

# Die Simulation quantitativer Merkmale durch Gene mit biochemisch definierbarer Wirkung

## I. Ein einfaches Modell

W. SEYFFERT

Institut für Genetik der Universität Tübingen

### Simulation of quantitative characters by genes with biochemically definable action

#### I. A simple model

**Summary.** The present paper contains the derivation of a simple model for estimating additive, dominant and epistatic contributions to a metric character by clearly defined genes. Formulas for 2 and 3 loci with 2 alleles each are given. If  $n$  loci are investigated, the model presupposes genetic material comprising all of the  $3^n$  possible genotypes on an isogenic background.

Interactions between genotypic components and environment are recognizable when the experiments are repeated in different environments.

By means of the estimated parameters the phenotypes can be clearly described.

Viele Eigenschaften höherer Organismen entziehen sich einer klassischen Mendel-Analyse, weil die in den Spaltungsnachkommenschaften zur Unterscheidung der Phänotypen erforderliche Diskontinuität scheinbar nicht gegeben ist. Statt dessen ist eine kontinuierliche Variation der Merkmale zu beobachten. Sie wird gemeinhin auf eine größere Anzahl von Genen mit mehr oder minder gleichartiger, kumulativer und oft stark umweltabhängiger Wirkung zurückgeführt (MATHER, 1949).

Viele dieser als quantitativ bezeichneten Merkmale haben für die Morphologie, Physiologie, Psychologie, Soziologie und auf jeden Fall für die Evolution der Organismen Bedeutung, manche von ihnen sind wirtschaftlich nutzbar. Sie verdienen daher unser unmittelbares Interesse.

Eine Eigentümlichkeit quantitativer Merkmale ist in ihrer Definition begründet: sie werden praktisch ohne Ausnahme mit den Einheiten unseres physikalischen Maßsystems erfaßt und beschrieben. Daher stellen sie manche Autoren als „metrische Eigenschaften“ den zumeist biochemisch definierbaren „qualitativen“ Merkmalen gegenüber.

Zur Untersuchung der Vererbungsweise metrischer Eigenschaften gibt es eine Reihe hochintellektueller biometrischer Methoden. Sie beruhen alle im Prinzip auf einer Aufgliederung der in verschiedenen Generationen vorgegebener Paarungssysteme gemessenen Varianzen in erbliche und nichterbliche Anteile und der erblichen Anteile in additive, dominanz- und interaktionsbedingte Komponenten (MATHER, 1949; COMSTOCK und ROBINSON, 1948, 1952; KEMPTHORNE, 1955; JINKS und HAYMAN, 1953, 1954). Der Aussagewert derartiger Analysenverfahren hängt von dem genetischen und mathematischen Modell ab, das ihnen zugrunde liegt. Es ist trivial aber notwendig zu erwähnen, daß ein mit Sicherheit komplexer Sachverhalt nur dann richtig erkannt und beschrieben werden kann, wenn das genetische Modell den natürlichen Gegebenheiten weitgehend entspricht und das

mathematische Modell mit einem Minimum an Vereinfachungen und Verallgemeinerungen auskommt.

In einem speziellen Fall konnte schon früher nachgewiesen werden, daß beide Bedingungen durchaus nicht immer erfüllt sind (SEYFFERT, 1960). Unsicherheiten in der Parameterschätzung und Fehlinterpretationen sind die Folge.

Es liegt daher nahe zu fragen, welches genetische Modell adäquat ist und welche Vereinfachungen und Verallgemeinerungen tragbar sind.

Ist es beispielsweise richtig anzunehmen, daß der Phänotyp vornehmlich von additiven Genwirkungen abhängt, daß die Inkremente verschiedener Gene gleiche Größenordnungen haben und daß nichtallele Interaktionen höherer Ordnung bedeutungslos sind und daher vernachlässigt werden können?

Antworten auf derartige Fragen sind nur aus Versuchen zu erhalten, in denen die an der Merkmalsbildung beteiligten Gene eindeutig definiert und mit geeigneten Methoden untersucht werden können. Merkwürdigerweise gibt es aber in der Literatur kaum Arbeiten, in denen die additiven, dominanz- und interaktionsbedingten Beiträge eines Systems definierter Gene zur Ausbildung eines Merkmals experimentell analysiert worden wären.

Das große und anspruchsvolle Gedankengebäude biometrischer Methoden zur Untersuchung der Vererbung quantitativer Merkmale ist demnach zumindest teilweise auf Annahmen errichtet worden, für die es keine relevanten Belege gibt.

Eindeutige Belege sind aber nur zu erwarten, wenn nicht nur die quantitativen, sondern zugleich auch die qualitativen Aspekte der Genwirkung untersucht werden; das heißt, daß es unerlässlich ist, biometrische Analysen durch biochemische Untersuchungen zu ergänzen.

In der mit der vorliegenden Arbeit beginnenden Reihe von Versuchen wird dieser neue Weg beschritten.

Ausgehend von der Einsicht, daß die Transskription und Translation der in der DNS enthaltenen Information bei allen Strukturgenen und zumindest bei einem Teil der regulierenden Gene in der gleichen Weise erfolgt, gibt es keinen Grund für die Annahme, daß zwischen Genen mit „qualitativer“ und „quantitativer“ Wirkung unterschieden werden muß.

Dementsprechend ist auch eine Einteilung in qualitative und quantitative Merkmale letztlich willkürlich. Sie ist lediglich der Ausdruck unseres Unvermögens, ein gegebenes Merkmal in einer Anzahl von Fällen anders als durch physikalische Messungen zu beschreiben. Dieses Unvermögen, diese Unspezifität der Messungen verhindern es, die an der Merkmalsbildung beteiligten Gene als Individuen zu erkennen.

Eine Änderung der Meßmethodik sollte daher die Kryptomerie der Polymerie metrischer Eigenschaften aufheben und es erlauben, kontinuierlich variierende Merkmale mit Hilfe geeigneter biochemischer Analyseverfahren auf eine Serie diskontinuierlicher Unterschiede, d. h. auf definierbare Gene, zurückzuführen. Eine Möglichkeit, die Richtigkeit dieser Annahmen zu überprüfen, liegt in der Umkehrung der Aussage: es sollte dann auch nachzuweisen sein, daß eine Serie von Genen mit biochemisch definierter Wirkung ein quantitatives Merkmal simulieren kann.

### Das Modell

Die Untersuchung quantitativer Beiträge definierter Gene zu einem allen gemeinsamen metrischen Merkmal setzt voraus, daß die Allele der beteiligten Loci in sämtlichen möglichen Kombinationen vorliegen. Nur unter dieser Bedingung sind alle nicht-allelen Interaktionen zu erfassen.

Sind  $n$  Loci mit je zwei Allelen gegeben, so resultieren daraus  $3^n$  verschiedene Genotypen.

Der erste experimentelle Schritt zur Verwirklichung dieser Kombinationen besteht darin, zunächst alle  $2^n$  möglichen Homozygoten darzustellen. Aus der diallelen Kreuzung der Homozygoten gehen insgesamt  $4^n$  Kombinationen hervor, deren Zusammensetzung einer  $F_2$  bei freier Genkombination entspricht. Heterozygote Genotypen treten, durch reziproke Kombinationen und analoge Kreuzungen bedingt, mehrfach auf. Die Gesamtzahl wiederholter Typen ist  $4^n - 3^n$ .

Eine wichtige Voraussetzung für eine unverfälschte Aussage des Versuches ist die Elimination des unterschiedlichen Einflusses aller im Versuch nicht variierten Gene. Der zweite experimentelle Schritt muß demnach darauf gerichtet sein, die Variation des genetischen Hintergrundes — zumindest soweit er das betreffende Merkmal beeinflußt — zu reduzieren, d. h. den Hintergrund isogen zu machen. Nur unter diesen Voraussetzungen ist das nachfolgend dargestellte Schätzverfahren für die verschiedenen Parameter gültig.

Da zunächst die Absicht besteht, die mathematisch-genetischen Modelle bestehender biometrischer Verfahren zu beurteilen, muß das abzuleitende Modell

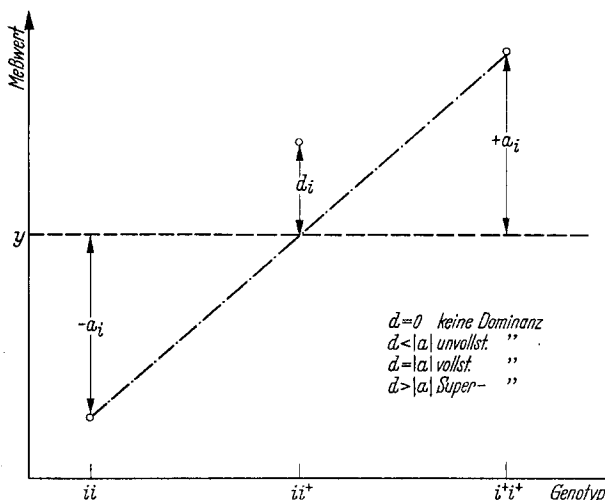


Abb. 1. Schematische Darstellung der additiven und dominanzbedingten Beiträge eines Allelenpaares.

den in der biometrischen Genetik üblichen homolog sein.

Die konventionelle Charakterisierung der Beiträge eines Allelenpaares ( $i, i^+$ ) erfolgt über die additive Komponente ( $a_i$ ), als die halbierte Differenz zwischen den Meßwerten der Homozygoten, und die Dominanzkomponente ( $d_i$ ), als die Differenz zwischen dem Meßwert der Heterozygoten und dem arithmetischen Mittel aus den Meßwerten der Homozygoten als Bezugsbasis ( $y$ ).

Diese Bezugsbasis ( $y$ ), häufig als Mittelelterwert bezeichnet, repräsentiert die Summe der Beiträge aller im Versuch nicht variierten Loci, d. h., den Beitrag des genetischen Hintergrundes.

Sofern sich die Untersuchung auf mehrere Loci erstreckt, sind zusätzlich nichtallele Interaktionen zu berücksichtigen. Grundsätzlich ist zwischen drei Interaktionstypen zu unterscheiden: nur additive Komponenten beteiligt ( $aa, aaa, \dots$ ), nur Dominanzkomponenten beteiligt ( $dd, ddd, \dots$ ) und Interaktionen zwischen Dominanz- und additiven Komponenten ( $ad, da, aad, ada, \dots$ ). Werden die Interaktionen mathematisch als Produkte dargestellt, deren Vorzeichen aus dem der beteiligten Faktoren resultieren, so ergibt sich für zwei Loci ( $i, i^+; j, j^+$ ) das nachfolgende Modell, das erstmals von CROW (unveröff.) in dieser Weise formuliert wurde:

$$\begin{pmatrix} iijj & iij+ & ii++ \\ i+jj & i+j+ & i+++ \\ ++jj & ++j+ & ++++ \end{pmatrix} = \begin{pmatrix} y-a-a+aa & y-a+d-ad & y-a+a-aa \\ y+d-a-da & y+d+d-dd & y+d+a-da \\ y+a-a-aa & y+a+d-ad & y+a+a-aa \end{pmatrix} \quad (1)$$

Diese 9 Gleichungen enthalten 9 Unbekannte. Somit kann jede Unbekannte eindeutig ermittelt und mit ihrer Hilfe jeder Phänotyp vollständig und unverwechselbar beschrieben werden.

Es muß aber vor dem Trugschluß gewarnt werden, man habe hiermit eine Möglichkeit, die biologische Wirkung der Gene und ihrer Kombinationen zu erfassen. Die mit Hilfe dieses Modells ermittelten Beiträge sind keinesfalls der unmittelbare Ausdruck der biologischen Aktivität der Gene und ihrer Primärprodukte. Alles, was dieses Modell leisten kann, ist eine exakte Beschreibung der Phänotypen auf Grund eines als additiv angenommenen Systems. Es ist ebensowenig wie andere Modelle geeignet, „Vorausagen über den Verlauf bestimmter biologischer Prozesse“ zu geben (BELLMANN und AHRENS, 1965).

Dennoch leistet dieses einfache Modell mehr als bisher verwendete, da es im Gegensatz zu anderen erlaubt, Beiträge einzelner Gene zum Phänotyp zu schätzen. Es ist von Koppelungserscheinungen unabhängig und kann theoretisch auf beliebig viele Loci ausgedehnt werden. Jedoch ist einer Erweiterung auf beispielsweise mehr als 4 Loci alsbald eine natürliche Grenze gesetzt, da das Gleichungssystem dann wegen der großen Zahl zu schätzender Parameter unhandlich wird.

### Das Schätzverfahren

Führt man zur Kennzeichnung der Meßwerte der Genotypen eine vereinfachte Schreibweise ein, indem man die Zahl der + Allele am jeweiligen Locus einsetzt, so gilt die Beziehung

$$\begin{pmatrix} iijj & iij+ & ii++ \\ i+jj & i+j+ & i+++ \\ ++jj & ++j+ & ++++ \end{pmatrix} = \begin{pmatrix} ij & ij & ij \\ 00 & 01 & 02 \\ 10 & 11 & 12 \\ 20 & 21 & 22 \end{pmatrix} \quad (1a)$$

$$a_i a_j a_k \quad \begin{array}{cccccccc} 8 & = & -1 & & 1 & & 1 & & -1 & & & & 1 & & -1 & & -1 & & 1 \end{array} \quad (5.8)$$

Wird die Ableitung in entsprechender Weise bis zur Diagonalform durchgeführt, resultieren daraus die Gleichungen (5). Aus Gründen der Platzersparnis werden sie in der unterteilten Form geschrieben. Die Symbole der  $a$ -,  $d$ - und Interaktionskomponenten auf der linken Seite der Gleichungen sind, die Diagonale der Unterteilung als Symmetrieachse benutzend, jeweils nach oben zu klappen.

Bei der Ableitung der vorstehenden Gleichungen wurden ausschließlich genetische Komponenten berücksichtigt. Metrische Merkmale zeigen jedoch eine starke Umweltabhängigkeit. Es interessiert daher besonders, wie hoch der Anteil umweltbedingter Leistungen ist und in welchem Ausmaß die verschiedenen genetischen Komponenten in Wechselwirkungen mit der Umwelt einbezogen sind.

Eine einfache Möglichkeit, diese Einflüsse zu erfassen, ergibt sich aus einer Wiederholung des Versuches unter veränderten Umweltbedingungen.

Bei der Ableitung der Ausgangsformeln gehen wir davon aus, daß es eine selbständige Wirkung der Umwelt ohne das Genom nicht geben kann. Das bedeutet, daß jede umweltbedingte Änderung des Phänotyps als Folge einer Interaktion einer oder mehrerer genetischer Komponenten mit der Umwelt verstanden werden muß.

Führen wir den Versuch unter zwei verschiedenen Umweltbedingungen (1, 2) durch und kennzeichnen die Umweltkomponente mit dem Symbol  $e$ , so erhalten wir für ein bifaktorielles Modell die nachfolgende Ausgangssituation (Gleichung 6):

$$\begin{pmatrix} ij.e & ij.e & ij.e \\ 00.1 & 01.1 & 02.1 \\ 10.1 & 11.1 & 12.1 \\ 20.1 & 21.1 & 22.1 \\ 00.2 & 01.2 & 02.2 \\ 10.2 & 11.2 & 12.2 \\ 20.2 & 21.2 & 22.2 \end{pmatrix} = \begin{pmatrix} i & j & ij & i & j & ij & i & j & ij & i & j & ij & i & j & ij & i & j & ij \\ y-a-a+aa-ye+ae+ae+aae & y-a+d-ad-ye+ae-de+ade & y-a+a-aa-ye+ae-ae+aae & y+d-a-da-ye-de+ae+dae & y+d+d+dd-ye-de-de-dde & y+d+a+da-ye-de-ae-dae & y+a-a-aa-ye-ae+ae+aae & y+a+d+ad-ye-ae-de-ade & y+a+a-aa-ye-ae-ae-aae & y-a-a+aa+ye+ae-ae+aae & y-a+d-ad+ye-ae+de-ade & y-a+a-aa+ye-ae+ae-aae & y+d-a-da+ye+de+ae-dae & y+d+d+dd+ye+de+de-dde & y+d+a+da+ye+de+ae+dae & y+a-a-aa+ye+ae-ae-aae & y+a+d+ad+ye+ae+de+ade & y+a+a-aa+ye+ae+ae+aae \end{pmatrix} \quad (6)$$

Die Ableitung bis zur Diagonalform ergibt die Gleichungen (7).

	$d$	$d$	$a$	$a$	$d$	$a$	$a$	$a$	$e$	$e$	$e$	$e$	$e$	$e$	$e$	$e$	$e$	$(e)$	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2		
$y$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$(i)$	0	0	0	1	1	1	2	2	2	0	0	0	1	1	1	2	2		
									$y$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$d$	$(j)$	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	
18	6	6	2																=	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
	12		4																=	-1	-1	-1	2	2	2	-1	-1	-1	-1	-1	2	2	2	-1	-1		
		12		4															=	-1	2	-1	-1	2	-1	-1	2	-1	-1	2	-1	-1	2	-1	2	-1	
			8																=	1	-2	1	-2	4	-2	1	-2	1	-2	4	-2	1	-2	1	-2	1	
				12	4														=	-1	-1	-1		1	1	1	-1	-1	-1		1	1	1				
					8														=	1	-2	1		-1	2	-1	1	-2	1		-1	2	-1				
						12	4												=	-1		1	-1		1	-1		1	-1		1	-1		1			
							8												=	1		-1	-2		2	1		-1	1		-1	-2		2	1		
								8											=	1		-1		-1	1		-1		-1	1		-1		1			
									18	6	6	2							=	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	
										12		4							=	1	1	1	-2	-2	-2	1	1	1	-1	-1	2	2	2	-1	-1		
											12		4						=	1	-2	1	1	-2	1	1	-2	1	-1	2	-1	-1	2	-1	-1	2	-1
												8							=	-1	2	-1	2	-4	2	-1	2	-1	1	-2	1	-2	4	-2	1	-2	1
													12	4					=	1	1	1		-1	-1	-1	-1	-1		1	1	1					
														8					=	-1	2	-1		1	-2	1	1	-2	1		-1	2	-1				
															12	4			=	1		-1	1		-1	1		-1	-1		1	-1		1			
																8			=	-1		1	2		-2	-1		1	1		-1	-2		2	1		
																	8		=	-1		1			1		-1	1		-1		-1		1			

schen Genetik in der Pflanzenzüchtung. 1. Teil: Die verschiedenen Formen der genetischen Variabilität und ihre Bedeutung in der Pflanzenzüchtung. Züchter 35, 156 bis 174 (1965). — 2. COMSTOCK, R. E., and H. F. ROBINSON: The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics 4, 254–266 (1948). — 3. COMSTOCK, R. E., and H. F. ROBINSON: Estimation of average dominance of genes. In: J. W. GOWEN, Heterosis, p. 494–516. Ames: Iowa State Coll. Press 1952. — 4. CROW, J. F.: Notes on Population Genetics. Mskr. Madison (1962) unveröff. — 5. JINKS, J. L., and

B. I. HAYMAN: The analysis of diallel crosses. Maize Genetics News Letter 27, 48–54 (1953). — 6. JINKS, J. L., and B. I. HAYMAN: The analysis of heritable variation in a diallel cross of *Nicotiana rustica* varieties. Genetics 39, 767–788 (1954). — 7. KEMPTHORNE, O.: The theoretical values of correlations between relatives in random mating populations. Genetics 40, 153–167 (1955). — 8. MATHER, K.: Biometrical Genetics. London: Methuen 1949. — 9. SEYFFERT, W.: Untersuchungen über die Vererbung quantitativer Charaktere an *Arabidopsis thaliana* (L.) Heynh. Z. f. Pflanzenzüchtung 42, 356–401 (1960).

## An Experimental Test of Quantitative Genetic Theory

J. F. KIDWELL<sup>1</sup> and O. KEMPTHORNE

Canada Dept. of Agriculture, Ottawa<sup>2</sup>, and Iowa State University, Ames<sup>3</sup>

**Summary.** An experiment was conducted with *Drosophila melanogaster* to test a theory of quantitative inheritance. The covariances among half and full sibs did not increase with increasing levels of  $F$ . The results pose the question "Why did inbreeding not behave as expected?" Some possible causes include: (1) selective elimination of particular genotypes during inbreeding, (2) inbreeding and/or selection within the reference population, (3) maternal age effects, (4) linkage. Linkage and selective elimination of some genotypes appear most likely to be of greatest importance.

A theory of quantitative inheritance applicable to diploid populations with an arbitrary number of segregating loci and arbitrary epistacy has been described in detail by KEMPTHORNE (1957, Ch. 19). This theory is due to FISHER (1918), WRIGHT (1935), COCKERHAM (1952, 1954), ANDERSON (1953), ANDERSON and KEMPTHORNE (1954), HORNER (1952), and KEMPTHORNE (1954). The formulation considered here does not take account of linkage, although COCKERHAM (1956) has extended it to include linkage in a special case and SCHNELL (1961) and VAN AARDE (1963) have made extensive theoretical investigations. The incorporation of linkage parameters in data interpretation presents considerable difficulties. KEMPTHORNE (1957, Ch. 20) also considered inbreeding in a diploid population with arbitrary epistacy, and no linkage. The formulation provides the basis for a reasonable experimental check on the theory based on absence of linkage, which is one of the main interpretive tools of much plant and animal breeding. The results of such an experiment are presented herein.

### The Theory and Test

The relevant essential features of the theory are briefly outlined for convenience. The original sources must be consulted for complete development.

The theoretical expectation for the partition of the total genetic variance, ( $\sigma_G^2$ ) is:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2 + \sigma_{AAA}^2 + \sigma_{AAD}^2 + \dots + \text{etc.},$$

where:

$\sigma_A^2$  is the variance due to average effects of genes.

$\sigma_D^2$  is the variance due to interactions among allelic gene effects (dominance).

$\sigma_{AA}^2$ ,  $\sigma_{AAA}^2$ , etc. is the variance due to interactions among average effects of two or more non-allelic genes (epistacy).

$\sigma_{DD}^2$ ,  $\sigma_{DDD}^2$  etc. is the variance due to interactions of two or more non-allelic dominance effects (epistacy).

$\sigma_{AD}^2$ ,  $\sigma_{AAD}^2$ , etc. is the variance due to interactions among dominance and average effects (epistacy).

Knowledge of these components is useful for many purposes, e.g., applied breeding studies and problems of evolution. They can be estimated from the covariances among relatives.

In a large panmictic population (specifically, if there is zero probability that the two genes possessed by an individual at any locus are identical by descent) it has been shown that:

$$\text{Cov (H.S.)} = 1/4 \sigma_A^2 + 1/16 \sigma_{AA}^2 + 1/64 \sigma_{AAA}^2 + \dots + \text{etc.}$$

$$\text{Cov (F.S.)} = 1/2 \sigma_A^2 + 1/4 \sigma_D^2 + 1/4 \sigma_{AA}^2 + 1/8 \sigma_{AD}^2 + 1/16 \sigma_{DD}^2 + 1/8 \sigma_{AAA}^2 + \dots + \text{etc.}$$

Covariances between half and full sibs can be estimated by use of a simple and well-known genetic experiment. Each of a number of sires,  $s$  (numbered from  $i = 1$  to  $s$ ) is mated to a random sample of dams ( $j$ ) where  $j$  runs from 1 to  $m_i$ . Each mating of sire  $i$  by dam  $j$  produces  $n_{ij}$  progeny. The observations can be represented by the model

$$y_{ijk} = \mu + s_i + d_{ij} + e_{ijk}$$

in which

$$E(s_i) = E(d_{ij}) = E(e_{ijk}) = 0$$

and the  $s$ ,  $d$ , and  $e$  quantities are uncorrelated and  $E(s_i^2) = \sigma_s^2$ ,  $E(d_{ij}^2) = \sigma_d^2$  and  $E(e_{ijk}^2) = \sigma_e^2$  then

$$E(Y_{ijk} - \mu)(Y_{ijk'} - \mu) = \sigma_s^2 + \sigma_d^2 = \text{Cov (F.S.)}, (k' \neq k)$$

$$E(Y_{ijk} - \mu)(Y_{ij'k} - \mu) = \sigma_s^2 = \text{Cov (H.S.)}, (k' \neq k), (j' \neq j)$$

$$E(Y_{ijk} - \mu)^2 = \sigma_s^2 + \sigma_d^2 + \sigma_e^2 = \sigma_p^2$$

Hence

$$\sigma_e^2 = \sigma_p^2 - \text{Cov (F.S.)},$$

<sup>1</sup> Present address: Division of Biological and Medical Sciences, Brown University, Providence, R.I.

<sup>2</sup> Contribution No. 213, Animal Research Institute, Ottawa, Ontario.

<sup>3</sup> Journal Paper No. 5354 of the Iowa Agricultural and Home Economics Experiment Station, Ames, Project No. 890.