A CONTRIBUTION TO THE PROBLEM OF TWO SAMPLES.

BY P. V. SUKHATME, B.Sc.,
Department of Applied Statistics, University College, London.

Received September 9, 1935.
(Communicated by Sir C. V. Raman, Kt., F.R.S., N.L.)

1. *Introductory.*

THE hypothesis, say H, that two samples of observations of a single variable have been drawn from the same population may be tested in a variety of ways. If there are strong grounds for believing that the populations sampled are approximately normal, then the only possible difference of importance will be between their means a_1 and a_2 or standard deviations σ_1 and σ_2; it is customary, therefore, to test separately for the significance of differences between (1) the sample variances σ_1 and σ_2 and (2) the sample means a_1 and a_2. If, however, both tests are on the border line of significance it may be difficult to decide what conclusions to draw; for this and other reasons there appear to be certain advantages in the availability of a single comprehensive test of the hypothesis of a common origin for the two samples. Such tests are frequently applied in statistical analysis; if the two samples contain many observations the hypothesis may be tested by applying the χ^2 test to the two series of grouped frequencies and so obtaining a single criterion to judge the probability of H. Again when a number of samples are available and it is wished to test whether the means in the several sampled populations are identical, a single test (whether in z or η^2 form) is applied rather than a separate test for each pair of samples.

J. Neyman and E. S. Pearson have discussed the use of such comprehensive tests in a number of problems. They have pointed out that in comparing two alternative tests they should in the first place be made equivalent, that is to say, adjusted so that in both cases the risk of rejection of H when it is true is the same; consideration should then be given to the relative efficiency of the two tests in rejecting the hypothesis when some alternative is true. They have shown that in certain cases there is a single test associated with a best critical region which is more efficient than any other from this last point of view.[1] In other cases no test with a best critical region common for all alternative hypotheses exists, but they have

[1] *Phil. Trans. Roy. Society, Series A*, 1933, **231.**

suggested the use of the principle of likelihood in picking out a test that may be described as associated with a good critical region.[2]

In the present problem they have suggested the use of a test[3] based on a criterion L defined below. The main object of this paper is to discuss the derivation and use of Tables of 5% and 1% probability levels for this criterion, but as a preliminary point it is of interest partly on theoretical grounds to compare this form of test with two alternative comprehensive tests of the same hypothesis.

2. Comparison of L test with two alternative two-sample tests.

If two samples of size n_1 and n_2 have been drawn randomly from two normal populations, then we may use as criteria to test separately the significance of the difference (1) between variances and (2) between means

$$1. \quad u = \frac{n_1 s_1^2}{n_1 s_1^2 + n_2 s_2^2} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (1)$$

$$2. \quad t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \quad \ldots \quad \ldots \quad (2)$$

where $\bar{x}_1$ and $\bar{x}_2$ are the means of two samples and

$$n_1 s_1^2 = \sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2 \,; \quad n_2 s_2^2 = \sum_{i=1}^{n_2} (x_i - \bar{x}_2)^2$$

If the population variances are the same, then the sampling distribution of u is

$$f_1(u) = \text{constant } u^{\frac{1}{2}(n_1-1)-1} (1-u)^{\frac{1}{2}(n_2-1)-1} \quad \ldots \quad \ldots \quad (3)$$

If both the population means and variances are the same and we write $n_1 + n_2 = \mathrm{N}$

$$f_2(t) = \text{constant} \left(1 + \frac{t^2}{\mathrm{N}-2}\right)^{-\frac{\mathrm{N}-1}{2}} \quad \ldots \quad \ldots \quad \ldots \quad (4)$$

Neyman and Pearson's criterion is

$$\mathrm{L} = \left(\frac{s_1}{s_0}\right)^{\frac{2n_1}{\mathrm{N}}} \left(\frac{s_2}{s_0}\right)^{\frac{2n_2}{\mathrm{N}}} = \frac{\mathrm{N}}{(n_1)^{\frac{n_1}{\mathrm{N}}} (n_2)^{\frac{n_2}{\mathrm{N}}}} \frac{u^{\frac{n_1}{\mathrm{N}}} (1-u)^{\frac{n_2}{\mathrm{N}}}}{\left(1 + \frac{t^2}{\mathrm{N}-2}\right)} \quad \ldots \quad \ldots \quad (5)$$

where $s_0^2 = \dfrac{\sum_{i=1}^{\mathrm{N}} (x_i - \bar{x}_0)^2}{\mathrm{N}}$, $\bar{x}_0$ being the mean of the combined sample of $\mathrm{N} = n_1 + n_2$ observations.

---

[2] *Biometrika*, **20A**, pp. 175–240 and 264-294.

[3] On the Problem of Two Samples (1930); On the Problem of k Samples (1931); *Bulletin de L'academie Polonaize des Sciences, Series A, Sciences Mathematiques.*

Fig. 1 is a diagram having co-ordinate axes t and u. The oval contour represents a member of (5) for which $n_1 = n_2 = 15$ and $L = \cdot 8028$, which has been so chosen that the chance is $\cdot 05$ of a point (t, u) falling outside the curve, if the hypothesis H is true. That is to say, if in drawing a sample of 15 from each of two normal populations, it is decided to reject H whenever the point (t, u) falls outside this contour, we shall reject H when in fact the populations have the same mean and standard deviations, 5 times in 100.

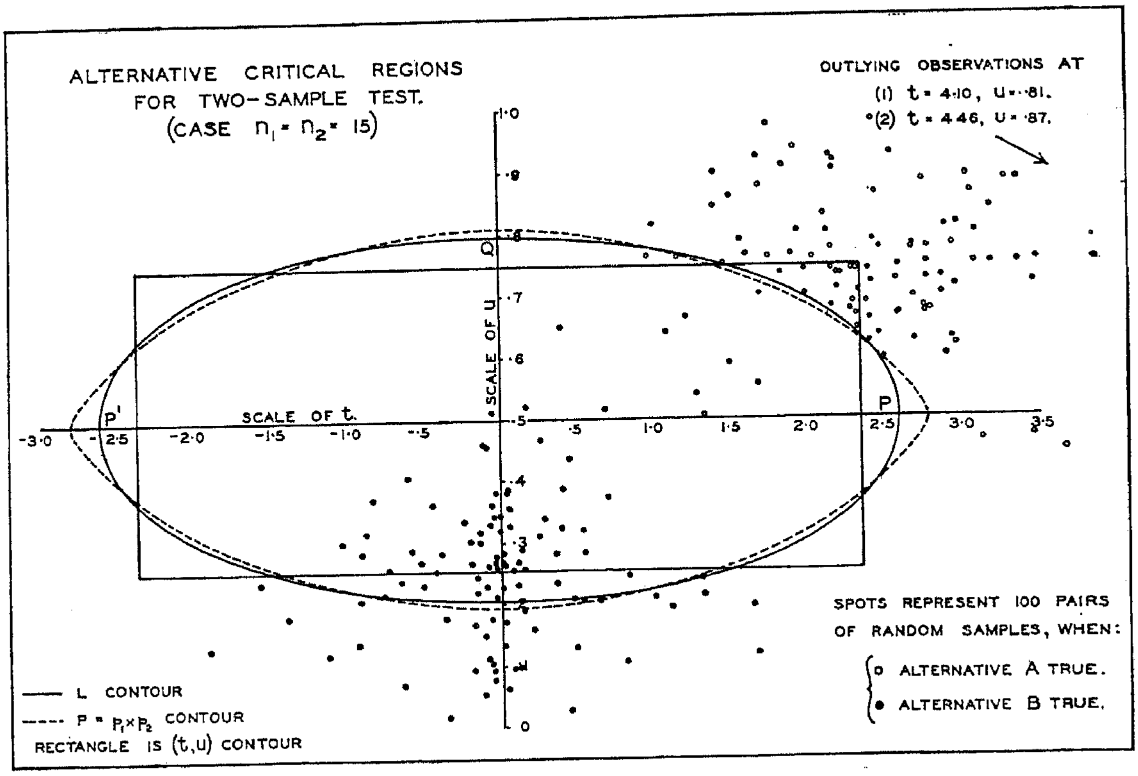

FIG. 1.

We may now use two further alternative contours:

(1) Decide to reject H unless both $u_1 \leqslant u \leqslant u_2$ and $t_1 \leqslant t \leqslant t_2$, *i.e.*, reject H when (t, u) lies outside a rectangle such as that shown in the figure. To make the test equivalent to the previous one, the limits must be chosen so that only 5% of possible sample points fall outside the rectangle if H be true. There will be an infinite number of ways of choosing these limits; in the present instance the limits for u and t are chosen in such a way that the chance of rejection of the hypothesis H_1 when it is true, is equal to the

chance of rejection of H_2 when it is true.[4] If ϵ denotes this chance of rejection, then it follows that the chance of falling inside the rectangle is $\cdot 95 = (1 - \epsilon)^2$, so that $\epsilon = \cdot 02532$.

(2) A third alternative test may be obtained on the lines suggested by R. A. Fisher[5] and K. Pearson.[6] If we write for the case $n_1 = n_2$

$$\left.\begin{aligned} p_1 &= \int_0^u f_1(u)du \qquad \text{if } u \leqslant \tfrac{1}{2} \\ p_1 &= \int_u^1 f_1(u)du \qquad \text{if } u > \tfrac{1}{2} \end{aligned}\right\} \qquad \ldots \qquad \ldots \qquad \ldots \quad (6)$$

and

$$\left.\begin{aligned} p_2 &= \int_{-\infty}^t f_2(t)dt \qquad \text{if } t \leqslant 0 \\ p_2 &= \int_t^{\infty} f_2(t)dt \qquad \text{if } t > 0 \end{aligned}\right\} \qquad \ldots \qquad \ldots \qquad \ldots \quad (7)$$

the criterion suggested is

$$P = p_1 \times p_2 \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \quad (8)$$

The region of rejection is that for which P is less than some specified value. The contour in the (t, u) field equivalent to that determined in the other two tests can be readily found from (8) on substituting for P its 5% point given by

$$P_{.05} = e^{-\frac{1}{2}\chi^2_{\cdot 05}} = \cdot 0087.$$

It is drawn as a dotted curve in the figure and differs in the present instance, when $n_1 = n_2$, only slightly from the L contour. It is likely that when $n_1 \neq n_2$ the two contours may differ much more but the case is not considered in this paper.

Having now obtained these equivalent contours associated with the use of these three criteria, it is of interest to consider their relative efficiency in rejecting the hypothesis H when the populations sampled are in fact different.

Since it can be shown[7] that there can be in this case no best critical region common for all alternatives to H, it may be expected that for some

[4] The hypotheses H_1 and H_2 are identical with those considered by Neyman and Pearson in their paper " On the Problem of Two Samples ". H_1 is the hypothesis that the samples have come from some two normal populations with a common variance and H_2 is the hypothesis that the samples have come from populations with a common mean, it being assumed that the populations have the same variance.

[5] *Statistical Methods for Research Workers*, Fourth Edition, p. 97.

[6] *Biometrika*, 1933, **25**, p. 379.

[7] The author has succeeded in proving this rigorously but as the proof extends over several pages it is not given here.

alternatives one criterion will be more efficient, for other alternatives a second and so on. That this is the case is shown in the diagram, where results are given for two series of samplings from populations not satisfying the conditions of the hypothesis.

		Population 1	Population 2
Alternative A.			
Mean	..	0·0	7·25
S.D.	..	10·0	5·75
Alternative B.			
Mean	..	0·0	0·0
S.D	..	10·0	5·75

100 pairs of samples of 15 were drawn in each case and the points (t, u) were plotted. It will be seen that the oval contours contain less circles and more dots than the rectangle. This means that while the L test and also the $(P = p_1 \times p_2)$ test are more likely than the (t, u) test to detect the fact that the sampled populations are different when these populations are as in A, the reverse is the case when the two populations are as in B. It is clear that there is here a matter for further theoretical investigation and the problem for the moment may be left with two questions. If there is no test with a common best critical region for all alternative hypotheses, on what principle can we choose from among other possible alternative tests? Will this principle be found to lead to the likelihood or L criterion, which in many problems already considered appears at any rate to satisfy our intuitional requirements?

3. *The Distribution of* L.

Suppose that the samples Σ_1, Σ_2 have been drawn at random from some normal populations with means a_1 and a_2 and standard deviations σ_1 and σ_2 respectively. If the hypothesis that Σ_1 and Σ_2 belong to the same normal populations be true then they must have been drawn from a common population with mean $= a$ and standard deviation $= \sigma$; and we know that the frequency function for the simultaneous variations in $\bar{x}_1$, $\bar{x}_2$, s_1 and s_2 is given by

$$\{\bar{x}_1,\ \bar{x}_2,\ s_1,\ s_2\} = \text{const.}\ (s_1)^{n_1-2}\ (s_2)^{n_2-2}\ e^{-\frac{n_1(\bar{x}_1-a)^2 + n_2(\bar{x}_2-a)^2 + n_1 s_1^2 + n_2 s_2^2}{2\sigma^2}} \quad .. \quad (9)$$

My aim will be to transform the variables and integrate for certain of them until we are left with the frequency function of two variables in terms of which the value of L can be expressed. Substituting first

$$\bar{x}_2' = \bar{x}_2 - a \text{ and } \zeta = \bar{x}_1 - \bar{x}_2 \quad .. \quad .. \quad .. \quad .. \quad (10)$$

we obtain the probability law of $\bar{x}_2'$, ζ, s_1 and s_2

$$f\{\bar{x}_2', \zeta, s_1, s_2\} = \text{const.}\ (s_1)^{n_1-2}\ (s_2)^{n_2-2}\ e^{-\frac{n_1 s_1^2 + n_2 s_2^2}{2\sigma^2}} \times$$

$$e^{-\frac{1}{2\sigma^2}\left\{(n_1+n_2)\left[\bar{x}_2' + \frac{n_1\zeta}{n_1+n_2}\right]^2 + \frac{n_1 n_2}{n_1+n_2}\zeta^2\right\}} \quad \ldots (11)$$

Integrating with respect to $\bar{x}_2'$ between the limits $-\infty$ and $+\infty$ and substituting

$$s_0^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2} + \frac{n_1 n_2}{(n_1+n_2)^2}\zeta^2 \quad \ldots \quad \ldots \quad \ldots (12)$$

we obtain

$$f\{s_0, s_1, s_2\} = \text{const.}\ \frac{(s_1)^{n_1-2}(s_2)^{n_2-2}\ e^{-\frac{n_1+n_2}{2\sigma^2}s_0^2}}{\left[1 - \frac{n_1\left(\frac{s_1}{s_0}\right)^2 + n_2\left(\frac{s_2}{s_0}\right)^2}{n_1+n_2}\right]^{\frac{1}{2}}} \quad \ldots \quad \ldots (13)$$

Lastly putting $s_1 = s_0\, x$ and $s_2 = s_0\, y$ .. .. (14)

and integrating for s_0 between the limits 0 and ∞, we get

$$f\{x, y\} = \text{const.}\ \frac{(x)^{n_1-2}\ (y)^{n_2-2}}{\sqrt{1 - \frac{n_1 x^2 + n_2 y^2}{n_1+n_2}}} \quad \ldots \quad \ldots \quad \ldots (15)$$

It follows from (5) and (14) that

$$L = (x)^{\frac{2n_1}{N}}\ (y)^{\frac{2n_2}{N}} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots (16)$$

Whence we obtain

$$f\{x, L\} = \text{const.}\ \frac{x^{2\left(\frac{n_1}{n_2}-1\right)}\ L^{\frac{N}{2}-\frac{N}{2n_2}-1}}{\sqrt{Nx^{\frac{2n_1}{n_2}} - n_1 x^{\frac{2n_1}{n_2}+2} - n_2 L^{\frac{N}{n_2}}}} \quad \ldots \quad \ldots (17)$$

Therefore

$$f(L) = \text{const.}\ L^{\frac{N}{2}-\frac{N}{2n_2}-1}\int_a^b \frac{x^{2\left(\frac{n_1}{n_2}-1\right)}}{\sqrt{Nx^{\frac{2n_1}{n_2}} - n_1 x^{\frac{2n_1}{n_2}+2} - n_2 L^{\frac{N}{n_2}}}}\,dx \quad \ldots (18)$$

where a, b denote the real roots of

$$Nx^{\frac{2n_1}{n_2}} - n_1 x^{\frac{2n_1}{n_2}+2} - n_2 L^{\frac{N}{n_2}} = 0 \quad \ldots \quad \ldots \quad \ldots (19)$$

When $n_1 = n_2 = n$, equation (18) may be written as

$$f(L) = \text{const.}\ L^{n-2}\int_a^b \frac{dx}{\sqrt{(x^2-a^2)\,(b^2-x^2)}} \quad \ldots \quad \ldots (20)$$

where a and b are now given by

$$\left.\begin{aligned} a &= \frac{1}{\sqrt{2}}\left\{\sqrt{1+L} - \sqrt{1-L}\right\} \\ \text{and } b &= \frac{1}{\sqrt{2}}\left\{\sqrt{1+L} + \sqrt{1-L}\right\} \end{aligned}\right\} \quad \ldots \quad \ldots \quad \ldots \quad (21)$$

On substituting $x^2 = a^2 \cos^2\phi + b^2 \sin^2\phi$ .. .. .. (22)

in (20) and using (21) we obtain

$$f(L) = \text{const.}\, \frac{L^{n-2}}{\sqrt{1+L}\left(1+\sqrt{\frac{1-L}{1+L}}\right)} \int_0 \frac{d\phi}{\sqrt{1 - \dfrac{4\sqrt{\frac{1-L}{1+L}}}{\left(1+\sqrt{\frac{1-L}{1+L}}\right)^2}\sin^2\phi}} \quad \ldots (23)$$

It will be noticed that the integral in equation (23) is an elliptic integral of the first order with modulus equal to $\sqrt{\dfrac{4\sqrt{\frac{1-L}{1+L}}}{\left(1+\sqrt{\frac{1-L}{1+L}}\right)^2}}$. The modulus appears to be very complex. I shall therefore try to connect this elliptic integral with another having a simple modulus.

Denote $\sqrt{\frac{1-L}{1+L}}$ by small k. Then the elliptic integral can be written as

$$\int_0^{\frac{\pi}{2}} \frac{d\phi}{\sqrt{1 - \frac{4k}{(1+k)^2}\sin^2\phi}} \quad \ldots \quad \ldots \quad \ldots \quad \ldots (24)$$

Substituting in (24)

$$2\phi = \theta + \sin^{-1}(k \sin\theta) \quad \ldots \quad \ldots \quad \ldots \quad \ldots (25)$$

we have

$$\int_0^{\frac{\pi}{2}} \frac{d\phi}{\sqrt{1 - \frac{4k}{(1+k)^2}\sin^2\phi}} = (1+k)\int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} \quad \ldots (26)$$

the latter being the elliptic integral with modulus k. Hence on evaluating the constant, (23) may be written as

$$f(L) = \frac{4\,\Gamma\left(n-\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\pi\,\Gamma(n-1)\,\Gamma\left(\frac{n-1}{2}\right)} \frac{L^{n-2}}{\sqrt{1+L}} \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} \quad \ldots (27)$$

where

$$k^2 = \frac{1-L}{1+L}.$$

This expression for the distribution of L is more simple and elegant than the one given in equation (23). It is of considerable interest to see how this form can readily be obtained by using relation (5) which may be written as

$$\mathrm{L} = \mathrm{L}_1 \times \mathrm{L}_2 \qquad \ldots \qquad (28)$$

where

$$\mathrm{L}_1 = \frac{\mathrm{N}}{(n_2)^{\frac{n_2}{\mathrm{N}}}\,(n_1)^{\frac{n_1}{\mathrm{N}}}}\, u^{\frac{n_1}{\mathrm{N}}}\,(1-u)^{\frac{n_2}{\mathrm{N}}} \qquad \ldots \qquad (29)$$

and

$$\mathrm{L}_2 = \left(1 + \frac{t^2}{\mathrm{N}-2}\right)^{-1} \qquad \ldots \qquad (30)$$

For this purpose it is essential to obtain the distribution of L_1, which can be easily done as follows :

The probability law of s_1^2 and s_2^2 is given by

$$f\{s_1^2,\, s_2^2\} = \text{const. } (s_1^2)^{\frac{n_1-3}{2}}\,(s_2^2)^{\frac{n_2-3}{2}}\, e^{-\frac{1}{2\sigma^2}(n_1 s_1^2 + n_2 s_2^2)} \qquad \ldots \qquad (31)$$

Substituting

$$\sqrt{\frac{n_1}{2}}\,\frac{s_1}{\sigma} = r\cos\beta\,;\ \sqrt{\frac{n_2}{2}}\,\frac{s_2}{\sigma} = r\sin\beta \qquad \ldots \qquad (32)$$

and integrating w.r.t. r between the limits 0 and ∞ we obtain

$$f(\beta) = \text{const. } (\cos\beta)^{n_1-2}\,(\sin\beta)^{n_2-2} \qquad \ldots \qquad (33)$$

It follows from (29) and (32) that

$$\mathrm{L}_1 = \mathrm{N}\left\{\frac{(\cos^2\beta)^{n_1}}{n_1^{\,n_1}}\ \frac{(\sin^2\beta)^{n_2}}{n_2^{\,n_2}}\right\}^{\frac{1}{\mathrm{N}}} \qquad \ldots \qquad (34)$$

On substituting (34) in (33) we get

$$f(\mathrm{L}_1) = \text{const. } \mathrm{L}_1^{\frac{\mathrm{N}}{2}-1}\,\{\sin\beta\cos\beta\,(n_2\cos^2\beta - n_1\sin^2\beta)\}^{-1} \ \ldots \ (35)$$

where β may be written in terms of L_1 from (34). When $n_1 = n_2$, (35) may be written on evaluating the constant, as

$$f(\mathrm{L}_1) = \frac{2\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)}\,(\mathrm{L}_1^2)^{\frac{n}{2}-1}\,(1-\mathrm{L}_1^2)^{-\frac{1}{2}} \qquad \ldots \qquad (36)$$

It will be noticed that L_1^2 is distributed in Pearson type I curve. This distribution has been recently deduced independently by Mr. Nayer from the distribution of u by using relation (29).[8]

The distribution of L_2 which can be easily deduced from the t distribution by using (30) is given by

$$f(\mathrm{L}_2) = \frac{\Gamma(n-\frac{1}{2})}{\Gamma(n-1)\,\Gamma(\frac{1}{2})}\,(\mathrm{L}_2)^{n-2}\,(1-\mathrm{L}_2)^{-\frac{1}{2}} \qquad \ldots \qquad (37)$$

---

[8] 'Practical Application of Neyman and Pearson's L_1 test.' Thesis for the degree of M.Sc. in the University of London, 1934.

It follows from (28) that the distribution of L will be that of $L_1 L_2$. From (36) and (37) we get

$$f(L_1 L_2) = \frac{2\,\Gamma(\frac{n}{2})\,\Gamma(n-\frac{1}{2})}{\pi\,\Gamma(\frac{n-1}{2})\,\Gamma(n-1)} L_1^{n-2}(1-L_1^2)^{-\frac{1}{2}} L_2^{n-2}(1-L_2)^{-\frac{1}{2}} \quad (38)$$

Substituting $L_1 = \frac{L}{L_2}$ in (38) and integrating w.r.t. L_2 between the limits L and 1, we have

$$f(L) = \frac{4}{\pi}\,\frac{\Gamma(n-\frac{1}{2})\,\Gamma(\frac{n}{2})}{\Gamma(n-1)\,\Gamma(\frac{n-1}{2})}\,\frac{L^{n-2}}{\sqrt{1+L}}\,K \quad \ldots \quad (39)$$

where K represents the elliptic integral

$$\int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} \quad \ldots \quad (40)$$

It will be noticed that (39) and (27) are identical.

4. *Methods of Evaluating the Probability Integral of* L.

(*a*) *The general case.*

To calculate the probability that L exceeds a given value L_0 it is necessary to integrate (15) so that

$$(x)^{\frac{2n_1}{N}}(y)^{\frac{2n_2}{N}} \geqslant L_0 \text{ and } n_1x^2 + n_2y^2 \leqslant N \quad \ldots \quad (41)$$

Solving for y we get $y = \sqrt{\dfrac{N-n_1x^2}{n_2}}$ the upper limit of y

and $$y = \left(\frac{L}{x^{\frac{2n_1}{N}}}\right)^{\frac{N}{2n_2}}$$ the lower limit of y.

If the corresponding roots of x be denoted by b and a then

$$p(L > L_0) = \sqrt{N}\;C\int_a^b dx \int_{\left(\frac{L}{x^{\frac{2n_1}{N}}}\right)^{\frac{N}{2n_2}}}^{\sqrt{\frac{N-n_1x^2}{n_2}}} \frac{x^{n_1-2}\,y^{n_2-2}}{\sqrt{N-n_1x^2-n_2y^2}}\,dy \quad \ldots \quad (42)$$

where

$$C = \frac{4(n_1)^{\frac{n_1-1}{2}} (n_2)^{\frac{n_2-1}{2}} \Gamma\left(\frac{n_1+n_2-1}{2}\right)}{\Gamma(\frac{1}{2})\,(n_1+n_2)^{\frac{n_1+n_2-2}{2}}\,\Gamma\left(\frac{n_1-1}{2}\right)\Gamma\left(\frac{n_2-1}{2}\right)} \quad .. \quad (43)$$

Substituting

$$y = \sqrt{\frac{\xi(N - n_1x^2)}{n_2}} \quad .. \quad .. \quad .. \quad .. \quad (44)$$

we obtain

$$p(L > L_0) = \frac{\sqrt{N}\, C}{2(n_2)^{\frac{n_2-1}{2}}} \int_a^b x^{n_1-2} (N - n_1x^2)^{\frac{n_2}{2}-1} dx \int_{\xi_1}^1 \xi^{\frac{n_2}{2}-1} (1-\xi)^{-\frac{1}{2}} d\xi .. \quad (45)$$

where

$$\xi_1 = \left(\frac{L^{\frac{N}{2n_2}}}{x^{\frac{n_1}{n_2}}}\right)^2 \times \frac{n_2}{N - n_1x^2} \quad .. \quad .. \quad .. \quad .. \quad (46)$$

Substituting further

$$\frac{n_1}{N} x^2 = \eta$$

$$p(L > L_0) = \frac{4}{\sqrt{\pi}} \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{n_1-1}{2}\right)\Gamma\left(\frac{n_2-1}{2}\right)} \int_{\frac{n_1}{N}a^2}^{\frac{n_1}{N}b^2} \eta^{\frac{n_1-3}{2}} (1-\eta)^{\frac{n_2-1}{2}} d\eta \times$$

$$\int_{\xi_1}^1 \xi^{\frac{n_2}{2}-1} (1-\xi)^{-\frac{1}{2}} d\xi \quad .. \quad .. \quad (47)$$

which can be evaluated on using the Tables of the Incomplete Beta Function.[9]

(*b*) *Case* $n_1 = n_2$.

The simple method of quadrature may be used with advantage when $n_1 = n_2 = n$. The main difficulty is the evaluation of the ordinates of L-frequency curve. Equation (39) shows that this depends on the evaluation of the elliptic integral K.

[9] Edited by Karl Pearson.

There are several Tables in use of the elliptic integrals. The Tables[10] by L. M. Milne Thomson, particularly, suit here well. In these Tables the values of K are given for all values of k^2 differing by $\cdot 01$. Interpolation into the Tables, however, becomes more and more difficult as k^2 increases. In such cases, the following method of evaluating K was adopted :—

Let $$\epsilon = \frac{1}{2}\,\frac{1-(1-k^2)^{\frac{1}{4}}}{1+(1-k^2)^{\frac{1}{4}}} \quad \ldots \quad (48)$$

then the nome q is given by

$$q = \epsilon + 2\epsilon^5 + 15\epsilon^9 + 150\epsilon^{13} + 1707\epsilon^{17} + 20910\epsilon^{21} + \ldots \quad \ldots \quad (49)$$

The series (49) enables us to calculate the value of q, corresponding to the desired value of k^2. By means of these values of q the function

$$\vartheta = 1 + 2q + 2q^4 + \ldots \quad \ldots \quad (50)$$

was calculated. The value of K is then easily obtained by multiplying ϑ^2 by $\frac{\pi}{2}$.

(c) *An approximate method.*

This method, suggested by Neyman and Pearson in their paper, consists in assuming Pearson type I curve to represent the probability law of L and obtaining the probability integral from the Tables of the Incomplete Beta Function. That is to say

$$f(\mathrm{L}) = \frac{\Gamma(m_1+m_2)}{\Gamma(m_1)\Gamma(m_2)}\,\mathrm{L}^{m_1-1}\,(1-\mathrm{L})^{m_2-1} \quad \ldots \quad (51)$$

represents the probability law of L, where the values of m_1 and m_2 are given by

$$m_1 = \frac{\mu_1'(\mu_1'-\mu_2')}{\mu_2'-\mu_1'^2}; \quad m_2 = \frac{(1-\mu_1')(\mu_1'-\mu_2')}{\mu_2'-\mu_1'^2} \quad \ldots \quad (52)$$

μ_1' and μ_2' being the first two moment coefficients of L about zero. It was, however, found in the present case that the values of m_2 differed only very slightly from unity and consequently the lower 5% and 1% values of L were immediately available from the following relations:

$$\mathrm{L}_{5\%} = (\cdot 05)^{\frac{1}{m_1}} \text{ and } \mathrm{L}_{1\%} = (\cdot 01)^{\frac{1}{m_2}} \quad \ldots \quad (53)$$

The lower limits are used because the criterion L is of such form that the hypothesis tested becomes less and less likely as L decreases from 1 towards 0.

---

[10] *Proceedings of the London Mathematical Society*, **33.**

5. *The Calculation of the Tables and an Illustration of their Use.*

Before using these simple relations in forming the Tables it was essential to examine the approximations involved. This was accomplished by comparing the values of $L_{5\%}$ and $L_{1\%}$ at (1) $n_1 = n_2 = 5$, 12, 20, 60 and (2) $n_1 = 5$, $n_2 = 15$, found by using (53) with those obtained by using the exact methods (*a*) and (*b*). The agreement between the two sets of values of L was found exact to three decimal places. Consequently the relation (53) was employed in forming the Tables. Tables I and II give these values for a number of different values of n_1 and n_2. Interpolation into the Tables for any pair of values of n_1 and n_2 less than 120 will be found fairly easy.

When either n_1 or n_2 exceeds 120, the percentage values of L can be obtained from the Tables of χ^2. It has been shown by Neyman and Pearson that if n_1 and n_2 are sufficiently large and we write $L^{\frac{N}{2}} = e^{-\frac{1}{2}\chi^2}$, then the frequency distribution of χ is given by

$$f(\chi) = \chi\, e^{-\frac{1}{2}\chi^2} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (54)$$

It follows that for large values of n_1 and n_2, approximations to the probability integral of L can be obtained from the Tables of the χ^2 integral. Using Fisher's Tables of χ^2, we find that the 5% and 1% values of χ^2 for two degrees of freedom are 5·991 and 9·210 respectively. Hence we get

$$L_{\cdot 05} = e^{-\frac{5\cdot 991}{n_1+n_2}} \text{ and } L_{\cdot 01} = e^{-\frac{9\cdot 210}{n_1+n_2}} \quad \ldots \quad \ldots \quad \ldots \quad (55)$$

On comparing the values of L obtained by this method with the exact ones, it was found that when n_1 and n_2 both exceeded 70 the agreement was excellent and differed only in the fourth place. The agreement, however, begins to fall off if the two samples are extremely unequal in size but it is very rarely that we may have to test two samples when one of them is thrice as large or larger than the other. I conclude therefore that if the average size of the two samples is sufficiently large and both the samples are of about the same size then $L_{\cdot 05}$ and $L_{\cdot 01}$ can be obtained by direct use of relation (55).

One very interesting point comes out of the above discussion. Equation (55) shows that the value of L does not depend on the values of n_1 and n_2 individually but on $n_1 + n_2$, or put in another way, on the average size of sample. This means that whatever be n_1 and n_2, so long as they are large and not very unequal, the sampling distribution of L is that which would be obtained if n_1 and n_2 were both equal to their average value $\frac{1}{2}(n_1+n_2)$. The importance of this result is considerably enhanced by the fact that it is

TABLE I.

5 per cent. values of L_1.

n_2 \ n_1	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
5	·442	·487	·522	·553	·579	·602	·641	·687	·742	·773	·808	·848	·892	·942	1·000
6	·487	·527	·559	·587	·611	·632	·667	·709	·758	·787	·820	·856	·897	·945	1·000
7	·522	·559	·589	·615	·637	·656	·688	·726	·772	·798	·828	·862	·901	·947	1·000
8	·553	·587	·615	·638	·658	·675	·705	·740	·782	·807	·835	·868	·905	·948	1·000
9	·579	·611	·637	·658	·676	·692	·720	·752	·792	·815	·841	·872	·907	·949	1·000
10	·602	·632	·656	·675	·692	·707	·733	·763	·800	·822	·847	·876	·910	·950	1·000
12	·641	·667	·688	·705	·720	·733	·755	·781	·813	·833	·856	·882	·913	·952	1·000
15	·687	·709	·726	·740	·752	·763	·781	·803	·830	·847	·866	·890	·918	·954	1·000
20	·742	·758	·772	·782	·792	·800	·813	·830	·852	·865	·881	·900	·924	·956	1·000
24	·773	·787	·798	·807	·815	·822	·833	·847	·865	·876	·890	·906	·928	·958	1·000
30	·808	·820	·828	·835	·841	·847	·856	·866	·881	·890	·901	·914	·933	·960	1·000
40	·848	·856	·862	·868	·872	·876	·882	·890	·900	·906	·914	·925	·940	·962	1·000
60	·892	·897	·901	·905	·907	·910	·913	·918	·924	·928	·933	·940	·950	·967	1·000
120	·942	·945	·947	·948	·949	·950	·952	·954	·956	·958	·960	·962	·967	·975	1·000
∞	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000

TABLE II. *1 per cent. values of* I_1.

n_2 \ n_1	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
5	·286	·331	·369	·402	·432	·459	·505	·562	·632	·673	·721	·776	·839	·913	1·000
6	·331	·374	·410	·441	·469	·494	·537	·589	·654	·692	·737	·788	·847	·917	1·000
7	·369	·410	·444	·473	·499	·522	·563	·611	·671	·707	·749	·797	·853	·919	1·000
8	·402	·441	·473	·501	·525	·547	·584	·629	·686	·720	·759	·804	·857	·921	1·000
9	·432	·469	·499	·525	·548	·568	·603	·645	·698	·730	·767	·810	·861	·923	1·000
10	·459	·494	·522	·547	·568	·587	·620	·660	·709	·739	·774	·815	·864	·925	1·000
12	·505	·537	·563	·584	·603	·620	·649	·684	·728	·755	·787	·824	·870	·927	1·000
15	·562	·589	·611	·629	·645	·660	·684	·714	·751	·775	·802	·836	·877	·930	1·000
20	·632	·654	·671	·686	·698	·709	·728	·751	·781	·800	·822	·850	·886	·933	1·000
24	·673	·692	·707	·720	·730	·739	·755	·775	·800	·816	·835	·860	·892	·935	1·000
30	·721	·737	·749	·759	·767	·774	·787	·802	·822	·835	·852	·872	·899	·939	1·000
40	·776	·788	·797	·804	·810	·815	·824	·836	·850	·860	·872	·888	·909	·943	1·000
60	·839	·847	·853	·857	·861	·864	·870	·877	·886	·892	·899	·909	·924	·949	1·000
120	·913	·917	·919	·921	·923	·925	·927	·930	·933	·935	·939	·943	·949	·962	1·000
∞	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000	1·000

easily generalised for the problem of k samples and is applicable alike to all the three criteria L, L_1 and L_2.

EXAMPLE: Two samples of skulls containing 15 and 13 individuals respectively give the following value for the mean and the standard deviation of the Cephalic Index. It is wished to test whether these two samples come from populations with a common Cephalic Index.

$$\text{Sample (1) } n_1 = 15\,;\ \bar{x}_1 = 76\cdot4067\,;\quad s_1^2 = 6\cdot6806$$
$$\text{Sample (2) } n_2 = 13\,;\ \bar{x}_2 = 73\cdot7077\,;\quad s_2^2 = 6\cdot9238$$

Substituting these values in

$$s_0^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2} + \frac{n_1 n_2}{(n_1 + n_2)^2} (\bar{x}_1 - \bar{x}_2)^2$$

we obtain

$$s_0^2 = 8\cdot6053\,;\ \frac{s_1^2}{s_0^2} = \cdot7763\,;\ \text{and}\ \frac{s_2^2}{s_0^2} = \cdot8046.$$

Since

$$\begin{aligned}\text{Log L} &= \frac{n_1}{\text{N}} \log\left(\frac{s_1^2}{s_0^2}\right) + \frac{n_2}{\text{N}} \log\left(\frac{s_2^2}{s_0^2}\right) \\ &= \tfrac{15}{28}\,(\bar{1}\cdot89003) + \tfrac{13}{28}\,(\bar{1}\cdot90558) = \bar{1}\cdot89725\end{aligned}$$

we have $\quad$ L $= \cdot789$.

Now to obtain the $L_{\cdot05}$ we shall have to interpolate into the Table II for $n_1 = 15$ and $n_2 = 13$. We have from the column headed $n_1 = 15$

n_2	$L_{\cdot05}$	$\delta^2 L_{\cdot05}$
20	·830	
15	·803	5
12	·781	4
10	·763	

The values 20, 15, 12 and 10 are in harmonic progression. Taking $u = 60/n_2$ as the new argument the value of u corresponding to $n_2 = 20$, 15, 12 and 10 becomes 3, 4, 5 and 6 respectively and the problem reduces to interpolating for $u = \frac{60}{13} = 4\cdot6154$ from the Table above which gives the values of $L_{\cdot05}$ corresponding to equidistant values of u. Actually on using Everett's simplified formula for interpolation it is found that $L_{\cdot05} = \cdot788$. The observed value is therefore very near to the 5% level. Since this is somewhat exceptional we shall do well to enquire whether the significance is due to difference in means or standard deviations of the populations. This leads us to apply the t and z tests. On using these it is found $t = 2\cdot633$ and $z = \cdot0234$, while the corresponding 5% values are $2\cdot056$ and $\cdot4649$ respectively. The value of t is thus definitely significant while that of z is not, suggesting thereby that it was very unlikely that the two series of skulls came from populations with a common mean Cephalic Index.

6. *Sensitiveness of L Test when applied to Non-Normal Data.*

It will be recalled that the application of L test presupposes that the samples are drawn from some normal population. Actually, however, we may sometimes deal with non-normal data and consequently it is necessary to know whether our test becomes invalid in this case.

There are two ways in which the L test may become invalid when the variation is non-normal.

(*a*) It will cease to be the most appropriate test.

(*b*) Its sampling distribution will no longer have the form given in equation (18).

It is difficult to measure the precise effect of (*a*) and (*b*); for, the task of devising tests appropriate to every possible form of non-normal variation and constructing the tables is an impossible one. It is possible, however, to obtain experimentally some idea regarding the validity of L test when applied to non-normal data. This could be accomplished by testing the goodness of fit of the L distribution for samples drawn from a series of non-normal populations. Table III gives the types of populations which have been considered. The types have been represented by curves of the Pearson system.

TABLE III.

Population curves	β_1	β_2	Standard deviation in terms of grouping unit
Type II	0	2·5	6·32
Type VII	0	4·1	5·67
Type III	0·2	3·3	5·00
Type I	1·0	3·8	12·37

The results of sampling carried out from these populations were kindly given to me by Dr. E. S. Pearson. It will be seen from the Table that one of the four populations is very skew, one is very leptokurtic while the remaining two are moderately non-normal.

Tables IV and V compare the observed distributions of L obtained from these experimental populations with the expected values calculated from the

TABLE IV.

$n_1 = 10\,;\; n_2 = 20.$

Limits of L	Normal Theory Frequencies	(grouped)	Frequencies in Samples from Experimental Populations (0·0, 2·5)	(grouped)	(0·0, 4·1)	(grouped)	(0·2, 3·3)	(grouped)	(1·0, 3·8)	(grouped)
·99—1·00	25·2		31		26		25		30	
·98— ·99	23·0		25		15		18		17	
·97— ·98	18·8		19		18		21		15	
·96— ·97	17·2		22		12		22		15	
·95— ·96	15·2		17		19		19		6	
·94— ·95	13·3		11		6		11		11	
·93— ·94	11·7		9		14		8		14	
·92— ·93	10·2		4		8		5		8	
·91— ·92	8·9		10		11		9		7	
·90— ·91	7·7		5		12		4		7	
·89— ·90	6·7	12·6	8	14	3	12	9	17	1	8
·88— ·89	5·9		6		9		8		7	
·87— ·88	5·1	9·5	7	13	6	11	3	11	5	12
·86— ·87	4·4		6		5		8		7	
·85— ·86	3·8	10·0	4	9	1	5	3	8	5	11
·84— ·85	3·3		2		2		2		3	
·83— ·84	2·9		3		2		3		3	
·82— ·83	2·5	8·0	1	5	4	12	6	12	4	12
·81— ·82	2·1		3		3		3		4	
·80— ·81	1·8		0		3		..		2	
·79— ·80	1·6		1		2		3		2	
·78— ·79	1·3	8·7	2	6	3	19	1	10	4	27
·77— ·78	1·1		1		3		1		2	
·76— ·77	1·0		1		2		..		4	
·75— ·76	·8		..		1		..		2	
·74— ·75	·7		..		..		2		2	
·73— ·74	·6		1		2		..		2	
·72— ·73	·5		..		..		2		..	
·71— ·72	·4		1		1		..		1	
·70— ·71	·4		..		1		..		2	
Less than ·70	1·9		..		6		4		8	
Totals	200·0		200		200		200		200	
Mean L	·9305		·9383		·9161		·9265		·9078	
σ_L	·0648		·0573		·0769		·0690		·0860	
χ^2	..		12·4		30·7		14·4		54·0	
$P(\chi^2)$	..		·5744		·0025		·4211		·000004	

law (27). The number of pairs of samples used are 200 when the sizes of two samples are 10 and 20 and also when they are each equal to 20. In the

TABLE V.

$n_1 = n_2 = 20.$

Limits of L	Normal Theory Frequencies		Frequencies in Samples from Experimental Populations (0·0, 2·5)		(0·0, 4·1)		(0·2, 3·3)		(1·0, 3·8)	
·99—1·00	34·2		33		28		30		31	
·98— ·99	28·6		34		20		24		28	
·97— ·98	23·8		18		18		31		25	
·96— ·97	20·0		21		20		18		23	
·95— ·96	16·6		13		14		18		15	
·94— ·95	13·8		16		15		14		10	
·93— ·94	11·4		16		16		10		4	
·92— ·93	9·4		12		8		11		11	
·91— ·92	7·8	14·2	4	14	6	16	7	12	8	13
·90— ·91	6·4		10		10		5		5	
·89— ·90	5·2	9·6	4	11	5	9	4	6	4	5
·88— ·89	4·4		7		4		2		1	
·87— ·88	3·6		4		3		5		4	
·86— ·87	2·8	8·8	..	5	4	15	3	13	1	10
·85— ·86	2·4		1		8		5		5	
·84— ·85	2·0		3		3		3		6	
·83— ·84	1·6		2		4		3		5	
·82— ·83	1·2		1		4		2		4	
·81— ·82	1·0		1		1		1		..	
·80— ·81	·8	9·6	..	7	1	21	..	13	1	25
·79— ·80	·7		..		3		1		..	
·78— ·79	·6		..		1		1		1	
·77— ·78	·5		..		1		..		..	
Less than ·77	1·2		..		3		2		8	
Totals	200·0		200		200		200		200	
Mean L	·9491		·9521		·9342		·9459		·9395	
σ_L	·0484		·0404		·0572		·0498		·0605	
χ^2	..		8·4		25·8		9·0		34·4	
$P(\chi^2)$	..		·6767		·0070		·6219		·0009	

lower part of the Tables are given the values of mean L, σ_L and the results of applying the (P, χ^2) test for goodness of fit.

The following points are suggested by an examination of the Tables :

(i) The agreement of the observed frequencies of L with those expected is satisfactory in the case of moderately non-normal distributions (0, 2·5) ; (·2, 3·3) ; while it is far from satisfactory in the case of the extremely skew or leptokurtic distributions.

(ii) A comparison of the means and the standard deviations of L suggests the same conclusions.

(iii) A comparison of the detailed frequencies shows that there is a tendency for too many low values of L to occur when the variation markedly differs from the normal. This is of particular interest because it is this tail of the sampling distribution which is of importance in tests of significance. In Table VI a comparison is made at about the levels $p_L = \cdot 05$ and $p_L = \cdot 02$ of the frequencies, theoretical and observed, of obtaining the value of L less than the value indicated in the third column of the Table.

TABLE VI.

Frequencies in two-hundred samples less than a given value of L.

Size of Samples		Value of L less than	Expected on Normal Theory	Observed in Experimental sampling from the given distributions			
n_1	n_2			(0·0, 2·5)	(0·0, 4·1)	(0·2, 3·3)	(1·0, 3·8)
10	20	·800	10·3	7	21	13	29
		·740	3·8	2	8	6	13
20	20	·850	9·6	7	21	13	25
		·810	3·8	0	9	4	10

The observed frequencies are of course subject to random sampling errors but it appears from Table VI that the difference (observed — expected frequency) increases both as the population β_1 and β_2 increase. For the moderately skew population (0·2, 3·3), however, little error appears to arise.

Thus on the whole it will be found that there is little to fear in the application of L test to data of moderately skew nature.

7. *Summary.*

(*a*) The use of a single comprehensive test of the hypothesis H of a common origin for the two samples is discussed. It is shown that a single best test of the hypothesis H does not exist. A comparison between L and two alternative two sample tests is made, which illustrates how the use of one test may be more efficient than the other with regard to a given alternative.

(*b*) The sampling distribution of L is obtained. It is found that the distribution involves an elliptic function of the first order when the samples are of equal size.

(*c*) Tables of the 5% and 1% levels of probability for L have been provided.

(*d*) Finally the question whether the test is suitable for application to non-normally distributed material is discussed. It is found that unless the distribution is extremely non-normal the test may be applied in practice.

In conclusion I like to express my thanks to Professor E. S. Pearson and Dr. J. Neyman for their advice and criticism while this paper was being prepared.

A Note on the Relation of the L, u *and* t *Tests. By Prof. E. S. Pearson.*

It will perhaps not be regarded as out of place if I add a few remarks to what Mr. Sukhatme has written regarding the alternative L and (t, u) tests, particularly in view of difficulty that arises in the interpretation of his skull measurement example. Here the sample point lies on the 5% L-contour but outside the 1% t-contour, and the case $n_1 = 15$, $n_2 = 13$ is very closely similar to that illustrated in Mr. Sukhatme's diagram with $n_1 = n_2 = 15$. If the reader turns to this diagram he will notice that a sample point (t, u) lying at P (or P′) on the 5% L-contour is extremely divergent judged from the point of view of t. In fact since for this point $t = 2 \cdot 62$, the chance of a more divergent positive value is less than ·01.[11] Similarly a sample at Q (or Q′)[12] is very exceptional (if H be true) judged from the point of view of u. If, however, in testing H we were to use the 5% level of the t-test whenever a sample point were in the direction of P or P′, the 5% level of the u (or z) test whenever in the direction of Q or Q′ and the L-contour only when the point was diagonally placed, we should in fact be using a far more stringent test than we may have intended. Following this procedure we should in the long run of experience be rejecting H when true in considerably more than 5% of samples, a result which follows because we are selecting the criterion with which to test H *after* we examined the sample data.

The real justification of using the t-test would appear to be that there were strong grounds for believing σ_1 to be equal to or nearly equal to σ_2, so that it was only of interest to test H_2.[13] The statistician who while believing

[11] Note that in the skull example $t = 2.63$ and $u = 0.53$.

[12] Q′ is the point where the curve cuts the u-axis nearer $u = 0$, and is not marked in the figure.

[13] See foot-note (4) for definition of H_2.

he was testing the hypothesis H, rejected this hypothesis when a sample point occurred near P or P′ would probably be influenced, perhaps unconsciously, by *a priori* considerations as to the likelihood of the certain alternatives. He would, for instance, be perhaps influenced by the common experience that population means α_1 and α_2 frequently differ while their standard deviations σ_1 and σ_2 are practically the same. Or from another point of view he would be regarding it as more important to detect a difference in means than a difference in standard deviations.[14]

In this action he might well be justified from the practical viewpoint, and Mr. Sukhatme's illustration serves to bring out admirably the difficulties connected with *a priori* probability that arise in an attempt to choose a good critical region for testing a statistical hypothesis when no best critical region exists. To recognise the difficulties is, however, an important first step towards meeting them. The intuitional basis upon which the practical statistician chooses the test he will employ is very frequently sound, but sometimes it is demonstrably at fault. As the principles of choice become more clearly outlined and generally understood these faults will cease, and we shall be left at any rate with a choice among good tests even if there is no agreement as to the best.

[14] The question of the influence of *a priori* knowledge on the contours of the sampling test was referred to by Dr. Neyman and myself in the paper introducing the likelihood criterion (*Biometrika*, **20A**, p. 190). The same point and a further one concerning the relative importance of different errors in judgment were discussed later in a paper entitled "The Testing of Statistical Hypothesis in Relation to Probabilities *a priori*." *Proc. Camb. Phil. Soc.*, **29**, pp. 492-510.