

A NEW FORMULA FOR PREDICTING THE SHRINKAGE OF THE COEFFICIENT OF MULTIPLE CORRELATION

By

DR. R. J. WHERRY
Cumberland University, Lebanon, Tennessee

With the perfection of the Doolittle Method for the solution of the constant values necessary for the multiple correlation and prediction technique, we may expect a constant increase in the use of this method in statistical practice. Theoretical statisticians have recognized for some time however that the multiple correlation coefficient, derived from a large number of independent variables, is apt to be deceptively large due to chance factors. When prediction equations derived in this manner are applied to subsequent sets of data, there is apt to be a rather large shrinkage in the resulting correlation coefficient obtained, as compared with the original observed multiple correlation coefficient. In order to avoid over optimism it is necessary to have some equation which will predict the most probable value of this shrinkage. The development of such a formula is the purpose of this paper.

The most promising formula of this type so far developed is the B. B. Smith formula, presented by M. J. B. Ezekial at the December, 1928, meeting of the American Mathematical Society held at Chicago. This formula is

$$(1) \quad \bar{R}^2 = 1 - \frac{1 - R^2}{1 - \frac{M}{N}} = \frac{NR^2 - M}{N - M}$$