Theor. Appl. Genet. 56, 119-131 (1980)

Random and Fixed Effects in Plant Genetics*

C.C. Cockerham**
Department of Statistics, North Carolina State University, Raleigh, North Carolina (USA)

Summary. A general model for any type of genetic entry is developed which takes into account both the factorial model of gene effects and the ancestral sources, whether inbred lines or outbred varieties, of the genes.

Utilizing the model, various genetic designs of fixed entries are explored for the estimation of genetic effects and the testing of genetic hypotheses. These designs consisted of generation means – parents, crosses, various types of backcrosses, and so on – stemming from one or more pairs of parents, and of hybrid combinations from factorial mating designs. Limitations, from the standpoint of genetic effects that can be estimated and genetic hypotheses that can be tested, are developed in considerable detail.

When entries from the factorial mating designs are considered to be random, attention is focused on the estimation of genetic variances, rather than effects, and on the concomitant changes in the tests of genetic hypotheses. While there is considerable improvement over fixed entries in the number of types of genetic variances that can be estimated, and of genetic hypotheses that can be tested, they are still very limited in contrast to what would be most desirable.

Key words: Genetic models – Means – Effects – Variances

Introduction

It is for estimation and hypothesis testing concerning gene action that I wish to review and compare some of the random and fixed entry approaches. Sometimes the breeder is working with populations such as varieties which are near-equilibrium populations. In this case many forms of breeding lead to individuals which can be considered to be random members of the population. The population then serves as a reference base with parameters for random members, and thus, the model for the individual with random gene effects. On the other hand, the breeder often has selected sets of material such as screened inbred lines or varieties, or such a varied collection of these, that it is hard to imagine that they could have any connection with some equilibrium population. Consequently, he tends to view his collection as a fixed set. Moreover, he may be interested only in possible derivatives of his material, and thus he has all of the genes of interest constituted in the material at hand, but of course not necessarily constituted in the best combinations.

All the methods to be considered are based on a general factorial model of gene effects for an unknown, but generally assumed to be not small, number of genes. These quantitative methods are often criticized on the basis that they are not critical in contrast to Mendelian methods for identifiable genes. Nothing could be truer. Much of the variation of interest to the plant breeder, however, is due to variation in genes which have not been identified, and moreover, are not easily identified.

On the other hand, some people appear to put great faith in quantitative methods, concluding and inferring far beyond the information available. Also, some methods simply do not provide the information asserted by the developers. I hope to develop and present some perspective in these matters.

* Paper No. 6018 of the Journal Series of the North Carolina Agricultural Research Service, Raleigh, North Carolina. This investigation was supported in part by NIH Research Grant No. GM 11546 from the National Institute of General Medical Sciences.

Preliminary: This paper was presented at the 7th International Biometric Conference. Since then it has come to various people's attention and I have been encouraged to give it a wider distribution. Except for editing, the paper is essentially as originally written.

0040-5752/80/0056/0119/$ 2.60

Models

The factorial model of gene effects (Kempthorne 1957) is reasonably complex even for genes at just two loci in diploids. It is presented here to show the basis for a general model for any type of genetic entry.

Notation	Description
x_i, x_j, y_k, y_l	Genes at loci x and y
$x_i x_j y_k y_l$	Genotype
Model	
$G_{ijkl} = \mu$	Genotypic value = mean
Additive and Dominance Effects	
$+ (a_i + a_j + d_{ij})$	+ (additive, a, and dominance, d, effects for locus x)
$+ (a_k + a_l + d_{kl})$	+ (effects for locus y)
[Epistatic (x by y) effects]	
$+ \{(aa)_{ik} + (aa)_{il} + (aa)_{jk} + (aa)_{jl}\}$	+ {additive by additive}
$+ \{(ad)_{ikl} + (ad)_{jkl} + (ad)_{kij} + (ad)_{lij}\}$	+ {additive by dominance}
$+ \{(dd)_{ijkl}\}$	+ {dominance by dominance}

It is convenient to summarize these effects, over an unknown number of loci, for individuals or entries – crosses, backcrosses, selfs and other generations – under consideration. This can be done along the lines of the factorial model of gene effects and in a manner which is descriptive of the genes as to their parental source. For additive effects we shall use $\sum_i \alpha_i A_i$, where i indexes the parental source (line, variety), $\alpha_i/2$ is the proportion of the genes in the entry under consideration from the ith source, A_i is the sum of the additive effects for the genes in a gamete from a parent of the ith source, $\sum_i \alpha_i = 2$. The corresponding term for dominance effects is $\sum_{i \leq j} \delta_{ij} D_{ij}$ where $\delta_{ij}(\delta_{ii})$ is the proportion of genotypes with alleles from parents from sources i and j(i), D_{ij} is the sum of the dominance effects for genes from the mating of a parent from source i with one from source j, and $\sum_{i \leq j} \delta_{ij} = 1$. The two sets of coefficients are related in that

$$\alpha_i = 2\delta_{ii} + \sum_j \delta_{ij}, \quad j \neq i.$$

It is only the δ's that have to be formulated for each type of entry. Now, a general model for the individual or entry can be written as

$$G = \mu + \sum_i \alpha_i A_i + \sum_{i \leq j} \delta_{ij} D_{ij} + (\sum_i \alpha_i A_i)^2 +$$

mean add. dom. add. by add.

$$+ (\sum_i \alpha_i A_i)(\sum_{i \leq j} \delta_{ij} D_{ij}) + (\sum_{i \leq j} \delta_{ij} D_{ij})^2 + \ldots .$$

add. by dom. dom. by dom.

The epistatic terms are further expanded to express their full content. For example,

$$(\sum_i \alpha_i A_i)^2 = \sum_i \alpha_i^2 (AA)_{ii} + 2 \sum_{i<j} \alpha_i \alpha_j (AA)_{ij}$$

and the additive by additive effects are distinguished as to the source of the genes. When coefficients such as 2 occur the effects are means of two distinct effects; in $(AA)_{ij}$ for x genes from i with y genes from j and for y genes from i with x genes from j. When all effects are added up for an entry the total should be

$$G = \mu + 2A + D + 4(AA) + 2(AD) + (DD) + \ldots .$$

Some examples will clarify the procedure and at first we consider the indexed parents to be homozygous lines. For the cross of parents i and j, $\delta_{ij} = 1$ and $\alpha_i = \alpha_j = \delta_{ij} = 1$.

$$G_{ij} = \mu + A_i + A_j + D_{ij} + (AA)_{ii} + 2(AA)_{ij} + (AA)_{jj} +$$
$$+ (AD)_{i(ij)} + (AD)_{j(ij)} + (DD)_{(ij)(ij)} + \ldots .$$

If we let j= i, we get the correct result for the ith parent, either in application to the above model or via $\delta_{ii} = 1$, $\alpha_i = 2\delta_{ii} = 2$,

$$G_i = G_{ii} = \mu + 2A_i + D_{ii} + 4(AA)_{ii} + 2(AD)_{i(ii)} +$$
$$+ (DD)_{(ii)(ii)} + \ldots .$$

Since the parents are homozygous, self-fertilization produces the same entry, $G_i^s = G_i$. Now consider the progeny from randomly mating members of the G_{ij}, denoted G_{ij}^r, or from selfing since the parents are homozygous. For $G_{ij}^s = G_{ij}^r$, $\delta_{ii} = \frac{1}{4}$, $\delta_{ij} = \frac{1}{2}$, $\delta_{jj} = \frac{1}{4}$, and the α's remain unchanged, of course. For an additional generation of self-fertilization, $G_{ij}^{ss} = G_{ij}^{rs}$, $\delta_{ii} = \frac{3}{8}$, $\delta_{ij} = \frac{1}{4}$ and $\delta_{jj} = \frac{3}{8}$, the δ's following the well-known consequences of inbreeding for any generation.

Next consider a three-way cross from three parents $G_{i(jk)}$. For this entry $\delta_{ij} = \delta_{ik} = \frac{1}{2}$, $\alpha_i = \delta_{ij} + \delta_{ik} = 1$, $\alpha_j = \delta_{ij} = \frac{1}{2}$, $\alpha_k = \delta_{ik} = \frac{1}{2}$. If the cross is selfed, $G_{i(jk)}^s$, $\delta_{ii} =$

$\delta_{ij} = \delta_{ik} = \frac{1}{4}$, $\delta_{jj} = \delta_{kk} = \frac{1}{8}$, but with random mating of members of $G_{i(jk)}$ to produce $G^r_{i(jk)}$, $\delta_{ii} = \delta_{ij} = \delta_{ik} = \frac{1}{4}$, $\delta_{jj} = \delta_{kk} = \frac{1}{16}$, $\delta_{jk} = \frac{1}{8}$. Further inbreeding is accommodated by operating on the δ's.

For a backcross, $G_{i(ij)}$, just let k = i in $G_{i(jk)}$ and $\delta_{ii} = \frac{1}{2}$, $\delta_{ij} = \frac{1}{2}$, $\alpha_i = \frac{3}{2}$, $\alpha_j = \frac{1}{2}$.

Higher degrees of crosses are easy to accommodate in principle. For $G_{(ij)(kl)}$, $\delta_{ik} = \delta_{il} = \delta_{jk} = \delta_{jl} = \frac{1}{4}$, and any of the parents may be set to be the same for generating crosses involving common parentage. Some of those involving common parentage reduce to other entries, of course. For example, $G_{(ii)(ii)} = G_{i(ii)} = G_{ii} = G_i$ and $G_{(ij)(ij)} = G^r_{ij} = G^s_{ij}$. From the cross $G_{i(j(kl)}$ with $\delta_{ij} = \frac{1}{2}$, $\delta_{ik} = \delta_{il} = \frac{1}{4}$ double and alternate backcrosses, and a host of other variations are easy to find.

The procedure should be clear by now how one can quickly write models involving any type of generation mean from a set of homozygous parents.

If the parents are not homozygous lines but are outbred populations (such as varieties) the same model representation can be used for any of the crosses, backcrosses and so on, as long as individuals are mated at random in making the crosses and no deliberate inbreeding such as self-fertilization is imposed. To distinguish this situation, we use H instead of G and note that we have the same operational simplicity for H as G except for G^s, i.e., $H_i = H_{ii} = H_{i(ii)} = H_{(ii)(ii)}$ or $H_{(ij)(ij)} = H^r_{ij}$ and so on. With inbred parents we had $G^s_{ij} = G^r_{ij}$, but this is not true of course for a variety. To take care of this feature for any H^s we introduce another type of δ, δ^s_{ij}, and another type of dominance effect D^s_{ii}. Rules in this case are not quite as straightforward. In going from H to H^s, the δ's for H are halved and the new δ^s_{ii}'s are apportioned appropriately as to parental sources as follows:

	H	H^s	H^{ss}
H_i	$\delta_{ii} = 1$	$\delta_{ii} = \delta^s_{ii} = \frac{1}{2}$	$\delta_{ii} = \frac{1}{4}$, $\delta^s_{ii} = \frac{3}{4}$
H_{ij}	$\delta_{ij} = 1$	$\delta_{ij} = \frac{1}{2}$, $\delta^s_{ii} = \delta^s_{jj} = \frac{1}{4}$	$\delta_{ij} = \frac{1}{4}$, $\delta^s_{ii} = \delta^s_{jj} = \frac{3}{8}$

The results for another generation of self-fertilization, H^{ss}, are also given. In all cases $\sum_i \delta^s_{ii} = F$ (the inbreeding coefficient), $\Sigma(\delta + \delta^s) = 1$ and $\alpha_i = 2(\delta_{ii} + \delta^s_{ii}) + \sum_j \delta_{ij}$ $(j \neq i)$. With selfing the model is more complex than for homozygous parents because of the introduction and expansion of D^s terms,

$$H^s_i = \mu + 2A_i + \tfrac{1}{2} D_{ii} + \tfrac{1}{2} D^s_{ii} + 4(AA)_{ii} + (AD)_{i(ii)} + (AD^s)_{i(ii)} + \tfrac{1}{4} (DD)_{(ii)(ii)} + \tfrac{1}{2} (DD^s)_{(ii)(ii)} + \tfrac{1}{4} (D^sD^s)_{(ii)(ii)} + \cdots .$$

A few other examples will illustrate the process. For H_{ij}, $\delta_{ij} = 1$ and for H^s_{ij}, $\delta^s_{ii} = \delta^s_{jj} = \frac{1}{4}$, $\delta_{ij} = \frac{1}{2}$; for H^r_{ij}, $\delta_{ii} = \delta_{jj} = \frac{1}{4}$, $\delta_{ij} = \frac{1}{2}$, and for H^{rs}_{ij}, $\delta_{ii} = \delta_{jj} = \frac{1}{8}$, $\delta_{ij} = \frac{1}{4}$, $\delta^s_{ii} = \delta^s_{jj} = \frac{1}{4}$; for $H_{i(jk)}$, $\delta_{ij} = \delta_{ik} = \frac{1}{2}$ and for $H^s_{i(jk)}$, $\delta_{ij} = \delta_{ik} = \delta^s_{ii} = \frac{1}{4}$, $\delta^s_{jj} = \delta^s_{kk} = \frac{1}{8}$. If an entry is self-fertilized an additional generation the operation of halving the δ's and introducing δ^s_{ii}'s is just repeated.

Although the new dominance effects introduced into the model involve only the parental identifications, the model is considerably more complicated. Eberhart and Gardner (1966) accounted for additive and dominance effects in their model for this situation by letting $a_i = 2A_i + D^s_{ii}$ and $d_i = D_{ii} - D^s_{ii}$. It seems preferable to maintain the model to correspond both with the factorial model of gene effects and the parental sources of the genes. While the model is overly parameterized, it is only in terms of the full model that one can explore estimable functions and the restrictions required to obtain estimators of certain effects. While some people are uncomfortable with dominance effects for homozygous lines, and rightly so if considering only homozygous lines, the effects must have enough definitional generality to accommodate the variety of genotypes to be encountered in the entire experimental population.

The expansion of the model into epistatic effects does not give the correct coefficients for these effects for linked genes in entries from parents which are themselves crosses. There is no simple way of including linkages in the formulation because the epistatic coefficients involve polynomial functions of linkage parameters which vary over sets of loci for each type of effect and it is the average that is used. These linkages do not affect the additive and dominance formulations. Linkage disequilibrium in parent varieties will also affect the results if there are epistatic effects but not with just dominance or additive effects.

Fixed Entries

For fixed entries one must cope with the specific effects for the genetic material at hand, whether in linear, quadratic, or higher order functions.

Generation Means

Two parents

The simplest example is one suggested by Mather (1949) in terms of scaling tests, although the basis was inherent in the work by Wright (1922). The experiment consists of two parents, G_1 and G_2, their cross, G_{12}, and the next generation from randomly mating G_{12} individuals (or selfing if the parents are homozygous), G^r_{12}. The models follow from the rules of the previous section.

$$G_1 = \mu + 2A_1 + D_{11} + 4(AA)_{11} + 2(AD)_{1(11)} + \ldots$$

$$G_2 = \mu + 2A_2 + D_{22} + 4(AA)_{22} + 2(AD)_{2(22)} + \ldots$$

$$G_{12} = \mu + A_1 + A_2 + D_{12} + (AA)_{11} + 2(AA)_{12} + (AA)_{22} + + (AD)_{1(12)} + (AD)_{2(12)} + \ldots$$

$$G^r_{12} = \mu + A_1 + A_2 + \frac{D_{11}}{4} + \frac{D_{12}}{2} + \frac{D_{22}}{4} + (AA)_{11} + + 2(AA)_{12} + (AA)_{22} + + \frac{(AD)_{1(11)}}{4} + \frac{(AD)_{1(12)}}{2} + \frac{(AD)_{1(22)}}{4} + + \frac{(AD)_{2(11)}}{4} + \frac{(AD)_{2(12)}}{2} + \frac{(AD)_{2(22)}}{4} + \ldots .$$

The (AD) interactions are included to show how these and higher order interactions are involved.

While this is a very simple example, it serves to illustrate most of the problems encountered. One problem is estimability of main effects when there are interactions (Finney 1948). In our situation, each successive term after A involves interactions of the genes. As is usually the case for estimation, one must enforce some conditions on the parameters in an overly parameterized model. In this connection, it is convenient to distinguish between *side conditions* placed on a category and *restrictions* placed on subsequent categories for estimability, as did Elston and Bush (1964). Side conditions placed on A's such as $A_1 = -A_2$ or $A_2 = 0$ do not affect the estimators of A since only the difference, $A_1 - A_2$, is estimable, but in general the restrictions (conditions) placed on higher order effects affect the estimators of lower order effects. Restrictions are often said to be placed so that the effects apply to the population of interest. We have four distinct populations. The arbitrariness is best exemplified in the placement of the mean, which through restrictions can be made to correspond to any one or any combination of the populations.

I shall advance an arbitrary approach. The approach is to fit successively terms in the model by least squares, beginning with the first, μ, ignoring the remaining terms and adjusting for previously fitted terms; generally called the forward Doolittle method. For the experiment under consideration, the side conditions $A_2 = -A_1$, $D_{22} = -D_{12} = D_{11}$ and $(AA)_{22} = -(AA)_{12} = (AA)_{11}$ are imposed at the appropriate stages, and the following orthogonal comparisons are found (hats denote observations and estimators):

	$\hat{G}_1$	$\hat{G}_2$	$\hat{G}_{12}$	$\hat{G}^r_{12}$
$\hat{\mu}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$\hat{A}$	$\frac{1}{4}$	$-\frac{1}{4}$	0	0
$\hat{D}$	$\frac{3}{11}$	$\frac{3}{11}$	$-\frac{5}{11}$	$-\frac{1}{11}$
$(\hat{AA})$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$-\frac{1}{2}$

We now take expectations of the estimators to see what functions of effects are involved,

$$\&\hat{\mu} = \mu + A_1 + A_2 + \tfrac{5}{16} D_{11} + \tfrac{3}{8} D_{12} + \tfrac{5}{16} D_{22} + \tfrac{3}{2} (AA)_{11} + + (AA)_{12} + \tfrac{3}{2} (AA)_{22} + \ldots$$

$$\&\hat{A} = \tfrac{1}{2} (A_1 - A_2) + \tfrac{1}{2} (D_{11} - D_{22}) + \tfrac{1}{2} [(AD)_{1(11)} - -(AD)_{2(22)}] + \ldots$$

$$\&\hat{D} = \tfrac{1}{4} (D_{11} - 2D_{12} + D_{22}) + \tfrac{6}{11} [(AA)_{11} - 2(AA)_{12} + + (AA)_{22}] + + \tfrac{1}{44} [23(AD)_{1(11)} - 22(AD)_{1(12)} - (AD)_{1(22)} - -(AD)_{2(11)} - 22(AD)_{2(12)} + 23(AD)_{2(22)}] + \ldots$$

$$\&(\hat{AA}) = \tfrac{1}{4} [(AA)_{11} - 2(AA)_{12} + (AA)_{22}] + \tfrac{1}{8} [(AD)_{1(11)} - -(AD)_{1(22)} - (AD)_{2(11)} + (AD)_{2(22)}] + \ldots .$$

For the leading term in each expression to constitute the expectation, then all other effects in the expression must sum to zero. This does not in general seem reasonable. The mean is for the average of the four generations in the experiment. The additive estimator is influenced by dominance and epistasis, the dominance estimator by epistasis but not additive, and the epistatic estimator by epistasis alone. The method is designed to do just that. Any test of significance of the hypothesis $\&(\hat{\ }) = 0$ is of course a function of the effects in $\&(\hat{\ })$. Only the test of the last hypothesis $\&(\hat{AA}) = 0$ is invariant with the fitting procedure.

An alternative procedure is to impose the same conditions as before, ignore effects of order (AD) and higher, and to fit the mean, additive, dominance, and additive by additive simultaneously by least squares. The results are

	$\hat{G}_1$	$\hat{G}_2$	$\hat{G}_{12}$	$\hat{G}^r_{12}$
$\hat{\mu}^*$	0	0	0	1
$\hat{A}^*$	$\frac{1}{4}$	$-\frac{1}{4}$	0	0
$\hat{D}^*$	0	0	−1	1
$(\hat{AA})^*$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$-\frac{1}{2}$

In this case the mean is for the G^r_{12}, that is $\&\hat{\mu}^* = G^r_{12}$. Neither the additive estimator nor the additive by additive estimator has changed, $\hat{A}^* = \hat{A}$ and $(\hat{AA})^* = (\hat{AA})$. The dominance estimator

$$\&\hat{D}^* = \tfrac{1}{4}(D_{11} - 2D_{12} + D_{22}) + \tfrac{1}{4}[(AD)_{1(11)} - 2(AD)_{1(12)} + (AD)_{1(22)} + (AD)_{2(11)} - 2(AD)_{2(12)} + (AD)_{2(22)}] + \ldots$$

now does not contain additive by additive effects, which is an advantage over the elimination method. However, the estimator is still a function of epistatic effects of the order (AD) and higher. Further, the tests of $\&\hat{D}^* = 0$ and $\&(\hat{AA})^* = 0$ are correlated. Given exactness of the model fitted, tests of $\&(\hat{\ })^* = 0$ are uniformly most powerful. In some sense this is nit-picking since only categorical functions of effects can be tested. This stems from the fact that the natural parameterization is much greater than the information available.

An extension of the experiment is to include advanced generations by self-fertilization, G_{12}^{rs} and so on, or backcross generations: first $G_{1(12)}$, $G_{2(12)}$; double $G_{1(1(12)}$, $G_{2(2(12)}$; alternate $G_{2(1(12)}$, $G_{1(2(12)}$; and so on; and/or self-fertilized generations of each. Hayman (1958) discusses the analysis in detail for a model including two-locus effects. By including more means, one can successively eliminate and test for higher order effects – one degree of freedom for (DD) if the two first backcrosses are included. There are decreasing returns from including additional generations, which can be seen by looking at the general model in more detail. This model encompasses all entries if the two parents are homozygous, and all but the generations derived by self-fertilization if the parents are equilibrium varieties.

$$G = \mu + \{\alpha_1 A_1 + \alpha_2 A_2\} + \{\delta_{11} D_{11} + \delta_{12} D_{12} + \delta_{22} D_{22}\} +$$
$$+ \{\alpha_1^2 (AA)_{11} + 2\alpha_1\alpha_2 (AA)_{12} + \alpha_2^2 (AA)_{22}\} +$$
$$+ \{\alpha_1\delta_{11}(AD)_{1(11)} + \alpha_1\delta_{12}(AD)_{1(12)} + \alpha_1\delta_{22}(AD)_{1(22)} + \alpha_2\delta_{11}(AD)_{2(11)} + \alpha_2\delta_{12}(AD)_{2(12)} + \alpha_2\delta_{22}(AD)_{2(22)}\} +$$
$$+ \{\delta_{11}^2 (DD)_{(11)(11)} + 2\delta_{11}\delta_{12}(DD)_{(11)(12)} + 2\delta_{11}\delta_{22}(DD)_{(11)(22)} + \delta_{12}^2 (DD)_{(12)(12)} + 2\delta_{12}\delta_{22}(DD)_{(12)(22)} + \delta_{22}^2 (DD)_{(22)(22)}\} + \ldots .$$

Since $\delta_{11} + \delta_{12} + \delta_{22} = 1$ for each entry the maximum variation in these is among the two parents and F_1. The other generations represent averages of these. Include in addition the functional correlation among the coefficients of different effects, most especially the fact that $\alpha_i = 2\delta_{ii} + \delta_{ij}$, and it is apparent that the successive elimination of categories removes a large portion of effects in the categories remaining. Also, it should be obvious that the effects removed, including higher order ones, remain in the category of removal.

One can clarify some of the functional relationships among the coefficients by reducing the model to a more manageable form through imposing restrictions on the effects. One can also examine the consequences of doing so which seems worthwhile at this juncture. By imposing restrictions of the form $A_2 = -A_1 = A^*$, $D_{22} = -D_{12} = D_{11} = D^*$, and so on,

$$G = \mu^* + \alpha^* A^* + \delta^* D^* + (\alpha^*)^2 (AA)^* + \alpha^*\delta^* (AD)^* + (\delta^*)^2 (DD)^* + \ldots$$

where $\alpha^* = \delta_{11} - \delta_{22}$ and $\delta^* = \delta_{11} + \delta_{22} - \delta_{12}$. Another way of viewing the result is that we have transformed the δ's to α^* and δ^*. That is,

$$\alpha_1 = 1 + \alpha^* \qquad \delta_{11} = \frac{1}{4} + \frac{\alpha^*}{2} + \frac{\delta^*}{4}$$

$$\alpha_2 = 1 - \alpha^* \qquad \delta_{12} = \frac{1}{2} - \frac{\delta^*}{2}$$

$$\delta_{22} = \frac{1}{4} - \frac{\alpha^*}{2} + \frac{\delta^*}{4}$$

If we make these substitutions into the original model we now see the constitution of the starred effects in terms of those for the original model.

$$\mu^* = \mu + A_1 + A_2 + \frac{D_{11}}{4} + \frac{D_{12}}{2} + \frac{D_{22}}{4} + (AA)_{11} + 2(AA)_{12} + (AA)_{22} + \ldots$$

$$A^* = A_1 - A_2 + \frac{1}{2}(D_{11} - D_{22}) + 2[(AA)_{11} - (AA)_{22}] + \frac{1}{4}[3(AD)_{1(11)} + 2(AD)_{1(12)} - (AD)_{1(22)} + (AD)_{2(11)} - 2(AD)_{2(12)} - 3(AD)_{2(22)}] + \ldots$$

$$D^* = \frac{1}{4}(D_{11} - 2D_{12} + D_{22}) + \frac{1}{4}[(AD)_{1(11)} - 2(AD)_{1(12)} + (AD)_{1(22)} + (AD)_{2(11)} - 2(AD)_{2(12)} + (AD)_{2(22)}] + \ldots$$

$$(AA)^* = (AA)_{11} - 2(AA)_{12} + (AA)_{22} + \frac{1}{2}[(AD)_{1(11)} - (AD)_{1(22)} - (AD)_{2(11)} + (AD)_{2(22)}] + \ldots .$$

These are sufficient to illustrate the procedure and the results. The leading term for each starred effect is of

course a function of effects with the same nomenclature, and the additional effects are of higher order interactions. The results may be viewed as a confounding of higher order effects. Nevertheless, the results show the influence on lower order effects of imposing restrictions on higher order effects. There is of course no good solution to the problem where the number of parameters exceeds the number of entries, and in an entangled manner. The coefficients of the starred effects are now much less correlated than before for many sets of entries. The procedure of successive elimination or of fitting effects simultaneously provides the same results as are obtained by operating on the general model.

Another approach, Sentz et al. (1954), to the analyses of these generation means is to consider only the means of parents, and of pairs of backcrosses, and then place all entries on a scale of heterozygosity ranging from zero for the parents to one for G_{12} (ranked on δ_{12}).

0	0.25	0.50	0.75	1.00
$\frac{1}{2}[G_1 + G_2]$	$\frac{1}{2}[G_{1(1(12)} + G_{2(2(12)}]$	G^r_{12}	$\frac{1}{2}[G_{1(2(12)} + G_{2(1(12)}]$	G_{12}
	G^{rs}_{12}	$\frac{1}{2}[G_{1(12)} + G_{2(12)}]$		
		G^{rr}_{12}		

Now, any deviations from linearity or differences among entries at the same level are due to epistatic effects. One may partition the deviations from linearity into quadratic and so on, but which of course are no more informative than the model estimators. This approach overlooks certain comparisons such as $G_{1(12)} - G_{2(12)} - [G_1 - G_2]/2$ which are functions of only epistatic effects. Except for these omissions, it provides the same composite test for epistasis. The method, of course, is directly related to that for inbreeding depression.

The effects of inbreeding, in particular self-fertilization, were summarized by Anderson and Kempthorne (1954). If we abstract out just successive generations from self-fertilization, the following expression holds in the absence of linkage,

$$S_t = S_2 - F_t D + F_t^2(DD) - F_t^3(DDD) + F_t^4(DDDD) - \ldots$$

whether the inital population is a single individual ($S_1 = G_{12}$) or an equilibrium population ($S_2 = H$). (t indexes the generation and F_t is the inbreeding coefficient relative to $F_2 = 0$, $F_1 = -1$.) Without epistasis there is a linear relationship among the generations, which is the same result as scaling entries according to level of heterozygosity because F measures relative degree of heterozygosity. Tests for epistasis are for deviations from linearity, but the limitations of the tests can be readily seen in this example since (1) only all dominance types of epistasis are involved, and (2) these effects are sums with all sorts of cancellation possible.

When the two parents are equilibrium varieties, self generations of each and of their H_{12} and H^r_{12} may be included in the experiment (Robinson and Cockerham 1961). Without the self generations the model is that given for homozygous parents, but with selfing, additional dominance terms must be introduced. Only dominance effects are included in the following table.

Original entries		Self entries	
H_1	: D_{11}	H^s_1	: $(D_{11} + D^s_{11})/2$
H_2	: D_{22}	H^s_2	: $(D_{22} + D^s_{22})/2$
$\frac{1}{2}(H_1 + H_2)$	: $\frac{D_{11} + D_{22}}{2}$	$\frac{1}{2}(H^s_1 + H^s_2)$	: $\frac{D_{11} + D_{22}}{4} + \frac{D^s_{11} + D^s_{22}}{4}$
H_{12}	: D_{12}	H^s_{12}	: $\frac{D_{12}}{2} + \frac{D^s_{11} + D^s_{22}}{4}$
H^r_{12}	: $\frac{D_{12}}{2} + \frac{D_{11} + D_{22}}{4}$	H^{rs}_{12}	: $\frac{D_{12}}{4} + \frac{D_{11} + D_{22}}{8} + \frac{D^s_{11} + D^s_{22}}{4}$

Note that $D_{12} - (D_{11} + D_{22})/2$ relates to heterosis, while $(D^s_{11} + D^s_{22} - D_{11} - D_{22})/2$ relates to inbreeding depression, but these become categorical descriptions when they are estimated, and a function of just dominance only when there is no epistasis.

If one considers just the two by three table (below dashed line) there are three degrees of freedom to test for epistatic effects, two for rows by columns (entry by originals vs. selfs) and one for quadratic effects among rows. Each includes additive by additive and higher order effects. Including selfs of each type of entry from equilibrium varieties does not increase the order of epistatic effects that can be eliminated and tested but contributes additional degrees of freedom at each stage.

Parents

The most important extension of these experiments is to include entries from a set of parents, that is, the parental entries, and all other types of entries for each pair of parents, generally called a diallel. A model encompassing any set of entries and a general analysis can be accomplished by the regression approach, although with considerable tedium in many cases. Hayman (1957) considered n homozygous parents and their $n(n-1)/2$ G_{ij}'s and G^r_{ij}'s. The parents could also be equilibrium varieties. After accounting for the mean, $(n-1)$ additive terms and $n(n-1)/2$ dominance terms, there are $n(n-1)/2$ deviations to test for epistasis composed of all types. Actually, the parents can be omitted and there are then only $n(n-3)/2$ deviations for epistasis.

When the parents are varieties and selfs are included, the following analysis was given by Gardner and Eberhart (1966), except they did not include H^{rs}_{ij}'s.

Diallel analysis (Varietal parents)	
Source	d.f.
Mean	1
Additive	n − 1
$(D^s_{ii} - D_{ii})$	n
$h_{ij} = D_{ij} - (D_{ii} + D_{jj})/2$	n(n − 1)/2
$\bar{h}$	1
h_i	(n − 1)
$(hh)_{ij}$	n(n − 3)/2
Epistasis	3n(n − 1)/2

When the H^{rs}_{ij}'s are not included n(n−1)/2 degrees of freedom are lost in the epistatic term. The parents may also be omitted with appropriate modification of the analysis. Gardner and Eberhart (1966) distinguished between inbreeding effects $(D^s_{ii} - D_{ii})$ (which could be partitioned into an average and deviations) and heterosis, h_{ij}, which was further partitioned into an average, and parent deviations and interactions. These categorical distinctions are descriptive, whatever the explanation in terms of gene action.

The experiments do not have to be confined to a diallel design but could be a factorial mating design between distinct sets of parents (Stuber and Moll 1969). One fits additive and dominance terms for each set. With G_{ij}'s and G^r_{ij}'s, it is the interaction of parents in the $(2G^r_{ij} - G_{ij})$ table, with or without parents, that provides the deviations for epistasis. It is easiest to illustrate for a factorial mating design without parental entries.

$(2G^r_{ij} - G_{ij})$ Table—Homozygous parents or varieties

	Parent 3	Parent 4
Parent 1	$\mu + A_1 + A_3 + \frac{D_{11} + D_{33}}{2} + \ldots$	$\mu + A_1 + A_4 + \frac{D_{11} + D_{44}}{2} + \ldots$
Parent 2	$\mu + A_2 + A_3 + \frac{D_{22} + D_{33}}{2} + \ldots$	$\mu + A_2 + A_4 + \frac{D_{22} + D_{44}}{2} + \ldots$

The same results are obtained for the $(2H^s_{ij} - H_{ij})$ table for varieties except the D's are D^s's. In either case the interaction comparison provides a test for epistasis since the additive and dominance effects cancel out.

Comments

In summary to this section on generation means, one can use the approach of successive elimination of effects or of fitting simultaneously the same set of effects. In either case tests are of categorical effects which include higher order effects. The situation is actually worse than it appears on the surface, because the effects in the categories are sums for an unknown number of genes and may be zero, or near so, from cancellation. These limitations can be severe for generations involving just a pair of parents, where the expectations of the particular epistatic comparisons may be zero yet there be lots of epistatic interactions among the genes. Hypotheses are limited to something like 'no demonstrable gene effects of certain types' rather than the more desired 'no gene effects of certain types.' If the test of effects at any stage by elimination is not significant, one cannot assume that the previous categories are free of these effects. For example, the test for epistasis may be insignificant, yet the categorical additive and dominance effects be entirely a function of epistasis. The limitations on the hypotheses would be much less severe when the analysis is of combined generations from several parents. In no case, however, are the estimators dependable for providing information about total gene effects or their subdivisions in the genetic material. The same problems of interpretation arise in the simultaneous fitting of effects.

While the experiments should involve more parents at the expense of more generations of material, they do somewhat different things. For an experiment constituted with a fixed number of generations, increasing its dimensions by using more parents (whether they are included in the experiment or not) decreases the limitations on the hypotheses to be tested and increases the power of the tests, but leads to no additional estimable functions or changes in the hypotheses that can be tested. Increasing the number of generations increases the number of testable hypotheses, but in a diminishing returns way because of the limitations in variation among the δ's. Just general tests for epistasis may be the most useful, since further partitioning and elimination of epistatic effects is accurate only for unlinked genes.

In all of the experiments, the parents, whether included in the experiment or not, must be in Hardy-Weinberg equilibrium. Homozygous parents are in this equilibrium. There is a tendency to conjecture that just any individual or collection of individuals may serve as parents. However, the comparison $G^s_{12} - (2G_{12} + G_{11} + G_{22})/4$ may involve dominance when the two individual parents are heterozygous. Linkage equilibrium was also assumed. (Homozygous parents are in linkage equilibrium.) Lack of linkage equilibrium in the varieties would affect only the epistatic effects.

Any time there are different types of entries they may have different error variances. Hayman (1958, 1960 and previously) and others have taken account of different error variances for each type of entry through a weighted least squares approach. This in principle mainly complicates the procedure, and chi-square tests are conditional on the estimated variances. An alternative, simple approach is to do an unweighted least squares and to parti-

tion the entry by replication or entry by environment, whichever is to serve as error, sums of squares into portions to correspond exactly to those for entries, which are valid χ^2's for testing the corresponding sums of squares among entries, whatever the true variances are. Of course, the power of tests is reduced in comparison to an overall pooled error, but the latter is not an acceptable alternative when the error variances differ to any extent.

Whether to use inbred lines or equilibrium varieties as parents may be dictated by the experimental material available. A pair of equilibrium varieties may be so nearly alike that the only differences found are due to inbreeding, and thus no information would be available on epistasis. Homozygous lines also can be similar, but their distinctness in genotype, with genes at frequencies of one, even when derived from the same population, suggests that they would generally be far superior to varieties in studies of this type. While the formal formulation for the two are similar it is emphasized again that the categorical effects are functions of gene differences in the material.

Hybrids

So far we have been concerned with pairwise combinations of genes among parents, which can be fractionated into various relative proportions of each parent by backcrossing, and each of these can be put further into different degrees of heterozygosity by random mating or selfing. There is the logical extension to three-way combinations (crosses), four-way combinations, and so on, each with a multitude of possible backcross and randomly mated generations, and each of which may be further inbred. A general formulation of the analysis for all these situations is of course impractical, but any particular experiment can be analyzed by regression analysis. There is merit in considering just the crosses, hybrids, because they are an integral part of many breeding programs, and because they lend themselves readily to identifiable partitioning in accordance with the mating design model.

While the parents could be varieties, there would seem to be no reason to use them, and we shall assume homozygous parents. Further, no parent shall enter more than once into any hybrid, to avoid backcrosses of any type.

Bauman (1959) introduced an experiment involving three parents, one of which he called a tester G_3, as a test for epistasis. The entries are G_{13}, G_{23} and $G_{3(12)}$, the latter being a three-way cross. Then,

$$\frac{G_{13} + G_{23}}{2} - G_{3(12)}$$

can be used to test for epistasis. Actually, each parent can be used as a tester to the other two, and by including the other single cross G_{12} and three-way hybrids $G_{2(13)}$ and $G_{1(23)}$, a combined test for epistasis with three degrees of freedom can be constructed.

Eberhart (1964) and Eberhart and Gardner (1966) found that by considering a joint analysis of single, three-way and double crosses from four parents, they could separate out fifteen degrees of freedom for epistasis, nine of which involved epistatic effects of an order higher than (AA).

Rawlings and Cockerham (1962a, b) developed analyses for all possible three-way crosses and double crosses from a single set of parents. While these analyses were developed in terms of random effects, only when the effects cancel out completely will they not appear in the expectations of mean squares. The analysis of three-way crosses provides tests for (AD) and higher order effects and that of the double crosses for (DD) and higher order effects.

The crosses do not have to be those for a single set of parents, but can be factorial among different sets of parents. The models and analyses are much easier to demonstrate for distinct sets of parents, where genetic effects are more exactly identified with the design effects. Also, the analyses of factorial designs are straightforward. The same models are used for single sets of parents; just the analyses are more complicated.

We shall designate the sets of parents A, B, C, ... with individual parents or lines as A_i, B_j, C_k, ... The same designations will be used for the design effects in the model for their progeny. The factorial mating design AB means the single cross progeny from mating each A_i parent to each B_j parent. The design model is

$$G_{ij} = \mu + A_i + B_j + (AB)_{ij}$$

and the design effects have the following translation in terms of genetic effects,

$$A_i = A_i + (AA)_{ii} + (AAA)_{iii} + \ldots$$

$$B_j = A_j + (AA)_{jj} + (AAA)_{jjj} + \ldots$$

$$(AB)_{ij} = D_{ij} + 2(AA)_{ij} + (AD)_{i(ij)} + (AD)_{j(ij)} + \\ + (DD)_{(ij)(ij)} + 3(AAA)_{iij} + 3(AAA)_{ijj} + \ldots .$$

The coefficients of certain effects are the number of distinct effects in the average. For example, there are two additive by additive effects for the genes x_iy_j and x_jy_i which are averaged in $(AA)_{ij}$.

For three-way crosses consider the mating design $A(BC)$,

$$G_{i(jk)} = \mu + A_i + B_j + C_k + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + \\ + (ABC)_{ijk}.$$

The translation of the design effects is

$$A_i = A_i + (AA)_{ii} + (AAA)_{iii} + \ldots$$

$$B_j = \tfrac{1}{2} A_j + \tfrac{1}{4} (AA)_{jj} + \tfrac{1}{8} (AAA)_{jjj} + \ldots$$

$$(AB)_{ij} = \tfrac{1}{2} D_{ij} + \tfrac{2}{2} (AA)_{ij} + \tfrac{1}{2} (AD)_{i(ij)} + \tfrac{1}{4} (AD)_{j(ij)} + \\ + \tfrac{1}{4} (DD)_{(ij)(ij)} + \tfrac{3}{2} (AAA)_{iij} + \tfrac{3}{4} (AAA)_{ijj} + \ldots$$

$$(BC)_{jk} = \tfrac{2}{4} (AA)_{jk} + \tfrac{3}{8} (AAA)_{jjk} + \tfrac{3}{8} (AAA)_{jkk} + \ldots$$

$$(ABC)_{ijk} = \tfrac{1}{4} (AD)_{k(ij)} + \tfrac{1}{4} (AD)_{j(ik)} + \tfrac{2}{4} (DD)_{(ij)(ik)} + \\ + \tfrac{6}{4} (AAA)_{ijk} + \ldots .$$

For C_k and $(AC)_{ik}$ substitute k for j in B_j and $(AB)_{ij}$, respectively. The numerators of the coefficients are the number of distinct effects in the average.

Double crosses, (AB) (CD), have more symmetry than three-way crosses.

$$G_{(ij)(kl)} = \mu + A_i + B_j + C_k + D_l + (AB)_{ij} + (AC)_{ik} + \\ + (AD)_{il} + (BC)_{jk} + (BD)_{jl} + (CD)_{kl} + \\ + (ABC)_{ijk} + (ABD)_{ijl} + (ACD)_{ikl} + \\ + (BCD)_{jkl} + (ABCD)_{ijkl}$$

$$A_i = \tfrac{1}{2} A_i + \tfrac{1}{4} (AA)_{ii} + \tfrac{1}{8} (AAA)_{iii} + \ldots$$

$$(AB)_{ij} = \tfrac{2}{4} (AA)_{ij} + \tfrac{3}{8} (AAA)_{iij} + \tfrac{3}{8} (AAA)_{ijj} + \ldots$$

$$(AC)_{ik} = \tfrac{1}{4} D_{ik} + \tfrac{2}{4} (AA)_{ik} + \tfrac{1}{8} (AD)_{i(ik)} + \tfrac{1}{8} (AD)_{k(ik)} \\ + \tfrac{1}{16} (DD)_{(ik)(ik)} + \tfrac{3}{8} (AAA)_{iik} + \\ + \tfrac{3}{8} (AAA)_{ikk} + \ldots$$

$$(ACD)_{ikl} = \tfrac{1}{8} (AD)_{k(il)} + \tfrac{1}{8} (AD)_{l(ik)} + \tfrac{2}{16} (DD)_{(ik)(il)} + \\ + \tfrac{6}{8} (AAA)_{ikl} + \ldots$$

$$(ABCD)_{ijkl} = \tfrac{2}{16} (DD)_{(ik)(jl)} + \tfrac{2}{16} (DD)_{(il)(jk)} + \\ + 0(AAA) + \ldots .$$

Interchangeable subscripts: A, B, C and D; (AB) and (CD); (AC), (AD), (BC) and (BD); and all three factor interactions.

By adding over the effects and taking into account the number in the sums one can see that they always add to 2A + D + 4(AA) + 2(AD) + (DD) + ... as they should for each progeny mean.

The identification with the design effects simplifies the analyses, because one can readily see what the elimination of design effects eliminates in terms of gene effects. For single crosses the elimination of A and B effects eliminates all additive and some additive types of epistatic effects, leaving in the deviations (AB) all of the dominance and most of the epistatic effects. In three-way hybrids, one can eliminte successively additive, $A + B + C$; dominance, $(AB) + (AC)$; and additive by additive, (BC); leaving in the remaining deviations, (ABC), additive by dominance, dominance by dominance and higher order epistatic effects. Thus, one can test for three-gene and higher order gene interactions, but there is a cost in the removal process, because only $\frac{1}{4}$ (AD), $\frac{1}{2}$ (DD), $\frac{3}{16}$ (AAA) + ... are left in the deviations. The process should be clear now, and in the double crosses, one can eliminate all three-gene interactions and test for fourgene, (DD), and higher order interactions.

Comments

Considerable clarification is also available about the categorical effects in each line of the analysis of variance. First, we can see the kinds of effects that are major contributors. It is correct, however, to say that A or B deviations are functions of just additive and additive types of epistatic effects? To do so is to restrict the interaction effects of the model to sum to zero in the usual ways for arriving at the analyses, and consequently to force the same summation restrictions over their contents in terms of gene effects, and exactly so for this experiment. While the major contributors to each category are clarified, there appears to be no more real justification to assume the categories to be completely free of other higher order gene effects than in the analysis of generation means, although in the latter case the gene effects had no neat organization.

There are a variety of other designs. In an experiment where a set of parents is mated to three-way hybrids [mating design $A(B(CD)]$ dominance by dominance interactions can be removed, with deviations of higher order effects. For the three-way and double crosses, there are three different orders or ways in which the parents can be combined, e.g., for three-way crosses, $A(BC)$, $B(CA)$ and $C(BA)$. Unfortunately, the inclusion of different orders only increases the degrees of freedom for some of the tests and does not essentially alter the types of effects that can be eliminated or tested for. There can be other logical categorical breakdowns of the effects such as A, (AB), (ABC) and so on as averages over orders and then interactions of these categories with orders.

There are also many variations of the designs when sets of parents are common. When a single set of parents is utilized we have (AA) the diallel, $A(AA)$ the triallel and $(AA)(AA)$ the quadriallel, these being the usual designs thought of for hybrids. The three different orders are required in these designs for a reasonably simple analysis. There are a host of other variations, $A(BB)$, $A(AB)$, $(AB)(CC)$, and so on.

The analysis given by Rawlings and Cockerham (1962a, b) for three-way and double cross hybrids was for average effects and effects by orders. The regression ap-

proach advocated here will lead to a different partitioning of some of the sums of squares, particularly for the three-way hybrids, but those for the higher order epistatic effects can be left the same or pooled.

Coanalyses of the different hybrid designs from the same sets of parents are also possible. The desirability of codesigns is a matter for investigation. The degree of epistatic effects for which one can test appears to be that available for the highest factor design. The inclusion of the lower factor designs provides more degrees of freedom for various tests, but again this may be at the expense of removing larger portions of the categories being tested.

When analysis is of a single design, linkage does not cause the same problem as in the case of generation means. Single crosses are not affected by linkage if the parents are homozygous, but they provide tests for only dominance and higher order effects. In the three-factor and four-factor designs each type of genetic effect is defined uniquely as to parents, and the effects can be written formally to include linkage parameters. Since the effects are individually eliminated, the kinds of effects being tested for at each stage remain unchanged. That is not to say that linkage will not affect the results, but that the nature of effects being tested remains the same. The epistatic effects are not the same from one type of hybrid to another from the same parents because of differing degrees of recombination, which is a disadvantage for codesigns. As far as just testing for epistatic effects, linkage can be ignored since the additive and dominance effects are not affected by linkage, as was pointed out for generation means.

Random Entries

So far we have been concerned only with gene effects as determined exactly by the genetic entries in the experiment. Quadratic functions (sums of squares or mean squares) are utilized as a means of making tests of certain composite hypotheses. Only the general nature of the gene effects in each case could be specified. Quadratic functions, on the other hand, provide a means for estimating the variances of these gene effects, or various components of genetic variance, providing appropriate experiments and assumptions can be made.

If parents, whether included in the experiment or not, are not assumed to be random samples, then quadratic functions have the same difficulties met with for linear differences and compounded with correlated gene effects. One may still want to estimate genetic variances in an advanced generation from a fixed set of parents. The same features are involved in the estimation of genetic variances in a population generated by a fixed set of parents as in the experiments from random parents and these will be considered together.

The estimation of genetic variances has been explored in considerable detail (Cockerham 1963). Only a few of the pertinent points will be presented.

To be emphasized is that estimates are of variances and covariances of effects in populations from which the experimental material are samples. The individual is the basic observational unit, whatever plot or other unit into which it may be summed or averaged, and it is the variances and covariances of the various types of individuals with which one must reckon. But to determine these variances and covariances of individuals, gene distributions and effects, and thus variances, must be defined for members of each kind of population. Then one can, in principle, formally define the variance covariance structure for all individuals in the sample in terms of variances and covariances of gene effects.

One assumption is made at the outset; linkage equilibrium in the population (or populations for different sets) of parents. That is not to say that this property must hold for the parents in the experiment, but for the population from which the parents are samples. For example, the property cannot hold for genes differing at two loci for any specific pair of parents but can on the average for random pairs of parents. Linkage equilibrium of genes in the population of parents also insures linkage equilibrium of genes in the cross populations or other generations, if parents are mated at random.

Linkage equilibrium refers to independence of genes at different loci. Inbreeding on the other hand causes correlations of allelic genes, and it was in this context that Professor Wright developed the inbreeding coefficient. Over different generations additive and dominance effects have different variances and are generally correlated. Harris (1964) found five quadratic functions required to express the covariances of inbred relatives for any number of alleles at a single locus. The number of quadratic functions required to accommodate just two-locus interactions makes the situation unduly complex. Consequently, it does not appear feasible to estimate genetic variances and covariances from a joint analysis of generations, if one wishes to include any generality in the model such as two-locus epistasis. There is no difficulty in estimating the covariances among relatives, the difficulty is in the interpretation of these covariances.

It is much simpler and more effective to concentrate on a single generation for the estimation of genetic variances (Cockerham 1963). The effects and variances are then defined for that generation but this is the problem that must be accommodated. Of the experiments considered for fixed entries, the hybrid designs are the most favorable for estimating genetic variances. We shall assume that parents in each set are samples from their respective linkage equilibrium populations, and then reduce the results to those when all parents are from the same population.

The gene effects and their variances are defined in terms of the hybrid populations, conditional on the frequencies of the genes in the parent populations, and for a model encompassing any number of alleles and loci. Formal expressions of these variances, which have been given by Stuber and Cockerham (1966), will not be written, but various sums of variances indicated in expectations.

For the *AB* design first consider the analysis in terms of random design effects.

Source	d.f.	Expectations of mean squares
A parents	n − 1	$\sigma^2_{AB} + m\sigma^2_A$
B parents	m − 1	$\sigma^2_{AB} + n\sigma^2_B$
$A \times B$	(n − 1) (m − 1)	σ^2_{AB}

Since various gene effects were identified with the design effects, the expectations of the design components of variance are expanded, taking into account the coefficients and the numbers of gene effects. Genes are all assumed to be independent so that all the effects are uncorrelated, and the square of any type of effect has an expectation equal to the variance of that type,

$$\sigma^2_A = (\sigma^2_A + \sigma^2_{AA} + \sigma^2_{AAA} + \ldots)_A$$
$$\sigma^2_B = (\sigma^2_A + \sigma^2_{AA} + \sigma^2_{AAA} + \ldots)_B$$
$$\sigma^2_{AB} = (\sigma^2_D + 2\sigma^2_{AA} + 2\sigma^2_{AD} + \sigma^2_{DD} + 6\sigma^2_{AAA} + \ldots)_{AB}$$

If the parents are from different populations, similarly designated components of the variance are not to be equated among rows. For example, $\sigma^2_{A_A}$ is the variance due to additive effects of genes in the cross from parental population *A*. Parental population *B* may have genes at different frequencies. Thus the variances are defined specifically into the three categories. It is arbitrary whether certain components such as the additive by additive be defined as $2\sigma^2_{AA_{AB}}$ or $\sigma^2_{AA_{AB}}$. The coefficients were included to indicate the result for all parents from a single population, in which case all variances of a type are summed into one.

The usual method for arriving at these expectations is in terms of covariances of relatives (Cockerham 1963). Let C_A be the covariance of individuals with only an *A* parent common and C_{AB} of different individuals with both parents common. Then

$$C = \alpha_A \sigma^2_{A_A} + \alpha_B \sigma^2_{A_B} + \alpha_A \alpha_B \sigma^2_{D_{AB}} + \alpha^2_A \sigma^2_{AA_A} + \alpha_A \alpha_B 2\sigma^2_{AA_{AB}} + \alpha^2_B \sigma^2_{AA_B} + \alpha_A \alpha^2_B \sigma^2_{AD_{BA}} + \alpha^2_A \alpha_B \sigma^2_{AD_{AB}} + \alpha^2_A \alpha^2_B \sigma^2_{DD_{AB}} + \ldots$$

where $\alpha_A = 1, \alpha_B = 0$ for C_A; $\alpha_A = 0, \alpha_B = 1$ for C_B; and $\alpha_A = 1, \alpha_B = 1$ for C_{AB}. The covariances are related to the design components of variance as follows:

$$\sigma^2_A = C_A, \quad \sigma^2_B = C_B, \quad \sigma^2_{AB} = C_{AB} - C_A - C_B.$$

Application of these covariances leads to the same expectations.

When both sets of parents are samples from the same population, then variances of the same type are equatable and are summed into a total, e.g., $\sigma^2_{A_A} = \sigma^2_{A_B} = \sigma^2_A/2$, $\sigma^2_{AA_A} = \sigma^2_{AA_B} = (2\sigma^2_{AA_{AB}})/2 = \sigma^2_{AA}/4$, and so on. Then the covariances of relatives have their usual expression for a single population,

$$C = \alpha\sigma^2_A + \delta\sigma^2_D + \alpha^2\sigma^2_{AA} + \alpha\delta\sigma^2_{AD} + \delta^2\sigma^2_D + \ldots$$

where $\alpha = (\alpha_A + \alpha_B)/2$, $\delta = \alpha_A\alpha_B$, and it is these covariances that are obtained in the diallel.

More details on gene effects, variances and covariances of relatives in population crosses are given by Schnell (1965) and Stuber and Cockerham (1966). The parents do not have to be homozygous, in which case the appropriate results are obtained by adjusting α_A and α_B.

We shall just outline the result for three-way hybrids. The analysis of variance and expectations of mean squares for the design effects are the usual ones.

Design *A(BC)*

Source	d.f.	Expectations of mean squares
A parents	p − 1	$\sigma^2_{ABC} + n\sigma^2_{AC} + m\sigma^2_{AB} + nm\sigma^2_A$
B grandparents	n − 1	$\sigma^2_{ABC} + p\sigma^2_{BC} + m\sigma^2_{AB} + mp\sigma^2_B$
C grandparents	m − 1	$\sigma^2_{ABC} + p\sigma^2_{BC} + n\sigma^2_{AC} + np\sigma^2_C$
$A \times B$	(p − 1) (n − 1)	$\sigma^2_{ABC} + m\sigma^2_{AB}$
$A \times C$	(p − 1) (m − 1)	$\sigma^2_{ABC} + n\sigma^2_{AC}$
$B \times C$	(n − 1) (m − 1)	$\sigma^2_{ABC} + p\sigma^2_{BC}$
$A \times B \times C$	(p − 1) (n − 1) (m − 1)	σ^2_{ABC}

$\sigma^2_A = C_A, \sigma^2_B = C_B, \sigma^2_C = C_C, \sigma^2_{AB} = C_{AB} - C_A - C_B$
$\sigma^2_{AC} = C_{AC} - C_A - C_C, \sigma^2_{BC} = C_{BC} - C_B - C_C$
$\sigma^2_{ABC} = C_{ABC} - C_{AC} - C_{AB} - C_{BC} + C_A + C_B + C_C$

At the bottom of the table the design variance components are translated into covariances of relatives, where the common parentages of the relatives are indicated by the subscripts. The general form for the covariance of relatives,

$$C = \alpha_A \sigma^2_{A_A} + \alpha_B \sigma^2_{A_B} + \alpha_C \sigma^2_{A_C} + \alpha_A \alpha_B \sigma^2_{D_{AB}} + \alpha_A \alpha_C \sigma^2_{D_{AC}} + \alpha^2_A \sigma^2_{AA_A} + \alpha_A \alpha_B 2\sigma^2_{AA_{AB}} + \alpha_A \alpha_C 2\sigma^2_{AA_{AC}} + \alpha^2_B \sigma^2_{AA_B} + \alpha_B \alpha_C \sigma^2_{AA_{BC}} + \alpha^2_C \sigma^2_{AA_C} + \ldots,$$

becomes very complicated and involves powers and products of all three α's in C_{ABC}. For this design $\alpha_A = 1, \alpha_B = \frac{1}{2}, \alpha_C = \frac{1}{2}$, and all α's except those with a subscript matching one for C are set to zero to find the covariance. The variances can also be found by taking expectation of the squares of the gene effects associated with the design effects.

$$\sigma^2_A = \left(\sigma^2_A + \sigma^2_{AA} + \sigma^2_{AAA} + \ldots\right)_A$$

$$\sigma^2_B = \left(\frac{\sigma^2_A}{4} + \frac{\sigma^2_{AA}}{16} + \frac{\sigma^2_{AAA}}{64} + \ldots\right)_B$$

$$\sigma^2_C = \left(\frac{\sigma^2_A}{4} + \frac{\sigma^2_{AA}}{16} + \frac{\sigma^2_{AAA}}{64} + \ldots\right)_C$$

$$\sigma^2_{AB} = \left(\frac{\sigma^2_D}{4} + \frac{2\sigma^2_{AA}}{4} + \frac{\sigma^2_{AD}}{4} + \frac{\sigma^2_{AD}}{16} + \frac{\sigma^2_{DD}}{16} + \frac{3\sigma^2_{AAA}}{4} + \frac{3\sigma^2_{AAA}}{16} + \ldots\right)_{AB}$$

$$\sigma^2_{AC} = \left(\frac{\sigma^2_D}{4} + \frac{2\sigma^2_{AA}}{4} + \frac{\sigma^2_{AD}}{4} + \frac{\sigma^2_{AD}}{16} + \frac{\sigma^2_{DD}}{16} + \frac{3\sigma^2_{AAA}}{4} + \frac{3\sigma^2_{AAA}}{16} + \ldots\right)_{AC}$$

$$\sigma^2_{BC} = \left(\frac{2\sigma^2_{AA}}{16} + \frac{3\sigma^2_{AAA}}{64} + \frac{3\sigma^2_{AAA}}{64} + \ldots\right)_{BC}$$

$$\sigma^2_{ABC} = \left(\frac{\sigma^2_{AD}}{16} + \frac{\sigma^2_{AD}}{16} + \frac{2\sigma^2_{DD}}{16} + \frac{6\sigma^2_{AAA}}{16} + \ldots\right)_{ABC}.$$

The variances have been separated so that the denominators correspond to products and powers of the α's.

When the parents are from the same population, the results may be expressed for total variances of each type in the same manner as for single crosses. Then,

$$C = \alpha\sigma^2_A + \delta\,\sigma^2_D + \alpha^2\,\sigma^2_{AA} + \alpha\delta\,\sigma^2_{AD} + \delta^2\,\sigma^2_{DD} + \ldots$$

where $\alpha = (\alpha_A + \alpha_B + \alpha_C)/2$, $\delta = \alpha_A(\alpha_B + \alpha_C)$, i.e., for C_A, $\alpha = \frac{1}{2}$, $\delta = 0$; for C_B and C_C, $\alpha = \frac{1}{4}$, $\delta = 0$; for C_{AB} and C_{AC}, $\alpha = \frac{3}{4}$, $\delta = \frac{1}{2}$, etc. This amounts to combining the genetic variance components of the same type and dividing each component by two to the power of the number of times that additive enters into its nomenclature. For example,

$$\sigma^2_A = \frac{\sigma^2_A}{2} + \frac{\sigma^2_{AA}}{4} + \frac{\sigma^2_{AAA}}{8} + \ldots$$

$$\sigma^2_{ABC} = \frac{\sigma^2_{AD}}{16} + \frac{\sigma^2_{DD}}{8} + \frac{3\sigma^2_{AAA}}{64} + \ldots.$$

The coefficients now represent the fractions of the total variances of each type in the population.

The extension of the analyses to the other hybrid designs is straightforward but tedious. Coanalyses among the designs from the same sets of parents are also possible. In particular, when the parents are samples from the same population, the covariances of all hybrids can be expressed as linear functions of components of genetic variance.

An additional assumption of no linkages has been assumed in some cases. Linkages do not affect the single cross results, or σ^2_A in design $A(BC)$. Linkages affect only the coefficients of the epistatic components and operate always to make the coefficients larger than those given.

Comments

One test of significance, that for the highest order effects for each design, is always the same for random or fixed effects. Here, the two procedures diverge. Genetic hypotheses that can be tested are confined to those for the design components of variance. Thus, there is still limited flexibility. One cannot, for example, test for additive or dominance variance alone because epistatic variances are always included. One can test for all additive types of epistatic variances, σ^2_{BC} in design $A(BC)$ or σ^2_{AB} and σ^2_{CD} in design $(AB)(CD)$. While all tests are of interest, some are more so than others.

When the parents are from the same population, and there is only one component of genetic variance of each type, one may be inclined to test stepwise up the table, eliminating those variances in the hypotheses which turned out to be not significant lower in the table. The problems in this case are the usual ones in making tests of significance.

One can now see for this model what gene effects are under test if just the error term were used to test each line in the table as would be the case for fixed effects. Indeed, the expectations of the mean squares involve functions of all types of gene effects except those previously eliminated.

An appeal of genetic variances is that one can obtain information on the portion of the total variance contributed by each type of effect. Unfortunately, the estimators involve complex packages, as indicated in the tests of design components of variance. Other estimators, such as linear functions of the design components of variance, can be constructed, but which can be accomplished in a more comprehensible manner in terms of covariances of relatives (which are found as linear functions of the design components of variance). In any case there are only so many not wholly dependent quadratic estimators, the number being less than or equal to the number of design

components of variance. One may restrict the genetic model to no more than the number of types of genetic variances that can be estimated. Alternatively, one may be satisfied to utilize estimators which are mostly additive, mostly dominance, mostly epistatic, and so on. Often not a great deal of confidence can be placed in the estimators when the standard errors are computed.

General Comments

The limitations in the tests for genetic effects in the analysis of generation means from a single pair of parents are obvious. The limitations are less severe of course in a joint analysis from several parents. Also, the parents do not have to be included, which is an advantage when one wants the information in terms of reasonably outbred material as is often the case with normally outcrossing species and the parents are homozygous lines. Some inbreeding or backcrossing must be accommodated always to provide the different generation means unless one uses three or more way combinations (hybrids) of parents.

For normally self-fertilizing species one may be more interested in the effects of genes as constituted in reasonably homozygous states, but again some differences in inbreeding are required. Also, there is often difficulty in crossing and in particular obtaining enough G_{ij} seed for an experiment. However, the $(2G_{ij}^{ss} - G_{ij}^{s})$ table can be used to test for epistasis just as illustrated for the $(2G_{ij}^{s} - G_{ij})$ table, but with smaller coefficients of many of the effects.

One wonders about the utility of varieties in such studies. Selected outbred varieties are often similar, but diverse collections can be put together. The inclusion of selfs of the varieties and their crosses allows the separation of heterotic and inbreeding depression effects but the inclusion of more varieties in the experiment at the expense of the selfs may provide a better test for epistasis, although with fewer degrees of freedom for the same total number of entries.

Where feasible, the three-way or higher degree hybrids, or in combination with the single crosses, have several advantages. They are all in the same noninbred generation. Gene effects are easily identified with the design effects, and the analyses are straightforward. Hybrids can be viewed from the standpoints of both random and fixed effects. Even though it may be hard to conceive of parental lines as a random sample of arbitrarily inbred lines from a randomly mating population, it may be possible to view them as a random sample from a population of selected lines in linkage equilibrium. There is certainly no harm in looking at the partitions of genetic variance, although one may not be able to set great store in them. They provide measures of relative amounts in contrast to presence or absence from tests of significance.

Literature

Anderson, V.L.; Kempthorne, O.: A model for the study of quantitative inheritance. Genetics **39**, 883-898 (1954)

Bauman, Loyal F.: Evidence of non-allelic gene interaction in determining yield, ear height, and kernel row number in corn. Agron. J. **51**, 531-534 (1959)

Cockerham, C. Clark: Estimation of genetic variances. Symposium on Statistical Genetics and Plant Breeding. NAS-NRC **982**, 53-94 (1963)

Eberhart, S.A.: Theoretical relations among single, three-way, and double cross hybrids. Biometrics **20**, 522-539 (1964)

Eberhart, S.A.; Gardner, C.O.: A general model for genetic effects. Biometrics **22**, 864-881 (1966)

Elston, R.C.; Bush, N.: The hypotheses that can be tested when there are interactions in an analysis of variance model. Biometrics **20**, 681-698 (1964)

Finney, D.J.: Main effects and interactions. J. Amer. Statist. Ass. **43**, 566-571 (1948)

Gardner, C.O.; Eberhart, S.A.: Analysis and interpretation of the variety cross diallel and related populations. Biometrics **22**, 439-452 (1966)

Harris, Dewey L.: Genotypic covariances between inbred relatives. Genetics **50**, 1319-1348 (1964)

Hayman, B.I.: Interaction, heterosis and diallel crosses. Genetics **42**, 336-355 (1957)

Hayman, B.I.: The separation of epistatic from additive and dominance variation in generation means. Heredity **12**, 371-390 (1958)

Hayman, B.I.: The theory and analysis of diallel crosses. III. Genetics **45**, 155-172 (1960)

Kempthorne, O.: An introduction to genetic statistics. New York: J. Wiley & Sons 1957

Mather, K.: Biometrical genetics. London: Methuen and Co. 1949

Rawlings, J.O; Cockerham, C. Clark: Triallel analysis. Crop Sc. **2**, 228-231 (1962a)

Rawlings, J.O; Cockerham, C. Clark: Analysis of double cross hybrid populations. Biometrics **18**, 229-244 (1962b)

Robinson, H.F.; Cockerham, C. Clark: Heterosis and inbreeding depression in populations involving two open-pollinated varieties of maize. Crop Sc. **1**, 68-71 (1961)

Schnell, F.W.: Die Covarianz zwischen Verwandten in einer genorthogonalen Population. I. Allgemeine Theorie. Biometr. Z. **7**, 1-49 (1965)

Sentz, J.C.; Robinson, H.F.; Comstock, R.E.: Relation between heterozygosis and performance in maize. Agronomy J. **46**, 514-520 (1954)

Stuber, C.W.; Cockerham, C. Clark: Gene effects and variances in hybrid populations. Genetics **54**, 1279-1286 (1966)

Stuber, C.W.; Moll, R.H.: Epistasis in maize (*Zea mays* L.). 1: F_1 hybrids and their S_1 progeny. Crop Sci. **9**, 124-127 (1969)

Wright, S.: The effects of inbreeding and crossbreeding on guinea pigs. USDA Bull. **1121** (1922)

Received August 22, 1979
Communicated by R.W. Allard

Dr. C.C. Cockerham
Department of Statistics
North Carolina State University
Raleigh, N. C. 27650 (USA)