96

THE GENETICAL INTERPRETATION OF STATISTICS OF THE THIRD DEGREE IN THE STUDY OF QUANTITATIVE INHERITANCE

R. A. FISHER,[1] F. R. IMMER,[2] AND OLOF TEDIN[3]
Rothamsted Experimental Station, Harpenden, Herts, England

Received July 2, 1931

For the past 20 years genetical methods have gradually been made more and more familiar to the practical breeders of plants and animals, upon whom the improvement for human use of the domesticated animals and cultivated plants finally depends. During this period it has become increasingly clear that the hereditary mechanism is well represented by the Mendelian scheme, as extended mainly by the work of the Drosophilists. It has been equally clear, however, that in all the practical problems of animal or plant improvement we are invariably faced with quantitative characters, which have shown themselves to be entirely intractable by the familiar genetical methods. These methods rest primarily upon the recognition of the effects of different single factors, and when these can be recognized the study of their effects in combination follows as a matter of routine. When individual factors cannot be recognized the analytic method of genetic study cannot even be commenced, and the question arises as to whether genetics as a science has any further resource to offer.

The successes of analytic genetics have been obtained mainly with the numerous deleterious recessives which are abundant in most species, with certain easily recognizable characters of practical importance to the plant or animal breeder and with fancy characters such as the crest, or silky plumage in the fowl, which, however attractive to fanciers, cannot be regarded as of general utility to mankind. The development of the quantitative characters on which practical utility is founded owes very little to genetic analysis except in so far as it has been demonstrated that it depends on a definite gene complex. This is implicit in the study of the individual effects of Mendelian factors without the means of evaluating the mass effects of a large number of minor factors severally influencing the utility character. We would stress, however, that the study of the metrical characters is not only of utilitarian interest. The nature of the heritable

[1] Chief Statistician at the ROTHAMSTED EXPERIMENTAL STATION, Harpenden, Herts, England.
[2] Fellow of the NATIONAL RESEARCH COUNCIL, and Associate Geneticist in the Office of Sugar Plants, UNITED STATES DEPARTMENT OF AGRICULTURE, University Farm, St. Paul, Minnesota.
[3] In charge of the breeding of barley and fodder roots at the WEIBULLSHOLM PLANT BREEDING STATION, Landskrona, Sweden.

elements which cause continuous or fluctuating variability must indeed be studied if progress in this direction is to be made possible; but such studies are also essential for an understanding of the evolutionary process by which organisms have been brought to their present state of organization and adaptation.

Most of the many workers who have attacked the problem of quantitative characters have carried their work far enough to verify that the heritable variability available was probably due to a large number of Mendelian factors, interacting generally in a cumulative manner. The principal criterion has been the greater variance of the F_2 as compared to the F_1 sample. These samples are, however, generally of very different magnitude, and care has not always been taken that the conditions of culture have been such as to make the comparison a valid one. It would seem essential for the purpose that the non-genetic causes of variability should be carefully equalized. With plants, the cultures to be compared should be grown in the same year, and if, as is probable, the F_2 is the larger culture, its variability should be estimated only within areas of the same size as that occupied by the F_1; for, with uniform seed, we can always obtain a higher variability by using a larger area.

In the case of this comparison of variance, a biometrical technique has been used to verify a genetical conclusion. In seeking for further points in the genetical situation which biometrical methods, combined with an adequate cultural technique, might be able to evaluate, it should be borne in mind that, as a school, the biometricians have shown themselves singularly unreceptive to genetical ideas. Methods which are genetically appropriate will not therefore be found ready made, and constants such as the correlation coefficient, which have been introduced with the highest biometrical testimonials, while they probably have, in suitable cases, an appropriate use, have assuredly done as much to confuse as they have to clarify the subject.

In studying the properties of a system of interacting factors it has been shown (FISHER 1918) that departures from the simple additive law of interaction will usually have effects somewhat similiar to non-heritable modifications. We may therefore be confident that, even if a strictly additive interaction is not exactly realized, the mass effects of segregation in a large number of factors will closely simulate those of simple cumulative systems. In such a system certain special quantities, of which the mean and the variance are examples, possess the remarkable property that each is simply compounded of contributions derived from the several factors acting singly. Thus the heritable variance observable among any group of

organisms may be regarded as the sum of the variances due to the individual factors. The portion of the variance which is heritable may be easily estimated from the covariances or mean products of the measurements of related individuals so that, without being able to recognize any single factor, we have a direct means of estimating their total contribution to the heritable variance.

This, however, is not enough to evaluate the selective potentialities of the population under examination. A number of further questions present themselves, most of which must at present remain unanswered. The same total variance might be contributed by a few factors each having a relatively large effect or by a multitude of smaller modifiers. In both cases progress can be made at once by selection, but whereas in the first case such progress will soon be accompanied by a decrease in the variance available, and will therefore soon be slowed down, in the second case progress can be continued much further in the same direction without the introduction of fresh material. Equally important, and fortunately less elusive, is the incidence of dominance; for mass selection will be far more successful in establishing recessives having a desirable effect than in establishing dominants of like effect, and this provides an obvious reason why, in material which has already, consciously or unconsciously, been much selected, the recessives are generally found to be variants in the direction which is judged to be disadvantageous.

We shall give some examples of the kind of data in which this bias in the prevalent direction of dominance, which selection must tend to introduce, appears to be shown. For the moment let us notice that effective biometrical methods of evaluating this bias will be of immediate practical value in the evaluation of the selective potentialities of a given population. From the purely scientific point of view it is also of importance that the dominance bias constitutes an existing record of the prevalent direction in which selection has acted in the immediate past. A geologist, by examining the population of individuals existing at a given horizon, might be able not only to specify the mean value and the variance of any measurement in this population, but might have a direct indication of the direction in which this measurement was in process of change.

STATISTICS OF THE THIRD DEGREE

In the study of the various methods by which the effects of biassed dominance may be brought to light, we shall be invariably led to the use of statistics of the third degree. Our knowledge of these quantities on the algebraic side is at present very incomplete. The birth of modern statistics

during the past generation may be typified by the transfer of attention from the totals and means characteristic of simple accountancy to the statistics of the second degree, the sums of squares and products, on which the whole apparatus of the calculus of correlations, or in more recent times of the analysis of variance and covariance, has been built. Statistics of the third and higher degrees have, of course, been used in fitting frequency curves and surfaces, but merely to evaluate empirical and arbitrarily chosen mathematical constants; and their practical inappropriateness for this purpose has been shown by their low efficiency in, for example, fitting the parameters of the Pearsonian curves. There is at present no comprehensive method of handling the statistics of the third degree analogous to the analysis of variance and covariance, to which nearly all work with second degree statistics can be reduced. Consequently, the methods we shall illustrate will probably be found to be capable of much improvement, and no exactitude can be claimed for the estimates of sampling error, or in consequence for the tests of significance. This, however, is a drawback which we may expect to be remedied with equal pace with the improvement of the experimental data, to which these methods may be applied.

In the case of statistics of the second degree we distinguish between the variance derived from the squares of the values of a single variate, and the covariance derived from the products of the values of the two different variates. The corresponding statistics of the third degree are of three kinds: (a) those derived from the cubes of the values of a single variate, (b) those derived from the product of one variate with the square of a second, and (c) those derived from the product of three different variates. It will be seen that all three types are, in different cases, of value. For each type we must throw our calculations in such a form as to obtain an unbiassed estimate of some parameter which, like the variance, satisfies the cumulative property, and which in consequence is interpretable in terms of the individual factors of the Mendelizing system.

For example, from a series of values of a single variate we can calculate the three statistics of the first, second and third degrees, namely,

$$k_1 = \frac{1}{n} S(x)$$

$$k_2 = \frac{1}{n-1}\left\{S(x^2) - \frac{1}{n} S^2(x)\right\}$$

$$k_3 = \frac{n}{(n-1)(n-2)}\left\{S(x^3) - \frac{3}{n} S(x^2)S(x) + \frac{2}{n^2} S^3(x)\right\}$$

where S( ) stands for summation over the sample observed, and n is the sample number, which are equivalent if $\bar{x}$ is the mean, to

$$k_1 = \bar{x}$$

$$k_2 = \frac{1}{n-1} S(x - \bar{x})^2$$

$$k_3 = \frac{n}{(n-1)(n-2)} S(x - \bar{x})^3.$$

Then it has been shown that k_1, k_2, k_3 are unbiassed estimates of quantities K_1, K_2, K_3 characteristic of the population sampled and possessing the cumulative property.

A. B. D. FORTUYN (1931, p. 163) gives eight seriations for the frequencies of different numbers of tailrings in different strains of mice, derived from *Mus musculus*, *Mus wagneri* and hybrids between these two forms. He was able to show that the variation in ring number was largely hereditary, for by selection from a common stock he obtained strains with average ring numbers, 142.6 and 216.2 respectively. Selection for high values of a variate should, when applied to a symmetrical population, generally shift the value of k_3 in the negative direction; equally, selection for low values should shift it in the positive direction. The amount of these changes will depend on the number of factors present. In an ideal case in which selection in opposite directions was applied to the F_2 from two homozygous lines, so that all pairs of allelomorphs were present initially in a 1:1 ratio, the ratio of the change in k_3 to that effected in the mean k_1 should be initially

$$\frac{-\frac{1}{2}S(d^4)}{S(d^2)}$$

where 2d is the difference between the homozygous forms in any one factor, and S stands for summation over the different factors. The rate at which the third moment is modified for a given change in the mean is evidently greater, other things being equal, the smaller the number of factors to the segregation of which the variance of F_2 is to be ascribed. As it stands it affords therefore a crude method of estimating or at least of setting a lower limit to the number of factors present. Although FORTUYN'S material was not formed as an F_2 from homozygous lines, it may be of interest to point out that the high selection line has in fact a negative k_3, though, on the number counted, not a significant value. Of the seven lines given, however, two, both with high ring numbers, do show significantly negative k_3, while a third with low ring number gives a k_3 which is significantly positive. The

phenomenon to be expected thus does not seem to be beyond attainable precision. The seven values are as follows:

TABLE 1

	k_1	k_3	g	n
Mus wagneri	138.8	−66.4	−.095±.228	113
L T M	142.6	−36.2	−.072±.229	111
B W T W	150.1	+2162.1	+.581±.199	149
W T W	159.0	+467.3	+.278±.182	179
Albino *Mus musculus*	189.3	−1194.0	−.526±.122	446
W M	195.0	−1667.6	−.584±.160	230
H T M	216.2	−238.5	−.258±.143	290

The best available test for the significance of k_3 (FISHER 1928) seems to be to calculate the ratio $g=k_3\,k_2^{-3/2}$; then for sampling from a normal population the true variance of g is $6n(n-1)/(n-2)(n+1)(n+3)$ and its distribution is, for samples over 100, sufficiently near to normality for significance to be inferred from the standard errors as shown in the table. The standard error will be used throughout this paper in testing significance.

This example is illustrative only of the type of biometrical effect which is not beyond experimental precision, by which direct information may be obtained as to the distribution of the heritable variance among the genetic factors present. The interpretation of any particular body of data for which this effect was measured could evidently be carried much further by estimating also such quantities as the heritable variance and those unsymmetrical effects ascribable to dominance. These latter will indeed, in practice, almost always complicate the interpretation of any data bearing on the size and number of the heritable factors.

THE EFFECTS OF DOMINANCE BIAS ON F_3 PROGENIES

The observational facts that the cross (F_1) between two strains frequently shows greater "vigour," or growth rate, than either parental type and that inbred lines frequently show a falling off in size, which is reversible by a single cross, may be interpreted either on the view that there is among the genetic factors present a pronounced bias in dominance, in the sense that greater size is more usually dominant to less size, than *vice versa*, or, on the contrary, that the heterozygote in a single factor is frequently larger than either of the corresponding homozygotes. These two views differ considerably in their practical consequences, but the contrast may be reduced to the quantitative question of the normal position of the heterozygote in a single factor relative to the two homozygotes.

Let us suppose that two homozygotes differing in any one factor differ on the average in the metrical character under observation by a quantity 2d, so that the mean values for the two homozygotes differ from an arbitrary origin (the mid-point between the two homozygotes) by +d and −d; we may then represent the average deviation of the heterozygote from the same origin by h. If h is generally positive, or at least generally positive for the more important factors, there will be in the system of factors considered a positive bias of dominance. The heterozygote for the whole group of factors will then exceed the mean of any two complementary homozygotes from which it might have been obtained. For any factor, if h lies between the limits −d and +d, dominance will be partial or incomplete, if it is equal to ±d dominance will be complete, but if it exceeds +d we shall have a case of superdominance. We shall consider how the biometrical data from F_3 progenies may be used to calculate whether the factors present as a whole have values of h which are positive, and if so whether there is evidence that they exceed the value d.

Although the main object of this paper is to call attention to the significance of various statistics of the third degree, yet it will be convenient here to state briefly some second degree results, which, though long known in principle, have not, we believe, been developed in a form convenient for experimental utilization. The three phases of any factor, if fully viable, may be expected in F_2 in the ratio 1:2:1. It easily follows that the contribution of such a factor to the variance in F_2 will be $\frac{1}{4}(2d^2+h^2)$, when the deviations are measured from the mean, $\frac{1}{2}h$. The total observable variance in F_2 does not, however, provide a satisfactory basis for evaluating directly the sum of these quantities, since, in all quantitative characters which are susceptible to environmental influences, a positive contribution will be made by environmental modification, and it does not appear that any experimental refinement could altogether eliminate this source of error. In the case of the covariance, on the other hand, the environmental deviations will be equally frequently positive and negative, and will only lower the precision of the result by increasing the quantities upon which the estimate of error is based. The covariance is calculated from $S(x-\bar{x})(y-\bar{y})/n-1$ where x and y are the two variates and n is the sample number. With much plant material two types of covariance may be fairly readily obtained: (1) the covariance between the F_2 parent and the mean of the F_3 progeny derived from it, the contribution of each independent factor to which is $\frac{1}{4}(2d^2+\frac{1}{2}h^2)$ and (2) the covariance of parent and progeny when the F_2 are crossed *inter se* at random, the value in this case being $\frac{1}{4}d^2$. From these two quantities we can, with precision limited only by the homozygosity of

the parent stocks, determine the part of the F_2 variance which is genotypic merely by taking twice their difference. What is equally interesting, the method shows a way of separately evaluating the ratio of the mean or average of the quantities h^2 to that of the quantities d^2, and so of discriminating between the hypothesis that the system of cumulative factors is one in which dominance is generally absent or slight, as is often assumed, and the hypothesis more generally favored when hybrid vigor is manifest that dominance is as complete in the quantitative factors as it generally is in factors which can be isolated for separate study.

Statistics of the second degree can obviously not distinguish whether h is positive or negative; they cannot therefore be used to investigate the extent to which dominance is biassed. Indications of this may be obtained from the first degree statistics, the means, as when an F_1 exceeds the mean of the parental values, and we infer that h is more frequently or more largely positive than negative. The comparisons of mean values, of which the most important is the difference between F_1 and F_2 is, however, a matter of some experimental difficulty, especially when the number of F_1 seeds is small, and, though they would be valuable in conjunction with other facts, by themselves they are not capable of more than a qualitative interpretation. From a set of F_3 progenies, however, it is possible to obtain three statistics of the third degree, which have a direct relevance. These are:

(1) The mean value of the k_3 from each of the progenies, to which each factor contributes $(-3/8)d^2h$.

(2) The covariance of the k_1 and k_2 in different F_3 progenies, to which each factor contributes $+h(2d^2+h^2)/32$.

(3) The k_3 of the means of different F_3 progenies, to which each factor contributes $(-3/8)d^2h$.

In respect of availability we shall show that no very extensive data are required to estimate the first of these quantities, while a larger number of progenies than in the examples to be given, with the exception of the barley data, though not an impossibly large number, would be needed to obtain good values for the second. The third would evidently be liable to large disturbances owing to the varying fertility of the areas upon which different progenies must be grown, and is in any case liable to much larger sampling errors than is the first. The first process will therefore always give the preferable value. What is important, however, is that, by a comparison of the mean value of k_3 with the covariance of k_1 and k_2, it is possible directly to distinguish between the views that an apparent effect of heterosis is due to ordinary dominance, either complete or incomplete, favoring the

larger values, in which case a homozygote may be established as vigorous as any heterozygote, and the alternative view that in many factors the heterozygote is more vigorous than either homozygote. If we multiply the covariance by 4 its value will be greater than, equal to, or less than the mean value of k_3 (with sign reversed) according as h (supposed positive) is greater than, equal to, or less than $+d$. With true superdominance four times the covariance should have a positive value exceeding the negative average value of k_3 within F_3 progenies, while if we are confronted only with a strong positive bias of the dominance, it should at most be equal to this value. If the system were of so simple a kind that in each factor the heterozygote was equal to the larger homozygote, we should find confirmation of the fact from the equality of $S(h^2)$ and $S(d^2)$, and should know that only by specific interactions could the average be raised above the level of the multiple heterozygote. Equally, if the covariance is less than this critical value, it is clear that the possibilities of mass selection have not been exhausted.

An example may be taken from the distribution of leaf length for 13 F_3 families of lettuce given by C. E. DURST (1930, p. 266). The mean leaf length in F_1 was greater than that of either parent, while the mean in F_2

TABLE 2

NUMBER OF INDIVIDUALS	k_1	k_2	k_3
21	4.286	3.014	1.427
8	6.000	13.143	4.572
39	7.590	1.143	−0.720
26	7.615	7.126	−2.381
17	7.647	3.742	−1.146
50	8.180	6.559	−3.122
25	8.480	10.927	−24.565
47	8.936	4.061	−2.660
13	9.539	5.102	−0.975
13	9.846	8.974	−20.700
35	9.886	8.104	−9.715
53	10.585	2.786	−5.631
11	12.000	8.200	−18.333

was slightly shorter than the larger leaved parent; by this indirect comparison the mean of F_2 may be judged to be about 2.2 units less than in F_1, one unit being 1.5 cm. The numbers of individuals and the values of k_1, k_2 and k_3 obtained for the 13 F_3 progenies are shown in table 2, differences of 1.5 cm being taken as one unit, the first unit being at 10 cm.

From these we find directly the mean value of k_3 to be -6.458 ± 2.532, a negative value, as the theory has indicated, which is statistically significant, but owing to the small number of families, not well determined numerically. For the covariance of k_1 and k_2 we have $+0.492$, which is, in accordance with the theory of cumulative factors, positive, though no statistical significance is to be attached to this fact as its standard error is as high as 2.017.

The k_3 of the means of the thirteen F_3 lines is found to be -3.476 ± 6.159, a negative value, again in agreement with the theory but not significant. The sampling variance of the mean k_3 has been obtained from the formula $\{1/(n-1)n\}\ S(k_3-\bar{k}_3)^2$, where n is the number of k_3's calculated, the variance of the k_3 of the F_3 means from $\{6n/(n-1)(n-2)\}\ k_2^3$ and the variance of the covariance of k_1 and k_2 from $(1/n-1)\{V(k_1)V(k_2)+V^2(k_1k_2)\}$ where $V(k_1)$, $V(k_2)$ stand for the estimated variance of k_1 and k_2 in different families and $V(k_1k_2)$ for their estimated covariance.

It will be observed that very different numbers of plants were obtained in the different F_3 families, and consequently the precision of the statistics derived from them must be expected to vary greatly. We may anticipate in general that the best possible theoretical estimates will be something between those obtained by giving, as above, equal weights to all families, and the corresponding values obtained by weighting each in proportion to the number of plants recorded. To ascertain whether in this case weighting would give appreciably increased precision, the values were recalculated on the latter system. The standard error of the mean k_3 was reduced from 2.532 to 2.087, and that of the covariance from 2.017 to 1.427, showing that it will be found a decided advantage in the use of such data for our present purpose if the F_3 progenies are approximately equal in number.

This advantage is emphasized equally by the circumstance that the statistics derived from the means of a limited number of plants will be affected by the k_3 of their distribution about their means. Both the covariance and the k_3 of the means require for this reason a correction algebraically equal to $-\bar{k}_3/s$ where s is the number of plants per family. With variable family numbers it will doubtless be sufficient in applying this correction to use the harmonic mean of the actual numbers, but the existence of a correction of this kind is a sufficient reason for keeping the numbers of the F_3 families as nearly constant as is conveniently possible.

The fourfold value of the covariance of k_1 and k_2 is here less than the mean k_3, indicating that h is generally less than $+d$; this is true even when the considerable correction $+.360$ is added to the crude value $+.492$. No significance can, however, in this case be attached to the comparison, con-

sidering the magnitude of the standard errors. It is of interest to note also that the k_2 of the F_2 progenies, 9.536, was almost twice that of the weighted k_2 of the F_3 progenies, that is, 5.468. The contribution of a single factor to the variance of the former would be $\frac{1}{4}(2d^2+h^2)$, and to the latter $\frac{1}{8}(2d^2+h^2)$. The comparison of the two variances thus suggests that a large proportion of the observed variance was in both cases genetic in origin.

For a critical study far more extensive data would be needed. The data given here are recognized to be inadequate except to point out how the problem may be attacked. Not only would more plants be required in F_2 and the F_3 lines but the number of F_3 lines should also be increased greatly in order that these should adequately sample the segregation in F_2. With the use of sufficiently extensive data it should be possible to reduce the standard errors to limits which would permit of exact comparisons between the different statistics and to fix the prevailing ratio of h and d within reasonable bounds. Replication, as a means of reducing the effect of soil variation, would be highly desirable.

Another example will be taken from published data on inheritance of a quantitative character in maize. EMERSON and EAST (1911, p. 77) presented data on inheritance of height of maize plants from a cross of Tom Thumb pop and Missouri dent, two open pollinated varieties. A guess mean was taken at 18 dcm, 1 dcm was taken as a unit and the following table computed from the sixteen different F_3 distributions given:

TABLE 3

NUMBER OF INDIVIDUALS	k_1	k_2	k_3
40	−7.225	2.076	0.597
114	−5.114	6.102	3.145
64	−6.016	2.524	0.106
65	−0.785	2.734	−3.155
90	2.011	4.685	−2.975
85	0.306	6.310	2.698
82	1.488	6.944	−4.716
85	0.635	6.162	−7.731
82	1.598	3.503	−0.268
95	4.432	5.567	−1.393
87	1.805	5.182	−5.226
149	3.745	3.070	−0.657
87	4.207	5.422	−2.954
93	4.774	4.764	4.943
87	5.862	2.981	−0.162
76	7.145	5.512	−0.011

The k_2 of the F_2 was 12.499 ± 1.415 and the k_3 of the F_2 was -15.694 ± 8.666. From the table we find the mean k_2 of the F_3 lines to be 4.596 ± 0.389 and the mean k_3 of the F_3 to be -1.110 ± 0.821. The latter is negative as expected for a dominance bias, but cannot be considered statistically significant with the data available. The covariance of k_1 and k_2 is found to be $+1.925 \pm 1.795$, a positive value as expected, but again not significant. Twice the value of the k_2 of the F_3 is here even less than the k_2 of the F_2. The variance of the mean k_2 of the F_3 lines was obtained from the formula $\{1/n(n-1)\}S(k_2-\bar{k}_2)^2$ where n is the number of F_3 families.

In the summary of the paper by EMERSON and EAST we find a statement relative to the "lack of skewness in the F_2 frequency distributions" for the crosses employed in studies on inheritance of height of plants. Taking the total distributions for F_2 in tables 25, 26, 27, 28 and 30 and the same for the F_2 in 1911 in table 29 we may determine statistically whether these distributions were or were not symmetrical. We find that $(k_3 k_2^{-3/2})$ was -0.105 ± 0.120, -0.254 ± 0.096, 0.028 ± 0.161, 0.012 ± 0.105, -0.355 ± 0.192 and -0.161 ± 0.106 in tables 25 to 30, respectively. A negative bias is indicated in four of the six crosses tested; although in the second alone is the skewness statistically significant, yet two other negative values are suggestively large.

Early studies of heterosis emphasized the fact that if dominance of growth factors were the explanation of hybrid vigor the F_2 distribution would be skew. Such was not the case, it was argued. With a large number of growth factors the F_2 distribution would tend toward the normal, and large numbers would be needed to demonstrate a negative bias. That such a negative bias is fairly common is demonstrated by the negative k_3 found from the data examined in this paper.

The maize data were obtained from a cross of two open pollinated varieties which were undoubtedly heterozygous in many of their factors for height. The data are given only to show how the problem may be attacked. Maize offers unusual possibilities for biometrical studies on quantitative inheritance. Some of the advantages of maize will be enumerated.

Selfed lines are already available which may be considered homozygous for the greater part of their growth factors. Crosses are easily made and a large number of seeds usually obtained. It would be possible, usually, to self a sufficient number of F_2 plants to obtain seed for tests in F_3 and to leave sufficient seed of the F_2 generation for comparison of the F_2 with the F_3 progenies in the following year. Since maize is very highly cross pollinated, seed from the various F_2 plants may be planted in an isolated plot also and allowed to cross *inter se*.

Parent lines, F_1, F_2, F_3 and lines from F_2 plants crossed *inter se* can then be grown in a single year in a replicated yield trial. The value of careful replication cannot be over-emphasized for such a study as a means of eliminating soil variation. Until more information is available on this point it may be suggested that, say, 50 to 100 plants in each of 100 F_3 lines ought to give sufficient data for a critical study. These might well be grown in five replicated plots of 10 to 20 plants each.

The contribution of a single factor difference to the following statistics may be noted:

Statistics of the second degree

Variance of F_2	$\frac{1}{4}(2d^2+h^2)$
Mean variance of F_3 progenies	$\frac{1}{8}(2d^2+h^2)$
Variance of means of F_3 progenies	$\frac{1}{4}(2d^2+\frac{1}{4}h^2)$
Covariance of F_2 parental value with mean of its F_3 offspring	$\frac{1}{4}(2d^2+\frac{1}{2}h^2)$
Covariance of F_2 parental value with mean of its biparental offspring	$\frac{1}{4}d^2$
Mean variance of biparental progenies	$\frac{1}{16}(4d^2+3h^2)$
Variance of means of biparental progenies	$\frac{1}{16}(4d^2+h^2)$
Mean variance of maternal progenies	$\frac{1}{8}(3d^2+2h^2)$
Variance of means of maternal progenies	$\frac{1}{8}d^2$

The biparental progenies are those obtained by crossing two F_2 plants. The maternal progenies are taken to be the progenies of plants exposed to open pollination by sister F_2 plants. Whether the multiplicity of measures also affords a method of measuring and eliminating the effects of incomplete linkage, we have not investigated; it would seem premature to attempt this until biometrical studies with visibly classifiable factors have shown what kind of disturbance is to be looked for.

Of the above statistics the four variances of individual values will usually be sensibly increased by environmental modification; the variances of the means could be freed from this bias, if the progenies are not grown on separate areas. The two covariances should be free from bias, and the experimental arrangement could be devoted solely to diminishing their sampling errors.

Statistics of the third degree

k_3 of F_2	$-\frac{3}{4}hd^2$
Mean k_3 of F_3 progenies	$-\frac{3}{8}hd^2$
Covariance of means and variances of F_3	$(h/32)\ (2d^2+h^2)$
Covariance of F_2 parental value and variance of F_3	$(h/16)\ (2d^2+h^2)$
k_3 of means of F_3 progenies	$-\frac{3}{8}hd^2$
Mean k_3 of biparental progenies	$-\frac{3}{16}hd^2$
Covariance of means and variances of biparental progenies	$-\frac{1}{8}hd^2$
Covariance of parental values and variances of biparental progenies	$-(h/32)\ (2d^2-h^2)$
Covariance of biparental progeny and biparental product	$-\frac{1}{8}hd^2$
k_3 of means of biparental progenies	$-\frac{3}{16}hd^2$
Mean k_3 of maternal progenies	$-\frac{3}{8}hd^2$

Covariance of means and variances of maternal progenies	$-\frac{1}{8}hd^2$
Covariance of parental values and variances of maternal progenies	$-\frac{3}{16}hd^2$
k_3 of means of maternal progenies	0

It will be noticed that the same quantities may be obtained experimentally in many different ways. This should afford a valuable empirical test as to the consistency of the genetical interpretation and will also prove valuable in detecting and eliminating the metrical bias, to be discussed later, with which statistics of the third degree, apparently with the exception of the covariance of the biparental progeny mean with the biparental product, may be affected. For the moment we need only notice that since this bias is always of the same sign, significant values of opposite signs cannot but indicate genuine genetical effects.

Some of these comparisons will be subject to greater errors than others. The previous examples show that the mean k_3 of F_3 lines is determined with much greater precision than the k_3 of the means, or the covariance of k_1 and k_2. One of the objects to be aimed at in exploratory work of this kind must always be to ascertain which biometrical values, of those which have an important genetical interpretation, can be ascertained in practice with a useful degree of precision.

In any case when both parental values are known as well as the mean of the progeny, the effect of this negative correlation of the progeny with the parental product may be exhibited by a purely biometrical procedure, one in which the relationships of the different parent stocks are not taken into consideration. We may in fact always calculate the regression of progeny values upon the three "independent" variates, maternal value, paternal value and parental product. Such a partial regression equation was calculated from data on row number from 46 backcrosses of maize kindly supplied by E. W. LINDSTROM. From data on row number of both parents and mean row number of the progeny the partial regression of mean progeny row number was found to be given by $0.4418x + 0.5148y - 0.0364p$ where x and y are the maternal and paternal row numbers and p is the parental product. The covariance of mean of progeny on the product of parental deviations was negative as expected, but the value shown for partial regression was not significant, having in this case a standard error $\pm .0369$. The regression of progeny on mother or father was about one-half as expected.

There are, in general, two important sources of disturbance in biometrical studies concerned with a study of quantitative inheritance. The first may be termed the dominance bias. When dominance favours the larger values the segregating generations will tend to have a negative skewness

(as measured by k_3). The extent of this bias will depend solely on whether dominance is complete or incomplete. The maximum skewness would be obtained when dominance is complete. With h greater or less than +d this skewness would be reduced. It is from the properties connecting h and d in different types of distributions and matings that these quantities may be estimated, as shown previously.

Another source of bias is encountered also very frequently. This may be termed the metrical bias, since it depends on the scale of measurements used. It is a not infrequent observation that the standard error of yield of a group of plots on a field of low fertility is often higher than that of plots on a high fertility field. The yields of plots from a low fertility field vary more with slight variations in soil fertility than do plots on a field which is already producing nearer to its maximum. The k_3 of yields from a group of plots varying in fertility would, in such cases, be negatively skew.

An example of this "inherent" negative bias will be taken from measurements of height of barley plants grown in different plots, with different nitrogen fertilizers, at the ROTHAMSTED EXPERIMENTAL STATION in 1928. The heights of sixteen plants were measured in each of 24 plots, the plants being selected at random within each plot. The mean height in cm to the auricle of the last expanded leaf for all observations was 37.6175 cm. The mean k_2 for the 24 plots of 16 plants each was 67.107 ± 2.078 and the mean k_3 of the same 24 plots was -194.099 ± 61.231. The covariance of k_1 and k_2 was -14.369 ± 13.248. A negative skewness is evident, indicating an inherent negative metrical bias. The covariance is also negative, again indicating the same type of disturbance. It is not significant, however. Since a normal variety of barley was used, and barley is very highly self fertilized, this bias cannot be attributed to genetic segregation for plant height factors.

These two sources of bias, the dominance bias and the metrical bias, will tend to counteract one another in the covariance of k_1 and k_2 of the lines in genetical studies of quantitative inheritance. The extent of the disturbance due to the metrical bias will depend somewhat on which statistics are used. It will be greatest in the covariance of k_1 and k_2 and will not be as serious a source of disturbance to the mean k_3.

Extensive data on height of barley plants from a cross of a two row (*Hordeum distichum nutans*) and a so-called six row variety (*Hordeum tetrastichum*) made by TEDIN and grown at Weibullsholm, Sweden, in 1930, will be used to illustrate the type of information which may be obtained from such data. This study was made entirely on F_3 material. Two hundred and seventy F_3 lines, segregating for the factors two row versus six row were

grown and individual height measurements obtained on 14,759 plants in these lines. The number of plants varied considerably in the different lines but weighting was not attempted.

The mean height of all the F_3 plants was 87.4 cm. Taking 5 cm as unit the mean k_2 of the 270 F_3 lines was 3.475 ± 0.077 and the mean k_3 -2.176 ± 0.204, a negative quantity in accordance with the theory that dominance of factors for plant height should give a negative k_3. This quantity was highly significant owing to the ample data on which it was based. Another third degree statistic was also available, that of the k_3 of the means of the F_3 lines. This was found to be -0.932 ± 0.567, a negative quantity as expected but not statistically significant. Since replication of the F_3 lines was not used this quantity would not be expected to be determined very accurately. The third statistic of the third degree calculated was the covariance of k_1 and k_2. This was found to be -0.564 ± 0.124, a negative quantity that was highly significant. If the dominance bias had been operating alone the covariance should be positive. The question arises as to whether the clearly significant negative value for the mean k_3 is to be ascribed to the co-operation of both measures of bias, or principally or entirely to the metrical factor.

For moderate bias it may be anticipated that the bias in the mean of the k_3 of the progenies will be proportional to six times the mean square of their variances, that is, to 81.8478, while the corresponding bias in the covariance of the means and variances of the families will be proportional to four times the product of the mean variance of the progenies and the variance of their means. This comes to 33.6060 or more than one-third of that found for the mean of k_3. Since the negative value found for the covariance, when corrected, as before explained, for the limited number measured in each family, is just less than a quarter of the mean value of k_3, we may conclude that the negative value of the covariance may be wholly accounted for by metrical bias without abolishing at the same time the negative values for k_3. In this case, however, it is clear that metrical bias has been a major factor in both values, and even the ample material measured would not allow us to attach significance to the residual genetic effects remaining after the metrical bias had been removed. It would, of course, be not surprising with a normally self fertilized plant, such as barley, that the pronounced bias in the dominance of genetic factors, comparable with that indicated for maize, should be absent or inconspicuous.

The average height of the 2 row, heterozygous (that is, for the 2 versus 6 row factor pair) and 6 row plants in the F_3 lines segregating for the 2 versus 6 row factor pair was found to be 88.2405, 88.0035 and 83.1190 cm,

respectively. The heterozygous group of plants cannot be considered significantly different in height from the 2 row plants. Apparently, therefore, as far as height factors linked with the 2 versus 6 row factor pair are concerned, no evidence of superdominance could be found. It seemed of interest to determine next the average variance (k_2) of height of the 2 row, heterozygous and 6 row plants when these were segregates in the same F_3 lines and grown on the same areas of land. The average k_2's were 87.6675, 84.1100 and 70.9825 square cm, respectively, for these three groups. The mean k_2 of the 2 row plants was 3.5575 ± 3.2275 square cm greater than the k_2 of the heterozygotes and 16.6850 ± 3.5950 square cm greater than that of the 6 row plants. The mean k_2 of the heterozygotes was 13.1275 ± 2.6500 square cm greater than the k_2 of the 6 row plants. The first of these differences is not statistically significant but the latter two are. When, however, the standard errors ($\sqrt{k_2}$) were expressed in percent of the mean heights, values of 10.61, 10.42 and 10.14 percent were obtained for the 2 row, heterozygous and 6 row groups. Apparently when the standard errors of the heights of these three classes of segregates, grown on the same small plots of land, were expressed in percent of the mean height the coefficient of variability was practically constant for the three groups.

SUMMARY

1. A genetical interpretation is given for various second and third moment statistics which are of use in studying quantitative inheritance.

2. Published data taken from lettuce and maize, and unpublished data from barley crosses are used to illustrate how the problem may be attacked. The special needs of data adequate for this purpose are illustrated, and certain possible precautions in planning the experiments are pointed out.

3. A study of the skewness of seven distributions for strains of mice selected for high and low tailring number indicated that the theoretical negative association between the statistics k_1 and k_3 in selected strains could probably be evaluated.

4. Formulae are given by which the effect of the dominance bias in the heterozygote in relation to the measurable characters of the homozygotes in F_2 or F_3 distributions or various types of crosses may be calculated.

5. The two common sources of bias (metrical and dominance) are discussed and data from a barley cross used to illustrate the results obtained when the former is of major importance.

6. Since the combined effect of the dominance and metrical biases may be obtained experimentally in many different ways, an empirical test of

the consistency of the genetical interpretations is available, as well as an opportunity of evaluating and eliminating the metrical bias.

7. Standard errors are given for the different statistics used.

LITERATURE CITED

DURST, CHARLES E., 1930 Inheritance in lettuce. Illinois Agric. Expt. Sta. Bull. 356.

EMERSON, R. A., and EAST, E. M. 1912 The inheritance of quantitative characters in maize. Nebraska Agric. Expt. Sta. Res. Bull. 2.

FISHER, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinb. **52:** 399–433.

1928 Moments and product moments of sampling distributions. Proc. Lond. Math. Soc. **30:** 199–238.

1930 The moments of the distribution for normal samples of measures of departure from normality. Proc. Roy. Soc. A **130:** 16–28.

FORTUYN, A. B. DROOGLEEVER, 1931 *Mus musculus* and *Mus wagneri* compared. I. The number of tailrings. Genetics **16:** 160–167.