THE CORRELATION BETWEEN THE SEX OF HUMAN SIBLINGS. I. THE CORRELATION IN THE GENERAL POPULATION*

J. ARTHUR HARRIS AND BORGHILD GUNSTAD
University of Minnesota, Minneapolis, Minnesota

Received November 13, 1929

TABLE OF CONTENTS

INTRODUCTION

Of recent years problems of sex—variously designated as sex determination, sex inheritance, sex linkage, etc.—have been extensively investigated by experimental methods. Interest in this field of experimentation has perhaps diverted attention from the possibilities of the application of statistical methods to certain problems and to certain kinds of data for which experimental methods as conventionally understood can not readily be used. Quite obviously the two methods of research are not mutually exclusive. Both may contribute to a more complete understanding of a complex problem.

The purpose of this paper is to consider the correlation between the sex of human siblings by means of methods which, as far as we are aware, have not heretofore been applied to this problem.

The sex of human offspring is generally assumed to be distributed by chance except for the fact that the proportion of male and female births is not equal. Let n_m be the number of male births and n_f be the number of female births in families of a total size of $n_m+n_f=n$. It is generally sup-

* Part of the cost of the accompanying tables is paid by the GALTON AND MENDEL MEMORIAL FUND.

posed that if $p_m = \Sigma(n_m)/\Sigma(n)$ be the probability of male births and $p_f = \Sigma(n_f)/\Sigma(n)$ be the probability of female births, the chances of the $s+1th$ birth in any given family being male will be p_m no matter what the sex of the s preceding children.

If this assumption be true, the distribution of the sexes in N families of constant size, n, should be given by the terms of the point binomial $N(p_m+p_f)^n$, with standard deviation of number of males (or of females) $\sqrt{np_m p_f}$. GEISSLER as early as 1889 recognized the importance of taking $p_m > p_f$ and gave $(p_m+p_f)^n$ for large series of German families. NEWCOMB (1904) also employed point binomials for representing the distribution of the number of males and females in families, but took $p_f = p_m$.

If the empirical distribution of number of males (or females) per family of size n is to be compared with the theoretical distribution as given by the point binomial with a view to determining whether the discrepancies between the two distributions are larger than those to be expected as the result of random sampling, some statistical criterion of goodness of fit must be employed. PEARSON'S χ^2, P test (1900) and ELDERTON'S tables for testing goodness of fit (1924) were not available to GEISSLER, but FISHER (1925, pp. 69–71) has shown that in the case of GEISSLER'S families of eight children the value of P (the probability that deviations as large as or larger than those actually observed may be reasonably supposed to have arisen from random sampling, and not to be confused with p_m or p_f as defined above) is small and that the actual standard deviation is larger than the theoretical.

FISHER notes the possibility of the presence of identical twins influencing the variance, but while he finds this factor inadequate to explain the differences between the observed and the theoretical standard deviation, he does not pursue the point further. The obvious explanation is that the sex of the offspring of the same parents is not wholly independent, but correlated.

METHODS

Given a series of families each of n children, six procedures for determining whether the sex of the offspring is distributed at random among the siblings are theoretically available.

(1) The actual distribution of number of males (or females) per family of given size may be compared with the theoretical distribution provided by the point binomial, as indicated above.

(2) The statistical constants of the actual distribution of number of

males (or females) per family of given size may be compared with the constants for the theoretical distribution.

(3) The relationship between the sex of any two births each occupying a definite place, say the *r*th and the *t*th, may be determined by the use of the four-fold table

Sex of *t*th child

		m	f	Total
Sex of *r*th child	m	n_{mm}	n_{mf}	$n_{mm}+n_{mf}$
	f	n_{fm}	n_{ff}	$n_{fm}+n_{ff}$
		$n_{mm}+n_{fm}$	$n_{mf}+n_{ff}$	

and the calculation of the correlation by methods applicable to such tables.

The applicability of the method is limited by the fact that in general the sequences of the sexes of human siblings are not given in the data, but only the numbers of each sex.

(4) If there be n siblings per family and if each be used once as a first and once as a second member of a pair we may regard the family as constituting a class and may obtain a four-fold intra-class (HARRIS 1913) correlation table $\Sigma[n(n-1)]$ entries, where Σ denotes summation for the N available families.

Such four-fold tables for the sex of members of the same sibship may be readily formed by methods indicated in a paper on the correlation between the fates of seeds planted in the same hill (HARRIS and NESS 1928) or from the moments of the frequency distributions by formulae given elsewhere (HARRIS, GUNSTAD and NESS 1930). The relationship between the sex of the children of the same family may then be expressed in terms of PEARSON'S equivalent probability correlation coefficient (1912).

(5) In cases in which the sex of any individual child, say the *s*th child, is definitely known and the total number of male and female children of the family is also available, the data may be represented in a bi-serial correlation surface, in which one variable is given in the alternative categories of male and female while the other is given in the quantitative terms of numbers of males or females in the $n-1$ remaining children of the family. Quite obviously such tables should be made for families of a given size, or precautions should be taken to express the number of males (or females) in the $n-1$ remaining children as ratios to $n-1$.

If the data permit this method of tabulation, it should be possible to determine correlation coefficients by some modification of PEARSON'S bi-serial method (1909).

(6) If the sex of the siblings of definite birth order be unknown but merely the total number of each sex in sibships of a given size, tables which are fundamentally symmetrical in nature but bi-serial in form may be constructed by determining the relationship between the sex of each of the n children classed in alternative categories and the sex of the remaining $n-1$ children of families of constant size considered on the quantitative scale of number of males (or females) per $n-1$ offspring.

Here, as in case (5), bi-serial theory must be applied.

These methods fall quite obviously into two groups. The first (methods 1 and 2) test for the existence of correlation between the sex of the members of the same family by determining the deviation of the number of either sex actually observed from the theoretical number, and expresses the closeness of agreement on an improbability scale. The methods of the second group (methods 3–6) express the differentiation of the families with respect to the tendency to produce an excess above the theoretical frequencies of children of either sex in terms of the correlation between the sex of the members of the same sibship.

No mathematical assumption concerning the nature of the "frequency distribution" of the character sex is necessary in the use of the first method. Given the probabilities p_m and p_f the point binomial is rigidly applicable as an expression of the theoretical distribution, on the assumption that the sex of the offspring is wholly uninfluenced by genetic or physiological variables peculiar to the family in which the offspring are produced. Practical limitations in the use of this method will be indicated below.

In the case of the four methods of the second group, certain underlying assumptions are necessary. Such assumptions are unavoidable when data can not be recorded on a quantitative scale. The methods of the group fall into two classes. The first involves a 2×2-fold tabulation of the data, while the second involves a $2\times n$-fold arrangement of the data. These last two must be treated by the bi-serial theory. Since certain difficulties in the modification of the theory to adapt it fully to present needs are still to be overcome, these methods will not be further discussed in this paper.

The correlations based on four-fold tables must be determined by the classical four-fold r method of PEARSON, or by the newer equivalent probability method also proposed by PEARSON (1912). Since the first method assumes the normality of distribution of the two variables, it has seemed desirable to place our reliance on the equivalent probability method which involves no assumption concerning the nature of the distribution. The calculation of the equivalent probability coefficient will be illustrated in greater detail below.

DATA ANALYZED AND RESULTS

The data here employed are drawn from the work of GEISSLER (1889), who has given records of the sex of 4,794,304 children. These are reported as born by 998,761 mothers, but these are weighted instead of actual numbers, since the full record of the mother and of her offspring is returned for each additional child reported during the ten year period 1876–1885. It must be noted that the series of families is not complete in that all cases in which only one child was born are omitted, and that the number of children is weighted, in that a mother who had borne two children before 1876 and who had borne three children during the decade would be recorded in families of three, four and five. These facts do not render the data unusable for present purposes, but it would be highly desirable to have the methods of analysis here developed applied to data for completed families—that is, to the records of numbers of children borne by mothers married for at least 20 or 25 years and having attained the age of 50 years.

Since GEISSLER's paper is relatively inaccessible we have rearranged his data to present the frequency distributions of number of males per family in the form of a correlation table between number of children born and number of males per family in table 1. All constants required in the present paper may be computed by the use of formulae to be given later from the data of this table.

Other series of data will be treated in a subsequent paper.

The frequency distribution of number of male children per family

The selection of the proper values of $p=\Sigma(n_m)/\Sigma(n)$ and $q=\Sigma(n_f)/\Sigma(n)$ to be used in calculating the terms of the point binomial presents no difficulty when only a single series of families of a given size is available, since the worker has no option but to use the actual numbers of each sex as given in his records. When families of different sizes are involved, the selection of the value of p requires some consideration.

GEISSLER determined the value of $p=0.514768$ by taking the 2,468,305 males and the 2,325,999 females of his families of from 2 to 30 children and adding to them the 114,609 males and the 108,719 females born as the first child in the families of two children. Thus he based his value of p on the 2,582,914 males out of the total 5,017,632 weighted births. The values of p and q thus determined he applied to families of from 2 to 12 children.

Elsewhere we have shown (HARRIS and GUNSTAD 1930) that there may be objection to GEISSLER'S procedure in obtaining the number of males and females born in families of one child, but at this date our only alternative is to utilize the value $p = 0.514768$ as given by him for the whole series.

While on first consideration it might seem that the value of p' determined from all of the available data is the most suitable value to be employed in calculating the theoretical numbers of children of each sex in families of a given size (because of the fact that the errors of random sampling will be smaller for p' derived from the sample containing all the individuals than for the values of p derived from the sub-samples representing families of any given size) this is not perfectly clearly the case.

If the sex ratio changes with the size of the family (as has been suggested in the literature), the value of p will also change. If this be true the employment of one constant value of p for the whole series will result in the using of a value which is too low for certain ranges of size of family and too high for other ranges of number of children per family.

The alternative method of procedure is, of course, to determine the values of p independently from the data of families of each size.

As far as we are aware there is no *a priori* theory for deciding which of the two procedures is the more logical. All that can be done to settle the point is to determine whether the ratios of male births to total births changes significantly from smaller to larger families, or to express this same relationship in some other way. This problem has been investigated (HARRIS and GUNSTAD 1930) with the result that, for GEISSLER'S data at least, there is a small and irregular but statistically significant increase in the proportion of males from the smaller to the larger sibships.

The results in the first section of table 2 are obtained from the constant value $p' = 0.514768$ while those in the second section are computed from the values for each individual family as given under the caption p.

The values of p tend to increase from the smaller to the larger families. Thus the values of $p - p'$ are negative for families of from 2 to 8 children and positive for families of from 9 to 18 children considered individually and for families of 19 to 30 children considered as a group. The methods of determining the ratios of $p - p'$ to their probable errors, $E_{(p-p')}$, and the significance of the differences have been discussed elsewhere (HARRIS and GUNSTAD 1930) and need not be considered here. For our present purposes the demonstration of a systematic trend in the values of p is of importance because it justifies the comparison of the actual numbers of children with two binomial distributions, one based on p and the other on p'.

The deviations of the empirical distributions of the number of males per family from the point binomials are expressed in terms of χ^2 and P.

Notwithstanding the fact that the point binomials are calculated in two ways, the values of P are low throughout. When the constant values of p' derived from all available data are employed the chances of the discrepancies between the observed and the theoretical distributions having arisen through random sampling are in all cases but one less than 2 in one hundred thousand. In 13 out of the 17 comparisons they are less than 1 in a million. When the value of p is determined independently for each size of family the highest value of P is only 0.0013 while 12 of the 17 values are less than $1/10^6$. In evaluating these results for the series as a whole it must not be forgotten that there is a certain, and unknown, amount of weighting due to two or more inclusions of some individuals. Since the distributions are calculated independently for each size of family, this limitation can not apply to the individual values of χ^2 and P.

Before closing this section a limitation of this method must be noted. As the size of the family increases the frequencies are distributed among an increasing number of classes. Thus there is a tendency for the values of χ^2 to become abnormally large through the influence of two purely statistical factors. First, the actual numbers of children must be recorded in integers whereas the theoretical numbers as given by the binomial may be given in fractions. Thus the ratio of the squared deviations of the observed from the theoretical numbers to the theoretical numbers may be abnormally large. Second, in the largest families, one or more classes may have no empirical frequencies. In such cases the class contributes the theoretical number of males to the value of χ^2. It is for this reason that the point binomial has not been computed for families of over 18 children. For families of this size the values of χ^2 are unquestionably too high.

Expression of deviation of observed from theoretical distribution in terms of difference in variability

An alternative method of expressing the results is to compare the squared standard deviation (the variance) of the empirical and theoretical distributions with regard to their probable error. The theoretical variance for number of males per family for families of size n is

$$\sigma'^2 = \mu_2' = npq.$$

We require to compare the observed squared standard deviation, μ_2 with the theoretical by evaluating $(\mu_2 - \mu_2') \pm E(\mu_2 - \mu_2')$. Since μ_2', represents a theoretical distribution to be used as a basis of comparison we take its probable error to be 0, and determine

$$(\mu_2 \pm \mu_2') \pm E_{\mu_2}.$$

As is well known (PEARSON 1903).

$$E_{\mu_2} = 0.67449\frac{\mu_4' - \mu_2'^2}{N}$$

where for samples containing two alternative classes only

$$\mu_2' = npq$$

$$\mu_4' = 3n^2p^2q^2 + npq(1 - 6pq)$$

and

$$\mu_4' - \mu_2'^2 = 2n^2p^2q^2 + npq(1 - 6pq).$$

These are easily evaluated. The results are given in table 3.

Here the first two columns give the size of the family and the total number of families in each class. The third, fourth and fifth columns give the empirical (μ_2) and the theoretical (μ_2') values of the second moment coefficients and the differences. The sixth column gives the ratio of ($\mu_2 - \mu_2'$) to its probable error. The final column gives $1/2(1-\alpha)$ or the probability of deviations as large as or larger than those actually observed having arisen through the errors of random sampling.

Since SHEPPARD'S tables of the probability integral (1902) do not give the values of deviation $>6\sigma$, we have had recourse to PEARSON'S (1912) table vi giving $-\log F$, where $F = \frac{1}{2}(1-\alpha)$ as defined by Sheppard.

Without exception the values of $\mu_2 - \mu_2'$ are positive. Since in the table of the probability integral,

$$\tfrac{1}{2}(1 - \alpha) = \int_x^\infty z\,dx$$

gives the probabilities of deviation of the proportion of males as large as or larger than those actually observed having arisen through the errors of random sampling, a glance at the values in the final column of table 3 shows that the chances of obtaining through the errors of random sampling differences between μ_2 and μ_2' as large as those actually found are in general exceedingly small.

The correlation between the sex of two consecutive children of the same mother

In only one case do GEISSLER'S (1889) data permit the determination of the correlation between the sex of two successive children of the same mother. This is possible in the case of families of two children. The frequencies for the 223,328 families and the routine of the correlation of the equivalent probability correlation coefficient are as follows.

Second Child

		Male	Female	Totals
First Child	Male	n_{mm} 59518	n_{mf} 55091	114609
	Female	n_{fm} 55196	n_{ff} 53523	108719
	Totals	114714	108614	223328

$$\chi^2 = \frac{(n_{mm}n_{ff} - n_{mf}n_{fm})^2 \cdot N}{(n_{mm} + n_{fm})(n_{mf} + n_{ff})(n_{mm} + n_{mf})(n_{fm} + n_{ff})}$$

$$= \frac{(144779078)^2 \times 223328}{(12460175871)(12459546396)} = 30.1529$$

$$\frac{1}{2}(1 + \alpha_1) = \frac{n_{mm} + n_{mf}}{N} = \frac{114609}{223328} = 0.513$$

$$\frac{1}{2}(1 + \alpha_2) = \frac{n_{mm} + n_{fm}}{N} = \frac{114714}{223328} = 0.514.$$

Interpolating from PEARSON's table V,

$$\chi_{\alpha_1} = 1.2536 \quad \chi_{\alpha_2} = 1.2537$$

$${}_0\sigma_r = \frac{1}{\sqrt{N}}\chi_{\alpha_1}\chi_{\alpha_2} = \frac{(1.2536)(1.2537)}{\sqrt{223328}} = \frac{1.57163832}{472.5759} = 0.00333$$

$$r_p{}^2(=){}_0\sigma_r{}^2\chi^2 = (0.0000110889)(30.1529) = 0.00033436$$

$$r_p(=) + 0.0183.$$

The equivalent probability intra-class correlation between the sex of children of the same family

In the determination of the intra-class equivalent probability correlation coefficients the first task is the formation of the 2×2-fold table. This is accomplished by use of the method of an earlier paper (HARRIS, GUNSTAD and NESS 1930) as follows:

Let n_m be the number of male births, and n_f the number of female births in families of total size $n_m + n_f = n$. Then the frequencies of the 4-fold table

	Male	Female	Totals
Male	n_{mm}	n_{mf}	$n_{mm}+n_{mf}$
Female	n_{fm}	n_{ff}	$n_{fm}+n_{ff}$
Totals	$n_{mm}+n_{fm}$	$n_{mf}+n_{ff}$	$\Sigma[n(n-1)]$

may be written in terms of the first and second moments, and product moments of the variables.

$$
\begin{aligned}
n_{mm} &= \Sigma\left[n_m(n_m - 1)\right] = \Sigma(n_m{}^2) - \Sigma(n_m)\\
n_{ff} &= \Sigma\left[n_f(n_f - 1)\right] = \Sigma(n_f{}^2) - \Sigma(n_f)\\
n_{mf} = n_{fm} &= \Sigma(n_f n_m) = \tfrac{1}{2}\left[\Sigma(n^2) - \Sigma(n_m{}^2) - \Sigma(n_f{}^2)\right].
\end{aligned}
$$

Applying these formulae to the determination of the 2×2-fold table for the relationship between the sex of the members of 95,390 families of 6 children each, for which the frequency distribution of number of males is given in table 1, we obtain the distribution shown in the following table

		Sex of "second" child		
		Male	Female	Totals
Sex of "first" child	Male	n_{mm} 758214	n_{mf} 712561	1470775
	Female	n_{mf} 712561	n_{ff} 678364	1390925
	Totals	1470775	1390925	2861700

The equivalent probability intra-class correlation may be determined (with a notation differing slightly from that used by PEARSON) as follows:

$$
\chi^2 = \frac{(n_{mm}\,n_{ff} - n_{mf}{}^2)^2 \cdot N}{[(n_{mm} + n_{mf})(n_{mf} + n_{ff})]^2} = \frac{(6601903175)^2 \times 2861700}{(2045737716875)^2} = 29.8032
$$

$$
\frac{1}{2}(1 + \alpha_1) = \frac{1}{2}(1 + \alpha_2) = \frac{n_{mm} + n_{mf}}{N} = \frac{1470775}{2861700} = 0.514.
$$

Interpolating from PEARSON'S table V,

$$
\chi_{\alpha_1} = \chi_{\alpha_2} = 1.2537
$$

$${}_0\sigma_r = \frac{1}{\sqrt{N}}\chi_{\alpha_1}\chi_{\alpha_2} = \frac{1.571764}{1691.6560} = 0.00093$$

$$r_p{}^2(=){}_0\sigma_r{}^2\cdot\chi^2 = (0.0000008649)(29.8032) = 0.00002578$$

$$r_p(=) + 0.0051$$

All of the data may be treated by the intra-class equivalent probability method. We have applied this procedure to individual families of 2 to 18 children, but have grouped together the results for the 141 families with 19 to 30 children, since the maximum number of any one of these classes is 77 for families of 19 children.

Table 4 gives the number of children per family, the frequency of each class of families, the entries (n_{mm}, n_{mf}, n_{fm}, and n_{ff}) of the four cells of the four-fold table, the value of χ^2 measuring the divergence between the actual frequencies and the theoretical frequencies of the four-fold distribution, and the equivalent probability correlation coefficients.

The correlation coefficients are of a very low order of magnitude, ranging from 0.0015 to 0.0824 but are positive in sign throughout. Since if there were no correlation between the sex of members of the same sibship the coefficients should be distributed about 0 with approximately equal numbers of positive and negative coefficients, we may note that if we ignore the magnitudes of the coefficients and consider only their uniform positive sign, we have about the same chance of obtaining these results by random sampling as of obtaining 18 consecutive heads in throwing a coin. This should occur about 4 times in a million.

The average values for the 17 coefficients for families of from 2 to 18 children is 0.0136. The coefficient for the 141 families with from 19 to 30 children is 0.0824. Combining all the data into a single four-fold distribution we find a coefficient of 0.0095.

The coefficient for the series as a whole is far lower than that for the families with from 19 to 30 children. It is quite probable that these very large families contain a relatively large proportion of twin births, many of which would be of the same sex. The low values for the series as a whole is determined by the enormous excess of records of families with small numbers of children. If we limit our attention to families of less than 9 children (for each class of which over 50,000 records are available) we find that the coefficients range from +0.0015 to +0.0182.

It is quite clear, therefore, that the sex of the members of the same family is not independent but correlated. Thus, certain parents have a definite tendency to produce families with a slight excess of males and others to produce sibships with a slight excess of females, both beyond their theoretical frequencies as determined by laws of chance.

DISCUSSION, SUMMARY AND CONCLUSIONS

This paper, which is one of a series on the sex of human offspring, presents the results of an investigation of the distribution of the two sexes in large series of German families of various sizes and of the correlation between the sex of members of the same sibship.

It is shown that the distribution of the numbers of the two sexes in families of a given size is not strictly in accord with the theory that the sex of human offspring is determined wholly by chance, but that there is a significant deviation from a chance distribution.

Methods of expressing this deviation on the mentally comprehensible correlation scale are suggested. It is shown that when this is expressed in terms of intra-class equivalent probability correlation there is a very low positive relationship, of the order $r = 0.01$, between the sex of members of the same sibship.

With respect to the relative value of the two methods of analysis here suggested we may note the following considerations.

On first thought the comparison of the actual number of males or females in families with a given number of children with the theoretical number as given by the point binomial would seem to give the most critical and valuable criterion. This would seem to be true for two reasons. First, the assumptions underlying the application of these methods are of the simplest, most fundamental, and admittedly sound kind. Second, each frequency of the empirical distribution is compared with a corresponding frequency of the point binomial.

Against this procedure two objections must be urged. First, the χ^2, P test becomes less reliable as the ratio of the number of classes to the number of individuals included in the empirical series becomes larger. Second, the values of P are often so nearly infinitesimal that they are mentally incomprehensible and non-comparable from series to series. These defects are apparently both eliminated by expression of the results in terms of equivalent probability intra-class correlation.

In the present case the results obtained by the two methods are wholly consistent as far as the biological generalizations to which they lead are concerned and differ only in the form and numerical values of the expressions. The uniform consistency of results for families of all sizes rather than the ratios of the individual correlation coefficients to their probable errors provides the basis for confidence in the statistical significance of the results.

As an incidental result of the present investigation we may note that it establishes empirically the significance of small values of the correlation coefficient.

Biologists often assert that correlation coefficients of the order $r=0.10$, $r=0.20$, etc. are "insignificant," "meaningless," "of no value," or "worthless." This is in part due to the fact that some biologists have not yet grasped the idea that statistical constants are to be judged by their ratios to their probable error and not by their absolute magnitude. It is in part due to the failure of most biologists to realize that with extensive data and refined methods of analysis biology may be a highly exact science, and that in consequence relationships of a very low order may be expressed with a high degree of confidence as to their validity.

In our opinion hypotheses as to the biological interpretation of these findings as to matters of fact should be held in abeyance until further quantitative evidence of various kinds is available. A suggestion that immediately presents itself is that the series of data, while primarily composed of single births, contains a certain proportion of twins and presumably a small proportion of triplets. These are generally known to show an abnormally high proportion of individuals of the same sex. Thus an attempt might be made to explain the present correlation as due to the mixture of highly correlated and uncorrelated materials. A discussion of the influence of the presence of these multiple births on the correlations must be reserved for a subsequent paper.

Finally we must emphasize the fact that sex in man presents a highly complicated problem or group of problems. The sex of all zygotes can not be ascertained, but only the sex of those zygotes which develop to an age of several months. Thus the investigation of the problem of prenatal mortality is necessary to a full interpretation of the present statistical results. Here, as in many other fields of biological investigation, individual series of data generally fail to provide all the information required. We must be content to postpone synthetic work until we shall have available a sufficient range of constants derived from various sources to make sound conclusions possible.

LITERATURE CITED

ELDERTON, W. PALIN, 1924 Tables for testing the goodness of fit of theory to observation. Biometrika 1: 155–163, 1902. Reprinted as table XII in PEARSON'S tables for statisticians and biometricians, Part I, London.

FISHER, R. A., 1925 Statistical methods for research workers. London: Oliver and Boyd.

GEISSLER, ARTHUR, 1889 Beiträge zur Frage des Geschlechtsverhältnisses der Geborenen. Zeitschr. K. Sächs. Statistischen Bureaus. 35: 1–24.

HARRIS, J. ARTHUR, 1913 On the calculation of intra-class and inter-class coefficients of correlation from class moments when the number of possible combinations is large. Biometrika **9**: 446–472.

HARRIS, J. ARTHUR, and GUNSTAD, BORGHILD, 1930 The problem of the relationship between the number and the sex of human offspring. Amer. Nat. (in press)

HARRIS, J. ARTHUR, GUNSTAD, BORGHILD, and NESS, MARIE M., 1930 The determination of intra-class and inter-class equivalent probability coefficients of correlation Amer. Nat. **64**: 115–141.

HARRIS, J. ARTHUR, and NESS, MARIE M., 1928 On the applicability of PEARSON's equivalent probability *r* method to the problem of seedling mortality in Sea Island, Egyptian and upland cotton. J. Agric. Res. **36**: 615–623.

NEWCOMB, SIMON, 1904 A statistical inquiry into the probability of causes of the production of sex in human offspring. Pub. Carnegie Instn. **11**: 1–34.

PEARSON, K., 1900 On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Phil. Mag. **50**: 167–175.

1903 On the probable errors of frequency constants. Biometrika **2**: 273–281.

1909 On a new method of determining correlation between a measured character, A, and a character B, of which only the percentage of cases wherein B exceeds(or falls short of) a given intensity is recorded for each grade of A. Biometrika **7**: 96–105.

1912 On a novel method of regarding the association of the variates classed solely in alternate categories. Drap. Co. Mem. Biom. **7**: 1–29, *Pls.* 1–2.

SHEPPARD, W. F., 1902 New tables of the probability integral. Biometrika **2**: 174–190.

TABLE 1

Distribution of nı n bcı of n.cles per sibship in GEISSLER'S *series of 998761 (weighted) families.*

Number of Males																										
	223328	179892	148903	120137	95390	72069	53680	38495	26500	16759	10690	6115	3332	1769	913	439	209	77	36	12	8	4	1	1	1	1	998761
21																						1					1
20																				1							1
18																					1						1
17																		1	1	..	1	1					4
16																1	3	..	1	1	..	1					7
15																3	2	6	..	..	2						13
14													1	2	8	5	6	1	2	..	..						25
13												4	11	21	21	23	14	6	3	2	..						105
12											7	24	60	49	40	32	15	8	5	3	1						244
11										14	60	108	146	109	80	44	20	7	4	1	1	1					595
10									43	159	298	338	276	203	117	68	34	8	6	3	1			1	1	1	1557
9								138	397	665	799	629	409	252	149	91	50	18	8				1				3606
8							342	981	1542	1611	1398	987	566	348	190	82	29	9	1								8086
7						786	2092	3204	3567	2899	2033	1234	675	335	137	46	14	6	1	1	1						17031
6					1943	4739	6678	6687	5674	3733	2360	1251	575	218	88	21	13	4	1								33985
5				4459	9933	12589	11929	9632	6363	3562	1821	849	328	145	48	12	6	2	1								61679
4			10634	20085	23274	20149	14959	9035	4868	2445	1198	428	173	53	20	6	..										107327
3		24923	38903	38665	30084	19103	10649	5652	2714	1164	521	182	88	19	8	3	3		2								172683
2	59518	68974	56025	36648	20736	10683	5331	2456	1058	405	160	69	20	12	5	2		1									262103
1	110287	65063	34705	16851	8012	3479	1485	630	239	89	29	10	4	2	1												240886
0	53523	20932	8636	3429	1408	541	215	80	35	13	6	2	..	1	1												88822
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	'22	23	27	28	29	30	

Number of Children per Family

TABLE 2

Values of χ^2 and P measuring the divergence of numbers of males from the numbers calculated from the point binomial on the basis of the value of p for individual families and on the basis of p' for the whole series.

SIZE OF FAMILIES	NUMBER OF FAMILIES	CONSTANT p'			VARIABLE p			
		χ^2	P	p	$p-p'$	$p-p'/E_{p-p'}$	χ^2	P
2	223328	33.4387	.000000	.513422	−.001346	2.79	30.1547	.000000
3	179892	17.3975	.000601	.514716	−.000052	0.12	17.3862	.000604
4	148903	31.9469	.000004	.513757	−.001011	2.46	29.4160	.000007
5	120137	29.4945	.000019	.514041	−.000727	1.78	28.1743	.000035
6	95390	53.0014	.000000	.513951	−.000817	1.94	51.9114	.000000
7	72069	46.7716	.000000	.514648	−.000120	0.27	46.7392	.000000
8	53680	91.8993	.000000	.514677	−.000091	0.18	91.8965	.000000
9	38495	84.5168	.000000	.515039	+.000271	0.49	84.2706	.000000
10	26500	78.8680	.000000	.517494	+.002726	4.27	71.7229	.000000
11	16759	68.5036	.000000	.516895	+.002127	2.75	65.1896	.000000
12	10690	92.8036	.000000	.516830	+.002062	2.21	100.0275	.000000
13	6115	80.4336	.000000	.519026	+.004258	3.58	74.5460	.000000
14	3332	136.8569	.000000	.520215	+.005447	3.50	125.2893	.000000
15	1769	75.6104	.000000	.521688	+.006920	3.35	91.6556	.000000
16	913	164.0801	.000000	.521495	+.006727	2.41	148.0915	.000000
17	439	56.7651	.000008	.535710	+.020942	5.36	40.5625	.001291
18	209	76.8111	.000000	.526581	+.011813	2.15	55.0391	.000038
19–30	141	..	..	.527866	+.013098	2.06	..	..

TABLE 3

Comparison of actual variance ($\mu_2=\sigma^2$) of number of males per family with theoretical variance ($=npq$).

SIZE OF FAMILY	NUMBER OF FAMILIES	μ_2	μ_2'	$\mu_2-\mu_2'$	$E(\mu_2-\mu'_2)$	$\frac{\mu_2-\mu_2'}{E(\mu_2-\mu_2')}$	½(1−α)
2	223328	.505447	.499640	+.005807	.000714	8.14	21×10^{-9}
3	179892	.757856	.749350	+.008506	.001377	6.18	15×10^{-6}
4	148903	1.008960	.999243	+.009717	.002139	4.54	11×10^{-4}
5	120137	1.253921	1.249014	+.004907	.003075	1.60	14×10^{-2}
6	95390	1.523017	1.498832	+.024185	.004226	5.72	57×10^{-6}
7	72069	1.790443	1.748498	+.041945	.005753	7.29	44×10^{-8}
8	53680	2.067417	1.998277	+.069140	.007697	8.98	67×10^{-11}
9	38495	2.370400	2.247964	+.122436	.010300	11.89	14×10^{-17}
10	26500	2.655890	2.496940	+.158950	.013882	11.45	52×10^{-16}
11	16759	2.957159	2.746860	+.210299	.019300	10.90	89×10^{-15}
12	10690	3.339853	2.996601	+.343252	.026472	12.97	1×10^{-18}
13	6115	3.633895	3.245294	+.388601	.038038	10.22	26×10^{-18}
14	3332	4.260543	3.494279	+.766264	.055649	13.77	71×10^{-22}
15	1769	4.409851	3.742944	+.666907	.082019	8.13	21×10^{-9}
16	913	4.836813	3.992607	+.844206	.122054	6.92	15×10^{-7}
17	439	4.965750	4.228322	+.737428	.186810	3.95	39×10^{-4}
18	209	5.742376	4.487282	+1.255094	.287780	4.36	16×10^{-4}

TABLE 4

Frequencies for 2×2-fold intra-class correlation tables, values of χ^2 derived from four-fold table, and equivalent probability coefficient (r_p) measuring the correlation between the sex of members of the same family.

SIZE OF FAMILIES	NUMBER OF FAMILIES	FREQUENCY OF 2×2-FOLD TABLE				χ^2	r_p
		n_{mm}	n_{mf}	n_{fm}	n_{ff}		
2	223328	119036	110287	110287	107046	60.3034	.0182
3	179892	287486	268074	268074	255718	34.7723	.0089
4	148903	473076	444924	444924	423912	18.7725	.0051
5	120137	635486	599622	599622	568010	2.3159	.0015
6	95390	758214	712561	712561	678364	29.8032	.0051
7	72069	804734	753052	753052	716060	48.3909	.0063
8	53680	800000	747161	747161	711758	73.4430	.0078
9	38495	739934	687570	687570	656566	128.4641	.0107
10	26500	642916	591308	591308	559468	119.3125	.0111
11	16759	496068	456822	456822	433778	108.0492	.0120
12	10690	380588	348701	348701	333090	153.0095	.0164
13	6115	259356	235764	235764	223056	94.9359	.0157
14	3332	166666	148805	148805	142148	172.5552	.0265
15	1769	102284	91518	91518	86170	60.1719	.0200
16	913	60362	53908	53908	50942	43.5393	.0222
17	439	34592	29376	29376	26064	14.1875	.0172
18	209	17996	15681	15681	14596	17.3119	.0259
19–30	141	15636	12737	12737	12804	148.0288	.0824
2–30	998761	6794430	6307871	6307871	5999550	927.5633	.0095