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.

is not expressible in simple form. However Fieller [5] has shown that a function R of v, namely $R = \frac{vm_x - m_y}{\sqrt{v^2\sigma_x^2 + \sigma_y^2}}$ will be nearly normally distributed with zero mean and unit variance, provided the probability of x being negative is small. In the given problem it follows that we can regard

$$(1)\qquad U = \frac{\left(1 - \frac{2}{9n_2}\right)F^{1/3} - \left(1 - \frac{2}{9n_1}\right)}{\sqrt{\frac{2}{9n_2}F^{2/3} + \frac{2}{9n_1}}},$$

as nearly normally distributed (with zero mean and unit variance) provided $n_2 \geq 3$, for with $n_2 = 3$ the probability of the denominator of $F^{1/3}$ being negative is only .0003. If it is desired to use the lower tail of the F distribution, then the statistic U should only be used if n_1 is also ≥ 3. Ordinarily, in most applications only the upper tail of the F distribution is used, and n_2, which corresponds to the number of degrees of freedom in the estimate of the error variance, will be much greater than 3.

The following tables show the degree of accuracy of the approximation. The exact value of F corresponding to various levels of significance are compared

P	$n_1 = 1,\quad n_2 = 10$	
	$t = \sqrt{F}$	
	Approximation	Exact Value
.20	1.37	1.37
.05	2.21	2.23
.01	3.16	3.17
.001	4.63	4.59
.0001	6.40	6.22

P	$n_1 = 4,\quad n_2 = 8$		$n_1 = 6,\quad n_2 = 12$	
	F		F	
	Approximation	Exact Value	Approximation	Exact Value
.99	.058	.068	.123	.130
.95	.161	.166	.248	.250
.80	.407	.406	.497	.496
.20	1.92	1.92	1.72	1.72
.05	3.84	3.84	3.00	3.00
.01	7.12	7.01	4.85	4.82
.001	15.38	14.39	8.58	8.38

with the approximate values, which are found by solving (1) for F by considering it as a quadratic equation in $F^{\frac{1}{3}}$. In these tables $P = \int_F^\infty \varphi(F)\, dF$, where $\varphi(F)$ is the probability distribution of F. The case $n_1 = 1$ is of special interest, since here $F = t^2$, where t has Student's distribution, and is shown separately.

REFERENCES

[1] R. A. FISHER and F. YATES, *Statistical Tables for Biological, Agricultural, and Medical Research*, London, 1938.

[2] KARL PEARSON (Editor), *Tables of the Incomplete Beta Function*, Biometric Laboratory, London, 1934.

[3] H. HOTELLING and L. R. FRANKEL, "The transformation of statistics to simplify their distribution," *Annals of Math. Stat.*, Vols. 8–9 (1937–38), pp. 87–96.

[4] E. B. WILSON and M. M. HILFERTY, "The distribution of chi-square," *National Acad. Sc. Proc.*, Vol. 17 (1931), pp. 684–688.

[5] E. C. FIELLER, "The distribution of the index in a normal bivariate population," *Biometrika*, Vol. 24 (1932), pp. 428–440.

NOTE ON THE DISTRIBUTION OF ROOTS OF A POLYNOMIAL WITH RANDOM COMPLEX COEFFICIENTS

BY M. A. GIRSHICK

United States Department of Agriculture

In order to obtain the distribution of roots of a polynomial with random complex coefficients, it was found convenient to employ a rather well known theorem on complex Jacobians. Since proofs of this theorem are not very plentiful in the literature, a brief and simple proof of it is presented in this note.

THEOREM: *Let n analytic functions be defined by*

$$w_p = u_p + iv_p = f_p(z_1, z_2, \cdots, z_n), \quad (p = 1, 2, \cdots, n), \tag{1}$$

where $z_p = x_p + iy_p$, $i = \sqrt{-1}$. *Let j denote the Jacobian of the transformation of the n complex variables defined by* (1). *That is*

$$j = \begin{vmatrix} \frac{\partial w_1}{\partial z_1} \cdots \frac{\partial w_1}{\partial z_n} \\ \cdots\cdots\cdots\cdots \\ \frac{\partial w_n}{\partial z_1} \cdots \frac{\partial w_n}{\partial z_n} \end{vmatrix}. \tag{2}$$

Let furthermore J denote the Jacobian of the transformation of the 2n real variables defined by the equations $u_p = u_p(x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_n)$ *and* $v_p = v_p(x_1, x_2, \cdots, x_n; y_1, y_2, \cdots, y_n)$, $(p = 1, 2, \cdots, n)$. *That is*

$$J = \begin{vmatrix} U_x & U_y \\ V_x & V_y \end{vmatrix}, \tag{3}$$