19

ON THE INTERPRETATION OF χ^2 FROM CONTINGENCY TABLES, AND THE CALCULATION OF P

Author's Note (CMS 5.86a)

This short paper, with all its juvenile inadequacies, yet did something to break the ice. Any reader who feels exasperated by its tentative and piecemeal character should remember that it had to find its way to publication past critics who, in the first place, could not believe that Pearson's work stood in need of correction, and who, if this had to be admitted, were sure that they themselves had corrected it.

The writer's point of view was that certain inconsistencies were manifest in Pearson's uses of the χ^2 test, and that certain other writers, notably Yule and Greenwood (1915) and Bowley (1920), had felt that something was wrong, without discovering what it was.

In the writer's thought, though not very explicitly in this paper, the mathematical distribution given by tables of χ^2 was that of the sum of n squares of variates normally and independently distributed about zero with unit variance. χ^2 in fact was the square of the distance of a random point from the centre of a homogeneous normal distribution in n dimensions. The number of dimensions, however, would be reduced by unity for every restriction upon deviations between expectation and observation, and it appeared that the inconsistencies in the literature could be straightened out if account were taken of the true number of degrees of freedom in which observation and expectation might in reality differ.

In the examples chosen the complete equivalence of different rational approaches, once the question of degrees of freedom is rectified, was still somewhat obscured by minor inconsistencies in the method of computation. These arise from the fact that the mathematical distribution is only realised in the limit for large samples, where non-linear restrictions tend to linearity.

 Journal of the Royal Statistical Society, 85: 87-94, (1922).

The treatment, in a footnote, of the multinomial distribution as a section of a multiple Poisson distribution of independent variates shows the most direct demonstration of the mathematical distribution tabled, as the sampling distribution of χ^2 defined as a measure of discrepancy between observation and hypothesis. This analytical artifice has, I hope, a lasting value.

The algebraic re-examination of Pearson's (1900) proof that the fitting of constants did *not* affect the distribution of χ^2 was set out in 1924 (Paper 34), in which the effects of inefficient fitting also are worked out.

On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P.

By R. A. Fisher, M.A., Rothamsted Experiment Station.

It is well known that the Pearsonian test of goodness of fit depends upon the calculation of the quantity χ^2 so defined that if m is the number of observations *expected* in any cell, and $m + x$ the number observed, then

$$\chi^2 = S\left(\frac{x^2}{m}\right),$$

the summation being extended to all the cells.

Pearson has shown (1) that *when the deviations are distributed with the sole restriction that their sum shall be zero*, the distribution of χ^2 is given by the Pearsonian curve of Type III,

$$df \propto \chi^{n'-2} e^{-\frac{1}{2}\chi^2} d\chi,$$

where n' is the number of cells.

We are not here concerned to criticise the general adequacy of the χ^2 test, which is certainly valid if the number of observations in each cell is large, but to emphasize the importance of the limitation italicized above. For the χ^2 test has been applied by Pearson and others to contingency tables, in which the sum of the deviations in any row or column is necessarily zero.

In these cases we shall show that Elderton's Tables of Goodness of Fit (2) may still be applied, but that the value of n' with which the table should be entered is not now equal to the number of cells, but to *one more than the number of degrees of freedom in the distribution.* Thus for a contingency table of r rows and c columns we should take $n' = (c - 1)(r - 1) + 1$ instead of $n' = cr$. This modification often makes a very great difference to the probability (P) that a given value of χ^2 should have been obtained by chance.

The most general way of proving this result consists in regarding the values of x (above) as independent co-ordinates in generalised space; then owing to the linear relations by which the deviations are restricted, for example that the marginal totals of the population should be equal to those observed, all possible sets of observations will lie relative to the centre of the distribution, specified by the assumed population, in a plane space, of the same number of dimensions as there are degrees of freedom. The frequency density at any point in this space is proportional to

$$e^{-\frac{1}{2}S\left(\frac{x^2}{\sigma^2}\right)}$$

when the sample is sufficiently great for the distribution of x to be regarded as normal, and where $\sigma_1, \sigma_2, \ldots$ represent the standard deviations of $x_1, x_2, \ldots$. To determine what values have to be assigned to the σ's when the x's are entirely independent, we must take account of the variation in the total number,

$$S(m + x) = N.$$

Since the different values of x are independent,

$$\sigma_N^2 = S(\sigma^2).$$

The variation of x may be regarded as due to two independent causes, namely the variation of N, and the variation of the proportion, which falls into any one compartment; we have therefore the series of equations,

$$\left.\begin{aligned} \sigma_1^2 &= p_1 q_1 \bar{N} + p_1^2 \sigma_N^2 \\ \sigma_2^2 &= p_2 q_2 \bar{N} + p_2^2 \sigma_N^2 \end{aligned}\right\} \qquad \text{(I)}$$

and so on, where p_1 is the chance of any observation falling in the cell (1).

Summing these, we find

$$\sigma_N^2 = \bar{N}\, S(pq) + \sigma_N^2\, S(p^2),$$

whence, since $S(p) = 1$ and $p - p^2 = pq$,

$$\sigma_N^2 = \bar{N}.$$

Substituting in (I),

$$\sigma_1^2 = (p_1q_1 + p_1^2)\,\overline{\mathrm{N}} = p_1\overline{\mathrm{N}} = m_1{}^*$$
$$\sigma_2^2 = (p_2q_2 + p_2^2)\,\overline{\mathrm{N}} = p_2\overline{\mathrm{N}} = m_2$$

and so on.

Whence

$$\mathrm{S}\left(\frac{x^2}{\sigma^2}\right) = \mathrm{S}\left(\frac{x^2}{m}\right) = \chi^2,$$

and the frequency density at any point in the generalised space is

$$e^{-\frac{1}{2}\chi^2}.$$

The surfaces of equal density are therefore the series of similar and coaxial ellipsoids, χ = constant; and since χ measures the linear dimensions of the corresponding ellipsoid, which by a homogeneous strain passes into a sphere, and since the plane space in

† * It is worth noting that the exact form of the distribution of N observations into a number of cells is given by the multinomial expansion,

$$\left(\frac{m_1}{\overline{\mathrm{N}}}k_1 + \frac{m_2}{\overline{\mathrm{N}}}k_2 + \quad . \quad . \quad . \quad .\right)^{\overline{\mathrm{N}}},$$

of which the coefficient of

$$k_1^{x_1}\ k_2^{x_2}\quad . \quad . \quad . \quad .$$

is the chance of the particular distribution,

$$x_1,\ x_2,\ .\ .\ .\ .$$

This may be regarded as a plane section of a distribution in which x_1, x_2, are *independently* distributed according to the Poisson series,

$$e^{-m_1}\left(1, m_1, \quad . \quad . \quad . \quad . \quad . \quad , \frac{m_1^{x_1}}{x_1\,!}, \quad . \quad . \quad . \quad . \quad .\right),$$

for in this case, N = S (x), will be distributed according to the series,

$$e^{-\overline{\mathrm{N}}}\left(1, \overline{\mathrm{N}}, \quad . \quad . \quad . \quad . \quad , \frac{\overline{\mathrm{N}}^{\mathrm{N}}}{\mathrm{N}\,!}, \quad . \quad . \quad . \quad . \quad . \quad \right)$$

and the chance of a given distribution, subject to the restriction $\mathrm{N} = \overline{\mathrm{N}}$ will be

$$\frac{\overline{\mathrm{N}}\,!\,e^{\overline{\mathrm{N}}}}{\overline{\mathrm{N}}^{\overline{\mathrm{N}}}}\left(\frac{m_1^{x_1}}{e^{m_1}x_1\,!}\right)\left(\frac{m_2^{x_2}}{e^{m_2}x_2\,!}\right)\quad . \quad . \quad \quad . \quad ,$$

which, since S $(m) = \overline{\mathrm{N}}$, reduces to

$$\frac{\overline{\mathrm{N}}\,!}{x_1\,!\,x_2\,!}\,.\ .\ .\ .\left(\frac{m_1}{\overline{\mathrm{N}}}\right)^{x_1}\left(\frac{m_2}{\overline{\mathrm{N}}}\right)^{x_2}\quad . \quad . \quad . \quad . \quad . \quad ,$$

the general term of the multinomial expansion.

This general case, however, in which the values of x may be small integers, extends beyond the range in which χ^2 may be considered a sufficient test of goodness of fit.

† For revised footnote, see page 345.

which the observations lie passes through the point, $\chi = 0$, the total frequency in the range of $d\chi$ must be proportional to

$$\chi^{n'-2} e^{-\frac{1}{2}\chi^2} d\chi,$$

where n' is one more than the number of the degrees of freedom.

Ex. 1.—In the fourfold table,

a	b	$a + b$
c	d	$c + d$
$a + c$	$b + d$	$a + b + c + d$

when the marginal totals are fixed, there remains only one degree of freedom. Consequently we must take n' equal to 2 and not to 4. We are thus led to perceive that χ is distributed so that,

$$df = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}\chi^2} d\chi.$$

This fact resolves a difficulty which has been felt with respect to the fourfold table. In 1915 Greenwood and Yule (3), using four fold tables to test the effect of inoculation against typhoid and cholera, follow Pearson in applying Elderton's table with $n' = 4$. They notice, however, that if we calculate the *proportion attacked* among the inoculated and among the uninoculated, thus

$$p = \frac{a}{a+b}, \quad p' = \frac{c}{c+d},$$

then the difference $p'-p$, compared to its probable error, should also give a test of independence; they find, in practice, that deviations which judged by the χ^2 test are not improbable, seem much less likely to occur when judged by the proportions attacked. While pointing out the difficulty, these authors judge it safer to apply the χ^2 test.

When we recognise that we should take $n' = 2$, the difficulty disappears, for the standard error of p is

$$\sqrt{\frac{(a+c)(b+d)}{(a+b+c+d)^2 (a+b)}},$$

and that of p' is

$$\sqrt{\frac{(a+c)(b+d)}{(a+b+c+d)^2 (c+d)}},$$

so that if

$$x = p' - p = \frac{bc - ad}{(a+b)(c+d)},$$

then

$$\frac{x^2}{\sigma_x^2} = \frac{(bc-ad)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)} = \chi^2,$$

and χ, for $n' = 2$, is, as we have shown above, distributed over the positive half of a normal curve, with unit standard deviation.

The two tests are, therefore, in reality identical when the test is rightly applied.

Dr. Bowley (5, 1921) has avoided this inconsistency by distinguishing the use of χ^2 in contingency tables from its use in testing goodness of fit. For the fourfold table he shows that if

$$x = \frac{ad - bc}{a+b+c+d},$$

then also

$$\frac{x^2}{\sigma_x^2} = \chi^2,$$

and consequently, χ being normally distributed, he uses the table of the probability integral. Thus three different tests of significant association in the fourfold table all lead to the same value of P, and this is what we should expect, since there is but one degree of freedom in the fourfold table, when the marginal totals are fixed.

It should be pointed out that certain of Pearson's "Tables for Statisticians and Biometricians," namely, Tables XVII, XIX and XX, together with XXII (*Abac* to determine r_P'), are all calculated on the assumption that $n' = 4$ in fourfold tables, and consequently should not be used when, as is almost always the case, the marginal proportions are obtained from the data.*

* I am indebted to Dr. Greenwood for pointing out to me that Pearson (6) has recognised that in some cases the value of n', with which Elderton's tables should be entered, ought to be reduced when linear restrictions are placed upon the observations. It would appear, however, that Pearson at that date drew a distinction between "linear relations imposed on the cell contents" and the restrictions which are introduced by our methods of reconstructing the hypothetical population from which the sample is regarded as drawn. Thus we find in Section 1 (p. 145) the introductory explanation, "Actually we find in the "sample M the number $m_{uv} \ldots \psi$, and the problem arises whether the system "represented by $m_{uv} \ldots \psi$ is so improbable that in the selected population M "the characteristics A, B, C . . . L, cannot be considered independent, "*i.e.* M is really not a random sample of the supposed population N. Clearly "the answer to this problem has already been given. We have to find the "value of χ^2" (stated in full notation for l variates), "and apply the tables

Ex. 2.—A further verification is possible in the case of the table with two rows and s columns,

f_1	f_2	. . . .	f_s	N
f_1'	f_2'	. . . .	f_s'	N′

" for 'goodness of fit.' Of course, in many cases the sampled population is " not known, and accordingly we can only put for " the marginal totals " the " values given by the sample itself, and test from this substitution the degree " of divergence from independence."

From this passage, and from the fact that throughout the paper no correction is suggested of the methods previously employed, and embodied in the *Tables for Statisticians* published only the year before, it is clear that Pearson did not recognise that in all cases linear restrictions imposed upon the frequencies of the sampled population, by our methods of reconstructing that population, have exactly the same effect upon the distributions of χ^2 as have restrictions placed upon the cell contents of the sample.

That the true distribution of χ^2 for the fourfold table was not recognised at this date may be inferred also from the fact that the criterion for differential death-rates, obtained *as an approximation* by very indirect methods, and applied correctly in a subsequent paper (7), namely :—

$$Q^2 = S\left\{\frac{a a'\left(\frac{d}{a}-\frac{d'}{a'}\right)^2}{(d+d')\left(1-\frac{d+d'}{a+a'}\right)}\right\}$$

the summation being taken over all age groups, when a, a', d, d' on the numbers exposed to risk and the numbers dying, in the two districts, follows at once from the fourfold table :—

	District A.	District B.	Total.
Surviving	$a - d$	$a' - d'$	$a + a' - d - d'$
Dying	d	d'	$d + d'$
Exposed to risk	a	a'	$a + a'$

for which

$$\chi^2 = \frac{a a'\left(\frac{d}{a}-\frac{d'}{a'}\right)^2}{(d+d')\left(1-\frac{d+d'}{a+a'}\right)}.$$

We thus obtain independent values of χ from the several age groups, and since χ for a fourfold table is normally distributed, the distribution of

$$Q^2 = S(\chi^2)$$

for u age groups must be exactly given by that of χ^2 in Elderton's tables, when $n' = u + 1$.

Treating this as a contingency table,

$$\chi^2 = S\left\{\frac{\left(f-\frac{f+f'}{N+N'}N\right)^2}{\frac{f+f'}{N+N'}N}+\frac{\left(f'-\frac{f+f'}{N+N'}N'\right)^2}{\frac{f+f'}{N+N'}N'}\right\}$$

the summation taken over all the columns.

Simplifying, we obtain,

$$\chi^2 = S\left\{\frac{NN'}{f+f'}\left(\frac{f}{N}-\frac{f'}{N'}\right)^2\right\},$$

while the number of degrees of freedom is $s-1$, so that we must enter Elderton's tables with $n' = s$.

Now Pearson (4) has developed a special test to be applied when we wish to know if two independent distributions are likely to be random samples from the same population. He arrives at the value of χ^2 obtained above by reducing the table to a simple series of s cells; so that this special method is in reality exactly the same as the direct application of χ^2 to the table, save that we take n' equal to s, and not to $2s$. This latter discrepancy is not, however, discussed in (4), or in the later paper (6), and the correct application of χ^2 to contingency tables of two or more variates has never been made clear.

Summary.

The χ^2 test may be applied to contingency tables, provided we take *not the number of cells* but *one more than the number of degrees of freedom* for n'.

So modified, the χ^2 test includes as special cases—

(i) the comparison of ratios in the fourfold table;

(ii) Pearson's method of comparison of distributed samples;

(iii) Pearson and Tocher's criterion of differential death rates.

The proof which we have given of the distribution of χ^2 is applicable, not only to contingency tables, but to all cases in which the frequencies observed are connected with those expected by a number of linear relations, beyond their restriction to the same total frequency. In taking the goodness of fit of a frequency curve fitted by means of four moments, the number of degrees of freedom has been reduced by 4, and since the four moments are linear functions of the class frequencies, we should take n' to be 4 less than the number of cells. In this case it should be noted that it is usual, and convenient, to calculate the moments from a finer graduation than that which we use in testing goodness of fit, and in consequence;

restricted plane region in which the observations lie will not pass exactly through the point $\chi = 0$; the distribution of χ, calculated from 4 less than the number of cells, will none the less be closely accurate even in these cases, and far more accurate than that obtained by putting n' equal to the number of cells.

In all cases, therefore, of applying the χ^2 test, it is necessary to take account of the number of degrees of freedom of the observations in relation to the expected distribution, to which they are compared; in cases where all the restrictions are of a linear character the correct distribution of χ may be found from Elderton's tables, or, if $n' = 2$, from a table of the probability integral, while in the case of restrictions of a non-linear character, Elderton's tables are no longer exactly applicable.

(1) K. Pearson (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag.*, l, 157.

(2) K. Pearson (1914). *Tables for Statisticians and Biometricians.* Cambridge University Press.

(3) G. U. Yule and M. Greenwood (1915). The statistics of antityphoid and anticholera inoculations, and the interpretation of such statistics in general. *Proc. Roy. Soc. of Medicine*, Section of Epidemiology and State Medicine, viii, 113.

(4) K. Pearson (1911). On the probability that two independent distributions are really samples of the same population. *Biometrika*, viii, 250.

(5) A. L. Bowley (1920). *Elements of Statistics*, p. 371. London: P. S. King & Sons.

(6) K. Pearson (1915). On the theories of multiple and partial contingency. *Biometrika*, xi, 145.

(7) K. Pearson and J. F. Tocher (1915). On criteria for the existence of differential death-rates. *Biometrika*, xi, 159.

Author's revised footnote (CMS 5.89a) - see page 339.

*It is worth noting that the exact form of the distribution of N observations into a number of cells is given by the multinomial expansion

$$\left(\frac{m_1}{M}k_1 + \frac{m_2}{M}k_2 + \cdots + \frac{m_s}{M}k_s\right)^N$$

of which the coefficient of

$$k_1^{a_1}k_2^{a_2}\cdots k_s^{a_s}$$

is the chance of the particular distribution

$$a_1, \quad a_2, \quad \cdots, \quad a_s.$$

This may be regarded as a plane section of a distribution in which $a_1, \cdots, a_s$ are *independently* distributed according to the Poisson series

$$e^{-m}\frac{m^a}{a!}$$

for all a, while the total,

$$\Sigma(a) = N,$$

will be distributed in the Poisson series,

$$e^{-M}\frac{M^N}{N!},$$

in which

$$\Sigma(m) = M.$$

Hence the chance of the given series of observed frequencies, subject to the restriction that the total number shall be N, will be

$$\frac{N!}{a_1!a_2!\cdots a_s!}\left(\frac{m_1}{M}\right)^{a_1}\left(\frac{m_2}{M}\right)^{a_2}\cdots\left(\frac{m_s}{M}\right)^{a_s},$$

the general term of the multinomial expansion.

This general case, however, in which the values of a may be small integers, extends beyond the range in which χ^2 may be considered an adequate test of goodness of fit.

Author's Addendum, 1947. As the sample is increased without limit, so are all the expectations m, and each Poisson distribution tends to normality with mean m and variance m. Hence

$$\sum\frac{(a-m)^2}{m}$$

is distributed in the limit as is the sum of the squares of s normal deviates distributed independently each with unit variance. The section of this distribution made by imposing the condition $M = N$ is of $s - 1$ dimensions, and further linear restrictions will each reduce the number of dimensions by unity.