NOTE ON THE WILCOXON-MANN-WHITNEY STATISTIC

By S. K. Zaremba

University College of Swansea

In a note recently published in these Annals, R. F. Potthoff [2], using the bounds for the variance of the Wilcoxon-Mann-Whitney statistic obtained by Z. W. Birnbaum and O. H. Klose [1], attempted to apply this statistic to a test for the coincidence of the medians of two random variables, each of which was assumed to be continuously and symmetrically distributed about its median. The test was claimed to be *consistent for practical purposes*. Unfortunately, consistency for practical purposes was not defined. However, Potthoff's note added some topicality to the question of the possible uses of the Wilcoxon-Mann-Whitney statistic and of *a priori* bounds for its variance.

In an earlier paper [3], I showed that the statistic in question could be used in in a test of the null hypothesis $P[X > Y] = \frac{1}{2}$ when X and Y were *any* random variables, and indeed when the relation "$>$" denoted non-metric preferences, which are relevant to psychology, market research, etc. I further showed that the upper bound obtained by Birnbaum and Klose for the variance of this statistic in the case of continuous distributions still applied under the null hypothesis not only in the case of discontinuous variables, but even in that of non-metric preferences, provided that these should be transitive. This is all we need in connection with the test, but a trivial modification of the proof is sufficient to cover the case when $P[X > Y]$ takes any positive value smaller than 1.

The greatest possible lower bound for the same variance was obtained by me, also for possibly non-metric preferences, under the assumption that the samples of X and Y were of the same size; however, an elaboration of my earlier argument proves that, at least when $P[X > Y] = \frac{1}{2}$, the lower bound obtained by Birnbaum and Klose in the case of continuous distributions applies to the case of non-metric preferences as well, let alone to that of discontinuous distributions. The crux of the argument depends on the following combinatorial lemma:

Lemma. *Let the* $m + n$ *objects* $x_1, \cdots, x_m, y_1, \cdots, y_n$ ($n \geqq m$; $n - m$ *even*) *be arbitrarily ordered, and let*

$$W = \sum_{i=1}^{m} \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} \xi_{ik}\xi_{il} + \sum_{k=1}^{n} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \xi_{ik}\xi_{jk},$$

where

$$\xi_{ik} = \tfrac{1}{2} \qquad \text{if } x_i \text{ precedes } y_k;$$
$$= -\tfrac{1}{2} \qquad \text{if } x_i \text{ follows } y_k;$$

then

$$W \geqq \tfrac{1}{8}m(m-2)(n-1) - \tfrac{1}{24}m(m-1)(m-2). \tag{1}$$

Received 7 August 1964; revised 23 November 1964.