RANK-SUM TESTS FOR DISPERSIONS[1]

By A. R. Ansari and R. A. Bradley[2]

Virginia Agricultural Experiment Station, Virginia Polytechnic Institute

1. Summary. This paper deals with non-parametric two-sample tests on dispersions. Two samples, X- and Y-samples of m and n independent observations from populations with continuous cumulative distribution functions $F(u)$ and $G(u)$ respectively, are considered. It is required for the basic test that the difference in locations (medians) of the two populations be known and, when this is so, the two samples may be adjusted to have equal locations. Taking these location parameters to be zero without loss of generality, we test the hypothesis that $G(u) \equiv F(u)$ against alternatives of the form $G(u) \equiv F(\theta u)$, $\theta \neq 1$. The two samples are ordered in a single joint array and ranks are assigned from each end of the joint array towards the middle. The statistic used is W, the sum of ranks for the X-sample.

The distribution of W is studied and tables of significant values of W are provided for $m + n \leqq 20$ and both upper- and lower-tail significance levels .005, .01, .025 and .05. The first four moments of W are developed and a normal approximation to the null distribution of W is devised.

Large-sample properties of the W-test are considered. A proof of limiting normality is based on a theorem of Chernoff and Savage. Consistency of the W-test is indicated and its relative efficiency in comparison with the variance-ratio F-test is obtained as $6/\pi^2$ when $F(u)$ is the normal distribution function.

Other non-parametric tests of dispersions are reviewed. The W-test is less efficient asymptotically than some of these other tests but is easier to apply, particularly with the tables provided.

A modified test is suggested for the case where the difference in population locations is not known. This involves replacing the two original samples by two corresponding samples of deviations from sample medians. The procedure of the W-test is applied to the two samples of deviations. The properties of the modified test have not been investigated except for a sampling study of rather limited scope. That study indicates that the moments of W for the modified test are not greatly different from those under the basic procedure.

2. Introduction. Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ represent two independent samples of sizes m and n of independent observations from two populations with continuous cumulative distribution functions (c.d.f.'s), $F(u)$ and $G(u)$

Received August 17, 1959; revised December 21, 1959.

[1] Research sponsored by the Statistics Branch, Office of Naval Research, U. S. Navy. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Present address: Department of Statistics, the Florida State University, Tallahassee, Florida.

respectively. We shall assume for the most part that the difference in location parameters (medians) $\mu_X - \mu_Y$ of the two populations is known and it may be taken to be zero for we may adjust the initial samples by subtracting $\mu_X - \mu_Y$ from each X-observation. The parameters μ_X and μ_Y need not be known, but no generality is lost in assuming $\mu_X = \mu_Y = 0$ in what follows. Then $F(u)$ and $G(u)$ are assumed to be of the same form and to differ at most in the value of a scale parameter θ, so that $G(u) \equiv F(\theta u)$. We develop a rank-order test of the null hypothesis,

$$(1) \qquad H_0: \ \theta = 1, \qquad \text{i.e.,} \quad G(u) \equiv F(u),$$

against either one-sided or two sided alternatives to H_0.

The two samples (adjusted by $\mu_X - \mu_Y$ if necessary) are ranked or ordered in a combined array represented by

$$(2) \qquad Z_1, \cdots, Z_{m+n}.$$

But ranks are assigned from both ends of (2), beginning with unity and working towards the center. If $m + n$ is even, we have the array of ranks

$$(3) \qquad 1, 2, 3, \cdots, (m+n)/2, \qquad (m+n)/2, \cdots, 3, 2, 1;$$

and, if $m + n$ is odd, we have

$$(4) \quad 1, 2, 3, \cdots, (m+n-1)/2, \qquad (m+n+1)/2, \qquad (m+n-1)/2, \cdots, 3, 2, 1.$$

The test statistic to be considered is

$$(5) \qquad W = \sum_X R(Z),$$

the sum of the ranks in (3) or (4) associated with the X-sample. An alternative form, equivalent to (5) and more useful in mathematical considerations, is

$$(6) \qquad W = \sum_{i=1}^{p} i\delta_i^x + \sum_{i=p+1}^{m+n} (m+n+1-i)\delta_i^x$$

where $\delta_i^x = 1$ if Z_i is an X-observation and $\delta_i^x = 0$ otherwise, and where

$$p = [(m+n+1)/2],$$

the largest integer in $(m + n + 1)/2$. Small values of W indicate larger dispersion for the X-sample and large values of W indicate larger dispersion for the Y-sample. Small values of W suggest that $\theta < 1$ and large values of W suggest that $\theta > 1$. The test based on W and its properties are discussed in the following sections.

Freund and Ansari [5] proposed the W-test and seem to have been the first to make such a proposal. David and Barton [4] presented a generalized procedure that includes the W-test as a special case but they did not investigate the properties of their method. Sukhatme ([13], [14]) has proposed two other rank order dispersion tests. The first of these is similar to the W-test but requires knowledge of μ_X and μ_Y, not simply of $\mu_X - \mu_Y$.

It is of interest for comparative purposes to note that Sukhatme's first test uses the statistic

$$T = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \chi(X_i, Y_j)$$

where

$$\chi(X, Y) = \begin{cases} 1, & \text{if } 0 < X < Y \text{ or } Y < X < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let $n_-(X)$, $n_-(Y)$, $n_+(X)$ and $n_+(Y)$ indicate numbers of negative and positive X- and Y- observations. Then

$$mnT = \sum_X R'(Z) - \tfrac{1}{2}[n_-(X)\{n_-(X) + 1\} + n_+(X)\{n_+(X) + 1)\}$$

where $\sum_X R'(Z)$ is the sum of ranks associated with the X-sample when the ranking is modified as in the array

$$1, 2, 3, \cdots, n_-, \qquad n_+, \cdots, 3, 2, 1,$$

$n_- = n_-(X) + n_-(Y)$, $n_+ = n_+(X) + n_+(Y)$. The statistic T depends on rankings of positive and negative observations separately and on the numbers of positive and negative values of X. Our statistic W, although a similar statistic, avoids attachment of any meaning to the zero point of the scale of X.

A statistic W', associated with W, could also be used for a test procedure; W' could be obtained directly by ranking from the center of the array (2) towards the two ends beginning with unities if $m + n$ is even and with a zero if $m + n$ is odd. Now W' is equivalent to W, since

$$W' = \tfrac{1}{2}m(m + n) + m - W \tag{7}$$

if $m + n$ is even and

$$W' = \tfrac{1}{2}(m + n + 1) - W \tag{8}$$

if $m + n$ is odd. W' may be preferred to W if a statistic is desired such that large values of the statistic occur with larger dispersion for the X-sample, but tests based on W and W' are equivalent in their properties.

Other nonparametric tests on dispersions are available.[3] We note papers by Rosenbaum [11], Kamat [6], Barton [2], Lehmann [7], Terry [15] and Mood [8]. In addition, tests have been proposed that are consistent against more varied departures from equality of two c.d.f.'s, but these will not be discussed. We do compare, when possible, the W-test with other tests on dispersions.

[3] Sidney Siegel and John W. Tukey [12] in a recent paper have a test similar to the W-test. They rank the array (2) as: 1, 4, 5, 8, 9, $\cdots$, 7, 6, 3, 2 and this permits use of existing tables.

In a final section we discuss what may be done if $\mu_X - \mu_Y$ is not known. The W-test cannot then be used directly because it is in some cases sensitive to differences in locations of the two populations.

3. The Null Distribution of W. When H_0 is true, $\theta = 1$ and $F(u) \equiv G(u)$. Under these conditions, each of $\binom{m+n}{m}$ distinct assignments of m X's to the $m + n$ positions in (2) is equally likely, and each such assignment yields a value of W. Probabilities of occurrence of each distinct value of W are obtained as the product of $1/\binom{m+n}{m}$ and the number of distinct assignments that yield that value of W. Tables for the cumulative distributions of W for various values of m and n may then be prepared.

Ansari [1] has prepared tables showing the complete distributions of W for $m = 2(1)10$, $m + n = 4(1)20$. In Table 1 we give only critical values of W for significance levels .995, .99, .975, .95, .05, .025, .01, and .005 taken from the complete tables. If $W_0(\alpha)$ is a critical value of W with significance level α,

$$P[W \geqq W_0(\alpha) \mid H_0] \leqq \alpha$$

for $\alpha \leqq .05$ and $P[W \leqq W_0(\alpha) \mid H_0] \leqq 1 - \alpha$ for $\alpha \geqq .95$.

Both a recursion formula and a frequency generating function have been derived to facilitate consideration of distributions of W. Let $f(W \mid m, n)$ denote the frequency of occurrence of W given the sample sizes m, n. (The corresponding probability is $P(W \mid m, n) = f(W \mid m, n)/\binom{m+n}{m}$.) The recursion formula for $m \geqq 2$ is

$$(9)\quad f(W \mid m, n+1) = f(W \mid m, n) + f(W - N - 1 \mid m - 1, n + 1)$$

where $m + n = 2N$ or $m + n = 2N + 1$ depending on whether $m + n$ is even or odd. Alternatively, (9) may be written

$$(10)\quad (m + n + 1)P(W \mid m, n + 1) = (n + 1)P(W \mid m, n) + mP(W - N - 1 \mid m - 1, n + 1).$$

The frequency generating function is

$$(11)\quad g(u, v) = \prod_{i=1}^{N} (1 + u^i v)^2 \qquad \text{if } m + n = 2N$$

$$= (1 + u^{N+1} v) \prod_{i=1}^{N} (1 + u^i v)^2 \qquad \text{if } m + n = 2N + 1.$$

The frequency function $f(W \mid m, n)$ is the coefficient of $u^W v^m$ in the expansions of (11).

The recursion formula is very nearly obvious in the form (9). Consider the

TABLE 1

Lower and Upper Significance Levels of W

Sample Sizes		Significance Levels							
m	n	.995	.99	.975	.95	.05	.025	.01	.005
2	5	—	—	—	2	—	—	—	—
2	6	—	—	—	2	8	—	—	—
2	7	—	—	—	2	9	—	—	—
2	8	—	—	2	2	10	10	—	—
2	9	—	—	2	2	11	11	—	—
2	10	—	—	2	2	12	12	—	—
2	11	—	—	2	2	13	13	—	—
2	12	—	—	2	2	14	14	—	—
2	13	—	2	2	3	14	15	—	—
2	14	—	2	2	3	15	16	16	—
2	15	—	2	2	3	16	17	—	—
2	16	—	2	2	3	17	17	18	—
2	17	—	2	2	3	18	19	—	—
2	18	—	2	2	3	19	19	20	—
3	5	—	—	—	4	11	—	—	—
3	6	—	—	4	4	13	13	—	—
3	7	—	—	4	5	13	14	—	—
3	8	—	—	4	5	15	16	16	—
3	9	—	4	4	5	16	17	17	—
3	10	—	4	5	5	17	18	18	19
3	11	—	4	5	6	18	19	20	—
3	12	4	4	5	6	20	21	22	22
3	13	4	4	5	6	21	22	23	23
3	14	4	5	6	7	22	23	24	25
3	15	4	5	6	7	23	24	25	26
3	16	4	5	6	7	24	25	27	28
3	17	4	5	6	8	25	26	28	29
4	4	—	—	6	6	14	14	—	—
4	5	—	6	6	7	14	16	—	—
4	6	6	6	7	7	17	17	18	18
4	7	6	6	7	8	19	19	20	—
4	8	6	6	7	8	20	21	22	22
4	9	6	7	8	9	21	22	23	24
4	10	7	7	8	9	23	24	25	25
4	11	7	7	9	10	24	26	27	27
4	12	7	8	9	10	26	27	28	29
4	13	7	8	9	11	27	29	30	31
4	14	8	9	10	11	29	30	31	32
4	15	8	9	10	12	30	32	33	34
4	16	8	9	11	12	32	33	35	36

TABLE 1—*Continued*

Sample Sizes		Significance Levels							
m	*n*	.995	.99	.975	.95	.05	.025	.01	.005
5	5	—	9	10	10	20	20	21	—
5	6	9	9	10	11	22	23	24	24
5	7	9	10	11	11	24	24	25	26
5	8	10	10	11	12	26	26	28	29
5	9	10	11	12	13	27	28	29	30
5	10	10	11	12	14	29	30	32	32
5	11	11	12	13	14	31	32	33	34
5	12	11	12	14	15	33	34	36	37
5	13	12	13	14	16	34	36	37	38
5	14	12	13	15	16	36	38	40	41
5	15	12	14	15	17	38	40	41	43
6	6	12	13	14	15	27	28	29	30
6	7	13	14	15	16	29	30	32	32
6	8	14	14	16	17	31	32	34	34
6	9	14	15	16	18	34	35	36	37
6	10	15	16	17	18	36	37	38	39
6	11	15	16	18	19	38	40	41	42
6	12	16	17	19	20	40	41	43	44
6	13	16	18	19	21	42	44	46	47
6	14	17	18	20	22	44	46	48	49
7	7	17	18	19	21	35	37	38	39
7	8	18	19	20	22	38	39	41	42
7	9	19	20	21	23	40	42	43	44
7	10	20	21	22	24	43	44	46	47
7	11	20	22	23	25	45	47	48	50
8	8	23	24	26	27	45	46	48	49
8	9	24	25	27	29	48	49	51	52
8	10	25	26	28	30	50	52	54	55
8	11	26	27	29	31	53	55	57	58
8	12	27	28	30	32	56	58	60	61
9	9	30	31	33	35	55	57	59	60
9	10	31	32	34	36	58	60	62	64
9	11	32	34	36	38	61	63	65	67
10	10	38	39	41	43	67	69	71	72

case with $m + n = 2N$. Note that $f(W \mid m, n + 1)$ is made up of two parts: the frequency of W when W does not contain the rank $N + 1$ and the frequency of W when W does contain $N + 1$. The first part is $f(W \mid m, n)$ and the second is $f(W - N - 1 \mid m - 1, n + 1)$ and (9) follows. The demonstration is similar when $m + n = 2N + 1$.

The frequency generating function is easily proved by mathematical induction. The fundamental part of the induction is proved using (9). We do not give details of the proof since it is easy but somewhat cumbersome, and since Barton and David also considered a generating function from which (11) results as a special case.

4. Moments and Approximate Distributions under H_0. The approximate distribution of W under H_0 is of interest for applications of the W-test beyond the scope of the prepared tables. We examine the moments of W on the hypothesis that all assignments of the X- and Y-observations to the array (2) are equally likely.

Suppose that $m + n = 2N$. Then

$$E(W) = mE_1(r) = m(m + n + 2)/4, \tag{12}$$

for we consider $E_1(r)$ as the expectation of an integer chosen at random from the first N integers and $(N + 1)/2 = (m + n + 2)/4$. We write

$$E(W^2) = mE_1(r^2) + m(m - 1)E(rs) \tag{13}$$

and

$$E(rs) = \left[2\binom{N}{2}E_1(rs) + \binom{N}{1}^2 E_2(rs)\right] \Big/ \binom{2N}{2}. \tag{14}$$

Now $E_1(r^2) = (N + 1)(2N + 1)/6$, the expectation of the square of an integer selected at random from the first N integers; $E_1(rs) = (3N + 2)(N + 1)/12$, the expectation of the product of two distinct integers selected at random from the first N integers; and $E_2(rs) = (N + 1)^2/4$, the expectation of the product of two integers selected at random separately from two sets of the first N integers. Coefficients of E_1 and E_2 in (14) are the appropriate weighting probabilities. Substitution in (14) and then (13) yields

$$\begin{aligned} E(W^2) = {} & [m(N + 1)(2N + 1)/6] \\ & + [m(m - 1)(N + 1)(3N^2 + N - 1)/\{6(2N - 1)\}]. \end{aligned} \tag{15}$$

Then, from (15) and (12) with replacement of N by $(m + n)/2$, we have

$$\mu_2 = \sigma_W^2 = mn(m + n - 2)(m + n + 2)/[48(m + n - 1)]. \tag{16}$$

Through similar arguments,

$$\mu_3 = 0 \tag{17}$$

and

$$\begin{aligned} \mu_4 = {} & \frac{mn\,(m + n + 2)}{3840\,(m + n - 3)(m + n - 2)(m + n - 1)}[5mn\,(m + n)^4 \\ & - 2(m^5 + 19m^4n + 52m^3n^2 + 52m^2n^3 + 19mn^4 + n^5) \\ & + 4(3m^4 + 16m^3n + 26m^2n^2 + 16mn^3 + 3n^4) \end{aligned} \tag{18}$$

$$- 4(6m^3 - 34m^2n - 34mn^2 + 6n^3)$$
$$- 16(2m^2 + 25mn + 2n^2) + 96(m + n)].$$

When $m + n = 2N + 1$, derivations of moments are slightly more complicated but similar. We obtain

$$E(W) = m(m + n + 1)^2/[4(m + n)], \tag{19}$$

$$\mu_2 = \sigma_W^2 = mn(m + n + 1)[3 + (m + n)^2]/[48(m + n)^2], \tag{20}$$

$$\mu_3 = mn(n - m)(m + n - 1)(m + n + 1)^2/[32(m + n - 2)(m + n)^3], \tag{21}$$

and

$$\mu_4 = \frac{mn(m + n + 1)}{3840\,(m + n - 2)(m + n)^4}[5mn(m + n)^6 - (2m^7 + 17m^6n + 57m^5n^2 + 100m^4n^3 + 100m^3n^4 + 57m^2n^5 + 17mn^6 + 2n^7) + 2(m^6 + 14m^5n + 47m^4n^2 + 68m^3n^3 + 47m^2n^4 + 14mn^5 + n^6) + 2(2m^5 - 35m^4n - 115m^3n^2 - 115m^2n^3 - 35mn^4 + 2n^5) + 15(4m^4 - m^3n - 10m^2n^2 - mn^3 + 4n^4) + 15(2m^3 + 9m^2n + 9mn^2 + 2n^3) - 30(m^2 - mn + n^2)]. \tag{22}$$

The moments above are sufficient to show that

$$\mu_3/\mu_2^{3/2} = 0, \qquad m + n = 2N, \tag{23}$$

$$\mu_3/\mu_2^{3/2} = O(N^{-3/2}), \qquad m + n = 2N + 1, \tag{24}$$

and

$$\mu_4/\mu_2^2 = 3 + O(N^{-1}), \quad m + n = 2N \text{ or } 2N + 1. \tag{25}$$

The limits are considered as $N \to \infty$ with m/n constant. These results suggest the use of

$$u = [W - E(W) \pm \tfrac{1}{2}]/\sigma_W \tag{26}$$

as a standard normal deviate for large m and n and with $E(W)$ and σ_W obtained from (12) and (16) or (19) and (20) as $m + n = 2N$ or $2N + 1$. As is often done in similar situations, $\frac{1}{2}$ in the numerator of (26) is a continuity correction with the sign chosen to diminish the numerator numerically. Comparisons with the exact distributions of W have shown that the use of the continuity correction is advantageous.

In Table 2 we have considered two situations: $m = 3$, $n = 11$ and $m = 7$, $n = 7$. Cumulative probabilities, $P(W \leqq W_0)$, are shown and the corresponding probabilities based on the normal approximation. It is seen that the normal approximation is quite useful at these values of m and n and somewhat better when $m = n$ than when $m \neq n$.

The Pearson system of frequency curves may be used to obtain somewhat

TABLE 2

Comparisons of $P(W \leqq W_0)$ Based on Exact and Normal Approximations to the Distributions of W*

$m = 3, n = 11$ (The distribution of W is symmetric about $W = 12$)

W_0	4	5	6	7	8	9	10	11
Exact Prob.	.0055	.0165	.0440	.0824	.1429	.2253	.3297	.4396
Normal Approx.	.0093	.0207	.0422	.0775	.1357	.2165	.3188	.4376

$m = 7, n = 7$ (The distribution of W is symmetric about $W = 28$)

W_0	16	17	18	19	20	21	22	23	24	25	26	27
Exact Prob.	.0006	.0017	.0052	.0122	.0256	.0466	.0804	.1270	.1894	.2652	.3537	.4493
Normal Approx.	.0016	.0035	.0073	.0141	.0232	.0478	.0794	.1233	.1742	.2604	.3502	.4487

* The continuity correction has been used.

better approximations to percentage points of the distribution of W, particularly when $m \neq n$. The statistic u in (26) is again computed but now we also require $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$. Table 42 in [9] is then entered with appropriate values of β_1 and β_2 and selected percentage points of the distribution of u are read from the table. Trial use of this method suggests that it is better than the normal approximation but we believe that the latter is sufficiently good for practical purposes when $m + n$ exceeds values in Table 1.

Ansari [1] has shown additional tables like Table 2 and also illustrated the use of the Pearsonian approximation.

5. Limiting Normality. Limiting normality of W is established through use of a theorem of Chernoff and Savage [3]. We first define their notation and then show how the theorem applies to W.

Chernoff and Savage consider two samples as we have done in Section 2. They define $m + n = N$, $\lambda_N = m/N$ and require that for all N the inequalities

$$0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1$$

hold for some $\lambda_0 \leqq \frac{1}{2}$. Sample c.d.f.'s are defined:

$$F_m(x) = (\text{number of } X_i \leqq x)/m, \qquad G_n(x) = (\text{number of } Y_i \leqq x)/n.$$

The combined sample c.d.f. is $H_N(x) = \lambda_N F_m(x) + (1 - \lambda_N)G_n(x)$; the combined population c.d.f. is $H(x) = \lambda_N F(x) + (1 - \lambda_N)G(x)$. A statistic T_N is defined in two equivalent ways. Firstly,

$$T_N = \int_{-\infty}^{\infty} J_N[H_N(x)]\, dF_m(x), \tag{27}$$

where J_N need be defined only at $1/N, 2/N, \cdots, N/N$ but may have its domain of definition extended to (0, 1) by a suitable convention. Secondly,

$$mT_N = \sum_{i=1}^{N} E_{Ni} Z_{Ni}, \tag{28}$$

where the E_{Ni} are given numbers and $Z_{Ni} = 1$ if Z_i is an X and $Z_{Ni} = 0$ otherwise. The theorem, subject to four conditions, states that

$$\lim_{N\to\infty} P[(T_N - \mu_N)/\sigma_N \leqq t] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \tag{29}$$

with μ_N and σ_N given in terms of quantities here defined.

Details on the application of the theorem are given in [1] but omitted here for brevity. W and T_N are associated with

$$T_N = W/mN. \tag{30}$$

The association follows when we define

$$J_N[H_N(x)] = \tfrac{1}{2} + \frac{1}{2N} - \left| \tfrac{1}{2} + \frac{1}{2N} - H_N(x) \right| \tag{31}$$

and

$$E_{Ni} = \tfrac{1}{2} + \frac{1}{2N} - \left| \tfrac{1}{2} + \frac{1}{2N} - \frac{i}{N} \right|, \qquad i = 1, \cdots, N. \tag{32}$$

The four conditions of the theorem may be checked except that the fourth does not hold when $H = \frac{1}{2}$, a point of measure zero and an exception permitted when the proof of the theorem is reviewed.

We may evaluate μ_N and σ_N^2 under H_0 where $F(x) \equiv G(x)$ and obtain $\mu_N = \frac{1}{4}$ and $\sigma_N^2 = n/(48\, mN)$, results asymptotically equivalent to (12) or (19) and (16) or (20) respectively. In practice, in applying the limiting normal distribution under H_0, we recommend the normal approximation outlined in Section 4.

The establishment of the limiting normality of W under H_0 in (1) and under alternatives with $\theta \neq 1$ is required in the following discussions of relative efficiencies.

6. Consistency of the W-Test. Consistency of the W-Test of (1),

$$H_0\colon\ \theta = 1, \qquad F(u) \equiv G(u),$$

against alternatives, H_a: $F(\theta u) \equiv G(u)$, $\theta \neq 1$, is indicated by the Chernoff-Savage Theorem. When H_0 is true, $T_N \xrightarrow{p} \frac{1}{4}$ since $\sigma_N^2 = n/(48\, mN) \to 0$ as $m, n \to \infty$ in constant ratio, or when λ_N is bounded as required by the theorem. When H_a is true, it can be seen from the theorem that $\sigma_N^2 \to 0$ and $T_N \xrightarrow{p} \mu_N$ with

$$|\mu_N - \tfrac{1}{4}| = (1 - \lambda_N) \int_{-\infty}^{\infty} |F(x) - F(\theta x)|\, dF(x) > 0, \qquad \theta \neq 1, \tag{33}$$

the last result depending on a zero median, $F(0) = \frac{1}{2}$. The test based on T_N is consistent and consequently the equivalent W-test is consistent against the alternatives indicated.

7. Relative Efficiencies. The relative efficiency of the W-test in comparison with other tests of dispersion may be obtained following the method of Pitman and Noether [10]. Local alternatives are considered and we define

$$\theta_N = 1 + \gamma/\sqrt{N}, \tag{33}$$

with $N = m + n$, θ_N replacing θ and now dependent on N as required by Noether.

The efficacy E_W of the W-test, or equivalently of the T_N-test, is required and is

$$E_W = [dE(T_N \mid \theta)/d\theta \mid_{\theta=1}]^2/\sigma_N^2(\theta)|_{\theta=1}\,. \tag{34}$$

E_W is evaluated through placement of $n/48mN$ in (34) for the denominator and through differentiation of μ_N with respect to θ which, with our definition of T_N, has the special form

$$E(T_N \mid \theta) = \mu_N = \int_{-\infty}^{\infty} \tfrac{1}{2} - | \tfrac{1}{2} - \lambda_N F(x) - (1 - \lambda_N)F(\theta x) \mid dF(x). \tag{35}$$

We differentiate under the integral sign and obtain

$$dE(T_N \mid \theta)/d\theta \mid_{\theta=1} = (1 - \lambda_N)\left[\int_{-\infty}^{0} xf^2(x)\,dx - \int_0^{\infty} xf^2(x)\,dx\right], \tag{36}$$

where $f(x)$ is the density function associated with $F(x)$. Now

$$E_W = \frac{48mn}{m+n}\left[\int_{-\infty}^{0} xf^2(x)\,dx - \int_0^{\infty} xf^2(x)\,dx\right]^2 \tag{37}$$

when λ_N is replaced by m/N.

E_W in (37) is correct except for terms $O(1/N)$, for the derivation was based on asymptotic results from the Chernoff-Savage Theorem. We have also obtained E_W more directly but the derivation is lengthy and will not be given in detail. We refer to the definition of W in (6) and consider, for illustration, the case with $m + n$ even. The random variables in (6) are the δ_i^x. Let $X_{[\alpha]}$ be the αth smallest X and then

$$\begin{aligned} P(\delta_i^x = 1) &= \sum_{\alpha=\max.(1,i-n)}^{\min.(i,m)} P(Z_i = x_{[\alpha]}) \\ &= \sum_{\alpha} [m!n!/\{(\alpha-1)!(m-\alpha)!(i-\alpha)!(n-1+\alpha)!\}] \\ &\quad \cdot \int_{-\infty}^{\infty} [F(x)]^{\alpha-1}[1 - F(x)]^{m-\alpha}[F(\theta x)]^{i-\alpha}[1 - F(\theta x)]^{n-i+\alpha} f(x)\,dx. \end{aligned} \tag{38}$$

From (38) and (6) expressions for $E(W \mid \theta)$ and $dE(W \mid \theta)/d\theta\,|_{\theta=1}$ may be obtained and reduced, the final reduction based on an interesting application of the method of steepest descent in the evaluation of integrals. We used the form for σ_W^2 in (16) and obtained a result asymptotically equivalent to (37). A form similar to (38) was also obtained with $m + n$ odd.

Noether set forth four conditions for the validity of the calculation of asymptotic relative efficiencies. The first three are easily checked for the W-test and the fourth involves uniform limiting normality which follows from the Chernoff-Savage Theorem.

The efficacy of the F-test for variances is

$$E_F = 4mn/[(m + n)(\beta_2 - 1)] \tag{39}$$

where

$$\beta_2 = \int_{-\infty}^{\infty} [x - E(x)]^4 \, dF(x) \Big/ \left[\int_{-\infty}^{\infty} \{x - E(x)\}^2 \, dF(x) \right]^2 \tag{40}$$

as described by Sukhatme [13]. The relative efficiency of the W-test to the F-test is $e_{WF} = \lim_{N\to\infty} E_W/E_F$ and reduces to

$$e_{WF} = 12(\beta_2 - 1) \left[\int_{-\infty}^{0} xf^2(x) \, dx - \int_{0}^{\infty} xf^2(x) \, dx \right]^2. \tag{41}$$

Special cases follow:

(i). If $f(x) = e^{-x^2/2}/\sqrt{2\pi}$, $e_{WF} = 6/\pi^2 = .61$.

(ii). If $f(x) = 1$, $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$, $e_{WF} = .60$.

(iii). If $f(x) = \frac{1}{2}e^{-|x|}$, $e_{WF} = .94$.

8. Other Tests of Dispersion. The W-test has the same relative efficiencies as Sukhatme's first test [13], as might be expected from their similarity.

Sukhatme [14] has proposed a second test. The statistic is

$$\begin{aligned} S = {} & \sum_{i=1}^{n} \sum_{\substack{j,k=1 \\ j \neq k}}^{m} Q(X_j, X_k, Y_i) + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \sum_{k=1}^{m} Q(X_k, Y_i, Y_j) \\ & + \frac{m + n - 2}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} K(X_j, Y_i), \end{aligned} \tag{42}$$

where

$$\begin{aligned} Q(u, v, w) &= 1 \quad \text{if} \quad 0 < u < w, 0 < v < w \quad \text{or} \quad w < u < 0, w < v < 0, \\ &= 0 \text{ otherwise} \end{aligned}$$

and

$$\begin{aligned} K(u, v) &= 1 \quad \text{if} \quad 0 < u < v \quad \text{or} \quad v < u < 0, \\ &= 0 \text{ otherwise.} \end{aligned}$$

The relative efficiency e_{SF} is

$$e_{SF} = \frac{720}{61} (\beta_2 - 1) \left[2 \int_{-\infty}^{\infty} xF(x)f^2(x) \, dx - \int_{-\infty}^{0} xf^2(x) \, dx \right]^2. \tag{43}$$

In the normal case (i), $e_{SF} = .69$; in the uniform case (ii), $e_{SF} = .80$; in the double exponential case (iii), $e_{SF} = 1.03$. Sukhatme's S-test requires knowledge of the locations of the two populations.

Mood [8] has proposed the statistic

$$M = \sum_{i=1}^{m} \left(r_i - \frac{m+n+1}{2}\right)^2 \tag{44}$$

where r_i is the rank of X_i in (2). The relative efficiency of Mood's test, as derived in [13], is

$$e_{MF} = 45(\beta_2 - 1)\left[2\int_{-\infty}^{\infty} xF(x)f^2(x)\,dx - \int_{-\infty}^{\infty} xf^2(x)\,dx\right]^2. \tag{45}$$

e_{MF} has values (i) .76, (ii) 1, and (iii) 1.08 respectively for the three distributions considered. Mood's test requires only knowledge of the relative locations of the two populations.

Lehmann ([7], pp. 173) has proposed a test that does not depend on knowledge of even the relative locations of the two populations but even the null distribution of his statistic is not distribution-free. The statistic, in a form given by Sukhatme [13], is

$$L = \sum_{i<j}\sum_{h<k} \phi(|X_i - X_j|, |Y_h - Y_k|) \Big/ \binom{m}{2}\binom{n}{2} \tag{46}$$

where $\phi(u, v) = 1$ if $u < v$ and $\phi(u, v) = 0$ otherwise. Relative efficiencies are not known since difficulties are introduced because of the dependency of the test on the natures of the populations sampled.

The properties of the David and Barton test are those of the W-test in the special case in which the two are equivalent.

Relative efficiencies have been shown for tests discussed in comparison with the F-test for variances. The relative efficiency of one rank test to another may be obtained from the ratio of the two relative efficiencies given.

The W-test is an improvement on Sukhatme's first test but is less efficient, though easier to use, than Sukhatme's second test. Mood's test is the most natural one against the background of normal-theory statistics and its efficiencies are the best. The W-test is somewhat easier to compute and with tables may be useful in many situations where a quick and easy test is desired.

9. Discussion. It is a disadvantage in the W-test that the relative locations of the two populations must be known. This disadvantage—or more serious ones—is also present for the other tests discussed. If the X- and Y-samples cannot be adjusted so that $\mu_X - \mu_Y = 0$, differences in locations seriously affect all of the tests of dispersion.

We would like to modify the W-test so that the X- and Y-samples are adjusted in locations on the basis of information from the sample itself. One possibility is to consider the sample medians $\tilde{X}$ and $\tilde{Y}$, the middle or averages of middle observations for odd- or even-size samples respectively, and let $U_i =$

TABLE 3

The Null Cumulative Distribution of $\tilde{W}$ for $m = n = 9$ from a Sampling Study and Corresponding Expected Frequencies for W Computed When $m = n = 8$

$\tilde{W}$	27	28	29	30	31	32	33	34	35	36
Observed Cum. Freq.	5	7	12	14	17	24	36	46	55	63
Frequency for W	3.6	5.7	8.7	12.6	17.5	23.5	30.3	37.9	45.9	54.1
$\tilde{W}$	37	38	39	40	41	42	43	44	—	48
Observed Cum. Freq.	69	80	84	87	91	95	97	99	—	100
Frequency for W	62.1	69.7	76.5	82.4	87.4	91.3	94.3	96.4	—	99.7

$X_i - \tilde{X}$ and $V_j = Y_j - \tilde{Y}$. A new array like (2) may be formed from the samples of U and V. We would like to again compute the W-statistic for this new array and refer to the test based on it as the $\tilde{W}$-test. But the distributions of $\tilde{W}$ are unknown and very difficult to investigate since the U's and V's are not independent. It does appear that the $\tilde{W}$-test should have the same asymptotic properties as the W-test, but this has not been verified.

We would like to use the $\tilde{W}$-test as though it were the W-test. The appropriate way to do this seems to be to drop out the zero-values of U and V when they occur and to then proceed with the W-test on the reduced samples of U and V. A check on the appropriateness of this was made by a sampling study. One hundred pairs of samples with $m = n = 9$ were taken from a table of random normal deviates and reduced to samples of U's and V's of eight each; the distribution of $\tilde{W}$ obtained is shown in Table 3. The mean and variance of $\tilde{W}$ from the sampling study were respectively 35.3 and 19.7. The corresponding values for a W-test with $m = n = 8$ obtained from (12) and (16) are 36 and 22.5. This suggests that the normal approximation for the W-test may be used for the $\tilde{W}$-test, but this limited study is not conclusive.

Other generalizations may be possible and merit investigation. The W-test may be extended to several samples for a rank analogue to a test of homogeneity of variances. Problems associated with the largest or smallest scale parameter of a set of populations might be considered.

The problem of ties has not been investigated. The effects of ties could be studied, but we suggest that the usual procedure of giving a tied rank the average rank for the set of tied values should be adequate.

We have chosen to consider the test of H_0: $F(u) \equiv G(u)$ against alternatives $F(\theta u) \equiv G(u)$, $\theta \neq 1$ given that X- and Y-populations have, or may be adjusted to have, a common, but not necessarily known, location as measured by their medians. It is interesting to consider what may happen if other alternatives are met. Firstly, we note that the proof of consistency of the W-test in Section 6 is dependent on a common median (taken to be zero without loss of generality) for the two populations but that otherwise (33) applies for any $G(x) \not\equiv F(x)$

replacing $F(\theta x)$ in (33). Hence the W-test is consistent against a much wider class of alternatives although power should be best for the situation considered and this seems to be the important one. The W-test can lose its sensitivity for detecting differences between dispersions in the presence of differences between medians. If the difference between population medians is large and dispersions are relatively small, it can happen that all X-observations precede all Y-observations in (2). If $m = n$ also, then $W = E(W)$ and H_0 would not be rejected even if there are differences in dispersions. It is important, as stated in [5] and similarly in [12], that in general a rejection of H_0 may be attributed to differences in dispersions. There is one exception: if m is very small compared to n, a very small value of W could be due to a difference in medians but this should be immediately apparent from an inspection of the data.

A concluding remark may be made. There is some interest in possible forms that statistics may take. Sukhatme's statistic [13] may be adjusted to estimate the probability $P(|X| < |Y|)$. The situation is not so clear for the W-test. We note that $W - \text{Min. } W$ is a count of the numbers of X's nearer the combined sample median than Y's. If we consider $(W - \text{Min. } W)/mn$, asymptotically we have an estimate of $P(|X - \mu| < |Y - \mu|)$ when μ is the common median of the two populations. This asymptotic result does suggest that the W-test will be consistent against alternatives for which

$$P(|X - \mu| < |Y - \mu|) \neq \tfrac{1}{2}.$$

10. Acknowledgments. We are pleased to acknowledge the assistance given us by John E. Freund through his original proposal of the problem and his work with A. R. Ansari in [5]. R. E. Bargmann has also been most helpful through discussions on various points included in the paper. Discussions with I. R. Savage at the 1958 Summer Statistical Institute[4] were particularly stimulating.

REFERENCES

[1] ANSARI, A. R., *Two Way Rank-Sum Tests for Variances*, Ph.D. Thesis, Virginia Polytechnic Institute Library, Blacksburg, Virginia, (1959).

[2] BARTON, D. E., "The limiting distribution of Kamat's test statistic," *Biometrika*, Vol. 43 (1956), pp. 386–387.

[3] CHERNOFF, HERMAN AND SAVAGE, I. RICHARD, "Asymptotic normality and efficiency of certain non-parametric test statistics," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 972–994.

[4] DAVID, FLORENCE N. AND BARTON, D. E., "A test for birth-order effects," *Ann. Human Genetics*, Vol. 22 (1958), pp. 250–257.

[5] FREUND, J. E. AND ANSARI, A. R., *Two-way Rank-Sum Tests for Variances*, Technical Report No. 34 to Office of Ordnance Research and National Science Foundation, Virginia Polytechnic Institute, August (1957).

[6] KAMAT, A. R., "A two-sample distribution-free test," *Biometrika*, Vol. 43 (1956), pp. 377–385.

[4] A. R. Ansari wishes to acknowledge a grant from the National Science Foundation allowing him to participate in the Institute.

[7] LEHMANN, E. L., "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165–179.

[8] MOOD, A. M., "On the asymptotic efficiency of certain non-parametric two-sample tests," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 514–522.

[9] PEARSON, E. S. AND HARTLEY, H. O., *Biometrika Tables for Statisticians*, Vol. 1, Cambridge University Press, Cambridge, (1956).

[10] NOETHER, GOTTFRIED E., "On a theorem of Pitman," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 64–68.

[11] ROSENBAUM, S., Tables for a non-parametric test of dispersion, *Ann. Math. Stat.*, Vol. 24 (1953), pp. 663–668.

[12] SIEGEL, SIDNEY AND TUKEY, JOHN W., "A nonparametric sum of ranks procedure for relative spread in unpaired samples," *J. Amer. Stat. Assn.*, Vol. 55 (1960), pp. 429–445.

[13] SUKHATME, BALKRISHNA V., "On certain two-sample non-parametric tests for variances," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 188–194.

[14] SUKHATME, BALKRISHNA V., "A two-sample distribution-free test for variances," *Biometrika*, Vol. 45 (1958), pp. 544–548.

[15] TERRY, MILTON E., *Some Rank-Order Tests which are Most Powerful Against Specific Parametric Alternatives*, Ph.D. Thesis, University of North Carolina Library, Chapel Hill, North Carolina, (1951).