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.

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

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.

2. General features of the z distribution. The z distribution is always unimodal, asymmetrical if $n_1 \neq n_2$, and symmetrical if $n_1 = n_2$. We see that interchanging n_1 and n_2 is the same as replacing z by $-z$. Fisher [7] noted that the two parameter family of curves includes as special cases the normal curve, the χ^2 distribution, and Student's distribution. The mode is at $z = 0$, the maximum ordinate is

$$y(0) = \frac{2n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} (n_1 + n_2)^{-\frac{1}{2}(n_1+n_2)}$$

or approximately

$$(2.1) \qquad y(0) = \frac{1}{\sqrt{2\pi}} \left\{\frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}^{-\frac{1}{2}} \quad \text{for } n_1 \text{ and } n_2 \text{ large.}$$

The two points of inflection are

$$(2.2) \quad z = \tfrac{1}{2} \log \left\{\frac{n_1 n_2 + n_1 + n_2 \pm \sqrt{n_1^2 + n_2^2 + 2n_1^2 n_2 + 2n_1 n_2^2 + 2n_1 n_2}}{n_1 n_2}\right\}.$$

They are equidistant from the mode, a property also of the Pearson system of frequency curves [24]. Also $\lim_{z \to \pm\infty} z^n \frac{d^n y(z)}{dz^n} = 0.$

3. The moment generating function and seminvariants. The moment generating function of the z distribution is

$$(3.1) \quad M_z(\theta) = \left(\frac{n_2}{n_1}\right)^{\frac{1}{2}\theta} \frac{B\left(\frac{n_2 - \theta}{2}, \frac{n_1 + \theta}{2}\right)}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} = \left(\frac{n_2}{n_1}\right)^{\frac{1}{2}\theta} \frac{\Gamma\left(\frac{n_2 - \theta}{2}\right)\Gamma\left(\frac{n_1 + \theta}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}.$$

The seminvariants of Thiele are defined by the following identity in θ:

$$(3.2) \qquad \log M_x(\theta) = \lambda_1 \theta + \lambda_2 \frac{\theta^2}{2!} + \lambda_3 \frac{\theta^3}{3!} + \lambda_4 \frac{\theta^4}{4!} + \cdots.$$

To find λ_r we take the logarithm of the moment generating function, expand it in powers of θ and choose the coefficient of $\frac{\theta^r}{r!}$. A complete discussion of properties of seminvariants may be found elsewhere [4].

4. The seminvariants of z. Now by the following formulas [11] p. 38:

$$(4.1) \quad \log \Gamma(1 + x) = \frac{-s_1 x}{1} + \frac{s_2 x^2}{2} - \frac{s_3 x^3}{3} + \frac{s_4 x^4}{4} - \cdots, \qquad |x| < 1,$$

$$(4.2)\quad \log\Gamma(1-x) = s_1x + \frac{s_2x^2}{2} + \frac{s_3x^3}{3} + \frac{s_4x^4}{4} + \cdots, \qquad |x| < 1,$$

where in both formulas

$$s_1 = \lim_{n\to\infty}\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} - \log n\right),$$

$$s_n = \frac{1}{1^n} + \frac{1}{2^n} + \frac{1}{3^n} + \frac{1}{4^n} + \cdots, \qquad n \geqq 2.$$

Also

$$(4.3)\quad \log B(\tfrac{1}{2}[1+x], \tfrac{1}{2}) = \log\pi - \sigma_1 x + \sigma_2\frac{x^2}{2} - \sigma_3\frac{x^3}{3} + \sigma_4\frac{x^4}{4} - \cdots, \qquad |x| < 1,$$

where

$$\sigma_n = \frac{1}{1^n} - \frac{1}{2^n} + \frac{1}{3^n} - \frac{1}{4^n} + \cdots, \qquad n \geqq 1$$

and

$$\sigma_n = \left(1 - \frac{1}{2^{n-1}}\right)s_n, \qquad n \geqq 2.$$

Hence from (4.1) and (4.3)

$$(4.4)\quad \begin{aligned}\log\Gamma\left(\frac{1+x}{2}\right) = \tfrac{1}{2}\log\pi - x\left(\sigma_1 + \frac{s_1}{2}\right) + \frac{x^2}{2}\left(\sigma_2 + \frac{s_2}{2^2}\right) \\ - \frac{x^3}{3}\left(\sigma_3 + \frac{s_3}{2^3}\right) + \frac{x^4}{4}\left(\sigma_4 + \frac{s_4}{2^4}\right) - \cdots.\end{aligned}$$

Since $\sigma_n = \left(1 - \frac{1}{2^{n-1}}\right)s_n$, $n \geqq 2$, we may write (4.4) as

$$(4.5)\quad \log\Gamma\left(\frac{1+x}{2}\right) = \tfrac{1}{2}\log\pi - x\left(\sigma_1 + \frac{s_1}{2}\right) + \sum_{k=2}^{\infty}\frac{(-1)^k x^k}{k}\left(1 - \frac{1}{2^k}\right)s_k.$$

From (3.1)

$$(4.6)\quad \begin{aligned}\log M_z(\theta) = \log\Gamma\left(\frac{n_2-\theta}{2}\right) + \log\Gamma\left(\frac{n_1+\theta}{2}\right) \\ + \frac{\theta}{2}(\log n_2 - \log n_1) - \log\Gamma\left(\frac{n_1}{2}\right) - \log\Gamma\left(\frac{n_2}{2}\right).\end{aligned}$$

The results assume slightly different forms for (A) n_1 and n_2 each even; (B) n_1 and n_2 each odd; (C) n_1 even, n_2 odd; (D) n_1 odd, n_2 even. The general formula for $\lambda_{r:z}$ for all cases is

$$(4.7) \qquad \lambda_{r:z} = \sum_{k=0}^{\infty}\left\{\frac{(-1)^r(r-1)!}{(n_1+2k)^r} + \frac{(r-1)!}{(n_2+2k)^r}\right\}, \qquad r \geqq 2.$$

This result is not so useful from the point of view of numerical applications as the formulas which follow.

5. Case A, n_1 and n_2 each even.

From (4.6)

$$(5.1) \qquad \log\Gamma\left(\frac{n_2-\theta}{2}\right) = \log\left(\frac{n_2-2-\theta}{2}\right) + \log\left(\frac{n_2-4-\theta}{2}\right) + \cdots + \log\left(1-\frac{\theta}{2}\right) + \log\Gamma\left(1-\frac{\theta}{2}\right).$$

Now $\log\left(1-\frac{\theta}{n_2-2}\right) = -\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{\theta}{n_2-2}\right)^k$. There will be $\frac{n_2}{2} - 1$ series of this sort, and only one series of the type $\log\Gamma\left(1-\frac{\theta}{2}\right) = \sum_{k=1}^{\infty}\frac{s_k}{k}\left(\frac{\theta}{2}\right)^k$ as given by (4.1). In the above expansion and those succeeding, terms not involving θ are omitted, since such terms are not needed in finding the seminvariants of z. The series $\log\Gamma\left(1-\frac{\theta}{2}\right)$ will always occur. Then

$$(5.2) \qquad \log\Gamma\left(\frac{n_2-\theta}{2}\right) = -\sum_{k=1}^{\infty}\frac{1}{k}\left[\left(\frac{\theta}{n_2-2}\right)^k + \left(\frac{\theta}{n_2-4}\right)^k + \cdots + \left(\frac{\theta}{2}\right)^k - s_k\left(\frac{\theta}{2}\right)^k\right],$$

or

$$(5.3) \qquad \log\Gamma\left(\frac{n_2-\theta}{2}\right) = \sum_{k=1}^{\infty}\frac{s_k}{k}\left(\frac{\theta}{2}\right)^k - \sum_{k=1}^{\infty}\frac{1}{k}\sum_{l=1}^{\frac{1}{2}n_2-1}\left(\frac{\theta}{2l}\right)^k.$$

We remark that the double sum is zero if $n_2 = 2$. Similarly

$$(5.4) \qquad \log\Gamma\left(\frac{n_1+\theta}{2}\right) = \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\left\{\left(\frac{\theta}{n_1-2}\right)^k + \left(\frac{\theta}{n_1-4}\right)^k + \cdots + \left(\frac{\theta}{2}\right)^k - s_k\left(\frac{\theta}{2}\right)^k\right\},$$

or

$$(5.5) \qquad \log\Gamma\left(\frac{n_1+\theta}{2}\right) = \sum_{k=1}^{\infty}\frac{(-1)^k}{k}s_k\left(\frac{\theta}{2}\right)^k + \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\sum_{l=1}^{l=\frac{1}{2}n_1-1}\left(\frac{\theta}{2l}\right)^k.$$

By use of (5.3) and (5.5) we have for the seminvariants of z, when n_1 and n_2 are even

$$(5.6) \qquad \lambda_{r:z} = \frac{(r-1)!}{2^r}\left\{\left(s_r - \sum_{k=1}^{\frac{1}{2}n_2-1}\frac{1}{k^r}\right) + (-1)^r\left(s_r - \sum_{k=1}^{\frac{1}{2}n_1-1}\frac{1}{k^r}\right)\right\}, \qquad r \geqq 2.$$

For $\lambda_{1:z} = \bar{z}$ we have by (4.6), (4.3), and (4.5)

$$\lambda_{1:z} = \frac{1}{2}\left[\left(\log n_2 - \sum_{k=1}^{\frac{1}{2}n_2-1} \frac{1}{k}\right) - \left(\log n_1 - \sum_{k=1}^{\frac{1}{2}n_1-1} \frac{1}{k}\right)\right]. \tag{5.7}$$

6. Case B, n_1 and n_2 odd. We have

$$\begin{aligned}\log \Gamma\left(\frac{n_2 - \theta}{2}\right) &= \log\left(\frac{n_2 - 2 - \theta}{2}\right) + \log\left(\frac{n_2 - 4 - \theta}{2}\right) + \cdots \\ &\qquad + \log\left(\frac{1 - \theta}{2}\right) + \log \Gamma\left(\frac{1 - \theta}{2}\right).\end{aligned} \tag{6.1}$$

Expanding $\log \Gamma\left(\frac{1-\theta}{2}\right)$ by (4.5)

$$\begin{aligned}\log \Gamma\left(\frac{n_2 - \theta}{2}\right) &= -\left[\sum_{k=1}^{\infty} \frac{\theta^k}{k(n_2 - 2)^k} + \frac{\theta^k}{k(n_2 - 4)^k} + \cdots + \frac{\theta^k}{k}\right] \\ &\qquad + \theta\left(\sigma_1 + \frac{s_1}{2}\right) + \sum_{k=2}^{\infty} \frac{\theta^k}{k}\left(1 - \frac{1}{2^k}\right)s_k.\end{aligned} \tag{6.2}$$

However $s_k\left(1 - \frac{1}{2^k}\right) = \frac{1}{1^k} + \frac{1}{3^k} + \frac{1}{5^k} + \frac{1}{7^k} + \cdots$, $k > 1$, which we shall denote hereafter by t_k. Hence (6.2) becomes

$$\log \Gamma\left(\frac{n_2 - \theta}{2}\right) = \theta\left(\sigma_1 + \frac{s_1}{2}\right) + \sum_{k=2}^{\infty} \frac{\theta^k}{k} t_k - \sum_{k=1}^{\infty} \frac{1}{k} \sum_{l=0}^{\frac{1}{2}(n_2-3)} \left(\frac{\theta}{2l + 1}\right)^k. \tag{6.3}$$

Also

$$\begin{aligned}\log \Gamma\left(\frac{n_1 + \theta}{2}\right) &= \log\left(\frac{n_1 + \theta - 2}{2}\right) + \log\left(\frac{n_1 + \theta - 4}{2}\right) + \cdots \\ &\qquad + \log\left(\frac{1 + \theta}{2}\right) + \log \Gamma\left(\frac{1 + \theta}{2}\right),\end{aligned} \tag{6.4}$$

and

$$\begin{aligned}\log \Gamma\left(\frac{n_1 + \theta}{2}\right) &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}\left[\frac{\theta^k}{(n_1 - 2)^k} + \frac{\theta^k}{(n_1 - 4)^k} + \cdots + \frac{\theta^k}{1}\right] \\ &\qquad - \theta\left(\sigma_1 + \frac{s_1}{2}\right) + \sum_{k=2}^{\infty} \frac{(-1)^k}{k} \theta^k t_k.\end{aligned} \tag{6.5}$$

$$\begin{aligned}\log \Gamma\left(\frac{n_1 + \theta}{2}\right) &= -\theta\left(\sigma_1 + \frac{s_1}{2}\right) \\ &\qquad + \sum_{k=2}^{\infty} \frac{(-1)^k \theta^k}{k} t_k + \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \sum_{l=0}^{l=\frac{1}{2}(n_1-3)} \frac{\theta^k}{(2l + 1)^k}.\end{aligned} \tag{6.6}$$

Combining both these results (6.3) and (6.6) we have

$$\lambda_{r:z} = (r-1)!\left\{\left(t_r - \sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{(2k+1)^r}\right) + (-1)^r\left(t_r - \sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{(2k+1)^r}\right)\right\}, \qquad r \geqq 2. \tag{6.7}$$

$$\lambda_{1:z} = \bar{z} = \left(\frac{1}{2}\log n_2 - \sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{2k+1}\right) - \left(\frac{1}{2}\log n_1 - \sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{2k+1}\right). \tag{6.8}$$

7. Cases C, D, and values of s_k, σ_k, t_k. The formulas for case C, n_1 even, n_2 odd are

$$\lambda_{r:z} = (r-1)!\left\{\left(t_r - \sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{(2k+1)^r}\right) + \frac{(-1)^r}{2^r}\left(s_r - \sum_{k=1}^{\frac{1}{2}n_1-1} \frac{1}{k^r}\right)\right\}, \qquad r \geqq 2. \tag{7.1}$$

$$\lambda_{1:z} = \bar{z} = \frac{1}{2}\log\frac{n_2}{n_1} + \frac{1}{2}\sum_{k=1}^{\frac{1}{2}n_1-1}\frac{1}{k} - \sum_{k=0}^{\frac{1}{2}(n_2-3)}\frac{1}{2k+1} + \sigma_1. \tag{7.2}$$

The results for case D, n_1 odd, n_2 even are

$$\lambda_{r:z} = (r-1)!\left\{\frac{1}{2^r}\left(s_r - \sum_{k=1}^{\frac{1}{2}n_2-1} \frac{1}{k^r}\right) + (-1)^r\left(t_r - \sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{(2k+1)^r}\right)\right\}, \qquad r \geqq 2. \tag{7.3}$$

$$\lambda_{1:z} = \bar{z} = \frac{1}{2}\log\frac{n_2}{n_1} - \sigma_1 + \sum_{k=0}^{\frac{1}{2}(n_1-3)}\frac{1}{2k+1} - \frac{1}{2}\sum_{k=1}^{\frac{1}{2}n_2-1}\frac{1}{k}. \tag{7.4}$$

We list the numerical values of s_k and t_k, $k \leqq 10$. The values of s_k are from Stieltjes [20],

(7.5)	s_1	$= 0.57721\ 56649$	(7.6)	σ_1	$= \log 2 = 0.69317\ 0206$
	s_2	$= 1.64493\ 40668$		t_2	$= 1.23370\ 00550$
	s_3	$= 1.20205\ 69032$		t_3	$= 1.05179\ 97903$
	s_4	$= 1.08232\ 32337$		t_4	$= 1.01467\ 80316$
	s_5	$= 1.03692\ 77551$		t_5	$= 1.00452\ 37628$
	s_6	$= 1.01734\ 30620$		t_6	$= 1.00144\ 70767$
	s_7	$= 1.00834\ 92774$		t_7	$= 1.00047\ 15487$
	s_8	$= 1.00407\ 73562$		t_8	$= 1.00015\ 51790$
	s_9	$= 1.00200\ 83928$		t_9	$= 1.00005\ 13452$
	s_{10}	$= 1.00099\ 45751$		t_{10}	$= 1.00001\ 70413$

By means of the formula $t_k = s_k\left(1 - \frac{1}{2^k}\right)$, $k > 1$, t_k was calculated from s_k. From the well known results for the Zeta function of Riemann $\zeta(s)$, [22], (p. 265, p. 267),

$$\zeta_s = s_k = \sum_{k=1}^{\infty}\frac{1}{k^s} = \frac{1}{\Gamma(s)}\int_0^{\infty}\frac{x^{s-1}e^{-x}}{1-e^{-x}}\,dx, \qquad s \geqq 1, \qquad k > 1. \tag{7.7}$$

$$\sigma_s = \left(1 - \frac{1}{2^{s-1}}\right)\zeta(s) = \frac{1}{\Gamma(s)}\int_0^\infty \frac{x^{s-1}}{e^x + 1}\,dx, \qquad \text{and} \tag{7.8}$$

$$t_s = \zeta(s)\left(1 - \frac{1}{2^s}\right). \tag{7.9}$$

8. The mean of the z distribution. From our previous formulas for $\bar{z}$ we prove that if $n_1 = n_2$, $\bar{z} = 0$, and $\bar{z} < 0$ for $n_2 > n_1$, $\bar{z} > 0$ for $n_1 > n_2$. The maximum absolute value of $\lambda_{1:z}$ will occur when $n_1 = 1$, $n_2 = \infty$, or $n_1 = \infty$, $n_2 = 1$, and from (7.4) or (6.8) we have $\max |\lambda_{1:z}| = \frac{s_1}{2} + \frac{1}{2}\log 2 = .6352$.

9. Formulas for $\lambda_{2:z}$, $\mu_{2:z}$, $\lambda_{3:z}$, $\mu_{3:z}$, $\lambda_{4:z}$, and $\mu_{4:z}$. We have four cases from (5.6), (6.7), (7.1), (7.3):

$$\begin{aligned}\lambda_{2:z} &= \frac{1}{4}\left[2s_2 - \sum_{k=1}^{\frac{1}{2}(n_1-2)} \frac{1}{k^2} - \sum_{k=1}^{\frac{1}{2}(n_2-2)} \frac{1}{k^2}\right] \\ &= .822467 - \frac{1}{4}\left(\sum_{k=1}^{\frac{1}{2}(n_1-2)} \frac{1}{k^2} + \sum_{k=1}^{\frac{1}{2}(n_2-2)} \frac{1}{k^2}\right), \qquad n_1, n_2 \text{ even.}\end{aligned} \tag{9.1}$$

$$\lambda_{2:z} = 2.467401 - \frac{1}{4}\left(\sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{(k+\frac{1}{2})^2} + \sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{(k+\frac{1}{2})^2}\right), \qquad n_1, n_2 \text{ odd.} \tag{9.2}$$

$$\lambda_{2:z} = 1.644934 - \frac{1}{4}\left(\sum_{k=1}^{\frac{1}{2}(n_1-1)} \frac{1}{k^2} + \sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{(k+\frac{1}{2})^2}\right), \qquad n_1 \text{ even, } n_2 \text{ odd.} \tag{9.3}$$

$$\lambda_{2:z} = 1.644934 - \frac{1}{4}\left(\sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{(k+\frac{1}{2})^2} + \sum_{k=1}^{\frac{1}{2}(n_2-2)} \frac{1}{k^2}\right), \qquad n_1 \text{ odd, } n_2 \text{ even.} \tag{9.4}$$

In all cases of course $\lambda_{2:z} > 0$ and moreover $\lambda_{2:z} \to 0$ as n_1 and $n_2 \to \infty$. We list

$$\lambda_{3:z} = \frac{1}{4}\left(\sum_{k=1}^{\frac{1}{2}n_1-1} \frac{1}{k^3} - \sum_{k=1}^{\frac{1}{2}n_2-1} \frac{1}{k^3}\right), \qquad n_1, n_2 \text{ even.} \tag{9.5}$$

$$\lambda_{3:z} = \frac{1}{4}\left(\sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{(k+\frac{1}{2})^3} - \sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{(k+\frac{1}{2})^3}\right), \qquad n_1, n_2 \text{ odd.} \tag{9.6}$$

$$\lambda_{3:z} = 1.803085 + \frac{1}{4}\left(\sum_{k=1}^{\frac{1}{2}(n_1-2)} \frac{1}{k^3} - \sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{(k+\frac{1}{2})^3}\right), \qquad n_1 \text{ even, } n_2 \text{ odd.} \tag{9.7}$$

$$\lambda_{3:z} = -1.803085 + \frac{1}{4}\left(\sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{(k+\frac{1}{2})^3} - \sum_{k=1}^{\frac{1}{2}(n_2-2)} \frac{1}{k^3}\right), \qquad n_1 \text{ odd, } n_2 \text{ even.} \tag{9.8}$$

$$\lambda_{4:z} = .811742 - \frac{3}{8}\left(\sum_{k=1}^{\frac{1}{2}n_2-1} \frac{1}{k^4} + \sum_{k=1}^{\frac{1}{2}n_1-1} \frac{1}{k^4}\right), \qquad n_1, n_2 \text{ even.} \tag{9.9}$$

$$\lambda_{4:z} = 12.17614 - 6\left(\sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{(2k+1)^4} + \sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{(2k+1)^4}\right), \qquad n_1, n_2 \text{ odd.} \tag{9.10}$$

$$(9.11)\quad \lambda_{4:z} = 6.493939 - 6\left(\sum_{k=0}^{\frac{1}{2}(n_2-3)} \frac{1}{(2k+1)^4} + \sum_{k=1}^{\frac{1}{2}n_1-1} \frac{1}{k^4}\right), \qquad n_1 \text{ even, } n_2 \text{ odd.}$$

$$(9.12)\quad \lambda_{4:z} = 6.493939 - 6\left(\sum_{k=1}^{\frac{1}{2}n_2-2} \frac{1}{k^4} + \sum_{k=0}^{\frac{1}{2}(n_1-3)} \frac{1}{(2k+1)^4}\right), \qquad n_1 \text{ odd, } n_2 \text{ even.}$$

We see $\lambda_{r:z} > 0$ whenever r is even. If r is odd $\lambda_{r:z} < 0$ if $n_2 > n_1$, and $\lambda_{r:z} > 0$ if $n_1 > n_2$. Also $\mu_{r:z} > 0$, $n_1 > n_2$, r odd, greater than one. Similarly $\mu_{r:z} < 0$, r odd > 1, $n_2 > n_1$.

10. Skewness, excess, and values of α_n. We take for our measure of skewness $\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\lambda_3}{\lambda_2^{3/2}}$. For $n_2 > n_1$, $\alpha_3 < 0$. Further the skewness increases negatively if n_1 remains constant as $n_2 \to \infty$. Thus negative skewness will be a maximum for $n_2 = \infty$, $n_1 = 1$, and positive skewness will be a maximum when $n_2 = 1$, $n_1 = \infty$. The absolute value of maximum α_3 is

$$(10.1)\qquad | \alpha_3 | = \left| \frac{2t_3}{t_2^{3/2}} \right| = 1.5351.$$

As our measure of kurtosis we use $\alpha_4 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{\lambda_4}{\lambda_2^2}$. As a measure of excess, E, we use $E = \alpha_4 - 3 = \frac{\lambda_4}{\lambda_2^2}$. The excess is always positive.

11. Approximations for $\lambda_{r:z}$ by the Euler-Maclaurin sum formula. The exact results given previously for the seminvariants become unwieldy for n_1 and n_2 large. Hence we develop useful approximations for the seminvariants, and give the maximum error of the approximation. We find first our results for $\lambda_{r:z}$ when n_1 and n_2 are even and $r > 1$. We begin with (5.6)

$$\lambda_{r:z} = \frac{(r-1)!}{2^r}\left\{\left(s_r - \sum_{k=1}^{\frac{1}{2}n_2-1} \frac{1}{k^r}\right) + (-1)^r\left(s_r - \sum_{k=1}^{\frac{1}{2}n_1-1} \frac{1}{k^r}\right)\right\}$$

and rewrite this as

$$(11.1)\qquad \lambda_{r:z} = \frac{(r-1)!}{2^r}\left\{\sum_{k=\frac{1}{2}n_2}^{\infty} \frac{1}{k^r} + (-1)^r \sum_{k=\frac{1}{2}n_1}^{\infty} \frac{1}{k^r}\right\}.$$

Now find the two sums of (11.1) by the Euler-Maclaurin sum formula [21] using the first three terms, and obtain

$$(11.2)\qquad \begin{aligned} \lambda_{r:z} = \frac{(r-2)!}{2}\Bigg[&\left(\frac{n_2+r-1}{n_2^r} + (-1)^r \frac{n_1+r-1}{n_1^r}\right) \\ &+ \frac{r(r-1)}{3}\left(\frac{1}{n_2^{r+1}} + \frac{(-1)^r}{n_1^{r+1}}\right) \\ &- \frac{r(r-1)(r+1)(r+2)}{45}\left(\frac{1}{n_2^{r+3}} + \frac{(-1)^r}{n_1^{r+3}}\right)\Bigg]. \end{aligned}$$

We use the following theorem [10] (p. 539), to find the error:

If $f(x)$ is of constant sign for $x > 0$, and together with all of its derivatives, tends monotonely to zero as $x \to \infty$, Euler's summation formula may be stated in the simplified form

$$\sum_{x=0}^{n} f_x = \int_0^n f(x)\,dx + \tfrac{1}{2}(f_n + f_0) + \frac{B_2}{2!}(f_n' - f_0') + \cdots$$
$$+ \frac{(-1)^{k-1}B_{2k}}{(2k)!}(f_n^{(2k-1)} - f_0^{(2k-1)}) + \frac{\theta B_{2k+2}}{(2k+2)!}(f_n^{(2k+1)} - f_0^{(2k+1)})$$

where $0 < \theta < 1$ and $B_2 = 1/6$, $B_4 = 1/30$, $B_6 = 1/42$, $B_8 = 1/30$, $B_{10} = 5/66$, etc. If we use

$$\lambda_{r:z} = \frac{(r-2)!}{2}\left(\frac{n_2 + r - 1}{n_2^r} + (-1)^r \frac{n_1 + r - 1}{n_1^r}\right), \tag{11.3}$$

then the error committed is of the same sign and less than

$$\frac{r!}{3!}\left\{\frac{1}{n_2^{r+1}} + \frac{(-1)^r}{n_1^{r+1}}\right\}.$$

If we take

$$\lambda_{r:z} = \frac{(r-2)!}{2}\left[\left(\frac{n_2 + r - 1}{n_2^r} + (-1)^r \frac{n_1 + r - 1}{n_1^r}\right) + \frac{r(r-1)}{3}\left(\frac{1}{n_2^{r+1}} + \frac{(-1)^r}{n_1^{r+1}}\right)\right], \tag{11.4}$$

then our error is less than, and has the same sign as

$$-\frac{(r+2)!}{90}\left\{\frac{1}{n_2^{r+3}} + \frac{(-1)^r}{n_1^{r+3}}\right\}.$$

Finally if we use (11.2), our error has the same sign as, and is less than

$$\frac{(r+4)!}{945}\left\{\frac{1}{n_2^{r+5}} + \frac{(-1)^r}{n_1^{r+5}}\right\}.$$

12. Approximations for other values of n_1 and n_2, $r > 1$. Now in case n_1 and n_2 are odd we have from (6.7)

$$\lambda_{r:z} = (r-1)!\left\{\sum_{k=\frac{1}{2}(n_2-1)}^{\infty} \frac{1}{(2k+1)^r} + (-1)^r \sum_{k=\frac{1}{2}(n_1-1)}^{\infty} \frac{1}{(2k+1)^r}\right\}. \tag{12.1}$$

Applying the Euler-Maclaurin sum formula to each of the sums in (12.1) we are led to exactly the same results given in paragraph (11). The other cases are obvious combinations of the sums in (11.1) and (12.1), and so for all values of n_1 and n_2 the approximate results for $\lambda_{r:z}$, $r > 1$ are

$$\lambda_{r:z} = \frac{(r-2)!}{2}\left\{\frac{n_2+r-1}{n_2^r} + (-1)^r\frac{n_1+r-1}{n_1^r}\right\} \tag{12.2}$$
$$+\frac{r!}{6}\left\{\frac{1}{n_2^{r+1}} + \frac{(-1)^r}{n_1^{r+1}}\right\} - \frac{(r+2)!}{90}\left\{\frac{1}{n_2^{r+3}} + \frac{(-1)^r}{n_1^{r+3}}\right\}.$$

Formulas (11.1) and (12.1) prove the result previously given for $\lambda_{r:z}$ (4.7).

13. The approximate values of $\lambda_{1:z}$. From (5.7)

$$\lambda_{1:z} = \frac{1}{2}\left[\left(\log n_2 - \sum_{k=1}^{\frac{1}{2}n_2-1}\frac{1}{k}\right) - \left(\log n_1 - \sum_{k=1}^{\frac{1}{2}n_1-1}\frac{1}{k}\right)\right], \qquad n_1 \text{ and } n_2 \text{ even.}$$

We use the Euler-Maclaurin sum formula on the sum

$$\sum_{k=1}^{\frac{1}{2}n_2-1}\frac{1}{k} = \left\{\sum_{k=0}^{\frac{1}{2}n_2-1}\left(\frac{1}{k+1}\right) - \frac{2}{n_2}\right\}$$

and the similar sum involved in $\lambda_{1:z}$. Hence we have

$$\lambda_{1:z} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) + \frac{1}{6}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right) - \frac{1}{15}\left(\frac{1}{n_2^4} - \frac{1}{n_1^4}\right), \qquad n_1, n_2 > 2. \tag{13.1}$$

The errors committed by using one, two, or three terms of (13.1) are less than, and of the same sign respectively as

$$\frac{1}{6}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right), \quad -\frac{1}{15}\left(\frac{1}{n_2^4} - \frac{1}{n_1^4}\right), \quad \frac{8}{63}\left(\frac{1}{n_2^6} - \frac{1}{n_1^6}\right).$$

For n_1 and n_2 both odd we find the same result as (13.1). The restriction n_1, $n_2 > 2$, may easily be replaced by $n_1, n_2 \geqq 2$ (for n_1, n_2 even) and $n_1, n_2 \geqq 1$ (for n_1, n_2 both odd). When n_1 is odd, n_2 even, the formula is again the same as (13.1) if n_1 and n_2 are sufficiently large; but if n_1 and n_2 are small we find in this case

$$\lambda_{1:z} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) + \frac{1}{6}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right) - \frac{1}{15}\left(\frac{1}{n_2^4} - \frac{1}{n_1^4}\right) \tag{13.2}$$
$$+\frac{1}{2}\left(1 - \frac{1}{2}\right) + \frac{1}{6}\left(1 - \frac{1}{4}\right) - \frac{1}{15}\left(1 - \frac{1}{16}\right) - \frac{1}{2}\log 2.$$

Another method of finding (12.2) would have been to use the asymptotic expression for $\log \Gamma(x)$.

14. Approximate values of $\lambda_{r:z}$ for values of r. We list the approximate values of $\lambda_{r:z}$ to three terms.

$$\lambda_{1:z} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) + \frac{1}{6}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right) - \frac{1}{15}\left(\frac{1}{n_2^4} - \frac{1}{n_1^4}\right)$$

$$(14.1)\quad \begin{aligned}
\lambda_{2:z} &= \frac{1}{2}\left(\frac{n_2+1}{n_2^2} + \frac{n_1+1}{n_1^2}\right) + \frac{1}{3}\left(\frac{1}{n_2^3} + \frac{1}{n_1^3}\right) - \frac{4}{15}\left(\frac{1}{n_2^5} + \frac{1}{n_1^5}\right)\\
\lambda_{3:z} &= \frac{1}{2}\left(\frac{n_2+2}{n_2^3} - \frac{n_1+2}{n_1^3}\right) + \left(\frac{1}{n_2^4} - \frac{1}{n_1^4}\right) - \frac{4}{3}\left(\frac{1}{n_2^6} - \frac{1}{n_1^6}\right)\\
\lambda_{4:z} &= \left(\frac{n_2+3}{n_2^4} + \frac{n_1+3}{n_1^4}\right) + 4\left(\frac{1}{n_2^5} + \frac{1}{n_1^5}\right) - 8\left(\frac{1}{n_2^7} + \frac{1}{n_1^7}\right)\\
\lambda_{5:z} &= 3\left(\frac{n_2+4}{n_2^5} - \frac{n_1+4}{n_1^5}\right) + 20\left(\frac{1}{n_2^6} - \frac{1}{n_1^6}\right) - 56\left(\frac{1}{n_2^8} - \frac{1}{n_1^8}\right)\\
\lambda_{6:z} &= 12\left(\frac{n_2+5}{n_2^6} + \frac{n_1+5}{n_1^6}\right) + 120\left(\frac{1}{n_2^7} + \frac{1}{n_1^7}\right) - 448\left(\frac{1}{n_2^9} + \frac{1}{n_1^9}\right).
\end{aligned}$$

The approximate values given by Cornish and Fisher [8] (p. 319), are similar, but have fewer terms. Cornish and Fisher give no remainder term. From (14.1) and (12.2) we see the maximum absolute values of $\lambda_{2r+1:z}$, $r \geqq 1$, occur when $n_2 = \infty$, $n_1 = 1$, or $n_2 = 1$, $n_1 = \infty$. Similarly $\lambda_{2r:z}$, $r \geqq 1$, has its maximum value for $n_1 = n_2 = 1$. The standard seminvariants of z are defined $\xi_{r:z} = \frac{\lambda_r}{\lambda_2^{\frac{1}{2}r}}$, $r \geqq 2$. We also note that for $n_2 > n_1$, $\xi_{2r+1:z} < 0$, $r \geqq 1$ and hence $\alpha_{2r+1} < 0$ also where $\alpha_n = \frac{\mu_n}{\mu_2^{\frac{1}{2}n}}$. Moreover the maximum absolute values of $\xi_{2r:z}$ and $\xi_{2r+1:z}$ occur when $n_1 = 1$, $n_2 = \infty$ or $n_2 = 1$, $n_1 = \infty$; and also for α_{2r} and α_{2r+1}. Approximately then

$$(14.2)\qquad \max \xi_{r:z} = (-1)^r \frac{(r-1)!}{2}, \qquad r \geqq 2.$$

The exact value for maximum $\alpha_{4:z}$ is $3 + \frac{6t_4}{t_2^2} = 7.07$.

15. Approach to normality of the *z* distribution. We prove the theorem: The distribution of z approaches normality as n_1 and $n_2 \rightarrow \infty$ in any manner whatever, with $\bar{z} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right)$, $\sigma_z^2 = \frac{1}{2}\left(\frac{1}{n_2} + \frac{1}{n_1}\right)$. We also find an upper bound of the absolute value of the difference between the z distribution and the function determined by the approximate seminvariants of z when n_1 and n_2 become large. To prove the theorem we start with the original distribution of z, and find when n_1 and n_2 are large,

$$(15.1)\qquad P(z) = \frac{1}{\sqrt{2\pi}\,\sigma_z}\left\{\frac{n_1+n_2}{n_1e^{2z}+n_2}\right\}^{\frac{1}{2}(n_1+n_2)} e^{n_1z}\,dz.$$

We change to standard units $z = t\sigma_z + \bar{z}$, then

$$(15.2)\qquad P(t) = \frac{1}{\sqrt{2\pi}}\left\{\frac{n_1 + n_2}{n_1 e^{2t\sigma+2\bar{z}} + n_2}\right\}^{\frac{1}{2}(n_1+n_2)} e^{n_1 t\sigma + n_1 \bar{z}}\, dt, \qquad -\infty < t < \infty.$$

We rewrite this as

$$(15.3)\qquad P(t) = \frac{1}{\sqrt{2\pi}}\left\{\frac{n_1 + n_2}{n_1 e^{2n_2(t\sigma+\bar{z})/(n_1+n_2)} + n_2 e^{-2n_1(t\sigma+\bar{z})/(n_1+n_2)}}\right\}^{\frac{1}{2}(n_1+n_2)} dt.$$

Expand $n_1 e^{2n_2(t\sigma+\bar{z})/(n_1+n_2)}$ and $n_2 e^{-2n_1(t\sigma+\bar{z})/(n_1+n_2)}$ and add term by term. Divide this result by $n_1 + n_2$ from the numerator of $P(t)$ to obtain

$$(15.4)\qquad 1 + \frac{2n_1 n_2(t\sigma + \bar{z})^2}{(n_1 + n_2)^2} + O_1\left\{\frac{1}{(n_1 + n_2)^{\frac{3}{2}}}\right\}.$$

Hence

$$(15.5)\qquad P(t) = \frac{1}{\sqrt{2\pi}}\left\{1 + \frac{2n_1 n_2(t\sigma + \bar{z})^2}{(n_1 + n_2)^2}\right\}^{-\frac{1}{2}(n_1+n_2)} dt.$$

We evaluate (15.5) for n_1 and n_2 large by using logarithms.

$$-\frac{n_1 + n_2}{2}\log\left\{1 + \frac{2n_1 n_2(t\sigma + \bar{z})^2}{(n_1 + n_2)^2}\right\}$$
$$= -\frac{n_1 + n_2}{2}\left[\left\{\frac{2n_1 n_2(t\sigma + \bar{z})^2}{(n_1 + n_2)^2}\right\} - \frac{1}{2}\left\{\frac{2n_1 n_2(t\sigma + \bar{z})^2}{(n_1 + n_2)^2}\right\}^2 + \sum_{r=3}^{\infty}\frac{(-1)^{r+1}}{r}\left\{\frac{2n_1 n_2(t\sigma + \bar{z})^2}{(n_1 + n_2)^2}\right\}^r\right].$$

This gives

$$-\frac{\sigma^{-2}}{2}(t^2\sigma^2 + 2t\sigma\bar{z} + \bar{z}^2) + \frac{n_1^2 n_2^2}{(n_1 + n_2)^3}(t\sigma + \bar{z})^4 + \sum_{r=3}^{\infty}(-1)^r\frac{\{2n_1 n_2(t\sigma + \bar{z})^2\}^r}{2r(n_1 + n_2)^{2r-1}}.$$

We rèduce this then to

$$-\frac{t^2}{2} - \sigma^{-1}\bar{z}t - \frac{(\bar{z}\sigma^{-1})^2}{2} + \frac{1}{2}\left\{\frac{2n_1^2 n_2^2}{(n_1 + n_2)^2}\right\}\frac{(t\sigma + \bar{z})^4}{n_1 + n_2}$$

+ terms involved in the above summation. Let $U = \sigma^{-1}\bar{z} < \sigma$. Since $\lim_{n_1,n_2\to\infty} \sigma = 0$, $\lim_{n_1,n_2\to\infty} U = 0$. Similarly $\lim_{n_1,n_2\to\infty} \frac{\bar{z}^2\sigma^{-2}}{2} = \lim_{n_1,n_2\to\infty} \frac{U^2}{2} = 0$. Consider $\frac{n_1^2 n_2^2}{(n_1 + n_2)^3}(t\sigma + \bar{z})^4 = \frac{\sigma^{-4}(t\sigma + \bar{z})^4}{4(n_1 + n_2)} = \frac{(t + U)^4}{4(n_1 + n_2)}$. Hence $\lim_{n_1,n_2\to\infty}\frac{(t + U)^4}{4(n_1 + n_2)} = 0$. In like fashion

$$\sum_{r=3}^{\infty}\frac{(-1)^r}{2r}\left\{\frac{2n_1 n_2}{n_1 + n_2}\right\}^r\frac{(t\sigma + \bar{z})^{2r}}{(n_1 + n_2)^{r-1}} = \sum_{r=3}^{\infty}\frac{(-1)^r\sigma^{-2r}(t\sigma + \bar{z})^{2r}}{2r(n_1 + n_2)^{r-1}}.$$

Now clearly from our previous discussion for $r = 2$, we see

$$\lim_{n_1, n_2 \to \infty} \sum_{r=3}^{\infty} \frac{(-1)^r}{2r} \frac{\sigma^{-2r}(t\sigma + \bar{z})^{2r}}{(n_1 + n_2)^{r-1}} = 0.$$

This completes the proof.

We now consider the function, $f(z)$, determined by the approximate seminvariants of z. We start with

$$\lambda_{1:z} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) \text{ and } \lambda_{r:z} = \frac{(r-2)!}{2}\left\{\frac{n_2 + r - 1}{n_2^r} + (-1)^r \frac{n_1 + r - 1}{n_1^r}\right\}, \quad r > 1,$$

from (12.2) using only the first term. We may easily prove then that as n_1 and n_2 approach infinity in any manner whatever the function $f(z)$ represents a normal frequency distribution with

$$\bar{z} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) \text{ and } \mu_{2:z} = \frac{1}{2}\left(\frac{n_2 + 1}{n_2^2} + \frac{n_1 + 1}{n_1^2}\right).$$

This further shows the identity of $f(z)$ and $y(z)$ in the limit as n_1 and $n_2 \to \infty$.

Since the moment generating function of $f(z)$ is

$$\left(1 - \frac{\theta}{n_2}\right)^{\frac{1}{2}(n_2 - 1 - \theta)} \left(1 + \frac{\theta}{n_1}\right)^{\frac{1}{2}(n_1 - 1 + \theta)}$$

we have

$$(15.6) \qquad f(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\theta z} \left(1 - \frac{i\theta}{n_2}\right)^{\frac{1}{2}(n_2 - 1 - i\theta)} \left(1 + \frac{i\theta}{n_1}\right)^{\frac{1}{2}(n_1 - 1 + i\theta)} d\theta.$$

I have not been able to evaluate (15.6). We instead shall find an upper bound to the difference $| f(z) - y(z) |$ as n_1 and n_2 become large. We form $f(z) - y(z)$. Then by use of Stirling's formula for $n!$ with the remainder term and by the Fourier Integral Theorem,

$$(15.7) \qquad | f(z) - y(z) | \leqq (e^{\beta_3/6n_1 + \beta_4/6n_2} - 1)y(z) \text{ where } 0 < \beta_3 < 1, 0 < \beta_4 < 1,$$

and

$$(15.8) \qquad \lim_{n_1, n_2 \to \infty} | f(z) - y(z) | = 0, \text{ and for this case } f(z) = y(z).$$

Of course (15.7) furnishes the upper bound of the absolute value between the frequency distribution of z and the function determined by the approximate seminvariants of z for any values of n_1 and n_2.

Up to this point we have assumed that there exists a function determined by the seminvariants

$$\lambda_{1:z} = \frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right) \text{ and } \lambda_{r:z} = \frac{(r-2)!}{2}\left\{\frac{n_2 + r + 1}{n_2^r} + (-1)^r \frac{n_1 + r - 1}{n_1^r}\right\}, \quad r > 1.$$

This may readily be proved by using the following theorem [18] (p. 536): The determined character of the moments problem for an infinite interval is insured if $\sum_{n=1}^{\infty} c_{2n}^{-1/2n}$ diverges $\left(c_n = \int_{-\infty}^{\infty} x^n\, dF(x)\right)$.

16. The Pearson types of approximating curve. In discussing the types of the Pearson system which may be expected to approximate the z distribution we shall use the results of H. C. Carver [1], and the further exposition of C. C. Craig [3]. To find the Pearson type we compute $\delta = \dfrac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3}$. We shall find it convenient to use the approximations $\alpha_3 = \dfrac{\sqrt{2}\,(n_1 - n_2)}{\sqrt{n_1 n_2(n_1 + n_2)}}$ and $\alpha_4 = 3 + 4\,\dfrac{(n_1^2 - n_1 n_2 + n_2^2)}{n_1 n_2(n_1 + n_2)}$ to obtain

$$(16.1) \qquad \delta = \frac{(n_1 + n_2)^2}{3n_1^2 n_2 + 3n_1 n_2^2 + 2n_1^2 - 2n_1 n_2 + 2n_2^2},$$

and consequently $0 < \delta \leqq \frac{1}{2}$. The only possibilities are Types IV, VII, VI, or V since the greatest value of α_3^2 by (14.1) is 2.3565. Now if $n_1 = n_2$, we have Type VII, since $\alpha_3 = 0$, $\delta > 0$. In all other cases we shall have Types IV, V, or VI according as $\alpha_3^2 < 4\delta(\delta + 2)$, $\alpha_3^2 = 4\delta(\delta + 2)$, $\alpha_3^2 > 4\delta(\delta + 2)$. We neglect δ^2. Hence $\alpha_3^2 < 8\delta$ implies

$$(16.2) \qquad \begin{aligned} n_2^4(n_1 - 2) + n_2^3(15n_1^2 + 6n_1) + n_2^2(15n_1^3 - 8n_1^2) \\ + n_2(n_1^4 + 6n_1^3) - 2n_1^4 > 0. \end{aligned}$$

A simple investigation reveals then the following results:

$$(16.3) \qquad \begin{aligned} &\text{Type IV for } n_1,\ n_2 \geqq 2,\ n_1 \neq n_2. \\ &\text{Type IV for } n_1 = 1,\ 1 \leqq n_2 \leqq 21;\ \text{or } n_2 = 1,\ 1 \leqq n_1 \leqq 21. \\ &\text{Type VI for } n_1 = 1,\ n_2 > 22. \\ &\qquad\quad\ \text{for } n_2 = 1,\ n_1 > 22. \\ &\text{Type VII for } n_1 = n_2. \end{aligned}$$

Clearly the z distribution has features comparable to Type IV since both have infinite range. However, Type IV is irksome to fit in practice.

17. The Type III approximating curve, the logarithmic curve, and the Gram-Charlier Type A. The criterion for Type III is $\delta = 0$, $\alpha_3 \neq 0$. We see that as n_1 and n_2 increase the value of δ will decrease. Even for small values of n_1 and n_2 Type III will furnish a fair approximation to the z distribution. For example $n_1 = 10$, $n_2 = 5$, $\delta = .094$. The advantage of the Type III approximation rests on the fact that Salvosa's tables may be used. From the chart in [16] since $\alpha_3^2 \leqq 2.3565$, we are assured that the approximating Type III curve is bell shaped. For $n_1 = 1, 2, n_2 =$ any value, this approximation is not all that could be desired, although even in such cases it does have value. We note that Type III has limited range at one extreme $\left(-\frac{2}{\alpha_3}, \infty\right)$ while the range of the z distribution is $(-\infty, \infty)$. Salvosa's tables extend as far as $\alpha_3 = 1.1$, and since max $\alpha_3 = 1.5351$, we see in some cases, and these only for $n_1 = 1$, n_2 large, we shall be obliged to make use of Pearson's *Tables of the Incomplete Gamma Function* [14]. The logarithmic frequency curve

$$f(u) = \frac{1}{\sqrt{2\pi}\, c(u - a)} \exp\left[-\frac{1}{2c^2}\left(\log\frac{u - a}{b}\right)^2\right]$$

will be useful in approximating the z distribution. While it has been discussed by many authors we shall follow Pae-Tsi Yuan [23], where a full bibliography may be found. In our discussion we use the $\beta_1 = \alpha_3^2$, $\beta_2 = \alpha_4$ chart of the Pearson system as given by S. J. Pretorius [16] (p. 147), since the logarithmic frequency locus connecting α_3^2 and α_4 is already drawn in. The justification of this curve for fitting is due to the fact that in the β_1, β_2 chart of the Pearson system as given by S. J. Pretorius [16] (p. 147), the logarithmic frequency locus lies in the Type VI region between the Type III locus and the Type V locus, and consequently closer to the Type IV region than Type III itself does. Hence since Type III fits fairly well under certain conditions and Type IV fits well we can expect the same for the logarithmic curve. Furthermore when α_3 is small the logarithmic curve is similar to Type III [23] (p. 42), and as α_3 becomes larger, $\alpha_3 = 1$, the difference between the two types is pronounced. However, it is just when α_3 becomes large in the region $n_1 = 1$, $n_2 \geqq 22$ that we find the logarithmic curves give a fine fit, since in such cases the point (α_3^2, β_2) lies practically on the logarithmic locus [16]. To fit the curve [23] (pp. 37, 48, 49), we find the values of the three parameters a, b, c. To find c we solve the equation $w^3 + 3w^2 - (4 + \alpha_{3:z}^2) = 0$ for w using the table [23] (p. 48) given by Pae-Tsi Yuan. Knowing w we can easily solve for

$$(17.1) \qquad \begin{aligned} c &= (\log w)^{\frac{1}{2}}, & b &= \left(\frac{w + 2}{\alpha_{3:z}}\right) w^{-\frac{1}{2}} \sigma_z, \\ a &= \bar{z} - \frac{(w + 2)\sigma_z}{\alpha_{3:z}}, & t &= \frac{z - \bar{z}}{\sigma_z} = \frac{e^{xc - \frac{1}{2}c^2} - 1}{(e^{c^2} - 1)^{\frac{1}{2}}} \end{aligned}$$

where the value of x must be obtained from the table of areas under the normal curve, if the z distribution is approximated by use of areas.

Since the Gram-Charlier Type A series generally approximates a Pearson Type IV fairly well when α_3^2 is not too large, it is to be expected that the Type A series will approximate the z distribution in those cases when $n_1 = n_2$, and also when α_3^2 is not too large.

18. Levels of significance and approximation methods. We shall apply the results of the previous paragraphs to the determination of the value of z for any level of significance α, i.e. the value of z such that $\int_{-\infty}^{z} y(z)\, dz = 1 - \alpha$. We have such levels as the median (the 50% point of significance), the 20%, 5%, 1%, and .1% points as given in [9]. Where these tables apply there is no need for other methods. It would be desirable to extend the results for any level of significance whatever. The methods which we shall use are (1) the logarithmic frequency curve, (2) the Gram-Charlier Type A, and (3) the Type III approximation. For finding the levels of significance by the Incomplete Beta function, the reader is referred to [13], (p. lviii, topic (viii)). The logarithmic curve is very simple to use in conjunction with the table of areas under the normal curve. From Pae-Tsi Yuan we have

$$(18.1) \qquad t = \frac{e^{xc - \frac{1}{2}c^2} - 1}{(e^{c^2} - 1)^{\frac{1}{2}}}, \qquad \text{where } (e^{c^2} - 1)^{\frac{1}{2}}$$

takes the same sign as α_3. The value of x is obtained from the table of the normal curve, 1.64 for the 5% level, 2.33 for the 1% level; the value of c is obtained from w (17.1), and consequently the value of t (18.1). Then we have if z_α = value of z for any level of significance, $t = \frac{z_\alpha - \bar{z}}{\sigma_z}$ to solve for z_α, where $\bar{z}$, and σ_z are the values of the mean and standard deviation of z as given by the proper formulas in (5), (6), (7). We illustrate with examples:

(18.2) 5% point of z, $n_1 = \infty$, $n_2 = 1$. $\alpha_3 = 1.5351$, $w = 1.2264$, $x = 1.64$, $t = 1.88$, $\bar{z} = .6352$, $\sigma_z = 1.11$, and as a result $z_{5\%} = 2.72$. Fisher [9] gives 2.7693.

We can also find $z_{5\%}$ easily for $n_1 = 1$, $n_2 = \infty$. Here $\alpha_3 = -1.5351$, $w = 1.2264$, $x = -1.64$, $t = 1.197$, $\bar{z} = -.6352$, $\sigma_z = 1.11$, $z_{5\%} = .694$ compared with Fisher [9] $z_{5\%} = 6729$.

(18.3) 1% point for $n_1 = 4$, $n_2 = 8$, $\bar{z} = -.0701$, $\sigma_z = .4819$, $\alpha_{3:z} = -.3619$, $w = 1.0144$, $t = 2.17$ and $z_{1\%} = .976$, while the accurate result is .9734.

From experience the values of z for any level of significance obtained by the logarithmic frequency curve will possess an error less than 2% of the true value of z for the level of significance if n_1 and n_2 are greater than twenty. It would seem that for other values of n_1 and n_2 the error could not be greater than 10%, and usually would be much less.

19. The Gram-Charlier Type A. We take the series in the form

$$F(t) = \varphi(t) + A_3\varphi^{(3)}(t) + A_4\varphi^{IV}(t), \qquad \varphi(t) = \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}}$$

$$t = \frac{z - \bar{z}}{\sigma_z}, \qquad A_3 = \frac{-\alpha_{3:z}}{3!}, \qquad A_4 = \frac{\alpha_{4:z}}{4!}.$$

Some examples follow.

(19.1) We use the material of (18.3) and employ three terms of $F(t)$. $\bar{z} = -.0701$, $\sigma_z = .4819$, $\lambda_{3:z} = -.0405$, $\lambda_{4:z} = .0336$, $A_3 = .06032$, $A_4 = .02596$.

Fitting $F(t)$ by ordinates we have $t = 2.17$, and consequently $z = .976$.

(19.2) We take $n_1 = n_2 = 5$, $\bar{z} = 0$, $\sigma_z = .4952$, $\lambda_{3:z} = 0$, $\lambda_{4:z} = .02798$, $A_3 = 0$, $A_4 = .01939$.

5% point: By ordinates $t = 1.57$, $z_{5\%} = .777$, while Fisher gives .8097.
1% point: By ordinates $t = 2.325$, $z_{1\%} = 1.15$, while Fisher gives 1.1974.

(19.3) We take $n_1 = 3$, $n_2 = 20$, $\bar{z} = -.15909$, $\sigma_z = .5099$, $\lambda_{3:z} = -.10222$, $\lambda_{4:z} = .08822$, $A_3 = .12854$, $A_4 = .05438$. By ordinates $t = 1.523$, $z_{5\%} = .618$, Fisher gives .5654. $t = 1.989$, $z_{1\%} = .855$, Fisher gives .7985. The Gram-Charlier Type A is recommended only for $n_1 = n_2$ and n_1, $n_2 \geqq 20$.

20. Type III approximation, the median, and 5% point. Since for Type III the median, m_z, is approximately two-thirds of the distance from the mode to the median if α_3 is moderate [12], [6], then we have further assuming n_1, $n_2 \geqq 20$.

$$m_z = \frac{1}{3}\left(\frac{1}{n_1} - \frac{1}{n_2}\right) + \frac{1}{9}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right). \tag{20.1}$$

From experience this result will furnish an accuracy with an error less than 2% of the true value in the range above indicated.

$$t_{5\%} = 1.6437 + .2760\alpha_3 - .04506\alpha_3^2. \tag{20.2}$$

This was found by use of Salvosa's tables and for $\alpha_3 > 1.1$ by [14].

$$z_{5\%} = \sigma_z[1.644 + .2760\alpha_{3:z} - .0451\alpha_{3:z}^2] + \bar{z}. \tag{20.3}$$

We illustrate the use of (20.3) with some examples.

(20.4) $n_1 = n_2 = 1$, $\sigma_z = 1.5706$, $\alpha_{3:z} = 0$, $\bar{z} = 0$, $z_{5\%} = 2.582$,

while the accurate value is $z_{5\%} = 2.5421$.

(20.5) $n_1 = \infty$, $n_2 = 1$, $\alpha_3 = 1.5351$, $\bar{z} = .6352$, $\sigma_z = 1.11$, $z_{5\%} = 2.81$. The accurate value is 2.7693.

(20.6) $n_1 = n_2 = 5$, $\sigma_z = .4952$, $\alpha_{3:z} = 0$, $\bar{z} = 0$, $z_{5\%} = .8141$, while the accurate value is $z_{5\%} = .8097$.

(20.7) $n_1 = 4$, $n_2 = 8$, $\bar{z} = -.0701$, $\sigma_z = .4819$, $\alpha_3 = -.3619$, $z_{5\%} = .6712$, while the accurate value is .6725.

(20.8) $n_1 = 1$, $n_2 = 10$, $\bar{z} = -.5835$, $\sigma_z = 1.1353$, $\alpha_3 = -1.4333$, $z_{5\%} = .7283$, while the accurate value is .8012.

In a future paper exactly the same methods will be used for any per cent point of z whatever in order to compare with the results of W. G. Cochran [2]. If

n_1 and n_2 are large we may use the approximate formulas for σ_z, $\alpha_{3:z}$, and $\bar{z}$ to obtain to the order of σ_z^3,

$$(20.9)\quad z_{5\%} = 1.644\sigma_z + .7760\left(\frac{1}{n_2} - \frac{1}{n_1}\right), \quad \text{where} \quad \sigma_z = \sqrt{\frac{1}{2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right)}.$$

We expand Fisher's result [9]
$z_{5\%} = \dfrac{1.6449}{\sqrt{n-1}} + .7843\left(\dfrac{1}{n_2} - \dfrac{1}{n_1}\right)$ by the binomial theorem, where $h = \dfrac{1}{\sigma_z^2}$, to obtain a comparable result

$$(20.10)\qquad z_{5\%} = 1.645\sigma_z + .7843\left(\frac{1}{n_2} - \frac{1}{n_1}\right).$$

The numerical examples given in this chapter illustrate unfavorable cases as well as favorable ones.

21. The distribution of *F*. Historically Snedecor [19] was the first to use F for e^{2z}. We find

$$(21.1)\qquad P(F) = \frac{n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2}}{\mathrm{B}\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \frac{F^{\frac{1}{2}n_1 - 1}}{(n_1 F + n_2)^{\frac{1}{2}(n_1+n_2)}} dF, \qquad 0 \leqq F \leqq \infty.$$

The distribution of F is J shaped if $n_1 \leqq 2$, and bell shaped for $n_1 > 2$, and for $n_1 > 2$ one mode exists, $F_0 = \dfrac{n_2(n_1 - 2)}{n_1(n_2 + 2)}$. The two points of inflection, which exist for $n_1 \geqq 4$, are equidistant from the mode. The moments are

$$\mu_m' = \left(\frac{n_2}{n_1}\right)^m \frac{\Gamma\left(\frac{n_1 + 2m}{2}\right)\Gamma\left(\frac{n_2 - 2m}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}, \qquad n_2 > 2m$$

$$\bar{F} = \frac{n_2}{n_2 - 2}, \qquad n_2 > 2, \qquad \mu_2 = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)} \sim 2\left(\frac{1}{n_1} + \frac{1}{n_2}\right),$$

$$\alpha_{3:F} \sim \frac{2\sqrt{2}(2n_1 + n_2)}{\sqrt{n_1 n_2(n_1 + n_2)}}.$$

The exact results for μ_3, μ_4, α_3, and α_4 are omitted because of length. We have the theorem that as $n_1, n_2 \to \infty$ in any manner whatever the distribution of F approaches normality with mean $\bar{F} = 1$, $\sigma_F = \sqrt{2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$. The proof is omitted. The only type of approximating curve of any value is Type III. Of course the distribution of F is Type VI. No tables exist for Type VI. Furthermore the F distribution approaches the Type III function so slowly as to make most approximations of little value unless $\alpha_{3:F} \leqq 1.1$. Other possible parameters are $\theta = \frac{n_1(n_2 + 1)}{n_2(n_1 + 1)} F$, and $H = \frac{n_1 F}{n_2 + n_1 F}$, [13]. Since $| \alpha_{3:H} | = 2 | \alpha_{3:z} |$ approximately we see that the distribution of H is more skewed than that of z. We mention briefly also $S_1^2 - S_2^2$ where $S_1^2 = \frac{n_1}{N_1} s_1^2$, $S_2^2 = \frac{n_2}{N_2} s_2^2$. Clearly z, F, θ, and H give equivalent levels of significance. This is not true for z and $S_1^2 - S_2^2$.

Finally, since $F = \frac{s_1^2}{s_2^2}$, it may be interpreted as a quotient [5]. When the moments of F do not exist, it is due to the distribution function of s_2^2.

22. Conclusion. We have found the seminvariants for the *z* distribution, and approximations for them. Type III, and the logarithmic normal frequency functions are shown to be excellent approximations to the *z* distribution. The approach to normality for the *z* distribution is proved. A formula is given for finding the 5% level of significance for *z*. The *F* distribution is studied along the same lines. As far as the construction of tables for levels of significance is concerned, the *z* distribution is much easier to use. My sincerest thanks are due Professor C. C. Craig for his helpful guidance and many suggestions.

BIBLIOGRAPHY

[1] H. C. CARVER, *Handbook of Mathematical Statistics*, H. L. Rietz, ed., Boston: Houghton-Mifflin Co., 1924. Chapter on frequency curves.

[2] W. G. COCHRAN, "Note on an approximate formula for the significance levels of *z*," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 93–95.

[3] C. C. CRAIG, "A new exposition and chart for the Pearson system of frequency curves," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 16–28.

[4] C. C. CRAIG, "An application of Thiele's semi-invariants to the sampling problem," *Metron*, Vol. 7 (1928–29), pp. 3–74.

[5] C. C. CRAIG, "The frequency function of *y*/*x*," *Annals of Math.*, Second Series, Vol. 30 (1929), pp. 471–486.

[6] A. T. DOODSON, "Relation of a mode, median, and mean in a frequency curve," *Biometrika*, Vol. 11 (1917), p. 425.

[7] R. A. FISHER, "On a distribution yielding the error functions of several well known statistics," *Proc. International Math. Cong.*, 1924, Toronto, Vol. 2, pp. 805–813.

[8] R. A. FISHER, and E. A. CORNISH, "Moments and cumulants in the specification of distributions," *Revue de l'Institut International de Statistics*, 5th year, pp. 307–20, 1937, La Hague.

[9] R. A. FISHER, and YATES, *Statistical Tables for Biological, Agricultural, and Medical Research*, London: Oliver and Boyd, 1938.

[10] K. KNOPP, *Theory and Application of Infinite Series*, English translation, Edinburgh: Blackie and Son, 1928.

[11] N. NIELSEN, *Handbuch der Theorie der Gamma Functionen*, Leipzig: Teubner, 1906.

[12] C. A. OLSHEN, "Transformation of the Pearson Type III distribution," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 176–200.

[13] K. PEARSON (Editor), *Tables of the Incomplete Beta Function*, London: Biometrika Office, University College, London, 1934.

[14] K. PEARSON (Editor), *Tables of the Incomplete Gamma Function*, London: His Majesty's Stationery Office, 1922.

[15] K. Pearson, S. A. Stouffer, and F. N. David, "Further applications in statistics of the $T_m(x)$ Bessel function," *Biometrika*, Vol. 24 (1932), pp. 293–350.

[16] S. J. Pretorius, "Skew bivariate frequency curves examined in the light of numerical illustrations," *Biometrika*, Vol. 22, (1930–31).

[17] L. R. Salvosa, "Tables of Pearson's Type III function," *Annals of Math. Stat.*, Vol. 1 (1930), pp. 191–8 et seq.

[18] J. Shohat, and M. Frechet, "A proof of the generalized second limit-theorem in the theory of probability," *Trans. Am. Math. Soc.*, Vol. 33 (1931), pp. 531–43.

[19] G. W. Snedecor, *Calculation and Interpretation of the Analysis of Variance and Covariance*, Ames, Iowa: Collegiate Press.

[20] T. J. Stieltjes, "Tables des valeurs des sommes $s_k = \sum_{n=1}^{\infty} n^{-k}$," *Acta Math.*, Vol. 10, pp. 299–302.

[21] Whittaker and Robinson, *The Calculus of Observations*, Edinburgh: Blackie and Son, second edition, p. 135.

[22] Whittaker and Watson, *Modern Analysis*, 4th edition, London: Cambridge University Press, 1935.

[23] Pae-Tsi Yuan, "On the logarithmic frequency distribution and the semi-logarithmic correlation surface," *Annals of Math. Stat.*, Vol. 4, (1933).

[24] R. T. Zoch, "Some interesting features of frequency curves," *Annals of Math. Stat.*, Vol. 4, (1935), pp. 1–10.