Biometrika Trust

NOTES ON THE HISTORY OF CORRELATION.

Being a paper read to the Society of Biometricians and Mathematical Statisticians, June 14, 1920.

By KARL PEARSON, F.R.S.

(1) As I have often stated, Laplace anticipated Gauss by some 40 years. In his memoir of 1783, *Histoire de l'Académie*, pp. 423—467, he gives the expression for the probability integral

$$\frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{1}{2}x^2} dx$$

and suggests (p. 433) its tabulation as a useful task. It is clear that to do this is to recognise the existence of the probability-curve

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

or in its doubly projected form

$$y = \frac{N}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}.$$

Laplace's investigation while not proceeding from the very simple axioms of Gauss, which lead directly to the above equation, is more satisfactory than Gauss' because we see better the nature of the approximations by which the curve is reached and get hints of how to generalise it. Many years ago I called the Laplace-Gaussian curve the *normal* curve, which name, while it avoids an international question of priority, has the disadvantage of leading people to believe that all other distributions of frequency are in one sense or another 'abnormal.' That belief is, of course, not justifiable. It has led many writers to try and force all frequency by aid of one or another process of distortion into a 'normal' curve.

Gauss starting with a normal curve as the law of distribution of errors reached at once the method of least squares. To understand the origin of the correlational calculus we must really go back to Gauss' fundamental memoirs on least squares, namely the *Theoria combinationis observationum erroribus minimis obnoxiae* of 1823 and the *Supplementum* of 1826.

We observe or measure *directly* a certain number of quantities a, b, c, d, ... Each of these quantities is supposed by Gauss to be independent and to follow the normal law. The combined probability of the system* is accordingly

$$P \propto e^{-\frac{1}{2}\left\{\left(\frac{a-\bar{a}}{\sigma_a}\right)^2 + \left(\frac{b-\bar{b}}{\sigma_b}\right)^2 + \left(\frac{c-\bar{c}}{\sigma_c}\right)^2 + ...\right\}},$$

or the product of the independent probabilities, where $\sigma_a, \sigma_b, \sigma_c, ...$ are the variability in errors of a, b, c, ... and $\bar{a}$, $\bar{b}$, $\bar{c}$, ... the means. This probability will be a maximum when

$$u^2_{a,\, b,\, c,} \;\; = S\left(\frac{a-\bar{a}}{\sigma_a}\right)^2$$

is a minimum. This is really the principle of weighted least squares. Its validity depends upon the normal law of distribution of error. Without this law holding it may be a utile method, but we have no means of proving it the 'best.'

The investigator in Gauss' case is, however, not interested in the quantities observed, but in certain indirectly ascertained quantities x_1, x_2, ... x_n which are functions of them. Thus

$$x_1 = f_1(a, b, c, ...),$$
$$x_2 = f_2(a, b, c, ...),$$
$$... = ...,$$

where $f_1, f_2, ...$ are known functions. Now Gauss cannot as a rule express from these general equations a, b, c, ... in terms of x_1, x_2, ... x_n.

He assumes that all of them differ slightly from their mean or 'true' values and accordingly expands by Taylor's theorem and reaches the result†

$$x_1 - \bar{x}_1 = \alpha_1(a - \bar{a}) + \beta_1(b - \bar{b}) + \gamma_1(c - \bar{c}) + ...,$$
$$x_2 - \bar{x}_2 = \alpha_2(a - \bar{a}) + \beta_2(b - \bar{b}) + \gamma_2(c - \bar{c}) + ...,$$

where the α, β, γ, ... are $\dfrac{df}{da}$, $\dfrac{df}{db}$, $\dfrac{df}{dc}$, ... and can be ascertained *a priori*. Clearly Gauss supposes that a linear relationship is adequate, in other words he replaces statistical differentials by mathematical differentials, a step he does not really justify.

From these linear equations we can find the $a - \bar{a}$, $b - \bar{b}$, $c - \bar{c}$, ... in terms of the indirectly observed variables $x_1 - \bar{x}_1$, $x_2 - \bar{x}_2$, $x_3 - \bar{x}_3$, ... by solution in determinantal form, say

$$a - \bar{a} = A_1(x_1 - \bar{x}_1) + B_1(x_2 - \bar{x}_2) + C_1(x_3 - \bar{x}_3)...,$$
$$b - \bar{b} = A_2(x_1 - \bar{x}_1) + B_2(x_2 - \bar{x}_2) + C_2(x_3 - \bar{x}_3)....$$

Substituting in u^2 we find

$$u^2_{x_1,\, x_2,\, ...} = S\left(\frac{A_1^2}{\sigma_a^2}\right)(x_1 - \bar{x}_1)^2 + S\left(\frac{B_1^2}{\sigma_a^2}\right)(x_2 - \bar{x}_2)^2 + 2S\left(\frac{A_1 B_1}{\sigma_a^2}\right)(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) + ...$$

Hence the probability of x_1, x_2, ... occurring is

$$P \propto e^{-\frac{1}{2}u^2_{x_1,\, x_2,\, ...}}.$$

---

* I use throughout notation which I assume now-a-days to be more familiar than that of Gauss

† $\bar{a}$, $\bar{b}$, $\bar{c}$, ... are actually in Gauss' method approximate or *guessed* solutions not means, but this does not affect the general nature of the discussion.

This is a normal surface which contains *the product terms*. As we now interpret it we say that the x's are *correlated* variates. And in this sense Gauss in 1823 reached the normal surface of n correlated variates. But he does not seek to express all his relations in terms of the S.D.'s σ_{x_1}, σ_{x_2}, σ_{x_3}, ... and the correlations r_{12}, r_{23}, ... of these variates. These x-variates are not for Gauss, nor for those who immediately followed him, the *directly* observed quantities. What he is seeking is the expression for σ_x, or the probable error of an indirectly observed variate in terms of

$$S\left(\frac{A_1^2}{\sigma_a^2}\right), \quad S\left(\frac{B_1^2}{\sigma_a^2}\right), \quad S\left(\frac{A_1 B_1}{\sigma_a^2}\right), \quad \ldots$$

In this case A, B, C are ratios of minors and determinants of the α, β, γ, ... which are Gauss' known quantities. His object therefore is to express σ_x not from direct observations but in terms of α, β, γ, ... through the sums of determinantal terms.

Writers on Least Squares and Adjustment of Observations then take w any function of x_1, x_2, ... x_n, i.e.

$$w = F(x_1, x_2, \ldots x_n),$$

express the relation in a linear form, i.e.

$$w - \overline{w} = \lambda_1 (x_1 - \bar{x}_1) + \lambda_2 (x_2 - \bar{x}_2) + \ldots,$$

and then, to find σ_w^2, go through lengthy analysis to determine

$$\text{Mean } (x_1 - \bar{x}_1)^2, \quad \text{Mean } (x_2 - \bar{x}_2)^2, \quad \text{Mean } (x_1 - \bar{x}_1)(x_2 - \bar{x}_2), \text{ etc.}$$

in terms of the original α, β, γ, ... There is not a word in their innumerable treatises that what is really being sought are the mutual correlations of a system of correlated variables. The mere using of the notation of the correlational calculus throws a flood of light into the mazes of the theory of errors of observation. There is much more in the theory of least squares than I have stated; there are equations of conditions—the angle and side equations of geodesy, etc.—these only complicate the matter. The point is this: that the Gaussian treatment leads (i) to a non-correlated surface for the directly observed variates, (ii) to a correlation surface for the *indirectly* observed variates. This occurrence of product terms arises from the geometrical relations between the two classes of variates, and not from an organic relation between the indirectly observed variates appearing on our direct measurement of them.

It will be seen that Gauss' treatment is almost the inverse of our modern conceptions of correlation. For him the *observed* variables are independent, for us the observed variables are associated or correlated. For him the non-observed variables are correlated owing to their known geometrical relations with observed variables; for us the unobservable variables may be supposed to be uncorrelated causes, and to be connected by unknown functional relations with the correlated variables. In short there is no trace in Gauss' work of observed physical variables being—apart from equations of condition—associated organically which is the fundamental conception of correlation.

(2) The next important work to be considered is that of August Bravais. It is entitled "Sur les probabilités des erreurs de situation d'un point." It was published in the *Mémoires présentés par divers savants à l'Académie royale des Sciences de l'Institut de France*, T. IX. Paris, 1846, pp. 256—332. It appears, however, to have been reported favourably upon in 1838*. Bravais was in many respects a remarkable man. Essentially a geologist he wrote also on astronomy, physics, meteorology and the theory of probabilities. He made a voyage to Lapland for geodesic purposes and took the opportunity of measuring a number of Lapp skulls! He had a width of action most sympathetic to the biometrician.

Writing in 1895 of the history of correlation I said:

" The fundamental theorems of correlation were for the first time and almost exhaustively discussed by Bravais [Title as above of his memoir] nearly half a century ago. He deals completely with the correlation of two and three variables." Then speaking of Galton's coefficient of correlation I say: "This indeed appears in Bravais' work, but a single symbol is not used for it. It will be found of great value in the present discussion. In 1892 Professor Edgeworth, also unconscious of Bravais' memoir, dealt in a paper on 'Correlated Averages' with correlation for three variables (*Phil. Mag.* Vol. XXXIV. 1892, pp. 194—204).' He obtained results identical with Bravais', although expressed in terms of 'Galton's functions'" [i.e. coefficients of correlation].

Again later, p. 287, in giving the fundamental equation for the correlation of three variates I wrote: "This agrees with Bravais' result, except that he writes for r_1, r_2, r_3 the values $\Sigma (yz)/(n\sigma_2\sigma_3)$ etc., which we have shown to be the best values (see *loc. cit.* p. 267)." Again on p. 301 I write before proving the general theorem of multiple correlation: "*Edgeworth's Theorem.* We may stay for a moment over the results above to deduce Professor Edgeworth's Theorem," with the footnote, "Briefly stated with some rather disturbing printer's errors in the 'Phil. Mag.' Vol. XXXIV. p. 201, 1892."

Now all these statements if they were correct would indicate that Bravais discovered correlation before Galton and that Edgeworth first published the form of the multiple correlation surface. They have been accepted by later writers, notably Mr Yule in his manual of statistics, who writes (p. 188):

"Bravais introduced the product-sum, but not a single symbol for a coefficient of correlation. Sir Francis Galton developed the practical method, determining his coefficient (Galton's function as it was termed at first) graphically. Edgeworth developed the theoretical side further and Pearson introduced the product-sum formula."

Now I regret to say that nearly the whole of the above statements are hopelessly incorrect. Bravais has no claim, whatever, to supplant Francis Galton as the discoverer of the correlational calculus. For the most part he is simply taking a very special case of the Gaussian analysis, and nowhere on p. 267 of his memoir can I now find that he has used the expressions for the correlation symbols without

---

* *Comptes rendus*, T. VII. p. 77.

their names. Again Edgeworth did *not* obtain results identical with Bravais', he went on a route of his own to find the true multiple correlation surface and gave as I said in 1895 only doubtful results. But I fear they were not all due to printer's errors. On re-examining his memoir 25 years later I think he harnessed imperfect mathematical analysis to a jolting car and drove it into an Irish bog on his road, and that it was doubtful analysis not errors of printing which led to his obscure conclusions. I was scarcely justified in 1895 in calling the multiple regression result Edgeworth's Theorem. He had tried in 1892 to solve the problem, and he can hardly be said to have succeeded properly. It is very difficult to explain now how my errors of ascription came about, still less possible is it to understand why later writers have not corrected my false history, but merely repeated it.

As far as I can remember what happened at all, it was as follows. I know that I was immensely excited by Galton's book of 1889—*Natural Inheritance*—and that I read a paper on it in the year of its appearance. In 1891—2 I lectured popularly on probability at Gresham College, taking skew whist contours as illustrations of correlation. In 1892 I lectured on variation, in 1893 on correlation to research students at University College, the material being afterwards published as the first four of my *Phil. Trans.* memoirs on evolution. At this time I dealt with correlation and worked out the general theory for three*, four and ultimately n variables. The field was very wide and I was far too excited to stop to investigate properly what other people had done. I wanted to reach new results and apply them. Accordingly I did not examine carefully either Bravais or Edgeworth, and when I came to put my lecture notes on correlation into written form, probably asked somebody who attended the lectures to examine the papers and say what was in them. Only when I now come back to the papers of Bravais and Edgeworth do I realise not only that I did grave injustice to others, but made most misleading statements which have been spread broadcast by the text-book writers.

(3) Let us now examine Bravais' memoir. He commences by stating that he is going to measure the errors of the determination of the coordinates x, y, z of a point in space. These coordinates are not measured directly but are functions of the observed elements $a, b, c, \ldots$, and he puts

$$x = \phi\,(a,\,b,\,c,\,\ldots),$$
$$y = \psi\,(a,\,b,\,c,\,\ldots),$$
$$z = \chi\,(a,\,b,\,c,\,\ldots).$$

He then expands x, y, z linearly in terms of a, b, c assuming that mathematical differentials may be used for errors; thus he writes

$$\delta x = A\,\delta a + B\,\delta b + C\,\delta c + \ldots,$$
$$\delta y = A'\,\delta a + B'\,\delta b + C'\,\delta c + \ldots,$$
$$\delta z = A''\,\delta a + B''\,\delta b + C''\,\delta c + \ldots.$$

He tells us that the A, B, C are differential coefficients, i.e. of the known functions ϕ, ψ, χ, and that to justify the neglect of higher powers and products we must get

* Published in the *R. S. Proc.* Vol. LVIII. p. 241, 1895.

rid of constant sources of error which arise chiefly from vices of method of observation, ignorance of physical laws, etc. That they can be removed by increasing the number of our observations, and in surveying—which he has essentially in mind—by using the repeating circle, which destroys the majority of constant errors and lessens the influence of variable causes by the fact itself of repeating the observed angles. It is clear that he is thinking solely of theodolite work, and that his x, y, z are Gauss' indirectly observed quantities, his directly observed quantities being angles and bases a, b, c, ...

He now changes his notation; he uses x, y, z for the errors δx, δy, δz, and m, n, p for δa, δb, δc, and takes equations

$$x = Am + Bn + Cp + \dots$$

and calls x, y, z the *dependent* variables, m, n, p the *independent* variables. He says that Laplace has shown that a variation of x between x and $x + \delta x$ will be of the form

$$\sqrt{\frac{h_x}{\pi}}\, e^{-h_x x^2} dx,$$

where h_x is given by

$$\frac{1}{h_x} = \frac{A^2}{h_m} + \frac{B^2}{h_n} + \frac{C^2}{h_p} + \dots.$$

It is therefore clear that he supposes that his *observed* quantities m, n, p, ... are *uncorrelated* in our sense of the word. In fact he gives for two and three variates the expressions

$$\sqrt{\frac{h_m h_n}{\pi . \pi}}\, e^{-(h_m m^2 + h_n n^2)} dm\, dn \quad \text{(p. 261),}$$

$$\sqrt{\frac{h_m h_n h_p}{\pi . \pi . \pi}}\, e^{-(h_m m^2 + h_n n^2 + h_p p^2)} dm\, dn\, dp \quad \text{(p. 264).}$$

There is obviously not a single step, not a line in this, which does not occur in Gauss, except that Gauss would use

$$u^2 = h_m m^2 + h_n n^2 + h_p p^2$$

and not trouble to state that the probability was given by the exponential.

Now Gauss' problem was to express the variability of x in terms of the variability of the observed quantities a, b, c, ... or m, n, p, and of the differential coefficients A, B, C. This is absolutely the same as Bravais' problem, and Bravais' treatment goes very little further than Gauss'—indeed it is essentially narrower as while Gauss neither limits the number of his variables nor their nature, Bravais treats only of *position* in space.

I will now give the value of the expression Bravais reaches for his surface of two dimensions, expressing by $d^2 w$ the briquette of frequency on $dx\, dy$:

$$\frac{d^2 w}{dx\, dy} = \frac{1}{\pi}\, \frac{1}{\left\{ \Sigma\, \frac{(AB' - A'B)^2}{h_m h_n} \right\}^{\frac{1}{2}}}\, e^{- \dfrac{\left\{ x^2 \Sigma \frac{A'^2}{h_m} - 2xy\, \Sigma \frac{AA'}{h_m} + y^2 \Sigma \frac{A^2}{h_m} \right\}}{\Sigma \left(\frac{(AB' - A'B)^2}{h_m h_n} \right)}} \quad \text{(p. 272).}$$

Now, if you take

$$x = Am + Bn + \dots,$$
$$y = A'm + B'n + \dots,$$
$$\sigma_x^2 = A^2\sigma_m^2 + B^2\sigma_n^2 + \dots,$$
$$\sigma_y^2 = A'^2\sigma_m^2 + B'^2\sigma_n^2 + \dots,$$
$$\text{Mean}\,(\delta x\,\delta y) = r_{xy}\sigma_x\sigma_y = AA'\sigma_m^2 + BB'\sigma_n^2 + \dots.$$

Whence

$$\sigma_x^2\sigma_y^2(1 - r_{xy}^2) = \Sigma\,(A^2B'^2 + A'^2B^2 - 2AA'BB')\,\sigma_n^2\sigma_m^2$$
$$= \Sigma\,(AB' - A'B)^2\,\sigma_n^2\sigma_m^2,$$

whence, remembering $h_m = \dfrac{1}{2\sigma_m^2},\quad h_n = \dfrac{1}{2\sigma_n^2},$

we easily deduce the

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

of our familiar notation.

But this is precisely what Bravais does *not* do, and for the simple reason that his x, y, z are not variables which he has directly determined and for which he can directly find σ_x, σ_y and r_{xy}. He is merely seeking to express the variability of x and y in terms of the directly determined constants and certain differential co-efficients. This is one of the fundamental problems of the Method of Least Squares and had already been solved by Gauss. Bravais adds so far *nothing whatever* to Gauss' solution of 20 years earlier. If Bravais discovered correlation, then Gauss had done so previously.

As a matter of fact while the above expression shows how a hasty examination of Bravais' memoir might lead one to believe he had reached the correlation surface, he was in fact occupied with an entirely different problem, one which was really only a particular case of Gauss' earlier and more comprehensive work.

We cannot pass over, however, the really valuable portion of Bravais' memoir. It lies in this: Having got his coefficients of x and y in terms of the differential coefficients A, B, C, ... he writes the surface

$$\frac{K}{\pi}\,e^{-K^2(a_1x^2 - 2\beta_0xy + a_2y^2)},$$

and then discusses the properties of a surface of which the contours are

$$ax^2 + 2cxy + by^2 = D,$$

i.e. the familiar ellipses of our normal surface. He gets the conjugate of x-axes as the locus of maximum y's and determines the probability of points lying in certain areas—bounded by similar ellipses or in angular sectors. He gets the line $x = -\dfrac{c}{a}y$, which corresponds to Galton's regression-line. But this is not a result of observing x and y and determining their association, but of the fact that x and y

are functions of certain independent and directly observed quantities. When he thinks of c and a at all, it is not in terms of observations on x and y but of the differential coefficients A, B, C of the geometrical relations between position in space and the angles by which that position is found.

Next we come to his surface of three variates and the treatment is identical. He writes

$$z = \frac{G}{\pi^{\frac{3}{2}}} e^{-(ax^2 + by^2 + cz^2 + 2ezy + 2fxz + 2gxy)} \quad \text{(p. 296)},$$

and his primary object is to determine a, b, c, e, f, g in terms of the differential coefficients A, B, C and the variabilities of the observed independent variates.

Thus he gives

$$a = G^2 \Sigma \frac{1}{h_m h_n} \{(A'B'')\}^2, \quad e = G^2 \Sigma \frac{1}{h_m h_n} \{(A''B)(A''B')\}, \quad \frac{1}{G^2} = \Sigma \frac{\{(AB'C'')\}^2}{h_m h_n h_p}.$$

There is throughout merely the standpoint of the Gaussian method of treating errors of observation, and if we are to attribute any discovery of the idea of correlation to Bravais we must with the same confidence assert that Gauss was the primary originator of the whole idea. To my mind this is absurd*. In the case of both these distinguished men the quantities they were observing were *absolutely independent*; they neither of them had the least idea of correlation between observed quantities. The product terms in their expressions—never analysed in the sense of correlation—arise solely not from organic relationships, but from the *geometrical* relationships which exist between their observed quantities and the indirectly observed quantities they deduce from them. Bravais himself (p. 331) says that the application of his results are narrowly circumscribed by the nature of his assumptions—astronomy and the great geodesic surveys alone provide sufficiently accurate material. As far as Gauss and Bravais are concerned we must, I think, hold that they contributed nothing of real importance to the problem of correlation, and that my statement of 1895 was a totally erroneous one.

The same criticism applies to all the treatment of the normal surfaces by later writers, which are described at very considerable length by Czuber in his *Theorie der Beobachtungfehler*, Leipzig, 1891. In all cases the variables are *indirectly* observed quantities and the product terms arise because they are mathematically supposed to be linear functions of the directly observed, but quite independent variables. That the directly measured quantities might themselves be correlated does not seem to have occurred to the many writers on the theory of observations.

As far as I am aware there is nothing to record on our subject beyond the work of the writers on the theory of observations referred to above until we reach Francis Galton himself. His first statement of his ideas was in a lecture at the

---

* I feel quite certain that if any one had told either Gauss or Bravais that $\Sigma(ab)$ for their *observed* measurements need not be zero, they would have been laughed out of court, as the astronomers now laugh at us, when we assert that their measurements of different stellar magnitudes are very probably correlated !

Royal Institution, Feb. 9, 1877. He had found it very difficult to collect human material for two generations and after careful consideration selected sweet pea seeds. These seeds were both measured and weighed and actually observations were taken on foliage and length of pod although as far as I am aware Galton never published the reductions of the latter. As he himself writes in 1885: "It was anthropological evidence that I desired, caring only for the seeds as means of throwing light on heredity in man. I tried in vain for a long and weary time to obtain it in sufficient abundance, and my failure was a cogent motive, together with others, in inducing me to make an offer of prizes for family records, which was largely responded to, and furnished me last year with what I wanted*."

The title of Galton's R. I. lecture was *Typical Laws of Heredity in Man.* Here for the first time appears a numerical measure r of what is termed 'reversion' and which Galton later termed 'regression.' This r is the source of our symbol for the correlation coefficient, which was really the first letter of 'reversion' not of 'regression.' The main results are given in a mathematical appendix†. Galton works with the modulus—i.e. our $\sqrt{2}\sigma$—probably because the tables of the probability integral were then given in the modulus as argument. But we can at once convert into more customary notation. Thus we find the now familiar result

$$v = \sqrt{1 - r^2}\, c_1,$$

or, translating his symbols:

Variability of family $= \sqrt{1 - r^2} \times$ variability of general population.

Galton had already reached the idea of homoscedasticity in the arrays of offspring. "I was certainly astonished to find the variability of the produce of the little seeds to be equal to that of the big ones; but so it was and I thankfully accept the fact, for had it been otherwise, I cannot imagine, from theoretical considerations, how the typical problem could be solved" (p. 10).

Next Galton supposes the mean taken of both parents and notes that the "variability of the parentage," what he would have called later the mid-parentage, $= \dfrac{1}{\sqrt{2}}$ variability of either parent. He has not yet reached the idea of reducing one sex to the standard of the other, and the result is only true, if we have to deal with characters not sexually differentiated.

Now we come to the test point‡:

"Reversion"—Galton tells us, p. 10—"is the tendency of the ideal mean filial type to depart from the parental type, reverting to what may be roughly and perhaps fairly described as the average ancestral type. If family variability had been the only process in simple descent that affected the characteristics of a sample the dispersion of the race from its mean ideal type would indefinitely increase with

---

* Address to Anthropological Section, *B. A. Report*, 1885, p. 1207.

† Royal Institution of Great Britain, Friday, February 9, 1877.

‡ Let the reader remember that these words were spoken just 40 years ago, and that they waited 12 to bring forth fruit!

the number of generations, but reversion checks this increase, and brings it to a standstill."
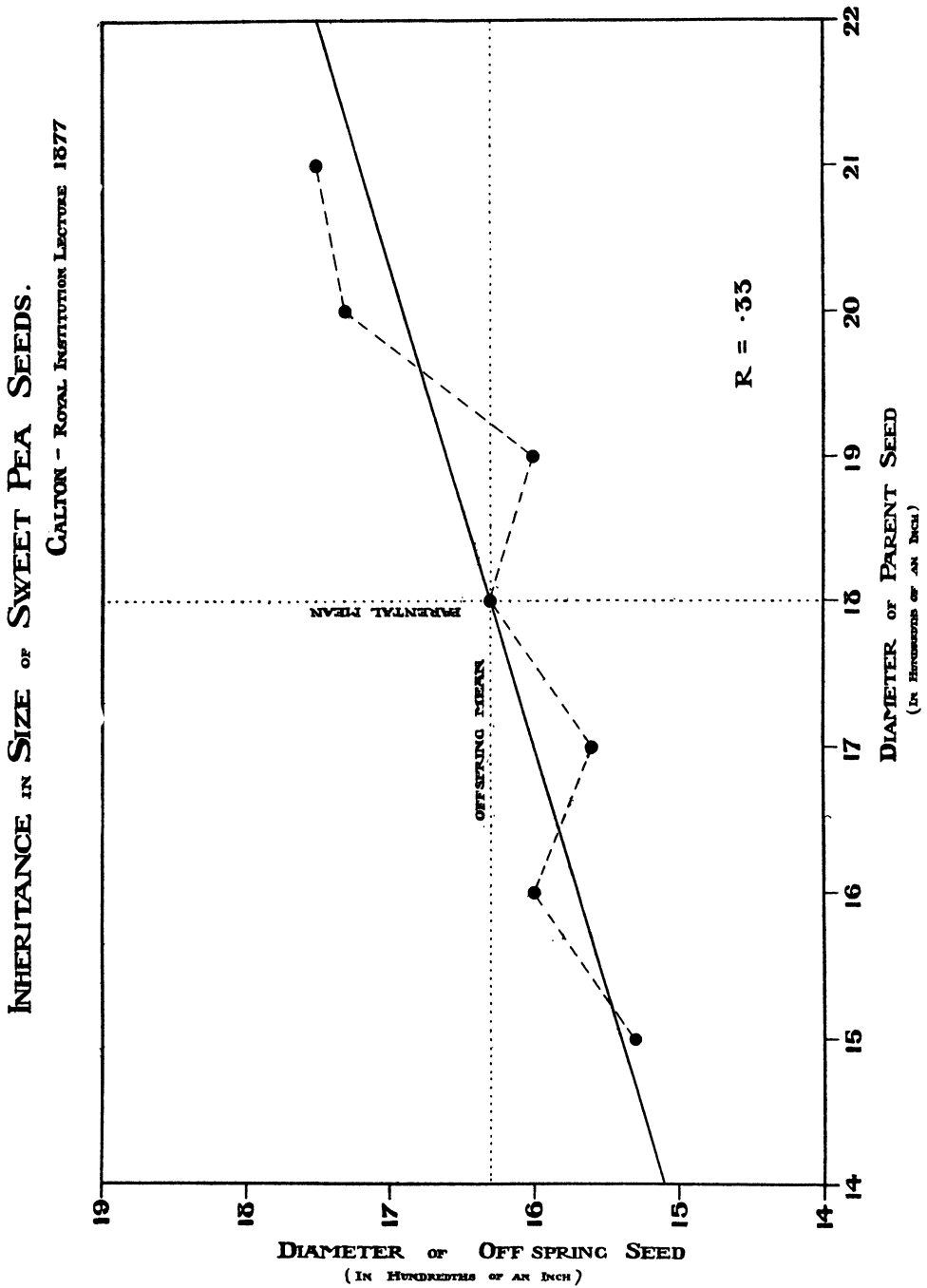
Galton's proof assuming homoscedasticity is of a very simple nature. Let the reversion be λx, where x is the parental character. Then the mean variability of the offspring generation

$$= \sigma^2 (1 - r^2) + \lambda^2 (\text{mean } x^2)$$
$$= \sigma^2 (1 - r^2) + \lambda^2 \sigma^2.$$

Therefore unless $\lambda = r$ the population cannot remain stable. Or without any hypothesis as to normality, only on the basis of linearity of reversion, homoscedasticity and stability the Galton coefficient r of reversion must be equal to the r which gives the reduction of the 'family variability.' Thus the lecture of 1877, while it contains points which later work was to clear up, still in the main lines gives on the data for size in sweet pea seeds the fundamental properties of the regression line. I have worked out Galton's data for sweet peas and show you a diagram of the result which Miss A. Davin has prepared for me. The parent seed was of course selected seed, and Galton took 100 of each parental grade and determined the mean of the offspring, which of course were non-selected seed, i.e. not seedsman's seed. Galton fixes the regression in round numbers at $\frac{1}{3}$, I make it slightly larger. In any case the regression coefficient is small, if we consider the sweet pea, as Galton did, as self-fertilising. It has been so proclaimed in several botanical investigations on heredity in the sweet pea. But in 1907 I watched a row of sweet peas and observed *Megachile Willughbiela*, the leaf-cutting bee, in quite considerable numbers visiting the flowers. The Superintendent of the R. H. S.'s garden at Wisley also replied to an inquiry that he had no doubt some English insect cross-fertilised sweet peas because in trying new sorts the gardeners had to place the rows in different parts of the garden to reduce the risk of cross-fertilisation. Darwin's statement[*] that "in this country it"—the sweet pea—"seems invariably to fertilise itself," appears open to question. Galton's coefficient may therefore, although low, be not so low as it appears on the assumption of self-fertilisation.

The next few years Galton was occupied in collecting material for further investigation of regression and heredity. He had established his Anthropometric Laboratory at South Kensington and by offering prizes obtained his Records of Family Faculties. The first-fruits of these data are to be found in his Presidential Address to the Anthropological Section of the British Association at Aberdeen in 1885. The part of this Address dealing with regression was considerably extended in a paper read to the Anthropological Institution in the same year. Galton now deals with the inheritance of stature and transmutes female to male stature before determining his mid-parentages. He does this, not as we should do now by multiplying by the ratio of paternal and maternal standard deviations, but by the multiplying factor of mean statures 1·08. This is roughly permissible if the coefficients of variation for the two sexes are the same as they very nearly are for stature. In this paper we have the first published diagram of the *two* regression

---

[*] *Cross and Self-fertilisation of Plants*, 1878, p. 153.

INHERITANCE IN SIZE OF SWEET PEA SEEDS.

GALTON - ROYAL INSTITUTION LECTURE 1877

R = ·35

PARENTAL MEAN

OFFSPRING MEAN

DIAMETER OF PARENT SEED
(IN HUNDREDTHS OF AN INCH)

DIAMETER OF OFF SPRING SEED
(IN HUNDREDTHS OF AN INCH)

3—2

lines and the first correlation table (of stature in parentage and offspring) as we should now call it.

Also Galton gives the diagram* which indicates how he discovered observationally the form of the normal frequency surface. He proceeded to smooth his correlation
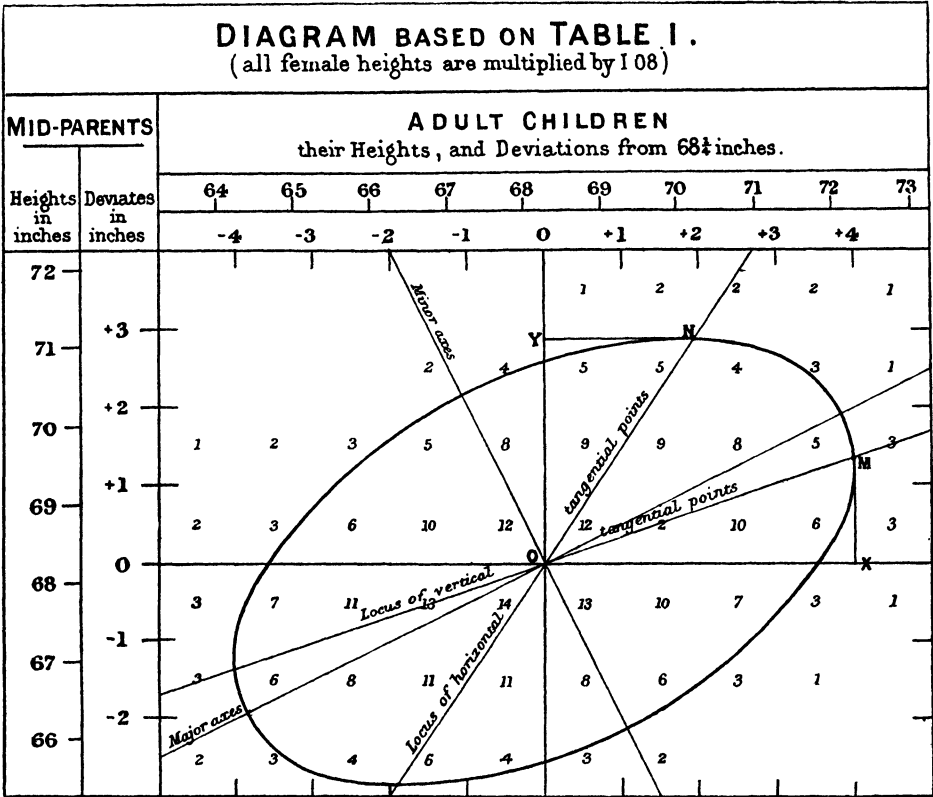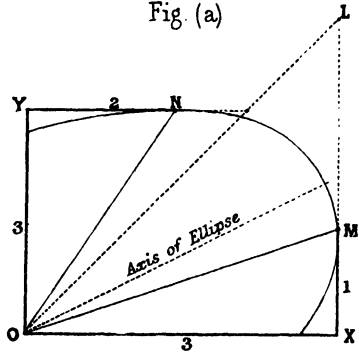

Fig. (a)

table by taking the mean of four adjacent cells, and then drawing contour lines through points of the same frequency. He found such contour lines were a system of concentric similar and similarly placed ellipsoids and that the regression lines

* Reproduced here by permission of the Royal Anthropological Institute.

were what the mathematician terms the conjugate diameters of the variate axes. He discovered that the sections parallel to the variate axes were 'apparently' normal curves of equal S.D. but that this S.D. was reduced and bore a constant ratio to the S.D. of the general population. He knew 8 years earlier the relation of $\sigma \sqrt{1 - r^2}$ to the 'reversion' coefficient r. That Galton should have evolved all this from his observations is to my mind one of the most noteworthy scientific discoveries arising from pure analysis of observations.

Why Galton did not at once write down the equation to his surface as

$$z \propto e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}} \times e^{-\frac{1}{2}\frac{1}{\sigma_x^2}\frac{1}{(1-r^2)}\left(x - r\frac{\sigma_x}{\sigma_y}y\right)^2}$$

has always been a puzzle to me. Actually he carried the problem, stated in the language of probability, to Mr J. D. Hamilton Dickson, a mathematician of Peterhouse, Cambridge, who after stating the wording of Galton's problem, wrote down the answer substantially as above in the fourth line of his memoir*! The fact is that Galton's statement of his problem, involving as it did the assumption of normal distribution, homoscedasticity and linear regression, provided the answer the moment his results were read in symbols. The explanation of Galton's action possibly lies in the fact that Galton was very modest and throughout his life underrated his own mathematical powers.

Thus in 1885 Galton had completed the theory of bi-variate normal correlation. The next stage in the theory of correlation, multi-variate correlation, was directly indicated by the general problem of ancestry. As is now well known the fundamental regression equation is

$$\frac{x_0 - \bar{x}_0}{\sigma_0} = -\Sigma\left(\frac{R_{0s}}{R_{00}}\frac{x_s - \bar{x}_s}{\sigma_s}\right) \quad \dots\dots\dots\dots\dots\dots\dots\text{(i)},$$

where R_{pq} is the p, q minor of the determinant

$$R = \begin{vmatrix} 1 & r_{01} & r_{02} \dots r_{0n} \\ r_{10} & 1 & r_{12} \dots r_{1n} \\ \dots\dots\dots\dots\dots\dots\dots \\ r_{n0} & r_{n1} & r_{n2} \dots 1 \end{vmatrix}$$

and the variability of the array is

$$\sigma_0\sqrt{\frac{R}{R_{00}}} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\text{(ii)}.$$

Galton endeavoured to reach this by a short cut, and thus evolved his law of ancestral heredity. This was a brilliant and suggestive step, but he was not able to state the conditions under which it is theoretically correct or bring forward data at that time to confirm its observational accuracy.

---

* R. S. Proc. Vol. XL. p. 63, 1886. Galton himself writes (B. A. Report, 1885, p. 1211), "I may be permitted to say that I never felt such a glow of loyalty and respect towards the sovereignty and magnificent sway of mathematical analysis as when his answer reached me, confirming, by purely mathematical reasoning, my various and laborious statistical conclusions with far more minuteness than I had dared to hope, for the original data ran somewhat roughly, and I had to smooth them with tender caution."

Another feature of Galton's work at this time must be noted. He worked with the median instead of the mean, and he used probable errors or quartile values instead of standard deviations. Further, to obtain r, he somewhat laboriously expressed both variates in terms of their quartile deviations: thus r became the slope of his regression line. It was then determined by graphically fitting a good line, or from certain chosen arrays. Thus he worked with somewhat primitive statistical tools, and the wonder is that he achieved as much by their aid as he did.

Given A and B with regression r_{ab}, B and C with regression r_{bc}, Galton assumed $r_{ac}=r_{ab}\times r_{bc}$ to obtain his kinship relations. A nephew is the son of a brother. Hence

r for uncle and nephew $= r$ for brothers $\times r$ for father and son.

This of course is incorrect; it implies the vanishing of the corresponding partial correlation coefficient. Again, I think, his mid-parental correlation is not theoretically consonant with his parental correlations.

Another noteworthy point of the 1877 R. I. and the 1885 A. I. papers is the ample provision of mechanical models to illustrate by dropping shot or seeds the properties of bi-variate frequency. One wonders whether these elaborate quincunxes have been preserved and if so where they are at the present time. I reproduce one of them by permission from the *Journal of the Royal Institution*.
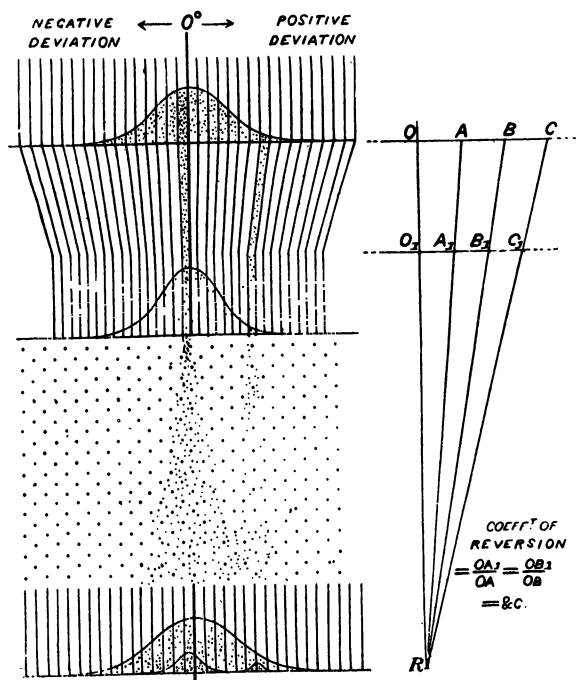
In 1886 Galton published a paper in the *Royal Society Proceedings** on "Family Likeness in Stature." This contains Hamilton Dickson's note and further data from Galton's *Family Records*.

* Vol. XL. pp. 42—66.

He gives under the headings " Mean regression w," and " Quartile of individual variability " the coefficients of correlation of various pairs of relations : Midparent and Offspring, Brothers, Fathers and Sons, Uncles and Nephews, Grandparents and Grandsons, but he does not realise that on the theory of multiple regression there are certain inconsistencies in his values. I do not think that there is much additional contribution to theory in this paper.

In 1888, however, Galton took a great step forward. He recognised that the whole statistical apparatus he had evolved for the treatment of the problem of heredity had a vastly wider significance. In a paper read to the Royal Society on December 5, 1888*, entitled " Correlations and their Measurement chiefly from Anthropometric Data," the term correlation first appears in our subject. Thus Galton's opening lines run :

" Co-relation or correlation of structure " is a phrase much used in biology, and not least in that branch of it which refers to heredity, and the idea is even more frequent than the phrase ; but I am not aware of any previous attempt to define it clearly, to trace its mode of action, or to show how to measure its degree.

Two variable organs are said to be correlated when the variation of the one is accompanied on the average by more or less variation of the other, and in the same direction (p. 135).

The last words seem to us now out of place, but Galton had not yet reached the idea of negative correlation. Also the balance is still swinging between 'co-relation' and 'correlation' although it has ultimately fallen to the more weighty word. How clearly Galton grasped the essence of correlation may be shown by the following sentences which might have saved many ingenious later investigators thinking they had made an important discovery. " It is easy to see that co-relation must be the consequence of the variations of the two organs being partly due to common causes. If they were wholly due to common causes, the co-relation would be perfect, as is approximately the case with the symmetrically disposed parts of the body. If they were in no respect due to common causes, the co-relation would be *nil*. Between these two extremes are an endless number of intermediate cases, and it will be shown how the closeness of co-relation in any particular case admits of being expressed by a single number " (p. 135). This single number it is needless to say is our present coefficient of correlation. Galton drops now the w of his 1886 work and returns to the r of his 1877 lecture, and the symbol r has remained to the present day.

Galton's process is the same as in the heredity problem. He used median and quartile and reduces the deviations to their respective quartiles as unit. He then smooths his means of arrays, draws a line to represent them and reads off its slope as r. He thus determines seven correlations which he here terms " indices of corre-lation†." They are between Stature and Cubit, Stature and Head Length, Stature and Middle Finger Length, Cubit and Middle Finger Length, Head Length and Head Breadth, Stature and Height of Knee, Cubit and Height of Knee. He fully

* R. S. Proc. Vol. XLV. pp. 135—145.

† On p. 143 r, the index of co-relation, is identified with the ' regression ' or ' reversion ' of Galton's earlier papers.

realises (1) that r is the same when obtained from either variate as 'relative,' (2) that r is always less than unity, (3) that r measures the closeness of co-relation, and (4) provides the regression line (p. 145).

On p. 144 the term "partial co-relation" is used but hardly in our modern sense although Galton is feeling his way towards multiple correlation. One problem he gives on p. 144 perhaps deserves mention, namely, if the n variates be expressed in terms of their quartiles then the quartile variability of their sum is $\sqrt{n}$ if they are independent and n if they be "rigidly and perfectly co-related." "The actual value would be almost always somewhere intermediate between these extremes, and would give the information that is wanted."

What Galton needs is the "multiple correlation coefficient," i.e.

$$\sqrt{1 - \frac{R}{R_{00}}},$$

but he is not yet on the right track for it.

In 1889 appeared Galton's book *Natural Inheritance* embodying most of the work we have discussed in the earlier memoirs of 1877 to 1888. Beyond this Galton did not carry the subject of correlation. He, in my opinion to-day, created it; there is nothing in the memoirs of Gauss or Bravais that really antedates his discoveries. They were dealing with the relatively narrow problem of determining the probable errors of indirectly observed quantities deduced from independent or uncorrelated directly observed quantities. The product-terms that arise in their investigations were expressed in terms of differential coefficients; they were not treated as a means of determining organic relationships between directly measured variates. Galton, starting from the organic relationship between parent and off-spring, passed to the idea of a coefficient measuring the correlation of all pairs of organs, and thence to the 'organic' relationship of all sorts of factors. If you think Galton did not appreciate the width of his new methods you must turn to the last paragraph of his *Introduction* to the *Natural Inheritance*.

"The conclusions cannot, however, be intelligently presented in an introductory chapter. They depend on ideas that must first be well comprehended, and which are now novel to the large majority of readers and unfamiliar to all. But those who care to brace themselves to a sustained effort, need not feel much regret that the road to be travelled over is indirect, and does not admit of being mapped beforehand in a way they can clearly understand. It is full of interest of its own. It familiarizes us with the measurement of variability, and with curious laws of chance *that apply to a vast diversity of social subjects*. This part of the inquiry may be said to run along a road on a high level that affords wide views in un-expected directions, and from which easy descents may be made to totally different goals to those we have now to reach. I have a great subject to write upon..." (p. 2).

Galton realised as fully as any of us now the width of application that would open up to the new calculus of correlation, and what easy descents there would be from the "high level road" to strange goals. His notebooks of this period show

that he was applying correlation and the regression line in a variety of ways thus to the relation between wing and tail length in birds, to fertility and to disease. His advance was chiefly hampered by the restriction of his data and the need for organised observers and computers.

The publication of *Natural Inheritance* provided Francis Galton with at least three recruits for the field of correlation: Weldon, Edgeworth and myself.

Weldon started in 1889 measuring the organs of shrimps at Plymouth and he was able to announce early in 1890—the letter is now in the glass case in our library here—the first correlation coefficients, or as he termed them "Galton Functions," between organs in shrimps. This was rapidly followed by his work on crabs, and the attempt to show that Galton functions were the same for all local races of the same species. In his first paper on shrimps Weldon writes[*]:

"In making this investigation I have had the great privilege of being constantly advised and helped in every possible way by Mr Galton. My ignorance of statistical methods was so great that without Mr Galton's constant help, given by letter at the expenditure of a very great amount of time and trouble, this paper would never have been written."

The pupil, however, was soon to outdistance the master in the width of his theoretical knowledge. A second paper on the shrimp followed in 1892[†], and this deals more closely with the correlations. Weldon now replaces medians by means in both marginal totals and arrays. He still uses probable errors or quartiles, and goes through the laborious process of reducing each deviation to the probable error. He uses r "in accordance with Mr Galton's notation" to represent the constant which measures the "degree of correlation" between organs. I think, but it is not quite clear, that he determined his probable error from the mean error, not from the quartile. He then determined r from each individual array and took the mean value of these r's as the true r. He accordingly introduced a greatly increased accuracy into the computing of correlation. He dealt with five local races of shrimps and found correlations for 22 pairs of organs. His regression diagram, p. 8, is still an admirable sample of this type of work. The correlations between post-spinous portion and total carapace length may be cited as illustrations of what Weldon and Galton were testing:

Plymouth (1000)	$r = 0.81$
Southport (800)	$r = 0.85$
Roscoff (500)	$r = 0.80$
Sheerness (380)	$r = 0.85$
Helder (300)	$r = 0.83$

The suggestion that r has the same value for all races of the same species was supposed to be confirmed by these results. We now realise that without a knowledge of the probable error of r, such a statement is illusory. But it was this very series of values which led to the investigation of the probable error of r and so to the extension of the correlational calculus.

[*] *R. S. Proc.* Vol. XLVII. p. 445, 1890.   [†] *R. S. Proc.* Vol. XLI. p. 2.

In this paper Weldon also published for the first time with due appreciation of their meaning *negative* correlation coefficients. In conclusion Weldon remarks: " A large series of such specific constants would give an altogether new kind of knowledge of the physiological connexion between the various organs of animals; while a study of those relations which remain constant through large groups of species would give an idea, attainable at present in no other way, of the functional correlations between various organs which have led to the establishment of the great subdivisions of the animal kingdom " (p. 11). In these lines we can read the starting-point of biometry as applied to other types of life than man.

I will not keep you longer over Weldon's contributions than to say that in 1893 appeared his third statistical paper* on " Correlated Variations in Naples and Plymouth shore Crabs." Weldon dealt with 23 pairs of organs in both Naples and Plymouth races. He proposes to call r "Galton's function"†. The paper shows that the 23 values of r at Plymouth and Naples are fairly close, but was again inconclusive because the significance of the differences could not be ascertained without a knowledge of the probable error of r.

We may next turn to Edgeworth, whose fundamental paper is that on " Correlated Averages" which appeared in the *Philosophical Magazine* of August, 1892, pp. 190—204. Edgeworth starts by referring to Galton's memoir of 1888 and Weldon's of 1892 on shrimps. He assumes for the probability that any particular values $x_1, x_2, \dots$ shall occur

$$\Pi = J e^{-R} dx_1 dx_2 dx_3 \dots,$$

where R is

$$= p_1 (x_1 - \bar{x}_1)^2 + p_2 (x_2 - \bar{x}_2)^2 + \dots + 2 q_{12} (x_1 - \bar{x}_1)(x_2 - \bar{x}_2) + \dots.$$

He does not justify this assumption but hopes to do so in a subsequent paper. He states that Galton by the happy device of measuring each deviation by the corresponding quartile had reduced in the case of two variates

$$R = \frac{x_1^2}{1 - \rho^2} - \frac{2 \rho x_1 x_2}{1 - \rho^2} + \frac{x_2^2}{1 - \rho^2}$$

to the discovery of a single constant ρ. This is hardly accurate; to reduce the expression R to the above it would be needful to measure not in terms of the quartile but of $\sqrt{2}$ S.D., which is I think sometimes termed the 'modulus‡. Edgeworth replaces Galton's "Index of Co-relation" and Weldon's "Galton's Function" by the term "coefficient of correlation." He then proceeds to weaken down Weldon's process of finding a mean r by suggesting that it will be adequate to find it by taking some of the ratios of 'subject' and mean 'relative' instead of the whole series. I look upon this suggestion as a distinctly retrogressive step.

---

* *R. S. Proc.* Vol. LIV. pp. 318—329.

† " The importance of this constant in all attempts to deal with the problems of animal variation was first pointed out by Mr Galton...and I would suggest that the constant whose changes he has investigated and whose importance he has indicated, may fitly be known as 'Galton's function,'" p. 325.

‡ Edgeworth appears to realise this on p. 194, but he did not go back and correct his statement of p. 190.

Our object should be to find the 'best value' for r and not how it may be most easily determined at the obvious cost of accuracy.

Although I am unable to follow some of Edgeworth's notation, he undoubtedly reaches something like the correct value for the correlation surface of three variates. In his notation

$$R = \Delta \left\{ (1 - \rho_{23}{}^2)\, x_1{}^2 + (1 - \rho_{31}{}^2)\, x_2{}^2 + (1 - \rho_{12}{}^2)\, x_3{}^2 \right.$$
$$\left. - 2x_1 x_2 \left(\rho_{12} - \rho_{13}\rho_{23} \right) - 2x_2 x_3 \left(\rho_{23} - \rho_{21}\rho_{31} \right) - 2x_3 x_1 \left(\rho_{31} - \rho_{32}\rho_{12} \right) \right\},$$

where
$$\Delta^{-1} = \left\{ (1 - \rho_{13}{}^2)(1 - \rho_{12}{}^2) - (\rho_{12}\rho_{23} - \rho_{13})^2 \right\}$$

according to him, but the factor $1 - \rho_{13}{}^2$ should be replaced by $1 - \rho_{23}{}^2$ I think.

Even with this change I am unable to reach the value he gives in Galton's case of
$$\rho_{12} = \cdot 8, \quad \rho_{13} = \cdot 9, \quad \rho_{23} = \cdot 8,$$
for these seem to give
$$\Delta = 9 \cdot 9305,$$
whereas Edgeworth's value is $16 \cdot 129$.

I do not grasp his equation at the foot of p. 196, nor follow how the equation at the top of p. 197 follows from it.

Lastly we come to p. 201 where we should expect to find the general regression equation. Edgeworth tells us that the reasoning is quite general and accordingly we ought to anticipate that his results whatever they are would give our accepted values
$$p_{ss} = \frac{R_{ss}}{R} \quad \text{and} \quad p_{ss'} = \frac{R_{ss'}}{R},$$

where R is the determinant of the correlations. Instead of this simple rule Edgeworth sums up in the middle of the page with equations
$$\Delta \rho_{13} = + \Delta^3 \left(\rho_{24}\rho_{31}\rho_{42} \right),$$
$$\Delta \rho_{14} = - \Delta^3 \left(\rho_{21}\rho_{32}\rho_{43} \right),$$
$$\vdots \qquad \vdots \qquad \vdots$$

There is no explanation of what the symbolism means, and I cannot interpret it, so as to provide the requisite generalisation for n variates.

On the other hand while unable to interpret Edgeworth's general analysis I agree in the case of four variates with the only two terms I have taken the trouble to test in his numerical illustration of this case,
$$\rho_{12} = \frac{1}{\sqrt{2}}, \quad \rho_{13} = \sqrt{\frac{1}{3}}, \quad \rho_{14} = \sqrt{\frac{1}{4}}, \quad \rho_{23} = \sqrt{\frac{2}{3}}, \quad \rho_{24} = \sqrt{\frac{2}{4}}, \quad \rho_{34} = \sqrt{\frac{3}{4}},$$
namely 2 as the coefficient of $x_1{}^2$ and $-2\sqrt{2}$ as the coefficient of $x_1 x_2$, my R being $\frac{1}{24}$, R_{11} being $\frac{1}{12}$ and R_{12} being $\frac{-\sqrt{2}}{24}$. Edgeworth does not provide the needful external constant of the frequency surface, i.e.
$$\frac{N}{(2\pi)^n \sigma_1 \sigma_2 \ldots \sigma_n} \frac{1}{\sqrt{R}}.$$

I should sum up Edgeworth's work of 1892 by saying that he left the problem of multiple correlation at least in a very incomplete state. He probably knew what he was seeking himself, but he did not give the requisite attention to the wording or printing of his memoir to make it clear to others, and accordingly in looking back at the matter now I am very doubtful whether in 1895 I ought to have called the problem of multiple correlation, "Edgeworth's Problem." He certainly did not put the answer to it in a form in which the statistician with a customary amount of mathematical training could determine the form of the surface for n variates, as soon as their S.D.'s and correlations had been calculated. I think I am justified in saying this for I have not to my recollection come across any treatment of multiple correlation which starts from Edgeworth's paper or uses his notation.

It will be seen from what has gone before that in 1892 the next steps to be taken were clearly indicated. They were, I think,

(*a*) The abolition of the median and quartile processes as too inexact for accurate statistics.

(*b*) The replacement of the laborious processes of dividing by the quartiles and averaging the deduced values of r, by a direct and if possible 'best' method of finding r.

(*c*) The determination of the probable errors of r as found by the 'best' and other methods.

(*d*) The expression of the multiple correlation surface in an adequate and simple form.

These problems were solved by Dr Sheppard or myself before the end of 1897.

Closely associated with these problems arose the question of generalising correlation. Why should the distribution be Gaussian, why should the regression curve be linear?

As early as 1893 I dealt with quite a number of correlation tables for long series and was able to demonstrate

(i) by applying Galton's process of drawing contours of equal frequency that most smooth and definite systems of contours can arise from long series, obviously mathematical families of curves, which are (*a*) ovaloid, not ellipsoid, and (*b*) which do not possess—like the normal surface contours—more than one axis of symmetry,

(ii) that regression curves can be quite smooth mathematical curves differing widely from straight lines,

(iii) that in cases wherein (i) and (ii) hold, homoscedasticity is not the rule.

I obtained differential equations to such systems, but for more than 25 years while often returning to them, have failed to obtain their integration.

This seems to me the desideratum of the theory of correlation at the present time: the discovery of an appropriate system of surfaces, which will give bi-variate skew frequency. We want to free ourselves from the limitations of the normal surface, as we have from the normal curve of errors.

As early as 1897 Mr G. U. Yule*, then my assistant, made an attempt in this direction. He fitted a line or plane by the method of least squares to a swarm of points, and this has been extended later to n-variates and is one of the best ways of reaching the multiple regression equations and the coefficient of multiple correlation†. Now while these methods are convenient or utile, we may gravely doubt whether they are more accurate *theoretically* than the assumption of a normal distribution. Are we not making a fetish of the method of least squares as others made a fetish of the normal distribution? For how shall we determine that we are getting a 'best fit' to our system by the method of least squares?

If we are fitting a curve $y = f(x, c_1, c_2, c_3)$
to a series of observations we can only assert that least square methods are theoretically accurate on the assumption that our observations of y for a given x obey the normal law of errors. That is the proof which Gauss gave of his method and I personally know no other. *Theoretically* therefore to have justification for using the method of least squares to fit a line or plane to a swarm of points we must assume the arrays to follow a normal distribution. If they do not, we may defend least squares as likely to give a fairly good result but we cannot demonstrate its accuracy. Hence in disregarding normal distributions and claiming great generality for our correlation by merely using the principle of least squares, we are really depriving that principle of the basis of its theoretical accuracy, and the apparent generalisation has been gained merely at the expense of theoretical validity. Take other distributions of deviations for the arrays and the method of least squares is not the one which will naturally arise from making the combined probability a maximum. It is by no means clear therefore that Mr Yule's generalisation indicates the real line of future advance.

I have endeavoured to indicate in this paper the broad outline of the early history of correlation which has now a most extensive literature. It is a long step from Francis Galton's 'reversion' in sweet pea seeds to the full theory of multiple correlation, which we now know to be identical with the spherical trigonometry of high-dimensioned space, the total correlation coefficients being the cosines of the edges of the polyhedra and the partial correlation coefficients the cosines of the polyhedral angles. But to find the correlation of the health of a child with the number of people per room while you render neutral its age, the health of its parents, the wages of its father, and the habits of its mother, is no less vital a problem than Galton's correlation of character in parent and offspring. It requires indeed more mathematics, but the mathematics are not there for the joy of the analyst but because they are essential to the solution. It is the transition from the mill as pestle and mortar to the mill with steam driven grain crushing steel rollers. But the inventor of milling was the person who bruised grain between two stones, and Galton was the man who discovered the highway across this new country with what he aptly terms "its easy descents to different goals."

* *Journal of Royal Statistical Society*, Vol. LX. Part iv, p. 3.

† *Biometrika*, Vol. VIII. p. 438. The method adopted in the paper is not that of fitting a generalised plane by least squares, but of making a generalised correlation coefficient take its maximum value. It appeals only to the rules of the differential calculus and not to the method of least squares, or indirectly to Gauss' law of errors.