THE 5% SIGNIFICANCE LEVELS FOR SUMS OF SQUARES OF RANK DIFFERENCES AND A CORRECTION

By Edwin G. Olds

Carnegie Institute of Technology

About ten years ago this author published a paper [1], containing tables for use in testing the significance of the rank correlation coefficient. In a paper on non-parametric tests, [2, p. 316] Scheffé remarks that it would be desirable to have these tables extended by inclusion of the 5% values. When the computation was begun it was noted that a necessary formula was given incorrectly. The main purpose of this note is to correct the formula and to extend Table V, [1, p. 148]. Incidentally, a minor addition for Table III, [1, p. 143] will be supplied.

The formula for the rank correlation coefficient, r', is given by

$$r' = 1 - \frac{6 \sum d^2}{n^3 - n},$$

where n is the number of individuals ranked and $\Sigma d^2 = \sum_{i=1}^{n} d_i^2$ (d_i being the rank difference for the ith individual). As noted in the original paper, the null hypothesis, $r' = 0$, is equivalent to the hypothesis $\Sigma d^2 = (n^3 - n)/6$, and the latter hypothesis is slightly more convenient to test. Scheffé's remark seems to be directed at Table V, which gives, for $11 \leq n \leq 30$, pairs of values between which Σd^2 has a probability, P, of being included. Values are tabled for $P = .99$, .98, .96, .90 and .80. The necessary values for $P = .95$ are given below and can easily be copied in the left-hand margin of the original Table [1, p. 148]. These values, as in the previous case, have been calculated by using the fact that

$$x = \frac{\sum d^2}{2} - \frac{n^3 - n}{12}$$

has an approximately normal distribution with a mean of zero and a variance of $(n - 1)[n(n + 1)/12]^2$. In the original paper, [1, p. 142] the denominator in the bracketed part of the variance was printed as 6, instead of 12.

In this author's original paper the exact frequencies of sums of squares of rank differences were given for $n = 2$ to $n = 7$ inclusive, [1, p. 139]. The same results, together with the results for $n = 8$, were obtained (independently) by Kendall and others and published some months later, [3, p. 255]. Therefore, it is possible to extend slightly the comparison of approximating functions given in Table III, [1, p. 143]. Using Kendall s results for $n = 8$ it is found that when the approximations obtained by using a Pearson Type II curve are compared with exact results the average and maximum differences of cumulatives are .0013 and .0067 respectively. When approximations are made by using the normal curve the corresponding errors are .0081 and .0163.