COMPARING K POPULATIONS WITH LINEAR RANK STATISTICS

Dennis D. Boos*

Department of Statistics
North Carolina State University

Institute of Statistics Mimeo Series # 1613

October, 1982

*Visiting the Department of Statistics
University of California, Berkeley

ABSTRACT

Comparing k Populations with Linear Rank Statistics

A table of linear rank statistics is used to identify types of differences among populations in the one-way layout. The 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. General comparisons are formed from row and column sums of squares of the basic table and have approximate chi-squared distributions under the null hypothesis.

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 the use of a table of linear rank statistics derived from Cramér-von Mises statistics which provide reasonably high power of detection and simplicity of interpretation as well as distribution-free null distributions. The derivation and original motivation is given in Section 2 and is essentially due to Pettitt (1976).

The basic table of interest is as follows.

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}	$T_1=\sum_1^k\left(\frac{N-n_i}{N}\right)T_{i1}^2$
Scale	Mood = $(u-\frac{1}{2})^2 - 1/12$	T_{12}	T_{22}	...	T_{k2}	$T_2=\sum_1^k\left(\frac{N-n_i}{N}\right)T_{i2}^2$
Skewness	$20(u-\frac{1}{2})^3 - 3(u-\frac{1}{2})$	T_{13}	T_{23}	...	T_{k3}	$T_3=\sum_1^k\left(\frac{N-n_i}{N}\right)T_{i3}^2$
Kurtosis	$210(u-\frac{1}{2})^4 - 45(u-\frac{1}{2})^2 + 9/8$	T_{14}	T_{24}	...	T_{k4}	$T_4=\sum_1^k\left(\frac{N-n_i}{N}\right)T_{i4}^2$
		$\sum_{p=1}^4 T_{1p}^2$	$\sum_{p=1}^4 T_{2p}^2$		$\sum_{p=1}^4 T_{kp}^2$	$T=\sum_{p=1}^4 T_p$

The first two rows are familiar Wilcoxon and Mood linear 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 $T_{i2} < 0$ (>0) indicates that the ith sample has smaller (larger) scale than the combined other samples. The row weighted sums 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 $T_{i3} < 0$ (>0), then the ith sample is left (right) skewed *relative* to the other samples combined. Likewise, if $T_{i4} < 0$ (>0), then the ith sample is shorter (longer) tailed than the others combined. 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 $\sum_{p=1}^{4} T_{ip}^2$ converges to a chi-squared distribution with four degrees of freedom (χ_4^2). The k-sample statistics T_p converge under H_0 to a χ_{k-1}^2, and the overall statistic ΣT_p converges to a $\chi_{4(k-1)}^2$. The column sums $\sum_{p=1}^{4} T_{ip}^2$ are general statistics for deciding if there are differences between the ith sample and all the others combined, and the overall statistic ΣT_p is a global comparison of all the samples.

Sample moments $(\bar{X}_i, \hat{\sigma}_i^2 = n_i^{-1}\Sigma(X_{ij}-\bar{X}_i)^2$, $\sqrt{b_{1i}} = n_i^{-1}\Sigma(X_{ij}-\bar{X}_i)/\hat{\sigma}_i^3$, $b_{2i} = n_i^{-1}\Sigma(X_{ij}-\bar{X}_i)^4/\hat{\sigma}_i^4)$ are standard descriptive alternatives to $(T_{i1}, T_{i2}, T_{i3}, T_{i4})$. The linear rank statistics have an advantage in that they are distribution-free under H_0, not nearly as sensitive to outliers, and are easier to combine into general comparisons such as $\sum_{p=1}^{4} T_{ip}^2$ and $\sum_{p=1}^{4} T_p$. The main disadvantage of the linear rank statistics is that one component's result may affect the performance of another. For example, the effect of location differences on non-parametric scale tests is well known (Moses, 1963). Alignment by subtracting off location estimators upsets the distribution-free property, though an *asymptotic* distribution-free property will still hold for the scale components if the populations are symmetric (see Randles, 1982). Further alignment is possible by dividing by a scale estimator after subtracting off the location estimator. In general, without alignment one should be cautious about interpreting results of other components when one component is fairly large (see Ex. B, Sect. 5).

The paper is organized as follows. Section 2 traces the origin of the proposed methods. Section 3 uses moment calculations to show that the asymptotic null distributions are reached fairly quickly. Pitman efficiencies are employed in Section 4 to justify the association of each component with a particular type of alternative. Section 5 contains examples and Section 6 is a brief summary. Some readers may want to go directly to Section 5.

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

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

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 (1958) 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) $A^{*2}_{n_1 n_2} = (N/n_2)\int_0^1 x_n^2(t)w(t)dt$ can be expressed as

$$A^{*2}_{n_1 n_2} = \frac{N}{n_2}\sum_{j=1}^{\infty}\frac{B_j^2}{j(j+1)} ,$$

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, $A^{*2}_{n_1 n_2}$ 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 $A^{*2}_{n_1 n_2} = B_1^2/2 + B_2^2/6 + B_3^2/12 + B_4^2/20 + \sum_{j=5}^{\infty} B_j^2/(j(j+1))$ 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 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^2_ℓ distributions. The proposed linear rank statistics discussed in the Introduction are explicitly given by $T_{ip} = \sum_{\ell=1}^{n_i} a_{Nip}(R_{\ell i})$, p=1,4, where

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

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

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

However, 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

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

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.

3. NULL DISTRIBUTIONS

As discussed in the Introduction, all the proposed 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 3 moments of S are well known (see, e.g., Randles and Wolfe, 1979, Ch. 8) to be

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

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

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

The fourth central moment does not seem to appear in the literature. However, using shortcuts suggested by Ron Randles (personal communication) as well as tedious algebra, the following expression was obtained.

$$E(S-ES)^4 = \frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum_{j=1}^{N} [C_N(j)-\bar{C}_N]^4 \sum_{i=1}^{N} [a_N(j)-\bar{a}_N]^4$$

$$- \frac{3}{(N-2)(N-3)} \left[\left(\sum_{j=1}^{N} [C_N(j)-\bar{C}_N]^2 \right)^2 \sum_{j=1}^{N} [a_N(j)-\bar{a}_N]^4 \right.$$

$$\left. + \left(\sum_{j=1}^{N} [a_N(j)-\bar{a}_N]^2 \right)^2 \sum_{j=1}^{N} [C_N(j)-\bar{C}_N]^4 \right]$$

$$+ \frac{3(N^2-3N+3)}{N(N-1)(N-2)(N-3)} \left(\sum_{j=1}^{N} [C_N(j)-\bar{C}_N]^2 \right)^2 \left(\sum_{j=1}^{N} [a_N(j)-a_N]^2 \right)^2 .$$

For the two-sample statistics T_{ip}, $C_N(j)=1$ if R_j is from the j^{th} 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 1.

--- INSERT TABLE 1 HERE ---

Since all the β_2 values are less than 3.0, use of standard normal critical values will result in conservative tests when $\sqrt{\beta_1} = 0$. This need not be the case when $\sqrt{\beta_1} > 0$. For example, at $N=20$ and $n_i=5$ Pearson curve approximations (Johnson, et al., 1963) indicate that we should use 1.94, 2.02, 1.94, and 1.98 respectively in place of 1.96 for two-sided $\alpha=.05$ level tests. In general, Table 1 suggests that the normal approximation be used for all but the smallest sample sizes.

The column summary statistics $\sum_{p=1}^{4} T_{ip}^2$ 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 the scores (2.2), tedious calculations yield

$$E\left(\sum_{p=1}^{4} T_{ip}^2\right)^2 = \sum_{p=1}^{4} \beta_2(p) + 2\left[N(N+1)\sum_{j=1}^{N}[C_N(j)-\bar{C}_N]^4 - 3(N-1)\left(\sum_{j=1}^{N}[C_N(j)-\bar{C}_N]^2\right)^2\right]$$

$$\times \sum_{p<q} \sum_{j=1}^{N}[a_{Nip}(j)-\bar{a}_{Nip}]^2[a_{Niq}(j)-\bar{a}_{Niq}]^2/[(N-1)(N-2)(N-3)]$$

$$+ 2\left[(N^2-3N+6)\left(\sum_{j=1}^{N}[C_N(j)-\bar{C}_N]^2\right)^2 - (N^2-N+6)\sum_{j=1}^{N}[C_N(j)-\bar{C}_N]^4\right]$$

$$\times \sum_{p<q} \sum_{j=1}^{N}[a_{Nip}(j)-\bar{a}_{Nip}]^2 \sum_{j=1}^{N}[a_{Niq}(j)-\bar{a}_{Niq}]^2/[(N-1)(N-2)(N-3)] ,$$

where $\beta_2(p)$ is the coefficient of kurtosis of T_{ip}. For the combinations of N and n_i given in Table 1, the variance of ΣT_{ip}^2 is 4.462, 4.091, 5.678, 5.267, 6.765, 6.545, 7.366, and 7.262 respectively. These values suggest that use of χ_4^2 critical values (variance = 8) leads to conservative tests.

The other statistics of interest T_p and $T = \Sigma T_p$ are even more difficult to investigate with moment calculations. Of course, some work on null distributions is available for the Kruskal-Wallis T_1 and Mood's T_2 (see Conover, 1980, Sect. 5.2 and 5.3).

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 2 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 2 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 (T_{i1}) and Mood (T_{i2}) ARE's at the normal and Laplace are well known. The t_5 results are surprisingly good. The *low* values of T_{i3} for location and T_{i4} 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,

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

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 2 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. T_{i3} and T_{i4} 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." T_{i3} and T_{i4} are naturally optimal at the Legendre and reasonably efficient at the Fourier alternatives. The T_{i1} and T_{i2} values are all appropriately low in the second half of Table 2.

5. 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 . The linear rank analysis yields $(T_{11}, T_{12}, T_{13}, T_{14}, \sum_{p=1}^{4} T_{1p}^2)$ $= (-1.60, -1.35, -1.75, -1.15, 8.75)$, using midranks to break ties. Thus, there seems to be a complex difference in populations with at least the suggestion that February is smaller for all four types of departures. (Recall that the T_{ip} should be compared to a standard normal and ΣT_{ip}^2 to a χ_4^2 .) The overall statistic 8.75 has P-value .07. 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 numbers were 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 indicates that the original skewness and kurtosis values may in fact be associated with the location and scale differences.

B) The data in Table 5 was given by Oskamp (1962) and used by Draper (1982) to illustrate robust regression methods. The numbers are actually percentages of correct predictions of patient disorders by three different groups. The analysis is given in Table 3. There is a decreasing trend

--- insert Table 3 here ---

in location and also possible skewness differences. Staff are clearly different from the combination of trainees and undergraduates and undergraduates are different from the combination of staff and trainees. It might also be of interest to look at the three two-sample comparisons. The means (73.02, 70.52, 69.62) are in agreement with the Wilcoxon statistics (3.27,-.88,-2.21), but the standard deviations (2.56,3.28, 4.00) are in reverse order to the Mood statistics (1.49,-1.09,-.44). This is an example of location effects disturbing the scale statistics. After subtracting off trimmed means, the Mood row becomes (-.71,-.22, .82,T_2 = .85) and is now in agreement with the standard deviations (but not near significance). The Skewness and Kurtosis rows become (-.43,.60,-.17,T_3 = .39) and (-2.25,.17,1.94,T_4 = 5.92) which suggests differences in tail length rather than skewness. In Table 3 midranks were used for ties without adjusting mean and variance formulas since the statistics would only change a small amount.

C) Boos (1982) analyzes the ratio of sale prices to assessed value of residential property in Fitchburg, Mass., in 1979. This data provided by A. R. Manson is found in Table 6 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. As mentioned in Boos (1982), 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. Table 4 gives the linear rank analysis. Here, the first column answers an important question:

--- INSERT TABLE 4 HERE ---

how does the population of single family dwellings compare with the combined population of two or more family dwellings? There are clear location and scale differences. If single family dwellings are left out, the last column of Table 5 becomes $(2.95, 1.03, 11.25, 1.87, 17.11)^T$ which indicates that the two, three, and four or more family dwellings are similar except for skewness. Another analysis performed was to subtract off the individual medians and recompute Table 5. The T_p column becomes $(1.86, 58.72, 12.25, 2.05)^T$ which emphasizes the scale differences. A final analysis was to subtract off the medians and divide by the interquartile ranges. The T_p column then becomes $(3.38, 2.47, 19.46, 4.75)^T$ giving further evidence of skewness differences.

6. SUMMARY

Linear rank statistics can be used to explore differences among populations which are more general than location and scale. They are easy to interpret and have null distribution-free distributions which converge fairly quickly to standard normals. Two types of summary statistics are available in the proposed table. The row sums of squares are the usual k-sample rank tests for specific types of departures - location, scale, skewness, and kurtosis. The column sums of squares may be viewed as "Neyman-Barton smooth" type tests for equality of populations which are alternatives to generalized Cramér-von Mises statistics with the advantage of limiting chi-squared distributions.

Table 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 2. Pitman Efficiencies of the Components

	Location Alternative				Scale Alternative			
Family	T_{i1}	T_{i2}	T_{i3}	T_{i4}	T_{i1}	T_{i2}	T_{i3}	T_{i4}
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	0	.05	0	0	.67	0	.13

	Skewness Alternative				Kurtosis Alternative			
Type	T_{i1}	T_{i2}	T_{i3}	T_{i4}	T_{i1}	T_{i2}	T_{i3}	T_{i4}
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 3. Linear Rank Analysis of Oskamp Data

	Staff	Trainees	Undergraduates	T_p
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 4. Linear Rank Analysis of Fitchburg Data

	Single Family	Two Family	Three Family	Four or More Family	T_p
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 5. Percentage Accuracy of Predictions From Oskamp (1962)

Staff, $n_1 = 21$			Trainees, $n_2 = 23$				Undergraduates, $n_3 = 28$			
68.5	72.0	75.0	62.5	70.0	71.5	74.5	58.0	68.5	71.0	72.0
69.0	73.0	75.0	63.0	70.0	71.5	74.5	60.0	69.0	71.5	72.5
69.0	73.5	75.0	66.0	70.0	73.0		64.5	69.0	71.5	73.0
70.5	73.5	75.5	68.5	70.5	73.5		65.5	69.0	71.5	74.0
70.5	74.0	76.0	69.0	70.5	74.0		66.0	69.5	71.5	74.0
70.5	74.0	76.5	69.5	71.0	74.0		66.5	70.0	72.0	74.0
71.5	74.5	76.5	69.5	71.5	74.0		68.5	70.5	72.0	74.5

Table 6. 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 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.

Boos, D. D. (1982), "Estimating Nonparametric Parameters from Stratified Samples," preprint.

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 Test of Fit," Annals of Statistics, 1, 380-387.

Draper, D. (1982), "Rank-Based Robust Analysis of Linear Models," presented at the IMS annual meeting, Cincinnati.

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.

Johnson, N. L., Nixon, E., Amos, D. E., and Pearson, E. S. (1963), "Table of Percentage Points of Pearson Curves for Given $\sqrt{\beta_1}$ and β_2, Expressed in Standard Measure," Biometrika, 50, 459-478.

Kiefer, J. (1958), "K-Sample Analogues of the Kolmogorov-Smirnov and Cramér-v. Mises Tests," Annals of Mathematical Statistics, 29, 420-447.

Moses, L. E. (1963), "Rank Tests of Dispersion," Annals of Mathematical Statistics, 34, 973-83.

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 Statistic," Biometrika, 63, 161-168.

Randles, R. H. (1982), "On Rank Tests Applied to Residuals," preprint.

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

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