A STUDY OF R. A. FISHER'S z DISTRIBUTION AND THE RELATED F DISTRIBUTION[1]

By Leo A. Aroian

Hunter College

1. Nature of the problem. Consider two samples of N_1 and N_2 drawings, each sample drawn from one of two populations consisting of variates normally distributed with equal population variances σ^2. We define the two sample means $\bar{x}_1 = \frac{\sum_{i=1}^{N_1} x_i}{N_1}$, $\bar{x}_2 = \frac{\sum_{j=1}^{N_2} x_j}{N_2}$, x_i's and x_j's independent variates. We calculate from the two samples

$$s_1^2 = \frac{\sum_{i=1}^{N_1} (x_i - \bar{x}_1)^2}{n_1} \quad \text{and} \quad s_2^2 = \frac{\sum_{j=1}^{N_2} (x_j - \bar{x}_2)^2}{n_2}, \qquad n_1 = N_1 - 1,\ n_2 = N_2 - 1.$$

The distribution of $z = \frac{1}{2} \log \frac{s_1^2}{s_2^2}$ is well known.

$$P(z) = \frac{2 n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2}}{\mathrm{B}\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \frac{e^{n_1 z}}{(n_1 e^{2z} + n_2)^{\frac{1}{2}(n_1+n_2)}} dz. \tag{1.1}$$

We shall denote the ordinates by $y(z)$. The purpose of this study is to discuss the seminvariants of the z distribution and also to find useful approximations for them; to show that as n_1 and n_2 approach infinity in any manner whatever the distribution of z approaches normality; to find the upper bound of the absolute value of the difference between the distribution function of z and the function determined by the approximate seminvariants of the distribution of z for n_1 and n_2 large; to approximate the z distribution by the Type III distribution, the Gram-Charlier Type A series, and the logarithmic frequency curve; and finally to investigate the same properties with respect to the F distribution, where $F = e^{2z} = \frac{s_1^2}{s_2^2}$. The non-existence of the moments of F for certain values of n_1 and n_2 is noted and explained on the basis of the distribution of the quotient $\frac{y}{x}$.

---

[1] Presented to the American Mathematical Society, September 10, 1938, New York City in part; and to the Institute December 27, 1939 at Philadelphia.