RANK ANALYSIS OF k SAMPLES

by

Dennis D. Boos

Institute of Statistics Mimeograph Series No. 1670

August 1985

ABSTRACT

A table of linear rank statistics is used to construct new tests in the one-way layout. The underlying score functions are related to Legendre polynomials and arise naturally in Pettitt's (1976) analysis of generalized Anderson-Darling statistics. The first two functions are the familiar Wilcoxon and Mood scores and the third and fourth are associated with skewness and kurtosis differences. Summary statistics are formed from row and column sums of squares of the basic table and have approximate chi-squared distributions under the null hypothesis. When location and/or scale differences are present, the observations must be standardized before interpreting some of the statistics. New results are given for the asymptotic distribution of linear rank statistics under these adjustments.

KEY WORDS AND PHRASES: Linear rank statistics, Anderson-Darling statistics, one-way layout, skewness, kurtosis.

1. INTRODUCTION

The fixed effects ANOVA which tests equality of means is often the natural first analysis when confronted with k independent samples, $X_{11},\ldots,X_{1n_1},X_{21},\ldots,X_{2n_2},\ldots,X_{k1},\ldots,X_{kn_k}$. Secondary hypotheses might concern equality of variances and/or normality assumptions. Similar goals can be pursued without normality assumptions using nonparametric tests such as the Kruskal-Wallis or Mood's test for scale. In some situations, however, one may be interested in more general comparisons beyond location and scale. Nonparametric tests such as generalized Kolmogorov-Smirnov and Cramér-von Mises statistics (see Section 2) are available, but "significant" differences are often hard to interpret. I would like to propose a table of linear rank statistics to be used as the building blocks for test statistics which have reasonably high power of detection and simplicity of interpretation as well as distribution-free null distributions. The derivation and original motivation are given in Section 2 and are essentially due to Pettitt (1976).

The basic linear rank statistics and summary statistics are given in Table 1.1. The first two rows are familiar Wilcoxon and Mood linear

---Insert Table 1.1 here---

rank statistics. That is, T_{i1} is Wilcoxon's two-sample rank sum test in standardized form for comparing group i with all the other groups combined. If $T_{i1} < 0$ (>0), then sample i tends to be shifted to the left (right) of the combined other samples. Likewise, T_{i2} is Mood's test in standardized form, and if the populations have equal location then

$T_{i2} < 0$ (>0) indicates that the ith sample has smaller (larger) scale than the combined other samples. The row weighted sum of squares are the usual Kruskal-Wallis and Mood's k-sample tests. T_{i3} and T_{i4} are new linear rank statistics designed to detect differences in skewness and kurtosis of the populations. If the populations have equal location and scale, then $T_{i3} < 0$ (>0) indicates that the ith sample is left (right) skewed *relative* to the other samples combined and $T_{i4} < 0$ (>0) indicates that the ith sample is shorter (longer) tailed than the others. T_{i3} and T_{i4} have been carefully chosen (see (2.2)) so that under H_0: all k populations are equal, the set $(T_{i1}, T_{i2}, T_{i3}, T_{i4})$ are uncorrelated and each entry has mean 0 and variance 1. Well-known linear rank statistic theory shows that under H_0 each T_{ij} converges in distribution to a standard normal as $\min(n_1,\ldots,n_k) \to \infty$, and each $T_{COLi} = \sum_{p=1}^{4} T_{ip}^{2}$ converges to a chi-squared distribution with four degress of freedom (χ_4^2). The k-sample statistics converge under H_0 to a χ_{k-1}^{2}, and the global statistic T_{GLOBE} converges to a $\chi_{4(k-1)}^{2}$.

Our main interest lies in the column and row summary statistics and the overall statistic T_{GLOBE}. The latter is a new omnibus statistic for detecting general differences among the k populations. It is an alternative to the k-sample Cramér-von Mises and Kolmogorov-Smirnov nonparametric tests with two advantages: the null limiting chi-squared distribution is very accessible and the results are simpler to interpret. T_{GLOBE} is both a weighted sum of the column statistics as shown in Table 1.1 and also the simple sum $T_{WILC} + T_{MOOD} + T_{SKEW} + T_{KURT}$ of the row summary statistics. The column summary statistics indicate which populations seem most different from the combination of the others, and the row summary statistics suggest the types of differences which exist between the populations. However, these row summary statistics must be interpreted with caution. It is well known that location differences will upset nonparametric tests for scale such as T_{MOOD} (Moses, 1963). These differences will also affect T_{SKEW} and T_{KURT}. Alignment by subtracting location estimators upsets the distribution-free property, though an *asymptotic* distribution-free property will still hold for the scale components T_{i2} and T_{MOOD} and for the kurtosis components T_{i4} and T_{KURT} if the populations are symmetric (e.g., Jureckova, 1979). Further alignment is possible by dividing by a scale estimator after subtracting off the location estimator. The affect of such alignment is studied in Section 3.

I suggest the following use of the statistics from Table 1.1. 1) Look at the summary statistics T_{COLi} and the overall statistic T_{GLOBE}. They are always valid for deciding whether any differences exist among the k populations and which populations are most different. 2) Look at the row summaries. Alignment may be necessary for T_{MOOD}, T_{SKEW}, and T_{KURT}. If T_{WILC} is large, then T_{MOOD} should be recomputed after subtracting off location estimators. If T_{MOOD} is large, then T_{SKEW} and T_{KURT} should be recomputed after alignment by scale estimators. 3) Look at the individual linear rank statistics T_{ij}. They can be useful for seeing trends across the k populations, but parameter estimates or tests for ordered alternatives may be more useful for that purpose. The individual T_{ij} can also be used to compare pairs of populations, or one may want to recompute Table 1.1 separately for each pair of interest.

The paper is organized as follows. Section 2 traces the origin of the proposed statistics. Section 3 investigates the null distributions of the distribution-free statistics with moment calculations and limited Monte Carlo work and also studies the asymptotic distribution of the aligned statistics. A new theorem for asymptotic normality of aligned rank statistics is given there. Pitman efficiencies are employed in Section 4 to justify the association of each component with a particular type of alternative. Section 5 gives Monte Carlo power results in a variety of different $k = 2$ and $k = 4$ situations, and comparisons are made with three other k-sample nonparametric tests. Section 6 contains three real data examples.

2. MOTIVATION

In analogy with the treatment sum of squares $\sum_1^k n_i(\bar{X}_i-\bar{\bar{X}})^2$ for one-way ANOVA, consider the weighted Cramér-von Mises statistic

$$\sum_{i=1}^{k} n_i \int_{-\infty}^{\infty} [F_{n_i}(x)-H_N(x)]^2 w(H_N(x))\,dH_N(x) \tag{2.1}$$

which measures deviations of the individual empirical distribution functions F_{n_i} from their weighted average $H_N=N^{-1}\sum_1^k n_i F_{n_i}$, $N=n_1+\ldots+n_k$. The usual ANOVA is designed for detecting location differences, whereas (2.1) is sensitive to all types of differences. Kiefer (1959) studied (2.1) for the standard Cramér-von Mises weight function $w(t)=1$ and Pettitt (1976) studied the Anderson-Darling weight function $w(t)=[t(1-t)]^{-1}$. For this latter weight function and $k=2$, Pettitt showed that a "continuous" version of (2.1) $AD = (N/n_2)\int_0^1 x_n^2(t)w(t)dt$ can be expressed as

$$AD = \sum_{j=1}^{\infty} \frac{B_j^2}{j(j+1)} ,$$

where the B_j are standardized versions of linear rank statistics $\sum_{i=1}^{n_1} a_N(R_{1i})$ with R_{1i} the rank of X_{1i} in the combined sample X_{11}, $\cdots, X_{1n_1}, X_{21}, \cdots, X_{2n_2}$. Here $x_n(t)$ is the empirical process formed from $R_{11}/(N+1), \cdots, R_{1n_1}/(N+1)$ and the form of the B_j is derived from the j^{th} Legendre polynomial of the first kind. In particular, B_1 is a linear function of Wilcoxon's rank sum statistic with score $a_N(i)=i$ and B_2 is a linear function of Mood's statistic with score $a_N(i)=[i-(N+1)/2]^2$. Thus, AD is a weighted sum of squares of standardized linear rank statistics which are asymptotically independent (because of the orthogonality of the Legendre polynomials), and each of which relates to a special kind of alternative. That is, B_1 is related to location,

B_2 to scale, and (as shown in Sect. 4) B_3 to skewness and B_4 to kurtosis, etc. These results are analogous to the one sample results of Durbin and Knott (1972), Stephens (1974), and Durbin, Knott, and Taylor (1975) who suggest that the first few components contain most of the information. The omnibus Anderson-Darling statistic $AD = B_1^2/2 + B_2^2/6 + B_3^2/12 + B_4^2/20 + \sum_{j=5}^{\infty} B_j^2/(j(j+1))$ gives the most weight to location, the next most weight to scale, etc., as one would typically want in terms of importance of alternatives. However, the limiting null distribution is not as accessible as that of the "Neyman-Barton smooth" type tests (see Miller and Quesenberry, 1979) which give equal weight for a finite number of B_j, i.e., $\sum_{j=1}^{\ell} B_j^2$. These latter statistics are not true omnibus tests (not consistent against all alternatives), but they focus on the most important alternatives and have null limiting χ_ℓ^2 distributions. A version of $B_1^2 + B_2^2$ has been mentioned by Duran, Tsai, and Lewis (1976). In this paper we study a version of $B_1^2 + B_2^2 + B_3^2 + B_4^2$. The proposed linear rank statistics discussed in the Introduction are explicitly given by $T_{ip} = \sum_{\ell=1}^{n_i} a_{Nip}(R_{i\ell})$, $p = 1,\ldots,4$, where

$$
\begin{aligned}
a_{Ni1}(\ell) &= \left(\frac{12}{n_i(N-n_i)(N+1)}\right)^{\frac{1}{2}}\left(\ell-\frac{N+1}{2}\right) \\
a_{Ni2}(\ell) &= \left(\frac{180}{n_i(N-n_i)(N+1)(N^2-4)}\right)^{\frac{1}{2}}\left((\ell-\frac{N+1}{2})^2-(\frac{N^2-1}{12})\right) \\
a_{Ni3}(\ell) &= \left(\frac{7}{n_i(N-n_i)(N+1)(N^2-4)(N^2-9)}\right)^{\frac{1}{2}}\left(20(\ell-\frac{N+1}{2})^3-(3N^2-7)(\ell-\frac{N+1}{2})\right) \\
a_{Ni4}(\ell) &= \left(\frac{1}{n_i(N-n_i)(N+1)(N^2-4)(N^2-9)(N^2-16)}\right)^{\frac{1}{2}}\left(210(\ell-\frac{N+1}{2})^4\right. \\
&\qquad \left. -15(3N^2-13)(\ell-\frac{N+1}{2})^2+\frac{9}{8}(N^2-9)(N^2-1)\right) .
\end{aligned}
\tag{2.2}
$$

The above linear rank statistics are associated with Legendre polynomials because of the expansion for $w(t) = [t(1-t)]^{-1}$. If $w(t) = 1$ is used, then the expansion in a Fourier series suggests the scores

$$a_{Nip}(\ell) = (-1)^p(2)^{\frac{1}{2}}\cos(p\pi\ell/(N+1)) \quad , \; p = 1, 4 \; . \tag{2.3}$$

However, Pitman efficiency calculations indicate that the power properties of the tests for $p > 1$ are not as good as (2.2) in typical situations. If $w(t) = [\phi(t)]^{-2}$, where $\phi(t)$ is the standard normal density (see De Wet and Venter, 1973), then the expansion is in terms of Hermite polynomials $H_p(x)$ and suggests the scores

$$a_{Nip}(\ell) = H_p(\Phi^{-1}(\ell/(N+1))) \quad , \; p = 1, \; 4 \; , \tag{2.4}$$

where Φ is the standard normal distribution function. T_{i1} and T_{i2} are then the familiar van der Waerden and Klotz tests. Unfortunately, in moderate size samples T_{i1} and T_{i3} and T_{i2} and T_{i4} are highly correlated, although a quadratic form involving the covariance matrix could be constructed.

3. NULL DISTRIBUTIONS

A. Finite Sample Moments for Nonaligned Statistics

As discussed in the Introduction, all the proposed nonaligned statistics are distribution-free and can easily be shown to have asymptotic normal or chi-squared distributions. However, there is always the practical question as to how good these asymptotic approximations are for small samples. The null distributions of the two-sample linear rank statistics T_{ip} may be written down in straightforward (but lengthy) fashion. I have chosen instead to calculate their skewness and kurtosis coefficients $\sqrt{\beta_1} = E(Y-\mu)^3/\sigma^3$ and $\beta_2 = E(Y-\mu)^4/\sigma^4$ for a variety of N and n_i values.

The general form of linear rank statistics is $S = \sum_{j=1}^{N} c_N(j) a_N(R_j)$ where R_j is the rank of the j^{th} observation. Let $\bar{c}_N = N^{-1}\sum_1^N c_N(j)$ and $\bar{a}_N = N^{-1}\sum_1^N a_N(j)$. Under the assumption that the rank vector $(R_1,\cdots,R_N)$ is uniformly distributed over the integers $(1,\cdots,N)$, the first 4 moments of S are given by (see, e.g., Randles and Wolfe, 1979, Ch. 8, Hajek and Sidak, 1967, p. 82)

$$E\,S = N\,\bar{c}_N\,\bar{a}_N$$

$$E(S-E\,S)^2 = \frac{1}{N-1}\sum_{j=1}^{N}[c_N(j)-\bar{c}_N]^2\sum_{j=1}^{N}[a_N(j)-\bar{a}_N]^2$$

$$E(S-E\,S)^3 = \frac{N}{(N-1)(N-2)}\sum_{j=1}^{N}[c_N(j)-\bar{c}_N]^3\sum_{j=1}^{N}[a_N(j)-\bar{a}_N]^3 \;.$$

$$\begin{aligned} E(S-ES)^4 = {} & \frac{N(N+1)}{(N-1)(N-2)(N-3)}\sum_{j=1}^{N}[c_N(j)-\bar{c}_N]^4\sum_{i=1}^{N}[a_N(j)-\bar{a}_N]^4 \\ & - \frac{3}{(N-2)(N-3)}\left[\left(\sum_{j=1}^{N}[c_N(j)-\bar{c}_N]^2\right)^2\sum_{j=1}^{N}[a_N(j)-\bar{a}_N]^4 \right. \\ & \left. + \left(\sum_{j=1}^{N}[a_N(j)-\bar{a}_N]^2\right)^2\sum_{j=1}^{N}[c_N(j)-\bar{c}_N]^4\right] \\ & + \frac{3(N^2-3N+3)}{N(N-1)(N-2)(N-3)}\left(\sum_{j=1}^{N}[c_N(j)-\bar{c}_N]^2\right)^2\left(\sum_{j=1}^{N}[a_N(j)-a_N]^2\right)^2 \;. \end{aligned}$$

For the two-sample statistics T_{ip}, $C_N(j) = 1$ if R_j is from the jth group and is 0 otherwise. Using the above moment formulas and the scores (2.2), the coefficients $\sqrt{\beta_1}$ and β_2 were calculated for a variety of N and n_i and are displayed in Table 3.1. Since all the β_2 values are less than 3.0, it

--- INSERT TABLE 3.1 HERE ---

seems likely that use of standard normal critical values will result in conservative tests when $\sqrt{\beta_1} = 0$. However, due to the discreteness of the rank tests, this need not be the case for small samples. For example, at $n_1 = 10$ and $n_2 = 10$ the square of the standardized Wilcoxon statistic satisfies $P(T_{WILC} \geq (1.645)^2) = .1052$. In general, Table 3.1 supports use of the normal approximation with the individual rank statistics whenever n_i and $N - n_i \geq 10$.

The column summary statistics $T_{COLi} = \sum_{p=1}^{4} T_{ip}^2$ converge in distribution under the null hypothesis to a χ_4^2. Here the moments are harder to calculate. The first moment is obviously 4 due to the standardization of the individual T_{ip}. For scores like (2.2) where the individual linear rank statistics are uncorrelated with variance = 1, tedious calculations yield

$$
\begin{aligned}
E(T_{COLi})^2 = &\sum_{p=1}^{4} \beta_2(p) + 2\left[N(N+1)\sum_{j=1}^{N}[C_N(j)-\bar{C}_N]^4 - 3(N-1)\left(\sum_{j=1}^{N}[C_N(j)-\bar{C}_N]^2\right)^2\right] \\
&\times \sum_{p<q}\sum_{j=1}^{N}[a_{Nip}(j)-\bar{a}_{Nip}]^2[a_{Niq}(j)-\bar{a}_{Niq}]^2/[(N-1)(N-2)(N-3)] \\
&+ 2\left[(N^2-3N+6)\left(\sum_{j=1}^{N}[C_N(j)-\bar{C}_N]^2\right)^2 - (N^2-N+6)\sum_{j=1}^{N}[C_N(j)-\bar{C}_N]^4\right] \\
&\times \sum_{p<q}\sum_{j=1}^{N}[a_{Nip}(j)-\bar{a}_{Nip}]^2\sum_{j=1}^{N}[a_{Niq}(j)-\bar{a}_{Niq}]^2/[(N-1)(N-2)(N-3)] ,
\end{aligned}
$$

where $\beta_2(p)$ is the coefficient of kurtosis of T_{ip}. For the combinations of N and n_i given in Table 3.1, the variance of T_{COLi} is 4.462, 4.09., 5.678, 5.267, 6.765, 6.546, 7.367, and 7.264 respectively. These values suggest that use of χ_4^2 critical values (variance = 8) leads to conservative tests. However, the simple adjustment $T_{ADJi} = [(T_{COLi} - 4)/\sqrt{Var(T_{COLi})}]\sqrt{8} + 4$ seems to give percentiles close to that of the χ_4^2 distribution. For example, at $\alpha = .10$ the critical value from a χ_4^2 distribution is 7.78, and Monte Carlo replication of the adjusted statistic at $(n_1 = 10,\ n_2 = 10)$, $(n_1 = 10,\ n_2 = 30)$, and $(n_1 = 20,\ n_2 = 20)$ gave estimated critical values of 7.95, 7.83, and 7.88, respectively.

Moment calculations for T_{GLOBE} when $k > 2$ are very messy, but we can anticipate from the $k = 2$ case above that use of chi-squared critical values will lead to conservative tests. However, the above adjusted column summary statistics may also be added to yield the adjusted overall statistic

$$T_{ADJ} = \sum_{i=1}^{k} \left(\frac{N-n_i}{N}\right) T_{ADJi} \ .$$

Table 3.2 shows that T_{ADJ} is an improvement over T_{GLOBE} but still slightly conservative.

Lastly, we could calculate second moments of the k-sample tests T_{WILC}, etc., but other work (see references at the end of Sections 5.2 and 5.3 of Conover, 1980) and the results in Table 3.2 suggest that the χ_{k-1}^2 limiting distributions are reached fairly quickly. Approximations to the distribution function of the row summary statistics may be obtained from Robinson (1980).

--- INSERT TABLE 3.2 HERE ---

B. Asymptotic Distributions Under Alignment

Here we are interested in the null distribution of the linear rank statistics after adjusting within each sample for location differences and then after adjusting for both location and scale differences. Most of the results given are asymptotic, and for simplicity we only study the two-sample case.

Consider the aligned linear rank statistic

$$S(\hat{\theta}) = \frac{1}{n_i} \sum_{j=1}^{n_1} \phi(\hat{R}_{1j}/(N+1)) \ , \tag{3.1}$$

where $\hat{R}_{1j}$ is either

1) the rank of $X_{1j} - \hat{\mu}_1$ among

$$X_{11} - \hat{\mu}_1, \ldots, X_{1n_1} - \hat{\mu}_1, \quad X_{21} - \hat{\mu}_2, \ldots, X_{2n_2} - \hat{\mu}_2$$

or

2) the rank of $(X_{1j} - \hat{\mu}_1)/\hat{\sigma}_1$ among

$$(X_{11} - \hat{\mu}_1)/\hat{\sigma}_1, \ldots, (X_{1n_1} - \hat{\mu}_1)/\hat{\sigma}_1, \quad (X_{21} - \hat{\mu}_2)/\hat{\sigma}_2, \ldots, (X_{2n_2} - \hat{\mu}_2)/\hat{\sigma}_2.$$

The score functions ϕ_p of interest are the asymptotic versions of (2.2), i.e., the Legendre polynomials $\phi_2(t) = (180)^{1/2}[(t - \frac{1}{2})^2 - 1/12]$, $\phi_3(t) = (7)^{1/2}[20(t - \frac{1}{2})^3 - 3(t - \frac{1}{2})]$, and $\phi_4(t) = 210(t - \frac{1}{2})^4 - 45(t - \frac{1}{2})^2 + 9/8$. In case 1) Jureckova (1979) has given conditions for $(Nn_1/n_2)^{1/2}[S(\hat{\theta}) - \mu(\theta_0)]$ to be asymptotically normal with mean 0 and variance $1 - 2I_1I_2 + A^2I_2^2$, where $I_1 = \int_0^1 \phi(u)\ell(F^{-1}(u))du$, $I_2 = \int_0^1 \phi(u)\phi(u,f)du$, $A^2 = \int_{-\infty}^{\infty} \ell^2(x)dF(x) - (\int_{-\infty}^{\infty} \ell(x)dF(x))^2$, the first sample has density $f(x - \mu_1)$, the second sample has density $f(x - \mu_2)$, $\ell(x)$ is the influence curve of μ_1 and μ_2 at $F(x) = \int_{-\infty}^{x} f(y)dy$, and $\phi(u,f) = -f'(F^{-1}(u))/f(F^{-1}(u))$.

Table 3.3 gives the asymptotic standard deviations for ϕ_2, ϕ_3, and ϕ_4

---Insert Table 3.3 here ---

when means and medians are respectively substracted from the observations. The distributions listed are the normal, the t distribution with 5 degrees of freedom (t_5), the Laplace with density $f(x) = \frac{1}{2}\exp(-|x|)$, the extreme value with distribution function $f(x) = \exp(-\exp(-x))$, and the lognormal whose log transform has variance $= (.5)^2$. When the distributions are symmetric, the results for ϕ_2 and ϕ_4 are the same as for the nonaligned statistics. In these cases only the Laplace at ϕ_3 in Table 3.2 is disturbing. The results for the two skewed distributions are more disturbing, especially when testing for scale. However, the lognormal with parameter $\sigma = .5$ has skewness coefficient $\sqrt{\beta_1} = 1.75$ and the adverse effects seem to be monotone in $\sqrt{\beta_1}$. Thus, if heavy skewness is present in both samples, it is advisable to transform the data before aligning. (Of course, monotone transformations do not affect the nonaligned statistics.) Another method of aligning is to adjust $\hat{\mu}_1$ (or $\hat{\mu}_2$) until the Wilcoxon statistic is zero. This is the same as subtracting the Hodges-Lehmann shift estimator median $\{X_{1i} - X_{2j},\ 1 \leq i \leq n_1,\ 1 \leq j \leq n_2\}$ from the sample 1 observations. In this case $\ell(x)$ is proportional to $F(x) - \frac{1}{2}$ and $I_1 \equiv 0$ for any Legendre score function, and the asymptotic variance is just $1 + I_2^2/12(\int f^2(y)dy)^2$. In Table 3.3 the asymptotic standard deviations are under the column headed by "H-L", and we see that this approach gives the best results. The k-sample extension is to use the residuals from R-estimation (e.g., McKean and Hettmansperger, 1978) with Wilcoxon (ϕ_1) scores. If the distributions are symmetric, a much easier computational method which is asymptotically equivalent is to subtract off the one-sample Hodges-Lehmann estimators given by median $\{(X_{ij} + X_{ik})/2,\ 1 \leq j \leq k \leq n_i\}$.

For the case 2) adjustment of both location and scale there are no results available at present. However, the approach of Randles (1982) may be used along with Chernoff and Savage (1958) results to yield a fairly general theorem for rank statistics with estimated parameters. The case 2) results will then follow in a corollary.

Leg $g_1(x,\theta)$ and $g_2(x,\theta)$ be transformations such that $g_1(X_{11},\theta)$ and $g_2(X_{21},\theta)$ have distribution functions $F_1(x,\theta)$ and $F_2(x,\theta)$. We are interested in linear rank statistics of the form (3.1) where $\hat{R}_{ij}$ is the rank of $g_1(X_{1j},\hat{\theta})$ among $g_1(X_{11},\hat{\theta}),\ldots,g_1(X_{1n_1},\hat{\theta})$, $g_2(X_{21},\hat{\theta}),\ldots,g_2(X_{2n_2},\hat{\theta})$. The Chernoff-Savage representation of the asymptotic mean of $S(\theta)$ is

$$\mu_N(\theta) = \int_{-\infty}^{\infty} \phi(H(x,\theta))dF_1(x,\theta) \ ,$$

where $H(x,\theta) = \lambda_N F_1(x,\theta) + (1 - \lambda_N)F_2(x,\theta)$ with $\lambda_N = n_1/N$. We write $\mu(\theta)$ when λ_N is replaced by its limit λ. Following the approach in Section 3 of Randles (1982), write $S(\hat{\theta}) - \mu_N(\theta_0)$ as

$$S(\hat{\theta}) - \mu_N(\hat{\theta}) - [B_{1N}(\hat{\theta}) + B_{2N}(\hat{\theta})] \tag{3.2a}$$

$$+ B_{1N}(\hat{\theta}) + B_{2N}(\hat{\theta}) - [B_{1N}(\theta_0) + B_{2N}(\theta_0)] \tag{3.2b}$$

$$+ B_{1N}(\theta_0) + B_{2n}(\theta_0) + \mu_N(\hat{\theta}) - \mu_N(\theta_0) \quad , \tag{3.2c}$$

where from Chernoff and Savage (1958, p. 976) we have

$$B_{1N}(\theta) + B_{2N}(\theta) = (1 - \lambda_N)\ \{\frac{1}{n_1} \sum_{i=1}^{n_1} [B(X_{1i},\theta) - EB(X_{1i},\theta)]$$

$$- \frac{1}{n_2} \sum_{j=1}^{n_2} [B^*(X_{2i},\theta) - EB^*(X_{2i},\theta)]\}$$

with

$$B(x,\theta) = \int_{x_0}^{x} \phi'(H(x,\theta))dF_2(x,\theta)$$

$$B^*(x,\theta) = \int_{x_0}^{x} \phi'(H(x,\theta))dF_1(x,\theta).$$

Randles method is to show that (3.2a) and (3.2b) are $o_p(N^{-1/2})$ and that (3.2c) is asymptotically normal. Here we are able to use uniform convergence results of Chernoff and Savage in place of the uniform differential method that Randles employs. Let θ have dimension p.

THEOREM 3.1. Let $X_{11},\ldots,X_{1n_1}$ and $X_{21},\ldots,X_{2n2}$ be independent samples such that $g_1(X_{1i},\theta)$ and $g_2(X_{1j},\theta)$ have continuous distribution functions $F_1(x,\theta)$ and $F_2(x,\theta)$ for all θ in a neighborhood N_0 of θ_0. Let ϕ be a score function such that $J_N(i/N) \equiv \phi(i/(N+1))$ satisfies conditions 1-4 of Theorem 1 of Chernoff and Savage (1958). Let $S(\theta) = n_1^{-1}\Sigma\phi(R_{1j}/(N+1)$ where R_{1j} is the rank of $g_1(X_{1j},\theta)$ among $g_1(X_{11},\theta),\ldots,g_1(X_{1n_1},\theta)$, $g_2(X_{21},\theta),\ldots,g_2(X_{2n_2},\theta)$. Assume that n_1,n_2, and $N \to \infty$ with $\lambda_N = n_1/N \to \lambda$, $0 < \lambda < 1$. Then

1) (Chernoff-Savage uniform approximation) for any $\varepsilon > 0$

$$P\{\sup_{\theta\varepsilon N_0} |S(\theta) - \mu_N(\theta) - [B_{1n}(\theta) + B_{2N}(\theta)]| > \frac{\varepsilon}{N^{1/2}}\} \to 0 .$$

2) if $N^{1/2}(\hat{\theta} - \theta_0) = O_p(1)$, then 1) implies that (3.2a) $= o_p(N^{-1/2})$.

3) if $B(x,\theta)$ and $B^*(x,\theta)$ satisfy Condition 2.3 of Randles (1982) uniformly for λ_N in a neighborhood of λ and $N^{1/2}(\hat{\theta} - \theta_0) = O_p(1)$, then (3.2b) $= o_p(N^{-1/2})$.

4) if in addition to the conditions of 3) above we assume that $\mu_N(\theta)$ has a gradient $\Delta\mu_N(\theta)$ which converges uniformly in a neighborhood θ_0 to the nonzero gradient $\Delta\mu(\theta_0)$ and

$$(Nn_1/n_2)^{1/2}(B_{1N}(\theta_0) + B_{2N}(\theta_0), \hat{\theta} - \theta) \xrightarrow{d} N_{p+1}(0,\Sigma) ,$$

then

$$(Nn_1/n_2)^{1/2}(S(\hat{\theta}) - \mu_N(\theta_0)) \xrightarrow{d} N(0,D'\Sigma D)$$

provided $D'\Sigma D > 0$, where $D' = (1,\Delta\mu(\theta_0))$.

REMARKS. If $F_1(x,\theta_0) = F_2(x,\theta_0)$, then $\mu_N(\theta_0) = \mu(\theta_0) = \int_0^1 \phi(t)dt$. Conclusion 1) is directly from the Chernoff-Savage proof. 2) follows the argument of Lemma 3.6 of Randles (1982). 3) is from Theorem 2.8 of Randles (1982) since $B_{1n} + B_{2N}$ is a sum of two U-statistics, with the slight complication that $B(x,\theta)$ and $B^*(x,\theta)$ depend on λ_N. 4) is similar to Lemma 3.7B of Randles (1982) with again the complication that $\mu_N(\theta)$ depends on λ_N.

Theorem 3.1 is of course more general than the work of Jureckova (1979) and Fligner and Hettmansperger (1979) since they only consider location alignment. But even in this case Theorem 3.1 provides new results. For example, if ϕ has two derivatives which are bounded and F has a bounded density f, then Theorem 3.1 holds whereas Jureckova requires f to have finite Fisher information.

Now we apply Theorem 3.1 for the case 2) situation where $\theta = (\mu_1,\mu_2,\sigma_1,\sigma_2)$ and $g_1(x,\theta) = (x-\mu_1)/\sigma_1$ and $g_2(x,\theta) = (x-\mu_2)/\sigma_2$. Note that the corollary applies to Legendre score functions since they have bounded derivatives of all orders.

COROLLARY. Let $X_{11},\ldots,X_{1n_1}$ and $X_{21},\ldots,X_{2n_2}$ be independent samples with distribution functions $F_1(x) = F((x-\mu_{10})/\sigma_{10})$ and $F_2(x) = F((x-\mu_{20})/\sigma_{20})$ where F has a bounded density f with finite first moment. Let $S(\hat{\theta})$ be given for the case 2) location-scale adjustment by (3.1) where ϕ is assumed to have two bounded derivatives and standardized so that $\int_0^1 \phi(t)dt = 0$ and $\int_0^1 \phi^2(t)dt = 1$. Suppose that

$$\hat{\mu}_i - \mu_{i0} = \frac{\sigma_{i0}}{n_i}\sum_{j=1}^{n_i} \ell_1\left(\frac{X_{ij}-\mu_{i0}}{\sigma_{i0}}\right) + o_p(n_i^{-1/2}) \ , \quad i = 1,2 \ ,$$

and

$$\hat{\sigma}_i - \sigma_{i0} = \frac{\sigma_{i0}}{n_i}\sum_{j=1}^{n_i} \ell_2\left(\frac{X_{ij}-\mu_{i0}}{\sigma_{i0}}\right) + o_p(n_i^{-1/2}), \quad i = 1,2 \ ,$$

where $\int \ell_j(x)dF(x) = 0$ and $0 < \int \ell_j^2(x)dF(x) < \infty$. If n_1,n_2 and $N\to\infty$ such that $n_1/N \to \lambda$, $0 < \lambda < 1$, then

$$(Nn_1/n_2)^{1/2}S(\hat{\theta}) \xrightarrow{d} N(0,D'\Sigma D) \ ,$$

where

$$D'\Sigma D = 1 - 2I_1I_2 + A_1^2I_2^2 - 2I_{11}I_{22} + 2A_{12}I_2I_{22} + A_2^2I_{22}^2 \ ,$$

with

$$I_1 = \int_0^1 \phi(t)\ell_1(F^{-1}(t))dt \qquad I_{11} = \int_0^1 \phi(t)\ell_2(F^{-1}(t))dt$$

$$I_2 = \int_0^1 \phi'(t)f(F^{-1}(t))dt \qquad I_{22} = \int_0^1 \phi'(t)f(F^{-1}(t))F^{-1}(t)dt$$

$$A_1^2 = \int_{-\infty}^{\infty} \ell_1^2(x)dF(x) \qquad A_2^2 = \int_{-\infty}^{\infty} \ell_2^2(x)dF(x)$$

$$A_{12} = \int_{-\infty}^{\infty} \ell_1(x)\ell_2(x)dF(x) \ .$$

Table 3.4 gives asymptotic standard deviations when sample means and standard deviations are used as estimators. The ϕ_3 results are the same for symmetric distributions when aligning for only location. Also included in

---Insert Table 3.4 Here---

Table 3.4 are some Monte Carlo estimates based on 5000 independent replications using sample 10% trimmed means and standard deviations as well as the usual means and standard deviations. Note that at $n_1 = n_2 = 20$ the kurtosis tests are conservative when using means and standard deviations and liberal (except for the extreme value) when using the trimmed estimators.

To summarize a bit, we note that the nice distribution-free character of the statistics discussed in Subsection 3A is not preserved after alignment. If the underlying distribution is symmetric then the scale and kurtosis tests are asymptotically distribution-free after subtracting off location estimators. However, even among symmetric distributions the shape of the distribution can have a large effect on the skewness test. The Laplace standard deviation of 1.24 in Table 3.3 is one example. For the same situation the uniform distribution is worse (s.d. = 1.83) but the triangular (sum of two uniforms) is much better (s.d. = 1.07), so that it is hard to anticipate how liberal the skewness test using standard normal critical values will be. In general, one should be cautious in interpreting the p-values for the skewness test after aligning for location and also for the kurtosis test after aligning for both location and scale.

4. PITMAN EFFICIENCY

Pitman asymptotic relative efficiencies (ARE's) are straightforward to calculate for linear rank statistics and help answer two questions about the components T_{ip}:

1) With what kinds of alternatives are the T_{ip} associated?
2) For small departures from the null value, how does each T_{ip} compare with the best possible test?

The results are of course asymptotic but experience has shown that ARE's are reliable indicators of what to expect in finite samples. For two-sample linear rank statistics the Pitman ARE is $[\int_0^1 \phi(u)\phi_{opt}(u)du]^2$, where $\phi(u)$ is the underlying score function standardized so that $\int_0^1 \phi(u)du = 0$ and $\int_0^1 \phi^2(u)du = 1$ and $\phi_{opt}(u)$ is the standardized score function which maximizes the efficacy under the particular type of alternative at hand (e.g., see Sect. 9.2 and 9.3 of Randles and Wolfe, 1979).

Table 4.1 lists the Pitman ARE in the two-sample case for a variety of location, scale, skewness, and kurtosis type alternatives. Numerical integration was used for some of the calculations. Location alternatives $F(x)$ versus $F(x-\theta)$ and scale alternatives $F(x)$ versus $F(x/(1+\theta))$ are well known and form the first part of the table. Here the base

--- INSERT TABLE 4.1 HERE ---

distribution functions $F(x)$ considered were the standard normal, the t distribution with 5 degrees of freedom (t_5), the Laplace with density $f(x) = \frac{1}{2}\exp(-|x|)$, and the extreme value distribution with $F(x) = \exp(-\exp(-x))$. The Wilcoxon (ϕ_1) and Mood (ϕ_2) ARE's at the normal and Laplace are well known. The t_5 results are surprisingly good. The *low* values of ϕ_3 for location and ϕ_4 for scale are *desirable*.

Durbin, Knott, and Taylor (1975) suggest two types of skewness alternatives. The first one is based on Edgeworth expansions of densities using Hermite polynomials and has density

$$f_\theta(x) = [1 + \theta(3x-x^3)](2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2} .$$

The second is based on Fourier expansions and has density

$$f_\theta(x) = [1 + \theta \sin(3\pi F(x))]f(x) ,$$

where $F(x)$ is an arbitrary distribution function with density $f(x)$. Using the same idea, a third type based on Legendre polynomials was added,

$$f_\theta(x) = [1 + \theta(20(F(x)-1/2)^3 - 3(F(x)-1/2))]f(x) .$$

Although the first density is connected to the normal distribution, the latter two are nonparametric so that the results do not depend on the form of $F(x)$. For kurtosis alternatives Durbin, Knott, and Taylor (1975) suggest

$$f_\theta(x) = [1 + \theta(x^4-6x^2+3)](2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}$$

and

$$f_\theta(x) = [1 + \theta \sin(4\pi F(x))]f(x) .$$

The Legendre type is

$$f_\theta(x) = [1 + \theta(210(F(x) - 1/2)^4 - 45(F(x) - 1/2)^2 + 9/8)]f(x) .$$

In Table 4.1 these alternatives are listed by the type of expansions used - Hermite, Fourier, or Legendre. The optimal score functions are those associated with a_{Ni3} and a_{Ni4} of (2.4) for the Hermite, of (2.3) for the Fourier, and of (2.2) for the Legendre. ϕ_3 and ϕ_4 are not very effective for the Hermite alternatives. As mentioned by Durbin, Knott, and Taylor (1975), these alternatives are "heavily dominated by behavior in the tails." ϕ_3 and ϕ_4 are naturally optimal at the Legendre and reasonably efficient at the Fourier alternatives. The ϕ_1 and ϕ_2 values are all appropriately low in the second half of Table 4.1.

5. MONTE CARLO POWER ESTIMATES

Finite sample power properties of the proposed statistics were studied in a limited Monte Carlo experiment. For the two-sample case ($n_1 = 20$, $n_2 = 20$) 15 different population pairs were considered and are listed in Table 5.1. These are basically a variety of location and scale

--- INSERT TABLE 5.1 HERE ---

shifts involving the normal distribution, the t distribution with 5 degrees of freedom, and the extreme value distribution $F(x) = \exp(-\exp(-x))$. Each shift is measured in mean and variance units. For example EV(.5, 2.25) stands for an extreme value distribution with mean = .5 and variance = 2.25. In each situation 5000 independent Monte Carlo sets of independent samples were generated, and the proportion of test rejections is reported. Three alternative omnibus tests were included for comparison with T_{GLOBE}. The first is the "continuous" version of the Anderson-Darling statistic mentioned in Section 2

$$AD = -(N/n_1n_2)\{n_1^2 - 2n_1^2 \log(N+1)$$

$$+ \sum_{j=1}^{n_1} (2j-1)[\log(R_{1(j)}) + \log(N+1-R_{1(n_1+1-j)})]\} ,$$

where $R_{1(1)},\ldots,R_{1(n_1)}$ are the ranks in the first sample ordered from smallest to largest. AD has the same asymptotic distribution as the one-sample Anderson-Darling test for uniformity (see Pettitt, 1976 and 1979). Another competitor is the "continuous" Cramér-von Mises statistic

$$CVM = \sum_{j=1}^{n_1} [\frac{R_{1(j)}}{N+1} - \frac{(j-\frac{1}{2})}{n_1}]^2 + \frac{1}{12n_1}$$

which has the same asymptotic distribution as the usual (unweighted) Cramér-von Mises test for uniformity. The third test is the two-sample Kolmogorov-Smirnov test

$$KS = \sup_x |F_{n_1}(x) - F_{n_2}(x)| ,$$

where F_{n_1} and F_{n_2} are the empirical distribution functions of the two samples. Asymptotic critical values were used in Table 5.1 for $T_{WILC} - T_{KURT}(\chi_1^2)$ and for AD and CVM. The percentiles of T_{GLOBE} were estimated by Monte Carlo although the simple adjustment suggested in Section 3 works well, and the KS test was randomized to get an exact $\alpha = .10$ test using the tables of Kim and Jennrich (1970). A check on these critical values using standard normal samples gave empirical powers of (.105, .102, .104, .098) for $T_{WILC} - T_{KURT}$ and (.102, .094, .105, .103) for T_{GLOBE}, AD, CVM, and KS with standard deviation $[(.9)(.1)/5000]^{1/2} = .004$.

REMARKS ON TABLE 5.1

1) The individual components appear to detecting the appropriate alternatives except for T_{GLOBE}. Apparently the kurtosis difference which T_4 measures is not big enough for these pairs of distributions.

2) Among the omnibus tests, T_{GLOBE} is best for the pure scale situation, for the "different types-equal location situation," and for the EV(0,1) versus $t_5(.5, 2.25)$. The other omnibus tests dominate T_{GLOBE} for pure location shift, for the t_5 location-scale shift, and for the N(0,1) versus $t_5(.5, 2.25)$. The tests are approximately comparable in the other situations.

3) 10% trimmed means and trimmed standard deviations were used to align the samples. The location alignment is helpful in detecting scale differences in the location-shift models and in detecting skewness differences in the "different types - location-shift situation." The tendency for aligned tests to be too liberal (true $\alpha > .10$) shows up most clearly in the last two columns of Table 5.1 for the extreme value distribution.

Another design studied is the four-sample problem with $n_1 = 10$, $n_2 = 30$, $n_3 = 10$, and $n_4 = 30$. The k-sample versions are computed for all but the KS by adding the individual two-sample statistics weighted by $(N-n_i)/N$ (see also Kiefer, 1959). Asymptotic distributions are equal to the k-1 convolution of distributions obtained in the two-sample case. For CVM and KS the critical values were obtained from Kiefer (1959). For $T_{WILC} - T_{KURT}$ the χ_3^2 was used, and the critical values for T_{GLOBE} and AD were obtained by Monte Carlo. A check with all samples from the standard normal gave empirical power of (.096, .089, .105, .103) for $T_{WILC} - T_{KURT}$ and (.107, .101, .095, .099) for T_{GLOBE}, AD, CVM, and KS. Table 5.2 lists the empirical power for 5 situations involving the normal and extreme value distribution. T_{GLOBE} dominates the other omnibus statistics for all but the last siutation. Alignment is once again useful for detecting scale and skewness differences.

--- INSERT TABLE 5.2 HERE ---

6. EXAMPLES

A) Barnett and Eisen (1982) compare rainfall at New York's Central Park for February and August of 1950-1979. In this two-sample problem we have $n_1 = n_2 = 30$ (see their Table 3 for the actual data). They note that a variety of nonparametric tests are not significant at the $\alpha = .10$ level but that their D statistic has a p-value of .04. A stem and leaf plot is quite useful and reveals that August has two large values 9.37 and 13.82 which dominate the comparison of the usual sample statistics $(\bar{X}, s, \sqrt{b_1}, b_2) = (3.34, 1.28, .14, 2.02)$ for February and (4.37, 2.64, 1.63, 6.79) for August. The initial linear rank analysis is resistant to outliers and yields $(T_{11}, T_{12}, T_{13}, T_{14}, T_{GLOBE}) =$ (-1.60, -1.35, -1.75, -1.15, 8.75) using midranks to break ties. The summary statistic 8.75 is adjusted by its null standard deviation 2.65 to get 9.07 and from a χ^2_4 distribution has p-value ≈ .06. Since the standardized Wilcoxon statistic is fairly large, we must first align for location before interpreting the other components. After subtracting off the 10% symmetrically trimmed means 3.31 and 4.07, we get $(T_{11}, T_{12}, T_{13}, T_{14}) = (.22, -2.44, -1.24, .58)$ which gives stronger evidence for a scale difference. The data was then divided by 10% symmetrically trimmed standard deviations .96 and 1.42 to get $(T_{11}, T_{12}, T_{13}, T_{14}) = (-.04, .06, -.93, -.11)$. This last result suggests that the population differences are mainly in location and scale.

B) Figure 6.1 is a stem and leaf plot of data given by Oskamp (1962) and used by Lehmann (1975, p. 254) to illustrate k-sample rank procedures. The numbers are actually percentages of correct predictions of patient

---INSERT FIGURE 6.1 HERE---

disorders by three different groups - staff and trainees at VA hospitals and undergraduate psychology majors. The initial linear rank analysis is given in Table 6.1. Clearly, there is a location difference

--- INSERT TABLE 6.1 HERE ---

(T_{WILC} = 11.08, $p < .005$), staff are different from the combination of trainees and undergraduates (T_{COL1} = 16.88, $p < .005$), and undergraduates appear to be different from the combination of staff and trainees (T_{COL3} = 8.09, $p \approx .07$ from adjusted χ_4^2). However, a two-sample comparison (not shown) of trainees and undergraduates is not close to significance ($p > .5$). After subtracting off (approximately) 10% symmetrically trimmed means 73.1, 70.9, and 70.2, respectively, the scale row is (-.72, -.23, .88, T_{MOOD} = .87), the skewness row is (-.43, .59, -.18, T_{SKEW} = .39), and the kurtosis row is (-2.27, .16, 1.93, T_{KURT} = 5.93). Thus, the analysis suggests no scale or skewness differences but rather kurtosis differences (T_{KURT} = 5.93, $p \approx .05$).

C) The data in Table 6.4 is the ratio (multiplied by 100) of assessed value to sale price or residential property in Fitchburg, Massachusetts, in 1979. This data was provided by A. R. Manson and consists of four groups: single family dwellings, n_1 = 219, two family dwellings, n_2 = 87, three family dwellings, n_3 = 62, and four or more family dwellings, n_4 = 28. These are not true random samples but are actually all the residential property which sold as "arm's-length" transactions during 1979 in Fitchburg. However, we shall treat them as iid random samples from infinite populations for ease of inference and illustration. Note that there are a number of large observations which force the use of robust procedures or at least some ad hoc data screening. Table 6.2

--- INSERT TABLE 6.2 HERE ---

gives the initial linear rank analysis. The first column shows that there is a clear difference between single family dwellings and the combined population of two or more family dwellings. After substracting medians, the single family column becomes $(-.83, -7.47, -1.32, -.67)^T$, and after further dividing by interquartile ranges, the first column is $(-1.34, -.16, -3.26, -1.45)^T$. Thus, the difference between the first group and the other three consists mainly of location, scale, and to a lesser degree skewness differences. When single family dwellings are deleted, the rank analysis is given by Table 6.3. There appears to be

---INSERT TABLE 6.3 HERE---

skewness differences (11.25, $p < .005$) and the individual skewness statistics (3.26, -1.79, -2.12) suggest a decreasing trend in right-tailed skewness. A final analysis of groups three and four by themselves revealed no significant differences.

Table 1.1. Linear Rank Statistics

Alternative	Underlying Score Function	Group 1	Group 2	...	Group k	k-Sample Tests
Location	Wilcoxon = $u - 1/2$	T_{11}	T_{21}	...	T_{k1}	$T_{WILC} = \sum_1^k (\frac{N-n_i}{N}) T_{i1}^2$
Scale	Mood = $(u - 1/2)^2 - 1/12$	T_{12}	T_{22}	...	T_{k2}	$T_{MOOD} = \sum_1^k (\frac{N-n_i}{N}) T_{i2}^2$
Skewness	$20(u - 1/2)^3 - 3(u - 1/2)$	T_{13}	T_{23}	...	T_{k3}	$T_{SKEW} = \sum_1^k (\frac{N-n_i}{N}) T_{i3}^2$
Kurtosis	$210(u - 1/2)^4 - 45(u - 1/2)^2 + 9/8$	T_{14}	T_{24}	...	T_{k4}	$T_{KURT} = \sum_1^k (\frac{N-n_i}{N}) T_{i4}^2$
		$T_{COL1} = \sum_{p=1}^{4} T_{1p}^2$	$T_{COL2} = \sum_{p=1}^{4} T_{2p}^2$	...	$T_{COLk} = \sum_{p=1}^{4} T_{kp}^2$	$T_{GLOBE} = \sum_1^k (\frac{N-n_i}{N}) T_{COLi}$

Table 3.1. Coefficients of Skewness and Kurtosis for the Null Distribution of Linear Rank Statistics Having Scores (2.2)

		T_{i1} (Wilcoxon)		T_{i2} (Mood)		T_{i3}		T_{i4}	
N	n_i	$\sqrt{\beta_1}$	β_2	$\sqrt{\beta_1}$	β_2	$\sqrt{\beta_1}$	β_2	$\sqrt{\beta_1}$	β_2
12	3	0	2.56	.22	2.55	0	2.56	.01	2.57
12	6	0	2.69	0	2.65	0	2.70	0	2.78
20	5	0	2.74	.17	2.73	0	2.73	.09	2.73
20	10	0	2.82	0	2.79	0	2.78	0	2.81
40	10	0	2.87	.12	2.86	0	2.86	.08	2.86
40	20	0	2.91	0	2.89	0	2.88	0	2.88
80	20	0	2.93	.08	2.93	0	2.93	.06	2.93
80	40	0	2.95	0	2.95	0	2.94	0	2.94

NOTE: $\sqrt{\beta_1} = E(Y-\mu)^3/\sigma^3$, $\beta_2 = E(Y-\mu)^4/\sigma^4$

Table 3.2. Estimated P(Type I error) of Summary Statistics Using Chi-Squared Critical values at α = .10.

n_1	n_2	n_3	n_4	T_{WILC}	T_{MOOD}	T_{SKEW}	T_{KURT}	T_{GLOBE}	T_{ADJ}
5	5	5	5	.091	.088	.089	.094	.053	.084
10	10	10	10	.093	.097	.095	.092	.081	.094
5	15	5	15	.092	.093	.096	.099	.082	.092
20	20	20	20	.091	.085	.095	.097	.082	.088
10	30	10	30	.093	.102	.097	.104	.095	.100

Note: Entries have standard deviation $\approx[(.1)(.9)/5000]^{1/2} = .004$.

Table 3.3. Asymptotic Standard Deviations of Linear Rank Statistics after Adjusting for Location

Score Function	ϕ_2 (Scale)			ϕ_3 (Skewness)			ϕ_4 (Kurtosis)		
Family / Estimator	Mean	Median	H-L	Mean	Median	H-L	Mean	Median	H-L
Normal	1.0	1.0	1.0	.98	1.10	1.02	1.0	1.0	1.0
t_5	1.0	1.0	1.0	1.02	.94	1.00	1.0	1.0	1.0
Laplace	1.0	1.0	1.0	1.24	.94	1.07	1.0	1.0	1.0
Extreme Value	1.19	1.14	1.09	.98	1.15	1.03	1.04	1.02	1.01
Lognormal (σ=.5)	1.44	1.29	1.18	1.01	1.27	1.08	1.12	1.08	1.05

Note: "H-L" refers to the two-sample Hodges-Lehmann shift estimator median $\{X_{1i} - X_{2j},\ 1 \le i \le n_1,\ 1 \le j \le n_2\}$.

Table 3.4. Standard Deviations of Linear Rank Statistics after Adjusting for Location + Scale

Score Function	ϕ_3 (Skewness)			ϕ_4 (Kurtosis)		
Estimators	Mean and S.D.		Trimmed (α=.1) Mean and S.D.	Mean and S.D.		Trimmed (α=.1) Mean and S.D.
Family	$n_1=\infty$ $n_2=\infty$	$n_1=20$ $n_2=20$	$n_1=20$ $n_2=20$	$n_1=\infty$ $n_2=\infty$	$n_1=20$ $n_2=20$	$n_1=20$ $n_2=20$
Normal	.98	1.04	1.05	.93	.97	1.10
t_5	1.02	1.09	1.04	.94	.96	1.05
Laplace	1.24	1.20	1.11	.91	.96	1.04
Extreme Value	1.10	1.11	1.10	.94	.95	.91
Lognormal (σ=.5)	1.31	1.19	1.17	1.08	.94	1.10

Note: At $n_1 = 20$, $n_2 = 20$ entries are Monte Carlo estimates with standard deviation $\approx .01$.

Table 4.1. Pitman Efficiencies of the Components

Family	Location Alternative				Scale Alternative			
	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_1	ϕ_2	ϕ_3	ϕ_4
Normal	.955	0	.03	0	0	.76	0	.14
t_5	.99	0	.00	0	0	.93	0	.05
Laplace	.75	0	.11	0	0	.87	0	.06
Extreme Value	.75	.14	.05	.02	.03	.67	.03	.13

Type	Skewness Alternative				Kurtosis Alternative			
	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_1	ϕ_2	ϕ_3	ϕ_4
Hermite	.03	0	.42	0	0	.14	0	.09
Fourier	.01	0	.80	0	0	.08	0	.62
Legendre	0	0	1.00	0	0	0	0	1.00

Table 5.1. Empirical Power in Two-Sample Case, $n_1 = 20$, $n_2 = 20$.

Populations		Components				Omnibus Tests				Aligned for Location			Aligned for Location & Scale	
1	2	T_{WICL}	T_{MOOD}	T_{SKEW}	T_{KURT}	T_{GLOBE}	AD	CVM	KS	T_{MOOD}	T_{SKEW}	T_{KURT}	T_{SKEW}	T_{KURT}
Location Shift														
N(0,1)	N(.5,1)	.45	.08	.09	.08	.27	.41	.43	.37	.11	.11	.10	.11	.14
t_5(0,1)	t_5(.5,1)	.53	.08	.09	.08	.33	.49	.52	.46	.11	.11	.11	.11	.11
EV(0,1	EV(.5,1)	.52	.15	.08	.07	.36	.50	.51	.46	.17	.11	.11	.13	.14
Scale Shift														
N(0,1)	N(0,2.25)	.11	.42	.08	.11	.28	.17	.16	.16	.44	.08	.12	.12	.14
t_5(0,1)	t_5(0,2.25)	.11	.36	.09	.09	.23	.16	.15	.16	.39	.09	.10	.12	.11
EV(0,1)	EV(0,2.25)	.12	.46	.08	.11	.31	.19	.18	.18	.47	.09	.11	.13	.14
Location-Scale Shift														
N(0,1)	N(.5,2.25)	.32	.38	.07	.10	.37	.37	.36	.34	.44	.08	.12	.11	.14
t_5(0,1)	t_5(.5,2.25)	.39	.32	.09	.08	.39	.42	.44	.41	.39	.09	.10	.12	.11
EV(0,1)	EV(.5,2.25)	.28	.25	.09	.10	.27	.29	.31	.29	.45	.09	.11	.14	.13
Different types - Equal Location and Scale														
N(0,1)	EV(0,1)	.12	.12	.22	.09	.17	.12	.13	.13	.15	.23	.09	.23	.13
N(0,1)	t_5(0,1)	.11	.14	.10	.11	.13	.10	.11	.11	.15	.11	.11	.11	.14
EV(0,1)	t_5(0,1)	.12	.12	.19	.10	.17	.11	.13	.13	.13	.20	.10	.22	.12
Different Types - Location-Scale Shift														
N(0,1)	EV(.5,2.25)	.22	.21	.17	.08	.23	.23	.22	.20	.32	.17	.10	.23	.12
N(0,1)	t_5(.5,2.25)	.37	.22	.08	.11	.31	.38	.39	.36	.26	.09	.14	.12	.14
EV(0,1)	t_5(.5,2.25)	.22	.34	.13	.09	.30	.26	.25	.25	.46	.15	.08	.23	.13

Note: Entries are the proportion of test rejections in 5000 Monte Carlo replications. Standard deviation $\leq (4 \cdot 5000)^{-1/2} = .007$. N(a,b), t_5(a,b), and EV(a,b) stand for normal, t_5, and extreme value distributions with mean a and variance b.

Table 5.2. Empirical Power in Four-Sample Case, $n_1 = 10$, $n_2 = 30$, $n_3 = 10$, $n_4 = 30$. Nominal Level $\alpha = .10$

Populations				Components				Omnibus Tests				Aligned for Location			Aligned for Location & Scale	
1	2	3	4	T_{WILC}	T_{MOOD}	T_{SKEW}	T_{KURT}	T_{GLOBE}	AD	CVM	KS	T_{MOOD}	T_{SKEW}	T_{KURT}	T_{SKEW}	T_{KURT}
N(0,1)	N(.5,1)	N(0,2.25)	N(.5,2.25)	.32	.50	.10	.12	.49	.45	.39	.33	.54	.07	.13	.10	.14
EV(0,1)	EV(.5,1)	EV(0,2.25)	EV(.5,2.25)	.41	.59	.14	.14	.61	.57	.49	.44	.59	.08	.13	.13	.13
N(0,1)	N(0,1)	EV(0,1)	EV(0,1)	.11	.12	.23	.10	.19	.13	.12	.13	.17	.22	.08	.23	.11
N(0,1)	N(0,1)	EV(.5,2.25)	EV(.5,2.25)	.24	.23	.19	.09	.31	.30	.25	.22	.39	.15	.11	.24	.11
N(0,1)	N(.5,2.25)	EV(0,1)	EV(.5,2.25)	.25	.24	.19	.08	.30	.30	.31	.31	.47	.15	.07	.24	.11

Note: Entries are the proportion of test rejections in 5000 Monte Carlo replications. Standard deviation $\leq (4 \cdot 5000)^{-1/2} = .007$. N(a,b) = normal distribution with mean and variance b. EV(a,b) = extreme value distribution with mean a and variance b.

Figure 6.1. Percentage Accuracy of Predictions from Oskamp (1962)

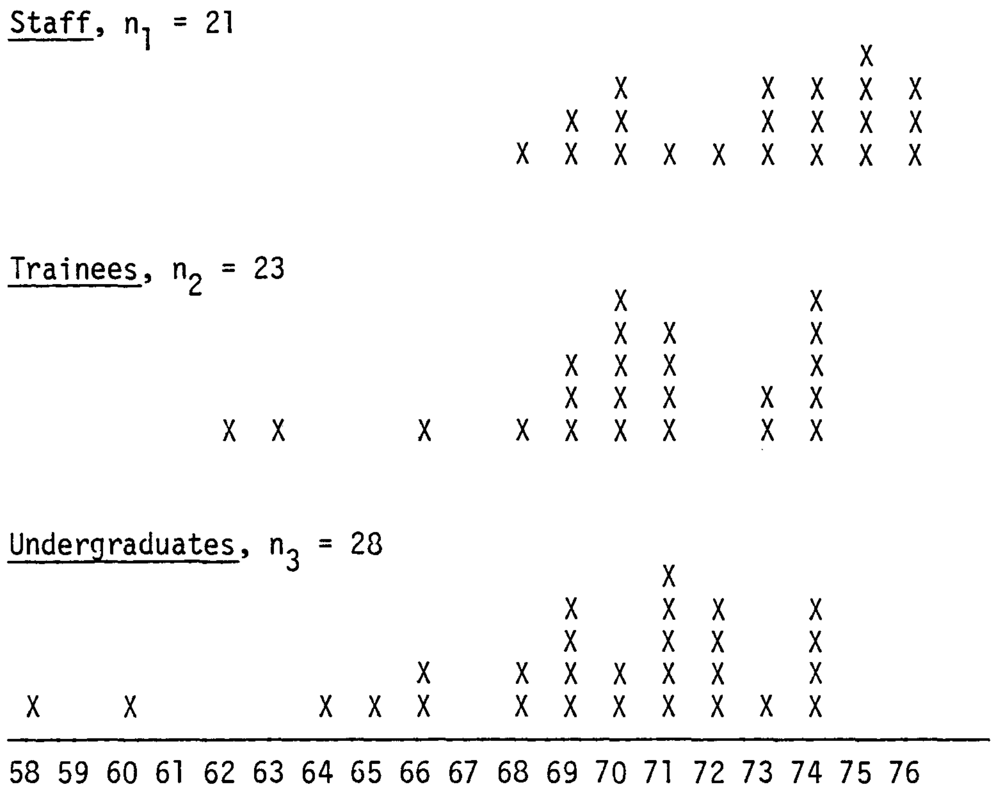

Table 6.1. Linear Rank Analysis of Oskamp Data (no alignment)

	Staff	Trainees	Undergraduates	
Wilcoxon	3.27	-.88	-2.21	11.08
Mood	1.49	-1.09	-.44	2.50
Skewness	1.93	-.13	-1.74	4.53
Kurtosis	.45	-.61	.04	.40
	16.88	2.36	8.09	18.51

Null distributions: Entries ≈ standard normal, column S.S. ≈ χ_4^2, row S.S. ≈ χ_2^2.

Table 6.2. Linear Rank Analysis of Fitchburg Data (no alignment)

	Single Family	Two Family	Three Family	Four or More Family	
Wilcoxon	-6.38	2.24	3.94	3.18	44.57
Mood	-5.31	2.23	2.53	3.11	30.85
Skewness	-.66	2.33	-.98	-1.10	6.37
Kurtosis	-.48	.90	-.60	.33	1.13
	69.55	16.22	23.21	21.09	82.92

Null distributions: Entries ≈ standard normal, column S.S. ≈ χ_4^2, row S.S. ≈ χ_3^2.

Table 6.3. Linear Rank Analysis of Fitchburg Data Without Single Family Dwellings (no alignment)

	Two Family	Three Family	Four or More Family	
Wilcoxon	-1.60	.73	1.25	2.95
Mood	.66	-1.00	.42	1.03
Skewness	3.26	-1.79	-2.12	11.25
Kurtosis	-.72	-.29	1.36	1.87
	14.11	4.84	8.07	17.11

Table 6.4. Ratios of Sale Price to Assessed Value of Residential Property in Fitchburg, Massachusetts, 1979

Group 1 - Single Family Dwellings, $n_1 = 219$

30.90	64.52	68.82	72.03	75.69	78.21	80.77	84.30	87.05	91.98	98.77
41.43	64.76	68.85	72.10	75.92	78.44	81.12	84.43	87.45	92.11	98.91
43.40	65.04	69.13	72.10	75.94	78.48	81.37	84.68	87.73	93.07	99.08
45.85	65.26	69.19	72.24	75.95	78.62	81.48	84.68	87.83	93.12	99.09
49.92	65.41	69.62	72.68	76.56	78.78	81.50	84.84	88.11	93.29	99.32
54.52	65.60	69.69	72.90	76.84	79.50	81.54	84.96	88.57	93.33	100.00
55.17	66.13	69.71	73.56	76.98	79.56	81.86	85.07	88.65	93.51	101.08
57.34	66.29	69.72	73.75	77.23	79.58	81.88	85.07	89.00	94.42	101.64
58.76	66.33	69.73	73.76	77.32	79.63	82.05	85.59	89.26	94.51	102.21
59.53	66.54	69.86	73.84	77.38	79.64	82.32	85.68	89.48	94.57	105.90
59.58	66.57	69.91	73.85	77.43	79.70	82.39	85.71	89.82	95.03	108.09
60.05	66.71	70.23	74.13	77.54	79.85	83.10	86.05	89.83	95.23	110.28
60.16	67.12	70.62	74.31	77.66	79.91	83.58	86.12	90.09	95.36	110.39
60.86	67.44	70.72	74.55	77.69	79.94	83.65	86.39	90.26	95.37	110.67
61.38	67.46	70.79	74.64	77.78	79.95	83.69	86.47	90.54	95.50	116.71
62.50	68.00	70.87	74.95	77.80	80.46	83.78	86.65	90.60	96.24	121.10
62.93	68.08	70.97	75.10	77.92	80.53	83.95	86.70	90.62	96.63	155.83
63.64	68.21	71.46	75.32	78.11	80.53	84.00	86.84	90.62	96.88	158.82
63.74	68.36	71.63	75.33	78.12	80.75	84.02	86.95	90.97	97.31	171.57
63.90	68.60	71.86	75.42	78.14	80.76	84.16	86.96	91.93	98.75	

Table 6.4 continued

Group 2 - Two Family Dwellings, $n_2 = 87$

54.83	67.04	72.56	76.40	79.97	84.36	93.54	96.96	113.38	139.30	184.80
55.59	67.40	73.42	76.57	80.15	84.80	93.73	97.58	114.30	139.79	203.10
56.85	67.52	74.01	76.64	80.42	84.81	94.04	97.67	115.60	140.07	233.17
58.28	69.02	74.72	77.36	81.22	85.68	95.16	99.27	116.36	150.06	258.24
59.03	69.80	75.00	78.00	81.24	86.39	96.12	100.11	127.00	166.00	464.17
59.77	70.52	75.46	78.40	82.42	87.54	96.16	100.45	128.79	169.57	729.98
63.87	71.53	75.62	78.84	82.81	90.07	96.44	100.70	130.00	172.41	2148.94
64.30	72.31	76.12	79.58	83.77	90.33	96.44	112.03	132.63	181.31	

Group 3 - Three Family Dwellings, $n_3 = 62$

25.62	67.26	77.67	83.05	87.21	95.16	105.48	111.00	124.50	136.79	196.48
51.01	67.48	78.53	83.10	88.62	98.09	106.04	111.02	124.85	140.20	201.29
52.84	70.43	82.00	84.63	89.15	99.55	107.30	111.70	125.71	141.73	
62.38	72.85	82.24	85.43	89.17	102.48	108.19	116.11	129.50	164.11	
63.61	73.09	82.31	86.88	92.93	103.08	108.40	116.47	133.92	175.20	
63.92	75.42	82.64	87.04	92.95	104.36	108.73	121.12	136.31	195.14	

Group 4 - Four or More Family Dwellings, $n_4 = 28$

22.05	69.56	84.90	89.52	98.33	107.46	118.88	125.10	142.57	2854.00
37.79	71.32	85.76	90.00	102.39	112.83	119.91	129.14	152.34	
43.54	76.84	88.77	97.40	102.50	115.50	119.91	131.95	487.48	

REFERENCES

Barnett, A., and Eisen, E. (1982), "A Quartile Test for Differences in Distribution," *Journal of the American Statistical Association*, 77, 47-51.

Chernoff, H., and Savage, I. R. (1958), "Asymptotic Normality and Efficiency of Certain Nonparametric Test Statistics," *Annals of Mathematical Statistics*, 29, 972-994.

Conover, W. J. (1980), *Practical Nonparametric Statistics*, New York: Wiley.

De Wet, T., and Venter, J. H. (1973), "Asymptotic Distributions for Quadratic Forms with Applications to Tests of Fit," *Annals of Statistics*, 1, 380-387.

Duran, B. S., Tsai, W. S., and Lewis, T. O. (1976), "A Class of Location-Scale Nonparametric Tests," *Biometrika*, 63, 173-176.

Durbin, J., and Knott, M. (1972), "Components of Cramér-von Mises Statistics. I," *Journal of the Royal Statistical Society, Ser. B.*, 34, 290-307.

Durbin, J., Knott, M., and Taylor, C. C. (1975), "Components of Cramér-von Mises Statistics. II," *Journal of the Royal Statistical Society, Ser. B.*, 37, 216-237.

Fligner, A. F., and Hettmansperger, T. P. (1979), "On the Use of Conditional Asymptotic Normality," *Journal of the Royal Statistical Society, Ser. B*, 41, 178-183.

Hajek, J., and Sidak, Z. (1967), *Theory of Rank Tests*, New York: Academic Press.

Jureckova, J. (1979), "Nuissance Medians in Rank Testing Scale," in *Contributions to Statistics: Jaroslav Hajek Memorial Volume*, ed. Jana Jureckova, Holland: Reidell.

Kiefer, J. (1959), "K-Sample Analogues of the Kolmogorov-Smirnov and Cramér-v. Mises Tests," *Annals of Mathematical Statistics*, 29, 420-447.

Kim, P. J., and Jennrich, R. I. (1970), "Tables of the Exact Sampling Distribution of the Two-Sample Kolmogorov-Smirnov Criterion $D_{mn}(m \leq n)$, in *Selected Tables in Mathematical Statistics*, Vol. I, eds. H. L. Harter and D. B. Owen, Chicago: Markham.

Lehmann, E. L. (1975), *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day.

McKean, J. W., and Hettmansperger, T. P. (1978), "A Robust Analysis of the General Linear Model Based on One Step R-Estimates, *Biometrika*, 65, 571-580.

Miller, F. L., and Quesenberry, C. P. (1979), "Power Studies of Tests for Uniformity, II," *Communications in Statistics-Simulation and Computation*, B8, 271-290.

Moses. L. E. (1963), "Rank Tests of Dispersion," *Annals of Mathematical Statistics*, 34, 973-983.

Oskamp, S. (1962), "The Relationship of Clinical Experience and Training Methods to Several Criteria of Clinical Prediction," *Psychological Monographs*, Vol. 47, No. 547.

Pettitt, A. N. (1976), "A Two-Sample Anderson-Darling Rank Statstic," *Biometrika*, 63, 161-168.

__________(1979), "Two-Sample Cramér-von Mises Type Rank Statistics," *Journal of the Royal Statistical Society, Ser. B*, 46-53.

Randles, R. H. (1982), "On the Asymptotic Normality of Statistics with Estimated Parameters," *Annals of Statistics*, 10, 462-474.

Randles, R. H., and Wolfe, D. A. (1979), "*Introduction to the Theory of Nonparametric Statistics*, New York: Wiley.

Robinson, J. (1980), "An Asymptotic Expansion for Permutation Tests with Several Samples," *Annals of Statistics*, 8, 851-864.

Stephens, M. A. (1974), "Components of Goodness-of-Fit Statistics," *Annals de l'Institut Henri Poincaré, Ser. B.*, 10, 37-54.