THE DISTRIBUTION OF THE MULTIPLE CORRELATION COEFFICIENT IN PERIODOGRAM ANALYSIS

By D. M. Starkey

1. **Geometrical interpretation of the problem.** We begin with a summary of some recent work by Hotelling, in a form relevant to this particular problem.[1] He suggests that the general question of finding the distribution of the multiple correlation coefficient corresponding to a fitted regression of y upon x may be solved by evaluating definite integrals corresponding to invariants of certain curves, surfaces, etc. For the purposes of illustration we may consider the case of fitting the relation

$$Y = a + bf(x, k, \epsilon)$$

where f is an arbitrary function, and a, b, k, ϵ are constants, to the observations y, where we are given n values of $y, y_1, y_2, \cdots, y_n$ and the corresponding values of $x, x_1, \cdots, x_n$. We shall postulate that the y's are independent and normally distributed about a certain mean and that the regression may be fitted by means of the principle of least squares.

We must minimize the sum of squares

$$\sum_{\alpha=1}^{\alpha=n} (y_\alpha - Y_\alpha)^2 = \sum_{\alpha=1}^{\alpha=n} [y_\alpha - a - bf(x_\alpha, k, \epsilon)]^2$$

and hence we differentiate with respect to a, obtaining the first condition for a minimum

$$\sum_{\alpha=1}^{\alpha=n} [y_\alpha - a - bf(x_\alpha, k, \epsilon)] = 0.$$

In the following, all summations take place over a range $\alpha = 1$ to n. Then we have

$$a = \bar{y} - b\bar{f}$$

where

$$\bar{y} = \frac{\Sigma y_\alpha}{n}, \qquad \bar{f} = \frac{\Sigma f(x_\alpha, k, \epsilon)}{n}$$

Thus we minimize the sum of squares

$$\Sigma[(y_\alpha - \bar{y}) - b(f(x_\alpha, k, \epsilon) - \bar{f})]^2$$

[1] Harold Hotelling, "Tubes and spheres in n-spaces, and a class of statistical problem", *American Journal of Mathematics*, April, 1939.