RANK ANALYSIS OF COVARIANCE

by

Dana Quade

University of North Carolina

Institute of Statistics Mimeo Series No. 483

July 1966

This research was supported by the National Institutes of Health Grants No. GM-10397 and GM-12868-01.

DEPARTMENT OF BIOSTATISTICS

UNIVERSITY OF NORTH CAROLINA

Chapel Hill, N. C.

RANK ANALYSIS OF COVARIANCE[1]

Dana Quade
University of North Carolina

Various methods are discussed for the problem of comparing two or more populations with respect to a response variable Y in the presence of a (possibly multivariate) concomitant variable X - a situation in which the standard method is the classical one-way analysis of covariance. A method based on ranks is developed.

1. INTRODUCTION

Suppose that from each of m populations we have a random sample of observations (Y_{ia}, X_{ia}), where Y_{ia} is the univariate response of the *ith* observation in the *ath* sample $(1 \leq i \leq n_a, 1 \leq a \leq m)$, and X_{ia} is the corresponding value of a concomitant variable, possibly multivariate, whose marginal probability distribution is the same in each population. The problem is then to test the hypothesis H_o that the conditional distribution of Y given X is also the same for each population, where the alternatives of interest are those which imply that some populations tend to have greater values of Y then others for all fixed values of X.

For such situations let us invoke a very general principle which may be described as follows. If the hypothesis is true then the populations are all identical and the samples can be pooled. Use the pooled sample to determine a relationship by which Y can be predicted from X. Then compare

[1]Supported by National Institutes of Health Grants No. GM-10397 and GM-12868-01.

each observed response Y_{ia} with the value which would be predicted for it from the corresponding X_{ia}, and assign it a score Z_{ia}, positive if Y_{ia} is greater than predicted, and negative if smaller. Finally, use the scores to perform an ordinary one-way analysis of variance for comparing the populations.

As a particular instance of this general method I propose a procedure called rank analysis of covariance. Assume, in order to make ranking possible, that each variate has been measured on at least an ordinal scale; continuity is not required, however, and even a dichotomy is permitted as an extreme case. So let the rank of Y_{ia} among all the $N = \Sigma n_a$ observed values of Y be $R_{ia} + (N+1)/2$, where the term $(N+1)/2$ has been inserted for convenience so that $\Sigma\Sigma R_{ia} = 0$, thus correcting the ranks for their mean; use "average ranks" in case of ties, and (for definiteness) rank from the smallest first. Similarly, if X is actually a p-variate variable $(X^{(1)},X^{(2)},\ldots,X^{(p)})$, let $C_{ia}^{(k)} + (N+1)/2$ be the rank of $X_{ia}^{(k)}$ among the N observed values of $X^{(k)}$. Then characterize the relationship between Y and X by performing an ordinary multiple linear regression of R on $C^{(1)},C^{(2)},\ldots,C^{(p)}$; calculate fitted values $\hat{R}_{ia}$, and assign as scores the residuals from this regression of ranks: i.e., let

$$Z_{ia} = R_{ia} - \hat{R}_{ia} .$$

Finally, to test the hypothesis of identical conditional distributions of Y on X use as criterion the variance ratio

$$VR = \frac{(N-m)\ \sum_a(\sum_i Z_{ia})^2/n_a}{(m-1)[\sum_{ai}\sum Z_{ia}^2 - \sum_a(\sum_i Z_{ia})^2/n_a]} ,$$

comparing it with the critical value of an F with (m-1, N-m) degrees of freedom. (Note that no explicit correction for the mean is required in VR since $\sum_{a}\sum_{i} Z_{ia} = \bar{\bar{Z}} = 0$.)

As a numerical example consider the (artificial) data of Table 1.

Table 1

Sample	Y	X_1	X_2	R	C_1	C_2	$\hat{R}$	Z
	16	26	12	-7	-1	-3	-1.72	-5.28
	60	10	21	-3	-5	0	-2.90	- .10
1	82	42	24	-1	3	1	2.12	-3.12
	126	49	29	3	5	3	4.04	-1.04
	137	55	34	4	6	4	5.00	-1.00
	44	21	17	-4	-2	-1	-1.54	-2.46
	67	28	2	-2	0	-6	-2.28	.28
2	87	5	40	0	-6	7	- .82	.82
	100	12	38	1	-4	6	- .04	1.04
	142	58	36	5	7	5	5.96	- .96
	17	1	8	-6	-7	-5	-5.96	- .04
	28	19	1	-5	-3	-7	-4.40	- .60
3	105	41	9	2	2	-4	- .36	2.36
	149	48	28	6	4	2	3.08	2.92
	160	35	16	7	1	-2	- .18	7.18
Sum	1320	450	315	0	0	0	0	0
Sum of squares	149282	18296	8977	280	280	280	165.48	114.52

Write R for the N×1 matrix of corrected ranks of responses, and C for the N×p matrix (here p=2) of corrected ranks of concomitant variates; then the N×1 matrix of fitted values is

$$\hat{R} = C(C'C)^{-1}C'R ,$$

where for this example

$$C'C = \begin{pmatrix} 280 & 70 \\ 70 & 280 \end{pmatrix}, \qquad C'R = \begin{pmatrix} 189 \\ 147 \end{pmatrix}, \qquad (C'C)^{-1}C'R = \begin{pmatrix} .58 \\ .38 \end{pmatrix},$$

and Table 1 shows $\hat{R}$ and $Z = R - \hat{R}$. The totals of the scores for the m=3 samples are

$$\sum_i Z_{i1} = -10.54 , \qquad \sum_i Z_{i2} = -1.28 , \qquad \sum_i Z_{i3} = 11.82 ,$$

and $\Sigma\Sigma Z_{ia}^2 = 114.52$, so that finally VR = 4.73, corresponding to an aposteriori level of significance $P \doteq .03$ using the F distribution with (2,12) degrees of freedom. It may be noted that application of the usual preliminary procedures produces no evidence that the assumptions required for the parametric analysis of covariance are not satisfied, and the standard test yields $F(2,10) = 3.38$, $P > .075$.

2. THEORY

Let us continue with the case where the concomitant variable X is p-dimensional. For any vector $\underline{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_p)'$ let $T(\underline{\lambda})$ be the test which rejects H_o for large values of

$$VR(\underline{\lambda}) = \frac{(N-m) \sum_a \left(\sum_i Z_{ia}(\underline{\lambda})\right)^2 / n_a}{(m-1) \left[\sum_{ai}\sum Z_{ia}^2(\underline{\lambda}) - \sum_a \left(\sum_i Z_{ia}(\underline{\lambda})\right)^2 / n_a \right]}$$

where

$$Z_{ia}(\underline{\lambda}) = R_{ia} - \sum_k \lambda_k C_{ia}^{(k)} .$$

Rank analysis of covariance is then the special case $T(\underline{t})$ where $\underline{t} = (C'C)^{-1}C'R$. But for the present let us suppose that $\underline{\lambda}$ is fixed (and hence in particular is <u>not</u> $\underline{t}$); and consider the mean and variance of a score $Z_{ia}(\underline{\lambda})$.

Let

$$\psi(u) = \begin{cases} \tfrac{1}{2} & u > 0 \\ 0 & u = 0 \\ -\tfrac{1}{2} & u < 0 \end{cases} ,$$

so that

$$R_{ia} = \sum_{bj}\sum \psi(Y_{ia}-Y_{jb}) .$$

Now

$$E[\psi(Y_{ia}-Y_{jb})] = \tfrac{1}{2}P\{Y_a > Y_b\} - \tfrac{1}{2}P\{Y_a < Y_b\}$$

where Y_a and Y_b are independent random observations from the marginal distributions of Y in the a<u>th</u> and b<u>th</u> populations, respectively; write

$$E[\psi(Y_{ia}-Y_{jb})] = \Theta_{ab} ,$$

and note that

$$\Theta_{ab} + \Theta_{ba} = 0, \quad \Theta_{aa} = 0, \quad \text{for } 1 \le a, b \le m.$$

Define

$$\Theta_a = \sum_b p_b \Theta_{ab} , \quad \text{where} \quad p_b = n_b/N;$$

then

$$E[R_{ia}] = N\Theta_a , \quad 1 \le a \le m.$$

Similarly, write

$$C_{ia}^{(k)} = \sum_{bj}\sum \psi(X_{ia}^{(k)} - X_{jb}^{(k)}) ;$$

but

$$E[C_{ia}^{(k)}] \equiv 0$$

since $X_{ia}^{(k)}$ and $X_{jb}^{(k)}$ are random observations from identical populations for any a and b. Hence, for all $\underline{\lambda}$,

$$E[Z_{ia}(\underline{\lambda})] = N\Theta_a .$$

Now let $(N^2-1)\Sigma/12$ be the p×p variance matrix of the ranks of the concomitant variates; if these variates are continuous then the diagonal entries of Σ will be equal to 1 and Σ will in fact be the correlation matrix; at any rate the diagonal entries will be no greater than 1. Let also

$(N^2-1)\sigma_a^2/12$ be the variance of the rank of a response from the ath population, for $1 \leq a \leq m$, and let the vector $(N^2-1)\underline{\eta}_a/12$ give the covariances of the ranks of the response with the concomitant variates. Then clearly

$$\mathrm{Var}[Z_{ia}(\underline{\lambda})] = \frac{N^2-1}{12}[\sigma_a^2 - 2\underline{\lambda}'\underline{\eta}_a + \underline{\lambda}'\Sigma\underline{\lambda}] .$$

Write

$$\zeta(\underline{\lambda}) = \frac{1}{12}\sum_a P_a[\sigma_a^2 - 2\underline{\lambda}'\underline{\eta}_a + \underline{\lambda}'\Sigma\underline{\lambda}] ;$$

then under the null hypothesis H_o

$$\mathrm{Var}[Z_{ia}(\underline{\lambda})] = (N^2-1)\zeta(\underline{\lambda})$$

for all a.

If we define

$$\phi\Big((Y_{ia},X_{ia}),(Y_{jb},X_{jb});\ \underline{\lambda}\Big) = \psi(Y_{ia}-Y_{jb}) - \sum_k \lambda_k\psi(X_{ia}^{(k)}-X_{jb}^{(k)} \quad ,$$

then we can express the typical score as

$$Z_{ia}(\underline{\lambda}) = \sum_{bj}\sum\phi\Big((Y_{ia},X_{ia}),(Y_{jb},X_{jb});\ \underline{\lambda}\Big) ;$$

note that because

$$\phi\Big((Y_{ia},X_{ia}),(Y_{jb},X_{jb});\underline{\lambda}\Big) + \phi\Big((Y_{jb},X_{jb}),(Y_{ia},X_{ia});\underline{\lambda}\Big) \equiv 0$$

we have

$$\sum_j\phi\Big((Y_{ia},X_{ia}),(Y_{ja},X_{ja});\underline{\lambda}\Big) = 0$$

also. We now obtain immediately

<u>Theorem 1</u>. Let $N \to \infty$ with $n_a = Np_a$ for fixed $p_a > 0$, $1 \leq a \leq m$; then

$$\frac{s_N^2(\underline{\lambda})}{(N-1)^2} \to \zeta(\underline{\lambda}) - \sum_a p_a \Theta_a^2 \quad \text{in probability,}$$

where $s_N^2(\underline{\lambda}) = \frac{1}{N-m} \left[\sum_a \sum_i Z_{ia}^2(\underline{\lambda}) - \sum_a \left(\sum_i Z_{ia}(\underline{\lambda})\right)^2 / n_a \right]$

and

<u>Theorem 2</u>. Let $N \to \infty$ with $n_a = Np_a$ for fixed $p_a > 0$, $1 \leq a \leq m$, and assume that H_o is true and that $\zeta > 0$; then for every fixed $\underline{\lambda}$ the random variable $VR(\underline{\lambda})$ has asymptotically the F-distribution with (m-1, N-m) degrees of freedom; and the test $T(\underline{\lambda})$ is consistent against any alternative for which $\sum_a p_a \Theta_a^2 > 0$.

These results follow from Theorems 5,6, and 7 of Quade [7], noting that his assumption B2 is satisfied since

$$\left|\phi\left((Y_{ia},X_{ia}),(Y_{jb},X_{jb});\underline{\lambda}\right)\right| \leq \tfrac{1}{2}[1 + \sum_k |\lambda_k|] .$$

Now consider the asymptotic relative efficiency (ARE) of the test $T(\underline{\lambda})$ in the Pitman sense, under the assumption that for each N the alternative H_N is true, where under H_N

$$\Theta_{ab} = \frac{\delta_{ab}}{\sqrt{N}} + o\left(\frac{1}{\sqrt{N}}\right) \qquad \text{for } 1 \leq a, b \leq m .$$

The standard test for comparison will be $T(\underline{0})$, which is easily seen to be a modified version of the familiar Kruskal-Wallis test (KW). (The two tests are related through the equation

$$VR(\underline{0}) = \frac{(N-m)H}{(m-1)(N-1-H)}$$

where

$$H = \frac{12}{N(N+1)} \sum_a (\sum_i R_{ia})^2/n_a$$

is the Kruskal-Wallis statistic, usually treated as a χ^2 with (m-1) degrees of freedom; asymptotically, in particular with respect to ARE, the tests are entirely equivalent.) Supposing also that

$$\text{Var}[Z_{ia}(\underline{\lambda})] = (N^2-1)[\zeta(\underline{\lambda}) + 0(\frac{1}{\sqrt{N}})]$$

for all ia under the sequence of alternatives $\{H_N\}$, the asymptotic distribution of $VR(\underline{\lambda})$ will then be noncentral F with (m-1, N-m) degrees of freedom and noncentrality parameter

$$\Delta(\underline{\lambda}) = \sum_a p_a \delta_a^2/\zeta(\underline{\lambda})$$

where $\delta_a = \sum_b p_b \delta_{ab}$ for $1 \leq a \leq m$. In such cases the ARE is the ratio of the noncentrality parameters: i.e.

$$\text{ARE of } T(\underline{\lambda}) \text{ w.r.t. } T(\underline{0}) \quad (\text{or KW}) = \frac{\Delta(\underline{\lambda})}{\Delta(\underline{0})} = \frac{\sigma^2}{\sigma^2 - 2\underline{\lambda}'\underline{\eta} + \underline{\lambda}'\Sigma\underline{\lambda}}$$

where σ^2 and $\underline{\eta}$ are the values of σ_a^2 and $\underline{\eta}_a$ common to all populations under H_o. Clearly the most efficient test of this type is one in which $\underline{\lambda}$ has been chosen to minimize $(\underline{\lambda}'\Sigma\underline{\lambda} - 2\underline{\lambda}'\underline{\eta})$. In particular, if Σ^{-1} exists, then one should take

$$\underline{\lambda} = \underline{\tau} = \Sigma^{-1}\underline{\eta} .$$

The ARE of $T(\underline{\tau})$ with respect to $T(\underline{0})$ or the Kruskal-Wallis test is then

$$\text{ARE}(T(\underline{\tau})/\text{KW}) = \frac{\sigma^2}{\sigma^2 - \underline{\eta}\Sigma^{-1}\underline{\eta}} = \frac{1}{1-R_S^2} ,$$

say, where R_S might be called the Spearman multiple correlation between the response and the concomitant variates.

The ARE of the test $T(\underline{\tau})$ may also be considered with respect to the classical parametric analysis of covariance procedure in the situation where the latter is appropriate, namely, when the conditional distributions of Y given X are normal with constant variance and with

$$E[Y_{ia}|X_{ia}] = \alpha_a + \sum_k \beta_k X_{ia}^{(k)} \quad ,$$

where under H_o

$$\alpha_a = \alpha \qquad \text{for all } a$$

but under H_N

$$\alpha_a = \alpha + \frac{\delta_a^*}{\sqrt{N}} + o\left(\frac{1}{\sqrt{N}}\right).$$

Then the ARE of the parametric analysis of variance with respect to the parametric analysis of covariance is

$$\text{ARE}(AV/AC) = 1 - R_P^2 \quad ,$$

where R_P is the Pearson product-moment multiple correlation between the response and the concomitant variates; and the ARE of $T(\underline{0})$ or the Kruskal-Wallis test with respect ot the parametric analysis of variance is

$$\text{ARE}(KW/AV) = 3/\pi \;;$$

hence the ARE of the test $T(\underline{\tau})$ with respect to the parametric analysis of covariance is

$$\begin{aligned}\text{ARE}(T(\underline{\tau})/AC) &= \text{ARE}(T(\underline{\tau})/KW) \times \text{ARE}(KW/AV) \times \text{ARE}(AV/AC) \\ &= \frac{3(1-R_P^2)}{\pi(1-R_S^2)} \quad .\end{aligned}$$

When the concomitant variable is univariate this reduces to

$$\text{ARE}(T(\underline{\tau})/AC) = \frac{3(1-\rho_P^2)}{\pi(1-\rho_S^2)}$$

where ρ_P and ρ_S are the Pearson and Spearman correlations between response and concomitant variable, related through the equation

$$\rho_P = 2\sin(\pi\rho_S/6).$$

This ARE has maximum value $3/\pi = .955$ at $\rho_P = \rho_S = 0$ and decreases to $\sqrt{3}/2 = .866$ as ρ_P and ρ_S approach 1 or -1.

In general, of course, the vector $\underline{\tau}$ is unknown; it would appear that the most reasonable thing to do then would be to estimate it. The obvious estimates of Σ and $\underline{\eta}$ are

$$\hat{\Sigma} = \frac{12\ C'C}{N(N^2-1)}\ ,\qquad \hat{\underline{\eta}} = \frac{12\ C'R}{N(N^2-1)}$$

where C and R are as defined at the end of Section 1. Hence, if C'C is nonsingular, take

$$\hat{\underline{\tau}} = (C'C)^{-1}C'R = \underline{t}\quad :$$

i.e., use the rank analysis of covariance test as there proposed. (If C'C is singular take any vector $\underline{t}$ such as to minimize the expression $\underline{t}'C'C\underline{t} - 2\underline{t}'C'R$.)

It may be noted that the elements of $\hat{\Sigma}$ and $\hat{\eta}$ are U-statistics, and the elements of $\underline{t}$ are continuous functions of them. Thus

$$E[\phi((Y_{ia},X_{ia}),(Y_{jb},X_{jb});\underline{\tau}) - \phi((Y_{ia},X_{ia}),(Y_{jb},X_{jb});\underline{t})]^2 = 0\left(\frac{1}{N}\right),$$

which is the Assumption D required for Theorem 8 of [7]; hence it follows that the statistic VR($\underline{t}$) in which the elements of $\underline{\tau}$ have been estimated from the data also has asymptotically the F-distribution with (m-1, N-m) degrees of freedom under the hypothesis. However, the test is valid and consistent for any choice of the vector $\underline{\lambda}$, although its efficiency is reduced as $\underline{\lambda}$ departs from $\underline{\tau}$.

3. RELATED WORK

Previously published work related to the problem of nonparametric analysis of covariance seems to be limited to the situation where there are only two populations to be compared and the concomitant variable is univariate: i.e., m=2, p=1. For purposes of comparison it will be convenient to set down the formulas for rank analysis of covariance in that special case. Corresponding to the bivariate observation (Y_{ia},X_{ia}) we then assign the score

$$Z_{ia} = R_{ia} - t_S C_{ia}$$

for $i=1,2,\ldots,n_a$, $a=1,2$, where

$$t_S = \frac{\Sigma\Sigma R_{ia} C_{ia}}{\Sigma\Sigma C_{ia}^2}$$

is the observed Spearman regression coefficient for the response Y on the concomitant variable X. Let

$$W_a = \sum_i Z_{ia} \ , \quad a = 1,2;$$

then $W_1 = -W_2$ and the variance ratio is

$$VR = \frac{N(N-2)W_1^2}{n_1 n_2 \Sigma\Sigma Z_{ia}^2 - NW_1^2} \ .$$

If there are no ties in the sample we have

$$\Sigma\Sigma Z_{ia}^2 = N(N^2-1)(1-r_S^2)/12$$

and

$$VR = \frac{12(N-2)W_1^2}{n_1 n_2 (N^2-1)(1-r_S^2) - 12W_1^2} \ .$$

At the Fourth Berkeley Symposium, David and Fix [5] proposed several methods based on ranks, including three essentially different ones, for testing the homogeneity of two bivariate random variables (Y_1,X_1) and (Y_2,X_2).

They did not require that the marginal distributions of X_1 and X_2 be the same, but they did assume that the two regression functions are parallel lines: i.e. that

$$E[Y_a|X_a] = \alpha_a + \beta X_a \qquad a = 1,2.$$

For testing

$$H_{o1} : E[Y_1] - E[Y_2] = 0$$

they propose the criterion

$$T_1 = \frac{W_1\sqrt{12}}{n_1\sqrt{N^2-1}} .$$

For small samples they suggest basing the test on the permutation distribution of T_1 given the N pairs of ranks (R_{ia}, C_{ia}), and they carry out the necessary computations for a number of examples. For larger samples they conclude that under H_{o1} the statistic

$$\frac{T_1\sqrt{n_1(N-1)}}{\sqrt{n_2(1-r_S^2)}} = \frac{W_1\sqrt{12}}{\sqrt{n_1 n_2(N+1)(1-r_S^2)}}$$

may be treated as a normal deviate. This proposal is essentially rank analysis of covariance - limited to the case m=2, p=1 - with a somewhat different approximation to the null-hypothesis distribution of the test criterion.

For testing the hypothesis

$$H_{o2} : E[Y_1] - E[Y_2] = E[X_1] - E[X_2] = 0$$

against general alternatives, David and Fix propose the criterion

$$T_2 = \frac{12}{n_1(N^2-1)} \{\Sigma R_{i1}^2 - 2r_S \Sigma R_{i1} C_{i1} + \Sigma C_{i1}^2\} ,$$

and they conclude that

$$R = \frac{n_1(N-1)T_2}{n_2(1-r_S^2)}$$

is distributed approximately as χ^2 with 2 degrees of freedom, at least for n_1 and n_2 both at least equal to 10. The statistic R has been proposed again, apparently independently, by Chatterjee and Sen [4] for the general bivariate two-sample location problem. However, the very generality of the alternatives contemplated by this procedure suggests that it will have relatively low power under our assumption of identical marginals for X_1 and X_2, and we do not discuss it further.

The third proposal of David and Fix, made "on common-sense grounds", is to use a test criterion (their ϕ or Θ) equivalent to

$$T_3 = \Sigma R_{i1} - \Sigma C_{i1} ;$$

under H_{o2} T_3 may be treated as approximately normally distributed with mean 0 and variance $n_1 n_2 (N+1)(1-r_S)/6$. Note that the test based on T_3 is the same as the test which would be called T(1) in the notation of Section 2.

It turns out, as will be seen , that this last test is also closely related to a procedure which was proposed by Bross [3] for the restricted situation where the response is a simple dichotomy, (although its extension to more general responses is immediate). Bross' test, which he calls COVAST ("Covariable Adjusted Sign Test"), has an interesting intuitive justification. Assume that the response and the concomitant variable are positively associated within any one population. Now let (Y_1,X_1) and (Y_2,X_2) be random observations from populations 1 and 2 respectively. Then this pair of observations may be called <u>concordant</u> if $(Y_1-Y_2)(X_1-X_2) > 0$, i.e., if $Y_1 > Y_2$, $X_1 > X_2$ or if $Y_1 < Y_2$, $X_1 < X_2$. The occurrence of such pairs might be explained on the basis of the general positive association between Y and X, without supposing any difference between the two populations. But if $Y_1 > Y_2$, $X_1 < X_2$ then the pair is <u>discordant</u> (Bross calls it an <u>inversion</u>) and provides a clear-cut bit of evidence indicating that responses in population 1 tend to be greater than in population 2. Similarly, if $Y_1 < Y_2$, $X_1 > X_2$ the pair indicates that population 2 tends to have greater responses. Consider all n_1n_2 intertreatment pairs, let I_1 and I_2 be the number of them which are inversions indicating greater responses in population 1 and 2 respectively. Then an intuitively reasonable test is to reject the hypothesis of homogeneity if I_1 and I_2 are too disparate.

For small samples Bross suggests using the permutation distribution of the statistic (I_1-I_2). He then attempts to show that, under a "grand null hypothesis", the statistic

$$\text{COVAST} = \frac{12(I_1-I_2)^2}{(N+4)(I_1+I_2)}$$

will have asymptotically the χ^2-distribution with 1 degree of freedom for large samples. Unfortunately, this proof contains an error (pointed out to me by N. L. Johnson) in that the unknown variance of (I_1-I_2) is replaced by an estimate of its variance conditioned upon the numbers of concordant and discordant pairs. The effect of this error is such that COVAST may simply be divided by 4 in order to make the stated asymptotic null hypothesis distribution correct; however, Bross' "grand null hypothesis" implies not only that the populations are identically the same but also that the response and covariable are independent, and his estimate of variance is applicable only in that situation.

If, however, we define

$$\phi^*\big((Y_{ia},X_{ia}),(Y_{jb},X_{jb})\big) = \begin{cases} 1 & \text{if } Y_{ia} > Y_{jb},\ X_{ia} < X_{jb} \\ -1 & \text{if } Y_{ia} < Y_{jb},\ X_{ia} > X_{jb} \\ 0 & \text{otherwise} \end{cases}$$

and assign scores

$$Z^*_{ia} = \sum_{bj}\sum \phi^*\big((Y_{ia},X_{ia}),(Y_{jb},X_{jb})\big)$$

for $i=1,2,\ldots,n_a$, $a=1,2$, then it is not difficult to verify that

$$I_1 - I_2 = \sum_i Z^*_{i1} .$$

The hypothesis of homogeneity may now be tested by performing an analysis of variance on the scores Z^*, the proof of asymptotic validity being exactly the same as for the scores $Z(\underset{\sim}{\lambda})$ considered previously. This solves Bross' distributional problem.

On the other hand, considering more carefully the test $T(1)$, note that

$$Z_{ia}(1) = \sum_{bj}\sum \phi\big((Y_{ia},X_{ia}),(Y_{jb},X_{jb});1)\big)$$

where

$$\phi\big((Y_{ia},X_{ia}),(Y_{jb},X_{jb});1\big) = \begin{cases} 1 & \text{if } Y_{ia} > Y_{jb},\ X_{ia} < X_{jb} \\ \tfrac{1}{2} & \text{if } Y_{ia} > Y_{jb},\ X_{ia} = X_{jb} \\ & \text{or } Y_{ia} = Y_{jb},\ X_{ia} < X_{jb} \\ 0 & \text{if } Y_{ia} > Y_{jb},\ X_{ia} > X_{jb} \\ & \text{or } Y_{ia} = Y_{jb},\ X_{ia} = X_{jb} \\ & \text{or } Y_{ia} < Y_{jb},\ X_{ia} < X_{jb} \\ -\tfrac{1}{2} & \text{if } Y_{ia} = Y_{jb},\ X_{ia} > X_{jb} \\ & \text{or } Y_{ia} < Y_{jb},\ X_{ia} = X_{jb} \\ -1 & \text{if } Y_{ia} < Y_{jb},\ X_{ia} > X_{jb} \end{cases}$$

and

$$\sum_i Z_{i1}(1) = I_1 + \tfrac{1}{2}P_1 - \tfrac{1}{2}P_2 - I_2$$

where P_1 and P_2 are the number of "partial inversions" indicating greater responses in populations 1 and 2 respectively. Since the test T(1) can also be expressed in terms of ranks it is generally simpler to use than that of Bross (although not in the case of a dichotomous response); but it is not easy to say which may be preferable if grounds other than simplicity of computation are considered. The difference in interpretation is that T(1) gives half-weight to the partial inversions which Bross discards entirely. Of course, for efficiency against alternatives of the type considered in Section 2 the best choice is the rank analysis of covariance test, i.e., $T(r_S)$ if there are no ties.

In Table 2 all three tests are illustrated using the data which Bross presented in [3]. The observations are (Y,X), where X is the birth weight of a baby with hyaline membrane disease, Y is the outcome (R, or 1, if "recovered"; D, or 0, if "died"), and the two populations or treatments are UK (urokinase activated human plasmin) and PL (placebo). Ties on birth weight were broken by using a second variable (X-ray findings) and the results of this are indicated by adding (+) or (-) in the table. The reader may find it instructive to check the computation of a few scores. Bross' test gives $I_1 - I_2 = 21$: the variance ratio, computed as described above, turns out to be 8.93 with (1,23) degrees of freedom, corresponding to a one-tailed probability level of .003; the level from the permutation distribution, according to Bross' calculations, is .004. The test T(1) gives $(I_1+\tfrac{1}{2}P_1-\tfrac{1}{2}P_2-I_2) = 26$, with variance ratio 2.55, which is not significant. Rank analysis of covariance gives $W_1 = 28.98$, with variance ratio 6.25 and corresponding one-tailed probability level about .01.

Table 2

Weight X	Treatment	Outcome Y	Scores: Bross	Scores: T(1)	Scores: Rank Analysis of Covariance
1.08	UK	D	0	3.5	-3.65
1.13	UK	R	7	15.0	8.44
1.14	PL	D	-1	1.5	-4.46
1.20	UK	R	6	13.0	7.63
1.30	UK	R	6	12.0	7.23
1.40	PL	D	-3	- 1.5	-5.67
1.59	UK	D	-3	- 2.5	-6.08
1.69	UK	R	4	9.0	6.02
1.88	PL	D	-4	- 4.5	-6.88
2.00 (-)	PL	D	-4	- 5.5	-7.29
2.00	PL	D	-4	- 6.5	-7.69
2.00 (+)	PL	R	1	5.0	4.40
2.10	PL	R	1	4.0	4.00
2.13	UK	R	1	3.0	3.60
2.14	PL	D	-7	-10.5	-9.31
2.15	UK	R	0	1.0	2.79
2.18	PL	R	0	0.0	2.38
2.30	UK	R	0	-1.0	1.98
2.40	UK	R	0	-2.0	1.58
2.44	PL	R	0	-3.0	1.17
2.50 (-)	UK	R	0	-4.0	.77
2.50 (+)	PL	R	0	-5.0	.37
2.70 (-)	UK	R	0	-6.0	- .04
2.70	UK	R	0	-7.0	- .44
2.70 (+)	UK	R	0	-8.0	- .85

4. DISCUSSION

The problem which has been considered is that of comparing two or more populations with respect to a response variable Y in the presence of a (possibly multivariate) concomitant variable X. In such situations the standard method is the parametric analysis of covariance. The assumptions required for strict validity of this classical procedure are that under H_o the conditional distribution of Y given X be (1) normal with (2) expectation linearly dependent on X, and (3) variance independent of X. The effects of violations of the first two of these have been considered by Atiqullah [1], who concluded that the parametric analysis of covariance is much less robust than the corresponding analysis of variance, and in particular that the requirement of linearity may be critical. In practice one may spend considerable time in a preliminary checking of assumptions; and if it is found that they have been violated, although this may be a result of some interest in itself, it is often essentially of secondary importance relative to the desired overall comparison.

Rank analysis of covariance avoids all three of the above assumptions, and yet, as has been shown, it is still fairly efficient when they are satisfied. On the other hand, it has been assumed throughout this paper that the concomitant variable X has the same distribution in each population, and this requirement is crucial; thus, since the parametric method does have any such assumption, it is applicable in certain situations where the rank method is not. Usually this assumption would be established non-statistically, from considerations of logic, although presumably it could be checked by a preliminary analysis of the concomitant variable. Another disadvantage of the rank method at this time appears to be its unknown behavior for small samples, say with less than five to ten observations per group.

The results of a rank analysis of covariance may be interpreted in terms of the probability that a response chosen at random from one population will exceed a response chosen at random from another. Such a concept seems at least as simple and useful as the usual formulation in terms of means. (The two interpretations are entirely equivalent when the assumptions of the parametric method are satisfied.) It is a commmon approach to problems of this sort, when it seems that a parametric analysis cannot be justified with the data as they stand, to search for a suitable transformation. Guidelines for such a search have been almost entirely heuristic, however; and if a successful transformation is found, its use with the parametric analysis may involve considerable additional complication not only in computation but also in interpretation. Of course, the method proposed here might be considered as an example of this approach in which the transformation is that of ranking. It is interesting to note that exactly the same interpretation may be given to a regression of the ranked responses on the concomitant variates after applying any transformation whatever, including ranking, or none, and not necessarily the same to each.

The general principle described in Section 1 is very widely applicable indeed. The tests of form $T(\underline{\lambda})$, including rank analysis of covariance in particular, and the closely related test of Bross, are all examples of it; and further examples will be considered below.

The following "quick and dirty" method for the case of one concomitant variate is based on a proposal of David and Fix [5] which they in turn state is a variant of a test due to Mood. Draw the scatter diagram of the pooled sample of N observations. Fit a curve to these data by any convenient method - by eye, if you like - and assign to each individual observation the score +1 if it lies above the fitted curve, and -1 if it lies below. Now (and not before!) determine the sample from which each observation was drawn and perform an analysis of variance on the scores. (This test is asymptotically equivalent to a χ^2-test on a contingency table in which there are two rows, corresponding to observations above and below the fitted curve, and m columns, corresponding to the different samples; it may be verified that

$$VR = \frac{(N-m)X^2}{(m-1)(N-X^2)}$$

where VR is the variance ratio statistic of the analysis of variance and X^2 is the statistic of the χ^2-test.) The preceding method could of course be "slowed down" and "cleaned up" by carefully specifying how to fit the curve. David and Fix suggest fitting a straight line which, when considered together with a vertical line drawn through the median of the pooled sample of concomitant variables, divides the diagram into four sectors each containing the same number of observations. However, any method of fitting, including by eye, will allow a valid test of the hypothesis of homogeneity if it is done without any knowledge as to the distribution of the observations among the samples and if the additional requirement is imposed that the proportions of observations on each side of the fitted curve not deviate too greatly from one-half. A proof of this may be derived easily from Theorem 2 of [7]. The method could obviously be extended to the case of more than one concomitant variate but would then lose much of its appeal.

Still another approach, commonly used when there are large samples and only two treatments to be compared, is to pick out pairs of observations which are "matched" on the concomitant variates and then analyse these by one of the standard techniques for paired comparisons. This method has the advantage of allowing the covariates to be completely arbitrary in form, even purely nominal, and also of allowing any arbitrary relationships among covariates and response. On the other hand, it appears to lack power relative to the preceding methods if some structure can be imposed on the data. For further discussion of these points see the recent paper by Billewicz [2]. I am now in the process of developing a broad class of methods which apply the general principle described above in conjunction with the idea of matching.

Finally, Elashoff and Govindarajulu [6] have recently developed another method in which I understand that they assume linear regression and obtain a nonparametric estimate of the slope; but I have not yet seen any of their results.

REFERENCES

[1] Atiqullah, M., "The robustness of the covariance analysis of a one-way classification," Biometrika, 51 (1964), 365-372.

[2] Billewicz, W. Z., "The efficiency of matched samples: an empirical investigation," Biometrics, 21 (1965), 623-644.

[3] Bross, Irwin D. J., "Taking a covariable into account," Journal of the American Statistical Association, 59 (1964), 725-736.

[4] Chatterjee, S. K., and Sen, "Non-parametric tests for the bivariate two-sample location problem," Calcutta Statistical Association Bulletin, 13 (1964), 18-58.

[5] David, F. N., and Fix, Evelyn, "Rank correlation and regression in a nonnormal surface," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, I (1961), 177-197.

[6] Elashoff, Robert, and Govindarajulu, Zakkula, "Nonparametric covariance analysis I: testing methods," presented to the Western Regional Meeting of the Institute of Mathematical Statistics, July, 1965.

[7] Quade, Dana, "On analysis of variance for the k-sample problem," Institute of Statistics, University of North Carolina, Mimeo Series No. 453. To appear in the Annals of Mathematical Statistics.