STOCHASTIC SIMULATION OF A SELECTION EXPERIMENT IN MAIZE

by

Sophronia W. Ward

Dissertation submitted to the Graduate Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

Klaus Hinkelmann, Chairman

Jesse C. Arnold

Raymond H. Myers

Clarence Genter

John M. White

March, 1978

Blacksburg, Virginia

ACKNOWLEDGMENTS

The research reported in this dissertation was conducted at Virginia Polytechnic Institute and State University, Blacksburg, Virginia, and was supported by Dr. Clarence Genter formerly of the Department of Agronomy, VPI & SU.

The author wishes to express her sincere appreciation to Dr. Klaus Hinkelmann of the Department of Statistics for his patience and guidance during her years of graduate study. Appreciation is extended to Dr. Genter for his support of this project, both educational and financial.

Thanks are expressed to other members of the committee, in particular to Dr. Raymond Myers of the Department of Statistics and Dr. John White of the Department of Dairy Science for their helpful suggestions. Thanks are also expressed to of the Statistical Consulting Center for her assistance in the computing area and to for her careful typing of this dissertation.

Special thanks are extended to the author's parents, and and to her children, and for their support and understanding during the years spent in graduate school.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

TABLE OF CONTENTS (cont.)

LIST OF TABLES

LIST OF TABLES (cont.)

LIST OF TABLES (cont.)

LIST OF TABLES (cont.)

LIST OF FIGURES

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

LIST OF FIGURES (cont.)

Chapter I

INTRODUCTION

Population improvement and the genetics of corn yield are of major concern to researchers involved in the development of higher yielding inbred lines of corn for use in hybrids. In order to produce higher yielding inbred lines and hybrids, an understanding of genetic mechanisms controlling yield is important. The behavior of this complexly inherited trait is the subject of much discussion, since the effectiveness of any selection scheme used to increase yield depends upon what type of gene action is involved. If partial to complete dominance is the prevalent gene action controlling yield, then some type of recurrent family selection is desirable.

The major distinguishing feature of recurrent selection is the intercrossing of selected individuals each cycle to allow for recombination of alleles present in the selections. Although the basic outline of recurrent selection, as used in many plant breeding programs, was first described by Jenkins (1940), the term was first used by Hull (1945) to describe selection schemes which follow a self-select-intercross pattern in each cycle of selection. The basic outline of a recurrent selection program can be described as follows:

1. Self-pollinate the plants in the population of interest.
2. Evaluate and select the superior plants.
3. Grow out selfed seed from the superior plants and make all possible crosses.
4. Repeat Steps 1-3 with the plants in the new population.

The basis of evaluation and selection varies with the type of recurrent selection. Descriptions of the four basic types can be found in articles by Jenkins (1940), Hull (1945), and Robinson, et al. (1955).

Advantages of recurrent selection over other types of selection are a slower rate of inbreeding combined with the possible recombination of favorable alleles in the intercross generation. In addition to allowing for recombination, intercrossing insures a slow approach to homozygosity of the population which is more favorable to the fixation of desirable alleles. Under recurrent selection population improvement is possible, "while maintaining a high degree of genetic variability in the population" (Genter, 1967). In general, recurrent selection "provides for a much more accurate genetic control..." (Sprague, 1952).

Theoretical consideration of the joint effect of several factors or genetic parameters, such as population size, selection intensity, linkage, or environmental variance, on the behavior of any quantitatively inherited trait such as yield of corn is extremely difficult except in the simpliest situations. As Crosby (1973) states, the "complexity of the mathematical techniques required increases exponentially rather than linearly with biological complexity." In some problems, simplifying assumptions can be made which lead to manageable mathematical expressions. Where appropriate, approximations can be useful in understanding biological relationships. However, one must realize "that approximations, even apparently reasonable ones, may lead to biological nonsense..." (Crosby, 1973).

Biological problems which lead to extremely complex mathematical expressions or for which approximations are not appropriate require

other techniques of investigation. One such technique is that of simulation. A model of the problem is developed and the results of many repetitions of an experiment using the model are averaged and analyzed. Particular attention is given to the values assigned to the parameters in the model in order that the results of the simulation study are biologically meaningful. Simulation of a realistic situation can be more enlightening than the theoretical investigation of a very simple unrealistic situation.

Simulation studies of mathematically intractable biological problems became feasible with the advent of high speed computers. Using Monte Carlo techniques, the effects of selection intensity, population size, heritability, and linkage on fixation of alleles and response to selection have been investigated by Fraser (1957a, 1957b), Martin and Cockerham (1960), and Qureshi and Kempthorne (1968). The effects of different gene action models have been investigated by Fraser (1960a, 1960b, 1960c), Gill (1965a, 1965c), Young (1966), Qureshi, et. al. (1968), and Qureshi (1968). Comparisons of Monte Carlo results with theoretical predictions have been made by Gill (1965b) and Young (1966). All of these studies aid in the understanding of the effects of genetic parameters in selection experiments; but, it must be kept in mind that "the empirical procedure is properly a tool" to the development of a formulated theory of the behavior of genetic parameters in selection (Martin and Cockerham, 1960).

A reasonable extension of the previously reported simulation studies is to model and simulate an experiment that is in progress and to compare simulated and experimental results. This thesis will focus

on a particular recurrent selection experiment described by Genter (1976). Specifically, we shall be concerned with the development of a genetic model to describe the experiment and the results of simulation runs of the model on a computer.

Two goals of the original experiment are (1) an increase in the frequency of desirable alleles contributing to yield and (2) the development of a high yielding inbred line of maize. Simulation of this experiment will help answer questions concerning the long range behavior of the loci contributing to yield. Some of these questions are: (1) Is it more effective to select a smaller proportion of the population each cycle? (2) Under what conditions are desirable alleles lost in the population? (3) How long will population improvement continue? (4) What conditions affect the rate of population improvement? (5) Can a maximum yield be attained for the population? (6) Can a completely homozygous high yielding population emerge? and (7) How fast does the coefficient of inbreeding increase?

A description of the experiment to be modeled, definitions and implications of genetic terms, and the development of the model are contained in Chapter II. Genetic parameters relevant to the selection experiment are identified and values for these parameters are chosen. Evidence from the literature supporting the particular values chosen for several of these parameters is presented. Other parameter values were determined by the nature of the selection experiment itself. A description of the simulation program and discussion of the techniques employed is included in Chapter III. The design used to investigate the parameters used in the simulation program is discussed in Chapter

IV. In addition, the procedures used to analyze the simulation results are presented.

In Chapter V the results of the simulation program are presented and analyses of those results are discussed. Comparisons are made with the experimental results for the first four cycles of selection. Analyses of variance of yield and percent homozygosity for different cycles of selection are included and the significant genetic parameters and interactions are discussed in the context of the original experiment. Behavior of yield and percent homozygosity through 30 cycles of selection is described using autoregression analysis. Situations in which the less desirable alleles became fixed in the population are identified and discussed.

A theoretical discussion of the inbreeding coefficient and its relationship to percent homozygosity is presented in Chapter VI. The limitations of existing theory, exact and approximate, are discussed. Upper and lower bounds for the minimum and maximum inbreeding in a recurrent selection program are included for the one locus case. Genetic variance components for the gene action models used are presented in Chapter VII. Partitioning of the total genetic variance for simple gene action models with linkage is included. A summary and conclusions are presented in Chapter VIII as well as suggestions for further research based on the study of recurrent selection.

Chapter II

MODELING THE SELECTION EXPERIMENT

2.1 Introduction

Any type of selection experiment has as its goal the improvement of a population for a specific trait or combination of traits. Through analyses of the results of selection, some knowledge of the genetic nature of the trait under consideration is gained. In maize, yield is controlled by many loci distributed throughout the 10 pairs of chromosomes. The behavior of yield over time and which genetic factors most influence this behavior will be investigated. Since a theoretical investigation is too complex, a mathematical model will be developed for a particular recurrent selection experiment (Genter, 1976), and investigated by Monte Carlo techniques.

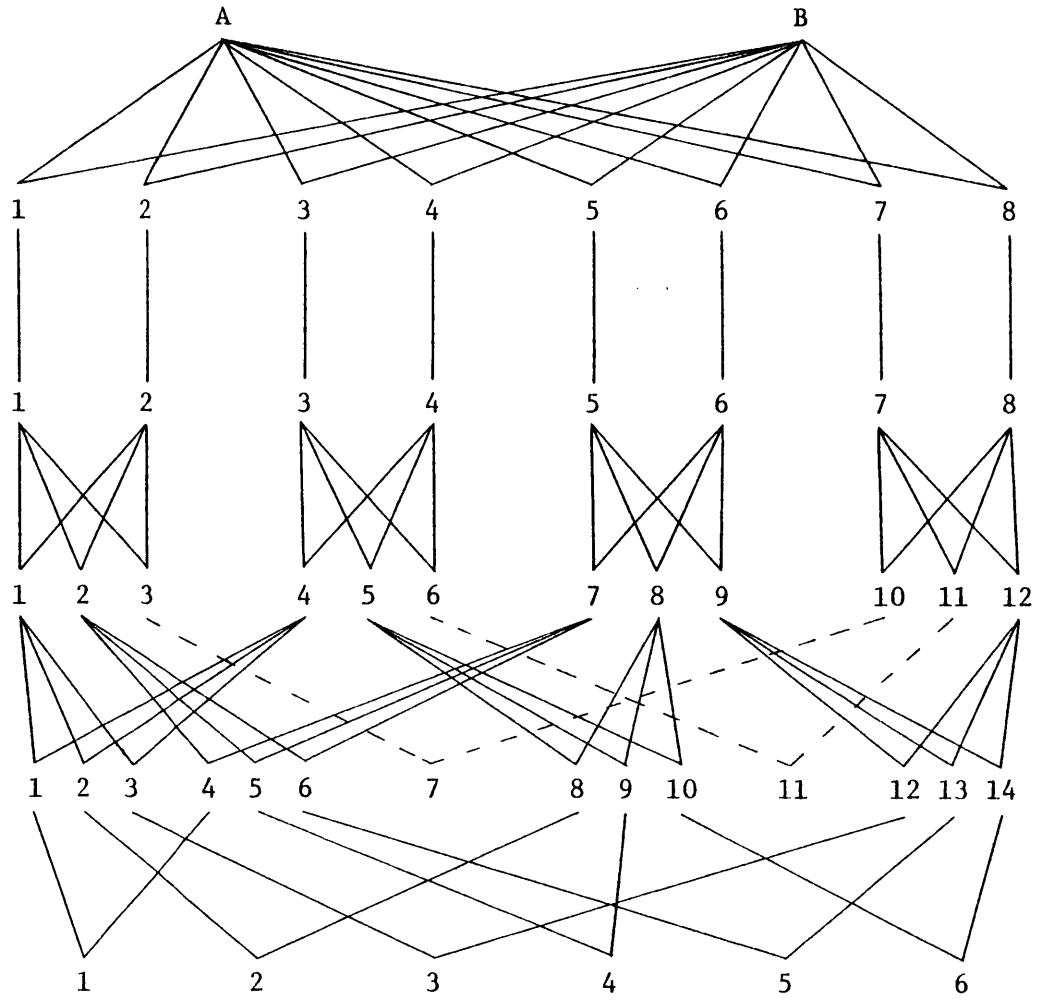
2.2 Description of the Selection Experiment

Starting with the F_1 hybrid of two unrelated inbred lines of maize and using recurrent selection for progeny yield, population improvement has been measured for four cycles. The two parental populations, Va 17 and Va 29, had been inbred for more than 10 generations and for all practical purposes are considered completely homozygous. All F_1 individuals are thus identical genotypically and have at most two alleles at each locus. They have two alleles at those loci at which the parental populations have different alleles. These loci are referred to as heterozygous. The loci for which the parental populations had the same allele are referred to as homozygous.

The F_1 plants were self-fertilized to produce the F_2 population. Since heterozygous loci have two alleles with frequencies of 0.5, the resulting F_2 population has the same characteristics as one which results from the cross-fertilization of the F_1 individuals. The coefficient of inbreeding at a given locus in the F_2 population is 0.5 because the original inbred lines each have an inbreeding coefficient of 1.0 and are unrelated to each other.

Random pairs of F_2 plants were crossed and seed from each cross was grown out to determine the average progeny yield. The N crosses with the highest average progeny yield were selected as parents of the next generation. Remnant seeds from the selected crosses were planted and a diallel mating design without selfing was used for intercrossing. All possible matings of plants from the different selected crosses were made using each plant in only one cross. This intercrossing scheme results in $\frac{N(N-1)}{2}$ crosses from which N crosses are selected to produce the next generation.

Selection based on average progeny yield and intercrossing offspring of the selected crosses has been continued for four cycles. Figure 2.1 is a diagram of the mating and selection scheme just described with N=4. The inbred parental populations are A and B. Individuals in the next five generations are designated by numbers. Only those individuals actually involved in the matings are included in the diagram. The progeny whose yields determined the crosses which were selected are omitted since they are used for testing purposes only.



Parental Populations

F_1

F_2= F_1 selfed

F_3= F_2× F_2

F_4= F_3× F_3

F_5= F_4× F_4

Figure 2.1 Mating and Selection Scheme

2.3 Development of the Model

2.3.1 Assumptions

In the development of a model to be used for simulating this recurrent selection experiment, relevant genetic and selection "parameters" must be identified and values for them must be chosen. Some simplifying assumptions must be made in order to obtain a model that can be analyzed. When making simplifying assumptions, the reality of the biological situation must be kept in mind. Obviously a simple model can at best be only an approximation to the experimental situation; however, while more complex models might better describe the actual experiment, they would be difficult to analyze and interpret. Understanding relatively simple models can give insight into the real-life situation. Even the simple model developed in this thesis is more complicated than any that can be treated theoretically without making still further assumptions.

In order to facilitate the investigation and interpretation of the effects of the genetic and environmental parameters on the progress due to selection, the following assumptions are made: (1) no mutation, (2) no linkage interference, (3) normal meiotic behavior, (4) no differential viability or fertility, and (5) no genotype by environment interaction. Although these assumptions most likely are not satisfied in the maize population being modeled, the consequences should be minimal relative to the parameters chosen for investigation. The most serious violation of the experimental situation is assuming no genotype by environment interaction. While this type of interaction does occur in nature, we are more interested in determining the effects of other factors. An additional assumption of the additivity of the

genotypic and environmental contributions to the phenotypic value or yield is discussed in Section 2.3.2(iv).

2.3.2 Design Parameters

The behavior of a population undergoing selection depends upon many factors both genetic and environmental. Five factors which we call design parameters have been chosen for investigation by simulation. These parameters are (i) selection intensity, (ii) number of loci that are heterozygous in the F_1, (iii) recombination value, (iv) environmental effect, and (v) gene action model. Values chosen for each of these parameters are based on results from previous experiments with maize as well as the procedures used in the actual experiment we are modeling:

(i) *Selection intensity* is related to the proportion of the population selected to produce the next generation. For the diallel mating procedure used in this experiment, the proportion of the population selected varies inversely according to the number of crosses, N, selected each cycle. Selection of N crosses from a population of $\frac{N(N-1)}{2}$ crosses gives a selected proportion of $\frac{2}{N-1}$. Thus, as the number of selected crosses increases, the proportion of the population selected decreases. In the actual experiment the number of crosses selected varied from a high of 20 in cycle 0 to a low of 12 in cycle 2. We chose values of N=10 and N=20 for this investigation to represent the range of values used in the original experiment. This is also a practical range for N in a recurrent selection experiment. Jenkins (1940) suggests the use of not less than 10 selected crosses, but the

population size increases rapidly to unmanageable numbers as N increases. With N=10, the next generation will have a total of 45 crosses from which to select the 10 with the highest average progeny yield, representing 22 percent of the population. When N=20, the next generation will have a total of 190 crosses from which the most productive 20 are selected, representing 11 percent of the population.

(ii) *The number of loci which are heterozygous in the* F_1, n_h, is the number of loci for which the parental lines had different alleles. Only those loci which are heterozygous will be affected by segregation during gametic formation and recombination of gametes to form new individuals. Heterozygous loci will be subject to selection pressure and the frequencies of the two alleles will change from cycle to cycle. Those loci which are homozygous in the F_1 will contribute the same allele to every gamete throughout the entire selection procedure and will not change with selection. Values of 50 and 100 heterozygous loci in the F_1 have been chosen for investigation, with the total number of loci contributing to yield taken to be 100. Maize has 10 pairs of chromosomes and for simulation purposes the 100 loci are considered to be distributed equally on the 10 pairs of chromosomes. The values chosen for n_h, the number of heterozygous loci, represent 50 percent and 100 percent of the total number of loci contributing to yield. When 50 percent of the loci are heterozygous, only the first five loci on each chromosome are segregating which has the effect of reducing the length of each chromosome from 10 loci to 5 loci. Using these values we can determine what effect the percentage of the total number of loci, the allelic frequencies of which can be changed by selection,

has on the improvement of the population and the approach to homozygosity.

(iii) *Recombination value* is a measure of the spacing of loci linked on the same chromosome pair. The number of crossovers which occur between the homologous chromosomes in the formation of gametes is a function of the distance between loci. The probability of a recombination between two loci is called the recombination value. For simplicity we assume a constant value, r, for the probability of a recombination between any two adjacent loci on the same chromosome in the genotype. This assumption means that the loci are equally spaced along the chromosome. We further assume that no crossover interference occurs; that is, all recombinations occur independently of each other. Even though those loci which have been identified in maize as contributing to yield are not equally spaced on the 10 chromosomes, it is conceivable that the average recombination value for all pairs of yield loci in the genotype will be approximately the same whether or not the loci are equally spaced. For this reason it is of interest to compare the average recombination values over the chromosome and genotype for the values chosen to be used in the simulation study.

The average recombination value for all loci linked together is a function of r and the number of loci on the same chromosome. The probability of a crossover between any two loci, i and j, on a chromosome which are m spaces apart can be obtained by expanding the expression $(r + (1-r))^m$ and summing the odd terms. The result is

$$r_m = \sum_{\text{all odd } x}^{m} \binom{m}{x} r^x (1-r)^{m-x}$$
$$= \frac{1}{2}\left(1 - (1-2r)^m\right) . \qquad (2.1)$$

The average recombination value between all possible pairs of loci on the same chromosome is

$$\bar{r}_c = \frac{\sum_{m=1}^{n-1} (n-m)\left(\frac{1}{2}(1-(1-2r)^m)\right)}{\frac{n(n-1)}{2}}$$
$$= \frac{\sum_{m=1}^{n-1} (n-m)\, r_m}{\frac{n(n-1)}{2}} \qquad (2.2)$$

where n is the total number of loci linked on the chromosome and m is the spacing between pairs of loci. $\bar{r}_c$ is a weighted average of the values of r_m for m=1,2,...,n-1. There are a total of $\frac{n(n-1)}{2}$ possible pairs of loci to be considered and there are n-m pairs which are m spaces apart.

If more than one chromosome is involved and the number of loci on each chromosome is the same, the average recombination value between any two loci in the entire genotype can be evaluated. This is

$$\bar{r}_g = \frac{2(n-1)\,\bar{r}_c + n(\ell-1)}{n\ell-1} \qquad (2.3)$$

where $\bar{r}_c$ is the average recombination value for pairs of loci on the same chromosome, n is the number of loci on each chromosome and ℓ is

the number of chromosomes.

Results presented by Moll, et al. (1964) indicate that chromosomes with 6 loci contributing to yield behave as if they have a constant recombination value of 0.2 and chromosomes with 11 loci contributing to yield behave as if they have a constant recombination value of 0.1. As we mentioned in Section 2.3.2(ii), for n_h=100 each of the 10 chromosomes have 10 heterozygous loci and for the purposes of selection the chromosomes are 10 loci in length. For n_h=50 the first five loci on each of the 10 chromosomes are heterozygous and for the purposes of selection the chromosomes are 5 loci in length. Using this information, we have chosen values for r of 0.2 and 0.05 to cover the range of reasonable values and allow for comparison of the effects of moderate vs. tight linkage. Values for $\bar{r}_c$ and $\bar{r}_g$ for 10 loci on each of 10 chromosomes are given in Table 2.1 for r=0.2 and r=0.05. As is evident from the table, the difference in r has a greater effect on $\bar{r}_c$ than on $\bar{r}_g$. Thus the average recombination value over the entire genotype is not very different for the two values of r.

(iv) *Environmental effects* are everpresent factors in the measurement of yield, and the environmental variance, σ_E^2, will be used as a measure of these effects. We assume that the yield or phenotypic value of individuals is the sum of a genetic component and an environmental component and that these components are independent and uncorrelated with each other. The phenotypic value and its variance are given by

TABLE 2.1

Average Recombination Values
for 10 loci on Each of 10 Chromosomes

	r=0.2	r=0.05
$\bar{r}_c$	0.375	0.158
$\bar{r}_g$	0.489	0.469

$$Y = G + E,$$

and (2.4)

$$\sigma_Y^2 = \sigma_G^2 + \sigma_E^2 .$$

The genotypic variance, σ_G^2, is computed from the genotypic values of individuals in the population (see (v)). Environmental variances, σ_E^2, of 0 and 60 were chosen for investigation. When σ_E^2=0, exact genotypic evaluations can be made. This is obviously not a realistic situation, but is included for comparative purposes. σ_E^2=60 represents a situation in which the coefficient of variation is approximately 10 percent for average yields of 75 in the F_1 based on 25 progeny. Coefficients of variation for average yield in the actual experiment ranged from 8 to 12 percent (Genter, 1976).

(v) The genotypic value, G, of individuals can be expressed in many different ways as a function of the loci involved. *Gene action models* for which the genotypic value can be expressed as a function of the number of loci in different phases are easy to represent mathematically and are particularly useful in simulation. Properties of gene action models of this type are given by Hill (1963). With 2 alleles possible at each segregating locus, there are only three possible phases which will be designated (+,+), (+,-), and (-,-). Symmetric models in which each locus has the same effect and the frequency of the more favorable allele (+) is the same for all loci have been discussed by Horner and Kempthorne (1955), Horner (1956), and Hill (1963). Hill, et al. (1963) presented results for situations in which the loci are divided into two or more subgenotypes. Within

each subgenotype, all loci follow the same gene action model with the frequency of the + allele the same at all loci.

Some classical gene action models, such as complementary, duplicate factor, multiplicative, and optimum number, are discussed by Horner (1956). The gene action models we will use are constructed from three basic models considered by Hill, et al. (1963). These models are (i) a linear model, (ii) a bottleneck model, and (iii) a model which incorporates a major gene with modifiers. In each of these basic models, the genotypic value, G, of an individual is expressed as a function of x_2, x_1, and x_0, the number of (+,+), (+,-), and (-,-) loci, respectively, in the genotype or subgenotype.

For the linear model, the genotypic value is given by

$$G = ax_2 + dx_1 - ax_0 \ . \tag{2.5}$$

The choices of a and d determine what type of linear model is being used. Complete dominance is represented by d=a, partial dominance by $0<d<a$, overdominance by $d>a$, and additivity by d=0. We can have several subgenotypes, each of which is controlled by a linear model with different values of a and d. The relative effect of the (+,+) phase in each subgenotype will be found by comparing the values of a for each subgenotype model. In any linear model there is no interaction between loci; and consequently, the genotypic variance is comprised of only additive and dominance variance components.

The bottleneck model has two subgenotypes. In the first subgenotype, the loci contribute to the genotypic value according to a linear model. The loci in the second subgenotype control the contri-

bution from the linear subgenotype and behave according to a threshold model. If the total number of (+,+), and (+,-) loci in the threshold subgenotype is greater than or equal to a threshold value, τ, then the linear subgenotype is expressed. The genotypic value for the bottleneck model is then given by

$$G = (ax_{12}+dx_{11}-ax_{10}) \sum_{t=\tau}^{n_2} 0^{|x_{22}+x_{21}-t|} \tag{2.6}$$

where a and d are as defined earlier; x_{i2}, x_{i1}, and x_{i0} represent the numbers of (+,+), (+,-), and (-,-) loci in the i^{th} subgenotype; n_2 is the total number of loci in the second or threshold subgenotype; and τ is the value of the threshold. Loci in the two different subgenotypes can interact with each other; therefore, the genotypic variance will include, in addition to additive and dominance components, epistatic components through order n_2. A more detailed discussion of variance components is given in Chapter VII.

There are two subgenotypes in the model which incorporates a major gene with modifiers. One subgenotype consists of a single locus which is called a major gene. This locus contributes a constant amount, U, to the genotypic value. The modifier loci in the second subgenotype follow a linear model and are expressed only when the major gene is homozygous for the less desirable allele. Additionally, in this situation a constant value, V, is subtracted from the genotypic value representing a large detrimental effect. The contribution from the modifier subgenotype alters the detrimental effect of the major gene when it is in the (-,-) phase.

For the major gene with modifies model, the genotypic value is given by

$$G = U - (V - ax_{22} - dx_{21} + ax_{20})\ 0^{|x_{12}+x_{11}|} \qquad (2.7)$$

where U, V, a, d, x_{i2}, x_{i1}, and x_{i0} are as defined previously. As in the bottleneck model, it is possible for the loci in this model to interact with each other. In this situation the genotypic variance includes epistatic components of order 2.

In the simulation study, we used three gene action models -- simple linear, multiple linear, and epistatic. In the simple linear model, all 100 loci in the genotype contribute in the same way according to a linear model. The multiple linear model has four subgenotypes, each with 25 loci. Within each of these four subgenotypes, the loci follow a linear model with different values of a and d. All three basic gene action models are included in the six subgenotypes of the epistatic model. There are four linear subgenotypes in addition to the bottleneck and major gene with modifiers subgenotypes. This model is so named because it is the only model in which interaction of loci, and therefore epistatic variation, can occur.

Values chosen for a were somewhat arbitrary, but those chosen for d were based on estimates reported in the literature of the average degree of dominance, $\bar{d}$, for loci contributing to yield. One method of estimation, based on the ratio of mean squares from the analysis of variance of the appropriate experimental design, is presented by Comstock and Robinson (1948,1952). Using this method, Robinson, et al. (1949) and Gardner, et al. (1953) report estimates of $\bar{d}$ in the over-

dominant range, but indicate that these estimates may be biased upward by linkage disequilibrium, in particular, repulsion phase linkages. It has been pointed out that so called pseudo-overdominance has the same significance as true overdominance; however, after several cycles of selection linkage blocks may be broken up and the cumulative effect of the loci may not be in the overdominant range.

Since most researchers would agree that "the choice of method to be used for the genetic improvement of crop plants is dependent upon the type of gene action involved," (Lindsey, et al., 1962), a great deal of effort has gone into determining the predominant type of gene action controlling corn yield. In an experiment reported by Penney, et al. (1962), the "predominant type of selection appears to have been for genes exhibiting complete or partial dominance or largely additive effects." Compton, et al. (1965) report that "additive gene action with partial to complete dominance is adequate to explain most of the results found in the literature on corn." According to Moll and Robinson (1967), the "average dominance for genes affecting yield of ear corn is in the range of partial to complete." Robinson, et al. (1958), Gardner and Lonnquist (1959), Lonnquist (1953), Jinks (1955), and Penney and Eberhart (1971) all support the hypothesis of partial to complete dominance of alleles at loci controlling yield.

Little evidence exists to support an epistatic gene action model. Bauman (1959) and Gorsline (1961) report evidence of epistasis in yield of maize. However, the type or amount of epistasis could not be determined. On the other hand, Compton, et al. (1965) have not found epistasis to be an important source of variation. Stuber, et al. (1966)

report that "epistatic effects contribute little to the genetic variability of the characters studied." Epistatic variation is such a small fraction of the total genotypic variation present even in epistatic gene action models that it is difficult to measure. In addition, linkage can bias estimates of epistatic variance components and lead to a lack of significance in tests of hypothesis. For this reason and for purposes of comparison with the non-epistatic (linear) models it seemed worthwhile to include an epistatic model in the study.

In the linear subgenotypes of all three models adopted for use in the simulation study, the values chosen for d represent partial dominance of the more favorable (+) allele. For the 4 linear subgenotypes in the multiple linear and epistatic models, values of a were chosen such that the largest effect of loci in one of the subgenotypes was 4 times that of the smallest effect. The specific values for the constants in each of the three gene action models used in this study were chosen from a large number of possible sets of values after a preliminary investigation. A summary of the models with the specific values for the constants is presented in Table 2.2. In the preliminary investigation, these three models gave results similar to those from the original experiment and appeared diverse enough to detect differences in the investigation of the long term behavior of yield and the approach to homozygosity.

All possible combinations of the values of the 5 design parameters chosen for investigation resulted in 48 different situations to simulate. Since replication is essential in a Monte Carlo study, we chose to simulate only 24 of the 48 possible situations in order to reduce the amount of computer time and cost. The method used to choose the

TABLE 2.2

Summary of Gene Action Models Used in the Simulation Study

Simple Linear

$$G = x_2 + 0.75y_1 - x_0$$

Multiple Linear

$$G = \sum_{i=1}^{4} (a_i x_{i2} + d_i x_{i1} - a_i x_{i0})$$

i	a_i	d_i	n_i
1	2.0	1.75	25
2	1.5	1.25	25
3	1.0	0.75	25
4	0.5	0.25	25

Epistatic

$$G = \sum_{i=1}^{4} (a_i x_{i2} + d_i x_{i1} - a_i x_{i0}) + (a_5 x_{52} + d_5 x_{51} - a_5 x_{50}) \sum_{t=\tau}^{n_6} 0^{|x_{62}+x_{61}-t|} + U - (V - a_8 x_{82} - d_8 x_{81} + a_8 x_{80})\, 0^{x_{72}+x_{71}}$$

i	a_i	d_i	n_i	
1	2.0	1.75	20	
2	1.5	1.25	20	
3	1.0	0.75	20	
4	0.5	0.25	30	
5	1.0	1.0	7	
6			3	$\tau = 2$
7			1	$U = 1$ $V = 10$
8	1.0	1.0	9	

particular 24 combinations of parameter values is discussed in Chapter IV. Four replications of each of the chosen situations were run giving a total of 96 simulation runs. Details of the computer program used are given in the next chapter.

In many problems, simulation is used to compare several types of estimators. Under these circumstances the average of a large number of replications of a particular situation and the variance of this average are important for comparative purposes. However, in this dissertation, the results of the simulation program will be used as guidelines for the interpretation of the behavior of variables over time and to aid in understanding the importance of different values of design parameters in a recurrent selection experiment. For these reasons in addition to economic considerations, only four replications of each design parameter combination have been run and the simulated values should not be taken as actual estimates.

Chapter III

THE SIMULATION PROCEDURE

3.1 Introduction

Simulation of the previously described recurrent selection scheme was performed on an IBM-370 computer system. Generation of the F_1 genotype, generation of gametes, mating of individuals, generation of offspring, and selection of high yielding crosses were accomplished using random numbers generated as power residues with a kernel of 65539. For a discussion of the properties of this random number generator see IBM Manual C20-8011. These random numbers, u_i, follow a uniform distribution on the interval (0,1). If an event, E_i, has a probability of occurrance equal to p, then E_i is realized when $0 \leq u_i \leq p$ and $\bar{E}_i$ is realized when $p < u_i \leq 1$. For example, in the generation of gametes each of the two alleles at the first locus of an individual has a probability of 0.5 of being chosen. A uniform random number is generated and if $u \leq 0.5$, the first allele is chosen for the gamete; if $u > 0.5$, the second allele is chosen.

Starting values for the uniform random number generator were chosen from a table of random numbers. Each of the 96 simulation runs had a unique seed or starting value which was an odd integer with 6 or more digits and the initial 500 random numbers in each run were discarded. Genotypic values were computed for each individual according to the models of gene action discussed in the previous chapter. Yield or phenotypic value was calculated as $Y = G + \sigma_E z$, where z is a standard normal random variable and σ_E^2 is the environmental variance. The z's

are generated from uniform random variables using the following transformation given by Box and Muller (1958):

$$z_1 = \sqrt{-2 \ln u_1} \sin 2\pi u_2$$

and (3.1)

$$z_2 = \sqrt{-2 \ln u_1} \cos 2\pi u_2 \ .$$

If u_1 and u_2 are uniform random variables on the interval (0,1) then z_1 and z_2 are independent standard normal random variables. Therefore, we are using the transformation in (3.1) to generate standard normal random variables from uniform random variables in the computer. The computation of the z_i's is incorporated into a subroutine. The first time the subroutine is used, two values of z are computed from two values of u. One of the values of z is used and the other is stored for subsequent use.

3.2 Programming Techniques and Computer Representation

Genotypes are represented in the computer as 100 ordered pairs of zeros and ones corresponding to the 100 loci which determine yield. The more desirable (+) allele is represented by a one and the less desirable (-) allele, by a zero. The columns of this 100 x 2 array represent the homologous pairs of chromosomes. Length of the 10 pairs of chromosomes, as measured by the number of loci assigned to each pair, can vary; but the total number of loci must equal 100. Spacing of the loci along the homologous pairs of chromosomes is controlled by the recombination value.

Gametes are generated from individuals by selecting one of the two alleles present at each locus and storing it in a one-dimensional array of length 100. For each chromosome pair, the initial allele is selected

with a probability of 0.5, and selection of successive alleles is controlled by the recombination value. The probability that the allele chosen from the i^{th} locus of a given chromosome pair comes from the same homologous chromosome as the allele chosen from the $(i-1)^{st}$ locus is 1-r, and the probability of a crossover between the $(i-1)^{st}$ and i^{th} loci is r. Individuals are formed by storing 2 gametic vectors of zeros and ones in a 100 x 2 array.

Pairs of individuals, representing matings or crosses, are stored in three-dimensional arrays of size 2 x 100 x 2. Offspring are generated by pairing two random gametes: one from each parent or individual comprising the cross. In the F_2 generation (selection cycle zero), $\binom{N}{2}$ random pairs of individuals are formed from which N pairs will be selected to be parents of the next generation. In succeeding cycles of selection, matings are accomplished by intercrossing offspring from each of the N selected crosses. N-1 offspring are generated from each selected cross giving a total of N(N-1) individuals, which are used to form new crosses. Each mating involves 2 offspring from different selected crosses and no individual is used in more than one mating. Thus we again have $\binom{N}{2}$ crosses and the selection procedure is repeated.

We will use the same mating and selection procedure as practiced in the actual experiment. The diallel mating scheme is comparable to random mating since there are an equal number of matings of individuals in each selected group; but finite population size and selection will alter the results expected for an infinite random mating population. The number of progeny used to evaluate each cross is fixed at 25. Each simulation run will continue for 30 cycles of selection or until the

population reaches complete homozygosity, whichever comes first.

The selection procedure for each cycle or generation involves choosing N crosses to be the parents of the next generation. The basis of selection is the progeny average yield or phenotypic value and the N crosses with the highest progeny average yield are selected. In the computer simulation program, selection of N crosses from $\binom{N}{2}$ crosses is accomplished in the following manner.

1. For the first N crosses, generate offspring; then compute and store the progeny average genotypic value and yield, the number of + alleles at each locus, the number of homozygous loci, and the parental genotypes.
2. For the (N+1)st cross, generate offspring, compute the progeny average yield and compare this average with the N average yields previously stored. If the progeny average yield for the (N+1)st cross is greater than the minimum progeny average yield previously stored, calculate information listed in Step 1 for the (N+1)st cross; otherwise, go to Step 4.
3. Store the new information calculated in Step 2 in the place of the information corresponding to the cross which had the minimum yield.
4. Repeat Step 2 with the (N+2)nd and subsequent crosses until all $\binom{N}{2}$ crosses have been tested. At this point the N progeny average yields stored will be the highest ones.

Genotypes of the N selected crosses are stored in a four-dimensional array of size N x 2 x 100 x 2.

From the genotypes of the N selected crosses, (N-1) new progeny are generated for each cross. Intercrossing of these N(N-1) individuals so that each individual contributes to one and only one mating is illustrated in Figure 3.1 for N=4. This is accomplished in the computer program as follows.

1. Generate new offspring from the first and second selected crosses. These two new offspring form the first mating in the next cycle of selection.
2. Generate new offspring from the first and third selected crosses forming the second mating.
3. Continue until the (N-1)st offspring from the first selected cross has formed a mating with an offspring from the Nth selected cross.
4. Generate new offspring from the second and third selected crosses to form a mating.
5. Continue generating offspring from selected crosses and pairing the offspring to form new matings until each selected cross has contributed N-1 offspring to N-1 matings in the next generation. Intercrossing and selection are continued until the entire population is homozygous for all 100 loci or until a specified number of cycles of selection have been completed.

Each simulation run begins with the generation of the F_1 or hybrid genotype. Control cards are used to specify the number and spacing of heterozygous loci on each of the 10 pairs of chromosomes. Heterozygous loci are represented by the ordered pairs (0,1) or (1,0), each with a probability of 0.5. If the total number of heterozygous loci is less

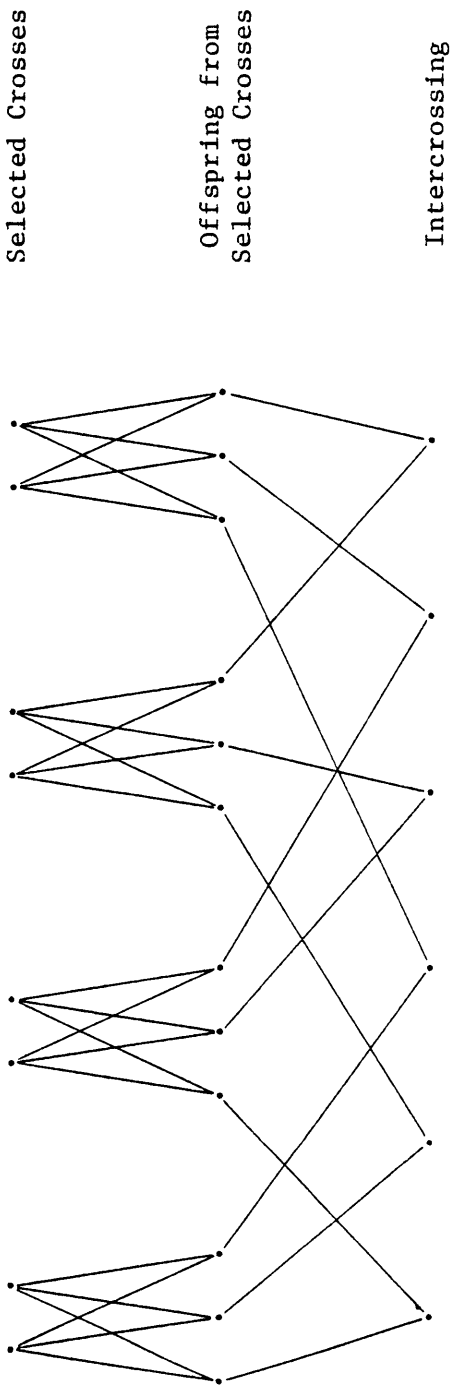


Figure 3.1 Intercrossing of Offspring from Selected Crosses

than 100, the remaining loci are homozygous. The specific designation for a homozygous locus in the F_1 is controlled by the generation of a uniform random number. If $u \leq 0.25$, the locus will be homozygous for the less desirable (-) allele and is designated (0,0). Otherwise, the locus will be homozygous for the more desirable (+) allele and is designated (1,1). Since we are not allowing for mutations to occur, all loci that are homozygous in the F_1 generation remain in the same homozygous state for all individuals in subsequent generations.

Genotypic values of individuals are computed from the genotype according to specific models of gene action. Each of the basic gene action models described in Chapter II is a function of the number of (+,+), (+,-), and (-,-) loci in the genotype for which the machine representations are (1,1), (1,0), and (0,0). Any gene action model which is a combination of the basic gene action models may be used. Up to six different linear subgenotypes may be included in creating a complex gene action model. The number of loci in each subgenotype as well as the specific gene action model for that subgenotype are specified on control cards. Assignment of particular loci to a given subgenotype is done within the program. Each locus is assigned a uniform random number and the random numbers are ranked. The loci associated with the lowest n_1 random numbers are assigned to subgenotype 1. The loci associated with the next n_2 random numbers are assigned to subgenotype 2. This assignment of loci to subgenotypes continues until all 100 loci have been allocated to one of the subgenotypes. A control array of these assignments is stored and used each time a genotypic value is calculated.

3.3 Output From the Simulation Program

Initial output from the simulation program consists of the control array for computing genotypic values, the F_1 genotype and genotypic value, F_1 average yield, F_2 average genotypic value, F_2 average yield, F_2 average yield as a percent of the F_1 average yield, F_2 genotypic variance, and the variance of the F_2 yield. Values assigned to the genetic design parameters, the seed for the uniform random number generator, and identification of the simulation run are also printed.

After each cycle of selection, information is printed for all crosses and selected crosses. Progeny average yield and average yield as a percent of the F_1 average yield are printed for each cross. Progeny average genotypic value, progeny average yield, average yield as a percent of the F_1 average yield, minimum and maximum average yield as a percent of F_1 average yield, genotypic variance, the variance of yield, the percent of homozygous loci, and the frequency of the more desirable (+) allele for each locus are calculated using combined information from all progeny of all crosses and printed for each cycle.

Additional information is printed for the selected crosses. Progeny average genotypic value and the percent of homozygous loci for the progeny of each selected cross are printed. Progeny average genotypic value and average yield, average yield as a percent of the F_1 average yield, and the percent of homozygous loci are calculated and printed using all progeny of selected crosses.

The flow chart in Figure 3.2 illustrates the basic steps involved in the simulation program. The steps in block A are performed only at the beginning of each simulation run. Block B contains the steps for

printing all the information after each cycle of selection. Block C includes the steps for the intercrossing and selection procedures which are repeated each cycle.

All calculations with the exception of the variances are performed using single precision. All intermediate steps involved in the computation of the genotypic and phenotypic variances were performed using double precision.

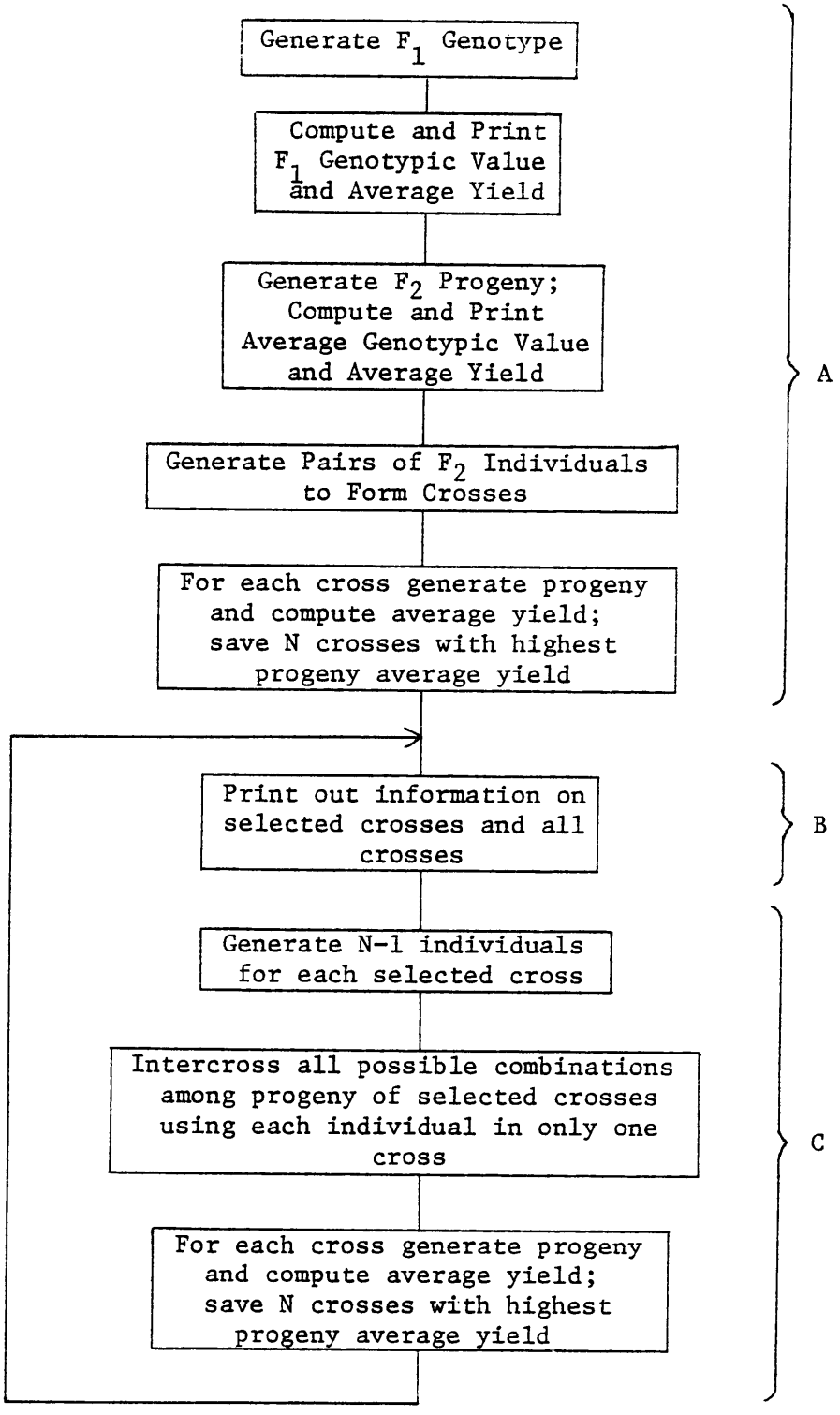


Figure 3.2. Flow Chart for Simulation Program

Chapter IV

STATISTICAL DESIGN AND ANALYSIS PROCEDURES

4.1 Introduction

The model developed in Chapter II for this recurrent selection experiment is a function of five design parameters. The computer execution time required for each simulation run vaired with different combinations of parameter values. For example, the values chosen for selection intensity have a large effect on the time requirements. With N = 10, the execution time for a single simulation run averaged 20 minutes; but for N = 20, the average execution time tripled to 60 minutes. In any simulation study, replication of each situation is important. Since the computer time and cost were limiting factors in the total number of simulation runs which could be performed, only a fraction of the possible combinations of parameter values have been simulated. The design used to select this subset of combinations of parameter values is discussed in Section 4.2.

A very large amount of output was generated by the simulation program and statistical treatment is required to analyze the results and make inferences concerning the selection experiment. The statistical analyses used to analyze the simulation output are presented in Section 4.3.

4.2 Statistical Design

Four of the five design parameters (factors) chosen for investigation in this simulation study have been assigned two values (levels) each. These parameters are selection intensity (N), recombination

value (r), number of loci which are heterozygous in the F_1 (n_h), and environmental variance (σ_E^2). The fifth parameter, gene action model, has 3 levels. Each one of the three gene action models summarized in Table 2.2 constitutes a different level. In all there are $2^4 \times 3 = 48$ combinations of values of the design parameters of which 24 are used in the simulation program. If too few combinations are used, then the effects of some of the design parameters cannot be tested.

The specific combinations chosen represent a $\frac{1}{2}$ - fraction of a 2^4 factorial design, where the 4 two-level factors are as given above, combined with the three gene action models, i.e. a $2^{4-1} \times 3$ fractional factorial. A summary of the 5 factors, the design parameter notation, and the levels assigned to each factor are contained in Table 4.1. The 24 combinations of parameter values to be investigated are listed in Table 4.2. The 2^{4-1} fractional factorial is obtained by using the four-factor interaction as the defining contrast. This means that the main effects are confounded with three-factor interactions and three of the two-factor interactions are confounded with the other three two-factor interactions. Table 4.3 gives the effects and their aliases for this 2^{4-1} fractional factorial design.

4.3 Statistical Analyses

Two types of analysis, analysis of variance and regression analysis, will be used in analyzing the output of the simulation program. The variables which will be analyzed are percent homozygosity (H), genotypic value (G), and yield (Y). All of the observations of H, G, and Y used in these analyses are averages of progeny in a given cycle of

TABLE 4.1

Notation and Values of the Five Design Parameters

Factor	Design Parameter	Symbol	Subscript	Subscript Value	Parameter Value
SI	Selection Intensity	N	i	1	N=10
				2	N=20
RV	Recombination Value	r	j	1	r=0.2
				2	r=0.05
NHL	Number of heterozygous loci in the F_1	n_h	k	1	n_h=50
				2	n_h=100
EV	Environmental Variance	σ_E^2	ℓ	1	σ_E^2=0
				2	σ_E^2=60
GAM	Gene Action Model	M	m	1	M_1=Simple Linear
				2	M_2=Multiple Linear
				3	M_3=Epistatic

TABLE 4.2

Parameter Values for the 24 Simulation Runs

N	r	n_h	σ_E^2	Gene Action Model
10	0.2	50	0	Simple Linear Multiple Linear Epistatic
		100	60	Simple Linear Multiple Linear Epistatic
	0.05	50	60	Simple Linear Multiple Linear Epistatic
		100	0	Simple Linear Multiple Linear Epistatic
20	0.2	50	60	Simple Linear Multiple Linear Epistatic
		100	0	Simple Linear Multiple Linear Epistatic
	0.05	50	0	Simple Linear Multiple Linear Epistatic
		100	60	Simple Linear Multiple Linear Epistatic

TABLE 4.3

Effects and Their Aliases for a ½-Fraction of a 2^4 Factorial Design with the 4-Factor Interaction as the Defining Contrast

Effect	Alias
SI	(RV)×(NHL)×(EV)
RV	(SI)×(NHL)×(EV)
NHL	(SI)×(RV)×(EV)
EV	(SI)×(RV)×(NHL)
(SI)×(RV)	(NHL)×(EV)
(SI)×(NHL)	(RV)×(EV)
(SI)×(EV)	(RV)×(NHL)

SI = Selection Intensity

RV = Recombination Value

NHL = Number of Heterozygous Loci

EV = Environmental Variance

selection. A superscript A indicates that the progeny from all $\frac{N(N-1)}{2}$ crosses are included in the average while a superscript S indicates that only the progeny from the N selected crosses are included in the average. Subscripts are used to indicate which values of the design parameters were used as well as to which replication and selection cycle the observation belongs, e.g. observations for yield will be denoted by $Y^{A}_{ijk\ell mnt}$ or $Y^{S}_{ijk\ell mnt}$, where the letters i, j, k, ℓ, and m are used for the five design parameters, n indicates the replication number, and t indicates the cycle of selection. For example, $Y^{S}_{1,1,2,1,3,1,15}$ is the average yield of the progeny of the selected corsses for N = 10, r = 0.2, n_h = 100, σ^2_E = 0, the epistatic gene action model, the first replication of this particular combination of parameter values and the 15th cycle of selection. A summary of the specific parameter values associated with the subscript values is given in Table 4.2.

Percent homozygosity (H), is the percentage of loci which are homozygous in the progeny each cycle of selection and for each combination of i, j, k, ℓ, m, and n, H is given by

$$H^{A}_{t} = \frac{\sum_{\text{all progeny}} (x_2+x_0)}{25n_h(\frac{N(N-1)}{2})} \tag{4.1}$$

or

$$H^{S}_{t} = \frac{\sum'_{\text{all progeny}} (x_2+x_0)}{25\ n_h\ N} \tag{4.2}$$

where x_2 and x_0 are the number of (+,+) and (−,−) loci, respectively, 25 is the number of progeny generated for each cross, n_h is the number of heterozygous loci in the F_1, N is the number of crosses selected each cycle and Σ' indicates only selected crosses are used in the summation. When $n_h = 50$, only one-half of the total number of loci contributing to yield are heterozygous in the F_1. Those loci which are homozygous in the F_1 remain homozygous throughout the selection procedure and do not contribute to the change in homozygosity. Therefore, only those loci which are heterozygous in the F_1 are included in the calculation of (4.1) and (4.2). Further discussion of percent homozygosity and its relationship to the coefficient of inbreeding can be found in Chapter VI. Genotypic value (G^A and G^S) and yield (Y^A and Y^S) were defined in Chapter II.

One other variable to be analyzed is a function of two values which are calculated each cycle of selection for the entire population. The ratio of the genotypic variance, σ_G^2, to the phenotypic variance, σ_Y^2, is referred to as heritability in the broad sense. We will use $h_B^2 = \sigma_G^2/\sigma_Y^2$ as an estimate of heritability in the broad sense and analyze this variable for the selection procedure.

Analyses of variance are run for selection cycles 5, 10, 15, 20, and the last cycle of selection, cycle T. In each analysis there are 96 observations, one from each simulation run. Since there is confounding of effects and interactions in this factorial design, we must assume some of the interactions to be zero in order to test the significance of any main effect except gene action model and any two-factor interaction involving parameters other than gene action model. The

particular two-factor interactions assumed to be zero depend upon which variable is being analyzed. For all analyses, the two-factor interactions involving gene action model can be tested and all interactions involving three or more factors are assumed to be zero.

In the analysis of percent homozygosity, the two-factor interactions involving EV with SI, RV, and NHL are assumed to be zero. It seems reasonable to assume that the effect of EV on the percent homozygosity at any cycle in the selection program will be the same regardless of the levels of SI, RV, or NHL. Furthermore, an interaction effect of the different levels of SI, RV, and NHL with each other would be understandable in genetic terms. The model for H^A and H^S then is

$$\begin{aligned} H_{ijk\ell mn(t)} = \mu &+ SI_{i(t)} + RV_{j(t)} + NHL_{k(t)} + EV_{\ell(t)} + GAM_{m(t)} \\ &+ (SI\times RV)_{ij(t)} + (SI\times NHL)_{ik(t)} + (RV\times NHL)_{jk(t)} \\ &+ (SI\times GAM)_{im(t)} + (RV\times GAM)_{jm(t)} + (NHL\times GAM)_{km(t)} \\ &+ (EV\times GAM)_{\ell m(t)} + R_{n(t)} + \eta_{ijk\ell mn(t)} , \\ i,j,k,\ell=1,2 \quad m=1,2,3 \quad &\text{and} \quad n=1,2,3,4 , \end{aligned} \tag{4.3}$$

where $SI_{i(t)}$, $RV_{j(t)}$, $NHL_{k(t)}$, $EV_{\ell(t)}$, and $GAM_{m(t)}$ represent the main effects of the design parameters selection intensity, recombination value, number of heterozygous loci in the F_1, environmental variance and gene action model, respectively; $(SI\times RV)_{ij(t)}$, $(SI\times NHL)_{ik(t)}$, $(RV\times NHL)_{jk(t)}$, $(SI\times GAM)_{im(t)}$, $(RV\times GAM)_{jm(t)}$, $(NHL\times GAM)_{km(t)}$, and $(EV\times GAM)_{\ell m(t)}$ represent the two-factor interactions; $R_{n(t)}$ represents the replication effect; and $\eta_{ijk\ell mn(t)}$ represents the error term.

A separate analysis is performed for cycles 5, 10, 15, 20, and T; therefore, the subscript t is enclosed in parentheses to indicate that it is not a variable in the model.

For the analysis of genotypic value, yield and heritability in the broad sense, the two-factor interactions involving parameters other than selection intensity or gene action model are assumed to be zero. The two-factor interactions affected by this assumption are NHL × EV, RV × EV and RV × NHL and they are confounded with SI × RV, SI × NHL, and SI × EV, respectively. Selection intensity is the one parameter which is under the control of the experimenter and the effect of two-factor interactions involving SI with RV, NHL, and EV may be informative in the determination of which value of N should be used in a recurrent selection experiment. Furthermore, we do not anticipate the effect of RV and NHL on genotypic value, yield, or heritability in the broad sense to be different for different values of EV. Finally, the effect of r on these variables is not expected to be different for the two values of NHL.

The model used in the analysis of genotypic value is

$$\begin{aligned} G_{ijk\ell mn(t)} = \mu &+ SI_{i(t)} + RV_{j(t)} + NHL_{k(t)} + EV_{\ell(t)} + GAM_{m(t)} \\ &+ (SI\times RV)_{ij(t)} + (SI\times NHL)_{ik(t)} + (SI\times EV)_{i\ell(t)} \\ &+ (SI\times GAM)_{im(t)} + (RV\times GAM)_{jm(t)} + (NHL\times GAM)_{km(t)} \\ &+ (EV\times GAM)_{\ell m(t)} + R_{n(t)} + \eta_{ijk\ell mn(t)} , \end{aligned}$$

$$i,j,k,\ell=1,2 \quad m=1,2,3 \quad \text{and} \quad n=1,2,3,4 , \qquad (4.4)$$

where the symbols are as defined above. The only difference is that the two-factor interaction SI × EV is included in the model for genotypic value and the two-factor interaction RV × NHL is assumed to be zero. This same model is used for the analyses of yield and heritability in the broad sense.

Univariate, rather than multivariate, analyses of variance were performed on these variables even though there are obvious relationships between the variables because we are interested in looking at the effects of the design parameters on each variable separately.

Regression analysis is used to describe the behavior of percent homozygosity, genotypic value, yield, and heritability in the broad sense for the duration of the selection experiment for each of the 24 combinations of the design parameters. For this recurrent selection scheme in which the individuals chosen to produce the next generation are a specific subset of the population, the characteristics of the new population are not independent of the previous population. The assumptions necessary to use ordinary least squares regression analysis are not satisfied because the residuals are possibly correlated.

A test for the correlation of error terms for each regression was performed using the Durbin-Watson test statistic

$$D = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \quad . \tag{4.5}$$

The e_t are residuals obtained from fitting an ordinary least squares regression line to the observations. The null and alternative hypotheses for this test are

$$H_0: \ \rho = 0$$

and (4.6)

$$H_1: \ \rho > 0 \ .$$

The majority of these tests indicated a significant correlation between successive residual terms; although some tests were non-significant or inconclusive. However, because of the way in which the data are generated and the relationship of the population in generation t to the population in generation t-1, a first order autoregressive model is assumed for all of the regression analyses.

The general model for each combination of design parameters is

$$y_{(ijk\ell m)\cdot t} = f(t) + \varepsilon_{(ijk\ell m)\cdot t} \ , \quad t=0,1,2,\ldots,T, \tag{4.7}$$

where $y_{(ijk\ell m)\cdot t}$ is the average over the four replications of the dependent variable under consideration, f(t) is a polynomial in t, $\varepsilon_{(ijk\ell m)\cdot t}$ is the residual term, T is the last cycle of selection and $(ijk\ell m)$ represents a particular combination of design parameters. The $\varepsilon_{(ijk\ell m)\cdot t}$, t=0,1,2,...,T, are assumed to have the structure

$$\varepsilon_{(ijk\ell m)\cdot t} = \rho \ \varepsilon_{(ijk\ell m)\cdot t-1} + u_{(ijk\ell m)\cdot t} \ , \tag{4.8}$$

where ρ is the autocorrelation parameter with $|\rho| < 1$, $\varepsilon_{(ijk\ell m)\cdot t}$ and $\varepsilon_{(ijk\ell m)\cdot t-1}$ are the residual terms for the observations in generation

t and t-1, respectively, and the $u_{(ijk\ell m)\cdot t}$ are independent and identically distributed as a random variable u_t with $E(u_t) = 0$ and $\sigma^2(u_t) = \sigma^2$. In this first order autoregressive process on the residuals, writing $\varepsilon_{(ijk\ell m)\cdot t}$ as ε_t, we have

$$E(\varepsilon_t) = 0$$

and (4.9)

$$\sigma^2(\varepsilon_t) = \frac{\sigma^2}{1-\rho^2} \quad .$$

Estimation of the regression parameters in the model (4.7) is accomplished after reparameterization of the model. Further details of this procedure are discussed in Neter and Wasserman (1974). The autocorrelation parameter, ρ, is estimated by

$$\hat{\rho} = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=2}^{T} e^2_{t-1}} \quad , \tag{4.10}$$

where the e_t are calculated from the ordinary least squares fit of the observations to the model in (4.7). Correlations between residuals more than one time period apart are assumed to be zero.

The polynomial function of t, f(t), depends upon the dependent variable being characterized. For percent homozygosity we assume a cubic function of t,

$$H_{(ijk\ell m)\cdot t} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \varepsilon_{(ijk\ell m)\cdot t} \quad , \quad t=1,2,\ldots,T. \tag{4.11}$$

For genotypic value, yield, and heritability in the broad sense, we assume a quadratic function of t. For genotypic value, this function is

$$G_{(ijk\ell m)\cdot t} = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_{(ijk\ell m)\cdot t} \quad , \quad t=1,2,\ldots,T. \qquad (4.12)$$

If each simulation run had continued until the population was completely homozygous, different functional forms of t might better characterize the behavior of these variables over time; however, we are concerned only with describing the behavior of the dependent variables over at most 30 cycles of selection which have been simulated.

Analysis of variance is used to determine the significance of the design parameters on these variables at specific cycles in the selection program and regression analysis is used to characterize the behavior of each variable for the duration of the selection program. The regression equations reported are those which give the best fit to the simulation output and are not necessarily those which would result if the selection program had continued for a longer period of time. An understanding of the behavior of population variables over the time span of the selection program rather than future predictions is the purpose of the regression analyses. Results of these analyses as well as comparisons of simulated and experimental observations are presented in the next chapter.

Chapter V

ANALYSES AND RESULTS OF THE SIMULATED DATA

5.1 Introduction

The output from the simulation program is used for two principle purposes. First, to compare population values from the first four cycles of selection with the observed values from the original experiment, i.e. to determine whether or not the simulated results behave in a similar manner as the actual results. Second, to evaluate the long term behavior of certain population characteristics (variables): (i) percent homozygosity (H), (ii) genotypic value (G), (iii) yield (Y), and (iv) heritability in the broad sense (h_B^2). Comparisons of observed and simulated results are discussed in Section 5.2.

In addition to the statistical analyses, observations concerning the fixation of alleles in the population at the conclusion of the selection program are of interest. The frequency of the more desirable (+) allele at each locus is printed for the entire progeny population each cycle. Using this information for cycle T, it is possible to determine which loci have become fixed for each of the two alleles and to which subgenotypes these loci belong. Fixation of alleles and the number of selection cycles necessary to achieve complete homozygosity of the population are closely related and are discussed together in Section 5.3. Results of the analyses of variance are presented in Section 5.4 and results of the regression analyses in Section 5.5. A summary of this chapter is given in Section 5.6.

5.2 Comparisons of Experimental and Simulated Data for Yield and Coefficient of Inbreeding

Results of the first four cycles of selection of the actual experiment are given by Genter (1976). For comparative purposes, all phenotypic means have been expressed as a percent of the F_1 phenotypic value and the values for the simulated data have been averaged over all 96 simulation runs. Initial comparisons were made with the simulation output averaged over the 32 runs for each gene action model separately; however, these individual averages were very close together so final comparisons are made using the overall averages.

Table 5.1 contains the observed and simulated average yields for cycles zero through four. Also the high and low yields for each cycle are included. Figure 5.1 illustrates the comparison of observed and simulated phenotypic means. In the F_2 generation, selection cycle zero, the observed average was 59.9 and the average of the simulated data was 59.0. The average low yield for the simulated data was 36.9 compared to an observed value of 31.6. The average high yield of the simulated data was 77.2 compared to an observed value of 91.6. Out of all 96 simulation runs, the maximum high yield was 95.5. The largest difference between observed and simulated values occurred in cycle one for the low yield.

Through the first four cycles of selection, the observed mean yield increased 20.5 percent of the F_1 value while the simulated results increased 18.8 percent. The observed low yield increased 26.7 percent compared to an increase of 27.9 percent for the simulated results. For the high yield, there was an observed increase of 8.2 percent and

TABLE 5.1

Comparison of Observed and Simulated Results for Four Cycles of Selection

Selection Cycle		Yield		
		Mean	High	Low
0	Obs.	59.9	91.6	31.6
	Sim.	59.0	77.2	36.9
1	Obs.	62.7	90.0	33.6
	Sim.	66.0	80.3	49.8
2	Obs.	64.4	88.0	45.4
	Sim.	70.2	83.7	55.7
3	Obs.	77.7	87.3	68.3
	Sim.	74.2	86.7	60.0
4	Obs.	80.4	99.8	58.3
	Sim.	77.8	89.6	64.8

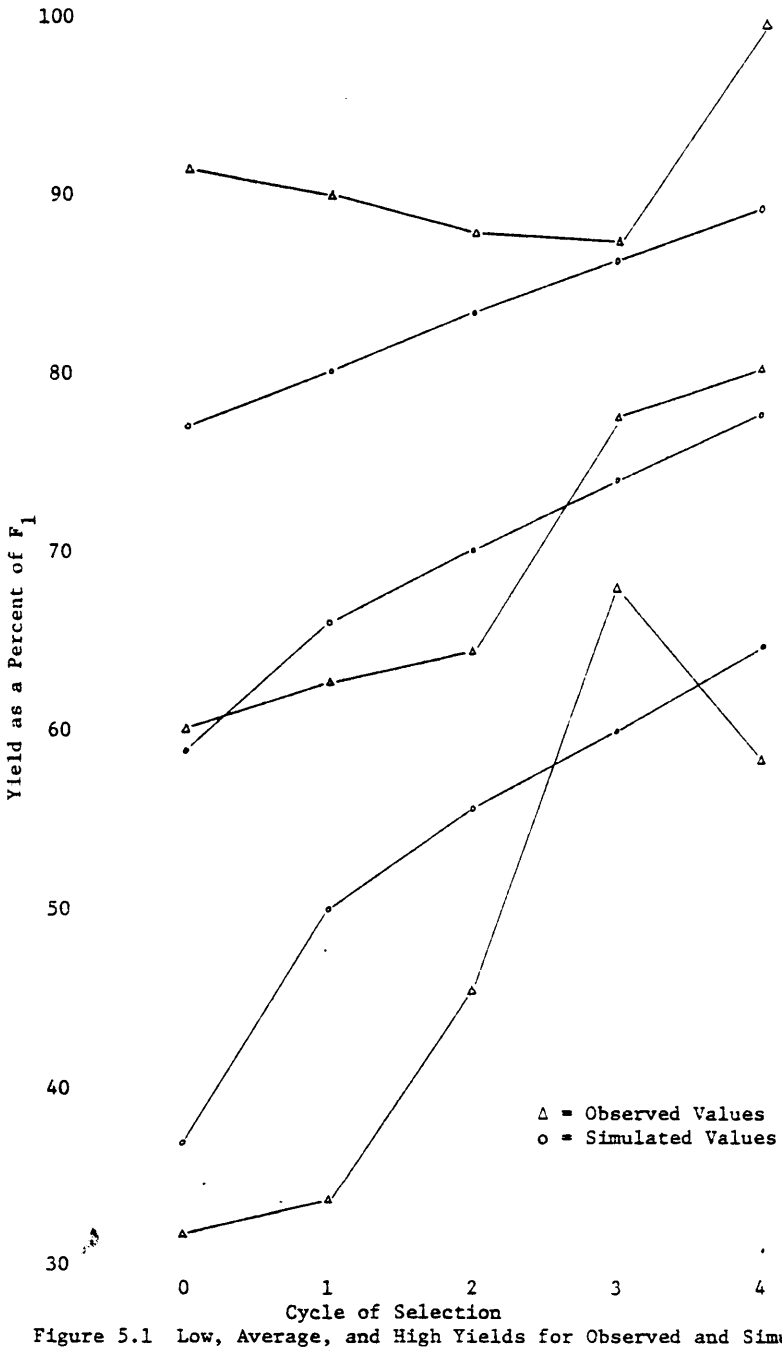


Figure 5.1 Low, Average, and High Yields for Observed and Simulated Values for Selection Cycles 0-4

a 12.4 percent increase for the simulated data.

The difference between experimental and simulated observations could not be tested statistically because the variances of individual observations in each cycle were not available for the experimental results. Although an appropriate test of hypothesis for this difference does not exist, some statistical indication of the agreement of simulated and experimental values can be provided by the construction of tolerance intervals for the observed yields.

Each of the 24 combinations of parameter values used in this simulation study represents a possible model for the recurrent selection experiment. Therefore, the 96 simulated values from the four replications of these 24 models can be considered observations from the population of models under investigation. In the construction of tolerance intervals, we will assume the mean and standard deviation of these 96 values to be the true population parameters. Each actual observation is the realization of a single experimental situation and the construction of tolerance intervals will provide an answer to the question: Do the experimental observations fall within the simulated values representing the population of models considered?

All tolerance intervals will be of the form

$$\bar{Y}_{\ldots\ldots t} \pm 1.96\ s\ ,\ t=0,\ldots,4 \qquad (5.1)$$

where $\bar{Y}_{\ldots\ldots t}$ and s are the mean and standard deviation of the 96 simulated values for each particular situation: low, average, or high yield in selection cycles 0-4. In every instance, tolerance intervals

so constructed for low, average, and high yields each cycle contain the relevant experimental value. This comparison of observed and simulated values indicates that the range of values used in modeling this recurrent selection experiment will give results comparable to the actual situation. Graphs of the tolerance intervals for low, average, and high yields in selection cycles 0-4 are given in the Appendix.

Comparisons of percent homozygosity from the simulation program with the observed value for the average coefficient of inbreeding reveal a close agreement of these two measurements on the average. No tests of significance were performed nor were tolerance intervals constructed because different techniques of measurement were used for each of the quantities. One other difficulty is that for the actual experiment, the observed coefficient of inbreeding for cycles zero through three does not include all of the selected crosses in those cycles, but only those crosses which contribute to the selected crosses of cycle four. As for yield, the simulated values used for comparison are the average of all 96 runs.

For this recurrent selection program, all individuals appearing after the F_1 generation and possessing identical alleles at a locus which was heterozygous in the F_1 have alleles that are identical by descent. In the simulation program, it is possible to observe every locus in each individual and determine whether the locus is homozygous ((+,+) or (-,-)) or heterozygous (+,-). The proportion of homozygous loci in a later generation is then a measure of the probability of alleles being identical by descent. The formula used to calculate percent homozygosity is given in (4.1) and (4.2). In the experimental

maize population, homozygosity also implies that the alleles are identical by descent, but it is not possible to observe which loci are heterozygous and which are homozygous. Under these circumstances the only quantity that can be computed is the one locus coefficient of inbreeding. This quantity is based on Malécot's coefficient of parentage using the probability that alleles are identical by descent from a distant ancestor. The coefficient of inbreeding is a theoretical value which we would expect to observe if the population were infinitely large. Thus, for a single locus and an infinitely large population, the coefficient of inbreeding is equal to the percent homozygosity in the population. Further discussion of the relationship between percent homozygosity and the coefficient of inbreeding is contained in Chapter VI.

Table 5.2 gives the observed coefficient of inbreeding and the simulated percent homozygosity for the first four cycles of selection. The minimum value for the coefficient of inbreeding for any individual in the F_2 or later generations is 0.5. Therefore, the values actually observed from the selection experiment cannot be lower than 0.5. The values for percent homozygosity are based on samples from a theoretical distribution which has an expected value of 0.5 in cycle zero. Hence, due to sampling variation, we can have simulated values which are less than the theoretical minimum, but the average over a large number of loci and progeny should be close to the theoretical expectation.

For the mean of all crosses, the greatest difference between observed and simulated values is 0.0094 and this occurs in cycle four. After cycle zero, the high and low values compare favorably for the

TABLE 5.2

Comparison of Observed Coefficient of Inbreeding and Percent Homozygosity from Simulation Study

Selection Cycle		Mean	Low	High
0	Obs.	0.5	0.5	0.5
	Sim.	0.5004	0.4789	0.5292
1	Obs.	0.5	0.5	0.5
	Sim.	0.5056	0.4906	0.5210
2	Obs.	0.5193	0.5	0.5625
	Sim.	0.5230	0.4995	0.5565
3	Obs.	0.5446	0.5	0.6094
	Sim.	0.5387	0.5019	0.5877
4	Obs.	0.5657	0.5078	0.6445
	Sim.	0.5563	0.5242	0.6359

simulated and experimental values. The largest difference (0.0164) in the high values occurred in cycle four and the largest difference in the low values occurred in cycle three (0.0217). A graph of the values in Table 5.2 is given in Figure 5.2.

While one must be careful in the interpretation of these comparisons, we can see that over the five selection cycles compared these two measures are in close agreement on the average. The low and high values differ more because of sampling variation in the simulated values and the fact that the observations from the actual experiment are single observations from one repetition of an experiment.

5.3 Homozygosity, Fixation, and Maximum Yield

Two characteristics which are closely related and of interest in this investigation are: (i) the length of time or number of selection cycles necessary for a population to reach complete homozygosity and (ii) the fixation of advantageous alleles in the population. If the predominant mode of gene action is linear with partial to complete dominance and homozygosity of the superior allele is desirable, then the desired end result of a selection program for a specific trait in a particular population is complete homozygosity for the most desirable allele at all loci contributing to that trait. In order to achieve this result, we must have sufficient selection pressure for progress to occur without causing the fixation of a less desirable (-) allele. If selection intensity is too high some of the - alleles which are linked to + alleles on the same chromosome will become fixed and the resulting homozygous population will not be fixed at the maximum

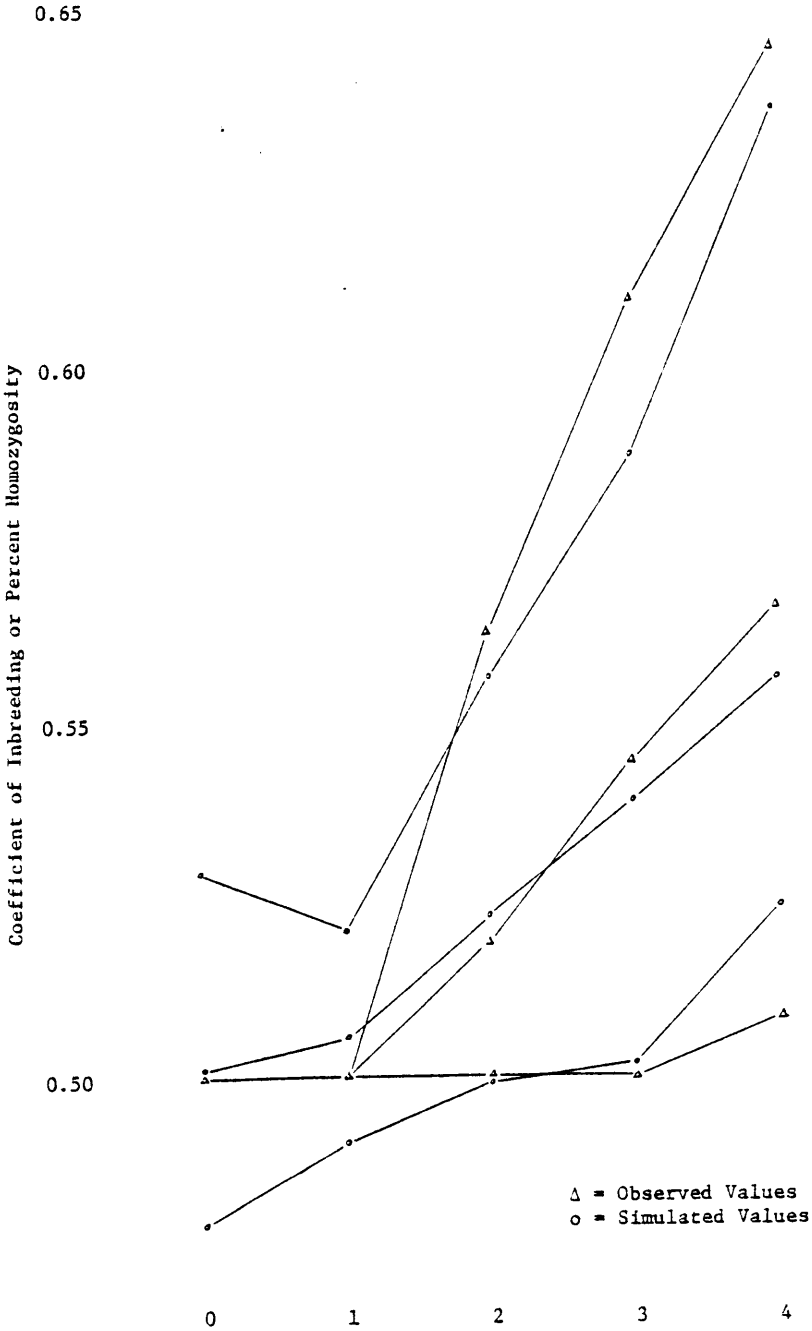


Figure 5.2 Observed Coefficient of Inbreeding and Simulated Percent Homozygosity for Selection Cycles 0-4

genotypic value. Therefore, recombination of alleles at different loci linked on the same chromosome must be sufficiently high to reduce the probability of the fixation of a - allele.

The final selection cycle in the simulation program will be designated as cycle T. For those simulation runs in which the population reached complete homozygosity before 30 cycles of selection, T is less than 30; otherwise, T is equal to 30. Tables 5.3 and 5.4 give the number of cycles of selection (T), the average genotypic value for the last cycle of selection (G_T), and the maximum genotypic value (G_{MAX}) for each replication of the 24 combinations of design parameters. The maximum genotypic value for any run is that value computed from the gene action model assuming all loci that are heterozygous in the F_1 have become fixed for the + allele. Replications of the runs with n_h=50 have different maxima because of the procedure which controls the homozygous loci in the F_1 and the assignment of loci to subgenotypes.

In Table 5.3, the results are given for all simulation runs for which $\sigma_E^2 = 0$. This represents an ideal situation in which the genotypic values and phenotypic values or yields are identical for all individuals and can be used as a reference for the interpretation of the effect of environmental variance. Under these circumstances exact evaluation of the genotype is possible and selection of superior genotypes is more effective. For the simple linear and multiple linear gene action models complete homozygosity of the population was achieved in less than 30 cycles of selection when r = 0.2 or n_h = 50. Only one replication of the multiple linear model with N = 20, n_h = 50, and $\sigma_E^2 = 0$ did not reach the genotypic maximum with complete homozygosity. All other runs which

TABLE, 5.3

T, G_T, and G_{MAX} for All Replications with $\sigma_E^2=0$

N	r	n_h	Rep	Model: Simple Linear			Model: Multiple Linear			Model: Epistatic[1]		
				T	G_T	G_{MAX}	T	G_T	G_{MAX}	T	G_T	G_{MAX}
10	0.2	50	1	24	66	66	24	94	94	30	71	71
			2	23	70	70	25	103	103	30	76.95	77
			3	24	68	68	25	108	108	30	80	80
			4	24	86	86	24	84	84	30	83	83
	0.05	100	1	30	92.93	100	30	115.42	125	30	105.51	108
			2	30	92.14	100	30	112.20	125	30	103.60	108
			3	30	93.74	100	30	116.49	125	30	106.36	108
			4	30	92.93	100	30	110.18	125	30	102.79	108
20	0.2	100	1	25	100	100	27	125	125	30	107.97	108
			2	26	100	100	28	125	125	30	107.90	108
			3	27	100	100	26	125	125	30	107.79	108
			4	26	100	100	28	125	125	30	107.95	108
	0.05	50	1	22	80	80	25	100	100	30	78.85	79
			2	22	80	80	21	100	101	30	66	66
			3	21	84	84	20	96	96	30	75	75
			4	20	80	80	23	86	86	30	86	86

[1]The maximum genotypic value can be achieved without complete homozygosity.

TABLE 5.4

T. G_T, and G_{MAX} for All Replications with σ_E^2=60

N	r	n_h	Rep	Model: Simple Linear			Multiple Linear			Epistatic		
				T	G_T	G_{MAX}	T	G_T	G_{MAX}	T	G_T	G_{MAX}
20	0.2	100	1	30	92.91	100	30	115.56	125	30	101.20	108
			2	30	93.58	100	30	116.27	125	30	102.66	108
			3	30	93.68	100	30	120.08	125	30	103.61	108
			4	30	97.13	100	30	118.37	125	30	99.63	108
	0.05	50	1	30	73.78	76	30	88.76	92	30	90.75	91
			2	30	74.46	78	30	85.4	88	30	93.54	97
			3	30	80.01	32	30	97.59	101	30	65.42	69
20	0.2	50	1	30	69.69	70	30	89.75	90	30	76.67	77
			2	30	83.73	84	30	101.52	102	30	82.59	83
			3	30	75.66	76	30	87.51	88	30	66.67	67
			4	30	73.69	74	30	80.73	81	30	79.86	80
	0.05	100	1	30	97.65	100	30	121.81	125	30	104.14	108
			2	30	96.83	100	30	119.10	125	30	105.88	108
			3	30	96.31	100	30	121.95	125	30	104.31	108
			4	30	96.23	100	30	120.68	125	30	104.72	108

reached complete homozygosity did so for the maximum genotypic value. For both linear models, the runs with N = 20, r = 0.05, n_h = 50, and σ_E^2 = 0, reached complete homozygosity more rapidly on the average than the other situations. One of these runs achieved the maximum genotypic value for the entire population after 20 cycles of selection.

All of the runs with N = 10, r = 0.05, and n_h = 100 failed to reach complete homozygosity or the maximum genotypic value by 30 cycles of selection for all three gene action models. The combination of low selection intensity, tight linkage, and more heterozygous loci subject to selection pressure results in a slower approach to complete homozygosity and the possibility that even when the population does become completely homozygous some of the loci will be fixed for the less desirable allele and consequently the maximum genotypic value will not be reached.

For the epistatic gene action model, none of the runs reached complete homozygosity in 30 cycles of selection. Six of the eight runs with n_h = 50 did reach the maximum genotypic value, but the populations were not completely homozygous. This situation occurs when the major gene in the subgenotype of the epistatic model which incorporates a major gene with modifiers becomes homozygous for the + allele. The remaining loci in the major gene subgenotype are no longer contributing to the genotypic value and consequently the allelic frequencies are fluctuating randomly.

The results for runs with σ_E^2 = 60 are given in Table 5.4. None of these runs reached complete homozygosity or the maximum genotypic value in 30 cycles of selection. Values for percent homozygosity for

these runs after 30 cycles of selection are given in Table 5.5. When EV is greater than zero and selection is based on phenotypic value, it is possible that the selected crosses do not have the most superior genotypes. For each of the three gene action models the runs with $N = 20$, $r = 0.2$, and $n_h = 50$ came the closest to reaching the maximum genotypic value and the highest percent homozygosity on the average.

The fixation of advantageous alleles in a population has been investigated from a theoretical approach (Kimura, 1957) as well as with the use of Monte Carlo techniques (Qureshi and Kempthorne, 1968). Theoretical investigations have been for situations with only one locus in which one allele has a slight selective advantage over the others. Even for this simple situation, exact analytical treatment has proved too complex and approximations are used. With the gene action models under consideration for this recurrent selection program, the selective advantage of the more favorable allele is considerably greater than can be accommodated by existing approximations. In addition, we are considering a genotype with a total of 100 loci and not just one locus. Therefore, simulation of this complex model will provide information on the fixation of advantageous alleles for a situation which could not easily be investigated theoretically.

Tables 5.6 and 5.7 give the results of the fixation of alleles for simulation runs with $n_h = 50$ and $n_h = 100$, respectively. Included in these tables are the number of loci which became fixed for the + allele, the number of loci which became fixed for the - allele, and the value of the percent homozygosity for the final cycle of selection (H_T^A). In all runs which reached complete homozygosity before T = 30, except one

TABLE 5.5

Percent Homozygosity when T=30 for All Runs with σ_E^2=60

N	r	n_h	Rep	Model: Simple Linear	Model: Multiple Linear	Model: Epistatic
10	0.2	100	1	0.8684	0.8946	0.8473
			2	0.878	0.8826	0.8740
			3	0.8692	0.9344	0.8902
			4	0.9311	0.8764	0.8342
	0.05	50	1	0.9264	0.9020	0.9862
			2	0.8674	0.8938	0.8860
			3	0.9102	0.8718	0.9697
			4	0.9024	0.9268	0.8837
20	0.2	50	1	0.9806	0.9835	0.9662
			2	0.9824	0.9704	0.9370
			3	0.9791	0.9690	0.9492
			4	0.9833	0.9803	0.9545
	0.05	100	1	0.9405	0.9354	0.9071
			2	0.9305	0.9078	0.9274
			3	0.9099	0.9273	0.8916
			4	0.9035	0.9063	0.9057

TABLE 5.6

Number of Loci Fixed for the + Allele and the - Allele and Percent Homozygosity when n_h=50

N	r	σ_E^2	Rep	Model: Simple Linear +	Simple Linear -	Simple Linear H_T^A	Multiple Linear +	Multiple Linear -	Multiple Linear H_T^A	Epistatic +	Epistatic -	Epistatic H_T^A
10	0.2	0	1	50		1.0	50		1.0	43	2	0.96
			2	50		1.0	50		1.0	39		0.93
			3	50		1.0	50		1.0	43	1	0.96
			4	50		1.0	50		1.0	41	3	0.97
	0.05	60	1	37		0.9264	31	1	0.9020	44	1	0.9862
			2	25		0.8674	30		0.8938	31	1	0.8860
			3	28		0.9102	25		0.8718	24	1	0.8687
			4	23		0.9024	32	2	0.9268	27	1	0.8827
20	0.2	60	1	42		0.9806	41		0.9825	40		0.9662
			2	40		0.9824	40		0.9704	31	1	0.9370
			3	39		0.9791	38		0.9690	36		0.9492
			4	45		0.9833	37		0.9803	40		0.9545
	0.05	0	1	50		1.0	50		1.0	44	3	0.98
			2	50		1.0	49	1	1.0	48		0.99
			3	50		1.0	50		1.0	47		0.98
			4	50		1.0	50		1.0	43	3	0.98

TABLE 5.7

Number of Loci Fixed for the + Allele and the - Allele and Percent Homozygosity when n_h=100

N	r	σ_E^2	Rep	Model: Simple Linear			Model: Multiple Linear			Model: Epistatic		
				+	-	H_T^A	+	-	H_T^A	+	-	H_T^A
10	0.2	60	1	48		0.8684	59	2	0.8946	40		0.8473
			2	52		0.878	57	1	0.8826	46		0.8740
			3	50		0.8692	65	2	0.9344	54	1	0.8902
			4	64		0.9311	52		0.8764	42	3	0.8342
	0.05	0	1	57		0.88	47	1	0.85	65	2	0.93
			2	51		0.89	45	2	0.85	63	3	0.92
			3	48		0.87	53		0.87	62	4	0.94
			4	51		0.88	41	2	0.84	56	1	0.90
20	0.2	0	1	100		1.0	100		1.0	88		0.96
			2	100		1.0	100		1.0	86	1	0.96
			3	100		1.0	100		1.0	85	1	0.95
			4	100		1.0	100		1.0	86	1	0.97
	0.05	60	1	59		0.9405	58		0.9354	48	2	0.9071
			2	53		0.9305	57		0.9078	53	1	0.9274
			3	42		0.9099	57		0.9273	46	2	0.8916
			4	42		0.9085	46		0.9065	49	1	0.9057

replication of N = 20, r = 0.05, n_h = 50, σ_E^2 = 0 with the multiple linear model, the + allele was fixed for each locus that was originally heterozygous in the F_1. As previously indicated, these runs also reached the maximum genotypic value.

For the one run which reached complete homozygosity without achieving the maximum genotypic value, it is of interest to determine in which of the subgenotypes a locus became fixed for the - allele. The multiple linear model contains four linear subgenotypes which will be designated L1, L2, L3, and L4, according to the notation in Table 2.2. Output from the simulation program indicates that one locus in subgenotype L4 became homozygous for the - allele. The loci in the L4 subgenotype contribute the smallest value to the genotypic value of each individual.

For the simple linear gene action model, none of the loci became fixed for the - allele. The quantity H_T^A indicates how close to complete homozygosity the population is. For example, in the case of N = 10, n_h = 100, and σ_E^2 = 0, the average percent homozygosity was 0.88 after 30 cycles of selection and approximately 50 percent of the loci were fixed for the + allele. Although the remaining loci were still heterozygous in some individuals in the population, the frequency of either the + or - allele for the majority of these heterozygous loci approaches 1.0. Thus, the value of H_T^A is relatively high. With σ_E^2 = 60, between 10 and 58 percent of the loci which were heterozygous in the F_1 were still heterozygous after 30 cycles of selection.

The results for the multiple linear model are very similar to the simple linear model with respect to the fixation of the + allele. Just

as in the simple linear model, the same three situations with $\sigma_E^2 = 0$ resulted in all loci fixed for the more desirable allele with the previously mentioned exception. In 8 of the remaining 31 simulation runs with the multiple linear model, the less desirable allele became fixed at one or two of the loci. For each of these 8 runs, N = 10 and the loci involved were in the subgenotypes with the smallest contribution to the genotypic value, subgenotypes L3 and L4. Those runs with N = 10, $n_h = 100$, and $\sigma_E^2 = 0$ failed to reach complete homozygosity by T = 30. Slightly more than 50 percent of the loci were still heterozygous and the percent homozygosity had reached an average of 0.85. Between 18 and 54 percent of the loci were still heterozygous for the runs with $\sigma_E^2 = 60$.

All runs with the epistatic model failed to reach complete fixation of the genotype. On the average a greater proportion of the loci which were heterozygous in the F_1 became fixed for those situations with $\sigma_E^2 = 0$. Fixation of the - allele for one or more loci occurred in at least one replication of each situation with the epistatic gene action model. Fixation of the - allele occurred more often for N = 10 than for N = 20, and for r = 0.05 than for r = 0.2. A total of 40 loci in all runs combined were fixed for the - allele and only 7, or 17.5%, of these loci were not in an epistatic subgenotype. These loci were assigned to L4, the linear subgenotype with the smallest contribution to the genotypic value. Four loci in the threshold portion of the bottleneck subgenotype and 29 loci in the modifier portion of the subgenotype which incorporates a major gene with modifiers were fixed for the - allele. These subgenotypes are explained in detail in Chapter II.

In each of the epistatic subgenotypes, it is possible to have a situation in which a constant value is contributed to the genotypic value and selection has no effect on this contribution. As an example, consider the bottleneck subgenotype. The threshold portion of this subgenotype has three loci and the linear portion has 7 loci. Since the threshold value, τ, has been set at two, the 7 loci in the linear portion of this subgenotype cannot contribute to the genotypic value if more than one of the threshold loci is homozygous for the - allele. If the population becomes homozygous for the - allele for more than one of the threshold loci, then selection no longer has any effect on the allelic frequencies at the loci in the bottleneck subgenotype. For those loci which are heterozygous, the allelic frequencies will fluctuate at random and prevent the population from becoming completely homozygous.

A similar situation occurs in the subgenotype with the major gene. Once the major gene becomes homozygous throughout the population for the + allele, the 9 loci which are modifiers of the major gene will no longer contribute to the genotypic value of any individual in the population. Thus these loci are not affected by selection and the allelic frequencies vary randomly. Investigation of the simulation output revealed that for this subgenotype the major gene became fixed for the + allele after the first few cycles of selection for those runs in which the major gene was heterozygous in the F_1. Once this occurred it was still possible for the population to reach the maximum genotypic value without being completely homozygous.

The 40 loci for which the - allele became fixed represent only 1.67% of the total number of loci which were heterozygous in the F_1 in

the 32 runs with the epistatic gene action model. The contribution for most of these loci to the genotypic was small and the effect of losing the + allele is minimal. A total of 1590 or 66.25% of the loci were fixed for the + allele and approximately 32% of the loci were still heterozygous. Thus the majority of the heterozygous loci which have the greatest contribution to an individual's genotypic value are either fixed for the + allele or are still heterozygous and the + allele has not been lost.

Investigation of the fixation of the more desirable allele in this selection program using an epistatic gene action model has shown that the epistatic subgenotypes tended to become fixed after a few cycles of selection. Once the epistatic contribution is fixed, only the linear subgenotypes can be changed with selection. While it does not seem unreasonable that gene action is truly epistatic, after several cycles of selection a situation can exist in which all heterozygous loci subject to selection pressure behave according to a linear model regardless of the epistatic nature of the original model. Perhaps this is part of the reason that epistasis is difficult to detect in selection programs and is therefore not considered to be an important factor in the gene action of many quantitative traits.

5.4 Results of the Analyses of Variance: Effects of Design Parameters

Results of the analyses of variance described in Chapter IV are presented in Tables 5.8 to 5.18. Each table contains the mean squares for each effect and interaction in the model for selection cycles 5, 10, 15, 20, and T, the last cycle of selection. The last cycle of selection

is the 30th cycle or the cycle in which the population became completely homozygous, whichever occurred first. Values of T can be found in Tables 5.3 and 5.4 for each simulation run.

Mean squares for percent homozygosity in *selected* crosses are presented in Tables 5.8, 5.9, and 5.10. Table 5.8 gives the mean squares for the analyses with all of the factors included. When all of the runs are analyzed together, the significance of some of the interactions involving gene action model may be masked since it is evident from Tables 5.3 and 5.4 that percent homozygosity behaves differently for the two values of EV. Each main effect except GAM became increasingly more significant as the selection procedure progressed through cycle 20. The effect of RV is not significant in cycle 5, but is highly significant in cycle 10. No interactions are significant in cycles 5 and 10, but SI × RV and RV × GAM are significant in cycles 15 and 20. The interaction of RV × NHL is significant only in cycle 20.

Since the last cycle is not the same for all runs we must be careful in comparing these results with the results from the analyses of cycles 5, 10, 15, and 20. Those runs for which the last cycle of selection was less than 30 all reached complete homozygosity of the population in the last cycle and all have $\sigma_E^2 = 0$. In the last cycle analysis, all main effects are highly significant as well as the 2-factor interactions SI × NHL and RV × GAM.

Tables 5.9 and 5.10 give the mean squares for the analyses of percent homozygosity in selected crosses for each value of EV. Similar trends are evident in both tables for the 3 main effects of SI, RV, and NHL. However, there is a noticeable difference in the significance of

GAM and the 2-factor interactions for the different values of EV. None of the effects are significant when $\sigma_E^2 = 60$, but all are highly significant in the last cycle analysis for runs with $\sigma_E^2 = 0$. This is not unexpected because as we saw in Tables 5.3 to 5.7, percent homozygosity has a similar behavior for each gene action model when $\sigma_E^2 = 60$; however, when $\sigma_E^2 = 0$ the behavior of percent homozygosity is different for different types of gene action.

The analyses for percent homozygosity of *all* crosses is presented in Tables 5.11 to 5.13. The significance of the main effects is the same as in the analysis of percent homozygosity of selected crosses with the exception of recombination value in cycles 5 and 10. The interaction RV × GAM is significant from cycle 15 to the last cycle. In the last cycle analysis, the main effect GAM and the 2-factor interactions SI × RV and SI × NHL are significant in addition to the ones previously mentioned.

A comparison of the analyses given in Tables 5.12 and 5.13 gives similar results as the comparison of Tables 5.9 and 5.10. Just as for percent homozygosity in selected crosses, separate analyses of percent homozygosity in all crosses for the 2 values of EV indicated that GAM or any 2-factor interactions are significant only when $\sigma_E^2 = 0$.

The analyses of genotypic value are presented in Tables 5.14 and 5.15. All main effects are highly significant for the analyses of genotypic value of selected crosses and all crosses. In order to combine the results of every simulation run, genotypic value is expressed as a percent of the maximum genotypic value, G_{MAX}, in each run. It may seem inconsistent that environmental variance is a significant factor

in the analysis of a variable which does not depend on environmental variance; however, the population each cycle depends upon phenotypic values or yields from the previous cycle and thus the value of EV does affect the genotypic value at different cycles in the selection program.

Analyses of genotypic value of selected crosses are presented in Table 5.14. In addition to the highly significant main effects, the interaction SI × NHL is highly significant throughout the selection process. As the selection program progresses, more of the 2-factor interactions exhibit significance. In cycle 20 all interactions except EV × GAM are significant. Three interactions in the last cycle are significant: SI × NHL, SI × GAM, and RV × GAM.

Results of the analyses of genotypic value for all crosses, shown in Table 5.15, are similar to those for the selected crosses. Again, the interaction SI × NHL is significant throughout the selection program.

In the analyses of variance of yield or phenotypic value, yield is expressed as a percent of the maximum genotypic value in each run. This may cause the values to exceed 100% since an environmental deviation is added to the genotypic value to obtain the value for yield. All main effects are highly significant in every cycle analyzed for yield of selected crosses (Table 5.16) and yield of all crosses (Table 5.17) except EV for cycle 5 in the selected crosses. Interaction SI × NHL is significant for both variables in every cycle analyzed. In the early stages of the selection program, only one or two of the interactions involving selection intensity are significant; however, in cycle 20 more of the interactions, particularly those involving gene

action model, have a significant effect on the yield.

Heritability in the broad sense was analyzed for those runs with $\sigma_E^2 = 60$. Table 5.18 gives the mean squares for these analyses. All main effects are highly significant for cycles 5 and 10. Only SI and NHL are significant in the later cycles. RV has no effect on heritability in the broad sense from cycle 15 to the end of the selection program and GAM is not significant from cycle 20 to the end of the selection program. Several interactions are significant in the early cycles of the selection program, but none are significant in cycle 20 or cycle T.

As the selection program continues, the individuals produced each generation become more similar to each other genotypically and thus, σ_G^2 decreases over time. Since $h_B^2 = \frac{\sigma_G^2}{\sigma_Y^2}$, where $\sigma_Y^2 = \sigma_G^2 + \sigma_E^2$ and σ_E^2 is a constant for the simulation runs, the behavior of σ_G^2 determines the behavior of h_B^2. When the population becomes completely homozygous for all loci, $\sigma_G^2 = 0$ and consequently $h_B^2 = 0$. The genotypic variance in a particular population is a function of the gene action model, the number of loci, and the recombination value. In the early cycles of selection, σ_G^2 is larger and the effect of the design parameters on σ_G^2 greater than in the later cycles. Therefore, as the genotypic variance decreases and the individuals in the population become similar genotypically, the effect of the factors on σ_G^2 and h_B^2 decreases.

The analyses of variance indicate that percent homozygosity for selected or all crosses is very much affected by the different values used for the design parameters. Examination of the means for the relevant cycles of selection shows that in general there is a greater

increase in homozygosity under more intense selection (N=20), a higher recombination value (r=0.2), fewer heterozygous loci (n_h=50), and no environmental error (σ^2_E=0). These results are not unexpected and are intuitively reasonable for the assumed biological model. With the higher selection intensity, the proportion of the population selected each cycle is 0.11. Since this proportion consists of individuals which are genotypically superior, as a consequence the frequency of the + allele at each locus is higher for these individuals than for the proportion of the population which is selected under the lower selection intensity (N = 10, i.e. a selected proportion of 0.22). As the frequency of the + allele for each locus increases with selection, the amount of homozygosity also increases. Therefore, as more intense selection causes a more rapid increase in the frequency of + alleles, percent homozygosity also increases more rapidly.

Differences in recombination value are significant after the selection program has been in progress for a number of cycles. With a larger probability for the recombination of alleles between adjacent loci, the linkage of + and - alleles at different loci on the same chromosome will be replaced with all + alleles linked together more quickly than with a smaller recombination value. Thus the amount of homozygosity increases with this increase in frequency of + alleles at more loci.

Exact identification of the genotype is possible with $\sigma^2_E = 0$ and allows for a higher rate of increase in the frequency of the + alleles thereby increasing the percent homozygosity more rapidly than in a situation with $\sigma^2_E = 60$. Individuals selected when there is a positive

environmental variance may be superior because of a positive environmental contribution rather than an increase in + alleles.

Finally, the type of gene action is significant only after a large number of cycles of selection. The most likely explanation for this result has already been discussed in section 5.3 under fixation of alleles. The allelic frequencies of several loci in the epistatic subgenotypes are fluctuating at random and the contribution from these loci to the amount of homozygosity is not increasing as a result of selection.

Interpretation of the analysis of variance results for percent homozygosity has been in large part in terms of the increase in frequency of the more desirable allele at each locus in the genotype which is heterozygous and therefore subject to selection pressure. While it is possible for percent homozygosity to increase with the decrease in frequency of + alleles, the selection of individuals is based on high genotypic value which results from an increase in the number of + alleles.

The model assumed for these analyses gave the best fit in cycle 20 for all but one of the analyses. For each analysis of percent homozygosity the error mean square increased through cycle 15 and declined in cycles 20 and cycle T. The variation in the values of error mean square does not appear to be related to the ability of the model to explain the data. No interaction was significant for all cycles of the selection program analyzed. Those interactions which were significant in one or more of the cycles analyzed do not affect the significance or interpretation of the main effects. The interaction of recombination

value by gene action model, which is significant in cycles 15, 20, and cycle T, indicates that the difference in percent homozygosity for the 2 values of r depends on which gene action model is used. For the two linear gene action models (M 1 and M 2) this difference is larger than for the epistatic gene action model (M 3).

The analyses of genotypic value and yield give similar results. The total genotypic or phenotypic variation in each cycle decreases as the number of selection cycles increases. This is reflected in Tables 5.14 to 5.17 by the decrease in the error mean square as the selection program progresses. However, the decrease in the error mean square is not a reflection of how well the model fits the data. In fact, the poorest fit occurs in selection cycle T with the smallest mean square error. This situation is probably caused by the fact that the last cycle of selection is not the same for all simulation runs.

Both genotypic value and yield increase more rapidly with more intense selection (N=20), larger recombination value (r=0.02), fewer heterozygous loci (n_h=50), no environmental error (σ^2_E=0), and linear gene action models (M 1 and M 2). As with the increase in percent homozygosity, the increase in genotypic value and yield are related to the increase in the frequency of + alleles at all loci which were heterozygous in the F_1. As the number of + alleles increases, genotypic value and yield also increase. In addition to selection intensity, recombination value, number of heterozygous loci and environmental error, gene action model is highly significant throughout the selection program. In the early cycles of selection the genotypic value or yield is smaller for the simple linear model than for the multiple linear or epistatic

gene action models. As the selection program progresses, the values for the two linear models become closer to each other, but values for the epistatic gene action model are higher when averaged over all appropriate simulation runs.

An investigation of the significant interactions in cycles 15 and 20 indicates that some combinations of parameter values contribute to a larger increase in genotypic value and yield. For example, the interaction of selection intensity by number of heterozygous loci is significant for all cycles analyzed. The combination of high selection intensity (N=20) and smaller number of heterozygous loci (n_h=50) results in a more rapid increase in genotypic value and yield than any other combination of N with n_h. From an investigation of the means of these variables for different combinations of the design parameters, we find that the type of interaction which occurs is one in which the factors enhance each other. The presence of significant interaction terms in the model does not mean that we should ignore the significance of the individual effects of the design parameters. Even with the significant interactions taken into account, the main effects are still significant.

The number of heterozygous loci in the F_1 has had a very significant effect on all of the variables throughout the selection program. When all 100 loci in the genotype are heterozygous, the average genotypic value of the F_2 population, expressed as a percent of G_{MAX}, is approximately 20 percentage points lower than when 50 loci are heterozygous in the F_1. In the former situation, none of the loci contributing to yield have common alleles in the two original inbred lines;

therefore, for gene action models with partial dominance, the average genotypic value of the F_2 population will be about 50% of the F_1 hybrid. If 50 loci are heterozygous in the F_1, then one-half of the loci contributing to yield have the same allele in the two inbred lines and the contribution to the value of yield from these homozygous loci is a constant throughout the entire selection process. Thus, the average genotypic value in the F_2 population is a higher percentage of the F_1 genotypic value for the same gene action models. The mechanisms controlling yield in corn are not as simple as proposed in the study; but observing the effect on genotypic value, yield, percent homozygosity, and heritability in the broad sense of all loci being heterozygous and only one-half of the loci being heterozygous can give an indication of the relative importance of knowing what portion of the total number of loci contributing to yield will be subject to selection pressure.

The only one of the five design parameters that an experimenter can control is that of selection intensity. Environmental error can be controlled to some extent, but not eliminated completely. The more accurate an investigator can make the evaluations of superior genotypes, the more rapid the increase in yield will be for each cycle of selection. Based on the results of this simulation study, the selection of N = 20 crosses in each cycle of selection will increase the response to selection without resulting in a large increase in the number of loci fixed for the - allele.

TABLE 5.8

Mean Squares for ANOVA of Percent Homozygosity in Selected Crosses for Cycles 5, 10, 15, 20 and T

Source of Variation	d.f.	Cycle 5	Cycle 10	Cycle 15	Cycle 20	Cycle T
Reps	3	0.12	0.23	0.35	0.75	0.66
SI	1	8.42**	62.50**	136.61**	185.60**	66.24**
RV	1	0.78	14.88**	33.09**	60.09**	20.88**
NHL	1	18.08**	102.17**	211.88**	239.21**	44.71**
EV	1	2.76*	8.72**	57.64**	148.53**	43.26**
GAM	2	0.80	0.46	0.28	0.80	2.52**
SI × RV	1	0.20	0.44	4.96*	17.44**	1.31
RV × NHL	1	0.36	0.55	0.12	2.92*	0.00
SI × NHL	1	0.11	1.68	0.74	2.41	5.13**
SI × GAM	2	0.98	0.86	1.10	1.02	1.06
RV × GAM	2	0.29	0.82	3.88*	3.24*	4.17**
NHL × GAM	2	0.73	0.49	0.49	0.77	0.66
EV × GAM	2	0.08	0.23	1.98	1.15	0.13
Error	75	0.55	0.65	0.94	0.69	0.44

* $p<0.05$

** $p<0.01$

TABLE 5.9

Mean Squares for ANOVA of Percent Homozygosity in Selected Crosses for Cycles 5, 10, 15, 20, and T when $\sigma_E^2=0$

Source of Variation	d.f.	Cycle 5	Cycle 10	Cycle 15	Cycle 20	Cycle T
Reps	3	0.33	0.38	0.27	0.01	0.09
SI	1	2.64*	25.64**	64.36**	117.52**	32.93**
RV	1	0.15	3.28*	11.97**	43.28**	23.35**
NHL	1	11.06**	58.03**	140.84**	192.93**	30.66**
GAM	2	0.26	0.42	1.88	1.41*	1.00**
SI × GAM	2	0.99	0.40	0.21	0.44	1.76**
RV × GAM	2	0.49	1.80	4.10*	2.49**	4.89**
NHL × GAM	2	0.21	0.69	0.99	2.28**	2.32**
Error	33	0.51	0.60	0.89	0.27	0.09

* $p<0.05$

** $p<0.01$

TABLE 5.10

Mean Squares for ANOVA of Percent Homozygosity in Selected Crosses for Cycles 5, 10, 15, 20, and T when $\sigma_E^2=60$

Source of Variation	d.f.	Cycle 5	Cycle 10	Cycle 15	Cycle 20	Cycle T
Reps	3	0.86	0.55	1.49	1.57	0.90
SI	1	6.14**	37.41**	72.36**	70.99**	33.31**
RV	1	0.74	13.27**	21.86**	19.21**	2.66*
NHL	1	7.23**	44.58**	75.99**	63.73**	15.35**
GAM	2	0.63	0.27	0.38	0.54	1.65
SI × GAM	2	0.17	0.49	1.25	0.61	0.13
RV × GAM	2	0.58	0.13	0.97	1.12	0.46
NHL × GAM	2	0.73	0.15	0.02	0.63	0.23
Error	33	0.56	0.72	0.99	1.06	0.64

* $p<0.05$

** $p<0.01$

TABLE 5.11

Mean Squares for ANOVA of Percent Homozygosity in All Crosses for Cycles 5, 10, 15, 20, and T

Source of Variation	d.f.	Cycle 5	Cycle 10	Cycle 15	Cycle 20	Cycle T
Reps	3	0.03	0.06	0.36	0.73	0.44
SI	1	7.88**	41.06**	110.29**	158.87**	67.83**
RV	1	1.87*	3.11*	19.71**	42.53**	24.01**
NHL	1	16.14**	83.00**	169.72**	220.94**	49.61**
EV	1	1.67*	6.07**	36.74**	119.65**	49.11**
GAM	2	0.41	0.70	0.16	0.45	2.28**
SI × RV	1	0.44	0.31	2.47	18.00**	1.81*
RV × NHL	1	0.01	0.52	0.26	2.29	0.10
SI × NHL	1	0.02	0.67	0.85	0.66	5.60**
SI × GAM	2	0.69	0.65	1.09	1.29	1.13
RV × GAM	2	0.38	0.81	2.96*	3.04*	4.82**
NHL × GAM	2	0.26	0.56	0.56	0.32	0.44
EV × GAM	2	0.06	0.09	0.80	0.85	0.13
Error	75	0.35	0.54	0.70	0.69	0.39

* $p<0.05$

** $p<0.01$

TABLE 5.12

Mean Squares for ANOVA of Percent Homozygosity in All Crosses for Cycles 5, 10, 15, 20, and T when $\sigma_E^2=0$

Source of Variation	d.f.	Cycle 5	Cycle 10	Cycle 15	Cycle 20	Cycle T
Reps	3	0.27	0.31	0.28	0.04	0.05
SI	1	3.72**	16.16**	49.94**	99.65**	36.53**
RV	1	1.16	0.45	6.18**	26.88**	26.40**
NHL	1	10.96**	46.69**	106.55**	182.55**	35.20**
GAM	2	0.11	0.19	0.66	0.85	1.04**
SI × GAM	2	0.76	0.24	0.23	0.38	1.68**
RV × GAM	2	0.21	0.74	2.84*	2.23**	4.69**
NHL × GAM	2	0.01	0.45	0.74	1.38*	1.96**
Error	33	0.35	0.53	0.69	0.40	0.08

* $p<0.05$

** $p<0.01$

TABLE 5.13

Mean Squares for ANOVA of Percent Homozygosity in All Crosses for Cycles 5, 10, 15, 20, and T when $\sigma_E^2=60$

Source of Variation	d.f.	Cycle 5	Cycle 10	Cycle 15	Cycle 20	Cycle T
Reps	3	0.51	0.37	1.06	1.82	0.65
SI	1	4.16**	25.43**	60.61**	61.50**	31.40**
RV	1	0.74	3.33*	14.39**	16.31**	3.21*
NHL	1	5.62**	36.62**	65.64**	56.40**	16.23**
GAM	2	0.36	0.59	0.30	0.44	1.37
SI × GAM	2	0.09	0.49	1.11	1.01	0.14
RV × GAM	2	0.72	0.35	0.96	1.13	0.88
NHL × GAM	2	0.37	0.51	0.32	0.39	0.23
Error	33	0.32	0.60	0.71	0.95	0.59

* $p<0.05$

** $p<0.01$

TABLE 5.14

Mean Squares for ANOVA of Genotypic Value Expressed as a Percent of G_{MAX} for Cycles 5, 10, 15, 20, and T for Selected Crosses

Source of Variation	d.f.	Cycle 5	Cycle 10	Cycle 15	Cycle 20	Cycle T
Reps	3	0.22	0.57	1.66	1.17	1.41
SI	1	900.45**	1307.99**	1209.57**	787.53**	145.12**
RV	1	329.54**	655.77**	551.85**	358.43**	42.52**
NHL	1	7051.67**	4587.60**	2573.96**	1204.44**	120.83**
EV	1	111.20**	244.07**	304.59**	206.66**	44.75**
GAM	2	189.36**	145.03**	68.24**	30.42**	6.16**
SI × RV	1	14.52	1.89	10.17	21.52**	1.01
SI × NHL	1	69.67**	90.57**	150.55**	177.53**	34.59**
SI × EV	1	6.59	14.41	27.86**	30.82**	0.00
SI × GAM	2	6.95	4.73	6.92	9.97*	4.22*
RV × GAM	2	1.15	10.95	11.59	10.83**	4.64*
NHL × GAM	2	5.36	13.26*	14.62*	14.73**	2.65
EV × GAM	2	7.94	8.07	10.16*	4.88	2.27
Error	75	6.30	4.00	3.00	2.19	1.01

* $p<0.05$

** $p<0.01$

TABLE 5.15

Mean Squares for ANOVA of Genotypic Value Expressed as a Percent of G_{MAX} for Cycles 5, 10, 15, 20 and T for All Crosses

Source of Variation	d.f.	Cycle 5	Cycle 10	Cycle 15	Cycle 20	Cycle T
Reps	3	0.65	0.27	1.77	1.32	1.75
SI	1	681.14**	1382.99**	1439.39**	1014.38**	216.38**
RV	1	351.92**	775.32**	684.39**	476.26**	72.76**
NHL	1	7957.32**	5513.65**	3341.41**	1717.95**	198.25**
EV	1	86.82**	233.33**	300.18**	268.85**	58.46**
GAM	2	128.39**	141.28**	81.20**	35.48**	6.84**
SI × RV	1	20.45	3.91	8.41	28.64**	2.46
SI × NHL	1	45.38*	74.66*	141.99**	189.43**	55.65**
SI × EV	1	5.67	7.82	29.49**	41.29**	0.67
SI × GAM	2	13.00	7.61	7.32	8.35	5.41*
RV × GAM	2	1.37	10.12	14.53*	12.84*	6.58**
NHL × GAM	2	5.28	16.13	13.33*	15.72**	3.80
EV × GAM	2	7.40	9.74	11.91*	7.19	3.01
Error	75	7.58	5.67	3.63	2.79	1.30

* $p<0.05$

** $p<0.01$

TABLE 5.16

Mean Squares for ANOVA of Yield Expressed as a Percent of C_{MAX} for Cycles 5, 10, 15, 20 and T for Selected Crosses

Source of Variation	d.f.	Cycle 5	Cycle 10	Cycle 15	Cycle 20	Cycle T
Reps	3	0.14	0.97	2.63	1.91	1.05
SI	1	985.35**	1481.37**	1365.47**	934.48**	200.83**
RV	1	339.10**	705.39**	593.26**	386.18**	51.33**
NHL	1	7329.64**	4945.20**	2871.60**	1416.56**	184.70**
EV	1	27.73*	82.09**	86.61**	25.73**	20.57**
GAM	2	180.80**	143.28**	69.96**	35.52**	10.39**
SI × RV	1	29.70*	15.73	0.11	2.91	2.54
SI × NHL	1	65.37**	73.37**	130.06**	158.88**	27.43**
SI × EV	1	1.40	2.17	9.64	9.28	4.72*
SI × GAM	2	8.46	6.27	6.38	9.80*	3.38*
RV × GAM	2	0.51	8.88	12.96*	11.98**	4.94*
NHL × GAM	2	4.78	14.87*	13.51*	13.59**	2.02
EV × GAM	2	9.15	6.02	8.37	2.74	1.37
Error	75	6.45	4.06	3.05	2.33	1.07

* $p<0.05$

** $p<0.01$

TABLE 5.17

Mean Squares for ANOVA of Yield Expressed as a Percent of G_{MAX} for Cycles 5, 10, 15, 20 and T for All Crosses

Source of Variation	d.f.	Cycle 5	Cycle 10	Cycle 15	Cycle 20	Cycle T
Reps	3	0.56	0.30	1.99	1.16	1.42
SI	1	691.20**	1397.57**	1438.37**	1020.92**	217.02**
RV	1	350.68**	768.82**	686.45**	478.25**	73.23**
NHL	1	7966.49**	5554.81**	3340.94**	1712.72**	202.18**
EV	1	88.62**	234.25**	297.43**	265.83**	57.27**
GAM	2	129.23**	144.52**	79.70**	37.15**	7.48**
SI × RV	1	20.91	5.08	8.43	29.32**	2.05
SI × NHL	1	45.83*	76.70**	141.06**	188.17**	55.24**
SI × EV	1	4.79	6.76	29.63**	39.98**	0.64
SI × GAM	2	13.07	7.84	7.97	8.64*	5.19*
RV × GAM	2	1.40	10.23	15.17*	14.38**	6.80**
NHL × GAM	2	5.15	17.54*	12.94*	14.22**	3.35
EV × GAM	2	7.17	8.75	12.93*	6.86	2.55
Error	75	7.80	5.57	3.64	2.73	1.28

* $p<0.05$

** $p<0.01$

TABLE 5.18

Mean Squares for ANOVA of h_B^2 in Cycles 5, 10, 15, 20 and T for All Crosses when $\sigma_E^2=60$

Source of Variation	d.f.	Cycle 5	Cycle 10	Cycle 15	Cycle 20	Cycle T
Reps	3	0.84	1.58	1.79	0.23	0.11
SI	1	15.88**	72.27**	99.43**	56.67**	11.95**
RV	1	12.00**	12.55**	4.51	1.33	0.50
NHL	1	523.80**	566.29**	405.81**	174.58**	19.45**
GAM	2	62.98**	24.46**	9.00**	0.55	0.07
SI × GAM	2	0.29	1.97	1.76	0.84	0.02
RV × GAM	2	2.74*	0.56	0.22	0.72	0.11
NHL × GAM	2	4.32**	5.10*	5.56*	0.43	0.04
Error	33	0.79	1.16	1.10	0.86	0.30

* $p<0.05$

** $p<0.01$

5.5 Results of the Regression Analyses: Behavior of Variables Over Time

In this section the results of the regression analyses over time are presented for percent homozygosity, genotypic value, phenotypic value or yield, and heritability in the broad sense. The analyses were performed according to the procedure discussed in Chapter IV.

Regression coefficients for percent homozygosity are presented in Tables 5.19 to 5.22. Figures 5.3 to 5.18 are graphs of the percent homozygosity averaged over the four replications of each combination of the design parameter values. All three gene action models for a specific combination of the other four factors are graphed together. For the majority of the combinations, a cubic function

$$H_{(ijk\ell m)\cdot t} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \varepsilon_{(ijk\ell m)\cdot t} \quad , \; t=1,2,\ldots,T \qquad (5.2)$$

gave the best fit for the simulated data. In the selected crosses with $\sigma_E^2 = 0$ and $r = 0.05$, two of the epistatic situations did not have a significant coefficient for the cubic term. Figures 5.3 to 5.6 are graphical representations of the simulated data which is analyzed in Table 5.19.

The cubic nature of the equation which describes the simulated data is evident from Figures 5.3, 5.5, and 5.6. From Figure 5.4 it is easily seen that the case of low selection intensity, tight linkage, and all loci heterozygous in the F_1 results in a slower approach to homozygosity. For two out of the three gene action models in this situation, the correlation between adjacent residuals was not significant. For all three gene action models, percent homozygosity has the same type of behavior for a given combination of N, r, n_h, and σ_E^2. A greater

similarity in the behavior occurs in the early cycles of selection for several combinations of N, r, n_h, and σ_E^2. Through the first 10 cycles with N = 20 and r = 0.2, the results for the three models are very close.

For the situations with $\sigma_E^2 = 60$, three of the four cases (see Figures 5.7, 5.8, and 5.10) behave in a similar way. Again, in each situation the three gene action models have similar behavior, and are closest to each other when N = 20, r = 0.05, $n_h = 100$, and $\sigma_E^2 = 60$ as can be seen from Figure 5.10. For this situation, the cubic coefficients are nonsignificant and two of the gene action models have nonsignificant correlations between adjacent residuals. All of the situations with N = 10 have nonsignificant correlation between adjacent residuals.

As can be observed from the graphs, the percent homozygosity increases slowly at first, more rapidly in the middle cycles, and then more slowly again after about 20 cycles of selection. The rate of increase of the percent homozygosity from cycle to cycle is dependent upon the gene action model and the selection pressure. With partial dominance, a large number of heterozygous loci will be represented in the selected crosses in the early cycles of selection, allelic frequencies will change slowly, and the increase in percent homozygosity will be slow at first. As the allelic frequencies shift away from 0.5 and more loci in the selected crosses become homozygous, percent homozygosity will increase more rapidly.

For the selected crosses in cycle zero, the range of average percent homozygosity for the 96 simulation runs is 0.424 to 0.492 with

a mean of 0.456. These selected crosses are less homozygous than the population from which they were selected. This is a result of the partial dominance exhibited in the linear subgenotypes of the gene action models under investigation.

Results of the regression analyses (Table 5.21) and graphs of the simulated data (Figures 5.11 to 5.14) for percent homozygosity of all crosses with $\sigma_E^2 = 0$ are similar to those for percent homozygosity of selected crosses. A linear function of cycle number is the best characterization of the results for the simple and multiple linear gene action models with N = 10, r = 0.05 and $n_h = 100$. Two of the situations have a nonsignificant correlation coefficient between adjacent residuals. As with percent homozygosity in selected crosses, for percent homozygosity of all crosses all three gene action models behave similarly for a particular combination of the other factors. A slower steady approach to complete homozygosity occurs with low selection intensity and tight linkage.

The results for percent homozygosity when $\sigma_E^2 = 60$ are given in Table 5.22 with the data graphed in Figures 5.15 to 5.18. In each situation, all three models behave similarly with the closest agreement in the early cycles of selection. Four out of the five cases in which the correlation between adjacent residual terms is nonsignificant are with N = 10. The multiple linear model with N = 10 and r = 0.05 was the only situation in which the cubic term was not significant. For cycle zero, the average percent homozygosity of all crosses ranged from 0.495 to 0.507 with an overall average of 0.5004. The expected value in cycle zero is 0.5, and deviations are due to finite population size.

As expected for either value of EV, complete homozygosity of the population is approached most rapidly when selection intensity is high and only 50 loci are heterozygous in the F_1 genotype. It would be difficult to compare the values of the regression coefficients for the different combinations of design parameters since several of the fitted equations are linear or quadratic rather than all cubic. More meaningful comparisons can be made by looking at the graphs of the output of the simulation program.

Regression coefficients for the genotypic value of selected crosses and all crosses are given in Tables 5.23 to 5.26 and graphs of the genotypic value averaged over the four replications of each combination of the design parameter values are presented in Figures 5.19 to 5.34. The quadratic function of cycle number discussed in Chapter IV does explain the behavior of genotypic value for the length of this selection program; however, in general it did not give as good a fit for the epistatic gene action model as for the simple and multiple linear gene action models. For both selected and all crosses the poorest fit of this quadratic model occurred for the situation with N = 20, r = 0.2, n_h = 50, and σ_E^2 = 60 for all three gene action models. The maximum genotypic value was achieved the earliest when N = 20, r = 0.05, n_h = 50, and σ_E^2 = 0 as is evident from Figures 5.22 and 5.30; but the most rapid approach to the maximum genotypic value in early selection cycles occurs for N = 20, r = 0.2, n_h = 100, and σ_E^2 = 0 (see Figure 5.29).

In all situations with n_h = 50, the genotypic value for the initial cycle of selection measured as a percent of the maximum genotypic value is between 63 and 71 percent for selected crosses and between 56 and

63 percent for all crosses. When $n_h = 100$, the range of genotypic values in cycle zero is 45 to 54 percent for selected crosses and 38 to 43 percent for all crosses. As was pointed out in the discussion of the significance of n_h in Section 5.4, the number of loci which are heterozygous in the initial population and are contributing to a trait which is the object of a selection program will have a large effect on the ultimate outcome of the selection program. For $n_h = 50$, the average genotypic value in cycle zero is almost 20 percent higher than for $n_h = 100$. This difference would have been even greater had the homozygous loci in the F_1 population all been homozygous for the more favorable allele.

For gene action models with partial dominance of the more favorable allele at the majority of the loci, the maximum genotypic value occurs when all loci are homozygous for the more favorable allele. Thus, the larger the value of n_h, the greater the change which can be accomplished by a selection program. On the other hand, in those situations with the smaller value of n_h the maximum genotypic value was reached earlier than in the situations with the larger value.

In all situations, the behavior of the genotypic value of selected crosses and all crosses is similar. The values for selected crosses are higher than for all crosses in the same cycle of selection as we would expect from the nature of the selection program. Similar behavior is exhibited by each of the three gene action models for a given combination of the other design parameters.

All of these regression analyses were performed under the assumption of correlation between adjacent residual terms because each new popula-

tion is generated from selected individuals of the previous generation. In 8 of the 24 regressions of genotypic value for selected crosses and 12 of the 24 regressions of genotypic value of all crosses, the correlation coefficient was not significant.

Tables 5.27 and 5.28 contain the regression coefficients for yield of selected crosses and yield of all crosses, respectively. In these analyses as in the analyses of variance, the yield for each situation has been expressed as a percent of the maximum genotypic value. Figures 5.35 to 5.42 are graphs of the means of the 4 replications for each combination of values of the design parameters. Only the results for situations with $\sigma_E^2 = 60$ are presented since yield and genotypic value are identical when $\sigma_E^2 = 60$. As one might expect, these tables and figures are almost identical to the results for genotypic value with $\sigma_E^2 = 60$, particularly in the early cycles of selection. As the population average genotypic value approaches the maximum genotypic value, the average yield surpasses the maximum genotypic value because the selected individuals have a positive environmental contribution. Therefore, it is possible for the average yield to be greater than 100 percent. (See Figure 5.37).

Results of the analyses of heritability in the broad sense are presented in Table 5.29. Graphs of the means of the simulated data for the four replications of each combination of design parameter values appear in Figures 5.43 to 5.46. This variable is calculated only for situations with $\sigma_E^2 = 60$ using data on the entire population each cycle. In each situation the multiple linear and epistatic gene action models behaved more alike and the simple linear gene action model was different.

When the selection intensity is lower (Figures 5.43 and 5.44) there is more fluctuation in the data than when the selection intensity is higher (Figures 5.45 and 5.46). In all of these situations, heritability in the broad sense for the simple linear gene action model is lower than for the other two gene action models in the early cycles of selection. This indicates that there is a higher proportion of genetic variance in the multiple linear and epistatic gene action models than in the simple linear gene action model. Further discussion of the importance of the genetic variance in a selection program such as the one under investigation is presented in Chapter VII.

From these regression analyses, we find that percent homozygosity is a cubic function of cycle number for the majority of the combinations of design parameter values. In the early cycles of selection percent homozygosity increases slowly because selection of superior genotypes is causing small changes in the frequency of + alleles at each heterozygous locus. After several cycles of selection have been accomplished and some recombination of alleles has occurred, percent homozygosity increases more rapidly each cycle. As the population approaches complete homozygosity the change in percent homozygosity decreases. At this point in a selection program very little progress is being made each cycle.

The quadratic function used in the analyses of genotypic value and yield is adequate to explain the behavior of these variables for the length of the selection program; however, if the selection program were extended, a different function might better characterize the behavior of genotypic value and yield.

5.6 Summary

Experimental and simulated observations from this selection program compare favorably to each other for the limited number of selection cycles included in the comparison. In the early cycles of selection, more variation occurs for both experimental and simulated observations; therefore, it would be desirable to extend these comparisons over additional cycles.

Fixation of less desirable alleles in the population occurred in several of the simulation runs; and in the majority of these situations, the subgenotype involved in either the multiple linear or the epistatic gene action model was the one with the smallest contribution to the genotypic value. The simulation results indicate that sufficient progress can be made in a recurrent selection program with N = 20 with only minimal loss of the more desirable alleles.

For most of the variables analyzed, the gene action model did not have a significant effect on the results of the selection program. Other factors, in particular selection intensity and number of heterozygous loci in the F_1, are more important for this specific selection program. Additional discussion of the results of the simulation study is presented in Chapter VIII.

TABLE 5.19

Regression Coefficients for Percent Homozygosity for Selected Crosses when $\sigma_E^2=0$

No. of Crosses Selected	r	n_h	Gene Action Model	$\hat{\beta}_0$	$\hat{\beta}_1$ [1]	$\hat{\beta}_2$ [2]	$\hat{\beta}_3$ [3]	$\hat{\rho}$	$\hat{\sigma}^2$ [2]	R^2
10	0.2	50	Simple Linear	0.474	1.16	15.01	-44.10	0.322†	0.274	0.999
			Multiple Linear	0.472	1.92	8.09	-28.92	0.520	0.272	0.997
			Epistatic	0.468	1.58	9.02	-29.83	0.653	0.307	0.994
	0.05	100	Simple Linear	0.442	1.93	-3.00	4.91*	0.090†	0.364	0.998
			Multiple Linear	0.452	1.83	-3.39	6.04*	0.486	0.186	0.996
			Epistatic	0.449	1.86	-0.85		0.195†	0.169	0.999
20	0.2	100	Simple Linear	0.484	0.77	16.06	-43.62	0.610	0.114	0.999
			Multiple Linear	0.474	1.01	12.19	-32.13	0.488	0.230	0.998
			Epistatic	0.476	1.33	9.78	-29.58	0.501	0.197	0.998
	0.05	50	Simple Linear	0.445	2.18	9.90*	-36.71*	0.523	0.953	0.992
			Multiple Linear	0.436	2.49	7.31	-32.80	0.629	0.268	0.997
			Epistatic	0.421	4.22	-7.86		0.891	0.544	0.966

† not significant

* $p<0.05$

[1] multiplied by 10^2

[2] multiplied by 10^4

[3] multiplied by 10^6

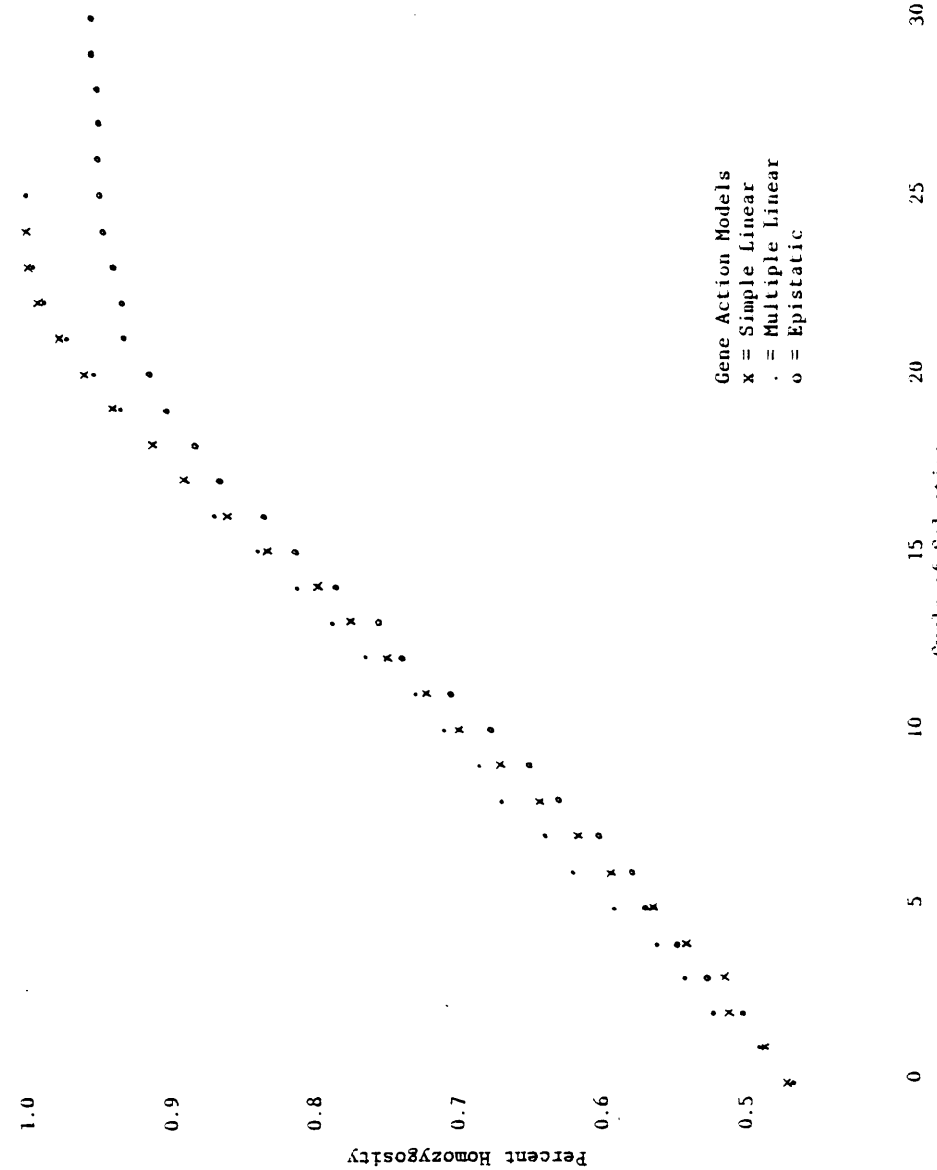


Figure 5.3 Percent Homozygosity of Selected Crosses with N=10, r=0.2, n_h=50, σ^2_E=0

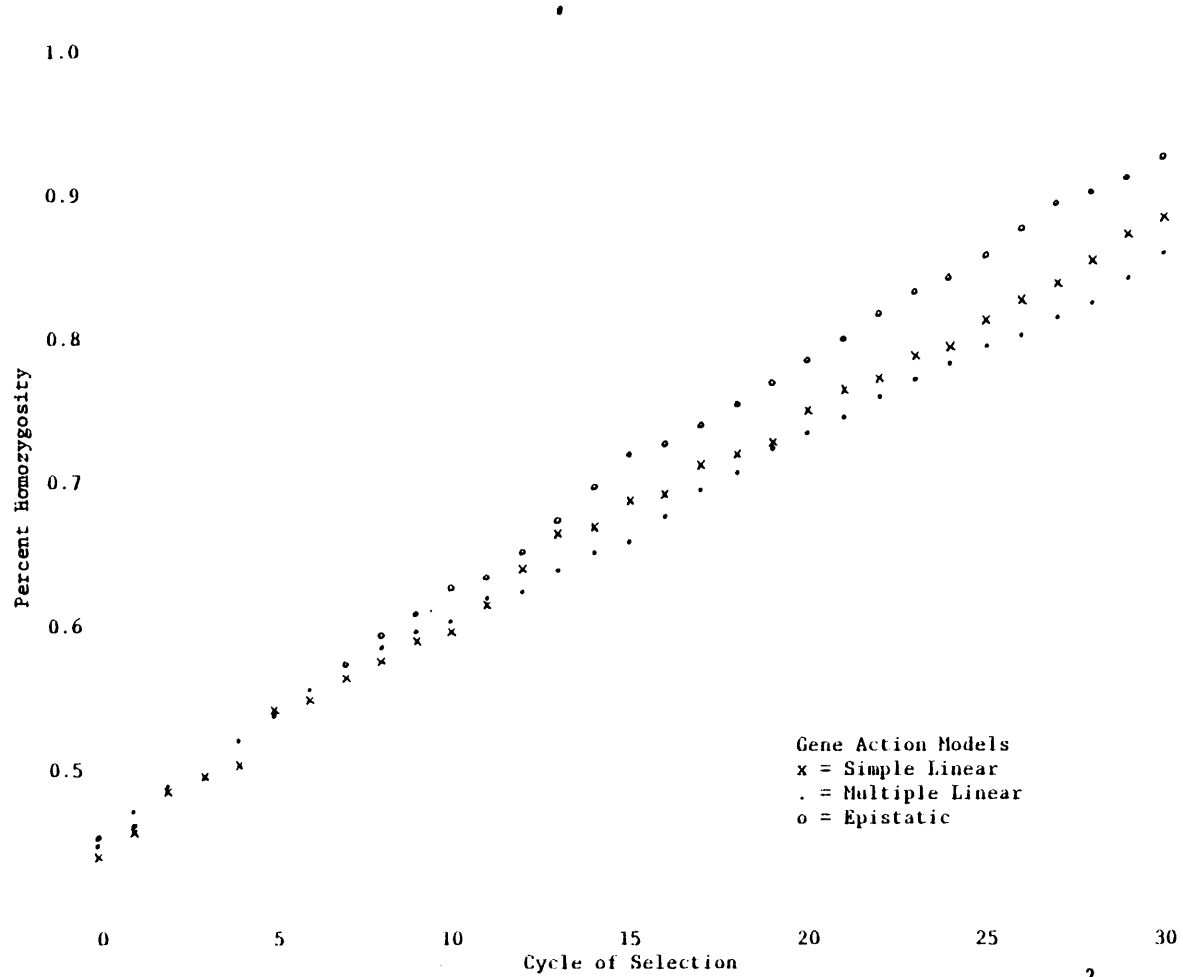


Figure 5.4 Percent Homozygosity of Selected Crosses with N=10, r=0.05, n_h=100, σ_E^2=0

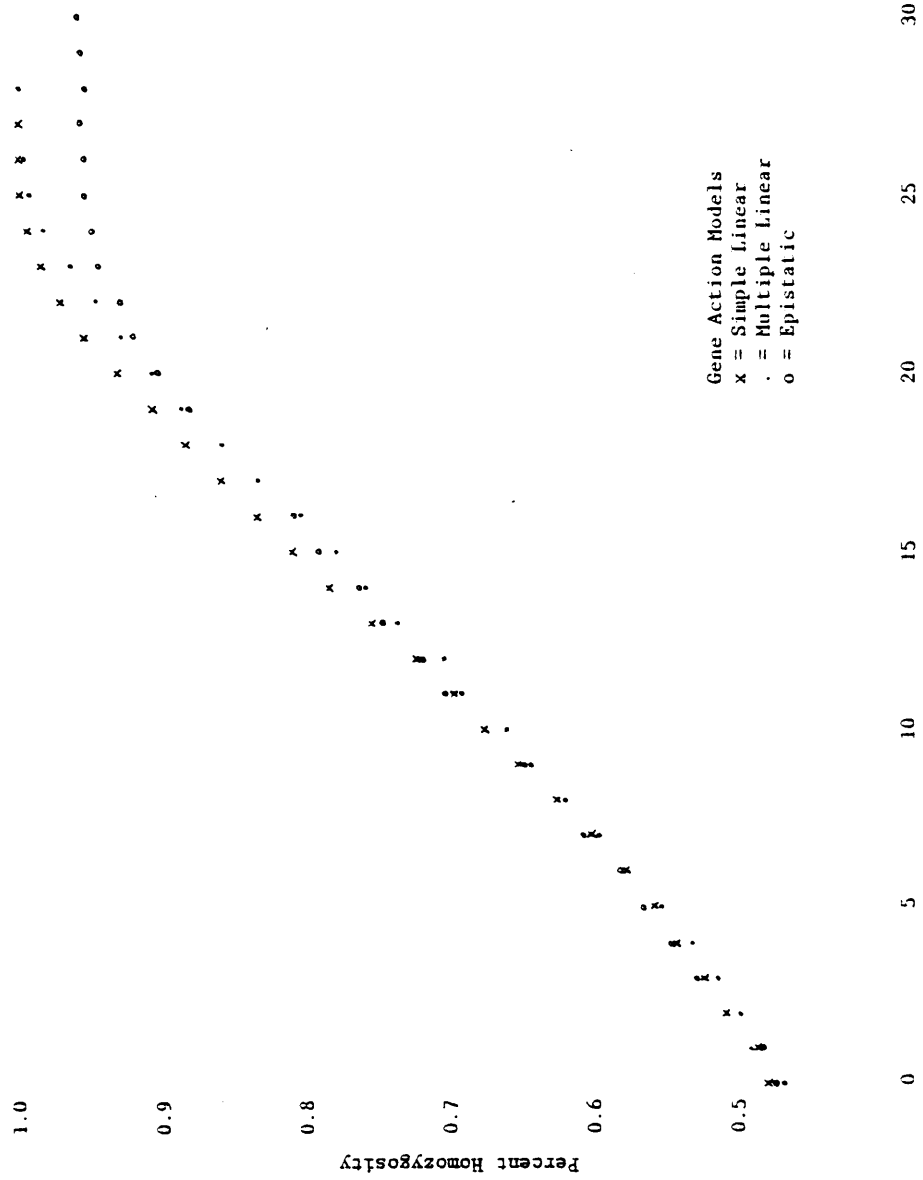


Figure 5.5 Percent Homozygosity of Selected Crosses with N=20, r=0.2, n_h=100, σ^2_E=0

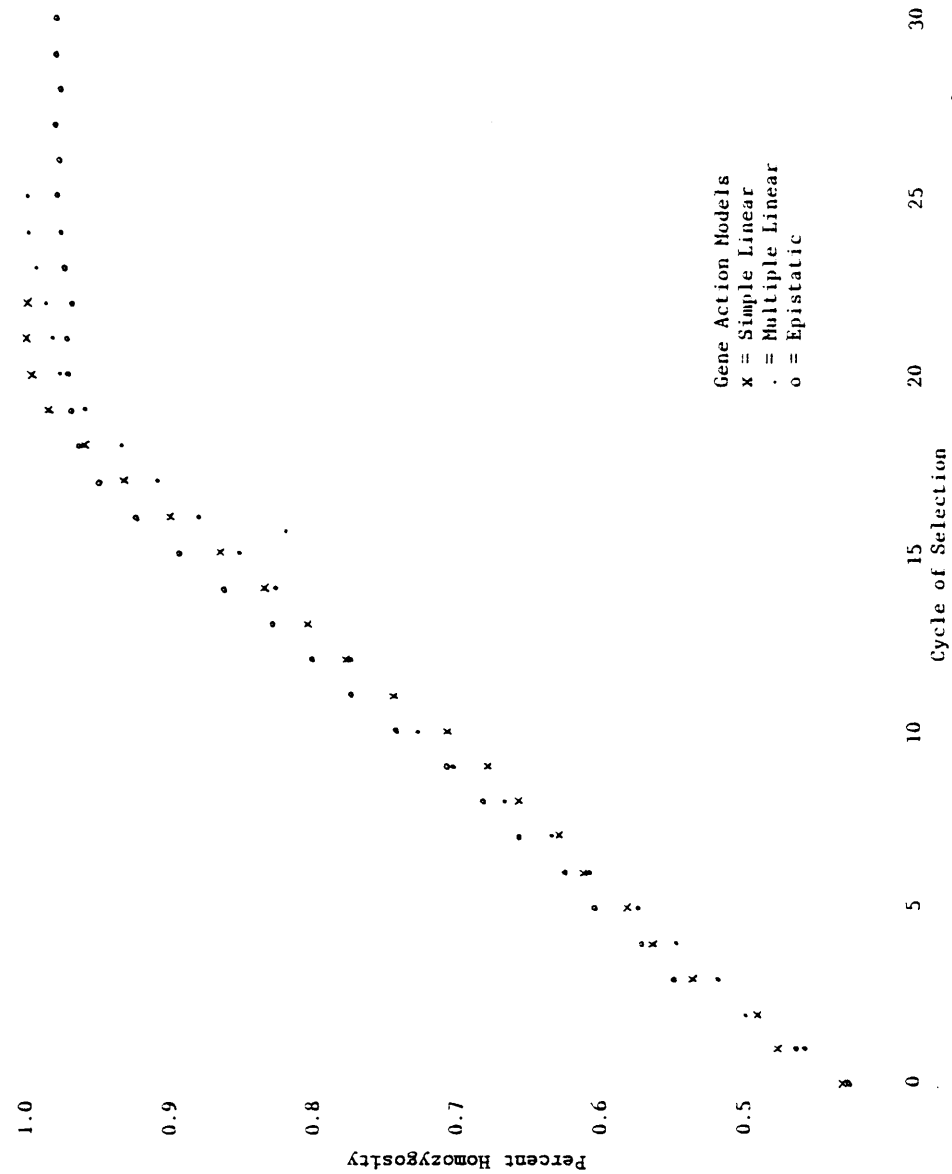


Figure 5.6 Percent Homozygosity of Selected Crosses with N=20, r=0.05, n_h=50, σ^2_E=0

TABLE 5.20

Regression Coefficients for Percent Homozygosity for Selected Crosses when $\sigma_E^2=60$

No. of Crosses Selected	r	n_h	Gene Action Model	$\hat{\beta}_0$	$\hat{\beta}_1$ [1]	$\hat{\beta}_2$ [2]	$\hat{\beta}_3$ [3]	$\hat{\rho}$	$\hat{\sigma}^2$ [2]	R^2
10	0.2	100	Simple Linear	0.473	1.25	2.28	-6.05	-0.093†	0.191	0.999
			Multiple Linear	0.476	1.01	5.26	-13.24	0.040†	0.290	0.999
			Epistatic	0.471	1.14	2.89	-7.78	0.019†	0.117	0.999
	0.05	50	Simple Linear	0.437	2.02	-1.52		0.016†	0.374	0.998
			Multiple Linear	0.448	2.64	-5.31	5.38*	0.155†	0.321	0.998
			Epistatic	0.467	1.90	-0.08†	-4.37	-0.320†	0.376	0.999
20	0.2	50	Simple Linear	0.454	2.82	0.52†	-13.62	0.755	0.246	0.993
			Multiple Linear	0.463	2.64	0.59†	-12.43	0.669	0.326	0.993
			Epistatic	0.454	3.36	-5.66		0.644	0.253	0.995
	0.05	100	Simple Linear	0.440	2.14	-1.72		-0.145†	0.121	0.999
			Multiple Linear	0.445	2.17	-1.84		0.539	0.266	0.996
			Epistatic	0.444	2.19	-2.08		0.253†	0.141	0.999

† not significant

* $p<0.05$

[1] multiplied by 10^2

[2] multiplied by 10^4

[3] multiplied by 10^6

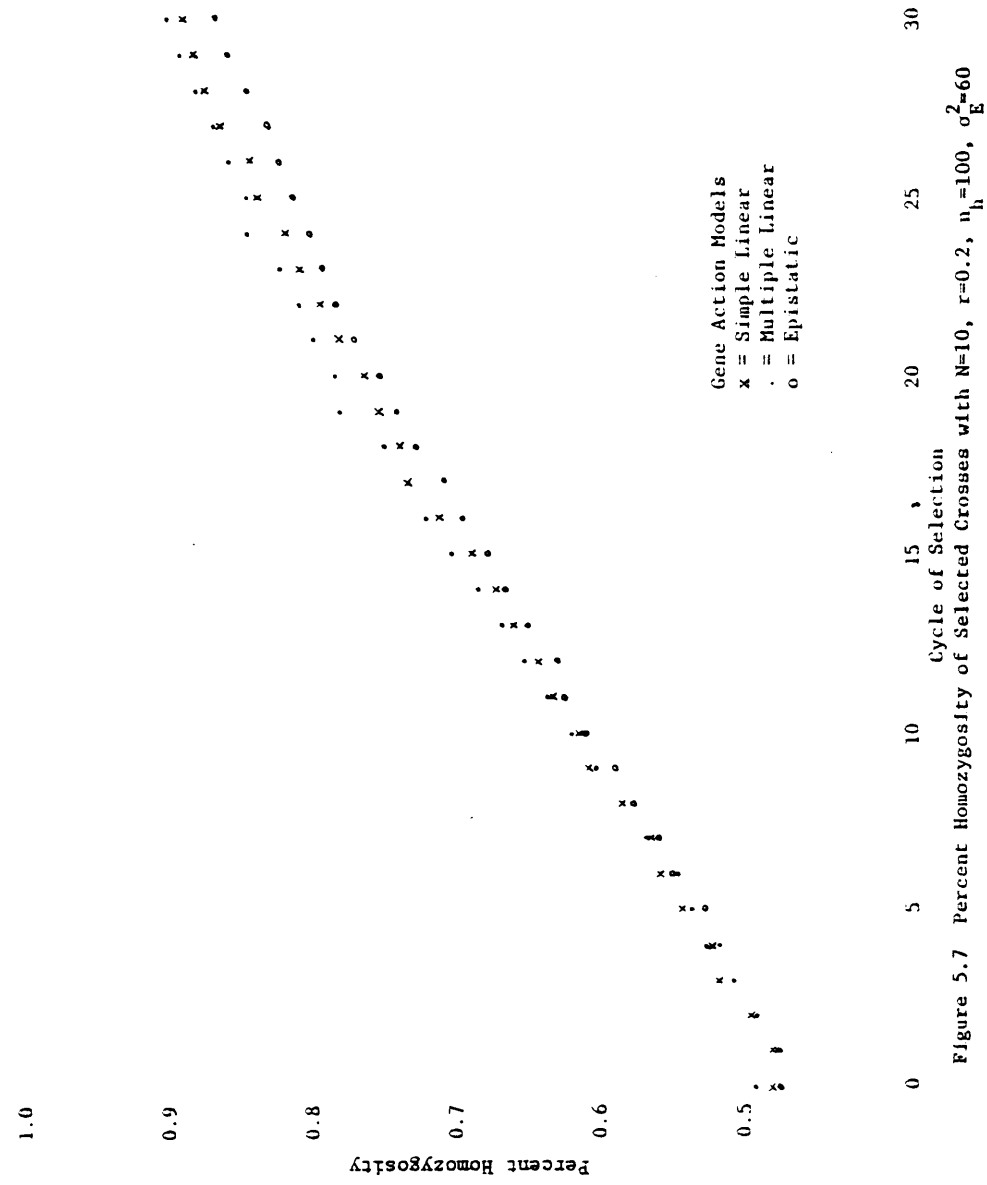


Figure 5.7 Percent Homozygosity of Selected Crosses with N=10, r=0.2, μ_h=100, σ^2_E=60

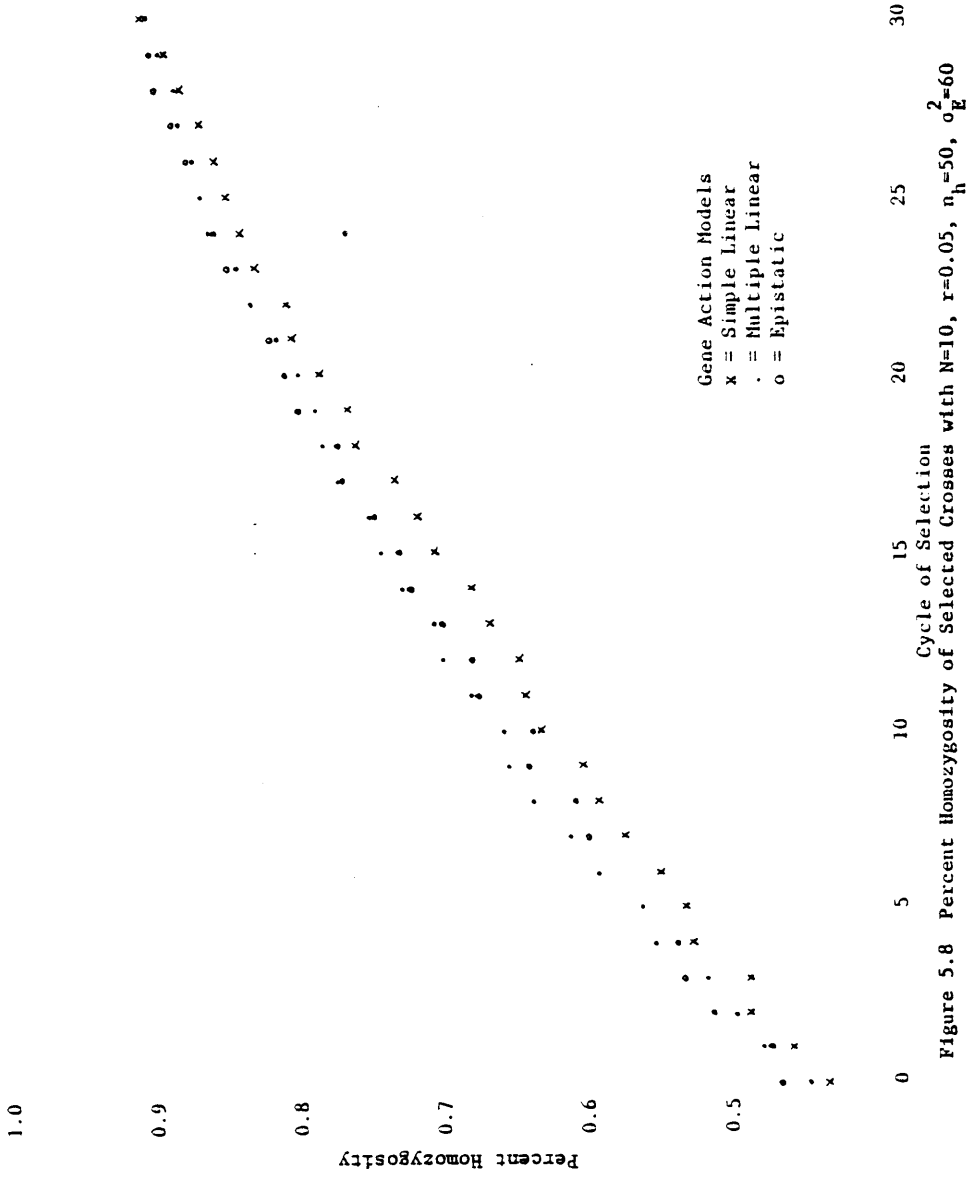


Figure 5.8 Percent Homozygosity of Selected Crosses with N=10, r=0.05, n_h=50, σ^2_E=60

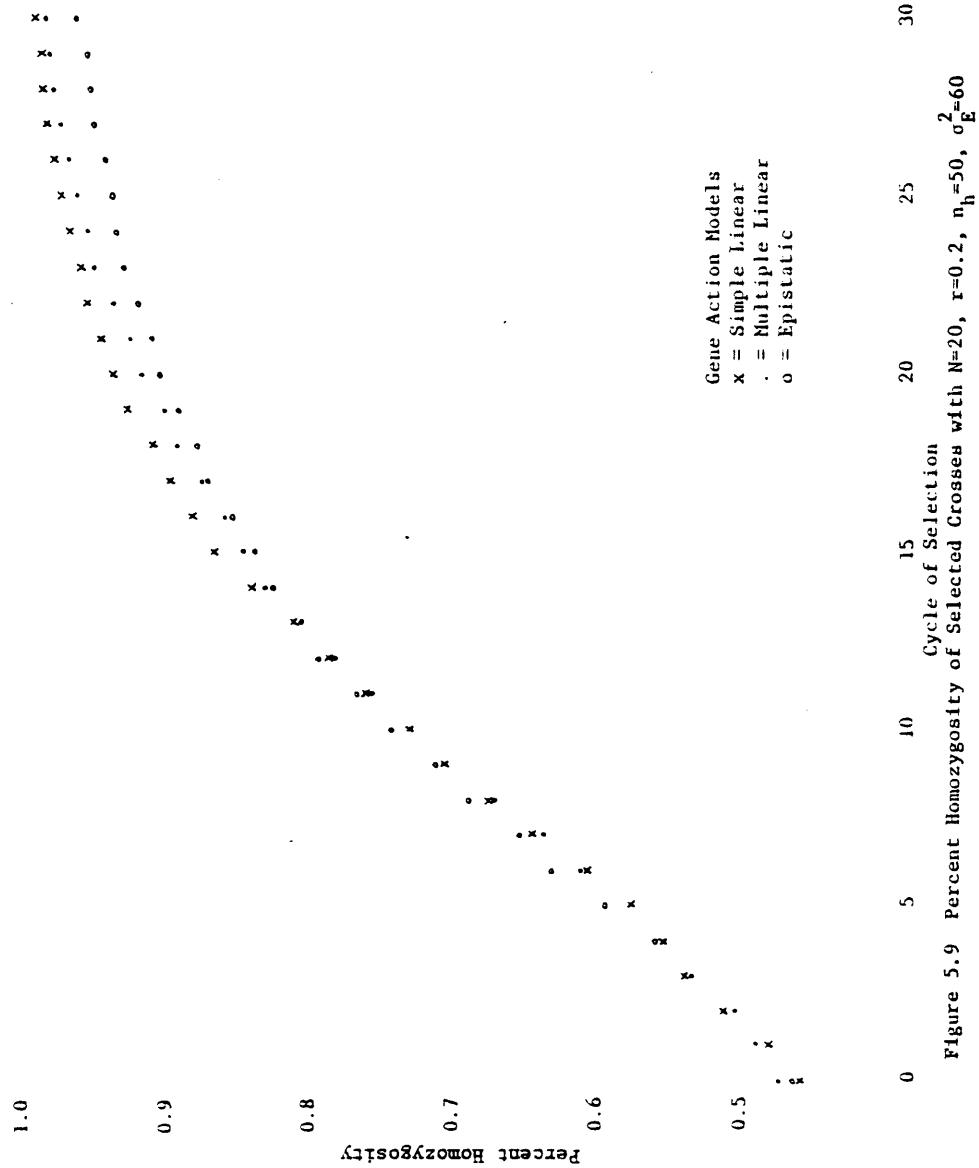


Figure 5.9 Percent Homozygosity of Selected Crosses with N=20, r=0.2, n_h=50, σ^2_E=60

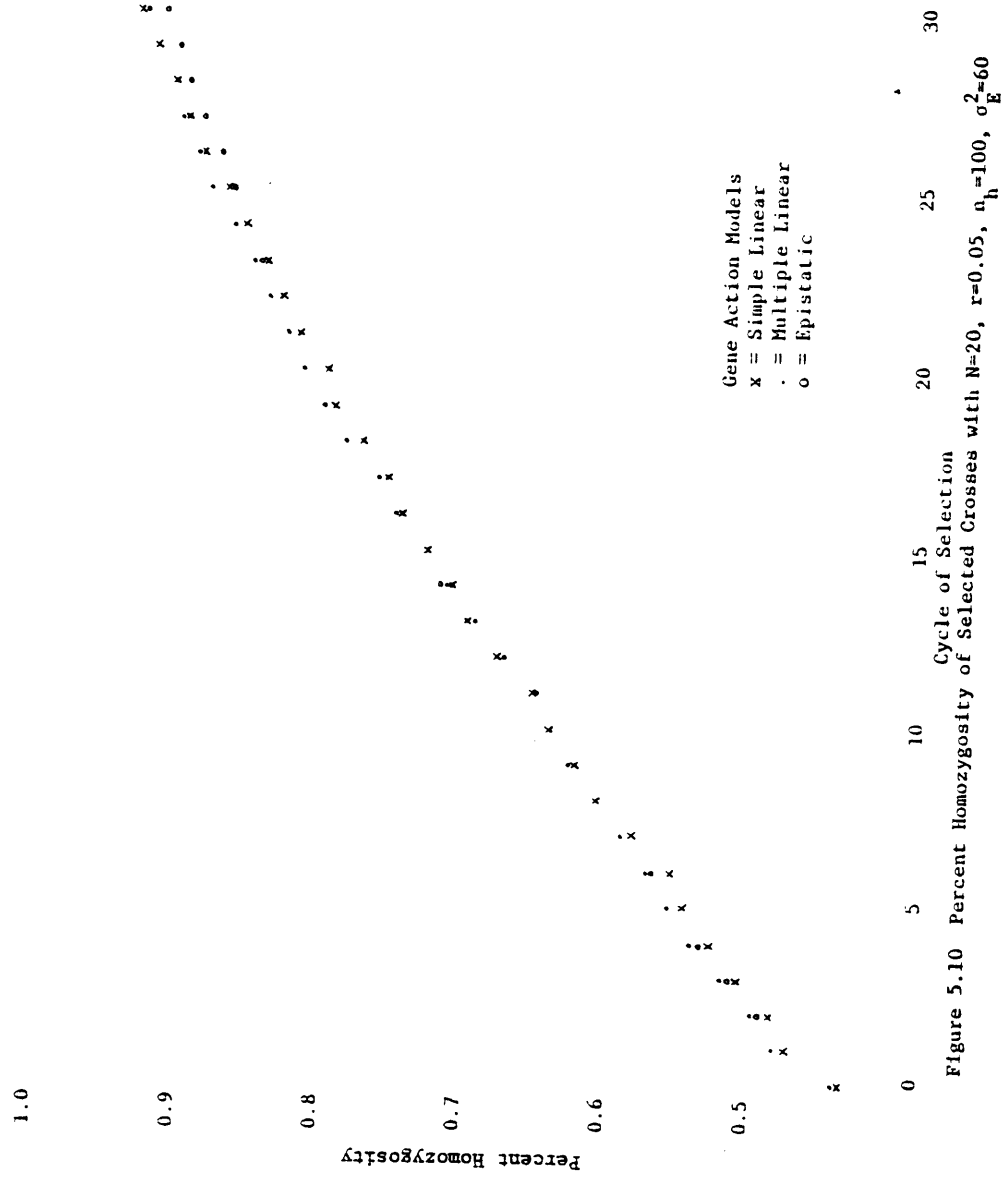


Figure 5.10 Percent Homozygosity of Selected Crosses with N=20, r=0.05, n_h=100, σ^2_E=60

TABLE 5.21

Regression Coefficients for Percent Homozygosity for All Crosses when $\sigma_E^2=0$

No. of Crosses Selected	r	n_h	Gene Action Model	$\hat{\beta}_0$	$\hat{\beta}_1$ [1]	$\hat{\beta}_2$ [2]	$\hat{\beta}_3$ [3]	$\hat{\rho}$	$\hat{\sigma}^2$ [2]	R^2
10	0.2	50	Simple Linear	0.501	0.842	14.12	-37.09	0.457	0.085	0.999
			Multiple Linear	0.496	1.46	9.33	-28.11	0.060†	0.235	0.999
			Epistatic	0.498	0.98	11.81	-33.57	0.496	0.202	0.998
	0.05	100	Simple Linear	0.494	1.28			0.459	0.130	0.997
			Multiple Linear	0.502	1.18			0.429	0.151	0.997
			Epistatic	0.491	1.29	1.90*	-4.79	0.305†	0.148	0.999
20	0.2	100	Simple Linear	0.504	0.54	15.24	-38.43	0.610	0.069	0.999
			Multiple Linear	0.500	0.63	12.67	-30.36	0.656	0.077	0.999
			Epistatic	0.499	0.87	11.38	-30.59	0.662	0.062	0.999
	0.05	50	Simple Linear	0.507	1.36	10.88	-30.53	0.548	0.378	0.996
			Multiple Linear	0.498	1.45	11.31	-36.35	0.571	0.193	0.998
			Epistatic	0.493	2.28	4.78†	-23.43	0.761	0.482	0.984

† not significant

* $p<0.05$

[1] multiplied by 10^2

[2] multiplied by 10^4

[3] multiplied by 10^6

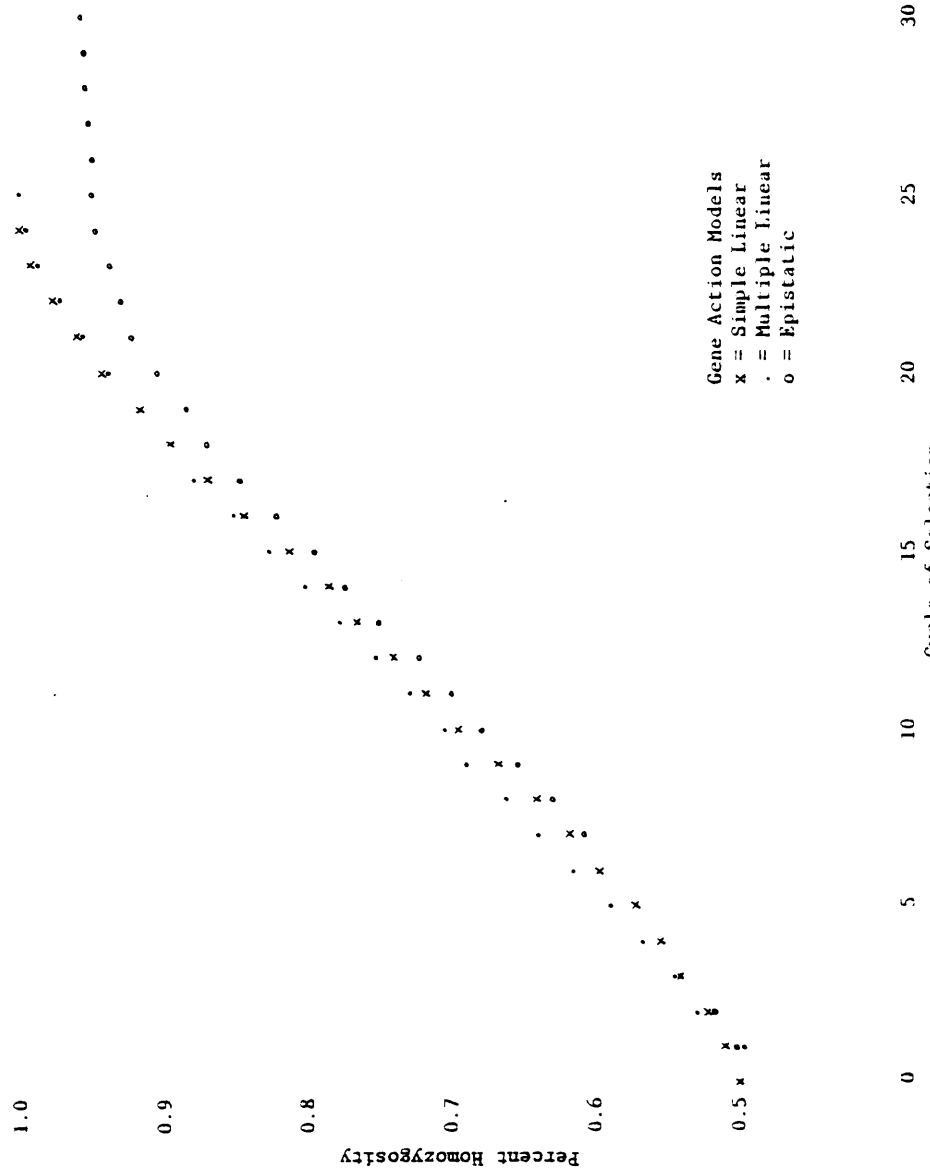


Figure 5.11 Percent Homozygosity of All Crosses with N=10, r=0.2, n_h=50, σ^2_E=0

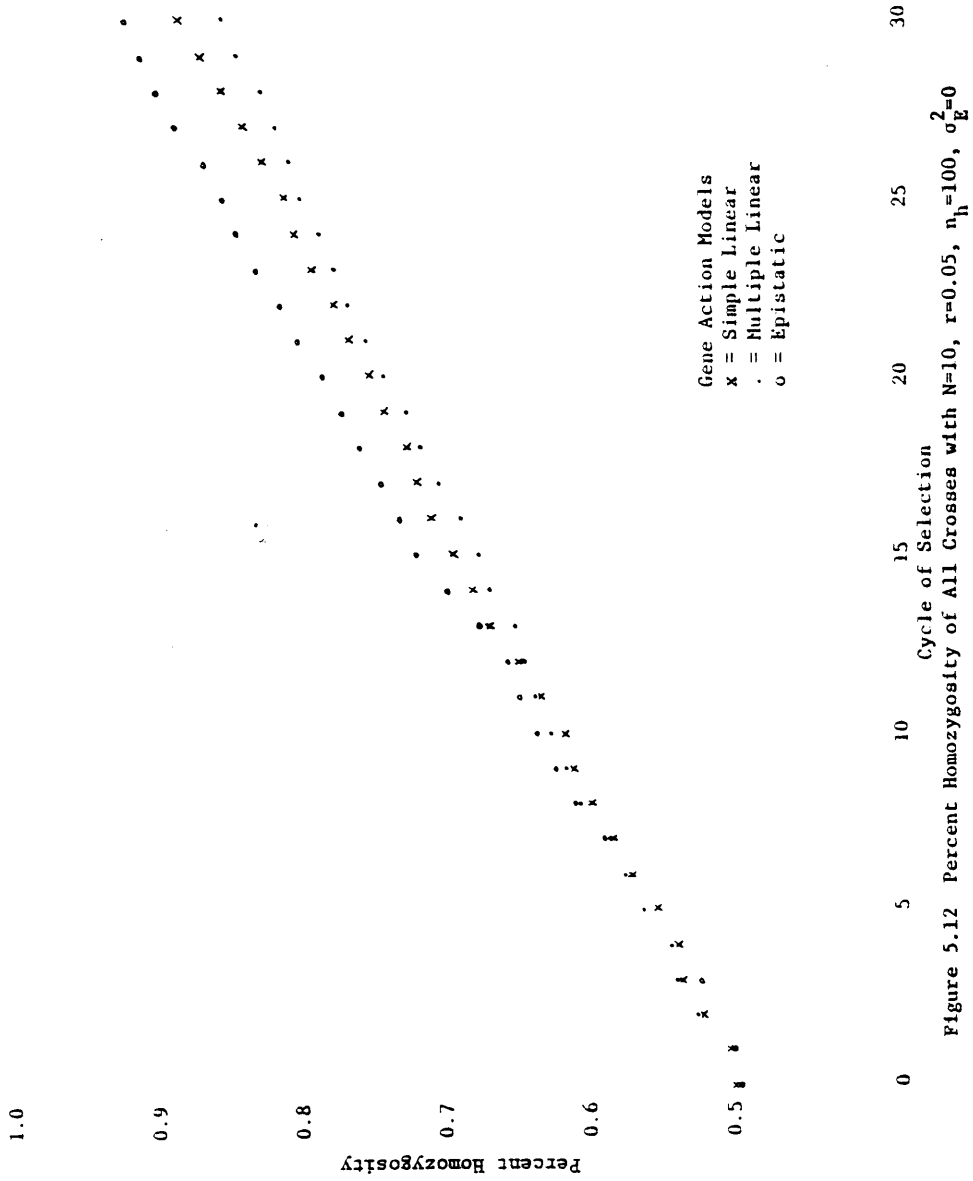


Figure 5.12 Percent Homozygosity of All Crosses with N=10, r=0.05, n_h=100, σ^2_E=0

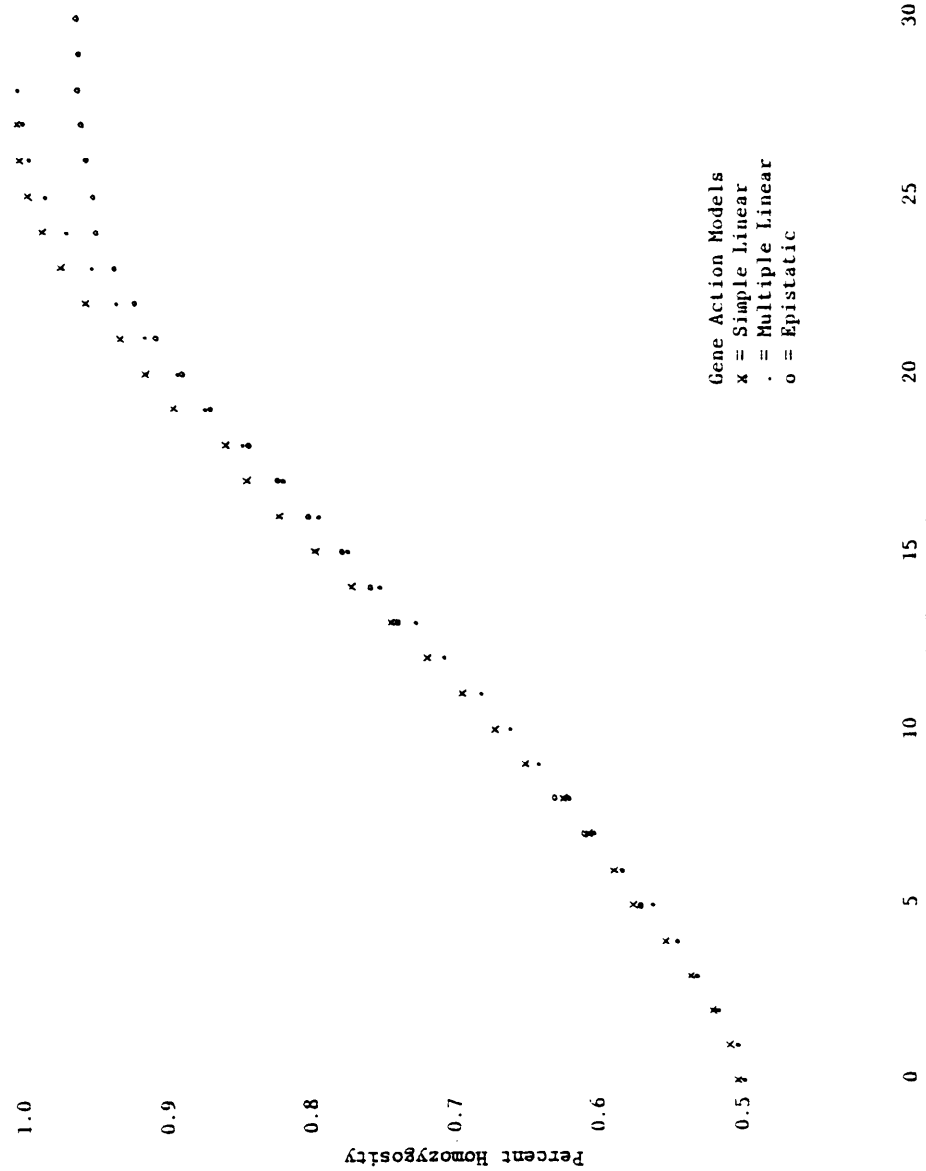


Figure 5.13 Percent Homozygosity of All Crosses with N=20, r=0.2, n_h=100, σ^2_E=0

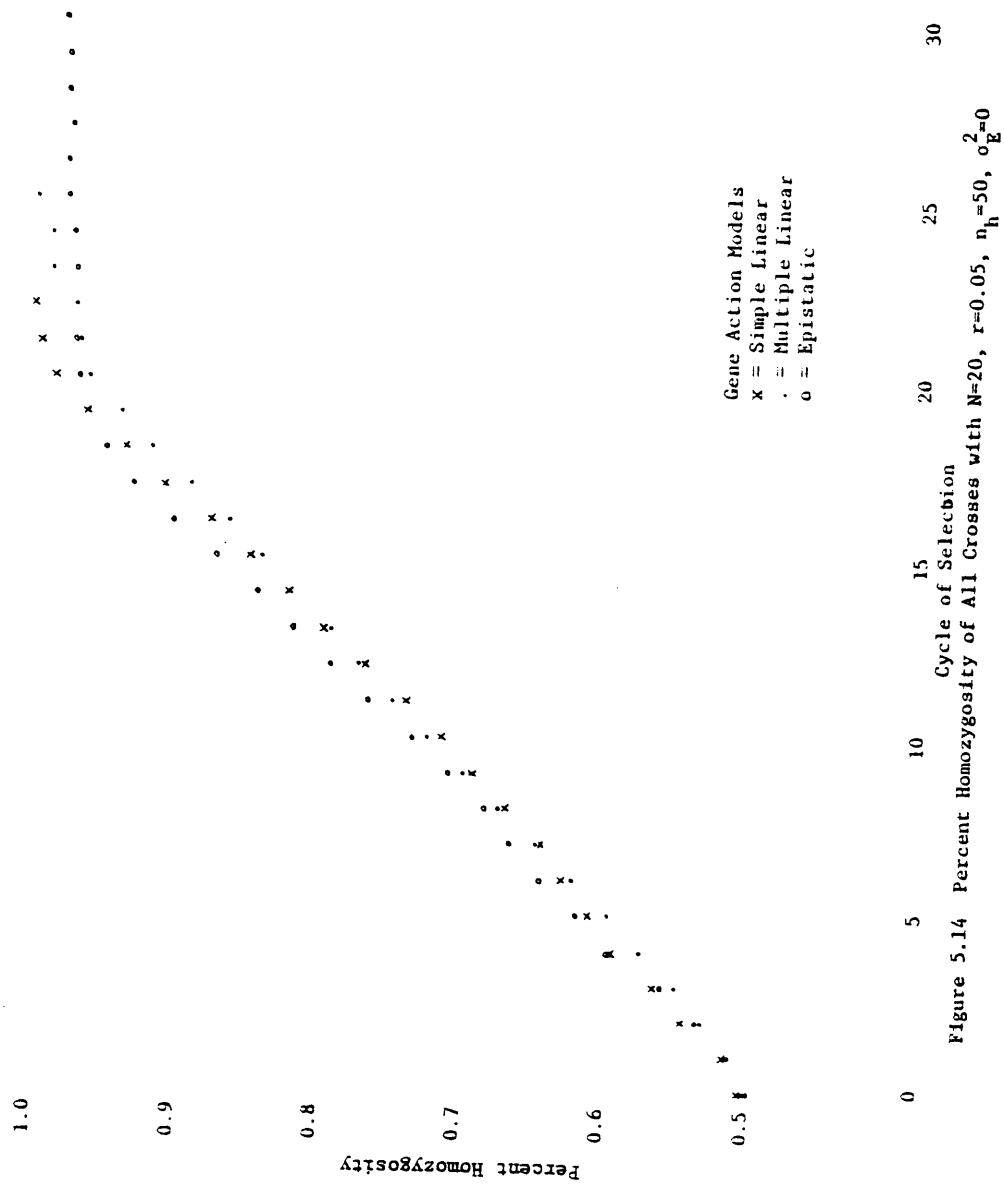


Figure 5.14 Percent Homozygosity of All Crosses with N=20, r=0.05, n_h=50, σ^2_E=0

TABLE 5.22

Regression Coefficients for Percent Homozygosity for All Crosses when $\sigma_E^2=60$

No. of Crosses Selected	r	n_h	Gene Action Model	$\hat{\beta}_0$	$\hat{\beta}_1$ [1]	$\hat{\beta}_2$ [2]	$\hat{\beta}_3$ [3]	$\hat{\rho}$	$\hat{\sigma}^2$ [2]	R^2
10	0.2	100	Simple Linear	0.493	1.07	2.70	-6.26	0.431	0.065	0.999
			Multiple Linear	0.493	0.87	5.34	-12.75	-0.040†	0.185	0.999
			Epistatic	0.498	0.80	4.31	-9.96	0.113†	0.097	0.999
	0.05	50	Simple Linear	0.495	1.22	2.42	-6.72	0.227†	0.209	0.998
			Multiple Linear	0.493	2.01	-2.16		0.501	0.187	0.996
			Epistatic	0.501	1.62	0.34†	-4.08	-0.077†	0.099	0.999
20	0.2	50	Simple Linear	0.491	1.81	6.48	-23.78	0.707	0.265	0.993
			Multiple Linear	0.494	1.84	4.92	-19.25	0.710	0.300	0.992
			Epistatic	0.490	2.33	0.80†	-11.67	0.682	0.279	0.992
	0.05	100	Simple Linear	0.495	1.37	1.91	-5.83	0.347	0.072	0.999
			Multiple Linear	0.501	1.30	2.50	-7.21	0.660	0.065	0.998
			Epistatic	0.498	1.37	1.93	-6.53	0.261†	0.072	0.999

† not significant

[1] multiplied by 10^2

[2] multiplied by 10^4

[3] multiplied by 10^6

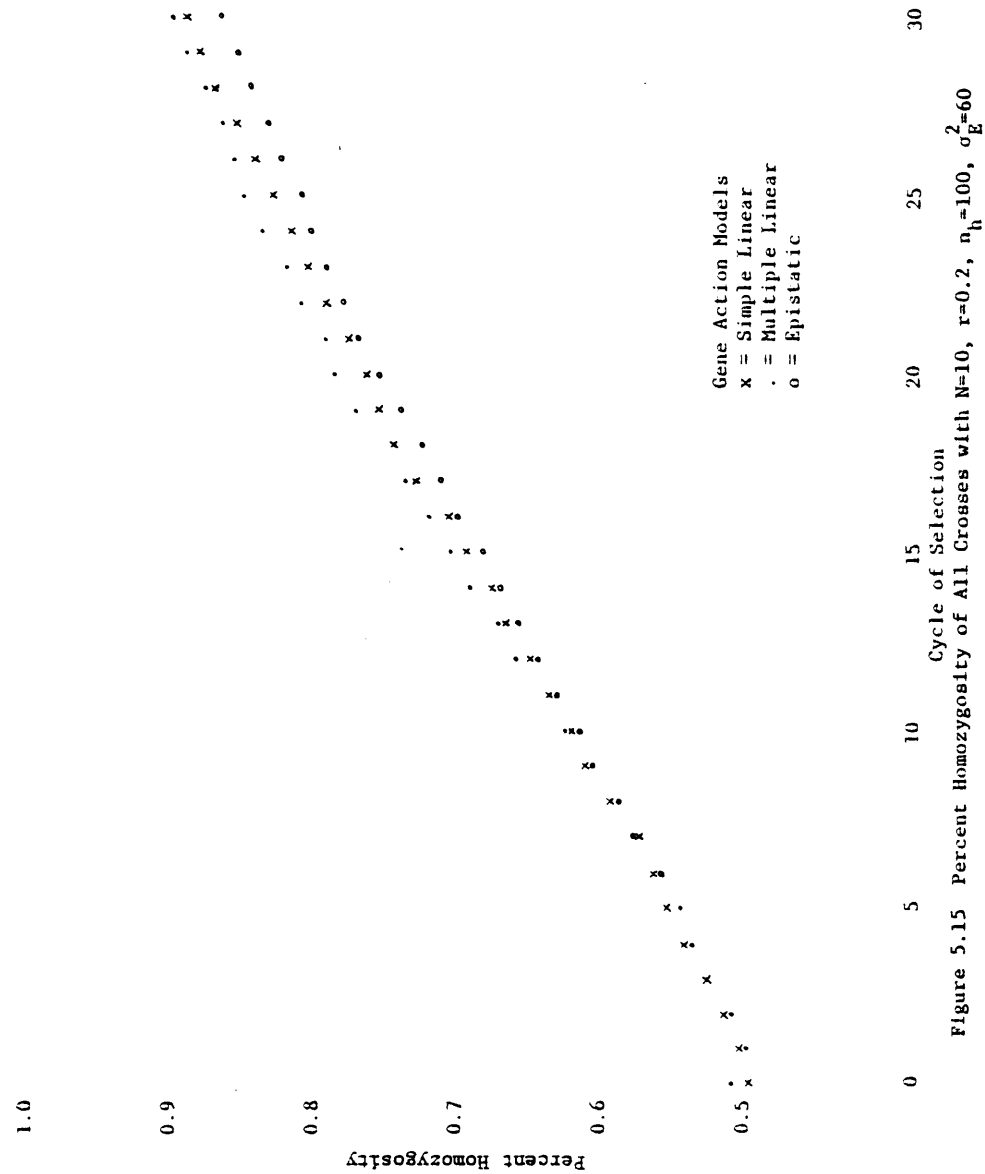


Figure 5.15 Percent Homozygosity of All Crosses with N=10, r=0.2, n_h=100, σ^2_E=60

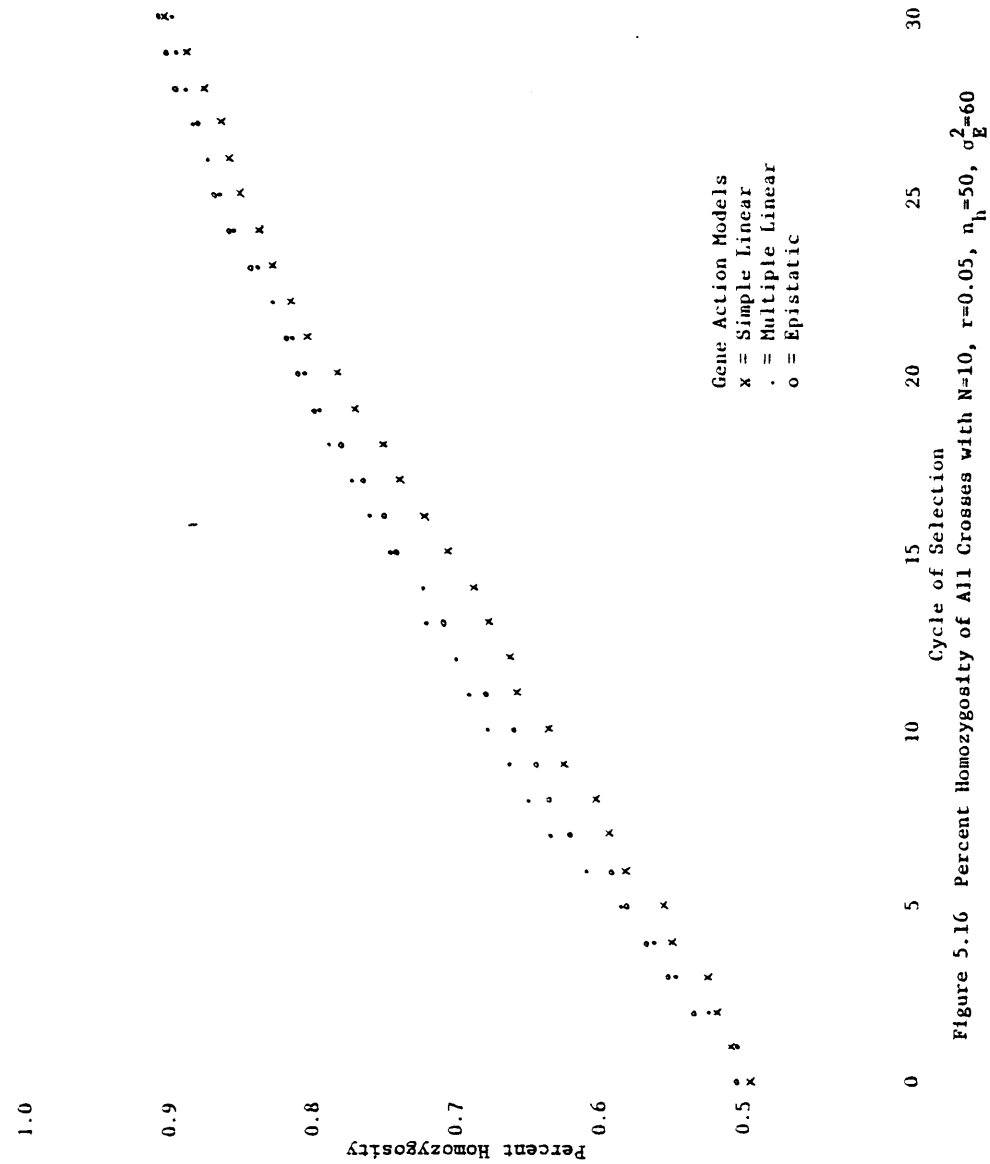


Figure 5.16 Percent Homozygosity of All Crosses with N=10, r=0.05, n_h=50, σ^2_E=60

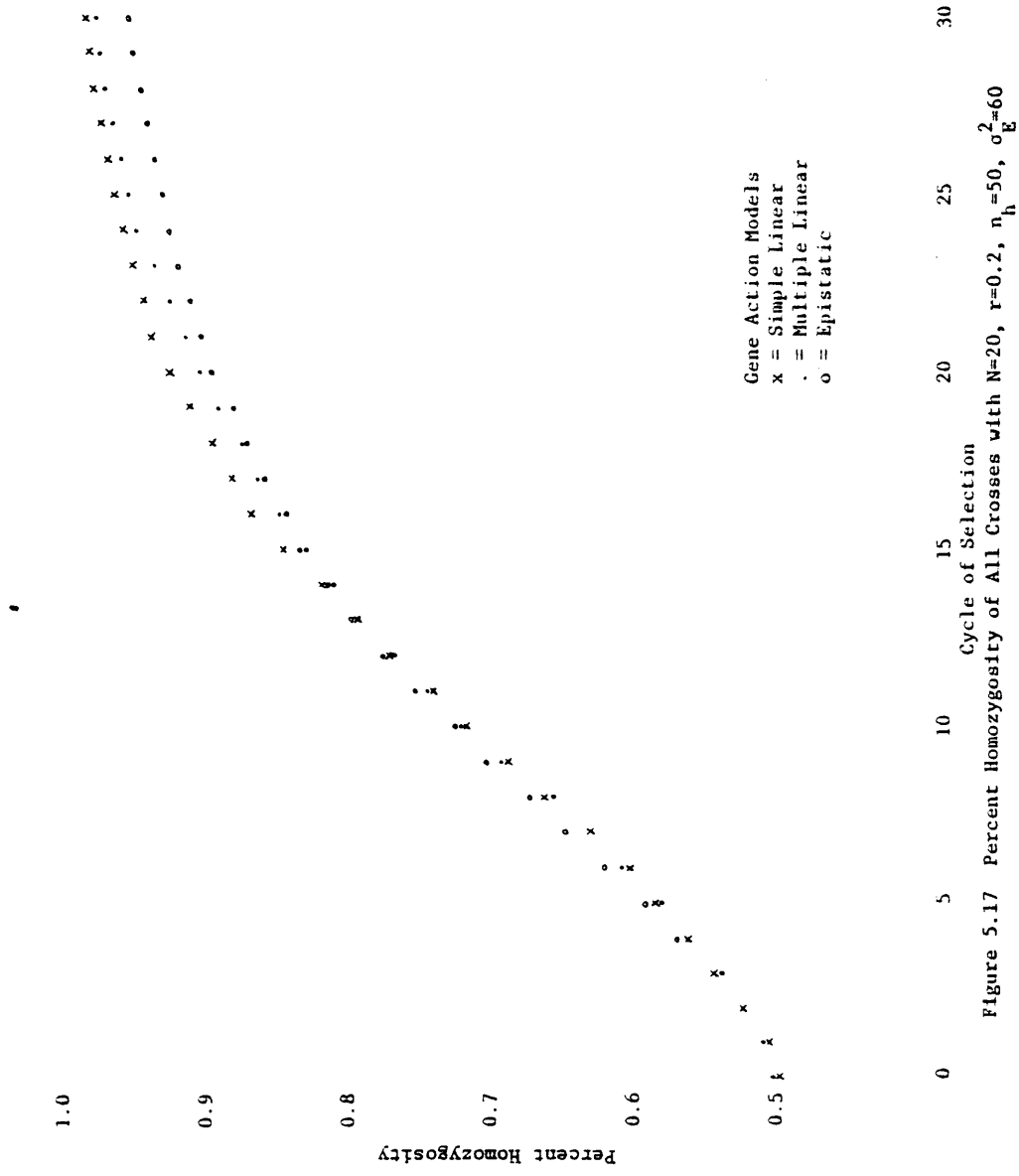


Figure 5.17 Percent Homozygosity of All Crosses with N=20, r=0.2, n_h=50, σ^2_E=60

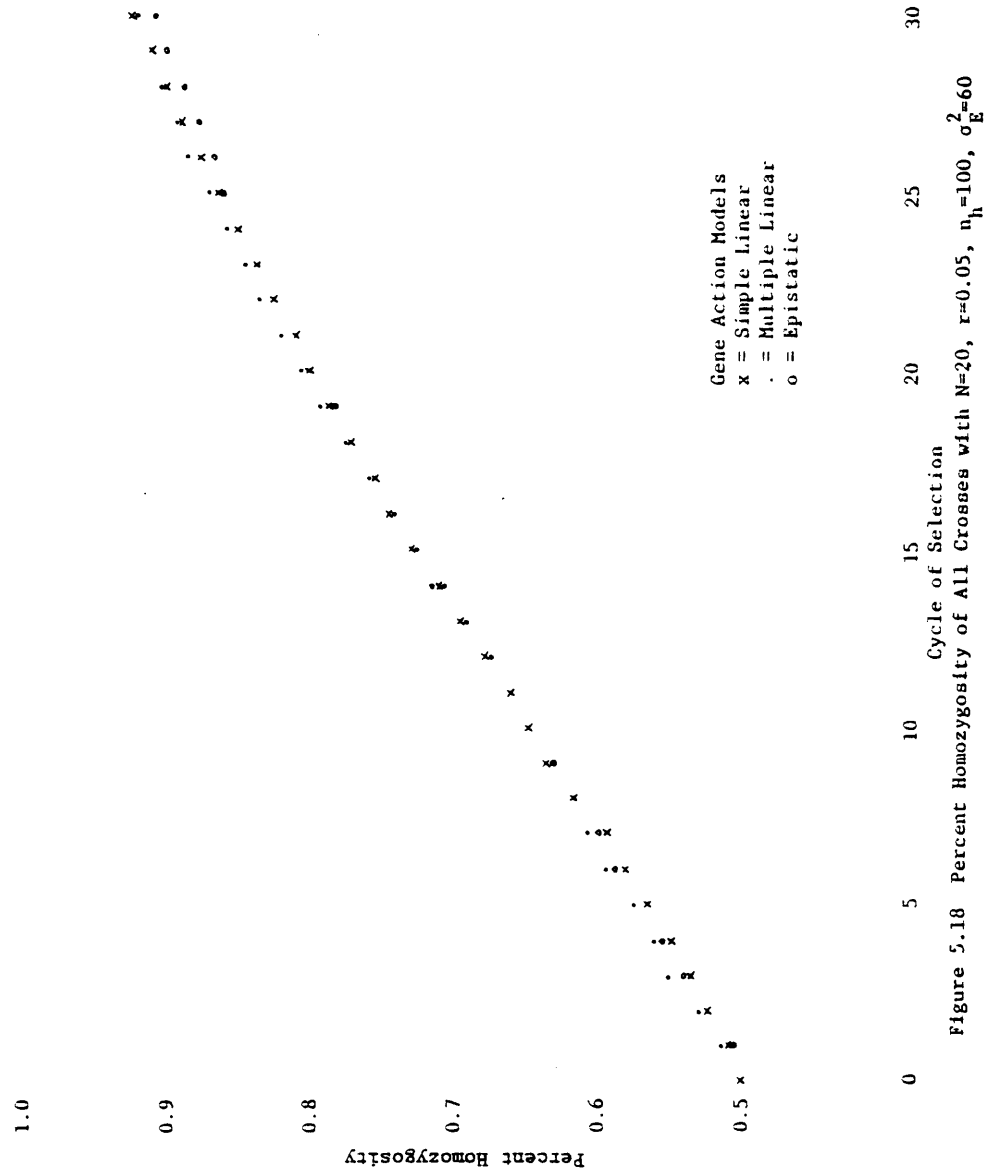


Figure 5.18 Percent Homozygosity of All Crosses with N=20, r=0.05, n_h=100, σ^2_E=60

TABLE 5.23

Regression Coefficients for Genotypic Value Expressed as a Percent of G_{MAX} for Selected Crosses when $\sigma_E^2=0$

No. of Crosses Selected	r	n_h	Gene Action Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$	$\hat{\sigma}^2$	R^2
10	0.2	50	Simple Linear	63.08	3.18	-0.068	0.241	0.091	0.999
			Multiple Linear	68.97	2.92	-0.068	0.446	0.399	0.990
			Epistatic	69.74	2.75	-0.060	0.650	0.370	0.980
	0.05	100	Simple Linear	45.94	2.19	-0.019	0.395	0.177	0.998
			Multiple Linear	50.02	2.17	-0.025	0.345	0.110	0.999
			Epistatic	49.99	2.85	-0.042	0.274	0.136	0.999
20	0.2	100	Simple Linear	47.98	4.42	-0.096	0.675	0.115	0.998
			Multiple Linear	51.73	4.09	-0.086	0.756	0.219	0.992
			Epistatic	53.46	4.10	-0.087	0.765	0.453	0.982
	0.05	50	Simple Linear	69.80	2.71	-0.060	0.634	0.078	0.996
			Multiple Linear	69.68	2.68	-0.059	0.078	0.056	0.999
			Epistatic	70.88	2.71	-0.059	0.823	0.194	0.975

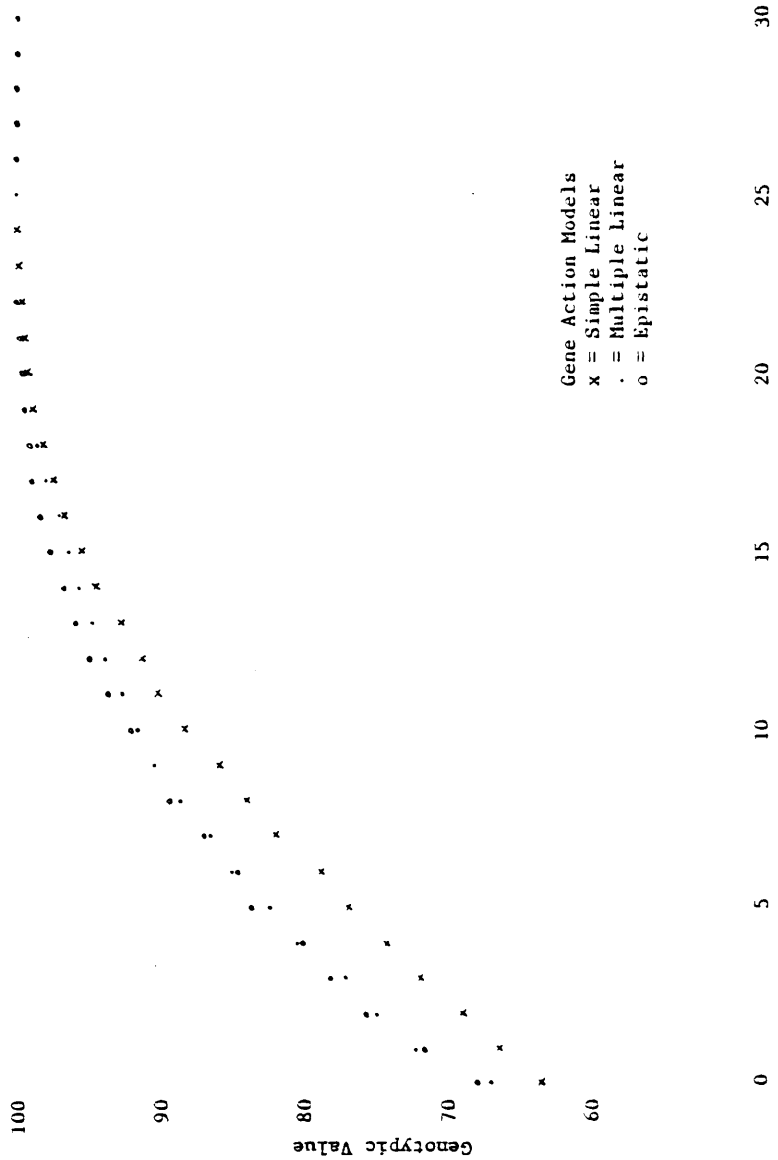


Figure 5.19 Genotypic Value of Selected Crosses with N=10, r=0.2, n_l=50, σ^2_E=0

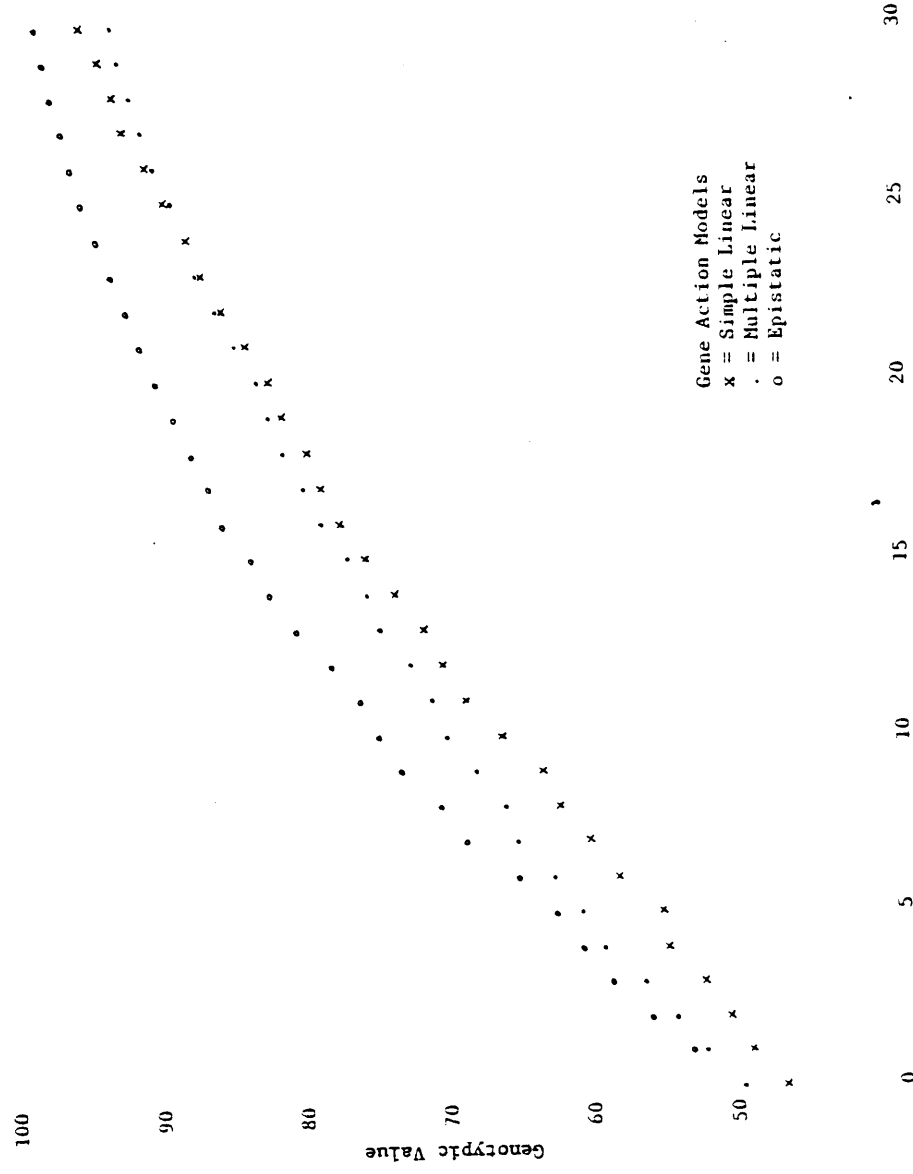


Figure 5.20 Genotypic Value of Selected Crosses with N=10, r=0.05, n_h=100, σ^2_E=0

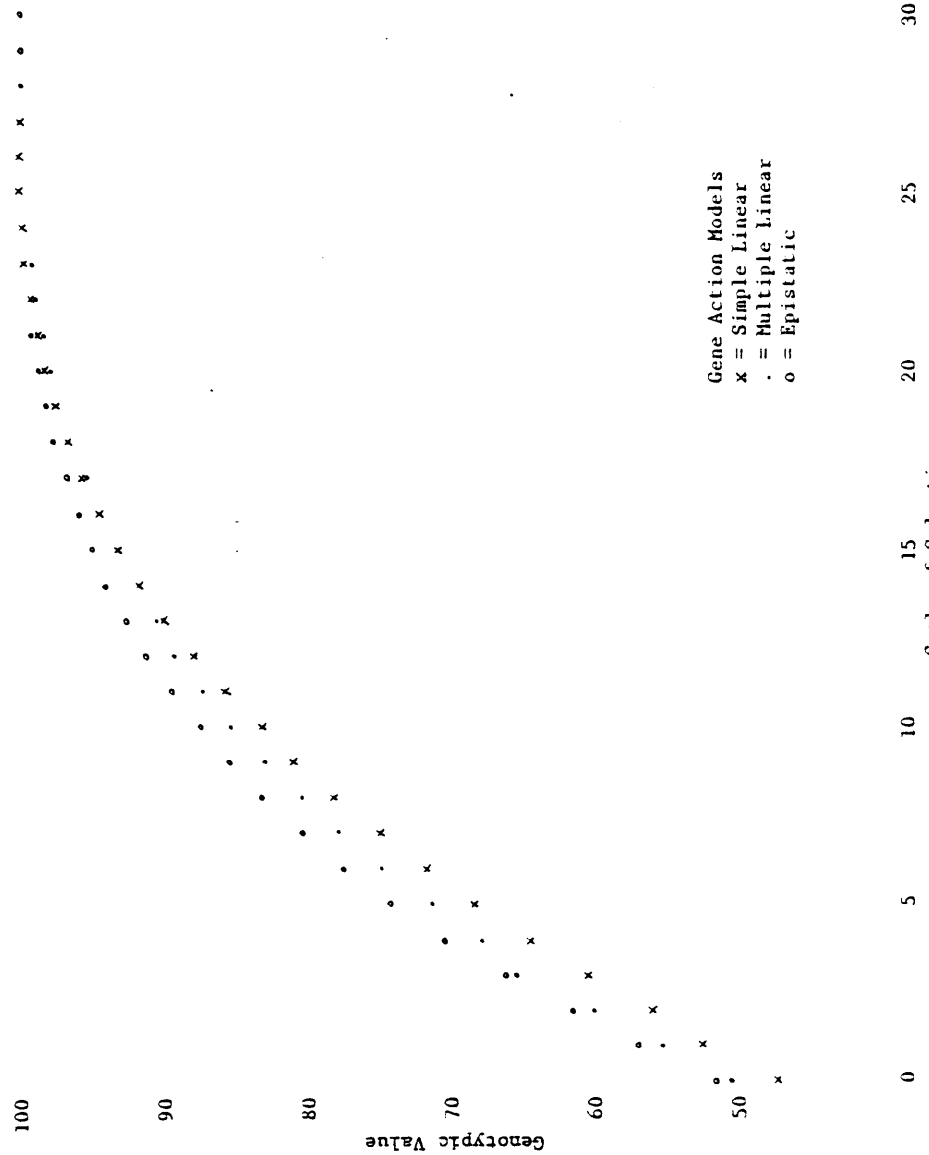


Figure 5.21 Genotypic Value of Selected Crosses with N=20, r=0.2, n_h=100, σ^2_E=0

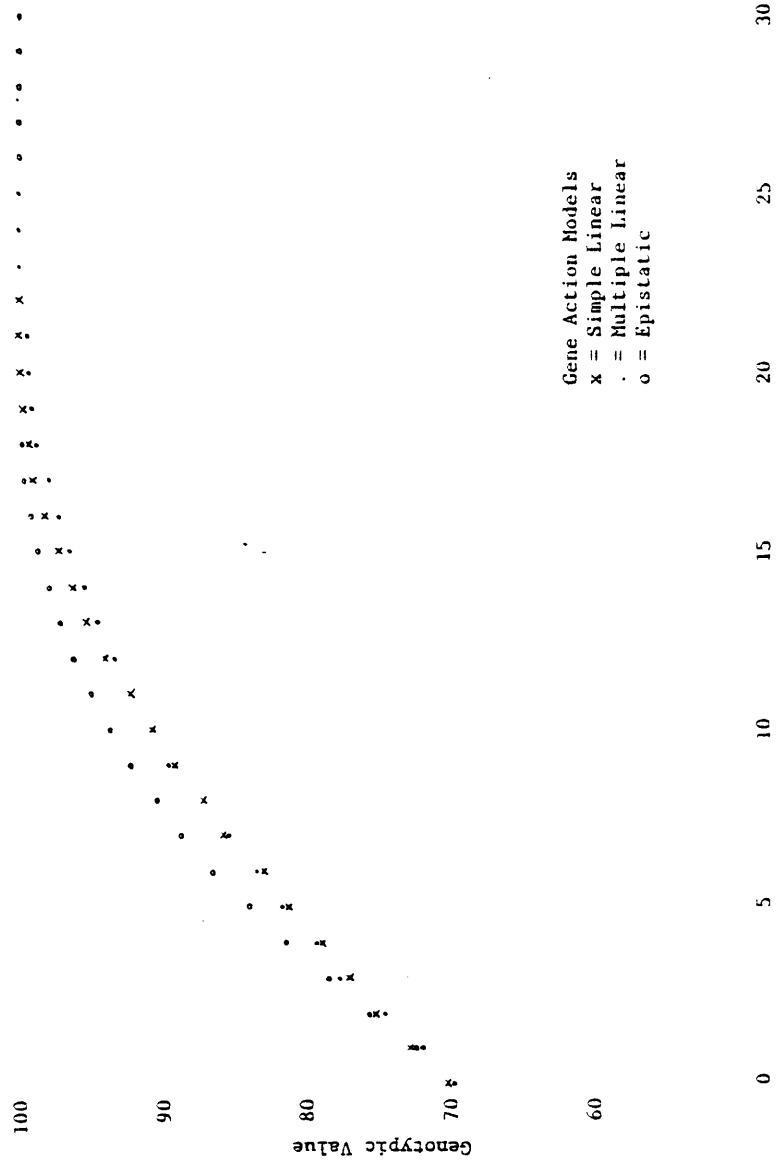


Figure 5.22 Genotypic Value of Selected Crosses with N=20, r=0.05, n_h=50, σ^2_E=0

TABLE 5.24

Regression Coefficients for Genotypic Value Expressed as a Percent of G_{MAX} for Selected Crosses when $\sigma_E^2=60$

No. of Crosses Selected	r	n_h	Gene Action Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$	$\hat{\sigma}^2$	R^2
10	0.2	100	Simple Linear	45.04	2.89	-0.041	0.549	0.125	0.999
			Multiple Linear	47.75	3.07	-0.051	0.390	0.306	0.997
			Epistatic	48.83	2.90	-0.046	0.081	0.159	0.999
	0.05	50	Simple Linear	65.75	1.86	-0.027	0.040	0.109	0.999
			Multiple Linear	68.87	1.76	-0.029	0.322	0.200	0.995
			Epistatic	68.47	1.87	-0.031	0.286	0.127	0.997
20	0.2	50	Simple Linear	45.94	2.19	-0.019	0.395	0.177	0.998
			Multiple Linear	50.02	2.17	-0.025	0.345	0.110	0.999
			Epistatic	49.99	2.85	-0.042	0.274	0.136	0.999
	0.05	100	Simple Linear	67.87	2.84	-0.061	0.802	0.284	0.971
			Multiple Linear	69.09	2.83	-0.062	0.753	0.634	0.949
			Epistatic	69.58	2.85	-0.064	0.792	0.670	0.934

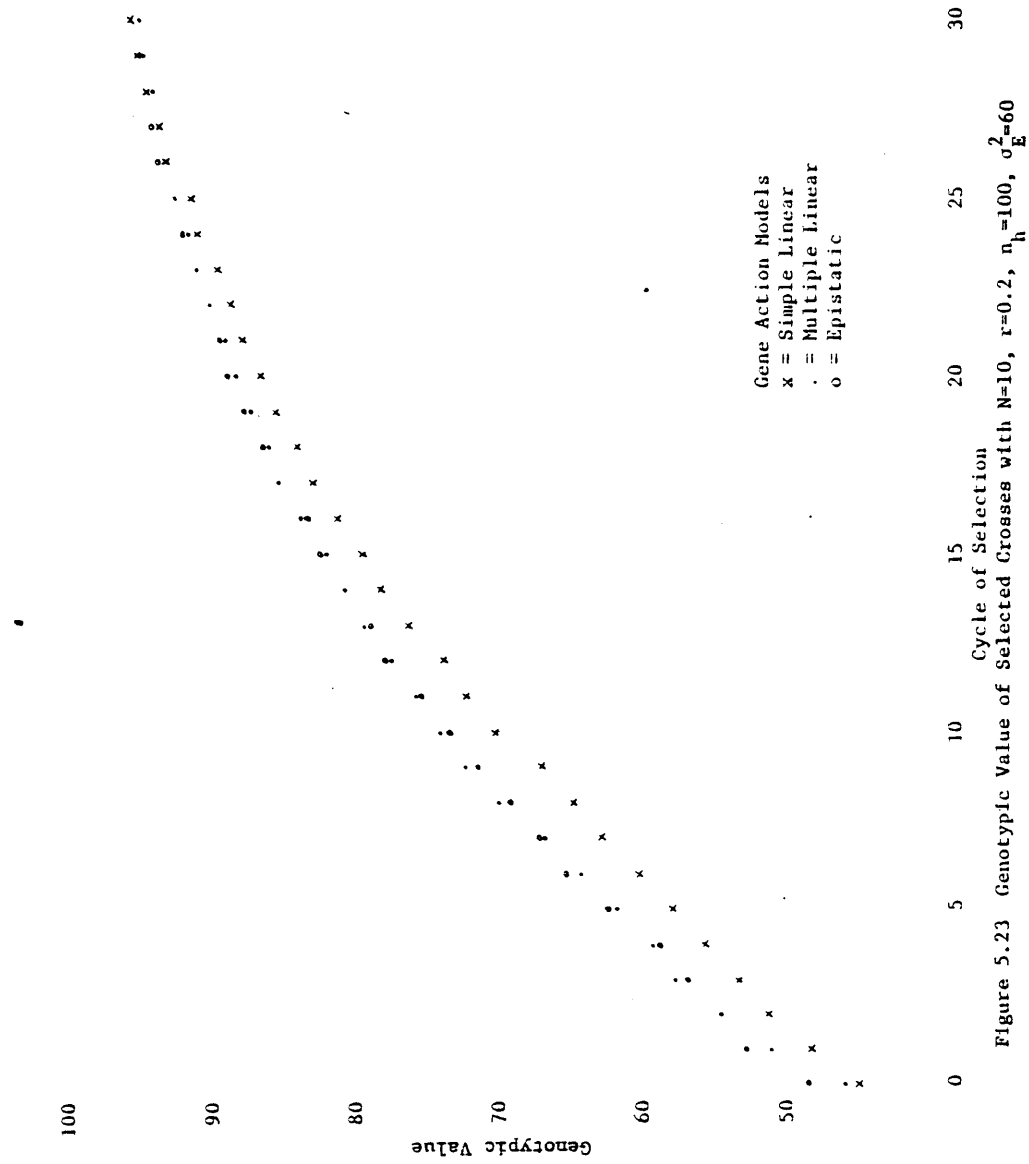


Figure 5.23 Genotypic Value of Selected Crosses with N=10, r=0.2, n_h=100, σ^2_E=60

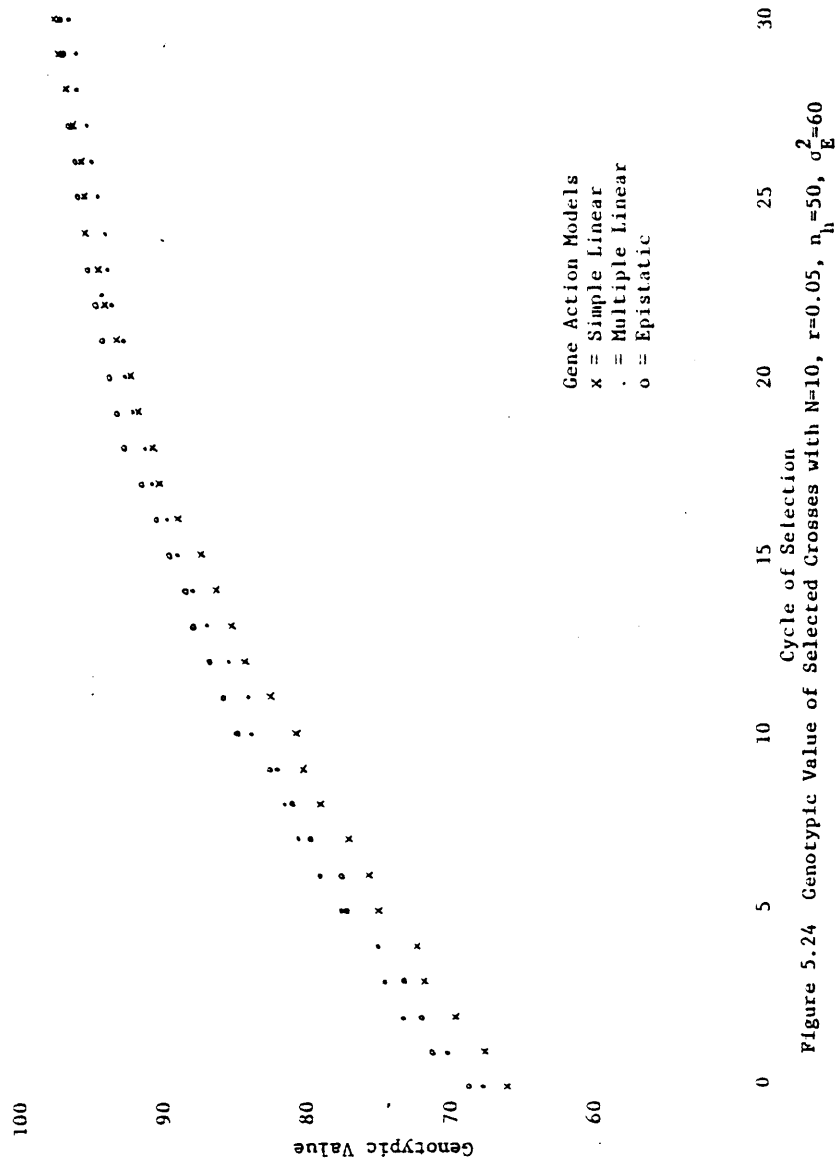


Figure 5.24 Genotypic Value of Selected Crosses with N=10, r=0.05, n_h=50, σ^2_E=60

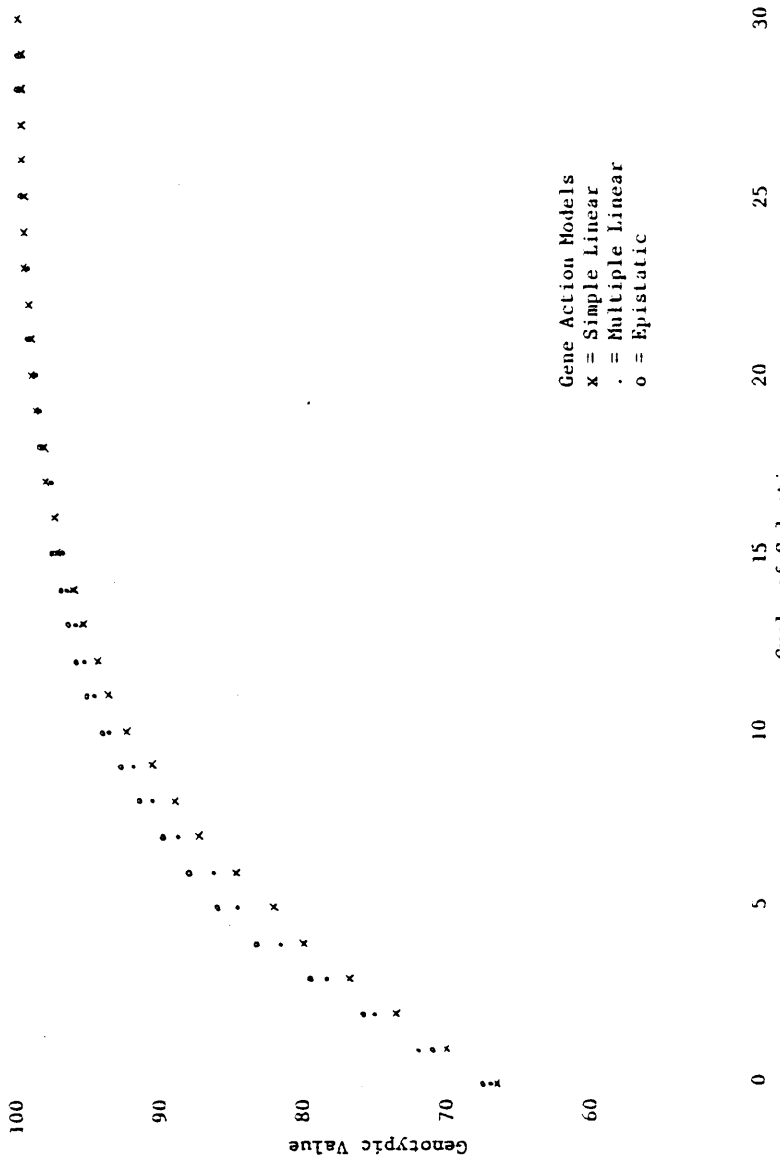


Figure 5.25 Genotypic Value of Selected Crosses with N=20, r=0.2, n_h=50, σ^2_E=60

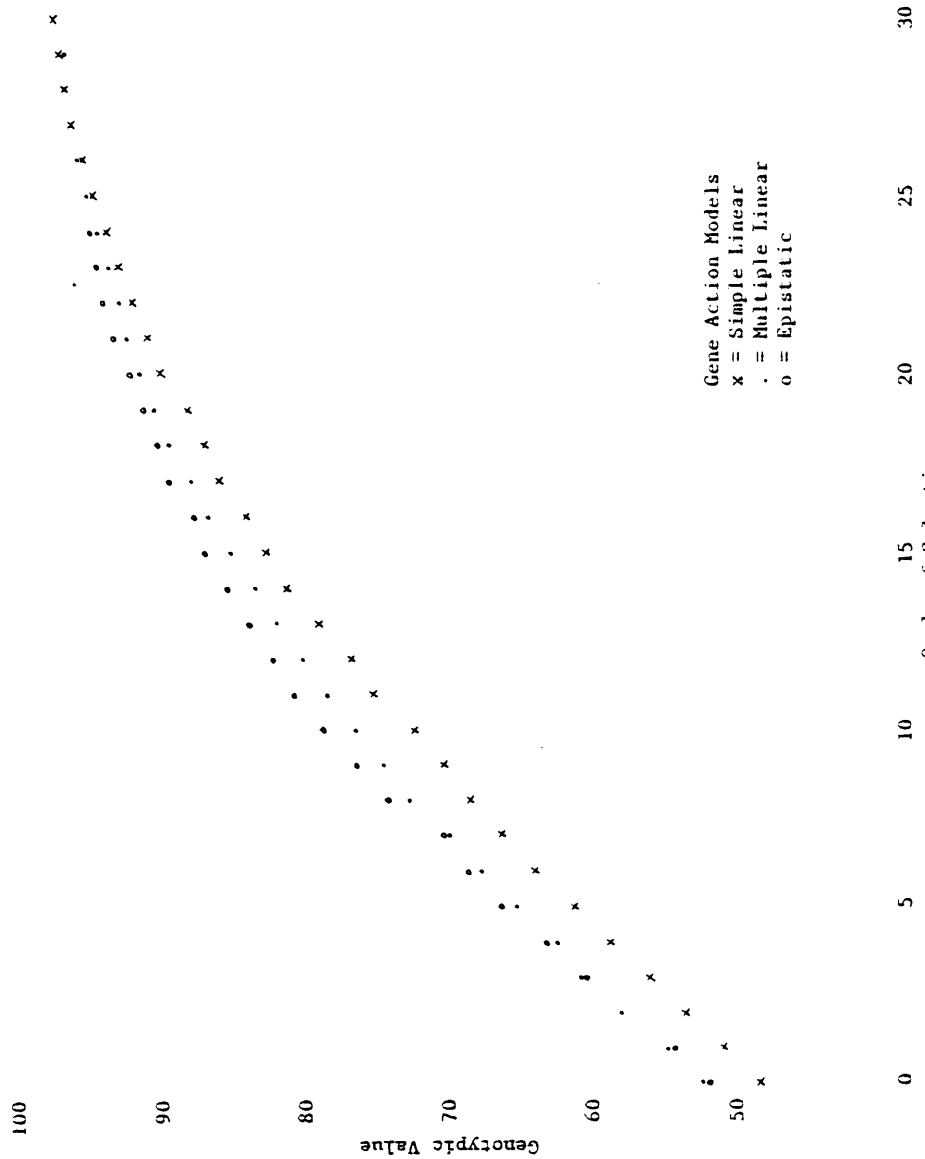


Figure 5.26 Genotypic Value of Selected Crosses with N=20, r=0.05, n_h=100, σ^2_E=60

TABLE 5.25

Regression Coefficients for Genotypic Value Expressed as a Percent of G_{MAX} for All Crosses when $\sigma_E^2=0$

No. of Crosses Selected	r	n_h	Gene Action Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$	$\hat{\sigma}^2$	R^2
10	0.2	50	Simple Linear	56.84	3.52	-0.072	0.223	0.121	0.999
			Multiple Linear	62.26	3.43	-0.079	0.555	0.429	0.990
			Epistatic	62.67	3.27	-0.069	0.553	0.491	0.988
	0.05	100	Simple Linear	38.61	2.54	-0.025	0.235	0.162	0.999
			Multiple Linear	42.07	2.57	-0.032	0.440	0.318	0.996
			Epistatic	41.49	3.27	-0.048	0.295	0.342	0.998
20	0.2	100	Simple Linear	39.57	4.93	-0.101	0.336	0.382	0.998
			Multiple Linear	42.37	4.74	-0.098	0.624	0.554	0.992
			Epistatic	42.99	4.91	-0.103	0.652	1.105	0.983
	0.05	50	Simple Linear	62.32	3.25	-0.069	0.152	0.160	0.999
			Multiple Linear	60.95	3.38	-0.074	0.325	0.142	0.998
			Epistatic	60.28	3.68	-0.081	0.670	0.592	0.980

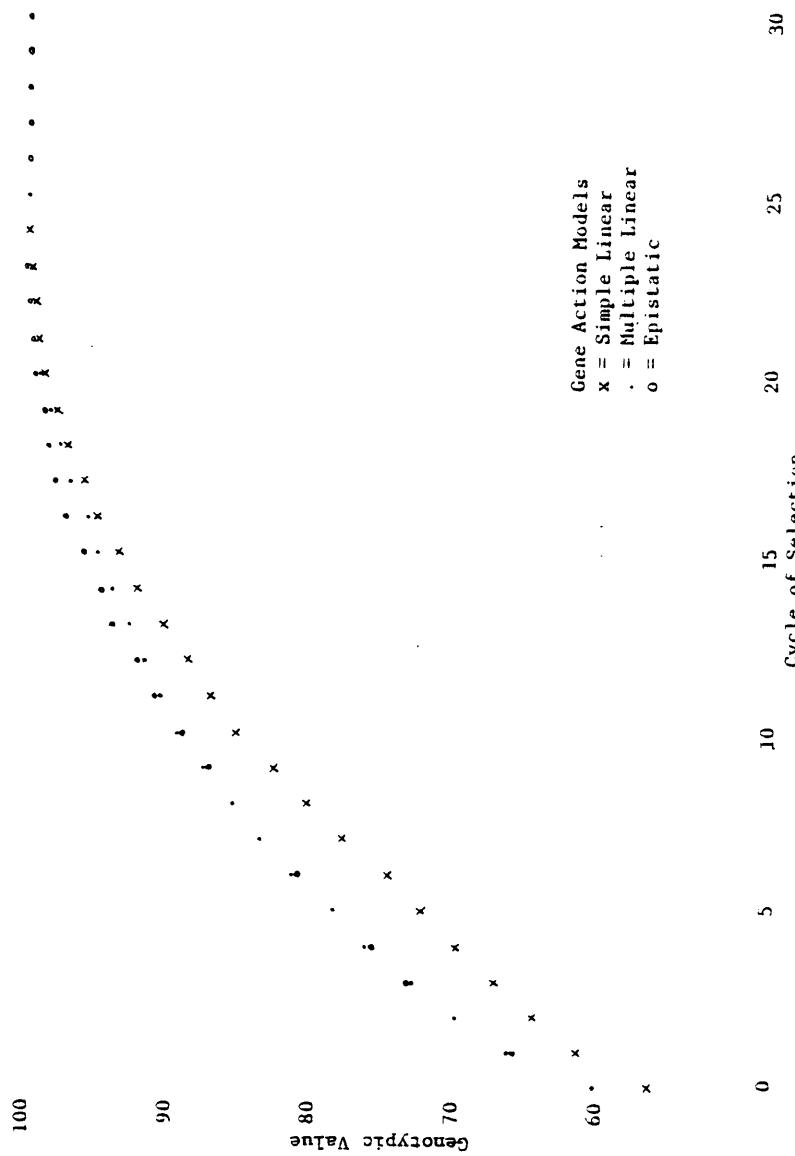


Figure 5.27 Genotypic Value of All Crosses with N=10, r=0.2, n_h=50, σ^2_E=0

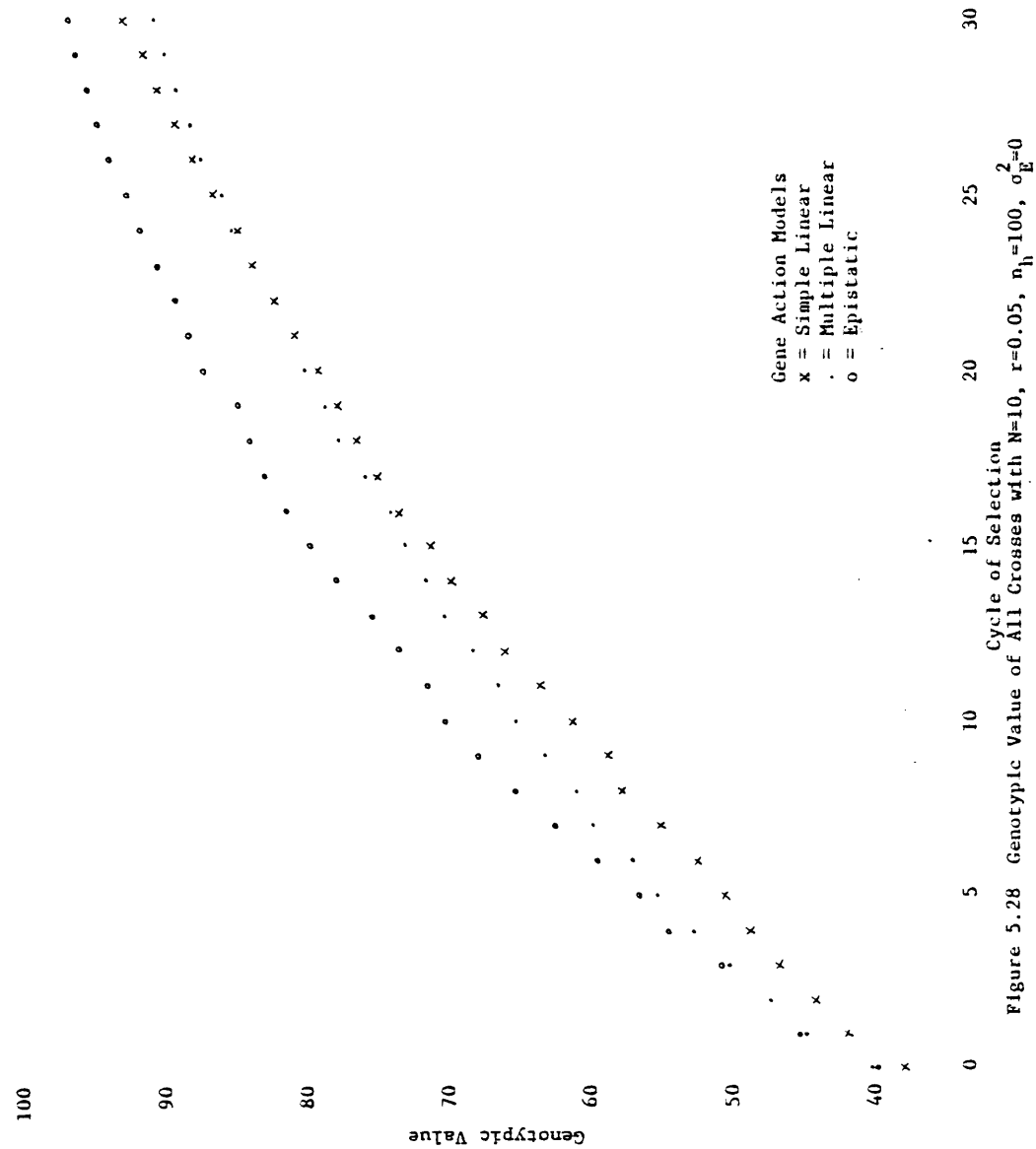


Figure 5.28 Genotypic Value of All Crosses with N=10, r=0.05, n_h=100, σ^2_E=0

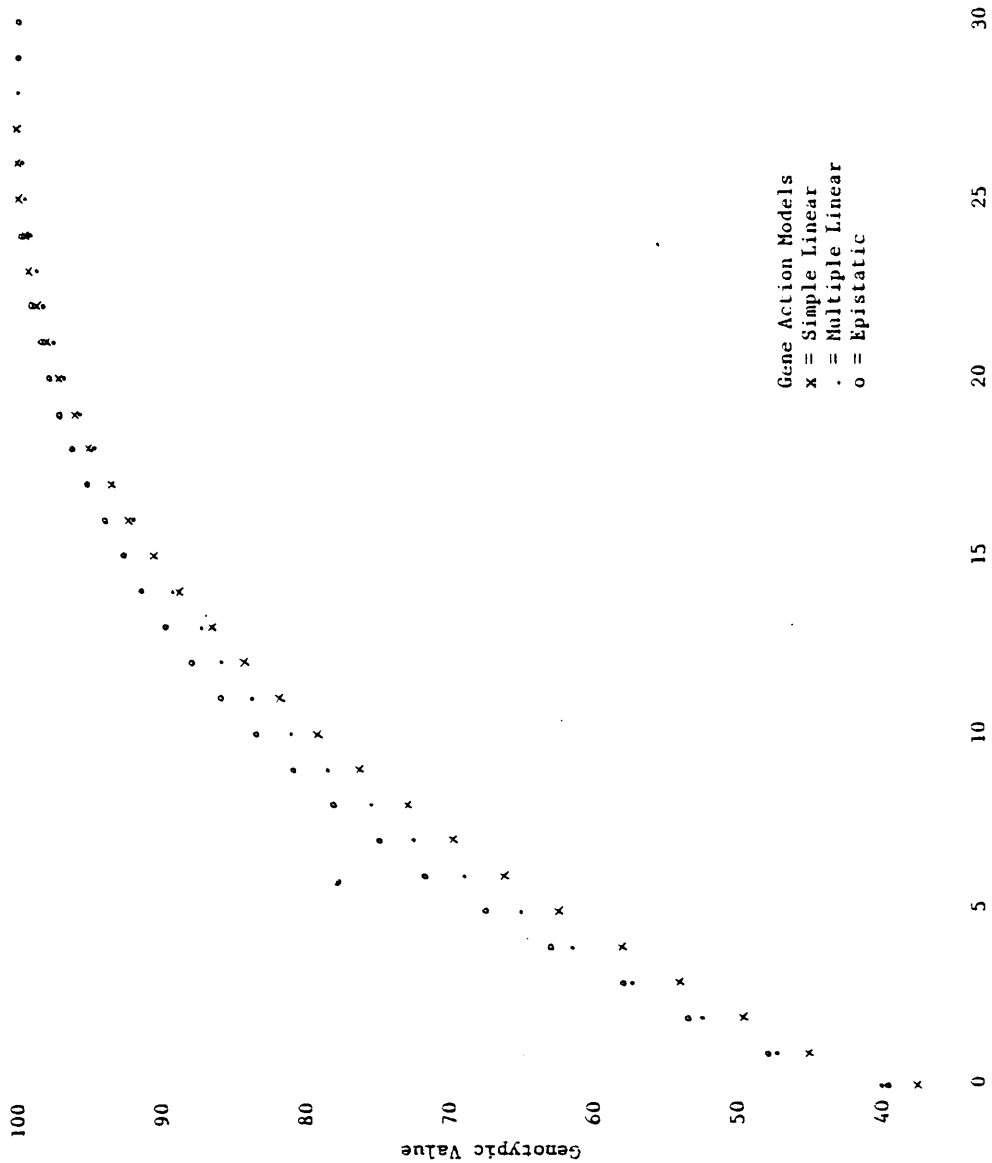


Figure 5.29 Genotypic Value of All Crosses with N=20, r=0.2, n_h=100, σ^2_E=0

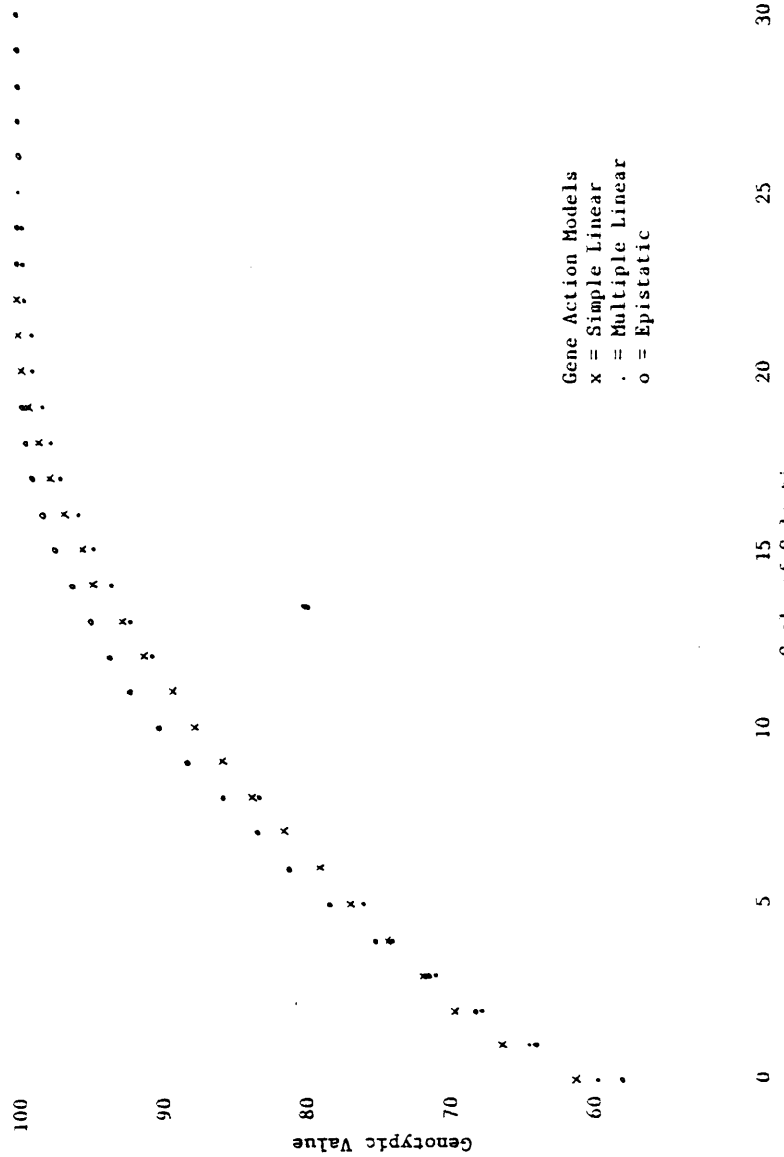


Figure 5.30 Genotypic Value of All Crosses with N=20, r=0.05, n_h=50, σ^2_E=0

TABLE 5.26

Regression Coefficients for Genotypic Value Expressed as a Percent of G_{MAX} for All Crosses when $\sigma_E^2=60$

No. of Crosses Selected	r	n_h	Gene Action Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$	$\hat{\sigma}^2$	R^2
10	0.2	100	Simple Linear	38.65	3.18	-0.045	0.389	0.214	0.998
			Multiple Linear	41.46	3.33	-0.053	0.252	0.360	0.998
			Epistatic	41.33	3.26	-0.051	-0.030	0.414	0.999
	0.05	50	Simple Linear	60.38	2.15	-0.031	0.215	0.125	0.999
			Multiple Linear	61.94	2.25	-0.038	0.424	0.456	0.990
			Epistatic	61.85	2.27	-0.037	-0.080	0.296	0.998
20	0.2	50	Simple Linear	60.74	3.44	-0.073	0.748	0.476	0.976
			Multiple Linear	61.02	3.53	-0.077	0.738	0.890	0.957
			Epistatic	60.03	3.76	-0.084	0.724	1.499	0.938
	0.05	100	Simple Linear	39.02	3.46	-0.051	0.240	0.124	0.999
			Multiple Linear	42.37	3.52	-0.058	0.286	0.260	0.998
			Epistatic	40.90	3.91	-0.070	0.561	0.348	0.996

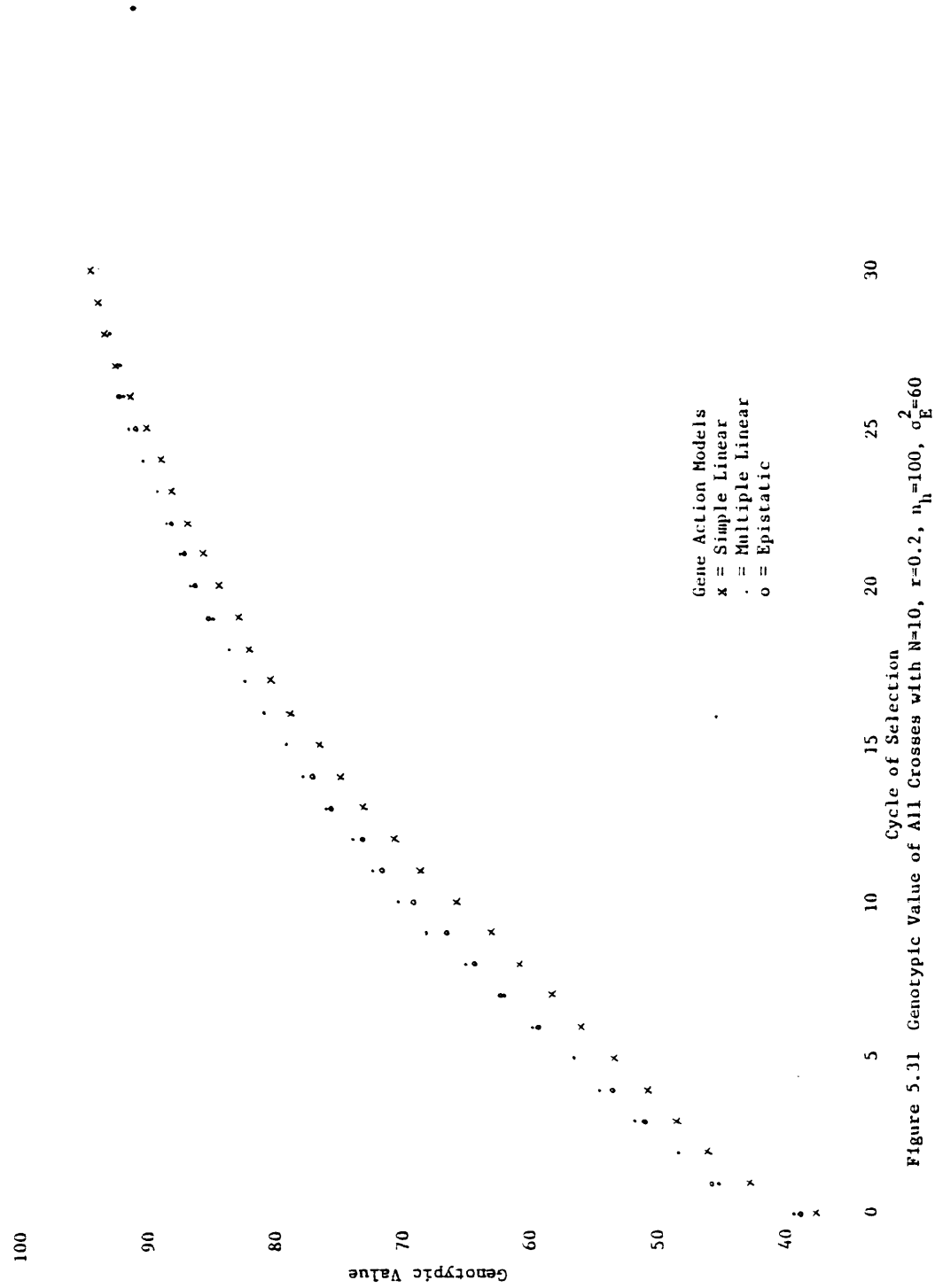


Figure 5.31 Genotypic Value of All Crosses with N=10, r=0.2, n_h=100, σ^2_E=60

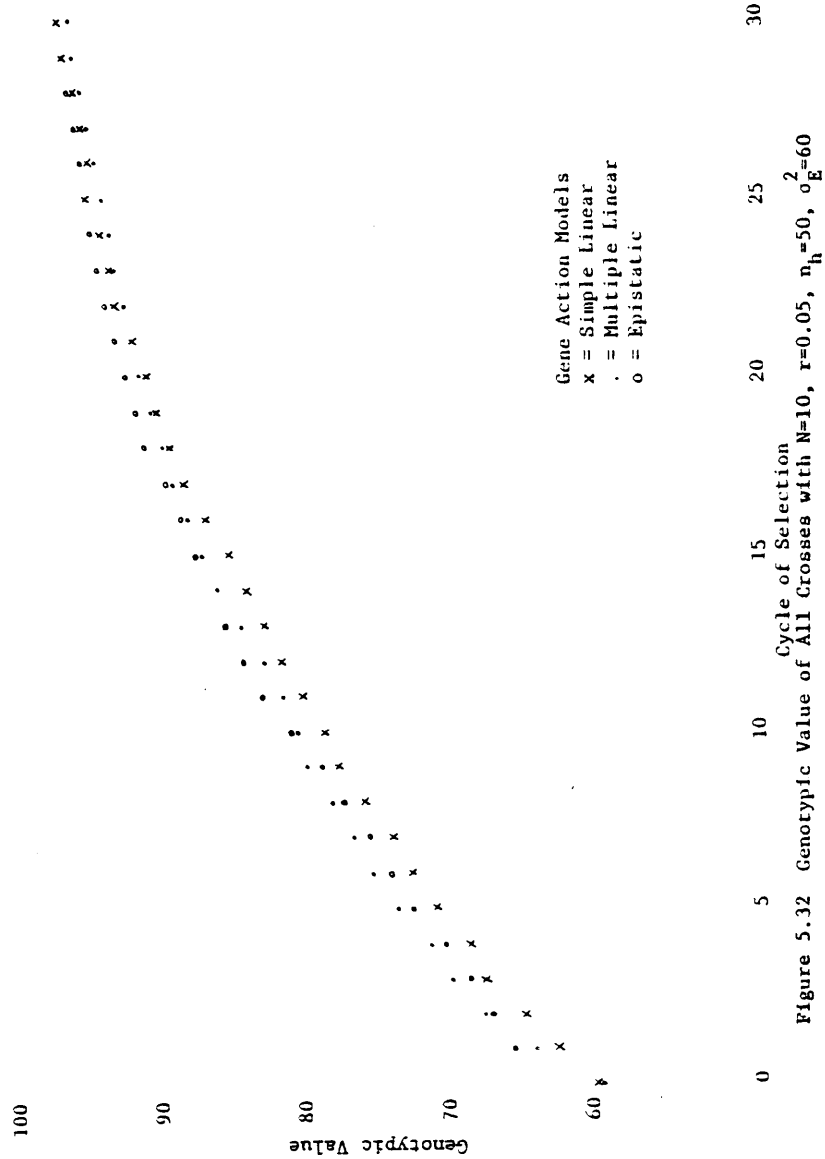


Figure 5.32 Genotypic Value of All Crosses with N=10, r=0.05, n_h=50, σ^2_E=60

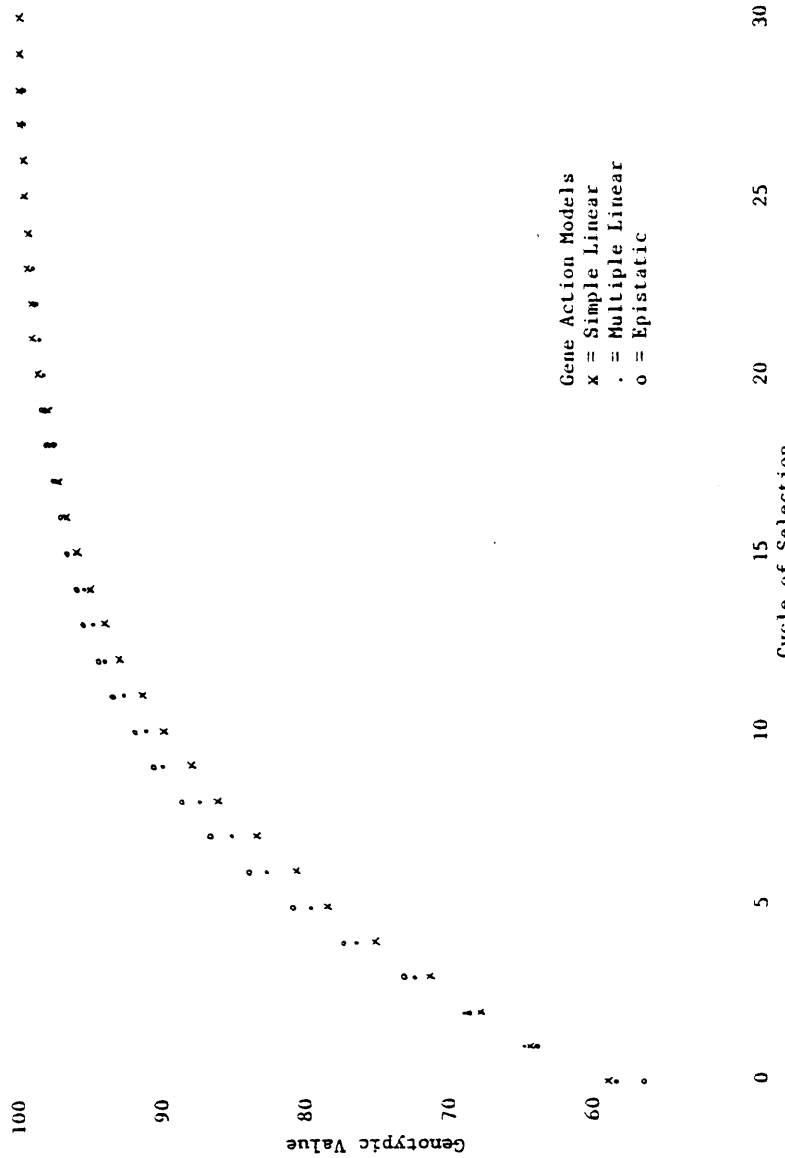


Figure 5.33 Genotypic Value of All Crosses with N=20, r=0.2, n_h=50, σ^2_E=60

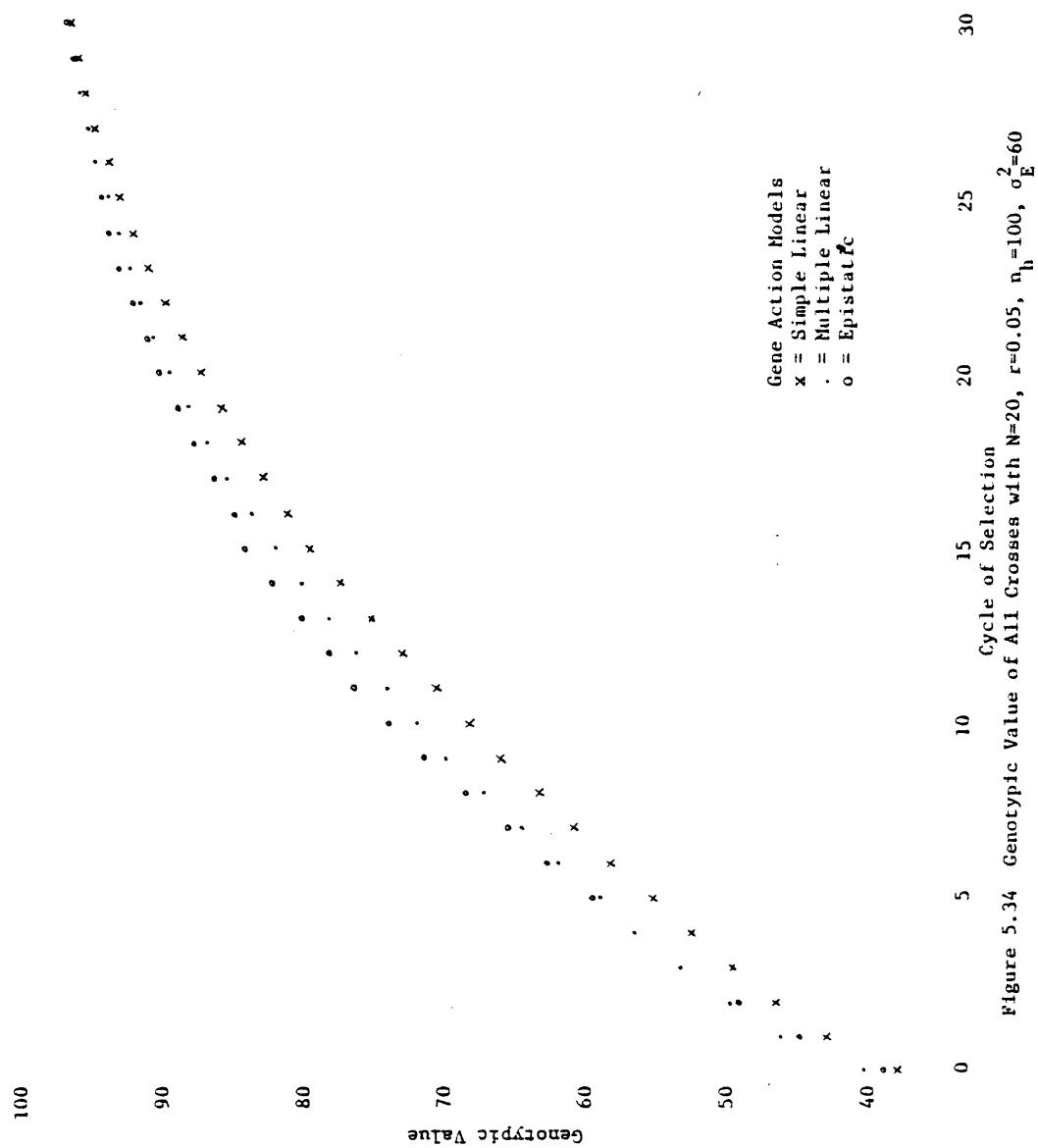


Figure 5.34 Genotypic Value of All Crosses with N=20, r=0.05, n_h=100, σ^2_E=60

TABLE 5.27

Regression Coefficients for Yield Expressed as a Percent of G_{MAX} for Selected Crosses when $\sigma_E^2=60$

No. of Crosses Selected	r	n_h	Gene Action Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$	$\hat{\sigma}^2$	R^2
10	0.2	100	Simple Linear	45.69	2.90	-0.040	0.625	0.126	0.998
			Multiple Linear	48.27	3.07	-0.050	0.325	0.356	0.997
			Epistatic	49.24	2.92	-0.045	0.238	0.145	0.999
	0.05	50	Simple Linear	66.81	1.89	-0.026	-0.035	0.120	0.999
			Multiple Linear	69.33	1.85	-0.030	0.411	0.260	0.992
			Epistatic	69.10	1.94	-0.031	0.342	0.187	0.996
20	0.2	50	Simple Linear	69.26	3.01	-0.064	0.767	0.351	0.974
			Multiple Linear	70.09	2.96	-0.064	0.755	0.588	0.957
			Epistatic	70.81	3.02	-0.067	0.782	0.742	0.939
	0.05	100	Simple Linear	48.52	3.00	-0.043	0.416	0.055	0.999
			Multiple Linear	52.42	2.98	-0.048	0.251	0.035	0.999
			Epistatic	52.42	3.20	-0.055	0.681	0.108	0.997

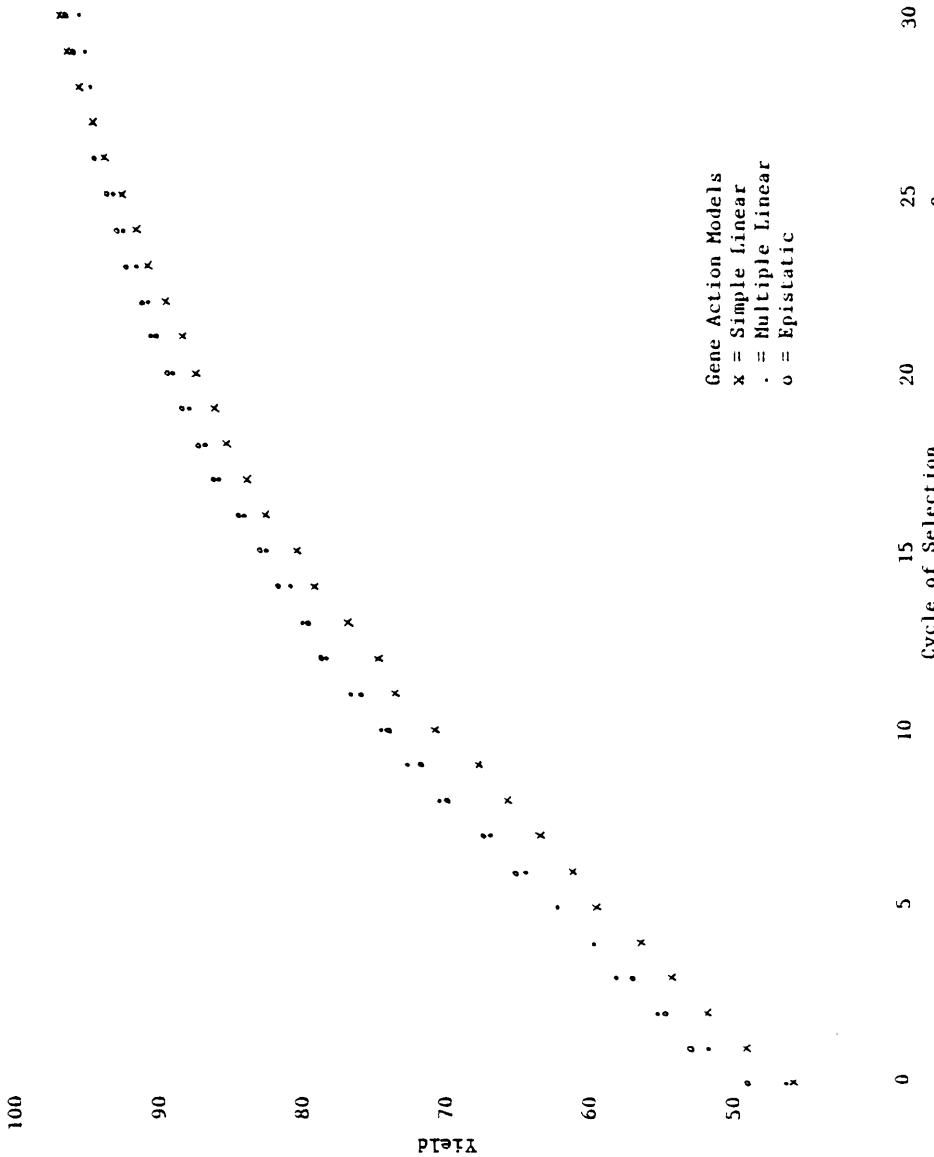


Figure 5.35 Yield of Selected Crosses with N=10, r=0.2, u_h=100, σ^2_E=60

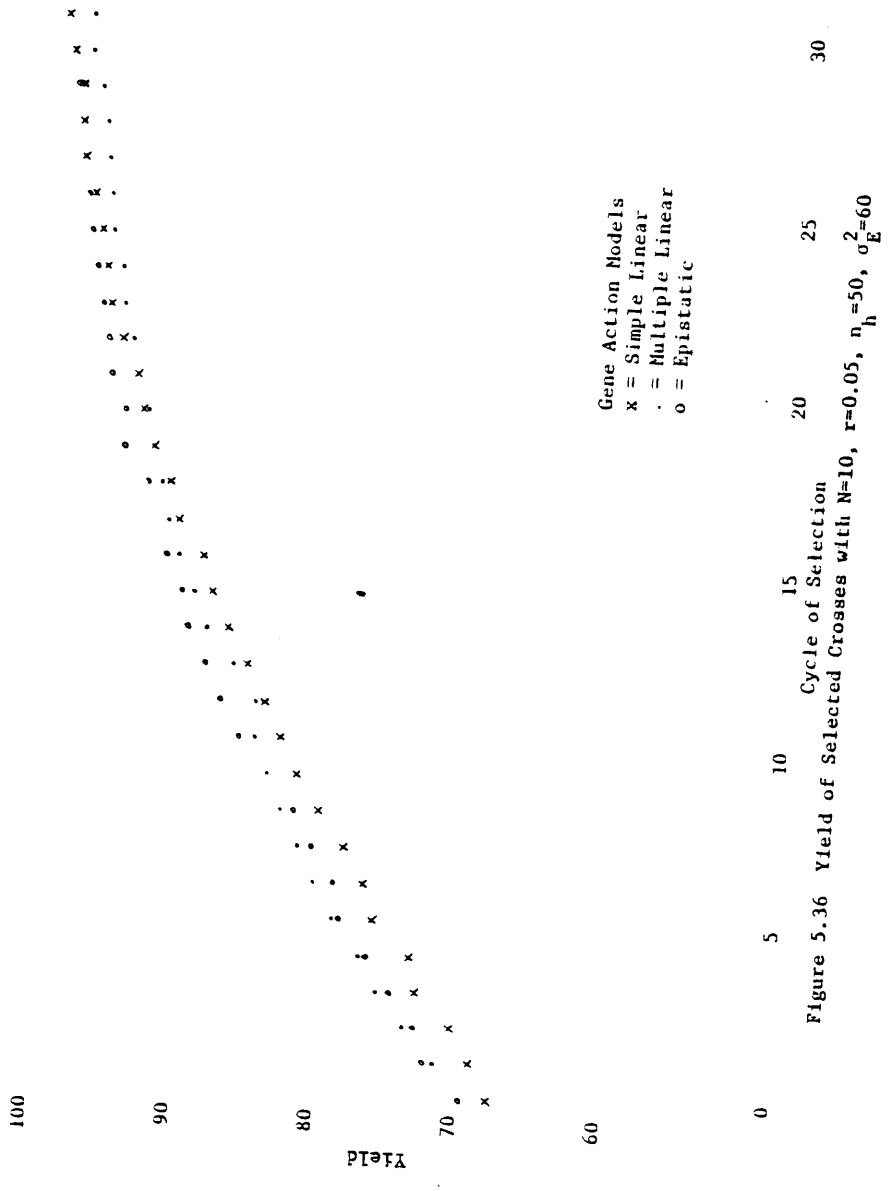


Figure 5.36 Yield of Selected Crosses with N=10, r=0.05, n_h=50, σ^2_E=60

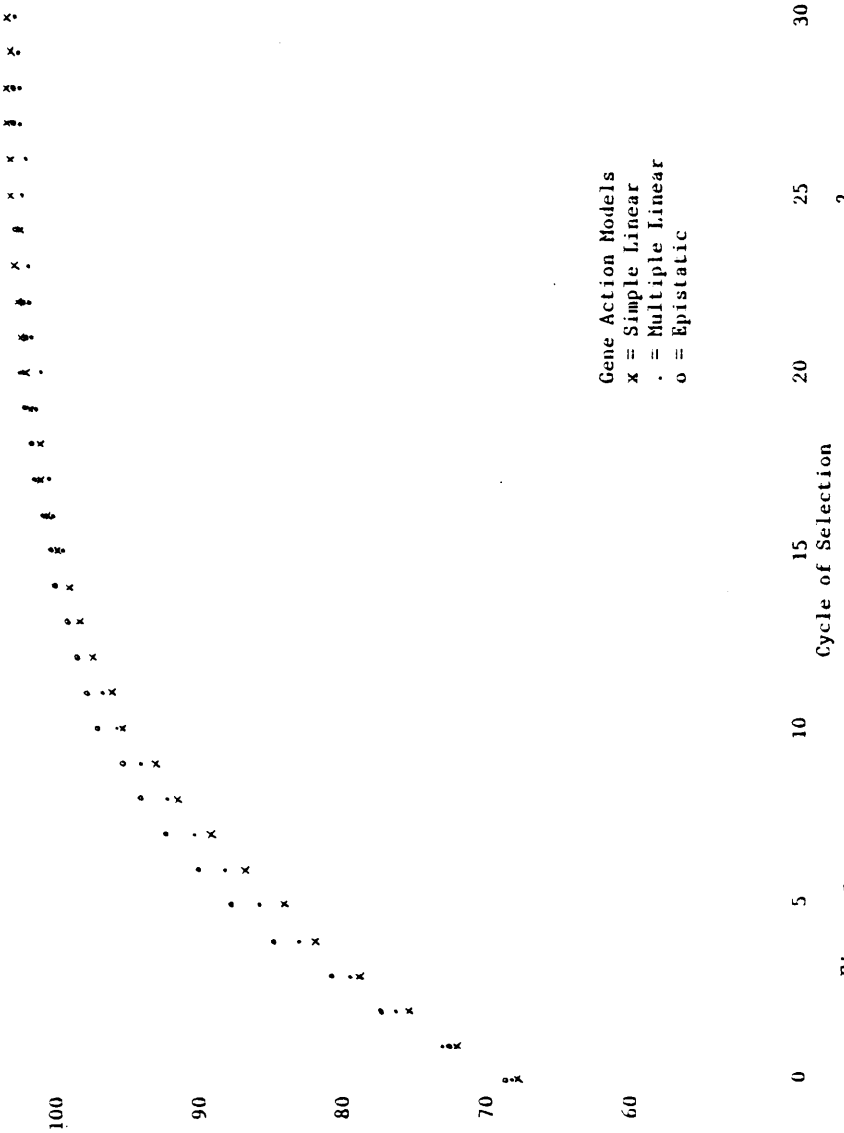


Figure 5.37 Yield of Selected Crosses with N=20, r=0.2, n_h=50, σ^2_E=60

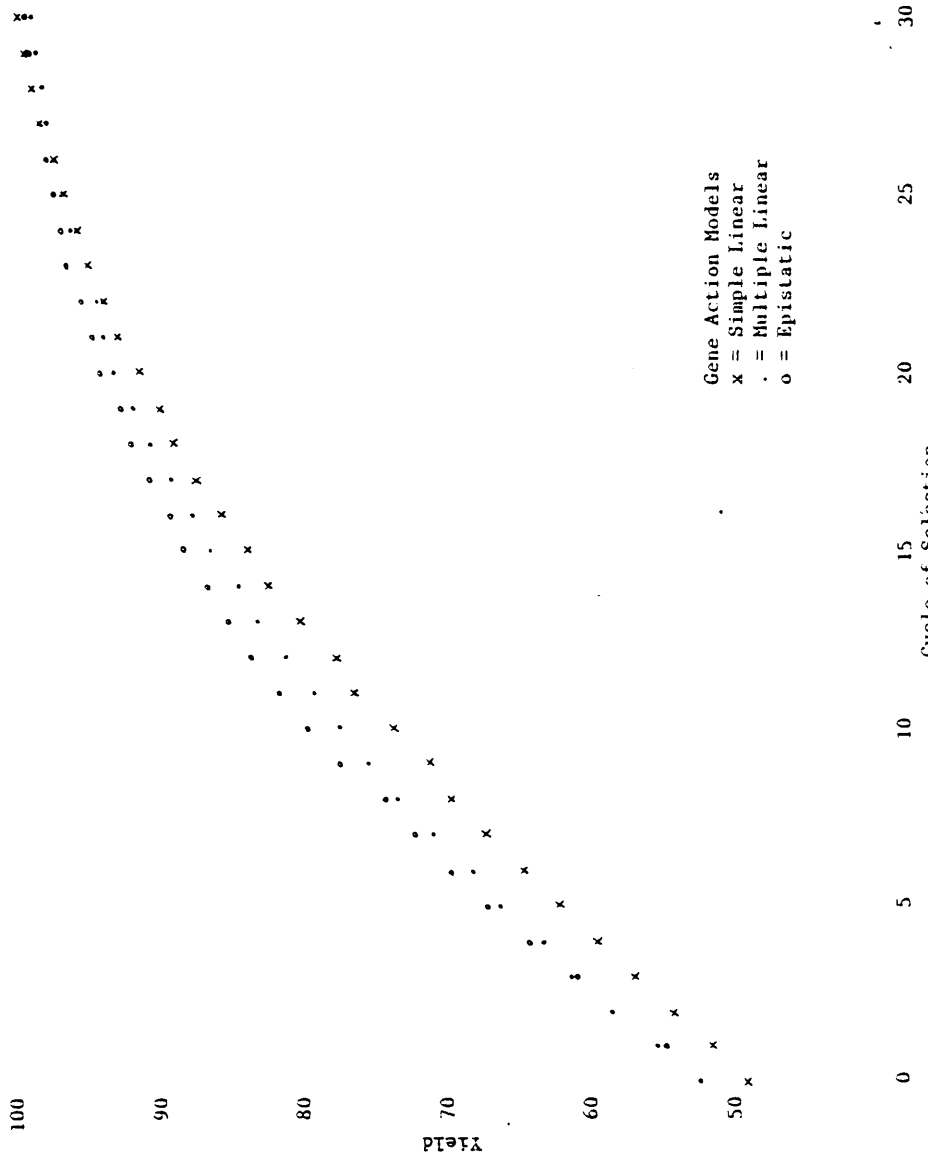


Figure 5.38 Yield of Selected Crosses with N=20, r=0.05, n_h=100, σ^2_E=60

TABLE 5.28

Regression Coefficients for Yield Expressed as a Percent of G_{MAX} for All Crosses when $\sigma_E^2=60$

No. of Crosses Selected	r	n_h	Gene Action Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$	$\hat{\sigma}^2$	R^2
10	0.2	100	Simple Linear	38.69	3.18	-0.044	0.432	0.193	0.998
			Multiple Linear	41.48	3.32	-0.053	0.203	0.394	0.998
			Epistatic	41.21	3.27	-0.051	-0.047	0.370	0.999
	0.05	50	Simple Linear	60.29	2.16	-0.031	0.250	0.115	0.999
			Multiple Linear	61.94	2.24	-0.038	0.407	0.502	0.989
			Epistatic	61.87	2.27	-0.037	-0.059	0.385	0.997
20	0.2	50	Simple Linear	60.75	3.44	-0.073	0.761	0.439	0.976
			Multiple Linear	60.97	3.54	-0.077	0.737	0.890	0.957
			Epistatic	60.04	3.75	-0.084	0.721	1.565	0.937
	0.05	100	Simple Linear	39.02	3.45	-0.051	0.208	0.123	0.999
			Multiple Linear	42.37	3.53	-0.058	0.277	0.261	0.999
			Epistatic	40.86	3.92	-0.070	0.561	0.350	0.996

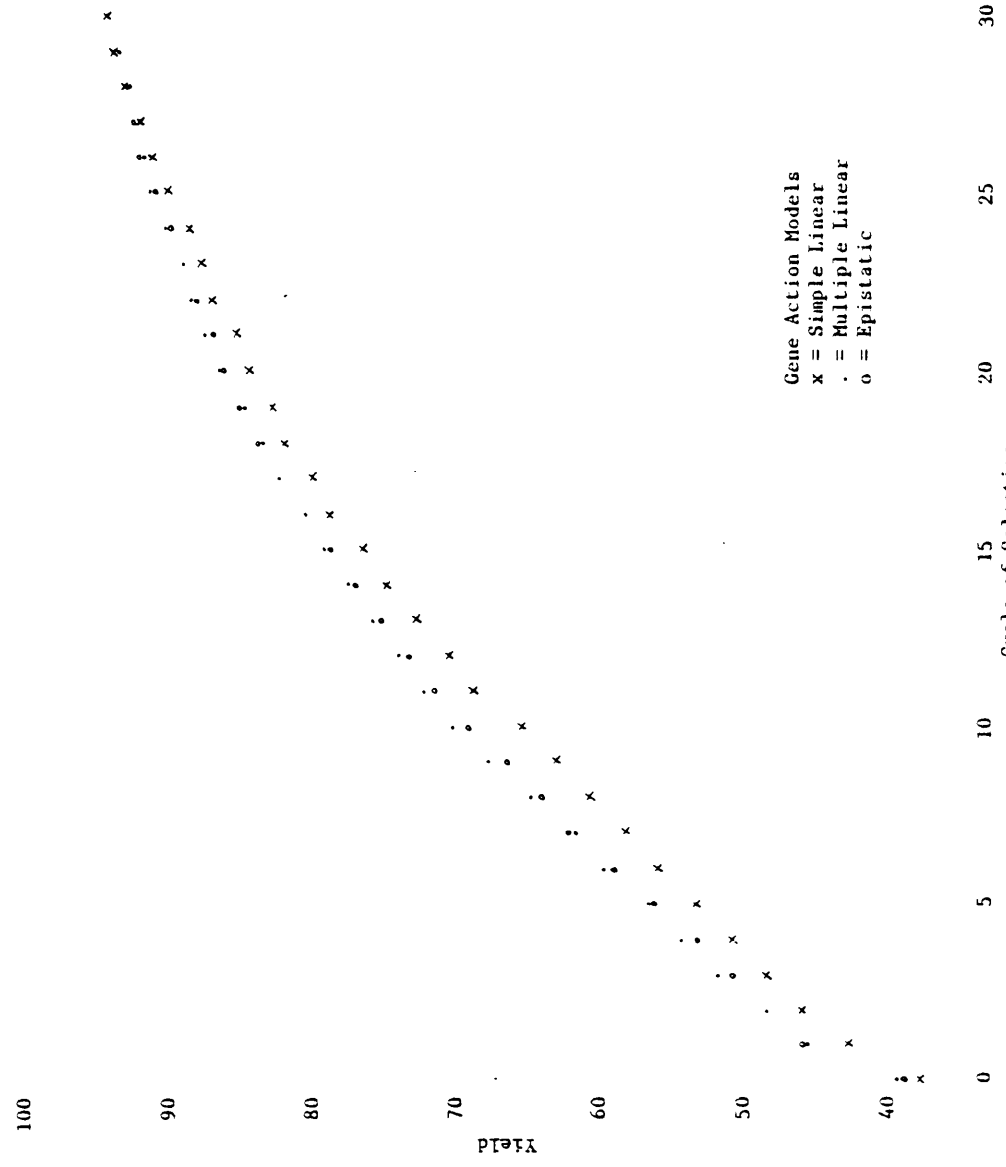


Figure 5.39 Yield of All Crosses with N=10, r=0.2, n_h=100, σ^2_E=60

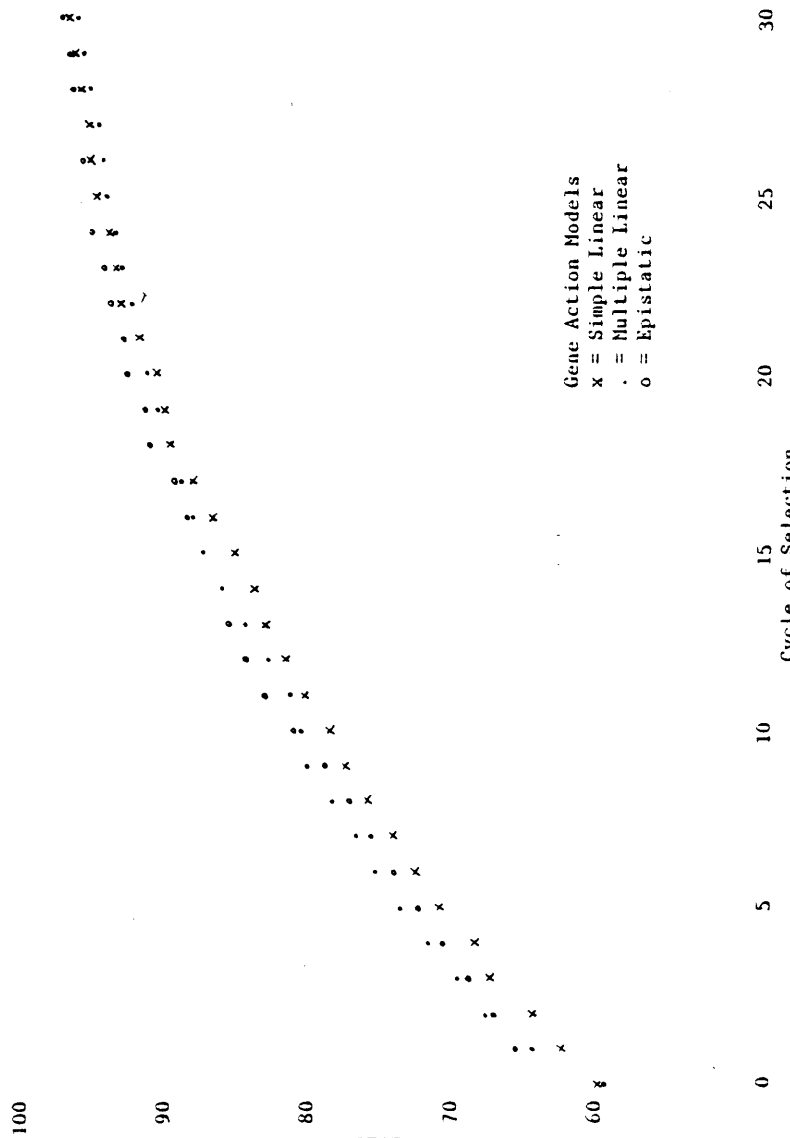


Figure 5.40 Yield of All Crosses with N=10, r=0.05, n_h=50, σ^2_E=60

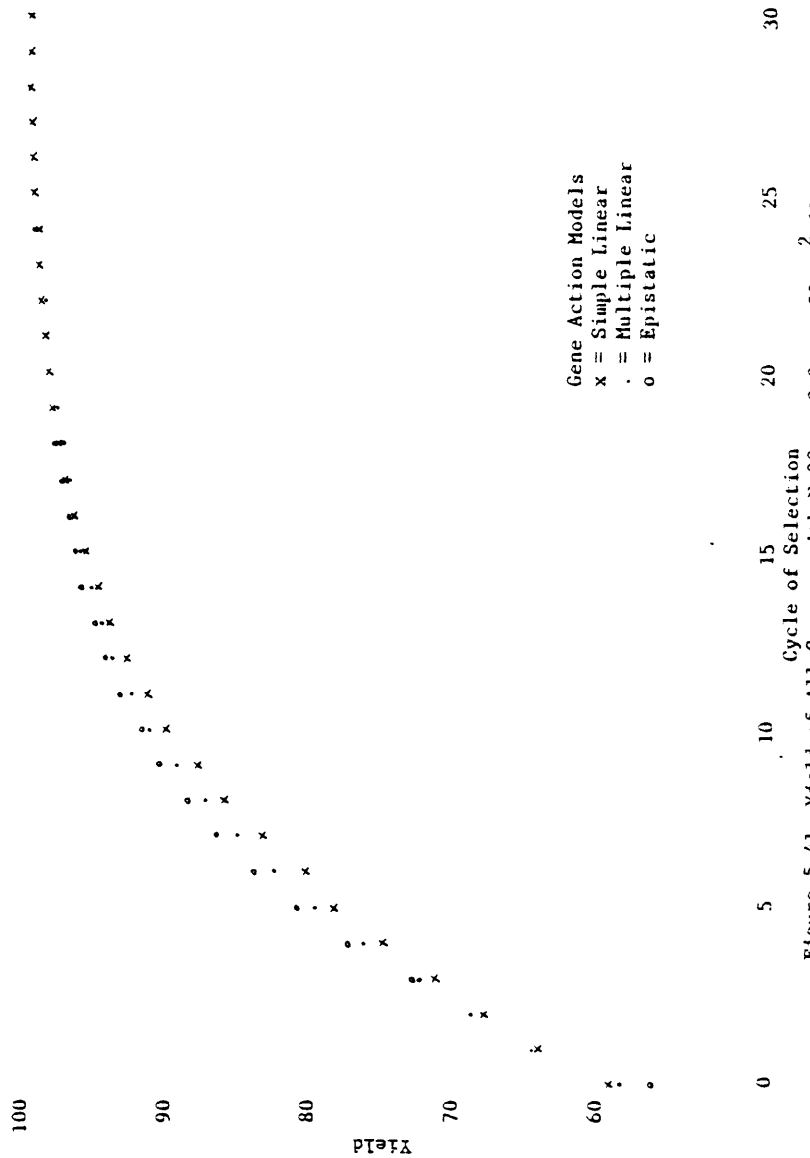


Figure 5.41 Yield of All Crosses with N=20, r=0.2, n_h=50, σ^2_E=60

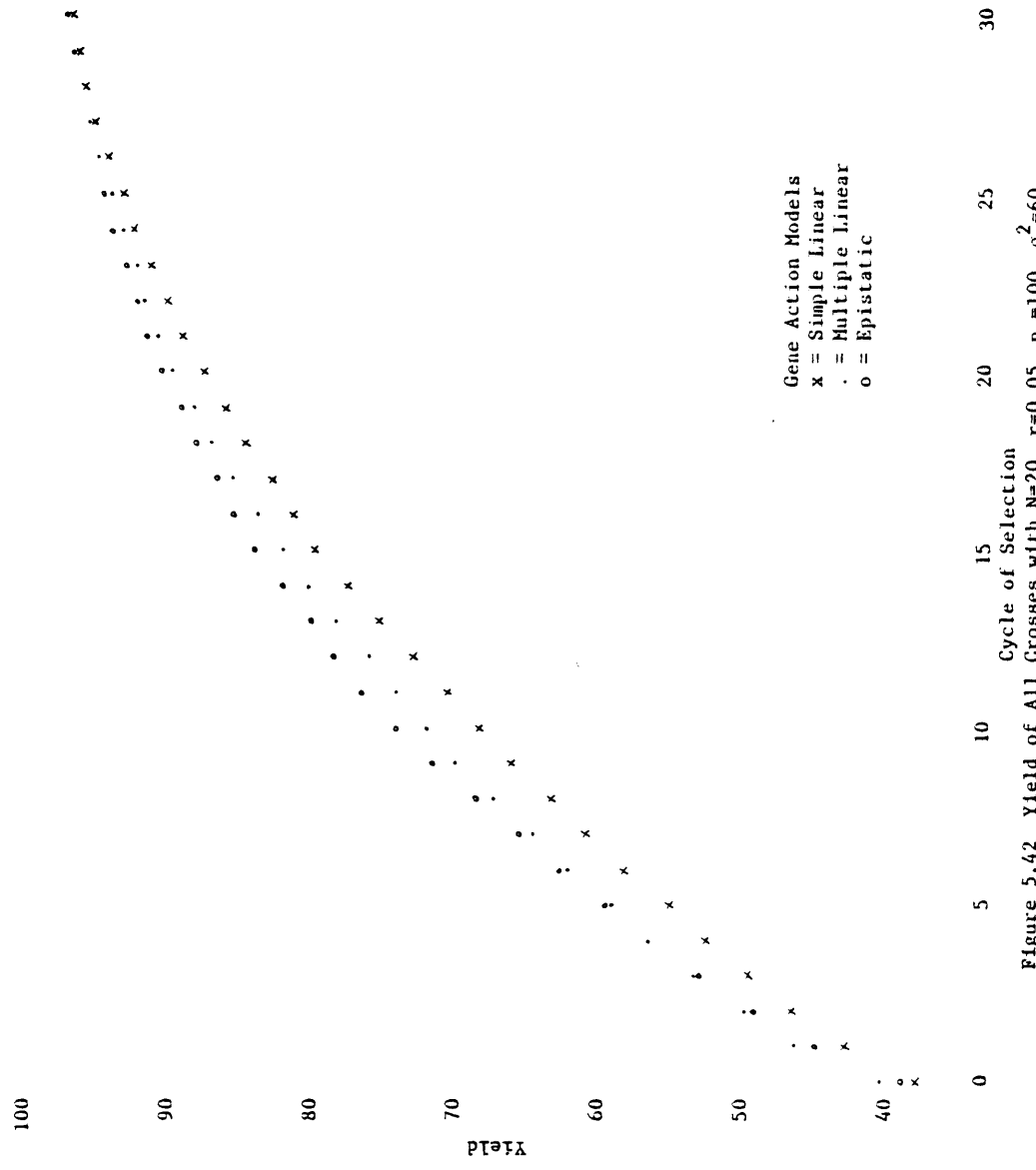


Figure 5.42 Yield of All Crosses with N=20, r=0.05, n_h=100, σ^2_E=60

TABLE 5.29

Regression Coefficients for h_B^2 for All Crosses when $\sigma_E^2=60$

No. of Crosses Selected	r	n_h	Gene Action Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$	$\hat{\sigma}^2$	R^2
10	0.2	100	Simple Linear	0.509	-1.94	1.64	0.080	2.01	0.988
			Multiple Linear	0.679	-2.48	1.31	0.447	2.23	0.984
			Epistatic	0.655	-2.48	1.87	-0.134	1.64	0.996
	0.05	50	Simple Linear	0.381	-2.29	3.92	0.452	1.53	0.965
			Multiple Linear	0.571	-3.92	7.26	0.539	2.32	0.972
			Epistatic	0.530	-3.45	6.06	0.285	3.16	0.978
20	0.2	50	Simple Linear	0.334	-2.81	5.85	0.369	0.53	0.990
			Multiple Linear	0.495	-4.32	9.14	0.730	0.90	0.974
			Epistatic	0.476	-4.32	9.44	0.739	1.27	0.959
	0.05	100	Simple Linear	0.607	-3.44	5.29	0.512	1.50	0.984
			Multiple Linear	0.760	-4.06	5.46	0.611	1.05	0.990
			Epistatic	0.742	-4.43	7.02	0.247	1.11	0.996

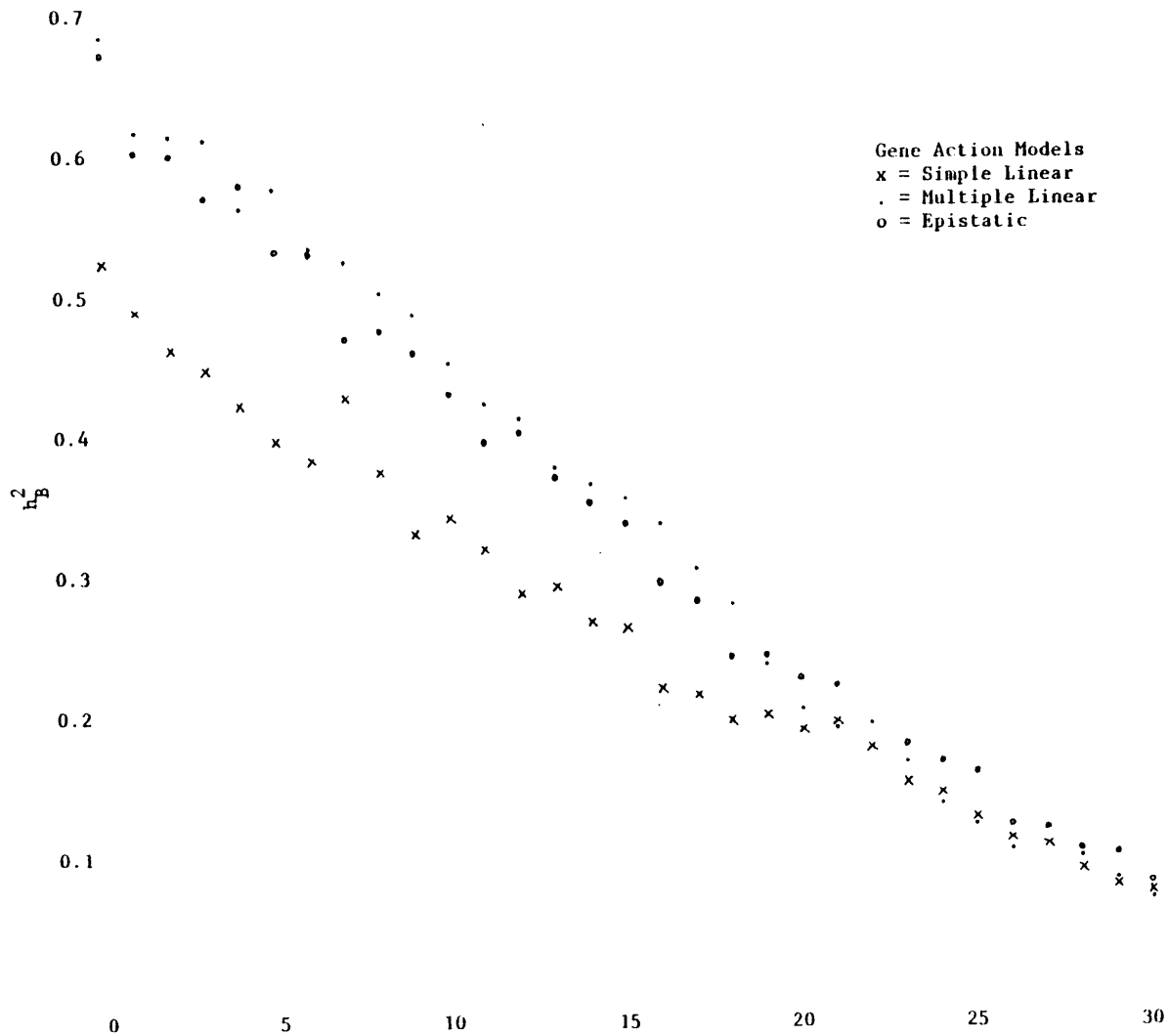


Figure 5.43 h_B^2 of All Crosses with N=10, r=0.2, n_h=100, σ_E^2=60

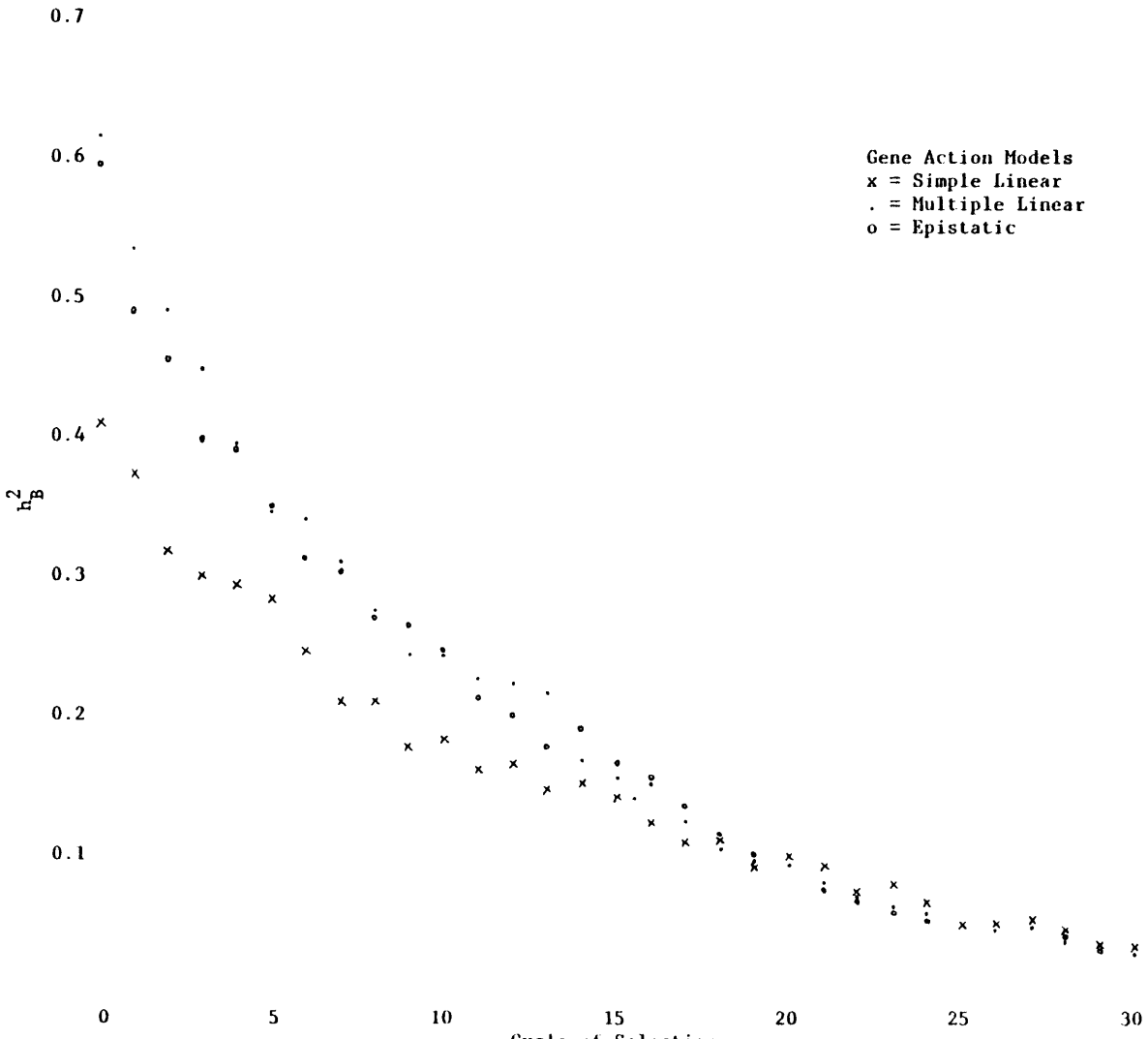


Figure 5.44 h_B^2 of All Crosses with N=10, r=0.05, n_h=50, o_E^2=60

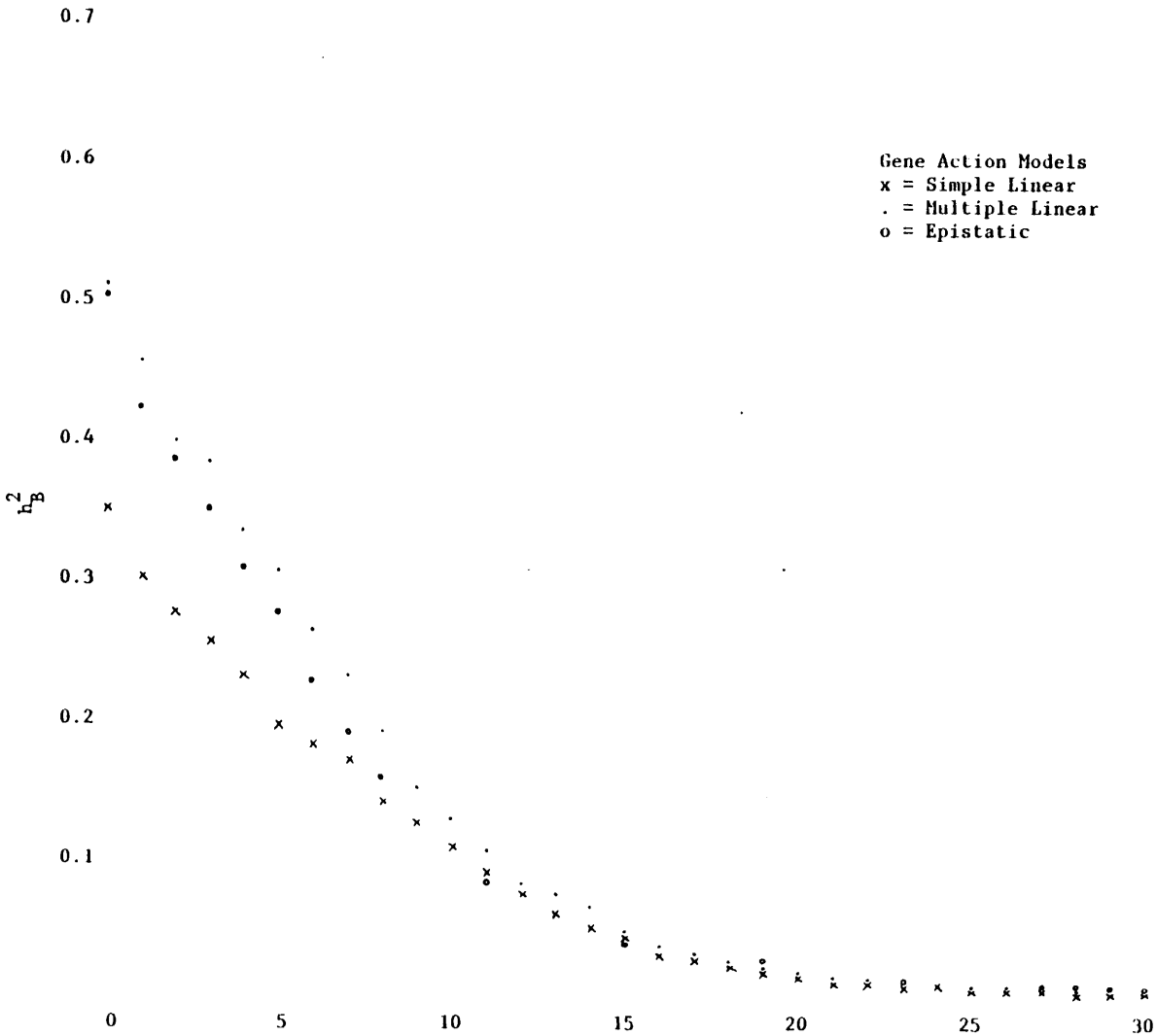


Figure 5.45 h_B^2 of All Crosses with N=20, r=0.2, n_h=50, σ_E^2=60

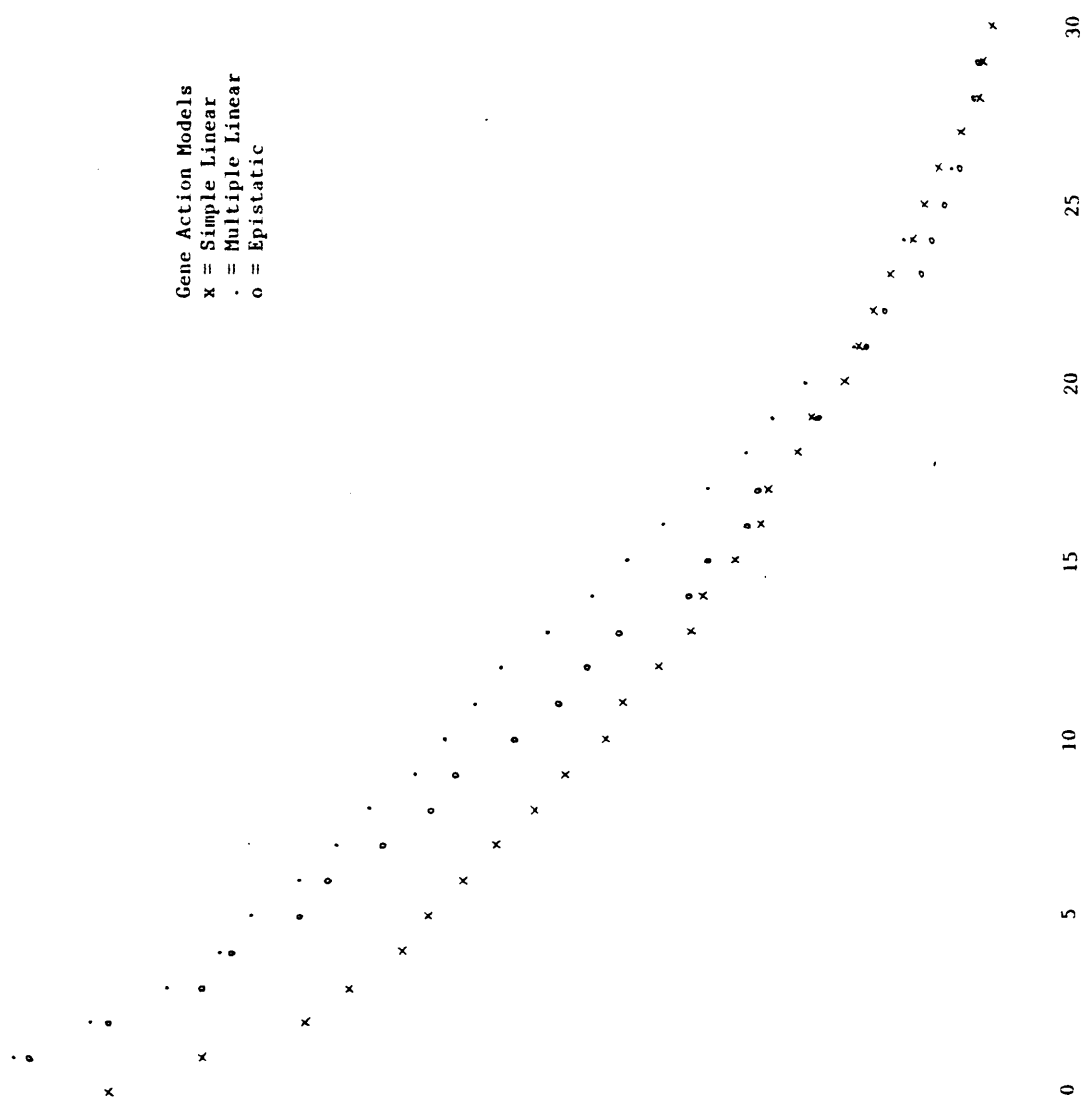


Figure 5.46 h^2_B of All Crosses with N=20, r=0.05, n_h=100, σ^2_E=60

Chapter VI

PERCENT HOMOZYGOSITY AND THE COEFFICIENT OF INBREEDING

6.1 Percent Homozygosity as a Measure of the Coefficient of Inbreeding

For the recurrent selection program under investigation one of the objects is to produce a population which is completely homozygous for the more favorable (+) allele at each locus controlling yield. In such a population each individual would contribute the maximum genotypic value under the gene action models considered in this study. Selection is effective only on those loci which have not become fixed for either allele; therefore, we need a measure of how homozygous the population is with respect to the yield loci.

In the actual experiment, the only measure of homozygosity which can be calculated is the one locus coefficient of inbreeding. This measure is not entirely satisfactory because the genetics of yield involves more than one locus, linkage of loci, and a large selective advantage of the + allele at each locus. Therefore, in this chapter, we will first review the one locus coefficient of inbreeding and how it relates to percent homozygosity in Section 6.2. Theoretical and Monte Carlo investigations will be discussed in Section 6.3. In Section 6.4, the general development of recurrence relationships for the coefficient of inbreeding will be presented using a simple example. Section 6.5 contains the development of the recurrence relationships for Genter's selection experiment which incorporate the practical situation of N-1 individuals produced by each selected cross. Upper and lower bounds for the selection schemes with maximum inbreeding and

minimum inbreeding are presented in Section 6.6. A comparison of theoretical results and results from both the original experiment and the simulation study are presented in Section 6.7.

6.2 The One Locus Coefficient of Inbreeding and Its Relationship to Percent Homozygosity

Consider an individual in an advanced cycle of selection in the selection program. For this individual, some of the loci which were heterozygous in the F_1 will be homozygous and the rest will still be heterozygous. At each of the homozygous loci the two alleles are alike (+,+) or (-,-). Two alleles at any locus may be alike because they are alike in state or alike (identical) by descent. Alleles which are identical by descent are copies of the same allele in an ancestor common to the individual's parents and alleles which are alike in state do not have this property. The one locus coefficient of inbreeding measures the probability that individuals have alleles at a locus which are identical by descent.

Since this selection program begins with two unrelated inbred lines for which we assume the coefficient of inbreeding is 1.0 in each line, when a locus which was heterozygous in the F_1 later becomes homozygous in an individual the two alleles at that locus are identical by descent. Hence, the percentage of loci which were heterozygous in the F_1 and have become homozygous in the population at a later time is used as an estimate of the probability that the two alleles at a locus are identical by descent.

In the actual experiment, the average one-locus coefficient of inbreeding is calculated from the pedigree each cycle of selection. This method can at best be only an approximation of the true probability that alleles are identical by descent because multiple loci, linkage, and selection are not taken into account. Selection can operate to increase this probability while linkage can have the opposite effect. Another factor to be accounted for is the increase in homozygosity due to finite population size.

It is possible to extend the concept of the coefficient of inbreeding to multiple loci under recurrent selection. However, an analytic technique which would take into consideration the large selective advantages of alleles present in this model does not exist. Under the circumstances of this model, an analytical solution to the behavior of the coefficient of inbreeding would be extremely difficult, if not impossible.

The use of percent homozygosity as a measure of the coefficient of inbreeding involves only those loci which were heterozygous in the F_1. Suppose there are a total of n_h heterozygous loci in the F_1 population and let n_t be the number of heterozygous loci in cycle t. Then we can express the percent homozygosity in cycle t as

$$H_t = 1 - \frac{n_t}{n_h} \quad . \tag{6.1}$$

On the other hand, we may define the coefficient of inbreeding in cycle t to be

$$f_t = \frac{1}{n_h} \left(\sum_{i=1}^{n_h} \text{Prob}(x_i = y_i) \right) \tag{6.2}$$

which may be estimated by H_t; i.e.

$$H_t = \hat{f}_t \ . \tag{6.3}$$

For the one locus case in diploid organisms, there is a direct relationship between the percent homozygosity, H, and the coefficient of inbreeding, f, (Kempthorne, 1957):

$$H = f + (1 - f)U \ , \tag{6.4}$$

where U is the probability that two alleles are alike in state, but not copies of an allele from a common ancestor. For the program investigated in this thesis, U = 0 and therefore H = f. In other words, the average coefficient of inbreeding calculated from the selected individuals in the actual experiment is a measure of the average amount of homozygosity in the population.

6.3 Theoretical and Monte Carlo Investigations of the Coefficient of Inbreeding

It is quite obvious that the coefficient of inbreeding and the amount of homozygosity in the population is important in any selection program. Allelic frequencies for heterozygous loci will change with selection. The effectiveness of selection in changing allelic frequencies will be greater for frequencies near 0.5. However, it must be kept in mind that other factors such as the number of heterozygous loci,

linkage of heterozygous loci, and the population size will affect how allelic frequencies change with selection and as a consequence how percent homozygosity changes with selection. As a population becomes more homozygous, selection is not as effective in changing allelic frequencies and percent homozygosity changes more slowly. Therefore, an understanding of the behavior of percent homozygosity for specific types of selection programs is desirable.

An accurate measure of percent homozygosity based on the number and distribution of heterozygous loci contributing to a specific trait requires the capability of observing each individual at the chromosome level. For the foreseeable future this is not a realistic approach and the theoretical one-locus coefficient of inbreeding based on relationships among relatives provides a guide line to the percent homozygosity. While direct observation of the amount of homozygosity is not possible for complex genetic situations, investigation of the behavior of this variable through analytical or Monte Carlo techniques has been done.

As was indicated in Section 5.3, the probability of a locus becoming fixed for a particular allele is a function of the selective advantage of that allele and its frequency in the population. Once an allele becomes fixed in a population, the locus in question is homozygous throughout the population. Therefore, one approach to the investigation of the behavior of homozygosity is to look at the probability that an allele becomes fixed in the population. Approximations of the probability of ultimate fixation of an allele with a slight selective advantage for the one-locus case have been derived by Kimura (1957) and further investigations are reported by Robertson (1960), Ewens (1963),

and Latter (1965).

For the gene action models under investigation in this simulation study, the selective advantage of the + allele is much larger than can be accommodated by these approximations. Also, we are interested in the fixation of the + allele at many loci which belong to different gene action models. Since even in the simpliest cases exact theoretical investigation is not feasible and approximations are necessary, more complex situations can only be investigated using Monte Carlo techniques. Qureshi and Kempthorne (1968), for example, present results of a Monte Carlo investigation of the fixation of alleles in a polygenic system with 40 loci.

Two papers by Cain and Hinkelmann (1970, 1972) present results of theoretical investigations of the coefficient of inbreeding in recurrent selection programs. For both the one locus case (Cain and Hinkelmann, 1970) and the case of m-linked loci (Cain and Hinkelmann, 1972), recurrence relationships are developed relating the coefficient of inbreeding in a given cycle of selection to the coefficient of inbreeding in previous cycles. These relationships are derived for three selection schemes: (a) selection for minimum inbreeding, (b) selection for maximum inbreeding, and (c) random selection. Further results are presented for the one locus case and the case of two linked loci by Choy and Weir (1976).

The relationships derived by Cain and Hinkelmann (1970, 1972) and Choy and Weir (1976) are applicable to recurrent selection programs which follow a self-select-intercross pattern for each cycle of selection. In addition, only a situation in which one offspring is produced

by each mating is considered. Finally, the selection criteria considered is based on controlling the amount of inbreeding and not on gene action model and number and distribution of loci.

6.4 General Development of the Recurrence Relationships for the Coefficient of Inbreeding

The development of theoretical recurrence relationships for the coefficient of inbreeding is base on relationships among the individuals selected to be parents of the succeeding generation. The one-locus coefficient of inbreeding of an individual X is defined as the probability that the two alleles at a given locus are identical by descent from an ancestor and is measured by Malecot's (1948) coefficient of parentage of the individual's parents. The pedigree in Figure 6.1 will be used to illustrate these two measures. By definition, if X has the genotype (x_1, x_2), the coefficient of inbreeding of X is

$$f_X = \text{Prob}(x_1 = x_2) \tag{6.5}$$

and this is equal to r_{CD}, the coefficient of parentage between C and D. The coefficient of parentage of individuals C and D is, by definition,

$$\begin{aligned} r_{CD} &= \frac{1}{4}\left(P(c_1=d_1) + P(c_1=d_2) + P(c_2=d_1) + P(c_2=d_2)\right) \\ &= \frac{1}{4}\left(r_{AA} + r_{AB} + r_{AB} + r_{BB}\right) \qquad (6.6) \\ &= \frac{1}{4}\left(\frac{1}{2}(1+f_A) + 2r_{AB} + \frac{1}{2}(1+f_B)\right) \\ &= \frac{1}{4} + \frac{1}{8}(f_A + f_B) + \frac{1}{2} r_{AB} \,. \qquad (6.7) \end{aligned}$$

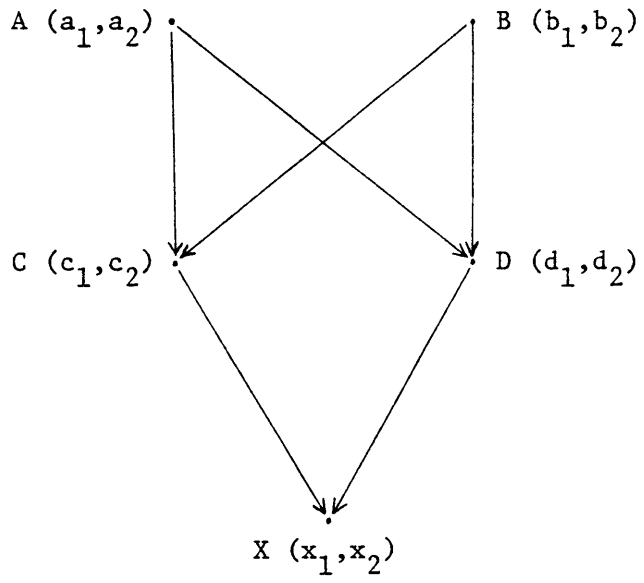


Figure 6.1 Pedigree of Individual X

From the definition of the coefficient of inbreeding and the coefficient of parentage, it is evident that $r_{AB} = f_C = f_D$ and therefore the coefficient of inbreeding of individual X can be expressed as a function of the coefficients of inbreeding of its ancestors. For further explanation see Kempthorne (1957) or Li (1955).

If we assume that $f_A = f_B$, then the coefficient of inbreeding of X can be expressed as a function of the average coefficient of inbreeding in the 2 previous generations. Thus,

$$f_3 = \frac{1}{4} + \frac{1}{4} f_1 + \frac{1}{2} f_2 , \tag{6.8}$$

where f_i represents the average coefficient of inbreeding in the i^{th} generation. In the next two sections, we use this approach in the development of recurrence equations for the average coefficient of inbreeding in each cycle of the recurrent selection program under investigation.

6.5 Recurrence Equations for Recurrent Selection without a Self-Fertilization Generation in each Cycle

For certain regular mating and selection systems, recurrence relationships can be developed relating the coefficient of inbreeding in one cycle to that of previous cycles. In the development of a recurrence formula for a recurrent selection experiment without a self-fertilization cycle and N-1 individuals produced by each mating, two situations must be considered. These situations can be illustrated by considering the individuals A_1 and A_2 of cycle t in Figure 6.2, which can be considered an extension of Figure 2.1. By tracing paths

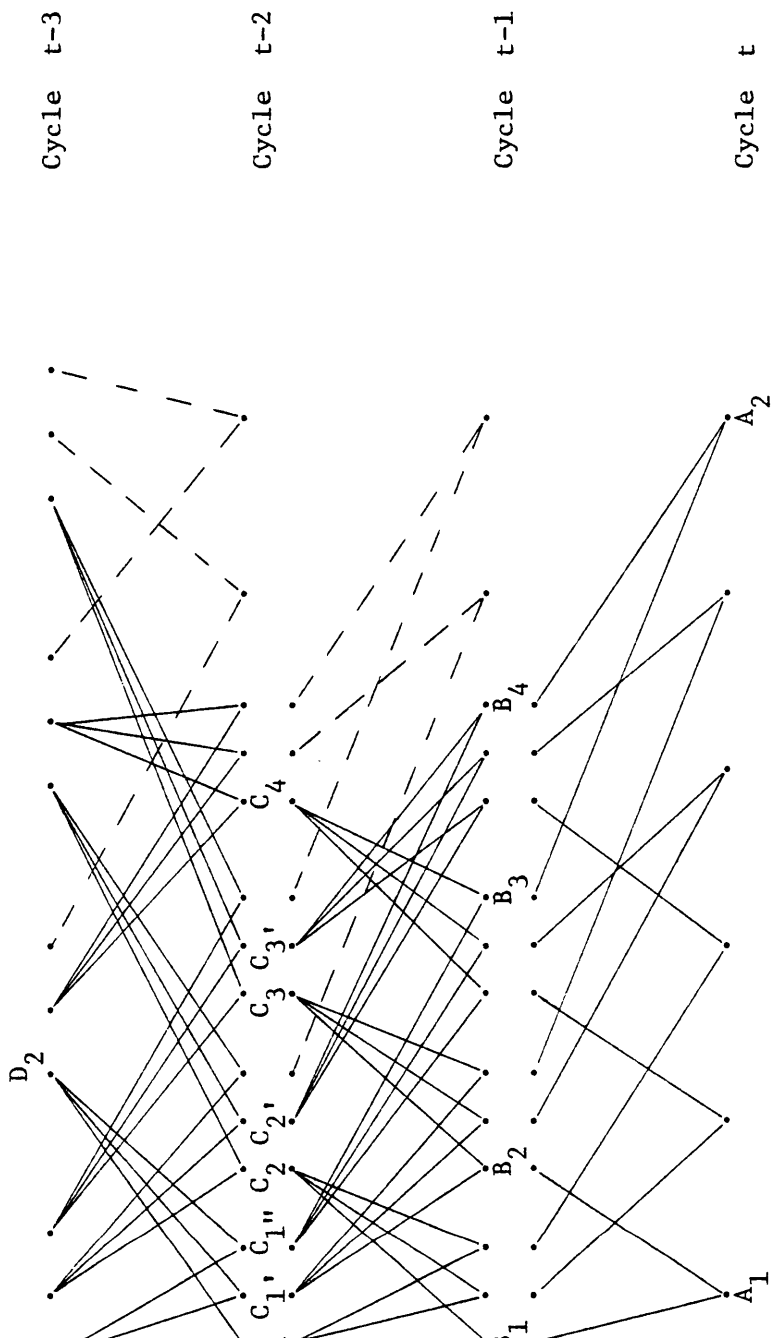


Figure 6.2 Recurrent Selection without a Self-Fertilization Generation Each Cycle: N = 4 and N-1 Individuals Are Produced by Each Mating

back from A_1 to the grandparents in cycle t-2, we find that 3 selected groups are represented and two of the grandparents are full sibs. Individual A_2 can be traced to grandparents representing four selected groups in cycle t-2. Expansion of the coefficient of inbreeding for these two individuals gives

$$f_{A_1} = r_{B_1B_2} = \frac{1}{4}\left(r_{C_1C_1'} + r_{C_1C_3} + r_{C_2C_1'} + r_{C_2C_3}\right) , \qquad (6.9)$$

and

$$f_{A_2} = r_{B_3B_4} = \frac{1}{4}\left(r_{C_1,C_2} + r_{C_1,C_3} + r_{C_4C_2} + r_{C_4C_3}\right) . \qquad (6.10)$$

Since C_1 and C_1' are full sibs just as C and D are in Figure 6.1,

$$r_{C_1C_1'} = \frac{1}{4} + \frac{1}{8}\left(f_{D_1} + f_{D_2}\right) + \frac{1}{2} r_{D_1D_2} . \qquad (6.11)$$

If we assume that $f_{D_1} = f_{D_2}$, then

$$r_{C_1C_1'} = \frac{1}{4}\left(1 + f_{t-3}\right) + \frac{1}{2} f_{t-2} . \qquad (6.12)$$

If we also assume that the coefficients of parentage between individuals in different groups of the same cycle are equal, then we can write

$$f_{A_1} = \frac{1}{16} + \frac{1}{16} f_{t-3} + \frac{1}{8} f_{t-2} + \frac{3}{4} f_{t-1} , \qquad (6.13)$$

and

$$f_{A_2} = f_{t-1} . \qquad (6.14)$$

Thus, an individual in cycle t whose grandparents in cycle t-2 represent four selected groups will have the same coefficient of inbreeding as individuals in cycle t-1. An individual in cycle t, whose grandparents represent only three selected groups in cycle t-2, will have a coefficient of inbreeding which is a function of the coefficient of inbreeding in the three previous cycles of selection.

If we let P_3 represent the probability that an individual in generation t has grandparents in three selected groups of cycle t-2 and P_4 represent the probability that an individual in cycle t has grandparents in four selected groups in cycle t-2, then a general expression for the coefficient of inbreeding in cycle t is, using (6.13) and (6.14),

$$f_t = P_3(1/16 + 1/16\ f_{t-3} + 1/8\ f_{t-2} + 3/4\ f_{t-1}) + P_4 f_{t-1} . \quad (6.15)$$

Our assumption that the coefficient of parentage between individuals in different groups of the same cycle are equal was necessary to the development of a recurrence relationship for the coefficient of inbreeding, but it is not necessarily a valid assumption. More accurate relationships would incorporate information from all previous cycles, but this is not always available.

Even though we cannot give exact recurrence equations, we can develop relationships which give upper and lower bounds for the coefficient of inbreeding when selection is for minimum inbreeding or for maximum inbreeding. Comparisons of these recurrence relationships applied to the selection program under investigation and the simulation results may aid in the interpretation of the behavior of the coefficient

of inbreeding or percent homozygosity. The probabilities, P_3 and P_4, in (6.15) depend upon the particular selection scheme, random selection, selection for minimum inbreeding, or selection for maximum inbreeding.

Values of P_3 and P_4 calculated for other recurrence selection schemes by Choy and Weir (1976) are applicable to this recurrent selection program without a self-fertilization generation each cycle. For random selection each cycle,

$$P_3 = \frac{4}{N+1}$$

and (6.16)

$$P_4 = \frac{N-3}{N+1} \quad .$$

If selection is performed for minimum inbreeding, each selected group in cycle t-2 contributes equally to the selected groups in cycle t-1 and consequently

$$P_3 = \frac{2}{N-1}$$

and (6.17)

$$P_4 = \frac{N-3}{N-1} \quad .$$

When selection for maximum inbreeding is practiced, the following situation occurs: (i) one of the groups in cycle t-2 contributes N-1 times to the selected groups in cycle t-1, (ii) two groups in cycle t-2 contribute twice to the selected groups in cycle t-1, and (iii) the remaining N-3 groups in cycle t-2 contribute only once each to the selected groups in cycle t-1. Under selection for maximum inbreeding,

of the $\binom{N}{2}$ possible pairs of individuals in the selected groups of cycle t-1, the probability that the parents of a given pair represent three groups in cycle t-2 is

$$P_3 = \frac{N^2-3N+6}{N(N-1)} \quad . \tag{6.18}$$

The probability that the parents of a pair of individuals in cycle t-1 represent four groups in cycle t-2 is

$$P_4 = \frac{2(N-3)}{N(N-1)} \quad . \tag{6.19}$$

Using these probabilities in equation (6.15) will give a recurrence formula for the coefficient of inbreeding for different selection schemes in a recurrent selection program without a self-fertilization generation each cycle of selection.

6.6 Upper and Lower Bounds on the Recurrence Relationships for Selection for Minimum and Maximum Inbreeding

The recurrence equation (6.15) together with the appropriate values of P_3 and P_4 will give a lower bound on the coefficient of inbreeding under selection for maximum inbreeding and an upper bound on the coefficient of inbreeding under selection for minimum inbreeding. This is a direct consequence of the assumption that the coefficient of parentage of all pairs of individuals in each cycle is the same. Another assumption which was made, but not explicitly stated, is that the average coefficient of inbreeding is the same for the selected groups as for the entire population each cycle.

A recurrence equation that will give an upper bound for the selection for maximum inbreeding can be found by considering the maximum coefficient of parentage between individuals in different selected groups of a given cycle of selection. A recurrence equation for the maximum coefficient of parentage can be derived from (6.9) and (6.11). Expressing $r_{C_1C_{1'}}$ as a function of the maximum coefficient of parentage in previous cycles yields

$$r_{C_1C_{1'}} = \frac{1}{4} + \frac{1}{4} r_{MAX}^{t-4} + \frac{1}{2} r_{MAX}^{t-3} \quad . \tag{6.20}$$

Replacing each coefficient of parentage of individuals in different groups in (6.9) with r_{MAX}^{t-2} and combining (6.9) with (6.20) gives

$$r_{MAX}^{t-1} = \frac{1}{16} + \frac{1}{16} r_{MAX}^{t-4} + \frac{1}{8} r_{MAX}^{t-3} + \frac{3}{4} r_{MAX}^{t-2} \quad . \tag{6.21}$$

Using the probabilities in (6.18) and (6.19), a recurrence equation for the upper bound on the coefficient of inbreeding in cycle t is given by

$$f_{MAX}^{t} = P_3 \, r_{MAX}^{t-1} + P_4 \, r_{MAX}^{t-2} \quad . \tag{6.22}$$

Examination of the actual coefficient of inbreeding from a pedigree under selection for maximum inbreeding will give values between the lower and upper bounds each cycle as indicated in Table 6.1.

Substitution of values for P_3 and P_4 from (6.17) into (6.15) will give an equation which gives an upper bound for the average coefficient of inbreeding under selection for minimum inbreeding. This specific selection scheme was not of interest with respect to the recurrent

TABLE 6.1

Upper and Lower Bounds for the Coefficient of Inbreeding when Selection is for Maximum Inbreeding Compared to the Exact Value when N=4

Cycle Number	Recurrence Formula		Exact Maximum
	Lower Bound	Upper Bound	
0	0.5	0.5	0.5
1	0.5	0.5	0.5
2	0.5260	0.5260	0.5260
3	0.5466	0.5508	0.5482
4	0.5657	0.5726	0.5693
5	0.5843	0.5937	0.5903
6	0.6020	0.6137	0.6103

selection program under investigation, but an attempt was made to develop a relationship for a lower bound on the average coefficient of inbreeding in this case. This has proved unsuccessful to date except for the trivial lower bound of 0.5 for each cycle of selection.

In the next section, observed, simulated, and theoretical values for the average coefficient of inbreeding and percent homozygosity are compared for the situations investigated in this simulation study.

6.7 Comparison of Theoretical and Observed Results

In the absense of selection, the coefficient of inbreeding or percent homozygosity will increase because of finite population size. The amount of heterozygosity each generation is decreased by the fraction

$$\lambda = \frac{N-1 + \sqrt{N^2+1}}{2N}$$

where N is the population size in each generation (Crow and Kimura, 1970). The effect of finite population size on the behavior of percent homozygosity is greater when N is small. Figure 6.3 compares the behavior of the one-locus coefficient of inbreeding under selection for maximum inbreeding and of percent homozygosity in a finite population of size N = 20 with the values observed for the actual experiment and the simulation results. By cycle three, the average values for the observed and simulated results are between the values for finite population size and the upper bound for selection for maximum inbreeding. Results for additional cycles of selection are needed before a definite pattern can be established.

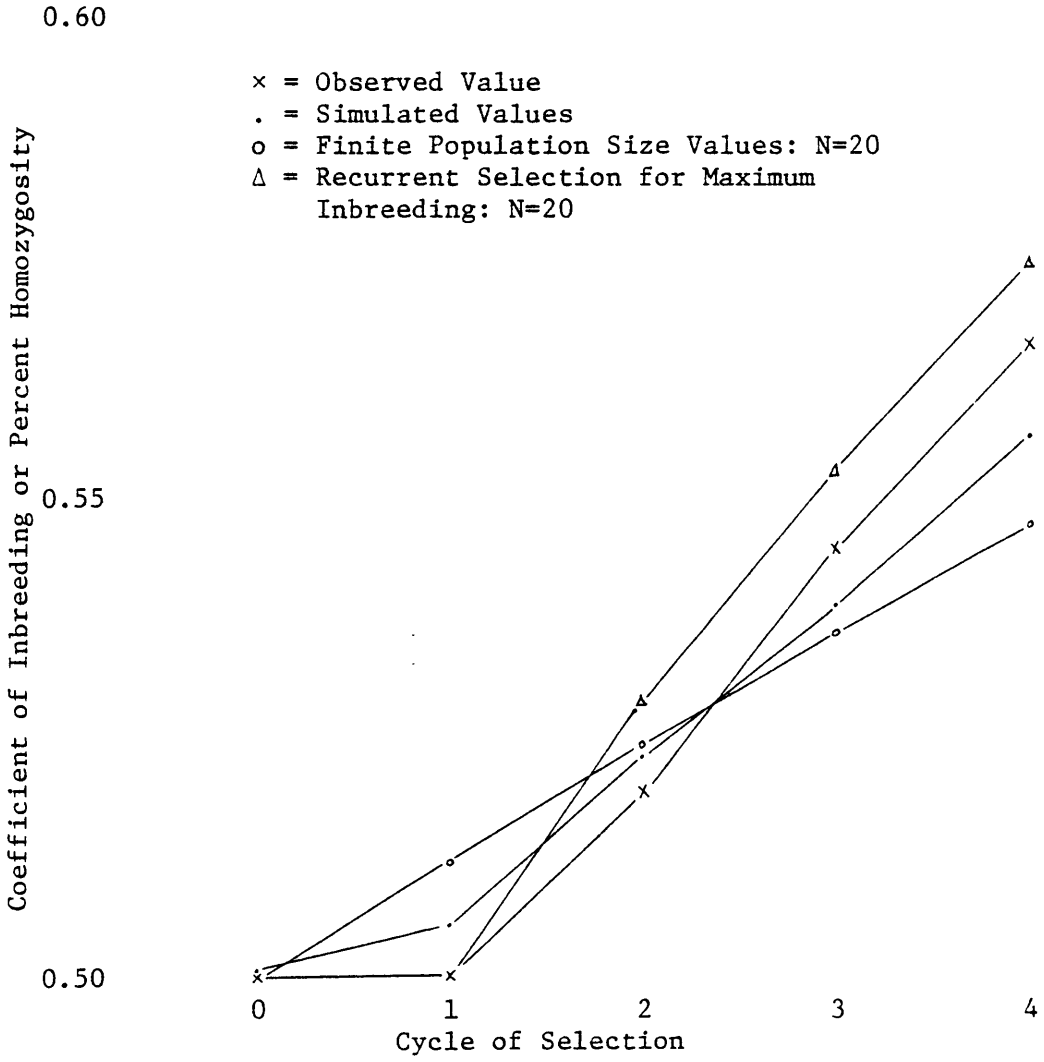


Figure 6.3 Average Coefficient of Inbreeding or Percent Homozygosity for Selection Cycles 0-4

The theoretical behavior of the upper bound on the average coefficient of inbreeding under maximum inbreeding and the change in homozygosity due to finite population size are illustrated in Figures 6.4 and 6.5 for N = 10 and N = 20, respectively, for the length of the simulated selection program. When N = 10, both of these curves are very similar; however, when N = 20, the change in homozygosity due to finite population size is lower than the upper bound for maximum inbreeding.

These curves have been compared with the behavior of percent homozygosity in the simulation study. For some of the simulation runs, the behavior of percent homozygosity over time, as illustrated in Figures 5.3 to 5.18 and the theoretical curves in Figures 6.4 and 6.5 are similar. In the majority of these situations, $n_h = 100$. For simulation runs with N = 10 and results similar to the values for finite population size, the simulated values were slightly higher until cycle 20 or 25 and then became lower. The most favorable comparisons occurred for the simulation runs with N = 10, r = 0.05, $n_h = 50$, and $\sigma_E^2 = 60$ and with N = 20, r = 0.05, $n_h = 100$, and $\sigma_E^2 = 60$. Thus in situations with many loci contributing to a trait such as yield and tight linkage of these loci, the upper bound for the average coefficient of inbreeding under selection for maximum inbreeding will provide a good approximation to the behavior of the change in homozygosity during the selection program.

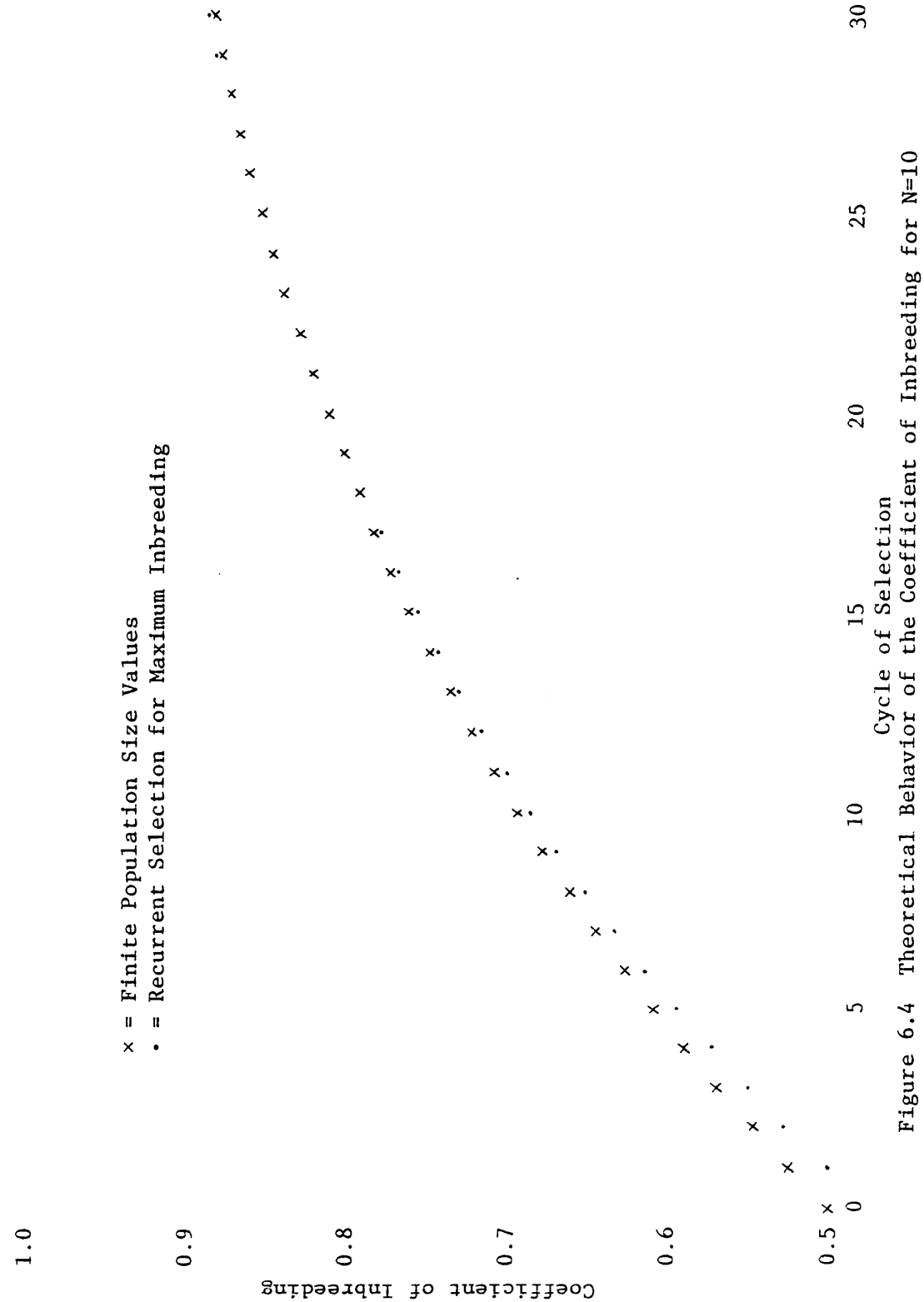


Figure 6.4 Theoretical Behavior of the Coefficient of Inbreeding for N=10

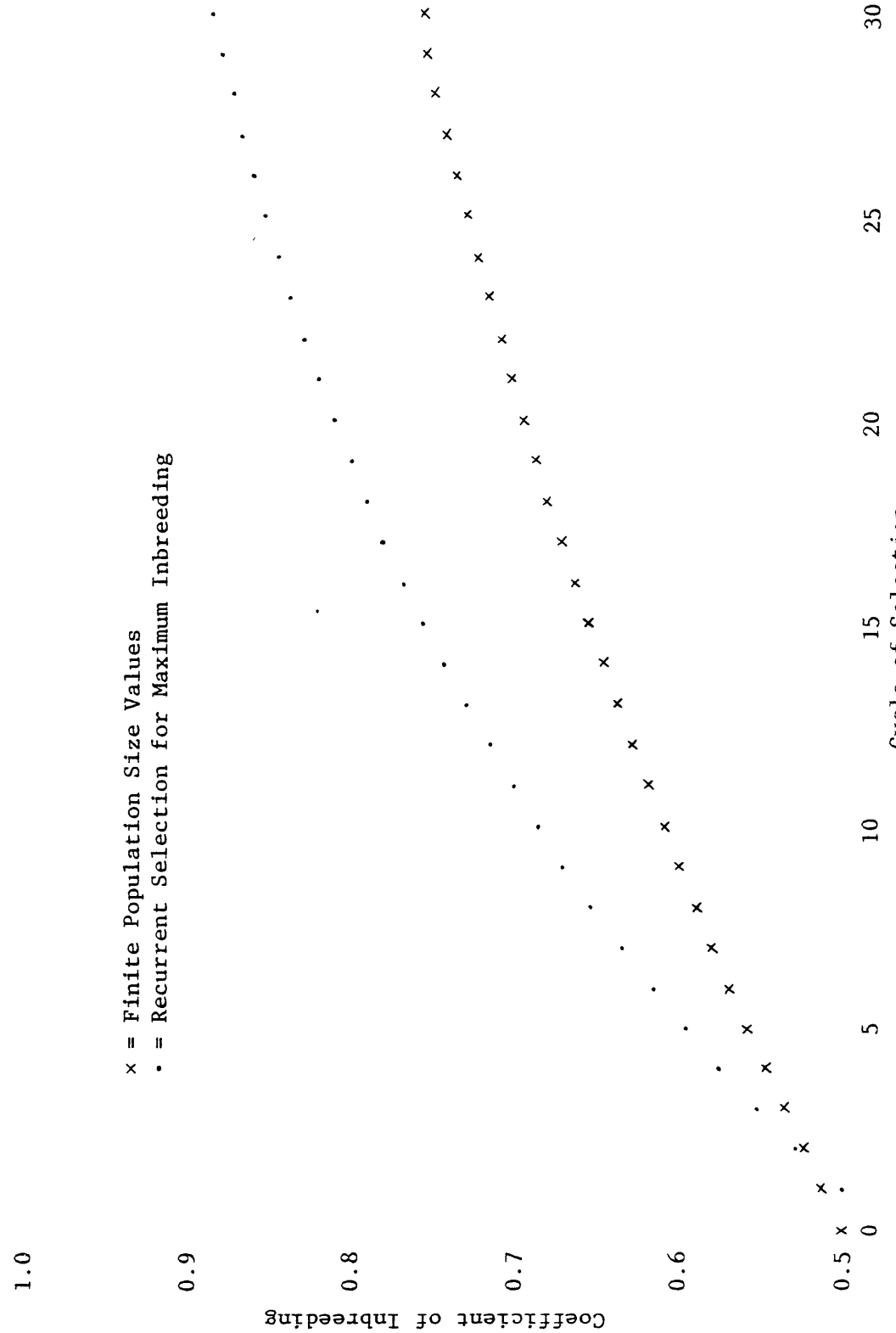


Figure 6.5 Theoretical Behavior of the Coefficient of Inbreeding for N=20

Chapter VII

GENETIC VARIANCE COMPONENTS

7.1 Introduction

Partitioning of the genotypic variance into various components provides useful information in a selection program. The effectiveness of a selection program, for example, can be measured by the amount of additive genetic variance in a population. The other components, such as dominance and epistatic components, may help generally in the determination of what type of gene action is involved in the trait of interest. For the gene action models used in this simulation study, evaluation of the possible genotypic variance components will help in the understanding of the effect of different types of gene action on selection. In addition, the relative size of the epistatic variance components in the epistatic gene action model may indicate why there is difficulty in detecting epistasis in actual populations. The partitioning of the genotypic variance for the gene action models in this study is presented in Section 7.2.

Linkage of loci which contribute to a particular trait such as yield increases the genotypic variance associated with that trait. Since the most realistic gene action models for quantitative traits will include linkage of loci, we have considered an extenstion of the formulas used to partition the genotypic variance under complete independence of all loci to include the case of m-linked loci. A discussion of the specific situations considered is presented in Section 7.3.

7.2 Partitioning of the Genotypic Variance for Symmetric Gene Action Models

The total genotypic variance can be partitioned into components representing (a) the effects of the individual alleles at each locus, (b) the interaction of alleles at the same locus and (c) the interaction of alleles at different loci. A general derivation for this partitioning of the genotypic variance into meaningful variance components is given by Kempthorne (1954). In general, the genotypic variance can be expressed as

$$\sigma_G^2 = \sum_{r,s} \sigma^2_{A^r D^s} , \tag{7.1}$$

where the subscripts r and s satisfy the inequality, $1 \leq r+s \leq m$, and m is the total number of loci. For m = 2

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2$$

or (7.2)

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2 ,$$

where $\sigma_I^2 = \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2$. The additive variance, σ_A^2, represents the variance of the effects of individual alleles at each locus; the dominance variance, σ_D^2, represents the variance of the interaction of alleles at the same locus; and the epistatic variance, σ_I^2, represents the variance of the interaction of alleles at different loci. For m loci, σ_G^2 can be partitioned into $\frac{(m+1)(m+2)}{2} - 1$ variance components.

An extension of the partitioning of the genotypic variance to include symmetric models of gene action is presented by Horner and Kempthorne (1955) and Horner (1956). Symmetric gene action models, such as the ones investigated in this simulation study, are models for which the genotypic value can be expressed simply as a function of the number of loci in each of three phases, (+,+), (+,-), or (-,-). It is assumed that the population is mating at random, there are two alleles at each locus, no linkage, and all loci follow the same gene action model. Expressions are derived for the case where the gene frequency is 0.5 at all loci (Horner and Kempthorne, 1955) and for the case where the gene frequency of the more favorable (+) allele is the same at all loci, but not necessarily 0.5 (Horner, 1956). In a paper by Hill, et al. (1963), the previous work was extended to include a situation in which the loci are divided into several subgenotypes, each with a different gene action model.

Expressions for the variance components for the gene action models used in this simulation study are given in Table 7.1. These expressions are an application of equations developed by Hill, et al. (1963) to the gene action models in Table 2.2. The F_2 generation for this recurrent selection program has only two alleles at each segregating locus and the frequency of each allele is 0.5 for all loci. Evaluation of the genotypic variance components for the F_2 generation for all three gene action models with $n_h = 100$ is given in Table 7.2. The simple linear and multiple linear gene action models give rise to only additive and dominance components of variance, while the epistatic model yields variance components up through order three. More specifically, in the

TABLE 7.1

Expressions for Evaluation of Genotypic Variance Components for the Simple Linear, Multiple Linear, and Epistatic Gene Action Models

Simple Linear Model

General Formulas	F_2 Generation ($p_0=p_1=0.5$)
$\mu = n(ap_1+d(2)p_0p_1-ap_0^2)$	$\mu = \frac{n}{2} d$
$\sigma_A^2 = 2np_0p_1(a-d(p_1-p_0))^2$	$\sigma_A^2 = \frac{n}{2} a^2$
$\sigma_D^2 = 4np_0^2p_1^2(d)^2$	$\sigma_D^2 = \frac{n}{4} d^2$

Multiple Linear Model

General Formulas	F_2 Generation ($p_0=p_1=0.5$)
$\mu = \sum_{i=1}^{4} n_i(a_ip_{i1}^2+d_i(2)p_{i0}p_{i1}-a_ip_{i0}^2)$	$\mu = \frac{1}{2} \sum_{i=1}^{4} n_i d_i$
$\sigma_A^2 = 2 \sum_{i=1}^{4} n_ip_{i0}p_{i1}(a_i-di(p_{i1}-p_{i0}))^2$	$\sigma_A^2 = \frac{1}{2} \sum_{i=1}^{4} n_i a_i^2$
$\sigma_D^2 = 4 \sum_{i=1}^{4} n_ip_{i0}^2p_{i1}^2d_i^2$	$\sigma_D^2 = \frac{1}{4} \sum_{i=1}^{4} n_i d_i^2$

Epistatic Model

General Formulas

$$\mu = \sum_{i=1}^{4} n_i(a_ip_{i1}^2+d_i2p_{i0}p_{i1}-a_ip_{i0}^2) + n_5(a_5p_{51}^2+d_52p_{50}p_{51}-a_5p_{50}^2) \times$$

$$\left(\sum_{t=\tau}^{n_6} \binom{n_6}{n_6-t} (p_{60}^2)^{n_6-t} (1-p_{60}^2)^t \right)$$

$$+ U - (V-n_8(a_8p_{81}^2+d_82p_{80}p_{81}-a_8p_{80}^2)p_{70}^2$$

TABLE 7.1 (cont.)

Epistatic Model

General Formulas

$$\sigma_A^2 = 8\sum_{i=1}^{4} n_i p_{i0}(-\tfrac{1}{2}(a_i - d_i(p_{i1}-p_{i0})))^2$$

$$+ n_5 8p_{50}p_{51}(-\tfrac{1}{2}(a_5 - d_5(p_{51}-p_{50}))(\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6}{t}(p_{60}^2)^s(1-p_{60}^2)^{n_6-s}))^2$$

$$+ n_6 8p_{60}p_{61}((n_5(a_5p_{51}+d_5 2p_{50}p_{51}-a_5p_{50}) \times$$

$$(-\frac{p_{60}}{2}\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-1}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-1-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s}))^2$$

$$+ n_7 8p_{70}p_{71}(-(V-n_8(a_8p_{81}^2+d_8 2p_{80}p_{81}-a_8p_{80}^2))\frac{p_{70}}{2})^2$$

$$+ n_8 8p_{80}p_{81}(-(-\tfrac{1}{2}(a_8-d_8(p_{81}-p_{80})))p_{70}^2)^2$$

$$\sigma_D^2 = \sum_{i=1}^{4} n_i(4p_{i0}p_{i1})^2(-\tfrac{1}{2}d_i)^2 + n_5(4p_{50}p_{51})^2 \times$$

$$(-\tfrac{1}{2}d_5\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-s})^2$$

$$+ n_6(4p_{60}p_{61})^2(n_5(a_5p_{51}^2+d_5 2p_{50}p_{51}-a_5p_{50}^2) \times$$

$$(-\tfrac{1}{4}\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-1}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-1-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s})^2$$

$$+ n_7(4p_{70}p_{71})^2(-(V-n_8(a_8p_{81}^2+d_8^2p_{80}p_{81}-a_8p_{80}^2))\tfrac{1}{4})^2$$

$$+ n_8(4p_{80}p_{81})^2(-(\tfrac{1}{2}d_8)p_{70}^2)^2$$

TABLE 7.1 (cont.)

Epistatic Model

General Formulas

$$\sigma^2_{AA} = n_5(8p_{50}p_{51})n_6(8p_{60}p_{61})(-\tfrac{1}{2}(a_5-d_5(p_{51}-p_{50})) \times$$

$$(-\frac{p_{60}}{2}\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-1}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-1-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s}))^2$$

$$+\frac{n_6(n_6-1)}{2}(8p_{60}p_{61})^2(n_5(a_5p_{51}^2+d_5 2p_{50}p_{51}-a_5p_{50}^2)(\frac{p_{60}}{2})^2$$

$$\times\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(p_{60}^2)^s(1-p_{60})^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s})^2$$

$$\sigma^2_{AD} = n_5(8p_{50}p_{51})n_6(4p_{60}p_{61})^2(-\tfrac{1}{2}(a_5-d_5(p_{51}-p_{50})) \times$$

$$(-\tfrac{1}{4}\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-1}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-1-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s}))^2$$

$$+ n_5(4p_{50}p_{51})^2n_6(8p_{60}p_{61}) \times$$

$$(-\tfrac{1}{2}d_5(-\frac{p_{60}}{2}\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-1}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-1-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s}))^2$$

$$+\frac{n_6(n_6-1)}{2}(8p_{60}p_{61})(4p_{60}p_{61})^2(n_5(a_5p_{51}^2+d_5 2p_{50}p_{51}-a_5p_{50}^2) \times$$

$$(\frac{p_{60}}{2}(\tfrac{1}{4})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s}))^2$$

TABLE 7.1 (cont.)

Epistatic Model

General Formulas

$$\sigma^2_{DD} = n_5(4p_{50}p_{51})^2 n_6(4p_{60}p_{61})^2 \times$$

$$(-\tfrac{1}{2}d_5(-\tfrac{1}{4})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-1}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-1-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s})^2$$

$$+\frac{n_6(n_6-1)}{2}(4p_{60}p_{61})^4(n_5(a_5p_{51}^2+d_5 2p_{50}p_{51}-a_5p_{50}^2)\times$$

$$((\tfrac{1}{4})^2\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s}))^2$$

$$+ n_7(4p_{70}p_{71})^2 n_8(4p_{80}p_{81})^2(-\frac{1}{8}d_8)^2$$

$$\sigma^2_{AAA} = \frac{n_6(n_6-1)(n_6-2)}{3!}(8p_{60}p_{61})^3(n_5(a_5p_{51}^2+d_5 2p_{50}p_{51}-a_5p_{50}^2)\times$$

$$(-(\frac{p_{60}}{2})^3\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-3}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-3-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s}))^2$$

$$+ n_5(8p_{50}p_{51})\frac{n_6(n_6-1)}{2}(8p_{60}p_{61})^2\times$$

$$(-\tfrac{1}{2}(a_5-d_5(p_{51}-p_{50}))(\frac{p_{60}^2}{2})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-2-s}\times$$

$$(-1)^{n_6-t-s}\binom{1}{n_6-t-s})^2$$

TABLE 7.1 (cont.)

Epistatic Model

General Formulas

$$\sigma^2_{AAD} = n_5(8p_{50}p_{51})\frac{n_6(n_6-1)}{2}(8p_{60}p_{61})(4p_{60}p_{61})^2(-\tfrac{1}{2}(a_5-d_5(p_{51}-p_{50}))\times$$

$$(\frac{p_{60}}{2}(\tfrac{1}{4})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s}))^2$$

$$+ n_5(4p_{50}p_{51})^2\frac{n_6(n_6-1)}{2}(8p_{60}p_{61})^2\times$$

$$(-\tfrac{1}{2}d_5(\frac{p_{60}}{2})^2\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s})^2$$

$$+\frac{n_6(n_6-1)(n_6-2)}{3!}(8p_{60}p_{61})^2(4p_{60}p_{61})^2\times$$

$$(n_5(a_5p_{51}^2+d_5 2p_{50}p_{51}-a_5p_{50}^2)(-1)(\frac{p_{60}}{2})^2(\tfrac{1}{4})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-3}{s}(p_{60}^2)^s\times$$

$$(1-p_{60}^2)^{n_6-3-s}(-1)^{n_6-t-s}\binom{3}{n_6-t-s})^2$$

TABLE 7.1 (cont.)

Epistatic Model

General Formulas

$$\sigma^2_{ADD} = n_5(8p_{50}p_{51})\frac{n_6(n_6-1)}{2}(4p_{60}p_{61})^4(-\tfrac{1}{2}(a_5-d_5(p_{51}-p_{50}))\times$$

$$(\tfrac{1}{4})^2\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s})^2$$

$$+ n_5(4p_{50}p_{51})^2\frac{n_6(n_6-1)}{2}(8p_{60}p_{61})(4p_{60}p_{61})^2\times$$

$$(-\tfrac{1}{2}d_5(\frac{p_{60}}{2})(\tfrac{1}{4})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s})^2$$

$$+\frac{n_6(n_6-1)(n_6-2)}{3!}(8p_{60}p_{61})(4p_{60}p_{61})^4(n_5(a_5p_{51}^2+d_5{}^2p_{50}p_{51}-a_5p_{50}^2)\times$$

$$(-1)(\frac{p_{60}}{2})(\tfrac{1}{4})^2\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-3}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-3-s}(-1)^{n_6-t-s}\binom{3}{n_6-t-s})^2$$

$$\sigma^2_{DDD} = n_5(4p_{50}p_{51})^2\frac{n_6(n_6-1)}{2}(4p_{60}p_{61})^4(-\tfrac{1}{2}d_5\times$$

$$(\tfrac{1}{4})^2\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s})^2$$

$$+\frac{n_6(n_6-1)(n_6-2)}{3!}(4p_{60}p_{61})^6(n_5(a_5p_{51}^2+d_5{}^2p_{50}p_{51}-a_5p_{50}^2)\times$$

$$(-1)(\tfrac{1}{4})^3\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-3}{s}(p_{60}^2)^s(1-p_{60}^2)^{n_6-3-s}(-1)^{n_6-t-s}\binom{3}{n_6-t-s})^2$$

TABLE 7.1 (cont.)

Epistatic Model

F_2 Generation ($p_0=p_1=0.5$)

$$\mu = \tfrac{1}{2}\sum_{i=1}^{4} n_i d_i + \tfrac{1}{2}n_5 d_5 \left(\sum_{t=}^{n_6} \binom{n_6}{n_6-t} (\tfrac{1}{4})^{n_6-t} (\tfrac{3}{4})^{t}\right) + U - (V - n_8(\tfrac{1}{2}d_8))\tfrac{1}{4}$$

$$\sigma_A^2 = 2\sum_{i=1}^{4} n_i(-\tfrac{1}{2}a_i)^2 + 2n_6(n_5(\tfrac{1}{2})d_5(-1)(\tfrac{1}{4}) \times$$

$$\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t} \binom{n_6-1}{s}(\tfrac{1}{4})^{s}(\tfrac{3}{4})^{n_6-1-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s})^2$$

$$+ 2n_5\left(-\tfrac{1}{2}a_5 \sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t} \binom{n_6}{s}(\tfrac{1}{4})^2(\tfrac{3}{4})^{n_6-s}\right)^2 + 2n_7(-(V-n_8(\tfrac{1}{2}d_8))\tfrac{1}{4})^2$$

$$+ 2n_8\left(\tfrac{1}{8}a_8 \atop \tfrac{1}{2}a_8(\tfrac{1}{4})\right)^2$$

$$\sigma_D^2 = \sum_{i=1}^{4} n_i(-\tfrac{1}{2}d_i)^2 + n_5\left(-\tfrac{1}{2}d_5 \sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t} \binom{n_6}{s}(\tfrac{1}{4})^{s}(\tfrac{3}{4})^{n_6-s}\right)^2$$

$$+ n_6\left(n_5(\tfrac{1}{2}d_5)(-\tfrac{1}{4}) \sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t} \binom{n_6-1}{s}(\tfrac{1}{4})^{s}(\tfrac{3}{4})^{n_6-1-s}\binom{1}{n_6-t-s}\right)^2$$

$$+ n_7(-(V-n_8(\tfrac{1}{2}d_8))\tfrac{1}{4})^2 + n_8(-\tfrac{1}{2}d_8(\tfrac{1}{4}))^2$$

$$\sigma_{AA}^2 = 2n_5 2n_6\left(-\tfrac{1}{2}a_5(-\tfrac{1}{4}) \sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t} \binom{n_6-1}{s}(\tfrac{1}{4})^{s}(\tfrac{3}{4})^{n_6-1-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s}\right)^2$$

$$+ 4\,\frac{n_6(n_6-1)}{2}\left(n_5(\tfrac{1}{4}d_5)(\tfrac{1}{4})^2 \sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t} \binom{n_6-1}{s}(\tfrac{1}{4})^{s}\right.$$

$$\left.(\tfrac{3}{4})^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s}\right)^2$$

TABLE 7.1 (cont.)

Epistatic Model

F_2 Generation ($p_0=p_1=0.5$)

$$\sigma^2_{AD} = 2n_5n_6(-\tfrac{1}{2}(a_5)(-\tfrac{1}{4})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-1}{s}(\tfrac{1}{4})^s(\tfrac{3}{4})^{n_6-1-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s})^2$$

$$+ n_5 2n_6(-\tfrac{1}{2}d_5(-\tfrac{1}{4})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-1}{s}(\tfrac{1}{4})^s(\tfrac{3}{4})^{n_6-1-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s})^2$$

$$+ 2\frac{n_6(n_6-1)}{2}(n_5(\tfrac{1}{2}d_5)(\tfrac{1}{4})(\tfrac{1}{4})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(\tfrac{1}{4})^s \times$$

$$(\tfrac{3}{4})^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s})^2$$

$$+ 2n_7n_8(-\tfrac{1}{8}d_8)^2 + n_7 2n_8(-\tfrac{1}{8}a_8)^2$$

$$\sigma^2_{DD} = n_5n_6(-\tfrac{1}{2}d_5(-\tfrac{1}{4})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-1}{s}(\tfrac{1}{4})^s(\tfrac{3}{4})^{n_6-1-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s})^2$$

$$+ \frac{n_6(n_6-1)}{2}(n_5(\tfrac{1}{2}d_5)(\tfrac{1}{4})^2\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(\tfrac{1}{4})^s$$

$$(\tfrac{3}{4})^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s})^2$$

$$+ n_7n_8(-\tfrac{1}{8}d_8)^2$$

TABLE 7.1 (cont.)

Epistatic Model

F_2 Generation ($p_0=p_1=0.5$)

$$\sigma^2_{AAA} = \frac{n_6(n_6-1)(n_6-2)}{3!}(2)^3(n_5(\tfrac{1}{2}d_5)(-1)(\tfrac{1}{4})^3 \sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-3}{s}(\tfrac{1}{4})^s \times$$

$$(\tfrac{3}{4})^{n_6-3-s}(-1)^{n_6-t-s}\binom{1}{n_6-t-s})^2$$

$$+ 2n_5\frac{n_6(n_6-1)}{2}(2)^2(-\tfrac{1}{2}a_5(\tfrac{1}{8})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(\tfrac{1}{4})^s \times$$

$$(\tfrac{3}{4})^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s})^2$$

$$\sigma^2_{AAD} = 2n_5\frac{n_6(n_6-1)}{2}(2)(-\tfrac{1}{2}a_5(\tfrac{1}{8})\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(\tfrac{1}{4})^s \times$$

$$(\tfrac{3}{4})^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s})^2$$

$$+ n_5\frac{n_6(n_6-1)}{2}(2)^2(-\tfrac{1}{2}d_5(\tfrac{1}{4})^2\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-2}{s}(\tfrac{1}{4})^s \times$$

$$(\tfrac{3}{4})^{n_6-2-s}(-1)^{n_6-t-s}\binom{2}{n_6-t-s})^2$$

$$+ \frac{n_6(n_6-1)(n_6-2)}{3!}(2)^2(n_5(\tfrac{1}{2}d_5)(-1)(\tfrac{1}{4})^3\sum_{t=\tau}^{n_6}\sum_{s=0}^{n_6-t}\binom{n_6-3}{s}(\tfrac{1}{4})^s \times$$

$$(\tfrac{3}{4})^{n_6-3-s}(-1)^{n_6t-s}\binom{3}{n_6-t-s})^2$$

TABLE 7.1 (cont.)

Epistatic Model

F_2 Generation ($p_0=p_1=0.5$)

$$\sigma^2_{ADD} = 2n_5 \frac{n_6(n_6-1)}{2} (-\tfrac{1}{2}a_5(\tfrac{1}{4})^2 \sum_{t=\tau}^{n_6} \sum_{s=0}^{n_6-t} \binom{n_6-2}{s} (\tfrac{1}{4})^s \times$$

$$(\tfrac{3}{4})^{n_6-2-s} (-1)^{n_6-t-s} \binom{2}{n_6-t-s})^2$$

$$+ n_5 \frac{n_6(n_6-1)}{2} (2)(-\tfrac{1}{2}d_5(\tfrac{1}{4})^2 \sum_{t=\tau}^{n_6} \sum_{s=0}^{n_6-t} \binom{n_6-2}{s} (\tfrac{1}{4})^s \times$$

$$(\tfrac{3}{4})^{n_6-2-s} (-1)^{n_6-t-s} \binom{2}{n_6-t-s})^2$$

$$+ \frac{n_6(n_6-1)(n_6-2)}{3!} (2)(n_5(\tfrac{1}{2}d_5)(-1)(\tfrac{1}{4})^3 \sum_{t=\tau}^{n_6} \sum_{s=0}^{n_6-t} \binom{n_6-3}{s} (\tfrac{1}{4})^s \times$$

$$(\tfrac{3}{4})^{n_6-3-s} (-1)^{n_6-t-s} \binom{3}{n_6-t-s})^2$$

$$\sigma^2_{DDD} = n_5 \frac{n_6(n_6-1)}{2} (-\tfrac{1}{2}d_5(\tfrac{1}{4})^2 \sum_{t=\tau}^{n_6} \sum_{s=0}^{n_6-t} \binom{n_6-2}{s} (\tfrac{1}{4})^s \times$$

$$(\tfrac{3}{4})^{n_6-2-s} (-1)^{n_6-t-s} \binom{2}{n_6-t-s})^2$$

$$+ \frac{n_6(n_6-1)(n_6-2)}{3!} (n_5(\tfrac{1}{2}d_5)(-1)(\tfrac{1}{4})^3 \sum_{t=\tau}^{n_6} \sum_{s=0}^{n_6-t} \binom{n_6-3}{s} (\tfrac{1}{4})^s \times$$

$$(\tfrac{3}{4})^{n_6-3-s} (-1)^{n_6-t-s} \binom{3}{n_6-t-s})^2$$

TABLE 7.2

Evaluation of Genotypic Variance Components for the Simple Linear, Multiple Linear, and Epistatic Gene Action Models for the F_2 Generation

	Gene Action Model		
	Simple Linear	Multiple Linear	Epistatic
μ	37.5	50	42.086
σ^2_G	64.0625	126.5625	113.1275
σ^2_A	50	93.75	81.4388
σ^2_D	14.0625	32.8125	29.4694
σ^2_I			2.2193
σ^2_{AA}			.8906
σ^2_{AD}			.8188
σ^2_{DD}			.2227
σ^2_{AAA}			.1367
σ^2_{AAD}			.0889
σ^2_{ADD}			.0444
σ^2_{DDD}			.0171
σ^2_E	60	60	60
σ^2_P	124.0625	186.5625	173.1275

epistatic model only the bottleneck subgenotype and the subgenotype which incorporates a major gene with modifiers will contribute to the epistatic variance components.

For the simple linear gene action model, 78 percent of the total genotypic variance is attributable to additive effects of alleles, while 74 percent of the genotypic variance is attributable to additive effects of alleles in the multiple linear model. The total genotypic variance is greater for the multiple linear gene action model than for the simple linear model. The total genotypic variance in the F_2 generation for the epistatic gene action model is less than for the multiple linear model. σ_A^2 accounts for approximately 72 percent of the total genotypic variance, σ_D^2 accounts for approximately 26 percent of σ_G^2, and the sum of all the epistatic variance components accounts for just under 2 percent of σ_G^2. For this epistatic gene action model, more than 98 percent of the total genotypic variation is due to nonepistatic effects of alleles and under circumstances such as this it would be very difficult to detect epistasis in an actual population. This type of situation may be part of the reason that there is little evidence supporting epistasis in yield of maize in the literature.

7.3 Variance Components for Gene Action Models with Linkage

In the previous section, the partitioning of the genotypic variance was accomplished using formulas developed under the assumption that the loci were not linked. Generalization of the previous formulas to include the situation of m-linked loci for symmetric gene action models has not been successful, but the formulas appropriate for some special cases

have been worked out. One special case is applicable to gene action models composed entirely of linear subgenotypes, such as the simple linear and multiple linear gene action models used in this simulation study.

The F_1 hybrid individuals in this recurrent selection program are all alike genotypically and hence the genotypic variance for this generation is zero. If we consider a situation with m-linked loci, the frequency of the gametes produced by these hybrid individuals will be a function of the linkage parameter. A more convenient linkage parameter, $\lambda = 1 - 2r$, will be used in this development of formulas to partition the genotypic variance in the presence of linkage (see Schnell, 1963). Recall from Chapter II that r is the recombination value between adjacent loci on a chromosome, and has possible values $0 \leq r \leq 0.5$. A value of $r = 0$ or $\lambda = 1.0$ corresponds to complete linkage of the loci and a value of $r = 0.5$ or $\lambda = 0.0$ corresponds to complete independence of the loci under consideration. Table 7.3 gives the frequencies of the different types of gametes produced for $m = 2$ and $m = 3$.

The F_2 genotypes and frequencies for $m = 2$ are given in Table 7.4. The genotypic value for each of these genotypes for two simple gene action models is also included. These two models are (a) a linear model,

$$G = a\, x_2 + d\, x_1 - a\, x_0 \quad , \tag{7.3}$$

and (b) an epistatic model,

$$G = (a\, x_{12} + d\, x_{11} - a\, x_{10})\, 0^{x_{20}} \quad . \tag{7.4}$$

TABLE 7.3

Frequencies of Gametes Produced by the F_1 Generation for m=2 and m=3

m=2		m=3	
Gamete	Freq.	Gamete	Freq.
+ +	$\frac{1+\lambda}{4}$	+ + +	$\frac{(1+\lambda)^2}{8}$
+ -	$\frac{1-\lambda}{4}$	+ + -	$\frac{1-\lambda^2}{8}$
- +	$\frac{1-\lambda}{4}$	+ - +	$\frac{(1-\lambda)^2}{8}$
- -	$\frac{1+\lambda}{4}$	- + +	$\frac{1-\lambda^2}{8}$
		+ - -	$\frac{1-\lambda^2}{8}$
		- + -	$\frac{(1-\lambda)^2}{8}$
		- - +	$\frac{1-\lambda^2}{8}$
		- - -	$\frac{(1+\lambda)^2}{8}$

TABLE 7.4

Genotypes, Genotypic Frequencies, and Genotypic Values for Two Gene Action Models for the F_2 Generation with m=2

Genotype	Freq.	Genotypic Value	
		Linear Model	Epistatic Model
$\frac{+\ +}{+\ +}$	$(\frac{1+\lambda}{4})^2$	2a	a
$\frac{+\ +}{+\ -}$	$\frac{1-\lambda^2}{8}$	a+d	a
$\frac{+\ +}{-\ +}$	$\frac{1-\lambda^2}{8}$	a+d	d
$\frac{+\ +}{-\ -}$	$\frac{(1+\lambda)^2}{8}$	2d	d
$\frac{+\ -}{+\ -}$	$(\frac{1-\lambda}{4})^2$	0	0
$\frac{+\ -}{-\ +}$	$\frac{(1-\lambda)^2}{8}$	2d	d
$\frac{+\ -}{-\ -}$	$\frac{1-\lambda^2}{8}$	d-a	0
$\frac{-\ +}{-\ +}$	$(\frac{1-\lambda}{4})^2$	0	-a
$\frac{-\ +}{-\ -}$	$\frac{1-\lambda^2}{8}$	d-a	-a
$\frac{-\ -}{-\ -}$	$(\frac{1+\lambda}{4})^2$	-2a	0

The values of a, d, and the x's are defined in Chapter II.

For the linear model, the mean and variance of the genotypic values are

$$\mu = d$$

and

$$\sigma_G^2 = (1+\lambda)a^2 + (1+\lambda^2)\frac{1}{2}d^2 \quad . \tag{7.5}$$

If $\lambda = 0$, the values of μ and σ_G^2 are

$$\mu = d$$

and

$$\sigma_G^2 = a^2 + \frac{1}{2}d^2 \ , \tag{7.6}$$

which are the values obtained when the loci are not linked. For m = 3, the mean and variance of the genotypic values are

$$\mu = \frac{3}{2}d$$

and (7.7)

$$\sigma_G^2 = (\frac{3}{2} + 2\lambda + \lambda^2)a^2 + (\frac{3}{2} + 2\lambda^2 + \lambda^4)\frac{1}{2}d^2 \quad .$$

For m = 4, the genotypic mean and variance are

$$\mu = 2d$$

and (7.8)

$$\sigma_G^2 = (2 + 3\lambda + 2\lambda^2 + \lambda^3)a^2 + (2 + 3\lambda^2 + 2\lambda^4 + \lambda^6)\frac{1}{2}d^2 \quad .$$

Although a formal proof has not been worked out, the previous expressions suggest that a general form for the genotypic mean and variance for a

linear model with m-linked loci is

$$\mu = \frac{m}{2} d$$

and (7.9)

$$\sigma_G^2 = (\frac{m}{2} + \sum_{i=1}^{m-1} (m-i)\lambda^i)a^2 + (\frac{m}{2} + \sum_{i=1}^{m-1} (m-i)\lambda^{2i}) \frac{1}{2} d^2 \quad , \quad m \geq 2$$

which reduces to

$$\mu = \frac{m}{2} d$$

and (7.10)

$$\sigma_G^2 = \frac{m}{2} a^2 + \frac{m}{4} d^2 \quad , \; m \geq 2$$

for complete independence of the loci (λ=0). In equations (7.9), the first term of σ_G^2 represents σ_A^2 and the second term σ_D^2. There are no epistatic variance components present in linear gene action models.

The genotypic variances is an increasing function of the number of linked loci and reaches a maximum for λ = 1.0. In this situation, the alleles linked together are transmitted as a unit to the next generation and there is no recombination of alleles at different loci. The maximum value for σ_G^2 is

$$\sigma_G^2 = (\frac{m}{2} + \sum_{i=1}^{m-1} (m-i))a^2 + (\frac{m}{2} + \sum_{i=1}^{m-1} (m-i)) \frac{1}{2} d^2$$

$$= (\frac{m}{2} + \frac{m(m-1)}{2}) (a^2 + \frac{1}{2} d^2)$$

$$= \frac{m^2}{2} (a^2 + \frac{1}{2} d^2) \; . \qquad (7.11)$$

Table 7.5 gives the values for the genotypic variance for the simple linear and multiple linear gene action models using both values of r which were investigated in the simulation program as well as linkage values representing complete linkage and complete independence. These values have been calculated assuming all 100 loci are linked together on the same chromosome. Obviously this is not the case, but for $\lambda = 0.6$ (r=0.2), loci which are more than 10 spaces apart on the chromosome are for all practical purposes independent. For the simple linear model, the genotypic variance under tight linkage is over four times the value for loose linkage and for the multiple linear model, the genotypic variance with tight linkage is almost 3.5 times the value for loose linkage. The genotypic variance for either model with tight linkage is less than one-half of the maximum possible value.

In both the simple linear and multiple linear gene action models in the range of recombination values used in this investigation, the proportion of total genotypic variance which is accounted for by additive effects of alleles increased from 78 to 87 percent for the simple linear model and from 74 to 84 percent for the multiple linear model. Thus, as the total genotypic variance increases with an increase in λ (a decrease in r), the proportion of σ^2_G attributable to σ^2_A also increases.

Extension of the formulas to partition the genotypic variance under linkage and a linear gene action model to future generations in this recurrent selection program has been done for the case of two linked loci and random mating of the population each generation. For m = 2, the gametic frequencies differ from the case with no linkage by a

TABLE 7.5

Genotypic Variance for the Simple Linear and Multiple Linear Gene Action Models for Complete Independence (r=0), the Values of RV Used in the Simulation Study (r=0.2 and r=0.05), and Complete Linkage (r=1)

Simple Linear Model

	λ=0.0	λ=0.6	λ=0.9	λ=1.0
σ^2_A	50	196.25	860	5000
σ^2_D	14.0625	29.6356	127.6514	1406.25
σ^2_G	64.0625	225.8856	987.6514	6406.25

Multiple Linear Model

	λ=0.0	λ=0.6	λ=0.9	λ=1.0
σ^2_A	93.75	346.8737	1154.706	2343.75
σ^2_D	32.8125	67.4194	253.9862	820.3125
σ^2_G	126.5625	414.2931	1408.6922	3164.0625

factor of $\frac{\lambda}{4}$ (see Table 7.2). Each generation of random mating reduces this disequilibrium value by a fraction $\frac{1+\lambda}{2}$. Therefore, the gametic frequencies for the F_k generation are

$$\begin{aligned} f(+,+) &= \frac{1}{4} + \frac{\lambda}{4}\left(\frac{1+\lambda}{2}\right)^{k-2} \\ f(+,-) &= \frac{1}{4} - \frac{\lambda}{4}\left(\frac{1+\lambda}{2}\right)^{k-2} \\ f(-,+) &= \frac{1}{4} - \frac{\lambda}{4}\left(\frac{1+\lambda}{2}\right)^{k-2} \\ f(-,-) &= \frac{1}{4} + \frac{\lambda}{4}\left(\frac{1+\lambda}{2}\right)^{k-2} \end{aligned} \tag{7.12}$$

for $k \geq 2$. In the case of complete linkage, $\lambda = 1.0$, the recombinant gametes (+,-) and (-,+) have frequencies of 0.0, and only the parental gametes (+,+) and (-,-) are produced each generation.

The genotypic mean and variance of the F_3 generation are

$$\mu = d$$

and (7.13)

$$\sigma_G^2 = \left(1 + \lambda\left(\frac{1+\lambda}{2}\right)\right)a^2 + \left(1 + \lambda^2\left(\frac{1+\lambda}{2}\right)^2\right)\frac{1}{2}d^2 .$$

By induction, we find that the values of the genotypic mean and variance for the F_k generation are

$$\mu = d$$

and (7.14)

$$\sigma_G^2 = \left(1 + \lambda\left(\frac{1+\lambda}{2}\right)^{k-2}\right)a^2 + \left(1 + \lambda^2\left(\frac{1+\lambda}{2}\right)^{2(k-2)}\right)\frac{1}{2}d^2 , \quad k \geq 2 .$$

If $\lambda = 0$ or $\lambda = 1$, the genotypic variance is constant for all generations of random mating. For $\lambda = 0$, equation (7.14) reduces to equation (7.6); and for $\lambda = 1$, equation (7.14) reduces to

$$\sigma_G^2 = 2a^2 + d^2 \ . \tag{7.15}$$

When $0 \leq \lambda < 1$, the factor $\frac{1+\lambda}{2}$ is less than 1 and the contribution of the effect of linkage to the genotypic variance approaches zero as the number of generations of random mating increases. Therefore, after a large number of generations of random mating, the effect of linkage on the genotypic variance is zero. Although the genotypic variance is a function of the linkage parameter, λ, the genotypic mean for the linear model is independent of λ.

One other gene action model was investigated for the effects of linkage on the genotypic mean and variance. In this very simple epistatic model (see equation (7.4)), one locus follows a linear model and the second locus acts as a control. This is the very simpliest form of a bottleneck model. The genotypic values for the F_2 generation are given in Table 7.3. The genotypic mean and variance are

$$\mu = \frac{3}{8} d + \frac{\lambda}{4} a + \frac{\lambda^2}{8} d$$

and (7.16)

$$\sigma_G^2 = \frac{3}{8} a^2 + \frac{15}{64} d^2 - \frac{3}{16} \lambda^2 a^2 + \frac{1}{32} \lambda^2 d^2 - \frac{1}{64} \lambda^4 d^2 - \frac{3}{16} \lambda ad - \frac{1}{16} \lambda^2 ad \ .$$

Evaluation of μ and σ_G^2 without linkage gives

$$\mu = \frac{3}{8} d$$

and (7.17)

$$\sigma_G^2 = \frac{3}{8} a^2 + \frac{15}{64} d^2 .$$

Thus, in the epistatic model, the linkage among loci affects both the mean and the variance of the genotypic values. In the F_3 generation,

$$\mu = \frac{3}{8} d + \frac{\lambda}{4} \left(\frac{1+\lambda}{2}\right) a + \lambda^2 \left(\frac{1+\lambda}{2}\right)^2 d . \quad (7.18)$$

In general, for the k^{th} generation,

$$\mu = \frac{3}{8} d + \frac{\lambda}{4} \left(\frac{1+\lambda}{2}\right)^{k-2} a + \frac{1}{2} \left(\frac{\lambda}{4}\right)^2 \left(\frac{1+\lambda}{2}\right)^{2(k-2)} d . \quad (7.19)$$

The expression for σ_G^2 is complicated and attempts to generalize to the k^{th} generation have not been successful. If the number of generations of random mating were large and $\lambda < 1.0$, then μ would approach $\frac{3}{8} d$. Even though the mean is inflated by the linkage parameter, the effects of linkage decrease with continued random mating.

While an extension of the formulas developed for partitioning the genotypic variance for symmetric gene action models has been limited to very simple models, one of these simple models is applicable to two of the three gene action models considered in this simulation study. The original selection experiment was designed in such a way that in the F_2 generation only two alleles were possible at each heterozygous locus and each allele had a frequency of 0.5. Thus a partition of σ_G^2 for a

linear gene action model with allelic frequencies equal to 0.5 can provide information concerning the effect of linkage on the values of σ^2_A and σ^2_D in the F_2 generation of this selection experiment. Generalization of the effects of linkage to more complex situations and other symmetric gene action models could provide useful information for later generations of this selection experiment as well as other types of selection programs.

Chapter VIII

CONCLUSIONS AND FURTHER RESEARCH

As an investigative tool, simulation has been shown to be of value for use with complex mathematical problems. In this thesis, we have applied Monte Carlo techniques to the investigation of a specific genetic recurrent selection experiment (Genter, 1976). The simulation of a biological problem involves the development of a model and repeated runs of the model on a high speed computer. In the development of the model to be simulated, an attempt was made to cover a wide range of possible values for each parameter with the actual biological situation somewhere in between. Simplifying assumptions were made and where possible, values for the parameters in the model were chosen according to the results of experimental investigations.

The original experiment had two major objectives: (1) the increase in the frequency of desirable alleles in the population contributing to yield and (2) the development of a high yielding inbred line of maize. Of particular interest from the simulation study was the behavior of yield and the coefficient of inbreeding over a large number of cycles of selection and the fixation in the population of desirable alleles.

Simulated observations for this selection experiment were averaged over all simulation runs and compared with actual observations from the first four cycles of selection of this experiment for yield and coefficient of inbreeding. In the early cycles of selection there is a large amount of genetic variance which is evident from the comparisons of experimental and simulated values. The gene action models used in

this study were chosen because their behavior gave results similar to those of the original experiment and supportive evidence was available from the literature. A more accurate assessment of the model can be made as observations from additional cycles of selection become available.

Investigation of the behavior of the favorable allele (+) at each locus provides information concerning the fixation of alleles in a situation where the selective advantage of the + allele is large. In the absence of a theoretical explanation for the probability of fixation of alleles under the circumstances of this model, simulation can help identify those situations which are favorable to the fixation of the alleles with a large selective advantage.

In the simple linear model, none of the desirable alleles were lost after 30 cycles of selection. In some cases the population was completely homozygous for the + allele, but in other cases there were still loci which were segregating in the population. Several of the situations with the multiple linear gene action model resulted in the fixation of the less desirable allele; however, this occurred for the subgenotype with the smallest contribution to the genotypic value. More loci in the epistatic model were fixed for the - allele, but again these loci were assigned to subgenotypes which contribute at most a small amount to the genotypic value. As indicated in Section 5.3, some of the loci in the epistatic subgenotypes were fixed for the - allele or still segregating, but were not contributing to the genotypic value.

From the simulation results, we find that it is possible for the population to achieve the maximum genotypic value in a recurrent selec-

tion program. In the linear models, the maximum genotypic value was reached when exact identification of the genotype was possible ($\sigma_E^2=0$). With an environmental contribution ($\sigma_E^2=60$), values of yield which exceed the genotypic maximum can occur; however, this does not necessarily indicate that all loci are fixed for the + allele. Large positive environmental contributions to yield will increase the average yield over the average genotypic value.

In the absence of environmental effects, population improvement will continue until the population becomes completely homozygous. The maximum genotypic value is reached if all of the loci are homozygous for the more desirable allele. Even if some loci become fixed for the less desirable allele, once the population becomes completely homozygous, no further improvement with selection is possible.

All of the design parameters considered in this model affect the rate of population improvement. A high selection intensity (N=20), loose linkage (r=0.2), fewer loci which are heterozygous in the F_1 ($n_h=50$), no environmental effects ($\sigma_E^2=0$), and linear gene action models (M1 and M2) contribute to the most rapid increase in yield. Selection of a smaller proportion of the population (a selection intensity of 11% with N=20) will result in a more rapid increase in yield without losing the desirable alleles at loci contributing to yield. Only under extreme conditions of tight linkage (r=0.05) and all loci heterozygous in the F_1 ($n_h=100$) should there be a concern with the fixation of less desirable alleles. This situation appears to occur more often for the epistatic subgenotypes in M3.

The coefficient of inbreeding in this experiment, as measured by the proportion of homozygous loci in the population each cycle, increases very slowly for the first few cycles of selection. Thereafter, the rate of increase depends upon the values of the other factors. For most situations investigated, the behavior of percent homozygosity followed a cubic function of cycle number. In the early cycles of selection, the selected crosses are more heterozygous than the population as a whole. Partial dominance of the genotype and a large number of loci contributing to yield will account for the amount of heterozygosity. After several cycles of selection, percent homozygosity begins to increase more rapidly for the selected crosses than for the population as a whole.

The behavior of percent homozygosity in the population for this selection program under the gene action models used is not at all comparable to the one-locus coefficient of inbreeding. Even though these values from the actual experiment and the simulation output compared favorably for the first four cycles of selection; after cycle one, the coefficient of inbreeding increases more rapidly than the value of percent homozygosity. An attempt to develop a single-valued measure of the coefficient of inbreeding for m loci was not successful.

For the duration of this selection study, the behavior of yield followed a quadratic function of cycle number. Since there is an upper limit for the genotypic value in the population, the quadratic function would not be adequate over a much longer period of time than the 30 cycles of selection of the simulation investigation. Under those circumstances, an exponential function might be more appropriate.

Inclusion of an epistatic gene action model in the investigation gives some insight into the effect of epistatic variance components. The epistatic variance is a very small proportion of the total genotypic variance and would be difficult to measure in an experimental situation with a large environmental error. Linkage increases the genotypic variance, but formulas for the partitioning of variance components for symmetric gene action models with linkage were possible only for extremely simple models.

Further research on this problem could take two different approaches: (1) research related to specific experiments and (2) research related to the effects of linkage on genetic variance components. As new observations become available, further comparisons could be made with the simulated results. More theoretical investigations need to be made in an attempt to describe the behavior of variables over time under complex models. The simulation results can be used as a reference for the theoretical expressions and may be of help in deriving approximations which would be applicable to selection experiments with a large selective advantage of the desirable allele.

The effects of linkage on the values of genetic variance components for symmetric gene action models will help in the understanding of how linkage affects different types of variance components for different types of models. For example, linkage increases the genotypic variance, but do all variance components increase proportionally or do some increase more rapidly than others? What happens to the epistatic variance when linkage is present? If linkage is high, is the epistatic variance proportionally higher? If so, then the detection of the presence of

epistasis in the population might be easier.

Correlation of the type of gene action to the results of an actual experiment involves an understanding of the behavior of variance components and how these can be estimated. Knowledge of the behavior of genotypic variance components in special cases will aid in the interpretation of the estimation of genotypic variance components from an experiment and will facilitate the interpretation of these experimental results. For example, the relationship of the change in quantities such as general and specific combining ability to the change in genetic variance components throughout a selection program could be investigated. An understanding of this relationship might be of value in the design of selection experiments.

BIBLIOGRAPHY

1. Allard, R.W., *Principles of Plant Breeding*, John Wiley and Sons, Inc., New York, 1960.

2. Anderson, V.L., and O. Kempthorne, "A Model for the Study of Quantitative Inheritance," *Genetics*, Vol. 39:280-295, 1954.

3. Bauman, Loyal F., "Evidence of Non-allelic Gene Interaction in Determining Yield, Ear Height, and Kernal Row Number in Corn," *Agronomy Journal*, 51:531-534, 1959.

4. Box, G.E.P., and Mervin E. Muller, "A Note on the Generation of Random Normal Deviates," *Annals of Mathematical Statistics*, 29:610-611, 1958.

5. Cain, R.L., and K. Hinkelmann, "Coefficients of Inbreeding and Homozygosity in Recurrent Selection: The One-Locus Case," *Theoretical and Applied Genetics*, 40:327-335, 1970.

6. Cain, R.L., and K. Hinkelmann, "Coefficients of Inbreeding and Homozygosity in Recurrent Selection: The Case of m-Linked Loci," *Theoretical and Applied Genetics*, 42:196-207, 1972.

7. Choy, S.C., and B.S. Weir, "Two Locus Inbreeding Measures for Recurrent Selection," N.C. Agriculture Experiment Station, 1976.

8. Cockerham, C. Clark, "Implications of Genetic Variances in a Hybrid Breeding Program," *Crop Science*, 1:47-52, 1961.

9. Compton, W.A., C.O. Gardner, and J.H. Lonnquist, "Genetic Variability in Two Open-Pollinated Varieties of Corn (Zea Mays L.) and Their F_1 Progenies," *Crop Science*, 5:505-508, 1965.

10. Comstock, R.E., and H.F. Robinson, "The Components of Genetic Variance in Populations of Biparental Progenies and Their Use in Estimation the Average Degree of Dominance," *Biometrics*, 4:254-266, 1948.

11. Comstock, R.E., and H.F. Robinson, Chapter 30, *Heterosis*, John W. Gowen, ed., Hafner Publishing Company, New York, 1952.

12. Crosby, Jack L., *Computer Simulation in Genetics*, John Wiley and Sons, Inc., New York, 1973.

13. Crow, James F., and Motoo Kimura, *An Introduction to Population Genetics Theory*, Harper and Row, New York, 1970.

14. Falconer, D.S., *Introduction to Quantitative Genetics*, The Ronald Press Company, New York, 1960.

15. Fraser, A.S., "Simulation of Genetic Systems by Automatic Digital Computers: I. Introduction," *Aust. J. Biol. Sci.*, 10:484-491, 1957.

16. Fraser, A.S., "Simulation of Genetic Systems by Automatic Digital Computers: II. Effects of Linkage of Rates of Advance Under Selection," *Aust. J. Biol. Sci.*, 10:492-499, 1957.

17. Fraser, A.S., "Simulation of Genetic Systems by Automatic Digital Computers: VI. Epistasis," *Aust. J. Biol. Sci.*, 13:150-162, 1960.

18. Fraser, A.S., "Simulation of Genetic Systems by Automatic Digital Computers: VII. Effects of Reproductive Rate and Intensity of Selection on Genetic Structure," *Aust. J. Biol. Sci.*, 13:344-350, 1960.

19. Fraser, A.S., Chapter 8, *Biometrical Genetics*, O. Kempthorne, ed., Pergammon Press, New York, 1960.

20. Fraser, A.S., and Donald Burnell, *Computer Models in Genetics*, John Wiley and Sons, Inc. New York, 1973.

21. Gardner, C.O., et al., "Dominance of Genes Controlling Quantitative Characters in Maize," *Agronomy Journal*, 45:186-191, 1953.

22. Gardner, C.O., and J.H. Lonnquist, "Linkage and the Degree of Dominance of Genes Controlling Quantitative Characters in Maize," *Agronomy Journal*, 51:524-528, 1959.

23. Genter, C.F., "Inbreeding Without Inbreeding Depression," Proc. 22nd Hybrid Corn Industry Research Conference American Trade Association, Washington, D.C., Dec., 1967.

24. Genter, C.F., "Yields of S_1 Lines from Original and Advanced Synthetic Varieties of Maize," *Crop Science*, 11:821-824, 1971.

25. Genter, C.F., "Recurrent Selection for Yield in the F_2 of a Maize Single Cross," *Crop Science*, 16:350-352, 1976.

26. Genter, C.F., and S.A. Eberhart, "Performance of Original and Advanced Maize Populations and Their Diallel Crosses," *Crop Science*, 14:881-885, 1974.

27. Gill, J.L., "Effects of Finite Size on Selection Advance in Simulated Genetic Populations," *Aust. J. Biol. Sci.*, 18:599-617, 1965.

28. Gill, J.L., "A Monte Carlo Evaluation of Predicted Selection Response," *Aust. J. Biol. Sci.*, 18:999-1007, 1965.

29. Gill, J.L., "Selection and Linkage in Simulated Genetic Populations," *Aust. J. Biol. Sci.*, 18:1171-1187, 1965.

30. Gill, J.L., and B.A. Clemmer, "Effects of Selection and Linkage on Degree of Inbreeding," *Aust. J. Biol. Sci.*, 19:307-317, 1966.

31. Gorsline, G.W., "Phenotypic Epistasis for Ten Quantitative Characters in Maize," *Crop Science*, 1:55-58, 1961.

32. Hill, William G., "Properties of Simple Models of Quantitative Gene Action for Monte Carlo Studies," Technical Report Number MC 3, 1963.

33. Hill, William G., et al.,"Components of Variance for Theoretical Models of Quantitative Gene Action," Technical Report Number MC 4, 1963.

34. Hill, William G., and Alan Robertson, "The Effect of Linkage on Limits to Artificial Selection," *Genet. Res.*, 8:269-294, 1966.

35. Horner, T.W., "The Components of Variance in Symmetrical Random Mating Populations with the Frequency of the More Favorable Allele the Same at all Loci," *Iowa State College Journal of Science*, 31, 1:67-77, 1956.

36. Horner, T.W., and O. Kempthorne, "The Components of Variance and the Correlations Between Relatives in Symmetrical Random Mating Populations," *Genetics*, 40:310-320, 1955.

37. Hull, Fred T., "Recurrent Selection for Specific Combining Ability in Corn," *J. Am. Soc. Agron.*, 37:134-145, 1945.

38. Hull, Fred T., Chapter 28, *Heterosis*, John W. Gowen, ed., Hafner Publishing Company, New York, 1952.

39. Jenkins, Merle T., "The Segregation of Genes Affecting Yield of Grain in Maize," *J. Am. Soc. of Agron.*, 32:55-63, 1940.

40. Jinks, J.L., "A Survey of the Genetical Basis of Heterosis in a Variety of Diallel Crosses," *Heredity*, 9:223-238, 1955.

41. Kempthorne, O., *An Introduction to Genetic Statistics*, The Iowa State University Press, 1957.

42. Kimura, Motoo, "Process Leading to Quasi-Fixation of Genes in Natural Populations Due to Random Fluctuation of Selection Intensities," *Genetics*, 39:883-898, 1954.

43. Kimura, Motoo, "On the Probability of Fixation of Mutant Genes in a Population," *Genetics*, 47:713-719, 1962.

44. Li, C.C., *Population Genetics*, The University of Chicago Press, 1955.

45. Lindsey, M.F., et al., "Estimates of Genetic Variance in Open-Pollinated Varieties of Cornbelt Corn," *Crop Science*, 2:105-108, 1962.

46. Lonnquist, John, "Heterosis and Yield of Grain in Maize," *Agronomy Journal*, 45:539-542, 1953.

47. Madalena, F.E., and W.G. Hill, "Population Structure in Artificial Selection Programmes: Simulation Studies," *Genetical Research*, 20:75-99, 1972.

48. Malécot, G., *Les mathématiques de l'hérédité*, Masson et Cie, Paris, 1948.

49. Martin, Frank G., Jr., and C. Clark Cockerham, Chapter 5 *Biometrical Genetics*, O. Kempthorne, ed., Pergammon Press, New York, 1960.

50. Moll, R.H., et al., "Estimates of Genetic Variances and Level of Dominance in Maize," *Genetics*, 49:411-423, 1964.

51. Moll, R.H., and H.F. Robinson, "Quantitative Genetic Investigations of Yield of Maize," *Der Zuchter*, 37:192-199, 1967.

52. Neter, John, and William Wasserman, *Applied Linear Statistical Models*, Richard D. Irwin, Inc., 1974.

53. Penney, L.H., and S.A. Eberhart, "Twenty Years of Reciprocal Recurrent Selection with Two Synthetic Varieties of Maize," *Crop Science*, 11:900-903, 1971.

54. Penney. L.H., et al., "Types of Gene Action in Yield Heterosis in Maize," *Crop Science*, 2:341-344, 1962.

55. Qureshi, A.W., and O. Kempthorne, "On the Fixation of Genes of Large Effects Due to Continued Truncation Selection in Small Populations of Polygenic Systems with Linkage," *Theoret. App. Gen.*, 38:249-255, 1968.

56. Qureshi, A.W., et al., "The Role of Finite Population Size and Linkage in Response to Continued Truncation Selection: I. Additive Gene Action," *Theoret. App. Gen.*, 38:256-263, 1968.

57. Qureshi, A.W., "The Role of Finite Population Size and Linkage in Response to Continued Truncation Selection: II. Dominance and Overdominance," *Theoret. App. Gen.*, 38:264-270, 1968.

58. Richey, Frederick D., "Hybrid Vigor and Corn Breeding," *Jour. Am. Soc. Agron.*, 38:833-841, 1946.

59. Robertson, Alan, "Inbreeding in Artificial Selection Programmes," *Genet. Res.*, 2:189-194, 1961.

60. Robinson, H.F., et al., "Estimates of Heritability and the Degree of Dominance in Corn," *Agronomy Journal*, 41:353-359, 1949.

61. Robinson, H.F., et al., "Genetic Variances in Open-Pollinated Varieties of Corn." *Genetics*, 40:45-60, 1955.

62. Robinson, H.F., et al., "Dominance vs. Overdominance in Heterosis: Evidence from Crosses Between Open-Pollinated Varieties of Maize," *Am. Nat.*, 90:127-131, 1956.

63. Robinson, H.F., et al., "Joint Interpretation of Heterosis and Genetic Variances in Two Open-Pollinated Varieties of Corn and Their Cross," *Genetics*, 43:868-877, 1958.

64. Schnell, F.W., "The Covariance Between Relatives in the Presence of Linkage," Statistical Genetics and Plant Breeding Natl. Acad. Sci. Natl. Res. Council Pub 982, 468-483.

65. Shikata, Morikazu, "The Generalized Inbreeding Coefficient in a Markov Process," *JUSE* 9:1-10, 1962.

66. Sprague, G.F., Chapter 26, *Heterosis*, John W. Gowen, ed., Hafner Publishing Company, New York, 1952.

67. Sprague, G.F., and Philip A. Miller, "A Suggestion for Evaluating Current Concepts of the Genetic Mechanism of Heterosis in Corn," *Agronomy Journal*, 42:161-162, 1950.

68. Sprague, G.F., et al., "Additional Studies of the Relative Effectiveness of Two Systems of Selection for Oil Content of the Corn Kernel," *Agronomy Journal*, 44:329-331, 1952.

69. Sprague, G.F., et al., "Effect of Epistasis on Grain Yield in Maize," *Crop Science,* 2:205-208, 1962.

70. Stuber, C.W., et al., "Genetic Variances and Interralationships of Six Traits in a Hybrid Population of Zea Mays L.," *Crop Science,* 6:455-458, 1966.

71. Weir, B.S., and C. Clark Cockerham, "Group Inbreeding with Two Linked Loci," *Genetics,* 63:711-742, 1969.

72. Weir, B.S., and C. Clark Cockerham, "Pedigree Mating with Two Linked Loci," *Genetics*, 61:923-940, 1969.

73. Young, S.S.Y., "Computer Simulation of Directional Selection in Large Populations: I. The Programme, The Additive, and The Dominance Models," *Genetics,* 53:189-205, 1966.

APPENDIX

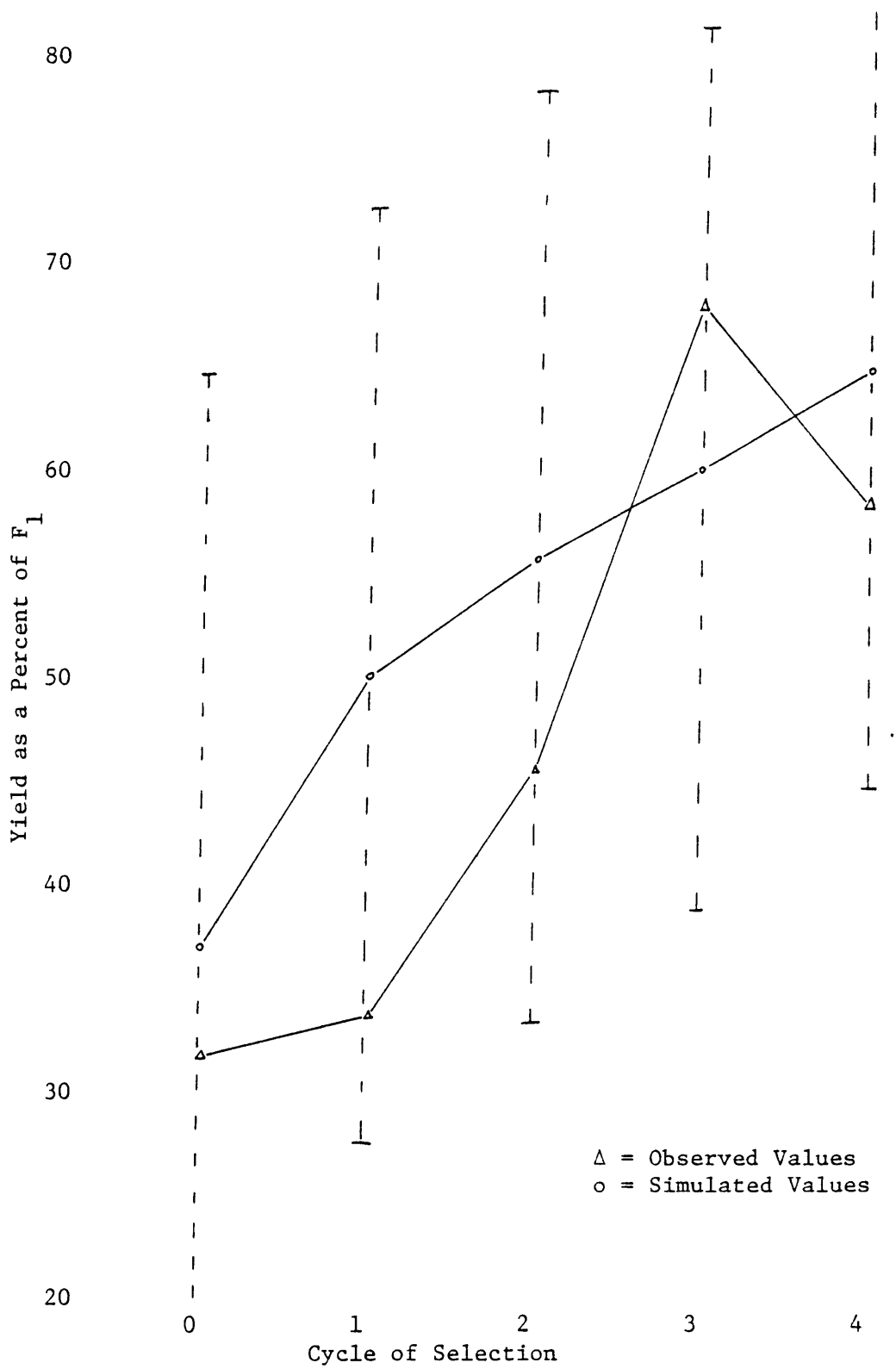


Figure 1 Tolerance Intervals on the Observed Values for Low Yield in Selection Cycles 0-4

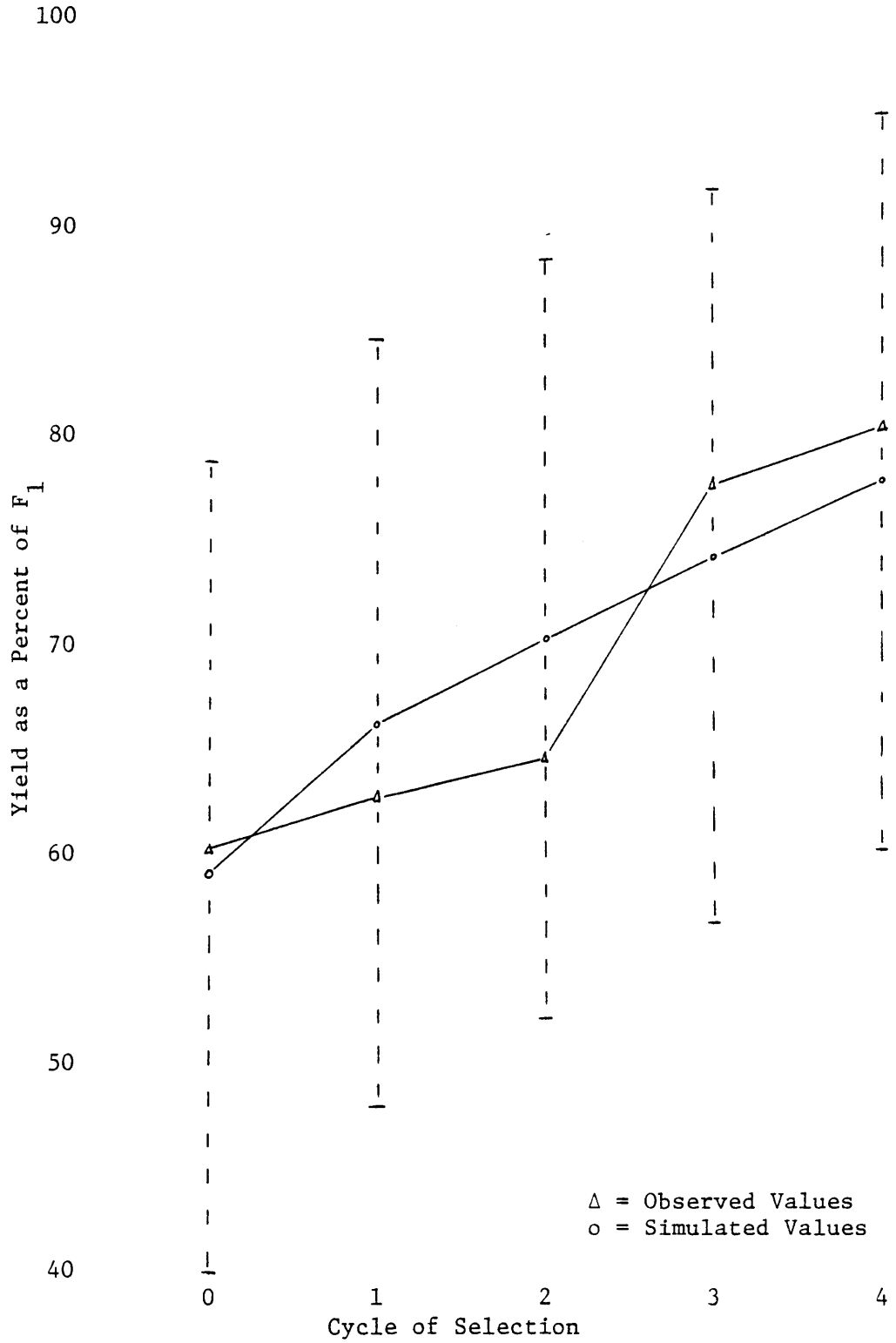


Figure 2 Tolerance Intervals on the Observed Values for Average Yield in Selection Cycles 0-4

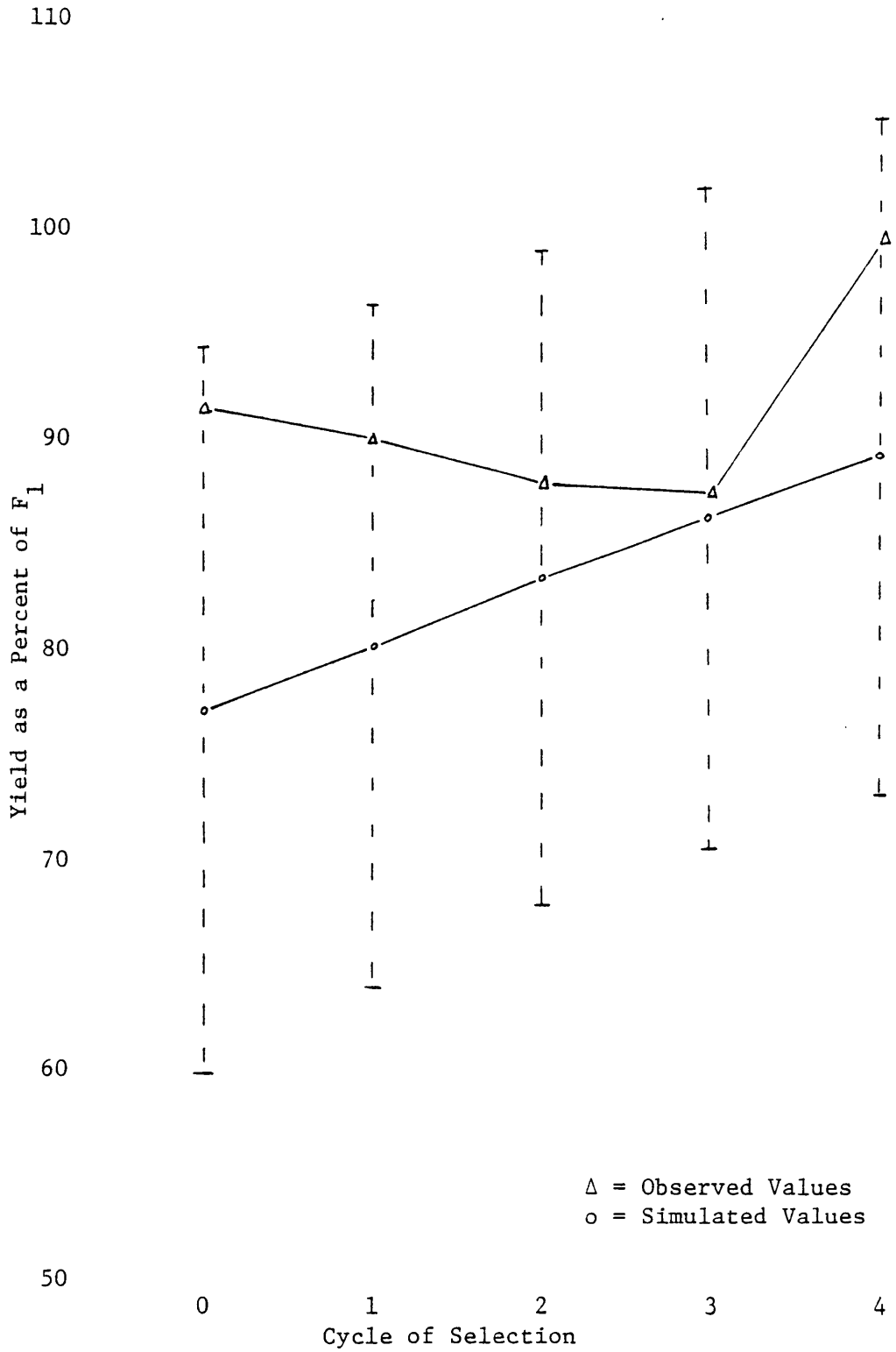


Figure 3 Tolerance Intervals on the Observed Values for High Yield in Selection Cycles 0-4

The vita has been removed from the scanned document

STOCHASTIC SIMULATION OF A SELECTION EXPERIMENT IN MAIZE

by

Sophronia W. Ward

(ABSTRACT)

A selection experiment in maize for increase in yield as described by Genter (1976. Recurrent selection for yield in the F_2 of a maize single cross. Crop Sci. 16:350-352) is simulated and the long term effects of selection intensity, recombination value, number of segregating loci, environmental variance, and gene action model on yield and approach to complete homozygosity are investigated. Selection intensities of 11% and 22%, recombination values of .2 and .05, 100 and 50 segregating loci, and environmental variances of 0 and 60 were chosen and the combinations used represent a 1/2-fraction of a 2^4 factorial design. Each of these 8 possible situations is combined with 3 gene action models to give 24 different situations for the simulation study. The first gene action model is linear with partial dominance for all loci. For the second and third gene action models, the genotype is partitioned into subgenotypes. The second gene action model has four subgenotypes, each of which is linear with partial dominance and different size effects. Six subgenotypes are combined in the third gene action model. Four of these are linear with partial dominance and different size effects, and the remaining two subgenotypes are epistatic.

In the simulation program, the genotype consists of 100 loci on 10 pairs of chromosomes, which contribute to the genotypic value

according to the gene action models used. Each run begins with the F_1 hybrid of two unrelated inbred populations and continues for 30 cycles of selection or until complete homozygosity of the genotype is reached. Genotypic and phenotypic means, genotypic and phenotypic variances, and percent homozygosity are obtained for each cycle of selection.

Compairson of experimental observations and simulated values averaged over all runs for the first four cycles of selection indicate the simulated results are in good agreement with the experimental values. Regression analyses of percent homozygosity, genotypic value, phenotypic value (yield), and heritability in the broad sense using cycle of selection as the independent variable are performed. In addition, the effects and 2-factor interactions of selection intenstiy, recombination value, number of segregating loci, environmental variance, and gene action model on approach to complete homozygosity and increase in yield are investigated using analysis of variance for cycles 5, 10, 15, 20, and the last cycle of selection.

A recurrence formula for average coefficient of inbreeding is developed for this selection scheme. Also, for each gene action model explicit expressions for partitioning the genotypic variance into additive, dominance, and all epistatic components are given and evaluated for the F_2 generation. Consequences of selection on general combining abilities are being investigated.