LETTERS TO THE EDITOR

The Distribution of the Root-Mean-Square of the Second type of the Multiple Correlation Co-efficient

The sampling distribution of R the multiple correlation coefficient of the one variate with $(p-1)$ other variates was worked out by Professor Fisher in 1928 under two different sets of conditions leading to different forms of distribution function. The first set of conditions consisted in the assumption that the bunch of readings in a sample for each of the $(p-1)$ variates remained constant from sample to sample, but only the bunch of readings for the first variate (whose multiple correlation with the $(p-1)$ other variates comes into the picture) changed from sample to sample ; there were also certain additional well-known assumptions which are common to the usual theory of linear regression. The resulting distribution of 'R' is also well known and was quoted by us in a recent issue of SCIENCE AND CULTURE[1]. This we shall call the first type of 'R' distribution. We found and announced in this Journal in the issue just mentioned that the root-mean-square of R is distributed (under the first type of conditions just referred to) in the same general manner (with only change of parameters) as 'R' itself. That is, the type of the distribution is conserved when we proceed to the mean. The second set of conditions consisted in the assumption that the variates constituted a multivariate normal population and that the bunch of readings for the $(p-1)$ other variates as also for the first variate changed from sample to sample. Under such assumptions it was found by Professor Fisher in 1928 that 'R' is distributed in the form

$$\text{Const. } R^{p-2}(1-R^2)^{\frac{n-p-2}{2}}\, dR.$$

$$\times\, {}_2F_1\left(\frac{n-1}{2},\ \frac{n-1}{2};\ \frac{p-1}{2},\ \beta^2 R^2\right) \quad . \quad . \quad (1)$$

where p is of course the numbers of characters, n is the size of the sample and β is the population value for the multiple correlation of the first variate with the $(p-1)$ others. This we call the second type of 'R' distribution.

Considering m such random samples each of size n from a multivariate normal population and denoting the root-mean square of the different 'R's of the different samples by $\bar{R}$, we have recently found that $\bar{R}$ is distributed in the form

$$\text{Const } \bar{R}^{m(p-1)-1}(1-\bar{R}^2)^{\frac{m(n-p)-2}{2}}$$

$$\times\, {}_2F_1\left[\frac{m(n-1)}{2}, \frac{m(n-1)}{2};\ \frac{m(p-2)+1}{2};\ \beta^2\bar{R}^2\right] d\,\bar{R} \quad (2)$$

Comparison of (1) and (2) shows that the root-mean-square of 'R' under the second type of conditions is distributed in the same general manner (with change of parameters) as 'R' itself. The property that the type of the distribution is conserved when we proceed to the mean holds therefore for both the two different types of 'R' distribution.

Statistical Laboratory,
Presidency College,
Calcutta, 6-6-1940.

S. N. Roy
Purnendu Bose

[1] S. N. Roy and Purnendu Bose, SCIENCE AND CULTURE, 5, 714, 1940.