect to an entirely different problem by Mazuy and Connor [4]. Scheffé-type tests for certain two-way ANOVA models with unequal variances are considered in [4].

2. RANDOMIZED TESTS

The tests of this paper, like Scheffé's test [7], will all be based on the ordinary t-distribution (except that for Problem B with general r the possibility of an F-test will also need to be considered). In this section we will seek optimal tests which have as many degrees of freedom for t as are readily available and which are based on minimum variance estimators, a method of attack closely related to that used by Scheffé [7] (which was to obtain a certain confidence interval of minimum expected length). For Problem A, however, it will turn out that we will have to seek a minimax estimator instead of a minimum variance estimator. The randomized tests of this section will all be optimal in the sense just indicated, whereas the non-randomized tests of Section 3 will all be sub-optimal to a lesser or greater degree. (In using the variance of an estimator as the criterion for judging a test, we are assuming implicitly that the variance of the estimator is closely related to the power of the test — an assumption which seems completely reasonable but which we will make no attempt to prove.)

2.1. Problem A for general r.

Let us first attempt to base a test of $H_{OA}(2)$ on the minimum variance unbiased estimator of $\alpha = \alpha_Z - \alpha_Y$. This estimator, which we will call $\hat{\alpha}^*$, can easily be found if we apply standard least squares theory to the homoscedastic variables $kY_1, kY_2, \ldots, kY_M, Z_1, Z_2, \ldots, Z_N$, where we define $k = \sigma_f/\sigma_e$. We need some more notation. Let $\bar{Y}, \bar{Z}, \bar{X}_\ell$, and $\bar{W}_\ell (\ell = 1, 2, \ldots, r)$ denote the means of the Y_i's, Z_j's, $X_{i\ell}$'s, and $W_{j\ell}$'s respectively. Let $\bar{X}$ and $\bar{W}$ be rx1 vectors containing the $\bar{X}_\ell$'s and $\bar{W}_\ell$'s respectively. Define $\bar{U}(rx1) = \bar{W} - \bar{X}$. Let $X(Mxr)$ and $W(Nxr)$ be matrices containing the $X_{i\ell}$'s and $W_{j\ell}$'s respectively. Define two matrices $x(Mxr) = X - j_M\bar{X}'$ and $w(Nxr) = W - j_N\bar{W}'$, where j denotes a column vector containing all 1's. Let $Y(Mx1)$ and $Z(Nx1)$ be vectors containing the Y_i's and Z_j's respectively. Define $K = k^2$. Then it turns out that

$$(5) \qquad \hat{\alpha}^* = (\bar{Z} - \bar{Y}) - \bar{U}'[w'w + K\, x'x]^{-1} (w' Z + K\, x' Y) \quad ,$$

and the variance of this minimum variance unbiased estimator is

$$(6) \quad \mathrm{var}(\hat{\alpha}^*) = \frac{\sigma_f^2}{N} + \frac{\sigma_e^2}{M} + \bar{U}' \left[\frac{1}{\sigma_f^2} w'w + \frac{1}{\sigma_e^2} x'x \right]^{-1} \bar{U} \qquad .$$

Unfortunately, however, (5) is not free of the unknown nuisance parameter K, and so it is not possible to use (5) either as an estimator of α or as a basis for constructing a test of $H_{OA}(2)$. This suggests that we should try a minimax approach. Consider a completely general linear function $\hat{\alpha}$ which is an unbiased estimator of α, i.e.,

(7) $$\hat{\alpha} = h'Z - g'Y \quad ,$$

where $g(M\text{x}1)$ and $h(N\text{x}1)$ must be chosen so as to satisfy

(8) $$g'j_M = 1, \quad h'j_N = 1, \quad g'X = h'W$$

in order to have $E(\hat{\alpha}) = \alpha$. Observe that

(9) $$\text{var}(\hat{\alpha}) = h'h\,\sigma_f^2 + g'g\,\sigma_e^2 \quad .$$

If we attempt to choose $\hat{\alpha}$ (7) so that its variance (9) is minimax, we run into trouble, because, no matter what selection is made for g and h, the variance (9) will be unbounded with respect to σ_e^2 and σ_f^2. However, in place of (9), let us consider an alternative loss function

(10) $$L = L(\hat{\alpha},K) = \frac{\text{var}(\hat{\alpha})}{\text{var}(\hat{\alpha}^*)} = \frac{K\,h'h + g'g}{(1/N)K + (1/M) + K\,\bar{U}'\,[w'w + K\,x'x]^{-1}\,\bar{U}}$$

$$= \frac{h'h + (1/K)\,g'g}{(1/N) + (1/MK) + (1/K)\,\bar{U}'\,[(1/K)w'w + x'x]^{-1}\,\bar{U}} \quad ,$$

which is the ratio of (9) to (6), i.e., the ratio of the variance of the estimator we select (7) to the variance of the ideal but impossible estimator (5). L (10) may thus be considered as the ratio of our actual "loss" to our smallest possible "loss", and hence seems to constitute an appropriate loss function itself. L must always be $\geq$ 1, but fortunately it is bounded above with respect to K; the closer L is to 1, the better.

What we will now do is to find an $\hat{\alpha}$ which minimizes the quantity $\max_K L(\hat{\alpha},K)$. For the time being, let us consider a class of estimators less general than (7), of the form

$$(11)\quad \hat{\alpha}_C = (\bar{Z} - \bar{Y}) - \bar{U}'[w'w + C\,x'x]^{-1}\,(w'\,Z + C\,x'\,Y) \qquad (C \geq 0).$$

Note that (11) is formally identical with (5), and indeed was suspected as being a fruitful source of an estimator precisely because of (5); however, K in (5) is an unknown nuisance parameter, whereas C in (11) is a number which we are required to choose. It is easy to show that $\hat{\alpha}_C$ (11) will be an unbiased estimator of α regardless of the choice of C. Its variance is

$$(12)\quad \text{var}(\hat{\alpha}_C) = \frac{\sigma_f^2}{N} + \frac{\sigma_e^2}{M} + \bar{U}'\,[w'w + C\,x'x]^{-1}\,[w'w\,\sigma_f^2 + C^2 x'x\,\sigma_e^2] \cdot [w'w + C\,x'x]^{-1}\,\bar{U}\,.$$

The ratio of (12) to (6) is

$$(13)\quad L(\hat{\alpha}_C,K) = \frac{(1/N)K + (1/M) + \bar{U}'[w'w+Cx'x]^{-1}[Kw'w+C^2x'x][w'w+Cx'x]^{-1}\,\bar{U}}{(1/N)K + (1/M) + K\,\bar{U}'[w'w + K\,x'x]^{-1}\,\bar{U}}$$

$$= \frac{(1/N) + (1/MK) + \bar{U}'[w'w + Cx'x]^{-1}[w'w + (C^2/K)x'x][w'w + Cx'x]^{-1}\bar{U}}{(1/N) + (1/MK) + (1/K)\,\bar{U}'\,[(1/K)w'w + x'x]^{-1}\,\bar{U}}.$$

Taking the limits in (13) as $K \to 0$ and as $K \to \infty$, we can write

$$(14a)\quad L(\hat{\alpha}_C,0) = 1 + MC^2\,\bar{U}'\,[w'w + C\,x'x]^{-1}\,x'x\,[w'w + Cx'x]^{-1}\,\bar{U}$$

and

$$(14b)\quad L(\hat{\alpha}_C,\infty) = 1 + N\,\bar{U}\,[w'w + Cx'x]^{-1}\,w'w\,[w'w + Cx'x]^{-1}\,\bar{U}\,.$$

Combining (13) and (14), and then utilizing the fact that $[w'w + K\,x'x]^{-1}$ is a positive definite matrix, we find that

$$L(\hat{\alpha}_C,K) = \frac{(1/M)\,L(\hat{\alpha}_C,0) + K(1/N)\,L(\hat{\alpha}_C,\infty)}{(1/M) + K(1/N) + K\,\bar{U}'[w'w + K\,x'x]^{-1}\bar{U}}$$

$$\leq \frac{1/M}{(1/M)+K(1/N)}\,L(\hat{\alpha}_C,0) + \frac{K(1/N)}{(1/M)+K(1/N)}\,L(\hat{\alpha}_C,\infty)$$

$$\leq \max\,[L(\hat{\alpha}_C,0),\ L(\hat{\alpha}_C,\infty)] \quad ,$$

from which it follows that

(15) $$\max_K\ L(\hat{\alpha}_C,K) = \max\,[L(\hat{\alpha}_C,0),\ L(\hat{\alpha}_C,\infty)] \quad .$$

We want to minimize (15) with respect to C. If (14a) is a strictly increasing function of C and (14b) a strictly decreasing function (which we will soon prove to be the case), then there will be a unique value of C(which we will call C*, say) for which (14a) equals (14b), i.e., there will be a unique C satisfying

(16) $$MC^2\bar{U}'\,[w'w + Cx'x]^{-1}x'x[w'w + Cx'x]^{-1}\,\bar{U} = N\,\bar{U}'[w'w + Cx'x]^{-1}w'w[w'w + Cx'x]^{-1}\bar{U},$$

and (15) will of course be minimized at the point $C = C^*$. Thus $\hat{\alpha}_{C^*}$ will be the minimax estimator in the class of estimators $\hat{\alpha}_C$ (11), and we will show later that $\hat{\alpha}_{C^*}$ is also minimax among all estimators of the more general class $\hat{\alpha}$ (7).

Right now, though, we investigate the two sides of (16) in order to establish the uniqueness of C*. If we define

(17) $$\mu = (M/N)C,\ G(rxr) = (1/M)\,x'x,\ \text{and}\ \ H(rxr) = (1/N)w'w \quad ,$$

then the two sides of (16) become

(18) $\bar{U}'[(1/\mu)H + G]^{-1} G[(1/\mu)H + G]^{-1}\bar{U} = \bar{U}'[H + \mu G]^{-1} H[H + \mu G]^{-1} \bar{U}$.

Since G and H are both symmetric positive definite, there exists a non-singular matrix $P(r\times r)$ such that

(19) $P'GP = I$ and $P'HP = D$, i.e., $G = (P')^{-1}P^{-1}$ and $H = (P')^{-1} D P^{-1}$,

where $D(r\times r)$ is a diagonal matrix whose diagonal elements (which are the characteristic roots of $G^{-1}H$) are all > 0 (we are appealing here to a standard matrix theorem, which is given, e.g., in [1, p. 341]). After application of (19), the two sides of (18) change to

(20) $$Q'[(1/\mu) D + I]^{-2}Q = Q'[D + \mu I]^{-2} D Q \quad ,$$

where $Q(r\times 1) = P' \bar{U}$. It is apparent that the left side of (20) [and hence of (16) also] is an increasing function and the right side a decreasing function of μ (and hence of C), which is what we were attempting to prove. Thus (16) has the unique solution C*.

We now return to the general class $\hat{\alpha}$ (7) and show that there is no $\hat{\alpha}$ for which $\max_K L(\hat{\alpha},K)$ is smaller than

(21) $$\max_K L(\hat{\alpha}_{C*},K) = L(\hat{\alpha}_{C*},0) = L(\hat{\alpha}_{C*},\infty) \quad .$$

Now clearly $\max_K L(\hat{\alpha}, K)$ is not less than

(22) $$\max [L(\hat{\alpha},0) , \quad L(\hat{\alpha},\infty)] = \max [M g'g, N h'h] ,$$

where (22) was written by taking the limits in (10) as $K \to 0$ and as $K \to \infty$. Thus it will suffice to show that the minimum of the right side of (22) [with respect to all (g,h) satisfying (8)] is $\geq$ (21). When (8) is observed, we may define

$$a(r\times 1) = X'g = W'h .$$

For any fixed value of a, we can minimize $Mg'g$ subject to the conditions $g'j_M=1$ and $g'X = a'$; the solution, which can be found rather simply by using Lagrangian multipliers, is

(23a) $$\min_g (M g'g|a) = 1 + M(a - \bar{X})' [x'x]^{-1} (a - \bar{X}) .$$

Similarly, for any fixed a, the minimum of $Nh'h$ subject to the conditions $h'j_N=1$ and $h'W = a'$ is

(23b) $$\min_h (Nh'h|a) = 1 + N(a - \bar{W})' [w'w]^{-1} (a - \bar{W}) .$$

Hence it follows from (23) that the minimum of the right side of (22) subject to the conditions (8) is

(24) $$1 + \min_a \max [(a - \bar{X})' G^{-1}(a - \bar{X}), (a - \bar{W})' H^{-1} (a - \bar{W})].$$

What we will now show is that (24) is equal to (21).

Note first that the minimum in (24) must occur at some value a such that

(25) $$(a - \bar{X})' G^{-1} (a - \bar{X}) = (a - \bar{W})' H^{-1} (a - \bar{W}) .$$

Hence we can find the minimum in (24) by using a Lagrangian multiplier and minimizing $(a - \bar{X})' G^{-1} (a - \bar{X})$ subject to the condition (25). Differentiating

$$(a - \bar{X})' G^{-1}(a - \bar{X}) - \lambda[(a - \bar{X})'G^{-1}(a - \bar{X}) - (a - \bar{W})' H^{-1}(a - \bar{W})]$$

with respect to a, we easily obtain the relation

(26) $$a = [(1 - \lambda)G^{-1} + \lambda H^{-1}]^{-1}[(1 - \lambda)G^{-1} \bar{X} + \lambda H^{-1} \bar{W}]$$

after setting the vector derivative equal to zero. Substituting (26) into (25), we find (after some matrix manipulations) that the two sides of (25) can then be written respectively as

$$(27)\quad \lambda^2\, \bar{U}'[(1-\lambda)H + \lambda G]^{-1}\, G[(1-\lambda)H + \lambda G]^{-1}\, \bar{U} \;=$$

$$(1-\lambda)^2\, \bar{U}'[(1-\lambda)H + \lambda G]^{-1} H[(1-\lambda)H + \lambda G]^{-1}\, \bar{U}\ .$$

But (27) is identical with (18), if we set $\lambda = \mu/(1 + \mu)$. Thus (25) is equal to (16) at the point of solution $C = C^*$ [i.e., at the point $\lambda = MC^*/(N + MC^*)$], from which it follows immediately that (24) is equal to (21). This completes the proof that $\hat{\alpha}_{C^*}$ is minimax among all estimators of the form (7).

By combining (16) and (12), we can write the variance of the estimator as

$$(28)\qquad \operatorname{var}(\hat{\alpha}_{C^*}) \;=\; s\left(\frac{\sigma_e^2}{M} + \frac{\sigma_f^2}{N} \right) \qquad ,$$

where s is the known constant

$$(29)\qquad s = 1 + N\, \bar{U}'\, [w'w + C^*x'x]^{-1}\, w'w\, [w'w + C^*x'x]^{-1}\, \bar{U} \qquad .$$

We now develop a t-test of $H_{0A}(2)$ based on the optimal estimator $\hat{\alpha}_{C^*}$. Note first that there exist matrices $E_X(M \times [M-r-1])$ and $E_W(N \times [M-r-1])$ such that

$$(30a)\qquad E'_X E_X = I, \qquad\qquad E'_W E_W = I$$

and

$$(30b)\qquad E'_X\, j_M = 0,\ E'_W\, j_N = 0,\ E'_X\, X = 0,\ E'_W\, W = 0$$

(where 0 denotes a null vector or matrix). Define an $(M - r - 1) \times 1$ vector

$$(31)\qquad V_A = M^{-\frac{1}{2}}\, E'_X\, Y \;+\; N^{-\frac{1}{2}}\, E'_W\, Z \qquad .$$

From (30) and (1) we see that V_A (31) is always $N(0, \sigma_A^2 I)$, where we are writing $\sigma_A^2 = (\sigma_e^2/M) + (\sigma_f^2/N)$. Hence $V_A'V_A/\sigma_A^2$ follows the χ^2-distribution with (M-r-1) d.f. Furthermore, it follows from (30b) that $\hat{\alpha}_{C*}$ [see (11)] and V_A (31) are independent. Since $(\hat{\alpha}_{C*} - \alpha)$ is $N(0, s\,\sigma_A^2)$ [see (28-29)], we conclude that

$$(32) \qquad t = \hat{\alpha}_{C*} / \left(\frac{s\, V_A' V_A}{M - r - 1} \right)^{\frac{1}{2}}$$

follows the t-distribution with (M-r-1) d.f. if $H_{0A}(2)$ is true. Hence (32) can be used to test $H_{0A}(2)$. Although our emphasis here is formally on the testing of a hypothesis, we can of course also obtain confidence bounds on α if we wish.

The numerator of (32) obviously is unique, since $\hat{\alpha}_{C*}$ is calculated from (11), taking $C = C*$ to be the solution of (16). But the denominator of (32) is not unique, inasmuch as the matrices E_X and E_W which appear in (31) can be chosen in infinitely many ways and still satisfy (30). Furthermore, for any choice of E_X and E_W, if the columns of either matrix are permuted, then (30) will always remain satisfied but V_A (31) (and also $V_A'V_A$) will generally have a different value for each permutation. This suggests that, once E_X and E_W have been obtained (by some process which itself would necessarily have to be arbitrary), then the columns of one of the matrices ought to be permuted by means of a random permutation.

We will not attempt to consider here the question of whether it is possible to get a t-test with more degrees of freedom than the (M-r-1) d.f. of (32). Rough preliminary investigation indicates, however, that this problem may involve mathematical difficulties whose effect could be such as to render the problem of little practical importance anyway.

2.2. Problem A for r=1.

For the special case r=1, the solution $C = C^*$ of (16) is simply

(33) $$C^* = N\,\sigma_W / M\sigma_X \quad ,$$

where we define $\sigma_W^2 = w'w/N$ and $\sigma_X^2 = x'x/M$. Hence, upon substitution of (33), (29) becomes

(34) $$s = 1 + \frac{(\overline{W} - \overline{X})^2}{(\sigma_W + \sigma_X)^2} \quad ;$$

and from (11) and (33) we obtain the formula

(35) $$\hat{\alpha}_{C^*} = (\overline{Z} - \overline{Y}) - (\overline{W} - \overline{X}) \left[\frac{(w'Z/N\sigma_W) + (x'Y/M\sigma_X)}{\sigma_W + \sigma_X} \right] \quad .$$

Thus, when $r = 1$, the t-statistic (32) for testing H_{OA} (2) is calculated by plugging (35), (34), and (31) into the formula (32), where the matrices E_X and E_W in (31) were selected so as to satisfy (30).

2.3. Problem B for r=1.

In tackling Problem A, we first tried to base a test of H_{OA} (2) on the minimum variance unbiased estimator of $\alpha = \alpha_Z - \alpha_Y$; but this didn't work, because the estimator (5) contains a nuisance parameter. For Problem B, however, no such difficulty as this is encountered: considering for the moment the case of general r and letting β_Y (rx1) and β_Z (rx1) denote vectors containing the $\beta_{Y\ell}$'s and $\beta_{Z\ell}$'s respectively, we find easily that the minimum variance unbiased estimators of the r elements of β (rx1) $= \beta_Z - \beta_Y$ are given by

(36) $$\hat{\beta}^* \text{ (rx1)} = [w'w]^{-1}\, w'Z - [x'x]^{-1}\, x'Y \quad ,$$

and (36) [unlike (5)] is of course free of the unknown nuisance parameter K. [Note here the obvious point that our hypothesis H_{OB} (4) is merely the same thing as $\beta = 0$.]

For the case $r=1$, it is an easy matter to discover a randomized t-test of H_{OB} (4) based on the estimator $\hat{\beta}^*$ (36). First note that

$$(37) \qquad \mathrm{var}\,(\hat{\beta}^*) = \frac{\sigma_e^2}{x'x} + \frac{\sigma_f^2}{w'w} = \sigma_B^2 \quad \text{(say)}$$

when $r=1$. Now we can find matrices E_X and E_W satisfying (30), following which we proceed analogously to (31) and define the $(M-2) \times 1$ vector

$$(38) \qquad V_B = M^{-\frac{1}{2}} \sigma_X^{-1} E_X' Y + N^{-\frac{1}{2}} \sigma_W^{-1} E_W' Z \qquad .$$

V_B (38) is $N(0, \sigma_B^2 I)$ and independent of $\hat{\beta}^*$ (36); also, $(\hat{\beta}^* - \beta)$ is $N(0, \sigma_B^2)$. Hence

$$(39) \qquad t = \hat{\beta}^* / \left(\frac{V_B' V_B}{M-2}\right)^{\frac{1}{2}} = \left(\frac{w'Z}{w'w} - \frac{x'Y}{x'x}\right) / \left(\frac{V_B' V_B}{M-2}\right)^{\frac{1}{2}}$$

follows the t-distribution with $(M-2)$ d.f. if H_{OB} (4) is true. Thus the t-statistic (39), which like (32) would evidently involve us with some randomization in the calculation of its denominator, may be used for testing H_{OB} (4) for the case $r=1$.

2.4. Problem B for general r.

In attempting to find a test of H_{OB} (4) for the case of general r, the natural approach to take might be to try to generalize (39) and obtain some sort of F-test analogous to (39). This method of attack seems to run into difficulty, however, because of the way that the nuisance parameters σ_e^2 and σ_f^2 are involved in the inverse of the $r \times r$ variance matrix

$$(40) \qquad \mathrm{var}\,(\hat{\beta}^*) = [x'x]^{-1} \sigma_e^2 + [w'w]^{-1} \sigma_f^2 \qquad .$$

By using an entirely different approach, though, it is definitely possible to develop an F-test [with r and $(M - 2r - 1)$ d.f.] of H_{OB} (4), by employing a device which will be mentioned briefly in Subsection 3.4. But we are deferring our consideration of such F-tests to Section 3 because of the fact that they generally would be constructed strictly as non-randomized tests rather than as randomized tests.

What we will propose here, however, is a randomized test based on a group of r (randomized) t-statistics each analogous to (39). Let $\hat{\beta}^*_\ell$ denote the ℓ-th element of the vector (36). Let $d_{X\ell}$ and $d_{W\ell}$ be defined by the equation

$$(41) \qquad \operatorname{var}(\hat{\beta}^*_\ell) = d^2_{X\ell}\,\sigma^2_e + d^2_{W\ell}\,\sigma^2_f \qquad ;$$

(41) represents the ℓ-th diagonal element of the matrix (40). We find matrices $E_{X\ell}$ ($M \times [M-r-1]$) and $E_{W\ell}$($N \times [M-r-1]$) satisfying (30), and then define [analogously to (38)] the $(M-r-1)\times 1$ vectors

$$(42) \qquad V_{B\ell} = d_{X\ell}\,E'_{X\ell}\,Y + d_{W\ell}\,E'_{W\ell}\,Z$$

for $\ell = 1, 2, \ldots, r$. (One could of course use the same two matrices for $E_{X\ell}$ and $E_{W\ell}$ for all r values of ℓ, but we attached the ℓ-subscript because we suspect the added flexibility may be advantageous.) For each ℓ, the (marginal) distribution of the statistic

$$(43) \qquad t_\ell = \hat{\beta}^*_\ell \Big/ \left(\frac{V'_{B\ell}\,V_{B\ell}}{M-r-1}\right)^{\frac{1}{2}}$$

will clearly be a t-distribution with $(M-r-1)$ d.f. if $\beta_{Y\ell} = \beta_{Z\ell}$.

At this point we utilize a simple device similar to one described (e.g.) by Dunn in [2]. Let $t_{\alpha/r}$ denote the point such that $100\,[1-(\alpha/r)]$ % of the t-distribution with $(M-r-1)$ d.f. lies between $-t_{\alpha/r}$ and $+t_{\alpha/r}$. Whatever might be the joint distribution of the r variables t_ℓ (43) under the null hypothesis (4), the probability that the r relations

$$(44) \qquad -t_{\alpha/r} \leq t_\ell \leq +t_{\alpha/r} \qquad (\ell = 1, 2, \ldots, r)$$

are simultaneously satisfied is $\geq 1 - \alpha$ if H_{OB} (4) is true. Hence we will have a test of size $\leq \alpha$ if we take as our critical region the complement of (44): i.e., our test is to reject H_{OB} (4) if one or more of the t_ℓ's (43) exceeds $t_{\alpha/r}$ in absolute value.

Simultaneous confidence bounds associated with this test (i.e., bounds on the r elements of β) are easily obtained.

3. NON-RANDOMIZED TESTS

All the tests of Section 2 we consider as being randomized tests, because of the fact that randomization should apparently play a role in the choice of the arbitrarily determined E-matrices appearing in (31), (38), and (42). Although certain optimal properties are associated with the tests of Section 2, many experimenters may nevertheless prefer not to use these tests, either because of a general objection to randomized (or otherwise arbitrary) test statistics, or else because of the computational labor required. Hence we try to develop, here in Section 3, some tests which can be used in lieu of the randomized tests of Section 2, and which (i) are non-randomized, (ii) require less computational labor, and (iii) come as close as possible to the optimal standards of the tests of the previous section. In form, the sub-optimal and non-randomized tests of the present section resemble strongly the optimal and randomized test of Scheffé [7], and in fact were inspired by the latter; but the resemblance is in form only and not in method of attack. The <u>method</u> of Scheffé [7] (i.e., his rationale for selecting a test) influenced the development of the previous section, whereas the <u>form</u> of Scheffé's test [7] influenced the present section.

3.1. Problem A for r=1.

The form of Scheffé's test [7] for the Behrens-Fisher problem is based on a randomized pairing of the two samples. In similar fashion, each test here in Section 3 will be based on a pairing of the two samples, but these pairings will all be non-randomized rather than randomized. For Problem A, we will develop a test of H_{0A} (2) based on the M variables

$$(45) \qquad T_{Ai} = -Y_i + (M/N)^{\frac{1}{2}} Z_{(i)} - (1/MN)^{\frac{1}{2}} \sum_{I=1}^{M} Z_{(I)} + (1/N) \sum_{J=1}^{N} Z_J \quad ,$$

where $Z_{(i)}$ denotes the Z-observation which is paired with Y_i $(i = 1,2,\ldots,M)$. Observe that (45) is formally equivalent with Scheffé's $(-d_i)$ (see p. 37 of [7]).

The M T_{Ai}'s (45) will be normal, mutually independent, and homoscedastic with variance

$$(46) \qquad \text{var}\,(T_{Ai}) = \sigma_e^2 + (M/N)\,\sigma_f^2 = M\,\sigma_A^2 \quad .$$

[Note the point that T_{Ai} is such that its variance (46) is equal to a known constant times (28).] Considering Problem A for the case of general r for the time being, we can write

$$(47) \qquad E(T_{Ai}) = \alpha + \beta_1 U_{i1} + \beta_2 U_{i2} + \ldots + \beta_r U_{ir} \qquad (i = 1,2,\ldots,M) ,$$

where we define

$$(48) \qquad U_{i\ell} = -X_{i\ell} + (M/N)^{\frac{1}{2}} W_{(i)\ell} - (1/MN)^{\frac{1}{2}} \sum_{I=1}^{M} W_{(I)\ell} + (1/N) \sum_{J=1}^{N} W_{J\ell} \quad ,$$

$(W_{(i)1}, W_{(i)2}, \ldots, W_{(i)r})$ being the set of r W-variables associated with $Z_{(i)}$. We now conclude that all the conditions for the general linear model (or, more specifically, for the multiple regression model) are fulfilled, and so we can obtain a t-test of H_{0A} (2) (i.e., of $\alpha = 0$) by applying standard multiple regression theory to the T_{Ai}'s (45) and $U_{i\ell}$'s (48).

Using $\hat{\alpha}'$ to denote the least-squares estimator of α under the model (47), we write the formula

$$(49) \qquad \hat{\alpha}' = (\bar{Z} - \bar{Y}) - \bar{U}' [u'u]^{-1} u' T_A \quad .$$

[The notation in (49) is analogous to that defined at the start of Sub-section 2.1: T_A (Mx1) contains the T_{Ai}'s (45); and $u(M\times r) = U - j_M \bar{U}'$, where $U(M\times r)$ contains the $U_{i\ell}$'s (48). Note that the mean of the T_{Ai}'s (45) is $\bar{T}_A = \bar{Z} - \bar{Y}$, and the r means of the $U_{i\ell}$'s (48) for the r values of ℓ (i.e., the r quantities $M^{-1} \Sigma_{i=1}^{M} U_{i\ell}$) are given by the previously-defined vector $\bar{U}$ (rx1) $= \bar{W} - \bar{X}$.] The variance of the estimator (49) is

$$(50) \qquad \text{var}(\hat{\alpha}') = \{(1/M) + \bar{U}' [u'u]^{-1} \bar{U}\} \text{ var}(T_{Ai}) \quad ,$$

which [by virtue of (46)] is equal to $s'\sigma_A^2$, where

$$(51) \qquad s' = 1 + M \bar{U}' [u'u]^{-1} \bar{U} \quad .$$

Using (49) and (50), we easily find that

$$(52) \qquad t = \frac{\{(\bar{Z} - \bar{Y}) - \bar{U}' [u'u]^{-1} u' T_A\} (M-r-1)^{\frac{1}{2}}}{\{(1/M) + \bar{U}' [u'u]^{-1} \bar{U}\}^{\frac{1}{2}} \{T_A' T_A - M \bar{T}_A^2 - T_A' u[u'u]^{-1} u' T_A\}^{\frac{1}{2}}}$$

is the t-statistic (d.f. = M-r-1) for testing H_{0A} (2). If we define W^o to be an Mxr matrix whose i-th row consists of $(W_{(i)1}, W_{(i)2}, \ldots, W_{(i)r})$ and then define $w^o(M\times r) = W^o - j_M \bar{W}^{o'}$ (where $\bar{W}^{o'}$ contains the column means of the matrix W^o), and if we let Z^o (Mx1) be a vector containing the M $Z_{(i)}$'s and let $\bar{Z}^o$ be their mean, then we can utilize (45) and (48) to write the following formulas which one might want to employ in computing t (52):

(53a) $u'u = x'x + (M/N)\, w^{o'} w^{o} - (M/N)^{\frac{1}{2}} x' w^{o} - (M/N)^{\frac{1}{2}} w^{o'} x$;

(53b) $u'T_A = x'Y + (M/N)\, w^{o'} Z^{o} - (M/N)^{\frac{1}{2}} x' Z^{o} - (M/N)^{\frac{1}{2}} w^{o'} Y$;

(53c) $T_A'T_A - M\bar{T}_A^2 = (Y'Y - M\bar{Y}^2) + (M/N)\,(Z^{o'}Z^{o} - M\,\bar{Z}^{o^2}) - 2(M/N)^{\frac{1}{2}}\,(Y'Z^{o} - M\,\bar{Y}\,\bar{Z}^{o})$.

This takes care of everything except for the question of how the pairing is to be done in the first place. The t-test using (52), which is a standard application of least squares theory based on the model (47), is of course valid for any pairing whatsoever. In the case of Scheffé's test [7] for the Behrens-Fisher problem, there is no reason for preferring one pairing to any other, and so a pairing has to be chosen at random. But in the present situation, by contrast, the different pairings will not all be of equal desirability. Therefore it would seem logical to choose that pairing which minimizes var $(\hat{\alpha}')$ (50).

For the special case r=1, to which we now finally turn our attention, this optimal pairing is not too hard to find. Looking at (50) and (53a), we see that, when r=1, var $(\hat{\alpha}')$ will be minimized for that pairing which maximizes

(54) $$(M/N)^{\frac{1}{2}} \sum_{i=1}^{M} (W_{(i)} - \bar{W}^{o})^2 - 2 \sum_{i=1}^{M} (X_i - \bar{X})\,(W_{(i)} - \bar{W}^{o}) \quad .$$

We pause briefly for some preliminaries. No generality will be lost if we assume that $X_i \leq X_{i+1}$ for all i and $W_j \leq W_{j+1}$ for all j. We will go further than this, however, and assume also that no two X_i's are alike and no two W_j's are alike, so that $X_1 < X_2 < \ldots < X_M$ and $W_1 < W_2 < \ldots < W_N$. [If this assumption of no "ties" is not observed, then some device (preferably something other than randomization, if possible) may have to be used to break the ties.]

In order to find the pairing which maximizes (54), we first establish two initial steps:

(I.) Of the N W_j's, $(N-M)$ will be unpaired and M will be paired; given the group of M W_j's which is to be paired, the best pairing is the one such that $W_{(1)} > W_{(2)} > \ldots > W_{(M)}$ (i.e., the W's are arranged in order opposite to the X's).

<u>Proof.</u> It will be sufficient to show that, if the M W_j's are arranged in any order other than the one just indicated, then it is possible to find a better pairing. If the W's are not in order opposite to the X's, then there exists (i, I) such that $i < I$ (i.e., $X_i < X_I$) and $W_{(i)} < W_{(I)}$. If we consider now the pairing which results from trading $W_{(i)}$ with $W_{(I)}$ and keeping everything else the same, it is not hard to demonstrate that (54) must be larger for the new pairing than for the original pairing. That completes the proof.

(II.) The $(N-M)$ unpaired W_j's must be consecutive. In other words, this means [thanks to (I.)] that the optimal pairing is of the form

$$(55) \qquad W_{(i)} = W_{N+1-i} \quad , \qquad 1 \leq i \leq \nu$$
$$= W_{M+1-i} \quad , \qquad \nu+1 \leq i \leq M \quad ,$$

where ν is yet to be determined.

<u>Proof.</u> If the $(N-M)$ unpaired W's are <u>not</u> consecutive, then there exist two integers $j < J$ such that W_J is paired but W_j and W_{J+1} are both unpaired. We assume that the paired W's are arranged in optimal order [see (I.)]. Then, considering all the W's except W_J as fixed, we may write (54) as a function of W_J, say $\varphi(W_J)$. Clearly, φ is quadratic in W_J, i.e., $\varphi(W_J) = \gamma_2 W_J^2 + \gamma_1 W_J + \gamma_0$ (say). Furthermore, $\gamma_2 = (M-1)/(MN)^2$. Since $\gamma_2 > 0$ and since $W_j < W_J < W_{J+1}$, it

follows that either $\varphi(W_j) > \varphi(W_J)$ or else $\varphi(W_{J+1}) > \varphi(W_J)$ (or both). Hence (54) can be improved upon by moving W_J from the paired to the unpaired group and then replacing it with W_j or W_{J+1} (one if not the other). Thus we conclude that a non-consecutive choice of the unpaired W's cannot be optimal, and so the optimal pairing must be of the form (55).

It remains only to determine ν in (55). Let θ_ν denote the value of (54) for the pairing (55). Then it is easy to show that

$$(56) \qquad \theta_{\nu+1} - \theta_\nu = 2(M-1)\,(MN)^{-\frac{1}{2}}\,(W_{N-\nu} - W_{M-\nu})\,\delta_\nu \qquad ,$$

where

$$(57) \qquad \delta_\nu = \frac{W_{N-\nu} + W_{M-\nu}}{2} - \frac{\sum_{j=1}^{M-\nu-1} W_j + \sum_{j=N-\nu+1}^{N} W_j}{M-1} - \frac{(MN)^{\frac{1}{2}}}{M-1}(X_{\nu+1} - \bar{X}) \;.$$

The coefficient of δ_ν in (56) is > 0, and δ_ν itself (57) is clearly a decreasing function of ν which turns from positive to negative somewhere between $\nu = 0$ and $\nu = M - 1$. Hence the optimal ν to use in (55) is

(58) $\qquad \nu$ (optimal) = the smallest integer ν such that δ_ν (57) is < 0.

Although this value of ν (58) should not be too hard to find, the experimenter might prefer to settle for a slightly sub-optimal ν in order to reduce the calculations required. It would be perfectly legitimate, e.g., to arbitrarily set $\nu = M/2$ if M is even or $= (M-1)/2$ (say) if M is odd.

For (52) as well as for the other tests to be given here in Section 3, it is no problem to obtain confidence bounds associated with the tests, even though (just as in Section 2) we are putting our main expository emphasis on the tests themselves.

Our non-randomized t-test of H_{0A} (2) for r=1, which is specified by (52-53) with the pairing made according to (55, 58), will of course be somewhat less than optimal in comparison with the randomized t-test of Section 2. Since $\mathrm{var}(\hat{\alpha}')$ (50)

and var $(\hat{\alpha}_{C*})$ (28) are each equal to σ_A^2 times a known constant, it is possible to compare the two tests by comparing these two known constants (s' and s respectively). For the case r=1, s is given by (34); and, by using (51) and (53a), we can write s' (for r=1) in the form

$$(59) \qquad s' = 1 + \frac{(\overline{W} - \overline{X})^2}{\sigma_X^2 + \sigma_{W^o}^2 - 2\rho\,\sigma_X\,\sigma_{W^o}} ,$$

where $\sigma_{W^o}^2 = w^{o\prime} w^o/N$ (note that $\sigma_{W^o}^2 \leq \sigma_W^2$) and where ρ is the correlation coefficient between the X_i's and the $W_{(i)}$'s. Offhand, it appears that ρ should be quite close to -1, and that $\sigma_{W^o}^2$ should not be much less than σ_W^2 unless N is considerably larger than M. Based on this, it looks as though s' (59) would normally be very little greater than s (34). Thus we would conclude that, for the case r=1, the randomized t-test of H_{OA} usually has little to recommend it over the non-randomized t-test (since the latter has the two advantages of being non-randomized and easier to calculate).

3.2. Problem A for general r.

Except for the question of the pairing, the non-randomized t-test for Problem A for general r was already completely determined in Sub-section 3.1 [see (52)]. In choosing a pairing, our objective will be to make $\overline{U}' [u'u]^{-1} \overline{U}$ as small as possible . From a strictly theoretical standpoint, this presents no problem at all: there are only a finite number of possible pairings [N!/(N-M)!, in fact], and we simply select the one for which $\overline{U}'[u'u]^{-1} \overline{U}$ is a minimum (assuming that the minimum is unique).

From a practical standpoint, however, the computation involved in evaluating

$\overline{U}'[u'u]^{-1}\overline{U}$ for $N!/(N-M)!$ different pairings would generally be prohibitive, except that perhaps in some cases it might be feasible with a large computer. We have not discovered a direct systematic way of finding the optimal pairing for the case of general r as we did for the case $r=1$.

Hence it may be necessary to settle for a somewhat sub-optimal pairing in order to avoid massive calculations. Various artificial techniques for obtaining such a pairing could undoubtedly be devised. We mention here one possibility which appears intuitively to be a reasonable technique. Define

$$\xi(i,j) = \sum_{\ell=1}^{r} \left[\frac{(\overline{W}_\ell - \overline{X}_\ell)}{(M/N)^{\frac{1}{2}} (W_{j\ell} - \overline{W}_\ell) - (X_{i\ell} - \overline{X}_\ell)} \right]^2 . \tag{60}$$

Let $\xi(i_1,j_1)$ denote the smallest of the MN $\xi(i,j)$'s. Now discard all of the $\xi(i,j)$'s for which $i = i_1$ or $j = j_1$, and let $\xi(i_2,j_2)$ denote the smallest of the remaining $(M-1)(N-1)$ $\xi(i,j)$'s. We continue in similar fashion, discarding all of the $\xi(i,j)$'s for which $i = i_2$ or $j = j_2$ and then letting $\xi(i_3, j_3)$ denote the smallest of the remaining $(M-2)(N-2)$ $\xi(i,j)$'s. Keeping on like this, we end up finally with the M couples (i_1, j_1), (i_2,j_2), ..., (i_M, j_M), and we then pair the two samples by setting $W_{(i_I)} = W_{j_I}$ $(I = 1,2,\ldots,M)$.

The technique just described is certainly feasible with a computer, and in some cases the MN $\xi(i,j)$'s (60) could be calculated by hand. It would appear that the technique should generally lead us to a fairly low value of s' (51), but in order to find out how effective it (or any other technique) really is, we would of course have to compare s' (51) with s (29).

From (51) and (29) it appears that, in general, the problem of pairing will be less crucial (and also the values of s' and s will be closer to 1) the smaller the elements of $\overline{U}$ are in absolute value. Thus the experimenter should normally try to make the elements of $\bar{U}$ as small as possible in absolute value if he has any control over the matter.

3.3 Problem B for r=1.

Since what we do in this sub-section will resemble in many ways what we did in Sub-section 3.1, we will omit some of the details. A natural mode of attack in trying to obtain a non-randomized t-test of H_{OB} (4) for r=1 is to utilize the M independent variables

$$(61) \qquad T_{Bi} = \mp (1/M^{\frac{1}{2}} \sigma_X)Y_i + (1/N^{\frac{1}{2}} \sigma_W) [Z_{(i)} - (1/M) \sum_{I=1}^{M} Z_{(I)} + (1/MN)^{\frac{1}{2}} \sum_{J=1}^{N} Z_J],$$

for these T_{Bi}'s (61) have a formula analogous to (45) and have variance equal to σ_B^2 (37). The ($\mp$) in (61) means that either sign may be chosen; the choice of sign would generally be made in line with the aim of getting the lowest possible variance of the estimator of $\beta = \beta_Z - \beta_Y$. The expectation of each T_{Bi} (61) can be written as a linear function of β_Y, β_Z, and a third parameter.

The t-test of H_{OB} (4) which results from this model has (M-3) d.f. and uses the statistic

$$(62) \qquad t = \frac{\hat{\beta} \ (M - 3)^{\frac{1}{2}}}{[\psi(R,\rho)]^{\frac{1}{2}} [T_B' T_B - M \bar{T}_B^2 - \hat{\beta}_Y(u_Y' T_B) - \hat{\beta}_Z(u_Z' T_B)]^{\frac{1}{2}}} ,$$

where

$$(63) \qquad \hat{\beta} = \hat{\beta}_Z - \hat{\beta}_Y ,$$

$$(64a) \qquad \hat{\beta}_Y = (1 - \rho^2)^{-1} [(u'_Y T_B) \pm \rho R (u_Z' T_B)] ,$$

$$(64b) \qquad \hat{\beta}_Z = (1 - \rho^2)^{-1} [\pm \rho R (u_Y' T_B) + R^2 (u_Z' T_B)] ,$$

(65a) $$T_B' T_B - M \bar{T}_B^2 = \frac{Y'Y - M \bar{Y}^2}{M \sigma_X^2} + \frac{Z^{o\prime} Z^o - M \bar{Z}^{o2}}{N \sigma_W^2} \mp 2 \frac{Y'Z^o - M \bar{Y} \bar{Z}^o}{(MN)^{\frac{1}{2}} \sigma_X \sigma_W} ,$$

(65b) $$u'_Y T_B = (1/M \sigma_X^2) x' Y \mp (1/MN)^{\frac{1}{2}} (1/\sigma_X \sigma_W) x' Z^o ,$$

(65c) $$u'_Z T_B = \mp (1/MN)^{\frac{1}{2}} (1/\sigma_X \sigma_W) w^{o\prime} Y + (1/N \sigma_W^2) w^{o\prime} Z^o ,$$

(66) $$\psi(R,\rho) = \frac{1 + R^2 \mp 2 \rho R}{1 - \rho^2} = \frac{(R-1)^2}{2(1 \mp \rho)} + \frac{(R+1)^2}{2(1 \pm \rho)} = \frac{\mathrm{var}(\hat{\beta})}{\sigma_B^2} ,$$

and

(67) $$R = \sigma_W/\sigma_{W^o} = (w'w/w^{o\prime} w^o)^{\frac{1}{2}} .$$

The optimal choice of pairing and of $(\mp)$ sign for this t-test [there are $2N!/(N-M)!$ possible choices altogether] would be the one for which $\psi(R,\rho)$ (66) is a minimum.

This minimum ψ could be located by (e.g.) evaluating all $2N!/(N-M)!$ possibilities, but there appears to be no simple systematic way of locating it. Nor does there seem to be available a simple and completely-defined method for arriving even at an approximate solution. The difficulty stems partly from the tricky nature of the function $\psi(R,\rho)$. For fixed R, $\psi(R, \rho)$ (66) is minimal at the point $\rho = \pm 1/R$ [at which point we have $\psi(R, \pm 1/R) = R^2$]; in the short interval on the ρ-axis from $\pm 1/R$ to ± 1, $\psi(R,\rho)$ increases rapidly and without bound, thereby creating something of a trap.

If it is not feasible to locate the exact minimum of ψ, one might be inclined to use the following pairing method to obtain an approximate solution:

(i) Choose the group of $(N-M)$ unpaired W_j's in such a way as to minimize R (67) (i.e., maximize $w^{o'}w^{o}$). This minimum R is attained by taking the unpaired W_j's to be $(W_{M-\nu+1}, W_{M-\nu+2}, \ldots, W_{N-\nu})$, where ν is the smallest integer such that

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

is < 0 [proof is along the same lines as with (II.) and (56-58) of Sub-section 3.1].

(ii) Pair the M X's and the M chosen W's in such a way as to maximize $|\rho|$. This maximum $|\rho|$ (which we will designate by $|\rho_o|$) is achieved by arranging the X's and W's in either the same or else the opposite order [proof is like that for (I.) of Sub-section 3.1], depending on which order gives the larger $|\rho|$; then the upper + or - sign (for same order) or lower sign (for opposite order) is selected everywhere in (61, 64-66).

This pairing method would appear to be quite satisfactory so long as $|\rho_o|$ turns out to be $\leq 1/R$. But if $|\rho_o|$ exceeds $1/R$ (an event which may or may not be much more than just a theoretical possibility), then ψ could become enormous. Of course a pairing could be found which would bring ρ close to $1/R$ in absolute value, but it would be necessary to devise a completely-defined technique for arriving at such a pairing.

In order to provide a simple and well-defined device for circumventing the problem just raised, we propose the following addendum to the pairing method (i-ii) which will now make the method completely clean:

(iii) In case $|\rho_o|$ turns out to be $> 1/R$, substitute $\sigma_{W^o}|\rho_o|^{-1}$ for σ_W wherever the latter appears in (61, 65), and replace R by $|\rho_o|^{-1}$ wherever R appears in (62, 64, 66). [Everything else in (61-66) remains the same, and the validity of the resulting new t-test is easily verified.]

Under this change (iii), the variance of the new estimator of β is

$$(68) \qquad \psi(|\rho_o|^{-1}, \rho_o) \ \text{var}\,(\text{new } T_{Bi}) = \rho_o^{-2}\left[\frac{\sigma_e^2}{x'x} + (\rho_o^2 R^2)\frac{\sigma_f^2}{w'w}\right].$$

Therefore, the danger of ψ exploding no longer exists; however, (68) is not precisely equal to a known constant times σ_B^2, which may be a bit of a drawback.

Note that everything simplifies considerably in the special case M=N. When M=N, R (67) is automatically equal to 1, so that (i) and (iii) become vacuous. Minimizing $\psi(1,\rho) = 2/(1 \pm \rho)$ then becomes a question merely of determining [see (ii)] whether the X's and W's should be arranged for pairing in the same order or the opposite order to obtain the maximum $|\rho|$.

The type of non-randomized t-test proposed in this sub-section [regardless of whether it is based on the scheme (i - iii), or on some technique to locate the pairing which gives the exact minimum of $\psi(R,\rho)$] will be slightly inferior to the randomized t-test of Sub-section 2.3 in two different respects: there will be one less degree of freedom [(M-3) versus (M-2)], and the variance of the estimator of β will generally be slightly larger [but not much larger, since $\psi(R,\rho)$ should (for the case $|\rho_o| \leq 1/R$) normally be close to 1]. However, it would appear that these two slight disadvantages would usually be much more than offset by the two advantages of the test of the present sub-section: it is easier to calculate [if the scheme (i - iii) is used] and is non-randomized.

3.4. Problem B for general r.

In principle, it is easy enough to generalize the non-randomized t-statistic (62) and obtain a set of r non-randomized t-statistics akin to the set (43) of randomized t's, in order to test H_{OB} (4) for general r by using (as in Sub-section 2.4) a critical region of size $\leq \alpha$ which is constructed as the union of r critical regions each of size α/r. More specifically, the ℓ-th such critical region ($\ell = 1, 2, \ldots, r$) may be based on the non-randomized t-test (d.f. = M - 2r - 1) of $\beta_{Y\ell} = \beta_{Z\ell}$ that is obtained by applying standard regression theory to the M independent variables

$$(69) \quad T_{Bi\ell} = \mp\ d_{X\ell} Y_i + d_{W\ell} [Z_{(i\ell)} - (1/M) \sum_{I=1}^{M} Z_{(I\ell)} + (1/MN)^{\frac{1}{2}} \sum_{J=1}^{N} Z_J] \quad ,$$

which are so defined as to be homoscedastic with variance equal to (41). [The subscript ℓ is included in $Z_{(i\ell)}$ in (69) in order to allow a different pairing to be used for each of the r values of ℓ.] Thus each one of the r groups of M $T_{Bi\ell}$'s (69) is used to compute one t-statistic (the ℓ-th group of M $T_{Bi\ell}$'s being tailored specifically for the t-test of $\beta_{Y\ell} = \beta_{Z\ell}$), in a way similar to that in which the group of M T_{Bi}'s (61) is used to compute the t-statistic (62). We will not spell out here the detailed formula for the r t-statistics, but this formula is obtained [just as (62-67) was obtained] by a straightforward application of ordinary multiple regression methods. Once the set of r non-randomized t-statistics has been calculated, the remainder of the procedure for testing H_{OB} (4) is of course exactly the same as in Sub-section 2.4.

There remains, however, the problem of how to choose (for each ℓ) the pairing and the ($\mp$) sign. As before, this is no problem at all in a strictly theoretical sense, because we can simply select (for each ℓ) that particular one of the

$2N!/(N-M)!$ possibilities which minimizes the variance of the estimator of $(\beta_{Z\ell} - \beta_{Y\ell})$ (assuming that the minimum is unique). But from a practical standpoint, such a lengthy technique as this would usually be out of the question. The best alternative method we have to offer is to apply (for each ℓ separately) steps (i - ii) of Sub-section 3.3: for $\ell=2$, for example, we choose the pairing and $(\mp)$ sign by applying (i-ii) of Subsection 3.3 to the M X_{i2}'s and N W_{j2}'s. However, it is not clear whether this method will generally give a test with satisfactory power; we might anticipate possible trouble on the basis of the potential explosiveness of the function $\psi(R, \rho)$ in Sub-section 3.3, but as yet we have found no way to generalize (iii) of Sub-section 3.3 in order to provide a means for averting such trouble.

Using a set of r non-randomized t-statistics is not the only way of getting a non-randomized test of H_{OB} (4) for general r. There exists a second way, which was already alluded to briefly in Section 2: we can develop a non-randomized F-test, with d.f. r and $(M - 2r - 1)$. Such an F-test could be based on M variables of the form (69) [but now we would drop the subscript ℓ in (69) everywhere that it appears]; the F-statistic would be calculated via standard regression formulas. But the major problem would be to figure out a satisfactory way of choosing the pairing, the $(\mp)$ sign, and the value of d_W/d_X. Thus we are faced again with the same type of problem which we have encountered many times throughout this paper: we have available to us a huge (or infinite) number of possible tests, all of which are valid; but in order to select from among these a single test, we must have a selection method which will be well-defined, which will not be too cumbersome to use in practice, and which will produce a test for which the power (or some other related criterion) is close to the best possible. We will not attempt to tackle here the unresolved questions pertaining to the matter of how an F-test should be selected for testing H_{OB} (4). This looks like a difficult problem, and, in fact, even the choice of a criterion (by which to judge different possible F-tests) may present difficulties; furthermore, it should be remarked that the F-test approach may or may not lead ultimately to a better test than the approach based on r t-statistics. [Incidentally, one might wish to consider a type of F-test which is more general than the type based on the M variables of the form (69): it obviously is possible to construct a set of M linear functions of the Y's and Z's which are homoscedastic and mutually independent, but which are not of the form (69); and F-tests of H_{OB} (4) can be based on this generalized kind of linear set.]

We should note that, under either of the non-randomized approaches suggested in this section (F-statistic or r t-statistics), there are only $(M - 2r - 1)$ d.f. for error, as compared with $(M-r-1)$ under the randomized approach of Sub-section 2.4. This may be a serious difference if r is relatively large and/or if M is relatively small.

For the case $r=1$, we feel confident that, for both Problems A and B, our non-randomized tests of Section 3 will almost always be preferable to our randomized tests of Section 2. For the case of Problem A for general r, the situation regarding the non-randomized test of Sub-section 3.2 based on the $\xi(i,j)$'s (60) certainly requires further investigation (perhaps of a practical rather than of a theoretical nature), but it should not be too surprising if this non-randomized test turns out in most cases to be entirely satisfactory (in comparison with the randomized test of Section 2.1). For the case of Problem B for general r, however, the randomized test of Section 2.4 may often have some pronounced advantages over any non-randomized test which is presently available, since we were not able to arrive at any very decisive results pertaining to non-randomized tests here in this final sub-section; in fact, the questions raised in this final sub-section seem to provide the greatest opportunity for further study.

4. ACKNOWLEDGMENT

The author wishes to thank Dr. Frederic M. Lord of Educational Testing Service for suggesting the problems concerning comparisons of two regression lines when the variances may be unequal.

REFERENCES

[1] Anderson, T. W., An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, 1958.

[2] Dunn, Olive J., "Multiple Comparisons Among Means", Journal of the American Statistical Association, 56 (1961), 52-64.

[3] Hájek, Jaroslav, "Inequalities for the Generalized Student's Distribution and Their Applications," Selected Translations in Mathematical Statistics and Probability, 2 (1962), 63-74.

[4] Mazuy, Kay K., and Connor, W. S., "Student's t in a Two-Way Classification with Unequal Variances," Technical Report No. 3, Statistics Research Division, Research Triangle Institute, Durham, North Carolina, 1963.

[5] Potthoff, Richard F., "A Test of Whether Two Regression Lines Are Parallel When the Variances May Be Unequal," Mimeo Series No. 331, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina, 1962.

[6] Potthoff, Richard F., "A Test of Whether Two Parallel Regression Lines Are the Same When the Variances May Be Unequal," Mimeo Series No. 334, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina, 1962.

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

[8] Welch, B. L., "The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved," Biometrika, 34 (1947), 28-35.