DISTRIBUTIONS OF SUMS OF SQUARES OF RANK DIFFERENCES FOR SMALL NUMBERS OF INDIVIDUALS[1]

By E. G. Olds

I. INTRODUCTION

In a recent article,[2] reporting the results of research under a grant-in-aid from the Carnegie Corporation of New York, Hotelling and Pabst have given a comprehensive treatment of the theory and application of rank correlation and have contributed significantly to existing knowledge on the subject. It is not the purpose of this note to evaluate their contribution but to attempt the solution of a problem they suggest.

In §3[3] they have given the well-known formula for rank correlation, $r' = 1 - \frac{6\Sigma d^2}{n^3 - n}$, 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). In §5 the question of the significance of r' in small samples has been considered from the following point of view; if the value of r', obtained from a comparison of the ranks of n individuals as a possible measure of the relation between two attributes, is such that there exists a high probability that it could have occurred by virtue of a chance rearrangement of the n individuals, then the value of r' does not furnish a significant indication of relationship. Then one test of the significance of a particular value of r' is to note whether it has a probability less than P (P equal to .01 or, less stringently, equal to .05) of occurring because of a chance re-ranking.

To apply this test it is necessary to have some information regarding the distribution of r' for the chance rearrangements of the numbers from 1 to n. Hotelling and Pabst have given the distribution of r' for the cases, $n = 2, 3, 4$. They have noted that the distribution is symmetrical for each value of n and that it has a range from -1 to 1. From a consideration of the probabilities corresponding to $\Sigma d^2 = 0, 2, 4, 6$, they have discussed the significance of values of r' for $n = 5, 6, 7$. In §8 they have stated, "Another problem is to find convenient and accurate approximations to the distribution of r', for moderate values of n, with close limits of error. A table calculated along the lines suggested in §5 would be very useful." This statement, along with the interest manifested by others in private communications, has led to the investigation reported in this paper.

---

[1] Presented to the American Mathematical Society, December 29, 1936.

[2] Harold Hotelling and Margaret Pabst, *Rank Correlation and Tests of Significance Involving No Assumption of Normality*, Annals of Mathematical Statistics, Vol. VII, 1936, pp. 29-43.

[3] Loc. cit.