Journal of the Korean
Statistical Society
Vol. 23, No. 1, 1994

# A Comparison of Distribution-free Two-sample Procedures Based on Placements or Ranks

**Dongjae Kim** [1]

## ABSTRACT

We discussed a comparison of distribution-free two-sample procedures based on placements or ranks. Iterative asymptotic distribution of both two-sample procedures is studied and small sample Monte Carlo simulation results are presented. Also, we proposed the Hodges-Lehmann type location estimator based on linear placement statistics.

**KEYWORDS:** Distribution-free test, Linear rank statistic, Linear placement statistic, Hodges-Lehmann location estimator

## 1. INTRODUCTION

One of the most commonly encountered statistical problems is that of determining whether two independent random samples arise from a common underlying distribution. Test procedures for the two-sample setting that are valid when no assumptions except continuity are placed on the forms of the underlying distributions are referred to as nonparametric or distribution-free

[1] Department of Biostatistics, Catholic University Medical College, 505 Banpo-Dong, Seocho-Gu, Seoul 137-701, Korea

procedures.

Many such methods are based on the relative ranks of the sample observations from one population among the combined set of sample observations from both populations. One particular collection of these rank tests is associated with the class of linear rank statistics (see, for example, Chapter 8 of Randles and Wolfe (1979)). This class includes the Wilcoxon (1945) rank sum test and the Mood (1950) median test.

A second method for constructing distribution-free tests in this setting utilizes the placements of the sample observations from one population among the sample items of the other population. The class of linear placement tests introduced by Orban and Wolfe (1982) is one such collection of placement tests. They showed that the classes of linear rank tests and linear placement tests are almost mutually exclusive, as the Mann-Whitney (1947) and Wilcoxon (1945) statistics lead to the only test procedure with an equivalent form in each of the classes.

In this paper we compare the asymptotic behavior of the classes of linear rank statistics and linear placement statistics as one of the sample sizes goes to infinity. A Monte Carlo simulation study of small sample power in a variety of setting is performed for some of these procedures. Finally, we investigate the Hodges-Lehmann location parameter estimators associated with the class of linear placement statistics.

## 2. A REVIEW

Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be independent random samples from populations with distribution functions $F(x)$ and $G(y)$, respectively. Let $R_1, R_2, \cdots, R_n$ be the ranks of $Y_1, Y_2, \cdots, Y_n$, respectively, among the combined set of $N = m + n$ sample observations from both populations. The class of two-sample linear rank statistics corresponds to statistics of the form

$$S_{n,m}^R = \sum_{i=1}^{n} a_N(R_i), \tag{2.1}$$

where $a_N(1), a_N(2), \cdots, a_N(N)$ is a set of N constants that are not all the same. These constants are called the scores.

Many statisticians have contributed to the literature on two-sample linear rank statistics. Wilcoxon (1945) suggested the rank sum test and the statistic for a two-sample linear rank median test is attributed to Mood (1950) and Westenberg (1948). The use of expected value normal scores in the two-sample problem was first proposed by Fisher and Yates (1938) and later studies by Terry (1952). The quantile normal scores were developed by van der Waerden (1952). Linear rank tests for scale have been proposed by Mood (1954), Ansari and Bradley (1960), and Klotz (1962), among others.

Numerous theorems can be used to establish the asymptotic normality of a properly standardized linear rank statistic under both the null and appropriate alternative hypotheses. Chernoff and Savage (1958) established the classic limit theorem, as $\min(m, n) \to \infty$, of linear rank statistics. Further refinements on the conditions of that theorem were developed by Govindarajulu, Le Cam and Raghavachari (1966). Hajek (1968) proved limiting normality when the square intergrable score function in continuous subject to mild regularity conditions on the underlying distributions. Similar results with weaker restrictions on the score function and different conditions on the underlying distributions were obtained Pyke and Shorak (1968) and Dupac and Hajek (1969).

For the more recently developed class of linear placement statistics, let $U_1, U_2, \cdots, U_n$ be the random variables defined by

$$mU_i = [\text{number of} \quad \text{X's} \leq Y_i], \tag{2.2}$$

$i = 1, 2, \cdots, n$. We refer to $U_i$ as the placement of $Y_i$ among the X's. The class of two-sample linear placement statistics then consists of the form

$$S^P_{n,m} = \sum_{i=1}^{n} \varphi_m(U_i), \tag{2.3}$$

where $\varphi_m(x)$ is any real-valued function defined on [0,1].

The exact distribution of the placements $U = (U_1, U_2, \cdots, U_n)$ and the first two null ($F \equiv G$) moments of a general linear placement statistic were studied by Orban and Wolfe (1982). Also, they investigated the asymptotic distribution of a linear placement statistic as the single sample size m goes to infinity. They required the following two assumptions on the scoring function $\varphi_m(x)$ in order to insure convergence, as $m \to \infty$, of the associated statistic.

**Assumption 1** Let $\varphi(x)$ be a real-valued function on [0,1], with at most a finite number of discontinuities. Let $\Psi = \{d_1, d_2, \cdots, \}$ be the discontinuity set of $\varphi(x)$.

**Assumption 2** $\{\varphi_m(x)\}$ is a sequence of real-valued functions on [0,1] that converges uniformly in $x$ to $\varphi(x)$ on every closed interval $[a,b] \subset [0,1] - \Psi$.

**Lemma 1.** (Orban and Wolfe(1982)) Let $[a,b] \subset [0,1] - \Psi$ and $0 < \delta < (b-a)/2$ be given. If $\varphi(x)$ and $\{\varphi_m(x)\}$ satisfy Assumptions 1 and 2, then for each $\varepsilon > 0$,

$$P\{\, |\varphi_m[F_m(y)] - \varphi[F(y)]\,| < \varepsilon\} \to 1$$

as $m \to \infty$, uniformly for $y \in \{\, y \,|\, a + \delta < F(y) < b - \delta\}$.

A third assumption insures that the uniform convergence of $\varphi_m[F_m(y)] - \varphi[F(y)]$ to zero occurs on an sufficiently large subset of the support for the distribution of Y.

**Assumption 3** The distribution functions $F(x)$ and $G(y)$ satisfy

$$\int_{F(y)\in\Psi} d\,G(y) = 0,$$

where $\Psi$ is the discontinuity set for $\varphi(x)$.

**Theorem 1.** (Orban and Wolfe(1982)) Let $S^P_{n,m}$ be a sequence of linear placement statistics with scoring functions $\{\varphi_m(x)\}$ satisfying Assumptions 1 and 2. If $F(x)$ and $G(y)$ satisfy Assumption 3, then

$$S^P_{n,m} - S^P_n \to 0 \quad \text{in probability}, \qquad \text{as} \quad m \to \infty,$$

where $S_n^P = \sum_{i=1}^{n} \varphi[F(Y_i)]$ and $\varphi(x) = \lim_{m\to\infty} \varphi_m(x)$. Moreover,

$$\lim_{m\to\infty} P[\, S_{n,m}^P \leq x\,] = P[\, S_n^P \leq x\,].$$

Fligner and Policello (1981) introduced robust rank tests using placement of $Y_i$ among the $X$'s and placement of $X_i$ among the $Y$'s for dealing with the standard Behrens-Fisher problem.

## 3. ONE-SAMPLE LIMIT AND ITERATIVE ASYMPTOTIC DISTRIBUTION

In this section the one-sample asymptotic ($m \to \infty$) distribution of a linear rank statistic $S_{n,m}^R$ is established under certain restrictions on the scoring function $a_N(x)$. A large-sample approximation to the exact null distribution of $S_{n,m}^R$ is drived. Also, we compare the one-sample limiting distributions of analogous linear rank and linear placement statistics. For this purpose, we require assumption for the linear rank scoring function $a_N(x)$.

**Assumption 4** $a_N(i) = b_N\, \phi(\, i\,/(N+1)) + d_N$, where $d_N$ and $b_N$ are constants for every N and $\phi(x)$ is any real-valued function on [0,1].

Under Assumption 4, the linear rank test procedure associated with $S_{n,m}^R$ is equivalent to the test based on the statistic

$$S_{n,m}^{R*} = \sum_{i=1}^{n} \phi(\frac{R_i}{N+1}) \tag{3.1}$$

**Theorem 2.** Let $[S_{n,m}^{R*}]_{m=1}^{\infty}$ be a sequence of test statistics of the form (3.1) with scoring function $\phi(x)$ satisfying Assumption 1 and Assumption 2. If $F(x)$ and $G(x)$ satisfy Assumption 3, then

$$S_{n,m}^{R*} - S_n^{R*} \to 0 \quad \text{in probability}$$

as $m \to \infty$, where $S_{n,m}^{R*} = \sum_{i=1}^{n} \phi[F(Y_i)]$. Moreover,

$$\lim_{n\to\infty} P\,[\,S^{R*}_{n,m} \le x\,] = P\,[\,S^{R*}_{n} \le x\,].$$

Theorem 2 is direct application results from Theorem 1 and tells us that the test procedure based on the test statistic $S^{R*}_{n,m}$ (3.1) with $\phi(x) = \varphi(x)$ is asymptotically ($m \to \infty$) equivalent to the analogous linear placement test procedure based on $S^{P}_{n,m}$ (2.3). Therefore, a linear placement test procedure and its linear rank analogue (with the same scoring function) are equivalent in the sense of their one-sample limiting ($m \to \infty$) distribution.

Our next concern is the asymptotic distribution of $S^P_n = \sum_{i=1}^{n} \varphi[F(Y_i)]$ as a sequence in $n$. Since $Y_1, Y_2, \cdots, Y_n$ are mutually independent and identically distributed, then so are $\varphi[F(Y_i)], i = 1, 2, \cdots, n$. Applying the Central Limit Theorem, we see that asymptotic ($n \to \infty$) distribution of the standardized $S^P_n$ is standard normal distribution ; that is,

$$\frac{S^P_n - n\overline{\varphi}}{\sqrt{nV_\varphi}} \to n(0,1) \quad \text{in distribution} \tag{3.2}$$

as $n \to \infty$, where

$$\overline{\varphi} = E(\varphi\,[F(Y_i)] = \int \varphi\,[F(y)]\,d\,G(y) \tag{3.3}$$

$$\text{and} \qquad V_\varphi = \mathrm{Var}(\varphi\,[F(Y_i)] = \int [\varphi\,[F(y)] - \overline{\varphi}\,]^2 d\,G(y) \tag{3.4}$$

**Definition 1.** Let $\{T_{n,m}\}$ be a sequence of statistics depending on $m$ and $n$, and let $F_{n,m}(x)$ be the cumulative distribution function for $T_{n,m}$. We say that $T_{n,m}$ has an iternative asymptotic distribution with cumulative distribution function $F(x)$ if $\lim_{n\to\infty}\lim_{m\to\infty} F_{n,m}(x)$ at all points of continuity of $F(x)$.

From Theorem 1 and Theorem 2, we have that if $\phi(x) = \varphi(x)$ then

$$\frac{S^{R*}_{n,m} - n\overline{\varphi}}{\sqrt{nV_\varphi}} - \frac{S^{R*}_{n} - n\overline{\varphi}}{\sqrt{nV_\varphi}} \to 0 \quad \text{in probability}$$

$$\text{and} \quad \frac{S^{P}_{n,m} - n\overline{\varphi}}{\sqrt{nV_\varphi}} - \frac{S^{P}_{n} - n\overline{\varphi}}{\sqrt{nV_\varphi}} \to 0 \quad \text{in probability}$$

as $m \to \infty$. These facts, in conjunction with (3.2), imply that $\frac{S^{R*}_{n,m} - n\overline{\varphi}}{\sqrt{nV_\varphi}}$ and $\frac{S^{P}_{n,m} - n\overline{\varphi}}{\sqrt{nV_\varphi}}$ have the same asymptotic standard normal distribution with iterative sence that $m \to \infty$ and then $n \to \infty$. This common asymptotic distribution can be used to obtain an approximate expression for the power function of a given linear rank or linear placement statistic as a guideline if $m$ is sufficiently larger than $n$.

**Example 1.** For the two-sample location problem, define $G(x) = F(x-\theta)$ and assume that $\varphi(x)$ is a nondecreasing function in $x$. For testing $H_0 : \theta = 0$ versus $H_1 : \theta > 0$, we would reject $H_0$ for large values of either the linear rank or linear placement statistic, $S^P_{n,m}$ (or $S^{R*}_{n,m}$), associated with $\varphi(x)$, and the iterative asymptotic upper $100\alpha$-th critical value for either of the associated hypothesis tests is $S_\alpha = n\overline{\varphi_0} + Z_\alpha\sqrt{nV_{\varphi_0}}$. Where $Z_\alpha$ is the upper $100\alpha$-th percentile of the standard normal distribution and $\overline{\varphi_0}$ and $V_{\varphi_0}$ are given by equations (3.3) and (3.4) under $H_0 : \theta = 0$. Since $V_{\varphi_0}$ is approximately equal to the variance of $\varphi(F(Y_i))$ under an arbitrary alternative hypothesis, the approximate power function, $\beta$, for either of these tests is

$$\begin{aligned}\beta(\theta) &= P(S^P_{n,m}(\text{or}S^{R*}_{n,m}) > S_\alpha : H_1 \text{ is true})\\ &\approx P(Z \geq \frac{n\overline{\varphi_0} + Z_\alpha\sqrt{nV_{\varphi_0}} - n\overline{\varphi_\theta}}{\sqrt{nV_{\varphi_0}}})\\ &= 1 - \Phi(Z_\alpha + \frac{\sqrt{N}(\overline{\varphi_0} - \overline{\varphi_\theta})}{\sqrt{V_{\varphi_0}}}),\end{aligned}$$

where $\Phi$ is the distribution function of the standard normal distribution. Now, if $(\varphi F)'(t) = d\varphi\left[F(y)\right]/dt$ exists and is continuous for every real number $t$, then

$$\begin{aligned}\frac{\sqrt{n}\,(\overline{\varphi_0} - \overline{\varphi_\theta})}{\sqrt{V_{\varphi_0}}} &= \frac{\sqrt{n}\,[\int \varphi(F(y))\,dF(y) - \int \varphi(F(y))\,dF(y-\theta)]}{\sqrt{\int(\varphi(F(y)) - \overline{\varphi_0})^2 dF(y)}}\\ &= \frac{\sqrt{n}\,[\int \varphi(F(y))\,dF(y) - \int \varphi(F(y+\theta))\,dF(y)]}{\sqrt{\int(\varphi(F(y)) - \overline{\varphi_0})^2 dF(y)}}\end{aligned}$$

$$\approx \frac{-\theta\sqrt{n}\,[\int(\varphi F)'(y)\,d\,F(y)]}{\sqrt{\int(\varphi(F(y)) - \overline{\varphi_0})^2 d\,F(y)}},$$

where this final step follows from a first order Taylor series expansion. Therefore, the approximate iterative asymptotic power for both test procedures in this setting is

$$\beta(\theta) \approx 1 - \Phi[z_\alpha - \frac{-\theta\sqrt{n}[\int(\varphi F)'(y)\,d\,F(y)]}{\sqrt{\int(\varphi(F(y)) - \overline{\varphi_0})^2 d\,F(y)}}],$$

which clearly depends on the level $\alpha$, the alternative value of the location parameter $\theta$, the scoring function $\varphi$, and the underlying distribution F.

Now, in order to compare the empirical powers for general alternatives, we conducted a Monte Carlo simulation study. In this investigation we considered only the normal score function N and the exponential score function E. Remember that the linear placement and linear rank procedures are equivalent for the identity score function.

For our Monte Carlo study, we used five different sample size configurations, $(m,n) = (10,5),(15,5),(8,8),(5,10)$, and (5,15), and alternative $p^* =$ 0.5, 0.6, 0.7, 0.8, 0.9, where $p^* = 1 - F(-\theta)$, along with significance level $\alpha = 0.05$.

For each of the parameter settings studied, the International Mathematical and Statistical Libraries (IMSL) routines RNNOR, RNEXP and RNCHY were employed to generate random samples from normal, exponential and Cauchy distributions, respectively. For the double exponential distribution, the probability integral transformation and the routine RNUN were used to generate the sample data.

In each case, we used 10,000 replications in obtaining the various power estimates and relative power estimates (RPE), where

$$\text{RPE} = \frac{\text{simulated power estimate for placement procedure}}{\text{simulated power estimate for analogous rank procedure}}$$

These simulated relative power estimates for the members of the two classes considered in this section are presented in Table 1. The designated alternatives configurations correspond to values of $p^*$.

**Table 1.** Power comparisons of analogous two-sample linear rank and linear placement tests for normal scores(N) and exponential scores(E). The table entries are RPE values.

| | | $p^* = 0.5$ | | $p^*=0.6$ | | $p^* = 0.7$ | | $p^*=0.8$ | | $p^* = 0.9$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| m | n | N | E | N | E | N | E | N | E | N | E |
| Normal Distribution | | | | | | | | | | | |
| 10 | 5 | 1.02 | 1.02 | 1.03 | 1.02 | 1.02 | 1.01 | 1.01 | 1.01 | 1.00 | 1.00 |
| 15 | 5 | 0.98 | 0.95 | 1.00 | 0.96 | 0.99 | 0.98 | 1.00 | 0.99 | 1.00 | 1.00 |
| 8 | 8 | 0.99 | 0.95 | 0.96 | 0.95 | 0.98 | 0.97 | 0.99 | 0.98 | 1.00 | 1.00 |
| 5 | 10 | 0.96 | 0.97 | 0.97 | 0.97 | 0.97 | 0.99 | 0.98 | 0.99 | 0.99 | 1.00 |
| 5 | 15 | 0.97 | 1.02 | 1.00 | 1.00 | 0.99 | 1.00 | 0.98 | 1.00 | 1.00 | 1.00 |
| Exponential Distribution | | | | | | | | | | | |
| 10 | 5 | 1.03 | 1.04 | 0.99 | 1.00 | 0.97 | 1.02 | 0.98 | 1.03 | 1.00 | 1.02 |
| 15 | 5 | 0.97 | 0.96 | 0.96 | 0.94 | 0.96 | 0.98 | 0.98 | 1.02 | 1.00 | 1.03 |
| 8 | 8 | 0.94 | 0.92 | 0.91 | 0.94 | 0.91 | 1.00 | 0.95 | 1.06 | 0.99 | 1.09 |
| 5 | 10 | 0.98 | 0.97 | 0.89 | 0.98 | 0.89 | 1.01 | 0.94 | 1.05 | 0.98 | 1.05 |
| 5 | 15 | 0.99 | 1.04 | 0.86 | 0.98 | 0.86 | 0.99 | 0.91 | 1.02 | 0.98 | 1.03 |
| Cauchy Distribution | | | | | | | | | | | |
| 10 | 5 | 1.02 | 1.02 | 1.02 | 1.05 | 1.04 | 1.06 | 1.03 | 1.06 | 1.02 | 1.04 |
| 15 | 5 | 0.98 | 0.94 | 1.00 | 1.01 | 1.02 | 1.04 | 1.03 | 1.05 | 1.02 | 1.02 |
| 8 | 8 | 0.98 | 0.95 | 1.01 | 0.99 | 1.03 | 1.06 | 1.05 | 1.10 | 1.04 | 1.14 |
| 5 | 10 | 1.02 | 0.96 | 1.04 | 1.00 | 1.05 | 1.01 | 1.04 | 1.02 | 1.03 | 1.02 |
| 5 | 15 | 1.02 | 1.01 | 1.07 | 1.03 | 1.07 | 1.01 | 1.03 | 1.01 | 1.00 | 1.01 |
| Double Exponential Distribution | | | | | | | | | | | |
| 10 | 5 | 1.04 | 1.04 | 1.04 | 1.04 | 1.03 | 1.03 | 1.02 | 1.03 | 1.01 | 1.02 |
| 15 | 5 | 0.99 | 0.97 | 0.99 | 0.98 | 1.01 | 1.01 | 1.01 | 1.02 | 1.01 | 1.01 |
| 8 | 8 | 0.97 | 0.95 | 0.98 | 0.98 | 1.00 | 1.00 | 1.01 | 1.03 | 1.01 | 1.03 |
| 5 | 10 | 0.97 | 0.97 | 1.00 | 0.99 | 1.01 | 0.99 | 1.01 | 1.00 | 1.00 | 1.01 |
| 5 | 15 | 1.02 | 1.02 | 1.05 | 1.02 | 1.03 | 1.01 | 1.01 | 1.00 | 0.99 | 1.01 |

The simulation results suggest several conclusions. The corresponding members of the two classes of distribution-free two-sample procedures are generally equivalent for all studied configurations and underlying distributions. However, we did find some trends in this investigation. If the parameter $\theta$

goes far from 0, the RPE values tend to be the largest. Also, for sample sizes $(m,n) = (15,5),(8,8)$, and (10,5), the distribution-free two-sample procedure based on placements are better, while the reverse is true for other sample size configurations. Therefore it appears that the placement test is better than the analogous rank test if $m$ is large relative to $n$. Finally, the placement test appears to be better than the rank test under the Cauchy and double exponential distributions, indicating that a distribution-free two-sample procedure based on placements is a viable alternative to the analogous linear rank procedure for heavy-tailed, symmetric underlying distributions.

## 4. HODGES-LEHMANN LOCATION ESTIMATORS BASED ON PLACEMENTS

In this section we consider a procedure for obtaining a point estimator for the location parameter from a linear placement statistic. This important technique was first proposed by Hodges and Lehmann (1963) for rank tests in the one-sample location setting and then extended to the two-sample location problem. Our estimators correspond to direct application of their technique to two-sample linear placement statistics.

Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be independent random samples from continuous distributions with cumulative distribution functions $F(x)$ and $F(x-\theta)$, respectively. Let $S(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n)$ be a linear placement statistic with a nondecreasing scoring function $\varphi_m(x)$ for testing $H_0; \theta = 0$ versus $H_1 : \theta > 0$ ; that is,

$$S(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n) = \sum_{i=1}^{n} \varphi_m(U_i),$$

where $U_i$ is defined in (2.2), $i = 1, 2, \cdots, n$.

In order to develop the Hodges-Lehmann location estimator associated with $S(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n)$, we need to verify the following three conditions on the linear placement statistic.

A1 $H_0 : \theta = 0$ is rejected for large value of $S(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n)$

A2 $S(x_1, x_2, \cdots, x_m\,;\, y_1 + h, \cdots, y_n + h)$ is nondecreasing function of $h$ for each $S(x_1, \cdots, x_m\,;\, y_1, \cdots, y_n)$

A3 When $\theta = 0$, the distribution of $S(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n)$ is symmetric about some value $\xi$ for every continuous $F(x)$.

Conditions A1 and A2 follow from the facts thst $\varphi_m$ is a nondecreasing function and that $U_i$ based on $y_i$ is no greater than the same $U_i$ based on $y_i + h, j = 1, 2, \cdots, n$, and $h > 0$.

**Theorem 3.** If the scoring function $\varphi_m(x)$ satisfies $\varphi_m(1-x) = \xi^* - \varphi_m(x)$ for some constant $\xi^*$ and for all $x \in [0,1]$, then the linear placement statistic $S(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n)$ satisfies Condition A3.

**Proof.** If $H_0 : \theta = 0$ is true, we see from Orban and Wolfe (1982) that

$$P_0\,[\,m\,U = (r_1, r_2, \cdots, r_n)\,] = \frac{m!\prod_{j=0}^{m} t_j!}{(m+n)!},$$

for any vector $\boldsymbol{r}$ containing $t_j$ values of $j$, $j = 0, 1, \cdots, m$, with $0 \le t_j \le n$ and $\sum_{j=0}^{m} t_j = n$ ; the null probability is 0, otherwise. Let $\boldsymbol{r}' = (m - r_1, m - r_2, \cdots, m - r_n)$. If $\boldsymbol{r}$ contains $t_j$ values of $j$, then $\boldsymbol{r}'$ contains $t_j$ values of $m - j, j = 1, 2, \cdots, m$ with $0 \le t_j \le n$ and $\sum_{j=0}^{m} t_j = n$. This implies that

$$P_0(m\,U = \boldsymbol{r}') = P_0(m\,U = \boldsymbol{r}), \tag{4.1}$$

for any vector $\boldsymbol{r}$. Also we know that $m\,U = \boldsymbol{r}'$ if and only if $m\,U' = \boldsymbol{r}$, where $U = (U_1, U_2, \cdots, U_n)$ and $U' = (1 - U_1, 1 - U_2, \cdots, 1 - U_n)$. It then follows from (8) that

$$P_0(m\,U = \boldsymbol{r}) = P_0(m\,U' = \boldsymbol{r}),$$

for any vector $\boldsymbol{r}$. Therefore,

$$(U_1, U_2, \cdots, U_n) \overset{d}{=} (1-U_1, 1-U_2, \cdots, 1-U_n),$$

where $\overset{d}{=}$ stands for equal in distribution. From the fact that $\varphi_m$ is a measurable function, we see from Theorem 1.3.7 in Randles and Wolfe (1979) that

$$\sum_{i=1}^{n} \varphi_m(U_i) = \sum_{i=1}^{n} \varphi_m(1-U_i).$$

Using the condition that $\varphi_m(1-x) = \xi^* - \varphi_m(x)$, we obtain

$$S(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n) = \\ \xi^* n - S(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n).$$

Thus, the null distribution of $S(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n)$ is symmetric about $\delta = \frac{\xi^* n}{2}$, and Condition A3 is satisfied.

The linear placement test statistic thus satisfies all three conditions, and the associated Hodges-Lehmann estimator of $\theta$ is given by

$$\hat{\theta} = \frac{\theta^* + \theta^{**}}{2}, \tag{4.2}$$

where $\theta^* = \sup\{\theta : S(X_1, X_2, \cdots, X_m\,;\, Y_1-\theta, \cdots, Y_n-\theta) > \xi\}$ and
$\theta^{**} = \inf\{\theta : S(X_1, X_2, \cdots, X_m\,;\, Y_1-\theta, Y_2-\theta, \cdots, Y_n-\theta) < \xi\}$

**Example 2.** Let $\varphi_m(x) = x$. Then $S(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n) = \sum_{i=1}^{n} U_i$ is proportional to the number of positive D's, where $D_{(1)} \leq D_{(2)} \leq \cdots \leq D_{(mn)}$ are the ordered values of the differences $Y_j - X_i$, $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, n$. It follows that $S(X_1, X_2, \cdots, X_m\,;\, Y_1-\theta, \cdots, Y_n-\theta)$ is proportional to the number of D's greater than $\theta$. With arguments similar to those in Example 7.1.8 in Randles and Wolfe (1979), it follows, as expected, that the resulting Hodges-Lehmann estimator for $\theta$ is

$$\hat{\theta} = \text{median}\,[Y_j - X_i \,:\, i = 1, 2, \cdots, m \quad \text{and} \quad j = 1, 2, \cdots, n].$$

**Example 3.** Let $\varphi_m(x) = 1$ if $x > 1/2\,;\, 0$ otherwise. Then $S(X_1, X_2, \cdots, X_m\,; Y_1, Y_2, \cdots, Y_n) = \sum_{i=1}^{n} \varphi_m(U_i)$ is equal to the number of Y's greater than median$\{X_1, X_2, \cdots, X_m\}$.

Thus $S(X_1, X_2, \cdots, X_m\,;\, Y_1 - \theta, \cdots, Y_n - \theta)$ is equal to the number of $D^{*'}$s greater than $\theta$, where $D_j^* = Y_j -$ median$\{X_1, X_2, \cdots, X_m\}, j = 1, 2, \cdots, n$. The associated Hodges-Lehmann estimator for $\theta$ is

$$\hat{\theta} = \text{median}\,[D_j^*\,;\, j = 1, 2, \cdots, n]$$

$$= \text{median}\{Y_1, Y_2, \cdots, Y_n\} - \text{median}\{X_1, X_2, \cdots, X_m\}$$

using similar arguments to those in Example 2. On the other hand, the Hodges-Lehmann estimator for $\theta$ based on the corresponding linear rank statistic with the same scoring function is

$$\hat{\theta}^* = \text{median}\,\{X_1, X_2, \cdots, X_m\,\} - \text{median}\,\{X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n\,\}$$

Therefore, as expected, the Hodges-Lehmann location estimator based on a linear placement statistic is not always equal to the one based on the analogous linear rank statistic.

We now turn our attention briefly to the exact distributional properties for Hodges-Lehmann estimators based on linear placement statistics. Using arguments similar to those of Lehmann 7.2.18., Theorem 7.2.21 and Corollary 7.2.31 in Randles and Wolfe (1979), a linear placement Hodges-Lehmann estimator has the following results.

**Theorem 4.** Let $\hat{\theta}(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n)$ be the Hodges-Lehmann estimator (4.2) associated with a test statistic $S(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n)$ with a nondecreasing scoring function $\varphi_m(x)$ satisfying $\varphi_m(1 - x) = \xi^* - \varphi_m(x)$ for all $x$. Then $\hat{\theta}$ is a shift statistic that is, it satisfies $\hat{\theta}(x_1, \cdots, x_m\,;\, y_i + k, \cdots, y_n + k) = \hat{\theta}(x_1, \cdots, x_m\,;\, y_1, \cdots, y_n) + k$ for all $(x_1, \cdots, x_m\,;\, y_1, \cdots, y_n)$ and $k$.

**Theorem 5.** Let $\hat{\theta}(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n)$ be as in Theorem 4. If $F(x)$ corresponds to a distribution that is symmetric about some value $\eta$, then $\hat{\theta}(X_1, X_2, \cdots, X_m\,;\, Y_1, Y_2, \cdots, Y_n)$ is symmetrically distributed about $\theta$.

**Theorem 6.** For the setting of Theorem 4, if

$$P_0(S(X_1, X_2, \cdots, X_m\, ;\, Y_1, Y_2, \cdots, Y_n) = \delta) = 0,$$

where $\delta = \xi^* n/2$ then $\hat{\theta}(X_1, X_2, \cdots, X_m\, ;\, Y_1, Y_2, \cdots, Y_n)$ is median unbiased for $\theta$.

## 5. ACKNOWLEDGEMENTS

The author would like to thank anonymous referees for helpful comments.

## REFERENCES

( 1) Ansari, A. R. and Bradley, R. A. (1960). Rank-sum tests for dispersions. *Annals of Mathematical Statistics* 31, 1174-1189.

( 2) Chernoff, H. and Savage, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Annals of Mathematical Statistics* 29, 972-994.

( 3) Dupac, V. and Hajek, J. (1969). Asymptotic normality of simple linear rank statistics under alternative II. *Annals of Mathematical Society* 40, 1992-2017.

( 4) Fisher, R. A. and Yates, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research*, 1st ed. Oliver & Boyd, Edinburgh.

( 5) Flinger, M. A. and Policello, G. E. II. (1981). Robust rank procedures for the Behrens-Fisher problem. *Journal of the American Statistical Association* 76, 162-168.

( 6) Govindarajulu, Z., LeCam, L. and Raghavachari, M. (1966). Generalizations of theorems of Chernoff and Savage on the asymptotic normality of test statistics. *Proc. Fifth Berkeley Symp.* 1, 609-638.

( 7) Hajek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Annals of Mathematical Statistics* 39, 325-346.

( 8) Hodgess, J. L. Jr.and Lehmann, E. L. (1963). Estimates of loction based on rank tests. *Annals of Mathematical Statistics* 34, 598-611.

( 9) Klotz, J. (1962). Nonparametric tests for scale. *Annals of Mathematical Statistics* 33, 498-512.

(10) Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18, 50-60.

(11) Mood, A. M. (1950). *Introduction to the Theory of Statistics.* McGraw-Hill, New York.

(12) Mood, A. M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *Annals of Mathematical Statistics* 25, 514-522.

(13) Orban, J. and Wolfe, D. A. (1982). A class of distribution-free two-sample tests based on placements. *Journal of the American Statistical Association* 77, 666-671.

(14) Pyke, R. and Shorack, G. R. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Annals of Mathematical Statistics* 39, 755-771

(15) Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the Theory of Nonparametric Statistics.* John Wiley & Sons, New York.

(16) Terry, M. E. (1952). Some rank order tests which are most powerful against specific parametric alternatives. *Annals of Mathematical Statistics* 23, 346-366.

(17) Van der Waerden, B. L. (1952). Order tests for the two-sample problem and their power. *Indagationes Mathematicae* 14, 453-458. Correction (1953) *Indagationes Mathematicae* 15, 80.

(18) Westenberg, J. (1948). Significance test for median and interquartile range in sample from continuous populations of any form. *Proc. Koninkl. Ned. Akad. Wetenshap.* 51, 353-261.

(19) Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics* 1, 80-83.