14

ON THE "PROBABLE ERROR" OF A COEFFICIENT OF CORRELATION DEDUCED FROM A SMALL SAMPLE

Author's Note (CMS 1.2a)

This is the second of three papers dealing with the sampling errors of correlation coefficients covering the cases (*i*) "The frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, Vol. 10, pp. 507–521, 1915.

Here the method of defining the sample by the coordinates of a point in Euclidean hyperspace was introduced, and it was shown that the exact sampling distribution could be obtained. The practical application of these results appears in the second paper in 1921, here referred to. It was concerned primarily to show that the sampling distribution was different for intraclass and interclass correlations, and to give the exact solution for the former, comparable with that given for the latter in 1915. The special simplicity of the solution in the intraclass case was one of the foundations of the recognition of the z-distribution, and the analysis of variance, in terms of which it would now be treated.

The third of these papers appeared in 1924, also in *Metron:* (*iii*) "The distribution of the partial correlation coefficient," *Metron*, Vol. 3, pp. 329–332, 1924.

It shows that the effect of the elimination of variates by partial correlation is simply to reduce the effective size of the sample by unity for each independent variate eliminated.

This group of three papers is part of a larger series appearing from 1915 to 1928, in which exact solutions were found for a variety of problems of distribution, and the corresponding tests of significance developed. Although many of these problems had been approached using statistics called correlation coefficients (e.g., Biserial r, Biserial η, etc.), yet it appears that the term was widely misapplied, and the problems themselves are now simply treated as comparisons of means, tests of heterogeneity, or regression problems.

In the final section this paper contains a discussion of the bearing of the new exact solutions of distribution problems on the nature of

Metron, 1: 3-32, (1921).

inductive inference. This is of interest for comparison with the paper on "Inverse probability," published 1930, nine years later. In 1930 the notion of fiducial probability is first introduced using as example the distribution found in this paper. In view of this later development the statement, "We can know nothing of the probability of hypotheses or hypothetical quantities," is seen to be hasty and erroneous, in the light of the different type of argument later developed. It will be understood, however, as referring only to the Bayesian probabilities *a posteriori*.

Introduction.

The problems of theoretical statistics fall into two main classes: *a)* To discover what quantities are required for the adequate description of a population, which involves the discussion of mathematical expressions by which frequency distributions may be represented: *b)* To determine how much information, and of what kind, respecting these population-values is afforded by a random sample, or series of random samples.

Problems of the second class require for their adequate discussion a knowledge of the distribution in random samples of specified size, of the statistical derivates used to estimate or evaluate the population-characters. Thus in calculating the correlation from a sample we are making an estimate of the correlation in a theoretical infinite population from which the sample is drawn. We wish to make the best possible estimate and to know as accurately as possible how far the estimate may be relied upon. To this end we seek to know the distribution of the values obtained when samples are drawn from an infinite population, whatever the value of the correlation in such population may be.

Some years ago, the writer applied a novel method of geometrical representation to problems of random sampling which had excited attention in the pages of *Biometrika*. He was thereby enabled to give the exact form of the curve of distribution in random samples of the coefficient of correlation when the latter was calculated in the ordinary way. His formulae emphasised the fact that in the neighbourhood of ± 1, the curves become extremely skew, even for large samples, and change their form so rapidly that the ordinary statement of the « probable error » is practically valueless. It was accordingly suggested that the variable r was unsuitable for expressing the accuracy of an observed correlation in these regions but that, by a simple transformation, a variable might be obtained the sampling curves of which are practically normal and of constant standard deviation.

Since that time, the curves of random sampling which arise from other methods of calculation have been examined, viz., those appropriate to intra-class correlations, represented by symmetrical tables. It was found that these curves while different from those previously obtained, could be rendered approximately normal by a similar transformation and that the discrepancies between the sampling curves derived by the two methods were greater than the departure of either from the normal form. The advantage of the transformation therefore became more apparent, it not only enabled an intelligible statement of the « probable error » to be made, even for high correlations, but permitted, and in the simplest possible manner, the making of allowance for the method of calculation.

In the former paper it was found, by applying a method previously developed, that the « most likely » value of the correlation of the population was, numerically, slightly smaller than that of the sample. This conclusion was adversely criticised in *Biometrika*, apparently on the incorrect assumption that I had deduced it from Bayes theorem. It will be shown in this paper that when the sampling curves are rendered approximately normal, the correction I had proposed is equal to the distance between the population-value and the mid-point of the sampling curve and is accordingly no more than the correction of a constant bias introduced by the method of calculation. No assumption as to *a priori* probability is involved.

The exact forms of a variety of frequency curves serve as an adequate basis for discussing methods by which a theoretical or ideal quantity, such as the population-value of the correlation coefficient, may be estimated from a sample. The attempt made by Bayes, upon which the determination of « inverse probabilities » rests, admittedly depended upon an arbitrary assumption, so that the whole method has been widely discredited; yet the very concept of a frequency curve, or surface, implies an infinite ideal population, the properties of which can only be estimated from samples. In my opinion, two radically distinct concepts have been confused under the name of « probability » and only by sharply distinguishing these can we state accurately what information a sample does give us respecting the population from which it is drawn.

1. *The curve of random sampling for « intraclass » correlations.*

In the calculation of fraternal correlations, and others of a like nature, in which the mean and standard deviation are presumed to be the same for both variables, it is usual to obtain the common mean and standard deviation from the whole of the observations. This procedure may be expected to give more accurate results than the use of separate means and standard deviations, when no distinction is made as to which of one pair of observations corresponds to each of another pair; this expectation is justified in that the probable error of such correlations is somewhat less than that of a similar correlation drawn from the same number of pairs of independent quantities, but the curve of random sampling is also affected in other ways.

If $x_1, x_1', x_2, x_2', \dots x_n, x_n'$ be n pairs of observations of this kind, then

$$2n\,\bar{x} = S\,(x + x')$$
$$2n\mu^2 = S\,\{(x - \bar{x})^2 + (x' - \bar{x})^2\}$$
$$n\mu^2 r = S\,(x - \bar{x})(x' - \bar{x})$$

are the equations which determine the statistics $\bar{x}$, μ and r. Correlations found in this way have been termed intraclass correlations. In other cases of intraclass correlations the observations may occur in sets of 3, 4 etc., but these cases will not be considered in detail.

The exact form of the curve of random sampling may be obtained by the method previously used (Fisher 1915) to determine that of correlations of the ordinary type, in which the means and standard deviations of the two variates are calculated separately. In this method a sample is represented by a point in generalised space, the separate measurements being the coordinates of the point.

If m, σ and ρ be the values of the mean, standard deviation and correlation of the hypothetical infinite Gaussian population from which the sample is drawn, then the frequency with which any pair of values fall into specified infinitesimal ranges, is

$$df = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2\sigma^2(1-\rho^2)}\{(x-m)^2 - 2\rho(x-m)(x'-m) + (x'-m)^2\}}\, dx\,dx';$$

hence the chance that all the observations of the sample have given values is

$$\text{(I)}\quad \frac{1}{(2\pi\sigma^2\sqrt{1-\rho^2})^n} e^{-\frac{1}{2\sigma^2(1-\rho^2)} S\{(x-m)^2-2\rho(x-m)(x'-m)+(x'-m)^2\}} dx_1 dx'_1 dx_2 dx'_2 \ldots dx_n dx'_n.$$

The representative point lies on a plane space, specified by $\bar{x}$, and on a generalised sphere of $(2n-1)$ dimensions lying in that space, $\mu\sqrt{2n}$ being the radius of that sphere. Its position upon the sphere is further restricted to a region at a fixed angular distance, θ, from the space of $(n-1)$ dimensions, specified by the equations

$$x_1 = x'_1,\ x_2 = x'_2, \ldots\ldots, x_n = x'_n.$$

This angular distance depends upon r in such a way that

$$r = \cos 2\theta$$

Hence the volume element $dx_1\, dx'_1 \ldots . dx_n\, dx'_n$ may, neglecting a constant multiplier, be replaced by

$$d\bar{x}\,\mu^{2n-2}\, d\mu \sin^{n-1}\theta \cos^{n-2}\theta\, d\theta,$$

or by

$$d\bar{x}\,\mu^{2n-2}\, d\mu\, (1-r)^{\frac{n-2}{2}} (1+r)^{\frac{n-3}{2}} dr.$$

The frequency element, simplified by substituting the derivates $\bar{x}$, μ and r for the coordinates, and by ignoring constant factors, may now be written

$$\text{(II)}\quad e^{-\frac{2n}{2\sigma^2(1-\rho^2)}\{(\bar{x}-m)^2(1-\rho)+\mu^2(1-\rho r)\}} d\bar{x}\,\mu^{2n-2} d\mu (1-r)^{\frac{n-2}{2}} (1+r)^{\frac{n-3}{2}} dr$$

an expression which specifies the relative frequencies with which any assigned values of $\bar{x}$, μ and r will occur in the process of random sampling.

The factor

$$e^{\frac{-2n}{2\sigma^2(1+\rho)}(\bar{x}-m)^2} d\bar{x}$$

involves only the variable $\bar{x}$, and shows it to be normally distributed about its mean, independently of the other derivates. Its accuracy depends upon ρ, and is increased without limit as ρ approaches -1, being reduced, however, when ρ is positive.

 7

Since we are concerned merely with the distribution of r, the remaining factor,

$$\text{(III)}\qquad e^{-\frac{n}{\sigma^2}\cdot\frac{1-\rho r}{1-\rho^2}\cdot\mu^2}\,\mu^{2n-2}\,d\mu.\,(1-r)^{\frac{n-2}{2}}(1+r)^{\frac{n-3}{2}}\,dr,$$

must be integrated with respect to μ, from 0 to ∞. It is easy to see that the only factors of the integral which involve r, must be

$$\text{(IV}a\text{)}\qquad (1-\rho r)^{-\left(n-\frac{1}{2}\right)}(1-r)^{\frac{n-2}{2}}(1+r)^{\frac{n-3}{2}}\,dr$$

an expression which gives the relative frequency of occurrence of different values of r.

For sets of 3 the corresponding expression is

$$\text{(IV}b\text{)}\qquad (1+\rho-2\rho r)^{-\frac{3n-1}{2}}(1-r)^{\frac{2n-2}{2}}\left(\frac{1}{2}+r\right)^{\frac{n-3}{2}}\,dr$$

in which it will be noticed that r cannot fall below $-\frac{1}{2}$. Indeed in general for sets of s it cannot fall below $-\frac{1}{s-1}$, although there is no restriction upon positive correlations. The expression for the distribution of the correlation within groups of s being

$$\text{(IV}c\text{)}\qquad (1+\overline{s-2}\,\rho-\overline{s-1}\,\rho r)^{-\frac{sn-1}{2}}(1-r)^{\frac{\overline{s-1}\,n-2}{2}}\left(\frac{1}{s-1}+r\right)^{\frac{n-3}{2}}\,dr$$

2. *The transformation of the curves of random sampling of correlations within classes of 2.*

It was noticed in a previous paper (FISHER, 1915) that the curve of sampling of the correlation coefficient becomes extremely skew towards the ends of its range, and in these regions changes its form rapidly as ρ is changed. It was suggested that this group of curves could be reduced both to approximate normality and to approximate constancy of the probable error, by the transformation

$$\text{(V)}\qquad \left\{\begin{array}{l}\rho=\tanh\zeta\\ r=\tanh z\end{array}\right.$$

This transformation has recently been applied (FISHER, 1919) to the measure of resemblance employed by THORNDIKE (1905) in his investigations of twins. If x and y are the deviations from their means of the measurements of two twins, THORNDIKE took

$$r = \frac{2xy}{x^2 + y^2}$$

as a measure of resemblance. If ρ is the true value of the correlation the curve of random sampling of r is

$$\frac{\sqrt{1-\rho^2}}{\pi} \cdot \frac{dr}{(1-\rho r)\sqrt{1-r^2}}$$

a curve not unlike that of the correlation derived from 3 pairs of observations, having infinite values at the extremes, and an antimode. This curve changes its form rapidly as ρ is changed. On applying the transformation (V) the curve becomes

$$\frac{1}{\pi} \operatorname{sech}(z-\zeta)\, dz,$$

a symmetrical curve centred at ζ. This curve is of absolutely invariant form for different values of ζ. For large values of $|z-\zeta|$ the curve falls off exponentially; it is markedly leptokurtic, having

$$\beta_2 = 5.$$

If we apply this transformation to expression (IVa) we obtain

$$\text{(VI)} \qquad \frac{\overline{n-\frac{3}{2}}\,!\,^{(*)}}{\overline{n-2}\,!\,\sqrt{2\pi}} \operatorname{sech}^{n-\frac{1}{2}}(z-\zeta) \,.\, e^{-\frac{1}{2}(z-\zeta)}\, dz$$

after inserting the requisite constant factor. This curve in also absolutely constant in form for all values of ζ, the effect of changing ζ being merely to shift it bodily. The curve also tends to normality as n increases with rapidity sufficient to justify the use of the probable error as an adequate representation of its distribution due to sampling.

(*) The symbol $x!$ is here used as equivalent to $\Gamma(x+1)$ whether x be an integer or not.

The approach to normality may be well shown by PEARSON's method of expansion (*Cooperative Study*, 1916). Writing x for $z-\zeta$, the ordinate may be expressed as

$$y=\frac{\overline{n-\frac{3}{2}}!}{\overline{n-2}!\,\sqrt{2\pi}}\,e^{-\frac{n-1}{2}x^2}\left\{1+\frac{n-1}{12}x^4-\left(\frac{n-1}{45}x^6-\frac{\overline{n-1}^2}{288}x^8\right)+\dots\right\}$$
$$\left\{1-\frac{1}{2}x-\frac{x^2}{8}+\frac{5x^3}{48}+\frac{17x^4}{384}\dots\right\}$$

whence may be obtained the moments about $x=0$

$$\bar{z}-\zeta=\mu_1'=-\frac{1}{2\overline{n-1}}\left\{1+\frac{1}{2\overline{n-1}}\dots\right\}$$
$$\mu_2'=\frac{1}{n-1}\left\{1+\frac{3}{4\overline{n-1}}+\frac{5}{12\,\overline{n-1}^2}\dots\right\}$$
$$\mu_3'=-\frac{3}{2\overline{n-1}^2}\left\{1+\frac{7}{4\overline{n-1}}\dots\right\}$$
$$\mu_4'=\frac{1}{(n-1)^2}\left\{3+\frac{13}{2\overline{n-1}}+\frac{145}{16\,\overline{n-1}^2}\dots\right\}$$

The distance of the mean, $\bar{z}$, from the population value ζ, though of higher order than the standard deviation is yet of lower order than the distance of the mean from the mode. For the mode is where

$$x=-\tanh^{-1}\frac{1}{2n-1};$$

that is

$$\check{z}-\zeta=-\frac{1}{2n-1}-\frac{1}{3(2n-1)^3}\dots\dots$$
$$=-\frac{1}{2\overline{n-1}}\left\{1-\frac{1}{2\overline{n-1}}\dots\right\}$$

Evidently the mean, mode and median of this curve approach each other more rapidly than they approach the population value. The optimum value of ζ for a given z, that is to say, that value of ζ which gives z with the greatest frequency, is given by the equation

$$\hat{\zeta}-z=\tanh^{-1}\frac{1}{2n-1}$$

The curve being invariant in form, it follows that the modal value of z, is also that value for which ζ is an optimum. Now the median observation for a given population value, and the optimum population value for a given observation are, unlike the mode and the mean, unchanged for all transformations; they differ approximately by

$$\frac{1}{3(n-1)^2}$$

in the scale of z, and a transformation which renders the sampling curve invariant is bound to bring the mean and mode into agreement to the same order. By virtue of this agreement adequate allowance may be made for the bias towards negative correlations displayed by small samples. Referred to the mean, the moments are

$$\mu_2 = \frac{1}{n-1}\left\{1 + \frac{1}{2\overline{n-1}} + \frac{1}{6\,\overline{n-1}^2}\cdots\right\}$$

$$\mu_3 = -\frac{1}{(n-1)^3}\cdots\cdots$$

$$\mu_4 = \frac{1}{(n-1)^2}\left\{3 + \frac{5}{(n-1)} + \frac{19}{4\,\overline{n-1}^2}\cdots\right\}$$

so that

$$\beta_1 = \frac{1}{(n-1)^3}\cdots\cdots$$

$$\beta_2 = 3 + \frac{2}{n-1} + \frac{1}{(n-1)^2}\cdots\cdots$$

The curve therefore becomes symmetrical with extreme rapidity, remains slightly leptokurtic, but is sensibly normal for all but the smallest possible samples.

The standard deviation is independent of ρ, and very nearly agrees with the formula

$$\frac{1}{\sqrt{n-{}^3/_2}}$$

the probable error may therefore be read off as χ_1, in Table V of *Tables for Statisticians* (PEARSON, 1914) interpolating for the half integer.

It is easy to assign a reason for the negative bias of correlations of this type; the mean value of the variate deviations are not in these cases reduced independently to zero, but are equal and of opposite sign, thus automatically introducing a small negative term into the product moment. The correction needed, being of higher order than the probable error, does not sensibly improve the accuracy of a single determination, but may become of importance in comparing averages based on samples of different sizes, or calculated by different methods, as are parental and fraternal correlations.

Since the method I should adopt in obtaining the population value and its probable error from an observation of a small sample differs materially from that developed in the *Cooperative Study* (1916) a number of brief examples may serve to illustrate the application of the principles here laid down.

Ex. I. — A correlation 0.6000 is derived from 13 « fraternal » pairs of observations, find the population value with its probable error.

Using the transformation $r = \tanh z$, we have

	r	z
Calculated value	+ 0.6000	+ 0.6930
Correction $\frac{1}{2n-1}$		+ 0.0417
Population value	+ 0.6259	+ 0.7347
Probable error of z for given ζ . .		$\pm$ 0.1989
Lower quartile	+ 0.4898	+ 0.5358
Higher quartile	+ 0.7323	+ 0.9336

The quartile distances are 0.1361 and 0.1064; their mean 0.1213 is in fair agreement with 0.1238 derived from the formula $0.67449\frac{1-r^2}{\sqrt{n-1}}$. The difference between them reveals the skewness of the curve and the impossibility of judging accurately, from the values of r alone, the probability of a difference of observations of two of three times the probable error.

Ex. II. — Estimate the probability that the above observation corresponds to a true value (1) 0.3000, (2) 0.9500.

	r	z	Differences	Standard error	Ratio	P
Hypothesis I.	0.3000	0.3095	0.4252	0.2887	1.47	0.142
Sample	0.6259	0.7347				
Hypothesis II.	0.9500	1.8318	1.0971	0.2887	3.80	0.00014

Thus the value 0.95 although, when measured on the r scale, nearer to the population value derived from the sample, is roughly 1000 times less likely than 0.30. The value of P cannot be taken very accurately from the probability integral of the normal curve, since the true curve in z is slightly leptokurtic ($\beta_2 = 3.17$) and extreme deviations are therefore more common than they would be if the distribution were strictly normal. However, the increase in accuracy in expressing the probable error in terms of z, the curve of which is invariant in form and approximately normal, instead of terms of r, the curve of which may be very skew and variable in form, even for high values of n, is sufficiently striking.

It would of course be possible to render the statement of the probable error of r less misleading, by writing $\begin{smallmatrix}+0.1064\\-0.1361\end{smallmatrix}$ instead of ± 0.1213; that is by stating the actual quartile distances. Such a change, though certainly more accurate, and giving at any rate a danger signal as to the nature of the distribution, does not describe it effectively. Although two numbers are given, they contain less information than the single probable error when the distribution is normal; nor can they be used with the same facility for those crucial comparisons, for which probable errors are necessary.

3. *The corresponding transformation for simple interclass correlations.*

If the data, as is commonly the case, consist of pairs of values of two variables x and y, which are not presumed to have the same mean or standard deviation, then the distribution of the correlation coefficient r, has been shown (Fisher 1915) to be

$$df = \frac{n-2}{\pi}(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}\int_0^\infty \frac{dt}{(\cosh t - \rho r)^{n-1}} \cdot dr$$

The application of the transformation (V) to this series of curves, though not leading to so simple a curve as that indicated by expression (VI) will be found to be attended by the same practical advantages. The element of frequency becomes

$$df = \frac{n-2}{\pi}\operatorname{sech}^{n-1}\zeta \operatorname{sech}^{n-2} z \int_0^\infty \frac{dt}{(\cosh t - \rho r)^{n-1}}\, dz$$

which may be expanded in the same manner as before, writing x for $z-\zeta$; then

$$*\quad y=\frac{n-2}{\sqrt{2\pi\,\overline{n-1}}}e^{-\frac{n-1}{2}x^2}\Big\{1+\frac{1}{2}\rho x+\Big(\frac{2+\rho}{8\,\overline{n-1}}+\frac{4-\rho^2}{8}x^2+\frac{n-1}{12}x^4\Big)$$

$$+\rho x\Big(\frac{4-\rho^2}{16\,\overline{n-1}}+\frac{4+3\rho^2}{48}x^2+\frac{n-1}{24}x^4\Big)$$

$$+\Big(\frac{4+12\rho^2+9\rho^4}{128\,\overline{n-1}^2}+\frac{8-2\rho^2+3\rho^4}{64\,\overline{n-1}}x^2+\frac{8+4\rho^2-5\rho^4}{128}x^4+\frac{28-15\rho^2}{1440}x^6\,\overline{n-1}+\frac{\overline{n-1}^2}{288}x^8\Big)+\ldots\Big\}$$

The form of the transformed curve involves ρ and is thus not absolutely constant in shape. Taking moments about, $x=0$, we have,

$$\bar{z}-\zeta=\mu_1'=\frac{\rho}{2(n-1)}\Big\{1+\frac{1+\rho^2}{8\,\overline{n-1}}\ldots\Big\}$$

$$\mu_2'=\frac{1}{n-1}\Big\{1+\frac{8-\rho^2}{4\,\overline{n-1}}+\frac{88-9\rho^2-9\rho^4}{24\,\overline{n-1}^2}\ldots\Big\}$$

$$\mu_3'=\frac{3\rho}{2(n-1)^2}\Big\{1+\frac{13+2\rho^2}{8\,\overline{n-1}}\ldots\Big\}$$

$$\mu_4'=\frac{1}{(n-1)^2}\Big\{3+\frac{28-3\rho^2}{2\,\overline{n-1}}+\frac{736-84\rho^2-51\rho^4}{16\,\overline{n-1}^2}\ldots\Big\}$$

As with the previous case, the median value of z, differs from that for which ζ is optimum, only by terms in $\frac{1}{(n-1)^2}$ and higher orders; and this transformation brings the mean and mode into similar agreement with those invariant quantities. For the optimum value of ρ has been shown (*Cooperative Study*, 1916) to be

$$\hat{\rho}=r-\frac{r(1-r^2)}{2\,\overline{n-1}}\Big(1-\frac{1-5r^2}{4\,\overline{n-1}}\Big),$$

and the median value of r is, to the same approximation,

$$\rho+\frac{\rho(1-\rho^2)}{2\,\overline{n-1}}\Big(1+\frac{9-14\rho^2}{6\,\overline{n-1}}\Big);$$

both of which agree, as far as the term $\frac{1}{n-1}$ with the correction

$$-\frac{r}{2\,\overline{n-1}}$$

required to bring the mean of the z curve into agreement with the population value.

* For $2+\rho$, read $2+\rho^2$.

 14

As before this correction is of a higher order than the standard deviation, and therefore does not add significantly to the accuracy of a single determination; its function is to allow for the size of the sample, and for the method of calculation, in cases where accurate comparisons are required between correlations or averages of correlations.

This correction unlike that obtained in section 2, changes its sign with r; it always reduces the observed value of r numerically, being positive when r is negative.

Taking now moments about the mean

$$\mu_2 = \frac{1}{n-1}\left\{1 + \frac{4-\rho^2}{2\,\overline{n-1}} + \frac{176 - 21\,\rho^2 - 21\,\rho^4}{48\,\overline{n-1}^2} \cdots\right\}$$

$$\mu_3 = \frac{\rho\left(\rho^2 - \frac{9}{16}\right)}{(n-1)^3} \cdots\cdots$$

$$\mu_4 = \frac{1}{\overline{n-1}^2}\left\{3 + \frac{224 - 48\,\rho^2 - 3\,\rho^4}{16\,\overline{n-1}} + \frac{1472 - 228\,\rho^2 - 141\,\rho^4 - 3\,\rho^6}{32\,\overline{n-1}^2} \cdots\right\}$$

giving

$$\beta_1 = \frac{\rho^2}{\overline{n-1}^3}\left(\rho^2 - \frac{9}{16}\right)^2 \cdots$$

$$\beta_2 = 3 + \frac{32 - 3\,\rho^4}{16\,\overline{n-1}} + \frac{128 + 112\,\rho^2 - 57\,\rho^4 - 9\,\rho^6}{32\,\overline{n-1}^2} \cdots$$

The curves therefore tend to normality as rapidly as those for correlations of the fraternal type. In addition to containing the divisor $(n-1)^3$, β_1 vanishes absolutely at the origin, and at two points which, as n increases, approach the limits $\pm$ 0.75. β_2 approaches its normal value at approximately the same rate as before. The value of ρ is seen to have very little influence on these curves (See Fig. 2); the weight of an observation is increased to a trifling extent from $n-3$ to $n - 2\,^1/_2$ as ρ passes from 0 to $\pm$ 1; except for the smallest samples β_1 is to be neglected for all values of ρ, and when n is small enough for the leptokurtosis of the curves to be appreciable, it is reduced by less than 10 % by the highest possible value of ρ.

When expressed in terms of z, the curve of random sampling is therefore sufficiently normal and constant in deviation to be adequately represented by a probable error. This may be obtained from the same table as before entered with the value $n-3$.

The increased accuracy of assuming a common mean and standard deviation for the variables, when this assumption is justified, is thus equivalent to a gain of from 1 ½ pairs to 1 pair of observations or from 3 to 2 measurements according to the value of ρ.

The application of these methods is illustrated by the following examples.

Ex. III. — In a sample of 25 pairs only of parent and child the correlation for a certain character was found to be 0.6000. What is the most reasonable value to give ρ in the sampled population, and what is its probable error?

Using the transformation $r = \tanh z$ we have

	r	z
Calculated value	0.6000	0.6930
Correction.		—0.0125
Population value	0.5918	0.6805
Probable error		±0.1438
Lower quartile.	0.4905	0.5367
Higher quartile.	0.6774	0.8243

This example was taken by the writers of the *Cooperative Study* to illustrate the supposed shortcomings of my formula equivalent to that used above, for the optimum value of the correlation. Their comments upon my methods imply such a serious misunderstanding of my meaning that a brief reply is necessary. The following passage well illustrates their attitude (p. 358).

« If we distributed our ignorance equally the result would be that stated on p. 357, i. e.

$$0.59194$$

But, in applying Bayes' Theorem to this case, to what result of experience do we appeal? Clearly the only result of experience by which we could justify this « equal distribution of ignorance » would be the accumulative experience that in past series the correlation of parent and child had taken with equal frequency of occurrence every value from —1 to +1. To appeal to such a result is absurd; Bayes' Theorem ought only to be used where we have in past experience, as for example in the case of probabilities and other statistical ratios, met with every admissible value with roughly equal frequency. There is no such experience in this case. On the contrary the mean value of ρ for very long series of frequencies of 1000 and upwards is known to be + 0.46 and the range is hardly more than 0.40 to 0.52 ». Applying some formulae the value

$$0.46225$$

is finally obtained, which as the authors justly remark, is « a totally different 'most likely value' from that obtained by 'equally distributing' our ignorance ».

From this passage a reader, who did not refer to my paper, which had appeared in the previous year, and to which the *Cooperative Study* was called an « Appendix », might imagine that I had used BOOLE's ironical phrase, « equal distribution of ignorance », and that I had appealed to « BAYES' theorem ». I must therefore state that I did neither. What is more important is that what I previously termed the 'most likely value', which I now, for greater precision, term the 'optimum' value of ρ, for a given observed r, is merely that value of ρ for which the observed r occurs with greatest frequency; it is obtained by making a maximum df, the frequency of occurrence of the observed value (FISHER 1915, p. 520).

It therefore involves no assumption whatsoever as to the probable distribution of ρ. The writers of the *Cooperative Study* appear to suppose that it depends upon the assumption that in past experience equal intervals, dr, of the range of possible correlations have received equal numbers of observed parental correlations. As a matter of fact the above analysis, in which we have used z instead of r, leads to exactly the same value of the optimum. Does the validity of the optimum therefore depend upon equal numbers of parental correlations having occurred in equal intervals dz? If so, it should be noted that this is inconsistant with an equal distribution in the scale of r, for

$$dz = \frac{dr}{1 - r^2}$$

As a matter of fact, as I pointed out in 1912 (FISHER, 1912) the optimum is obtained by a criterion which is absolutely independent of any assumption respecting the *a priori* probability of any particular value. It is therefore the correct value to use when we wish for the best value *for the given data*, unbiassed by any *a priori* presuppositions.

Though I am reluctant to criticise the distinguished statisticians who put their names to the *Cooperative Study*, I do not consider their treatment of this example justifiable. A correlation 0.6000 is calculated from the sample. I suggest that the exact form of the curve of random sampling indicates that in small samples the correlation, positive or negative, is likely to be exaggerated, and therefore to correct for this effect, the best value to take is 0.5918. The writers of the *Cooperative Study* apparently imagine that my method depends upon « BAYES' Theorem » (*), or upon an assumption that our experience of parental correlations is equally distributed on the r scale, (and therefore not so on the scale of any of the innumerable functions of r, such as z, which might equally be used to measure correlation), and consequently alter my method by adopting what they consider to be a better *a priori* assumption as to the distribution of ρ. This they enforce with such rigour that a sample which expresses the value 0.6000 has its message so modified in transmission that it is finally reported as 0.462 at a distance of 0.002 only above that value which is assumed *a priori* to be most probable!

In my opinion 0.462 cannot be regarded as the correlation of the sample at all. It is in fact a kind of average value made up of the value for the sample and the values previously obtained from the other samples of different populations, measured in respect of different characters, but which have in common that they all refer to parents and children. As an average it has, of course, some value, though it could be obtained more simply and more accurately by regarding it as such. Regarded as the contribution which this sample makes to our knowledge of parental correlation it is simply misleading; its value depends almost wholly upon the preconcieved opinions of the computer and scarcely at all upon the actual data supplied to him.

Ex. IV. — A second sample of 13 from a similar population gives a correlation 0.7, what is the weighted mean of these two values?

	r	z	weight	Correlation (z) × weight
First sample .	0.6000	0.6930		
Correction . .		— 0.0125		
		0.6805	22	14.971
Second sample	0.7000	0.8673		
		— 0.0292		
		0.8381	10	8.381
Mean	0.6229	0.7297	32	23.352

(*) More properly upon a postulate analogous to that required to demonstrate BAYES Rule (BAYES, p. 371).

The weight of each sample measured on the scale of z is taken to be $(n-3)$, since the increase in weight due to the magnitude of ρ is less than 0.25 in both cases, and can indeed always be ignored save when exceptional refinement of methods is employed. The weights to be attached to values of z may therefore be taken straight from the numbers of observations. By using the z scale the changes in the variability of r, which render all means of values of r more or less inaccurate, are avoided. The mean value of r obtained thus, from that of z, represents with high accuracy the value of the correlation of the population which would yield two such samples with the greatest frequency.

4. *The probable error of intraclass correlations in general.*

The expression (IVc) gives the distribution of the correlation coefficient derived from sets of s observations, every pair of each set being given a place on the correlation table. Without examining in detail the approach of these curves to the normal form, they may be used to obtain adequate expressions for the probable errors of such correlations. This is the more necessary because all statisticians who have hitherto used these correlations, including the writer, have employed formulae which are not even approximately correct.

The transformation which rectifies this group of curves is

$$\text{VII}\quad \begin{cases} 2\overline{s-1}\,r = \overline{s-2} + s\tanh(z-\varphi) \\ 2\overline{s-1}\,\rho = \overline{s-2} + s\tanh(\zeta-\varphi)\,; \\ \text{in which}\qquad \tanh\varphi = \dfrac{s-2}{s}\,; \end{cases}$$

this transformation like (V) is applicable without labour merely by the use of a table of hyperbolic tangents. Substituting in (IVc), and writing x for $(z-\zeta)$, the curve becomes

$$\text{VIII}\quad \frac{\frac{sn-3}{2}!}{2^{\frac{sn-3}{2}}\cdot\frac{\overline{s-1}\,n-2}{2}!\,\frac{n-3}{2}!}\,e^{\frac{\overline{s-2}\,n+1}{2}(x-\varphi)}\,\operatorname{sech}^{\frac{sn-1}{2}}(x-\varphi)\,dx$$

the constant multiplier having been inserted to reduce the area of the curve to unity.

 19

This curve like its special case when $s=2$, is absolutely constant in form irrespective of the value of ρ; consequently the value of r for which ρ is an optimum agrees absolutely with the mode. Indeed this series of curves only differs from those hitherto considered in one material respect. The positive and negative infinite values of x, which give the limiting values of r, correspond to the values

$$-\frac{1}{s-1} \text{ and } +1$$

instead of

$$-1 \text{ and } +1$$

For the mode we obtain

$$\tanh(x-\varphi) = -\frac{\overline{s-2}\, n+1}{sn-1} = -\frac{s-2}{s} - \frac{2\overline{s-1}}{s\,\overline{sn-1}}$$

whence, since

$$\tanh \varphi = \frac{s-2}{s}$$

we have

$$\tanh x = -\frac{s}{2\overline{sn-1}} \Big/ \left(1 - \frac{s-2}{2\overline{sn-1}}\right)$$

The correction necessary to allow for the bias towards negative values in small samples is therefore, irrespective of s

$$+\frac{1}{2n}$$

From the second differential at the mode, we obtain as a first approximation to the weight of an observation

$$\frac{2\overline{s-1}}{s} n:$$

this value though constant on the scale of z, leads to very different values of the probable error of r.

If the peculiar nature of this distribution be ignored, and it is treated as an ordinary correlation, the standard error, will be approximately

$$\frac{1-r^2}{\sqrt{n}}$$

The term $(1-r^2)$ is $\frac{dr}{dz}$ of transformation (V); increments of z and r are nearly equal in the neighbourhood of zero, and z is equally accelerated as r approaches $+1$ and -1. In transformation (VII) the relation between z and r is linear, not in the neighbourhood of zero, but in that of the point

$$z = \varphi$$
$$r = \frac{s-2}{2\,\overline{s-1}};$$

this is in fact the central point of the range of possible values. In this neighbourhood

$$\frac{dr}{dz} = \frac{s}{2\,\overline{s-1}}$$

the standard error therefore in this region (where $z - \varphi$ is small) is

$$\frac{1}{\sqrt{n}}\left(\frac{s}{2\,\overline{s-1}}\right)^{3/2}$$

For correlations in the neighbourhood of 0.5, the increase of the numbers of members of a class can never increase the accuracy beyond that obtained from 8 times as many pairs. If the observer has the choice, he will occupy his time more profitably in observing an increased number of small classes, when, as in the investigations on homotyposis, correlations of this order are expected.

In the neighbourhood of zero, the case is quite different. As s is increased the zero is approached more and more closely by the end of the range. The distribution curve of r becomes extremely skew, so that the probable error of r becomes a very inadequate index of the curve of sampling. However, if we suppose n increased indefinitely for a given value of s, we may obtain the standard error as before.

$$\frac{dr}{dz} = \frac{s}{2\,\overline{s-1}}\operatorname{sech}^2\varphi = \frac{2}{s}$$

so that the standard error is

$$\frac{1}{\sqrt{n}}\left(\frac{2}{s\,\overline{s-1}}\right)^{\frac{1}{2}}$$

in other words, the probable error in the neighbourhood of zero is the value obtained by treating every possible pair as an independent pair of observations.

The general formula, for the standard error of r, which includes these particular cases, is

$$(1-r)\left(\frac{1}{s-1}+r\right)\sqrt{\frac{2\,\overline{s-1}}{sn}}$$

Ex. V. — The following values were given (HARRIS, 1916) for the correlations between « ovules failing » in different pods of the same tree (*Cercis Canadensis*); 100 pods were taken from each tree.

Meramec Highlands,	60	trees	0.0684	[0.0087]
Lawrence, Kansas,	22	trees	0.0858	[0.0143]

These are the lowest values recorded; the probable errors given by HARRIS are placed in parentheses. The author states « There can be no reasonable question of the statistical trustworthiness of all the direct homotypic correlations. The lowest is that for ovules failing per pod and this is in all cases 6 or more times its probable error ».

For both series

$$s = 100$$

$$\tanh \varphi = \frac{s-2}{s} = 0.98$$

$$\varphi = 2.2975.$$

Hence

r	$1-r$	tanh $\overline{z-\varphi}$	$\overline{z-\varphi}$	z	weight	Probable error	Ratio $z/p.e.$
0.0684	0.9316	−0.8446	−1.2370	1.0605	118.8	0.06188	17.14
0.0858	0.9142	−0.8101	−1.1273	1.1702	43.56	0.10221	11.45

The values of r are 7.86 and 6.00 times the probable errors given by HARRIS; the true ratios are 17.14 and 11.45. Thus HARRIS is amply justified in regarding these values as significant.

The formulae used above are

$$\tanh \overline{z-\varphi} = 1 - \frac{2\overline{s-1}}{s}(1-r)$$

and for the weight

$$\frac{2(s-1)}{s} n;$$

for such large values of s it would indeed be sufficient to use $2n$ as the weight.

Ex. VI. — The following values are given in the same table for the correlation between number of ovules in different pods on the same tree.

Meramec Highlands	60	trees	+ 0.3527	[0.0076]
Lawrence, Kansas	22	trees	+ 0.3999	[0.0121]

From these values we obtain the difference

$$0.0472 \pm [0.0142]$$

which might well be regarded as significant.

In this case we shall use the correction $+\frac{1}{2n}$

$1-r$	$\tanh(z-\varphi)$	$\overline{z-\varphi}$	$\frac{1}{2n}$	$\hat{\zeta}-\varphi$	difference	
0.6473	—0.2817	—0.2896	+0.0083	—0.2813	0.1135	±0.1195
0.6001	—0.1882	—0.1905	+0.0227	—0.1678		

In spite of the correction which tends to increase the difference, the latter is now quite insignificant. The probable errors in this region are about three times as great as those stated. The accuracy of these higher values is but little increased by using such a large number of pods from each tree.

Ex. VII. In *Homotyposis in the Vegetable Kingdom* (PEARSON 1900) there are many examples of correlations found from groups of correlated values. The uncertainty of the probable errors is indicated by placing them in parentheses. The correlation between the numbers of leaflets in different leaves from the same ash tree, will serve as a good example.

$s = 26$ $\varphi = .9231$

	Buckinghamshire $n = 108$		Dorsetshire $n = 120$		Monmouthshire $n = 100$	
	r	$z-\varphi$	r	$z-\varphi$	r	$z-\varphi$
Correlations	+0.3743	—0.2062	+0.3964	—0.1622	+0.4047	—0.1458
Probable error		±0.0466		±0.0444		±0.0486
Lower quartile	+0.3513	—0.2528	+0.3740	—0.2066	+0.3802	—0.1944
Upper quartile	+0.3977	—0.1596	+0.4191	—0.1178	+0.4296	—0.0972
Mean quartile distance	0.0232		0.0225		+0.0247	
Probable error given	[0.0109]		[0.0102]		[0.0111]	
$\frac{0.67449}{\sqrt{8n}}$	0.0228		0.0218		0.0238	

All these lie in the region where $\overline{z-\varphi}$ is small, and where an increased number of trees rather than an increased number of leaves per tree is required to give greater accuracy. The estimated probable errors are less than half their true values; while the value found by entering the probable error table (PEARSON, 1914) with $8n$ gives a good approximation. Since s is not very large the latter approximation is too small when z is close to φ and is more nearly accurate at some distance from the centre; it would, of course, be much too large at greater distances, as in the neighbourhood of zero.

The interquartile ranges of all these observations overlap, the observed differences are therefore quite insignificant. Pearson's conclusion that they are « fairly alike » is thus greatly strengthened.

Ex. VIII. — An example of an observed correlation near to zero is afforded by the resemblances of human twins. It occurred to the writer (FISHER, 1919) that if there really were two types of twins differing greatly in their degree of resemblance, a positive correlation should be found between the likeness of the same pair of twins in different traits. The degree of correlation to be expected depends on the proportionate numbers of the two types, and the correlations of the two types. For the group of measurements available (THORNDIKE, 1905) it was estimated at + 0.18. If only one type of twins were present it should be very nearly zero. The measure of resemblance used was that explained in Section 2 of this paper.

The value found from 39 pairs of twins each measured in 6 traits was $-0.016 \pm [0.048]$ the probable error having been calculated on a basis of 195 independent pairs. Using transformation VII we have

$$\varphi = .8047$$

	r	z	Weight	Probable error
Observation	— 0.016	— 0.0497	65	0.0837
Hypothesis	0.18	+ 0.4201		

The probable error was therefore greatly exaggerated, but the observation is still sensibly zero. Its difference from the point

$$\rho = 0.18$$

is now much more significantly apparent; for using the original estimate, this difference is only 4.1 times its probable error, for which $P = 0.0051$, while correctly measured it is 5.6 times its probable error, for which $P = 0.000,14$.

The evidence in favour of a single type of origin for this group of twins is thus stronger than I had previously imagined.

 24

Note on the confusion between Bayes' *Rule and my method of the evaluation of the optimum.*

My treatment of this problem differs radically from that of Bayes. Bayes (1763) attempted to find, by observing a sample, the actual *probability* that the population value lay in any given range. In the present instance the complete solution of this problem would be to find the probability integral of the distribution of ρ. Such a problem is indeterminate without knowing the statistical mechanism under which different values of ρ come into existence; it cannot be solved from the data supplied by a sample, or any number of samples, of the population. What we can find from a sample is the *likelihood* of any particular value of ρ, if we define the likelihood as a quantity proportional to the probability that, from a population having that particular value of ρ, a sample having the observed value r, should be obtained.

So defined, probability and likelihood are quantities of an entirely different nature. Probability is transformable as a differential element; thus if the probability that r falls in the range dr

$$y \quad dr$$

and we use the transformation

$$r = \tanh z$$

then the probability that z falls into the range dz, is

$$y \operatorname{sech}^2 z \, dz.$$

The likelihood of a particular value of ρ on the other hand, is equal to the likelihood of the corresponding value of ζ, being unchanged by any transformation. It is not a differential element, and is incapable of integration. Again the *mode* of a frequency curve may be arbitrarily shifted by transforming the variable, in terms of which the observations are measured, as by transformation (V) the mode of the distribution of observations has been brought into close approximation to the median; but no transformation can alter the value of the *optimum*, or in any way affect the likelihood of any suggested value of ρ.

Numerically the likelihood may be measured in terms of its maximum value; the likelihood of the optimum being taken as unity.

This must not be confused with the statement the probability of a supposition is unity, which would amount to certainty that the hypothetical supposition were true. «Probable errors» attached to hypothetical quantities should not be interpreted as giving any information as to the probability that the quantity lies within any particular limits. When the sampling curves are normal and equivariant the «quartiles» obtained by adding and sub-tracting the probable error, express in reality the limits within which the likelihood exceeds 0.796,542, within twice, thrice and four times the probable error the values of the likelihood exceed 0.402,577, 0.129,098, and 0.026.267; within once, twice and thrice the standard error, they exceed 0.606,051, 0.135,335 and 0.011,109.

The concepts of probability and likelihood are applicable to two mutually exclusive categories of quantities.

We may discuss the probability of occurrence of quantities which can be observed or deduced from observations, in relation to any hypotheses which may be suggested to explain these observations. We can know nothing of the probability of hypotheses or hypothetical quantities. On the other hand we may ascertain the likelihood of hypotheses and hypothetical quantities by calculation from observations: while to speak of the likelihood (as here defined) of an observable quantity has no meaning.

Rothamsted Experimental Station
Harpenden. England. October 1920.

REFERENCES

BAYES (1763) — *An essay towards solving a problem in the doctrine of chances.* « Phil. Trans. » LIII pp. 370-418.

H. E. SOPER, A. W. YOUNG, B. M. CAVE, A. LEE and K. PEARSON, *Cooperative Study* (1916) — *On the distribution of the correlation coefficient in small samples.* « Biom. » XI. 328-413.

R. A. FISHER (1912) — *On an absolute criterion for fitting frequency curves.* « Messenger of Maths. ». 1912.

R. A. FISHER (1915) — *Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population.* « Biom. » X, 507-521.

R. A. FISHER (1919) — *The genesis of twins.* « Genetics » IV, 489-499.

J. A. HARRIS (1916) — *A contribution to the problem of homotyposis.* « Biom. » XI, 204.

K. PEARSON (1900) — *Homotyposis in the Vegetable Kingdom.* « Phil. Trans. » A. Vol. 197. 1901. p. 295.

K. PEARSON (1914) — *Tables for statisticians and biometricians.* Table V. Camb. Univ. Press.

E. L. THORNDIKE (1905) — *Measurements of twins.* « Columbia Univ. Archives of Philosophy, Psychology and Scientific Methods ». N. 1.

26

Table I.

$$z = \tanh^{-1} r = \frac{1}{2}(\log \overline{1+r} - \log \overline{1-r})$$

For interpolation to seven decimal places apply Everett's Central Difference Formula:

$$f(a+\theta\omega) = \theta f_1 + \varphi f_0 - \frac{\theta\varphi}{6}\left\{(\theta+1)\Delta^2 f_1 + (\varphi+1)\Delta^2 f_0\right\}$$

$$+ \frac{\theta\varphi(\theta+1)(\varphi+1)}{120}\left\{(\theta+2)\Delta^4 f_1 + (\varphi+2)\Delta^4 f_0\right\};$$

in which ω is the increment in r, $f_0 = f(a)$, $f_1 = f(a+\omega)$, and $\varphi = 1-\theta$.

r	z	$\Delta^2 z$	r	z	$\Delta^2 z$	$\Delta^4 z$
.00	0	0	.50	.549,306,1	1779	3
.01	.010,000,3	21	.51	.562,729,8	1863	7
.02	.020,002.7	39	.52	.576,339,8	1954	5
.03	.030,009,0	61	.53	.590,145,2	2050	7
.04	.040,021,4	79	.54	.604,155,6	2153	6
.05	.050,041,7	102	.55	.618.381,3	2262	6
.06	.060,072,2	120	.56	.632.833,2	2377	11
.07	.070,114,7	141	.57	.647,522,8	2503	6
.08	.080,171,3	163	.58	.662,462,7	2635	10
.09	.090,244,2	182	.59	.677,666,1	2777	12
.10	.100,335,3	205	.60	.693,147,2	2931	10
.11	.110,446,9	225	.61	.708,921,4	3095	14
.12	.120,581,0	248	.62	.725.005,1	3273	15
.13	.130,739,9	268	.63	.741,416,1	3466	15
.14	.140,925,6	291	.64	.758,173,7	3674	17
.15	.151,140,4	315	.65	.775,298,7	3899	22
.16	.161,386,7	337	.66	.792.813,6	4146	21
.17	.171,666,7	360	.67	.810,743,1	4414	27
.18	.181,982,7	385	.68	.829,114,0	4709	25
.19	.192,337,2	409	.69	.847,955,8	5029	38
.20	.202,732,6	433	.70	.867.300.5	5387	32
.21	.213,171,3	461	.71	.887,183,9	5777	46
.22	.223,656,1	486	.72	.907,645,0	6213	47
.23	.234,189,5	517	.73	.928,727,4	6696	58
.24	.244,774,1	541	.74	.950.479,4	7237	65
.25	.255,412,8	569	.75	.972,955,1	7843	78
.26	.266,108,4	598	.76	.996,215,1	8527	89
.27	.276,863,8	629	.77	1.020,327,8	9300	112
.28	.287,682,1	659	.78	1.045,370,5	10185	124
.29	.298,566,3	691	.79	1.071,431,7	11194	158
.30	.309,519,6	725	.80	1.098,612,3	12361	190
.31	.320,545,4	759	,81	1.127,029,0	13718	229
.32	.331,647,1	795	.82	1.156,817,5	15304	292
.33	.342,828,3	830	.83	1.188,136,4	17182	362
.34	.354,092,5	871	.84	1.221,173.5	19422	464
.35	.365,443,8	908	.85	1.256,152,8	22126	600
.36	.376,885,9	951	.86	1.293,344,7	25430	798
.37	.388,423,1	994	.87	1.333,079,6	29532	1067
.38	.400,059,7	1037	.88	1.375,767,7	34701	1484
.39	.411,800,0	1086	.89	1.421,925,9	41354	2106
.40	.423,648,9	1134	.90	1.472,219,5	50113	3104
.41	.435,611,2	1185	.91	1,527,524,4		
.42	.447,692,0	1239	.92	1.589,026,9		
.43	.459,896,7	1294	.93	1.658,390,0		
.44	.472,230,8	1354	.94	1.738,049,3		
.45	.484,700,3	1415	.95	1.831,780,8		
.46	.497,311,3	1480	,96	1.945,910,1		
.47	.510,070,3	1550	.97	2.092,295,7		
.48	.522,984,3	1620	.98	2.297,559,9		
.49	.536,060,3	1698	.99	2.646,652,4		

TABLE I (*continued*).

r	z	$\Delta^2 z$	r	z	$\Delta^2 z$	$\Delta^4 z$
.900	1.472,219,5	498	.950	1.831,780,8	1999	6
.901	1.477,507,7	510	.951	1.842,138,5	2082	4
.902	1.482,846,9	518	.952	1.852,704,4	2169	6
.903	1.488,237,9	531	.953	1.863,487,2	2262	8
.904	1.493,682,0	541	.954	1.874,496,2	2363	4
.905	1.499,180,2	553	.955	1.885,741,5	2468	9
.906	1.504,733,7	564	.956	1.897,233,6	2582	8
.907	1.510,343,6	576	.957	1.908,983,9	2704	8
.908	1.516,011,1	591	.958	1.921,004,6	2834	9
.909	1.521,737,7	601	.959	1.933,308,7	2973	14
.910	1.527,524,4	617	.960	1.945,910,1	3126	7
.911	1.533,372,8	629	.961	1.958,824,1	3286	17
.912	1.539,284,1	645	.962	1.972,066,7	3463	13
.913	1.545,259,9	659	.963	1.985,655,6	3653	14
.914	1.551,301,6	675	.964	1.999,609,8	3857	22
.915	1.557,410,8	691	.965	2.013,949,7	4083	16
.916	1.563,589,1	707	.966	2.028,697,9	4325	26
.917	1.569,838,1	725	.967	2.043,878,6	4593	22
.918	1.576,159,6	741	.968	2.059,518,6	4883	32
.919	1.582,555,2	761	.969	2.075,646,9	5205	30
.920	1.589,026,9	781	.970	2.092,295,7	5557	39
.921	1.595,576,7	799	.971	2.109,500,2	5948	41
.922	1.602,206,4	821	.972	2.127,299,5	6380	50
.923	1.608,918,2	841	.973	2.145,736,8	6862	57
.924	1.615,714,1	866	.974	2.164,860,3	7401	65
.925	1.622,596,6	886	.975	2.184,723,9	8005	78
.926	1.629,567,7	913	.976	2.205,388,0	8687	90
.927	1.636,630,1	937	.977	2.226,920,8	9459	110
.928	1.643,786,2	963	.978	2.249,399,5	10341	125
.929	1.651,038,6	990	.979	2.272,912,3	11348	160
.930	1.658,390,0	1020	.980	2.297,559,9	12515	187
.931	1.665,843,4	1049	.981	2.323,459,0	13869	231
.932	1.673,401,7	1079	.982	2.350,745,0	15454	291
.933	1.681,067,9	1114	.983	2.379,576,4	17330	362
.934	1.688,845,5	1146	.984	2.410,140,8	19568	465
.935	1.696,737,7	1182	.985	2.442,662,0	22271	599
.936	1.704,748,1	1220	.986	2.477,410,3	25573	799
.937	1.712,880,5	1258	.987	2.514,715,9	29674	1066
.938	1.721,138,7	1300	.988	2.554,988,9	34841	1485
.939	1.729,526,9	1342	.989	2.598,746,0	41493	2106
.940	1.738,049,3	1389	.990	2.646,652,4	50251	3101
.941	1.746,710,6	1434	.991	2.699,583,9	62110	4773
.942	1.755,515,3	1486	.992	2.758,726,4	78742	7720
.943	1.764,468,6	1537	.993	2.825,743,1	103094	13588
.944	1.773,575,6	1594	.994	2.903,069,2	141034	25134
.945	1.782,842,0	1652	.995	2.994,480,7	204108	55510
.946	1.792,273,6	1713	.996	3.106,303,0	322692	147648
.947	1.801,876,5	1780	.997	3.250,394,5	588914	583273
.948	1.811,657,4	1847	.998	3.453,377,4	1438409	
.949	1.821,623,0	1922	.999	3.800,201,2		

Above .990, for accuracy to seven decimal places, and generally above 995, the exact formula, $z = \frac{1}{2}(\log_e \overline{1+r} - \log_e \overline{1-r})$, should be employed.

Table II.

$$\varphi = \tanh^{-1} \frac{s-2}{s} = \frac{1}{2} \log \overline{s-1}$$

s	φ	s	φ	s	φ	s	φ
1	—	26	1.609,437,9	51	1.956,011,5	76	2.158,744,1
2	0	27	1.629,048,3	52	1.965,912,8	77	2.165,366,7
3	.346,573,6	28	1.647,918,4	53	1.975,621,9	78	2.171,902,7
4	.549,306,1	29	1.666,102,3	54	1.985,146,0	79	2.178,354,4
5	.693,147,2	30	1.683,647,9	55	1.994,492,0	80	2.184,723,9
6	.804,719,0	31	1.700,598,7	56	2.003,666,6	81	2.191,013,3
7	.895,879,7	32	1.716,993,6	57	2.012,675,8	82	2,197,224,6
8	.972,955,1	33	1.732,868,0	58	2.021,525,6	83	2.203,359,6
9	1.039,720,8	34	1.748,253,8	59	2.030,221,5	84	2.209,420,3
10	1.098,612,3	35	1.763,180,3	60	2.038,768,7	85	2.215,408,4
11	1.151,292,5	36	1.777,674,0	61	2.047,172,3	86	2.221,325,6
12	1.198.947,6	37	1.791,759,5	62	2.055.436,9	87	2,227,173,6
13	1.242,453,3	38	1.805,459,0	63	2.063,567,2	88	2.232,954,1
14	1.282.474,7	39	1.818,793,1	64	2.071,567.4	89	2,238,668,4
15	1.319,528,7	40	1.831,780,8	65	2.079,441,5	90	2,244,318,2
16	1.354,025,1	41	1.844,439,7	66	2.087,193,6	91	2.249,904,8
17	1.386,294,4	42	1.856,786,0	67	2.094,827,4	92	2.255,429,8
18	1.416,606,7	43	1.868,834,8	68	2.102.346,3	93	2.260,894,3
19	1.445,185,9	44	1.880,600,1	69	2.109,753,9	94	2.266,299,7
20	1.472,219,5	45	1.892,094,8	70	2.117,053,3	95	2.271,647,4
21	1.497,866,1	46	1.903,331,2	71	2,124,247,6	96	2.276,938,4
22	1,522,261,2	47	1.914,320,7	72	2,131,339,9	97	2.282,174,1
23	1.545,521,2	48	1.925,073,8	73	2.138,333,1	98	2.287,355,5
24	1,567.747,1	49	1 935,600,5	74	2.145,229,7	99	2.292,483,7
25	1.589,026,9	50	1.945,910,1	75	2.152,032,5	100	2.297,559,9

FIG. 1. *Correlation between classes of 1*

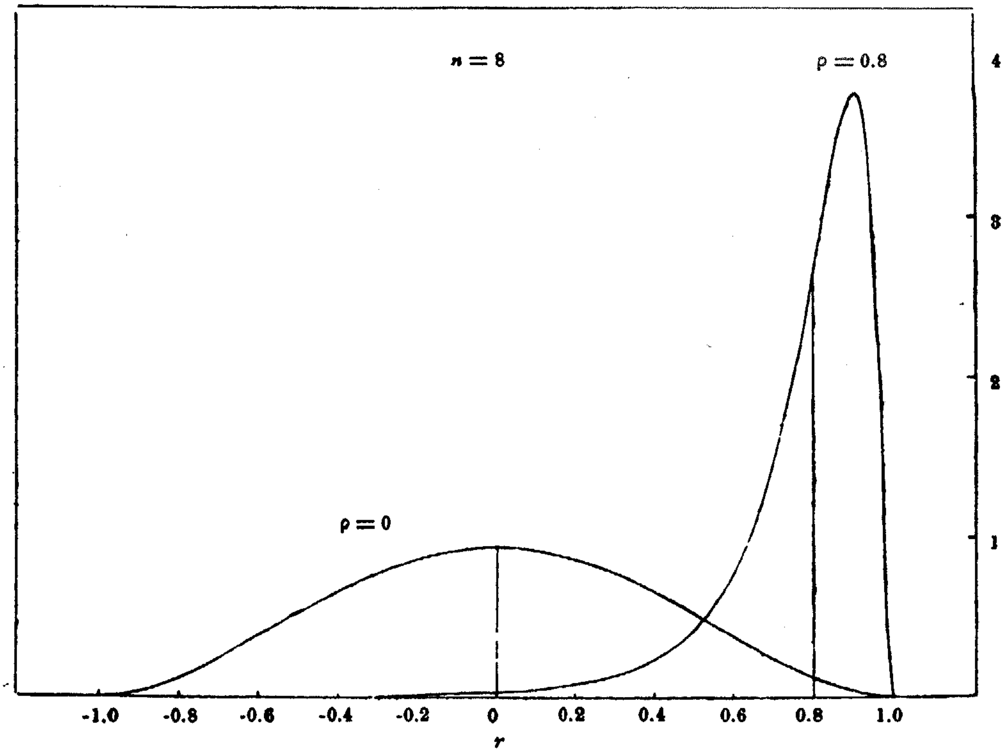


FIG. 3. *Correlation within classes of 2*

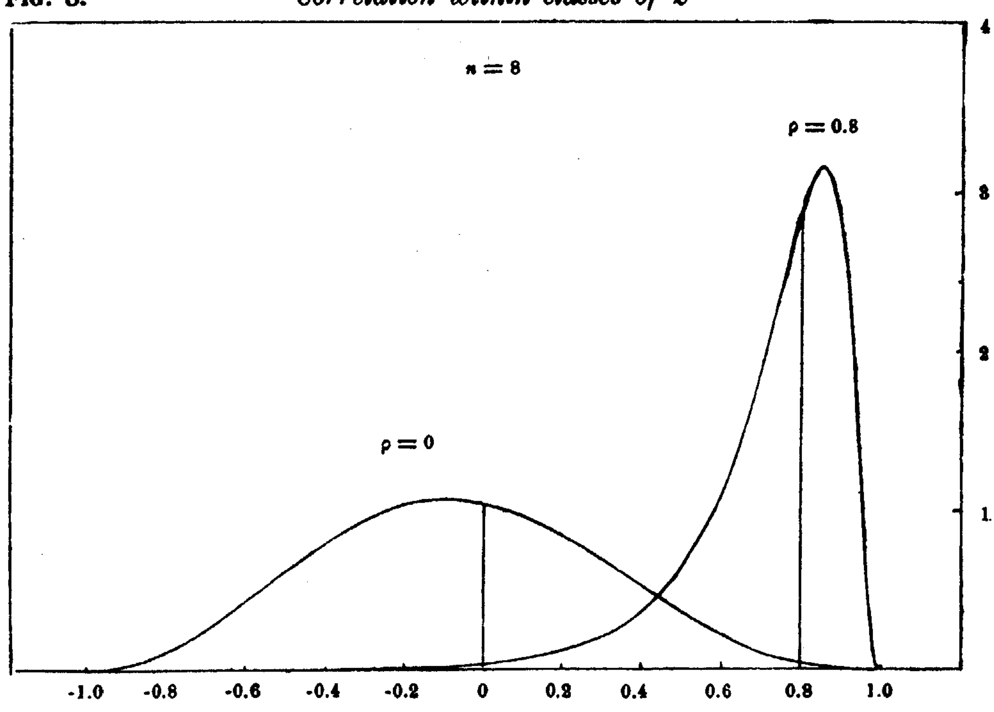


FIG. 2. *Correlation between classes of 1*

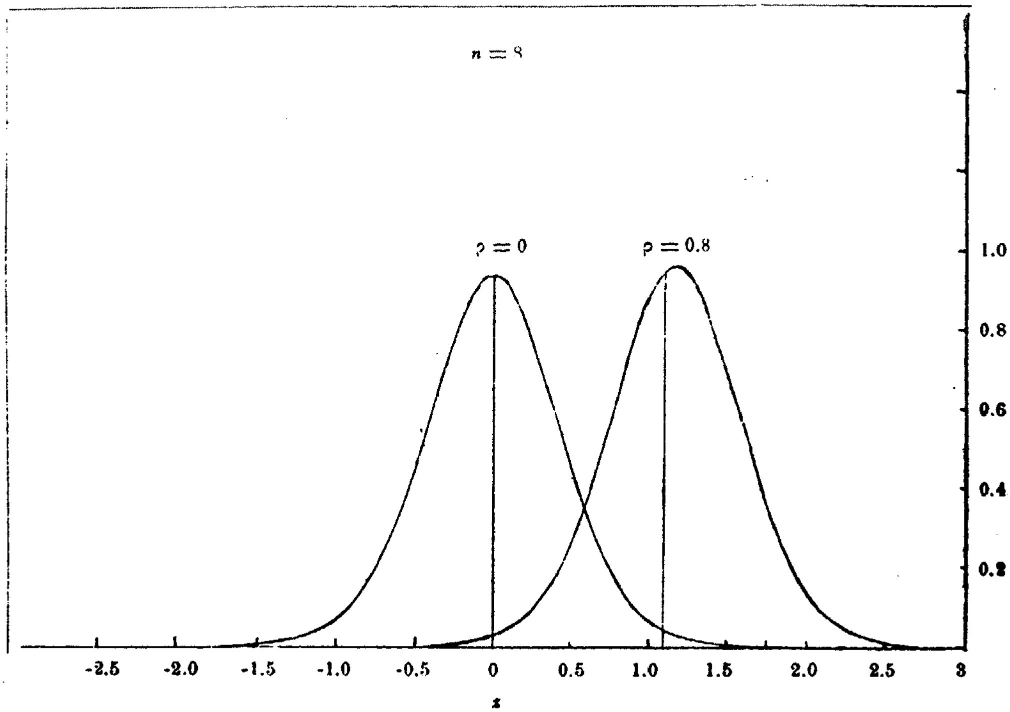
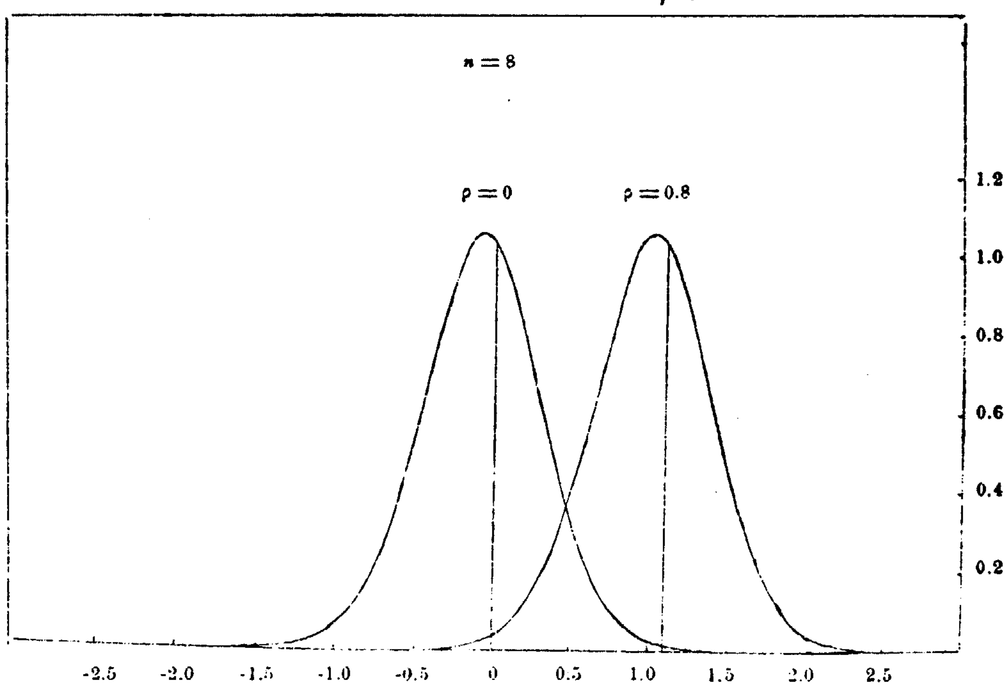


FIG. 4. *Correlation within classes of 2*



EXPLANATION OF DIAGRAMS

The diagrams illustrate the curves of random sampling of the correlation coefficient. Figures 1 and 2 illustrate the ordinary case in which it is not assumed that the two variables have the same mean and standard deviation; these have been called interclass correlations, the number in each class being unity. In each case the number of pairs of observations is 8. Figures 3 and 4 give the corresponding curves for intraclass correlations within classes of 2; as when measurements of pairs of brothers are arranged in a symmetrical table. In figures 1 and 3 it will be seen that when these correlations are distributed over the scale of r, the curves are far from normal even when $\rho = 0$, and when $\rho = 0.8$ they become extremely skew; the probable error, based on the standard deviation of these curves, gives no adequate notion of the chances of random sampling; partly because the curves are skew, and may be infinitely skew even for high values of n; partly because the curves change greatly in standard deviation, and in form, as ρ changes. The population value, ρ, is marked in each case by an ordinate.

In figures 2 and 4 are shown the curves of random sampling on the scale of z; to the eye these curves appear symmetrical, and even for 8 pairs of observations are sufficiently near to normality to be effectively represented by a probable error. For the intraclass correlations the probable error, and indeed the entire curve is identical for all values of ρ; for the interclass correlations the probable error is somewhat larger, and is slightly variable; the higher mode of the curve for $\rho = 0.8$ indicating that the standard error of high correlations is slightly exaggerated by the formula $\frac{1}{\sqrt{n-3}}$. The greater accuracy obtained by assuming a common mean and standard deviation, when this assumption is justified, is shown in figures 1 and 3, by the higher mode of the curve for $\rho = 0$, but it is entirely masked in the curve $\rho = 0.8$, by the negative bias of the interclass correlation and the positive bias of the interclass[*] correlation, which owing to the rapid changes of form of the curves on the r scale, obscure the increased

* For interclass, read intraclass.

 32

* accuracy of the interclass value. In figure 3 that negative bias is seen from the displaced mode of the curve $\rho = 0$.

In figures 2 and 4 the nature of the bias due to the method of calculation is clearly shown, for although the curves are all to the eye symmetrical, the population value is centrally placed only for $\rho = 0$, of the interclass curves; for these the bias is proportional to ρ, being negative when ρ is negative; the mean mode, median and the value of z for which ρ is an optimum all lie close together at the centre of the curve, exceeding the population value by about $\frac{\rho}{2n-1}$, which is the correction necessary for this case. For the intraclass correlations the bias is negative and independent of ρ, and is adequately allowed for by adding $\frac{1}{2n-1}$ to all observed values of z. These corrections are not derived from any supposed distribution of ρ, or ζ, the distribution of which is regarded as completely unknown; and about which it is most undesirable to make assumptions, if an objective value for the correlation is to be obtained from the sample.

It should be clear that the correction adds little to the likelihood of an individual value, but is needed when accurate comparisons are made between correlations, and especially averages of correlations, which have perhaps been calculated from samples of different sizes, or by different methods.

* For interclass, read intraclass.