Reprinted from

SANKHYĀ : THE INDIAN JOURNAL OF STATISTICS

VOL. 1. PART 1. JUNE, 1933

EDITED BY P. C. MAHALANOBIS

TABLES FOR THE APPLICATION OF L-TESTS

BY

PROF. P. C. MAHALANOBIS.

ART PRESS
20, BRITISH INDIAN
STREET, CALCUTTA

SANKHYA: THE INDIAN JOURNAL OF STATISTICS.

Vol. 1. Part 1. June, 1933.

Sankhyā is published with the help of the Indian Statistical Institute and in consultation with the following Journal Committee appointed by the Institute.

Each volume of *Sankhyā* will contain about 400 pages, with plates and tables, and will be issued in approximately four equal parts.

The Journal is supplied free to each member of the Indian Statistical Institute. The subscription price including packing and postage for non-members, payable in advance, is Rs. 20/- per volume ; single issues Rs. 6/- including postage. Rates for foreign countries to be calculated on the basis of 32 shillings for each volume, and 9 shillings for each number. Subscriptions should be sent to the Art Press, 20, British Indian Street, Calcutta.

Papers for publication and books and reprints for review should be sent to the Editor, Prof. P. C. Mahalanobis, Presidency College, Calcutta.

It is a condition of publication in *Sankhyā* that the paper shall not already have been published elsewhere, and will not be reprinted without leave of the Editor. Manuscripts should be written on one side of the paper, and should be preferably typewritten. It is necessary that the raw data or measurements on which the paper is based, and detailed information of their sources, should be furnished in every case. It is desirable that the distributions or correlation charts from which statistical constants are calculated should accompany the paper. Tables should be numbered consecutively with Arabic figures without brackets (Table 1, Table 45, etc.), and should have short explanatory headings. Diagrams and drawings should also be numbered consecutively with Arabic figures, and should be sent in a finished state suitable for direct reproduction with all lettering in pencil. References to tables and diagrams should be given within brackets in the text. The list of literature cited should be arranged alphabetically according to authors' names. In the case of Journals the volume number, year and page references, and in the case of printed books the name of the author and the publisher, the date of publication, and the number of the edition or impression should always be stated clearly. Abbreviations, local names, vernacular measures, etc., should be written very legibly (preferably in block letters) and should be explained where they first occur in the text. A detailed Note for Contributors has been prepared and may be obtained on application to the Editor.

Contributors receive 25 copies of their papers free. The cost of additional reprints will depend upon the number required ; for 50 additional copies the rate is Rs. 6/- for each sheet of eight pages or Rs. 2/- per page for part of a sheet, with an extra charge for plates printed on special paper. Additional copies should be ordered at the time of returning the final proof.

TABLES FOR L-TESTS.

By P. C. MAHALANOBIS.

Introduction.*

E. S. Pearson and J. Neyman[1] have recently considered the problem of testing whether a number of given samples all belong to the same normal population or are statistically differentiated in their mean values or variabilities. They have considered three cases:—

(1) The hypothesis H_0 that the samples belong to normal populations having the same mean value and the same standard deviation.

(2) The hypothesis H_1 that the samples come from populations having the same standard deviation, it being immaterial whether the mean values are equal or unequal.

(3) The hypothesis H_2 that the mean values are appreciably equal, it being assumed that the standard deviations are also equal.

They have given three different statistics which I am calling L_0, L_1, and L_2 to test the above three hypotheses respectively.

Let n_t, m_t and s_t^2 be the size, the mean value and the variance of the t^{th} sample Then

$$n_t . m_t = S(x), \qquad n_t . s_t^2 = S(x - m_t)^2 \qquad \dots (1)$$

where S represents a summation for all n_t values of x. Let there be k such samples.

The average variance s_a^2 within all samples is defined by

$$N = \Sigma (n_t), \qquad N . s_a^2 = \Sigma (n_t . s_t^2) \qquad \dots (2)$$

where N is the total number of individual observations available, and Σ represents a summation for all k samples.

Finally the general mean, m_0, and the general variance, s_0^2 are defined by

$$N . m_0 = \Sigma S(x), \qquad N . s_0^2 = \Sigma S (x - m_0)^2 \qquad \dots (3)$$

where ΣS represents a summation for all N values of x.

It will be noticed that s_0^2 is the 'total' variance, and s_a^2 the mean variance 'within samples' ordinarily used in the analysis of variance.

The working formulas take a sample form when the size of the sample is the same for all samples, *i.e.*

$$n_1 = n_2 \quad \dots \quad = n_k = n, \qquad N = n . k \qquad \dots (4)$$

---

*Discussed at a meeting of the Indian Statistical Institute held in the Presidency College, Calcutta, on Friday the 10th December, 1932.

[1] "On the Problem of k Samples." *Bulletin de l'Académie Polonaise des Sciences et des Lettres, Série A*, 1931, pp. 460-481.

Pearson and Neyman's formulas can in this case be written in the following form:—

$$L_0=\left\{\frac{s_1^2.\ s_2^2.\ s_3^2\ \ldots\ldots\ldots\ldots\ s_k^2}{s_0^2.\ s_0^2.\ s_0^2\ \ldots\ldots\ldots\ldots\ s_0^2}\right\}^{1/k} \qquad \ldots \qquad \ldots \qquad (5)$$

$$L_1=\left\{\frac{s_1^2.\ s_2^2.\ s_3^2\ \ldots\ldots\ldots\ldots\ s_k^2}{s_a^2.\ s_a^2.\ s_a^2\ \ldots\ldots\ldots\ldots\ s_a^2}\right\}^{1/k} \qquad \ldots \qquad \ldots \qquad (6)$$

$$L_2=s_a^2/s_0^2 \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad (7)$$

It is convenient to introduce the geometric mean of the variances defined by

$$s_g^2=(s_1^2.\ s_2^2.\ s_3^2\ \ldots\ldots\ldots\ldots\ s_k^2)^{1/k} \qquad \ldots \qquad \ldots \qquad \ldots \qquad (8)$$

$$\log (s_g^2)=(1/k)\ [\log s_1^2 + \log s_2^2 + \ldots\ldots\ldots\ldots \log s_k^2] \qquad \ldots \qquad (8{\cdot}1)$$

$$\log L_0=\log (s_g^2)-\log (s_0^2) \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad (5{\cdot}1)$$

$$\log L_1=\log (s_g^2)-\log (s_a^2) \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad (6{\cdot}1)$$

$$\log L_2=\log (s_a^2)-\log (s_0^2) \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad (7{\cdot}1)$$

It will be noticed that

$$\log L_0=\log L_1+\log L_2 \qquad \text{or} \qquad L_0=(L_1).\ (L_2) \qquad \ldots \qquad (9)$$

The significance of L_0, L_1, and L_2 can be stated in words. If hypothesis H_0 is true, that is, if all k samples are drawn from a population having the same mean value and the same variance than L_0 will be approximately equal to unity. On the other hand if L_0 is appreciably less than unity, we should conclude that either the mean values or the variabilities or both are different in the different samples.

In the same way if hypothesis H_1 is true, that is, if the k samples are drawn from populations having the same standard deviation, then the value of L_1 will be sensibly equal to unity. Thus if L_1 is appreciably less than unity, it would be reasonable to infer that the standard deviations of the different samples are not equal. (For the purposes of this test, it is immaterial whether mean values are identical or different).

Finally if hypothesis H_2 is true, that is, if the samples come from populations having the same mean value, (it being also assumed that they also have the same variability) then the value of L_2 will be equal to unity. Thus when L_2 is appreciably lower than unity the mean values cannot be considered to be equal.

It can be easily shown that $L_2=(1-\eta^2)$ where η is Karl Pearson's correlation ratio. Further it can be shown that

$$L_2=1\Big/\left\{1+\frac{k-1}{N-k}\cdot e^{2z}\right\} \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad (10)$$

where z is the function used in Fisher's test for the analysis of variance. It will be noticed

$$L_2 \rightarrow (N-k)/(N-1), \text{ as } z\rightarrow 0 \text{ and } L_2\rightarrow 0, \text{ as } z\rightarrow \infty \qquad \ldots \qquad (10.1)$$

The statistic L_2 thus simply furnishes an alternative form of Fisher's z-test, and need not be considered here further.

The exact distributions of L_0 and L_1 are not known, but Pearson and Neyman have shown that on certain assumptions the p^{th} moment-coefficient, μ'_p for L_0 and L_1 can be obtained approximately by the following equations.

Writing $N=n.k$, $a=(n-1)/2$, $g=a+(p/k)$, $d=(N-1)/2$, and $e=N-k)/2$, we have

$$\mu'_p\,(L_0) = k^p.\,[\Gamma(g)/\Gamma(a)]^k\,.\,[\Gamma(d)/\Gamma(d+p)] \quad \ldots \quad \ldots \quad (11)$$

$$\mu'_p\,(L_1) = k^p.\,[\Gamma(g)/\Gamma(a)]^k\,.\,[\Gamma(e)/\Gamma(e+p)] \quad \ldots \quad \ldots \quad (12)$$

Putting $p=1$, 2, 3, and 4, it is thus possible to get the mean value $(=\mu'_1)$, the variance $(=\mu_2)$ and also $\beta_1=\mu_3^2/\mu_2^3$ and $\beta_2=\mu_4/\mu_2^2$ for any assigned value of n and k. The probability of occurrence of L_1 can then be obtained by quadrature, but the process will be extremely troublesome. Pearson and Neyman have, however, shown that on certain plausible assumptions the 5% and 1% points for L_0 and L_1 can be obtained approximately in the following way. One pair of values of n_1 and n_2 are calculated for both L_0 and L_1 with the help of equation (13). The corresponding 5% and 1% values of z are next found by interpolation from Fisher's z-table (Statistical Methods for Research Workers, 4th edition, 1932, Tables V and VI). Finally the 5% and 1% values of L_1 are obtained by using these interpolated values of z in equation (14).

$$n_1=\frac{2(1-\mu'_1)\,(\mu'_1-\mu'_2)}{\mu'_2-(\mu'_1)^2} \qquad n_2=\frac{2\mu'_1\,(\mu'_2-\mu'_1)}{\mu'_2-(\mu'_1)^2} \quad \ldots \quad \ldots \quad (13)$$

$$L_1 = n_2/(n_2+n_1.e^{2z}) \quad \ldots \quad \ldots \quad \ldots \quad (14)$$

Construction of the Tables.

Writing $N=n.k$, $a=(n-1)/2$, $b=a+(1/k)$, $c=a+(2/k)$, $d=(N-1)/2$, and $e=(N-k)/2$ in equations (11) and (12), and

$$A \equiv k.[\Gamma(b)/\Gamma(a)]^k \qquad B \equiv k^2[\Gamma(c)/\Gamma(a)]^k \quad \ldots \quad (15)$$

and putting $p=1$, and 2 for μ'_1, μ'_2 respectively we get the following values:—

$$\mu'_1(L_0) = A/d\ , \qquad \mu'_2(L_0) = B/d.(d+1) \quad \ldots \quad \ldots \quad (16)$$

$$\mu'_1(L_1) = A/e\ , \qquad \mu'_2(L_1) = B/e.(e+1) \quad \ldots \quad \ldots \quad (17)$$

We have used values of $n=2$, 3, 4, 5, 10, 15, 20, 30, 40, 50 and $k=2$, 3, 4, 5, 10, 20, 25, 50.

The values of a will therefore vary from 0·5 to 24·5, and hence of b from 0·52 to 25·0, and of c from 0·54 to 25·5. The values of $\log \Gamma(a)$, $\log \Gamma(b)$, and $\log \Gamma(c)$ can therefore be directly obtained from Tracts for Computers* No. IX except for 9 values between 0 and 1 which can be easily calculated with the help of the same table.

It was thus easy to obtain the values of μ'_1 and μ'_2 and hence the value of $\mu_2=\mu'_2-(\mu_1)^2$. Here μ'_1 represents the mean value of L_1 and $\sqrt{\mu_2}$ the standard deviation of L_1. These values are given in Table 1 (for L_0) and Table 2 (for L_1).

---

*Log $\Gamma(x)$ from $x=1$ to 50·9 by intervals of ·01 by John Brownlee Edited by Karl Pearson. Cambridge University Press, 1923.

The next step was to calculate n_1 and n_2 with the help of equation (13). This work was done on a Brunsviga machine retaining 7 decimal places in μ'_1 and μ'_2. The corresponding values of 5% and 1% z were then determined by harmonic interpolation from Fisher's Tables‡.

Finally L was calculated from equation (14) which can be put more conveniently in the form :—

$$L = 1/[1+(n_1/n_2).\ e^{2z}] \quad \ldots \quad \ldots \quad \ldots \quad (14{\cdot}1)$$

This equation lends itself easily to the use of a table of Addition-Substraction logarithms§, which gives log $(a+b)$ directly from knowledge of log (a) and log (b). Putting $a=1$ or log $(a)=0$, and $b=(n_1/n_2)\ .\ e^{2z}$ it will be noticed further that

$$\log L = -\log(1+b) \quad \ldots \quad \ldots \quad \ldots \quad (14{\cdot}2)$$

$$\log(b) = \log(n_1) - \log(n_2) + 2z.\ \log(e) \quad \ldots \quad \ldots \quad (14{\cdot}3)$$

The 5% and 1% values of L_0 and L_1 are given in Tables 3, 4, 5, and 6.

In an earlier paper Pearson and Neyman gave a short Table of the 5% and 10% values of a connected statistics λ_H for $k=2$, i.e., for two samples with n (the size of the two samples) = 5, 10, 20, 50 and ∞. This Table has been reproduced as Table XXXVII(b) on pp. 223-224 of Tables for Statisticians and Biometricians, Part II, edited by Karl Pearson. But $\lambda_H = L_0^n$, when $n_1=n_2=n$, and we find that the values tabled here agree satisfactorily with the values given by Pearson and Neyman for $n=10$, 20, and 50 for the 5% point, and $n=5$, 20, and 50 for the 1% point. I am unable to reconcile the discrepancies in the values† for $n=5$ for the 5% point, and for $n=10$ for the 1% point.

In using the present Tables it is necessary to remember that equations (13) and (14) can only furnish approximate results. Neyman and Pearson are of opinion (p. 478) that the approximate values may be inadequate for practical purposes when $n=2$, and $n=3$. Also the value of n must not be too large (p. 475). The approximate character of the results probably explains the anomalies in the values of L_0 and L_1 for $n=2$ and $n=3$.

Apart from the theoretical consideration discussed above, the uncertainties of interpolation in the z-tables (especially for large values of n_1 and n_2) introduce appreciable errors. It is not unlikely, therefore, that the values given in the present tables are in error to the extent of two or three units in the third place of decimal. I am having this question examined carefully ; I have retained four figures for the present pending such scrutiny.

---

‡R. A. Fisher, "Statistical Methods for Research Workers" Oliver and Boyd, 4th edition, 1932.

§Addition-Substraction Logarithms to Five Decimal Figures by L. M. Berkeley. White Book and Supply Co., New York City, 1930.

†For $n=5$, the 5% point from Neyman and Pearson's value of λ_H is 0·4411 against 0·4424 in the present Table, and for $n=10$, the 1% point is 0·5863 against 0·5870 in the present Table.

Table 1—Mean Value and Standard Deviation of L_0

	n=2		n=3		n=4		n=5		n=10	
k	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.
2	·4244	·2942	·6283	·2497	·7276	·2051	·7854	·1721	·8963	·0934
3	·3100	·2268	·5341	·2141	·6523	·1821	·7234	·1553	·8639	·0865
4	·2611	·1864	·4909	·1874	·6169	·1623	·6938	·1396	·8481	·0787
5	·2341	·1602	·4661	·1680	·5963	·1472	·6766	·1273	·8387	·0723
10	·1845	·1022	·4188	·1185	·5566	·1064	·6429	·0930	·8202	·0537
20	·1618	·0676	·3962	·0834	·5374	·0759	·6266	·0667	·8111	·0388
25	·1574	·0595	·3918	·0745	·5336	·0680	·6233	·0598	·8093	·0349
50	·1488	·0408	·3830	·0525	·5261	·0482	·6169	·0425	·8057	·0248

	n=15		n=20		n=30		n=40		n=50	
k	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.
2	·9317	·0638	·9491	·0484	·9663	·0326	·9748	·0246	·9799	·0197
3	·9098	·0595	·9326	·0453	·9552	·0306	·9665	·0231	·9732	·0186
4	·8991	·0543	·9244	·0414	·9498	·0280	·9624	·0212	·9699	·0170
5	·8927	·0500	·9196	·0381	·9465	·0258	·9599	·0196	·9679	·0157
10	·8800	·0372	·9100	·0285	·9400	·0193	·9550	·0146	·9640	·0118
20	·8738	·0270	·9052	·0207	·9368	·0138	·9525	·0106	·9620	·0087
25	·8725	·0243	·9043	·0186	·9361	·0126	·9521	·0096	·9616	·0080
50	·8700	·0173	·9024	·0145	·9348	·0091	·9511	·0073	·9608	·0061

TABLE 2—MEAN VALUE AND STANDARD DEVIATION OF L_1

	n=2		n=3		n=4		n=5		n=10	
k	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.
2	·6366	·3078	·7834	·2232	·8488	·1717	·8836	·1889	·9461	·0704
3	·5166	·2780	·7121	·2115	·7972	·1640	·8439	·1327	·9279	·0670
4	·4569	·2478	·6750	·1947	·7711	·1516	·8239	·1228	·9188	·0619
5	·4213	·2238	·6525	·1800	·7553	·1406	·8119	·1139	·9133	·0573
10	·3506	·1582	·6073	·1352	·7236	·1064	·7876	·0863	·9022	·0433
20	·3155	·1104	·5845	·0983	·7076	·0777	·7754	·0630	·8967	·0315
25	·3085	·0983	·5799	·0883	·7043	·0699	·7729	·0567	·8956	·0285
50	·2946	·0674	·5707	·0631	·6979	·0499	·7630	·0406	·8934	·0203

	n=15		n=20		n=30		n=40		n=50	
k	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.	Mean	S. D.
2	·9649	·0471	·9740	·0353	·9829	·0236	·9873	·0177	·9898	·0141
3	·9532	·0447	·9653	·0335	·9772	·0223	·9830	·0167	·9865	·0133
4	·9473	·0412	·9610	·0308	·9743	·0205	·9809	·0154	·9848	·0123
5	·9437	·0381	·9584	·0285	·9726	·0190	·9796	·0142	·9837	·0113
10	·9366	·0287	·9531	·0214	·9692	·0143	·9770	·0107	·9817	·0086
20	·9330	·0210	·9505	·0157	·9674	·0101	·9758	·0078	·9809	·0064
25	·9323	·0189	·9500	·0141	·9671	·0094	·9755	·0070	·9805	·0060
50	·9309	·0135	·9489	·0103	·9664	·0068	·9750	·0057	·9801	·0048

ILLUSTRATIONS.

Example 1. Five samples of ten observations each are given in the following Table. It is desired to test whether all the five samples may be considered to have been drawn from the same universe.

(1)	(2)	(3)	(4)	(5)
107·87	110·31	109·71	111·09	111·03
109·64	108·89	110·29	111·06	111·74
109·67	110·67	111·43	108 93	109·16
109·84	112·69	109·93	111·27	108·32
109·68	110·60	109·49	110·43	111·27
111·24	109·36	109·65	108·65	109·59
108·96	110·36	111·28	109·33	110·06
111·22	111·87	110·26	111·00	110·72
110·00	111·64	110·56	110·39	109·82
109·70	109·42	110·34	109·41	109·66

The variances for each sample are directly calculated, and the logarithms entered in tabular form as shown below.

Sample	Variance (s^2)	log (s^2)
1	0·8626	$\bar{1}$·93580
2	1·3060	0·11594
3	0·3871	$\bar{1}$·58787
4	0·8783	$\bar{1}$·94362
5	0·9880	$\bar{1}$·99475
Sum	4·4220	$\bar{1}$·57798
Mean	0·8844	$\bar{1}$·91560

The mean variance within samples is thus $s_a^2=0{\cdot}8844$. Also since s_g^2 is the geometric mean of the variances, log $(s_g^2)=\bar{1}{\cdot}91560$. Further the general variance, $(s_o^2)=0{\cdot}9512$. Hence log $(s_g^2)=\bar{1}{\cdot}91560$, log $(s_a^2)=\bar{1}{\cdot}94664$, log $(s_o^2)=\bar{1}{\cdot}97827$. We then have

$$\log (L_0)=\log (s_g^2)-\log (s_o^2)=\bar{1}{\cdot}91560-\bar{1}{\cdot}97827=\bar{1}{\cdot}93732$$

$$\log (L_1)=\log (s_a^2)-\log (s_o^2)=\bar{1}{\cdot}91560-\bar{1}{\cdot}94664=\bar{1}{\cdot}96895$$

$$\log (L_2)=\log (s_a^2)-\log (s_o^2)=\bar{1}{\cdot}94664-\bar{1}{\cdot}97827=\bar{1}{\cdot}96837$$

We get finally: $L_0=0{\cdot}8656$, $L_1=0{\cdot}9310$, $L_2=0{\cdot}9298$.

From Table 3, we find that for $n=10$, $k=5$, the 5% point for $L_0=0{\cdot}7057$, and from Table 4, the 1% point for $L_0=0{\cdot}6367$. The observed value of $L_0=0{\cdot}8656$ cannot be considered significantly less than unity. The samples may, therefore, be considered to have been drawn from the same universe.

The test for L_1 also naturally fails. From Tables 5 and 6 we find that the 5% and 1% points for L_1 are 0·8025 and 0·7350 respectively, while the observed value is 0·9310.

The analysis of variance given below shows the absence of any heterogeneity in mean values. This is corroborated by the high value of $L_2 = 0{\cdot}9298$.

Variation			D. F.	Sum of Squares	Variance
Between samples	...	...	4	8·4318	0·8353
Within samples	...	...	45	44·2187	0·9826
Total		...	49	47·5600	...

The data actually represented random samples (drawn with the help of Random Numbers*) from the same normal population with a mean value of 110, and standard deviation = 1.

Example 2. Consider a second set of 5 samples of size 10 each.

(1)	(2)	(3)	(4)	(5)
207·87	212·31	218·71	217·09	219·03
209·64	210 89	214·29	217·06	219·74
209·67	212·67	215·43	214·93	217·16
209·84	214·69	213·93	217·27	216·32
209·68	212·60	213·49	216·43	219·27
210·24	211·36	213·65	214·65	217·59
208 96	212,36	214·56	215·93	218·06
211·22	213·87	215·28	217·00	218·72
210 00	213·64	214·26	216·39	217·62
209·70	211·42	214·34	215·41	217·66

The variances within samples are given below. The value of $s_0^2 = 9{\cdot}3800$.

Sample	Variance (s^2)	log (s^2)
1	0·6610	$\bar{1}{\cdot}82020$
2	1·3060	0·11594
3	0·3871	$\bar{1}{\cdot}58782$
4	0·7718	$\bar{1}{\cdot}88750$
5	1·0043	0·00186
Sum	4·1302	$\bar{1}{\cdot}41332$
Mean	0·8260	$\bar{1}{\cdot}88266$

The L-statistics are then easily calculated.

$\log (s_g^2) = \bar{1}{\cdot}88267$	$\log L_0 = \bar{2}{\cdot}91046$	$L_0 = 0{\cdot}0814$
$\log (s_a^2) = 1{\cdot}91698$	$\log L_1 = \bar{1}{\cdot}96569$	$L_1 = 0{\cdot}9240$
$\log (s_0^2) = 0{\cdot}97220$	$\log L_2 = \bar{2}{\cdot}94478$	$L_2 = 0{\cdot}0881$

The observed value of $L_0 = 0{\cdot}0814$ is far below the 1% point (0·6367) showing the existence of heterogeneity.

The observed value of $L_1 = 0{\cdot}9240$ on the other hand is considerably higher than the expected 5% value 0·8025. The variances, therefore, may be considered to be significantly same in all the samples.

*Tracts for Computers No. XV, Random Sampling Numbers by L. H. C. Tippett (Edited by Karl Pearson, Cambridge University Press, 1927).

The heterogeneity of the samples is then clearly due to differences in mean values, and Fisher's z-test can be legitimately applied in this case. The analysis of variance is given below.

Variation	D. F.	Sum of Squares	Variance
Between samples	4	427·6984	106·9246
Within samples	45	41·3011	0 .9178
Total ...	49	468·9995	...

$z=2·3789$, $x=118·68$.

The ratio of the variances is $x=s_1^2/s_2^2=118·68$ while the 1% value of the ratio is less than 4·018.§ The differences between samples are clearly significant.

The samples were actually drawn from populations with mean values 210, 212, 214, 216 and 218 respectively but with the same variance $\sigma^2=1$.

Example 3. Consider a third example.

(1)	(2)	(3)	(4)	(5)
50·62	49·12	54·35	55·16	49·79
47·77	50·87	54·22	58·69	51·20
51·32	54·29	45·72	45·79	50·06
55·38	49·79	55·08	41·61	51·30
51·20	48·48	51·73	56·33	48·78
48·71	48·94	44·58	47·94	50·56
50·72	51·69	49·73	50·28	50·08
53·73	53·85	53·98	53·58	51·06
53·29	50·77	51·55	48·08	49·39
48·83	51·03	47·63	48·31	52·62

The variances within samples one given below. The value of $s_0^2=9·6339$.

Sample	Variance (s^2)	log (s^2)
1	1·1004	0·04155
2	5·2386	0·71922
3	3·4949	0·54343
4	12·9813	1·11332
5	25·0678	1·39912
Sum	47·8830	3·81664
Mean	9·5766	0·76333

The L-statistics now become:—

log (s_g^2) = 0·72333	log L_0 = $\bar{1}$·73952	L_0 = 0·5489
log (s_a^2) = 0·98121	log L_1 = $\bar{1}$·74212	L_1 = 0·5522
log (s_0^2) = 0·98380	log L_2 = $\bar{1}$·99741	L_2 = 0·9940

§This is obtained from "The Auxiliary Tables for Fisher's z-test" (*Ind. Jour. Agricultural Science*, Vol. II, Part 6, December, 1932, pp. 679—693) published by the present writer.

The observed value of $z=2·3789$, while the 1% value is less than 0·6954 for $n_1=4$, $n_2=30$ (R. A. Fisher: Statistical Methods for Research Workers, 4th edition, 1932. Table VI, p. 224).

Both L_0 and L_1 are significantly less than unity. The samples are clearly differentiated, and the variances of the different samples must also be considered different. The high value of L_2 suggests that the mean values are not differentiated, and Fisher's z-test fails completely as will be seen from the analysis of variance.

Variation	D. F.	Sum of Squares	Variance
Between samples	4	2·8682	0·7171
Within samples	45	478·8302	10·6407
Total ...	49	481·6984	...

It will be noticed that the variance between mean values is actually less than the average variance within samples.

The hypothetical populations from which the samples were drawn had the same mean value 50, but the standard deviations were equal to 1, 2, 3, 4, and 5 respectively.

Example 4. We shall consider another example.

(1)	(2)	(3)	(4)	(5)
349·79	352·62	353·12	360·35	363·16
351·20	349·77	354·87	360·22	366·69
350·06	353·32	358·29	351·72	353·79
351·30	357·38	353·79	361·08	349·61
348·78	353·20	352·48	357·73	364·33
350·56	350·71	352·94	350·58	355·94
350·08	352·72	355·69	353·73	358·28
351·06	355·73	357·85	359·98	361·58
349·39	355·29	354·77	357·55	356·08
352·62	350·83	355·03	353·63	356·31

The variances within samples are found to be the same as in Example 3, but the general variance $s_0^2 = 17·5428$. So that we now have :—

$$\log (s_g^2) = 0·72333 \qquad \log L_0 = \bar{1}·47923 \qquad L_0 = 0·3015$$
$$\log (s_a^2) = 0·98121 \qquad \log L_1 = \bar{1}·74212 \qquad L_1 = 0·5522$$
$$\log (s_0^2) = 1·24410 \qquad \log L_2 = \bar{1}·73711 \qquad L_2 = 0·5459$$

Both L_0 and L_1 are significantly less than unity. The z-test is also positive as shown below‡.

Variation	D. F.	Sum of Squares	Variance
Between samples	4	398·3082	99·5771
Within samples	45	478·8302	10·6407
Total ...	49	877·1384	...

$z = 1·1181 \quad x = 9·36.$

‡The observed value of z is 1·1181, while the 1% per cent. value is less than 0·6954.

The samples are clearly differentiated both in their mean values and in their variabilities. Actually they were drawn from the following populations:—$m_1=350$, $\sigma_1=1$; $m_2=352$, $\sigma_2=2$; $m_3=354$, $\sigma_3=3$; $m_4=356$, $\sigma_4=4$; and $m_5=358$, $\sigma_5=5$.

Example 5. The results given below were obtained by Mr. K. V. Joshi of the Cotton Research Laboratory, Surat, in his experiments on the effect of the application of manures in different months on the production of "Buds," "Flowers," "Bolls," and the ratio of "Flowers: Buds," "Bolls: Flowers," and "Bolls: Buds" in the Cotton plant in Surat.*

DATA RELATING TO THE COTTON PLANT (SAMPLES OF 20)

	Buds		Flowers		Bolls	
	Mean	Variance	Mean	Variance	Mean	Variance
Control	254·0	16900·25	99·3	1466·35	31·6	129·01
July—manure ...	380·2	7077·30	138·6	11074·05	39·7	77·39
August—manure ...	257·3	6791·42	118·7	1375·89	34·7	104·28
General ...	297·2	13779·78	118·9	4955·73	35·3	115·15

	Flowers : Buds		Bolls : Flowers		Bolls : Buds	
	Mean	Variance	Mean	Variance	Mean	Variance
Control	39·0	104·65	31·9	13·00	12·4	8·79
July—manure ..	36·5	109·65	28·6	26·05	10·4	11·09
August—manure ...	46·1	89·26	29·2	29·83	13·5	4·53
General ..	40·5	101·43	29·9	24·73	12·1	10·03

The calculated values of L_0, L_1, L_2 and z are given below.‡ Here $n=20$ and $k=3$. The 5% and 1% points are also shown for comparison.§

	L_0	L_1	L_2	z
Buds	**0·6771**	0·9046	**0·7486**	**1·1207**
Flowers	**0·5684**	**0·6000**	0·9478	0·2215
Bolls	0·8802	0·9788	0·8998	0·5718
Flowers : Buds ...	**0·7558**	0·8988	**0·8409**	**0·8250**
Bolls : Flowers ...	0·8741	0·9465	0·9286	0·4185
Bolls : Buds ...	**0·7579**	0·9287	**0·8176**	**0·9179**
5% values	0·8450	0·8980	(0·8999)	0·5764
1% values	0·7900	0·8476	(0·8508)	0·8065

*P. C. Mahalanobis: "Effect of Fertilizers on the Variability of the Yield and Rate of Shedding of Buds, Flowers, and Bolls in the Cotton Plant in Surat" (*Ind. Jour. Agri. Science*, Vol. III, Part I, February, 1933).

‡Significant values are shown in heavy type.

§The 5% and 1% points for L_2 were obtained indirectly from the corresponding 5% and 1% points for z and give only approximate values for purposes of comparison.

The L_0-test shows that there is no appreciable effect in the case of "Bolls," and the proportion of "Bolls: Flowers," while differences in "Flowers," "Bolls" and the proportion of "Flowers: Buds," and "Bolls: Buds" are significant. The L_1-test next shows that in the case of "Flowers" the variabilities are appreciably affected, but there is no effect on the mean values. Finally, the z-test reveals that the differentiation in the case "Buds," "Flowers: Buds," and "Bolls: Buds" arise through a change in mean values (but not in variabilities).

General Procedure.

The above examples suggest the following systematic procedure.

(1) First apply the L_0-test. If the test is negative, *i.e.*, if L_0 is found to be sensibly equal to unity (*i.e.*, not significantly different from 1) then the samples may be considered to have been drawn from the same population having the same mean value and the same standard deviation.

(2) If the L_0-test is positive, *i.e.*, if L_0 is significantly less than unity then the samples must be considered to be statistically differentiated.

(3) In this case next use the L_1-test. If L_1 is sensibly equal to unity, then the variabilities may be considered to be equal. Fisher's z-test (which can be now used with safety) should be next applied and would reveal the differences in mean values.

(4) On the other hand if L_1 is significantly less than unity, the variabilities cannot be considered to be the same. If the z-test is also positive, then the samples must be considered to be differing in both mean values and variabilities.

(5) Finally when the existence of differentiation in either mean values or variabilities has been detected, the analysis may be pushed further by a comparison of the different samples *in pairs* with the help of (*a*) Fisher's t-test for mean values, and (*b*) the z-test for variabilities.

Conclusion.

There are certain definite advantages in the use of the L-statistics in case of more than two samples. The L_0-test shows immediately whether there exists any significant differentiation in either mean values or variabilities. The use of the L_1-test and the z-test would then locate the cause of differentiation. Further the use of the t-test (for mean values), or the z-test (for variabilities) require the splitting up of the whole group of k samples into $k(k-1)/2$ pairs for purposes of comparison, so that $k(k-1)/2$ separate results are obtained in each case some of which may be in agreement while others are discrepant. The L_0 and L_1 (and L_2) tests on the other hand refer to all the k samples taken together, and each of them furnish a kind of over-all criterion for the group of samples as a whole.

The heavy work of computation in preparing the present Tables was done by Mr. Jitendramohan Sen Gupta, B.Sc., and the results were checked by Mr. Subhendu Sekhar Bose, M.Sc., and other workers of the Statistical Laboratory, Presidency College, Calcutta.

TABLE 3.—FIVE PER CENT. VALUES OF L_0

Values of *k* (number of samples)	Values of *n* (size of samples.) 2	3	4	5	10	15	20	30	40	50
2	·0190	·1719	·3277	·4424	·7071	·8028	·8515	·9007	·9254	·9403
3	·0217	·1745	·3205	·4290	·6977	·7950	·8450	·8960	·9216	·9372
4	·0258	·1846	·3328	·4416	·7008	·7968	·8463	·8967	·9222	·9376
5	·0301	·1946	·3421	·4511	·7057	·8000	·8483	·8983	·9234	·9384
10	·0470	·2296	·3792	·4857	·7262	·8142	·8595	·9057	·9294	·9432
20	·0656	·2640	·4149	·5193	·7549	·8358	·8766	·9183	·9383	·9511
25	·0721	·2744	·4204	·5338	·7635	·8422	·8816	·9210	·9408	·9519
50	·0878	·3036	·4644	·5731	·7865	·8550	·8891	·9275	·9449	·9554

TABLE 4.—ONE PER CENT. VALUES OF L_0.

Values of *k* (number of samples)	Values of *n* (size of samples.) 2	3	4	5	10	15	20	30	40	50
2	·0023	·0590	·1804	·2856	·5870	·7185	·7811	·8516	·8875	·9097
3	·0024	·0856	·2053	·3112	·6043	·7253	·7900	·8575	·8922	·9133
4	·0065	·1031	·2298	·3353	·6219	·7382	·8001	·8644	·8975	·9176
5	·0092	·1180	·2478	·3558	·6367	·7490	·8086	·8702	·9018	·9210
10	·0229	·1679	·3093	·4171	·6804	·7809	·8335	·8877	·9154	·9321
20	·0422	·2166	·3656	·4736	·7286	·8180	·8632	·9098	·9316	·9443
25	·0495	·2315	·3814	·4954	·7428	·8285	·8710	·9142	·9356	·9478
50	·0688	·2729	·4389	·5464	·7727	·8486	·8833	·9243	·9422	·9531

TABLE 5.—FIVE PER CENT. VALUES OF L_1.

Values of k (number of samples)	Values of n (size of samples.) 2	3	4	5	10	15	20	30	40	50
2	·0723	·3107	·4782	·5842	·7985	·8673	·9014	·9349	·9512	·9612
3	·0704	·3040	·4696	·5755	·7925	·8632	·8980	·9325	·9495	·9598
4	·0753	·3152	·4800	·5849	·7970	·8662	·9005	·9341	·9506	·9608
5	·0825	·3278	·4915	·5950	·8025	·8699	·9032	·9358	·9519	·9618
10	·1185	·3738	·5341	·6318	·8228	·8813	·9135	·9427	·9573	·9659
20	·1472	·4191	·5697	·6658	·8417	·8961	·9227	·9498	·9619	·9694
25	·1578	·4320	·5841	·6747	·8453	·8989	·9250	·9508	·9630	·9697
50	·1878	·4723	·6278	·7137	·8688	·9150	·9365	·9584	·9672	·9730

TABLE 6.—ONE PER CENT. VALUES OF L_1.

Values of k (number of samples)	Values of n (size of samples) 2	3	4	5	10	15	20	30	40	50
2	·0126	·1361	·2818	·3855	·6782	·7821	·8359	·8902	·9171	·9340
3	·0169	·1615	·3138	·4291	·6992	·7976	·8476	·8981	·9234	·9388
4	·0233	·1876	·3414	·4594	·7194	·8118	·8586	·9056	·9291	·9434
5	·0285	·2101	·3703	·4838	·7350	·8228	·8645	·9113	·9334	·9470
10	·0624	·2838	·4483	·5556	·7791	·8537	·8908	·9273	·9458	·9565
20	·0997	·3524	·5138	·6142	·8131	·8768	·9082	·9404	·9546	·9633
25	·1129	·3718	·5304	·6287	·8201	·8817	·9121	·9415	·9565	·9640
50	·1517	·4328	·5915	·6672	·8576	·9077	·9307	·9548	·9643	·9697

INDIAN STATISTICAL INSTITUTE

The Indian Statistical Institute is a Society registered under the Indian Societies Registration Act (Act No. XXI of 1860).

Extracts from the Memorandum of Association.

II. The objects of the Society shall be:—

(1) To promote the study of Statistics both pure and applied and allied subjects.

(2) To provide for research and instruction for the advancement of the study and dissemination of knowledge of Statistics and allied subjects.

* * * * * * *

IV. The membership of the Institute and its constituent bodies shall be open to all persons irrespective of sex, nationality, race, creed or class.

V. The Institute shall not distribute any part of its income or profit as dividend, gift, or bonus or in any other way among its members, provided nevertheless that office-bearers, workers, and other employees shall not be debarred from receiving remuneration or participating in grants because of their also being members of the Institute.

The Institute consists of Associates, Ordinary and Life Members, Fellows, Honorary Members and Patrons. The supreme control including the power of making rules is vested in the members in General Meeting assembled. The President and other office-bearers are elected annually. The management of the affairs of the Institute is in the hands of a Council elected annually by the members of the Institute. There is a standing Technical Committee consisting of members who are specially qualified by their knowledge of statistics to undertake the direction or review of special investigations and technical problems. Members of the Technical Committee are appointed by the Technical Committee itself, and it is provided in the Rules that no opinion in technical questions shall be given on behalf of the Institute without the approval of the Technical Committee.

Membership Fee Rs. 15/- per year includes subscription to *Sankhyā: The Indian Journal of Statistics.*

Detailed information and Membership Forms may be obtained on application to the Hon. Secretary, Prof. P. C. Mahalanobis, Presidency College, Calcutta.