Statistical Techniques in QUANTITATIVE GENETICS

J P JAIN

Statistical Techniques in QUANTITATIVE GENETICS

J P JAIN
Indian Agricultural Statistics Research Institute
New Delhi

Tata McGraw-Hill Publishing Company Limited
NEW DELHI

Editorial Supervisor: Usha Tankha
Production Supervisor: M S Phogat

Published by Tata McGraw-Hill Publishing Company Limited
12/4 Asaf Ali Road, New Delhi 110 002 and
printed at Urvashi Press, 49/3 Vaidwara, Meerut 250 002

Preface

The present book grew out of the increasingly evident need for synthesizing the basic principles of quantitative genetics with the concepts and methods of animal breeding. It presents an up-to-date account of quantitative genetics in a logically integrated and practical form.

This book is intended to serve as a textbook of quantitative genetics for students majoring either in animal or plant breeding as also for those intending to pursue careers as professional statisticians with agricultural universities or such other organizations engaged in plant and animal breeding research. The application of the basic principles of quantitative genetics is restricted to problems in animal breeding and therefore those who intend to become specialists in animal breeding will find in this book all they require including an outline of breeding plans and techniques for measuring genetic progress. For others, barring a few chapters, it will serve as a general textbook on quantitative genetics.

A characteristic feature of this book is that the derivation of almost all the formulae are demonstrated in full for a better understanding of their implications and an appreciation of the subject. Accordingly it is hoped this presentation will help stimulate further interest in the subject and inspire a study of the original literature. In pursuit of this aim a comprehensive list of over 800 references to published works, including those not cited in the text but have a direct bearing on the topics covered, is included at the end of this book. To distinguish references to research articles from those to books, manuals, monographs and tables, the latter are indicated by asterisks in the text.

The level of presentation in this book presumes that the reader has had courses in statistics and population genetics.

In addition to the excellent books by Lerner, Lush, Kempthorne, Falconer, Mather, Mather and Jinks, Li and Crow and Kimura, I have, during my study of the subject and writing of this book, been influenced by the works of a large number of authors notably Dickerson, Cockerham, Robertson, Wright, Griffing, Searle and Henderson; I feel deeply grateful to them all.

I wish to express my special gratitude and indebtedness to Shri V. N. Amble, Ex-Chief Executive Officer, National Sample Survey Organization,

New Delhi for initiating me into the subject of statistical genetics and kindling interest for research in animal breeding. I would like to express my gratitude to Dr. Daroga Singh, Ex-Director, Indian Agricultural Statistics Research Institute, New Delhi for timely advice on writing such a textbook. I am also grateful to Dr. Prem Narain, Director, IASRI with whom I have been working for the last couple of years. The association with him has broadened my professional knowledge and experience which has been helpful in the present work.

Finally, I would like to express my thanks and appreciation to Prof. S. R. Searle and to the editors of *Heredity* and *J. Dairy Science* for kind permission to use the material in Tables 7.1 and 11.7 and in Fig. 10.1.

J P JAIN

Contents

1

Quantitative Inheritance

1.1 INTRODUCTION

Quantitative genetics is concerned with the inheritance of those characters that differ among individuals in degree rather than in kind, quantitative rather than qualitative. The familiar examples of characters that exhibit differences in degree are growth rates in many species, milk production in dairy cattle, fleece weight in sheep, etc. and the examples of characters that exhibit differences in kind are coat colour, major lethals, plumage colour, comb shape, winglessness, etc. In other words, the quantitative characters show a continuous range of variability from one extreme to the other without natural discontinuities. Qualitative characters, in contrast, divide individuals into distinct or clear-cut categories. Because the quantitative characters can be specified accurately only in terms of metrics such as length, time, weight, or proportion, they are also referred to as metrical characters. This chapter is chiefly concerned with the inheritance of quantitative characters with a brief description of the inheritance of qualitative characters and the distinction between the two, an understanding of which is of fundamental significance in the application of genetics to animal and plant breeding.

1.2 QUALITATIVE INHERITANCE

The qualitative traits are governed by one or a few pairs of genes which have phenotypic effects so large that genotypes can be classified accurately. Environmental differences have relatively little effect on these traits as compared with genetic differences. In view of the large enough effects of these genes, so as to permit their ready identification, they are sometimes referred to as major genes. The science of genetics of characters, controlled by major genes, is called classical or Mendelian genetics, after its founder (Mendel, 1866). The essence of Mendelism is that inheritance is by particles, called genes, and that these are present in pairs, one member of each pair having come from each parent. Further, each gene maintains its identity generation after generation. When the individual reproduces, it transmits to each off-

spring one or the other, but not both, of the genes in each pair it possesses. Thus, the parents give to each offspring only a sample half of their own inheritance. The laws of chance govern this sampling, subject to the restriction that each sample must contain one gene of every pair. All the genes thus show segregation, and those which are borne in the same chromosome show linkage with one another. The effects of these genes are so large that the phenotypes can be classed into a small number of easily recognizable, qualitative groups. The variation exhibited by qualitative traits is thus one of discontinuous type. The inheritance of qualitative traits is studied by comparing the observed numbers in different groups with the Mendelian ratios expected on the hypothesis about their mode of inheritance. This type of analysis of the data is spoken of as either Mendelian analysis or genetical analysis.

1.3 QUANTITATIVE INHERITANCE

Since quantitative characters, such as growth or milk production, cannot be explained in simple Mendelian ratios produced by a few genes, it remained a matter of great controversy for some time whether or not to accept the Mendelian particulate mode of inheritance to explain the continuous variation exhibited by quantitative traits. This sort of variation remained the object of active and vigorous study in the last part of the nineteenth century in England by Galton (1888) and his successor Karl Pearson (1897, 1905). These workers could establish, by means of correlation and regression studies on a number of characters in man and dogs, that such variation was at least partly heritable but failed to reconcile their results with the Mendelian mode of inheritance. In 1906, Yule suggested that continuous quantitative variation might be produced by a multitude of individual genes, each with a small effect on the measured character. Soon after Yule's proposal, the classical experiments of Nilsson-Ehle (1909), a Swedish plant breeder, on colour in wheat and that of East (1915), an American scientist, on flower size in tobacco together with the results of Johannsen's (1903, 1909) purelines experiments on seed weight of beans led to the formulation of a multiple factor hypothesis which explains the inheritance of quantitative traits. According to this hypothesis, the quantitative characters are governed by a large number of minor genes, which Mather (1943) called polygenes, which are inherited in accordance with the Mendelian principles, each having a small, similar and cumulative effect, and whose effects are highly susceptible to environmental modifications. Such inheritance is also called polygenic inheritance. It should be noted that the basic premise of the multiple factor hypothesis is that the genes controlling the quantitative traits are subject to the same laws of transmission and have the same general properties as the genes, whose transmission and properties are displayed by qualitative characters. However, owing to the simultaneous segregation of a number of genes in conjunction with the influence of environmental forces it is not possible to follow genes individually in their transmission from generation to genera-

tion. Nor is it possible, with polygenic inheritance, to determine the number of genes segregating in the offspring of a cross by the usual hybridization experiment. The reasons are that first, the effects of individual genes on the phenotype may be obscured by variation due to environment; and second, that the effects of the genes may not be simply additive.

An important question that arises in connection with polygenic inheritance is, how it can explain continuous variation when the genes controlling the character are inherited in the Mendelian fashion. There are two reasons for this: one is the simultaneous segregation of many genes affecting the character, and the other is the superimposition of truly continuous variation arising from non-genetic causes. Consider, for example, skin colour in humans which, say, is controlled by two unlinked genes of equal and additive effect (Davenport, 1913). Suppose that there is no dominance and the two allelomorphs of each gene are equally common. Then if the segregation of these genes were the only cause of variation, the genotypes and phenotypes of the different generations would be as follows:

P	Black skin *AABB*	×		White skin *aabb*
F_1		Medium (mulatto) *AaBb*		
F_2		1 *AABB*	Black	
		2 *AABb*	Dark	
		1 *AAbb*	Medium	1 Black
		2 *AaBB*	Dark	4 Dark
		4 *AaBb*	Medium	→ 6 Medium
		2 *Aabb*	Light	4 Light
		1 *aaBB*	Medium	1 White
		2 *aaBb*	Light	
		1 *aabb*	White	

Thus, in the F_2 generation with two genes there are five different phenotypic classes with a continuous gradation between white and black with an intermediate level of expression being the most common. If there were n gene pairs segregating, each with two alleles and without dominance, the frequencies of phenotypes in an F_2 can be calculated on the basis of binomial distribution: $(\frac{1}{2}A + \frac{1}{2}a)^{2n}$, the phenotypic expression is reckoned proportional to the number of capital letters in the genotype. With three genes segregating, there would be seven phenotypic classes, with four genes segregating, there would be nine phenotypic classes, and with n gene pairs there would be $(2n + 1)$ classes. Thus with the increase in the number of segregating genes, the number of phenotypic classes would increase and it would then be more difficult to recognize the discontinuities. The variation would nearly be continuous and would approach the normal frequency curve in view of

the binomial distribution tending to a normal distribution as n is increased. In addition, superimposition on this variation due to non-heritable causes will further smoothen out the small differences between the members of different classes and the resulting variation as we see would then be truly continuous.

Another important question that arises in connection with polygenic inheritance is that when Mendelian ratios are not exhibited by quantitative characters, how then are we sure that the genes controlling them are inherited in the Mendelian fashion? Mendelian inheritance, as is known, is characterized by two properties which the genes must show in their transmission from one generation to the next. These are segregation and linkage. If then, we can establish that the determinants of heritable continuous variation exhibit both these properties, we cannot avoid the conclusion that they are nuclear genes and conform to Mendelian principles. That the polygenes show both segregation and linkage can be seen from the following empirical considerations.

1.3.1 Segregation of Polygenes

The experiments dealing with inbred lines of a species and their crosses show that the variability within the F_1 population is comparable in magnitude to that within the two parental populations. This conforms to Mendelian expectation, since all individuals in a hybrid between two pure lines should be identical genetically and the variations in the parental and F_1 populations is therefore wholly non-heritable.

Further, F_2 generations are more variable than the parental and F_1 generations. Although the F_2 generation cannot be expected to be immune from environmental effects, there is no reason why it should be more susceptible to environment influence than the parental generations or the F_1. The excess of variability in F_2 can then be explained due to segregation and recombination of genes which is expected in the F_2 in accordance with Mendelian principles.

In the generations succeeding F_2, the variability of a family observed is either the same or less than that of the family from which it came, but not greater. This again is in conformity with the Mendelian expectations, as with the increase in homozygosity within the families due to continued self-fertilization, the genetic variability reduces in each succeeding generation. It can thus be concluded that polygenes show segregation in the way shown by Mendel's principles.

In addition, the empirical results with quantitative traits also show that the mean performance of F_1's from reciprocal crosses are alike. Male and female parents contribute equally to F_1 as would be expected with nuclear inheritance.

1.3.2 Linkage of Polygenes

Polygenes show linkage with major genes as well as with one another. The first evidence of linkage of polygenes and major genes is provided by

the results of the classical experiment which Sax (1923) conducted on dwarf beans. He made a cross between a strain with large coloured beans and another which had white beans. Seed size showed itself to be continuously variable in character but pigmentation proved to be due to a single gene difference, the F_2 giving a ratio of 3 : 1 of coloured and white beans. By growing F_3 progenies, the coloured F_2 plants were further classified into homozygotes and heterozygotes. These appeared in the ratio of 1 : 2. The average bean weight in the three classes of F_2 plants gave the following data (Table 1.1).

Table 1.1 Average weights of beans (in CG) from coloured and white members of an F_2

No. of plants	Colour constitution	Average bean weight
45	Coloured PP	30.7 ± 0.6
80	Coloured Pp	28.3 ± 0.3
41	pp	26.4 ± 0.5

The standard errors show the differences in average weights of seeds in the three classes to be significant—the PP plants had the largest seeds and the pp plants the smallest. Clearly the average weight is correlated with the number of P alleles present. This experiment, however, cannot be taken as giving conclusive evidence of linkage because the effect could be due to a pleiotropic secondary effect of P itself. For an experiment showing conclusive evidence of linkage between polygenes and major genes, see Rasmusson (1935).

Establishing the linkage relationships between the members of the polygenic system is quite complex as the members of such systems seldom produce individually identifiable phenotypic effects. Special techniques of chromosome assay, due to Morgan (1913*) are used in such studies. For an excellent account of the chromosome assay experiment conducted by Mather and Harrison (1949) which establishes the linkage between members of the polygenic system see Mather and Jinks (1971*, Chapter 1).

Thus, from the evidence of the properties, characteristic of nuclear borne genes which the polygenes also exhibit, we may conclude that the polygenes are subject to the same laws of transmission as the genes that are carried on the chromosomes.

1.4 METHODS OF STUDY OF QUANTITATIVE VARIATION

Although quantitative genetics is an extension of Mendelian genetics, the methods of study employed in the two are different. In qualitative genetics, the individuals can be classified into a number of distinct descriptive classes and as such the qualitative variation can be studied by simply counting the

numbers of individuals in different classes and testing their agreement with those expected on the hypothesis of Mendelian segregation. The methods most commonly employed in Mendelian analysis are binomial test, normal test and X^2 test. Since in quantitative inheritance, the phenotypes are not distinct and separate but exhibit continuous variation, the methods of Mendelian analysis are inappropriate. In quantitative variation, each individual is represented by a numerical value and consequently statistical methods available for the study of continuous variation would be appropriate. The properties of such populations as is well known are specified in terms of means, variances and covariances. And since the frequency distributions of most metric characters approximate more or less closely to normal curves, it is possible to make use of the properties of normal distribution and to apply the appropriate statistical techniques for drawing valid inferences.

At this stage a brief note on the historical development of quantitative genetics is due. The theoretical basis of this subject was established by the innovative work of R. A. Fisher in 1918 linking biometry and genetics, followed closely by the classical works of Sewall Wright (1921) and Haldane (1924–32). On the foundation laid by this great trio, Lush (1937*) and his co-workers built a coherent theory of scientific animal breeding programmes. In the succeeding years Kempthorne (1957*) contributed many statistical concepts and methods for research in quantitative genetics integrating the approaches of Fisher, Haldane, Wright, Malécot and Cockerham. Later notable contributors to the subject are too many to record individually without being biased.

1.5 QUANTITATIVE AND QUALITATIVE TRAITS

That both continuous and discontinuous variation is observed in characters, such as stature, indicates that the distinction between qualitative and quantitative characters is not absolute. Stature in animals and plants, it is true, is usually controlled by a polygenic system, but a single major gene may produce a dwarf animal or a dwarf plant. The difference between the two types of characters depends not only on the magnitude of effect of individual genes but also on the relative importance of heredity and environment in producing the final phenotype. It is therefore apparent that the key to progress in the analysis of quantitative characters lies in evaluating the relative contribution of these two causal agents to the total variability, the study of which forms the subject matter of several chapters.

2

Causal Components of Phenotypic Value

2.1 INTRODUCTION

Quantitative traits, as discussed in the previous chapter, are governed by many pairs of minor genes, called polygenes, having small and cumulative effects which are highly susceptible to environmental modifications. Polygenes acting together in a variety of ways in a given environment give rise to the expression of the trait the measure of which is the phenotypic value of that trait. A thorough knowledge of the different types of gene action involved and the subdivision of the phenotypic value into component parts attributable to different causes is important in the study of the genetic properties of populations.

2.2 BROAD CHARACTERIZATION OF PHENOTYPIC VALUE

The observed value of a trait measured on an individual is the phenotypic value of that individual. This is due in part to the genes which the individual receives from its parents and the environment in which it is raised. The particular assemblage of genes possessed by the individual is its genotype and the sum total effect of this assemblage of genes is called the genotypic value. The environment comprises all the non-genetic circumstances that influence the phenotypic value.

For considering the division of phenotypic value into components attributable to the influence of genotype and environment, let us assume that individuals of genotypes $G_1, G_2, \ldots, G_m$ are observed in a set of all types of environments $E_1, E_2, \ldots, E_n$ (strictly speaking a continuum) and that the phenotypic or observed value of G_i in E_j is x_{ij}. The phenotypic expression of these genotypes in different environments can be set out as in Table 2.1.

Table 2.1 Phenotypic values of different genotypes in different environments

Genotype	Environment						Average
	E_1	E_2	...	E_j	...	E_n	
G_1	x_{11}	x_{12}	...	x_{1j}	...	x_{1n}	$\bar{x}_{1.}$
G_2	x_{21}	x_{22}	...	x_{2j}	...	x_{2n}	$\bar{x}_{2.}$
⋮	⋮	⋮		⋮		⋮	⋮
G_i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{in}	$\bar{x}_{i.}$
⋮	⋮	⋮		⋮		⋮	⋮
G_m	x_{m1}	x_{m2}	...	x_{mj}	...	x_{mn}	$\bar{x}_{m.}$
Average	$\bar{x}_{.1}$	$\bar{x}_{.2}$	...	$\bar{x}_{.j}$	...	$\bar{x}_{.n}$	$\bar{x}_{..}$

We define

(i) $\bar{x}_{i.}$ as the average of the phenotypic values of genotype G_i over all possible environments as the genotypic value of that genotype;

(ii) $(\bar{x}_{.j} - \bar{x}_{..})$ as the environmental deviation corresponding to environment E_j; and

(iii) $(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})$ as the interaction between the ith genotype and jth environment in which the individual has been raised.

Any phenotypic value, x_{ij}, may then be represented by the following identity,

$$x_{ij} \equiv \bar{x}_{i.} + (\bar{x}_{.j} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})$$

This leads to the general formula

$$P = G + E + I_{GE} \tag{2.1}$$

where P is the phenotypic value, G is the genotypic value, E is the environmental deviation, and I_{GE} is the interaction between genotype and environment. Every characteristic, thus, is both hereditary and environmental and is the joint product of the two. The genes cannot develop the characteristic unless they have the proper environment, and no amount of attention to the environment will cause the characteristic to develop unless the necessary genes are present. If either the genes or the environment are changed, the characteristic which results from their interactions may be changed.

In many genetical studies it is assumed that environmental deviations and genotypic values are independent of each other; in other words that there is no interaction between genotypic value and environmental deviations. This assumption is not always justifiable. Some genotypes may perform more satisfactorily in one environment than they do in another. Very often some genotypes may even fail to exhibit the inherited conditions unless proper environment is provided. Thus, this type of interaction is quite

important but can be isolated and measured only under rather controlled populations. In normal circumstances, therefore, it is best regarded as part of the environment. Breeders who are concerned with the improvement of their stock can avoid the complication of genotype-environment interaction by producing and selecting breeding stock under the same conditions in which offspring will be produced. Biometricians, on the other hand, try to eliminate or minimize the effect of interaction by the appropriate choice of scale of measurement. For all practical purposes we, therefore, write

$$P = G + E \tag{2.2}$$

Since the mean environmental deviation in the population as a whole is zero, the mean phenotypic value is equal to the mean genotypic value. The term population mean then refers equally to the phenotypic as well as to genotypic value of the population.

2.3 ADDITIVE AND NON-ADDITIVE GENETIC EFFECTS

The many genes responsible for expression of quantitative traits may act in an additive or in a non-additive manner. The non-additive type of gene action can be of two sorts, the first sort is the interaction between allelic genes caused by dominance and the second sort is the interactions of genes which are not allelic to each other. For brevity these latter are called epistacy (or epistasis).

2.3.1 Additive Gene Action

If the action of the gene is like adding brick upon brick, as in the construction of a building, the action is called additive. Here adding one more brick (gene) makes exactly the same increase in the weight of a brick pile, regardless of the number of types of bricks the pile already contains. Each additional gene contributes equally and their effects accumulate.

More specifically, additive gene action means that the effect of substituting a gene for its allele is the same no matter what other genes are present. For example, the effect of substituting A_2 for A_1 is the same whether it occurs in a homozygote A_1A_1 or heterozygote A_1A_2. That is, $(y_{11} - y_{12}) = (y_{12} - y_{22})$, where y_{ij} is the genotypic value of A_iA_j genotype. This implies that dominance is absent. The effect is also the same regardless of the genetic constitution at other loci; hence there is no epistacy. Thus, additive gene action is synonymous with no dominance and no epistacy.

Average Effect

If genes act strictly in an additive fashion then the effect a gene has on the genotypic value is the effect of that gene substitution, and this effect is the same in all genotypes. But all genes do not act in this simple manner and therefore there is no way of knowing exactly what each gene does in every