A CLASS OF TESTS BASED ON U-STATISTICS FOR THE SEVERAL SAMPLE PROBLEM

by

Jayant V. Deshpande

University of North Carolina

Institute of Statistics Mimeo Series No. 506

January 1967

This research was supported by the U. S. Air Force Office of Scientific Research, Grant no. AF-AFOSR-760-65.

DEPARTMENT OF STATISTICS

University of North Carolina

Chapel Hill, N. C.

1. Summary and Introduction:

Let $x_{i1}, \ldots, x_{in_i}$ be a sample of real observations from the i-th population, with cumulative distribution function (c.d.f.), $F_i(x)$, $i = 1,2,\ldots,c$. Let the c samples be independent and the F's be continuous. In this paper we shall consider tests for the null hypothesis

$$H_0: \quad F_1(x) = F_2(x) = \ldots = F_c(x) = F(x), \text{ say.}$$

In particular we will be interested in tests for the situations when one of the following two assumptions is considered reasonable. (i) The F's have identical functional form and if at all, the differences are only in their location parameters i.e. $F_i(x) = F(x-\mu_i)$, μ_i being any real number. In such a situation, a test for H_0 is, in effect, a test for the hypothesis $\mu_1 = \mu_2 = \ldots = \mu_c$. (ii) Again assuming identical functional form and assuming that the differences, if any, to be among the scale parameters, i.e. $F_i(x) = F(x\theta_i)$ θ_i being real and positive, a test of H_0 reduces to a test of the hypothesis $\theta_1 = \theta_2 = \ldots = \theta_c$.

Let c-plets be formed by selecting one observation from each of the c samples. The total number of c-plets that can be formed in this way is $\prod_{i=1}^{c} n_i$. Let v_{ij} be the number of c-plets in which the observation selected from the i-th sample is larger than exactly (j-1) observations (and smaller than the other (c-j) observations.) Since the F's are assumed to be continuous, the probability of the existance of ties is zero. Let us define $u_{ij} = v_{ij} / \prod_{i=1}^{c} n_i$. It is seen that these u's are the generalized version [13] to c samples of Hoeffding's U-statistics [8].

Let us have $N = \sum_{i=1}^{c} n_i$, $p_i = n_i/N$, $L_i = \sum_{j=1}^{c} a_j u_{ij}$, where a's are real constants (not all equal) and

$$A = \sum_{j=1}^{c} \sum_{\ell=1}^{c} a_j a_\ell \left\{ \frac{\binom{c-1}{j-1}\binom{c-1}{\ell-1}}{\binom{2c-2}{j+\ell-2}(2c-1)} - \frac{1}{c^2} \right\}.$$

Then we define a class of statistics $\mathcal{L}$ as

$$\mathcal{L} = \frac{N(c-1)^2}{A\ c^2} \left[\sum_{i=1}^{c} p_i\ L_i^2 - \left(\sum_{i=1}^{c} p_i L_i \right)^2 \right].$$

With each of the members of this class we associate a test: Reject H_0 at a significance level α if $\mathcal{L}$ exceeds a predetermined constant $\mathcal{L}_\alpha$. We, later in this paper, show that under H_0, $\mathcal{L}$ is distributed as χ^2 with c-1 degrees of freedom in the limit as $N \rightarrow \infty$. Hence for sufficiently large N, $\mathcal{L}_\alpha$ may be approximated by the corresponding significance point of the relevant χ^2 distribution.

Tests proposed by Bhapkar [2],[3], Sugiura, [12], and the author [5], [6] may be seen to be members of this class. These tests have been proposed on heuristic grounds and on considerations of mathematical simplicity (at least, so far as the tests proposed by the present author are concerned). In this paper it is attempted to give unified treatment of the asymptotic distribution of all the u_{ij} (i, j = 1,2,...,c) and to explore tests based on them for the null hypothesis H_0 against alternative hypotheses of interest. The earlier test statistics, mentioned above, were constructed, taking into account the magnitude of u's under the null hypothesis and the alternative hypothesis of interest. a's were selected in such a way as to emphasize the difference between the two. In this paper we shall obtain a's which maximize the noncentrality parameter of the non central χ^2 distribution, which $\mathcal{L}$ has under alternative hypotheses. Under the null hypothesis, the noncentrality parameter is, of course, zero. Since the maximum of the noncentrality parameter depends on F; this approach will lead us to the test which will have maximum asymptotic relative efficiency (in the Pitman sense) amongst this class of tests against a specified alternative for a particular F. This test may be used if the particular F is suspected as likely.

We find a condition involving a's and F's for the consistency of tests in this class against specified alternatives. $\mathcal{L}$ is shown to have, asymptotically, a noncentral X^2 distribution with c-1 degrees of freedom. Under certain specified alternative hypotheses. The a's which give the maximum of the noncentrality parameter for a given alternative hypothesis are found. These a's depend on the original distribution F. The tests with these a's defining L_i, will have maximum asymptotic relative efficiency (ARE) amongst this class of tests, against the specified alternative and for the particular distribution. Some ARE 's of members of this class with respect to the Kruskal's H test [9] are computed.

2. Distribution of u_{ij} under H_0:

In this section we prove

Theorem 2.1 Let X_{ij} be independent random variables with continuous c.d.f. $F_i(x)$, $i = 1,2,\ldots,c$, $j = 1,2,\ldots, n_i$. Then if $F_1(x) = \ldots = F_c(x) = F(x)$, the distribution of $w_{ij} = N^{1/2}(u_{ij} - \frac{1}{c})$, in the limit as $N \to \infty$ in such a way as p_i, $(i= 1,2,\ldots,c)$ remain fixed, is normal with mean zero and the elements of the covariance matrix Σ given by

$$\sigma_{ij,k\ell} = \text{cov}\ (w_{ij},\ w_{k\ell}) \text{ where}$$

$$
(2.1)\begin{cases}
\sigma_{ij,i\ell} = \dfrac{1}{(c-1)^2}\left[\dfrac{\binom{c-1}{j-1}\binom{c-1}{\ell-1}}{\binom{2c-2}{j+\ell-2}(2c-1)} - \dfrac{1}{c^2}\right]\left[\dfrac{(c-1)^2}{p_i} + \sum_{r\neq i}\dfrac{1}{pr}\right] \\
\text{and} \\
\underset{i\neq k}{\sigma_{ij,k\ell}} = \dfrac{1}{(c-1)^2}\left[\dfrac{\binom{c-1}{j-1}\binom{c-1}{\ell-1}}{\binom{2c-2}{j+\ell-2}(2c-1)} - \dfrac{1}{c^2}\right]\left[\sum_{r=1}^{c}\dfrac{1}{p_r} - \dfrac{c}{p_i} - \dfrac{c}{p_k}\right].
\end{cases}
$$

u_{ij}, N and p_i being the same as defined in section 1.

<u>Proof</u>: Let us define

$$
\Phi_{ij}(x_{1t_1}, x_{2t_2}, \ldots, x_{ct_c}) = 1 \text{ if } x_{it_i} \text{ is larger than exactly } (j-1) \text{ other } x\text{'s}
$$
$$
= 0 \text{ otherwise.}
$$

Then it is seen that

$$
(2.2) \qquad u_{ij} = \frac{1}{\pi\, n_i} \sum_{t_1=1}^{n_1} \sum_{t_2=1}^{n_2} - - - \sum_{t_c=1}^{n_c} \Phi_{ij}(x_{1t_1}, x_{2t_2}, \ldots, x_{ct_c})
$$

for $i, j = 1,2,\ldots,c$.

Obviously u_{ij} are U - statistics generalized to the c - sample case. They, therefore, under H_0 have asymptotically, as $N \to \infty$ in such a way that p_i remain constant, a multivariate normal distribution with

$$
\mathcal{E}(u_{ij}) = \mathcal{E}(\Phi_{ij}) = \eta_{ij} \text{ say and}
$$

$$
(2.3) \qquad \lim_{N\to\infty} N \operatorname{Cov}(u_{ij}, u_{k\ell}) = \sum_{r=1}^{c} \frac{1}{p_r} \xi^{(r)}_{ij,k\ell}, \text{ where}
$$

$$
\xi^{(r)}_{ij,k\ell} = \mathcal{E}[\Phi_{ij}(X_1, X_2, \ldots, X_r, \ldots, X_c)\ \Phi_{k\ell}(X_1', X_2', \ldots, X_r, \ldots, X_c')] - \eta_{ij}\,\eta_{k\ell}.
$$

We evaluate these quantities.

$$\eta_{ij} = \mathcal{E}[\Phi_{ij}(X_{ct_c}, X_{1t_1}, X_{2t_2}, \ldots, X_{ct_c})]$$

$$= \text{Prob } [X_{it_i} \text{ is larger than } (j-1)\ X\text{'s and smaller than } (c-j)\ X\text{'s}]$$

$$= \frac{1}{c}\ .$$

$$\xi^{(i)}_{ij,i\ell} = \mathcal{E}[\Phi_{ij}(X_1, \ldots, X_i, \ldots, X_c) \cdot \Phi_{i\ell}(X'_1, \ldots, X_i, \ldots, X'_c)] - \eta_{ij}\eta_{i\ell}$$

$$= \text{Prob } [X_i \text{ is larger than } (j-1)\ X\text{'s and } (\ell-1)\ X'\text{'s and smaller than } (c-j)\ X\text{'s and } (c-\ell)\ X'\text{'s}] - \frac{1}{c^2}$$

(2.4) $$= \frac{\binom{c-1}{j-1}\binom{c-1}{\ell-1}}{\binom{2c-2}{j+\ell-2}(2c-1)} - \frac{1}{c^2}\ .$$

$\xi^{(r)}_{ij,k\ell}$, $r \neq i,k$

$$= \mathcal{E}[\Phi_{ij}(X_1, \ldots, X_r, \ldots, X_c).\ \Phi_{k\ell}(X'_1, \ldots, X_r, \ldots, X'_c)] - \frac{1}{c^2}$$

$$= \mathcal{E}[\{\int_{-\infty}^{X_r} \binom{c-2}{j-1}[F(x)]^{j-1}[1-F(x)]^{c-j-1}dF(x) + \int_{X_r}^{\infty} \binom{c-2}{j-2}[F(x)]^{j-2}[1-F(x)]^{c-j}\, df(x)\}$$

$$\cdot \{\int_{-\infty}^{X_r} \binom{c-2}{\ell-1}[F(x)]^{\ell-1}[1-F(x)]^{c-\ell-1}dF(x) + \int_{X_r}^{\infty} \binom{c-2}{\ell-2}[F(x)]^{\ell-2}[1-F(x)]^{c-\ell}\, dF(x)\}] - \frac{1}{c^2}$$

$$= \mathcal{E}[\{I_1 + I_2\} \cdot \{I_3 + I_4\}] - \frac{1}{c^2}, \text{ say.}$$

It is seen that

$$I_1 + I_2 = \frac{1}{(c-1)}\{1 - \binom{c-1}{j-1}[F(x)]^{j-1}[1-F(x)]^{c-j}\}$$

and

$$I_3 + I_4 = \frac{1}{(c-1)}\{1 - \binom{c-1}{\ell-1}[F(x)]^{\ell-1}[1-F(x)]^{c-\ell}\}$$

$$\therefore \ \xi^{(r)}_{ij,k\ell} \underset{r \neq i,k}{} = \int_{-\infty}^{\infty} \frac{1}{(c-1)^2} \{1 - \binom{c-1}{j-1}[F(x)]^{j-1}[1-F(x)]^{c-j}\}$$

$$\{1 - \binom{c-1}{\ell-1}[F(x)]^{\ell-1}(1-F(x)]^{c-\ell}\} dF(x)$$

(2.5) $$= \frac{1}{(c-1)^2}\left[\frac{\binom{c-1}{j-1}\binom{c-1}{\ell-1}}{\binom{2c-2}{j+\ell-2}(2c-1)} - \frac{2}{c} + 1\right] - \frac{1}{c^2}.$$

And lastly

$$\xi^{(i)}_{ij,k\ell} = \mathcal{E}[\Phi_{ij}(X_1,\ldots,X_i,\ldots,X_c)\ \Phi_{k\ell}(X_1',\ldots,X_i,\ldots,X_c')] - \frac{1}{c^2}.$$

$$= \mathcal{E}[\binom{c-1}{j-1}[F(x)]^{j-1}[1-F(x)]^{c-j}$$

$$\cdot\{\int_{-\infty}^{X_i} \binom{c-2}{\ell-1}[F(x)]^{\ell-1}(1-F(x)]^{c-\ell-1} dF(x) +$$

$$\int_{X_i}^{\infty} \binom{c-2}{\ell-2}[F(x)]^{\ell-2} [1-F(x)]^{c-\ell} dF(x)\}] - \frac{1}{c^2}$$

$$= \int_{-\infty}^{\infty} \frac{1}{(c-1)} \{\binom{c-1}{j-1}[F(x)]^{j-1}[1-F(x)]^{c-j} -$$

$$\binom{c-1}{j-1}\binom{c-1}{\ell-1}[F(x)]^{j+\ell-2} [1 - F(x)]^{2\ell-j-\ell}\} dF(x) - \frac{1}{c^2}$$

(2.6) $$= \frac{1}{(c-1)}\left[\frac{1}{c} - \frac{\binom{c-1}{j-1}\binom{c-1}{\ell-1}}{\binom{2c-2}{j+\ell-2}(2c-1)}\right] - \frac{1}{c^2}.$$

It is seen that $\xi^{(i)}_{ij,k\ell} = \xi^{(k)}_{ij,k\ell}$.

From (2.3) we write

$$\lim_{N\to\infty} N\,\mathrm{Cov}(u_{ij},u_{i\ell}) = \frac{1}{p_i}\,\xi^{(i)}_{ij,i\ell} + \sum_{r\neq i}\frac{1}{p_r}\,\xi^{(r)}_{ij,i\ell}$$

Substituting from (2.4) and (2.5) we obtain

(2.7) $$= \frac{1}{(c-1)^2}\left[\frac{\binom{c-1}{j-1}\binom{c-1}{\ell-1}}{\binom{2c-2}{j+\ell-2}(2c-1)} - \frac{1}{c^2}\right]\left[\frac{(c-1)^2}{p_i} + \sum_{r\neq i}\frac{1}{p_r}\right]$$

Similarly we write using (2.5) and (2.6)

(2.8) $$\lim_{N\to\infty} N\,\mathrm{Cov}(u_{ij},u_{k\ell})_{i\neq k} = \frac{1}{(c-1)^2}\left[\frac{\binom{c-1}{j-1}\binom{c-1}{\ell-1}}{\binom{2c-2}{j+\ell-2}(2c-1)} - \frac{1}{c^2}\right]$$

$$\left[\sum_{r=1}^{c}\frac{1}{p_r} - \frac{c}{p_i} - \frac{c}{p_k}\right].$$

Hence we conclude that w_{ij} is, in the limit distributed normally with zero means and the elements of covariance matrix given by 2·1.

3. The statistic $\mathcal{L}$ and the class of tests.

Let us consider linear forms L_i of u_{ij} for i = 1,2,...,c. We define

(3.1) $$L_i = a_1 u_{i1} + a_2\, u_{i2} + \ldots + a_c\, u_{ic}.$$

It is assumed that a_j are not all equal and are real constants.

$$(3.2) \qquad \mathcal{E}(L_i) = \sum_{j=1}^{c} a_j \, \mathcal{E}(u_{ij}) = \frac{1}{c} \sum_{j=1}^{c} a_j = \bar{a} \text{ say}$$

Let $\lambda_{ik} = \lim_{N \to \infty} N \operatorname{Cov}(L_i, L_k)$. for $i, k = 1,2,\ldots,c$.

$$\text{Then } \lambda_{ii} = \sum_{j=1}^{c} \sum_{j=1}^{c} a_j a_\ell \, \sigma_{ij,i\ell}$$

$$= \frac{1}{(c-1)^2} \left[\frac{(c-1)^2}{p_i} + \sum_{r \neq i} \frac{1}{p_r} \right] \left[\sum_{j=1}^{c} \sum_{\ell=1}^{c} a_j a_\ell \left\{ \frac{\binom{c-1}{j-1}\binom{c-1}{\ell-1}}{\binom{2c-2}{j+\ell-2}(2c-1)} - \frac{1}{c^2} \right\} \right]$$

And

$$\lambda_{\substack{ik \\ i \neq k}} = \sum_{j=1}^{c} \sum_{\ell=1}^{c} a_j a_\ell \, \sigma_{ij,k\ell}$$

$$(3.3) \qquad = \frac{1}{(c-1)^2} \left[\sum_{r=1}^{c} \frac{1}{p_r} - \frac{c}{p_i} - \frac{c}{p_k} \right] \left[\sum_{j=1}^{c} \sum_{\ell=1} a_j a_\ell \left\{ \frac{\binom{c-1}{j-1}\binom{c-1}{\ell-1}}{\binom{2c-2}{j+\ell-2}(2c-1)} - \frac{1}{c^2} \right\} \right]$$

(3.2) and (3.3) can be rewritten using the definition of A given in (1.1) as

$$(3.4) \qquad \begin{cases} \lambda_{ii} = \dfrac{A}{(c-1)^2} \left[\dfrac{(c-1)^2}{p_i} + \displaystyle\sum_{r \neq i} \dfrac{1}{p_r} \right] & \text{and} \\[2ex] \lambda_{ik} = \dfrac{A}{(c-1)^2} \left[\displaystyle\sum_{r=1}^{c} \dfrac{1}{p_r} - \dfrac{c}{p_i} - \dfrac{c}{p_k} \right]. & \end{cases}$$

Hence we conclude that $N^{1/2}(\underset{\sim}{L} - \bar{a}\underset{\sim}{J})$ has, in the limit as $N \to \infty$, a multivariate normal distribution with mean vector $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{\Lambda}$. Here $\underset{\sim}{L}' = (L_1, L_2, \ldots, L_c)_{1xc}$, $\underset{\sim}{J}' = (1,1,\ldots 1)_{1xc}$, $\underset{\sim}{0}' = (0,0,\ldots,0)_{1xc}$ and

$$\underset{\sim}{\Lambda} = (\lambda_{ik})_{cxc} \quad i,k = 1,2,\ldots,c.$$

The multivariate distribution is singular since $\Sigma L_i = \Sigma a_i = K$. In fact it may, trivially, be observed that $\underset{\sim}{\Lambda}\,\underset{\sim}{J} = \underset{\sim}{0}$. We consider the distribution of $N^{1/2}(\underset{\sim}{L}_0 - \bar{a}\, J_0)$. It is nonsingular with $\underset{\sim}{0}_0$ mean and $\underset{\sim}{\Lambda}_0$ as covariance matrix. Here $\underset{\sim}{L}_0' = (L_1, \ldots L_{c-1})_{1xc-1}$, $\underset{\sim}{J}_0' = (1,1,\ldots,1)_{1xc-1}$, $\underset{\sim}{0}_0' = (0,\ldots,0)_{1xc-1}$ and $\underset{\sim}{\Lambda}_0 = (\lambda_{ik})_{(c-1)x(c-1)}$ $i,k = 1,2,\ldots,c-1$.

Therefore $\mathcal{L} = N(\underset{\sim}{L}_0 - \bar{a}\,\underset{\sim}{J}_0)'\ \underset{\sim}{\Lambda}_0^{-1}\ (\underset{\sim}{L}_0 - \bar{a}\underset{\sim}{J}_0)$ is distributed under the null hypothesis as $N \to \infty$, as a χ^2 variate with c-1 degrees of freedom. Following Bhapkar [2], we simplify and obtain

$$\mathcal{L} = \frac{(c-1)^2 N}{c^2 A}\left[\left\{\sum_{i=1}^{c} p_i (L_i - \bar{a})^2\right\} - \left\{\sum_{i=1}^{c} p_i (L_i - \bar{a})\right\}^2\right]$$

$$(3.5) \qquad = \frac{(c-1)^2 N}{c^2 A}\left[\sum_{i=1}^{c} p_i L_i^2 - \left\{\sum_{i=1}^{c} p_i L_i\right\}^2\right].$$

We have proved the following theorem:

<u>Theorem 3.1</u> If $F_1 = F_2 = \ldots = F_c$ and $n_i = N p_i$ where the p_i are fixed numbers such that $\sum_{i=1}^{c} p_i = 1$, then the statistic $\mathcal{L}$, as defined in (3.5) above, for any real a_j such that they are not all equal, has a limiting χ^2 distribution with c-1 degrees of freedom.

It may be noted that Bhapkar's V [2] and W [3] statistics, Sugiura's V_{rs} and D_{rs} [12] statistics and the L [5] and D[6] statistics proposed by the author are obtained from $\mathcal{L}$ by choosing appropriate a_j.

4. Consistency of tests based on $\mathcal{L}$:

In this section we give a condition for the consistency of tests based on $\mathcal{L}$, using a lemma of Bhapkar [2] which we quote below.

Lemma (Bhapkar): Let $\eta^{(i)} = f^{(i)}(F_1,F_2,\ldots,F_c)$, $i = 1,2,\ldots,g$, be real valued functions such that $f^{(i)}(F_1,F_2,\ldots,F_c) = \eta_0^{(i)}$ for all $(F,F,\ldots,F)$ in a class C_0. Let $T^{(i)}_{n_1,n_2,\ldots,n_c} = t^{(i)}(X_{11},\ldots,X_{1n_1};\ \ldots\ ;$ $X_{c1},\ldots,X_{cn_c})$, $i = 1,2,\ldots,g$ be a sequence of real valued statistics such that $T^{(i)}_{n_1,n_2,\ldots,n_c}$ tends to $\eta^{(i)}$ in prob. as $\min(n_1,n_2,\ldots,n_c) \to \infty$. Suppose that at least one $f^{(i)}(F_1,F_2,\ldots,F_c) \neq \eta_0^{(i)}$ for all $F_1,F_2,\ldots,F_c$ in a class C_1. Further let $W_{n_1,n_2,\ldots,n_c} = w(T^{(1)}_{n_1,n_2,\ldots,n_c},\ldots,T^{(g)}_{n_1,n_2,\ldots,n_c})$ be a non negative function which is zero if and only if $T^{(i)}_{n_1,n_2,\ldots,n_c} = \eta_0^{(1)}$ for all i. Then the sequence of tests which rejects when $W_{n_1,n_2,\ldots,n_c} > d_{n_1,n_2,\ldots,n_c}$ is consistent for testing

H: C_0 at every fixed level of significance against the alternative H': C_1.

Let us take

$$\eta^{(i)} = \sum_{j=1}^{c} a_j \cdot P[X_i \text{ is greater than } (j-1)\ X\text{'s and smaller than } (c-j)\ X\text{'s}].$$

Where X_i are independent random variables with distribution functions F_i. Let $T^{(i)}_{n_1,n_2,\ldots,n_c} = L_i$ for $i = 1,2,\ldots,c$. Then the convergence in probability of L_i to $\eta^{(i)}$ as $\min(n_1, n_2, \ldots, n_c) \to \infty$, follows since $N^{1/2}(L_i - \eta^{(i)})$ has, in the limit, normal distribution with mean 0 and finite variance.

We have

$$(4.1) \qquad \eta_0^{(i)} = \sum_{j=1}^{c} a_j \, \mathcal{E}(\Phi_{ij} \mid H_0) = \frac{1}{c} \sum_{j=1}^{c} a_j = \bar{a}.$$

For the class C_1 we have

$$(4.2) \qquad \eta^{(i)} = \sum_{j=1}^{c} a_j \left\{ \int_{-\infty}^{\infty} \Sigma^* \prod_{(j-1)\text{ terms}} F_r(x) \prod_{(c-j)\text{ terms}} [1 - F_s(x)] dF_i(x) \right\}$$

Here Σ^* indicates summation over all possible choices of (j-1) F's out of (c-1) F's (all except F_i).

Hence we conclude that tests of the type which reject $H: F_1 = F_2 = \ldots = F_c$ if

$$\mathcal{L} > \mathcal{L}_\alpha$$

are consistent for all F_i, $i = 1, 2, \ldots, c$, such that $\eta^{(i)}$ defined by (4.2) is different from $\bar{a}$ for at least one i. It may be noted that $\mathcal{L}$ is a nonnegative function of L_i and equal to zero only when $L_i = \bar{a}$ for each i.

This condition does not appear to be too restrictive.

5. Distribution of u_{ij} and $\mathcal{L}$ under alternative hypotheses

In this section we derive the limiting distribution of u_{ij} and $\mathcal{L}$ under the following two sequences of alternative hypothesis

$$(5.1) \qquad H_{L_n}: \; F_i(x) = F(x - n^{-1/2}\theta_i) \text{ and}$$

$$(5.2) \qquad H_{S_n}: \; F_i(x) = F(x(1 + n^{-1/2}\delta_i)).$$

Here n is given by the relation $n_i = ns_i$ where s_i are fixed integers, all θ_i are not equal, $\delta_i > 0$ for each i and all δ_i are not equal.

Theorem 5.1: (a) w_{ij}, as defined in theorem 2.1, has jointly in the limit as $n \to \infty$, under H_{L_n} multivariate normal distribution, with means

$$(5.3) \qquad \eta_{ij}^{L_n} = \left(\sum_{i=1}^{c} s_i\right)^{1/2} \sum_{r=1}^{c} (\theta_r - \theta_i) \left\{ \binom{c-2}{j-1} \int_{-\infty}^{\infty} f^2(x)[F(x)]^{j-1}[1-F(x)]^{c-j-1} dx \right.$$

$$\left. - \binom{c-2}{j-2} \int_{-\infty}^{\infty} f^2(x)\, [F(x)]^{j-2}[1 - F(x)]^{c-j} dx \right\}$$

and elements of the covariance matrix given by (2.1) under the following two conditions.

(i) F is absolutely continuous with derivative f and

(ii) There exists a function g such that

$$|[f(y+h) - f(y)]/h| \leq g(y) \quad \text{for small } h \text{ and}$$

$$\int_{-\infty}^{\infty} g(y)\, f(y)\, dy < \infty .$$

(b) Under H_{L_n} and the above two conditions $\mathcal{L}$ has, in the limit as $n \to \infty$, noncentral χ^2 distribution with c-1 degrees of freedom and noncentrality parameter given by

$$(5.4) \qquad \mu_{L_n} = \frac{(\underset{\sim}{a}'\,\underset{\sim}{b})^2}{A} \sum s_i\, (\theta_i - \bar{\theta})^2 \quad \text{where}$$

$$\underset{\sim}{a}' = (a_1, a_2, \ldots, a_c), \quad \underset{\sim}{b}' = (b_1, b_2, \ldots, b_c),$$

$$b_j = \binom{c-1}{j-1} \{(c-j) \int_{-\infty}^{\infty} f^2(x)[F(x)]^{j-1}[1-F(x)]^{c-j-1}dx$$

$$-(j-1) \int_{-\infty}^{\infty} f^2(x) [F(x)]^{j-2} [1 - F(x)]^{c-j}dx\}$$

and A as defined in (1.1) and $\overline{\theta} = \Sigma s_i\theta_i / \Sigma s_i$

Proof. (a)

$$\mathcal{E}(u_{ij}|H_{L_n}) = \Sigma' \int_{-\infty}^{\infty} \pi_{(j-1)\text{ terms}} F(x-n^{-1/2}\theta_r) \pi_{(c-j)\text{ terms}} [1 - F(x-n^{-1/2}\theta_k)]$$

$$dF(x-n^{-1/2}\theta_i)$$

{In the above expression Σ' indicates summation over all possible choices of (j-1) F's out of (c-1) F's (all except F_i)}.

$$= \Sigma' \int_{-\infty}^{\infty} \pi_{(j-1)\text{ terms}} F(x-n^{-1/2}\theta_r + n^{-1/2}\theta_i)$$

$$\pi_{(c-j)\text{ terms}} [1-F(x-n^{-1/2}\theta_k + n^{-1/2}\theta_i)] \, d\, F(x).$$

We expand F in the above expression in a Taylor series around x upto first term and the remainder. Then using condition (ii) of the theorem we may write

$$= \Sigma' \int_{-\infty}^{\infty} \{[F(x)]^{j-1}[1 - F(x)]^{c-j} - n^{-1/2} f(x) [F(x)]^{j-2}$$

$$[1-F(x)]^{c-j} \sum_{(j-1)\text{ terms}} (\theta_r - \theta_i)$$

$$+ n^{-1/2} f(x) [F(x)]^{j-1} [1-F(x)]^{c-j-1} \sum_{(c-j)\text{ terms}} (\theta_k - \theta_i) \}dF(x) + o(n^{-1})$$

$$= \frac{1}{c} + n^{-1/2} \sum_{r=1}^{c} (\theta_r-\theta_i) \{\binom{c-2}{j-1} \int_{-\infty}^{\infty} f^2(x)[F(x)]^{j-1}[1-F(x)]^{c-j-1}dx$$

$$- \binom{c-2}{j-2} \int_{-\infty}^{\infty} f^2(x)\,[F(x)]^{j-2}[1-F(x)]^{c-j}\,dx\} + O(n^{-1})$$

Proceeding on exactly similar lines and using conditions (i) and (ii) of the theorem we see that

$$\lim_{N \to \infty} N \operatorname{Cov}(u_{ij}, u_{k\ell} \mid H_{L_n}) = \lim_{N \to \infty} N \operatorname{Cov}(u_{ij}, u_{k\ell} \mid H_0)$$

Hence part (a) of the theorem follows.

(b) It follows that $\mathcal{L}$, in the limit as $n \to \infty$ is distributed as noncentral X^2 with c-1 degrees of freedom and the noncentrality parameter

$$\frac{(c-1)^2}{A} \sum_{i=1}^{c} s_i(\theta_i - \overline{\theta})^2 \Big[\sum_{j=1}^{c} a_j \Big\{\binom{c-2}{j-1} \int_{-\infty}^{\infty} f^2(x)[F(x)]^{j-1}[1-F(x)]^{c-j-1}dx$$

$$-\binom{c-2}{j-2} \int_{-\infty}^{\infty} f^2(x)\,[t - F(x)]^{j-2}\,[1 - F(x)]^{c-j}\,dx\Big\}$$

which further simplifies to (5.4)

<u>Theorem 5.2</u>: (a) w_{ij} as defined in theorem 2.1 has jointly in the limit as $n \to \infty$, under H_{S_n}, multivariate normal distribution, with means

$$(5.5) \qquad \eta_{ij}^{S_n} = \Big(\sum_{i=1}^{c} s_i\Big)^{1/2} \sum_{r=1}^{c} (\delta_r - \delta_i)\,\Big\{\binom{c-2}{j-2} \int_{-\infty}^{\infty} x\, f^2(x)\,[F(x)]^{j-2}$$

$$[1-F(x)]^{c-j}\,dx - \binom{c-2}{j-1} \int_{-\infty}^{\infty} x\, f^2(x)\,[F(x)]^{j-1}[1-F(x)]^{c-j-1}dx\Big\}$$

and the elements of the covariance matrix given by (2.1) under the following conditions:

(i) F is absolutely continuous with derivative f.

(ii) There exists a function g such that $|[f(x) - f(x+hx)]/h| \leq g(x)$ for small h

and $\int_{-\infty}^{\infty} [x\, g(x)]^i f(x)\, dx < \infty$ for $i = 1,2,\ldots, 2c-1$.

and (iii) There exists $A < \infty$ such that

$$P_F\, [|x\, f(x)| < A] = 1.$$

(b) Under H_{S_n} and the above three conditions, $\mathcal{L}$ has, in the limit as $n \to \infty$, a noncentral χ^2 distribution with c-1 degrees of freedom and noncentrality parameter given by

$$(5.6) \qquad \mu_{S_n} = \frac{(\underset{\sim}{a}'\, \underset{\sim}{d})^2}{A} \sum_{i=1}^{c} s_i (\delta_i - \bar{\delta})^2 \qquad \text{where}$$

$\underset{\sim}{d}' = (d_1,\ldots,d_c)$ with

$$d_j = \binom{c-1}{j-1} \{(j-1) \int_{-\infty}^{\infty} x\, f^2(x)\, [F(x)]^{j-2} [1-F(x)]^{c-j}\, dx$$

$$- (c-j) \int_{-\infty}^{\infty} x\, f^2(x)\, [F(x)]^{j-1} [1-F(x)]^{c-j-1} dx\}$$

and $\bar{\delta} = \Sigma\, s_i\, \delta_i / \Sigma\, s_i$.

$\underset{\sim}{a}$ and A are same as in theorem 5.1.

<u>Proof</u>. (a)

$$\mathcal{E}(u_{ij}\, |H_{S_n}) = \Sigma' \int_{-\infty}^{\infty} \prod_{(j-1)\text{ terms}} [F\ (x)] \prod_{(c-j)\text{ terms}} [1-F_k(x)]\, dF_i(x)$$

(Σ' indicates summation as in theorem 5.1)

$$= \Sigma' \int_{-\infty}^{\infty} \prod_{(j-1) \text{ terms}} [F(x(1 + n^{-1/2} \delta r)(1+n^{-1/2} \delta_i)^{-1})]$$

$$\prod_{(c-j) \text{ terms}} [1 - F(x(1+n^{-1/2} \delta_k)(1+n^{-1/2}\delta_i)^{-1})] dF(x)$$

[We now expand F in a Taylor series around x upto first term and the remainder. This is permissible in view of the condition (i) in the theorem. Then using conditions (ii) and (iii) we write]

$$= \Sigma' \int_{-\infty}^{\infty} \{[F(x)]^{j-1}[1 - F(x)]^{c-j} + n^{-1/2} x f(x) [F(x)]^{j-2}[1 - F(x)]^{c-j}$$

$$\sum_{(j-1) \text{ terms}} (\delta_r - \delta_i)(1+n^{-1/2} \delta_i)^{-1} - n^{-1/2} x f(x)[F(x)]^{j-1}[1 - F(x)]^{c-j-1}$$

$$\sum_{(c-j) \text{ terms}} (\delta_r-\delta_i)(1+n^{-1/2}\delta_i)^{-1}\} dF(x) + O(n^{-1})$$

$$= \frac{1}{c} + n^{-1/2} \sum_{r=1}^{c} (\delta_r-\delta_i) \{\binom{c-2}{j-2} \int_{-\infty}^{\infty} x f^2(x) [F(x)]^{j-2} [1 - F(x)]^{c-j} dx$$

$$- \binom{c-2}{j-1} \int_{-\infty}^{\infty} x f^2(x) [F(x)]^{j-1} [1 -F(x)]^{c-j-1}\} dx + O(n^{-1}).$$

After lengthy derivation on similar lines we obtain that

$$\lim_{N \to \infty} N \operatorname{Cov}(u_{ij}, u_{k\ell} \mid H_{S_n}) = \lim_{N \to \infty} N \operatorname{Cov}(u_{ij}, u_{k\ell} \mid H_0).$$

Hence part (a) of the theorem is proved.

(b) It follows easily that $\mathcal{L}$, in the limit as $n \to \infty$, is distributed as noncentral χ^2 with c-1 degrees of freedom and the noncentrality parameter which simplifies to

$$\frac{(c-1)^2}{A} \sum_{i=1}^{c} s_i(\delta_i-\bar{\delta})^2 \{ \sum_{j=1}^{c} a_j \{ \binom{c-2}{j-2} \int_{-\infty}^{\infty} x f^2(x) [F(x)]^{j-2}$$

$$[1 -F(x)]^{c-j} dx - \binom{c-2}{j-1} \int_{-\infty}^{\infty} xf^2(x)[F(x)]^{j-1}[1-F(x)]^{c-j-1}dx\}$$

which can be seen to be equal to

$$\mu_{S_n} = \frac{(\underset{\sim}{a}'\underset{\sim}{d})^2}{A} \sum_{i=1}^{c} s_i(\delta_i - \bar{\delta})^2.$$

6. Asymptotic Relative Efficiency:

We know from Hannan's [7] and Andrews's [1] work that the asymptotic relative efficiency (ARE), in the Pitman sense, of one test with respect to another, is equal to the ratio of the noncentrality parameters of the two test statistics, provided that they are asymptotically distributed as noncentral χ^2 variates with the same degrees of freedom under the given sequence of alternative hypotheses (e.g. H_{L_n} or H_{S_n}). Hence to obtain a test statistic, from the class of statistics $\mathcal{L}$, which will have maximum ARE, we need to maximize the noncentrality parameter over all real $\underset{\sim}{a}$, for any given sequence of hypotheses.

We know that under H_{L_n} and H_{S_n}, $\mathcal{L}$ is, in the limit, distributed as noncentral χ^2 with c-1 degrees of freedom and noncentrality parameters given by μ_{L_n} and μ_{S_n} respectively. Let us first take μ_{L_n}

$$\mu_{L_n} = \frac{(\underset{\sim}{a}'\underset{\sim}{b})^2}{A} \Sigma s_i(\theta_i-\bar{\theta})^2$$

We need to maximize only $\frac{(\underset{\sim}{a}'\underset{\sim}{b})^2}{A}$ since the other factor does not involve a.

Let us define $\underset{\sim}{D} = (d_{ij})_{c \times c}$ $i,j, = 1,2,\ldots,c.$

where $$d_{ij} = \frac{\binom{c-1}{i-1}\binom{c-1}{j-1}}{\binom{2c-2}{i+j-2}(2c-1)} - \frac{1}{c^2}$$

and $\underset{\sim}{D}_0 = (d_{ij})_{c-1 \times c-1}$ $i, j = 1,2,\ldots, c-1$

It is then obvious that $A = \underset{\sim}{a}' \underset{\sim}{D} \underset{\sim}{a}$.

We see that $K_i\underset{\sim}{D}$, where K_i is a constant depending on i, is the covariance matrix of u_{ij}, $j = 1,2,\ldots,c$. As we know that $\sum_{j=1}^{c} u_{ij} = 1$, $\underset{\sim}{D}$ is singular.

Since this is the only one restriction on the u_{ij} ($j = 1,2,\ldots,c$). $\underset{\sim}{D}_0$ is nonsingular and of course, positive definite. We note that $\sum_{i=1}^{c} d_{ij} = \sum_{j=1}^{c} d_{ij} = \sum_{j=1}^{c} b_j = 0$. In view of these we may assume, without loss of generality that $\sum_{i=1}^{c} a_i = 0$, since the value of $\underset{\sim}{a}'\underset{\sim}{b}$ or $\underset{\sim}{a}'\underset{\sim}{D}\underset{\sim}{a}$ remains unchanged even if $\underset{\sim}{a}$ is replaced by $\underset{\sim}{a} - \bar{a}\underset{\sim}{J}$.

It may be further seen that $\underset{\sim}{a}'\underset{\sim}{b} = \underset{\sim}{a}'_0\underset{\sim}{b}_0$ and $\underset{\sim}{a}'\underset{\sim}{D}\underset{\sim}{a} = \underset{\sim}{a}'_0 \underset{\sim}{E} \underset{\sim}{a}_0$, where $\underset{\sim}{a}'_0 = (a_1,a_2,\ldots,a_{c-1})$, $\underset{\sim}{b}'_0 = (b_1-b_c,b_2-b_c,\ldots,b_{c-1}-b_c)$ and $\underset{\sim}{E} = (e_{ij})$ $i, j = 1,2,\ldots,c-1$, with

$$e_{ij} = (d_{ij} - d_{ic} - d_{cj} + d_{cc}).$$

Using the facts that $\sum_i d_{ij} = \sum_j d_{ij} = 0$, it is seen that

$$\underset{\sim}{E} = \underset{\sim}{T}\, \underset{\sim}{D}_0\, \underset{\sim}{T} \quad \text{where } T = (t_{ij}) \quad i,j = 1,2,\ldots,c-1 \text{ and}$$

$$t_{ii} = 2 \quad \text{and } t_{ij} = 1 \text{ if } i \neq j.$$

Therefore $\underset{\sim}{E}$ is positive definite.

Since now we have that

$$\frac{(\underset{\sim}{a}'\underset{\sim}{b})^2}{\underset{\sim}{a}'\underset{\sim}{D}\,\underset{\sim}{a}} = \frac{(\underset{\sim}{a}_0'\underset{\sim}{b}_0)^2}{\underset{\sim}{a}_0'\underset{\sim}{E}\,\underset{\sim}{a}_0}$$, using Cauchy's inequality it may be seen that

$$\frac{(\underset{\sim}{a}_0'\underset{\sim}{b}_0)^2}{\underset{\sim}{a}_0'\underset{\sim}{E}\,\underset{\sim}{a}_0} \leq \underset{\sim}{b}_0'\underset{\sim}{E}^{-1}\underset{\sim}{b}_0$$ for all real a and the equality is attained

whenever $\underset{\sim}{a}_0 \,\alpha\, \underset{\sim}{E}^{-1}\underset{\sim}{b}_0$. On similar lines it can be proved that

$$\frac{(\underset{\sim}{a}'\underset{\sim}{d})^2}{\underset{\sim}{a}'\underset{\sim}{D}\,\underset{\sim}{a}} \leq \underset{\sim}{d}_0'\underset{\sim}{E}^{-1}\underset{\sim}{d}_0 \quad \text{where } \underset{\sim}{d}_0' = (d_1 - d_c, \ldots d_{c-1} - d_c).$$

Hence we have proved the following theorem.

<u>Theorem 6.1</u>: (a) The maximum of μ_{L_n} for all real a is $\Sigma\, s_i(\theta_i - \bar{\theta})^2$. $(\underset{\sim}{b}_0'\underset{\sim}{E}^{-1}\underset{\sim}{b}_0)$ and is obtained when $\underset{\sim}{a}_0 \,\alpha\, \underset{\sim}{E}^{-1}\underset{\sim}{b}_0$ and $a_c = -\sum_{i=1}^{c-1} a_i$.

(b) The maximum of μ_{S_n} for all real a is $\Sigma\, s_i(\delta_i - \bar{\delta})^2(\underset{\sim}{d}'_0\underset{\sim}{E}^{-1}\underset{\sim}{d}_0)$ and is obtained when $\underset{\sim}{a}_0 \,\alpha\, \underset{\sim}{E}^{-1}\underset{\sim}{d}_0$ and $a_c = -\sum_{i=1}^{c-1} a_i$.

Here $\underset{\sim}{a}_0$, $\underset{\sim}{b}_0$, $\underset{\sim}{d}_0$ and $\underset{\sim}{E}$ are the same as defined earlier in this section.

Though $\underset{\sim}{E}^{-1}$ may be computed for any specified c, it has not been possible to compute it for general c in a simple and attractive form. Below it is given for c = 2,3.

For $c = 2$, $\underset{\sim}{E}$ is a scalar and $\underset{\sim}{E}^{-1} = 3$.

In this case $b_1 = \int_{-\infty}^{\infty} f^2(x)\,dx$ and $b_2 = -\int_{-\infty}^{\infty} f^2(x)\,dx$.

$$(6.2) \qquad \therefore \quad \underset{\sim}{b}_0' \underset{\sim}{E}^{-1} \underset{\sim}{b}_0 \sum_{i=1}^{2} s_i(\theta_i-\bar{\theta})^2 = 12\left[\int_{-\infty}^{\infty} f^2(x)\,dx\right]^2 \sum_{i=1}^{2} s_i(\theta_i-\bar{\theta})^2$$

is the maximum of μ_{L_n} under the sequence H_{L_n} of alternative hypotheses. Hence for $c = 2$, the maximum ARE amongst that of any test belonging to this class is equal to one w.r.t. the Wilcoxon-Mann-Whitney [14],[11] U test is Kruskal's H test [9] restricted to two samples. This ARE is attained for any a_1 and a_2 so long as they are not equal.

For $c = 2$, under H_{S_n}, $d_1 = -\int_{-\infty}^{\infty} x\,f^2(x)\,dx$ and $d_2 = \int_{-\infty}^{\infty} x\,f^2(x)\,dx$.

Hence the maximum of μ_{S_n} is

$$(6.3) \qquad 12\left[\int_{-\infty}^{\infty} x\,f^2(x)\,dx\right]^2 \sum_{i-1}^{2} s_i(\delta_i-\bar{\delta})^2$$

This too is the same as the corresponding expression given by the above mentioned two tests. The maximum is attained for any a_1 and a_2 as long as they are distinct.

For $c = 3$, we have

$$\underset{\sim}{E}^{-1} = \begin{bmatrix} 8 & -10 \\ -10 & 20 \end{bmatrix}$$

Under H_{L_n} we can write the maximum of μ_{L_n} as

$$4[2(b_1-b_3)^2 - 5(b_1-b_3)(b_2-b_3) + 5(b_2-b_3)^2] \sum_{i=1}^{3} s_i(\theta_i-\bar{\theta})^2$$

which is attained when

$$(6.5)\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \alpha \begin{pmatrix} 8b_1-10b_2+2b_3 \\ -10b_1+20b_2-10b_3 \\ 2b_1-10b_2+8b_3 \end{pmatrix} \quad \text{i.e.} \quad \alpha \begin{pmatrix} 4b_1-5b_2+b_3 \\ -5b_1+10b_2-5b_3 \\ b_1-5b_2+4b_3 \end{pmatrix}$$

For all distributions symmetric about zero, we have $b_1 = -b_3$ and $b_2 = 0$. Hence, in such case we have the maximum of μ_{L_n} as $12\, b_1^2 \sum_{i=1}^{3} s_i(\theta_i-\overline{\theta})^2$. b_1^2 may further be seen to be equal to $[\int_{-\infty}^{\infty} f^2(x)\, dx]^2$ for symmetric distributions. This maximum is attained for $a_1 = k$, $a_2 = 0$ and $a_3 = -k$ for any real k. This test, therefore, coincides with the L test [5] proposed by the author with $k = -1$, for three samples.

For exporential distribution, we have the maximum of μ_{L_n} as $8 \sum_{i=1}^{3} s_i(\theta_i-\overline{\theta})^2$ which is attained when $a_1 = 4k$, $a_2 = -5k$ and $a_3 = k$ for any real k. The ARE of this test with respect to Kruskal's H test for $c = 3$ and for exponential distribution is 2.66 for H_{L_n}.

For H_{S_n} we can obtain the a's which maximize μ_{S_n} and its maximum value by replacing b_i by d_i in (6.5) and (6.4).

We know, for distributions symmetric about zero, $d_1 = d_3$ · hence for such distributions, the maximum of μ_{S_n} is

$$(6.6) \qquad 720\, [\int_{-\infty}^{\infty} x\, f^2(x)\, F(x)\, dx]^2 \sum_{i=1}^{3} s_i(\delta_i-\overline{\delta})^2 \text{ which is attained}$$

for $a_1 = k$, $a_2 = -2k$ and $a_3 = k$ for any real k. In particular, for the normal distribution we have the above expression equal to $1.524 \sum_{i=1}^{3} s_i(\delta_i-\overline{\delta})^2$.

The ARE of this test w.r.t. a test proposed by Lehmann [10] (pp.273-275) is ·76 which is same as the ARE of the D test proposed by the author [6].

The maximum of μ_{S_n} for exponential distribution is seen to be

$(^8/9) \sum_{i-1}^{3} s_i(\delta_i - \bar{\delta})^2$ and is attained for $a_1 = -2k$, $a_2 = -5k$ and $a_3 = 7\ k$, for any real k. The ARE of this test w.r.t. Bhapkar's V test [2] (which is more efficient against H_{S_n} for exponential distribution than the D test) is 1·60.

7. Remarks:

It is noticed that for c = 3, the 'best' tests that we obtain out of this class for normal (or more generally, any symmetric continuous distribution) distribution are those which have been already proposed on heuristic grounds. It is conjectured that for c > 3 the 'best' tests may not coincide with existing tests. For nonsymmetric distributions (e.g. exponential) the 'best' test even for c = 3, is different from any of the previously proposed tests.

It is possible to take a different approach to construct tests. Bhapkar [3] has constructed a test for the c sample problem based on pairs (X_i, X_j) of observations where X_i and X_j are from different samples. Chatterjee [4] has proposed a test based on triplets (X_i, X_j, X_k) of observations for the same problem. Let us consider t-plets formed by t (< c) observations such that each of them represents a distinct sample. Let us define a function

(7.1) $\qquad \phi^t_{\alpha_i j}(x_{\alpha_1 s_{\alpha_1}}, \ldots, x_{\alpha_t s_{\alpha_t}}) = m_j$ whenever $x_{\alpha_i s_{\alpha_i}}$ is larger than

exactly (j-1) x's and smaller than the rest.

Here $(\alpha_1,\ldots,\alpha_t)$ are t members of $(1,2,\ldots,c)$. To assure symmetry between all the c samples we will have to consider a function based collectively on all the t-plets in which an observation from the α_i th sample occurs.

Let us consider

$$\Sigma' \Phi^t_{\alpha_i j}(x_{\alpha_1 s_{\alpha_1}},\ldots,x_{\alpha_t s_{\alpha_t}})$$ where Σ' indicates summation over all the Φ's in which $x_{\alpha_i s_{\alpha_i}}$ occurs and $(x_{\alpha_1 s_{\alpha_1}},\ldots,x_{\alpha_t s_{\alpha_t}})$ are from $(x_{1t_1}, x_{2t_2},\ldots,x_{ct_c})$. If $x_{i\alpha_i}$ is larger than exactly k-1 other x's in $(x_{1t}, x_{2t_2},\ldots,x_{ct_c})$ we shall have

$$\Sigma' \Phi^t_{\alpha_i,j}(x_{\alpha_1 s_{\alpha_1}},\ldots,x_{\alpha_t s_{\alpha_t}}) = \sum_{j=1}^{t} \binom{k-1}{j-1}\binom{c-k}{t-j} m_j$$

with the conventions that $\binom{0}{0} = 1$ and $\binom{x}{y} = 0$ whenever $y > x$.

Hence it seems likely that we shall develop identical tests if we base them on t-plets using the function defined in (7.1) or if we base them on c-plets using the function (7.2) defined below.

(7.2) $$\Phi_{ik}(x_{1t_1}\ x_{2t_2},\ldots,x_{ct_c}) = \sum_{j=1}^{t} \binom{k-1}{j-1}\binom{c-k}{t-j} m_j \text{ whenever}$$

x_{it_i} is larger than exactly k-1 x's

$= 0$ otherwise

8. Acknowledgement: I am grateful to Professor Wassily Hoeffding for useful discussions and constructive comments during this work.

References.

[1] Andrews, Fred C., (1954), "Asymptotic behaviour of some rank tests for analysis of variance," Ann. Math. Stat., 25, pp. 724-736.

[2] Bhapkar, V. P., (1961), "A nonparametric test for the problem of several samples," Ann. Math. Stat., 32, pp. 1108-1117.

[3] Bhapkar, V. P., (1964), "A nonparametric test for the several sample location problem," Uni. of North Carolina Institute of Statistics, Mimeo Series no. 411.

[4] Chatterjee, S. K., (1966), "A multisample nonparametric test based on U-statistics," Calcutta Stat. Ass. Bull., 15, pp. 109-119.

[5] Deshpande, Jayant V., (1965), "A nonparametric test based on U-statistics for the problem of several samples," J. Indian Stat. Ass., 3, pp. 20-29.

[6] Deshpande, Jayant V., (1965), "Some nonparametric tests of statistical hypotheses," Dissertation submitted for the Ph.D. degree of the University of Poona.

[7] Hannan, E. J., (1956), "The asymptotic powers of certain tests based on multiple correlation," J. Roy. Stat. Soc., Ser. B., 18, pp. 227-233.

[8] Hoeffding, Wassily, (1948), "A class of statistics with asymptotically normal distributions," Ann. Math. Stat, 19, pp. 293-325.

[9] Kruskal, W. H., (1952), "A nonparametric test for the several sample problem," Ann. Math. Stat., 23, pp. 525-540.

[10] Lehmann, E. L., (1959), "Testing of Statistical Hypotheses," John Wiley and Sons, New York.

[11] Mann, H. B. and Whitney, D. R., (1947), "On a certain test whether one of the two random variables is stochastically larger than the other," Ann. Math. Stat., 18, pp. 50-60.

[12] Sugiura, Nariaki, (1965), "Multisample and multivariate nonparametric tests based on U - statistics and their asymptotic efficiencies"; Osaka J. Math., 2, pp. 385-426.

[13] Sukhatme, B. V., (1958), "Testing the hypothesis that two populations differ only in location": Ann. Math. Stat., 29, pp. 60-78.

[14] Wilcoxon, F., (1945), "Individual comparisons by ranking methods": Biometrics Bulletin (now Biometrics), 1, pp. 80-83.