RANK CORRELATION AND REGRESSION IN A NONNORMAL SURFACE

F. N. DAVID[1] AND EVELYN FIX[2]
UNIVERSITY OF CALIFORNIA, BERKELEY

1. The problem

The problem, seemingly simple, which led to the work reported in this paper, was this: to find a "usable" test for a difference in position of two regression lines if it is known that the two lines have the same slope but it is also known that the marginal distributions are extremely skewed and one is loath to use transformations and normal theory may be inadequate. More precisely, we consider the following situation. Suppose that there are two sets of bivariate correlated observations, n_1 in the first set and n_2 in the second set, with $n_1 + n_2 = n$. Neither the functional form of the bivariate population (or populations) nor any of the parameters descriptive of the population (populations) from which the observations have been drawn is known. However, the number of observations (between 20 and 50) is sufficient to indicate that the marginal distributions may be extremely skew Type III and the regression may be linear. Furthermore, from the description of the experiments generating the bivariate observations, it seems reasonable to assume that, if indeed they do come from different populations, the regression lines of the parent populations have the same slope with the difference occurring in the intercepts or position of the lines. The data may be analyzed in a variety of ways, including the orthodox one of transforming the data and using normal theory tests. Here we discuss the possibility of using tests based on the rank correlation coefficient attributed to Spearman. We propose the use of these tests in their conditional form. In order that the criteria proposed may be compared with others, we suggest a bivariate functional form for the population, which we have been unable to find discussed elsewhere and which enables us to throw some light on the behavior of the mean value of Spearman's rho when the population is nonnormal.

2. Rank correlation in the normal correlation distribution

Before proceeding with the discussion of the problem outlined, it is useful to recall what is known about Spearman's rho in the case of the bivariate normal

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-43).
[2] This investigation was supported (in part) by a research grant (No. RG-3666) from the National Institutes of Health, Public Health Service.