

A MULTIVARIATE TEST FOR TIME TREND WITH  
PHARMACEUTICAL APPLICATIONS

by

E. Jacquelin Dietz and Timothy J. Killeen

Department of Statistics  
North Carolina State University

Institute of Statistics Mimeo Series #1195

September, 1978

A Multivariate Test for Time Trend with  
Pharmaceutical Applications

by

E. Jacquelin Dietz and Timothy J. Killeen

During a drug trial, each subject may have many blood constituents measured at regular time intervals. The traditional method of evaluating such data using "normal ranges" has undergone much controversy in the clinical chemistry literature. We propose a new multivariate test for time trend based on Kendall's  $\tau$  statistic as a way to test for change in blood constituents over the course of a drug experiment. The new test avoids the unrealistic assumptions and repeated testing problem of the normal range method.

KEY WORDS: Time trend; Multivariate trend test; Nonparametric tests; Normal range.

E. Jacquelin Dietz is Assistant Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27650. Timothy J. Killeen is Associate Professor, Department of Statistics, University of Connecticut, Storrs, CT 06268. The work of the first author was supported by a doctoral dissertation fellowship from the Pfizer Pharmaceutical Company. She wishes to thank Dr. David S. Salsburg of Pfizer, who provided much of the motivation for this research.

## 1. INTRODUCTION

During many drug trials, each subject's condition is monitored by weekly or monthly blood tests. As many as 20 or 30 blood constituents may be measured at each of 12 or 15 time points. Such data is traditionally evaluated by the normal range method, which compares a subject's value for a constituent to values from a reference population of healthy people. The normal range for a variable is a tolerance interval  $\bar{X} \pm ks$ , where  $\bar{X}$  and  $s$  are the mean and standard deviation of a sample from the reference population. The constant  $k$  depends upon the sample size  $n$ , the population proportion  $p$  to be covered by the interval, and the desired confidence. Values of  $k$  are calculated under the assumption that each variable has a Gaussian distribution in the reference population. In practice,  $k$  is often set equal to two, the limiting value when  $p$  equals .95 and  $n \rightarrow \infty$ . In the drug trial, each subject's value for each constituent at each time point is compared to the corresponding normal range.

The normal range concept has undergone much controversy in the clinical chemistry literature. Problems associated with the use of a reference population include the medical dilemma of defining health and the practical problem of obtaining blood chemistry data from large numbers of people (Healy 1969). Many blood constituents show age and sex variation, necessitating separate normal ranges for different segments of the population and compounding the problem of finding healthy subjects (Harris 1974). Another objection to the traditional normal range concerns the assumption that the variables are normally distributed. Many authors (Elveback, Guillier, and Keating 1970; Mainland 1971) argue that many such variables have skewed distributions.

Performing a univariate test on each variable at each time point results in many false positives (Schoen and Brooks 1970). Since each value falling outside the normal range requires further examination, much time and effort can be expended on such values. Finally, the normal range method ignores an important hypothesis in a drug trial. Presumably the experimenter wishes to know whether the drug has had an effect, and thus whether blood constituents have changed over the trial. Harris (1975) and Nadel et al. (1969) suggest comparing an individual's blood chemistry values to that person's previous values, instead of to a reference population.

Alternative procedures for analyzing blood chemistry data have been proposed. Multivariate normal tolerance regions reduce the amount of repeated testing, but assume multivariate normality and employ a reference population (Winkel, Lyngbye, and Jørgensen 1972). "Normal ranges" based on nonparametric percentile estimates avoid the normality assumption, but involve repeated testing and a reference population (Herrera 1958).

We propose a new multivariate test for time trend that is applicable to blood chemistry data in a drug trial. This test assumes only the continuity of the variables. Since the procedure is multivariate and incorporates all the time points, the repeated testing problem is eliminated. The procedure tests for change in the variables over the course of the experiment, and by comparing an individual's values only to his past values, avoids the problems associated with the use of a reference population.

## 2. UNIVARIATE TREND TEST

We first describe a univariate test for time trend due to Mann (1945). Let  $Y_1, Y_2, \dots, Y_n$  be a sequence of continuous observations, ordered over time. We wish to test  $H_0$ , the hypothesis that the observations are randomly ordered, versus the alternative of a monotone trend over time. The null hypothesis implies that all  $n!$  time orders of the observations are equally likely. Let

$$c_{ij} = +1, \text{ if } Y_i < Y_j, \\ = -1, \text{ if } Y_i > Y_j.$$

Then, under  $H_0$ , the test statistic  $K_Y = \sum_{i < j} c_{ij}$  has mean 0 and variance  $\sigma^2 = n(n-1)(2n+5)/18$ .  $K_Y/\sigma$  is asymptotically  $N(0,1)$ ; its exact null distribution can be found by enumeration for small sample sizes. Note that the Mann statistic depends only on the ranks of the observations.

It is well known that Mann's test is a special case of Kendall's  $\tau$  test for correlation (Kendall 1970). In general, Kendall's procedure is used for bivariate observations  $(U_i, V_i)$ ,  $i=1,2,\dots,n$ , to test  $H_0: \tau = 0$  versus  $H_1: \tau \neq 0$ , where

$$\tau = P\{(V_j - V_i)(U_j - U_i) > 0\} - P\{(V_j - V_i)(U_j - U_i) < 0\}, \quad i < j.$$

Kendall's test is based on the statistic

$$K_{UV} = \sum_{i < j} d_{ij} \tag{2.1}$$

where

$$\begin{aligned} d_{ij} &= +1, \text{ if } (V_j - V_i)(U_j - U_i) > 0, \\ &= -1, \text{ if } (V_j - V_i)(U_j - U_i) < 0. \end{aligned}$$

For the Mann test,  $V_i = i, i=1,2,\dots,n$ .

### 3. BIVARIATE TREND TEST

We now extend the Mann test to the bivariate case. Let  $(X_i, Y_i), i=1,2,\dots,n$ , be a sequence of continuous bivariate observations, ordered over time. We wish to test  $H_0$ , the hypothesis that the observations are randomly ordered, versus the alternative of a monotone time trend in one or both variables. Like the univariate test, the bivariate procedure depends only on the ranks of the observations. Let  $R_i$  be the rank of  $X_i$  among the  $X$ 's and  $S_i$  the rank of  $Y_i$  among the  $Y$ 's.

The bivariate test statistic is a function of two univariate Mann statistics,  $K_X$  and  $K_Y$ , for the  $X$ 's and  $Y$ 's respectively. Under  $H_0$ ,  $K_X/\sigma$  and  $K_Y/\sigma$ , where  $\sigma^2 = n(n-1)(2n+5)/18$ , are each asymptotically  $N(0,1)$ ; they are in general dependent, the relationship between them depending upon the underlying bivariate distribution of  $X$  and  $Y$ . The following theorem is proved in Appendix A.

Theorem 1: Let  $\rho_K$  denote the conditional null correlation between  $K_X$  and  $K_Y$ , given the  $(R_i, S_i), i=1,2,\dots,n$ . Then

$$\rho_K = \{6K_{XY} + 24 \sum_{i=1}^n R_i S_i - 6n(n+1)^2\} / \{n(n-1)(2n+5)\},$$

where  $K_{XY}$  is the Kendall statistic (2.1).

If  $R_i = S_i$  for all  $i$ , then  $\rho_K = 1$ ; if  $R_i = n+1-S_i$  for all  $i$ , then  $\rho_K = -1$ . In either case, the bivariate problem reduces to the univariate one, and the Mann trend test is applicable. When  $-1 < \rho_K < 1$ , we use the following theorem to define a bivariate test statistic.

Theorem 2: Under  $H_0$ , if  $-1 < \rho_K < 1$ , then  $(K_X/\sigma \ K_Y/\sigma)'$  is asymptotically

$$N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_K \\ \rho_K & 1 \end{pmatrix} \right),$$

and

$$(K_X/\sigma \ K_Y/\sigma) \begin{pmatrix} 1 & \rho_K \\ \rho_K & 1 \end{pmatrix}^{-1} \begin{pmatrix} K_X/\sigma \\ K_Y/\sigma \end{pmatrix} \quad (2.2)$$

is asymptotically  $\chi^2(2)$ .

Proof: We use a central limit theorem of Noether (1970) to prove that  $K_X/\sigma$  and  $K_Y/\sigma$  are asymptotically bivariate normal. It is sufficient to show that  $aK_X + bK_Y$ , properly standardized, is asymptotically normal for all nonzero  $a$  and  $b$ . Noether's theorem concerns statistics of the form  $T = \sum_{i \in I} \sum_{j \in J} v_{ij}$ , where the  $v_{ij}$  are uniformly bounded and  $v_{ij}$  is independent of  $v_{gh}$  if neither  $i$  nor  $j$  equals  $g$  or  $h$ . The theorem states that  $(T - E(T))/\text{Var}^{1/2}(T)$  is asymptotically  $N(0,1)$  if  $\text{Var}(T)$  is of order  $n^3$ . We have  $T = aK_X + bK_Y = \sum_{i=1}^{n-1} \sum_{j=i+1}^n v_{ij}$ , where



$$\begin{aligned}
v_{ij} &= a+b, \text{ if } X_j > X_i, Y_j > Y_i, \\
&= -a-b, \text{ if } X_j < X_i, Y_j < Y_i, \\
&= a-b, \text{ if } X_j > X_i, Y_j < Y_i, \\
&= -a+b, \text{ if } X_j < X_i, Y_j > Y_i.
\end{aligned}$$

Then

$$\begin{aligned}
\text{Var}(T) &= a^2 \text{Var}(K_X) + b^2 \text{Var}(K_Y) + 2ab\rho_K (\text{Var}(K_X)\text{Var}(K_Y))^{\frac{1}{2}} = \\
&\sigma^2(a^2 + b^2 + 2ab\rho_K),
\end{aligned}$$

which is of order  $n^3$ , as required by the theorem.

The second part of the theorem follows directly from the first part.

Thus the quadratic form (2.2) is an asymptotically distribution-free test statistic for  $H_0$  in the bivariate case.

It is interesting to note that  $\rho_K$  is a linear combination of two well-known measures of correlation between the  $X_i$  and  $Y_i$ . Kendall's  $\hat{\tau}$ , the point estimate of  $\tau$ , is defined as  $2K_{XY}/n(n-1)$ . Spearman's rho, given by

$$r_S = 12 \frac{\sum_{i=1}^n \{R_i - (n+1)/2\} \{S_i - (n+1)/2\}}{n(n^2-1)},$$

is the classical sample correlation coefficient calculated using ranks.

Algebraic manipulation yields  $\rho_K = \alpha \hat{\tau} + (1-\alpha)r_S$ , where  $\alpha = 3/(2n+5)$ .

Thus  $\rho_K \rightarrow r_S$  everywhere.

## 4. MULTIVARIATE TREND TEST

We now extend the test for time trend to the multivariate case. Let the matrix

$$\underline{X} = \begin{pmatrix} X_{11} & X_{12} & \cdot & \cdot & \cdot & X_{1p} \\ X_{21} & X_{22} & \cdot & \cdot & \cdot & X_{2p} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ X_{n1} & X_{n2} & \cdot & \cdot & \cdot & X_{np} \end{pmatrix}$$

denote a sequence of continuous  $p$ -vectors, observed at times  $1, 2, \dots, n$ . We wish to test  $H_0$ , the null hypothesis that the  $p$ -vectors are randomly ordered, versus the alternative of a monotone trend in one or more of the  $p$  variables. Replace  $\underline{X}$  by the matrix of ranks

$$\underline{R} = \begin{pmatrix} R_{11} & R_{12} & \cdot & \cdot & \cdot & R_{1p} \\ R_{21} & R_{22} & \cdot & \cdot & \cdot & R_{2p} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ R_{n1} & R_{n2} & \cdot & \cdot & \cdot & R_{np} \end{pmatrix}$$

where the  $n$  observations for each coordinate are ranked among themselves. Thus each column of  $\underline{R}$  is a permutation of  $(1, 2, \dots, n)$ . Let  $K_i$  be the Kendall statistic for the ranks  $R_{1i}, R_{2i}, \dots, R_{ni}$  and the times  $1, 2, \dots, n, i=1, 2, \dots, p$ . Under  $H_0$ , each  $K_i/\sigma$  is asymptotically  $N(0,1)$ ,

where  $\sigma^2 = n(n-1)(2n+5)/18$ . Noether's theorem can be used as before to prove that the  $K_i/\sigma$  are asymptotically  $p$ -variate normal, since a sufficient condition for multivariate normality is that every linear combination of the  $K_i/\sigma$  be asymptotically normal.

Given the  $n$   $p$ -vectors  $(R_{j1}, R_{j2}, \dots, R_{jp})$ ,  $j=1, 2, \dots, n$ , i.e., the rows of the matrix  $\underline{R}$ ,  $H_0$  implies that all  $n!$  time orderings of the rows are equally likely. The conditional null covariance,  $\sigma_{ij}$ , between  $K_i$  and  $K_j$ , given the rows of  $\underline{R}$ , equals the covariance defined in the bivariate case for the  $i$ th and  $j$ th variables. Thus,

$$\sigma_{ij} = K_{ij}/3 + 4 \sum_{\ell=1}^n R_{\ell i} R_{\ell j} / 3 - n(n+1)^2/3,$$

where  $K_{ij}$  is the Kendall statistic for the  $i$ th and  $j$ th coordinates.

Therefore, under  $H_0$ , the quadratic form

$$(K_1, K_2, \dots, K_p) \begin{pmatrix} \sigma^2 & \sigma_{12} & \cdot & \cdot & \cdot & \sigma_{1p} \\ \sigma_{12} & \sigma^2 & \cdot & \cdot & \cdot & \sigma_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{1p} & \sigma_{2p} & \cdot & \cdot & \cdot & \sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} K_1 \\ K_2 \\ \cdot \\ \cdot \\ K_p \end{pmatrix}$$

is asymptotically  $\chi^2(p)$ , and is an asymptotically distribution-free statistic for testing  $H_0$ .

## 5. EMPIRICAL LEVEL FOR SMALL SAMPLES

Since the multivariate test for time trend has no parametric or nonparametric competitors, efficiency calculations or power comparisons are not possible. To examine the adequacy of the  $\chi^2$  approximation for the null distribution, the bivariate test was performed on samples of size 12 and 20 from the  $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$  distribution with  $\rho = 0, .3, .6, \text{ and } .9$ . For each bivariate observation desired, two independent  $N(0,1)$  variables,  $X$  and  $Y$ , were generated using the Scientific Subroutine Package Program GAUSS.  $\begin{pmatrix} X \\ Y \end{pmatrix}$  was then multiplied by

$$\frac{1}{2} \begin{pmatrix} (1+\rho)^{\frac{1}{2}} + (1-\rho)^{\frac{1}{2}} & (1+\rho)^{\frac{1}{2}} - (1-\rho)^{\frac{1}{2}} \\ (1+\rho)^{\frac{1}{2}} - (1-\rho)^{\frac{1}{2}} & (1+\rho)^{\frac{1}{2}} + (1-\rho)^{\frac{1}{2}} \end{pmatrix}$$

to produce variables with the desired covariance matrix. This computer work was performed on the IBM 360/65 at the University of Connecticut Computer Center.

The nominal significance level for each test, based on the  $\chi^2(2)$  distribution, was .05. Each value of the empirical level shown in the table is based on 2000 samples. In every case, the empirical level is conservative; it appears to be increasing in  $n$  and decreasing in  $\rho$ . Thus, for the values of  $n$  typically encountered in the analysis of blood chemistry data from drug trials, the trend test is likely to be conservative.

Empirical Level of Trend Test				
n	$\rho$			
	0.0	0.3	0.6	0.9
12	.036	.033	.036	.031
20	.050	.047	.042	.034

APPENDIX: CALCULATION OF  $\rho_K$ 

We now calculate  $\rho_K$ , the conditional null correlation between  $K_X$  and  $K_Y$ , given the  $(R_i, S_i)$ ,  $i=1,2,\dots,n$ . Recall that  $(R_i, S_i)$  is the rank pair observed at time  $i$ ,  $i=1,2,\dots,n$ . Let  $a_i$  be the Y rank associated with the X rank of  $i$ ; let  $t_i$  be the time at which  $(i, a_i)$  was observed. Conditioning on the  $(R_i, S_i)$  is equivalent to conditioning on the  $a_i$ . Under  $H_0$ , given the  $a_i$ ,  $(t_1, t_2, \dots, t_n)$  assumes the  $n!$  permutations of  $(1, 2, \dots, n)$  with equal probability.

$K_X$ , the Kendall statistic for  $R_1, R_2, \dots, R_n$  and the times  $1, 2, \dots, n$ , is, equivalently, the Kendall statistic for the X ranks  $1, 2, \dots, n$  and the times  $t_1, t_2, \dots, t_n$ . Thus  $K_X = \sum_{i < j} c_{ij}$ , where

$$\begin{aligned} c_{ij} &= +1, \text{ if } t_i < t_j, \\ &= -1, \text{ if } t_i > t_j. \end{aligned}$$

$K_Y$  is the Kendall statistic for  $S_1, S_2, \dots, S_n$  and the times  $1, 2, \dots, n$ , or, equivalently, for the Y ranks  $a_1, a_2, \dots, a_n$  and the times  $t_1, t_2, \dots, t_n$ . If the  $a_i$  are arranged in increasing order, the corresponding times are  $t_{a_1}, t_{a_2}, \dots, t_{a_n}$ . Thus  $K_Y$  is also the Kendall statistic for the ranks  $1, 2, \dots, n$  and the times  $t_{a_1}, t_{a_2}, \dots, t_{a_n}$ .

Let

$$\begin{aligned} b_{ij} &= +1, \text{ if } t_{a_i} < t_{a_j}, \\ &= -1, \text{ if } t_{a_i} > t_{a_j}; \end{aligned}$$

then  $K_Y = \sum_{i < j} b_{ij}$ .

The conditional covariance between  $K_X$  and  $K_Y$  is

$$E(K_X K_Y | (R_i, S_i), i=1, 2, \dots, n) = E(K_X K_Y | a_1, a_2, \dots, a_n),$$

where the expectation is taken with respect to the probability measure that assigns mass  $1/n!$  to each possible value of  $(t_1, t_2, \dots, t_n)$ . To simplify notation,  $a_1, a_2, \dots, a_n$  will be omitted in the following conditional expectations. We have

$$E(K_X K_Y) = E(\sum_{i < j} c_{ij} \sum_{i < j} b_{ij}) = \sum_{i < j} \sum_{k < \ell} E(b_{ij} c_{k\ell}).$$

To evaluate  $E(b_{ij} c_{k\ell})$ , we consider seven cases, depending upon the values of  $a_i, a_j, k$ , and  $\ell$ . We say that  $(t_{a_i}, t_{a_j})$  is concordant with  $(t_k, t_\ell)$  if  $\text{sgn}(t_{a_i} - t_{a_j}) = \text{sgn}(t_k - t_\ell)$ ; otherwise the pairs are discordant. For values of  $(t_1, t_2, \dots, t_n)$  such that  $(t_{a_i}, t_{a_j})$  and  $(t_k, t_\ell)$  are concordant (discordant),  $b_{ij} c_{k\ell}$  equals  $+1$  ( $-1$ ). The seven cases and the value of  $E(b_{ij} c_{k\ell})$  for each are as follows.

1.  $a_i = k, a_j = \ell$ .  $(t_{a_i}, t_{a_j})$  is always concordant with  $(t_k, t_\ell)$ .

$$E(b_{ij} c_{k\ell}) = 1.$$

2.  $a_i = \ell, a_j = k$ .  $(t_{a_i}, t_{a_j})$  is always discordant with  $(t_k, t_\ell)$ .

$$E(b_{ij} c_{k\ell}) = -1.$$

3.  $a_i = k, a_j \neq \ell$ . Partition the  $n!$  values of  $(t_1, t_2, \dots, t_n)$  into

six groups of equal size, depending upon the ranks of  $t_{a_i} = t_k$ ,

$t_{a_j}$ , and  $t_\ell$  among themselves. These ranks determine the

concordance or discordance of  $(t_{a_i}, t_{a_j})$  and  $(t_k, t_l) = (t_{a_i}, t_l)$ ,

and thus the value of  $b_{ij}c_{kl}$ .

$t_{a_i}$	$t_{a_j}$	$t_l$	$b_{ij}c_{kl}$
1	2	3	1
1	3	2	1
2	1	3	-1
2	3	1	-1
3	1	2	1
3	2	1	1

Thus  $E(b_{ij}c_{kl}) = (4(1) + 2(-1))/6 = 1/3$ .

4.  $a_i \neq k, a_j = l$ . This case is similar to (3).  $E(b_{ij}c_{kl}) = 1/3$ .
5.  $a_i = l, a_j \neq k$ . Again, partition the  $n!$  values of  $(t_1, t_2, \dots, t_n)$  into six groups, depending on the ranks of  $t_{a_i}, t_{a_j} = t_l$ , and  $t_k$ .

$t_{a_i}$	$t_{a_j}$	$t_k$	$b_{ij}c_{kl}$
1	2	3	-1
1	3	2	-1
2	1	3	1
2	3	1	1
3	1	2	-1
3	2	1	-1

Thus  $E(b_{ij}c_{kl}) = (4(-1) + 2(1))/6 = -1/3$ .

6.  $a_i \neq l, a_j = k$ . This case is similar to (5).  $E(b_{ij}c_{kl}) = -1/3$ .

7.  $a_i \neq k, a_i \neq l, a_j \neq k, a_j \neq l$ .  $b_{ij}$  and  $c_{kl}$  are independent,

each with mean 0.  $E(b_{ij}c_{kl}) = 0$ .



For the given  $(a_1, a_2, \dots, a_n)$ , we count the number of times each case appears in the sum  $\sum_{i < j} \sum_{k < l} E(b_{ij} c_{kl})$ . For an  $i < j$  for which  $a_i < a_j$ , the values of  $k < l$  fall into the seven cases as follows.

Case	$E(b_{ij} c_{kl})$	Frequency
1	1	1
3	1/3	$n - a_i - 1$
4	1/3	$a_j - 2$
5	-1/3	$a_i - 1$
6	-1/3	$n - a_j$

Therefore,

$$\begin{aligned} & \sum_{i < j, a_i < a_j} \sum_{k < l} E(b_{ij} c_{kl}) = \\ & \sum_{i < j, a_i < a_j} \{1 + (n - a_i - 1)/3 + (a_j - 2)/3 - (a_i - 1)/3 - (n - a_j)/3\} = \\ & \sum_{i < j, a_i < a_j} \{1/3 + 2(a_j - a_i)/3\}. \end{aligned}$$

For an  $i < j$  for which  $a_i > a_j$ , we have the following partition of  $k < l$  into cases:

Case	$E(b_{ij} c_{kl})$	Frequency
2	-1	1
3	1/3	$n - a_i$
4	1/3	$a_j - 1$
5	-1/3	$a_i - 2$
6	-1/3	$n - a_j - 1$

Therefore,

$$\begin{aligned} & \sum_{i < j, a_i > a_j} \sum_{k < l} E(b_{ij} c_{kl}) = \\ & \sum_{i < j, a_i > a_j} \{-1 + (n-a_i)/3 + (a_j-1)/3 - (a_i-2)/3 - (n-a_j-1)/3\} = \\ & \sum_{i < j, a_i > a_j} \{-1/3 + 2(a_j-a_i)/3\}. \end{aligned}$$

Finally,

$$E(K_{XY}) = \sum_{i < j, a_i < a_j} \{1/3 + 2(a_j-a_i)/3\} + \sum_{i < j, a_i > a_j} \{-1/3 + 2(a_j-a_i)/3\} =$$

$$K_{XY}/3 + 2 \sum_{i < j} (a_j - a_i)/3,$$

where  $K_{XY}$  is Kendall's statistic for  $a_1, a_2, \dots, a_n$  and the integers  $1, 2, \dots, n$ , or equivalently, for the  $X_i$  and  $Y_i$ . Rewriting  $\sum_{i < j} (a_j - a_i)$  as

$$2 \sum_{i=1}^n ia_i - n(n+1)^2/2 = 2 \sum_{i=1}^n R_i S_i - n(n+1)^2/2$$

gives

$$E(K_{XY}) = K_{XY}/3 + 4 \sum_{i=1}^n R_i S_i/3 - n(n+1)^2/3.$$

Thus

$$\rho_K = E(K_{XY})/\sigma^2 = \{6K_{XY} + 24 \sum_{i=1}^n R_i S_i - 6n(n+1)^2\} / \{n(n-1)(2n+5)\},$$

as stated in Theorem 1.

## REFERENCES

- Elveback, L. R., Guillier, C. L., and Keating, F. R., Jr. (1970), "Health, Normality, and the Ghost of Gauss," Journal of the American Medical Association, 211, 69-75.
- Harris, Eugene K. (1974), "Effects of Intra- and Interindividual Variation on the Appropriate Use of Normal Ranges," Clinical Chemistry, 20, 1535-1542.
- Harris, Eugene K. (1975), "Some Theory of Reference Values. I. Stratified (Categorized) Normal Ranges and a Method for Following an Individual's Clinical Laboratory Values," Clinical Chemistry, 21, 1457-1464.
- Healy, M. J. R. (1969), "Normal Values from a Statistical Viewpoint," Bulletin de l'Academie Royale de Medecine de Belgique, 9, 703-718.
- Herrera, Lee (1958), "The Precision of Percentiles in Establishing Normal Limits in Medicine," Journal of Laboratory and Clinical Medicine, 52, 34-42.
- Kendall, Maurice G. (1970), Rank Correlation Methods, London: Charles Griffin & Company Limited.
- Mainland, Donald (1971), "Remarks on Clinical 'Norms'," Reprinted in Clinical Chemistry, 17, 267-274.
- Mann, Henry B. (1945), "Nonparametric Tests Against Trend," Econometrica, 13, 245-259.
- Nadel, E., Mantel, N., Kaplan, N., and Altschuler, H. (1969), "Individuality in Serum Lactic Dehydrogenase (LDH) Activity in Aged Persons: Significance for Cancer Diagnosis," Annals of the New York Academy of Sciences, 161, 581-601.
- Noether, Gottfried E. (1970), "A Central Limit Theorem with Nonparametric Applications," Annals of Mathematical Statistics, 41, 1753-1755.
- Schoen, I., and Brooks, S. H. (1970), "Judgment Based on 95% Confidence Limits: A Statistical Dilemma Involving Multitest Screening and Proficiency Testing of Multiple Specimens," American Journal of Clinical Pathology, 53, 190-193.
- Winkel, P., Lyngbye, J., and Jørgensen, K. (1972), "The Normal Region - a Multivariate Problem," Scandinavian Journal of Clinical and Laboratory Investigation, 30, 339-344.