CHAPTER IV

THE STATISTICAL STUDY OF HEREDITY

In studying heredity, either of two methods may be adopted. We may either choose a character and observe or measure its development in a large number of parents and in their children, and so deduce the average extent of resemblance between parents and children for that character; or we may consider a number of individual cases separately, and endeavour to discover the manner in which the character appears in the children who have parents or ancestors possessing it. With regard to the first method Prof. Pearson has written 'We must proceed from inheritance in the mass to inheritance in narrower and narrower classes, rather than attempt to build up general rules on the observation of individual instances.' And '...the very nature of the distribution ...seems to indicate that we are dealing with that sphere of indefinitely numerous small causes, which in so many other instances has shown itself only amenable to the calculus of chance, and not to

the analysis of the individual instance' [25, 'Math.
Contrib. III.'   *Phil. Trans. Roy. Soc.* A, 1896, p. 255].
The second method on the other hand has been used
in cases where the causes of variation appear to be
few and definite, and seeks to isolate these causes.
The first method is thus clearly adapted especially to
characters which vary continuously and which can
be measured ; the second to characters which vary
discontinuously and can be sharply separated into
classes.   The first method gives on the whole the
average intensity of inheritance, but little information
with regard to its probable development in individual
cases ; the second attempts to answer the question
in what manner the character will be distributed
among the offspring in any family.

   The founder of the modern statistical, or as it is
now often called, the biometrical study of heredity was
Sir Francis Galton, and its leading exponents have
been Professor Karl Pearson and the late Professor
Weldon.   In this chapter an attempt will be made to
explain the fundamental principles on which the
biometric methods rest, and to outline the chief
results obtained; the methods themselves frequently
require mathematics of an advanced order, and for
the study of them the reader is referred to the books
and papers dealing with the subject mentioned in the
bibliography.

   It has already been seen that in the case of a

character which varies continuously about a mean or
mode, the greater the divergence from the mode in
either direction, the fewer will be the individuals
showing that divergence. In the case of human
stature, if the modal height of a population is 68
inches, there will be fewer men of 64 or 72 inches
than of 66 or 70, and still fewer of 63 or 73 inches.
If now the sons of all the men having a given diver-
gence were measured, and it were found that they
averaged as great a divergence from the mode as
their fathers, it is clear that on the average the
height of the sons would equal that of their fathers.
This does not mean that every son would exactly
resemble his father in stature, but the sons would
vary about the paternal stature equally above and
below it, and when plotted in a curve their statures
would make a curve having the paternal stature as
its mode. The average stature of the sons would
then be completely determined by the stature of the
fathers. If on the other hand the stature of the
father had no relation with that of his sons, it is
clear that the statures of the sons of fathers of any
height would vary about the mean of the general
population considered. In practice it is found that
the modal value for sons of fathers of a given height
is between the height of their fathers and the mode
of the general population. That is to say, if the fathers
diverge a given amount from the general mode, their

sons will on the average diverge less ; they will vary about a modal value lying between the general mode and the fathers' measurement.   This fact is called 'regression.' It sometimes seems paradoxical to those who have not considered it that the mean deviation of children from the general mode is always less than that of their parents.   But of course it does not mean that all sons of tall fathers will be shorter than their fathers ; some will be as tall or taller, but the sons of a number of fathers of given stature will vary about a mode lying between the fathers' stature and the mode of the whole population.

Now it is plain that the amount of regression is a measure of the intensity of inheritance ; if the modes for sons of fathers of every deviation have deviations nearly as great as those of the fathers, the intensity of inheritance would be high ; if the modes for the sons deviate but slightly from the general mode, whatever be the deviation of the fathers, the intensity would be low.   A definite case may make this clearer. Suppose the modal stature of the population is 68 inches ; it might then be found that for fathers of 64 inches (i.e. deviating 4 inches below), the height of the sons ranged from 61 to 72 inches.   If, however, their modal value had a deviation only slightly less than the fathers' deviation, say with a mode at 65 inches, the regression would be slight and the intensity of inheritance high ; if the sons' mode had a deviation

much less than the fathers', say at 67 inches, the
regression on the general mean would be considerable
and the intensity of inheritance low.   If then we can
find means of determining the ratio between the
deviation of sons in general and the deviation of their
parents, we shall have a measure of the intensity of
inheritance for the character considered.   This ratio
is called the 'coefficient of correlation' between
father and son for that character.   It should be
noticed that correlation simply means that two
quantities vary in relation to each other ; the corre-
lation between parents and children is a convenient
method of estimating the intensity of inheritance,
but correlation exists between any two related
variables, e.g. between the measurements of two
limbs in the same individual, such as an arm and a
leg, or between the numbers obtained in successive
throws of dice, if not all the dice are picked up for
the second throw.   The correlation between the same
measurement in brothers may be used as a measure
of inheritance, for two brothers resemble each other
more than two chance individuals because they are
children of the same parents.

   The principle of obtaining a coefficient of corre-
lation between father and sons is as follows.   It will
be convenient to assume that the variability of the
character considered is normal, i.e. that the frequency
curve falls evenly on either side of the mode, so that

the mode is identical with the mean. Stature in inches may be taken as an example. A large number of fathers and a son of each are measured to the nearest inch ; it can then be found what is the average measurement of sons for fathers of each
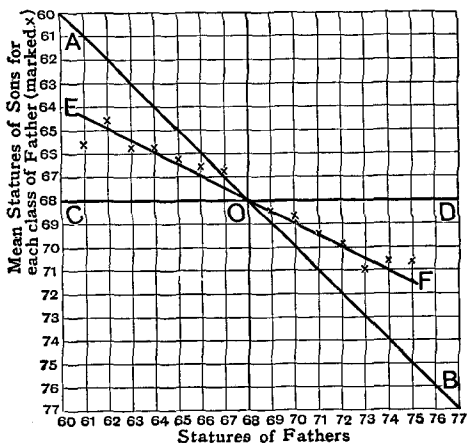

Fig. 6. Diagram of correlation between fathers and sons.

height from the lowest value to the highest. It will be found that the mean deviation of the sons from the mean of the population is less than the deviation of the fathers for each class of fathers. The average ratio between the mean deviations of the sons to the

deviations of the fathers is then the coefficient of correlation between father and son for this character[1]. This is more clearly seen in diagram form.

If a square is made with its sides divided into equal lengths corresponding to equal increments in stature from 60 to 76 inches, the top may represent the scale of statures of fathers and the side the scale of mean statures of sons for each class of fathers. If, then, there were complete correlation between fathers and sons, the mean stature of sons of fathers 62 inches high would be 62, of fathers of 63 inches, 63, of 64, 64 and so on. If on the other hand there were no correlation, the means of the sons of every class of father would be the mean of the population (68).

In the first case the line joining the points representing the means of the sons would be a diagonal running from corner to corner (AB), in the second case a horizontal line running across the middle (CD). But if the correlation is between these extremes the line would lie between the diagonal and the horizontal (EF), and the greater the correlation the steeper would be the slope of EF. The steepness of this line is thus a measure of correlation, and since all these lines pass through O in the middle of the square, the

----

[1] It is assumed throughout that the variability of the sons is similar to that of the fathers. If their variability were different this would have to be allowed for. The variation is also assumed to be normal, so that the mode in each case coincides with the mean.

slope is measured by the size of the angle EOC. The angle made by the diagonal at O is 45°, the tangent of which is 1 (unity). If there were no correlation the angle would vanish, EF coinciding with CD, and the correlation coefficient would be 0. Intermediates are represented by the value of the tangent of the angle EOC. In practice it rarely happens that the points representing the means of the sons for each class of fathers lie in a perfectly straight line; when they approach it closely the correlation is called 'linear'; when they depart from it considerably, it is called 'skew.'

Table of intensity of parental inheritance in different species. (From Pearson.)

Species	Character	Mean value	Number of Pairs used
Man	Stature	·506	4886
	Span	·459	4873
	Forearm	·418	4866
	Eye Colour	·495	4000
Horse	Coat Colour	·522	4350
Basset Hound	Coat Colour	·524	823
Greyhound	Coat Colour	·507	9279
Aphis	Ratio of right Antenna	·439	368
(*Hyalopterus*	to Frontal Breadth		
trirhodus)	(non-sexual reproduction)		
Water-flea	Ratio of Basal Joint of	466	96
(*Daphnia magna*)	Antenna to Body length		
	(non-sexual reproduction)		

Prof. Pearson and his collaborators have worked out the correlation between parent and child for a number of measurable characters in Man, Animals, and Plants, and they find that the numbers group themselves about a value not far from 0·48, varying from 0·42 to 0·52.   That is to say, on the average the offspring deviate from the mean about half as much as the parent.

The parental correlation hitherto discussed has taken no account of the second parent, for if individuals mate at random the one parent may be considered alone, and the second will *on the average* have the mean value for the general population.   But it is clear that one may take for the parental value in each class the mean of the two parents (making allowance for any difference in measurement due to sex), and plot the means of the sons (or daughters) against the classes so produced.   The value derived from taking the mean of father and mother is called the 'mid-parent,' and the correlation so arrived at would give the measure of resemblance between children and their mid-parents.   This is naturally higher than the correlation observed when only one parent is considered; for if both parents deviate in the same direction from the mode of the population, the children will average a greater deviation than if only one does so, and still more than if one deviates in one direction, the other in the opposite.   We thus

obtain a measure of the amount contributed to the offspring by the two parents together, but even now we do not find the correlation complete (1·0) because the contributions from previous ancestors have also to be taken into account.

Galton was the first to introduce the idea of the 'mid-parent,' and he went on to attempt to estimate the average contribution to the children from each generation of ancestors. Since the correlation between offspring and mid-parent is not complete, part of the heritage, which is not visibly present in the parents, must be contributed from more distant ancestors. Galton concluded from the data he collected that on the average half the heritage of an individual may be taken as derived from the two parents, one quarter from the four grandparents, one eighth from the great grandparents, and so on, the whole series $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$—$)$ adding up to unity. Pearson estimates the average correlation between offspring and one parent, as about ·5, of offspring with a grandparent as ·33, with a great grandparent as ·22, the correlation coefficient with an ancestor of each generation being $\frac{2}{3}$ of that of the next below; these numbers, however, are not in any way comparable with Galton's series ·5, ·25, ·125, etc. Galton attempted to estimate the amount of the heritage received from the 'mid-ancestor' of each generation independently of what was received from other generations; but in the

metaphor of bequests of property, he calculated that
of the total heritage of an individual, half on the
average was bequeathed by the parents, one quarter
by the grandparents direct to the grandchild and so
on.   Pearson's series ·5, ·33, ·22 etc. gives the average
measure of resemblance between children and an
ancestor of each generation, which is clearly a totally
different thing.   From this series he has worked out
figures corresponding to Galton's, making the series
·6244, ·1988, ·0630, i.e. he finds that the parental
bequest is greater and the ancestral bequests less
than Galton estimated.   From the results obtained
first by Galton and later by Pearson has been formu-
lated the 'Law of Ancestral Heredity,' which has
been stated in various forms, perhaps the most
general being 'the mean character of the offspring
can be calculated with the more exactness, the
more extensive our knowledge of the corresponding
characters of the Ancestry' (Yule [44]). But it should
be noted that there is an important difference between
Galton's original statement of the law, and the later
statements of Prof. Pearson.   Galton wrote that 'the
two parents between them contribute on the average
one-half of each inherited faculty, each of them
contributing one-quarter.   The four grandparents
contribute between them one-quarter, or each of
them one-sixteenth ; and so on.'   He regarded this
as a physiological statement of the way faculties

were transmitted, while Pearson, in his later writings at least, regards the law simply as a statistical description of what is found when large numbers are observed in mass.

It has been mentioned that the characters which especially lend themselves to statistical treatment are those which vary continuously and which can be accurately measured, but Prof. Pearson has applied similar methods to discontinuous characters, which can be classified into groups but not measured, for example *coat-colour in horses*. He finds as the results of his enquiries that the inheritance of such characters can be stated in terms similar to those obtained with measurable characters, so that the principle of ancestral correlation leading up to the law of ancestral heredity may be applied to these characters also.   But whatever may be the case with *characters which vary continuously, it will be seen* below that discontinuous characters are commonly alternative in their inheritance, i.e. there is no blending, but the offspring exhibit one or other only ; and in some at least of these cases, the character of the offspring cannot be calculated with any more exactness if the ancestry is known than *if it is not.* Such instances show clearly that although the law may be statistically true when applied to considerable populations, it gives us no clue to the physiological

processes which determine the transmission of char-
acters from one generation to another.

Another argument that has been used against the
physiological validity of the law of ancestral heredity
is based on the work of Johannsen and others who
have obtained results similar to his in other cases.
Johannsen worked at the inheritance of weight of
seeds in beans and in barley, and self-fertilised the
plants investigated for a series of generations so as
to isolate what he calls 'pure lines.' He found that
in beans, for example, the seed-weights of a mixed
population gave a normal frequency curve—the
weights varied continuously and evenly about a mean
value. The beans on an individual plant when the
flowers are self-fertilised also form a normal curve
about a mean, but this mean is not necessarily identical
with that of the race in general. If now the flowers
on such an individual are self-fertilised, and the beans
produced are sown, the mean weight of the beans on
all the daughter plants will be identical with the
mean of the beans on the parent, i.e. among the
offspring produced by self-fertilisation there is no
regression towards the mean of the race. It thus
makes no difference whether large or small seeds are
chosen *within the pure line* ; the mean weight of the
seeds on plants grown from the smallest and largest of
the parental beans (seeds) is in each case equal to the

parental mean.   Selection, therefore, within the pure line has no effect in altering the mean weight of the seeds, for the differences in seed weight within the line are not inherited.   The probable cause of this is that the differences between the seeds on a self-fertilised plant are due to the action of external circumstances ; the position of the beans in the pod or the position of the pods on the plant cause differences in the nutrition which allow some beans to grow larger than others.   These differences are 'acquired characters,' and we have here additional evidence that such are not inherited.   It is to variation of this type that the term 'fluctuation' is applied by some authorities.

It is clear then that if selection is made among beans harvested from a mixed population, on the whole the larger beans will belong to pure lines having a higher mean, and thus selection for a few generations will isolate pure lines having a high value, and the mean of successive generations will rise until the largest pure lines have been isolated. Beyond that point further selection will have no effect.   This is precisely the result arrived at by Prof. Pearson from a study of selection within a mixed population ; the mean will rise rapidly on the first selection, more slowly later, until in very few generations it reaches a point at which selection has no appreciable effect.   Pearson calculates that if

selection now ceases, the selected race will very
slowly revert towards the mean of the general popu-
lation. But, as has been seen, this conclusion is based
on the assumption that continuous variation is due to
the concurrent action of an indefinite number of small
independent causes. If, however, Johannsen is correct,
we may divide these causes into two classes: the
causes which induce 'fluctuation' as explained above,
which agree with Pearson's requirements, and the
cause or causes which give rise to the difference
between one pure line and another. Now this second
group may conceivably consist in a single factor of
the nature of a small 'mutation,' and if so, by isolating
the pure line this factor is also isolated, and no return
towards the mean of the general population need
take place. According to Johannsen this isolation
can be effected in one generation by selecting the
self-fertilised plants which have the highest average
yield, instead of selecting the heaviest beans them-
selves.

We thus obtain by experiments such as those of
Johannsen a new conception of the possible nature of
continuous variation; it may be due partly to 'fluctua-
tion' brought about by the action of environment
and not inherited, partly to a series of small step-
wise 'mutations,' each of which owing to fluctuation
overlaps the next, and can only be isolated when it is
possible to breed pure lines. It should be said that

there is as yet no certainty that this account of continuous variation is sufficient to cover all cases; it is a suggestion of possibility rather than a statement of fact.

We have seen that there is reason to believe that the Law of Ancestral Inheritance is true only when applied to a large number of individuals considered in mass, or, as it has been put, that it is a statistical rather than a physiological law. In individual cases it is not true that the offspring need be influenced by ancestors beyond the parents, but in other cases, as will be seen in dealing with Mendelian heredity, these ancestors have important effects, so that statistically it is possible to say what is the average influence of the ancestors of any generation upon the offspring. Now in cases where it is possible to define rigidly single characters, much more is learned from the physiological than from the statistical method, but where no such rigid separation of characters is possible the statistical law is the only one that can be applied. This is particularly the case in characters which vary continuously, or where the categories into which the character falls overlap one another, as for example in Johannsen's beans. Further, the statistical method is frequently the only one which is available when experiment is impossible and when our knowledge of the facts is based solely on numerical data from observed cases, and this of

course applies especially to inheritance in Man,
where experimental evidence is not available.

By collecting family histories of distinguished
men, Galton showed long ago [15] that exceptional
mental qualities were inherited; and this work
has recently been much extended and made more
definite by Professor Pearson and his school. It
is commonly believed that exceptionally gifted men
do not have distinguished sons, but this like many
other popular beliefs is only partly true. It
has been seen that if an individual deviates a
certain amount from the general mean, his children
will on the average deviate less, because when the
whole ancestry is taken into account, the effect of
previous generations is to cause regression on the
mean of the population. And since the theory of
regression depends on the assumption that variation
is due to the existence of a large number of inde-
pendent causes acting concurrently, it is unlikely
that among the limited number of offspring of one
exceptional man any one child will unite in himself
the same combination of factors as went to make up
the father's character. Further, it is improbable that
an unusually gifted man will marry a wife equally
gifted in the same manner, and the mother's influence
on the children is closely similar to that of the father.
It cannot therefore be expected that all great men
should have equally great sons, but they are far more

likely to have exceptional sons than are mediocre
men, and if the mother is also exceptional in the
same direction this probability is greatly increased.

In the last few years the intensity of inheritance
in such characters has been given numerical ex-
pression. Professor Pearson, after working out the
statistical laws of inheritance in many physical char-
acters of man, animals and plants, has applied the
same methods to what are called the mental and
moral attributes. Characters were chosen such as
vivacity, popularity, conscientiousness, temper, ability,
hand-writing, which were estimated by reports from
school-teachers on the children in their schools;
and also intellectual ability as shown in university
examinations or by the position in a public school
at a particular age (Schuster [10])[1]. All these, when
investigated by the same methods as were devised
for the coat-colour of horses or eye-colour in man,
are found to give results closely in accord with those
obtained for physical features. The conclusion is
therefore reached that not only bodily characters,
but also those of the mind are essentially determined
by the hereditary endowment received from the
parents. This result is of great importance practi-

[1] In these characters the resemblance between parent and child
cannot of course be estimated directly, but it has been pointed out
above that the resemblance between brothers may be used as a test
of the intensity of heredity.

cally; it shows how little room is left in the development of the individual for the effects of environment even on the intellect or mind in the broadest sense of the word; no doubt the direction which intellectual development takes is to a considerable extent determined by circumstances, but the kind of mind is irrevocably decided before the child is born. Still less is there room for the inheritance of the mental acquirements made by the individual during his life, and hence the hopes held out of improving the race by education and by special care of the dull or feeble-minded are illusory, except in so far as they improve the *tradition.* Just as the welfare of the race may be increased by an invention which is handed on from generation to generation, so the good effects of education or other improved conditions may be handed on, but this is not heredity. The father may educate his children because he himself was educated, but the mental powers of his children will be the same whether he had a good education or none[1]. And the effects of special care given to the weakly or feeble-minded may be absolutely harmful to the race, if the improvement so effected leads to more frequent

[1] Of course education is a necessary condition for the full development of the mental powers, but at present we have no evidence that it can add potentialities not present at birth. The subject is more fully discussed in Chapters VII and VIII.

marriage among such unfortunates than would otherwise be the case, for then an increased number of defective children may be born, and the race-average be lowered. Hence has arisen the study known as 'Eugenics,' the study, that is, of the methods by which the race may be improved both physically and mentally. The whole trend of the results obtained is that in order to produce exceptionally gifted men in both body and mind, those with high development of the characters desired should be encouraged to marry; and that to prevent the production of the weakly and feeble-minded, the only method is to prevent such from having offspring. It is admitted that at present these things hardly come within 'practical politics,' but there is little doubt that the nation which first finds a way to make them practical will in a very short time be the leader of the world.