RANK TESTS OF DISPERSION[1]

By Lincoln E. Moses

Stanford University

Introduction and summary. In a recent paper [1] Ansari and Bradley have shown the equivalence of two rank tests for comparing dispersion, one test due to Barton and David [2], the other to Ansari and Freund, and have provided tables of the exact distribution. They observe that Siegel and Tukey have proposed [11] a similar test which permits use of existing tables. They also exhibit the mean of the limiting normal distribution under the alternative hypothesis. Later Klotz [7] established the equivalence of all these tests. In the present paper it is shown that

(1) Any of these tests is consistent against differences in dispersion if the two distributions have a common median and differ in a scale parameter, and under some less restrictive circumstances. But without such restrictions bizarre asymptotic behavior can arise—including good sensitivity against translation for some non-symmetric densities. One (not very natural) example is offered in which the test constructed for rejection if one of two scale parameters is the larger, actually turns out to be consistent against that parameter's being the smaller of the two.

(2) No rank test (i.e., a test invariant under strictly increasing transformation of the scale) can hope to be a satisfactory test against dispersion alternatives without some sort of strong restrictions (e.g., equal or known medians) being placed on the class of admissible distribution pairs.

(3) Box [3] has proposed testing equality of variances by applying the t test to the logarithms of variances computed within small subgroups. He indicates how such tests should be robust (though not of exact size). Distribution free tests of exact size can be constructed by applying a rank test in place of the t test. Wilcoxon's test applied to variances-within-triads has asymptotic efficiency .5 against normal alternatives. If the two samples each have 9 observations then the exact power is readily calculated and "efficiency" is again about .5.

1. Definitions. Let $x_1, \cdots, x_m$ be independent observations on a variate with absolutely continuous distribution function $F(t)$ and density $f(t)$. Let the sample c.d.f. of these m observations be denoted by $F_m(t)$. Let $y_1, \cdots, y_n$ be independent observations on a variate with absolutely continuous distribution function $G(t)$ and density $g(t)$. Let the sample c.d.f. of these n observations be denoted by $G_n(t)$.

For testing the hypothesis $F = G$ against alternatives that the distributions differ in "dispersion," Ansari, Bradley [1] and Freund have proposed a test

Received April 10, 1961; revised February 18, 1963.

[1] This work was done with the support of a fellowship of the John Simon Guggenheim Memorial Foundation, and while the author was situated in the Department of Social Medicine, University of Oxford.

which is described as follows. The combined sample is arranged in order from least to greatest. Rank 1 is assigned to each of the two outermost observations, Rank 2 to the next ones toward the center from them, etc. If the sum of the sample sizes is an even number, say $2p$, then the ranking ends with the two observations in the middle each receiving rank p. The test statistic, W, is the sum of the ranks belonging to $y_1, \cdots, y_n$. If W is small the hypothesis that $F = G$ is rejected in favor of the alternative that G has greater dispersion than F. (The alternative that F has greater dispersion than G would be favored by a large value of W. A two-sided form of the test rejects for W too large or too small.)

Ansari and Bradley show that the W test is equivalent to one independently proposed by Barton and David [2]. Klotz [7] shows that the W test is equivalent to one proposed by Siegel and Tukey [11]. The discussion of all these tests will be done here in terms of the W test.

2. Asymptotic behavior of the tests. In [1] it is shown that W may be represented as

$$W = nN\left\{\frac{1}{2} - n^{-1}\sum_{j=1}^{n}\left|H_{mn}(y_j) - \frac{1}{2}\right| + \theta(N^{-1})\right\}, \tag{1}$$

where we write θ to denote a generic quantity of magnitude not exceeding one, and where $H_{mn}(t) = [m/(m+n)]F_m(t) + [n/(m+n)]G_n(t)$ is the sample c.d.f. of the combined sample. Now define $\lambda_N = m/(m+n)$ and impose the condition that $\lambda_N = \lambda + \theta(N^{-1})$ for some fixed $0 < \lambda < 1$. Then we can rewrite (1) as

$$W/(nN) = \tfrac{1}{2} - n^{-1}\sum_{j=1}^{n} |\lambda F_m(y_j) + (1-\lambda)G_n(y_j) - \tfrac{1}{2} + \theta(N^{-1})| + \theta(N^{-1}). \tag{1'}$$

From the theory of the Kolmogorov-Smirnov test we have $\sup |F_m(t) - F(t)| = O_p(N^{-\frac{1}{2}})$ and $\sup |G_n(t) - G(t)| = O_p(N^{-\frac{1}{2}})$. Thus we may write (1′) as

$$W/(nN) = \tfrac{1}{2} - n^{-1}\sum_{j=1}^{n} |\lambda F(y_j) + (1-\lambda)G(y_j) - \tfrac{1}{2}| + O_p(N^{-\frac{1}{2}}). \tag{1''}$$

In (1″) we see $W/(nN)$ exhibited as the sum of two random quantities. The second coverages to zero in probability; the first is the average of n identically distributed bounded random variables. The sum of the two components therefore converges in probability to the expectation of the first, which we may call μ^*,

$$\mu^* = \frac{1}{2} - \int |\lambda F(t) + (1-\lambda)G(t) - \tfrac{1}{2}|\, dG(t). \tag{2}$$

It is also shown in [1] that, under our conditions on λ_N, if $F = G$, then,

$$\text{Var}\,(W/nN) = [\lambda(1-\lambda)](48N)^{-1}(1 + O(n^{-1})). \tag{3}$$

From (2) and (3) follows the consistency of the two-sided W test against alternatives giving non-zero values of

$$(4)\qquad \mu^* - \frac{1}{4} = \frac{1}{4} - \int_{-\infty}^{\infty} \left| \lambda F(t) + (1 - \lambda)G(t) - \frac{1}{2} \right| dG(t).$$

It also follows from results in [1] that the W test is not consistent against pairs (F, G) for which $\mu^* = \frac{1}{4}$. That is, for any (F, G) for which $\mu^* = \frac{1}{4}$, and given any $\epsilon > 0$ there is a level of significance $\alpha(\epsilon)$ such that the probability of rejecting H_0 at level $\alpha(\epsilon)$ does not tend to one (and indeed, remains less than ϵ) as $N \to \infty$ with $m/(m + n) \to \lambda$.

The asymptotic behavior of the test can now be explored by studying the quantity $\mu^* - \frac{1}{4}$ in certain examples. The examples presented have been chosen, not because they are likely to arise in practice, but rather because they display possibilities which inhere in the test under discussion.

Example 1. *Densities on disjoint ranges.* Let the distributions of x and y be completely disjoint in the sense that there exists a number A for which $F(A) = 1$ and $G(A) = 0$. Then the value of μ^* is different from $\frac{1}{4}$ unless $\lambda = \frac{1}{2}$, for

$$\mu^* - \frac{1}{4} = \frac{1}{4} - \int_{-\infty}^{\infty} |\lambda F(t) + (1 - \lambda)G(t) - \tfrac{1}{2}| \, dG(t)$$

$$= \frac{1}{4} - \int_{A}^{\infty} |\lambda + (1 = \lambda)G(t) - \tfrac{1}{2}| \, dG(t)$$

$$= \begin{cases} \dfrac{1 - 2\lambda}{4} \dfrac{\lambda}{1 - \lambda} & \text{if } \lambda < \frac{1}{2} \\[2ex] \dfrac{1 - 2\lambda}{4} & \text{if } \lambda > \frac{1}{2}, \end{cases}$$

and it is easily seen that the first expression is positive for $\lambda < \frac{1}{2}$ and the second is negative for $\frac{1}{2} < \lambda < 1$. Thus in very large samples the hypothesis of equal dispersion will be rejected (with high probability) or not, solely as the two samples are of different or equal size, and without any regard to the dispersion in the populations at all, if the distributions are "disjoint" as defined here.

Example 2. *Translation parameter.* Let $G(t) = F(t - a)$. Then from (4) we have

$$\frac{1}{2} - \mu^* = \int_{-\infty}^{\infty} |\lambda F(t) + (1 - \lambda)F(t - a) - \tfrac{1}{2}| \, dF(t - a)$$

$$= \int_{-\infty}^{\infty} | \lambda F(u + a) + (1 - \lambda)F(u) - \tfrac{1}{2}| \, dF(u).$$

If we now assume that the density $f(t)$ is bounded, positive at the median, ν, and possesses almost everywhere a derivative $f'(t)$ which is bounded, we may expand the integral in a Taylor's series in a, and obtain

$$(5)\qquad \mu^* - \frac{1}{4} = \lambda a \left[\int_{-\infty}^{\nu} f^2(u) \, du - \int_{\nu}^{\infty} f^2(u) \, du \right] + O(a^2).$$

From (5) we may infer that

(a) If $f(t)$ is symmetric then the two integrals are equal, $\mu^* - \frac{1}{4}$ differs from 0 by an amount of smaller order than a, and the efficiency of the test as defined by Pitman [4], [6] is then zero against translation.

(b) If the two integrals are unequal, then the two-sided test is consistent against small translation of one distribution with respect to the other, and the efficiency of the test is positive for these alternatives. This efficiency can even be large as the following example shows. Let

$$\begin{aligned} f(t) &= \tfrac{1}{2} && 0 \leqq t \leqq 1 \\ &= 1/2(A - 1) && 1 < t \leqq A \\ &= 0 && \text{otherwise.} \end{aligned}$$

Then corresponding to a translation $aN^{-\frac{1}{2}}$ we have, in accordance with (5)

$$\mu^* - \frac{1}{4} = \lambda \frac{a}{N^{\frac{1}{2}}} \left[\frac{1}{4} - \frac{1}{4(A - 1)}\right] = \frac{\lambda a}{4N^{\frac{1}{2}}} \frac{A - 2}{A - 1},$$

and this divided by σ_0 the (null) standard error of $W/(nN)$ is

$$\frac{\mu^* - \frac{1}{4}}{\sigma_0} = \frac{\frac{\lambda a}{4N^{\frac{1}{2}}} \frac{A - 2}{A - 1}}{\left(\frac{\lambda}{1 - \lambda} \frac{1}{48N}\right)^{\frac{1}{2}}} = [\lambda(1 - \lambda)]^{\frac{1}{2}} \frac{A - 2}{A - 1} 3^{\frac{1}{2}} a.$$

If instead we used the Wilcoxon test to detect this translation, then the comparable non-centrality parameter (corresponding to the Mann and Whitney U statistic divided by nN) is

$$\begin{aligned} \frac{\lambda \frac{a}{N^{\frac{1}{2}}} \int_{-\infty}^{\infty} f^2(t)\, dt}{\{[\lambda/(1 - \lambda)](1/12N)\}^{\frac{1}{2}}} &= [\lambda(1 - \lambda)]^{\frac{1}{2}} a 12^{\frac{1}{2}} \left[\frac{1}{4} + \frac{1}{4(A - 1)}\right] \\ &= [\lambda(1 - \lambda)]^{\frac{1}{2}} \frac{A - 2}{A - 1} 3^{\frac{1}{2}} a \left(\frac{A}{2(A - 2)}\right). \end{aligned}$$

The square of the term in parentheses is the Pitman relative efficiency of the Wilcoxon test to the dispersion test, and we see that, for A very large, the dispersion test is approximately four times as sensitive against small translations as is the Wilcoxon test.

EXAMPLE 3. *Scale parameter*. Let $G(t) = F(t/b)$. Then from (4) we have

$$\begin{aligned} \tfrac{1}{2} - \mu^*(b) &= \int_{-\infty}^{\infty} |\lambda F(t) + (1 - \lambda)F(t/b) - \tfrac{1}{2}|\, dF(t/b) \\ &= \int_{-\infty}^{\infty} |\lambda F(bu) + (1 - \lambda)F(u) - \tfrac{1}{2}| F(u)\, du. \end{aligned}$$

Now writing $b = 1 + \beta(\beta > 0)$ and imposing upon f the requirements that $f'(u)$ exists almost everywhere bounded, that $uf'(u)$ and $u^2f'(u)$ are also bounded, and that $f(\nu) > 0$, we may expand the integral in a Taylor's series in β and arrive at

$$\mu^* - \tfrac{1}{4} = \lambda\beta\left[\int_{-\infty}^{\nu} uf^2(u)\,du - \int_{\nu}^{\infty} uf^2(u)\,du\right] + O(\beta^2). \tag{6}$$

The parameter in (6), if negative, implies consistency against $\beta > 0$ of the one-sided test which rejects for small W. Small W results from the observations y being far from the median of the joint sample, i.e., "more spread out," and $\beta > 0$ implies y has a greater dispersion than x. Then the consistency of the one-sided test against $\beta > 0$ is satisfactory, or unsatisfactory as $\mu^* - \frac{1}{4}$ is negative or positive. If f is symmetric around zero then ν is zero and both of the integrals make a negative contribution to the coefficient of β; in this case the consistency of the one-sided test is "satisfactory." But unsatisfactory consistency can occur as in the following case:

$$\begin{aligned} f(x) &= (1/\sigma)(2/\pi)^{\frac{1}{2}} \exp\{-\tfrac{1}{2}(x/\sigma - A)^2\} &&\text{if } x \geqq A\sigma > 0 \\ &= 0 &&\text{if } x < A\sigma. \end{aligned} \tag{7}$$

Then σ is a scale parameter for this distribution (although A is not a location parameter). Tables of the normal distribution show that $\nu = A\sigma + .675\sigma$, whence

$$\begin{aligned} \mu^* - \tfrac{1}{4} &= \lambda\beta\left[\frac{1}{\sigma^2}\frac{2}{\pi}\int_{A\sigma}^{A\sigma+.675\sigma} u \exp\left[-(u - A\sigma)^2/\sigma^2\right] du \right. \\ &\qquad \left. - \frac{2}{\sigma^2\pi}\int_{A\sigma+.675\sigma}^{\infty} u \exp\left[-(u - A\sigma)^2/\sigma^2\right] du\right] \\ &= \lambda\beta\frac{2}{\pi}\left[\int_0^{.675} (z + A)e^{-z^2}\,dz - \int_{.675}^{\infty} (z + A)e^{-z^2}\,dz\right] \\ &= \lambda\beta\frac{2}{\pi}\left[\int_0^{.675} ze^{-z^2}\,dz - \int_{.675}^{\infty} ze^{-z^2}\,dz + A\int_0^{.675} e^{-z^2}\,dz - A\int_{.675}^{\infty} e^{-z^2}\,dz\right], \end{aligned}$$

and this expression is negative for small A and positive for large A. Thus for distributions of the family (7) whether the test is consistent satisfactorily or unsatisfactorily with regard to the scale parameter σ depends upon the value of the parameter A.

This result is not unnatural, in the light of what we have already found about translation of non-symmetric densities, for if A is large then a small increase in σ, say $\Delta\sigma$, not only increases the dispersion, but also shifts the distribution to the right by $A\Delta\sigma$.

EXAMPLE 4. F and G are two distributions with the same median ν.

In this case we write

$$\begin{aligned} \tfrac{1}{2} - \mu^* &= \int_{-\infty}^{\nu} \left[\tfrac{1}{2} - \lambda F(t) - (1 - \lambda)G(t)\right] dG(t) \\ &\qquad + \int_{\nu}^{\infty} \left[\lambda F(t) + (1 - \lambda)G(t) - \tfrac{1}{2}\right] dG(t) \end{aligned}$$

$$= (1 - \lambda)/4 + \lambda \left\{\frac{1}{4} - \int_{-\infty}^{\nu} F(t)\, dG(t) - \int_{\nu}^{\infty} (1 - F(t))\, dG(t) + \tfrac{1}{2}\cdot\tfrac{1}{2}\right\}$$

$$= \tfrac{1}{4} + \lambda(\tfrac{1}{4} - P(x < y < \nu) - P(\nu < y < x)).$$

Now observing that if $F = G$ the two indicated probabilities are each $\frac{1}{8}$, we may write $\mu^* - \frac{1}{4} = \lambda(\Delta_L + \Delta_R)$, where $P(x < y < \nu) = \frac{1}{8} + \Delta_L$ and $P(\nu < y < x) = \frac{1}{8} + \Delta_R$. We see that in case the medians of x and y are equal, the consistency of the test depends upon $P(y$ lies between x and the median) which is a very natural dispersion parameter.

In this example, and some of those preceding it, the non-centrality parameter contains λ as a coefficient. This results from having chosen to study the statistic W with the normalization $W/(nN)$ which is not symmetric in m and n. Reference to the expression for the variance in (3) above shows that the non-centrality parameter for $W/(nN)$ expressed as a multiple of its standard deviation (under the null hypothesis) involves λ through the expression $[(\lambda)(1 - \lambda)]^{\frac{1}{2}}$ which has the symmetry which is usual in two-sample problems.

3. General remarks about rank tests for dispersion. By a two-sample rank test is usually meant a test based upon the ranks of the observations when arranged from least to greatest in one ranking. Such tests have properties advantageous in various circumstances. One property is that the value of such a statistic is left unchanged by any strictly increasing transformation of the scale of measurement, and thus the distribution of the test statistic and the properties of the test are unaltered by such transformations.

The burden of this section is to show that this feature of rank tests makes it hopeless to try to devise a satisfactory rank test of dispersion without greatly restricting the class of distributions to which it is to be applied, at least if one adopts either of two rather general concepts of what is meant by "dispersion."

Dispersion in the first sense relates to closeness—on the average, and in some specified sense—of independent observations within pairs (or larger clusters). Some parameters of this first type, which contrast the dispersions of two populations are

$$\Pr(|y_1 - y_2| > |x_1 - x_2|), \qquad E|y_1 - y_2| - E|x_1 - x_2|,$$

$$E[\text{range}(y_1, y_2, \cdots, y_k)] - E[\text{range}(x_1, x_2, \cdots, x_k)],$$

$$E(y_1 - y_2)^2 - E(x_1 - x_2)^2 = 2(\sigma_y^2 - \sigma_x^2).$$

Dispersion in the second sense relates to closeness—on the average, and in some specified sense—of a typical observation to some central number associated with the distribution. Some parameters of this second type which contrast the dispersions of two populations are:

$$\Pr(|y - \nu_y| > |x - \nu_x|), \qquad E|y - \nu_y| - E|x - \nu_x|,$$

$$E(y - \mu_y)^2 - E(x - \mu_x)^2 = \sigma_y^2 - \sigma_x^2.$$

Both of these notions of dispersion depend upon *distance*. Though strict monotonic transformations leave *order* unchanged, they do distort distance. Accordingly we might suspect that, except possibly under restrictive circumstances, a rank test could hardly serve well against dispersion alternatives.

It is easy to see that if two absolutely continuous distribution functions F, G are disjoint in the sense of Example 1, then no rank test can satisfactorily contrast their dispersion. The reason is that the distribution of ranks is the same for the pair $(F(x), G(x))$ as for the pair $(F(Ty), G(Tx))$ where T is any strictly increasing transformation of the real axis. In particular, T may be chosen to "stretch out" (in any of the above senses) the range of one of the random variables and to leave the other unaffected. Then the behavior of any rank test at all would be entirely unrelated to the relative dispersions of the populations. Somewhat similar conclusions can be obtained by essentially the same argument if one of the distributions has any mass lying entirely to the right, or entirely to the left of the other.

We offer an example of the same phenomenon for a pair of distributions absolutely continuous with respect to one another. They have exactly equal dispersions in any of the above senses—or indeed in any usual sense at all. We then show two transformations which leave the distribution of ranks unaltered; the first one provides the one distribution with greater dispersion in all the above senses while the second provides the other with the greater. For $0 \leqq t \leqq 1$ let $f_x(t) = 2(1 - t)$ and $g_y(t) = 2t$ and let both densities be zero for t outside the unit interval. These densities are oppositely-facing right triangles on the same base.

The first transformation replaces t by e^{-u}. This amounts to replacing x by $x' = -\log x$ and y by $y' = -\log y$.

Under this transformation the two densities become $f_{x'}^{(1)}(u) = 2(e^{-u} - e^{-2u})$ and $g_{y'}^{(1)}(u) = 2e^{-2u}$. These densities have variances $\sigma_{x'}^2 = \frac{5}{4}$, $\sigma_{y'}^2 = \frac{1}{4}$ and

$$P(|x_1' - x_2'| > |y_1' - y_2'|) = \tfrac{26}{36}.$$

If instead we transform t into $1 - e^{-w}$ the new densities, of $x'' = -\log(1 - x)$ and $y'' = -\log(1 - y)$ are $f_{x''}^{(2)}(w) = 2e^{-2w}$ and $g_{y''}^{(2)}(w) = w(e^{-w} - e^{-2w})$ and the relations above are now reversed. But under each of three pairs of densities $(f, g)(f^{(1)}, g^{(1)})$ and $(f^{(2)}, g^{(2)})$ the distribution of any rank statistic whatever is exactly the same.

The author is indebted to Wassily Hoeffding for the following observations with regard to this example. The random variable x' is distributed as $y' + z$ where z has density $\exp(-u)$, $u > 0$, and is independent of y'. Hence, x' will have greater dispersion than y' for any definition of dispersion such that (1) the dispersion of any random variable not certainly a constant is positive, (2) the dispersion of the sum of two independent random variables with positive dispersions is greater than the dispersion of either summand.

4. Some rank-like tests for dispersion. If one abandons the effort to construct a two-sample rank test of dispersion which will behave well in spite of translation, etc., he may still seek for a test having the convenience and robustness often enjoyed by rank tests. Two lines of approach are natural—they correspond to the two characterizations of dispersion we have already used.

First, one might apply a rank test to pseudo-observations constructed by replacing each observation by its value minus the median of its sample. This has been proposed as being suitable, at least in large samples, by Mood [9] and Ansari and Bradley [1]. Justifying such a modified test for large samples would require showing that so long as $F = G$ (that is, provided the null hypothesis held) there was a limiting distribution for the test statistic, independent of F; the test would then be called "asymptotically distribution-free." Sukhatme [12] has shown that in fact Mood's test so modified is not asymptotically distribution-free. Ansari and Bradley [1] show that their modified test is not distribution-free in small samples, and leave as an open question whether it is asymptotically distribution-free.

Second, imitating a proposal of Box [3], we might break the samples up into small, equal, exclusive, and exhaustive random subsets, compute for each such subset the value of a dispersion statistic (e.g., range or sample variance) and apply such a rank test as Wilcoxon's to these statistics. This kind of approach seems to be wasteful of data since only the variation within subsets is utilized, that between subsets being ignored. However, Lehmann's attempt [8] to use all the information on dispersion resulted in a test statistic which is not only very ponderous to compute but is not distribution-free. In view of the shortcomings of rank tests and of median-adjusted rank tests for this problem, it seems worthwhile to explore the rank-like dispersion tests of the Box sort. (The term "rank-like" is used because the tests are *not* invariant under distortion of the scale of measurement, but do employ rank scores.)

The most readily interpreted test of this form would apply the Wilcoxon test to the intra-pair differences

$$u_1 = |x_1 - x_2|, \qquad u_2 = |x_3 - x_4| \text{ etc.}$$

and

$$v_1 = |y_1 - y_2|, \qquad v_2 = |y_3 - y_4| \text{ etc.}$$

In this case the test statistic is essentially an estimate of $P(|x_i - x_j| > |y_i - y_j|)$ and this is a very natural dispersion parameter.

The rank-like test most easily investigated for small sample behavior under normal alternatives is one based on triads of observations. Turning now to that investigation, let

$$\xi = \tfrac{1}{2}\sum_1^3 (x - \bar{x})^2 \qquad \eta = \tfrac{1}{2}\sum_1^3 (y - \bar{y})^2.$$

Let the original samples be $x_1, \cdots, x_{3m}$ and $y_1, \cdots, y_{3n}$. Let the observed values of ξ and η be $u_1, u_2, \cdots, u_m$ and $v_1, v_2, \cdots, v_n$. The test rejects the hypothesis that x and y have equal dispersion if the sum of the ranks of the v's in the common ranking of the u's and v's is too far from its expectation.

The test is consistent against alternatives of the form $P(\xi > \eta) \neq \frac{1}{2}$. The probability appearing on the left-hand side is a fairly readily interpreted dispersion parameter.

If the observations come from normal distributions with parameters (μ_x, σ_x^2) and (μ_y, σ_y^2) respectively, then the above inequality is equivalent to $\sigma_x^2 \neq \sigma_y^2$. Further, because the distribution of the sample variance of three independent observations from the normal distribution is given by the exponential distribution, it is relatively easy to investigate small sample, as well as large sample behavior of the test. The case $3m = 3n = 9$ and $\alpha = .05$ is worked out below.

TEST: at level .05 reject H_0 (accept the alternative hypothesis that $\sigma_y^2 > \sigma_x^2$) if $\min(v_1, v_2, v_3) > \max(u_1, u_2, u_3)$

$$1 - \beta = \Pr\{\text{rej } H_0 | \sigma_y^2 = k\sigma_x^2|\} = 3\int_0^\infty (1 - e^{-x/\sigma^2})^3 e^{-3x/k\sigma^2} \frac{dx}{k\sigma^2}$$

$$= 3\int_0^\infty (1 - e^{-ku})^3 e^{-3u}\, du = 3\left(\frac{1}{3} - \frac{3}{3+k} + \frac{3}{3+2k} - \frac{1}{3(1+k)}\right).$$

Similar calculations can be carried through for the case $3m = 3n = 12$. An .05 test in this case must involve some randomization since there are 70 (equally likely under H_0) outcomes. The Wilcoxon test calls for rejection with probability $\frac{3}{4}$ if either of the orders yielding a rank sum for the u's of 12 is obtained and for certain rejection if the sum is 10 or 11. A more powerful test studied by Savage [10] is got by rejecting certainly for the rank sums 10 and 11 and for one of the two orders with rank sum of 11, rejecting with probability $\frac{1}{2}$ if the other appears.

TABLE 1

Values of k corresponding to stated power of several tests for comparing two variances

Test		Power				
		.10	.25	.50	.75	.90
	N_x, N_y					
F test	4, 4	1.721	3.938	9.277	21.85	50.01
	5, 5	1.555	3.095	6.388	14.19	26.24
	6, 6	1.463	2.665	5.050	9.569	17.44
	N_x, N_y					
Rank-like test	9, 9 (W or S)	1.50	3.00	6.90	18.04	51.10
	12, 12 (W)	1.40	2.48	4.98	11.35	28.95
	12, 12 (S)	1.40	2.46	4.87	10.73	26.06

For each of these rank tests it is possible to find the value of k, defined above as σ_y^2/σ_x^2, for which the power assumes any specified value. Such values are displayed in Table 1, W denoting Wilcoxon and S denoting Savage. Values of k corresponding to the same power values for a .05-level F test based upon samples of equal size $N_x = N_y$ are also shown. (The latter values can be read directly from table 8.3 of [5].) It is seen that the rank-like test has power comparable to that of the F test based on approximately half as many observations, and that the test is less satisfactory for "distant alternatives," that is, for large values of k.

The asymptotic efficiency for normal alternatives is readily evaluated and found to be .5.

This two-sample test can be extended to the k-sample dispersion problem by applying the Kruskal-Wallis H test to the ranks of the variances of triads in the k samples. In small samples the exact size of a rank-like test can be a powerful reason for using it. In large samples it may well be more convenient (but not necessarily more efficient) to apply a Box-test than a rank-like adaptation of it.

For both classes of tests the question of the best subgroup size remains open. And in the case of the tests proposed here there also remains the question of which index of dispersion—range, variance, average deviation, etc.,—should be used within subsamples.

If considerations of asymptotic efficiency for normal distributions are important to the user then he could construct a test related to the present one just as the Fisher-Yates-Hoeffding-Terry [13] test is related to the Wilcoxon test. This modification would entail the loss of ease of interpretation of the consistency parameter. It can be conjectured that such a test would have asymptotic efficiency $\frac{2}{3}$. The most powerful test of any kind based on $u_1, \cdots, u_m, v_1, \cdots, v_n$ would have efficiency $\frac{2}{3}$ in comparison with the likelihood ratio test. The locally most powerful rank test should asymptotically attain this ceiling for alternatives very near to H_0.

REFERENCES

[1] Ansari, A. R. and Bradley, R. A. (1960). Rank sum tests for dispersion. *Ann. Math. Statist.* **31** 1174–1189.

[2] Barton, D. E. and David, F. N. (1958). A test for birth order effects. *Ann. Human Genetics* **22** 250–257.

[3] Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika* **40** 318–335.

[4] Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* **29** 972–994.

[5] Eisenhart, Churchill (1947). Planning and interpreting experiments for comparing two standard deviations. *Selected Techniques of Statistical Analysis* (C. Eisenhart, M. V. Hastay and W. A. Wallis, eds.), pp. 267–318. McGraw-Hill, New York.

[6] Hodges, J. L., Jr. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist.* **27** 324–335.

[7] Klotz, Jerome (1962). Nonparametric tests for scale. *Ann. Math. Statist.* **33** 498–512.

[8] Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. *Ann. Math. Statist.* **22** 165–179.

[9] Mood, A. M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *Ann. Math. Statist.* **25** 514–522.
[10] Savage, I. R. (1956). Contributions to the theory of rank-order statistics: two-sample case. *Ann. Math. Statist.* **27** 590–615.
[11] Siegel, S. and Tukey, J. W. (1960). A non-parametric sum of ranks procedure for relative spread in unpaired samples. *J. Amer. Statist. Assoc.* **55** 429–445.
[12] Sukhatme, B. V. (1958). Testing in hypothesis that two populations differ only in location. *Ann. Math. Statist.* **29** 60–78.
[13] Terry, M. E. (1952). Some rank order tests which are most powerful against specific parametric alternatives. *Ann. Math. Statist.* **23** 346–366.