ON AN EXACT TEST OF ASSOCIATION BETWEEN THE OCCURRENCE OF THUNDERSTORM AND AN ABNORMAL IONISATION

By S. S. BOSE AND P. C. MAHALANOBIS.

J. N. Bhar and P. Syam[1] had recently sent us some data relating to the occurrence of thunderstorms and the ionisation of the Kennelly-Heaviside layer near Calcutta and had invited our opinion as to the existence of any statistical association between them. When simultaneous observations on two variables are given in quantitative measure, the most efficient measure of association is Coefficient of Correlation. But if the variates are given in categories, Kearl Pearson[2] had shown that an appropriate measure of association can be obtained by the mean square contingency χ^2. Using this $P(\chi^2)$ test we found a significant association between the occurrence of thunderstorms and the existence of abnormal ionisation, and these results were published by Bhar and Syam in a recent paper in the *Philosophical Magazine*[1].

R. A. Fisher has however pointed out one serious objection to applying Pearson's $P(\chi^2)$ test: "The distribution of χ^2 is a continuous distribution. The distribution of frequencies must, however, be always discontinuous. Consequently, the use of χ^2 in the comparison of observed with expected frequencies can only have approximate accuracy. It was in order to avoid the iregularities produced by small numbers that it has been stipulated that in no group shall the expected number be less than five"[3] We shall discuss in this note the application of more exact tests to Bhar and Syam's data.

One of the contingency tables (Table V, p. 525) given in their paper is shown below.

TABLE 1. FREQUENCY DISTRIBUTION OF THUNDERSTORM AND IONISATION.

(*a*) Occasions of No-Thunderstorm.

	Normal	Abnormal	Total
No Magnetic disturbance	79	22	101
Magnetic disturbance	32	17	49
Total	111	39	150

(*b*) Occasions of Thunderstorm.

No Magnetic disturbance	1	20	21
Magnetic disturbance	1	18	19
Total	2	38	40

The χ^2 test was obtained as follows:—

TABLE 2. VALUES OF $P(\chi^2)$ AND $P(\chi_1^2)$.

	N	χ^2	C^1	$P(\chi)^2$	χ_1^2	$P(\chi_1^2)$	P
No Thunder storm	150	2·859	·137	·90	2·227	·136	·189
Thunder storm	40	0·005	·012	·942	0·427	·514	·780

The values of $P(\chi^2)$ are high and of the order of ninety per cent. or more, so that the above data do not point to any significant association between magnetic disturbance and abnormal ionisation.

The problem may be now examined from the point of view of fundamental theorems of probability. If we write a four-fold table in the following form:—

	A	Not-A	Total
B	a	b	N-n
Not-B	c	d	n
Total	N-n'	n'	N

then the sucessive probabilities from d=o to d=n are known to be proportional to the terms of the hypergeometric series $F(-n, -n', N-n-n'+1, 1)$. Alternatively stated the probability corresponding to any term (a, b, c, d) is given by

$$\frac{(N-n)!\ n!\ (N-n')!\ n'!}{N!\ a!\ b!\ c!\ d!}$$

The probability of having no abnormal ionisation when there is magnetic disturbance is obtained by putting d=o, but keeping the marginal totals n=49, and n'=39 the same in Table 1. We then have a=62, b=39, c=49, and d=o. The value of P(o) is then easily calculated

$$P(o)=\frac{101!\ 49!\ 111!\ 39!}{150!\ 62!\ 39!\ 49!\ o!}=\cdot 00000027.$$

The probabilities of having 0, 1, 2, ... 39 abnormal ionisations when there is magnetic disturbance were calculated in the same way and are given in column 2 of Table 3, and the probabilities of having an assigned number of abnorml ionisations or more are given in column 3 of the same table. We notice that the probability of having 17 or more abnormal ionisations is 0.0696. If we add up the two tail regions we have P=0.1392. The value of P corresponding to $\chi^2=2.8582$ is 9.0908 which is clearly an underestimate.

This direct method of calculations though rigorous is very laborious. F Yates has however shown that a very good approximation is obtained "by computing the values of χ^2 for deviations half a unit less than the true deviations."[4] This is what Yates has called the "correction for continuity'.' Thus the corrected χ^2 is given by

$$\chi^2=\frac{(78.5\times 16.5—22.5\times 32.5)^2\times 150}{101\times 49\times 111\times 39}=2.2318$$

which gives $P(\chi_1^2)=0.1336$. This is in satisfactory agreement with the exact value of 0.1392.

The values of χ^2 and χ_1^2 corrected for continuity are shown in Table 2. The difference between the two values of P are considerable in the cases of both "thunder-storms" and "no-thunderstorms". The exact value of P obtained from the hyper-geometric series is shown in the last column of Table 2. In the case of "no-thunderstorms," the value of $P(\chi_1^2)=0.136$ agrees closely with the exact value of 0.139, but in the case of thunderstorms the value of $P(\chi_1^2)=.514$ against an exact value of 0.780.

From Table 1 (b), we calculate the following probability:—

		P	
Nubmer of ab-normal ionisation on magnetic days	17	.2192	
	18	.5115	.7808
	19	.2693	
		1.0000	

TABLE 3. PROBABILITY OF HAVING 0, 1, 2, ... ABNORMAL IONISATIONS WHEN THERE IS MAGNETIC STORM.

No. of Abnormal Ionisations	Probability of Ionisation	Probability of assigned Ionisation or more	No. of Abnormal Ionisations	Probability of Ionisation	Probability of assigned Ionisation or more
0	0·0000		12	·1519	·6858
1	·0000		13	·1557	·5339
2	·0000		14	·1370	·3782
3	·0000		15	·1038	·2412
4	·0002	1·0000	16	·0678	·1374
5	·0010	·9998	17	·0383	·0696
6	·0038	·9988	18	·0187	·0313
7	·0113	·9950	19	·0079	·0126
8	·0271	·9837	20	·0036	·0047
9	·0540	·9566	21	·0009	·0011
10	·0900	·9026	22	·0002	·0002
11	·1268	·8126	23	·0000	·0000

Hence, the probability of having 18 or more abnormal ionisations on magnetically disturbed days is .7807 and the question of adding the other tail region does not arise. The values of $P(\chi^2)=.9420$, and $P(\chi_1^2)=0.5138$; and both are equally unreliable. This shows the inefficiency of χ^2 and χ_1^2 in regard to contingency tables with very small frequencies and skew distributions. In the present case, there is however no difficulty in interpreting the results, for the values of P are too high to indicate any significant association between abnormal ionisation and magnetic disturbance.

THUNDERSTORMS AND IONISATION.

Investigations by Appleton, Naismith and others[5] have shown the existence of a correlation between troposphere disturbance and ionisation in upper air. Bhar and Syam's observations may be arranged in the form of 2 × 2 tables as follows:

TABLE 4. FREQUENCY DISTRIBUTION OF IONISATION VALUES.

(a) Days without magnetic disturbance.

	Normal	Abnormal	Total
NoThunderstorm	79	22	101
Thunderstorm	1	20	21
Total	80	42	122

(b) Days with magnetic disturbance.

	Normal	Abnormal	Total
NoThunderstorm	32	17	49
Thunderstorm	1	18	19
Total	33	35	68

The results of the $P(\chi^2)$ test of independence between thunderstorm and abnormal ionisation, is shown in Table 5.

TABLE 5. VALUES OF P AND χ^2 AND OF P AND χ_1^2.

	χ^2	$P(\chi^2)$	χ_1^2	$P(\chi_1^2)$
No Magnetic disturbance	41·26	$<10^{-9}$	38·10	$<10^{-9}$
Magnetic disturbance	19·28	$<10^{-5}$	17·36	$<10^{-5}$

The $P(\chi^2)$ values, which are also shown in Table 3, are exceedingly small, so that thunderstorms and abnormal ionisation cannot be considered to be independent. The occurrence of thunderstorms and the existence of abnormal ionisation are thus significantly associated with each other, and this is true of days both with and without magnetic disturbance.

THUNDERSTORMS AND MAGNETIC DISTURBANCE.

Finally, we may test the association between the occurrence of a thunderstorm and a magnetic disturbance. The contingency table is shown in Table 6.

TABLE 6. FREQUENCY DISTRIBUTION OF IONISATION VALUES.

(a) Normal	No Thunderstorm	Thunderstorm	Total
Magnetic Disturbance	82	1	33
No Magnetic Disturbance	79	1	80
Total	111	2	118

(b) Abnormal	No Thunderstorm	Thunderstorm	Total
Magnetic Disturbance	17	18	85
No Magnetic Disturbance	22	20	42
Total	39	38	77

The value of $P(\chi^2)$ and $P(\chi_1^2)$ together with exact values of P are shown in Table 7.

TABLE 7. VALUES OF P AND χ^2 AND OF P AND χ_1^2.

	χ^2	$P(\chi^2)$	χ_1^2	$P(\chi_1^2)$	P
Normal Ionisation	0·426	0·516	·017	0·895	[·5006]
Abnormal Ionisation	·110	0·874	·011	0·920	·9171

The high values of P indicate that thunderstorms and magnetic disturbance are independent so far as the present observations are concerned.

CONCLUSIONS.

In testing independence with 2 × 2 tables, if the frequencies are small, the $P(\chi^2)$ test, in general, gives underestimated probabilities. If the distribution is not very skew, Yates' correction for continuity gives good approximation. If it is skew, it is always desirable to use the exact test expressed in terms of the hypergeometric series. The application of this exact test to Bhar and Syam's observations shows that a significant association exists between the occurrence of thunderstorms and abnormal ionisation in the upper air; but there is no appreciable connexion between magnetic disturbances and abnormal ionisation or between magnetic disturbances and thunderstorms.

REFERENCES.

(1) BHAR, J. N. and SYAM, P.: Effect of Thunderstorm and Magnetic Storms on the Ionisation of the Kennelly-Heaviside Layer. *Phil. Mag.* Vol. 23, April, 1937. No. 155, p. 513-528.

(2) PEARSON, K.: On the Theory of Contingency and its Relation to Association and Normal Correlation. *Draper's Company Research Memoirs, Biometric Series,* I, pp. 1-55 (1904).

(3) FISHER, R. A.: *Statistical Methods for Research Workers,* 6th edition, p. 97.

(4) YATES, F.: Contingency Tables involving Small Numbers and the χ^2 test. *Supp. Jour. Roy. Stat. Soc.,* Vol. I (1), 1934, pp. 217-235.

(5) APPLETON, E. V., and NAISMITH, R.: Weekly Measurement of Upper Atmospheric Ionisation. *Proc. Roy. Soc., A,* Vol. 150 (1936), p. 685.

November, 1937
Statistical Laoratory, Calcutta.