DEPARTMENT OF APPLIED MATHEMATICS,
UNIVERSITY COLLEGE, UNIVERSITY OF LONDON.

DRAPERS' COMPANY RESEARCH MEMOIRS.

BIOMETRIC SERIES, I.

MATHEMATICAL CONTRIBUTIONS TO THE THEORY OF EVOLUTION.

XIII. ON THE THEORY OF CONTINGENCY AND ITS RELATION TO ASSOCIATION AND NORMAL CORRELATION.

BY

KARL PEARSON, F.R.S.

[WITH TWO DIAGRAMS.]

LONDON:
PUBLISHED BY DULAU AND CO., 37 SOHO SQUARE, W.
1904.

Price Four Shillings.

DEPARTMENT OF APPLIED MATHEMATICS,
UNIVERSITY COLLEGE, UNIVERSITY OF LONDON.

DRAPERS' COMPANY RESEARCH MEMOIRS.

BIOMETRIC SERIES, I.

MATHEMATICAL CONTRIBUTIONS TO THE THEORY OF EVOLUTION.

XIII. ON THE THEORY OF CONTINGENCY AND ITS RELATION TO ASSOCIATION AND NORMAL CORRELATION.

BY

KARL PEARSON, F.R.S.

[WITH TWO DIAGRAMS.]

LONDON:
PUBLISHED BY DULAU AND CO., 37 SOHO SQUARE, W.
1904.

M.W

Price Four Shillings.

In March, 1903, *the Worshipful Company of Drapers announced their intention of granting* £1,000 *to the University of London to be devoted to the furtherance of research and higher work at University College. After consultation between the University and College authorities, the Drapers' Company presented* £1,000 *to the University to assist the statistical work and higher teaching of the Department of Applied Mathematics. It seemed desirable to commemorate this—probably, first occasion on which a great City Company has directly endowed higher research work in mathematical science—by the issue of a special series of memoirs in the preparation of which the Department has been largely assisted by the grant. Such is the aim of the present series of "Drapers' Company Research Memoirs."*

K. P.

On Contingency and its Relation to Association and Normal Correlation.

By KARL PEARSON, *F.R.S.*

CONTENTS.

		Page
(1.)	Introductory. On classification and scales	1
(2.)	On the conception of contingency	5
(3.)	On the relation between mean square contingency and normal correlation	7
(4.)	On the relation between mean contingency and normal correlation	9
(5.)	The two contingency coefficients	15
(6.)	Influence of overfine grouping	16
(7.)	Probable errors and error correlations of sub-contingencies	17
(8.)	Change of order of grouping unimportant in the general case of linear regression	19
(9.)	Relation of contingency to association	21
(10.)	Theory of multiple contingency	22
(11.)	Illustrations:—	27
	A.—Inheritance of stature in man	27
	B.—Colour inheritance in greyhounds	30
	C.—Hair colours in man	31
	D.—Occupational or professional contingency in relatives	32
(12.)	General conclusions	34

(1.) *Introduction.*

IN dealing with the problem of the relationship of attributes, not capable of quantitative measurement, it has been usual to classify the two attributes into a number of groups, $A_1, A_2, A_3, \ldots A_s$ and $B_1, B_2, B_3, \ldots B_t$. In this manner a table has been formed containing s columns and t rows, or $s \times t$ compartments. The total frequency of the population, or of the "universe" under consideration, to use the logician's phrase, is then distributed into sub-groups corresponding to these $s \times t$ compartments. In simple cases of association, as in that of the presence of the vaccination cicatrix and the recovery from an attack of smallpox, s and t are both equal to two, and we have a simple four-fold division of the universe. In other cases we have higher numbers, as when we classify the human eye into eight colour classes and correlate these classes with six or more classes for hair colour. We may even run up to as many as 18 to 25 classes for each attribute when we table the coat colours of thoroughbred horses or pedigree dogs in the case of pairs of blood relatives.

Hitherto, in order to obtain a measure of the degree of correlation or association, we have proceeded on the assumption that it was necessary to arrange the system of classes like $A_1, A_2, \ldots A_s$ in some order, which corresponded to a real quantitative scale in the attribute, although we were unable to use this scale directly. Thus one arranged eye-colours in what appeared to correspond to a scale of varying amounts of orange pigment; the coat colours of horses were arranged in an order corresponding fairly to what an artist would call their "value." I even analysed hair tints by photographic processes. In all such cases the order seemed of vital importance. Once this order was settled, the methods of my memoir* on the correlation of characters not quantitatively measurable could be applied—the actual scale corresponding to the classification could be deduced, and we were able, on the assumption of normal frequency, to actually plot the regression lines for the correlation of a variety of attributes.† The conception, however, of order in the classification was at times very hampering. Take three broad classes like those for human temper—*quick tempered, good natured*, and *sullen*; it is difficult to grasp the exact meaning of a quantitative scale at the basis of this classification, and it is not obvious that the right order is necessarily that with good-natured in the middle. Or, again, take the case of human hair; omitting the brown reds, we can get a practically continuous series of shades from jet black to flaxen, and from flaxen with increasing red up to the deepest reds. Only the brown reds come in and upset the system! We seem, therefore, forced to take a double scale, first one of black, and then one of red pigment. Or, again, take the coat colour of greyhounds; these are classified into as many as 40 fairly narrow groups, and we can arrange these groups in ascending order of red, or black, or other pigmentation. We have more than one possible scale.

Now in recent work on such things as temper in man, eye colour in man, and hair colour in man or other animals, I have proceeded to arrange my groups in two or three different orders, and to calculate the correlation on the basis of these different orders. The results for the different orders came out in rather striking agreement, and the first sort of conclusion that one was tempted to draw was, for example, that the inheritance of pigmentation was strikingly alike for all pigments. But the agreement was in some cases far closer than one is accustomed to find when one compares the inheritance of directly measurable characters, and I soon became convinced that owing to some important theoretical law hitherto overlooked, the order of the groups by which we classify our attributes is a matter of no importance when we are determining correlation. The group order is all important for variation, it has practically no influence on correlation. We may put sullen tempers where we please in regard to quick and good-natured; we may place the shades of red hair at either end of the hair scale or in the middle, and the inheritance coefficient will come

* 'Phil. Trans.,' A, vol. 195, pp. 1–47.

† For example, for health and ability and for the correlation of the psychical and physical characters, see the "Fourth Annual Huxley Lecture," 'Journal of the Anthropological Institute,' vol. 33, pp. 194–195.

out nearly the same in value. Nay, we may go further, and classify finger prints like Mr. GALTON into "tents," "arches," "whorls," "croziers," &c., &c., and still be able to find a numerical value of the degree of resemblance between two blood relatives,. although any arrangements of such groups into a possible quantitative scale may be inconceivable. The object of this present paper is to deal with this novel conception of what I have termed *contingency*, and to see its relation to our older notions of association and normal correlation. The great value of the idea of contingency for economic, social, and biometric statistics seems to me to lie in the fact that it frees us from the need of determining scales before classifying our attributes. I shall endeavour to illustrate the importance of this freedom in the illustrations which follow the theoretical treatment of the subject.

(2.) *On the Conception of Contingency.*

In mathematical treatises on algebra a definition is usually given of independent probability. If p be the probability of any event, and q the probability of a second event, then the two events are said to be independent, if the probability of the combined event be $p \times q$. Now let A be any attribute or character and let it be classified into the groups $A_1, A_2, \ldots A_s$, and let the total number of individuals examined be N, and let the numbers which fall into these groups be $n_1, n_2, \ldots n_s$ respectively. Then the probability of an individual falling into one or other of these groups is given by n_1/N, $n_2/N, \ldots n_s/N$ respectively. Now suppose the same population to be classified by any other attribute into the groups $B_1, B_2, \ldots B_t$, and the group frequencies of the N individuals to be $m_1, m_2, \ldots m_t$ respectively. The probability of an individual falling into these groups will be respectively m_1/N, m_2/N, $m_3/N, \ldots m_t/N$. Accordingly the number of combinations of B_v with A_u to be expected on the theory of independent probability if N pairs of attributes are examined is

$$N \times \frac{n_u}{N} \times \frac{m_v}{N} = \frac{n_u \cdot m_v}{N} = \nu_{uv}, \text{ say.}$$

Let the number actually observed be n_{uv}. Then, allowing for the errors of random sampling,

$$n_{uv} - \frac{n_u m_v}{N} = n_{uv} - \nu_{uv}$$

is the deviation from independent probability in the occurrence of the groups A_u, B_v. Clearly the total deviation of the whole classification system from independent probability must be some function of the $n_{uv} - \nu_{uv}$ quantities for the whole table. I term any measure of the total deviation of the classification from independent probability a measure of its *contingency*. Clearly the greater the contingency, the greater must be the amount of association or of correlation between the two attributes, for such association or correlation is solely a measure from another standpoint of the degree of deviation from independence of occurrence.

Now it must be quite clear that if we make our measurement of contingency any function whatever of such quantities as $n_{uv} - \nu_{uv}$, its magnitude will be absolutely independent of the order of classification, *i.e.*, its value will be unchanged if we re-arrange the A's and the B's in any manner whatever. This is the fundamental gain of this new conception of contingency. But precisely as we can measure position or acceleration in a great variety of ways, so it is possible to measure contingency. We must try to select out of these ways those which: (a) bring contingency into line with the customary notions of correlation and association; and (b) permit of not too laborious calculations leading to the required measure.

We will consider these points at some length. I have shown in a paper,* "On Deviations from the Probable in a Correlated System of Variables," that if m'_1, $m'_2, \ldots m'_n$ be any system of observed frequencies and $m_1 . m_2, \ldots m_n$ be any system of theoretical frequencies known *à priori*, then if

$$\chi^2 = \text{Sum}\left\{\frac{(m'_q - m_q)^2}{m_q}\right\} \text{ from } q = 0 \text{ to } n$$

be calculated, we can deduce a quantity P from χ^2 which is the probability that in any trial a system $m''_1, m''_2, \ldots m''_n$ of observed frequencies will occur, which deviates more from $m_1, m_2, \ldots m_n$ than the actually observed system does. Tables have been worked out by Mr. PALIN ELDERTON giving the value of P, for a considerable range of values of χ^2 and n, and have been published in 'Biometrika.'†

Now it will be obvious that if we want to measure contingency, we really want to measure the deviation of the observed results from independent probability, and therefore if we take $m_1, m_2, \ldots m_n$ to correspond to the system ν_{uv} and $m'_1, m'_2, \ldots m'_n$ to correspond to the actually observed system n_{uv},

$$\chi^2 = \text{S}\left\{\frac{(n_{uv} - \nu_{uv})^2}{\nu_{uv}}\right\} \quad \ldots \ldots \ldots \ldots \quad \text{(i.)},$$

will be a proper quantity to calculate, and P would measure how far the observed system is or is not compatible with a basis of independent probability. If P be large the chances are in favour of the system arising from independent probability; if P be small there is certainly association between the attributes. Hence $1 - \text{P}$ would be a proper measure of the contingency. I propose to call $1 - \text{P}$ the *contingency grade*. Further, it is convenient to have a name for a function closely related to χ^2. I shall call

$$\phi^2 = \chi^2/\text{N} \quad \ldots \ldots \ldots \ldots \quad \text{(ii.)}$$

the *mean square contingency*.

* 'Phil. Mag.,' July, 1900, pp. 157–175.

† Vol. I., p. 155.

It will be seen that, in the method by which we have approached the problem, we have not had to consider the question of the sign of the contingency like $n_{uv} - \nu_{uv}$, our mean square contingency is based on a summation of squares extending to all the $s \times t$ compartments of the table. But if we treat now of quantities like $n_{uv} - \nu_{uv}$ their total sum must be zero, since for the whole table

$$S(n_{uv}) = N = S(\nu_{uv}).$$

Let us suppose that the symbol Σ refers to a summation of all the *positive* contingencies, and let

$$\psi = \Sigma(n_{uv} - \nu_{uv})/N \quad \ldots\ldots\ldots\ldots \quad \text{(iii.)},$$

then ψ shall be spoken of as the *mean contingency*. Clearly any functions of either ϕ^2 or ψ would serve to measure the contingency. We shall be guided in our choice of such functions by considering what are the values of ϕ^2 and ψ in the case of normal correlation.

(3.) *On the Relation between Mean Square Contingency and Normal Correlation.*

Let x and y denote the deviations from their respective means of two characters or attributes, of which σ_x, σ_y are the standard deviations and r is the correlation. Then if we assume a normal distribution of frequency, $z_0\,\delta x\,\delta y$ would be the frequency of individual pairs falling between x and $x + \delta x$, y and $y + \delta y$, where

$$z_0 = \frac{N}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} \quad \ldots\ldots\ldots\ldots \quad \text{(iv.)},$$

on the assumption of independent probability, and $z\,\delta x\,\delta y$, where

$$z = \frac{N}{2\pi\sqrt{1-r^2}\,\sigma_x\sigma_y} e^{-\frac{1}{2}\frac{1}{1-r^2}\left(\frac{x^2}{\sigma_x^2}-\frac{2rxy}{\sigma_x\sigma_y}+\frac{y^2}{\sigma_y^2}\right)} \quad \ldots\ldots\ldots \quad \text{(v.)},$$

on the assumption of contingent probability.

We then have at once

$$\phi^2 = S\left\{\frac{(z\,\delta x\,\delta y - z_0\,\delta x\,\delta y)^2}{Nz_0\,\delta x\,\delta y}\right\} = S\left\{\frac{(z-z_0)^2}{Nz_0}\,\delta x\,\delta y\right\},$$

and we have only to insert the values of z and z_0, given by (iv.) and (v.), and integrate all over the plane of x, y, to find the mean square contingency.

Now, if $ac > b^2$, we know that

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(ax^2-2bxy+cy^2)}\,dx\,dy = \frac{1}{\sqrt{ac-b^2}} \quad \ldots\ldots \quad \text{(vi.)}.$$

This is all we need, for if $x = \sigma_x x'$, $y = \sigma_y y'$:

$$\phi^2 = \frac{1}{\mathrm{N}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(\frac{z^2}{z_0} - 2z + z_0 \right) dx'\, dy'$$

$$= \frac{1}{2\pi} \left\{ \frac{1}{1 - r^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left\{x'^2 \frac{1+r^2}{1-r^2} - \frac{4rx'y'}{1-r^2} + y'^2 \frac{1+r^2}{1-r^2}\right\}} dx'\, dy' \right.$$

$$- \frac{2}{\sqrt{1 - r^2}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left\{x'^2 \frac{1}{1-r^2} - \frac{2rx'y'}{1-r^2} + y'^2 \frac{1}{1-r^2}\right\}} dx'\, dy'$$

$$\left. + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x'^2 + y'^2)} dx'\, dy' \right\}$$

$$= \frac{1}{1 - r^2} \frac{1}{\sqrt{\left(\frac{1 + r^2}{1 - r^2}\right)^2 - \frac{4r^2}{(1 - r^2)^2}}} - \frac{2}{\sqrt{1 - r^2}} \frac{1}{\sqrt{\frac{1}{(1 - r^2)^2} - \frac{r^2}{(1 - r^2)^2}}} + 1 \quad . \quad \text{(vii.)},$$

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

Thus the mean square contingency is simply $r^2/(1-r^2)$. Or,

$$r = \pm \sqrt{\frac{\phi^2}{1 + \phi^2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(viii.)}.$$

Thus the relationship between mean square contingency and correlation in the case of normal frequency is of an extremely simple character.

We see at once :—

(i.) That since the mean square contingency is absolutely independent of the arrangement of our classes, the coefficient of correlation is also entirely independent of the arrangement of our classes on the basis of any assumed order or scale.

(ii.) Provided our classes are sufficiently small to allow of us legitimately replacing by groupings over small areas the theoretical integrations, the coefficient of correlation can be found from the mean square contingency.

We have thus an entirely new method of finding correlation in the case of quantitatively non-measurable characters. It assumes, however, that our classification-groups are sufficiently numerous and their contents sufficiently small to justify us in supposing that the contingency has reached a definite limit. Clearly in working in the future by the contingency method, we shall have to adopt rather more numerous classes, and they should not contain too irregular proportions of individuals, but we can then afford to drop any question of scale or order of grouping.

It may be asked whether this method of deriving the correlation from the contingency cannot replace the earlier method of deducing the correlation by the fourfold division of the material. The answer is that in some cases it can do so very advantageously, but it is very far from doing so in all. The contingency found from a fourfold table is a perfectly real and very proper measure of the deviation of its material from independent probability. But if this mean square contingency be substituted in equation (viii.), it will not give us the correlation. The proper mean square contingency to give us the correlation must be based on a sufficiently *large* number of classes. When, however, we take, say, 20 classes for each attribute, we have 400 terms to deal with in calculating ϕ^2, and although the result might then possibly give a more accurate value for the correlation than that found from a fourfold division, yet the labour of determining it is far greater and may be excessive. Further, the simple classification into two or three groups may be all we are able to make at all, or all we can conveniently make. Hence the new conception of contingency, while illuminating the whole subject—especially as demonstrating that the correlation is independent of scale or grouping, does not do away with the older method of the fourfold division. I propose to call the expression

$$\sqrt{\frac{\phi^2}{1+\phi^2}},$$

the *first coefficient of contingency.*

We note that with small enough classes the coefficient of contingency becomes the coefficient of correlation. Accordingly, with a view of lessening the number of coefficients in use, I adopt the following convention: Any expression or function of either the mean square contingency (ϕ^2) or the mean contingency (ψ) (or indeed of any other measure of the contignency), which, when the grouping is sufficiently small, is theoretically equal to the coefficient of correlation—on the hypothesis of normal frequency—shall be termed a coefficient of contingency. All such coefficients of contingency must, on the same hypothesis, become equal on a sufficiently small grouping, and they will scarcely differ widely from each other when the frequency is not absolutely normal and the grouping is merely moderately small. These points will be illustrated later.

(4.) *On the Relation of Mean Contingency to Normal Correlation.*

A great deal of the labour of finding either the coefficient of contingency or the coefficient of correlation by the method of mean square contingency when the groups are numerous, depends upon the squaring of the contingencies and dividing by the frequency to be expected on the basis of independent probabilities. The whole of this labour is escaped, if we work with the mean contingency instead of the mean square contingency; further, since in this case we only sum for the positive contingencies, neglecting the negative, we have usually to deal with only, or often less than, a moiety of the terms involved in calculating ϕ^2. On the other hand, there is no simple relation between the correlation and the mean contingency such as we have found between correlation and mean square contingency in equation (viii.) above.

The relation is far more complex and is only expressible in the form of integrals reducible by quadratures. Still, for practical purposes we rarely want the coefficient of contingency to more than two decimal places. Hence, if the integral be evaluated for the coefficient proceeding by equal intervals, we can plot a curve giving the value of the coefficient of contingency in terms of the mean contingency, and this will be sufficiently accurate to enable us to read off the former in terms of the latter to the required degree of accuracy. The enquiry also brings out some other points not without interest.*

To investigate the curve which in a normal correlation surface separates on the plane of xy areas of positive from areas of negative contingency.

The frequency due to independent probability will be equal to that due to the actual contingent probability when

$$\frac{\mathrm{N}}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} = \frac{\mathrm{N}}{2\pi\sigma_x\sigma_y}\frac{1}{\sqrt{1-r^2}} e^{-\frac{1}{2}\frac{1}{1-r^2}\left(\frac{x^2}{\sigma_x^2}-\frac{2rxy}{\sigma_x\sigma_y}+\frac{y^2}{\sigma_y^2}\right)},$$

where r is the coefficient of correlation, or of contingency.

Clearly

$$(1-r^2)\log_e(1-r^2) = -r^2\left\{\frac{x^2}{\sigma_x^2}-\frac{2xy}{r\sigma_x\sigma_y}+\frac{y^2}{\sigma_y^2}\right\} \quad . \quad . \quad . \quad . \quad (\text{ix.}).$$

Since r is always less than unity, this curve is clearly a hyperbola, which possesses several interesting properties. We see at once that all the contingency of one sense is grouped into the space between the two branches of this hyperbola, and that the contingency of the other sense is grouped into the two separate spaces inside the two branches. Thus contingency of either sense is for normal correlation *continuous*, and abrupt changes of sign in the contingency—beyond the limits of random sampling—are not to be expected.

By testing on actual correlation tables I find this hyperbola comes out in a fairly marked manner, in fact, quite as significantly as the elliptic contours of equal frequency.

I propose to consider the properties of this zero contingency hyperbola—it forms the curve along which two really contingent events have a frequency identical with their independent probability.

Consider the two families of curves:

$$\frac{x^2}{\sigma_x^2}-\frac{2rxy}{\sigma_x\sigma_y}+\frac{y^2}{\sigma_y^2}=\alpha \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{x.}),$$

$$\frac{x^2}{\sigma_x^2}-\frac{2}{r}\frac{xy}{\sigma_x\sigma_y}+\frac{y^2}{\sigma_y^2}=\beta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{xi.}).$$

* I have to heartily thank my assistant, Dr. L. N. G. FILON, for the substance of the first part of the investigation given below, down to equation (xiii.). I owe the calculation and plotting of the curves $u = e^{-\kappa \sec\theta}$ to my assistant, Mr. J. C. M. GARNETT.

Since r is always < 1, the α family form a set of concentric, similar, and similarly-situated ellipses, and the β family a set of concentric, similar, and similarly-situated hyperbolas. Any conic having double contact with the hyperbola β_0, of zero contingency defined by (ix.), at the ends of a diameter $y = mx$, has for its equation

$$\frac{x^2}{\sigma_x^2} - \frac{2}{r}\frac{xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2} + \lambda(y - mx)^2 = \beta_0.$$

If this be identical with an ellipse α, we have, by comparing coefficients and eliminating λ and m,

$$\beta_0^2/\alpha^2 = 1/r^2.$$

Consequently $\alpha = \pm r\beta_0$, the sign being determined from the fact that α must always be positive for real ellipses.

Now the ordinate z of the normal frequency surface is given by

$$z = \frac{N}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\, e^{-\frac{1}{2(1-r^2)}\left\{\frac{x^2}{\sigma_x^2} - \frac{2rxy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2}\right\}},$$

$$= \frac{N}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\, e^{-\frac{\alpha}{2(1-r^2)}},$$

and to find the mean contingency we must determine the whole volume lying inside the *two* branches of the above hyperbola, integrating on *both* sides of the line of contact of the families of hyperbolas and ellipses.*

We have $\iint \frac{z}{N}\, dx\, dy$ over this area

$$= I_r = \frac{4}{2\pi}\frac{1}{\sigma_x\sigma_y}\frac{1}{\sqrt{1-r^2}}\int_{r\beta_0}^{\infty} d\alpha \int_{\beta_0}^{\alpha/r} \frac{e^{-\frac{\alpha}{2(1-r^2)}}}{J}\, d\beta,$$

where

$$J = \frac{\delta(\alpha, \beta)}{\delta(x, y)} = -\frac{4(1-r^2)}{r\sigma_x\sigma_y}\left(\frac{x^2}{\sigma_x^2} - \frac{y^2}{\sigma_y^2}\right)$$

from (x.) and (xi.).

But from (x.) and (xi.)

$$\left\{\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)^2 - \frac{4x^2y^2}{\sigma_x^2\sigma_y^2}\right\}(1-r^2)^2 = (\alpha^2 - \beta^2 r^2)(1-r^2).$$

Or, choosing the signs to make J positive, we have

$$J = \frac{4\sqrt{1-r^2}}{r\sigma_x\sigma_y}\sqrt{\alpha^2 - \beta^2 r^2}.$$

* The ellipses and hyperbolas have common pairs of conjugate diameters; one line of contact is one of the asymptotes of the hyperbola $\frac{x^2}{\sigma_x^2} - \frac{y^2}{\sigma_y^2} = 1$; and tangents at an intersection point of any of the family of ellipses with any of the family of hyperbolas are respectively parallel to conjugate diameters of this hyperbola. These geometrical properties, however, need not detain us here.

Thus the required integral is

$$I_r = \frac{2r}{4\pi(1-r^2)}\int_{r\beta_0}^{\infty} d\alpha \int_{\beta_0}^{\alpha/r} d\beta \frac{e^{-\frac{\alpha}{2(1-r^2)}}}{\sqrt{\alpha^2-\beta^2 r^2}},$$

$$= \frac{1}{2\pi(1-r^2)}\int_{r\beta_0}^{\infty} \cos^{-1}\frac{\beta_0 r}{\alpha}\, e^{-\frac{\alpha}{2(1-r^2)}}\, d\alpha.$$

To simplify put, using (ix.),

$$\alpha = \beta_0 r \sec\theta, \quad k = \frac{\beta_0 r}{2(1-r^2)} = -\frac{1}{2r}\log_e(1-r^2) \quad . \quad . \quad . \quad . \quad \text{(xii.)},$$

where k will always be positive, since $r < 1$.

We have

$$I_r = \frac{k}{\pi}\int_0^{\frac{\pi}{2}} e^{-k\sec\theta}\,\theta\sec\theta\tan\theta\, d\theta,$$

or, integrating by parts,

$$I_r = \frac{1}{\pi}\int_0^{\pi/2} e^{-k\sec\theta}\, d\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xiii.)}.$$

The curves $u = e^{-k\sec\theta}$ were then plotted with our coordinatograph for a series of values of k or r on a large scale, drawn in with a spline and integrated with a Coradi compensating planimeter. The values of I_r resulting are tabled on p. 15.

We have next to investigate what is the volume NQ_r of the surface of independent probability

$$z_0 = \frac{N}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)},$$

which falls within the same hyperbola of contingency. We shall then have in $Q_r - I_r$ the required value of ψ, the mean contingency on the basis of normal correlation. We have

$$Q_r = \frac{1}{2\pi\sigma_x\sigma_y}\iint e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} dx\, dy$$

taken over the space inside the two branches of the hyperbola

$$\frac{x^2}{\sigma_x^2} - \frac{2xy}{r\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2} = \beta_0.$$

Write $x = x'\sigma_x$, $y = y'\sigma_y$, and we have

$$x'^2 - 2\frac{x'y'}{r} + y'^2 = \beta_0.$$

Transform to polars, $\rho\cos\theta = x'$, $\rho\sin\theta = y'$,

$$\rho^2 = \frac{r\beta_0}{r - \sin 2\theta}.$$

This shows us that the axes are given by $\theta = \frac{\pi}{4}$ and $\frac{\pi}{2} + \frac{\pi}{4}$, or are a and b, where

$$a^2 = r\beta_0/(1 + r), \quad b^2 = r\beta_0/(1 - r).$$

Take these axes as axes of coordinates. Then we have to integrate

$$Q_r = \frac{1}{\pi}\iint e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy,$$

over the area inside *one* branch of the hyperbola

$$(1 + r)\,x^2 - (1 - r)\,y^2 = r\beta_0 \quad \text{(xiv.)}.$$

Let

$$\left.\begin{aligned} x^2 + y^2 &= \alpha, \\ x^2 - y^2 + r\,(x^2 + y^2) &= r\beta \end{aligned}\right\} \quad \text{(xv.)},$$

and let us transfer the integrations to α and β.

We have

$$x^2 = \tfrac{1}{2}\{\alpha - r\,(\alpha - \beta)\},$$
$$y^2 = \tfrac{1}{2}\{\alpha + r\,(\alpha - \beta)\},$$

and

$$Q_r = \frac{2}{\pi}\iint \frac{e^{-\frac{1}{2}\alpha}}{J}\,d\alpha\,d\beta,$$

over one-half one branch of the hyperbola.

$$J = \frac{d\alpha}{dy}\frac{d\beta}{dx} - \frac{d\alpha}{dx}\frac{d\beta}{dy} = \frac{8yx}{r} = \frac{4}{r}\sqrt{\alpha^2 - r^2(\alpha - \beta)^2}.$$

Thus we have

$$Q_r = \frac{r}{2\pi}\int_{\frac{r\beta_0}{1+r}}^{\infty} d\alpha e^{-\frac{1}{2}\alpha}\int_{\beta_0}^{\frac{1+r}{r}\alpha} \frac{d\beta}{\sqrt{\alpha^2 - r^2(\alpha - \beta)^2}} \quad \text{(xvi.)}.$$

The limits are obtained from the consideration, easily seen on a figure, that for a given α we must integrate from $\beta = \beta_0$, the given hyperbola, to $\beta = \frac{1+r}{r}\alpha$, the *touching* hyperbola; and then for α we must take every circle from that touching β_0, *i.e.*, $\alpha = r\beta_0/(1 + r)$ up to infinity.

We will first integrate with regard to β, and put

$$r\,(\alpha - \beta) = -\,\alpha \sin\phi.$$

This gives, when $\beta = (1 + r)\,\alpha/r$, $\phi = \frac{1}{2}\pi$; and when

$$\beta = \beta_0, \quad \phi = \sin^{-1}\frac{r\,(\beta_0 - \alpha)}{\alpha} = \phi_0.$$

Thus we find

$$Q_r = \frac{1}{2\pi}\int_{\frac{r\beta_0}{1+r}}^{\infty} d\alpha e^{-\frac{1}{2}\alpha}\int_{\phi_0}^{\frac{1}{2}\pi} d\phi = \frac{1}{2\pi}\int_{\frac{r\beta_0}{1+r}}^{\infty} \cos^{-1}\frac{r(\beta_0-\alpha)}{\alpha} e^{-\frac{1}{2}\alpha} d\alpha \quad . \quad . \quad \text{(xvii.)}.$$

Take

$$\cos\chi = r(\beta_0 - \alpha)/\alpha,$$

then

$$\alpha = \infty, \qquad \cos\chi = -r,$$

$$\alpha = r\beta_0/(1+r), \quad \cos\chi = 1.$$

Hence

$$Q_r = \frac{1}{2\pi}\int_0^{\cos^{-1}(-r)} \chi e^{-\frac{1}{2}\frac{r\beta_0}{r+\cos\chi}} \frac{r\beta_0 \sin\chi}{(r+\cos\chi)^2} d\chi,$$

$$= \frac{1}{\pi}\int_0^{\cos^{-1}(-r)} e^{-\frac{1}{2}\frac{r\beta_0}{r+\cos\chi}} d\chi,$$

observing that the term between the limits vanishes at both.

Take

$$\cos\theta = (r+\cos\chi)/(r+1).$$

Then

$$\chi = 0, \qquad \theta = 0,$$

$$\chi = \cos^{-1}(-r), \quad \theta = \tfrac{1}{2}\pi.$$

Thus we find finally, after some reductions,

$$Q_r = \frac{1}{\pi}\int_0^{\frac{1}{2}\pi} e^{-\kappa\sec\theta}\sqrt{\frac{1+\cos\theta}{\epsilon+\cos\theta}}\, d\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xviii.)},$$

where

$$\left.\begin{aligned} \epsilon &= (1-r)/(1+r), \\ \kappa &= \tfrac{1}{2}\frac{r\beta_0}{1+r} = -\tfrac{1}{2}\frac{1-r}{r}\log_e(1-r^2) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \text{(xix.)},$$

$= (1-r)\,k$, of the integral I_r.

Tables were now formed of ϵ and κ and the ordinates of the curves

$$v = e^{-\kappa\sec\theta}\sqrt{\frac{1+\cos\theta}{\epsilon+\cos\theta}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(xx.)}$$

calculated.* These ordinates were plotted on a large scale by aid of a Coradi coordinatograph and the resulting curves integrated as before, the values of Q_r thus found are given with the values of I_r and ψ in the table below. I believe this table gives the mean contingency in terms of the correlation true to at least three places of decimals. The u and v curves are both interesting analytically and subject to rather curious changes of type. We were aided in plotting them by calculating, where

* I owe the calculation of these ordinates to Dr. ALICE LEE.

needful, $\frac{du}{d\theta}$ and $\frac{dv}{d\theta}$. Finally, the values of r were plotted by my demonstrator, Mr. L. W. ATCHERLEY, to the corresponding values of ψ. Thus a curve was obtained, which enables us to read off the correlation from the contingency correct to at least two places of decimals—sufficient for nearly all practical purposes.

TABLE I.—Table of Integrals I_r, Q_r, and the Contingency ψ for Values of r.

r.	I_r.	Q_r.	ψ.
0·00	·5000	·5000	·0000
·05	·4620	·4762	·0142
·10	·4342	·4652	·0310
·20	·3895	·4536	·0641
·30	·3501	·4498	·0996
·40	·3162	·4547	·1385
·50	·2830	·4643	·1813
·60	·2489	·4814	·2325
·70	·2128	·5106	·2978
·80	·1700	·5524	·3824
·90	·1186	·6279	·5093
·95	·0796	·7009	·6213
1·00	·0000	1·0000	1·0000

Diagram I. at the end of this memoir will therefore serve for most purposes of interpolation, and it will be seen that now that the integrals have been evaluated and the diagram constructed, the correlation can be very easily found from mean contingency. But the method seems to me distinctly inferior to that of mean square contingency, and this for much the same reasons that mean error calculations are inferior to mean square error work in curve fitting. Further, the grade of contingency can be found at once from a knowledge of mean square contingency, and whatever be the distribution is a significant and interpretable constant. This is only true of the correlation deduced from mean contingency if the distribution be normal.

(5.) To sum up our results so far :—

We have, if

n_{uv} be the actual frequency of a group in the population, N which combines the characters A_u and B_v, ν_{uv} be the frequency of this group on the hypothesis of independent probability, then

$n_{uv} - \nu_{uv}$ is simply a sub-contingency,

$S\left\{\frac{(n_{uv}-\nu_{uv})^2}{\nu_{uv}}\right\} = \chi^2$ may be termed the square contingency,

$S\left\{\frac{(n_{uv}-\nu_{uv})^2}{N\nu_{uv}}\right\} = \phi^2$ is the mean square contingency,

$\Sigma\left(\frac{n_{uv}-\nu_{uv}}{N}\right) = \psi$, where Σ is the sum for positive (or negative) sub-contingencies only, is the mean contingency.

Any one of these expressions is a measure of the deviation of the system from independent probability, and therefore of the amount of association or correlation between the characters or attributes involved. But any function of these expressions is also a proper measure. Such functions are :—

(*a.*) The contingency grade. This is $1 - P$, where P is to be found from χ^2 by aid of the tables for "goodness of fit." See 'Biometrika,' vol. 1, pp. 155, *et seq.*

(*b.*) The mean square contingency coefficient $= C_1$, where

$$C_1 = \sqrt{\frac{\phi^2}{1 + \phi^2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{xxi.}).$$

(*c.*) The mean contingency coefficient $= C_2$, where C_2 is to be found from the table on p. 15 or from Diagram I. at the end of this memoir.

In the case of sufficiently small grouping and normal correlation we have

$$C_1 = C_2 = \text{coefficient of correlation.}$$

But it must not be forgotten that this is essentially a limiting, not a general case. Nevertheless the approach to equality of the two contingency coefficients will be a good measure of the normality of the distribution and the suitability as to smallness of our elements of grouping.

(6.) A little experience of actual working, however, shows that in practice it is perfectly easy to overshoot the mark in fineness of grouping. Suppose that in dealing with 1000 cattle we find a single instance of a calf inscribed as "mulberry," say the offspring of a red cow by a dark fawn bull. Now if there be 30 dark fawn bulls, the independent probability of a dark fawn bull having a mulberry offspring is ·03. Hence the sub-contingency for a ♂ parent-offspring table $= 1 - {\cdot}03 = {\cdot}97$, and the corresponding contribution to the square contingency will be $({\cdot}97)^2/03$, or is upwards of 31. The fact is, that when we come to very fine groupings we get at once into difficulties owing to our having to record by *units* only. Suppose "mulberry" calves actually had no relation to any special parentage, but were rare anomalies occurring once among 1000 calves, or perhaps were merely an odd breeder's fancy description, then a unit cannot be divided in the proportions of the colour parentage, it must fall into some one colour parentage group. The result is that a few isolated individuals will give large contributions to the mean square contingency. The above example is purely hypothetical, but similar cases have actually occurred in dealing with colour problems by the contingency method. They are exactly similar to those which occur when dealing with outlying individuals by the test for "goodness of fit." In a frequency distribution we proceed only by units, but the theory gives fractional values of the frequency; hence in forming the value of χ^2 to measure goodness of fit, one or two unit "outliers," although not improbable as far as the *whole* of the tail of a curve is concerned, may be exceedingly improbable if considered from the standpoint of the actual group in which they do occur. This point must be carefully borne in mind in actual practice, for by sufficient refinement of grouping, *i.e.*, till we reduce certain groups to a single individual or two, the mean square contingency can be increased in a remarkable manner.

(7.) Of course this is merely saying that the probable errors of the sub-contingencies increase largely when we make ν_{uv} very small. Unfortunately I have not yet succeeded in determining the probable errors of the contingency coefficients. If c_{uv} be the contingency, determined by

$$c_{uv} = n_{uv} - \frac{n_u n_v}{N},$$

and $\Sigma_{c_{uv}}$ its standard deviation for random sampling, I find

$$\Sigma_{c_{uv}} = n_{uv}\left(1 - \frac{n_{uv}}{N}\right) + \frac{n_u n_v}{N^2}\left(n_u + n_v - \frac{4n_u n_v}{N}\right) - 2\frac{n_{uv}}{N}\left(n_u + n_v - \frac{3n_u n_v}{N}\right). \quad \text{(xxii.)},$$

so that the probable error of any individual contingency $= \cdot 67449\, \Sigma_{c_{uv}}$ is determined.

Further, if $R_{c_{uv}c_{u'v'}}$ be the correlation between errors due to random sampling in two contingencies c_{uv} and $c_{u'v'}$, *not* belonging to either the same row or column,

$$\Sigma_{c_{uv}}\Sigma_{c_{u'v'}}R_{c_{uv}c_{u'v'}} = -\frac{n_{uv}n_{u'v'}}{N} + 2\frac{n_{uv}n_{u'}n_{v'} + n_{u'v'}n_u n_v}{N^2}$$
$$+ \frac{n_{uv'}n_v n_{u'} + n_{u'v}n_{v'}n_u}{N^2} - 4\frac{n_u n_{u'} n_v n_{v'}}{N^3} \quad \text{(xxiii.)}.$$

Similarly we find for the correlation of errors of two contingencies of the same column, $R_{c_{uv}c_{uv'}}$, the result

$$\Sigma_{c_{uv}}\Sigma_{c_{uv'}}R_{c_{uv}}R_{c_{uv'}} = -\frac{n_{uv}n_{uv'}}{N} - \frac{n_v n_{uv'} + n_{v'}n_{uv}}{N}\left(1 - \frac{3n_u}{N}\right)$$
$$+ \frac{n_u n_v n_{v'}}{N^2}\left(1 - \frac{4n_u}{N}\right) \quad \text{(xxiv.)},$$

and for errors of two contingencies of the same row,

$$\Sigma_{c_{uv}}\Sigma_{c_{u'v}}R_{c_{uv}c_{u'v}} = -\frac{n_{uv}n_{u'v}}{N} - \frac{n_u n_{u'v} + n_{u'}n_{uv}}{N}\left(1 - \frac{3n_v}{N}\right)$$
$$+ \frac{n_v n_u n_{u'}}{N^2}\left(1 - \frac{4n_v}{N}\right). \quad \text{(xxv.)}.$$

Results (xxii.) to (xxv.) enable us to find the probable errors and the error correlations for any individual contingencies which will arise from random sampling, and are so far of value; but when we attempt to find the general expression for the probable error of either the mean or mean square contingency, it becomes so complex that there appears little hope of deducing a simple result. Arithmetically the problem might be solved at the expense of rather troublesome numerical calculations if the number of sub-groups was not very large. A general and simple expression for the probable error of ψ or ϕ^2 involving ψ or ϕ^2 only does not appear likely to exist, and an expression involving all the sub-group frequencies would be very troublesome for computation. Practically the errors of the contingency coefficients may be fairly reasonably taken to lie between the probable errors of r as found by a fourfold division of a table and by the product method, approaching the latter more closely as the number of sub-groups is sufficiently increased. With the experience of probable errors of fourfold tables before us we may, I think, safely take the probable error of a contingency coefficient C for rough judgments to be less than

$$2 \times \cdot 67449 \frac{1 - C^2}{\sqrt{n}},$$

i.e., double the probable error of a correlation coefficient found from the product moment. At the same time we must distinctly be cautious, remembering the difficulty as to isolated units referred to in the previous section.

We may look at the probable error of the contingency from another standpoint. Taking the mean squared contingency, we have

$$1 + \phi^2 = \frac{1}{1 - r^2}.$$

Therefore

$$\delta\phi^2 = \frac{2r}{(1 - r^2)^2} \delta r,$$

and accordingly, if Σ_{ϕ^2}, Σ_r be the standard deviations in errors of ϕ^2 and r,

$$\Sigma_{\phi^2} = \frac{2r}{1 - r^2} \Sigma_r = \frac{2r}{(1 - r^2)^2} \frac{1 - r^2}{\sqrt{N}} *$$

$$= \frac{2}{\sqrt{N}} \frac{r}{1 - r^2} = \frac{2}{\sqrt{N}} \sqrt{\phi^2 (1 + \phi^2)}.$$

Hence if we were to determine ϕ^2 from r, the probable error of ϕ^2 would be given by

$$\text{Probable error of } \phi^2 = \cdot 67449 \frac{2}{\sqrt{N}} \sqrt{(1 + \phi^2) \phi^2}.$$

Or, we can put it into the more useful form,

$$\text{Percentage probable error of } \phi^2 = \frac{1 \cdot 34898}{\sqrt{N}} \sqrt{\frac{1 + \phi^2}{\phi^2}} \quad . \quad . \quad (\text{xxvi.}).$$

Thus the percentage probable error increases rapidly as the contingency gets smaller.

* 'Phil. Trans.,' A, vol. 191, p. 242.

Of course, the probable error of ϕ^2 as found from r is not necessarily the same as the probable error of ϕ^2 found directly, but it may serve as a guide to its approximate value.

If it were the same, the probable error of r as found from ϕ^2 would be $\cdot 67449/\{(1+\phi^2)\sqrt{\mathrm{N}}\}$, a result, as indicated in the previous paragraph, much too small, except possibly for very successful systems of grouping.

(8.) To find under what other condition than normal correlation small changes in the order of grouping will not affect the value of the correlation.

Let us assume the unit of grouping to be very small, but not necessarily the same for all groups. Let the two characters or attributes be x and y, and suppose n_s to be the total frequency of individuals in the range $y_s - \epsilon$ to $y_s + \epsilon$, and n_{s+1} to be the total frequency in the range $y_{s+1} - \epsilon'$ to $y_{s+1} + \epsilon'$. Let $y_{s+1} - y_s = \epsilon + \epsilon' = h$ be so small that its square may be neglected. Let $\bar{x}$, $\bar{y}$ be the mean values of the characters, N the total frequency. We will find the changes in the moments and constants supposing the array n_s and n_{s+1} interchanged in position.

Clearly $\delta\bar{x} = 0$ and $\delta\sigma_x = 0$.

$$\mathrm{N}\,(\bar{y} + \delta\bar{y}) = \mathrm{S}\,(y_s n_s) + h\,(n_s - n_{s+1}),$$

or,

$$\delta\bar{y} = h\,(n_s - n_{s+1})/\mathrm{N}.$$

$$\mathrm{N}\,(\sigma_y + \delta\sigma_y)^2 = \mathrm{S}\,(y_s^2 n_s) + 2h\,(y_s n_s - y_{s+1} n_{s+1}) - \mathrm{N}\,(\bar{y} + \delta\bar{y})^2,*$$

$$2\sigma_y\,\delta\sigma_y = 2h\,(y_s n_s - y_{s+1} n_{s+1}) - 2\mathrm{N}\bar{y}\,\delta\bar{y},$$

$$\frac{\delta\sigma_y}{\sigma_y} = \frac{h}{\sigma_y^2}\,\frac{(y_s - \bar{y})\,n_s - (y_{s+1} - \bar{y})\,n_{s+1}}{\mathrm{N}}.$$

Next if

$$\mathrm{P} = \mathrm{S}\,(xy) - \mathrm{N}\bar{y}\,\bar{x},$$

$$\mathrm{P} + \delta\mathrm{P} = \mathrm{S}\,(xy) + h\,(n_s\bar{x}_s - n_{s+1}\bar{x}_{s+1}) - \mathrm{N}\bar{y}\,\bar{x} - \mathrm{N}\bar{x}\,\delta\bar{y},$$

or,

$$\delta\mathrm{P} = h\,\{n_s\,(\bar{x}_s - \bar{x}) - n_{s+1}\,(\bar{x}_{s+1} - \bar{x})\},$$

where $\bar{x}_s$ and $\bar{x}_{s+1}$ are the means of the arrays n_s and n_{s+1}.

But if r be the correlation coefficient of x and y characters,

$$r = \frac{\mathrm{P}}{\mathrm{N}\sigma_x\sigma_y}.$$

Therefore

$$\frac{\delta r}{r} = \frac{\delta\mathrm{P}}{\mathrm{P}} - \frac{\delta\sigma_x}{\sigma_x} - \frac{\delta\sigma_y}{\sigma_y},$$

* It must be noted here that the squares of the change in $\bar{y}$ and σ_y are neglected. Hence the changes must not be so great that $\delta\bar{y}$ and $\delta\sigma_y$ are sensibly as compared with $\bar{y}$ and σ_y.

and substituting the above values,

$$\frac{\delta r}{r} = h\left\{\frac{n_s(\bar{x}_s - \bar{x}) - n_{s+1}(\bar{x}_{s+1} - \bar{x})}{\mathrm{P}} - \frac{(y_s - \bar{y})\,n_s - (y_{s+1} - \bar{y})\,n_{s+1}}{\mathrm{N}\sigma_y^2}\right\}.$$

If this is to vanish for any value of s and h, it will be sufficient, since

$$\mathrm{P} = r \times \mathrm{N}\sigma_x\sigma_y,$$

$$\bar{x}_s - \bar{x} = \frac{r\sigma_x}{\sigma_y}(y_s - \bar{y}),$$

and

$$\bar{x}_{s+1} - \bar{x} = \frac{r\sigma_x}{\sigma_y}(y_{s+1} - \bar{y}).$$

Or, if the mean $\bar{x}_m$ of any y_m-array of individuals be determined by

$$\bar{x}_m - \bar{x} = \frac{r\sigma_x}{\sigma_y}(y_m - \bar{y}).$$

But this is the condition for linear regression.

Hence we conclude that in any correlated system of variables, obeying the law of linear regression, we can, without sensibly modifying the correlation, interchange two adjacent y-arrays (*e.g.*, two rows of the correlation table), provided the grouping be fine. But if we can interchange any two adjacent y-arrays, we can, by a repetition of such changes, interchange any two y-arrays whatever; and a precisely similar statement must be valid for any two x-arrays (*e.g.*, two columns of the correlation table). Hence, given a sufficiently small system of grouping, we may state that in all cases of linear regression the actual order of the scales is immaterial as far as the determination of the correlation is concerned.

The practical importance of this result would appear to be great, for it frees us when dealing with scale orders from the need for supposing normal frequency; the indifference of the scale order when determining correlation is still true, provided the regression is linear; and this linearity of regression is not only found from observation to be very general—for example, in inheritance problems*—but follows from theory itself in the case of various hypotheses.†

In actual practice, of course, the degree of fineness of the grouping is limited by many considerations, and hence it will often be better to proceed by the fourfold division method, taking that division where possible at a very distinct classification. But the general principle now demonstrated will enable us in future to pay much less

* See "The Laws of Inheritance in Man.—I. Inheritance of the Physical Characters," 'Biometrika,' vol. 2, pp. 362–3; also "Inheritance of Mental and Moral Characters in Man," 'Huxley Memorial Lecture,' 1903. 'Journal of the Anthropological Institute,' vol. 33, pp. 185–7.

† "Contributions to the Theory of Evolution.—XII. On a Generalised Theory of Alternative Inheritance, with special reference to MENDEL's Laws." 'Phil. Trans.,' A, vol. 203, p. 85.

attention to the actual order chosen for the scales if we are dealing with a class of characters for which we may reasonably presume the regression to be sensibly linear.

(9.) If we take the crudest possible division of our material into only four groups, thus :—

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

corresponding to what Mr. YULE has termed the *association* of two attributes, we have at once

$$\psi = \frac{2(ab - cd)}{N^2} \quad \text{(xxvii.)},$$

$$\phi^2 = \frac{(ab - cd)^2}{(a + d)(c + b)(a + c)(d + b)} \quad \text{(xxviii.)}.$$

Now it is clear that in this case ϕ^2 reduces to r_{hk}^2, where r_{hk} is the correlation between errors in the position of the means of the two characters under consideration, as determined by a fourfold table, and $\frac{1}{2}\psi$ is in this simple case what I have defined as the transfer per unit of total frequency.* Both are expressions intimately connected with the conception of association, and have already been discussed in relation to it.† The coefficients, C_1 and C_2, of contingency—either of which might serve as a measure of the association—will not in this simple case, however, be necessarily even approximately equal to each other, still less to either the coefficient of correlation or Mr. YULE's coefficient of association.‡

It is worth while illustrating this on a numerical example. Taking the small-pox returns for the epidemic of 1890, we have :—

Cicatrix.	Recoveries.	Deaths.	Totals.
Present . . .	1562	42	1604
Absent . . .	383	94	477
Totals. . . .	1945	136	2081

* 'Phil. Trans.,' A, vol. 195, pp. 12 and 14.

† *Ibid.*, p. 15 *et seq.*

‡ 'Phil. Trans.,' A, vol. 194, p. 272.

These give us $\phi^2 = \cdot 0845$, $\chi^2 = 175 \cdot 76$, $\psi = \cdot 0604$. From these we find

$C_1 = \cdot 279$, $C_2 = \cdot 190$.
YULE's coefficient of association $= \cdot 803$.
Coefficient of correlation by fourfold division $= \cdot 595$.
Grade of contingency $= 1 - P$,* where $P = 718/10^{40}$.

Now so far as numerical values go these things are all totally different. C_1, C_2, and the coefficient of association depend very largely on where the fourfold division is taken.† It is extremely difficult to use them therefore for comparative purposes. On the other hand, the coefficient of correlation with the assumption, however, of normality is free of this restriction; it brings us into line with other things for comparative purposes. The grade of contingency is also independent in a sense of the division, *i.e.*, it has a definite physical meaning. What it tells us is this, that the deviation from independent probability in the relation between result, a case of small-pox and presence or absence of cicatrix is such that the above table could only arise 718 times in 10^{40} cases if the two events were absolutely independent.

If, instead of a table like the above, we take a number of alternative possibilities for each attribute, the coefficient of association loses its uniqueness of meaning; C_1 and C_2 still retain their significance, and as the number of alternatives become greater, merge in the coefficient of correlation. The grade of contingency, on the other hand, retains the same perfectly definite meaning throughout. I think this statement may serve as some warning of the caution needful in using the coefficients now introduced. The degree of approach of both C_1 and C_2 to the correlation must be studied for each special class of cases, and only when this has been done will their use be really legitimate and effective.

(10.) *On the Relation between Multiple Contingency and Multiple Normal Correlation.*

Suppose instead of a single correlation table we have a multiple correlation system. Such a system is well illustrated by the cabinet at Scotland Yard, which contains the measurements of habitual criminals on the old system of body measurements, now discarded in favour of a finger-print index. We have in this case a division of the cabinet into 3 compartments, which mark a threefold division of long, medium, and

* When the number of groups = 4, we have ('Phil. Mag.,' vol. 50, p. 157 *et seq.*) :—

$$P = \sqrt{\frac{2}{\pi}} \int_{\chi}^{\infty} e^{-\frac{1}{2}\chi^2} d\chi + \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}\chi^2} \chi.$$

$$= \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}\chi^2} \chi \left\{1 - \frac{1}{\chi^2} + \frac{1}{\chi^4} - \frac{3}{\chi^6} + \frac{15}{\chi^8}\right\},$$

whence P is easily found if χ^2 be large.

† YULE, *loc. cit.*, p. 276.

short head lengths. Each of these vertical divisions is then sub-divided horizontally into three divisions giving the corresponding divisions for head breadth; each of these head-breadth divisions has three drawers for large, moderate, and small face breadths. Each drawer is sub-divided into three sections for three finger groups, and these again into compartments for cubit groups, and so on. If this be carried out for the seven characters dealt with, we should have ultimately 3^7 sub-groups forming a multiple correlation system of the 7^{th} order.* We may ask what is the mean square contingency of such a system and to what extent does it diverge from an independent probability system? Of course, for an ideal anthropometric index system the divergence should be very slight.

Let $x_1, x_2 \ldots x_n$ be the n variables of a multiple normal correlation surface, to which the equation is

$$z = \frac{\text{N}}{(2\pi)^{\frac{1}{2}n}\sigma_1\sigma_2\ldots\sigma_n\sqrt{\text{R}}} \text{ expt. } -\tfrac{1}{2}\left\{\text{S}_1\left(\frac{\text{R}_{pp}}{\text{R}}\frac{x_p^2}{\sigma_p^2}\right) + 2\text{S}_2\left(\frac{\text{R}_{pq}}{\text{R}}\frac{x_p x_q}{\sigma_p\sigma_q}\right)\right\}.$$

Here $\sigma_1, \sigma_2 \ldots \sigma_n$ are the standard deviations of the n variables; S_1 denotes a sum of all values of p from 1 to n, S_2 a sum of all unlike values of p and q from 1 to n; while R is the determinant

$$\begin{vmatrix} 1, & r_{12}, & r_{13}, & \cdot & \cdot & \cdot & \cdot & r_{1n} \\ r_{21}, & 1, & r_{23}, & \cdot & \cdot & \cdot & \cdot & r_{2n} \\ r_{31}, & r_{32}, & 1, & \cdot & \cdot & \cdot & \cdot & r_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{n1}, & r_{n2}, & r_{n3}, & \cdot & \cdot & \cdot & \cdot & 1 \end{vmatrix}$$

and R_{st} is the minor corresponding to the constituent r_{st}, and the r's are the correlation coefficients.†

Now if ϕ^2 be the mean square contingency, we have

$$\phi^2 = \frac{1}{\text{N}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{(z-z_0)^2}{z_0}\,dx_1\,dx_2\ldots dx_n,$$

where z_0 = value of z when all the r's are zero, or

$$z_0 = \frac{\text{N}}{(2\pi)^{\frac{1}{2}n}\sigma_1\sigma_2\ldots\sigma_n} \text{ expt. } -\tfrac{1}{2}\left\{\text{S}_1\left(\frac{x_p^2}{\sigma_p^2}\right)\right\}.$$

Thus we have, writing $x_p = \sigma_p x'_p$, etc.,

$$\phi^2 = \frac{1}{(2\pi)^{\frac{1}{2}n}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\left(\frac{\zeta^2}{\zeta_0} - 2\zeta + \zeta_0\right)dx'_1\,dx'_2\ldots dx'_n,$$

* See MACDONELL, "On Criminal Anthropometry," 'Biometrika,' vol. 1, p. 205 *et seq.*

† 'Phil. Trans.,' A, vol. 187, p. 302, or *Ibid.*, A, vol. 200, pp. 3–8.

where

$$\zeta = \frac{1}{\sqrt{R}} \text{ expt.} - \tfrac{1}{2}\left\{S_1\left(\frac{R_{pp}}{R} x'^2_p\right) + 2S_2\left(\frac{R_{pq}}{R} x'_p x'_q\right)\right\},$$

$$\zeta_0 = \text{expt.} - \tfrac{1}{2}\{S_1(x'^2_p)\}.$$

Now

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(c_{11}x'^2_1 + c_{22}x'^2_2 + \ldots + 2c_{12}x'_1x'_2 + \ldots)}\, dx'_1\, dx'_2 \ldots dx'_n$$

$$= (2\pi)^{\frac{1}{2}n}/\sqrt{\Delta} \quad \ldots\ldots\ldots \quad \text{(xxix.)},$$

where

$$\Delta = \begin{vmatrix} c_{11}, & c_{12}, & c_{13}, & \ldots & c_{1n} \\ c_{21}, & c_{22}, & c_{23}, & \ldots & c_{2n} \\ c_{31}, & c_{32}, & c_{33}, & \ldots & c_{3n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ c_{n1}, & c_{n2}, & c_{n3}, & \ldots & c_{nn} \end{vmatrix}$$

We are now in a position to find all the integrals involved in the equation for ϕ^2, we have

$$\phi^2 = \frac{1}{R}\frac{1}{\sqrt{\Delta'}} - 2 + 1 = \frac{1}{R\sqrt{\Delta'}} - 1,$$

where

$$\Delta' = \begin{vmatrix} \frac{2R_{11}}{R} - 1, & \frac{2R_{12}}{R}, & \frac{2R_{13}}{R}, & \ldots & \frac{2R_{1n}}{R} \\ \frac{2R_{21}}{R}, & \frac{2R_{22}}{R} - 1, & \frac{2R_{23}}{R}, & \ldots & \frac{2R_{2n}}{R} \\ \frac{2R_{31}}{R}, & \frac{2R_{32}}{R}, & \frac{2R_{33}}{R} - 1, & \ldots & \frac{2R_{3n}}{R} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \frac{2R_{n1}}{R}, & \frac{2R_{n2}}{R}, & \frac{2R_{n3}}{R}, & \ldots & \frac{2R_{nn}}{R} - 1 \end{vmatrix}$$

To evaluate this determinant, we notice that since $r_{pp} = 1$, we have, if p and q be different,

$$R_{p1}r_{p1} + R_{p2}r_{p2} + R_{p3}r_{p3} + \ldots + R_{pn}r_{pn} = R,$$

$$R_{p1}r_{q1} + R_{p2}r_{q2} + R_{p3}r_{q3} + \ldots + R_{pn}r_{qn} = 0.$$

Hence

$$\frac{2R_{p1}}{R} r_{p1} + \frac{2R_{p2}}{R} r_{p2} + \frac{2R_{p3}}{R} r_{p3} + \ldots + \left(\frac{2R_{pp}}{R} - 1\right) r_{pp} + \ldots + \frac{2R_{pn}}{R} r_{pn} = 1$$

and

$$\frac{2R_{p1}}{R} r_{q1} + \frac{2R_{p2}}{R} r_{q2} + \frac{2R_{p3}}{R} r_{q3} + \ldots + \left(\frac{2R_{pp}}{R} - 1\right) r_{qp} + \ldots + \frac{2R_{pn}}{R} r_{qn} = -r_{qp}.$$

Now multiply the determinant Δ' by the determinant R, we find, using the above relations,

$$\Delta' R = \begin{vmatrix} 1, & -r_{12}, & -r_{13}, & \dots & -r_{1n} \\ -r_{21}, & 1, & -r_{23}, & \dots & -r_{2n} \\ -r_{31}, & -r_{32}, & 1, & \dots & -r_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ -r_{n1}, & -r_{n2}, & -r_{n3}, & \dots & 1, \end{vmatrix}$$

$$= R', \text{ say.}$$

Here R′ is R with the sign of all the correlations *changed.* Hence it follows that

$$\phi^2 = \frac{1}{\sqrt{RR'}} - 1 \quad \dots \dots \dots \quad \text{(xxx.)}.$$

Special Cases.

(i.) Simple correlation

$$R' = R = 1 - r_{12}^2, \quad \text{and} \quad \phi^2 = r_{12}^2/(1 - r_{12}^2), \text{ as before.}$$

(ii.) Triple correlation

$$R = 1 - r_{23}^2 - r_{31}^2 - r_{12}^2 + 2r_{23}r_{31}r_{12},$$

$$R' = 1 - r_{23}^2 - r_{31}^2 - r_{12}^2 - 2r_{23}r_{31}r_{12}.$$

$$\phi^2 = \frac{1}{\sqrt{(1 - r_{23}^2 - r_{31}^2 - r_{12}^2)^2 - 4r_{23}^2 r_{31}^2 r_{12}^2}} - 1.$$

(iii.) Quadruple correlation

$$RR' = \{1 - r_{12}^2 - r_{13}^2 - r_{14}^2 - r_{23}^2 - r_{24}^2 - r_{34}^2 + r_{12}^2 r_{34}^2 + r_{23}^2 r_{14}^2 + r_{13}^2 r_{24}^2 - 2(r_{12}r_{14}r_{23}r_{34} + r_{14}r_{13}r_{23}r_{24} + r_{12}r_{13}r_{24}r_{34})\}^2 - 4\{r_{23}r_{24}r_{34} + r_{34}r_{14}r_{13} + r_{12}r_{14}r_{24} + r_{12}r_{13}r_{23}\}^2,$$

and so on.

Clearly a condition has to be satisfied among the correlation coefficients, or the process by which we have deduced ϕ^2 is not legitimate. We must have Δ positive for equation (xxix.) to be true. Now, for *normal* correlation R must be real and positive, or the equation to the multiple correlation surfaces become imaginary. Hence it follows that Δ' must be positive, and therefore R′ must be positive. This seems to give a definite condition to be satisfied by the correlation coefficients, and in some cases rather narrow limits are enforced. For example, in the case of triple correlation we must have

$$1 - r_{23}^2 - r_{31}^2 - r_{12}^2 - 2r_{23}r_{31}r_{12}$$

positive, and this appears to reduce very considerably the possible values for the

TABLE II.—Stature of Father and Son.

Stature of Son.	Stature of Father.																	Totals.
	58·5–59·5.	59·5–60·5.	60·5–61·5.	61·5–62·5.	62·5–63·5.	63·5–64·5.	64·5–65·5.	65·5–66·5.	66·5–67·5.	67·5–68·5.	68·5–69·5.	69·5–70·5.	70·5–71·5.	71·5–72·5.	72·5–73·5.	73·5–74·5.	74·5–75·5.	
59·5–60·5	—	—	—	—	·50	·50	1	—	—	—	—	—	—	—	—	—	—	2·0
60·5–61·5	—	—	—	—	·50	—	—	—	1	—	—	—	—	—	—	—	—	1·5
61·5–62·5	—	·25	·25	—	·50	1	·25	·25	·50	·50	—	—	—	—	—	—	—	3·5
62·5–63·5	—	·25	·25	2·25	2·25	2	4	5	2·75	1·25	—	·25	·25	—	—	—	—	20·5
63·5–64·5	1	—	1·50	3·75	3	4·25	8	9·25	3	1·25	1·5	·75	1·25	—	—	—	—	38·5
64·5–65·5	2	*1	·50	2	3·25	9·50	13·50	10·75	7·50	5·50	3·50	2·50	—	—	—	—	—	61·5
65·5–66·5	—	·50	1	2·25	5·25	9·50	10	16·75	17·50	16	5·25	2	2·50	1	—	—	—	89·5
66·5–67·5	—	1·50	2	4·75	3·50	13·75	19·75	26·50	25·75	19·50	12·50	13·75	3·25	·50	1	—	—	148·0
67·5–68·5	—	—	1·50	2	7·50	10	10·25	24·25	31·50	23·50	29·50	13·25	8·50	9·50	2·25	—	—	173·5
68·5–69·5	—	—	1	—	5·25	5	12·75	18·25	16	24	29	21·50	10	3·50	2·25	—	1	149·5
69·5–70·5	—	—	—	—	1	2·50	5·75	18·75	11·75	19·50	22·50	19·50	14·50	6·25	3·50	1·50	1	128·0
70·5–71·5	—	—	—	—	—	3·25	5	8·75	10·75	19	14·75	20·75	10·75	8	5	1	1	108·0
71·5–72·5	—	—	—	—	—	·25	3	1·25	7	7·75	10·75	11·25	10	8·50	2·75	·50	—	63·0
72·5–73·5	—	—	—	—	—	—	·75	·75	2·50	7·50	6·50	6	7·50	6·25	3·25	·50	·50	42·0
73·5–74·5	—	—	—	—	1	—	1·50	1·50	—	5·25	2·25	2·50	6·50	3·25	3·25	—	2	29·0
74·5–75·5	—	—	—	—	—	—	—	—	—	1	2	—	2·50	·75	1·75	·50	—	8·5
75·5–76·5	—	—	—	—	—	—	—	—	—	1·25	·25	—	·50	1	1	—	—	4·0
76·5–77·5	—	—	—	—	—	—	—	—	—	1·25	·25	1	—	—	1·50	—	—	4·0
77·5–78·5	—	—	—	—	—	—	—	—	—	—	1	1	—	·25	·75	—	—	3·0
78·5–79·5	—	—	—	—	—	—	—	—	—	—	—	—	—	·25	·25	—	—	·5
Totals.	3	3·5	8	17	33·5	61·5	95·5	142	137·5	154	141·5	116	78	49	28·5	4	5·5	1078

* By a printer's error unfortunately given as ·1 in 'Biometrika.'

correlationship of three characters.* The source of this novel condition appears to lie in the integration of the term ζ^2/ζ_0, and this is only possible by use of equation (i.), provided the surface $Z = \zeta^2/\zeta_0$ has "ellipsoidal" contours. If it has not, we may get the subject of integration becoming infinite with one or other of the x's, and consequently, although both ζ and ζ_0 vanish at ∞, ζ^2/ζ_0 may not do so, *i.e.*, the mean square contingency tends in certain tracks to become indefinitely large. In fact, our method of deducing multiple contingency from the normal correlation coefficients is only valid provided the system is not only a possible correlation system with the given values of the coefficients, but also when these coefficients all have their signs reversed.

(11.) *Illustrations.* A.—*Stature in Father and Son.*

Table II. gives the distribution of 1078 cases of stature in father and son.† The correlation r, as found from the product moment in the usual way, is ·514.

I propose to consider the approach of C_1 and C_2 to r as we increase the fineness of the grouping. Clearly it would involve extreme labour to work out the contingencies—especially the mean square contingency—for the table as it stands.

To begin with I classed in three inch groups and got the following table, in which the figures in brackets are the independent probabilities.

TABLE III.—Stature of Father and Son in Inches.

		Stature of Father.							
		58·5–61·5.	61·5–64·5.	64·5–67·5.	67·5–70·5.	70·5–73·5.	73·5–76·5.	Totals.	Chances.
Stature of Son.	58·5–61·5	— (·05)	1·5 (·36)	2 (1·20)	— (1·32)	— (·50)	— (·03)	3·5	·0032
	61·5–64·5	3·5 (·84)	19 (6·50)	33 (21·75)	5·5 (23·87)	1·5 (9·02)	— (·55)	62·5	·0580
	64·5–67·5	8·5 (4·02)	53·75 (31·07)	148 (104·03)	80·5 (114·15)	8·25 (43·14)	— (2·64)	299	·2774
	67·5–70·5	2·5 (6·07)	33·25 (46·86)	149·25 (156·90)	202·25 (172·17)	60·25 (65·06)	3·5 (3·97)	451	·4184
	70·5–73·5	— (2·87)	3·5 (22·13)	39·75 (74·10)	104·25 (81·31)	62 (30·73)	3·5 (1·88)	213	·1976
	73·5–76·5	— (·56)	1 (4·31)	3 (14·44)	14·5 (15·84)	20·5 (5·99)	2·5 (·37)	41·5	·0385
	76·5–79·5	— (·10)	— (·77)	— (2·59)	4·5 (2·84)	3 (1·07)	— (·07)	7·5	·0069
	Totals . .	14·5	112	375	411·5	155·5	9·5	1078	1·0000

* For example, if ·5 be the value of parental correlation, then the correlation of two brothers could not exceed ·5 without making R′ negative.

† See 'Biometrika,' vol 2, p. 415.

The independent probabilities were found by multiplying the "chances" of a son occurring in each group by the totals for each group of fathers. Taking the difference of the observed sub-group frequencies and the independent probability frequencies, we have $N \times \psi = 205{\cdot}62$ from the positive and $= - 205{\cdot}66$ from the negative differences, a quite good agreement. Hence we find $\psi = {\cdot}1908$.

Using Diagram I. we have

$$C_2 = {\cdot}522.$$

Proceeding now to the mean square contingency obtained by squaring all the above found contingencies, dividing each by the independent probability frequency and summing, we find

$$\phi^2 = {\cdot}2755,$$

whence

$$C_1 = {\cdot}465.$$

The value of C_1 is clearly too small. We must infer that our grouping was not fine enough. Accordingly in Table IV. I have re-arranged the matter in 2-inch groupings, and have then in the same manner proceeded to find ψ and ϕ^2. In this case I found $\psi = {\cdot}2013$, and thus

$$C_2 = {\cdot}542,$$

while

$$\phi^2 = {\cdot}3568,$$

and

$$C_1 = {\cdot}513.$$

I thus conclude that the grouping is now fine enough to give C_1 and C_2 approximately equal to the correlation.*

* *i.e.*, within the probable error of that result.

TABLE IV.—Stature of Father and Son in Inches.

		Stature of Father.										
		58·5–60·5.	60·5–62·5.	62·5–64·5.	64·5–66·5.	66·5–68·5.	68·5–70·5.	70·5–72·5.	72·5–74·5.	74·5–76·5.	Totals.	Chances.
Stature of Son.	59·5–61·5	— (·02)	— (·08)	1·5 (·31)	1 (·77)	1 (·95)	— (·84)	— (·41)	— (·11)	— (·02)	3·5	·00325
	61·5–63·5	·5 (·14)	2·75 (·56)	5·75 (2·11)	9·5 (5·29)	5 (6·49)	·25 (5·73)	·25 (2·83)	— (·72)	— (·12)	24	·02226
	63·5–65·5	4 (·60)	7·75 (2·32)	20 (8·81)	41·5 (22·03)	17·25 (27·04)	8·25 (23·89)	1·25 (11·78)	— (3·01)	— (·51)	100	·09276
	65·5–67·5	2 (1·43)	10 (5·51)	32 (20·93)	73 (52·33)	78·75 (64·22)	33·5 (56·73)	7·25 (27·98)	1 (7·16)	— (1·21)	237·5	·22032
	67·5–69·5	— (1·95)	4·5 (7·49)	27·75 (28·46)	65·5 (71·16)	95 (87·34)	93·25 (77·15)	31·5 (38·03)	4·5 (9·74)	1 (1·65)	323	·29963
	69·5–71·5	— (1·42)	— (5·47)	6·75 (20·80)	38·25 (51·99)	61 (63·82)	77·5 (56·37)	39·5 (27·80)	11 (7·11)	2 (1·20)	236	·21892
	71·5–73·5	— (·63)	— (2·44)	·25 (9·5)	5·75 (23·13)	24·75 (28·39)	34·5 (25·08)	32·25 (12·37)	7 (3·17)	·5 (·54)	105	·09740
	73·5–75·5	— (·23)	— (·87)	1 (3·31)	3 (8·26)	6·25 (10·14)	6·75 (8·96)	13 (4·42)	5·5 (1·13)	2 (·19)	37·5	·03479
	75·5–77·5	— (·05)	— (·19)	— (·70)	— (1·76)	2·5 (2·16)	1·5 (1·91)	1·5 (·94)	2·5 (·24)	— (·04)	8	·00742
	77·5–79·5	— (·02)	— (·08)	— (·31)	— (·77)	— (·95)	2 (·84)	·5 (·41)	1 (·11)	— (·02)	3·5	·00325
	Totals.	6·5	25	95	237·5	291·5	257·5	127	32·5	5·5	1078	1·00000

To show the effect of too fine a grouping, I worked out the mean contingency for the inch grouping in Table II. There resulted

$$\psi = \cdot 2309, \text{ giving } C_2 = \cdot 597.$$

I therefore conclude that with sufficiently fine grouping the new method of contingency will give contingency coefficients sensibly equal to the correlation coefficient. But that with over fine grouping, the effect of individual units scattered here and there at random over the table, becomes influential and exaggerates the value of the correlation. Hence, when a correlation table can be formed and worked in the old ways, there is little doubt that it is safer to do so, and the labour will hardly be sensibly greater, at least when compared with the method of mean square contingency. I have not faced the labour required to determine the mean square contingency of the table with 340 sub-groups. Dr. LEE has worked out the mean square contingency for a table with 400 sub-groups, and we do not think it desirable to deal with a table of more than 10^2 to 15^2 entries again. Still the mean square contingency coefficient will hardly be as great on the full table as the mean contingency coefficient.

The following table gives the results :—

COMPARISON of Methods of Finding Correlation.

No. of groupings.	Mean contingency.	Mean square contingency.	Fourfold division.	Correlation table.
42	·522	·465	(Mean of six divisions)*	—
90	·542	·513	·550	—
340	·597	—	—	·514

Thus the first contingency method approaches the fourfold, the second, the ordinary correlation method.

Diagram II. at the end of this memoir gives the hyperbola of zero contingency for this case, calculated on the basis of the correlation coefficient being ·514. The means and standard deviations are :—

Father	67″·698,	2″·7048,
Son	68″·661,	2″·7321,

and the equation to the hyperbola referred to the means as the origin is

$$x^2 - 3{\cdot}8522yx + {\cdot}9801y^2 = 6{\cdot}2510.$$

The shaded squares are those of positive contingency. It will be seen that the hyperbola separates fairly well areas of positive, from areas of negative contingency. In most cases where there is an invasion across the boundary, the contingencies hardly differ from zero by amounts greater than the probable errors due to random sampling.

Illustration B.—*Data from Colour Inheritance in Greyhounds.*

In the previous example we have dealt with material in which contingency methods were directly comparable as to result with the correlation found by the "best" or product method process. In this illustration I deal with matter which can only provide a correlation to be found by the fourfold division process for comparison with the contingency coefficients. The data from which this illustration is drawn were extracted by Miss A. BARRINGTON from the 'Greyhound Studbook.' We deal with the inheritance of red and black pigments in the coat colour. I have selected six cases of the resemblance of brethren from *different* litters to compare the methods on. Tables were formed giving 16 to 25 contingency sub-groups of varying degrees of pigment, and these were worked out (*a*) by Miss BARRINGTON herself for the mean square contingency, (*b*) by myself for the mean contingency, and (*c*) by Dr. A. LEE

* See 'Phil. Trans.,' A, vol. 195, p. 42. The values range from ·521 to ·594, or almost the same range as we obtain from the mean contingency results.

for the fourfold correlation results. The results reached are given in the accompanying table. It is desirable to state that the number dealt with was about 1000 pairs of brethren in each case.

TABLE V.—Fraternal Resemblance of Greyhounds from Different Litters.

Character.	C_1, Mean Square Contingency.	C_2, Mean Contingency.	r, Fourfold Table.
Red in brothers	·478	·695	·456
„ sisters	·528	·612	·620
„ sister and brother	·488	·615	·450
Black in brothers	·512	·615	·558
„ sisters	·482	·632	·552
„ sister and brother	·502	·622	·593
Mean	·498	·632	·538
Mean deviation from mean . .	·016	·032	·057

We see at once from this table that the method of mean square contingency gives far more uniform results than either the mean contingency method or the fourfold division method. The average given by it is close to what we have found for fraternal resemblance, *i.e.*, ·5, in other cases, and within fairly close limits, all six cases now give ·5. The mean contingency gives results more divergent among themselves, but less so than those of the fourfold division method; their average, however, diverges most from what we have found in other cases.

The lesson, I think, to be learnt from this is: That the mean square contingency coefficient, although more laborious to find, is better than the mean contingency coefficient. That even with only 16 to 25 contingency sub-groups we may deduce results comparable with those obtained by fourfold divisions. But that it is probably *always* necessary to check a series by a certain number of fourfold division workings, for such are the only test that we have not got too crude a grouping reducing the contingency coefficient below the correlation value, or too fine a grouping introducing the difficulty already referred to (see p. 16), of magnifying the contingency coefficient owing to anomalous units.

Illustration C.—Hair Colour in Man.

I take the subject of hair colour because it is one in which doubts have been raised as to the order of pigments in a scale.

The following table gives the resemblance of pairs of brothers in hair colour :—

TABLE VI.

Second Brother.	First Brother. Red.	Fair.	Brown.	Dark.	Jet Black.	Totals.
Red	30·5	23	16	12	—	81·5
Fair	23	416	158	67·75	·25	665
Brown	16	158	394	98·25	8·25	674·5
Dark	12	67·75	98·25	328·5	19	525·5
Jet Black	—	·25	8·25	19	10	37·5
Totals	81·5	665	674·5	525·5	37·5	1984

The correlation found by taking the mean of four four-fold table divisions was ·621.* This result is based on the above scale order. We will now see what difference will arise if we work by contingency, so that the *scale order is absolutely indifferent, e.g.*, red might follow jet black.

We find

$$\phi^2 = \cdot 603896,$$

and accordingly $C_1 = \cdot 614$, a result within the limits of the probable error identical with the value of r found from the four-fold division method.

This illustration confirms the opinion I have already expressed, *i.e.*, that if the contingency be calculated for 16 to 36 sub-groups we shall obtain by the method of mean square contingency satisfactory results; *i.e.*, values close to the coefficient of correlation as found by product moment or four-fold division methods. In this case, as in others, I find the mean contingency far inferior to the mean square contingency.

My experience seems to show that about 25 sub-groups is the distribution to be aimed at; 9 is too few. Thus I worked out the relationship of temper in sisters for three-fold division—sullen, good-tempered, quick-tempered—or for 9 sub-groups. The method of mean contingency gave ·44 and of mean squared contingency ·36. Both far too small, as I find from each of four four-fold divisions a result of about ·5.

Illustration D.—*On Occupational or Professional Correlation between Relatives.*

I take as a final illustration a case in which any idea of scale is practically inconceivable, and yet one in which it is of considerable interest to measure the deviation from independent probability. It belongs to a class of problems in which I hope this new method of contingency will be fruitful of result. In classifying men into occupational and professional groups, we clearly cannot do so on the basis of any

* "Huxley Memorial Lecture," 'Journal of Anthropological Institute,' vol. 33, pp. 197 and 215.

scale which will put the army, church, and bar in any special order. On the other hand, it becomes of special interest to determine how far tastes and preferences for particular callings in life run in families. Miss EMILY PERRIN has undertaken a lengthy investigation of this kind, and has provided me with the pure contingency table given as Table VII. The occupations of 775 fathers and sons are here classed in broad general groups, which can be arranged purely alphabetically. More minute divisions and data for other series of relatives will be published later by Miss PERRIN, and it is not my present purpose to anticipate her conclusions, but merely to suggest the valuable applications which may be made of the novel methods to pure contingency results. What is the numerical measure of the relationship in pursuit between father and son, and how far is it removed from a mere chance relationship?

TABLE VII.—Contingency between Occupations of Fathers and Sons.

Nature of occupation.		Occupation of Son.														
		Army.	Art.	Teacher, Clerk, Civil Servant.	Crafts.	Divinity.	Agriculture.	Landownership.	Law.	Literature.	Commerce.	Medicine.	Navy.	Politics and Court.	Scholarship and Science.	Totals.
Occupation of Father.	Army	28	—	4	—	—	—	1	3	3	—	3	1	5	2	50
	Art	2	51	1	1	2	—	—	1	2	—	—	—	1	1	62
	Teacher, Clerk, Civil Servant	6	5	7	—	9	1	3	6	4	2	1	1	2	7	54
	Crafts	—	12	—	6	5	—	—	1	7	1	2	—	—	10	44
	Divinity	5	5	2	1	54	—	—	6	9	4	12	3	1	13	115
	Agriculture	—	2	3	—	3	—	—	1	4	1	4	2	1	5	26
	Landownership	17	1	4	—	14	—	6	11	4	1	3	3	17	7	88
	Law	3	5	6	—	6	—	2	18	13	1	1	1	8	5	69
	Literature	—	1	1	—	4	—	—	1	4	—	2	1	1	4	19
	Commerce	12	16	4	1	15	—	—	5	13	11	6	1	7	15	106
	Medicine	—	4	2	—	1	—	—	—	3	—	20	—	5	6	41
	Navy	1	3	1	—	—	—	1	—	1	1	1	6	2	1	18
	Politics and Court	5	—	2	—	3	—	1	8	1	2	2	3	23	1	51
	Scholarship and Science	5	3	—	2	6	—	1	3	1	—	—	1	1	9	32
	Totals	84	108	37	11	122	1	15	64	69	24	57	23	74	86	775

Miss PERRIN has extracted this first series from the 'Dictionary of National Biography'; hence she has, as a rule, tabled the distinguished, or at least moderately distinguished, sons of less distinguished fathers. It is, for example, not easy to win any form of distinction in agriculture. For this reason the distribution of occupations for sons differs widely from that of the occupations for fathers. There has accordingly been selection of the second generation, which undoubtedly must influence the result, *i.e.*, tend to weaken the observed relationship.

Working out the 196 contingencies, squaring, dividing by the independent probability frequencies, summing and averaging, I find for the mean square contingency

$$\phi^2 = 1{\cdot}299206,$$

whence

$$\phi^2/(1 + \phi^2) = {\cdot}393794,$$

and the coefficient of mean square contingency = ·6275. This would correspond to the correlation in occupation between father and son. Now if occupation were settled solely by fitness or taste, and these characters were inherited as other human faculties, we should expect the correlation between father and son to be about ·46.* Or, roughly, the hereditary relationship is increased by about $\frac{1}{3}$ in the matter of occupation. Remembering what we have noted as to selection above, the real increment is probably somewhat larger than this. Roughly, however, we may conclude from Miss PERRIN's data that about $\frac{3}{4}$ of the observed resemblance in occupation between father and son is due to hereditary influences, and the remaining $\frac{1}{4}$ to environmental effect. These numbers are subject to revision when Miss PERRIN's data are more ample and have been more fully analysed and discussed.

(12.) *General Conclusions.*

The general conception of contingency developed in this memoir I consider in the first place of *theoretical* importance. Its practical applications are not negligible, but are, for reasons given below, of less importance than might *à priori* be supposed.

(A.) In the first place, the conception of contingency enables us at once to generalise the notion of the association of two attributes developed by Mr. YULE. We can class individuals not into two alternate groups, but into as many groups with exclusive attributes as we please, and either the mean contingency or the mean square contingency will enable us to see the extent to which two such systems are contingent or non-contingent.

(B.) This result enables us to start from the mathematical theory of independent probability as developed in the elementary text books, and build up from it a generalised theory of association, or, as I term it, *contingency*. We reach the notion of a pure contingency table, in which the order of the sub-groups is of no importance whatever.

(C.) We then investigate the relation of contingency to normal correlation, and find that with normal frequency distributions both contingency coefficients pass with sufficiently fine grouping into the well-known correlation coefficient. Since, however,

* 'Biometrika,' vol. 2, p. 379.

the contingency is independent of the order of grouping, we conclude that when we are dealing with alternative and exclusive sub-attributes, we need not insist on the importance of any particular order or scale for the arrangement of the sub-groups.

(D.) This conception can be extended from normal correlation to any distribution with linear regression; small changes (*i.e.*, such that the sum of their squares may be neglected as compared with the square of mean or standard deviation) may be made in the order of grouping without affecting the correlation coefficient.

(E.) The results (C) and (D) are not so fruitful for practical working as might at first sight appear, for they depend in practice on the legitimacy of replacing finite integrals by sums over a series of varying areas, where no quadrature formula is available. If we, to meet the difficulty, make a very great number of small classes, the calculation, especially of the mean square contingency, becomes excessively laborious. Further, since in observation individuals go by units, casual individuals, which may fairly represent the total frequency of a considerable area, will be found on some one or other isolated small area, and thus increase out of all proportion the contingency. The like difficulty occurs when we deal with outlying individuals in the case of frequency curves, only it is immensely exaggerated in the case of frequency surfaces.

(F.) It is thus not desirable in actual practice to take too many or too fine sub-groupings. It is found, under these conditions, that the correlation coefficient as determined by the product moment or fourfold division methods is approximated to more closely in the case of the contingency coefficient found from mean square contingency than in the case of that found from mean contingency. Probably 16 to 25 contingency sub-groups will give fairly good results in the case of mean square contingency, but for each particular type of investigation it appears desirable to check the number of groups proper for the purpose by comparison with the results of test fourfold division correlations. Under such conditions it appears likely that very steady and consistent results will be obtained from mean square contingency.

(G.) Finally, contingency may be applied—of course, at first tentatively and with caution—in the consideration of a whole class of problems in which no attempt at a scale or order of sub-groups is possible, in short, where alphabetical order is as good as any other. For example, it would seem to be available in a vast range of problems of exclusive and alternative inheritance.