NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

AN APPROXIMATE NORMALIZATION OF THE ANALYSIS OF VARIANCE DISTRIBUTION

By Edward Paulson[1]

Columbia University

The statistic $F = s_1^2/s_2^2$, where s_1^2 and s_2^2 are two independent estimates of the same variance, has played an essential part in modern statistical theory. All tests of significance involving the testing of a linear hypothesis, which includes the analysis of variance and covariance and multiple regression problems, can be reduced to finding the probability integral of the F distribution. This distribution (and the equivalent distribution of $z = \frac{1}{2} \log F$) has so far been directly tabulated only for the 20, 5, 1, and 0.1 percent levels of significance [1]. To find the critical value of F for some other probability level would require the use of Pearson's extensive triple-entry tables [2], which is not very convenient to use for this purpose, and in addition is inadequate for some ranges of the parameters.

It therefore appears that it might be of some practical value to have an approximate method of determining the critical values of F for other probability levels. A solution will be given based on a modified statistic U, a function of F, so selected as to tend to have a nearly normal distribution with zero mean and unit variance. This normalized statistic will have the additional advantage that further tests are possible with normalized variates, as pointed out by Hotelling and Frankel [3].

F can be written in the form

$$F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2},$$

where χ_1^2 and χ_2^2 have the chi-square distribution with n_1 and n_2 degrees of freedom respectively. It is known from the work of Wilson and Hilferty [4] that $\left(\frac{\chi^2}{n}\right)^{\frac{1}{3}}$ is nearly normally distributed with mean $1 - 2/9n$ and variance $2/9n$. An obvious approach to the problem of securing an approximation to the F distribution is to regard $F^{\frac{1}{3}}$ as the ratio of two normally distributed variates. In general the distribution of the ratio $v = y/x$ where y and x are normally and independently distributed with means m_y and m_x and standard deviations σ_y and σ_x

[1] Work done under a grant-in-aid from the Carnegie Corporation of New York.