ON THE SAMPLING DISTRIBUTION OF THE MULTIPLE CORRELATION COEFFICIENT

By S. S. Wilks*

The problem of finding the distribution of the multiple correlation coefficient in samples from a normal population with a non-zero multiple correlation coefficient was solved in 1928 by Fisher[1] by the application of geometrical methods. In his derivation he used the facts that the population value ρ of the multiple correlation coefficient is invariant under linear transformations of the independent variates, and that the distribution of the multiple correlation coefficient is independent of all population parameters except ρ.

In this paper it will be shown that the distribution of the multiple correlation coefficient can be derived directly from Wishart's[2] generalized product moment distribution without making use of geometrical notions and the property of the invariance of ρ under linear transformations of the independent variates. Furthermore, it will not be necessary to show that the distribution will be independent of all population parameters except ρ.

The population value of the multiple correlation coefficient between a variate x_1 and a set of variates $x_2, x_3, \cdots x_n$ is the ordinary correlation coefficient between x_1 and that linear function of the variates $x_2, x_3, \cdots x_n$ which will make this correlation a maximum. It can be expressed as $\rho^2 = 1 - \frac{\Delta}{\Delta_1}$ where Δ is the determinant of the correlations among all of the

*National Research Fellow in Mathematics.

[1]R. A. Fisher, The general sampling distribution of the multiple correlation coefficient, Proceedings of the Royal Society of London, series A, vol. 121 (1928), pp. 654-73.

[2]John Wishart, The generalized product moment distribution in samples from a normal multivariate population, Biometrika, vol. 20A (1928) pp. 32-52.