ANALYSIS OF THE FULL DIALLEL CROSS

by

Ernest S. Hilderbrand

B. S., Kansas State University, 1965

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics and
Statistical Laboratory

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1966

Approved by:

Stanley Hearden

Major Professor

TABLE OF CONTENTS

INTRODUCTION

A full diallel cross is the set of p^2 possible single crosses and selfs between p inbred homozygous parental lines. The animal or plant breeder needs the full diallel cross in order to determine whether crossing per se is of value in improving productivity. Furthermore, he needs to determine the relative importance of certain types of specific combining ability and to indicate whether extensive crossing is needed to exploit non-additive genetic variation. Thus, the purpose of the diallel cross is to investigate the types and magnitudes of variability that contribute to differences among the p inbred parental lines.

The p^2 combinations may be displayed in a p x p table where x_{ii} represents the mean value of the i^{th} inbred line, x_{ij} represents the mean value of the F_1 generation resulting from crossing the i^{th} and j^{th} inbred lines, and x_{ji} represents the mean value of the F_1 generation of the reciprocal cross. Hence, the p^2 combinations may be divided into three groups: (1) the p parental lines, (2) the set of $1/2p(p-1)$ F_1's, and (3) the set of $1/2p(p-1)$ reciprocals.

There have been many different approaches to the analysis of the diallel cross depending upon the mathematical model used. This makes it very difficult for the experimenter to select one of the many methods available for his particular needs. Yates (1947) presented the orginal analysis. This was modified and stated in genetic parameters by Jinks and Hayman (1953) and improved further by Jinks (1954) and Hayman (1954 a, b). Kempthorne (1956) discussed the analysis in terms of the variances of the inbred parents and the crossbred offspring, and the covariances between

parents and offspring. Griffing (1956) discussed the analysis in terms of general combining ability (g. c. a.) and specific combining ability (s. c. a.) as defined by Sprague and Tatum (1942). Henderson (1948, 1952) discussed the analysis in terms of combining ability variances and maternal effects. Jinks (1954) and Jinks and Broadhurst (1963) discussed still another analysis concerned with maternal effects. Wearden (1964) summerized and compared most of the different preceding analyses in connection with two new more realistic models in order that the experimenter might use the diallel cross more effectively.

The purpose of this report is to discuss the analysis of the full diallel cross replicated in a randomized complete block design with special emphasis on the results of Wearden's (1964, 1965) work. Two methods of sampling may be applied to the diallel models. Four analyses will be considered for each of the sampling methods and model combinations.

DESIGN

A randomized complete block design with replication was chosen because of its common usage and its ease of application to the full diallel cross. Various other designs may be used quite successfully, but some form of replication is necessary for an estimate of random variation. It is assumed that there are p^2 matings each of which is assigned at random within each of the r replicates. The source of variation, degrees of freedom, and expectations of mean squares for such a design are:

SOURCE	d.f.	E (M. S.)
REPLICATES	r-1	$\sigma^2 + p^2\sigma_B^2$
MATINGS	p^2-1	$\sigma^2 + r\sigma_M^2$
RANDOM VARIATION	(r-1)(p^2-1)	σ^2

The purpose of the analysis of a full diallel cross is to partition the p^2-1 degrees of freedom, the corresponding sum of squares, and variances among matings into meaningful and useful genetical components or effects. To accomplish this, a linear model is necessary which will give reliable estimates of the magnitudes of these various genetical components or effects when used with the appropriate analysis.

MODELS

In diploid species the female and male parents both contribute equally to the nuclear genetic composition of the zygote, but the biological contribution of the female parent is usually greater than that of the male

parent. Maternal effects are possible through cytoplasmic inheritance (if one accepts this theory), or may exist because the female gamate is usually larger than the male gamate. Also the zygote usually receives nutrition for development either directly or indirectly from the female parent, and finally, it is usually the female parent who feeds and cares for the young after they are born. Wearden (1964) has presented the following maternal effects model (m. m. e.),

$$X_{ijK} = \mu + g_i + g_j + m_j + s_{ij} + b_K + \epsilon_{ijK} ,$$

which includes a single linear term, m_j, to account for these various maternal effects.

Often the progeny of the ij^{th} cross differ greatly from the progeny of the ji^{th} cross and maternal effects do not always account for these differences. Sex linkage or heterotic effects are probably the best explanation for these effects. Wearden (1964) has also presented the following reciprocal effects model (m. r. e.),

$$X_{ijK} = \mu + g_i + g_j + s_{ij} + r_{ij} + b_K + \epsilon_{ijK} ,$$

which takes into account these reciprocal effects.

Griffing (1956) discussed the model

$$X_{ijK} = \mu + g_i + g_j' + s_{ij}' + r_{ij} + b_K + \epsilon_{ijK}$$

for the full diallel cross which does not account for maternal effects and always accounts for reciprocal effects whether they exist or not. Hayman (1954 a, b) discussed the model

$$X_{ijK} = \mu + g_i + g'_j + s_{ij} + m'_i - m'_j + r_{ij} + b_K + \epsilon_{ijK}$$

which accounts for additive variation between the parents and for maternal effects. In order to account for the dominance effects Hayman partitioned the s_{ij} component into the components b_1, b_2, and b_3 where

b_1 = a general heterotic effect common to all biparental progeny.

b_2 = a unique heterotic effect shared by all biparental progeny from a given parent.

b_3 = a third form of dominance accounting for the remaining genetic differences or the fortuitous combination of genes.

Kempthorne (1956) has discussed Hayman's g_i, g_j, and s_{ij} further in statistical genetic terms.

The components in the above four mathematical models are defined as follows:

X_{ijK} = the observation of the cross between the i^{th} paternal line and the j^{th} maternal line in the K^{th} replicate.

μ = a mean value common to all matings to which inferences can be made from this set of p^2 crosses.

g_i = the common genic contribution of the i^{th} paternal line (mean deviation from the grand mean due to the i^{th} paternal line).

g_j = the common genic contribution of the j^{th} maternal line (mean deviation from the grand mean due to the j^{th} maternal line).

g_j' = the general combining ability effect of the j^{th} maternal line which is confounded with maternal effects if they exist.

m_j = the maternal contribution of the j^{th} maternal line.

m_j' = the difference between the effect of the j^{th} parental line used as a male parent and as a female parent.

s_{ij} = the interaction between the genetic contribution of the i^{th} line and that of the j^{th} line such that $s_{ij} = s_{ji}$.

s_{ij}' = the specific combining ability effect for the cross between the i^{th} paternal line and that of the j^{th} maternal line.

r_{ij} = the additional effect of using the i^{th} line as a male parent and the j^{th} line as a female parent such that $r_{ij} = 0$ and $r_{ij} = -r_{ji}$.

b_K = the effect of the K^{th} replicate. It is usually assumed that the b_K are normally, identically, and independently distributed random variables with zero mean and variance σ_B^2.

ϵ_{ijK} = the random effect peculiar to the ij^{th} cross in the K^{th} replicate. It is assumed that the ϵ_{ijK} are normally, identically, and independently distributed random variables with zero mean and variance σ^2.

The sampling distributions of g_i, g_j, g_j', m_j, m_j', s_{ij}, s_{ij}' and r_{ij} depend upon the sampling method used and are discussed under sampling methods.

Sprague and Tatum (1942) defined general combining ability (g. c. a.) as "the average performance of a line in hybrid combination," and specific combining ability (s. c. a.) as "those cases in which certain combinations do relatively better or worse than would be expected on the basis of the average performance of the lines involved." Wearden's (1964) definition of g_i is not that of general combining ability as defined by Griffing (1956). Furthermore, Wearden's definition of s_{ij} does not fit Griffing's s'_{ij}. Griffing defined these effects as deviations from the mean of the F_1's while Wearden defined these effects as deviations from the weighted means of the p inbred lines and p(p-1) crossbred lines. Although the differences between Wearden's g_i and general combining ability are probably trivial, his s_{ij} in addition to specific combining ability contains a component for the average deviation of the crosses from their respective midparents. This additional component is Hayman's "mean dominance deviation." He has developed a method for computing the sum of squares due to this effect and has indicated the appropriate variance ratio to test this "mean dominance deviation" effect for significance.

RANDOM SAMPLING EFFECTS

Sometimes it is impossible to include all parental lines in an experiment as a species is often easily subdivided into readily identifiable subgroups such as clones, varieties, families, herds, or flocks. A random sample is taken from the parent population and inferences are not made so much about the individual lines in the sample, but are made about the parental population parameters. Eisenhart's (1947) Model II describes

the situation where the p parental lines are a random sample of some inbred homozygous parent population.

Under the assumption of random sampling, μ is the mean of the parent population and the expectations of all components of the model are zero. While the expectations of these components squared are respectively σ_g^2, σ_m^2, σ_s^2, σ_B^2, and σ^2. The reciprocal effects are fixed in the sense that $r_{ii} = 0$ and $r_{ij} = -r_{ji}$, but it is assumed that the $1/2\ p(p-1)$ pairs of r_{ij} come from a very large population of such pairs of effects. Therefore the expectation of r_{ij} is zero and the expectation of the square of r_{ij} is σ_r^2. In order to make valid variance ratio tests of significance under the assumption of random sampling where μ is a constant and all other effects are random variables, it is necessary to assume that these random variables are normally, identically, and independently distributed with zero means and variance σ_θ^2 where θ = g, m, s, and r respectively.

FIXED SAMPLING EFFECTS

The experimenter often encounters the situation where the p parental lines are regarded as being the population about which inferences are to be made. Eisenhart's (1947) Model I describes the situation where the parental lines are a fixed sample.

When the experimenter uses a fixed set of parental lines, μ is the mean of all possible replications of the experiment. Fixed effects mean that the components of the model sum to zero for all i and j except for r_{ij} which is assumed to sum to zero only for each ij combination. Thus, the effects b_K and ϵ_{ijK} are random variables while g_i, g_j, m_j, s_{ij}, and

r_{ij} are constants. The expectations of the constant components in the model are the constants themselves. The expectations of the squares of the above constants are the constants squared. Statistically, these constants have no variance although an average of the squared effects may be computed. This average squared effect corresponds to the expected mean square in the analysis of variance for fixed effects experiments. It is computed by dividing the sum of the squared constant by the appropriate degrees of freedom for that term in the analysis of variance table. These averages of squared deviations will be called "variances of fixed effects" and will be denoted by $\tilde{\sigma}^2_\theta$. Which is the same σ^2_θ notation used with random sampling with a tilde superscript.

ANALYSIS OF THE FULL DIALLEL CROSS

The four analyses to be considered for the full diallel cross are (1) Hayman (1954 a, b) in terms of the Wearden (1964) models with Hayman's subdivision of the b sum of squares, (2) the p x p factorial used by Jinks and Broadhurst (1963), (3) the Wearden (1964) interpretation of the sum of squares given by Henderson (1952), and (4) Griffing's (1956) analysis for his model. Modified or partial diallel crosses will not be considered and the reader is referred to Griffing's (1956) discussion for this situation.

Each analysis will be developed for a single replicate of the full diallel cross to keep the mathematical symbolism to a minimum. In order to expand the analyses to r replicates, crosses are summed over all replicates and the indicated analysis is performed on these totals. Thus, all divisors for the sums of squares must be multiplied by r and all of the coefficients of all variance components except σ^2 in the expectations of mean squares must be multiplied by r.

SUMS OF SQUARES AND EXPECTATIONS FOR RANDOM SAMPLING METHODS

It is necessary to compute eleven different sums of squares in order to partition the p^2-1 degrees of freedom and the corresponding sums of squares for matings according to the four analyses under consideration. Table 1 presents the sums of squares and their expectations under both of Wearden's models for a single replicate.

In table 1, a dot indicates the summation from 1 to p over the values with the omitted subscript, and the capital sigmas indicate the summation

over all values of the i, j, and ij combinations. Letters with a prime indicate that the sums of squares express variation from the orgin while letters without a prime indicate that the sums of squares express variability from the mean. The symbol 1 is used for the correction factor since one degree of freedom is associated with this value.

Table 2 gives the degrees of freedom, sums of squares, and expectations of mean squares for both the maternal effects model and reciprocal effects model in conjunction with the Hayman, factorial, and Henderson analyses, and Griffing's model used with his analysis.

The Hayman analysis for the maternal effects model gives a valid variance ratio test (F - test) for the significance of maternal effects as verified by the ratio c expected mean square / d expected mean square. The ratio b expected mean square / d expected mean square indicates a valid variance ratio test for the significance of the variance due to genetic interaction. There is no valid variance ratio test for detecting a significant genic variance. The ratio b_1 expected mean square / d expected mean square indicates a valid variance ratio test for the significance of general heterosis. The valid variance ratio test for the significance of nested heterosis is verified by the ratio b_2 expected mean square / d expected mean square. The ratio b_3 expected mean square / d expected mean square demonstrates a valid test for the significance of the fortuitous combination of genes.

The reciprocal effects model for the Hayman analysis has the same expectations for the c and d mean squares which allows these two terms to be pooled. With replication, the significance of genetic interaction variance may be tested, as shown by the ratio b expected mean square /

Random variation expected mean square. The valid variance ratio test for the significance of reciprocal effects is shown by the ratio pooled c and d expected mean square / Random variation expected mean square. The significance of average heterosis among lines, nested heterosis, and the fortuitous combination of genes may be tested validly with the respective ratios, b_1 mean square, b_2 mean square, and b_3 mean square by Random variation mean square. An exact variance ratio test does not exist for the significance of genic variance, but the ratio a mean square / b mean square is a conservative test of genic variance. A conservative test is one where the actual probability of a type I error is less than the α - level given in a table of critical values. This test is approximate since the coefficient of σ_s^2 in the b expected mean square is approximately equal to the coefficient of σ_s^2 in the a expected mean square.

Valid variance ratio tests exist only for the maternal effects model when the factorial analysis is used. Significant maternal effects are detected by the ratio Maternal mean square / Paternal mean square. The ratio Paternal expected mean square / M x P expected mean square indicates a valid variance ratio test of significant genic variance. Replication is necessary to test the significance of genetic interaction variance. This is tested by the ratio M x P mean square / Random variation mean square.

The factorial analysis used in conjunction with the reciprocal effects model is of no use because there are no valid or even good approximate variance ratio tests. Also, it is not possible to estimate the variance

components as the expectations of the Maternal and Paternal mean squares are the same. This results in trying to solve two independent equations for three unknown variance components.

The analysis which Wearden has inferred from Henderson's equations for the estimation of variance components is particularly intended for the maternal effects model. The significance of maternal effects is tested validly by the ratio Dams mean square / Sires mean square. Genetic interaction can be tested for significance as inferred by the ratio Crosses expected mean square / Remainder expected mean square. There is no exact test of genic variation. The ratio Sires mean square / Crosses mean square is a conservative test of genic variation. This test may be too conservative though if σ_g^2 is very small. The Crosses term is the same as the b term in the Hayman analysis. Thus it may be partitioned into general heterosis, nested heterosis, and the fortuitous combination of genes. These effects are tested for significance as discussed under the Hayman analysis.

There is no valid variance ratio test for genic variance when the Henderson analysis is used in conjunction with the reciprocal effects model. Valid variance ratio tests exist for testing the significance of reciprocal effects and genetic interaction variance when there is replication. These effects are tested with the ratios Remainder mean square and Crosses mean square by Random variation mean square respectively. Again Crosses may be partitioned into Hayman's three b components. These effects are tested for significance as indicated in the Hayman analysis for the reciprocal effects model.

Replication is necessary for any variance ratio tests to be valid under the Griffing analysis. The significance of general combining

ability effects and specific combining ability effects are tested with the respective ratios (g. c. a.) mean square and (s. c. a.) mean square by Random variation mean square. The ratio Reciprocal effects mean square / Random variation mean square is a valid variance ratio test of reciprocal effects. Griffing's analysis of the full diallel cross does not estimate the variances of general combining ability and specific combining ability as defined by Sprague and Tatum (1942). They defined these terms about the mean of the F_1 generation or the mean of the crossbreds. Griffing's analysis results in these terms being computed from the weighted mean or the mean of the purebreds and crossbreds. If maternal effects exist, they will be confounded with reciprocal effects since Griffing's reciprocal effects term combines Hayman's c and d terms under the maternal effects model.

TABLE 1

Coefficients for variation from the orgin in a single replicate of a diallel cross under the assumption that the parental lines constitute a random sample

UMS OF SQUARES	SYMBOL	μ^2	σ^2_g	σ^2_s	σ^2	σ^2_m	σ^2_r
$\sum x_{ij}^2$	T'	p^2	$2p(p+1)$	p^2	p^2	p^2	$p(p-1)$
$\sum(X_{i.})^2/p$	p'	p^2	$p(p+3)$	p	p	p	$(p-1)$
$\sum(X_{ij})^2/p$	M'	p^2	$p(p+3)$	p	p	p^2	$(p-1)$
$_{..}{}^2/p^2$	1	p^2	$4p$	1	1	p	0
$(X_{i.}+X_{.i})^2/2p$	G'	$2p^2$	$2p(p+3)$	$2p$	$(p+1)$	$p(p+3)/2$	0
$\sum_{>j}(X_{ij}+X_{ji})^2/2+\sum X_{ii}^2$	C'	p^2	$2p(p+1)$	p^2	$p(p+1)/2$	$p(p+1)/2$	0
$(X_{i.}-X_{.i})^2/2p$	R	0	0	0	$(p-1)$	$p(p-1)/2$	$2(p-1)$
$\sum_{>j}(X_{ij}-X_{ji})^2/2$	D	0	0	0	$p(p-1)/2$	$p(p-1)/2$	0
$(_{..}-pX_{.})^2/p^2(p-1)$	B_1	0	0	1	1	0	0
$(_{i.}+X_{.i}-pX_{ii})^2/p(p-2)$	B_2	0	0	$(p-1)$	$(p-1)$	0	0
$\sum(2X_{..}-pX_{.})^2/p^2(p-2)$							
$\sum_{>j}(X_{ij}+X_{ji})^2/4+\sum X_{ii}^2$	B_3	0	0	$\frac{p(p^2-5p+5)}{p-2}$	$p(p-3)/2$	0	0
$\sum(X_{i.}+X_{.i}-2X_{ii})^2/2(p-2)$							
$X_{..}-X_{.})^2/(p-1)(p-2)$							

TABLE 2

Expectations of mean squares for a single replicate of a p^2 diallel cross under the assumption of random sampling of parental lines

SOURCE	df	Ss	Maternal effects model E(M.S.)	Reciprocal effects model E(M.S.)
Hayman Analysis				
a (Parental lines)	(p-1)	G'-2(1)	$\sigma^2+\frac{p}{2}\sigma_m^2+2\sigma_s^2+2p\sigma_g^2$	$\sigma^2+2\sigma_s^2+2p\sigma_g^2$
b (Genetic interaction)	1/2p(p-1)	C'-G'+1	$\sigma^2 + 2(\frac{p-1}{p})\sigma_s^2$	$\sigma^2 + 2(\frac{p-1}{p})\sigma_s^2$
b_1	1	B_1	$\sigma^2 + \sigma_s^2$	$\sigma^2 + \sigma_s^2$
b_2	(p-1)	B_2	$\sigma^2 + \sigma_s^2$	$\sigma^2 + \sigma_s^2$
b_3	p/2(p-3)	B_3	$\sigma^2 + \frac{2(p^2-5p+5)}{(p-3)(p-2)}\sigma_s^2$	$\sigma^2 + \frac{2(p^2-5p+5)}{(p-3)(p-2)}\sigma_s^2$
c (Average maternal effects)	(p-1)	R	$\sigma^2 + \frac{p}{2}\sigma_m^2$	$\sigma^2 + 2\sigma_r^2$
d (Reciprocal effects)	1/2(p-1)(p-2)	D-R	σ^2	$\sigma^2 + 2\sigma_r^2$

TABLE 2 CONTINUED

SOURCE	df	Ss	Maternal effects model E(M.S.)	Reciprocal effects model E(M.S.)
Factorial Analysis				
Maternal strains	(p-1)	M'-1	$\sigma^2 + \sigma_s^2 + p\sigma_g^2 + p\sigma_m^2$	$\sigma^2 + \sigma_r^2 + \sigma_s^2 + p\sigma_g^2$
Paternal strains	(p-1)	P'-1	$\sigma^2 + \sigma_s^2 + p\sigma_g^2$	$\sigma^2 + \sigma_r^2 + \sigma_s^2 + p\sigma_g^2$
M x P	(p-1)(p-1)	T'-P'-M'+1	$\sigma^2 + \sigma_s^2$	$\sigma^2 + (\frac{p-1}{p-1})\sigma_r^2 + \sigma_s^2$
Henderson Analysis				
Dams	(p-1)	M'-1	$\sigma^2 + \sigma_s^2 + p\sigma_g^2 + p\sigma_m^2$	$\sigma^2 + \sigma_r^2 + \sigma_s^2 + p\sigma_g^2$
Sires	(p-1)	P'-1	$\sigma^2 + \sigma_s^2 + p\sigma_g^2$	$\sigma^2 + \sigma_r^2 + \sigma_s^2 + p\sigma_g^2$
Crosses	1/2p(p-1)	C'-G'+1	$\sigma^2 + 2(\frac{p-1}{p})\sigma_s^2$	$\sigma^2 + \frac{2(p-1)}{p}\sigma_s^2$
Remainder	1/2(p-1)(p-2)	D-R	σ^2	$\sigma^2 + 2\sigma_r^2$

TABLE 2 CONTINUED

SOURCE	df	Ss	Griffing model E(M.S.)
Griffing Analysis			
General combining ability (g.c.a.)	p-1	G' - 2(1)	$\sigma^2 + \frac{2(p-1)}{p}\sigma_s^2 + 2p\sigma_g^2$
Specific combining ability (s.c.a.)	1/2p(p-1)	C' - G' +1	$\sigma^2 + \frac{2(p^2-p+1}{p^2}\sigma_s^2$
Reciprocal effects	1/2p(p-1)	D	$\sigma^2 + 2\sigma_r^2$

SUMS OF SQUARES AND EXPECTATIONS FOR FIXED SAMPLING METHODS

Table 3 gives the sums of squares and their expectations for the situation when the p parental lines are a fixed sample and inferences are just to be made about these fixed parental lines. The symbolism and notation used in this table is the same as that of table 1.

The coefficients in table 3 differ in many cases from those in table 1. This is because fixed effects when summed over all values add to zero. This also introduces a negative correlation, $-1/(p-1)$, within any set of p constants.

Table 4 gives the degrees of freedom, method of computing sums of squares, and expectations of mean squares for the Hayman, factorial, and Henderson analyses used with Wearden's models. It also gives this information for the Griffing analysis using his model.

The Hayman analysis was especially designed for experiments involving a fixed set of parental lines. Under the maternal effects model genic variation is tested for significance with the ratio a mean square / d mean square. The valid variance ratio test (F - test) for the significance of genetic interaction variance is indicated by the ratio b expected mean square / d expected mean square. Significance of maternal effects is tested by the ratio c mean square / d mean square. General heterosis may be tested for significance with the ratio b_1 mean square / d mean square. The ratio b_2 expected mean square / d expected mean square indicates a valid variance ratio test for the significance of nested heterosis. The ratio b_3 mean square / d mean square is used to test the significance of the fortuitous combination of genes.

All variance components may be tested by the Random variation expected mean square to yield exact variance ratio tests when the Hayman analysis for the reciprocal effects model is replicated. General heterosis, nested heterosis, and the fortuitous combination of genes may also be tested validly by the Random variation expected mean square.

The factorial analysis for the maternal effects model gives exact variance ratio tests. Maternal effects can be tested for significance as seen by the ratio Maternal effects expected mean square / Paternal effects expected mean square. With replication significance of genic variance is detected by the ratio Paternal effects mean square / Random variation mean square. The ratio M x P expected mean square / Random variation expected mean square indicates the test for significant genetic interaction variance.

As in the case of random sampling, the factorial analysis for the reciprocal effects model is of little value. No valid variance ratio tests exist and it is not possible to estimate the variance components for the same reasons as discussed under random sampling.

The Henderson analysis for the maternal effects model results in exact variance ratio tests when the parental lines are fixed. Maternal effects are tested for significance by the ratio Dams mean square / Sires mean square. Significant genic variation and genetic interaction variance are detected respectively by the ratios Sires mean square and Crosses mean square by Remainder mean square. Crosses may be partitioned into average heterosis among lines, nested heterosis, and the fortuitous combination of genes. These effects are tested for significance as indicated in the Hayman analysis for the fixed effects model.

The Henderson analysis for the reciprocal effects model does not give an exact test for genic variation. But a very conservative test for genic variation is made by the ratio Sires mean square / Remainder mean square. Reciprocal effects and genetic interaction variance may be tested for significance by the ratios Remainder mean square and Crosses mean square by Random variation mean square respectively with replication. Again Crosses may be partitioned into Hayman's three b components. These components are tested for significance as discussed in the Hayman analysis for the fixed reciprocal effects model.

Griffing's analysis for his model yields exact variance ratio tests when there is replication. General combining ability effects and specific combining ability effects can be tested for significance as seen by the Random variation expected mean square. Reciprocal effects are also tested by the Random variation mean square. Again Griffing's analysis for the full diallel cross does not estimate the variance of general combining ability and specific combining ability as defined by Sprague and Tatum (1942). The reason for this is the same as discussed under Griffing's analysis for random sampling methods.

TABLE 3

Coefficients of variation from the orgin in a single replicate of a full diallel ross under the assumption that the parental lines are a fixed sample

UMS OF SQUARES	SYMBOL	μ^2	σ_g^2	σ_s^2	σ^2	σ_m^2	σ_r^2
X_{ij}^2	T'	p^2	$2p^2$	p^2	p^2	p^2	$p(p-1)$
$(X_{i.})^2/p$	P'	p^2	p^2	0	0	0	$p-1$
$(X_{ij})^2/p$	M'	p^2	p^2	0	p	p^2	$p-1$
$^2_{..}/p^2$	1	p^2	0	0	1	0	0
$(X_{i.}+X_{.i})^2/2p$	G'	$2p^2$	$2p^2$	0	$p+1$	$p^2/2$	0
$(X_{ij}+X_{ji})^2/2+\Sigma X_{ii}^2$	C'	p^2	$2p^2$	p^2	$p(p+1)/2$	$p^2/2$	0
$(X_{i.}-X_{.i})^2/2p$	R	0	0	0	$p-1$	$p^2/2$	$2(p-1)$
$(X_{ij}-X_{ji})^2/2$	D	0	0	0	$p(p-1)/2$	$p^2/2$	$p(p-1)$
$(_{i.}-pX_{.})^2/p^2(p-1)$	B_1	0	0	$p/(p-1)$	1	0	0
$(X_{i.}+X_{.i}-pX_{ii})^2/p(p-2)$	B_2	0	0	$\frac{p(p-1)}{p-2}$	$(p-1)$	0	0
$2X_{..}-pX_{.})^2/p^2(p-2)$							
$X_{ij}+X_{ji})^2/4+\Sigma X_{ii}^2$	B_3	0	0	$\frac{p(p^3-4p^2+3p+1)}{(p-1)(p-2)}$	$(p-3)/2$	0	0
$(X_{i.}+X_{.i}-2X_{ii})^2/2(p-2)$							
$X_{..}-X_{.})^2/(p-1)(p-2)$							

TABLE 4

Expectations of mean squares for a single replicate of a p^2 diallel cross under the assumption that the p parental lines are a fixed sample

SOURCE	df	Ss	Maternal effects model E(M.S.)	Reciprocal effects model E(M.S.)
Hayman Analysis				
a (Parental lines)	(p-1)	G'-2(1)	$\sigma^2+1/2p\tilde{\sigma}^2_m+2p\tilde{\sigma}^2_g$	$\sigma^2+2p\tilde{\sigma}^2_g$
b (Genetic interaction)	1/2p(p-1)	C'-G'+1	$\sigma^2 + \frac{2p}{p-1}\tilde{\sigma}^2_s$	$\sigma^2 + \frac{2p}{p-1}\tilde{\sigma}^2_s$
b_1	1	B_1	$\sigma^2 + \tilde{\sigma}^2_s$	$\sigma^2 + \tilde{\sigma}^2_s$
b_2	(p-1)	B_2	$\sigma^2 + \frac{p}{p-2}\tilde{\sigma}^2_s$	$\sigma^2 + \frac{p}{p-2}\tilde{\sigma}^2_s$
b_3	1/2p(p-3)	B_3*	$\sigma^2 + \frac{2(p^3-4p^2+3p+1)}{(p-1)(p-2)(p-3)}\tilde{\sigma}^2_s$	$\sigma^2 + \frac{2(p^3-4p^2+3p+1)}{(p-1)(p-2)(p-3)}\tilde{\sigma}^2_s$
c (Average maternal effects)	(p-1)	R	$\sigma^2 + 1/2p\tilde{\sigma}^2_m$	$\sigma^2 + 2\tilde{\sigma}^2_r$
d (Reciprocal effects)	1/2(p-1)(p-2)	D-R	σ^2	$\sigma^2 + 2\tilde{\sigma}^2_r$

* Also $B_3 = C - B_1 - B_2$

TABLE 4 CONTINUED

SOURCE	df	Ss	Maternal effects model E(M.S.)	Reciprocal effects model E(M.S.)
Factorial Analysis				
Maternal strains	(p-1)	M'-1	$\sigma^2 + p\tilde{\sigma}^2_g + p\tilde{\sigma}^2_m$	$\sigma^2 + \tilde{\sigma}^2_r + p\tilde{\sigma}^2_g$
Paternal strains	(p-1)	P'-1	$\sigma^2 + p\tilde{\sigma}^2_g$	$\sigma^2 + \tilde{\sigma}^2_r + p\tilde{\sigma}^2_g$
M x P	(p-1)(p-1)	T'-M'-P'+1	$\sigma^2 + \tilde{\sigma}^2_s$	$\sigma^2 + \frac{p-2}{p-1}\tilde{\sigma}^2_r + \tilde{\sigma}^2_s$
Henderson Analysis				
Dams	(p-1)	M'-1	$\sigma^2 + p\tilde{\sigma}^2_g + p\tilde{\sigma}^2_m$	$\sigma^2 + \tilde{\sigma}^2_r + p\tilde{\sigma}^2_g$
Sires	(p-1)	P'-1	$\sigma^2 + p\tilde{\sigma}^2_g$	$\sigma^2 + \tilde{\sigma}^2_r + p\tilde{\sigma}^2_g$
Crosses	1/2p(p-1)	C'-G'+1	$\sigma^2 + \frac{2p}{p-1}\tilde{\sigma}^2_s$	$\sigma^2 + \frac{2p}{p-1}\tilde{\sigma}^2_s$
Remainder	1/2(p-1)(p-2)	D-R	σ^2	$\sigma^2 + 2\tilde{\sigma}^2_r$

TABLE 4 CONTINUED

SOURCE	df	Ss	Griffing model E(M.S.)
Griffing Analysis			
General combining ability (g.c.a.)	(p-1)	G' - 2(1)	$\sigma^2 + 2p\tilde{\sigma}^2_g$
Specific combining ability (s.c.a.)	1/2p(p-1)	C' - G' + 1	$\sigma^2 + 2p/(p-1)\ \tilde{\sigma}^2_s$
Reciprocal effects	1/2p(p-1)	D	$\sigma^2 + 2\tilde{\sigma}^2_r$

REPLICATION AND POWER

Replication is necessary in the analysis of the full diallel cross regardless of the model used. It is usually needed to provide an estimate of the random variation σ^2. If a comparison within the p^2 matings results in an independent estimate of σ^2, then this σ^2 may be pooled with the σ^2 resulting from replication giving greater denominator degrees of freedom, and thus power, where this source of variation is the testing term for tests of significance.

The relative power of the different variance ratio tests many times determines the analysis used if the experimenter is interested in analyzing the significance of one particular variance. The following discussion concerning the relative power assumes that there is replication and that all estimates of random variation σ^2 are pooled. The b (Genetic interaction) mean square of Hayman's analysis and Crosses mean square of the Henderson analysis give the more powerful test of σ^2_s for both of Wearden's models under both sampling methods. For a fixed set of parental lines the Hayman analysis gives the most powerful test of σ^2_g for the reciprocal effects model. The factorial analysis or Henderson analysis are probably more powerful than the Hayman analysis in testing σ^2_g for the maternal effects model. The Hayman analysis is probably less powerful since the presence of the coefficient for σ^2_m in the a mean square for the maternal effects model requires the use of the c mean square as a test term resulting in a loss of denominator degrees of freedom for the variance ratio test. The larger the relative size of σ^2_m and the greater the loss in degrees of freedom, the more adversely affected will be the power in testing σ^2_g

under the Hayman analysis for the fixed maternal effects model. The pooling of the c and d mean squares and testing them with the Random variation mean square gives the most powerful test of σ_r^2 under the reciprocal effects model for both sampling methods. Griffing's Reciprocal effects term is the same as the Hayman pooled c and d terms under the reciprocal effects model. Therefore, testing Reciprocal effects by Random variation is also the most powerful test of σ_r^2 .

LEAST SQUARES ESTIMATES FOR WEARDEN'S MODELS

The least squares estimates for the components of the maternal effects model,

$$X_{ij} = \mu + g_i + g_j + m_j + s_{ij} + e_{ij} ,$$

are obtained under the assumption that g_i, g_j, m_j, and s_{ij} are fixed variables. Then the expectations of the partial derivatives taken with respect to the various components of the error sum of squares,

$$\Sigma(X_{ij} - \mu - g_i - g_j - m_j - s_{ij})^2 ,$$

are set identically equal to zero. This results in the following set of normal equations.

$$\partial \Sigma(X_{ij} - \mu - g_i - g_j - m_j - s_{ij})^2 / \partial \mu$$

$$= 2 \Sigma(X_{ij} - \mu - g_i - g_j - m_j - s_{ij})$$

$$= \Sigma X_{ij} - p^2\hat{\mu} - p\Sigma\hat{g}_j - p\Sigma\hat{m}_j - \Sigma\hat{s}_{ij} \equiv 0.$$

$$\partial \Sigma(X_{ij} - \mu - g_i - g_j - m_j - s_{ij})^2 / \partial g_i$$

$$= 2 \Sigma(X_{ij} - \mu - g_i - g_j - m_j - s_{ij})$$

$$= \Sigma X_{ij} - p\hat{\mu} - p\hat{g}_i - \Sigma\hat{g}_j - \Sigma\hat{m}_j - \Sigma\hat{s}_{ij} \equiv 0.$$

$$\partial \Sigma(X_{ij} - \mu - g_i - g_j - m_j - s_{ij})^2 / \partial m_j$$

$$= 2 \Sigma(X_{ij} - \mu - g_i - g_j - m_j - s_{ij})$$

$$= \Sigma X_{ij} - p\hat{\mu} - \Sigma \hat{g}_i - p\hat{g}_j - p\hat{m}_j - \Sigma \hat{s}_{ij} \equiv 0.$$

The condition that $s_{ij} = s_{ji}$ must be utilized in the following manner to obtain a normal equation for s_{ij}.

$$\partial (\Sigma\varepsilon_{ij}^2 + \Sigma\varepsilon_{ji}^2) / \partial s_{ij}$$

$$= \partial(\Sigma(X_{ij}-\mu-g_i-g_j-m_j-s_{ij})^2 + \Sigma(X_{ji}-\mu-g_j-g_i-m_i-s_{ij})^2)/\partial s_{ij}$$

$$= 2((X_{ij}-\mu-g_i-g_j-m_j-s_{ij}) + (X_{ji}-\mu-g_j-g_i-m_i-s_{ji}))$$

$$= X_{ij} + X_{ji} - 2\hat{\mu} - 2\hat{g}_i - 2\hat{g}_j - \hat{m}_j - \hat{m}_i - 2\hat{s}_{ij} \equiv 0.$$

The assumption that g_i, g_j, m_j, and s_{ij} are fixed variables implies the restrictions that

$$\Sigma\hat{g}_i = \Sigma\hat{g}_j = \Sigma\hat{m}_j = \Sigma\hat{s}_{ij} = 0.$$

The least squares estimates of the various effects are then obtained by applying the above restrictions to the normal equations when solving them. Therefore, the least squares estimates are:

$$\hat{\mu} = X_{..}/p^2 = \bar{X}_{..} ,$$

$$\hat{g}_i = (X_{i.} - p\hat{\mu})/p = \bar{X}_{i.} - \bar{X}_{..} ,$$

$$\hat{g}_j = \hat{g}_i = \bar{X}_{j.} - \bar{X}_{..}$$

because it is assumed that the genic contributions of the maternal and

paternal lines are the same,

$$\hat{m}_j = (X_{ij} - p\hat{\mu} - p\hat{g}_j) / p = \bar{X}_{.j} - \bar{X}_{..} - \bar{X}_{j.} + \bar{X}_{..}$$

$$= \bar{X}_{.j} - \bar{X}_{j.} , \text{ and}$$

$$\hat{s}_{ij} = (X_{ij} + X_{ji} - 2\hat{\mu} - 2\hat{g}_i - 2\hat{g}_j - \hat{m}_j - \hat{m}_i) / 2$$

$$= (X_{ij} + X_{ji} - X_{i.} - \bar{X}_{.i} - \bar{X}_{j.} - \bar{X}_{.j}) / 2 + \bar{X}_{..} .$$

The least square estimates for the reciprocal effects model,

$$X_{ij} = \mu + g_i + g_j + s_{ij} + r_{ij} + \epsilon_{ij} ,$$

are obtained in a similar way. The normal equations for the reciprocal effects model are:

$$\partial \Sigma(X_{ij} - \mu - g_i - g_j - s_{ij} - r_{ij})^2 / \partial \mu$$

$$= 2 \Sigma(X_{ij} - \mu - g_i - g_j - s_{ij} - r_{ij})$$

$$= \Sigma X_{ij} - p^2\hat{\mu} - p\Sigma\hat{g}_i - p\Sigma\hat{g}_j - \Sigma\hat{s}_{ij} - \Sigma\hat{r}_{ij} \equiv 0.$$

$$\partial \Sigma(X_{ij} - \mu - g_i - g_j - s_{ij} - r_{ij})^2 / \partial g_i$$

$$= 2 \Sigma(X_{ij} - \mu - g_i - g_j - s_{ij} - r_{ij})$$

$$= \Sigma X_{ij} - p\hat{\mu} - p\hat{g}_i - \Sigma \hat{g}_j - \Sigma \hat{s}_{ij} - \Sigma \hat{r}_{ij} \equiv 0.$$

$$\partial \Sigma(X_{ij} - \mu - g_i - g_j - s_{ij} - r_{ij})^2 \;/\; \partial g_j$$

$$= 2 \Sigma(X_{ij} - \mu - g_i - g_j - s_{ij} - r_{ij})$$

$$= \Sigma X_{ij} - p\hat{\mu} - \Sigma \hat{g}_i - p\hat{g}_j - \Sigma \hat{s}_{ij} - \Sigma \hat{r}_{ij} \equiv 0.$$

Again, the condition that $s_{ij} = s_{ji}$ must be utilized to obtain a normal equation for s_{ij}.

$$\partial (\Sigma \epsilon^2_{ij} + \Sigma \epsilon^2_{ji}) \;/\; \partial s_{ij}$$

$$= \partial (\Sigma(X_{ij}-\mu-g_i-g_j-s_{ij}-r_{ij})^2 + \Sigma(X_{ji}-\mu-g_j-g_i-s_{ji}-r_{ji})^2)/\partial s_{ij}$$

$$= 2 \left((X_{ij}-\mu-g_i-g_j-s_{ij}-r_{ij}) + (X_{ji}-\mu-g_j-g_i-s_{ji}-r_{ji}) \right)$$

$$= X_{ij} + X_{ji} - 2\hat{\mu} - 2\hat{g}_i - 2\hat{g}_j - 2\hat{s}_{ij} \equiv 0.$$

Also, the condition that $r_{ij} = -r_{ji}$ is necessary to obtain a normal equation for r_{ij}.

$$\partial (\Sigma \epsilon^2_{ij} - \Sigma \epsilon^2_{ji}) \;/\; \partial r_{ij}$$

$$= \partial(\Sigma(X_{ij}-\mu-g_i-g_j-s_{ij}-r_{ij})^2 - \Sigma(X_{ji}-\mu-g_j-g_i-s_{ji}-r_{ji})^2)/\partial r_{ij}$$

$$= 2((X_{ij}-\mu-g_i-g_j-s_{ij}-r_{ij}) - (X_{ji}-\mu-g_j-g_i-s_{ji}-r_{ji}))$$

$$= X_{ij} - X_{ji} - 2\hat{r}_{ij} \equiv 0.$$

The restriction,

$$\Sigma \hat{g}_i = \Sigma \hat{g}_j = \Sigma \hat{s}_{ij} = \Sigma \hat{r}_{ij} = 0,$$

is applied to the above normal equations when solving them for the least squares estimates of the various effects. Thus, the least squares estimates for the components of the reciprocal effects model are:

$$\hat{\mu} = X_{..}/p = \bar{X}_{..} \quad ,$$

$$\hat{g}_i = (X_{i.} - p\hat{\mu}) = \bar{X}_{i.} - \bar{X}_{..} \quad ,$$

$$\hat{g}_j = (X_{.j} - p\hat{\mu}) = \bar{X}_{.j} - \bar{X}_{..} \quad ,$$

$$\hat{s}_{ij} = (X_{ij} + X_{ji} - 2\hat{\mu} - 2\hat{g}_i - 2\hat{g}_j) / 2$$

$$= (X_{ij} + X_{ji})/2 - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..} \quad , \text{ and}$$

$$\hat{r}_{ij} = (X_{ij} - X_{ji})/2 \quad .$$

The least square estimates for components of Wearden's (1964) maternal effects model, Wearden's reciprocal effects model, and Griffing's (1956) reciprocal effects model are summerized in table 5. The least squares estimates for the components of Griffing's model are derived in a similar manner.

TABLE 5

Least squares estimates of the effects of Wearden's maternal effects model, reciprocal effects model, and Griffing's reciprocal effects model

EFFECT ESTIMATE	MATERNAL EFFECTS MODEL	RECIPROCAL EFFECTS MODEL	GRIFFING'S MODEL
μ	$\bar{X}_{..}$	$\bar{X}_{..}$	$\bar{X}_{..}$
g_i	$\bar{X}_{i.} - \bar{X}_{..}$	$\bar{X}_{i.} - \bar{X}_{..}$	$(\bar{X}_{i.}+\bar{X}_{.i})/2-\bar{X}_{..}$
g_j	$\bar{X}_{j.} - \bar{X}_{..}$	$\bar{X}_{.j} - \bar{X}_{..}$	———
m_j	$\bar{X}_{.j} - \bar{X}_{j.}$	———	———
s_{ij}	$(X_{ij}+X_{ji}-\bar{X}_{i.}-\bar{X}_{.j}+\bar{X}_{..})/2+\bar{X}_{..}$	$(X_{ij}+X_{ji})/2 -\bar{X}_{i.}-\bar{X}_{.j}+\bar{X}_{..}$	$(X_{ij}+X_{ji}-X_{i.}-X_{.i}$ $-\bar{X}_{j.}-\bar{X}_{.j})/2+\bar{X}_{..}$
r_{ij}	———	$(X_{ij}-X_{ji})/2$	$(X_{ij}-X_{ji})/2$

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to his major professor, Dr. Stanley Wearden, for his advice and assistance during the preparation of this report.

The help of Dr. Leslie F. Marcus and Dr. John D. Wheat in reviewing the manuscript is gratefully acknowledged.

BIBLIOGRAPHY

Eisenhart, C. (1947). "The Assumptions Underlying the Analysis of Variance." Biometrics, 3, 1-21.

Griffing, B. (1956). "Concept of General and Specific Combining Ability in Relation to Diallel Crossing Systems." Australian Journal of Biological Sciences, 9, 463-493.

Hayman, B. I. (1954a). "The Analysis of Variance of Diallel Crosses." Biometrics, 10, 235-244.

Hayman B. I. (1954b). "The Theory and Analysis of Diallel Crosses." Genetics, 39, 789-809.

Henderson, C. R. (1948). "Estimation of General, Specific, and Maternal Combining Abilities in Crosses Among Inbred Lines of Swine." Unpublished Ph.D. Thesis, Iowa State College.

Henderson, C. R. (1952). "Specific and General Combining Ability." Heterosis. Iowa State College Press, Ames.

Jinks, J. L., and Hayman, B. I. (1953). "Analysis of Diallel Crosses." Maize Genetics Cooperation News Letter, 27, 48-54.

Jinks, J. L. (1954). "The Analysis of Continuous Variation in a Diallel Cross of Nicotiana Rustica Varieties." Genetics, 39, 767-788.

Jinks, J. L., and Broadhurst, P. L. (1963). "Diallel Analysis of Litter Size and Body Weight in Rats." Heredity, 18, 319-336.

Kempthorne, O. (1956). "The Theory of the Diallel Cross." Genetics, 41, 451-459.

Sprague, G. F., and Tatum, L. A. (1942). "General vs. Specific Combining Ability in Single Crosses of Corn." J. Am. Soc. Agron., 34, 923-932.

Wearden, S. (1964). "AlternativeAnalyses of the Diallel Cross." Heredity, 9, 669-680.

Wearden, S., Tindell, D., and Craign, J. V. (1965). "Use of a Full Diallel Cross to Estimate General and Specific Combining Ability in Chickens." Poultry Science, 44, 1043-1053.

Yates, F. (1947). "The Analysis of Data from All Possible Reciprocal Crosses Between a Set of Parental Lines." Heredity, 1, 287-302.

ANALYSIS OF THE FULL DIALLEL CROSS

by

Ernest S. Hilderbrand

B.S., Kansas State University, 1965

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirement for the degree

MASTER OF SCIENCE

Department of Statistics and
Statistical Laboratory

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1966

ABSTRACT

A full diallel cross is the set of p^2 possible single crosses and selfs between p inbred homozygous parental lines. The animal or plant breeder needs the full diallel cross in order to determine whether crossing per se is of value in improving productivity. Furthermore, he needs to determine the relative importance of certain types of specific combining ability and to indicate whether extensive crossing is needed to exploit non-additive genetic variation. Thus, the purpose of the analysis of the full diallel cross is to investigate the types and magnitudes of variability that contribute to differences among the p inbred homozygous parental lines.

The experimental design discussed is the randomized complete block design with replication. Four statistical models are applied to this design: Wearden's maternal effects model, Wearden's reciprocal effects model, Hayman's additive effects model, and Griffing's reciprocal effects model. The purpose, advantages, and disadvantages of each model is discussed.

Two methods of sampling, fixed and random, are considered. The sampling method has a profound effect on the analyses and interpretation of the above models.

Four analyses of the full diallel cross are presented for both of Wearden's models and Griffing's model under both sampling methods; namely, Hayman's, the factorial, Henderson's, and Griffing's. The purpose, advantages, and disadvantages of each analysis is discussed for each model and sampling method.

The least squares estimates for the components of Wearden's maternal effects model and reciprocal effects model are presented. Power and replication of the four analyses are discussed in order that the experimenter may select the best combination of model, analysis, and sampling method to accomplish his purpose.