THE STANDARD ERROR OF GINI'S MEAN DIFFERENCE

By Z. A. Lomnicki

Polish University College, London

A general expression for the standard error of Gini's mean difference g was given in a paper under the same title by U. S. Nair [1]. See also [2], pp. 216–217.

The object of this note is to deduce in a more direct way a simpler formula for the variance of this statistic. The expression obtained is equivalent to that given by Nair except for an additional term overlooked in his final formula. The simplification is due to the fact that, for the evaluation of the expected values of g and g^2, it is not necessary to arrange the sample values in ascending order of magnitude as done by Nair.

Let n be the size of the sample, $f(x)$ the probability density function of the parent population, μ the mean and σ^2 the variance of x in the parent and let

$$F(x) = \int_{-\infty}^{x} f(t)\, dt, \qquad Z(x) = \int_{-\infty}^{x} tf(t)\, dt.$$

From the definition

$$g = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} | x_i - x_j | \tag{1}$$

(where the values x_i are not in order of magnitude but are numbered as they appear in the sample), we have

$$\begin{aligned} E(g) &= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} E(| x_i - x_j |) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} | x - y | f(x)f(y)\, dx\, dy = \Delta, \end{aligned} \tag{2}$$

where Δ is the mean difference (parameter) of the parent population. It is easy to check that Δ can also be written

$$\Delta = 2 \int_{-\infty}^{+\infty} \{xF(x) - Z(x)\} f(x)\, dx = 2 \int_{-\infty}^{\infty} xf(x)(2F(x) - 1)\, dx. \tag{3}$$

In order to find $E(g^2)$ let us write

$$g^2 = \frac{4}{n^2(n-1)^2} \{\sum (x_i - x_j)^2 + 2\sum | x_i - x_j | | x_i - x_k | + 2\sum | x_i - x_j | | x_k - x_l |\}. \tag{4}$$

The first sum should be read as the double sum extended to all pairs of different subscripts i, j, and has $n(n-1)/2$ terms; the second as a triple sum extended to all combinations of two pairs (i, j), (i, k) of different subscripts i, j, k and has $n(n-1)(n-2)/2$ terms; the third as a quadruple sum extended to all com-