Mathematical Centre
Statistical Department
2de Boerhaavestraat 49
Amsterdam.

Head of Department: Prof.Dr D.van Dantzig,
Chief of Statistical Consultation:Prof.Dr J.Hemelrijk.

Report S 92 (VP 2)

<u>Prepublication</u>.

A non-parametric k sample test and its connection with the H-test.

by

T.J.Terpstra.

August 1952.

1. Introduction and summary.

In this paper a non-parametric k-sample test is given for the hypothesis H_0, that k independently distributed random variables $\underline{x}_1, \ldots, \underline{x}_k$ [1]) have the same continuous distribution function. In the test use is made of WILCOXON's statistic $\underline{\mathcal{U}}$, as defined by MANN and WHITNEY (cf. [1] and [3]) for comparing two samples. If $\underline{x}_{h,\xi}$ $(\xi \leqq n_h)$ are the observations taken from $\underline{x}_h$ [2]) and $\{\underline{\mathcal{U}}_{h,j}\}$ $(h<j)$ are the number of pairs (ξ,η) $(\xi \leqq n_h, \eta \leqq n_j)$ with $\underline{x}_{h,\xi} > x_{j,\eta}$, the test is based on the statistic

$$\underline{T}^2 = 12 \sum_{h<j} \frac{\tilde{\underline{\mathcal{U}}}_{h,j}^2}{n_h n_j} - \frac{12}{N+1} \sum_i \frac{\tilde{\underline{\mathcal{U}}}_i^2}{n_i}, \tag{1.1}$$

where

$$\tilde{\underline{\mathcal{U}}}_{h,j} = \underline{\mathcal{U}}_{h,j} - \tfrac{1}{2} n_h n_j ,$$

$$\tilde{\underline{\mathcal{U}}}_i = \sum_{h<i} \tilde{\underline{\mathcal{U}}}_{h,i} - \sum_{i<j} \tilde{\underline{\mathcal{U}}}_{i,j}, \tag{1.2}$$

and

$$N = \sum_i n_i .$$

By means of a recurrence relation for the simultaneous probability distribution of $\{\underline{\mathcal{U}}_{h,j}\}$, this distribution is shown to be asymptotically normal, from which it follows that $\underline{T}^2$ is asymptotically distributed as $\underline{\chi}^2$ with $\nu = \binom{k}{2} - 1$ degrees of freedom.

For large n_i the hypothesis H_0 will thus be rejected with a confidence $\geqq 1-\alpha$, if the observed $T^2 \geqq \chi^2_\alpha$, whereas χ^2_α is defined by

$$P[\underline{\chi}^2 \geqq \chi^2_\alpha] = \alpha .$$

Analogous to KRUSKAL and WALLIS [8], it may be expected that for small n_i the incomplete Γ-function and incomplete B-function are adequate approximations of the exact distribution of $\underline{T}^2$. At moment numerical calculationsare carried out to confirm this.

In the last section a connection is given between the T^2-test and the H-test, by means of which the statistic

$$\underline{H} = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{\tilde{\underline{\mathcal{U}}}_i^2}{n_i}, \tag{1.3}$$

---

1) Random variables will be distinguished from numbers (e.g. the value they take in an experiment) by underlining them.

2) If not explicitly mentioned h and j take the values $1, \ldots, k$, whereas ξ, η, when occurring as second suffices, run through the values $1, \ldots, n_h$ or $1, \ldots, n_j$, if h or j is the first suffix.

is shown to be asymptotically distributed as $\underline{\chi}^2$ with $\nu = k-1$ degrees of freedom [3]).

2. The simultaneous probability distribution of $\{\underline{u}_{h,j}\}$ and its moments.

Because of the continuity of the distribution function of the variables $\underline{x}_h$, all observations may be assumed to be different from each other and can thus be arranged in order of increasing magnitude.

If $T_{n_1,\dots,n_k}\{\underline{u}_{h,j}\}$ is the number of sequences $\{\underline{x}_{h,\xi}\}$, in which for each h and $j>h$ an $\underline{x}_{j,\eta}$ precedes an $\underline{x}_{h,\xi}$ $\underline{u}_{h,j}$ times, we obtain, by omitting the last observation in each of these sequences, the recurrence relation

$$T_{n_1,\dots,n_k}\{\underline{u}_{h,j}\} = \sum_{i=1}^{k} T_{n_1,\dots,n_i-1,\dots,n_k}\{\underline{u}_{h,j}^{(i)}\}, \tag{2.1}$$

where

$$\underline{u}_{h,j}^{(i)} = \underline{u}_{h,j} - \delta_h^i n_j \text{, if } h<j,\ \delta_h^i \tag{2.2}$$

being KRONECKER's symbol.

If H_0 is true, any of the $\frac{(n_1+\dots+n_k)!}{n_1!\dots n_k!}$ different sequences has equal probability, consequently

$$P_{n_1,\dots,n_k}\{\underline{u}_{h,j}\} = \sum_{i=1}^{k} \frac{n_i}{n_1+\dots+n_k} P_{n_1,\dots,n_i-1,\dots,n_k}\{\underline{u}_{h,j}^{(i)}\}, \tag{2.3}$$

with initial conditions

$P_{n_1,\dots,n_k}\{\underline{u}_{h,j}\} = 0$, if an $\underline{u}_{h,j}<0$ or an $n_i<0$,

and

$$P_{0,\dots,n_i,\dots,0}\{\underline{u}_{h,j}\} = \begin{cases} 0, & \text{if an } \underline{u}_{h,j} \neq 0 \\ 1, & \text{if all } \underline{u}_{h,j} = 0. \end{cases}$$

As (under H_0) $\mathcal{E}\underline{u}_{h,j} = \frac{1}{2}n_h n_j$ (cf. [3]), we obtain from (2.2) for each i, putting $\underline{\tilde{u}}_{h,j} = \underline{u}_{h,j} - \mathcal{E}\underline{u}_{h,j}$ and $\underline{\tilde{u}}_{h,j}^{(i)} = \underline{u}_{h,j}^{(i)} - \mathcal{E}\underline{u}_{h,j}^{(i)}$:

$$\underline{\tilde{u}}_{h,j} = \underline{\tilde{u}}_{h,j}^{(i)}, \text{ if } h,j \neq i,$$

$$\underline{\tilde{u}}_{i,j} = \underline{\tilde{u}}_{i,j}^{(i)} + \tfrac{1}{2}n_j,\quad i<j,$$

$$\underline{\tilde{u}}_{h,i} = \underline{\tilde{u}}_{h,i}^{(i)} - \tfrac{1}{2}n_h,\quad h<i,$$

Multiplying (2.3) by

$$\prod_{h<j} \underline{\tilde{u}}_{h,j}^{r_{h,j}} = \prod_{\substack{h<j\\ h,j\neq i}} \underline{\tilde{u}}_{h,j} \prod_{h<i} (\underline{\tilde{u}}_{h,i}^{(i)} - \tfrac{1}{2}n_h)^{r_{h,i}} \prod_{i<j} (\underline{u}_{i,j}^{(i)} + \tfrac{1}{2}n_j)^{r_{i,j}},$$

and using the binomial expansion, we obtain the following recurrence relation for the higher moments:

----

3) This theorem is independently proven by W.H.KRUSKAL and W.A.WALLIS (cf. [8]).

$$\mathcal{E}_{n_1,\dots,n_k} \prod_{h<j} \underline{\tilde{U}}_{h,j}^{\,\tau_{h,j}} = \sum_{i=1}^{k} \frac{n_i}{n} \sum_{\alpha_{1,i}=0}^{\tau_{1,i}} \cdots \sum_{\alpha_{i-1,i}=0}^{\tau_{i-1,i}} \; \sum_{\alpha_{i,i+1}=0}^{\tau_{i,i+1}} \cdots \sum_{\alpha_{i,k}=0}^{\tau_{i,k}} \left\{ \prod_{h<i} \binom{\tau_{h,i}}{\alpha_{h,i}} \left(\frac{-n_h}{2}\right)^{\tau_{h,i}-\alpha_{h,i}} \right\} .$$

$$\cdot \left\{ \prod_{i<j} \binom{\tau_{i,j}}{\alpha_{i,j}} \left(\frac{n_j}{2}\right)^{\tau_{i,j}-\alpha_{i,j}} \right\} . \mathcal{E}_{n_1,\dots,n_i-1,\dots,n_k} \prod_{h<i} \underline{\tilde{U}}_{h,i}^{\,\alpha_{h,i}} . \tag{2.4}$$

$$. \prod_{i<j} \underline{\tilde{U}}_{i,j}^{\,\alpha_{i,j}} \prod_{\substack{h<j \\ h,j \neq i}} \underline{\tilde{U}}_{h,j}^{\,\tau_{h,j}} .$$

3. The limit-distribution of $\{\underline{U}_{h,j}\}$.

Let $F(n_1, \dots, n_k)$ be a polynomial of integers $n_1, \dots, n_k$, then we define (cf. [4]) the operator:

$$\psi F(n_1,\dots,n_k) \overset{df.}{=} \sum_{i=1}^{k} n_i \{ F(n_1,\dots,n_i,\dots,n_k) - F(n_1,\dots,n_i-1,\dots,n_k) \} . \tag{3.1}$$

Then we have:

Lemma 1: If $F(n_1, \dots, n_k)$ is a polynomial in n_i and $\psi F(n_1, \dots, n_k)$ is a polynomial of degree λ_i in n_i for all i, then $F(n_1, \dots, n_k)$ is a polynomial of degree λ_i in n_i for all i.

Lemma 2: If $P_\lambda(n_1, \dots, n_k)$ and $Q_\lambda(n_1, \dots, n_k)$ are polynomials of degree λ in all n_i and if the variables n_i tend to ∞, so that

$$\lim \frac{\psi P_\lambda(n_1, \dots, n_k)}{Q_\lambda(n_1, \dots, n_k)} = c, \text{ then } \lim \frac{P_\lambda(n_1, \dots, n_k)}{Q_\lambda(n_1, \dots, n_k)} = \frac{c}{\lambda} .$$

Proof:

Putting

$$F(n_1,\dots,n_k) \overset{def}{=} \sum_{\alpha_1=0}^{\lambda_1} \cdots \sum_{\alpha_k=0}^{\lambda_k} A_{\alpha_1,\dots,\alpha_k} n_1^{\alpha_1} \cdots n_k^{\alpha_k} ,$$

it follows from definition (3.1):

$$\psi F(n_1,\dots,n_k) = \sum_{i=1}^{k} n_i \sum_{\alpha_1=0}^{\lambda_1} \cdots \sum_{\alpha_k=0}^{\lambda_k} A_{\alpha_1,\dots,\alpha_k} n_1^{\alpha_1} \cdots n_{i-1}^{\alpha_{i-1}} n_{i+1}^{\alpha_{i+1}} \cdots n_k^{\alpha_k} .$$

$$. \sum_{\alpha_i'=0}^{\alpha_i-1} (-1)^{\alpha_i-\alpha_i'-1} \binom{\alpha_i}{\alpha_i'} n_i^{\alpha_i'} .$$

$$= \sum_{\beta_1=0}^{\lambda_1} \cdots \sum_{\beta_k=0}^{\lambda_k} B_{\beta_1,\dots,\beta_k} n_1^{\beta_1} \cdots n_k^{\beta_k} ,$$

where $B_{\lambda_1,\dots,\lambda_k} = \sum_{i=1}^{k} \lambda_i A_{\lambda_1,\dots,\lambda_k} = \lambda A_{\lambda_1,\dots,\lambda_k}$.

Proof of lemma 1 and 2:

From $B_{\lambda_1,\dots,\lambda_k} \neq 0$ it follows $A_{\lambda_1,\dots,\lambda_k} \neq 0$.

Lemma 3: All even moments $\mathcal{E}_{n_1,\dots,n_k} \prod_{h<j} \underline{\tilde{U}}_{h,j}^{\,\tau_{h,j}}$, with

$\sum_{h<j} \tau_{h,j} = 2R \; (R=0,1,2,\dots)$, are of degree $\leq 3R$ in all n_i [4]).

---

4) Because of the symmetry of the distribution, all odd moments are zero.

From definition (3.1) and (2.4) we obtain

$$\psi \mathcal{E}_{n_1,\dots,n_k} \prod_{h<j} \underline{\tilde{U}}_{h,j}^{\tau_{h,j}} = \sum_{i=1}^{k} n_i \underset{\left(\sum\limits_{h<i}\alpha_{h,i}+\sum\limits_{i<j}\alpha_{i,j}<\sum\limits_{h<i}\tau_{h,i}+\sum\limits_{i<j}\tau_{i,j}\right)}{\sum_{\alpha_{1,i}=0}^{\tau_{1,i}} \cdots \sum_{\alpha_{i-1,i}=0}^{\tau_{i-1,i}} \sum_{\alpha_{i,i+1}=0}^{\tau_{i,i+1}} \cdots \sum_{\alpha_{i,k}=0}^{\tau_{i,k}}} \left\{ \prod_{h<i} \binom{\tau_{h,i}}{\alpha_{h,i}} \left(\frac{-n_h}{2}\right)^{\tau_{h,i}-\alpha_{h,i}} \right\} \cdot \tag{3.2}$$

$$\cdot \left\{ \prod_{i<j} \binom{\tau_{i,j}}{\alpha_{i,j}} \left(\frac{n_j}{2}\right)^{\tau_{i,j}-\alpha_{i,j}} \right\} \cdot \mathcal{E}_{n_1,\dots,n_i-1,\dots,n_k} \prod_{h<i} \underline{\tilde{U}}_{h,i}^{\alpha_{h,i}} \cdot \prod_{i<j} \underline{\tilde{U}}_{i,j}^{\alpha_{i,j}} \cdot \prod_{h<j} \underline{\tilde{U}}_{h,j}^{\tau_{h,j}}$$

For $R=1$, the moments of second order are of degree 3 in all n_i. In this case we have (cf. [1], [3] and [4])

$$\mathcal{E}_{n_1,\dots,n_k} \underline{\tilde{U}}_{h,j}^{2} = \tfrac{1}{12} n_h n_j (n_h+n_j+1),$$

$\mathcal{E}_{n_1,\dots,n_k} \underline{\tilde{U}}_{h,j} \cdot \underline{\tilde{U}}_{l,m} = 0$, if h, j, l and m are different from each other,

$$\mathcal{E}_{n_1,\dots,n_k} \underline{\tilde{U}}_{h',i} \cdot \underline{\tilde{U}}_{h,i} = \tfrac{1}{12} n_{h'} n_h n_i, \tag{3.3}$$

$$\mathcal{E}_{n_1,\dots,n_k} \underline{\tilde{U}}_{i,j} \cdot \underline{\tilde{U}}_{i,j'} = \tfrac{1}{12} n_i n_j n_{j'},$$

$$\mathcal{E}_{n_1,\dots,n_k} \underline{\tilde{U}}_{h,i} \cdot \underline{\tilde{U}}_{i,j} = -\tfrac{1}{12} n_h n_i n_j .$$

If we assume all moments of order $2R<2R_0$ to be of degree $\leqq 3R$ in all n_i, it follows from (3.2) and Lemma 1, that the moments of order $2R_0$ are of degree $\leqq 3R_0$.

Putting

$$\underline{W} = \prod_{h<j} \underline{\tilde{U}}_{h,j}^{\tau_{h,j}},$$

we **thus** obtain

$$\psi \mathcal{E}_{n_1,\dots,n_k} \underline{W} = \sum_{i=1}^{k} n_i \Big\{ \sum_{h<i} \binom{\tau_{h,i}}{\tau_{h,i}-2} \left(\frac{-n_h}{2}\right)^2 \mathcal{E}_{n_1,\dots,n_i-1,\dots,n_k} \underline{W} \cdot \underline{\tilde{U}}_{h,i}^{-2}$$

$$+ \sum_{h'<h<i} \binom{\tau_{h',i}}{\tau_{h',i}-1} \binom{\tau_{h,i}}{\tau_{h,i}-1} \frac{n_{h'} n_h}{4} \mathcal{E}_{n_1,\dots,n_i-1,\dots,n_k} \underline{W} \cdot \underline{\tilde{U}}_{h',i}^{-1} \cdot \underline{\tilde{U}}_{h,i}^{-1}$$

$$+ \sum_{i<j} \binom{\tau_{i,j}}{\tau_{i,j}-2} \left(\frac{n_j}{2}\right)^2 \mathcal{E}_{n_1,\dots,n_i-1,\dots,n_k} \underline{W} \cdot \underline{\tilde{U}}_{i,j}^{-2}$$

$$+ \sum_{i<j<j'} \binom{\tau_{i,j}}{\tau_{i,j}-1} \binom{\tau_{i,j'}}{\tau_{i,j'}-1} \frac{n_j n_{j'}}{4} \mathcal{E}_{n_1,\dots,n_i-1,\dots,n_k} \underline{W} \cdot \underline{\tilde{U}}_{i,j}^{-1} \cdot \underline{\tilde{U}}_{i,j'}^{-1} \tag{3.4}$$

$$-\sum_{h<i<j}\binom{\tau_{h,i}}{\tau_{h,i}-1}\binom{\tau_{i,j}}{\tau_{i,j}-1}\frac{n_h n_j}{4}\,\mathcal{E}_{n_1,\dots,n_i-1,\dots,n_k}\,\underline{W}\cdot\underline{\tilde{U}}_{h,i}^{-1}\cdot\underline{\tilde{U}}_{i,j}^{-1}\Big\} \tag{3.4}$$

$$+\,P_{3R-1}(n_1,\dots,n_k),$$

where $P_{3R-1}(n_1,\dots,n_k)$ is a polynomial of degree $3R-1$ in $n_1,\dots,n_k$.

By reduction of (3.4) we obtain

$$\begin{aligned}\psi\,\mathcal{E}_{n_1,\dots,n_k}\underline{W} = \sum_{i=1}^{k}\Big\{&\frac{1}{8}\sum_{h<i} n_h n_i(n_h+n_i)\tau_{h,i}(\tau_{h,i}-1)\,\mathcal{E}_{n_1,\dots,n_k}\underline{W}\cdot\underline{\tilde{U}}_{h,i}^{-2}\\ &+\frac{1}{4}\sum_{h'<h<i} n_{h'}n_h n_i\cdot\tau_{h',i}\cdot\tau_{h,i}\,\mathcal{E}_{n_1,\dots,n_k}\underline{W}\cdot\underline{\tilde{U}}_{h',i}^{-1}\cdot\underline{\tilde{U}}_{h,i}^{-1}\\ &+\frac{1}{4}\sum_{i<j<j'} n_i n_j n_{j'}\cdot\tau_{i,j}\cdot\tau_{i,j'}\cdot\mathcal{E}_{n_1,\dots,n_k}\underline{W}\cdot\underline{\tilde{U}}_{i,j}^{-1}\cdot\underline{\tilde{U}}_{i,j'}^{-1}\\ &-\frac{1}{4}\sum_{h<i<j} n_h n_i n_j\cdot\tau_{h,i}\cdot\tau_{i,j}\cdot\mathcal{E}_{n_1,\dots,n_k}\underline{W}\cdot\underline{\tilde{U}}_{h,i}^{-1}\cdot\underline{\tilde{U}}_{i,j}^{-1}\Big\}\\ &+P_{3R-1}(n_1,\dots,n_k).\end{aligned} \tag{3.5}$$

We now define

$$\sigma_{h,j}\overset{df.}{=}\sqrt{\mathcal{E}_{n_1,\dots,n_k}\underline{\tilde{U}}_{h,j}^{2}} = \sqrt{\tfrac{1}{12}n_h n_j(n_h+n_j+1)}\ ,$$

$$\lambda_{n_1,\dots,n_k}^{\{\tau_{h,j}\}}\overset{df.}{=}\mathcal{E}_{n_1,\dots,n_k}\prod_{h<j}\left(\frac{\underline{\tilde{U}}_{h,j}}{\sigma_{h,j}}\right)^{\tau_{h,j}}, \tag{3.6}$$

$$\lambda_{n_1,\dots,n_k}^{\tau_{h',i}-1,\,\tau_{h,i}-1}\overset{df.}{=}\mathcal{E}_{n_1,\dots,n_k}\Big\{\prod_{h<j}\left(\frac{\underline{\tilde{U}}_{h,j}}{\sigma_{h,j}}\right)^{\tau_{h,j}}\Big\}\cdot\left(\frac{\underline{\tilde{U}}_{h',i}}{\sigma_{h',i}}\right)^{-1}\left(\frac{\underline{\tilde{U}}_{h,i}}{\sigma_{h,i}}\right)^{-1},\ \text{etc.}$$

For $n_1,\dots,n_k$ tending to ∞ in such a way that

$$\lim\mathcal{E}_{n_1,\dots,n_k}\frac{\underline{\tilde{U}}_{h',i}}{\sigma_{h',i}}\cdot\frac{\underline{\tilde{U}}_{h,i}}{\sigma_{h,i}} = \lim\sqrt{\frac{n_{h'}n_h}{(n_{h'}+n_i+1)(n_h+n_i+1)}} = \varrho_{h',i;\,h,i}\ ,\ \text{etc.} \tag{3.7}$$

and denoting the limiting values of $\lambda_{n_1,\dots,n_k}^{\{\tau_{h,j}\}}$ and $\lambda_{n_1,\dots,n_k}^{\tau_{h',i}-1,\,\tau_{h,i}-1}$, etc., by $\lambda^{\{\tau_{h,j}\}}$ resp. $\lambda^{\tau_{h',i}-1,\,\tau_{h,i}-1}$, etc., it follows from (3.5)

$$\lim \frac{\varphi\,\mathcal{E}_{n_1,\dots,n_k} \prod_{h<j} \underline{\tilde{u}}_{h,j}^{\tau_{h,j}}}{\prod_{h<j} \sigma_{h,j}^{\tau_{h,j}}} = \frac{3}{2}\sum_{i=1}^{k}\Big\{\sum_{h<i} \tau_{h,i}(\tau_{h,i}-1)\,\lambda^{\tau_{h,i}-2}$$

$$+\sum_{h'<h<i} \tau_{h',i}\cdot\tau_{h,i}\cdot 2\varrho_{h',i;h,i}\cdot\lambda^{\tau_{h',i}-1,\tau_{h,i}-1}+\sum_{i<j<j'} \tau_{i,j}\cdot\tau_{i,j'}\cdot 2\varrho_{i,j;i,j'}\cdot\lambda^{\tau_{i,j}-1,\tau_{i,j'}-1} \qquad (3.8)$$

$$-\sum_{h<i<j} \tau_{h,i}\cdot\tau_{i,j}\cdot 2\varrho_{h,i;i,j}\cdot\lambda^{\tau_{h,i}-1,\tau_{i,j}-1}\Big\}.$$

It can now be proven by induction that the moments $\lambda^{\{\tau_{h,j}\}}$ are identical with the moments $\lambda'^{\{\tau_{h,j}\}}$ of a multinormal distribution of variables $\{\underline{u}'_{h,j}\}$, each with mean 0 and variance 1, and correlation-coefficients ϱ as defined by (3.7). The moment-generating function φ of this distribution is given by (cf. [2]).

$$\varphi(t_{1,2},\dots,t_{k-1,k}) = \exp\frac{1}{2}\Big\{\sum_{h<j} t_{h,j}^2 + 2\sum_{h'<h<i} \varrho_{h',i;h,i}\, t_{h',i}\, t_{h,i} +$$

$$+ 2\sum_{i<j<j'} \varrho_{i,j;i,j'}\, t_{i,j}\, t_{i,j'} - 2\sum_{h<i<j} \varrho_{h,i;i,j}\, t_{h,i}\, t_{i,j}\Big\} \qquad (3.8^a)$$

From this expression and its definition

$$\varphi(t_{1,2},\dots,t_{k-1,k}) = \mathcal{E}\, e^{\prod_{h<j} \underline{u}'_{h,j}\cdot t_{h,j}},$$

we obtain

$$\lambda'^{\{\tau_{h,j}\}} = 2^{-R}\prod_{h<j}\tau_{h,j}!\sum_{\{a\}} \frac{\prod_{h'<h<i}(2\varrho_{h',i;h,i})^{a_{h',i;h,i}} \prod_{i<j<j'}(2\varrho_{i,j;i,j'})^{a_{i,j;i,j'}} \prod_{h<i<j}(2\varrho_{h,i;i,j})^{a_{h,i;i,j}}}{\prod_{h<j} a_{h,j}!\prod_{h'<h<i} a_{h',i;h,i}!\prod_{i<j<j'} a_{i,j;i,j'}!\prod_{h<i<j} a_{h,i;i,j}!} \qquad (3.9)$$

where the summation is performed over all $a_{h,j}$ etc., satisfying the relations

$$\tau_{h,j} = 2a_{h,j} + \sum_{h'<h}(a_{h',h;h,j} + a_{h',j;h,j}) + \sum_{h<i<j}(a_{h,i;h,j} + a_{h,j;i,j})$$

$$+\sum_{j<j'}(a_{h,j;h,j'} + a_{h,j;j,j'}), \qquad (h<j).$$

For $R=1$, all moments $\lambda^{\{\tau_{h,j}\}}$ of second order satisfy (3.9). If we assume this to hold for $R<R_0$ it follows from (3.8) that

$$\lim \frac{\psi \mathcal{E}_{n_1,\ldots,n_k} \prod_{h<j} \tilde{\underline{U}}_{h,j}^{\tau_{h,j}}}{\prod_{h<j} \sigma_{h,j}^{\tau_{h,j}}} = 3R\lambda'^{\{\tau_{h,j}\}} ,$$

consequently, according to Lemma 2,

$$\lambda^{\{\tau_{h,j}\}} = \lim \mathcal{E}_{n_1,\ldots,n_k} \prod_{h<j} \left(\frac{\tilde{\underline{U}}_{h,j}}{\sigma_{h,j}}\right)^{\tau_{h,j}} = \lambda'^{\{\tau_{h,j}\}}, \quad \text{q.e.d.}$$

We then have

<u>Theorem I</u>: The distribution of the set of variables $\{\tilde{\underline{U}}_{h,j}\}$ is asymptotically equivalent with a multinormal distribution with covariance-matrix given by (3.3).

Denoting the covariance-matrix of the set $\{\tilde{\underline{U}}_{h,j}\}$ by $\| \sigma_{h,j;\,l,m} \|$, it follows from (3.3)

$$| \sigma_{h,j;\,l,m} | = 12^{-\binom{k}{2}} \{(n_1 \ldots n_k)(n_1 + \ldots + n_k + 1)\}^{k-1},$$

and the inverse matrix $| \sigma^{h,j;\,l,m} |$ is given by

$$\sigma^{h,j;\,h,j} = \frac{12(N+1-n_h-n_i)}{n_h n_i (N+1)},$$

$\sigma^{h,j;\,l,m} = 0$, if h, j, l and m are different from each other,

$$\sigma^{h,i;\,h,i} = \frac{-12}{n_i(N+1)},$$

$$\sigma^{i,j;\,i,j'} = \frac{-12}{n_i(N+1)},$$

$$\sigma^{h,i;\,i,j} = \frac{12}{n_i(N+1)},$$

where $N = \sum_i n_i$.

Because of the asymptotic normality of the set $\{\tilde{\underline{U}}_{h,j}\}$, we can state

<u>Theorem II</u>: The distribution of the variable T^2, defined by :

$$\underline{T}^2 = \sum_{h<j} \sigma^{h,j;h,j}\, \underline{\tilde{U}}_{h,j}^2 + 2\sum_{h'<h<i} \sigma^{h',i;h,i}\, \underline{\tilde{U}}_{h',i}\cdot \underline{\tilde{U}}_{h,i} + 2\sum_{i<j<j'} \sigma^{i,j;i,j'}\, \underline{\tilde{U}}_{i,j}\cdot \underline{\tilde{U}}_{i,j'}$$

$$+ 2\sum_{h<i<j} \sigma^{h,i;i,j}\, \underline{\tilde{U}}_{h,i}\cdot \underline{\tilde{U}}_{i,j}$$

$$= 12\sum_{h<j} \frac{\underline{\tilde{U}}_{h,j}^2}{n_h n_j} - \frac{12}{N+1}\sum_{i} \frac{\underline{\tilde{U}}_i^2}{n_i},$$

where $\underline{\tilde{U}}_i = \sum_{h<i} \underline{\tilde{U}}_{h,i} - \sum_{i<j} \underline{\tilde{U}}_{i,j}$,

and $N = \sum_i n_i$,

is asymptotically equivalent with a X^2-distribution with $\nu = \binom{k}{2}$ degrees of freedom.

4. Some theorems about the H-test.

4.1. Introduction.

Instead of comparing each pair of samples with each other as in the foregoing test, by application of the H-test each sample is compared with all other samples together by means of WILCOXON's statistic. For the sample $\underline{x}_{i,\xi}$ $(\xi \leq n_i)$, taken from $\underline{x}_i$, this statistic (denoted by $\underline{U}_i$) is equal to the number of pairs (ξ,η) $(\xi \leq n_i, \eta \leq n_j, j \neq i)$ with $\underline{x}_{i,\xi} > \underline{x}_{j,\eta}$.
In his paper RIJKOORT [7] conjectured that the distribution of the variable

$$\underline{H}^* = \frac{12(k-1)}{(N+1)(N^2 - \sum n_i^2)} \sum_{i=1}^{k} \underline{\tilde{U}}_i^2 \tag{4.1.1}$$

is asymptotically equivalent with a X^2-distribution with $\nu = k-1$ degrees of freedom.
In the following it will be shown that the simultaneous distribution of the set $\{\underline{U}_i\}$ is asymptotically normal under H_0, from which it follows, that under H_0 the statistics H and H^*, defined by (1.3) and (4.1.1) are asymptotically distributed as $\underline{X}^2$, with $\nu = k-1$ degrees of freedom.

4.2. The simultaneous probability-distribution of $\{\underline{U}_i\}$.

In the same way as in section 2 the following recurrence relation is obtained for the simultaneous probability distribution of $\{\underline{U}_i\}$:

$$P_{n_1,\dots,n_k}\{\underline{\mathcal{U}}_i\} = \sum_{h=1}^{k} \frac{n_h}{N} P_{n_1,\dots,n_h-1,\dots,n_k}\{\underline{\mathcal{U}}_i^{(h)}\},$$

where $\underline{\mathcal{U}}_i^{(h)} = \begin{cases} \underline{\mathcal{U}}_i, \text{ if } h \neq i, \\ \underline{\mathcal{U}}_i - (N-n_i), \text{ if } h = i. \end{cases}$,

with initial conditions

$$P_{n_1,\dots,n_k}\{\underline{\mathcal{U}}_i\} = 0, \text{ if an } \underline{\mathcal{U}}_i < 0 \text{ or an } n_i < 0,$$

$$P_{0,\dots,n_h,\dots,0}\{\underline{\mathcal{U}}_i\} = \begin{cases} 0, \text{ if an } \underline{\mathcal{U}}_i \neq 0, \\ 1, \text{ if all } \underline{\mathcal{U}}_i = 0. \end{cases}$$

4.3. The asymptotic distribution of $\{\underline{\mathcal{U}}_i\}$.

If $\underline{\mathcal{U}}_{\{h\}+\{j\}}$ denotes WILCOXON's statistic, obtained by comparing the samples taken from $\underline{x}_h$ and $\underline{x}_j$ together with all other samples together, we have

$$\underline{\mathcal{U}}_{\{h\}+\{j\}} = \underline{\mathcal{U}}_h + \underline{\mathcal{U}}_j - n_h n_j . \tag{4.3.1}$$

As

$$Var\ \underline{\mathcal{U}}_i = \frac{1}{12} n_i (N-n_i)(N+1), \tag{4.3.2}$$

and

$$Var\ \underline{\mathcal{U}}_{\{h\}+\{j\}} = \frac{1}{12}(n_h+n_j)(N-n_h-n_j)(N+1),\quad h \neq j,$$

it follows from (4.3.1) and (4.3.2)

$$Cov(\underline{\mathcal{U}}_h, \underline{\mathcal{U}}_j) = -\frac{1}{12} n_h n_j (N+1),\quad h \neq j.$$

As $\sum_i \underline{\mathcal{U}}_i = \sum_{h<j} n_h n_j$, we need only consider the simultaneous distribution of $k-1$ variables $\underline{\mathcal{U}}_i$, e.g. $\underline{\mathcal{U}}_1, \dots, \underline{\mathcal{U}}_{k-1}$.
Denoting the reduced variables $\underline{\mathcal{U}}_i$ by $\tilde{\underline{\mathcal{U}}}_i$, we have

$$\tilde{\underline{\mathcal{U}}}_i = \underline{\mathcal{U}}_i - \frac{1}{2} n_i (N-n_i)$$

and

$$\tilde{\underline{\mathcal{U}}}_i = -\sum_{h<i} \tilde{\underline{\mathcal{U}}}_{h,i} + \sum_{i<j} \tilde{\underline{\mathcal{U}}}_{i,j} . \tag{4.3.3}$$

The asymptotic momenting generating function of the reduced variables $\{\tilde{\underline{\mathcal{U}}}_i\}$ thus follows from the asymptotic momenting generating function of the variables $\{\tilde{\underline{\mathcal{U}}}_{h,j}\}$, which is given by (3.8a).
We then obtain

$$\varphi(t_1,\dots,t_{k-1}) = exp\, \frac{1}{2}\Big\{\sum_{i=1}^{k-1} \sigma_{i,i} t_i^2 + 2\sum_{h<j} \sigma_{h,j} t_h t_j\Big\}, \tag{4.3.4}$$

where the covariance-matrix $\| \sigma_{h,j} \|$ is given by

$$\sigma_{i,i} = \frac{1}{12} n_i(N-n_i)(N+1),$$

(4.3.5)

$$\sigma_{h,j} = -\frac{1}{12} n_h n_j (N+1), \quad (h \neq j).$$

The determinant of this matrix is given by

$$D = |\sigma_{h,j}| = \left\{ \frac{1}{12}(N+1) \right\}^{k-1} n_1 \dots n_k N^{k-2},$$

and the minors $\Delta_{i,i}$ and $\Delta_{h,j}$ by

$$\Delta_{i,i} = \left\{ \frac{1}{12}(N+1) \right\}^{k-2} n_1 \dots n_{i-1} n_{i+1} \dots n_{k-1} (n_k + n_i) N^{k-3},$$

(4.3.6)

$$\Delta_{h,j} = \left\{ \frac{1}{12}(N+1) \right\}^{k-2} n_1 \dots n_{k-1} N^{k-3}.$$

If all $n_i \neq 0$, $\|\sigma_{h,j}\|$ is thus positive definite, from which it follows that (4.3.4) is the momenting generating function of a multi-normal distribution (cf. [2]). So we can state the following

Theorem III: The distribution of the set of variables $\{\tilde{\underline{U}}_i\}$ is asymptotically equivalent with a multinormal distribution with covariance-matrix given by (4.3.5).

Denoting the inverse matrix of $\|\sigma_{h,j}\|$ by $\|\sigma^{h,j}\|$, we have

$$\sigma^{i,i} = 12 \frac{n_i + n_k}{n_i n_k N(N+1)},$$

and

$$\sigma^{h,j} = \frac{12}{n_k N(N+1)}.$$

Because of the asymptotic normality of the simultaneous distribution of the variables $\{\tilde{\underline{U}}_i\}$, we can state

Theorem IV: The distribution of the variable $\underline{U}^2$, defined by

$$\underline{U}^2 = \sum_{i=1}^{k-1} \sigma^{i,i} \tilde{\underline{U}}_i^2 + 2 \sum_{\substack{h<j \\ j=2}}^{k-1} \sigma^{h,j} \tilde{\underline{U}}_h \tilde{\underline{U}}_j$$

$$= \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{\tilde{\underline{U}}_i^2}{n_i},$$

is asymptotically equivalent with a χ^2-distribution with $\nu = k-1$ degrees of freedom .